UNVEILING RECOMBINANT CELL LINE MYSTERIES:
NANOPORE CAS9 SEQUENCING FOR INTEGRATION SITE ANALYSIS (ISA) IN CHO CELLS

PathoQuest

iDTECT® INTEGRATION SITE ANALYSIS

Crispin ZAVALA-ALVARADO[1], Stephanie LEVON[1], Meissa BARBOUCHE[1], Sandrine MOREIRA[1], Samriti MIDHA[2], Richard FETHERSTON[2], Klaudia CIURKOT[2], Phill JAMES[2], Sean MCKENZIE[2], Veronica FOWLER[2] and Colette CÔTÉ[1]

[1]PathoQuest, Paris, France; [2]Oxford Nanopore Technologies, Oxford, United Kingdom

Introduction

Biologics have evolved significantly over the past decade, yet many are still highly dependent on genetically engineered cell lines for manufacturing. Historically, this has included the use of rodent cell lines like Chinese hamster ovary (CHO) cells for monoclonal antibody production. However, with the emergence of a wide range of novel modalities and, similarly, mechanisms for modifying genomes of various species for therapeutic and production purposes, a paradigm shift for biologics manufacturing has occurred.

Despite this changing landscape, for decades the tests required to characterize these engineered cell lines, as required by global health authorities and regulatory agencies, have consisted of a complex collection of aging, low-resolution assays. These have typically included:

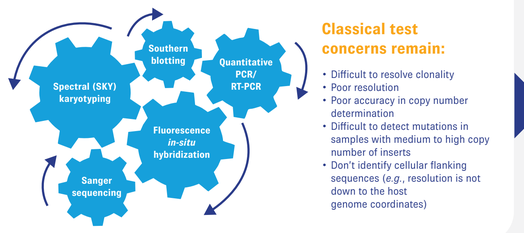
- **Sanger sequencing** which has traditionally been used to confirm the nucleotide sequence identity of the integrated expression cassette ('vector') or modified sequence, as well as the expressed mRNA copies.
- **Southern blotting, fluorescence *in situ* hybridization (FISH) and/or spectral karyotyping (SKY analysis)** which have all been used to grossly locate the inserted or modified sequence within the host genome.
- **Quantitative polymerase chain reaction (qPCR)**-based methods, including real-time or droplet digital PCR for the determination of the copy number of the inserted vector or modified sequence within the host genome.

While some advancements in testing have been made, the industry still longs for a simplified testing package.

Traditional integration site analysis testing

Traditional test package consists of multiple, aging assays to assess:

- Copy number
- Large structural changes/ gross configuration
- Total number of integration sites
- Identity verification of the expression cassette



Classical test concerns remain:

- Difficult to resolve clonality
- Poor resolution
- Poor accuracy in copy number determination
- Difficult to detect mutations in samples with medium to high copy number of inserts
- Don't identify cellular flanking sequences (*e.g.*, resolution is not down to the host genome coordinates)

Simplified solution

Goal: one test package that combines and replaces several assays

Advantages:

- Simplified testing
- Improved resolution
  - Within the host genome (characterization of insertion site flanking sequences)
  - Within the vector (sequence identity confirmation)

The desire for a simplified testing solution is often complicated by the varying nature of testing requirements outlined in these global guidance documents (e.g., ICH Q5B [1995], focusing on characterization of the expression construct; ICH Q5D [1997], focusing on characterization of the cell substrate itself; and the European Pharmacopeia [*EU Phar*], Chapter 5.14 [2006]). For example, ICH Q5B requires characterization of large structural changes within the vector sequence. While a vector identity test (typically by Sanger sequencing) is implied in this document, it is not mandated. *EU Phar* 5.14 requires the determination of the copy number, gross configuration, and sequence identity verification of the vector. In contrast to ICH Q5B, *EU Phar* 5.14 does not require a determination of the number or nature of the integration sites, vector sequence variant detection or other characterization. Despite these disparities, the intent of these documents is clear: utilize testing strategies that provide the highest resolution of characterization, to ensure not only product safety but also safety to the patient.

