Analysis of publicly available datasets to produce novel findings with clinical relevance

A PhD Thesis for the Degree of

Doctor of Philosophy

in

Computational Biology

School of Biosciences,

University of Kent

Katie-May McLaughlin

2021

Declaration

No part of this thesis has been submitted in support of an application for any degree or other qualification of the University of Kent or any other University or Institute of learning.

Name: Katie-May McLaughlin

Date: 06/07/2021

Abstract

Advances in high-throughput sequencing technologies have facilitated the generation of large-scale genomic and pharmacogenomic databases. Such databases represent an important source of multi-platform data and a critical resource for biomedical research. Moreover, the computational tools available to analyse such 'big data' have evolved substantially in recent years. Here, we have utilised various open-access data resources for our cancer and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)/coronavirus disease 2019 (COVID-19)-related research. Specifically, in our cancer studies, we have correlated expression of genes differentially expressed in response to the phosphorylation status of phosphoprotein enriched in astrocytes 15 (PEA-15) in cisplatin-treated SKOV-3 ovarian cancer cell lines with survival of cisplatin-treated patients. We have also investigated the role of deoxynucleoside triphosphate triphosphohydrolase SAMHD1 in influencing drug sensitivity and cancer patient survival using data from both cell line and clinical studies. In our SARS-CoV-2/COVID-19 studies, we have used structural data to predict the impact of differentially conserved amino acid positions (DCPs) between SARS-CoV and SARS-CoV-2 on the function of SARS-CoV-2 proteins. We have also used transcriptomic and proteomic datasets of SARS-CoV-2-infected cells and patients to identify links between pathways of COVID-19 clinicopathogenesis and deregulation of genes involved in those pathways. Our computational approach demonstrates how publicly accessible data can not only be used to complement *in vitro* investigations, but also to generate novel findings with clinical significance.

Acknowledgements

I would like to thank my supervisors, Mark Wass and Martin Michaelis, for all their support and guidance throughout my PhD. It has been a great privilege to know two such eminent scientists, to work alongside you and to learn from you. I am extremely grateful that you gave me the opportunity to study for a PhD – it is something that I would never have imagined I could possibly do. I cannot thank you both enough for showing me what it means to be a scientist and how to think around complex problems. It has been brilliant to work with people so devoted to their work – your dedication and enthusiasm is a constant inspiration to me.

I would also like to thank all of the members of the Wass/Michaelis lab for their wisdom, encouragement and friendship. I simply could not have got through this without you. I would like to particularly thank Ian Reddin for getting me started with everything, helping me navigate the cancer databases and for being so patient with me!

I would also like to thank Tim Fenton for his kind help with database access and data processing.

I would like to thank my family for their constant support throughout the last four years. It all goes without saying. I love you.

Finally, I would like to thank my dogs. If it weren't for the constant barking, licking, running about, jumping on my computer and demanding cuddles, I don't know how I would have coped. Thank you.

Table of Contents

List of Figures

Figure 1.3. Number of datasets associated with SARS-CoV-2 between February 2020 and October 2020

Figure 2.1. (**A**) Expression of hemagglutinin (HA)-tagged phosphoprotein enriched in astrocytes (PEA-15) in SKOV-3 cells after transfection with the HA-tagged empty vector (EV), PEA-15AA (AA) and PEA-15DD (DD). GAPDH was used as a loading control. (**B**) Cisplatin sensitivity (pEC50, mean ± SEM, n = 8) of transfected SKOV-3-EV (EV), SKOV-3-AA (AA) and SKOV-3-DD (DD) cells. *** $p < 0.001$, n.s. = not significant.

Figure 2.2. A representative Western blot of phosphorylated extracellular signal-regulated kinase1/2 (p-ERK1/2) expression in nuclear and cytosolic fractions of the SKOV-3 cells transfected with empty vector (EV), PEA-15AA (AA) and PEA-15DD (DD). GAPDH and Lamin B1 were used as the markers and loading controls of cytosolic (C) and nuclear fractions (N), respectively.

Figure 2.3. Heatmap of the transcriptome-wide ClariomTM S array, regulated genes with fold change cut-off at 2.0 for differentially expressed genes and a p-value cut-off at 0.05 are shown.

Figure 2.4. Heatmap indicating the relationship between low expression of the indicated genes and sensitivity/outcome, favourable (low cisplatin EC50 in SKOV-3-AA cells or prolonged survival of cisplatin-treated patients, indicated in yellow) or unfavourable (high cisplatin EC50 in SKOV-3-AA cells or reduced survival of cisplatin-treated patients, indicated in blue), based on the comparison of gene expression between SKOV-3-AA and EV- or PEA-15DD-transfected variants and TCGA data.

Figure 2.5. Venn diagram representing the exclusively and commonly regulated genes in different transfected cells upon cisplatin exposure. The diagram shows the total number of genes affected by cisplatin exposure in empty vector—(EV), PEA-15AA—(AA) and PEA-15DD-transfected—(DD) cells.

Figure 2.6. Twenty-one biological pathways significantly affected by cisplatin treatment in SKOV-3-AA cells, listed according to the significance level (log 2 base) in a descending order.

Figure 2.7. Representative Western blots and the corresponding densitometric quantification (mean ± SEM, n = 3) of (**A**) the relative uridine diphosphate-glucuronyl transferase (UGT)1A expression and (**B**) the relative nuclear factor erythroid 2-related factor 2 (Nrf2) expression in empty vector—(EV), PEA-15AA—(AA) and PEA-15DD-transfected—

(DD) cells after treatment with 15 µM cisplatin (+Pt) for 24 h and in untreated transfected SKOV-3 cells. GAPDH was used as a loading control. * $p < 0.05$, ** $p < 0.01$.

Figure 2.8. Representative Western blots and the corresponding densitometric quantification (mean ± SEM, n = 3) of (**A**) the relative Nrf2 expression and (**B**) the relative UGT1A expression in the transfected untreated SKOV-3 cells (Ctrl), after exposure to 15 µM cisplatin (Pt), to 20 µM retinoic acid (RA) and after co-incubation with 20 µM retinoic acid and 15 µM cisplatin (Pt + RA) for 24 h are shown. GAPDH was used as a loading control. *$p < 0.05$, ** $p < 0.01$, n.s. = not significant.

Figure 2.9. Sensitivity of SKOV-3 cells (pEC50, mean ± SEM, n = 9–10) of cisplatin alone (Pt), and upon co-incubation with 20 µM retinoic acid (Pt + RA) was determined over 48 h. *** $p < 0.001$.

Figure S2.1. (**A**) Expression of hemagglutinin (HA)-tagged PEA-15 in EFO27^{r}CDDP^{2000} cells after transfection with the HA-tagged empty vector (EV), PEA-15AA (AA) and PEA-15DD (DD). GAPDH was used as a loading control. (**B**) Cisplatin cytotoxicity (pEC_{50}, mean ± SEM, *n* = 8) in nontransfected EFO27^{r}CDDP^{2000} cells, cells transfected with empty vector (EV), with PEA-15AA (AA), and with PEA-15DD (DD). ***$p < 0.001$, n.s. = not significant.

Figure 3.1. SAMHD1 levels differ between T-ALL and B-ALL.

Comparison of SAMHD1 expression (mRNA abundance) levels in T-ALL and B-ALL cell lines from the CTRP, CCLE, and GDSC (**A**) and in blasts from leukaemia patients (**B**). (**C**) Comparison of the expression of other genes known to affect nucleoside analogue activity based on CTRP data. Respective CCLE and GDSC data are provided in Figure S3.2. *p-values for the comparison B-ALL vs. T-ALL.

Figure 3.2. Comparison of nelarabine (CTRP) and cytarabine (CTRP, GDSC) sensitivity between B-ALL and T-ALL cell lines and correlation of SAMHD1 mRNA levels with the nelarabine and cytarabine sensitivity (expressed as AUC) across all B-ALL and T-ALL cell lines.

Pearson's r values and respective p-values are provided. Respective data on the correlation of SAMHD1 expression with drug sensitivity exclusively for B-ALL and T-ALL cell lines are provided in Figure S3.3 (nelarabine) and Figure S3.4 (cytarabine).

Figure 3.3. SAMHD1 protein and mRNA levels in the RCCL panel of B-ALL and T-ALL cell lines.

(**A**) Representative Western blots indicating protein levels of total SAMHD1 and phosphorylated SAMHD1 (p-SAMHD1). GAPDH was used as loading control. (**B**) Quantitative

SAMHD1 protein levels are shown as means ± SD from three independent experiments (quantified using near-infrared Western blot images to determine the ratio SAMHD1/ GAPDH relative to the positive control THP-1, an acute myeloid leukaemia cell line characterised by high cellular SAMHD1 levels (Schneider et al., 2017). SAMHD1 mRNA abundance levels are shown as means ± SD from three technical replicates (as determined by qPCR, relative to cell line ROS-50) in B-ALL and T-ALL cell lines. Unpaired two-tailed Student's t-tests were used to compare means (represented as horizontal lines ± SEM) of SAMHD1 protein or mRNA levels in B-ALL and T-ALL cells.

Figure 3.4. AraG and cytarabine concentrations that reduce the viability of the RCCL ALL cell lines by 50% (IC50) and correlation of the IC50s with the cellular SAMHD1 protein or mRNA levels.

Numerical data are provided in Supplementary Data 3.4. Closed circles and error bars represent means ± SD of three independent experiments, each performed in three technical replicates. Linear regression analyses were performed using GraphPad Prism.

Figure 3.5. Effect of SAMHD1 on nelarabine and cytarabine sensitivity in ALL and AML cells.

(**A**) Dose-response curves of AraG- and cytarabine-treated ALL cell lines in the absence or presence of Vpx virus-like particles (cause SAMHD1 depletion), or Vpr virus-like particles (negative) controls. Concentrations that reduce ALL cell viability by 50% (IC50s) and Western blots confirming SAMHD1 depletion are provided. Each symbol represents the mean ± SD of three technical replicates of one representative experiment out of three. (**B**) Effects of AraG and cytarabine on AML cell lines in the absence or presence of functional SAMHD1. In the SAMHD1-expressing AML cell line THP-1, the SAMHD1 gene was disrupted by CRISPR/Cas9 (THP1-KO). The non-SAMHD1 expressing AML cell line HEL and the non-SAMHD1 expressing ALL cell line JURKAT were transduced with wild-type SAMHD1 (SAMHD1_WT) or the triphosphohydrolase-defective SAMHD1 mutant D311A (SAMHD1_D311A). Dose-response curves, drug concentrations that reduce cell viability by 50% (IC50s), and Western blots confirming SAMHD1 protein levels are provided. Each symbol represents the mean ± SD of three independent experiments, each performed in three technical replicates.

Figure 3.6. SAMHD1 promoter methylation in ALL cell lines.

(**A**) Analysis of SAMHD1 promoter methylation in the RCCL cell line panel through amplification of a single PCR product (993-bp) corresponding to the promoter sequence after HpaII digestion. A 0.25-kb fragment of the GAPDH gene lacking HpaII sites was PCR-amplified

using the same template DNA served as loading control. (**B**) GDSC data indicating SAMHD1 promoter methylation in B-ALL and T-ALL cell lines and correlation of SAMHD1 promoter methylation and SAMHD1 expression across all ALL cell lines. (**C**) GDSC data indicating the level of global methylation in B-ALL and T-ALL cell lines.

Figure S3.1. Gene expression profiles and nelarabine sensitivity in acute lymphoblastic leukaemia (ALL) cell lines. (**A**) Nelarabine sensitivity expressed as area under the curve (AUC) in T-cell precursor ALL (T-ALL) and B-ALL cell lines from CTRP. (**B**) Heatmap illustrating expression patterns of genes differentially regulated between T-ALL and B-ALL cell lines based on CTRP data. Heatmaps displaying the expression of all genes in the CTRP ALL cell lines and those displaying gene expression in the ALL cell lines in the CCLE and GDSC datasets are provided in Figure S3.1D-H. Individual gene expression values are presented in Supplementary Data 1. (**D**) Heatmap illustrating expression (mRNA abundance) of all genes in B- vs. T-ALL cells based on CTRP data. (**E**) Heatmap illustrating expression (mRNA abundance) of differentially regulated genes in B- vs. T-ALL cells based on CCLE data. (**F**) Heatmap illustrating expression (mRNA abundance) of all genes in B- vs. T-ALL cells based on CCLE data. (**G**) Heatmap illustrating expression (mRNA abundance) of differentially regulated genes in B- vs. T-ALL cells based on GDSC data. (**H**) Heatmap illustrating expression (mRNA abundance) of all genes in B- vs. T-ALL cells based on GDSC data.

Figure S3.2A. SAMHD1 expression in ALL patients with different immunophenotypes. P values for comparisons between individual groups.

Figure S3.2B. SAMHD1 expression in ALL patients with different genotypes. P values for comparisons between individual groups (1=B-other, 2=Ph pos, 3=Ph-like, 4=KMT2A, 5=T-ALL, 6=Burkitt, 7=TCF3, 8=ETV6, 9=Hyperdip.

Figure S3.3. Expression of genes known to be potentially involved in nucleoside analogue activity in B-ALL and T-ALL cell lines in the CCLE and GDSC.

Figure S3.3. Expression of genes known to be potentially involved in nucleoside analogue activity in patient-derived B-ALL and T-ALL cells.

Figure S3.4. Correlation of the expression of genes (mRNA abundance) known to affect nucleoside analogue activity to the nelarabine sensitivity (expressed as AUC) across all ALL, the B-ALL and the T-ALL cell lines based on CTRP data. Pearson's r values and respective p-values are provided.

Figure S3.5. Comparison of SAMHD1 expression (mRNA abundance) levels in acute myeloid leukaemia (AML), B-cell acute lymphoblastic leukaemia (B-ALL), T-cell acute lymphoblastic leukaemia (T-ALL) cells in CTRP, CCLE, and GDSC. Respective p-values are provided.

Figure S3.6. Correlations of SAMHD1 expression (mRNA abundance) with the cytarabine AUC exclusively in B-ALL and T-ALL cell lines based on CTRP and GDSC data. Pearson's r values and respective p-values are provided.

Figure S3.7. SAMHD1 protein levels in the RCCL panel of B-ALL and T-ALL cell lines. Representative Western blots indicating protein levels of total SAMHD1 and GAPDH in 23 cell lines of the RCCL panel, which were run on the same gel and blotted on the same membrane to confirm the representativeness of the blots provided in Figure 3.3A. Figure 3.3A is provided for comparison.

Figure S3.8. Correlations of SAMHD1 protein and mRNA levels determined in the RCCL cell lines with the SAMHD1 expression data derived from the CTRP, CCLE, and GDSC among the cell lines that are represented in both respective datasets. Pearson's r values and respective p-values are provided.

Figure S3.9. Correlations of the nelarabine AUCs derived from the CTRP and the AraG IC50 values determined in the RCCL panel across the ALL cell lines present in both datasets. Pearson's r values and respective p-values are provided.

Figure S3.10. Uncropped Western blots and agarose gels.

Figure 4.1. Effect of *SAMHD1* expression in cancer patients. (**A**) Kaplan Meier plots indicating survival in cancer patients with tumours characterised by high or low *SAMHD1* expression (as determined by best separation) across all patients in the TCGA and TARGET databases. P-values were determined by log-rank test. (**B**) Pie charts indicating the number of cancer types for which high SAMHD1 expression was associated with increased survival, reduced survival, or not significantly associated with survival based on data from the TCGA and TARGET databases. Data are presented in Table S4.1 and Table S4.2.

Figure 4.2. *SAMHD1* expression levels and 5-year survival rates in dependence of sex based on TCGA data. (**A**) Kaplan Meier plots indicating sex-specific survival in cancer patients with tumours characterised by high or low *SAMHD1* expression (as determined by best separation). P-values were determined by log-rank test. (**B**) Heatmap indicating the association of *SAMHD1* expression and 5-year survival rates (blue: high SAMHD1 associated with higher survival rates, yellow: low SAMHD1 associated with higher survival rates). (**C**)

Heatmap indicating cancer entities in which high *SAMHD1* expression (blue) or low *SAMHD1* expression (yellow) is significantly ($p<0.05$) associated with higher 5-year survival rates. (**D**) Cancer entities in which SAMHD1 displays a trend towards differing roles by sex. Blue indicates higher survival rates in patients with tumours with high SAMHD1 levels, yellow in patients with low SAMHD1 levels. Abbreviations: ACC, Adrenocortical carcinoma; BLCA, Bladder Urothelial Carcinoma; BRCA, Breast invasive carcinoma; CESC, Cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, Cholangiocarcinoma; COAD, Colon adenocarcinoma; DLBC, Lymphoid Neoplasm Diffuse Large B-cell Lymphoma; ESCA, Oesophageal carcinoma; GBM, Glioblastoma multiforme; HNSC, Head and Neck squamous cell carcinoma; KICH, Kidney Chromophobe; KIRC, Kidney renal clear cell carcinoma; KIRP, Kidney renal papillary cell carcinoma; LAML, Acute Myeloid Leukaemia; LGG, Low Grade Glioma; LIHC, Liver hepatocellular carcinoma; LUAD, Lung adenocarcinoma; LUSC, Lung squamous cell carcinoma; MESO, Mesothelioma; OV, Ovarian serous cystadenocarcinoma; PAAD, Pancreatic adenocarcinoma; PCPG, Pheochromocytoma and Paraganglioma; PRAD, Prostate adenocarcinoma; READ, Rectum adenocarcinoma; SARC, Sarcoma; SKCM, Skin Cutaneous Melanoma; STAD, Stomach adenocarcinoma; TGCT, Testicular Germ Cell Tumours; THCA, Thyroid carcinoma; THYM, Thymoma; UCEC, Uterine Corpus Endometrial Carcinoma; UCS, Uterine Carcinosarcoma; UVM, Uveal Melanoma.

Figure 4.3. *SAMHD1* expression levels and 5-year survival rates in dependence of sex based on TARGET data. (**A**) Kaplan Meier plots indicating sex-specific survival in cancer patients with tumours characterised by high or low *SAMHD1* expression (as determined by best separation). P-values were determined by log-rank test. (**B**) Heatmap indicating the association of *SAMHD1* expression and 5-year survival rates (blue: high SAMHD1 associated with higher survival rates, yellow: low SAMHD1 associated with higher survival rates). (**C**) Heatmap indicating cancer entities in which high *SAMHD1* expression (blue) or low *SAMHD1* expression (yellow) is significantly ($p<0.05$) associated with higher 5-year survival rates.

Figure 4.4. Kaplan-Meier plots indicating survival in cancer patients of different race with tumours characterised by high or low *SAMHD1* expression (as determined by best separation) based on TCGA data. P-values were determined by log-rank test.

Stratifying of patients in the TARGET database according to race provided some trends, which may point towards differences, but the numbers are too low to draw firm

conclusions (Figure 4.5, Figure S4.5, Table S4.6). No significant differences were detected between the SAMHD1 levels in the different race groups (Figure S4.6).

Figure 4.5. Kaplan-Meier plots indicating survival in cancer patients of different race with tumours characterised by high or low *SAMHD1* expression (as determined by best separation) based on TARGET data. P-values were determined by log-rank test.

Figure 4.6. Tumour suppressor and oncogenic effects of SAMHD1 in different cancer types, as suggested by SAMHD1 levels in tumour tissues vs. matched normal tissues (Tumour vs. control) or the comparison of 5-year survival in patients with SAMHD1 high or low tumours (SAMHD1 high vs. low). Higher SAMHD1 levels in matched normal tissues were interpreted as tumour suppressor activity, while higher SAMHD1 levels in tumour tissues as indication of oncogenic effects. Higher 5-year survival in patients with SAMHD1 high tumours was construed as sign of tumour suppressor activity, higher 5-year survival in patients with SAMHD1 low tumours indication of oncogenic effects. (**A**) Data for all available comparisons. (**B**) Data for entities, in which at least the difference for one comparison reached statistical significance. (**C**) Data for entities, in which the difference for both comparisons reached statistical significance.

Figure 4.7. Inverse correlation between SAMHD1 promotor methylation levels or miRNA levels and SAMHD1 expression based on TCGA data. (**A**) Correlation between SAMHD1 promotor methylation levels and SAMHD1 expression across all patients and in THYM patients, which displayed the strongest inverse correlation across all cancer types. Data for all cancer types are presented in Table S4.8. (**B**) Correlation of mir-23b with SAMHD1 expression across all patients and of mir-30c-1 with SAMHD1 in THYM. mir-23b was the miRNA that displayed the strongest inverse correlation with SAMHD1 across all patients. The inverse correlation between mir-30c-1 and SAMHD1 was the strongest among all miRNAs in all cancer types. Data for all significant inverse correlations of miRNAs and SAMHD1 across all cancer types are provided in Table S4.10.

Figure 4.8. *SAMHD1* mutations in cancer tissues. (**A**) *SAMHD1* was mutated 201 times, including 175 non-synonymous mutations, which puts *SAMHD1* within the 15.3% of most commonly mutated genes. (**B**) Survival in patients with and without *SAMHD1* mutant tumours. (**C**) Survival in UCEC (cancer type with the most *SAMHD1* mutations) patients with and without *SAMHD1* mutant tumours. (**D**) Survival in patients with tumours with or without mutations in *TP53*, the most commonly mutated tumour suppressor genes. (**E**) Lollipop plot indicating locations of missense mutations in *SAMHD1*. Residues predicted to

be involved in ligand binding are labelled in bold. (**F**) nonsynonymous mutations mapped (coloured red) onto the SAMHD1 protein structure (Protein Data Bank identifier 6DWD (Knecht et al., 2018) with bound clofarabine hydrochloride (indicated in cyan) and magnesium ion (green). The image on the left shows the full structure and on the right the active site is displayed. Yellow dashed lines indicate hydrogen bonds between mutated residues and ligand.

Figure S4.1. Comparison of SAMHD1 expression levels as expressed by fragments per kilobase of transcript per million mapped reads (FPKM) after removal of sex-specific cancer types and of BRCA (breast invasive carcinoma), for which only a very small fraction of samples (12/ 1,089 tumour tissue samples, 1/ 113 matched normal tissue samples) was derived from males. P-values were determined by Mann-Whitney U (Wilcoxon) test for independent groups.

Figure S4.1. Comparison of *SAMHD1* expression levels as expressed by fragments per kilobase of transcript per million mapped reads (FPKM) in tumours from the TARGET database. P-values determined by Mann-Whitney U (Wilcoxon) test for independent groups.

Figure S4.3A. Heatmap indicating the association of SAMHD1 expression and 5-year survival rates (blue: high SAMHD1 associated with higher survival rates, yellow: low SAMHD1 associated with higher survival rates).

Figure S4.3B. Heatmap indicating cancer entities in which high SAMHD1 expression (blue) or low SAMHD1 expression (yellow) is significantly ($p<0.05$) associated with higher 5-year survival rates.

Figure S4.4. SAMHD1 levels in patients of different race in TCGA. Comparison of all groups was performed using the Kruskal-Wallis test. Individual group comparisons were performed using the Mann-Whitney U (Wilcoxon) rank sum test.

Figure S4.3. Heatmap indicating the association of SAMHD1 expression and 5-year survival rates (blue: high SAMHD1 associated with higher survival rates, yellow: low SAMHD1 associated with higher survival rates). (**A**) All cancer types independently of significance level. (**B**) Cancer types in which at least comparison resulted in a significant ($p<0.05$) difference.

Figure S4.6. SAMHD1 levels in patients of different race in the TARGET database. Comparison of all groups was performed using the Kruskal-Wallis test. Individual group comparisons were performed using the Mann-Whitney U (Wilcoxon) rank sum test.

Figure S4.7. Summary of the findings of the literature review for articles containing data on the role of SAMHD1 during oncogenesis. Articles were identified by using the search term "(((Cancer) OR (tumor) OR (tumour))) AND (SAMHD1)" in PubMed (https://pubmed.ncbi.nlm.nih.gov) on 17th June 2021.

Figure 5.1. SARS-CoV-2 and SARS-CoV replication in cell culture. (**A**) Cytopathogenic effect (CPE) formation 48 h post-infection in MOI 0.01-infected Caco2, CL14, DLD-1 and HT29 cells. Representative images showing immunostaining for double-stranded RNA (indicates virus replication) and quantification of virus genomes by qPCR are presented in Figure S5.3. (**B**) CPE formation in SARS-CoV and SARS-CoV-2 (MOI 0.01)-infected ACE2-negative 293 cells and 293 cells stably expressing ACE2 cells (293/ACE2) 48 h post-infection. Immunostaining for double-stranded RNA and quantification of virus genomes by qPCR is shown in Supplementary Figure S5.4. (**C**) Western blots indicating cellular ACE2 and TMPRSS2 protein levels in uninfected cells. Uncropped blots are provided in Figure S5.6. (**D**) A sequence view of the DCPs in the vicinity of the S two cleavage sites and an image of the R815 cleavage site and closely located DCPs. S is cleaved and activated by TMPRSS2. (**E**) Concentration-dependent effects of the TMPRSS2 inhibitors camostat and nafamostat on SARS-CoV-2- and SARS-CoV-induced cytopathogenic effect (CPE) formation determined 48 h post-infection in Caco2 infected at an MOI of 0.01 using a phase contrast microscope. Similar effects were observed in CL14 cells (Figure S4.6). Values are presented as means ± S.D. (n = 3)

Figure 5.2. SARS-CoV-2 and SARS-CoV S interaction with ACE2. (A–D)Differentially conserved positions in the Spike protein. (**A**) A sequence view of the DCPs present in the Spike protein, with an inset showing the receptor binding domain. (**B**) The S interface with ACE2 (cyan). The ACE2 interface is shown in blue spheres, DCPs in red. (**C**) The V404 = K417 DCP. (**D**) The R426 = N439 DCP, the left image shows SARS-CoV S R426, the image on the right show the equivalent N439 in SARS-CoV-2 S. (**E**) SARS-CoV residues associated with altering ACE2 affinity and the residues at these positions in SARS-CoV-2 S. (**F**) Cytopathogenic effect (CPE) formation in SARS-CoV-2 and SARS-CoV (MOI 0.01)-infected Caco2 cells in the presence of antibodies directed against ACE2 or DPP4 (MERS-CoV receptor) 48 h post-infection

Figure S5.1. The BLOSUM scores for the amino acid substitutions present in the SDPs. A graph is plotted that combines all of the proteins and one for each of the individual proteins that were analysed.

Figure S5.2. Overview of modelled DCPs. DCPs with likely functional effect are indicated by arrows and labelled. Structural model shown is indicated in brackets. DCPs likely to have an effect are coloured red; DCPs with a possible effect are shown in orange; and DCPs unlikely to have an effect are coloured yellow. Please refer to table S7 for full details of structural analysis of each DCP.

Figure S5.3. SARS-CoV-2 and SARS-CoV susceptibility of cell lines. (**A**) Representative images showing MOI 0.01-infected cells immunostained for double-stranded RNA 48h post infection. (**B**) Quantification of virus genomes by qPCR at different time points post infection (p.i.). Values are presented as means ± S.D. (n =3).

Figure S5.4. SARS-CoV-2 and SARS-CoV replication in 293 cells stably expressing ACE2 cells (293/ACE2). (**A**) Immunostaining for double-stranded RNA (indicating virus replication) in SARS-CoV-2 and SARS-CoV (MOI 0.01)-infected 293/ACE2 cells 48h post infection. (**B**) Quantification of virus genomes by qPCR in SARS-CoV-2 and SARS-CoV (MOI 0.01)-infected 293/ACE2 cells 48h post infection. Values are presented as means ± S.D. (n =3).

Figure S5.5. Uncropped Western blots for Figure 5.1C. 293/ACE2 cells served as positive control for ACE2. * Protein quantification

Figure S5.6. Role of TMPRSS2-mediated S cleavage in SARS-CoV-2 and SARS-Co-V replication. Concentration-dependent effects of the TMPRSS2 inhibitors camostat and nafamostat on SARS-CoV-2- and SARS-CoV-induced cytopathogenic effect (CPE) formation determined 48h post infection in CL14 cells infected at an MOI of 0.01. Values are presented as means ± S.D. (n =3).

Figure 6.1. SARS-CoV-2 infection is associated with increased CD47 levels. (**A**) CD47 protein abundance in uninfected (control) and SARS-CoV-2-infected (virus) Caco-2 cells (data derived from (Bojkova, Klann et al., 2020). P-values were determined by two-sided Student's t-test. (**B**) CD47 and SARS-CoV-2 N protein levels and virus titers (genomic RNA determined by PCR) in SARS-CoV-2 strain FFM7 (MOI 1)-infected air-liquid interface cultures of primary human bronchial epithelial (HBE) cells and SARS-CoV-2 strain FFM7 (MOI 0.1)-infected Calu-3 cells. Uncropped blots are provided in Figure S6.1. (**C**) CD47 mRNA levels in post mortem samples

from COVID-19 patients (data derived from (Blanco-Melo et al., 2020)). P-values were determined by two-sided Student's t-test.

Figure 6.2. SARS-CoV-2 infection increases SIRPα in primary human monocytes. (**A**) SARS-CoV-2 strain FFM7 (MOI 1) infection of primary human monocytes does not result in the production of genomic viral RNA as detected by PCR. (**B**) SARS-CoV-2 strain FFM7 (MOI 1)-infected primary human monocytes display enhanced SIRPα levels. Uncropped blots are provided in Figure S6.4.

Figure 6.3. Results of the PubMed (https://pubmed.ncbi.nlm.nih.gov) literature search for "CD47 aging" (**A**) and "CD47 hypertension" (**B**). (**C**) Overview figure of the data derived from the literature searches. Age-related increased CD47 levels may contribute to pathogenic conditions associated with severe COVID-19.

Figure 6.4. Results of the PubMed (https://pubmed.ncbi.nlm.nih.gov) literature search for "CD47 diabetes" (**A**). (**B**) Overview figure of the data derived from the literature search. Hyperglycemia- and diabetes-induced increased CD47 levels may contribute to immune escape of SARS-CoV-2-infected cells.

Figure S6.1. Uncropped Western blots to Figure 6.1. Bands are indicated by frames. Numbers indicate quantification results.

Figure S6.2. CD47 mRNA levels in SARS-CoV-2-infected Calu-3 cells (data derived from (Blanco-Melo et al., 2020)). P-values were determined by two-sided Student's t-test.

Figure S6.3. CD47 levels in SARS-CoV-2 (MOI 0.1)-infected Caco2 cells as determined by flow cytometry (FACSCanto II, BD Biosciences). Cells were fixed with 4% formaldehyde (10 minutes) and then stained for CD47 using a PE-labelled CD47 antibody (Miltenyi, # 130-123-754, 1:50 dilution). Isotype REA Control Antibody (S) (human IgG1, PE-labelled, Miltenyi, # 130-113-438, 1:50 dilution) was used as control.

Figure S6.4. Uncropped Western blots to Figure 6.2.

Figure 7.1. Gene products anticipated to be of potential relevance for COVID-19-related coagulopathy, based on genes with a role in coagulation that are differentially expressed between females and males (Table S7.1) and whose expression correlates with age (Table S7.2). Candidate gene products were either (**A**) procoagulants (ADAMTS13, F11, HGFAC, KLKB1), which display higher expression in males than in females and increase with age, or (**B**) anticoagulants (C1QTNF1, SERPINA5), which display higher expression in females than in

males and decrease with age. A complete list of the relevant genes overlapping between Table S1 and Table S2 is presented in Table S3. p-values were determined by two-sided Student's t-test. * p-value < 0.05.

Figure 7.2. SARS-CoV-2-induced (derived from (Bojkova et al., 2020)) transferrin (TF) expression and age- and gender-specific expression of the procoagulant TF and its antagonist SERPINC1/antithrombin based on GTEx data. Data are presented as violin blots to indicate the distribution of individual values and as bar charts to facilitate comparisons. TF displayed higher levels in SARS-CoV-2-infected cells than in non-infected cells. Moreover, TF expression and the expression ration of TF and SERPINC1 increased with age and were higher in females than in males. (**A**) TF protein abundance in uninfected (control) and SARS-CoV-2-infected (virus) Caco-2 cells. p-values are the result of a two-sided Student's t-test. (**B**) TF expression (TPM) in females and males across six age groups. p-values were calculated using the Wilcoxon rank sum test for independent groups. (**C**) SERPINC1 expression (TPM) in females and males across six age groups. p-values were calculated using the Wilcoxon rank sum test for independent groups. (D) Ratio of TF/SERPINC1 expression (TPM) in females and males across six age groups. p-values were calculated using the Wilcoxon rank sum test for independent groups. * p-value < 0.05.

Figure 8.1. Anti-severe acute respiratory syndrome virus 2 (SARS-CoV-2) effects of aprotinin and SERPINA1/alpha-1 antitrypsin. (**A**) Concentration-dependent effects of aprotinin and SERPINA1/alpha-1 antitrypsin on SARS-CoV-2-induced cytopathogenic effect (CPE) formation determined 48 h post-infection in Caco2 cells infected at a multiplicity of infection (MOI) of 0.01 with the three different SARS-CoV-2 isolates. The viability of the Caco2 cells was 84.3 ± 2.7% relative to the untreated control in the presence of 20 µM of aprotinin. (**B**) Immunostaining for the SARS-CoV-2 S protein in aprotinin- and SERPINA1/alpha-1 antitrypsin-treated Caco2 cells infected at an MOI of 0.01 with the three different SARS-CoV-2 isolates as determined 48 h post-infection. The protease inhibitors were tested at four concentrations in 1:4 dilution steps ranging from 20 to 0.3125 µM. A quantification is provided in Figure S8.1. (**C**) Copy numbers of genomic RNA in Caco2 cells infected with different SARS-CoV-2 isolates (MOI of 0.01) in response to treatment with aprotinin or SERPINA1/alpha-1 antitrypsin as determined 48 h post-infection. FFM1, 1/Human/2020/Frankfurt; FFM2, 2/Human/2020/Frankfurt; FFM6, 6/Human/2020/Frankfurt.

Figure 8.2. Effects of aprotinin on SARS-CoV-2-induced caspase 3/7 activation. Caspase 3/7 activity was determined in Caco2 cells infected with different SARS-CoV-2 isolates (MOI of 0.01) 48 h post-infection.

Figure 8.3. Anti-SARS-CoV-2 effects of aprotinin when administered post-infection. For post-infection experiments, the cells were incubated with the virus for a one-hour adsorption period. Then, the cells were washed three times in PBS prior to the addition of the drug. (**A**) The effects of aprotinin and the RNA polymerase inhibitor remdesivir (a positive control drug that interferes with virus replication after virus entry) on virus replication as determined by qPCR in SARS-CoV-2/FFM1 (MOI of 0.1)-infected Caco2 cells 8 h post-infection (after approximately one round of replication). * $p < 0.05$ as determined by one-way ANOVA and Dunnett's multiple comparison test. (**B**) The effects of aprotinin on cytopathogenic effect (CPE) formation in SARS-CoV-2/FFM1 (MOI of 0.01)-infected Caco2 cells were determined 48 h post-infection.

Figure 8.4. Regulation of host cell protease inhibitors in SARS-CoV-2-infected cells. (**A**) Total protein levels based on a publicly available proteomics dataset (Bojkova et al., 2020), indicating cellular levels of endogenous protease inhibitors in SARS-CoV-2 (MOI of 1)-infected Caco2 cells 2 h and 24 h post-infection. Data were normalized using summed intensity normalization for sample loading, followed by internal reference scaling and trimmed mean of M normalization. * p-values as determined using a two-sided Student's t-test. (**B**) Mean protein translation of endogenous protease inhibitors in arbitrary units (AU) (normalized and corrected summed peptide spectrum matches (PSMs) were averaged) in SARS-CoV-2 (MOI of 1)-infected Caco2 cells 2 h and 24 h post-infection based on a publicly available translatome dataset (Bojkova et al., 2020). * p-values as determined using a two-sided Student's t-test.

Figure 8.5. Antiviral effects of aprotinin in SARS-CoV-2-infected air–liquid interface (ALI) cultures from primary bronchial epithelial cells. (**A**) Abundance of the SARS-CoV-2 proteins N (nucleocapsid) and M (membrane) in primary bronchial epithelial cell ALI cultures infected with SARS-CoV-2/FFM1 (MOI of 1) in the presence or absence of aprotinin (20 µM) as determined 5 days post-infection by multiplexed mass spectrometry analysis using acquisition targeting of previously identified viral peptides modified with TMTpro. The detailed data are presented in Table S1. (**B**) Western blots indicating cellular SARS-CoV-2 N and TMPRSS2 levels in primary bronchial epithelial cell ALI cultures infected with SARS-CoV-2/7/Human/2020/Frankfurt (FFM7) (MOI of 1) in the presence or absence of aprotinin as

detected 5 days post infection. GAPDH was served as the loading control. Uncropped Western blots are shown in Figure S5.

Figure S8.1. Quantification of immunostaining for the spike protein in SARS-CoV-2-infected (isolates FFM1, FFM2, FFM6) Caco2 cells with and without treatment of aprotinin or SERPINA1/ alpha-1 antitrypsin (prolastin) presented in Figure 8.2B.

Figure S8.2. Trypsin inhibition by aprotinin and SERPINA1/ alpha-1 antitrypsin. Nearly confluent Caco2 cell cultures were washed three times with PBS and incubated with 400µg/ mL trypsin alone or in combination with aprotinin 20µM or SERPINA1 20µM for 2h.

Figure S8.2. Trypsin inhibition by aprotinin and SERPINA1/ alpha-1 antitrypsin. Nearly confluent A549 cell cultures were washed three times with PBS and incubated with 400µg/ mL trypsin alone or in combination with aprotinin 20µM or SERPINA1 20µM for 2h.

Figure S8.3. Quantification of immunostaining for the spike protein in SARS-CoV-2/FFM7 (MOI 0.01)-infected Calu-3 cells 48h post infection in response to aprotinin treatment.

Figure S8.4. Uncropped Western blots corresponding to Figure 8.4B. Quantification was performed by laser-induced fluorescence using an infrared scanner (Odyssey, Li-Cor Biosciences) and Image Studio version 3.1 software.

List of Tables

Table S5.1. Protein structures used for structural analysis obtained from the Protein Data Bank.

Table S5.2. Structural models generated by Phyre2 and used for structural analysis. Where structures were not available from the Protein Data Bank, the structures were modelled.

Table S5.3. Criteria used for classifying proposed effect on protein structure and function within the structural analysis.

Table S5.4. Specificity Determining Positions (DCPs) identified between SARS-CoV and SARS-CoV-2.

Table S5.5. Analysis of DCPs present in the SARS-CoV and SARS-CoV-2 Spike protein interface with human ACE2.

Table S5.6. Structural analysis of DCPs in SARS-CoV and SARS-CoV-2

Table 6.1. Evidence supporting a link between aging and/ or hypertension and increased CD47 levels.

Table 6.2. Evidence supporting a link between diabetes and increased CD47 levels.

Table S6.1. Literature search for 'CD47 aging'. Literature search performed using PubMed (https://pubmed.ncbi.nlm.nih.gov) on 17th February 2021

Table S6.2 Literature search for 'CD47 hypertension'. Literature search performed using PubMed (https://pubmed.ncbi.nlm.nih.gov) on 18th February 2021

Table S6.3. Literature search for 'CD47 diabetes'. Literature search performed using PubMed (https://pubmed.ncbi.nlm.nih.gov) on 19th February 2021

Table S6.4. Literature search for 'CD47 obesity'. Literature search performed using PubMed (https://pubmed.ncbi.nlm.nih.gov) on 22nd February 2021

Table 7.1. Candidate gene products that may be involved in COVID-19-related coagulopathy. Candidates were either (**A**) procoagulants (ADAMTS13, F11, HGFAC, KLKB1), which display a lower expression in females than in males and increase with age, or (**B**) anticoagulants (C1QTNF1, SERPINA5), which display higher expression in females than in males and decrease with age.

Table S7.1. Genes associated with the GO term "Blood Coagulation" (GO:0007596) in males vs. females

Table S7.2. Correlation of the expression of coagulation-associated genes with age

Table S7.3. Genes with relevant functions whose expression levels correlate with age and differ between males and females

Table S7.4. Genes differentially regulated in SARS-CoV-2-infected Caco2 cells

Table S7.5. Interaction partners of SARS-CoV-2 proteins.

Table 8.1. Aprotinin concentrations that reduce SARS-CoV-2-induced cytopathogenic effect (CPE) formation, SARS-CoV-2 spike (S) levels, and SARS-CoV-2-induced caspase 3/7 activation by 50% (IC50) as determined in Caco2 cells infected with different SARS-CoV-2 isolates (MOI of 0.01) 48 h post-infection.

Table S8.1. Detailed mass spectrometry data

List of Supplementary Data Files

Supplementary Data 3.1. Gene transcripts differentially regulated (mRNA abundance) between T-ALL and B-ALL cell lines based on data derived from GDSC, CCLE, or CTRP.

Supplementary Data 3.2. Pathway analysis using the PANTHER database to identify differentially regulated processes based on genes differentially regulated between B-ALL and T-ALL cell lines in the CTRP, CCLE, and GDSC.

Supplementary Data 3.3. Genes whose expression is directly or inversely correlated with the nelarabine AUC based on CTRP data.

Supplementary Data 3.4. Genes whose expression is directly or inversely correlated with the nelarabine AUC based on CTRP data.

Supplementary Data 3.5. Data underlying graphs.

Table S4.1. 5-year survival in cancer patients with tumours displaying high or or low SAMHD1 levels as determined by best separation of data derived from TCGA.

Table S4.2. 5-year survival in cancer patients with tumours displaying high or low SAMHD1 levels as determined by best separation of data derived from TARGET.

Table S4.3. 5-year survival in cancer patients divided by sex with tumours displaying high or low SAMHD1 levels as determined by best separation of data derived from TCGA.

Table S4.4. 5-year survival in cancer patients divided by sex with tumours displaying high or low SAMHD1 levels as determined by best separation of data derived from TARGET.

Table S4.5. 5-year survival in cancer patients divided by race with tumours displaying high or low SAMHD1 levels as determined by best separation of data derived from TCGA.

Table S4.6. 5-year survival in cancer patients divided by race with tumours displaying high or low SAMHD1 levels as determined by best separation of data derived from TARGET.

Table S4.7. Median SAMHD1 levels (FPKM, fragments per kilobase of transcript per million mapped reads) in tumour and control tissues

Table S4.8. Correlation between *SAMHD1* expression levels and SAMHD1 promotor methylation in different cancer types based on data derived from TCGA.

Table S4.9. miRNAs documented to regulate SAMHD1.

Table S4.10. Correlation of miRNAs and *SAMHD1* expression in different cancer types.

Table S4.11. TCGA mutation data.

Table S4.12. 5-year survival in patients with SAMHD1 mutant or wild-type tumours based on TCGA data.

Table S4.13. Literature search performed in PubMed (https://pubmed.ncbi.nlm.nih.gov) on 17th June 2021 using the search term "(((Cancer) OR (tumor) OR (tumour))) AND (SAMHD1)".

List of Abbreviations

3CL^{pro}	3C-Like Protease
ABCB1	ATP Binding Cassette Subfamily B Member 1
ABL	Abelson Murine Leukaemia
ACC	Adrenocortical Carcinoma
ACE2	Angiotensin-Converting Enzyme 2
AKT3	AKT Serine/Threonine Kinase 3
ALK	Anaplastic Lymphoma Kinase
ALL	Acute Lymphoblastic Leukaemia

AML	Acute Myeloid Leukaemia
AraG	Arabinosylguanine
ASNS	Asparagine Synthetase
AUC	Area Under the Curve
AVPdb	Antiviral Peptide Database
BALF	Bronchoalveolar Lavage Fluid
B-ALL	B-Acute Lymphoblastic Leukaemia
Bat-SL	Bat SARS-Like
BCL2	B-Cell Lymphoma 2
BCR	Breakpoint Cluster Region
BET	Bromodomain and Extraterminal Domain
BLAST	Basic Local Alignment Search Tool
BLCA	Bladder Urothelial Carcinoma
BLOSUM	Blocks Substitution Matrix
BLVRB	Biliverdin Reductase B
BRAF	B-Raf Proto-Oncogene, Serine/Threonine Kinase
BRCA	Breast Invasive Carcinoma
BRD4	Bromodomain-Containing Protein 4
Cas9	CRISPR-Associated Protein 9
CCLE	Cancer Cell Line Encyclopaedia
CD147	Cluster of Differentiation 147
CD47	Cluster of Differentiation 47
CESC	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma
CFE	Cancer Functional Event
CHOL	Cholangiocarcinoma
COAD	Colon Adenocarcinoma
COCA	Cluster-Of-Cluster-Assignments
COVID-19	Coronavirus Disease 2019
CPE	Cytopathogenic Effect
CRC	Colorectal Cancer
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CRP	C-Reactive Protein
CtIP	C-Terminal Binding Protein-Interacting Protein
CTNNB1	Catenin Beta 1

CTRP	Cancer Therapeutics Response Portal
CTSL	Cathepsin L
DCP	Differentially Conserved Position
DIC	Disseminated Intravascular Coagulation
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleoside Triphosphate
EGFR	Epidermal Growth Factor Receptor
EMBOSS	European Molecular Biology Open Software Suite
E protein	Envelope protein
ER	Endoplasmic Reticulum
ERBB2	Erb-B2 Receptor Tyrosine Kinase 2
ERGIC	Endoplasmic Reticulum-Golgi Apparatus Intermediate Compartment
ERK	Extracellular Signal-Related Protein Kinase
ESCA	Oesophageal Carcinoma
FDA	United States Food and Drug Administration
FDR	False Discovery Rate
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
GBM	Glioblastoma Multiforme
gCSI	Genentech Cell Line Screening Initiative
GDSC	Genomics of Drug Sensitivity in Cancer
GEO	Gene Expression Omnibus
GISAID	Global Initiative on Sharing All Influenza Data
GNAS	Guanine Nucleotide Binding Protein Alpha Stimulating
GO	Gene Ontology
GRP78	78 kDa glucose-regulated protein
GSH	Glutathione
GSSG	Glutathione Disulphide
GTEx	Genotype-Tissue Expression
HA	Haemagglutinin
HCoV	Human Coronavirus
HGSC	High-Grade Serous Ovarian Carcinoma
HNSC	Head and Neck Squamous Cell Carcinoma
HPV	Human Papillomavirus

HSP90	Heat Shock Protein 90
ICGC	International Cancer Genome Consortium
IDWAS	Imputed Drug-Wide Association Study
IFN	Interferon
IGF-1	Insulin-Like Growth Factor 1
IGF1R	Insulin-Like Growth Factor 1 Receptor
IKZF1	Ikaros family zinc finger 1
IL-6	Interleukin 6
I-TASSER	Iterative Threading Assembly Refinement
kDa	Kilodalton
KICH	Kidney Chromophobe
KIRC	Kidney Renal Clear Cell Carcinoma
KIRP	Kidney Renal Papillary Cell Carcinoma
LAML	Acute Myeloid Leukaemia
LIHC	Liver Hepatocellular Carcinoma
LGG	Low Grade Glioma
LINCS	Library of Integrated Network-Based Cellular Signatures
LRRK2	Leucine-Rich Repeat Kinase 2
LUAD	Lung Adenocarcinoma
LUSC	Lung Squamous Cell Carcinoma
MDM4	Murine Double Minute 4
MERS-CoV	Middle East Respiratory Syndrome Coronavirus
MESO	Mesothelioma
miRNA	MicroRNA
MGMT	O-6-Methylguanine-DNA Methyltransferase
MLH1	MutL Homolog 1
MRE11	Meiotic Recombination 11
mRNA	Messenger RNA
MOI	Multiplicity Of Infection
M^{Pro}	Main Protease
M protein	Membrane protein
NADP	Nicotinamide Adenine Dinucleotide Phosphate
NAMPT	Nicotinamide Phosphoribosyltransferase
NCBI	National Center for Biotechnology Information

NCI	National Cancer Institute
NERVTAG	The New and Emerging Respiratory Virus Threats Advisory Group
NHBE	Normal Human Bronchial Epithelial cells
NMR	Nuclear Magnetic Resonance
NOTCH1	Notch Receptor 1
N protein	Nucleoprotein
NQO1	NAD(P)H Quinone Dehydrogenase 1
NRF2	Nuclear Factor Erythroid 2-Related Factor 2
NSCLC	Non-Small-Cell Lung Cancer
NSP	Non-Structural Protein
OCG	Office of Cancer Genomics
ORF	Open Reading Frame
OV	Ovarian Serous Cystadenocarcinoma
PAAD	Pancreatic Adenocarcinoma
PANTHER	Protein Analysis Through Evolutionary Relationships
PARP	Poly (ADP-Ribose) Polymerase
PBMC	Peripheral Blood Mononuclear Cells
PCA	Principal Component Analysis
PCPG	Pheochromocytoma and Paraganglioma
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEA-15	Proliferation And Apoptosis Adaptor Protein 15
PI3Kβ	Phosphatidylinositol 4,5-Bisphosphate 3-Kinase Catalytic Subunit Beta Isoform
PL^{pro}	Papain-Like Protease
PMS2	Post Meiotic Segregation Increased 2
POLR2A	RNA Polymerase II Subunit A
PRAD	Prostate Adenocarcinoma
PSA	Prostate-Specific Antigen
RAP-MS	RNA Antisense Purification with Mass Spectrometry
RBD	Receptor Binding Domain
RCC	Renal Cell Carcinoma
RCCL	Resistant Cancer Cell Line Collection
READ	Rectum Adenocarcinoma

RMA	Robust Multichip Average
RNA	Ribonucleic Acid
RNA-seq	RNA sequencing
RNP	Ribonucleocapsid
RNR	Ribonucleotide Reductase
RTC	Replicase-Transcriptase Complex
SAMHD1	Sterile Alpha Motif and Histidine/Aspartic Acid Domain-Containing Protein 1
SARC	Sarcoma
SARS-CoV	Severe Acute Respiratory Syndrome Coronavirus
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SCLIP	Superior Cervical Ganglion 10-Like Protein
SDP	Specificity Determining Position
SETD2	SET Domain Containing 2
SF3B1	Splicing Factor 3B Subunit 1
shRNA	Short Hairpin RNA
siRNA	Small Interfering RNA
SKCM	Skin Cutaneous Melanoma
SLC2A3	Solute Carrier Family 2 Member 3
SLC35F2	Solute Carrier Family 35 Member F2
SLFN11	Schlafen Family Member 11
SNP	Single Nucleotide Polymorphism
SOX10	SRY-Box Transcription Factor 10
S protein	Spike protein
STAD	Stomach Adenocarcinoma
T-ALL	T-Acute Lymphoblastic Leukaemia
TARGET	Therapeutically Applicable Research to Generate Effective Treatments
TARID	TCF21 antisense RNA inducing promoter demethylation
TCF21	Transcription Factor 21
TCGA	The Cancer Genome Atlas
TF	Transferrin
TGCT	Testicular Germ Cell Tumour
THCA	Thyroid Carcinoma
THYM	Thymoma

TMPRSS2	Transmembrane Protease Serine 2
TNF	Tumour Necrosis Factor
TOP1	DNA Topoisomerase 1
TP53	Tumour Protein P53
TPM	Transcripts Per Million
tRNA	Transfer RNA
UCEC	Uterine Corpus Endometrial Carcinoma
UCS	Uterine Carcinosarcoma
UVM	Uveal Melanoma
VEGF	Vascular Endothelial Growth Factor
VEGFR	Vascular Endothelial Growth Factor Receptor
VUI	Variant Under Investigation
WHO	World Health Organisation

Chapter 1: Introduction

Access to genomic datasets from sources ranging from large, publicly funded projects to small, independent research groups has expanded rapidly in recent years. As high-throughput technologies have become easier and cheaper to conduct, the availability of data types including gene expression, protein abundance, copy number, mutation, methylation and miRNA expression has increased significantly. This, coupled with the development of sophisticated bioinformatics methodologies for analysing such data, has represented an important step in the pursuit of precision medicine. It is widely acknowledged that patient response to treatment can be substantially affected by genetic features, whether inherited or the result of mutations, and targeted therapies are becoming increasingly important in a wide variety of clinical settings. Moreover, the ability to rapidly produce data which can be made publicly available has been of critical importance in the context of the Coronavirus disease 2019 (COVID-19) pandemic.

In this work, we have made use of numerous publicly available data resources to answer biological questions relating to cancer and to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)/COVID-19. Our *in silico* investigations have complemented the *in vitro* studies of our collaborators (and vice versa), which we have integrated to produce novel findings with a high degree of validity and scientific and clinical relevance. We have used a variety of bioinformatics resources to perform our analyses, to which we have applied statistical methodologies to evaluate the significance of our results and make conclusions. Our work demonstrates the diversity of applications for large-scale open-access databases and the benefit that databases of large numbers of cell lines or patients has can have on our ability to produce reliable results. We have also highlighted the importance of the contribution of data shared from smaller research studies to the wealth of information accessible in the public domain.

1.1. Use of publicly available data for the analysis of cancer

1.1.1. Overview of cancer

According to the World Health Organisation, cancer is currently the second most common cause of death globally, accounting for one in six deaths annually (World Health Organization, 2021). One study estimated 1,898,160 new cases and 608,570 deaths from any cancer type per day in in the United States, with prostate, lung and colorectal cancers accounting for 46% of all cases (Siegel et al., 2021). Recent increases in cancer incidences have been reported in a number of studies. In the United Kingdom, increases of between 67% in the incidence of uterine cancer and 375% in the incidence of melanoma were identified between 1980 and 2013 (using data from Cancer Research UK). This was attributed to the introduction of screening programs and increased public awareness of cancer prompting earlier diagnosis of cancers which may have otherwise gone undetected (for example, before death due to another cause), as well as increases in incidental detection (for example, due to the increase in the number of CT scans performed for other diseases). Importantly, the study did not show similarly significant increases in mortality in the majority of cases, which the authors propose could be the result of over-diagnosis or improved treatments (Oke et al., 2018). In another study (using data from the National Cancer Registration and Analysis Service via the Office for National Statistics), an increase in the age-standardised rates across 23 cancer types of 783.29 to 808.13 per 100,000 individuals was identified between 1993 and 2014. Moreover, they showed a decrease in age-standardised mortality rates from 542.69 to 403.92 per 100,000 individuals (Smittenaar et al., 2016). More recently, a study in the United States (using data derived from the Surveillance, Epidemiology, and End Results Program, the National Program of Cancer Registries, the North American Association of Central Cancer Registries and the National Center for Health Statistics) showed that, in men, cancer incidences have slowly declined and then stabilised since the significant increases that were observed up to the 1990s (coinciding with the introduction of prostate-specific antigen (PSA) testing). They attributed this decline to reductions in the use of screening programs (i.e. to avoid over-diagnosis) and changes in population behaviour (for example, fewer men smoking resulting in reductions in lung cancer incidence). In women, they show that incidences have remained stable since the 1970s but slightly increased in recent years, which they attributed to decreases in fertility and increases in body weight resulting in increases in breast cancer incidences. In both males and females, they reported a decline in colorectal cancer incidences but increases in the incidence of kidney cancer, pancreatic cancer, melanoma and upper respiratory tract cancers. They also noted a steady decline in mortality rates corresponding to an overall decline of 31% between

1991 and 2018, also attributed to early diagnosis and improved treatments, as in the UK studies (Siegel et al., 2021).

Notably, each study (Oke et al., 2018; Siegel et al., 2021; Smittenaar et al., 2016) highlights cancer-specific differences in the relationship between incidence and mortality – for example, the UK studies (Oke et al., 2018; Smittenaar et al., 2016) show that increasing incidences of cervical cancer in recent decades have coincided with reduced mortality, whereas increasing incidences of liver cancer have correlated with increased mortality in the same timeframe. This could reflect the successes for certain cancers in prevention (e.g. the human papillomavirus vaccine (HPV)), early detection and the development of increasingly effective treatments (Aref-Adib & Freeman-Wang, 2016), which are lacking for other cancers (e.g. small hepatic lesions can be difficult to diagnose by ultrasound scan, and while α-foetoprotein is reportedly the only biomarker used for surveillance of hepatocellular carcinoma, it is prone to false positives), as well as differences in exposure to risk factors (e.g. alcohol-related cirrhosis and non-alcoholic steatohepatitis resulting from diabetes or metabolic syndrome is an increasingly prevalent risk factor for liver cancer in Western societies) (Llovet et al., 2021). Interestingly, in contrast to the UK studies, the study of US statistics showed that survival from cervical cancer has not improved in line with that of other common cancers in the last five decades, which the authors proposed to be the result of increases in the incidences of cervical adenocarcinomas (which cytology screening is less effective in detecting than cervical squamous cell carcinomas) as well as low rates of HPV vaccination in the US. They also report that declines in mortality from prostate cancer, which had coincided with the introduction of PSA testing, had ceased due to reduced PSA testing following concerns of over-diagnosis, also resulting in increases in late-stage diagnoses (Siegel et al., 2021). Meanwhile, a significant overall increase in the number of deaths from cancer between 2014 and 2035 has also been projected, which was proposed to be due to increases in population size and age (Smittenaar et al., 2016). This highlights the urgency for the development of early interventions and better targeted therapies.

In the process of tumorigenesis, cells undergo a series of genetic changes which collectively confer growth advantage. The nature of these changes has been proposed to convey the following features, facilitated by the 'enabling characteristics' of genome instability/mutation and tumour-promoting inflammation (Hanahan & Weinberg, 2000; Hanahan & Weinberg, 2011):

- sustaining proliferative signalling

- evading growth suppressors
- resisting cell death
- enabling replicative immortality
- inducing angiogenesis
- activating invasion and metastasis
- reprogramming of energy metabolism
- evading immune destruction

Cytotoxic chemotherapeutic agents and radiation therapies have been used for decades in the treatment of cancer. However, systemic therapies are associated with side effects with both immediate and chronic consequences to patients, ranging from nausea and vomiting to life-threatening multi-organ toxicity (Nurgali et al., 2018). It is therefore of critical importance to select treatment protocols based on the risk-benefit ratio – for example, it has been known that treatment-refractory T-acute lymphoblastic leukaemias respond better to treatment with the nucleoside analogue drug nelarabine compared with B-acute lymphoblastic leukaemia. Given the risk of neurotoxicity associated with nelarabine therapy, knowledge of lineage-specific differences in drug response is therefore extremely important (Kadia & Gandhi, 2017). We investigate this further in chapter 3, identifying expression of the deoxynucleoside triphosphate triphosphohydrolase SAMHD1 (sterile alpha motif and histidine/aspartic acid domain containing protein 1) as a possible factor influencing differential nelarabine sensitivity (since SAMHD1 has been previously shown to detoxify nucleoside analogue drugs including nelarabine (Herold, N., Rudd, Sanjiv, Kutzner, Bladh et al., 2017; Knecht et al., 2018)). By better understanding the pathways driving cancer progression and the mechanisms of action of anti-cancer agents, better therapy choices can be made. Drugs which target specific pathways can also be developed which not only have enhanced efficacy over traditional chemotherapeutics but may also have less severe side effects and fewer long-term sequelae.

Each of the above 'hallmarks' of cancer therefore presents a potential therapeutic opportunity, many of which are already being exploited for the treatment of various cancer types. For example, protein kinase inhibitors such as imatinib, gefitinib and erlotinib block the phosphorylation by serine, threonine or tyrosine kinases of target proteins, thereby preventing the transduction of signals stimulating cell cycle regulation, survival and proliferation (Kannaiyan & Mahadevan, 2018). Angiogenesis is also targeted by drugs such as bevacizumab, a recombinant humanized monoclonal antibody which binds to vascular

endothelial growth factor A (VEGF-A) and inhibits its interaction with vascular endothelial growth factor receptors 1 and 2 (VEGFR-1 and VEGFR-2), thereby attenuating signals which would promote proliferation of vascular endothelial cells and subsequent blood vessel formation (Keating, 2014). Many other putative drug targets remain at the pre-clinical stage of research, such as phosphoprotein enriched in astrocytes 15 (PEA-15), the phosphorylation status of which has been shown to determine its ability to retain extracellular signal-regulated kinase 1 and 2 (ERK1/2) in the cytoplasm and thereby attenuate proliferative signalling (Bartholomeusz et al., 2008; Sulzmaier, F. et al., 2012). The role of PEA-15 in cancer and drug response is explored further in chapter 2.

During the development and use of both cytotoxic and targeted cancer drug therapies, intratumour heterogeneity has been identified as a major barrier to the development of effective treatments (Marusyk et al., 2020; Mirzayans & Murray, 2020; Ramón Y Cajal et al., 2020). The genomic instability observed among cancer cells results in the accumulation over time of populations of genetically diverse cells. Those cells harbouring mutations which convey favourable phenotypic consequences (which can include properties mediating drug resistance) will be selected for as the tumour progresses. In addition to innate resistance, mutations, changes in gene expression, changes in the activation of target genes and even changes to the tumour microenvironment *post* drug or radiotherapy can contribute to the development of adaptive resistance, rendering therapies which initially appeared to work to be less effective – as described by Hanahan and Weinberg, loss of any of the 'hallmark capabilities' that cancer cells exhibit may be compensated for by greater reliance on other features, and redundancy of signalling pathways between each of the capabilities can also mean that drug-induced selective pressures stimulate heterogeneous tumour cell populations to adapt.

Genomic, epigenomic, transcriptomic and proteomic analyses are key to the improvement in our understanding of cancer. As the public dissemination of datasets becomes commonplace, analysis of these features by researchers from disciplines across the scientific community, from biochemists to computer scientists to statisticians, has the potential to broaden our understanding of the disease, enabling rigorous interrogation of the data and analysis from many different perspectives. This may in turn translate into the improvement of existing therapeutic interventions and the development of novel, more effective treatments.

1.1.2. Datasets available for the analysis of cancer

A large number of publicly available genomics datasets relate to cancer. This is unsurprising, given that cancer is a heterogenous group of diseases which affect a relatively large proportion of the population (~39.5% of individuals are expected to develop cancer at some point in their life (National Cancer Institute, 2020)). Many of the best-know examples include large-scale pharmacogenomic datasets of cancer cell line data, such as the National Cancer Institute-60 Human Tumor Cell Lines Screen (NCI-60) (Alley et al., 1988), the Cancer Cell Line Encyclopaedia (CCLE) (Barretina et al., 2012), the Cancer Therapeutics Response Portal (CTRP) (Basu et al., 2013; Rees et al., 2016; Seashore-Ludlow et al., 2015) and the Genomics of Drug Sensitivity in Cancer (GDSC) (Garnett et al., 2012; Iorio et al., 2016; Yang, W. et al., 2013), while tumour sample genomic and clinical data are available from sources such as the Cancer Genome Atlas (TCGA) and Therapeutically Applicable Research To Generate Effective Treatments (TARGET) databases. Other resources, such as cBioPortal (Cerami et al., 2012; Gao, J. et al., 2013) and the NCI's Genomic Data Commons (National Cancer Institute, 2021) collate multi-platform data from projects such as those listed above along with datasets from smaller studies which are often concerned with a single cancer type. Data generated as part of independent research projects can also be submitted to online repositories such as the Gene Expression Omnibus (GEO) (National Center for Biotechnology Information, 2021), including gene expression, methylation and single nucleotide polymorphism (SNP) array data, which can then be utilised by the wider research community. Notably, a significant proportion of the datasets submitted to GEO are also cancer related. To illustrate this, we performed a search of the GEO database to compare the number of cancer-related uploads with that of other human diseases (Figure 1.1). Viral and bacterial infections (similarly diverse groups of diseases), other leading causes of mortality (Rana et al., 2020) as well as selected common genetic diseases (Parkinson's disease (DeMaagd & Philip, 2015), Down's syndrome (Kazemi et al., 2016), cystic fibrosis (Almughem et al., 2020), thalassemia (Marengo-Rowe, 2007)) were included in the search:

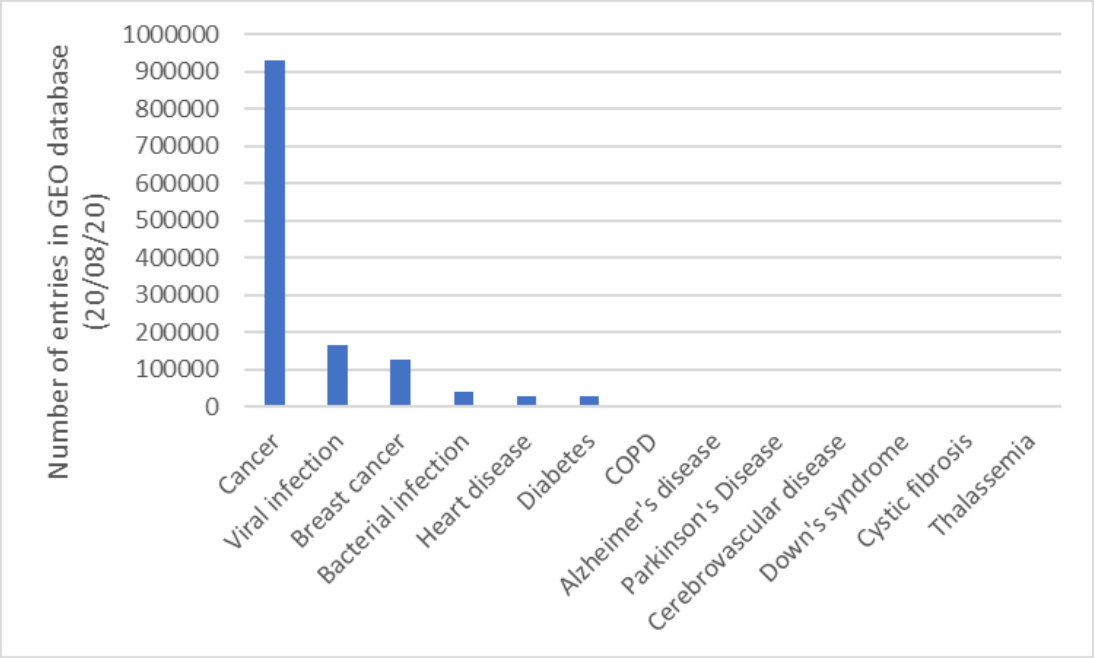


Figure 1.1. Number of articles associated with different diseases deposited in GEO as of 20/08/2020

Of note, data from studies pertaining only to breast cancer were the third most frequently uploaded. This indicates that it may not simply be a result of the large number of different cancer types that cancer-related datasets are the most common.

1.1.3. Authors demonstrate the value of large-scale cancer genomics datasets

1.1.3.1. Cell line data

Pharmacogenomic screens involving hundreds of cell lines from multiple lineages treated with hundreds of different drugs and small molecules have become an extremely important resource for the analysis of cancer drug response. The results from such screens can not only elucidate mechanisms of action for drugs which are already in clinical use, but they can also help to inform the development of novel drugs which may have therapeutic benefit to patients. They can also aid in the stratification of patients likely to respond well to certain treatments, which could ultimately help to avoid unnecessarily subjecting patients to drug-associated risks and side effects and could also lead to improvements in therapeutic outcomes. To highlight the value of these resources, database releases are often accompanied by publications demonstrating the flexibility of the data and their potential applications in cancer research.

The CCLE is an ongoing collaboration between the Broad Institute and the Novartis Institutes for Biomedical Research and its Genomics Institute of the Novartis Research Foundation which began in 2008. It was established for the purpose of characterising large numbers of human cancer cell lines to relate genomic features with drug response. The project currently includes data for 1457 cell lines from 39 lineages, including gene expression, protein expression, miRNA expression, copy number and mutation data. In their flagship study, researchers from this project first compared genomic features (gene expression, mutation and copy number variation) of cancer cell lines in their database with primary tumour data from external datasets. They reported strong positive correlations between analogous features in the datasets, indicating that cell lines may provide an appropriate proxy for the analysis of tumour biology. They were able to identify multiple genomic features predicting cell line drug response, such as high insulin-like growth factor 1 receptor (*IGF1R*) expression correlating with sensitivity to insulin-like growth factor (IGF-1) receptor inhibitor AEW541 in multiple myeloma cell lines and high schlafen family member 11 (*SLFN11*) expression predicting sensitivity to deoxyribonucleic acid (DNA) topoisomerase 1 (TOP1) inhibitors irinotecan and topotecan across cell lines from multiple lineages. They also validated this latter finding using data from the NCI-60 cell line panel (Barretina et al., 2012).

A more recent addition to the CCLE database included metabolic profiling of cancer cell lines, such as quantification of the abundance of nitrogenous compounds and cationic species, sugars, organic acids and lipids. They found significant lineage-specific differences in the abundance of different metabolites, as well as differences based on presence of features such as mutations, copy number variations and DNA methylation status. They also showed how the different metabolic profiles in the cancer cell lines could be associated with differences in gene dependencies in these cell lines – for instance, they noted how cells with a high abundance of redox metabolites (e.g. glutathione (GSH), glutathione disulphide (GSSG) and nicotinamide adenine dinucleotide phosphate ($NADP^+$)) were sensitive to knockout of nuclear factor erythroid 2–related factor 2 (*NRF2*, a transcription factor which regulates expression of antioxidant proteins). They also found that cells with aberrantly low levels of asparagine were more sensitive to knockout of asparagine synthetase (*ASNS*), and that cells expressing low levels of *ASNS* were significantly more sensitive to asparagine depletion in the media. Moreover, they showed that cells expressing low *ASNS* were more sensitive to asparaginase treatment, both *in vitro* and *in vivo* (Li, H. et al., 2019).

The authors expanded further on the existing CCLE database by including ribonucleic acid sequencing (RNA-seq), whole exome sequencing, whole genome sequencing, reverse-phase

protein array, RNA splicing, DNA methylation, microRNA (miRNA) expression and histone modification data for subsets of CCLE cell lines. The publication associated with these additions involved further comparisons with external cancer gene dependency databases. For example, they showed that the sensitivity of melanoma cell lines to SRY-box transcription factor 10 (*SOX10*) knockdown was paralleled by hypomethylation of the *SOX10* promoter in these cell lines, and that high expression of mir-215 was associated with catenin beta 1 (CTNNB1) dependency in stomach and colon cancer cell lines (which they also validated using the TCGA dataset, which showed enrichment of mir-215 in gastric cancers) (Ghandi et al., 2019).

Similarly to the CCLE, the developers of the CTRP database (also based at the Broad Institute) demonstrated the benefit of using large numbers of drugs and small molecules (n = 355) to generate sensitivity measurements across a subset of extensively characterised cancer cell lines (n = 242) from the CCLE. They validated their methods by identifying known associations between genetic mutations and drug sensitivities (e.g. cell lines mutant for B-Raf proto-oncogene, serine/threonine kinase (*BRAF*) were sensitive to P-0850, an analogue of vemurafenib, a BRAF-V600E inhibitor), and demonstrated the ability to identify novel drug-gene associations – some lineage-specific – with potential therapeutic implications. For instance, they found that epidermal growth factor receptor (*EGFR*)-mutant cells were resistant to multiple drugs which inhibit nicotinamide phosphoribosyltransferase (NAMPT, the rate-limiting enzyme of the NAD biosynthetic pathway). They also predicted sensitivity to navitoclax based on mutation and copy number variation, finding that *CTNNB1* mutations resulting in increased β-catenin expression correlated with sensitivity to navitoclax (Basu et al., 2013).

In an expansion of the previous version of the CTRP (cell lines n = 860, drug compounds n = 481), the authors described a novel method of analysis which integrated sensitivity data for multiple cancer cell lines to multiple compounds, thereby not only enabling the identification of common features of the cell lines which may contribute to sensitivity, but also facilitating the discovery of mechanisms of action for unannotated compounds. They revealed known drug-gene feature associations (e.g. cell lines mutant for *BRAF*-V600 showed sensitivity to BRAF inhibitors such as selumetinib) as well as novel associations (e.g. they found that LRRK2-in-1, a leucine-rich repeat kinase 2 (LRRK2) inhibitor, also functioned as a bromodomain inhibitor) (Seashore-Ludlow et al., 2015).

The same CTRP researchers also showed, in a separate publication, how basal gene expression can be used to uncover previously unknown mechanisms of action for small molecule compounds based on the correlation of gene expression with cancer cell line sensitivity. They also demonstrated how correlation-based analysis can identify mechanisms of drug resistance. For example, after controlling for confounders (e.g. cell line lineage, epithelial vs mesenchymal state, etc.), they found that high expression of biliverdin reductase B (*BLVRB*) was associated with resistance to B-cell lymphoma 2 (BCL2) inhibitor obatoclax, and *in vitro* validation suggested that BLVRB may modify obatoclax and prevent its interaction with BCL2-family proteins. Furthermore, they correlated high expression of solute carrier family 35 member F2 (*SLC35F2*) with sensitivity to YM-155, and high expression of ATP binding cassette subfamily B member 1 (*ABCB1*) with resistance to YM-155, indicating that the drug is a substrate for both transporters and its efficacy thus depends on its cellular uptake and efflux (Rees et al., 2016).

Meanwhile, the GDSC database was developed by a separate group (the Cancer Genome Project at the Wellcome Trust Sanger Institute), and initially included genomic profiling of 639 of cancer cell lines (some overlapping with the CCLE) along with sensitivity measurements to 130 drugs. They identified associations between tissue of origin and drug sensitivity and also between gene mutations and drug sensitivity. Again, some of the associations pointed to known mechanisms of drug action (e.g. breakpoint cluster region-Abelson murine leukaemia (*BCR-ABL*)-positive cell lines were sensitive to novel and approved ABL inhibitors such as nilotinib), whereas others revealed previously unknown associations (e.g. cell lines expressing mutant notch receptor 1 (*NOTCH1*) displayed sensitivity to BCL2 inhibitor navitoclax, which they suggested could be due to the increased expression of BCL2 family proteins in these cells). Similarly to the CCLE study, they also demonstrated how multiple cellular features can be integrated to identify markers of sensitivity (Garnett et al., 2012). A follow-up paper outlined how researchers can query the GDSC database using online resources (Yang, W. et al., 2013). The group later used tumour sample mutation, copy number and methylation data (from sources including the TCGA) to generate a map of 'cancer functional events' (CFEs), which they then compared with their cell line pharmacogenomic data. This enabled them to identify, for example, that mutations in tumour protein P53 (*TP53*) were associated with sensitivity to mitomycin C in bladder cancer cell lines, which has relevance given the use of this drug in treating bladder cancers. Moreover, they validated their results using data from the CCLE and CTRP (Iorio et al., 2016).

Critically, each of these studies has displayed the usefulness of cell lines as model systems reflecting *in vivo* biology. The CCLE researchers found strong positive correlations between multiple genomic features in the CCLE data with data from primary tumour samples from independent datasets. Similarly, the GDSC researchers found significant agreement between the tumour data and the cell lines, especially when considering tumours and cell lines with the same tissue of origin. They did, however, highlight the importance of using large numbers of cell lines to retain statistical significance in the associations between features of cancer and drug responses (Iorio et al., 2016; Rees et al., 2016).

1.1.3.2. Patient data

In addition to cell line databases, data from patient tumour samples represent a critical resource for cancer genomic analysis. The TCGA project was founded in 2006 as a collaboration between the National Cancer Institute and the National Human Genome Research Institute. They have performed multidimensional characterisation of human tumour samples from 33 different cancer types in order to better our understanding of the molecular features underlying different cancers. The data have been widely utilised throughout the cancer research community to enable, for example, the identification of tumorigenic pathways and potential therapeutic targets. A number of publications have been associated with the TCGA datasets, both regarding individual cancer types and pan-cancer analyses. The first involved an analysis of aberrant copy number, gene expression and DNA methylation from 206 glioblastoma samples. They found 453 non-synonymous SNPs in 223 genes, including several within genes involved in DNA damage repair. They also found multiple previously identified and novel copy number alterations (e.g. amplification of AKT serine/threonine kinase 3 (*AKT3*)) and copy-neutral loss of heterozygosity in the region including the *TP53* gene. Moreover, they identified that deficient mismatch repair in samples characterised by methylation of the O-6-methylguanine-DNA methyltransferase (*MGMT*) gene from patients treated with DNA alkylating agents was associated with substantially greater GC to AT transitions at non-CpG dinucleotides compared with CpG dinucleotides (The Cancer Genome Atlas Research Network, 2008). Meanwhile, the pan-cancer analyses published by the TCGA research group highlight the common features observed between cancer types, the improvement in statistical power associated with larger sample sizes (i.e. by aggregation of data from samples derived from multiple cancer types) and the limitations associated with considering only individual cancer entities when attempting to identify cancer driver genes. They also note the importance of accounting for data that was

generated using different platforms and correcting for batch effects when integrating data from different sources (The Cancer Genome Atlas Research Network, 2013).

Similarly, the TARGET database is a resource of tumour sample data, in this case derived from paediatric patients. The project has been developed by the Office of Cancer Genomics (OCG) with the Cancer Therapy Evaluation Program at the National Cancer Institute (National Cancer Institute, NCI), in collaboration with research institutions and hospitals across Central and North America. The project also aims to characterise genetic alterations and aberrant pathways associated with different cancer types in order to identify therapeutic targets. To date, samples derived from five cancer types have been extensively investigated by gene expression, copy number, methylation, miRNA and mutation analyses. Numerous publications have been associated with TARGET data, including an analysis of high-risk B-cell acute lymphoblastic leukaemia which revealed poor outcomes to be associated with deletion of and missense, frameshift and nonsense mutations in Ikaros family zinc finger 1 (*IKZF1*) (encoding zinc finger protein Ikaros, which regulates lymphocyte differentiation) (Medeiros, 2009), an analysis of metastatic neuroblastoma which revealed surprisingly low frequencies of recurrent somatic mutations but noted potentially pathogenic germline variants in six genes (Pugh et al., 2013), and an analysis of clear cell sarcoma of the kidney which revealed downregulation of transcription factor 21 (TCF21) antisense RNA inducing promoter demethylation (*TARID*, a long non-coding RNA which induces promoter demethylation) and concomitant hypermethylation of the gene encoding transcription factor TCF21, a regulator of kidney cell differentiation (Gooskens et al., 2015). Pan-cancer analyses have also been performed by the group, which highlighted the importance of considering paediatric cancers independently of adult cancers, given divergent features such as lower median somatic mutation rate and differing cancer driver genes (Ma et al., 2018).

The studies outlined above exemplify how the availability of multi-platform data from large numbers of cancer patients is a significant development in cancer research. Not only do these databases provide a useful source of cancer genomics data for analysis, but they also enable direct associations between features of cancer and patient outcomes to be identified. Moreover, many of these databases incorporate data on drug treatments, patient characteristics and demographic characteristics (such as age, sex, ethnicity and lifestyle factors) so they can be taken into account when performing analyses of the data and testing hypotheses.

1.1.4. Open access to (pharmaco)genomic datasets enables researchers to investigate cancer drug response and predict patient survival

By releasing pharmacogenomic screening data for use by researchers from different fields, novel perspectives on the content and potential applications for the data are gained. In some instances, researchers choose to analyse individual genomic features in tandem with drug response data, which can be a simple and rapid method to examine mechanisms of action of chemotherapeutic agents and to predict cellular responses. It has been previously proposed that gene expression quantification can predict drug response better than other genomic features, such as recurrently aberrant copy number segments and mutations in cancer associated genes (Iorio et al., 2016). Indeed, it has been demonstrated that gene expression data can be correlated with cell line drug response to discover novel drug mechanisms of action and to identify markers of drug sensitivity – for example, the authors of the CTRP database showed that cell line sensitivity to tanespimycin (a heat shock protein 90 (HSP90) inhibitor) correlated with high expression of NAD(P)H Quinone Dehydrogenase 1 (*NQO1*), a gene that has been previously shown to be involved in the activation of the drug. Their results also gave mechanistic explanation for the relative insensitivity of the cell lines to other HSP90 inhibitors (Rees et al., 2016). In another study, high expression of *VEGFA* was shown to correlate with resistance to the anaplastic lymphoma kinase (ALK) inhibitor TAE684 in non-small cell lung cancer and soft tissue cancer-derived cell lines in both the CCLE and GDSC databases. They note that VEGF-A has been shown to be induced under hypoxic conditions and that previous studies have shown hypoxia to be associated with the development of resistance to ALK inhibitors, which provides evidence supporting their findings (Qin, Y. et al., 2017). Such results therefore suggest that correlating basal gene expression with cell line drug response has the potential to identify important markers of drug response, which may have therapeutic implications.

Other studies have demonstrated the importance of large databases for predicting drug response of untreated cell lines – for example, data from the CCLE and GDSC databases were used to generate linear weighted models to predict drug response based on the gene expression profiles of cell lines (i.e. reasoning that cell lines with similar gene expression profiles may respond similarly to a given drug, and thus giving these higher weights in the model) and to predict response to a novel drug based on the structural similarity of the drug to other drugs and the known response of cell lines to those similar drugs (i.e. reasoning that

drugs which are structurally similar may affect cell lines in a similar manner). They also combined these models to improve their predictive power (Zhang et al., 2015).

The availability of clinical data from patients (such as from the TCGA and TARGET databases) also enables the prediction of patient survival based on individual genomic features, which could in turn be considered as therapeutic targets. For example, one study found that high expression of solute carrier family 2 member 3 (*SLC2A3*) was significantly associated with reduced overall survival and disease-free survival in samples of colorectal carcinoma from the TCGA and from a dataset from an independent research group available in the GEO database (Kim, E. et al., 2019). In another study, while mutations in known cancer driver genes were reported to lack prognostic ability, copy number alterations of numerous genes were found to be prognostic in various cancer types both in the TCGA and in an independent cohort from cBioPortal and the International Cancer Genome Consortium (ICGC) – for example, novel associations between murine double minute 4 protein (*MDM4*) amplification and prostate cancer and between *NOTCH2* amplifications and melanoma were identified (Smith, J. C. & Sheltzer, 2018).

Meanwhile, integration of multiple data types can further enhance our understanding of disease processes and better inform the development of targeted therapies. In one study, data from five separate platforms (copy number, methylation, messenger RNA (mRNA) expression, miRNA expression and protein expression) from the TCGA database were clustered in order to enhance identification of disease subtypes above their tissue-of-origin annotations. They noted that their 'Cluster-Of-Cluster-Assignments' (COCA) algorithm, when added to a multivariate model in addition to tissue-of-origin and clinical features, significantly enhanced the prediction of survival outcomes. They also found that COCA subtypes correlated with the mutation status of numerous genes, copy number variation and gene expression profiles. Moreover, their clusters revealed similarities between distinct tumour types (for example, copy number profiles of their basal-like breast cancer subtype reflected most closely that of their ovarian cancer subtype), as well as diversity within certain other tumour types (for example, bladder cancer samples clustered into seven of the 11 main COCA groups, displaying substantial differences in multiple genomic features (Hoadley et al., 2014). This study revealed the importance of considering genomic features when classifying tumours, which could in turn facilitate better stratification of patients when selecting treatments.

The prediction of cancer cell line or patient response to a drug is also frequently approached using machine learning. Methods such as elastic net regression and random forest algorithms are commonly used to integrate multiple data types (e.g. gene expression, mutation, copy number variation) in order to predict response to individual drugs or drug families – for example, the authors of the CCLE, CTRP and GDSC databases each employed machine learning in their flagship studies (Barretina et al., 2012; Basu et al., 2013; Garnett et al., 2012). Biomarkers of *in vivo* drug response have also been identified using machine learning-based methods. One such method, named 'IDWAS' (imputed drug-wide association study), involved the generation of a regression model from GDSC basal gene expression and drug response data, which was applied to data from the TCGA to predict therapy response in patients. They demonstrated the validity of their method with their prediction that sensitivity to Erb-B2 receptor tyrosine kinase 2 (ERBB2) inhibitor lapatinib was greater in ERBB2+ individuals and increased with increasing *ERBB2* copy number (Geeleher et al., 2017). In another study, clustering molecular subtypes based on features such as miRNA, mRNA and protein expression revealed the influence of certain markers on patient survival outcomes, both within individual cancer types and also between different cancer types. For example, kidney cancer patient survival could be predicted from a machine learning model trained using ovarian cancer somatic copy number alteration data. They also noted how amalgamating somatic mutation data across cancer types enabled the identification of low-frequency mutations within genes which are known therapeutic targets or that have been previously found to have clinical relevance, which may have been otherwise overlooked (Yuan et al., 2014). As publicly available genomics databases are expanded to include greater numbers of cell lines of different lineages, greater numbers of patient samples from different tissues and a greater variety of genomic features from each, the ability to assemble improved training sets will increase, which will be used to develop models with higher accuracy and better predictive power.

1.1.5. Publicly available datasets can supplement *in vitro* and *in vivo* investigation

In addition to providing a resource for the generation of novel findings, publicly available datasets are frequently used to supplement *in vitro* and *in vivo* studies and as the basis for further investigation. Many studies use analysis of pharmacogenomic databases as the

starting point for their investigations, which they can then expand upon and/or experimentally validate. For example, in one study, hemizygous deletion of RNA polymerase II subunit A (*POLR2A*) was found to accompany *TP53* deletion in most samples from colorectal cancer (CRC) patients in the TCGA, and *POLR2A* expression was strongly correlated with its copy number in the TCGA and CCLE databases, which they also found in tumour and matched normal samples from colorectal cancer patients. These findings prompted further investigation into the sensitivity of CRC cells to inhibition of POLR2A and its therapeutic implications (Liu, Y. et al., 2015). In another study, GDSC drug response data was used to show that renal cell carcinoma (RCC) cell lines harbouring SET domain containing 2 (*SETD2*) mutations were significantly more sensitive to phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit beta isoform (PI3Kβ) inhibitors than cells which were wildtype for *SETD2*. They also used data from RCC patients from the TCGA to determine that *SETD2* was frequently deleted, mutated or downregulated, and that deletion of MutL homolog 1 (*MLH1* (encoding a component of the mismatch repair machinery) frequently accompanied that of *SETD2*. Given these findings, MLH1 status was investigated in *SETD2* mutant A498 RCC cell lines treated with a PI3Kβ inhibitor *in vitro*, which revealed that loss of MLH1 may confer drug resistance via downregulation of post meiotic segregation increased 2 (PMS2), another mismatch repair protein (Feng, C. et al., 2015). Elsewhere, copy number variants associated with response to tyrosine kinase inhibitors were identified using the GDSC database, their association with survival was calculated using the TCGA kidney renal clear cell carcinoma (KIRC) database, and the results were validated with *in vitro* assays – for example, copy number gain of guanine nucleotide binding protein alpha stimulating (*GNAS*) was associated with resistance to cabozantinib and reduced patient survival, while overexpression of GNAS in clear cell renal cell carcinoma cells resulted in increased proliferation, invasion and migration (Li, Y. et al., 2020).

Other researchers choose to follow their experimental findings with database analyses. For instance, a small interfering RNA (siRNA) screen of two patient-derived ovarian cancer cell lines identified that knockdown of bromodomain-containing protein 4 (*BRD4*) sensitised cells to poly (ADP-ribose) polymerase (PARP) inhibitor rucaparib, and that combination of rucaparib with bromodomain and extraterminal domain (BET) inhibitors synergistically reduced cell viability. To further investigate these results, they used the TCGA database to determine that greater levels of aneuploidy were observed in high-grade serous ovarian carcinoma (HGSC) patients with high BRD4. They also found mutual exclusivity between high BRD4 and breast cancer 1/2 (*BRCA1/2*) mutation in TCGA HGSC patients, while the GDSC data

showed that cell lines with *BRCA2* mutations were significantly more sensitive to BET inhibitors than cell lines carrying wildtype *BRCA2*. Use of these cell line and patient databases therefore provided support for their suggestion that high expression of *BRD4* may facilitate adaptation to genomic instability in ovarian cancer cells, which can be targeted with a combination of PARP and BET inhibitors (Lui et al., 2020). In another recent study, response to YM155 was analysed in cell lines from the CTRP and GDSC studies, which showed that cell lines expressing high levels of *ABCB1* were relatively more resistant to YM155 and cell lines expressing high levels of *SLC35F2* were relatively more sensitive to YM155 treatment, which provided further evidence supporting their initial finding that *ABCB1* expression was increased and *SLC35F2* expression was decreased in YM155-adapted UKF-NB-3 cell lines (Michaelis et al., 2020).

Meanwhile, others integrate *in vitro*, *in vivo* and *in silico* investigation to build a comprehensive body of evidence for the hypothesis being tested. For example, one study investigated the effect of a silent poly (ADP-ribose) polymerase 1 (*PARP1*) single nucleotide polymorphism on mRNA secondary structure, ribosome inhibition and ultimately protein structure, and its effect on sensitivity to PARP inhibitors. In addition to biochemical assays, they analysed mutant vs wildtype samples from the TCGA breast invasive carcinoma (BRCA) cohort, as well as drug response data from mutant vs wildtype cell lines in the CTRP and GDSC databases. They found that presence of a synonymous variant in *PARP1* in TCGA BRCA patients was associated with significantly lower *PARP1* expression than in patients expressing the wildtype allele, and their *in vitro* investigation also revealed *PARP1* expression to be lower in COV362 cells (which carry the mutant *PARP1* allele) compared with SKOV3 cells (which carry the wildtype *PARP1* allele). In addition, they found that cell lines in the CTRP and GDSC which carry the silent mutation were more sensitive to PARP inhibitors olaparib and veliparib, which they also validated with further *in vitro* assays (Cashman et al., 2020).

1.2. Use of publicly available data for the analysis of SARS-CoV-2

1.2.1. Overview of SARS-CoV-2

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of the novel coronavirus disease COVID-19, was first identified in December 2019 in the Hubei Province of Wuhan, China (Zhu, N. et al., 2020). It is one of seven coronaviruses known to infect humans (along with alphacoronaviruses human coronavirus (HCoV)-229E and HCoV-NL63, and betacoronaviruses HCoV-OC43, HCoVHKU1, severe acute respiratory syndrome coronavirus (SARS-CoV), and Middle East respiratory syndrome coronavirus (MERS-CoV) (Wang, H. et al., 2020)), and has been found to be most closely related to bat coronaviruses (bat SARS-like (bat-SL)-CoVZC45, sequence identity 87.99%; bat-SL-CoVZXC21, sequence identity 87.23%), and less so to SARS-CoV (~79% identity) and MERS-CoV (~60% identity) (Lu et al., 2020). Despite sequence similarities and use of the same host receptor (angiotensin-converting enzyme 2 (ACE2) (Hoffmann, Kleine-Weber et al., 2020; Letko et al., 2020; Zhou, P. et al., 2020)), there are notable differences in the pathogenesis of SARS-CoV-2 compared with related viruses. Notably, as of the end of 2020, there had been over 79,673,754 cases and 1,761,381 deaths worldwide due to COVID-19 (covid19.who.int), compared with around 8,000 cases and 774 deaths between November 2002 and July 2003 due to the SARS-CoV outbreak (nhs.uk/conditions/sars/).

SARS-CoV-2 entry into host lung epithelial cells is mediated through the binding of the viral Spike (S) protein with the ACE2 receptor. Notably, a study using data derived from three transcriptomics databases and from nine single-cell RNA-seq datasets revealed limited ACE2 expression in the respiratory tract (mainly in alveolar type 2 cells), while its highest expression was reported to be in the intestinal tract, with high expression also detected in the kidney, testis, gallbladder and heart. This could explain some of the clinical manifestations of disease that have been observed relating to these tissues, discussed below. Similar results for ACE2 protein expression levels were obtained using publicly available mass spectrometry data and by immunohistochemistry and Western blot by the authors. The authors suggest this could imply alternative mechanisms of SARS-CoV-2 entry (for example, via a different receptor or via a receptor-independent mechanism) or could result from immune-driven local upregulation of ACE2 in respiratory epithelial cells in response to infection (Hikmet et al., 2020). Others have suggested a potential role for co-receptors which may facilitate SARS-CoV-2 S binding to ACE2 (for example, heparan sulphate, neuropilins or sialic acid) such as 78 kilodalton (kDa) glucose-regulated protein (GRP78) or cluster of differentiation 147 (CD147) (Zamorano Cuervo & Grandvaux, 2020).

The S protein is present on the viral surface as a trimer comprising the surface-exposed S1 receptor binding subunit (containing the receptor binding domain (RBD)) and the

transmembrane S2 fusion subunit (Shang, J. et al., 2020; Xia, S. et al., 2020). The RBD is reported to cycle between 'up' and 'down' conformational states, with the latter mediating immune evasion and the former facilitating ACE2 receptor binding (Shang, J. et al., 2020). To facilitate membrane fusion, priming of the Spike protein by cellular proteases including the proprotein convertase furin and the serine protease transmembrane protease serine 2 (TMPRSS2) is required. Between S1 and S2 subunits, a multibasic four amino acid insertion was identified as a cleavage site for furin, which is absent from SARS-CoV and other SARS-related coronaviruses (Walls et al., 2020). Cleavage at this site by furin, which takes place in the endoplasmic reticulum-Golgi apparatus intermediate compartment during virion assembly, has been shown to promote proteolytic cleavage of the S protein and viral membrane fusion, potentially by making a further cleavage site located just upstream of the fusion peptide, the S2' site, more accessible to TMPRSS2, which cleaves the S protein on the cell surface (Bestle et al., 2020; Hoffmann et al., 2020; Papa et al., 2021; Tang, T. et al., 2020; Xia, S. et al., 2020). Cleavage at the S1/S2 site followed by further cleavage at the S2' site facilitates fusion of viral and host membranes and subsequent release of the viral genome into the host cell. Inhibition of TMPRSS2 using serine protease inhibitors such as camostat and nafamostat has been previously shown to inhibit SARS-CoV-2 entry (Hoffmann et al., 2020; Hoffmann, Schroeder, Kleine-Weber, Müller, Drosten, & Pöhlmann, 2020b), and in chapter 8 we show that another serine protease inhibitor, aprotinin, also inhibits SARS-CoV-2 entry at therapeutic concentrations and therefore has potential to be used as an antiviral treatment. Lysosomal cathepsins have also been proposed to play a role in mediating SARS-CoV-2 viral entry (Shang, J. et al., 2020), and cathepsin L (CTSL) has been shown to further cleave SARS-CoV-2 S protein following cleavage by furin, while knockdown by siRNA has been shown to inhibit SARS-CoV-2 pseudovirus entry into Huh7 cells (Zhao, M. et al., 2021). It has been suggested that, in the absence of TMPRSS2, SARS-CoV-2 may enter cells via an endocytic pathway within endolysosomes, following which CTSL may cleave the S2' subunit and stimulate viral fusion with endosomal membranes and subsequent release of the viral genome (Murgolo et al., 2021; Tang, T. et al., 2020). The ability to use different entry mechanisms therefore expands the tropism of the SARS-CoV-2 virus and could enhance its infectivity and transmissibility.

Upon release of the SARS-CoV-2 genome into the host cell cytoplasm, two overlapping open reading frames (ORFs), ORF1a and ORF1b, are translated on host ribosomes into polyprotein 1a (pp1a) and polyprotein 1b (pp1b), which are processed to produce 16 non-structural proteins (NSP1-11 from pp1a and NSP1-10 and NSP12-16 from pp1b) by viral proteases NSP3

(papain-like protease (PL^{pro})) and NSP5 (3C-like protease ($3CL^{pro}$) or main protease (M^{pro})). These 16 NSPs together form the viral replicase-transcriptase complex (RTC), with NSP1 involved in inhibiting host cell translation, degrading host mRNA and inhibition of interferon (IFN) signalling and NSPs 2-11 reportedly functioning in support of viral replication (such as by acting as replication cofactors, double membrane vesicle formation and mediating immune escape), whereas NSPs 12-16 are directly involved in viral replication. In particular, NSP12 acts as the RNA-dependent RNA polymerase, synthesising viral RNA, with NSP7 and NSP8 acting as cofactors and NSP14 providing proofreading 3'-5' exonuclease activity (Perlman & Netland, 2009; Romano et al., 2020; V'kovski et al., 2021). The RTC also mediates transcription of the remainder of the viral genome, which encodes the structural proteins S, envelope (E), membrane (M) and nucleocapsid (N), interspersed with accessory proteins. Following translation, the structural proteins are translocated to the endoplasmic reticulum (ER) membranes and move through the ERGIC, where interaction with newly formed viral RNA results in the formation of mature virions (Perlman & Netland, 2009; V'kovski et al., 2021). The N protein packages the viral genome into helical nucleocapsid structure (RNP), the M protein determines the shape of the viral envelope and interacts with E to facilitate virion assembly, while the E protein forms a viroporin which facilitates viral assembly and exit (Satarker & Nampoothiri, 2020). One study showed that, instead of using the biosynthetic secretory pathway, β-coronaviruses such as SARS-CoV-2 exit the host cell within deacidified lysosomes, which also facilitates immune escape by perturbing the antigen presentation pathway (Ghosh et al., 2020). Meanwhile, the accessory proteins perform diverse functions involving immune escape and interferon antagonism (Orf3a, Orf3b, Orf6, Orf7a, Orf8, Orf9b), cytokine signalling (Orf3a, Orf9c) and apoptosis (Orf3a) (Redondo et al., 2021). Orf3a has also been proposed to be responsible for the deacidification of lysosomes used for viral exit (Ghosh et al., 2020).

SARS-CoV-2 infection therefore relies on a complex process of receptor binding and proteolytic cleavage by cellular proteases to facilitate membrane fusion and viral entry before the virus can utilise host protein synthesis machinery to produce the proteins required by the virus for replication and subsequent exit from the host cell. A schematic of SARS-CoV-2 entry, replication and exit is provided in Figure 1.2.

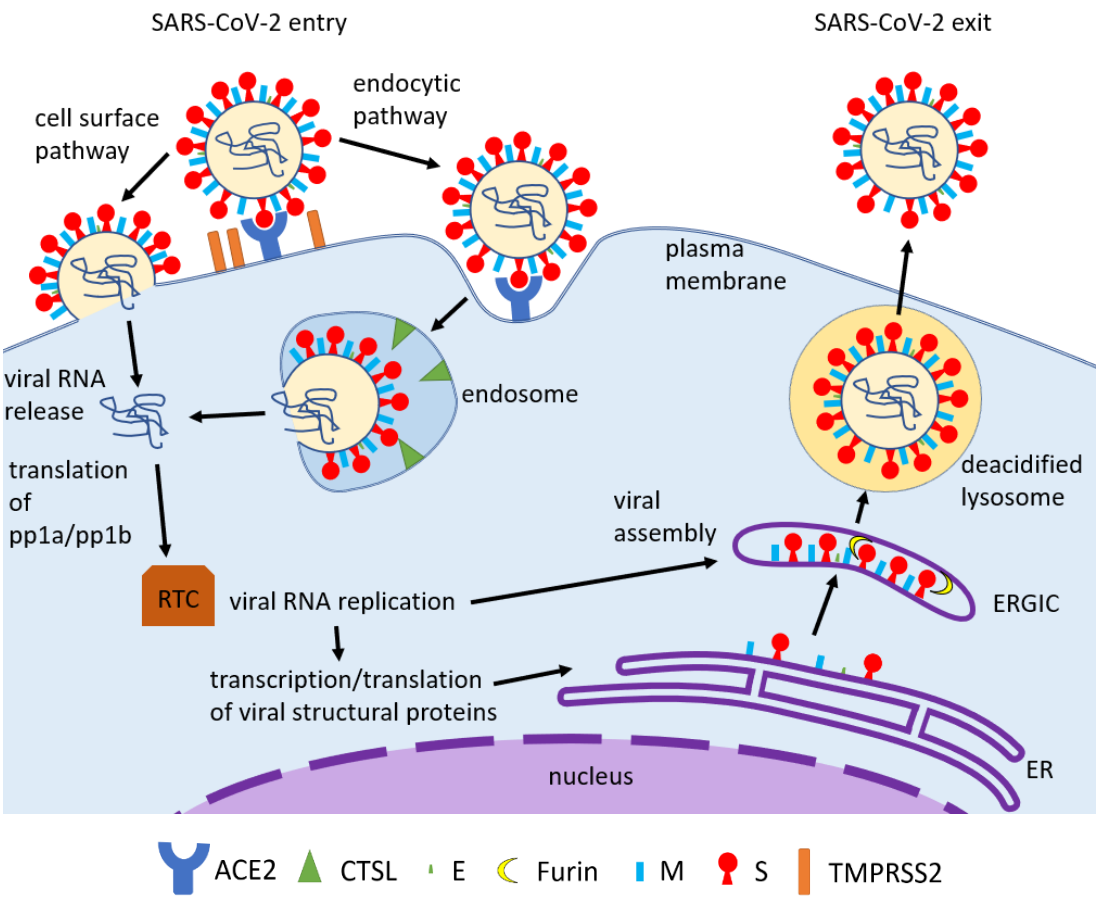


Figure 1.2. Overview of SARS-CoV-2 viral entry, replication and exit. Following binding to receptor ACE2, release of viral genome occurs via the cell surface pathway (membrane fusion at plasma membrane) or via the endocytic pathway (membrane fusion in endosome). Polyproteins pp1a/pp1b are translated and processed by viral proteases to form RTC. Viral genome is replicated and further proteins (structural and accessory) are transcribed and translated before structural proteins are inserted into ER membranes. Viral RNA interaction with structural proteins, virion assembly and Spike protein cleavage by proprotein convertase furin occurs in ERGIC. Mature virions exit the host cell in deacidified lysosomes. ACE2 = ACE2 angiotensin converting enzyme 2. CTSL = Cathepsin L. E = Envelope protein. ER = Endoplasmic reticulum. M = Membrane protein. RTC = Replicase-transcriptase complex. S = Spike protein. TMPRSS2 = Transmembrane protease serine 2.

The clinical manifestations of COVID-19 range from asymptomatic infection to life-threatening multiorgan failure and death. The risk of developing severe disease has been shown to correlate with certain demographic factors (such as older age (Liu, Y. et al., 2020), being male (Kelada et al., 2020; Peckham et al., 2020) and being from black, Asian or ethnic minority background (Patel et al., 2020; Sze et al., 2020)) as well as a number of comorbidities (including hypertension, diabetes and coronary heart disease (Zhou, F. et al., 2020)). Mild to

moderate cases typically present as flu-like symptoms (fever, cough, headache, myalgia, fatigue) as well as respiratory distress, gastrointestinal symptoms, sore throat and loss of sense of taste and/or smell. In severe cases, patients can also develop symptoms resembling acute respiratory distress syndrome (ARDS) (Li, X. & Ma, 2020), neurological symptoms (Whittaker et al., 2020), cardiac injury (Mitrani et al., 2020), acute kidney injury (Chaibi et al., 2020) and laboratory findings including lymphocytopoenia and elevated C-reactive protein (Esakandari et al., 2020). Substantial release of proinflammatory cytokines such as IL-6, IL2R and TNFα indicative of a 'cytokine storm' has been frequently observed, which can lead to massive infiltration of immune cells leading to tissue damage and multiple organ failure (Olwal et al., 2021; Qin, C. et al., 2020). In one study, 100% of fatal cases developed COVID-19-related sepsis, compared with 42% of non-fatal cases (Zhou, F. et al., 2020). Autopsy reports have shown frequent pulmonary consolidation with diffuse alveolar damage, hyaline membrane formation and hyperplasia of type II pneumocytes (Borczuk et al., 2020; Tian et al., 2020). In patients with a longer duration of disease, large vessel thrombi as well as microthrombi within arterioles and capillaries were frequently observed (Borczuk et al., 2020; Fahmy et al., 2021). Coagulopathy is recognised as a significant feature of severe COVID-19 and has been proposed to result from the release of proinflammatory cytokines, which is known to be associated with hypercoagulation, for example, by increasing fibrinogen and platelet production and via induction of the expression of tissue factor, which plays a critical role in the initiation of coagulation (Hadid et al., 2020; Mezalek et al., 2020). We explore COVID-19-related coagulopathy further in chapter 7, in which we discuss a possible role for transferrin (an iron-binding glycoprotein which has recently been implicated as an important clotting regulator (Tang, X., Zhang et al., 2020)) in mediating enhanced coagulation in severe cases of infection.

