Spring 5-12-2016

Multi-omic modeling of translational efficiency for synthetic gene design

Joseph Longworth
University of Sheffield, j.longworth@sheffield.ac.uk

Javier Gonzalez
University of Sheffield

Paul Dobson
University of Manchester

Josselin Noirel
Conservatoire National des Arts et Métiers

Neil Lawrence
University of Sheffield

MULTI-OMIC MODELING OF TRANSLATIONAL EFFICIENCY FOR SYNTHETIC GENE DESIGN

Joseph Longworth, University of Sheffield
j.longworth@sheffield.ac.uk
Javier Gonzalez, University of Sheffield
Paul Dobson, University of Manchester
Josselin Noirel, Conservatoire National des Arts et Métiers
Neil Lawrence, University of Sheffield
Mark Dickman, University of Sheffield
David James, University of Sheffield

Key Words: Translation, Omics, Modeling, Gene Design, SILAC words.

Controlled expression of recombinant genes in CHO cells for advanced cell engineering will require precise, coordinated control of the synthetic processes that underpin the production of specific recombinant products or the optimal stoichiometry of functional effector proteins for multigene engineering applications. Although control of recombinant gene transcription in CHO host cells is now possible, technologies that enable control of recombinant mRNA translation rate are lacking. This is undesirable as in eukaryotic cells, cellular mRNA concentration itself may only explain a relatively small proportion of the variation in cellular protein abundance; mRNA translation rate is by far the most important contributor to cellular protein concentration.

We have taken a top-down, genome-scale computational modeling approach to develop computational design tools that enable control of recombinant gene translational activity in CHO cells. Through a combination of pulsed stable isotope labelling of amino acids in cell culture (pSILAC) and RNA-Seq based analysis of the CHO cell transcriptome we quantified the translational efficiency of > 4000 mRNAs.

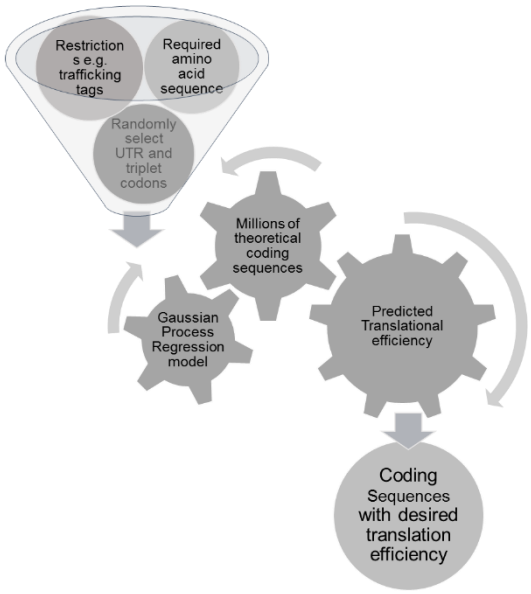

Figure 1 – Diagram illustrating the concept of gene design

Based on informatic reconstruction of CHO mRNAs (to include untranslated and coding sequences) we built and trained a gaussian process regression model using over 250 defined mRNA sequence features to enable validated *in silico* prediction of mRNA translational efficiency in CHO cells from mRNA sequence.

Using this genome-scale empirical modeling we created a computational gene analysis and design platform that permits both prediction of the translational efficiency of natural and recombinant mRNAs in CHO cells and *de novo* design of synthetic mRNAs with predictable translational activity.

This platform will be employed to (i) maximize the efficiency of recombinant mRNA translation for easy-to-express proteins, (ii) optimize the rate of mRNA translation for difficult-to-express proteins and (iii) control the stoichiometry of product synthesis in multigene expression systems.