Subcloning induces changes in the DNA-methylation pattern of outgrowing Chinese hamster ovary cell colonies

Marcus Weinguny1,2,† | Gerald Klanert1,† | Peter Eisenhut1,2 | Isac Lee3 | Winston Timp3 | Nicole Borth1,2

1 ACIB Gmbh, Austrian Centre of Industrial Biotechnology, Vienna, Austria
2 Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Vienna, Austria
3 Department of Biomedical Engineering, Johns Hopkins University, Baltimore, USA

Correspondence
Dr. Nicole Borth, Department of Biotechnology, University of Natural Resources and Life Sciences Vienna, Muthgasse 18, 1190 Vienna, Austria.
Email: nicole.borth@boku.ac.at

†Marcus Weinguny and Gerald Klanert contributed equally to this study.

Abstract
Chinese hamster ovary (CHO) cells are the most extensively used mammalian production system for biologics intended for use in humans. A critical step in the establishment of production cell lines is single cell cloning, with the objective of achieving high productivity and product quality. Despite general use, knowledge of the effects of this process is limited. Importantly, single cell cloned cells display a wide array of observed phenotypes, which so far was attributed to the instability and variability of the CHO genome. In this study we present data indicating that the emergence of diverse phenotypes during single cell cloning is associated with changes in DNA methylation patterns and transcriptomes that occur during the subcloning process. The DNA methylation pattern of each analyzed subclone, randomly picked from all outgrowing clones of the experiment, had unique changes preferentially found in regulatory regions of the genome such as enhancers, and de-enriched in actively transcribed sequences (not including the respective promoters), indicating that these changes resulted in adaptations of the relative gene expression pattern. The transcriptome of each subclone also had a significant number of individual changes. These results indicate that epigenetic regulation is a hidden, but important player in cell line development with a major role in the establishment of high performing clones with improved characteristics for bioprocessing.

KEYWORDS
Chinese hamster ovary, CHO, DNA methylation, single cell cloning, subclones, subcloning, transcriptome

1 INTRODUCTION

Worldwide, Chinese hamster ovary (CHO) cells are the most relevant production organism for biotherapeutic proteins for use in humans due to their ability to produce human-like glycosylation and the availability of selection and amplification systems.[27,58] For manufacture of a therapeutic protein it is necessary to generate production clones, which (I) reproducibly produce the protein in high quantities and

Abbreviations: ccd, cumulative cell days; ccv, cumulative cell volume; CHO, Chinese hamster ovary; DEG, differentially expressed gene; DMG, differentially methylated gene; DMR, differentially methylated regions; GO, gene ontology; GSEA, Gene Set Enrichment Analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; MCB, master cell bank; maxTiter, maximum Titer; PCA, Principal Component Analysis; qP, specific productivity; RNA-Seq, RNA Sequencing; TSS, transcription start site; VCD, Viable cell Density; VCV, Viable cell volume; μ, growth rate

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. Biotechnology Journal published by Wiley-VCH GmbH
quality; (II) grow reasonably fast and to high cell densities; and (III) were generated according to the guidelines of the regulatory agencies. A prerequisite there is the subcloning of cells with subsequent screening for further cell line development. This is done to remove low producers that may have a growth advantage over high producers and thus lead to loss in culture productivity over time, and to ensure consistent product quality, a result of the known variability and genomic instability of in-vitro cultivated cells. While most steps in cell line development have been optimized, streamlined, and automated by the industry, leading to significant reduction in timelines, the time required for subcloning is an unsurmountable limiting factor due to the time single cells need to grow into a colony of sufficient size for further expansion.

Subcloning can be achieved via limited dilution, which is essentially based on the statistical likelihood that due to the dilution used, there will be one cell per well. Other tools are available, such as fluorescent activated cell sorting, to enrich for high producers and further assure clonality. With the advancement in automation and microfluidic systems, various instruments for subcloning, such as cell printing and light-induced sorting, as well as instruments for additional image verification of clonality have been described. Additionally, media additives to promote colony outgrowth were developed.

These tools aim to ensure either the isolation of a verified single cell or the enrichment of rare high producers. However, insufficient scientific attention has been paid to the underlying reason to perform subcloning at all, which is the inherently high variation in subclone phenotypes even when derived from a parent population that had already been subcloned. In the pre-omics era, a common dictum in the field was that the high genomic variation in CHO cells (or any fast growing cell line, for that matter) is the source of this variation and that during subcloning one simply selects for individual cells with specific mutations that, thus, provide individual phenotypes to each subclone. However, a recent study that investigated the oxidative phosphorylation and energy metabolism in a variety of sequenced CHO cells showed that cells with specific phenotypes are characterized by a mixture of mutations in relevant genes and changes in their expression patterns. While it is likely that changes in expression levels are also caused by sequence mutations in the relevant regulatory regions for a given gene, an additional and as likely explanation is that many differences in phenotype across different cell lines, lineages and subclones are also controlled by changes in the epigenetic code of cells. This was indeed shown to be the case after adaptation of cell lines to new culture conditions and altered medium composition, as well as for a cell line selected specifically for higher transient productivity.