There are notable differences in the clinical characteristics of the diseases caused by SARS-CoV-2 compared to SARS-CoV, which caused a major outbreak in 2002. Although both viruses infect the lower respiratory tract causing pneumonia with acute respiratory distress, sepsis and multi-organ failure, the fatality rate from SARS was substantially higher than that of COVID-19 (~10% compared with ~2%, respectively (Wang, C. et al., 2021)). SARS-CoV-2 transmission has also been observed from individuals with asymptomatic infection (Johansson, M. A. et al., 2021), whereas little or no asymptomatic transmission was observed for SARS (Wilder-Smith et al., 2005). Gastrointestinal symptoms and have been shown to be less common in COVID-19 compared to SARS, and laboratory findings suggest that patients with COVID-19 are less likely to develop thrombocytopaenia (Zhu, Z. et al., 2020). These

clinical differences are likely to be the result of structural and functional differences in the SARS-CoV and SARS-CoV-2 viruses, which also translate into differing host responses and therefore differing outcomes. For example, the furin cleavage site between S1 and S2 subunits of the Spike protein, which is present in SARS-CoV-2 but absent in SARS-CoV (Shang et al., 2020; Hoffman et al., 2020), may be associated with more efficient viral spread in SARS-CoV-2. The SARS-CoV Spike protein has also been found to bind with lower affinity to the ACE2 receptor compared with SARS-CoV-2, which could also underly the differences observed in their infectivity (Yan, R. et al., 2020). Moreover, as discussed, the Spike protein of SARS-CoV has also been shown to exist predominantly in the 'up' conformation compared with the 'down' conformation in SARS-CoV-2, which has been suggested to mediate immune evasion (Rossi et al., 2020). Further differences in immune evasion between SARS-CoV and SARS-CoV-2 have been identified in relation to interferon antagonism – for example, NSP1, NSP6 and ORF3b of SARS-CoV-2 have been shown to inhibit type I IFN signalling more potently than the equivalent proteins in SARS-CoV (Konno et al., 2020; Xia, H. et al., 2020). We investigate another potential mechanism of SARS-CoV-2 immune evasion (by upregulation of cluster of differentiation 47 (CD47), a marker of self that has been implicated in immune evasion in cancer (Chao et al., 2012)) in chapter 6. This chapter also includes discussion of the previous associations that have been made between CD47 expression and some of the common comorbidities of COVID-19. Moreover, we also investigate the sequence and structural differences between SARS-CoV and SARS-CoV-2 in detail and discuss possible mechanisms determining the differences in their clinical phenotypes further in chapter 5.

1.2.2. Datasets available for the analysis of SARS-CoV-2

The SARS-CoV-2 viral genome was first released by an international consortium on the 10th January 2020 (Holmes, 2020) and deposited in GenBank (Benson et al., 2013) (accession MN908947), a large online resource collating publicly available nucleotide sequence data from thousands of independent investigations. Four additional sequences from other patients which had >99.9% sequence homology were isolated, and all five were deposited between the 10th and 11th January in the Global Initiative on Sharing Avian Influenza Data (GISAID) database (Shu, Y. & McCauley, 2017). This platform was initially developed to facilitate the sharing of sequencing data from studies of influenza viruses but has now also

become a critical resource for the sharing of SARS-CoV-2 sequencing data. An article associated with the findings was published on the 3rd February 2020 (Zhou, P. et al., 2020). At the time of writing (21/08/2020), 16,557 nucleotide sequences of the SARS-CoV-2 virus have been released (11,123 complete sequences, 5,434 partial sequences).

In addition to sequence data, structures of the SARS-CoV-2 proteins have been deposited in the Protein Data Bank (PDB) (Berman et al., 2000), a large online repository of publicly accessible 3D structural data for proteins, DNA and RNA from x-ray crystallography, electron microscopy and nuclear magnetic resonance (NMR) studies. The SARS-CoV-2 structures available in the PDB include unliganded and uncomplexed protein structures as well as viral proteins in complex with other SARS-CoV-2 proteins, with ACE2, with antibodies or bound to drug compounds. Furthermore, as described for publications relating to cancer, a rapidly increasing number of datasets from studies into SARS-CoV-2 has been deposited into the GEO database (1,339 as of 24/10/2020, Figure 1.3), which includes not only genomic and transcriptomic datasets from studies of the SARS-CoV-2 virus itself, but data relating to host cell response to the virus.

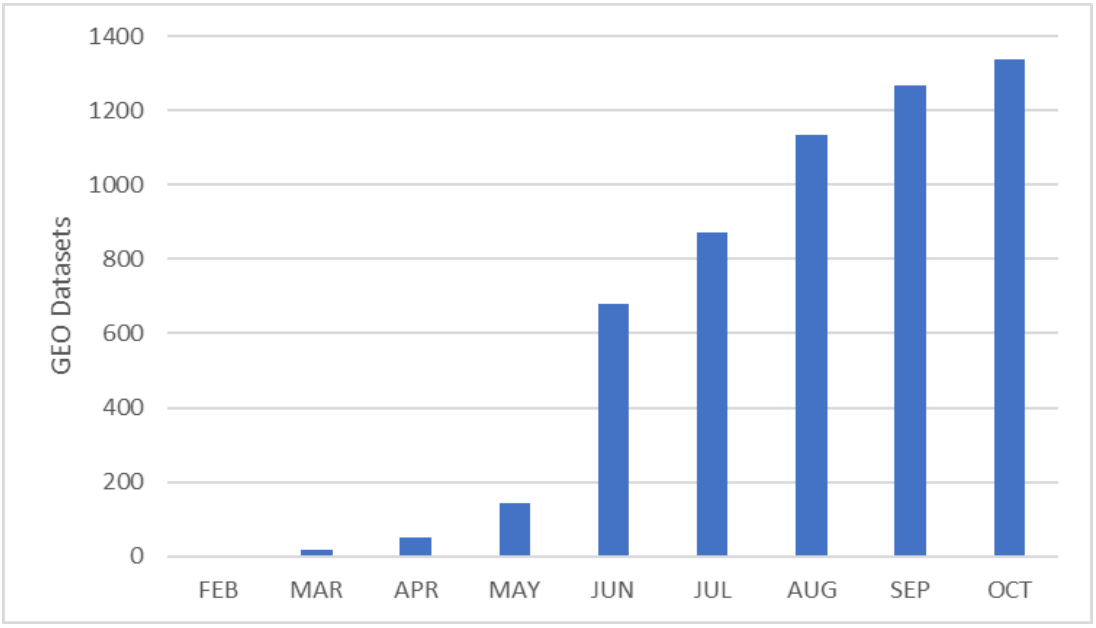


Figure 1.3. Number of datasets associated with SARS-CoV-2 between February 2020 and October 2020

1.2.3. Authors demonstrate the value of their data and its public release in the context of a global pandemic

The first crystal structure of a SARS-CoV-2 protein, the main protease protein (M^{pro}) in complex with inhibitor N3, was deposited in the PDB in January 2020 (PDB accession: 6LU7). In their associated publication, the authors described how the availability of the structure can facilitate the rapid identification of potential inhibitors of the M^{pro} protein, which may in turn have clinical relevance (Jin, Z. et al., 2020). Similarly, the first crystal structure of the SARS-CoV-2 receptor binding domain in complex with ACE2 was deposited in March 2020 (PDB accession: 6VW1). The authors identified the interacting residues and demonstrated how differences in binding affinities compared with that of SARS-CoV:ACE2 may affect the strength of the complex. They also remarked how their study may inform the development of novel treatments by highlighting potentially relevant epitopes that could be targeted by monoclonal antibodies, and how the results could even aid in the development of a receptor binding domain vaccine (Shang, J. et al., 2020).

The global host cellular response to viral infection has also been the focus of a number of investigations for which the data have been publicly released. For example, a proteomics study of SARS-CoV-2-infected Caco-2 cells revealed enrichment of components of the host translation machinery post infection at both the translatome and proteome level, as well as increases in proteins involved in RNA modification and carbon metabolism. For example, they investigated whether inhibition of splicing factor 3B subunit 1 (SF3B1) or of hexokinase may inhibit viral replication. Since inhibition of splicing and glycolysis did appear to prevent viral replication, the authors concluded that targeting of these pathways could represent a potential therapeutic strategy for COVID-19 (Bojkova et al., 2020). In another study, researchers infected A549 cells exogenously expressing ACE2 with the virus, both at a high multiplicity of infection (MOI) (2) and a low MOI (0.2), as well as CALU-3 cells, which they reported to be more susceptible to SARS-CoV-2 infection than wildtype A549 cells. They also infected ferrets with SARS-CoV-2 and quantified differential protein expression in samples from the trachea at day 3 post infection and intranasal samples at 14 days post infection. In addition, they compared the transcriptional profile of post-mortem lung samples from COVID-19 patients with samples from healthy lung biopsies. Importantly, their results revealed significant differences in responses to SARS-CoV-2 infection compared with infection with other viruses (e.g. respiratory syncytial virus and influenza A virus), especially pertaining to low expression of interferons I and III combined with significant increases in chemokine signalling (Blanco-Melo et al., 2020). Elsewhere, transcriptome analyses were performed comparing healthy peripheral blood mononuclear cells (PBMC) and bronchoalveolar lavage fluid (BALF) cells compared with SARS-CoV-2-infected cells. This

revealed similar perturbations to those observed in the study by Bojkova et al., such as upregulation of host transcriptional machinery and increased protein targeting to the endoplasmic reticulum membrane, as well as downregulation of proteins involved in response to stress (Xiong et al., 2020). Another group infected Vero E6 cells with SARS-CoV-2 and performed tandem mass spectrometry at intervals up to 7 days post-infection to identify differentially expressed proteins, also revealing deregulation of pathways such as membrane trafficking and lipid metabolism (Grenga et al., 2020). Datasets of host transcriptional and translational response to SARS-CoV-2 infection therefore provide a rich source of information for further study, which can be utilised in the identification of pathways which may be targeted for therapeutic intervention.

Meanwhile, other studies have investigated the interactions *between* host and viral proteins post viral entry. These publications have included the release of datasets which have enabled the scientific community to gain a better understanding of the pathogenesis of the SARS-CoV-2 virus and also provided critical information to suggest further potential therapeutic targets. In one notable study, human proteins that interact directly with SARS-CoV-2 proteins were identified by affinity purification and mass spectrometry to produce a list of 332 high confidence interactors. Consistent with the differentially expressed genes observed in other studies post-infection, they found a large proportion of interactions between viral proteins and host proteins involved in membrane trafficking, translation and regulation of the immune system. They also identified a number of drugs which could potentially target host-viral interactions, such as zotatifin, an inhibitor of eukaryotic initiation factor 4A (eIF4A), which they had found to interact with non-structural protein 9 (NSP9 which, based on 97% homology with NSP9 of SARS-CoV, is thought to be important for viral replication (Littler et al., 2020)) (Gordon et al., 2020). This study had already been cited by 450 other articles at the time of writing (03/01/2021) (National Library of Medicine, 2021). In another study, host-viral protein-protein interactions were analysed by overexpression of plasmids encoding SARS-CoV-2 encoded genes in HEK293 cells, affinity purification, liquid chromatography and mass spectrometry. This identified 295 interactions between viral proteins and host proteins, and again involved in pathways including intermembrane trafficking, metabolic processes, stress response and inflammation, as well as mRNA transport and nucleotide excision repair. 45 of the host proteins they identified were also found to bind with SARS-CoV-2 proteins in the Gordon et al. dataset. In addition, they investigated intra-viral protein-protein interactions using yeast-two hybrid screens and co-immunoprecipitation, arguing that disruption of these interactions could present another opportunity for therapeutic

intervention. They also quantified proteome changes between healthy and COVID-19-infected PBMC samples from patients, for example revealing increased expression of proteins involved in blood coagulation and decreased expression of proteins involved in T-cell activation in severely affected patients (Li, J. et al., 2020). Another group transduced A549 cells with haemagglutinin (HA)-tagged SARS-CoV and SARS-CoV-2 protein-expressing lentiviruses and performed affinity purification and mass spectrometry, which identified 1484 interactions between viral and cellular proteins associated with immune, stress and DNA damage response. They also quantified mRNA/protein abundance, phosphorylation and ubiquitination in SARS-CoV-2-infected ACE2-expressing A549 cells, which also revealed perturbations in these pathways (Stukalov et al., 2020).

In addition to quantification of changes in protein abundance following SARS-CoV-2 infection, another group identified host and viral protein interactions with viral RNA by infection of human liver cell line Huh7 with SARS-CoV-2 followed by RNA antisense purification with mass spectrometry (RAP-MS). 13 SARS-CoV-2 proteins and 104 human proteins were identified to interact with SARS-CoV-2 RNA. Notably, they compared their data with that of Gordon et al., and found only 10 of the 332 human proteins that interacted with SARS-CoV-2 proteins to also bind to SARS-CoV-2 RNA. They also identified significant enrichment of proteins interacting with SARS-CoV-2 RNA involved in translation, DNA damage response and membrane targeting, and described how their interaction network can suggest potential therapeutic targets, as in other studies outlined above. Furthermore, they performed detailed investigations into the importance of host protein/viral RNA interactions on viral pathogenesis (e.g. by performing knockout studies of human proteins using clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated protein 9 (Cas9)) (Schmidt et al., 2020).

Taken together, each of these studies for which data regarding the host cellular transcriptional and translational response to SARS-CoV-2 infection as well as host-viral interactions have been publicly released represents an important source of information for the identification of potential therapeutic targets, which is especially significant in the context of a pandemic of a novel disease for which few treatment options exist.

1.2.4. Publicly available data improves our understanding of SARS-CoV-2 and facilitates the identification of potential drug treatments

As described, the first sequences of the SARS-CoV-2 genome were deposited into databases that were either publicly accessible (GenBank (Benson et al., 2013)) or accessible to the scientific community (GISAID (Shu, Y. & McCauley, 2017)) within days of the discovery of the virus. These sequences and subsequent releases of sequence and structural data have prompted substantial research efforts by institutions worldwide into the nature of the virus, its relation to other coronaviruses and how it interacts with its host. A number of studies have focused on the entry receptor ACE2 – for example, one study investigated the potential for zoonotic transmission of the virus by comparing sequences of ACE2 orthologues from 410 vertebrate species with human ACE2 at known SARS-CoV-2 Spike binding residues. They also performed mutagenesis of variant residues from a selection of ACE2 orthologues using the PDB structure 6M0J to further analyse the nature of the host-viral interactions (Damas et al., 2020). Meanwhile, another group performed homology modelling of ACE2 structures for nine species, using sequences from the National Center for Biotechnology Information (NCBI) and using the human ACE2 structure (6M17 and 2AJF) as a template. This revealed differing residue interactions with the SARS-CoV-2 RBD, which they suggested may underlie the divergent susceptibility to SARS-CoV-2 infection observed in different species. In addition, sequences of transmembrane protease serine 2 (TMPRSS2, a serine protease which facilitates priming of the SARS-CoV-2 S protein and subsequent viral entry) from a number of species were aligned with that of the human sequence, and a model for human TMPRSS2 was generated using the protein structure prediction server Iterative Threating Assembly Refinement (I-TASSER). This enabled, for example, the identification of residues constituting a pocket surrounding the catalytic triad of the peptidase S1 domain, which was found to be identical in each of the species studied. The results led the authors to conclude that TMPRSS2 is unlikely to determine host susceptibility to SARS-CoV-2 (Brooke & Prischi, 2020).

Meanwhile, others have made use of available SARS-CoV-2 structures and the structures of relevant host proteins to simulate binding of drugs and peptide inhibitors. For example, one study used a publicly available antiviral peptide database (AVPdb (Qureshi et al., 2013)) to perform *in silico* docking of a selection of 51 peptides to the SARS-CoV-2 RBD, also using the PDB structure 6M0J. They were then able to perform molecular dynamics simulations and structure-activity relationship analysis on the peptides with the highest affinity for the RBD (Chowdhury et al., 2020). In another study, United States Food and Drug Administration (FDA) approved antiviral and antimalarial drugs were screened against the structure of ACE2 and of SARS-CoV-2 3C-like protease (Hussien & Abdelaziz, 2020). Similarly, ZINC15, another compound library (Sterling & Irwin, 2015) was used for a docking study of covalent inhibitors

of cysteine and serine proteases (cathepsins B and L and TMPRSS2, respectively), which could prevent the priming of the SARS-CoV-2 S protein for host cell entry (Li, Q. et al., 2020).

Numerous studies have also made use of the host-viral protein interaction network released by Gordon et al. (Gordon et al., 2020). For example, this dataset has been used to predict which existing drugs may disrupt the interactions between SARS-CoV-2 and host proteins and therefore be repurposed to treat COVID-19 infection (Das et al., 2020; Kowalewski & Ray, 2020), to compare with other predicted virus-host interactions (e.g. between predicted viral miRNAs and their human gene targets (Saçar Demirci & Adan, 2020)), to supplement further analyses (e.g. regarding the impact of SARS-CoV-2 on mitochondrial function (Singh et al., 2020)) and even to characterise subsets of interacting proteins in terms of their implications in tumorigenesis (Süt, 2020). The latter study is an example of how research into SARS-CoV-2/COVID-19 could have wider implications even extending into other areas of biological and clinical interest.

As discussed previously, a key source of valuable information regarding the impact of SARS-CoV-2 infection on the host has been transcriptomics and proteomics datasets from both patient-derived samples and studies of SARS-CoV-2-infected cell lines. In one study, publicly available transcriptomic data from a lung sample of a COVID-19 patient was analysed in combination with transcriptomic data from SARS-CoV-2 and SARS-CoV-infected normal human bronchial epithelial (NHBE) and 2B4 cell lines, respectively. Each dataset was obtained from the GEO database. This enabled the identification of pathways deregulated uniquely in the lung sample of the COVID-19 patient, such as those associated with surfactant and cholesterol metabolism. They also analysed the Gordon et al. dataset alongside a similar dataset of the SARS virus-host interactome generated using yeast two-hybrid screening (Pfefferle et al., 2011), and used these data to predict possible drugs and therapies targeting systems affected by SARS-CoV-2 infection, such as lung surfactants (Islam & Khan, 2020). Another study involved gene set enrichment analysis of interactome data from Gordon et al., transcriptome data from Blanco-Melo et al. and Xiong et al. and proteome data from Bojkova et al., as well as data from a resource that mines scientific literature to identify connections between genes (and their variants) and human diseases (Bauer-Mehren et al., 2010) and which now includes a data collection relating to SARS-CoV-2. The integration of data from these sources also facilitated the identification functional pathways affected by SARS-CoV-2 infection and drugs targeting those pathways – for instance, they identified that deregulation of mRNA splicing could be targeted by a number of antineoplastic agents, such as mitomycin and cisplatin (Barh et al., 2020).

In addition to utilising data that has been generated during the course of the COVID-19 pandemic, numerous research groups have seized the opportunity to mine large databases of compounds for their ability to target processes involved in the pathogenesis of the virus. The possibility of repurposing existing drugs (which have already undergone clinical trials and have been assessed for their safety and efficacy) has also been of great interest, especially given the urgency to find novel treatments. Targeting of the interaction between SARS-CoV-2 Spike protein and human ACE2 has been one common approach for the identification of possible drug interventions. One study used two of the 'drug-like chemical libraries' from MTiOpenScreen resource (Drugs-lib, a database of 7,173 approved drugs, and iPPI-Lib, a database of 51,232 drug-like molecules which target protein-protein interactions (Lagarde et al., 2018)) to identify potential inhibitors which bind in the same location as ACE2 in the RBD of SARS-CoV-2. They also used binding prediction software (RASPD (Mukherjee & Jayaram, 2013)), which incorporates data from a large compound library (ZINC (Irwin & Shoichet, 2005)). Each of these databases enabled the identification of several potential SARS-CoV-2-S RBD-binding compounds, including four antiviral compounds licensed and one in clinical trials for the treatment of hepatitis C (Behloul et al., 2020).

Before the SARS-CoV-2-host interactome had been experimentally examined, researchers sought to broaden the scope of investigation outside of Spike-ACE2 interactions using pre-existing data. In one study, expression of ACE2 was correlated with expression of other genes using data from lung adenocarcinoma (LUAD) patients in the TCGA, lung tissue samples of individuals in the Genotype-Tissue Expression (GTEx) database (a database of genomic data from post-mortem samples of 980 donors (GTEx Consortium, 2013)), and data from two GEO datasets of healthy volunteers. Generation of a protein-protein interaction network from the correlated genes enabled the identification of a number of existing drugs which target these pathways (Cava et al., 2020). Elsewhere, human proteins that were known to interact with other coronaviruses (four human coronaviruses, one murine and one avian) or proteins known to be involved in pathways related to coronavirus infection were identified from the literature. Existing drugs targeting these proteins and/or their functional pathways were then identified using multiple large drug databases (Zhou, Y. et al., 2020).

Another research group reasoned that existing drug response data could be used to identify drugs able to reverse the effects of lung injury caused by SARS-CoV-2-mediated inhibition of ACE2. Transcriptome data from lung tissue of deceased COVID-19 patients as well as data from lung tissue of healthy donors were obtained from GEO (from Blanco-Melo et al., discussed previously), and differential gene expression between COVID-19-infected and non-

infected donors was calculated. Meanwhile, they used the Library of Integrated Network-Based Cellular Signatures (LINCS) L1000 database, which provides gene expression data from human cell lines treated with perturbagens (including small molecules and FDA approved drugs) (Subramanian et al., 2017) to identify compounds which were associated with the inverse expression of the lung injury-associated genes, as well as 'concordantly expressed genes' (i.e. those which were upregulated in disease but downregulated by drug treatment, and vice versa) (He & Garmire, 2020). Such studies demonstrate the flexibility of open-access data, and how different data types and resources can be integrated in novel ways to make potentially significant findings.

1.2.5. Publicly available data enables the monitoring of and rapid response to SARS-CoV-2

Rapid sequencing and continual monitoring of mutations within the SARS-CoV-2 genome has been critically important during the COVID-19 pandemic, facilitating the early characterisation of the virus and enabling variants to be tracked. The GISAID and GenBank databases have been particularly important platforms for researchers to deposit sequences and retrieve sequences for analysis. For example, a study conducted on the earliest released sequences in GISAID identified that the genome comprises 14 ORFs which encode 27 proteins (15 NSPs, four structural proteins and eight accessory proteins). They also identified 380 amino acid substitutions between SARS-CoV-2 and SARS and SARS-like bat CoVs, none of which were of residues known to directly interact with ACE2, but six of which were located within the receptor binding domain (Wu, A. et al., 2020). Variants emerging in the viral genome began to be analysed as early as March 2020 using the sequencing data from both GenBank and GISAID (Pachetti et al., 2020; Wang, C. et al., 2020), and correlations with differing infectivity rates and survival outcomes across different populations have been performed. For example, a study of sequences retrieved from GenBank revealed that the presence of 4715L-type ORF1ab protein and 614G-type S protein, both of which were less prevalent among Asian populations, was associated with increased fatality rates (Toyoshima et al., 2020). GISAID has also been used to estimate the replicative advantage of a novel strain of SARS-CoV-2 (variant under investigation (VUI)-202012/01) (Grabowski et al., 2021), and also to enable the analysis of the impact of novel variants on disease severity and patient outcomes (Voss et al., 2020). Moreover, when it was discovered that mink in farms in

Denmark and the Netherlands had become infected with SARS-CoV-2, sequences of viral genomes isolated from mink and from humans infected in geographically similar regions were quickly uploaded to GISAID (Hammer et al., 2020; Oreshkova et al., 2020) and examined by the scientific community. One study included the use of these data to investigate the mechanisms of viral adaptation to a new host, such as by identifying variants emerging independently in multiple samples (van Dorp et al., 2020).

Open-access resources have also played a role in the generation of clinical recommendations during the pandemic. A notable example is the use of the University of Liverpool COVID-19 Drug Interactions resource (University of Liverpool, 2021), which was referred to in guidelines for the administration of dexamethasone in conjunction with remdesivir (Department of Health and Social Care, 2020). Although the resource gives no empirical evidence for any interaction (or lack of interaction) between the two drugs, it provides rationale for their safe co-administration. In addition, the minutes of a New and Emerging Respiratory Virus Threats Advisory Group (NERVTAG) meeting (NERVTAG Bird Table 8, 9th October 2020) include mention of a preprint which showed that exogenous heparin can inhibit SARS-CoV-2 entry by competing for binding to cell surface heparan sulphate. Notably, the study cited had involved the use of publicly available SARS-CoV-2 crystal structures to investigate the electrostatic potential across the surface of the Spike protein and to identify a putative heparan sulphate binding site (Clausen et al., 2020).

1.3. Thesis overview

This thesis presents six studies which all make use of open-access data to make novel biological findings. The databases and data resources used in each study are summarised in Table 1.1. A brief summary of each study is given below.

Chapter 1: *Introduction*

An introduction to the concept of publicly accessible data, discussing the data resources available to the scientific community and highlighting the scope and flexibility of application of these resources.

Chapter 2: *Non-Phosphorylatable PEA-15 Sensitises SKOV-3 Ovarian Cancer Cell Lines to Cisplatin*

An analysis of the relationship between the phosphorylation status of the astrocytic phosphoprotein PEA-15 and the sensitivity of ovarian cancer cell lines to the platinum-based chemotherapeutic drug cisplatin.

Contribution

- Acquired relevant data from the TCGA database
- Performed all analyses relating to use of the TCGA data
- Produced summary data table of TCGA results
- Assisted in the preparation of the manuscript for publication

Chapter 3: *SAMHD1 is a key regulator of the lineage-specific response of acute lymphoblastic leukaemias to nelarabine*

An analysis of the relationship between the expression of deoxynucleoside triphosphate triphosphohydrolase SAMHD1 in T-acute lymphoblastic leukaemia compared with B-acute lymphoblastic leukaemia and sensitivity to the nucleoside analogue drug nelarabine.

Contribution

- Acquired all relevant data from the CCLE, CTRP and GDSC databases
- Performed all analyses relating to use of the CCLE, CTRP and GDSC data
- Produced all graphs and data tables of results from each analysis of CCLE, CTRP and GDSC data
- Assisted in the preparation of the manuscript for publication

Chapter 4: *Multifaceted roles of SAMHD1 in cancer*

An analysis of the expression and regulation of SAMHD1 and its impact on the survival of patients in the TCGA and TARGET databases.

Contribution

- Acquired data from the TCGA and TARGET databases
- Performed all analyses of the data
- Produced figures and tables of results from the analysis of the data
- Assisted in the preparation of the manuscript for publication

Chapter 5: *Differentially conserved amino acid positions may reflect differences in SARS-CoV and SARS-CoV-2 behaviour*

An analysis of the genomic variations between SARS-CoV and SARS-CoV-2 which may be responsible for the differences in infectivity and mortality of the diseases they cause.

Contribution

- Acquired structural data from the PDB database
- Performed all structural analyses using PyMOL
- Produced figures and data tables of results from the structural analysis
- Assisted in the preparation of the manuscript for publication

Chapter 6: *A potential role of the CD47-SIRPalpha axis in COVID-19 pathogenesis*

An analysis of the transmembrane protein CD47 and its association with several COVID-19 risk factors which indicate that CD47 may act as a biomarker for the identification of patients likely to develop severe disease as a result of SARS-CoV-2 infection.

Contribution

- Acquired all relevant data released by Bojkova et al. (2020) and Blanco-Melo et al. (2020)
- Performed analyses of the transcriptomic and proteomic data
- Produced graphs from analyses of transcriptomic and proteomic data
- Assisted in the preparation of the manuscript for publication

Chapter 7: *COVID-19-related coagulopathy, is transferrin a missing link?*

An analysis of the coagulation-associated genes which are differentially expressed in response to SARS-CoV-2 infection including the iron transport protein transferrin, whose expression also correlates with COVID-19 risk factors and which could be associated with the disorder of coagulation observed in severe cases of COVID-19.

Contribution

- Acquired all relevant data released by Bojkova et al. (2020), Gordon et al. (2020) and data from the GTEx database
- Acquired data and performed all analyses relating to use of the GO ontology database
- Performed all analyses of the transcriptomic and proteomic data
- Produced all graphs and data tables of results from each analysis
- Assisted in the preparation of the manuscript for publication

Chapter 8: *Aprotinin inhibits SARS-CoV-2 replication*

An analysis of the potential for protease inhibitor aprotinin to compensate for the SARS-CoV-2-mediated downregulation of endogenous protease inhibitors, thereby inhibiting SARS-CoV-2 entry and replication.

Contribution

- Acquired all relevant data released by Bojkova et al. (2020)
- Performed analyses of the transcriptomic and proteomic data
- Assisted in the preparation of the manuscript for publication

Chapter 9: *Discussion*

An appraisal of the significance of the findings from each of the manuscripts presented in chapters 2-8, giving perspectives on their scientific and clinical relevance.

Chapter 10: *Conclusions*

A concluding statement highlighting the relevance and significance of this work.

Appendix

Supplementary materials for each of the manuscripts presented in chapters 2-8.

Table 1.1. Databases and data resources used for each study in this thesis

Database/resource	Chapter(s)	Data used in study	Reference
The Cancer Genome Atlas (TCGA)	2, 4	Transcriptomics (RNA-seq), methylation, miRNA expression, mutation, demographics, survival data	(The Cancer Genome Atlas Research Network, 2008; The Cancer Genome Atlas Research Network, 2013)
The Cancer Cell Line Encyclopaedia (CCLE)	3	Transcriptomics (microarray)	(Barretina et al., 2012)
The Cancer Therapeutics Response Portal (CTRP)	3	Transcriptomics (microarray), drug response	(Basu et al., 2013; Rees et al., 2016; Seashore-Ludlow et al., 2015)
Genomics of Drug Sensitivity in Cancer (GDSC)	3	Transcriptomics (microarray), drug response, methylation	(Garnett et al., 2012; Iorio et al., 2016; Yang, W. et al., 2013)

(Herold, T. et al., 2017)	3	Transcriptomics (microarray)	(Herold, T. et al., 2017)
Therapeutically Applicable Research to Generate Effective Treatments (TARGET)	4	Transcriptomics (RNA-seq), demographics, survival data	(https://ocg.cancer.gov/programs/target)
Global Initiative on Sharing Avian Influenza Data (GISAID)	5	Genomic sequence data	(Shu, Y. & McCauley, 2017)
Protein Data Bank (PDB)	5	Protein structural data	(Berman et al., 2000)
Virus Pathogen Database and Analysis Resource (VIPR)	5	Genomic sequence data	(Pickett, Sadat et al., 2012; Pickett, Greer et al., 2012)
(Blanco-Melo et al., 2020)	6	Transcriptomics (RNA-seq)	(Blanco-Melo et al., 2020)
(Bojkova et al., 2020)	6, 7, 8	Translatome and proteome proteomics (liquid chromatography-tandem mass spectrometry (LC-MS/MS))	(Bojkova et al., 2020)
Gene Ontology (GO) database	7	List of genes annotated with GO term “blood coagulation” (GO:0007596)	(The Gene Ontology Consortium, 2019)

The Genotype-Tissue Expression Project (GTEx)	7	Transcriptomics (RNA-seq)	(GTEx Consortium, 2013)
IntAct Molecular Interaction Database (IntAct)	7	List of human proteins interacting with SARS-CoV-2 proteins, derived from literature/user submissions	(https://www.ebi.ac.uk/intact/)
(Gordon et al., 2020)	7	List of human proteins interacting with SARS-CoV-2 proteins, identified by affinity chromatography/mass spectrometry	(Gordon et al., 2020)

Chapter 2: Non-phosphorylatable PEA-15 sensitises SKOV-3 ovarian cancer cells to cisplatin

Shahana Dilruba[1], Alessia Grondana[1], Anke C Schiedel[2], Naoto T Ueno[3], Chandra Bartholomeusz[3], Jindrich Cinatl Jr[4], Katie-May McLaughlin[5], Mark N Wass[5], Martin Michaelis[5], Ganna V Kalayda[1]

[1]Department of Clinical Pharmacy, Institute of Pharmacy, University of Bonn, 53113 Bonn, Germany.

[2]Department of Pharmaceutical and Medicinal Chemistry, Institute of Pharmacy, University of Bonn, 53113 Bonn, Germany.

[3]Section of Translational Breast Cancer Research, Department of Breast Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

[4]Institute of Medical Virology, Goethe University Hospital Frankfurt, 60323 Frankfurt/Main, Germany.

[5]Industrial Biotechnology Centre and School of Biosciences, School of Biosciences, University of Kent, Canterbury CT2 7NZ, UK.

Published: Cells. 2020 Feb 24;9(2):515. doi: https://doi.org/10.3390/cells9020515

Abstract

The efficacy of cisplatin-based chemotherapy in ovarian cancer is often limited by the development of drug resistance. In most ovarian cancer cells, cisplatin activates extracellular signal-regulated kinase1/2 (ERK1/2) signalling. Phosphoprotein enriched in astrocytes (PEA-15) is a ubiquitously expressed protein, capable of sequestering ERK1/2 in the cytoplasm and inhibiting cell proliferation. This and other functions of PEA-15 are regulated by its phosphorylation status. In this study, the relevance of PEA-15 phosphorylation state for cisplatin sensitivity of ovarian carcinoma cells was examined. The results of MTT-assays indicated that overexpression of PEA-15AA (a non-phosphorylatable variant) sensitised SKOV-3 cells to cisplatin. Phosphomimetic PEA-15DD did not affect cell sensitivity to the drug. While PEA-15DD facilitates nuclear translocation of activated ERK1/2, PEA-15AA acts to sequester the kinase in the cytoplasm as shown by Western blot. Microarray data indicated deregulation of thirteen genes in PEA-15AA-transfected cells compared to non-transfected or PEA-15DD-transfected variants. Data derived from The Cancer Genome Atlas (TCGA) showed that the expression of seven of these genes including *EGR1* (early growth

response protein 1) and *FLNA* (filamin A) significantly correlated with the therapy outcome in cisplatin-treated cancer patients. Further analysis indicated the relevance of nuclear factor erythroid 2-related factor 2/antioxidant response element (Nrf2/ARE) signalling for the favourable effect of PEA-15AA on cisplatin sensitivity. The results warrant further evaluation of the PEA-15 phosphorylation status as a potential candidate biomarker of response to cisplatin-based chemotherapy.

2.1. Introduction

Platinum-based drugs have been used to treat ovarian cancer since the late 1970s and cisplatin, followed by carboplatin-based combinations, has been the standard of care for over 15 years. As most patients relapse and ultimately succumb to ovarian cancer, new strategies are urgently required to improve survival. Cisplatin is believed to exert its cytotoxic effects via its interaction with DNA and formation of DNA adducts, primarily intrastrand crosslinks. This initiates signal transduction pathways involving among others Ataxia Telangiectasia Mutated (ATM) protein, p53, p73 and mitogen-activated protein kinases (MAPK), eventually resulting in cancer cell apoptosis (Dilruba & Kalayda, 2016). However, DNA damage-mediated apoptotic signals can be diminished leading to the development of resistance, which represents a major restraint of cisplatin-based chemotherapy (Siddik, 2003).

Among the constituents of DNA damage signalling, mitogen-activated protein kinases (MAPKs) are of particular interest for their diverse functions in modulating cell death machineries. MAPKs are a family of structurally-related serine/threonine protein kinases that coordinate various extracellular stimuli to regulate cell growth and survival (Chang & Karin, 2001; Johnson & Lapadat, 2002; Marshall, 1995). There are three major subfamilies of MAPKs: extracellular signal-regulated kinase (ERK1/2), stress-activated protein kinase (SAPK)/c-Jun N-terminal kinase (JNK) and p38 MAPK (Siddik, 2003).

ERK1 and ERK2 are homologous isoforms that share the same substrate specificity in vitro (Gille et al., 1992). These 44- and 42-kDa proteins that phosphorylate a multitude of protein substrates (Boulton et al., 1990; Boulton et al., 1991) share 85% of amino acid identity. In resting conditions, ERK1/2 is anchored in the cytoplasm by its association with MEK1/2 (Fukuda et al., 1997), the microtubule network (Reszka et al., 1995) or with phosphatases, which contain a nuclear export signal (NES) (Karlsson et al., 2004). Upon stimulation, ERK1/2

becomes phosphorylated at threonine and tyrosine residues resulting in its separation from MEK1/2. Then ERK1/2 translocates into the nucleus by passive diffusion of the monomer (Adachi et al., 1999), active transport of the dimer (Whitehurst et al., 2002) or by a direct interaction of ERK1/2 with the nuclear pore complex (Matsubayashi et al., 2001). Upon translocation, activated ERK1/2 phosphorylates the ternary complex factors Elk-1, Sap-1a and TIF-IA (growth dependent transcription initiation factor) (Chen, R. H. et al., 1992; Lenormand et al., 1993; Zhao, J. et al., 2003). Phosphorylation of Elk-1 enhances transcription of growth-related proteins, such as c-Fos (Marais et al., 1993).

Phosphoprotein enriched in astrocytes -15 kDa (PEA-15) is a small scaffold protein, ubiquitously expressed and highly conserved among mammals (Danziger et al., 1995; Estelles et al., 1996; Ramos et al., 1998). It is involved in the regulation of several cellular functions, including glucose metabolism, cell proliferation, apoptosis and survival (Fiory et al., 2009). Its expression has been shown to be elevated in tumours, including human glioma and mammary carcinomas (Glading et al., 2007; Hao et al., 2001), and in cell lines derived from human larynx, cervix and skin tumours (Formisano et al., 2005; Glading et al., 2007). PEA-15 may function as a tumour promoter or suppressor, regulating both proliferation and apoptosis (Sulzmaier, F. et al., 2012).

PEA-15 has two phosphorylation sites at Ser104 and Ser116 and is preferentially phosphorylated by protein kinase C (PKC) at Ser104 and by calcium/calmodulin-dependent protein kinase II (CaMKII) or Akt at Ser116 (Araujo et al., 1993; Estelles et al., 1996; Kubes et al., 1998). PEA-15 can bind both ERK1/2 and phosphorylated ERK1/2 (p-ERK1/2) with equal affinity (Callaway et al., 2007). PEA-15 phosphorylation releases ERK1/2 resulting in its translocation to the nucleus and activation of the nuclear transcription factor Elk-1 as well as other transcription factors promoting cell proliferation. As PEA-15 has a nuclear export sequence, it is almost exclusively confined to the cytoplasm (Formstecher et al., 2001). PEA-15-mediated cytoplasmic sequestration of ERK1/2 was reported to suppress tumourigenicity in ovarian cancer by diminishing the activity of Elk-1 (Bartholomeusz et al., 2006). Another study found that PEA-15 expression inhibited cell proliferation by autophagy involving ERK1/2 activation (Bartholomeusz et al., 2008). The functions of PEA-15 are tightly regulated by its phosphorylation status (Fiory et al., 2009). The impact of PEA-15 phosphorylation was investigated by Lee et al. (Lee, J. et al., 2012) in ovarian cancer tissue samples revealing that tissues from high-grade ovarian tumour were significantly more likely than adjacent normal tissues to express PEA-15 phosphorylated at both sites. The authors used phosphomimetic and non-phosphorylatable PEA-15 mutants where the two serine residues of PEA-15 at 104

and 116 positions were either replaced by two aspartic acid (PEA-15DD) or two alanine (PEA-15AA) residues, respectively. The non-phosphorylatable PEA-15AA exerted a more pronounced antitumourigenic effect in ovarian cancer than did the phosphomimetic PEA-15DD (Lee, J. et al., 2012). To study the role of PEA-15 phosphorylation in paclitaxel sensitivity in ovarian cancer, PEA-15AA and PEA-15DD were overexpressed in SKOV-3.ip1, OVTOKO and HEY cells. All three cell lines showed enhanced sensitivity to paclitaxel when phosphomimetic PEA-15DD was overexpressed, while nonphosphorylatable PEA-15AA augmented resistance to paclitaxel (Xie et al., 2013).

The aim of this work was to investigate the influence of PEA-15 and its phosphorylation status on cisplatin sensitivity in ovarian carcinoma cells. We show that the non-phosphorylatable PEA-15AA sensitises ovarian cancer cells to cisplatin. The results warrant further evaluation of the phosphorylation state of PEA-15 in order to consider it as a potential biomarker of tumour sensitivity to cisplatin.

2.2. Methods

2.2.1. Chemicals

3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) was purchased from AppliChem (Darmstadt, Germany), dimethylsulfoxide (DMSO) was ordered from Riedel-de Haën, (Seelze, Germany), bovine serum albumin (BSA), all-trans-retinoic acid and cisplatin were obtained from Sigma-Aldrich (Steinheim, Germany), foetal calf serum, penicillin-streptomycin solution, IMDM™ Medium, trypsin-EDTA solution were ordered from PAN™ Biotech (Aidenbach, Germany) and ultrapure water was obtained using a Purelab Plus ™ system from ELGA Labwater (Celle, Germany).

2.2.2. Cell Lines and Cell Culture

The SKOV-3 ovarian carcinoma cell line (ATCC® HTB77™) was from American Type Culture Collection (ATCC). The cisplatin-resistant ovarian carcinoma cell line EFO27^{r}CDDP^{2000} was derived from the Resistant Cancer Cell Line (RCCL) collection (www.kent.ac.uk/stms/cmp/RCCL/RCCLabout.html). Cell backups were frozen with 10% DMSO. Cells were grown as monolayers in IMDM™ medium supplemented with 10% foetal calf serum, 100 IU/mL penicillin, 0.1 mg/mL streptomycin in a humidified atmosphere containing 5% CO2. The medium of EFO27^{r}CDDP^{2000} cells was supplemented with 2 µg/mL

cisplatin. Every ten passages a new backup of cells was thawed in order to avoid alterations in cell features during cultivation.

2.2.3. Plasmid Transfection

Cells were transfected with plasmids of interest using K2® transfection system (Biontex Laboratories GmbH, Munich, Germany) according to the manufacturer's instruction. Plasmids used in this study include three DNA constructs of pcDNA3-HAR36 containing two mutated PEA-15 protein versions and the empty vector. These plasmids were kindly provided by Prof. Naoto T. Ueno (Breast Medical Oncology, MD Anderson Cancer Center, University of Texas, Houston, TX, USA). The constructs were originally made by Ramos et al. (Ramos et al., 1998). The two mutated versions of PEA-15 protein were PEA-15AA, in which two serine residues at 104 and 116 positions were replaced with two alanine (A) residues, and PEA-15DD, in which the same serine residues were replaced with aspartic acid (D). Cells were used for protein expression and chemosensitivity analysis 48 h after transfection.

2.2.4. Cell Fractionation

Nuclear/cytosol fractionation kit (Biovision Inc., Milpitas, CA, USA) was used for separating the cytosolic and the nuclear fractions of the transfected cells according to the manufacturer's protocol. All steps of the fractionation were performed on ice. To analyse the efficiency of the fractionation, Western blot analysis was performed for each fraction (nuclear or cytosolic) as described below. As a marker for the cytosolic fraction, GAPDH was used. For the nuclear fraction, Lamin B1 was used as an indicator of the purity of the nuclear fraction. No cross-contamination was assumed when the indicator was detected only in the expected fraction.

2.2.5. Western Blot

Cells were fractionated as mentioned above or lysed in lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 0.1% SDS, 1 mM NaF, 2 mM Na_3VO_4, 70% NP-40, 0.5% sodium deoxycholate, 6.24 mM benzamidine and 0.5 mM PMSF, all from Sigma-Aldrich, Steinheim, Germany) for 30 min on ice. Then the cells were sonicated on ice using a sonicator with the settings: 60% power, 5 s pulse, 30 s interval, 3 pulses per sample. The suspensions were centrifuged for 5 min at 18,620× *g* at 4 °C, and the protein content in the supernatants was measured using the bicinchoninic acid assay (BCA™ Protein Assay Kit, Pierce, Rockford, IL, USA) (Smith, P. K. et al., 1985). Samples containing 30 µg total protein were subjected to electrophoresis in 12% SDS-polyacrylamide gel and transferred to a PVDF membrane (Carl Roth GmbH, Karlsruhe, Germany). The membranes were blocked in 5% milk powder in TBS-T (0.2% Tween-20) for 1

h, rinsed three times with TBS-T and incubated at 4 °C overnight with primary antibodies diluted in TBS-T with 5% BSA. After washing three times with TBS-T, incubation with the secondary antibody diluted 1:1000 in TBS-T with 5% milk powder for 1.5 h followed. The monoclonal mouse antibody against UGT1A (diluted 1:1000) sc-271268, the polyclonal rabbit antibody against Nrf2 (diluted 1:500) sc-722 were from Santa Cruz Biotechnology, Heidelberg, Germany, the monoclonal mouse antibody against HA MMS-101-P (diluted 1:500) was received from Covance Inc., PA, USA. The secondary HRP-conjugated goat anti-rabbit (diluted 1:1000) SBA-4030-05 was obtained from Biozol Diagnostica Vertrieb GmbH, Eching, Germany, the rabbit polyclonal antibody against p-ERK1/2 (Thr202/Tyr204) 9101 (diluted 1:1000) was ordered from Cell Signaling Technology Europe B.V., Frankfurt (Main), Germany, and the Peroxidase AffiniPure goat anti-mouse (diluted 1:5000) 115-035-003 was from Jackson ImmunoResearch Europe Ltd., Cambridgeshire, UK. The detection was performed using a Molecular Imager ChemiDoc™ XRS⁺ System from Bio-Rad Laboratories GmbH, Munich, Germany. After subsequent triple washing with TBS-T, the membranes were incubated for 30 min. with the rabbit antibody against GAPDH (GTX100118, Biozol Diagnostica Vertrieb GmbH, Eching, Germany) diluted 1:20000 or for 1 h with the rabbit polyclonal antibody to Lamin B1 (GTX-103292, Biozol Diagnostica Vertrieb GmbH, Eching, Germany) diluted 1:1000 in TBS-T with 5% BSA. After rinsing with TBS-T, the incubation with the secondary antibody and detection were performed as described above. Densitometric analysis was performed using ImageLab™ 5.1 software (Bio-Rad Laboratories, Hercules, CA, USA).

2.2.6. MTT Assay

Cell sensitivity to cisplatin, retinoic acid or their combination was assessed using an MTT-based assay (Alley et al., 1988). In brief, cells were seeded in 96-well microtiter plates (1×10^4 cells/well) and allowed to attach overnight. Then medium was removed and nine subsequent dilutions of retinoic acid or cisplatin in medium were added to the cells in triplicate (100 µL/well). For the combination treatment, cisplatin dilutions each contained 20 µM retinoic acid. After 47 h of incubation, 20 µL of a 5 mg/mL MTT solution in phosphate buffered saline (PBS) was added to each well, and the cells were incubated at 37 °C for 1 h. The supernatant was discarded, and the formazan crystals formed were dissolved in 100 µL DMSO. Absorbance of the dye was measured at 570 nm with background subtraction at 690 nm using a Multiskan Ascent® microtiter plate reader (Thermo Fisher Scientific, Langenselbold, Germany). The results were analysed and the pEC_{50} values (pEC_{50} = -log EC_{50}, EC_{50} = half maximal effective concentration) were estimated for each independent

experiment with the GraphPad Prism™ 6 analysis software package (GraphPad Software, San Diego, CA, USA) using non-linear regression (sigmoidal dose response, variable slope). The mean pEC_{50} values were calculated from the results of several independent experiments and used for the determination of the respective EC_{50} values.

2.2.7. cDNA Microarray Analysis

SKOV-3 cells were transfected with PEA-15-HA- (empty vector, EV), PEA-15AA- or PEA-15DD-containing plasmids, respectively, as described above. Twenty-four hours after transfection, cells were treated with 15 µM cisplatin (this concentration corresponds to the EC_{50} of cisplatin in EV-transfected cells as measured after 48 h of incubation) for 24 h and then the total RNA of the cells was extracted using my-Budget RNA mini kit following the instructions provided by the manufacturer. Next generation transcriptome-wide gene-level expression profiling using Clariom™ S assay was performed by Life and Brain GmbH (Bonn, Germany). The microarray data have been deposited into the Gene Expression Omnibus database under the accession number GSE144041.

The raw data of the array were collected as CEL files and analysed using the Transcriptome Analysis Console (TAC 4.1, Thermofisher Scientific, Waltham, MA, USA) software. The gene expression was analysed with the Gene Level Signal Space Transformation-Robust Multi-Chip Analysis (SST-RMA) summarization method. Data obtained from the microarray were normalised by the robust multiarray average method (Irizarry et al., 2003). A probe set was considered expressed if ≥50% samples had DABG (Detected Above Background) values below DABG threshold ($p < 0.05$). Statistical significance of the differences in gene expression was analysed using Limma (Ritchie et al., 2015). Differential expression was assumed at a p value <0.05 and a fold change in expression ≥2 or ≤−2.

2.2.8. Correlation of Tumour Gene Expression Levels with the Survival of Cisplatin-Treated Patients in The Cancer Genome Atlas (TCGA)

Gene expression data from patient tumours was derived from The Cancer Genome Atlas (TCGA) (Liu, J. et al., 2018; The Cancer Genome Atlas Research Network, 2008) via the GDC Data Portal (https://portal.gdc.cancer.gov). The Bioconductor R package TCGAbiolinks was used to obtain corresponding clinical data. Tumour gene expression data and cisplatin response data were available for 779 patients representing 23 different cancer types (adrenocortical carcinoma, n = 2; bladder urothelial carcinoma, n = 78; breast invasive carcinoma, n = 2; cervical squamous cell carcinoma and endocervical adenocarcinoma, n = 122; cholangiocarcinoma, n = 4; lymphoid neoplasm diffuse large b-cell lymphoma, n = 2;

oesophageal carcinoma, *n* = 14; glioblastoma multiforme, *n* = 7; head and neck squamous cell carcinoma, *n* = 94; kidney renal papillary cell carcinoma, *n* = 1; liver hepatocellular carcinoma, *n* = 4; lung adenocarcinoma, *n* = 84; lung squamous cell carcinoma, *n* = 72; mesothelioma, *n* = 34; ovarian serous cystadenocarcinoma, *n* = 115; pancreatic adenocarcinoma, *n* = 2; sarcoma, *n* = 3; skin cutaneous melanoma, *n* = 9; stomach adenocarcinoma, *n* = 44; testicular germ cell tumours, *n* = 53; thymoma, *n* = 5; uterine corpus endometrial carcinoma, *n* = 21; uterine carcinosarcoma, *n* = 7). Overall survival (OS) was defined as days to last follow-up or death as previously described (Ng, S. W. et al., 2016).

Genes whose expression was undetected (FPKM = 0, FPKM – fragments per kilobase of exon model per million reads mapped) in >10% samples were removed from the investigation as previously described (Cai, L. et al., 2017). Patients were stratified by cancer type, and the median expression for each gene was calculated. Genes with expression values above the median were considered highly expressed, while those below the median were considered to have low expression. Patients were then re-amalgamated before being stratified by drug treatment for pan-cancer analysis.

Cox proportional hazards regression was used to calculate the hazard ratio for cohorts expressing high (above-median) vs. low (below-median) levels of a given gene. Calculations were performed using the R survminer and survival packages. p-values in each case are the result of a log rank (Mantel-Cox) test, which assesses whether there is a significant difference between the survival of two independent groups. Hazard ratios quoted refer to values for 'low' (below median) expression for each gene in the model, with values >1 indicating increased hazard (i.e., reduced OS) and values <1 indicating decreased hazard (i.e., increased OS). Multiple test correction was again performed for each of the samples (Benjamini-Hochberg method (Benjamini & Hochberg, 1995)). The false discovery rate (FDR) values are thresholds that indicate whether p-values are below or above the respective significance levels.

2.2.9. Statistical Analysis

GraphPad Prism™ 6 was used for analysing and plotting the data. Protein expression data were compared using one-way ANOVA with a Holm-Sidak post-test. Unpaired *t*-test was used to compare pEC_{50} values. *p* values of <0.05 were considered significant.

2.3. Results

2.3.1. PEA-15AA Sensitised SKOV-3 Cells to Cisplatin

In order to assess if the phosphorylation status of PEA-15 affects cisplatin sensitivity, the non-phosphorylatable PEA-15AA (further AA) and the phosphomimetic PEA-15DD (further DD) were overexpressed in SKOV-3 cells by liposome-mediated transient transfection. Control cells were transfected with the empty vector (further EV). The efficiency of the transfection was confirmed by Western blot (Figure 2.1A). PEA-15AA overexpression significantly increased the sensitivity of SKOV-3 cells to cisplatin compared to the empty vector-transfected cells (Figure 2.1B). This was not the case upon PEA-15DD overexpression. The EC_{50} in SKOV-3-AA cells was lower (EC_{50} = 9.9 µM) than in SKOV-3-EV cells (14.9 µM), while SKOV-3-DD cells (EC_{50} = 14.3 µM) showed similar cisplatin sensitivity as the SKOV-3-EV controls.

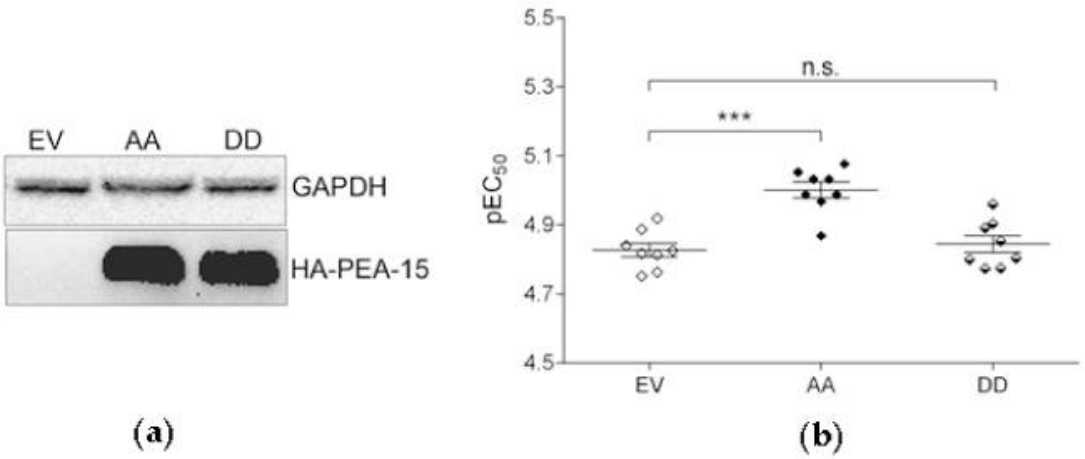


Figure 2.1. (**A**) Expression of hemagglutinin (HA)-tagged phosphoprotein enriched in astrocytes (PEA-15) in SKOV-3 cells after transfection with the HA-tagged empty vector (EV), PEA-15AA (AA) and PEA-15DD (DD). GAPDH was used as a loading control. (**B**) Cisplatin sensitivity (pEC_{50}, mean ± SEM, *n* = 8) of transfected SKOV-3-EV (EV), SKOV-3-AA (AA) and SKOV-3-DD (DD) cells. *** $p < 0.001$, n.s. = not significant.

In addition, we confirmed the effect of PEA-15AA on cisplatin sensitivity in the cisplatin-resistant EFO27^{r}CDDP^{2000} ovarian carcinoma cell line. Non-phosphorylatable PEA-15 significantly sensitised EFO27^{r}CDDP^{2000} cells to the platinum drug (Figure S2.1). The EC_{50} value after PEA-15AA transfection was 25.1 µM and thus lower than that of 33.1 µM in cells transfected with the empty vector. Overexpression of PEA-15DD did not have any significant effect (EC_{50} = 30.9 µM) compared to the empty-vector control.

2.3.2. PEA-15AA-Transfected SKOV-3 Cells Contain more Cytosolic p-ERK1/2 than PEA-15DD-Transfected Cells

EV-, PEA-15AA- and PEA-15DD-transfected SKOV-3 cells were fractionated to separate the nuclear and cytosolic fractions. Figure 2.2 presents relative expression of activated ERK in the cytosolic and nuclear fractions of the same samples following transfection with different mutants of PEA-15, loaded on the same gel. In SKOV-3-AA cells, the cytosolic fraction contained more p-ERK1/2 than the corresponding nuclear fraction of the same cell lysate, while in SKOV-3-DD cells the opposite was observed (Figure 2.2). This provides a proof of concept that PEA-15 unphosphorylated at both Ser104 and Ser116 (PEA-15AA) retains p-ERK1/2 in the cytoplasm of SKOV-3-AA cells while the phosphomimetic PEA-15DD does not keep p-ERK1/2 in the cytoplasm, promoting an increase in p-ERK1/2 nuclear accumulation in SKOV-3-DD cells.

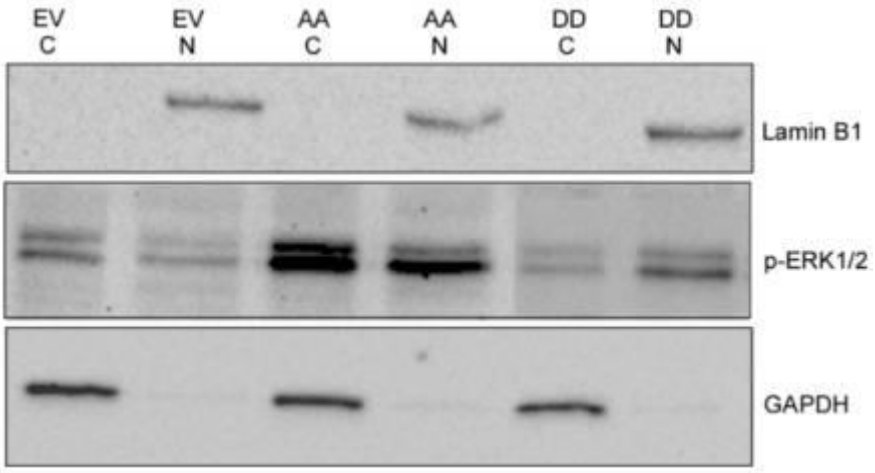


Figure 2.2. A representative Western blot of phosphorylated extracellular signal-regulated kinase1/2 (p-ERK1/2) expression in nuclear and cytosolic fractions of the SKOV-3 cells transfected with empty vector (EV), PEA-15AA (AA) and PEA-15DD (DD). GAPDH and Lamin B1 were used as the markers and loading controls of cytosolic (C) and nuclear fractions (N), respectively.

2.3.3. Differentially Expressed Genes in Transfected Cells

As PEA-15 has many other functions inside the cells beside controlling ERK1/2 localisation, a microarray analysis was warranted to identify genes, which may account for enhanced cisplatin sensitivity in PEA-15AA-transfected cells. Clariom^{TM} S assay was performed to analyse the differences in gene expression in untreated and cisplatin-treated SKOV-3-EV, SKOV-3-AA and SKOV-3-DD cells. Genes were considered to be differentially expressed when the fold change was ≥2 and the *p* value was <0.05. The resulting heat map of regulated genes shows clear clustering between treatment conditions (Figure 2.3).

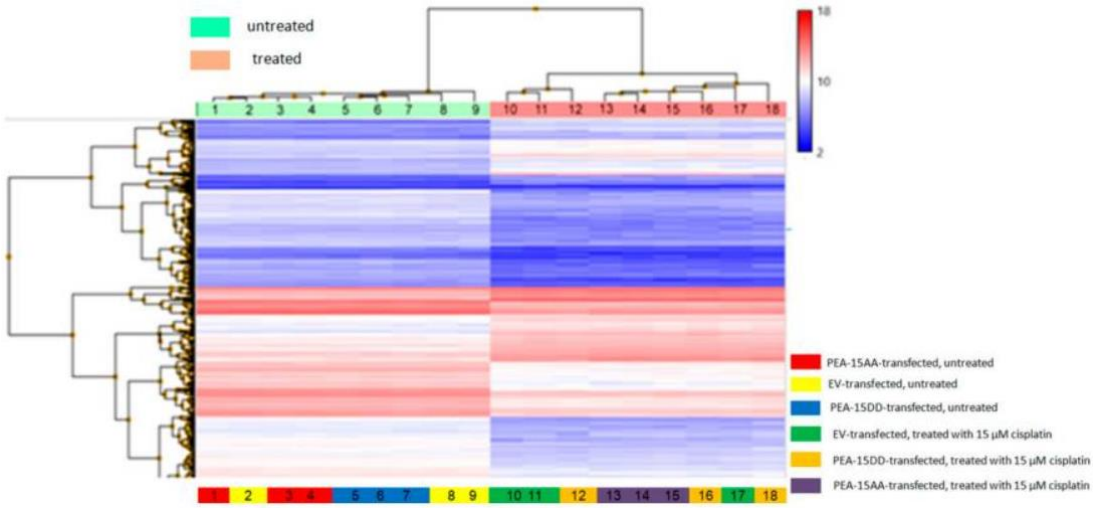


Figure 2.3. Heatmap of the transcriptome-wide Clariom^{TM} S array, regulated genes with fold change cut-off at 2.0 for differentially expressed genes and a *p*-value cut-off at 0.05 are shown.

Table 2.1 shows the number of differentially expressed genes in EV-, PEA-15AA- and PEA-15DD-transfected cells after cisplatin exposure and in untreated transfected cells. Prior to cisplatin exposure, only three genes were differentially regulated between untreated SKOV-3-EV and untreated SKOV-3-AA. Between SKOV-3-EV and SKOV-3-DD, the number of differentially expressed genes was 18, while for SKOV-3-AA and SKOV-3-DD the number was 10. Following cisplatin treatment, 4430 genes were differentially regulated in EV-transfected cells, while in PEA-15AA- and PEA-15DD-transfected cells the numbers were 4196 and 4110, respectively.

Table 2.1. Number of differentially expressed genes, compared as treatment condition 1 vs. condition 2 with at least two-fold up- or down-regulation with $p < 0.05$.

Treatment Condition 1	Treatment Condition 2	Number of Genes
SKOV-3-EV, untreated	SKOV-3- AA, untreated	3
SKOV-3-EV, untreated	SKOV-3-DD, untreated	18
SKOV-3-AA, untreated	SKOV-3- DD, untreated	10
SKOV-3-EV, untreated	SKOV-3-EV, 15 µM cisplatin, 24 h	4430
SKOV-3-AA, untreated	SKOV-3-AA, 15 µM cisplatin, 24 h	4197
SKOV-3-DD, untreated	SKOV-3-DD, 15 µM cisplatin, 24 h	4110

2.3.4. Correlation of Genes Differentially Expressed in the Comparisons Untreated SKOV-3-AA vs. SKOV-3-EV and SKOV-3-AA vs. SKOV-3-DD Cells with the Survival of Cisplatin-Treated Patients in The Cancer Genome Atlas (TCGA)

Next, we selected genes, which were differentially regulated in comparisons: SKOV-3-AA vs. SKOV-3-EV and SKOV-3-AA vs. SKOV-3-DD. Then we checked the expression levels of those 13 genes in the tumours of cisplatin-treated cancer patients and correlated them to patient survival (Table 2.2).

Table 2.2. Correlation of the expression of genes, differentially expressed between SKOV-3-AA and SKOV-3-EV or SKOV-3-DD cells, in tumours of cisplatin-treated patients (n = 779) and patient survival.

Gene	Fold Change	Hazard Ratio [1]	*p*-Value	FDR (thr. 0.2) [2]	FDR (thr. 0.05)
EGR1	−2.22	0.7068	0.00322 **	0.034099	0.008525
NAV3	−2.18	0.7062	0.00317 **	0.033948	0.008487
GPRC5C	2.14	1.4745	0.000971 **	0.023936	0.005984

Gene	Fold change	Hazard ratio 1	p-value	FDR (thr. 0.2) 2	FDR (thr. 0.05)
TSTD3	2.16	0.7545	0.0162 *	0.053933	0.013483
ELFN2	2.09	0.9036	0.385	0.139261	0.034815
LOC100287225	−2.37	n.a. [3]	n.a.	n.a.	n.a.
TIPARP	−2.19	0.7505	0.015 *	0.052694	0.013174
PRKG1	−2.00	0.7532	0.0154 *	0.053205	0.013301
ND6	−2.07	0.8492	0.163	0.10641	0.026602
RBM26	−2.42	0.8951	0.342	0.133674	0.033418
FLNA	−2.86	0.7523	0.0152 *	0.052997	0.013249
MAVS	−2.24	0.6656	0.000569 **	0.020231	0.005058
GTSF1L	−2.07	n.a.	n.a.	n.a.	n.a.

[1] Hazard ratio at low gene expression levels in tumour tissue; [2] thr. = false discovery rate threshold determined according to Benjamini-Hochberg (Benjamini & Hochberg, 1995); [3] n.a. = not applicable; * = *p*-value below 0.05 and significant at FDR = 0.2; ** = *p*-value below 0.05 and significant at FDR = 0.05.

Patient survival is expressed as the hazard ratio at low (below median) expression of the respective gene in tumour tissue. A hazard ratio >1 means that the overall survival is reduced in patients with tumours that display low expression of the respective gene. A hazard ratio <1 indicates prolonged survival in patients, whose tumours display low expression of the respective genes. When we prepared a heatmap, in which we directly compared the effect of low expression of the investigated genes on cisplatin sensitivity or patient survival, there was substantial overlap as illustrated in Figure 2.4. The expression levels of nine of the eleven genes, for which gene expression data was available in the TCGA, were associated with beneficial (low EC_{50}, prolonged survival) or poor (high EC_{50}, shorter survival) outcome in the same way in both datasets (Table 2.2, Figure 2.4). There was a significant correlation (at a FDR of 0.2) for seven of these nine genes between their expression level in tumour tissue

and patient survival: early growth response protein 1 (*EGR1*), neuron navigator 3 (*NAV3*), G Protein-Coupled Receptor Class C Group 5 Member C (*GPRC5C*), TCDD-inducible poly [ADP-ribose] polymerase (*TIPARP*), cGMP-dependent protein kinase 1 (*PRKG1*), filamin A (*FLNA*), mitochondrial antiviral-signalling protein (*MAVS*).

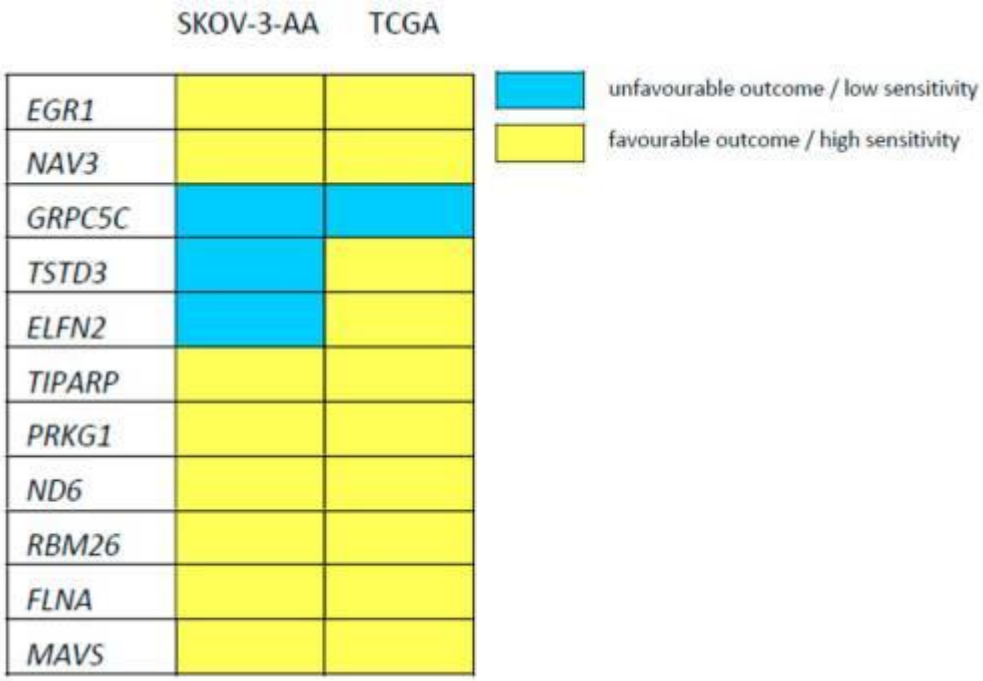


Figure 2.4. Heatmap indicating the relationship between low expression of the indicated genes and sensitivity/outcome, favourable (low cisplatin EC_{50} in SKOV-3-AA cells or prolonged survival of cisplatin-treated patients, indicated in yellow) or unfavourable (high cisplatin EC_{50} in SKOV-3-AA cells or reduced survival of cisplatin-treated patients, indicated in blue), based on the comparison of gene expression between SKOV-3-AA and EV- or PEA-15DD-transfected variants and TCGA data.

2.3.5. Pathway Analysis for the Genes Exclusively Regulated in SKOV-3-AA Cells Following Cisplatin Treatment

Since cisplatin treatment had a major impact on gene expression, we created a Venn diagram to define the genes exclusively regulated upon cisplatin treatment in different transfected cells. For all three cell types, 3039 differentially regulated genes were common. However, 717 genes were exclusively regulated after cisplatin exposure in SKOV-3-EV cells, while in SKOV-3-AA and SKOV-3-DD cells 444 genes and 383 genes were exclusively differentially expressed, respectively (Figure 2.5). The 444 genes that were found to be exclusively regulated in SKOV-3-AA cells upon cisplatin exposure may contribute to the increased

cisplatin sensitivity in these cells. A pathway analysis based on Wikipathways revealed 21 pathways to be significantly affected in response to cisplatin treatment in SKOV-3-AA cells (Figure 2.6).

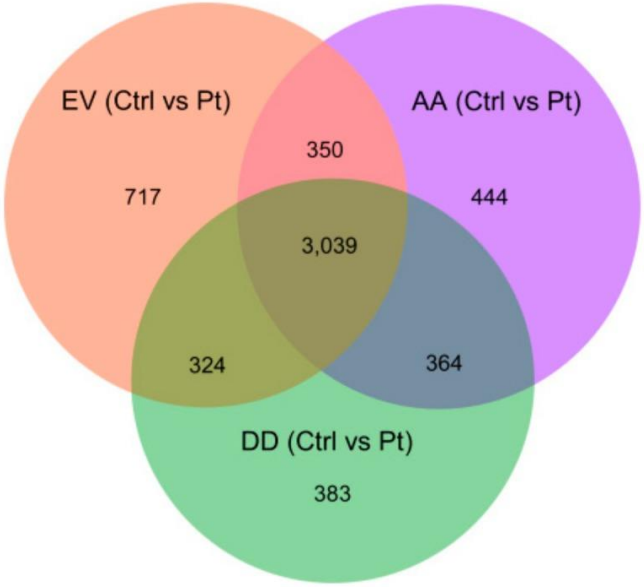


Figure 2.5. Venn diagram representing the exclusively and commonly regulated genes in different transfected cells upon cisplatin exposure. The diagram shows the total number of genes affected by cisplatin exposure in empty vector—(EV), PEA-15AA—(AA) and PEA-15DD-transfected—(DD) cells.

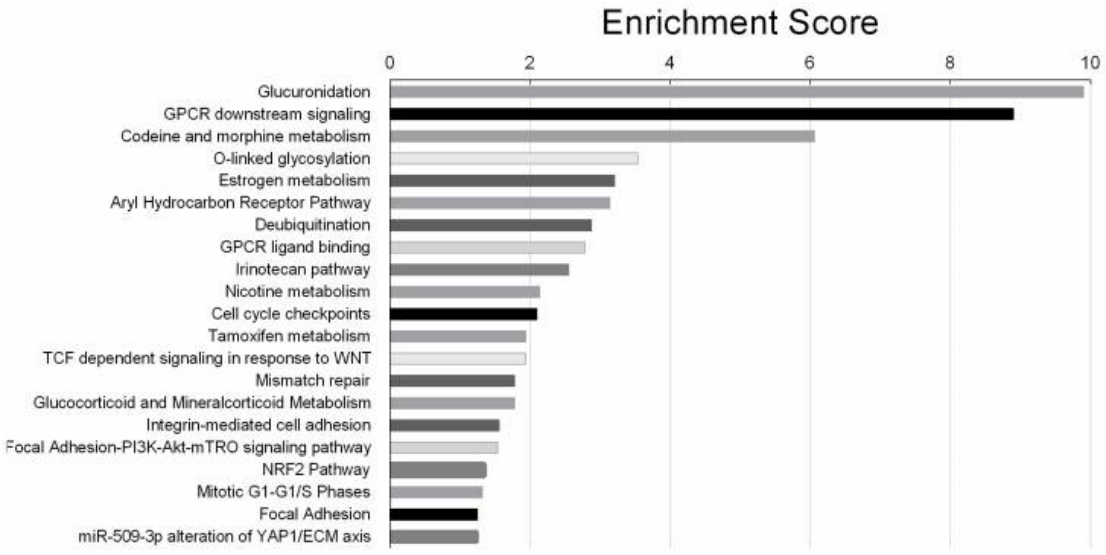


Figure 2.6. Twenty-one biological pathways significantly affected by cisplatin treatment in SKOV-3-AA cells, listed according to the significance level (log 2 base) in a descending order.

The most significantly affected pathway was the glucuronidation pathway with a significance value of 9.9 (log 2 base). Glucuronidation is the process of metabolizing substances such as drugs, pollutants, bilirubin, androgens, estrogens, glucocorticoids, fatty acids and bile acids. In glucuronidation process, the glucuronic acid of a uridine diphosphate (UDP) glucuronic acid is transferred to a substrate by UDP-glucuronyl transferase (UGT). The resulting substrates, glucuronides are more soluble in water and are excreted from body with urine and faeces. Among the other significant pathways, the Nrf2 pathway is of importance, as it is an upstream regulator of the UGTs. In addition, this pathway was previously found to influence cisplatin sensitivity in cancer cells (Furfaro et al., 2015).

2.3.6. Evaluation of the Responsible Genes Within the Affected Pathways

2.3.6.1. UGT1A and Nrf2 Pathway

All ten *UGT1A* isoforms were downregulated in SKOV-3-AA cells after cisplatin exposure according to the microarray data. Western blot analysis indicated that the expression of UGT1A protein in EV-, PEA-15AA- and PEA-15DD-transfected cells decreased in response to cisplatin to 55.63% (p = 0.02), 38.65% (p = 0.0013) and 48.27% (p = 0.01) of their basal level in untreated cells, respectively (Figure 2.7A). The greatest extent of reduction after cisplatin exposure was thus observed after PEA-15AA transfection (61.35% in SKOV-3-AA compared to 44.37% for SKOV-3-EV and 51.73% for SKOV-3-DD, along with much lower p-value).

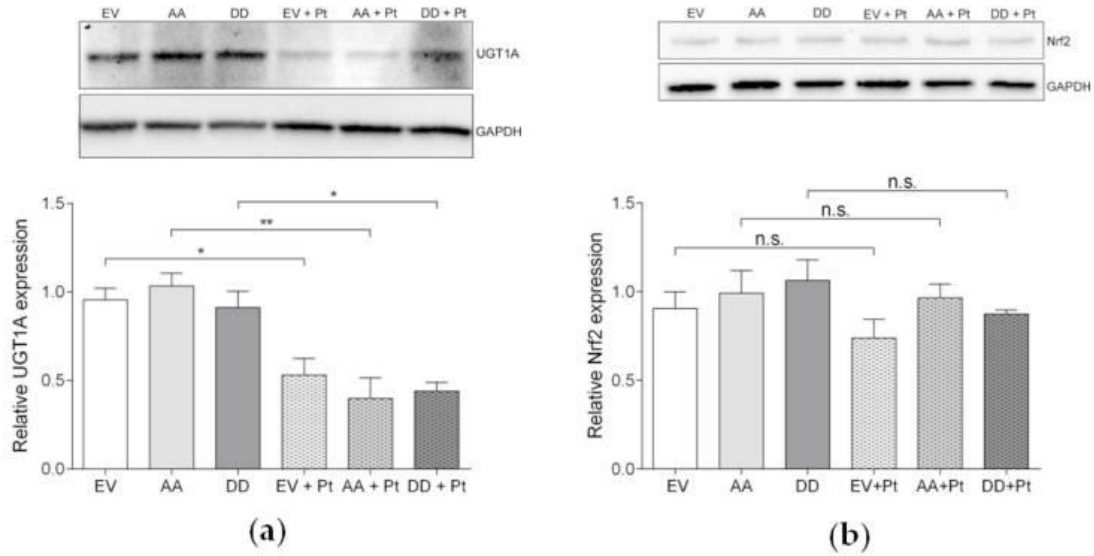


Figure 2.7. Representative Western blots and the corresponding densitometric quantification (mean ± SEM, *n* = 3) of (**A**) the relative uridine diphosphate-glucuronyl transferase (UGT)1A expression and (**B**) the relative nuclear factor erythroid 2-related factor 2 (Nrf2) expression in empty vector—(EV), PEA-15AA—(AA) and PEA-15DD-transfected—(DD) cells after treatment with 15 µM cisplatin (+Pt) for 24 h and in untreated transfected SKOV-3 cells. GAPDH was used as a loading control. * $p < 0.05$, ** $p < 0.01$.

The *UGT1A* family is also annotated in other pathways significantly affected by cisplatin treatment in SKOV-3-AA cells, including codeine and morphine metabolism, estrogen metabolism, aryl hydrocarbon receptor pathway, irinotecan pathway, nicotine metabolism, tamoxifen metabolism and Nrf2 pathway. Nrf2 is the upstream regulator of UGT1A expression. The array did not show any differential regulation of *Nrf2*, but some downstream genes of Nrf2 were differentially expressed. Therefore, the expression of Nrf2 was analysed on the protein level by Western blot in transfected cells before and after exposure to cisplatin (Figure 2.7B). The results showed a slight decrease in the overall Nrf2 levels after cisplatin exposure in all investigated cells. However, the differences were not statistically significant conforming to the result of the array.

2.3.6.2. Retinoic Acid, an Inhibitor of Nrf2/ARE Pathway, Increases Cisplatin Sensitivity

Since UGT1A expression was affected by cisplatin in SKOV-3-AA cells, while Nrf2 expression was not, we hypothesised that Nrf2-associated downstream signalling via the Nrf2/ARE pathway may contribute to the sensitisation of PEA-15AA-transfected cells to cisplatin. All-trans retinoic acid (further retinoic acid), which is known to reduce the Nrf2-mediated induction of ARE-driven genes (Atencia et al., 1994), was used to investigate this phenomenon further.

The EC_{50} value of retinoic acid was 216 µM (pEC_{50} = 3.660 ± 0.003, mean ± SEM, *n* = 6) after 48 h of incubation. In order to investigate the influence of the compound on cisplatin sensitivity, 20 µM of retinoic acid was used in combination with cisplatin in SKOV-3 cells as this concentration was lower than the EC_{10} of retinoic acid and therefore not toxic to the cells.

Retinoic acid did not affect Nrf2 protein levels in all investigated cells independent of cisplatin treatment (Figure 2.8A). However, retinoic acid significantly reduced UGT1A levels alone and in combination with cisplatin (Figure 2.8B). This suggests that retinoic acid interferes with UGT1A expression via effects on Nrf2 signalling (but without directly affecting

Nrf2 levels) similarly to PEA-15AA overexpression. In agreement, retinoic acid also significantly sensitised SKOV-3 cells to cisplatin (Figure 2.9). The EC_{50} value for cisplatin decreased from 32.6 µM to 13.9 µM in the presence of 20 µM retinoic acid. It should be noted that this experiment was conducted with non-transfected SKOV-3 cells, which cannot be directly compared with empty vector-transfected cells. Generally, after transfection process, cells became more sensitive to cisplatin. A similar effect of transfection procedure was also noticed in EFO27^{r}CDDP^{2000} cells (Figure S2.1).

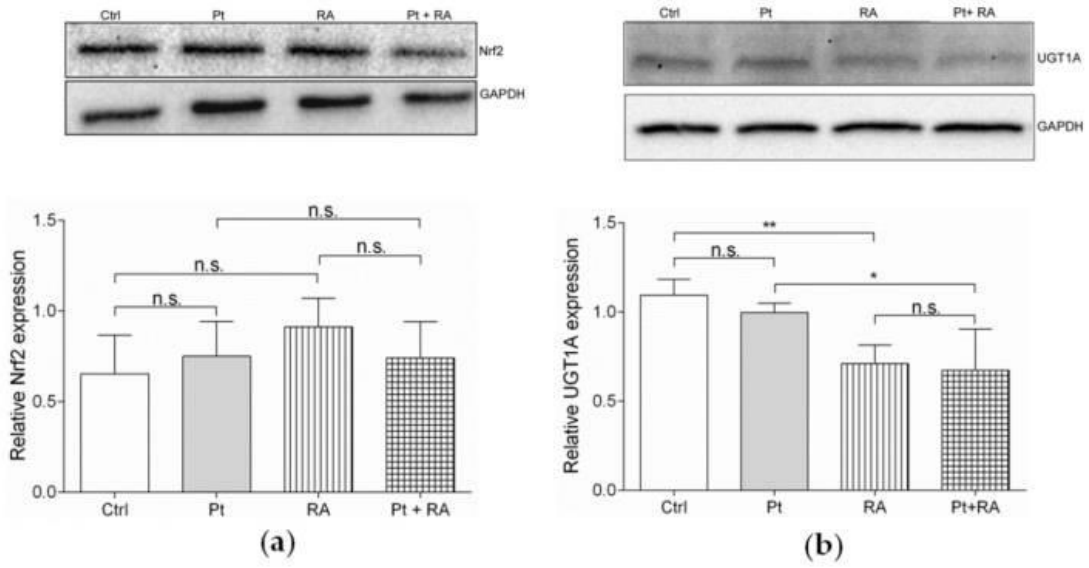


Figure 2.8. Representative Western blots and the corresponding densitometric quantification (mean ± SEM, *n* = 3) of (**A**) the relative Nrf2 expression and (**B**) the relative UGT1A expression in the transfected untreated SKOV-3 cells (Ctrl), after exposure to 15 µM cisplatin (Pt), to 20 µM retinoic acid (RA) and after co-incubation with 20 µM retinoic acid and 15 µM cisplatin (Pt + RA) for 24 h are shown. GAPDH was used as a loading control. *$p < 0.05$, ** $p < 0.01$, n.s. = not significant.

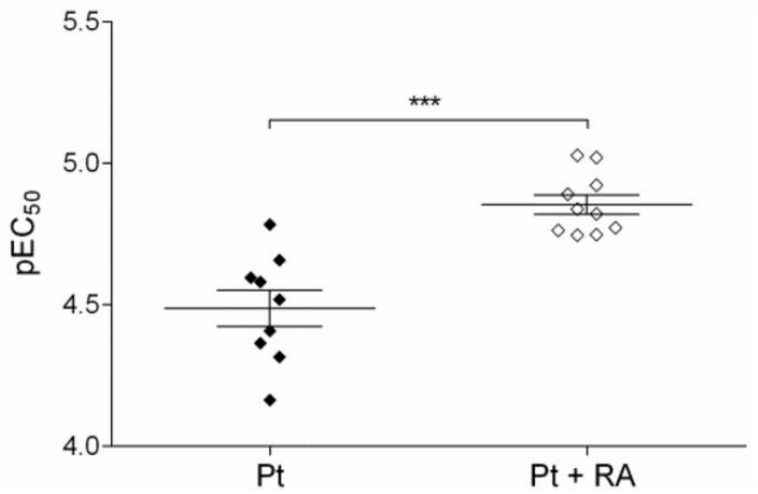


Figure 2.9. Sensitivity of SKOV-3 cells (pEC_{50}, mean ± SEM, *n* = 9–10) of cisplatin alone (Pt), and upon co-incubation with 20 µM retinoic acid (Pt + RA) was determined over 48 h. *** $p < 0.001$.

2.4. Discussion

Our results show that overexpression of PEA-15AA, a PEA-15 variant that cannot be phosphorylated at Ser104 and Ser116, increases the sensitivity of SKOV-3 ovarian cancer cells to cisplatin. The modest absolute effect of PEA-15 phosphorylation status on pEC50 may result from counteracting signalling pathways. Nevertheless, the finding was confirmed in an additional cell line model.

Previously, PEA-15 overexpression was found to induce cisplatin resistance and knockdown of PEA-15 was shown to enhance sensitivity to cisplatin in colorectal carcinoma cell line (Funke et al., 2013). In that case, the wild-type PEA-15 was overexpressed, with both regulatory serine residues (Ser104 and Ser116) open to be phosphorylated. Therefore, the observed effects cannot be attributed to a single factor, i.e., overexpression of PEA-15. Our results indicate that PEA-15 phosphorylation status critically determines the cancer cell response to cisplatin. Non-phosphorylatable PEA-15AA had been previously found to inhibit ovarian cancer tumourigenicity and progression by blocking beta-catenin (Lee, J. et al., 2012). However, we did not observe direct effects of PEA-15AA on cell growth kinetics. PEA-15 phosphorylation at Ser116 rendered glioblastoma cells resistant to glucose-deprivation mediated cell death (Eckert et al., 2008), which is in line with our finding that unphosphorylatable PEA-15 can potentiate cytotoxic effects. However, it appears to be not the case for all kinds of toxic stimuli. Interestingly, PEA-15AA had been shown to reduce

ovarian cancer cell sensitivity to paclitaxel (Xie et al., 2013), a drug that interferes with microtubule disassembly (Xiao et al., 2006), whereas cisplatin acts in a different fashion directly inducing DNA damage (Dilruba & Kalayda, 2016). PEA-15 phosphorylation status affects several signalling pathways and apparently, depending on the mode of action, different downstream events gain more importance accounting for the diverging effects on cell sensitivity to different drugs.

In agreement with previous reports, PEA-15AA retained activated ERK1/2 in the cytoplasm. However, due to the diverse cellular functions of PEA-15, a microarray analysis was warranted to investigate the underlying key players of sensitisation of SKOV-3 cells to cisplatin by PEA-15AA.

In untreated cells, PEA-15-AA overexpression resulted in the differential regulation of thirteen genes between SKOV-3-AA and SKOV-3-EV or SKOV-3-DD cells. For eleven of these genes, data from cisplatin-treated patients was available in TCGA, and the expression of seven genes significantly correlated with the outcome of cisplatin therapy in the same way as expected from our data. Some of these genes are of particular interest in the context of ERK1/2 signalling. Early growth response protein 1 (*EGR1*) is a downstream nuclear target of Elk-1, which itself is regulated by activated ERK1/2 upon nuclear translocation (Gregg & Fraizer, 2011; Shi, Q. et al., 2014), suggesting that PEA-15AA-mediated retention of activated ERK1/2 in the cytosol results in reduced *EGR1* expression. Low EGR1 levels have also been reported to correlate with high cisplatin sensitivity in primary glioma cells (Calogero et al., 2011). *FLNA*, coding for filamin A, a non-muscle actin filament cross-linking protein, is a downstream target of the nuclear transcription factor Sp1 (D'Addario et al., 2006), which is also regulated by ERK1/2 (Milanini-Mongiat et al., 2002). Reduced *FLNA* levels were observed in PEA-15AA-transfected cells in this study, and also correlated with increased cisplatin sensitivity in a mouse xenograft model (Yue et al., 2012) and with improved overall survival in patients (Jin, Y. Z. et al., 2016) according to previous reports. Filamin A was also found to promote cancer growth (Savoy & Ghosh, 2013) and to play an important role in repair of DNA damage (Evans & Relling, 1999).

A total of 444 genes were exclusively differentially regulated in cisplatin-treated SKOV-3-AA cells with UGT1A family representing the most significantly affected pathway (glucuronidation). UGTs catalyse the addition of a β-glucuronic acid moiety to a variety of nucleophilic sites of xenobiotics and endogenous compounds including bilirubin, steroids, bile acids, drugs and other carcinogenic and toxic compounds (Xie et al., 2015), and have

been linked to drug resistance in cancer. The downregulation of UGT1A enzymes upon cisplatin exposure is likely to be a consequence of suppression of upstream signalling via Nrf2 pathway, which was also affected in response to cisplatin treatment. However, effects did not involve Nrf2 itself but seemed to be rather mediated by downstream signalling events. Retinoic acid, a known inhibitor of Nrf2/ARE signalling reduced UGT1A levels and increased cisplatin sensitivity of SKOV-3 cells independently of Nrf2 expression in a similar fashion as PEA-15AA overexpression.

PEA-15 was already evaluated for treating advanced breast cancer in mice. To target PEA-15 in advanced breast tumours, Xie et al. developed a breast cancer specific construct (T-VISA) composed of the human telomerase reverse transcriptase (hTERT; T) promoter and a versatile transgene amplification vector VISA (VP16-GAL4-WPRE integrated systemic amplifier). T-VISA-PEA-15 was found to be highly specific, to selectively express PEA-15 in breast cancer cells, and to induce cancer cell killing in vitro and in vivo without affecting normal cells (Xie et al., 2015). A similar construct with PEA-15AA would be of interest for ovarian cancer in order to investigate if the efficacy of cisplatin treatment can be improved.

In conclusion, non-phosphorylatable PEA-15AA increases cisplatin sensitivity of ovarian cancer cells, suggesting that the phosphorylation status of PEA-15 could be evaluated as a biomarker to predict the responsiveness of ovarian tumours to cisplatin treatment. For this purpose, correlation of PEA-15 phosphorylation with therapy outcome in high-grade serous ovarian cancer is needed, as this subtype is the most aggressive and has the worst prognosis (Domcke et al., 2013).

Chapter 3: SAMHD1 is a key regulator of the lineage-specific response of acute lymphoblastic leukaemias to nelarabine

Tamara Rothenburger[1], Katie-May McLaughlin[2], Tobias Herold[3,4], Constanze Schneider[1,5], Thomas Oellerich[6,7,8], Florian Rothweiler[1], Andrew Feber[9], Tim R Fenton[2], Mark N Wass[2], Oliver T Keppler[10], Martin Michaelis[11], Jindrich Cinatl Jr[12]

[1]Institut für Medizinische Virologie, Klinikum der Goethe-Universität, Paul Ehrlich-Straße 40, 60596, Frankfurt am Main, Germany.

[2]School of Biosciences, University of Kent, Canterbury, CT2 7NJ, UK.

[3]Department of Medicine III, University Hospital, LMU Munich, Marchioninistraße 15, 81377, Munich, Germany.

[4]Research Unit Apoptosis in Hematopoietic Stem Cells, Helmholtz Zentrum München, German Research Center for Environmental Health (HMGU), Feodor-Lynenstraße 21, 81377, Munich, Germany.

[5]Department of Medicine II, Hematology/Oncology, Goethe-Universität, Frankfurt am Main, Germany; Frankfurt Cancer Institute, Goethe University, Theodor-Stern-Kai 7, 60590, Frankfurt am Main, Germany.

[6]Department of Medicine II, Hematology/Oncology, Goethe-Universität, Theodor-Stern-Kai 7, 60590, Frankfurt am Main, Germany.

[7]Frankfurt Cancer Institute, Goethe University, Theodor-Stern-Kai 7, 60590, Frankfurt am Main, Germany.

[8]German Cancer Consortium/German Cancer Research Center, Im Neuenheimer Feld 280, 69120, Heidelberg, Germany.

[9]Division of Surgery and Interventional Science, University College London, Gower Street, London, WC1E 6BT, UK.

[10]Faculty of Medicine, Max von Pettenkofer Institute, Virology, LMU München, Pettenkoferstraße 9a, 80336, Munich, Germany.

[11]School of Biosciences, University of Kent, Canterbury, CT2 7NJ, UK. M.Michaelis@kent.ac.uk.

[12]Institut für Medizinische Virologie, Klinikum der Goethe-Universität, Paul Ehrlich-Straße 40, 60596, Frankfurt am Main, Germany. Cinatl@em.uni-frankfurt.de.

Published: Communications Biology. 2020 Jun 24;3(1):324. doi: doi: https://doi.org/10.1038/s42003-020-1052-8.

Abstract

The nucleoside analogue nelarabine, the prodrug of arabinosylguanine (AraG), is effective against T-cell acute lymphoblastic leukaemia (T-ALL) but not against B-cell ALL (B-ALL). The underlying mechanisms have remained elusive. Here, data from pharmacogenomics studies and a panel of ALL cell lines reveal an inverse correlation between nelarabine sensitivity and the expression of *SAMHD1*, which can hydrolyse and inactivate triphosphorylated nucleoside analogues. Lower SAMHD1 abundance is detected in T-ALL than in B-ALL in cell lines and patient-derived leukaemic blasts. Mechanistically, T-ALL cells display increased *SAMHD1* promoter methylation without increased global DNA methylation. SAMHD1 depletion sensitises B-ALL cells to AraG, while ectopic SAMHD1 expression in SAMHD1-null T-ALL cells induces AraG resistance. SAMHD1 has a larger impact on nelarabine/AraG than on cytarabine in ALL cells. Opposite effects are observed in acute myeloid leukaemia cells, indicating entity-specific differences. In conclusion, *SAMHD1* promoter methylation and, in turn, *SAMHD1* expression levels determine ALL cell response to nelarabine.

3.1. Introduction

Acute lymphoblastic leukaemia (ALL) cells originate from precursor lymphoid T- (T-ALL) and B-cells (B-ALL). In children, ALL is the most common cancer associated with high cure rates of about 85%. In adults, ALL accounts for 15–25% of acute leukaemias and is associated with a less favourable outcome (Coccaro et al., 2019; Follini et al., 2019; Mohseni et al., 2018; Pavlovic et al., 2019). Among ALLs, T-ALL is responsible for ~15% of paediatric ALLs and 25% of adult ALLs (Follini et al., 2019). Nelarabine displays selective activity in T-ALL over B-ALL and is used for the treatment of relapsed and refractory T-ALL but not routinely for the treatment of B-ALL (Cohen et al., 1983; Follini et al., 2019; Jabbour et al., 2015; Jabbour et al., 2018; Kadia & Gandhi, 2017; Kantarjian et al., 2017; Marks & Rowntree, 2017; Shewach & Mitchell, 1989; Teachey & O'Connor, 2020). However, the molecular mechanisms underlying this difference remain elusive. Moreover, nelarabine therapy can be associated with irreversible life-threatening neurotoxicity (Berg et al., 2005; Kadia & Gandhi, 2017). Hence, biomarkers indicating patients who are most likely to benefit from nelarabine therapy are needed.

Here, we used an approach that combined data from the large pharmacogenomics screens Cancer Therapeutics Response Portal (CTRP) (Basu et al., 2013), Cancer Cell Line

Encyclopedia (CCLE) (Barretina et al., 2012), and Genomics of Drug Sensitivity in Cancer (GDSC) (Yang, W. et al., 2013) with data from an ALL cell line panel derived from the Resistant Cancer Cell Line (RCCL) collection and patient data to investigate the mechanisms underlying the discrepancy in the nelarabine sensitivity between T-ALL and B-ALL.

The results show that low expression of Sterile alpha motif and histidine-aspartic acid domain-containing protein 1 (SAMHD1) in T-ALL cells is a key determinant of nelarabine sensitivity and that SAMHD1 is a potential biomarker and therapeutic target for the improvement of nelarabine-based therapies for both T-ALL and B-ALL patients.

3.2. Methods

3.2.1. Analysis of data from pharmacogenomics screens

ALL cell line drug sensitivity data and RMA-normalised gene expression values were obtained from the CCLE (2015 release, https://portals.broadinstitute.org/ccle), which contains data from 34 ALL cell lines (18 B-ALL and 16 T-ALL) (Barretina et al., 2012), GDSC (2016 release, https://www.cancerrxgene.org/; 21 B-ALL/ 17 T-ALL cell lines) (Yang, W. et al., 2013), and CTRP (version 2, 2015 release, https://ocg.cancer.gov/programs/ctd2/data-portal; 11 B-ALL/ 13 T-ALL cell lines) (Basu et al., 2013).

Gene expression was compared using the Mann-Whitney U (Wilcoxon) test for independent groups. Multiple test correction of p-values was performed using the Benjamini–Hochberg (BH) procedure (Benjamini & Hochberg, 1995), with a false discovery rate (FDR) of 0.05 (BH = (rank/n) × FDR, where n = the total number of genes compared). Gene expression levels between B-ALL and T-ALL cell lines were visualised using the ggboxplot function in R. Heatmaps showing gene expression levels were generated using the ggplot2 package in R.

In all, 36 of the ALL cell lines (19 B-ALL, 17 T-ALL) in the GDSC and 22 ALL cell lines (10 B-ALL, 12 T-ALL) in the CTRP were treated with cytarabine. 23 ALL cell lines (11 B-ALL, 12 T-ALL) in the CTRP were treated with nelarabine. Scatter plots and their associated Pearson correlations for each drug AUCs against gene expression were calculated using the ggplot2 package in R.

Expression of 18,542 genes was correlated with the nelarabine AUC in ALL cell lines and SAMHD1 expression was correlated with the AUC values of 441 drugs tested in ALL cell lines using the CTRPv2 dataset. Pearson correlation coefficients were calculated using the cor.test

function in R. *P*-values for each correlation were ranked and multiple test correction was performed (Benjamini–Hochberg procedure, FDR = 0.05).

Pathway analysis was performed using the PANTHER (version 14.1) Overrepresentation Test (Mi et al., 2019) based on genes significantly differentially expressed in B-ALL and T-ALL cell lines after Benjamini–Hochberg p-value correction (FDR = 0.05). Fisher's exact test was applied to calculate over- vs. underrepresentation of classes. Heatmaps were prepared using the ggplot2 package in R.

Beta values for CpG sites in the SAMHD1 promoter derived from the GDSC (Gene Expression Omnibus ID GSE68379) were correlated with SAMHD1 expression in ALL cell lines.

3.2.2. Analysis of patient data

SAMHD1 gene expression was analyzed in publicly available Microarray data of 306 primary adult B- and T-ALL patients (Gene Expression Omnibus ID GSE66006) (Herold, T. et al., 2017). The median percentage of leukemic cells in the samples was 90%.

3.2.3. Drugs

Cytarabine was purchased from Tocris Biosciences (via Bio-Techne GmbH, Wiesbaden, Germany), AraG from Jena Bioscience (Jena, Germany).

3.2.4. Cell lines

The human ALL cell lines 697, ALL-SIL, BALL-1, CTV-1, GRANTA-452, HAL-01, HSB-2, JURKAT, KE-37, MHH-CALL-4, MN-60, MOLT-4, MOLT-16, NALM-6, NALM-16, P12-ICHIKAWA, REH, ROS-50, RPMI-8402, RS4;11, SEM, TANOUE, and TOM-1 and the AML cell lines THP-1 and HEL were obtained from DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, Braunschweig, Germany). The ALL cell lines CCRF-CEM and JJHan were received from ATCC (Manassas, VA, US) and the ALL cell line KARPAS231 from Cambridge Enterprise Ltd. (Cambridge, UK).

THP-1 cells deficient in SAMHD1 (THP-1 KO) and control cells (THP-1 Ctr.) were generated using CRISPR/Cas9 approach as previously described (Oellerich et al., 2019; Schneider et al., 2017; Wittmann et al., 2015). THP-1 cells were plated at a density of 2×10^5 cells per ml. After 24 h, 2.5×10^6 cells were resuspended in 250 µl Opti-MEM, mixed with 5 µg CRISPR/Cas plasmid DNA, and electroporated in a 4-mm cuvette using an exponential pulse at 250 V and 950 mF utilizing a Gene Pulser electroporation device (Bio-Rad Laboratories). We used a plasmid encoding a CMV-mCherry-Cas9 expression cassette and a human *SAMHD1* gene specific gRNA driven by the U6 promoter. An early coding exon of the *SAMHD1* gene was

targeted using the following gRNA construct: 5'-CGGAAGGGGTGTTTGAGGGG-3'. Cells were allowed to recover for 2 days in six-well plates filled with 4 ml medium per well before being FACS sorted for mCherry-expression on a BD FACS Aria III (BD Biosciences). For subsequent limiting dilution cloning, cells were plated at a density of 5, 10, or 20 cells per well of nine round-bottom 96-well plates and grown for 2 weeks. Plates were scanned for absorption at 600 nm and growing clones were identified using custom software and picked and duplicated by a Biomek FXp (Beckman Coulter) liquid handling system.

The HEL and JURKAT SAMHD1-WT and SAMHD1-D311A cell lines were generated by co-transfection of the packaging vector pPAX2 (Addgene), either pHR-SAMHD1-WT or pHR-SAMHD1-D311A and a plasmid encoding VSV-G, as previously described (Oellerich et al., 2019). All cell lines were routinely tested for Mycoplasma, using the MycoAlert PLUS assay kit from Lonza, and were authenticated by short tandem repeat profiling, as described elsewhere.

All cell lines were cultured in IMDM (Biochrom) supplemented with 10% FBS (Sigma-Aldrich), 4 mM L-glutamine (Sigma-Aldrich), 100 IU per ml penicillin (Sigma-Aldrich), and 100 µg per ml streptomycin (Sigma-Aldrich) at 37 °C in a humidified 5% CO_2 incubator.

3.2.5. Viability assay

Cell viability was determined by 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay modified after Mosmann (Mosmann, 1983), as previously described (Onafuye et al., 2019). Cells suspended in 100 µL cell culture medium were plated per well in 96-well plates and incubated in the presence of various drug concentrations for 96 h. Then, 25 µL of MTT solution (2 mg/mL (w/v) in PBS) were added per well, and the plates were incubated at 37 °C for an additional 4 h. After this, the cells were lysed using 200 µL of a buffer containing 20% (w/v) sodium dodecylsulfate in 50% (v/v) N,N-dimethylformamide with the pH adjusted to 4.7 at 37 °C for 4 h. Absorbance was determined at 570 nm for each well using a 96-well multiscanner. After subtracting of the background absorption, the results are expressed as percentage viability relative to control cultures which received no drug. Drug concentrations that inhibited cell viability by 50% (IC50) were determined using CalcuSyn (Biosoft, Cambridge, UK).

3.2.6. Western blotting

Western blotting was performed as previously described (Schneider et al., 2017). Cells were lysed in Triton X-100 sample buffer and proteins separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Proteins were blotted on a nitrocellulose membrane

(Thermo Scientific). The following primary antibodies were used at the indicated dilutions: SAMHD1 (Proteintech, 12586-1-AP, 1:1,000), β-actin (BioVision, 3598R-100, 1:5,000), pSAMHD1 (Cell Signaling, 89930S, 1:1,000), and GAPDH (Trevigen, 2275-PC-10C, 1:5,000). Visualisation and quantification were performed using IRDye-labeled secondary antibodies (LI-COR Biotechnology, IRDye®800CW Goat anti-Rabbit, 926-32211, 1:40,000) according to the manufacturer's instructions. Band volume analysis was conducted by Odyssey LICOR. Uncropped blots are presented in Figure S3.10. SAMHD1 quantification was performed using a protein extract of the AML cell line THP-1 as internal control (Figure S3.10, Figure S3.5).

3.2.7. mRNA analysis

RNA extraction and TaqMan-based mRNA quantification of SAMHD1 (assay no. Hs00210019_m1) and RNaseP (TaqMan® RNaseP Assay (A30065)) as endogenous reference control were performed according to the manufactures protocol (Applied Biosystems). Total RNA was extracted using the RNeasy Kit from Qiagen and stored at −80 °C until use. Relative quantitative PCR analyses were performed on the QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems). SAMHD1 mRNA expression levels were quantified by using the ΔΔCt method with RNaseP mRNA as an endogenous reference control. All samples were run in triplicate. Data analysis was conducted using the QuantStudio System Software (Applied Biosystems).

3.2.8. *SAMHD1* promoter methylation

SAMHD1 promoter methylation was determined as previously described (Oellerich et al., 2019). *SAMHD1* promoter contains five *HpaII* sites surrounding the transcription start site (de Silva et al., 2013). Methylation of the *HpaII* sites in the *SAMHD1* promoter would prevent digestion by the *HpaII*, and the intact sequence would serve as a template for PCR amplification using *SAMHD1* promoter-specific primers that flank the *HpaII* sites. To measure methylation of the *SAMHD1* promoter genomic DNA was treated with the methylation-sensitive HpaII endonuclease or left untreated as described previously with some modifications (de Silva et al., 2013). PM3.fwd: TTCCGCCTCATTCGTCCTTG and PM3.rev: GGTTCTCGGGCTGTCATCG were used as *SAMHD1* promoter-specific primers. A single PCR product (993-bp) corresponding to the *SAMHD1* promoter sequence was obtained from untreated genomic DNA and treated DNA from cells with methylated but not from cells with unmethylated *SAMHD1* promoter. To serve as input control, a 0.25-kb fragment of the *GAPDH* gene lacking *HpaII* sites was PCR-amplified using the same template DNA (de Silva et al., 2013). Uncropped agarose gels are shown in Figure S3.10.

3.2.9. SAMHD1 depletion using Vpx virus-like particles

Cells were spinoculated with VSV-G pseudotyped virus-like particles carrying either Vpx or Vpr from SIVmac251, produced by co-transfection of 293T cells with pSIV3 + gag pol expression plasmids and a plasmid encoding VSV-G as previously described (de Silva et al., 2013; Schneider et al., 2017).

