SEMMELWEIS EGYETEM

DOKTORI ISKOLA

Ph.D. értekezések

3293.

BARSI SZILVIA

Celluláris és molekuláris élettan

című program

Programvezető: Dr. Hunyady László, egyetemi tanár

Témavezetők: Dr. Szalai Bence, tudományos munkatárs

Dr. Hunyady László, egyetemi tanár

SYSTEMS BIOLOGY APPLICATION OF PERTURBATION GENE EXPRESSION PROFILES

PhD thesis

Szilvia Barsi

Molecular Medicine Doctoral School

Semmelweis University

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

Official reviewers: Erzsébet Fichó, 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

Table of Contents

List of Abbreviations

AUC	Area under the curve
BKNN	Batch balanced k-nearest neighbors
CCC	Cell-cell communication
Ch25h	Cholesterol 25-Hydroxylase
CMap	Connectivity Map
CLUE	CMap linked user environment
CoV	Coronavirus
COVID-19	Coronavirus Disease 2019
CRISPR	Clustered regularly interspaced short palindromic repeats
DEG	Differentially expressed genes
FPR	False positive rate
GEO	Gene Expression Omnibus
GSEA	Gene Set Enrichment Analysis
GTEx	Genotype-Tissue Expression
H5N1	Highly pathogenic avian influenza virus A
HMG-CoA	3-hydroxy-3-methylglutaryl coenzyme A
HPIV	Human parainfluenza virus
IAV	Influenza A Virus
IC	Intracellular
ICB	Immune checkpoint blockade
IRF	Interferon Regulatory Factor
JAK	Janus Kinase
LIANA	LIgand-receptor ANalysis frAmework
LINCS	Library of Integrated Network-Based Cellular Signatures
NFkB	Nuclear Factor-kappa B
NIH	National Institute of Health
MERS	Middle East Respiratory Syndrome
mTOR	Mammalian target of rapamycin
mRNA	Messenger ribonucleic acid

MβCD	methyl β-cyclodextrin
MWU	Mann-Whitney U
OLS	Ordinary Least Squares
OS	Overall survival
PD-1	Programmed Cell Death Protein 1
PD-L1	Programmed Cell Death Ligand 1
PM	Plasma membrane
PROGENy	Pathway RespOnsive GENes
RF	Random Forest
RIDDEN	Receptor actIvity Data Driven inferENce
RIG-I	Retinoic acid-inducible gene I
RNA	Ribonucleic acid
RNA-Seq	RNA-Sequencing
ROC	Receiver-operating characteristic curve
RSV	Respiratory Syncytial Virus
SARS-CoV	Severe Acute Respiratory Syndrome Coronavirus
shRNA	Short hairpin RNA
SMILES	Simplified molecular-input line-entry system
SREBF1/2	Sterol Regulatory Element Binding Transcription Factor 1/2
STAT	Signal Transducer and Activator of Transcription
TNFa	Tumour Necrosis Factor alpha
TF	Transcription factor
TPR	True positive rate
UMAP	Uniform Manifold Approximation and Projection
VIPER	Virtual Inference of Protein-activity by Enriched Regulon analysis

1. INTRODUCTION

1.1. Perturbation signatures

Understanding the complex functions and mechanisms of cells is essential for deciphering how they maintain homeostasis, adapt to their environment, respond to external stimuli, and develop diseases. Systems biology addresses these challenges by integrating experimental data with computational modeling to map and interpret the complex interaction networks within a cell. System-level measurement techniques, such as RNA sequencing (RNA-Seq) and proteomics, generate vast datasets, but these have specific limitations. For example, transcriptomics, which measures messenger RNA (mRNA) abundance, may not align with the activation of functional proteins due to various regulatory mechanisms. However, measuring the activity of a protein is low-throughput, capturing only one functional state under a specified condition, and does not provide system-level insights.

To overcome the limitations of these measurement techniques, computational methods have been developed to provide more biologically relevant and interpretable insights into high-throughput data, such as pathway activity prediction tools or drug target identification using similarity-based or machine learning-based methods. These methods utilize large-scale transcriptomic data and integrate prior biological knowledge to infer functional outcomes from observed changes in gene expression. Most powerful methods rely on perturbation transcriptomics profiles, where cellular responses to genetic or chemical perturbations are measured. By the perturbations, we can capture the downstream effects of the perturbed protein or gene in the form of gene expression signatures. Fortunately, there are large community efforts to provide publicly available datasets for the research community to help the deeper understanding of cellular processes and responses. On public databases, like Gene Expression Omnibus (GEO) (1) and ArrayExpress (2), baseline and treatment gene expression data are also deposited, however, these experiments are conducted in laboratories worldwide and follow various protocols, so it is challenging to harmonize them into a comprehensive data source. Furthermore, baseline experiments, where the gene expression is measured in the untreated samples, do not provide insight into how specific genes, proteins,

or pathways influence downstream effects. To understand causal relationships, it is essential to modulate gene expression to uncover the downstream consequences.

Advances in transcriptomics and high-throughput screening enable the cost-effective generation of perturbation signatures. The Connectivity Map (CMap) presented a low-cost, high-throughput, reduced representation expression profiling method, the L1000 method, as part of the National Institute of Health (NIH) Library of Integrated Network-Based Cellular Signatures (LINCS) Consortium (3). This library contains perturbation-induced molecular and cellular signatures on a large scale. Data generation followed common data standards, allowing for easy data harmonization. Chemical and genomic perturbagens are also applied to various cells and time points to enable the investigation of wide-range perturbation effects, like diverse inhibition and activation levels, knock-down, knock-out, or overexpression. LINCS L1000 technology enables the measurement of 978 landmark genes that can explain ~90% of the variance of total transcriptomics difference, and with the usage of this, the remainder of the transcriptome is inferred. The library contains over a million profiles. Since then, several applications of the LINCS have been demonstrated to identify repurposing candidates from drugs that can reverse the expression profiles. The similarity in perturbation signatures between drug effects and diseased signatures can suggest repurposing opportunities without requiring prior knowledge of direct interactions. Several significant findings include the identification of a cyclin-dependent kinase inhibitor that can attenuate the endotoxemic process in sepsis (4), an inhibitor targeting diabetic kidney disease (5), and compounds capable of reversing a mutant p53-MYC–dependent signature in head and neck squamous cell carcinoma (6).

