Title

Restoration of DNA repair mitigates genome instability and increases productivity of Chinese hamster ovary cells.

Journal

Biotechnology and Bioengineering, 119(3)

Authors

Spahn, Philipp

Zhang, Xiaolin

Hu, Qing

et al.

Publication Date

2022-03-01

DOI

10.1002/bit.28016

Peer reviewed

HHS Public Access

Author manuscript

Biotechnol Bioeng. Author manuscript; available in PMC 2023 March 01.

Published in final edited form as:

Biotechnol Bioeng. 2022 March ; 119(3): 963–982. doi:10.1002/bit.28016.

Restoration of DNA Repair Mitigates Genome Instability and Increases Productivity of Chinese Hamster Ovary Cells

Philipp N. Spahn[1,2,#], **Xiaolin Zhang**[3,#], **Qing Hu**[4], **Huiming Lu**[5], **Nathaniel K. Hamaker**[3], **Hooman Hefzi**[6], **Shangzhong Li**[6], **Chih-Chung Kuo**[6], **Yingxiang Huang**[1], **Jamie C. Lee**[1,6], **Anthony J. Davis**[5], **Peter Ly**[4], **Kelvin H. Lee**[3,*], **Nathan E. Lewis**[1,2,6,*]

[1] Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093

[2] The Novo Nordisk Foundation Center for Biosustainability at the University of California, San Diego School of Medicine, San Diego, La Jolla, CA 92093

[3] Department of Chemical and Biomolecular Engineering, University of Delaware, Newark, DE 19711

[4] Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX 75390

[5] Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX 75390

[6] Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093

Abstract

Chinese Hamster Ovary (CHO) cells are the primary host for manufacturing of therapeutic proteins. However, productivity loss is a major problem and is associated with genome instability, as chromosomal aberrations reduce transgene copy number and decrease protein expression. We analyzed whole-genome sequencing data from 11 CHO cell lines and found deleterious single-nucleotide variants in DNA repair genes. Comparison with primary Chinese hamster cells confirmed DNA repair to be compromised in CHO. Correction of key DNA repair genes by SNV reversal or expression of intact cDNAs successfully improved DNA repair and mitigated karyotypic instability. Moreover, overexpression of intact copies of *LIG4* and *XRCC6*

[*] These authors jointly supervised this work: Kelvin H. Lee, khl@udel.edu, 302-831-0344, Nathan E. Lewis, n4lewis@eng.ucsd.edu, 858-997-5844.

[#]These authors contributed equally: Philipp N. Spahn, Xiaolin Zhang

Author Contributions

PNS, XZ, KHL, NEL designed the project. QH, PL performed and analyzed karyotype experiments. HL, AJD performed and analyzed immunoblotting experiments. NKH, JCL assisted with experiments. HH, SL, YH generated SNV data. CCK analyzed RNA-seq data. PNS, XZ performed and analyzed all remaining experiments. PNS, XZ, KHL, NEL wrote the paper. PNS, XZ contributed equally to this work.

Competing Interests

PNS, XZ, KHL, and NEL are listed as inventors on two pending patent applications by the University of Delaware and the University of California, related to observations in this manuscript. The remaining authors declare no competing interests. All authors have approved the final manuscript.

Code Availability

Code used for this study can be found at:
https://github.com/LewisLabUCSD/NGS-Pipeline
https://github.com/LewisLabUCSD/cho_dna_variant_call
https://github.com/PhilippSpahn/Confocal_Image_Processing

in a CHO cell line expressing secreted alkaline phosphatase mitigated transgene copy loss and improved protein titer retention. These results show that correction of DNA repair genes yields improvements in genome stability in CHO, and provide new opportunities for cell line development for sustainable protein expression.

Keywords

protein expression; cell line instability; CHO; DNA repair

Introduction

Chinese Hamster Ovary (CHO) cells have been the leading expression system for the industrial production of therapeutic proteins for over 40 years, and projections show they will maintain this dominant position into the foreseeable future because they produce >80% of therapeutic proteins approved between 2014–2018 corresponding to more than $100 billion of products per year (Walsh, 2018). Steady improvements in cell line development, media formulation, and bioprocessing now enable production yields exceeding 10 g/L from a fed-batch culture. Emerging resources, including the CHO and hamster genome sequence (Hilliard, MacDonald, & Lee, 2020; Lewis et al., 2013; Rupp et al., 2018; Xu et al., 2011; Yusufi et al., 2017) and genome editing tools (J. S. Lee, Grav, Lewis, & Kildegaard, 2015) now allow researchers to rely less on largely empirical, “trial-and-error” approaches to CHO cell line development, and move towards a more rational engineering approach, in pursuit of optimized CHO lines with tailored, superior attributes (Kildegaard, Baycin-Hizal, Lewis, & Betenbaugh, 2013; Stolfa et al., 2018).

However, cell line instability, i.e. the propensity of a cell line to lose industrially relevant properties over time, remains a challenging problem in the field as it can result in loss of titer, changes in product quality, or changes in cell growth potentially impacting manufacturing and drug supply. Thus, considerable effort is expended to ensure that production clones are stable for use in fed-batch production - for example, that a given clone retains 70% of its titer after 60 generations in culture which would be deemed a “stable” producer (Bailey, Hatton, Field, & Dickson, 2012). However, as the field moves toward perfusion and continuous culture, cell line stability remains a key challenge to overcome.

Production instability can arise from transcriptional transgene silencing through epigenetic mechanisms, such as promoter methylation (Chusainow et al., 2009; M. Kim, O’Callaghan, Droms, & James, 2011; Moritz, Woltering, Becker, & Göpfert, 2016; Mutskov & Felsenfeld, 2004; Paredes, Park, Jeong, Yoon, & Baek, 2013; Veith, Ziehr, MacLeod, & Reamon-Buettner, 2016; Yang, Mariati, Chusainow, & Yap, 2010). However, while such epigenetic transcriptional silencing was associated with production instability in some studies, a loss in transgene copy number due to genome instability (i.e. the rapid accumulation of mutations and chromosome aberrations) appears to be the predominant cause of production instability (Baik & Lee, 2016; Bandyopadhyay et al., 2019; Barnes, Bentley, & Dickson, 2003; Beckmann et al., 2012; Fann, Guirgis, Chen, Lao, & Piret, 2000; Hammill, Welles, & Carson, 2000; M. Kim et al., 2011; S. J. Kim, Kim, Ryu, Hong, & Lee, 1998; Veith et al.,

2016), although it is very possible that these phenomena are not mutually exclusive. The inability to maintain genome integrity will negatively affect production stability in most, if not all, production cell lines during long-term culture. Because massive transgene expression imposes a high metabolic demand on the host, transgenes are effectively under negative selection in the bioreactor, causing the emergence of non-producing subpopulations that can outgrow the remaining producers in the pool, resulting in titer decline.

Genome instability is common across cancers, and can arise from defects in DNA damage repair (Khanna & Jackson, 2001). In CHO, genome instability was first reported in the 1970s when direct observations of CHO chromosomes revealed a divergence from the Chinese hamster (*Cricetulus griseus*) karyotype and variability in karyotypes among CHO clones (Deaven & Petersen, 1973). These karyotype variations occur regardless of growth condition, and do not differ markedly between pooled and clonal populations (Feichtinger et al., 2016; Vcelar et al., 2018; Wurm & Wurm, 2017). Loss of chromosomal material and improper chromosome fusions (translocations) are caused by improperly repaired double-strand breaks (DSBs) which result from metabolic processes, such as attack by free radicals or collapsed DNA replication forks (Van Gent, Hoeijmakers, & Kanaar, 2001). Given that primary cells from the Chinese hamster do not show genome instability, DSB repair capacity is likely compromised in CHO. Thus, identifying and restoring deficient DSB repair genes in CHO cells could effectively enhance the DSB repair system, mitigate genome instability, and improve product titer stability.

To date, few studies have addressed the persistent genome instability problem in CHO cells, and it has been challenging to develop effective counterstrategies. Detailed quantification of chromosomal instabilities in production cell lines has indicated that certain chromosome sites are less prone to instability than others (Hilliard & Lee, 2021; Kaas, Kristensen, Betenbaugh, & Andersen, 2015). This observation has suggested that transgene loss may be avoided by targeting transgenes to these stable chromosomal areas, an option now possible through the development of targeted transgene integration techniques (Gaidukov et al., 2018; Hamaker & Lee, 2018; J. S. Lee, Kallehauge, Pedersen, & Kildegaard, 2015). Further studies used gene knock-outs (*ATR* and *BRCA1*, respectively) to increase product titer by increasing transgene copy number amplification (K. H. Lee, Onitsuka, Honda, Ohtake, & Omasa, 2013; Matsuyama, Yamano, Kawamura, & Omasa, 2017), but whether these knock-outs are able to sustain high production in long-term culture has remained unclear. Thus, a pressing need remains for novel approaches to mitigate production instability stemming from DSBs. Although the mechanistic connections between production instability and DNA DSBs are evident, the field has not systematically explored the engineering of DNA repair as a possible means to reduce transgene loss in CHO. Inactivation of *ATR* was shown to result in an increase in transgene copies during the amplification phase, but also a less rigid cell cycle control and higher chromosomal instability (K. H. Lee et al., 2013), which might exacerbate production instability in the long run. Therefore, rather than inactivating DNA repair genes for short-term gains, rescue of deficient DNA repair could constitute a promising approach to achieve long-term improvement in stability.

Here we investigate the relationship between DSB repair and genome instability in CHO cells. We describe the landscape of mutations in DNA repair genes across a panel of 11

commonly used CHO lines, and show that DSB repair capacity is compromised in CHO compared to the Chinese hamster. We then show that restoration of key DSB repair genes yields significant improvements in repair capacity and genome stability. Finally, we show in a proof-of-concept experiment that correction of DNA repair genes *XRCC6* and *LIG4* improves the stability of transgene copy number and protein titer during extended culture of a CHO cell line secreting alkaline phosphatase (SEAP).

Results

CHO cells harbor numerous mutations in DNA repair genes and are deficient in DSB repair

To test if DSB repair is deficient in CHO cells, we compared DSB repair between CHO-K1 (a commonly used ancestral Chinese Hamster Ovary cell line) and Chinese hamster *(Cricetulus griseus)* fibroblasts, freshly extracted from lung tissue. The endogenous DSB level was compared by measuring the intensity of γH2AX foci. γH2AX denotes phosphorylated histone H2AX in the chromatin area surrounding a DSB which often extends several megabases from the break site, visible as a focus in confocal microscopy and is a well-established marker of DSBs (Rogakou, Boon, Redon, & Bonner, 1999). CHO-K1 showed significantly elevated levels of γH2AX signal under untreated conditions, indicating a higher endogenous level of unrepaired DSB compared to hamster fibroblasts (Fig. 1a,c). To compare the DSB repair capacity in both samples, we induced DSBs by treating the cells with 5 Gy of ionizing radiation, and analyzed γH2AX signal at multiple time-points during recovery. As early as 5 min post irradiation, a strong increase in γH2AX signal was evident indicating the activation of the DNA damage response that further intensified until 60 min post irradiation. Noticeably, CHO-K1 cells showed a much stronger increase in γH2AX signal (Fig. 1a,c), consistent with a higher number of DSBs and hence slower DSB repair. After 24 h of recovery, γH2AX signal had almost vanished in the fibroblasts, indicating complete repair of the irradiation-induced lesions, while a strong signal was still visible in CHO-K1 (Fig. 1a,c). We also noted a higher degree of nuclear fragmentation in CHO-K1 cells at t=24h (Fig. 1b,c), indicative of higher rates of apoptosis. Thus, DSB repair is compromised in CHO-K1 cells, compared to genomically stable hamster fibroblasts. We reasoned that, given the high overall burden of mutations in the CHO genome (Lewis et al., 2013), mutations in DNA repair genes would likely contribute to this DNA repair deficiency.

To assess the extent of mutations in genes affecting DNA repair, we analyzed whole-genome sequencing (WGS) data of 11 CHO cell lines, including those commonly used for cell line development in biopharmaceutical production (e.g. CHO-S, CHO-XB11, CHO-DG44) (Table 1). We extracted a list of genes related to DNA repair from a curated database (Wood, Mitchell, & Lindahl, 2005), and aligned the respective CHO homologs to the Chinese hamster genome (Lewis et al., 2013; Rupp et al., 2018). We looked for non-synonymous single-nucleotide variants (SNVs) and short indels (<10 bp), located in exons, and used a phylogeny-based score to predict the deleteriousness of these mutations on gene function (PROVEAN score) (Choi, Sims, Murphy, Miller, & Chan, 2012). All 11 CHO lines were heavily affected by SNVs in DNA repair genes (Suppl. Fig. 1). We identified 157 mutations in 14 ontology classes related to genome stability (Fig. 2a). Using copy-number information derived from the WGS datasets, we retained only mutations which had undergone a loss of

heterozygosity due to their higher likelihood of causing loss-of-function phenotypes. After this filtering, the most widespread and most deleterious mutations occurred in genes related to non-homologous end-joining (NHEJ), DNA-damage sensing, and in helicases (Fig. 2b: box). All of these ontology classes are highly relevant for the maintenance of chromosome stability, so we reasoned that correction of mutations in these genes could be a way to mitigate chromosomal instability in CHO, by enhancing DNA repair capacity, in particular the repair of DSBs.

