Received: 20 August 2023 | Revised: 25 October 2023 | Accepted: 12 November 2023

DOI: 10.1002/biot.202300425

Biotechnology Journal

RESEARCH ARTICLE

Prediction of CHO cell line stability using expression of DNA repair genes

Lauren T. Cordova[1] | **Hussain Dahodwala**[1,2] | **Rebecca Cooley**[3] | **Kelvin H. Lee**[1,2]

[1]Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, Delaware, USA

[2]National Institute for Innovation in Manufacturing Biopharmaceuticals, Newark, Delaware, USA

[3]Pfizer, Inc, 875 Chesterfield Pkwy W, Chesterfield, Missouri, USA

Correspondence
Kelvin H. Lee, Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE 19713, USA.
Email: KHL@udel.edu

Funding information
National Science Foundation, Grant/Award Number: 1624698; National Institute of Standards and Technology, Grant/Award Number: 70NANB17H002

Abstract

Chinese hamster ovary (CHO) cells are essential to biopharmaceutical manufacturing and production instability, the loss of productivity over time, is a long-standing challenge in the industry. Accurate prediction of cell line stability could enable efficient screening to identify clones suitable for manufacturing saving significant time and costs. DNA repair genes may offer biomarkers to address this need. In this study, over 40 cell lines representing various host lineages from three companies/organizations were evaluated for expression of five DNA repair genes (*Fam35a*, *Lig4*, *Palb2*, *Pari*, and *Xrcc6*). Expression measured in cells with less than 30 population doubling levels (PDLs) was correlated to stability profiles at 60+ PDL. Principal component analysis identified markers which separate stable and unstable CHO-DG44 cell lines. Notably, two genes, *Lig4* and *Xrcc6*, showed higher expression in unstable CHO-DG44 cell lines with copy number loss identified as the mechanism of production instability. Expression levels across all cell ages showed lower DNA repair gene expression was associated with increased cell age. Collectively, DNA repair genes provide critical insight into long-term behavior of CHO cells and their expression levels have potential to predict cell line stability in certain cases.

KEYWORDS
biomarkers, biopharmaceuticals, bioprocess development, bioprocess engineering, CHO cells, CHO production instability, Lig4, Xrcc6

1 | INTRODUCTION

Chinese hamster ovary (CHO) cells are the primary cell type used for manufacturing of therapeutic proteins, especially monoclonal antibodies (mAbs). CHO cells are used for manufacturing of 70% of all recombinant proteins and all mAbs approved in the past 8 years.[1] These biopharmaceuticals generate more than $100 billion in revenue per year with at least four mAb products reaching blockbuster status within the first year of approval as of 2016.[2,3] CHO cells possess a variety of desirable features for biopharmaceutical manufacturing such as robust growth in suspension culture, the ability to generate human-compatible glycosylation patterns, and multi-gram titer expression. The use of CHO cells can also be attributed to inherent genome plasticity which enables the selection of cells with desirable phenotypes. However, this same genome plasticity, can also result in chromosomal rearrangements and unstable mAb production during long-term culture durations once desirable phenotypes are achieved.[1] One particular example of this issue is in the selection of clones expressing high titers. Genome plasticity facilitates screening of clone pools to identify cells with high titers. However, the stability of the titer must be evaluated over 60 generations.[3,4] Clonal development can be accelerated with new methods for predicting cell line stability.

Previous work has identified genetic and epigenetic mechanisms including copy number loss, gene silencing, and histone modification as contributors to cell line instability.[4,5] Multiple of these

Abbreviations: AMBIC, advanced mammalian biomanufacturing innovation center; CHO, Chinese hamster ovary; HR, homologous recombination; IVCC, integrated viable cell count; mAb, monoclonal antibody; MTX, methotrexate; NHEJ, non-homologous end joining; PCA, principal component analysis; PDL, population doubling level; qP, cell-specific productivity; VCD, viable cell density.

mechanisms may act together to cause cell line instability which can also be impacted by phenotypes inherent to the host cell line. Additionally, the integration location of transgene cassettes can influence expression as some locations are considered "hot spots" capable of high expression. Despite these diverse mechanisms for cell line instability, early prediction of cell line instability is of great interest for the biopharmaceutical process development community. In the search for these predictive markers, multiple studies have sought mechanistic or broad expression-based screens but were limited to a single CHO host cell lineage.[4,6,7] Since the first isolation of CHO cells in 1958, adaptation, mutagenesis, and genetic engineering have been widely applied resulting in vast genomic differences attributed to each lineage.[1,8–10] CHO-K1, CHO-DG44, CHO-DUXB11, and CHO-GS represent the most widely used CHO cell lines in biopharmaceutical production.[1] Although these cell lines are all CHO-derived, their behavior with respect to production capacity and cell line stability can vary across host lineages, due to the use of mutagens in their creation and the methods employed in final selection. Therefore, identification of universal, early biomarkers of cell line instability across multiple CHO cell hosts would enable significant time and resource savings during biopharmaceutical development.

Genome instability, which can serve as one mechanism of production instability across CHO cell lines, occurs primarily through chromosomal rearrangement.[11] This instability is inherent to the CHO genome as compared to the ancestral Chinese hamster genome and may be due to improper DNA double strand break repair.[1,3] The two most widely studied pathways for DNA repair are homologous recombination (HR) and non-homologous end joining (NHEJ) with recent work identifying microhomology-mediated end-joining as a third alternative mechanism.[12,13] Detailed information about DNA repair pathways and behavior specifically within CHO cells is extensively reviewed elsewhere.[10,14–16] Five key genes studied in this work, which show mutations between the CHO and Chinese hamster genome, *Fam35a*, *Lig4*, *Palb2*, *Pari*, and *Xrcc6*, span activity within both HR and NHEJ pathways. *Fam35a* acts as part of the shieldin complex to initiate NHEJ DNA repair while *Lig4* is responsible for the final ligation step to repair DNA breaks.[17–19] The *Xrcc6* gene which encodes the KU70 protein is a member of the x-ray repair cross-complementing group responsible for NHEJ.[20] Conversely, *Palb2* and *Pari* are directly involved in HR-directed repair mechanisms with *Pari* expression as a rate-limiting factor in CHO cells.[18] Collectively, these five genes and others involved in DNA double strand break repair have been implicated in human cancers.[20–22] and may also provide critical insight into CHO cell line instability.

In this work, expression of five DNA repair genes (*Fam35a*, *Lig4*, *Palb2*, *Pari*, and *Xrcc6*) at early cell ages was measured and associated with relative productivity at PDL60. Through a partnership with the Advanced Mammalian Biomanufacturing Innovation Center (AMBIC), analysis of industrially-relevant cell lines ultimately identified *Lig4* and *Xrcc6* as potential biomarkers for stability prediction specifically within the CHO-DG44 host lineage. Further study of these cell lines confirmed stability trends and identified copy number loss as the main cause of production instability in cell lines aged without selection pressure. This work provides insight into DNA repair gene expression during cell aging and presents preliminary biomarkers for early prediction of CHO cell production instability.