Besides the LINCS library, there are other sources of perturbation signatures, but these usually cover only a smaller part of the molecular functions, like cytokine transcriptional responses of the cell lines or *in vivo* samples. A recently published method and resource, known as CytoSig (7), collects cytokine stimulation signatures, where cell lines are perturbed with cytokines, from various public databases such as GEO. The authors developed a model to predict the cytokine signaling activities from bulk and single-cell transcriptomic profiles using this atlas of transcriptional patterns induced by cytokines. Computational methods applying this approach help to generate novel biological insights.

In recent years, along with *in vitro* response data, new datasets have emerged that measure the cellular responses of *in vivo* samples to various perturbations. The Immune Dictionary is a large dataset that measures how different immune cell types respond to cytokines in mouse lymph nodes at the single-cell level. (8). This study enhances our understanding of the activation states of different immune cell types. In addition, cell-cell communication (CCC) is essential for coordinating immune responses among different immune cells. Cytokines are the ligands in this process, binding to specific receptors on target cells to trigger immune responses. The Immune Dictionary and the CytoSig allow for a broader understanding of the roles cytokines play in these immune responses.

Additional methods have been developed to improve the accessibility of single-cell perturbation profiles, such as Perturb-Seq, which combines CRISPR-based genetic perturbations with single-cell RNA sequencing (9) or MIX-Seq, which enables multiplexed transcriptional profiling of post-perturbation responses across many samples (10), showing the potential benefits of these in the field of functional genomics and drug discovery (11).

Although these datasets can serve as a basis for research, they all have their limitations. LINCS L1000 profiles, however, reduce the cost of measurements, are the largest bulk perturbation dataset yet, and have shown that the whole transcriptome can be inferred from them, but still have limited transcriptome coverage as only the nearly 1000 genes are actually measured, which decreases the reliability of the remaining segment of their transcriptome. Other bulk or single-cell perturbation profiles either capture one segment of biology or are not measured across different cell types from various tissues and various perturbations or modalities. The Immune Dictionary focuses on immune cell types and cytokine transcriptional responses. CytoSig aggregates bulk transcriptional response datasets from various public sources with experiments conducted in multiple laboratories, while Perturb-Seq is a methodology employed across different research centers. Despite the growing adoption and potential of Perturb-Seq, both approaches share the limitation of requiring data harmonization to ensure effective integration across diverse sources.

1.2. **Signature similarity in drug repurposing**

De novo drug development is a consuming and risky strategy for the pharmacological industry, meaning over a billion dollars (12,13) and more than 10 years (13,14) are spent on development and preclinical investigations, clinical trials, including failure costs. Repurposing drugs for new indications results in significant savings in costs and time during this process, additionally, available clinical data at the beginning of the development results in reduced risk for further progress (15). Advancements in computational biology, measurement techniques, and growing high-throughput experimental data sources enable researchers to identify novel applications of the existing drugs. Comparing the drug treatment-induced and disease transcriptomic signatures has proven to be an effective approach for repurposing drugs for new therapeutic indications (16). The concept of transcriptome signature reversal is a general approach to determining whether a drug can reverse a disease-related gene expression signature. If a drug can induce gene expression that is opposite the disease signature, it has a potential to reverse gene expression changes associated with the disease, and thus leading to the reversal of the disease phenotype (17). This approach is shown to be effective in the field of oncology (17), inflammatory (18), metabolic disorders (18), central nervous system disorders (19), or infectious diseases, like drug repurposing against Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infection (20–22). Additionally, when infected cells activate adaptive antiviral mechanisms, like interferon pathways, inhibiting these pathways may lead to different effects on viral replication, possibly resulting in an increase rather than just a decrease in viral levels (23). Besides identifying potentially effective drugs for repurposing, gaining insight into the mechanisms of action is essential. Furthermore, validating the findings would be crucial. Unfortunately, many studies lack this validation (23).

1.3. **Mechanistic understanding**

Besides measuring expression levels of particular genes or proteins, it is crucial to understand the processes that link expression levels and changes between different conditions to the associated phenotypes, whether as a cause or consequence. However, there are cases where differential gene expression, such as upregulation of a gene, correlates with active

protein, for instance, the Cholesterol 25-Hydroxylase (Ch25h) expression change correlates with the activity of the enzyme upon angiotensin II stimulation in vascular smooth muscle cells (24), measurement of the mRNA level often does not correlate with protein abundance (25). The levels of proteins in a steady state are not directly linked to the levels of mRNA expressed. The time needed for transcription, processing, and transport of mRNA is significantly shorter than the time required to reach a new steady-state level of protein (26). Thus, single gene expressions do not provide insight into the processes present in the cell as a result of cell-cell interactions, stimulation, or drug treatment.

To interpret high-dimensional gene expression profiles and gain more insight into cellular processes and phenotypes, several computational tools and databases have been developed. The most widely used database is the Molecular Signatures Database (27), which contains annotated gene sets for gene set enrichment analysis (GSEA). This enrichment analysis provides insight into ranked gene expression by identifying gene sets exhibiting significant skew in their ranks. The annotated gene sets encompass various pathway annotations, biological processes, cellular components, or molecular functions (28). These sets use prior biological knowledge derived from published information about biochemical pathways or co-expression (29). The method and prior knowledge have their limitations, the method maps transcript expression to signaling proteins while neglecting the effects of post-translational modifications or the topology of pathways (30,31) and does not model activities of pathways based on the weighted mode of regulation (32). To address these challenges, methods were developed that consider these weights of regulations, such as VIPER (33) or linear models (32,34) and footprint-based methods that besides considering the weights, focus on functional protein targets or pathway-regulated genes (30,31,35). Footprint-based methods infer the activity of a protein or pathway by capturing the expression pattern of the genes it regulates, rather than measuring the abundance of the protein or gene itself. This approach provides a more biologically meaningful insight into protein function (35). Furthermore, combining these with perturbation signatures has the advantage of measuring the initial gene expression response to a specific stimulus in a given cellular context, which directly links upstream functional proteins to their target genes.

1.4. **Drug repurposing in the COVID-19 pandemic**

COVID-19 pandemic outbreak emerged in 2019 caused by highly infectious SARS-CoV-2 coronavirus. The number of infections has grown rapidly, posing a significant threat to public health (36) by causing severe acute respiratory syndrome and other diseases. Therefore, the scientific community united to explore potential therapeutics for the virus and disease. In addition to vaccines, the promising strategy has been drug repurposing, as available safety profiles for the therapeutics facilitate the approval process for the new indication and help quicker access to treatments (37).