Correction of DNA repair genes improves DNA damage response in CHO

The cluster of most pervasive, most deleterious mutations (Fig. 2b: box) comprised six SNVs in four genes (*ATM* R2830H, *RAD1* E125G, *WRN* R879Q, *WRN* V1096A, *PRKDC* D1641N, *PRKDC* S3419G). In mammalian DNA repair, the protein kinase *Ataxia-telangiectasia mutated* (*ATM*) is among the first factors to become activated upon DSBs, and thus represents a key upstream player in orchestrating the damage response to DSBs (Fig. 2c) (Shiloh, 2006). All 11 CHO cell lines carry a SNV in the catalytic domain (R2830H, Fig. 3a) which is predicted to be deleterious according to PROVEAN scoring. We thus decided that this SNV would make a promising target for gene correction.
DSB repair occurs through two main pathways: non-homologous end-joining (NHEJ) and homology-directed repair (HDR) (Jensen & Rothenberg, 2020). In HDR, a sequence template homologous to the damaged strand is utilized to produce an error-free repair of the DNA lesion. In NHEJ, DNA ends from both sides of the lesion are ligated without using homology information which often implies loss of base pairs. While HDR is clearly preferable for maintaining sequence integrity, its activation requires homologous sequences which are only available during S- and G2-phase, but unavailable during the longer G1 phase. In addition, homozygous mutations in HDR genes were less widespread among CHO cell lines (Fig. 2b). Thus, we reasoned that restoration of NHEJ would be of primary importance to mitigate chromosomal instability and transgene loss in CHO. The gene carrying the most widespread and deleterious mutations within the NHEJ ontology was *PRKDC* which constitutes the catalytic subunit of the DNA-PK (DNA-dependent protein kinase) which forms a stable complex with DNA and the Ku heterodimer at the DSB site and bridges the loose chromosome ends required for NHEJ (Davis & Chen, 2013). In our dataset, all 11 CHO cell lines carry two SNVs in *PRKDC*, both predicted to be deleterious according to PROVEAN. Thus, we reasoned that *PRKDC* would be another high-priority target for gene correction efforts. We prioritized SNV D1641N in the FAT domain over S3419G due its PROVEAN score indicating higher deleteriousness (Fig. 3b).

Using CRISPR/Cas9-mediated gene correction protocols (Integrated DNA Technologies), we generated a clonal CHO-K1 cell line with a successful reversal of R2830H in *ATM* (hereafter referred to as CHO-K1 ATM^+) (Fig. 3a). From this population, we generated a sub-clone with a successful reversal of D1641N in *PRKDC* (hereafter referred to as CHO-K1 ATM^+ $PRKDC^+$) (Fig. 3b). These reversals were done in succession to assess the cumulative effect of DNA repair improvements. Gene expression analysis revealed that *ATM* levels in CHO-K1 WT (carrying the R2830H mutation) were slightly upregulated compared to Chinese hamster fibroblasts (Fig. 3c). Interestingly, the reversal of R2830H reduced *ATM* transcript abundance back to levels comparable to hamster fibroblasts (Fig.

3c,e). Immunoblotting of ATM after treatment with 10 Gy X-ray radiation confirmed elevated ATM phosphorylation and abundance in CHO-K1 WT, which was reduced in the engineered cell lines, resembling the lower levels seen in hamster fibroblasts (Fig. 3d). *PRKDC* transcript and protein levels were slightly reduced in CHO-K1 WT compared to hamster fibroblasts but showed minor changes in the engineered cell lines (Fig. 3c–e). RNA-seq did not reveal major additional differences in the engineered ATM^+ and ATM^+ $PRKDC^+$ cell lines, and gene set enrichment analysis did not identify significantly up-/downregulated pathways (Fig. 3e), indicating that the reversals did not cause major changes in the transcriptome.

In addition to the high-priority gene targets *ATM* and *PRKDC*, we decided to also investigate *XRCC6* and *LIG4*, despite the heterozygosity of the SNVs in these genes (Fig. 3f,g), because these genes represent key players in the NHEJ pathway. *XRCC6* codes for Ku70, which forms a heterodimer with Ku80 and is required for the sensing of DSB ends. Once bound to the DSB, the Ku heterodimer serves as an important “tool belt” to bridge the DSB and allows for the binding of additional repair proteins. *LIG4* is the ligase responsible for the terminal step in NHEJ (Chang, Pannunzio, Adachi, & Lieber, 2017). Given the feasible length of their cDNAs (1.8 kb and 2.7 kb, respectively), we generated cell lines harboring intact Chinese hamster copies of these genes (hereafter referred to as CMV::XRCC6 and CMV::LIG4, Fig. 3h) to test whether overexpression of these NHEJ repair factors would improve DSB repair.

Chromosomal aberrations, often seen in CHO cell lines of high passage number (Vcelar et al., 2018), are caused by incorrect joining of non-corresponding ends after two separate DSBs (Van Gent et al., 2001). The EJ5-GFP reporter cassette models two such simultaneous separate DSBs to measure the frequency with which the incorrect ends are joined (Bennardo, Cheng, Huang, & Stark, 2008). Briefly, two DSBs are induced flanking a large spacer that separates a CMV promoter from a GFP reading frame. If the two distal, i.e. incorrect, ends are joined, the loss of the spacer triggers GFP expression (Fig. 4a–c). Sorting GFP-positive cells thus allow the quantification of DSB repair capability and the estimation of the propensity of a cell line to develop chromosomal aberrations. We validated the assay in CHO-K1 WT cells treated with KU-60019, a highly effective small-molecule inhibitor against the ATM kinase. Growing CHO-K1 in 3 μM KU-60019 caused a significant increase in GFP signal compared to untreated CHO-K1 WT, indicating higher rates of distal end-joining (Fig. 4d, Suppl. Fig. 2a), consistent with previous results obtained in mouse stem cells (Bennardo & Stark, 2010). Interestingly, CHO-K1 ATM^+ showed a significant decrease in GFP signal, indicating a successful reduction of distal end-joining events (Fig. 4d; Suppl. Fig. 2a). Further reduction in distal end-joining was obtained in ATM^+ $PRKDC^+$ (Fig. 4d; Suppl. Fig. 2a) indicating that combinatorial correction of these genes confers cumulative improvements in DSB repair. Considering that the *ATM* R2830H reversal also reverted *ATM* expression back to Chinese hamster levels (Fig. 3c), we tested whether reduction of *ATM* expression alone could be responsible for the reduction in distal-end joining. However, reduction of *ATM* transcript in CHO-K1 WT with an siRNA did not yield a major reduction in distal-end joining (Fig. 4d,e; Suppl. Fig. 2b), consistent with the role of ATM to suppress distal end-joining in mice (Bennardo & Stark, 2010). We thus conclude that reversal of the R2830H mutation in *ATM*, and not the mere reduction in transcript level, is responsible

for the improved DSB repair capability and the reduction in distal-end joining in CHO. In accordance with this, a control cell line harboring a frameshift mutation in *ATM* showed reduction in transcript and a strong increase in GFP signal indicating increased distal end-joining (Fig. 4d,e; Suppl. Fig. 2b,d). Finally, CHO-K1 cells overexpressing Chinese hamster *XRCC6* or *LIG4* copies showed a reduction in GFP signal indicating improved fidelity of DSB repair, but to a lesser extent than ATM$^+$ PRKDC$^+$ (Fig. 4d; Suppl. Fig. 2c).

The EJ5-GFP reporter quantifies the frequency of incorrect ("distal") end-joining of DSBs, but does not allow for an assessment of persistent DSBs, since failed end-joining would also yield a GFP-negative cell and would thus be indistinguishable from a correct ("proximal") end-joining event. Thus, in order to also compare the frequency of unrepaired DSBs, we analyzed γH2AX foci. After exposure to 5 Gy of ionizing radiation, all engineered CHO-K1 cell lines showed a faster clearance of γH2AX foci during recovery, indicating faster repair of DSBs (Fig. 4f,g). Together, these data show that correction of DNA repair genes successfully improves both fidelity and speed of DSB repair in CHO-K1 cells.

Restoration of DNA repair improves genome stability in CHO

Incorrect end-joining of DSBs can lead to chromosomal aberrations, ultimately driving transgene copy number loss (Baik & Lee, 2016). We thus asked whether the improvements in the DNA damage response in the engineered CHO cell lines would improve the integrity of their genome. For this, we analyzed single-cells by low-voltage electrophoresis where the length of the resulting DNA tail is an indicator of the amount of genome fragmentation (neutral comet assay). After exposing cells to 10 Gy of ionizing radiation, we noticed significant increase in tail length across all samples in cells lysed 5 min after exposure, indicating successful fragmentation of the genome (Fig. 5a,b). Over a timespan of 30 minutes after exposure, the majority of DSBs are typically repaired, visible as shorter comet tails (Olive & Banáth, 2006). Interestingly, the ATM$^+$ and ATM$^+$ PRKDC$^+$ cell lines showed significantly shorter tails after 30 minutes of recovery compared to CHO-K1 WT, indicating a lower level of genome fragmentation (Fig. 5a). In the case of ATM$^+$ PRKDC$^+$ this increased genome stability was evident under even higher doses of radiation (Suppl. Fig. 2e,f). In the absence of external genotoxic stress, a certain amount of endogenous DSBs can arise, e.g. as a consequence of defects during DNA replication (Aguilera & García-Muse, 2013). Interestingly, also in the absence of genotoxic stress, ATM$^+$ and ATM$^+$ PRKDC$^+$ cell lines showed shorter DNA tails (Fig. 5a), consistent with a lower level of persistent DBS breaks and improved genome stability under standard culturing conditions. We also observed slightly shorter DNA tails in CMV::XRCC6 and CMV::LIG4 cell lines under untreated conditions, but no major differences after irradiation and recovery (Fig. 5b), possibly because the SNVs in these genes are only heterozygous in CHO-K1 WT (Fig. 3f,g). Together, these results show that correction of key genes in DNA damage sensing and the NHEJ pathway can enhance genome stability in CHO-K1.

We thus wondered whether these engineered cell lines would show a mitigation of karyotype instability, a common problem in long-term culturing of CHO cells (Bandyopadhyay et al., 2019; Rouiller et al., 2015; Vcelar et al., 2018). We focused on ATM$^+$ and ATM$^+$ PRKDC$^+$

for this analysis, given the more pronounced improvement in genome stability in these cell lines.

Karyotypic analyses by multi-color DNA fluorescence *in situ* hybridization (FISH) revealed that CHO-K1 WT cells harbored several pre-existing chromosomal aberrations, including translocations of chromosomes #3, #6, and #7, whole-chromosome duplication of #4, and loss of the X chromosome ("main karyotype", Fig. 5c). To determine whether DNA repair correction improves chromosome stability over time, these cell lines were passaged three times per week over a 60-day period. We then used multi-color DNA FISH on metaphase spreads to evaluate how frequently WT and engineered cell lines acquired a *de novo* numerical or structural chromosome abnormality that deviates from the main karyotype. Whereas CHO-K1 WT acquired many numerical and structural variations (Fig. 5d,e), ATM^+ and ATM^+ $PRKDC^+$ cell lines exhibited significantly fewer structural chromosomal abnormalities (Fig. 5e), consistent with improved DNA repair and genome stability. In contrast, KU-60019-mediated inhibition of ATM as a control resulted in a considerable increase in the frequency of metaphases with a rearrangement (~75%), indicative of an accelerated accumulation of structural variants over the 60 day period (Fig. 5d,e).

In summary, while CHO cells carry a high mutational burden in DNA repair genes, correction of only a few of these mutations can yield considerable improvements in DSB repair fidelity and speed, as well as in karyotype stabilization. This indicates that DNA repair engineering could be an effective strategy to stabilize phenotypic properties of CHO cell lines, and most importantly the expression rates of recombinant protein.