The epigenome is a network of various mechanisms that regulate gene expression without changing the underlying nucleotide sequence. The most prominent mechanisms are chromatin accessibility as controlled by the methylation of cytosines in the DNA (also known as methylome) and diverse modifications of the N-terminal tails of histones. The presence and/or combination of such modifications are able to control whether or not a gene is expressed and to what level of transcription. As a specific phenotype is not likely to be determined by a single mutation or the change in expression of a single gene, the combination of genetic and epigenetic variation actually provides cells with a wide space of possible transcriptomes and resulting phenotypes across the 10–15,000 expressed coding and the even more non-coding genes. This wide space of variation in transcript expression levels would explain the known adaptability of CHO cells and the success of process and cell line development that the biopharmaceutical industry has seen over the last 30 years.

The two mechanisms of epigenetic modifications addressed above have different kinetics and thus serve different functionalities in cells. While histone modifications enable rapid responses to culture environment and the maintenance of stable as well as differentially regulated gene expression; DNA-methylation is a more long term mechanism where the information "learned" during adaptation can be passed on to daughter cells. In the present study, we therefore focused on DNA-methylation as the "inheritable" epigenetic mark and aimed to understand its role in relation to the behavior and the emergence of specific phenotypes of individual subclones and their resulting diversity. Therefore, an already established subclone producing an Epo-Fc fusion protein was single cell sorted, the first 36 outgrowing subclones expanded and their batch behavior tested twice over 8 weeks. Six subclones that represented the full range of phenotypes and stability behavior were chosen for detailed characterization of their transcriptome and methylome in comparison to the parental cell line.

2 | METHODS

2.1 | Cell culture

A previously established CHO DUKX B11 subclone producing EPO-Fc was routinely cultivated in TPP TubeSpin bioreactors (TPP Techno Plastic Products, Switzerland) in CD CHO media (Thermo Fisher Scientific, USA) supplemented with 8 mM L-Glutamine (Sigma-Aldrich, USA), 960 µM Methotrexate (Sigma-Aldrich) and 0.2% Anti Clumping Agent (Thermo Fisher Scientific), at 37°C, 7% CO₂, 85% humidity and 220 rpm shaking with a shaking diameter of 25 mm. Cells were passaged every 3–4 days.

2.2 | Subcloning

25 x 96-well-plates (Greiner Bio-One, Austria, Cat.#: 655086) were prepared with 200 µL medium per well, further supplemented with 1:100 Penicillin-Streptomycin (10,000 U mL⁻¹, 10 mg mL⁻¹ Streptomycin, VWR Chemicals, USA), CHO DUKX B11 cells (passage 6 after thawing) were sorted by a MoFlo Astrios (Beckmann Coulter, USA), using a 488 nm laser to determine forward (FSC) and side scatter (SSC). Sorted single cells were statically incubated at 37°C, 7% CO₂ and 85% humidity for 3 weeks. Afterwards, 68 randomly chosen colonies were chosen for further investigation. Colonies were roughly the same shape and size as observed by the naked eye. Cells were...
transferred to the inner wells of 3 x 48-well plates (Greiner Bio-One) with 0.8 mL of appropriate Medium, and incubated at 37°C, 7% CO₂, 85% humidity and 300 rpm shaking with a shaking diameter of 25 mm. After one passage, the number of colonies was reduced to 48 clones, selecting for growing cell lines. A week later cells were transferred to TPP TubeSpin bioreactors, by centrifuging plates at 180 rpm for 8 min, removing the supernatant and resuspending cells in 1 mL of fresh medium which was then transferred into tubes already containing 9 mL fresh medium.

2.3 Stability study

The design of the study is presented in Figure 1a. In short, cells were kept in culture, as described above, for 53 days. The first batch evaluation was done one passage after the transfer to TPP TubeSpin bioreactors. A second batch was performed at the end of the 53 days. During that time, 12 cultures were lost, resulting in 36 clones that could be used for further investigation. In parallel, three replicates of freshly thawed CHO DUKX B11, passage 6 post thawing, were used to perform the same stability study, to determine parental phenotype changes.

2.4 Batch cultures

Cell lines were seeded at 0.2 x 10⁶ cells mL⁻¹ in 20 mL TPP TubeSpin bioreactors and incubated as described above. Viable cell concentration (VCD) and viability were determined each day using a ViCell XR 2.04 (Beckman Coulter) based on the Trypan blue exclusion method. RNA and gDNA were sampled on day 2 and 5. For RNA, 5 x 10⁶ cells were spun down at 200 x g for 8 min, supernatant removed and the pellet dissolved in 600 μL TRI Reagent (Merck KGaA, Germany) and stored at −80°C for later isolation. gDNA was isolated using the DNeasy Blood & Tissue Kit (Qiagen, Netherlands) according to the manufacturer’s instructions and stored at −80°C. Product titer was quantified each day. Cells were spun down at 180 x g for 8 min and the supernatant frozen at −20°C for later quantification. Batches were characterized using an in-house R package vicellR version 0.1.9.²³

FIGURE 1 Results of CHO DUKX B11 aging study. (A) Study design. After single cell sorting, cells were expanded, and a first batch experiment was performed. Cells were kept in culture for 53 days at which end a second batch experiment was performed. (B) Change of maximum titer between batches, blue line: trendline of all subclones. (C) μP and μ deviations between both batch experiments in % as calculated by dividing the respective value of batch 2 by that of batch 1.
2.5 | EPO-Fc quantification

EPO-Fc concentration was quantified using the Octet RED96e (FORTÉBIO, USA), equipped with Dip and Read Protein A Biosensors (Pall Corp, USA) according to the manufacturer’s recommendations. Quantification was performed relatively to Trastuzumab (BioVision, USA), as no EPO-Fc standard was commercially available.

