Multi-omics data integration for the discovery of COVID-19 drug targets

Tyrone Chen[1], Kim-Anh Lê Cao[2], Sonika Tyagi[1]

1 School of Biological Sciences, Monash University, Australia; **2** Melbourne Integrative Genomics, School of Mathematics and Statistics, University of Melbourne, Australia

tyrone.chen@monash.edu 0000-0002-9207-0385

INTRODUCTION

- We hypothesised that to study complex phenotypes such as COVID-19, integrating multi-omics data would reveal signals that were not possible by single-omics data alone.
- To test our hypothesis, we used a medically relevant multi-omics dataset.
- Comparing single-omics and multi-omics data showed that data integration performed better than single-omics analyses (data not shown).

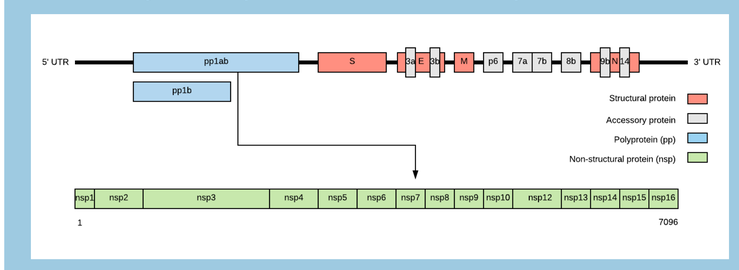

Figure 1: Structure of the SARS-Cov-2 genome

DATA

- We obtained time-series data for proteome and translatome[1].
- A CaCo-2 cell line was grown in vitro as three separate batches, for a total of n=3 per treatment condition.
- Cells were sampled at four timepoints for proteome and translatome data.
- 6381 features were present in the original proteome data and 2715 features were present in the original translatome data.

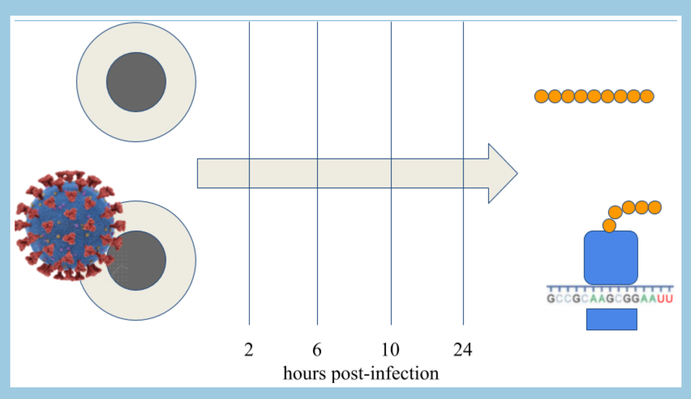

Figure 2: Infected CaCo-2 cells were sampled for proteome and translatome data at 2/6/10/24 hours post-infection.

References:

1. Bojkova, D., Klann, K., Koch, B. et al. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. Nature (2020).
2. Singh, A., Shannon, C. P., Gautier, N. et al. DIABLO: an integrative approach for identifying key molecular drivers from multi-omics assays, Bioinformatics, Volume 35, Issue 17, 1 September 2019, Pages 3055–3062.
3. Ursu, O., Holmes, J., Bologa, C. G. et al. DrugCentral 2018: an update, Nucleic Acids Research, Volume 47, Issue D1, 08 January 2019, Pages D963–D970.
4. Wold, H. (1966) Estimation of Principal Components and Related Models by Iterative Least Squares. In: Krishnaiah, P.R., Ed., Multivariate Analysis, Academic Press, New York, 391-420.
5. Liquet, B., Cao, K.L., Hocini, H. et al. A novel approach for biomarker selection and the integration of repeated measures experiments from two assays. BMC Bioinformatics 13, 325 (2012).
6. Szklarczyk D., Gable A. L., Lyon D., et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 2019 Jan; 47:D607-613.
7. Jassal B., Matthews L., Viteri G., et al. The reactome pathway knowledgebase. Nucleic Acids Res. 2020 Jan 8;48(D1):D498-D503.