Restoration of DNA repair improves titer stability in a producing CHO cell line

Based on the notion that genome instability can disrupt the maintenance of high protein titers in industrial biomanufacturing, we reasoned that genome stabilization could counteract this problem by slowing the loss of transgene copies caused by chromosome instability. Thus, the results above support the idea that correction of DNA repair genes could help stabilize protein titers. We applied this strategy in CHO-SEAP, an adherent cell line expressing human secreted alkaline phosphatase (SEAP) (Hayduk & Lee, 2005). CHO-SEAP carries identical SNVs as CHO-K1 in *ATM*, *PRKDC*, *XRCC6*, and *LIG4*, with the exception that D869N does not occur in CHO-SEAP. Guided by our results for non-producing CHO-K1, we generated CHO-SEAP cell lines with corrected DNA repair genes, using SNV reversal or lentiviral integration of cloned Chinese hamster cDNAs, as described above. Surprisingly, in contrast to our observations in CHO-K1, chemical inhibition of ATM did not lead to an increase but instead a slight decrease in distal end-joining in CHO-SEAP (Fig. 6a; Suppl. Fig. 2g). This led us to conclude that the effect of *ATM* correction might be cell-line specific (see Discussion), and we focused on *XRCC6* and *LIG4* instead. These cell lines, CHO-SEAP CMV::XRCC6 and CMV::LIG4, showed significantly improved DSB repair fidelity, evident as a reduction in distal end-joining in the EJ5-GFP assay (Fig. 6a; Suppl. Fig. 2g,h), and lower levels of γH2AX foci during recovery after irradiation, with *XRCC6* overexpression causing a more pronounced effect (Fig. 6b). Curiously, reversal of *ATM* R2830H in CHO-SEAP CMV::XRCC6 did not yield further improvements, but

instead caused an increase in distal end-joining compared to CMV::XRCC6 (Fig. 6a), consistent with our observations for ATM inhibition.

Because both *XRCC6* and *LIG4* overexpression had beneficial effects on DSB repair in CHO-SEAP, similarly to our results in CHO-K1, we wondered whether these engineered cell lines would maintain higher copy numbers of SEAP transgene in long-term culture. We grew clonal populations of CHO-SEAP WT, CMV::XRCC6, and CMV::LIG4 cell lines and cultivated them in parallel over a 74-day period after which both SEAP titer and SEAP transgene copy number was compared to the start date of the experiment. Prior to the start of the experiment, cells were cultured in 5 μM methotrexate (MTX) for 1 week to select for high SEAP copy number amplification, after which MTX was absent from the growth medium. MTX is a competitive inhibitor of dihydrofolate reductase, an essential metabolic enzyme, which is co-expressed with the transgenic SEAP locus (Hayduk & Lee, 2005). In all cell lines, SEAP titer retention (i.e. titer at day 74, relative to day 0) correlated significantly with SEAP copy number retention (Fig. 6c), highlighting the importance of genome stability as a determinant of cell line productivity. CHO-SEAP WT cells retained about 75% of the initial SEAP titer on average, but a considerable fraction of clones showed dramatic titer decline to as low as 10% (Fig. 6c). Clones overexpressing *LIG4* showed improvements in stability of both titer and copy number although results were only marginally statistically significant (Fig. 6c,d). *XRCC6* overexpression, however, gained significant stabilization of both SEAP titer and copy number, retaining about 90% SEAP titer on average, and showing fewer cases of dramatic titer decline, with the minimum titer retention at 30% (Fig. 6c,d).

Thus, correction of DNA repair genes, in particular of the NHEJ pathway in CHO cells, improves DSB repair, genome stability, and transgene copy number and protein titers. Consequently, these results point to a novel rational cell engineering strategy to enhance protein expression in CHO cells.

Discussion

The CHO-based platform in biologics manufacturing is well-established, but an important unsolved problem is genome instability, which causes a population derived from a clonal cell to display genetic heterogeneity (Dahodwala & Lee, 2019). Apart from a few studies identifying impaired repair pathways (Goth-Goldstein, 1980; Shen, Zdzienicka, Mohrenweiser, Thompson, & Thelen, 1998), this is the first report documenting the full extent of mutations affecting DNA repair genes across CHO cell lines. Moreover, while reactivation of silenced DNA repair genes has been successfully implemented before in CHO (Jeggo & Holliday, 1986), restoration of DNA repair capacity has not yet been systematically explored to mitigate genome instability in cell line development. This study is the first report to show that restoring DNA repair function through genome editing ameliorates genome stability in CHO. Furthermore, we showed that despite the high mutational burden in DNA repair genes, restoration of just a single gene can yield significant improvements in genome integrity. This makes DNA repair restoration a powerful and feasible novel addition to the cell line engineering toolbox. Our dataset of mutated DNA repair genes opens up a plethora of options for future studies, targeting single genes or

combinations of genes to develop cell lines for biopharmaceutical manufacturing. While effective alternative approaches have been described to increase CHO cell productivity, such as overexpression of key metabolic genes (Berger et al., 2020), suppression of apoptosis (Xiong et al., 2019), or design of engineered promoters (Nguyen et al., 2019), restoration of DNA repair tackles the mechanistic root of genome instability and could enable long-lasting stability improvements. Beyond protein expression, restoration of DNA repair genes could prove effective in other aspects of cell line engineering, for example by improving rates of targeted gene integration or gene correction in CHO, relying on homology-directed repair (Bosshard, Duroy, & Mermod, 2019). Also, while the effects of gene correction may vary across cell lines (see below), the concept presented in this study applies likewise to other CHO cell lines and would be of particular interest for suspension cell lines used in protein manufacturing (e.g., CHO-S). It also likely applies to other immortalized mammalian cell lines beyond CHO.

As shown here, improvement of DSB repair capacity appears to occur in an incremental fashion when combinations of DNA repair genes are being restored, provided these genes work synergistically, as *ATM* and *PRKDC* do in CHO-K1, for example, evident from the gradual improvement in repair capacity when combining both gene corrections. However, selecting the optimal gene targets, or combination of gene targets, for DNA repair correction is far from trivial. While the methods of target prioritization used in this study, such as curated (human) DNA repair ontology databases and phylogeny-based deleteriousness scoring, proved effective, they only represent guidelines rather than definitive predictors, and target selection must be carried out on a case-by-case basis. For example, *XRCC6* overexpression yielded very positive outcomes on stabilization in CHO-SEAP although the SNV (Q606H) occurs only heterozygously (and only received a moderate PROVEAN score). On the other hand, expression level analysis of DNA repair genes through RNA-sequencing might be convenient, given its routine use in CHO cell line engineering, but our report shows that selection of targets based on low expression levels could also be potentially misleading since, for example, *ATM* appears to be upregulated in CHO-K1 WT despite its deleterious SNV. Possibly, the R2830H mutation could stabilize the RNA, or overexpression of the hypomorphic protein could be the result of a feedback mechanism since, interestingly, chemical inhibition of ATM protein also resulted in strong upregulation of *ATM* transcript (Fig. 4e). ATM promoter activity has been found to be regulated by DNA damage stimuli (Gueven et al., 2006). Thus, the higher basal level of DSBs in CHO (in particular after ATM inhibition) could drive ATM upregulation at the transcriptional level.

Our study also revealed that correction of the same genes can yield different outcomes in different cell lines. Other studies reported similar unexpected effects upon inhibition of DNA repair genes in CHO, when it was found, for example, that *MRE11* knock-down can increase HR efficiency, despite its essential role in HR (Bosshard et al., 2019). Given the divergent genomes of different CHO cell lineages (Lewis et al., 2013) and the complex, intertwined nature of the mammalian DSB repair cascade, results from one cell line may not necessarily apply likewise to others. In mammals, DSB repair follows a “decision tree” (Scully, Panday, Elango, & Willis, 2019) where pathway choice is largely determined by the severity of the DNA lesion. In particular, the core NHEJ pathway can act independently of ATM, but ATM plays an essential role in initiating repair of lesions requiring a higher degree of pre-

processing and more advanced repair pathways, such as homology-directed repair (HDR), alternative end-joining (aEJ), or the Fanconi anemia (FA) pathway (Scully et al., 2019). Clearly, for this repair "choice" to be effective, genes in these pathways downstream of ATM need to be functional. It is possible that CHO-K1 has retained higher functionality of these pathways (and thus benefits from ATM correction), while in CHO-SEAP these pathways might be so heavily affected by mutations that ATM activation yields no net improvement in DSB repair, or even yields a net negative effect by shifting repair away from the (possibly less affected) core NHEJ pathway. Our data show slightly more deleterious SNVs in HDR or FA pathways in CHO-SEAP (a DXB11 derivative) compared to CHO-K1 (Suppl. Fig. 1), but a deeper characterization of the repair capacity of these DNA repair pathways in different CHO cell lines would be necessary and would constitute a valuable resource to enable engineering strategies tailored to specific CHO cell lines. In addition, developing a high-throughput screening protocol would be powerful for identifying optimal gene targets for correction. The EJ5-GFP cell line described here could serve as an excellent discovery tool for flow-cytometry screens when combined with gene activation libraries (CRISPRa) (Karottki et al., 2020) or base-editors (Hanna et al., 2021). We point out, however, that the EJ5-GFP assay is designed to measure "total" NHEJ (Bennardo et al., 2008). Additional end-joining assays would be required to differentiate between sub-classes of NHEJ, e.g. canonical and alternative end-joining (Gunn & Stark M., 2009).

Finally, it is important to note that engineering of high-producer CHO lines through transgene copy number amplification (e.g. through selection with increasing concentrations of methotrexate), actually requires a plastic genome and thus implicitly relies on genome instability. Isolation of high-producing clones might thus require a harsher selection phase in a stabilized genome. However, the expression of corrected DNA repair genes could also be controlled by inducible promoters that would be turned on only after a high producer has been isolated. Alternatively, another implementation would be to utilize dedicated transcriptional "safe-harbor" hotspots for transgene expression (Chi, Zheng, Jiang, Chen-Tsai, & Kong, 2019), rather than random integration sites. Either strategy would enable the engineering of a DNA repair-enhanced parental cell line from which specific expression hosts could be derived.

In conclusion, this study provides the a thorough documentation of the deficient DNA repair machinery in CHO cells, and constitutes a proof-of-concept of the notion of DNA repair restoration as a powerful method for cell line development in industrial protein expression to ensure stable upstream manufacturing processes.

Materials & Methods

Whole-genome sequencing analysis

Whole-genome sequencing data of 11 CHO cell lines (Lewis et al., 2013; Rupp et al., 2018) (Table 1) were downloaded from Genbank (https://www.ncbi.nlm.nih.gov/sra), together with the Chinese hamster genome (Lewis et al., 2013). Raw sequencing reads were pre-processed using FastQC (Andrews, 2010) for quality control and Trimmomatic (Bolger, Lohse, & Usadel, 2014) to remove low-quality base pairs and adapters. For mapping, the RefSeq CHO genome assembly GCF_000223151.1 was used. Transcript sequences of DNA repair

genes (Wood et al., 2005) extracted from the reference genome assembly are known to have a low proportion of mismatches compared to the higher quality RefSeq transcripts. Thus, to ensure the reference genome has the correct sequences of DNA repair genes, RefSeq transcript sequences of DNA repair genes were merged into one scaffold separated by 500 letter N gaps. In cases of multiple isoforms, the transcript with the longest coding sequence was selected. Genomic coding regions of corresponding DNA repair genes were masked. Then the sequencing reads were mapped against the new reference genome using BWA (Li & Durbin, 2009). Non-synonymous SNVs and InDels (including heterozygosity) were identified using the *gatk3*.5 software package (McKenna, Hanna, Banks, & DePristo, 2010) using standard parameters and annotated using SnpEff (Cingolani, Platts, Wang, & Lu, 2012). SnpSift (Cingolani, Patel, et al., 2012) was used to filter genes with ontologies related to DNA repair (Wood et al., 2005). Deleteriousness of mutations was predicted based on the PROVEAN score (Choi et al., 2012). Protein-protein interaction networks were generated with STRING (Szklarczyk et al., 2015).

RNA sequencing and differential expression analysis

RNA was extracted in from triplicate flasks (RNeasy, Qiagen) and stored at −80°C before sequencing. RNA-sequencing data quality was assessed using FastQC (Andrews, 2010). Adapter sequences and low-quality bases were trimmed using Trimmomatic (Bolger et al., 2014). Reads were then quasi-mapped to the CHO-K1 genome (Lewis et al., 2013; Rupp et al., 2018) and quantified with Salmon (Patro, Duggal, Love, Irizarry, & Kingsford, 2017) using default parameters. Raw gene expression data were then normalized and analyzed for differential expression using DESeq2 (Love, Anders, & Huber, 2014). After Benjamini-Hochberg FDR correction, genes with adjusted p-values less than 0.05 and fold changes greater than $\log_2(1.5)$ were considered differentially expressed.