Computational approaches, like machine learning and other *in silico* drug discovery approaches, were implemented for repurposing drugs. The existing large amount of data deposited in public repositories helps this investigation. Besides this data, in 2020, at the beginning of pandemic, host transcriptional response signatures following SARS-CoV-2 infection were also deposited (38,39). Furthermore, ChEMBL released a large-scale drug screening study to identify potential anti-SARS-CoV-2 agents in cell-based assays (40,41) to contribute to this effort. These datasets give way to drug repurposing strategies by comparing the transcriptomic signatures induced by active anti-SARS-CoV-2 agents to the infection-induced signatures to identify potential drugs (17).

1.5. **Cell-cell communication**

The communication between cells is essential for functioning and coordinating complex processes within and between them, such as maintaining homeostasis, cell growth, differentiation, or immune interactions (42). Receptors are the key proteins that can receive signals from the environment and initiate processes in response to them. In the case of viral infections, the cells initiate immune responses and release cytokines, the ligands that bind to cytokine receptors to further promote immune response in the surrounding cells (43,44). Dysregulation of the receptor activation, caused by altered ligand binding, mutation, or overexpression of the receptor, can lead to various diseases. Receptors that initiate altered signaling are potential therapeutic targets. Developing methods that identify receptor activity states within cells can help to understand the cellular processes and explain their phenotype. Several computational tools have been developed, including CellChat (45), CellPhoneDB

(46), to analyze CCC. Most methods utilize prior knowledge of intercellular interactions and statistical methods to prioritize them to identify potential communication events (47). However, they assume that co-expression implies signaling and do not evaluate whether these interactions result in functional downstream responses. Some other approaches, such as the NicheNet (48) models CCC by integrating prior knowledge of gene regulatory networks with gene expression from interacting cells.

These CCC methods may provide mechanistic insight into cell-cell communication-induced processes, opening the opportunity to identify disease-associated changes in these interactions.

2. OBJECTIVES

The objectives of this study are to utilize perturbation gene expression signatures and demonstrate their applicability and potential to advance drug repurposing strategies and gain mechanistic insights into the processes of communicating cells. First, we aim to investigate an infectious disease, the SARS-CoV-2 infection, focusing on identifying potential mechanisms of action of *in vitro* effective antiviral drugs. Second, we aim to develop a computational tool for gaining mechanistic insights into processes induced by cell-cell communication by inferring receptor activities from transcriptomic profiles.

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. Collection and preparation of public datasets

3.1.1. LINCS perturbation signatures

Level 5 gene expression profiles from the LINCS-L1000 dataset were obtained from the Data Library of the CMap linked user environment (CLUE) (3). We downloaded perturbation data from 5 modalities: shRNA, CRISPR, drug treatment, ligand stimulation, and overexpression. We calculated consensus gene expression signatures for each drug (for antiviral drugs) or each perturbation-cell line pair (for receptors and ligands) using the moderated-z score (MODZ) method. Briefly, weighted averages of z-score vectors of perturbation signatures are summarized to a single differential expression vector based on Spearman correlation, representing a consensus signature of the given perturbation. We calculated the consensus signature using landmark genes (978), which are the measured genes from the L1000 profile for antiviral drug investigation and RIDDEN construction. Additionally, we used the inferred genes (11350+978) for benchmarking the RIDDEN.

3.1.2. Selecting drug signatures from ChEMBL for anti-SARS-CoV-2

We obtained ChEMBL (41) 27 SARS-CoV-2 release dataset (40) to collect drugs that showed *in vitro* anti-SARS-CoV-2 activity. We aligned drugs from LINCS-L1000 with ChEMBL drugs from this release using drug names and the simplified molecular-input line-entry system (SMILES).

3.1.3. Download respiratory virus-induced signatures from GEO

We downloaded virus infection *in vitro* microarray experiments from GEO, with accession numbers GSE28166 (Highly pathogenic avian influenza virus A, H5N1) (49), GSE37571 (Influenza), GSE33267 (SARS-CoV-1) (50), GSE56677 (51) and GSE45042 Middle East respiratory syndrome–related coronavirus (MERS-CoV) (52) and RNA-Seq profiles of SARS-CoV-2 and other virus-infected human cell lines with accession numbers GSE147507 (SARS-CoV-2, Respiratory syncytial virus (RSV), Influenza A Virus (IAV), Human parainfluenza virus (HPIV)) (39) and GSE148729 (SARS-CoV-1 and 2) (38). The

raw microarray profiles were processed with *limma (53)*, and RNA-Seq profiles with *DESeq2* (54) R packages to calculate differential expression profiles between mock and infected samples.

3.1.4. Selecting receptor and ligand signatures using OmniPath

We queried the OmniPath database (55) using the OmniPath R package for ligand and receptor interaction networks (56) that are compiled from the LIgand-receptor ANalysis frAmework (LIANA) CCC resource (47). We filtered for a curated list of known interactions. Based on this, we selected receptor and ligand perturbation signatures from the 5 modalities from LINCS for the construction of the receptor activity prediction model.

3.1.5. CytoSig and ImmuneDictionary collection

For the RIDDEN model predictive performance evaluation, we downloaded the cytokine stimulation signatures from 2 resources, the CytoSig (7) and the ImmuneDictionary (8). The CytoSig contains gene expression profiles gathered from public databases, including transcriptional responses of cytokines across different cell lines from *in vitro* assays. The ImmuneDictionary contains responses of 17 immune cell types to 86 cytokines measured in *in vivo* experiments on mouse lymph nodes by single-cell RNA-Seq. We used gene-wise z-scores for normalization, and we aggregated the cell signatures into average cytokine perturbation signatures of the immune cell types of ImmuneDictionary. We excluded signatures that contain only a few differentially expressed genes (8), as they do not capture the responses to perturbations. For benchmarking the RIDDEN model, we selected the overlap between these cytokines and LINCS cytokines and their receptors, where the interaction is defined based on the OmniPath intercellular interaction network (56).

3.1.6. Transcriptional profiles of patients

We obtained gene expression profiles of pretreatment samples and the overall survival of patients to nivolumab, a PD1 inhibitor, and everolimus, an mTOR inhibitor, from the study (57).

We downloaded 7 patients' raw single-cell transcriptomic profiles with clear cell renal cell carcinoma (58). We prepared the data following the method described in the publication (59,60), with the difference of using the Batch balanced k-nearest neighbors

(BBKNN) algorithm (61) for batch correction, and we modified the annotation based on grouping cell types defined by different marker genes into broader cell type categories. We visualized the cells using Uniform Manifold Approximation and Projection (UMAP) 2-D projection.