2 | MATERIALS AND METHODS

2.1 | Cell lines and culture information

The CHO-DG44 cell lines used in this work were generously provided by the National Institutes of Health as part of the AMBIC consortia. Three cell lines, designated here as cell lines A, B, and C, were isolated from a single cell cloning campaign and express the same monoclonal antibody construct. These cell lines were grown in ActiPro media (Cytiva) with 6 mM glutamine using 125 mL shake flasks, an operating volume of 20–30 mL, and a shake speed of 135 rpm (25 mm incubator throw diameter). Culture conditions were 37°C with 5% CO_2 and 80% relative humidity using a seeding density of 0.3–0.4 million cells per mL. To generate various cell line ages, doubling times were calculated at each passage and cultures were maintained until the cumulative population doubling level (PDL) reached the desired age. Cell count and viability were measured every 3–4 days using a Vi-Cell XR (Beckman) prior to passage and PDL at each passage was calculated using initial and final viable cell counts.[23,24] Frozen banks were created at approximately PDL 20, 35, 60, and 90. Cultures were aged in the presence and absence of 30 μM methotrexate (MTX) selection to generate a total of six different cell line populations for detailed analysis. Additional cell lines were analyzed off-site for DNA repair gene expression by industry partners and included CHO-GS and CHO-K1 lineages. Cultivation methods for these cell lines including growth media, temperature and shaking speed varied, and were not disclosed by industry partners. The CHO-K1 cell vial for use as an RNA internal standard was generated using the same culture conditions outlined above for growth in ActiPro media. Cell banks were created using cells from passage three and frozen in 10% DMSO.

2.2 | Fed-batch culture conditions

For accurate comparison of production instability, all DG44-based cell lines were grown in fed-batch culture using the same basal media as outlined above with the CellBoost 7a/7b system (Cytiva). Feeds were added daily starting on day 3 and maintained at 3% vol 7a (0.3% vol 7b) until culture termination. Concentrated glucose solution (200 g L^{-1}) was also added daily after day 5 at 33 μL to supplement the glucose provided in the 7a feed solution to reach approximately 5 g L^{-1} total glucose at each feeding To assess all ages and required conditions within a single experiment, fed-batch culture was performed in bioreactor spin tubes with an operating volume of 10 mL and a shake speed of 200 rpm (incubator throw diameter 25 mm). Daily sampling volume was matched with daily feed/glucose addition to minimize volume loss. Cell count and viability were analyzed daily for these experiments with cultures terminated when viability fell below 70%.

Titer from fed-batch cultures was quantified using an Octet Red 96e with Protein A biosensors. Samples were diluted 10-fold with

phosphate buffered saline containing 0.1% bovine serum albumin and quantified using Octet Calibrators (1–700 μg mL^{-1}, Sartorius). MAb titer was assessed every other day and at culture termination. Overall cell specific productivity was determined as the slope of titer plotted as a function of integrated viable cell count (IVCC). Daily IVCC was calculated using the trapezoid approximation of daily viable cell density measurements.

2.3 | Quantification of DNA repair gene and mAb expression

Gene expression was quantified from bulk RNA extracted using the RNeasy Mini Kit (Qiagen). For CHO-DG44 samples, an aliquot of 5 million cells was used for RNA extraction at fed-batch culture onset. For all other cell lines, RNA extraction and qPCR analysis was performed off-site by each industry partner following the protocol provided in Document S1. RNA extraction was followed by DNase treatment with a DNA-free kit (Invitrogen) and clean up using RNeasy MinElute Cleanup Kit (Qiagen). RNA concentration was measured and diluted to 20 ng μL^{-1} with 40 ng loaded per sample. Reverse transcription and quantification were completed using the TaqMan RNA-to-Ct 1-Step kit (Applied Biosystems) according to the manufacturer's protocol. PrimeTime PCR assays (Integrated DNA Technologies) used the FAM fluorophore. Primer sequences for all DNA repair genes, *Actr5* as a housekeeping gene and mAb heavy/light chain can be found in Table S1.

Relative expression was quantified using the ΔΔCt method comparing to a standard CHO-K1 RNA isolated from a frozen cell bank. This reference material was provided to participating sites to enable normalization and comparison across cell lines and locations. The CHO-K1 RNA sample was directly isolated from a frozen vial with three washes using 5 mL sterile Phosphate Buffered Saline prior to RNA extraction to completely remove residual cryoprotectant. Raw qPCR expression data was analyzed using templated spreadsheets, collected, and aggregated for broader analysis. qPCR expression data was collected in technical triplicate with isolated, divergent values removed from the average Ct calculation across the dataset. For mAb expression of CHO-DG44 clones within the aging study, light and heavy chain expression was determined using the ΔΔCt method comparing to PDL20 conditions. Data was further normalized to the mAb light chain in cell line A to allow for easier comparison across cell lines.

2.4 | MAb copy number quantification

Copy number of the light and heavy chain of the IgG antibody expressed by the CHO-DG44 cell lines was quantified using TaqMan Applied Master Mix (Applied Biosystems) with genomic DNA. At initiation of fed-batch culture, a sample containing 5 million cells was collected from each cell line/age. Genomic DNA was isolated using a QIAmp DNA Mini Kit (Qiagen) with RNase and Proteinase K added according to manufacturer's instructions. DNA solutions were diluted to 15 ng μL^{-1} with 75 ng loaded per sample well. PrimeTime PCR assays

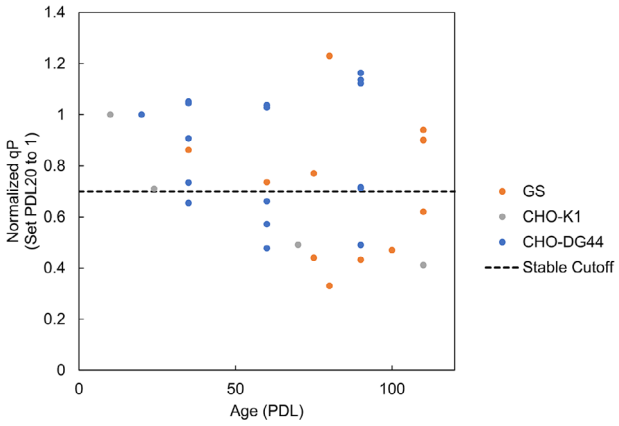

FIGURE 1 Cell age and relative productivity of cell culture samples considered in this study. The full dataset included 44 total samples across multiple ages and CHO cell lineages. Relative productivity of 70% is shown as the stability cut-off with a dashed line.

