



Kent Academic Repository

Roobol, Joanne, Stanley, Maria, Zielewicz, Laura, Wei, Wei, Allen, Martin J., Zhang, Lin and Smales, C. Mark (2026) *Identification and engineering of Chinese hamster ovary (CHO) cell genes whose expression is correlated with age-related growth rate acceleration*. Journal of Biological Engineering . ISSN 1754-1611.

Downloaded from

<https://kar.kent.ac.uk/116188/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://doi.org/10.1186/s13036-026-00765-3>

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY-NC-ND (Attribution-NonCommercial-NoDerivatives)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in **Title of Journal** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

Identification and engineering of Chinese hamster ovary (CHO) cell genes whose expression is correlated with age-related growth rate acceleration

Joanne Roobol · Maria Stanley · Laura Zielewicz · Wei Wei · Martin J. Allen · Lin Zhang · C. Mark Smales 

Received: 28 May 2026 / Accepted: 19 August 2026
© The Author(s) 2026

Abstract

Generational age-related Chinese hamster ovary (CHO) cell growth changes are undesirable for large-scale manufacturing cell lines. An age-related high growth phenotype in CHO cells has previously been shown to be associated with a Chromosome 2 (Chr2) duplication event in late generation cells. We have evaluated the contribution of a subset of the 123 genes involved in the duplication event in relation to the high growth phenotype of late generation cells (>90 generations post-cloning). We show that this age correlated high growth phenotype associated with a Chr2 duplication event in CHO cells can be reversed by single knockdown of several transcripts that are upregulated as a result of the duplication event. Further, one of these genes, *Cxxc1*, when overexpressed in low growth phenotype cells, results in the generation of high growth phenotype cell pools. Thus, *Cxxc1* overexpression engineering in CHO cells can impart a high growth phenotype that correlates with those observed in Chr2 duplication event cell lines. Furthermore, this impact can be reversed by subsequent knockdown of *Cxxc1* transcript levels such that the cells revert to a low growth phenotype. The high growth phenotype of *Cxxc1* overexpressing engineered cell pools was shown to be stable for 100 generations, and although the high growth phenotype was associated with a reduced cell specific productivity, volumetric titre was maintained/compensated for by the increased cell number. We propose targeted cell engineering of CHO cells to overexpress *Cxxc1* may be a potential strategy to overcome the issue of age-related genetic instability associated growth phenotype to generate new host cell lines with a stable, high growth phenotype.

Keywords CHO cells · Age-related growth phenotype instability · Chromosome 2 duplication · *Cxxc1* · Cell and metabolic engineering

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.

Joanne Roobol and Maria Stanley are joint first authors.



Background

Mammalian cells, and in particular Chinese hamster ovary (CHO) cells, are the leading platform for the multi-billion dollar biotherapeutic protein market. CHO cells dominate the production market partially due to their robust adaptable nature allowing large production suspension cultures to be grown to high cell densities yielding multi-g/litre of high-quality recombinant protein [1]. CHO cells excel in their ability to produce high quality and titre of secreted recombinant protein products for therapeutic applications [2, 3], partially due to the ‘plasticity’ of these cells that allows them to adapt to different culture processes and the demands of recombinant protein production [4–9]. The flipside of this ‘plasticity’ is the potential for genomic instability which can lead to genotypic and/or phenotypic heterogeneity that can ultimately result in a clonal recombinant cell line having to be rejected late in cell line development. Additionally, and possibly more consequential, is the possibility of manufacturing runs being rejected if phenotypic instability results in product that is not within acceptable product quality limits. Instability can be associated with changes in growth phenotype, product expression, and product quality [10]. Such instability can arise from genomic rearrangements, epigenetic regulation and/or due to outgrowth of sub-populations of cells that have acquired a genetic alteration [4, 6, 8, 11–14].

A number of recent reports have identified growth associated genes and pathways in CHO cells that impact growth phenotypes using both genome-wide screens and engineering approaches. For example, Xiong et al. used a pooled CRISPR screen for a negative selection screen that identify 1980 genes that affected CHO cell proliferation [15]. Marzluf et al. used an elegant genome wide CRISPR knock-out screen to identify 235 putative growth-enhancing gene knockouts and 3,110 genes essential for cell fitness, including the chromatin-associated transcription factor *Cxxc1* [16]. Shin et al. have also recently reported a genome scale pooled CRISPR knockout screen in CHO cells identifying essential genes for cell fitness and those correlated with recombinant protein production [17]. Collectively these reports indicate that CHO cell growth phenotypes are associated with multi-gene expression signatures. Further, Masuda et al. have recently reported the development of a novel, fast growing CHO derived cell line (CHO-MK) with a doubling time of approximately 10 h, showing that achieving such rapid growth is possible in in vitro culture CHO cells, although the mechanisms by which the growth phenotype is sustained are yet to be elucidated [18].

The growing knowledge of key determinants of CHO cell growth phenotypes and of recombinant protein production has resulted in attempts to engineer these cells to manipulate these phenotypes to enhance cell growth and recombinant protein productivity. Torres et al. used an anti-apoptosis engineering approach to prolong culture alongside a multi-target inducible engineering approach to accelerate growth and then activate cell cycle arrest once sufficient cell numbers had been achieved to deliver enhanced viable cell concentrations, qP and product yields [19]. The authors ‘tuned’ growth and culture longevity by firstly knocking out pro-apoptotic proteins Bax and Bak using a CRISPR/Cas9 approach, secondly layering inducible cMYC expression upon this to allow rapid cell proliferation, and finally adding an inducible BLIMP1 ‘brake’ to attenuate cell proliferation and initiate cell cycle arrest. This exquisite system allowed the tuning of cell growth and productivity to maximise both [19]. In a parallel approach, Latorre et al. first generated a recombinant EPO-producing CHO cell line and then co-overexpressed two transcription factors, MYC and XBP1s, that resulted in increased viable cell concentration and enhanced EPO production [20]. The enhanced growth was ascribed to the over-expression of MYC as in the study of Torres et al.

With regard to the issue of CHO cell line product and growth instability, a number of mechanisms associated with these changes have been described. Silencing of recombinant gene expression is attributed to, amongst other mechanisms, gene silencing, epigenetic methylation events, or loss of recombinant gene copies, eventually resulting in multiple phenotypic outcomes [5, 6, 8, 10–14, 21–25]. Long term culture is reported to result in age-related instability phenotypes whereby CHO cells adapt to prioritize proliferation and survival that can result in growth instability [10]. In particular, CHO cells at late passage number in long term culture showed an early decline in culture viability and little stationary phase whilst there was an upregulation in the expression of genes

such as MAPK/ERK and PI3K-Akt involved in cell proliferation or survival pathways [10]. With direct relevance to this study, Huhn et al. [7] reported on genomic instability of CHO cells, describing particular genomic regions as being profoundly diverse, where smaller chromosomes (7–10 and X) were more labile and heterochromatic and larger chromosomes, especially Chr1, were relatively stable with less recombination and associated genetic changes. The authors also described a haploid reduction of a Chromosome 2 (Chr2) region after a culture adaptation that was consistently inherited suggesting a common adaptation mechanism. Interestingly, they reported that Chr2 is significantly enriched with genes associated with CHO suspension culture e.g. cytoskeleton rearrangement and insulin receptor signalling [7].

Age-related or long term culturing changes in growth phenotype in cultured mammalian cells are often considered indicative of clonal instability and to be avoided [10]. Wei et al. have also reported on an age-related high growth phenotype that is associated with a Chromosome 2 duplication event in late generation cells [24]. The authors identified a duplication region from Chromosome 2 in later generation cells (defined by >90 divisions from cloning in Wei et al., [24]) compared to early generation cells (defined by <30 divisions from cloning). This gene copy number (GCN) variation event, identified by whole genome sequencing, consisted of a continuous segment which spans 50.2 Mbp or 10.9% of the Chinese hamster Chr2 (NC_048595.1:145.8 M-196.0 M). It was determined by ddPCR that pre-aged cells were haploid (~1 copy) and late generation cells were diploid (~2 copies). Cells with the gene duplication also showed an approximately 2-fold increase in the transcripts from these genes compared to the non-aged early generation, haploid cells. Moreover, a subset of 123 genes in the Chr2 amplified segment were determined to be commonly overexpressed in high growth phenotype cell lines. These late generation cells were associated with a high growth phenotype and were independent of the CHO host, and presence or type of recombinant product [24].

The purpose of this study was to investigate the contribution of a subset of the 123 genes involved in the Chr2 duplication event in relation to the high growth phenotype of late generation cells (>90 divisions). We hypothesised that some of the duplicated genes were likely to contribute more to the associated high growth phenotype and that manipulation of these in high or low growth phenotype cells would result in the ‘switching’ of cells from one growth phenotype to the other. We also hypothesised that identification of such age-correlated signature genes would allow targeted and rational cell engineering to design and deliver cell lines to promote stability, preventing low/high growth phenotype changes, which could be applied to the development of stable growth phenotype host cell lines. Through a combined approach of characterization of low and high growth phenotypes and investigating the impact of manipulation of selected Chr2 duplicated genes by knockdown and overexpression, we aimed to identify such target genes, evaluate the impact of manipulating these on low and high growth phenotype cells, and address the hypotheses outlined above.

Materials and methods

Shake flask cell culture

Suspension Pfizer C32 and C15 cells (previously described in Wei et al. [24]) were grown in commercially available CD CHO medium (Gibco 10743-029) with no additives. These cells express a model monoclonal antibody (mAb) and are derived from the same host but show different growth phenotypes as a result of an age-related gene copy number variation (duplication) in the C15 cell line. The cells were from stock generated in the previous study by Wei et al. [24] and assumed to be identical to those reported in that study. We did not reconfirm Chr2 duplication upon recovery of the cells from cryo-storage, but did reconfirm that Chr2 duplication region gene transcripts were increased approximately two-fold in expression in high growth phenotype cells compared to the low growth phenotype in line with the previous study as reported below. Derivatives of the cells engineered to overexpress key target genes were grown with hygromycin selection in CD CHO medium containing 1 mg/mL hygromycin. Cells were grown at 37 °C with shaking at 140 rpm in a 5% CO₂ atmosphere. Cells were seeded at

either 0.2 or 0.3×10^6 viable cells/mL in 125 mL Erlenmeyer flasks in 30 mL of CD CHO medium and were passaged every 3–4 days respectively. For growth studies, cells were seeded at 0.3×10^6 viable cells/mL in 30 mL CD CHO medium (and with 1 mg/mL hygromycin as appropriate) in 125 mL Erlenmeyer flasks incubated at 37 °C with shaking at 140 rpm in a 5% CO₂ atmosphere. For fed-batch cultures, liquid fed (CHO CD EfficientFeed B, Gibco) was added at 10% v/v on days 3, 6 and 9 of culture. Samples were removed to determine viable cell concentration and cell diameter using a ViCell XR instrument (Beckman Coulter) and for other analyses as detailed. Generation number (or population doublings) for the stability study were calculated as;

Generations = $\log(N_p/N_0)/\log 2$.

where N_0 = viable cell number at seeding and N_p = viable cell number at passage.

Total cumulative generations/generation number (or population doublings) = sum of all population doublings.

24-Deep Well Plate (24-DWP) cell culture

Cells were grown in 24-DWPs (Enzyscreen) in batch culture mode using the Duetz sandwich-cover system (Enzyscreen) with shaking at 300 rpm in a 5% CO₂ atmosphere at 37 °C for up to 11 days in a Kuhner shaking incubator. Cells were seeded at 0.3×10^6 viable cells/mL in a 2.5 mL culture volume of CD CHO medium. Sampling was undertaken such that no more than 10% of the total volume (0.25 mL) was removed from any well throughout culture.

qRT-PCR transcript analysis

For relative transcript amount analysis, 1×10^6 viable cells were harvested for RNA extraction (Qiagen, product 74104 & 79654) and subsequently treated for gDNA removal (Promega, M6101). mRNA levels were determined by qRT-PCR relative to the expression of *Cog1* as a house keeping gene using the NEB Luna Universal One-step RT-qPCR kit (E3005X) and a Thermo Fisher QuantStudio™ 3 cycler. A summary of primers used in this study for qRT-PCR analysis are listed in Supplementary Table 1.

Proteomic analysis

Filter aided sample preparation (FASP) was used for the preparation of samples for proteomic studies [26]. Lysis buffer (200 µL of 0.1 M Tris-HCl, pH 7.6, 0.1 M DTT, 4% w/v SDS) was added to each cell pellet (5×10^6 cells) and pipetted up and down to lyse the cells. Cell lysis was further aided by sonication with a probe sonicator (Sonics Vibracell) (3×30 s burst with 30 s rest between at 55% amplitude, all on ice). Lysates were then heated to 56 °C for 20 min before centrifugation at 13,000 rpm for 5 min to clarify the lysate. 40 µL of protein extract was then mixed with 400 µL of 8 M urea and then transferred to a Microcon10 kDa centrifugal filter (Sigma, MRCPT010) and centrifuged at 14000xg for 60 min at 20 °C. The flowthrough was discarded and the concentrate in the column diluted with 400 µL wash buffer 1 (8 M urea, 0.1 M Tris-HCl, pH 8.5) and centrifuged at 14000xg for 60 min at 20 °C. The flowthrough was discarded and the concentrate alkylated by the addition of 200 µL freshly prepared alkylation solution (0.05 M Iodoacetamide, 8 M Urea, 0.1 M Tris-HCl, pH 8.5). This was incubated in the dark at room temperature for 20 min followed by centrifugation at 14000xg for 60 min at 20 °C. The flowthrough was discarded, 200 µL of wash buffer 2 added (8 M urea, 0.1 M Tris-HCl pH 7.9) and the column centrifuged at 14000xg for 60 min at 20 °C. This step was then repeated. The Microcon 10 kDa centrifugal filter was then transferred to a new collection tube. The concentrate was diluted with 100 µL of freshly prepared 50 mM ammonium bicarbonate. Sequencing-grade trypsin (2 µg, Promega V5111) was added followed by 1 µL of 1% (v/v) ProteaseMax surfactant (Promega, V2071). The sample was then incubated at 37 °C for 3 h. A further 2 µg of trypsin was then added and incubated overnight at room temperature. The filter was then centrifuged at 14000xg for 40 min at 20 °C. The filtrate in the collection tube was retained, the filter rinsed with 50 µL 0.5 M NaCl and centrifuged at 14000xg for 20 min at 20 °C and the filtrate retained. This step was repeated once more.

Trifluoroacetic acid (TFA) (2.5%) was then added to each peptide sample to a final concentration of 0.1% (v/v) TFA. Samples were cleaned up using Pierce 100 μ L C₁₈ tips (ThermoFisher, 87784) according to manufacturer's instructions, evaporated to dryness in a vacuum centrifuge and then resuspended in 15 μ L of 5% (v/v) acetonitrile (ACN) and 0.1% (v/v) formic acid. Peptide concentration was estimated on a NanodropTM spectrophotometer and diluted to 1 μ g/ μ L in 5% (v/v) acetonitrile, 0.1% (v/v) formic acid. Samples were stored at -80 °C before nano LC-MS/MS analysis.