3.2. Investigation of anti-SARS-CoV-2 compounds signatures and mechanism of action

3.2.1. Functional genomics analysis

We used computational tools to get insights into the cellular processes, we inferred pathway activities using PROGENy (R package progeny) (30,31) and inferred transcription factor (TF) activities using DoRothEA, a curated resource of signed TF-target interactions (regulons) (62) and VIPER algorithm (33) of the different virus-induced signatures and LINCS drug perturbation signatures. PROGENy inferred z-scores of pathway activities by using 1000 permutations of gene labels for background distribution. In DoRothEA, TFs are assigned to confidence levels from A to E, and we selected the high-quality regulons (A, B, C levels).

3.2.2. Signature similarity for repurposing

We used Spearman's correlation (scipy Python library) to calculate the signature similarity in expression signatures of shared genes between virus-infected samples and antiviral drug treatment samples.

3.2.3. Machine learning-based methods

We predicted effective drugs against SARS-CoV-2 using a Random Forest (RF) Classifier algorithm (scikit-learn Python library) on the TF activity scores of the drug-treated signatures. The classifier was trained with default parameters, except for number of trees, which was set to 300. Other default settings included the Gini impurity criterion, bootstrap sampling enabled, and no restriction on maximum tree depth. We used 100 random training sets, each containing 50% of the effective and 50% of the non-effective drugs. We calculated the mean importance of the features (TFs) as well as the mean probabilities of antiviral

activity for each drug. To assess the predictive performance of both similarity-based and machine learning-based predictions, we performed Receiver Operating Characteristic (ROC) analysis, where the positive class consisted of the antiviral drugs from the LINCS database that had measured evidence for the anti-SARS-CoV-2 effect (in ChEMBL (40)) and the negative class contained the rest of the drugs. ROC curves were computed for each of the 100 different validation sets to compare machine learning-based methods with similarity-based methods, considering the signature similarity scores of the corresponding drugs.

3.2.4. **Image analysis**

Automatic confocal microscopy images from an experiment conducted by the co-authors of (23) were used to determine the cholesterol sensor ratio between the plasma membrane (PM) and the cytosol in HEK293A cells, where cells were transfected with fluorescent cholesterol sensor. Images were segmented by *CellPose* (63), a deep learning-based segmentation method (CellPose Python library). The cytoplasm marker channel was used for filtering high-quality images using Laplace filtering and used for the input of the CellPose model with parameter cell diameter set to greater than 200 pixels and channel set to greyscale. To determine the PM outer boundary and cytoplasm boundary, we applied a binary dilation algorithm with a default structure and 5 iterations and a binary erosion algorithm with a default structure and 10 iterations, respectively, after determining the cell boundaries by CellPose. The PM boundary was obtained by subtracting the cytoplasm boundary from the outer boundary. The log2 ratio of the mean PM and mean intracellular (IC) cholesterol sensor (D4H) fluorescence intensities (log2(PM/IC)) for each cell in the D4H channel was calculated to examine the cholesterol distribution at the PM. The following ordinary least squares (OLS) linear regression (statsmodels Python package) model was used to determine the association between the intensity ratio and change: log2(PM/IC) ~ Time + Time: Drug + Experiment, where Time denotes the elapsed time after drug treatment, the Drug represents the used drug with DMSO as reference and Experiment denote the individual experiments (n=3).

3.3. RIDDEN

3.3.1. Model construction for receptor activity prediction

We used the following OLS regression models (statsmodels Python package) to estimate the relationship between receptor perturbation and gene expression: $g_i = \beta_{0i} + \beta_{i,j}\, r_j + \varepsilon_i$, where r_j is a vector of samples (cell-receptor perturbation pairs), where each value denotes the perturbation of receptor j in the sample. (1 if stimulation, -1 if inhibition, and 0 if no perturbation). The curated ligand-receptor interaction prior knowledge from OmniPath (56) was used to find the corresponding receptors of the ligands and translate the ligand perturbation into receptor perturbation. The g_i represents the gene expression values (from LINCS consensus signatures), where the values are the expression level of a specific gene in the sample. The β_{0i} is the intercept for gene i and $\beta_{i,j}$ is the coefficient representing the association between the receptor j perturbation and the expression of gene i, ε_i is the error term for gene i. We merged the computed coefficients into a receptor-gene parameter matrix, the RIDDEN matrix, which represents the relationship between the gene expression change under different conditions (perturbation type, direction, cell line) and the receptor perturbation.

3.3.2. Inference of receptor activities using the RIDDEN

Receptor activities can be inferred from a transcriptional profile by calculating the dot product of the gene expression of the sample and the RIDDEN matrix. The RIDDEN tool calculates a z-score of receptor activities by generating a background distribution through 1000 permutations of gene labels as the reference.

3.3.3. Cross-validation of receptor signatures

We fitted linear models (OLS statsmodel Python package) for each of the five modalities separately using the same method described in 3.3.1. and used the parameter matrices to infer receptor activities from the perturbation signature matrices of the other modalities as described in 3.3.2. We calculated ROC Area under the curve (AUC) to assess the performance of predicting which receptor was perturbed in the sample. Inhibiting and

activating perturbation were handled separately, as described in the publication (64). For the final RIDDEN model, we filtered for receptors with high predictive performance based on the following criteria and fitted linear models for the remaining receptors, including all modalities. We used receptors that had greater than or equal to 0.6 ROC AUC in at least one prediction. We did not include cases where the CRISPR model predicted shRNA perturbation and conversely, because if the receptor is not present in the cell, the receptor knock-out or knock-down will give us incorrect information about the potential gene expression change upon receptor perturbation. The filtering resulted in 229 different receptors.

3.3.4. **Assigning confidence levels to receptors**

We assigned confidence levels to the receptors. We randomly split the dataset five times in half, trained a model on the training set, and predicted receptor activities on the test set. On each split, we used the Mann-Whitney U (MWU) test to determine whether there is a significant difference in receptor activities between the perturbed and non-perturbed conditions. Inhibitory and stimulatory perturbations were analyzed separately. The mean of the receptors was calculated, including all splits. Confidence levels from A to E were assigned to the receptors based on the aggregated p-values of the MWU test.

