Analytical Subcloning of a Clonal Cell Line Demonstrates Cellular Heterogeneity That Does Not Impact Process Consistency or Robustness

John J. Scarcelli (ORCID)

Cell Line Development, Biotherapeutics Pharmaceutical Sciences, Pfizer Inc., Andover, MA, 01810

Megan Hone

Cell Line Development, Biotherapeutics Pharmaceutical Sciences, Pfizer Inc., Andover, MA, 01810

Kathryn Beal

Cell Line Development, Biotherapeutics Pharmaceutical Sciences, Pfizer Inc., Andover, MA, 01810

Alejaida Ortega

Cell Line Development, Biotherapeutics Pharmaceutical Sciences, Pfizer Inc., Andover, MA, 01810

Bruno Figueroa

Culture Process Development, Biotherapeutics Pharmaceutical Sciences, Pfizer Inc., Andover, MA, 01810

Jason A. Starkey

Analytical Research and Development, Biotherapeutics Pharmaceutical Sciences, Pfizer Inc., Chesterfield, MO, 63017

Karin Anderson

Cell Line Development, Biotherapeutics Pharmaceutical Sciences, Pfizer Inc., Andover, MA, 01810

During development of a cell line intended to support production of an IgG2 monoclonal antibody, a sequence variant caused by a genetic mutation was identified in the bulk drug substance. Gene copy number analysis together with the level of the observed variant in genomic DNA indicated that the master cell bank was a mixed population of cells; some harboring the variant copy and some mutation free. Since the cell bank had been single-cell cloned, this variant could be used as a biomarker to demonstrate either that the bank was not derived from a single cell, or that the variant was a result of a post-cloning genetic event, leading to a mixed population of cells. The sequence variant was only present in a small percentage of subclones, confirming the hypothesis that the cell bank was indeed a mixed population. Interrogation of subclones via Southern blot analysis revealed that almost all subclones had very similar transgene integrant structures, suggesting that the cell bank was likely derived from a single cell, and the cellular event that yielded the sequence variant was a post-cloning event. Further, there were likely several other post-cloning events that impacted transgene loci, leading to a population of related, yet genetically distinct cells comprising the cell bank. Despite this, the heterogeneous bank performed consistently in a bioprocess across generational age with comparable product quality. These results experimentally demonstrate the heterogeneity of a cell bank derived from a single cell, and its relationship to process consistency. © 2018 The Authors Biotechnology Progress published by Wiley Periodicals, Inc. on behalf of American Institute of Chemical Engineers *Biotechnol. Prog.*, 34:602–612, 2018
Keywords: cell line development, clonality, sequence variant

Introduction

Recombinant protein therapeutics, or biologics, are an important class of pharmaceuticals that have transformed the way new therapies are developed for a variety of diseases. A majority of recombinant protein therapeutics are produced in immortalized mammalian cells, in particular Chinese hamster ovary (CHO) cells.[1] CHO cells are able to grow to high cell densities in suspension culture, can be grown in chemically-defined medium, and properly fold, assemble, and modify complex mammalian proteins. These properties have enabled

CHO cells to become the platform of choice for the production of recombinant protein therapeutics.[2]

The general procedure to generate a CHO cell line expressing a recombinant therapeutic is well-established: DNA expression plasmid(s) encoding the transgene(s) of the biologic as well as a selectable marker is transfected into a host cell.[3] The resulting culture is a pool of cells that are typically the product of multiple independent transfection events. Cells arising from different transfection events can differ in the number and chromosomal loci of the integrated expression plasmid(s). These differences can potentially lead to a range of growth characteristics, transgene expression, genetic stability, responses to manufacturing processes, and ultimately differences in final drug substance product quality attributes. In an effort to move forward with cell populations that are less heterogeneous and identify highly productive cell lines, selected cellular pools are subjected to single cell cloning. In these terms, cloning simply means to generate cell lines whose populations are each derived from a single cell; when pools are cloned, cells of the pool are diluted such that there is no more than one cell per vessel (e.g., well of a tissue culture plate). Due to random post-cloning genetic events such as mutations and chromosomal rearrangements, cloning does not result in genetically identical cellular populations. Therefore, when referring to a cell population as a "clone," "clonal," or to assurance of "clonality," it is really referring to having been derived from a single cell, or the assurance that a cell population has been derived from a single cell, as these cellular populations cannot be "clones" or "clonal." These cell lines are allowed to accumulate to a cell mass that can be analyzed for several characteristics (growth, expression, stability, response to feed, product quality, etc.). Clones having desirable attributes are further analyzed, and ultimately a single cloned cell line is used to generate a master cell bank (MCB). As the use of recombinant cells to produce protein therapeutics expanded, ICH guidelines were established to provide broad instruction on standards for the derivation of cell lines to be used in manufacture of biologics.[4] Among the guidance is the requirement for a cloning step, specifically, "for recombinant products, the cell substrate is the transfected cell containing the desired sequences, which has been cloned from a single cell progenitor." This guidance remains relevant, even 20 years later as advancements in cell culture process and drug substance characterization have been realized. Recently, various health authorities have requested greater detail around the cloning step, including (but not limited to), the method used for cloning, calculation of a probability that the cell bank is clonal, and images of the clone at the single cell stage. Cell banks having a low assurance of clonal derivation may ultimately lead to changes in the manufacturing control strategy, which could add cost to product manufacturing, and potentially delays to approval for safe and efficacious therapeutics.[5]

An extensive body of literature suggests that a genetically identical population of cells in a culture originating from a single cell to be a biological impossibility. First, the process of cell division itself can serve as a source of cell population heterogeneity, and is well characterized in the field of cancer biology.[6] In the clonal evolution model, tumors originate from a mutated single cell. As the cell divides, additional mutations arise that result in additional subpopulations.[7] This degree of heterogeneity was identified several years ago in karyotyping performed on several types of mammalian tissue culture cells, where a wide range of chromosome numbers were found even in lines reported as being derived from single cells.[8] The CHO cell was originally characterized as being hardy, reliable, and having grown in continuous cultivation with no observable change in karyotype for several passages.[9] Throughout the past 60 years, differences in cultivation techniques and selection for more favorable growth conditions (e.g., growth in serum free medium) are potential sources of cellular stress, which could in theory lead to mutation. Over this time, the karyotype of the CHO cell has been significantly altered to the point that chromosomal heterogeneity has been observed in cells of a clonal population.[10-13] Previous studies have used genetic sequence variants (unintended amino acid substitutions in a percentage of the bulk drug substance) to gain a better understanding of the genetic plasticity present in cloned cell lines.[14] Intraclonal heterogeneity has also been previously observed phenotypically in clonal CHO cells, with reporter gene expression of clonal cell lines mirroring the level of heterogeneity of a mixed transfection pool only 18 days after cloning.[15]

Although there is extensive literature documenting the complexity surrounding chromosomal abnormalities in cultured CHO cells, to our knowledge there are no reports examining transgene cassette consistency (SNPs, integrant structures, etc.) in CHO cells derived from a single cell, and the relationship of cellular heterogeneity with consistency (cell culture performance and product quality profiles) in an at-scale biomanufacturing process. The purpose of this study is to explore the relationship between cell line clonality (i.e., single cell derivation), cellular heterogeneity, and biomanufacturing consistency. An IgG2 monoclonal antibody (mAb) expressing cell line was generated from what was believed to be derived from a single cell. Later in product development, a low level sequence variant was detected in both the final mAb product and at the genetic level (both DNA and RNA). This cellular population was used to examine the relationship between clonality, population diversity, and bioprocess performance. Using two different cloning methods, the MCB was recloned, with the resulting cell lines representing subclones. The subclones were analyzed for the presence of the sequence variant, as well as for the integration pattern of the heavy chain (HC) and light chain (LC) transgenes via Southern blotting. We also assessed the consistency of the performance of the cell bank in a biomanufacturing process across cellular generational age. Our results demonstrate a range of both genetic and phenotypic heterogeneity in the recombinant protein expressing transgenes of a cell bank derived from a single cell. Despite the observed heterogeneity, we also observed robust performance of the cell bank across a large generational age window.