For mass spectrometry analysis, 1 μ g of peptide sample was spiked with 50 fmol yeast enolase peptide mix (Waters, 186002325) for absolute quantification. Peptides were separated on a HSS T3 Acquity column (Waters, 186008816) 75 μ m i.d. x 15 cm (1.8 μ m, 100 Å) using an Acquity M-Class UPLC (Waters); elution was performed with a linear gradient from 5 to 40% B over 90 min (solvent A, 0.1% (v/v) formic acid, solvent B, 0.1% (v/v) formic acid in acetonitrile) at a flow rate of 0.3 μ L/min and the eluate directed via a nanospray source into a Synapt G2-Si mass spectrometry instrument (Waters) with data collected in UDMSE mode.

The mass spectrometry data were imported into Progenesis QI for Proteomics (Non-Linear Dynamics) and peptides assigned using the reference proteome “*Cricetulus griseus* (Chinese hamster) (*Cricetulus barabensis griseus*) (CHO K1 cell line)” (Proteome ID UP000001075 – accessed 16-12-2021) from UniProt using an MSe Search algorithm (Ion Accounting/Apex3D) with a false discovery rate (FDR) of 4%. We used a more lenient FDR of 0.04 (as used by others for CHO cell growth phenotype proteomic mapping [27] and in other fields [28]) as opposed to the widely used 0.01 accepting this would give a higher number of false positives but potentially generate more identifications that could give a broader biological insight into the mechanisms involved. Automatic alignment was performed, and peak picking undertaken, by assessing all runs with a minimum chromatographic peak width of 0.1 min and maximum ion charge of 5. Following peak picking, MSe fragmentation data was used to automatically identify peptides by the Progenesis software, accounting for trypsin cleavage, carbamidomethyl modifications to cysteine residues and methionine oxidation. Maximum protein mass was set to 250 kDa with a maximum of one missed cleavage allowed. For peptide and protein assignments, a minimum of three fragments per peptide was required and a minimum of five fragments per protein. All assigned proteins contained at least one unique peptide. Absolute quantification using Hi-N was conducted, using the 3 most abundant peptides for each protein, with 50 fmol yeast enolase as calibrant (accession number P00924). Statistical analysis was performed within Progenesis Qi for proteomics using a one-way ANOVA. An adjusted p-value (q-value) was returned using an optimised FDR approach.

siRNA knockdown experiments

siRNA knockdown was performed on the same day of seeding cells. A summary of siRNA oligos are detailed in Supplementary Table 2. All Stars negative control siRNA (Qiagen 1027281) and Rad21 siRNA oligos were used as controls. The Rad21 siRNA oligos were used as a positive control and redesigned based on the screening study by [29] to initially optimize the knockdown protocol. Knockdown in 24-DWPs was carried out by seeding 0.3×10^6 viable cells/mL in 2.25 mL of CD CHO medium per well. A mixture of 1 μ L RNAiMAX (Fisher 10514953) and 15 nM siRNA (a smart pool composed of 3 siRNA at 5 nM each) were made up in 250 μ L CD CHO medium and incubated for 10 min at room temperature. The siRNA/ RNAiMAX complex was then added to wells in a dropwise motion and the plates then incubated in a Kuhner incubator at 37 °C with shaking at 300 rpm in a 5% CO₂ atmosphere.

Knockdown in 125 mL Erlenmeyer shake flasks was carried out by seeding 0.3×10^6 viable cells/mL in 27 mL CD CHO medium per flask. A mixture of 12 μ L of RNAiMAX and 15 nM siRNA (a smart pool composed of 3 siRNA at 5 nM each) were made in 3 mL CD CHO medium and incubated for 10 min at room temperature. The siRNA/ RNAiMAX complex was then added to the wells in a dropwise motion. The flasks were then incubated in a Kuhner incubator at 37 °C with shaking at 140 rpm in a 5% CO₂ environment.

SDS-PAGE and western blot analysis

1×10^6 viable cells were pelleted and lysed in 100 μ L of ice-cold lysis buffer (20 mM HEPES-NaOH, pH 7.2, 100 mM NaCl, 10 mM Na β -glycerophosphate, 0.5% Nonidet-P40 with 50 mM NaF, 1 mM activated Na₃ VO₄, 10 mg/mL leupeptin, 2 mg/mL pepstatin and 0.2 mM PMSF added just before use). 10 μ g of reduced whole lysate protein sample was run on 10% SDS-PAGE and western transfer undertaken onto nitrocellulose as previously described [3].

Antibodies for western blotting were sourced from the following. Sigma; β -actin (A5441). Cell Signaling; Cops5 (6895), H3K4me3 (9751 S). Abcam; Cxxc1 (ab198977). Anti-tubulin (Woods et al., 1989) was a kind gift from Professor Keith Gull, (University of Oxford, UK). eGFP (3E1) was a kind gift from CRUK. Secondary antibodies for immunoblot detection were anti-whole IgG (mouse or rabbit)-HRP conjugates (Sigma) and were detected by ECL (GE Healthcare) detection.

Generation of a bi-cistronic expression plasmid with a hygromycin-eGFP fusion gene and cloning of target genes

A Pfizer proprietary plasmid designed for site-directed recombination into the host cell genome using Flp recombinase was used in this study (although the Flp recombinase activity was not ultimately used). The plasmid harbours two multiple cloning sites (MCS) under the control of hCMV-IE-intronA promoters and a blasticidin resistance cassette. The cell lines used in this study (C15 and C32) have genes for the light chain and heavy chain of a mAb integrated into the genome landing pads. We developed a system with an alternative selection marker in MCS2 (HygR-eGFP fusion) and MSC1 contained the Chr2 target genes of interest.

The fusion protein gene was cloned into MSC2, HygR-eGFP with a GS GS linker between the hygromycin resistance gene and eGFP gene. The fusion protein gene was generated using overlap extension PCR using Phusion™ High-Fidelity DNA polymerase (ThermoFisher, F530S). In PCR round 1, HygR and eGFP fragments were amplified from plasmids containing these genes using primers encoding extensions complementary to the other fragment (Supplementary Table 3). PCR products were separated by agarose gel electrophoresis, the DNA fragments excised from the gel and DNA extracted using a gel extraction kit (Qiagen, 28704). The fusion protein gene was then generated and amplified in PCR round 2 using the products of PCR round 1 as templates. The PCR products were separated by agarose gel electrophoresis and DNA bands of the correct size were excised from the gel. The fusion PCR products were digested with restriction enzymes *Bam*HI (ThermoFisher, FD0054) and *Xho*I (ThermoFisher, FD0694) and ligated into MCS2 of the destination vector, which was also digested with *Bam*HI and *Xho*I, using T4 DNA ligase (Promega, M1801). Colonies obtained after transformation of competent DH5 α *E.coli* were used to amplify plasmid DNA (Qiagen, 27104). Successful cloning of the HygR-eGFP fusion was confirmed by Sanger sequencing (Genewiz).

Target genes for further investigation (arrowed in Fig. 5) were selected from the siRNA knockdown screen based upon the following criteria; (1) All those genes whereby knockdown, regardless of knockdown efficiency/amount, had resulted in relative growth reduction to 0.5 or less of the control (*Atp5f1a*, *Rrs1*, *Lman1*, *Cxxc1*, *Seh1l*, *Rpl17*, *Rrps14*, *Rps20*), (2) Two examples where knockdown was approximately 75% or better and there was little/no impact on growth compared to the control (*LOC100758579*, *Psmg2*), (3) Examples across all screened genes whereby there was modest knockdown (expression of gene between 0.25 and 0.35 of control levels) and modest impact on growth (between 0.85 and 0.60 relative to control; *Napg*, *Cops5*, *Nars1*), (4) An example of little knockdown (0.74 relative to control) and modest impact on growth (0.71 of control, *Rbm22*). Target genes (encoded protein function detailed in Supplementary Table 4) were amplified by PCR from cDNA generated from the C15E cell line (Primers in Supplementary Table 3) and cloned into MCS1 using Sequence and Ligation Independent Cloning (SLIC) [30]. Briefly, the vector was digested with *Nhe*I (ThermoFisher, FD0973), and the linearised plasmid purified (Qiagen, 28704). Target genes were amplified from cDNA by PCR using primers with 15 bp homology to the plasmid backbone digested with *Nhe*I. Linearised vector and the target gene PCR product

were incubated at room temperature with T4 DNA polymerase (NEB, M0203S) for 2.5 min before being placed on ice for 15 min followed by transformation of competent DH5α *E. coli*. The *mCherry* gene was cloned into MCS1 as a control and to allow determination of the number of cells expressing both mCherry and eGFP.

Generation of stable cell pools overexpressing target Chr2 genes

To generate stable cell pools overexpressing target Chr2 genes, the target gene (MCS1) /hygro-eGFP fusion selection marker (MCS2) plasmids were linearized with *NruI* (NEB, R3192), purified by ethanol precipitation and dissolved in TE for electroporation. Electroporation was carried out on a Bio-Rad Genepulser Xcell electroporator. In triplicate, 20 µg of linearized plasmid in 100 µL of TE and 1×10^7 host cells were electroporated at 300 V and 900 µF in a 0.4 cm gap cuvette. The cells were then transferred into 20 mL of CD CHO medium in T75 flasks. The flasks were incubated at 37 °C in a static incubator in a 5% CO₂ gas environment. 24 h post-transfection, 2 mL of CD CHO medium with hygromycin to result in a 1 mg/mL final concentration of hygromycin was added. The triplicate transfections were then pooled and cell counts carried out using a ViCell-XR instrument 14 days post-transfection to assess recovery. Once viable cell number and culture viability was deemed to have recovered sufficiently, cells were transferred to shake flasks and treated as described under cell maintenance.

Flow cytometry

The percentage of cells expressing eGFP following transfection with eGFP or a hygR-eGFP fusion plasmid was determined by flow cytometry on a FACS-Jazz flow cytometer (BD Biosciences).

Antibody titre measurement

Supernatant antibody titres were measured using a Amperia™ antibody quantification system (Abselion) using human IgG1 kappa-HRP (Southern Biotech, 0151 K-05) as the detection antibody. Supernatants were diluted in PBST and the concentration of antibody calculated from a standard curve generated using Tocilizumab 0.93–120 µg/mL.

Results

A CHOK1 age-related cell growth phenotype previously reported in proprietary medium, fed-batch culture in bioreactors is reproducible in shake flasks and 24 deep well plates in commercial medium and feed

Initially we set out to establish a system that would reproduce the previously reported age-correlated chromosome 2 duplication and growth rate acceleration observed under fed-batch culture conditions using proprietary medium and feeds [24]. For this, we used 4 model cell lines from Pfizer expressing the same monoclonal antibody (mAb); these are designated C15E, C15L, C32E, C32L. C32 and C15 refer to clonal recombinant cell lines derived from the same CHOK1 host cell line and expressing the same mAb. Cell lines noted with E and L refer to early and late generation numbers with late versions being aged compared to the early generation cell line (25 generations for E v 100 generations for L post-cloning as described in Wei et al. [24]). C15 exhibited high growth phenotype and chromosome 2 duplication event at early generation, whilst the C32 cell line exhibited low growth phenotype and lacked the chromosome 2 duplication at early generation. The details by which the C32 and C15 cell lines were generated is outlined in Wei et al. [24].

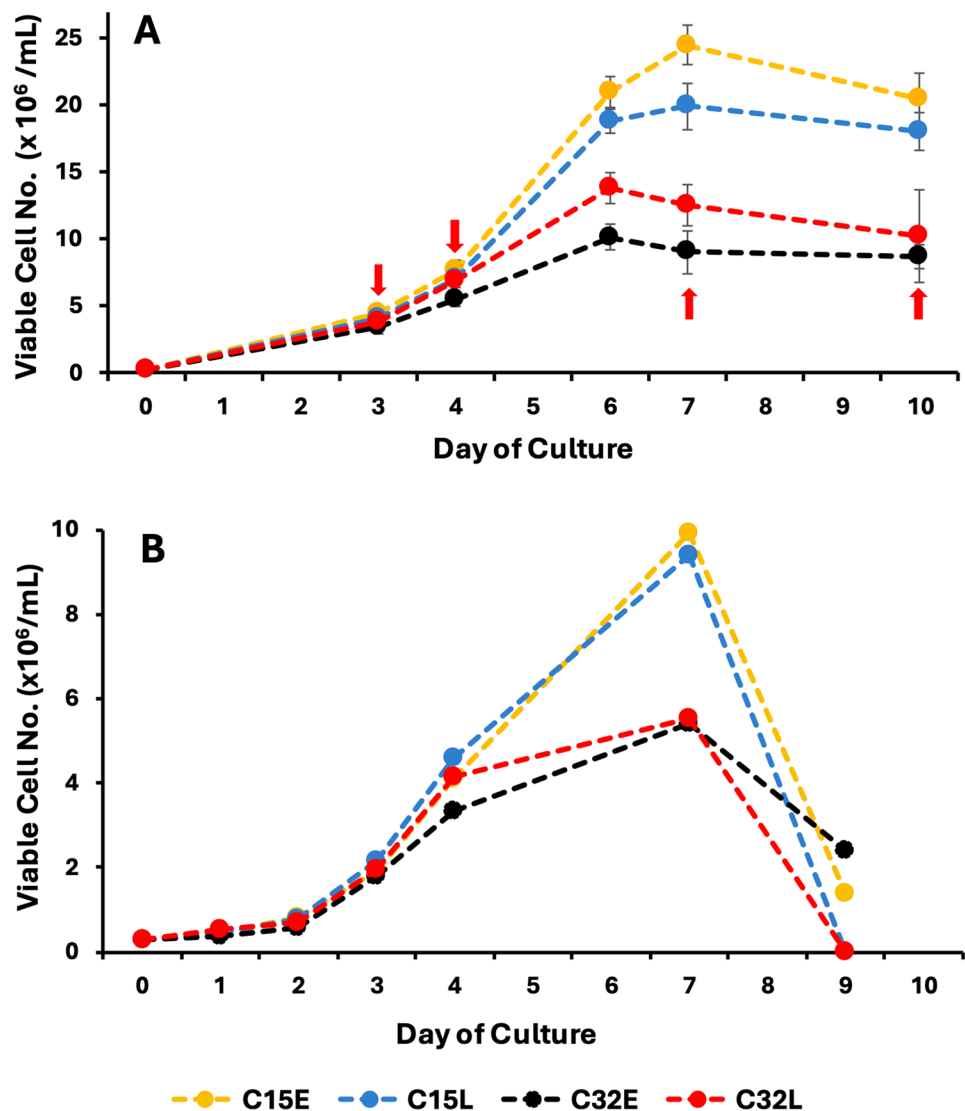
When grown using the commercial medium and feed under fed-batch conditions, the age and GCN growth phenotypes observed mirrored those from the previous study and those expected based upon the presence/absence

of the chromosome 2 duplication event that was shown to correlate with growth phenotype. The two C15 cell lines reached a viable cell concentration approximately twice that of the C32 cell lines with the late generation C32 cell line (C32L) reaching a higher viable cell concentration than the early (C32E) (Fig. 1A). The difference in growth phenotype was not obvious until after the first feed on day 3 of culture, was pronounced by day 6, and was then maintained until the end of the culture on day 10 (Fig. 1A). Cell and supernatant samples were collected on days 3, 7 and 10 (and day 4 for proteomic analysis) of culture for subsequent analysis. These growth differences were also observed under batch culture conditions in both shake flasks (data not shown) and in a 24 deep well plate (DWP) format (Fig. 1B). Days at which statistically significant differences in viable cell number were observed are reported in Supplementary Table 5.