However, illegitimate and potentially insidious modifications can and will happen:

- Can activate oncogenes
- Can lead to inadvertent gene disruption
- Can modify pseudogenes
- Can cause infectious or other adverse or transformative changes to cells
- Can ultimately impact a patient

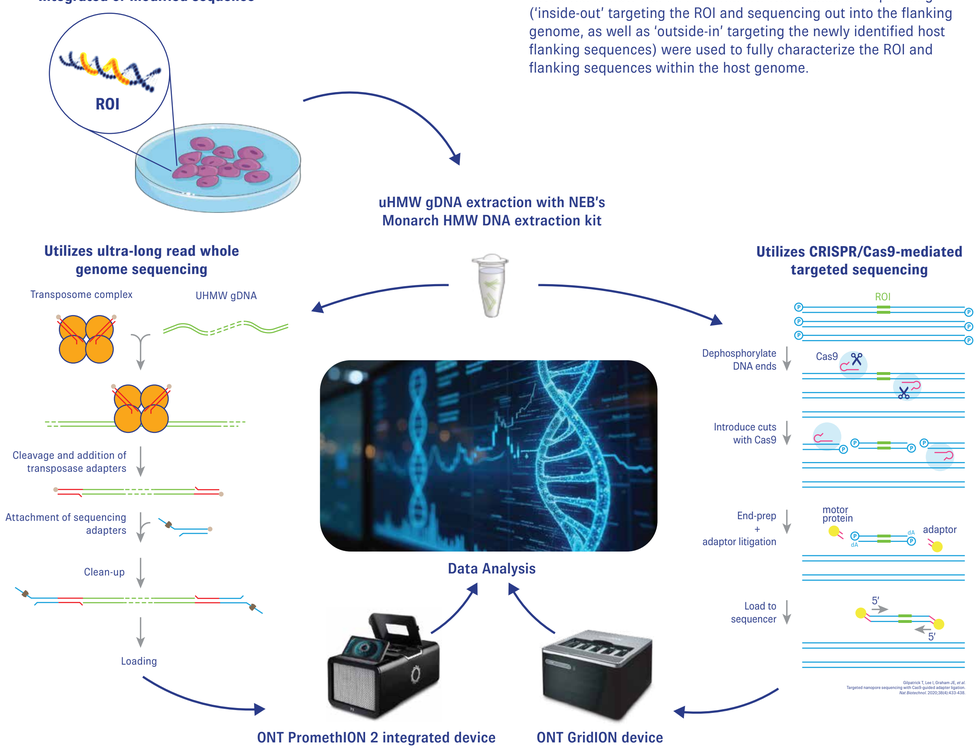
The one technology, available today, that is fully capable of achieving the desired level of resolution in a rapid and simplified format is Next Generation Sequencing (NGS). Here we showcase two studies using engineered CHO cells to demonstrate the power and flexibility of NGS to revolutionize these testing requirements.

Methodology

Both an ultra-long read whole genome sequencing (WGS: panel A, below) and CRISPR/Cas9-mediated targeted approach (panel B, below) utilizing Oxford Nanopore Technologies' (ONT) sequencing chemistries and platforms were used to characterize engineered CHO cell lines used for manufacturing. High molecular weight (HMW) DNA was extracted from each test sample using NEB's Monarch HMW DNA Extraction Kit. Libraries were prepared and sequenced as detailed below. The data was then analyzed using PathoQuest's proprietary data analysis pipelines for cell line integration site characterization. Characterization included the determination of the location of the integration site(s) within the host genome, as well as a determination of the copy number and general configuration of the inserted sequence.

A For the WGS approach, ultra-high molecular weight DNA (>200 kb) was prepared for sequencing on the PromethION P2i device using ONT's Ultra-Long DNA Sequencing Kit V14 and R10.4.1 flow cell.

B For the targeted approach, libraries were prepared using ONT's Cas9 Sequencing Kit. Briefly, short guide RNAs were combined with the CRISPR/Cas9 complex to first target the region of interest (ROI) within the host cell line genome for cleavage and to generate the sequencing priming site. Sequencing was then performed using ONT's R9.4.1 flow cell on the GridION. Two rounds of sequencing ('inside-out' targeting the ROI and sequencing out into the flanking genome, as well as 'outside-in' targeting the newly identified host flanking sequences) were used to fully characterize the ROI and flanking sequences within the host genome.