Results

The presence of a low level genetic variant suggests a mixed population of cells

During clinical progression of an IgG2 subclass mAb (herein called mAb 1), an in-depth analysis of the production MCB was performed as part of laboratory scale cell line and process characterization. This included extensive mRNA and genomic DNA sequencing of the HC and LC transgenes. Polymerase chain reactions (PCRs) intended to amplify the HC and LC transgenes were performed in triplicate on genomic DNA isolated from MCB cells using a high fidelity DNA polymerase. The HC and LC PCR products were gel purified and ligated into a bacterial cloning vector and transformed into *Escherichia coli*. Thirty two independent colonies from each bacterial transformation (representing each PCR) were chosen for plasmid purification and sequencing.

Successful sequencing reactions occurred for 94 independent DNA sub-clones, for both the HC and LC. Of the 94 HC subclones sequenced, 3 revealed an ACC to ATC codon change (~3.2%), which would result in a threonine to isoleucine amino acid substitution at position 253 (T253I; Table 1). In addition to analysis of genomic DNA, a similar clonal sequencing method was also performed on mRNA extracted from MCB cells. Clonal sequencing using RNA as a starting material resulted in 2 out of 383 *E. coli* sub-clones sequenced containing the T253I mutation, confirming the low level variant. There was no detectable evidence of other HC or LC sequence variants.

To enable a more comprehensive assessment, a droplet digital PCR (ddPCR) method was developed to quantify levels of this sequence relative to total HC gene sequence/transcript. This method utilized a FAM-tagged fluorescent probe designed to specifically detect only the T253I codon-containing gene/transcript and a HEX-tagged fluorescent probe that will detect all mAb 1 HC genes/transcripts. The abundance of the T253I species was determined as a percentage of total mAb 1 HC gene/transcript. To develop our method, three mAb 1 MCB vials were independently thawed into separate shake flasks and after 3 days, genomic DNA (gDNA) and RNA were isolated and used to assess levels of the T253I variant species. The sequence variant containing gene was found in ~4% of the gDNA isolates, similar to what was observed using clonal sequencing. Using RNA as a starting material, a transcript variant level of 0.5% was observed, again mirroring what was observed with RNA clonal sequencing. Interestingly, the percentage of variant containing transcripts were approximately eight-fold lower than the percentage of variant containing gDNA, suggesting that the rate of transcription of the gene harboring the variant is lower relative to wild-type copies. Reversed-phase high-performance liquid chromatography (RP-HPLC) subunit/domain analysis of the final mAb 1 bulk drug substance
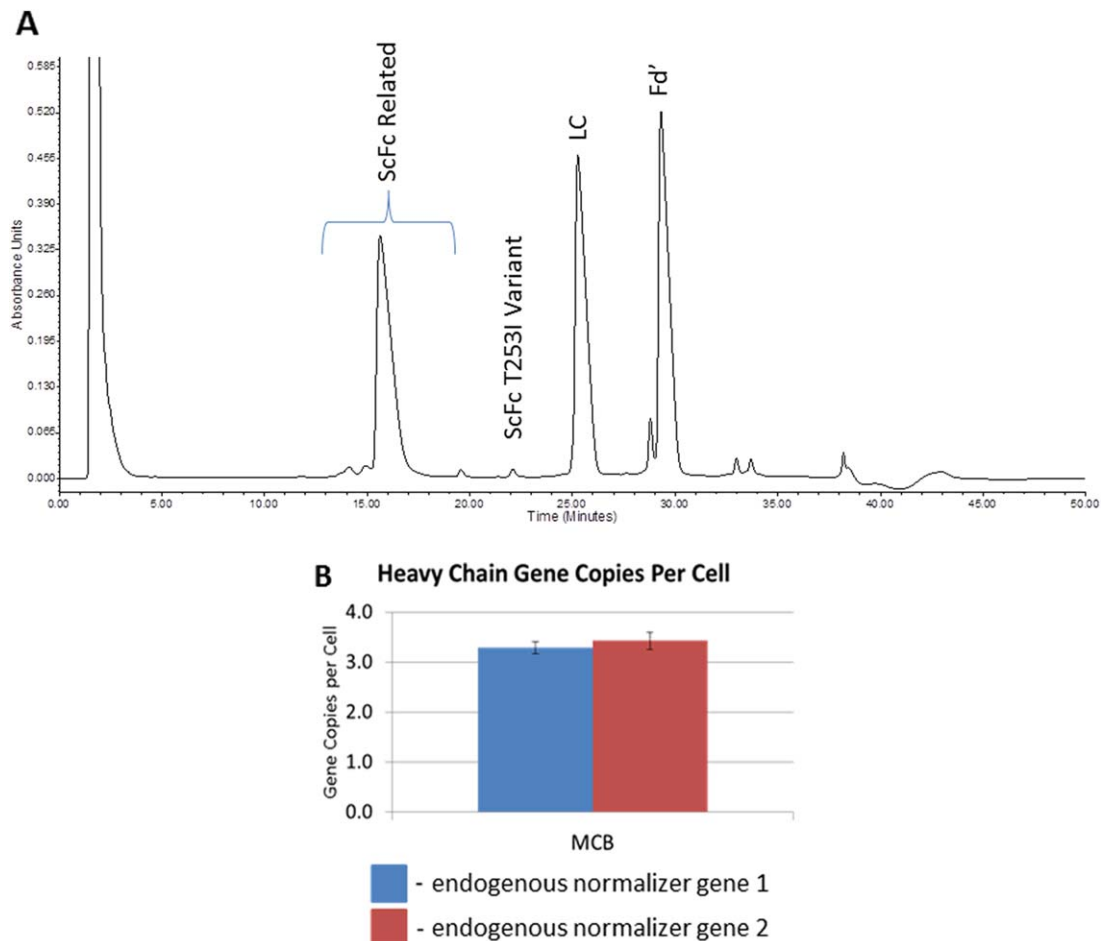
Table 1. mAb 1 Sequence Variant Levels

mAb 1 HC Sequence Variant		
Material	Position	Level
Nucleotide-gDNA	C814T	ddPCR – 4.0% +/− 0.8% CS – 3.2%
Nucleotide-mRNA	C814T	ddPCR – 0.5% +/− 0.1% CS – 0.5%
Bulk Drug Substance	T253I	1.3% – <0.2%

mAb 1 Master Cell Bank sequence variant location and levels within the heavy chain coding region in the MCB nucleic acid sequence and the bulk drug substance. The range in level presented for the variant in the bulk drug substance represents the range observed across several different drug substance lots examined.

Abbreviations: CS, clonal sequencing result; ddPCR, droplet digital PCR result.

