

University of Kent

School of Biosciences

Investigating mechanisms of resistance to serdemetan in neuroblastoma

Eden Alexandra Laing

Submitted September 2022

Thesis for MSc by Research in Computational Biology

Abstract

Neuroblastoma is the most abundant extra-cranial tumor in pediatric patients. Though outlook for low-risk patients has significantly improved in the last decade, those classified as high-risk face uncertain outcomes. Drug-resistance is the principal issue, causing relapse and death for approximately 50% of high-risk patients. MDM2 inhibitors are under investigation as drug candidates for neuroblastoma. They induce p53 anticarcinogenic activity by disrupting the negative feedback interaction between MDM2 and p53, and are known to induce TP53 mutations as a resistance mechanism. Serdemetan, introduced as an MDM2 inhibitor, is one such drug where the likelihood of resistance development is high. Therefore it serves as a useful model for studying resistance formation. Investigation into how resistance develops is critical to circumnavigating it and improving high-risk patient survival rates. To address this, exome data from serdemetan-adapted sublines of the neuroblastoma cell line UKF-NB-3 were compared to those derived from the parental UKF-NB-3 cell line. Critical variants and pathways effected were identified and the degree of similarity between the sublines was established. Sublines containing TP53 null mutations were found to contain a high degree of similarity in variants acquired after serdemetan treatment, while those wild type for this gene were more similar. Three pathways were determined to be critically mutated, with alterations likely leading to resistance formation: 1) the cholesterol transport pathway, 2) the p53 signaling pathway, and 3) the xenobiotic clearance pathway. The involvement of multiple pathways indicates that there is not a one-size-fits-all mechanism of action through which resistance to serdemetan develops. However, mutations to the cholesterol transport pathway were deemed to be the most crucial to resistance formation, as the presence of these alone were enough to induce it. These findings implicate the cholesterol transport pathway in serdemetan's mechanism of action, confirming its' mechanisms exist beyond just modulation of MDM2 activity.

Word Count: 23366 Page Number: 109

Table of Contents

ABSTRACT	1
LIST OF FIGURES	3
LIST OF TABLES	4
ACKNOWLEDGEMENTS.....	5
ABBREVIATIONS	5
CHAPTER 1: LITERARY REVIEW.....	7
1.1 NEUROBLASTOMA STATISTICS AND BACKGROUND	7
1.2 GENES ASSOCIATED WITH NEUROBLASTOMA	8
1.3 TREATMENT OF NEUROBLASTOMA.....	10
1.4 RESISTANCE TO SYSTEMIC THERAPY	11
1.5 TARGETING MDM2 AND P53	13
1.6 SERDEMETAN BACKGROUND AND MECHANISMS OF ACTION	15
1.7 NEXT GENERATION SEQUENCING AS A SOLUTION.....	17
1.8 OVERVIEW AND AIMS OF THESIS	19
CHAPTER 2: METHODOLOGY	19
2.1 PIPELINE FOR CALLING AND FILTERING SINGLE NUCLEOTIDE VARIANTS.....	19
2.1.1 Variant calling	21
2.1.2 Variant filtering	24
2.1.3 Categorizing variants as pre-existing or acquired	26
2.2 COMPUTATIONAL ANALYSIS	28
2.2.1 Investigation of subpopulation selection.....	29
2.2.2 Gene ontology	30
2.2.3 Visualization and effect predication of mutations using PyMOL.....	31
CHAPTER 3: SUBLINE EVOLUTION LEADING TO RESISTANCE FORMATION.....	31
3.1 RESULTS	31
3.1.1 Comparing the number of variants between resistant sublines.....	32
3.1.2 Comparison of variants shared between all serdemetan-adapted cell lines.....	34
3.1.3 Comparison of variants shared between serdemetan-adapted cell line pairs	37
3.1.4 Variant allele frequency change: an indicator for subpopulation selection	41
3.2.5 Mutational profiles and signatures	43
3.1.6 Circular visualization of variants	45
CHAPTER 4: FUNCTIONAL ANALYSIS OF THE VARIANTS AND VARIANT PATHWAYS.....	46
4.1 RESULTS	46
4.1.1 Mechanisms of resistance to serdemetan	48
4.1.2 Functional categories affected as identified by KEGG Mapper	55
4.1.3 Hierarchical clustering of common pathways	62
4.1.4 Pathways affected as identified by KEGG Mapper	67
4.1.5 Investigations into shared variant consequence	76
CHAPTER 5: DISCUSSION	96
5.1 SUBLINE EVOLUTION LEADING TO RESISTANCE FORMATION	96
5.2 FUNCTIONAL ANALYSIS OF THE VARIANTS AND VARIANT PATHWAYS	97
5.3 SUMMARY OF MAIN FINDINGS AND CONCLUDING REMARKS	98
5.2 FUTURE STUDIES	99
BIBLIOGRAPHY	100

List of Figures

Figure 1: MYCN pathway overview in neuroblastoma

Figure 2: Mechanisms of intrinsic and acquired resistance in cancerous tumours

Figure 3: The mechanism of action of serdemetan within the MYCN pathway

Figure 4: Potential off target effects of serdemetan

Figure 5: Subline selection protocol

Figure 6: Variant calling pipeline

Figure 7: Variant filtering steps and specifications

Figure 8: Workflow for classification of variants

Figure 9: Categorization of variants

Figure 10: Scatterplot of the number of variants per classification type

Figure 11: Comparison of shared variants between all serdemetan-resistant UFK-NB-3 sublines

Figure 12: Comparison of shared variants between pairs of serdemetan-resistant UFK-NB-3 sublines

Figure 13: Shift in variant allele frequency between UFK-NB-3 and UFK-NB-3 serdemetan-adapted cell lines

Figure 14: Mutational profile heatmap

Figure 15: Circular visualization of variants

Figure 16: The p53 signalling pathway

Figure 17: The Mdm2-HIF1 α Axis

Figure 18: The cholesterol metabolism pathway

Figure 19: Functional categories affected by serdemetan resistance as identified by KEGG MAPPER

Figure 20: Hierarchical clustering of serdemetan resistance UKF-NB-3 sublines shared variants mapped to KEGG Pathways

Figure 21: Pathways affected by serdemetan resistance as identified by KEGG MAPPER

Figure 22: PyMOL renderings of APOB missense variants, S to Y, before and after mutation

Figure 23: PyMOL renderings of APOB cysteine residue at position 88 prior to mutation

Figure 24: PyMOL renderings of ABCA10 missense variants, V to L, before and after mutation

Figure 25: PyMOL renderings of TP53 missense variants, C176S, before and after mutation

Figure 26: PyMOL renderings of TP53 missense variants, C44S and C83S, before and after mutation

Figure 27: PyMOL renderings of the TP53 missense variants, Y to C, in subline V before and after mutation

Figure 28: PyMOL renderings of PLA2G4F arginine residue at position 367 prior to mutation

Figure 29: PyMOL renderings of AHR missense variants, T to M, before and after mutation

Figure 30: PyMOL renderings of UGT2B7 missense variants, D to H, before and after mutation as well as renderings of UGT2B7 leucine residue at position 201 prior to stop codon mutation

Figure 31: PyMOL renderings of UGT2B4 missense variants, P to S, before and after mutation

List of Tables

Table 1. UKF-NB-3 subline similarities and differences

Table 2. Results from the variant effect predictor MTBAN for the S162Y variant in APOB

Table 3. Results from the variant effect predictor MTBAN for the V1226L variant in APOB

Table 4. Results from the variant effect predictor MTBAN for the variants C176S and Y205C in p53

Table 5. Results from the variant effect predictor MTBAN for the variant T578M in AHR

Table 6. Results from the variant effect predictor MTBAN for the variant D99H in UGT2B7

Table 7. Results from the variant effect predictor MTBAN for the variant P69S in UGT2B4

Acknowledgements

I would like to thank my supervisors, Dr. Mark Wass and Professor Martin Michaelis, for their invaluable guidance during the completion of this project. I could not have done this without their support, insight, and advice. I would also like to thank Magdalena Antczak for the preliminary work she did to construct the computational pipeline.

Thank you to my lab group members for always being open to tackling any queries we've had together and the comradery that kept spirits high.

I would like to especially thank my parents. This would have been impossible to complete without your emotional and financial support. Thank you for always loving and believing in me. I am eternally grateful and so blessed to have you on my team.

Abbreviations

AA	Amino Acid
ABCA1	ATP-binding Cassette Subfamily A Member-1
ABCA10	ATP-binding cassette subfamily A member 10
ABCD2	ATP-binding cassette subfamily D member 2
ABCG1	ATP-binding Cassette Subfamily G Member-1
AHR	Aryl Hydrocarbon Receptor
ALK	Anaplastic Lymphoma Kinase
APOB	Apolipoprotein B

APOB-100	Apolipoprotein B100
ARG	Arginine
ASP	Aspartic Acid
BAM	Binary Alignment Map
BCF	Binary Variant Call Format
CDS	Protein Coding Sequence
COSMIC	Catalogue of Somatic Mutations in Cancer
CYS	Cysteine
GATK	Genome Analysis Toolkit
gnomAD	Genome Aggregation Database
GRCh37	Genome Reference Consortium Human Reference Genome 37
HDL-c	High-Density Lipoprotein
HIF1 α	Hypoxia Inducible Factor 1 Alpha
HIS	Histidine
INDEL	Insertions and Deletions
LEU	Leucine
LDL-c	Low-Density Lipoprotein
MDM2	Murine Double Minute 2
MET	Methionine
MTBAN	Mutation Effect Predictor using the Temporal Convolutional Network and the Born-Again Networks
NGS	Next Generation Sequencing
PLA2G4F	Phospholipase A2 Group IVF
PRO	Proline
RCCL	Resistant Cancer Cell Line Collection

SAM	Sequence Alignment Map
SER	Serine
SNV	Single Nucleotide Variants
SR-BI	Scavenger Receptor Class B Member 1
TCGA	The Cancer Genome Atlas
THR	Threonine
TYR	Tyrosine
UGT	UDP-glucuronosyltransferase
VAF	Variant Allele Frequency
VAL	Valine
VCF	Variant Calling Format
VEP	Ensembl Variant Effect Predictor
VLDL	Very Low Density Lipoproteins
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing

Chapter 1: Literary Review

1.1 Neuroblastoma statistics and background

Neuroblastomas are malignant tumours of the sympathetic nervous system (1,2). They present as an oncologic dilemma within the sphere of paediatric cancer biology, with 90% of patients diagnosed before the age of five (3,4). As the most abundant paediatric extra-cranial solid tumour (8-10%), they are a leading cause of cancer-related deaths in children aged one to four years old (3,5–7). Neuroblastoma tumours are responsible for 15% of cancer-related deaths in paediatric patients (3,4,7–9). Manifesting within a broad clinical range, treatment and prognosis is correlated to patient age and degree of cell differentiation (1,3,5,7,9,10).

Prognosis predictions are formulated according to the International Neuroblastoma Pathology Classification, using clinical parameters INRG imaging stage, age, DNA ploidy/MYCN amplification and stage based on pathological/histological features (4,5,7). Based on these criteria, patients are split into the pre-treatment groups very low-, low-, intermediate-, high- and ultra-high-risk groups (1–3). Approximately 40% of patients present with high-risk neuroblastoma which is difficult to cure (4), including children over the age of 18 months with metastatic disease and those with tumours harbouring MYCN amplification (1,6). Infants below the age of 18 months tend to have disease that spontaneously regresses or is treatable with surgery alone - whereas older patients are at higher risk for metastasized, treatment resistant and aggressive neuroblastomas (4). Where disease has metastasized, secondary malignant growths are primarily located in the bone marrow, bone, lymph nodes, liver or subcutaneous tissue (5). Outcomes from treatment vary greatly, from low-risk patients displaying tumour regression to high-risk patients where the likelihood of reoccurrence and mortality remain high (between 40-50% survival rate) (3,4,7,9). Current research is focused on further improving survival rates for this risk group, with treatment options being revised and improved (7). Genes commonly associated with the development of neuroblastoma, with a focus on high-risk patients, will be discussed in the next section.

1.2 Genes associated with neuroblastoma

The most frequently observed mutations found in neuroblastoma are: MYCN oncogene amplification, 1p and 11p deletion, 17q gain, tyrosine kinase domain alteration of anaplastic lymphoma kinase (ALK), PTPN11, ATRX, proximal alterations of the TERT gene, chromatin remodelling genes, neuritogenesis-specific genes and NRAS-MAPK pathway genes (2,7,11). For more than 30 years, studies have pointed to the amplification of oncogene MYCN, a member of the MYC proto-oncogene family, as one of the principle biological markers observed in neuroblastoma (1,5,7,12). Present in 20-25% of cases, the genomic

status of this gene has been included as part of patient risk stratification and clinical decision making (5). MYCN functions as a transcription factor and is able to trans activate MDM2, one of the downstream targets of MYCN, preventing the tumour suppressor activity of p53 (3). An overview of the MYCN pathway as seen in neuroblastoma is shown in Figure 1. The overexpression of MYCN in neuroblastoma has been linked to rapid tumour progression and particularly poor patient prognosis and is often observed in high-risk neuroblastoma patients – the focus group for this study (1,5). Furthermore, MYCN overexpression has been found to sensitise neuroblastoma cells to MDM2 inhibitors (1). Another area of interest is the anaplastic lymphoma kinase (ALK), a receptor tyrosine kinase whose gene is commonly mutated in genetically predisposed patients (1). ALK mutations have been found to occur in approximately 10% of high-risk neuroblastoma cases (7). It has also been confirmed as a target of the transcription factor MYCN – linking it closely to poor patient outcomes (1). Previous studies examining the ALK gene have found that the frequency of these mutations increase when examining relapsed patients – indicating that ALK may play a role in resistance development (5). PHOX2B has also been identified as an inheritable gene responsible for many familial neuroblastoma cases (2). Unlike in most other cancers, the tumour suppressor gene - p53 - is rarely mutated in neuroblastoma tumours at the time of diagnosis (2). Treatment of neuroblastoma depends largely on the types of mutations and degree of MYCN amplification found in patient tumours (1).

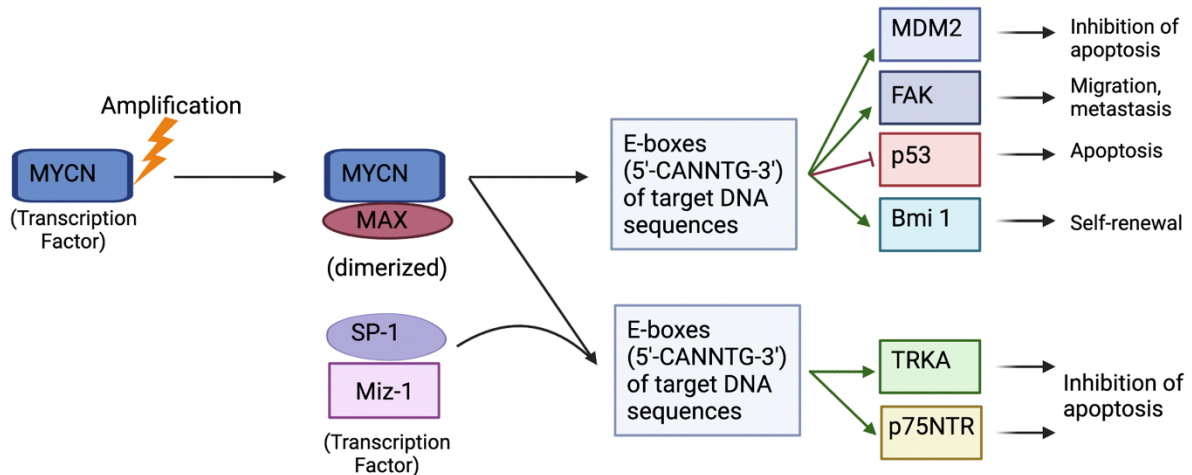


Figure 1 MYCN pathway overview in neuroblastoma

MYCN is often amplified in cancer, leading to overexpression of the transcription factor MYCN. This transcription factor dimerizes with Max to bind target DNA sequences, or E-boxes (with the sequence 5'-CANNTG-3'), to regulate transcription of genes involved in cell growth and proliferation (12–14). Created with BioRender.com.

1.3 Treatment of Neuroblastoma

The existing standard treatment of neuroblastomas involves multimodal chemotherapy, radiation, immunotherapy and surgery (3,5). These treatments are split into three stages – induction, consolidation, and post-consolidation/maintenance therapy – lasting approximately 18 months (7). During the induction stage of treatment, patients receive 5-8 rounds of intensive chemotherapy (7). The goal of multimodal chemotherapy is to attain patient remission by decreasing the tumour burden and enabling resection (3). Surgery is only recommended after adjuvant therapies near the end of the induction phase, to reduced risk of incomplete resection and surgical morbidity, however it is limited to open biopsy in high-risk cases (3,7). The consolidation stage involves high-dose chemotherapy and radiation therapy with the aim of eliminating any residual disease (7). The final stage, post-consolidation and maintenance, has been implemented as an attempt to prevent relapse from occurring and eradicate all disease if the previous stages failed to (7). The inclusion of this stage has contributed to improved survival rates (7). Over the past 30 years there has been sizable improvements to patient survival rates, with an increase in 5-year survival rates from 52% to

74% (5). This is largely due to the inroads made improving cure rates in the low-risk patient group (3). However there is mounting evidence that these methods are unsuccessful when presented with high-risk aggressive and drug resistance neuroblastomas (3). Where the degree of cell differentiation is high and systemic treatment needed, prevalence of relapse is common (approximately 50-60%) with only modest improvements to mortality rates in the last 30 years (1,3). With such little improvement observed for this patient group, investigations into new treatment approaches to decrease relapse and prevent disease progression are necessary (4,5). Investigations should be focused on preventing drug resistance development, and patient relapse, during systemic therapy.

1.4 Resistance to systemic therapy

Resistance development to systemic therapies after initial treatment proves to be the biggest obstacle to improved prognosis in patients with advanced disease and represents an unmet clinical need (10,15). The development of drug resistance can lead to the growth of existing tumours, further invasion of neighbouring tissues or eventual metastasis to remote tissues, resulting in patient death (16). Many patients in the high-risk category who relapse due to residual resistant disease eventually die from disease progression (3). Relapse and a lack of response to treatment can result from two distinct types of resistance – acquired or intrinsic (15,17). These mechanisms of resistance differ conceptually – and understanding and accounting for each is critical to preventing treatment failure (15). Acquired resistance develops as a result of directed evolution due to pressures exerted from systemic therapies (15). In these cases, patients initially respond well to treatment, however, once developing drug resistant cancer tumour cells therapy fails (15). This resistance can develop as a result of genetic or epigenetic aberrations resulting from treatment or from the expansion of small subpopulations of tumour cells not eliminated during treatment that enable regrowth of a drug resistant tumour (17,18). The latter is thought to occur due to the great degree of molecular

cell heterogeneity found in cancerous tumours (19). On the other hand, intrinsic resistance exists as a consequence of pre-existent resistance of patient tumour cells arising prior to treatment that render cancer cells unresponsive to it (15). This means that intrinsically resistance cells will not respond to treatment from the outset (15). The mechanistic differences of these two kinds of drug resistance in cancers are outlaid in Figure 2. Previous studies have found that the overall count of mutations in neuroblastoma biopsies increases from diagnosis to the time of relapse (1). It is plausible that these mutations could develop as a form of acquired drug resistance, whereby the drug administered could increase the underlying mutation rate of the tumour cells or apply selective pressure for drug-resistant subpopulations of cells. An increased mutation rate would provide increased selection opportunity, with selective pressures ultimately allowing for resistance to emerge during the treatment period. Identification of resistance biomarkers and development of new therapy targets is necessary to combat both intrinsic and acquired drug resistance – where established treatment options are no longer viable (15,20). High throughput analysis of drug resistance neuroblastoma cell lines could help with the identification of such biomarkers (15). The study of acquired resistance has proved difficult in the past due to most cancer cell lines being developed from tumors at the time of patient diagnosis (15). Prior to study, the treatment of a drug susceptible neuroblastoma cell line until resistance develops could provide a means to examine acquired resistance mechanisms not previously studied, more aptly depicting the evolutionary process leading to resistance development (15). Additionally it could allow for the development of a resistance model that reflects drug-induced intra-tumor heterogeneity – a complex area of cancer formation requiring more investigation (15). Different MDM2 inhibitors are under investigation for the treatment of neuroblastomas, however their connection to rapid drug resistance progression needs to be addressed (21).

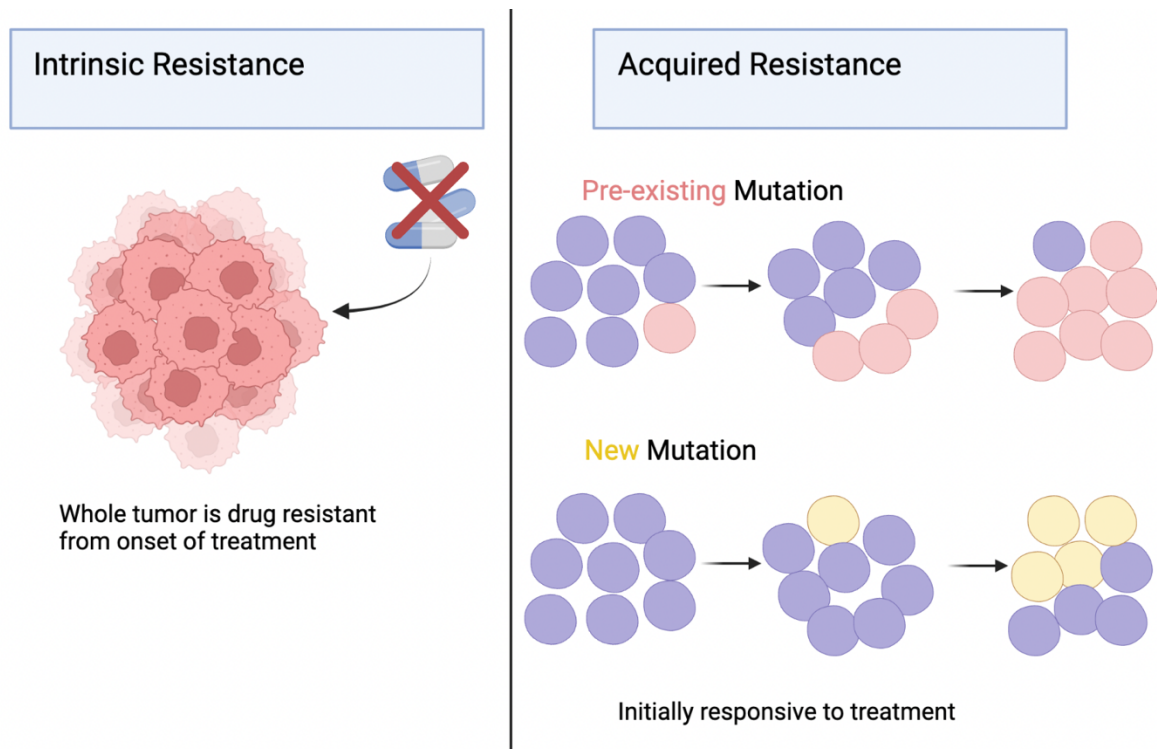


Figure 2 Mechanisms of intrinsic and acquired resistance in cancerous tumours

Illustration of the defining differences between intrinsic and acquired resistance formation in cancerous tumours, including the difference mechanisms by which acquired resistance develops. Created with BioRender.com.

1.5 Targeting MDM2 and p53

Among the genomic abnormalities responsible for neuroblastoma development, multiplication of oncogenes and loss of tumor suppressor abilities play a substantial role in tumorigenesis (8). The transcription factor p53, encoded by the gene TP53, is a tumor suppressor which is known to play a key role in all types of human cancers (15). Activation of p53 is induced by several stress signals, including DNA damage, oxidative stress and activated oncogenes (22,23). The p53 protein is employed as a transcriptional activator of p53-regulated genes (24). This results the following major outputs; cell cycle arrest, genetic instability, inhibition of blood-vessel formation, and cellular senescence or apoptosis (22,23). Murine double minute 2 (MDM2) inhibitors have been investigated for the treatment of TP53 wild-type neuroblastoma (1,3). MDM2 has been identified as an endogenous negative regulator of p53 (3,15,23). Enhanced MDM2 activity leads to inhibition of the p53 pathway and contributes to tumor formation (8,25). Under normal conditions, MDM2 physically binds

to p53 as part of a negative feedback loop. Once bound, MDM2 ubiquitinates p53 to signal for its degradation in the cytosol, preventing p53 migration into the nucleus (15,25). This leads to a regulated decrease in p53 activity (23). In neuroblastoma MDM2 function is often dysregulated with MDM2 being amplified and overexpressed, even in neuroblastomas without MDM2 gene amplification (3). MDM2 inhibitors allow for the activation of p53 by disrupting the interaction between MDM2 and p53 – allowing p53 to reach its gene targets (15). Mutations to the p53-producing TP53 gene rarely occur in neuroblastoma at the time of diagnosis (1,3,26). However, there is evidence of TP53 mutations and inactivation of the p53 pathway frequently occurring at the time of patient relapse, despite wild-type TP53 alleles originally being recorded in the patient (15,26). In these cases, this change could contribute to the development of chemoresistance and point to this altered pathway as a potential clinical target (1). Ultimately, the result of this chemoresistance is poor patient response to treatment and poor prognosis in patients with high-risk neuroblastoma (24). MDM2 inhibitors were initially hoped to activate p53 without causing genotoxic and cytotoxic stress to normal cells as other chemotherapeutic agents might do – though these findings have recently been challenged (27,28). Their use prevents p53 degradation in the cytosol, allowing for it to enter the nucleus and trigger the necessary pathways to eliminate tumor cells, and as such were thought to cause less genomic instability along the way (27). In addition, elevated MDM2 expression has been found to promote multidrug resistance in neuroblastoma cells, and therefore MDM2 inhibitors were hoped to reduced formation of resistant cells (8,29). However, this has not been found to be the case. While effective for initial treatment, treatment with MDM2 inhibitors has been found to induce TP53 mutations with drug-resistance developing rapidly thereafter (29). Therefore, investigating the mechanisms by which MDM2 inhibitor-induced resistance comes about provides a promising approach to improving neuroblastoma therapy, especially in cases of advanced disease. Serdemetan is a

MDM2 inhibitor that has been shown to be effective in the treatment of p53-null tumors, suggesting the possibility of an alternative parallel mechanism of action at play (30,31). Since the mechanism of action of MDM2 inhibitors in general is thought to be based primarily on their ability to signal p53 degradation, investigation into alternative pathways affected could help to explain this phenomenon.

1.6 Serdemetan background and mechanisms of action

Serdemetan (JNJ-26854165) is a novel small molecule originally isolated by Johnson & Johnson Pharmaceutical based on its ability to induce p53 activity (32). Its initially proposed mechanism of action as a MDM2-p53 antagonist involved preventing the association of MDM2 with MDM2 substrates such as p53, as indicated in Figure 3 (1).

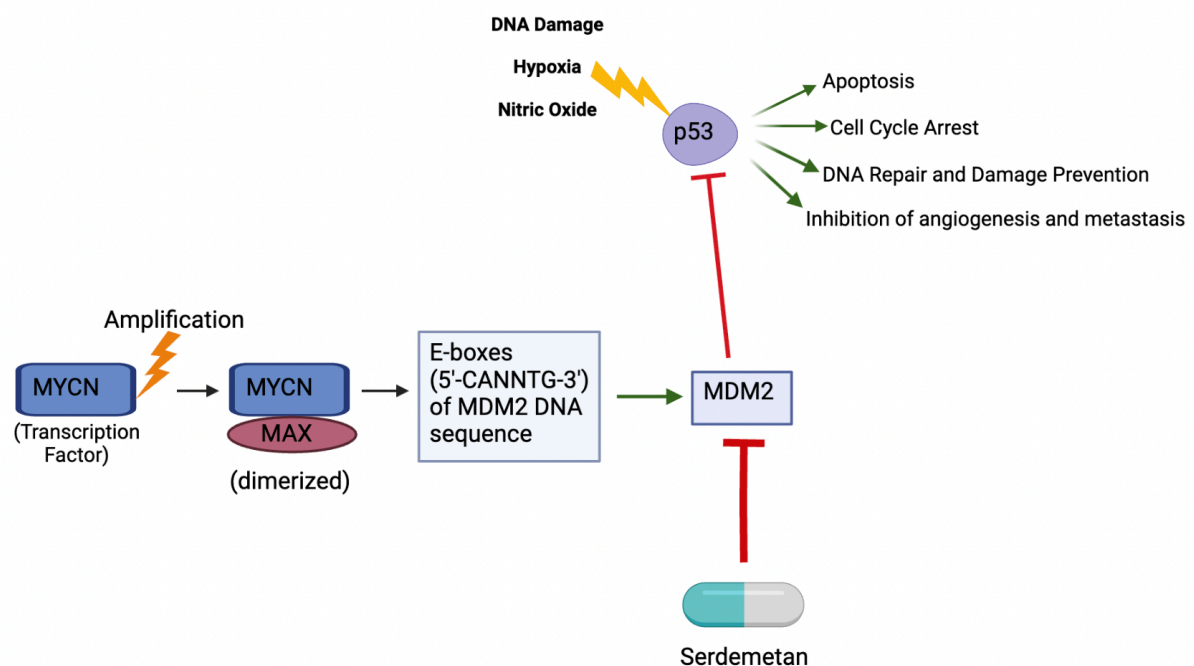


Figure 3 The mechanism of action of serdemetan within the MYCN pathway

Illustration of the known mechanism of action of serdemetan within the MYCN pathway in the treatment of neuroblastoma (12–14,33). Created with BioRender.com.

However, more recent studies point to an alternative molecular mechanism of action that might be MDM2-independent or p53-independent (31). Whereas initial studies of serdemetan suggested activity as an inhibitor of MDM2, subsequent studies showed antiproliferative

activity in both p53 wild-type and mutant tumor models, with equivalent potencies regardless of p53 status (31,34,35). How this activity fits in with a MDM2-mediated mechanism of action is not clear. One of many possible mechanisms proposed is the complex formation of Mdm2 and hypoxia inducible factor 1 alpha (HIF1 α), a transcription factor activated in response to hypoxic stimuli. Hypoxia inducible factor 1 alpha promotes angiogenesis and upregulates metabolic genes, which are necessary to sustain tumor cells (35). There is also suggestion of an alternate mechanism of action that runs parallel to the MDM2-mediated mechanism of action of serdemetan. A previous study found serdemetan additionally functions as an inhibitor of cholesterol transport, whereby it prevents the elevated cholesterol requirements of malignant cells from being fulfilled leading to eventual cell death (31). These potential off target effects of serdemetan, as identified in the literature, are summarized in Figure 4 below.

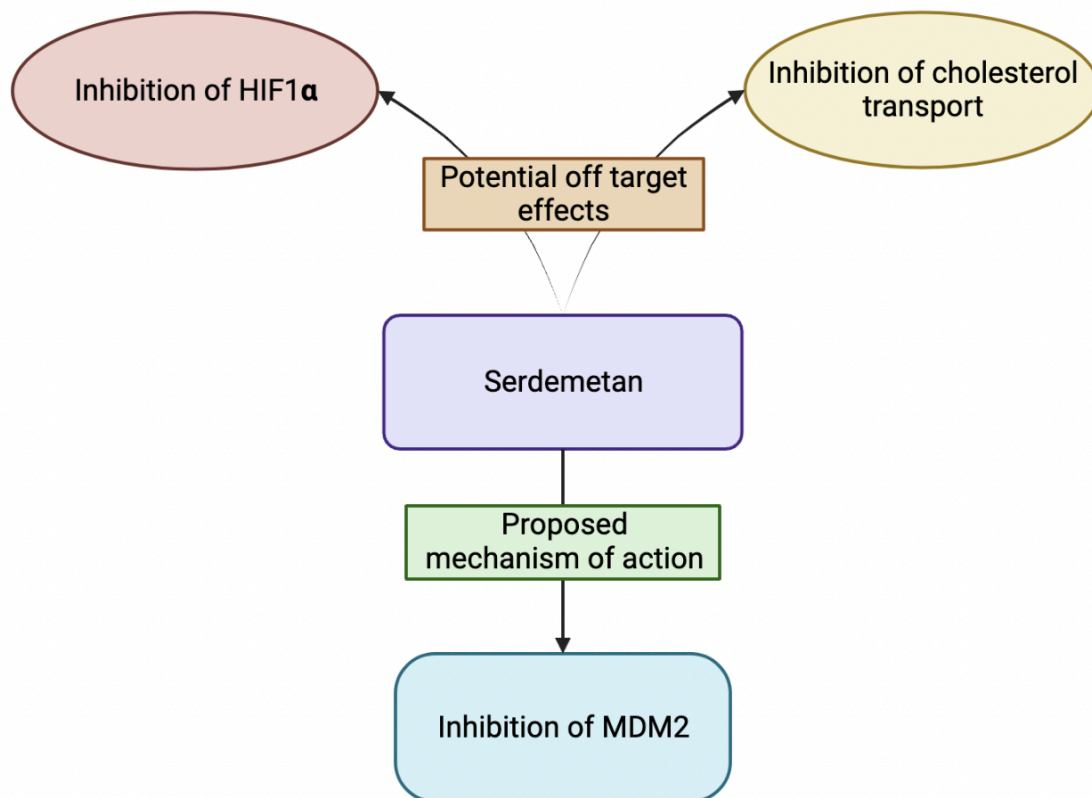


Figure 4 Potential off target effects of serdemetan

Figure displaying the potential off target effects of serdemetan by which an alternative mechanism of action for the treatment of neuroblastoma might occur (36,37). Created with BioRender.com.