The high growth phenotype observed in fed-batch shake flask correlates with increased chromosome 2 gene transcript expression throughout culture

Previously it was reported that the age-correlated growth phenotype was associated with the enhanced transcript expression (approximately 2-fold) of 123 genes mapped to the chromosome 2 duplication event [24]. In order to confirm that the fed-batch culture system developed during this study was reproduced at both the phenotypic and

Fig. 1 Growth phenotype profiles of CHO cell lines used in this study. **(A)** Fed-batch shake flask data illustrating growth differences between hosts. Statistically significant differences in viable cell number were evident on days 6 and 7 between high and low growth phenotypes (ANOVA with post hoc Tukey HSD, $p < 0.01$, see Supplementary Table 5). Feed was added on day 3 and 6. Sampling on days 3, 7 and 10 for analysis was undertaken as indicated by arrows. An additional sample was taken on day 4 for proteomic analysis. $n = 4$, error bars show ± 1 SD. **(B)** Growth phenotype of the same cell lines as in **(A)** grown in batch culture in a 24 deep well plate (24-DWP) in 2.5 mL volume when seeded at 0.3×10^6 cells/mL. 24-DWPs were used for screening assays as indicated. $n = 2$



transcript level, we applied qRT-PCR analysis to a subset of genes based on potential function, previous transcript analysis in the original study, expression at different abundancies and being dispersed across the duplicated segment. This analysis confirmed that the target chromosome 2 duplication transcripts investigated were typically and consistently found to be increased in expression in the C15 cell lines by approximately 2-fold compared to the C32E cell line on day 3 of fed-batch culture (Fig. 2 and Supplementary Fig. 1), even though there was no discernible change in growth phenotype this early in culture (Fig. 1A). Of note, when we investigated a control transcript used in this study, *Rad21* which is not a chromosome 2 duplication gene (selected based upon the study of [29]), there was no obvious relationship between growth phenotype and the amount of *Rad21* transcript amounts between C15 and C32 cell lines on day 3. We therefore deemed *Rad21* an appropriate control for our siRNA knockdown optimization (see below).

We then extended the study to investigate the expression profile of the selected chromosome 2 target transcripts across the fed-batch culture. In all cases the approximate 2-fold increase in transcript amounts in the C15 cell lines (with the duplication event) compared to the C32E cell line (without) was maintained at day 7 of culture. This was not observed for the non-chromosome 2 *Rad21* transcript in line with the day 3 data (Fig. 2). Interestingly, for a number of the chromosome 2 targets (*Rnmt*, *Pias2*, *Sehl1*, *Cops5*, *Cxxc1*) there was an increase in the expression of the transcripts in the C15 cell lines at day 10 of culture with a 3–6-fold increase in expression compared to the day 3 C32E control cell line. In all cases the expression of the chromosome 2 target transcripts in the C15 cell lines was maintained at elevated levels at day 10 of fed-batch culture compared to the C32E control (Fig. 2 and Supplementary Fig. 1).

Proteomic analysis shows that protein expression of specific, but not all, chromosome 2 gene copy number (GCN) signature target genes from the age-related high growth phenotype cell lines reflects increased transcript expression

We next applied mass spectrometry-based proteomics to map protein expression from chromosome 2 GCN signature genes in the low and high growth phenotype CHO cell lines. The proteome profile of high and low growth phenotype cultures were analysed on days 3, 4, 7 and 10 of fed-batch culture (Fig. 1A). Absolute abundance was estimated using 50 fmol yeast enolase peptide mix as a calibrant protein. Significant differential expression was determined by ANOVA at $p < 0.05$ (or with an adjusted p -value (q -value) using an optimised FDR approach $q < 0.05$). Uniprot accession numbers were converted to entrez gene IDs using DAVID (DAVID Functional Annotation Bioinformatics Microarray Analysis (ncifcrf.gov)) for cross reference to the chromosome 2 upregulated genes. Approximately 75% of Uniprot protein accession numbers were successfully converted to gene IDs and approximately 30% of the 123 duplicated chromosome 2 genes were amongst these detected. All chromosome 2 duplication event associated proteins positively identified are listed in Tables 1, 2 and 3. As expected, not all proteins encoded by the chromosome 2 duplication region were identified in the proteomic study, likely due to low or no expression, poor ionization of peptides, technical limitations/quantitative limits, sensitivity and potential incorrect or absent annotation in the database.

Volcano plots comparing protein abundance in low (C32E) and high (C15E) growth phenotype cell lines (Student t-test, two-sided with Benjamini-Hochberg adjustment for multiple comparisons) identified a number of chromosome 2 duplication encoded proteins that were upregulated (> 1.5 fold) in the high growth phenotype ($p < 0.05$) across culture in line with the general 2-fold increase in transcript amounts (Fig. 3B). The chromosome 2 duplication event encoded proteins with significantly higher abundance in the high growth phenotype compared to the lower growth phenotype are summarized in Table 1. Figure 3A highlights a number of proteins derived from the Chr2 target duplication genes detected in the study that were significantly more abundant in C15E high growth phenotype than in the C32E low growth phenotype cell line. However, there was not an increase in abundance of all proteins encoded by chromosome 2 duplication genes as shown in Fig. 3A (and Supplementary Figs. 2–5), and thus the correlation between an increase in transcript amount due to the duplication event and

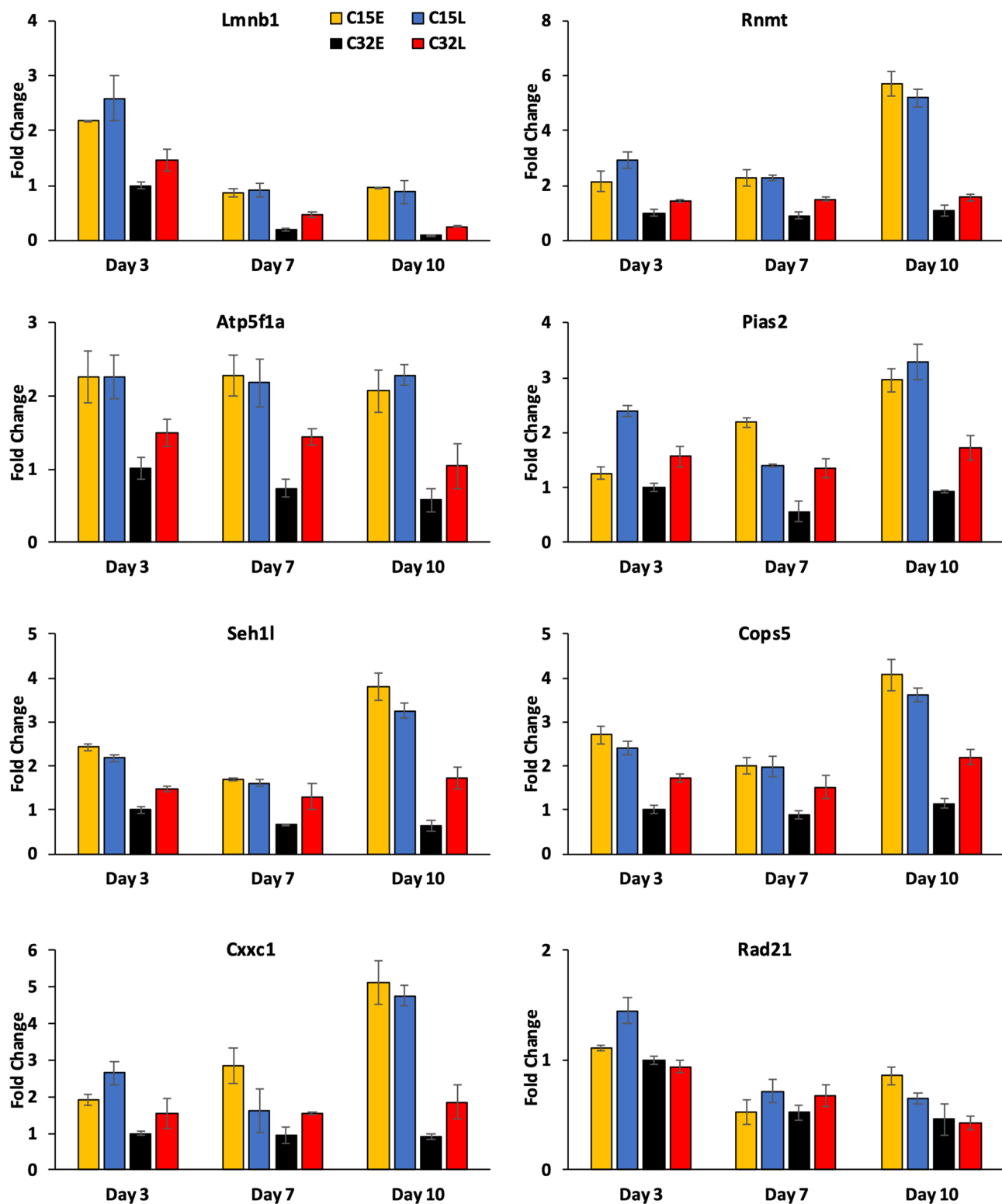


Fig. 2 Transcript (mRNA) level fingerprints of specific chromosome 2 duplication targets and a control Rad21 target in age-correlated CHO cell low (C32) and high (C15) growth phenotypes throughout different phases of fed-batch culture as indicated in Fig. 1. Samples were taken of host cells on day 3, 7, and 10 of culture, and the relative amounts of the indicated target transcripts determined by qRT-PCR. Fold change in transcript amounts is shown relative to day 3, C32E samples. $n = 3$, error bars show ± 1 SD

Table 1 Differentially expressed chromosome 2 duplicated encoded proteins with higher abundance in high growth phenotype cells compared to low growth phenotype cells

Accession	Description (Gene)	Day3 Fold change	Day4 Fold change	Day7 Fold change	Day10 Fold change
G3HMD8	tubulin beta 6 class V (Tubb6)	1.756	2.17	NS	NS
G3HDS7	syndecan binding protein (Sdcbp)	3.06	4.93	NS	NS
G3H5G6	thioredoxin like 1 (Txnl1)	2.18	NS	3.32	3.58
G3HKA5	sorting nexin 2(Snx2)	1.88	NS	NS	1.93
G3I1P0	isochorismatase domain containing 1 (Isoc1)	2.02	2.32	1.54	NS
G3HHE3	leucyl-tRNA synthetase 1 (Lars1)	1.57	ND	1.97	2.01
G3HKB0	peptidylprolyl isomerase C (Ppic)	2.13	3.80	2.21	ND
G3H5H2	NSF attachment protein gamma (Napg)	1.65	1.95	NS	NS
G3H1B7	proline rich coiled-coil 1 (Prcc1)	NS	5.17	3.62	2.81
G3HUI7	lectin, mannose binding 1 (Lman1)	NS	1.70	NS	2.60
G3HSQ5	RNA guanine-7 methyltransferase (Rnmt)	1.88	NS	NS	NS
G3HVE2	MCC regulator of WNT signaling pathway (Mcc)	ND	3.38	ND	ND
G3H1B2	aldehyde dehydrogenase 7 family member A1 (Aldh7a1)	NS	2.03	1.77	NS
G3I477	RAB27B, member RAS oncogene family (Rab27b)	NS	2.99	NS	ND
G3H1B3	lamin B1 (Lmnbl)	NS	NS	2.82	2.50
G3HTR9	malic enzyme 2 (Me2)	NS	NS	2.31	2.76
G3HTM9	eukaryotic translation initiation factor 1 A (LOC100758579)	NS	NS	NS	2.28
G3I2D3	ribosomal protein S20 (Rps20)	NS	NS	NS	2.09
G3HPU1	aspartate beta-hydroxylase (Asph)	ND	NS	NS	1.66
G3HAE2	ATP synthase F1 subunit alpha (Atp5f1a)	NS	NS	2.26	NS
G3HS34	platelet derived growth factor receptor beta (Pdgfrb)	ND	NS	1.51	NS
G3H5F7	asparaginyl-tRNA synthetase 1 (Nars1)	NS	NS	2.94	NS

NS=no significant change in expression/not differentially expressed. ND=not detected

Table 2 Differentially expressed chromosome 2 duplicated encoded proteins with lower abundance in high growth phenotype cells compared to low growth phenotype cells

Accession	Description (Gene)	Day3 Fold change	Day4 Fold change	Day7 Fold change	Day10 Fold change
G3IAC6	interferon-inducible GTPase 1 (LOC100760047)	ND	0.22	NS	NS
G3HTM9	eukaryotic translation initiation factor 1 A (LOC100758579)	NS	NS	0.42	NS
G3H5H2	NSF attachment protein gamma (Napg)	NS	NS	0.38	NS
G3ICE3	COP9 signalosome subunit 5 (Cops5)	ND	ND	ND	0.58

NS=no significant change in expression/not differentially expressed. ND=not detected

subsequent protein abundance was not conserved across all Chr2 targets. Indeed, in a small number of cases the high growth phenotype cells had decreased abundance of Chr2 duplicated encoded proteins compared to the low growth phenotype (Table 2). This is further demonstrated in Fig. 4, whereby the transcript and associated protein abundance of selected Chr2 targets are compared side-by-side. We note that of the two key Chr2 targets investigated in detail below, Cops5 and Cxxc1, only Cops5 was detected in the proteomic analysis, although both were observed by western blot (see sections below). Cxxc1 is a transcription factor and likely to be expressed at low levels, potentially explaining why we did not observe this protein in the proteomic data. Collectively, these data show that although duplication of many of the Chr2 genes was associated with an approximate 2-fold increase in transcript and protein abundance of the products the genes code for, this was not the case for all duplicated genes.