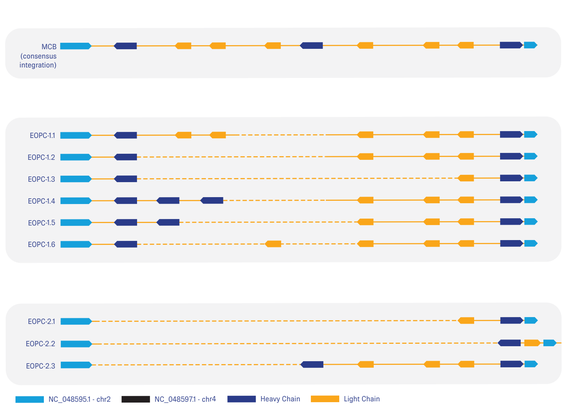


Results

While CHO cell lines have long been considered the workhorse of the biopharmaceutical industry, their genomes are notoriously plastic, resulting in significant rearrangement of the chromosomes which can be extremely difficult to resolve properly using traditional methods. To assess the utility of these NGS approaches, multiple engineered CHO cell lines containing inserted sequences of interest were characterized using the targeted and/or WGS approaches detailed herein. These included CHO master cell banks (MCBs) and end of production cells (EOPCs) predicted to contain either simple (low copy number) or complex (high copy number) insertions.

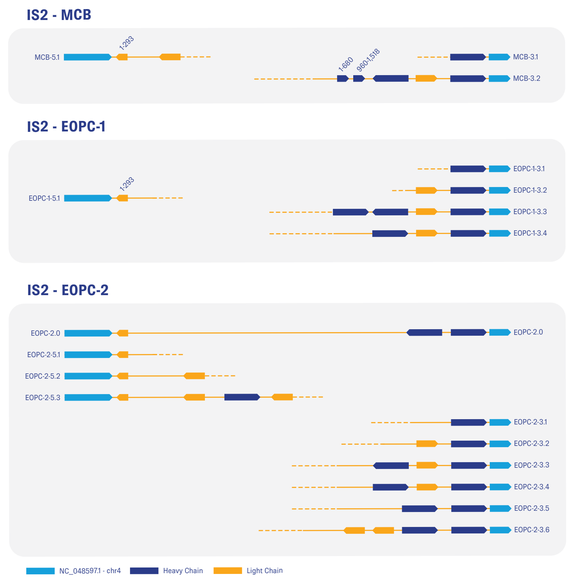
Low-complexity CHO cell line demonstrating variable loss of copy number

In this example, two (2) rounds of CRISPR/Cas9-targeted ONT sequencing were used to characterize inserted heavy and light chain sequences in an engineered CHO MCB and two (2) associated EOPCs. A total of two (2) integration sites (IS1, left panel and IS2, right panel) were identified in both samples, the first on chromosome 2 and the second on chromosome 4. In each case, both loss of and variability in copy number of the heavy and light chains were observed.



Schematic of the first integration site (on chromosome 2) in the EOPC samples compared to the MCB. The image shown in the top panel reflects the integration of the heavy (dark blue) and light (orange) chains relative to the host sequence (light blue) in the MCB. The images shown in the middle and bottom panels depict the position and orientation of the heavy and light chains relative to the host in the EOPC samples.

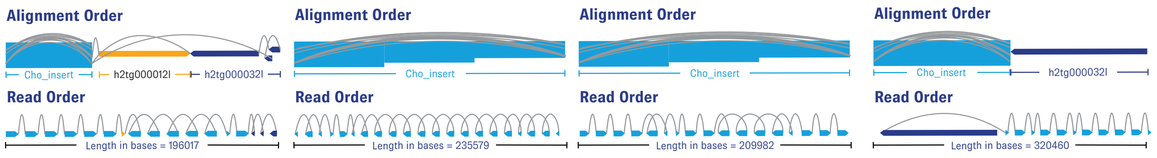
These profiles are based on the characterization of reads that span the 5′ (proximal) host sequence, through the vector insertion and back into the 3′ or distal host sequence. In these images, the dotted line represents copy number losses relative to the MCB integration. It is not possible to identify exactly which heavy or light chain was lost; the images are representative of the number of heavy and light chain copies observed. In all instances, the 3′ terminal heavy chain is truncated, similar to what was observed for the MCB. While the general configuration is relatively maintained, variability in copy number is evident in both EOPC samples.