3.3.5. **Prediction performance comparison**

We compared the predictive performance of RIDDEN with CytoSig (7) in predicting perturbed cytokines or cytokine receptors. For ground truth datasets, we used CytoSig's cytokine perturbation bulk transcriptomic profiles and ImmuneDictionary's cytokine perturbation single-cell transcriptomic profiles. The comparison was conducted in the following scenarios: 1) RIDDEN model's performance on CytoSig dataset, 2) CytoSig model's performance on RIDDEN's cytokine and cytokine receptor dataset with signature containing inferred and landmark genes for equitable comparison, 3) RIDDEN's performance on ImmuneDictionary dataset, and 4) CytoSig model's performance on ImmuneDictionary dataset. We evaluated the performance of the models with ROC AUC and used separate ROC curves for each receptor. We handled different perturbation directions separately (inhibitory and stimulatory), and the receptor perturbations indicated the positive

values. We used overlapping cytokines and cytokine receptors between the 2 resources for each comparison. We translated between ligand and receptor perturbations using LIANA's ligand-receptor interaction resource. (47,56,65)

3.3.6. Patient's survival associations with receptor activity and gene expression

We inferred receptor activities from bulk transcriptomic profiles of nivolumab and everolimus-treated samples (57) (described in Methods - Transcriptional profiles of patients section) using RIDDEN. For investigation of the associations of the receptor activity and gene expression with patient's overall survival, we performed Cox regression analysis using the lifelines Python package (66) with the following equation: $h(t|x) = h_0(t) * \exp(\beta * X)$, where $h(t|x)$ denotes the hazard function, $h_0(t)$ is a baseline hazard, β is the coefficient and X denotes the receptor activity or the gene expression. Additionally, we used the log-rank test, in which we separated patient groups by the mean of the gene expression or receptor activity.

3.3.7. Application availability

The RIDDEN tool is available at https://github.com/basvaat/RIDDEN_tool.

4. RESULTS

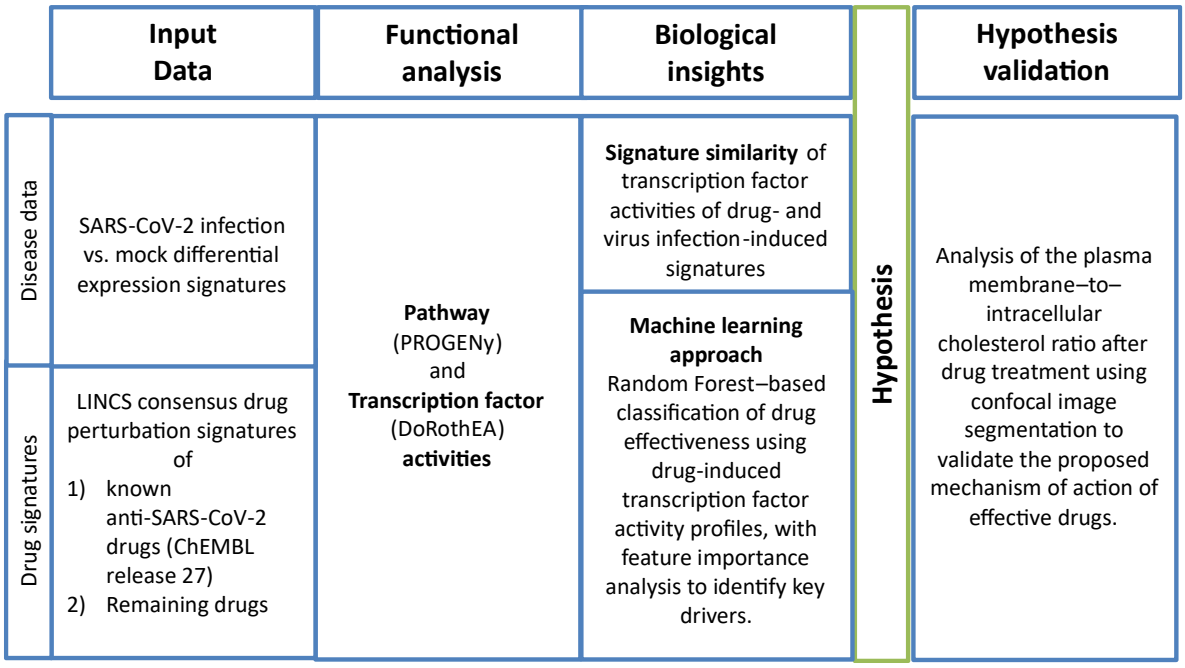

Figure 1. Pathway and TF activities were inferred from virus- and drug-induced transcriptomic signatures to gain mechanistic insights into effective antiviral drug mechanisms of action or effects of SARS-CoV-2 infection. Signature similarity and machine learning analyses were then applied to link these mechanisms to drug effectiveness, generating hypotheses that were subsequently validated by confocal image segmentation and quantitative analysis.

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

We investigated how cells responded to SARS-CoV-2 infection and to drugs that were reported to be effective against the virus. We used drug perturbation gene expression profiles to gain insights into the mechanism of action that makes these drugs effective.

4.2. Functional genomic analysis of SARS-CoV-2-infected cells suggests an adaptive response

We analyzed two lung epithelial cancer cell lines, the Calu-3 and A549 SARS-CoV-2 infection-induced gene expression signatures (infected-control) (38,39). Using PROGENy

and DoRothEA, we inferred pathway and TF activities (see Methods 3.2.1.). Increased TNFa and NFkB pathway activities were observed in both cell lines, while the JAK-STAT pathway was active only in the infected Calu-3 cell lines (Figure 2A). Accordingly, NFkB and STAT TFs were strongly activated, in addition to the IRF TF. The cell growth-related TFs, the E2Fs, and Myc showed decreased activity. Interestingly, Sterol regulatory element-binding factors 1/2 (SREBF1/2), the key transcriptional regulators of cholesterol synthesis, were also inactive in the infected cells (Figure 2B). The activated pathways and TFs are known to be involved in the host innate immunity against viral infection (67), furthermore, TFs with decreased activity are part of the interferon response (68,69).

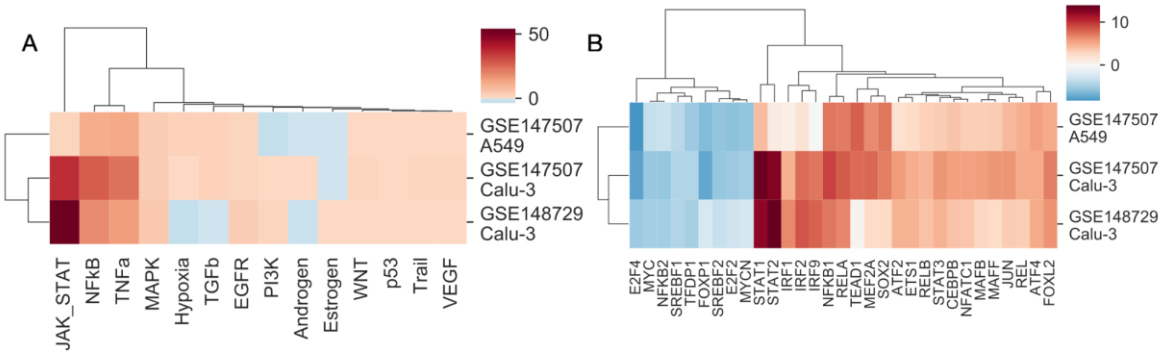

Figure 2. Transcription factor and pathway activities of the SARS-CoV-2-infected Calu-3 and A549 cell lines.