3.2.10. Statistics and reproducibility

Statistical data analyses were performed in GraphPad Prism version 7. Population means were compared using unpaired two-tailed Student's *t*-tests. Data are presented as means ± standard deviation (S.D.). Specific information on the number and nature of replicates is provided in the figure legends. Correlation analyses were performed using linear regression in GraphPad Prism resulting in r^2 as a measure for goodness-of-fit and the *P* value, which is calculated from an F test, indicating whether the slope is significantly different from zero.

3.3. Results

3.3.1. Gene expression comparison between T-ALL and B-ALL cells

To identify potential differences between T-ALL and B-ALL that may explain the observed discrepancies in nelarabine sensitivity, we started by analysing data derived from the large pharmacogenomics databases CCLE (Barretina et al., 2012), Cancer Therapeutics Response Portal (CTRP) (Basu et al., 2013), and Genomics of Drug Sensitivity in Cancer (GDSC) (Yang, W. et al., 2013). The CCLE contained data derived from 34 leukaemia cell lines (18 B-ALL, 16 T-ALL) and the GDSC from 38 leukaemia cell lines (21 B-ALL, 17 T-ALL), with an overlap of 19 cell lines (Table S3.1). The CCLE and CTRP used the same cell line panel for their studies (Basu et al., 2013).

Nelarabine was tested in 24 ALL (11 B-ALL, 13-T-ALL) cell lines in the CTRP (Table S3.2). In agreement with the available literature (Beesley et al., 2007; Cohen et al., 1983; Homminga et al., 2011; Shewach & Mitchell, 1989), nelarabine displayed higher activity in T-ALL than in B-ALL cell lines (Figure S3.1A, Table S3.2).

Initially, we compared transcriptomics data (mRNA abundance) between T-ALL and B-ALL cell lines. Substantial proportions of transcripts displayed significant differences ($P < 0.05$) in their abundance levels between T-ALL and B-ALL cells in the GDSC (3,998/ 22.5% of 17,735

transcripts), CCLE (8,498/ 42.1% of 20,172 transcripts), and CTRP (4,507/ 24.3% of 18,539 transcripts) (Supplementary Data 3.1). Gene expression heatmaps illustrated these differences (Figure S3.1), but manual analysis of the top differentially regulated genes did not result in the identification of candidate genes, whose expression seemed likely to be responsible for the observed differences in nelarabine sensitivity.

A pathway analysis using the PANTHER database (Protein ANalysis THrough Evolutionary Relationships, http://pantherdb.org) (Mi et al., 2019) also did not reveal processes that may underlie the increased nelarabine sensitivity of T-ALL cells (Figure S3.1, Supplementary Data 3.2). As expected, B-cell- and T-cell-specific processes featured prominently among the most strongly differentially regulated pathways.

3.3.2. SAMHD1 levels correlate with nelarabine resistance

The correlation of transcriptomics data with the nelarabine drug response, represented as AUC, identified *SAMHD1* as the gene, whose expression displayed the most significant direct correlation (Supplementary Data 3.3). Analysis of *SAMHD1* expression exclusively in either the B-ALL or T-ALL subset also showed a highly significant direct correlation with the nelarabine AUC (Supplementary Data 3.3). Furthermore, when we correlated drug AUCs with *SAMHD1* expression, nelarabine displayed the most significant direct correlation with *SAMHD1* expression across all ALL cell lines, the second most significant direct correlation with *SAMHD1* expression in the B-ALL cell lines, and the third most significant direct correlation with *SAMHD1* expression in the T-ALL cell lines (Supplementary Data 3.4).

3.3.3. SAMHD1 levels are lower in T-ALL than in B-ALL cells

SAMHD1 is a deoxynucleotide triphosphate (dNTP) hydrolase that cleaves physiological dNTPs and triphosphorylated nucleoside analogues (Amie et al., 2013; Ballana et al., 2014; Hollenbaugh et al., 2017; Knecht et al., 2018; Mauney & Hollis, 2018). It was previously shown to interfere with the activity of anti-cancer nucleoside analogues including nelarabine (Herold, N. et al., 2017; Hollenbaugh et al., 2017; Knecht et al., 2018). If SAMHD1 was responsible for the differences observed in nelarabine sensitivity between T-ALL and B-ALL, T-ALL cells would be expected to express lower levels of *SAMHD1*. Indeed, the *SAMHD1* expression (mRNA abundance) levels were significantly lower in T-ALL than in B-ALL cell lines in all three databases (Figure 3.1A). Similar findings were detected in a gene expression dataset derived from blasts of 306 ALL (222 B-ALL, 84 T-ALL) patients (Herold, T. et al., 2014; Herold, T. et al., 2017) (Figure 3.1B). Further analysis revealed a reduced expression of *SAMHD1* in T-ALL in general but more pronounced in the thymic and mature

immunophenotypic subtype (Figure S3.2A). On the genetic level, some B-ALL subgroups like for example Philadelphia (Ph)-like patients display a gene expression pattern of *SAMHD1* that is equally low as seen in T-ALL (Figure S3.2B).

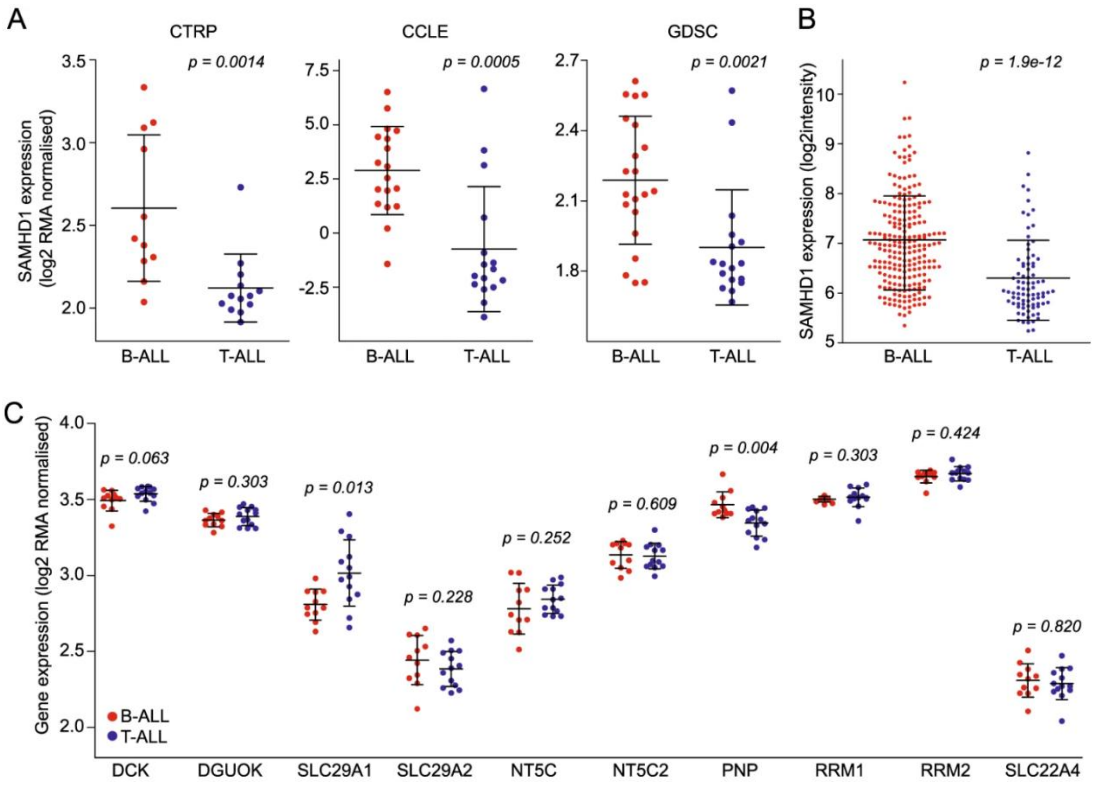


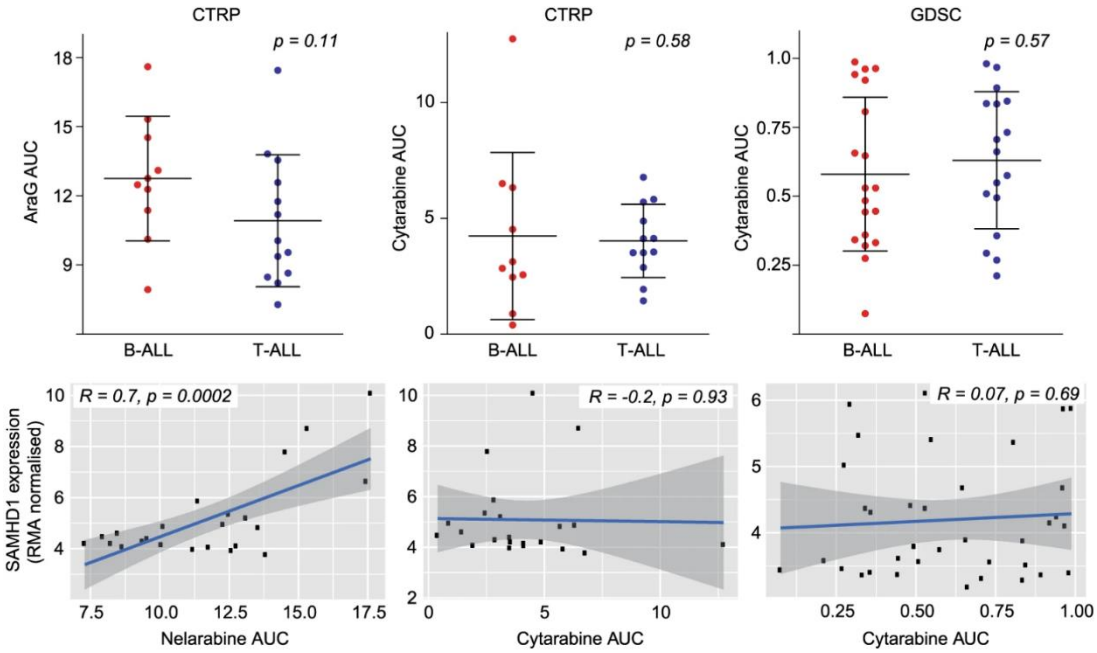
Figure 3.1. SAMHD1 levels differ between T-ALL and B-ALL.

Comparison of SAMHD1 expression (mRNA abundance) levels in T-ALL and B-ALL cell lines from the CTRP, CCLE, and GDSC (**A**) and in blasts from leukaemia patients (**B**). (**C**) Comparison of the expression of other genes known to affect nucleoside analogue activity based on CTRP data. Respective CCLE and GDSC data are provided in Figure S3.2. **p*-values for the comparison B-ALL vs. T-ALL.

3.3.4. SAMHD1 is a determinant of nelarabine sensitivity in ALL

A number of other gene products have been described to be involved in the transport, activation, and metabolism of nucleoside analogues such as nelarabine, including DCK, DGUOK, SLC29A1 (ENT1), SLC29A2 (ENT2), NT5C, NT5C2, PNP, RRM1, RRM2 and SLC22A4 (OCTN1) (Drenberg et al., 2017; Homminga et al., 2011). While statistically significant differences in the expression of some of the respective genes were noted between B-ALL and T-ALL cell lines in some of the three datasets, none was consistent across all three and none

was as robust as in the expression of *SAMHD1* (Figure 3.1c, Figure S3.3). In patient samples, SAMHD1 also displayed the most significant difference in expression levels between B-ALL



and T-ALL (Figure S3.3). Moreover, only the expression of *SAMHD1* correlated with the nelarabine AUC in the CTRP dataset (Figure 3.2, Figure S3.4). This shows that SAMHD1 is a critical determinant of nelarabine efficacy in ALL and that low SAMHD1 levels critically contribute to the specific nelarabine sensitivity of T-ALL cells.

Figure 3.2. Comparison of nelarabine (CTRP) and cytarabine (CTRP, GDSC) sensitivity between B-ALL and T-ALL cell lines and correlation of SAMHD1 mRNA levels with the nelarabine and cytarabine sensitivity (expressed as AUC) across all B-ALL and T-ALL cell lines.

Pearson's r values and respective p-values are provided. Respective data on the correlation of *SAMHD1* expression with drug sensitivity exclusively for B-ALL and T-ALL cell lines are provided in Figure S3.3 (nelarabine) and Figure S3.4 (cytarabine).

3.3.5. SAMHD1 is no determinant of cytarabine sensitivity in ALL

Cellular SAMHD1 levels have previously been shown to critically determine cytarabine efficacy in acute myeloid leukaemia (AML) cells (Hollenbaugh et al., 2017; Knecht et al., 2018; Schneider et al., 2017) and *SAMHD1* expression levels are lower in T-ALL than in AML cells (Figure S3.5). The CTRP and GDSC contained data on cytarabine activity. In contrast to AML cells, however, there was no difference in the cytarabine sensitivity between B-ALL and T-ALL cell lines and no correlation between *SAMHD1* expression and cytarabine sensitivity in

ALL cells (Figure 3.2, Figure S3.6). Hence, the effect of SAMHD1 on nucleoside analogue activity depends on the tissue context.

3.3.6. SAMHD1 mRNA levels reflect protein levels in ALL cell lines

To further investigate the role of SAMHD1 on nelarabine and cytarabine efficacy in ALL, we assembled a panel consisting of 15 B-ALL and 11 T-ALL cell lines from the RCCL collection (Michaelis et al., 2019) (Table S3.3). Firstly, we investigated the extent to which cellular SAMHD1 mRNA levels are indicative of cellular protein levels. Western blot analyses confirmed that the RCCL T-ALL cell lines generally display lower SAMHD1 protein levels than the RCCL B-ALL cell lines (Figure 3.3a, Figure S3.7). However, quantitative western blot analysis and quantitative PCR (qPCR) showed that cellular SAMHD1 mRNA levels do not always directly correlate with cellular SAMHD1 protein levels (Figure 3.3B). This is likely to reflect the complexity of the regulation of protein levels, which are determined by transcription and translation efficacy, factors that control mRNA stability (e.g. microRNAs and proteins that control mRNA degradation), and post-translational modifications that promote (proteasomal) protein degradation (Bicknell & Ricci, 2017; Dikic, 2017; Jaén et al., 2018; Ko & Dixon, 2018; Radhakrishnan & Green, 2016; Wolf & Menssen, 2018). Moreover, mutations may affect SAMHD1 function, as demonstrated in patients with chronic lymphocytic leukaemia and colorectal cancer (Clifford et al., 2014; Rentoft et al., 2016). However, the only ALL cell line with a SAMHD1 mutation was Jurkat, which harboured an R611* nonsense mutation based on GDSC data. SAMHD1 mRNA and protein levels in the RCCL are correlated with SAMHD1 mRNA levels in the corresponding cell lines from CTRP, CCLE and GDSC (Figure S3.8). Hence, SAMHD1 mRNA levels, largely predict SAMHD1 protein levels, which is in line with previous findings (Schneider et al., 2017).

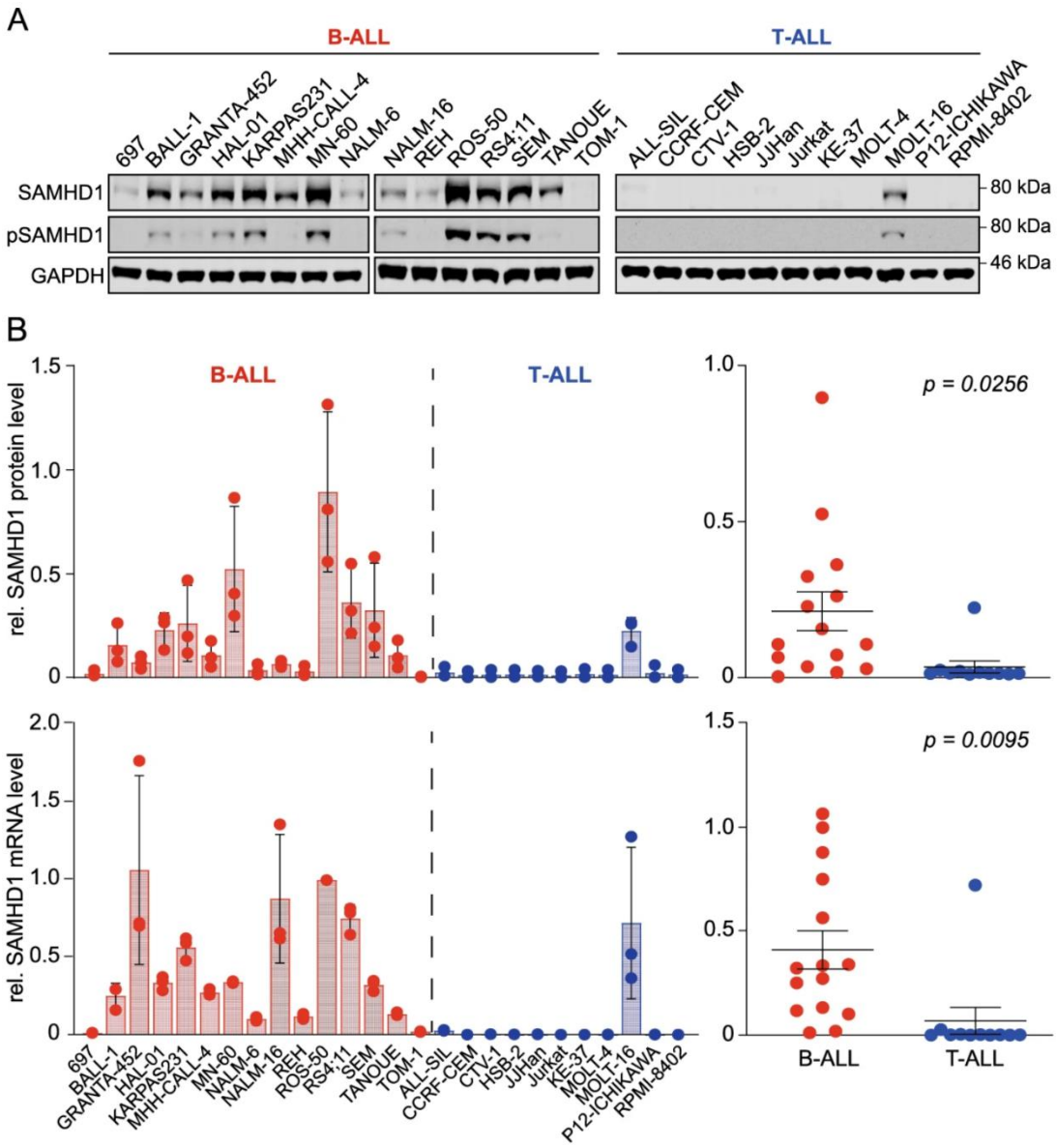


Figure 3.3. SAMHD1 protein and mRNA levels in the RCCL panel of B-ALL and T-ALL cell lines.

(**A**) Representative Western blots indicating protein levels of total SAMHD1 and phosphorylated SAMHD1 (p-SAMHD1). GAPDH was used as loading control. (**B**) Quantitative SAMHD1 protein levels are shown as means ± SD from three independent experiments (quantified using near-infrared Western blot images to determine the ratio SAMHD1/ GAPDH relative to the positive control THP-1, an acute myeloid leukaemia cell line characterised by high cellular SAMHD1 levels (Schneider et al., 2017). SAMHD1 mRNA abundance levels are shown as means ± SD from three technical replicates (as determined by qPCR, relative to cell line ROS-50) in B-ALL and T-ALL cell lines. Unpaired two-tailed Student's *t*-tests were used to compare means (represented as horizontal lines ± SEM) of SAMHD1 protein or mRNA levels in B-ALL and T-ALL cells.

Next, the sensitivity of the RCCL ALL cell lines was tested against arabinosylguanine (AraG), the product of the prodrug nelarabine (Curbo & Karlsson, 2006) and cytarabine. The results were in agreement with the CTRP data showing that T-ALL cell lines were significantly more sensitive to AraG than B-ALL cell lines (Figure 3.4, Table S3.3). Notably, there was a significant correlation between the nelarabine AUCs in the CTRP and the AraG IC50s in the RCCL panel among the cell lines that were present in both datasets (Figure S3.9). In contrast to the CTRP and GDSC data that had not indicated a difference between the cytarabine sensitivity of T-ALL- and B-ALL- cells, T-ALL cell lines displayed a trend indicating increased sensitivity to cytarabine ($P = 0.055$) (Figure 3.4). SAMHD1 protein levels displayed a significant correlation with the AraG concentrations that reduced cell viability by 50% (IC50) in all ALL cell lines and the lineage-specific subanalyses (Figure 3.4). In contrast, a significant correlation between SAMHD1 protein levels and cytarabine activity was only detected across all ALL cell lines but not when only B-ALL or T-ALL cell lines were considered (Figure 3.4). SAMHD1 mRNA levels were correlated with the AraG IC50 across all ALL cell lines and T-ALL cell lines but not B-ALL cell lines ($P = 0.1335$) (Figure 3.4). No significant correl ation was detected between the SAMHD1 mRNA levels and the cytarabine IC50 in the RCCL ALL cell lines (Figure 3.4).

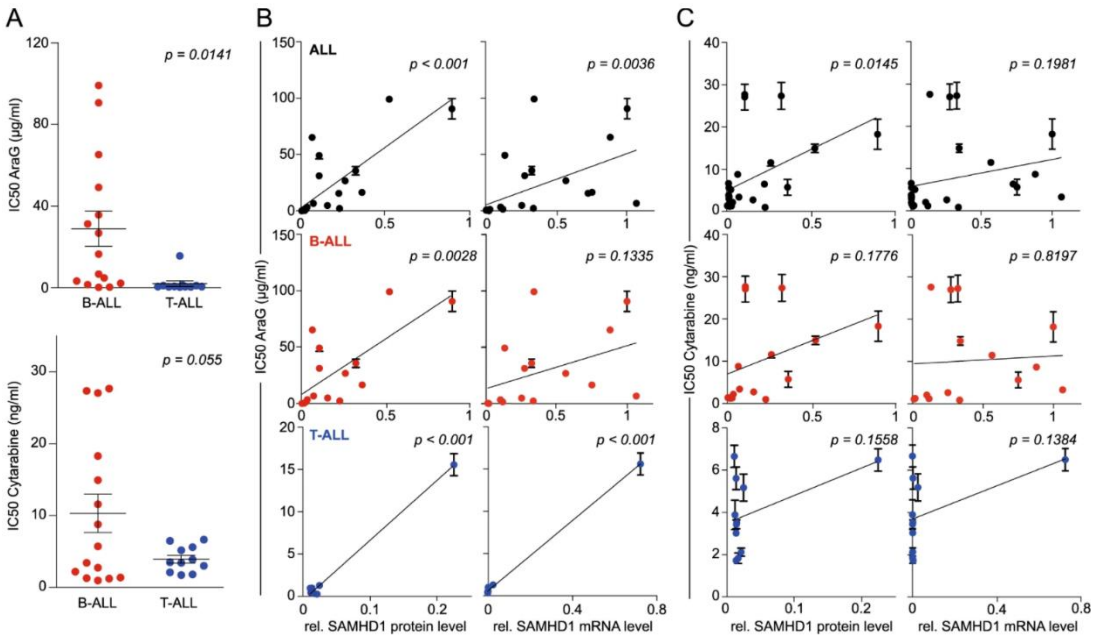


Figure 3.4. AraG and cytarabine concentrations that reduce the viability of the RCCL ALL cell lines by 50% (IC50) and correlation of the IC50s with the cellular SAMHD1 protein or mRNA levels.

Numerical data are provided in Supplementary Data 3.4. Closed circles and error bars represent means ± SD of three independent experiments, each performed in three technical replicates. Linear regression analyses were performed using GraphPad Prism.

Taken together, these results confirm the CTRP data in showing that cellular SAMHD1 levels determine ALL sensitivity against AraG, the product of nelarabine, and that low SAMHD1 levels in T-ALL cells are associated with specific nelarabine/ Ara-G activity in this lineage. In contrast to the CTRP and GDSC data, the additional experimental analyses in the RCCL ALL cell line panel suggest that SAMHD1 levels may also affect cytarabine activity in ALL, albeit to a lower degree than AraG activity.

3.3.7. SAMHD1 depletion sensitises ALL cells to AraG

To further investigate the functional role of SAMHD1 in determining AraG and cytarabine activity in ALL cells, we depleted SAMHD1 using virus-like particles containing Vpx as previously described (Baldauf et al., 2012; Schneider et al., 2017). Vpx is a protein encoded by HIV-2 and certain SIV strains that mediates proteasomal SAMHD1 degradation (Baldauf et al., 2012; Hrecka et al., 2011; Laguette et al., 2011). Vpx virus-like particles resulted in the sensitisation of ALL cells to AraG and cytarabine but exerted much more pronounced effects on the activity of AraG (Figure 3.5A).

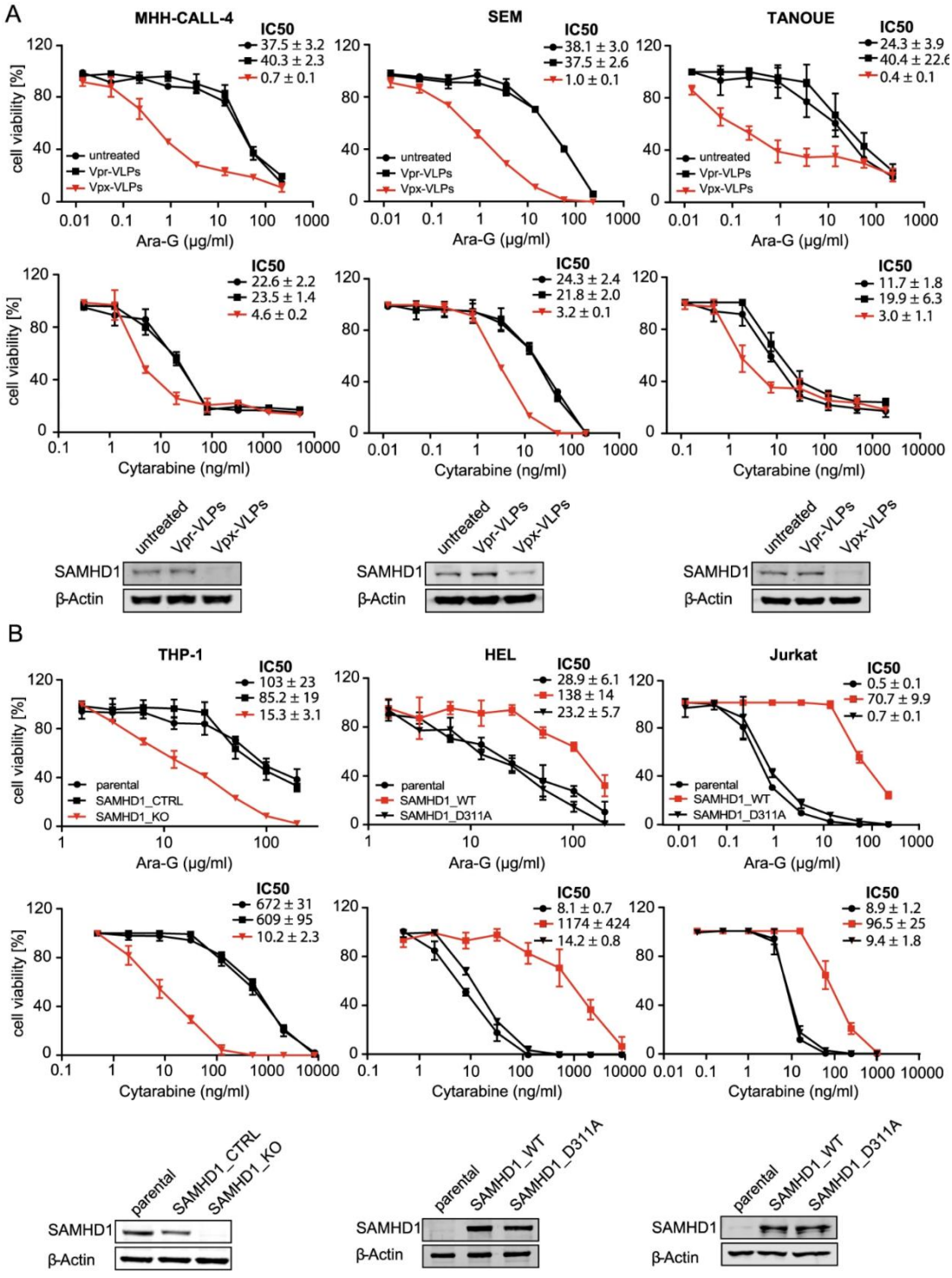


Figure 3.5. Effect of SAMHD1 on nelarabine and cytarabine sensitivity in ALL and AML cells.

(**A**) Dose-response curves of AraG- and cytarabine-treated ALL cell lines in the absence or presence of Vpx virus-like particles (cause SAMHD1 depletion), or Vpr virus-like particles (negative) controls. Concentrations that reduce ALL cell viability by 50% (IC50s) and Western blots confirming SAMHD1 depletion are provided. Each symbol represents the mean ± SD of three technical replicates of one representative experiment out of three. (**B**) Effects of AraG and cytarabine on AML cell lines in the

absence or presence of functional SAMHD1. In the *SAMHD1*-expressing AML cell line THP-1, the *SAMHD1* gene was disrupted by CRISPR/Cas9 (THP1-KO). The non-*SAMHD1* expressing AML cell line HEL and the non-*SAMHD1* expressing ALL cell line JURKAT were transduced with wild-type *SAMHD1* (SAMHD1_WT) or the triphosphohydrolase-defective *SAMHD1* mutant D311A (SAMHD1_D311A). Dose-response curves, drug concentrations that reduce cell viability by 50% (IC50s), and Western blots confirming SAMHD1 protein levels are provided. Each symbol represents the mean ± SD of three independent experiments, each performed in three technical replicates.

In the ALL cell lines MHH-CALL-4, SEM, and TANOUE, the AraG IC50s were between 37.5- and 101-fold lower following exposure to Vpx virus-like particles compared to Vpr virus-like particles, which served as negative controls. In contrast, Vpx virus-like particles only reduced the cytarabine IC50s by 5- and 7-fold lower in these cell lines.

3.3.8. Different role of SAMHD1 as resistance factor in ALL and AML

In AML cells, SAMHD1 has been described as a critical regulator of cytarabine activity (Schneider et al., 2017). Since Vpx virus-like particle-mediated SAMHD1 depletion had resulted in a more pronounced sensitisation of ALL cells to AraG than to cytarabine, we further compared the effect of the presence or absence of functional SAMHD1 on the activity of these structurally related nucleoside analogues in these two types of acute leukaemia. Cell models included the *SAMHD1*-expressing AML cell line THP-1 and its subline, in which the *SAMHD1* gene had been disrupted by CRISPR/Cas9 (THP-1- KO). Further, we investigated the *SAMHD1* low/ non- expressing cell lines HEL (AML) and Jurkat (T-ALL) and their respective sublines transduced either with wild-type (WT) *SAMHD1* or the triphosphohydrolase-defective mutant *SAMHD1-D311A*. In the AML cell lines, absence of functional SAMHD1 was associated with a 60-fold (THP-1/ THP-1-KO) and 6583-fold (HEL-SAMHD1_WT/HEL-SAMHD1_D311A) sensitisation to cytarabine, but only a 5.6- and 6.0-fold sensitisation to AraG (Figure 3.5B). The T-ALL cell line Jurkat-SAMHD1_D311A was 101 times more sensitive to AraG than Jurkat-SAMHD1_WT, while JURKAT-SAMHD1_D311A was only 10 times more sensitive to cytarabine (Figure 3.5B). In summary, SAMHD1 activity critically regulates cytarabine activity but has a much lower impact on AraG in AML cells. The opposite effect is observed in ALL cells, in which SAMHD1 crucially determines AraG activity but exerts substantially less pronounced effects on cytarabine activity. This further confirms that the cellular background critically determines the importance of SAMHD1 as regulator of nucleoside activity.

3.3.9. High SAMHD1 promoter methylation in T-ALL cell lines

SAMHD1 levels may be regulated by *SAMHD1* promoter methylation in leukaemia cells (de Silva et al., 2013; Oellerich et al., 2019). Therefore, we compared *SAMHD1* promoter methylation in T-ALL and B-ALL cell lines through amplification of a single PCR product (993-bp) corresponding to the promoter sequence after HpaII digestion. Results indicated that the *SAMHD1* promoter was methylated in all T-ALL cell lines but one (MOLT-16) (Figure 3.6A), which was the only T-ALL cell line characterised by high SAMHD1 mRNA and protein levels (Figure 3.3) and low AraG sensitivity (Table S3.3). In contrast, *SAMHD1* promoter methylation was only observed in two out of 15 B-ALL cell lines (NALM-6, TOM-1) (Figure 3.6A). In agreement, *SAMHD1* promoter methylation was also significantly higher in T-ALL than in B-ALL cells in the GDSC and inversely correlated with *SAMHD1* expression (Figure 3.6B). Notably, global DNA methylation did not differ between T-ALL and B-ALL cell lines (Figure 3.6C), suggesting lineage-specific differences. Taken together, this suggests that the differences in cellular SAMHD1 levels observed between T-ALL and B-ALL cell lines are to a large extent caused by differences in *SAMHD1* promoter methylation.

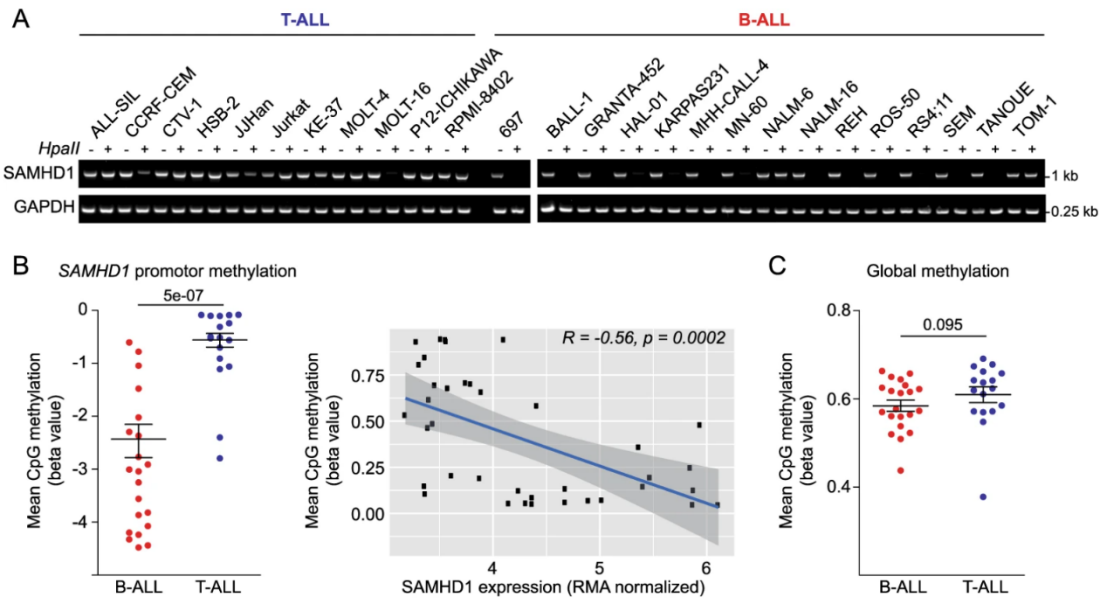


Figure 3.6. SAMHD1 promoter methylation in ALL cell lines.

(**A**) Analysis of *SAMHD1* promoter methylation in the RCCL cell line panel through amplification of a single PCR product (993-bp) corresponding to the promoter sequence after HpaII digestion. A 0.25-kb fragment of the GAPDH gene lacking HpaII sites was PCR-amplified using the same template DNA served as loading control. (**B**) GDSC data indicating *SAMHD1* promoter methylation in B-ALL and T-ALL cell lines and correlation of *SAMHD1* promoter methylation and *SAMHD1* expression across all ALL cell lines. (**C**) GDSC data indicating the level of global methylation in B-ALL and T-ALL cell lines.

3.4. Discussion

Similar chemotherapeutic agents are used to treat T-ALL and B-ALL. However, nelarabine is specifically used for relapsed T-ALL (Follini et al., 2019; Jabbour et al., 2015; Kadia & Gandhi, 2017; Kantarjian et al., 2017; Marks & Rowntree, 2017; Zwaan et al., 2017). Although it had been known for decades that nelarabine is more active in T-ALL than in B-ALL cells (Shewach & Mitchell, 1989; Verhoef & Fridland, 1985), the underlying mechanisms had remained elusive.

Here, we used an approach combining data derived from large pharmacogenomics screens (CTRP, CCLE, GDSC), an RCCL-derived ALL cell line panel, and patient data and found that cellular SAMHD1 levels critically determined ALL cell sensitivity to nelarabine and AraG. Nelarabine is metabolised into AraG, which is then triphosphorylated by cellular kinases into the active form (Curbo & Karlsson, 2006). SAMHD1 is a deoxynucleotide triphosphate (dNTP) hydrolase that cleaves and inactivates triphosphorylated nucleoside analogues including triphosphorylated AraG (Hollenbaugh et al., 2017; Knecht et al., 2018; Mauney & Hollis, 2018; Oellerich et al., 2019). Moreover, T-ALL cells were characterised by substantially lower SAMHD1 levels than B-ALL cells. Previous studies had demonstrated an association between AraG efficacy and AraG triphosphate levels in leukaemia cells, but the mechanism determining AraG triphosphate levels had remained unknown (Akahane et al., 2019; Homminga et al., 2011; Shewach & Mitchell, 1989; Verhoef & Fridland, 1985). Hence, SAMHD1 is the missing link explaining the discrepancy in nelarabine sensitivity between T-ALL and B-ALL.

Notably, SAMHD1 has also been shown to promote DNA damage repair including damage induced by the topoisomerase inhibitors camptothecin and etoposide (Daddacha et al., 2017a). Thus, SAMHD1-mediated repair of nelarabine/ AraG-induced DNA damage may potentially also contribute to the increased nelarabine/ AraG resistance associated with high SAMHD1 levels in ALL cells. However, *SAMHD1* expression was not associated with generally increased resistance to DNA damaging agents in the CTRP (Supplementary Data 3.4). The AUC of the PARP inhibitor veliparib was correlated with *SAMHD1* expression, but the AUC of the PARP inhibitor olaparib was not. The AUCs of etoposide and other prominent DNA damaging agents such as the alkylating agents temozolomide, ifosfamide, and dacarbazine and the nucleoside analogue 5-fluorouracil also were also not correlated with SAMHD1 expression, and the AUCs of the alkylating agents cyclophosphamide and chlorambucil, the

nucleoside analogue gemcitabine, the DNA cross-linker mitomycin C, and the topoisomerase inhibitor doxorubicin displayed a significant inverse correlation to *SAMHD1* expression (Supplementary Data 3.4). These data do not suggest that SAMHD1 interferes with the effects of anti-cancer drugs predominantly via promotion of DNA damage repair.

Data derived from the RCCL ALL cell line panel as well as from the GDSC indicate that the differences in *SAMHD1* expression observed between T-ALL and B-ALL cells are at least in part the consequence of higher *SAMHD1* promoter methylation in T-ALL than in B-ALL cells. Thus, *SAMHD1* expression levels and *SAMHD1* promoter methylation are potential biomarkers of nelarabine sensitivity that deserve further clinical investigation. Based on our current data, patients suffering from ALL characterised by high *SAMHD1* expression are unlikely to benefit from therapy using nelarabine and may be better treated with ribose-based thiopurines that are no SAMHD1 substrates, such as 6-thioguanine or 6-mercaptopurine (Wu, C. & Li, 2018).

SAMHD1 depletion sensitised ALL cells to AraG, indicating that SAMHD1 may also serve as a therapeutic target to improve nelarabine therapies in ALL patients. Notably, both T-ALL and B-ALL patients may benefit from SAMHD1 inhibition in combination with nelarabine therapy. Interestingly, the effect of SAMHD1 on the activity of nucleoside analogues varied substantially between different forms of leukaemia. SAMHD1 was previously shown to critically determine the activity of the nucleoside analogue cytarabine in AML (Hollenbaugh et al., 2017; Knecht et al., 2018; Schneider et al., 2017). Compared to the pronounced effects of SAMHD1 on nelarabine/ AraG activity in ALL, however, SAMHD1 exerted only minor effects on the activity of cytarabine in this leukaemia type. Interestingly, the situation was reversed in AML cells, where SAMHD1 critically affected cytarabine activity but had much lower impact on AraG. These findings are important, because they illustrate that, despite a general trend in the biomedical community towards tumour-agnostic approaches, which consider cancer-specific alterations independently of the cancer type (Huang, F. W. & Feng, 2019; Luoh & Flaherty, 2018), a much more in depth understanding of the molecular make-up of cancer cells will be required, before therapy decisions can be entirely based on molecular markers without taking the cancer entity into consideration.

In conclusion, our data indicate that cellular SAMHD1 levels critically determine ALL cell sensitivity to nelarabine/ AraG and that T-ALL cells display lower SAMHD1 levels than B-ALL. This provides a solution to a decades old conundrum providing a mechanistic explanation for the higher nelarabine sensitivity of T-ALL cells compared to B-ALL cells. Hence, SAMHD1 has

potential as a biomarker for the more accurate identification of ALL patients, who are likely to benefit from nelarabine therapy. Moreover, SAMHD1 is a therapeutic target for the design of improved nelarabine-based treatment strategies for ALL patients.

Chapter 4: Multifaceted roles of SAMHD1 in cancer

Katie-May McLaughlin[1], Jindrich Cinatl jr.[2]*, Mark N. Wass[1]*, Martin Michaelis[1]*

[1] School of Biosciences, University of Kent, Canterbury, UK

[2] Institute of Medical Virology, Goethe-University, Frankfurt am Main, Germany

Submitted to Review Commons for pre-journal peer review. Available at bioRxiv: https://www.biorxiv.org/content/10.1101/2021.07.03.451003v1

Abstract

SAMHD1 is discussed as a tumour suppressor protein, but its potential role in cancer has only been investigated in very few cancer types. Here, we performed a systematic analysis of the TCGA (adult cancer) and TARGET (paediatric cancer) databases, the results of which did not suggest that SAMHD1 should be regarded as a *bona fide* tumour suppressor. SAMHD1 mutations that interfere with SAMHD1 function were not associated with poor outcome, which would be expected for a tumour suppressor. High SAMHD1 tumour levels were associated with increased survival in some cancer entities and reduced survival in others. Moreover, the data suggested differences in the role of SAMHD1 between males and females and between different races. Often, there was no significant relationship between SAMHD1 levels and cancer outcome. Taken together, our results indicate that SAMHD1 may exert pro- or anti-tumourigenic effects and that SAMHD1 is involved in the oncogenic process in a minority of cancer cases. These findings seem to be in disaccord with a perception and narrative forming in the field suggesting that SAMHD1 is a tumour suppressor. A systematic literature review confirmed that most of the available scientific articles focus on a potential role of SAMHD1 as a tumour suppressor. The reasons for this remain unclear but may include confirmation bias and publication bias. Our findings emphasise that hypotheses, perceptions, and assumptions need to be continuously challenged by using all available data and evidence.

4.1. Introduction

Sterile α motif and histidine-aspartic domain containing protein 1 (SAMHD1) was initially discovered in dendritic cells and named dendritic cell-derived IFN-γ induced protein (DCIP) (Coggins et al., 2020; Li, N. et al., 2000). SAMHD1 is indeed involved in the regulation of interferon signalling, and SAMHD1 mutations are associated with Aicardi-Goutieres syndrome, an autoimmune inflammatory disorder characterised by a dysregulated interferon response (Coggins et al., 2020; Mauney & Hollis, 2018; Rice, G. I. et al., 2009).

In the meantime, SAMHD1 has been shown to exert a range of additional functions (Chen, Z. et al., 2021; Coggins et al., 2020; Mauney & Hollis, 2018). As a deoxynucleoside triphosphate hydrolase (dNTPase), that cleaves deoxynucleoside triphosphates (dNTPs) into deoxynucleosides and triphosphate, SAMHD1 plays, together with enzymes that catalyse dNTP biosynthesis, an important role in the maintenance of balanced cellular dNTP pools (Chen, Z. et al., 2021; Coggins et al., 2020; Mauney & Hollis, 2018). Since imbalances in cellular dNTP pools affect cell cycle regulation and DNA stability, SAMHD1 is also involved in the regulation of these processes (Chen, Z. et al., 2021).

In addition to controlling cellular dNTP levels, SAMHD1 has been shown to maintain genome integrity by a range of further mechanisms, including maintenance of telomere integrity, inhibition of LINE-1 retrotransposons, facilitation of homologous recombination-mediated double-strand break repair and DNA end joining, and prevention of R-loop formation at transcription-replication conflict regions (Akimova et al., 2021; Chen, Z. et al., 2021; Herold, N., Rudd, Sanjiv, Kutzner, Myrberg et al., 2017; Park et al., 2021). Additionally, low SAMHD1 levels have been detected in chronic lymphocytic leukaemia (CLL), lung cancer, cutaneous T-cell lymphoma, AML, colorectal cancer, and Hodgkin lymphoma. Moreover, loss-of-function SAMHD1 mutations have been described in cancer types including CLL and colorectal cancer (Chen, Z. et al., 2021; Coggins et al., 2020; Herold, N. et al., 2017; Mauney & Hollis, 2018). Due to these observations, SAMHD1 is being considered as a tumour suppressor protein.

However, the potential role of SAMHD1 in cancer diseases is more complex. It also recognises and cleaves the triphosphorylated, active forms of a range of anti-cancer nucleoside analogues. In this context, SAMHD1 has been described as a clinically relevant resistance factor in acute myeloid leukaemia (AML) and acute lymphoblastic leukaemia (ALL) against nucleoside analogues including cytarabine, decitabine, and nelarabine (Herold, N. et al., 2017; Knecht et al., 2018; Oellerich et al., 2019; Rothenburger et al., 2020; Schneider et al., 2017).

So far, the potential tumour suppressor activity of SAMHD1 has only been investigated in a few cancer types. To establish a broader understanding of the role of SAMHD1 in cancer, we here performed a systematic analysis of mutation data, gene expression data, and cancer patient survival data provided by The Cancer Genome Atlas (TCGA) (The Cancer Genome Atlas Research Network, 2008) and the Therapeutically Applicable Research To Generate Effective Treatments (TARGET) (https://ocg.cancer.gov/programs/target) databases. The TCGA provided data from 9,703 patients with 33 different types of adult cancer and the TARGET database from 1,091 patients with seven different paediatric cancer types.

4.2. Materials and Methods

4.2.1. Gene Expression and Clinical Data

Gene expression data (FPKM values) from patient tumours were derived from The Cancer Genome Atlas (TCGA) (The Cancer Genome Atlas Research Network, 2008) via the GDC Data Portal (https://portal.gdc.cancer.gov). The Bioconductor R package TCGAbiolinks was used to obtain corresponding clinical data. Primary tumour gene expression data and clinical response data were available for 9,572 patients (5,037 female, 4,535 male) with 33 different cancer types. Ages at diagnosis ranged from 14 to 90 (median age at diagnosis = 61, no data for 113 patients). Data were also downloaded for 694 matched normal tissue samples.

Gene expression (RPKM) values and clinical data were extracted for patients in the TARGET database from the National Cancer Institute Office of Cancer Genomics TARGET data matrix (https://ocg.cancer.gov/programs/target/data-matrix). Primary tumour sample data was available for a total of 1,091 patients in TARGET (470 females and 593 males) with seven cancer types. Ages at diagnosis ranged from six days to 32.41 years (median age at diagnosis was 5.4 years (1976 days)).

Tumour vs normal sample gene expression was compared using the wilcox.test function in R, which performs the Mann Whitney U test for independent groups. Pairwise comparisons were made using the Wilcoxon Signed Rank test. Pie charts were generated using ggplot2.

4.2.2. Methylation and miRNA data

TCGA methylation beta values and miRNA expression values (reads per million miRNA mapped) were downloaded from the GDC Data Portal (https://portal.gdc.cancer.gov). Mean methylation beta values for each CpG site in the SAMHD1 promoter for which data were

available (cg02078758, cg00642209, cg16430572, cg09128050, cg12099051, cg18861300, cg11094122, cg22769031, cg23888977, cg09717261, cg24951864, cg06097592, cg22583967, cg10804363 and cg12517061) were calculated per individual. Expression data for miRNAs which were listed in DIANA-TarBase v8 (Karagkouni et al., 2018) as being experimentally validated to positively interact with SAMHD1 (n=21) along with eight miRNAs shown in previous experiments to target SAMHD1 (Jin, C. et al., 2014; Kohnken et al., 2017; Pilakka-Kanthikeel et al., 2015; Riess et al., 2017) were extracted for analysis. Scatter plots and associated Pearson correlations for methylation and miRNA expression with SAMHD1 expression were calculated using the ggplot2 package in R.

4.2.3. Survival analyses

Cox proportional hazards regression was used to calculate the hazard ratio for cohorts expressing high levels of SAMHD1. Overall survival (OS) was defined as days to last follow-up or death, as previously described (Ng, S. W. et al., 2016). Calculations were performed using the R survminer and survival packages. The 'surv_cutpoint' function was used to identify the optimal expression cut-off point to give the lowest p-value for high vs low expression. We permitted the cut-off to be only between the 20th and 80th percentiles of gene expression values, as described by previously (Uhlen et al., 2017).

Kaplan-Meier survival curves were generated using R package ggsurvplot. P-values in each case were the result of a log rank (Mantel-Cox) test, which assesses whether there is a significant difference between the survival of two independent groups. Hazard ratios quoted refer to values for 'low' (below the calculated optimal cut-off) expression for each gene in the model, with values >1 indicating increased hazard (i.e. reduced OS) and values <1 indicating decreased hazard (i.e. increased OS).

4.2.4. Mutation Data and Variant Effect Prediction

Mutation data for 10,149 TCGA patients were downloaded from the GDC Data Portal (https://portal.gdc.cancer.gov). A dot plot displaying mutation frequencies of 21,156 genes was generated using ggplot2.

In order to assess the potential impact of mutations in SAMHD1, we used the online tool Variant Effect Predictor (VEP) (McLaren et al., 2016) to obtain reports from SIFT (Sim et al., 2012), PolyPhen-2 (Adzhubei et al., 2010) , Condel (González-Pérez & López-Bigas, 2011) and CADD (Kircher et al., 2014).

A lollipop plot of SAMHD1 mutations was generated using the cBioPortal MutationMapper tool (https://www.cbioportal.org/mutation_mapper) (Cerami et al., 2012).

4.2.5. Literature review

Relevant articles were identified on 17th June 2021 by using the search term "(((Cancer) OR (tumor) OR (tumour))) AND (SAMHD1)" in PubMed (https://pubmed.ncbi.nlm.nih.gov) on the basis of the principles outlined in the PRISMA guidelines (http://prisma-statement.org). Articles in English were included into the analysis, when they contained original data on the role of SAMHD1 in cancer. Moreover, reviews that discussed the potential impact of SAMHD1 on cancer were used to analyse conceptions and the predominant narrative in the field. Two reviewers independently analysed articles for relevant information and then agreed a list of relevant articles.

4.3. Results

4.3.1. High *SAMHD1* expression is not consistently associated with increased survival

Although SAMHD1 has recently been considered as a tumour suppressor protein (Coggins et al., 2020; Herold, N. et al., 2017; Mauney & Hollis, 2018), high *SAMHD1* expression in tumour tissues was not associated with favourable outcomes across all patients in the TCGA (Figure 4.1A). In contrast, high *SAMHD1* expression was associated with favourable outcome across all patients in the paediatric cancer database TARGET (Figure 4.1A).

Considering the individual cancer categories in the TCGA database, high *SAMHD1* expression was significantly associated with increased survival in nine out 33 of cancer categories (Figure 4.1B, Table S4.1) and with poor outcome in five cancer categories (Figure 4.1B, Table S4.1). This indicates that the role of SAMHD1 differs between cancer types and that it does not always function as a tumour suppressor. Similarly, high SAMHD1 expression was significantly correlated with longer survival in only one cancer type (osteosarcoma) in the TARGET database but with reduced survival in two others (acute lymphoblastic leukaemia, Wilm's tumour) (Figure 4.1B, Table S4.2).

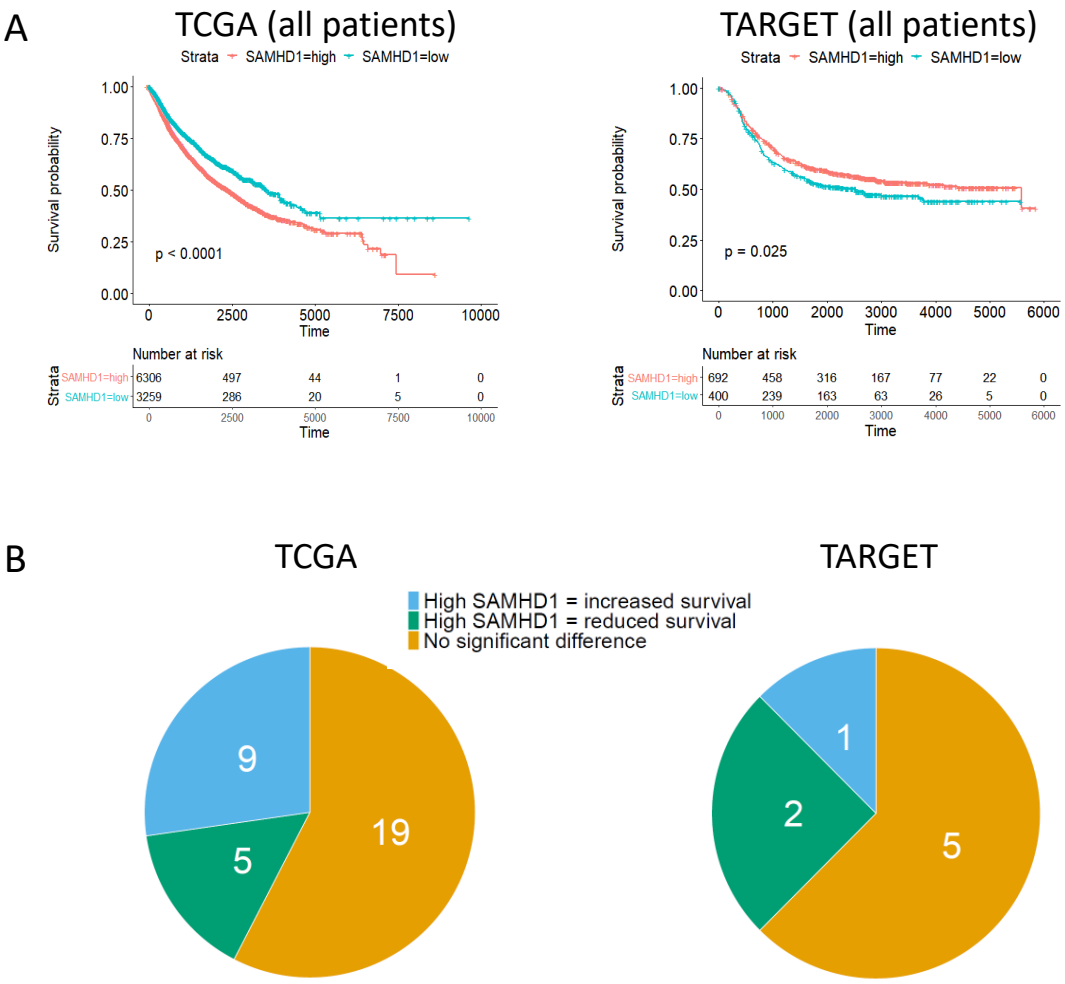


Figure 4.1. Effect of *SAMHD1* expression in cancer patients. (**A**) Kaplan Meier plots indicating survival in cancer patients with tumours characterised by high or low *SAMHD1* expression (as determined by best separation) across all patients in the TCGA and TARGET databases. P-values were determined by log-rank test. (**B**) Pie charts indicating the number of cancer types for which high SAMHD1 expression was associated with increased survival, reduced survival, or not significantly associated with survival based on data from the TCGA and TARGET databases. Data are presented in Table S4.1 and Table S4.2.

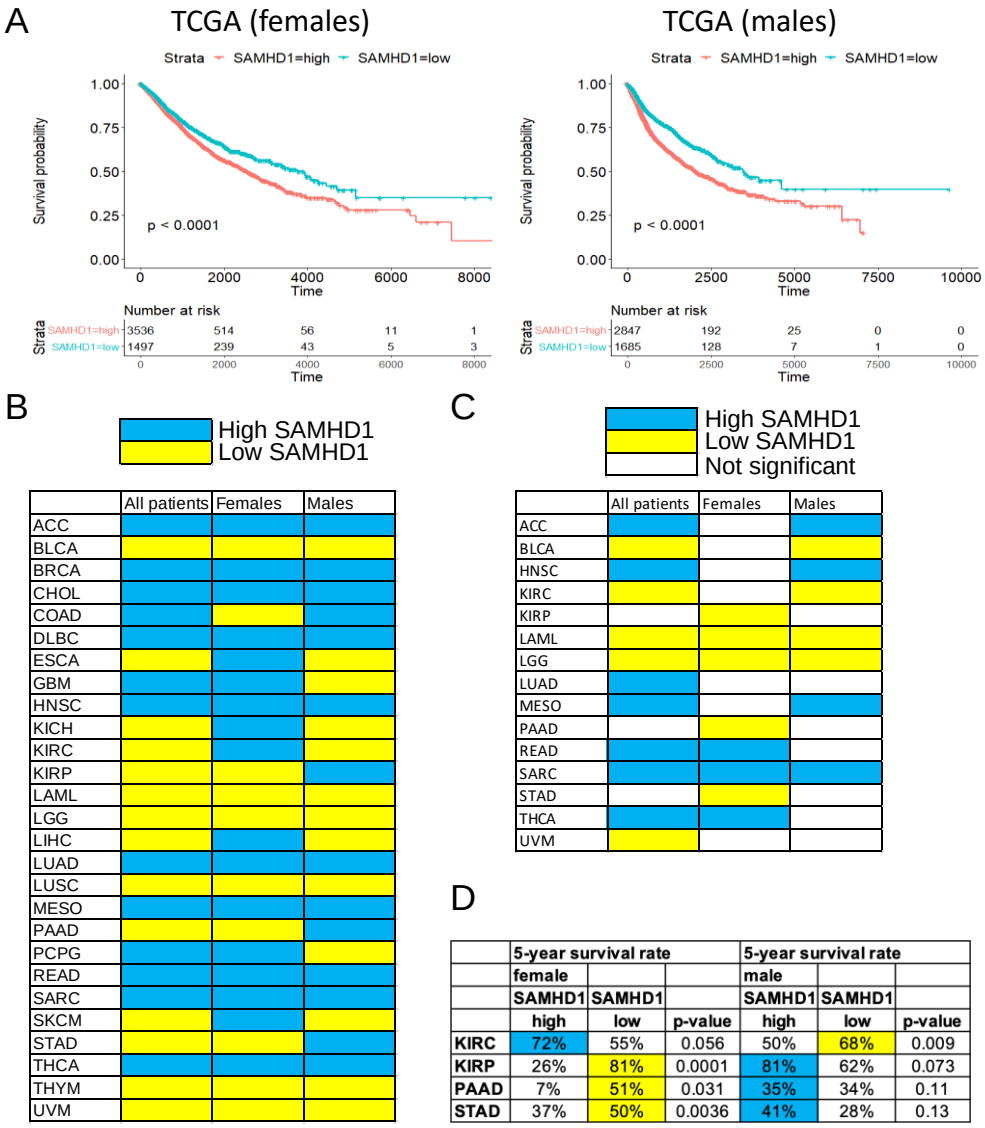
4.3.2. Role of *SAMHD1* expression in the context of sex

Next, we analysed the role of SAMHD1 in males and females in TCGA and TARGET. In TCGA, there were no sex-specific differences with regard to the association of SAMHD1 with survival time across all cancer types (Figure 4.2A). However, some discrepancies became visible upon the comparison of the role of SAMHD1 in the 27 cancer entities that occur in both females and males (Figure 4.2B, Table S4.3). When we did not consider statistical significance levels, high SAMHD1 levels were associated with higher 5-year survival rates in

13 cancer entities across all patients, in 17 cancer entities in female patients, and in 14 cancer entities in male patients (Figure 4.2B, Table S4.3).

When we only considered comparisons in which the 5-year survival rates were significantly different ($p<0.05$) between high and low SAMHD1-expressing tumours for at least one comparison (across all patients, in females, and/ or males), differences reached significance for only one sex in ten cancer types (Figure 4.2C, Table S4.3). Consistent findings were obtained for three cancer types (LAML, LGG, SARC, all abbreviations for cancer entities are provided in Table S4.1 and the legend of Figure 4.2).

Although differences did not reach our cut-off value for statistical significance ($p < 0.05$), trends were detected indicating opposite effects of SAMHD1 on disease outcome between the sexes in four cancer entities, (Figure 4.2D, Table S4.3). In kidney renal clear cell carcinoma (KIRC), low SAMHD1 levels were associated with a higher 5-year survival rate (68%) than high SAMHD1 levels (50%) in males ($p=0.009$). In contrast, high SAMHD1 levels were related to higher survival in females (72% vs. 55%), with the p-value being close to significance ($p=0.056$). In three other cancer types (KIRP, PAAD, STAD), low SAMHD1 levels were associated with higher 5-year survival in females and with lower 5-year survival in males (Figure 4.2D, Table S4.3).



	All patients	Females	Males
ACC			
BLCA			
BRCA			
CHOL			
COAD			
DLBC			
ESCA			
GBM			
HNSC			
KICH			
KIRC			
KIRP			
LAML			
LGG			
LIHC			
LUAD			
LUSC			
MESO			
PAAD			
PCPG			
READ			
SARC			
SKCM			
STAD			
THCA			
THYM			
UVM			

	All patients	Females	Males
ACC			
BLCA			
HNSC			
KIRC			
KIRP			
LAML			
LGG			
LUAD			
MESO			
PAAD			
READ			
SARC			
STAD			
THCA			
UVM			

D

	5-year survival rate			5-year survival rate		
	female			male		
	SAMHD1 high	SAMHD1 low	p-value	SAMHD1 high	SAMHD1 low	p-value
KIRC	72%	55%	0.056	50%	68%	0.009
KIRP	26%	81%	0.0001	81%	62%	0.073
PAAD	7%	51%	0.031	35%	34%	0.11
STAD	37%	50%	0.0036	41%	28%	0.13

Figure 4.2. *SAMHD1* expression levels and 5-year survival rates in dependence of sex based on TCGA data. (**A**) Kaplan Meier plots indicating sex-specific survival in cancer patients with tumours characterised by high or low *SAMHD1* expression (as determined by best separation). P-values were determined by log-rank test. (**B**) Heatmap indicating the association of *SAMHD1* expression and 5-year survival rates (blue: high SAMHD1 associated with higher survival rates, yellow: low SAMHD1 associated with higher survival rates). (**C**) Heatmap indicating cancer entities in which high *SAMHD1* expression (blue) or low *SAMHD1* expression (yellow) is significantly ($p<0.05$) associated with higher 5-year survival rates. (**D**) Cancer entities in which SAMHD1 displays a trend towards differing roles by sex. Blue indicates higher survival rates in patients with tumours with high SAMHD1 levels, yellow in patients with low SAMHD1 levels. Abbreviations: ACC, Adrenocortical carcinoma; BLCA, Bladder Urothelial Carcinoma; BRCA, Breast invasive carcinoma; CESC, Cervical squamous cell carcinoma and endocervical adenocarcinoma; CHOL, Cholangiocarcinoma; COAD, Colon adenocarcinoma; DLBC,

Lymphoid Neoplasm Diffuse Large B-cell Lymphoma; ESCA, Oesophageal carcinoma; GBM, Glioblastoma multiforme; HNSC, Head and Neck squamous cell carcinoma; KICH, Kidney Chromophobe; KIRC, Kidney renal clear cell carcinoma; KIRP, Kidney renal papillary cell carcinoma; LAML, Acute Myeloid Leukaemia; LGG, Low Grade Glioma; LIHC, Liver hepatocellular carcinoma; LUAD, Lung adenocarcinoma; LUSC, Lung squamous cell carcinoma; MESO, Mesothelioma; OV, Ovarian serous cystadenocarcinoma; PAAD, Pancreatic adenocarcinoma; PCPG, Pheochromocytoma and Paraganglioma; PRAD, Prostate adenocarcinoma; READ, Rectum adenocarcinoma; SARC, Sarcoma; SKCM, Skin Cutaneous Melanoma; STAD, Stomach adenocarcinoma; TGCT, Testicular Germ Cell Tumours; THCA, Thyroid carcinoma; THYM, Thymoma; UCEC, Uterine Corpus Endometrial Carcinoma; UCS, Uterine Carcinosarcoma; UVM, Uveal Melanoma.

In TARGET, high SAMHD1 levels were significantly associated with increased survival across all cancer types in females (Figure 4.3A) but not in males (Figure 4.3B). For seven paediatric cancer types, data were available for both sexes. When we did not consider statistical significance levels, higher 5-year survival rates were recorded for SAMHD1 patients with SAMHD1 high tumours across all patients, in one entity in female patients, and in four entities in male patients (Figure 4.3B, Table S4.4).

When we only considered cancer types in which the 5-year survival rates were significantly different ($p<0.05$) between high and low SAMHD1-expressing tumours for at least one comparison (across all patients, in females, and/ or males), differences reached significance only for females in two cancer types (Figure 4.3C, Table S4.4). Taken together, these data suggest that the role of SAMHD1 in cancer may differ between the sexes in some cancer types.

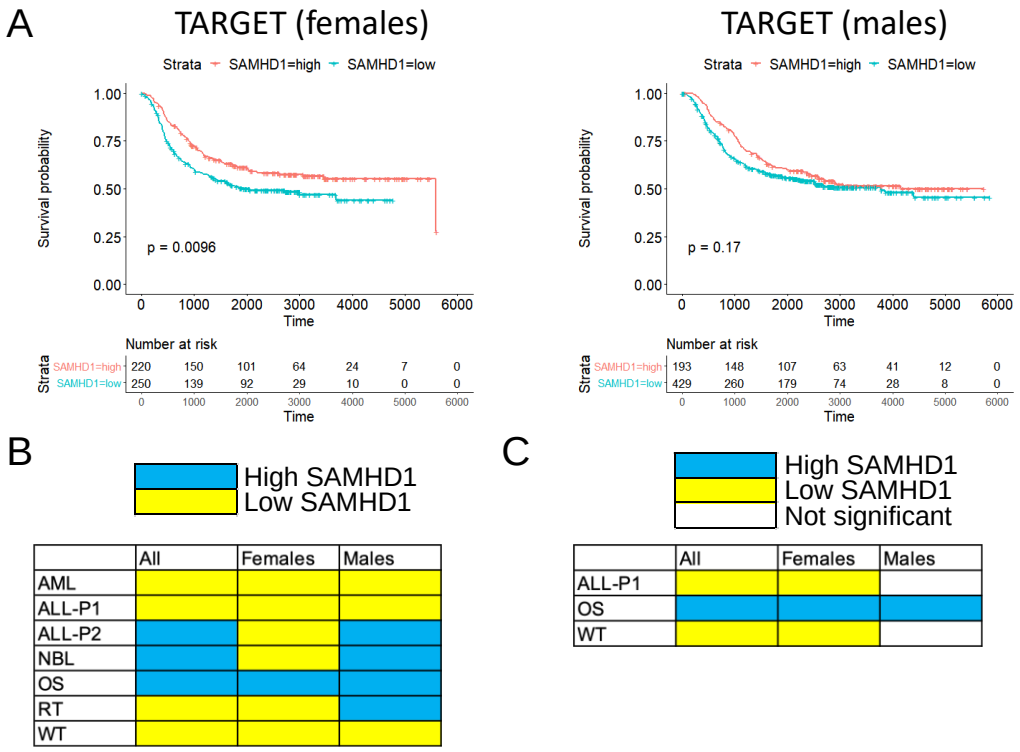


Figure 4.3. *SAMHD1* expression levels and 5-year survival rates in dependence of sex based on TARGET data. (**A**) Kaplan Meier plots indicating sex-specific survival in cancer patients with tumours characterised by high or low *SAMHD1* expression (as determined by best separation). P-values were determined by log-rank test. (**B**) Heatmap indicating the association of *SAMHD1* expression and 5-year survival rates (blue: high SAMHD1 associated with higher survival rates, yellow: low SAMHD1 associated with higher survival rates). (**C**) Heatmap indicating cancer entities in which high *SAMHD1* expression (blue) or low *SAMHD1* expression (yellow) is significantly ($p<0.05$) associated with higher 5-year survival rates.

Notably, the potentially different roles between SAMHD1 in female and male cancer patients do not appear to be the consequence of sex-specific discrepancies in *SAMHD1* expression. In TCGA, there were no significant sex-specific differences in *SAMHD1* expression between tumour samples and matched normal tissue samples from females and males, when we excluded sex-specific cancer types (CESC, OV, PRAD, TGCT, UCEC, UCS) and BRCA (only 12 out of 1,089 tumour tissue samples from male patients, only one out of 113 matched normal tissue samples from a male) (Figure S4.1). SAMHD1 expression was also not significantly different in males and females in the TARGET database (Figure S4.2).

4.3.3. Role of *SAMHD1* expression in the context of race

The vast majority of data in TCGA and TARGET are derived from white individuals, which reduces the significance of the race-related data. In TCGA, high SAMHD1 expression was associated with reduced overall survival in white patients (Figure 4.4). This reflects the findings obtained across all patients (Figure 4.1A) and probably that 6,834 (82%) out of 8,319 patients, for whom race data are available, are reported to be white. Apart from this, a significant difference in outcome in dependence of tumour SAMHD1 levels was only detected in Native Hawaiian or other Pacific islander patients, in whom high SAMHD1 was associated with improved survival (Figure 4.4). However, only 13 individuals fell into this category. Cancer-type specific comparisons did not reveal significant differences in SAMHD1-related outcomes between racial groups (Figure S4.3, Table S4.5), which may be due to the low numbers of patients in most of the categories (Figure 4.4). SAMHD1 levels were generally similar between the different race groups (Figure S4.4). Only Native Hawaiian or other Pacific islander patients displayed increased levels (Figure S4.4).

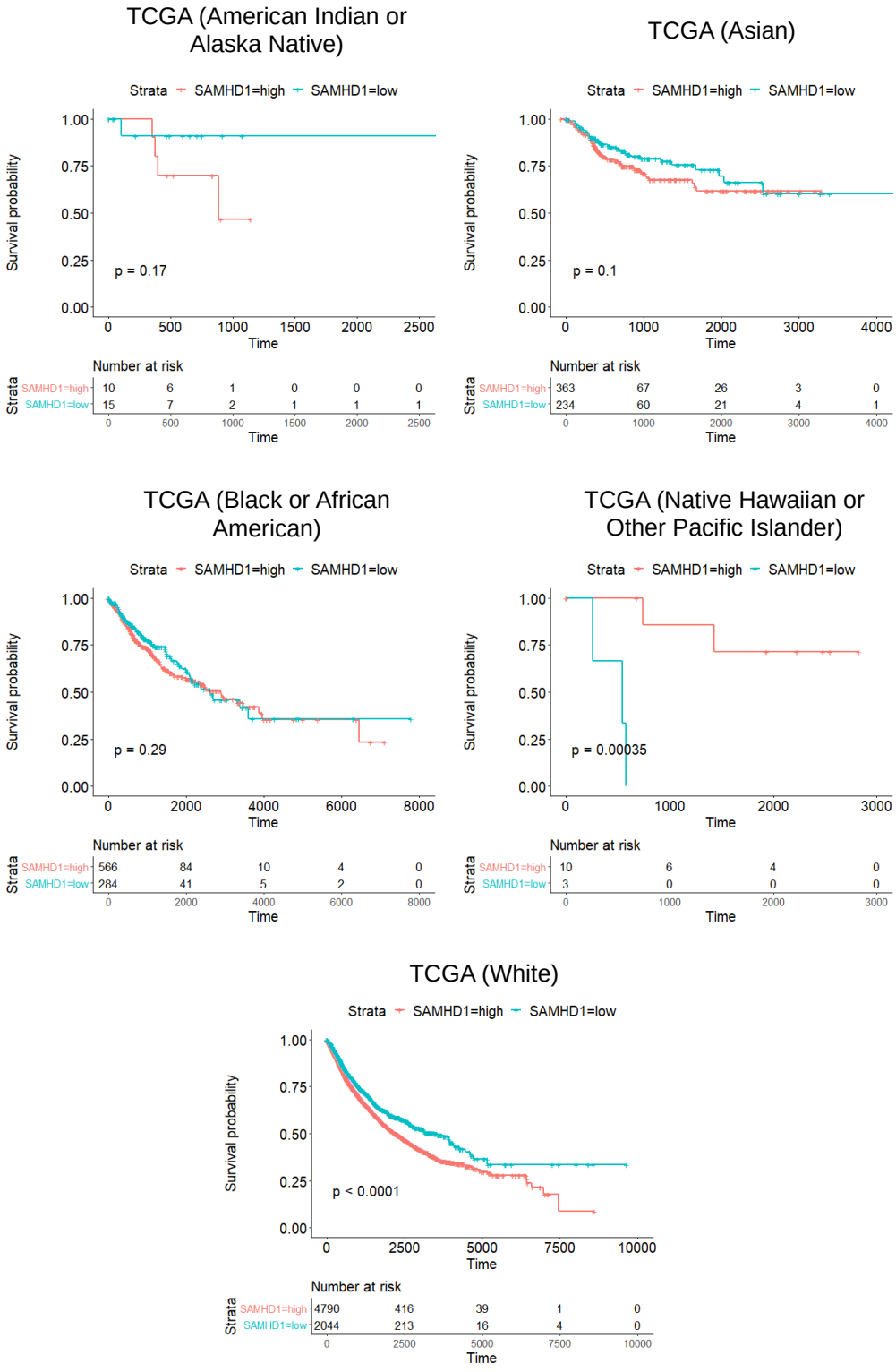


Figure 4.4. Kaplan-Meier plots indicating survival in cancer patients of different race with tumours characterised by high or low *SAMHD1* expression (as determined by best separation) based on TCGA data. P-values were determined by log-rank test.

Stratifying of patients in the TARGET database according to race provided some trends, which may point towards differences, but the numbers are too low to draw firm conclusions (Figure 4.5, Figure S4.5, Table S4.6). No significant differences were detected between the SAMHD1 levels in the different race groups (Figure S4.6).

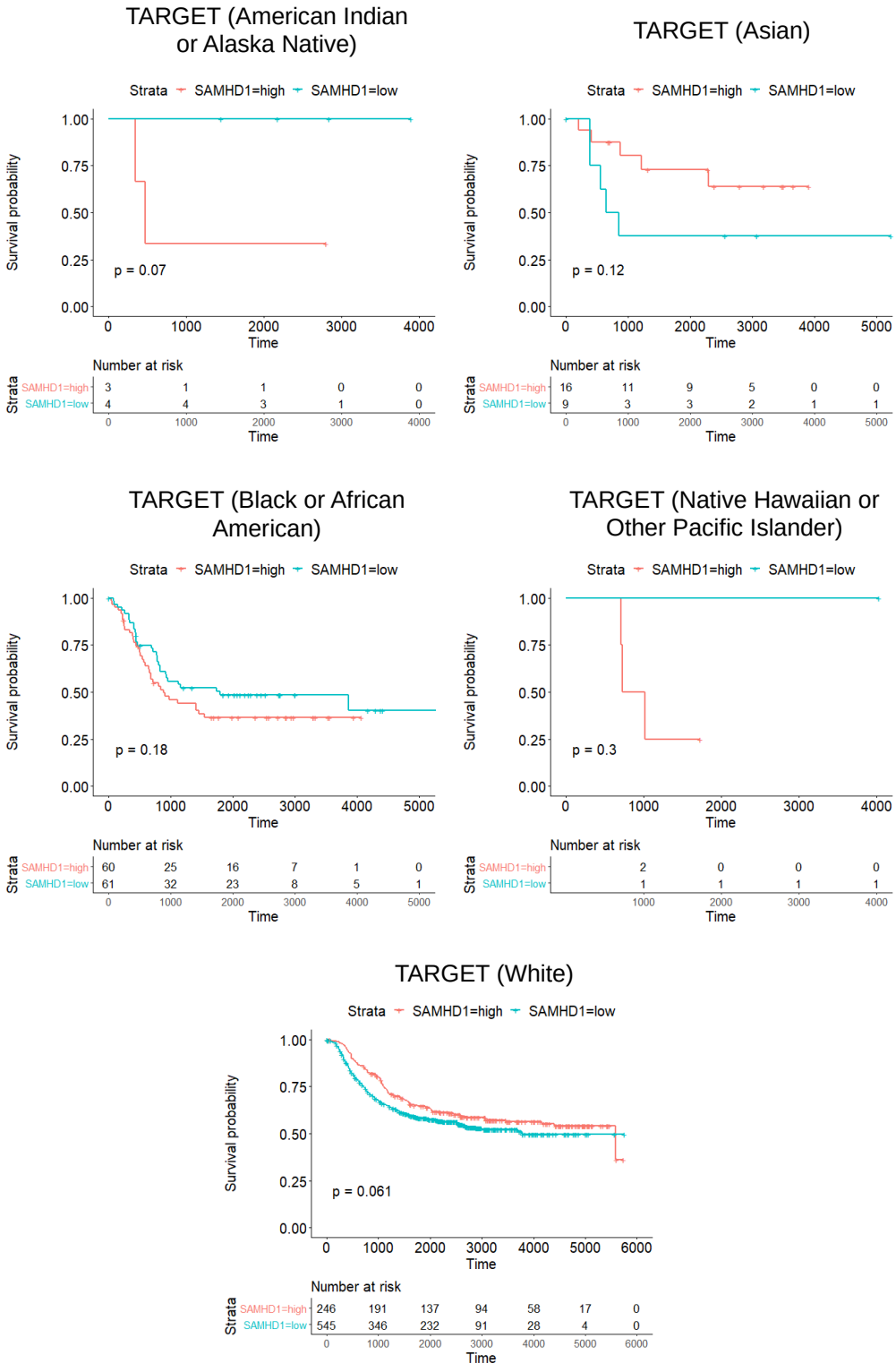


Figure 4.5. Kaplan-Meier plots indicating survival in cancer patients of different race with tumours characterised by high or low *SAMHD1* expression (as determined by best separation) based on TARGET data. P-values were determined by log-rank test.

4.3.4. *SAMHD1* expression in tumour vs matched normal samples

To further investigate the potential role of SAMHD1 in cancer, we next compared *SAMHD1* expression data in tumour tissue and matched normal samples, which were available for 695 patients and 21 cancer types in TCGA. Across all patients, there was no significant difference between the *SAMHD1* FPKM (fragments per kilobase of transcript per million mapped reads) values of tumour samples and matched normal samples (Wilcoxon signed-rank test p-value = 0.14). However, when stratifying by cancer type, SAMHD1 levels significantly differed ($p<0.05$) between tumour samples and matched control samples in seven cancer types (Table S4.7). SAMHD1 was higher in matched control samples suggesting tumour suppressor activity in three cancer types (BLCA, LUAD, LUSC) and higher in tumour samples from four cancer types (KIRC, KIRP, LIHC, STAD) suggesting oncogenic action (Table S4.7).

Next, we compared the results on potential tumour suppressor or oncogenic functions of SAMHD1 from tumour and matched normal tissues (Table 4.7) to those obtained from analysing 5-year survival in cancer patients with SAMHD1 low or high tumours (Figure 4.2; Table S4.1). When we did not consider statistical significance levels, SAMHD1 levels were higher in control tissues suggesting tumour suppressor activity in ten cancer entities (Figure 4.6A, Table S4.7). In twelve of the 21 cancer types, both SAMHD1 levels in tumour and matched normal tissues and the relationship of 5-year survival and tumour SAMHD1 levels indicated a similar role of SAMHD1, i.e. tumour suppressor (higher *SAMHD1* expression in matched normal tissue, higher 5-year survival in patients with SAMHD1 high tumours) or oncogenic (higher *SAMHD1* expression in tumour tissues, higher 5-year survival in patients with SAMHD1 low tumours) activity (Figure 4.6A, Table S4.7).

In the next step, only cancer entities were considered for which at least one of the comparisons had resulted in a statistically significant ($p<0.05$) difference, leaving 13 cancer types (Figure 4.6B, Table S4.7). In seven of these 13 cancer types, the anticipated role of SAMHD1 (tumour suppressor or oncogenic) coincided between both comparisons (Figure 4.6B, Table S4.7).

In only three cancer entities (BLCA, KIRC, LUAD), the differences reached statistical significance for both comparisons (Figure 4.6C, Table S4.7). SAMHD1 consistently displayed oncogenic activity in KIRC (Kidney renal clear cell carcinoma) and tumour suppressor activity in LUAD (Lung adenocarcinoma). In BLCA (Bladder Urothelial Carcinoma), higher SAMD1 levels in matched normal tissue samples suggested tumour suppressor activity, whereas higher 5-year survival in patients with SAMHD1 low tumours suggested oncogenic effects

(Figure 4.6C, Table S4.7). Hence, SAMHD1 may exert oncogenic activity in KIRC and tumour suppressor activity in LUAD, but clear evidence is lacking for other cancer entities.

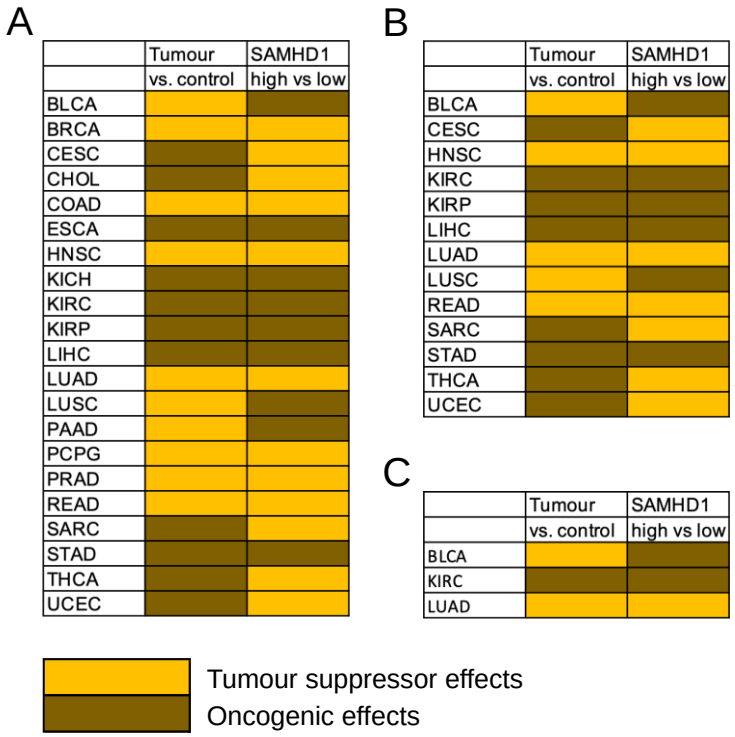


Figure 4.6. Tumour suppressor and oncogenic effects of SAMHD1 in different cancer types, as suggested by SAMHD1 levels in tumour tissues vs. matched normal tissues (Tumour vs. control) or the comparison of 5-year survival in patients with SAMHD1 high or low tumours (SAMHD1 high vs. low). Higher SAMHD1 levels in matched normal tissues were interpreted as tumour suppressor activity, while higher SAMHD1 levels in tumour tissues as indication of oncogenic effects. Higher 5-year survival in patients with SAMHD1 high tumours was construed as sign of tumour suppressor activity, higher 5-year survival in patients with SAMHD1 low tumours indication of oncogenic effects. (**A**) Data for all available comparisons. (**B**) Data for entities, in which at least the difference for one comparison reached statistical significance. (**C**) Data for entities, in which the difference for both comparisons reached statistical significance.

4.3.5. SAMHD1 regulation by methylation and miRNAs

Promotor methylation and miRNAs have been described to be involved in SAMHD1 regulation (Chen, Z. et al., 2021; de Silva et al., 2013; Kohnken et al., 2017). Tumour and normal sample SAMHD1 expression and promoter methylation beta values were available for 18 cancer types in TCGA. SAMHD1 promotor methylation significantly inversely

correlated with SAMHD1 expression levels across all patients, but the correlation coefficient was moderate and the relationship appears weak (Figure 4.7A).

When we looked at the individual cancer types, an inverse correlation between SAMHD1 expression and promotor methylation was detected in 18 cancer entities (Table S4.8). In eleven of these cancer types, the inverse correlations displayed p-values < 0.05 (Table S4.8). The cancer types with the strongest inverse correlations between SAMHD1 expression and promotor methylation were TGCT (Testicular Germ Cell Tumours), THYM (Thymoma), and CHOL (Cholangiocarcinoma) (Figure 4.7A). There were also 15 cancer entities with a direct correlation between SAMHD1 expression and promotor methylation, only four of which were associated with a p-value < 0.05. These data suggest that promotor methylation is one SAMHD1 regulation mechanism among others and that its role differs between cancer types.

Eight miRNAs (mir-30a, mir-155, and six subtypes of mir-181) have been described to be involved in SAMHD1 regulation (Jin, C. et al., 2014; Kohnken et al., 2017; Pilakka-Kanthikeel et al., 2015; Riess et al., 2017). Moreover, 21 miRNAs were indicated to interact with SAMHD1 in the DIANA-TarBase v8 (http://www.microrna.gr/tarbase) (Karagkouni et al., 2018), an online resource that lists experimentally validated miRNA/mRNA interactions. After the removal of overlaps, this resulted in a list of 28 miRNAs with a documented effect on SAMHD1 (Table S4.9).

Each of these 28 miRNAs were found to be inversely correlated with SAMHD1 expression in between two (mir-155) and all 28 (mir-23b and mir-183) cancer entities (Table S4.10). Six miRNAs (mir-23b, mir-30a, mir-192, mir-181d, mir-218-1, mir-218-2) were significantly ($p<0.05$) inversely correlated with SAMHD1 across all patients, with mir-23b showing the strongest inverse correlation (Figure 4.7B, Table S4.10). The strongest inverse correlation in a cancer type was detected between mir-23b and SAMHD1 in TGCT ($R= -0.54$, $p=7.37x10^{-13}$) (Figure 4.7B, Table S4.10).

Taken together, SAMHD1 levels are determined by complex regulation mechanisms that include promotor methylation and miRNAs, together with post-translational modifications such as phosphorylation and acetylation that have also been described (Chen, Z. et al., 2021; Coggins et al., 2020).

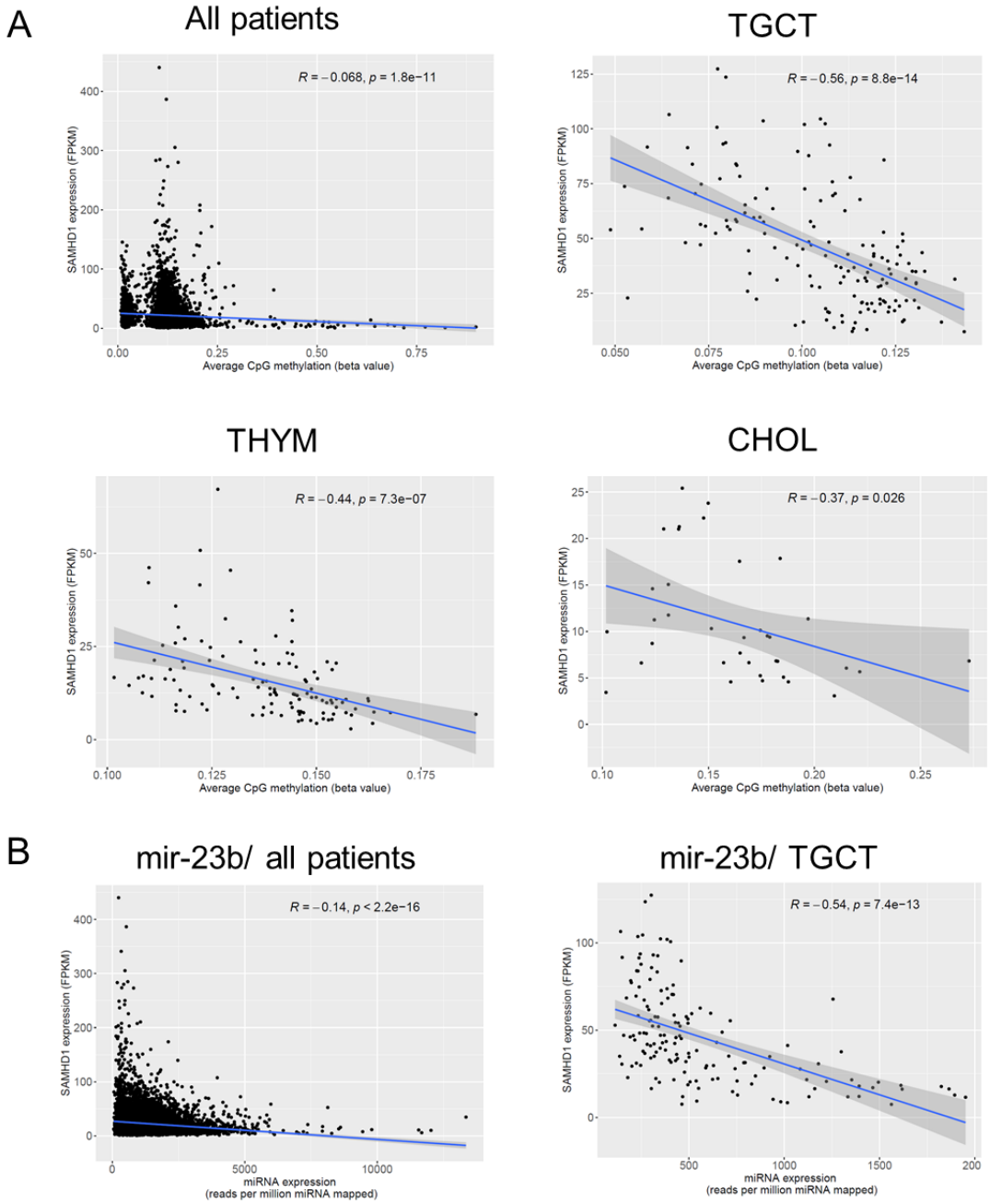


Figure 4.7. Inverse correlation between SAMHD1 promotor methylation levels or miRNA levels and SAMHD1 expression based on TCGA data. (**A**) Correlation between SAMHD1 promotor methylation levels and SAMHD1 expression across all patients and in THYM patients, which displayed the strongest inverse correlation across all cancer types. Data for all cancer types are presented in Table S4.8. (**B**) Correlation of mir-23b with SAMHD1 expression across all patients and of mir-30c-1 with SAMHD1 in THYM. mir-23b was the miRNA that displayed the strongest inverse correlation with SAMHD1 across all patients. The inverse correlation between mir-30c-1 and SAMHD1 was the strongest among all miRNAs in all cancer types. Data for all significant inverse correlations of miRNAs and SAMHD1 across all cancer types are provided in Table S4.10.

4.3.6. *SAMHD1* mutations and patient survival

SAMHD1 mutations have been described in cancers including chronic lymphocytic leukaemia, T-cell prolymphocytic leukaemia, mantle cell lymphoma, cutaneous T-cell lymphoma and colon cancer (Amin et al., 2016; Bühler et al., 2021; Burns et al., 2018; Clifford et al., 2014; Guièze et al., 2015; Johansson, P. et al., 2018; Merati et al., 2015; Nadeu et al., 2020; Rentoft et al., 2016; Roider et al., 2021).

Mutation data was available for 10,149 patients in the TCGA. 15,351 out of 21,156 genes harboured at least one non-synonymous mutation in one patient (Table S4.11). The three most commonly mutated genes were *TTN*, *MUC16*, and *TP53* (Table S4.11). *TTN* and *MUC16* encode the two longest human proteins (36,800 and 14,500 amino acids, respectively) that are frequently found mutated. Mutations in these genes are commonly regarded not to be of functional relevance and removed as artefacts or used as indicators of the mutational burden of tumours, while *TP53* is known to be the most commonly mutated tumour suppressor gene (Kim, Y. A. et al., 2017; Lawrence et al., 2013; Levine, 2020; Oh et al., 2020; Wang, X. et al., 2020; Yang, Y. et al., 2020).

In total, *SAMHD1* was mutated 201 times, including 175 non-synonymous mutations in 159 patients (1.57% of patients for whom mutation data was available) (Table S4.11). This places *SAMHD1* within the top 15.3% of most commonly mutated genes (Figure 4.8A, Table S4.11). Among the 135 patients with *SAMHD1* mutant tumours for whom survival data were available, *SAMHD1* mutations were associated with superior outcome (Figure 4.8B, Table S4.12). In 18 of the 25 cancer types, in which *SAMHD1* mutations were detected, 5-year survival was higher in patients with *SAMHD1* mutant tumours (Table S4.12). However, the significance of these data is limited due to the low number of SAMHD1 mutations. Notably, the p-value (0.07) was close to significance in UCEC, the cancer type with the most SAMHD1 mutations (35/ 6.6% out of 527), in which 93.2% of patients with SAMHD1 mutant cancers survived for five years, in contrast to 76.1% of the 492 UCEC patients with SAMHD1 wild-type cancers (Figure 4.8C, Table S4.12).

Although it is not possible to draw firm conclusions from these data, they do not support a general tumour suppressive role of SAMHD1, as mutations in tumour suppressor genes would rather be expected to result in shorter survival. For example, mutations in *TP53*, the most commonly mutated tumour suppressor gene (Levine, 2020), were associated with reduced survival (Figure 4.8D).

4.3.7. *SAMHD1* mutations are likely to be deleterious

Twenty-nine of the mutations are likely to result in a loss of function, including 11 stop-gain, 11 frameshift, six splice site and one stop loss mutation. While 21 mutations were located in untranslated regions (six 5' UTR, 15 3' UTR), four were in introns, three in-frame, 25 were synonymous, with the remaining 104 resulting in nonsynonymous mutations.

50 mutations had already been described in cancer cells or were present in positions that had been found mutated in cancer cells (Table S4.11). Three of the *SAMHD1* mutations identified in the TCGA (R143C/ UCEC patient, R145Q/ COAD patient, R290H/ STAD patient) were loss-of-function mutations associated with Aicardi-Goutières syndrome (Coggins et al., 2020; Mauney & Hollis, 2018; Rice, G. I. et al., 2009; UniProt Consortium, 2021). 18 nonsynonymous mutations occurred at positions demonstrated to be important for SAMHD1 function by mutagenesis studies according to UniProt (Figure 4.8E, Figure 4.8F, Table S4.11). This was supported by structural analysis which showed that ten non-synonymous mutations were located around the SAMHD1 active ligand binding sites (Figure 4.8F).

Next, the SAMHD1 non-synonymous variants were analysed using SIFT (Sim et al., 2012), PolyPhen-2 (Ng, P. C. & Henikoff, 2001; Sim et al., 2012; Vaser et al., 2016), Condel (González-Pérez & López-Bigas, 2011), and CADD (Kircher et al., 2014) to predict if they are likely to have an effect on protein function (Table 4.11). Approximately half of the amino acid changes were predicted to have a significant impact on SAMHD1 function (SIFT: 63/104 (60.6%), Polyphen-2: 50/104 (48.1%) and Condel: 54/104 (51.9%)). 72 of these variants also had a scaled CADD score of >20, which rates a variant among the top 1% of the most deleterious changes. Five variants displayed CADD scores >30 (Figure 4.8E). 39 variants had a SIFT rating of 'tolerated', a PolyPhen-2 rating of 'benign' and a Condel rating of 'neutral', of which 13 had a scaled CADD score of <10. Predictions for the remaining 17 amino acid changes were inconsistent (i.e. contrasting SIFT, PolyPhen-2 and Condel predictions) (Table 4.11).

Taken together, many of the mutations appear to affect SAMHD1 function. However, loss-of-function should typically be associated with reduced survival in tumour suppressor genes.

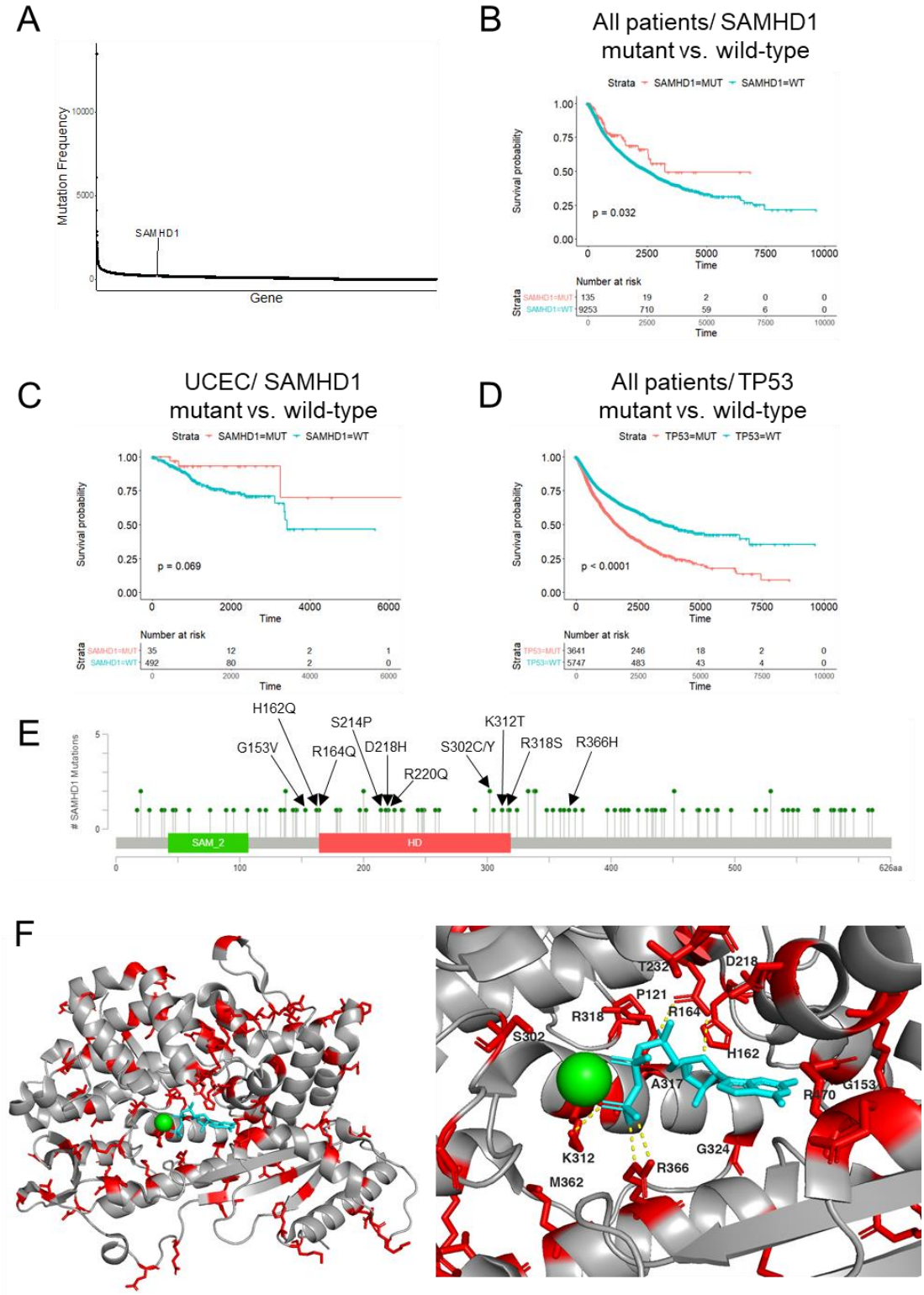


Figure 4.8. *SAMHD1* mutations in cancer tissues. (**A**) *SAMHD1* was mutated 201 times, including 175 non-synonymous mutations, which puts *SAMHD1* within the 15.3% of most commonly mutated genes. (**B**) Survival in patients with and without *SAMHD1* mutant tumours. (**C**) Survival in UCEC (cancer type with the most *SAMHD1* mutations) patients with and without *SAMHD1* mutant tumours. (**D**) Survival in patients with tumours with or without mutations in *TP53*, the most commonly mutated tumour suppressor genes. (**E**) Lollipop plot indicating locations of missense mutations in *SAMHD1*. Residues predicted to be involved in ligand binding are labelled in bold. (**F**) nonsynonymous mutations mapped (coloured red) onto the SAMHD1 protein structure (Protein Data Bank identifier 6DWD (Knecht et al., 2018) with bound clofarabine hydrochloride (indicated in cyan) and magnesium ion (green). The image on the left shows the full structure and on the right the active site is displayed. Yellow dashed lines indicate hydrogen bonds between mutated residues and ligand.

4.3.8. Literature review SAMHD1 and cancer

Our analysis of TCGA and TARGET data do not suggest that SAMHD1 generally functions as a tumour suppressor protein. Neither *SAMHD1* mutations nor low SAMHD1 levels were consistently associated with reduced survival. However, SAMHD1 is discussed as a potential tumour suppressor protein in the literature (Chen, Z. et al., 2021; Herold, N. et al., 2017). Next, we performed a systematic review to compare our findings to those from the literature and to gain further insights into the narrative underlying the perceived role of SAMHD1 in cancer.

The literature search was performed in PubMed (https://pubmed.ncbi.nlm.nih.gov) on 17th June 2021 using the search term "(((Cancer) OR (tumor) OR (tumour))) AND (SAMHD1)". It resulted in 150 hits, including 35 articles with relevant original data and 15 relevant secondary literature articles (reviews, editorials, comments) (Figure S4.7, Table S4.13).

The first articles reported on a potential role of SAMHD1 in cancer in 2013 (Clifford et al., 2014; de Silva et al., 2014; Shi, Y. et al., 2014). The first one reported on low SAMHD1 levels in patients with Sézary syndrome, an aggressive subtype of cutaneous T-cell lymphoma, due to SAMHD1 methylation (de Silva et al., 2014), while the second paper described SAMHD1 variants associated with hepatitis B virus- and hepatitis C virus-induced hepatocellular carcinoma (Shi, Y. et al., 2014). The third paper found SAMHD1 mutations in chronic lymphocytic leukaemia and proposed that these mutations promote leukaemia development by affecting SAMHD1-mediated DNA repair (Clifford et al., 2014).

Among the 35 articles that reported on an association between SAMHD1 and cancer, four did not (entirely) support the 'SAMHD1 is a tumour suppressor' narrative (Kodigepalli et al., 2018; Shang, Z. et al., 2018; Xagoraris et al., 2021; Yang, C. A. et al., 2017). One study correlated high *SAMHD1* levels with metastasis formation in colorectal cancer (Yang, C. A. et al., 2017). Kodigepalli et al., (2018) reported that SAMHD1 knock-out increases acute myeloid leukaemia cell proliferation via PI3K signalling but inhibits tumourigenesis potentially due to a lack of SAMHD1-mediated TNFalpha suppression. Notably, the title of this study exclusively focused on the inhibitory effects of SAMHD1 on leukaemia cell proliferation and did not mention its role in tumourigenesis (Kodigepalli et al., 2018). One study detected a SAMHD1 increase upon lung cancer progression (Shang, Z. et al., 2018), and the most recent study, correlated the presence of SAMHD1 in Hodgkin lymphoma cells with unfavourable outcome (Xagoraris et al., 2021). Notably, this study (Xagoraris et al., 2021) even referred to SAMHD1 as "novel tumour suppressor" in the title, although the study rather indicated an oncogenic role of SAMHD1.

The remaining articles largely focused on SAMHD1 mutations and reduced SAMHD1 levels in different cancer types as well as on SAMHD1's potential role as a tumour suppressor involved in DNA repair (Table S4.13). *SAMHD1* mutations were detected in patients with hepatocellular carcinoma (Shi, Y. et al., 2014), chronic lymphocytic leukaemia (Amin et al., 2016; Burns et al., 2018; Clifford et al., 2014; Guièze et al., 2015; Kim, J. A. et al., 2016), cutaneous T-cell lymphoma (Merati et al., 2015), colorectal cancer (Rentoft et al., 2016), T-cell prolymphocytic leukaemia (Johansson, P. et al., 2018), acute myeloid leukaemia (Zhu, K. W. et al., 2018), and mantle cell lymphoma (Bühler et al., 2021; Nadeu et al., 2020).