Cloning of Chinese hamster genes

Chinese hamster (*Cricetulus griseus*) tissue (lung, liver) was a gift from George Yerganian. Fibroblasts were extracted from lung tissue as previously described (Seluanov, Vaidya, & Gorbunova, 2010). RNA extraction (RNeasy, Qiagen), cDNA synthesis (SuperScriptIII, Invitrogen), and cDNA purification with ethanol precipitation were carried out using standard protocols. 1 μL purified total cDNA (100–200 ng) per reaction was used to amplify Chinese hamster genes through high-fidelity PCR (Q5, New England Biolabs). For CHO-K1, DNA repair genes were cloned into a pLJM1 backbone (Addgene #19319) using restriction/ligation (New England Biolabs). For CHO-SEAP a vector backbone fragment was obtained by PCR amplification from plasmid pcDNA3.1/zeo(+) (ThermoFisher), and plasmids expressing DNA repair genes were constructed via Gibson assembly of gene fragment(s) and the vector fragment following the manufacturer's instructions (New England Biolabs).

Cell culture and cell line generation

CHO-K1 cells (ATCC: CCL-61) and CHO-SEAP cells (Hayduk & Lee, 2005) were cultured in F-12K medium (Gibco), or Iscove's Modified Dulbecco's Medium (IMDM, Gibco), respectively, supplemented with 10% (v/v) fetal bovine serum (FBS, Corning) and 1% (v/v) penicillin/streptomycin (Gibco) at 37°C under an atmosphere of 5% CO_2. Cells

were passaged every 2–3 days. The EJ5-GFP reporter cassette (Addgene #44026) was genomically integrated through XhoI-linearization, electroporation (Neon, ThermoFisher) and selection with puromycin (7 μg/mL, Omega Scientific) and hygromycin (300 μg/mL, Omega Scientific) for two weeks, followed by isolation of clonal populations. ATM^+ and ATM^+ $PRKDC^+$ cell lines were generated through electroporation (Neon, ThermoFisher) of a Cas9:tracrRNA:sgRNA ribonucleoprotein particle (Integrated DNA Technologies), targeting R2830H in *ATM*, or D1641N in *PRKDC*, with homology donor oligos. Clones were screened by PCR and restriction for the presence of a novel TaqI site (*ATM*) or BamHI site (*PRKDC*), caused by introduction of silent SNVs into the homology donor sequences, and verified by Sanger sequencing (Eton Biosciences). Sanger sequencing data was deconvoluted using the ICE Analysis Tool (Synthego). CMV::XRCC6 and CMV::LIG4 cell lines were generated by transfection (Lipofectamine LTX, Invitrogen) of the corresponding expression plasmids and subsequent selection with puromycin (7 μg/mL, Omega Scientific) or Zeocin (200 μg/mL, Gibco), respectively, followed by isolation of clonal populations. All cells were maintained under combined puromycin/hygromycin selection throughout the experiments to avoid loss of the EJ5-GFP insertion. ATM was inhibited with KU-60019 (3 μM, Selleckchem).

Irradiation

For immunofluorescence analysis and comet assay, irradiation of cells was carried out on a X-RSD 320 Precision X-Ray machine (2.87 Gy/min). After exposure, cells were allowed to recover for the indicated time periods at 37°C with 5% CO_2.

Immunofluorescence analysis

Cells were seeded on chambered slides (MatTek) and allowed to adhere overnight. After irradiation and the indicated recovery time, cells were fixated in 4% paraformaldehyde (ThermoFisher) for 10 min, washed in PBS (Gibco) for 2 min, and permeabilized with 0.5% Triton-X (Amresco) for 5 min, followed by washing for 5 min in PBS. After blocking with 5% goat-serum (Millipore-Sigma) for 1h, cells were incubated in anti-γH2AX antibody (Cell Signaling Technology, Rabbit #9718) at 1:1000 dilution for 1h, washed three times in PBS-T (=0.1% Triton-X in PBS) for 5 min, and incubated with DyLight 488 goat-anti-rabbit (ThermoFisher) at 1:1000 dilution for 1h in the dark. After three washes in PBS-T for 5 min, cells were mounted in anti-fade mounting medium containing DAPI (Vectashield Vibrance, Vector Laboratories). Samples were analyzed on a SP8 confocal microscope (Leica) with gain/offset settings kept identical between samples.

Comet assays

Neutral comet assays were carried out following the manufacturer's protocol (Trevigen) with 30 min electrophoresis at 1 V/cm in TBE-buffer. Nuclei were stained with Vista DNA Green at 1:10000 dilution (Abcam). Slides were analyzed on an Axio Imager 2 (Zeiss) with gain/offset settings kept identical between samples.

Image processing

Immunofluorescence images were processed with MATLAB (The MathWorks) using custom scripts. Briefly, DAPI images were first segmented into binary images to identify nuclei, and then used to mask the corresponding Alexa488 images to extract nuclear area, perimeter, total fluorescence intensity, and circularity (= 4*area*pi / perimeter^2). Comet assay images were cropped into segments of 1–3 nuclei per image and then analyzed using the OpenComet plug-in (www.cometbio.org/index.html) for ImageJ (NIH) (background subtraction enabled with head detection by “brightest region”).

EJ5-GFP flow cytometry assays

The Cas9-sgEJ5 plasmid was constructed by ligation of two sgRNAs, targeting the EJ5-GFP cassette (Addgene #44026), into pX333 (Addgene #64073), and subsequent DrdI/KpnI-subcloning of the entire dual sgRNA expression cassette into pSpCas9(BB)-2A-miRFP670 (Addgene #91854). sgRNA editing efficacy and specificity were tested in vitro by incubating PCR products of both targets sites with a Cas9-sgRNA ribonucleoprotein particle (Integrated DNA Technologies) at 37°C for 15 min, followed by incubating with Proteinase K (Integrated DNA Technologies) for 10 min at room temperature (Fig. 4c). To run the EJ5-GFP assay, cells were transfected with 1 μg of Cas9-sgEJ5 plasmid (Lipofectamine LTX, Invitrogen; 12-well format) in triplicates. After 30 h cells were detached (Trypsin 0.05%, Gibco), resuspended in 250 μL DPBS (Gibco), and analyzed on a Canto II flow cytometer (BD Biosciences). Untransfected cells served as negative control to define proper gates in the APC and FITC channels for miRFP and GFP expression, respectively. Distal end-joining events were identified through Boolean gating, as shown in Fig. 4b. At least three different replicates from independent Cas9-sgEJ5 transfections were pooled for each genotype. Fluorescence intensities were log10-transformed, and the median of the pooled WT sample was subtracted form all datapoints for normalization. The full distributions of log10 fluorescence in all samples after gating is shown in Suppl. Fig. 2a–c, g–h. All flow cytometry data was analyzed in FlowJo (BD Biosciences), followed by statistical analysis in Prism (GraphPad) or R (The R Project for Statistical Computing).

SEAP quantification

Selected cell lines were seeded at 1×10^5 cells/mL and subcultured for four passages with 5 μM MTX in 24-well plates. Then, cells were subcultured every three or four days for 74 days without MTX selection. Cell lines expressing *XRCC6* and *LIG4* were continuously maintained in media with 200 μg/mL Zeocin (Gibco). Cell lines at Day 0 and Day 74 were seeded at 1×10^5 cells/mL in 6-well plates containing 3 mL growth media per well without MTX. When the cell density reached ~80% confluency, culture supernatant samples were collected for SEAP quantification (PhosphaLight, AppliedBiosystems). Alkaline phosphatase from human placenta (Millipore-Sigma) was assayed and used as a standard. Gene copy number was determined by Real-Time PCR using TaqMan Fast Advanced Master Mix (ThermoFisher) performed on a 7500 fast system (Applied Biosystems) according to the manufacturer’s protocol. PCR primers and TaqMan probes were designed and synthesized by IDT (Integrated DNA Technologies). Genomic DNA was extracted from Day 0 and Day 74 cell samples (DNeasy, Qiagen). The diluted genomic DNA of one

CHO-SEAP clone (400 ng, 200 ng, 100 ng, 50 ng, 25 ng, 0 ng) was used to make a standard curve. Relative SEAP gene copy number in each cell sample was quantified in duplicate using 100 ng genomic DNA as a template with β-Actin as a reference.

Quantitative PCR

Quantitative PCR was performed on a CFX96 Touch Deep Well system (BioRad) (melting/annealing/elongation: 20s. 40 cycles) with 50 ng purified cDNA per sample and 4 technical replicates per reaction. GAPDH expression was used for reference. Analysis was performed following standard protocols (Pfaffl, 2001) with pairwise comparisons to each reference replicate. Each experiment was repeated with cDNA from a different batch for validation.

Immunoblotting analysis

Cells were irradiated and immunoblotted as previously described (Lu, Saha, Beckmann, Hendrickson, & Davis, 2019). Specificially, cells were irradiated with γ-rays generated by a Mark 1 137Cs irradiator (JL Shepherd and Associates) with a total dose of 10 Gy. Following mock treatment or irradiation, the cells were allowed to recover for the time indicated in the figure legends, washed three times with ice-cold phosphate-buffered saline (PBS), and then whole-cell extracts were generated by resuspending the cells in 1× radioimmunoprecipitation assay (RIPA) buffer (Cell Signaling Technology), sonicated, and then incubated on ice for 20 min. After centrifugation at 14,000 rpm at 4°C for 30 min to clear the lysate, the protein concentration of each sample was measured using the Pierce BCA Protein Assay kit (Thermo Fisher). Each sample (30 μg) was separated via sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and then transferred to a polyvinylidene difluoride (PVDF) membrane. The PVDF membrane was blocked using 5% non-fat milk and the membranes were then incubated with primary antibodies for 2 h at room temperature. The following commercial antibodies were used: anti-ATM (Abcam, ab109027), anti-ATM phospho-S1981 (Abcam, ab81292), anti-phospho-H2AX (S139) (EMD Millipore, 05–636), anti-γH2AX (Cell Signaling Technology, 7631), and anti-tubulin (Sigma, T5168). In-house produced antibodies include mouse monoclonal antibodies against DNA-PKcs (25–4). The membranes were washed four times with tris-buffered saline-Tween (TBS-T), then incubated with secondary antibodies at room temperature for 2 h. Secondary antibodies included anti-mouse IgG (HRP-linked) and anti-rabbit IgG (HRP-linked), which were purchased from Cell Signaling Technology (7026, 7074). Following the washing of the membranes four times with TBS-T, the immunoblots were reacted using Pierce™ enhanced chemiluminescent (ECL) Western Blotting Substrate (Thermo Fisher), exposed to Blue X-ray films (Light Labs), and then developed using a Protec X-ray film processor.

Karyotype analysis

Metaphase spreads were prepared as previously described (Vcelar et al., 2018). For multi-color DNA fluorescence *in situ* hybridization (FISH) using hamster-specific FISH probes (12XCHamster mFISH probe kit, MetaSystems), slides were denatured in 0.07N NaOH at room temperature for 1 minute and FISH probes were denatured at 75°C for 5 minutes. FISH probes were then applied to chromosome spreads, sealed with a coverslip, and incubated in a humidified chamber at 37°C overnight. Slides were then washed with 0.4×

SSC at 72°C for 2 minutes and 2× SSC, with 0.05% Tween-20 at room temperature for 1 minute. Slides were gently rinsed in water, air dried, and stained with DAPI. Images were acquired using a Zeiss Axio Imager Z2 equipped with a Metafer Slide Scanning System and analyzed using Isis (MetaSystems) software. For karyotypic analyses, the most abundant karyotype across samples was defined as the representative ("main") karyotype, and deviations from this karyotype were scored as a numerical alteration (whole-chromosomal aneuploidy) and/or structural alteration (e.g. inter-chromosomal rearrangement, visible deletion). Aberrant karyotypes (Fig. 5d,e) were defined as karyotypes showing at least one structural (or numerical) deviation from the main karyotype.