Figure 1. The mAb 1 sequence variant is the result of a mixed population of cells, some which harbor a genetic level sequence variant. (A) RP-HPLC subunit domain analysis of mAb 1 drug substance. ScFc: single chain fragment crystalizable domain; Fd′: N-terminal domain of the HC.; T253I Variant: T253I containing ScFc domain (B) Heavy chain gene copies per cell, using two different endogenous genes for normalization.
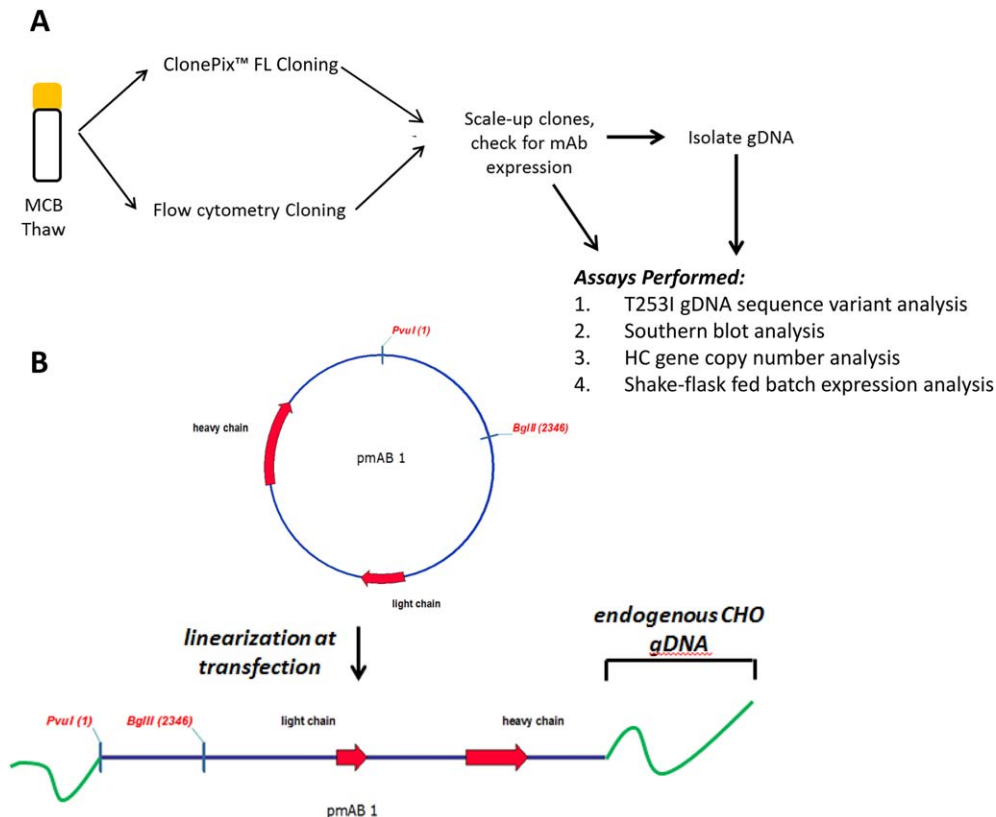
Figure 2. **MCB Subcloning and Southern blot strategy. (A) The MCB subcloning/characterization process. Assays performed on the subclones included T253I variant level analysis in gDNA, Southern blot analysis, heavy chain (HC) gene copy number analysis, and shake-flask fed batch mAb expression analysis. (B) Restriction fragment length polymorphism analysis strategy. Prior to the transfection that generated the MCB, the pmAb 1 expression plasmid was linearized with *PvuI*. DNA from subclones were digested with *BglII* and subjected to Southern blot analysis, probing for either the heavy chain or light chain. The resulting restriction fragment patterns are compared to the MCB.**

independently confirmed the presence of the T253I variant at a low level, similar to that of the transcript (Figure 1A). As the MCB was passaged to accrue generational age and analyzed, the percentage of the sequence variant containing gene decreased over time, which is an observation made previously in other sequence variant containing cell lines.[14]

Prior to MCB generation, this cell line was cloned using a Quixell micromanipulator; an automated cell transfer system where the transfer of a single cell is confirmed microscopically.[16] The protocol used to clone this cell bank was qualified to yield a greater than 96% probability of any individual cell line being derived from a single cell. A possible explanation for a low level sequence variant is that each cell in the population contains numerous integrated transgenes, a small percent of which contain the variant. For example, if the HC gene is present at 25 copies per cell, it is possible that the low level variant is 1 of the 25 (4%) copies. A gene copy number less than 25 would indicate that the variant is in a sub-population of the cells. The gene copy number per cell of the mAb 1 HC gene was 3.4 as determined using ddPCR (Figure 1B). This observation, together with the observed 4% level of the T253I variant at the gDNA level strongly suggests that the MCB is a mixed population of cells, some containing at least one copy of the HC variant containing gene, and some cells with no copies of the variant. It should be noted that during cell line generation for the mAb 1 cell line used to produce the MCB, no amplification step was employed.

Analytical subcloning to determine cell bank clonality

There are two possible causes of the heterogeneous MCB cell population containing the variant HC gene: (i) the cell bank was not derived from a single cell or (ii) after single cell cloning but prior to the establishment of the MCB, the nucleotide mutation occurred in one cell, and the cell propagated into a sub-population of the MCB. To determine which of these two scenarios occurred, the MCB was single cell cloned using two different cloning methods (Figure 2A). The first method used was flow cytometry. After sorting, two plates were stained with dye and sacrificed to determine the single cell plating efficiency. Of the 192 wells assessed, 122 were occupied with cells, and only 2 of those 122 wells occupied had more than one cell; a 98.3% single cell efficiency. The second method utilized the ClonePix FL instrument. The ClonePix method involves suspension of cells in a semi-solid matrix, after which cells are incubated for several days in order to generate colonies. Colonies are then picked by the instrument and moved to 96-well plates. It is not possible to determine single cell plating efficiency since several cells for each clone are moved into a well. Thus, the probability of clonality could not be established for this method.