(Integrated DNA Technologies) which contain preformulated primer and probe sequences were used with quantification determined by comparison to a gBlock (Integrated DNA Technology) standard curve of known copy number. To minimize material usage, the housekeeping gene *Actr5* used a HEX fluorophore to enable multiplexing with the FAM fluorophore on the heavy and light chain mAb probe sets (see Table S1 for primer/probe sequences). Amplification followed manufacturer's instructions for TaqMan Applied Master Mix with a CFX96 Real-Time PCR Detection System (BioRad). Relative copy number was generated by normalizing copy numbers to those at PDL20 for each cell line/condition.

2.5 | Dataset from multiple industrial CHO hosts

Information about specific stability of cell lines used in this study was aggregated by various companies and organizations (Figure 1). These cell lines span a variety of ages covering relevant timeframes from vial thaw to production bioreactor (PDL60). For some cell lines, information was available at longer cell ages up to PDL110. For comparison across different cell line platforms, cell specific productivity (qP) was normalized to the youngest cell age available (PDL0-20). Stable cell lines are defined as maintaining a cell specific productivity greater than 70% of the initial value (i.e., greater than 0.7 normalized qP) (Figure 1).[25] This full dataset represented a total of 44 cell line samples with 53% of the dataset designated as stable and 58%, 32%, and 11% of the dataset representing the CHO-GS, DG-44, and K1 host lineages, respectively. Within each CHO host lineage, there is an equal distribution of stable and unstable cell lines.

2.6 | Statistical methods and analysis

All data for fed-batch culture represents the mean and standard deviation from two biological replicates. All qPCR analysis was performed

in biological duplicate and technical triplicate with the average Ct value and corresponding standard deviation used in calculations. All statistical analysis was performed in R (4.2.0) using RStudio version 2022.02.1. Principle Component Analysis (PCA) used the factoextra package and box plots were generated using the ggplot2 package.

3 | RESULTS

3.1 | DNA repair gene expression in early age relevant cell lines

The CHO genome has been characterized by reduced copy number or mutations in key DNA repair pathways.[3] To interrogate the relationship between these DNA repair genes and CHO cell line stability, each cell line was evaluated for expression of five DNA repair genes (*Fam35a*, *Lig4*, *Palb2*, *Pari*, and *Xrcc6*). These genes were selected for analysis because the CHO genome contains at least one nucleotide mutation within the gene which changes an amino acid in the protein sequence as compared to the Chinese hamster genome which might impact function. Additionally, *Lig4* and *Xrcc6* have been previously explored within the context of DNA repair and production instability in CHO.[3] Only cell line samples with PDL between 0 and 30 were considered for this portion of the study. A total of 19 cell lines (dataset shown in SI Table 2) were evaluated for DNA repair gene expression with 10 cell lines classified as stable (relative qP at PDL60 greater than 70%).[25] Principal Component Analysis (PCA) applied to the full dataset with cell line stability as the separation metric showed overlapping PCA 95% confidence interval ellipses suggesting that there is no linear combination of the five DNA repair gene expression that sufficiently explains population differences (i.e., serve as an early predictor of cell line instability) (Figure 2A).

Because most samples were from either the CHO-DG44 or CHO-GS hosts, we next sought to use host lineage host as a grouping metric for PCA analysis.[26] (Figure 2B). The 95% confidence ellipses show distinct separation of the two host cell line groups indicating characteristic DNA repair gene expression within each CHO host cell line, with all five DNA repair genes contributing to the loading of the first two PCA dimensions (Table S3). These findings are consistent with the notion that each host cell lineage has a distinctive expression profile of these DNA repair genes, as might be expected due to the distinct developmental stages required to generate each lineage.

The early stability dataset was split into subsets for the two most abundant lineages (CHO-GS, CHO-DG44). PCA of the CHO-GS host showed nearly complete overlap of the data (Figure 2C) while clear separation was observed between the stable and unstable CHO-DG44 populations (Figure 2D). From this dataset of CHO-GS cell line samples, long term cell line stability is likely not attributable to expression of these five DNA repair genes in any linear combination. For this lineage, cell line stability may be attributable to expression of different genes or other phenotypic signatures. For the CHO-DG44 samples, PCA dimension 2 completely separates stable and unstable populations with 12.8% and 73.1% loading from *Lig4* and *Xrcc6*, respectively (Table S3).

Low expression of both *Lig4* and *Xrcc6* in CHO-DG44 hosts was observed in stable cell populations while elevated expression was seen in unstable populations (Figure 3). DNA repair gene expression of *Lig4* and *Xrcc6* in the CHO-GS/K1 host cell lines along with expression of *Fam35a*, *Palb2*, and *Pari* for all cell lines show between a 0.3 and 0.5-fold change in expression relative to the CHO-K1 standard sample (Figures S1–S4). *Fam35a*, *Palb2*, and *Pari* (Figures S2–S4) show similar expression levels ($p > 0.05$) between the stable and unstable populations in both CHO-DG44 and other CHO host cell lines.

To further probe the ability of DNA repair genes to predict cell line stability a priori, linear regression was applied to predict relative qP from relative RNA expression of *Lig4* and *Xrcc6*. The regression fit (Relative qP = 1.45–0.35* Relative *Lig4* RNA expression – 0.4*Relative *Xrcc6* expression) was statistically significant ($p < 0.05$) with these two variables explaining 94% of the variation in the data as shown in Figure S5 ($R^2 = 0.94$, adjusted $R^2 = 0.91$). This relationship suggests that *Lig4* and *Xrcc6* have potential as marker genes to screen CHO-DG44 based cell lines for long-term production stability.

3.2 | Aging study of CHO-DG44 clones

A study using the CHO-DG44 cell lines was performed to determine the mechanisms of production stability in these cell lines. Cells were aged to PDL 20, 35, 60, and 90 in the presence or absence of 30 μM MTX, banked and thawed for simultaneous fed-batch screening of all conditions. Addition of MTX provides selection pressure to maintain the mAb construct and dihyrofolate reductase (DHFR) copy number. Without the DHFR activity, the presence of MTX prevents synthesis of essential nucleotide precursors. While addition of MTX can maintain construct copy number, use of this compound in industrial biomanufacturing is not desirable. The three cell lines evaluated here, in both the presence or absence of MTX, showed consistent growth behavior reaching a peak viable cell density (VCD) between 15 and 25 million cells mL^{-1} between day 6 and 8 of culture (Figure S6). Culture VCD and viability dropped rapidly at the end of culture with cell lines aged in MTX surviving longer than their non-selected counterparts (Figure S7). PDL90 cultures had the shortest fed-batch culture duration compared to younger cell age cultures often terminating at least a day earlier, regardless of the cell line/selection conditions (Figure S7).