2.6 | DNA methylation analysis

The gDNA of six selected clones (first batch, D2) and three replicates of CHO DUKX B11, passage 6 post thawing were bisulfite treated using the EZ DNA Methylation-Lightning Kit (Zymo Research, USA) according to the manufacturer’s instructions. The library was prepared using Accel-NGS Methyl-Seq DNA Library Kit (Swift Biosciences, USA) with Accel-NGS Unique Dual Indexing Kit (Swift Biosciences) according to the manufacturer’s instructions. Libraries were analyzed by Illumina NovaSeq S1 with 300 cycles at The Genetic Resources Core Facility of Johns Hopkins University, USA. Sequencing depth for each sample is shown in Table 2.

Raw reads were processed using Trim-galore 0.6.0,[34] with a quality cut-off of 28 and trimming at the 5′ end of 5 bp of read 1 and 15 bp of read 2. Additionally, 5′ bp were removed at the 3′ end on both reads. Processed reads were aligned paired-end mode to the Chinese hamster genome[48] using the Bismark v0.22.1 pipeline (non-default parameters: N = 1, score_min = L,0-0.6).[26] Moreover, Bismark was used to remove duplicate reads and to generate methylation profiles with default settings. Qualimap v.2.2.2 was used to check raw read quality[40] and sequencing depth (Table 2). Due to the lack of replicates differential analysis was performed using DSS-single.[62] DSS setting comprised a smoothing span of 500 bp with a minimum DMR length of 50 bp with ≥4 CpGs and p < 0.05 (Wald test). DMR intersection upset plot was generated using the R-package ComplexHeatmap.[16] Raw methylation data was acquired using the R-package bsseq.[18]

Loci were filtered for PCA and methylation plotting for a coverage of >0 in all samples.

Chromatin state enrichment analysis was performed using data acquired from[48] accessible on http://cgr-referencegenome.boku.ac.at/ using the line "TP_3(42hr)". DMRs were assigned to the chromatin state using the intersect command of the R package GenomicRanges[29]. Enrichment was determined using the following ratio:

\[
\text{Chromatin Enrichment} = \frac{\sum l_{\text{chromDMR}} / \sum l_{\text{DMR}}}{\sum l_{\text{chromGenome}} / \sum l_{\text{Genome}}}
\]

where \(l_{\text{chromDMR}}\) is the length of intersections of each reported chromatin states with identified DMRs, \(l_{\text{DMR}}\) is the length of all DMRs identified, \(l_{\text{chromGenome}}\) the length of each reported chromatin state identified in the reference genome and \(l_{\text{Genome}}\) the length of the whole reference genome.

For Euclidian distance and uniform loci analysis, parental replicate reads were summed up and filtered together with subclone reads for loci with coverage >0. Circular diagrams were generated using the R package circlize.[17] Sequencing data has the accession number PRJEB38542 and a browser is available at www.BorthLabCHOresources.boku.ac.at.

2.7 | Impact of aging

To investigate the impact of methylation during cell line aging, data from a previous CHO methylome study[11] was processed using the above pipeline. Four datasets were used: 1. data acquired from a master cell bank (MCB) sample of the CHO-K1 cell pool, 2. data from the same cell line after maintenance in culture for 6 months, 3. data of a high productivity (qP) subclone, isolated from the MCB pool, and 4. data of the same clone after 3 months in cultivation. DMR analysis was performed between the MCB samples and the respective subclone samples to evaluate the impact of aging. Additionally, DMR analysis was performed between the starting MCB sample and the starting subclone sample, to investigate the impact of phenotypic selection. The only difference to above was during trimming, where a quality cut off of 28, and trimming at the 5′ end of 10 bp and 3′ end of 1 bp were chosen to ensure read quality for the different data set.

2.8 | RNA sequencing

Total RNA was isolated from the D5 samples using Direct-zol RNA mini prep kit (Zymo Research) according to the manufacturers instruction. Singular samples of the six selected clones were submitted for sequencing as two technical replicates, whereas the parental cell line was analyzed in two biological replicates.

Low quality reads and adapters of the raw sequences were trimmed using Trimmomatic 0.36.[4] Processed reads were mapped to the Chinese hamster genome[48] using HiSat2 2.1.0.[22] Mapped reads were counted using the HTSeq python package.[2] DESeq2 R package 1.24.0[22] was used to analyze read counts. Differential expression analysis was performed using the DESeq function of the package. Differentially expressed genes between samples were analyzed using the log2 fold change threshold 0 and BH adjusted p-value 0.05. Genes with a foldchange difference of ≥1.5 and BH p-value < 0.05 were deemed significantly differentially expressed. For further analysis, counts were normalized using the DESeq2’s variance stabilizing transformation (vst-normalization).