In our TCGA analysis performed above (Figure S4.8, Table S4.12), there is rather a trend towards higher 5-year survival rates among hepatocellular carcinoma (LIHC) patients with *SAMHD1* mutant tumours, although the significance of the data is limited due to low numbers (Table S4.12). All four patients with *SAMHD1* mutant hepatocellular carcinoma survived for five years, while only 51.4% of 359 patients with *SAMHD1* wild-type hepatocellular carcinomas survived for five years. In colorectal adenocarcinoma (COAD), there was no noticeable difference between the survival of patients with *SAMHD1* mutant and *SAMHD1* wild-type tumours (Table S4.12). Only in rectal adenocarcinoma (READ), a trend suggested that patients with *SAMHD1* mutant tumours may have a worse outcome. None out of five patients with *SAMHD1* mutant tumours survived for five years, while 54.9% of 130 patients with *SAMHD1* wild-type tumours did (Table S4.12).

Acute myeloid leukaemia (LAML) was the only cancer type in which a significant difference was detected between patients with *SAMHD1* mutant and wild-type cancer cells. Patients with *SAMHD1* mutant leukaemia cells had a higher 5-year survival rate (Table S4.8). However, this is most probably not due to general oncogenic activity, but because lack of SAMHD1 function results in a higher activity of nucleoside analogues including cytarabine and decitabine that are SAMHD1 substrates and commonly used for LAML treatment (Oellerich et al., 2019; Schneider et al., 2017).

The study on colorectal cancer (Rentoft et al., 2016) was the only one that had used TCGA data. However, it only used TCGA data to identify mutations, but did not compare survival in patients with and without *SAMHD1* mutations.

SAMHD1 expression levels have been suggested to impact on cutaneous T-cell lymphoma (de Silva et al., 2014), lung cancer (Shang, Z. et al., 2018; Wang, J. L. et al., 2014), colorectal cancer (Yang, C. A. et al., 2017), and acute myeloid leukaemia (Jiang et al., 2020). Low SAMHD1 levels were described in cutaneous T-cell lymphoma and acute myeloid leukaemia cells (Jiang et al., 2020), supporting a potential role as a tumour suppressor. TCGA did not contain data on *SAMHD1* expression in cutaneous T-cell lymphoma or acute myeloid leukaemia cells relative to control cells.

In lung cancer, conflicting results were reported. One study found that SAMHD1 is down regulated in lung cancer by methylation and inhibits tumour cell proliferation (Wang, J. L. et al., 2014). The other study reported that SAMHD1 levels increase in the serum of lung cancer patients upon progression (Shang, Z. et al., 2018). Our analysis of SAMHD1 data found significantly higher 5-year survival rates in patients with tumours displaying high *SAMHD1* expression levels (Table S4.1) supporting the first study. Notably, elevated SAMHD1 in the serum of lung cancer patients may not have been derived from cancer tissue.

In colorectal cancer, low SAMHD1 levels were detected in tumour tissues relative to adjacent control tissues (Yang, C. A. et al., 2017), which agrees with a tumour suppressor function. However, higher SAMHD1 levels were associated with metastasis formation (Yang, C. A. et al., 2017), rather supporting an oncogenic role. The study included the analysis of TCGA data on colorectal cancer (Yang, C. A. et al., 2017), but no systematic analysis of SAMHD1 across different cancer entities.

The 15 relevant secondary literature articles all had narratives focussing on the potential role of SAMHD1 as a tumour suppressor (Table S4.13).

4.4. Discussion

SAMHD1 has been suggested to exert tumour suppressor functions due its role in maintaining genome integrity and as an inhibitor of uncontrolled proliferation (Chen, Z. et al., 2021; Herold, N. et al., 2017). However, our analysis of TCGA and TARGET data does not suggest that SAMHD1 should be regarded as a *bona fide* tumour suppressor. Notably, SAMHD1 mutations that interfere with SAMHD1 function were not associated with poor outcome, which is something that would be expected from a tumour suppressor. In agreement, no increased cancer formation has been described in SAMHD1-deficient animal models (Kohnken et al., 2017).

Our results rather indicated that changes in SAMHD1 are involved in the oncogenic process in a minority of cases and that it may exert pro- or antitumourigenic effects in different cancer types (and perhaps individual tumours). Moreover, the role of SAMHD1 may differ between the sexes and different races. These findings also show that our understanding of the processes underlying cancer needs to improve further, before a broad paradigm shift towards tumour-agnostic approaches (Danesi et al., 2021) can become a reality.

Notably, the interpretation of our findings may be affected by SAMHD1 being a triphosphohydrolase that cleaves and inactivates the triphosphorylated forms of a number of nucleoside analogues including cytarabine, decitabine, and nelarabine (Oellerich et al., 2019; Rothenburger et al., 2020; Schneider et al., 2017). However, most cancer diseases are not treated with SAMHD1 substrates. Notably, KIRC (kidney renal clear cell carcinoma), the only cancer in which 5-year survival is significantly lower in SAMHD1 high tumours and SAMHD1 levels are significantly higher in tumour than in control tissues, suggesting an oncogenic role of SAMHD1, is not treated with nucleoside analogues (Geynisman et al., 2021). Hence, the absence of tumour suppressor activity and/ or oncogenic effects cannot simply be explained by SAMHD1-mediated inactivation of nucleoside analogue substrates.

Our findings demonstrating that SAMHD1 plays multifaceted (and often, if any, minor) roles in cancer seem to be in disaccord with a perception and narrative forming in the field suggesting that SAMHD1 is a tumour suppressor (Chen, Z. et al., 2021; Herold, N. et al., 2017). A systematic review confirmed that most of the available literature focuses on a potential role of SAMHD1 as a tumour suppressor. Among 35 original articles on the role of SAMHD1 in cancer, 31 discussed a potential tumour suppressor function and three potential oncogenic effects. One article reported both potential tumour suppressor and oncogenic activity, but only mentioned the anticipated tumour suppressor effects in the title

(Kodigepalli et al., 2018). All 15 secondary literature articles (reviews, editorials, comments) had a narrative built around SAMHD1 being a candidate tumour suppressor.

The narrative that SAMHD1 is a tumour suppressor has formed since 2013 around findings in a limited number of cancer entities (Table S4.13). Three reasons may contribute to the perpetuation of such a narrative without much scrutiny. Firstly, SAMHD1 has been described to maintain genome integrity by a range of different mechanisms (Akimova et al., 2021; Chen, Z. et al., 2021; Herold, N. et al., 2017; Park et al., 2021). Hence, a potential tumour suppressor role is plausible and convincing. Further research will have to show why SAMHD1's multifaceted roles in DNA repair do not translate into a consistent and general tumour suppressor function.

The second potential reason is confirmation bias. Scientists (like everybody else) tend to accept findings that support their own experiences, assumptions, and perceptions and to disregard evidence that challenges them (Letrud & Hernes, 2019; Yanai & Lercher, 2021). Thus, researchers are more likely to look for data that support their hypothesis and not for those that contradict it. Notably, one study referred to SAMHD1 as "novel tumour suppressor" in the title, although *SAMHD1* expression was described as an adverse prognostic factor in Hodgkin lymphoma (Xagoraris et al., 2021).

The final potential reason is publication bias, i.e. a focus on 'positive' findings that are easier to publish in more prestigious journals than 'negative' findings (Begley & Ioannidis, 2015; Marks-Anglin & Chen, 2020; Nissen et al., 2016; Wass et al., 2019). In the case of studies investigating a potential role of SAMHD1 in cancer, this means that some studies that did not find a relationship between SAMHD1 and cancer may simply not have been published and that the publicly available data may not reflect all available data on the subject.

In conclusion, SAMHD1 can play multifaceted roles in cancer that may differ between different cancer types, the sexes, and races. In contradiction to the predominant narrative, SAMHD1 may exert oncogenic as well as tumour suppressor activity and may often be a minor (if any) player in carcinogenesis. Our findings emphasise that hypotheses, perceptions, and assumptions need to be continuously challenged by using all available data and evidence. In this context, it is important that all data are actually published and made available, even if they are not deemed particularly exciting by researchers. Finally, the increasing number of available data and databases should be effectively used to inform and challenge our research and research findings.

Chapter 5: Differentially conserved amino acid positions may reflect differences in SARS-CoV-2 and SARS-CoV behaviour

Denisa Bojkova[1,#], Jake E McGreig[2,#], Katie-May McLaughlin[2,#,] Stuart G Masterson[2,#], Magdalena Antczak[2] , Marek Widera[3] , Verena Krähling[3] , Sandra Ciesek[1,4], Mark N Wass[2], Martin Michaelis[2], Jindrich Cinatl Jr[1]

[1]Institute for Medical Virology, University Hospital, Goethe University Frankfurt am Main, Germany.

[2]School of Biosciences, University of Kent, Canterbury, UK.

[3]Institute of Virology, Biomedical Research Center (BMFZ), Philipps University Marburg, Germany.

[4]German Center for Infection Research, DZIF, Braunschweig, Germany

[#]equal contribution

Published: Bioinformatics. 09 February 2021; btab094. https://doi.org/10.1093/bioinformatics/btab094

Abstract

SARS-CoV-2 is a novel coronavirus currently causing a pandemic. Here, we performed a combined in-silico and cell culture comparison of SARS-CoV-2 and the closely related SARS-CoV. Many amino acid positions are differentially conserved between SARS-CoV-2 and SARS-CoV, which reflects the discrepancies in virus behaviour, i.e. more effective human-to-human transmission of SARS-CoV-2 and higher mortality associated with SARS-CoV. Variations in the S protein (mediates virus entry) were associated with differences in its interaction with ACE2 (cellular S receptor) and sensitivity to TMPRSS2 (enables virus entry via S cleavage) inhibition. Anti-ACE2 antibodies more strongly inhibited SARS-CoV than SARS-CoV-2 infection, probably due to a stronger SARS-CoV-2 S-ACE2 affinity relative to SARS-CoV S. Moreover, SARS-CoV-2 and SARS-CoV displayed differences in cell tropism. Cellular ACE2 and TMPRSS2 levels did not indicate susceptibility to SARS-CoV-2. In conclusion, we identified genomic variation between SARS-CoV-2 and SARS-CoV that may reflect the differences in their clinical and biological behaviour.

5.1. Introduction

In December 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a novel betacoronavirus, was identified that causes a respiratory disease and pneumonia called coronavirus disease 19 (COVID-19) (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020; Zhu, N. et al., 2020). As of 22nd of December 2020, 77 801 721 confirmed COVID-19 cases and 1 713 109 COVID-19 deaths have been reported (Dong et al., 2020). Since 2002, SARS-CoV-2 is the third betacoronavirus, after severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV), that has caused a substantial outbreak associated with significant mortality (Wu, A. et al., 2020).

SARS-CoV-2 is closely related to SARS-CoV (Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020; Wu, A. et al., 2020). Entry of both viruses is mediated via interaction of the viral Spike (S) protein with the cellular receptor ACE2, and both viruses depend on S activation by cellular proteases, in particular by TMPRSS2 (Cui et al., 2019; Wan et al., 2020; Wrapp et al., 2020; Wu, A. et al., 2020; Yan, R. et al., 2020). Despite these similarities, the diseases caused by SARS-CoV-2 (COVID-19) and SARS-CoV (SARS) differ. According to WHO, the SARS-CoV outbreak resulted in 8098 confirmed and suspected cases and 774 deaths, equalling a mortality rate of 9.6% (www.who.int). Estimated mortality rates for SARS-CoV-2 are below 1% (Borges do Nascimento et al., 2020). SARS-CoV was only spread by symptomatic patients with severe disease (Cheng, V. C. et al., 2013). In contrast, SARS-CoV-2 has been reported to be transmitted by individuals who are asymptomatic during the incubation period or who do not develop symptoms at all (Rivett et al., 2020).

We have developed an approach to identify sequence-associated phenotypic differences between related viruses based on the identification of differentially conserved amino acid sequence positions (DCPs) and in silico modelling of protein structures (Martell et al., 2019; Pappalardo et al., 2016). Conserved amino acid positions are likely to be of functional relevance, and differential conservation may indicate functional differences and they have been widely used for the analysis of protein families (Das et al., 2015; Rausell et al., 2010). Here, we used this method to identify differentially conserved positions that may explain phenotypic differences between SARS-CoV-2 and SARS-CoV. These data were combined with data derived from virus-infected cells.

5.2. Materials and methods

5.2.1 Structural analysis

Sequences for each of the SARS-CoV-2 proteins were obtained from the GISAID resource. The protein sequences were then filtered for sequences from human hosts with high coverage, and sequences with spans of X's were removed. The number of sequences retained after filtering for each protein is shown in Table S5.4. Fifty-three SARS-CoV genome sequences derived from human hosts were downloaded from VIPR (Pickett, Sadat et al., 2012; Pickett, Greer et al., 2012). Open Reading Frames (ORFs) were extracted using EMBOSS getorf (Rice, P. et al., 2000) and matched to known proteins using BLAST. Fragments and mismatches were discarded. To match the ORF1ab non-structural proteins, a BLAST database of the sequences from the SARS non-structural proteins was generated and the SARS-CoV-2 ORF1ab searched against it. The sequences for each protein were then aligned using ClustalO (Sievers et al., 2011) with default settings.

Conserved positions were identified by calculating the Jensen-Shannon divergence score (Capra & Singh, 2007) for each position in the multiple sequence alignment in virus. Differing alignment positions with conservation score >0.8 for both species were considered as differentially conserved positions (DCPs).

SARS-CoV-2 and SARS-CoV protein structures were downloaded from the Protein Data Bank (PDB; Table S5.1) (Armstrong et al., 2020). Where structures were not available, they were modelled using Phyre2 ((Kelley et al., 2015); Table S5.2). Where Phyre2 did not generate a confident model, structural models from AlphaFold were used (Senior et al., 2020). Ligand binding sites were modelled using 3DLigandSite (Wass et al., 2010). DCPs were mapped onto protein structures using PyMOL. Exposed (solvent-accessible) and buried (solvent-inaccessible) residues were identified using Python module findSurfaceResidues with default parameters. Amino acid changes at DCPs were manually analysed for their potential impact on protein structure and function based on the presence or absence of hydrogen bonding, changes in hydrogen bonding capacity and changes in charge in SARS-CoV compared with SARS-CoV-2 proteins. Where models were unavailable, mutagenesis was performed within PyMOL to assess the potential impact of the amino acid changes. The structural analysis grouped DCPs into six different categories based on the effect that they were proposed to have. These include 'unlikely', 'possible' and 'likely'. The possible and likely categories were split into three and two subgroups respectively depending on the type of effect (Table S5.3).

5.2.2 Cell culture

The Caco2 cell line was obtained from DSMZ (Braunschweig, Germany). The cells were grown at 37°C in minimal essential medium (MEM) supplemented with 10% foetal bovine serum (FBS), 100 IU/ml penicillin, and 100 µg/mL of streptomycin. 293 cells (PD-02-01; MicrobixBisosystems Inc.) and 293/ACE2 cells (Kamitani et al., 2006) (kindly provided by Shinji Makino, UTMB, Galveston, Texas) were cultured in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% FBS, 50 IU/mL penicillin and 50 µg/mL streptomycin. Selection of 293/ACE2 cells constitutively expressing human angiotensin-converting enzyme 2 (ACE2) was performed by addition of 12 µg/mL blasticidin. All culture reagents were purchased from Sigma (Munich, Germany). Cells were regularly authenticated by short tandem repeat (STR) analysis and tested for mycoplasma contamination.

5.2.3 Virus infection

The isolate SARS-CoV-2/1/Human/2020/Frankfurt (Hoehl et al., 2020) was cultivated in Caco2 cells as previously described for SARS-CoV strain FFM-1 (Cinatl Jr et al., 2004). Virus titres were determined as TCID50/ml in confluent cells in 96-well microtitre plates (Cinatl Jr et al., 2005; Cinatl et al., 2003).

5.2.4 Western blot

Western blotting was performed as previously described (Schneider et al., 2017). Briefly, cells were lysed using Triton-X-100 sample buffer, and proteins were separated by SDS-PAGE. Proteins were blotted on a nitrocellulose membrane (Thermo Scientific). Detection occurred by using specific antibodies against β-actin (1:2500 dilution, Sigma-Aldrich, Munich, Germany), ACE2 and TMPRSS2 (both 1:1000 dilution, abcam, Cambridge, UK) followed by incubation with IRDye-labeled secondary antibodies (LI-COR Biotechnology, IRDye®800CW Goat anti-Rabbit, 926-32211, 1:40 000) according to the manufacturer's instructions. Protein bands were visualized by laser-induced fluorescence using infrared scanner for protein quantification (Odyssey, Li-Cor Biosciences, Lincoln, NE, USA).

5.2.5 Receptor blocking experiments

SARS-CoV/SARS-CoV-2 receptor blocking experiments were adapted from Cinatl et al (2004) (Cinatl Jr et al., 2004). Caco2 cells were pre-treated for 30 min at 37°C with goat antibodies directed against the human ACE2 or DDP4 ectodomain (R&D Systems, Wiesbaden-Nordenstadt, Germany). Then, cells were washed three times with PBS and infected with SARS-CoV-2 at MOI 0.01. Cytopathogenic effects were monitored 48 h post-infection.

Cytopathogenic effect (CPE) was assessed visually by light microscopy by two independent laboratory technicians 48 h after infection (Cinatl et al., 2003).

5.2.6 Antiviral assay

Confluent cell cultures were infected with SARS-CoV-2 or SARS-CoV in 96-well plates at MOI 0.01 in the absence or presence of drug. Cytopathogenic effect (CPE) was assessed visually by light microscopy by two independent investigators 48 h post-infection (Cinatl et al., 2003).

5.2.7 Viability assay

Cell viability was determined by 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay modified after Mosmann (Mosmann, 1983), as previously described (Onafuye et al., 2019).

5.2.8 Qpcr

SARS-CoV-2 and SARS-CoV RNA was isolated from cell culture supernatants using AVL buffer and the QIAamp Viral RNA Kit (Qiagen) according to the manufacturer's instructions. RNA was subjected to OneStepqRT-PCR analysis using the SYBR green based Luna Universal One-Step RT-qPCR Kit (New England Biolabs) and a CFX96 Real-Time System, C1000 Touch Thermal Cycler. Primers were adapted from the WHO protocol (Corman et al., 2020) targeting the open reading frame for RNA-dependent RNA polymerase (RdRp) of both SARS-CoV-2 and SARS-CoV: RdRP_SARSr-F2 (GTGARATGGTCATGTGTGGCGG) and RdRP_SARSr-R1 (CARATGTTAAASACACTATTAGCATA) using 0.4 µM per reaction. RNA copies/ml were determined by standard curves which were using plasmid DNA (pEX-A128-RdRP) harbouring the corresponding amplicon regions for SARS-CoV-2 RdRP target sequence (GenBank Accession numberNC_045512). For each condition, three biological replicates were used. Mean and standard deviation were calculated for each group.

5.3. Results

5.3.1 Determination of differentially conserved positions (DCPs)

Coronavirus genomes harbour single-stranded positive sense RNA (+ssRNA) of about 30 kilobases in length, which contain six or more open reading frames (ORFs) (Cui et al., 2019; Wu, A. et al., 2020). The SARS-CoV-2 genome has a size of approximately 29.8 kilobases and was annotated to encode 14 ORFs and 27 proteins (Wu, A. et al., 2020). Two ORFs at the 5'-terminus (ORF1a, ORF1ab) encode the polyproteins pp1a and pp1b, which comprise 15 non-

structural proteins (nsps), the nsps 1 to 10 and 12–16 (Wu, A. et al., 2020). Additionally, SARS-CoV-2 encodes four structural proteins (S, E, M, N) and eight accessory proteins (3a, 3b, p6, 7a, 7b, 8b, 9b, orf14) (Wu, A. et al., 2020). This set-up resembles that of SARS-CoV. The 8a protein in SARS-CoV is absent in SARS-CoV-2. 8 b is longer in SARS-CoV-2 (121 amino acids) than in SARS-CoV (84 amino acids), while 3 b is shorter in SARS-CoV-2 (22 amino acids) than in SARS-CoV (154 amino acids) (Wu, A. et al., 2020).

To identify genomic differences between SARS-CoV-2 and SARS-CoV that may affect the structure and function of the encoded virus proteins, we identified differentially conserved amino acid positions (DCPs) (Rausell et al., 2010) and determined their potential impact by in silico modelling (Martell et al., 2019; Pappalardo et al., 2016).

In the reference sequences of the 22 SARS-CoV-2 virus proteins that could be compared with SARS-CoV, 1393 positions encoded different amino acids. 891 (64%, 9% of all SARS-CoV-2 genome residues) of these positions were DCPs (Table S5.2). Most of the amino acid substitutions at DCPs appear to be fairly conservative as demonstrated by the average BLOSUM substitution score of 0.32 (median 0; Figure S5.1) and with 69% of them having a score of 0 or greater (the higher the score the more frequently such amino acid substitutions are observed naturally in evolution). 46% of DCPs represent conservative changes where amino acid properties are retained (e.g. change between two hydrophobic amino acids), 18% represented polar—hydrophobic substitutions, and <10% were changes between charged amino acids (Table S5.3).

Six of the SARS-CoV-2 proteins have a higher proportion of DCPs, S, 3a, p6, nsp2, nsp3 (papain-like protease), and nsp4 with 14.82%, 11.68%, 9.52%, 21.38%, 17.9% and 10.8% of their residues being DCPs, respectively (Table S5.4). Very few DCPs were observed in the envelope (E) protein and most of remaining non-structural proteins encoded by ORF1ab. For example, no residues in the helicase and <4% of residues in the RNA-directed RNA polymerase, 2'-O-Methyltransferase, nsp8 and nsp9 are DCPs (Table S5.1).

We were able to map 572 DCPs onto protein structures (Figure S5.2, Table S5.5 and S4.6). Nearly all of the mapped DCPs occur on the protein surface (86%), with only 34 DCPs buried within the protein, primarily in S and the papain-like protease (nsp3) (Table S5.3). We propose that 49 DCPs are likely to result in structural/functional differences between SARS-CoV and SARS-CoV-2 proteins. A further 259 could result in some change. The remaining 264 DCPs seem unlikely to have a substantial functional impact (Table S5.3).

5.3.2 Differentially conserved positions (DCPs) in interferon antagonists

At least 10 SARS-CoV proteins have roles in interferon antagonism (Totura & Baric, 2012). Two of these proteins, p6 and the papain-like protease (nsp3), contain many DCPs, two have very few DCPs (nsp7 and nsp16), five have intermediate numbers of DCPs (nsp14, nsp1, nsp15, N and M), while p3b is not encoded by SARS-CoV-2. Initial studies have identified a difference in the interferon inhibition between SARS-CoV and SARS-CoV-2 (Lokugamage et al., 2020). Thus, it is possible that especially the DCPs in p6 and the papain-like protease may have an effect on interferon inhibition.

5.3.3 Differences in cell tropism between SARS-CoV-2 and SARS

Next, we elucidated whether the substantial number of DCPs results in different phenotypes in cell culture, using the cell lines Caco2, CL14 (susceptible to SARS-CoV infection), HT-29 and DLD-1 (non-susceptible) (Cinatl Jr et al., 2004). Analogously to SARS-CoV infection, SARS-CoV-2 replication was detected in Caco2 and CL14 cells, but not in HT-29 or DLD-1 cells, as shown by cytopathogenic effects (CPE) (Figure 5.1A), staining for double-stranded RNA (Figure S5.3A) and viral genomic RNA levels (Figure S5.3B).

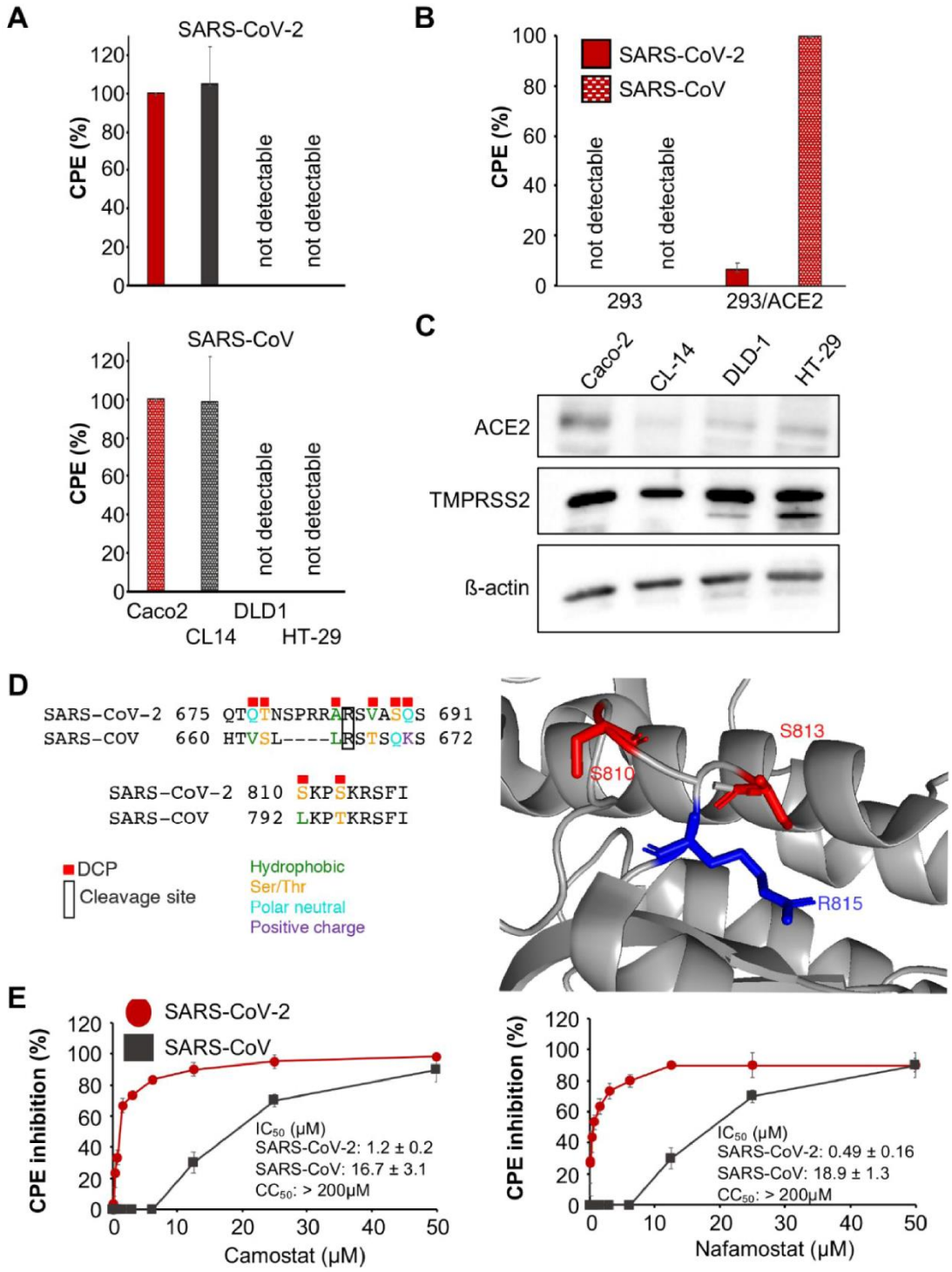


Figure 5.1. SARS-CoV-2 and SARS-CoV replication in cell culture. (**A**) Cytopathogenic effect (CPE) formation 48 h post-infection in MOI 0.01-infected Caco2, CL14, DLD-1 and HT29 cells. Representative images showing immunostaining for double-stranded RNA (indicates virus replication) and quantification of virus genomes by qPCR are presented in Figure S5.3. (**B**) CPE formation in SARS-CoV and SARS-CoV-2 (MOI 0.01)-infected ACE2-negative 293 cells and 293 cells stably expressing ACE2 cells (293/ACE2) 48 h post-infection. Immunostaining for double-stranded RNA and quantification of virus

genomes by qPCR is shown in Figure S5.4. (**C**) Western blots indicating cellular ACE2 and TMPRSS2 protein levels in uninfected cells. Uncropped blots are provided in Figure S5.5. (**D**) A sequence view of the DCPs in the vicinity of the S two cleavage sites and an image of the R815 cleavage site and closely located DCPs. S is cleaved and activated by TMPRSS2. (**E**) Concentration-dependent effects of the TMPRSS2 inhibitors camostat and nafamostat on SARS-CoV-2- and SARS-CoV-induced cytopathogenic effect (CPE) formation determined 48 h post-infection in Caco2 infected at an MOI of 0.01 using a phase contrast microscope. Similar effects were observed in CL14 cells (Figure S5.6). Values are presented as means ± S.D. (n = 3)

However, ACE2-expressing 293 cells differed in their susceptibility to SARS-CoV-2 and SARS-CoV (Figure 5.1B, Figure S5.4). ACE2 has been identified as a cellular receptor for both SARS-CoV-2 and SARS-CoV (Cui et al., 2019; Hoffmann et al., 2020; Walls et al., 2020; Wan et al., 2020; Wrapp et al., 2020; Wu, A. et al., 2020; Yan, R. et al., 2020). Unmodified 293 cells are not susceptible to SARS-CoV infection due to a lack of ACE2 expression. However, 293 cells that stably express ACE2 (293/ACE2) support SARS-CoV infection (Kamitani et al., 2006). As expected, infection of 293 cells with SARS-CoV or SARS-CoV-2 did not result in detectable cytopathogenic effect (CPE) (Figure 5.1B), but a SARS-CoV-induced CPE was detected in 293/ACE2 cells (Figure 5.1B). In contrast, 293/ACE2 cells displayed limited permissiveness to SARS-CoV-2 infection (Figure 5.1B). Staining for double-stranded RNA (Figure S5.4A) and detection of viral genomic RNA copies (Figure S5.4B) confirmed these findings. Hence, the ACE2 status does not reliably predict cell sensitivity to SARS-CoV-2. Indeed, CL-14 was characterized by lower ACE2 levels than DLD-1 and HT29 (Figure 5.1C).

SARS-CoV-2 and SARS-CoV cell entry depends on S cleavage by transmembrane serine protease 2 (TMPRSS2) (Hoffmann et al., 2020; Hoffmann, Schroeder, Kleine-Weber, Müller, Drosten, & Pöhlmann, 2020a; Zhou, Y. et al., 2015). However, the non-SARS-CoV-2 susceptible and susceptible cell lines displayed similar TMPRSS2 levels (Figure 5.1C). Thus, cellular TMPRSS2 levels do also not reliable predict cell susceptibility to SARS-CoV-2.

5.3.4 Differences between SARS-CoV-2 and SARS-CoV S (Spike) protein cleavage sites and sensitivity to protease inhibitors

R667 and R797 are the critical cleavage sites in SARS-CoV S that are recognized by TMPRSS2 (Simmons et al., 2013; Zhou, Y. et al., 2015). These cleavage sites are conserved in SARS-CoV-2 (R685 and R815) (Figure 5.1D). However, there is a four amino acid insertion in SARS-CoV-2 S prior to R685 and many of the residues close to R685 are DCPs (V663 = Q677, S664 = T678,

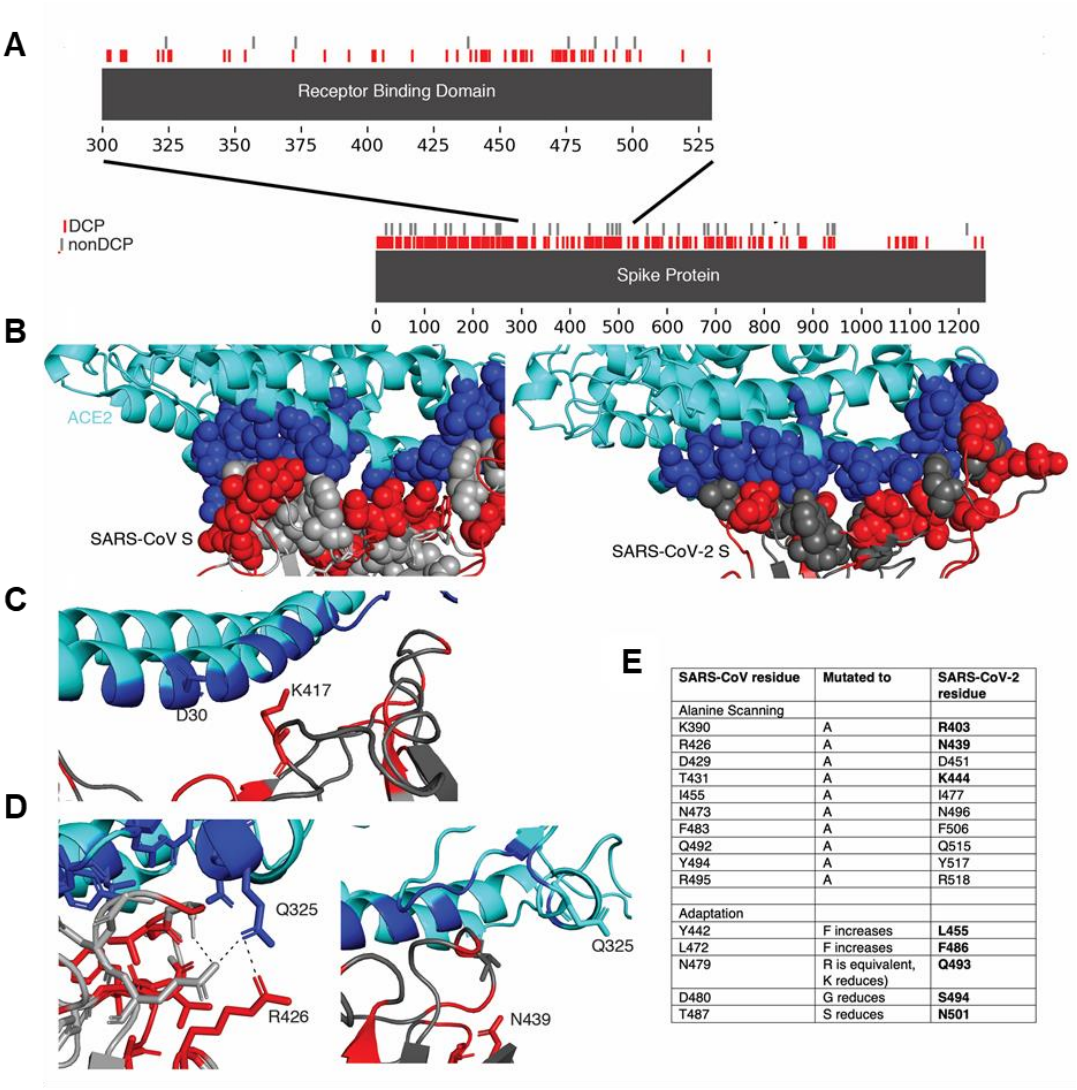
T669 = V687, Q671 = S689, K672 = Q690 DCPs are represented by the SARS-CoV residue followed by the SARS-CoV-2 residue) (Figure 5.1D). The R815 cleavage site has two DCPs in close proximity (L792 = S810, T795 = S813) (Figure 5.1D). Around the R685 cleavage site two DCPs retain polar side chains (S664 = T678, Q671 = S689), while the others represent larger changes between hydrophobic and polar side chains (V663 = Q677, T669 = V687) and one changes from a positive charge to a polar side chain (K672 = Q690). While around the R815 cleavage site, one substitution is conservative (T795 = S813) and the other is a hydrophobic to polar change (L792 = S810).

These changes are likely to impact on TMPRSS2-mediated S cleavage. Indeed, SARS-CoV-2 was more sensitive than SARS-CoV to inhibition by the serine protease inhibitors camostat and nafamostat (Figure 5.1E, Figure S5.6), which are known to inhibit TMPRSS2-mediated S cleavage and virus entry (Hoffmann et al., 2020; Hoffmann et al., 2020a; Zhou, Y. et al., 2015). This confirms that the observed differences in the amino acid sequence of S have functional consequences.

5.3.5 Differences between SARS-CoV-2 and SARS-CoV S interaction with ACE2

Our computational analysis detected further interesting changes in the S protein. SARS-CoV-2 S is 77.46% sequence identical to the SARS-CoV S and many of the remaining positions are DCPs (186 residues) (Table S5.1).

The SARS-CoV S receptor binding domain (residues 306-527, equivalent to 328-550 in SARS-CoV-2) is enriched in DCPs, containing 43 DCPs (19% of residues). Nine of the 24 SARS-CoV S residues in direct contact with ACE2 were DCPs (Figure 5.2A, Table S5.4). Five of these DCPs represent conservative substitutions in amino acid (hydrophobic—hydrophobic or polar-polar), two hydrophobic -polar substitutions, one positive charge to polar change, while the ninth is substitution between a hydrophobic and positively charged amino acid (Table S5.5).



SARS-CoV residue	Mutated to	SARS-CoV-2 residue
Alanine Scanning		
K390	A	**R403**
R426	A	**N439**
D429	A	D451
T431	A	**K444**
I455	A	I477
N473	A	N496
F483	A	F506
Q492	A	Q515
Y494	A	Y517
R495	A	R518
Adaptation		
Y442	F increases	**L455**
L472	F increases	**F486**
N479	R is equivalent, K reduces)	**Q493**
D480	G reduces	**S494**
T487	S reduces	**N501**

Figure 5.2. SARS-CoV-2 and SARS-CoV S interaction with ACE2. (**A–D**)Differentially conserved positions in the Spike protein. (**A**) A sequence view of the DCPs present in the Spike protein, with an inset showing the receptor binding domain. (**B**) The S interface with ACE2 (cyan). The ACE2 interface is shown in blue spheres, DCPs in red. (**C**) The V404 = K417 DCP. (**D**) The R426 = N439 DCP, the left image

shows SARS-CoV S R426, the image on the right show the equivalent N439 in SARS-CoV-2 S. (**E**) SARS-CoV residues associated with altering ACE2 affinity and the residues at these positions in SARS-CoV-2 S. (**F**) Cytopathogenic effect (CPE) formation in SARS-CoV-2 and SARS-CoV (MOI 0.01)-infected Caco2 cells in the presence of antibodies directed against ACE2 or DPP4 (MERS-CoV receptor) 48 h post-infection

Analysis of the DCPs using the SARS-CoV and SARS-CoV-2 S protein complexes with ACE2 (Song et al., 2018; Yan, R. et al., 2020) identified runs of DCPs (A430-T433, F460-A471) in surface loops forming part of the S-ACE2 interface and resulted in different conformations in SARS-CoV-2 S compared to SARS-CoV S (Figure 5.2A, 4.2B). Two DCPs remove intramolecular hydrogen bonding within the spike protein in SARS-CoV-2 (Table S5.4) and three DCPs (R426 = N439, N479 = QQ493, Y484 = Q498) are residues that form hydrogen bonds with ACE2. For two of these positions, hydrogen bonding with ACE2 is present with both S proteins, but for R426 = N439 hydrogen bonding with ACE2 is only observed with SARS-CoV S. N439 in SARS-CoV-2 S is not present in the interface and the sidechain points away from the interface. Further, analysis of the SARS-CoV-2 S-ACE2 complex highlighted important roles of the V404 = K417 DCP, where K417 in SARS-CoV-2 S is able to form a salt bridge with ACE2 D30 (Figure 5.2C, 4.2D) (Yan, R. et al., 2020).

Alanine scanning (Chakraborti et al., 2005) and adaptation experiments (Wan et al., 2020) have identified 16 SARS-CoV S residues impacting on the binding affinity with ACE2. For all five residues identified from adaptation studies and four of the 11 identified by alanine scanning experiments, different amino acids are present in SARS-CoV-2 S (Figure 5.2E), highlighting the difference in the interaction with ACE2.

In agreement with our structural analysis, we detected differences in the effects of an anti-ACE2 antibody on SARS-CoV-2 and SARS-CoV infection. Antibodies directed against ACE2 were previously shown to inhibit SARS-CoV replication (Li, W. et al., 2003). In line with this, an anti-ACE2 antibody inhibited SARS-CoV infection in Caco2 cells (Figure 5.2F). In contrast, the anti-ACE2 antibody displayed limited activity against SARS-CoV-2 infection (Figure 5.2F). This shows that it is more difficult to antagonize SARS-CoV-2 infection with anti-ACE2 antibodies and supports previous findings indicating a stronger binding affinity of SARS-CoV-2 S to ACE2 compared to SARS-CoV S (Walls et al., 2020; Wrapp et al., 2020). As anticipated, antibodies directed against DPP4, the MERS-CoV receptor (Cui et al., 2019; de Wit et al., 2016), did not interfere with SARS-CoV or SARS-CoV-2 infection (Figure 5.2F).

5.4. Discussion

Here, we performed an in-silico analysis of the effects of differentially conserved amino acid positions (DCPs) between SARS-CoV-2 and SARS-CoV proteins on virus protein structure and function in combination with a comparison of wild-type SARS-CoV-2 and SARS-CoV in cell culture.

We identified 891 DCPs, which represents 64% of the amino acid positions that differ between SARS-CoV-2 and SARS-CoV and nearly 9% of all residues encoded by the SARS-CoV genome. 49 of these DCPs are likely to have a structural and functional impact. The DCPs are not equally distributed between the proteins. DCPs are enriched in S, 3a, p6, nsp2, papain-like protease and nsp4, but very few DCPs are present in the envelope (E) protein and most of the remaining non-structural proteins encoded by ORF1ab. This indicates that the individual proteins differ in their tolerance to sequence changes and/or their exposure to selection pressure exerted by the host environment.

The large proportion of DCPs reflects the differences in the clinical behaviour of SARS-CoV-2 and SARS-CoV. Mortality associated with SARS-CoV is higher than that associated with SARS-CoV-2 (Borges do Nascimento et al., 2020; Cui et al., 2019). SARS-CoV causes a disease of the lower respiratory tract. Infected individuals are only contagious when they experience symptoms (de Wit et al., 2016). SARS-CoV-2 is present in the upper respiratory tract and can be readily transmitted prior to the onset of symptoms. Mild but infectious cases may substantially contribute to its spread (Rivett et al., 2020).

Although further research will be required to elucidate in detail, which DCPs are responsible for which differences in virus behaviour, our analysis has already provided important clues. Both viruses use ACE2 as a receptor and are activated by the transmembrane serine protease TMPRSS2 (Cui et al., 2019; Hoffmann et al., 2020; Li, W. et al., 2003; Walls et al., 2020; Wan et al., 2020; Wrapp et al., 2020; Yan, R. et al., 2020). Our results show, however, that the ACE2 and the TMPRSS2 status are not sufficient to predict cells susceptibility to SARS-CoV-2 or SARS-CoV. The cell line CL14 supported SARS-CoV-2 replication, although it displayed lower ACE2 levels and similar TMPRSS2 levels to non-susceptible DLD-1 and HT29 cells. Thus, attempts to identify SARS-CoV-2 target cells based on the ACE2 status (Luan et al., 2020; Qiu et al., 2020; Xu, H. et al., 2020) need to be considered with caution.

As previously described (Kamitani et al., 2006), ACE2 expression rendered SARS-CoV non-permissive 293 cells susceptible to SARS-CoV. However, ACE2 expression had a substantially lower impact on SARS-CoV-2 infection. This suggests the presence of further host cell factors that determine SARS-CoV-2 susceptibility. Based on our sequence analysis, DCPs in the viral interferon antagonists may contribute to the differences observed in the cellular tropism of SARS-CoV-2 and SARS-CoV.

Our computational analysis detected DCPs in the ACE2-binding domain of S, which are likely to impact S-ACE2 binding. In agreement, an anti-ACE2 antibody displayed higher efficacy against SARS-CoV than against SARS-CoV-2, illustrating the differences between SARS-CoV-2 S and SARS-CoV S interaction with ACE2. This probably reflects an increased SARS-CoV-2 S affinity to ACE2 compared to SARS-CoV S (Wrapp et al., 2020), which may be more difficult to antagonize.

To mediate virus entry, S needs to be cleaved by host cell proteases, in particular by TMPRSS2 (Hoffmann et al., 2020; Hoffmann et al., 2020a; Zhou, Y. et al., 2015). The S cleavage sites are conserved between SARS-CoV-2 and SARS-CoV. However, we found DCPs in close vicinity to the S cleavage sites, which are likely to affect S cleavage by host cell enzymes and/or the activity of protease inhibitors on S cleavage. Indeed, the serine protease inhibitors camostat and nafamostat, which interfere with S cleavage (Hoffmann et al., 2020; Hoffmann et al., 2020a), displayed increased activity against SARS-CoV-2 infection than against SARS-CoV infection, confirming the functional relevance of the DCPs.

In conclusion, our in-silico study revealed a substantial number of differentially conserved amino acid positions in the SARS-CoV-2 and SARS-CoV proteins. In agreement, cell culture experiments indicated differences in the cell tropism of these two viruses and showed that cellular ACE2 and TMPRSS2 levels do not reliably indicate cell susceptibility to SARS-CoV-2. Moreover, we identified DCPs in S that are associated with differences in the interaction with ACE2 and increased SARS-CoV-2 sensitivity to the protease inhibitors camostat and nafamostat relative to SARS-CoV.

Chapter 6: A potential role of the CD47-SIRPalpha axis in COVID-19 pathogenesis

Katie-May McLaughlin[1,#], Denisa Bojkova[2,#], Joshua D. Kandler[2,#], Marco Bechtel[2], Philipp Reus[2], Trang Le[2], Florian Rothweiler[2], Julian U. G. Wagner[3], Andreas Weigert, Sandra Ciesek[2,5,6], Mark N. Wass[2], Martin Michaelis[2], Jindrich Cinatl Jr[2]

[1]School of Biosciences, University of Kent, Canterbury, UK; km625@kent.ac.uk (K.M.); M.N.Wass@kent.ac.uk (M.N.W); M.Michaelis@kent.ac.uk (M.M.)

[2]Institute for Medical Virology, University Hospital, Goethe University Frankfurt am Main, Germany; Marco.Bechtel@kgu.de (M.B.); Denisa.Bojkova@kgu.de (D.B.); Joshua.Kandler@kgu.de (J.D.K.); philipp.reus@kgu.de (P.R.); letrang1211@gmail.com (T.L.); Sandra.ciesek@kgu.de (S.C.); Cinatl@em.uni-frankfurt.de (J.C.jr.)

[3]Institute for Cardiovascular Regeneration, Goethe University, Theodor Stern Kai 7, Frankfurt, Germany; German Center for Cardiovascular Research (DZHK), Frankfurt, Germany; Faculty for Biological Sciences, Goethe University, Frankfurt, Germany; j.wagner@med.uni-frankfurt.de (J.U.G.W.)

[4]Faculty of Medicine, Institute of Biochemistry I, Goethe-University, Frankfurt am Main, Germany; weigert@biochem.uni-frankfurt.de (A.W.)

[5]German Center for Infection Research, DZIF, External partner site Frankfurt am Main, Germany; Sandra.ciesek@kgu.de (S.C.)

[6]Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch Translational Medicine und Pharmacology, Frankfurt, Germany; Sandra.ciesek@kgu.de (S.C.)

[#]Equal contribution

Submitted to MDPI Pathogens (Viral Pathogens) for peer review

Abstract

The coronavirus SARS-CoV-2 is the cause of the ongoing COVID-19 pandemic. Most SARS-CoV-2 infections are mild or even asymptomatic. However, a small fraction of infected individuals develops severe, life-threatening disease, which is caused by an uncontrolled immune response resulting in hyperinflammation. However, the factors predisposing individuals to severe disease remain poorly understood. Here, we show in a range of model systems and data from post mortem samples that SARS-CoV-2 infection results in increased

levels of CD47, which is known to mediate immune escape in cancer and virus-infected cells. Systematic literature searches further indicated that known risk factors such as older age and diabetes are associated with increased CD47 levels. High CD47 levels contribute to vascular disease, vasoconstriction, and hypertension, conditions which may predispose SARS-CoV-2-infected individuals to COVID-19-related complications such as pulmonary hypertension, lung fibrosis, myocardial injury, stroke, and acute kidney injury. Hence, age-related and virus-induced CD47 expression is a candidate mechanism potentially contributing to severe COVID-19 and also a therapeutic target, which may be addressed by antibodies and small molecules.

6.1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is causing the ongoing coronavirus disease 2019 (COVID-19) outbreak (Chilamakuri & Agarwal, 2021; Hokello et al., 2020), which has resulted in more than 120 million confirmed cases and more than 2.6 million confirmed COVID-19-associated deaths so far (Dong et al., 2020).

The first COVID-19 vaccines have been developed (Chilamakuri & Agarwal, 2021), and their roll-out has started in many countries. However, it will take a significant time until large parts of the world population will be vaccinated, and there is growing concern about the emergence of escape variants that can bypass immunity conferred by the current vaccines and previous SARS-CoV-2 infections (Andreano et al., 2020; Kemp et al., 2020; Liu, Z. et al., 2020; Sabino et al., 2021; Weisblum et al., 2020; Wibmer et al., 2021). Thus, for the foreseeable future there will be a need for improved COVID-19 therapies.

Currently, the therapeutic options for COVID-19 are still very limited (Chilamakuri & Agarwal, 2021; Rebold et al., 2021). COVID-19 therapies can either directly inhibit SARS-CoV-2 replication or target other COVID-19-associated pathophysiological processes, such as corticosteroids that are anticipated to control COVID-19-related cytokine storm and hyperinflammation (Pum et al., 2021). Dexamethasone and potentially other corticosteroids increase survival in patients who depend on oxygen support (RECOVERY Collaborative Group et al., 2021; WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group et al., 2020). In a controlled open-label trial, dexamethasone reduced mortality in patients receiving oxygen with (from 41.1% to 29.3%) or without (from 26.2% to 23.3%) mechanical ventilation, but increased mortality in patients not requiring oxygen support (RECOVERY

Collaborative Group et al., 2021). Other immunomodulatory therapy candidates are being tested, but conclusive results are pending (Rebold et al., 2021). Further COVID-19 therapeutics under investigations include anticoagulants that target COVID-19-induced systemic coagulation and thrombosis (coagulopathy) (Hadid et al., 2020).

However, it would be much better to have effective antiviral treatments that reliably prevent COVID-19 disease progression to a stage when immunomodulators and anticoagulants are needed. The antiviral drug remdesivir was initially described to reduce recovery time from 15 to ten days and 29-day mortality from 15.2% to 11.4% (Beigel et al., 2020; Rebold et al., 2021). However, other trials did not confirm this and conclusive evidence on the efficacy of remdesivir remains to be established (Rebold et al., 2021). The JAK inhibitor baricitinib, which interferes with cytokine signaling, was reported to improve therapy outcomes in combination with remdesivir in a double-blind, randomised, placebo-controlled trial, in which patients were either treated with remdesivir plus baricitinib or remdesivir plus placebo (Kalil et al., 2021). Moreover, convalescent sera and monoclonal antibodies are under clinical investigation for COVID-19 treatment (Devarasetti et al., 2021; Tuccori et al., 2020).

Ideally, antiviral therapies are used early in the disease course to prevent disease progression to the later immunopathology-driven stages (Weinreich et al., 2021). However, only a small proportion of patients develops severe disease (Salzberger et al., 2020). Therefore, a better understanding of the underlying processes is required to identify patients, who will develop severe disease, as early as possible.

Here, we investigated the potential role of the ubiquitously expressed cell surface glycoprotein CD47 in severe COVID-19. CD47 is the receptor of thrombospondin-1 (THBS1) and the counter-receptor for signal regulatory protein-α (SIRPα). CD47 interaction with SIRPα inhibits the activation of macrophages and dendritic cells and thrombospondin-1/ CD47 signaling inhibits T cell activation (Cham et al., 2020; Kaur et al., 2020). High CD47 expression prevents immune recognition of cancer and virus-infected cells (Cham et al., 2020; Kaur et al., 2020).

6.2. Methods

6.2.1 Cell culture

Calu-3 cells (ATCC) were grown at 37°C in minimal essential medium (MEM) supplemented with 10% fetal bovine serum (FBS), 100 IU/mL penicillin, and 100 µg/mL of streptomycin. All

culture reagents were purchased from Sigma. Cells were regularly authenticated by short tandem repeat (STR) analysis and tested for mycoplasma contamination.

Primary human bronchial epithelial cells were purchased from ScienceCell. For differentiation to air-liquid interface (ALI) cultures, the cells were thawed and passaged once in PneumaCult-Ex Medium (StemCell technologies) and then seeded on transwell inserts (12 well plate, Sarstedt) at $4x10^4$ cells/insert. Once cell layers reached confluency, medium on the apical side of the transwell was removed, and medium in the basal chamber was replaced with PneumaCult ALI Maintenance Medium (StemCell Technologies) including Antibiotic/Antimycotic solution (Sigma Aldrich) and MycoZap Plus PR (Lonza). During a period of four weeks, medium was changed and cell layers were washed with PBS every other day. Criteria for successful differentiation were the development of ciliated cells and ciliary movement, an increase in transepithelial electric resistance indicative of the formation of tight junctions, and mucus production.

Human monocytes were isolated from buffy coats of healthy donors (RK-Blutspendedienst Baden-Württemberg-Hessen, Institut für Transfusionsmedizin und Immunhämatologie Frankfurt am Main, Germany). After centrifugation on Ficoll (Pancoll, PAN-Biotech) density gradient, mononuclear cells were collected from the interface, washed with PBS, and plated on cell culture dishes (Cell+, Saarstedt) in RPMI1640 (Gibco) supplemented with 100 IU/mL penicillin and 100 µg/mL of streptomycin. After incubation for 90 minutes (37°C, 5% CO2), non-adherent cells were removed, and the medium was changed to RPMI1640 supplemented with 100 IU/mL penicillin, 100 µg/mL of streptomycin, and 3% human serum (RK-Blutspendedienst Baden-Württemberg-Hessen, Institut für Transfusionsmedizin und Immunhämatologie Frankfurt am Main, Germany).

6.2.2 Virus infection

SARS-CoV-2/7/Human/2020/Frankfurt (SARS-CoV-2/FFM7) was isolated and cultivated in Caco2 cells (DSMZ) as previously described (Hoehl et al., 2020; Toptan et al., 2020). Virus titers were determined as TCID50/ml in confluent cells in 96-well microtiter plates (Cinatl Jr et al., 2005; Cinatl et al., 2003).

Monocytes were infected at an MOI of 1 with SARS-CoV-2/FFM7 for 2 hours. After infection, cells were washed three times with PBS and subsequently cultivated in RPMI1640 (Gibco) supplemented with 100 IU/mL penicillin and 100 µg/mL of streptomycin.

6.2.3 Western blot

Cells were lysed using Triton-X-100 sample buffer, and proteins were separated by SDS-PAGE. Detection occurred by using specific antibodies against CD47 (1:100 dilution, CD47 Antibody, anti-human, Biotin, REAfinity™, # 130-101-343, Miltenyi Biotec), SARS-CoV-2 N (1:1000 dilution, SARS-CoV-2 Nucleocapsid Antibody, Rabbit MAb, #40143-R019, Sino Biological), SIRPα (1:1000 dilution, SIRPα/SHPS1 (D6I3M) Rabbit mAb #13379, Cell Signaling), and GAPDH (1:1000 dilution, Anti-G3PDH Human Polyclonal Antibody, #2275-PC-100, Trevigen). Protein bands were visualized and quantified by laser-induced fluorescence using infrared scanner for protein quantification (Odyssey, Li-Cor Biosciences).

6.2.4. qPCR

SARS-CoV-2 RNA from cell culture supernatant samples was isolated using AVL buffer and the QIAamp Viral RNA Kit (Qiagen) according to the manufacturer's instructions. SARS-CoV-2 RNA from cell lysates was isolated using RTL Buffer and the RNeasy Mini Kit (Qiagen) according to the manufacturer´s instructions. Absorbance-based quantification of the RNA yield was performed using the Genesys 10S UV-Vis Spectrophotometer (Thermo Scientific). RNA was subjected to OneStep qRT-PCR analysis using the Luna Universal One-Step RT-qPCR Kit (New England Biolabs) and a CFX96 Real-Time System, C1000 Touch Thermal Cycler. Primers were adapted from the WHO protocol29 targeting the open reading frame for RNA-dependent RNA polymerase (RdRp): RdRP_SARSr-F2 (GTG ARA TGG TCA TGT GTG GCG G) and RdRP_SARSr-R1 (CAR ATG TTA AAS ACA CTA TTA GCA TA) using 0.4 µM per reaction. Standard curves were created using plasmid DNA (pEX-A128-RdRP) harboring the corresponding amplicon regions for RdRP target sequence according to GenBank Accession number NC_045512. For each condition three biological replicates were used. Mean and standard deviation were calculated for each group.

6.2.5 Data acquisition and analysis

Normalized protein abundance data from SARS-CoV-2-infected Caco-2 cells were derived from a recent publication (Bojkova et al., 2020). Data were subsequently normalized using summed intensity normalization for sample loading, followed by internal reference scaling and Trimmed mean of M normalization. Mean protein abundance was plotted using the function *ggdotplot* of the R package ggpubr. P-values were determined by two-sided student's t-test

Raw read counts from post-mortem samples of two COVID-19 patients and two healthy controls, as well as mock infected and SARS-CoV-2-infected Calu-3 cells, were derived from

a recent publication (Blanco-Melo et al., 2020) via the gene expression omnibus (GEO) database (accession: GSE147507) and processed using DESeq2. Normalized gene counts were plotted using the function *ggdotplot* of the R package ggpubr. P-values were determined by two-sided student's t-test.

6.2.6 Literature review

Relevant articles were identified by using the search terms 'CD47 aging', 'CD47 hypertension', 'CD47 diabetes', and 'CD47 obesity' in PubMed (https://pubmed.ncbi.nlm.nih.gov) on the basis of the principles outlined in the PRISMA guidelines (http://prisma-statement.org). Articles in English were included into the analysis, when they contained original data on the influence of aging, diabetes, diabetes, or obesity on CD47 expression levels and/ or the relevance of CD47 with regard to pathological conditions observed in severe COVID-19. Two reviewers independently analyzed articles for relevant information and then agreed a list of relevant articles.

6.3. Results

6.3.1 SARS-CoV-2 infection results in enhanced CD47 expression

A publicly available proteomics dataset (Bojkova et al., 2020) indicated increased CD47 expression in SARS-CoV-2-infected Caco2 colorectal carcinoma cells (Figure 6.1A). We also detected enhanced CD47 levels in SARS-CoV-2-infected primary human bronchial epithelial cells (HBE) grown in air liquid interface (ALI) cultures (Bojkova, Bechtel et al., 2020) and Calu-3 lung cancer cells (Figure 6.1B). Analysis of transcriptomics data from another study also indicated increased CD47 levels in SARS-CoV-2-infected Calu-3 cells (Figure S6.2) and in post mortem lung samples from COVID-19 patients (Figure 6.1C) (Blanco-Melo et al., 2020). Flow cytometry analysis confirmed increased CD47 levels in SARS-CoV-2-infected Caco2 cells (Figure S6.3).

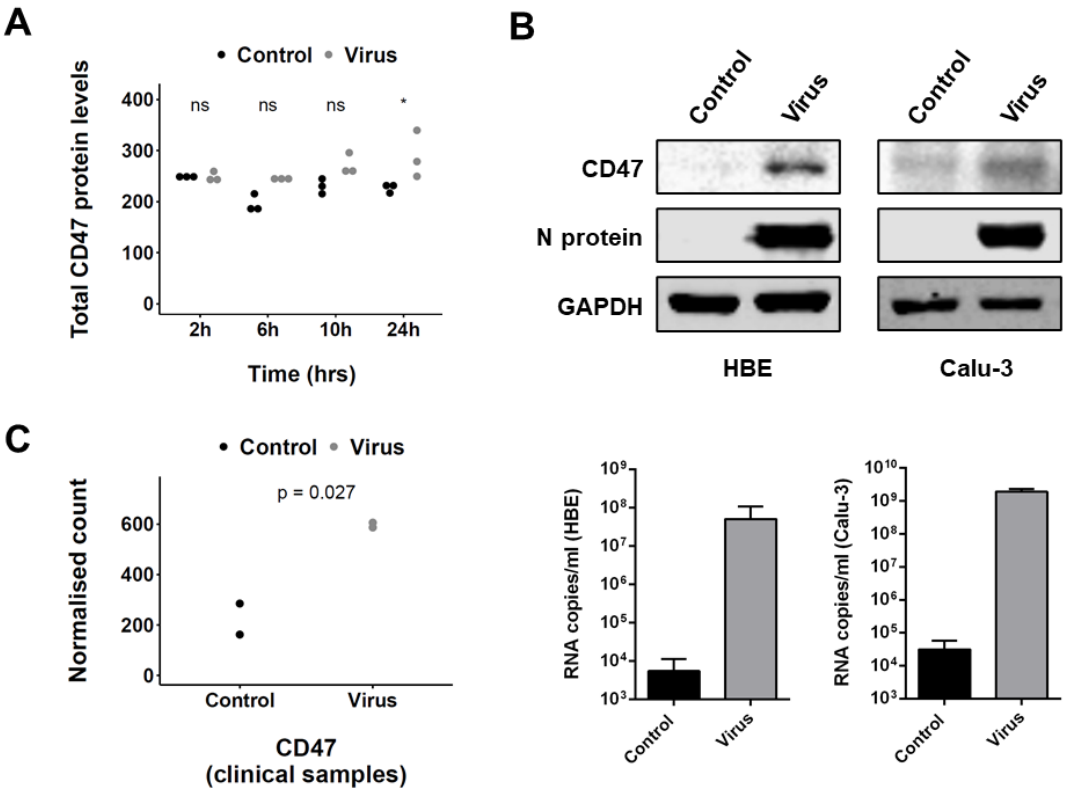


Figure 6.1. SARS-CoV-2 infection is associated with increased CD47 levels. (**A**) CD47 protein abundance in uninfected (control) and SARS-CoV-2-infected (virus) Caco-2 cells (data derived from (Bojkova et al., 2020). P-values were determined by one-way ANOVA with post-hoc Tukey HSD test. (**B**) CD47 and SARS-CoV-2 N protein levels and virus titers (genomic RNA determined by PCR) in SARS-CoV-2 strain FFM7 (MOI 1)-infected air-liquid interface cultures of primary human bronchial epithelial (HBE) cells and SARS-CoV-2 strain FFM7 (MOI 0.1)-infected Calu-3 cells. Uncropped blots are provided in Figure S6.1. (**C**) CD47 mRNA levels in post mortem samples from COVID-19 patients (data derived from (Blanco-Melo et al., 2020)). P-values were determined by two-sided Student's t-test.

6.3.2 Increased SIRPα levels in SARS-CoV-2-infected monocytes

CD47 inhibits the activity of innate immune cells via interaction with SIRPα (Cham et al., 2020; Kaur et al., 2020). Hence, we next investigated whether SARS-CoV-2 infection of monocytes may impact SIRPα levels. SARS-CoV-2 did not result in a productive infection of primary human monocytes as indicated by a lack of an increase in genomic RNA levels (Figure 6.2A). However, SARS-CoV-2 infection resulted in increased SIRPα levels in primary monocytes from three different individuals (Figure 6.2B). Hence, SARS-CoV-2 may interfere with both players of the CD47- SIRPα axis.

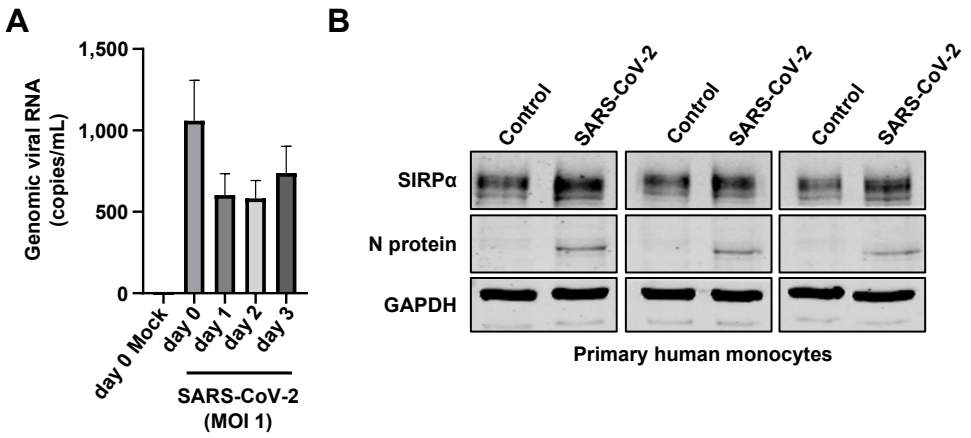


Figure 6.2. SARS-CoV-2 infection increases SIRPα in primary human monocytes. (**A**) SARS-CoV-2 strain FFM7 (MOI 1) infection of primary human monocytes does not result in the production of genomic viral RNA as detected by PCR. (**B**) SARS-CoV-2 strain FFM7 (MOI 1)-infected primary human monocytes display enhanced SIRPα levels. Uncropped blots are provided in Figure S6.4.

6.4. Discussion

Here, we show that SARS-CoV-2 infection is associated with increased CD47 expression in a range of model systems and in post mortem samples from COVID-19 patients. CD47 exerts immunosuppressive activity via interaction with SIRPα in immune cells and as a thrombospondin-1 receptor (Cham et al., 2020; Kaur et al., 2020). In this context, human CD47 expression is discussed as a strategy to enable the xenotransplantation of organs from pigs to humans (Cooper et al., 2019; Hosny et al., 2021). Moreover, high CD47 expression is an immune escape mechanism observed on cancer cells, and anti-CD47 antibodies are under investigation as cancer immunotherapeutics (Feng, R. et al., 2020; Kaur et al., 2020). Due its immunosuppressive action, CD47 expression is also discussed as a target for the treatment of viral and bacterial pathogens including SARS-CoV-2 (Cham et al., 2020; Oronsky et al., 2020; Tal et al., 2020). It has been demonstrated that cells infected with different viruses display enhanced CD47 levels, which function as a "don't eat me" signal, which interferes with the immune recognition of virus-infected cells (Tal et al., 2020). Thus, our data indicating increased CD47 levels in a range of SARS-CoV-2 infection models and clinical samples further support the potential role of CD47 as a drug target for the mediation of a more effective antiviral immune response.

Moreover, we found that, although SARS-CoV-2 did not replicate in primary human monocytes, it increased the levels of the CD47 binding partner SIRPα in these cells. Hence, SARS-CoV-2 infection may affect immune recognition of SARS-CoV-2-infected cells by upregulating both players of the CD47- SIRPα axis.

Older age, diabetes, and obesity are known risk factors for COVID-19 morbidity and mortality (Hokello et al., 2020; Shah et al., 2021). To further investigate a potential role of CD47 in severe COVID-19, we performed systematic literature searches on the relationship of CD47 and the known COVID-19 risk factors 'ageing', 'diabetes', and 'obesity'.

A literature search in PubMed (https://pubmed.ncbi.nlm.nih.gov, 17th February 2020) using the terms 'CD47' and aging' resulted in 62 hits (Table S6.1). Eight of these articles contained information that support a link between age-related increased CD47 levels and an elevated risk of severe COVID-19 (Figure 6.3A, Table S6.1). One article suggested that alpha-tocopherol reduced age-associated streptococcus pneumoniae lung infection in mice by CD47 downregulation (Bou Ghanem et al., 2015), which is in accordance with the known immunosuppressive functions of CD47 (Cham et al., 2020; Kaur et al., 2020).

The remaining seven articles reported on age-related increased CD47 levels in vascular cells that are associated with reduced vasodilatation and blood flow (Table S6.1), as CD47 signaling inhibits NO-mediated activation of soluble guanylate cyclase and in turn vasodilatation (Isenberg et al., 2008; Miller et al., 2010). Since reduced vasodilatation can cause hypertension (Touyz et al., 2018), we performed a follow-up literature search using the search terms "CD47 hypertension" (Table S6.2). This resulted in 20 hits, including a further six relevant studies (Figure 6.3B, Table S6.2). The evidence supporting a link between aging and/ or hypertension and increased CD47 levels is summarized in Table 6.1.

Initial experiments showed that loss or inhibition of CD47 prevented age- and diet-induced vasculopathy and reduced damage caused by ischemic injury in mice (Isenberg et al., 2007). CD47-deficient mice indicated that CD47 functions as a vasopressor and were also shown to be leaner and to display enhanced physical performance and a more efficient metabolism (Frazier et al., 2011; Isenberg et al., 2009). In agreement, CD47 was upregulated in clinical pulmonary hypertension and contributed to pulmonary arterial vasculopathy and dysfunction in mouse models (Bauer et al., 2012; Rogers, Sharifi-Sanjani et al., 2017). Age-related increased CD47 levels further affected peripheral blood flow and wound healing in mice (Rogers, Roberts et al., 2013) and NO-mediated vasodilatation of coronary arterioles of rats (Nevitt et al., 2016). Moreover, thrombospondin-1/ CD47 signaling was shown to induce

ageing-associated senescence in endothelial cells (Gao, Q. et al., 2016; Meijles et al., 2017) and age-associated deterioration in angiogenesis, blood flow, and glucose homeostasis (Ghimire et al., 2020).

Increased CD47 levels were also detected in the lung of a sickle cell disease patient with pulmonary arterial hypertension, and vasculopathy and pulmonary hypertension were reduced in a CD47-null mouse model of sickle cell disease (Novelli et al., 2019; Rogers, Yao et al., 2013). Finally, anti-CD47 antibodies reversed fibrosis in various organs in mouse models (Wernig et al., 2017), which may be relevant in the context of COVID-19-associated pulmonary fibrosis (Leeming et al., 2021).

In addition to immunosuppressive activity, ageing-related increased CD47 levels may thus be involved in vascular disease, vasoconstriction, and hypertension and predispose COVID-19 patients to related pathologies such as pulmonary hypertension, lung fibrosis, myocardial injury, stroke, and acute kidney injury (Cruz Rodriguez et al., 2020; Fabrizi et al., 2020; Karmouty-Quintana et al., 2020; Leeming et al., 2021; Maile et al., 2008; Rogers, Ghimire et al., 2017; Sanghvi et al., 2021; Scutelnic & Heldner, 2020; Soto-Pantoja et al., 2013).

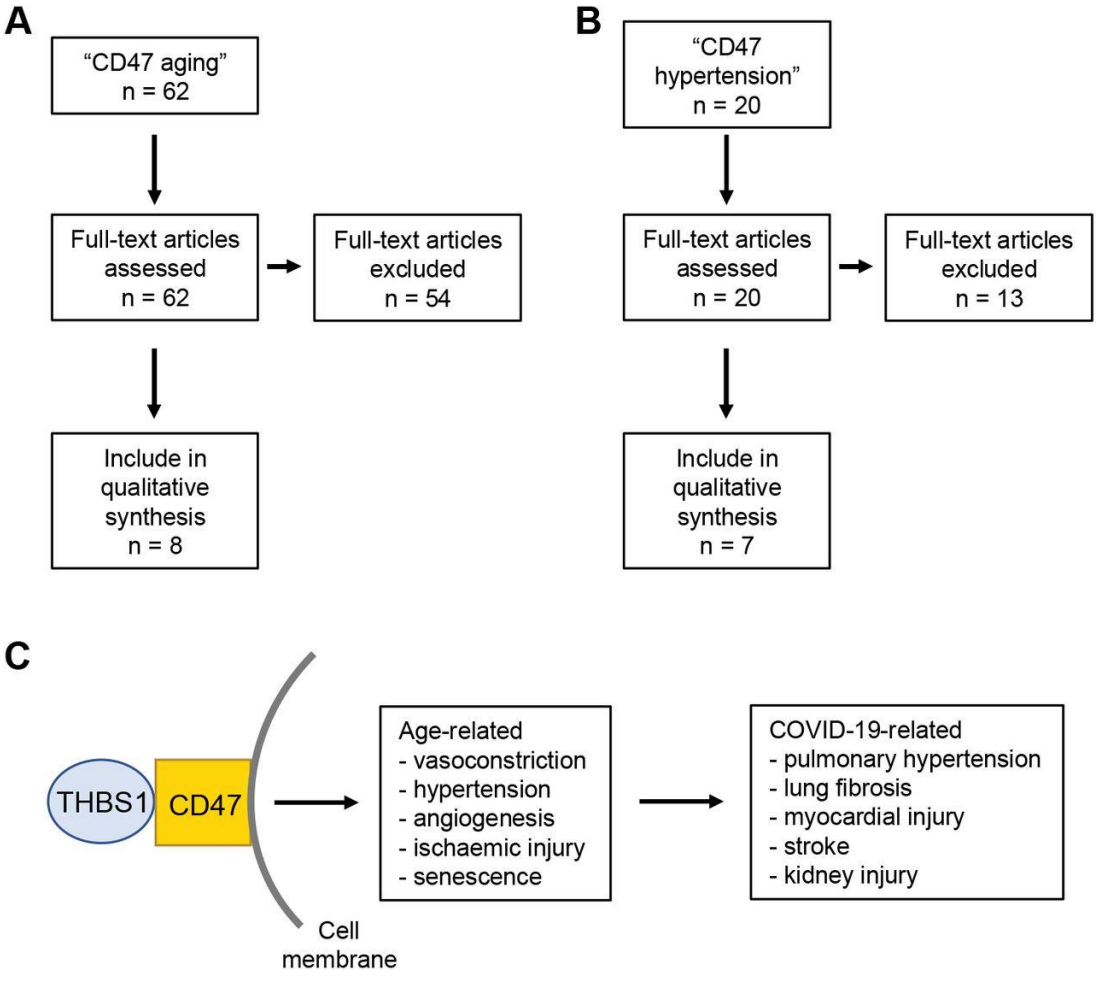


Figure 6.3. Results of the PubMed (https://pubmed.ncbi.nlm.nih.gov) literature search for "CD47 aging" (**A**) and "CD47 hypertension" (**B**). (**C**) Overview figure of the data derived from the literature searches. Age-related increased CD47 levels may contribute to pathogenic conditions associated with severe COVID-19.

Table 6.1. Evidence supporting a link between aging and/ or hypertension and increased CD47 levels.

Reference	Link between aging and/ or hypertension and increased CD47 levels
(Bou Ghanem et al., 2015)	CD47 downregulation may be involved in the alpha-tocopheol-mediated inhibition of age-associated streprococcus pneumoniae lung infection in mice
(Isenberg et al., 2007)	Blocking thrombospondin-1/CD47 signaling alleviates deleterious effects of aging on tissue responses to ischemia
(Isenberg et al., 2009)	CD47 null mice indicate that CD47 functions as vasopressor
(Frazier et al., 2011)	CD47-null mice are leaner - loss of signaling from the TSP1-CD47 system promotes accumulation of normally functioning mitochondria in a tissue-specific and age-dependent fashion leading to enhanced physical performance, lower reactive oxygen species production and more efficient metabolism
(Bauer et al., 2012)	High CD47 levels promote pulmonary arterial hypertension in lungs from humans and mice
(Rogers et al., 2017)	TSP1-CD47 signaling is upregulated in clinical pulmonary hypertension and contributes to pulmonary arterial vasculopathy and dysfunction
(Rogers et al., 2013)	Increased THBS1/ CD47 signalling contributes to reduced skin blood flow and wound healing in aged mice
(Nevitt et al., 2016)	CD47 blocks NO-mediated vasodilatation
(Gao, Q. et al., 2016)	THBS1/ CD47 signalling drives endothelial cell senescence
(Meijles et al., 2017)	TSP1 promotes ageing-associated human and mouse endothelial cell senescence through CD47
(Ghimire et al., 2020)	Increased CD47 expression causes age-associated deterioration in angiogenesis, blood flow and glucose homeostasis
(Rogers et al., 2013)	Increased CD47 levels in the lung of a sickle cell disease patient with pulmonary arterial hypertension relative to control tissues
(Novelli et al., 2019)	Pulmonary hypertension reduced in a CD47-null mouse model of sickle cell disease
(Wernig et al., 2017)	Anti-CD47 antibodies reversed fibrosis in various organs in mouse models

Diabetes has also been associated with an increased risk of severe COVID-19 and COVID-19-related death (Shah et al., 2021). A PubMed search for "CD47 diabetes" produced 47 hits, nine of which reported on increased CD47 levels in response to hyperglycemia and/ or diabetes (Figure 6.4, Table S6.3).

Hyperglycemia protected CD47 from cleavage resulting in increased CD47 levels (Allen et al., 2009; Maile et al., 2008; Maile et al., 2009; Maile et al., 2010). In agreement, increased CD47

levels were detected in various cell types and tissues in rat diabetes models and diabetes patients (Abdul-Rahman et al., 2012; Abu El-Asrar et al., 2013; Bitar, 2019; Maile et al., 2012; Wang, J. M. et al., 2014) (Table 6.2). Therefore, diabetes-induced increased CD47 levels may interfere with the recognition of SARS-CoV-2-infected cells by the immune system (Cham et al., 2020; Kaur et al., 2020).

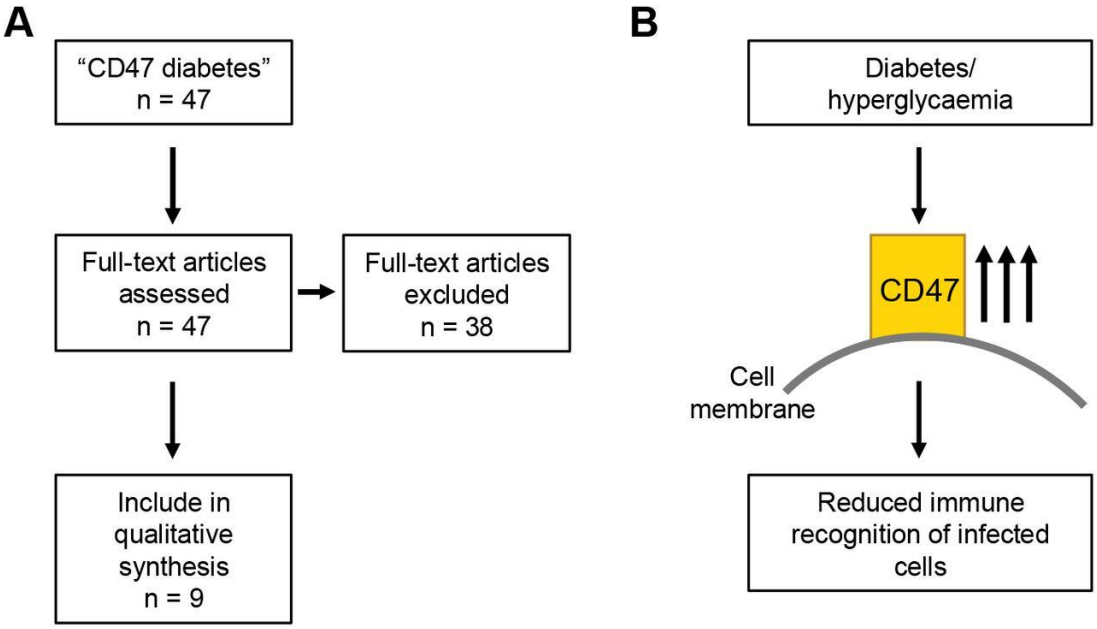


Figure 6.4. Results of the PubMed (https://pubmed.ncbi.nlm.nih.gov) literature search for "CD47 diabetes" (**A**). (**B**) Overview figure of the data derived from the literature search. Hyperglycemia- and diabetes-induced increased CD47 levels may contribute to immune escape of SARS-CoV-2-infected cells.

Table 6.2. Evidence supporting a link between diabetes and increased CD47 levels.

Reference	**Link between aging and/ or hypertension and increased CD47 levels**
(Maile et al., 2008)	Hyperglycemia protects CD47 from cleavage
(Allen et al., 2009)	Hyperglycemia protects CD47 from cleavage
(Maile et al., 2009)	Hyperglycemia protects CD47 from cleavage
(Maile et al., 2010)	Hyperglycemia protects CD47 from cleavage
(Maile et al., 2012)	CD47 is involved in pathophysiological changes in retinal cells in response to hyperglycemia in cell culture and rats
(Abdul-Rahman et al., 2012)	Elevated CD47 mRNA levels both in the hippocampus and prefrontal cortex of type 2 diabetes rat model

(Abu El-Asrar et al., 2013)	Increased levels of CD47 in epiretinal membranes with active neovascularisation in proliferative diabetic retinopathy
(Wang, J. M. et al., 2014)	Increased THBS1/ CD47 signalling in bone marrow–derived angiogenic cells in a rat diabetes model
(Bitar, 2019)	Increased diabetes-associated CD47 levels inhibit angiogenesis and wound healing in a diabetes model in rats

As obesity is another risk factor for severe COVID-19 (Shah et al., 2021), we also performed a PubMed search for "CD47 obesity", which resulted in eight hits, two of which provided potentially relevant information (Table S6.4). Results indicated that CD47-deficient mice were leaner, probably as a consequence of elevated lipolysis (Maimaitiyiming et al., 2015; Norman-Burgdolf et al., 2020). Hence, low CD47 levels may be associated both with lower weight and increased immune recognition of virus-infected cells (Cham et al., 2020; Kaur et al., 2020; Maimaitiyiming et al., 2015; Norman-Burgdolf et al., 2020), but there is no direct evidence suggesting that obesity may also directly increase CD47 expression. However, obesity may at least indirectly contribute to enhanced CD47 levels as risk factor for diabetes and hypertension (Shah et al., 2021).

Taken together, increased CD47 levels associated with aging, hypertension, and hyperglycemia and diabetes may predispose individuals to severe COVID-19.

6.5. Conclusions

Severe COVID-19 disease is the consequence of hyperinflammation ('cytokine storm') in response to SARS-CoV-2 infection (Cooper et al., 2019; Feng, R. et al., 2020; Hosny et al., 2021). Hence, the optimal time window for antiviral intervention is as early as possible to prevent disease progression to severe stages driven by immunopathology (Weinreich et al., 2021). Since the vast majority of cases are mild or even asymptomatic (Salzberger et al., 2020), an improved understanding of the processes underlying severe COVID-19 is required for the early identification of patients at high risk.

Here, we investigated a potential role of CD47 expression in determining COVID-19 severity. SARS-CoV-2 infection results in enhanced CD47 expression, which is known to interfere with the host immune response, and SARS-CoV-2 elevated the CD47 binding partner SIRPα on monocytes. Moreover, CD47 levels are elevated in groups at high risk from COVID-19 such as older individuals and individuals with hypertension and/ or diabetes. Thus, high CD47 levels may predispose these groups to severe COVID-19. Additionally, CD47 is a potential

therapeutic target, which can be addressed with antibodies and small molecules (Cham et al., 2020; Kaur et al., 2020; Oronsky et al., 2020; Tal et al., 2020). Targeting SIRPα is an alternative therapeutic option that may be associated with a favorable pharmacokinetic and safety profile, as SIRPα is restricted to monocytes and macrophages (Kuo et al., 2020).

Chapter 7: COVID-19-related coagulopathy—is transferrin a missing link?

Katie-May McLaughlin[1], Marco Bechtel[2], Denisa Bojkova[2], Christian Münch[3,4,5], Sandra Ciesek[2,6,7], Mark N Wass[1], Martin Michaelis[1], Jindrich Cinatl Jr[2]

[1]School of Biosciences, University of Kent, Canterbury CT2 7NJ, UK.

[2]Institute for Medical Virology, University Hospital, Goethe University, 60596 Frankfurt am Main, Germany.

[3]Institute of Biochemistry II, Faculty of Medicine, Goethe University, 60590 Frankfurt am Main, Germany.

[4]Frankfurt Cancer Institute, Goethe University, 60590 Frankfurt am Main, Germany.

[5]Cardio-pulmonary Institute, Goethe University, 60590 Frankfurt am Main, Germany.

[6]German Center for Infection Research, DZIF, External Partner Site, 60590 Frankfurt am Main, Germany.

[7]Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch Translational Medicine und Pharmacology, 60590 Frankfurt am Main, Germany.

Published: Diagnostics (Basel). 2020 Jul 30;10(8):539. doi: doi: https://doi.org/10.3390/diagnostics10080539.

Abstract

SARS-CoV-2 is the causative agent of COVID-19. Severe COVID-19 disease has been associated with disseminated intravascular coagulation and thrombosis, but the mechanisms underlying COVID-19-related coagulopathy remain unknown. The risk of severe COVID-19 disease is higher in males than in females and increases with age. To identify gene products that may contribute to COVID-19-related coagulopathy, we analyzed the expression of genes associated with the Gene Ontology (GO) term “blood coagulation” in the Genotype-Tissue Expression (GTEx) database and identified four procoagulants, whose expression is higher in males and increases with age (ADAMTS13, F11, HGFAC, KLKB1), and two anticoagulants, whose expression is higher in females and decreases with age (C1QTNF1, SERPINA5). However, the expression of none of these genes was regulated in a proteomics dataset of SARS-CoV-2-infected cells and none of the proteins have been identified as a binding partner

of SARS-CoV-2 proteins. Hence, they may rather generally predispose individuals to thrombosis without directly contributing to COVID-19-related coagulopathy. In contrast, the expression of the procoagulant transferrin (not associated to the GO term “blood coagulation”) was higher in males, increased with age, and was upregulated upon SARS-CoV-2 infection. Hence, transferrin warrants further examination in ongoing clinic-pathological investigations.

7.1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the causative agent of the ongoing coronavirus disease 2019 (COVID-19) outbreak (Marchandot et al., 2020; Zhu, N. et al., 2020). SARS-CoV-2 was first detected in December 2019 in the Chinese city Wuhan and has since spread around the world. The John Hopkins University Coronavirus Resource Center (Dong et al., 2020) currently reports more than 16 million confirmed COVID-19 cases and more than 650,000 confirmed COVID-19-related deaths. The majority of SARS-CoV-2 infections are mild with a substantial proportion of asymptomatic cases. However, SARS-CoV-2 causes in some patients severe life-threatening multi-organ disease (Guan et al., 2020; Iba et al., 2020; Marchandot et al., 2020; Tang, N. et al., 2020).

Severe COVID-19 disease has been associated with disseminated intravascular coagulation and thrombosis (Guan et al., 2020; Iba et al., 2020; Marchandot et al., 2020; Tang, N. et al., 2020), but the mechanisms underlying COVID-19-related coagulopathy remain unknown. It is known, however, that the risk of severe and fatal COVID-19 disease is higher in males than in females and that it increases with age (Borges do Nascimento et al., 2020). Similarly, the risk of coagulation-related pathologies and thrombosis increases with age and is further enhanced in males (Di Minno et al., 2016; Previtali et al., 2011). Thus, gene products that (1) are involved in coagulation, (2) change with age, (3) differ in their levels between females and males, and (4) are regulated in response to SARS-CoV-2 infection represent candidate factors that may contribute to COVID-19-related coagulopathy and disease severity.

To identify such candidate factors that may be involved in COVID-19-related coagulopathy, we here performed a combined analysis of a proteomics dataset derived from SARS-CoV-2-infected cells (Bojkova et al., 2020), of a dataset of host cell proteins found to bind to SARS-CoV-2 proteins (Gordon et al., 2020), and of human gene expression data from the Genotype-Tissue Expression (GTEx) database (GTEx Consortium, 2013).

7.2. Methods

7.2.1. Data Acquisition

Gene Ontology is an initiative that annotates genes with functions (The Gene Ontology Consortium, 2019). Genes associated with the Gene Ontology (GO) term "Blood Coagulation" (GO:0007596) were identified using the online database AmiGO 2 (The Gene Ontology Consortium, 2019). This generated a list of 335 unique genes annotated with 23 unique terms (including "Blood Coagulation" and 22 child terms) for further analysis.

Gene expression data (transcripts per million, TPM) and clinical data for 980 individuals (17,382 samples from 30 tissues) were downloaded from the GTEx Portal (https://www.gtexportal.org/home/datasets; GTEx Project, version 8). We also used normalized protein abundance data from a recent publication (Bojkova et al., 2020) in which protein abundance in uninfected and SARS-CoV-2-infected Caco-2 (SARS-CoV-2-susceptible colorectal cancer cell line) cells was quantified. Data were subsequently normalized using summed intensity normalization for sample loading, followed by internal reference scaling and Trimmed mean of M normalization.

We also queried the EBI IntAct database (https://www.ebi.ac.uk/intact/) for "annot:dataset"coronavirus"", and filtered for interactions between human proteins and SARS-CoV-2 proteins. These SARS-CoV-2-interacting proteins were derived from a study by Gordon et al. (Gordon et al., 2020), in which 29 SARS-CoV-2 proteins were cloned, tagged, and expressed in HEK293T cells, and incubated for 40 h prior to affinity purification and identification of binding partners by mass spectrometry. This resulted in 332 high confidence human protein interactors for 26 of the SARS-CoV-2 proteins.

7.2.2. Data Analysis

Analyses were performed using R3.6.1. Linear models were generated to estimate the relationship between gene expression and age using the base R function lm, which generated p-values indicating the significance of the relationship. Models with a p-value <0.05 were considered significant. Plots were generated using the R package ggplot2. Mean protein abundance (for proteomics data) and median gene expression TPM (for GTEx data) was plotted using the function ggviolin. p-values indicating the significance of the difference between gene expression/protein abundance in males and females in each given age group were the result of a Wilcoxon rank sum test for independent groups. For the proteome data,

we performed a two-sided student's t-test. Boxplots comparing gene expression in males and females were generated using the function ggboxplot, for which p-values were the result of a Wilcoxon rank sum test for independent groups.

7.3. Results

7.3.1. Identification of Genes that May Be Associated with an Increased Coagulation Risk in Males and at an Older Age

Using AmiGO 2 (GTEx Consortium, 2013), we identified 335 genes, which are associated with the GO term "blood coagulation" (GO:0007596). Since the risk of severe COVID-19 disease increases with age and is higher in males than in females (Borges do Nascimento et al., 2020), coagulation-associated genes which may be relevant in the context of COVID-19-related coagulopathy would be expected to differ in their expression between females and males and change in their expression with age.

An analysis of the genes associated with the GO term "blood coagulation" using the Genotype-Tissue Expression (GTEx) database (GTEx Consortium, 2013) resulted in 256 coagulation-associated genes, that are differently expressed between females and males (Table S7.1) and 237 genes whose expression changed with age (Table S7.2). These lists included many genes, whose products are involved in the regulation of upstream processes, which may be linked to coagulation in certain cell types and under certain circumstances but are not core players directly involved in the actual coagulation process. Examples include members of major signaling cascades such as PI3K or MAPK signaling (Tables S6.1 and S6.2). Hence, the functions of these genes were manually annotated to identify candidate genes, whose products act as procoagulants and anticoagulants (Tables S6.1 and S6.2). This resulted in a list of 49 overlapping genes, whose products are directly involved in coagulation (Table S7.3).

Two groups of genes were considered as candidates, whose products may increase or reduce the risk of COVID-19-related coagulopathy: (1) procoagulants that display higher expression in males and increase in their expression with age and (2) anticoagulants that display higher expression in females and decrease in their expression with age. According to these criteria, we found four procoagulants (ADAMTS13, F11, HGFAC, KLKB1) and two anticoagulants (C1QTNF1, SERPINA5), whose expression may predispose males and older individuals to

COVID-19-related coagulopathy and severe COVID-19 disease (Table 7.1, Figure 7.1, Table S7.3).

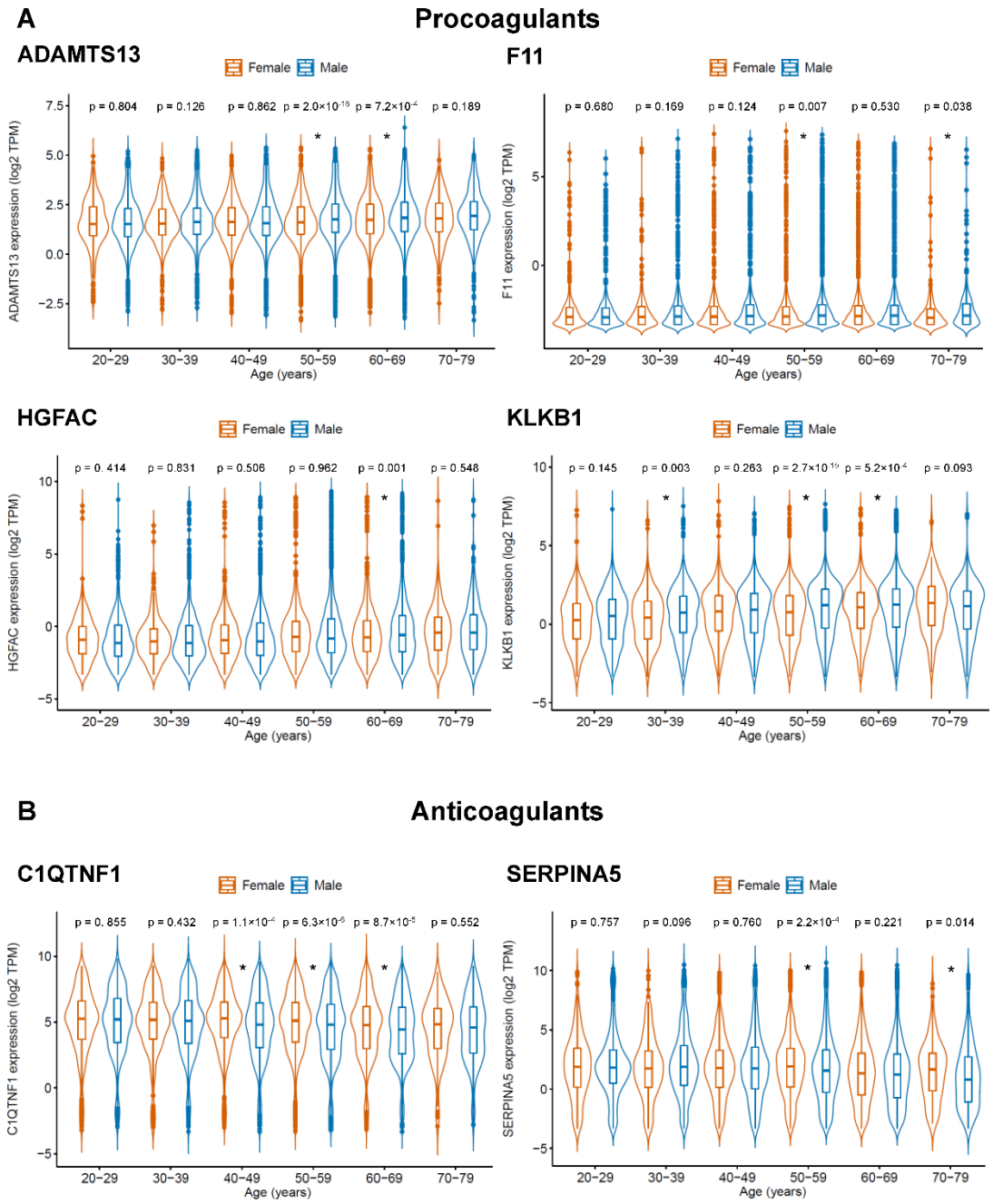


Figure 7.1. Gene products anticipated to be of potential relevance for COVID-19-related coagulopathy, based on genes with a role in coagulation that are differentially expressed between females and males (Table S7.1) and whose expression correlates with age (Table S7.2). Candidate gene products were either (**A**) procoagulants (ADAMTS13, F11, HGFAC, KLKB1), which display higher expression in males than in females and increase with age, or (**B**) anticoagulants (C1QTNF1, SERPINA5), which display

higher expression in females than in males and decrease with age. A complete list of the relevant genes overlapping between Table S7.1 and Table S7.2 is presented in Table S7.3. *p*-values were determined by two-sided Student's t-test. * *p*-value < 0.05.

Table 7.1. Candidate gene products that may be involved in COVID-19-related coagulopathy. Candidates were either (**A**) procoagulants (ADAMTS13, F11, HGFAC, KLKB1), which display a lower expression in females than in males and increase with age, or (**B**) anticoagulants (C1QTNF1, SERPINA5), which display higher expression in females than in males and decrease with age.

(A) Procoagulants				
	Female vs. male		**Age-associated expression**	
	Relative expression	***p*-value ***	**Direction**	***p*-value**
ADAMTS13	low	8.9×10^{-8}	increase	1.6×10^{-11}
F11	low	3.5×10^{-4}	increase	$<2.2 \times 10^{-16}$
HGFAC	low	0.032	increase	0.049
KLKB1	low	$<2.2 \times 10^{-16}$	increase	4.3×10^{-6}
(B) Anticoagulants				
	Female vs. male		**Age-associated expression**	
	Relative expression	***p*-value**	**Direction**	***p*-value**
C1QTNF1	high	6.7×10^{-13}	decrease	2.6×10^{-16}
SERPINA5	high	3.7×10^{-3}	decrease	2.8×10^{-6}

* *p*-value < 0.05 were considered as significantly different.

7.3.2. No Overlap between COVID-19-Related Coagulopathy Predisposition Genes and SARS-CoV-2-Associated Genes

Next, we investigated whether there is a known relationship between the candidate genes, whose products may predispose individuals to severe COVID-19 disease, and SARS-CoV-2

infection. For this, we used a proteomics and translatome dataset derived from SARS-CoV-2-infected and non-infected cells (Bojkova et al., 2020) (Table S7.4) and 332 high confidence human SARS-CoV-2 interactor proteins, which had been identified by expressing 29 tagged SARS-CoV-2 proteins in HEK293T cells, followed by affinity purification and mass spectrometric identification of binding partners (Gordon et al., 2020) (Table S7.4). However, none of our six candidates were shown to be regulated by or interact with SARS-CoV-2 (Tables S6.4 and S6.5).

7.3.3. Transferrin May Be Involved in COVID-19-Related Coagulopathy

While searching manually for additional candidates potentially involved in COVID-19-related coagulopathy, we found transferrin to be upregulated in SARS-CoV-2-infected cells relative to non-infected cells (Bojkova et al., 2020) (Figure 7.2A). Transferrin is a glycoprotein circulating in the blood that is best known for its function as an iron carrier. It binds to cellular transferrin receptors and delivers iron by receptor-mediated endocytosis (Kawabata, 2019; Luck & Mason, 2012). However, transferrin has also been shown to increase coagulation independent of its role as an iron transporter by interfering with antithrombin/SERPINC1-mediated inhibition of coagulation proteases including thrombin and factor XIIa (Tang, X. et al., 2020). Hence, there might be a link between transferrin levels and coagulation in COVID-19 patients.

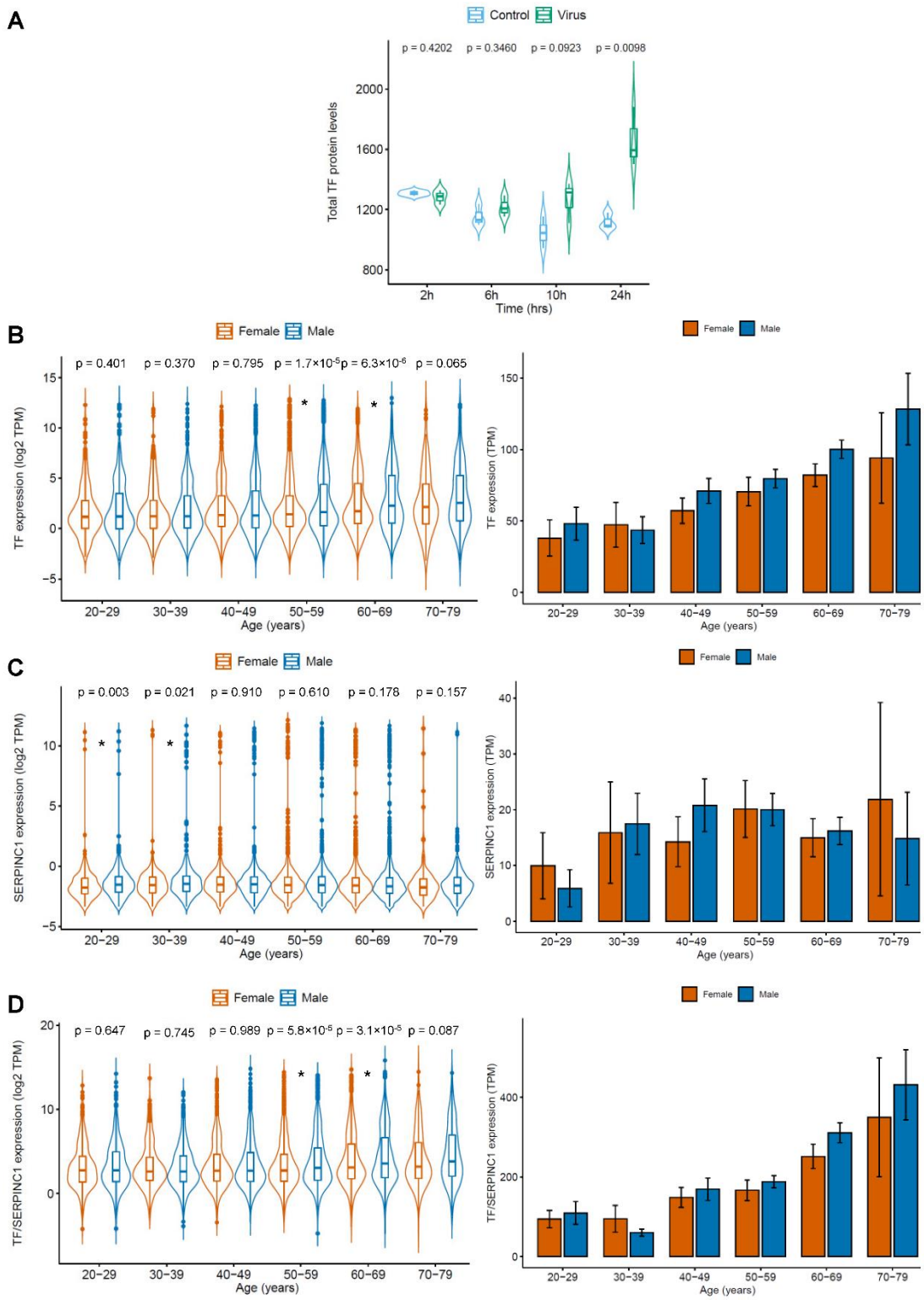


Figure 7.2. SARS-CoV-2-induced (derived from (Bojkova et al., 2020)) transferrin (TF) expression and age- and gender-specific expression of the procoagulant TF and its antagonist SERPINC1/antithrombin based on GTEx data. Data are presented as violin blots to indicate the distribution of individual values and as bar charts to facilitate comparisons. TF displayed higher levels in SARS-CoV-2-infected cells than in non-infected cells. Moreover, TF expression and the expression ration of TF and SERPINC1

increased with age and were higher in females than in males. (**A**) TF protein abundance in uninfected (control) and SARS-CoV-2-infected (virus) Caco-2 cells. *p*-values are the result of a two-sided Student's *t*-test. (**B**) TF expression (TPM) in females and males across six age groups. *p*-values were calculated using the Wilcoxon rank sum test for independent groups. (**C**) SERPINC1 expression (TPM) in females and males across six age groups. *p*-values were calculated using the Wilcoxon rank sum test for independent groups. (**D**) Ratio of TF/SERPINC1 expression (TPM) in females and males across six age groups. *p*-values were calculated using the Wilcoxon rank sum test for independent groups. * *p*-value < 0.05.

GTEx data indicated that transferrin expression increased with age and was higher in males than in females (Figure 7.2B). In contrast, expression of its antagonist antithrombin did not increase with age and was similar in females and males (Figure 7.2C). Thus, the transferrin/antithrombin ratio increases with age and is higher in males than in females (Figure 7.2D). This correlates with the risk of severe and fatal COVID-19 disease, which is higher in males than in females and also increases with age (Borges do Nascimento et al., 2020). Hence, an increased transferrin/antithrombin ratio may contribute to COVID-19-related coagulopathy and more severe disease in older patients, in particular in males.

Transferrin was not included in the list of SARS-CoV-2-interacting proteins (Gordon et al., 2020) (Table S7.4). This suggests that transferrin is regulated in response to SARS-CoV-2 infection but does not directly interact with SARS-CoV-2 proteins.

7.4. Discussion

Severe COVID-19 disease is associated with intravascular coagulation and thrombosis (COVID-19-related coagulopathy) (Buja et al., 2020; Guan et al., 2020; Iba et al., 2020; Marchandot et al., 2020; Tang, N. et al., 2020) and the risk of severe disease increases with age and is higher in males than in females (Borges do Nascimento et al., 2020). To identify factors involved in coagulation that may contribute to COVID-19-related coagulopathy, we used genes associated with the GO term "blood coagulation" and the GTEx database resulting in four procoagulants, whose expression was higher in males than in females and increased with age (ADAMTS13, F11, HGFAC, KLKB1), and two anticoagulants, whose expression was higher in females and decreased with age (C1QTNF1, SERPINA5).

