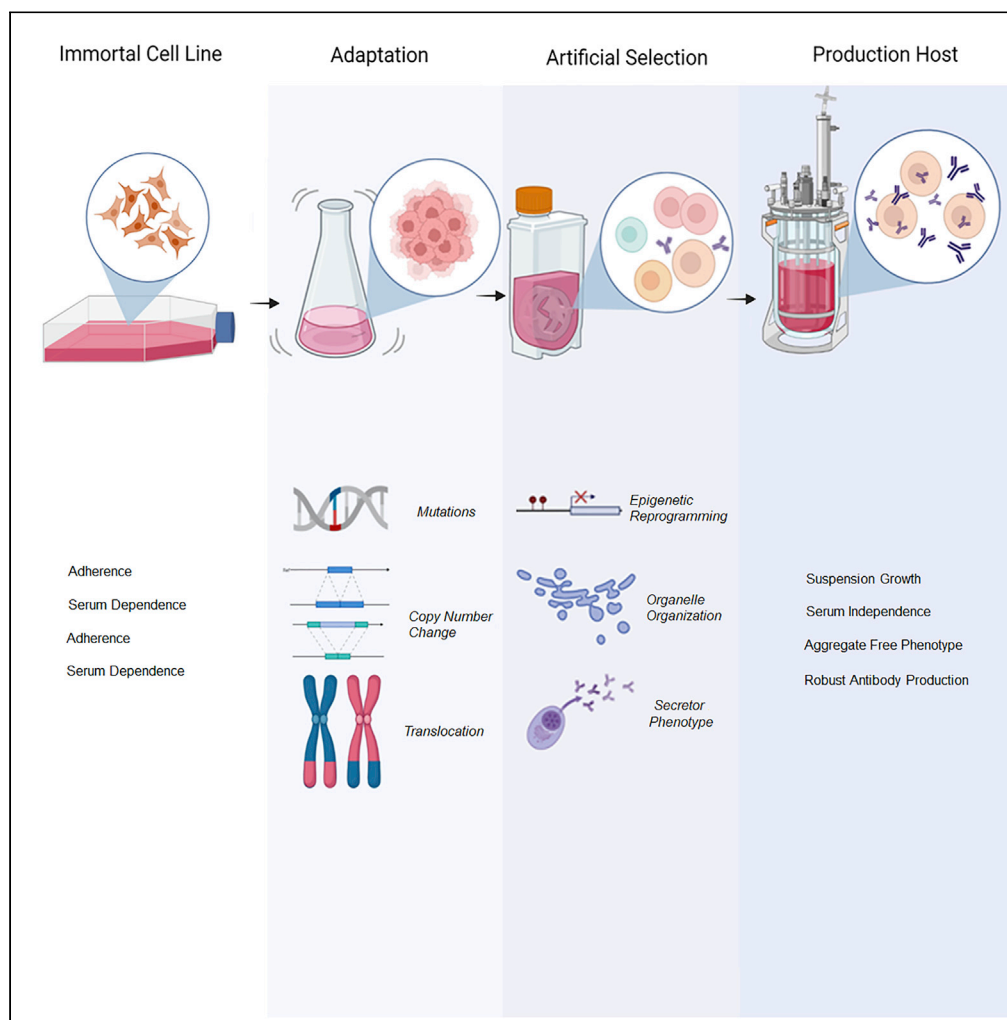


Article

Chromosomal instability drives convergent and divergent evolution toward advantageous inherited traits in mammalian CHO bioproduction lineages



Steve Huhn, Meiping Chang, Amit Kumar, ..., Henry Lin, Gregg Nyberg, Zhimei Du

zhimeidu@gmail.com

Highlights

Inheritance of genetic and epigenetic CHO programs driving antibody production

Identification of CHO genetic circuits associated with cell adaptation and evolution

Discovery of chromosome rearrangements associated with bioproduction

Cell engineering and synthetic biology strategies for improving CHO bioprocess

Huhn et al., iScience 25, 104074
April 15, 2022 © 2022 The Authors.
<https://doi.org/10.1016/j.isci.2022.104074>



Article

Chromosomal instability drives convergent and divergent evolution toward advantageous inherited traits in mammalian CHO bioproduction lineages

Steve Huhn,^{1,3,4} Meiping Chang,^{1,3} Amit Kumar,^{1,3} Ren Liu,¹ Bo Jiang,¹ Michael Betenbaugh,² Henry Lin,¹ Gregg Nyberg,¹ and Zhimei Du^{1,*}

SUMMARY

Genetic instability of Chinese hamster ovary (CHO) cells is implicated in production inconsistency through poorly defined mechanisms. Using a multi-omics approach, we analyzed the variations of CHO lineages derived from CHO-K1 cells. We identify an equilibrium between random genetic variation of the CHO genome and heritable traits driven by culture conditions, selection criteria, and genetic linkage. These inherited changes are associated with the selection pressures related to serum removal, suspension culture transition, protein expression, and secretion. We observed that a haploid reduction of a Chromosome 2 region after serum-free, suspension adaptation, was consistently inherited, suggesting common adaptation mechanisms. Genetic variations also included ~200 insertions/deletions, ~1000 single-nucleotide polymorphisms, and ~300–2000 copy number variations, which were exacerbated after gene editing. In addition, heterochromatic chromosomes were preferentially lost as cells continuously evolved. Together, these observations demonstrate a highly plastic signature for adapted CHO cells and paves the way towards future host cell engineering.

INTRODUCTION

Biopharmaceuticals, from monoclonal antibodies to large peptides, represent a market share of hundreds of billions annually worldwide (Butler and Meneses-Acosta, 2012; Greber and Fussenegger, 2007; Omasa et al., 2010; Hacker et al., 2009). Unlike small-molecule drugs, biologics are more complex in size and structure and are produced from living cells. From DNA to protein product, additional characteristics from various posttranslational modifications further define biologic activity. These molecule-specific characteristics are known as critical quality attributes (CQAs). CQAs that result from posttranslational modifications, such as glycosylation, are generated by intracellular machinery within production cell lines. Thus, despite identical product sequences, CQAs can still vary considerably lot-to-lot, suggesting that cell-based production machinery is highly variant. Further investigation of this variability at both the cellular and molecular levels is critical in developing rational product quality control strategies.

Chinese hamster ovary (CHO) cells continue to be the primary industrial protein production workhorse due to their attractive ability to adapt to various culture conditions (Kim et al., 2012; Karthik et al., 2007). First isolated by Theodore Puck in 1956, CHO cells were originally derived from *Cricetulus griseus* tissue grown *in vitro* (Karthik et al., 2007; Wurm and Hacker, 2011). A singular clone derived from immortalized progenitors, deemed CHO-K1, is the ancestor for the majority of CHO hosts used in manufacturing (Karthik et al., 2007; Wurm and Hacker, 2011; Xu et al., 2011). Nevertheless, modern CHO cell lines significantly diverge from CHO-K1 and markedly vary between institutions (Wurm and Hacker, 2011; Lewis et al., 2013). Thus, the term CHO represents not one, but many subcultures with diverse growth, expression yield, and protein product quality features (Wurm and Hacker, 2011; Lewis et al., 2013; Lakshmanan et al., 2019; Wurm, 2017). On the other hand, this plasticity represents a significant disadvantage during biotherapeutic manufacturing and results in lot-to-lot variability in cell culture performance. The plasticity of CHO cells concurrently imbues genomic instability (Dhiman

¹Process Cell Sciences, Biologics Process R&D, Merck & Co., Inc., Kenilworth, NJ 07033, USA

²Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD 21218, USA

³These authors contributed equally

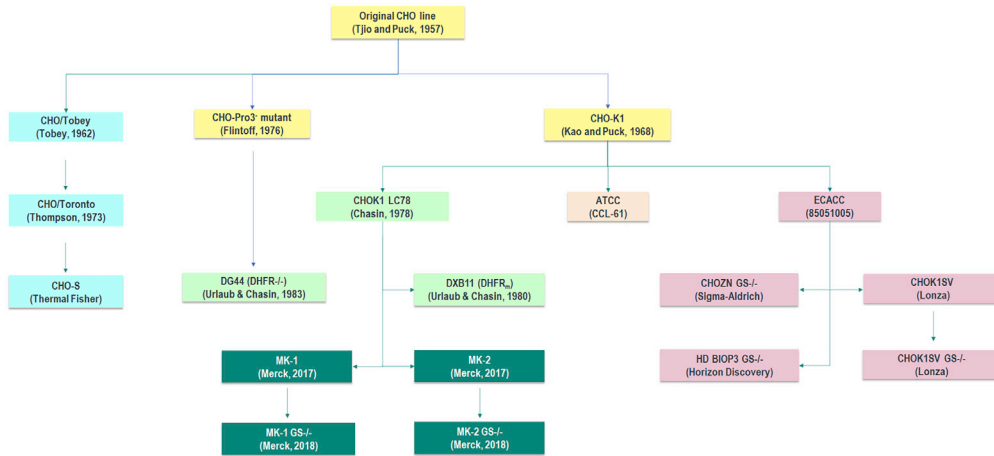
⁴Lead contact

*Correspondence: zhimeidu@gmail.com

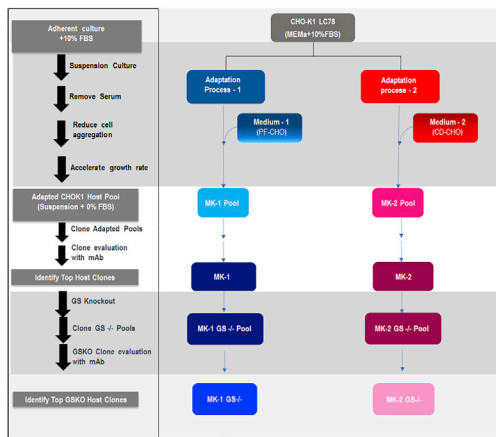
<https://doi.org/10.1016/j.isci.2022.104074>



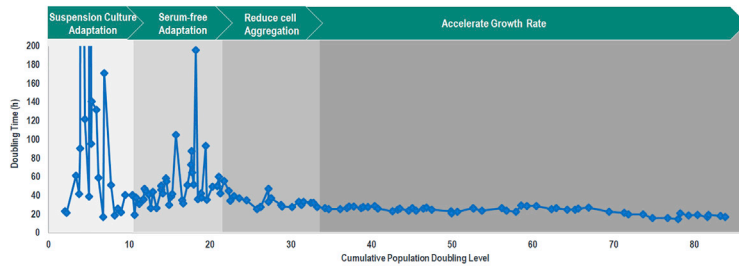
A



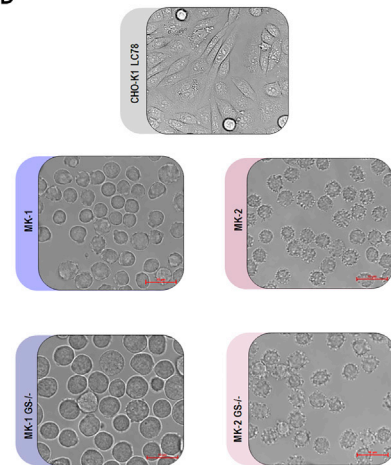
B



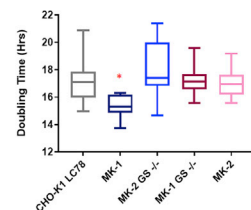
C



D



E



F

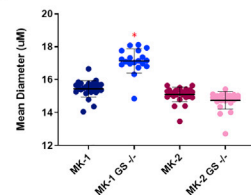


Figure 1. History and adaptation of CHO-K1 cultures into serum-independent suspension cells

(A) CHO cells represent many subspecies from independent laboratories. A family tree depicting the source material for independent CHO lineages is depicted, with the MK-1 and MK-2 hosts and derivative GS knockout hosts shown in dark green.

(B) CHO-K1 LC78 Cells were adapted into chemically defined media by independent methods to generate two unique host cells. MK-2 cells were generated by titrating amounts of serum over time in chemically defined media (CD-CHO/MEM-alpha mix. Alternatively, MK-1 cells were produced by slowly titrating soy-hydrolysate proficient (PF-CHO). Once stable pools were established, and the doubling time normalized, single cell clones were generated using FACs. These clones were then scaled, banks were prepared, and clones were transfected with recombinant antibody. Following a fed-batch production assay, the top producers, deemed "MK-1" and "MK-2," were identified, and these host cells were thawed. These hosts were then transfected with ZFN mRNA and used to generate GS^{-/-} pools. These GS^{-/-} pools were then cloned and ranked for protein expression as above to yield MK-1 GS^{-/-} or MK-2 GS^{-/-} host lines. See also [Figure S1](#).

(C) Representative cell doubling times during the adaptation process of MK-1 are graphed in the bottom panel.

(D) Cells were visualized at 100 \times magnification using an inverted light microscope.

These cells were passaged in shake flasks and the doubling time (E) and mean diameter (F) of each lineage were recorded over 10–15 passages (data is represented as boxplots for >30 cells where the centroid represents the median and whiskers represent the min and max, or dot plots where each dot represents one cell.) Scale bars represent 10 microns in length. Red asterisks represent significance ($p < 0.05$) versus all other data points using Student's *t* test with unequal variance.