In total, 192 wells containing Clonepix subclones and 160 wells containing flow cytometry subclones were screened for mAb 1 titer. Of the 192 ClonePix subclones, 166 were expressing mAb 1 (data not shown). The ClonePix method used here did not incorporate the use of fluorescence to enrich for producing subclones, they were only picked by
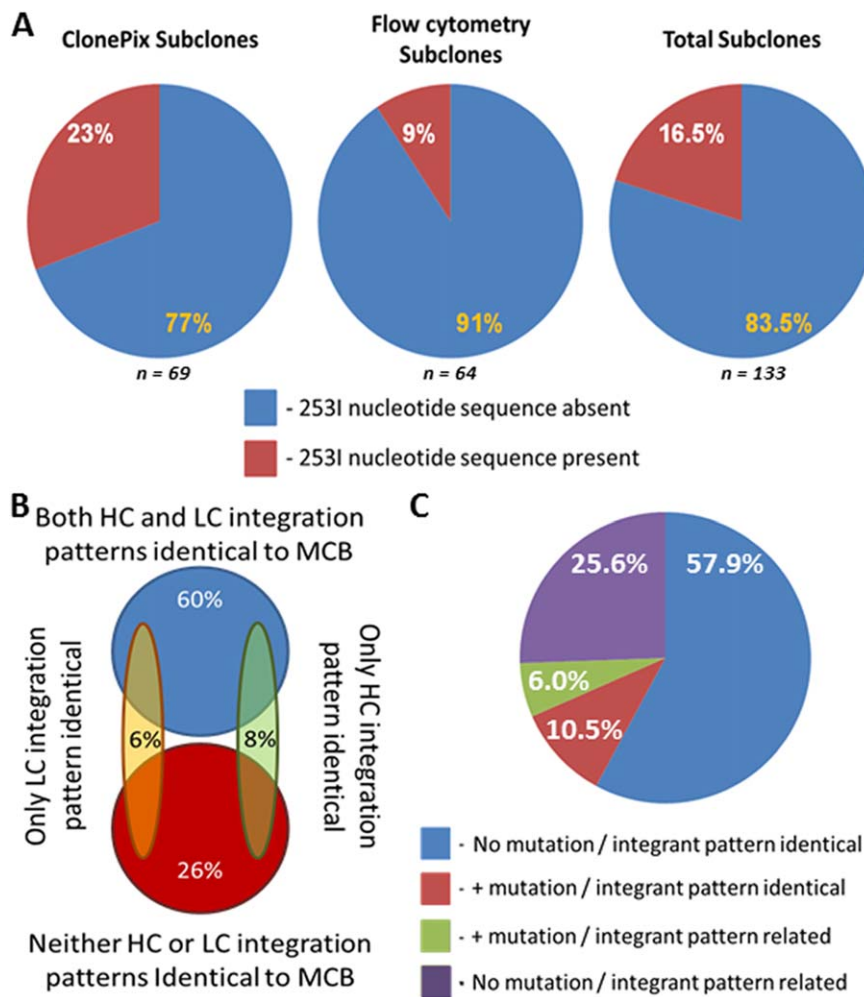
Figure 3. Analytical Subcloning sequence variant and integration pattern data summary. (A) Percentages of subclones harboring the genetic sequence variant, separated by cloning method. (B) Percentages of subclones having either an identical or related integrant banding pattern to the MCB when probing for either heavy chain (HC) or light chain coding sequence. (C) Percentages of subclones that match/do not match the MCB HC Southern banding pattern versus presence/absence of T253I sequence variant.

white light to ensure that colonies chosen were well separated and of uniform shape. Interestingly, all 160 of the analyzed flow cytometry colonies were expressing mAb 1 (data not shown). For further analysis, 69 ClonePixFL expressing subclones and 64 flow cytometry expressing subclones were randomly chosen for scale up to shake flasks, followed by genomic DNA isolation to test for the presence of the gene containing the T253I variant via ddPCR. Of the 69 ClonePix clones that were analyzed, 16 contain the T253I sequence variant, while 6 of the 64 flow cytometry clones contain the variant (∼16.5% of all subclones contained the variant, Figure 3A). This result supports our hypothesis that the MCB is a heterogeneous population of cells with respect to the sequence variant, as the variant was not detected in all subclones.

Southern blot analysis of the subclones and MCB was used to investigate MCB heterogeneity and determine if there is correlation between the sequence variant and any other observed heterogeneity. During the early stages of MCB cell line generation, the mAb 1 expressing plasmid was linearized via digestion with the restriction endonuclease *PvuI* prior to transfection and integration into the host cell genome. As shown in Figure 2B, digestion of gDNA from MCB subclones with the restriction endonuclease *BglII* will

yield genomic restriction fragments containing both a portion of the expression plasmid, as well as host genomic sequences beyond the *PvuI* site at the 3′ end of the fragment. Given that the expression plasmid integrated randomly into the host cell genome during cell line generation, the number of integration events, the position(s) of the *BglII* site(s) adjacent to the integrated expression plasmid(s), as well as the total size of the restriction fragment(s), will also be random. However, since this is a stable integration the number and sizes of the restriction fragments after integration will remain static. If the production cell line used to produce the MCB was derived from a single cell, all of the resulting subclones should have restriction fragment patterns (banding signatures) related to that of the MCB. It is noteworthy that other groups have observed differences in chromosome architecture over time, which could lead to differences in integrant banding patterns.[10,11] The MCB HC 3′ integration banding signature is comprised of two bands, both migrating between 9 and 23 kilobases (Supporting Information Figure S1A). The LC 3′ banding signature also includes two bands migrating between 9 and 23 kb, plus a third band migrating at ∼6 kb (Supporting Information Figure S1B). Genomic DNA was digested from all 133 subclones with *BglII*, and digests were analyzed via Southern blotting, probing for both the
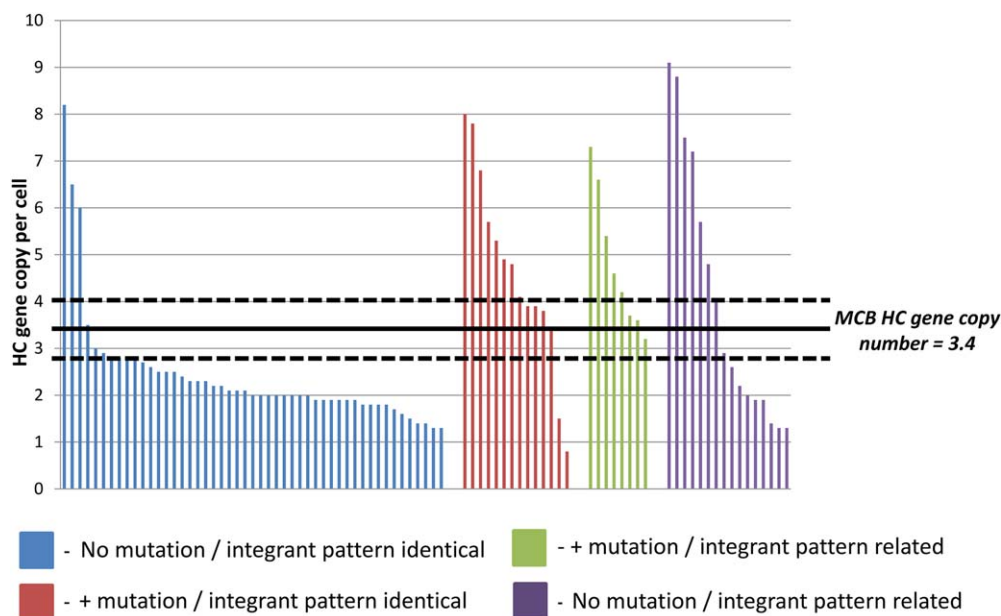
Figure 4. Heavy chain gene copy analysis. HC gene copy per cell results for the subclones are grouped by presence/absence of the T253I genetic variant, and HC integrant Southern blot result. HC gene copy number for the mAb 1 MCB is ∼3.4 copies per cell; dashed lines represent a three standard deviation range.

HC and the LC, and the banding pattern was compared to that of the MCB. A subclone with an identical banding pattern was one that had only bands migrating at the same positions as the MCB, while a subclone with a related banding pattern was one that had at least one band migrating at the same position as the MCB. Of the 133 subclones analyzed, 132 had a banding signature that was related to the MCB. The one subclone that was not related to the MCB had a single band for both the HC and LC integration Southern blots. While this band did not migrate with any of the bands in the MCB in either the HC or LC blots, it did co-migrate with bands of other subclones that had a related pattern to the MCB, suggesting that this subclone was still likely related to the other subclones, as well as the MCB. Since all 133 subclones displayed a degree of relatedness to the MCB, the data strongly suggests that the MCB was derived from a single cell.

Examination of post-single cell cloning heterogeneity

While nearly all subclones analyzed by Southern blot analysis appear to be related to the MCB, ∼40% of the subclones displayed banding patterns that did not exactly match the MCB (Figure 3B). Of those that did not match, some matched for either the HC (11 of 133 total subclones) or the LC (8 of 133 subclones), and some matched for neither HC nor LC (34 of 133 subclones) (Figure 3B). These results suggest that in addition to the presence/absence of the T253I variant, there are other signatures of cellular heterogeneity pertaining to the HC and LC expressing transgenes, specifically with respect to minor rearrangements of integrant structure.