However, these candidate factors were not found to be regulated in SARS-CoV-2-infected cells (Bojkova et al., 2020) or among proteins known to interact with SARS-CoV-2 proteins (Gordon et al., 2020). Hence, they may rather generally predispose individuals to coagulopathy, which is known to be higher in males and to increase with age (Di Minno et al., 2016; Previtali et al., 2011), than being directly involved in the disease processes mediated by SARS-CoV-2 infection. Therefore, these factors may also be relevant in age- and gender-related thrombosis formation beyond COVID-19.

Our manual further investigations identified transferrin as a candidate factor, which may be involved in COVID-19-related coagulopathy. The expression of transferrin, a known procoagulant (Tang, X. et al., 2020), was upregulated in SARS-CoV-2-infected cells, increased with age, and was higher in males than in females. Moreover, a rise of transferrin levels was observed in patients during COVID-19 disease progression (Bolondi et al., 2020). Transferrin is an iron carrier protein that circulates and delivers iron to cells via transferrin receptor binding followed by receptor-mediated endocytosis (Kawabata, 2019; Luck & Mason, 2012). However, it also promotes coagulation by iron-independent mechanisms as an inhibitor of antithrombin, which interferes with the prothrombotic activity of coagulation proteases such as thrombin and factor XIIa (Tang, X. et al., 2020).

Transferrin is primarily produced in the liver. However, (SARS-CoV-2-induced) locally produced transferrin may contribute to COVID-19 pathology, even independent of circulating transferrin levels (Kawabata, 2019; Luck & Mason, 2012; Lum et al., 1986; Mateos et al., 1998; McClain et al., 2018; Murakami et al., 2019; Yang, F. et al., 1997; Zakin et al., 2002). For example, transferrin is produced in the brain (Zakin et al., 2002), and high transferrin levels have been associated with hypercoagulability and ischemic stroke (Zakin et al., 2002). Stroke is a significant complication in COVID-19 (Tang, X. et al., 2020) and is much more common in COVID-19 than, for example, in influenza patients (Koralnik & Tyler, 2020). Both ischemic and hemorrhagic strokes are observed in COVID-19 patients (Tang, X. et al., 2020). Notably, transferrin may not only contribute to ischemic strokes via inducing coagulation (Merkler et al., 2020), it may also increase the brain injury associated with hemorrhagic strokes by facilitating cellular iron uptake (Nakamura et al., 2005). High transferrin levels have also been associated with diabetes and metabolic syndrome (Fumeron et al., 2006; McClain et al., 2018; Vari et al., 2007; Zhao, H. Q. et al., 2015), which are known risk factors for severe COVID-19 disease.

7.5. Conclusions

In conclusion, the role of transferrin in the course of COVID-19 disease and in particular of COVID-19-related coagulopathy should be considered and further examined in ongoing clinico-pathological investigations. If the role of transferrin is confirmed in the pathogenesis of severe COVID-19 disease and in COVID-19-related coagulopathy, it is a candidate diagnostic marker for the monitoring of COVID-19 progression and may guide the use of anticoagulants in COVID-19 patients.

Chapter 8: Aprotinin inhibits SARS-CoV-2 replication

Denisa Bojkova[1], Marco Bechtel[1], Katie-May McLaughlin[2], Jake E McGreig[2], Kevin Klann[3], Carla Bellinghausen[4], Gernot Rohde[4,] Danny Jonigk[5,6], Peter Braubach[5,6], Sandra Ciesek[1,7,8], Christian Münch[3,9,10], Mark N Wass[2], Martin Michaelis[2], Jindrich Cinatl Jr[1]

[1]Institute for Medical Virology, University Hospital, Goethe University, 60596 Frankfurt am Main, Germany.

[2]School of Biosciences, University of Kent, Canterbury CT2 7NJ, UK.

[3]Faculty of Medicine, Institute of Biochemistry II, Goethe University, 60590 Frankfurt am Main, Germany.

[4]Department of Respiratory Medicine and Allergology, University Hospital, Goethe University, 60590 Frankfurt am Main, Germany.

5Institute of Pathology, Hannover Medical School (MHH), 30625 Hannover, Germany.

[6]Biomedical Research in Endstage and Obstructive Lung Disease Hannover (BREATH), The German Center for Lung Research (Deutsches Zentrum für Lungenforschung, DZL), Hannover Medical School (MHH), 30625 Hannover, Germany.

[7]German Center for Infection Research, DZIF, External Partner Site, 60596 Frankfurt am Main, Germany.

[8]Fraunhofer Institute for Molecular Biology and Applied Ecology (IME), Branch Translational Medicine und Pharmacology, 60596 Frankfurt am Main, Germany.

[9]Frankfurt Cancer Institute, Goethe University, 60596 Frankfurt am Main, Germany.

[10]Cardio-Pulmonary Institute, Goethe University, 60590 Frankfurt am Main, Germany.

Published: Cells. 2020 Oct 30;9(11):2377. doi: https://doi.org/10.3390/cells9112377.

Abstract

Severe acute respiratory syndrome virus 2 (SARS-CoV-2) is the cause of the current coronavirus disease 19 (COVID-19) pandemic. Protease inhibitors are under consideration as virus entry inhibitors that prevent the cleavage of the coronavirus spike (S) protein by cellular proteases. Herein, we showed that the protease inhibitor aprotinin (but not the protease inhibitor SERPINA1/alpha-1 antitrypsin) inhibited SARS-CoV-2 replication in therapeutically achievable concentrations. An analysis of proteomics and translatome data indicated that SARS-CoV-2 replication is associated with a downregulation of host cell protease inhibitors.

Hence, aprotinin may compensate for downregulated host cell proteases during later virus replication cycles. Aprotinin displayed anti-SARS-CoV-2 activity in different cell types (Caco2, Calu-3, and primary bronchial epithelial cell air–liquid interface cultures) and against four virus isolates. In conclusion, therapeutic aprotinin concentrations exert anti-SARS-CoV-2 activity. An approved aprotinin aerosol may have potential for the early local control of SARS-CoV-2 replication and the prevention of COVID-19 progression to a severe, systemic disease.

8.1. Introduction

Severe acute respiratory syndrome virus 2 (SARS-CoV-2), a novel betacoronavirus, causes a respiratory disease and pneumonia called coronavirus disease 19 (COVID-19) and is the cause of a current pandemic responsible for millions of cases and hundreds of thousands of deaths (Chen, N. et al., 2020; Coronaviridae Study Group of the International Committee on Taxonomy of Viruses, 2020; Dong et al., 2020; Lu et al., 2020; Wu, F. et al., 2020; Zhou, P. et al., 2020; Zhu, N. et al., 2020). Drugs for the treatment of COVID-19 are urgently needed.

Cell entry of coronaviruses is mediated by the interaction of the viral spike (S) protein with their host cell receptors, which differ between different coronaviruses (Cui et al., 2019). For example, Middle East respiratory syndrome coronavirus (MERS-CoV) uses dipeptidyl peptidase 4 (DPP4) as a cellular receptor (Cui et al., 2019). Host cell entry of SARS-CoV-2 and of the closely related severe acute respiratory syndrome virus (SARS-CoV) is mediated by angiotensin-converting enzyme 2 (ACE2) (Cui et al., 2019; Hoffmann et al., 2020; Matsuyama et al., 2020). S binding to ACE2 depends on S cleavage at three sites (S1, S2, and S2') by host cell proteases, typically by the transmembrane serine protease 2 (TMPRSS2), and can be inhibited by serine protease inhibitors (Hoffmann et al., 2020; Matsuyama et al., 2020). Camostat was the first serine protease inhibitor that was shown to inhibit TMPRSS2 (Hoffmann et al., 2020). Subsequently, additional TMPRSS2 inhibitors, including nafamostat and Arbidol derivatives, were demonstrated to interfere with SARS-CoV-2 internalization into host cells (Choudhary & Silakari, 2020; Hoffmann et al., 2020a; Yamamoto et al., 2020).

Aprotinin is a serine protease inhibitor, which has previously been shown to inhibit TMPRSS2 and has been suggested as a treatment option for influenza viruses and coronaviruses (Shen, L. W. et al., 2017; Zhirnov et al., 2011). Herein, we investigated the effects of aprotinin against SARS-CoV-2.

8.2. Materials and Methods

8.2.1. Drugs

SERPINA1/alpha-1 antitrypsin (Prolastin) was obtained from Grifols (Barcelona, Spain). Aprotinin was purchased from Sigma-Aldrich (Darmstadt, Germany)).

8.2.2. Cell Culture

The Caco2 cell line was obtained from DSMZ (Braunschweig, Germany), and Calu-3 from ATCC (Manassas, VA, US). The cells were grown at 37 °C in minimal essential medium (MEM) supplemented with 10% fetal bovine serum (FBS), 100 IU/mL of penicillin, and 100 µg/mL of streptomycin. All culture reagents were purchased from Sigma-Aldrich. Cells were regularly authenticated by short tandem repeat (STR) analysis and tested for mycoplasma contamination.

Lung tissue for the isolation of primary epithelial cells was provided by the Hannover Medical School, Institute of Pathology (Hannover, Germany). The use of tissue was approved by the ethics committee of the Hannover Medical School (MHH, Hannover, Germany, number 2701–2015) and was in compliance with The Code of Ethics of the World Medical Association. Primary bronchial epithelial cells were isolated from the lung explant tissue of a patient with lung emphysema as described previously (Van Wetering et al., 2000). All patients or their next of kin gave written informed consent for the use of their lung tissue for research. Basal cells were expanded in Keratinocyte-SFM medium supplemented with bovine pituitary extract (25 µg/mL), human recombinant epidermal growth factor (0.2 ng/mL, all from Gibco, Schwerte, Germany), isoproterenol (1 nM, Sigma), Antibiotic/Antimycotic Solution (Sigma-Aldrich), and MycoZap Plus PR (Lonza, Cologne, Germany) and cryopreserved until further use.

For differentiation, the cells were thawed and passaged once in PneumaCult-Ex Medium (StemCell Technologies, Cologne, Germany) and then seeded on transwell inserts (12-well plate, Sarstedt, Nümbrecht, Germany) at 4 × 104 cells/insert. Once the cell layers reached confluency, the medium on the apical side of the transwell was removed, and medium in the basal chamber was replaced with PneumaCult ALI Maintenance Medium (StemCell Technologies), including Antibiotic/Antimycotic Solution (Sigma-Aldrich) and MycoZap Plus PR (Lonza). During a period of four weeks, the medium was changed and the cell layers were washed with PBS every other day. Criteria for successful differentiation were the

development of ciliated cells and ciliary movement, an increase in transepithelial electric resistance indicative of the formation of tight junctions, and mucus production.

8.2.3. Virus Infection

The isolates SARS-CoV-2/1/Human/2020/Frankfurt (SARS-CoV-2/FFM1), SARS-CoV-2/2/Human/2020/Frankfurt (SARS-CoV-2/FFM2), SARS-CoV-2/6/Human/2020/Frankfurt (SARS-CoV-2/FFM6), and SARS-CoV-2/7/Human/2020/Frankfurt (SARS-CoV-2/FFM7) were isolated and cultivated in Caco2 cells as previously described (Hoehl et al., 2020; Toptan et al., 2020). Virus titers were determined as TCID50/mL in confluent cells in 96-well microtiter plates (Cinatl Jr et al., 2005; Cinatl et al., 2003).

8.2.4. Antiviral Assay

Confluent cell cultures were infected with SARS-CoV-2 in 96-well plates at a multiplicity of infection (MOI) of 0.01 in the absence or presence of the drug. The cytopathogenic effect (CPE) was assessed visually 48 h post-infection (Cinatl et al., 2003). Concentrations that inhibited CPE formation by 50% (IC50) were determined using CalcuSyn (Biosoft, Cambridge, UK).

8.2.5. Viability Assay

Cell viability was determined by 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay modified after Mosman (Mosmann, 1983), as previously described (Onafuye et al., 2019). Confluent cell cultures in 96-well plates were incubated with the drug for 48 h. Then, 25 µL of MTT solution (2 mg/mL (w/v) in PBS) were added per well, and the plates were incubated at 37 °C for an additional 4 h. After this, the cells were lysed using 200 µL of a buffer containing 20% (w/v) sodium dodecylsulfate and 50% (v/v) N,N-dimethylformamide with the pH adjusted to 4.7 at 37 °C for 4 h. Absorbance was determined at 570 nm for each well using a 96-well multiscanner (Tecan, Crailsheim, Germany). After subtracting of the background absorption, the results are expressed as percentage viability relative to control cultures that received no drug. Drug concentrations that inhibited cell viability by 50% (CC50) were determined using CalcuSyn (Biosoft).

8.2.6. Immunostaining for SARS-CoV-2 S Protein

Immunostaining was performed as previously described (Cinatl Jr et al., 1995), using a monoclonal antibody directed against SARS-CoV-2 S protein (1:1500 dilution, Sino Biological, Eschborn, Germany) 24 h post-infection.

8.2.7. Caspase 3/7 Activation

Caspase 3/7 activation was determined using the Caspase-Glo® 3/7 Assay (Promega, Walldorf, Germany) according to the manufacturer's instructions.

8.2.8. qPCR

SARS-CoV-2 RNA from the cell culture supernatant samples was isolated using AVL buffer and the QIAamp Viral RNA Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Absorbance-based quantification of the RNA yield was performed using the Genesys 10S UV-Vis Spectrophotometer (Thermo Fisher Scientific, Dreieich, Germany). RNA was subjected to OneStep qRT-PCR analysis using the Luna Universal One-Step RT-qPCR Kit (New England Biolabs, Frankfurt am Main, Germany) and a CFX96 Real-Time System, C1000 Touch Thermal Cycler (Bio-Rad, Feldkirchen, Germany). Primers were adapted from the WHO protocol29 targeting the open reading frame for RNA-dependent RNA polymerase (RdRp): RdRP_SARSr-F2 (GTG ARA TGG TCA TGT GTG GCG G) and RdRP_SARSr-R1 (CAR ATG TTA AAS ACA CTA TTA GCA TA) using 0.4 µM per reaction. Standard curves were created using plasmid DNA (pEX-A128-RdRP) harboring the corresponding amplicon regions for RdRp target sequence according to GenBank Accession number NC_045512. For each condition, three biological replicates were used. The mean and standard deviation were calculated for each group.

8.2.9. Western Blot

Cells were lysed using Triton-X-100 sample buffer (Sigma-Aldrich), and proteins were separated by SDS-PAGE. Detection occurred by using specific antibodies against SARS-CoV-2 N (1:1000 dilution, SARS-CoV-2 Nucleocapsid Antibody, Rabbit monoclonal antibody (Mab), #40143-R019, Sino Biological), ACE2 (1:500 dilution, Anti-ACE2 antibody, #ab15348, Abcam, Berlin, Germany), TMPRSS2 (1:1000 dilution, Recombinant Anti-TMPRSS2 antibody [EPR3861], #ab92323, Abcam), and GAPDH (1:1000 dilution, Anti-G3PDH Human Polyclonal Antibody, #2275-PC-100, Trevigen, Wiesbaden, Germany). Protein bands were visualized by laser-induced fluorescence using an infrared scanner for protein quantification (Odyssey, Li-Cor Biosciences, Bad Homburg, Germany).

8.2.10. Sample Preparation for LC–MS

Preparation of samples was performed as previously described (Klann et al., 2020) and labeled with TMTpro multiplexing reagents.

8.2.11. Targeted Analysis by SPS–MS3

Mass spectrometry data were acquired in centroid mode on an Orbitrap Fusion Lumos mass spectrometer hyphenated to an easy-nLC 1200 nano HPLC system using a nanoFlex ion source (ThermoFisher Scientific) applying a spray voltage of 2.6 kV with the transfer tube heated to 300 °C and a funnel RF of 30%. Internal mass calibration was enabled (lock mass 445.12003 m/z). Peptides were separated on a self-made, 32 cm long, 75 µm ID fused-silica column, packed in house with 1.9 µm C18 particles (ReproSil-Pur, Dr. Maisch, Ammerbuch-Entringen, Germany) and heated to 50 °C using an integrated column oven (Sonation, Biberach, Germany). The HPLC solvents consisted of 0.1% formic acid in water (Buffer A) and 0.1% formic acid with 80% acetonitrile in water (Buffer B).

Dependent scans were performed on precursors matching a mass list of viral peptides modified with TMTpro reagents and their charge states (mass tolerance was set to 5 ppm for matching precursors). Peptides were eluted by a non-linear gradient from 5% to 40% B over 30 min, followed by a step-wise increase to 95% B in 6 min, which was held for another 9 min. Full scan MS spectra (350–1500 m/z) were acquired with a resolution of 120,000 at m/z 200, a maximum injection time of 100 ms, and an automatic gain control (AGC) target value of 4 × 105. The 10 most intense precursors matching the target list per full scan were selected for fragmentation (“Top 10”) and isolated with a quadrupole isolation window of 0.4 Th. MS2 scans were performed in the Orbitrap using a maximum injection time of 300 ms, an AGC target value of 1.5 × 104, and fragmented using HCD with a normalized collision energy (NCE) of 35% and a fixed first mass of 110 m/z. Repeated sequencing of already acquired precursors was limited by setting a dynamic exclusion of 20 s and 10 ppm and advanced peak determination was deactivated.

8.2.12. Data Analysis

RAW data was processed with Proteome Discoverer 2.4 software. HCD-fragmented spectra were searched against a SARS-CoV-2 proteome FASTA file (UniProt pre-realease) by SequestHT and the false discovery rate (FDR) was calculated using a target/decoy-based approach. TMTpro reporter abundances were extracted and used for plotting and statistical analysis.

8.2.13. Data Availability

The mass spectrometry proteomics data were deposited to the ProteomeXchange Consortium via the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD019950.

8.3. Results

8.3.1. The Protease Inhibitor Aprotinin Exerts Superior Anti-SARS-CoV-2 Activity Relative to the Endogenous Protease Inhibitor SERPINA1/alpha-1 Antitrypsin

We compared the anti-SARS-CoV-2 activity of aprotinin (Shen, L. W. et al., 2017; Solun & Shoenfeld, 2020) and SERPINA1/alpha-1 antitrypsin, an endogenous protease inhibitor that is available as a pharmaceutical preparation for the treatment of alpha-1 antitrypsin deficiency (Strnad et al., 2020), against three different SARS-CoV-2 isolates from two lineages (L: SARS-CoV-2/FFM1 and SARS-CoV-2/FFM2; GR: SARS-CoV-2/FFM6) (Toptan et al., 2020). SARS-CoV-2/FFM1 and SARS-CoV-2/FFM2 were isolated from patients in Hubei province in China, while SARS-CoV/FFM6 was derived from an Italian patient (Toptan et al., 2020).

The aprotinin concentrations that inhibited the formation of cytopathogenic effects (CPEs) by 50% (IC50) in SARS-CoV-2-infected Caco2 cells ranged from 0.81 µM (SARS-CoV-2/FFM2) to 1.03 µM (SARS-CoV-2/FFM1) across the three tested SARS-CoV-2 isolates, whereas SERPINA1/alpha-1 antitrypsin did not show significant antiviral effects in the tested concentrations up to 20 µM (Figure 8.1A). Similar effects were observed by cell staining for SARS-CoV-2 S protein (Figure 8.1B and Figure S8.1, Table 8.1). Quantification of genomic SARS-CoV-2 RNA using qPCR confirmed that aprotinin inhibits SARS-CoV-2 replication (Figure 8.1C). Aprotinin (20 µM) reduced the genomic RNA levels of SARS-CoV-2/FFM1 by 900-fold, those of SARS-CoV-2/FFM2 by 237-fold, and those of SARS-CoV-2/FFM6 by 584-fold.

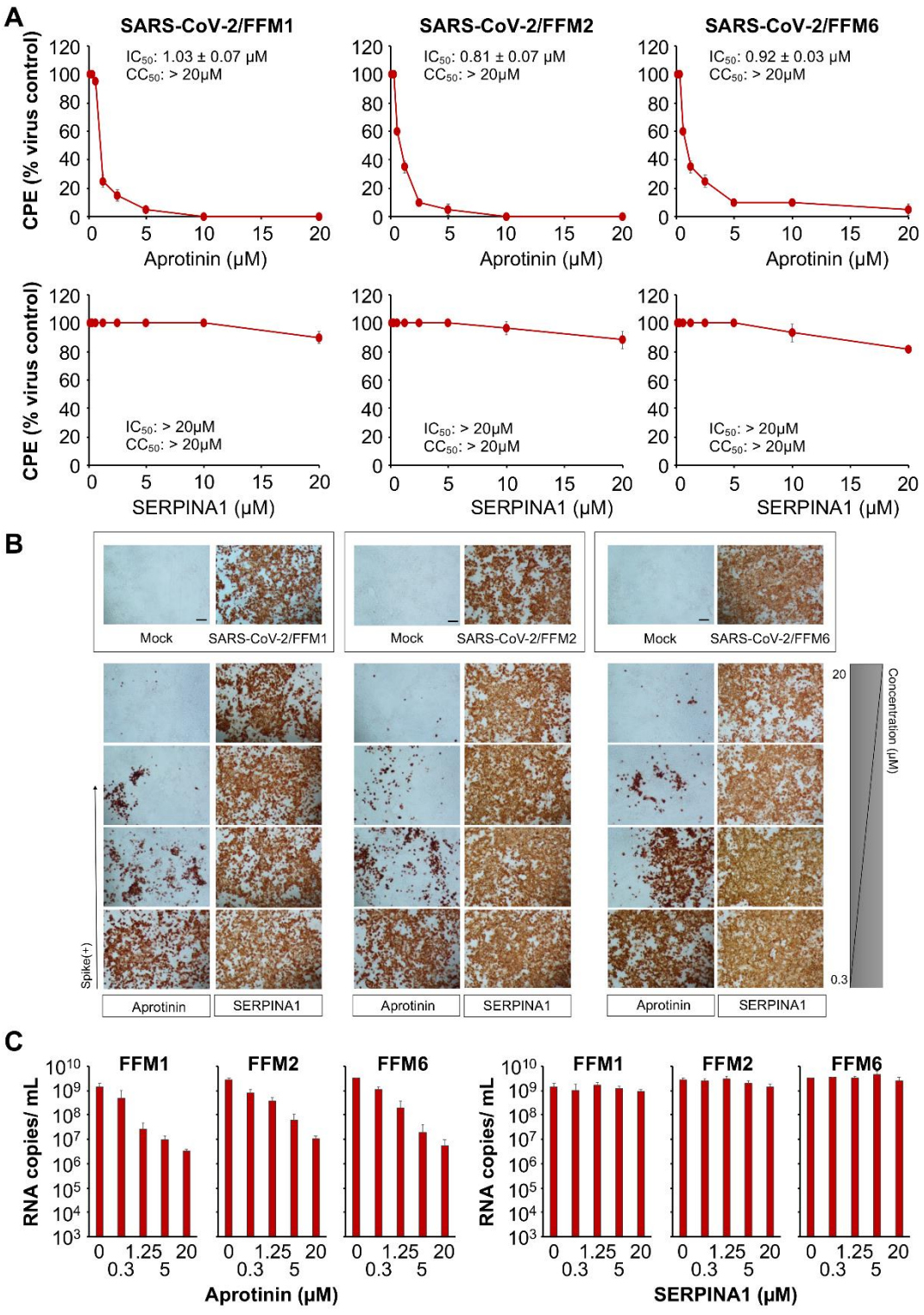


Figure 8.1. Anti-severe acute respiratory syndrome virus 2 (SARS-CoV-2) effects of aprotinin and SERPINA1/alpha-1 antitrypsin. (**A**) Concentration-dependent effects of aprotinin and SERPINA1/alpha-1 antitrypsin on SARS-CoV-2-induced cytopathogenic effect (CPE) formation determined 48 h post-infection in Caco2 cells infected at a multiplicity of infection (MOI) of 0.01 with the three different SARS-CoV-2 isolates. The viability of the Caco2 cells was 84.3 ± 2.7% relative to the untreated control

in the presence of 20 µM of aprotinin. (**B**) Immunostaining for the SARS-CoV-2 S protein in aprotinin- and SERPINA1/alpha-1 antitrypsin-treated Caco2 cells infected at an MOI of 0.01 with the three different SARS-CoV-2 isolates as determined 48 h post-infection. The protease inhibitors were tested at four concentrations in 1:4 dilution steps ranging from 20 to 0.3125 µM. A quantification is provided in Figure S8.1. (**C**) Copy numbers of genomic RNA in Caco2 cells infected with different SARS-CoV-2 isolates (MOI of 0.01) in response to treatment with aprotinin or SERPINA1/alpha-1 antitrypsin as determined 48 h post-infection. FFM1, 1/Human/2020/Frankfurt; FFM2, 2/Human/2020/Frankfurt; FFM6, 6/Human/2020/Frankfurt.

Table 8.1. Aprotinin concentrations that reduce SARS-CoV-2-induced cytopathogenic effect (CPE) formation, SARS-CoV-2 spike (S) levels, and SARS-CoV-2-induced caspase 3/7 activation by 50% (IC50) as determined in Caco2 cells infected with different SARS-CoV-2 isolates (MOI of 0.01) 48 h post-infection.

	IC_{50} (µM)		
	FFM1	**FFM2**	**FFM6**
CPE formation	1.03 ± 0.07	0.81 ± 0.07	0.92 ± 0.03
S levels	0.79 ± 0.15	1.04 ± 0.21	1.65 ± 0.30
Caspase 3/7 activation	0.41 ± 0.25	0.32 ± 0.09	0.73 ± 0.40

Both aprotinin and SERPINA1/alpha-1 antitrypsin are trypsin inhibitors (Gettins, 2002; Solun & Shoenfeld, 2020). To verify the integrity of the used protease inhibitor samples, we tested their capacity to antagonize trypsin and enable Caco2 and A549 cell adhesion. The results confirmed that both protease inhibitors are active (Figure S8.2). Taken together, these findings indicate differences in the protease inhibitor spectrum of aprotinin and SERPINA1/alpha-1 antitrypsin that result in different effects on SARS-CoV-2 replication.

8.3.2. Quantification of the Antiviral Effects of Aprotinin by Measuring SARS-CoV-2-Induced Caspase 3/7 Activation

Different viruses, including SARS-CoV-2, have been shown to induce caspase 3 activation (Li, S. et al., 2020; Michaelis et al., 2007; Ren et al., 2020; Xu, M. et al., 2016), and virus-induced caspase 3 activation has been used as read-out in assays that quantify the antiviral effects of

drug candidates (Xu, M. et al., 2016). Hence, we used the Caspase-Glo® 3/7 Assay (Promega) as an additional quantitative method to determine the anti-SARS-CoV-2 activity of aprotinin. The results confirmed those obtained by CPE formation and S expression resulting in similar IC50 values (Figure 8.2, Table 8.1).

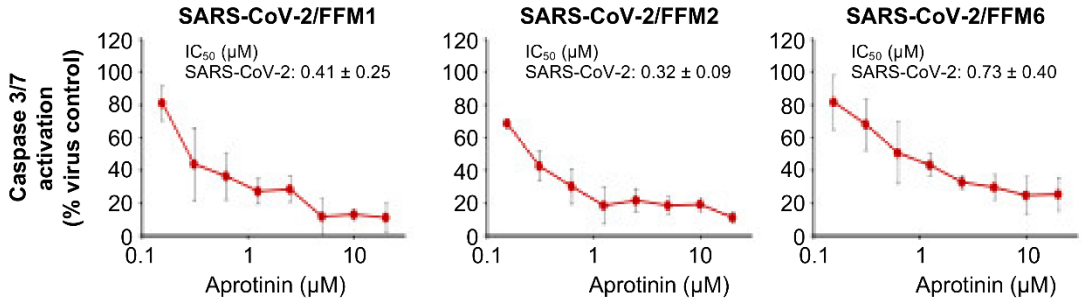


Figure 8.2. Effects of aprotinin on SARS-CoV-2-induced caspase 3/7 activation. Caspase 3/7 activity was determined in Caco2 cells infected with different SARS-CoV-2 isolates (MOI of 0.01) 48 h post-infection.

8.3.3. Aprotinin Inhibits Virus Entry

Protease inhibitors were suggested to interfere with SARS-CoV-2 replication predominantly as entry inhibitors that prevent S cleavage and activation (Shen, L. W. et al., 2017). In agreement, aprotinin addition after a one-hour adsorption period did not significantly interfere with SARS-CoV-2 replication in one round of a replication assay, in which virus titers were determined 8 h post-infection with an MOI of 0.1 (Figure 8.3A). In contrast, remdesivir, which was anticipated to interfere with the replication of the viral genome, inhibited SARS-CoV-2 replication when added post-infection (Figure 8.3A).

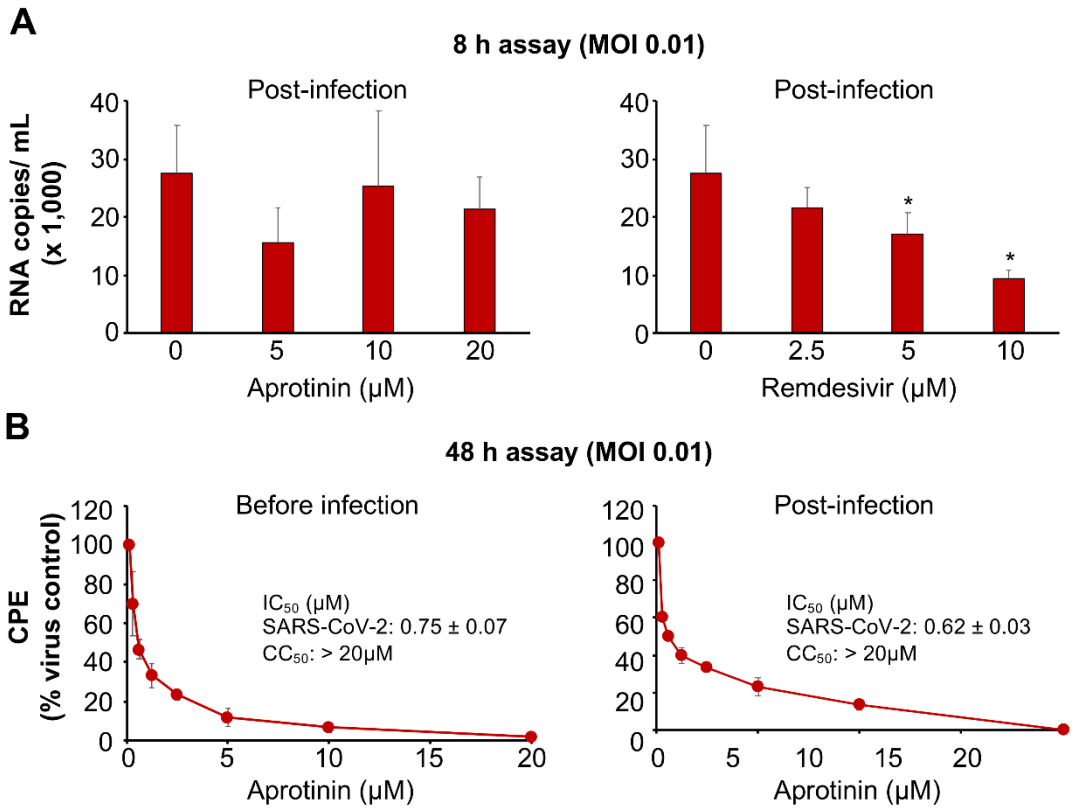


Figure 8.3. Anti-SARS-CoV-2 effects of aprotinin when administered post-infection. For post-infection experiments, the cells were incubated with the virus for a one-hour adsorption period. Then, the cells were washed three times in PBS prior to the addition of the drug. (**A**) The effects of aprotinin and the RNA polymerase inhibitor remdesivir (a positive control drug that interferes with virus replication after virus entry) on virus replication as determined by qPCR in SARS-CoV-2/FFM1 (MOI of 0.1)-infected Caco2 cells 8 h post-infection (after approximately one round of replication). * $p < 0.05$ as determined by one-way ANOVA and Dunnett's multiple comparison test. (**B**) The effects of aprotinin on cytopathogenic effect (CPE) formation in SARS-CoV-2/FFM1 (MOI of 0.01)-infected Caco2 cells were determined 48 h post-infection.

8.3.4. Aprotinin May Interfere with SARS-CoV-2-Mediated Downregulation of Host Cell Protease Inhibitors

Notably, aprotinin exerted similar anti-SARS-CoV-2 effects when added before or after infection of Caco2 cells with a lower MOI (0.01) in a 48 h assay (Figure 8.3B). In this format, aprotinin probably inhibits the later rounds of SARS-CoV-2 replication, but other mechanisms may also contribute.

Host cell protease inhibitors interfere with the activity of proteases such as TMPRSS2 (Esumi et al., 2015; Straus et al., 2020) that mediate SARS-CoV-2 cell entry by cleaving and activating the viral S protein (Hoffmann et al., 2020; Hoffmann et al., 2020a; Yamamoto et al., 2020). An analysis of the effect of SARS-CoV-2 infection on host cell protease inhibitors using proteomics data from SARS-CoV-2-infected Caco2 cells (Bojkova et al., 2020) showed that the endogenous protease inhibitors SPINT1 (Kunitz-type protease inhibitor 1), SPINT2 (Kunitz-type protease inhibitor 2), and SERPINA1 (alpha-1-antitrypsin) are present at lower levels in SARS-CoV-2-infected cells than in non-infected control cells 24 h post-infection (Figure 8.4A). Translatome data from the same dataset (Bojkova et al., 2020) indicated that the translation of SERPINA1 and SPINT2 (but not that of SPINT1) is also reduced in SARS-CoV-2-infected cells (Figure 8.4B). Hence, SARS-CoV-2 infection results in the downregulation of endogenous protease inhibitors, which may support SARS-CoV-2 replication. Thus, compensation for downregulated endogenous protease inhibitors may contribute to the antiviral effects of aprotinin.

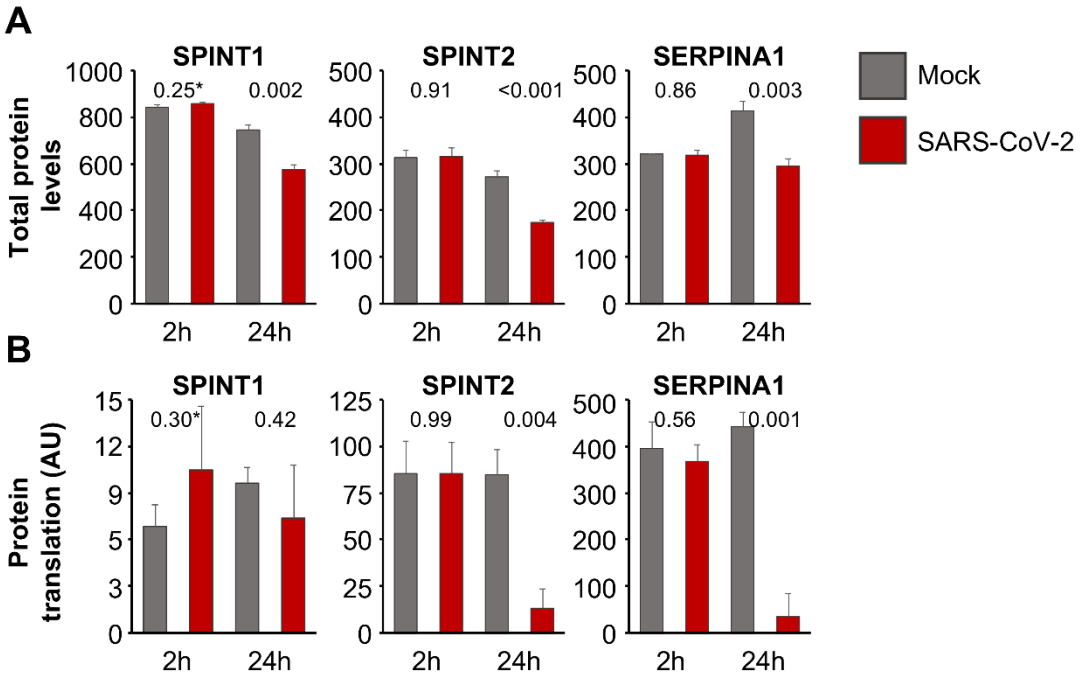


Figure 8.4. Regulation of host cell protease inhibitors in SARS-CoV-2-infected cells. **(A)** Total protein levels based on a publicly available proteomics dataset (Bojkova et al., 2020), indicating cellular levels of endogenous protease inhibitors in SARS-CoV-2 (MOI of 1)-infected Caco2 cells 2 h and 24 h post-infection. Data were normalized using summed intensity normalization for sample loading, followed by internal reference scaling and trimmed mean of M normalization. * p-values as determined using a two-sided Student's t-test. **(B)** Mean protein translation of endogenous protease inhibitors in

arbitrary units (AU) (normalized and corrected summed peptide spectrum matches (PSMs) were averaged) in SARS-CoV-2 (MOI of 1)-infected Caco2 cells 2 h and 24 h post-infection based on a publicly available translatome dataset (Bojkova et al., 2020). * p-values as determined using a two-sided Student's t-test.

8.3.5. Aprotinin Exerts Anti-SARS-CoV-2 Activity in Air–Liquid Interface (ALI) Cultures from Primary Bronchial Epithelial Cells

We also investigated the effects of aprotinin in SARS-CoV-2-infected air–liquid interface (ALI) cultures from primary bronchial epithelial cells. A targeted proteomics assay demonstrated that aprotinin 20 µM suppressed the expression of the SARS-CoV-2 proteins N (nucleocapsid protein) and M (membrane protein) in SARS-CoV-2-infected ALI cultures (Figure 8.5A, Table S8.1). The results for N were confirmed by Western blots in the ALI cultures infected with SARS-CoV-2/FFM7 (Figure 8.5B). SARS-CoV-2/FFM7 (G lineage) is an alternative isolate derived from a patient from Israel (Toptan et al., 2020). Aprotinin also suppressed SARS-CoV-2 S expression in SARS-CoV-2/FFM7-infected Calu-3 lung adenocarcinoma cells (Figure S8.3).

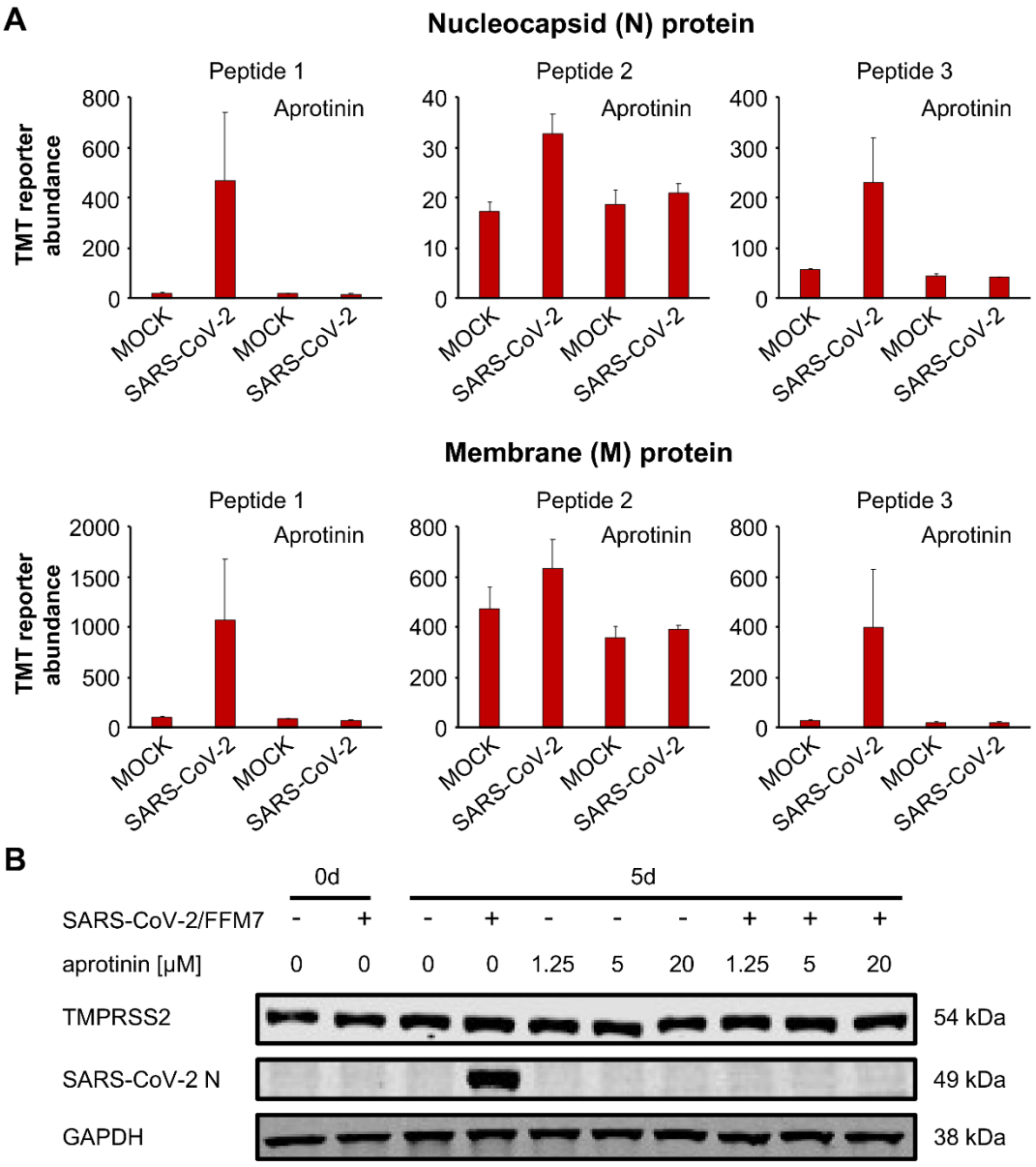


Figure 8.5. Antiviral effects of aprotinin in SARS-CoV-2-infected air–liquid interface (ALI) cultures from primary bronchial epithelial cells. (**A**) Abundance of the SARS-CoV-2 proteins N (nucleocapsid) and M (membrane) in primary bronchial epithelial cell ALI cultures infected with SARS-CoV-2/FFM1 (MOI of 1) in the presence or absence of aprotinin (20 µM) as determined 5 days post-infection by multiplexed mass spectrometry analysis using acquisition targeting of previously identified viral peptides modified with TMTpro. The detailed data are presented in Table S8.1. (**B**) Western blots indicating cellular SARS-CoV-2 N and TMPRSS2 levels in primary bronchial epithelial cell ALI cultures infected with SARS-CoV-2/7/Human/2020/Frankfurt (FFM7) (MOI of 1) in the presence or absence of aprotinin as detected 5 days post infection. GAPDH was served as the loading control. Uncropped Western blots are shown in Figure S8.4.

8.4. Discussion

Herein, we showed that aprotinin inhibits SARS-CoV-2 replication predominantly as an entry inhibitor, probably via interfering with SARS-CoV-2 S activation by TMPRSS2. Notably, SERPINA1/alpha-1 antitrypsin, which is available as a pharmaceutical preparation for the treatment of alpha-1 antitrypsin deficiency (Strnad et al., 2020), did not inhibit SARS-CoV-2 replication in the same concentration range. Further investigations will have to elucidate the differences between aprotinin and SERPINA1/alpha-1 antitrypsin that are responsible for the discrepancy in anti-SARS-CoV-2 activity. Notably, SERPINA1/alpha-1 antitrypsin was shown to inhibit TMPRSS2 in an enzymatic assay and is suggested as an antiviral treatment for COVID-19 (Azouz et al., 2020). A clinical trial testing SERPINA1/alpha-1 antitrypsin for the treatment of COVID-19 has recently been started (ClinicalTrials.gov Identifier: NCT04385836). Based on our data, however, SERPINA1/alpha-1 antitrypsin is not expected to exert direct antiviral effects in COVID-19 patients. Our findings also indicate that antiviral therapy candidates should be tested for their effects on complete replication-competent viruses in permissive cells.

Aprotinin exerted anti-SARS-CoV-2 effects in three cell culture models (Caco2, Calu-3, and air–liquid interface cultures from primary bronchial epithelial cells) and against three SARS-CoV-2 strains (FFM1, FFM2, FFM6, and FFM7). Notably, another study became available during the revision of our manuscript that detected anti-SARS-CoV-2 activity of aprotinin in Calu-3 cells (Bestle et al., 2020). Our findings are also in agreement with studies that reported other TMPRSS2 inhibitors to inhibit SARS-CoV-2 entry and replication (Choudhary & Silakari, 2020; Hoffmann et al., 2020a; Yamamoto et al., 2020). In addition, furin has been shown to cleave and activate SARS-CoV-2 S and furin inhibitors have been demonstrated to exert anti-SARS-CoV-2 effects (Cheng, Y. W. et al., 2020).

Endogenous protease inhibitors may interfere with the activation of virus surface proteins such as S by host cell proteases (Esumi et al., 2015; Hoffmann et al., 2020; Hoffmann et al., 2020a; Straus et al., 2020; Yamamoto et al., 2020). Our analysis of proteomics and translatome data from SARS-CoV-2-infected Caco2 cells (Bojkova et al., 2020) revealed a downregulation of endogenous protease inhibitors in response to SARS-CoV-2 infection, which may contribute to efficient SARS-CoV-2 replication. In addition to entry inhibition, compensation for downregulated endogenous proteases may, hence, further contribute to the antiviral activity of aprotinin during later rounds of SARS-CoV-2 replication.

The clinical potency of aprotinin is typically measured in kallikrein inhibitor units (KIUs) (Levy et al., 1994; Zhirnov et al., 2011). Therapeutic aprotinin plasma levels were described to reach 147 ± 61 KIU/mL after the administration of 1,000,000 KIU (Levy et al., 1994). Moreover, an aerosol preparation of aprotinin, which is likely to result in increased local aprotinin concentrations in the lung, is approved for the treatment of influenza in Russia (Zhirnov et al., 2011). The aprotinin IC50 values for SARS-CoV-2-induced CPE formation, S expression, and apoptosis induction ranged from 0.32 to 1.65 µM, which is equivalent to 4.0 KIU and 20.6 KIU, respectively. Hence, aprotinin interferes with SARS-CoV-2 infection in therapeutically achievable concentrations.

Aprotinin exerts pro- and antithrombotic effects by balancing fibrinolysis and thrombus formation and is approved for the prevention of blood loss during surgery. It interferes with the fibrinolysis of established thrombi by plasmin, but also inhibits contact-activated thrombus formation in the blood stream (Dietrich, 1996; Kuitunen et al., 2005; Solun & Shoenfeld, 2020; Terrell et al., 1996). Late-stage, severe COVID-19 disease has been associated with disseminated intravascular coagulation and thrombosis (COVID-19-related coagulopathy) (Marchandot et al., 2020). Based on the available data, it is not clear whether aprotinin may exert pro- or antithrombotic effects in patients suffering from COVID-19-related coagulopathy. Thus, aprotinin would have to be considered with care for such patients.

However, antiviral treatment may anyway be of limited impact in late-stage COVID-19 disease, during which, damage is anticipated to be largely caused by immunopathology and not by virus replication (Kuitunen et al., 2005; Lega et al., 2020; Polycarpou et al., 2020). Hence, the main potential of antiviral drugs may lie in the early treatment of COVID-19 patients to suppress virus replication and, through this, to prevent COVID-19 progression into a severe, life-threatening disease. Local aprotinin therapy of the airways and the lungs using an aerosol, which is clinically approved in Russia and has been reported to be very well tolerated in influenza patients (Zhirnov et al., 2011), may have particular potential as such an antiviral treatment for early stage COVID-19 disease. Notably, aprotinin may additionally prevent the very early stages of lung injury by inhibition of matrix metalloproteinases and, in turn, of the cytokine storm that eventually results in severe, systemic COVID-19 disease (Solun & Shoenfeld, 2020).

8.5. Conclusions

In conclusion, therapeutic aprotinin concentrations inhibit SARS-CoV-2 replication as entry inhibitors and by compensating for downregulated cellular protease inhibitors during later replication cycles. Local treatment of the respiratory tract using an aprotinin aerosol, which is approved in Russia for the treatment of influenza (Zhirnov et al., 2011), may be a particularly promising strategy to suppress virus replication and lung injury early and to prevent COVID-19 progression into a severe, systemic disease.

Chapter 9: Discussion

The manuscripts presented in this thesis have each demonstrated the potential for publicly available data to be used to contribute to new knowledge in different areas of biological interest. Here, such data have been used to investigate mechanisms of drug resistance in cancer, as well as for the characterisation of the novel coronavirus SARS-CoV-2 and COVID-19 to produce novel findings with scientific significance and clinical relevance. The methods we have used can also be easily extrapolated to other areas of research. In this chapter, the context in which each manuscript should be considered is discussed, the implications of the work are presented, and perspectives on how these findings and similar works may be expanded upon in the future are proposed.

9.1. Investigating drug resistance in cancer

Drug resistance is one of the major challenges faced in the treatment of cancer. It is estimated that around 90% of cancer treatment failures are due to drug resistance (Bukowski et al., 2020; Mansoori et al., 2017; Wang, X. et al., 2019), which can be the result of mechanisms such as drug efflux, reduced drug uptake, drug detoxification, alteration of the drug target, DNA damage repair or resistance to apoptosis. With the use of several large-scale pharmacogenomic databases, we have been able to examine a number of these factors and generate findings which could be of critical importance when determining patient treatment protocols.

Ovarian cancer is the seventh most common type of cancer diagnosed globally, and it is the leading cause of mortality among all gynaecological cancers (Coburn et al., 2017). Late diagnosis of ovarian cancer is a significant issue, with over 75% of cases identified at an advanced stage due to the early stages of disease being asymptomatic (Doubeni et al., 2016). Moreover, while most cases respond to platinum-based chemotherapeutic agents initially, the majority become refractory to treatment (Damia & Broggini, 2019). It is therefore of critical importance that we develop a better understanding of the mechanisms driving resistance in order to find ways to block these pathways or to find alternative treatments.

Given its role as a regulator of diverse cellular signalling pathways and its implication in numerous diseases (including cancer, diabetes, cardiovascular disease, neurodegenerative diseases, polycystic ovarian syndrome (Greig & Nixon, 2014)), the impact of differential PEA-

15 expression on the expression of other genes has been widely investigated (Funke et al., 2013; Shin et al., 2015; Sulzmaier, F. J. et al., 2012). The significance of PEA-15 phosphorylation status has also been studied in relation to a number of biological processes, such as the serum starvation response (Quintavalle et al., 2014), expression of β-catenin (a cell adhesion protein and transducer of Wnt signalling) and Ki-67 (a marker of cell proliferation) (Lee, J. et al., 2012) and expression of superior cervical ganglion 10 (SCG10)-like protein (SCLIP, a microtubule destabilising protein) (Xie et al., 2013).

In chapter 2, we investigated the impact of PEA-15 phosphorylation status on the sensitivity of SKOV-3 ovarian cancer cells to the platinum-based chemotherapeutic agent cisplatin. Transcriptome-wide expression analysis of SKOV-3 cells transfected with PEA-15-AA, PEA-15-DD or PEA-15-EV enabled the identification of genes which were differentially expressed depending on PEA-15 phosphorylation status, and TCGA data was then used to investigate the impact of expression of these differentially regulated genes on survival of cisplatin-treated patients. This enabled us to infer mechanisms by which PEA-15 phosphorylation status (and potentially the associated ERK1/2 sequestration) may influence cisplatin sensitivity. Our study appears to be the only one to date which has been able to identify a correlation between PEA-15 phosphorylation status and sensitivity of ovarian cancer cells to cisplatin, enabling us to consider it a potential biomarker to predict treatment outcomes. This is particularly important, as cisplatin is widely used as a first-line therapy for ovarian cancer (Helm & States, 2009; Pokhriyal et al., 2019).

Using the TCGA provided us with a unique opportunity to include data from a large cohort of cisplatin-treated patients complete with gene expression and survival outcome statistics, which may have been otherwise time-consuming and costly to obtain. Another advantage to our approach was that it eliminated bias towards any specific pathway (i.e. we did not set out to analyse perturbation of the expression of any particular genes in response to PEA-15 phosphorylation). We were able to identify that nine of the 11 genes which were significantly differentially expressed in untreated SKOV-3-AA cells compared with SKOV-3-DD cells were also associated with corresponding survival outcomes in cisplatin-treated TCGA patients (i.e. genes which were upregulated were also associated with improved survival when their expression was high, while genes which were downregulated were associated with improved survival when their expression was low). Further analysis of these differentially expressed genes could reveal important pathways affecting the development of resistance to cisplatin in ovarian cancers, which could in turn have therapeutic implications.

Drug resistance is also an important issue in the treatment of acute lymphoblastic leukaemia (ALL). ALL involves the abnormal proliferation of a clonal population of lymphoid precursor cells in the bone marrow. While 75% of cases arise from the malignant transformation of B-cell progenitor cells, the remainder arise from T-cell progenitors (Terwilliger & Abdul-Hay, 2017). Outcomes appear to worsen with increasing age, with one clinical trial finding five-year survival rates of 41% in patients under 55 and 21% in patients over 55 (Sive et al., 2012). It is well recognised that nelarabine, the prodrug of arabinosylguanine (Ara-G), is effective in treating T-ALL but less effective in B-ALL (Verhoef & Fridland, 1985), and a number of studies have found greater accumulation of Ara-G in T-ALL cells compared with B-ALL cells (Gandhi et al., 1998; Rodriguez Jr et al., 2003). However, the reason for this difference has been previously unclear. Given the risks associated with nelarabine treatment (including, as previously highlighted, potentially lethal neurotoxicity), identifying a biomarker which could aid in the stratification of patients likely to benefit from the treatment has substantial clinical importance.

In the paper presented in chapter 3, we were able to shed light on the possible mechanism driving this difference in nelarabine response. We revealed how differential regulation of SAMHD1, by virtue of its ability to detoxify certain nucleoside analogue drugs, was able to influence differential drug response in T-acute lymphoblastic leukaemia (T-ALL) compared with B-acute lymphoblastic leukaemia (B-ALL). To achieve this, we combined results from wet-lab investigations with analysis of pharmacogenomics data from the GDSC, CCLE and CTRP databases, as well as transcriptomic data from a cohort of ALL patients. We were not only able to identify that *SAMHD1* expression was lower in T-ALL cells compared with B-ALL cells, but we also showed that methylation of the *SAMHD1* promoter was a likely factor influencing the differential expression observed. Our results again demonstrated that cell line pharmacogenomic databases can be used for the validation of *in vitro* investigations, and that they can be used to generate findings which could ultimately inform treatment protocols.

It is acknowledged that the efficacy of nucleoside analogue drugs such as nelarabine depends in part on the accumulation of the drug within the cell, which enables the drug to out-compete deoxynucleotide triphosphates for incorporation into DNA. Our results give specific rationale for stratifying patients based on their *SAMHD1* expression levels to determine in which cases nelarabine would be most likely to be effective. For example, a recent study investigated use of nelarabine with fludarabine (which inhibits ribonucleotide reductase, thereby reducing the accumulation of competing deoxyguanosine triphosphates) and

etoposide (a topoisomerase II inhibitor which arrests the cell cycle in G1/S phase, when DNA synthesis inhibitors are most effective) (Kumamoto et al., 2020). Combination of this approach with stratification by *SAMHD1* expression level could further enhance treatment efficacy when using nelarabine. This could enable clinicians to identify patients who may better respond to an alternative therapy or to a nucleoside analogue which is not a substrate for SAMHD1.

9.2. Further investigations into SAMHD1 and its impact on cancer patient survival

As discussed, a major physiological role for SAMHD1 has been described as its cell cycle-dependent regulation of dNTP pools through its deoxynucleoside triphosphohydrolase activity in G1 phase, which acts in opposition to the de novo dNTP synthesis catalysed by ribonucleotide reductase (RNR) in S phase (Franzolin et al., 2013). This activity has been proposed to contribute to the maintenance of genome stability, as imbalanced dNTP levels lead to reduced replicative fidelity and impaired mismatch repair (Pai & Kearsey, 2017). In addition to its G1 phase activity, SAMHD1 has also been shown to play a direct role in mediating DNA damage repair during S phase, where it localises to double strand breaks and promotes DNA end resection via its interaction with C-terminal binding protein-interacting protein (CtIP) and meiotic recombination 11 (MRE11), thereby facilitating homologous recombination (Daddacha et al., 2017a). Notably, methylation-dependent downregulation of SAMHD1 and mutations which affect its tumour-suppressive dNTPase activity have been associated with the development of a number of different types of cancer (Clifford et al., 2014; de Silva et al., 2014; Johansson, P. et al., 2018; Rentoft et al., 2016), while overexpression of *SAMHD1* has been associated with reduced proliferation of lung cancer A549 cells (Wang, J. L. et al., 2014). Findings such as these have contributed to the perception of SAMHD1 as a tumour suppressor (Coggins et al., 2020; Herold, N. et al., 2017; Kohnken et al., 2015; Mauney & Hollis, 2018).

Certain aspects of SAMHD1 function also appear to oppose its potential tumour-suppressive activity. For example, the role of SAMHD1 in genome maintenance is widely accepted to present significant challenges in the treatment of cancer. As discussed, SAMHD1 can detoxify nucleoside analogue drugs with broad substrate specificity (Herold, N., Rudd, Ljungblad et al., 2017; Schneider et al., 2017) separate from its documented activity in enhancing the efficacy of nucleoside analogue reverse transcriptase inhibitors via depletion of competing dNTPs (Amie et al., 2013; Coggins et al., 2020). Moreover, the involvement of SAMHD1 in mediating DNA damage repair could also represent another mechanism by which SAMDH1

opposes the activity of anti-cancer drugs, such as DNA damaging agents (Daddacha et al., 2017b).

Post-translational modifications of SAMHD1 could also influence whether it plays a tumour-promoting role. For example, it has been suggested that acetylation at K405 may lead to enhanced dNTPase activity and depletion of dNTP pools in G1 phase, thereby promoting transition to S phase and thus increasing cancer cell proliferation (Coggins et al., 2020; Lee, E. J. et al., 2017). Meanwhile, reports into the effect of SAMHD1 phosphorylation are conflicting – some have found that phosphorylation at T592 destabilises the tetramer and diminishes its dNTPase activity (Tang, C. et al., 2015; Yan, J. et al., 2015), while others have found phosphorylation to have no impact on its catalytic activity (Welbourn et al., 2013; White et al., 2013). Interestingly, however, it has been shown that phosphorylation status may play a critical role in the DNA damage repair function of SAMHD1 independently of its dNTPase activity – phosphomimetic T592E mutant SAMHD1 was found to promote replication fork progression and to restore DNA end resection in HEK293T in which *SAMHD1* had been silenced using short hairpin RNA (shRNA) (Coquel et al., 2018). Elsewhere, it has been shown that oxidation of SAMHD1 results in the formation of intrachain disulphide bonds, inhibition of tetramerization and a reduction in its catalytic activity (Mauney et al., 2017), which could have significant implications given the redox dysregulation present in many cancers (Acharya et al., 2010; Perillo et al., 2020; Purohit et al., 2019).

Given how differences in *SAMHD1* expression and regulation, mutation status and exposure to nucleoside analogue drugs contribute to a complex role for this triphosphohydrolase in influencing cancer progression, we were interested to explore the evidence further. In chapter 4, we have shown that SAMHD1 is widely regarded as a tumour suppressor that can detoxify certain nucleoside analogue drugs (Chen, Z. et al., 2021; Herold, N. et al., 2017). On the other hand, we also highlighted a number of studies which noted a potential tumour-promoting role for SAMHD1 independent of its drug-inactivating function (Kodigepalli et al., 2018; Shang, Z. et al., 2018; Xagoraris et al., 2021; Yang, C. A. et al., 2017). We related these findings to our own investigation into the role of SAMHD1 in cancer, in which we took an unbiased approach in assessing whether it may be considered a tumour suppressor or tumour promoter based on publicly available cancer patient data.

The large volume of data available for the hundreds of cancer patients included the TCGA and TARGET studies enabled us to perform multiple analyses of *SAMHD1* expression and its regulation in relation to patient survival. We found substantial differences in the survival

outcomes of patients expressing high or low *SAMHD1*, depending on cancer type. While this could reflect differences in treatment regimes for different cancers (for example, lymphoid cancers are often treated with nucleoside analogues), it could also reflect tissue-specific SAMHD1 activity which has yet to be fully understood. The differential regulation of SAMHD1 by promoter methylation and miRNAs we observed in patients with different cancer types could in part underly these differences and warrants further investigation. Moreover, the differences observed in outcomes for male and female patients expressing high or low *SAMHD1* gives yet further support for an individualised approach to cancer treatment, since factors such as sex-specific gene expression or differing hormone levels may affect the impact of SAMHD1 on patient survival and/or response to therapy.

Our investigations into the impact of *SAMHD1* mutations in TCGA patients also further challenged the notion that SAMHD1 may be considered a tumour suppressor. Despite previous evidence which suggests that loss-of-function mutations in *SAMHD1* or methylation-dependent downregulation of SAMHD1 is associated with the development of various types of cancer, our finding that expression of mutant *SAMHD1* is associated with improved survival outcomes compared with expression of wildtype *SAMHD1* does not support this. We supported this conclusion by showing that expression of mutant *TP53* (a well-known frequently mutated tumour suppressor (Kim, Y. A. et al., 2017; Lawrence et al., 2013; Levine, 2020; Wang, X. et al., 2020)) among the same cohort of patients was associated with reduced overall survival. This study, alongside our work in chapter 3, has therefore highlighted the need for a holistic approach to the investigation of the role of SAMHD1 and its diverse functions, as well as the need for careful clinical decision-making when considering *SAMHD1* expression and/or mutation status as a factor influencing cancer treatment.

9.3. Characterisation of the novel coronavirus SARS-CoV-2

While complete conservation of an amino acid position is generally simple to identify and easy to interpret, positions with functional significance which are conserved within protein families but vary between them are less easily distinguished. Building on early sequence, phylogeny and structure-based approaches to determining functionally significant differentially conserved positions (Casari et al., 1995; del Sol Mesa et al., 2003; Hannenhalli & Russell, 2000; Lichtarge et al., 1996; Pazos & Sternberg, 2004; Pazos et al., 2006), a more recent principal component analysis (PCA)-like approach (‘multiple correspondence

analysis') based on the simultaneous identification of protein sub-families and 'specificity determining positions' (SDPs) was described. This study revealed how the presence of SDPs in and around protein and ligand binding sites closely reflected substrate specificity and, by extension, the functional evolution of a protein (Rausell et al., 2010). The methodology of this study was employed in recent research which involved the analysis of *Ebolavirus* genomes to explain the differences in pathogenicity between the related viruses (Martell et al., 2019; Pappalardo et al., 2016). We employed a similar approach in the work presented in chapter 5 – by combining analysis of multiple sequence alignments of publicly available SARS-CoV and SARS-CoV-2 sequences with manual analysis of publicly available structural data for each viral protein, we have identified how amino acid substitutions that consistently differed between SARS-CoV and SARS-CoV-2 (at differentially conserved positions, or DCPs) could affect the differences observed in their cell tropism, receptor binding affinity and sensitivity to protease inhibitors.

Differential conservation of selected functional residues in a protein arise due to selective pressures necessitating functional adaptation to a novel environment or host. We have highlighted the SARS-CoV-2 viral proteins which may be under the greatest selective pressure and therefore harbour the highest density of DCPs. Notably, we observed enrichment of DCPs in the Spike glycoprotein – two of the vaccines currently approved for use in the UK are mRNA-based vaccines which encode the Spike glycoprotein (Pfizer (Polack et al., 2020), Moderna (Baden et al., 2021)), while the third consists of an adenoviral vector containing the Spike glycoprotein gene (AstraZeneca (Voysey et al., 2020)). Variants arising within this viral protein have therefore been of concern due to the potential for a reduction in the efficacy of these vaccines. Conversely, our results revealed relatively few DCPs present in the envelope (E) protein, which is also located on the surface of the SARS-CoV-2 virus and could therefore present an alternative vaccine target. The notion that the E protein could be considered a potential vaccine target is not without precedent – another study also identified a lower frequency of mutations in this viral protein compared to the Spike glycoprotein, and also suggested that the introduction of inactivating mutations in the ion channel structure of the E protein, which plays a role in viral pathogenicity, could provide the basis for the development of an inactivated or live attenuated vaccine (Sarkar & Saha, 2020). Importantly, the lower mutability of this viral protein could suggest a low risk of reversion to a pathogenic strain.

It has been suggested that SARS-CoV and SARS-CoV-2 differ in the efficiency with which they suppress type I interferon (IFN) responses (Konno et al., 2020; Xia, H. et al., 2020). We noted

a significant number of DCPs in SARS-CoV-2 proteins (e.g. orf6) which have been previously found to play a role in interferon antagonism in the equivalent proteins in SARS-CoV (Totura & Baric, 2012) and in SARS-CoV-2 (Yuen et al., 2020). Notably, a number of studies have suggested significant clinical implications for interferon antagonism in SARS-CoV-2 – one study showed significantly impaired type I interferon responses in the most severe cases of COVID-19, with IFN-β mRNA and protein levels undetectable in all infected patients (Hadjadj et al., 2020). Moreover, a randomised double-blind placebo-controlled phase 2 trial showed that COVID-19 patients treated with nebulised IFN-β-1a had a higher probability of improvement and faster recovery time than patients who received a placebo (Monk et al., 2020). Our results could therefore prompt further investigation of DCP-enriched interferon-inhibiting proteins in SARS-CoV-2, which could reveal the mechanisms driving the enhanced interferon antagonism observed in critically ill patients.

Our analysis of DCP-enriched regions of the SARS-CoV-2 proteins has also complemented our *in vitro* findings and those of other groups. We have identified DCPs which may have a specific effect on the infectivity of the SARS-CoV-2 virus compared with SARS-CoV, such as those in proximity to the Spike glycoprotein RBD which may contribute to its enhanced affinity for ACE2. As discussed, targeting this interaction has been the focus of many efforts to develop novel treatments and vaccines for COVID-19, and identifying compounds which sufficiently overcome high affinity Spike/ACE2 binding has been a significant challenge. Monoclonal antibodies purified from the convalescent plasma of SARS-CoV-2 patients which target the Spike RBD have demonstrated efficacy in blocking interaction with ACE2 *in vitro* (Chen, X. et al., 2020) and in mouse models (Wu, Y. et al., 2020), and several have entered phase 2/3 trials (Tuccori et al., 2020). Meanwhile, one group proposed that multiple high-affinity monoclonal antibodies directed against multiple epitopes of SARS-CoV-2 Spike protein could more effectively inhibit viral adhesion and entry than a single antibody against a single epitope (Khatri et al., 2020). In an alternative approach, structural data of Spike protein RBD/ACE2 interactions have been used as a template for the computer-aided design of novel peptide or miniprotein inhibitors, which can be optimised for high-affinity binding (Cao et al., 2020; Huang, X. et al., 2020). The DCPs we identified and which were also shown in adaptation studies to affect Spike protein affinity for ACE2 (Wan et al., 2020) could therefore be useful for reference in the design of novel antiviral peptides. Additionally, we found several DCPs in regions important for cleavage of proteases, which complemented our *in vitro* findings suggesting different responses to protease inhibition in SARS-CoV-2

compared to SARS-CoV. We further investigated the impact of protease inhibition on SARS-CoV-2 infectivity in chapter 8, discussed below.

This study has demonstrated how sequencing data made publicly available to the scientific community can be used to make significant inferences regarding the nature of a novel virus. In the context of a global pandemic in which the rapid identification of therapeutic and vaccine targets is paramount, the ability to narrow the search for such targets using 'big data' approaches represents a critical development in modern drug discovery.

9.4. Determining mechanisms driving severe COVID-19 and identifying potential therapeutic interventions

Several factors have been associated with an increased risk of developing severe COVID-19. These include demographic characteristics, such as age, being male and being from black, Asian and ethnic minority backgrounds (Kelada et al., 2020; Liu, Y. et al., 2020; Patel et al., 2020; Peckham et al., 2020; Sze et al., 2020), as well as health conditions such as hypertension, diabetes, cardiovascular disease, cerebrovascular disease, respiratory disease, renal disease and cancer (Wolff et al., 2021; Zhou, F. et al., 2020). In the papers presented in chapters 5 and 6, we were able to integrate the analysis of a number of publicly available transcriptomic and proteomics datasets to further investigate the role of a number of these risk factors in the development of severe COVID-19.

The mechanisms linking severe COVID-19 to its known risk factors are incompletely understood, but current evidence frequently points to altered immune function predisposing individuals to more severe disease. In terms of the demographic risk factors, it has been noted that studies of other viral infections have demonstrated hormonal differences between males and females affecting immune responses, such that females are capable of mounting a more effective antiviral immune response compared with males. For example, oestrogen has been found to regulate the activity of numerous cells of both the innate and adaptive immune systems, inducing cytokine and chemokine signalling via regulation of neutrophil activity, inducing dendritic cell maturation and enhancing T and B cell activity (Khan & Ansar Ahmed, 2016; Klein & Flanagan, 2016). Specific differences in response to SARS-CoV-2 infection have also been observed, such as higher expression of ACE2 in males (Sama et al., 2020) and higher levels of terminally differentiated and activated T cells in female COVID-19 patients compared with males (Takahashi et al., 2020). Meanwhile, the link between severe COVID-19 and age is unclear, although it has been proposed that factors

such as age-related immune system, transcriptional and epigenetic changes may underlie the association (Mueller et al., 2020).

The mechanism by which hypertension is associated with increased risk of severe disease is also unclear, but it has also been proposed to relate to dysregulation of the immune system (Guzik et al., 2020; Kamyshnyi et al., 2020), for example, through increases in cluster of differentiation 8 (CD8)+ T cells which are both immunosenescent (i.e. have a reduced ability to recognise and response to antigens) and proinflammatory (i.e. capable of producing proinflammatory cytokines, such as IFN-γ and tumour necrosis factor (TNF)-α) (Youn et al., 2013). The involvement of ACE2 in the renin-angiotensin-aldosterone system (a key pathway regulating blood pressure) and the use of ACE inhibitors among hypertensive patients (which has been suggested to result in increased ACE2 expression) has also been highlighted, although no causal link with severe disease has yet been established (Danser et al., 2020; Kamyshnyi et al., 2020).

Greater susceptibility to infections as a result of dysregulation of the immune system in both type 1 and type 2 diabetes, especially when glucose levels are poorly controlled, has also been widely documented (Critchley et al., 2018; Graves & Kayal, 2008; Muller et al., 2005). One review noted that factors such as reduced activity of natural killer cells, variable neutrophil numbers, impaired chemotaxis and phagocytosis and chronic low-level inflammation resulting from increased activation of CD4+ and CD8+ T cells have been proposed to render patients with diabetes more susceptible to the 'cytokine storm' observed in patients with severe COVID-19 (Erener, 2020). They also suggested that an increased propensity towards hypercoagulation (e.g. due to altered plasma concentrations of coagulation proteins and metal ions and increased platelet activation) (Sobczak & Stewart, 2019) and reduced pulmonary function (Anandhalakshmi et al., 2013) could contribute towards the increased risk of severe disease observed among diabetic patients (Erener, 2020).

An increased risk of severe COVID-19 has also been observed in obese patients (Cai, Q. et al., 2020; Hamer et al., 2020; Kalligeros et al., 2020). This has been suggested to relate to the association between excess adipose tissue and chronic inflammation due to recruitment of macrophages producing cytokines such as TNF-α and interleukin 6 (IL-6) and a reduction in serum concentrations of the anti-inflammatory hormone adiponectin, as well as endothelial dysfunction and impaired fibrinolysis predisposing patients to thromboses (Lockhart & O'Rahilly, 2020).

For each of these risk factors, however, their frequent co-existence presents difficulty in interpreting their individual impact on propensity toward severe disease – for example, obesity is frequently linked with hypertension and diabetes (Pantalone et al., 2017). It has also been argued that age cannot be considered in isolation of the increased likelihood of possessing certain other known risk factors as an individual ages (Romero Starke et al., 2020). Similarly, while there is consensus that outcomes for COVID-19 patients with hypertension are worse than for patients without hypertension, it is unclear whether it can be considered an independent risk factor (Huang, S. et al., 2020; Pan et al., 2020; Wei et al., 2021).

There is a need to develop reliable methods to predict which COVID-19 patients are most likely to develop severe disease, which can aid in the stratification of patients for appropriate treatment. Several groups have used demographic characteristics and comorbidities (as well as clinical and laboratory observations) to make such predictions – for example, one group generated an equation to predict outcome severity which incorporated patient age, sex, dyspnoea, immunocompromise, chronic kidney disease and hyperlipidaemia (Ryan et al., 2020). Others performed Cox proportional hazards regression on clinical data from a cohort of 419 COVID-19 patients to identify independent risk factors (age, comorbidities, albumin levels and C-reactive protein (CRP) levels), which they then used to generate a nomogram and associated scoring system to aid in the prediction of severe disease based on presence of those risk factors (Dai et al., 2020). Similarly, another group developed a machine learning method trained on data from 105 potential risk factors which could generate predictions of disease severity based on data from up to eight input variables (Wongvibulsin et al., 2021).

As discussed previously (chapter 1.2.4), analyses of the transcriptomic and proteomic changes that occur as a result of infection with SARS-CoV-2 have been widely used to identify deregulated pathways which can be therapeutically targeted to reduce infection or progression to severe disease. Such data can also be used for the identification of biomarkers which could be used alongside demographic characteristics and clinical measurements that are routinely taken (such as CRP and albumin levels) to predict which patients may be at risk of severe disease. For example, one study involved quantification of proteins and metabolites in the sera of patients with severe COVID-19 as well as control subjects, which enabled the identification of 22 proteins and seven metabolites whose expression could be used to accurately predict severe cases (Shen, B. et al., 2020). Another study also used plasma proteomic profiling of COVID-19 patients to identify single proteins and combinations of proteins which could predict disease outcome (Shu, T. et al., 2020), while others identified 27 putative biomarkers which were differentially expressed depending on COVID-19 severity

(Messner et al., 2020). The results from each of these studies included both novel and known biomarkers of viral infection, such as CRP.

In the paper presented in chapter 6, we were able to integrate data from two independent sources (proteomics data from Bojkova et al. (Bojkova et al., 2020) and transcriptomics data from Blanco-Melo et al. (Blanco-Melo et al., 2020)) with data obtained from wet-lab investigations to provide strong evidence supporting our hypothesis that CD47 expression levels may be used as a biomarker for increased risk of developing severe COVID-19. Given the immense body of literature freely available online, we were also able to conduct extensive literature reviews incorporating large numbers of studies relating to CD47 and its role in a number of the known comorbid conditions of severe COVID-19. Our results demonstrate how the substantial quantity of transcriptomics and proteomics data that have been generated during the COVID-19 pandemic can be used to identify specific biomarkers of disease severity which can be experimentally validated and whose role can be explained through prior knowledge of their functions in other contexts.

An important feature of severe COVID-19 which has a significant impact on mortality is abnormal blood coagulation, or coagulopathy. Studies have reported several markers of coagulopathy which have been correlated with worse outcomes – for example, higher prothrombin time and elevated D-dimer (a fibrin degradation product and marker of thrombosis) has been observed in patients in intensive care (Huang, C. et al., 2020) and (in addition to increased activated partial thromboplastin time) in non-survivors compared with survivors (Tang, N. et al., 2020; Zhou, F. et al., 2020). Thrombocytopaenia has also been observed in a number of cases (Zhou, F. et al., 2020). In one study, it was suggested that the clinical presentations of 71.4% of non-survivors were consistent with a form of disseminated intravascular coagulation (DIC), whereas only 0.6% of survivors exhibited similar laboratory findings (Tang, N. et al., 2020). The DIC exhibited in COVID-19 patients has been suggested to differ from sepsis-related DIC, which would be typically marked by less extreme increases in D-dimer and more prominent thrombocytopaenia (Levi et al., 2020) – indeed, one study showed patients with more severe disease exhibited higher increases in platelet levels compared with less severely ill patients, which correlated with longer hospital stays (Qu et al., 2020).

In chapter 7, the mechanisms driving the disorder of blood coagulation observed in certain COVID-19 patients were investigated by combining analysis of data from four different sources – the GTEx database, the Gene Ontology Resource (Ashburner et al., 2000; Gene

Ontology Consortium, 2021), proteomics data from Bojkova et al. (Bojkova et al., 2020) and interactome data from Gordon et al. (Gordon et al., 2020). Not only does this study demonstrate, similar to the study presented in chapter 6, the usefulness of publicly available proteomics data for unveiling mechanisms of SARS-CoV-2 pathogenicity, but it has also demonstrated the value of pre-existing data resources for the investigation of a novel disease. With our simple approach, we were able to use basal gene expression data in combination with gene function annotations to identify possible links between expression of genes involved in blood coagulation, such as transferrin, and selected risk factors for COVID-19 (increasing age and being male).

In addition to understanding factors contributing to the development of severe COVID-19, the identification of drugs which may inhibit viral infection or prevent progression to severe disease is a critical issue, and one which has been the focus of many researchers using various approaches during the COVID-19 pandemic. As described in chapter 1.2.4, drug repurposing has been an attractive option, given the convenience of prior knowledge of drug targets and their likely effects as well as their safety and potency. The widespread clinical use of the steroid dexamethasone (shown to reduce deaths by a third in ventilated patients and a fifth in patients receiving oxygen therapy (Mahase, 2020)) is a notable example.

By revisiting the proteomics dataset released by Bojkova et al. (Bojkova et al., 2020), in chapter 8 we were able to identify SARS-CoV-2-induced downregulation of host cell protease inhibitors, which our *in vitro* studies suggested could be counteracted by the protease inhibitor aprotinin. As discussed, aprotinin has been previously shown to inhibit TMPRSS2 activity – for example, MDCK cells engineered to express TMPRSS2 were susceptible to H1N1 influenza virus infection in the absence of protease inhibitors (as TMPRSS2 was able to cleave the receptor binding glycoprotein haemagglutinin), and viral replication and spread was reduced by treatment with aprotinin (Böttcher et al., 2009). As it has been shown that the interaction between SARS-CoV-2 S protein and ACE2 requires its proteolytic cleavage and activation by host cell proteases such as TMPRSS2, our findings have significant implications for the possible benefit of repurposing aprotinin for the treatment COVID-19. As highlighted, an aerosol preparation of aprotinin has been shown to be effective against influenza infection in mice (Ovcharenko & Zhirnov, 1994) and double-blind randomised clinical trials showed similar benefit (Zhirnov et al., 2011), which suggests its efficacy in targeting viral replication at therapeutic concentrations. Of note, an open-label observational trial assessing the efficacy and safety of treatment of moderate-to-severe COVID-19 patients with

aprotinin began in August 2020, and results are pending (ClinicalTrials.gov Identifier: NCT04527133).

Several other studies have also suggested the possible benefit of using protease inhibitors as a prophylactic measure against SARS-CoV-2 infection, including one which showed that camostat mesylate inhibits TMPRSS2-mediated S priming (Hoffmann et al., 2020), one which found that nafamostat mesylate was ten times more effective in inhibiting viral fusion than camostat mesylate (Yamamoto et al., 2020) and one which suggested the efficacy of treatment with nafamostat mesylate in combination with the antiviral favipiravir (Doi et al., 2020). Another group suggested repurposing of a specific TMPRSS2 inhibitor, bromhexine, which acts as a mucolytic and has been used to treat respiratory conditions associated with excess mucus production (Maggio & Corsini, 2020). A trial of bromhexine in hospitalised COVID-19 patients later showed that treatment with the drug was associated with reduced admission to intensive care, reduced requirement for mechanical ventilation and reduced mortality, as well as faster recovery from symptoms such as cough and dyspnoea and lower serum CRP levels after two weeks (Ansarin et al., 2020). Our results could therefore have significant clinical relevance, given the clear emphasis on the possible benefit of using protease inhibitors in the treatment of COVID-19.

In each of our analyses of publicly available proteomics and transcriptomics datasets, we once again demonstrated the value of data sharing within the research community during the course of the pandemic. Notably, our use of the same databases for multiple separate analyses is indicative of their flexibility and diversity of applications. Our findings could stimulate further research into the potential for CD47 and transferrin to be used as biomarkers of severe disease and as possible therapeutic targets, and they also provided useful evidence to suggest the potential benefit of aprotinin in targeting some of the key mechanisms of SARS-CoV-2 pathogenesis, which could also be investigated further in a clinical setting. In the case of transferrin, the results could even have implications for other diseases involving deregulation of blood coagulation.

9.5. Limitations to the use of publicly available data

Despite the advantages to the use of large publicly available datasets of cell line and patient data, its limitations have been widely documented. Many of the issues stem from the challenges inherent in laboratory and clinical investigations and are therefore not restricted to open-access databases. Not only can data be noisy (for example, low frequency genomic

variants may be difficult to detect), susceptible to batch effects (for example, assays performed under subtly different environmental conditions could yield variable results which are difficult to compare) and insufficient for performing lineage-specific analyses (due to small sample sizes), data tend to only provide measurements at a single time point (i.e. they are a snapshot of the physical state of the cells), and may therefore fail to completely reflect the nature of dynamic features such as gene expression. These issues can seriously impact the reliability and validity of the data and the conclusions drawn from results.

A notable limitation to our investigation in chapter 3 was the small numbers of acute lymphoblastic leukaemia cell lines included in the databases (CCLE, CTRP and GDSC). This highlights a significant issue for large-scale pharmacogenomics databases overall, whereby certain cancer subtypes are less well represented in such databases than others. For example, the GDSC database includes gene expression data for 66 adenocarcinoma-type non-small-cell lung cancer (NSCLC) samples, but it only includes data for one sample from an adrenal gland tumour. Until databases including larger numbers of cell lines and tumour samples from different sub-lineages are produced, the statistical power of analyses such as these can be improved by normalisation methods which enable cross-sample comparisons to be made (i.e. to increase the pool of samples for each condition). This would also enable the results to be compared with new data in subsequent studies. We were nevertheless able to generate significant results using these data independently. Moreover, the consilience observed between the results from each of the different resources used gives us further confidence in our conclusions.

Another issue which may have affected our analysis of data from genomics studies is cell line genetic drift. Repeated passage of cell lines could mean they no longer reflect the characteristics of the tumour of origin, making interpretation of the results and predictions regarding drug sensitivity of particular cell lines difficult. For example, one study of the CCLE, GDSC and Genentech Cell Line Screening Initiative (gCSI) database, showed that between 4.5-6.1% of the entire genomes of the same cell lines had been affected by genetic drift (Quevedo et al., 2020). While it has been proposed that such drift may mainly affect passenger mutations (since correlation between hotspot somatic variants between the CCLE and GDSC was found to be high) (Ghandi et al., 2019), others have demonstrated the impact that genetic divergence can have on cell phenotype – for example, OVCAR-3 cells from different collections of the NCI-60 cell line panel have been shown to exhibit significantly different responses to treatment with doxorubicin, as well as significantly different protein levels of asparagine synthase but similar levels of mRNA (Lorenzi et al., 2009). The possibility

of genetic drift affecting cell line sensitivity to nelarabine in our study in chapter 3 must therefore be taken into account when interpreting the results. However, the agreement of the *in silico* results with our *in vitro* findings, generated using a number of different cell lines from the same lineages, suggests this issue may not significantly affect our results.

Different methods to handle incomplete dose-response curves have also been recognised as a potential issue affecting large pharmacogenomic datasets. For example, in the GDSC data, 126 of the 427 cell lines treated with paclitaxel had their IC50 values extrapolated beyond the top tested drug concentration of 0.1024μM. In the CCLE data, 503 cell lines were treated with paclitaxel over a dose range of 0.0025-8μM – however, for 34 of these cell lines, the maximum IC50 value was capped at 8μM. These differing approaches, along with inadequate standardisation of pharmacological assay, have been proposed to explain the lack of agreement observed between the CCLE and GDSC in previous investigations (Haibe-Kains et al., 2013). The CTRP dataset avoids this problem by providing only AUC (area under the curve) values instead of IC50 values. AUC has been proposed to provide a superior measure of drug response that incorporates data for both the potency and efficacy of the drug and can be more easily compared across cell lines, providing that that cell lines are tested using the same drug concentration ranges (Pozdeyev et al., 2016). We made use of AUC values from the CTRP and GDSC in our study in chapter 3, which correlated well with our *in vitro* investigations.

Concerns regarding the reliability and reproducibility of genomic data from large scale pharmacogenomic screens (a 'reproducibility crisis') have also been raised (Begley & Ioannidis, 2015; Hirsch & Schildknecht, 2019; Wass et al., 2019). Criticism for such databases often surrounds differing practices and methodologies for handling and processing cell lines. For example, limitations in storage capacity by institutions housing large collections of cell lines can necessitate passage of cells which, if repeatedly performed, can result in genetic alterations as described previously. This also highlights the importance of ensuring that, when obtaining cell lines for analysis from other institutions or cell culture banks, sufficient information regarding cell passage number is provided (Hirsch & Schildknecht, 2019). Similarly, cell line lineage is a critical factor to ascertain, as misclassification of cell lineage is another issue which has affected the use of cell lines. A well-known example of this is MDA-MB-435, which was initially considered to be of breast cancer origin but was later suggested to derive entirely from a melanoma cell line (due to mislabelling of the sample or contamination of the breast cancer cells with the melanoma cells (Rae et al., 2007)). However, MDA-MB-435 was still being classed as a breast cancer cell line and used in studies

in the 2010s (Prasad & Gopalan, 2015). Given the reliance of our study in chapter 3 on the ability to differentiate T-acute lymphoblastic leukaemia from B-acute lymphoblastic leukaemia, such discrepancies could seriously impact the conclusions drawn from the results. However, the cell lines included in these pharmacogenomic screens have all undergone extensive characterisation by the researchers responsible for them, and so we can be confident that this issue is unlikely to affect our study.

Analysis of data from databases of donated patient tissue samples (such as the TCGA, TARGET and GTEx) has also been associated with certain obstacles. For example, tumour sample purity can be affected by differing amounts of infiltration by stromal and immune cells. One study estimated TCGA tumour sample purity across all cancer types to be 75.3±18.9%, with inter-tumour variability proposed to be the result of differing mutation rates – for example, lung adenocarcinomas and lung squamous cell carcinomas were found to have both the lowest tumour purity and the highest frequency of mutations compared to all other TCGA cancer types investigated. They suggested that this may be due to greater inflammation (i.e. infiltrating immune cells) in the microenvironment of these tumours driving enhanced mutation. On the other hand, they also proposed that the ease with which tumour cells can be isolated from surrounding cells during sample collection can differ based on cancer type and grade. They also found that tumour purity can impact genomic analyses (such as gene expression correlations and molecular subtyping) (Aran et al., 2015), which may be relevant to our studies in chapters 2, 4 and 7. In future studies, it may improve the validity and reliability of results to adjust for tumour purity before performing analyses using data from tumour samples.

As for cell line pharmacogenomic databases, small sample sizes of cancer patient data (especially for individual cancer types) remain a significant issue, and incomplete data for certain patients (clinical data, certain types of genomic data or combinations of different data types) can further reduce the pool available for analysis. For example, our study in chapter 2 was particularly focused on investigating the role of PEA-15 in cisplatin response in ovarian cancer, but our TCGA analyses were not entity-specific. 772 cisplatin-treated patients were included in the TCGA database; however, only 114 had ovarian cancer. In order to improve the statistical power of the survival analyses, we therefore decided to include the full cisplatin-treated TCGA cohort. However, it may have been useful to perform the analyses only in cisplatin-treated ovarian cancer patients, as this approach may have enhanced the specificity of the observed associations. In addition, patients are typically treated with multiple cytotoxic and/or targeted therapies, and the majority of cisplatin-treated patients

included in our analysis were also treated with a number of other drugs. This could clearly confound results when attempting to consider the impact of only one drug and must therefore be taken into consideration when interpreting the results.

It is also noteworthy that research using open-access data is often limited in the patient phenotypes that can be investigated due to restrictions on the available clinical data – for example, it would have been useful to analyse expression of transferrin and CD47 in relation to the risk factors of COVID-19 using the GTEx dataset for a more comprehensive analysis in chapters 6 and 7. Meanwhile, in some cases, it is also impossible to distinguish between patients that have died as a result of their cancer or due to unrelated causes, which may affect interpretation of our results in chapter 4.

Another potentially problematic issue to consider when analysing large volumes of data is the application of arbitrary cut-offs during analysis. For example, in analysing 'high' vs 'low' gene expression, we used the median of normalised expression (FPKM) across each sample in chapter 2 (with 'high' expression considered to be above median and 'low' expression to be below median), whereas we used a computationally calculated 'best separation' method (i.e. the cut-off which gives the lowest p-value) for our analyses in chapter 4. Our ability to determine high vs low expression was therefore limited by the data we had available, and so expression levels should be considered relative to patients within the same sample. As discussed previously, larger sample sizes would improve the validity of the results by improving the statistical power. Meanwhile, in our study in chapter 5, we considered residues at equivalent alignment positions which differed between the two viruses but were >80% conserved within their respective viral genomes to be differentially conserved. This assumption may oversimplify the situation, potentially resulting in positions with functional significance being overlooked. However, with availability of larger numbers of sequences, it will become possible to make more accurate prediction of which residues may be considered differentially conserved.

Another important issue relevant to a number of the papers presented in this work is the lack of correlation between mRNA and protein levels frequently reported (Koussounadis et al., 2015; Maier et al., 2009). Factors which affect translation of an mRNA transcript (including *cis*-regulatory elements such as promoters and enhancers, regulation of translation elongation factors by phosphorylation, transfer RNA (tRNA) abundances, miRNAs, mRNA transcript stability) along with protein degradation (by the ubiquitin-proteasome pathway or by lysosomal proteolysis) dictate the relative abundances of mRNA

transcripts and their corresponding proteins within a cell at any given time (de Sousa Abreu et al., 2009). It has been shown that mRNA and protein levels correlate better for certain classes of genes than others (such as those encoding proteins involved in translation or metabolic processes (Nicolet & Monika C. Wolkers, 2020), or those encoding cell cycle-regulated proteins, for which mRNA and protein synthesis and protein degradation are coordinated (Ly et al., 2014)). We addressed this issue in the study presented in chapter 3, in which cell line mRNA levels (determined by microarray) were compared with protein levels (determined by quantitative Western blot and quantitative PCR) and found some differences in SAMHD1 mRNA and protein levels in certain cell lines, but overall expression levels appeared moderately correlated (both within *in vitro* data and between *in vitro* and *in silico* data). However, for our studies using patient transcriptomic data from the TCGA, TARGET and GTEx databases (chapters 2, 4, 6 and 7), our inferences of clinical significance were based on the assumption that mRNA levels will reflect protein levels. Further validation of our results will therefore be necessary in order to be confident in any conclusions made from these studies.

It should also be noted that, while the patient sample-derived data used in these studies were generated using next-generation sequencing technologies (such as RNA-seq for gene expression quantification), the cell line data was generated using microarray technology. Microarrays have been associated with a number of issues (such as difficulties in normalisation and correcting for batch effects, noisy data, low accuracy and specificity, incorrect probe assignment to genes and limitations in the probes included on a single microarray) (Jaksik et al., 2015), many of which are overcome by RNA-seq technologies – for example, RNA-seq is better able to detect transcripts with low expression, it displays greater reproducibility, and the method also allows for the identification of novel transcripts since it does not rely on the use of a reference genome (van der Kloet et al., 2020). As the technology becomes more cost-effective to use, RNA-seq is becoming the preferred method for gene expression analysis. For example, newer updates of the CCLE databases are indeed including expression data generated using the RNA-seq platform (Ghandi et al., 2019).

A number of our studies also involved manual analysis of data which may also have several limitations, given its subjectivity and the potential for human error. In particular, the structural analyses carried out (including that of the SAMHD1 substrate binding pocket in chapter 4 and of each of the available structures of SARS-CoV and SARS-CoV-2 proteins in chapter 5) relied on use of molecular graphics software (i.e. static protein structures) to make conclusions regarding the impact of amino acid variants on protein function. While

supplementary data is available regarding the impact of SAMHD1 active site mutations (from extensive in vitro studies described in the literature, as described), little is yet known about the impact of variants on SARS-CoV-2 structure and function. Moreover, considering the impact of amino acid substitutions independently of their known interactions (for example, with other SARS-CoV-2 proteins if the protein normally exists in a complex) could have resulted in variants with substantial structural or functional impact being predicted to be benign. As more is discovered about the interactions of the SARS-CoV-2 viral proteins with other viral and host proteins and as more structures become available which model these interactions, it will be possible to make more confident predictions about the consequences of variants observed within the SARS-CoV-2 genome.

Chapter 10: Conclusions

Substantial progress has been made in recent years to incorporate computational methods into biological research. Tasks that might have taken years to complete can now be completed in days, and this has had a major impact on the ability to deliver effective medical insights and interventions quickly and affordably. The drive towards the availability of public data has undoubtedly complemented this development, as increasingly powerful and efficient methods to analyse the vast quantities of available data are required.

This project has highlighted the flexibility of a computational approach when adapting to novel scenarios and responding to different research demands. It is clear that the analysis of large databases has been made considerably more feasible by the implementation of *in silico* methods – for example, it is now possible to compare features of disease from different tissues of origin, between different populations of individuals, and to relate these variations to the differences in response to treatment. This in turn has prompted researchers to produce larger and larger datasets with the knowledge that these data can be rapidly processed and even analysed with minimal human input.

References

Abdul-Rahman, O., Sasvari-Szekely, M., Ver, A., Rosta, K., Szasz, B. K., Kereszturi, E., & Keszler, G. (2012). Altered gene expression profiles in the hippocampus and prefrontal cortex of type 2 diabetic rats. *BMC Genomics, 13*(81) doi:https://doi.org/10.1186/1471-2164-13-81

Abu El-Asrar, A. M., Nawaz, M. I., Ola, M. S., De Hertogh, G., Opdenakker, G., & Geboes, K. (2013). Expression of thrombospondin-2 as a marker in proliferative diabetic retinopathy. *Acta Ophthalmologica, 91*(3), e169-e177. doi:https://doi.org/10.1111/aos.12035

Acharya, A., Das, I., Chandhok, D., & Saha, T. (2010). Redox regulation in cancer: A double-edged sword with therapeutic potential. *Oxidative Medicine and Cellular Longevity, 3*(1), 23-34. doi:https://doi.org/10.4161/oxim.3.1.10095

Adachi, M., Fukuda, M., & Nishida, E. (1999). Two co-existing mechanisms for nuclear import of MAP kinase: Passive diffusion of a monomer and active transport of a dimer. *EMBO Journal, 18*, 5347-5358. doi:https://doi.org/10.1093/emboj/18.19.5347

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., . . . Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods, 7*(4), 248-249. doi:https://doi.org/10.1038/nmeth0410-248

Akahane, K., Murakami, Y., Kagami, K., Abe, M., Harama, D., Shinohara, T., . . . Inukai, T. (2019). High ENT1 and DCK gene expression levels are a potential biomarker to predict favorable response to nelarabine therapy in T-cell acute lymphoblastic leukemia. *Hematological Oncology, 37*(4), 516-519. doi:https://doi.org/10.1002/hon.2654

Akimova, E., Gassner, F. J., Schubert, M., Rebhandl, S., Arzt, C., Rauscher, S., . . . Geisberger, R. (2021). SAMHD1 restrains aberrant nucleotide insertions at repair junctions generated by DNA end joining. *Nucleic Acids Research, 49*(5), 2598-2608. doi:https://doi.org/10.1093/nar/gkab051

Allen, L. B., Capps, B. E., Miller, E. C., Clemmons, D. R., & Maile, L. A. (2009). Glucose-oxidized low-density lipoproteins enhance insulin-like growth factor I-stimulated smooth

muscle cell proliferation by inhibiting integrin-associated protein cleavage. *Endocrinology, 150*(3), 1321-1329. doi:https://doi.org/10.1210/en.2008-1090

Alley, M. C., Scudiero, D. A., Monks, P. A., Hursey, M. L., Czerwinski, M. J., Fine, D. L., . . . Boyd, M. R. (1988). Feasibility of drug screening with panels of human tumor cell lines using a microculture tetrazolium assay. *Cancer Research, 48*(3), 589-601.

Almughem, F. A., Aldossary, A. M., Tawfik, E. A., Alomary, M. N., Alharbi, W. S., Alshahrani, M. Y., & Alshehri, A. A. (2020). Cystic fibrosis: Overview of the current development trends and innovative therapeutic strategies. *Pharmaceutics, 12*(7), 616. doi:https://doi.org/10.3390/pharmaceutics12070616

Amie, S. M., Daly, M. B., Noble, E., Schinazi, R. F., Bambara, R. A., & Kim, B. (2013). Anti-HIV host factor SAMHD1 regulates viral sensitivity to nucleoside reverse transcriptase inhibitors via modulation of cellular deoxyribonucleoside triphosphate (dNTP) levels. *The Journal of Biological Chemistry, 288*(28), 20683-20691. doi:https://doi.org/10.1074/jbc.M113.472159

Amin, N. A., Seymour, E., Saiya-Cork, K., Parkin, B., Shedden, K., & Malek, S. N. (2016). A quantitative analysis of subclonal and clonal gene mutations before and after therapy in chronic lymphocytic leukemia. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research, 22*(17), 4525-4535. doi:https://doi.org/10.1158/1078-0432.CCR-15-3103

Anandhalakshmi, S., Manikandan, S., Ganeshkumar, P., & Ramachandran, C. (2013). Alveolar gas exchange and pulmonary functions in patients with type II diabetes mellitus. *Journal of Clinical and Diagnostic Research, 7*(9), 1874-1877. doi:https://doi.org/10.7860/JCDR/2013/6550.3339

Andreano, E., Piccini, G., Licastro, D., Casalino, L., Johnson, N. V., Paciello, I., . . . Rappuoli, R. (2020). SARS-CoV-2 escape in vitro from a highly neutralizing COVID-19 convalescent plasma. *bioRxiv,* doi:https://doi.org/10.1101/2020.12.28.424451

Ansarin, K., Tolouian, R., Ardalan, M., Taghizadieh, A., Varshochi, M., Teimouri, S., . . . Chapman, K. R. (2020). Effect of bromhexine on clinical outcomes and mortality in COVID-19 patients: A randomized clinical trial. *BioImpacts : BI, 10*(4), 209-215. doi:https://doi.org/10.34172/bi.2020.27

Aran, D., Sirota, M., & Butte, A. (2015). Systematic pan-cancer analysis of tumour purity. *Nature Communications, 6*, 8971. doi:https://doi.org/10.1038/ncomms9971

Araujo, H., Danziger, N., Cordier, J., Glowinski, J., & Chneiweiss, H. (1993). Characterization of PEA-15, a major substrate for protein kinase C in astrocytes. *Journal of Biological Chemistry, 268*(8), 5911-5920.