Pathway enrichment was performed using Gene Set analysis (GSEA), gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG). GSEA analysis was performed using GSEA 4.0.3.[57,52] GSEA Preranked settings were 1000 permutations, use of c2.cp,v7.0, symbols.gmt as geneset and no collapse. GO profiling was done using the command gost of the R-package gProfiler.[45] with the options: organism = "cgpicr", domain_scope = "annotated", correction_method = "gsCS". Identified genes were mapped to the KEGG pathways[20,21] using the R package KEGGprorifer,[66] as annotated to identify significantly enriched pathways. Significance was tested
Transcriptome characterization

Stable High (SH)

Stability study and subclone phenotypic

Full genome DNA-methylation pattern

RESULTS

Downstream computational analysis

Table 1: Phenotypical classification of chosen CHO DBX 11 subclones

| Sample | Classified phenotype | Abbreviations |
|--------|----------------------|---------------|
| 1-B3   | Stable qP, low expression | Stable Low (SL) |
| 1-C7   | Stable qP, low-medium expression | Stable Low-Medium (SLM) |
| 1-C3   | Stable qP, medium expression | Stable Med (SM) |
| 1-E2   | Stable qP, high expression | Stable High (SH) |
| 1-C6   | Decreased qP, medium-low expression | Decreased Medium (DM) |
| 1-E7   | Increased qP, high expression | Increased High (IH) |

using a hypergeometric test with \( q < 0.05 \) using a Benjamini-Hochberg correction.

2.9 Downstream computational analysis

Downstream analysis was done in the statistical software R version 3.6.2.\[44\] Plotting was done using ggplot2.\[61\] Upset plots were generated using the R-package UpSetR.\[18\]

3 RESULTS

3.1 Stability study and subclone phenotypic diversity

To investigate the effects of subcloning on cellular phenotypes, a previously subcloned EPO-Fc producer cell line\[28\] was re-subcloned, in the expectation that an already subcloned cell line has a more uniform genetic landscape than a cell pool and thus should yield more homogenous subclones than a pool. After outgrowth of randomly selected colonies which showed reasonable growth and expansion after 3 weeks, 36 subclones were maintained in culture for 53 days (Figure 1a). Batch experiments at the beginning and the end of this stability study were used to evaluate the phenotype of clones. Unexpectedly, the clones performed better during the second batch (Figure S1–S5), whereas the parental cell line showed the anticipated productivity loss.\[41\] This lower performance of the recently subcloned S1–S5, compared to the parental cell line, may have been due to the cells not yet being fully recovered after subcloning, a stress that the parental pool was not exposed to. Based on the maximum titer measured and the deviation between the two batch cultivations (Figure 1b and c), four clones were selected along the diagonal to cover the entire range of stable productivity, as well as two subclones that deviated from the diagonal due to either increased or decreased productivity (Table 1).

3.2 Transcriptome characterization

For transcriptional characterization of both the parental and the second generation subclones, RNA-Seq samples were taken during the exponential phase of the first batch. Using DESeq2, we measured the Euclidian distance between gene expression patterns which revealed two main clusters. One contained the parental clone and SL, while the other five subclones are closer to each other than to the parent or the low-productivity clone SL (Figure 2a). Performing a principal component analysis (PCA) shows that all replicates match closely and that each subclone can be distinguished (Figure 2b). Moreover, the same pattern as in the distance matrix is visible: Along PC1, most of the subclones (SM, SLM, SH, DM, IH) are separated from the parent and SL, while these two are clearly separated along PC2. Performing differentially expressed gene (DEG) analysis between the parental cell lines and isolated subclones (Figure S6 and Table S1) showed that the 5 subclones that clustered in the PCA share a significant number of up- and downregulated genes (Figure 2c), which are not DE in SL. However, each of the subclones also shows a large number of unique DEG, indicating individual clone specific transcriptome changes. One has to keep in mind that, due to the differences in growth behavior, the samples used for RNA-Seq for some of the subclones were already during the late stage of exponential growth (SM, SH), while the other samples, including the ones from the parent, were still mid-exponential (Figure S3). Nevertheless, closely spaced transcriptome data from a previous study showed that there are little changes in transcriptome throughout the exponential phase of a batch culture\[19\] and no distinct clustering of these two samples was observed in the PCA or the distance matrix (Figure 2a and b). This supports that there is no systematic difference between the samples due to variation in culture stage or growth and strengthens the conclusion of individual, subclone specific changes.

To understand the underlying function of these transcriptional changes, we performed pathway analysis via GSEA (Figure 2d and Table S2). This revealed that the subclones share many upregulated pathways, even though the specific genes in each pathway may be different. Clones SL, SLM and SH also show a substantial number of unique pathways that are significantly enriched. On the other hand, downregulated pathways are more unique for each subclone with only one enriched pathway (Cytokine_Receptor_Interaction) downregulated in all 6 subclones (Figure S7). We also applied first generation enriched pathway analysis, namely GO enrichment via g:Profiler and KEGG Analysis (Figure S8 and Tables S3 and S4). GO enrichment showed a similar picture as GSEA. Although some pathways are shared by all subclones, a majority is either unique or only shared by a subset of subclones. The same was seen in the KEGG analysis. Overall, DEG and enriched pathway analysis revealed that, although the subclones share some changes in their respective transcriptomes, each of them is also characterized by a noticeable number of unique changes. Interestingly, these shared pathways include several that are related to cell division (e.g., several cell cycle checkpoints), indicating that these support clonal outgrowth, as well as several pathways that are related to DNA repair.

