Identifying Stable Hotspots in the CHO Genome for Therapeutic Protein Production

Bill Hilliard
Advisor: Kelvin H. Lee
Committee Members: Eleftherios T. Papoutsakis, Wilfred Chen, Nicholas Sandoval

---

Chinese hamster ovary (CHO) cell lines used to produce commercial quantities of therapeutic proteins commonly exhibit a decrease in productivity over time in culture, a phenomenon termed production instability. Random integration of transgenes encoding the protein of interest into locations in the CHO genome that are vulnerable to genetic and epigenetic instability can lead to production instability due to copy number loss and epigenetic silencing. These cell line development challenges can be overcome by using targeted integration to insert transgenes at pre-defined genomic loci, often called "hotspots," that are transcriptionally permissive and have enhanced stability relative to the rest of the genome. However, the use of targeted integration for industrial CHO cell line development currently requires significant upfront effort to identify hotspot loci capable of supporting multi-gram per liter therapeutic protein production from a limited number of transgene copies.

Characterization of the CHO epigenome can aid in the identification of hotspots that maximize transcription from a single integrated transgene copy and maintain their desirable properties in diverse cell lines and bioprocessing conditions. To support this effort, a chromosome-scale Chinese hamster reference genome suitable for genome-scale comparative analysis of different CHO cell lines was developed by scaffolding a previous genome assembly version using high-throughput chromosome conformation capture. Large "safe harbor" regions characterized by transcriptionally permissive three-dimensional chromatin structures with enhanced genetic and epigenetic stability were then identified by comparative multi-omics analysis of two CHO-K1 cell lines. To further pinpoint hotspot locations at base-pair resolution, we measured CMV promoter-driven transgene expression strength and stability simultaneously at thousands of integration sites using a pooled high-throughput screening method in which a barcoded reporter library was randomly integrated into the genome and assayed by next-generation sequencing. The resulting genome-scale dataset captured the entire range of position-dependent expression theoretically possible from a single transgene copy. Integration sites exhibiting maximal transcriptional output were associated with a limited set of genetic and epigenetic properties defining hotspot regions of approximately 10kb genome-wide. Cell lines generated by retargeting eight hotspot candidates from the high-throughput screen consistently exhibited higher transgene mRNA expression than a commercially viable hotspot. Additionally, monoclonal antibody productivity from a novel hotspot was improved by more than two-fold, reaching a volumetric productivity of 2g/L, by incorporating chromatin-modifying elements isolated from the CHO genome into the transgene expression cassette. These findings will be valuable resources for targeted integration platform development within the CHO community.