In addition to monitoring cell growth and viability, samples were periodically collected for mAb titer quantitation. All cell lines at PDL20 produced a maximum titer between 1.2 and 1.6 g L^{-1} (Figure S8). Time course samples during growth show significant mAb accumulation during the fed-batch portion (day 3 onward) in all cultures. For the conditions lacking MTX selection, lower titer is observed as cells age (Figure S8). For cell lines A and C, titer fell with increasing age while cell line B showed consistent titer at about 1 g L^{-1} for PDL20 and 35 with a 40% decrease in PDL60 and 90 conditions (Figure S8). Without the MTX selection pressure, production instability impacts each of these

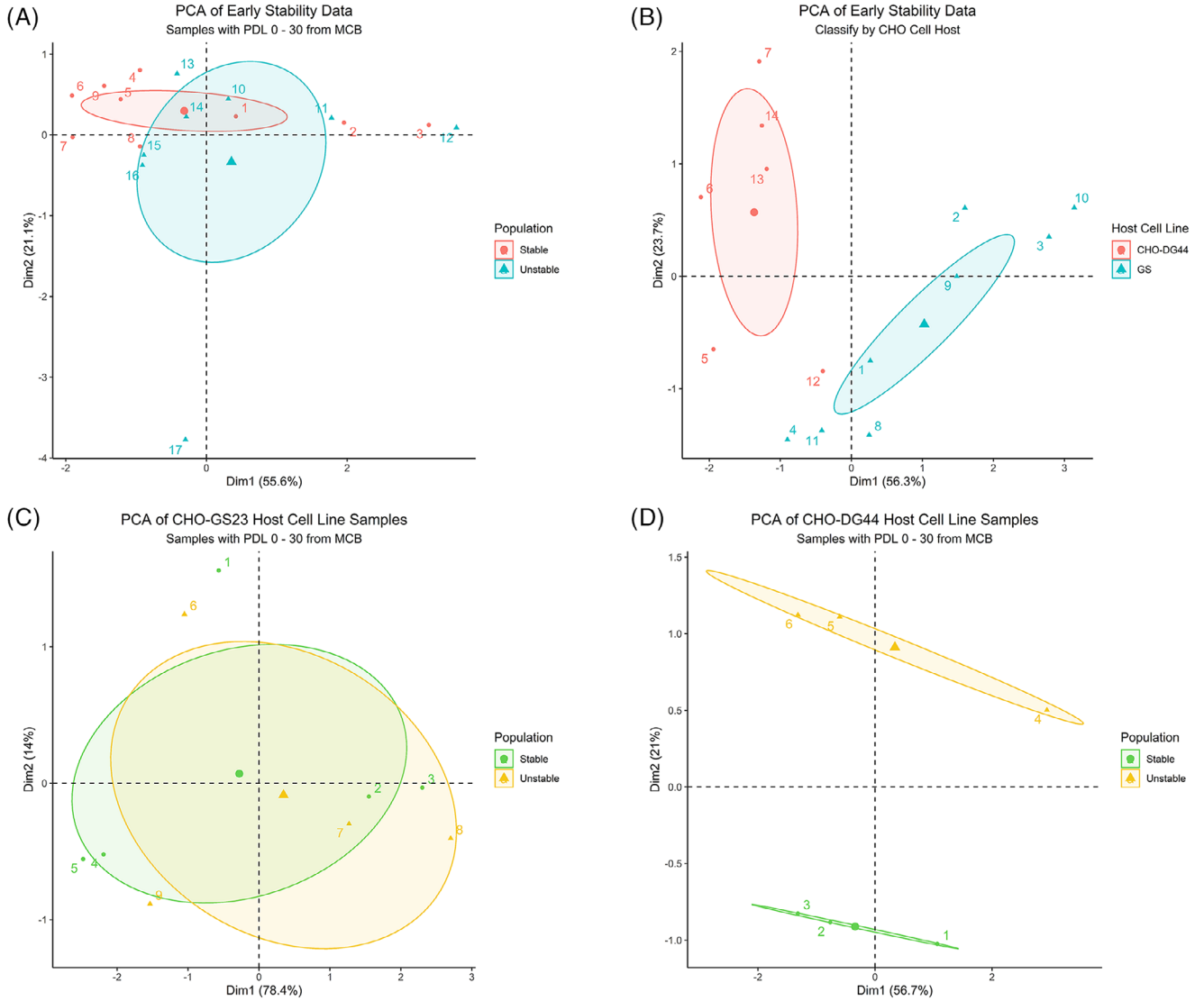

FIGURE 2 Principal Component Analysis of DNA repair genes toward predicting production instability. Analysis of data from young cell lines (≤PDL30) to find characteristic groupings of variables shows overlap in all data (A) and separation when considering host cell line (B). Each host cell line in this work has a characteristic set of DNA repair gene expression levels. Analysis of each subgroup of host cell lines CHO-GS (C) and CHO-DG44 (D) indicate that characteristic/predictable behavior between stable and unstable cell lines is observed in only the DG44 subset.

cell lines. For cell lines maintained with MTX selection during the aging process, titer was nearly identical across all cell ages.

From the VCD and titer data, cell-specific productivity (qP) was calculated for each condition and normalized to the corresponding PDL20 condition for comparison across cell lines (Figure 4A). Consistent with the titer results, cultures without MTX selection show a drop in normalized qP with cell age. All three cell lines demonstrate behavior consistent with production instability; cell line B reached a normalized productivity of 66% at PDL60 which is maintained at PDL90 while cell line C had a normalized productivity of 50% by PDL35. As expected, MTX selection maintained a high normalized qP for all cell ages with a 12%–16% increase from PDL20 to PDL90 (Figure 4A).

DNA copy number and RNA expression of the mAb cassette were quantified, as loss of either can indicate production instability has occurred within CHO cells. For the cell lines aged in the absence of MTX selection, cell lines A and C show loss in light chain copy number at PDL35 (Figure 4B) consistent with the observed decreased in relative qP (Figure 4A). A reduction in light chain copy number was seen in cell line B by PDL60 (Figure 4B). Despite these losses, the mAb heavy chain copy number was maintained at above 60% relative copy number for all cell line ages (Figure 4C). For cultures aged in MTX, the light chain and heavy chain copy numbers remained constant with age for cell lines B and C (Figure 4BC) while the maintenance of MTX selection pressure may have resulted in a duplication of the light chain expression cassette between PDL60 and PDL90 for cell line A (Figure 4B).