DNA oligos

Primers

EJ5-GFP Insertion	AGCCTCTGTTCCACATACACT / CCAGCCACCACCTTCTGATA
EJ5-GFP Spacer 5' site	CGTGCTGGTTATTGTGCTGT / GATGTGGAATGTGTGAGGCC
EJ5-GFP Spacer 3' site	AAGAACGAGATCAGCAGCCT / GCTGAACTTGTGGCCGTTTA
ATM R2830H	AGAGGTGTCCAGGCCAAGTT / GAGCTAACAATCAGCACGAACA
PRKDC D1641N	AGAACCAGTTGCTGTAGTCTTGT / CCTGTGTGGTGATGGTGCATA
Cloning of *C.gri. XRCC6 (NheI / EcoRI overhangs)*	TTATGCTAGCCCTTCTGTCCCTTTGGCTCG / TTATGAATTCTAAGTAGGTGGTCTGGCTGC
Cloning of *C.gri. LIG4 (AgeI / EcoRI overhangs)*	TTATACCGGTGTTTATGTTGGAGCCGGCTT / TTATGAATTCGGATAGCCTGCTGTTGTTCTG
Subcloning of sgRNA expression locus (pX333)	ACGACCTACACCGAACTGAG / AGGTCATGTACTGGGCACAA
SEAP copy number	TCGTATTTCATGTCTCCAGGC / AGAATCTGGTGCAGGAATGG Probe: 5'−6-FAM-CATGAGATG-ZEN-GGTCACAGACGGGT-3'-AbkFQ
β-Actin copy number	ACACCTTCTACAACGAGCTG / CCTGAATGGCTACGTACATGG Probe: 5'−6-FAM-TCTGGGTCA-ZEN-TCTTTTCACGGTTGGC-3'-ABkFQ
GAPDH qPCR	GCAAAGTCATCCCAGAGCTG / AAGGTGGAAGAGTGGGAGTC
XRCC6 qPCR	ATGGAGTCTGAGCAAGTGGT / TGCTTCTTCGGTCCACTCTT
LIG4 qPCR	AGGCTGAGCGTTCGGAGA / CATCCTTGCCTTCTCTTGGT
ATM qPCR	AGCTAGATGAATGTGCTCTCCG / TCCTAGGCCTCCCATCATTTTG
PRKDC qPCR	ACTCGCCATGAAACTCCTGA / CTGGTTACAAGAGGCCTGGA

All primers were designed using Primer3 (Koressaar & Remm, 2007).

sgRNAs

Targeting *ATM* R2830H	AGTTTTCCATTACTTCTGCA
Targeting *PRKDC* D1641N	TGGCCAGGCTCTTACAGCTG
sgEJ5_5'	AACAGGGTAATAATTCTACC
sgEJ5_3'	TAACAGGGTAATGGATCCAC

ssDNA Oligos

ATM R2830H homology donor	GTTTCTCAAACCAAACAGCTGGGTCCAAGAATTTTTCCATACAAAAATATCGAAAAACTGGTTCAAAGTTTTGGCAAATAGTC
PRKDC D1641N homology donor	CATTGCTCCTGCAGAGGAAAGGCAGTGCCTGCAATCATTGGATCCTAGCTGTAAGAGCCTGGCCAATGGACTCCTGGAGTTAG

Statistical analysis

Statistical tests and sample sizes are indicated in the figure legends. All tests were two-sided and carried out in Prism (GraphPad) or R (The R Project for Statistical Computing). The non-parametric Wilcoxon-Mann-Whitney test was the preferred test in order to avoid assumptions about data normality. Significance levels for all tests shown in the figures are: # $p<0.1$, * $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$, ns: $p\geq0.1$.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to express our gratitude to George Yerganian for sharing primary Chinese hamster tissue. We thank Bjørn Voldborg and Alexandra Hoffmeyer for sharing cell lines and for RNA sequencing services, and Nicole Borth for sharing metaphase preparation protocols. We thank Mojtaba Samoudi for assistance with fibroblast isolation. We thank Francisco Diaz, Jennifer Santini and Marcy Erb for technical assistance. We also thank Philip Gordts for constructive feedback, and John Ruano-Salguero and Jong Youn Baik for comments on the manuscript. This work was supported by NIGMS (R35 GM119850, NEL), the Novo Nordisk Foundation (NNF20SA0066621, NEL), NSF (grants 1736123 and 1412365, KHL), NIH (CA162804 and GM047251, AJD). Laser-scanning confocal microscopy was supported by NINDS P30 NS047101.

Grants supporting this work:

NIGMS R35 GM119850 (NEL)

Novo Nordisk Foundation NNF20SA0066621 (NEL)

NSF 1736123 and 1412365 (KHL)

NIH CA162804 and GM047251 (AJD)

NINDS P30 NS047101 (UCSD)

Data Availability

Any data generated or analyzed during this study, associated protocols, materials within the manuscript, and references to public databases are included in the article or are available from the corresponding author on request. Source data are provided with this paper. Whole-genome sequencing data from CHO-S (DTU) and CHO-Z are property of the Technical University of Denmark and are available upon request.

References

Aguilera A, & García-Muse T (2013). Causes of genome instability. Annual Review of Genetics, 47, 1–32. 10.1146/annurev-genet-111212-133232

Andrews S (2010). fastQC: A quality control tool for high throughput sequence data. Retrieved from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Baik JY, & Lee KH (2016). A framework to quantify karyotype variation associated with CHO production instability. Biotechnology and Bioengineering, 1–24. 10.1002/bit.26231

Bailey LA, Hatton D, Field R, & Dickson AJ (2012). Determination of Chinese hamster ovary cell line stability and recombinant antibody expression during long-term culture. Biotechnology and Bioengineering, 109(8), 2093–2103. 10.1002/bit.24485 [PubMed: 22896849]

Bandyopadhyay AA, O'Brien SA, Zhao L, Fu HY, Vishwanathan N, & Hu WS (2019). Recurring genomic structural variation leads to clonal instability and loss of productivity. Biotechnology and Bioengineering, 116(1), 41–53. 10.1002/bit.26823 [PubMed: 30144379]

Barnes LM, Bentley CM, & Dickson AJ (2003). Stability of protein production from recombinant mammalian cells. Biotechnology and Bioengineering, 81(6), 631–639. 10.1002/bit.10517 [PubMed: 12529877]

Beckmann TF, Krämer O, Klausing S, Heinrich C, Thüte T, Büntemeyer H, … Noll T (2012). Effects of high passage cultivation on CHO cells: A global analysis. Applied Microbiology and Biotechnology, 94(3), 659–671. 10.1007/s00253-011-3806-1 [PubMed: 22331235]

Bennardo N, Cheng A, Huang N, & Stark JM (2008). Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. PLoS Genetics, 4(6). 10.1371/journal.pgen.1000110

Bennardo N, & Stark JM (2010). ATM limits incorrect end utilization during non-homologous end joining of multiple chromosome breaks. PLoS Genetics, 6(11), 16–18. 10.1371/journal.pgen.1001194

Berger A, Le Fourn V, Masternak J, Regamey A, Bodenmann I, Girod PA, & Mermod N (2020). Overexpression of transcription factor Foxa1 and target genes remediate therapeutic protein production bottlenecks in Chinese hamster ovary cells. Biotechnology and Bioengineering, 117(4), 1101–1116. 10.1002/bit.27274 [PubMed: 31956982]

Bolger AM, Lohse M, & Usadel B (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics, 30(15), 2114–2120. 10.1093/bioinformatics/btu170 [PubMed: 24695404]

Bosshard S, Duroy PO, & Mermod N (2019). A role for alternative end-joining factors in homologous recombination and genome editing in Chinese hamster ovary cells. DNA Repair, 82(August), 102691. 10.1016/j.dnarep.2019.102691 [PubMed: 31476574]

Chang HHY, Pannunzio NR, Adachi N, & Lieber MR (2017). Non-homologous DNA end joining and alternative pathways to double-strand break repair. Nature Reviews Molecular Cell Biology, 18(8), 495–506. 10.1038/nrm.2017.48 [PubMed: 28512351]

Chi X, Zheng Q, Jiang R, Chen-Tsai RY, & Kong LJ (2019). A system for site-specific integration of transgenes in mammalian cells. PLoS ONE, 14(7), 1–14. 10.1371/journal.pone.0219842

Choi Y, Sims GE, Murphy S, Miller JR, & Chan AP (2012). Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS ONE, 7(10). 10.1371/journal.pone.0046688

Chusainow J, Yang YS, Yeo JHM, Ton PC, Asvadi P, Wong NSC, & Yap MGS (2009). A study of monoclonal antibody-producing CHO cell lines: What makes a stable high producer? Biotechnology and Bioengineering, 102(4), 1182–1196. 10.1002/bit.22158 [PubMed: 18979540]

Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, & Lu X (2012). Using Drosophila melanogaster as a model for genotoxic chemical mutational studies with a new program, SnpSift. Frontiers in Genetics, 3(MAR), 1–9. 10.3389/fgene.2012.00035 [PubMed: 22303408]

Cingolani P, Platts A, Wang LL, & Lu X (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly, 6(2), 1–13. 10.4161/fly.19695

Dahodwala H, & Lee KH (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(August 2018), 128–137. 10.1016/j.copbio.2019.01.011 [PubMed: 30826670]

Davis AJ, & Chen DJ (2013). DNA double strand break repair via non-homologous end-joining. Translational Cancer Research, 2(3), 130–143. 10.3978/j.issn.2218-676X.2013.04.02 [PubMed: 24000320]

Deaven LL, & Petersen DF (1973). The chromosomes of CHO, an aneuploid Chinese hamster cell line: G-band, C-band, and autoradiographic analyses. Chromosoma, 41(2), 129–144. 10.1007/BF00319690 [PubMed: 4120885]

Fann CH, Guirgis F, Chen G, Lao MS, & Piret JM (2000). Limitations to the amplification and stability of human tissue-type plasminogen activator expression by Chinese hamster ovary cells. Biotechnology and Bioengineering, 69(2), 204–212. 10.1002/(SICI)1097-0290(20000720)69:2<204::AID-BIT9>3.0.CO;2-Z [PubMed: 10861399]

Feichtinger J, Hernández I, Fischer C, Hanscho M, Auer N, Hackl M, … Borth N (2016). Comprehensive genome and epigenome characterization of CHO cells in response to evolutionary pressures and over time. Biotechnology and Bioengineering, 113(10), 2241–2253. 10.1002/bit.25990 [PubMed: 27072894]

Gaidukov L, Wroblewska L, Teague B, Nelson T, Zhang X, Liu Y, … Weiss R (2018). A multi-landing pad DNA integration platform for mammalian cell engineering. Nucleic Acids Research, 46(8), 4072–4086. 10.1093/nar/gky216 [PubMed: 29617873]

Goth-Goldstein R (1980). Inability of Chinese Hamster Ovary Cells to Excise O6-Alkylguanine. Cancer Research, 40(7), 2623–2624. [PubMed: 7388815]

Gueven N, Fukao T, Luff J, Paterson C, Kay G, Kondo N, & Lavin MF (2006). Regulation of the Atm promoter in vivo. Genes Chromosomes and Cancer, 45(1), 61–71. 10.1002/gcc.20267 [PubMed: 16180236]

Gunn A, & Stark M,J (2009). I-SceI-Based Assays to Examine Distinct Repair Outcomes of Mammalian Chromosomal Double Strand Breaks. In DNA Repair Protocols (Vol. 531, p. 588). 10.1006/cbir.1999.0447

Hamaker NK, & Lee KH (2018). Site-specific integration ushers in a new era of precise CHO cell line engineering. Current Opinion in Chemical Engineering, 22, 152–160. 10.1016/j.coche.2018.09.011 [PubMed: 31086757]

Hammill L, Welles J, & Carson GR (2000). The gel microdrop secretion assay: Identification of a low productivity subpopulation arising during the production of human antibody in CHO cells. Cytotechnology, 34(1–2), 27–37. 10.1023/A:1008186113245 [PubMed: 19003378]

Hanna RE, Hegde M, Fagre CR, DeWeirdt PC, Sangree AK, Szegletes Z, … Doench JG (2021). Massively parallel assessment of human variants with base editor screens. Cell, 184(4), 1064–1080.e20. 10.1016/j.cell.2021.01.012 [PubMed: 33606977]

Hayduk EJ, & Lee KH (2005). Cytochalasin D can improve heterologous protein productivity in adherent Chinese hamster ovary cells. Biotechnology and Bioengineering, 90(3), 354–364. 10.1002/bit.20438 [PubMed: 15772946]

Hilliard W, & Lee KH (2021). Systematic identification of safe harbor regions in the CHO genome through a comprehensive epigenome analysis. Biotechnology and Bioengineering, 118(2), 659–675. 10.1002/bit.27599 [PubMed: 33049068]

Hilliard W, MacDonald ML, & Lee KH (2020). Chromosome-scale scaffolds for the Chinese hamster reference genome assembly to facilitate the study of the CHO epigenome. Biotechnology and Bioengineering, 117(8), 2331–2339. 10.1002/bit.27432 [PubMed: 32410221]