To further examine the degree of cellular heterogeneity for this cell bank, we determined if there was a correlation between HC banding pattern and the presence of the T253I variant. All MCB subclones that contained the T253I variant were examined for the presence of a 3' integration banding pattern that matched the MCB. Little correlation between the

HC integration banding pattern and the presence of a sequence variant is observed (Figure 3C), suggesting that the events leading to the rearrangements observed for the transgene expression cassettes were different than the event(s) leading to the generation of the sequence variant. Interestingly, subclones with integration patterns matching the MCB were observed both with and without the sequence variant, further suggesting that the presence of the sequence variant containing cell population were not the result of failure to derive the MCB from a single cell, and rather the result of a post-cloning genetic event.

A subset of the subclones was analyzed for HC gene copy number (Figure 4). This subset of clones represents those with and without the T253I variant, as well as clones with matching and related HC integrant banding patterns. The coefficient of variation for the mAb 1 HC gene copy assay is 4.8%, and the standard deviation was 0.2 copies per cell. Interestingly, the majority of subclones had HC gene copy numbers significantly different from that of the MCB, even when using a three standard deviation range with respect to the mAb 1 HC gene copy number assay (Figure 4). The vast majority of subclones lacking the T253I variant coding sequence had HC gene copy numbers lower than the MCB, while subclones harboring the variant coding sequence had higher gene copy numbers. This result suggests that a mutation causing the sequence variant may have been associated with a gene duplication event very early after single cell cloning. The above results confirm a wide array of genetic heterogeneity within the mAb 1 clonal MCB.

Phenotypic performance for subclones with integration pattern identical to the MCB

There were 32 subclones of the MCB identified that were (i) sequence variant free, and (ii) had HC and LC integration patterns identical to the MCB. To evaluate phenotypic performance, these 32 subclones were subjected to a 12-day
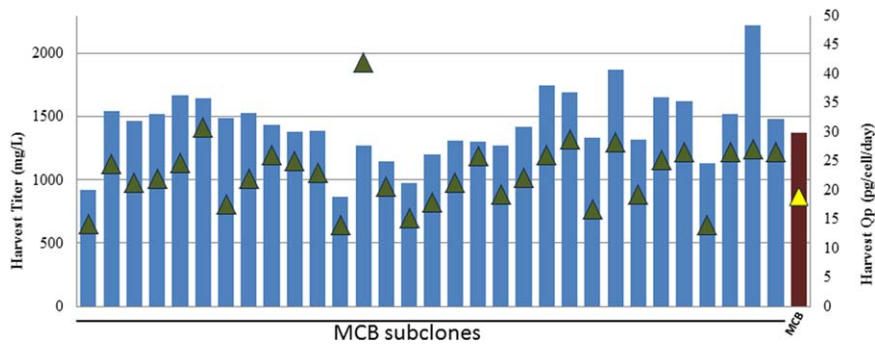
Figure 5. Shake flask fed batch analysis of a subset of subclones. Several subclones that were sequence variant free and had identical heavy chain and light chain integration Southern blot banding patterns were subjected to a 12-day fed batch expression analysis. The MCB was also included as a control. Harvest titer (bars) and harvest specific productivity (Qp, triangles) are presented.
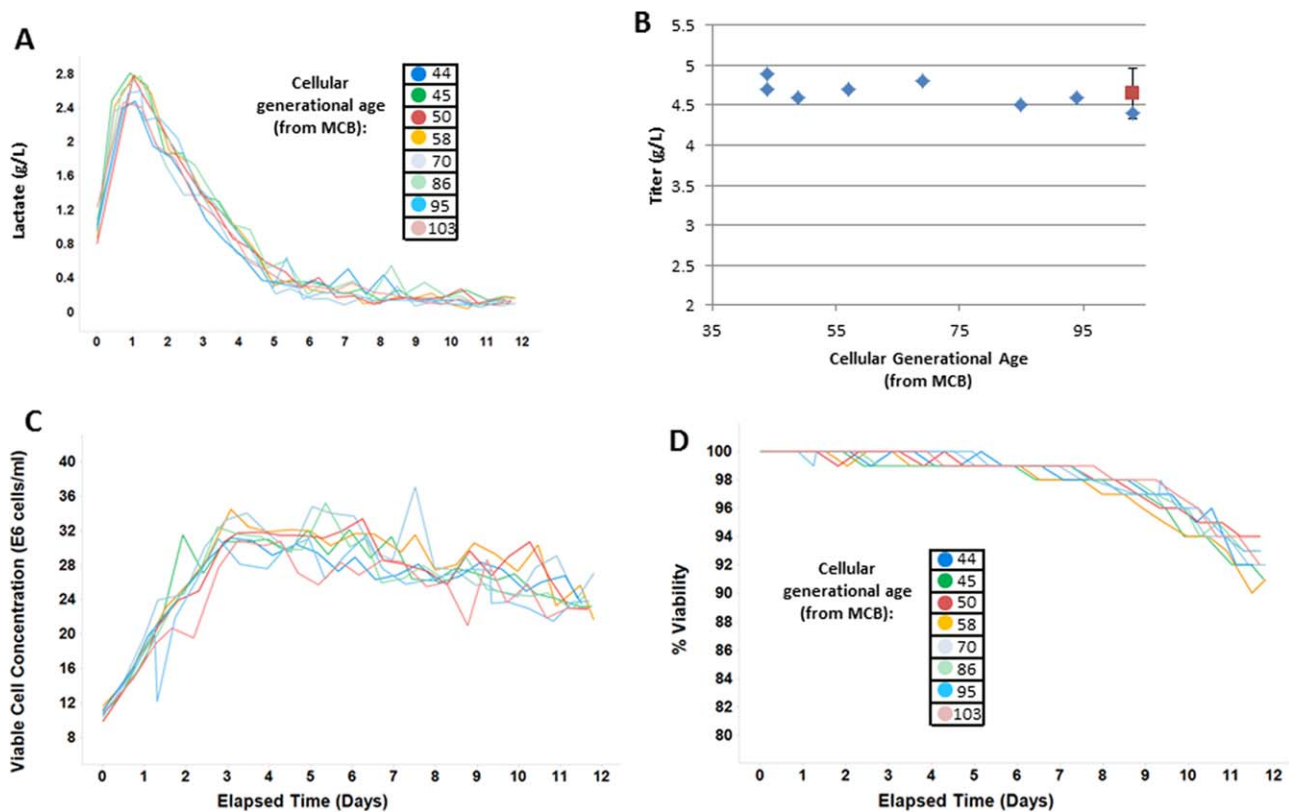

Figure 6. The heterogeneous MCB performs consistently in a bioprocess over a large cellular generational age window. (A) Lactate accumulation over days in culture, (B) harvest titer, (C) viable cell concentration, and (D) viability for the mAb 1 working cell bank using the same at scale upstream process across a range of cellular generational age. For (B), blue diamonds represent harvest titer for eight drug substance batches, while the red square is the average titer for all eight batches. Error bars represent two standard deviations of the average.

shake flask fed batch analysis (Figure 5). In spite of the fact that these cells were subcloned from the same MCB, and were similar with respect to LC and HC integrant structure, a range of expression performance was still observed. The subclones ranged in titer from 0.9 to 2.2 g/l, and specific productivity (Qp) ranged from 14 to 42 pg/cell/day. In the same experiment, titer for the MCB in the fed batch was 1.4 g/l, and the Qp was 19 pg/cell/day. This experiment demonstrates that even subclones with matching HC and LC Southern blot integrant structures can exhibit different phenotypic performance, and that other changes are occurring independent of transgene genetic heterogeneity.