In a parallel analysis, RNA expression of the mAb light and heavy chain indicated expression was lower in cell lines in non-MTX aged conditions as compared to the MTX aged conditions (Figure 4D). A decrease in light chain expression was observed in cell lines B and C when aged without MTX selection(Figure 4D) but did not show

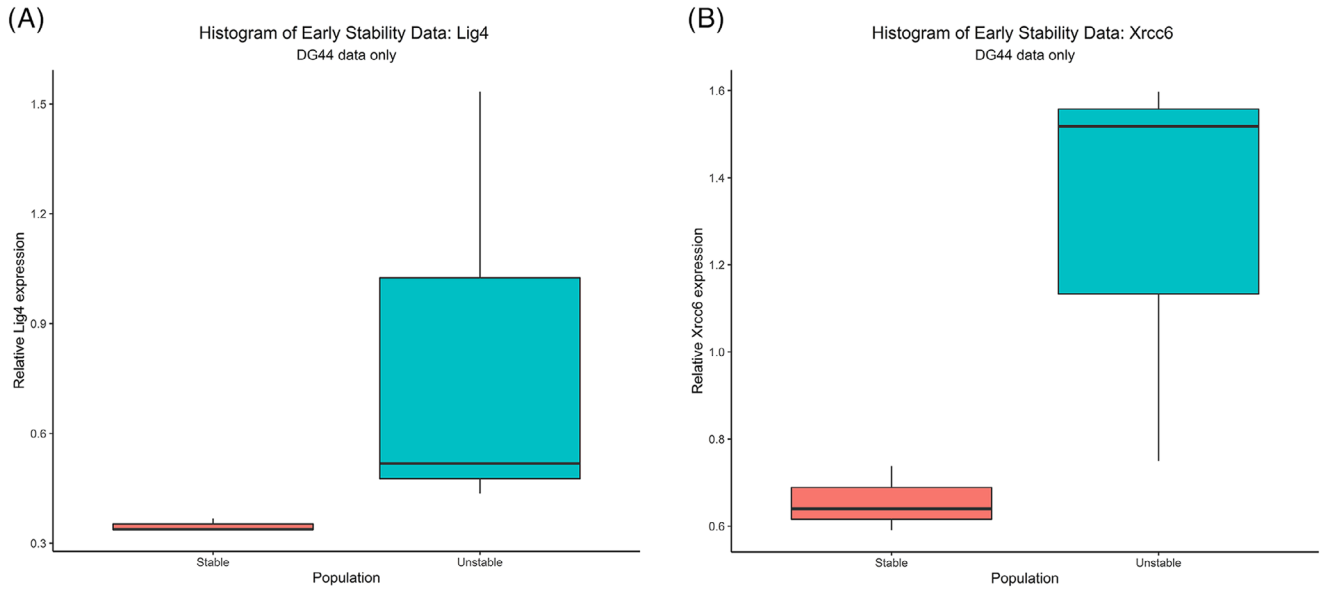

FIGURE 3 Box plot of relative expression of Lig4 and Xrcc6 DNA repair genes in early age of industrial CHO-DG44 cell lines. Stable cell lines were defined by ≥ 70% relative productivity at PDL60 and considers only samples between PDL0 and 30 with normalization to housekeeping gene Actr5 and a CHO-K1 RNA standard.

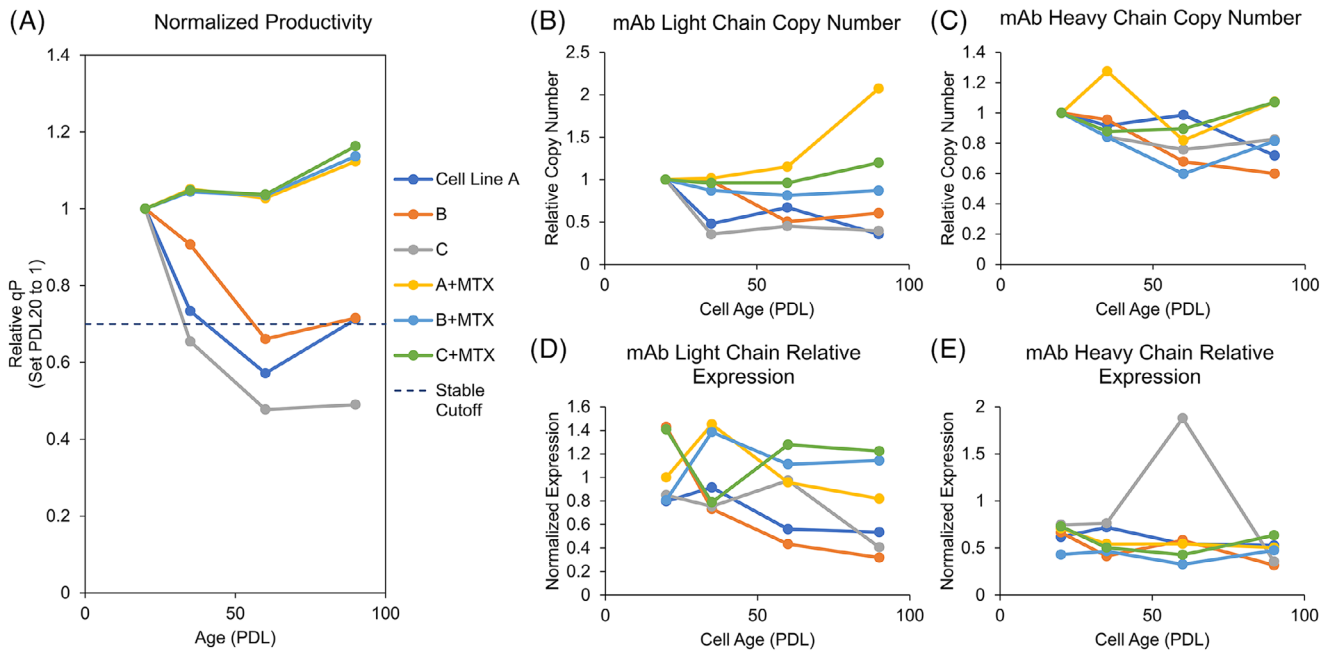

FIGURE 4 Relative cell specific productivity, mAb copy number and mRNA expression of CHO-DG44 cell lines. (A) Cell specific productivity (qP) shows production instability occurs without maintenance of MTX with the limit of stability (70%) shown as a dashed line. Relative copy number of the mAb light chain (B) and heavy chain (C) demonstrate decline with age without selection mirroring relative qP. mRNA expression of the light chain (D) decreases with age in cell lines aged without MTX while lower expression was observed for heavy chain mRNA expression (E). Points represent average of two replicates normalized to the PDL20 value.

the same correlation with relative qP as light chain copy number (Figure 4B). For all samples, light chain expression was maintained higher than heavy chain expression (Figure 4D,E). Maintenance of MTX selection led to constant mRNA expression (both heavy and light chain) that did not change with age except for the likely outlier in cell line C at PDL60 (Figure 4E). Overall, copy number loss is the mechanism for production instability within the studied CHO-DG44 cell lines.