Aref-Adib, M., & Freeman-Wang, T. (2016). Cervical cancer prevention and screening: The role of human papillomavirus testing. *The Obstetrician and Gynaecologist, 18*(4), 251-263. doi:https://doi.org/10.1111/tog.12279

Armstrong, D. R., Berrisford, J. M., Conroy, M. J., Gutmanas, A., Anyango, S., Choudhary, P., . . . Velankar, S. (2020). PDBe: Improved findability of macromolecular structure data in the PDB. *Nucleic Acids Research, 48*(D1), D335-D343. doi:https://doi.org/10.1093/nar/gkz990

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., . . . Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. the gene ontology consortium. *Nature Genetics, 25*(1), 25-29. doi:https://doi.org/10.1038/75556

Atencia, R., Garcia-Sanz, M., Unda, F., & Arechaga, J. (1994). Apoptosis during retinoic acid-induced differentiation of F9 embryonal carcinoma cells. *Experimental Cell Research, 214*(2), 663-667. doi:https://doi.org/10.1006/excr.1994.1304

Azouz, N. P., Klingler, A. M., & Rothenberg, M. E. (2020). Alpha 1 antitrypsin is an inhibitor of the SARS-CoV2–Priming protease TMPRSS2. *bioRxiv,* doi:https://doi.org/10.1101/2020.05.04.077826

Baden, L. R., El Sahly, H. M., Essink, B., Kotloff, K., Frey, S., Novak, R., . . . Zaks, T. (2021). Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. *The New England Journal of Medicine, 384*(5), 403-416. doi:https://doi.org/10.1056/NEJMoa2035389

Baldauf, H. M., Pan, X., Erikson, E., Schmidt, S., Daddacha, W., Burggraf, M., . . . Keppler, O. T. (2012). SAMHD1 restricts HIV-1 infection in resting CD4(+) T cells. *Nature Medicine, 18*(11), 1682-1687. doi:https://doi.org/10.1038/nm.2964

Ballana, E., Badia, R., Terradas, G., Torres-Torronteras, J., Ruiz, A., Pauls, E., . . . Esté, J. A. (2014). SAMHD1 specifically affects the antiviral potency of thymidine analog HIV reverse

transcriptase inhibitors. *Antimicrobial Agents and Chemotherapy, 58*(8), 4804-4813. doi:https://doi.org/10.1128/AAC.03145-14

Barh, D., Tiwari, S., Weener, M. E., Azevedo, V., Góes-Neto, A., Gromiha, M. M., & Ghosh, P. (2020). Multi-omics-based identification of SARS-CoV-2 infection biology and candidate drugs against COVID-19. *Computers in Biology and Medicine, 126*, 104051. doi:https://doi.org/10.1016/j.compbiomed.2020.104051

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A. A., Kim, S., . . . Garraway, L. A. (2012). The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature, 483*(7391), 603-607. doi:https://doi.org/10.1038/nature11003

Bartholomeusz, C., Itamochi, H., Nitt, M., Saya, H., Ginsberg, M. H., & Ueno, N. T. (2006). Antitumor effect of E1A in ovarian cancer by cytoplasmic sequestration of activated ERK by PEA-15. *Oncogene, 25*, 79-90. doi:https://doi.org/10.1038/sj.onc.1209014

Bartholomeusz, C., Rosen, D., Wei, C., Kazansky, A., Yamasaki, F., Takahashi, T., . . . Ueno, N. T. (2008). PEA-15 induces autophagy in human ovarian cancer cells and is associated with prolonged overall survival. *Cancer Research, 68*(22), 9302-9310. doi:https://doi.org/10.1158/0008-5472.CAN-08-2592

Basu, A., Bodycombe, N. E., Cheah, J. H., Price, E. V., Liu, K., Schaefer, G. I., . . . Schreiber, S. L. (2013). An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. *Cell, 154*(5), 1151-1161. doi:https://doi.org/10.1016/j.cell.2013.08.003

Bauer, P. M., Bauer, E. M., Rogers, N. M., Yao, M., Feijoo-Cuaresma, M., Pilewski, J. M., . . . Isenberg, J. S. (2012). Activated CD47 promotes pulmonary arterial hypertension through targeting caveolin-1. *Cardiovascular Research, 92*(4), 682-693. doi:https://doi.org/10.1093/cvr/cvr356

Bauer-Mehren, A., Rautschka, M., Sanz, F., & Furlong, L. I. (2010). DisGeNET: A cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. *Bioinformatics (Oxford, England), 26*(22), 2924-2926. doi:https://doi.org/10.1093/bioinformatics/btq538

Beesley, A. H., Palmer, M. L., Ford, J., Weller, R. E., Cummings, A. J., Freitas, J. R., . . . Kees, U. R. (2007). In vitro cytotoxicity of nelarabine, clofarabine and flavopiridol in paediatric acute lymphoblastic leukaemia. *British Journal of Haematology, 137*(2), 109-116. doi:https://doi.org/10.1111/j.1365-2141.2007.06527.x

Begley, C. G., & Ioannidis, J. P. (2015). Reproducibility in science: Improving the standard for basic and preclinical research. *Circulation Research, 116*(1), 116-126. doi:https://doi.org/10.1161/CIRCRESAHA.114.303819

Behloul, N., Baha, S., Guo, Y., Yang, Z., Shi, R., & Meng, J. (2020). In silico identification of strong binders of the SARS-CoV-2 receptor-binding domain. *European Journal of Pharmacology, 173701* doi:https://doi.org/10.1016/j.ejphar.2020.173701

Beigel, J. H., Tomashek, K. M., Dodd, L. E., Mehta, A. K., Zingman, B. S., Kalil, A. C., . . . ACTT-1 Study Group Members. (2020). Remdesivir for the treatment of covid-19 - final report. *The New England Journal of Medicine, 383*(19), 1813-1826. doi:https://doi.org/10.1056/NEJMoa2007764

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B, 57*(1), 289-300. doi:https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research, 41 (Database issue)*, D36-D42. doi:https://doi.org/10.1093/nar/gks1195

Berg, S. L., Blaney, S. M., Devidas, M., Lampkin, T. A., Murgo, A., Bernstein, M., . . . Children's Oncology Group. (2005). Phase II study of nelarabine (compound 506U78) in children and young adults with refractory T-cell malignancies: A report from the children's oncology group. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology, 23*(15), 3376-3382. doi:https://doi.org/10.1200/JCO.2005.03.426

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., . . . Bourne, P. E. (2000). The protein data bank . *Nucleic Acids Research, 28*(1), 235-242. doi:https://doi.org/10.1093/nar/28.1.235

Bestle, D., Heindl, M. R., Limburg, H., Van Lam van, T., Pilgram, O., Moulton, H., . . . Böttcher-Friebertshäuser, E. (2020). TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells. *Life Science Alliance, 3*(9), e202000786. doi:https://doi.org/10.26508/lsa.202000786

Bicknell, A. A., & Ricci, E. P. (2017). When mRNA translation meets decay. *Biochemical Society Transactions, 45*(2), 339-351. doi:https://doi.org/10.1042/BST20160243

Bitar, M. S. (2019). Diabetes impairs angiogenesis and induces endothelial cell senescence by up-regulating thrombospondin-CD47-dependent signaling. *International Journal of Molecular Sciences, 20*(3), 673. doi:https://doi.org/10.3390/ijms20030673

Blanco-Melo, D., Nilsson-Payant, B. E., Liu, W. C., Uhl, S., Hoagland, D., Møller, R., . . . tenOever, B. R. (2020). Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell, 181*(5), 1036-1045. doi:https://doi.org/10.1016/j.cell.2020.04.026

Bojkova, D., Bechtel, M., McLaughlin, K., McGreig, J. E., Klann, K., Bellinghausen, C., . . . Cinatl Jr, J. (2020). Aprotinin inhibits SARS-CoV-2 replication. *Cells, 9*(11), 2377. doi:https://doi.org/10.3390/cells9112377

Bojkova, D., Klann, K., Koch, B., Widera, M., Krause, D., Ciesek, S., . . . Münch, C. (2020). Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature, 583*, 469-472. doi:https://doi.org/10.1038/s41586-020-2332-7

Bolondi, G., Russo, E., Gamberini, E., Circelli, A., Meca, M. C. C., Brogi, E., . . . Agnoletti, V. (2020). Iron metabolism and lymphocyte characterisation during covid-19 infection in ICU patients: An observational cohort study. *World Journal of Emergency Surgery, 15*, 41. doi:https://doi.org/10.1186/s13017-020-00323-2

Borczuk, A. C., Salvatore, S. P., Seshan, S. V., Patel, S. S., Bussel, J. B., Mostyka, M., . . . Beasley, M. B. (2020). COVID-19 pulmonary pathology: A multi-institutional autopsy cohort from italy and new york city. *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc., 33*(11), 2156-2168. doi:https://doi.org/10.1038/s41379-020-00661-1

Borges do Nascimento, I. J., Cacic, N., Abdulazeem, H. M., von Groote, T. C., Jayarajah, U., Weerasekara, I., . . . Marcolino, M. S. (2020). Novel coronavirus infection (COVID-19) in

humans: A scoping review and meta-analysis. *Journal of Clinical Medicine, 9*(4), 941. doi:https://doi.org/10.3390/jcm9040941

Böttcher, E., Freuer, C., Steinmetzer, T., Klenk, H. D., & Garten, W. (2009). MDCK cells that express proteases TMPRSS2 and HAT provide a cell system to propagate influenza viruses in the absence of trypsin and to study cleavage of HA and its inhibition. *Vaccine, 27*(45), 6324-6329. doi:https://doi.org/10.1016/j.vaccine.2009.03.029

Bou Ghanem, E. N., Clark, S., Du, X., Wu, D., Camilli, A., Leong, J. M., & Meydani, S. N. (2015). The α-tocopherol form of vitamin E reverses age-associated susceptibility to streptococcus pneumoniae lung infection by modulating pulmonary neutrophil recruitment. *Journal of Immunology (Baltimore, Md. : 1950), 194*(3), 1090-1099. doi:https://doi.org/10.4049/jimmunol.1402401

Boulton, T. G., Nye, S. H., Robbins, D. J., Ip, N. Y., Radziejewska, E., Morgenbesser, S. D., . . . Yancopoulos, G. D. (1991). ERKs: A family of protein-serine/threonine kinases that are activated and tyrosine phosphorylated in response to insulin and NGF. *Cell, 65*, 663-675. doi:https://doi.org/10.1016/0092-8674(91)90098-j

Boulton, T. G., Yancopoulos, G. D., Gregory, J. S., Slaughter, C., Moomaw, C., Hsu, J., & Cobb, M. H. (1990). An insulin-stimulated protein kinase similar to yeast kinases involved in cell cycle control. *Science, 249*(4964), 64-67. doi:https://doi.org/10.1126/science.2164259

Brooke, G. N., & Prischi, F. (2020). Structural and functional modelling of SARS-CoV-2 entry in animal models. *Scientific Reports, 10*, 15917. doi:https://doi.org/10.1038/s41598-020-72528-z

Bühler, M. M., Lu, J., Scheinost, S., Nadeu, F., Roos-Weil, D., Hensel, M., . . . Zenz, T. (2021). SAMHD1 mutations in mantle cell lymphoma are recurrent and confer in vitro resistance to nucleoside analogues. *Leukemia Research, 107*(106608) doi:https://doi.org/10.1016/j.leukres.2021.106608

Buja, L. M., Wolf, D. A., Zhao, B., Akkanti, B., McDonald, M., Lelenwa, L., . . . Kar, B. (2020). The emerging spectrum of cardiopulmonary pathology of the coronavirus disease 2019 (COVID-19): Report of 3 autopsies from houston, texas, and review of autopsy findings from other united states cities. *Cardiovascular Pathololgy, 48*, 107233. doi:https://doi.org/10.1016/j.carpath.2020.107233

Bukowski, K., Kciuk, M., & Kontek, R. (2020). Mechanisms of multidrug resistance in cancer chemotherapy. *International Journal of Molecular Sciences, 21*(9), 3233. doi:https://doi.org/10.3390/ijms21093233

Burns, A., Alsolami, R., Becq, J., Stamatopoulos, B., Timbs, A., Bruce, D., . . . Schuh, A. (2018). Whole-genome sequencing of chronic lymphocytic leukaemia reveals distinct differences in the mutational landscape between IgHVmut and IgHVunmut subgroups. *Leukemia, 32*(2), 332-342. doi:https://doi.org/10.1038/leu.2017.177

Cai, L., Li, Q., Du, Y., Yun, J., Xie, Y., DeBerardinis, R. J., & Xiao, G. (2017). Genomic regression analysis of coordinated expression. *Nature Communications, 8*(1), 2187. doi:https://doi.org/10.1038/s41467-017-02181-0

Cai, Q., Chen, F., Wang, T., Luo, F., Liu, X., Wu, Q., . . . Xu, L. (2020). Obesity and COVID-19 severity in a designated hospital in shenzhen, china. *Diabetes Care, 43*(7), 1392-1398. doi:https://doi.org/10.2337/dc20-0576

Callaway, K., Abramczyk, O., Martin, L., & Dalby, K. N. (2007). The anti-apoptotic protein PEA-15 is a tight binding inhibitor of ERK1 and ERK2, which blocks docking interactions at the D-recruitment site. *Biochemistry, 46*(32), 9187-9198. doi:https://doi.org/10.1021/bi700206u

Calogero, A., Porcellini, A., Lombari, V., Fabbiano, C., Arcella, A., Miscusi, M., . . . Ragona, G. (2011). Sensitivity to cisplatin in primary cell lines derived from human glioma correlates with levels of EGR-1 expression. *Cancer Cell International, 11*, 5. doi:https://doi.org/10.1186/1475-2867-11-5

Cao, L., Goreshnik, I., Coventry, B., Case, J. B., Miller, L., Kozodoy, L., . . . Baker, D. (2020). De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. *Science, 370*(6515), 426-431. doi:https://doi.org/10.1126/science.abd9909

Capra, J. A., & Singh, M. (2007). Predicting functionally important residues from sequence conservation. *Bioinformatics (Oxford, England), 23*(15), 1875-1882. doi:https://doi.org/10.1093/bioinformatics/btm270

Casari, G., Sander, C., & Valencia, A. (1995). A method to predict functional residues in proteins. *Nature Structural and Molecular Biology, 2*, 171-178. doi:https://doi.org/10.1038/nsb0295-171

Cashman, R., Zilberberg, A., Priel, A., Philip, H., Varvak, A., Jacob, A., . . . Efroni, S. (2020). A single nucleotide variant of human PARP1 determines response to PARP inhibitors. *NPJ Precision Oncology, 4*, 10. doi:https://doi.org/10.1038/s41698-020-0113-2

Cava, C., Bertoli, G., & Castiglioni, I. (2020). In silico discovery of candidate drugs against covid-19. *Viruses, 12*(4), 404. doi:https://doi.org/10.3390/v12040404

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., . . . Schultz, N. (2012). The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discovery, 2*, 401. doi:https://doi.org/10.1158/2159-8290.CD-12-0095

Chaibi, K., Dao, M., Pham, T., Gumucio-Sanguino, V. D., Di Paolo, F. A., Pavot, A., . . . Gaudry, S. (2020). Severe acute kidney injury in patients with COVID-19 and acute respiratory distress syndrome. *American Journal of Respiratory and Critical Care Medicine, 202*(9), 1299-1301. doi:https://doi.org/10.1164/rccm.202005-1524LE

Chakraborti, S., Prabakaran, P., Xiao, X., & Dimitrov, D. S. (2005). The SARS coronavirus S glycoprotein receptor binding domain: Fine mapping and functional characterization. *Virology Journal, 2*(73) doi:https://doi.org/10.1186/1743-422X-2-73

Cham, L. B., Adomati, T., Li, F., Ali, M., & Lang, K. S. (2020). CD47 as a potential target to therapy for infectious diseases. *Antibodies (Basel, Switzerland), 9*(3), 44. doi:https://doi.org/10.3390/antib9030044

Chang, L., & Karin, M. (2001). Mammalian MAP kinase signalling cascades. *Nature, 410*, 37-40. doi:https://doi.org/10.1038/35065000

Chao, M. P., Weissman, I. L., & Majeti, R. (2012). The CD47-SIRPα pathway in cancer immune evasion and potential therapeutic implications. *Current Opinion in Immunology, 24*(2), 225-232. doi:https://doi.org/10.1016/j.coi.2012.01.010

Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., . . . Zhang, L. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in wuhan, china: A descriptive study. *The Lancet, 395*(10223), 507-513. doi:https://doi.org/10.1016/S0140-6736(20)30211-7

Chen, R. H., Sarnecki, C., & Blenis, J. (1992). Nuclear localization and regulation of erk- and rsk-encoded protein kinases. *Molecular and Cellular Biology, 12*(3), 915-927. doi:https://doi.org/10.1128/mcb.12.3.915

Chen, X., Li, R., Pan, Z., Qian, C., Yang, Y., You, R., . . . Ye, L. (2020). Human monoclonal antibodies block the binding of SARS-CoV-2 spike protein to angiotensin converting enzyme 2 receptor. *Cellular and Molecular Immunology, 17*(6), 647-649. doi:https://doi.org/10.1038/s41423-020-0426-7

Chen, Z., Hu, J., Ying, S., & Xu, A. (2021). Dual roles of SAMHD1 in tumor development and chemoresistance to anticancer drugs. *Oncology Letters, 21*(6), 451. doi:https://doi.org/10.3892/ol.2021.12712

Cheng, V. C., Chan, J. F., To, K. K., & Yuen, K. Y. (2013). Clinical management and infection control of SARS: Lessons learned. *Antiviral Research, 100*(2), 407-419. doi:https://doi.org/10.1016/j.antiviral.2013.08.016

Cheng, Y. W., Chao, T. L., Li, C. L., Chiu, M. F., Kao, H. C., Wang, S. H., . . . Yeh, S. H. (2020). Furin inhibitors block SARS-CoV-2 spike protein cleavage to suppress virus production and cytopathic effects. *Cell Reports, 33*(2), 108254. doi:https://doi.org/10.1016/j.celrep.2020.108254

Chilamakuri, R., & Agarwal, S. (2021). COVID-19: Characteristics and therapeutics. *Cells, 10*(2), 206. doi:https://doi.org/10.3390/cells10020206

Choudhary, S., & Silakari, O. (2020). Scaffold morphing of arbidol (umifenovir) in search of multi-targeting therapy halting the interaction of SARS-CoV-2 with ACE2 and other proteases involved in COVID-19. *Virus Research, 289*(198146) doi:https://doi.org/10.1016/j.virusres.2020.198146

Chowdhury, S. M., Talukder, S. A., Khan, A. M., Afrin, N., Ali, M. A., Islam, R., . . . Halim, M. A. (2020). Antiviral peptides as promising therapeutics against SARS-CoV-2. *The Journal of Physical Chemistry B, 124*(44), 9785-9792. doi:https://doi.org/10.1021/acs.jpcb.0c05621

Cinatl Jr, J., Cinatl, J., Weber, B., Rabenau, H., Gümbel, H. O., Chenot, J. F., . . . Doerr, H. W. (1995). In vitro inhibition of human cytomegalovirus replication in human foreskin fibroblasts and endothelial cells by ascorbic acid 2-phosphate. *Antiviral Research, 27*(4), 405-418. doi:https://doi.org/10.1016/0166-3542(95)00024-G

Cinatl Jr, J., Hoever, G., Morgenstern, B., Preiser, W., Vogel, J. U., Hofmann, W. K., . . . Doerr, H. W. (2004). Infection of cultured intestinal epithelial cells with severe acute respiratory syndrome coronavirus. *Cellular and Molecular Life Sciences : CMLS, 61*(16), 2100-2112. doi:https://doi.org/10.1007/s00018-004-4222-9

Cinatl Jr, J., Michaelis, M., Morgenstern, B., & Doerr, H. W. (2005). High-dose hydrocortisone reduces expression of the pro-inflammatory chemokines CXCL8 and CXCL10 in SARS coronavirus-infected intestinal cells. *International Journal of Molecular Medicine, 15*(2), 323-327.

Cinatl, J., Morgenstern, B., Bauer, G., Chandra, P., Rabenau, H., & Doerr, H. W. (2003). Glycyrrhizin, an active component of liquorice roots, and replication of SARS-associated coronavirus. *The Lancet, 361*(9374), 2045-2046. doi:https://doi.org/10.1016/S0140-6736(03)13615-X

Clausen, T. M., Sandoval, D. R., Spliid, C. B., Pihl, J., Painter, C. D., Thacker, B. E., . . . Esko, J. D. (2020). SARS-CoV-2 infection depends on cellular heparan sulfate and ACE2. *bioRxiv, 2020.07.14.201616* doi:https://doi.org/10.1101/2020.07.14.201616

Clifford, R., Louis, T., Robbe, P., Ackroyd, S., Burns, A., Timbs, A. T., . . . Schuh, A. (2014). SAMHD1 is mutated recurrently in chronic lymphocytic leukemia and is involved in response to DNA damage. *Blood, 123*(7), 1021-1031. doi:https://doi.org/10.1182/blood-2013-04-490847

Coburn, S. B., Bray, F., Sherman, M. E., & Trabert, B. (2017). International patterns and trends in ovarian cancer incidence, overall and by histologic subtype. *International Journal of Cancer, 140*(11), 2451-2460.

Coccaro, N., Anelli, L., Zagaria, A., Specchia, G., & Albano, F. (2019). Next-generation sequencing in acute lymphoblastic leukemia. *International Journal of Molecular Sciences, 20*(12), 2929. doi:https://doi.org/10.3390/ijms20122929

Coggins, S. A., Mahboubi, B., Schinazi, R. F., & Kim, B. (2020). SAMHD1 functions and human diseases. *Viruses, 12*(4), 382. doi:https://doi.org/10.3390/v12040382

Cohen, A., Lee, J. W., & Gelfand, E. W. (1983). Selective toxicity of deoxyguanosine and arabinosyl guanine for T-leukemic cells. *Blood, 61*(4), 660-666.

Cooper, D., Hara, H., Iwase, H., Yamamoto, T., Li, Q., Ezzelarab, M., . . . Ayares, D. (2019). Justification of specific genetic modifications in pigs for clinical organ xenotransplantation. *Xenotransplantation, 26*(4), e12516. doi:https://doi.org/10.1111/xen.12516

Coquel, F., Silva, M. J., Técher, H., Zadorozhny, K., Sharma, S., Nieminuszczy, J., . . . Pasero, P. (2018). SAMHD1 acts at stalled replication forks to prevent interferon induction. *Nature, 557*(7703), 57-61. doi:https://doi.org/10.1038/s41586-018-0050-1

Corman, V. M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D. K., . . . Drosten, C. (2020). Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveillance : Bulletin Europeen Sur Les Maladies Transmissibles = European Communicable Disease Bulletin, 25*(3), 2000045. doi:https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045

Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. (2020). The species severe acute respiratory syndrome-related coronavirus: Classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology, 5*, 536-544. doi:https://doi.org/10.1038/s41564-020-0695-z

Critchley, J. A., Carey, I. M., Harris, T., DeWilde, S., Hosking, F. J., & Cook, D. G. (2018). Glycemic control and risk of infections among people with type 1 or type 2 diabetes in a large primary care cohort study. *Diabetes Care, 41*(10), 2127-2135. doi:https://doi.org/10.2337/dc18-0287

Cruz Rodriguez, J. B., Lange, R. A., & Mukherjee, D. (2020). Gamut of cardiac manifestations and complications of COVID-19: A contemporary review. *Journal of Investigative Medicine :*

The Official Publication of the American Federation for Clinical Research, 68(8), 1334-1340. doi:https://doi.org/10.1136/jim-2020-001592

Cui, J., Li, F., & Shi, Z. L. (2019). Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology, 17*(3), 181-192. doi:https://doi.org/10.1038/s41579-018-0118-9

Curbo, S., & Karlsson, A. (2006). Nelarabine: A new purine analog in the treatment of hematologic malignancies. *Reviews on Recent Clinical Trials, 1*(3), 185-192. doi:https://doi.org/10.2174/157488706778250104

D'Addario, M., Arora, P. D., & McCulloch, C. A. (2006). Role of p38 in stress activation of Sp1. *Gene, 379*, 51-61. doi:https://doi.org/10.1016/j.gene.2006.04.012

Daddacha, W., Koyen, A. E., Bastien, A. J., Head, P. E., Dhere, V. R., Nabeta, G. N., . . . Yu, D. S. (2017a). SAMHD1 promotes DNA end resection to facilitate DNA repair by homologous recombination. *Cell Reports, 20*(8), 1921-1935. doi:https://doi.org/10.1016/j.celrep.2017.08.008

Daddacha, W., Koyen, A. E., Bastien, A. J., Head, P. E., Dhere, V. R., Nabeta, G. N., . . . Yu, D. S. (2017b). SAMHD1 promotes DNA end resection to facilitate DNA repair by homologous recombination. *Cell Reports, 20*(8), 1921-1935. doi:https://doi.org/10.1016/j.celrep.2017.08.008

Dai, Z., Zeng, D., Cui, D., Wang, D., Feng, Y., Shi, Y., . . . Lu, H. (2020). Prediction of COVID-19 patients at high risk of progression to severe disease. *Frontiers in Public Health, 8*(574915) doi:https://doi.org/10.3389/fpubh.2020.574915

Damas, J., Hughes, G. M., Keough, K. C., Painter, C. A., Persky, N. S., Corbo, M., . . . Lewin, H. A. (2020). Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. *Proceedings of the National Academy of Sciences, 117*(36), 22311-22322. doi:https://doi.org/10.1073/pnas.2010146117

Damia, G., & Broggini, M. (2019). Platinum resistance in ovarian cancer: Role of DNA repair. *Cancers, 11*(1), 119. doi:https://doi.org/10.3390/cancers11010119

Danesi, R., Fogli, S., Indraccolo, S., Del Re, M., Dei Tos, A. P., Leoncini, L., . . . Conte, P. (2021). Druggable targets meet oncogenic drivers: Opportunities and limitations of target-

based classification of tumors and the role of molecular tumor boards. *ESMO Open, 6*(2), 100040. doi:https://doi.org/10.1016/j.esmoop.2020.100040

Danser, A., Epstein, M., & Batlle, D. (2020). Renin-angiotensin system blockers and the COVID-19 pandemic: At present there is no evidence to abandon renin-angiotensin system blockers. *Hypertension (Dallas, Tex. : 1979), 75*(6), 1382-1385. doi:https://doi.org/10.1161/HYPERTENSIONAHA.120.15082

Danziger, N., Yokoyama, M., Jay, T., Cordier, J., Glowinski, J., & Chneiweiss, H. (1995). Cellular expression, developmental regulation, and phylogenic conservation of PEA-15, the astrocytic major phosphoprotein and protein kinase C substrate. *Journal of Neurochemistry, 64*(3), 1016-1025. doi:https://doi.org/10.1046/j.1471-4159.1995.64031016.x

Das, S., Camphausen, K., & Shankavaram, U. (2020). In silico drug repurposing to combat COVID-19 based on pharmacogenomics of patient transcriptomic data. *Research Square, rs.3.rs-39128* doi:https://doi.org/10.21203/rs.3.rs-39128/v1

Das, S., Lee, D., Sillitoe, I., Dawson, N. L., Lees, J. G., & Orengo, C. A. (2015). Functional classification of CATH superfamilies: A domain-based approach for protein function annotation. *Bioinformatics (Oxford, England), 31*(21), 3460-3467. doi:https://doi.org/10.1093/bioinformatics/btv398

de Silva, S., Hoy, H., Hake, T. S., Wong, H. K., Porcu, P., & Wu, L. (2013). Promoter methylation regulates SAMHD1 gene expression in human CD4+ T cells. *The Journal of Biological Chemistry, 288*(13), 9284-9292. doi:https://doi.org/10.1074/jbc.M112.447201

de Silva, S., Wang, F., Hake, T. S., Porcu, P., Wong, H. K., & Wu, L. (2014). Downregulation of SAMHD1 expression correlates with promoter DNA methylation in sézary syndrome patients. *The Journal of Investigative Dermatology, 134*(2), 562-565. doi:https://doi.org/10.1038/jid.2013.311

de Sousa Abreu, R., Penalva, L. O., Marcotte, E. M., & Vogel, C. (2009). Global signatures of protein and mRNA expression levels. *Molecular bioSystems, 5*(12), 1512-1526. doi:https://doi.org/10.1039/b908315d

de Wit, E., van Doremalen, N., Falzarano, D., & Munster, V. J. (2016). SARS and MERS: Recent insights into emerging coronaviruses. *Nature Reviews Microbiology, 14*(8), 523-534. doi:https://doi.org/10.1038/nrmicro.2016.81

del Sol Mesa, A., Pazos, F., & Valencia, A. (2003). Automatic methods for predicting functionally important residues. *Journal of Molecular Biology, 326*(4), 1289-1302. doi:https://doi.org/10.1016/s0022-2836(02)01451-1

DeMaagd, G., & Philip, A. (2015). Parkinson's disease and its management: Part 1: Disease entity, risk factors, pathophysiology, clinical presentation, and diagnosis. *P & T : A Peer-Reviewed Journal for Formulary Management, 40*(8), 504-532.

Department of Health and Social Care. (2020). In MHRA, NHS(Eds.), *COVID-19 therapeutic alert* doi:Alert ref: CEM/CMO/2020/026

Devarasetti, P. K., Rajasekhar, L., Baisya, R., Sreejitha, K. S., & Vardhan, Y. K. (2021). A review of COVID-19 convalescent plasma use in COVID-19 with focus on proof of efficacy. *Immunologic Research, 69*(1), 18-25. doi:https://doi.org/10.1007/s12026-020-09169-x

Di Minno, M. N., Ambrosino, P., Ambrosini, F., Tremoli, E., Di Minno, G., & Dentali, F. (2016). Prevalence of deep vein thrombosis and pulmonary embolism in patients with superficial vein thrombosis: A systematic review and meta-analysis. *Journal of Thrombosis and Haemostasis : JTH, 14*(5), 964-972. doi:https://doi.org/10.1111/jth.13279

Dietrich, W. (1996). Reducing thrombin formation during cardiopulmonary bypass: Is there a benefit of the additional anticoagulant action of aprotinin? *Journal of Cardiovascular Pharmacology, 27*(1), S50-S57. doi:https://doi.org/10.1097/00005344-199600001-00011

Dikic, I. (2017). Proteasomal and autophagic degradation systems. *Annual Review of Biochemistry, 86*, 193-224. doi:https://doi.org/10.1146/annurev-biochem-061516-044908

Dilruba, S., & Kalayda, G. V. (2016). Platinum-based drugs: Past, present and future. *Cancer Chemotherapy and Pharmacology, 77*, 1103-1124. doi:https://doi.org/10.1007/s00280-016-2976-z

Doi, K., Ikeda, M., Hayase, N., Moriya, K., Morimura, N., & COVID-UTH Study Group. (2020). Nafamostat mesylate treatment in combination with favipiravir for patients critically ill with

covid-19: A case series. *Critical Care, 24*(392) doi:https//doi.org/10.1186/s13054-020-03078-z

Domcke, S., Sinha, R., Levine, D. A., Sander, C., & Schultz, N. (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. *Nature Communications, 4*, 2126. doi:https://doi.org/10.1038/ncomms3126

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases, 20*(5), 533-534. doi:https://doi.org/10.1016/S1473-3099(20)30120-1

Doubeni, C. A., Doubeni, A. R., & Myers, A. E. (2016). Diagnosis and management of ovarian cancer. *American Family Physician, 93*(11), 937-944.

Drenberg, C. D., Gibson, A. A., Pounds, S. B., Shi, L., Rhinehart, D. P., Li, L., . . . Sparreboom, A. (2017). OCTN1 is a high-affinity carrier of nucleoside analogues. *Cancer Research, 7*(8), 2102-2111. doi:https://doi.org/10.1158/0008-5472.CAN-16-2548

Eckert, A., Böck, B. C., Tagscherer, K. E., Haas, T. L., Grund, K., Sykora, J., . . . W Roth, W. (2008). The PEA-15/PED protein protects glioblastoma cells from glucose deprivation induced apoptosis via the ERK/MAP kinase pathway. *Oncogene, 27*, 1155-1166. doi:https://doi.org/10.1038/sj.onc.1210732

Erener, S. (2020). Diabetes, infection risk and COVID-19. *Molecular Metabolism, 39*, 101044. doi:https://doi.org/10.1016/j.molmet.2020.101044

Esakandari, H., Nabi-Afjadi, M., Fakkari-Afjadi, J., Farahmandian, N., Miresmaeili, S. M., & Bahreini, E. (2020). A comprehensive review of COVID-19 characteristics. *Biological Procedures Online, 22*, 19. doi:https://doi.org/10.1186/s12575-020-00128-2

Estelles, A., Yokoyama, M., Nothias, F., Vincent, J. D., Glowinski, J., & Vernier, P. (1996). The major astrocytic phosphoprotein PEA-15 is encoded by two mRNAs conserved on their full length in mouse and human. *Journal of Biological Chemistry, 271*(25), 14800-14806. doi:https://doi.org/10.1074/jbc.271.25.14800

Esumi, M., Ishibashi, M., Yamaguchi, H., Nakajima, S., Tai, Y., Kikuta, S., . . . Wakita, T. (2015). Transmembrane serine protease TMPRSS2 activates hepatitis C virus infection. *Hepatology, 61*(2), 437-446. doi:https://doi.org/10.1002/hep.27426

Evans, W. E., & Relling, M. V. (1999). Pharmacogenomics: Translating functional genomics into rational therapeutics. *Science (New York, N.Y.), 286*(5439), 487-491. doi:https://doi.org/10.1126/science.286.5439.487

Fabrizi, F., Alfieri, C. M., Cerutti, R., Lunghi, G., & Messa, P. (2020). COVID-19 and acute kidney injury: A systematic review and meta-analysis. *Pathogens (Basel, Switzerland), 9*(12), 1052. doi:https://doi.org/10.3390/pathogens9121052

Fahmy, O. H., Daas, F. M., Salunkhe, V., Petrey, J. L., Cosar, E. F., Ramirez, J., & Akca, O. (2021). Is microthrombosis the main pathology in coronavirus disease 2019 severity?-A systematic review of the postmortem pathologic findings. *Critical Care Explorations, 3*(5), e0427. doi:https://doi.org/10.1097/CCE.0000000000000427

Feng, C., Ding, G., Jiang, H., Ding., Q., & Wen, H. (2015). Loss of MLH1 confers resistance to PI3Kβ inhibitors in renal clear cell carcinoma with SETD2 mutation. *Tumor Biology, 36*, 3457-3464. doi:https://doi.org/10.1007/s13277-014-2981-y

Feng, R., Zhao, H., Xu, J., & Shen, C. (2020). CD47: The next checkpoint target for cancer immunotherapy. *Critical Reviews in Oncology/Hematology, 152*, 103014. doi:https://doi.org/10.1016/j.critrevonc.2020.103014

Fiory, F., Formisano, P., Perruolo, G., & Beguinot, F. (2009). PED/PEA-15, a multifunctional protein controlling cell survival and glucose metabolism. *American Journal of Physiology - Endocrinology and Metabolism, 297*(3), E592-E601. doi:https://doi.org/10.1152/ajpendo.00228.2009

Follini, E., Marchesini, M., & Roti, G. (2019). Strategies to overcome resistance mechanisms in T-cell acute lymphoblastic leukemia. *International Journal of Molecular Sciences, 20*(12), 3021. doi:https://doi.org/10.3390/ijms20123021

Formisano, P., Perruolo, G., Libertini, S., Santopietro, S., Troncone, G., Raciti, G. A., . . . Beguinot, F. (2005). Raised expression of the antiapoptotic protein ped/pea-15 increases

susceptibility to chemically induced skin tumor development. *Oncogene, 24*(47), 7012-7021. doi:https://doi.org/10.1038/sj.onc.1208871

Formstecher, E., Ramos, J. W., Fauquet, M., Calderwood, D. A., Hsieh, J. C., Canton, B., . . . Chneiweiss, H. (2001). PEA-15 mediates cytoplasmic sequestration of ERK MAP kinase. *Developmental Cell, 1*(2), 239-250. doi:https://doi.org/10.1016/s1534-5807(01)00035-1

Franzolin, E., Pontarin, G., Rampazzo, C., Miazzi, C., Ferraro, P., Palumbo, E., . . . Bianchi, V. (2013). The deoxynucleotide triphosphohydrolase SAMHD1 is a major regulator of DNA precursor pools in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America, 110*(35), 14272-14277. doi:https://doi.org/10.1073/pnas.1312033110

Frazier, E. P., Isenberg, J. S., Shiva, S., Zhao, L., Schlesinger, P., Dimitry, J., . . . Frazier, W. A. (2011). Age-dependent regulation of skeletal muscle mitochondria by the thrombospondin-1 receptor CD47. *Matrix Biology : Journal of the International Society for Matrix Biology, 30*(2), 154-161. doi:https://doi.org/10.1016/j.matbio.2010.12.004

Fukuda, M., Gotoh, Y., & Nishida, E. (1997). Interaction of MAP kinase with MAP kinase kinase: Its possible role in the control of nucleocytoplasmic transport of MAP kinase. *EMBO Journal, 16*(8), 1901-1908. doi:https://doi.org/10.1093/emboj/16.8.1901

Fumeron, F., Péan, F., Driss, F., Balkau, B., Tichet, J., Marre, M., . . . Insulin Resistance Syndrome (DESIR) Study Group. (2006). Ferritin and transferrin are both predictive of the onset of hyperglycemia in men and women over 3 years: The data from an epidemiological study on the insulin resistance syndrome (DESIR) study. *Diabetes Care, 29*(9), 2090-2094. doi:https://doi.org/10.2337/dc06-0093

Funke, V., Lehmann-Koch, J., Bickeböller, M., Benner, A., Tagscherer, K. E., Grund, K., . . . Roth, W. (2013). The PEA-15/PED protein regulates cellular survival and invasiveness in colorectal carcinomas. *Cancer Letters, 335*(2), 431-440. doi:https://doi.org/10.1016/j.canlet.2013.02.053

Furfaro, A. L., Traverso, N., Domenicotti, C., Piras, S., Moretta, L., Marinari, U. M., . . . Nitti, M. (2015). The Nrf2/HO-1 axis in cancer cell growth and chemoresistance. *Oxidative Medicine and Cellular Longevity, 2016*(1958174) doi:https://doi.org/10.1155/2016/1958174

Gandhi, V., Plunkett, W., Rodriguez Jr, C. O., Nowak, B. J., Du, M., Ayres, M., . . . Keating, M. J. (1998). Compound GW506U78 in refractory hematologic malignancies: Relationship between cellular pharmacokinetics and clinical response. *Journal of Clinical Oncology, 16*(11), 3607-3615. doi:https://doi.org/10.1200/JCO.1998.16.11.3607

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., . . . Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Science Signalling, 6*(269), pl1. doi:https://doi.org/10.1126/scisignal.2004088

Gao, Q., Chen, K., Gao, L., Zheng, Y., & Yang, Y. G. (2016). Thrombospondin-1 signaling through CD47 inhibits cell cycle progression and induces senescence in endothelial cells. *Cell Death and Disease, 7*(9), e2368. doi:https://doi.org/10.1038/cddis.2016.155

Garnett, M. J., Edelman, E. J., Heidorn, S. J., Greenman, C. D., Dastur, A., Lau, K. W., . . . Benes, C. H. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature, 483*(7391), 570-575. doi:https://doi.org/10.1038/nature11005

Geeleher, P., Zhang, Z., Wang, F., Gruener, R. F., Nath, A., Morrison, G., . . . Huang, R. S. (2017). Discovering novel pharmacogenomic biomarkers by imputing drug response in cancer patients from large genomics studies. *Genome Research, 27*(10), 1743-1751. doi:https://doi.org/10.1101/gr.221077.117

Gene Ontology Consortium. (2021). The gene ontology resource: Enriching a GOld mine. *Nucleic Acids Research, 49*(D1), D325-D334. doi:https://doi.org/10.1093/nar/gkaa1113

Gettins, P. G. (2002). Serpin structure, mechanism, and function. *Chemical Reviews, 102*(12), 4751-4804. doi:https://doi.org/10.1021/cr010170+

Geynisman, D. M., Maranchie, J. K., Ball, M. W., Bratslavsky, G., & Singer, E. A. (2021). A 25 year perspective on the evolution and advances in an understanding of the biology, evaluation and treatment of kidney cancer. *Urologic Oncology, S1078-1439*(21), 00194-0. doi:https://doi.org/10.1016/j.urolonc.2021.04.038

Ghandi, M., Huang, F. W., Jané-Valbuena, J., Kryukov, G. V., Lo, C. C., McDonald, E. R., . . . Sellers, W. R. (2019). Next-generation characterization of the cancer cell line encyclopedia. *Nature, 569*(7757), 503-508. doi:https://doi.org/10.1038/s41586-019-1186-3

Ghimire, K., Li, Y., Chiba, T., Julovi, S. M., Li, J., Ross, M. A., . . . Rogers, N. M. (2020). CD47 promotes age-associated deterioration in angiogenesis, blood flow and glucose homeostasis. *Cells, 9*(7), 1695. doi:https://doi.org/10.3390/cells9071695

Ghosh, S., Dellibovi-Ragheb, T. A., Kerviel, A., Pak, E., Qiu, Q., Fisher, M., . . . Altan-Bonnet, N. (2020). B-coronaviruses use lysosomes for egress instead of the biosynthetic secretory pathway. *Cell, 183*(6), 1520-1535.e14. doi:https://doi.org/10.1016/j.cell.2020.10.039

Gille, H., Sharrocks, A. D., & Shaw, P. E. (1992). Phosphorylation of transcription factor p62TCF by MAP kinase stimulates ternary complex formation at c-fos promoter. *Nature, 358*, 414-417. doi:https://doi.org/10.1038/358414a0

Glading, A., Koziol, J. A., Krueger, J., & Ginsberg, M. H. (2007). PEA-15 inhibits tumor cell invasion by binding to extracellular signal-regulated kinase 1/2. *Cancer Research, 67*(4), 1536-1544. doi:https://doi.org/10.1158/0008-5472.CAN-06-1378

González-Pérez, A., & López-Bigas, N. (2011). Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, condel. *American Journal of Human Genetics, 88*(4), 440-449. doi:https://doi.org/10.1016/j.ajhg.2011.03.004

Gooskens, S. L., Gadd, S., Guidry Auvil, J. M., Gerhard, D. S., Khan, J., Patidar, R., . . . Perlman, E. J. (2015). TCF21 hypermethylation in genetically quiescent clear cell sarcoma of the kidney. *Oncotarget, 6*(18), 15828-15841. doi:https://doi.org/10.18632/oncotarget.4682

Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., . . . Krogan, N. J. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature, 583*, 459-468. doi:https://doi.org/10.1038/s41586-020-2286-9

Grabowski, F., Preibisch, G., Kochanczyk, M., & Lipniacki, T. (2021). SARS−CoV−2 variant under investigation 202012/01 has more than twofold replicative advantage. *medRxiv, 2020.12.28.20248906* doi:https://doi.org/10.1101/2020.12.28.20248906

Graves, D. T., & Kayal, R. A. (2008). Diabetic complications and dysregulated innate immunity. *Frontiers in Bioscience : A Journal and Virtual Library, 13*, 1227-1239. doi:https://doi.org/10.2741/2757

Gregg, J., & Fraizer, G. (2011). Transcriptional regulation of EGR1 by EGF and the ERK signaling pathway in prostate cancer cells. *Genes & Cancer, 2*(9), 900-909. doi:https://doi.org/10.1177/1947601911431885

Greig, F. H., & Nixon, G. F. (2014). Phosphoprotein enriched in astrocytes (PEA)-15: A potential therapeutic target in multiple disease states. *Pharmacology & Therapeutics, 143*(3), 265-274. doi:https://doi.org/10.1016/j.pharmthera.2014.03.006

Grenga, L., Gallais, F., Pible, O., Gaillard, J., Gouveia, D., Batina, H., . . . Armengaud, J. (2020). Shotgun proteomics analysis of SARS-CoV-2-infected cells and how it can optimize whole viral particle antigen production for vaccines. *Emerging Microbes & Infections, 9*(1), 1712-1721. doi:https://doi.org/10.1080/22221751.2020.1791737

GTEx Consortium. (2013). The genotype-tissue expression (GTEx) project. *Nature Genetics, 45*, 580-585. doi:https://doi.org/10.1038/ng.2653

Guan, W. J., Ni, Z. Y., Hu, Y., Liang, W. H., Ou, C. Q., He, J. X., . . . China Medical Treatment Expert Group for Covid-19. (2020). Clinical characteristics of coronavirus disease 2019 in china. *New England Journal of Medicine, 382*(18), 1708-1720. doi:https://doi.org/10.1056/NEJMoa2002032

Guièze, R., Robbe, P., Clifford, R., de Guibert, S., Pereira, B., Timbs, A., . . . Schuh, A. (2015). Presence of multiple recurrent mutations confers poor trial outcome of relapsed/refractory CLL. *Blood, 126*(18), 2110-2117. doi:https://doi.org/10.1182/blood-2015-05-647578

Guzik, T. J., Mohiddin, S. A., Dimarco, A., Patel, V., Savvatis, K., Marelli-Berg, F. M., . . . McInnes, I. B. (2020). COVID-19 and the cardiovascular system: Implications for risk assessment, diagnosis, and treatment options. *Cardiovascular Research, 116*(10), 1666-1687. doi:https://doi.org/10.1093/cvr/cvaa106

Hadid, T., Kafri, Z., & Al-Katib, A. (2020). Coagulation and anticoagulation in COVID-19. *Blood Reviews, 100761* doi:https://doi.org/10.1016/j.blre.2020.100761

Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., . . . Terrier, B. (2020). Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science (New York, N.Y.), 369*(6504), 718-724. doi:https://doi.org/10.1126/science.abc6027

Haibe-Kains, B., El-Hachem, N., Birkbak, N. J., Jin, A. C., Beck, A. H., Aerts, H. J., & Quackenbush, J. (2013). Inconsistency in large pharmacogenomic studies. *Nature, 504*(7480), 389-393. doi:https://doi.org/10.1038/nature12831

Hamer, M., Gale, C. R., Kivimäki, M., & Batty, G. D. (2020). Overweight, obesity, and risk of hospitalization for COVID-19: A community-based cohort study of adults in the united kingdom. *Proceedings of the National Academy of Sciences of the United States of America, 117*(35), 21011-21013. doi:https://doi.org/10.1073/pnas.2011086117

Hammer, A. S., Quaade, M. L., Rasmussen, T. B., Fonager, J., Rasmussen, M., Mundbjerg, K., . . . Bøtner, A. (2020). SARS-CoV-2 transmission between mink (neovison vison) and humans, denmark. *Emerging Infectious Diseases, 27*(2), 10.3201/eid2702.203794. doi:https://doi.org/10.3201/eid2702.203794

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell, 100*(1), 57-70. doi:https://doi.org/10.1016/S0092-8674(00)81683-9

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: The next generation. *Cell, 144*(5), 646-674. doi:https://doi.org/10.1016/j.cell.2011.02.013

Hannenhalli, S. S., & Russell, R. B. (2000). Analysis and prediction of functional sub-types from protein sequence alignments. *Journal of Molecular Biology, 303*(1), 61-76. doi:https://doi.org/10.1006/jmbi.2000.4036

Hao, C., Beguinot, F., Condorelli, G., Trencia, A., Van Meir, E. G., Yong, V. W., . . . Petruk, K. C. (2001). Induction and intracellular regulation of tumor necrosis factor related apoptosis-inducing ligand (TRAIL) mediated apoptosis in human malignant glioma cells. *Cancer Research, 61*(3), 1162-1170.

He, B., & Garmire, L. (2020). Prediction of repurposed drugs for treating lung injury in COVID-19. *F1000Research, 9*, 609. doi:https://doi.org/10.12688/f1000research.23996.2

Helm, C. W., & States, J. C. (2009). Enhancing the efficacy of cisplatin in ovarian cancer treatment - could arsenic have a role. *Journal of Ovarian Research, 2*(2) doi:https://doi.org/10.1186/1757-2215-2-2

Herold, N., Rudd, S. G., Ljungblad, L., Sanjiv, K., Myrberg, I. H., Paulin, C. B., . . . Schaller, T. (2017). Targeting SAMHD1 with the vpx protein to improve cytarabine therapy for hematological malignancies. *Nature Medicine, 23*(2), 256-263. doi:https://doi.org/10.1038/nm.4265

Herold, N., Rudd, S. G., Sanjiv, K., Kutzner, J., Bladh, J., Paulin, C., . . . Schaller, T. (2017). SAMHD1 protects cancer cells from various nucleoside-based antimetabolites. *Cell Cycle (Georgetown, Tex.), 16*(11), 1029-1038. doi:https://doi.org/10.1080/15384101.2017.1314407

Herold, N., Rudd, S. G., Sanjiv, K., Kutzner, J., Myrberg, I. H., Paulin, C., . . . Schaller, T. (2017). With me or against me: Tumor suppressor and drug resistance activities of SAMHD1. *Experimental Hematology, 52*, 32-39. doi:https://doi.org/10.1016/j.exphem.2017.05.001

Herold, T., Baldus, C. D., & Gökbuget, N. (2014). Ph-like acute lymphoblastic leukemia in older adults. *The New England Journal of Medicine, 371*(23), 2235. doi:https://doi.org/10.1056/NEJMc1412123#SA1

Herold, T., Schneider, S., Metzeler, K. H., Neumann, M., Hartmann, L., Roberts, K. G., . . . Gökbuget, N. (2017). Adults with philadelphia chromosome-like acute lymphoblastic leukemia frequently have IGH-CRLF2 and JAK2 mutations, persistence of minimal residual disease and poor prognosis. *Haematologica, 102*(1), 130-138. doi:https://doi.org/10.3324/haematol.2015.136366

Hikmet, F., Méar, L., Edvinsson, Å, Micke, P., Uhlén, M., & Lindskog, C. (2020). The protein expression profile of ACE2 in human tissues. *Molecular Systems Biology, 16*(7), e9610. doi:https://doi.org/10.15252/msb.20209610

Hirsch, C., & Schildknecht, S. (2019). In vitro research reproducibility: Keeping up high standards. *Frontiers in Pharmacology, 10*, 1484. doi:https://doi.org/10.3389/fphar.2019.01484

Hoadley, K. A., Yau, C., Wolf, D. M., Cherniack, A. D., Tamborero, D., Ng, S., . . . Stuart, J. M. (2014). Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell, 158*(4), 929-944. doi:https://doi.org/10.1016/j.cell.2014.06.049

Hoehl, S., Berger, A., Kortenbusch, M., Cinatl Jr, J., Bojkova, D., Rabenau, H., . . . Ciesek, S. (2020). Evidence of SARS-CoV-2 infection in returning travelers from wuhan, china. *New England Journal of Medicine, 382*(13), 1278-1280. doi:https://doi.org/10.1056/NEJMc2001899

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., . . . Pöhlmann, S. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell, 181*(2), 271-280. doi:https://doi.org/10.1016/j.cell.2020.02.052

Hoffmann, M., Schroeder, S., Kleine-Weber, H., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020a). Nafamostat mesylate blocks activation of SARS-CoV-2: New treatment option for COVID-19. *Antimicrobial Agents and Chemotherapy, 64*(6), e00754-20. doi:https://doi.org/10.1128/AAC.00754-20

Hoffmann, M., Schroeder, S., Kleine-Weber, H., Müller, M. A., Drosten, C., & Pöhlmann, S. (2020b). Nafamostat mesylate blocks activation of SARS-CoV-2: New treatment option for COVID-19. *Antimicrobial Agents and Chemotherapy, 64*(e00754-20) doi:https://doi.org/10.1128/AAC.00754-20

Hokello, J., Sharma, A. L., Shukla, G. C., & Tyagi, M. (2020). A narrative review on the basic and clinical aspects of the novel SARS-CoV-2, the etiologic agent of COVID-19. *Annals of Translational Medicine, 8*(24), 1686. doi:https://doi.org/10.21037/atm-20-5272

Hollenbaugh, J. A., Shelton, J., Tao, S., Amiralaei, S., Liu, P., Lu, X., . . . Kim, B. (2017). Substrates and inhibitors of SAMHD1. *PLoS One, 12*(1), e0169052. doi:https://doi.org/10.1371/journal.pone.0169052

Holmes, E. (2020). Novel 2019 coronavirus genome. Retrieved from www.virological.org/t/novel-2019-coronavirus-genome/319

Homminga, I., Zwaan, C. M., Manz, C. Y., Parker, C., Bantia, S., Smits, W. K., . . . Meijerink, J. P. (2011). In vitro efficacy of forodesine and nelarabine (ara-G) in pediatric leukemia. *Blood, 118*(8), 2184-2190. doi:https://doi.org/10.1182/blood-2011-02-337840

Hosny, N., Matson, A. W., Kumbha, R., Steinhoff, M., Sushil Rao, J., El-Abaseri, T. B., . . . Burlak, C. (2021). 3'UTR enhances hCD47 cell surface expression, self-signal function, and

reduces ER stress in porcine fibroblasts. *Xenotransplantation, 28*(1), e12641. doi:https://doi.org/10.1111/xen.12641

Hrecka, K., Hao, C., Gierszewska, M., Swanson, S. K., Kesik-Brodacka, M., Srivastava, S., . . . Skowronski, J. (2011). Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature, 474*(7353), 658-661. doi:https://doi.org/10.1038/nature10195

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., . . . Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *Lancet (London, England), 395*(10223), 497-506. doi:https://doi.org/10.1016/S0140-6736(20)30183-5

Huang, F. W., & Feng, F. Y. (2019). A tumor-agnostic NTRK (TRK) inhibitor. *Cell, 177*(1), 8. doi:https://doi.org/10.1016/j.cell.2019.02.049

Huang, S., Wang, J., Liu, F., Liu, J., Cao, G., Yang, C., . . . Xiong, B. (2020). COVID-19 patients with hypertension have more severe disease: A multicenter retrospective observational study. *Hypertension Research : Official Journal of the Japanese Society of Hypertension, 43*(8), 824-831. doi:https://doi.org/10.1038/s41440-020-0485-2

Huang, X., Pearce, R., & Zhang, Y. (2020). De novo design of protein peptides to block association of the SARS-CoV-2 spike protein with human ACE2. *Aging, 12*(2), 11263-11276. doi:https://doi.org/10.18632/aging.103416

Hussien, M. A., & Abdelaziz, A. (2020). Molecular docking suggests repurposing of brincidofovir as a potential drug targeting SARS-CoV-2 ACE2 receptor and main protease. *Network Modeling and Analysis in Health Informatics and Bioinformatics, 9*(1), 56. doi:https://doi.org/10.1007/s13721-020-00263-6

Iba, T., Levy, J. H., Levi, M., & Thachil, J. (2020). Coagulopathy in COVID-19. *Journal of Thrombosis and Haemostasis : JTH, 18*(9), 2103-2109. doi:https://doi.org/10.1111/jth.14975

Iorio, F., Knijnenburg, T. A., Vis, D. J., Bignell, G. R., Menden, M. P., Schubert, M., . . . Garnett, M. J. (2016). A landscape of pharmacogenomic interactions in cancer. *Cell, 166*(3), 740-754. doi:https://doi.org/10.1016/j.cell.2016.06.017

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., & Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics, 4*(2), 249-264. doi:https://doi.org/10.1093/biostatistics/4.2.249

Irwin, J. J., & Shoichet, B. K. (2005). ZINC--a free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling, 45*(1), 177-182. doi:https://doi.org/10.1021/ci049714+

Isenberg, J. S., Frazier, W. A., & Roberts, D. D. (2008). Thrombospondin-1: A physiological regulator of nitric oxide signaling. *Cellular and Molecular Life Sciences : CMLS, 65*(5), 728-742. doi:https://doi.org/10.1007/s00018-007-7488-x

Isenberg, J. S., Hyodo, F., Pappan, L. K., Abu-Asab, M., Tsokos, M., Krishna, M. C., . . . Roberts, D. D. (2007). Blocking thrombospondin-1/CD47 signaling alleviates deleterious effects of aging on tissue responses to ischemia. *Arteriosclerosis, Thrombosis and Vascular Biology, 27*(12), 2582-2588. doi:https://doi.org/10.1161/ATVBAHA.107.155390

Isenberg, J. S., Qin, Y., Maxhimer, J. B., Sipes, J. M., Despres, D., Schnermann, J., . . . Roberts, D. D. (2009). Thrombospondin-1 and CD47 regulate blood pressure and cardiac responses to vasoactive stress. *Matrix Biology : Journal of the International Society for Matrix Biology, 28*(2), 110-119. doi:https://doi.org/10.1016/j.matbio.2009.01.002

Islam, A. B. M. M. K., & Khan, M. A. (2020). Lung transcriptome of a COVID-19 patient and systems biology predictions suggest impaired surfactant production which may be druggable by surfactant therapy. *Scientific Reports, 10*(19395) doi:https://doi.org/10.1038/s41598-020-76404-8

Jabbour, E., Kantarjian, H., Ravandi, F., Thomas, D., Huang, X., Faderl, S., . . . O'Brien, S. (2015). Combination of hyper-CVAD with ponatinib as first-line therapy for patients with philadelphia chromosome-positive acute lymphoblastic leukaemia: A single-centre, phase 2 study. *The Lancet. Oncology, 16*(15), 1547-1555. doi:https://doi.org/10.1016/S1470-2045(15)00207-7

Jabbour, E., Short, N. J., Ravandi, F., Huang, X., Daver, N., DiNardo, C. D., . . . Kantarjian, H. (2018). Combination of hyper-CVAD with ponatinib as first-line therapy for patients with philadelphia chromosome-positive acute lymphoblastic leukaemia: Long-term follow-up of

a single-centre, phase 2 study. *The Lancet Haematology, 5*(12), e618-e627. doi:https://doi.org/10.1016/S2352-3026(18)30176-5

Jaén, R. I., Prieto, P., Casado, M., Martín-Sanz, P., & Boscá, L. (2018). Post-translational modifications of prostaglandin-endoperoxide synthase 2 in colorectal cancer: An update. *World Journal of Gastroenterology, 24*(48), 5454-5461. doi:https://doi.org/10.3748/wjg.v24.i48.5454

Jaksik, R., Iwanaszko, M., Rzeszowska-Wolny, J., & Kimmel, M. (2015). Microarray experiments and factors which affect their reliability. *Biology Direct, 10*, 46. doi:https://doi.org/10.1186/s13062-015-0077-2

Jiang, H., Li, C., Liu, Z., & Shengjing Hospital, H. (2020). Expression and relationship of SAMHD1 with other apoptotic and autophagic genes in acute myeloid leukemia patients. *Acta Haematologica, 143*(1), 51-59. doi:https://doi.org/10.1159/000500822

Jin, C., Peng, X., Liu, F., Cheng, L., Lu, X., Yao, H., . . . Wu, N. (2014). MicroRNA-181 expression regulates specific post-transcriptional level of SAMHD1 expression in vitro. *Biochemical and Biophysical Research Communications, 452*(3), 760-767. doi:https://doi.org/10.1016/j.bbrc.2014.08.151

Jin, Y. Z., Pei, C. Z., & Wen, L. Y. (2016). FLNA is a predictor of chemoresistance and poor survival in cervical cancer. *Biomarkers in Medicine, 10*(7), 711-719. doi:https://doi.org/10.2217/bmm-2016-0056

Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., . . . Yang, H. (2020). Structure of mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature, 582*(7811), 289-293. doi:https://doi.org/10.1038/s41586-020-2223-y

Johansson, M. A., Quandelacy, T. M., Kada, S., Prasad, P. V., Steele, M., Brooks, J. T., . . . Butler, J. C. (2021). SARS-CoV-2 transmission from people without COVID-19 symptoms. *JAMA Network Open, 4*(1), e2035057. doi:https://doi.org/10.1001/jamanetworkopen.2020.35057

Johansson, P., Klein-Hitpass, L., Choidas, A., Habenberger, P., Mahboubi, B., Kim, B., . . . Dürig, J. (2018). SAMHD1 is recurrently mutated in T-cell prolymphocytic leukemia. *Blood Cancer Journal, 8*(11) doi:https://doi.org/10.1038/s41408-017-0036-5

Johnson, G. L., & Lapadat, R. (2002). Mitogen-activated protein kinase pathways mediated by ERK, JNK, and p38 protein kinases. *Science, 298*, 1911-1912. doi:https://doi.org/10.1126/science.1072682

Kadia, T. M., & Gandhi, V. (2017). Nelarabine in the treatment of pediatric and adult patients with T-cell acute lymphoblastic leukemia and lymphoma. *Expert Review of Hematology, 10*(1), 1-8. doi:https://doi.org/10.1080/17474086.2017.1262757

Kalil, A. C., Patterson, T. F., Mehta, A. K., Tomashek, K. M., Wolfe, C. R., Ghazaryan, V., . . . ACTT-2 Study Group Members. (2021). Baricitinib plus remdesivir for hospitalized adults with covid-19. *The New England Journal of Medicine, 384*(9), 795-807. doi:https://doi.org/10.1056/NEJMoa2031994

Kalligeros, M., Shehadeh, F., Mylona, E. K., Benitez, G., Beckwith, C. G., Chan, P. A., & Mylonakis, E. (2020). Association of obesity with disease severity among patients with coronavirus disease 2019. *Obesity (Silver Spring, Md.), 28*(7), 1200-1204. doi:https://doi.org/10.1002/oby.22859

Kamitani, W., Narayanan, K., Huang, C., Lokugamage, K., Ikegami, T., Ito, N., . . . Makino, S. (2006). Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. *Proceedings of the National Academy of Sciences of the United States of America, 103*(34), 12885-12890. doi:https://doi.org/10.1073/pnas.0603144103

Kamyshnyi, A., Krynytska, I., Matskevych, V., Marushchak, M., & Lushchak, O. (2020). Arterial hypertension as a risk comorbidity associated with COVID-19 pathology. *International Journal of Hypertension, 8019360* doi:https://doi.org/10.1155/2020/8019360

Kannaiyan, R., & Mahadevan, D. (2018). A comprehensive review of protein kinase inhibitors for cancer therapy. *Expert Review of Anticancer Therapy, 18*(12), 1249-1270. doi:https://doi.org/10.1080/14737140.2018.1527688

Kantarjian, H. M., DeAngelo, D. J., Advani, A. S., Stelljes, M., Kebriaei, P., Cassaday, R. D., . . . Marks, D. I. (2017). Hepatic adverse event profile of inotuzumab ozogamicin in adult patients with relapsed or refractory acute lymphoblastic leukaemia: Results from the open-label, randomised, phase 3 INO-VATE study. *The Lancet Haematology, 4*(8), e387-e398. doi:https://doi.org/10.1016/S2352-3026(17)30103-5

Karagkouni, D., Paraskevopoulou, M. D., Chatzopoulos, S., Vlachos, I. S., Tastsoglou, S., Kanellos, I., . . . Hatzigeorgiou, A. G. (2018). DIANA-TarBase v8: A decade-long collection of experimentally supported miRNA-gene interactions. *Nucleic Acids Research, 46*(D1), D239-D245. doi:https://doi.org/10.1093/nar/gkx1141

Karlsson, M., Mathers, J., Dickinson, R. J., Mandl, M., & Keyse, S. M. (2004). Both nuclear cytoplasmic shuttling of the dual specificity phosphatase MKP-3 and its ability to anchor MAP kinase in the cytoplasm are mediated by a conserved nuclear export signal. *Journal of Biological Chemistry, 279*(40), 41882-41891. doi:https://doi.org/10.1074/jbc.M406720200

Karmouty-Quintana, H., Thandavarayan, R. A., Keller, S. P., Sahay, S., Pandit, L. M., & Akkanti, B. (2020). Emerging mechanisms of pulmonary vasoconstriction in SARS-CoV-2-induced acute respiratory distress syndrome (ARDS) and potential therapeutic targets. *International Journal of Molecular Sciences, 21*(21), 8081. doi:https://doi.org/10.3390/ijms21218081

Kaur, S., Cicalese, K. V., Bannerjee, R., & Roberts, D. D. (2020). Preclinical and clinical development of therapeutic antibodies targeting functions of CD47 in the tumor microenvironment. *Antibody Therapeutics, 3*(3), 179-192. doi:https://doi.org/10.1093/abt/tbaa017

Kawabata, H. (2019). Transferrin and transferrin receptors update. *Free Radical Biology and Medicine, 133*, 46-54. doi:https://doi.org/10.1016/j.freeradbiomed.2018.06.037

Kazemi, M., Salehi, M., & Kheirollahi, M. (2016). Down syndrome: Current status, challenges and future perspectives. *International Journal of Molecular and Cellular Medicine, 5*(3), 125-133.

Keating, G. M. (2014). Bevacizumab: A review of its use in advanced cancer. *Drugs, 74*, 1891-1925. doi:https://doi.org/10.1007/s40265-014-0302-9

Kelada, M., Anto, A., Dave, K., & Saleh, S. N. (2020). The role of sex in the risk of mortality from COVID-19 amongst adult patients: A systematic review. *Cureus, 12*(8), e10114. doi:https://doi.org/10.7759/cureus.10114

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., & Sternberg, M. J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols, 10*(6), 845-858. doi:https://doi.org/10.1038/nprot.2015.053

Kemp, S. A., Collier, D. A., Datir, R., Ferreira, I., Gayed, S., Jahun, A., . . . Gupta, R. K. (2020). Neutralising antibodies in spike mediated SARS-CoV-2 adaptation. *medRxiv,* doi:https://doi.org/10.1101/2020.12.05.20241927

Khan, D., & Ansar Ahmed, S. (2016). The immune system is a natural target for estrogen action: Opposing effects of estrogen in two prototypical autoimmune diseases. *Frontiers in Immunology, 6*(635) doi:https://doi.org/10.3389/fimmu.2015.00635

Khatri, I., Staal, F., & van Dongen, J. (2020). Blocking of the high-affinity interaction-synapse between SARS-CoV-2 spike and human ACE2 proteins likely requires multiple high-affinity antibodies: An immune perspective. *Frontiers in Immunology, 11*(570018) doi:https://doi.org/10.3389/fimmu.2020.570018

Kim, E., Jung, S., Park, W. S., Lee, J. H., Shin, R., Heo, S. C., . . . Chai, Y. J. (2019). Upregulation of SLC2A3 gene and prognosis in colorectal carcinoma: Analysis of TCGA data. *BMC Cancer, 19*(1), 302. doi:https://doi.org/10.1186/s12885-019-5475-x

Kim, J. A., Hwang, B., Park, S. N., Huh, S., Im, K., Choi, S., . . . Lee, D. S. (2016). Genomic profile of chronic lymphocytic leukemia in korea identified by targeted sequencing. *PLoS One, 11*(12), e0167641. doi:https://doi.org/10.1371/journal.pone.0167641

Kim, Y. A., Madan, S., & Przytycka, T. M. (2017). WeSME: Uncovering mutual exclusivity of cancer drivers and beyond. *Bioinformatics (Oxford, England), 33*(6), 814-821. doi:https://doi.org/10.1093/bioinformatics/btw242

Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., & Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics, 46*, 310-315. doi:https://doi.org/10.1038/ng.2892

Klann, K., Tascher, G., & Münch, C. (2020). Functional translatome proteomics reveal converging and dose-dependent regulation by mTORC1 and eIF2α. *Molecular Cell, 77*(4), 913-925. doi:https://doi.org/10.1016/j.molcel.2019.11.010

Klein, S., & Flanagan, K. (2016). Sex differences in immune responses. *Nature Reviews Immunology, 16*, 626-638. doi:https://doi.org/10.1038/nri.2016.90

Knecht, K. M., Buzovetsky, O., Schneider, C., Thomas, D., Srikanth, V., Kaderali, L., . . . Xiong, Y. (2018). The structural basis for cancer drug interactions with the catalytic and allosteric sites of SAMHD1. *Proceedings of the National Academy of Sciences of the United States of America, 115*(43), E10022-E10031. doi:https://doi.org/10.1073/pnas.1805593115

Ko, P. J., & Dixon, S. J. (2018). Protein palmitoylation and cancer. *EMBO Reports, 19*(10), e46666. doi:https://doi.org/10.15252/embr.201846666

Kodigepalli, K. M., Bonifati, S., Tirumuru, N., & Wu, L. (2018). SAMHD1 modulates in vitro proliferation of acute myeloid leukemia-derived THP-1 cells through the PI3K-akt-p27 axis. *Cell Cycle (Georgetown, Tex.), 17*(9), 1124-1137. doi:https://doi.org/10.1080/15384101.2018.1480218

Kohnken, R., Kodigepalli, K. M., Mishra, A., Porcu, P., & Wu, L. (2017). MicroRNA-181 contributes to downregulation of SAMHD1 expression in CD4+ T-cells derived from sèzary syndrome patients. *Leukemia Research, 52*, 58-66. doi:https://doi.org/10.1016/j.leukres.2016.11.010

Kohnken, R., Kodigepalli, K. M., & Wu, L. (2015). Regulation of deoxynucleotide metabolism in cancer: Novel mechanisms and therapeutic implications. *Molecular Cancer, 14*, 176. doi:https://doi.org/10.1186/s12943-015-0446-6

Konno, Y., Kimura, I., Uriu, K., Fukushi, M., Irie, T., Koyanagi, Y., . . . Sato, K. (2020). SARS-CoV-2 ORF3b is a potent interferon antagonist whose activity is increased by a naturally occurring elongation variant. *Cell Reports, 32*(12), 108185. doi:https://doi.org/10.1016/j.celrep.2020.108185

Koralnik, I. J., & Tyler, K. L. (2020). COVID-19: A global threat to the nervous system. *Annals of Neurology, 88*, 1-11. doi:https://doi.org/10.1002/ana.25807

Koussounadis, A., Langdon, S., Um, I., Harrison, D. J., & Smith, V. A. (2015). Relationship between differentially expressed mRNA and mRNA-protein correlations in a xenograft model system. *Scientific Reports, 5*, 10775. doi:https://doi.org/10.1038/srep10775

Kowalewski, J., & Ray, A. (2020). Predicting novel drugs for SARS-CoV-2 using machine learning from a >10 million chemical space. *Heliyon, 6*(8), e04639. doi:https://doi.org/10.1016/j.heliyon.2020.e04639

Kubes, M., Cordier, J., Glowinski, J., Girault, J. A., & Chneiweiss, H. (1998). Endothelin induces a calcium-dependent phosphorylation of PEA-15 in intact astrocytes: Identification of Ser104 and Ser116 phosphorylated, respectively, by protein kinase C and calcium/calmodulin kinase II in vitro. *Journal of Neurochemistry, 71*(3), 1307-1314. doi:https://doi.org/10.1046/j.1471-4159.1998.71031307.x

Kuitunen, A., Hiippala, S., Vahtera, E., Rasi, V., & Salmenperä, M. (2005). The effects of aprotinin and tranexamic acid on thrombin generation and fibrinolytic response after cardiac surgery. *Acta Anaesthesiologica Scandinavica, 49*(9), 1272-1279. doi:https://doi.org/10.1111/j.1399-6576.2005.00809.x

Kumamoto, T., Goto, H., Ogawa, C., Hori, T., Deguchi, T., Araki, T., . . . Toyoda, H. (2020). FLEND (nelarabine, fludarabine, and etoposide) for relapsed T-cell acute lymphoblastic leukemia in children: A report from japan children's cancer group. *International Journal of Hematology, 112*(5), 720-724. doi:https://doi.org/10.1007/s12185-020-02962-2

Kuo, T. C., Chen, A., Harrabi, O., Sockolosky, J. T., Zhang, A., Sangalang, E., . . . Wan, H. I. (2020). Targeting the myeloid checkpoint receptor SIRPα potentiates innate and adaptive immune responses to promote anti-tumor activity. *Journal of Hematology and Oncology, 13*(160) doi:https://doi.org/10.1186/s13045-020-00989-w

Lagarde, N., Rey, J., Gyulkhandanyan, A., Tufféry, P., Miteva, M. A., & Villoutreix, B. O. (2018). Online structure-based screening of purchasable approved drugs and natural compounds: Retrospective examples of drug repositioning on cancer targets. *Oncotarget, 9*(64), 32346-32361. doi:https://doi.org/10.18632/oncotarget.25966

Laguette, N., Sobhian, B., Casartelli, N., Ringeard, M., Chable-Bessia, C., Ségéral, E., . . . Benkirane, M. (2011). SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by vpx. *Nature, 474*(7353), 654-657. doi:https://doi.org/10.1038/nature10117

Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., . . . Getz, G. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature, 499*, 214-218. doi:https://doi.org/10.1038/nature12213

Lee, E. J., Seo, J. H., Park, J. H., Vo, T., An, S., Bae, S. J., . . . Kim, K. W. (2017). SAMHD1 acetylation enhances its deoxynucleotide triphosphohydrolase activity and promotes cancer cell proliferation. *Oncotarget, 8*(40), 68517-68529. doi:https://doi.org/10.18632/oncotarget.19704

Lee, J., Bartholomeusz, C., Krishnamurthy, S., Liu, P., Saso, H., LaFortune, T. A., . . . Ueno, N. T. (2012). PEA-15 unphosphorylated at both serine 104 and serine 116 inhibits ovarian cancer cell tumorigenicity and progression through blocking β-catenin. *Oncogenesis, 1*, e22. doi:https://doi.org/10.1038/oncsis.2012.22

Leeming, D. J., Genovese, F., Sand, J., Rasmussen, D., Christiansen, C., Jenkins, G., . . . Karsdal, M. A. (2021). Can biomarkers of extracellular matrix remodelling and wound healing be used to identify high risk patients infected with SARS-CoV-2?: Lessons learned from pulmonary fibrosis. *Respiratory Research, 22*(1), 38. doi:https://doi.org/10.1186/s12931-020-01590-y

Lega, S., Naviglio, S., Volpi, S., & Tommasini, A. (2020). Recent insight into SARS-CoV2 immunopathology and rationale for potential treatment and preventive strategies in COVID-19. *Vaccines, 8*(2), 224. doi:https://doi.org/10.3390/vaccines8020224

Lenormand, P., Sardet, C., Pages, G., L'Allemain, G., Brunet, A., & Pouyssegur, J. (1993). Growth factors induce nuclear translocation of MAP kinases (p42mapk and p44mapk) but not of their activator MAP kinase kinase (p45mapkk) in fibroblasts. *Journal of Cell Biology, 122*(5), 1079-1088. doi:https://doi.org/10.1083/jcb.122.5.1079

Letko, M., Marzi, A., & Munster, V. (2020). Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. *Nature Microbiology, 5*(4), 562-569. doi:https://doi.org/10.1038/s41564-020-0688-y

Letrud, K., & Hernes, S. (2019). Affirmative citation bias in scientific myth debunking: A three-in-one case study. *PLoS One, 14*(9), e0222213. doi:https://doi.org/10.1371/journal.pone.0222213

Levi, M., Thachil, J., Iba, T., & Levy, J. H. (2020). Coagulation abnormalities and thrombosis in patients with COVID-19. *The Lancet Haematology, 7*(6), e438-e440. doi:https://doi.org/10.1016/S2352-3026(20)30145-9

Levine, A. J. (2020). P53: 800 million years of evolution and 40 years of discovery. *Nature Reviews Cancer, 20*(8), 471-480. doi:https://doi.org/10.1038/s41568-020-0262-1

Levy, J. H., Bailey, J. M., & Salmenperä, M. (1994). Pharmacokinetics of aprotinin in preoperative cardiac surgical patients. *Anesthesiology, 80*, 1013-1018. doi:https://doi.org/10.1097/00000542-199405000-00010

Li, H., Ning, S., Ghandi, M., Kryukov, G. V., Gopal, S., Deik, A., . . . Sellers, W. R. (2019). The landscape of cancer cell line metabolism. *Nature Medicine, 25*(5), 850-860. doi:https://doi.org/10.1038/s41591-019-0404-8

Li, J., Guo, M., Tian, X., Wang, X., Yang, X., Wu, P., . . . Liang, Q. (2020). Virus-host interactome and proteomic survey reveal potential virulence factors influencing SARS-CoV-2 pathogenesis. *Med (New York, N.Y.), 10.1016/j.medj.2020.07.002* doi:https://doi.org/10.1016/j.medj.2020.07.002

Li, N., Zhang, W., & Cao, X. (2000). Identification of human homologue of mouse IFN-gamma induced protein from human dendritic cells. *Immunology Letters, 74*(3), 221-224. doi:https://doi.org/10.1016/s0165-2478(00)00276-5

Li, Q., Wang, Z., Zheng, Q., & Liua, S. (2020). Potential clinical drugs as covalent inhibitors of the priming proteases of the spike protein of SARS-CoV-2. *Computational and Structural Biotechnology Journal, 18*, 2200-2208. doi:https://doi.org/10.1016/j.csbj.2020.08.016

Li, S., Zhang, Y., Guan, Z., Li, H., Ye, M., Chen, X., . . . Peng, K. (2020). SARS-CoV-2 triggers inflammatory responses and cell death through caspase-8 activation. *Signal Transduction and Targeted Therapy, 5*, 235. doi:https://doi.org/10.1038/s41392-020-00334-0

Li, W., Moore, M. J., Vasilieva, N., Sui, J., Wong, S. K., Berne, M. A., . . . Farzan, M. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature, 426*(6965), 450-454. doi:https://doi.org/10.1038/nature02145

Li, X., & Ma, X. (2020). Acute respiratory failure in COVID-19: Is it “typical” ARDS? *Critical Care, 24*, 198. doi:https://doi.org/10.1186/s13054-020-02911-9

Li, Y., Shen, Y., Zhu, Z., Wen, H., & Feng, C. (2020). Comprehensive analysis of copy number variance and sensitivity to common targeted therapy in clear cell renal cell carcinoma: In silico analysis with in vitro validation. *Cancer Medicine, 9*(16), 6020-6029. doi:https://doi.org/10.1002/cam4.3281

Lichtarge, O., Bourne, H. R., & Cohen, F. E. (1996). An evolutionary trace method defines binding surfaces common to protein families. *Journal of Molecular Biology, 257*(2), 342-358. doi:https://doi.org/10.1006/jmbi.1996.0167

Littler, D. R., Gully, B. S., Colson, R. N., & Rossjohn, J. (2020). Crystal structure of the SARS-CoV-2 non-structural protein 9, Nsp9. *iScience, 23*(7), 101258. doi:https://doi.org/10.1016/j.isci.2020.101258

Liu, J., Lichtenberg, T., Hoadley, K. A., Poisson, L. M., Lazar, A. J., Cherniack, A. D., . . . Hu, H. (2018). Cancer genome atlas research network. an integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell, 173*(2), 400-416. doi:https://doi.org/10.1016/j.cell.2018.02.052

Liu, Y., Mao, B., Liang, S., Yang, J. W., Lu, H. W., Chai, Y. H., . . . Shanghai Clinical Treatment Experts Group for COVID-19. (2020). Association between age and clinical characteristics and outcomes of COVID-19. *The European Respiratory Journal, 55*(5), 2001112. doi:https://doi.org/10.1183/13993003.01112-2020

Liu, Y., Zhang, X., Han, C., Wan, G., Huang, X., Ivan, C., . . . Lu, X. (2015). TP53 loss creates therapeutic vulnerability in colorectal cancer. *Nature, 520*(7549), 697-701. doi:https://doi.org/10.1038/nature14418

Liu, Z., VanBlargan, L. A., Rothlauf, P. W., Bloyet, L. M., Chen, R. E., Stumpf, S., . . . Whelan, S. P. J. (2020). Landscape analysis of escape variants identifies SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *bioRxiv,* doi:https://doi.org/10.1101/2020.11.06.372037

Llovet, J. M., Kelley, R. K., Villanueva, A., Singal, A. G., Pikarsky, E., Roayaie, S., . . . Finn, R. S. (2021). Hepatocellular carcinoma. *Nature Reviews Disease Primers, 7*(6) doi:https://doi.org/10.1038/s41572-020-00240-3

Lockhart, S. M., & O'Rahilly, S. (2020). When two pandemics meet: Why is obesity associated with increased COVID-19 mortality? *Med (New York, N.Y.), 1*(1), 33-42. doi:https://doi.org/10.1016/j.medj.2020.06.005

Lokugamage, K. G., Hage, A., de Vries, M., Valero-Jimenez, A. M., Schindewolf, C., Dittmann, M., . . . Menachery, V. D. (2020). Type I interferon susceptibility distinguishes SARS-CoV-2 from SARS-CoV. *Journal of Virology, 94*(23), e01410-20. doi:https://doi.org/10.1128/JVI.01410-20

Lorenzi, P. L., Reinhold, W. C., Varma, S., Hutchinson, A. A., Pommier, Y., Chanock, S. J., & Weinstein, J. N. (2009). DNA fingerprinting of the NCI-60 cell line panel. *Molecular Cancer Therapeutics, 8*(4), 713-724. doi:https://doi.org/10.1158/1535-7163.MCT-08-0921

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., . . . Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *Lancet (London, England), 395*(10224), 565-574. doi:https://doi.org/10.1016/S0140-6736(20)30251-8

Luan, J., Lu, Y., Jin, X., & Zhang, L. (2020). Spike protein recognition of mammalian ACE2 predicts the host range and an optimized ACE2 for SARS-CoV-2 infection. *Biochemical and Biophysical Research Communications, 526*(1), 165-169. doi:https://doi.org/10.1016/j.bbrc.2020.03.047

Luck, A. N., & Mason, A. B. (2012). Transferrin-mediated cellular iron delivery. *Current Topics in Membranes, 69*, 3-35. doi:https://doi.org/10.1016/B978-0-12-394390-3.00001-X

Lui, G., Shaw, R., Schaub, F. X., Stork, I. N., Gurley, K. E., Bridgwater, C., . . . Grandori, C. (2020). BET, SRC, and BCL2 family inhibitors are synergistic drug combinations with PARP inhibitors in ovarian cancer. *EBioMedicine, 60*, 102988. doi:https://doi.org/10.1016/j.ebiom.2020.102988

Lum, J. B., Infante, A. J., Makker, D. M., Yang, F., & Bowman, B. H. (1986). Transferrin synthesis by inducer T lymphocytes. *The Journal of Clinical Investigation, 77*(3), 841-849. doi:https://doi.org/10.1172/JCI112381

Luoh, S. W., & Flaherty, K. T. (2018). When tissue is no longer the issue: Tissue-agnostic cancer therapy comes of age. *Annals of Internal Medicine, 169*(4), 233-239. doi:https://doi.org/10.7326/M17-2832

Ly, T., Ahmad, Y., Shlien, A., Soroka, D., Mills, A., Emanuele, M. J., . . . Lamond, A. I. (2014). A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *eLife, 3*, e01630. doi:https://doi.org/10.7554/eLife.01630

Ma, X., Liu, Y., Liu, Y., Alexandrov, L. B., Edmonson, M. N., Gawad, C., . . . Zhang, J. (2018). Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature, 555*(7696), 371-376. doi:https://doi.org/10.1038/nature25795

Maggio, R., & Corsini, G. U. (2020). Repurposing the mucolytic cough suppressant and TMPRSS2 protease inhibitor bromhexine for the prevention and management of SARS-CoV-2 infection. *Pharmacological Research, 157*(104837) doi:https://doi.org/10.1016/j.phrs.2020.104837

Mahase, E. (2020). Covid-19: Low dose steroid cuts death in ventilated patients by one third, trial finds. *Bmj, 369*, m2422. doi:https://doi.org/10.1136/bmj.m2422

Maier, T., Güell, M., & Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. *FEBS Letters, 583*(24), 3966-3973. doi:https://doi.org/10.1016/j.febslet.2009.10.036

Maile, L. A., Allen, L. B., Hanzaker, C. F., Gollahon, K. A., Dunbar, P., & Clemmons, D. R. (2010). Glucose regulation of thrombospondin and its role in the modulation of smooth muscle cell proliferation. *Experimental Diabetes Research, 617052* doi:https://doi.org/10.1155/2010/617052

Maile, L. A., Allen, L. B., Veluvolu, U., Capps, B. E., Busby, W. H., Rowland, M., & Clemmons, D. R. (2009). Identification of compounds that inhibit IGF-I signaling in hyperglycemia. *Experimental Diabetes Research, 267107* doi:https://doi.org/10.1155/2009/267107

Maile, L. A., Capps, B. E., Miller, E. C., Aday, A. W., & Clemmons, D. R. (2008). Integrin-associated protein association with SRC homology 2 domain containing tyrosine phosphatase substrate 1 regulates igf-I signaling in vivo. *Diabetes, 57*(10), 2637-2643. doi:https://doi.org/10.2337/db08-0326

Maile, L. A., Gollahon, K., Wai, C., Byfield, G., Hartnett, M. E., & Clemmons, D. (2012). Disruption of the association of integrin-associated protein (IAP) with tyrosine phosphatase non-receptor type substrate-1 (SHPS)-1 inhibits pathophysiological changes in retinal endothelial function in a rat model of diabetes. *Diabetologia, 55*(3), 835-844. doi:https://doi.org/10.1007/s00125-011-2416-x

Maimaitiyiming, H., Norman, H., Zhou, Q., & Wang, S. (2015). CD47 deficiency protects mice from diet-induced obesity and improves whole body glucose tolerance and insulin sensitivity. *Scientific Reports, 5*(8846) doi:https://doi.org/10.1038/srep08846

Mansoori, B., Mohammadi, A., Davudian, S., Shirjang, S., & Baradaran, B. (2017). The different mechanisms of cancer drug resistance: A brief review. *Advanced Pharmaceutical Bulletin, 7*(3), 339-348. doi:https://doi.org/10.15171/apb.2017.041

Marais, R., Wynne, J., & Treisman, R. (1993). The SRF accessory protein elk-1 contains a growth factor-regulated transcriptional activation domain. *Cell, 73*(2), 381-393. doi:https://doi.org/10.1016/0092-8674(93)90237-K

Marchandot, B., Sattler, L., Jesel, L., Matsushita, K., Schini-Kerth, V., Grunebaum, L., & Morel, O. (2020). COVID-19 related coagulopathy: A distinct entity? *Journal of Clinical Medicine, 9*(6), 1651. doi:https://doi.org/10.3390/jcm9061651

Marengo-Rowe, A. J. (2007). The thalassemias and related disorders. *Proceedings (Baylor University. Medical Center), 20*(1), 27-31. doi:https://doi.org/10.1080/08998280.2007.11928230

Marks, D. I., & Rowntree, C. (2017). Management of adults with T-cell lymphoblastic leukemia. *Blood, 129*(9), 1134-1142. doi:https://doi.org/10.1182/blood-2016-07-692608

Marks-Anglin, A., & Chen, Y. (2020). A historical review of publication bias. *Research Synthesis Methods, 11*(6), 725-742. doi:https://doi.org/10.1002/jrsm.1452

Marshall, C. J. (1995). Specificity of receptor tyrosine kinase signaling: Transient versus sustained extracellular signal-regulated kinase activation. *Cell, 80*, 179-185. doi:https://doi.org/10.1016/0092-8674(95)90401-8

Martell, H. J., Masterson, S. G., McGreig, J. E., Michaelis, M., & Wass, M. N. (2019). Is the bombali virus pathogenic in humans? . *Bioinformatics, 35*(19), 3553-3558. doi:https://doi.org/10.1093/bioinformatics/btz267

Marusyk, A., Janiszewska, M., & Polyak, K. (2020). Intratumor heterogeneity: The rosetta stone of therapy resistance. *Cancer Cell, 37*(4), 471-484. doi:https://doi.org/10.1016/j.ccell.2020.03.007

Mateos, F., Brock, J. H., & Pérez-Arellano, J. L. (1998). Iron metabolism in the lower respiratory tract. *Thorax, 53*(7), 94-600. doi:https://doi.org/10.1136/thx.53.7.594

Matsubayashi, Y., Fukuda, M., & Nishida, E. (2001). Evidence for existence of a nuclear pore complex-mediated, cytosol-independent pathway of nuclear translocation of ERK MAP kinase in permeabilized cells. *Journal of Biological Chemistry, 276*(45), 41755-41760. doi:https://doi.org/10.1074/jbc.M106012200

Matsuyama, S., Nao, N., Shirato, K., Kawase, M., Saito, S., Takayama, I., . . . Takeda, M. (2020). Enhanced isolation of SARS-CoV-2 by TMPRSS2-expressing cells. *Proceedings of the National Academy of Sciences of the United States of America, 117*(13), 7001-7003. doi:https://doi.org/10.1073/pnas.2002589117

Mauney, C. H., & Hollis, T. (2018). SAMHD1: Recurring roles in cell cycle, viral restriction, cancer, and innate immunity. *Autoimmunity, 51*(3), 96-110. doi:https://doi.org/10.1080/08916934.2018.1454912

Mauney, C. H., Rogers, L. C., Harris, R. S., Daniel, L. W., Devarie-Baez, N. O., Wu, H., . . . Hollis, T. (2017). The SAMHD1 dNTP triphosphohydrolase is controlled by a redox switch. *Antioxidants & Redox Signaling, 27*(16), 1317-1331. doi:https://doi.org/10.1089/ars.2016.6888

McClain, D. A., Sharma, N. K., Jain, S., Harrison, A., Salaye, L. N., Comeau, M. E., . . . Das, S. K. (2018). Adipose tissue transferrin and insulin resistance. *The Journal of Clinical*

Endocrinology and Metabolism, 103(11), 4197-4208. doi:https://doi.org/10.1210/jc.2018-00770

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R., Thormann, A., . . . Cunningham, F. (2016). The ensembl variant effect predictor. *Genome Biology, 17*(122) doi:https://doi.org/10.1186/s13059-016-0974-4

Medeiros, B. C. (2009). Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *The New England Journal of Medicine, 360*(17), 1787-1788. doi:https://doi.org/10.1056/NEJMc090454

Meijles, D. N., Sahoo, S., Al Ghouleh, I., Amaral, J. H., Bienes-Martinez, R., Knupp, H. E., . . . Pagano, P. J. (2017). The matricellular protein TSP1 promotes human and mouse endothelial cell senescence through CD47 and Nox1. *Science Signaling, 10*(501), eaaj1784. doi:https://doi.org/10.1126/scisignal.aaj1784

Merati, M., Buethe, D. J., Cooper, K. D., Honda, K. S., Wang, H., & Gerstenblith, M. R. (2015). Aggressive CD8(+) epidermotropic cutaneous T-cell lymphoma associated with homozygous mutation in SAMHD1. *JAAD Case Reports, 1*(4), 227-229. doi:https://doi.org/10.1016/j.jdcr.2015.05.003

Merkler, A. E., Parikh, N. S., Mir, S., Gupta, A., Kamel, H., Lin, E., . . . Navi, B. B. (2020). Risk of ischemic stroke in patients with covid-19 versus patients with influenza. *medRxiv,* doi:https://doi.org/10.1101/2020.05.18.20105494

Messner, C. B., Demichev, V., Wendisch, D., Michalick, L., White, M., Freiwald, A., . . . Ralser, M. (2020). Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection. *Cell Systems, 11*(1), 11-24. doi:https://doi.org/10.1016/j.cels.2020.05.012

Mezalek, Z. T., Khibri, H., Ammouri, W., Bouaouad, M., Haidour, S., Harmouche, H., . . . Adnaoui, M. (2020). COVID-19 associated coagulopathy and thrombotic complications. *Clinical and Applied Thrombosis/Hemostasis : Official Journal of the International Academy of Clinical and Applied Thrombosis/Hemostasis, 26*(1076029620948137) doi:https://doi.org/10.1177/1076029620948137

Mi, H., Muruganujan, A., Ebert, D., Huang, X., & Thomas, P. D. (2019). PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research, 47*(D1), D419-D426. doi:https://doi.org/10.1093/nar/gky1038

Michaelis, M., Kleinschmidt, M. C., Doerr, H. W., & Cinatl Jr, J. (2007). Minocycline inhibits west nile virus replication and apoptosis in human neuronal cells. *Journal of Antimicrobial Chemotherapy, 60*(5), 981-986. doi:https://doi.org/10.1093/jac/dkm307

Michaelis, M., Wass, M. N., & Cinatl Jr., J. (2019). Drug-adapted cancer cell lines as preclinical models of acquired resistance. *Cancer Drug Resistance, 2*, 447-456.

Michaelis, M., Wass, M. N., Reddin, I., Voges, Y., Rothweiler, F., Hehlgans, S., . . . Cinatl Jr, J. (2020). YM155-adapted cancer cell lines reveal drug-induced heterogeneity and enable the identification of biomarker candidates for the acquired resistance setting. *Cancers, 12*(5), 1080. doi:https://doi.org/10.3390/cancers12051080

Milanini-Mongiat, J., Pouysségur, J., & Pagès, G. (2002). Identification of two Sp1 phosphorylation sites for p42/p44 mitogen-activated protein kinases: Their implication in vascular endothelial growth factor gene transcription. *Journal of Biological Chemistry, 277*(23), 20631-20639. doi:https://doi.org/10.1074/jbc.M201753200

Miller, T. W., Isenberg, J. S., & Roberts, D. D. (2010). Thrombospondin-1 is an inhibitor of pharmacological activation of soluble guanylate cyclase. *British Journal of Pharmacology, 159*(7), 1542-1547. doi:https://doi.org/10.1111/j.1476-5381.2009.00631.x

Mirzayans, R., & Murray, D. (2020). Intratumor heterogeneity and therapy resistance: Contributions of dormancy, apoptosis reversal (anastasis) and cell fusion to disease recurrence. *International Journal of Molecular Sciences, 21*(4), 1308. doi:https://doi.org/10.3390/ijms21041308

Mitrani, R. D., Dabas, N., & Goldberger, J. J. (2020). COVID-19 cardiac injury: Implications for long-term surveillance and outcomes in survivors. *Heart Rhythm, 17*(11), 1984-1990. doi:https://doi.org/10.1016/j.hrthm.2020.06.026

Mohseni, M., Uludag, H., & Brandwein, J. M. (2018). Advances in biology of acute lymphoblastic leukemia (ALL) and therapeutic implications. *American Journal of Blood Research, 8*(4), 29-56.

Monk, P. D., Marsden, R. J., Tear, V. J., Brookes, J., Batten, T. N., Mankowski, M., . . . Wilkinson, T. M. A. (2020). Safety and efficacy of inhaled nebulised interferon beta-1a (SNG001) for treatment of SARS-CoV-2 infection: A randomised, double-blind, placebo-controlled, phase 2 trial. *The Lancet Respiratory Medicine, 9*(2), 196-206. doi:https://doi.org/10.1016/S2213-2600(20)30511-7

Mosmann, T. (1983). Rapid colorimetric assay for cellular growth and survival: Application to proliferation and cytotoxicity assays. . *Journal of Immunological Methods, 65*(1-2), 55-63. doi:https://doi.org/10.1016/0022-1759(83)90303-4

Mueller, A. L., McNamara, M. S., & Sinclair, D. A. (2020). Why does COVID-19 disproportionately affect older people? *Aging, 12*(10), 9959-9981. doi:https://doi.org/10.18632/aging.103344

Mukherjee, G., & Jayaram, B. (2013). A rapid identification of hit molecules for target proteins via physico-chemical descriptors. *Physical Chemistry Chemical Physics : PCCP, 15*(23), 9107-9116. doi:https://doi.org/10.1039/c3cp44697b

Muller, L. M., Gorter, K. J., Hak, E., Goudzwaard, W. L., Schellevis, F. G., Hoepelman, A. I., & Rutten, G. E. (2005). Increased risk of common infections in patients with type 1 and type 2 diabetes mellitus. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America, 41*(3), 281-288. doi:https://doi.org/10.1086/431587

Murakami, Y., Saito, K., Ito, H., & Hashimoto, Y. (2019). Transferrin isoforms in cerebrospinal fluid and their relation to neurological diseases. *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences, 95*(5), 198-210. doi:https://doi.org/10.2183/pjab.95.015

Murgolo, N., Therien, A. G., Howell, B., Klein, D., Koeplinger, K., Lieberman, L. A., . . . Olsen, D. B. (2021). SARS-CoV-2 tropism, entry, replication, and propagation: Considerations for drug discovery and development. *PLoS Pathogens, 17*(2), e1009225. doi:https://doi.org/10.1371/journal.ppat.1009225

Nadeu, F., Martin-Garcia, D., Clot, G., Díaz-Navarro, A., Duran-Ferrer, M., Navarro, A., . . . Campo, E. (2020). Genomic and epigenomic insights into the origin, pathogenesis, and clinical behavior of mantle cell lymphoma subtypes. *Blood, 136*(12), 1419-1432. doi:https://doi.org/10.1182/blood.2020005289

Nakamura, T., Xi, G., Park, J. W., Hua, Y., Hoff, J. T., & Keep, R. F. (2005). Holo-transferrin and thrombin can interact to cause brain damage. *Stroke, 36*, 348-352. doi:https://doi.org/10.1161/01.STR.0000153044.60858.1b

National Cancer Institute. (2020). Cancer statistics. Retrieved from https://www.cancer.gov/about-cancer/understanding/statistics

National Cancer Institute. (2021). Genomic data commons. Retrieved from https://gdc.cancer.gov/

National Center for Biotechnology Information. (2021). Gene expression omnibus (GEO). Retrieved from https://www.ncbi.nlm.nih.gov/geo/

National Library of Medicine. (2021). PubMed. Retrieved from pubmed.ncbi.nlm.nih.gov

Nevitt, C., McKenzie, G., Christian, K., Austin, J., Hencke, S., Hoying, J., & LeBlanc, A. (2016). Physiological levels of thrombospondin-1 decrease NO-dependent vasodilation in coronary microvessels from aged rats. *American Journal of Physiology. Heart and Circulatory Physiology, 310*(11), H1842-H1850. doi:https://doi.org/10.1152/ajpheart.00086.2016

Ng, P. C., & Henikoff, S. (2001). Predicting deleterious amino acid substitutions. *Genome Research, 11*(5), 863-874. doi:https://doi.org/10.1101/gr.176601

Ng, S. W., Mitchell, A., Kennedy, J. A., Chen, W. C., McLeod, J., Ibrahimova, N., . . . Wang, J. C. (2016). A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature, 540*, 433-437. doi:https://doi.org/10.1038/nature20598

Nicolet, B. P., & Monika C. Wolkers, M. C. (2020). Limited but gene-class specific correlation of mRNA and protein expression in human CD8+ T cells. *bioRxiv, 2020.04.21.053884* doi:https://doi.org/10.1101/2020.04.21.053884

Nissen, S. B., Magidson, T., Gross, K., & Bergstrom, C. T. (2016). Publication bias and the canonization of false facts. *eLife, 5*, e21451. doi:https://doi.org/10.7554/eLife.21451

Norman-Burgdolf, H., Li, D., Sullivan, P., & Wang, S. (2020). CD47 differentially regulates white and brown fat function. *Biology Open, 9*(12) doi:https://doi.org/10.1242/bio.056747

Novelli, E. M., Little-Ihrig, L., Knupp, H. E., Rogers, N. M., Yao, M., Baust, J. J., . . . Gladwin, M. T. (2019). Vascular TSP1-CD47 signaling promotes sickle cell-associated arterial vasculopathy and pulmonary hypertension in mice. *American Journal of Physiology. Lung Cellular and Molecular Physiology, 316*(6), L1150-L1164. doi:https://doi.org/10.1152/ajplung.00302.2018

Nurgali, K., Jagoe, R. T., & Abalo, R. (2018). Editorial: Adverse effects of cancer chemotherapy: Anything new to improve tolerance and reduce sequelae? *Frontiers in Pharmacology, 9*(245) doi:https://doi.org/10.3389/fphar.2018.00245

Oellerich, T., Schneider, C., Thomas, D., Knecht, K. M., Buzovetsky, O., Kaderali, L., . . . Cinatl Jr, J. (2019). Selective inactivation of hypomethylating agents by SAMHD1 provides a rationale for therapeutic stratification in AML. *Nature Communications, 10*(1), 3475. doi:https://doi.org/10.1038/s41467-019-11413-4

Oh, J., Jang, S. J., Kim, J., Sohn, I., Lee, J., Cho, E. J., . . . Sung, C. O. (2020). Spontaneous mutations in the single TTN gene represent high tumor mutation burden. *Npj Genomic Medicine, 5*, 33. doi:https://doi.org/10.1038/s41525-019-0107-6

Oke, J. L., O'Sullivan, J. W., Perera, R., & Nicholson, B. D. (2018). The mapping of cancer incidence and mortality trends in the UK from 1980-2013 reveals a potential for overdiagnosis. *Scientific Reports, 8*(1), 14663. doi:https://doi.org/10.1038/s41598-018-32844-x

Olwal, C. O., Nganyewo, N. N., Tapela, K., Djomkam Zune, A. L., Owoicho, O., Bediako, Y., & Duodu, S. (2021). Parallels in sepsis and COVID-19 conditions: Implications for managing severe COVID-19. *Frontiers in Immunology, 12*(602848) doi:https://doi.org/10.3389/fimmu.2021.602848

Onafuye, H., Pieper, S., Mulac, D., Cinatl Jr, J., Wass, M. N., Langer, K., & Michaelis, M. (2019). Doxorubicin-loaded human serum albumin nanoparticles overcome transporter-mediated drug resistance in drug-adapted cancer cells. *Beilstein Journal of Nanotechnology, 10*, 1707-1715. doi:https://doi.org/10.3762/bjnano.10.166

Oreshkova, N., Molenaar, R. J., Vreman, S., Harders, F., Oude Munnink, B. B., Hakze-van der Honing, R. W., . . . Stegeman, A. (2020). SARS-CoV-2 infection in farmed minks, the

netherlands, april and may 2020. *Euro Surveillance, 25*(23) doi:https://doi.org/10.2807/1560-7917.ES.2020.25.23.2001005

Oronsky, B., Knox, S., Cabrales, P., Oronsky, A., & Reid, T. R. (2020). Desperate times, desperate measures: The case for RRx-001 in the treatment of COVID-19. *Seminars in Oncology, 47*(5), 305-308. doi:https://doi.org/10.1053/j.seminoncol.2020.07.002

Ovcharenko, A. V., & Zhirnov, O. P. (1994). Aprotinin aerosol treatment of influenza and paramyxovirus bronchopneumonia of mice. *Antiviral Research, 23*(2), 107-118. doi:https://doi.org/10.1016/0166-3542(94)90038-8

Pachetti, M., Marini, B., Benedetti, F., Giudici, F., Mauro, E., Storici, P., . . . Ippodrino, R. (2020). Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Translational Medicine, 18*, 179. doi:https://doi.org/10.1186/s12967-020-02344-6

Pai, C. C., & Kearsey, S. E. (2017). A critical balance: dNTPs and the maintenance of genome stability. *Genes, 8*(2), 57. doi:https://doi.org/10.3390/genes8020057

Pan, W., Zhang, J., Wang, M., Ye, J., Xu, Y., Shen, B., . . . Wan, J. (2020). Clinical features of COVID-19 in patients with essential hypertension and the impacts of renin-angiotensin-aldosterone system inhibitors on the prognosis of COVID-19 patients. *Hypertension (Dallas, Tex : 1979), 76*(3), 732-741. doi:https://doi.org/10.1161/HYPERTENSIONAHA.120.15289

Pantalone, K. M., Hobbs, T. M., Chagin, K. M., Kong, S. X., Wells, B. J., Kattan, M. W., . . . Burguera, B. (2017). Prevalence and recognition of obesity and its associated comorbidities: Cross-sectional analysis of electronic health record data from a large US integrated health system. *BMJ Open, 7*(11), e017583. doi:https://doi.org/10.1136/bmjopen-2017-017583

Papa, G., Mallery, D. L., Albecka, A., Welch, L. G., Cattin-Ortolá, J., Luptak, J., . . . James, L. C. (2021). Furin cleavage of SARS-CoV-2 spike promotes but is not essential for infection and cell-cell fusion. *PLoS Pathogens, 17*(1), e1009246. doi:https://doi.org/10.1371/journal.ppat.1009246

Pappalardo, M., Juliá, M., Howard, M., Rossman, J. S., Michaelis, M., & Wass, M. N. (2016). Conserved differences in protein sequence determine the human pathogenicity of ebolaviruses. *Scientific Reports, 6*(23743) doi:https://doi.org/10.1038/srep23743

Park, K., Ryoo, J., Jeong, H., Kim, M., Lee, S., Hwang, S. Y., . . . Ahn, K. (2021). Aicardi-goutières syndrome-associated gene SAMHD1 preserves genome integrity by preventing R-loop formation at transcription-replication conflict regions. *PLoS Genetics, 17*(4), e1009523. doi:https://doi.org/10.1371/journal.pgen.1009523

Patel, A., Abdulaal, A., Ariyanayagam, D., Killington, K., Denny, S. J., Mughal, N., . . . Charani, E. (2020). Investigating the association between ethnicity and health outcomes in SARS-CoV-2 in a london secondary care population. *PLoS One, 15*(10), e0240960. doi:https://doi.org/10.1371/journal.pone.0240960

Pavlovic, S., Kotur, N., Stankovic, B., Zukic, B., Gasic, V., & Dokmanovic, L. (2019). Pharmacogenomic and pharmacotranscriptomic profiling of childhood acute lymphoblastic leukemia: Paving the way to personalized treatment. *Genes, 10*(3), 191. doi:https://doi.org/10.3390/genes10030191

Pazos, F., Rausell, A., & Valencia, A. (2006). Phylogeny-independent detection of functional residues. *Bioinformatics (Oxford, England), 22*(12), 1440-1448. doi:https://doi.org/10.1093/bioinformatics/btl104

Pazos, F., & Sternberg, M. J. (2004). Automated prediction of protein function and detection of functional sites from structure. *Proceedings of the National Academy of Sciences of the United States of America, 101*(41), 14754-14759. doi:https://doi.org/10.1073/pnas.0404569101

Peckham, H., de Gruijter, N. M., Raine, C., Radziszewska, A., Ciurtin, C., Wedderburn, L. R., . . . Deakin, C. T. (2020). Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. *Nature Communications, 11*(1), 6317. doi:https://doi.org/10.1038/s41467-020-19741-6

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., . . . Vizcaíno, J. A. (2019). The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Research, 47*(D1), D442-D450. doi:https://doi.org/10.1093/nar/gky1106

Perillo, B., Di Donato, M., Pezone, A., Di Zazzo, E., Giovannelli, P., Galasso, G., . . . Migliaccio, A. (2020). ROS in cancer therapy: The bright side of the moon. *Experimental & Molecular Medicin, 52*(2), 192-203. doi:https://doi.org/10.1038/s12276-020-0384-2

Perlman, S., & Netland, J. (2009). Coronaviruses post-SARS: Update on replication and pathogenesis. *Nature Reviews Microbiology, 7*(6), 439-450. doi:https://doi.org/10.1038/nrmicro2147

Pfefferle, S., Schöpf, J., Kögl, M., Friedel, C. C., Müller, M. A., Carbajo-Lozoya, J., . . . von Brunn, A. (2011). The SARS-coronavirus-host interactome: Identification of cyclophilins as target for pan-coronavirus inhibitors . *PLoS Pathogens, 7*(10), e1002331. doi:https://doi.org/10.1371/journal.ppat.1002331

Pickett, B. E., Greer, D. S., Zhang, Y., Stewart, L., Zhou, L., Sun, G., . . . Scheuermann, R. H. (2012). Virus pathogen database and analysis resource (ViPR): A comprehensive bioinformatics database and analysis resource for the coronavirus research community. *Viruses, 4*(11), 3209-3226. doi:https://doi.org/10.3390/v4113209

Pickett, B. E., Sadat, E. L., Zhang, Y., Noronha, J. M., Squires, R. B., Hunt, V., . . . Scheuermann, R. H. (2012). ViPR: An open bioinformatics database and analysis resource for virology research. *Nucleic Acids Research, 40 (Database issue)*, D593-D598. doi:https://doi.org/10.1093/nar/gkr859

Pilakka-Kanthikeel, S., Raymond, A., Atluri, V. S., Sagar, V., Saxena, S. K., Diaz, P., . . . Nair, M. (2015). Sterile alpha motif and histidine/aspartic acid domain-containing protein 1 (SAMHD1)-facilitated HIV restriction in astrocytes is regulated by miRNA-181a. *Journal of Neuroinflammation, 12*, 66. doi:https://doi.org/10.1186/s12974-015-0285-9

Pokhriyal, R., Hariprasad, R., Kumar, L., & Hariprasad, G. (2019). Chemotherapy resistance in advanced ovarian cancer patients. *Biomarkers in Cancer, 11*(1179299X19860815) doi:https://doi.org/10.1177/1179299X19860815

Polack, F. P., Thomas, S. J., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S., . . . C4591001 Clinical Trial Group. (2020). Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine. *The New England Journal of Medicine, 383*(27), 2603-2615. doi:https://doi.org/10.1056/NEJMoa2034577

Polycarpou, A., Howard, M., Farrar, C. A., Greenlaw, R., Fanelli, G., Wallis, R., . . . Sacks, S. (2020). Rationale for targeting complement in COVID-19. *EMBO Molecular Medicine, 12*, e202012642. doi:https://doi.org/10.15252/emmm.202012642

Pozdeyev, N., Yoo, M., Mackie, R., Schweppe, R. E., Tan, A. C., & Haugen, B. R. (2016). Integrating heterogeneous drug sensitivity data from cancer pharmacogenomic studies. *Oncotarget, 7*(32), 51619-51625. doi:https://doi.org/10.18632/oncotarget.10010

Prasad, V. V., & Gopalan, R. O. (2015). Continued use of MDA-MB-435, a melanoma cell line, as a model for human breast cancer, even in year, 2014. *NPJ Breast Cancer, 1*(15002) doi:https://doi.org/10.1038/npjbcancer.2015.2

Previtali, E., Bucciarelli, P., Passamonti, S. M., & Martinelli, I. (2011). Risk factors for venous and arterial thrombosis. *Blood Transfusion = Trasfusione Del Sangue, 9*(2), 120-138. doi:https://doi.org/10.2450/2010.0066-10

Pugh, T. J., Morozova, O., Attiyeh, E. F., Asgharzadeh, S., Wei, J. S., Auclair, D., . . . Maris, J. M. (2013). The genetic landscape of high-risk neuroblastoma. *Nature Genetics, 45*(3), 279-284. doi:https://doi.org/10.1038/ng.2529