Bioprocess performance of the heterogeneous cell bank

The bioprocess performance consistency of the mAb 1 MCB was assessed. When a non-clonal cell bank is used in a bioprocess, an expected outcome might be inconsistency over generational age of both phenotypic attributes as well as mAb product quality attributes. In order to maintain consistency, the control strategy may necessitate shortening of the cellular generational window used to run the bioprocess when a non-clonal cell bank is used.

The mAb 1 MCB was passaged for ∼20 generations and vials of cells were frozen in order to create a working cell bank (WCB). The WCB was used to generate several
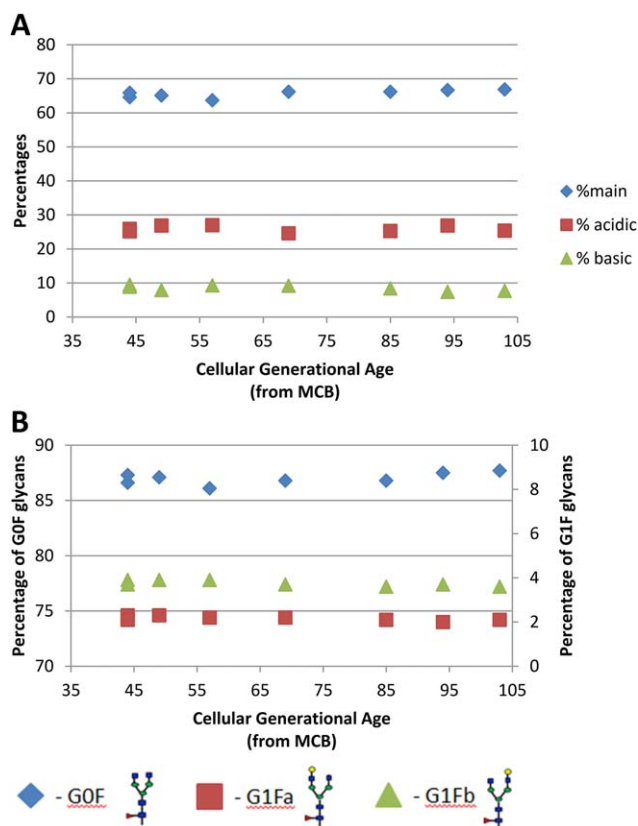
Figure 7. (A) Charge heterogeneity and (B) N-glycosylation of mAb 1 drug substance produced from the mAb 1 Working Cell Bank across a range of cellular generational ages.

independent 12,000 L development batches of drug substance. To assess the impact that generational age of this bank may have on process consistency, the performance of eight production batches were compared. These batches used the WCB passaged to various generational ages as an inoculum source for bioreactors using the same in-process parameters. Various bioprocess performance metrics were assessed to determine whether or not cellular age influences WCB performance (Figure 6 and Supporting Information Figure S2). Cell growth and viability were not impacted due to generational age. For harvest day titer (Figure 6B) and volumetric productivity (Supporting Information Figure S2), while there is a slight decrease for both of these across generational age, the values for the batches inoculated with the oldest generational age are still within two standard deviations of the average of all of the batches. A decrease of this magnitude is more likely due to variance between batches rather than caused by a drift in cellular populations. These results demonstrate that phenotypically, the mAb 1 cell bank performs consistently across generational age, up to at least 103 generations from MCB manufacture. It should be noted that the range of generational ages evaluated here is not typical; normally the generational age would be more limited (as short as ∼45 generations from the MCB), either due to phenotypic performance of the cell line, or lower expected commercial demand.

Product attributes of mAb 1 drug substance generated from the eight 12,000-L scale development batches were analyzed and compared for consistency and quality. The charge heterogeneity, N-linked glycan profile (Figure 7), as well as the percentage of high molecular mass species (HMMS) and antibody integrity/fragmentation (Supporting

Information Table S1) were measured. For all attributes examined, no correlation of changes in product quality with generational age of the cells from which the material was generated was observed. Taken together with the phenotypic data, the results demonstrate that the mAb 1 MCB is capable of consistent and robust performance in a large scale bioprocess.

Conclusions

The mAb 1 expressing MCB described here is a heterogeneous population of cells, initially determined through discovery of a low level sequence variant in the final drug substance. This was further confirmed by integration Southern blot and transgene copy number analysis of MCB subclones. While the MCB was likely derived from a single cell, there also appear to be other post-single cell cloning events that result in the MCB being comprised of several related, yet distinct, sub-populations of cells. In spite of the heterogeneous nature of the MCB, as well as several reported instances of genetic plasticity of CHO cells,[12,13,17] the bank performed robustly and consistently in a large scale bioprocess for several generations after establishment.

While this MCB was likely derived from a single cell, at the time of MCB generation the bank was no longer homogenous. This is not surprising given the fact that every cell division has the opportunity to introduce some genetic variability via mitosis and polymerase errors. For example, the error rate of a mammalian DNA polymerase with proofreading and mismatch repair capabilities is approximately one of every 1×10^9 bases.[18] Given the estimated size of the CHO cell genome (2.6×10^9 bases),[19] it is expected that the average cell division would yield ∼2.6 polymerase errors. Errors during mitosis are another opportunity for introduction of genetic variation. Given the integration Southern blot analysis results observed here for the subclones, as well as reports of gross aneuploidy in CHO cell lines,[17] this may be the more prevalent error source when considering the transgene integrant sites. The mAb 1 MCB accrued ∼115 generations in the time between single cell cloning and MCB production, (considering an average doubling time of 20.8 h and ∼100 days in culture between single cell cloning and MCB production). During this time, $\sim 2 \times 10^{34}$ cell divisions would have occurred, meaning that there is ample opportunity for mutation to arise post cloning.

Perhaps the cell culture environment is playing a role in the introduction of mutation. A cell culture flask is obviously a much different environment than the one native to a hamster ovary, and sequencing of the CHO cell genome revealed a marked divergence from the hamster genome.[10,11] During electroporation-mediated transfection the cellular membrane is damaged, resulting in lipid peroxidation and generation of reactive oxygen species.[20] Lipid peroxidation produces compounds capable of forming adducts that are known to damage DNA, which can introduce mutations.[21,22] Flow cytometry subjects cells to hydrodynamic forces, which may result in significant cell damage,[23] resulting in potential mutagenesis. One could argue that our observations here were due to mutagenesis the MCB was exposed to during analytical subcloning, and are therefore artifacts of the subcloning step. This however is unlikely, as two different subcloning methods (flow cytometry and ClonePix) were utilized which work by very different mechanisms. Some of the Southern integration blot patterns observed that did not

match the MCB for subclones cloned by both the ClonePix and flow cytometry appeared to be the same (data not shown), suggesting that the subpopulation having at least those particular banding pattern signatures arose prior to the independent MCB subcloning events.

Although sometimes used to augment assurance of clonality, the process of analytical subcloning is not without its drawbacks. Data generated from an analytical subcloning effort will likely reveal what was suspected initially and what has been demonstrated here; that the cell bank is indeed a mixed population of cells, whether or not the bank was originally derived from a single cell. Further, when data suggests that all subclones are related to the parental bank, the method provides support of single cell derivation. However when this does not happen, it is impossible to rule out that differences observed may be due to a cellular event that occurred after the original single cell cloning event, or even due to the process of analytical subcloning itself. Due to this, analytical subcloning is of questionable utility for routine monitoring of clonality assurance.