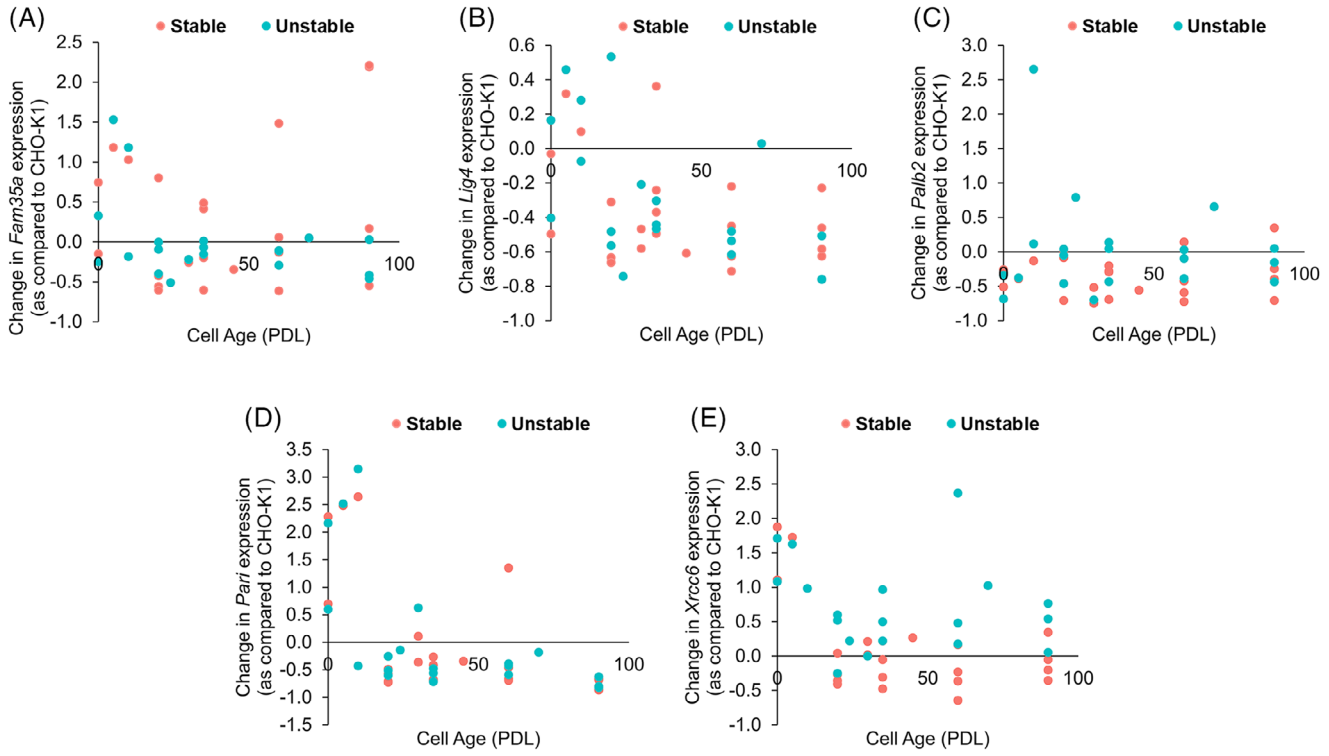

FIGURE 5 Expression profile of DNA repair genes as a function of cell age. All qPCR data across a span of ages was combined for this analysis. Expression decreases as cells age for Fam35a (A), Lig4 (B), Palb2 (C), Pari (D), and Xrcc6 (E).

3.3 | DNA repair gene expression profiles with cell age

We sought to combine all collected qPCR results for DNA repair genes across all sample ages and identify predictive profiles associated with cell aging. A total of 44 samples were available to model expression changes as a function of age (PDL) with 52% representing stable cell lines (Table S4). Expression of DNA repair genes generally decreased as cells aged when normalized to the CHO-K1 reference sample (Figure 5). Linear regressions fit to the full dataset for each gene showed a negative correlation between cell age and *Lig4* (Figure 5B) or *Pari* (Figure 5D) expression, where expression consistently fell in both stable and unstable populations (p-value < 0.005, Table S5). The other DSB repair genes, *Fam35a*, *Palb2*, and *Xrcc6* showed no overarching correlation with cell age (Figure 5ACE).

When considering the stable and unstable populations as separate groups for linear regression, *Fam35a* expression shows a slightly negative correlation specifically within unstable cell lines. Interestingly, *Xrcc6* expression rapidly drops beyond PDL20 for both populations with stable cell lines showing lower expression as compared to unstable cell lines. The *Xrcc6* expression in stable cell lines was significantly correlated with age (Table S5). Assessment of DNA repair gene expression, specifically *Xrcc6*, provides useful insight that could be leveraged throughout culture duration.

4 | DISCUSSION

In this work, DNA expression data was collected from several companies and organizations using different CHO cell line hosts to understand the feasibility of identifying biomarkers for predicting cell line stability. Over 40 mAb production cell line samples were analyzed for DNA repair gene expression at various cell ages (Figure 1). The five genes assessed here (*Fam35a*, *Lig4*, *Palb2*, *Pari*, and *Xrcc6*) are genes involved with different pathways for DNA repair and were selected based on mutations in their coding sequences between the Chinese hamster and CHO genomes. PCA analysis of expression patterns for these five DNA repair genes identified CHO host specific profiles with statistically significant classification of stable populations for CHO-DG44 samples (Figure 2). The distinct grouping between CHO-GS host cell lines and CHO-DG44 cell line is not surprising, as CHO-DG44 was generated through mutagenic selection and has a predominant history of chromosomal haploidy while CHO-GS was created through targeted genetic engineering.[1] The mutagenesis required to create the CHO-DG44 host cell line may be associated with deficiencies in the DNA repair machinery. Given the targeted nature of genetic modifications.[2] DNA repair machinery may be more intact within the CHO-GS cell line. With many possible mechanisms of instability including chromosomal rearrangement and epigenetic changes, additional work is needed to identify biomarkers withing the CHO-GS and CHO-K1 cell lineages.

Within the CHO-DG44 samples, *Lig4* and *Xrcc6* expression levels were elevated in unstable cell lines (Figure 3) while *Fam35a*, *Palb2*, and *Pari* expression did not change between the stable and unstable groups (Figure S2-S4). These findings support the importance of *Lig4* and *Xrcc6* in CHO cell line stability as previous work[3] has shown that expression of the Chinese hamster versions of these genes maintained higher average titer and transgene copy numbers over 74 days of culture. In human cells, both *Lig4* and *Xrcc6* have associations with cancer and other conditions which also display genome instability. Specifically, elevated *Lig4* levels have been associated with glioma and prostate cancer[21–22] while elevated *Xrcc6* expression has been studied within various cancer modalities from gastric, lung, and cervical cancers to osteosarcoma.[20] In the case of osteosarcoma, upregulation of *Xrcc6* is observed in cancerous cells as compared to adjacent non-tumor tissues.[20] The higher expression of *Xrcc6* seen in this work for unstable CHO cell lines in conjunction with previous findings further supports a conserved mechanism for genome destabilization via DNA repair pathways.