3.3 Full genome DNA-methylation pattern

Next, the full genome DNA-methylation pattern was sequenced and differentially methylated regions (DMRs) called. The coverage for the
FIGURE 2 RNA Sequencing results (A) Euclidean distance matrix of samples based on vst-normalized counts of all expressed genes. R1/R2 indicate replicates. (B) PCA of vst-normalized counts of all expressed genes. Overlapping symbols appear darker in color. (C) Overlapping differentially expressed genes for all subclones (D) Overlapping enriched pathways via GSEA for all subclones

TABLE 2 Sequencing depth of analyzed samples

| Sample | Coverage according to Qualimap v2.2.2 |
|--------|---------------------------------------|
| Parent-R1 | 13.40 |
| Parent-R2 | 9.79 |
| Parent-R3 | 11.54 |
| SL | 9.65 |
| SLM | 10.97 |
| SM | 7.93 |
| SH | 10.80 |
| DM | 9.45 |
| IH | 11.51 |

subclones and for each of the three replicates from the parental cell line ranged from 9.45x to 13.4x (Table 2). The PCA based on all identified CpG loci (Figure 3a) shows a similar pattern as observed for the transcriptome analysis. The parental replicates cluster closely, SL is clearly separated both from the parents and from most other subclones (SLM, SH, DM, IH) that form a cluster. Here, however, SM is also distant from the other subclones, indicating a distinct and unique methylation pattern. Furthermore, all samples have a similar level of overall DNA methylation (Figure 3b). Based on the RNA-seq data, the expressed and the non-expressed genes show the expected methylation pattern around the transcription start site (TSS) in all samples (Figure 3c): for expressed genes, the methylation is reduced, while for non-expressed genes it remains high. Single-DSS analysis to determine differentially methylated regions (DMR) revealed that the majority of DMRs in the subclones are hypermethylated (Figure 3d), where SL has almost one third more DMRs than the other subclones and an even higher percentage of hypermethylated DMRs. SM, although showing approximately the same total number of DMRs as subclones SLM, SH, DM and IH, displays more hypermethylation. Interestingly, a detailed look at these DMRs shows that there is only a small intersection of identified DMRs between all subclones (Figure 3e), while the majority is unique to each subclone.

To obtain an estimation of the impact of time in culture during the subcloning step on these results, we evaluated data from a previous study where the cell's DNA-methylation pattern had been reanalyzed after 4 and 6 months in culture in Figure S13. Using our same analysis pipeline we found that few changes occur during the time in culture. This is in marked contrast to the number of DMRs found...
between the parental cell pool and a subclone stringently selected for increased productivity where a much larger number of DMRs was detected, even exceeding those found in the present study after a single subcloning step without selection pressure. Also in line with a recent publication, aging seems to result in more hypomethylation, while subcloning increases the number of hyper-methylated CpGs.[36] Differentially methylated Genes (DMG) are defined as genes where at least one DMR is found in their promoter region (+1500 to -200 to the TSS). Interestingly, only a small overlap between DEG and DMG can be observed (Figure S9). However, the overlaps of DEG and DMG show the same trend: the more hypomethylated the DMR is, the higher is the observed log2Foldchange and vice versa (Figure S10). Seeing
that DMRs actually had only little direct impact on DEG, the location of the identified DMRs was investigated in detail. For this, previously published chromatin states for the CHO genome\(^{[48]}\) were used. DMRs were enriched in chromatin states annotated as regions flanking the TSS up—or downstream or enhancer regions (Figure S11a), which possibly has an impact on the expression level of distal, and thus unknown, genes. Clustering the Euclidean distance between all methylation loci shows that the parental cell line has similar distance to each of the subclones, with greater distances in between the subclones (Figure S11b). Plotting the 3D network revealed the parent sitting in the middle of the subclones, with SM and SL showing more distance to the other subclones and to each other (Figure S11c). This indicates that the parental subclone is indeed the starting cell line and that each subclone has developed in individual directions.

To determine whether these individual methylation patterns of the analyzed subclones are derived from adventitiously present variants in the parent population or whether they are due to active changes in methylation pattern that happened during the process of subcloning itself, a more detailed analysis of individual methylation loci was performed. Specifically we focused on those Cs in the parent genome where at the higher coverage of 35x all sequenced Cs were either fully methylated or unmethylated. From this subgroup we identified those that had a different methylation state in any subclone (Figure 4). Approx. 50% of Cs in the parent were uniformly methylated, and approx. half of these sites had a different methylation state in one or the other of the subclones, with the specific percentage of such differentially methylated sites in each subclone ranging from 5.2%—6.7% (Figure 4a). In each subclone, approx. 40% of such differentially methylated sites that had been uniformly methylated in the parent were unique to this specific subclone (Figure 4a and b). The pattern of how these uniformly methylated sites change in each subclone is presented in Figure 5 and in Figure S12. In each subclone, between 3.8% and 8.6% of these loci with originally uniform, but now different methylation were again either fully methylated or demethylated, while 8%—15% were methylated at 50%, which could indicate uniform methylation at one allele and demethylation at the other. Most importantly, the presence of uniformly methylated Cs that had the inverse methylation state in the parent indicates that these new methylation states in the subclone had not been present in the original population, but were modified during the process of subcloning.

