White paper

Targeted Locus Amplification and NGS combined with qPCR-based breakpoint analysis for the assurance of monoclonality in recombinant cell lines

Judith GM Bergboer[1], Martijn JE Kelder[1], Max van Min[1], Nika Tuta[2], Mitja Crček[2], Matjaz Vogelsang[2]

[1] Cergentis, Utrecht, The Netherlands
[2] Novartis, Biologics Technical Development TRD, Mengeš, Slovenia

Abstract

Recombinant protein therapeutics are routinely produced in Chinese hamster ovary (CHO) cells. Minimizing the heterogeneity within a Master Cell Bank (MCB) allows for a well-controlled process that is capable of the consistent manufacture of a product. Regulatory authorities therefore expect that clonal CHO cell lines are used. In this paper, we describe a rapid, reliable and cost-effective assessment of the probability of clonal derivation of recombinant cell populations by combining TLA and NGS with MCB-specific breakpoint qPCR assays and statistical analyses.

Introduction

Recombinant protein therapeutics, or biologics, are an important class of pharmaceuticals for which Chinese hamster ovary (CHO) cells are the most commonly used expression system. The process of developing a CHO cell line expressing a specific recombinant therapeutic is well-established: expression vector(s) encoding the transgene(s) of the therapeutic agent as well as a selectable marker are transfected into the host cell[1]. The resulting culture is a heterogeneous pool of cells that is, in case of random integration, typically the product of multiple independent integrations of expression vector(s) in the CHO genome. The next step is to select and grow those candidates that stably produce highest titers of the protein of interest. The top clone in this process leads to generation of the Master Cell Bank (MCB), which is used in the manufacturing of recombinant biologics.

A clonally derived MCB helps to ensure a robust production process and consistent product quality; FDA guidance[2] instruct cloning the cell substrate "*from a single cell progenitor*" during cell line development, while the EMA guidance stipulates that "*the cell substrate to be used for the production of the monoclonal antibodies should be a stable and continuous monoclonal cell line*[3]. Regulatory authorities therefore request a high assurance of clonality[4,5,6].

The FDA has recommended that two-rounds of limiting dilution cloning (LDC) at sufficiently low seeding densities (≤0.5 cells/well) provide acceptable probability that a cell line is clonal[4,7]. Other approaches have been developed and used either in combination with limiting dilution or as stand-alone methods[8]. These approaches include, but are not limited to, use of the ClonePix system[9], flow cytometry-mediated single cell sorting[10,11] and automated cell imaging systems[12]. Some ongoing clinical programs however employ legacy cell lines that were created before the industry had access to such practices and methods and may not satisfy current regulatory expectations for clonality when filing for market access.

Supporting evidence can be requested at several stages (e.g. IND or BLA) in the filing process. To provide supporting evidence the following additional tests can be considered: sub-clone analysis whereby a vial of the Master Cell Bank is plated as single cells (using LDC), expanded, and characterized using phenotypic analyses (e.g. cell doubling time, specific productivity etc.), product quality testing and genotypic analyses (e.g. fluorescence in situ hybridization (FISH) or Southern blotting) to evaluate individual integration sites[13].

Targeted Locus Amplification (TLA) combined with next-generation sequencing (NGS) allows for complete characterization of integration sites and the integrated transgene/vector sequence in any species[14]. This technology has been widely adapted by the pharmaceutical industry in various phases of CLD[15,16,17,18].

In this paper we describe a general and cost-effective approach to analytically assess the probability of monoclonal derivation of recombinant cell populations (a similar approach has been presented by Aebischer-Gumy *et al.*[16]). Using TLA combined with NGS, unique genetic features of the MCB can be identified, e.g. the breakpoint sequence between genome and the plasmid which characterizes the integration site or vector-vector junctions of an integrated concatemer. Clonally derived cell populations generated from the MCB can be analyzed by qPCR for the presence or absence of these unique genetic features. Compared to other cited technologies, such as Southern blotting and fluorescence in situ hybridization (FISH), qPCR breakpoint analyses allow for the analysis of a large number of monoclonal-derived cell populations for unique MCB-specific breakpoints. The methods and statistical analyses described in this paper therefore enable an efficient assessment of the probability of clonality.

Whilst we here describe the analysis of a CHO cell bank, the approach equally well applies to other cell types used in the production of biopharmaceuticals or viruses, such as those of human (e.g. HEK293) and murine (NS0 and Sp2/0) origins.