Schematic of the second integration site (on chromosome 4) in the EOPC samples (middle and bottom panels) compared to the MCB (top panel). The heavy (dark blue) and light (orange) chains are depicted relative to the host sequence (light blue). In these images, the dotted line represents where the reads terminate prior to reaching the distal host sequence. Even based on this truncated data, it is evident that the 5′ profile remains consistent among the samples; however, the 3′ end of the insertion varies considerably.

High-complexity CHO cell line demonstrating an extreme integration profile

In this example, multiple rounds of CRISPR/Cas9-targeted ONT sequencing were attempted to characterize an inserted sequence within an engineered CHO MCB (generously provided by Nicole Borth, BOKU University). However, this method was unable to fully characterize the insert (data not shown). Due to the unexpectedly highly complex nature of this clone, the ONT-based WGS approach was also applied to provide further insight into the nature of the integration event. As illustrated below, despite the presence of reads spanning well over 200 kb each (the longest at 320 kb), it was not possible to stitch these reads, containing dozens of perfect tandem repeats of the integrated sequence, into a complete integration site profile. Based on available WGS data, the predicted insertion size is ~ 350 – 400 kb with > 50 copies of the vector inserted into a single site in the CHO genome. While additional ultra-long read sequencing data is required to fully characterize this CHO clone (work in progress), this profile clearly demonstrates an extreme example of a very high copy number insert which, while challenging to characterize with current technologies, is impossible to characterize using traditional methods such as Sanger sequencing and spectral analysis.



Upstream sequence: Illustrated by a 196 kb read showing a small rearrangement in the host sequence followed by six copies of the vector (including a directional switch) and one truncation, an ALU element then six additional perfect head to tail copies of the vector.

Internal sequence: A 235 kb read entirely comprised of 200 kb of perfect head to tail tandem repeats of the vector sequence with no host sequence detected.

Internal sequence: A 210 kb read comprised of multiple complete and partial vector copies with directional switching, all in tandem.

Downstream sequence: A 320 kb read showing host sequence fused to multiple truncated and intact copies of the vector in tandem, all in a head to tail orientation.

Conclusion

Engineered CHO cell lines have long maintained their position as the workhorses of choice when it comes to biopharmaceutical manufacturing. For example, they grow well in suspension cultures, have short doubling times, are highly adaptable at scale, are very easy to engineer and have a high safety profile. However, these cell lines have proven to be highly challenging to characterize with constant rearrangements and duplications, and highly variable integration profiles (as evidenced above). Interestingly, in 2016, an industry paper* was published in *Biologicals* that outlined a rationale to shift the focus away from characterization of the production system and towards characterization of the product itself. Specifically, the paper concluded that:

- "referring to a production cell line as a "clone" or to the "clonality" of a manufacturing cell bank is misleading, as any population of these types of cells cultured for a length of time will accumulate genetic and phenotypic heterogeneity…
- that the primary focus for any biopharmaceutical manufacturing process should be on the product being produced rather than on the "clonality assurance" of the cell line used for its derivation. Similarly, the product is ultimately highly purified, and the focus should be on product and process consistency rather than on uncontrollable aspects of the cell lines used to produce the product…
- although the purpose of a cloning step is to facilitate the isolation of stable, highly-productive cell populations, and to minimize the genetic and phenotypic diversity within a cell line population, the key to ultimately identifying an appropriate production cell line is the genotypic and phenotypic (including the expressed product) characterization of the chosen cell lines post-cloning.
- And lastly, since it is the product, not the production cell line which is used to treat the patient, primary emphasis should be placed on the elements that comprise holistic integrated process development such as risk assessments, control strategies, process characterization and process validation."

Despite this clear positioning by industry leaders, a full 10 years later there has been no change to the guidance recommendations, and unfortunately, nor is one expected. Therefore, it is even more imperative to ensure that the most advanced technologies and methods, like NGS, and in particular, WGS using a sequencing platform like ONT that is capable of extremely long read generation continue to be advanced and used to provide the greatest level of characterization of these cell lines.

*Industry view on the relative importance of "clonality" of biopharmaceutical-producing cell lines - https://doi.org/10.1016/j.biologicals.2016.01.001

Contact us

PathoQuest's NGS solutions are designed to meet and exceed regulatory expectations and accelerate cell line characterization as part of your biosafety testing strategy.

Contact our experts to hear more about how PathoQuest's NGS testing portfolio can simplify and secure your biosafety testing strategy and accelerate your path to market… compliant, safer, and faster.

contact@pathoquest.com | www.pathoquest.com