SYSTEMS BIOLOGY APPLICATION OF PERTURBATION GENE EXPRESSION PROFILES

PhD thesis

Szilvia Barsi

Molecular Medicine Doctoral School
Semmelweis University

Supervisor: Bence Szalai, MD, PhD
László Hunyady, MD, DSc

Official reviewers: Fichó Erzsébet, PhD
László Imre Dobson, PhD

Head of the Complex Examination Committee:
Miklós Csala, MD, DSc

Members of the Complex Examination Committee:
Tamás Hegedűs, PhD
Zoltán Gáspári, DSc

Budapest, 2025

1. Introduction

Understanding how cells respond to genetic and chemical changes is essential to deciphering biological mechanisms and advancing drug discovery. Although transcriptomics measures mRNA abundance, it may not reliably reflect functional protein activity, limiting insight into cellular processes. Footprint-based computational methods address this gap by inferring protein activities from the patterns of regulated genes.

Systems biology can integrate high-throughput perturbation gene expression profiles, such as the LINCS L1000 signatures to investigate the downstream effects of modulating genes or proteins in various cellular contexts. This approach offers new opportunities for drug repurposing by comparing disease-induced and drug-induced transcriptional signatures to identify promising therapeutic candidates and understand mechanism of action.

Leveraging these perturbation signatures, computational tools have been developed to infer signaling activities and cell-cell communication from transcriptomic data, thereby overcoming the limitations of expression-based methods. These advances enhance mechanistic understanding of complex biological processes and support the identification of effective treatments by linking perturbations to functional outcomes.

2. Objectives

1. Investigation of SARS-CoV-2 host response and antiviral drug mechanisms by a signature similarity-based approach to find potential anti-SARS-CoV-2 drugs and gain insights into the in vitro effective antiviral drug mechanisms.
   - Gain insight into the host response to the SARS-CoV-2 infection in cellular assays by functional genomic analysis using computational tools.
   - Conduct functional genomic analysis of drug effects in cellular assays, considering drugs that have been reported to be effective against the SARS-CoV-2 virus.
   - Compare infection-induced signatures and drug-induced signatures in cell lines using signature similarity metric to identify potential anti-SARS-CoV-2 drugs by gaining insights into the mechanisms that can explain antiviral effects.
2. Identification and validation of the antiviral drug mechanism.
   - Evaluate the predictive potential of the signature similarity or machine learning-based approach for identifying *in vitro* effective antiviral drugs. Furthermore, to identify molecular features that most contribute to classifying effective or ineffective antiviral drugs, and to gain further insight into their mechanisms.

- Validate the cholesterol depletion effect of the selected antiviral drugs by determining the fluorescent cholesterol sensor ratio between the plasma membrane and cytosol in an *in vitro* assay, where the cholesterol localization is determined by automatic confocal microscopy imaging.

3. Development and evaluation of the RIDDEN (Receptor actIvity Data Driven inferENce) computational tool.
   - Develop a computational framework, RIDDEN, to infer receptor activities from transcriptomic data using perturbation signatures.
   - Benchmark RIDDEN against a state-of-the-art method, CytoSig, for predicting cytokine and cytokine receptor signaling activities.
   - Evaluate the performance of both RIDDEN and CytoSig using cytokine perturbation datasets as ground truth.
4. Demonstrating the RIDDEN's applicability to patient data.
   - Apply RIDDEN to the PD-1 inhibitor-treated patients' transcriptomic dataset to demonstrate its practical utility.
   - Demonstrate that RIDDEN provides insights into the cellular mechanisms of intercellular communication by investigating the association between predicted receptor activity and therapeutic responses.

3. Methods

3.1. Data collection and preparation

LINCS-L1000 level 5 perturbation gene expression data from five modalities (shRNA, CRISPR, drug treatment, ligand stimulation, overexpression) were obtained.

Anti-SARS-CoV-2 drug signatures were selected from ChEMBL's SARS-CoV-2 dataset. Respiratory virus infection signatures for viruses were downloaded from GEO and processed for differential expression.

Ligand-receptor interactions from curated OmniPath data were used to select receptor and ligand perturbation signatures.

Cytokine stimulation signatures were sourced from CytoSig and Immune Dictionary datasets, normalized by gene-wise z-scores and filtered.