Material and methods

Cell line generation

A stable monoclonal antibody (mAb)-producing Chinese hamster ovary (CHO) cell bank was generated using the DHFR/MTX selection process. Briefly, a linearized expression vector encoding the heavy and light chain genes was transfected into CHO parental cells via electroporation. Transfected cells were grown in selective growth medium at 36.5°C and 10% CO2 to recover stable integrants (i.e cells that have integrated the expression vector into their genome). Pools were single cell cloned with a limiting dilution approach combined with imaging. Cells were seeded at a final density of 0.3 vc/well in 96 well plates. Cells were expanded and assessed for productivity, growth and product quality. The top clone was used to generate primary seed lot (PSL), which was further scaled up to generate the Master Cell Bank (MCB).

TLA/NGS

TLA followed by NGS, as well as bioinformatical analyses were performed on PSL vial as described[14]. Region of interest was targeted using the transgene-specific primer set. TLA products were sequenced on an Illumina sequencer generating paired-end, 2x150 bp reads. Mapping was performed using BWA-SW (Smith-Waterman algorithm[19]) with the Chinese hamster genome sequence (GCF_003668045.1 assembly) as reference genome.

Analytical subcloning and DNA isolation

MCB vial was thawed and cells were cultivated in serum-free medium at 36.5°C and 10% CO2 before single-cell isolation was carried out with Cytena single-cell printer and imaging. Cells were dispensed into 96-well plates prefilled with 100µl serum-free growth medium. Cell imaging was performed at days 0, 1, 10, and 18 after single-cell deposition. Clones were expanded in 24DW plates and 200uL of each culture was used for DNA extraction using KAPA Express Extract kit following manufacturer's instructions. The remaining culture for each subclone was cryopreserved.

Quantitative PCR

DNA extracts were assessed for the presence of MCB specific integration site using qPCR. TaqMan assays targeting genome-vector junction site and a CHO-genome region (*GLUC* region was used to control for successful extraction of gDNA), respectively, were custom designed by Applied Biosystems. Quantitative PCR was performed in 10 µL total reaction volume using 2X TaqMan Universe PCR master mix. The following thermal parameters were used: UNG nuclease activation at 50°C for 2 minutes and initial denaturation at 95°C for 10 minutes, followed by 40 cycles at 95°C for 15 seconds, and 60°C for 1 minute. Only DNA extracts with Ct_{GLUC}<30 (indication of successful extraction of gDNA) were considered in the interpretation. Every reaction was performed in triplicates.

Statistical assessment of probability of clonality

The standard practice for setting an upper confidence bound for a fraction or number of non-conforming items of the American Society for Testing and Materials (ASTM E2334-09 (Eq. 1)[20]) presents a method for the setting of a confidence interval of an unknown rate of occurrence of cells with the unique genetic event on the basis of a number of samples tested and all found to have the unique genetic event.

The formula is therefore suited to determine the probability of clonality:

One-sided confidence interval for clonal derivation = $\sqrt[N]{1-C}$, in which N is tested populations and C the confidence interval used.

Table 2 shows the effects of increasing N on the probability of clonality using a 95% confidence interval. Supplementary **Table 1** shows the effects of increasing N on the probability of clonality using a 90%, 95% and 99% confidence interval, respectively.

Table 2: Calculations of one sided 95% confidence intervals for clonal derivation.

Number of clonally derived populations tested and found to conform	One-sided 95% confidence interval for non-clonal derivation	One-sided 95% confidence interval for clonal derivation
1	0.95	0.05
2	0.776	0.224
3	0.632	0.368
4	0.527	0.473
5	0.451	0.549
10	0.259	0.741
20	0.139	0.861
50	0.058	0.942
60	0.049	0.951
75	0.039	0.961
93	0.032	0.968
100	0.03	0.97
186	0.016	0.984

Results

TLA sites

A single integration site was observed in the analyzed cell culture, suggesting the occurrence of a single integration event of the transgene into genome of the CHO cell, from which the MCB originates. The integration site was observed on chromosome 9 in the Chinese hamster genome assembly and the genome-vector junction sequences described in **Table 3** were identified.

Junction	Sequence
NW_020822657.1 (picr_41_new): 15329882 (+) – transgene:17(+)	5′– TTTCAAGGCCTAGGGTAACACGTTTGGAATCAACTTCTTGTCTGCCAG AGACGGTGTCTAT||*NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN GC*
transgene:11290(+) – NW_020822657.1 (picr_41_new): : 15329887 (+)	5′– *NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN NNNNNNNNNNNNNNNNN*||AAATGTCTGTTTTTAGGTGGCAGACTTGTTTG GGGGGCAGAGTCTTGCTATGTGGGGACCTGGCTAGCTTGGAATACTAT AT

Table 3: Nucleotide sequences at genome-vector junction sites. Sequence orientation is presented with (-) or (+); transgene sequence is presented in italics, junction is marked (||). Transgene and CriGri genome positions (bp) at junctions are indicated. Sequences of TaqMan assays for detection of identified junction sites are marked in *green* (forward and reverse primer) and *blue* (probe). True transgene sequence at junction sites is consealed (ie. Ns are used instead).