Pum, A., Ennemoser, M., Adage, T., & Kungl, A. J. (2021). Cytokines and chemokines in SARS-CoV-2 infections-therapeutic strategies targeting cytokine storm. *Biomolecules, 11*(1), 91. doi:https://doi.org/10.3390/biom11010091

Purohit, V., Simeone, D. M., & Lyssiotis, C. A. (2019). Metabolic regulation of redox balance in cancer. *Cancers, 11*(7), 955. doi:https://doi.org/10.3390/cancers11070955

Qin, C., Zhou, L., Hu, Z., Zhang, S., Yang, S., Tao, Y., . . . Tian, D. S. (2020). Dysregulation of immune response in patients with coronavirus 2019 (COVID-19) in wuhan, china. *Clinical Infectious Diseases : An Official Publication of the Infectious Diseases Society of America, 71*(15), 762-768. doi:https://doi.org/10.1093/cid/ciaa248

Qin, Y., Conley, A. P., Grimm, E. A., & Roszik, J. (2017). A tool for discovering drug sensitivity and gene expression associations in cancer cells. *PLoS One, 12*(4), e0176763. doi:https://doi.org/10.1371/journal.pone.0176763

Qiu, Y., Zhao, Y. B., Wang, Q., Li, J. Y., Zhou, Z. J., Liao, C. H., & Ge, X. Y. (2020). Predicting the angiotensin converting enzyme 2 (ACE2) utilizing capability as the receptor of SARS-CoV-2. *Microbes and Infection, 22*(4-5), 221-225. doi:https://doi.org/10.1016/j.micinf.2020.03.003

Qu, R., Ling, Y., Zhang, Y. H., Wei, L. Y., Chen, X., Li, X. M., . . . Wang, Q. (2020). Platelet-to-lymphocyte ratio is associated with prognosis in patients with coronavirus disease-19. *Journal of Medical Virology, 92*(9), 1533-1541. doi:https://doi.org/10.1002/jmv.25767

Quevedo, R., Smirnov, P., Tkachuk, D., Ho, C., El-Hachem, N., Safikhani, Z., . . . Haibe-Kains, B. (2020). Assessment of genetic drift in large pharmacogenomic studies. *Cell Systems, 11*(4), 393-401. doi:https://doi.org/10.1016/j.cels.2020.08.012

Quintavalle, C., Di Costanzo, S., Zanca, C., Tasset, I., Fraldi, A., Incoronato, M., . . . Condorelli, G. (2014). Phosphorylation-regulated degradation of the tumor-suppressor form of PED by chaperone-mediated autophagy in lung cancer cells. *Journal of Cellular Physiology, 229*(10), 1359-1368. doi:https://doi.org/10.1002/jcp.24569

Qureshi, A., Thakur, N., Tandon, H., & Kumar, M. (2013). AVPdb: A database of experimentally validated antiviral peptides targeting medically important viruses. *Nucleic Acids Research, 42*(D1), D1147-D1153. doi:https://doi.org/10.1093/nar/gkt1191

Radhakrishnan, A., & Green, R. (2016). Connections underlying translation and mRNA stability. *Journal of Molecular Biology, 428*(18), 3558-3564. doi:https://doi.org/10.1016/j.jmb.2016.05.025

Rae, J. M., Creighton, C. J., Meck, J. M., Haddad, B. R., & Johnson, M. D. (2007). MDA-MB-435 cells are derived from M14 melanoma cells—a loss for breast cancer, but a boon for melanoma research. *Breast Cancer Research and Treatment, 104*, 13-19. doi:https://doi.org/10.1007/s10549-006-9392-8

Ramón Y Cajal, S., S., M., Capdevila, C., Aasen, T., De Mattos-Arruda, L., Diaz-Cano, S. J., . . . Castellví, J. (2020). Clinical implications of intratumor heterogeneity: Challenges and opportunities. *Journal of Molecular Medicine (Berlin, Germany), 98*(2), 161-177. doi:https://doi.org/10.1007/s00109-020-01874-2

Ramos, J. W., Kojima, T. K., Hughes, P. E., Fenczik, C. A., & Ginsberg, M. H. (1998). The death effector domain of PEA-15 is involved in its regulation of integrin activation. *Journal of Biological Chemistry, 273*, 33897-33900. doi:https://doi.org/10.1074/jbc.273.51.33897

Rana, J. S., Khan, S. S., Lloyd-Jones, D. M., & Sidney, S. (2020). Changes in mortality in top 10 causes of death from 2011 to 2018. *Journal of General Internal Medicine, 1-2* doi:https://doi.org/10.1007/s11606-020-06070-z

Rausell, A., Juan, D., Pazos, F., & Valencia, A. (2010). Protein interactions and ligand binding: From protein subfamilies to functional specificity. *Proceedings of the National Academy of Sciences of the United States of America, 107*(5), 1995-2000. doi:https://doi.org/10.1073/pnas.0908044107

Rebold, N., Holger, D., Alosaimy, S., Morrisette, T., & Rybak, M. (2021). COVID-19: Before the fall, an evidence-based narrative review of treatment options. *Infectious Diseases and Therapy, 10*(1), 93-113. doi:https://doi.org/10.1007/s40121-021-00399-6

RECOVERY Collaborative Group, Horby, P., Lim, W. S., Emberson, J. R., Mafham, M., Bell, J. L., . . . Landray, M. J. (2021). Dexamethasone in hospitalized patients with covid-19. *The New England Journal of Medicine, 384*(8), 693-704. doi:https://doi.org/10.1056/NEJMoa2021436

Redondo, N., Zaldívar-López, S., Garrido, J. J., & Montoya, M. (2021). SARS-CoV-2 accessory proteins in viral pathogenesis: Knowns and unknowns. *Frontiers in Immunology, 12*, 708264. doi:https://doi.org/10.3389/fimmu.2021.708264

Rees, M. G., Seashore-Ludlow, B., Cheah, J. H., Adams, D. J., Price, E. V., Gill, S., . . . Schreiber, S. L. (2016). Correlating chemical sensitivity and basal gene expression reveals mechanism of action. *Nature Chemical Biology, 12*(2), 109-116. doi:https://doi.org/10.1038/nchembio.1986

Ren, Y., Shu, T., Wu, D., Mu, J., Wang, C., Huang, M., . . . Zhou, X. (2020). The ORF3a protein of SARS-CoV-2 induces apoptosis in cells. *Cellular and Molecular Immunology, 17*, 881-883. doi:https://doi.org/10.1038/s41423-020-0485-9

Rentoft, M., Lindell, K., Tran, P., Chabes, A. L., Buckland, R. J., Watt, D. L., . . . Chabes, A. (2016). Heterozygous colon cancer-associated mutations of SAMHD1 have functional significance. *Proceedings of the National Academy of Sciences of the United States of America, 113*(17), 4723-4728. doi:https://doi.org/10.1073/pnas.1519128113

Reszka, A. A., Seger, R., Diltz, C. D., Krebs, E. G., & Fischer, E. H. (1995). Association of mitogen-activated protein kinase with the microtubule cytoskeleton. *Proceedings of the National Academy of Sciences of the United States of America, 92*(19), 8881-8885. doi:https://doi.org/10.1073/pnas.92.19.8881

Rice, G. I., Bond, J., Asipu, A., Brunette, R. L., Manfield, I. W., Carr, I. M., . . . Crow, Y. J. (2009). Mutations involved in aicardi-goutières syndrome implicate SAMHD1 as regulator of the innate immune response. *Nature Genetics, 41*(7), 829-832. doi:https://doi.org/10.1038/ng.373

Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: The european molecular biology open software suite. *Trend in Genetics : TIG, 16*(6), 276-277. doi:https://doi.org/10.1016/s0168-9525(00)02024-2

Riess, M., Fuchs, N. V., Idica, A., Hamdorf, M., Flory, E., Pedersen, I. M., & König, R. (2017). Interferons induce expression of SAMHD1 in monocytes through down-regulation of miR-181a and miR-30a. *The Journal of Biological Chemistry, 292*(1), 264-277. doi:https://doi.org/10.1074/jbc.M116.752584

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research, 43*(7), e47. doi:https://doi.org/10.1093/nar/gkv007

Rivett, L., Sridhar, S., Sparkes, D., Routledge, M., Jones, N. K., Forrest, S., . . . Weekes, M. P. (2020). Screening of healthcare workers for SARS-CoV-2 highlights the role of asymptomatic carriage in COVID-19 transmission. *eLife, 9*, e58728. doi:https://doi.org/10.7554/eLife.58728

Rodriguez Jr, C. O., Stellrecht, C. M., & Gandhi, V. (2003). Mechanisms for T-cell selective cytotoxicity of arabinosylguanine. *Blood, 102*(5), 1842-1848. doi:https://doi.org/10.1182/blood-2003-01-0317

Rogers, N. M., Ghimire, K., Calzada, M. J., & Isenberg, J. S. (2017). Matricellular protein thrombospondin-1 in pulmonary hypertension: Multiple pathways to disease. *Cardiovascular Research, 113*(8), 858-868. doi:https://doi.org/10.1093/cvr/cvx094

Rogers, N. M., Roberts, D. D., & Isenberg, J. S. (2013). Age-associated induction of cell membrane CD47 limits basal and temperature-induced changes in cutaneous blood flow. *Annals of Surgery, 258*(1), 184-191. doi:https://doi.org/10.1097/SLA.0b013e31827e52e1

Rogers, N. M., Sharifi-Sanjani, M., Yao, M., Ghimire, K., Bienes-Martinez, R., Mutchler, S. M., . . . Isenberg, J. S. (2017). TSP1-CD47 signaling is upregulated in clinical pulmonary hypertension and contributes to pulmonary arterial vasculopathy and dysfunction. *Cardiovascular Research, 113*(1), 15-29. doi:https://doi.org/10.1093/cvr/cvw218

Rogers, N. M., Yao, M., Sembrat, J., George, M. P., Knupp, H., Ross, M., . . . Isenberg, J. S. (2013). Cellular, pharmacological, and biophysical evaluation of explanted lungs from a patient with sickle cell disease and severe pulmonary arterial hypertension. *Pulmonary Circulation, 3*(4), 936-951. doi:https://doi.org/10.1086/674754

Roider, T., Wang, X., Hüttl, K., Müller-Tidow, C., Klapper, W., Rosenwald, A., . . . Dietrich, S. (2021). The impact of SAMHD1 expression and mutation status in mantle cell lymphoma: An analysis of the MCL younger and elderly trial. *International Journal of Cancer, 148*(1), 150-160. doi:https://doi.org/10.1002/ijc.33202

Romano, M., Ruggiero, A., Squeglia, F., Maga, G., & Berisio, R. (2020). A structural view of SARS-CoV-2 RNA replication machinery: RNA synthesis, proofreading and final capping. cells. *Cells, 9*(5), 1267. doi:https://doi.org/10.3390/cells9051267

Romero Starke, K., Petereit-Haack, G., Schubert, M., Kämpf, D., Schliebner, A., Hegewald, J., & Seidler, A. (2020). The age-related risk of severe outcomes due to COVID-19 infection: A rapid review, meta-analysis, and meta-regression. *International Journal of Environmental Research and Public Health, 17*(16), 5974. doi:https://doi.org/10.3390/ijerph17165974

Rossi, G. A., Sacco, O., Mancino, E., Cristiani, L., & Midulla, F. (2020). Differences and similarities between SARS-CoV and SARS-CoV-2: Spike receptor-binding domain recognition and host cell infection with support of cellular serine proteases. *Infection, 48*(5), 665-669. doi:https://doi.org/10.1007/s15010-020-01486-5

Rothenburger, T., McLaughlin, K., Herold, T., Schneider, C., Oellerich, T., Rothweiler, F., . . . Cinatl Jr., J. (2020). SAMHD1 is a key regulator of the lineage-specific response of acute lymphoblastic leukaemias to nelarabine. *Communications Biology, 3*(1), 324. doi:https://doi.org/10.1038/s42003-020-1052-8

Ryan, C., Minc, A., Caceres, J., Balsalobre, A., Dixit, A., Ng, B. K., . . . Fung, C. (2020). Predicting severe outcomes in covid-19 related illness using only patient demographics, comorbidities and symptoms. *The American Journal of Emergency Medicine, S0735-6757*(20), 30809-3. doi:https://doi.org/10.1016/j.ajem.2020.09.017

Sabino, E. C., Buss, L. F., Carvalho, M., Prete Jr, C. A., Crispim, M., Fraiji, N. A., . . . Faria, N. R. (2021). Resurgence of COVID-19 in manaus, brazil, despite high seroprevalence. *Lancet (London, England), 397*(10273), 452-455. doi:https://doi.org/10.1016/S0140-6736(21)00183-5

Saçar Demirci, M. D., & Adan, A. (2020). Computational analysis of microRNA-mediated interactions in SARS-CoV-2 infection. *PeerJ, 8*(e9369) doi:https://doi.org/10.7717/peerj.9369

Salzberger, B., Buder, F., Lampl, B., Ehrenstein, B., Hitzenbichler, F., Holzmann, T., . . . Hanses, F. (2020). Epidemiology of SARS-CoV-2. *Infection, 1–7* doi:https://doi.org/10.1007/s15010-020-01531-3

Sama, I. E., Ravera, A., Santema, B. T., van Goor, H., Ter Maaten, J. M., Cleland, J., . . . Voors, A. A. (2020). Circulating plasma concentrations of angiotensin-converting enzyme 2 in men and women with heart failure and effects of renin-angiotensin-aldosterone inhibitors. *European Heart Journal, 41*(19), 1810-1817. doi:https://doi.org/10.1093/eurheartj/ehaa373

Sanghvi, S. K., Schwarzman, L. S., & Nazir, N. T. (2021). Cardiac MRI and myocardial injury in COVID-19: Diagnosis, risk stratification and prognosis. *Diagnostics (Basel, Switzerland), 11*(1), 130. doi:https://doi.org/10.3390/diagnostics11010130

Sarkar, M., & Saha, S. (2020). Structural insight into the role of novel SARS-CoV-2 E protein: A potential target for vaccine development and other therapeutic strategies. *PLoS One, 15*(8), e0237300. doi:https://doi.org/10.1371/journal.pone.0237300

Satarker, S., & Nampoothiri, M. (2020). Structural proteins in severe acute respiratory syndrome coronavirus-2. *Archives of Medical Research, 51*(6), 482-491. doi:https://doi.org/10.1016/j.arcmed.2020.05.012

Savoy, R. M., & Ghosh, P. M. (2013). The dual role of filamin a in cancer: Can't live with (too much of) it, can't live without it. *Endocrine-Related Cancer, 20*(6), R341-R356. doi:https://doi.org/10.1530/ERC-13-0364

Schmidt, N., Lareau, C. A., Keshishian, H., Ganskih, S., Schneider, C., Hennig, T., . . . Munschauer, M. (2020). The SARS-CoV-2 RNA-protein interactome in infected human cells. *Nature Microbiology, 10.1038/s41564-020-00846-z* doi:https://doi.org/10.1038/s41564-020-00846-z

Schneider, C., Oellerich, T., Baldauf, H. M., Schwarz, S. M., Thomas, D., Flick, R., . . . Cinatl Jr, J. (2017). SAMHD1 is a biomarker for cytarabine response and a therapeutic target in acute myeloid leukemia. *Nature Medicine, 23*(2), 250-255. doi:https://doi.org/10.1038/nm.4255

Scutelnic, A., & Heldner, M. R. (2020). Vascular events, vascular disease and vascular risk factors-strongly intertwined with COVID-19. *Current Treatment Options in Neurology, 22*(11), 40. doi:https://doi.org/10.1007/s11940-020-00648-y

Seashore-Ludlow, B., Rees, M. G., Cheah, J. H., Cokol, M., Price, E. V., Coletti, M. E., . . . Schreiber, S. L. (2015). Harnessing connectivity in a large-scale small-molecule sensitivity dataset. *Cancer Discovery, 5*(11), 1210-1223. doi:https://doi.org/10.1158/2159-8290.CD-15-0235

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., . . . Hassabis, D. (2020). Improved protein structure prediction using potentials from deep learning. *Nature, 577*(7792), 706-710. doi:https://doi.org/10.1038/s41586-019-1923-7

Shah, H., Khan, M., Dhurandhar, N. V., & Hegde, V. (2021). The triumvirate: Why hypertension, obesity, and diabetes are risk factors for adverse effects in patients with COVID-19. *Acta Diabetologica, 1-13* doi:https://doi.org/10.1007/s00592-020-01636-z

Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., . . . Li, F. (2020). Structural basis of receptor recognition by SARS-CoV-2. *Nature, 581*, 221-224. doi:https://doi.org/10.1038/s41586-020-2179-y

Shang, Z., Qian, L., Liu, S., Niu, X., Qiao, Z., Sun, Y., . . . Xiao, H. (2018). Graphene oxide-facilitated comprehensive analysis of cellular nucleic acid binding proteins for lung cancer.

ACS Applied Materials and Interfaces, 10(21), 17756-17770. doi:https://doi.org/10.1021/acsami.8b05428

Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., . . . Guo, T. (2020). Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell, 182*(1), 59-72. doi:https://doi.org/10.1016/j.cell.2020.05.032

Shen, L. W., Mao, H. J., Wu, Y. L., Tanaka, Y., & Zhang, W. (2017). TMPRSS2: A potential target for treatment of influenza virus and coronavirus infections. *Biochimie, 142*, 1-10. doi:https://doi.org/10.1016/j.biochi.2017.07.016

Shewach, D. S., & Mitchell, B. S. (1989). Differential metabolism of 9-beta-D-arabinofuranosylguanine in human leukemic cells. *Cancer Research, 49*(23), 6498-6502.

Shi, Q., Sutariya, V., Bishayee, A., & Bhatia, D. (2014). Sequential activation of elk-1/egr-1/GADD45α by arsenic. *Oncotarget, 5*(11), 3862-3870. doi:https://doi.org/10.18632/oncotarget.1995

Shi, Y., Lv, G., Chu, Z., Piao, L., Liu, X., Wang, T., . . . Zhang, P. (2014). Identification of natural splice variants of SAMHD1 in virus-infected HCC. *Oncology Reports, 31*(2), 687-692. doi:https://doi.org/10.3892/or.2013.2895

Shin, M., Lee, K., Yang, E. G., Jeon, H., & Song, H. K. (2015). PEA-15 facilitates EGFR dephosphorylation via ERK sequestration at increased ER–PM contacts in TNBC cells. *FEBS Letters, 589*(9), 1033-1039. doi:https://doi.org/10.1016/j.febslet.2015.03.009

Shu, T., Ning, W., Wu, D., Xu, J., Han, Q., Huang, M., . . . Zhou, X. (2020). Plasma proteomics identify biomarkers and pathogenesis of COVID-19. *Immunity, 53*(5), 1108-1122. doi:https://doi.org/10.1016/j.immuni.2020.10.008

Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data – from vision to reality. *Euro Surveillance, 22*(13) doi:https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494

Siddik, Z. H. (2003). Cisplatin: Mode of cytotoxic action and molecular basis of resistance. *Oncogene, 22*, 7265-7279. doi:https://doi.org/10.1038/sj.onc.1206933

Siegel, R. L., Miller, K. D., Fuchs, H. E., & Jemal, A. (2021). Cancer Statistics, 2021. *CA: a cancer journal for clinicians*, *71*(1), 7–33. doi:https://doi.org/10.3322/caac.21654

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., . . . Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular Systems Biology, 7*(539) doi:https://doi.org/10.1038/msb.2011.75

Sim, N. L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., & Ng, P. C. (2012). SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research, 40*(Web Server issue), W452-W457. doi:https://doi.org/10.1093/nar/gks539

Simmons, G., Zmora, P., Gierer, S., Heurich, A., & Pöhlmann, S. (2013). Proteolytic activation of the SARS-coronavirus spike protein: Cutting enzymes at the cutting edge of antiviral research. *Antiviral Research, 100*(3), 605-614. doi:https://doi.org/10.1016/j.antiviral.2013.09.028

Singh, K. K., Chaubey, G., Chen, J. Y., & Suravajhala, P. (2020). Decoding SARS-CoV-2 hijacking of host mitochondria in COVID-19 pathogenesis. *American Journal of Physiology. Cell Physiology, 319*(2), C258-C267. doi:https://doi.org/10.1152/ajpcell.00224.2020

Sive, J. I., Buck, G., Fielding, A., Lazarus, H. M., Litzow, M. R., Luger, S., . . . Goldstone, A. H. (2012). Outcomes in older adults with acute lymphoblastic leukaemia (ALL): Results from the international MRC UKALL XII/ECOG2993 trial. *British Journal of Haematology, 157*(4), 463-471. doi:https://doi.org/10.1111/j.1365-2141.2012.09095.x

Smith, J. C., & Sheltzer, J. M. (2018). Systematic identification of mutations and copy number alterations associated with cancer patient prognosis. *eLife, 7*, e39217. doi:https://doi.org/10.7554/eLife.39217

Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., . . . Klenk, D. C. (1985). Measurement of protein using bicinchoninic acid. *Analytical Biochemistry, 150*(1), 76-85. doi:https://doi.org/10.1016/0003-2697(85)90442-7

Smittenaar, C. R., Petersen, K. A., Stewart, K., & Moitt, N. (2016). Cancer incidence and mortality projections in the UK until 2035. *British Journal of Cancer, 115*(9), 1147-1155. doi:https://doi.org/10.1038/bjc.2016.304

Sobczak, A., & Stewart, A. J. (2019). Coagulatory defects in type-1 and type-2 diabetes. *International Journal of Molecular Sciences, 20*(24), 6345. doi:https://doi.org/10.3390/ijms20246345

Solun, B., & Shoenfeld, Y. (2020). Inhibition of metalloproteinases in therapy for severe lung injury due to COVID-19. *Medicine in Drug Discovery, 7*, 100052. doi:https://doi.org/10.1016/j.medidd.2020.100052

Song, W., Gui, M., Wang, X., & Xiang, Y. (2018). Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2. *PLoS Pathogens, 14*(8), e1007236. doi:https://doi.org/10.1371/journal.ppat.1007236

Soto-Pantoja, D. R., Stein, E. V., Rogers, N. M., Sharifi-Sanjani, M., Isenberg, J. S., & Roberts, D. D. (2013). Therapeutic opportunities for targeting the ubiquitous cell surface receptor CD47. *Expert Opinion on Therapeutic Targets, 17*(1), 89-103. doi:https://doi.org/10.1517/14728222.2013.733699

Sterling, T., & Irwin, J. J. (2015). ZINC 15 – ligand discovery for everyone. *Journal of Chemical Information and Modeling, 55*(11), 2324-2337. doi:https://doi.org/10.1021/acs.jcim.5b00559

Straus, M. R., Kinder, J. T., Segall, M., Dutch, R. E., & Whittaker, G. R. (2020). SPINT2 inhibits proteases involved in activation of both influenza viruses and metapneumoviruses. *Virology, 543*, 43-53. doi:https://doi.org/10.1016/j.virol.2020.01.004

Strnad, P., McElvaney, N. G., & Lomas, D. A. (2020). Alpha(1)-antitrypsin deficiency. *New England Journal of Medicine, 382*(15), 1443-1455. doi:https://doi.org/10.1056/NEJMra1910234

Stukalov, A., Girault, V., Grass, V., Bergant, V., Karayel, O., Urban, C., . . . Pichlmair, A. (2020). Multi-level proteomics reveals host-perturbation strategies of SARS-CoV-2 and SARS-CoV. *bioRxiv, 2020.06.17.156455* doi:https://doi.org/10.1101/2020.06.17.156455

Subramanian, A., Narayan, R., Corsello, S. M., Peck, D. D., Natoli, T. E., Lu, X., . . . Golub, T. R. (2017). A next generation connectivity map: L1000 platform and the first 1,000,000 profiles. *Cell, 171*(6), 1437-1452. doi:https://doi.org/10.1016/j.cell.2017.10.049

Sulzmaier, F., Opoku-Ansah, J., & Ramos, J. W. (2012). Phosphorylation is the switch that turns PEA-15 from tumor suppressor to tumor promoter. *Small GTPases, 3*(3), 173-177. doi:https://doi.org/10.4161/sgtp.20021

Sulzmaier, F. J., Valmiki, M. K., Nelson, D. A., Caliva, M. J., Geerts, D., Matter, M. L., . . . Ramos, J. W. (2012). PEA-15 potentiates H-ras-mediated epithelial cell transformation through phospholipase D. *Oncogene, 31*(30), 3547-3560. doi:https://doi.org/10.1038/onc.2011.514

Süt, B. B. (2020). Molecular profiling of immune cell-enriched severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) interacting protein USP13. *Life Sciences, 258*, 118170. doi:https://doi.org/10.1016/j.lfs.2020.118170

Sze, S., Pan, D., Nevill, C. R., Gray, L. J., Martin, C. A., Nazareth, J., . . . Pareek, M. (2020). Ethnicity and clinical outcomes in COVID-19: A systematic review and meta-analysis. *The Lancet EClinicalMedicine, 29*(100630) doi:https://doi.org/10.1016/j.eclinm.2020.100630

Takahashi, T., Ellingson, M. K., Wong, P., Israelow, B., Lucas, C., Klein, J., . . . Iwasaki, A. (2020). Sex differences in immune responses that underlie COVID-19 disease outcomes. *Nature, 588*, 315-320. doi:https://doi.org/10.1038/s41586-020-2700-3

Tal, M. C., Torrez Dulgeroff, L. B., Myers, L., Cham, L. B., Mayer-Barber, K. D., Bohrer, A. C., . . . Hasenkrug, K. J. (2020). Upregulation of CD47 is a host checkpoint response to pathogen recognition. *mBio, 11*(3), e01293–20. doi:https://doi.org/10.1128/mBio.01293-20

Tang, C., Ji, X., Wu, L., & Xiong, Y. (2015). Impaired dNTPase activity of SAMHD1 by phosphomimetic mutation of thr-592. *The Journal of Biological Chemistry, 290*(44), 26352-26359. doi:https://doi.org/10.1074/jbc.M115.677435

Tang, N., Li, D., Wang, X., & Sun, Z. (2020). Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. *Journal of Thrombosis and Haemostasis : JTH, 18*(4), 844-847. doi:https://doi.org/10.1111/jth.14768

Tang, T., Bidon, M., Jaimes, J. A., Whittaker, G. R., & Daniel, S. (2020). Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral Research, 178*(104792) doi:https://doi.org/10.1016/j.antiviral.2020.104792

Tang, X., Fang, M., Cheng, R., Zhang, Z., Wang, Y., Shen, C., . . . Lai, R. (2020). Iron-deficiency and estrogen are associated with ischemic stroke by up-regulating transferrin to induce hypercoagulability. *Circulation Research, 127*, 651-663. doi:https://doi.org/10.1161/CIRCRESAHA.119.316453

Tang, X., Zhang, Z., Fang, M., Han, Y., Wang, G., Wang, S., . . . Lai, R. (2020). Transferrin plays a central role in coagulation balance by interacting with clotting factors. *Cell Research, 30*, 119-132. doi:https://doi.org/10.1038/s41422-019-0260-6

Teachey, D. T., & O'Connor, D. (2020). How I treat newly diagnosed T-cell acute lymphoblastic leukemia and T-cell lymphoblastic lymphoma in children. *Blood, 135*(3), 159-166. doi:https://doi.org/10.1182/blood.2019001557

Terrell, M. R., Walenga, J. M., Koza, M. J., & Pifarré, R. (1996). Efficacy of aprotinin with various anticoagulant agents in cardiopulmonary bypass. *The Annals of Thoracic Surgery, 62*(2), 506-511.

Terwilliger, T., & Abdul-Hay, M. (2017). Acute lymphoblastic leukemia: A comprehensive review and 2017 update. *Blood Cancer Journal, 7*(6), e577. doi:https://doi.org/10.1038/bcj.2017.53

The Cancer Genome Atlas Research Network. (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature, 455*, 1061-1068. doi:https://doi.org/10.1038/nature07385

The Cancer Genome Atlas Research Network. (2013). The cancer genome atlas pan-cancer analysis project. *Nature Genetics, 45*, 1113-1120. doi:https://doi.org/10.1038/ng.2764

The Gene Ontology Consortium. (2019). The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Research, 47*(D1), D330-D338. doi:https://doi.org/10.1093/nar/gky1055

Tian, S., Xiong, Y., Liu, H., Niu, L., Guo, J., Liao, M., & Xiao, S. Y. (2020). Pathological study of the 2019 novel coronavirus disease (COVID-19) through postmortem core biopsies. *Modern Pathology : An Official Journal of the United States and Canadian Academy of Pathology, Inc., 33*(6), 1007-1014. doi:https://doi.org/10.1038/s41379-020-0536-x

Toptan, T., Hoehl, S., Westhaus, S., Bojkova, D., Berger, A., Rotter, B., . . . Widera, M. (2020). Optimized qRT-PCR approach for the detection of intra- and extra-cellular SARS-CoV-2 RNAs. *International Journal of Molecular Sciences, 21*(12), 4396. doi:https://doi.org/10.3390/ijms21124396

Totura, A. L., & Baric, R. S. (2012). SARS coronavirus pathogenesis: Host innate immune responses and viral antagonism of interferon. *Current Opinion in Virology, 2*(3), 264-275. doi:https://doi.org/10.1016/j.coviro.2012.04.004

Touyz, R. M., Alves-Lopes, R., Rios, F. J., Camargo, L. L., Anagnostopoulou, A., Arner, A., & Montezano, A. C. (2018). Vascular smooth muscle contraction in hypertension. *Cardiovascular Research, 114*(4), 529-539. doi:https://doi.org/10.1093/cvr/cvy023

Toyoshima, Y., Nemoto, K., Matsumoto, S., Nakamura, Y., & Kiyotani, K. (2020). SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. *Journal of Human Genetics, 65*, 1075-1082. doi:https://doi.org/10.1038/s10038-020-0808-9

Tuccori, M., Ferraro, S., Convertino, I., Cappello, E., Valdiserra, G., Blandizzi, C., . . . Hoehl, S. (2020). Anti-SARS-CoV-2 neutralizing monoclonal antibodies: Clinical pipeline. *mAbs, 12*(1), 1854149. doi:https://doi.org/10.1080/19420862.2020.1854149

Uhlen, M., Zhang, C., Lee, S., Sjöstedt, E., Fagerberg, L., Bidkhori, G., . . . Ponten, F. (2017). A pathology atlas of the human cancer transcriptome. *Science, 357*(6352) doi:https://doi.org/10.1126/science.aan2507

UniProt Consortium. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research, 49*(D1), D480-D489. doi:https://doi.org/10.1093/nar/gkaa1100

University of Liverpool. (2021). COVID-19 drug interactions. Retrieved from https://www.covid19-druginteractions.org/

van der Kloet, F. M., Buurmans, J., Jonker, M. J., Smilde, A. K., & Westerhuis, J. A. (2020). Increased comparability between RNA-seq and microarray data by utilization of gene sets. *PLoS Computational Biology, 16*(9), e1008295. doi:https://doi.org/10.1371/journal.pcbi.1008295

van Dorp, L., Tan, C. C. S., Lam, S. D., Richard, D., Owen, C., Berchtold, D., . . . Balloux, F. (2020). Recurrent mutations in SARS-CoV-2 genomes isolated from mink point to rapid host-adaptation. *bioRxiv, 2020.11.16.384743* doi:https://doi.org/10.1101/2020.11.16.384743

Van Wetering, S., van der Linden, A. C., van Sterkenburg, M. A., de Boer, W. I., Kuijpers, A. L., Schalkwijk, J., & Hiemstra, P. S. (2000). Regulation of SLPI and elafin release from bronchial epithelial cells by neutrophil defensins. *American Journal of Physiology - Lung Cellular and Molecular Physiology, 278*(1), L51-L58. doi:https://doi.org/10.1152/ajplung.2000.278.1.L51

Vari, I. S., Balkau, B., Kettaneh, A., André, P., Tichet, J., Fumeron, F., . . . DESIR Study Group. (2007). Ferritin and transferrin are associated with metabolic syndrome abnormalities and their change over time in a general population: Data from an epidemiological study on the insulin resistance syndrome (DESIR). *Diabetes Care, 30*(7), 1795-1801. doi:https://doi.org/10.2337/dc06-2312

Vaser, R., Adusumalli, S., Leng, S. N., Sikic, M., & Ng, P. C. (2016). SIFT missense predictions for genomes. *Nature Protocols, 11*(1), 1-9. doi:https://doi.org/10.1038/nprot.2015.123

Verhoef, V., & Fridland, A. (1985). Metabolic basis of arabinonucleoside selectivity for human leukemic T- and B-lymphoblasts. *Cancer Research, 45*, 3646-3650.

V'kovski, P., Kratzel, A., Steiner, S., Stalder, H., & Thiel, V. (2021). Coronavirus biology and replication: Implications for SARS-CoV-2. *Nature Reviews Microbiology, 19*(3), 155-170. doi:https://doi.org/10.1038/s41579-020-00468-6

Voss, J. D., Skarzynski, M., McAuley, E. M., Maier, E. J., Gibbons, T., Fries, A. C., & Chapleau, R. R. (2020). Variants in SARS-CoV-2 associated with mild or severe outcome. *medRxiv, 2020.12.01.20242149* doi:https://doi.org/10.1101/2020.12.01.20242149

Voysey, M., Clemens, S., Madhi, S. A., Weckx, L. Y., Folegatti, P. M., Aley, P. K., . . . Oxford COVID Vaccine Trial Group. (2020). Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: An interim analysis of four randomised controlled trials in brazil, south africa, and the UK. *Lancet (London, England), 397*(10269), 99-111. doi:https://doi.org/10.1016/S0140-6736(20)32661-1

Walls, A. C., Park, Y. J., Tortorici, M. A., Wall, A., McGuire, A. T., & Veesler, D. (2020). Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell, 181*(2), 281-292. doi:https://doi.org/10.1016/j.cell.2020.02.058

Wan, Y., Shang, J., Graham, R., Baric, R. S., & Li, F. (2020). Receptor recognition by the novel coronavirus from wuhan: An analysis based on decade-long structural studies of SARS coronavirus. *Journal of Virology, 94*(7), e00127-20. doi:https://doi.org/10.1128/JVI.00127-20

Wang, C., Liu, Z., Chen, Z., Huang, X., Xu, M., He, T., & Zhang, Z. (2020). The establishment of reference sequence for SARS-CoV-2 and variation analysis. *Journal of Medical Virology, 92*(6), 667-674. doi:https://doi.org/10.1002/jmv.25762

Wang, C., Wang, Z., Wang, G., Lau, J. Y., Zhang, K., & Li, W. (2021). COVID-19 in early 2021: Current status and looking forward. *Signal Transduction and Targeted Therapy, 6*(1), 114. doi:https://doi.org/10.1038/s41392-021-00527-1

Wang, H., Li, X., Li, T., Zhang, S., Wang, L., Wu, X., & Liu, J. (2020). The genetic sequence, origin, and diagnosis of SARS-CoV-2. *European Journal of Clinical Microbiology & Infectious Diseases : Official Publication of the European Society of Clinical Microbiology, 39*(9), 1629-1635. doi:https://doi.org/10.1007/s10096-020-03899-4

Wang, J. L., Lu, F. Z., Shen, X. Y., Wu, Y., & Zhao, L. T. (2014). SAMHD1 is down regulated in lung cancer by methylation and inhibits tumor cell proliferation. *Biochemical and Biophysical Research Communications, 455*(3-4), 229-233. doi:https://doi.org/10.1016/j.bbrc.2014.10.153

Wang, J. M., Tao, J., Chen, D. D., Cai, J. J., Irani, K., Wang, Q., . . . Chen, A. F. (2014). MicroRNA miR-27b rescues bone marrow-derived angiogenic cell function and accelerates wound healing in type 2 diabetes mellitus. *Arteriosclerosis, Thrombosis and Vascular Biology, 34*(1), 99-109. doi:https://doi.org/10.1161/ATVBAHA.113.302104

Wang, X., Duanmu, J., Fu, X., Li, T., & Jiang, Q. (2020). Analyzing and validating the prognostic value and mechanism of colon cancer immune microenvironment. *Journal of Translational Medicine, 18*(1), 324. doi:https://doi.org/10.1186/s12967-020-02491-w

Wang, X., Zhang, H., & Chen, X. (2019). Drug resistance and combating drug resistance in cancer. *Cancer Drug Resistance, 2*, 141-160. doi:http://dx.doi.org/10.20517/cdr.2019.10

Wass, M. N., Kelley, L. A., & Sternberg, M. J. (2010). 3DLigandSite: Predicting ligand-binding sites using similar structures. *Nucleic Acids Research, 38 (Web Server issue)*, W469-W473. doi:https://doi.org/10.1093/nar/gkq406

Wass, M. N., Ray, L., & Michaelis, M. (2019). Understanding of researcher behavior is required to improve data reliability. *GigaScience, 8*(5), giz017. doi:https://doi.org/10.1093/gigascience/giz017

Wei, Z. Y., Qiao, R., Chen, J., Huang, J., Wu, H., Wang, W. J., . . . Geng, Y. J. (2021). The influence of pre-existing hypertension on coronavirus disease 2019 patients. *Epidemiology and Infection, 149*, e4. doi:https://doi.org/10.1017/S0950268820003118

Weinreich, D. M., Sivapalasingam, S., Norton, T., Ali, S., Gao, H., Bhore, R., . . . Trial Investigators. (2021). REGN-COV2, a neutralizing antibody cocktail, in outpatients with covid-19. *The New England Journal of Medicine, 384*(3), 238-251. doi:https://doi.org/10.1056/NEJMoa2035002

Weisblum, Y., Schmidt, F., Zhang, F., DaSilva, J., Poston, D., Lorenzi, J. C., . . . Bieniasz, P. D. (2020). Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *eLife, 9*, e61312. doi:https://doi.org/10.7554/eLife.61312

Welbourn, S., Dutta, S. M., Semmes, O. J., & Strebel, K. (2013). Restriction of virus infection but not catalytic dNTPase activity is regulated by phosphorylation of SAMHD1. *Journal of Virology, 87*(21), 11516-11524. doi:https://doi.org/10.1128/JVI.01642-13

Wernig, G., Chen, S. Y., Cui, L., Van Neste, C., Tsai, J. M., Kambham, N., . . . Weissman, I. L. (2017). Unifying mechanism for different fibrotic diseases. *Proceedings of the National Academy of Sciences of the United States of America, 114*(18), 4757-4762. doi:https://doi.org/10.1073/pnas.1621375114

White, T. E., Brandariz-Nuñez, A., Valle-Casuso, J. C., Amie, S., Nguyen, L. A., Kim, B., . . . Diaz-Griffero, F. (2013). The retroviral restriction ability of SAMHD1, but not its deoxynucleotide triphosphohydrolase activity, is regulated by phosphorylation . *Cell Host & Microbe, 13*(4), 441-451. doi:https://doi.org/10.1016/j.chom.2013.03.005

Whitehurst, A. W., Wilsbacher, J. L., You, Y., Luby-Phelps, K., Moore, M. S., & Cobb, M. H. (2002). ERK2 enters the nucleus by a carrier-independent mechanism. *Proceedings of the National Academy of Sciences of the United States of America, 99*(11), 7496-7501. doi:https://doi.org/10.1073/pnas.112495999

Whittaker, A., Anson, M., & Harky, A. (2020). Neurological manifestations of COVID-19: A systematic review and current update. *Acta Neurologica Scandinavica, 142*(1), 14-22. doi:https://doi.org/10.1111/ane.13266

WHO Rapid Evidence Appraisal for COVID-19 Therapies (REACT) Working Group, Sterne, J., Murthy, S., Diaz, J. V., Slutsky, A. S., Villar, J., . . . Marshall, J. C. (2020). Association between administration of systemic corticosteroids and mortality among critically ill patients with COVID-19: A meta-analysis. *Jama, 324*(13), 1330-1341. doi:https://doi.org/10.1001/jama.2020.17023

Wibmer, C. K., Ayres, F., Hermanus, T., Madzivhandila, M., Kgagudi, P., Lambson, B. E., . . . Moore, P. L. (2021). SARS-CoV-2 501Y.V2 escapes neutralization by south african COVID-19 donor plasma. *bioRxiv,* doi:https://doi.org/10.1101/2021.01.18.427166

Wilder-Smith, A., Teleman, M. D., Heng, B. H., Earnest, A., Ling, A. E., & Leo, Y. S. (2005). Asymptomatic SARS coronavirus infection among healthcare workers, singapore. *Emerging Infectious Diseases, 11*(7), 1142-1145. doi:https://doi.org/10.3201/eid1107.041165

Wittmann, S., Behrendt, R., Eissmann, K., Volkmann, B., Thomas, D., Ebert, T., . . . Gramberg, T. (2015). Phosphorylation of murine SAMHD1 regulates its antiretroviral activity. *Retrovirology, 12*(103) doi:https://doi.org/10.1186/s12977-015-0229-6

Wolf, D. H., & Menssen, R. (2018). Mechanisms of cell regulation - proteolysis, the big surprise. *FEBS Letters, 592*(15), 2515-2524. doi:https://doi.org/10.1002/1873-3468.13109

Wolff, D., Nee, S., Hickey, N. S., & Marschollek, M. (2021). Risk factors for covid-19 severity and fatality: A structured literature review. *Infection, 49*(1), 15-28. doi:https://doi.org/10.1007/s15010-020-01509-1

Wongvibulsin, S., Garibaldi, B. T., Antar, A., Wen, J., Wang, M. C., Gupta, A., . . . Robinson, M. L. (2021). Development of severe COVID-19 adaptive risk predictor (SCARP), a calculator

to predict severe disease or death in hospitalized patients with COVID-19. *Annals of Internal Medicine, M20-6754* doi:https://doi.org/10.7326/M20-6754

World Health Organization. (2021). Cancer. Retrieved from https://www.who.int/health-topics/cancer#tab=tab_1

Wrapp, D., Wang, N., Corbett, K. S., Goldsmith, J. A., Hsieh, C. L., Abiona, O., . . . McLellan, J. S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science (New York, N.Y.), 367*(6483), 1260-1263. doi:https://doi.org/10.1126/science.abb2507

Wu, A., Peng, Y., Huang, B., Ding, X., Wang, X., Niu, P., . . . Jiang, T. (2020). Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in china. *Cell Host & Microbe, 27*(3), 325-328. doi:https://doi.org/10.1016/j.chom.2020.02.001

Wu, C., & Li, W. (2018). Genomics and pharmacogenomics of pediatric acute lymphoblastic leukemia. *Critical Reviews in Oncology/Hematology, 126*, 100-111. doi:https://doi.org/10.1016/j.critrevonc.2018.04.002

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., . . . Zhang, Y. Z. (2020). A new coronavirus associated with human respiratory disease in china. *Nature, 579*(7798), 265-269. doi:https://doi.org/10.1038/s41586-020-2008-3

Wu, Y., Wang, F., Shen, C., Peng, W., Li, D., Zhao, C., . . . Liu, L. (2020). A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science (New York, N.Y.), 368*(6496), 1274-1278. doi:https://doi.org/10.1126/science.abc2241

Xagoraris, I., Vassilakopoulos, T. P., Drakos, E., Angelopoulou, M. K., Panitsas, F., Herold, N., . . . Rassidakis, G. Z. (2021). Expression of the novel tumour suppressor sterile alpha motif and HD domain-containing protein 1 is an independent adverse prognostic factor in classical hodgkin lymphoma. *British Journal of Haematology, 193*(3), 488-496. doi:https://doi.org/10.1111/bjh.17352

Xia, H., Cao, Z., Xie, X., Zhang, X., Chen, J. Y., Wang, H., . . . Shi, P. Y. (2020). Evasion of type I interferon by SARS-CoV-2. *Cell Reports, 33*(1), 108234. doi:https://doi.org/10.1016/j.celrep.2020.108234

Xia, S., Lan, Q., Su, S., Wang, X., Xu, W., Liu, Z., . . . Jiang, S. (2020). The role of furin cleavage site in SARS-CoV-2 spike protein-mediated membrane fusion in the presence or absence of trypsin. *Signal Transduction and Targeted Therapy, 5*(1), 92. doi:https://doi.org/10.1038/s41392-020-0184-0

Xiao, H., Verdier-Pinard, P., Fernandez-Fuentes, N., Burd, B., Angeletti, R., Fiser, A., . . . Orr, G. A. (2006). Insights into the mechanism of microtubule stabilization by taxol. *Proceedings of the National Academy of Sciences of the United States of America, 103*(27), 10166-10173. doi:https://doi.org/10.1073/pnas.0603704103

Xie, X., Bartholomeusz, C., Ahmed, A. A., Kazansky, A., Diao, L., Baggerly, K. A., . . . Ueno, N. T. (2013). Bisphosphorylated PEA-15 sensitizes ovarian cancer cells to paclitaxel by impairing the microtubule-destabilizing effect of SCLIP. *Molecular Cancer Therapeutics, 12*(6), 1099-1111. doi:https://doi.org/10.1158/1535-7163.MCT-12-0737

Xie, X., Tang, H., Liu, P., Kong, Y., Wu, M., Xiao, X., . . . Xie, X. (2015). Development of PEA-15 using a potent non-viral vector for therapeutic application in breast cancer. *Cancer Letters, 356*(2 Pt B), 374-381. doi:https://doi.org/10.1016/j.canlet.2014.09.033

Xiong, Y., Liu, Y., Cao, L., Wang, D., Guo, M., Jiang, A., . . . Chen, Y. (2020). Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerging Microbes & Infections, 9*(1), 761-770. doi:https://doi.org/10.1080/22221751.2020.1747363

Xu, H., Zhong, L., Deng, J., Peng, J., Dan, H., Zeng, X., . . . Chen, Q. (2020). High expression of ACE2 receptor of 2019-nCoV on the epithelial cells of oral mucosa. *International Journal of Oral Science, 12*(1), 8. doi:https://doi.org/10.1038/s41368-020-0074-x

Xu, M., Lee, E. M., Wen, Z., Cheng, Y., Huang, W. K., Qian, X., . . . Tang, H. (2016). Identification of small-molecule inhibitors of zika virus infection and induced neural cell death via a drug repurposing screen. *Nature Medicine, 22*, 1101-1107. doi:https://doi.org/10.1038/nm.4184

Yamamoto, M., Kiso, M., Sakai-Tagawa, Y., Iwatsuki-Horimoto, K., Imai, M., Takeda, M., . . . Inoue, J. (2020). The anticoagulant nafamostat potently inhibits SARS-CoV-2 S protein-mediated fusion in a cell fusion assay system and viral infection in vitro in a cell-type-dependent manner. *Viruses, 12*(6), 629. doi:https://doi.org/10.3390/v12060629

Yan, J., Hao, C., DeLucia, M., Swanson, S., Florens, L., Washburn, M. P., . . . Skowronski, J. (2015). CyclinA2-cyclin-dependent kinase regulates SAMHD1 protein phosphohydrolase domain. *The Journal of Biological Chemistry, 290*(21), 13279-13292. doi:https://doi.org/10.1074/jbc.M115.646588

Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., & Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science (New York, N.Y.), 367*(6485), 1444-1448. doi:https://doi.org/10.1126/science.abb2762

Yanai, I., & Lercher, M. (2021). Novel predictions arise from contradictions. *Genome Biology, 22*, 153. doi:https://doi.org/10.1186/s13059-021-02371-6

Yang, C. A., Huang, H. Y., Chang, Y. S., Lin, C. L., Lai, I. L., & Chang, J. G. (2017). DNA-sensing and nuclease gene expressions as markers for colorectal cancer progression. *Oncology, 92*(2), 115-124. doi:https://doi.org/10.1159/000452281

Yang, F., Friedrichs, W. E., & Coalson, J. J. (1997). Regulation of transferrin gene expression during lung development and injury. *American Journal of Physiology, 273*, L417-L426. doi:https://doi.org/10.1152/ajplung.1997.273.2.L417

Yang, W., Soares, J., Greninger, P., Edelman, E. J., Lightfoot, H., Forbes, S., . . . Garnett, M. J. (2013). Genomics of drug sensitivity in cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Research, 41 (Database issue)*, D955-D961. doi:https://doi.org/10.1093/nar/gks1111

Yang, Y., Zhang, J., Chen, Y., Xu, R., Zhao, Q., & Guo, W. (2020). MUC4, MUC16, and TTN genes mutation correlated with prognosis, and predicted tumor mutation burden and immunotherapy efficacy in gastric cancer and pan-cancer. *Clinical and Translational Medicine, 10*(4), e155. doi:https://doi.org/10.1002/ctm2.155

Youn, J. C., Yu, H. T., Lim, B. J., Koh, M. J., Lee, J., Chang, D. Y., . . . Park, S. (2013). Immunosenescent CD8+ T cells and C-X-C chemokine receptor type 3 chemokines are increased in human hypertension. *Hypertension (Dallas, Tex : 1979), 62*(1), 126-133. doi:https://doi.org/10.1161/HYPERTENSIONAHA.113.00689

Yuan, Y., Van Allen, E. M., Omberg, L., Wagle, N., Amin-Mansour, A., Sokolov, A., . . . Liang, H. (2014). Assessing the clinical utility of cancer genomic and proteomic data across tumor types. *Nature Biotechnology, 32*(7), 644-652. doi:https://doi.org/10.1038/nbt.2940

Yue, J., Lan, S., Yuan, C., & Shen, Z. (2012). Prognostic values of filamin-A status for topoisomerase II poison chemotherapy. *International Journal of Biological Sciences, 8*(4), 442-450. doi:https://doi.org/10.7150/ijbs.4155

Yuen, C. K., Lam, J. Y., Wong, W. M., Mak, L. F., Wang, X., Chu, H., . . . Kok, K. H. (2020). SARS-CoV-2 nsp13, nsp14, nsp15 and orf6 function as potent interferon antagonists. *Emerging Microbes & Infections, 9*(1), 1418-1428. doi:https://doi.org/10.1080/22221751.2020.1780953

Zakin, M. M., Baron, B., & Guillou, F. (2002). Regulation of the tissue-specific expression of transferrin gene. *Developmental Neuroscience, 24*(2-3), 222-226. doi:https://doi.org/10.1159/000065690

Zamorano Cuervo, N., & Grandvaux, N. (2020). ACE2: Evidence of role as entry receptor for SARS-CoV-2 and implications in comorbidities. *eLife, 9*, e61390. doi:https://doi.org/10.7554/eLife.61390

Zhang, N., Wang, H., Fang, Y., Wang, J., Zheng, X., & Liu, X. S. (2015). Predicting anticancer drug responses using a dual-layer integrated cell line-drug network model. *PLoS Computational Biology, 11*(9), e1004498. doi:https://doi.org/10.1371/journal.pcbi.1004498

Zhao, H. Q., Wu, H., Meng, R., Du, S., & Tao, S. J. (2015). Distribution of serum transferrin, and its associations with metabolic disorders among chinese: A nation-wide, health and nutrition survey. *Molecular Nutrition & Food Research, 59*(8), 1535-1540. doi:https://doi.org/10.1002/mnfr.201500038

Zhao, J., Yuan, X., Frodin, M., & Grummt, I. (2003). ERK-dependent phosphorylation of the transcription initiation factor TIF-IA is required for RNA polymerase I transcription and cell growth. *Molecular Cell, 11*(2), 405-413. doi:https://doi.org/10.1016/s1097-2765(03)00036-4

Zhao, M., Yang, W., Yang, F., Zhang, L., Huang, W., Hou, W., . . . Yang, J. (2021). Cathepsin L plays a key role in SARS-CoV-2 infection in humans and humanized mice and is a promising

target for new drug development. *Signal Transduction and Targeted Therapy, 6*, 134. doi:https://doi.org/10.1038/s41392-021-00558-8

Zhirnov, O. P., Klenk, H. D., & Wright, P. F. (2011). Aprotinin and similar protease inhibitors as drugs against influenza. *Antiviral Research, 92*(1), 27-36. doi:https://doi.org/10.1016/j.antiviral.2011.07.014

Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., . . . Cao, B. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in wuhan, china: A retrospective cohort study. *Lancet (London, England), 395*(10229), 1054-1062. doi:https://doi.org/10.1016/S0140-6736(20)30566-3

Zhou, P., Yang, X., Wang, X., Hu, B., Zhang, L., Zhang, W., . . . Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature, 579*(7798), 270-273. doi:https://doi.org/10.1038/s41586-020-2012-7

Zhou, Y., Hou, Y., Shen, J., Huang, Y., Martin, W., & Cheng, F. (2020). Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. *Cell Discovery, 6*(14) doi:https://doi.org/10.1038/s41421-020-0153-3

Zhou, Y., Vedantham, P., Lu, K., Agudelo, J., Carrion Jr., R., Nunneley, J. W., . . . Simmons, G. (2015). Protease inhibitors targeting coronavirus and filovirus entry. *Antiviral Research, 116*, 76-84. doi:https://doi.org/10.1016/j.antiviral.2015.01.011

Zhu, K. W., Chen, P., Zhang, D. Y., Yan, H., Liu, H., Cen, L. N., . . . Chen, X. P. (2018). Association of genetic polymorphisms in genes involved in ara-C and dNTP metabolism pathway with chemosensitivity and prognosis of adult acute myeloid leukemia (AML). *Journal of Translational Medicine, 16*(1), 90. doi:https://doi.org/10.1186/s12967-018-1463-1

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., . . . China Novel Coronavirus Investigating and Research Team. (2020). A novel coronavirus from patients with pneumonia in china, 2019. *The New England Journal of Medicine, 382*(8), 727-733. doi:https://doi.org/10.1056/NEJMoa2001017

Zhu, Z., Lian, X., Su, X., Wu, W., Marraro, G. A., & Zeng, Y. (2020). From SARS and MERS to COVID-19: A brief summary and comparison of severe acute respiratory infections caused

by three highly pathogenic human coronaviruses. *Respiratory Research, 21*(1), 224. doi:https://doi.org/10.1186/s12931-020-01479-w

Zwaan, C. M., Kowalczyk, J., Schmitt, C., Bielorai, B., Russo, M. W., Woessner, M., . . . Leverger, G. (2017). Safety and efficacy of nelarabine in children and young adults with relapsed or refractory T-lineage acute lymphoblastic leukaemia or T-lineage lymphoblastic lymphoma: Results of a phase 4 study. *British Journal of Haematology, 179*(2), 284-293. doi:https://doi.org/10.1111/bjh.14874

Appendix

Supplementary Materials for Chapter 2

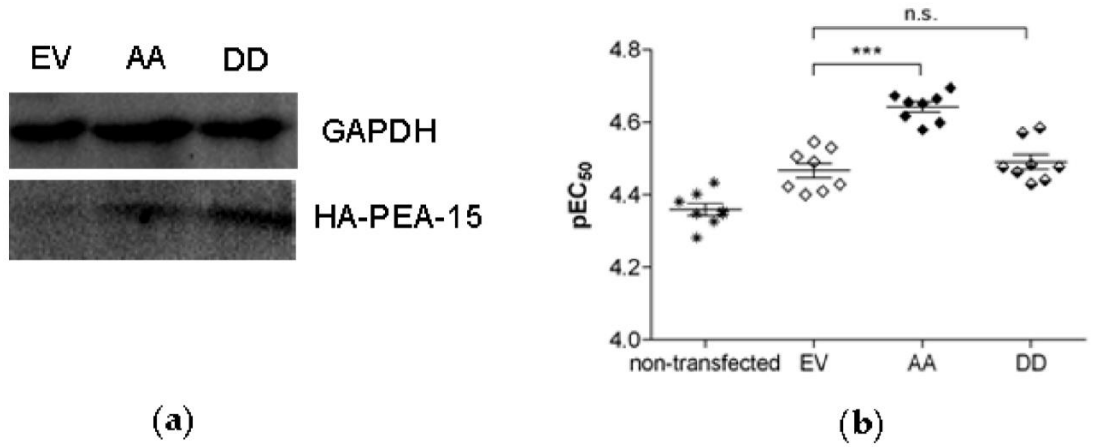


Figure S2.1. (**A**) Expression of hemagglutinin (HA)-tagged PEA-15 in EFO27^{r}CDDP^{2000} cells after transfection with the HA-tagged empty vector (EV), PEA-15AA (AA) and PEA-15DD (DD). GAPDH was used as a loading control. (**B**) Cisplatin cytotoxicity (pEC_{50}, mean ± SEM, *n* = 8) in nontransfected EFO27^{r}CDDP^{2000} cells, cells transfected with empty vector (EV), with PEA-15AA (AA), and with PEA-15DD (DD). ***p < 0.001, n.s. = not significant.

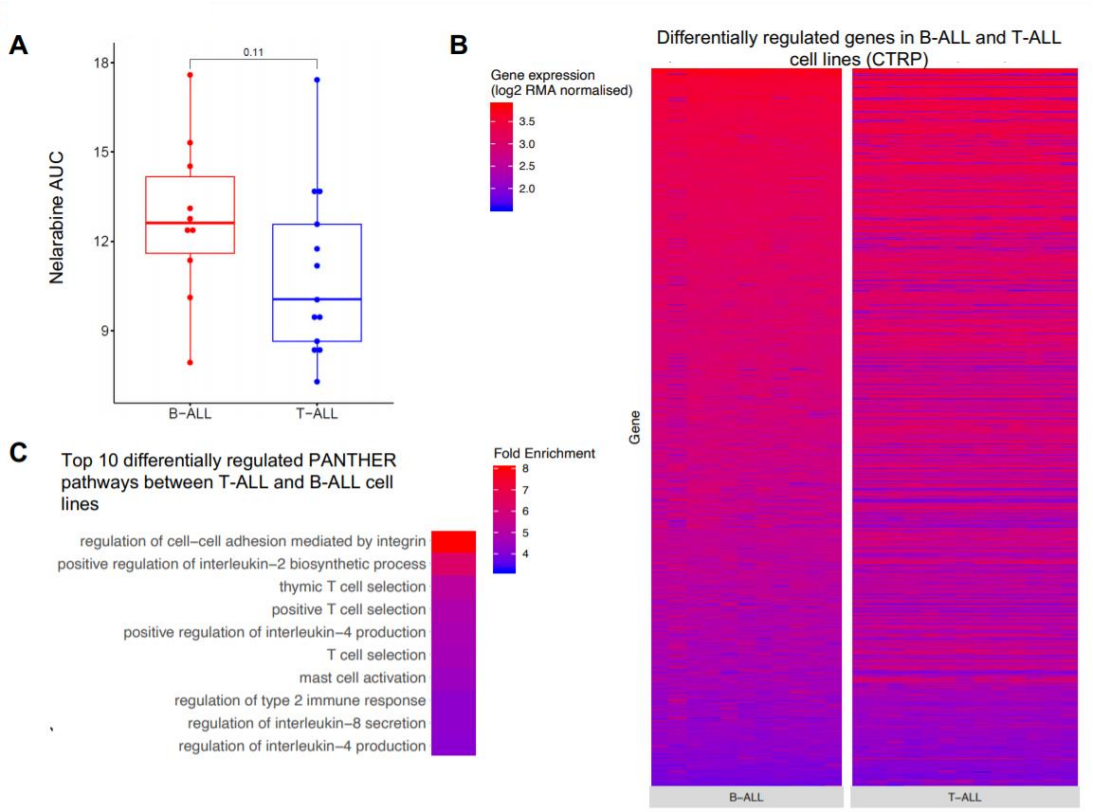
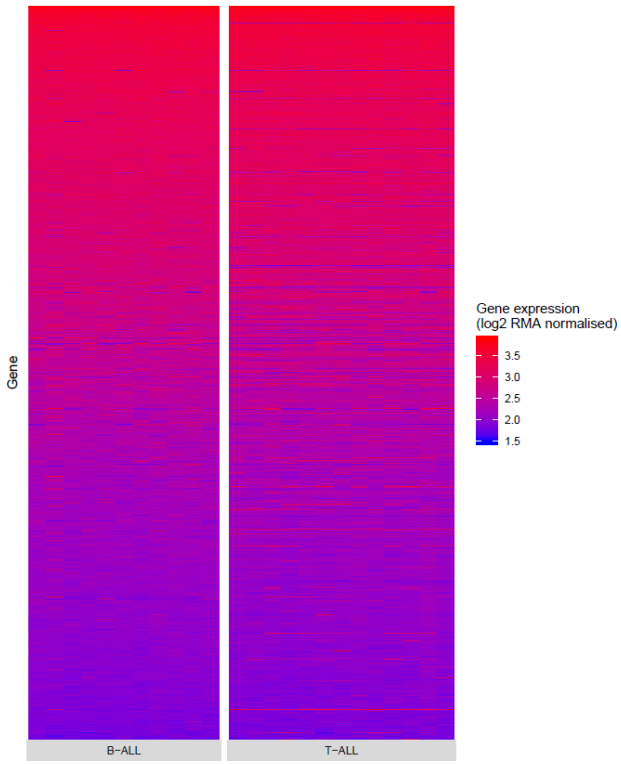
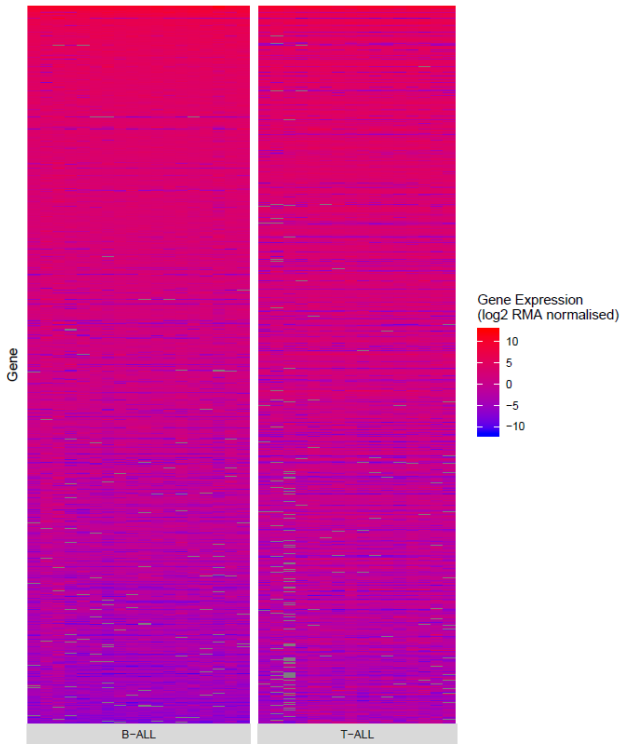


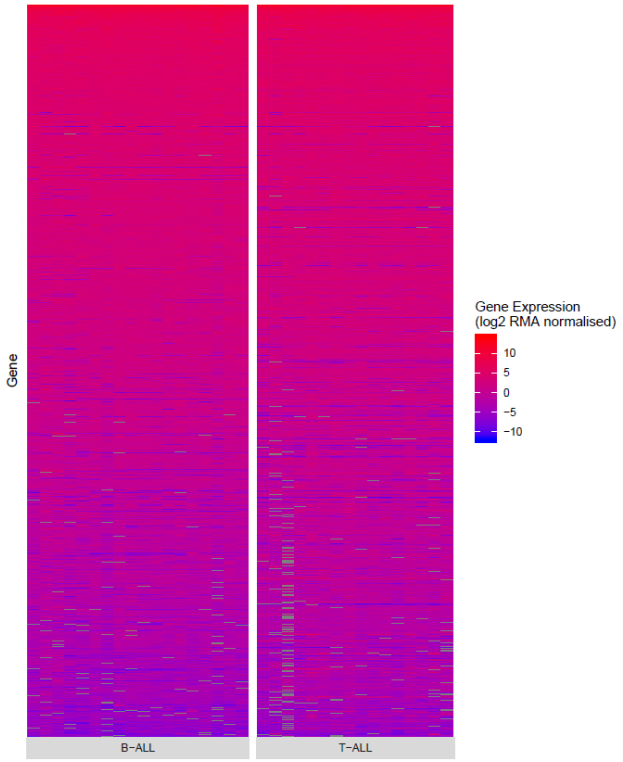
Figure S3.1. Gene expression profiles and nelarabine sensitivity in acute lymphoblastic leukaemia (ALL) cell lines. (**A**) Nelarabine sensitivity expressed as area under the curve (AUC) in T-cell precursor ALL (T-ALL) and B-ALL cell lines from CTRP. (**B**) Heatmap illustrating expression patterns of genes differentially regulated between T-ALL and B-ALL cell lines based on CTRP data. Heatmaps displaying the expression of all genes in the CTRP ALL cell lines and those displaying gene expression in the ALL cell lines in the CCLE and GDSC datasets are provided in Figure S3.1D-H. Individual gene expression values are presented in Supplementary Data 3.1.



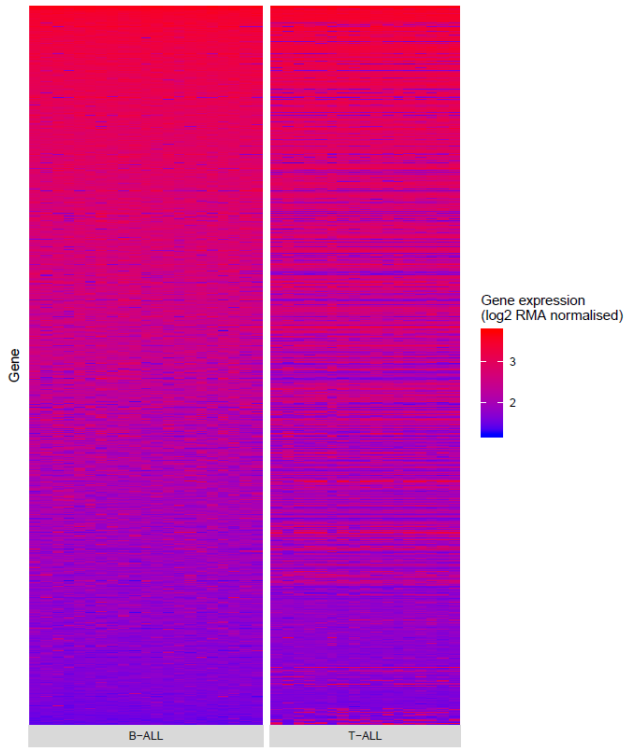
(**D**) Heatmap illustrating expression (mRNA abundance) of all genes in B- vs. T-ALL cells based on CTRP data.



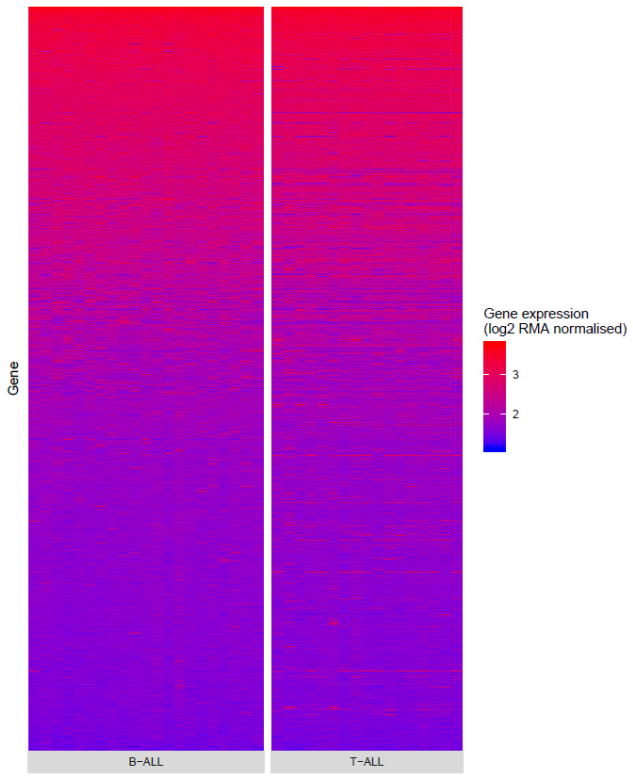
(**E**) Heatmap illustrating expression (mRNA abundance) of differentially regulated genes in B- vs. T-ALL cells based on CCLE data.



(**F**) Heatmap illustrating expression (mRNA abundance) of all genes in B- vs. T-ALL cells based on CCLE data.



(**G**) Heatmap illustrating expression (mRNA abundance) of differentially regulated genes in B- vs. T-ALL cells based on GDSC data.



(**H**) Heatmap illustrating expression (mRNA abundance) of all genes in B- vs. T-ALL cells based on GDSC data.

A

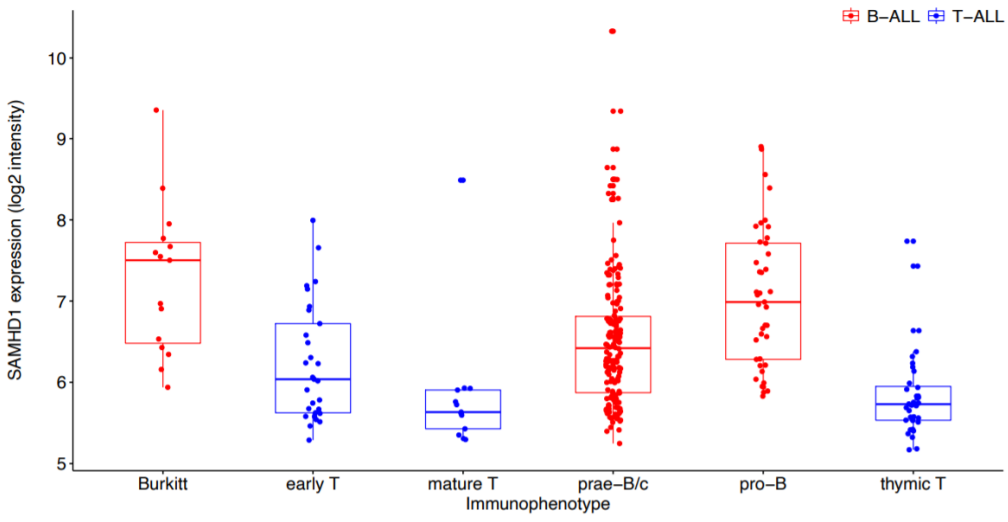


Figure S3.2A. SAMHD1 expression in ALL patients with different immunophenotypes.

	***P* value**
pre-B/ c-Burkitt	0.0067358
early T/ Burkitt	0.00073591
pro-B/ Burkitt	0.927323943
mature T/ Burkitt	5.00E-05
thy T-Burkitt	2.23E-07
early T/ pre-B/c	0.478671852
pro-B/ pre-B/c	0.00285102
mature T/ pre-B/c	0.040886675
thy T/ pre-B/c	9.73E-05
pro-B/ early T	0.000553081
mature T/ early T	0.68842299
thy T/ early T	0.379143369
mature T/ pro-B	5.40E-05
thy T/ pro-B	3.67E-09
thy T/ mature T	1

Figure S3.2A. SAMHD1 expression in ALL patients with different immunophenotypes. P values for comparisons between individual groups.

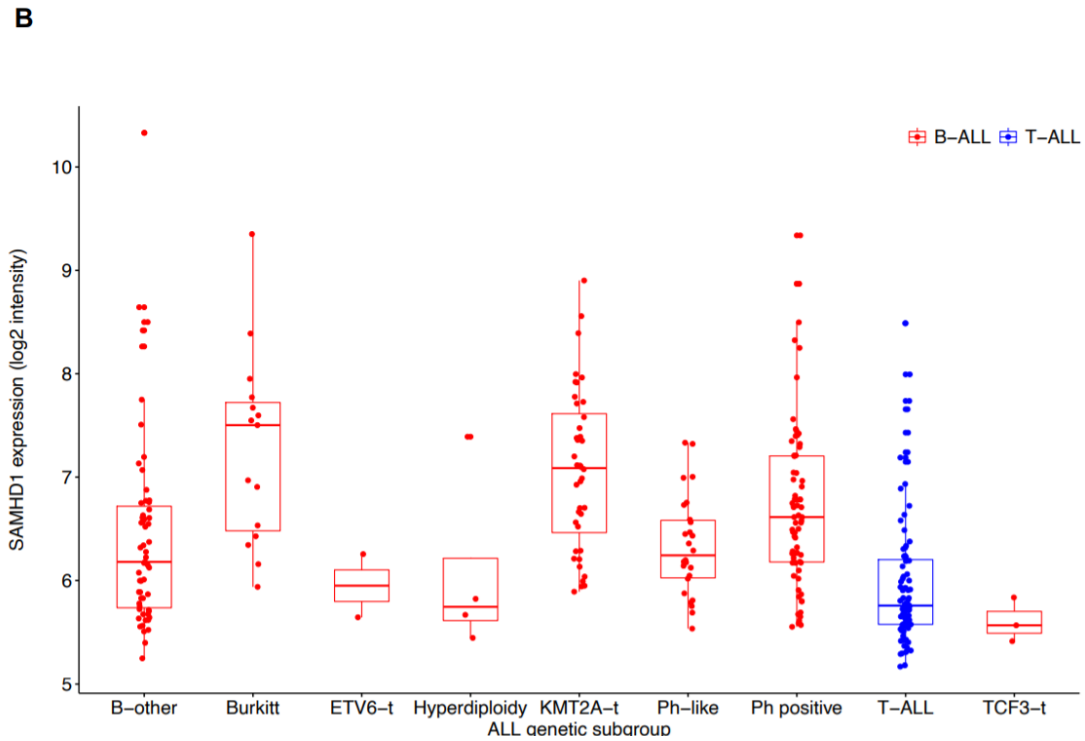


Figure S3.2B. SAMHD1 expression in ALL patients with different genotypes.

	***P* value**
2 vs. 1	0.38282375
3 vs. 1	0.999905
4 vs. 1	0.00230736
5 vs. 1	0.02880067
6 vs. 1	0.00619649
7 vs. 1	0.70123101
8 vs. 1	0.99542401
9 vs. 1	0.99530362

3 vs. 2	0.38981245
4 vs. 2	0.41595306
5 vs. 2	2.73E-07
6 vs. 2	0.26166224
7 vs. 2	0.2631177
8 vs. 2	0.8994844
9 vs. 2	0.79514518
4 vs. 3	0.00682233
5 vs. 3	0.55549758
6 vs. 3	0.0073478
7 vs. 3	0.84170864
8 vs. 3	0.99912995
9 vs. 3	0.99961577
5 vs. 4	1.60E-10
6 vs. 4	0.99429435
7 vs. 4	0.04940224
8 vs. 4	0.56219644
9 vs. 4	0.28423458
6 vs. 5	3.79E-07
7 vs. 5	0.99596858
8 vs. 5	1
9 vs. 5	0.99999962
7 vs. 6	0.02309725
8 vs. 6	0.37573601

9 vs. 6	0.14840814
8 vs. 7	0.99992001
9 vs. 7	0.99681029
9 vs. 8	0.99999994

Figure S3.2B. SAMHD1 expression in ALL patients with different genotypes. P values for comparisons between individual groups (1=B-other, 2=Ph pos, 3=Ph-like, 4=KMT2A, 5=T-ALL, 6=Burkitt, 7=TCF3, 8=ETV6, 9=Hyperdip.

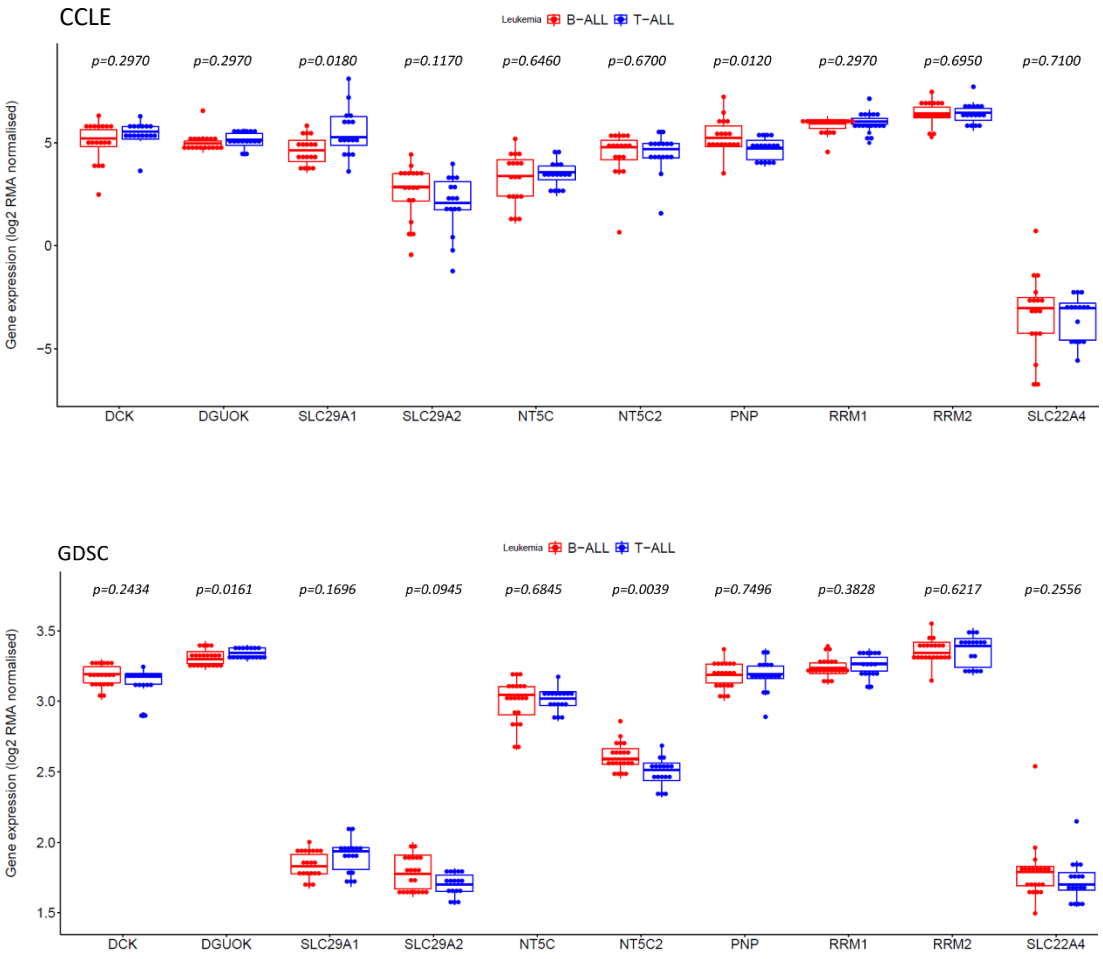


Figure S3.3. Expression of genes known to be potentially involved in nucleoside analogue activity in B-ALL and T-ALL cell lines in the CCLE and GDSC.

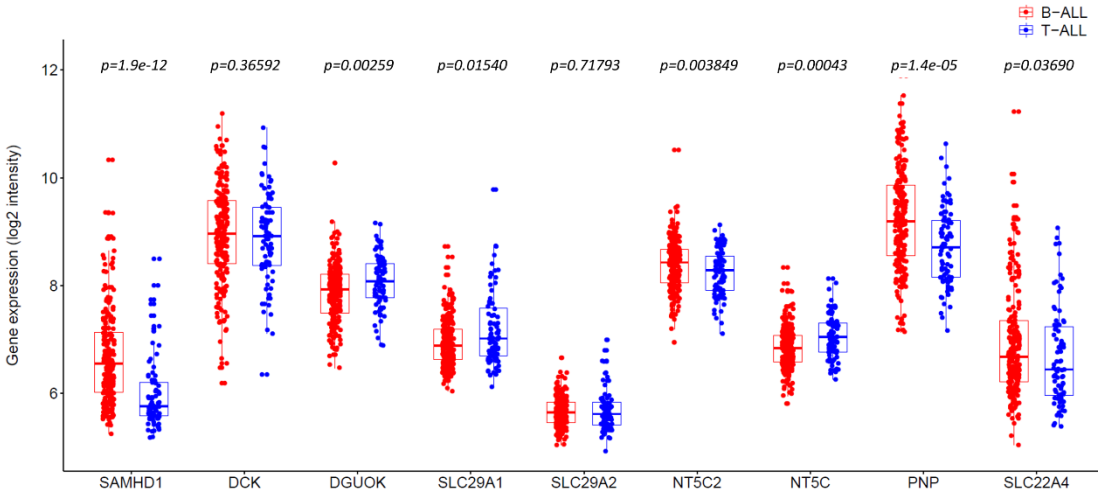
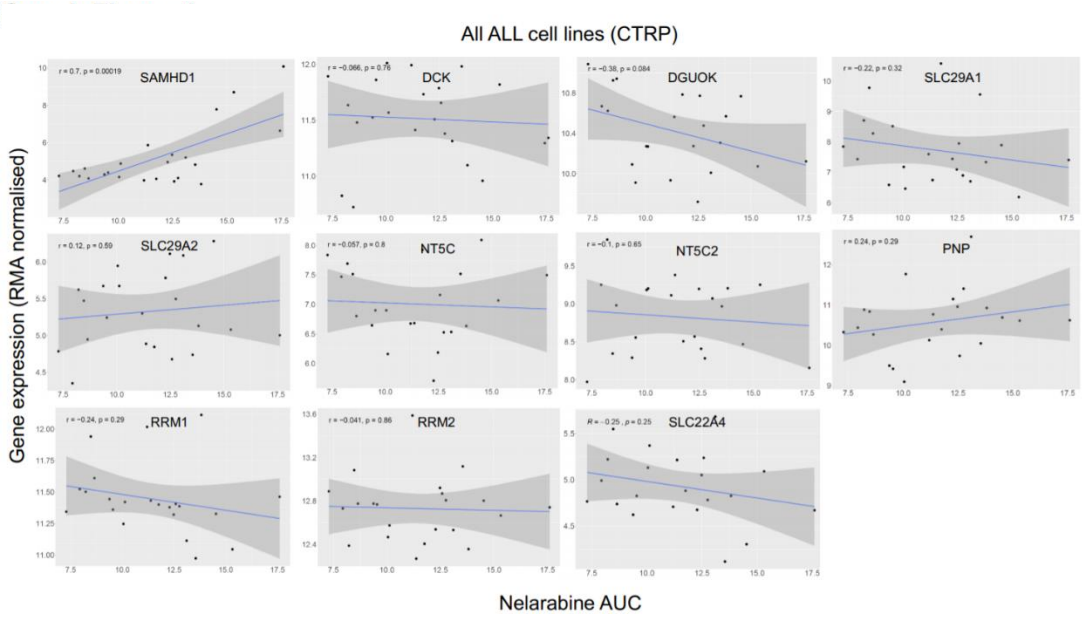


Figure S3.3. Expression of genes known to be potentially involved in nucleoside analogue activity in patient-derived B-ALL and T-ALL cells.



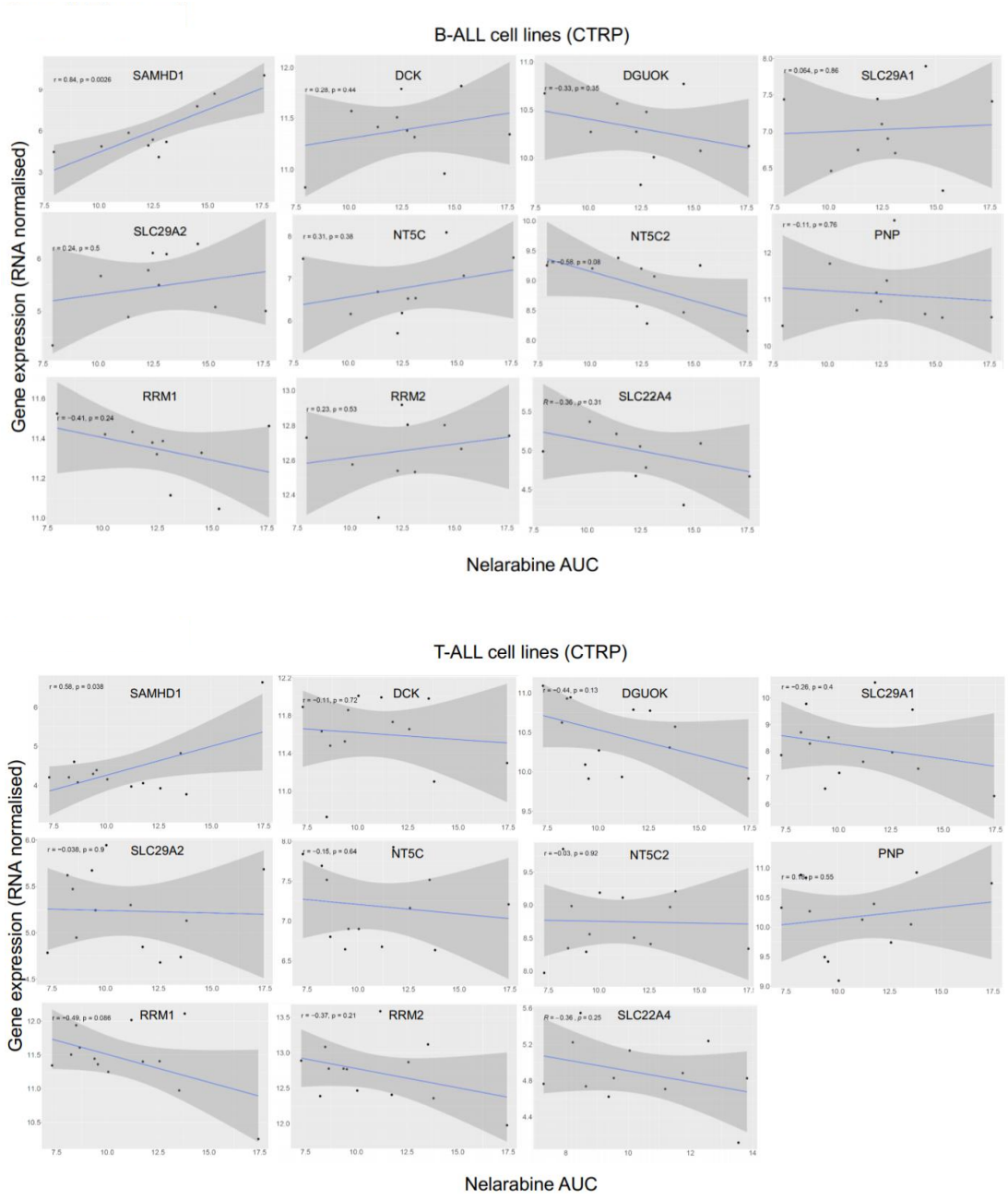


Figure S3.4. Correlation of the expression of genes (mRNA abundance) known to affect nucleoside analogue activity to the nelarabine sensitivity (expressed as AUC) across all ALL, the B-ALL and the T-ALL cell lines based on CTRP data. Pearson's r values and respective p-values are provided.

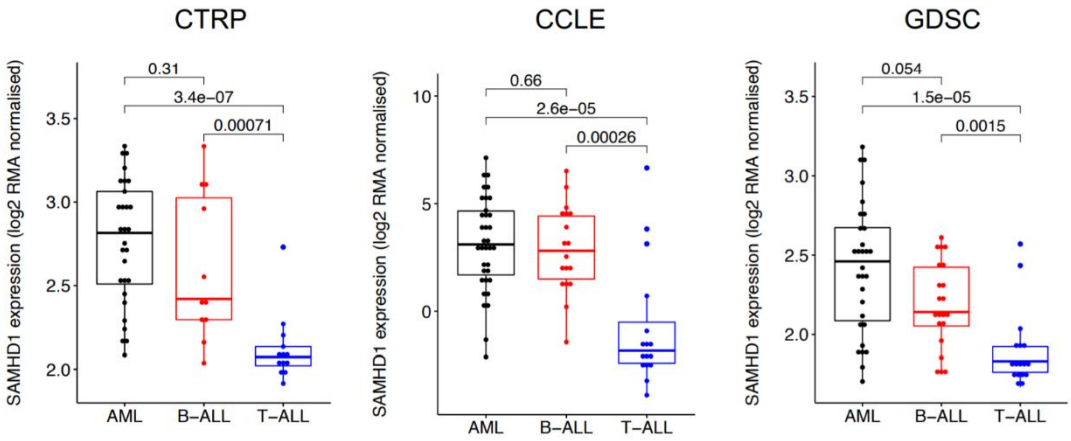


Figure S3.5. Comparison of SAMHD1 expression (mRNA abundance) levels in acute myeloid leukaemia (AML), B-cell acute lymphoblastic leukaemia (B-ALL), T-cell acute lymphoblastic leukaemia (T-ALL) cells in CTRP, CCLE, and GDSC. Respective p-values are provided.

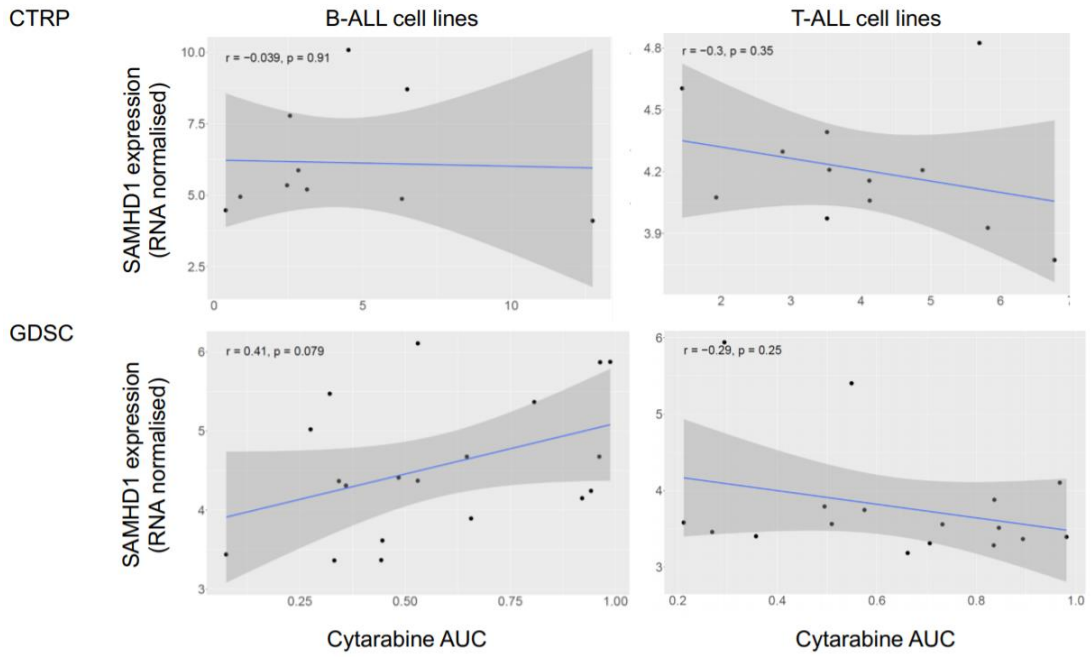


Figure S3.6. Correlations of SAMHD1 expression (mRNA abundance) with the cytarabine AUC exclusively in B-ALL and T-ALL cell lines based on CTRP and GDSC data. Pearson's r values and respective p-values are provided.

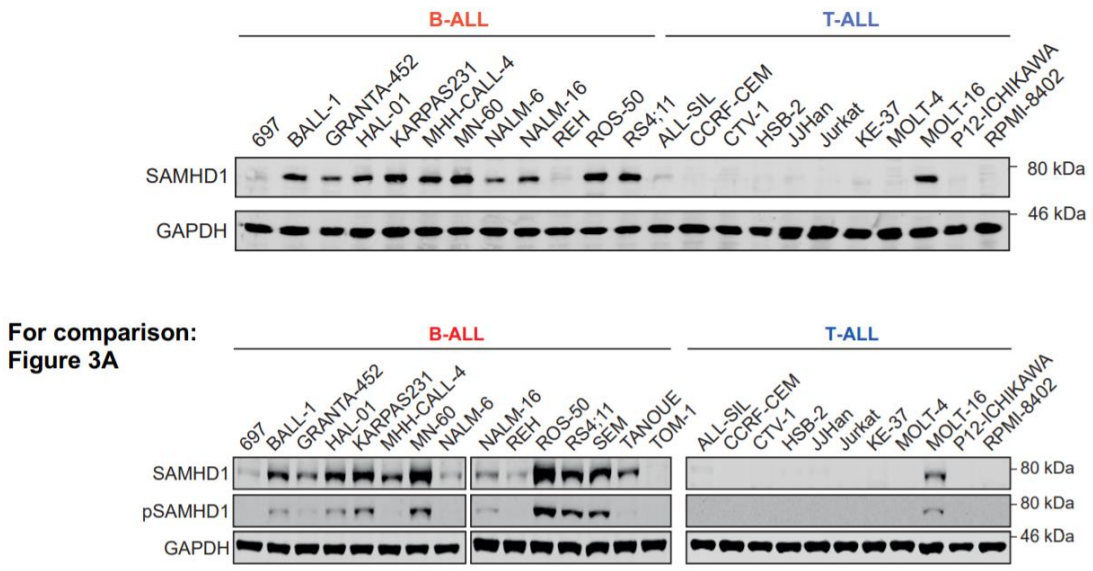


Figure S3.7. SAMHD1 protein levels in the RCCL panel of B-ALL and T-ALL cell lines. Representative Western blots indicating protein levels of total SAMHD1 and GAPDH in 23 cell lines of the RCCL panel, which were run on the same gel and blotted on the same membrane to confirm the representativeness of the blots provided in Figure 3.3A. Figure 3.3A is provided for comparison.

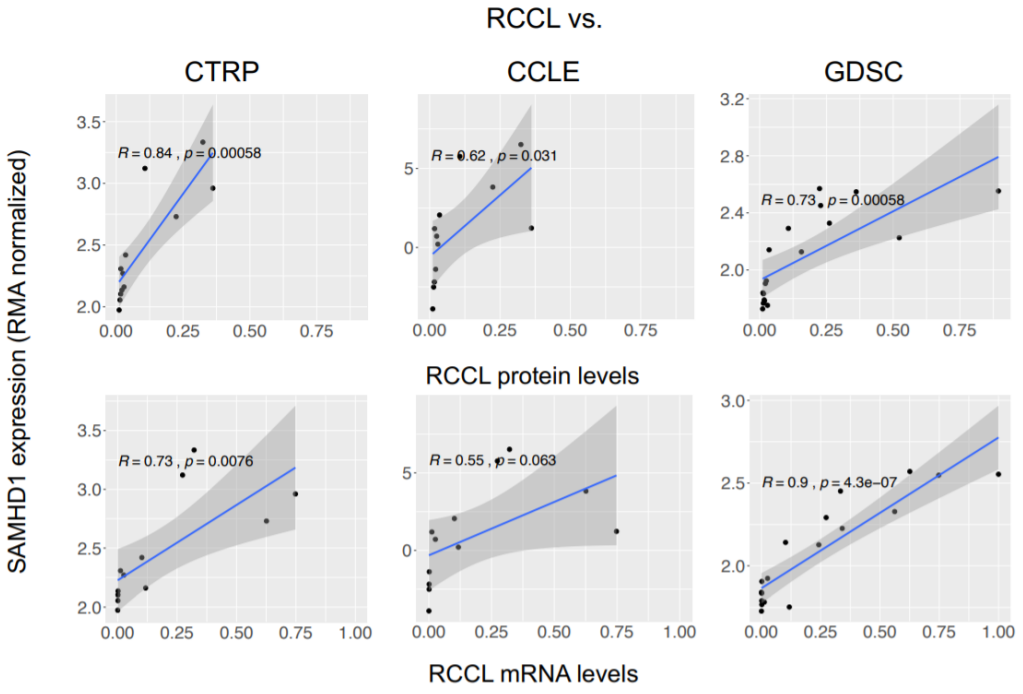


Figure S3.8. Correlations of SAMHD1 protein and mRNA levels determined in the RCCL cell lines with the SAMHD1 expression data derived from the CTRP, CCLE, and GDSC among the cell lines that are represented in both respective datasets. Pearson's r values and respective p-values are provided.

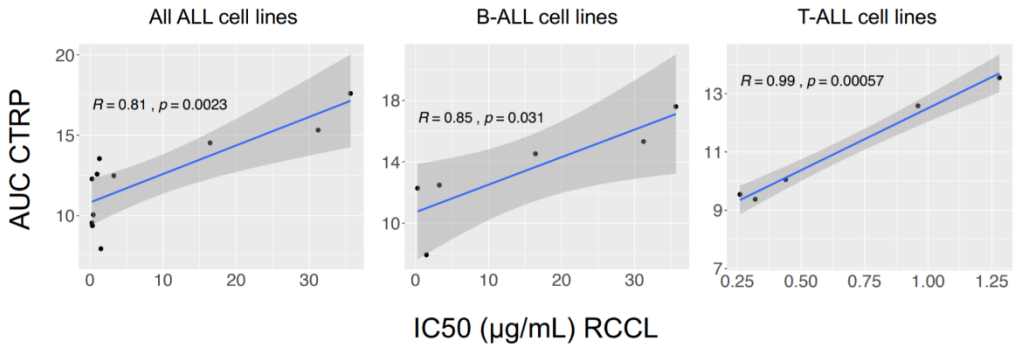
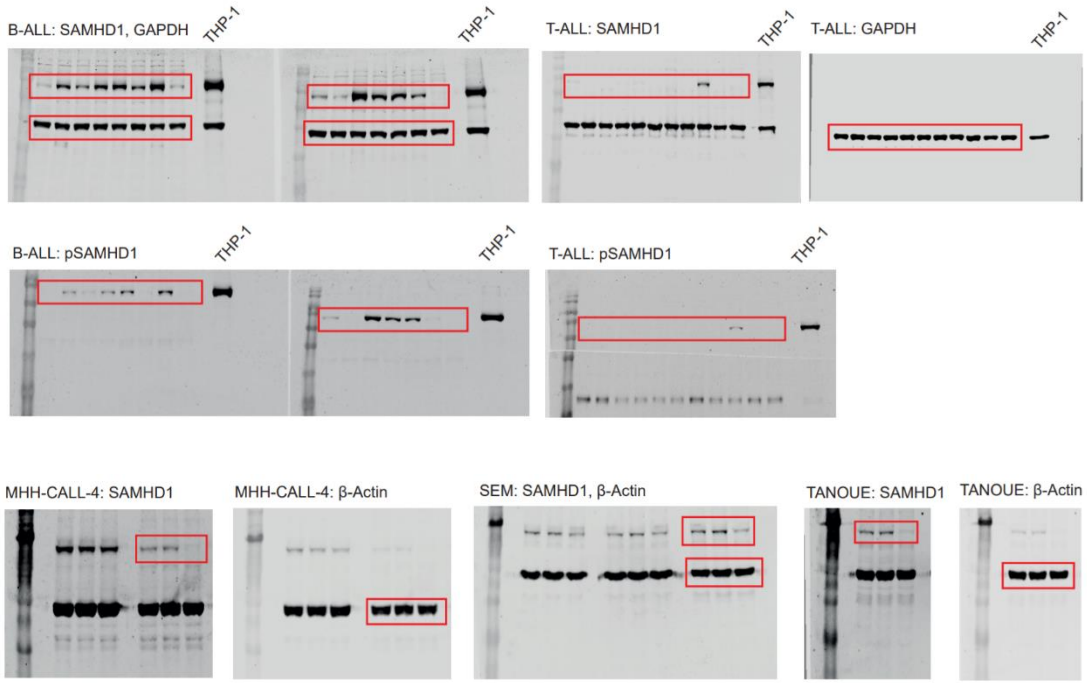


Figure S3.9. Correlations of the nelarabine AUCs derived from the CTRP and the AraG IC50 values determined in the RCCL panel across the ALL cell lines present in both datasets. Pearson's r values and respective p-values are provided.



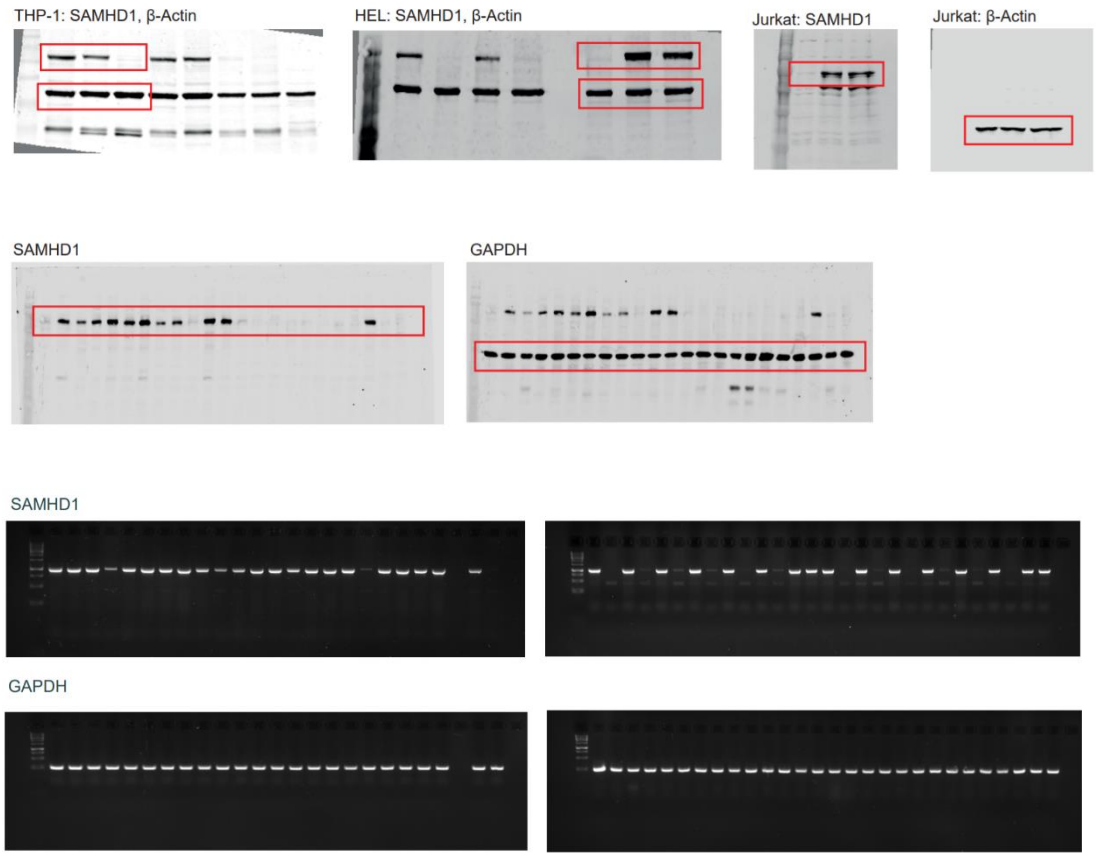


Figure S3.10. Uncropped Western blots and agarose gels.

Table S3.1. B-ALL and T-ALL cell lines in the CCLE and GDSC (overlaps highlighted in *italics*)

CCLE		**GDSC**	
Cell Line	**Lineage**	**Cell Line**	**Lineage**
697	B-ALL	*697*	B-ALL
A4FUK	B-ALL	ALL-PO	B-ALL
EHEB	B-ALL	BALL-1	B-ALL
HUNS1	B-ALL	GR-ST	B-ALL
JM-1	B-ALL	HAL-01	B-ALL
KASUMI-2	B-ALL	KARPAS-231	B-ALL
KOPN-8	B-ALL	*KOPN-8*	B-ALL

MHH-CALL2	B-ALL	LC4-1	B-ALL
MHH-CALL3	B-ALL	*MHH-CALL-2*	B-ALL
MHH-CALL4	B-ALL	*MHH-CALL-4*	B-ALL
MUTZ-5	B-ALL	MHH-PREB-1	B-ALL
NALM-19	B-ALL	MN-60	B-ALL
NALM-6	B-ALL	*NALM-6*	B-ALL
RCH-ACV	B-ALL	P30-OHK	B-ALL
REH	B-ALL	*RCH-ACV*	B-ALL
RS-411	B-ALL	*REH*	B-ALL
SEM	B-ALL	ROS-50	B-ALL
SUP-B15	B-ALL	*RS4-11*	B-ALL
		SUP-B15	B-ALL
		SUP-B8	B-ALL
		U-698-M	B-ALL
ALL-SIL	T-ALL	*ALL-SIL*	T-ALL
C8166	T-ALL	ATN-1	T-ALL
DND-41	T-ALL	BE-13	T-ALL
HPB-ALL	T-ALL	CCRF-CEM	T-ALL
JURKAT	T-ALL	*DND-41*	T-ALL
KE-37	T-ALL	HH	T-ALL
LOUCY	T-ALL	*JURKAT*	T-ALL
MOLT-13	T-ALL	KARPAS-45	T-ALL
MOLT-16	T-ALL	*KE-37*	T-ALL
MOLT-3	T-ALL	*LOUCY*	T-ALL

P12-ICHIKAWA	T-ALL	*MOLT-13*	T-ALL
PEER	T-ALL	*MOLT-16*	T-ALL
PF-382	T-ALL	MOLT-4	T-ALL
RPMI-8402	T-ALL	*P12-ICHIKAWA*	T-ALL
SUP-T11	T-ALL	*PF-382*	T-ALL
TALL-1	T-ALL	*RPMI-8402*	T-ALL
		SUP-T11	T-ALL

Table S3.2. B-ALL and T-ALL cell line sensitivity to nelarabine expressed as area under the curve (AUC) derived from CTRP.