Despite the observed heterogeneity, the cell bank was capable of consistent performance in a large scale bioprocess over a range of commercially representative generational ages. Cell bank heterogeneity can potentially lead to a subpopulation of cells either performing better under certain bioprocess parameters than other subpopulations, or a subpopulation of cells outgrowing the rest of the cell population over generational age.[17] These events can cause the bank as a whole to exhibit significant changes in process performance across different parameters or over cellular generational age. To ensure that bank heterogeneity does not impact process consistency and is therefore suitable for commercial manufacturing, it is good practice during later stage development to assess cell banks, both genetically and phenotypically in the intended commercial upstream manufacturing process (at laboratory scale, prior to process qualification), across generational age. Consistent bioprocess performance does not imply that the bank is homogeneous/clonal, it simply means that the amount of heterogeneity present is not enough to significantly impact process consistency. Observations of inherent heterogeneity in cell lines that were originally derived from a single cell, both in this study and throughout the literature,[10,12,17] calls into question the level of scrutiny one should place around the single cell cloning step when generating a cell line for eventual production of a biotherapeutic.

In conclusion, while single cell cloning is indeed a useful step for minimizing cell culture heterogeneity, our data demonstrates that it does not eliminate cellular heterogeneity. Focusing instead on the development of a robust bioprocess which yields consistent phenotypic performance and biotherapeutics product quality attributes will allow organizations to effectively ensure satisfactory product supply.

Materials and Methods

Cell culture

Routine cell culture passaging took place in CD-CHO media (Life Technologies, #10743-029) supplemented with 25 µM L-methionine sulfoximine (MSX, Millipore, GSS-1015-F). Cells were grown in incubators set to 36.5°C, 5% CO_2, 80% humidity, and shaking at 140 rpm with an orbital diameter of 1 inch. Single cell cloning via flow cytometry

was performed as previously.[24] The sort was performed on a FACS Aria II cell sorter equipped with FACSDiva v8.0.1 software. For ClonePix cloning, a ClonePix FL was used. Cells were cultured in 6-well plates containing semisolid medium. Colonies were screened by white light to ensure that they were distinct and well separated from other colonies.

Cell line generation for the production cell line used to produce the mAb 1 MCB consisted of electroporation of a CHO K1 derived host with a plasmid encoding both the mAb 1 HC and LC. Expressing cells were selected for using 50 µM MSX. Shortly after cloning, the concentration of MSX was reduced to 25 µM. No amplification step was ever performed.

Genomic DNA and RNA isolation

Genomic DNA sample preparation was performed using Qiagen DNeasy Blood and Tissue kit (Qiagen #69506). Total RNA sample preparation was performed using Qiagen RNeasy Kit (Qiagen #74106).

Clonal sequencing

Total RNA was isolated from mAb 1 MCB cells, then converted to cDNA using the Superscript III First Strand Synthesis System with oligo dT primers (Life Technologies, 18080-051). Each cDNA sample was used as template for PCR of the mAb 1 heavy and light chain open reading frames. Each PCR product was subcloned into a pBluescript vector and transformed into *E. coli*. Ninety-six colonies from each transformation event were used as template for Sanger sequencing to yield double-stranded sequence for each insert. Results were analyzed using Sequencher software.

Sequence variant and gene copy analysis using droplet digital PCR (ddPCR)

ddPCR methods were performed as described previously.[14] Briefly, 1 µg of gDNA template for each sample was digested with *CviQI*. Reaction mixes were made using ddPCR Supermix (Biorad, #186–3010) and primer/probe sets specific for either the mutation harboring HC gene and total HC gene (for sequence variant analysis) or the total HC gene and an endogenous CHO gene (for HC gene copy number analysis). For sequence variant analysis, subclones having a signal above the host negative control for the variant detecting probe were counted as positive for the presence of the variant. For HC gene copy number analysis, the *ACTR5* gene was routinely used as the endogenous gene for normalization, as it was previously determined to have two copies per cell by next generation sequencing and ddPCR. The ratio of the total heavy chain probe signal to *ACTR5* probe signal, multiplied by two yields the number of copies of the mAb gene in one cell. In Figure 1B, the CHO endogenous gene *COG1* was also used for normalization, which had also been determined to be present at two copies per cell.

Southern blot

5 µg for each genomic DNA sample were used per digest. As a control, expression plasmid DNA at 100% relative to MCB genomic DNA levels was spiked into 5 µg of host genomic DNA. These samples were digested in the same

manner, to allow comparison with the predicted expression plasmid fragments. The digested DNA samples were size-fractionated on 0.8% agarose gels, transferred to positively-charged nylon membranes and cross linked by UV irradiation. Membranes were pre-hybridized in a formamide-based hybridization buffer. They were subsequently hybridized in the same solution with a target-specific denatured [^{32}P]-labeled probe. Probe templates for mAb 1 HC and LC were generated by PCR using appropriate primer sets. Probes were labeled with dCTP, (α-^{32}P; Perkin Elmer, BLU513H) using the Prime-It RmT Random Primer Labeling Kit (Stratagene, 300392), followed by removal of free nucleotides using Quick Spin Columns for radiolabeled DNA purification Sephadex G-50 (Roche, 11–273-973-001). After hybridization, the membranes were taken through a series of increasingly stringent washes and then exposed to a phosphorimage screen.

Shake Flask Fed Batch

For fed-batch shake flask analysis, cells were seeded at 3×10^5 cells ml^{-1} in 250-ml shake flasks containing 50 ml of proprietary medium. Cells were fed starting on day 3 of the culture with a proprietary feed consisting of a mixture of amino acid and trace elements. Daily viabilities and viable cell concentrations were determined using a Vi-CELL automated cell viability analyzer. Antibody titers in the medium were determined by bio-layer interferometry.

Fed-Batch Production Bioreactor

Fed-batch Production bioreactors were performed at 12,000-L scale. The bioreactor was inoculated at 6–10×10^6 viable cells/ml using an N-1 perfusion bioreactor. Starting on day 0, concentrated proprietary feed media and glucose were added. All vessels maintained continuous control of pH using 1 N sodium carbonate. Daily offline gas and pH measurements were performed with a Blood Gas Analyzer. Dissolved oxygen was controlled at 40% of air saturation and temperature was controlled at 36.5°C. Viable cell density, glucose and lactate levels were determined for samples collected daily using a NOVA Flex BioProfile. Conditioned medium collected was analyzed for the monoclonal antibody titer by Protein A high performance liquid chromatography (HPLC).

Analytical Methods

For imaged capillary isoelectric focusing (iCE), protein samples were prepared and analyzed using an iCE280 Analyzer (ProteinSimple) per procedures described previously.[25] iCE separates protein species based on their charge differences in a pH gradient generated by ampholytes under the influence of an electric field. Protein charged species are focused within a capillary under DC voltage and detected at 280 nm with whole capillary imaging. For size exclusion HPLC (SEC), ~50 μg of each protein sample was injected onto a YMC-Pack Diol-200 size-exclusion column (300 × 8 mm) maintained at 30°C. High molecular mass species (HMMS) and monomer were separated using isocratic elution with a mobile phase containing 20 mM sodium phosphate and 400 mM sodium chloride, pH 7.2.