Next, the presence of the identified, MCB-specific integration site was assessed in DNA extracts from 60 analytical subclones using TaqMan assays targeting CriGri_Chr9:3752476-transgene:17 junction (*ASSAY1*) and *GLUC* region in the Chinese hamster genome, respectively. The presence of the NW_020822657.1 (picr_41_new): 15329882 – transgene:17 genome-vector junction site was confirmed in all analyzed subclones (**Figure 1**). This finding supports the monoclonal origin of the analyzed mAb producing cell bank at over 95% probability and with 95% confidence.

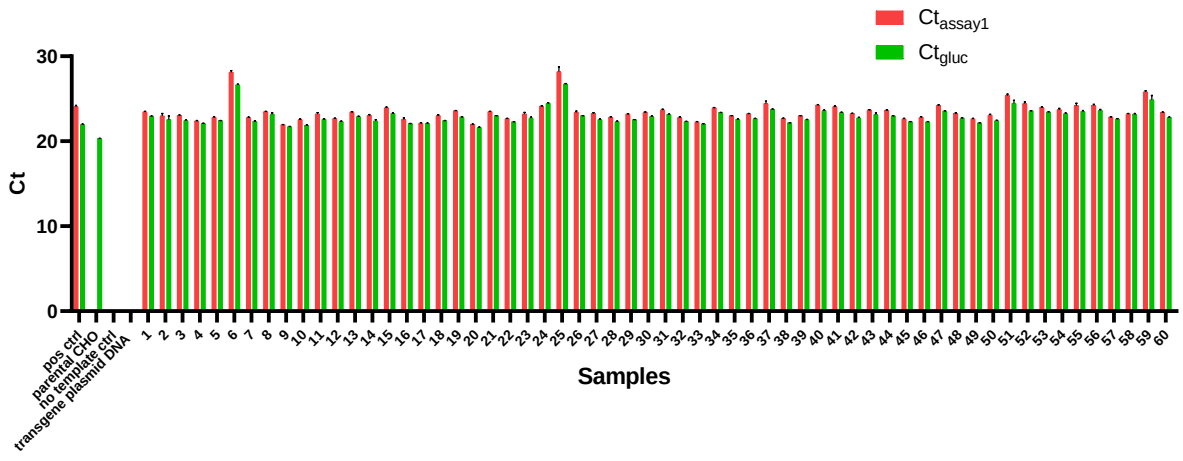

Figure 1: Confirming presence of MCB-specific junction site in all analyzed 60 MCB-derived analytical subclones The performance of both TaqMan assays was adequate as the following acceptance criteria were met: successful amplification of *GLUC* and *ASSAY1* in MCB (pos ctrl); successful amplification of *GLUC* in non-transfected parental CHO cell line (neg ctrl); no amplification in either the transgene plasmid DNA sample or no template control. Amplification of *ASSAY1* and *GLUC* was confirmed in all 60 analytical subclones.

Conclusion and discussion

Our work demonstrates that the use of TLA followed by NGS allows a detailed analysis of integrated transgenes, transgene integration sites and the identification of unique genetic features in a specific cell line. Cell bank homogeneity was assessed by testing populations clonally derived from the cell bank for the presence or absence of identified genetic features.

The intrinsic plasticity of the CHO genome[21,22,23] can result in the loss of specific genetic sequences of the MCB in subclones. This highlights the advantage of the analysis of at least 2 MCB specific breakpoints (**Table 4**). Clones with negative qPCR results can also be further evaluated using TLA to determine if they do share the MCB integration site. In addition, the evaluation of a subset of subclones over time using TLA and NGS provides information about the genetic stability of the integration site and integrated transgene sequences, which are key for a stable recombinant therapeutic protein production process.

Table 4: Potential outcomes from TLA and qPCR breakpoint analysis experiment

Event	Cause	Solution
Not all integration sites are identified in original MCB	Integration sites with partial integrated vector present in MCB	Perform TLA with multiple primer sets
MCB-specific breakpoint is not confirmed in at least one analytical subclone	Genetic instability of the subclone or MCB is not clonal	Use at least 2 breakpoints in the qPCR breakpoint analysis or Evaluate other MCB-specific integration site (if present) or Perform TLA on 'negative' subclone

In conclusion, we have described a cost-effective approach to analytically assess the probability of clonal derivation of recombinant cell populations, by combining TLA and NGS with MCB-specific breakpoint qPCR assays and statistical analyses.