Although the exact mechanism of action remains under investigation, the observed activities of serdemetan in preclinical models suggest that it may have efficacy in the treatment of cancer, so long as issues with resistance are overcome (21). Rapid development of serdemetan resistance following treatment limits clinical effectiveness – with evidence of serdemetan therapy inducing p53 mutations in previous experimental studies (21). Understanding how this resistance develops is critical to avoiding its development in the first place. High-throughput analysis of the genomes linked with serdemetan resistance might provide new insight into how to optimize these therapies (38).

1.7 Next generation sequencing as a solution

Resistance to chemotherapeutic agents can develop through a host of different mechanisms, including but not limited to changes in drug metabolism, increased drug efflux and decreased

uptake, drug target mutations and decreased drug target levels (17). These are often caused by genetic aberrations, such as single nucleotide variants, indels and structural alterations of chromosomes, that introduce functional and structural changes to proteins (5,34). In the past 18 years, the development of next generation sequencing (NGS) has allowed for the generation of unprecedented amounts of data, fundamentally altering the field of genomic research (39). NGS provides massively parallel technology that enables increased sequencing volume by allowing for the sequencing of millions of fragments in a genome simultaneously (40,41). This greatly improved on the previously much slower Sanger sequencing technology available, where only a single DNA fragment could be sequenced at a time (40,41). NGS techniques, such as whole genome and whole exome sequencing, have enabled comprehensive molecular characterization of cancer cells to allow for the identification of changes in their genomic landscape (39). This extends to cancer cells demonstrating resistance to drugs (38). NGS techniques have been and could further be applied when investigating neuroblastoma tumors, possibly leading to significant insights regarding neuroblastoma pathogenesis and resistance development (1). Of these techniques, whole genome sequencing (WGS) captures the full genome while whole exome sequencing captures solely the exons, the coding sequences making up approximately 1% of the human genome (42). Therefore, WES provides a more cost-effective approach, covering merely the actionable areas of the genome where variants might effect protein structure and function (43). Analysis of the entire exome could allow for the detection of genomic alterations associated with the molecular pathogenesis of drug-resistant neuroblastoma (9). Additionally, such analysis could aid in the definition of a NGS signature specific for serdemetan-resistant neuroblastoma (1). Because of the benefits of such techniques, WES strategies will be employed in this study to tackle its objectives.

1.8 Overview and aims of thesis

This study will aim to address the issue of drug resistance in high-risk neuroblastoma by exploiting state-of-the-art comparative genomics. In this study, sublines selected in parallel from the neuroblastoma parental cell line, UKF-NB-3, were treated with serdemetan until resistance developed. This method of subline development is significant as anti-cancer therapies similarly exert selective pressure on tumor cell populations affecting the clonal evolution processes and ultimately resulting in resistance development. This method could allow for the study of acquired resistance, as any similarities between the sublines would represent pre-existing resistant subpopulations in the heterogeneous parental cell line, providing a model system that accurately represents intra-tumor heterogeneity in a way that tumor biopsies would be unable to do. The exome sequences of four of the serdemetan-adapted UKF-NB-3 sublines generated will be analyzed comparatively to the serdemetan-sensitive parental UKF-NB-3 cell line in order to investigate the mechanisms underlying resistance formation to serdemetan, identify potential resistance biomarkers, and (ideally) to gain further insights into the mechanism of action of serdemetan.

Chapter 2: Methodology

2.1 Pipeline for calling and filtering single nucleotide variants

The UKF-NB-3 cell line was originally isolated from bone marrow metastases from a high-risk patient with MYCN-amplified stage 4 neuroblastoma (6). The UKF-NB-3 sublines, taken from the parental cell line prior to treatment, were established in parallel experiments through continuous exposure to stepwise increasing serdemetan concentrations until drug resistance was acquired (outlined in Figure 5). The final serdemetan concentration reached was 5 μ M for all UKF-NB-3 sublines. Such a method of resistance subline cultivation should provide a means for independent selection of pre-existing resistant cells from within the heterogeneous

parental cell line as well as provide a means for studying de novo variants that are acquired through treatment.

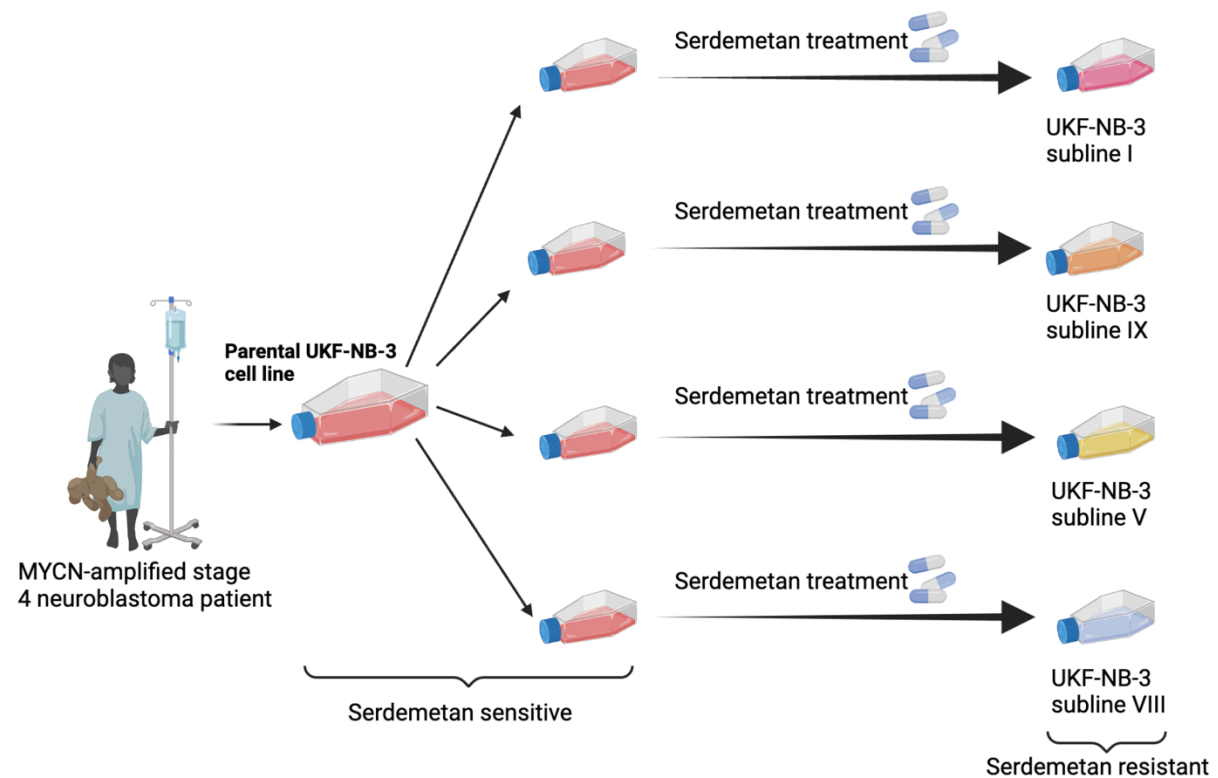


Figure 5 Subline selection protocol

Diagram outlining the selection protocol used to obtain the four independent serdemetan resistant sublines from the sensitive UKF-NB-3 parental cell line. Created with BioRender.com.

The exomes of the generated serdemetan-adapted UKF-NB-3 sublines were then sequenced using WES approaches. A table outlining details of the distinct sublines, including their TP53 status established during sequencing, is provided below (Table 1).

Table 1 UKF-NB-3 subline similarities and differences

UKF-NB-3 subline	Serdemetan concentration administered for resistance development	TP53 status
Subline I	5 μ M	wild type
Subline IX	5 μ M	mutant
Subline V	5 μ M	mutant
Subline VIII	5 μ M	wild type

An in-house pipeline was adapted and used to call and filter variants from the exome sequencing data. The pipeline is based on the Genome Analysis Toolkit (GATK) best practices for germline short nucleotide variant (SNV) and insertions and deletions (INDEL) discovery (44).

2.1.1 Variant calling

The first segment of the pipeline uses the tools indicated in Figure 6 below to call variants from the parental UKF-NB-3 cell line and serdemetan-adapted sublines' exome sequencing data.

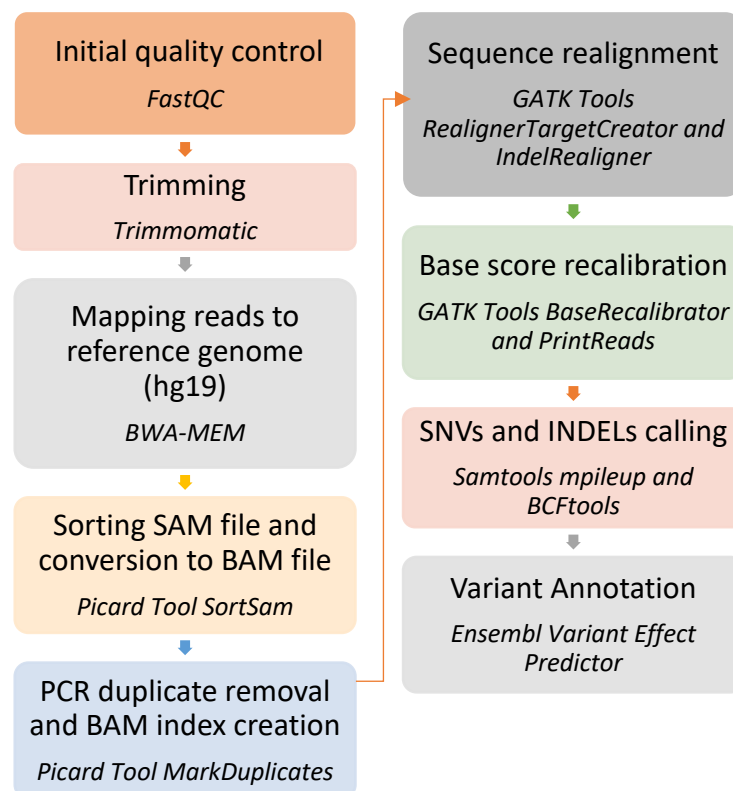


Figure 6 Variant calling pipeline

Steps in the in-house built computational pipeline to call variants in the parental and serdemetan-resistant UKF-NB-3 cell lines. Input into the pipeline is a FASTQ file containing reads from exome sequencing analysis. Output is the called variants with variant effect predictor annotation in a VCF file for each of the parental and serdemetan-resistant UKF-NB-3 cell lines. Each step is annotated with action performed and the tools used.

FASTQ files containing the reads from the exome sequencing data for each cell line were fed into the pipeline, resulting in an output of called variants, which were then filtered in the subsequent section of the pipeline. Figure 6 provides a complete overview of the first

segment of the pipeline, variant calling, including the steps and tools used to complete them. A more detailed description of each of the steps of this section will be provided in the subsections below.

2.1.1.1 Quality control of FASTQ files

FastQC was used to assess the initial quality of the raw sequencing data (45). Attention was focused on quality scores across all bases, GC and N content, reads length distribution and sequence duplications to identify low quality reads.

2.1.1.2 Trimming

Trimming of sequence reads was performed using Trimmomatic-0.38 to remove sequencing adaptors and filter out too short and bad quality reads. The default parameters of this tool were used: LEADING: 3 TRAILING: 3 SLIDINGWINDOW: 4:15 MILEN: 36.

2.1.1.3 Mapping trimmed reads to reference genome

Trimmed sequence reads were mapped to the Genome Reference Consortium human reference genome 37, hg19 (GRCh37) using Burrows-Wheeler Alignment Tool (BWA-MEM) with the algorithm bwa-0.7.17-mem. This generated Sequence Alignment Map (SAM) files containing read-group information. Parameters were used at the tools default settings; -M. Samtools flagstat was used to calculate and print statistics throughout each of the subsequent steps.

2.1.1.4 Sorting of SAM file, conversion to BAM file, and PCR duplicate removal

The Picard (v.2.17.10) SortSam tool was applied to order the output SAM file by coordinates according to the left-most mapping position of the read. The sorted SAM file was then converted to a Binary Alignment Map (BAM) format, and the Picard MarkDuplicates tool was used to detect and remove duplicate reads. Duplicates were removed to mitigate biases introduced during the data generation steps. Lastly, MarkDuplicates was used to build a BAM index file that reduced searching time through the generated BAM file.

2.1.1.5 Sequence realignment and base score recalibration

To minimise the number of reads falsely identified as SNVs due to poor local realignment, the sequence was realigned around INDELs, using GATK-3.7.0 RealignerTargetCreator and IndelRealigner tools. Next, the GATK-3.7.0 BaseRecalibrator and PrintReads tools were utilized to recalibrate the base scores of individual bases. This was done to adjust for base quality scores that could have been inaccurately estimated due to a systematic technical error of the sequencer.

2.1.1.6 Variant calling and VCF filtering

SNVs and INDELs were called using Samtools-1.7 with default parameters. First, mpileup created Binary Variant Call Format (BCF) files from the BAM files. The resulting BCF files were input into BCFtools to generate a Variant Calling Format (VCF) file containing the variants position data. Variants with the base Phred quality score below 30, which corresponds to a probability lower than 99.9% of a variant being called correctly, were marked as “low quality” using the filter option from Samtools bcfutils-1.6. Lastly, the VCF file was uncompressed and prepared for annotation.

2.1.1.7 Variant annotation with effect prediction

Variant annotation is the practice of ascribing functional information to DNA variants. To do this, the position of the variants, their effect and whether this effect was synonymous or non-synonymous, and likelihood of the variant being pathogenic (parameters SIFT and PolyPhen-2) was predicted using Ensembl Variant Effect Predictor (VEP) (Release 96). The Genome Aggregation Database (gnomAD) was used to calculate the frequency of the variant occurring in the population.

2.1.2 Variant filtering

The second segment of the pipeline includes steps used to filter the called variants produced in the previous section. An overview of how the called variants were filtered, including the selection criteria for this, is provided in Figure 7 below.

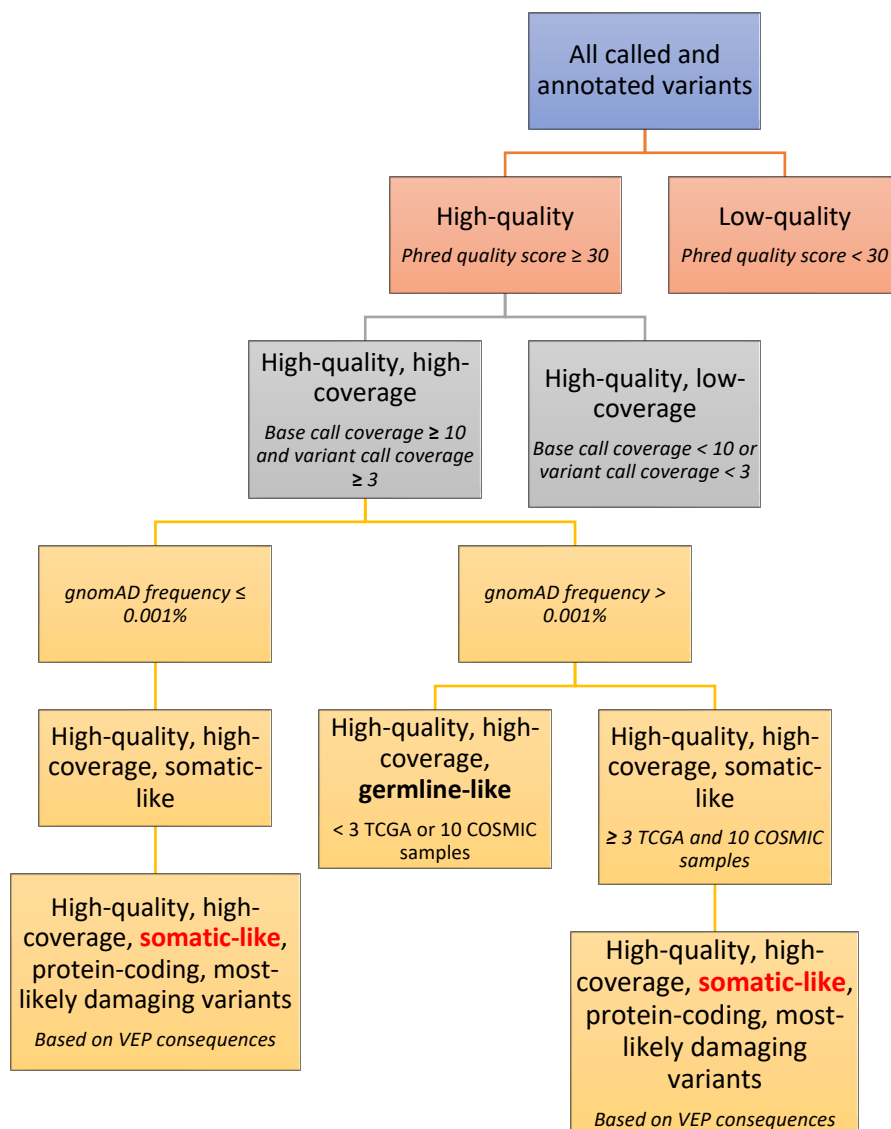


Figure 7 Variant filtering steps and specifications

Steps in the in-house built pipeline to filter variants in the parental and serdemetan-resistant UKF-NB-3 cell lines. Input into the pipeline is all called and annotated variants. Output is high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants for each of the parental and serdemetan-resistant UKF-NB-3 cell lines. Each step is annotated with the filtering categories and specifications.

The tools and parameter details of the filtering segment of the pipeline will be outlaid in greater detail in the subsequent sections below.

2.1.2.1 Filtering to obtain high-quality, high-coverage, somatic-like variants

Base call accuracy and coverage were examined first to select high-quality, high-coverage, somatic-like variants. Somatic-like variants were defined as variants that are rare in the human population presented in exomes and genomes of unrelated individuals from different geographical regions gathered into the gnomAD database. Variants were removed from the call set if the base called had a Phred quality score below 30 (less than 99.9% accuracy) and the variant was covered by less than 10 reads. The specific variant allele needed to be covered by a minimum of 3 reads for the variant to be retained. Following this, germline-like variants, defined as variants common in the human population presented in exomes and genomes of unrelated individuals from different geographical regions aggregated into the gnomAD database, were removed. Variants with a frequency above 0.001% in the gnomAD database were removed from the call set. An exception to this rule occurred when these variants additionally appeared in the Cancer Genome Atlas (TCGA) database and the Catalogue of Somatic Mutations in Cancer (COSMIC) database in at least 3 and 10 samples respectively.

2.1.2.2 Filtering to obtain protein-coding and most likely damaging variants

Additional filtering was performed based on the field ‘consequence of the variant’ from the VEP results. The consequence of the variant was determined as the effect that the variant had on the longest of the transcripts. Variants were retained if likely to be damaging to the protein. Most-likely damaging variants were defined as variants with a frameshift, incomplete terminal codons, splice acceptors, splice donors, a start codon or stop codon lost, a stop codon gained, missense variants, inframe insertions and inframe deletions. Variants were removed from the downstream analysis if determined to not alter the protein sequence or if outside of the sequencing scope (i.e. not found in the protein sequence, such as upstream variants).

2.1.3 Categorizing variants as pre-existing or acquired

Variants in the serdemetan-adapted cell line were classified depending on their presence or absence in the original parental cell line, as defined in the schematic of Figure 8.

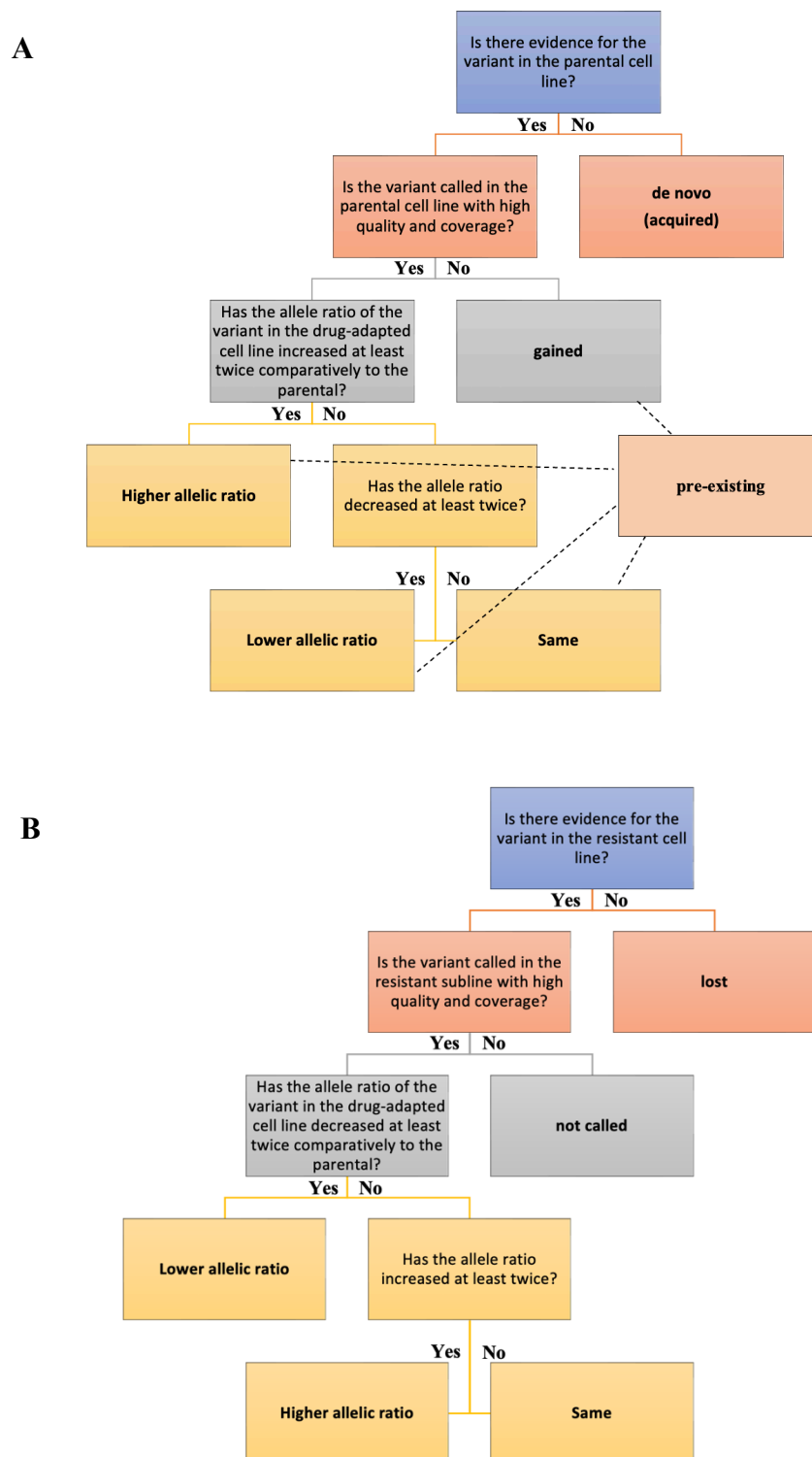


Figure 8 Workflow for classification of variants

Classification workflow of serdemetan-adapted cell line variants into categories **A**) de novo (acquired) variants or gained, higher allelic ratio, lower allelic ratio, and same variants (pre-existing) or **B**) lost, not called, lower allelic ratio, higher allelic ratio, and same.

The resistant cell lines were analyzed for changes in the variants from the parental cell line, to determine whether they showed intrinsic or acquired resistance. Acquired variants were split into categories: de novo, gained, and higher allelic ratio. Variants that were not retained in the drug-adapted cell line were similarly split into categories: lost, not called, and lower allelic ratio. Variants were labelled de novo if there was no evidence (no supporting reads) for them in the parental cell line. Gained variants were classified as those not called in the parental cell line, but with evidence of incidence in the original alignment file, or those called in the parental cell line, but with base quality scores below the Phred score threshold of 30. Variants that were classed as same, higher allelic ratio, and lower allelic ratio were called in both cell lines – the serdemetan-adapted sub-line as well as the parental cell line - and in both cell lines, their base quality score passed a threshold Phred score of 30. The three categories were distinguished from one another based on the ratio of the variant allele frequency of the drug-adapted sub-line to a variant allele frequency of the parental cell line. Variants with an increase of at least 2-fold in the variant allele frequency were classified as higher allelic ratio variants. Conversely, those with 2-fold decrease in the variant allele frequency between the drug-adapted and parental cell line were classified as lower allelic ratio variants. In cases where the ratio was not increased or decreased more than 2 times, the variant was classified as being the same between the two cell lines. The specifications of these categories are modelled in Figure 9.

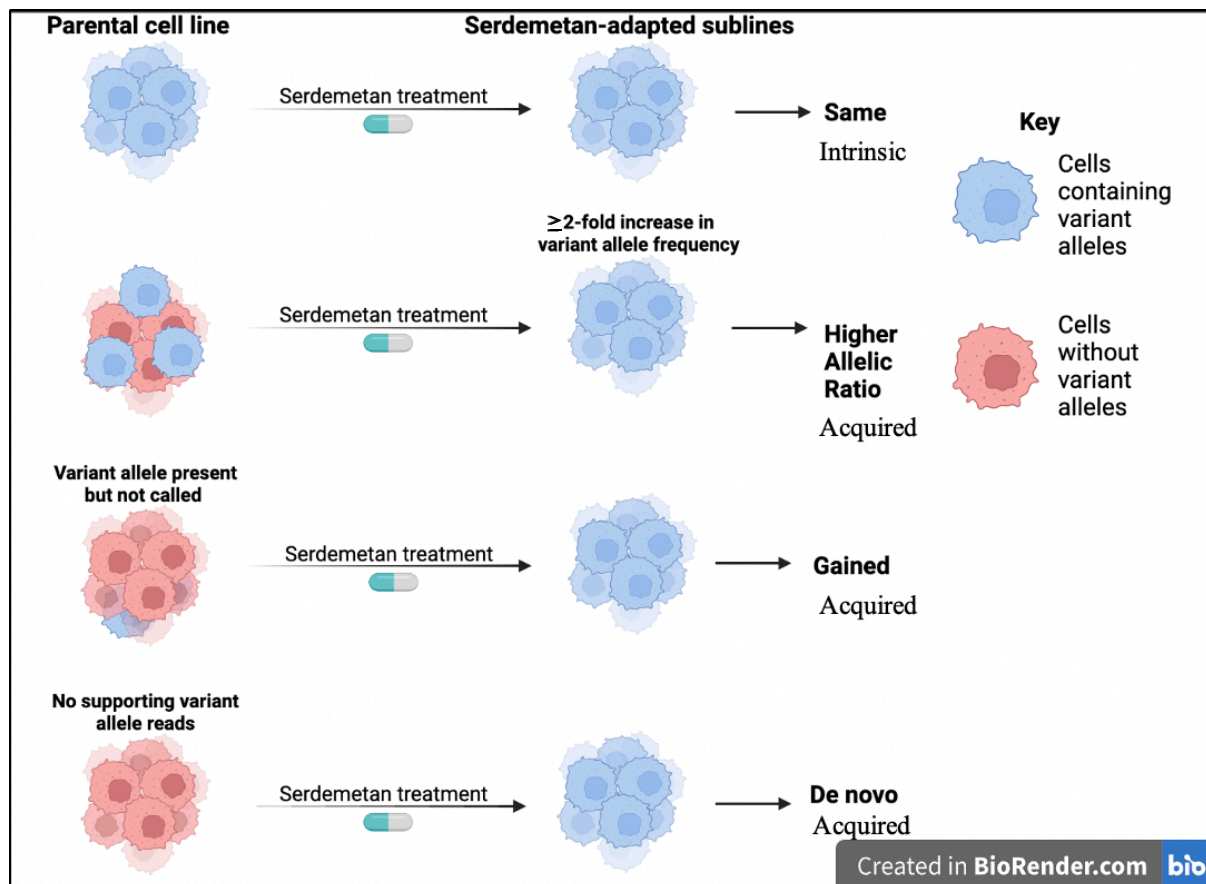


Figure 9 Categorization of variants

Overview of how variant allele frequency was used to classify the called variants of the serdemetan-adapted sublines into the categories: same, higher allelic ratio, gained, and de novo. Created with BioRender.com.

The categories of lost and not called were identified similarly as de novo and gained variants respectively, but instead were variants present of high quality and coverage in the parental cell line, but with no (lost) or poor (not called) evidence for them in the drug-adapted sublines.

2.2 Computational analysis

The computational analysis of the filtered and classified variants was divided into three parts: investigations into subpopulation selection during resistance formation, a functional analysis of the variants using Gene Ontology, and finally an analysis into the variant pathways themselves. These methods will be described in greater detail in the upcoming sections.

2.2.1 Investigation of subpopulation selection

When looking into serdemetan-adapted cell subpopulation selection, four major strategies were performed. Quantification of variants per subline, comparison of variants between sublines, analysis of variant allele frequency change, and generation of mutational profiles.

2.2.1.1 Quantifying variants per classification type

The number of variants defined as high-quality, high-coverage, somatic-like, protein-coding, and most-likely damaging per classification type was calculated for each subline and the results were displayed in a scatterplot. The scatterplot was generated using Python language with packages seaborn, pandas, matplotlib and numpy. The different serdemetan-resistant sublines were distinguished from each other by means of discretely colored dots.

2.2.1.2 Comparison of variants shared between serdemetan-adapted cell lines

Comparative studies of the resultant called variants from serdemetan-adapted sublines and subline pairs in different categories and at different levels of filtering was performed.

Heatmaps demonstrating the number of shared variants between serdemetan-adapted cell line pairs were generated using scripts made in the Python language with packages matplotlib, seaborn, pandas, and numpy. Additionally, Venn diagrams of the shared variants between all four of serdemetan-adapted cell lines were created with packages (tidyverse, hrbrthemes, tm, proustr, magrittr, dplyr, and VennDiagram) in RStudio using R language.

2.2.1.3 Variant allele frequency change: an indicator for subpopulation selection

Variants outside of the exome were filtered for and removed from analysis. Thus only those with the following consequences were examined: frameshift, splice acceptor or donor, incomplete terminal codon, stop or start lost, stop or start retained, stop gained, missense, inframe insertion or deletion, synonymous variants, 5' or 3' UTR variants, nonsense-mediated mRNA decay transcript, protein altering and coding sequence variants. The ratio of the sum of all alleles from variant calling to that of the forward and reverse non-reference

alleles was calculated as the variant allele frequency (VAF). The Python packages numpy and matplotlib.pyplot (function 'scatter') were used to generate scatterplots of the VAF change in each exonic variant in the parental cell line to the serdemetan-adapted cell lines. Statistical analysis involved performing a linear regression of VAF of exonic variants in the parental cell line and the serdemetan-adapted sublines using the python package scipy.stats with function 'linregress', while the function 'spearmanr' was used to calculate Spearman correlation coefficients.