4 | DISCUSSION

The process of subcloning and the ensuing wait for single cells to expand sufficiently to be transferred into larger culture vessels is one of the rate and time limiting steps in cell line development that is difficult to speed up. Subcloning is stressful, as documented by the low number of outgrowing colonies and in many cases by the necessity to supplement supportive proteins, hormones, and growth factors. A recent study revealed that a major contributor to this stress is the absence of cell-to-cell communication and, possibly, the lack of proteins and other cellular signals in the culture environment\(^{[59]}\). CHO cells come from an organism where they are used to being in constant contact and communication with other cells and tissues, and where a continuous network of signals is being transmitted to stimulate and support them.\(^{[50]}\) In the usual suspension culture environment where cells are present in large number, this communication network is generated by the cells themselves who secrete proteins and other communication tools such as exosomes. During subcloning, all of this is missing which conceivably puts cells under stress.

Despite this stressfulness and the long timelines required for subcloning, it has been a standard part of all cell line development platforms in the past 30 years. Invented for hybridoma technology\(^{[25]}\), where it was an essential part of generating cell lines that produce antibodies with unique binding sites, it later was used also for recombinant cell line technology, where the uniqueness of the product was given in any case by the used plasmid that was transfected. Nevertheless,
the technique had its advantage, as during transfection and selection, a wide range of productivities was generated in the cells within a population, where the high producers may have a significant growth disadvantage and thus could be outgrown by non- or low-producers.[67] In later years, the main focus of ensuring monoclonality was to maintain stability and reproducible product quality. This was based on the assumption that a subclone is genomically more homogenous than a cell pool,[60] even though several studies have shown that subclones of a subclone are as genomically diverse as subclones of the parent.[3,9,11,55] With respect to phenotypes, it was already well established that there is high diversity with respect to growth rate and specific productivity in subclones generated from an already subcloned cell line.[24,42,49,53]

Based on the here presented results it seems that, unbeknownst, but empirically successful, the process of subcloning itself made a significant contribution to our ability to isolate rare cells with outstanding properties throughout the thousands of cell screening programs performed during the last 30 years. If the isolation of cells and the consequential stress induced by this isolation, initiates a cellular program that actively alters the DNA methylation pattern of cells, randomly, but possibly with a preference for regulatory regions that will cause subtle changes in gene expression patterns, then these subtle changes, along with enabling the cells to survive and expand during subcloning, may also have adventitious effects that change other phenotypic traits, such as growth and productivity, as observed in the above mentioned studies. The data presented here would support such a hypothesis, even though definite proof will require a detailed understanding of the mechanisms that cells use to achieve such changes in their methylation pattern. It has been frequently proposed that within each CHO cell population different variants are present and that during subcloning one simply isolates one of these already existing variants. However, our results indicate that instead the epigenome is actively changed by the cells themselves. Focusing on completely methylated cytosines in the parental lines, each of the six subclones sequenced had a unique set of cytosines which were differentially methylated after subcloning. Given the coverage in the parental line (35x), if there were any cells present in the parental population with a different methylation state, they would have had to be present in less than 3% of the population. As this pattern was found across a high number of such sites, many of these unique for individual subclones, it is highly unlikely that precisely these unique combinations were present already in the parent population as rare variants. If only such cells had been able to grow out of the main population, then the subcloning efficiency would also have had to be less than 3% where in reality it is typically in the range of 30%–70%. Therefore, we suggest that our results support the hypothesis that these new methylation patterns were actively generated by the cells in response to the stress caused by isolation during the subcloning process itself.[59]

An important question in this context is at what stage such a process of re-methylation takes place and for how long it lasts. In the new subclones, a notable part of the new methylation pattern was again uniform throughout the population, with 3%–9% fully methylated or non-methylated. In addition, 8%–15% of the Cs were methylated in 50% of the reads. The status of 50% methylation can be caused either by one allele being homogenously methylated and one allele being demethylated in each individual cell, or by the presence in the population of 50% cells that have both alleles methylated and 50% with both alleles demethylated, or any mixture thereof that statistically results in 50% of the reads being methylated or not. In the first case, along with all those C sites that have changed their methylation state in 100% of the population, it is likely that the change occurred at the single cell state during subcloning, before the first division of the cell in the well occurred, thus resulting in a homogenous outcome.

It would therefore be an immediate response to the process of subcloning and the concomitant stress of having no communication to
other cells. It also may well be the required first step towards getting cells in shape for division. It is likely, however, that the search for a new and suitable transcriptome pattern that enables growth under solitary conditions continued at least for some time during the subsequent divisions, as the remaining differentially methylated Cs are not uniformly methylated, but have relative levels of methylation that range from 0.1 to 0.9 (Figure 5).

A closer look at the differential transcriptome of the 6 clones provides indications of what cells need to enable them to outgrow from single cell colonies. While most of the DE genes and also enriched pathways are unique (as one would expect if the above mentioned changes in epigenetics are random), there are also some that overlap between all subclones. These include cell growth, cell cycle progression, and cell-cell interactions, and thus are plausible pathways that enable outgrowth into new clones. They also align well with the pathways identified in a recent study to support more efficient clonal outgrowth in cell lines selected for this property. Of interest, in view of the know effect of subcloning to stabilize and “rejuvenate” cells, is the differential regulation of pathways and genes related to DNA repair.