Table 3 Chromosome 2-encoded duplicated proteins detected but no significant difference in abundance between high and low growth phenotypes

Accession	Description (Gene)
G3IA38	DNA polymerase iota (Poli)
G3HMS8	ribosomal protein L17 (Rpl17)
G3HMX5	proteasome assembly chaperone 2 (Psmg2)
G3HHY2	hydroxysteroid 17-beta dehydrogenase 4 (Hsd17b4)
G3GU30	60 S ribosomal protein L6-like (LOC100750478)
G3HHD9	protein phosphatase 2 regulatory subunit beta (Ppp2r2b)
G3IDA4	Proenkephalin (Penk)
G3HSY4	mitochondrial ribosomal protein L15 (Mrpl15)
G3H1B5	chromosome 2 C5orf63 homolog (C2H5orf63)
G3H8N9	SMAD family member 2 (Smad2)
G3HMW3	G protein subunit alpha L (Gnal)
G3HTS2	60 S ribosomal protein L7a-like (LOC100770169)
G3I100	60 S ribosomal protein L27 (LOC100768273)
G3I101	chondroitin sulfate synthase 3 (Chsy3)
G3I2D1	LYN proto-oncogene, Src family tyrosine kinase (Lyn)
G3HSY7	ATPase H+ transporting V1 subunit H (Atp6v1h)
G3IDL9	coiled-coil domain containing 192 (Ccdc192)
G3HSR1	SEH1 like nucleoporin (Seh1l)
G3HTM8	transmembrane p24 trafficking protein 7 (Tmed7)
G3HUJ3	phorbol-12-myristate-13-acetate-induced protein 1 (Pmaip1)
G3I8H1	alcohol dehydrogenase iron containing 1 (Adhfe1)
G3IAK8	glycerol kinase (LOC100764600)
G3HYK6	dihydropyrimidinase like 3 (Dpysl3)
G3HS48	interferon-inducible GTPase 1 (LOC100763791)
G3HMX5	proteasome assembly chaperone 2 (Psmg2)
G3I3R9	carboxypeptidase A6 (Cpa6)
G3IAC0	interferon-inducible GTPase 1(LOC100762377)
G3IA36	chromosome 2 C18orf54 homolog(C2H18orf54)

Functional screening by siRNA knockdown identifies target genes from the high growth phenotype chromosome 2 duplication signature whose knockdown impacts growth phenotype

We used an siRNA screen approach to identify genes from the Chr2 duplication event whose knockdown impacts the observed phenotype in high growth cell lines. A method for the transcript knockdown of target genes was established in a 24 deep well plate (24-DWP) format using the non-chromosome 2 control target transcript of Rad21 based on the previous study by Klanert et al. [29]. This optimisation step tested longevity of gene suppression and assessed conditions for minimal impact on culture viability with maximum gene knockdown (data not shown). The optimisation process established that day 5 post-siRNA transfection (120 h post-transfection) was best to evaluate if growth had been impacted whilst the siRNA knockdown was still present. We also confirmed that knockdown of Rad21 was maintained for at least 120 h in CHO cells (Supplementary Fig. 6). Rad21, whose knockdown has previously been shown to reduce cell growth [29], was thereafter used as a positive siRNA control in the 24 deep well plates whilst a negative scrambled siRNA control was also run on all plates.

Targets from the 123 genes involved in the previously reported Chr2 duplication event were selected to screen for impact on growth phenotype by siRNA knockdown based on RNAseq transcript analysis [24], confirmation by qRT-PCR to have increased transcript amounts under the conditions used in this study and reported above, and based upon the proteomics analysis carried out during this investigation. siRNA oligos were then designed for all target genes with 3 oligo sequences that targeted different regions of the sequence being designed (Supplementary Table 2). These oligos were then used as a pool in the 24 deep well plate system to screen the impact of knockdown of the target genes on the growth phenotype of the C15E high growth phenotype cell line. Samples were

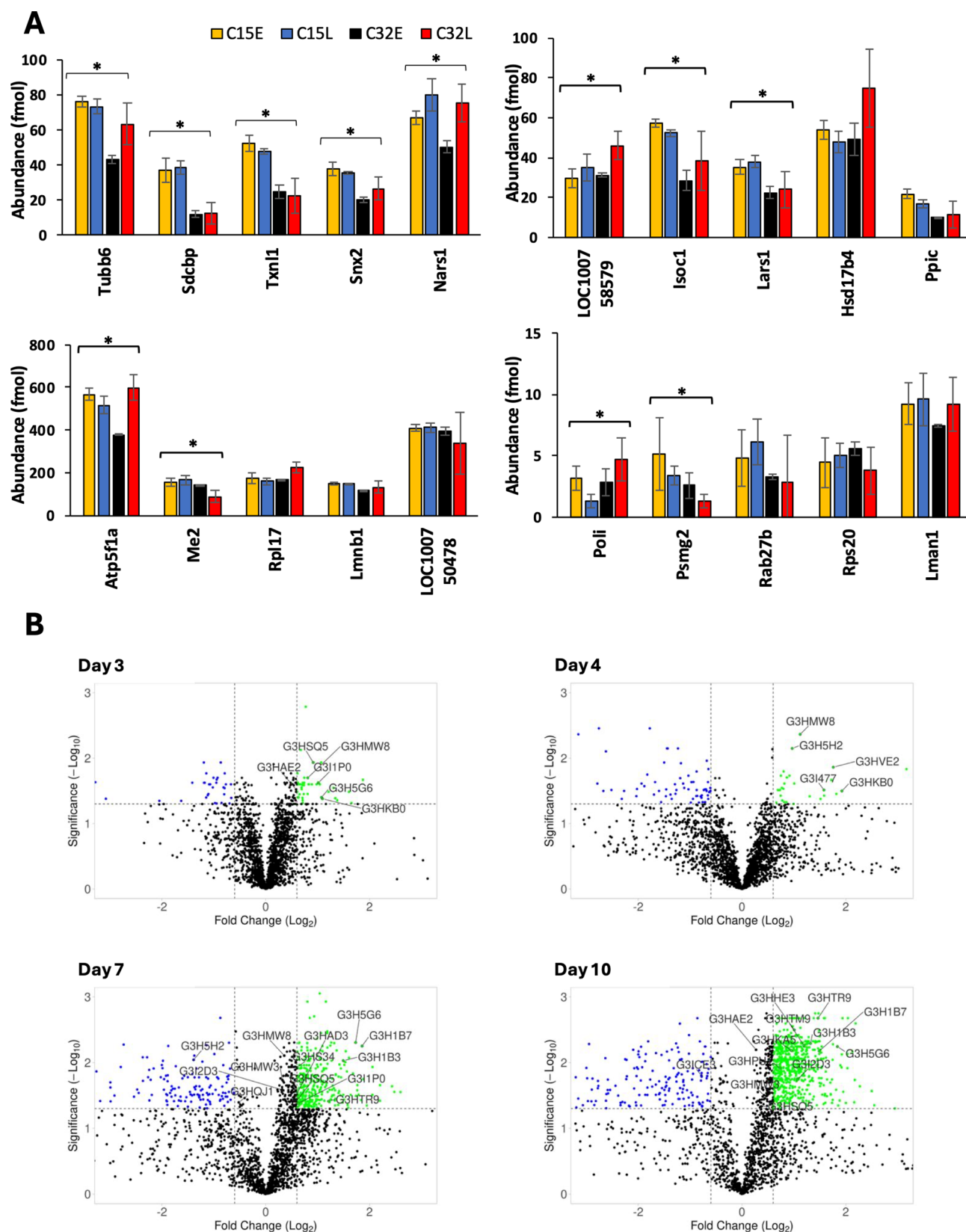
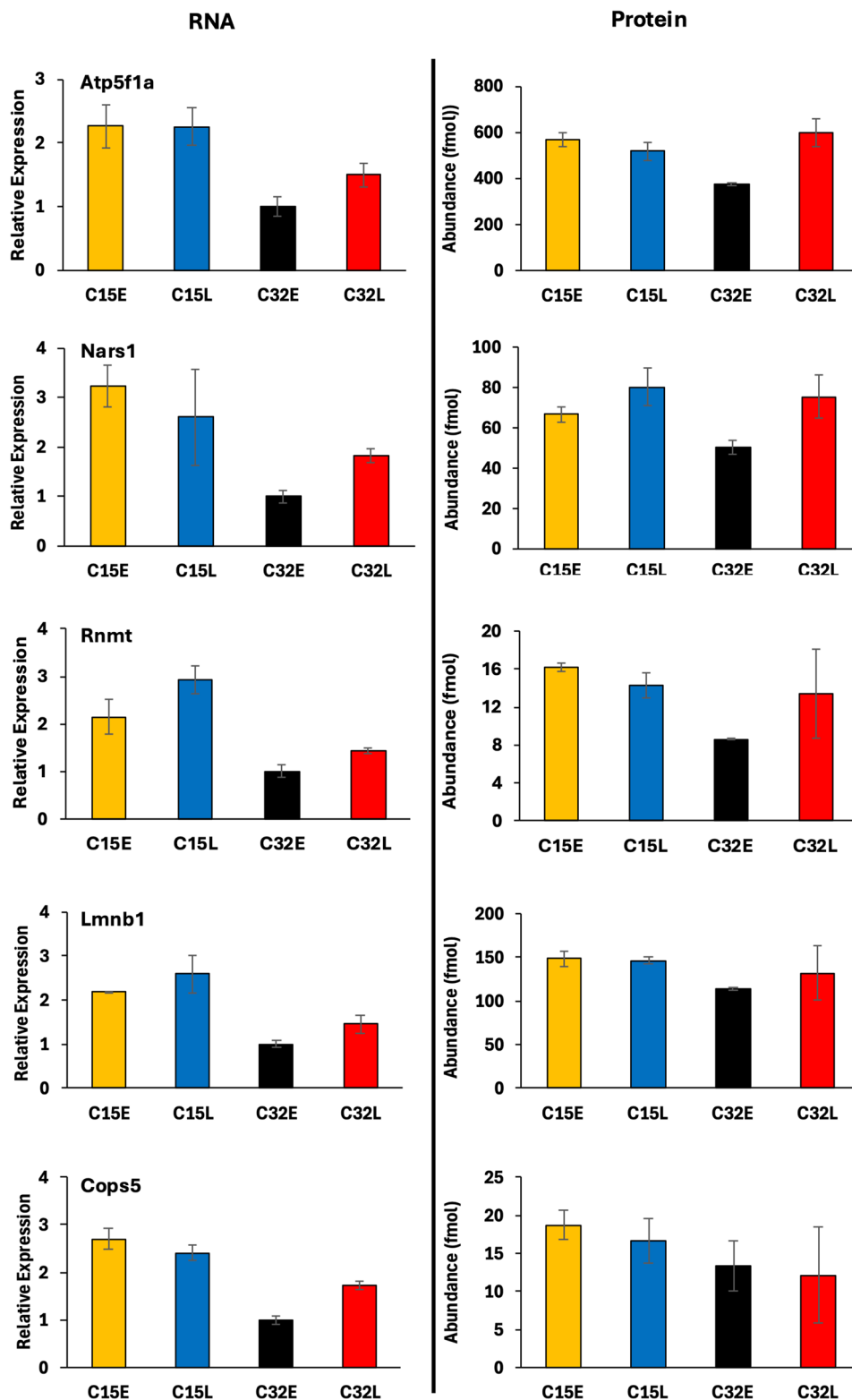


Fig. 3 Proteomic profiling of protein expression in CHO cell lines with different age-related growth phenotypes. **(A)** Protein abundance of chr_2 gene products in low and high growth phenotype cells (C15E, C15L, C32E and C32L) from day 3 of fed-batch culture measured by LC-MS/MS. Statistically significant differential expression across the 4 cell lines was determined by ANOVA. * $q < 0.05$ (FDR adjusted p value). Remaining data for days 3, 4, 7 and 10 provided in Supplementary Figs. 2–5. **(B)** Volcano plots of proteins identified by LC-MS/MS on samples from days 3, 4, 7 and 10 of growth, showing abundance in low (C32E) and high growth phenotypes (C15E). Each point represents the fold change between C32E and C15E abundance. The horizontal line represents a significance $p < 0.05$ (2-sided t -test, adjusted for multiple comparisons using the Benjamini-Hochberg procedure). The vertical lines represent fold change > 1.5 . Proteins with higher abundance in the high growth phenotype are in green. Proteins with higher abundance in low growth phenotype are in blue. Chr_2 protein products are labelled. (Plots generated using VolcanoPlot; “VolcanoPlot” is a web app for creating, exploring, labelling and

Fig. 4 Comparison of transcript and protein levels for selected chromosome 2 targets on day 3 of fed-batch culture from the different cell lines as indicated in Fig. 1. Transcript levels for all targets follow a similar pattern, however protein levels were more variable. $n=3$, error bars show ± 1 SD

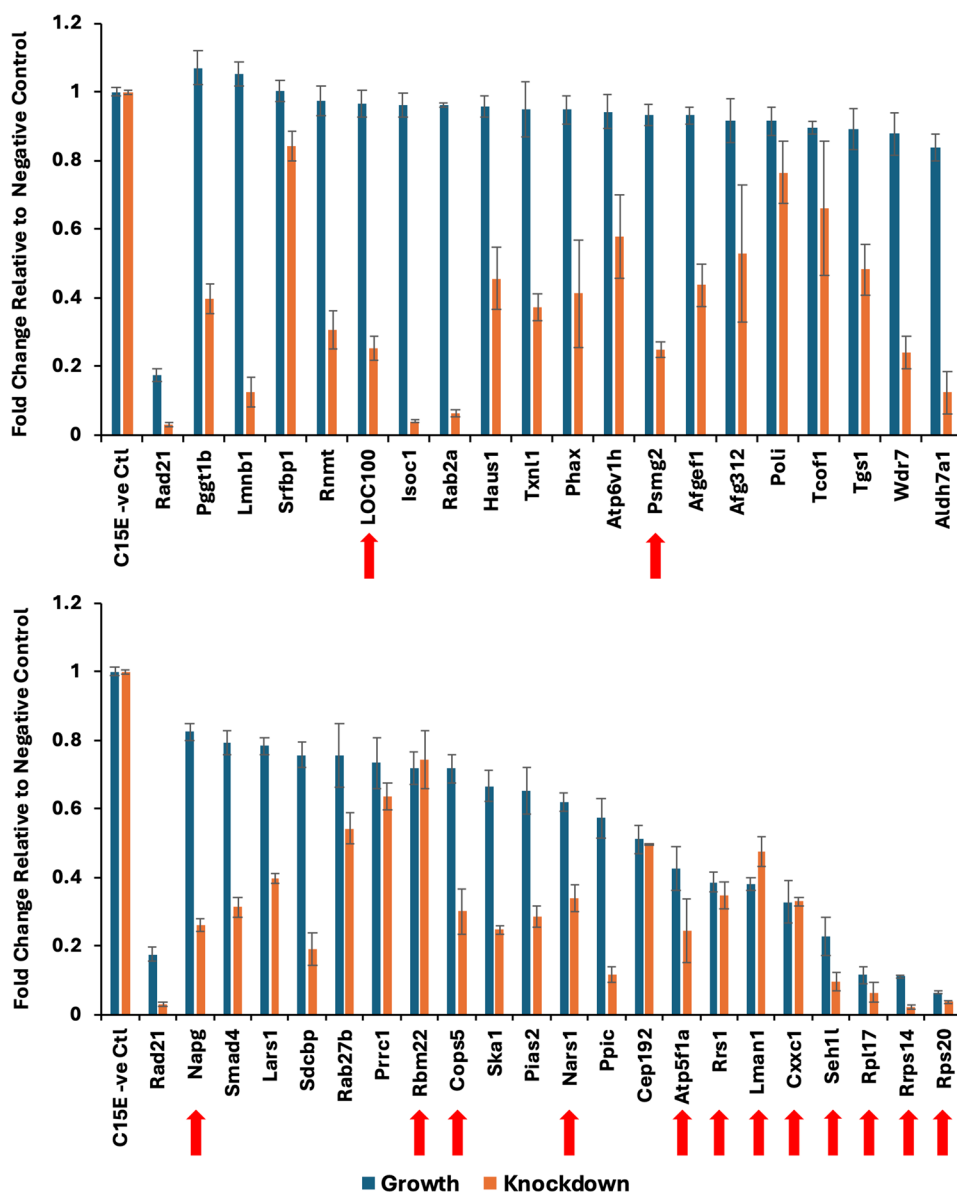


taken 5 days post-siRNA transfection for viable cell number, culture viability, and RNA extraction (for qRT-PCR based target transcript analysis).

qRT-PCR analysis of the knockdown efficiency for each target gene showed a wide range of knockdown had been achieved for the different genes when compared to the scrambled siRNA negative control (Fig. 5). The positive control Rad21 siRNA had >90% transcript knockdown and this was associated with in a large decrease in viable cell number (Fig. 5). When investigating the Chr2 target gene transcripts, there were groups where the knockdown was inefficient and there was no impact on growth as determined by the viable cell number on day 5 post-siRNA transfection (e.g. *Srfbp*, *Poli*, Fig. 5). There was also the reverse case where there was a large knock-down in transcript but no impact on cell number observed (e.g. *Lmnbl*, *Isoc1*, *Rab2a*, *Aldh7a1*, Fig. 5). From these data we concluded that transcripts in these groups either had little impact or the knockdown was insufficient to observe any impact using the screening approach and siRNAs described.