[et al., 2019](#); [Fan et al., 2012](#); [Kildegaard et al., 2013](#); [Kim et al., 2011](#)). This well-recognized issue in the biotech industry has obviated an emphasis on both genetic homogeneity through isogenic cloning and process operational control strategy. Single-cell cloning is thought to induce phenotypic homogeneity of the production culture and mitigate variability during manufacturing ([Wurm, 2017](#)). Despite this, significant phenotypic and genetic variation within clonal populations is still observed in continuous long-term cultures ([Scarcelli et al., 2018](#); [Vcelar et al., 2018](#); [O'Brien et al., 2020](#)). Furthermore, phenotypic variation cannot be reduced when recombinant DNA genomic integration events are homogenized through site-specific integration, suggesting that intrinsic genetic instability CHO is a causal factor of manufacturing inconsistency ([Hamaker and Lee, 2018](#)). This ultimately raises the question if developing a process control strategy for CHO is rational, as genetic variability is both unavoidable and ubiquitous. It is vital to further dissect the heritable and variable features of CHO cells at genetic and phenotypic levels to address this question.

In this study, we systematically investigate how genetic profiles and corresponding phenotypes are affected by both single-cell cloning and cell culture processes. Here, we employed two different strategies to derive serum-free, suspension, CHO cells from adherent CHO-K1 cultures ([Figures 1B and S1](#)). These adapted pools were then cloned, the resulting clones were transfected with the gene for Trastuzumab, and clones were then screened for recombinant antibody production. The recombinant clones were then ranked, and the top-producing derivative host cell lines were subsequently modified by Zinc Finger Nucleases to eliminate expression of glutamine synthetase (GS). Knock-out pools were then re-cloned, yielding GS-null derivative cell lines as shown in [Figure 1B](#). Through this framework, we identified a universally conserved haploid reduction of a region of Chromosome (Chr) 2 (36–60 MBp) that occurred after adaptation and was consistently inherited in the derived cell lines.

These findings suggest a common pathway for general adaptation to serum-free suspension culture. This Chr 2 region is significantly enriched in common pathways associated with CHO suspension culture, i.e. cytoskeletal rearrangement, and insulin receptor signaling. However, we also observed that derivative clones also spontaneously generated additional genetic modifications, such as coding polymorphisms and copy number variations, which occur *in cis* with these general adaptation-related traits. Inherited polymorphisms were associated with unique chromosome translocations and rearrangements, specific to the different adaptation strategies, suggesting functional importance. Similarly, a consistent reduction of Chrs 9, 10, and X, which are highly heterochromatic, was observed among all lineages. Analysis of all these inherited and unique genetic changes illuminated affected pathways such as the unfolded protein response (UPR), DNA repair, oxidative stress, protein translation, and cytoskeletal rearrangement. These changes may be associated with both cell growth and recombinant protein production.

Taken together, our study demonstrates that both adaptive evolution and random genetic drift act in concert to ensure enough CHO diversity exists to adapt to culture-mediated evolutionary pressures. Understanding genetic and the corresponding phenotypic patterns can improve our control strategy during cell line development and biologics manufacturing.

RESULTS

Directed evolution and development of new CHO host cell lines for recombinant protein production

Despite sharing a universal common ancestor, the CHO moniker encompasses thousands of divergent cell lines across the many laboratories of academia and the biopharma industry (Lewis et al., 2013; Wurm and Hacker, 2011) (Figure 1A). Commercial CHO lineages are selectively adapted for the stressors of large-scale bioreactors but the genetic forces that mold cells into protein production factories remain ill-defined (Feichtinger et al., 2016; Lewis et al., 2013; Xu et al., 2011). To better characterize the processes involved during host development, we leveraged recent internal host development efforts to better understand directed cell-evolution. The CHO-K1 adherent cell line, CHO-K1 LC78, derived directly from the Puck lab, was adapted to suspension culture with two different serum-free culture medias in a stepwise manner (Figure S1). Following the continuous removal of aggregated cell clumps and accelerating of the cell growth through continuous subculture for 1–3 months, cells were cloned, transfected, and screened for recombinant protein expression. Two host clones, MK-1 and MK-2, were selected based on desired phenotypic criteria, such as high growth rate, productivity, protein folding, and N-linked glycosylation efficiency (Figure 1B). These selected progenitor host clones vials were then thawed and further subjected to genetic knockout of the glutamine synthetase (GS) gene to enhance productivity by increasing selection stringency. The cells were again cloned, transfected, and screened for recombinant protein expression with the same selection phenotypic criteria as discussed earlier. The top producers hosts were identified as MK-1 GS $-/-$ and MK-2 GS $-/-$.

Throughout the adaptation process, cells experienced two major stressors that significantly impeded proliferation as shown in Figure 1C. These occurred during the transition to suspension culture and as serum was completely removed from the culture media. Serum contains more than 1000 components, including growth factors, lipids, carbohydrates, hormones, enzymes, and other undefined constituents, which are necessary to support cell growth and function (Tu et al., 2018). Replacement of serum with the two different chemically defined medias revealed a stunning diversification of cell morphologic and phenotypic traits. MK-1 and MK-2 lineages were unique in both cell morphology and growth, with MK-2 demonstrating a bud-like cell morphology (Figure 1D) and a significantly longer doubling time (Figure 1E). Following genetic ablation of the GS locus, some phenotypes drastically changed, such as growth rate and cell size (Figures 1E and 1F; MK-1 vs MK-1 GS $-/-$). On the other hand, some phenotypes, such as bud-like morphology, were maintained (Figure 1D; MK-2 vs MK-2 GS $-/-$).

Characterization of genomic plasticity in CHO lineages

Next, we evaluated the genetic characteristics of individual lineages. We noticed that the different adaptation processes resulted in considerable karyotypic variability across lineages (Figure 2A). Compared with parental CHO-K1 LC78, in which the majority of cells exhibited 20 chromosomes, MK-1 lost one chromosome, whereas MK-1 GS $-/-$ roughly doubled the genome ($n = 19$, and 38 for MK-1, and MK-1 GS $-/-$, respectively, Figure 2A). On the other hand, the numeric distribution of chromosomes was seemingly well maintained in MK-2 and MK-2 GS $-/-$ ($n = 20$, 20, and 20 for CHO-K1 LC78, MK-2, and MK-2 GS $-/-$, respectively, Figure 2A).

We next reasoned that numerical chromosome heterogeneity was insufficient to determine genomic stability of these cell lines. Chromosome number can appear to be constant while rearrangements and other structural changes occur (Auer et al., 2018; Brinkrolf et al., 2013; Vcelar et al., 2018). Other studies have noted that the configuration and copy number of rearranged chromosomes are associated with productivity in CHO (Yamano et al., 2015). Thus, we directed our attention to chromosome painting. Here, we detected chromosome recombination events, translocations, and deletions (Figure 2B). Although MK-2 and MK-2 GSKO $-/-$ lineages demonstrated amplification of Chr 2 and loss of Chr 10, MK-1 and MK-1 GSKO $-/-$ exhibited amplification of Chrs 4, 6, and 8 and a decrease of X (Figure 2B). Generally, we observed that loss of Chr 9 was ubiquitous, whereas Chr 1 was well maintained. Surprisingly, the ploidy of MK-1 GS $-/-$ lineage, which was initially suggested by spreading, demonstrated augmented amplification of specific genomic loci over others (notably Chr 4 through 6), instead of a complete doubling of progenitor genetic material (Figure 2B). In addition, more chromosome abnormalities and unusual cytogenetic structures were identified in this cell line, suggesting potential genetic stress, as will be discussed later.

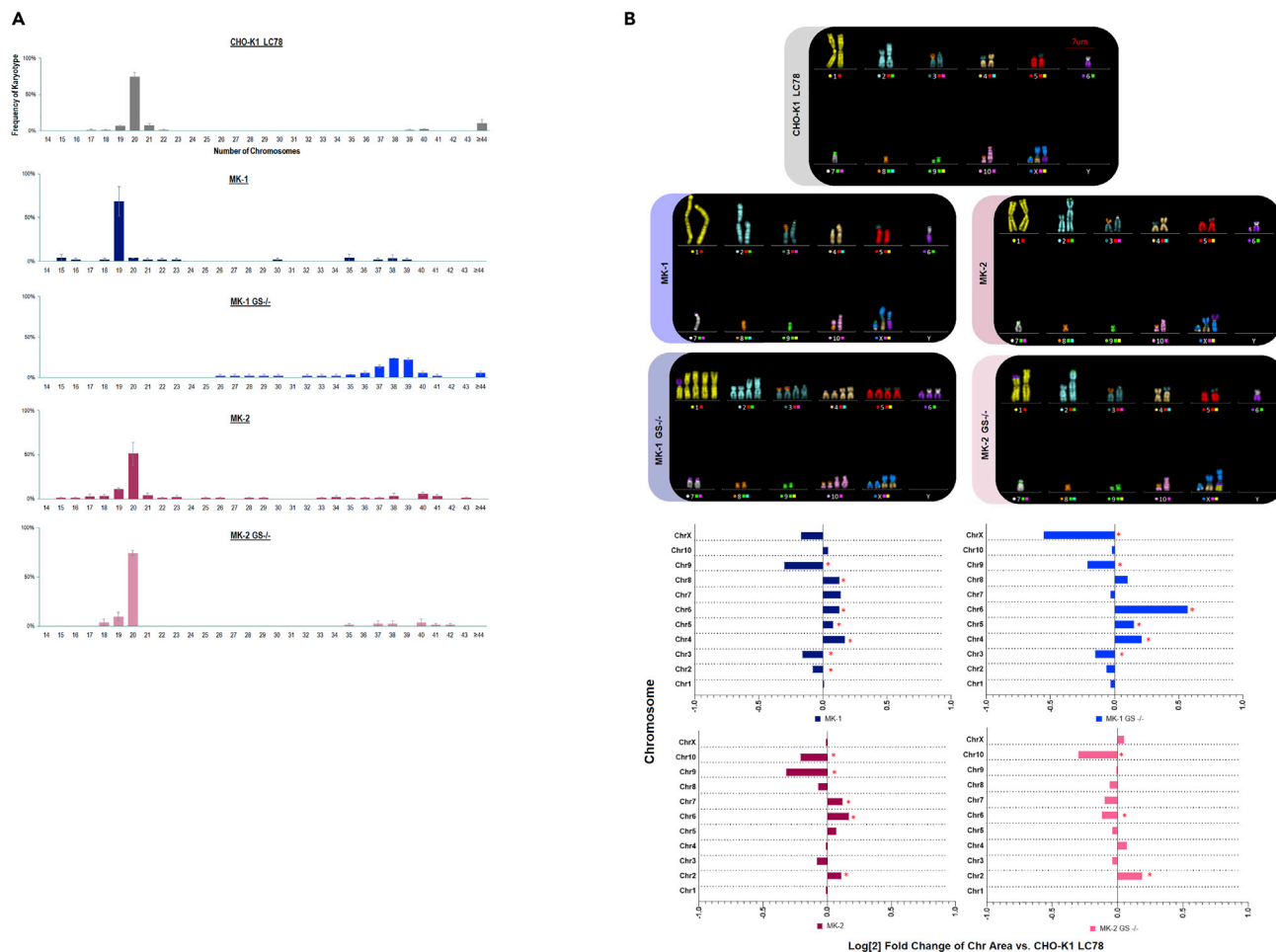


Figure 2. Characterization of host cell lineages by cytology

(A) Metaphase spreads were prepared from each cell line and the frequency distribution of metaphase chromosomes per cell was then counted and plotted in histogram form. Data from >30 cells is represented as mean \pm SD.

(B) Representative chromosome painting results from the five cell lines are depicted. Images representing the most common karyotype of >20 images and two experiments (right panel). The average distribution of individual chromosome pixels that occupied a cell's karyotype (genome) was averaged across >20 metaphases. This value was then represented as Log₂ fold changes versus CHO-K1 LC78 distribution of chromosome paints. The red asterisk represents values significantly changed using Student's t test with unequal variance ($p < 0.05$) versus the corresponding CHO-K1 LC78 chromosome (bottom panel).

To clearly differentiate common and unique structure features of cell lineages, we calculated the chromosome translocation frequency (Figure 3A) and represented it in heatmap form (Figures 3B and 3C). The majority of these chromosome translocations within the CHO-K1 LC78 line were well maintained in the progeny cell lineages, whereas translocations for individual lineages were also identified (Figures 3B and 3C). Some of the major translocations (appearing in 100% of adapted cells) may be resultant from adaptation to specific culture media. These translocations were stably inherited in the derivative GS^{-/-} cell lines. Noticeably, MK-1 and MK-1 GS^{-/-} demonstrated a unique translocation of Chr 2 and 9, whereas in MK-2 and MK-2 GS^{-/-} translocations, between Chr 4 and 7 were observed (Figure 3C). In recombinant hosts, these rearrangements were stable up to 75 generations (Figure S2). The inherited, stable nature of these rearrangements suggests process-dependent genetic drift events may be important to the characteristics of each lineage. In addition, besides inherited variations, new stable translocations (appearing in 100% of cells) were acquired in the GS^{-/-} progeny lineages, such as a Chr 1 and 6 fusion and Chr X and a 1 rearrangement in MK-1 GS^{-/-} (Figure 3C). Taken together, these data imply that culture conditions may spur enormous genetic and phenotypic plasticity in CHO cells.