2.2.1.4 Mutational signature analyses

High-quality, high-coverage, somatic-like, protein-coding mutations in VCF format for the parental and all drug-adapted sublines were submitted to the Mutational Signatures in Cancer (MuSiCa) tool, with the genome GRCh37 as a reference. The generated mutational profiles were taken and used to create a heatmap of the mutational signatures in RStudio, applying packages ggplot2, grid, gridExtra, and scales.

2.2.1.5 Circular visualization of variants

A circular map of the high-quality, high-coverage, somatic-like, protein-coding, most likely damaging variants from the UKF-NB-3 parental and serdemetan-adapted subline genomes was created using Circos-0.69-9 and labelled with Inkspace.

2.2.2 Gene ontology

List of pathways, proteins, genes and mutations known to be involved in the development of serdemetan resistance were determined using the following publications: (31,34,35,46).

Constituents and illustrations of the p53 pathway and the cholesterol transport pathway were acquired from the model pathways publicly available on the KEGG database (p53 signaling pathway - Homo sapiens (human) and cholesterol metabolism - Homo sapiens (human)

((47,48))). The KEGG Mapper tool: Search Pathway was used to detect the functional categories and pathways affected by de novo and lost high-quality, high-coverage, somatic-

like, protein-coding, most likely damaging variants with default settings. KEGG BRITE, an ontology database, is the reference database for BRITE mapping in KEGG Mapper and is a hierarchical classification system for biological objects such as genes, proteins and diseases (49,50). Results from the Pathway database was used to create a shared pathways heatmap and a cluster map depicting the hierarchical clustering of shared pathways in the Python language utilizing packages: matplotlib, seaborn, pandas, and numpy. BRITE databases were similarly used to create a shared functional categories heatmap using the same methods.

2.2.3 Visualization and effect predication of mutations using PyMOL

Mutations to proteins of interest identified by KEGG Mapper pathway analysis were visualized using PyMOL. Alphafold Protein Structure Database was used to obtain PDB files of non-mutated high-resolution crystal structures of the mutated proteins. The mutations were incorporated structurally using the mutagenesis wizard embedded in PyMOL software. The variants of interest were submitted to the online variant effect predictor: Mutation effect predictor using the Temporal convolutional network and the Born-Again Networks (MTBAN) and tables generated were used for analysis.

Chapter 3: Subline evolution leading to resistance formation

3.1 Results

Development of acquired drug resistance undermines the efficacy of therapy during the treatment of neuroblastoma (15). Analysis of multiple serdemetan-resistant sublines all derived from the same parental cell line could allow for detection of the critical variants required for resistance formation and inform as to whether serdemetan resistance develops from a singular subset of variants or if there are multiple possibilities by which acquired resistance arises. To explore how resistant neuroblastoma sublines evolve following serdemetan therapy, the exomes of four serdemetan resistant UKF-NB-3 sublines derived

from the same parental cell line were sequenced and their variants called and filtered. As part of the comparative studies of these resistant sublines, their variants were compared and analyzed so that overlaps between the sublines could be detected and the degree of similarity between the sublines could be defined. A surface level inspection of the sublines was performed where the number of variants in each classification category was calculated and combined into a scatterplot. Further, two methods of visualization were utilized when analyzing the overlap between shared variants. The first produced venn diagrams showing variant overlaps between all the sublines simultaneously while the second produced heatmaps that compared variants of the resistant sublines in pairs. Additionally, the changes in variant allele frequency were investigated for all the resistant sublines as compared to the parental cell line. This was considered as a possible indicator for how subpopulation selection occurred. Finally, two types of mutational profiles were created and analysed incorporating both the resistant sublines and the parental cell line. All this information together should provide useful insight into the degree of variant similarity required for resistance formation in neuroblastoma. It should also inform as to whether the serdemetan resistant sublines acquired resistance independently or developed from the same resistant subpopulation of cells in the original parental cell line.

3.1.1 Comparing the number of variants between resistant sublines

The first part of the comparative studies performed was quantifying how many high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants there were for each serdemetan-resistant subline respectfully. This was done for in each of the following different classification categories: de novo, gained, lower allele ratio, same, and higher allele ratio. These results were presented in a scatterplot (Figure 10) so that the general trends between sublines and categories could be ascertained.

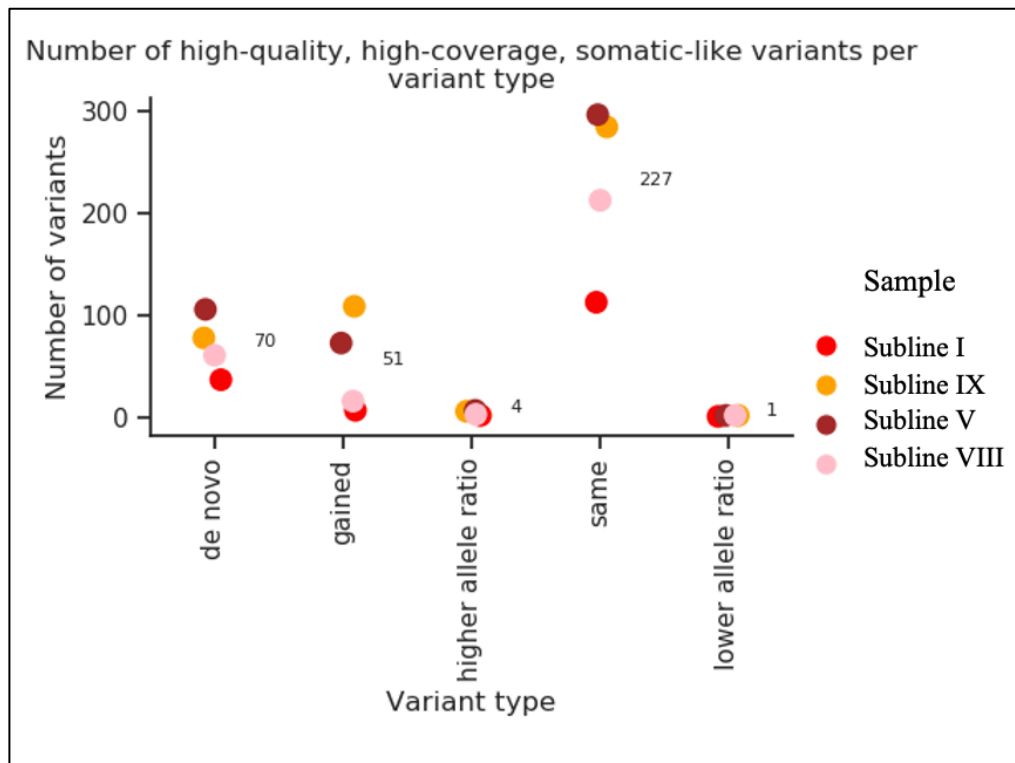


Figure 10 Scatterplot of the number of variants per classification type

A scatterplot showing the number of high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants for each serdemetan-resistant subline per classification category. Classification types include de novo, gained, higher allele ratio, same, and lower allele ratio.

Overall the serdemetan-adapted subline I contained the lowest number of high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants across all categories, with cell line VIII following closely behind. When analyzing the scatterplot for trends across the different sublines, the number of variants for each resistant subline seemed to follow a similar tendency across the different classification types. For all the categories, same was the category with the most variants detected, with a vast majority - a mean number of 227 variants across the sublines. The categories lower and higher allele ratio contained the lowest number of variants, with means of 1 and 4 respectively. The categories de novo and gained were the midline quantities calculated for all the resistant sublines – with a mean of 70 variants calculated for the de novo category and 51 for the gained category. These analogous patterns could be indicative of similar resistance mechanisms developing across the different serdemetan-adapted sublines. However, more investigations corroborating this are necessary

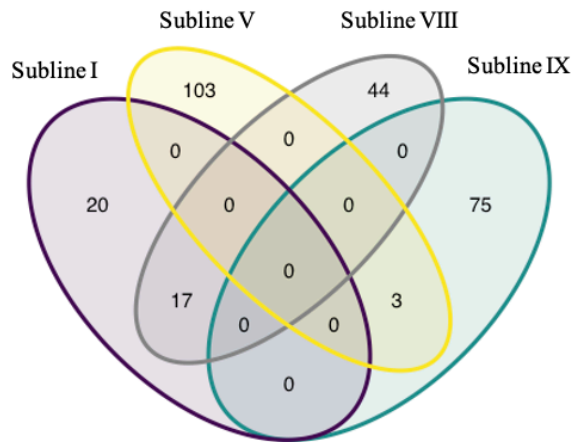
before making any such conclusions, as the variants for each subline may be different variants and not be occurring in the same pathways – meaning that different mechanisms of resistance could still have developed and be present. To assess whether the variants counted were shared or dissimilar variants, the commonalities between the detected variants were analyzed.

3.1.2 Comparison of variants shared between all serdemetan-adapted cell lines

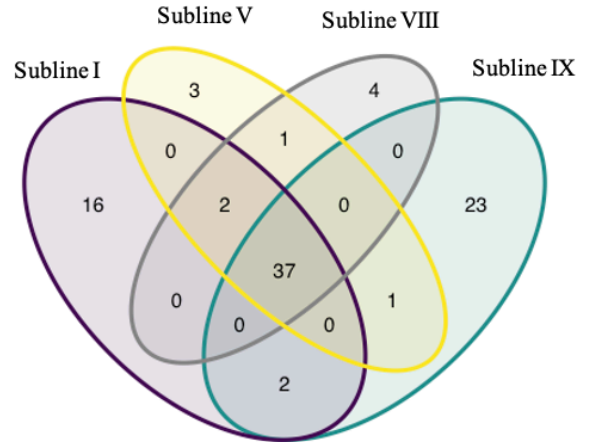
After quantifying the number of variants each subline held per classification type, overlaps between variants in the different serdemetan-adapted cell lines were compared in venn diagrams generated in RStudio (Figure 11). This was done for the variant categories de novo (no supporting reads in the parental line), gained (with base quality scores below 30 or not called but present in the original parental alignment file), lost (no supporting reads in the resistant subline) and not called (with base quality scores below 30 or not called but present in the original subline alignment file) with the highest degree of filtering looking at only the high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants. A focus was placed on the most stringently filtered variants in category de novo. This was done to identify possible common acquired resistance origins and identify the degree of similarity between the different resistant sublines.

A

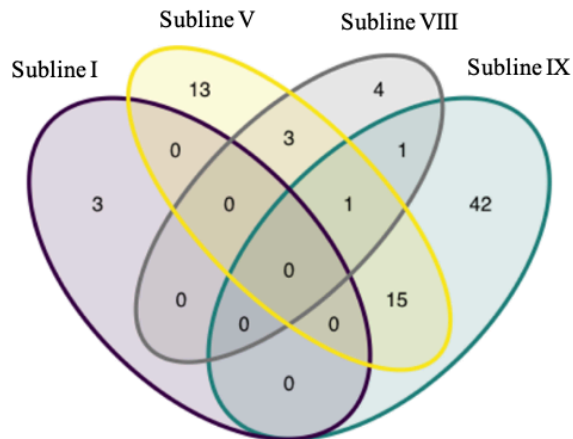
Shared de novo high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants

**B**

Shared lost high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants

**C**

Shared gained high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants

**D**

Shared not called high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants

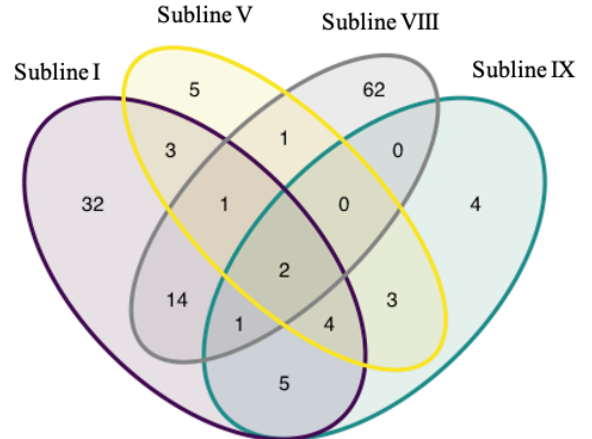


Figure 11 Comparison of shared variants between all serdemetan-resistant UFK-NB-3 sublines

Venn diagrams depicting the number of A) de novo, B) lost, C) gained, and D) not called high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants shared between all of UKF-NB-3 sublines adapted to serdemetan.

When looking at the shared de novo variants, approximately 26% of UFK-NB-3 sublines I and VIII variants were shared between them, while 2% of sublines IX and V variants were

shared (Figure 11A). No other overlaps between sublines were found and no common variants were detected in all four sublines. The lack of overlap between all the sublines could suggest that exposing the same cell line to the same drug in the same concentration may result in the development of distinct resistance mechanisms. However, the similarity detected between the subline pair I and VIII, as well as the slight similarity for pair IX and V, could be allusive of acquired resistance developing from the selection of one of two cell subpopulations with distinct mechanisms of serdemetan resistance from the original chemosensitive cell line. These shared *de novo* mutations were likely present in the parental cell line but not detected due to a lack of sequencing depth. This theory is further supported when analysing the category gained where the overlap between the variants of sublines IX and V increased to approximately 68%, supporting the idea of possible similarity in their acquired mechanism of resistance (Figure 11C). Due to the lesser degree of confidence in the gained category as acquired variants, similarity between the variants of these sublines could also be due to selection of the same pre-existing drug-resistant subpopulations rather than distinct adaptations after therapy – an example of intrinsic rather than acquired resistance (15). It is interesting to note that on average the TP53 mutated sublines, IX and V, shared the most similarity in their variant profiles, while the TP53 wild type sublines, I and VIII, had a greater overlap. When looking at the shared lost variants between the different sublines, there is a large amount of overlap between all the sublines – with 37 variants being shared (Figure 11B). This could point to the loss of certain variants playing a more important role in the formation of serdemetan resistance. However, this may not necessarily be the case as this could also result due to these variants being inconsequential in the development of the neuroblastoma tumour cells and therefore easily lost with no consequence. Investigating specifically what variants have been lost between the four sublines and studying the pathways affected is necessary to postulate whether this loss is an essential part of developing

serdemetan resistance. The not called category has less clear trends but sublines I and VIII, at approximately 15%, and sublines IX and V, at 33%, still have the greatest degree of similar variants (Figure 11D). The overlap between sublines I and IX follow closely behind at approximately 14%. More appears to be shared in regards to variants lost and not called (with an overlap of 2 variants between all the sublines) than for acquired variants – though whether this is relevant to resistance development remains to be seen. Examining how the variants are shared between the cell line pairs, as well as inspecting variant sharing at lower degrees of filtering, should provide more insight into the resistant subline relationships.

3.1.3 Comparison of variants shared between serdemetan-adapted cell line pairs

For a closer investigation of resistant subline variant sharing, heatmaps portraying the number of de novo high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants shared for a pair of UKF-NB-3 sublines adapted to serdemetan were generated (Figure 12). This was done for categories de novo, gained, lost and not called. Beyond this, heatmaps were created to illustrate variant sharing for the cell lines at earlier stages in of the variant filtering process. This was done to ascertain the penetration depth of trends in subline variant sharing observed for the highest filtering category.

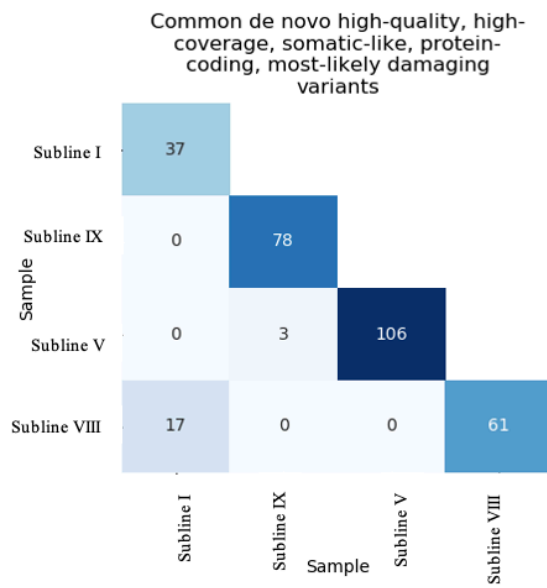
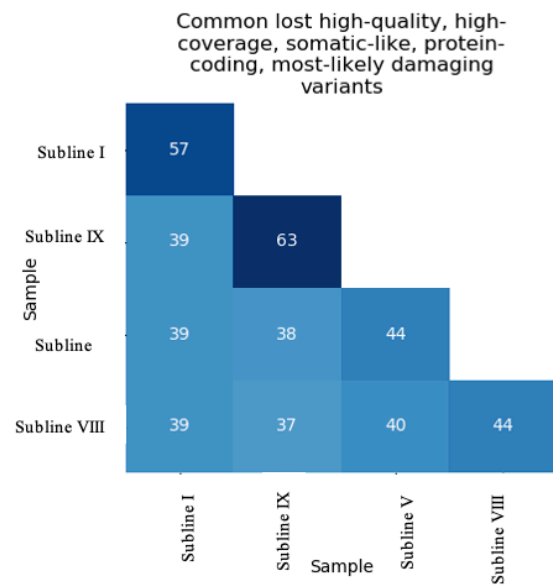
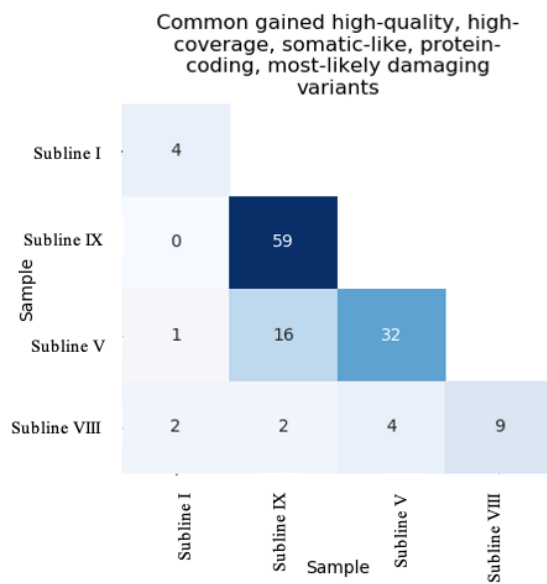
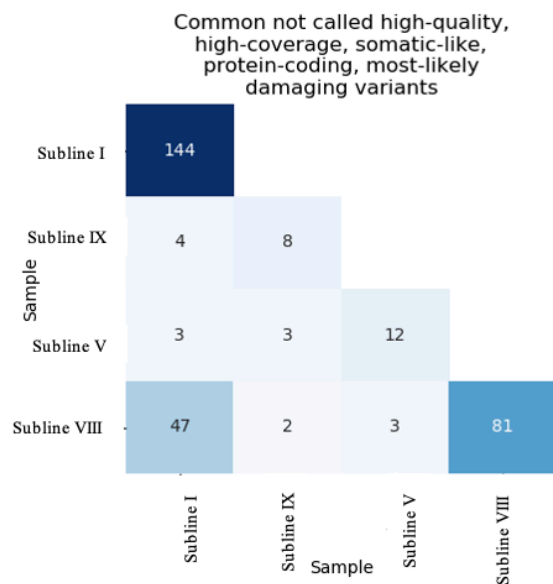
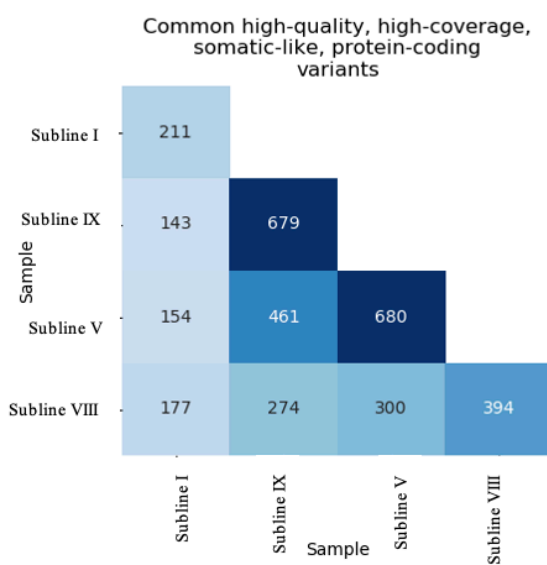
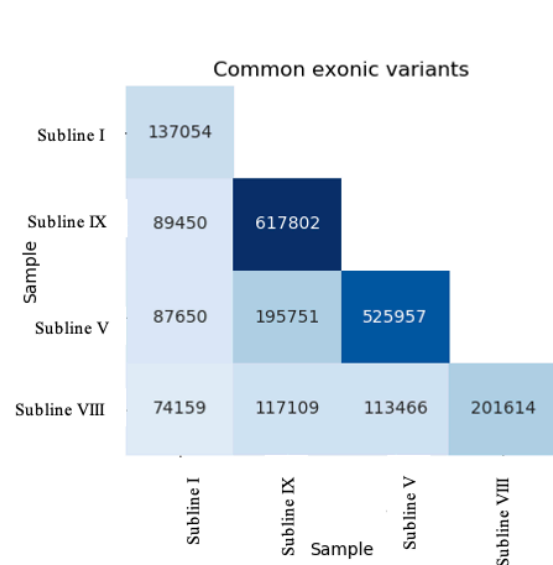
A**B****C****D****E****F**

Figure 12 Comparison of shared variants between pairs of serdemetan-resistant UFK-NB-3 sublines

Heatmaps depicting the number of A) de novo, B) lost, C) gained, D) not called high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants shared for a pair of UFK-NB-3 sublines adapted to serdemetan. The heatmaps of panel; E) shared high-quality, high-coverage, somatic-like, protein-coding variants, and F) shared exonic variants, show the number of variants called at earlier stages in the filtering process.

When examining the heatmaps, similar trends are observed as in the venn diagrams.

However, the similarity between the TP53 wild type and TP53 mutated cell line pairs is more obvious in the heatmaps. In the de novo variant heatmap, the overlap of variants between sublines I and VIII as well as sublines XI and V stand out as the only groups of variant sharing (Figure 12A). This is separation is less distinct for the gained category, as there is slightly more overlap of variants in other subline pairs (Figure 12C). However, when looking at the percentage of variants shared, these percentages are still greatest for subline pair XI and V, with the subline pair I and VIII following next. Mutual variants lost between the subline pairs are more commonly observed, with an average of 38.67 variants shared between each cell line pair (Figure 12B). In the next chapter, the pathways and variants shared here will be investigated to determine if the loss of these variants is significant to resistance development. The not called category heatmap is very similar to the gained heatmap (Figure 12D). There are more overlaps between cell lines for variant sharing, making the trends more difficult to decipher. Nevertheless, the trends become clearer and remain consistent when percentages are calculated for variants shared between sublines. Variants shared between the subline pair I and VIII is 21% of the sublines total number of variants while sharing between the subline pair IX and V is 15%. These are the biggest percentages of variant sharing by far, with the other overlaps between subline variant sharing all around only 2%, which could be due to chance rather than selection during resistance formation. This supports the theory that resistance formation in the neuroblastoma cell lines studied here resulted from the alteration of two distinct pathways, each of which has two sublines. It is not known whether this similarity of variants between the cell line pairs results from acquired resistance developing due to similar genetic aberrations occurring during treatment or due to the selection of the

same small subpopulation of resistant tumour cells not being eliminated during treatment and diverging from there.

To finish of these investigations, heatmaps examining the variants shared between cell line pairs at earlier stages of the pipeline process, before categorization and filtering of the variants, were generated. This was done to see when the trends observed between the cell line pairs could first be detected and if they occur prior to filtering for acquired variants. Two heatmaps were generated at these earlier stages. The first displays high-quality, high-coverage, somatic-like, protein-coding variants before filtering based on the consequence of the variant or categorization was performed (Figure 12E). In this heatmap, the trends observed previously begin to fade. While the resistant cell line pairs that shared the most variants were still sublines I and VIII, with a similarity of 29%, and sublines IX and V, with a similarity of 33%, the overlaps of shared variants between all the sublines are less distinct. There is an overlap of 26% between sublines VIII and XI and an overlap of 28% between sublines VIII and V. The second heatmap created included all the called exonic variants prior to any quality filtering or categorization (Figure 12F). In this heatmap, the trends in overlap between cell line pairs were consistent with those previously found however they were less distinct. The cell line pairs of sublines I and VIII and sublines IX and V still maintained the highest degree of overlap, with 22% and 17% variant overlaps respectfully. However, percentage overlaps between the other sublines were not far behind, with the sublines VIII and V having a 16% overlap, the sublines VIII and IX having a 14% overlap, the sublines IX and I having a 12% overlap, and the sublines I and V having a 13% overlap in variants. The lack of distinction between shared variants at these earlier filtering stages corroborates the idea that the variants causing resistance were likely acquired independently for the two groups of cell line pairs. However, since the trend in maintained, some resistance might have

been intrinsic to certain subpopulations of cells prior to treatment. To help clarify how serdemetan resistance formed in the sublines and determine whether it was pre-existing or not, variant allele frequencies (VAFs) were calculated and compared between the UKF-NB-3 parental cell line and each of the serdemetan-adapted sublines.

3.1.4 Variant allele frequency change: an indicator for subpopulation selection

To further study how resistant subpopulations were developed, VAFs across all the called exonic variants of the UKF-NB-3 cell line and the serdemetan-adapted cell lines were calculated. A VAF has been defined as “the percentage of sequence reads observed matching a specific DNA variant divided by the overall coverage at that locus” (51). The changes in VAF for each exonic variant from the UKF-NB-3 cell line resulting from the adaption of the sublines to serdemetan were examined in the form of Scatterplots (Figure 13).

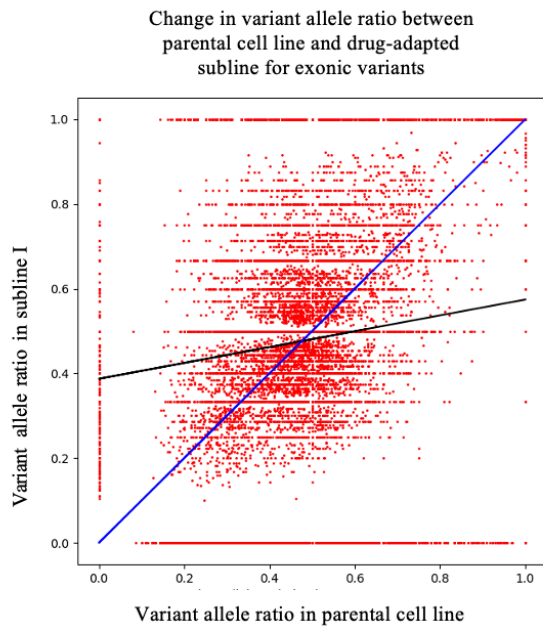
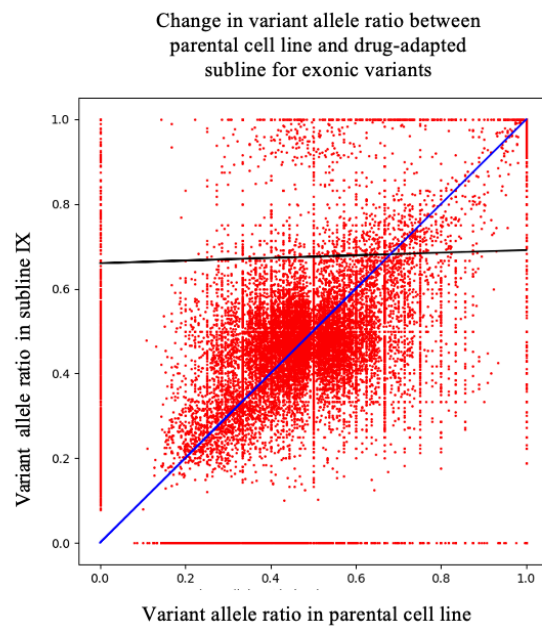
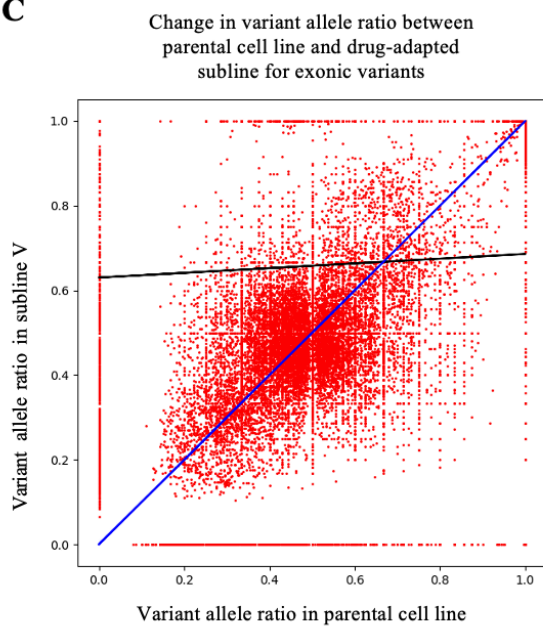
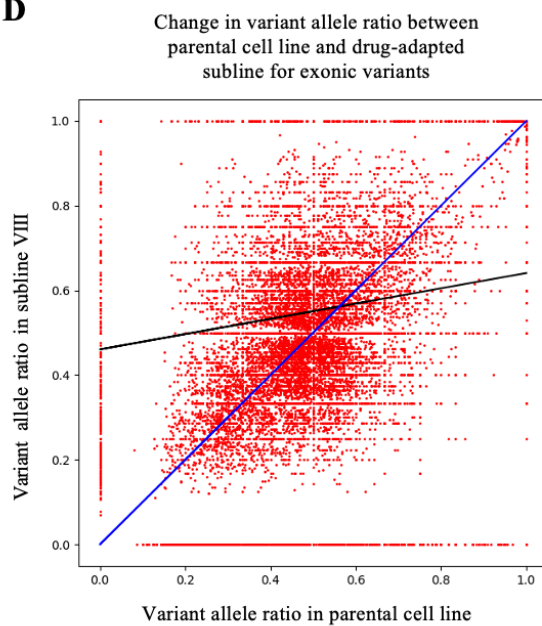
A**B****C****D**

Figure 13 Shift in variant allele frequency between UKF-NB-3 and UKF-NB-3 serdemetan-adapted cell lines

Scatterplots illustrating the change in VAF between the parental UKF-NB-3 cell line and the four serdemetan-adapted sublines A) I, B) IX, C) V and D) VIII. All exonic variants were included in the analysis. Each point on the scatterplot represents the VAF of the variant in the parental cell line (x-axis) and the corresponding VAF in the serdemetan-adapted subline (y-axis). When the VAF equals zero, the variant was not called. The black line shows the regression line for the data, and the blue line indicates the case where the VAF of the variant does not change between the parental and the serdemetan adapted subline.

Looking at how the VAFs change between the chemo sensitive parental cell line and the serdemetan resistant sublines will allow for an analysis of how much the variant alleles changed in response to resistance formation to serdemetan. A linear regression analysis was

performed, and spearman correlation coefficients calculated to evaluate the relationship between VAF in the parental cell lines and VAF in each serdemetan-adapted subline. The results of this process were very similar for all four serdemetan-adapted sublines, however greater changes in VAF were observed in the scatterplots for the sublines IX and V, pointing to a higher degree of mutation in the variants of these sublines as compared to the parental cell line (Figures 13B and 13C). This matches what was seen in the scatterplots depicting the number of variants per type for each subline (Figure 10). Additionally, all the sublines were altered significantly from the parental cell line after their adaptation to serdemetan. The similarity in the scatterplots generated for sublines I and VIII, as well as the ones for IX and V, support the hypothesis that the sublines within these subline pairs are related in terms of resistance mechanism and might have arisen from the same subpopulation of chemo resistant cells in the parental tumour. In the next section, mutational profiles of the different sublines were created and analysed to strengthen the understanding of the relationships between the sublines in serdemetan resistance development.

3.2.5 Mutational profiles and signatures

Several previous studies have stressed the importance of analysing the different expression patterns observed in neuroblastoma (1,5). The heterogeneous nature of neuroblastoma presentation makes the early prediction of patient risk at the time of diagnosis a foremost objective so that an optimal treatment strategy may be implemented (1,38). Generation of expression profiles may allow for early identification of patient risk group and point to the best actionable clinical course regarding tumour chemotherapeutic sensitivity, providing a window into personalized medical treatment of serdemetan-resistant neuroblastomas (1,5). Additionally, the generation of such profiles could be important for future study reference and help to close the gap between pharmacological and biological studies (5). Therefore, it was important in this study to generate profiles such as this for future clinical reference.

