A machine learning poly-omics classifier to improve protein production in CHO cells

Business Interaction Voucher funding from BioProNET has enabled Claudio Angione and colleagues from Teesside University to work with the Centre for Process Innovation to develop a computational method that could be used to predict and improve therapeutic protein production.

The challenge – Chinese hamster ovary (CHO) cells are the most commonly used cell line for the production of therapeutic proteins such as monoclonal antibodies. But like most mammalian cells, they are very inefficient at producing recombinant proteins.

Aims – The objective was to combine novel machine learning techniques with poly-omic analysis, to build a computational method that: (i) accurately identifies whether target cells have optimal conditions for producing the target protein; (ii) if not, predicts genetic modifications that will likely increase protein production.

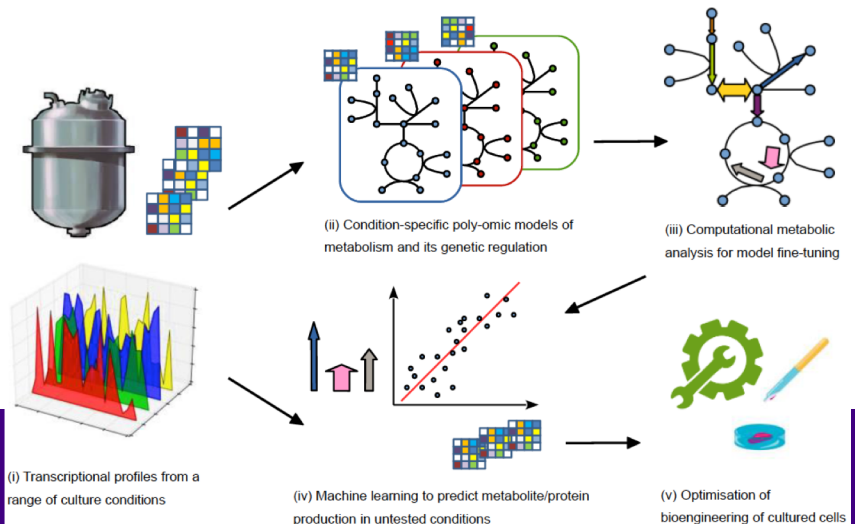
The collaboration – The Centre for Process Innovation (CPI) provided accesses to experimental bioprocessing data and were involved in key discussions.

Key findings – The researchers integrated experimental data at gene level with data generated *in silico* via a publicly available genome-scale metabolic model of a CHO cell within an integrated data-driven framework. They trained the poly-omics machine learning method using gene expression data from varying conditions and associated reaction rates in metabolic pathways, reconstructed *in silico*. They then evaluated this approach through a computational validation based on cross-validation, estimating the average prediction error in general settings. Importantly, they showed that metabolic predictions coupled with gene expression data can significantly improve estimations of lactate production based solely on gene expression.

It is anticipated that the predictive ability of pipeline will to vastly improve with additional omic data (from CPI). Moreover, although the validation focused on lactate production, the proposed methodological framework can be implemented around any target metabolite or protein.

Outcomes and next steps
The computational work was presented at IECM-2 2017 and highlighted on Phys.org:
https://sciforum.net/conference/iecm-2/paper/4993
http://tees.openrepository.com/tees/handle/10149/621547
https://phys.org/news/2017-12-machine-boost-protein-production-pharmaceuticals.html

The results will be extended with additional data provided by CPI, and validated in an upcoming project for which the researchers are seeking funding.



"As a result of this project, we are currently discussing with CPI further ways to incorporate machine learning in their pipelines."