Epigenetics is another mechanism for altering chromatin structure, which in turn controls genome accessibility and gene expression. To evaluate variations at the epigenetic level of individual lineages,

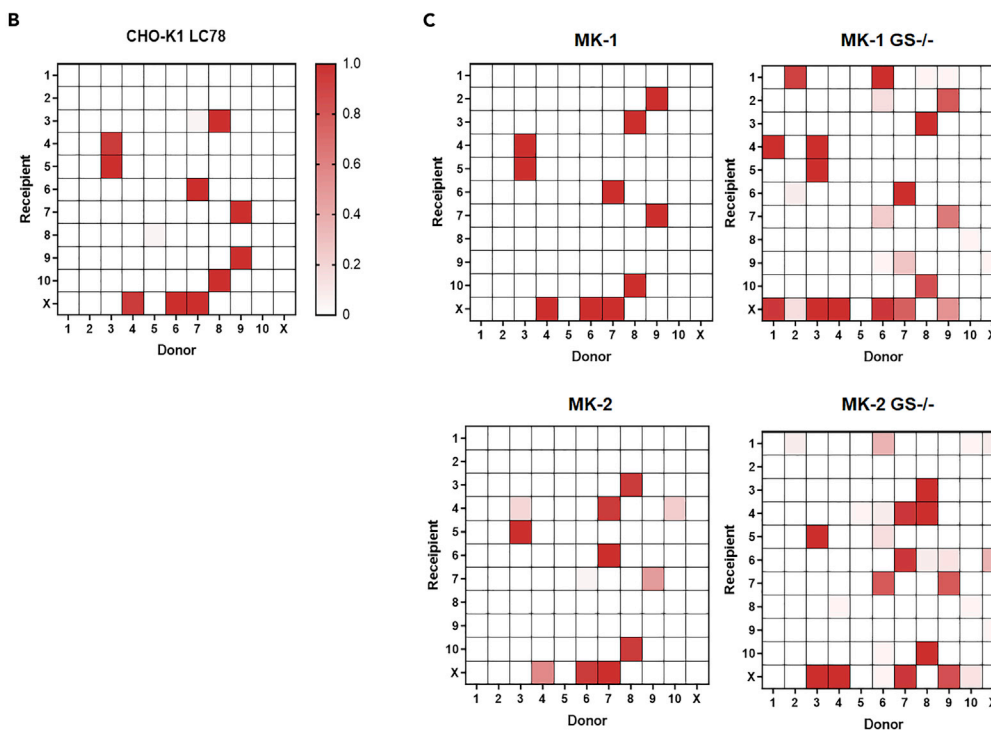
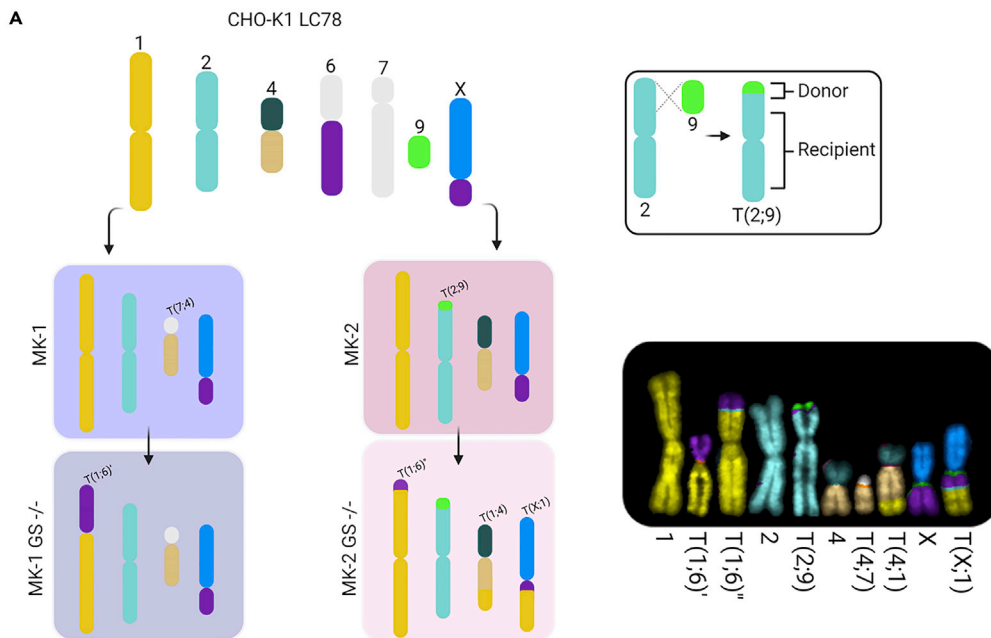


Figure 3. Characterization of host cell translocation events

(A) Schematic depicting how translocation events were quantified in CHO hosts. Recombination frequency involving donor or recipient chromosome exchanges (see materials and methods) were scored and averaged across >30 metaphase spreads in CHO-K1 LC78 (B) and individual developed new lineages (C). See also Figure S2.

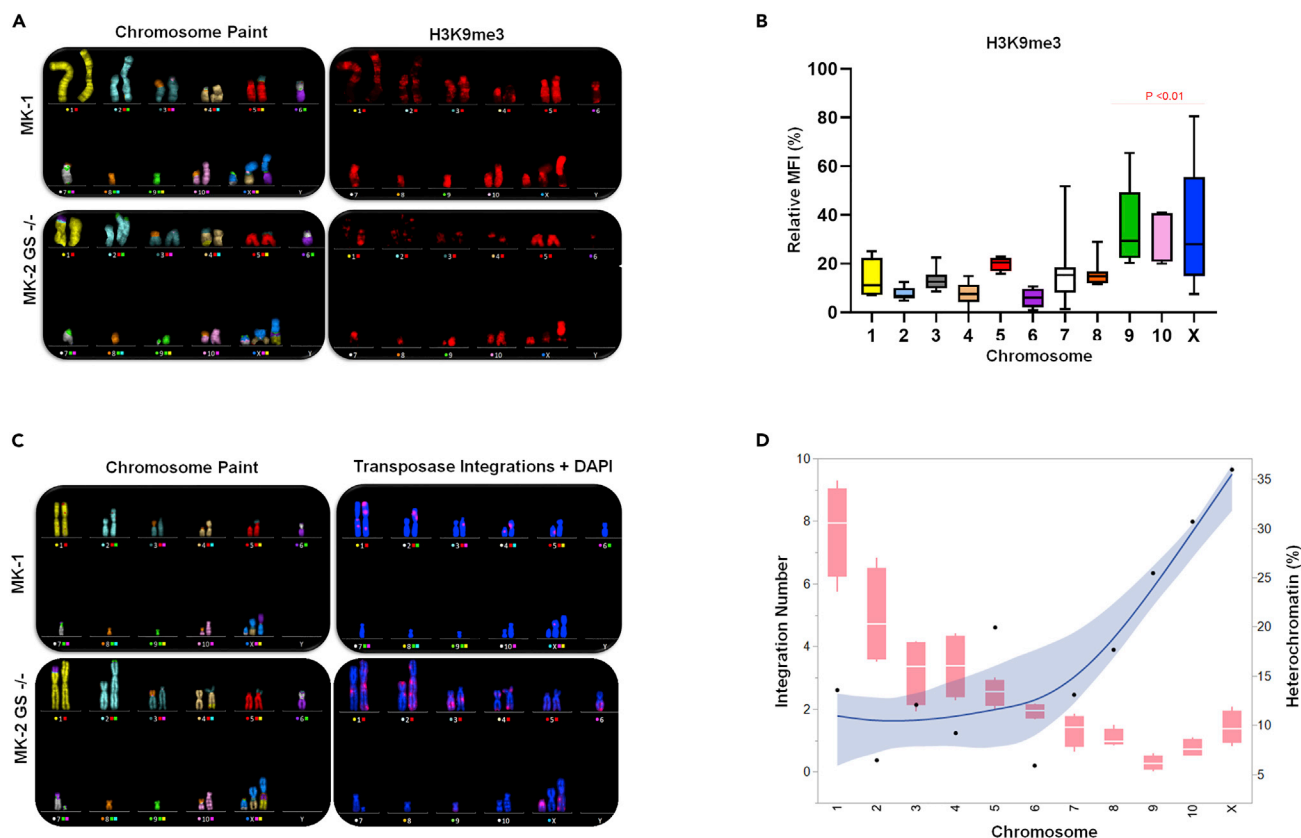


Figure 4. Association of heterochromatic regions with the CHO genome

(A) Formalin-fixed metaphase spreads were probed with whole-chromosome paints and H3K9me3.

(B) The mean intensity of H3K9me3 marker was quantified and represented in boxplot form. Each plot represents the average intensity of all chromosome components identified, including rearrangements, across 5–10 cells. Data are represented as boxplots where the centroid represents the median and whiskers represent the min and max. Chrs underneath the red bar (Chrs 9, 10, and X) indicate statistical significance ($p < 0.05$ using a Student's t test with unequal variance) versus those without a bar (Chrs 1–8).

(C) Host cells were transfected with recombinant DNA and transposase mRNA. Following selection and stable cell establishment, the recombinant pools were fixed and probed for recombinant DNA and chromosome paints.

(D) The number of integrations was counted in >20 cells and binned based on their chromosome identity and plotted as box-plots (left-y axis, see description of box-plots above) These data were overlaid with the content of H3K9me3 staining (blue trendline, right-y axis).

metaphase spreads were stained with the heterochromatic marker, H3K9Me3 (Figure 4A). This result demonstrated low intensity in Chr 1 through 3 but significant enrichment in Chr 9, 10, and X, suggesting enriched heterochromatin content in particular regions of the genome (Figure 4B). Because transposase-mediated integration prefers euchromatic regions of the genome for stable integration, transgenic loci should demonstrate recalcitrance to these H3K9Me3 stained regions (Buenrostro et al., 2015). Painted metaphase spreads from hosts with recombinant DNA demonstrated transposon-specific integration in Chr 1 through 3 but not in Chrs 9 and 10 (Figure 4C), which was instead inversely associated with H3K9Me3 staining (Figure 4D). Furthermore, an association between these heterochromatic regions and the regions lost through adaptation was noted (Figure 2B). This result suggests that genomic engineering or targeted integration may want to strategically avoid these chromosomes.

Artificial selection of adapted host lineages with desired recombinant protein productivity characteristics

The ultimate goal of host cell line development is to develop a lineage capable of high protein secretion in the context of industrial-scale reactors (Karthik et al., 2007). The productivity of adapted cell pools was evaluated by transfection with recombinant trastuzumab linked to a GS rescue gene. Recombinant cells were then selected in glutamine-free media supplemented with the GS inhibitor, MSX, and antibiotics. Here, we observed that doubling times were largely stabilized by five passages, or about two weeks for

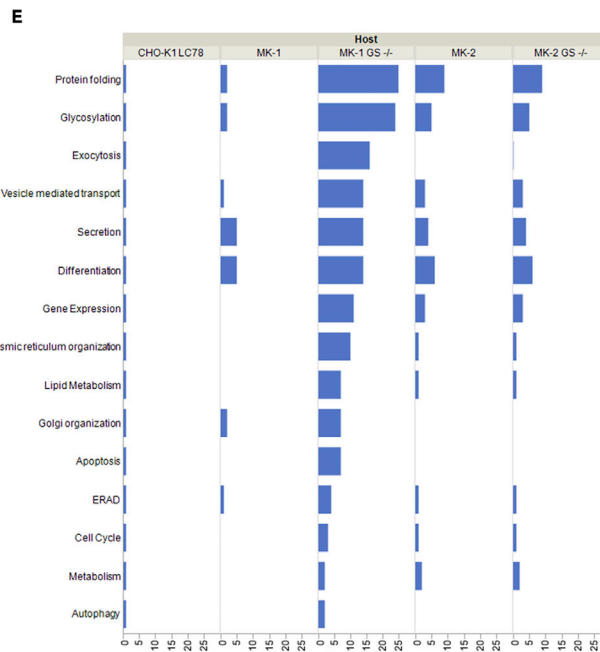
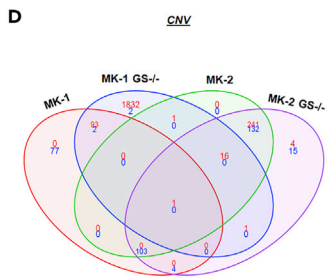
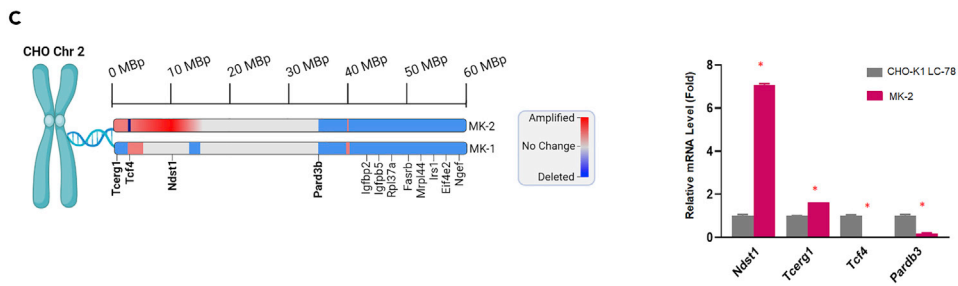
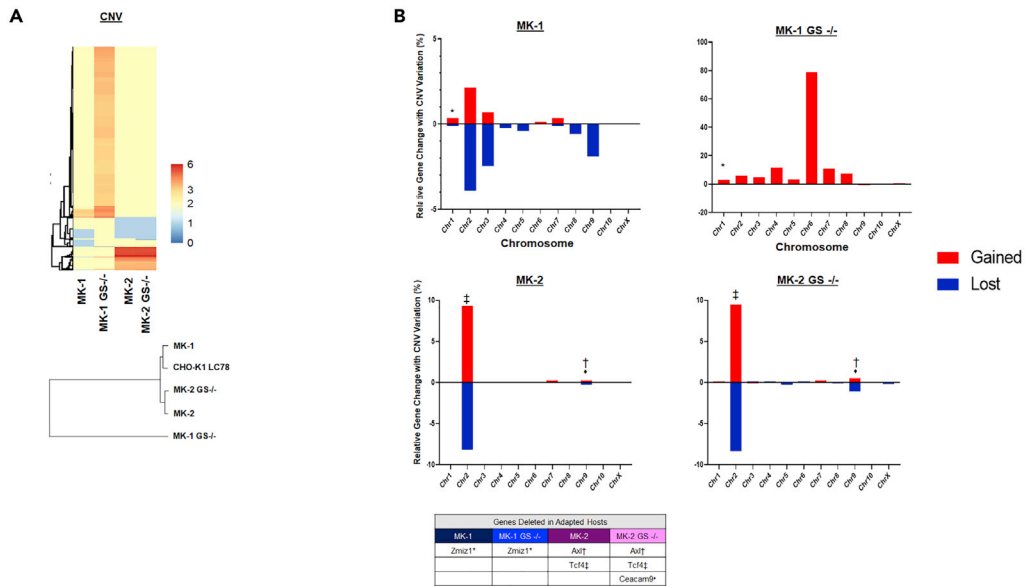


Figure 5. Genome-wide copy-number changes in adapted CHO cell cultures

Copy-number variation against CHO-K1 LC78 progenitors was detected using whole genome sequencing and CNVKit.