A) Heatmap of PROGENy pathway activities of the infected Calu-3 and A549 cell lines, with positive values indicating activation (red) and negative values indicating inhibition (blue).

B) Heatmap of DoRothEA TF activities of the infected Calu-3 and A549 cell lines, with positive values indicating activation and negative values indicating inhibition. Only transcription factors (TFs) exhibiting significant variations in activity (absolute normalized enrichment score greater than 4) are shown.

With the help of a signaling network contextualization tool, the CARNIVAL (70), our collaborator analyzed which upstream signaling pathways regulate the inferred activity changes using the Retinoic acid-inducible gene I (RIG-I)-like receptors, the key receptors in

virus recognition (71), as initiators of the signaling. This analysis showed that SARS-CoV-2 infection-induced RIG-I-like receptor activation can lead to the reported TF activity changes, with the AKT1 and MAPK1 intermediate nodes (23). Additionally, there is evidence that these TFs are modulated by infection (72–75).

4.3. Functional genomic analysis of drug-treated signatures shows similarities to infection-induced signatures in cells

Using the 4671 LINCS L1000 drug-induced consensus signatures and a set of *in vitro* effective anti-SARS-CoV-2 drugs from the ChEMBL database, including 47 overlapping drugs with LINCS (see Methods 3.1.2.), we performed a functional genomic analysis using PROGENy and DoRothEA. NFkB and TNFa pathways were activated by several *in vitro* effective drugs, including niclosamide, perhexiline, and digoxin. JAK-STAT was also activated by receptor tyrosine kinase inhibitors. (Figure 3A). The activation of these pathways can also be observed in the infection-induced signatures. Interestingly, TF activity analysis revealed that the SREBFs are activated by a large cluster of drugs (Figure 3B). We calculated the average TF activities for the 5 largest drug clusters and plotted them against the average TF activities of the 3 SARS-CoV-2 signatures. The general observation across the clusters is a prominent increase in STATs and a decrease in E2F4 TF activity. We observed a high or lower, but significant positive correlation between the drug-induced and diseased signatures in four out of the five clusters (Spearman's rho = 0.64, 0.14, 0.18, 0.58 and 0.04, p = $8.55*10^{-35}$, 0.0122, 0.00174, $3.14*10^{-27}$ and 0.484 from upper right to lower left panel in Figure 3C). In two drug clusters, we found high drug-induced activity of SREBF1/2, which is opposite to the inhibition of these TFs by SARS-CoV-2 infection (Figure 3C). We investigated the similarity of virus-induced signatures to LINCS by Spearman's correlation, which was shown to be an effective method for analyzing the similarity to LINCS L1000 data (3,76). We found that anti-SARS-CoV-2 drugs (ChEMBL drugs) have a higher similarity to infection signatures than other drugs (Figure 3D, Mann-Whitney U test p-value = $<1*10^{-200}$).

The findings indicate that drugs known to have effective anti-SARS-CoV-2 effects *in vitro* induce transcriptomic signatures and TF activity profiles similar to those induced by

SARS-CoV-2 infection. Additionally, a group of drugs demonstrates high activation of SREBF, which plays a key role in regulating cholesterol metabolism.

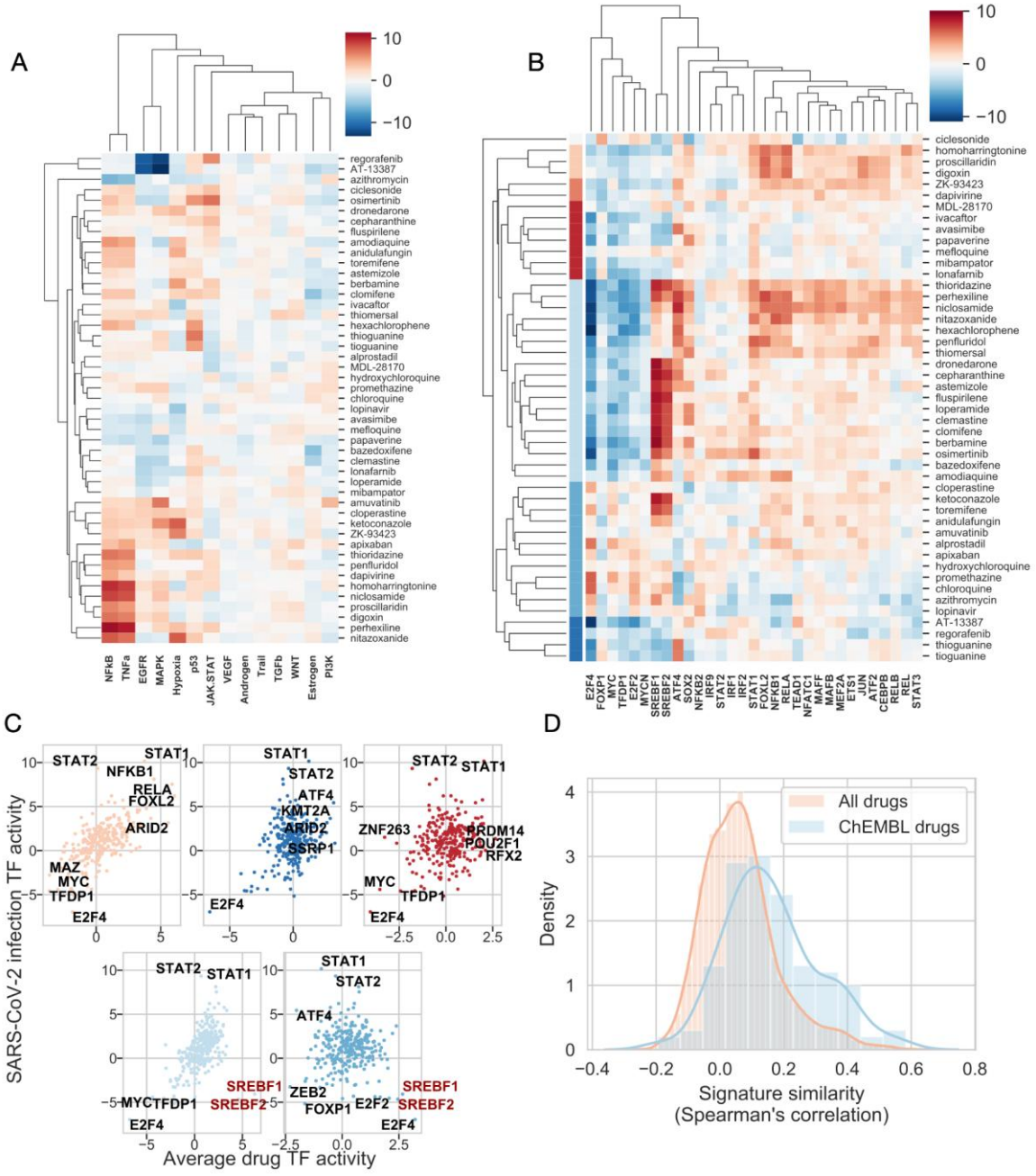

Figure 3. Functional genomic analysis of *in vitro* effective drug treatment (ChEMBL drugs).