Additionally, comparison of the mutational profiles generated for each resistant subline will provide insight into their degree of variant similarity. Two diverse kinds of profiles were generated, a mutational profile heatmap as well as a circular representation of the mutational profile which will be discussed in the next section (Figure 14 and 15 respectively).

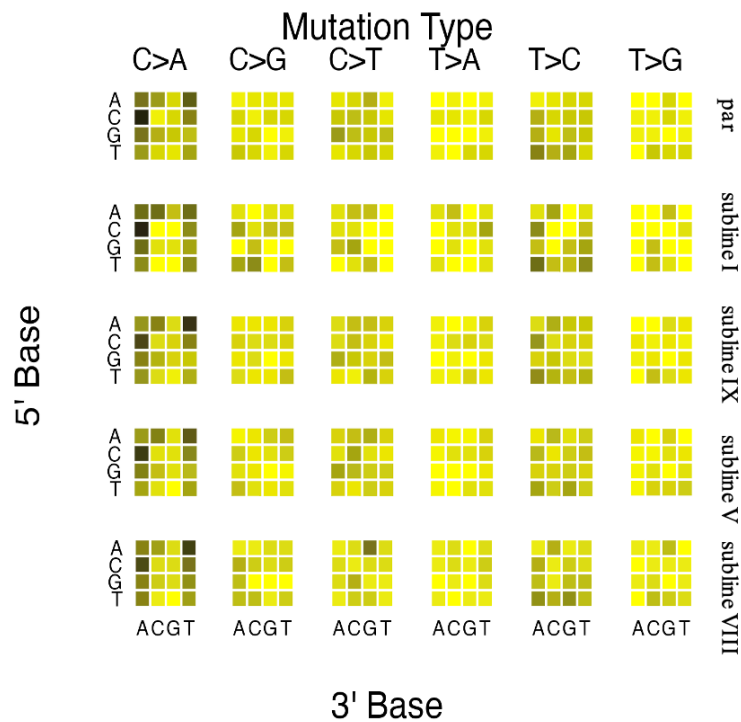


Figure 14 Mutational profile heatmap

Mutational signatures heatmap depicting mutation type frequency, in terms of base changes, for the parental UKF-NB-3 cell line and each of the serdemetan resistant sublines. The heatmap was generated using R language with packages ggplot2, grid, gridExtra, and scales based on signatures obtained from the web application MuSiCa after submitting the cell line whole exome sequencing data.

The mutational profile heatmap is a visual representation of the somatic mutation type frequency – with regards to base substitutions – observed for each subline as well as the parental cell line. Analysis of the mutational heatmap indicates further similarity between the sublines – in addition to that of the parental cell line (as expected as some similarity is expected to originate from selection of cells from within the parental tumour) - building on the similarity observed for the changes in VAF ratio. This could indicate that only a small number of the mutations detected are necessary to cause serdemetan resistance, and are in fact de novo from the parental chemo resistant cell line. Support for this is further provided

from the de novo venn diagram in Figure 11A that indicated that only a small number of de novo mutations are present in the resistant cell lines. From solely a visual analysis of the heatmap, the chemo resistant sublines I and VIII appear to have the most unique mutational profiles for base substitution from the parental cell line. This does not necessarily mean that these sublines are the most mutated of the resistant sublines, but rather that the types of mutations that they contain differ more drastically from the parental cell line than that observed for sublines IX and V. Additionally, the relationship between the resistant subline pair I and VIII, as well as the subline pair XI and V, is strengthened with regards to their mutational signature similarities.

3.1.6 Circular visualization of variants

A second mutational expression profile, the Circos plot illustrated in Figure 15, was created so that the frequency of variants in relation to their chromosomal position on the karyotype for GRCh37 (hg19) could be examined.

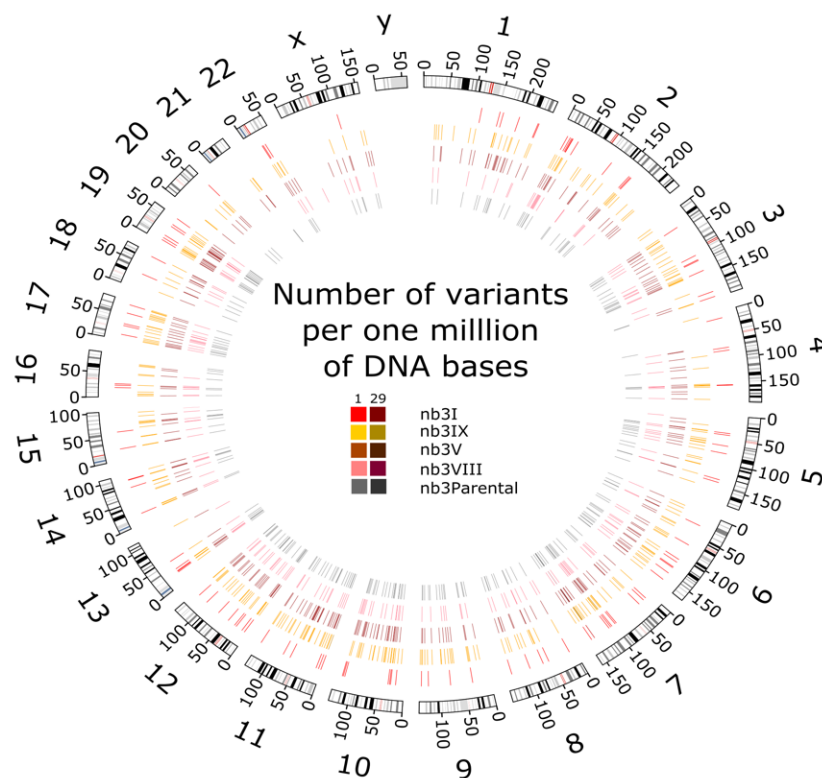


Figure 15 Circular visualization of variants

Circular representation of the cell line expression profiles showing high-quality, high-coverage, somatic-like, protein-coding, most likely damaging variants. Each coloured line signifies the number of variants per one million DNA bases. The differently coloured bands - bright and dark - represent low (with a minimum of 1) and high (with a maximum of 29) number of variants, respectively. The outer part of the plot demonstrates karyotype GRCh37 (hg19). DNA positions are marked every 50 million bases. The outer labels signify chromosome numbers.

Variants appeared on almost every chromosome for every cell line, with the only exception being a lack of variants on the Y chromosome. However, the absence of variants on this chromosome is likely due to tumour cells the cell lines were created from coming from a female patient. The chemo resistant sublines IX and V showed the greatest number of variants when compared to the parental and other sublines. This is consistent with what was observed for the variant number and changes in VAF ratio scatterplots (Figures 10 and 13 respectfully). The resistant subline I is the least mutated subline according to the Circos plot – and supported by the number of variants Scatterplot. This difference in the degree of mutation between the different sublines indicates that there is flexibility in regards to the number of variants required for serdemetan resistance to develop and that rather, the number required falls within a range rather than a set point. Taking the findings of this analysis, as well as all the prior analyses, into account is necessary to pull together a picture on how stringent the pathway to serdemetan resistance is and what that pathway (or pathways) may be.

Chapter 4: Functional analysis of the variants and variant pathways

4.1 Results

Investigations into variant gene ontology and the pathways affected by the variants were undertaken to help complete the picture of how serdemetan resistance develops during patient therapy. Currently little is known about how resistance to serdemetan develops and there is general confusion surrounding the mechanism of action of serdemetan (30). Though introduced as a MDM2-inhibitor, with p53-activating properties, a subsequent study has suggested an additional parallel mechanism of action as a cholesterol transport inhibitor

(31,52). Cholesterol transport is crucial for several cellular functions such as the formation of the plasma membrane and lipid rafts to steroid hormone production (31). This is especially true for malignant cells, where cholesterol requirements are elevated to match their heightened proliferation (31). Inhibition of cholesterol transport in neuroblastoma cells has been found to successfully hinder the growth of these cells and promote cell death – making the cholesterol transport pathway a suitable chemotherapeutic target (31). Should malignant cells gain variants that alter the targets of these cholesterol transport inhibitor drugs so that cholesterol is still transported at an elevated rate, the tumor would become resistant to treatment. Therefore, understanding these pathways and examining whether and where they have been altered in the serdemetan-resistant sublines of this study is of interest.

Additionally, variants introduced affecting the ability of serdemetan to act as an inhibitor of MDM2, the primary mechanism of action identified for serdemetan, should also be effective as a method of stimulating drug resistance (46). There is dispute about exactly what is most crucially effected in this pathway by serdemetan's inhibition of MDM2 (35). Investigating exactly where the mutations are introduced in this complex signaling pathway may help future studies to combat these variants causing drug-resistance. Whether variants effecting multiple different pathways, or just a singular pathway, are necessary for resistance development is to be determined. Investigations into the gene ontology of the resistant UKF-NB-3 sublines will provide information as to if any critical variants/pathways are altered in all the sublines or if there are multiple avenues available through which serdemetan resistance can be achieved. To examine gene expression differences between the parental UKF-NB-3 cell line and the serdemetan-adapted sublines, their specific gene lists were mapped to the biologic process GO classification as defined by KEGG Mapper. However before this undertaking, research of what previous literature said on the topic of serdemetan resistance and its mechanism of action was performed to determine what pathways are most

likely affected during serdemetan treatment, and therefore what would likely be altered during resistance development. This was when cross referenced with the results from KEGG Mapper so that key variants could be identified and analysed in PYMOL. Analysis in PYMOL served to inform as to whether a specific variant was likely to significantly affect the protein analysed, and therefore the pathway. Two main pathways, the MDM2-mediated pathway and the cholesterol pathway, have been identified as playing a part in serdemetan's mechanism of action, and therefore likely to be affected during resistance formation. These pathways will be discussed in the next sections of this chapter.

4.1.1 Mechanisms of resistance to serdemetan

The mechanism by which resistance to serdemetan develops is not clearly understood. This confusion is compounded by the lack of clarity surrounding the mechanism of action of serdemetan (53). Two pathways have been identified as part of the mechanism of action of serdemetan during treatment – inhibition of MDM2 and cholesterol transport (31,52). As inhibition of these pathways has been identified as crucial for serdemetan mechanism of action, if either contain *de novo* variants distinct from the chemo sensitive parental cell line it could reveal how resistance to serdemetan was established. The next three sections discuss possible critical mutations to the genes of these pathways that might be involved in serdemetan resistance development as indicated by previous studies. The first pathway to be discussed will be the p53 pathway – the pathway initially specified as being involved in serdemetan's mechanism of action (54).

4.1.1.1 Mutations in p53 signalling pathway

Serdemetan is widely understood to have an inhibitory effect on the action of MDM2, a p53 tumour suppressor protein inhibitor, as part of its mechanism of action (35). Initially, the primary action of serdemetan as a MDM2 inhibitor was believed to be suppressing tumour growth and inducing cell death through interrupting the association between MDM2 and p53,

thereby preventing ubiquitination of p53 by MDM2 (8,55). This action in turn inhibits the degradation of p53 in the cytosol, permitting the transportation of p53 to the nucleus (33). There it activates the transcription of several tumour suppressing genes – with apoptosis and cell cycle arrest of tumour cells being the principal outcomes (33,34,47). In a healthy cell, when certain stress signals are encountered that cause DNA damage, hypoxia, oxidative stress etc. p53 is activated (22,23). MDM2 functions as part of an autoregulatory negative feedback loop where-in it regulates the activity of p53 as a tumour suppressor protein by signalling for the decrease of the cellular levels of this protein once the required levels for signalling have been reached (56,57). The cellular levels of MDM2 are then sequentially regulated, with lowered p53 levels signalling for degradation of MDM2 (56). The N-terminus of MDM2 binds to p53 and suppresses its activity, while its C-terminus functions as an E3 ubiquitin ligase (34). An illustration of this pathway, adapted from figures in an article Kung et al. and the KEGG Pathway Database, is shown in Figure 16 (33,47).

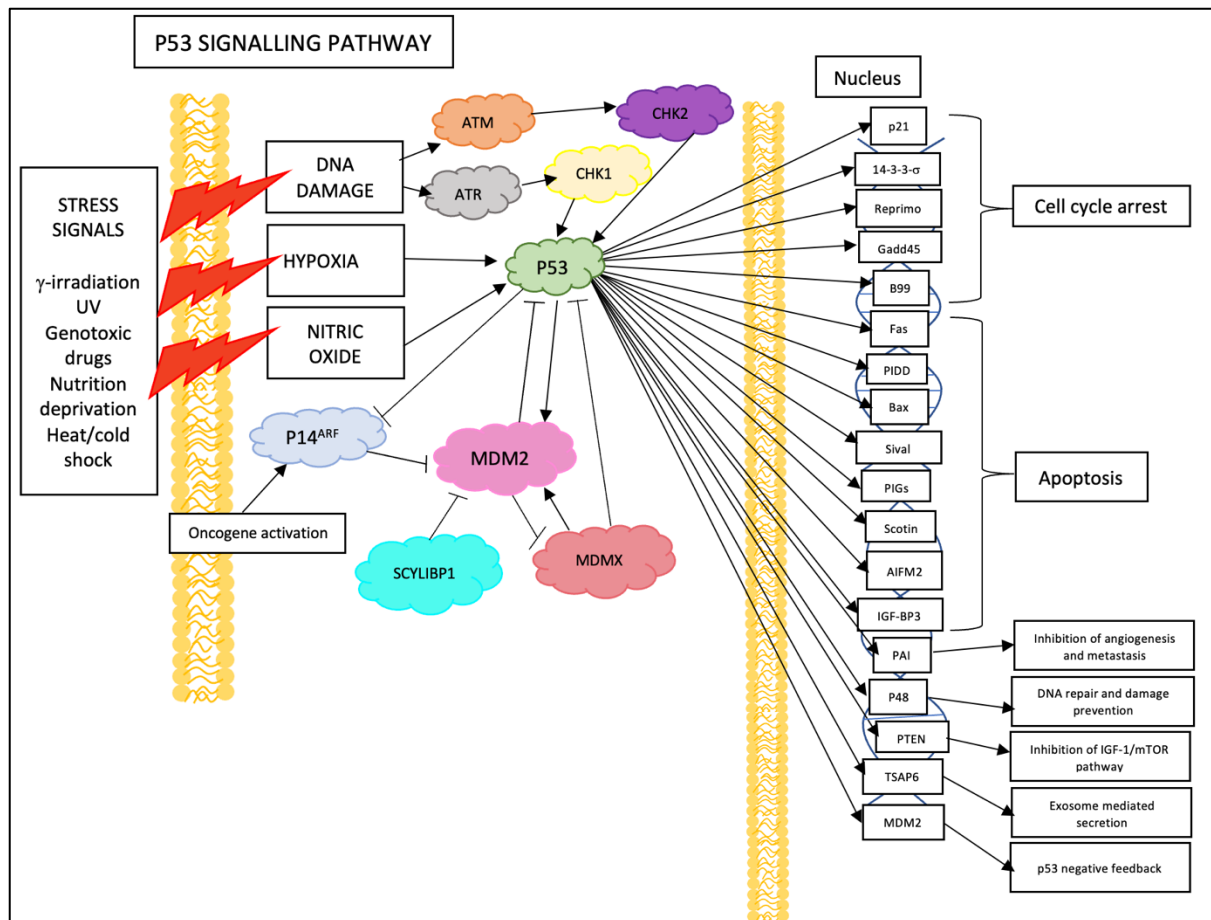


Figure 16 The p53 signalling pathway

Illustration of the p53 signalling pathway found in homo sapiens adapted from the KEGG Pathway Database and Kung et al.

In neuroblastoma malignancies it appears that while expression of the MDM2 oncogene is often amplified – mutations of the tumour suppressor p53 are rare (found in less than 2% of neuroblastoma tumours) despite their usual prevalence in other kinds of cancers (34). The 2% found are usually confined to cases of relapsed patients where drug resistance has developed (34). The most frequently observed mutations to the protein include: the substitution of cysteine to phenylalanine at codon 135 (135Phe) and cysteine to tyrosine at codon 135 (135Tyr) (34,58). Other p53 mutations that have been correlated with drug resistance include the mutation of codon 182 from cysteine to a stop codon, the substitution of codon 245 from glycine to serine, the substitution of codon 176 from cysteine to phenylalanine, the substitution of codon 342 from arginine to leucine, the substitution of codon 286 from glutamine to lysine, the substitution of codon 282 from arginine to tryptophan and finally the

duplication of exons 7-8-9 in TP53 (34,58). These have all been flagged as potential variants to look out for when analysing the mutations present in the serdemetan resistant sublines of this study.

In addition to the mutations of p53 itself, resistant neuroblastoma cell have also been found to have MDM2 amplification beyond the levels of chemo sensitive cell lines – reducing p53 function (34,59). This MDM2 activity amplification is often due to inactivation of ARF, the gene coding for MDM2 regulatory protein p14^{ARF}, through deletion or promoter methylation (34,60–62). In fact approximately half of relapsed patients are found to have mutations at either ARF or MDM2 locus (34). These mutations prevent p14^{ARF} from binding and impairing MDM2's regulatory function of p53 (34,62). These are clear examples of resistance forming mutations/alterations other than the classical TP53 mutations found in most other cancers. This will be discussed more extensively in the next section.

4.1.1.2 Alternate mutations in the MDM2 pathway

Serdemetan has been found to have equivalent potencies in both TP53 wild-type and mutated/null tumours (63,64). This indicates that there must be methods other than TP53 inactivation through which serdemetan is able to target tumour cells. In healthy cells MDM2 stabilizes HIF1 α under hypoxic conditions – allowing HIF1 α to perform its function as a promoter of angiogenesis and an up-regulator of metabolic genes (35). A study by Lehman et al. found that, under hypoxic conditions, serdemetan treatment prevented the stabilization of HIF1 α by MDM2 –subsequently decreasing the levels of its downstream targets: VEGF and several other glycolytic enzymes (35,37). This pathway is modelled in Figure 17. The prevention of HIF1 α stabilization by MDM2 occurred regardless of the cell line's p53 status (35). VEGF is an enzyme responsible for the re-direction of blood flow and nutrients to tumour cells under hypoxic conditions, enhancing cell proliferation and thereby stimulating

angiogenesis (35,65). Activation of this enzyme through the binding of VEGFR-2 leads to a cascade of different signalling pathways including: the MAPK signalling pathway, the PI3K-Akt signalling pathway, and the mTOR signalling pathway (65). Glycolytic enzymes found to decrease with serdemetan treatment included Enolase, Glut1, MMP2, and PGK1/2 (35,37). These enzymes are critical for the maintenance of glycolysis in tumour cells when exposed to hypoxic conditions (35). Should this pathway contain mutations that enable the stabilization of HIF1 α to perform its angiogenesis stimulatory functions in the presence of serdemetan – drug resistance would develop as a result. Therefore these and related genes should be flagged as possible sources of serdemetan resistance.

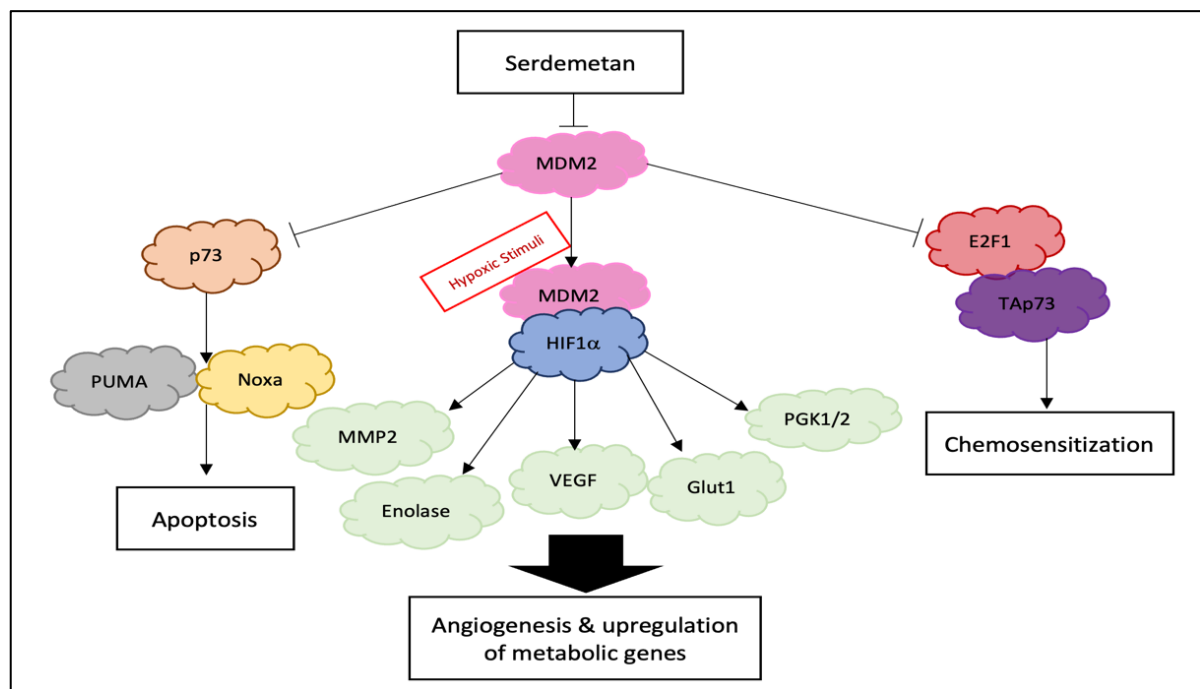


Figure 17 The Mdm2-HIF1 α Axis

An illustration of serdemetan modulation of the Mdm2-HIF1 α axis based on information and figures from Lehman et al., Zafar et al., Joshi et al., and Pflaum et al. (34,35,37,46).

Other possible p53-independent MDM2 pathways that might be mutated during resistance development include: the inhibition of TAp73 and E2F1, which induces G1 cell-cycle arrest and chemosensitizes cells, and the MDM2-p73 interaction, partly responsible for induction of apoptosis in neuroblastoma cells (34,46). p73 is a structural homolog of p53 with both

overlapping and distinct functions with a history of mutation in neuroblastoma (46,66).

However, serdemetan's effect on these pathways has not been confirmed – only the effect of Nutlin-3, a structurally different MDM2 inhibitor (34). Still, these pathways are important to consider when analysing the variants and pathways affected in this study's serdemetan resistant sublines as they may well be affected as demonstrated in Nutlin-3. These are just some of the pathways that have been shown to be affected by serdemetan treatment but are not inclusive of all the possible options when it comes to serdemetan-resistance development. The next section will discuss how variants in the cholesterol transport pathway may be equally, or more, significant in the formation of resistant sublines.

4.1.1.3 Mutations in cholesterol transport pathway

Beyond the studies investigating serdemetan as an MDM2 inhibitor, serdemetan has been linked to the inhibition of cholesterol transport through degradation of ATP-binding cassette subfamily A member-1 (ABCA1) cholesterol transporters (31,52). This ultimately results in the stimulation of cell death (31). Cancerous cells have commonly been found to exhibit atypical regulation of cholesterol-regulated genes – typically requiring elevated levels of cholesterol to sustain their enhanced proliferation (31). Common malignant genes requiring this elevated cholesterol requirement include those for protein kinase C- β protein, a segment of the B-cell receptor signalling complex, and 5-lipoxygenase (67,68). Because of its insolubility in plasma, cholesterol is transported in conjunction with either low-density lipoprotein (LDL-c), via the LDL receptor, or high-density lipoprotein (HDL-c), imported by scavenger receptor class B member 1 (SR-BI) and exported by ABCA1 and ATP-binding cassette subfamily G member-1 (ABCG1) (31,36). This alternate pathway could additionally explain how serdemetan has been found to be effective for treatment of neuroblastoma cell lines containing TP53 null mutations (31). In fact, a study by Uhr et al. has even suggested this to be the dominate mechanism of action of serdemetan (52). A previous study conducted

by Jones et al. found there was an accumulation of cholesterol within endosomes of cells due to an downregulation of ABCA1 expression in response to serdemetan treatment (31). In response, cells try to reverse this action by upregulating SREBF-1/2 and LXR- α/β , which subsequently increases ABCA1 transcription (31). These actions are inadequate to overturn the downregulation of ABCA1 and cholesterol continues to accumulate until reaching a cytotoxic level – triggering cell death (31). As part of the study, Jones et al. investigated the effects of serdemetan on resistant MEF cells (31). Here cells appeared to boost ABCA1 expression to circumnavigate the effect of serdemetan. This likely indicates that ABCA1 is a critical protein targeted by serdemetan, likely downregulated by treatment, and it will be of interest to investigate how it has been affected in the resistant sublines examined here (31). Jones et al. also noted that the effects observed in their could still be due to serdemetan having a more comprehensive influence on cholesterol metabolism and transport genes – so this whole pathway has been highlighted for review during pathway variant analysis (31). The cholesterol metabolic pathway, with the genes and proteins of interest displayed, is exhibited in Figure 18 (48).

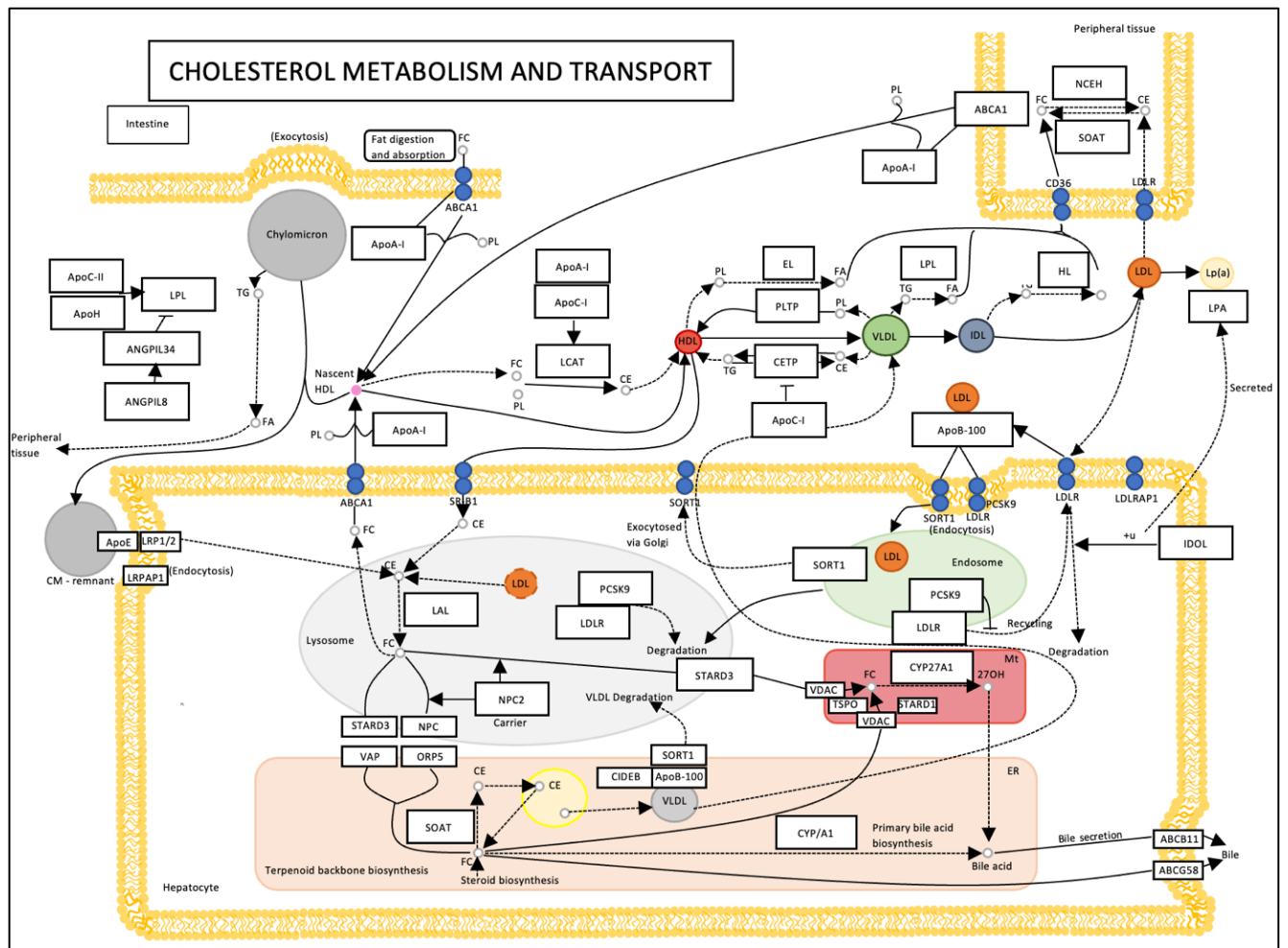


Figure 18 The cholesterol metabolism pathway

Illustration of the cholesterol metabolism pathway found in homo sapiens adapted from the cholesterol metabolism pathway modelled by the KEGG Pathway Database (48).

Overall, it seems that serdemetan has a lysomotrohic effect, inhibiting cholesterol trafficking and transport to force cholesterol accumulation and resultant apoptosis due to a build of cytotoxicity (31). Next, it was important to determine which pathways were being modified, and how exactly variants were being introduced during, during serdemetan resistance formation.

4.1.2 Functional categories affected as identified by KEGG Mapper

After completing research into the pathways expected to be modified during resistance formation, it was time to return focus to the serdemetan-resistant sublines of this study and see exactly where variants had been introduced. The first part of this investigation involved

determining the functional categories affected in the sublines, as identified and defined by KEGG Mapper. According to the description of KEGG mapper in the database “KEGG mapping is the process to map molecular objects (genes, proteins, small molecules, etc.) to molecular interaction/reaction/relation networks (KEGG pathway maps, BRITE hierarchies and KEGG modules). It is not simply an enrichment process; rather it is a set operation to generate a new set” (69). Taking the subset of mapped genes generated by KEGG mapper, heatmaps displaying the number of variants affected per functional category (as defined by KEGG Mapper) for the four different resistant sublines were created using Python language for all categories defined (Figure 19). The functional hierarchy KEGG Orthology (KO) is a database of functional orthologs manually defined in context of the KEGG molecular networks (70). As such, this category does not hold relevance for the analysis in this study and will not be discussed further.