Patient transcriptomic and survival data included nivolumab and everolimus-treated cohorts and renal cell carcinoma single-cell RNA-Seq data were obtained from GEO.

3.2. Functional genomics and analysis

Pathway activities were inferred using PROGENy; transcription factor activities were calculated with DoRothEA.

Signature similarity between virus-infected and drug-treated samples was quantified by Spearman's correlation of gene expression signatures. Random Forest classifiers were used on

TF activities to predict antiviral drug effectiveness, with cross-validation and feature importance analysis.

Image segmentation was performed, cytoplasm (IC) and plasma membrane (PM) boundaries were defined. The log2 ratio of PM to IC fluorescence intensity was calculated per cell. Ordinary least squares regression assessed associations of intensity ratios with treatment, time, and drug conditions.

3.3. RIDDEN model development

Linear regression models related receptor perturbation status (activation/inhibition/no perturbation) to gene expression changes, creating a receptor-gene coefficient matrix. Ligand perturbations were mapped to corresponding receptors using OmniPath interactions. Receptor activities in new samples were inferred by dot product of gene expression profiles and the coefficient matrix, with significance assessed by permutation-based z-scores.

3.4. Benchmarking and patient data analysis

RIDDEN performance was benchmarked against CytoSig using *in vitro* and *in vivo* cytokine perturbation datasets.

Associations between receptor activity or gene expression and patient survival were analyzed by Cox regression and log-rank tests on immune checkpoint blockade-treated patient transcriptomic data.

4. Results

4.1 A comparison of the disease and the drug-induced signatures reveals similar responses in the cells.

We analyzed host cell transcriptomic responses to SARS-CoV-2 infection and the effects of drugs that were reported to be effective against the virus, aiming to identify mechanisms of action and potential drug repurposing candidates using a signature similarity-based approach. We compared infection-induced and drug-induced gene expression signatures in lung epithelial cell lines, infer pathway and transcription factor (TF) activities, and assess whether drugs induce similar or opposite gene expression changes to the antiviral response.

Functional genomic analysis revealed that SARS-CoV-2 infection activates innate immune pathways, including TNFa, NFkB, JAK-STAT and suppresses the key cholesterol-regulating TFs, SREBF1/2, in host cells. Several antiviral effective drugs, identified from the ChEMBL database, induced TF activity profiles that were similar to those of SARS-CoV-2 infection. Notably, a group of drugs demonstrated high activation of SREBF TFs, which is in contrast to the virus-induced suppression of these factors. Additionally, we found that anti-SARS-CoV-2 drugs have a significantly higher signature similarity to infection signatures than other drugs.

4.2 Drug perturbation signature similarity to infection signatures predict the antiviral activity and identifies antiviral mechanisms of action.

We showed that SARS-CoV-2 infection and several effective drugs elicit similar cell responses, which led us to explore how we can predict drug effectiveness using signature similarity. We used signature-similarity approach and machine learning, using random forest models to predict effectiveness and evaluated the predictions using ROC analysis with ChEMBL *in vitro* effective anti-SARS-CoV-2 drugs as positive values. The signature similarity scores showed predictivity to effective drug in SARS-CoV-2-infected cell lines. The random forest effectively classified antiviral drugs, and highlighted SREBF1/2 as most important features. We performed analysis on confocal microscopy images of cell lines treated with SREBF-activating drugs identified in our analysis. These TFs are key regulators of cholesterol biosynthesis, which suggested a potential effect on cholesterol metabolism. We determined fluorescence intensity of cholesterol sensor in plasma membrane (PM) and cytoplasm marker (IC). These SREBF-activating drugs significantly decreased the PM/IC cholesterol ratio compared to control, confirming their cholesterol depleting effect from the plasma membrane.

4.3 Development and evaluation of the computational tool, RIDDEN, which infers receptor activities from ligand and receptor perturbation gene expression signatures.