A detailed look at where in the genome these changes took place reveals a strong de-enrichment for regions that are being actively transcribed (as defined by chromatin state annotations that are based on histone marks found)—apparently these regions are being exempt from changes in methylation. Other regions such as quiescent or silenced chromatin states, and also, surprisingly, promoter regions, are not affected, while there is strong enrichment for differential methylation in regions around the transcription start site and in regulatory and enhancer states. The absence of changes in promoter methylation explains the little overlap observed between differentially expressed genes and differentially methylated promoters. Together, this would indicate that phenotypes are mainly controlled by regulation of the strength of expression of a given panel of genes, rather than by complete shutdown of genes or activation of others that were previously silenced. Nevertheless, apart from the known function of promoter methylation as an ON/OFF switch for gene expression, the exact mechanisms and dependencies between DNA methylation, gene regulation and its local context in particular with respect to enhancer regions remain mostly unelucidated.

The caveat on these results is that the chromatin states used were determined for CHO cells of a different lineage than the one used in this study, although cultivated under the same conditions and in the same medium, so that the chromatin states in the current cell line may differ in some genomic regions. A recently in-house performed survey of published transcriptomes for a variety of CHO cells revealed, however, that the differences between most cell lines and subclones or timepoints are in the individual expression levels of genes. The list of genes that are expressed in different cell types is fairly comparable, with only few genes expressed uniquely in some lineages or states, and these typically found at low read counts (unpublished results). Thus, we expect that the majority of the used chromatin states is correctly assigned also for the cell line used in this study.

In conclusion, we here presented data that show that during the process of subcloning the DNA methylation pattern of outgrowing cells is altered in a random way to generate new epigenomes and transcriptomes in the resulting subclones. This process appears to be caused by the cells activating a yet unknown mechanism to actively and randomly change their DNA methylation pattern, resulting in new and unique patterns that were not present in the parental population. These changes contribute to the generation of new and diverse phenotypes in the subclones obtained and thus also support the screening for production clones with improved performance in bioprocessing. Due to the fact that DNA methylation patterns are copied during the division and therefore passed on to daughter cells, the thus obtained new patterns and associated phenotypes are then stable within the new subclone population, at least for a time, possibly until they are again modified due to some other stress situation arising that requires again a modification of the transcriptome.

ACKNOWLEDGMENTS

The authors want to thank Nikolaus Virgolini for support during the Aging study. Next Generation Sequencing was performed by the NGS Facility at the Vienna Biocenter Core Facilities (VBCF), member of the Vienna Biocenter (VBC), Austria. This work was supported through the COMET funding scheme managed by the Austrian Research Promotion Agency FFG with support by BMVIT, BMDW, SFG, Standortagentur Tirol, Government of Lower Austria und Vienna Business Agency. MW and PE received additional support by the PhD program BioToP (Biomolecular Technology of Proteins) funded by the Austrian Science Fund [FWF Project W1224]. MW received further funding by the Austrian Marshall Plan Foundation. Partial support was provided by the NSF AMBIC IUCRC center grant number 1624684 (WT, IL). The authors thank Mike Betenbaugh for initial scientific discussion.

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in NCBI, reference number [PRJEB38542].

AUTHOR CONTRIBUTIONS

Conceptualization, N.B and WT.; Methodology, M.W., G.K., P.E. and I.L.; Software, M.W. and I.L.; Investigation, M.W and G.K; Writing – Original Draft, M.W. and N.B.; Writing – Review & Editing, G.K., P.E., I.L. and W.T.; Funding Acquisition, N.B. and W.T.; Resources, N.B. and W.T; Supervision, N.B. and WT.