A

		Number of genes with de novo high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants																											
Subline I		0	1	0	0	5	2	0	0	3	5	2	1	2	0	0	0	27	0	0	0	2	0	0	2	1	0	0	0
Subline IX		0	0	0	1	9	4	2	0	2	19	3	3	3	1	1	4	63	1	0	7	1	0	0	5	1	2	0	1
Subline V		0	2	4	1	6	6	5	1	3	23	4	8	1	0	2	3	80	1	1	7	2	0	0	4	0	2	4	0
Subline VIII		2	2	0	0	6	5	2	1	0	15	5	1	2	0	1	1	46	1	0	4	3	1	1	1	0	1	0	0
	Amino acid related enzymes																												
	CD molecules																												
	Cell adhesion molecules																												
	Chaperones and folding catalysts																												
	Chromosome and associated proteins																												
	Cytoskeleton proteins																												
	DNA repair and recombination proteins																												
	DNA replication proteins																												
	Domain-containing proteins not elsewhere classified																												
	Enzymes																												
	Exosome																												
	G protein-coupled receptors																												
	Glycosaminoglycan binding proteins																												
	Glycosylphosphatidylinositol (GPI)-anchored proteins																												
	Glycosyltransferases																												
	Ion channels																												
	KEGG Orthology (KO)																												
	Lectins																												
	Lipid biosynthesis proteins																												
	Membrane trafficking																												
	Messenger RNA biogenesis																												
	Mitochondrial biogenesis																												
	Pattern recognition receptors																												
	Peptidases and inhibitors																												
	Proteasome																												
	Protein kinases																												
	Protein phosphatases and associated proteins																												
	Proteoglycans																												
	Ribosome																												
	Ribosome biogenesis																												
	Spliceosome																												
	Transcription factors																												
	Transcription machinery																												
	Transfer RNA biogenesis																												
	Translation factors																												
	Transporters																												
	Ubiquitin system																												
Brite functional hierarchy																													

Brite functional hierarchy

B

		Number of genes with gained high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants																									
Subline I		0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Subline IX		3	1	3	1	1	5	3	1	1	1	1	1	1	34	4	2	1	1	0	2	6	1	1	2	2	2
Subline V		2	0	0	0	0	2	0	0	0	0	1	0	0	11	2	0	2	0	1	0	1	0	0	1	0	0
Subline VIII		1	0	0	0	0	1	0	1	0	0	0	0	0	3	1	0	1	0	0	0	0	0	0	0	0	0
	CD molecules																										
	Cell adhesion molecules																										
	Chromosome and associated proteins																										
	Cytoskeleton proteins																										
	DNA repair and recombination proteins																										
	Enzymes																										
	Exosome																										
	G protein-coupled receptors																										
	GTP-binding proteins																										
	Glycosaminoglycan binding proteins																										
	Glycosyltransferases																										
	Ion channels																										
	KEGG Orthology (KO)																										
	Membrane trafficking																										
	Messenger RNA biogenesis																										
	Protein kinases																										
	Protein phosphatases and associated proteins																										
	Ribosome biogenesis																										
	Spliceosome																										
	Transcription factors																										
	Transcription machinery																										
	Transfer RNA biogenesis																										
	Transporters																										
	Ubiquitin system																										

Brite functional hierarchy

C

		Number of genes with lost high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants																											
Subline I		1	1	0	1	1	1	1	10	4	5	3	4	40	1	1	4	0	0	1	4	1	2	0	2	1	1	3	
Subline IX		0	1	1	6	0	1	1	3	14	6	3	3	4	52	1	1	3	0	2	1	3	1	2	1	4	1	1	3
Subline V		0	1	0	2	0	1	3	1	11	3	3	3	3	35	1	1	2	0	0	1	3	1	1	0	2	1	1	2
Subline VIII		0	1	0	0	0	1	2	1	9	3	3	3	4	35	1	1	3	1	0	1	3	1	1	1	2	1	1	2
		CD molecules	Cell adhesion molecules	Chaperones and folding catalysts	Chromosome and associated proteins	Cytokine receptors	Cytoskeleton proteins	DNA repair and recombination proteins	Domain-containing proteins not elsewhere classified	Enzymes	Exosome	G protein-coupled receptors	Glycosaminoglycan binding proteins	Ion channels	KEGG Orthology (KO)	Lectins	Lipid biosynthesis proteins	Membrane trafficking	Messenger RNA biogenesis	Mitochondrial biogenesis	Nuclear receptors	Peptidases and inhibitors	Protein kinases	Protein phosphatases and associated proteins	Spliceosome	Transcription factors	Transcription machinery	Transporters	Ubiquitin system
		Brite functional hierarchy																											

D

		Number of genes with not called high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants																									
Subline I		1	1	7	3	1	5	9	5	2	0	1	1	38	5	3	2	1	0	0	2	0	2	2	1	1	
Subline IX		0	0	2	1	0	1	1	0	1	0	0	0	5	0	1	0	0	0	0	0	0	1	0	0	0	
Subline V		0	0	1	0	0	1	2	0	2	0	0	2	11	0	2	1	0	0	0	0	1	2	0	0	1	
Subline VIII		0	0	0	2	0	2	3	2	1	1	0	3	16	2	2	0	0	1	1	0	0	1	0	0	1	
		CD molecules	Cell adhesion molecules	Chromosome and associated proteins	DNA repair and recombination proteins	DNA replication proteins	Domain-containing proteins not elsewhere classified	Enzymes	Exosome	G protein-coupled receptors	GTP-binding proteins	Glycosaminoglycan binding proteins	Ion channels	KEGG Orthology (KO)	Membrane trafficking	Messenger RNA biogenesis	Mitochondrial biogenesis	Peptidases and inhibitors	Protein kinases	Protein phosphatases and associated proteins	Proteoglycans	Ribosome biogenesis	Spliceosome	Transcription factors	Transporters	Ubiquitin system	
		Brite functional hierarchy																									

E

Number of genes with higher allelic ratio high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants

Subline I	0	0	0	1	1	2	0	0	1	0	0	0	0
Subline IX	0	1	0	2	0	4	0	1	0	1	0	1	1
Subline V	1	0	0	2	0	5	1	0	0	1	1	1	0
Subline VIII	0	0	1	1	0	3	0	0	0	1	1	0	0
	Cytoskeleton proteins	DNA repair and recombination proteins	Domain-containing proteins not elsewhere classified	Enzymes	Ion channels	KEGG Orthology (KO)	Messenger RNA biogenesis	Peptidases and inhibitors	Protein kinases	Protein phosphatases and associated proteins	Spliceosome	Transcription factors	Ubiquitin system

Brite functional hierarchy

F

Number of genes with lower allelic ratio high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants

Subline I	0	0	0	0	1	0	0	1
Subline IX	0	0	0	0	1	0	0	0
Subline V	0	0	1	1	1	0	0	0
Subline VIII	1	1	0	0	2	1	1	0
	Cytoskeleton proteins	Enzymes	Exosome	GTP-binding proteins	KEGG Orthology (KO)	Membrane trafficking	Protein kinases	Spliceosome

Brite functional hierarchy

1

	Number of genes with same high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants																															
Subline I	0	1	0	10	0	0	4	2	0	1	19	2	2	0	1	3	67	1	7	3	7	5	5	3	0	2	5	1	1	2	6	
Subline IX	1	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	3	0	1	0	0	0	1	0	0	0	0	0	0	0	0	
Subline V	4	5	2	30	1	2	12	7	1	11	48	12	9	1	2	10	183	1	20	7	11	9	9	5	2	3	13	4	1	6	12	
Subline VIII	2	3	0	28	1	1	8	5	0	6	36	6	7	1	2	9	135	1	11	4	9	8	7	4	1	3	13	4	1	2	9	
	CD molecules	Cell adhesion molecules	Chaperones and folding catalysts	Chromosome and associated proteins	Cytochrome P450	Cytokine receptors	Cytoskeleton proteins	DNA repair and recombination proteins	DNA replication proteins	Domain-containing proteins not elsewhere classified	Enzymes	Exosome	G protein-coupled receptors	Glycosaminoglycan binding proteins	Glycosyltransferases	Ion channels	KEGG Orthology (KO)	Lipid biosynthesis proteins	Membrane trafficking	Messenger RNA biogenesis	Mitochondrial biogenesis	Peptidases and inhibitors	Protein kinases	Protein phosphatases and associated proteins	Proteoglycans	Spliceosome	Transcription factors	Transcription machinery	Transfer RNA biogenesis	Transporters	Ubiquitin system	
	Brite functional hierarchy																															

Figure 19 Functional categories affected by serdemetan resistance as identified by KEGG MAPPER

Number of genes with high-quality, high-coverage, somatic-like, protein-coding, most likely damaging variants for the different categories: de novo (A), gained (B), lost (C), not called (D), higher allelic ratio (E), lower allelic ratio (F), and same (G). These were mapped according to the KEGG BRITE database functional hierarchies across the serdemetan-adapted UKF-NB-3 sublines I, IX, V, and VIII.

Genes containing de novo high-quality, high-coverage, somatic-like, protein-coding, most likely damaging variants were mapped to 37 Brite functional hierarchies. Upon analysis of the de novo heatmap (Figure 19A), the functional hierarchies containing the highest number of variants include: enzymes, membrane trafficking, chromosome and associated proteins, transcription factors, cytoskeleton proteins, exosome and G protein-coupled receptors. The functional category enzymes had the greatest number of variants with a total of 62. This is to be expected as mutations altering enzymes are likely to have a large effect on the pathway functionality – which is important for resistance to form. Additionally this category is quite broad, and therefore should include a large number of the variants - even those affecting different pathways. Functional hierarchies that appear quite significant to this study's analyses are chromosome and associated proteins, transcription factors, membrane trafficking, and G protein-coupled receptors. Should resistance originate from variants

forming in the MDM2/p53 signalling pathway – the functional categories expected to be flagged would be transcription factors and chromosome and associated proteins. Since there are many transcription factors indicated in this pathway, including p53 itself, modifications to this pathway would likely affect some of these proteins. Some of these transcription factors include: Dmp1, E2F1, YB-1, and YY1, as specified in previous literature (62). Additionally, if the resistance developed originates from variants in the cholesterol transport pathway – the functional categories: membrane trafficking, exosomes and G protein-coupled receptors would be expected to contain multiple variant hits – as these are categories cholesterol transporters, such as ABCA1 and ABCG1, would fall under. These results are promising and hint that the pathways suggested to be part of serdemetan’s mechanism of action in previous literature are likely to be accurate. When analysing the other categories of variants, specifically gained (Figure 19B), lost (Figure 19C), and not called (Figure 19D), the general trends from the de novo heatmap are maintained in terms of the functional categories with the most variants. The main different is the quantity of variants observed here, with the lost and de novo variant categories having the majority. Very few variants were identified in the higher allelic and lower allelic ratio categories so it is difficult to detect any trends in these groupings. The final category investigated was the variants that were detected at the same frequency (Figure 19G) in the resistant sublines as in the chemo sensitive UKF-NB-3 parental cell line. In this grouping there is still overlap with the other categories, with the same functional hierarchies mutated being: enzymes, exosome, chromosome and associated proteins, cytoskeleton proteins, G protein-coupled receptors, membrane trafficking and transcription factors. As these are the functional categories likely to be most involved in serdemetan’s mechanism of action, it is not surprising to have them be highly mutated even prior to resistance development. However, there were some differences in the same category, with new functional hierarchies showing higher levels of variants. These include: domain-

containing proteins not elsewhere specified, ion channels, mitochondrial biogenesis, peptidases and inhibitors, protein kinases, ubiquitin system. These classes of variants are expected with the development of neuroblastoma – especially after exposure to the MDM2 inhibitor serdemetan. Following this, clustermaps displaying the quantity of high-quality, high-coverage, somatic-like, protein-coding, most-likely damaging variants per pathway (as defined by KEGG Mapper) for the different categories defined earlier were created. These figures were analysed as an introductory inquiry into the pathways altered as serdemetan resistance forms and will be discussed in the next section.

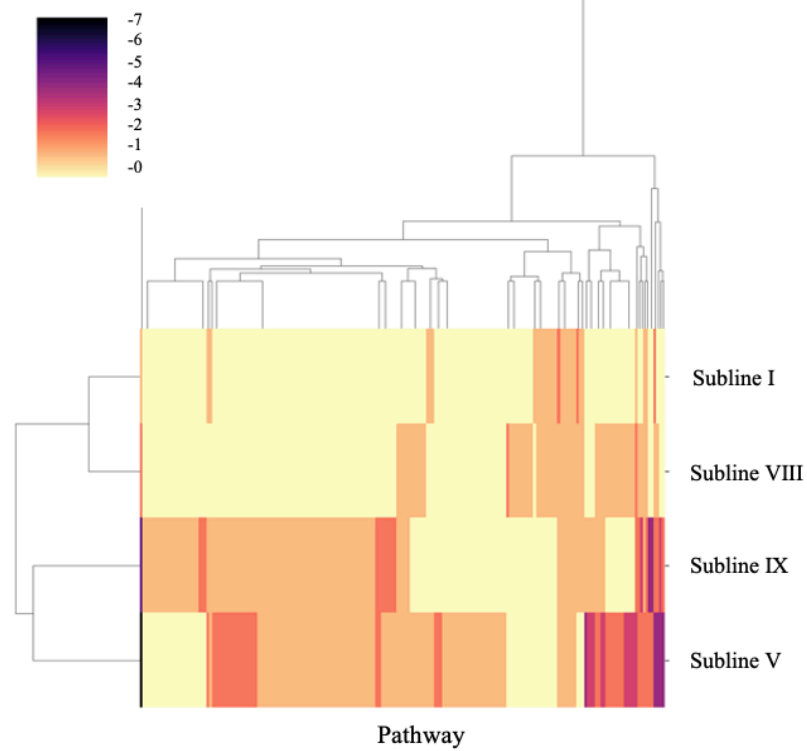
4.1.3 Hierarchical clustering of common pathways

The first step undertaken when examining the pathways mutated during serdemetan resistance development in the UKF-NB-3 sublines of this study was to create hierarchical clustermaps (Figure 20). These were constructed using python language in relation to the number of genes with high-quality, high-coverage, somatic-like, protein-coding, most likely damaging variants that were mapped to specific KEGG pathways/Brite functional hierarchies for the different variant categories defined. The results of this section corroborated what was found in the previous chapter, with sublines IX and V containing the greatest number of de novo variants and displaying very similar mutational profiles with regards to the pathways altered. As previously encountered, sublines I and VIII also had a great degree of similarity in their de novo clustermaps (Figure 20A) – with lower levels of mutations than the other two sublines. This suggests that the resistance mechanism developed in the IX-V subline pair differs from the one that developed in the I-VIII subline pair. Upon zooming out and looking at the variants in the same category (Figure 20G), the clustermaps begin to look more similar across the sublines. This is likely the case because the sublines originated from the same parental UKF-NB-3 cell line and only diverged from each other after serdemetan treatment ensued. The two groupings of the sublines could have, as earlier discussed, developed from

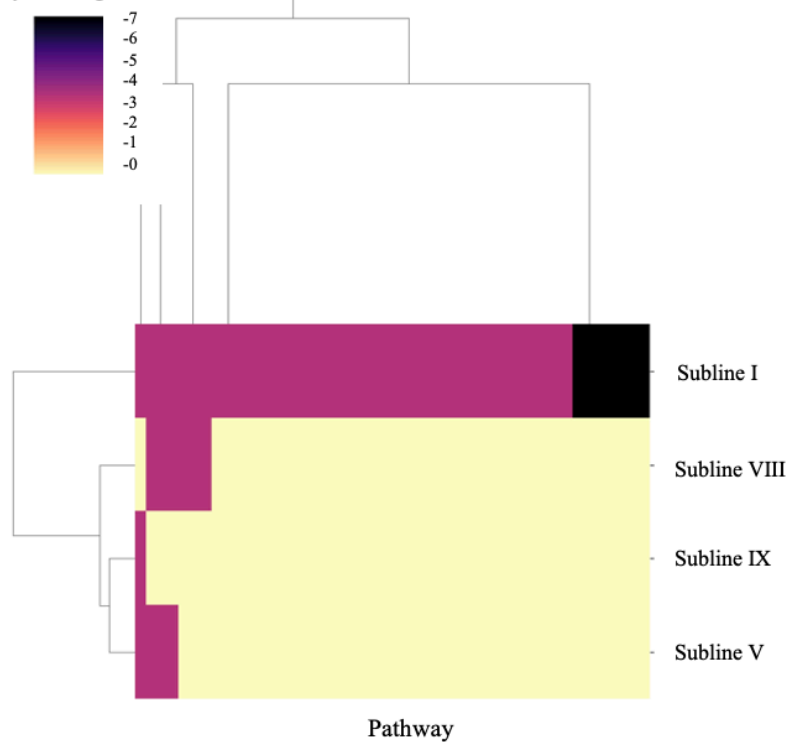
two distinct subpopulations of resistant cells being selected for during treatment – explaining the similarity between the sublines pairs. These would have further mutated apart during treatment but the same critical resistance variants would be retained in the subline groups. Few pathways appear affected in the gained category clustermap – and only subline IX having a notable number of variants (Figure 20B). This lack of variants carries through to the higher allelic ratio and lower allelic ratio clustermaps – and confirms what was visualized with the functional category heatmaps (Figures 20E and F). When looking at the clustermaps for category lost and not called, again there can be seen a notable number of variants and pathways affected (Figures 20C and D). This could either imply that resistance might have developed from the loss of particular variants or that there are several pathways that contain non critical variants that were not retained during serdemetan treatment simply due to chance. The second option appears likely, as the variants lost in the pathways of the sublines overlap substantially – without the separation between the two groups of sublines observed elsewhere. However, only analysing exactly which pathways have been affected in each subline across all the categories will reveal what is critically being mutated as resistance to serdemetan forms, leading this study into the next phase of research.

A

Number of genes with de novo
high-quality, high-coverage,
somatic-like, protein-coding,
most-likely damaging variants

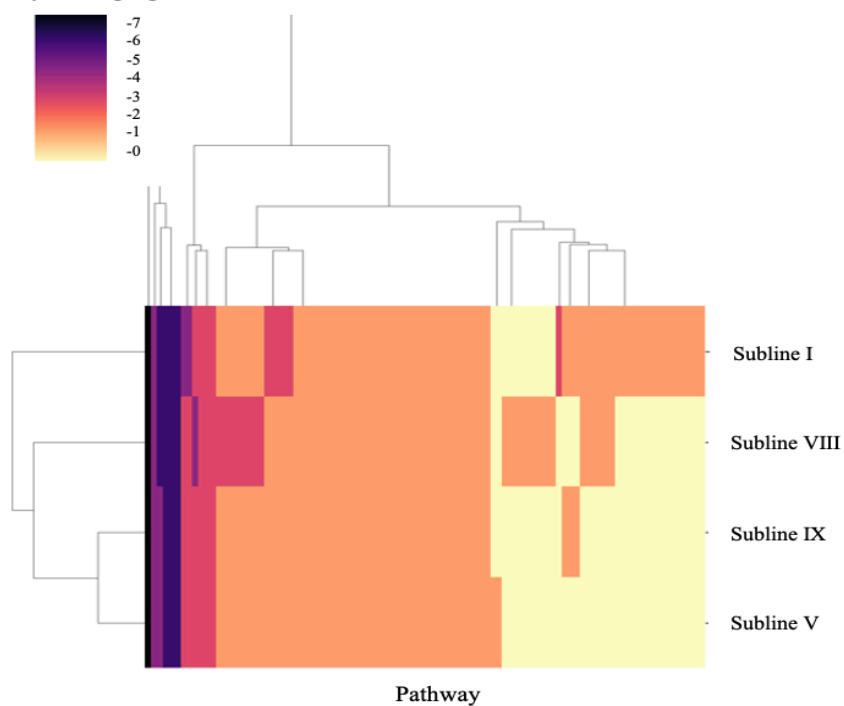
**B**

Number of genes with de novo
high-quality, high-coverage,
somatic-like, protein-coding,
most-likely damaging variants

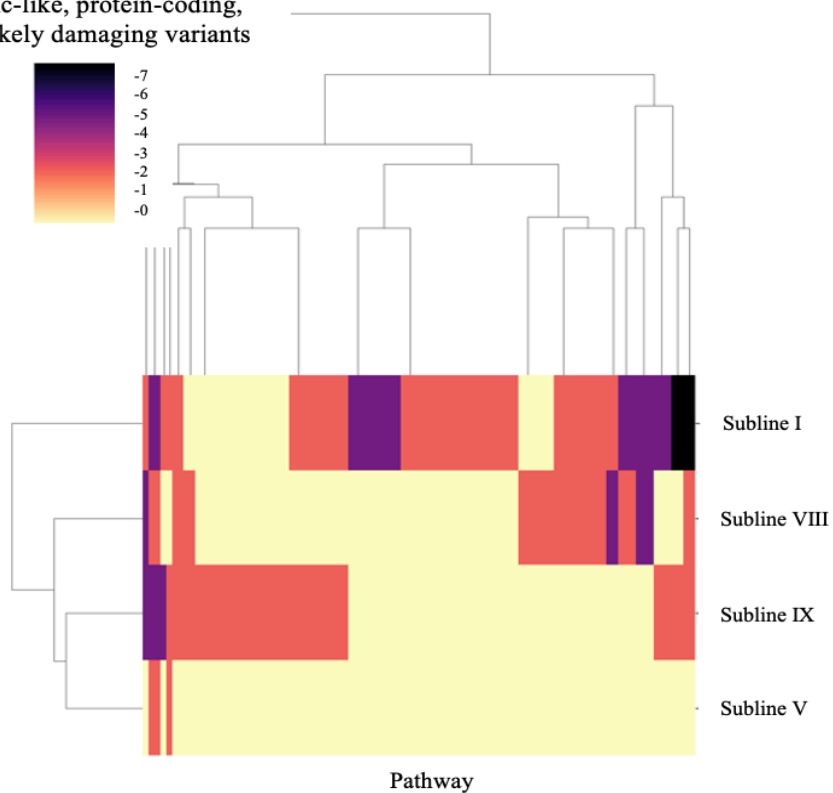


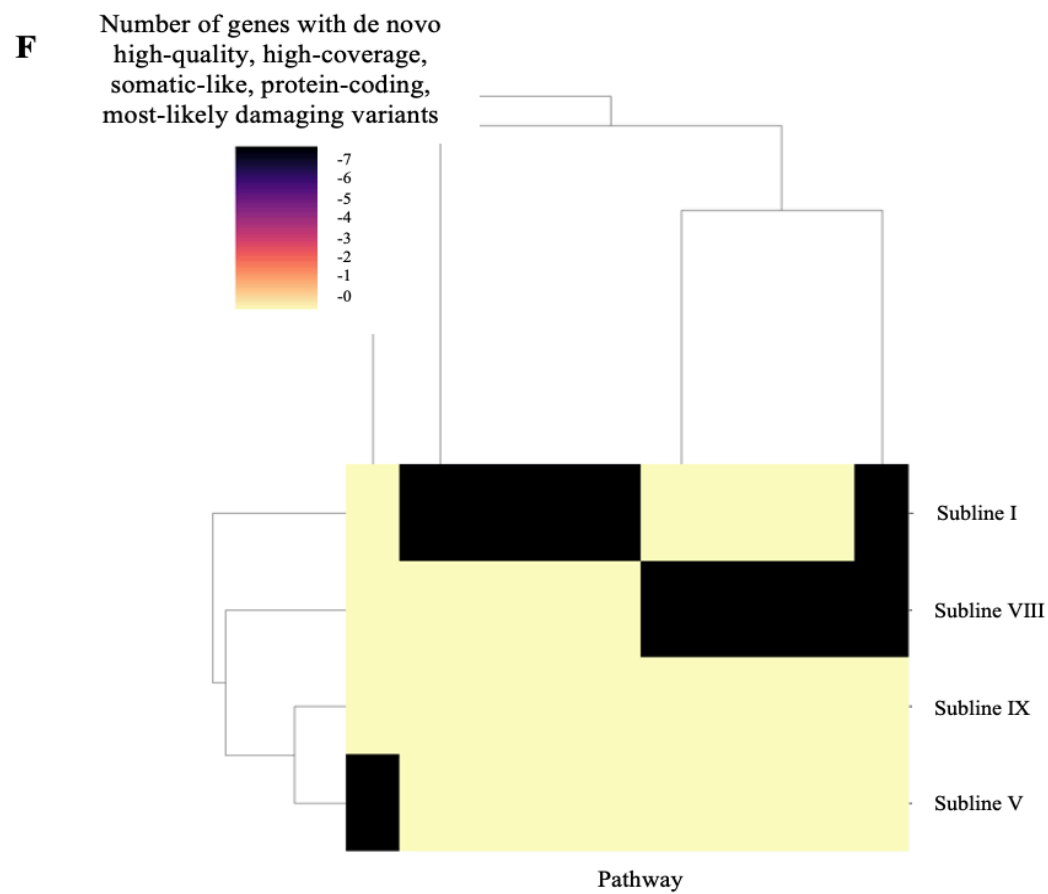
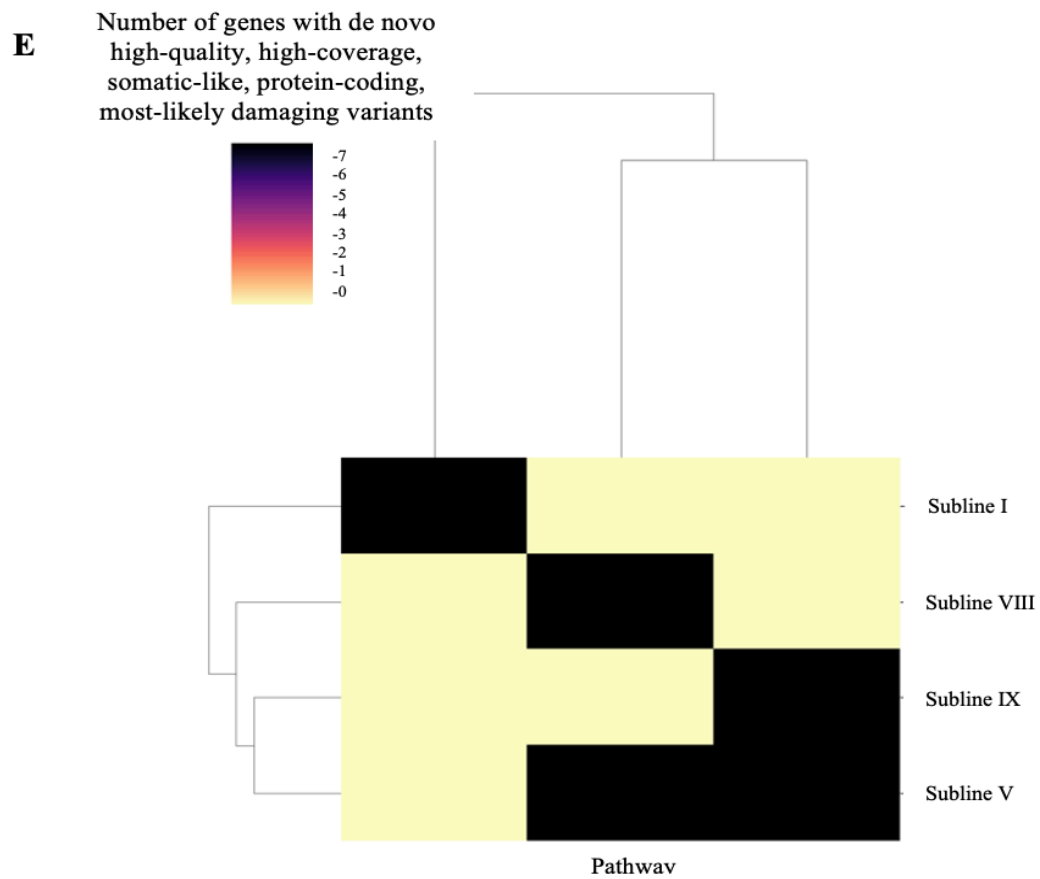
C

Number of genes with de novo
high-quality, high-coverage,
somatic-like, protein-coding,
most-likely damaging variants

**D**

Number of genes with de novo
high-quality, high-coverage,
somatic-like, protein-coding,
most-likely damaging variants





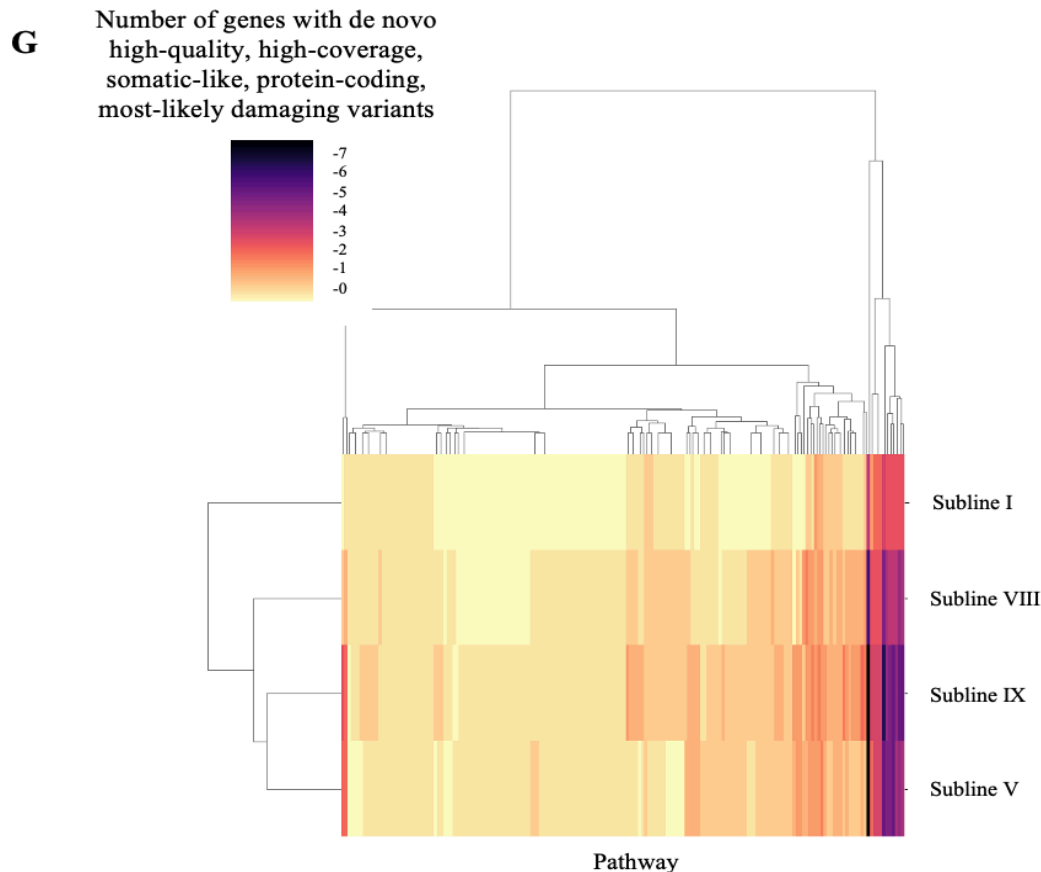


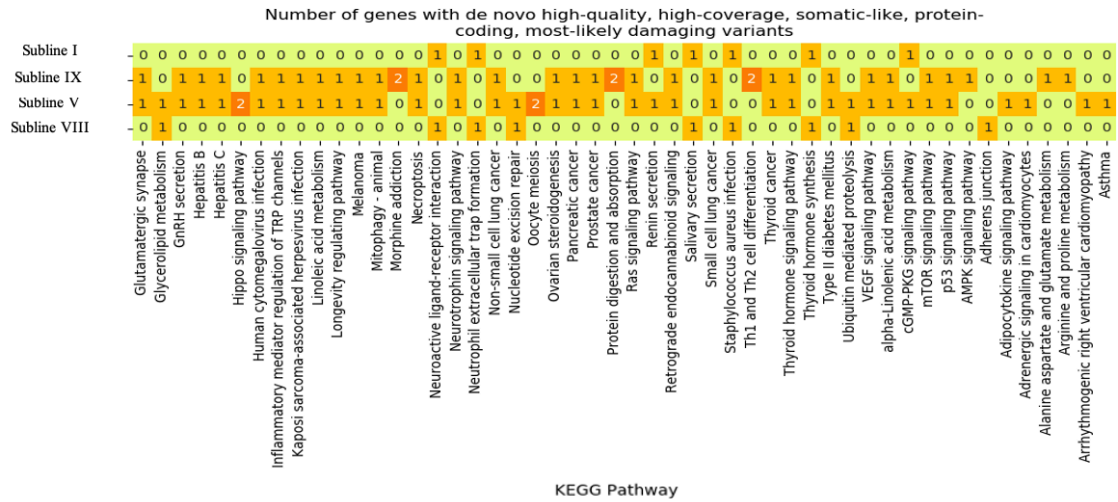
Figure 20 Hierarchical clustering of serdemetan resistance UKF-NB-3 sublines shared variants mapped to KEGG Pathways

Clustermaps displaying the hierarchical clustering of high-quality, high-coverage, somatic-like, protein-coding, most likely damaging variants according to their pathways as defined and identified by the KEGG Mapper Database. Clustermaps were made for categories de novo (A), gained (B), lost (C), not called (D), higher allelic ratio (E), lower allelic ratio (F), and same (G) displaying variants found across all four UKF-NB-3 serdemetan-resistant sublines.

4.1.4 Pathways affected as identified by KEGG Mapper

To investigate which pathways were most altered during serdemetan resistance development, pathways containing de novo, high-quality, high-coverage, somatic-like, protein-coding, most likely damaging variants for each of the serdemetan resistant sublines were identified by the KEGG Mapper tool. Similarly to the results generated and displayed for Figure 19, identifying the variant functional categories – the genes distinguished for each pathway are the result of a pathway enrichment analysis, meaning the genes mapping to the pathway are more than those expected by chance alone. The results from this process are displayed in heatmaps generated using python language (Figure 21).