The RIDDEN (Receptor actIvity Data Driven inferENce) computational tool was developed to infer receptor activities directly from transcriptomic data by leveraging large-scale ligand and receptor perturbation signatures from the LINCS L1000 dataset, combined with curated ligand-receptor interactions from OmniPath. Unlike other methods that rely on ligand and receptor expressions, RIDDEN utilizes receptor footprints, which capture the downstream transcriptional consequences of receptor activity change, rather than assuming that the presence of ligands and receptors directly reflects signaling activity. This approach enables the estimation of receptor activity from bulk and single-cell transcriptomic profiles. RIDDEN was benchmarked against CytoSig, a state-of-the-art method for predicting cytokine signaling activities, using both *in vitro* bulk cytokine perturbation profiles and *in vivo* single-cell immune response data from the Immune Dictionary as ground truths. RIDDEN demonstrated comparable or improved predictive performance, which highlights its applicability for understanding receptor-driven signaling in diverse biological contexts.

4.4 RIDDEN identifies biomarkers of cancer therapy response

We demonstrated the applicability of RIDDEN to infer receptor activities of patient samples and investigated how these are associated with cancer immune-checkpoint blockade therapy response. RIDDEN inferred receptor activities from pretreatment bulk gene expression data and revealed that PD-1 receptor activity was significantly associated with improved overall survival in these patients, whereas the gene expression levels of PD-1 and its ligand PD-L1 were not predictive. These results highlight how RIDDEN can reveal important signaling activities that are not apparent from gene expression alone. Moreover, single-cell analysis showed the PD-1 receptor activity can be inferred in specific cell types likely to be present in the tissue microenvironment, such as T cells and tumor–-associated macrophages, but absent in tumor and other stromal cells. This demonstrates RIDDEN's ability to provide mechanistic insights into intercellular communication and its relation to therapeutic response.

5. Conclusions

This work demonstrates applicability of perturbation gene expression signatures as a tool for drug repurposing and receptor activity inference, highlighting their value in uncovering mechanisms of action and guiding hypothesis generation.

First, I leveraged SARS-CoV-2 infection-induced gene expression signatures and compared them with drug treatment-induced perturbation profiles to gain insights into antiviral drug mechanism of action. In contrast to the classical signature-reversal approach, we found that effective antiviral drugs often mimic adaptive host responses, activating pathways like NFkB and JAK-STAT. Several of these drugs also modulated SREBF1/2, key regulators of lipid metabolism. Experimental validation using fluorescent cholesterol sensors confirmed that these drugs reduce plasma membrane cholesterol and that this depletion contributes to their antiviral effect. These findings refine our understanding of signature-based drug repurposing in viral contexts and highlight cholesterol modulation as a key antiviral mechanism.

I demonstrated the systematic use of chemical and genomic perturbation signatures. I developed RIDDEN (Receptor actIvity Data Driven inferENce), a computational tool that infers receptor activity from transcriptomic profiles of 229 receptors.

Unlike co-expression-based methods, RIDDEN leverages the downstream transcriptional response of receptor perturbation to provide insights into cell-cell communication. Benchmarking on *in vitro* and *in vivo* datasets showed that RIDDEN accurately reflects the transcriptional responses of the cytokine receptor modulation and performs comparably to, or better than, existing models in this context. When applied to a cancer immunotherapy dataset, RIDDEN identified receptor activities associated with treatment response, revealing biologically meaningful signals that gene expression is not able to capture.

These applications demonstrate the value of perturbation gene expression signatures and how they can serve as a bridge between high-dimensional data and interpretable biological insight, guiding hypothesis generation in drug discovery and advancing our understanding of complex cellular responses.

6. Bibliography of the candidate's publications

Publications related to the thesis:

I. Barsi, Szilvia ; Varga, Eszter ; Dimitrov, Daniel ; Saez-Rodriguez, Julio ; Hunyady, László ; Szalai, Bence
RIDDEN: Data-driven inference of receptor activity from transcriptomic data
PLOS COMPUTATIONAL BIOLOGY 21 : 6 Paper: e1013188 , 20 p. (2025) IF 3.6

II. Barsi, Szilvia ; Papp, Henrietta ; Valdeolivas, Alberto. ; Tóth, Dániel J. ; Kuczmog, Anett ; Madai, Mónika ; Hunyady, László ; Várnai, Péter. ; Saez-Rodriguez, Julio ; Jakab, Ferenc ; Szalai Bence
Computational drug repurposing against SARS-CoV-2 reveals plasma membrane cholesterol depletion as key factor of antiviral drug activity
PLOS COMPUTATIONAL BIOLOGY 18 : 4 Paper: e1010021 , 20 p. (2022) IF 4.3