(A) Heatmap of the observed copy-number across cell lines (top); dendrogram of the cell lines generated using the log₂ ratio of the copy-number against CHO-K1 LC78 (bottom).

(B) CNV Kit results were mapped to individual chromosomes. Amplified regions are depicted in red, deleted regions in blue. The percentage of genes changes out of all genes on the indicated chromosome is represented as a histogram. The genes completely deleted in each host are indicated in the table below the graph. The symbol adjacent to each gene is depicted in the histogram, to designate where each deleted gene maps (Zmzi1=^{*}, Axl=[†], Tcf4=[‡], Ceacam9=[◆]).

(C and D) Schematic diagram showing CNV in a segment (0–60 MBp) of the chromosome 2 in MK-1 and MK-2 (left panel). The amplified and deleted regions are colored in red and blue, respectively. Genes are indicated with tick marks across the segment (left panel). Genetic polymorphisms in these coding regions is analyzed in Figure S4. Subset of genes were then verified for copy-number quantification in CHO-K1 LC78 and MK-2 by qPCR and plotted in bar plot. The results represent three biological experiments, averaged with error bars as \pm SD (right panel). Red asterisk indicates a significant difference ($p < 0.05$) using Student's *t* test between CHO-K1 LC78 and MK-2. (D) Venn diagram of amplified and deleted genes in each adapted host versus CHO-K1 LC78. Blue text indicates the number of genes with reduced copy number, and red text represents the number of genes amplified.

(E) Bar graph of the number of genes that fall into the indicated ontology categories. Only genes with statistically significantly changed CNVs were used for plotting. See also Figure S3 for cell phenotype relating to genetic changes.

most cell lines, whereas MK-1 GS^{-/-} experienced a clear delay (Figure S3A). When cells were subjected to a fed-batch production assay, we detected that specific productivity (qP) was enhanced in the GS^{-/-} cell lines, highlighting the importance of gene editing when using the GS system (Figure S3B). Notably, a 3-fold increase in productivity was shown by using MK-1 GS^{-/-} cell line, and, in support of this, increased mRNA levels of both heavy and light chains in passaging culture were likewise detected in that same cell line (Figure S3C). This observed qP increase was maintained when adjusting for each cell on a per volume basis (S3B). In addition, this mechanism of high productivity was not linked to recombinant DNA copy number, as only a small increase in copy number occurred in both MK-1 and MK-2 GS^{-/-} as compared with their wild-type counterpart (Figure S3D). Antibody COA profiles varied from host to host. MK-1 and MK-1 GS^{-/-} host lineages showed a reduction in high-molecular-weight (HMW) species, suggesting improved protein assembly efficiency and increased matured glycan species, such as galactosylation (G1F, G2F, G1, G2) (Figure S3B). These data demonstrate that the adaptation-mediated genetic changes result in different performance attributes during bioprocess production.

Copy number variation associated with the adaptation and selection process

Identification of the constant and pliable regions within the CHO genome may yield important clues regarding the endogenous genes involved in bioproduction. Because our cytogenetic data showed chromosome amplification and reduction for all CHO cell lineages, we hypothesized that the endogenous gene profile, especially in copy number (CNV), might also be impacted (See Data S1: CNV Data). We thus compared genome-wide CNV between adapted cell lineages and the parental CHO-K1 LC78 (see Materials and Methods & Figure 5A). Through this approach, we noted that the MK-2 and MK-2 GS^{-/-} hosts experienced significant CNV changes to a large region of Chr 2 (Figure 5B) and approximately 61% of the CNV changes in MK-2 were inherited by MK-2 GS^{-/-} (Figure 5B). On the contrary, the MK-1 lineage demonstrated less amplification and more deletions when compared with MK-2 (66% of changed genes in MK-1 were deletions versus 46% in MK-2) across the genome, which were more specific to Chr 2, 3, and 9 (Figure 5B). Interestingly, only 5% of the CNV variations observed in MK-1 were shared with MK-1 GS^{-/-} (Figures 5A and 5B) and were located on Chr 2. We then focused on the Chr 2 regions with similar CNV changes, which were conserved among different lineages (Figure 5C). Remarkably, 62% of the genes lost in Chr 2 in both MK-1 and MK-2 were identical, suggesting common evolution pressures during the adaptation process in this study (Figure 5C). Notably, this region was enriched with Igfbp5, Igfbp2, and the Irs1 genes, which regulate cell response to insulin and insulin-like growth factor (IGF) (Figure 5C). In addition, many of the lost genes are among the negative regulators of growth and promoters of focal adhesion and differentiation, particularly in the mTOR, AKT, and PI3K pathways (data not shown). Taken together, these results suggest that culture process changes, such as serum removal, and specific selection criteria, such as suspension culture, may drive some of the common genetic changes identified earlier.

Besides the common features, different lineages also showed unique CNV profiles. Analysis of the deleted genes confined to Chr 3 exposed a subset of positive regulators of cell junction assembly, such as Mpp7 and Pard3, in both MK-1 and MK-1 GS^{-/-}. These genes were perhaps deleted as a result of a rigorous selection of nonadhesive cells (see Materials and Methods), and the deletion of a Chr 3 arm is visible in painted and quantified MK-1 and MK-1 GS^{-/-} karyotypes. Interestingly, this Chr 3 region is maintained in both MK-2 and MK-2 GS^{-/-} (Figures 5B and 5D). These lineages instead contain a uniquely amplified

Chr 2 region containing repressors of adherence junction assembly (Sema5a, Sema6a, Pdgfrb, and Csf1r) (Figures 5B and 5D). CNV variation was further stimulated by the gene editing process. Following the GS knockout through ZFN, MK-1 GS^{-/-} cells exhibited a doubling of nearly the entire genome, some of which was amplified nonproportionally (Figures 2, 5B and 5D). Of note, much of this overrepresented amplification was localized to Chr 6 (>80% of the genes localized to Chr 6 were amplified). The region was relatively enriched in factors implicated in endoplasmic reticulum stress, endoplasmic-reticulum-associated protein degradation (ERAD), and the unfolded protein response (UPR) (Figure 5E). Overall, both common and unique CNV profiles might be used as biomarkers to trace individual lineages or offer useful hints to improve cell lines via chromosomal engineering.

Polymorphic variations associated with the adaptation and selection process

In addition to CNV, genetic variants such as single nucleotide polymorphisms (SNPs) and small insertions/deletions (INDELs) in protein coding regions were also observed. We hypothesized that spontaneously acquired and maintained polymorphisms in developed cell lineages may be indicative of selection by evolutionary pressure. With respect to genetic drift, we observed that on average approximately 528 ± 117 coding SNPs and 93.5 ± 2.1 coding INDELs were spontaneously generated in either MK-1 or MK-2 following adaptation (Figure S4A). These mutations were unique, and not found within the progenitor, CHO-K1 LC78, suggesting they may be driven through adaptation. With respect to inheritance, approximately 50% of both the SNPs and INDELs were maintained in the corresponding GS^{-/-} progeny. Retention of specific SNPs and INDELs suggests that some of these spontaneously acquired mutations may be helpful to survive in the modified cell culture environments. In support of these data, gene set enrichment analysis using spontaneously acquired genetic variants from both MK-1 and MK-2 revealed a genetic network implicated in focal adhesion (such as Itgax, Lamb3, Vav1, Vwf, Pak2, Nid1, Rarb etc.). In addition, on average, 690 ± 36 coding SNPs and 157 ± 0.7 coding INDELs were identified in the GS^{-/-} lines but were not found in their progenitors nor in CHO-K1 LC78 (Data not shown), suggesting that artificial genomic manipulation may accelerate continuous genetic drift.

Although cell culture and genomic manipulation are drivers of genetic variation, selection criteria might also play a critical role in inheritance. To evaluate this hypothesis, we characterized the ontologies of the common genes containing polymorphisms exclusively in the MK-1 and MK-2, but not CHO-K1 LC78 cell lines. Because lineages were screened and selected based on adaptation to suspension culture, serum removal, and productivity, as expected, we identified spontaneously acquired genetic polymorphisms associated with adaptation and bioproduction. These examples include focal adhesion turnover and chromatin assembly (Histone H3-K9 methylation, positive regulation of chromatin binding, positive regulation of histone ubiquitination) (Figure S4B). Alternatively, the commonly conserved polymorphism among all cell lineages, including CHO-K1 LC78, included genes associated with vesicle-mediated transport and detoxification of reactive oxygen species (ROS; Figure S4C). Of notable interest, we identified four genes containing SNPs—Cat, Ncf2, Sod1, Txnrd2—all of which are involved in the removal of ROS (Figure S4C). The generation of ROS is a common issue during industrial bioproduction, affecting both cell health and product quality, highlighting the importance of this gene network (Chevallier et al., 2020). Inspection of SOD1, a gene well characterized by mutagenesis, revealed the presence of the K137M mutation in all cell lines, including CHO-K1 LC78, which abolishes a key acetyl lysine site. Acetylation of SOD1 has been shown to sensitize cancer cells to genotoxic agents, and this mutation may offer protection against oxidation within a bioreactor setting (Lin et al., 2015).

Inherited and novel gene expression signatures

Phenotypic changes can also be regulated through epigenetic mechanisms (Schimke, 1988; Borth and Hu, 2018; Vishwanathan et al., 2014). Through RNA-Seq (See Data S2: Transcriptomic Data), we observed both newly emergent and conserved transcriptomic changes across host lineages (Figures 6A and 6B). The majority of genes ($91\% \pm 3\%$ across all hosts) changed at the mRNA level did not have accompanied CNV changes, suggesting that most adaptation-based reprogramming occurs through transcriptional regulation (Figure 6C). Interestingly, we observed that 64% of the transcriptomic changes were inherited to the MK-2 GS^{-/-} clone from MK-2, whereas this number was 11.5% between the MK-1 and MK-1 GS^{-/-} clones. We reasoned that the genes with tandem changes across CNV and gene expression across different lineages may be especially important founder events, critical for host cell establishment (Namba et al., 1996). After performing this filtering step, we identified 18 genes localized to a contiguous region on Chr 2 (MBp: 44–59). Remarkably, this genetic subset experienced similar changes in both CNV and gene