C



D

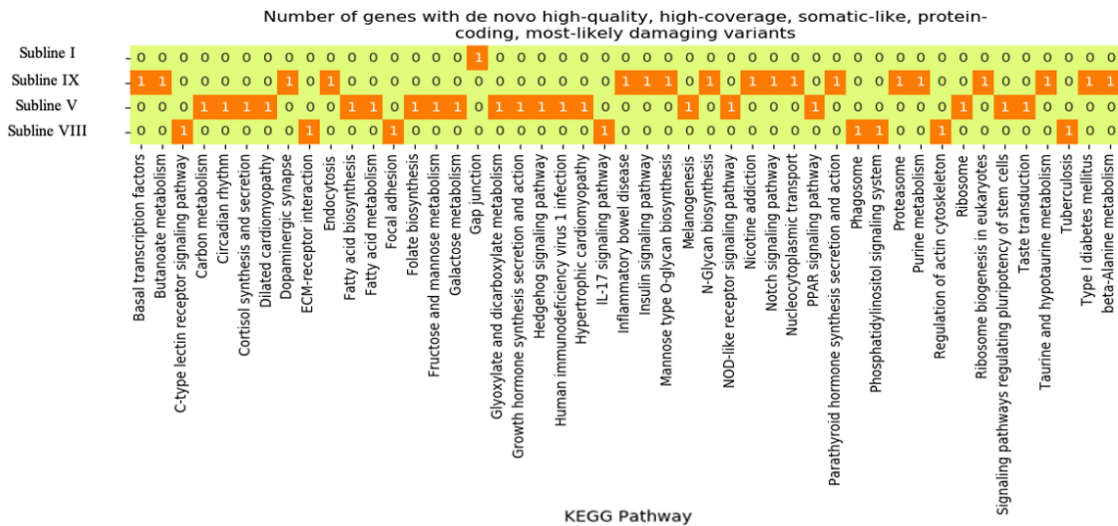


Figure 21 Pathways affected by serdemetan resistance as identified by KEGG MAPPER

Heatmaps displaying the number of genes with de novo, high-quality, high-coverage, somatic-like, protein-coding, most likely damaging variants mapped according to pathways identified by the KEGG Mapper tool across the serdemetan-adapted UKF-NB-3 sublines I, IX, V, and VIII. Pathways were divided into groups of 50, with those containing the highest number of variants in Figure 21A) followed by B), C), and then E).

Genes containing de novo high-quality, high-coverage, somatic-like, protein-coding, most likely damaging variants were mapped to 197 KEGG pathways (Figure 21 A, B, C and D).

For greater clarity, the results from KEGG Mapper were separated into four separate heatmaps, with the first fifty most mutated pathways displayed in Figure 21A, moving across

to those with the least variants displayed in Figure 21D, respectfully. When analysing the heatmaps, the relationship between the two groups of resistant subline pairs becomes more apparent – with the subline pair IX and V overlapping with 103 out of the 197 total pathways detected and subline I sharing 23 of its 29 flagged pathways with subline VIII. This reinforces the relationship between the subline pairs suggested in earlier chapters. It is interesting to note that the most effected pathways appear to be those involved in metabolism, with fifteen de novo variants in this pathway across the sublines. This is something that previous studies implicating the MDM2-HIF1 α axis pointed to as possibly being affected by the p53-independent portion of serdemetan’s mechanism of action (35). Should resistance originate from alterations to the MDM2-HIF α axis – there would be an expected raised number of metabolic pathway variants due to the vast number of signalling pathways triggered by this cascade. Additionally, this is defined as a global pathway map on KEGG Mapper and as such, broadly incorporates many other pathways under its umbrella – including the cholesterol metabolism pathway discussed earlier. This broad inclusion of other pathways further explains the high number of variants observed for this category. Upon further investigation of which variants were responsible for flagging genes in this category – none appear to be a part of significant pathways (as defined by previous literature) or overlapping between the different resistant neuroblastoma sublines. While some substantially mutated pathways, such as olfactory transduction and human papillomavirus infection, contain more gene variants than the next pathways discussed, these do not seem to be significant in resistant formation based on what is known about serdemetan. Therefore these pathways will not be discussed further.

The pathways most notably affected were decided according to the research previously conducted for serdemetan’s mechanism of action, the quantity of variants detected in the

pathway, and the degree of overlap between the resistant UKF-NB-3 sublines. These pathways were organized into three main categories: 1) those implicated in cholesterol and metabolism, 2) those implicated with cancer and angiogenesis, and 3) those matching neither criterion. Those implicated in cholesterol and metabolism include: chemical carcinogenesis – receptor activation, bile secretion, cholesterol metabolism, fat digestion and absorption, ABC transporters, Regulation of lipolysis in adipocytes, and lipid and atherosclerosis. Those implicated with cancer and angiogenesis include: cell cycle, PI3K-AKT signalling pathway, transcriptional misregulation in cancer, apoptosis, VEGF signalling pathway, p53 signalling pathway, mTOR signalling pathway, chemical carcinogenesis – DNA adducts, chemical carcinogenesis – receptor activation, and pathways in cancer. Those matching neither criterion include: choline metabolism in cancer, glycerophospholipid metabolism, cAMP signalling pathway, drug metabolism – cytochrome P450, drug metabolism – other enzymes, and TH17 cell differentiation.

When analysing the genes flagged in these pathways, a few stand out as significant for serdemetan resistance formation. The first de novo variant group of interest is the missense and stop gained variants altering the structural component of cholesterol (LDL) carrier lipoproteins: apolipoprotein B (APOB) gene (36,71). Note that apolipoprotein B100 (APOB-100) has been defined as synonymous with APOB and will be used interchangeably (71,72). Apolipoproteins are responsible for stabilization of lipoproteins and form the contact element for cellular receptors (36). Variants of this gene are responsible for flagging the pathways: cholesterol metabolism, fat digestion and absorption, vitamin digestion and absorption (as cholesterol is a precursor molecule to vitamin D), and lipid and atherosclerosis (36). Produced in the liver, APOB-100 has a requisite role in the construction of triglyceride-rich very low density lipoproteins (VLDL) – with each nascent VLDL particle containing a

molecule of APOB (71,73). VLDL particles are spherical macromolecule lipoproteins that allow cholesterol to be transported through plasma due to their hydrophilic coating (36). The APOB-100 fragment of the VLDL particles serves as the ligand for the LDL receptor, which mediates the transport of approximately three quarters of LDL from the plasma into the required tissues (71,73). Where and how this protein fits in the cholesterol metabolic pathway can be seen in Figure 18. This was the only protein that was consistently mutated across all the UKF-NB-3 resistant-sublines of this study –highlighting the possibility that mutagenesis of this protein is a requisite to serdemetan resistance. According to former studies, when previously mutated the gene variants often result in truncation of the protein, overall giving rise to very low LDL plasma concentrations and the condition of hypobetalipoproteinaemia (71). This is theorised as being due to low secretion rates of truncated APOB proteins or rapid clearance rates of the mutated protein from the plasma (71). However, there has also been evidence of specific APOB-100 mutations, those impacting the ability of LDL to bind to LDL receptors, being capable of inducing hypercholesterolaemia and premature atherosclerosis (71). In the next section, investigations into exactly where and what mutations are occurring to the APOB proteins of the studied sublines will be explored, as well as how these might affect the protein's functionality to bring about resistance.

Another protein group that was found mutated in some of the resistant sublines, and has been implicated in cholesterol metabolism, was the ABC transporters (31). The ABC genes are divided into seven separate subfamilies (designated A through G) (74). ATP-binding cassette subfamily A member 10 (ABCA10), a cholesterol-responsive gene and member of the ABCA transporter subfamily suggested to be involved in macrophage lipid homeostasis, was found mutated in subline IX (75). The majority of ABCA proteins appear to have some involvement with lipid trafficking across cell types – so there is a high probability of this protein being

involved in cholesterol transport (74). ATP-binding cassette subfamily D member 2 (ABCD2), a member of the ABCD subfamily involved in peroxisomal import of fatty acids/fatty acyl-CoAs, was found mutated in subline V (74,76). There are four identified ABC proteins included in subfamily D (1-4) – a subfamily of half-sized ABC transporters existent as homodimers (74,76). Considering the prominent role the ABCA1 transporters play in the cholesterol transport pathway during serdemetan resistance development – and the close relation of ABCA10 to ABCA1, it will be important to properly assess the impact of the mutations to this gene. The ABC transporter, ABCD2, does not appear to be linked to the cholesterol pathway or other pathways implicated in serdemetan mechanism of action. Therefore, only the variants for ABCA10 will be explored to a greater degree.

The next gene discussed, TP53, was flagged in multiple pathways, contained overlaps between two of the serdemetan-resistant sublines, and has been highlighted many times in previous literature as a crucial element affected during serdemetan's mechanism of action – with frequent mutations detected in relapsed neuroblastoma patients. Variants to this gene were flagged in the following pathways by KEGG Mapper: p53 signalling pathway, cell cycle, PI3K-AKT signalling pathway, transcriptional misregulation in cancer, apoptosis, lipid and atherosclerosis, cellular senescence and pathways in cancer. Though only two of the sublines, IX and V, contained variants for this gene, the significance of variants being introduced in this pathway cannot be ignored – especially due to how integrally entangled this protein with cancer regulation. Because the p53 signalling pathway has been the most widely confirmed pathway involved in serdemetan's mechanism of action, any de novo variants introduced to TP53 must be further investigated to determine their structural impact on the protein and if they might alter serdemetan's efficacy. Additionally, the possibility of

resistance developing through distinct mechanisms cannot be ruled out and therefore pathway overlaps between all four resistant UKF-NB-3 sublines is not required for further analysis.

Another gene commonly mutated in the subline pair IX and V is phospholipase A2 group IVF (PLA2G4F). Variants to this gene were flagged in the following pathways on KEGG Mapper: VEGF signalling pathway, choline metabolism in cancer, and glycerophospholipid metabolism. According to previous literature, phospholipase A2 is Ca^{2+} dependent enzyme that hydrolyses the sn-2 ester bond in glycerophospholipids (77). Phospholipids, along with apolipoproteins, are one of the main components of lipoproteins – helping to provide the hydrophilic coat that allows for transportation of hydrophobic cholesterol and triglycerides (36,78). This could suggest some involvement with cholesterol metabolism. This protein has also been identified as part of the VEGF signalling pathway – a pathway previously identified to be part of serdemetan's mechanism of action (35). Therefore, modifications to this protein – depending on where and how they occur – could be involved in serdemetan resistance formation. Further analysis is required to develop a more substantial hypothesis.

The aryl hydrocarbon receptor (AHR) protein is ligand-activated transcription factor historically known as a mediator of chemical toxicity, playing a role in normal vascular development and xenobiotic metabolism (79). This gene was identified as containing variants for the sublines I and VIII in the KEGG Mapper pathways: chemical carcinogenesis – receptor activation and TH17 cell differentiation. As part of its role in xenobiotic metabolism AHR plays an important part in the clearance of foreign chemicals, including drugs, from cells (79). This, in addition to its role in vascular development, make it a suitable candidate for mutations that induce drug resistance. Variants to the AHR gene that cause upregulation of its functions could allow for enhanced clearance of serdemetan from cells (preventing

serdemetan from performing its mechanism of action) as well as enhanced vascular development at tumour sites, promoting angiogenesis. A structural analysis of AHR with the variants will allow for better determination of this theory's accuracy.

Glucuronidation, the addition of glucuronic acid from UDP-glucuronic acid to another molecule, is a chief reaction involved in xenobiotic metabolism/elimination and the biotransformation of hydrophobic molecules, including steroid hormones, into glucuronide derivatives (80,81). The addition of this constituent is theorized to cause steric hinderance of the molecule targeted, eliminating binding affinity for the ligands receptor and leaving the molecule more polar, inactive and readily eliminable from circulation (80). The enzymes responsible for catalysing this reactions are UDP-glucuronosyltransferase (UGT) enzymes (80,81). The genes UGT2B4 and UGT2B7 were found to be mutated in the UKF-NB-3 serdemetan-resistant subline V, while just UGT2B7 was mutated for subline VIII. These genes were flagged as containing variants in the pathways: bile secretion, drug metabolism, and chemical carcinogenesis – DNA adducts. UGT2B4 exhibits glucuronidation activity toward 3R-diol, estriol (E3), and 4-hydroxyestrone while UGT2B7 conjugates a large variety of compounds, including oestrogens, androgens, morphine, AZT, and retinoic acid at position 3R-hydroxy (80,81). Should these genes contain variants able to increase the functionality of these enzymes – the rate of drug clearance would rise, developing into a resistance mechanism for the removal of serdemetan.

While there were variants to other pathways of interest, these were the only variants to be repeated between the serdemetan-resistant sublines for the different pathways identified by KEGG Mapper. However, the pathways affected by these variants (PI3K-AKT signalling pathway, transcription misregulation in cancer, mTOR signalling pathway, choline

metabolism in cancer, glycerophospholipid metabolism, regulation of lipolysis in adipocytes, and cAMP signalling pathway) have been noted and almost all of them appear to be involved in somehow with one of the serdemetan resistance mechanisms outlined earlier in the chapter. This suggests that there might be three partly independent mechanisms through which serdemetan resistance can form: 1) modulation of cholesterol metabolism – specifically with regards to transport, 2) modulation of the activity of p53, and 3) modulation of MDM2's p53-independent downstream signalling targets. In the next section, the genes specifically mentioned above will undergo a preliminary structural analysis in PyMOL.

4.1.5 Investigations into shared variant consequence

To determine if the variants to the genes defined as “significant” were likely to alter protein activity and further play a role in serdemetan resistance formation, renderings of the amino acid mutations were created and analysed in PyMOL. All unmutated 3D protein structures were obtained from the AlphaFold Protein Structure Database. Homo sapien 3D protein structures containing predictions with the highest degree of confidence were selected. These renderings were used to predict how variants might impact protein structure and activity. However, to make more formal conclusions, a molecular dynamics study of the variants would need to be performed. This is beyond the scope of this thesis and is suggested for future studies. Changes to hydrogen-bonding, amino acid properties, variant location and steric hinderance within the protein structure, and the type of mutation introduced were considered when drawing together conclusions. To add an additional layer of analysis, variants were submitted to the variant effect predictor: Mutation effect predictor using the Temporal convolutional network and the Born-Again Networks (MTBAN). This online tool allows for the prediction of the deleteriousness of a single amino acid variant and is based on a deep generative model (82).

The first variants investigated were those affecting the structure of the APOB-100 protein. During this analysis, focus will be placed on the position in the protein coding sequence (CDS) that variants occur. The CDS is the part of a gene that is fully translated into a protein, comprising of the exons of the gene, in addition to start and stop codons (83). The UKF-NB-3 serdemetan resistant sublines I and VIII contained the same missense variants of the APOB-100 protein. These variants altered the same CDS position: 485, with an adenosine base being substituted for a cytosine at this position. This base substitution resulted in an amino acid (aa) substitution of the serine (Ser) residue, originally coded by a UCU codon, to a tyrosine (Tyr) residue, now coded by a UAU codon, at aa position 162 (shown in Figure 22). The fact that the variants observed to this protein are at the exact same positions in these two sublines strongly suggests that the acquired resistance observed for these sublines originates from the same subpopulation of resistant cells in the original tumour that contained this variant.

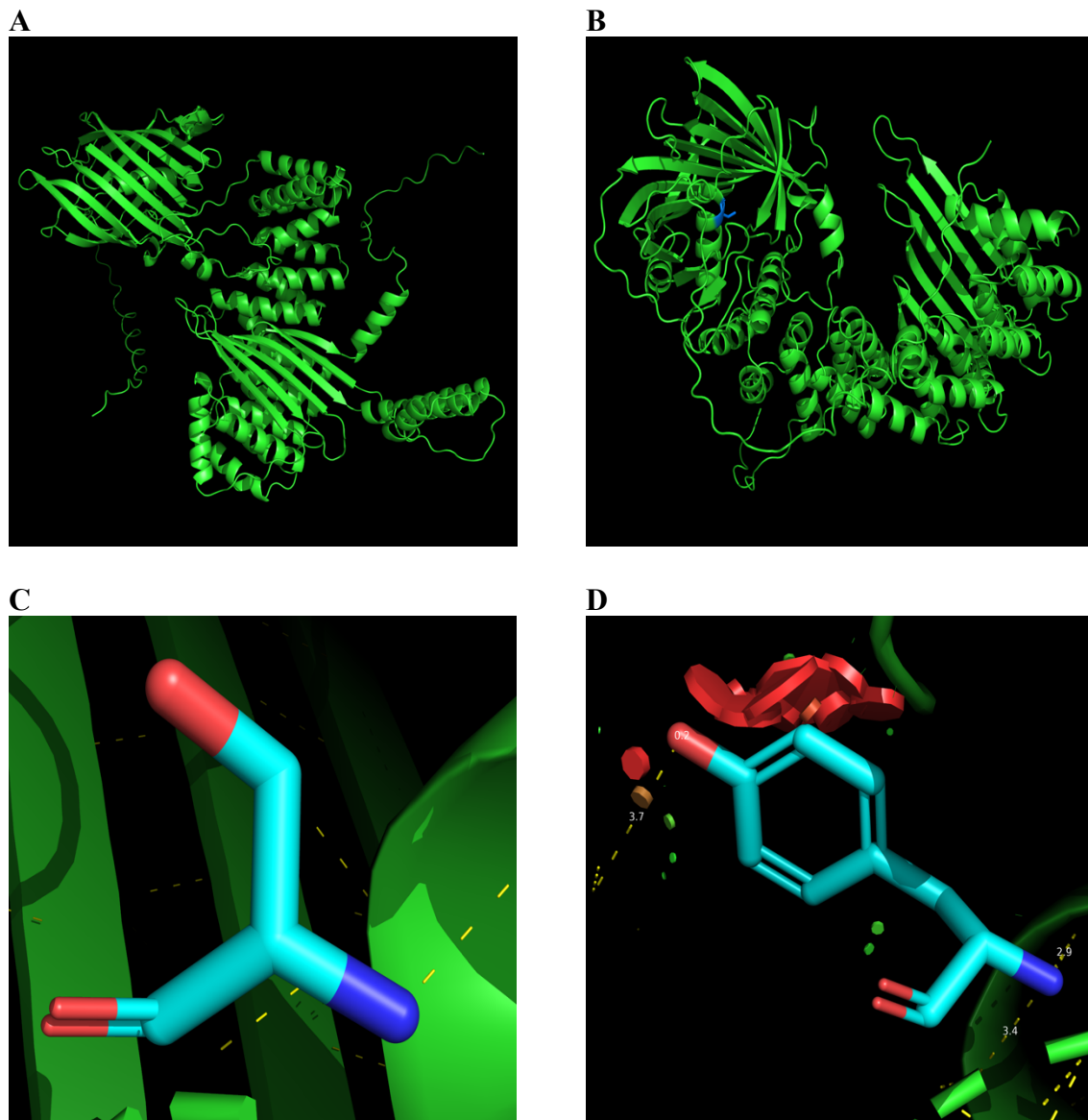


Figure 22 PyMOL renderings of APOB missense variants, S to Y, before and after mutation

PyMOL renderings of the APOB-100 protein in various states including: prior to mutation (A), with the location of residue to be mutated indicated (B), an up close view of the residue prior to mutation (C), and finally the mutated residue (D). Hydrogen bonding and sterically induced clashes are indicated in (C) and (D).

This amino acid substitution is likely a significant one as Ser, an aa with a polar side chain, is being mutated to Tyr, an aa with a hydrophobic side chain. Tyr additionally contains a side chain that is significantly larger than that of Ser. These differences are the basis of the aa's quite different properties, and upon their substitution, should have substantial effects on hydrogen bonding, and therefore protein functionality or stability, and introduce steric hinderance between adjacent aa side chains. When analysing the mutation in PyMOL, the

significance of this aa change becomes clearer. Firstly, there is an additional hydrogen bond added by the inclusion of Tyr in the place of Ser (Figure 22D). This hydrogen bond might decrease protein flexibility necessary for its action. Secondly, due to the large difference in aa side chain sizes, clashes between the mutated residue and those surrounding it have been introduced (Figure 22D). These clashes could destabilize the protein's structure. These things taken together suggest that this variant is likely to be deleterious to the protein's function – and therefore significant. This is further confirmed by the results from variant effect predictor MTBAN displayed in Table 2.

Table 2 Results from the variant effect predictor MTBAN for the S162Y variant in APOB

Uniprot accession: P04114				
Variant	Log probability score	Z-score	Probability of deleteriousness	Predicted label
S162Y	-35.9487	-1.3423	0.864	deleterious

The results of the predictor indicate that the variant is likely to be deleterious, with a probability of 86.4%. Therefore this prompts the conclusion that this variant alters APOB significantly and, because of the known function of this protein, likely infers some kind of resistance to the sublines to serdemetan.

The UKF-NB-3 serdemetan resistant subline IX contained a variant gene coding for the APOB-100 protein at the CDS position 264, where a cytosine base was substituted for an adenosine – introducing a stop gained mutation. The variant, located at aa 88, converted the cysteine (Cys), coded by the codon UGC, into the stop codon UGA (as shown in Figure 23).

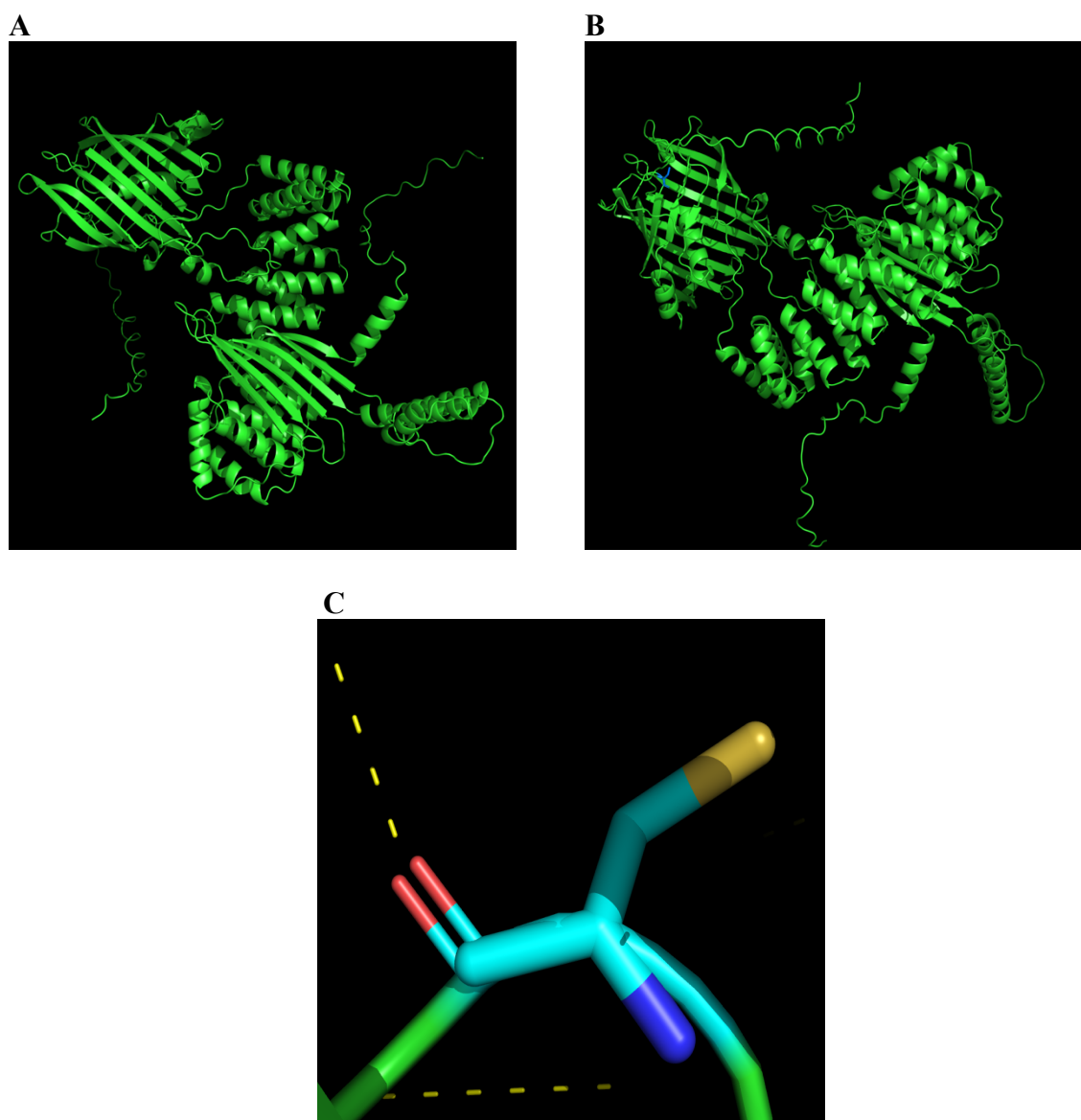


Figure 23 PyMOL renderings of APOB cysteine residue at position 88 prior to mutation

PyMOL renderings of the APOB-100 protein residue 88, cysteine, prior to mutation including the overall protein structure (A), the location of residue to be mutated (B), an up close view of the residue prior to mutation (C) with hydrogen bonding and sterically induced clashes indicated.

This mutation is introduced quite early in the protein's aa sequence, and would likely result in a significantly truncated protein. This would greatly impact how the protein functions, as well as its interactions within its environment – and might introduce some of the effects previously discussed such as increasing APOB-100 clearance rates from plasma and lowering the secretion rates of the truncated proteins. Therefore, this variant can be deemed significant, likely deleterious, and might be actively involved in the formation of resistance to serdemetan. How this fits in with serdemetan's mechanism of action is not clear, and further

studies into the molecular dynamics of this mutation are required. This variant was not submitted to MTBAN as the model only predicts on substitution variants and not deletions.

The final subline, UKF-NB-3 serdemetan resistant subline V, contained one missense variant in the CDS, switching the guanine originally at this position into a cytosine. This changed the codons of this sequence from GUG to CUG, converting the valine (Val) aa coded for at this location into a leucine (Leu) aa in the APOB-100 protein at aa position 1226. Unfortunately this part of APOB's protein structure has not yet been modelled by AlphaFold, preventing the generation of PyMOL renderings of the variant. However based on the properties of the two aa substituted being quite similar, it is difficult to say whether this variant will have a significant or destabilising effect on the function of APOB-100. To determine this probability without being able to model the protein, the variant was submitted to MTBAN. The variant effect predictor determined a probability of 75.23% that the variant introduced here would be deleterious (shown in Table 3). Modelling this mutation once the structure of APOB has been mapped out further will help when making further predictions about this mutation, but at this point, it seems that the mutations introduced to APOB are all deleterious, significant, and likely to be involved in serdemetan resistance development.

Table 3 Results from the variant effect predictor MTBAN for the V1226L variant in APOB

Uniprot accession: P04114				
Variant	Log probability score	Z-score	Probability of deleteriousness	Predicted label
V1226L	-22.384	-0.2001	0.7523	deleterious

The next variants of interest were ones occurring in the ABC transporter protein ABCA10. This gene was only mutated in the subline IX, with a missense variant resulting in the base substitution of guanine for thymine at position 1723 in the CDS. This variation at position 1723 results in the substitution of the aa 575 from a Val residue to a Leu residue, due to the codon change for this amino acid from GUA to UUA (seen in Figure 24).

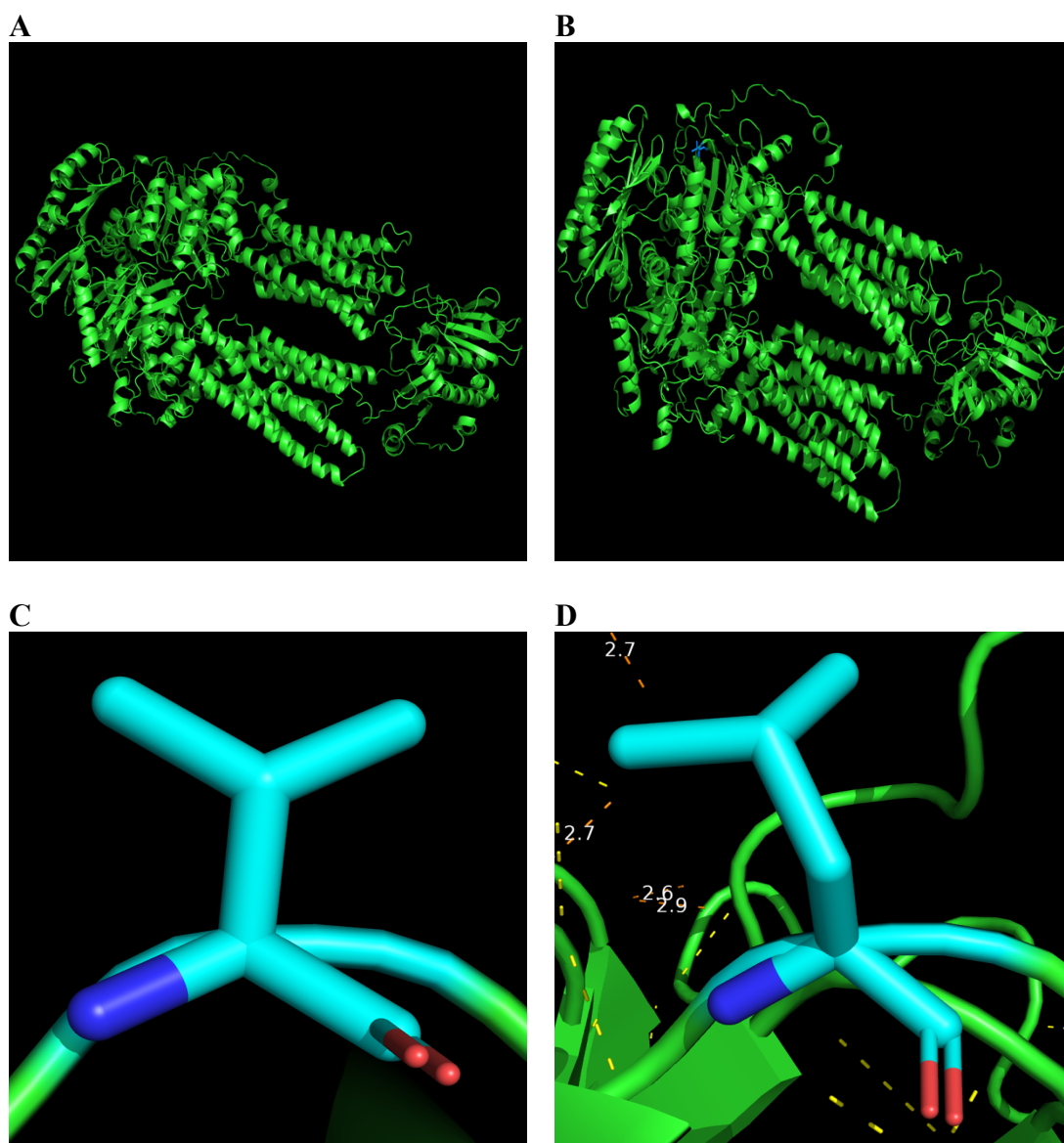


Figure 24 PyMOL renderings of ABCA10 missense variants, V to L, before and after mutation

PyMOL renderings of the ABCA10 transporter protein in various states including: prior to mutation (A), with the location of residue to be mutated indicated (B), an up close view of the residue prior to mutation (C), and finally the mutated residue (D). Hydrogen bonding and sterically induced clashes are indicated in (C) and (D).

As discussed previously, the aa Val and Leu are very similar in their properties and side chain structure – and therefore their substitution is not expected to introduce significant changes to ABCA10’s protein structure. And this is exactly what is seen when mutated in PyMOL. No major clashes are introduced and no hydrogen bonds are broken or formed. Therefore it is unlikely that this variant plays a significant role in serdemetan resistance development.

Though only two of the serdemetan resistant sublines contained mutations for the TP53 gene, the subline pair IX and V, each of these sublines contained multiple variants to the gene. Subline IX contained three TP53 missense variants in the CDS sequence. The first of these occurred at position 527 in the CDS, where a guanine base was substituted for a cytosine base, changing the codon at this position to UCC from UGC. This substitution resulted in aa 176 in the protein being mutated from a Cys to a Ser residue. This mutation can be seen in Figure 25 below.