Publications not related to the thesis:

I. Damouni, Amir ; Tóth, Dániel J. ; Barsi, Szilvia ; Nagy, Dániel Károly ; Kasbary, Alexander ; Hunyady, László ; Cserző, Miklós ; Várnai, Péter
Differential activation of the inositol 5-phosphatase SHIP2 by EGF and insulin signaling pathways
JOURNAL OF BIOLOGICAL CHEMISTRY 301 : 7 Paper: 110275 , 18 p. (2025) IF 3.9

II. Gém, Janka Borbála ; Kovács, Kinga Bernadett ; Barsi, Szilvia ; Hadadnejadtehrani, Saba ; Damouni, Amir ;

Turu, Gábor ; Tóth, András Dávid ; Várnai, Péter ; Hunyady, László ; Balla, András
Role of LMCD1 in the Long-Term Effects of Angiotensin II in Vascular Smooth Muscle Cells
INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES 26 : 9 Paper: 4053 , 21 p. (2025) IF 4.9

III. Kovács, Kinga Bernadett ; Szalai, Laura ; Szabó, Pál ; Gém, Janka Borbála ; Barsi, Szilvia ; Szalai, Bence ; Perey-Simon, Bernadett ; Turu, Gábor ; Tóth, András Dávid ; Várnai, Péter ; Hunyady, László ; András, Balla
An Unexpected Enzyme in Vascular Smooth Muscle Cells: Angiotensin II Upregulates Cholesterol-25-Hydroxylase Gene Expression
INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES 24 : 4 Paper: 3968 , 15 p. (2023) IF 4.9

IV. Széky, Balázs ; Mayer, Balázs ; Gyöngy, Miklós ; Hajdara, Anna ; Barsi, Szilvia ; Kárpáti, Sarolta ; Németh, Krisztián
Tri-Lineage Differentiation of NTERA2 Clone D1 Cells towards Neural, Hepatic and Osteogenic Lineages in Vitro
FOLIA BIOLOGICA 67 : 5-6 pp. 174-182. , 9 p. (2021)
IF 1.167

V. Barsi, Szilvia ; Szalai, Bence
Modeling in systems biology: Causal understanding before prediction?
Patterns 2 : 6 Paper: 100280 , 3 p. (2021) IF 0

VI. Mohammadzadeh, Amir ; Lakatos, Péter P. ; Balogh, Mihály ; Zádor, Ferenc ; Karádi, Dávid Árpád ; Zádori,

Zoltán S. ; Király, Kornél ; Galambos, Anna Rita ; Barsi, Szilvia ; Riba, Pál, Benyhe, Sándor ; Köles, László ; Tábi, Tamás; Szökő, Éva ; Hársing, Lászlo G. Jr ; Al-Khrasani, Mahmoud
Pharmacological Evidence on Augmented Antiallodynia Following Systemic Co-Treatment with GlyT-1 and GlyT-2 Inhibitors in Rat Neuropathic Pain Model
INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES 22 : 5 Paper: 2479 , 15 p. (2021) IF 6.208

VII. Zádor, Ferenc ; Mohammadzadeh, Amir ; Balogh, Mihány ; Zádori, Zoltán S. ; Király, Kornél ; Barsi, Szilvia ; Galambos, Anna R. ; László, Szilvia B. ; Hutka, Barbara ; Váradi, András ; Hosztafi, Sándor ; Riba, Pál ; Benyhe, Sándor ; Fürst, Susanna ; Al-Khrasani, Mahmoud
Comparisons of in Vivo and in Vitro Opioid Effects of Newly Synthesized 14-Methoxycodeine-6-O-sulfate and Codeine-6-O-sulfate
MOLECULES 25 : 6 Paper: 1370 , 22 p. (2020) IF 4.412

VIII. Al-Khrasani, Mahmoud ; Mohammadzadeh, Amir ; Balogh, Mihály ; Király, Kornél ; Barsi, Szilvia ; Hajnal, Benjamin ; Köles, László ; Zádori, Zoltán S ; Hársing, László G.
Glycine transporter inhibitors: a new avenue for managing neuropathic pain
BRAIN RESEARCH BULLETIN 152 pp. 143-158. , 16 p. (2019) IF 3.370

The cumulative impact factor: **36.757**