Jeggo PA, & Holliday R (1986). Azacytidine-induced reactivation of a DNA repair gene in Chinese hamster ovary cells. Molecular and Cellular Biology, 6(8), 2944–2949. 10.1128/mcb.6.8.2944 [PubMed: 2431295]

Jensen RB, & Rothenberg E (2020). Preserving genome integrity in human cells via DNA double-strand break repair. Molecular Biology of the Cell, 31, 859–865. 10.1091/mbc.E18-10-0668 [PubMed: 32286930]

Kaas CS, Kristensen C, Betenbaugh MJ, & Andersen MR (2015). Sequencing the CHO DXB11 genome reveals regional variations in genomic stability and haploidy. BMC Genomics, 16(1), 1–9. 10.1186/s12864-015-1391-x [PubMed: 25553907]

Karottki K. J. la C., Hefzi H, Xiong K, Shamie I, Hansen AH, Li S, … Lewis NE (2020). Awakening dormant glycosyltransferases in CHO cells with CRISPRa. Biotechnology and Bioengineering, 117(2), 593–598. 10.1002/bit.27199 [PubMed: 31631317]

Khanna KK, & Jackson SP (2001). DNA double-strand breaks: signaling, repair and the cancer connection. Nature Genetics, 27(3), 247–254. 10.1038/85798 [PubMed: 11242102]

Kildegaard HF, Baycin-Hizal D, Lewis NE, & Betenbaugh MJ (2013). The emerging CHO systems biology era: harnessing the 'omics revolution for biotechnology. Current Opinion in Biotechnology, 24(6), 1102–1107. 10.1016/j.copbio.2013.02.007 [PubMed: 23523260]

Kim M, O'Callaghan PM, Droms KA, & James DC (2011). A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. Biotechnology and Bioengineering, 108(10), 2434–2446. 10.1002/bit.23189 [PubMed: 21538334]

Kim SJ, Kim NS, Ryu CJ, Hong HJ, & Lee GM (1998). Characterization of Chimeric Antibody Producing CHO Cells in the Course of Dihydrofolate Reductase-Mediated Gene Amplification and Their Stability in the Absence of Selective Pressure. Biotechnology and Bioengineering, 58(1).

Koressaar T, & Remm M (2007). Enhancements and modifications of primer design program Primer3. Bioinformatics, 23(10), 1289–1291. 10.1093/bioinformatics/btm091 [PubMed: 17379693]

Lee JS, Grav LM, Lewis NE, & Kildegaard HF (2015). CRISPR/Cas9-mediated genome engineering of CHO cell factories: Application and perspectives. Biotechnology Journal, 10(7), 979–994. 10.1002/biot.201500082 [PubMed: 26058577]

Lee JS, Kallehauge TB, Pedersen LE, & Kildegaard HF (2015). Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. Scientific Reports, 1–11. 10.1038/srep08572

Lee KH, 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. 10.1007/s00253-013-4923-9 [PubMed: 23615744]

Lewis NE, Liu X, Li Y, Nagarajan H, Yerganian G, O'Brien E, … Palsson BO (2013). Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. Nature Biotechnology, 31(8), 759–765. 10.1038/nbt.2624

Li H, & Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25(14), 1754–1760. 10.1093/bioinformatics/btp324 [PubMed: 19451168]

Love M, Anders S, & Huber W (2014). Differential analysis of count data - the DESeq2 package. Genome Biology (Vol. 15). 10.1186/s13059-014-0550-8

Lu H, Saha J, Beckmann PJ, Hendrickson EA, & Davis AJ (2019). DNA-PKcs promotes chromatin decondensation to facilitate initiation of the DNA damage response. Nucleic Acids Research, 47(18), 9467–9479. 10.1093/nar/gkz694 [PubMed: 31396623]

Matsuyama R, Yamano N, Kawamura N, & Omasa T (2017). Lengthening of high-yield production levels of monoclonal antibody-producing Chinese hamster ovary cells by downregulation of breast cancer 1. Journal of Bioscience and Bioengineering, 123(3), 382–389. 10.1016/j.jbiosc.2016.09.006 [PubMed: 27742176]

McKenna A, Hanna M, Banks E, & DePristo M (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20(1), 1297–1303. 10.1101/gr.107524.110.20 [PubMed: 20644199]

Moritz B, Woltering L, Becker PB, & Göpfert U (2016). High levels of histone H3 acetylation at the CMV promoter are predictive of stable expression in Chinese hamster ovary cells. Biotechnology Progress, 32(3), 776–786. 10.1002/btpr.2271 [PubMed: 27073121]

Mutskov V, & Felsenfeld G (2004). Silencing of transgene transcription precedes methylation of promoter DNA and histone H3 lysine 9. EMBO Journal, 23(1), 138–149. 10.1038/sj.emboj.7600013

Nguyen LN, Baumann M, Dhiman H, Marx N, Schmieder V, Hussein M, … Borth N (2019). Novel Promoters Derived from Chinese Hamster Ovary Cells via In Silico and In Vitro Analysis. Biotechnology Journal, 14(11). 10.1002/biot.201900125

Olive PL, & Banáth JP (2006). The comet assay: A method to measure DNA damage in individual cells. Nature Protocols, 1(1), 23–29. 10.1038/nprot.2006.5 [PubMed: 17406208]

Paredes V, Park JS, Jeong Y, Yoon J, & Baek K (2013). Unstable expression of recombinant antibody during long-term culture of CHO cells is accompanied by histone H3 hypoacetylation. Biotechnology Letters, 35(7), 987–993. 10.1007/s10529-013-1168-8 [PubMed: 23468139]

Patro R, Duggal G, Love MI, Irizarry RA, & Kingsford C (2017). Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods, 14(4), 417–419. 10.1038/nmeth.4197 [PubMed: 28263959]

Pfaffl M (2001). A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acid Research, 29(9), 2069–2082. 10.1111/j.1365-2966.2012.21196.x

Rogakou EP, Boon C, Redon C, & Bonner WM (1999). Megabase chromatin domains involved in DNA double-strand breaks in vivo. Journal of Cell Biology, 146(5), 905–915. 10.1083/jcb.146.5.905

Rouiller Y, Kleuser B, Toso E, Palinksy W, Rossi M, Rossatto P, … Broly H (2015). Reciprocal Translocation Observed in End-of-Production Cells of a Commercial CHO-Based Process. PDA Journal of Pharmaceutical Science and Technology, 69(4), 540 LP–552. 10.5731/pdajpst.2015.01063 [PubMed: 26242789]

Rupp O, MacDonald ML, Li S, Dhiman H, Polson S, Griep S, … Lee KH (2018). A reference genome of the Chinese hamster based on a hybrid assembly strategy. Biotechnology and Bioengineering, 115(8), 2087–2100. 10.1002/bit.26722 [PubMed: 29704459]

Scully R, Panday A, Elango R, & Willis NA (2019). DNA double-strand break repair-pathway choice in somatic mammalian cells. Nature Reviews Molecular Cell Biology, 20(11), 698–714. 10.1038/s41580-019-0152-0 [PubMed: 31263220]

Seluanov A, Vaidya A, & Gorbunova V (2010). Establishing primary adult fibroblast cultures from rodents. Journal of Visualized Experiments, (44), 1–4. 10.3791/2033 [PubMed: 20164822]

Shen MR, Zdzienicka MZ, Mohrenweiser H, Thompson LH, & Thelen MP (1998). Mutations in hamster single-strand break repair gene XRCC1 causing defective DNA repair. Nucleic Acids Research, 26(4), 1032–1037. [PubMed: 9461464]

Shiloh Y (2006). The ATM-mediated DNA-damage response: taking shape. Trends in Biochemical Sciences, 31(7), 402–410. 10.1016/j.tibs.2006.05.004 [PubMed: 16774833]

Stolfa G, Smonskey MT, Boniface R, Hachmann AB, Gulde P, Joshi AD, … Campbell A (2018). CHO-Omics Review: The Impact of Current and Emerging Technologies on Chinese Hamster Ovary Based Bioproduction. Biotechnology Journal, 13(3), 1–14. 10.1002/biot.201700227

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, … von Mering C (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acid Research, 43(Database issue), D447–52. 10.1093/nar/gku1003

Van Gent DC, Hoeijmakers JHJ, & Kanaar R (2001). Chromosomal stability and the DNA double-stranded break connection. Nature Reviews Genetics, 2(3), 196–206. 10.1038/35056049

Vcelar S, Jadhav V, Melcher M, Auer N, Hrdina A, Sagmeister R, … Borth N (2018). Karyotype variation of CHO host cell lines over time in culture characterized by chromosome counting and chromosome painting. Biotechnology and Bioengineering, 115(1), 165–173. 10.1002/bit.26453 [PubMed: 28921524]

Veith N, Ziehr H, MacLeod RAF, & Reamon-Buettner SM (2016). Mechanisms underlying epigenetic and transcriptional heterogeneity in Chinese hamster ovary (CHO) cell lines. BMC Biotechnology, 16(1), 1–16. 10.1186/s12896-016-0238-0 [PubMed: 26729248]

Walsh G (2018). Biopharmaceutical benchmarks 2018. Nature Biotechnology, 24(7), 769–776. 10.1038/nbt.3040

Wood RD, Mitchell M, & Lindahl T (2005). Human DNA repair genes, 2005. Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis, 577(1–2 SPEC. ISS.), 275–283. 10.1016/j.mrfmmm.2005.03.007 [PubMed: 15922366]

Wurm F, & Wurm M (2017). Cloning of CHO Cells, Productivity and Genetic Stability—A Discussion. Processes, 5(2), 20. 10.3390/pr5020020

Xiong K, Marquart KF, la Cour Karottki KJ, Li S, Shamie I, Lee JS, … Kildegaard HF (2019). Reduced apoptosis in Chinese hamster ovary cells via optimized CRISPR interference. Biotechnology and Bioengineering, 116(7), 1813–1819. 10.1002/bit.26969 [PubMed: 30883679]

Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, … Wang J (2011). The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. Nature Biotechnology, 29(8), 735–741. 10.1038/nbt.1932

Yang Y, Mariati, Chusainow J, & Yap MGS (2010). DNA methylation contributes to loss in productivity of monoclonal antibody-producing CHO cell lines. Journal of Biotechnology, 147(3–4), 180–185. 10.1016/j.jbiotec.2010.04.004 [PubMed: 20430058]

Yusufi FNK, Lakshmanan M, Ho YS, Loo BLW, Ariyaratne P, Yang Y, … Lee D-Y (2017). Mammalian Systems Biotechnology Reveals Global Cellular Adaptations in a Recombinant CHO Cell Line. Cell Systems, 4(5), 530–542.e6. 10.1016/j.cels.2017.04.009 [PubMed: 28544881]

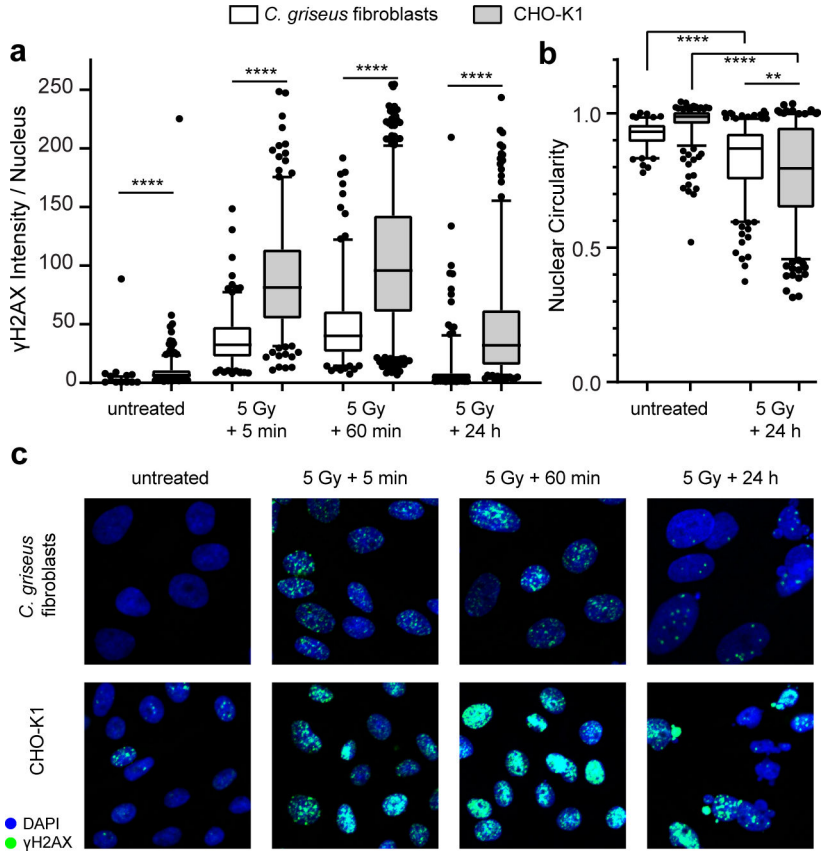

Fig. 1: DSB repair capacity in CHO cells.

