RNA-seq based expression analysis of the CHO cell protein secretion pathway

Lund, Anne Mathilde; Kaas, Christian Schrøder; Kildegaard, Helene Fastrup; Kristensen, Claus; Andersen, Mikael Rørdam

Publication date: 2013

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):
RNA-seq based expression analysis of the CHO cell protein secretion pathway

Anne Mathilde Lund¹,*, Christian Schrøder Kaas², Helene Fastrup Kildegaard³, Claus Kristensen², Mikael Rørdam Andersen¹

(1) Center for Microbial Biotechnology, Department of Systems Biology, Technical University of Denmark, Kgs. Lyngby, Denmark; (2) Current address: Novo Nordisk, Maaloev, Denmark; (3) Current address: Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Hoersholm, Denmark

*amalu@bio.dtu.dk

Introduction

The Chinese hamster ovary (CHO) cell-line is the predominant mammalian industrial cell line being used to produce recombinant therapeutic proteins. Although CHO cells have been used for more than 25 years, the genome sequence was first published in 2011. So far there have been limited studies of the cell biology of the CHO cell and the potential of cell line engineering. To elucidate the poorly understood cellular processes that control and limit recombinant protein production and secretion, a system-wide study was initiated to identify possible engineering targets relevant for therapeutic protein production.

Objectives

- Reconstruction of the complex cellular machineries of the early protein secretion pathway and unfolded protein responses by employing legacy knowledge of mouse
- By using RNA-seq data, a differential gene expression analysis of the constructed CHO secretion pathway would provide a unique possibility for identification of active components to increase the productivity of recombinant proteins.

Strategy of in silico reconstruction of CHO cell pathways

First a study was initiated for identifying mouse proteins in the literature associated with the secretion pathway and the ER stress responses. Amino acid sequences were used for BLASTp against the available RefSeq CHO-K1 genome. The identified genes could be used for extracting parts of the mapped RNA-seq data. However, only mapped genes

Proteins associated or linked to UPR pathway were identified by manually curate available literature on mouse models and cell lines. Furthermore, was the know interactions and dynamics of proteins in the UPR pathway mapped. Found proteins were used to identify CHO-K1 genes of the UPR pathway.

Reconstructed UPR pathway and RNA-Seq data analysis

Protein known to be involved in the apoptotic pathway and have a direct interaction is also seen to cluster across the samples in the RNA-Seq data. However, apoptosis can be induced through different path of the UPR pathway depending on the cause of stress, which is likely the reason for some of the apoptotic genes to cluster more closely to chaperones and folding proteins.

The CHO-K1 genome was sequenced recently and the annotation is at present insufficient, so more than 4 misreads was form each sample not possible to map.

Many of the chaperones commonly related to protein folding seem to be highly expressed in all cell lines, but a variation between producer strains in non-producers, exponential growth or stationary phase and stressed cells. E.g. the gene encoding CHOP protein is increased in expressed under stress and in stationary phase, which also has shown positive protein production by knock-down.

Perspectives

This preliminary study shows the possibilities for using RNA-seq data and cluster analysis to identify new gene clusters based on biological gene expression behaviour that can lead to a greater biological understanding of the important industrial cell line CHO-K1. It can also be used for identifying genetic targets for improvement of protein production by overexpression transcription factors or create knock-downs of growth inhibiting.

Design and methods for RNA-Seq samples

- RNA was extracted under different growth conditions and treatment from three different CHO cell lines
- Paired-end RNA sequencing was performed by AROS a/s on Illumina HiSeq 2000 platform with a sequencing depth of min. 35 mios reads

Overview of samples

<table>
<thead>
<tr>
<th>CHO-K1</th>
<th>CHO-K1_sg</th>
<th>CHO-DG44_FVIII</th>
</tr>
</thead>
<tbody>
<tr>
<td>exp. growth</td>
<td>stationary</td>
<td>exp. growth</td>
</tr>
<tr>
<td>1 samples</td>
<td>1 samples</td>
<td>stationary</td>
</tr>
<tr>
<td>control</td>
<td>(UNIEA)</td>
<td>1 sample</td>
</tr>
<tr>
<td>control</td>
<td>(UNIEA)</td>
<td>1 sample</td>
</tr>
<tr>
<td>NaBu1</td>
<td>2 samples</td>
<td>exponential growth</td>
</tr>
<tr>
<td>none</td>
<td>medium</td>
<td>high</td>
</tr>
<tr>
<td>5 samples</td>
<td>1 samples</td>
<td>2 samples</td>
</tr>
</tbody>
</table>

Workflow of RNA-Seq data handling and analysis

References