INTEGRATED ANALYSIS OF GENOMIC AND EPIGENOMIC INSTABILITY FOR CHO CELL LINE ENGINEERING

Sofie O'Brien, Department of Biomedical Engineering, University of Minnesota, Minneapolis, MN, USA
bluvs002@umn.edu
Arpan Bandyopadhyay, Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN, USA
Conor O'Brien, Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN, USA
Zion Lee, Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN, USA
Wei-Shou Hu, Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN, USA

Key Words: Genome instability, CGH, Sequencing, Karyotyping, ATAC-seq

Stability is an important factor in the development of cell lines for therapeutic protein production. In culture, the chromosome number and structure of Chinese Hamster Ovary (CHO) cells undergo rapid change. Over the course of cultivation, selection, and adaptation, chromosomal aberrations such as mutations, deletions, duplications, and other structural variants can accumulate. Some genomic regions may be more prone to such instability than others. When introducing exogenous genes for product formation or for engineering cell characteristics, it is critical to integrate into a stable region. A deeper understanding of the relationship between structure and stability is important for cell culture engineering.

We investigated the genome stability of CHO cell lines at the macroscopic and microscopic levels, as well as from the epigenetic and genetic perspective. At the macroscopic level, we examined chromosomal and karyotypic variation, observing that the progenies of single cell clones quickly developed widely distributed variants with different numbers and types of chromosomes. However, at the population level the karyotype and chromosomal number distribution remained in a similar range. Stability at the microscopic level was analyzed using a gene-coding region focused comparative genomic hybridization (CGH) microarray, allowing us to determine genomic variations in gene copy number. With CGH data for many parent-daughter relationships, including subclones and relationships between host and producing cell lines, we identified genome segment changes that happen commonly during cell line development and subcloning.

To further examine variation at the microscopic and genetic level, whole genome sequencing data of multiple CHO cell lines was used to identify structural variants, such as deletions, inversions, and duplications using the tools DELLY2 and LUMPY. Heterogeneity was present within each cell line and visible in the form of genome mosaicism. The effect of epigenetic modifications on the CHO genome was explored using the Assay for Transposase Accessible Chromatin Sequencing (ATAC-seq), which examines chromatin accessibility. ATAC-seq information was incorporated with transcriptional activity data using RNA-seq from multiple cell lines to identify inaccessible regions of the genome.

This integrated systems approach combining chromosome number, karyotyping, CGH, genome sequencing, ATAC-seq, and RNA-seq gives us insight into the heterogeneity and instability of CHO cells, allowing us to identify desirable and undesirable regions for gene integration. With this data, we can select sites ideal for targeted integration of transgenes as well as screen out potentially unstable cell lines developed using random integration.