A fed-batch study of CHO-DG44 clones identified mAb copy number loss as the primary mechanism for cell line instability (Figure 4). Only cell lines with MTX addition during the aging process displayed stable production which is expected although not ideal for manufacturing purposes. Cell lines A and B showed maintenance of heavy chain copy number with light chain copy number loss observed in all cultures aged without MTX selection. Copy number loss is a common mechanism of production instability particularly within the CHO-DG44 host cell line as a result of the random mutagenesis used to develop the cell line platform.[1,7] In addition to mutagenesis, gene amplification, often used to generate high expressing pools, has significant limitations as many cell lines with high transgene copy numbers subsequently lose these sequences through spontaneous elimination.[1,4,7] Copy number loss represents only one of many mechanisms which contribute to challenges in CHO cell line stability, which include chromosomal rearrangement, DNA methylation, transcriptional/translational silencing, and histone modification.[1,4,7,28]

Collection of additional data points across CHO cell ages from early culture to long-term (PDL100+) can continue to identify patterns associated with cell line stability. From the combined dataset which spans multiple ages and CHO host lineages (Figure 5), *Lig4* and *Pari* show decreased expression with cell age while *Xrcc6* levels are significantly lower in stable cell lines which can be used in assessing cell lines during aging. Expression of *Pari* is related to HR suggesting that lower expression would enable higher recombination frequencies making an unstable genome more likely.[18] As shown previously, *Lig4* and *Xrcc6* have distinct behavior useful toward predicting cell line instability.[3]

This work identifies possible biomarkers for CHO cell line instability for the CHO-DG44 lineage. Expression profiling efforts within CHO have also predicted relative qP from a dataset of 70 cell lines.[29] These efforts identified *Pari* expression as part of the qP predictive model[29] but did not identify impacts from other DNA repair genes. With a dataset of similar magnitude, this work identified *Lig4* and *Xrcc6* as critical players which suggests a subset of genes can be profiled to predict properties in CHO cells. Overall, these results suggest the potential for using gene expression-based biomarkers to predict cell line stability which could reduce the time and resource intensive aging studies of CHO production clones.

AUTHOR CONTRIBUTIONS

Lauren T. Cordova: Data curation; Formal analysis; Methodology; Writing – original draft; Writing – review & editing. Hussain Dahodwala: Conceptualization; Investigation; Writing – original draft; Writing – review & editing. Rebecca L. Cooley: Investigation; Methodology; Writing – review & editing. Kelvin H. Lee: Conceptualization; Funding acquisition; Project administration; Supervision; Writing – review & editing

ACKNOWLEDGMENTS

The authors thank Hunter Balduf for his assistance with this project. Additionally, the authors thank the National Institutes of Health for donating materials and additional Pfizer and Eli Lilly staff for supporting data collection as part of this work. This work was supported by the National Science Foundation Industrial-University Cooperative Research Center (IUCRC) program through the Advanced Mammalian Biomanufacturing Innovation Center (AMBIC) (IIP – 1624698) and AMBIC industrial membership fees and it was supported in part by the National Institute of Standards and Technology 70NANB17H002. KHL is named as an inventor on a patent owned by the University of Delaware related to DNA repair genes. The other authors declare no commercial or financial conflict of interest. This work does not contain data from experiments using animals or human studies.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Kelvin H. Lee https://orcid.org/0000-0003-0908-2981

REFERENCES

1. Dahodwala, H., & Lee, K. H. (2019). The fickle CHO: A review of the causes, implications, and potential alleviation of the CHO cell line instability problem. *Current Opinion in Biotechnology*, *60*, 128–137. https://doi.org/10.1016/J.COPBIO.2019.01.011
2. Jagschies, G. (2018). Selected biotherapeutics overview. *Biopharmaceutical Processing: Development, Design, and Implementation of Manufacturing Processes*, 59–72. https://doi.org/10.1016/B978-0-08-100623-8.00003-7
3. Spahn, P. N., Zhang, X., Hu, Q., Lu, H., Hamaker, N. K., Hefzi, H., Li, S., Kuo, C. C., Huang, Y., Lee, J. C., Davis, A. J., Ly, P., Lee, K. H., & Lewis, N. E. (2022). Restoration of DNA repair mitigates genome instability and increases productivity of Chinese hamster ovary cells. *Biotechnology and Bioengineering*, *119*(3), 963–982. https://doi.org/10.1002/BIT.28016
4. Kim, M., O'Callaghan, P. M., Droms, K. A., & James, D. C. (2011). A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. *Biotechnology and Bioengineering*, *108*(10), 2434–2446. https://doi.org/10.1002/BIT.23189