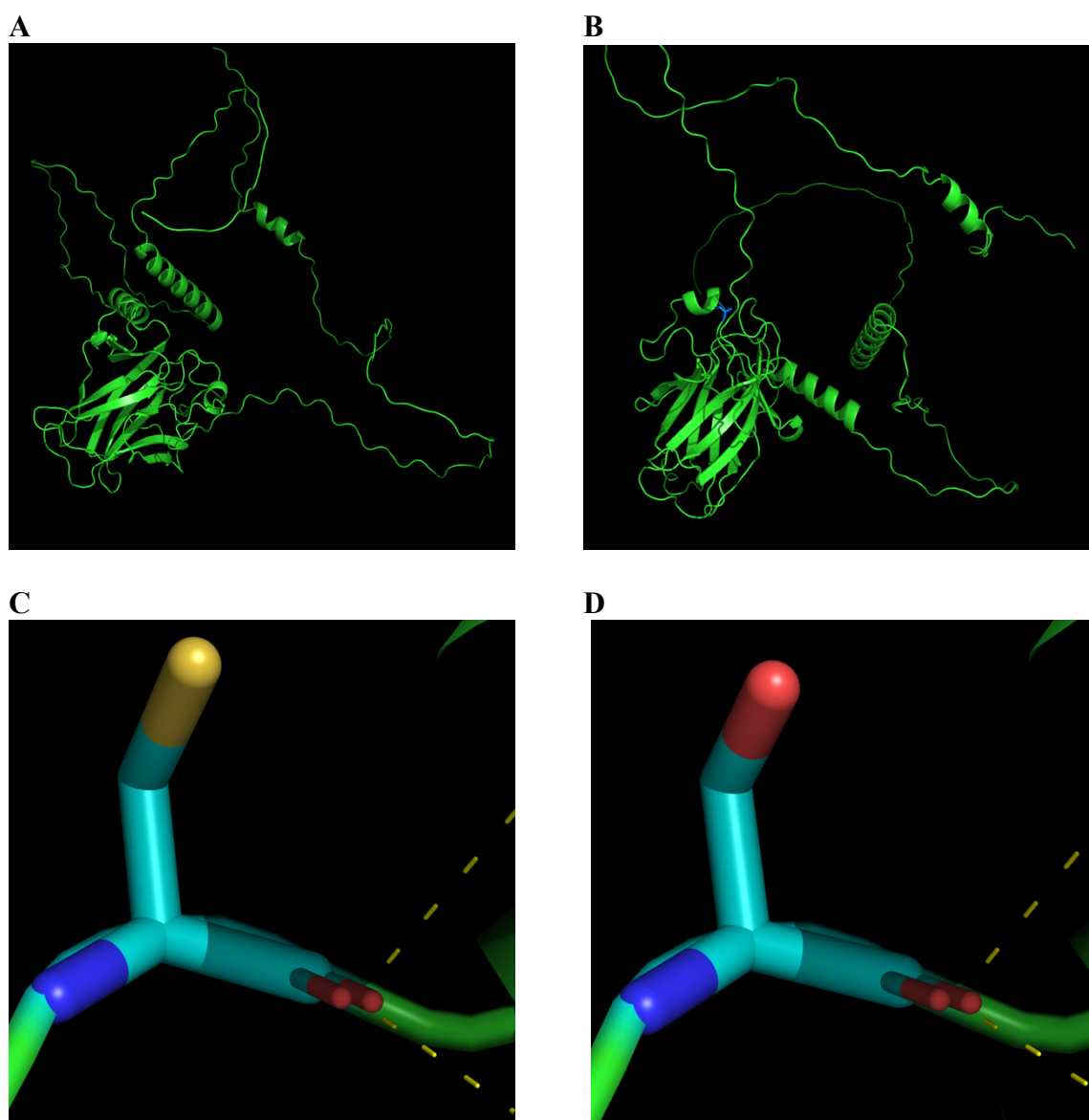


Figure 25 PyMOL renderings of TP53 missense variant, C176S, before and after mutation

PyMOL renderings of the p53 protein C176S mutation in various states including: prior to mutation (A), with the location of cysteine residue to be mutated indicated (B), an up close view of the residue prior to mutation (C), and finally the mutated serine residue (D). Hydrogen bonding and sterically induced clashes are indicated in (C) and (D).

Due to the similarity between these residues, this mutation is unlikely to have an effect on the activity of p53 – a surprising finding as this protein was expected to contain variants leading to resistance development due to its keen involvement in serdemetan’s mechanism of action.

This could suggest that perhaps alteration of the p53 signalling pathway is not how serdemetan resistance is forming, and rather that it might be more dependent on alterations affecting cholesterol transport. However, this is called into question when the results from the variant effect predictor MTBAN are analysed (shown in Table 3). The predictions yielded indicate that the C176S variant has a 75.23% probability of being deleterious. If the results of this tool are accurate, then p53 would be significantly altered by this variant – which is likely a null mutation – prompting resistance and carcinogenesis in the patients in which it occurs.

The last two missense mutations of TP53, Cys to Ser substitutions at aa positions 44 and 83, do not occur in P04637-1, the isoform of the protein chosen by UniProt as the canonical sequence, but rather appear in one of the 8 other isoforms of p53 that are produced by alternative promoter usage & alternative splicing (84). The mutation C44S (Figure 26A and B) can occur in 3 of the 8 isoforms recorded for p53, while evidence for the original aa in the C83S mutation (Figure 26C and D) occurs only in previously recorded computationally mapped potential isoform sequences – however evidence for it is still present in 4 of these instances. This indicates that the parental cell line likely contains such isoforms that are the basis for these mutations.

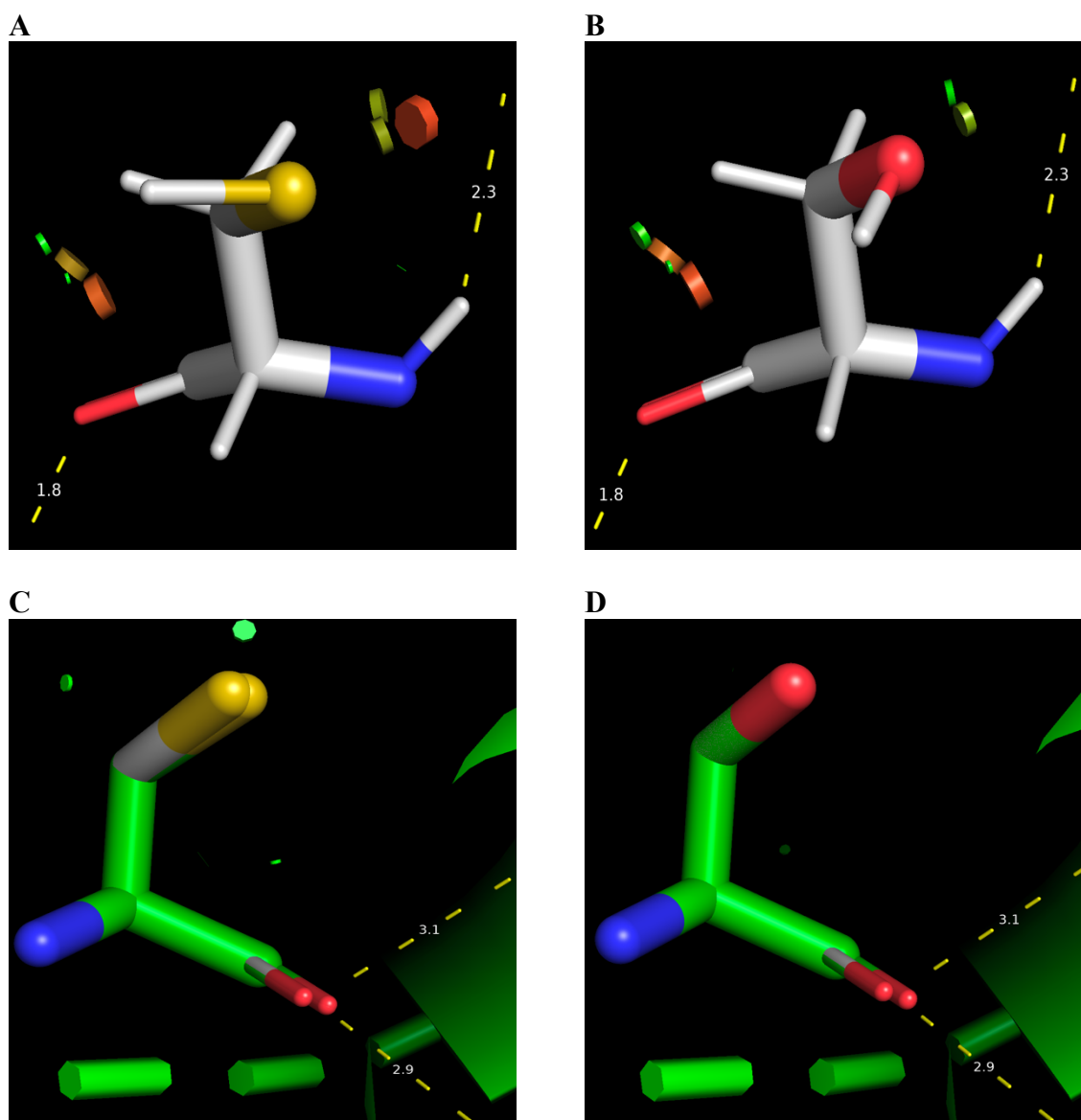


Figure 26 PyMOL renderings of TP53 missense variants, C44S and C83S, before and after mutation

PyMOL renderings of the p53 protein prior to the C44S mutation (A) and the C83S mutation (C), and after with the C44S mutated residue (B) and the C83S mutated residue (D). Hydrogen bonding and sterically induced clashes are indicated.

Since both of these variants involve the substitution of Ser for Cys, like the previously discussed C176S mutation, it is unlikely that these would have significant effects on protein structure due to the similarity between these two residues. This can clearly be seen in the PyMOL model for both the C44S and the C83S mutations, where the aa substitution has not introduced any additional clashes and not disrupted the existing hydrogen bonds.

Subline V also contained three TP53 missense variants, however these occurred at different locations from the ones encountered for subline IX. The base substitution of an adenosine to

a guanine at the CDS position 614 resulted in a change from Tyr, codon UAU, to Cys, codon UGU, at aa position 205 (shown in Figure 27A-D). The remaining missense variants, similarly causing the substitution of an aa Tyr residue to Cys, occurred at different aa residue locations – 73 and 112. Both variants introduced a guanine residue in the place of an adenosine, to change the aa codon from UAU to UGU. The first resulted from a base substitution at the CDS position 218, while the second occurred at the position 335. Like for the variants detected in subline IX, the last two mutations mentioned would not be able to occur in the isoform of p53 that contains the UniProt canonical aa sequence, as this isoform does not contain a tyrosine residue at that position. Instead these tyrosine residues appear in other isoforms listed on UniProt (84), isoforms that largely overlap with the ones that contain the residues needed for the IX subline mutations – further proof for their origin beginning from the same subpopulation of cells in the parental cell line.

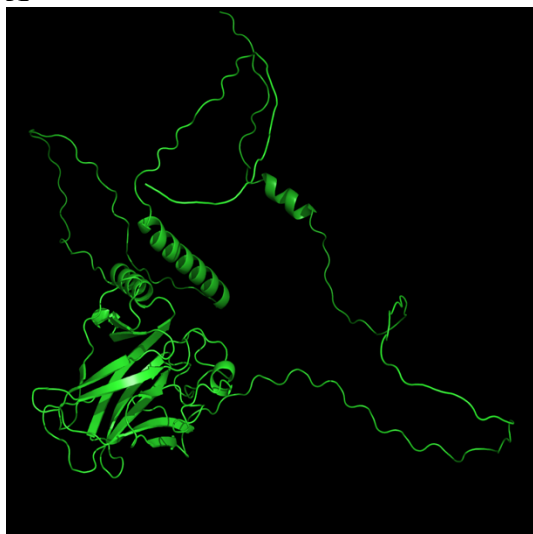
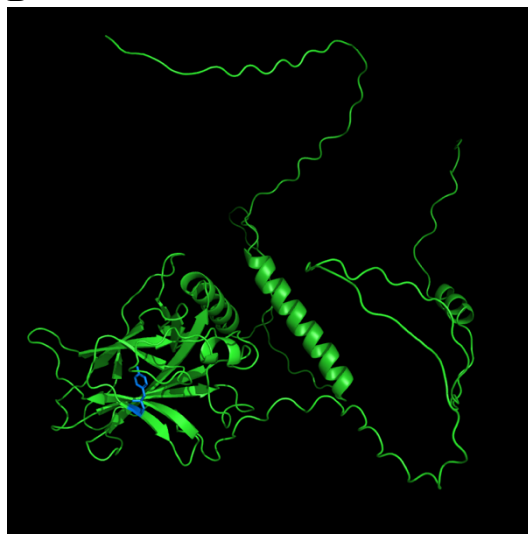
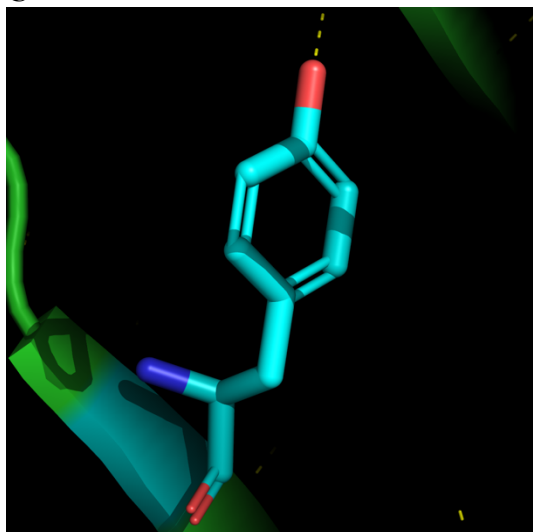
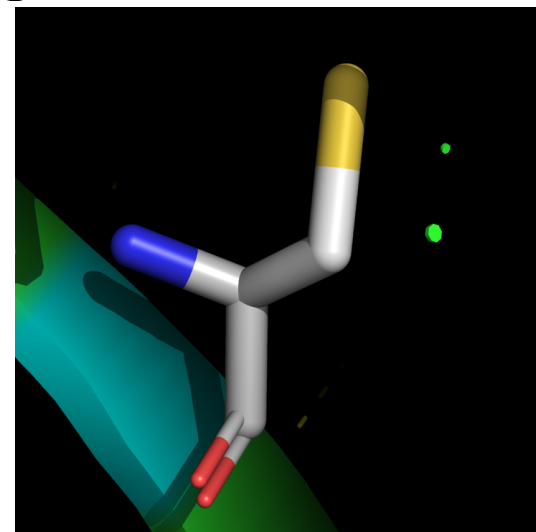
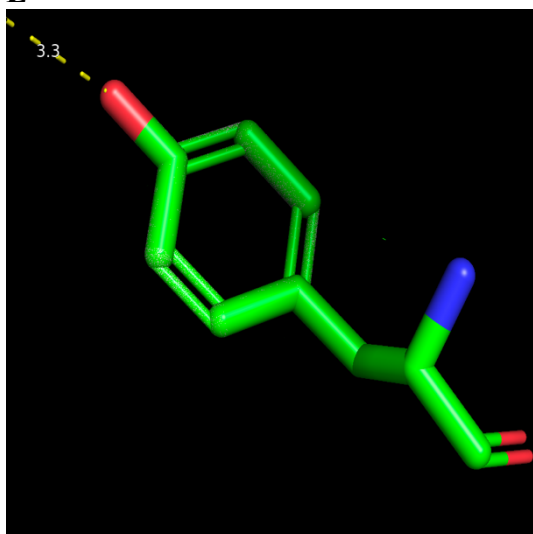
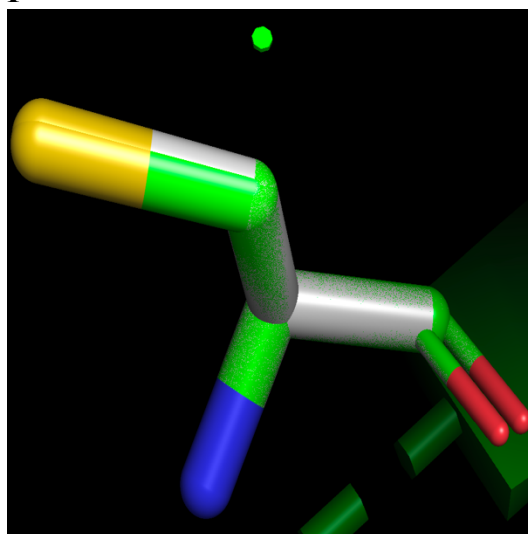
A**B****C****D****E****F**

Figure 27 PyMOL renderings of the TP53 missense variants, Y to C, in subline V before and after mutation

PyMOL renderings of the p53 protein in various states including: prior to mutation (A), with the location of residue to be mutated for the Y205C indicated (B), an up close view of the Tyr residue prior to the Y205C mutation (C), the Y205C mutated residue (D). The Tyr residue for the missense variant Y112C prior to mutation (E), and the Cys residue after the Y112C mutation (F). Hydrogen bonding and sterically induced clashes are indicated in (C) and (D).

The residue change for all these variants is quite significant, with the original aa, Tyr, containing a large hydrophobic side chain with an aromatic ring. The aa replacing this residue, Cys, is much smaller and the -OH group found at the end of Tyr has been replaced with a -SH group, which grants Cys its nucleophilic properties. The weighted impact of this change is supported by the PyMOL renderings of the missense variant changes Y205C and Y112C (Figure 27), however the substitution at Y73C was not rendered in PyMOL as there was no AlphaFold model available of the isotope containing the residue needed. In these figures, the change between residues from C to D and from E to F shows that a hydrogen bond has been lost. The deleterious nature of the Y205C variant was confirmed by MTBAN, with 82.72% likelihood (Table 4). Depending on where in the protein this mutation is located, this could induce a null mutation – rendering p53 inactive and encouraging patient relapse.

Table 4 Results from the variant effect predictor MTBAN for the variants C176S and Y205C in p53

Uniprot accession: P04637				
Variant	Log probability score	Z-score	Probability of deleteriousness	Predicted label
C176S	-15.43	-0.2577	0.7523	deleterious
Y205C	-20.2	-0.6948	0.8272	deleterious

The PLA2G4F gene was found to contain the same two frameshift mutations in both subline IX and V. This corroborates the theorised relationship between the serdemetan-resistant UKF-NB-3 sublines IX and V, where these sublines evolved from the same subpopulation of resistant cells in the original heterogeneous parental cell line – diverging from there. Similarly to what was observed for the APOB variants in sublines I and VIII, these mutations occur at equivalent aa positions 367 and 369 – with the frameshift induced by a nucleotide,

guanine, insertion following the bases coding for the arginine (Arg) residue there. Figure 28 shows the position of this Arg residue in the protein.

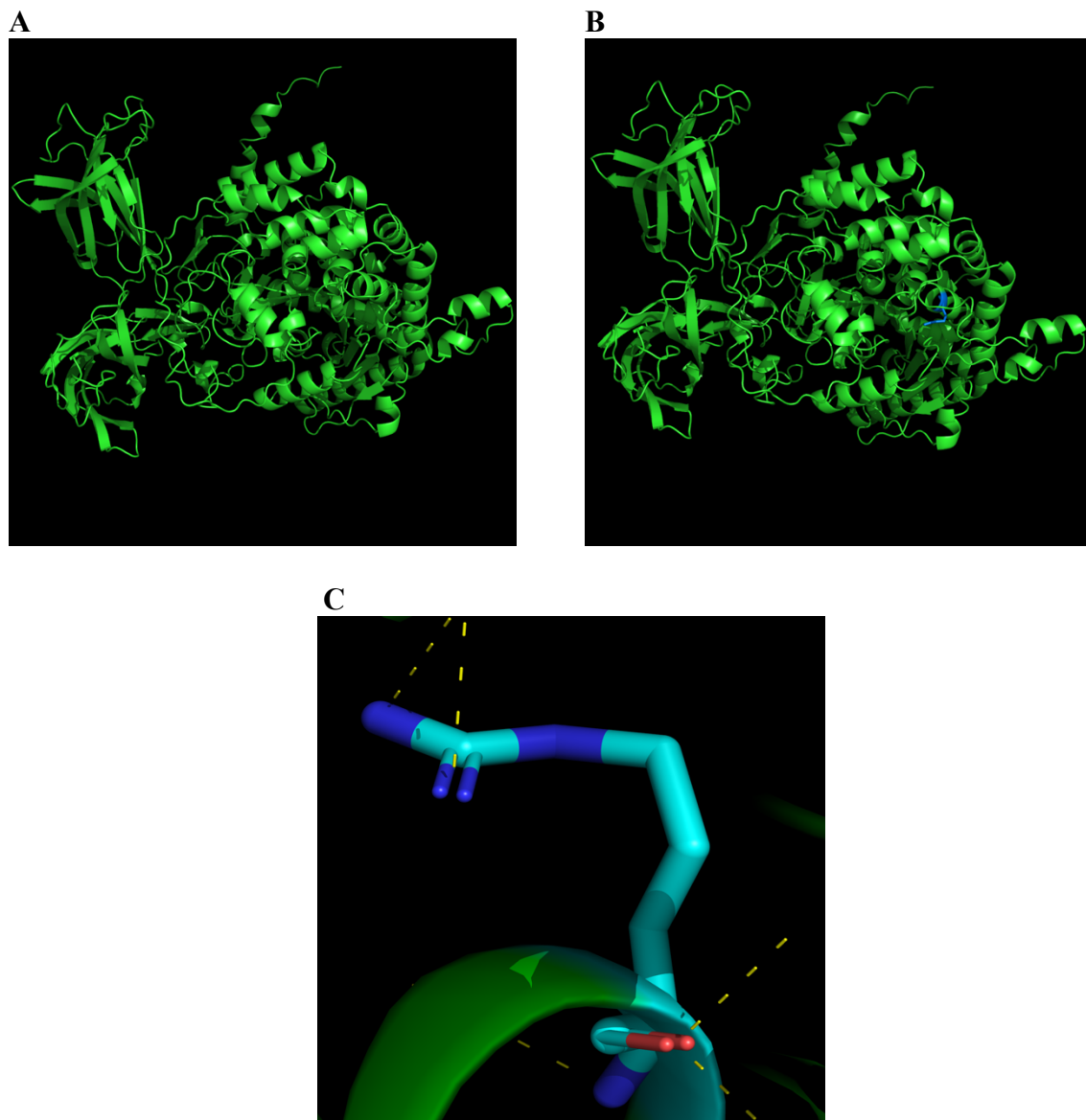


Figure 28 PyMOL renderings of PLA2G4F arginine residue at position 367 prior to mutation

PyMOL renderings of the PLA2G4F protein residue 367, arginine, prior to the introduction of a frameshift mutation. These rendering include the overall protein structure (A), the location of residue where the frameshift is introduced (B), an up close view of the residue interactions prior mutation (C), with hydrogen bonding and sterically induced clashes indicated.

The induction of a frameshift mutation at these points in a protein containing 849 residues means that approximately half of this proteins residues will be altered from the frameshift.

This would cause significant changes in protein structure and functionality – likely yielding inactive PLA2G4F proteins. Due to the integral part these phospholipids play in the formation of the hydrophilic coating of lipoproteins, this variant would likely cause some

degree of disruption to cholesterol transport. Should this result in blocking serdemetan's effect on the cholesterol transport pathway or allowing neuroblastoma cells to sequester cholesterol without increased toxicity – this mutation could play a role in serdemetan resistance formation.

The trend of specific variant overlapping observed for the speculated subline pairs continues with the mutations detected for AHR in subline I and VIII. These sublines contain the same missense variant, down to the CDS position, 1733, where the variant was introduced. Here a thymine was substituted for the original cytosine, changing the codon from ACG to AUG to code for a methionine (Met) residue instead of the original threonine at aa position 578 (shown in Figure 29). This is strong evidence confirming the likelihood that subline I and VIII originated from the same subpopulation of resistant cells in the parental cell line tumour.

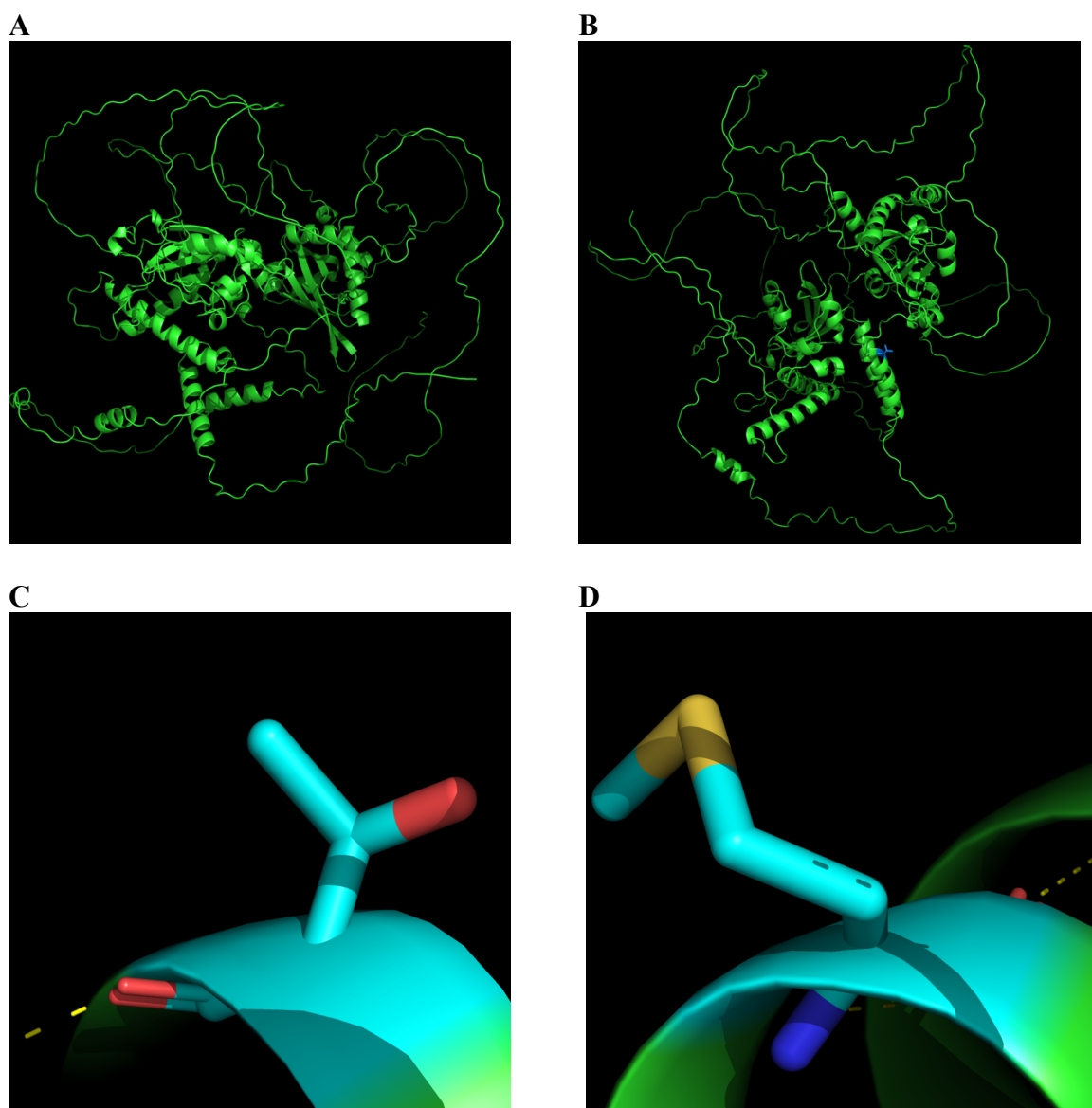


Figure 29 PyMOL renderings of AHR missense variants, T to M, before and after mutation

PyMOL renderings of the AHR protein in various states including: prior to mutation (A), with the location of residue to be mutated indicated (B), an up close view of the residue prior to mutation (C), and finally the mutated residue (D). Hydrogen bonding and sterically induced clashes are indicated in (C) and (D).

As the residues exchanged in this missense mutation differ considerably from one another in terms of their properties, it is predicted that this mutation will be deleterious to the function of AHR. Thr is a polar aa with a hydroxyl group on its side chain while Met is a nonpolar aa with hydrophobic properties. Met is also quite larger than Thr, a factor that might induce steric hinderance. When analysing the PyMOL renderings of this residue change, clashes induced by this variant are not evident. But the results from MTBAN confirm the likelihood of this residue change being deleterious – with a 75.23% chance (Table 5). Therefore the

possibility of this variant having an impactful presence cannot be ruled out and AHR may still well play a role in the resistance observed in this study's sublines (I and VIII specifically).

Table 5 Results from the variant effect predictor MTBAN for the variant T578M in AHR

Uniprot accession: P35869				
Variant	Log probability score	Z-score	Probability of deleteriousness	Predicted label
T578M	-25.0042	-0.2704	0.7523	deleterious

The last gene group that will be investigated in PyMOL are the UGT genes, UGT2B7 and UGT2B4. Surprisingly, UGT2B7 was found to contain variants in the sublines V and VIII, not one of our theorised subline pairs. However the variants occurring in each subline are wholly different, with a missense variant at CDS position 295 (GAC to CAC) changing the aspartic acid (Asp) aa at position 99 into a histidine (His) in subline V and a stop gained variant at CDS position 602 (UUA to UGA) changing the Leu at aa position 201 into a stop codon (shown in Figure 30). This indicates that the two variants could have developed independently from each other rather than evolving from the same resistant cell subpopulation – which is more in line with the current working theory.

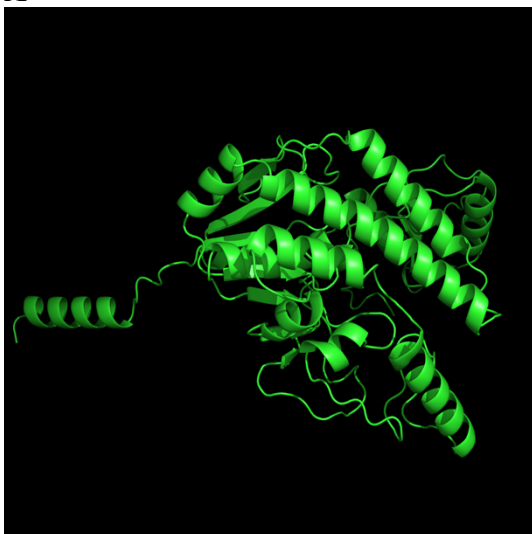
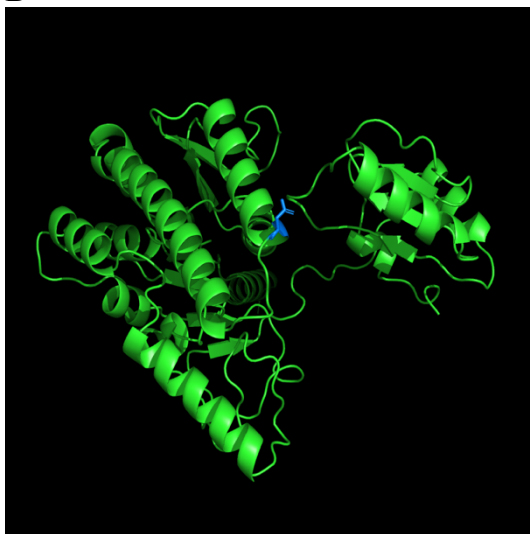
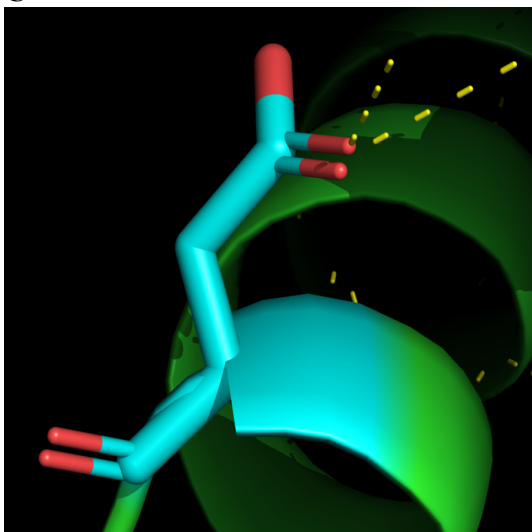
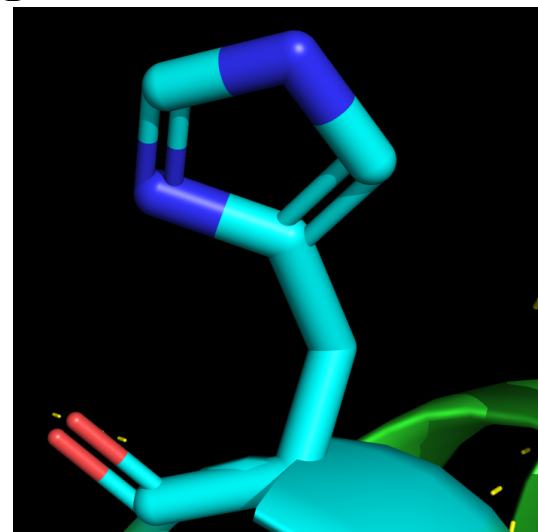
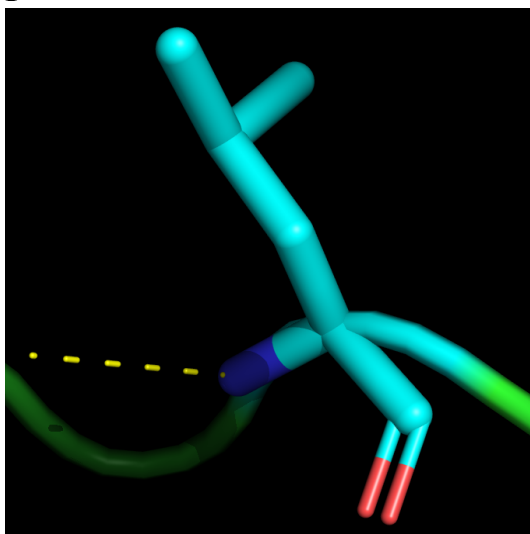
A**B****C****D****E****F**

Figure 30 PyMOL renderings of UGT2B7 missense variants, D to H, before and after mutation as well as renderings of UGT2B7 leucine residue at position 201 prior to stop codon mutation

PyMOL renderings of the UGT2B7 protein in various states including: prior to mutation (A), the location of residue to be mutated indicated for the D99H mutation (B), a up close view of the aspartic acid residue prior to mutation (C), the mutated histidine residue (D), the location of residue to be mutated indicated for the leucine to stop gained mutation (E), and finally a up close view of the leucine residue prior to mutation (F). Hydrogen bonding and sterically induced clashes are indicated in (C), (D) and (F).