(a) Quantification of γH2AX foci in CHO-K1 and Chinese hamster *(Cricetulus griseus)* fibroblasts after treatment with 5 Gy irradiation and the indicated recovery periods. Untreated cells shown to the left. Whiskers showing 5/95 percentiles. Wilcoxon-Mann-Whitney test. untreated: n=123/330; 5min: n=183/272; 60min: n=182/578; 24h: n=261/290. **(b)** Quantification of nuclear morphology as an indicator of nuclear fragmentation during apoptosis, after treatment with 5 Gy irradiation and 24 h recovery. Whiskers showing 5/95-quantiles. Wilcoxon-Mann-Whitney test. untreated: n=123/330; 24h: n=261/290; **(c)** Images showing immunostainings against γH2AX in CHO-K1 and Chinese hamster *(Cricetulus griseus)* fibroblasts corresponding to the conditions in (a). Cells counterstained with DAPI.

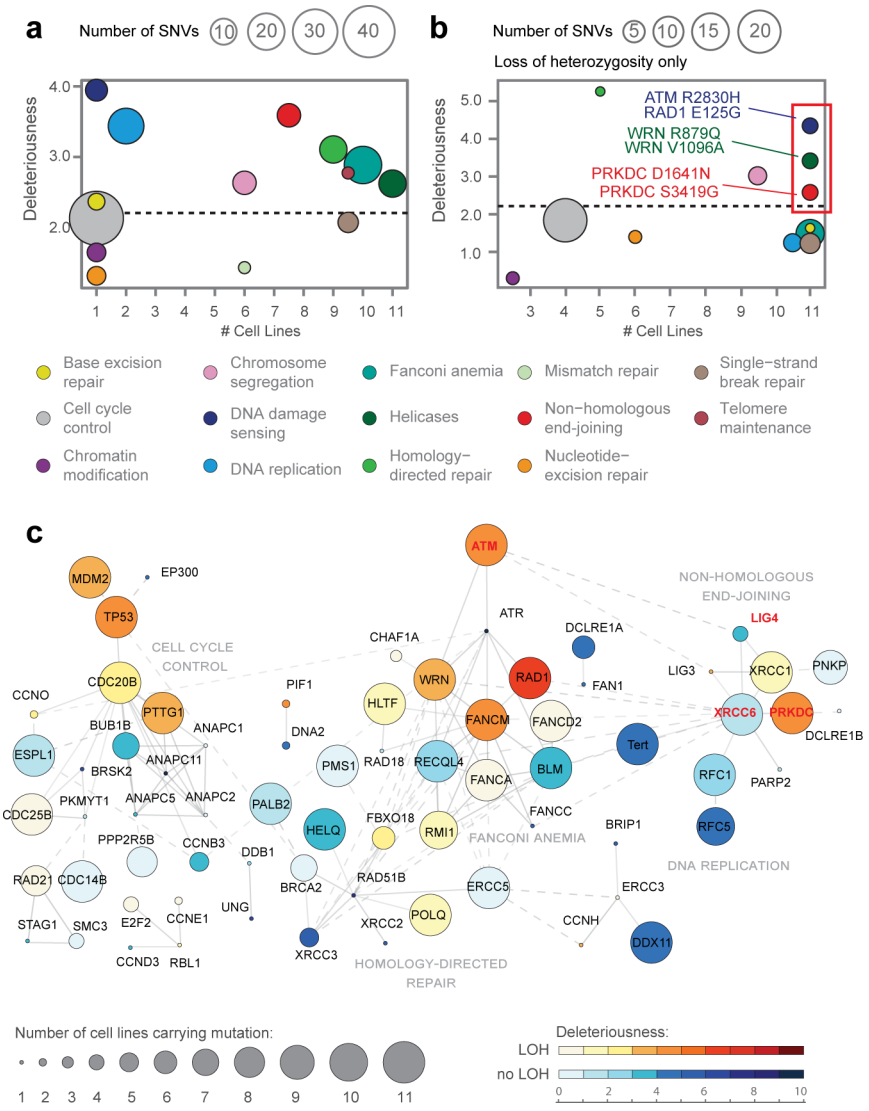

Fig. 2: DNA repair mutation landscape in CHO.

(a): Mutations found in whole-genome sequencing data from 11 major CHO cell lines, categorized into 14 gene ontology groups (bottom legend) related to genome stability. The number of CHO lines carrying each mutation and the predicted deleteriousness are averaged across all mutations in each ontology group (x-axis: median of cell lines affected; y-axis: average negative PROVEAN score). Dashed line indicates the recommended threshold (2.282) to separate neutral from detrimental SNVs (Choi et al., 2012). In total, 157 mutations are shown. **(b):** Subset of mutations shown in (a) that have undergone loss of heterozygosity, i.e. having lost the Chinese hamster wildtype allele at the respective locus (62 mutations total). **(c)**: Protein-protein interaction network of genes carrying mutations in at least one CHO cell line. Number of CHO cell lines carrying each mutation is displayed by bubble size. Color represents predicted deleteriousness (negative PROVEAN score) and loss of heterozygosity (LOH). In cases where multiple mutations were found in a gene, the one being more conserved across CHO lines is shown. The network is grouped into 4 clusters based on protein-protein interaction score with connections across clusters shown by

dashed lines. Genes not participating in a cluster were omitted for clarity. Genes selected for correction in this study are shown in red.

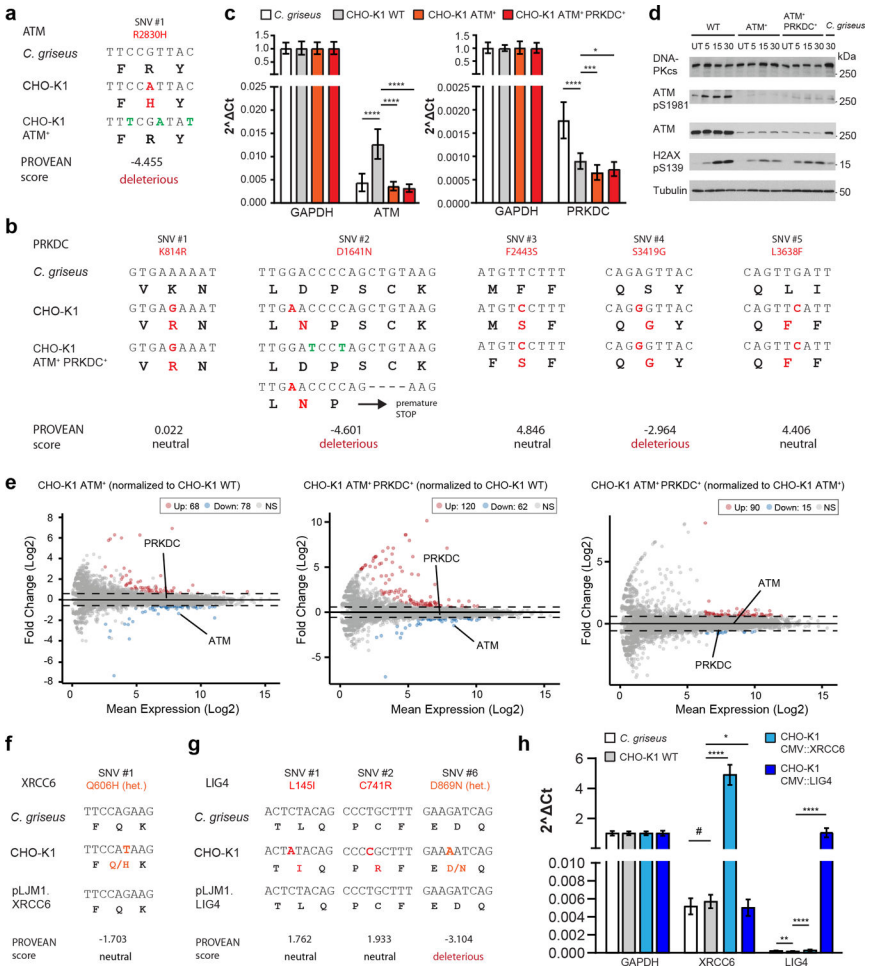

Fig. 3: Correction of DNA repair mutations in CHO.

(a) Alignment of *ATM* Sanger sequencing data from genomic DNA of Chinese hamster (*C. griseus*), CHO-K1 WT, and CHO-K1 ATM$^+$, and predicted deleteriousness (PROVEAN score). The additional silent SNVs in CHO-K1 ATM$^+$ were intentionally introduced during gene editing (see Methods). **(b)** Alignment of *PRKDC* Sanger sequencing data from genomic DNA of Chinese hamster (*C. griseus*), CHO-K1 WT, and CHO-K1 ATM$^+$ PRKDC$^+$. In ATM$^+$ PRKDC$^+$, gene editing introduced a frameshift mutation in one *PRKDC* gene copy leading to a premature stop, but to a successful reversal of the D1641N SNV in the second gene copy. The additional silent SNVs in ATM$^+$ PRKDC$^+$ were intentionally introduced during gene editing (see Methods). **(c)** Gene expression analysis (quantitative PCR) of *ATM* and *PRKDC* from cDNA of Chinese hamster (*C. griseus*), CHO-K1 WT, CHO-K1 ATM$^+$, and CHO-K1 ATM$^+$ PRKDC$^+$. GAPDH expression was used for normalization. Mean +/− S.D. Wilcoxon-Mann-Whitney test, n=16. **(d)** Immunoblots against the indicated antigens from cell extracts of CHO-K1, CHO-K1 ATM$^+$, CHO-K1 ATM$^+$ PRKDC$^+$, and Chinese hamster *(C. griseus)* fibroblasts, after exposure to 10 Gy of ionizing radiation and the indicated recovery time periods (in min), UT: untreated control. **(e)**: RNA-seq analysis of CHO-K1 ATM$^+$ and CHO-K1 ATM$^+$ PRKDC$^+$. *left*: Transcript fold-change (y-axis) in ATM$^+$ relative to WT, *middle*: Transcript fold-change (y-axis) in ATM$^+$

PRKDC$^+$ relative to WT. *right*: Transcript fold-change (y-axis) in ATM$^+$ PRKDC$^+$ relative to ATM$^+$. Significantly up-/downregulated genes highlighted in red/blue, NS: not significant. **(f)** Alignment of *XRCC6* Sanger sequencing data from genomic DNA of Chinese hamster *(C. griseus)*, CHO-K1 WT, and the pLJM1.XRCC6 plasmid used for generation of CHO-K1 CMV::XRCC6. CHO-K1 harbors a heterozygous SNV at Q606H. **(g)** Alignment of *LIG4* Sanger sequencing data from genomic DNA of Chinese hamster *(C. griseus)*, CHO-K1 WT, and the pLJM1.LIG4 plasmid used for generation of CHO-K1 CMV::LIG4. CHO-K1 harbors a heterozygous SNV at D869N, predicted to be deleterious by PROVEAN. **(h)** Quantitative PCR of *XRCC6* and *LIG4* from cDNA of Chinese hamster (*C. griseus*), CHO-K1 WT, CHO-K1 CMV::XRCC6, and CHO-K1 CMV::LIG4. GAPDH expression was used for normalization. Wilcoxon-Mann-Whitney test, n=16.