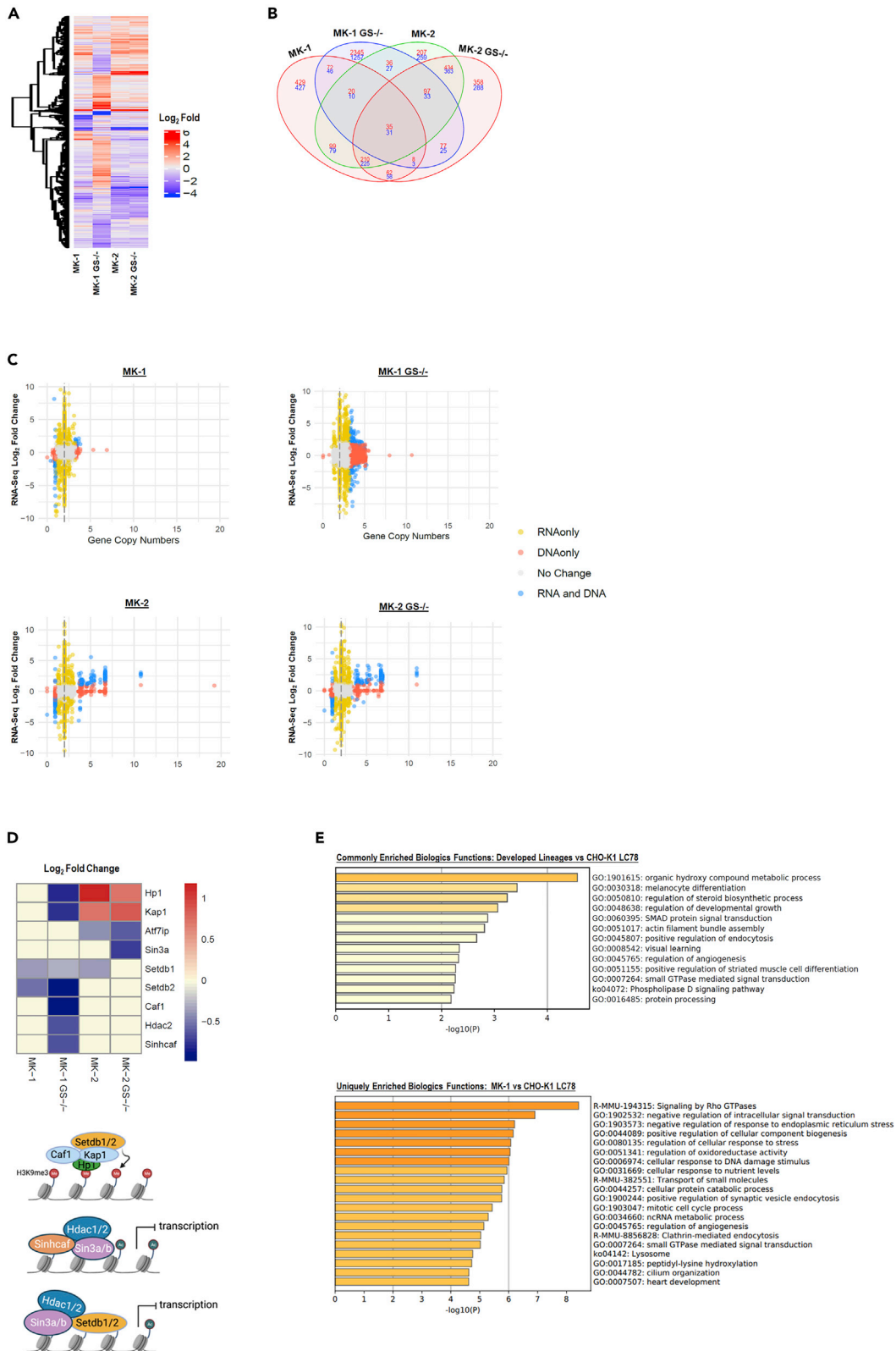


Figure 6. Comparison of transcriptome level changes

The significant differential expression changes are represented as Log2 fold change versus the CHO-K1 LC78 lineage. RNASeq counts were normalized, and the gene expression Log2 fold changes versus CHO-K1 LC78 in the four cell lines were visualized as (A) heatmap of genes significantly differentially expressed versus CHO-K1 LC78. The log2 fold change of RNA expression was plotted in the heatmap. Genes were clustered, and dendrogram was drawn with Euclidean distance and complete linkage algorithm.

(B) Venn diagram of significantly differentially expressed genes. The number of genes increased and decreased versus CHO-K1 LC78 is colored in red and blue, respectively.

(C) Scatterplot of RNA log2 fold change versus DNA copy number. Genes from host lineages were bucketed into four categories: genes significantly changed on both a transcriptomic and genomic level (blue dots), significantly changed in RNA expression only (yellow dots), significantly changed in DNA copy only (red dots) or no changes (grey dots). The gray dashed vertical line indicates a copy number of 2.

(D) Heatmap of RNA expression in key epigenetic regulators (top panel). Log2 fold change versus CHO-K1 LC78 was used for plotting the genes in the heatmap with red color indicating increased and blue for decreased RNA expression. Schematic diagram depicting how these regulators affect heterochromatin (bottom panel). DNA and chromatin are colored gray. Blue circles marked Ac represent histone acetylation, and red circles marked Me represent H3K9 methylation. Colored ovals represent different epigenetic regulators.

(E) Gene set enrichment analysis of differentially expressed genes versus CHO-K1 LC78 that are common (top panel) to all hosts or unique to MK-1 GS^{-/-} (bottom panel). X-axis represents the $-\log_{10}$ p-value of the enrichment, and the significantly enriched biological process and pathways are indicated on the y-axis. See also [Figure S3](#) for cell phenotype relating to genetic changes.

expression across the adaptation arms (haploid in both hosts and 50%–60% loss in expression versus CHO-K1 LC78). Here we observed clear enrichment of gene clusters implicated in protein translation (Eif4e2), focal adhesion complexes (Dock10, Ngef), as well as insulin signaling (Irs1, Igfbp2, Igfbp5) (see also [Figure S4B](#)). These gene ontology clusters appeared to be relevant to the selection strategies we employed to produce desired cell phenotypes. Dock10 and Ngef, for example, promote Cdc42-mediated cytoskeletal rearrangement and may enhance the suspension-like CHO phenotype. Likewise, the Igfbp family modulates the response to insulin and insulin-like growth factors. Removing serum from the media may hypersensitize cell lines to the growth factors utilized in chemically defined media, such as IGF1. Lastly, given that Eif4e2 is a major repressor of translation, its attenuation may be linked to selection for enhanced productivity ([Morita et al., 2012](#)).

Lastly, histone modification and DNA methylation are primary mechanisms for epigenetic control, which can affect chromatin structure and modulate transcriptional initiation. Expression patterns of key complex subunits of epigenetic modulation complexes, such as HP1, Setdb1, Sin3a, Hdac2, and KAP1, were different in the individual host cell lines. Alternatively, degrees of inheritance were observed in the lineages with same adaptation method ([Figure 6D](#)). In addition, conserved transcriptomic changes were detected among all adapted lineages involving genetic networks associated with actin filament bundle assembly, small GTPase assembly, and angiogenesis, all of which are involved in morphological changes ([Figure 6E](#)) ([Shridhar et al., 2017](#)). Furthermore, we also observed common changes to metabolism particularly with steroidal fatty acid synthesis and protein processing ([Figure 6E](#)). Specific changes were also observed. Chiefly, we identified an exclusive upregulation of the DNA damage stimulus response and negative regulation of the ER stress pathways in MK-1 GS^{-/-}, in agreement with the high productivity and high plasticity of this cell line ([Figure 6E](#)).

Deep sequencing and cytology revealed mechanistic hints regarding the unusually strong qP phenotype observed in the MK-1 GS^{-/-} cell line. These hints were further explored by mapping the transcripts observed in our genetic screens to signaling pathways ([Figure 7](#)). The unfolded protein response (UPR) is the primary signaling network activated in response to the accumulation of unfolded and/or misfolded protein in the ER endoplasmic reticulum (ER) ([Chakrabarti et al., 2011](#); [Du et al., 2013](#); [Schröder and Kaufman, 2005](#)). This pathway directly impacts productivity and product quality of therapeutic proteins ([Chakrabarti et al., 2011](#); [Du et al., 2013](#); [Schröder and Kaufman, 2005](#)). Here, we observed increased levels of the transcripts for critical factors, such as ATF6, XBP1, Sec61, BiP, HSP90, DNAJB9/11 and PDIA2,3, EDEM1/3, UGGT1, Kdelr1, and Cnx ([Figure 7B](#)), which act in tandem to improve secretion capacity, including protein folding and glycosylation efficiency ([Lyman and Schekman, 1997](#)).

Clonal variability in size occurs during cell line development but is not well mechanistically described. MK-1 GS demonstrated a significant increase in cell size and genome content ([Figures 1F, 2A and 2B](#)), suggesting the potential association of these two factors. In addition, the high levels of chromosome aberrations, translocations, and slow doubling time in this cell line ([Figures 1, 2, and 3](#)) also suggest genetic stress ([Figures 7C and 7D](#)). Consistent with these phenotypes, we observed overexpression of several critical genes implicated in homologous recombination (HR) and nonhomologous end-joining (HR and NHEJ) in

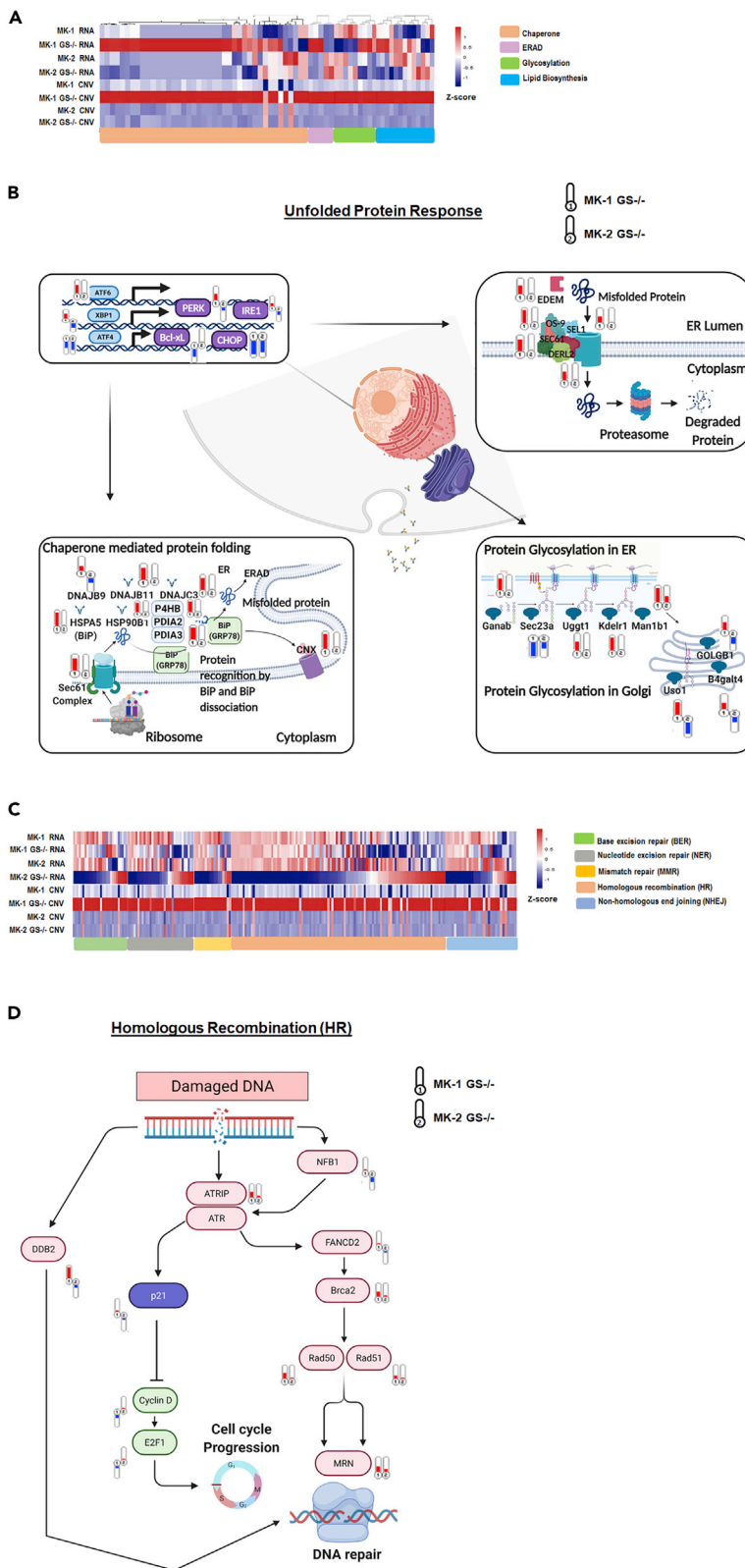


Figure 7. Diversity of UPR and DNA damage response in different host lineages

(A) Heatmap showing differences in RNA expression and CNV for genes involved in different function of UPR including chaperone-mediated protein folding, ERAD, and glycosylation. Note that the MK-1 GS^{-/-} cell line exhibited significantly upregulated expression at both DNA and RNA levels across all subsystems.

(B) Gene set enrichment analysis of representative gene set for UPR showing enriched UPR in MK-1 GS^{-/-} compared with MK-2 GS^{-/-} cell line. UPR pathway schematic showing upregulation of genes involved in protein folding, lipid biogenesis, ERAD, and glycosylation in MK-1 GS^{-/-} cell line (thermometer-1) compared with MK-2 GS^{-/-}

(thermometer-2). (C) Heatmap showing differences in gene expression and CNV for genes involved in DNA repair. Genes are categorized by the distinct repair pathway. Note that MK-1 GS^{-/-} cell line exhibited increased homologous recombination (HR) but not other pathways such as base excision repair (BER), nucleotide excision repair (NER), mismatch repair (MMR), or nonhomologous end joining (NHEJ). (D) Gene set enrichment analysis of representative gene set for homologous recombination (HR) genes in MK-1 GS^{-/-} compared with MK-2 GS^{-/-} cell line. A schematic depicting the HR pathway in MK-1 GS^{-/-} cell line (thermometer-1) compared with MK-2 GS^{-/-} (thermometer-2). Red and blue thermometers represent upregulation and downregulation of mRNA level, respectively. All CNV and RNA differential expression were compared with CHO-K1 LC78. See also [Figure S3](#) for cell phenotype relating to genetic changes.