Cell Line	**Lineage**	**AUC**
HPBALL	T-ALL	11.757
DND-41	T-ALL	7.29
SUPT-1	T-ALL	8.4742
JURKAT	T-ALL	12.58
PEER	T-ALL	13.818
PF-382	T-ALL	11.193
ALL-SIL	T-ALL	13.546
P12-ICHIKAWA	T-ALL	9.5418
RPMI-8402	T-ALL	10.052
MOLT-16	T-ALL	17.44
MOLT-13	T-ALL	8.6473
TALL-1	T-ALL	8.2253
KE-37	T-ALL	9.3765

SEM	B-ALL	17.602
RCH-ACV	B-ALL	12.759
MHH-CALL3	B-ALL	10.123
RS-411	B-ALL	14.526
MHH-CALL4	B-ALL	15.322
REH	B-ALL	7.9391
697	B-ALL	12.28
SUP-B15	B-ALL	11.372
KASUMI-2	B-ALL	13.107
NALM-6	B-ALL	12.483

Table S3.3. AraG and cytarabine concentrations that reduce B-ALL and T-ALL cell line sensitivity by 50% (IC50).

B-ALL	**IC50**	
Cell Line	**AraG (µg/mL)**	**Cytarabine (ng/mL)**
697	0.27 ± 0.01	1.23 ± 0.05
BALL-1	4.76 ± 0.50	2.74 ± 0.05
GRANTA-452	6.69 ± 0.70	3.42 ± 0.25
HAL-01	2.15 ± 0.49	0.98 ± 0.04
KARPAS231	26.73 ± 2.62	11.57 ± 0.77
MHH-CALL-4	31.22 ± 2.50	27.07 ± 3.05
MN-60	99.10 ± 1.62	14.95 ± 0.99
NALM-6	3.25 ± 0.28	2.19 ± 0.06
NALM-16	65.19 ± 2.72	8.79 ± 0.42

REH	1.48 ± 0.07	1.31 ± 0.17
ROS-50	90.62 ± 9.05	18.28 ± 3.57
RS4;11	16.42 ± 1.32	5.74 ± 1.89
SEM	35.64 ± 3.71	27.35 ± 3.18
TANOUE	49.14 ± 2.95	27.67 ± 0.60
TOM-1	0.10 ± 0.01	1.38 ± 0.03
T-ALL	**IC50**	
Cell Line	**AraG (µg/mL)**	**Cytarabine (ng/mL)**
ALL-SIL	1.28 ± 0.16	5.17 ± 0.64
CCRF-CEM	0.43 ± 0.02	3.88 ± 0.70
CTV-1	0.38 ± 0.07	1.72 ± 0.01
HSB-2	0.52 ± 0.05	3.51 ± 0.07
JJHan	0.97 ± 0.19	5.61 ± 0.53
Jurkat	0.96 ± 0.06	6.65 ± 0.52
KE-37	0.32 ± 0.11	1.83 ± 0.25
MOLT-4	0.46 ± 0.01	3.02 ± 0.10
MOLT-16	15.55 ± 1.30	6.48 ± 0.53
P12-ICHIKAWA	0.26 ± 0.01	2.12 ± 0.19
RPMI-8402	0.44 ± 0.01	3.43 ± 0.21

Supplementary Files:

Supplementary Data 3.1. Gene transcripts differentially regulated (mRNA abundance) between T-ALL and B-ALL cell lines based on data derived from GDSC, CCLE, or CTRP.

Link:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7314829/bin/42003_2020_1052_MOESM2_ESM.xlsx

Supplementary Data 3.2. Pathway analysis using the PANTHER database to identify differentially regulated processes based on genes differentially regulated between B-ALL and T-ALL cell lines in the CTRP, CCLE, and GDSC.

Link:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7314829/bin/42003_2020_1052_MOESM3_ESM.xlsx

Supplementary Data 3.3. Genes whose expression is directly or inversely correlated with the nelarabine AUC based on CTRP data.

Link:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7314829/bin/42003_2020_1052_MOESM4_ESM.xlsx

Supplementary Data 3.4. Genes whose expression is directly or inversely correlated with the nelarabine AUC based on CTRP data.

Link:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7314829/bin/42003_2020_1052_MOESM5_ESM.xlsx

Supplementary Data 3.5. Data underlying graphs.

Link:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7314829/bin/42003_2020_1052_MOESM6_ESM.xlsx

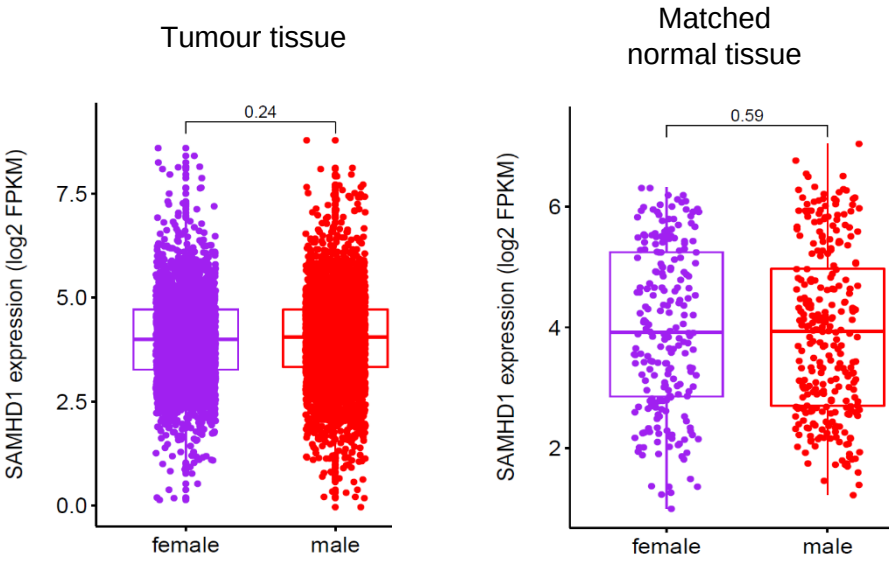


Figure S4.1. Comparison of SAMHD1 expression levels as expressed by fragments per kilobase of transcript per million mapped reads (FPKM) after removal of sex-specific cancer types and of BRCA (breast invasive carcinoma), for which only a very small fraction of samples (12/ 1,089 tumour tissue samples, 1/ 113 matched normal tissue samples) was derived from males. P-values were determined by Mann-Whitney U (Wilcoxon) test for independent groups.

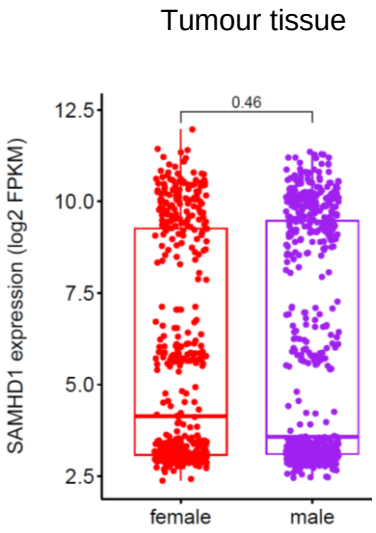


Figure S4.1. Comparison of *SAMHD1* expression levels as expressed by fragments per kilobase of transcript per million mapped reads (FPKM) in tumours from the TARGET database. P-values determined by Mann-Whitney U (Wilcoxon) test for independent groups.

A

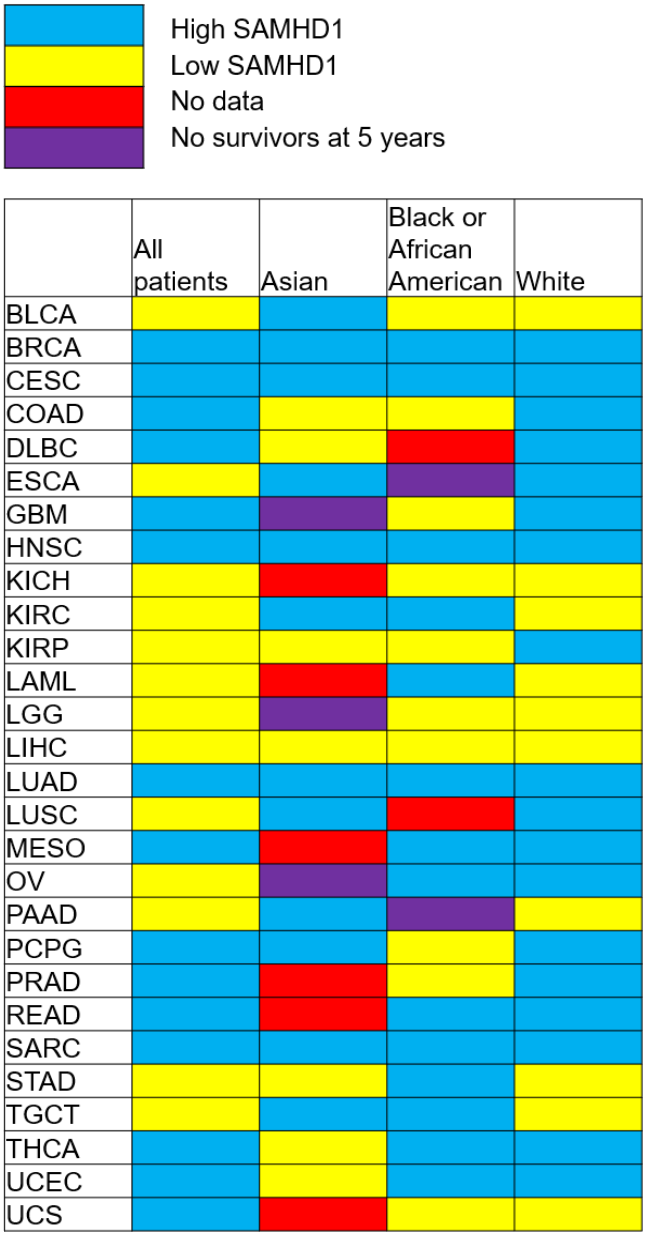
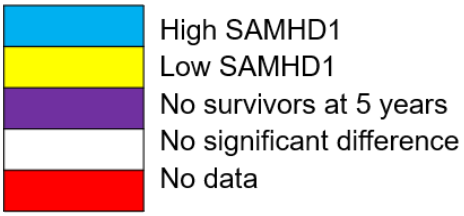


Figure S4.3A. Heatmap indicating the association of SAMHD1 expression and 5-year survival rates (blue: high SAMHD1 associated with higher survival rates, yellow: low SAMHD1 associated with higher survival rates).

B



	All patients	Asian	Black or African American	White
BLCA				
CESC				
HNSC				
KICH				
KIRC				
LAML				
LGG				
LUAD				
MESO				
OV				
PAAD				
PCPG				
PRAD				
READ				
SARC				
THCA				
UCEC				

Figure S4.3B. Heatmap indicating cancer entities in which high SAMHD1 expression (blue) or low SAMHD1 expression (yellow) is significantly ($p<0.05$) associated with higher 5-year survival rates.

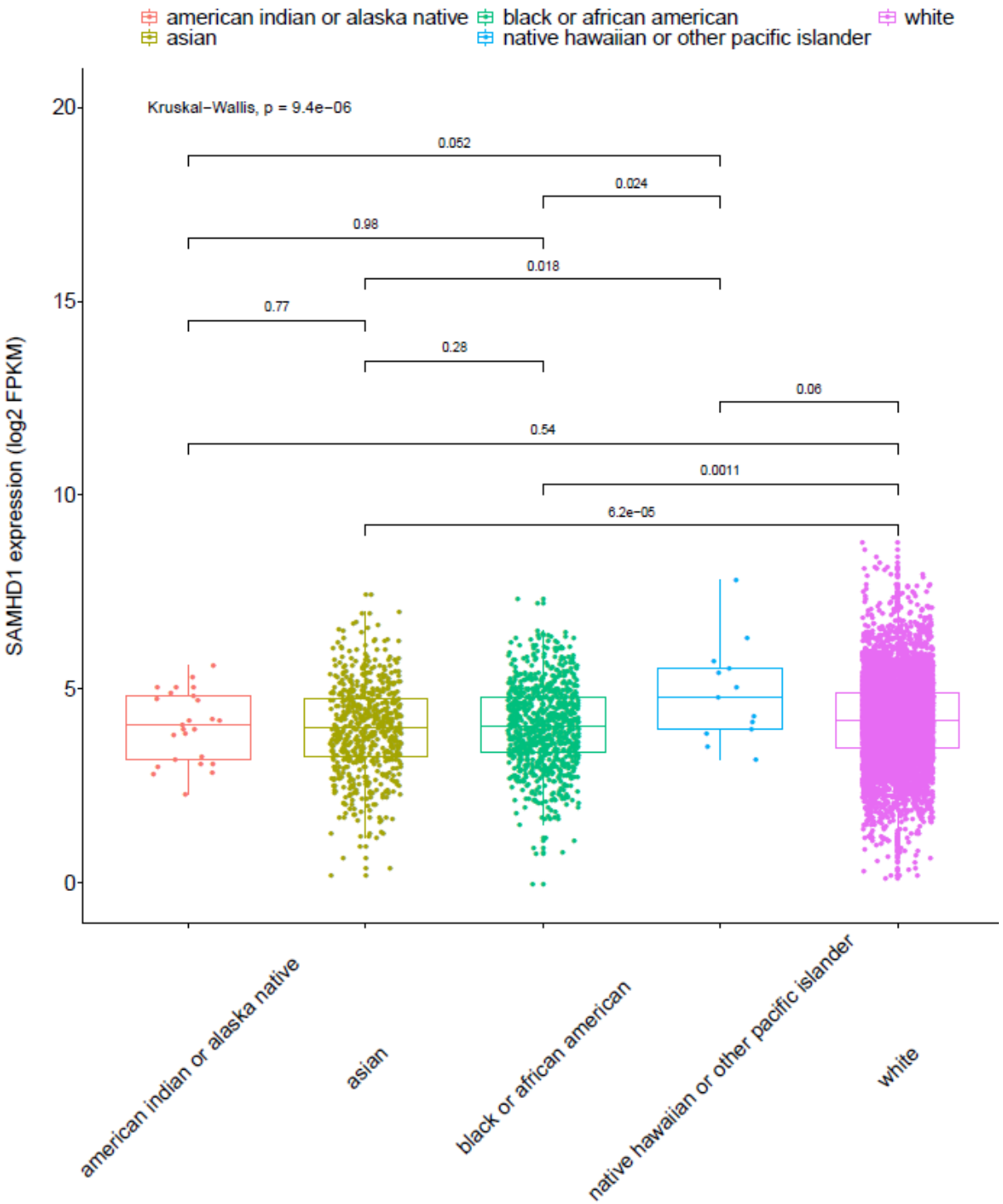


Figure S4.4. SAMHD1 levels in patients of different race in TCGA. Comparison of all groups was performed using the Kruskal-Wallis test. Individual group comparisons were performed using the Mann-Whitney U (Wilcoxon) rank sum test.

A

High SAMHD1
Low SAMHD1
No data
No survivors at 5 years

	All	Asian	Black or African American	White
AML				
ALL-P1				
ALL-P2				
NBL				
OS				
RT				
CCSK				
WT				

B

High SAMHD1
Low SAMHD1
No survivors at 5 years
No significant difference
No data

	All	Asian	Black or African American	White
ALL-P1				
ALL-P2				
OS				
RT				
WT				

Figure S4.3. Heatmap indicating the association of SAMHD1 expression and 5-year survival rates (blue: high SAMHD1 associated with higher survival rates, yellow: low SAMHD1 associated with higher survival rates). (**A**) All cancer types independently of significance level. (**B**) Cancer types in which at least comparison resulted in a significant ($p<0.05$) difference.

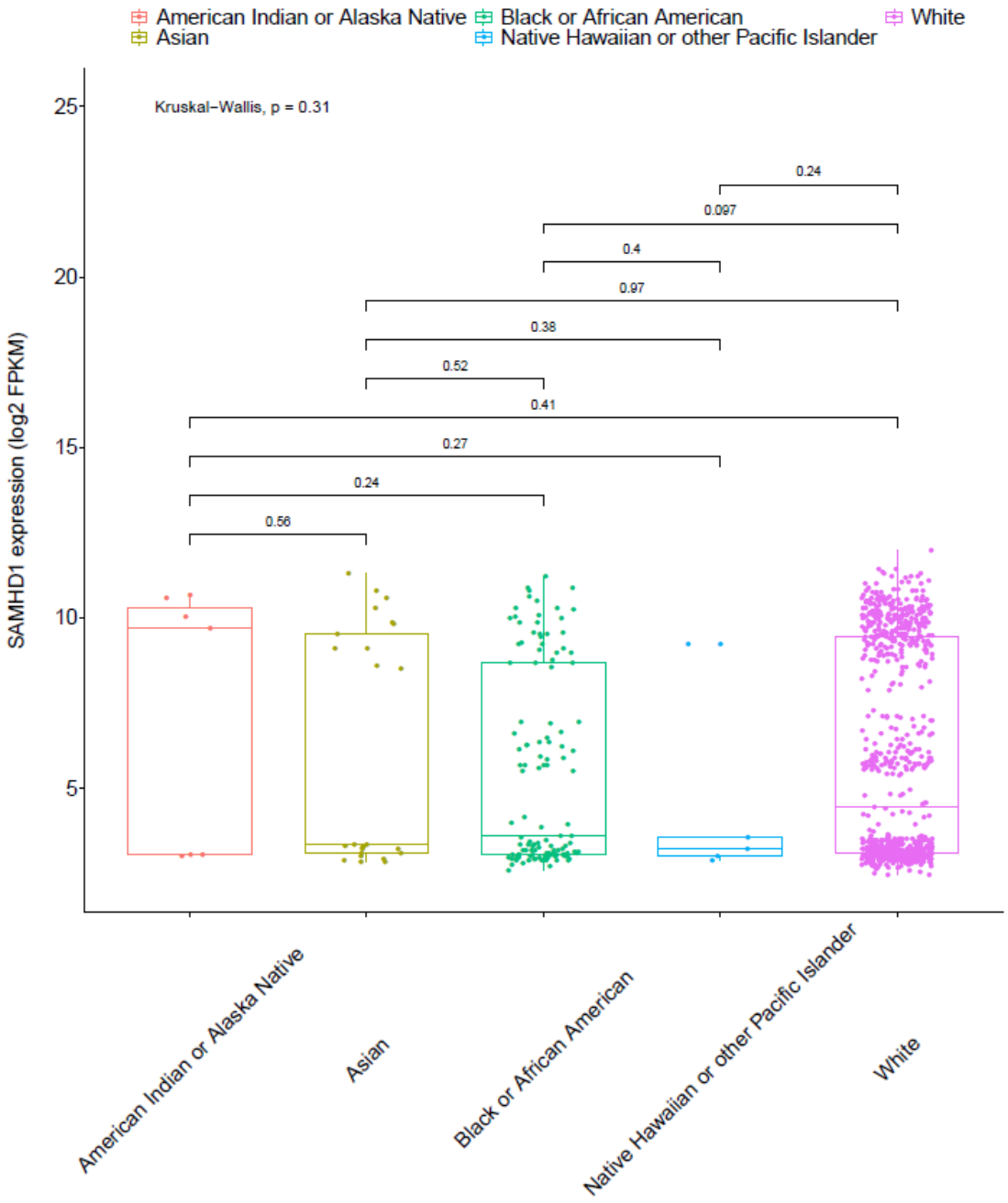


Figure S4.6. SAMHD1 levels in patients of different race in the TARGET database. Comparison of all groups was performed using the Kruskal-Wallis test. Individual group comparisons were performed using the Mann-Whitney U (Wilcoxon) rank sum test.

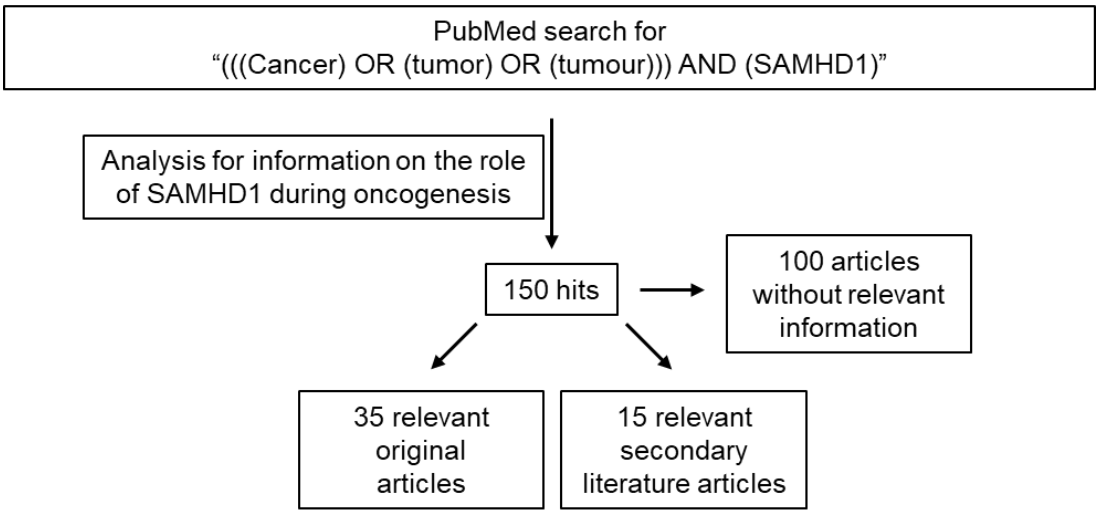


Figure S4.7. Summary of the findings of the literature review for articles containing data on the role of SAMHD1 during oncogenesis. Articles were identified by using the search term "(((Cancer) OR (tumor) OR (tumour))) AND (SAMHD1)" in PubMed (https://pubmed.ncbi.nlm.nih.gov) on 17th June 2021.

Supplementary Files:

Table S4.1. 5-year survival in cancer patients with tumours displaying high or low SAMHD1 levels as determined by best separation of data derived from TCGA.

Table S4.2. 5-year survival in cancer patients with tumours displaying high or low SAMHD1 levels as determined by best separation of data derived from TARGET.

Table S4.3. 5-year survival in cancer patients divided by sex with tumours displaying high or low SAMHD1 levels as determined by best separation of data derived from TCGA.

Table S4.4. 5-year survival in cancer patients divided by sex with tumours displaying high or low SAMHD1 levels as determined by best separation of data derived from TARGET.

Table S4.5. 5-year survival in cancer patients divided by race with tumours displaying high or low SAMHD1 levels as determined by best separation of data derived from TCGA.

Table S4.6. 5-year survival in cancer patients divided by race with tumours displaying high or low SAMHD1 levels as determined by best separation of data derived from TARGET.

Table S4.7. Median SAMHD1 levels (FPKM, fragments per kilobase of transcript per million mapped reads) in tumour and control tissues

Table S4.8. Correlation between *SAMHD1* expression levels and SAMHD1 promotor methylation in different cancer types based on data derived from TCGA.

Table S4.9. miRNAs documented to regulate SAMHD1.

Table S4.10. Correlation of miRNAs and *SAMHD1* expression in different cancer types.

Table S4.11. TCGA mutation data.

Table S4.12. 5-year survival in patients with SAMHD1 mutant or wild-type tumours based on TCGA data.

Table S4.13. Literature search performed in PubMed (https://pubmed.ncbi.nlm.nih.gov) on 17th June 2021 using the search term "(((Cancer) OR (tumor) OR (tumour))) AND (SAMHD1)".

Link:
https://www.biorxiv.org/content/biorxiv/early/2021/07/05/2021.07.03.451003/DC1/embed/media-1.zip?download=true

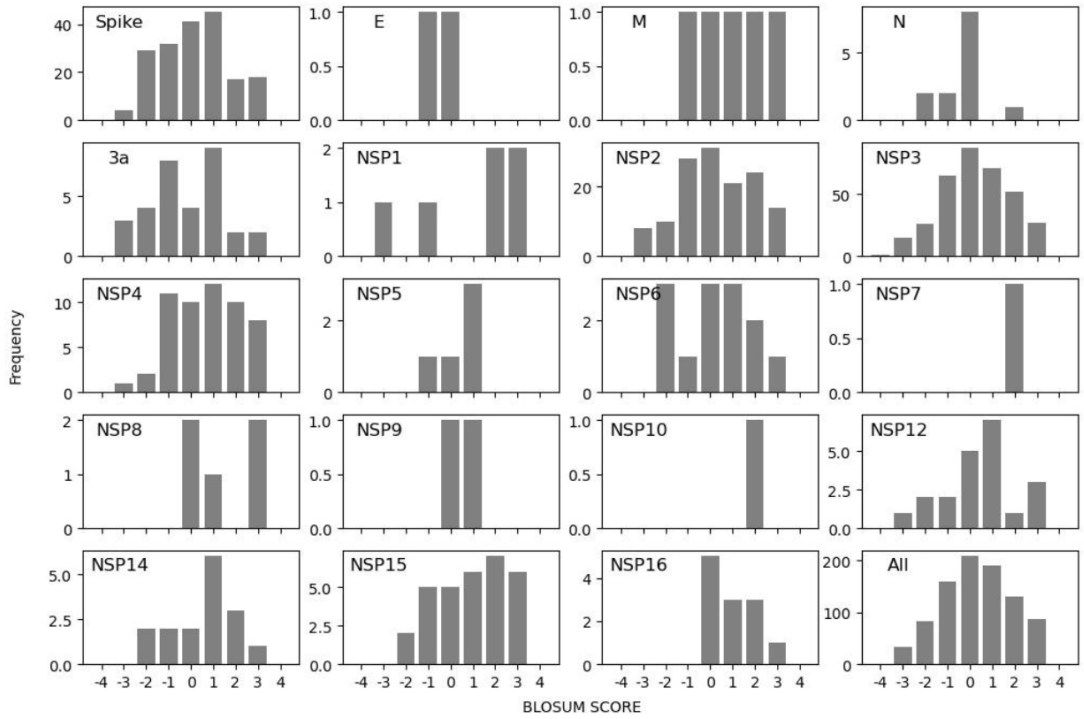
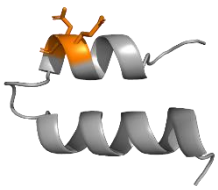
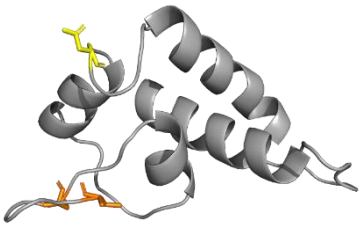


Figure S5.1. The BLOSUM scores for the amino acid substitutions present in the SDPs. A graph is plotted that combines all of the proteins and one for each of the individual proteins that were analysed.

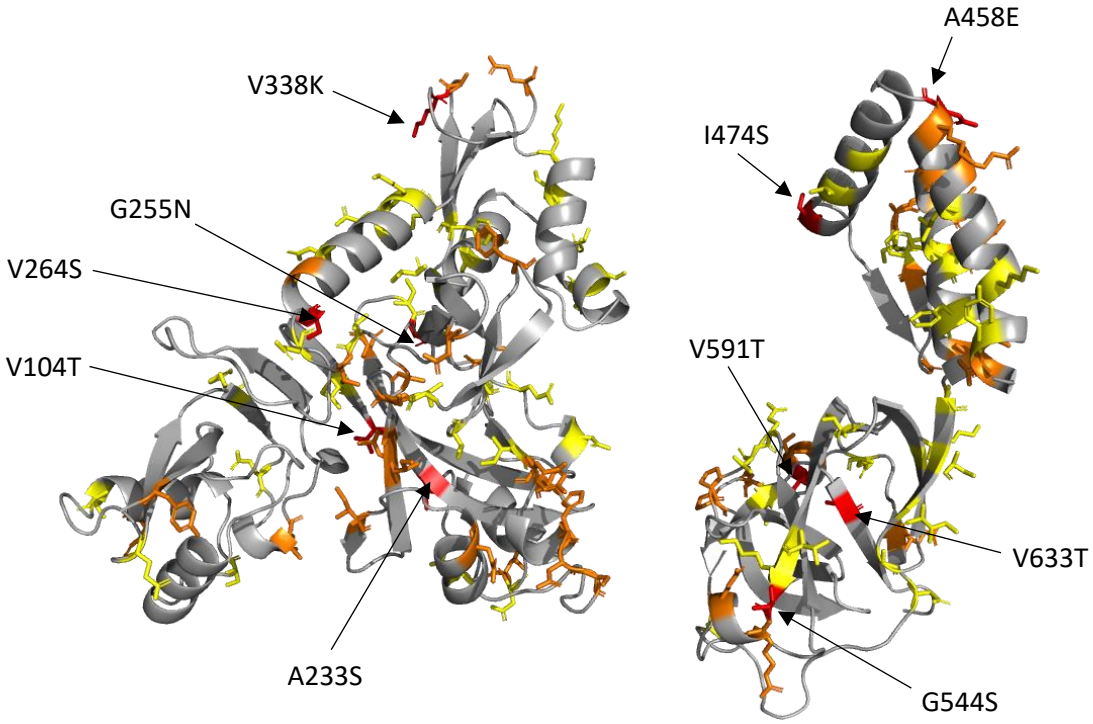
NSP1
(Phyre2 model d2gdta1)



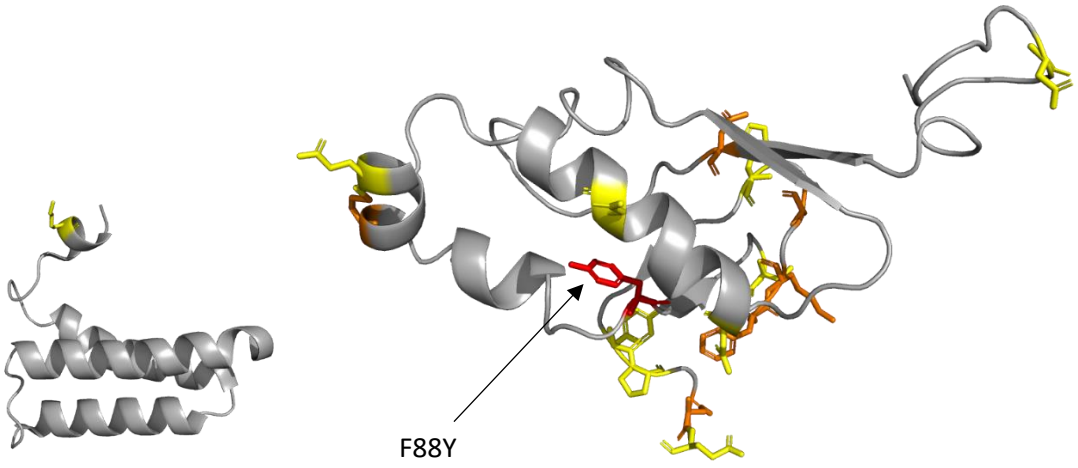
NSP1
(Phyre2 model c6zojj_)



NSP4
(Phyre2 model c3gzfD)

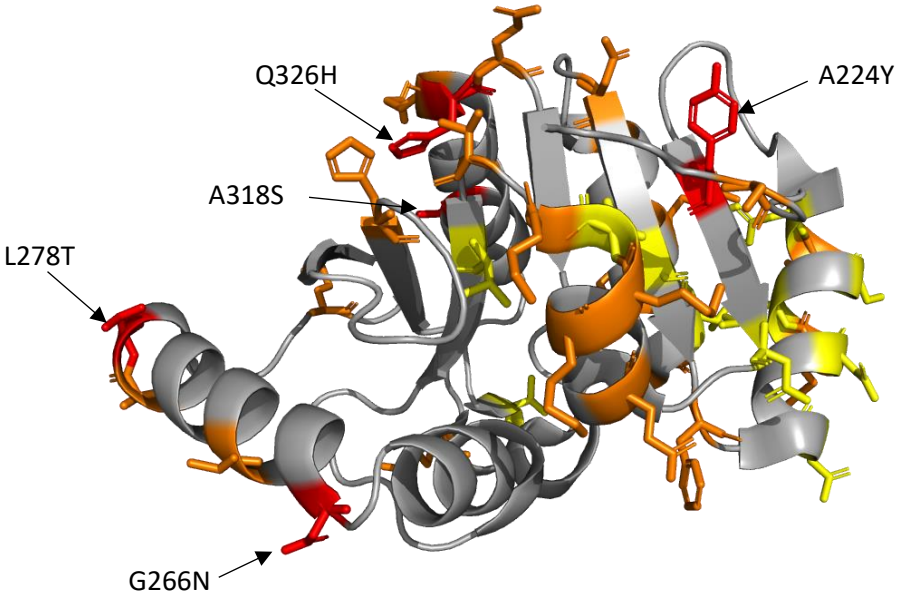


NSP2
(AlphaFold model)

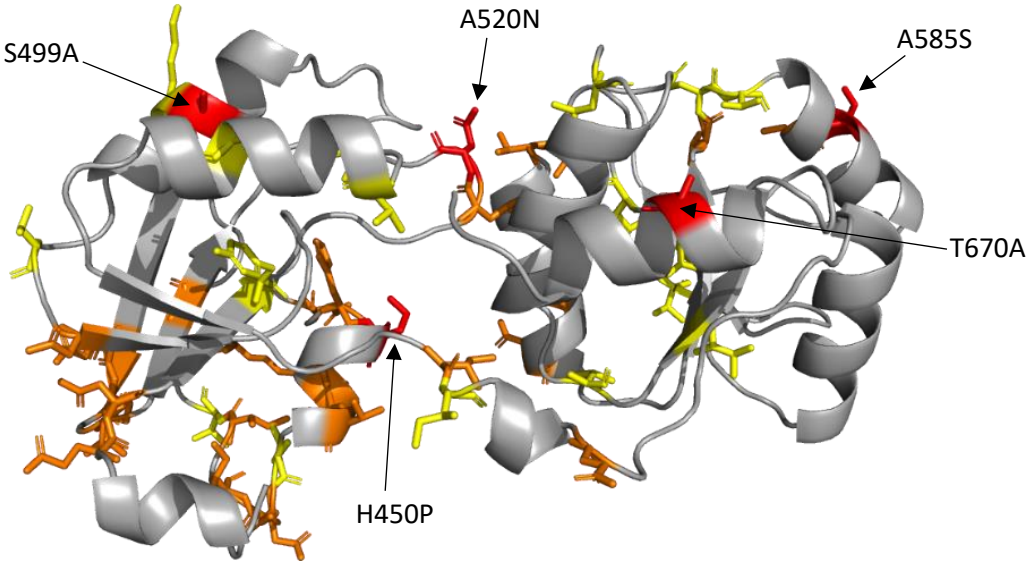


NSP7
(model 6xip)

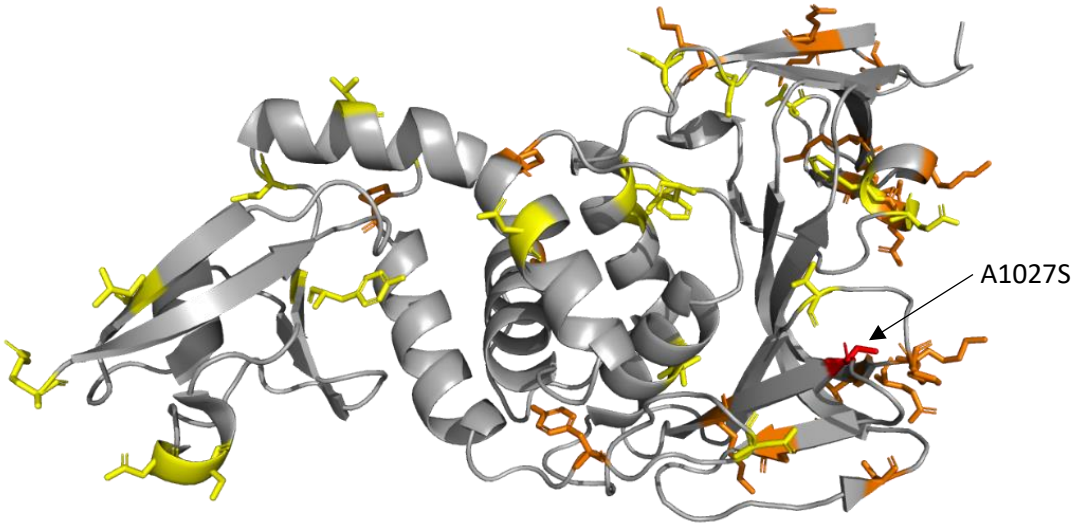
NSP3
(Phyre2 model d2gria1)



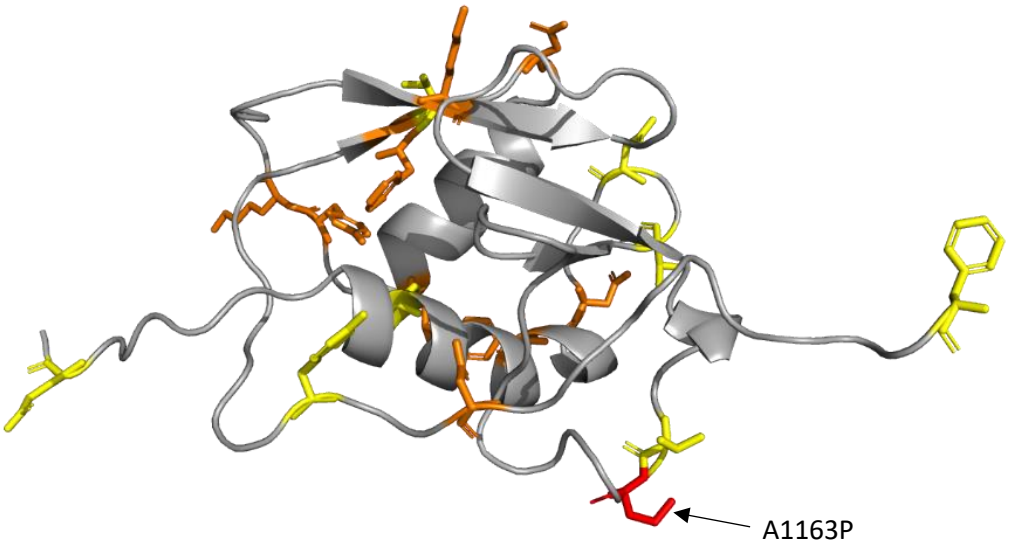
NSP3
(model 6vxs)



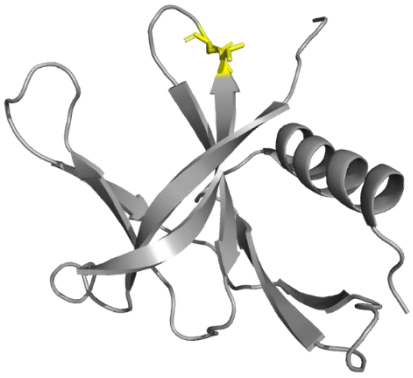
NSP3
(Phyre 2 model c2wctC)



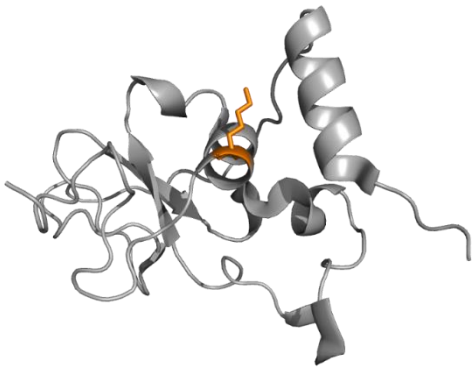
NSP3
(model 6w9c)



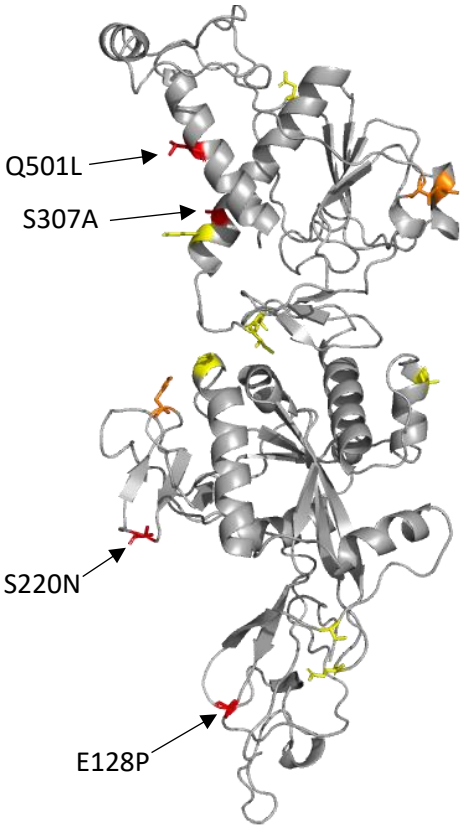
NSP3
(Phyre2 model c2k87A)



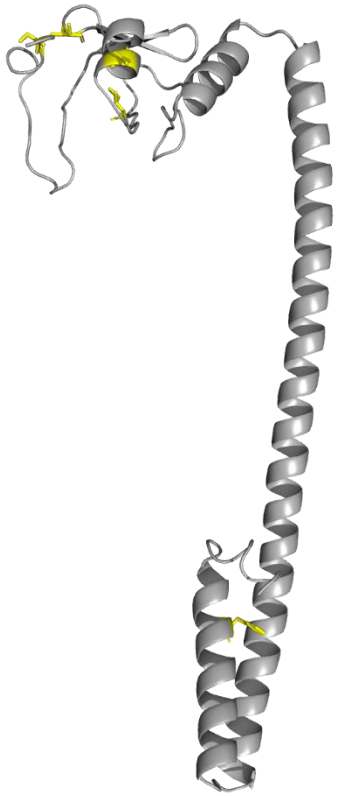
NSP9
(model 6w4b)



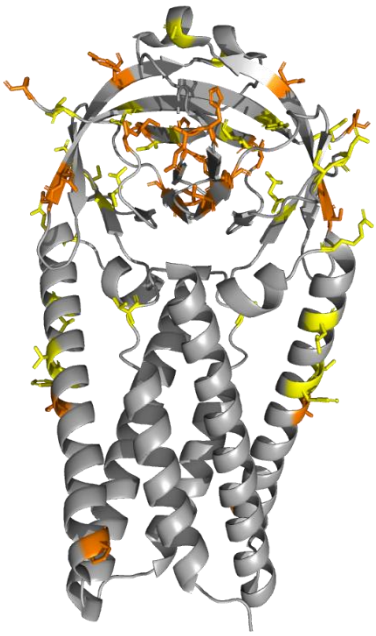
NSP10
(model 6w61)



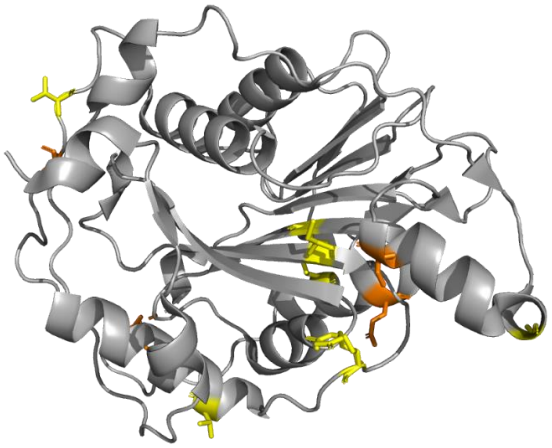
NSP14
(Phyre2 model c5c8sd)



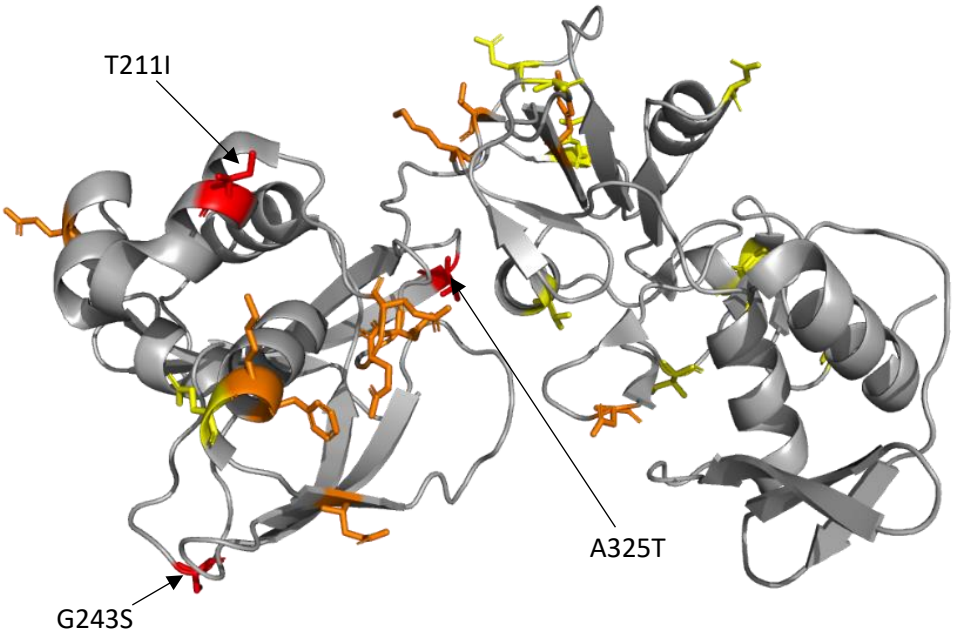
NSP8
(Phyre 2 model c2ahmG)



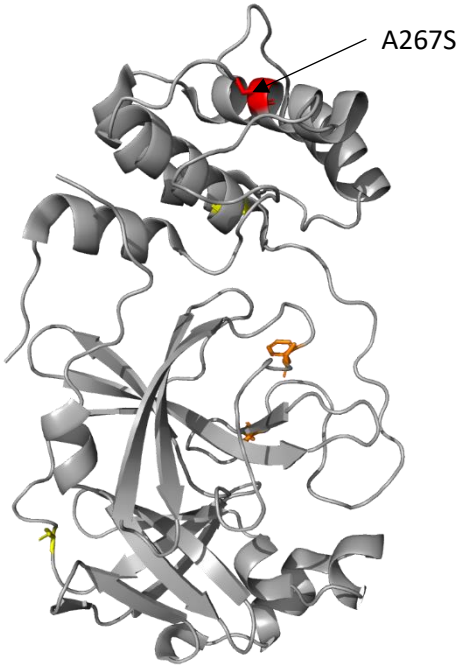
3a
(model 6xdc)



NSP16
(model 6w61)

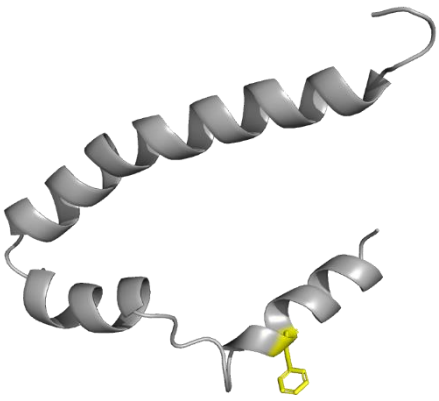


NSP15
(model 6vww)

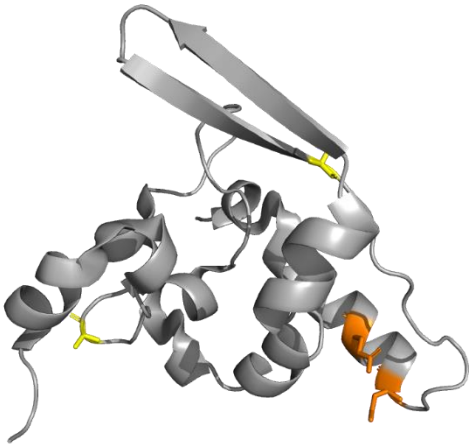


NSP5
(model 6y2e)

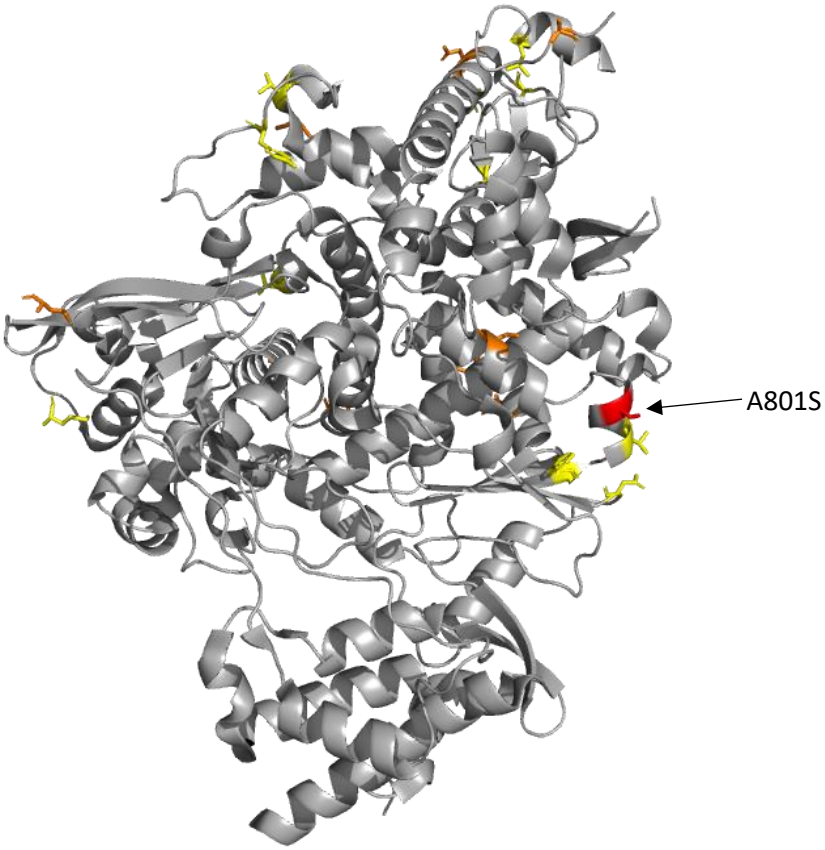
Nucleocapsid
(model 6m3m)



Envelope
(Phyre 2 model c5x29B)



Nucleocapsid
(model 6zco)



NSP12
(model 7bv2)

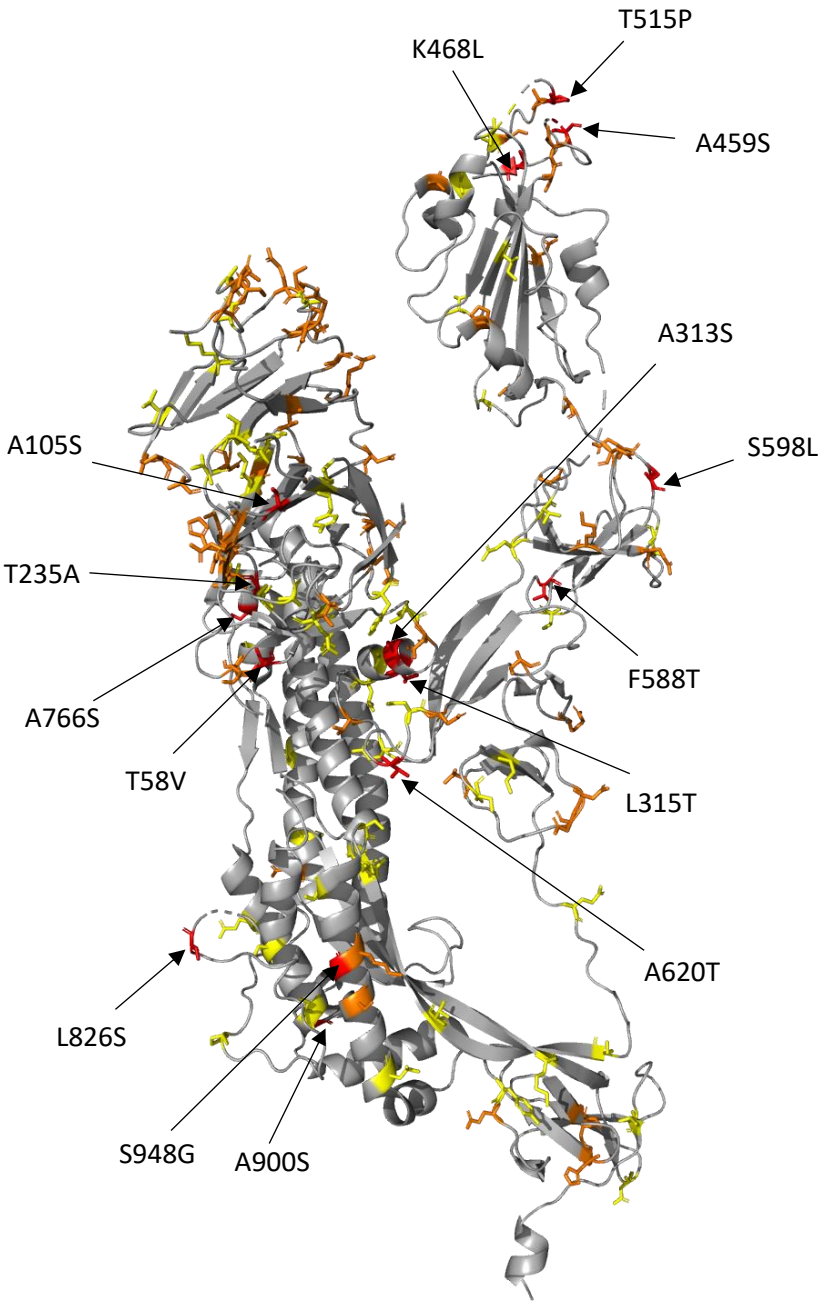


Figure S5.2. Overview of modelled DCPs. DCPs with likely functional effect are indicated by arrows and labelled. Structural model shown is indicated in brackets. DCPs likely to have an effect are coloured red; DCPs with a possible effect are shown in orange; and DCPs unlikely to have an effect are coloured yellow. Please refer to table S7 for full details of structural analysis of each DCP.

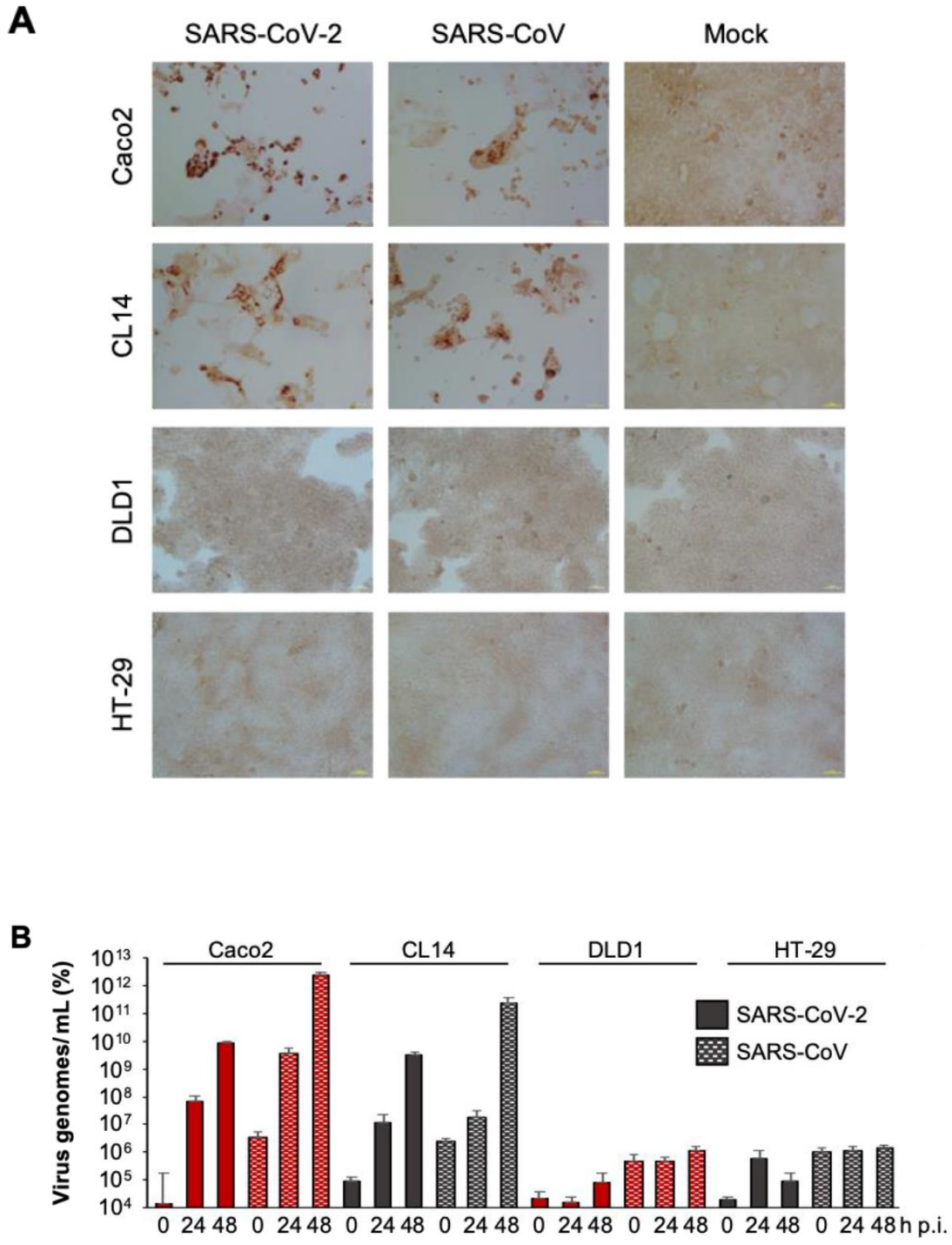


Figure S5.3. SARS-CoV-2 and SARS-CoV susceptibility of cell lines. (**A**) Representative images showing MOI 0.01-infected cells immunostained for double-stranded RNA 48h post infection. (**B**) Quantification of virus genomes by qPCR at different time points post infection (p.i.). Values are presented as means ± S.D. (n =3).

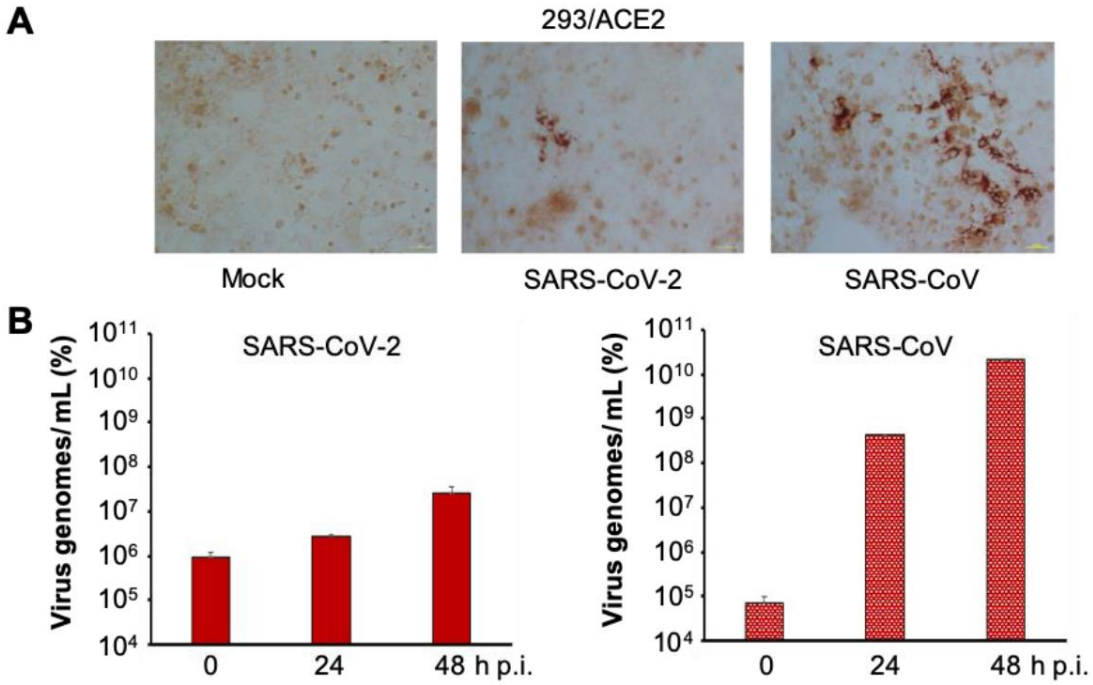
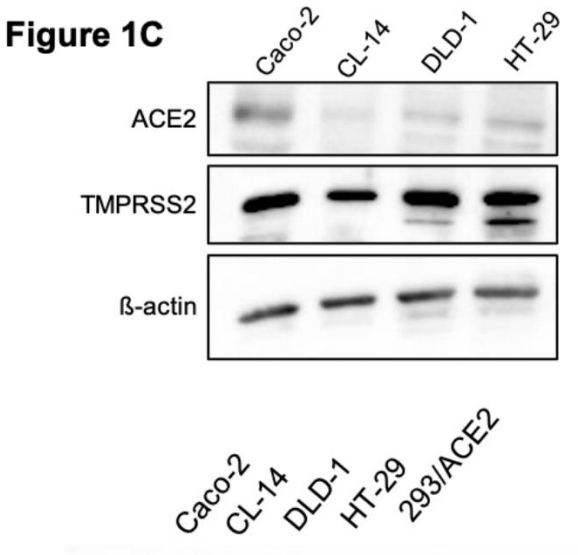


Figure S5.4. SARS-CoV-2 and SARS-CoV replication in 293 cells stably expressing ACE2 cells (293/ACE2). (**A**) Immunostaining for double-stranded RNA (indicating virus replication) in SARS-CoV-2 and SARS-CoV (MOI 0.01)-infected 293/ACE2 cells 48h post infection. (**B**) Quantification of virus genomes by qPCR in SARS-CoV-2 and SARS-CoV (MOI 0.01)-infected 293/ACE2 cells 48h post infection. Values are presented as means ± S.D. (n =3).



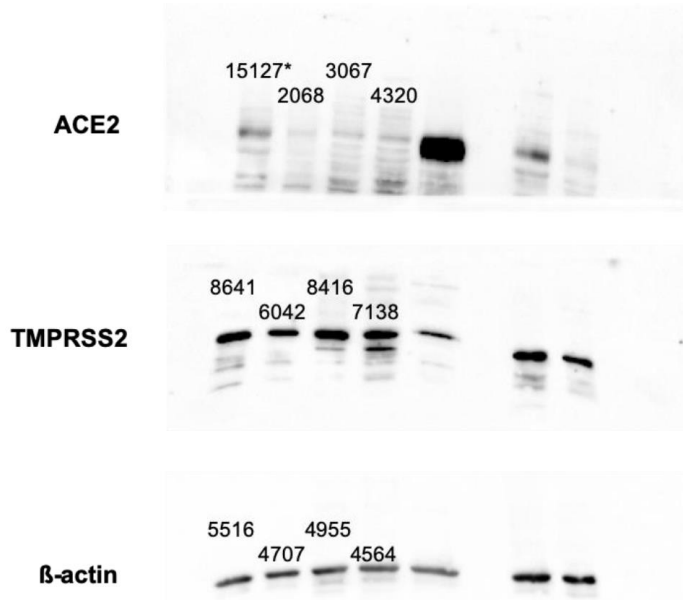


Figure S5.5. Uncropped Western blots for Figure 5.1C. 293/ACE2 cells served as positive control for ACE2. * Protein quantification

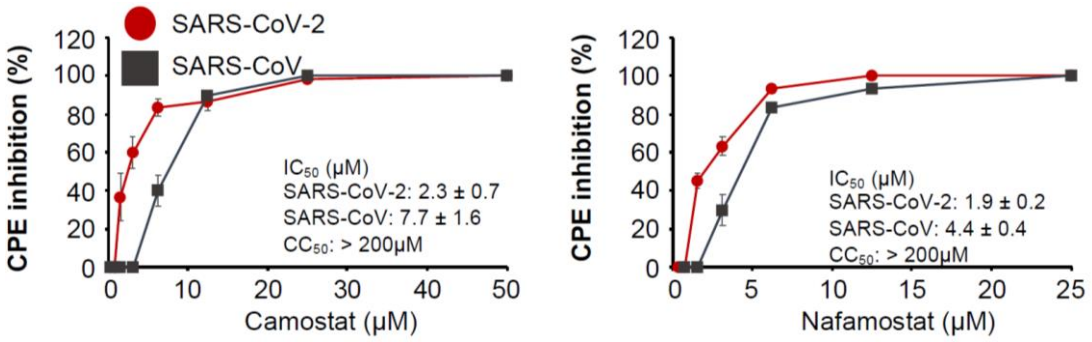


Figure S5.6. Role of TMPRSS2-mediated S cleavage in SARS-CoV-2 and SARS-Co-V replication. Concentration-dependent effects of the TMPRSS2 inhibitors camostat and nafamostat on SARS-CoV-2- and SARS-CoV-induced cytopathogenic effect (CPE) formation determined 48h post infection in CL14 cells infected at an MOI of 0.01. Values are presented as means ± S.D. (n =3).

Table S5.1. Protein structures used for structural analysis obtained from the Protein Data Bank.

SARS	PDB identifier	Protein	Residues
SARS-CoV	2hsx	NSP1	13-127
SARS-CoV	2gri	NSP3	1-111
SARS-CoV	2fav	NSP3	185-354
SARS-CoV-2	6vxs	NSP3	207-373
SARS-CoV	2w2g	NSP3	389-652
SARS-CoV	2kaf	NSP3	655-720
SARS-CoV	5y3e	NSP3	723-1036
SARS-CoV-2	6w9c	NSP3	748-1061
SARS-CoV	2k87	NSP3	1066-1180
SARS-CoV	2h2z	NSP5	1-306
SARS-CoV-2	6y2e	NSP5	1-306
SARS-CoV	6nur	NSP7	2-71
SARS-CoV-2	6xip	NSP7	1-70
SARS-CoV	2ahm	NSP8	1-190
SARS-CoV	2fyg	NSP10	10-132
SARS-CoV-2	6w61	NSP10	18-132
SARS-CoV	6nur	NSP12	41-819
SARS-CoV-2	7bv2	NSP12	31-929
SARS-CoV	5c8s	NSP14	1-525
SARS-CoV	2h85	NSP15	1-345
SARS-CoV-2	6vww	NSP15	1-346
SARS-CoV	2xyq	NSP16	1-290
SARS-CoV-2	6w61	NSP16	1-299
SARS-CoV	6acg	S:ACE2	18-1119
SARS-CoV-2	6m17	S:ACE2	336-518
SARS-CoV	5xlr	S	33-1120
SARS-CoV	5wrg	S	261-1058
SARS-CoV-2	6vsb	S	27-1146
SARS-CoV-2	6xdc	3a	40-238
SARS-CoV	5x29	E	8-65
SARS-CoV	1yo4	7a	16-99

SARS-CoV	1ssk	N	49-185
SARS-CoV-2	6m3m	N	48-173
SARS-CoV	2gib	N	270-366
SARS-CoV-2	6zco	N	248-364

Table S5.2. Structural models generated by Phyre2 and used for structural analysis. Where structures were not available from the Protein Data Bank, the structures were modelled.

SARS	**Template structure**	**Protein**	**Residues**	**Coverage**	**Confidence**	**Identity (%)**
SARS-CoV-2	2gdta1	NSP1	13-127	64	100	86
SARS-CoV	6zojj	NSP1	148-180	17	99.1	76
SARS-CoV-2	6zojj	NSP1	148-180	18	99.8	100
SARS-CoV-2	2gria1	NSP3	2-111	5	100	77
SARS-CoV-2	2acfa1	NSP3	207-373	8	100	74
SARS-CoV-2	2wctC	NSP3	425-676	12	100	76
SARS-CoV-2	2fe8B	NSP3	745-1058	16	100	82
SARS-CoV-2	2k87A	NSP3	1089-1203	5	100	82
SARS-CoV-2	3gzfD	NSP4	403-477	18	100	41
SARS-CoV-2	2duca1	NSP5	2-283	98	100	96
SARS-CoV-2	2ahmG	NSP8	1-175	95	100	97
SARS-CoV-2	1uw7A	NSP9	1-90	100	100	97

SARS-CoV-2	2g9tT	NSP10	9-116	86	100	98
SARS-CoV-2	6nusA	NSP12	118-909	87	100	97
SARS-CoV-2	5c8sD	NSP14	1-504	97	100	95
SARS-CoV	6xdcB	3a	40-238	72	100	77
SARS-CoV-2	5x29B	E	8-65	77	99.8	91
SARS-CoV-2	1yo4A	7a	16-98	67	100	91

Table S5.3. Criteria used for classifying proposed effect on protein structure and function within the structural analysis.

Effect	**Reason**
Unlikely	Conservative changes (between residues with the same polarity/charge) which do not affect ability to form hydrogen bonds with equivalent residues in SARS-CoV and SARS-CoV-2
Possible – conformational change	Changes which could affect the ability of a sidechain of a residue in a given position to form hydrogen bonds with equivalent residues in SARS-CoV and SARS-CoV-2 (e.g. gain/loss of polarity, substitution for larger/smaller sidechain) but no such effects are visible, or conservative changes (between residues with the same polarity/charge) which appear in the model to result in

	gain/loss of hydrogen bonding between equivalent residues in SARS-CoV and SARS-CoV-2 (but mutagenesis suggests hydrogen bonding is possible with sidechain rotation)
Possible – alteration of sidechain/ligand interactions	Changes which result in gain of charge/alter the charge of a sidechain for a residue in a given position
Possible – conformational change and alteration of sidechain/ligand interactions	Changes which affect the ability of a sidechain of a residue in a given position to form hydrogen bonds with equivalent residues in SARS-CoV and SARS-CoV-2 (e.g. gain/loss of polarity, substitution for larger/smaller sidechain) but no such effects are visible, and changes which result in gain of charge/alter the charge of a sidechain for a residue in a given position
Likely – conformational change	Changes which result in visible alteration in the conformation of a protein at a given location (e.g. through loss of hydrogen bonding between equivalent residues in SARS-CoV and SARS-CoV-2) and/or which result in the loss of capacity for hydrogen bonding
Likely – conformational change (and possible alteration of sidechain/ligand interactions)	Changes which result in visible alteration in the conformation of a protein at a given location (e.g.

	through loss of hydrogen bonding between equivalent residues in SARS-CoV and SARS-CoV-2, and/or which result in the loss of capacity for hydrogen bonding and which result in gain of charge/alter the charge of a sidechain for a residue in a given position)

Table S5.4. Specificity Determining Positions (DCPs) identified between SARS-CoV and SARS-CoV-2.

Protein (SARS-CoV)	**Protein (SARS-CoV-2)**	**Sequences in Dataset**	**Protein Length (SARS-CoV)**	**DCPs Identified**	**% of Residues DCPs**
S	S	73863	1255	186	14.82
3a	ORF3a	91214	274	32	11.68
3b		n/a	154		
E	E	94787	76	2	2.63
M	M	93860	221	15	2.26
6	6	94935	63	13	9.52
7a	7a	82940	122	0	0
7b	7b	n/a	44	NA	
8a/8b	8	n/a	39/84	NA	NA
9b		n/a	98	NA	
N	N	91609	422	13	3.08
	ORF10	n/a	n/a		
nsp1	nsp1	93621	180	6	3.33
nsp2	nsp2	88288	636	136	21.38
Nsp3	nsp3	75324	1922	344	17.90

nsp4	nsp4	89707	500	54	10.80
nsp5	nsp5	91731	306	5	1.63
nsp6	nsp6	93432	290	13	4.48
nsp7	nsp7	95038	83	1	1.20
nsp8	nsp8	94806	198	5	2.53
nsp9	nsp9	94970	113	2	1.77
nsp10	nsp10	92505	139	1	0.72
nsp12	nsp12	89874	932	21	2.25
nsp13	nsp13	91305	601	0	0
nsp14	nsp14	72306	527	16	3.04
nsp15	nsp15	85595	346	31	8.96
nsp16	nsp16	83565	298	12	4.03
Total				891	9.36

Table S5.5. Analysis of DCPs present in the SARS-CoV and SARS-CoV-2 Spike protein interface with human ACE2.

SDP	**SARS-CoV structural analysis**	**SARS-CoV-2 structural analysis**	**Effect?**
V404=K417	V404 is not in the interface	K417 is in the interface and could form a salt bridge with ACE-2 D30	Likely – new polar interaction within interface
R426=N439	Loss of hydrogen bond to ACE2 Gln325 due to shorter sidechain. N would still be able to form hydrogen bonds	N439 is located away from the interface site and so does not form a hydrogen bond with ACE2. Instead forms a hydrogen bond with S443 (also a DCP – A430=S443) which is likely to stabilise the	Likely – Loss of interface hydrogen bond.

		loop they are both part of.	
Y442=L455	Y422 forms hydrogen bond to backbone of W476 – loss could result in conformational change. The sidechain also contacts the backbone of ACE2 D30 and K31	L455 remains in interface and contacts ACE2 D30 and H34.	Likely – loss of intramolecular hydrogen bond
F460=Y473	Conservative change.	Introduction of OH group that can form hydrogen bonds. Y473 forms hydrogen bond with backbone of R457 and is closer to ACE2 T27 so potential to form hydrogen bond in interface.	Possible – introduction of hydrogen bond (could be with ACE2)
P462=A475	Located in a loop, could affect this conformation – many DCPs in this loop	Loop has different conformation.	Possible – Conformational change of loop
N479=Q493	Interface hydrogen bond formed with ACE2 H34 backbone. With a shorter sidechain this this may be lost in SARS-CoV-2.	Q493 forms a hydrogen bond with ACE2 E35 in this complex. So hydrogen bond is maintained but also different.	Possible – hydrogen bond with ACE2 retained but to different residue.
Y484=Q498	Y484 can form hydrogen bonds with ACE2 Gln42	Q498 maintains hydrogen bonds with ACE2 Gln42	Possible – change in residue forming

	(sidechain) and intramolecular H bonds with T433 (backbone), Y436 (sidechain).		hydrogen bonds with ACE2.

Supplementary Files:

Table S5.6. Structural analysis of DCPs in SARS-CoV and SARS-CoV-2

Link:

https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/bioinformatics/PAP/10.1093_bioinformatics_btab094/2/btab094_supplementary_data.zip?Expires=1625998633&Signature=yEUo7aFswjDlJ2J0BZuqT5wteb0dWc78MDQzrLGwtkWyWt6uaNzLl9Dg17DKRgDiHotgC8sN2I7h62rxtiA3RlPRE5aCr~qPU4QSmJABE-RLMe8r2myHS5XMsQbOb1VJtYWHUVDkfkNUp81R9lkGUQqKiD9sAgIYXurbyoJ~wJMk~IIFz8IyIJcgZ0SJiEeZUtMQRE8-pjm3d-xMix1bgj0dI9PKl8Iw8BugX4SGWjqmgoD9xeEUbTwZWvP32-fxYnxxi5fgMyuDWdcTEitPS6YAaYwf0q8kkI7XaSRXS41NZi0HhMRZiDMh2Aeo0eEADveEabsXUiSoJ5G~cLKhwg__&Key-Pair-Id=APKAIE5G5CRDK6RD3PGA

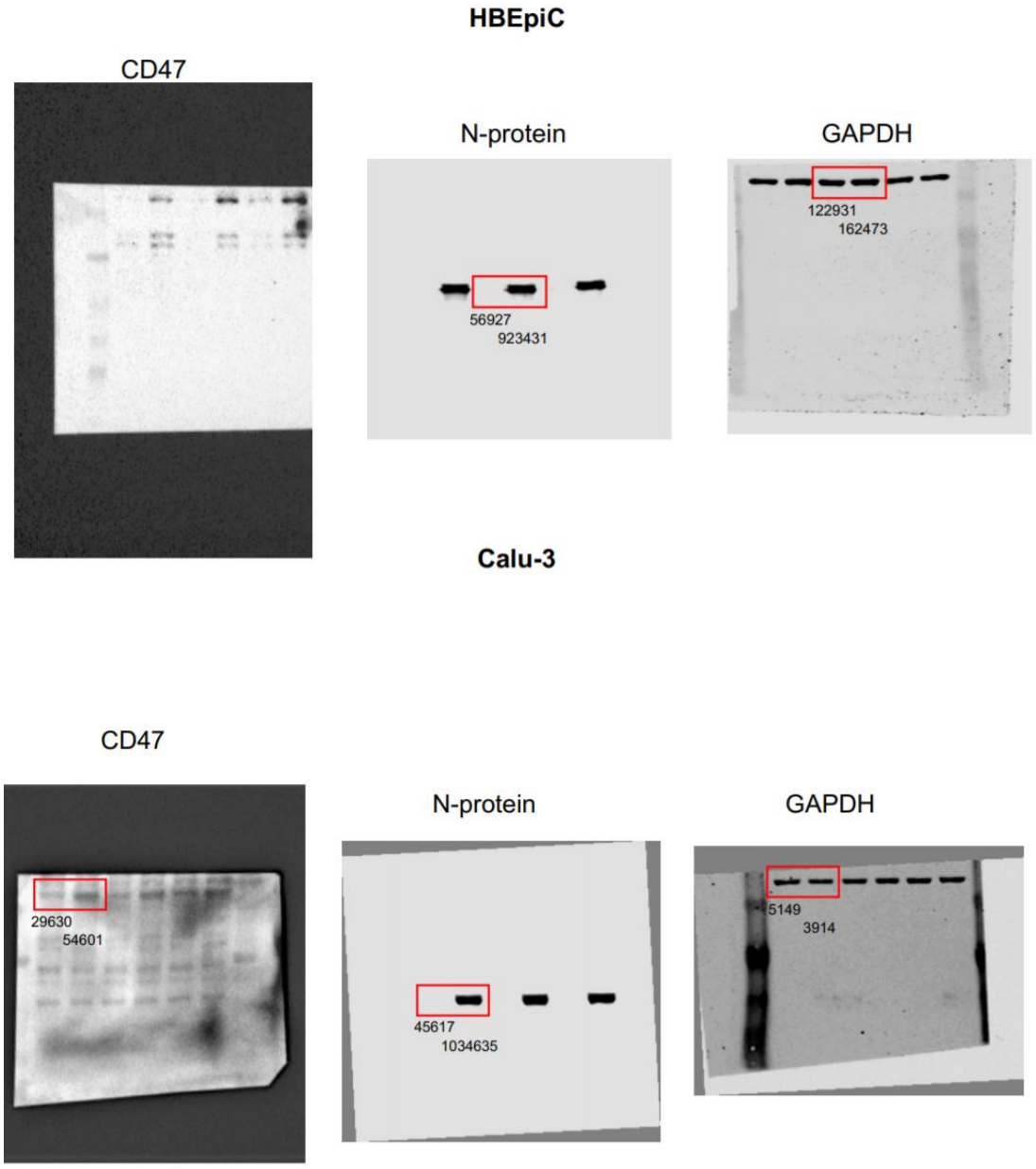


Figure S6.1. Uncropped Western blots to Figure 6.1. Bands are indicated by frames. Numbers indicate quantification results.

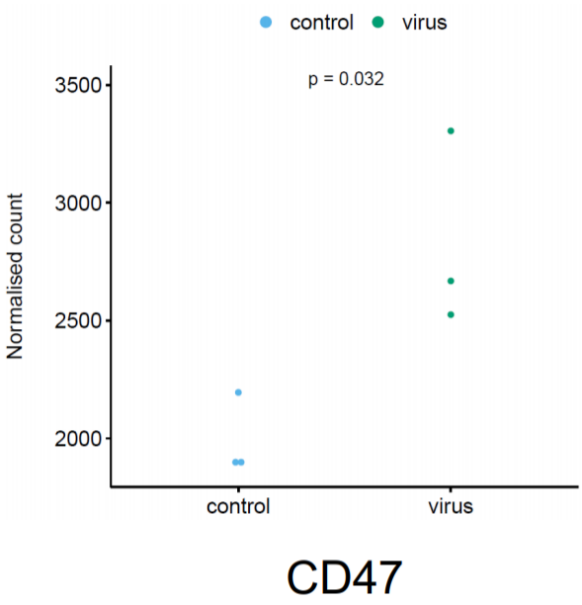


Figure S6.2. CD47 mRNA levels in SARS-CoV-2-infected Calu-3 cells (data derived from (Blanco-Melo et al., 2020)). P-values were determined by two-sided Student's t-test.

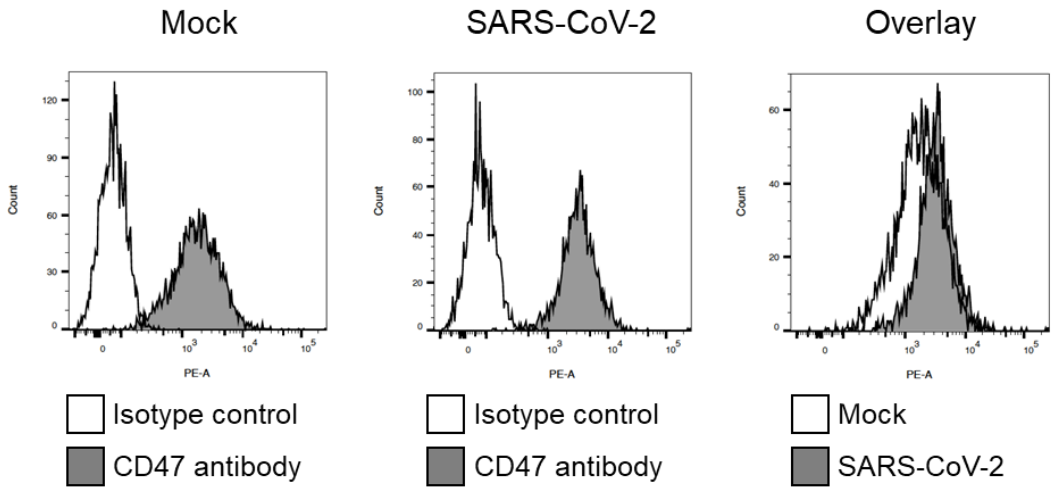


Figure S6.3. CD47 levels in SARS-CoV-2 (MOI 0.1)-infected Caco2 cells as determined by flow cytometry (FACSCanto II, BD Biosciences). Cells were fixed with 4% formaldehyde (10 minutes) and then stained for CD47 using a PE-labelled CD47 antibody (Miltenyi, # 130-123-754, 1:50 dilution). Isotype REA Control Antibody (S) (human IgG1, PE-labelled, Miltenyi, # 130-113-438, 1:50 dilution) was used as control.

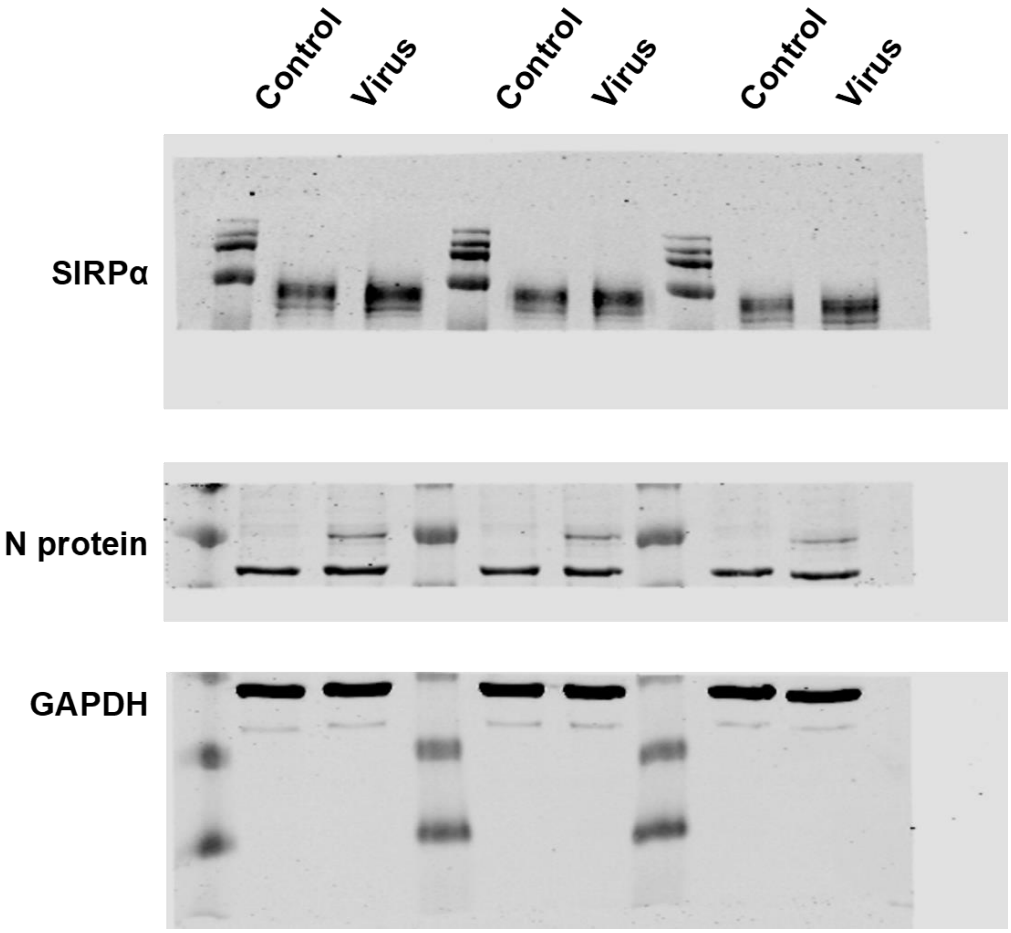


Figure S6.4. Uncropped Western blots to Figure 6.2.

Supplementary Files:

Table S6.1. Literature search for 'CD47 aging'. Literature search performed using PubMed (https://pubmed.ncbi.nlm.nih.gov) on 17th February 2021

Link:

https://www.biorxiv.org/content/biorxiv/early/2021/03/01/2021.03.01.433404/DC3/embed/media-3.xlsx?download=true

Table S6.2 Literature search for 'CD47 hypertension'. Literature search performed using PubMed (https://pubmed.ncbi.nlm.nih.gov) on 18th February 2021

Link:

https://www.biorxiv.org/content/biorxiv/early/2021/03/01/2021.03.01.433404/DC4/embed/media-4.xlsx?download=true

Table S6.3. Literature search for 'CD47 diabetes'. Literature search performed using PubMed (https://pubmed.ncbi.nlm.nih.gov) on 19th February 2021

Link:

https://www.biorxiv.org/content/biorxiv/early/2021/03/01/2021.03.01.433404/DC5/embed/media-5.xlsx?download=true

Table S6.4. Literature search for 'CD47 obesity'. Literature search performed using PubMed (https://pubmed.ncbi.nlm.nih.gov) on 22nd February 2021

Link:

https://www.biorxiv.org/content/biorxiv/early/2021/03/01/2021.03.01.433404/DC6/embed/media-6.xlsx?download=true

Supplementary Materials for Chapter 7

Supplementary Files:

Table S7.1. Genes associated with the GO term “Blood Coagulation” (GO:0007596) in males vs. females

Table S7.2. Correlation of the expression of coagulation-associated genes with age

Table S7.3. Genes with relevant functions whose expression levels correlate with age and differ between males and females

Table S7.4. Genes differentially regulated in SARS-CoV-2-infected Caco2 cells

Table S7.5. Interaction partners of SARS-CoV-2 proteins.

Link:

https://www.mdpi.com/2075-4418/10/8/539/s1

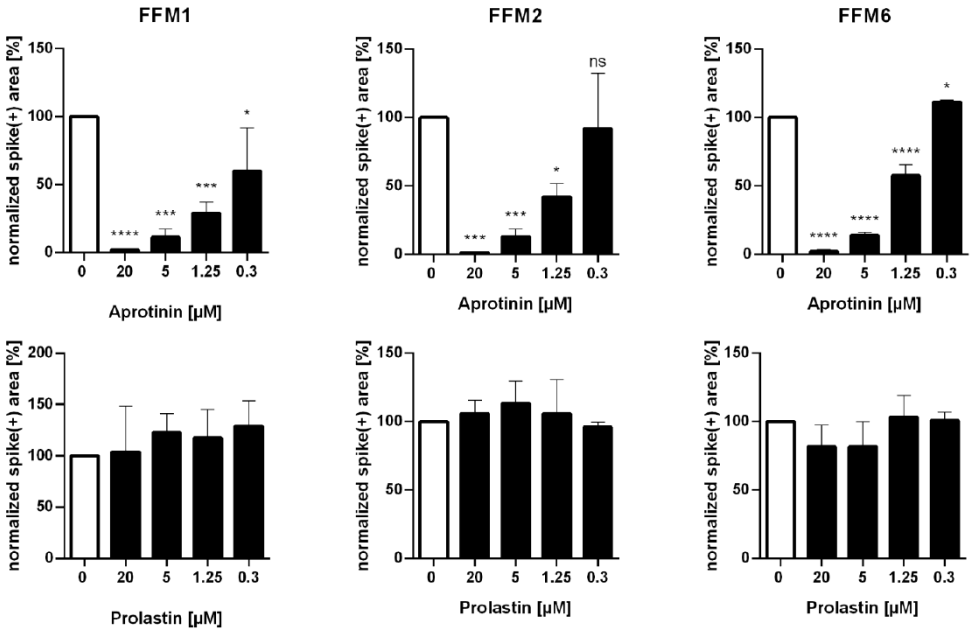
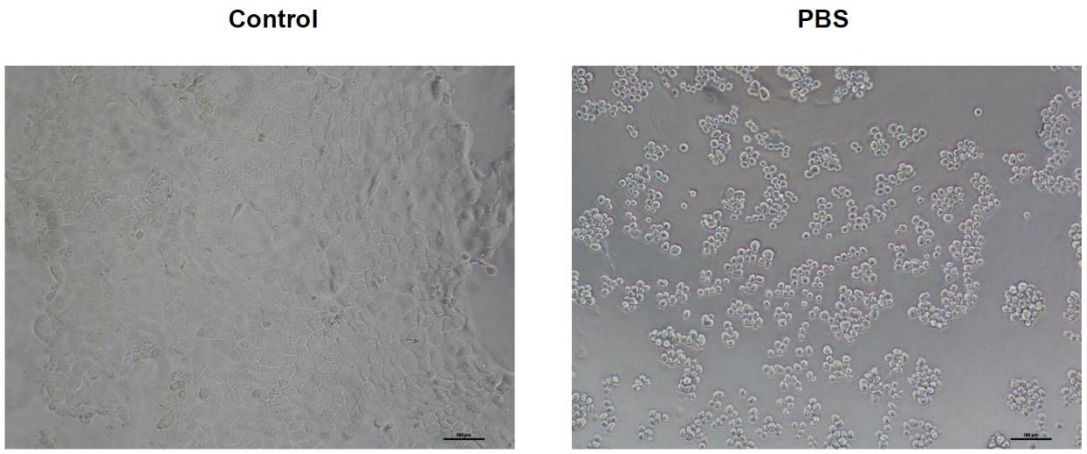


Figure S8.1. Quantification of immunostaining for the spike protein in SARS-CoV-2-infected (isolates FFM1, FFM2, FFM6) Caco2 cells with and without treatment of aprotinin or SERPINA1/ alpha-1 antitrypsin (prolastin) presented in Figure 8.2B.



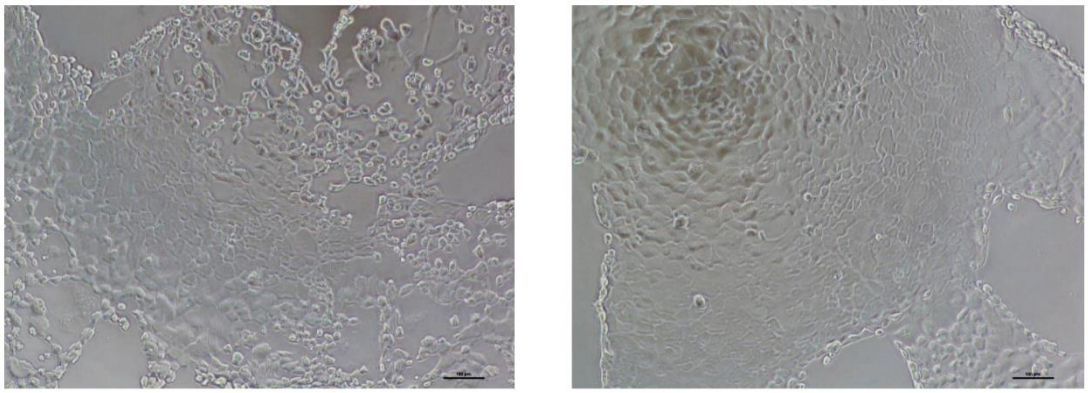


Figure S8.2. Trypsin inhibition by aprotinin and SERPINA1/ alpha-1 antitrypsin. Nearly confluent Caco2 cell cultures were washed three times with PBS and incubated with 400µg/ mL trypsin alone or in combination with aprotinin 20µM or SERPINA1 20µM for 2h.

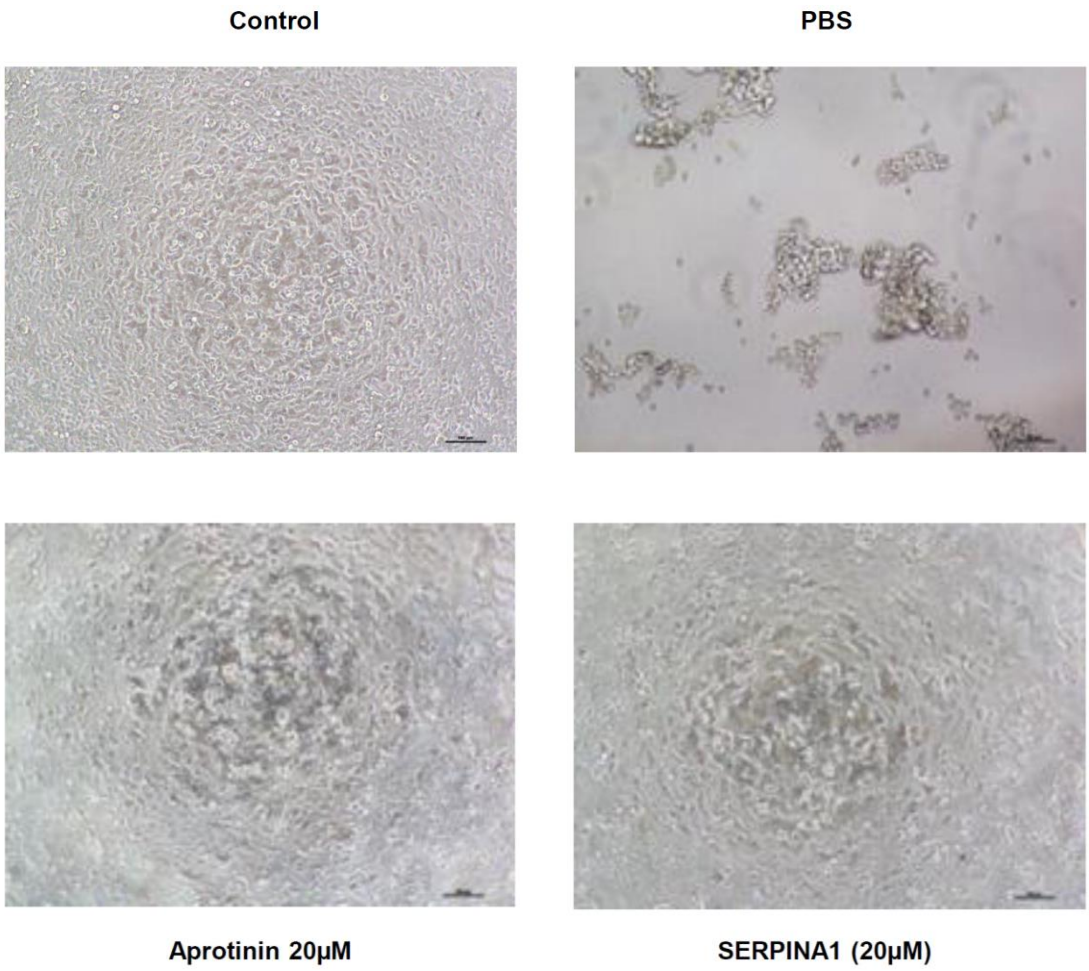


Figure S8.2. Trypsin inhibition by aprotinin and SERPINA1/ alpha-1 antitrypsin. Nearly confluent A549 cell cultures were washed three times with PBS and incubated with 400µg/ mL trypsin alone or in combination with aprotinin 20µM or SERPINA1 20µM for 2h.

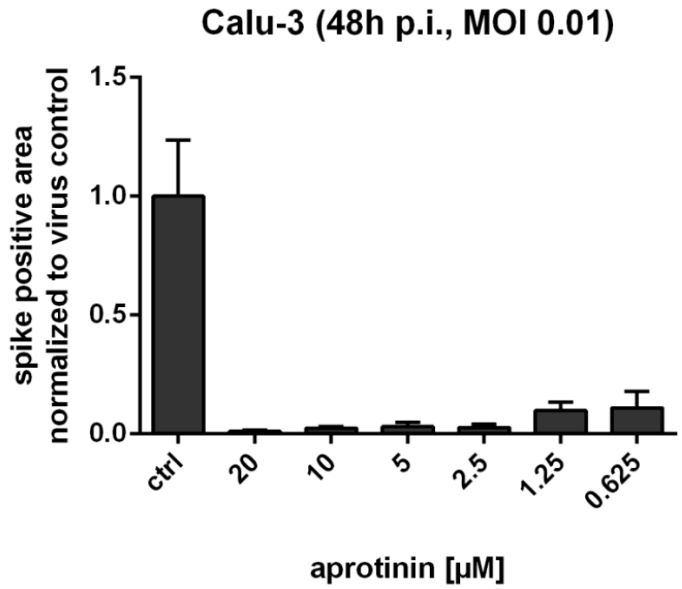
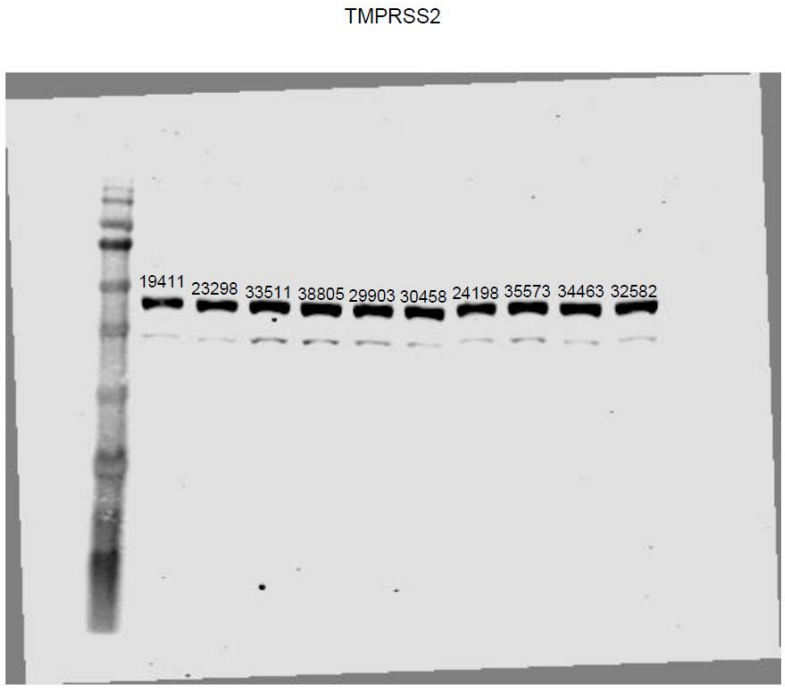
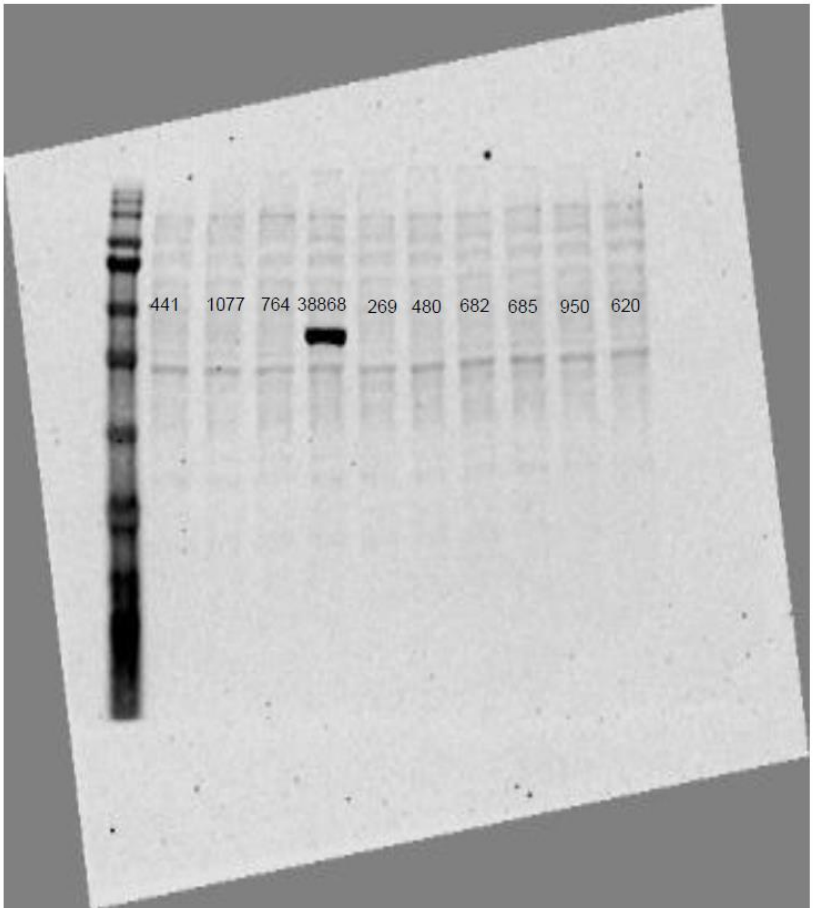


Figure S8.3. Quantification of immunostaining for the spike protein in SARS-CoV-2/FFM7 (MOI 0.01)-infected Calu-3 cells 48h post infection in response to aprotinin treatment.



N


441
1077
764
38868
269
480
682
685
950
620

GAPDH

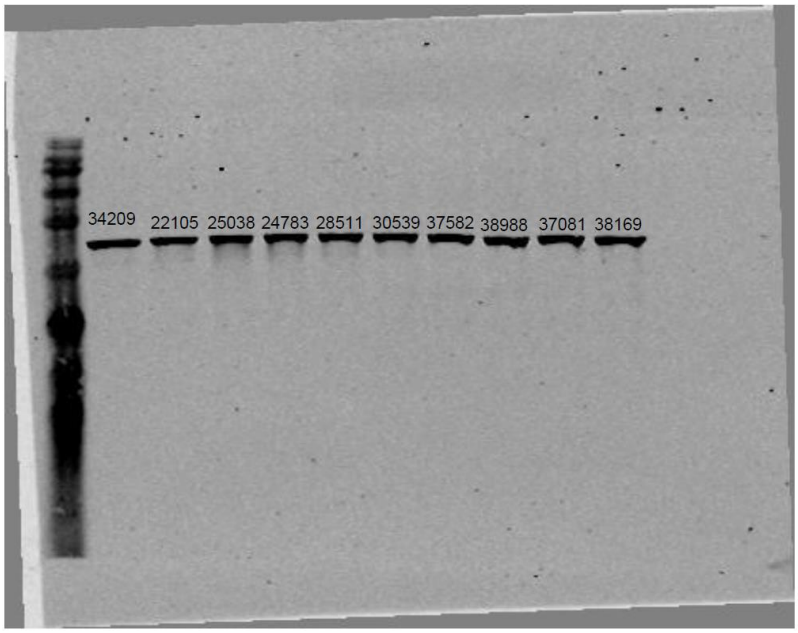

34209
22105
25038
24783
28511
30539
37582
38988
37081
38169

Figure S8.4. Uncropped Western blots corresponding to Figure 8.4B. Quantification was performed by laser-induced fluorescence using an infrared scanner (Odyssey, Li-Cor Biosciences) and Image Studio version 3.1 software.

Supplementary Files:

Table S8.1. Detailed mass spectrometry data

Link:

https://www.mdpi.com/2073-4409/9/11/2377/s1