There were also target transcripts that when knocked down had a large impact on growth (e.g. *Rsp20*, *Rrps14*, *Rpl17*, *Seh1l*) and others whereby the knockdown was not as efficient but there was an obvious impact on growth (e.g. *Lman1*, *Cxxc1*) (Fig. 5). From these data we selected 14 targets to investigate further (highlighted with

Fig. 5 siRNA knockdown of chromosome 2 selected targets and the impact on age-related high growth phenotype in the C15E high growth phenotype cell line. Cells were transfected with a pool of 3 oligo sequences that targeted different regions of each chromosome 2 target transcript and then 120 h post-siRNA transfection cells harvested. siRNA knockdown of targets was confirmed by qPCR whilst growth impact was monitored by determining the viable cell number and culture viability 120 h post-siRNA transfection using a Vicell instrument. The figure shows the C15E cell line as a control and Rad21 as a positive control. The change in transcript amount and growth for all targets is shown relative to a C15E negative scrambled siRNA control. Targets are ranked from left to right based on relative growth compared to the control upon siRNA knockdown. Red arrows indicate gene targets that were selected to be taken forward for cell engineering as described in the text. $n=3$, error bars show ± 1 SD. Note: Loc100 short for LOC100758579



a red arrow in Fig. 5), selecting a range of siRNA knockdown efficiency and growth impacts in line with the criteria outlined in the methods section. This included subsets from all groups described above. Interestingly, although there was a large decrease in the viable cell number observed upon knockdown of many targets, inferring an impact on growth phenotype, the culture viability remained high in almost all cases suggesting that the knockdown did not result in cell death but prevented or slowed cell proliferation. Cell size was also found to be approximately the same across the different knockdown targets screened (data not shown). Finally, in order to confirm that the results observed in the 24-DWP siRNA knockdown were consistent at the shake flask scale, we repeated the knockdown experiment in shake flasks with the controls (*Rad21* positive control, scrambled siRNA negative control) and 3 of the Chr2 target gene transcripts (*Seh1l*, *Pias2*, and *Rnmt*). An obvious impact on growth 5 days post-siRNA knockdown was observed for the control *Rad21* and for *Pias2* and *Seh1l*, in agreement with the 24-DWP data (Fig. 6). There was also no more than a subtle impact on growth upon *Rnmt* knockdown, again in agreement with the 24-DWP screen data (see Figs. 5 and 6).

Over-expression of Chr2 targets in low generation, low growth phenotype C32E cells and the impact on growth phenotype

In order to generate C32E low growth phenotype cells over-expressing selected individual Chr2 targets we used a Pfizer proprietary bi-cistronic expression vector with two multiple cloning sites (MCS), with each gene being driven by an independent CMV promoter. So that successfully transfected cells could be easily identified, we generated a hygromycin resistance-eGFP fusion protein encoding gene (with a small GSGS linker between the two genes) that was cloned in the second MCS (MCS2). The resulting vector was linearised and transfected into C32E mAb expressing cells and hygromycin selection added 24 h after transfection. As a control we transfected a pcDNA3.1 plasmid containing the eGFP gene driven by a CMV promoter. When the cell pools that recovered post-transfection were analysed by western blot for eGFP expression and qRT-PCR for eGFP transcript amounts there was a band of the expected size for each protein present in the cell lysates of the pools and transcript present in all those samples where it was expected (Supplementary Fig. 7). This confirmed that the hygromycin-eGFP fusion protein was able to confer hygromycin resistance. Flow cytometry analysis confirmed that eGFP fluorescence was also present and that when a second target gene was placed in MCS1, in this case mCherry, the majority of cells that recovered from transfection and selection expressed both mCherry and eGFP (Supplementary Fig. 8). We therefore used the hygromycin-eGFP fusion selection system with target genes cloned into MCS1 for the generation of C32E low growth phenotype cells over-expressing 14 selected Chr2 targets to determine if they had any impact on the growth phenotype of C32E.

We initially assessed the impact of over-expression of the 14 Chr2 targets individually on the low growth phenotype of the C32E cell line in the 24-DWP batch culture assay (Fig. 7A and Supplementary Fig. 9). To achieve this, C32E cell pools stably over-expressing each target were generated using the hygromycin-eGFP fusion selection system. These were then cultured for 11 days in the 24-DWP batch culture system and the viable cell concentration and culture viability monitored throughout (Fig. 7A and Supplementary Fig. 9). Overexpression at the transcript level of the targets was confirmed by qRT-PCR on Day 11 of culture from the 24 DWPs (and subsequently on Day 3 of culture from shake flask experiments, Supplementary Fig. 10 and Fig. 7C). As controls, the original C32E cell line and a cell pool engineered to over-express mCherry instead of a Chr2 target were included. In all cases the C32E and mCherry control cells reached similar maximum viable cell concentrations, although the mCherry expressing cell pool maintained viable cell number for a prolonged period compared to the C32E control (Fig. 7A and Supplementary Fig. 9). The overexpression of many of the Chr2 target genes investigated had no obvious impact on the cell growth phenotype in the 24-DWP assay and the profiles were similar to that of the mCherry control cell pool. The *Atp5f1a* and *Rps20* engineered pools did grow more rapidly and reach a higher maximum viable cell number, whilst the *Rrs1* engineered cells also eventually reached a higher maximum viable cell number that was maintained for longer than the controls (Supplementary Fig. 9). However,

Fig. 6 Validation of siRNA knock down of chromosome 2 target impacts on age-related growth phenotype in C15E cells in Erlenmeyer 125 mL shake flasks (**A** and **B**) compared to that observed in 24 deep well plates (**C**) 5 days (120 h) post-siRNA transfection. (**A**) Impact of siRNA knockdown on growth phenotype of C15E cells when cultured in shake flasks post-siRNA transfection over a 5-day time course. (**B**) and (**C**) reports the viable cell number of each condition when undertaken in shake flasks or 24-deep well plates, respectively. Cells, transfection reagent only; negative, scrambled siRNA control. $n=3$, error bars show ± 1 SD

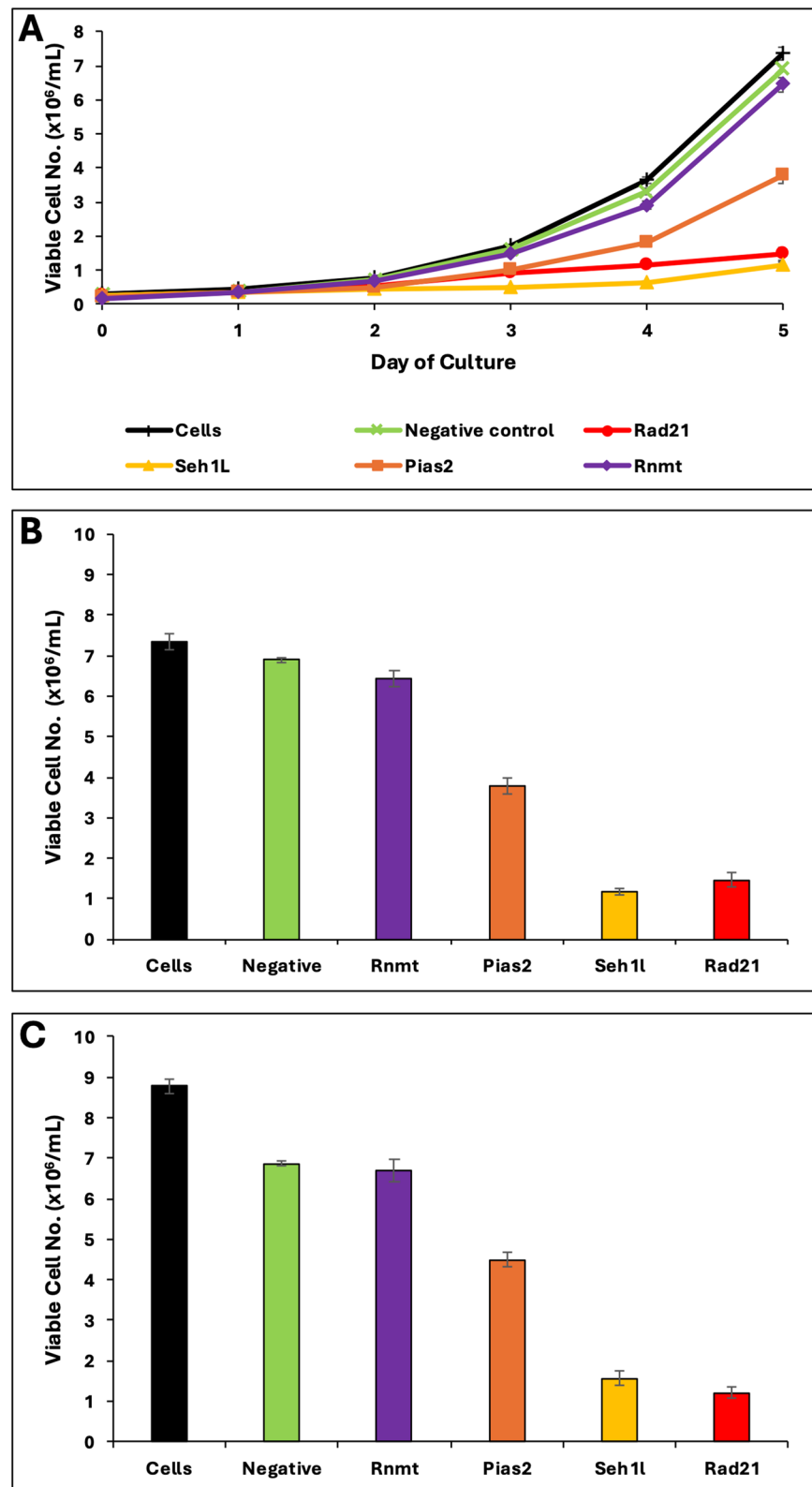
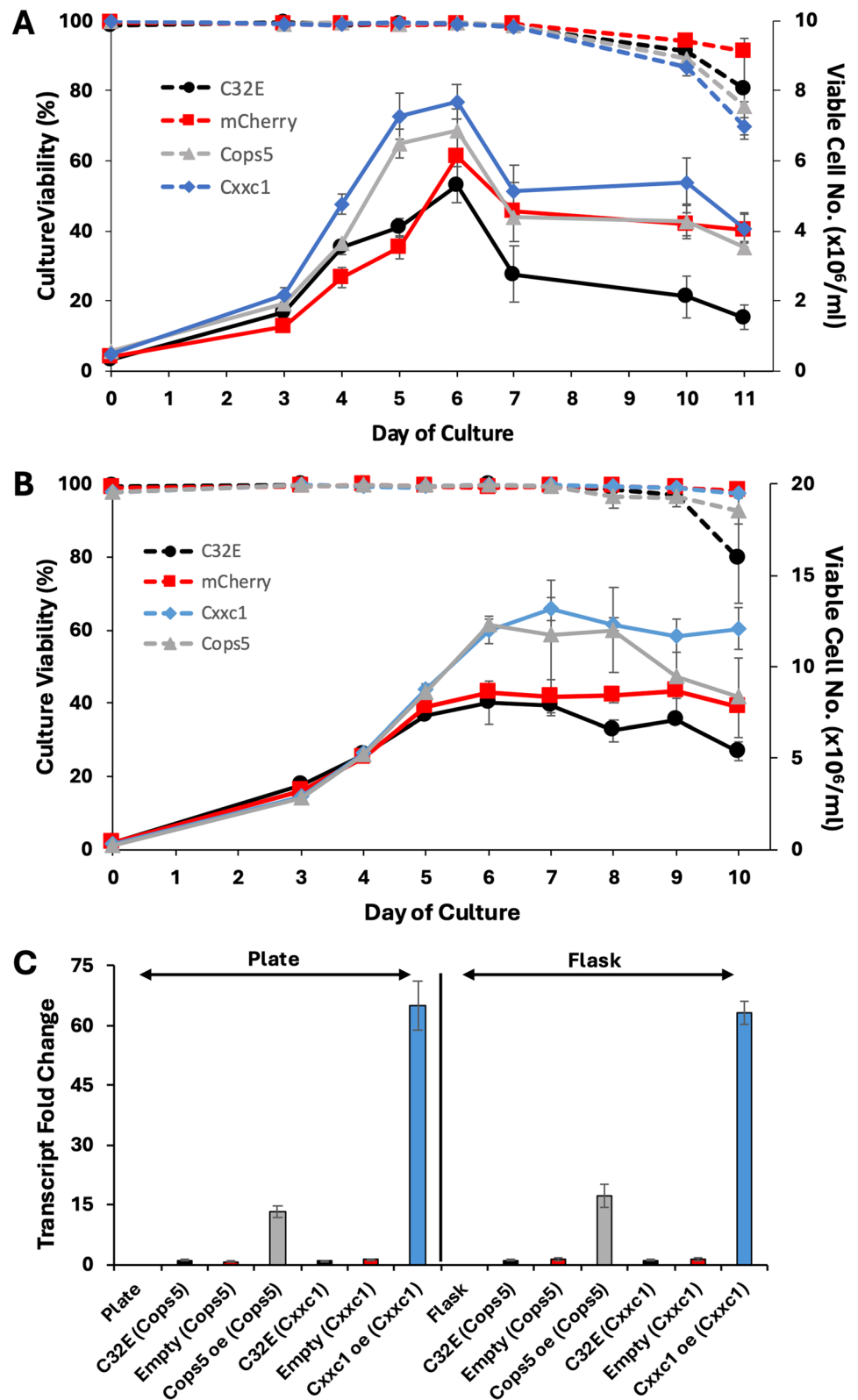


Fig. 7 Engineering of C32E low growth phenotype cells to overexpress Cops5 or Cxxc1 results in a high-growth phenotype cell pool. The original control C32E cell line, one engineered to express mCherry, and those engineered to overexpress Cxxc1 and Cops5 were cultured in 24-deep well plate format (A) or in 125 mL Erlenmeyer shake flasks (B). Cxxc1 and Cops5 engineered cells grew to a higher peak viable cell concentration than C32E untransfected cells consistent with a high growth phenotype compared to the C32E or control cells expressing mCherry in both 24-DWP (A) and shake flask (B) cultures. Statistically significant increases in viable cell number for Cxxc1 and Cops5 engineered cells compared to C32E untransfected and mCherry expressing cells were observed on day 5 ($p < 0.01$) in the DWP, and on days 5, 6 ($p < 0.01$) and day 8 ($p < 0.05$) in shake flask cultures (ANOVA with post hoc Tukey HSD, see Supplementary Table 1). Solid lines = viable cell number; Dashed lines = culture viability. (C) Transcript fold change relative to C32E control (Empty = mCherry) on day 11 of culture from 24-DWPs (A) and day 3 of culture from shake flasks (B). Each data point represents mean $\pm$ 1 SD ($n = 3$)



both Cops5 and Cxxc1 overexpressing C32E cell pools consistently grew faster and to a higher maximum viable cell concentration than the mCherry control in the 24-DWP format (Fig. 7A).