We identified 55 potential drug and target combinations. One of the drugs is already in clinical trials (etiposide). We also report 5 new drugs that are not currently under investigation (aspartic acid, asulacrine, carubicin, daunorubicinol, intoplicine).

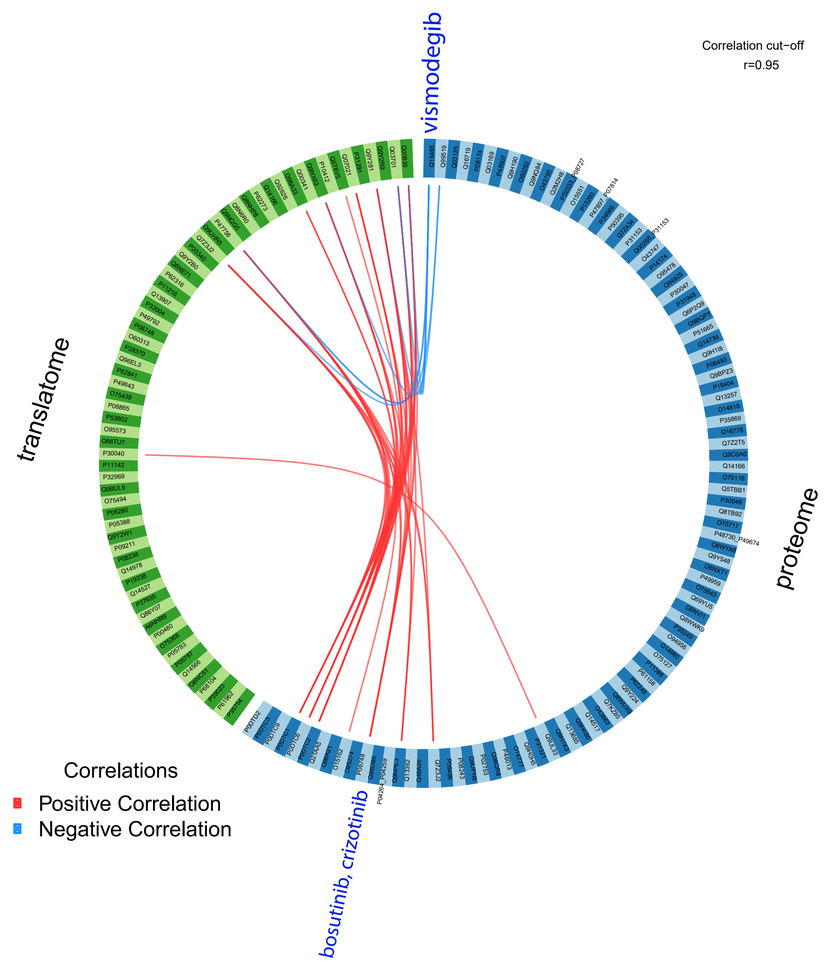

Figure 5: Using a component-based multivariate approach called DIABLO[2], we identify and select highly correlated features across the matched proteome and translatome datasets. Loadings from each component corresponding to low-level features were obtained. Strong positive correlations between low-level features are highlighted in red and strong negative correlations between low-level features are highlighted in blue. Within this set, we identified drug targets using DrugCentral[3] and two drug targets are highlighted.

FUTURE DIRECTIONS

- Although single-omics information is useful, we demonstrate that resolution improves with multi-omics approach on this small test case.
- We will perform similar integrative analyses for two additional matched datasets: [proteome, transcriptome] and [epitranscriptome, transcriptome]
- We will assess the reproducibility of our results.

METHODS AND RESULTS

- Missing values and a batch effect were observed.
- We corrected for this by filtering and imputing[4] values followed by a multilevel decomposition[5].
- 59% (1595 data points) of the data remained for translatome. Proteome data was not significantly affected.