MK-1 GS^{-/-} but not other DNA repair pathways, such as nucleotide excision repair (NER) and base excision repair (BER) ([Figure 7C](#)). In support of this, we observed significant upregulation of the upstream DNA damage sensors ATRIP, DDB2, and Nfbd1 in MK-1 GS^{-/-} ([Giglia-Mari et al., 2011](#)) ([Figure 7D](#)). Activation of these markers coincided with elevated Brca2, Rad50, and Rad51, genes implicated in double-strand break repair ([Figure 7D](#)). As the frequency of double-strand breaks increases, the cell cycle is suspended, and cells undergo arrest ([Giglia-Mari et al., 2011](#)). Accordingly, we observed increased transcripts levels of p21 and decreased Cyclin-D and E2F1 ([Barr et al., 2017](#)) ([Figure 7D](#)). Taken as a whole, these data support the hypothesis that the polyploidization and increased cell size of MK-1 GS^{-/-} confers a marked increase in protein folding and ER capacity. Simultaneously, such changes result in increased stress to the genome, associated with the DNA damage, and this also has implications in replication machinery. These stressors result in impaired cell-cycle progression as compared with the MK-1 progenitor.

DISCUSSION

Thus far, genetic instability of CHO cells has been observed via next-generation sequencing (NGS), physical mapping, and other approaches. However, it remains unclear how culture processes during cell line development impact cell line performance and subsequent product quality. Our results demonstrate that even though genetic drift of CHO cells continuously drives populations toward genetic variation instead of uniformity over time, some heritable traits in both genetic and phenotypic levels are sustained across different lineages. The observed data provide a framework to understand the mechanisms of adaptive plasticity in CHO cell lines, which enables rational process development to improve productivity and quality.

During our cell line development efforts, we observed a conserved series of genetic modifications that may be implicated in the genetic reprogramming steps that result in the serum-free, suspension-cell phenotype. This was manifested in structural and transcriptomic reduction of Chr 2 (among others) and was conserved in the derived GS^{-/-} lineages. In addition to common changes, clear radiation among the four independent cell lines was also observed, highlighted by unique genomic rearrangements, mutations, structural alterations, and transcriptomic changes. These inherited changes might impact cell phenotype and antibody characteristics, obviating high adaptive fitness in CHO. These data also provide context for the process-related changes that affect recombinant protein COAs in biopharmaceuticals. Although these differences are manifested in multiple independently derived lineages, future studies involving genome engineering will be required to validate whether the stress was causative or if the genomic variation was part of the random mutagenesis and plasticity of CHO cells.

Immortal cell lines are subjected to high mutagenesis rates *in vitro*, accelerating selection especially in the context of process development ([Brody et al., 2018](#); [Wurm, 2017](#)). This concept is analogous to niche filling in ecological setting ([Wurm, 2017](#)). These adapted phenotypes occur with independent mutations *in cis*, imbuing additional unique characteristics. In our work, for example, transcriptomic changes were inherited from MK-1 to MK-1 GS^{-/-} but occurred with two other unique changes, namely increased DNA damage and high productivity. The productive phenotype was obviously favorable, but the instability phenotype was not. It remains somewhat paradoxical that the instability observed in CHO cells is both a strength

and limitation of the system for biotechnology applications. For example, chromosomally unstable cells are intimately tied to productivity traits (Lai et al., 2015). Such cell lines are highly desired during cell line development phase but can represent a liability during the production and scale up, in which cell line stability and robustness is essential.

In the context of cell engineering, our work suggests several tantalizing hints with regard to the CHO genome. We describe that stability of particular genomic regions is profoundly diverse. Predominantly we observe that the smaller chromosomes, (mostly 7–10 and X) were both more labile and more heterochromatic in the adapted cell lines. Alternatively, many of the larger chromosomes, especially Chr 1, were relatively stable, experiencing relatively less recombination and associated genetic changes. It is possible that these genetically silent, superfluous regions are associated with irrelevant developmental programs, and their silencing and loss results in a growth advantage. These regions may contain genes essential to primary tissue but not immortal cell lines.

Of exceptional interest, in the highly productive and polyploid MK-1 GS^{-/-} line, we observed that the genome was highly amplified albeit not completely doubled, with some regions more selectively expanded in CNV than others. These hyper-amplified regions contained more content from Chr 6 and were accompanied by enrichment of UPR and ER stress genes. In the advent of chromosome transfer and biomarker establishment, these findings offer clear suggestions for rational cell engineering (Oshimura et al., 2015; Wlaschin et al., 2006).

We ultimately conclude that although tremendous plasticity and diversity of CHO hosts can be a challenge in process development, this inherent variability represents an opportunity to understand the impact of *in vitro* genetic drift, artificial selection, and adaptive fitness and ultimately to utilize it in biotechnology applications.

Limitations of the study

A limitation of this study is the number of cell lines analyzed. The adaptation strategy arms resulted in genomic analysis of a total of four clones of the many derived by the MK-1 and MK-2 pools. We acknowledge a small sample was used to support our conclusions. In addition, our work suggests but does not demonstrate that many omics changes contribute to phenotype. Future work involving genomic engineering and/or synthetic biology is required to validate our observations.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - Cell adaptation
 - Glutamine synthetase knockout and screening
 - Whole genome sequencing and analyses
 - RNA sequencing and analysis
 - Pathways analysis and statistical analysis
 - Cytogenetics
 - FISH
 - Probe generation
 - Chromosome painting
 - Gene expression analysis
 - Recombinant protein production assay
 - Genomic DNA copy number and gene expression quantification assays
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.104074>.

ACKNOWLEDGMENTS

We thank Dr. Larry Chasin for providing CHO-K1 LC78 cell line and manuscript review. We also thank Dr. Ying Huang for manuscript review.

AUTHOR CONTRIBUTIONS

S. Huhn, A. Kumar, B. Jiang, and R. Liu contributed to wet lab experiments, including cell culture and cytology. M. Chang and A. Kumar performed the bioinformatics. M. Betenbaugh, H. Lin, G. Nyberg, and Z. Du helped design the experiments and gave critical oversight.

DECLARATION OF INTERESTS

S. Huhn, A. Kumar, B. Jiang, R. Liu, H. Lin, G. Nyberg, and Z. Du are employees of Merck Sharp & Dohme Corp., a subsidiary of Merck & Co., Inc., Kenilworth, NJ, USA and potentially own stock and/or hold stock options in Merck & Co., Inc., Kenilworth, NJ, USA. Merck & Co., Inc provided support in the form of salaries for all authors but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the “author contributions” section.

Received: September 28, 2021

Revised: January 17, 2022

Accepted: March 11, 2022

Published: April 15, 2022

REFERENCES

- Auer, N., Hrdina, A., Hiremath, C., Vcelar, S., Baumann, M., Borth, N., and Jadhav, V. (2018). ChromaWizard: an open source image analysis software for multicolor fluorescence in situ hybridization analysis. *Cytometry A*. *93*, 749–754.
- Barr, A.R., Cooper, S., Heldt, F.S., Butera, F., Stoy, H., Mansfeld, J., Novák, B., and Bakal, C. (2017). DNA damage during S-phase mediates the proliferation-quiescence decision in the subsequent G1 via p21 expression. *Nat. Commun.* *8*, 14728.
- Borth, N., and Hu, W.S. (2018). Enhancing CHO by systems biotechnology. *Biotechnol. J.* *13*, e1800077.
- Brinkrolf, K., Rupp, O., Laux, H., Kollin, F., Ernst, W., Linke, B., Kofler, R., Romand, S., Hesse, F., Budach, W.E., et al. (2013). Chinese hamster genome sequenced from sorted chromosomes. *Nat. Biotechnol.* *31*, 694–695.
- Brody, Y., Kimmerling, R.J., Maruvka, Y.E., Benjamin, D., Elacqua, J.J., Haradhvala, N.J., Kim, J., Mouw, K.W., Frangaj, K., Koren, A., et al. (2018). Quantification of somatic mutation flow across individual cell division events by lineage sequencing. *Genome Res.* *28*, 1901–1918.
- Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* *109*, 21.29.1–21.29.9.
- Butler, M., and Meneses-Acosta, A. (2012). Recent advances in technology supporting biopharmaceutical production from mammalian cells. *Appl. Microbiol. Biotechnol.* *96*, 885–894.
- Chakrabarti, A., Chen, A.W., and Varner, J.D. (2011). A review of the mammalian unfolded protein response. *Biotechnol. Bioeng.* *108*, 2777–2793.
- Chevallier, V., Andersen, M.R., and Malphettes, L. (2020). Oxidative stress-alleviating strategies to improve recombinant protein production in CHO cells. *Biotechnol. Bioeng.* *117*, 1172–1186.
- Dhiman, H., Gerstl, M.P., Ruckerbauer, D., Hanscho, M., Himmelbauer, H., Clarke, C., Barron, N., Zanghellini, J., and Borth, N. (2019). Genetic and epigenetic variation across genes involved in energy metabolism and mitochondria of Chinese hamster ovary cell lines. *Biotechnol. J.* *14*, e1800681.
- Du, Z., Treiber, D., McCoy, R.E., Miller, A.K., Han, M., He, F., Domnitz, S., Heath, C., and Reddy, P. (2013). Non-invasive UPR monitoring system and its applications in CHO production cultures. *Biotechnol. Bioeng.* *110*, 2184–2194.
- Fan, L., Kadura, I., Krebs, L.E., Hatfield, C.C., Shaw, M.M., and Frye, C.C. (2012). Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells. *Biotechnol. Bioeng.* *109*, 1007–1015.
- Feichtinger, J., Hernández, I., Fischer, C., Hanscho, M., Auer, N., Hackl, M., Jadhav, V., Baumann, M., Krempl, P.M., Schmid, C., et al. (2016). Comprehensive genome and epigenome characterization of CHO cells in response to evolutionary pressures and over time. *Biotechnol. Bioeng.* *113*, 2241–2253.
- Giglia-Mari, G., Zotter, A., and Vermeulen, W. (2011). DNA damage response. *Cold Spring Harb Perspect. Biol.* *3*, a000745.
- Greber, D., and Fussenegger, M. (2007). Multi-gene engineering: simultaneous expression and knockdown of six genes off a single platform. *Biotechnol. Bioeng.* *96*, 821–834.
- Hacker, D.L., De Jesus, M., and Wurm, F.M. (2009). 25 years of recombinant proteins from reactor-grown cells - where do we go from here? *Biotechnol. Adv.* *27*, 1023–1027.
- Hamaker, N.K., and Lee, K.H. (2018). Site-specific integration ushers in a new era of precise CHO cell line engineering. *Curr. Opin. Chem. Eng.* *22*, 152–160.
- Huhn, S.C., Ou, Y., Kumar, A., Liu, R., and Du, Z. (2019). High throughput, efficacious gene editing & genome surveillance in Chinese hamster ovary cells. *PLoS One* *14*, e0218653.
- Huhn, S.C., Ou, Y., Tang, X., Jiang, B., Liu, R., Lin, H., and Du, Z. (2021). Improvement of the efficiency and quality in developing a new CHO host cell line. *Biotechnol. Prog.* *37*, e3185.
- Karthik, J., Wlaschin, K.F., Hu, W.S., and Yap, M.G.S. (2007). Recombinant protein therapeutics from CHO cells – 20 years and counting. *Chem. Eng. Prog.* *103*, 40–47.
- Kildegaard, H.F., Baycin-Hizal, D., Lewis, N.E., and Betenbaugh, M.J. (2013). The emerging CHO systems biology era: harnessing the ‘omics revolution for biotechnology. *Curr. Opin. Biotechnol.* *24*, 1102–1107.