ORCID

Nicole Borth https://orcid.org/0000-0001-6324-9338

REFERENCES

1. Ambrosi, C., Manzo, M., & Baubec, T. (2017). Dynamics and context-dependent roles of DNA methylation. Journal of Molecular Biology, Molecular Mechanisms of Stem Cells Pluripotency and Cell Fate Specification, 429, 1459–1475. https://doi.org/10.1016/j.jmb.2017.02.008
47. Rose, N. R., & Klose, R. J. (2014). Understanding the relationship between DNA methylation and histone lysine methylation. Biochim. Biophys. Acta BBA - Gene Regul. Mech., 1839, 1362–1372. https://doi.org/10.1016/j.bbagrm.2014.02.007
48. Rupp, O., MacDonald, M. L., Li, S., Dhiman, H., Polson, S., Grieb, S., Heffner, K., Hernandez, I., Brinkrolf, K., Jadav, V., Samoudi, M., Hao, H., Kingham, B., Goessmann, A., Betenbaugh, M. J., Lewis, N. E., Borth, N., & Lee, K. H. (2018). A reference genome of the Chinese hamster based on a hybrid assembly strategy. Biotechnology and Bioengineering, 115(8), 2087–2100. https://doi.org/10.1002/bit.26722
49. Scarcelli, J. J., Hone, M., Beal, K., Ortega, A., Figueroa, B., Starkey, J. A., & Anderson, K. (2018). Analytical subcloning of a clonal cell line demonstrates cellular heterogeneity that does not impact process consistency or robustness. Biotechnology Progress, 34, 602–612. https://doi.org/10.1002/btpr.2646
50. Schmitz, J., Noll, T., & Grünberger, A. (2019). Heterogeneity studies of mammalian cells for bioproduction: From tools to application. Trends in Biotechnology, 37, 645–660. https://doi.org/10.1016/j.tibtech.2018.11.007
51. Sleiman, R. J., Gray, P. P., McCall, M. N., Codamo, J., & Sunstrom, N.-A. (2008). Accelerated cell line development using two-color fluorescence activated cell sorting to select highly expressing antibody-producing clones. Biotechnology and Bioengineering, 99, 578–587. https://doi.org/10.1002/bit.21612
52. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., & Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences, 102, 15545–15550. https://doi.org/10.1073/pnas.0506580102
53. Tharmalingam, T., Barkhordarian, H., Tejeda, N., Daris, K., Yaghmour, S., Yam, P., Lu, F., Goudar, C., Munro, T., & Stevens, J. (2018). Characterization of phenotypic and genotypic diversity in subclones derived from a clonal cell line. Biotechnology Progress, 34, 613–623. https://doi.org/10.1002/btpr.2666
54. Vcelar, S., JadHAV, V., Melcher, M., Auer, N., Hrdina, A., Sägmeister, R., Heffner, K., Puklowski, A., Betenbaugh, M., Wenger, T., Leisch, F., Baumann, M., & Borth, N. (2018a). Karyotype variation of CHO host cell lines over time in culture characterized by chromosome counting and chromosome painting. Biotechnology and Bioengineering, 115, 165–173. https://doi.org/10.1002/bit.26453
55. Vcelar, S., Melcher, M., Auer, N., Hrdina, A., Puklowski, A., Leisch, F., JadHAV, V., Wenger, T., Baumann, M., & Borth, N. (2018b). Changes in chromosome counts and patterns in CHO cell lines upon generation of recombinant cell lines and subcloning. Biotechnology Journal, 13, 1700495. https://doi.org/10.1002/btj.201700495
56. Vito, D., Erikson, J., Skjådt, C., Weilguny, D., Rasmussen, S. K., & Smale, C. M. (2020). Defining IncRNAs correlated with CHO cell growth and IgG productivity by RNA-Seq. iScience, 23, 100785. https://doi.org/10.1016/j.isci.2019.100785
57. Vito, D., & Smale, C. M. (2020). The long non-coding RNA transcriptome landscape in CHO cells under batch and fed-batch conditions. Biotechnology Journal, 0, 1800122. https://doi.org/10.1002/btj.201800122
58. Walsh, G. (2018). Biopharmaceutical benchmarks 2018. Nature Biotechnology, 36, 1136–1145. https://doi.org/10.1038/nbt.4305
59. Weinguny, M., Klanert, G., Eisenhut, P., Johnson, A., Ivansson, D., Rupp, O., MacDonald, M. L., Li, S., Dhiman, H., Polson, S., Grieb, S., Heffner, K., Hernandez, I., Brinkrolf, K., Jadav, V., Samoudi, M., Hao, H., Kingham, B., Goessmann, A., Betenbaugh, M. J., Lewis, N. E., Borth, N., & Lee, K. H. (2018). Understanding the relationship between DNA methylation and histone lysine methylation. Biochim. Biophys. Acta BBA - Gene Regul. Mech., 1839, 1362–1372. https://doi.org/10.1016/j.bbagrm.2014.02.007
60. Welch, J. T., & Arden, N. S. (2019). Considering “clonality”: A regulatory perspective on the importance of the clonal derivation of mammalian cell banks in biopharmaceutical development. Biologicals, 62, 16–21. https://doi.org/10.1016/j.biologicals.2019.09.006
61. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag.
62. Wu, H., Xu, T., Feng, H., Chen, L., Li, B., Yao, B., Qin, Z., Jin, P., & Conneely, K. N. (2015). Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. Nucleic Acids Research, 43, e141. https://doi.org/10.1093/nar/gkv715
63. Wurm, F., & Wurm, M. (2017). Cloning of CHO Cells, productivity and genetic stability – A discussion. Processes, 5, 20. https://doi.org/10.3390/pr5020020
64. Xu, W., Yu, X., Zhang, J., Bhushan, S., Prasad, S., Prasad, K., Wu, F., Yuan, J., & Poon, H. F. (2019). Soy hydrolysate mimic autocrine growth factors effect of conditioned media to promote single CHO-K1 cell proliferation. Tissue & Cell, 58, 130–133. https://doi.org/10.1016/j.tice.2019.05.005

65. Yim, M., & Shaw, D. (2018). Achieving greater efficiency and higher confidence in single-cell cloning by combining cell printing and plate imaging technologies. Biotechnology Progress, 34, 1454–1459. https://doi.org/10.1002/btpr.2698

66. Zhao, S., Guo, Y., & Shyr, Y. (2019). KEGGprofile: An annotation and visualization package for multi-types and multi-groups expression data in KEGG pathway.

67. Zeyda, M., Borth, N., Kunert, R., & Katinger, H. (1999). Optimisation of sorting conditions for the selection of stable, high producing mammalian cell lines. Biotechnology Progress, 15(5), 953–957.

SUPPORTING INFORMATION

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

How to cite this article: Weinguny M, Klanert G, Eisenhut P, Lee I, Timp W, Borth N. Subcloning induces changes in the DNA-methylation pattern of outgrowing Chinese hamster ovary cell colonies. Biotechnol J. 2021;16:e2000350. https://doi.org/10.1002/biot.202000350