When focusing on the Asp to His mutation introduced a aa position 99 for subline V, the change between the properties of the residues is not huge. Both residues are large, polar molecules, with the biggest different being their opposite charges. When analysing the residue change in PyMOL, sterically there are not any clashes introduced however two hydrogen bonds are lost with the change. This change in hydrogen bonds – depending on where the bonds occurred within the protein – might be critical for protein stability. However, it is difficult to tell without performing a molecular dynamics study – which is beyond the scope of this thesis. To draw final conclusions about the significance of this mutation, the variant was submitted to MTBAN with the results of this process shown in Table 6. According to MTBAN predictions, this variant has a 24.61% change of being deleterious and therefore is more likely benign.

Table 6 Results from the variant effect predictor MTBAN for the variant D99H in UGT2B7

Uniprot accession: P16662				
Variant	Log probability score	Z-score	Probability of deleteriousness	Predicted label
D99H	-2.23	0.8587	0.2461	benign

Additionally, UGT2B4 was found mutated in subline V, containing a missense variant at aa position 69, from the change of codons from CCA to UCA at CDS position 205, that switched out the original proline (Pro) aa out for a Ser residue (Figure 31).

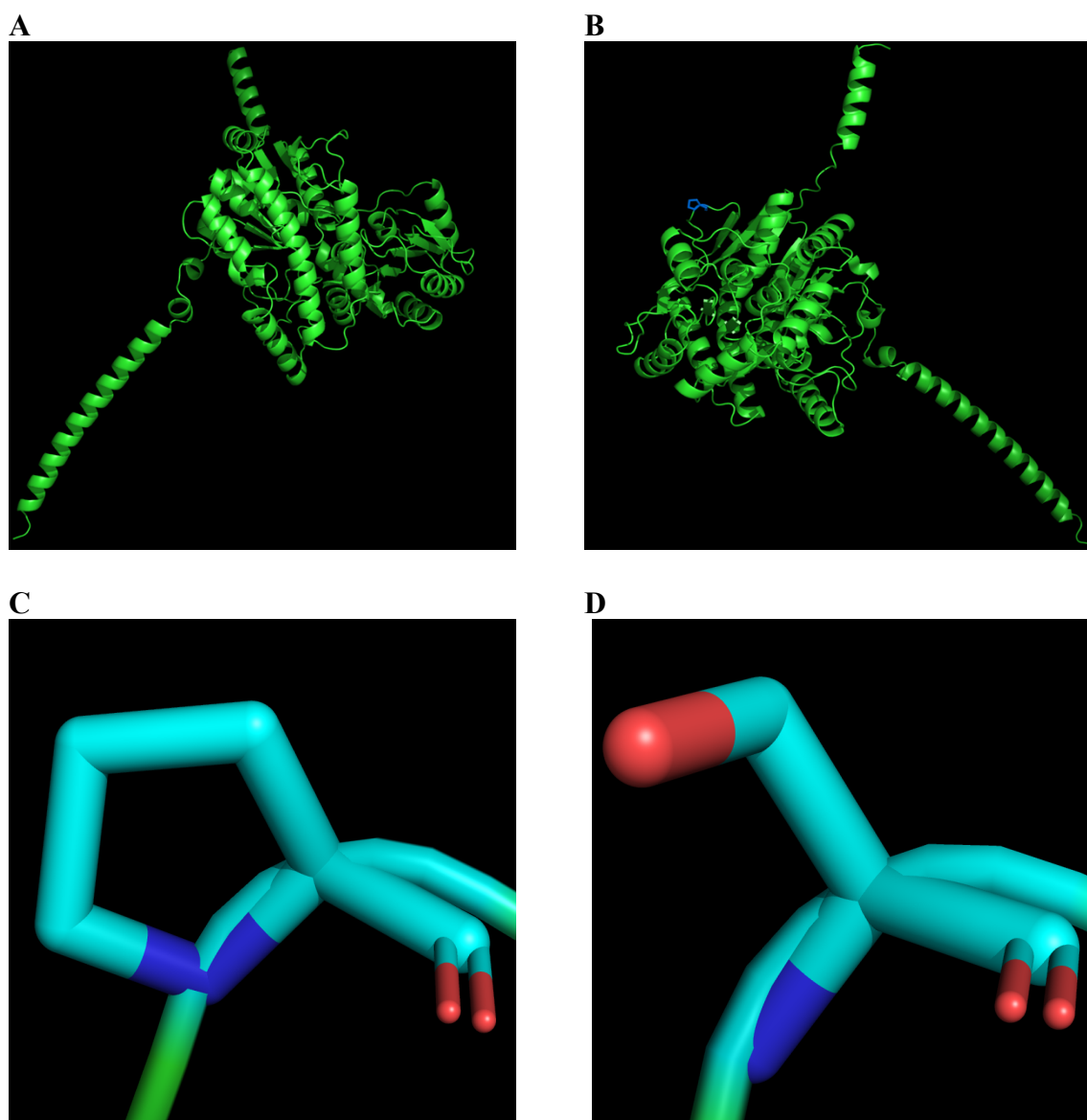


Figure 31 PyMOL renderings of UGT2B4 missense variants, P to S, before and after mutation

PyMOL renderings of the UGT2B4 protein in various states including: prior to mutation (A), the location of residue to be mutated indicated (B), a up close view of the residue prior to mutation (C), and finally the mutated residue (D). Hydrogen bonding and sterically induced clashes are indicated in (C) and (D).

To determine if this mutation was likely to alter protein functionality, the properties of Pro and Ser were considered. Pro is a small, nonpolar residue while Ser is a small polar residue. Both residues are not charged. Next to be considered for the PyMOL renderings of the variant in the protein affected (Figure 31). These indicated no major (visible) changes were made to the protein by this institution. Therefore, it seems that the residue change might have little impact on protein structure. The final piece to be considered before drawing conclusions about this variant was the results from MTBAN. The missense mutation changing Pro to Ser

was determined to be benign by MTBAN, with a probability of 75.39% (Table 7). Therefore, it was concluded that this variant did not seem to alter UGT2b4's structure enough to impact its activity and likely was not involved in the formation of serdemetan resistance.

Table 7 Results from the variant effect predictor MTBAN for the variant P69S in UGT2B4

Uniprot accession: P06133				
Variant	Log probability score	Z-score	Probability of deleteriousness	Predicted label
P69S	-1.2	0.9447	0.2461	benign

Chapter 5: Discussion

5.1 Subline evolution leading to resistance formation

The primary aim of the first results chapter of this thesis, subline evolution leading to resistance formation, was to determine how the different serdemetan resistant sublines evolved from the UKF-NB-3 parental cell line during drug therapy. Examining the degree of similarity between the variants of the different sublines, as well as how they differed from those of the parental cell line, was the main strategy employed towards this end. Two main details became apparent through this investigation. The first was that overall, the variants of all the resistant sublines had a low level of similarity across the board, especially with regard to lost variants. This indicates that there might be some essential overlap between how resistance develops in all of the sublines and that particular variants were consistently selected or lost from the same subpopulation of cells in the parental cell line. However, this could also be due to the variants that were established or lost being found in noncritical genes which are not essential to cellular survival and therefore are more likely to occur due to random mutations. The second detail was that the variants detected within the two resistant subline pairs defined were significantly similar. The first variant pair what was found to have

a large degree of variant overlap was sublines I and VIII. These two sublines were found to have similar numbers of variants within the distinct categories established (Figure 10), a substantial degree of overlap in the variants detected (especially in the category de novo) no matter on the degree of filtering analysed (Figures 11 and 12), similarity in the mutational profiles generated (Figure 14 and 15), as well as in how their VAF ratios changed as compared to the parental cell line (Figure 13). This was also true of the resistant subline pair IX and V. This trend divides the sublines into groups/pairs containing TP53 variants, sublines IX and V, and those that are wild type for p53, I and VIII. Further analysis into exactly how this similarity and TP53 status might involve resistance and what key variants were shared was performed in the next section of results. This further step of investigation was critical to understanding serdemetan resistance development and even serdemetan's mechanism of action itself.

5.2 Functional analysis of the variants and variant pathways

The primary aim of the second results chapter was to investigate variant gene ontology as well as the pathways affected by variants. This was done in order to piece together how serdemetan resistance might have developed in the neuroblastoma sublines of this study, possibly leading to insight how have it develops during patient therapy. The data suggested that there are three main mechanisms of resistance. The first is through modulation of cholesterol transport. The second is via the introduction of p53 null mutations. And the third pathway is drug clearance by AHR. However, since cholesterol transport variants existed independently from p53 and AHR variants in the resistant-sublines – and because all the resistant sublines contained significant variants to the proteins of this pathway– this was thought to possibly be the more crucial pathway modified during serdemetan resistance formation of the sublines investigated. The most notably mutated proteins identified included: APOB (mutated significantly in all the serdemetan resistant sublines), p53 (causing

deleterious mutations in sublines IX and V), PLA2G4F (mutated significantly in sublines IX and V), and AHR (containing deleterious mutations in sublines I and VIII).

5.3 Summary of main findings and concluding remarks

The first conclusion formed was that the resistant sublines had a general level of similarity to their variants, implying that there might be some critical overlap between how serdemetan resistance develops in all of the sublines. The second conclusion formed was that there was a greater degree of similarity between the variants detected within two resistant subline pairs, subline pair I and VIII and subline pair IX and V. This relationship was likely formed on the back of two distinct serdemetan-resistant subpopulations of cells that were selected for during serdemetan treatment and were present in the original chemo sensitive UKF-NB-3 cell line. In the second section of this study the critical overlap between the variants of the UKF-NB-3 serdemetan-resistant sublines was identified to be those involved in and modifying the cholesterol transport pathway – the pathway less reported in previous literature as being involved in serdemetan’s mechanism of action. The subline pair IX and V likely forms resistance via mutations to this pathway in conjunction with variants in the p53 signalling pathway – where null variants were detected. The subline pair I and VIII, however, likely formed resistance through mutations to the cholesterol transport pathway and through mutations to the AHR protein – a protein identified as crucial to drug clearance in the body. Critically mutated proteins were identified as: APOB, p53 PLA2G4F and AHR. It is important to note the possible pitfalls of this study, including the highlighting of variants and pathways that are potentially not relevant or causal to serdemetan resistance and merely occur due to chance and the large volume of data analysed. This analysis alone is not enough to definitively conclude that the highlighted variants are critical in serdemetan resistance development. Further investigations (discussed in the next section) would be required to supplement this study before drawing such conclusions.

Therefore, it can be concluded that the different serdemetan-resistant subline pairs appear to have developed the same underlying resistance mechanism, via modulation to the cholesterol transport pathway, but different accent resistance mechanisms that had little overlap in de novo variants between the two groups. This indicates there is not a “one-size-fits-all” mechanism for resistance development, but that there are some crucial elements that are needed for resistance to occur. Further, TP53 mutations do not appear to be a requirement for serdemetan resistance development, confirming the presence and significance of serdemetan’s p53 independent mechanism of action. TP53 mutated sublines shared more variants in common than the TP53 wild type sublines and vice versa. This suggests that there may have been two original serdemetan-resistant subpopulation of cells that were selected for during therapy and are the basis groups for the subline pairs correlated here. Additionally, TP53 mutant sublines showed the greatest degree of mutation. Further investigations into this may be pivotal to deepening understanding of how serdemetan works.

5.2 Future Studies

The most proximal analysis needed is a molecular dynamics study, investigating the variants highlighted here as being most significant. This would allow for conclusions to be drawn regarding how (or if) these variants impact protein structure to induce serdemetan resistance in neuroblastoma cells. This combined with proteomic approaches could further help to elucidate the intracellular pathways and molecules implicated in serdemetan resistance (1). Additionally, the performance of functional wet lab studies could supplement the findings provided here. Such research, including gene knock-out and PCR methods, could potentially validate the findings of this computational study.

In recent years, studies have also highlighted the importance of investigating expression levels of noncoding RNAs as well as genome-wide methylation patterns (5). A miRNA expression signatures study could provide further insights into the nature of serdemetan-resistance neuroblastoma.

Any and all of these methods would be complementary to the comparative computational study performed here, which outlined promising avenues for these studies to begin with.

Bibliography

1. Esposito MR, Aveic S, Seydel A, Tonini GP. Neuroblastoma treatment in the post-genomic era. *J Biomed Sci*. 2017;24(1):1–16.
2. Mitra S, Muralidharan SV, Marco M Di, Juvvuna PK, Kosalai ST, Reischl S, et al. Subcellular distribution of p53 by the p53-Responsive lncRNA NBAT1 Determines Chemotherapeutic Response in Neuroblastoma. *Cancer Res*. 2021;81(6):1457–71.
3. Colon NC, Chung DH. Neuroblastoma. *Adv Pediatr*. 2011;58(1):297–311.
4. Newman EA, Abdessalam S, Aldrink JH, Austin M, Heaton TE, Bruny J, et al. Update on neuroblastoma. *J Pediatr Surg* [Internet]. 2019;54(3):383–9. Available from: <https://doi.org/10.1016/j.jpedsurg.2018.09.004>
5. Schleiermacher G, Janoueix-Lerosey I, Delattre O. Recent insights into the biology of neuroblastoma. *Int J Cancer*. 2014;135(10):2249–61.
6. Michaelis M, Hinsch N, Michaelis UR, Rothweiler F, Simon T, Ilhelm Doerr HW, et al. Chemotherapy-associated angiogenesis in neuroblastoma tumors. *Am J Pathol* [Internet]. 2012;180(4):1370–7. Available from: <http://dx.doi.org/10.1016/j.ajpath.2011.12.011>

7. Smith V, Foster J. High-risk neuroblastoma treatment review. *Children*. 2018;5(9).
8. Wang W, Wang X, Rajaei M, Youn JY, Zafar A, Deokar H, et al. Targeting MDM2 for neuroblastoma therapy: In vitro and in vivo anticancer activity and mechanism of action. *Cancers (Basel)*. 2020;12(12):1–16.
9. De Preter K, Vandesompele J, Heimann P, Yigit N, Beckman S, Schramm A, et al. Human fetal neuroblast and neuroblastoma transcriptome analysis confirms neuroblast origin and highlights neuroblastoma candidate genes. *Genome Biol*. 2006;7(9).
10. Bowen KA, Chung DH. Recent advances in neuroblastoma. *Curr Opin Pediatr*. 2009;21(3):350–6.
11. Miles X, Vandevoorde C, Hunter A, Bolcaen J. MDM2/X Inhibitors as Radiosensitizers for Glioblastoma Targeted Therapy. *Front Oncol*. 2021;11(July):1–21.
12. Liu R, Shi P, Wang Z, Yuan C, Cui H. Molecular Mechanisms of MYCN Dysregulation in Cancers. *Front Oncol*. 2021;10(February):1–12.
13. Dang C V. MYC on the Path to Cancer. *Cell*. 2012;149(1):22–35.
14. Transcriptional misregulation in cancer [Internet]. KEGG PATHWAY. [cited 2023 Apr 15]. Available from: <https://www.genome.jp/pathway/hsa05202>
15. Michaelis M, Wass MN, Cinatl J. Drug-adapted cancer cell lines as preclinical models of acquired resistance. *Cancer Drug Resist*. 2019;2(3):447–56.
16. Mansoori B, Mohammadi A, Davudian S, Shirjang S, Baradaran B. The different mechanisms of cancer drug resistance: A brief review. *Adv Pharm Bull [Internet]*. 2017;7(3):339–48. Available from: <http://dx.doi.org/10.15171/apb.2017.041>
17. Cree IA, Charlton P. Molecular chess? Hallmarks of anti-cancer drug resistance. *BMC Cancer [Internet]*. 2017;17(1):1–8. Available from: <http://dx.doi.org/10.1186/s12885->

016-2999-1

18. Holohan C, Van Schaeybroek S, Longley DB, Johnston PG. Cancer drug resistance: an evolving paradigm. *Nat Rev Cancer*. 2013;13:714–26.
19. Swanton C. Intratumor heterogeneity: Evolution through space and time. *Cancer Res*. 2012;72(19):4875–82.
20. Fenton TR, Garrett MD, Wass MN, Michaelis M. What really matters - response and resistance in cancer therapy. *Cancer Drug Resist*. 2018;1(4):200–3.
21. Cinatl J, Speidel D, Hardcastle I, Michaelis M. Resistance acquisition to MDM2 inhibitors. *Biochem Soc Trans*. 2014;42(4):752–7.
22. Vogelstein B, Lane D, Levine A. Surfing p53 Network. *Nature*. 2000;408(6810):307–10.
23. Levine AJ, Hu W, Feng Z. The P53 pathway: What questions remain to be explored? *Cell Death Differ*. 2006;13(6):1027–36.
24. Trino S, De Luca L, Laurenzana I, Caivano A, Del Vecchio L, Martinelli G, et al. P53-MDM2 pathway: Evidences for a new targeted therapeutic approach in B-acute lymphoblastic leukemia. *Front Pharmacol*. 2016;7(DEC):1–7.
25. Bálint É, Vousden KH. Activation and activities of the p53 tumour suppressor protein. *Br J Cancer*. 2001;85(12):1813–23.
26. Moreno L, Caron H, Geoerger B, Eggert A, Schleiermacher G, Brock P, et al. Accelerating drug development for neuroblastoma - New Drug Development Strategy: an Innovative Therapies for Children with Cancer, European Network for Cancer Research in Children and Adolescents and International Society of Paediatric Oncology Europe Neu. *Expert Opin Drug Discov [Internet]*. 2017;12(8):801–11. Available from: <https://doi.org/10.1080/17460441.2017.1340269>

27. Michaelis M, Rothweiler F, Barth S, Cinat J, Van Rikxoort M, Löschmann N, et al. Adaptation of cancer cells from different entities to the MDM2 inhibitor nutlin-3 results in the emergence of p53-mutated multi-drug-resistant cancer cells. *Cell Death Dis.* 2011;2(12).
28. Brown CJ, Cheok CF, Verma CS, Lane DP. Reactivation of p53: from peptides to small molecules. *Trends Pharmacol Sci.* 2011;32(1):53–62.
29. Keshelava N, Zuo JJ, Chen P, Waidyaratne SN, Luna MC, Gomer CJ, et al. Loss of p53 function confers high-level multidrug resistance in neuroblastoma cell lines. *Cancer Res.* 2001;61(16):6185–93.
30. Chargari C, Leteur C, Angevin E, Bashir T, Schoentjes B, Arts J, et al. Preclinical assessment of JNJ-26854165 (Serdemetan), a novel tryptamine compound with radiosensitizing activity in vitro and in tumor xenografts. *Cancer Lett.* 2011;312(2):209–18.
31. Jones RJ, Gu D, Bjorklund CC, Kuitse I, Remaley AT, Bashir T, et al. The novel anti-cancer agent JNJ-26854165 induces cell death through inhibition of cholesterol transport and degradation of ABCA1. *J Pharmacol Exp Ther.* 2013;347(2):540.
32. Selivanova G. Wild type p53 reactivation: From lab bench to clinic. *FEBS Lett* [Internet]. 2014;588(16):2628–38. Available from: <http://dx.doi.org/10.1016/j.febslet.2014.03.049>
33. Kung CP, Weber JD. It's Getting Complicated—A Fresh Look at p53-MDM2-ARF Triangle in Tumorigenesis and Cancer Therapy. *Front Cell Dev Biol.* 2022;10(January):1–26.
34. Zafar A, Wang W, Liu G, Xian W, McKeon F, Zhou J, et al. Targeting the p53-MDM2 pathway for neuroblastoma therapy: Rays of hope. *Cancer Lett* [Internet].

- 2021;496(June 2020):16–29. Available from:
<https://doi.org/10.1016/j.canlet.2020.09.023>
35. Lehman JA, Hauck PM, Gendron JM, Batuello CN, Eitel JA, Albig A, et al. Serdemetan Antagonizes the Mdm2-HIF1 α Axis Leading to Decreased Levels of Glycolytic Enzymes. *PLoS One*. 2013;8(9):8–12.
 36. Röhrl C, Stangl H. Cholesterol metabolism—physiological regulation and pathophysiological deregulation by the endoplasmic reticulum. *Wiener Medizinische Wochenschrift*. 2018;168(11–12):280–5.
 37. Joshi S, Singh AR, Durden DL. MDM2 regulates hypoxic hypoxia-inducible factor 1 α stability in an E3 ligase, proteasome, and PTEN-phosphatidylinositol 3-kinase-AKT-dependent manner. *J Biol Chem*. 2014;289(33):22785–97.
 38. Xia X. Bioinformatics and Drug Discovery. *Curr Top Med Chem*. 2017;17(15):1709–26.
 39. Voelkerding K V., Dames SA, Durtschi JD. Next-generation sequencing:from basic research to diagnostics. *Clin Chem*. 2009;55(4):641–58.
 40. Key differences between next-generation sequencing and Sanger sequencing [Internet]. illumina. 2023. Available from:
<https://emea.illumina.com/science/technology/next-generation-sequencing/ngs-vs-sanger-sequencing.html#:~:text=The critical difference between Sanger,of genes at one time.>
 41. Behjati S, Tarpey PS. What is next generation sequencing? *Arch Dis Child Educ Pract Ed*. 2013;98(6):236–8.
 42. Warr A, Robert C, Hume D, Archibald A, Deeb N, Watson M. Exome sequencing: Current and future perspectives. *G3 Genes, Genomes, Genet*. 2015;5(8):1543–50.
 43. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole-genome

- sequencing is more powerful than whole-exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A*. 2015;112(17):5473–8.
44. Germline short variant discovery (SNPs + Indels). GATK.
 45. Andrews S. FastQC [Internet]. Babraham Bioinformatics. [cited 2022 Jun 12]. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
 46. Pflaum J, Schlosser S, Müller M. P53 family and cellular stress responses in cancer. *Front Oncol*. 2014;4(OCT):1–15.
 47. p53 signaling pathway - Homo sapiens (human) [Internet]. KEGG PATHWAY. [cited 2022 Aug 23]. Available from: <https://www.genome.jp/pathway/hsa04115>
 48. Cholesterol metabolism - Homo sapiens (human) [Internet]. KEGG PATHWAY. [cited 2022 Aug 23]. Available from: <https://www.genome.jp/pathway/hsa04979>
 49. Kanehisa M, Sato Y. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci*. 2020;29(1):28–35.
 50. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. *Nucleic Acids Res*. 2019;47(D1):D590–5.
 51. Strom SP. Current practices and guidelines for clinical next-generation sequencing oncology testing. *Cancer Biol Med*. 2016;13(1):3–11.
 52. Uhr K, Prager-van der Smissen WJC, Heine AAJ, Ozturk B, Smid M, Göhlmann HWH, et al. Understanding drugs in breast cancer through drug sensitivity screening. *Springerplus*. 2015;4(1).
 53. Smith MA, Gorlick R, Kolb EA, Lock R, Carol H, Maris JM, et al. Initial Testing of JNJ-26854165 (Serdemetan) by the Pediatric Preclinical Testing Program. *Pediatr Blood Cancer*. 2012;59(February):329–32.
 54. Taberero J, Dirix L, Schöfski P, Cervantes A, Lopez-Martin JA, Capdevila J, et al. A

- phase I first-in-human pharmacokinetic and pharmacodynamic study of serdemetan in patients with advanced solid tumors. *Clin Cancer Res*. 2011;17(19):6313–21.
55. Duffy MJ, Synnott NC, O’Grady S, Crown J. Targeting p53 for the treatment of cancer. *Semin Cancer Biol* [Internet]. 2022;79(July 2020):58–67. Available from: <https://doi.org/10.1016/j.semcancer.2020.07.005>
 56. Gupta A, Shah K, Oza MJ, Behl T. Reactivation of p53 gene by MDM2 inhibitors: A novel therapy for cancer treatment. *Biomed Pharmacother* [Internet]. 2019;109(October 2018):484–92. Available from: <https://doi.org/10.1016/j.biopha.2018.10.155>
 57. Lakoma A, Barbieri E, Agarwal S, Jackson J, Chen Z, Kim Y, et al. The MDM2 small-molecule inhibitor RG7388 leads to potent tumor inhibition in p53 wild-type neuroblastoma. *Cell Death Discov*. 2015;1(1):1–9.
 58. Tweddle DA, Malcom AJ, Brown N, Pearson ADJ, Lunec J. Evidence for the Development of p53 Mutations after Cytotoxic Therapy in a Neuroblastoma Cell Line. *Cancer Res*. 2001;61(1):8–13.
 59. Tweddle DA, Pearson ADJ, Haber M, Norris MD, Xue C, Flemming C, et al. The p53 pathway and its inactivation in neuroblastoma. *Cancer Lett*. 2003;197(1–2):93–8.
 60. Zhang Y, Xiong Y, Yarbrough WG. ARF promotes MDM2 degradation and stabilizes p53: ARF-INK4a locus deletion impairs both the Rb and p53 tumor suppression pathways. *Cell*. 1998;92(6):725–34.
 61. Chen L, Rousseau RF, Middleton SA, Nichols GL, Newell DR, Lunec J, et al. Pre-clinical evaluation of the MDM2-p53 antagonist RG7388 alone and in combination with chemotherapy in neuroblastoma. *Oncotarget*. 2015;6(12):10207–21.
 62. Inoue K, Fry EA, Frazier DP. Transcription factors that interact with p53 and Mdm2.

- Int J Cancer. 2016;138(7):1577–85.
63. Kim ES, Shohet JM. Reactivation of p53 via MDM2 inhibition. *Cell Death Dis*. 2015;6(10):1–2.
 64. Valter K, Zhivotovsky B, Gogvadze V. Cell death-based treatment of neuroblastoma. *Cell Death Dis* [Internet]. 2018;9(2):1–15. Available from: <http://dx.doi.org/10.1038/s41419-017-0060-1>
 65. Abhinand CS, Raju R, Soumya SJ, Arya PS, Sudhakaran PR. VEGF-A/VEGFR2 signaling network in endothelial cells relevant to angiogenesis. *J Cell Commun Signal* [Internet]. 2016;10(4):347–54. Available from: <http://dx.doi.org/10.1007/s12079-016-0352-8>
 66. Norris MD, Gilbert J, Smith SA, Marshall GM, Salwen H, Cohn SL, et al. Expression of the putative tumour suppressor gene, p73, in neuroblastoma and other childhood tumours. *Med Pediatr Oncol*. 2001;36(1):48–51.
 67. Mahshid Y, Lisy MR, Wang X, Spanbroek R, Flygare J, Christensson B, et al. High expression of 5-lipoxygenase in normal and malignant mantle zone B lymphocytes. *BMC Immunol*. 2009;10:1–12.
 68. Shinohara H isaaki, Kurosaki T. Comprehending the complex connection between PKCbeta, TAK1, and IKK in BCR signaling. *Immunol Rev*. 2009;232(1):300–18.
 69. KEGG Mapping [Internet]. Kanehisa Laboratories. 2019. Available from: <https://www.genome.jp/kegg/kegg1b.html>
 70. KO (KEGG ORTHOLOGY) Database [Internet]. KEGG. 2021. Available from: <https://www.genome.jp/kegg/ko.html>
 71. FARESE R V., LINTON MF, YOUNG SG. Apolipoprotein B gene mutations affecting cholesterol levels. *J Intern Med*. 1992;231(6):643–52.
 72. Apob [Internet]. MGI Alliance of genome resources. 2022 [cited 2022 Aug 30].

Available from: <http://www.informatics.jax.org/marker/MGI:88052>

73. Havel RJ, Kane JP. Introduction: Structure and Metabolism of Plasma Lipoproteins. In: Valle DL, Antonarakis S, Ballabio A, Beaudet AL, Mitchell GA, editors. The Online Metabolic and Molecular Bases of Inherited Disease [Internet]. New York, NY: McGraw-Hill Education; 2019. Available from: <http://ommbid.mhmedical.com/content.aspx?aid=1181441818>
74. Vasiliou V, Vasiliou K, Nebert DW. Human ATP-binding cassette (ABC) transporter family. 2009;3(3):281–90.
75. Wenzel JJ, Kaminski WE, Piehler A, Hiemerl S, Langmann T, Schmitz G. ABCA10, a novel cholesterol-regulated ABCA6-like ABC transporter. Biochem Biophys Res Commun. 2003;306(4):1089–98.
76. Okamoto T, Kawaguchi K, Watanabe S, Agustina R, Ikejima T, Ikeda K, et al. Characterization of human ATP-binding cassette protein subfamily D reconstituted into proteoliposomes. Biochem Biophys Res Commun. 2018;496(4):1122–7.
77. Ohto T, Uozumi N, Hirabayashi T, Shimizu T. Identification of novel cytosolic phospholipase A2s, murine cPLA2 δ , ϵ , and ζ , which form a gene cluster with cPLA2 β . J Biol Chem [Internet]. 2005;280(26):24576–83. Available from: <http://dx.doi.org/10.1074/jbc.M413711200>
78. Nakayama H, Iwabuchi K. Molecular Mechanisms Underlying the Immunological Activities of Glycosphingolipid-Enriched Lipid Rafts in Phagocytes. Ref Modul Chem Mol Sci Chem Eng. 2019;
79. Stevens EA, Mezrich JD, Bradfield CA. The aryl hydrocarbon receptor: A perspective on potential roles in the immune system. Immunology. 2009;127(3):299–311.
80. Lévesque E, Turgeon D, Carrier JS, Montminy V, Beaulieu M, Bélanger A. Isolation and

- characterization of the UGT2B28 cDNA encoding a novel human steroid conjugating UDP-glucuronosyltransferase. *Biochemistry*. 2001;40(13):3869–81.
81. Carrier J-SC, Turgeon D, Journault K, Hum DW, Bélanger A. Isolation and characterization of the human UGT2B7 gene. *Biochem Biophys Res Commun*. 2000;272(2):616–21.
 82. Kim HY, Jeon W, Kim D. An enhanced variant effect predictor based on a deep generative model and the Born-Again Networks. *Sci Rep*. 2021;11(1).
 83. Parnell LD. Advances in Technologies and Study Design. *Prog Mol Biol Transl Sci*. 2012;108:17–50.
 84. P04637 · P53_HUMAN [Internet]. UniProt. Available from: <https://www.uniprot.org/uniprotkb/P04637/entry#sequences>