The procedure for measuring intact and fragmented antibody levels was a previously described capillary gel electrophoresis technique.[26] Samples were denatured with SDS and heated with or without a reducing agent. The protein samples were reduced into heavy and light chains which were electrophoretically separated in a capillary containing sieving medium and detected using UV absorbance, using a Beckman Coulter Proteomelab PA800 instrument. The separation allows quantitation of the resolved heavy and light chains as well as size related impurities. The purity was reported as the time corrected total percent of heavy and light chains.

N-glycan profiling was performed using HILIC as described previously.[27] Briefly, N-glycans were released from the mAb using PNGase F, labeled with a fluorescent probe 2-aminobenzamide (2-AB), separated using hydrophilic interaction chromatography (HILIC) with a Waters Xbridge amide column, and quantified using fluorescence detection.

The quantitation of T235I sequence variant levels was performed by separating the Fc domains by HPLC following IdeS digestions.[28] Approximately, one unit of IdeS (FabRICATOR Genovis; 59 unit/μl) was added to each μg of protein. The protein/enzyme mixture was then diluted with a dilution buffer (50 mM sodium phosphate, 150 mM NaCl, pH 6.6) to a final protein concentration of 2 mg/ml and incubated at 37°C for 60 min. The IdeS digested sample was reduced by 125 mM DTT and 4 M GuHCl at 37°C for 45 min.

The RP HPLC separation of IdeS digested and DTT reduced samples was performed on a BEH300 C4 column (100 × 2.1 mm, 1.7 μm, Waters, Inc.). The column was held initially at 69.8% mobile phase A (0.1% TFA in 10% IPA) and 30.2% mobile phase B (0.1% TFA in 10% IPA and 90% ACN) for 2 min. Separation was achieved with a linear gradient from 30.2% B to 35.0% B in 30 min. The column temperature and flow rate were maintained at 65°C and 0.2 ml/min, respectively. Typically, 7.5 μg of sample was injected onto the column for analysis.

Literature Cited

1. Zhu J. Mammalian cell protein expression for biopharmaceutical production. *Biotechnol Adv.* 2012;30:1158–1170.
2. Mohan C, Kim YG, Koo J, Lee GM. Assessment of cell engineering strategies for improved therapeutic protein production in CHO cells. *Biotechnol J.* 2008;3:624–630.
3. Wurm FM. Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol.* 2004;22:1393–1398.
4. International Conference on Harmonization of Technical Requirement for Registration of Pharmaceuticals for Human Use, ICH tripartite guidelines for Biological products Q5D. 1997.
5. Frye C, Deshpande R, Estes S, Francissen K, Joly J, Lubiniecki A, Munro T, Russell R, Wang T, Anderson K. Industry view on the relative importance of "clonality" of biopharmaceutical-producing cell lines. *Biologicals.* 2016;44:117–122.
6. Greaves M, Maley CC. Clonal evolution in cancer. *Nature.* 2012;481:306–313.

7. Nowell PC. The clonal evolution of tumor cell populations. *Science*. 1976;194:23–28.

8. Hsu TC. Chromosomal evolution in cell populations. *Int Rev Cytol*. 1962;12:69–161.

9. Puck TT, Cieciura SJ, Robinson A. Genetics of somatic mammalian cells III. Long-term cultivation of euploid cells from hujan and animal subjects. *J Exp Med*. 1958;108:945–959.

10. Deaven LL, Petersen DF. The chromosomes of CHO, an aneuploid Chinese hamster cell line: G-band, C-band, and autoradiographic analyses. *Chromosoma*. 1973;41:129–144.

11. Wurm FM, Hacker D. First CHO genome. *Nat Biotechnol*. 2011;29:718–720.

12. Wurm FM. CHO quasispecies – Implications for manufacturing processes. *Processes*. 2013;1:296–311.

13. Vcelar S, Jadhav V, Melcher M, Auer N, Hrdina A, Sagmeister R, Heffner K, Puklowski A, Betenbaugh M, Wenger T, Leisch F, Baumann M, Borth N. Karyotype variation of CHO host cell lines over time in culture characterized by chromosome counting and chromosome painting. *Biotechnol Bioeng*. 2018;115:165–173.

14. Lin TJ, Beal KM, DeGruttola HS, Brennan S, Marzilli LA, Anderson K. Utilization of sequence variants as biomarkers to analyze population dynamics in cloned cell lines. *Biotechnol Bioeng*. 2017;114:1744–1752.

15. Pilbrough W, Munro TP, Gray P. Intraclonal protein expression heterogeneity in recombinant CHO cells. *PLoS One*. 2009;4:e8432.

16. Wewetzer K, Seilheimer B. Establishment of a single-step hybridoma cloning protocol using an automated cell transfer system: Comparison with limiting dilution. *J Immunol Meth*. 1995;179:71–76.

17. Wurm FM, Wurm MJ. Cloning of CHO cells, productivity and genetic stability – A discussion. *Processes*. 2017;5:20.

18. McCulloch SD, Kunkel TA. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res*. 2008;18:148–161.

19. Xu X, Nagarajan H, Lewis NE, Pan S, Cai Z, Liu X, Chen W, Xie M, Wang W, Hammond S, Andersen MR, Neff N, Passarelli B, Koh W, Fan HC, Wang J, Gui Y, Lee KH, Betenbaugh MJ, Quake SR, Famili I, Palsson BO, Wang J. The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat Biotechnol*. 2011;29:735–741.

20. Gabriel B, Teissie J. Generation of reactive-oxygen species induced by electropermeabilization of Chinese hamster ovary cells and their consequence on cell viability. *Eur J Biochem*. 1994;223:25–33.

21. Marnett LJ. Lipid peroxidation-DNA damage by malondialdehyde. *Mutat Res*. 1999;424:83–95.

22. Maccarrone M, Bladergroen MR, Rosato N, Finazzi AAF. Role of lipid peroxidation in electroporation-induced cell permeability. *Biochem Biophys Res Commun*. 1995;209:417–425.

23. Mollet M, Godoy-Silva R, Berdugo C, Chalmers JJ. Computer simulations of the energy dissipation rate in a fluorescence-activated cell sorter: Implications to cells. *Biotechnol Bioeng*. 2008;100:260–272.

24. Zhang L, Inniss MC, Han S, Moffat M, Jones H, Zhang B, Cox WL, Rance JR, Young RJ. Recombinase-mediated cassette exchange (RMCE) for monoclonal antibody expression in the commercially relevant CHOK1SV cell line. *Biotechnol Prog*. 2015;31:1645–1656.

25. He XZ, Que AH, Mo JJ. Analysis of charge heterogeneities in mAbs using imaged CE. *Electrophoresis*. 2009;30:714–722.

26. Lacher NA, Roberts RK, He Y, Cargill H, Kearns KM, Holovics H, Ruesch MN. Development, validation, and implementation of capillary gel electrophoresis as a replacement for SDS-PAGE for purity analysis of IgG2 mAbs. *J Sep Sci*. 2010;33:218–227.

27. Shang TQ, Saati A, Toler KN, Mo J, Li H, Matlosz T, Lin X, Schenk J, Ng C-K, Duffy T, Porter TJ, Rouse JC. Development and application of a robust N-glycan profiling method for heightened characterization of monoclonal antibodies and related glycoproteins. *J Pharm Sci*. 2014;103:1967–1978.

28. An Y, Zhang Y, Mueller HM, Shameem M, Chen X. A new tool for monoclonal antibody analysis: Application of IdeS proteolysis in IgG domain-specific characterization. *MAbs*. 2014;6:879–893.