A) Heatmap of the PROGENy pathway activities calculated from LINCS consensus signatures of the 47 ChEBML drugs. The red color indicates activation, while blue indicates inactive pathways.
B) Heatmap of TF activities (DoRothEA) calculated from LINCS consensus signatures of the 47 ChEMBL drugs. The red color indicates activation, while blue indicates inactive TF. The drugs were clustered based on TF activity profiles, and 5 clusters were identified. The clusters are indicated by color on the left side of the heatmap.
C) The scatterplots show the relationship between the average TF activity of drug-treated cell lines (calculated from the average of LINCS drug perturbation signatures of the *in vitro* effective ChEMBfL drugs) in the 5 largest drug clusters and the average TF activity profile of 3 SARS-CoV-2-infected cell lines. Top TFs are marked. Cluster colors match those in (B).
D) The plot illustrates the distribution of signature similarity between drug-induced signatures and average SARS-CoV-2-induced signatures for effective anti-SARS-CoV-2 drugs (ChEMBL drugs) and all LINCS-L1000 drugs (All drugs).

4.4. **Drug perturbation signatures and their similarity to infection signatures predict the antiviral activity**

We showed that SARS-CoV-2 infection and several effective drugs elicit similar responses, which led us to explore how we can predict drug effectiveness using signature similarity. We used unsupervised and supervised learning strategies to predict effectiveness and evaluated the predictions using ROC analysis by using ChEMBL *in vitro* effective anti-SARS-CoV-2 drugs as positive values.

The similarity scores (Methods 3.2.2.) show predictivity to effective drugs (Figure 4A) in SARS-CoV-2-infected cell lines (ROC AUCs: 0.75, 0.74, and 0.64 for GSE147507 A549, GSE147507 Calu-3, and GSE148729 Calu-3, respectively). Furthermore, we aimed to investigate if the signature similarity is predictive in the case of other respiratory viruses

infected Calu-3 and/or A549 cell lines, like other coronaviruses (SARS-CoV (50), MERS-CoV (51,52)), influenza (Influenza A (49)), H5N1, parainfluenza (HPIV) or RSV (Methods 3.1.3.). These infection signatures exhibited less predictive performance for the ChEMBL anti-SARS-CoV-2 drugs with ROC AUC values <0.7, except for one SARS and RSV signature with ROC AUCs 0.70 and 0.71, respectively Figure 4B), indicating the relative specificity of the similarity-based approaches for SARS-CoV-2.

We classified drugs into effectivity groups using the drug-induced TF activity profile as features with the supervised Random Forest (RF) algorithm (Methods 3.2.3.). The positive class consisted of the 47 effective ChEMBL drugs. We established a cross-validation method based on random subsampling and assessed its performance using the ROC AUC (Methods 3.2.3.). We compared the ROC AUC values of the RF-based model and similarity-based values of the three SARS-CoV-2 signatures (Figure 4C). The mean ROC AUC values are 0.72 and 0.68, 0.66, 0.57, respectively, indicating a slight improvement in the predictive performance of the machine-learning-based method (paired t-test p-values between machine learning and similarity-based methods: $3.02*10^{-7}$, $2.76*10^{-15}$, $4.89*10^{-15}$ for GSE147507 A549, Calu-3, and GSE148729 Calu-3 signatures, respectively). To get more mechanistic insights into the drug effects, we used RF classification and Gini importance. We identified SREBF1 and 2 as the most important TFs in predicting effective drugs (Figure 4D).