5. McBurney, M. W., Mai, T., Yang, X., & Jardine, K. (2002). Evidence for repeat-induced gene silencing in cultured mammalian cells: Inactivation of tandem repeats of transfected genes. *Experimental Cell Research*, *274*(1), 1–8. https://doi.org/10.1006/EXCR.2001.5443
6. Barnes, L. M., Bentley, C. M., & Dickson, A. J. (2003). Stability of protein production from recombinant mammalian cells. *Biotechnology and Bioengineering*, *81*(6), 631–639. https://doi.org/10.1002/BIT.10517
7. Dorai, H., Corisdeo, S., Ellis, D., Kinney, C., Chomo, M., Hawley-Nelson, P., Moore, G., Betenbaugh, M. J., & Ganguly, S. (2012). Early prediction of instability of chinese hamster ovary cell lines expressing recombinant antibodies and antibody-fusion proteins. *Biotechnology and Bioengineering*, *109*(4), 1016–1030. https://doi.org/10.1002/BIT.24367
8. PUCK, T. T., CIECIURA, S. J., & ROBINSON, A. (1958). Genetics of somatic mammalian cells. *Journal of Experimental Medicine*, *108*(6), 945. https://doi.org/10.1084/JEM.108.6.945
9. Szkodny, A. C., & Lee, K. H. (2022). Biopharmaceutical manufacturing: Historical perspectives and future directions. *Annual Review of Chemical and Biomolecular Engineering*, *13*, 141–165. https://doi.org/10.1146/ANNUREV-CHEMBIOENG-092220-125832
10. Wurm, F. M., & Wurm, M. J. (2017). Cloning of CHO cells, productivity and genetic stability—A discussion. *Processes 2017*, *5*(2), 20. https://doi.org/10.3390/PR5020020
11. Baik, J. Y., & Lee, K. H. (2017). A framework to quantify karyotype variation associated with CHO cell line instability at a single-cell level. *Biotechnology and Bioengineering*, *114*(5), 1045–1053. https://doi.org/10.1002/BIT.26231
12. Hamaker, N. K., & Lee, K. H. (2020). A site-specific integration reporter system that enables rapid evaluation of CRISPR/Cas9-mediated genome editing strategies in CHO cells. *Biotechnology Journal*, *15*(8), 2000057. https://doi.org/10.1002/BIOT.202000057
13. Kostyrko, K., Bosshard, S., Urban, Z., & Mermod, N. (2015). A role for homologous recombination proteins in cell cycle regulation. *Cell Cycle*, *14*(17), 2853–2861. https://doi.org/10.1080/15384101.2015.1049784
14. Lee, K. H., Onitsuka, M., Honda, K., Ohtake, H., & Omasa, T. (2013). Rapid construction of transgene-amplified CHO cell lines by cell cycle checkpoint engineering. *Applied Microbiology and Biotechnology*, *97*(13), 5731–5741. https://doi.org/10.1007/S00253-013-4923-9
15. Sancar, A., Lindsey-Boltz, L. A., Ünsal-Kaçmaz, K., & Linn, S. (2004). Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. *Annual Review of Chemical and Biomolecular Engineering.073723*, *73*, 39–85. https://doi.org/10.1146/ANNUREV.BIOCHEM.73.011303.073723
16. Khanna, K. K., & Jackson, S. P. (2001). DNA double-strand breaks: Signaling, repair and the cancer connection. *Nature Genetics*, *27*, 247–254. https://doi.org/10.1038/85798
17. Findlay, S., Heath, J., Luo, V. M., Malina, A., Morin, T., Coulombe, Y., Djerir, B., Li, Z., Samiei, A., Simo-Cheyou, E., Karam, M., Bagci, H., Rahat, D., Grapton, D., Lavoie, E. G., Dove, C., Khaled, H., Kuasne, H., Mann, K. K., … & Orthwein, A. (2018). SHLD 2/ FAM 35A co-operates with REV 7 to coordinate DNA double-strand break repair pathway choice. *EMBO Journal*, *37*, e100158. https://doi.org/10.15252/embj.2018100158
18. Bosshard, S., Duroy, P. O., & Mermod, N. (2019). A role for alternative end-joining factors in homologous recombination and genome editing in Chinese hamster ovary cells. *DNA Repair*, *82*, 102691. https://doi.org/10.1016/J.DNAREP.2019.102691
19. Wang, C., Sun, Z., Wang, M., Jiang, Z., Zhang, M., Cao, H., Luo, L., Qiao, C., Xiao, H., Chen, G., Li, X., Liu, J., Wei, Z., Shen, B., Wang, J., & Feng, J. (2022). Novel CRISPR/Cas9-mediated knockout of LIG4 increases efficiency of site-specific integration in Chinese hamster ovary cell line. *Biotechnology Letters*, *3*, 1063–1072. https://doi.org/10.1007/s10529-022-03282-7
20. Zhu, B., Cheng, D., Li, S., Zhou, S., & Yang, Q. (2016). High expression of XRCC6 promotes human osteosarcoma cell proliferation through the β-catenin/wnt signaling pathway and is associated with poor prognosis. *International Journal of Molecular Sciences*, *17*, 1188. https://doi.org/10.3390/IJMS17071188
21. Grupp, K., Roettger, L., Kluth, M., Hube-Magg, C., Simon, R., Lebok, P., Minner, S., Tsourlakis, M. C., Koop, C., Graefen, M., Adam, M., Haese, A., Wittmer, C., Sauter, G., Wilczak, W., Huland, H., Schlomm, T., Steurer, S., & Krech, T. (2015). Expression of DNA ligase IV is linked to poor prognosis and characterizes a subset of prostate cancers harboring TMPRSS2: ERG fusion and PTEN deletion. *Oncology Reports*, *34*(3), 1211–1220. https://doi.org/10.3892/OR.2015.4080
22. Liu, Y., Zhou, K., Zhang, H., Yin, Y. S., Chen, L., Xu, Z., Zhong, Y., Liu, H., Jin, L., Wei, Q., Huang, F., Lu, D., & Zhou, L. (2008). Polymorphisms of LIG4 and XRCC4 involved in the NHEJ pathway interact to modify risk of glioma. *Human Mutation*, *29*(3), 381–389. https://doi.org/10.1002/HUMU.20645
23. Hamaker, N. K., Min, L., & Lee, K. H. (2022). Comprehensive assessment of host cell protein expression after extended culture and bioreactor production of CHO cell lines. *Biotechnology and Bioengineering*, *119*(8), 2221–2238. https://doi.org/10.1002/BIT.28128
24. Cordova, L. T., Dahodwala, H., Elliott, K. S., Baik, J., Odenewelder, D. C., Nmagu, D., Skelton, B. A., Uy, L., Klaubert, S. R., Synoground, B. F., Chitwood, D. G., Dhara, V. G., Naik, H. M., Morris, C. S., Yoon, S., Betenbaugh, M., Coffman, J., Swartzwelder, F., Gilmeister, M., … & Lee, K. H. (2022). Generation of reference cell lines, media, and a process platform for CHO cell biomanufacturing. *Biotechnology and Bioengineering*, *120*(3), 715–725. https://doi.org/10.1002/BIT.28290
25. Bailey, L. A., Hatton, D., Field, R., & Dickson, A. J. (2012). Determination of Chinese hamster ovary cell line stability and recombinant antibody expression during long-term culture. *Biotechnology and Bioengineering*, *109*(8), 2093–2103. https://doi.org/10.1002/BIT.24485
26. Todorov, H., Fournier, D., & Gerber, S. (2018). Principal components analysis: Theory and application to gene expression data analysis. *Genomics and Computational Biology*, *4*(2), 100041.
27. CHOZN® Cell Line Engineering & Development. (n.d.). Retrieved October 17, 2023, from https://www.sigmaaldrich.com/US/en/services/custom-products/chozn-cell-line-engineering-and-development?gclid=Cj0KCQjwhL6pBhDjARIsAGx8D595FTmXOq0amSZJsPSvqaCKoBJFE7C2NSViq1K8XmZGgq1Vh6JIZEkaAq4JEALw_wcB
28. Yang, Y., Mariati, Chusainow, J., & Yap, M. G. S. (2010). DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. *Journal of Biotechnology*, *147*(3–4), 180–185. https://doi.org/10.1016/J.JBIOTEC.2010.04.004
29. Clarke, C., Doolan, P., Barron, N., Meleady, P., O'sullivan, F., Gammell, P., Melville, M., Leonard, M., & Clynes, M. (2011). Predicting cell-specific productivity from CHO gene expression. *Journal of Biotechnology*, *151*, 159–165. https://doi.org/10.1016/j.jbiotec.2010.11.016

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Cordova, L. T., Dahodwala, H., Cooley, R., & Lee, K. H. (2024). Prediction of CHO cell line stability using expression of DNA repair genes. *Biotechnology Journal*, *19*, e2300425. https://doi.org/10.1002/biot.202300425