We then set out to confirm that the impact on cell growth phenotype observed in the 24-DWP screen of selected Chr2 gene overexpressing C32E cell pools was replicated in fed-batch shake flask experiments. There was no obvious difference in the growth profile between Seh11 and Lman1 overexpressing pools and the C32E and mCherry expressing controls (Supplementary Fig. 11). On-the-other-hand, there was a small but discernible increase in the maximum viable cell concentration achieved in Rsp20, Atp5f1a, Rbm22 and Rpl17 over-expressing cells (Supplementary Fig. 11). Most prominent however, was an increase in growth phenotype observed in Cxxc1 and Cops5 over-expressing C32E cell pools where there was a clear and obvious increase in growth phenotype resulting in a higher maximum viable cell concentration (Fig. 7B). This increase was similar to the difference observed between the low growth and high growth phenotype cells (e.g. C15E v C32E). Our conclusion from these data was that overexpression of Cxxc1 and Cops5 results in a high growth phenotype in C32E cells that correlates with that observed in high growth phenotype C15E Chr2 duplication cell line. Days at which statistically significant differences in cell number were observed in Fig. 7A and B are reported in Supplementary Table 5.

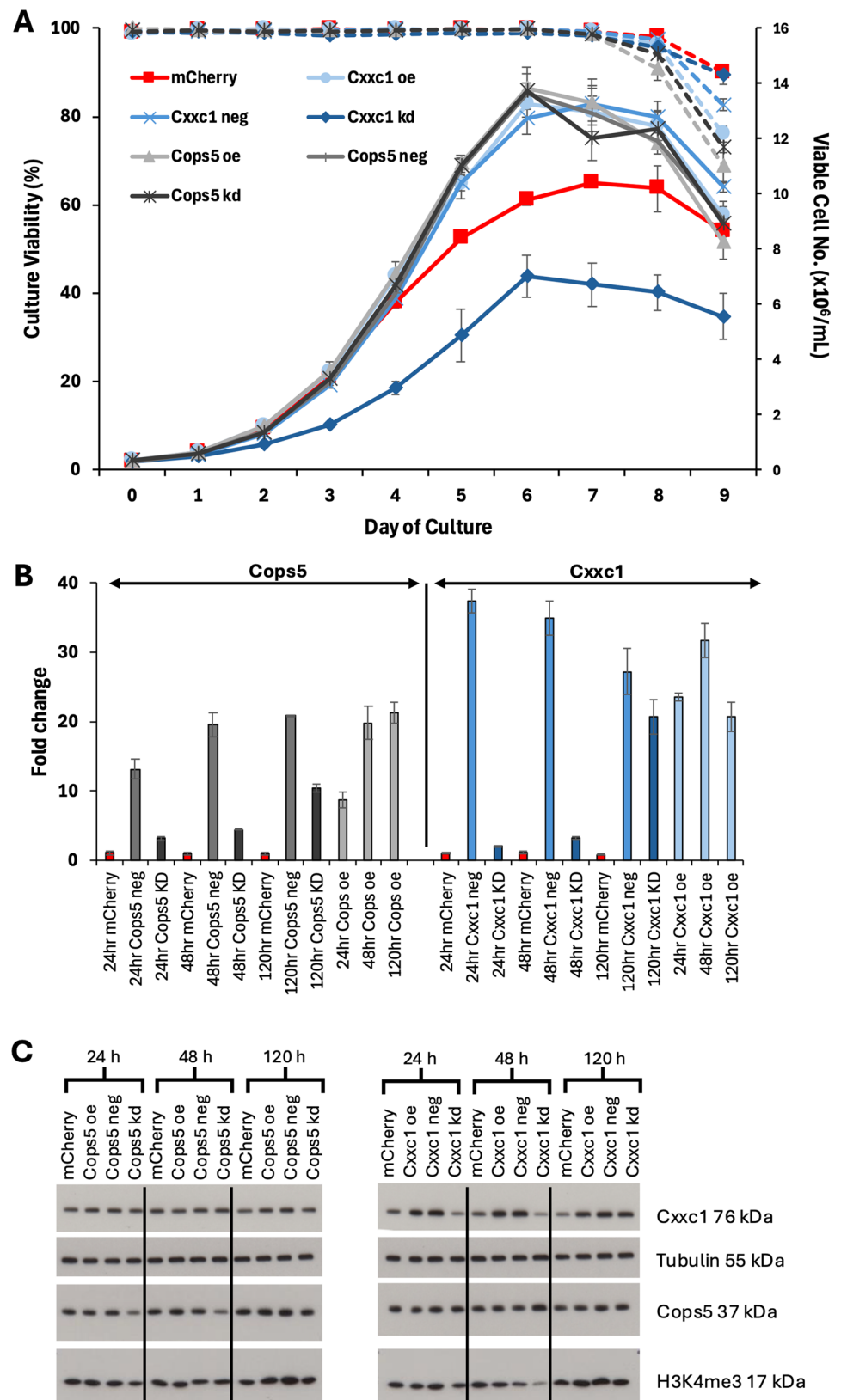
Over-expression or knockdown of Cxxc1 expression 'switches' CHO C32E cells from high growth to low growth phenotype

Having established that the overexpression of Cxxc1 and Cops5 in the C32E cell line resulted in a high growth phenotype we set out to further confirm that the manipulation of the expression of these genes can switch cells from a low to high growth phenotype (or vice versa). Initially we went back to the high growth phenotype C15E cell line (see Fig. 1A) and used siRNA to knockdown Cxxc1 and Cops5 expression. We then observed the resulting growth phenotype in fed-batch shake flask cultures, checking that knockdown persisted to at least day 5 of culture by qRT-PCR (Supplementary Fig. 12). Knockdown of Cxxc1 had a clear impact on C15E cell growth with the high growth phenotype abolished and growth following a low growth phenotype profile (Supplementary Fig. 12A). However, knockdown of Cops5, although resulting in a small change in growth profile, did not have such an obvious impact as in the case of Cxxc1 knockdown (Supplementary Fig. 12A). This again provided evidence that growth phenotype was affected by the amount of expression of these Chr2 genes and that knockdown of Cxxc1 had a greater impact on growth compared to Cops5.

We then took the C32E overexpressing Cxxc1 and Cops5 cell pools that exhibited a high growth phenotype and investigated if this could be reversed by siRNA knockdown of the over-expression in fed-batch shake flask cultures. Samples were taken throughout a 9-day culture to determine the impact on viable cell number, culture viability, relative amounts of the target transcripts across the first 5 days of culture, and for western blotting to determine the impact at the protein level on Cxxc1, Cops5 and H3K4me3 expression (Fig. 8). Again, the Cxxc1 and Cops5 over-expression cell pools showed an increased, high growth phenotype compared to the mCherry overexpressing control cell pool (Fig. 8A). When the overexpressing Cxxc1 and Cops5 C32E cell pools were transfected with a scrambled siRNA negative control there was no obvious impact on the cell growth phenotype.

siRNA knockdown of *Cops5* in the Cops5 overexpression cell pool also had no impact on the growth phenotype, in agreement with the limited impact observed when knockdown was undertaken in the C15E high growth phenotype cell line (Fig. 8A). Although there was a clear knockdown in transcript levels of *Cops5* in the overexpression cell pools, the amount of transcript remained elevated compared to the mCherry control (Fig. 8B) and thus it is possible knockdown was insufficient to reverse the effect of overexpression. Further, unlike the pattern observed when undertaking knockdown in the C15E cell line, a recovery of transcript amounts was more apparent by 120 h which is likely due to the overexpression mechanism of the plasmid overcoming the effect of siRNA knockdown (Fig. 8B). For Cops5 this was also apparent at the protein level where western blot analysis confirmed reduced expression at 24 and 48 h whilst at 120 h of culture this was less obvious (Fig. 8C and Supplementary Fig. 13).

Fig. 8 Knockdown of *Cxxc1* in C32E overexpressing cell pools reverses the high-growth impact imparted by *Cxxc1* engineering but not the knockdown of *Cops5* in *Cops5* engineered cells. **(A)** 125 mL Erlenmeyer shake flask growth culture profiles of C32E mCherry (control), *Cxxc1* overexpressing and *Cops5* overexpressing engineered C32E cell pools following transfection with a negative control scrambled siRNA (neg) or target (kd) siRNAs. Solid lines show the viable cell concentration; dashed lines show culture viability. **(B)** Transcript data reporting fold change in *Cops5* or *Cxxc1* transcript amounts relative to mCherry cell pool 24 h post-siRNA transfection. $n=3$, error bars show ± 1 SD. **(C)** Western blot analysis confirming knockdown of *Cxxc1* and *Cops5* at the protein level and profiling of H3K4me3 expression. Tubulin used as a loading control



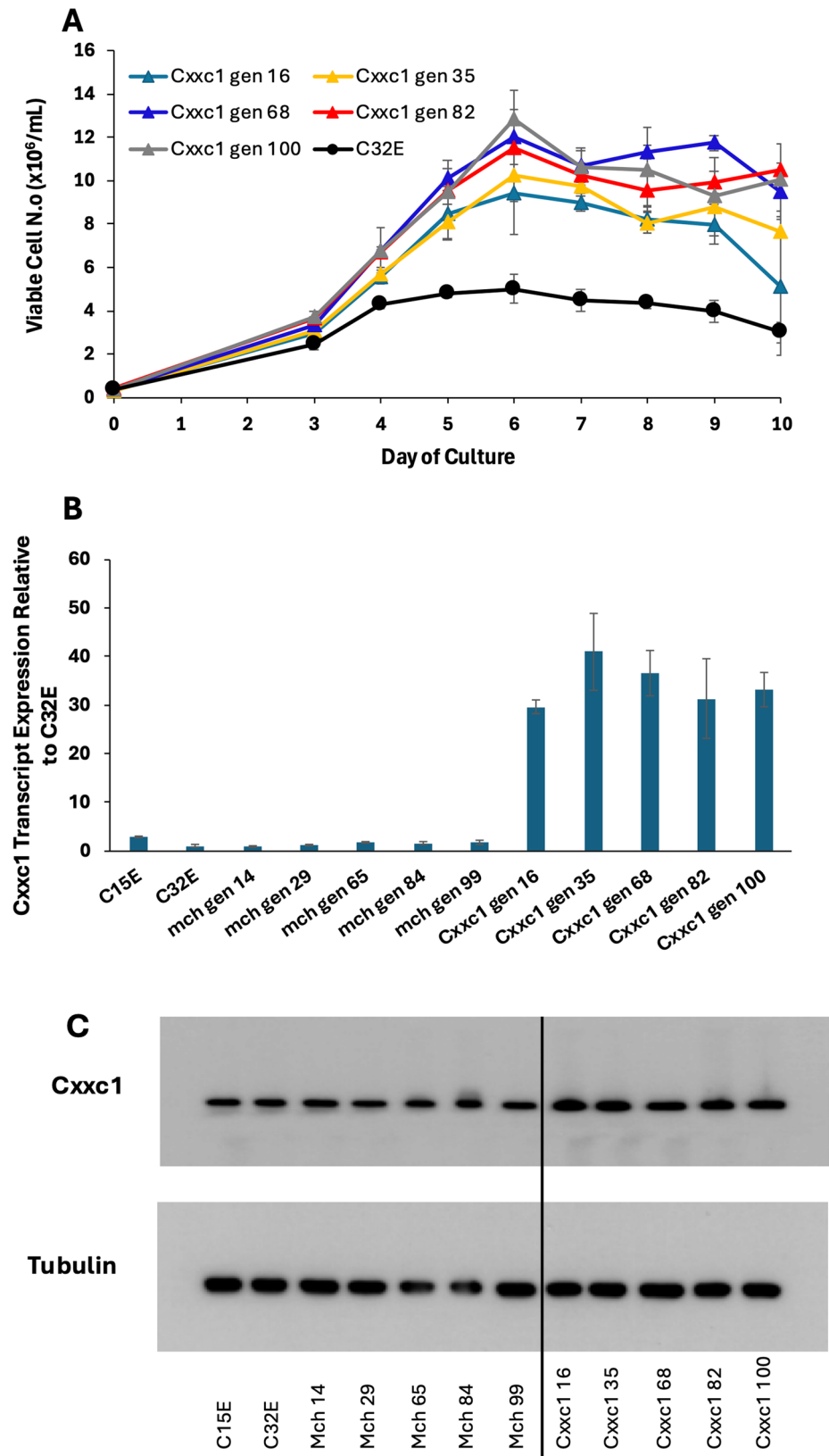
When siRNA knockdown of *Cxxc1* in the C32E *Cxxc1* overexpressing cell pools was undertaken there was an obvious impact on growth phenotype, with the growth profile reduced even compared to the mCherry control (Fig. 8A). Days at which statistically significant differences in cell number were observed in Fig. 8A are reported in Supplementary Table 5. Once again there was efficient knockdown at 24 and 48 h of culture, but this was lost by 120 h of culture (Fig. 8B). Despite this, the cell growth profile did not recover suggesting that *Cxxc1* expression is important to defining the overall growth profile as either low or high growth. Two possible interpretations of this result are that continuously elevated *Cxxc1* expression is required to achieve the high growth phenotype, or alternatively, that early transient knockdown selects for, or commits, the cell population to a low growth trajectory. Similar to *Cops5*, the *Cxxc1* transcript levels were reflected at the protein level as determined by western blot whereby there was reduced expression of *Cxxc1* at 24 and 48 h in the knockdown samples compared to the overexpression cell pool but this was not apparent at 120 h (Fig. 8C and Supplementary Fig. 13). All samples were also analysed for Histone 3.1 trimethyl-histone H3 (lys4) methylation by western blot. *Cxxc1* (also known as CFP1) contains a SET1 interaction domain required for some SET domain proteins to bind to histone H3K4 methyltransferases and such epigenetic regulation has been shown to influence proliferation [31] and cancer cell proliferation [32]. Whilst there were some subtle changes observed in expression of H3K4me3 in the *Cxxc1* overexpressing and knockdown cells, there was no obvious relationship between expression and growth phenotype (Fig. 8C and Supplementary Fig. 13). Although at day 5 there was a small increase in H3K4me3 for both *Cops5* and *Cxxc1* over-expressing cells, the subtle change suggests the growth-promoting effect of *Cxxc1* maybe independent of H3K4-methylation related epigenetic regulation, a finding which warrants further studies in the future.

