COMPREHENSIVE META-ANALYSIS OF THE CHO CODING TRANSCRIPTOME

Markus Riedl, acib – Austrian Centre of Industrial Biotechnology, Vienna, Austria; Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria
markusriedl@acib.at
Caterina Ruggeri, Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria
Nicolas Marx, Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria
Nicole Borth, Department of Biotechnology, University of Natural Resources and Life Sciences, Vienna, Austria

Key Words: CHO, Transcriptomics, RNA-seq, Omics, Meta-analysis

Chinese Hamster Ovary (CHO) cells are amongst the most important cell factories in biotechnology. Capable of incorporating complex post-translational modifications, they are invaluable in the production of biopharmaceuticals. They are steadily developed to improve yield and product quality through cellular engineering and by optimizing bioprocesses. High-throughput omics technologies have drastically influenced these endeavours by enabling a comprehensive molecular insight at multiple levels, leading to more rational and informed engineering decisions. Oftentimes, such datasets are created to elucidate a specific biological question, and the potential to put these individual datasets into greater context is left unused. With numerous RNA-sequencing datasets accumulated, we aim to seize the opportunity to conduct a large-scale meta-analysis of the CHO transcriptome to study gene expression across various cell lines and cell culture conditions.

We used publicly available RNA-sequencing datasets, available from the Sequence Read Archive (SRA) of the NCBI, as well as several datasets that were produced in-house. The datasets span a wide variety of different cell lines, culture conditions, and growth phases, including various antibody-producing cell lines as well as different sequencing approaches and coverage depths. This unique heterogeneity, however, requires careful considerations of each individual data processing step. Starting from raw reads in FASTQ format, we consistently process all datasets through a reproducible workflow that incorporates state-of-the-art bioinformatic tools. RNA-seq reads are mapped to the most recent reference genome assembly of the Chinese Hamster. Our approach involves novel batch adjustment techniques and sophisticated normalization methods in order to address the challenges arising from inhomogeneous data. Finally, custom-developed R scripts assess gene expression and correlation networks across several biological conditions of various cell lines.

Our work aims to elucidate the complex circuitry of the CHO transcriptome. We conduct a large-scale investigation of the gene expression patterns across various cell lines under different biological conditions. This meta-analysis of RNA-seq data is capable of revealing transcriptomic programs that cohere with industrially relevant phenotypes and/or differences that are specific to culture conditions or recombinant cell lines. On the long term, we aim for extensive insights into the regulatory circuits that ultimately determine gene expression, and in consequence the phenotype of CHO cells. With our approach, we demonstrate the vast potential of publicly available omics datasets for elucidating molecular and phenotypic relations on a large scale. The pipeline is end-to-end reproducible and can be easily adapted for similar intentions.