- Kim, J.Y., Kim, Y.G., and Lee, G.M. (2012). CHO cells in biotechnology for production of recombinant proteins: current state and further potential. *Appl. Microbiol. Biotechnol.* 93, 917–930.
- Kim, M., O’callaghan, P.M., Droms, K.A., and James, D.C. (2011). A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. *Biotechnol. Bioeng.* 108, 2434–2446.
- Lai, Y., Yamano, N., Onitsuka, M., and Omasa, T. (2015). Increased antibody productivity in Chinese hamster ovary cells through induction of chromosomal instability by cell fusion. *BMC Proc.* 9, P11.
- Lakshmanan, M., Kok, Y.J., Lee, A.P., Kyriakopoulos, S., Lim, H.L., Teo, G., Poh, S.L., Tang, W.Q., Hong, J., Tan, A.H., et al. (2019). Multi-omics profiling of CHO parental hosts reveals cell line-specific variations in bioprocessing traits. *Biotechnol. Bioeng.* 116, 2117–2129.
- Lewis, N.E., Liu, X., Li, Y., Nagarajan, H., Yerganian, G., O’Brien, E., Bordbar, A., Roth, A.M., Rosenbloom, J., Bian, C., et al. (2013). Genomic landscapes of Chinese hamster ovary cell lines as revealed by the *Cricetulus griseus* draft genome. *Nat. Biotechnol.* 31, 759–765.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, G.P.D.P. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Lin, C., Zeng, H., Lu, J., Xie, Z., Sun, W., Luo, C., Ding, J., Yuan, S., Geng, M., and Huang, M. (2015). Acetylation at lysine 71 inactivates superoxide dismutase 1 and sensitizes cancer cells to genotoxic agents. *Oncotarget* 6, 20578–20591.
- Lyman, S.K., and Schekman, R. (1997). Binding of secretory precursor polypeptides to a translocon subcomplex is regulated by BIP. *Cell* 88, 85–96.
- Morita, M., Ler, L.W., Fabian, M.R., Siddiqui, N., Mullin, M., Henderson, V.C., Alain, T., Fonseca, B.D., Karashchuk, G., Bennett, C.F., et al. (2012). A novel 4EHP-GIGYF2 translational repressor complex is essential for mammalian development. *Mol. Cell Biol* 32, 3585–3593.
- Namba, M., Mihara, K., and Fushimi, K. (1996). immortalization of human cells and its mechanisms. *Crit. Rev. Oncog* 7, 19–31.
- O’Brien, S.A., Ojha, J., Wu, P., and Hu, W.S. (2020). Multiplexed clonality verification of cell lines for protein biologic production. *Biotechnol. Prog.* 36, e2978.
- Omasa, T., Onitsuka, M., and Kim, W.D. (2010). Cell engineering and cultivation of Chinese hamster ovary (CHO) cells. *Curr. Pharm. Biotechnol.* 11, 233–240.
- Oshimura, M., Uno, N., Kazuki, Y., Katoh, M., and Inoue, T. (2015). A pathway from chromosome transfer to engineering resulting in human and mouse artificial chromosomes for a variety of applications to bio-medical challenges. *Chromosome Res.* 23, 111–133.
- Putri, G.H., Anders, S., Pyl, P.T., Pimanda, J.E., and Zanini, F. (2021). Analysing high-throughput sequencing data in Python with HTSeq 2.0. Preprint at arXiv. <https://doi.org/10.48550/arXiv.2112.00939>.
- Scarcelli, J.J., Hone, M., Beal, K., Ortega, A., Figueroa, B., Starkey, J.A., and Anderson, K. (2018). Analytical subcloning of a clonal cell line demonstrates cellular heterogeneity that does not impact process consistency or robustness. *Biotechnol. Prog.* 34, 602–612.
- Schimke, R.T. (1988). Gene amplification in cultured cells. *J. Biol. Chem.* 263, 5989–5992.
- Schröder, M., and Kaufman, R.J. (2005). ER stress and the unfolded protein response. *Mutat. Res.* 569, 29–63.
- Shridhar, S., Klanert, G., Auer, N., Hernandez-Lopez, I., Kařidula, M.M., Hackl, M., Grillari, J., Stralis-Pavese, N., Kreil, D.P., and Borth, N. (2017). Transcriptomic changes in CHO cells after adaptation to suspension growth in protein-free medium analysed by a species-specific microarray. *J. Biotechnol.* 257, 13–21.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A.* 102, 15545–15550.
- Talevich, E.A.-O., Shain, A.H., Botton, T., and Bastian, B.C. (2016). CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput. Biol.* 12, e1004873.
- Tu, C., Fiandalo, M.V., Pop, E., Stocking, J.J., Azabdaftari, G., Li, J., Wei, H., Ma, D., Qu, J., Mohler, J.L., et al. (2018). Proteomic analysis of charcoal-stripped fetal bovine serum reveals changes in the insulin-like growth factor signaling pathway. *J. Proteome Res.* 17, 2963–2977.
- Vcelar, S., Melcher, M., Auer, N., Hrdina, A., Puklowski, A., Leisch, F., Jadhav, V., Wenger, T., Baumann, M., and Borth, N. (2018). Changes in chromosome counts and patterns in CHO cell lines upon generation of recombinant cell lines and subcloning. *Biotechnol. J.* 13, e1700495.
- Vishwanathan, N., Le, H., Le, T., and Hu, W.S. (2014). Advancing biopharmaceutical process science through transcriptome analysis. *Curr. Opin. Biotechnol.* 30, 113–119.
- Wlaschin, K.F., Seth, G., and Hu, W.S. (2006). Toward genomic cell culture engineering. *Cytotechnology* 50, 121–140.
- Wurm, F.M. (2017). Cloning of CHO cells, productivity and genetic stability—a discussion. *Processes* 5, 20.
- Wurm, F.M., and Hacker, D. (2011). First CHO genome. *Nat. Biotechnol.* 29, 718–720.
- Xu, X., Nagarajan, H., Lewis, N.E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., Hammond, S., et al. (2011). The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat. Biotechnol.* 29, 735–741.
- Yamano, N., Kumamoto, T., Takahashi, M., Frank, J., Onitsuka, M., and Omasa, T. (2015). Stability difference of each chromosome in Chinese hamster ovary cell line. *BMC Proc.* 9, P1.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
H3K9Me3 Antibody	Abcam	Cat#8898
Anti-DIG POD	Roche	Cat#11207733910
Chemicals, peptides, and recombinant proteins		
FBS, New Zealand Origin	Cytiva	Cat#SH30406.02HI
Geneticin	Thermo	Cat#10131027
Salmon Sperm DNA	Invitrogen	Cat#AM9680
Cytiva HyClone™ PF-CHO™ LS	Cytiva	Cat#SH3035902
CD CHO Medium	Thermo	Cat#10743029
L-Glutamine	GIBCO	Cat#25030-81
HT supplement	GIBCO	Cat#11067030
L-Methionine sulfoximine	Millipore	Cat#GSS-1015-F
MEM alpha	Gibco	Cat#12571-063
20× SSC	Lonza	Cat#51205
12× Chinese hamster Probe	Metasystems	Cat# D-1526-060-DI
Buffer RLT	Qiagen	Cat#79216
Pepsin	Abcam	Cat#ab64201
RNAse A	Sigma	Cat#R6148
Hypotonic Solution	Genial Helix	Cat#GGS-JL-007
Accuprime Pfx	Invitrogen	Cat#12344032
Colcemid	Thermo	Cat#15212012
0.075M KCL	Thermo	Cat#10575090
Vectashield + DAPI	Vector Labs	Cat#H-1200
PBS	Gibco	Cat#10010-031
Critical commercial assays		
Gel Extraction Kit	Qiagen	Cat#28704
DNeasy Blood and Tissue Kit	Qiagen	Cat#69504
Super-Script IV VILO Kit	Thermo	Cat#11756050
Quanti-iT dsDNA Assay kit	Thermo	Cat#Q33120
Illumina TruSeq DNA PCR-Free Kit	Illumina	Cat#20015962
TSA Plus Kit	Perkin Elmer	Cat#NEL763001KT
RNAEasy Plus Micro	Qiagen	Cat#74034
TruSeq Stranded Total RNA Library Preparation Prep Gold Kit	Illumina	Cat#20020598
HiScribe™ T7 ARCA mRNA Kit	NEB	Cat#E2060S
GenElute Mammalian Genomic DNA Miniprep Kit	Sigma	Cat#G1N70
QuickExtract™ DNA Extraction Solution	Lucigen	Cat#QE09050
Megaclear Kit	Thermo	Cat#AM1908
BioAnalyzer DNA 7500 chip	Agilent	Cat#5067-1506
BioAnalyzer RNA 6000 Nano Chip	Agilent	Cat#5067-1511
Deposited data		
Processed CNV Data	This manuscript	Data S1
Processed RNASeq Data	This manuscript	Data S2

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
CriGri-PICR13	Ensemble	Database version 105.1
Experimental models: Cell lines		
CHO-K1	Dr. Lawrence Chasin	LC78
MK-1	This manuscript	N/A
MK-1 GS –/–	This manuscript	N/A
MK-2	This manuscript	N/A
MK-2 GS –/–	This manuscript	N/A
Oligonucleotides		
5'-GGGTGGCCCGTTTCATCT-3'	Sigma	GS Forward Primer
5'-CGTGACAACCTTCCCATATCACA-3'	Sigma	GS Reverse Primer
Recombinant DNA		
CompoZr® ZFN	Sigma	Cat#ZFNGS-1KT
Software and algorithms		
Prism	Graphpad	Version 8
ISIS	Metasystems	Version 5.5.10
R-Studio	RStudio Inc	Version 1.2.5033
CNVKit	TALEVICH, E. A.-O., SHAIN, A. H., BOTTON, T. & BASTIAN, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing	Version 2.0
HTSeq	PUTRI, G. H., ANDERS, S., PYL, P. T., PIMANDA, J. E. & ZANINI, F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. 2021	Version 0.6
GSEA	SUBRAMANIAN, A., TAMAYO, P., MOOTHA, V. K., MUKHERJEE, S., EBERT, B. L., GILLETTE, M. A., PAULOVICH, A., POMEROY, S. L., GOLUB, T. R., LANDER, E. S. & MESIROV, J. P. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. <i>Proc Natl Acad Sci U S A</i> , 102, 15,545–50	Version 4.2.2
GATK4	Broad Institute	Version 4
Haplomapper	Sentieon	Version 202,112.01
BWA	LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. <i>Bioinformatics</i> , 25, 1754–1760	Version 0.7.5
bcl2fastq'	Illumina	Version v1.8.4
MATLAB	Mathworks	R2019b
Samtools	LI, H., HANDSAKER, B., WYSOKER, A., FENNELLS, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & SUBGROUP, G. P. D. P. 2009. The Sequence Alignment/Map format and SAMtools. <i>Bioinformatics</i> , 25, 2078–2079	Version 1.2
Nsolver	Nanostring	Version 4.1
Seqprep	GitHub Link	Version 1.2
Other		
Hanabi PVI	ADS Biotech	Cat#HCMS-99-0001
Axiolmager Z2	Zeiss	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Neon Electroporator	Thermo	Cat#MPK5000
QX200 Droplet Digital PCR	BioRad	Cat#1864001
ViCell-XR	Beckman Coulter	Cat#731196
nCOUNTER FLEX	Nanostring	N/A
Nanodrop	Invitrogen	Cat#ND-2000
Bioanalyzer	Agilent	Cat#G2939BA
Rx Imola	Randox	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Steven C. Huhn (steven.huhn@merck.com).

Materials availability

Plasmids and cell lines generated in this study are restricted as they are under patent by Merck & Co., Inc.

Data and code availability

Processed whole-genome sequencing data and RNASeq data are available in the supplemental materials. Microscopy data reported in this paper will be shared by the lead contact upon request.

This study does not report original code

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Adherent CHO-K1 LC78 cells (a generous gift from Dr. Lawrence Chasin) were cultured in MEM α media supplemented with glutamine (Gibco, Cat: 12571-063) and 10% FBS in T-175 flasks. Once cells reached 90% confluency, they were trypsinized. Briefly, flasks were washed with 30 mL PBS, the PBS was decanted, and replaced with 10 mL 0.025% trypsin. Cells were incubated at 37°C for 3–5 min and were dislodged by gentle tapping. The trypsin was neutralized with an equivalent amount of complete media, transferred to a 50 mL conical tube and the cells were centrifuged at 200 \times g. Cells were resuspended by complete media, and an aliquot was counted via ViCell. Cells were then seeded at a density of 3 \times 10⁶ in flasks, and returned to the incubator at 37°C, 5% CO₂, 80% humidity.

MK-1, MK-1 GS $-/-$, MK-2, and MK-2 GS $-/-$ cells were grown in a Multitron shaker at 36.5°C, 5% CO₂, 80% humidity, and 140rpm in shakeflasks. MK-1 cells used PF-CHO LS while MK-2 cells used CDCHO. Cells were regularly seeded at 0.4 \times 10⁶ in shakeflasks and were subcultured every two days.

METHOD DETAILS

Cell adaptation

MEM α media was replaced with either CD-CHO medium (MK-2 lineages; Gibco, Cat: 10743029) or PF-CHO LS (MK-1 lineages; GE, Cat: SH30359.02) gradually in shake flasks until growth rate stabilized and aggregates were no longer present (see [Figure S1](#)). Cell aggregates were periodically removed via filtration every subsequent passage through a 35-micron strainer. The resultant serum and aggregate free pools were subsequently single cell cloned by FACS and deposited one cell per well into 96-well plates. Clones were scaled, banks were prepared, and clones were evaluated for mAb expression using Trastuzumab as a model. Briefly, 5 \times 10⁶ CHO cells were collected and mixed with 10 mg plasmid containing the sequence for Trastuzumab (<https://www.genome.jp/entry/D03257>). Cells were then processed for electroporation with the Neon transfection system (Thermo Fisher) using the manufacturer's CHO protocol (voltage: 1700, Width [ms]: 20, Pulse Number: 1. Following electroporation, transfected cells were placed in a static

incubator at 37°C for 2 days. Cells were then transferred to shake-flasks containing antibiotic and glutamine free media containing 12.5 μM MSX and G418 (400 μg/mL) to generate stable pools. Recombinant clones were then transferred to production media, and 14-day fed batch production (see “Recombinant Protein Production assay”). The recombinant clones exhibiting the highest mAb production were then identified as MK-1 and MK-2.

Cells from these banks were then thawed and processed for GS knockout in order to create GS-null cell lines as described below (Huhn et al., 2019, 2021).