During recombinant protein production, the productivity, as defined by recombinant protein titre, is a result of the cell growth phenotype during culture and the cell specific productivity (qP). Therefore, as the *Cxxc1* and *Cops5* overexpressing cell pools with the high growth phenotypes were derived from the low growth phenotype C32E cell line, we investigated whether the overexpression and associated change in growth phenotype, or subsequent knockdown, influence the antibody titre achieved from the cells from fed-batch shake flask cultures. Although there was a clear change in the growth phenotype upon *Cxxc1* and *Cops5* overexpression compared to the original C32E cell line and mCherry overexpression control, the antibody titre from the *Cxxc1* and *Cops5* pools was more-or-less the same as that from the original C32E cell line and mCherry control (Supplementary Fig. 14A). Further, when overexpression was reversed by siRNA knockdown, with a subsequent change to low growth phenotype for the *Cxxc1* overexpression cell pools, there was again little difference in the final antibody titre at the end of culture (Supplementary Fig. 14B).

Targeted *Cxxc1* engineered cell pools exhibit a stable high cell growth phenotype

We next sought to determine if *Cxxc1* overexpressing CHO cell pools showed a stable high-growth phenotype. We therefore aged the *Cxxc1* over-expression C32E cell pools through 100 generations, cryopreserving cells at different generations post recovery from transfection and selection for *Cxxc1* expression. We then revived these all simultaneously and undertook fed-batch shake flask cultures of the different generation cell pools and original low growth phenotype C32E cell line. All the *Cxxc1* cell pools of different generations maintained a high growth phenotype compared to the original C32E cell line and more-or-less consistent *Cxxc1* transcript and protein expression across the different generation numbers investigated (Fig. 9). Although there was some variation in the growth profile of the different generation number pools, and the lower generation pools (16 and 35) did not reach quite the same viable cell concentration as the higher generation pools (68, 82, 100), the general profiles were similar (Fig. 9). The *Cxxc1* generations are pools (not clonal) and therefore the growth profiles observed are indicative of consistent and robust growth phenotypes such that an age-associated change is prevented over prolonged culture periods as discussed below.

Fig. 9 C32E cell pools engineered to stably overexpress Cxxc1 have a stable high-growth phenotype over 100 generations compared to the original C32E cell line (A) and show consistent transcript (B) and protein (C) level expression of Cxxc1. C32E Cxxc1 engineered cell pools were aged to 100 generations. During this process cell pools were cryopreserved at each of the generation numbers indicated. All pools and the C32E were then revived simultaneously and after recovery from cryopreservation cultured under fed-batch conditions in 125 mL Erlenmeyer shake flasks. (A) Samples were taken on the days indicated and the viable cell number determined using a Vicell instrument. $n=3$, error bars show ± 1 SD. Cxxc1 transcript (B) and protein (C) analysis on day 3 of culture at the generation number indicated



Discussion

CHO cells excel in their ability to produce high quality and titre of secreted recombinant protein products for therapeutic applications, partially due to the ‘plasticity’ of these cells that allows them to adapt to different culture processes and demands of recombinant protein production. As outlined in the introduction, the flipside of this ‘plasticity’ is the potential for genomic instability that can lead to either, or both, genotypic and phenotypic heterogeneity that can result in a recombinant cell line having to be rejected late in cell line development, or even more damagingly, rejection of manufacturing runs/material as phenotypic attributes end up falling outside set acceptable criteria [33]. Wei et al. [24] recently reported that an age-correlated high cell growth phenotype was associated with the duplication of a 50.2 Mbp segment of chromosome 2 consisting of 123 specific genes whose transcripts were also upregulated in the aged, high growth phenotype cells lines. We initially showed that this phenotype could be replicated in 24-DWP and shake flask cultures under batch and fed-batch culture conditions using different, commercially available medium and feeds than in the original study. This confirmed that the phenotype was not only product independent as previously reported but also conserved across different culture formats and growth mediums, as was the associated increase in transcript levels. The establishment of these culture modes and processes allowed us to then screen individual genes in the duplicated Chr2 segment for any potential cause-and-effect relationship.

Before further investigating potential cause-and-effect, we extended the previous study to investigate the impact of the gene duplication in the high growth phenotype cells on the associated protein levels. Whilst we were not able to observe or identify all 123 proteins that are encoded by the Chr2 duplicated genes, of those identified, many but not all, also showed an increase in protein abundance that mirrored the gene duplication and transcript increases. We then used the transcript data, proteome data and knowledge of the role of individual genes to select a subset of 43 genes from the 123 to investigate the impact of siRNA knockdown on the growth phenotype.

siRNA knockdown was initially assessed in a 24-DWP format and growth phenotype used as the assay read-out. We recognise that this approach only investigated the potential impact of individual gene duplication and transcript increase on growth phenotype, and that in reality the high growth phenotype most likely arises from the combined effect of multiple transcript changes. Nevertheless, we hypothesised that there were a number of individual transcripts that made a major contribution towards defining the high growth phenotype and that knock-down of their increased expression would reduce the growth towards a low growth phenotype. This indeed was observed, with the knockdown of some transcripts resulting in a notable decrease in growth phenotype whilst others had little to no effect. Those transcripts whose knockdown did elicit an impact may be contributors to the Chr2 duplication high growth phenotype, but this data only allows a correlative link to be determined. Further, although we did not achieve efficient knockdown for all targets, we judged that we had identified sufficient targets with a notable impact for further investigation using the criteria outlined in the methods section. Using this criteria, we selected 14 genes to overexpress in the low growth phenotype C32E cell line and determine the impact on growth phenotype. These genes encoded for proteins with a range of functions related to cell growth including control of the cell cycle, tRNA synthesis, mRNA translation and its control, energy generation, and ribosomal proteins (Supplementary Table 4).

In order to generate C32E cell pools stably expressing the Chr2 target genes of interest, we used a hygromycin-eGFP fusion protein approach similar to that previously described [34]. This allowed us to generate stable cell pools that were then assayed for the impact on cell growth in 24-DWPs and fed-batch shake flask experiments. In all cases the resulting pools showed over-expression of the targets at the transcript level. The magnitude or ‘mean’ overexpression in pools was anywhere from 3–4-fold higher than in the original C32E cell line for some targets to more than 60-fold higher (Supplementary Fig. 10 and Supplementary Table 2 for detailed magnitude changes). We acknowledge this represents a larger increase than observed in the high growth phenotype Chr2 duplication cell lines whereby transcripts tended to be around 2-fold higher, nevertheless these were considered suitable for an initial investigation as to the impact of increased transcript expression on growth phenotype. Importantly,

although the magnitude of change at the transcript level were much higher in the *Cxxc1* engineered overexpression pools than the 2-fold in the Chr2 duplication event cells, we note that at the protein level *Cxxc1* expression was only 2–3.5 fold higher in the overexpression cell pools compared to the mCherry control (Supplementary Fig. 13). Further, protein expression was reduced to very similar amounts 24 and 48 h post-knockdown in the *Cxxc1* overexpression pools compared to the mCherry control pools before recovering at 120 h in line with the transcript profile of knockdown at 24/48 h and recovery by 120 h. Nevertheless, we acknowledge it is possible that the much larger magnitude overexpression of transcription factors such as *Cxxc1* in our overexpression cell pools may elicit changes to growth phenotypes through a mechanism different to those at the more modest 2-fold changes in the Chr2 duplication event cell lines.

Overexpression of many of the Chr2 targets had marginal to little impact on the C32E cell phenotype, however two targets, *Cops5* and *Cxxc1*, resulted in a growth phenotype similar to that observed for the high growth phenotype C15E cell line containing the Chr2 duplication event. When these two targets were subsequently knocked down again by siRNA in the overexpression cell pools, the *Cops5* overexpressing cell pools maintained their high growth phenotype, however the *Cxxc1* overexpressing cell pools reverted to a low growth phenotype. Thus, *Cxxc1* expression may be a molecular switch that can tune cell growth from a high to a low growth phenotype. Although the data here supports the conclusion that *Cxxc1* overexpression and its reversal can modulate growth phenotype, whether *Cxxc1* drives the phenotype in the Chr2 duplication event, is simply a contributor amongst a number of other Chr2 duplication genes to the high growth phenotype, or does not play any major role in the phenotype of Chr2 duplication high growth phenotype cells, cannot be ascertained from these data. Thus, whilst *Cxxc1* has been implicated in activating pathways controlling cell proliferation [35], and genome wide studies in CHO cells have shown *Cxxc1* to be an essential fitness gene [16], we do not suggest that *Cxxc1* is the sole contributor to the Chr2 duplication high growth phenotype. The evidence provided here, and the findings from previous studies, suggests that the duplication of this gene may be one contributor amongst others to the high growth phenotype observed in CHO Chr2 duplication event cell lines. Further, the duplication event is a polygenic event as opposed to the single-gene approach of *Cxxc1* overexpression. Other studies using genome wide functional approaches show that changes in the expression of single genes are often insufficient to modulate proliferation and that multi-gene interventions may be needed [16]. However, an advantage of engineering transcription factors such as *Cxxc1* that regulate a number of other genes to modulate proliferation is that in such cases single gene manipulation may be sufficient to modulate multiple genes that drive a phenotype. Indeed, both Torres et al. [19] and Latorre et al. [20] have used MYC transcription factor expression to drive rapid cell growth, setting a precedent that such approaches can be utilised to shift cell proliferation. It is noted that both these studies manipulated other genes to also enhance product titres as discussed in the introduction and in more detail below.

Based on literature reports of *Cxxc1* functioning as a chromatin regulator, we looked to determine if this was part of the mechanism by which *Cxxc1* overexpression drove a high growth phenotype. *Cxxc1*, also known as *CFP1*, is known to be involved in epigenetic regulation of gene expression, acting as a transcriptional activator as a component of the SETD1 complex involved in histone H3K4 trimethylation [36]. *Cxxc1* is reported to be important in maintaining genetic stability and cell function in rapidly dividing cells and its overexpression has been reported in various cancerous cell types with rapid cell growth characteristics [32, 37]. We therefore used western blot analysis to determine if Histone 3.1 trimethyl-histone H3 (lys4) methylation was increased in *Cxxc1* overexpressing C32E cells compared to the mCherry expressing C32E control. Although there were subtle differences observed, there was no clear association with growth phenotype and therefore we were unable to confirm from these data if *Cxxc1* may be acting via such a mechanism. Interestingly, a recent study by Lu et al. reported that SETDB1 knockdown in CHO cells can enhance recombinant protein production via an epigenetic transcriptional silencing mechanism by enhancing cell specific productivity [22]. Kim et al. also reported that knockout of the ATF7IP-SETDB1 complex in a bispecific antibody producing CHO cell line enhanced productivity but suppressed cell growth [38]. In the study reported here, we found that overexpression of *Cxxc1* resulted in a high growth phenotype but no change in volumetric mAb productivity, and thus a reduction in cell specific productivity, not in disagreement with the findings of Lu et al. [22] and Kim et al. [38].

We note that our engineering studies were undertaken on CHO cell lines already expressing a recombinant monoclonal antibody product and thus the data presented here does not confirm that Cxxc1 engineered host cell lines would be competitive producers. Nevertheless, although the focus of this study was not to improve cell specific productivity, but rather to investigate high growth phenotype, the results presented here show that there is a trade-off between cell specific productivity and the Cxxc1 engineered high growth phenotype. Despite this, there was no negative impact on the volumetric titre achieved, an important observation, as it was possible to achieve the same product titre, albeit from a greater biomass with lower cell specific productivity (qP), over the same process duration with the higher growth phenotype cell pools. It is possible that with further process optimisation and extended culture one could improve the qP or titre further from these high growth phenotype cells. An alternative approach could be to combine higher growth phenotype engineering with engineering that also seeks to enhance qP/product titre. As outlined in the introduction section, Torres et al. [19] and Latorre et al. [20] have both previously described such elegant strategies, using combinations of transcription factors that drive growth and productivity in a time dependent manner to enhance growth and product yields from CHO cell systems. These studies, focussed at achieving enhanced and rapid cell biomass, prolonged culture and increased product titre, could be aligned with our approach of engineering to achieve high growth and productivity phenotypes and preventing growth phenotype instability. The additional engineering to support productivity in these growth engineered cell lines explains the differences in productivity between the current study and the previous reports [19, 20].

Regardless of the mechanism of action, a key objective of this work was to identify Chr2 duplication targets that could be investigated in cell line engineering approaches to deliver cell lines with a consistent growth phenotype. We therefore determined the growth phenotype stability of the C32E Cxxc1 overexpressing engineered cell pools over 100 generations. We acknowledge that whilst the use of cell pools could prevent clonal variation being observed, it has been argued that using pools as opposed to clonal cell lines helps maintain the natural diversity of host cells and reduces clone-selection bias [16]. The previous report by Wei et al. [24] considered early generation, low growth phenotypes as those < 30 generations and establishment of late generation, high growth phenotypes as > 90. At all generations investigated, the C32E Cxxc1 overexpressing cell pools showed a consistent high growth phenotype and did not either revert to the original low growth phenotype or show a further increase in growth. Considering these experiments were undertaken in cell pools, this shows a remarkable stability for the previously unstable C32E cell line over 100 generations and evidence that the targeted engineering of Cxxc1 in CHO cell hosts may offer an opportunity to generate new host with a 'locked in' high growth phenotype. If this approach prevents growth instability during cell line development due to genetic instability, this would be advantageous for the generation of manufacturing production cell lines with inherent high growth phenotypes that do not exhibit age-related growth phenotypes. However, ensuring growth stability may compromise the cell specific productivity that a subsequent host cell line may deliver and thus a balance may need to be found between Cxxc1 expression, age-related growth instability, and cell specific productivity (as discussed above) in an engineering approach.