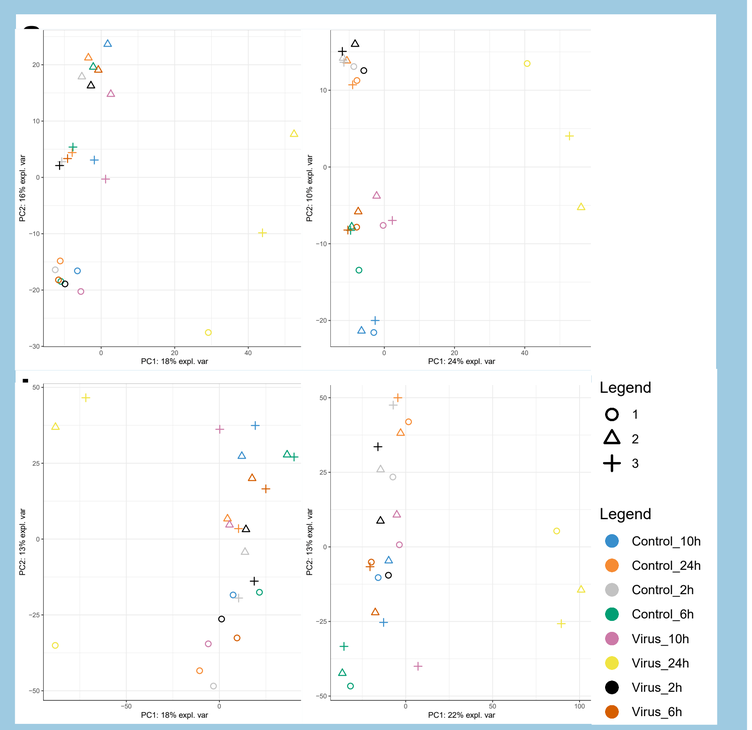

Figure 3: Principal component analysis sample plot before and after multilevel decomposition for (a) translatome and (b) proteome.

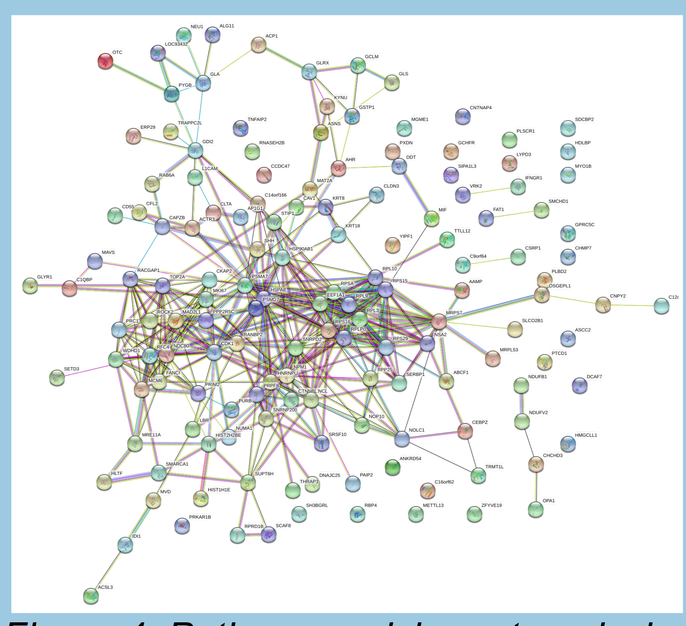

Figure 4: Pathway enrichment analysis using Reactome[6] and StringDB[7] with correlated proteome and translatome data.

The top 10 most enriched pathways included: **Metabolism of RNA, Infectious disease, Peptide chain elongation, Eukaryotic Translation, Influenza Life Cycle and Infection, Viral mRNA Translation, Metabolism, Influenza Viral RNA Transcription and Replication.**

School of Biological Sciences

Genetics, Genomics and Health

tyagilab

We acknowledge high performance computing support provided by the Monash eResearch Centre.
We acknowledge and pay respects to the Elders and Traditional Owners of the land on which our four Australian campuses stand.