Glutamine synthetase knockout and screening

CompoZr® ZFN mRNAs were prepared from two plasmids (ZFNGSA9075 and ZFNGSB9372, Sigma, ZFNGS) expressing a pair of ZFNs targeting CHO GS. The two plasmids were first linearized, followed by purification and *In Vitro* transcription using HiScribe™ T7 ARCA mRNA Kit (NEB). The two paired-ZFN mRNAs were purified using MegaClear Kit (Ambion), combined, and used for transfection. 5×10^6 CHO cells were collected and mixed with 5 μg ZFN mRNA mixture and processed for electroporation with the Neon transfection system (Thermo Fisher) using the manufacturer’s CHO protocol (voltage: 1700, Width [ms]: 20, Pulse Number: 1). Following electroporation, transfected cells were cold-shocked in a static incubator at 30°C for 2 days, and then placed at 37°C for 2 days. Cells were then cloned into 96-well plates through limiting dilution at 0.8 cell/well. Cells were then monitored by imaging (Cellviva, SynGene) until 40%–50% confluency was scored by Cellviva software (~10–14 days). Cells were then expanded in a continuous and stepwise manner based on growth (40–50% confluency) from plates to spin tubes, for 20 days. After 2 weeks, each well with colony was screened for GS gene disruption (knockout) through Sanger sequencing. For Sanger sequencing, genomic DNA from each clone was extracted by 50 μL of Quick Extract solution (Epicenter), followed by heating at 65°C for 15 min and then 95°C for 5 min. PCR reactions in this study used AccuPrime Pfx DNA Polymerase (Invitrogen) and an ABI Veriti thermocycler. The reactions proceeded identically to the manufacturer’s recommendations, except that 100 μL of reaction volume was used per reaction with 100 ng of input gDNA, using an annealing temperature of 68°C. Each reaction did not exceed 30 cycles. The GS gene fragment from each clone was amplified from its genomic DNA through PCR using forward primer (5′-GGGTGGCCCGTTTCATCT-3′) and reverse primer (5′-CGTGACAACCTTCCCATATCACA-3′). The PCR products were sent for PCR cleanup followed by Sanger sequencing using the reverse primer.

Whole genome sequencing and analyses

The genomic DNA was extracted using GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich, Catalog number G1N70) using manufacturer’s protocol. The extracted genomic DNA samples underwent quality check (QC), using Invitrogen Quanti-iT dsDNA assay and gel electrophoresis to determine DNA concentration and DNA quality, respectively. The samples passing QC were then used to generate libraries using the Illumina TruSeq DNA PCR-Free kit for DNA sample. The concentration and size range of the generated libraries were determined using the Quanti-iT dsDNA Assay kit and the Agilent 2100 BioAnalyzer DNA 7500 chip, respectively.

The libraries were sequenced using the Illumina HiSeq platforms, with read length of 2×150 bp for DNA library. 1% or 5% Phix control was spiked-in the library prior to sequencing. 100Gb of sequencing data was generated per DNA sample for an estimated ~40× depth of coverage.

Prior to data analysis, samples were demultiplexed using bcl2fastq-v.1.8.4, and adapter sequences were trimmed using Seqprep. For WGS analysis, paired end Illumina reads were aligned to CriGri-PICR13 using BWA-v.0.7.5 (Li and Durbin, 2009), Samtools-v.1.2 (Li et al., 2009), and duplicate removal with Picard. Subsequently, CNVkit was used for copy number variation analysis at gene level with threshold (-t) option set as 0 to maximize the number of detected genes and minimum probes (-m) option set as 3 returning genes with at least 3 probes. The log₂ ratio of the copy number versus CHOK1 progenitor cells was calculated with CNVkit (Talevich et al., 2016), with value greater than -1 but less than 0.7 set to be no change comparing to the CHO-K1 LC78 progenitor cells. Sentieon Haplotyper (<https://support.sentieon.com/manual/>) was used for SNP and INDEL variant detection, GATK4 (<https://gatk.broadinstitute.org/hc/en-us/articles/360036194592-Getting-started-with-GATK4>) was used for SNP and INDEL selection and low-quality SNP and INDEL filtering. Variant decomposition and normalization were done with vt (<https://genome.sph.umich.edu>) and variant annotation with vep99 (<https://useast.ensembl.org>).

RNA sequencing and analysis

The total cellular RNA of CHO cells was extracted using RNAeasy plus micro kit (Qiagen, 74034) using manufacturer's protocol. The extracted total RNA samples underwent quality check (QC) using NanoDrop ND-2000 spectrophotometer and the Agilent 2100 BioAnalyzer RNA 6000 Nano Chip to determine RNA concentration and RNA integrity, respectively. The samples passing QC were then used to generate libraries using the Illumina TruSeq Stranded Total RNA Library Preparation Prep Gold kit. The concentration and size range of the generated libraries were determined using the Quanti-iT dsDNA Assay kit and the Agilent 2100 BioAnalyzer, respectively. The libraries were sequenced using the Illumina HiSeq platforms, with read length of 2×100 bp for the RNA library. 1% or 5% Phix control was spiked-in the library prior to sequencing and 2×100 million reads were generated. Prior to data analysis, samples were demultiplexed using `bcl2fastq-v.1.8.4`, and adapter sequences were trimmed using `trim_galore_v0.3.3`. For RNAseq bioinformatics analysis, samples were aligned to CriGri-PICR13 using STAR_2_5 as pair. The alignment files were sorted by read names using `samtools-0.1.19`. Using sorted reads, raw counts were generated using HTSeq (Putri et al., 2021) version 0.6.0) and subsequent differential expression.

Pathways analysis and statistical analysis

Gene Set Enrichment analysis was performed with GSEA software and the Molecular Signature Database (mSigDB) (<http://www.broad.mit.edu/gsea/>), using custom gene sets for biological pathways of interest (Subramanian et al., 2005). The gene ontology (GO) IDs were mapped to the genes in the OMICS data; and the numbers of genes in each GO term were calculated for each OMICS data type. Heatmaps were generated using R packages, including `ggplot2` and `pheatmap`. Prism (Graphpad, CA) and JMP (SAS, NC) were utilized to calculate statistics.

Cytogenetics

Suspension CHO cells were treated with 0.2 ug/mL Colcemid for 4–5 h at 37C. Cells were then adjusted to $1 \times$ with Pre-Hypotonic Swelling Solution (Genial Helix, Cat: GGS-JL-007) and pelleted at 200xg for 10 min. All but 0.5 mL of the remaining culture media was aspirated, and cells were resuspended in the residual media. The remaining cell volume was adjusted dropwise with 1 mL of 0.075M KCL and then adjusted to 13 mL 0.075M KCL with a serological pipette. Cultures were then incubated 20 min, pelleted as above, resuspended in 0.5 mL residual KCL, and fixed with dropwise addition of 3:1 methanol acetic acid. The cultures were then washed three times in of 3:1 methanol acetic acid and aliquots of cells were dropped onto slides using a Hanabi-PVI cell-spreader. Slides were baked overnight at 50C before any analysis. For karyotyping, only cells with non-overlapping chromosomes were considered for quantitation.

FISH

Cytogenetic spreads were transferred to prewarmed PBS for 5–10 minutes and then subsequently transferred to 0.2 M HCL at 37C for 30 minutes. Following several washes, the spreads were digested with RNase A (Sigma, Cat: R6148), Pepsin (Abcam, Cat: ab64201), and then fixed in 1% paraformaldehyde for 10 minutes. The slides were then serially dehydrated in 80,90 & 100% ethanol and allowed to air dry.

Slides were then denatured in waterbath at 72C for 30 minutes in $2 \times$ SSC (Lonza, Cat: 51205), allowed to cool to room temperature for 20 minutes, and immediately transferred to coplin jars containing: $0.1 \times$ SSC, 0.07 N NaOH, $0.1 \times$ SSC at 4C and $2 \times$ SSC at 4C for 1 min each. The slides were then serially dehydrated as above, warmed to 37C and denatured probe containing 1–2.5 ng/uL probe DNA was applied to the slide. The slide was covered with a coverslip, sealed with rubber cement, and incubated at 37C for at least 18 h. Following hybridization, the slides were submerged in $2 \times$ SSC +0.1% tween, the rubber cement was removed, and coverslips were floated off. Slides were washed at 72C in $0.4 \times$ SSC for 2 minutes, and then washed in $2 \times$ SSC +0.1% tween for 5 minutes. Slides were then equilibrated in PBS, blocked in 0.5% Blocking Reagent (Perkin Elmer, Cat: FP1012), and incubated with anti-DIG POD (Roche, Cat: 11207733910 at 1:500) overnight. Slides were then developed using the TSA Plus Cyanine 3.5 (Perkin Elmer, Cat: NEL763001KT) exactly to manufacturer's specifications. Slides were mounted with Vectashield + DAPI (Vectashield, Cat: H-1200) before viewing on a Motorized Carl Zeiss Microscope Axiolmager Z2 using ISIS software (Metasystems, Inc).

Probe generation

1–5 μg of plasmid was digested using the Roche DIG Nick translation kit for 2–5 h, until an approximate probe size of ~ 500 bp was visualized on a 1.5% agarose gel. The probe was purified using the QIAquick Gel Extraction Kit (Qiagen, Cat: 28704 and 28706) and eluted in purified water. Prior to use, the probe was combined with 12.5 μg of salmon sperm DNA (Invitrogen, Cat: AM9680), adjusted to 0.5M NaCl and 70% ice-cold ethanol, and incubated at -20°C for 2 h. The probe was precipitated by centrifugation at 16,000 $\times g$ for 1 h at 4 $^\circ\text{C}$, washed twice in 70% ethanol, and dissolved by incubating in 5–10 μl of 100% formamide using a thermomixer.

Chromosome painting

Slides were prepared identically to those utilized in generic FISH, but probe was substituted for 12 \times Chinese Hamster Probe (Metasystems, Cat: D-1526-060-DI). Chromosome paints were visualized using ISIS software (Metasystems, Inc).

For quantitation, chromosomes were converted to 8-bit form prior to analysis and exported as TIFF files. A custom MATLAB script (version R2019b) was written to quantify the relative abundance of different pixel colors in each image, which corresponded to each chromosome. The amount of pixels from each chromosome was calculated as fraction of the entire genome (all chromosome areas). We normalized this number as a Log₂ fold change versus CHO-K1 LC78 vs. the corresponding CHO-K1 LC78 chromosome. Quantification reflects the capture of 20–30 metaphases across >2 experiments.

To calculate chromosome recombination frequency, recipient chromosomes were defined as a contiguous entity containing $>50\%$ of the original Wild-Type Chinese Hamster chromosome paint, and donors as those containing $<50\%$. Contiguous chromosomes were scored for the intersection of such donor and recipient sequences and heatmaps reflect the percentage of individual metaphases containing such translocations. 20–30 metaphases across >2 experiments stained with 12 \times Chinese Hamster chromosome paints were analyzed for recombination.

Gene expression analysis

5×10^6 cells were pelleted and lysed with RLT Buffer (Qiagen, Cat: 79216). Lysates were sonicated and normalized to 2800 cells/mL. Approximately 4.5 μl of cell lysate was used per custom probe set designed by Nanostrings (Nanostrings Inc, WA). The plexset-12 protocol was then followed according to the manufacturer's instructions. Cartridges were scanned using maximum sensitivity, and, for data analysis, the Nsolver (Nanostrings Inc, WA) software was used, with background thresholding set to a value of 20.

Recombinant protein production assay

Cells were seeded in in-house production media and glucose & lactate levels were measured daily using the RANDOX RX imola chemistry analyzer (Crumlin, UK). The measurement of growth, viability, osmolarity, glucose, and lactate were collected daily. Cell density and viability were measured a Vi-CELL cell counter (Beckman Coulter). Glucose and lactate levels were measured by using the RANDOX RX imola chemistry analyzer (Crumlin) or D-Fructose/D-Glucose Assay Kit (Megazyme). mAb production levels were determined by Protein-A HPLC. High molecular weight (HMW) and N-linked glycosylation were measured through SEC and HILIC assays, respectively (Huhn et al., 2021).

Genomic DNA copy number and gene expression quantification assays

Genomic DNAs were extracted from cells using DNeasy Blood and Tissue Kit (QIAGEN, 69,504). Total RNAs from CHO cells were extracted using the RNeasy Plus Mini Kit from QIAGEN (Qiagen, Catalog # 74034). cDNAs were prepared from the RNA samples by reverse transcription using the Super-Script IV VILO Master Mix (Thermo Fisher Scientific, Cat: 11756050). The QX200 Droplet Digital PCR (ddPCR) System (Bio-Rad, Hercules, CA) was used to determine the copy number of the recombinant DNA. Fluorescently-labeled oligo nucleotide probes for the ddPCR reactions were designed using the Primer Express Software (Applied Biosystems, Thermo Fisher Scientific, MA), and synthesized by Invitrogen (Thermo Fisher Scientific, MA). The results were normalized to B2M or HPRT. For gene expression studies, real-time PCR (qPCR) was performed using CFX Opus 96 Real-Time PCR Instrument (Bio-Rad Laboratories, Inc., PA) according to the manufacturer's specifications. mRNA amounts were normalized relative to GAPDH mRNA.

QUANTIFICATION AND STATISTICAL ANALYSIS

Normality of the data was evaluated and statistical tests was performed using Student's *t* test with unequal variance ($p < 0.05$) in Prism 8.4.3 (GraphPad Software). Data representation, can be viewed in the figure legends. Correlation was determined using heatmaps (Spearman) in Prism 8.4.3 (GraphPad Software) (significance level, $p < 0.05$). The statistics used in the bioinformatics analysis is described in the individual method sections above.