Conclusion

In conclusion, we have shown that an age-related high growth phenotype associated with a Chr2 duplication event in CHO cells can be reversed by knockdown of specific transcripts that are upregulated in their expression as a result of the duplication event. Further, one of these genes, Cxxc1, can be overexpressed in low growth phenotype cells to generate cells whose phenotype correlates to that observed in the Chr2 high growth phenotype cells, and this can be reversed by subsequent knockdown of Cxxc1 transcript levels. Although we do not propose that Cxxc1 is the sole contributor to the Chr2 age-related high growth phenotype, the data we present shows that Cxxc1 overexpression can produce a stable high growth phenotype that correlates with that of Chr2 duplication event cell lines. Further, we propose that (i) tuning Cxxc1 expression could be utilised to tune to either low or high growth phenotypes, and (ii) targeted cell engineering to overexpress Cxxc1 is a potential strategy to

overcome the issue of CHO cell age-related genetic instability associated growth phenotype to generate new host cell lines with a stable, high growth phenotype.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1186/s13036-026-00765-3>.

Acknowledgements The authors wish to thank Dr Erik S Barton (Pfizer) for useful discussions and suggestions.

Author contributions J.R.: Conceptualization, investigation, data acquisition, data interpretation, methodology, validation, writing – original draft, writing – review and editing. M.S.: Conceptualization, investigation, data acquisition, data interpretation, methodology, validation, writing – original draft, writing – review and editing. L.Zi., Conceptualization, analysis, data interpretation, methodology, project administration, writing – review and editing. W.W., Conceptualization, analysis, data interpretation, methodology, project administration, writing – review and editing. M.J.A., Analysis, data interpretation, methodology, writing – review and editing. L.Zh., Conceptualization, funding acquisition, analysis, data interpretation, methodology, project administration, writing – review and editing. C.M.S., Conceptualization, funding acquisition, analysis, data interpretation, methodology, validation, project administration, supervision, writing – original draft, writing – review and editing. All authors have read and approved the final submitted manuscript.

Funding We thank Pfizer for part funding the work described here.

Data availability All data generated or analysed during this study are included in this published article [and its supplementary information files] and/or are available from the corresponding author on reasonable request.

Declarations

Procedures All procedures performed on animals were in accordance with regulations and established guidelines and were reviewed and approved by an Institutional Animal Care and Use Committee or through an ethical review process.

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Walsh G, Walsh E. Biopharmaceutical benchmarks 2022. *Nat Biotechnol.* 2022;40(12):1722–60. <https://doi.org/10.1038/s41587-022-01582-x>.
2. Budge JD, Knight TJ, Povey J, Roobol J, Brown IR, Singh G, et al. Engineering of Chinese hamster ovary cell lipid metabolism results in an expanded ER and enhanced recombinant biotherapeutic protein production. *Metab Eng.* 2020;57:203–16. <https://doi.org/10.1016/j.ymben.2019.11.007>.
3. Roobol A, Roobol J, Smith ME, Carden MJ, Hershey JWB, Willis AE, et al. Engineered transient and stable overexpression of translation factors eIF3i and eIF3c in CHOK1 and HEK293 cells gives enhanced cell growth associated with increased c-Myc expression and increased recombinant protein synthesis. *Metab Eng.* 2020;59:98–105. <https://doi.org/10.1016/j.ymben.2020.02.001>.
4. Chang M, Kumar A, Kumar S, Huhn S, Timp W, Betenbaugh M, et al. Epigenetic comparison of CHO hosts and clones reveals divergent methylation and transcription patterns across lineages. *Biotech Bioeng.* 2022;119(4):1062–76. <https://doi.org/10.1002/bit.28036>.

5. Choi DH, Kim SJ, Song J, Park SY, Park CH, Lee J, et al. Exploring CHO cell stability during prolonged passaging via eXplainable AI driven flux balance analysis. *npj Syst Biol Appl.* 2026;12(1):36. <https://doi.org/10.1038/s41540-026-00660-z>.
6. Dahodwala H, Lee KH. The fickle CHO: a review of the causes, implications, and potential alleviation of the CHO cell line instability problem. *Curr Opin Biotechnol.* 2019;60:128–37. <https://doi.org/10.1016/j.copbio.2019.01.011>.
7. Huhn S, Chang M, Kumar A, Liu R, Jiang B, Betenbaugh M, et al. Chromosomal instability drives convergent and divergent evolution toward advantageous inherited traits in mammalian CHO bioproduction lineages. *iScience.* 2022;25(4):104074. <https://doi.org/10.1016/j.isci.2022.104074>.
8. Marx N, Eisenhut P, Weinguny M, Klanert G, Borth N. How to train your cell - Towards controlling phenotypes by harnessing the epigenome of Chinese hamster ovary production cell lines. *Biotechnol Adv.* 2022;56:107924. <https://doi.org/10.1016/j.biotechadv.2022.107924>.
9. Syddall KL, Fernandez-Martell A, Cartwright JF, Alexandru-Crivac CN, Hodgson A, Racher AJ, et al. Directed evolution of biomass intensive CHO cells by adaptation to sub-physiological temperature. *Metab Eng.* 2024;81:53–69. <https://doi.org/10.1016/j.ymben.2023.11.005>.
10. Torres M, Betts Z, Scholey R, Elvin M, Place S, Hayes A, et al. Long term culture promotes changes to growth, gene expression, and metabolism in CHO cells that are independent of production stability. *Biotech Bioeng.* 2023;120(9):2389–402. <https://doi.org/10.1002/bit.28399>.
11. Baik JY, Lee KH. Growth Rate Changes in CHO Host Cells Are Associated with Karyotypic Heterogeneity. *Biotechnol J.* 2018;13(3):1700230. <https://doi.org/10.1002/biot.201700230>.
12. Tzani I, Herrmann N, Carillo S, Spargo CA, Hagan R, Barron N, et al. Tracing production instability in a clonally derived CHO cell line using single-cell transcriptomics. *Biotech Bioeng.* 2021;118(5):2016–30. <https://doi.org/10.1002/bit.27715>.
13. Yamano-Adachi N, Hata H, Nakanishi Y, Omasa T. Effects of genome instability of parental CHO cell clones on chromosome number distribution and recombinant protein production in parent-derived subclones. *J Biosci Bioeng.* 2024;137(1):54–63. <https://doi.org/10.1016/j.jbiosc.2023.10.001>.
14. Zeh N, Schmidt M, Schulz P, Fischer S. The new frontier in CHO cell line development: From random to targeted transgene integration technologies. *Biotechnol Adv.* 2024;75:108402. <https://doi.org/10.1016/j.biotechadv.2024.108402>.
15. Xiong K, La Cour Karotki KJ, Hefzi H, Li S, Grav LM, Li S, et al. An optimized genome-wide, virus-free CRISPR screen for mammalian cells. *Cell Rep Methods.* 2021;1(4):100062. <https://doi.org/10.1016/j.crmeth.2021.100062>.
16. Marzluf J, Rattay M, Goertz M, Zehe C, Leroux AC. Genetic drivers of Chinese hamster ovary cell proliferation revealed by functional genomics. *iScience.* 2025;28(11):113847. <https://doi.org/10.1016/j.isci.2025.113847>.
17. Shin SW, Kim SH, Gasselin A, Lee GM, Lee JS. Comprehensive genome-scale CRISPR knockout screening of CHO cells. *Sci Data.* 2025;12(1):71. <https://doi.org/10.1038/s41597-025-04438-6>.
18. Masuda K, Kubota M, Nakazawa Y, Iwama C, Watanabe K, Ishikawa N, et al. Establishment of a novel cell line, CHO-MK, derived from Chinese hamster ovary tissues for biologics manufacturing. *J Biosci Bioeng.* 2024;137(6):471–9. <https://doi.org/10.1016/j.jbiosc.2024.02.005>.
19. Torres M, McConnaughie D, Akhtar S, Gaffney CE, Fievet B, Ingham C, et al. Engineering mammalian cell growth dynamics for biomanufacturing. *Metab Eng.* 2024;82:89–99. <https://doi.org/10.1016/j.ymben.2024.01.006>.
20. Latorre Y, Torres M, Vergara M, Berrios J, Sampayo MM, Gödecke N, et al. Engineering of Chinese hamster ovary cells for co-overexpressing MYC and XBPIs increased cell proliferation and recombinant EPO production. *Sci Rep.* 2023;13(1):1482. <https://doi.org/10.1038/s41598-023-28622-z>.
21. Hilliard W, Lee KH. Systematic identification of safe harbor regions in the CHO genome through a comprehensive epigenome analysis. *Biotech Bioeng.* 2021;118(2):659–75. <https://doi.org/10.1002/bit.27599>.
22. Lu JT, Zhang M, Gao YP, Yang LL, Cui ZM, Han MM, et al. SETDB1 knockdown boosts recombinant protein in CHO cells via epigenetic transcriptional silencing. *AMB Expr.* 2025;15(1):110. <https://doi.org/10.1186/s13568-025-01914-5>.
23. Spahn PN, Zhang X, Hu Q, Lu H, Hamaker NK, Hefzi H, et al. Restoration of DNA repair mitigates genome instability and increases productivity of Chinese hamster ovary cells. *Biotech Bioeng.* 2022;119(3):963–82. <https://doi.org/10.1002/bit.28016>.
24. Wei W, Zielewicz L, Zhang L. Duplication of a chromosome 2 segment in production CHO cell lines correlates with age-related growth improvement. *Biotechnol J.* 2024;19(2):2300407. <https://doi.org/10.1002/biot.202300407>.
25. Weinguny M, Klanert G, Eisenhut P, Lee I, Timp W, Borth N. Subcloning induces changes in the DNA-methylation pattern of outgrowing Chinese hamster ovary cell colonies. *Biotechnol J.* 2021;16(6):2000350. <https://doi.org/10.1002/biot.202000350>.
26. Wiśniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. *Nat Methods.* 2009;6(5):359–62. <https://doi.org/10.1038/nmeth.1322>.
27. Bryan L, Henry M, Kelly RM, Frye CC, Osborne MD, Clynes M, et al. Mapping the molecular basis for growth related phenotypes in industrial producer CHO cell lines using differential proteomic analysis. *BMC Biotechnol.* 2021;21(1):43. <https://doi.org/10.1186/s12896-021-00704-8>.

28. Chengan K, Hind C, Stanley M, Wand ME, Nagappa LK, Howland K, et al. A cell-free strategy for host-specific profiling of intracellular antibiotic sensitivity and resistance. *npj Antimicrob Resist*. 2023;1(1):16. <https://doi.org/10.1038/s44259-023-00018-z>.
29. Klanert G, Fernandez DJ, Weinguny M, Eisenhut P, Bühler E, Melcher M, et al. A cross-species whole genome siRNA screen in suspension-cultured Chinese hamster ovary cells identifies novel engineering targets. *Sci Rep*. 2019;9(1):8689. <https://doi.org/10.1038/s41598-019-45159-2>.
30. Jeong JY, Yim HS, Ryu JY, Lee HS, Lee JH, Seen DS, et al. One-Step Sequence- and Ligation-Independent Cloning as a Rapid and Versatile Cloning Method for Functional Genomics Studies. *Appl Environ Microbiol*. 2012;78(15):5440–3. <https://doi.org/10.1128/AEM.00844-12>.
31. Meng X, Zhu Y, Liu K, Wang Y, Liu X, Liu C, et al. CXXC-finger protein 1 associates with FOXP3 to stabilize homeostasis and suppressive functions of regulatory T cells. *eLife*. 2025;13:RP103417. <https://doi.org/10.7554/eLife.103417>.
32. Yang LQ, Hu HY, Han Y, Tang ZY, Gao J, Zhou QY, et al. CpG-binding protein CFP1 promotes ovarian cancer cell proliferation by regulating BST2 transcription. *Cancer Gene Ther*. 2022;29(12):1895–907. <https://doi.org/10.1038/s41417-022-00503-z>.
33. Dorai H, Corisdeo S, Ellis D, Kinney C, Chomo M, Hawley-Nelson P, et al. Early prediction of instability of chinese hamster ovary cell lines expressing recombinant antibodies and antibody-fusion proteins. *Biotech Bioeng*. 2012;109(4):1016–30. <https://doi.org/10.1002/bit.24367>.
34. Konishi Y, Karnan S, Takahashi M, Ota A, Damdindorj L, Hosokawa Y, et al. A System for the Measurement of Gene Targeting Efficiency in Human Cell Lines Using an Antibiotic Resistance—GFP Fusion Gene. *Biotechniques*. 2012;53(3):141–52. <https://doi.org/10.2144/0000113911>.
35. Sabbattini P, Trasanidis N, Taslim J, Whitwell H, Ng H, Chaidos A, et al. CXXC1 is required for IRF4 chromatin binding and expression and regulates cellular fitness in lenalidomide-sensitive and -resistant myeloma cells. *Blood*. 2025;146(Supplement 1):571–571. <https://doi.org/10.1182/blood-2025-571>.
36. Mahadevan J, Skalnik DG. Efficient differentiation of murine embryonic stem cells requires the binding of CXXC finger protein 1 to DNA or methylated histone H3-Lys4. *Gene*. 2016;594(1):1–9. <https://doi.org/10.1016/j.gene.2016.08.048>.
37. Yang Y, Yang Y, Chan K, Couture J. Analyzing the impact of CFP1 mutational landscape on epigenetic signaling. *FASEB J*. 2021;35(8):e21790. <https://doi.org/10.1096/fj.202100427R>.
38. Kim SH, Park JH, Shin S, Shin S, Chun D, Kim YG, et al. Genome-Wide CRISPR/Cas9 Screening Unveils a Novel Target ATF7IP–SETDB1 Complex for Enhancing Difficult-to-Express Protein Production. *ACS Synth Biol*. 2024;13(2):634–47. <https://doi.org/10.1021/acssynbio.3c00646>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Joanne Roobol¹ · Maria Stanley¹ · Laura Zielewicz² · Wei Wei² · Martin J. Allen² · Lin Zhang^{2,3} · C. Mark Smales^{1,4} 

✉ C. Mark Smales
c.m.smales@kent.ac.uk; mark.smales@nibr.ie

¹ School of Natural Sciences, Giles Lane, University of Kent, Canterbury, Kent CT2 7NJ, UK

² Cell Line Development, Biotherapeutics Pharmaceutical Sciences, Pfizer Inc, Andover, MA, USA

³ Bioprocess Research and Development, Merck & Co. Inc, Kenilworth, NJ, USA

⁴ National Institute for Bioprocessing Research and Training, Foster Avenue, Mount Merrion, Blackrock, Co., Dublin A94 X099, Ireland