References

1. Wurm FM. (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol.* 22:1393–1398.
2. ICH FDA (1998). Topic Q5D Quality of Biotechnological Products: Derivation and Characterisation of Cell Substrates Used for Production of Biotechnological/Biological Products. *CPMP/ICH/294/95*
3. EMA (2016). Development, production, characterisation and specifications for monoclonal antibodies and related products. *EMA/CHMP/BWP/532517/2008*
4. Kennett S. (2014). Establishing Clonal Cell Lines – A Regulatory Perspective Black Cell, Blue Cell, Old Cell, New Cell? *WCBP*
5. Novak R. (2017). Regulatory perspective on the evaluation of clonality of mammalian cell banks. *CDER/OPQ/OBP/DBRRI*
6. Welch J. (2017). Tilting at clones: A regulatory perspective on the importance of "Clonality" of mammalian cell banks. *CDER/OPQ/OBP/DBRRIV*
7. Wu P *et al.* (2018) Tools and methods for providing assurance of clonality for legacy cell lines. *Cell Culture Engineering XVI*
8. Gross A *et al.* (2015) Technologies for single-cell isolation. *Int J Mol Sci.* 16:16897–16919
9. Newman ENC and Whitney D. (2007) Rapid automated selection of mammalian cell colonies by cell surface protein expression. *Nat Methods.* 4, 462
10. Misaghi S *et al.* (2016) Slashing the timelines: Opting to generate high-titer clonal lines faster via viability-based single cell sorting. *Biotechnol Prog.* 32:198–207
11. DeMaria CT *et al* (2007). Accelerated clone selection for recombinant CHO CELLS using a FACS-based high-throughput screen. *Biotechnol Prog.* 23:465–472
12. Evans K *et al.* (2015). Assurance of monoclonality in one round of cloning through cell sorting for single cell deposition coupled with high resolution cell imaging. *Biotechnol Prog.* 31:1172–1178
13. Rawatt R. (2016) Regulatory Consideration for Biotechnology Products: Clonality of the Production Cell Bank. *Informa Life Sciences Annual Cell Line Development and Engineering Conference, 11 – 13 April, 2016*
14. De Vree P *et al.* (2014). Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping *Nature Biotechnology* 32: 1019-1025
15. Kaas CS *et al.*, (2015) Deep Sequencing Reveals Different Compositions of mRNA Transcribed From the F8 Gene in a Panel of FVIII-producing CHO Cell Lines. *Biotechnol J.* 10:1081-1089
16. Aebischer-Gumy C *et al.* (2018) Analytical Assessment of Clonal Derivation of eukaryotic/CHO Cell Populations. *J Biotechnol.* 20;17-26.
17. Boyd D *et al.* (2018) Isolation and Characterization of a Monoclonal Antibody Containing an Extra Heavy-Light Chain Fab Arm. *MAbs* 10:346-353.
18. Aeschlimann SH *et al.* (2019) Enhanced CHO Clone Screening: Application of Targeted Locus Amplification and Next-Generation Sequencing Technologies for Cell Line Development. *Biotechnol J.* 14:e1800371
19. Li H. and Durbin R. (2010) Fast and accurate long-read alignment with Burrows-Wheeler Transform. *Bioinformatics* 1:589-595.
20. ASTM E2334-09 (2018), Standard Practice for Setting an Upper Confidence Bound for a Fraction or Number of Non-Conforminga items, or a Rate of Occurrence for Non-Conformities, Using Attribute Data, When There is a Zero Response in the Sample, *ASTM International*, West Conshohocken, PA

21. Barnes LM *et al.* (2003). Stability of protein production from recombinant mammalian cells. *Biotechnol. Bioeng.* 81:631–639

22. Vcelar S *et al.* (2018). Changes in chromosome counts and patterns in CHO cell lines upon generation of recombinant cell lines and subcloning. *Biotechnol. J.* 13: e1700495

23. Wurm, F (2013). CHO quasispecies—implications for manufacturing processes. *Processes* 1:296–311

24. Frye C et al., (2016). Industry view on the relative importance of "clonality" of biopharmaceutical-producing cell lines. *Biologicals* 44, 117-122

Supplementary material

Supplementary Table 1: Calculations of one sided 90%, 95% and 99% confidence intervals for clonal derivation.

Number of clonally derived populations tested and found to conform	One sided 90% confidence interval for clonal derivation	One sided 95% confidence interval for clonal derivation	One sided 99% confidence interval for clonal derivation
1	0.100	0.050	0.010
5	0.631	0.549	0.398
10	0.794	0.741	0.631
20	0.891	0.861	0.794
30	0.926	0.905	0.858
40	0.944	0.928	0.891
50	0.955	0.942	0.912
60	0.962	0.951	0.926
70	0.968	0.958	0.936
80	0.972	0.963	0.944
90	0.975	0.967	0.950
100	0.977	0.970	0.955
200	0.989	0.985	0.977

Contact

Address: Cergentis B.V.
Yalelaan 62
3584 CM Utrecht
The Netherlands

Website: www.cergentis.com

Phone: +31 30 760 1636

General: info@cergentis.com
Sales: sales@cergentis.com