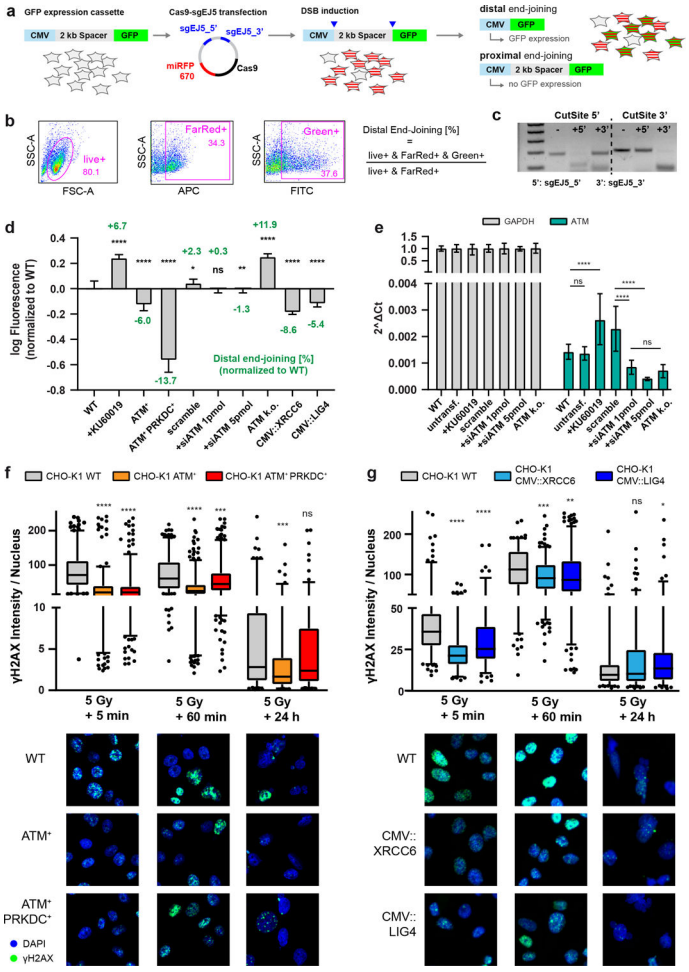

Fig. 4: Analysis of DNA repair in engineered CHO cell lines.

(a): The EJ5-GFP cassette comprises a CMV promoter, a GFP reading frame, and a 2 kb spacer that blocks GFP expression. Transfection with the Cas9-sgEJ5 plasmid, encoding a far-red fluorescent miRFP-tagged Cas9 and two sgRNAs, triggers two DSBs at the 5' and 3' ends of the spacer. Transfected (miRFP-positive) cells that join the proximal (i.e., correct) ends keep the spacer in place and remain GFP-negative while transfected cells that join the distal (i.e., incorrect) ends express GFP (assay modified from (Bennardo et al., 2008)). **(b):** Distal end-joining is quantified by relating the fraction of GFP/miRFP-positive cells to all miRFP-positive cells, using flow cytometry with the gates shown. **(c):** PCR products encompassing the 5' and 3' cut sites on the EJ5-GFP spacer, after incubation with either a Cas9:sgEJ5_5' or Cas9:sgEJ5_3' ribonucleoprotein particle (Integrated DNA Technologies). **(d):** EJ5-GFP assay on the CHO cell lines shown (+KU60019: CHO-K1 WT grown with 3 μM KU60019; scramble: transfection with 1 pmol scramble RNA; ATM k.o.: CHO-K1 harboring an *ATM* knock-out (Suppl. Fig. 2d)). Median of log10 GFP fluorescence (after subtraction of WT median) in miRFP-positive cells. Error bars = 95% confidence interval. Differences to WT in distal end-joining fraction indicated in green. Wilcoxon-Mann-Whitney test, n=8315 / 9019 / 9240 / 10474 / 12323 / 12197 / 11130 / 12039 / 7864 / 7775, pooled cell populations from at least three Cas9-sgEJ5 transfections.

Full distributions shown in Suppl. Fig. 2a–c. **(e):** Quantitative PCR of *ATM* from cDNA obtained after Cas9-sgEJ5 transfection (untransf.= untransfected CHO-K1 WT control). Mean +/− S.D. Wilcoxon-Mann-Whitney test, n=16. **(f):** Immunostainings against γH2AX in CHO-K1 WT, ATM^+, ATM^+ $PRKDC^+$, after treatment with 5 Gy irradiation and the indicated recovery periods. Whiskers showing 5/95-quantiles. Wilcoxon-Mann-Whitney test, n=229 / 275 / 292 / 260 / 355 / 321 / 141 / 112 / 153. **(g):** Immunostainings against γH2AX in CHO-K1 WT, CMV::XRCC6, and CMV::LIG4 after treatment with 5 Gy irradiation and the indicated recovery periods. Cells counterstained with DAPI. Whiskers showing 5/95-quantiles. Wilcoxon-Mann-Whitney test, n=186 / 132 / 117 / 166 / 231 / 254 / 150 / 178 / 131.

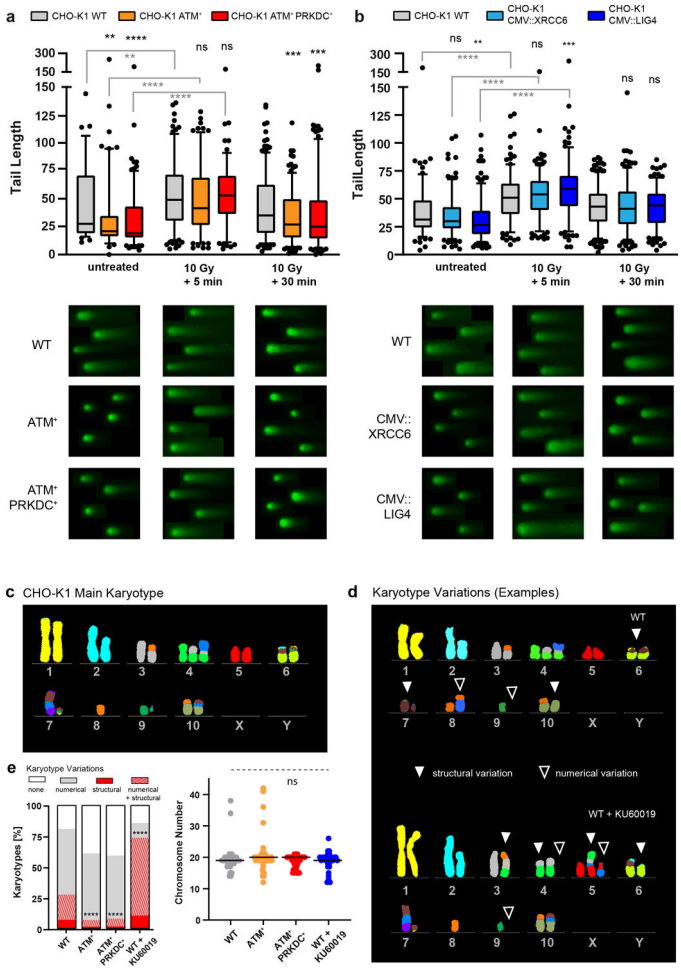

Fig. 5: Analysis of genome stability in engineered CHO cell lines.

(a): Analysis of DNA tails as an indicator of genome fragmentation in CHO-K1 WT, ATM$^+$, and ATM$^+$ PRKDC$^+$ cell lines using electrophoresis in a low-melting agar (comet assay) after 10 Gy of ionizing radiation and the indicated recovery periods. Untreated controls shown on the left. Whiskers showing 5/95-quantiles. Wilcoxon-Mann-Whitney test, n=78 / 123 / 169 / 179 / 146 / 152 / 292 / 217 / 215. Composite example images shown below. DNA stained with Vista Green. **(b):** Analysis of DNA tails in CHO-K1 WT, CMV::XRCC6, and CMV::LIG4 cell lines using electrophoresis in a low-melting agar (comet assay) after 10 Gy of ionizing radiation and the indicated recovery periods. Untreated controls shown on the left. Whiskers showing 5/95-quantiles. Wilcoxon-Mann-Whitney test, n=158 / 216 / 248 / 202 / 205 / 187 / 320 / 259 / 202. Composite example images shown below. DNA stained with Vista Green. **(c):** Main CHO-K1 karyotype identified by multi-color DNA FISH. **(d):** Examples of karyotypes deviating from the WT karyotype (top) and WT supplemented with 3 μM KU-60019 (bottom) following long-term culture. Open arrows indicate a numerical variation (gain or loss of a chromosome) and closed arrows indicate a structural variation (e.g. translocations, deletions). **(e):** *Left*: Quantification of metaphase spreads by multi-color DNA FISH from the indicated cell lines with one or more chromosomal abnormalities relative to the main karyotype. Data are depicted as the mean of two experiments (WT

n=60; ATM+ n=58; $ATM^{+}PRKDC^{+}$ n=58; WT+KU60019 n=27 metaphases). Binomial test on all structural variations (combined red and red-striped fractions). *Right:* Total number of chromosomes per karyotype. Bar = median. Non-parametric ANOVA (Kruskal-Wallis test).

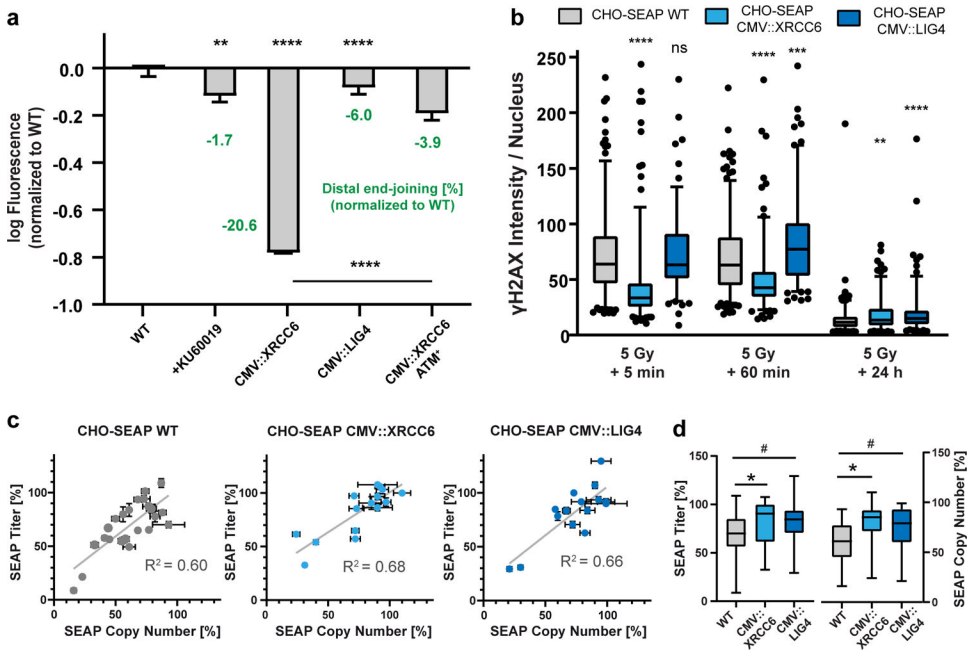

Fig. 6: DNA repair and protein titer in a producing CHO cell line.

(a): EJ5-GFP assay on the CHO-SEAP cell lines shown (+KU60019: CHO-SEAP WT grown with 3 μM KU-60019). Median of log10 GFP fluorescence (after subtraction of WT median) in miRFP-positive cells. Error bars = 95% confidence interval. Differences to WT in distal end-joining fraction indicated in green. Wilcoxon-Mann-Whitney test, n=7687 / 7136 / 14006 / 6126 / 9363, pooled cell populations from three Cas9-sgEJ5 transfections. Full distributions shown in Suppl. Fig. 2g,h. **(b):** Immunostainings against γH2AX in CHO-SEAP WT, and two representative clones of CHO-SEAP CMV::XRCC6 and CMV::LIG4 from the long-term study shown in (c). Cells were treated with 5 Gy irradiation and allowed to recover for the indicated time periods. Whiskers showing 5/95-quantiles. Wilcoxon-Mann-Whitney test, n=213 / 218 / 137 / 251 / 170 / 142 / 257 / 157 / 146. **(c):** SEAP titer change versus SEAP gene copy number change in clonal populations of the indicated genotypes over a long-term culture period of 74 days. Percentages indicate retention of SEAP titer and copy number at day 74 relative to day 0. **(d):** Boxplots of the data shown in (c). Whiskers showing min/max. Wilcoxon-Mann-Whitney test. n=23 / 15 / 16.

Table 1:

Genome sequences of CHO cell lines analyzed in this study.

Cell line	Origin	NCBI SRA#
CHO-K1 (adherent)	ATCC	SRP045758
CHO-K1 (adherent)	ECACC	SRS406579
CHO-K1/SF (adherent)	ECACC	SRS406580
CHO protein-free (suspension)	ECACC	SRS406578
CHO-DG44 (adherent)	Life Technologies	SRS406582
CHO-S (suspension)	Life Technologies	SRS406581
CHO-S (suspension)	Clone from the Technical University of Denmark (derived from Life Technologies)	(unpublished)
C0101 (suspension)	Undisclosed company (Drug producing cell line derived from CHO-S from Life Technologies)	SRX258098
CHO-Z (suspension)	Clone from the Technical University of Denmark (Serum-free suspension adapted clone derived from an ECACC CHO-K1 clone)	(unpublished)
CHO-DXB11 (adherent)	Clone from the Technical University of Denmark	SRX689758
CHO-pgsA-745 (adherent)	ATCC	SRS1203612