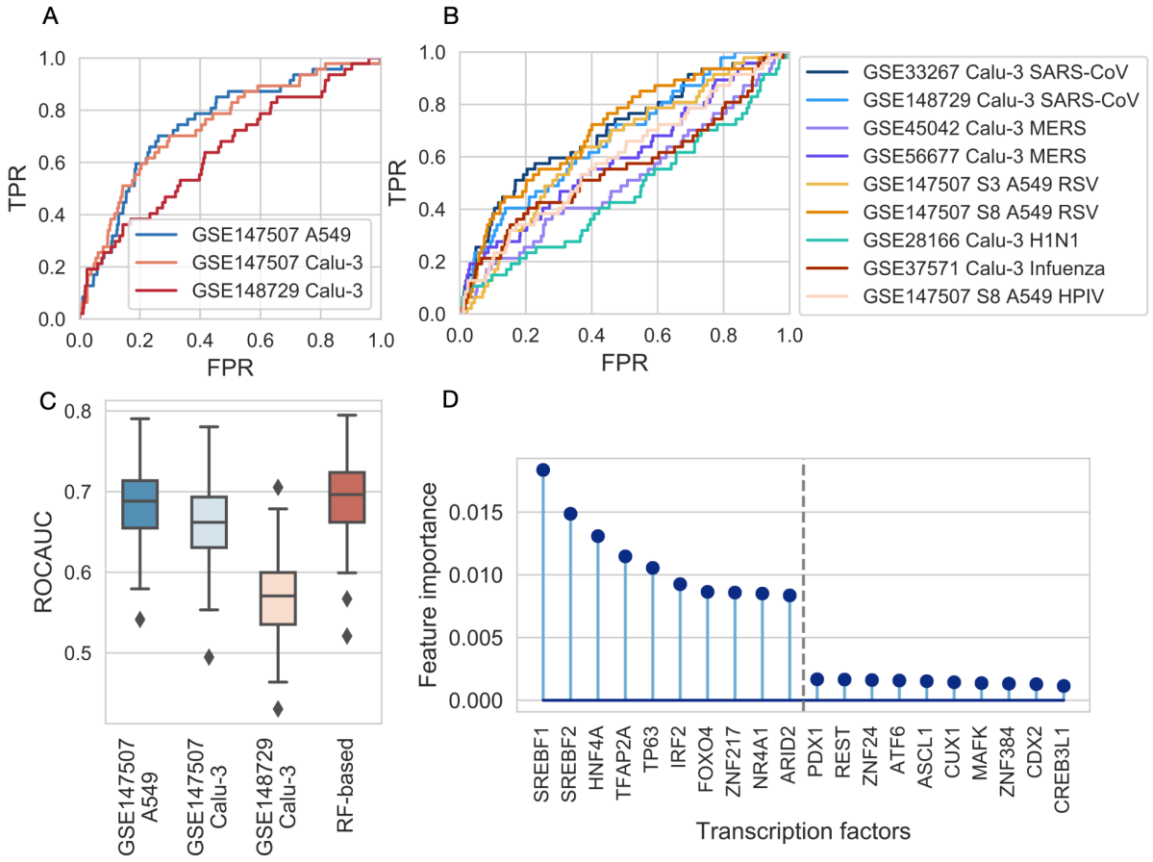

Figure 4. Similarity-based and machine-learning-based methods predict effective drugs

A) ROC curves evaluating the performance of similarity-based predictions using drug–SARS-CoV-2 infection signature similarity. Known *in vitro* effective anti-SARS-CoV-2 drugs from the ChEMBL dataset were used as the positive class. FPR refers to False Positive Rate, TPR refers to True Positive Rate

B) ROC curves evaluating similarity-based predictions using drug–other respiratory virus infection signature similarity. Known in vitro effective anti-SARS-CoV-2 drugs from the ChEMBL dataset were used as the positive class.

C) ROC AUC values of the similarity-based values and the RF Classification algorithm predicting the effective drugs from ChEMBL. Random subsampling and cross-validation were used to predict the effectivity. Boxplots represent the median (central line), first and third quartile (box),

minimum and maximum non-outlier values (whiskers), and outliers (diamonds).

D) The top and bottom 10 features ranked by Gini importance in the RF classification model trained on TF activity of drug perturbation signatures and predicting *in vitro* effectiveness against SARS-CoV-2.

With the unsupervised approach, we achieved similar predictive performance to the similarity-based approach and, interestingly, found that the SREBFs contribute the most to predicting drug antiviral effectiveness.

4.5. *In vitro* validation of plasma membrane depletion by SREBF-activating drugs and their antiviral effect

We demonstrated that SARS-CoV-2 infection inhibits SREBF TFs, while a range of anti-SARS-CoV-2 drugs activate them. This shows a contrasting effect compared to other TFs that are similarly activated or inhibited by both conditions in cells.

SREBF TF can be activated by the reduction of the plasma membrane and endoplasmic reticulum cholesterol levels. Its activation promotes the production of cholesterol and other lipid-synthesizing enzymes (77). Consequently, reduced activity of SREBF caused by virus infection may lead to decreased cholesterol synthesis, hindering viral replication and/or viral entry into cells (69). Thus, this effect can be considered as an adaptive response of the host cell (Figure 5). Based on our hypothesis, some *in vitro* effective drugs directly reduce plasma membrane cholesterol, which contributes to an antiviral effect. This mechanism may lead to an increased SREBF activity due to the lower cholesterol levels in the plasma membrane (Figure 5).

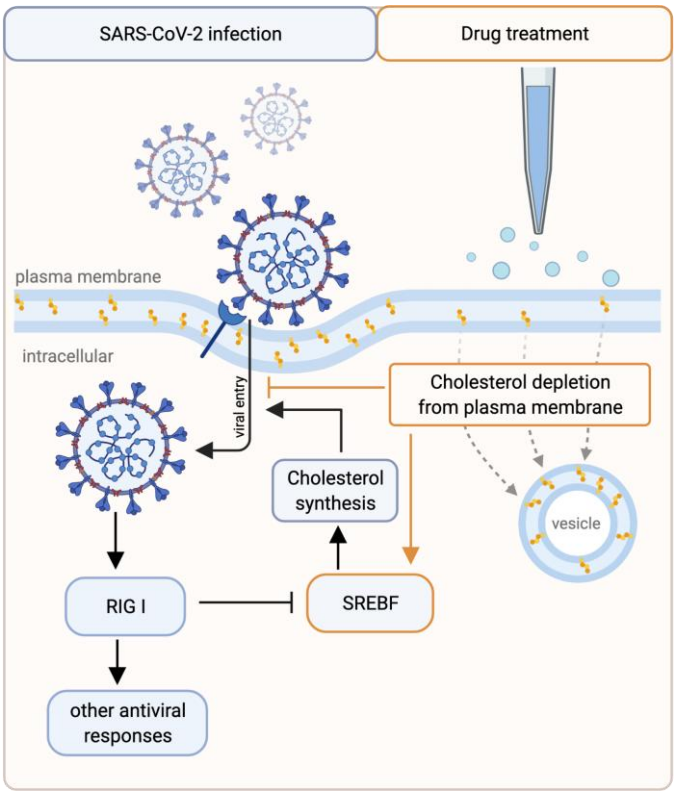

Figure 5. Schematic figure of the hypothesis that antiviral drugs block virus entry into cells by cholesterol depletion from plasma membrane, and are leading to a compensatory increased SREBF1/2 activity. The effects caused by viral infections are indicated by black arrows on the left side, whereas the changes caused by drugs are shown with orange arrows on the right side. RIG I denotes retinoic acid-inducible gene-I-like receptors, which are pattern-recognition receptors that sense viral RNA and trigger antiviral responses, like interferon signaling.

To confirm this hypothesis, our colleagues conducted an *in vitro* experiment using automated confocal microscopy imaging on HEK293A cells that were infected with a fluorescent cholesterol sensor, D4H-mVenus (78,79), and cytoplasmic Cerulean as a cytosolic marker. The cells were treated with negative control DMSO, a plasma membrane cholesterol-depleting compound methyl β-cyclodextrin (MβCD) as a positive control, and three drugs: loperamide, amiodarone, and chlorpromazine, which were selected based on their *in silico* SREBF-activating effects (all drugs were used in 10 μM final concentration).

Amiodarone was one of the top-predicted drugs of the Random Forest model (Figure 6A), while loperamide and chlorpromazine are part of the ChEMBL anti-SARS-CoV-2 drug set (Figure 6B). In the experiment, rosuvastatin was also applied, which does not directly influence plasma membrane cholesterol, but inhibits 3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA) reductase, a de novo cholesterol synthesis limiting enzyme.

For a systematic and unbiased analysis of PM cholesterol changes, I calculated the ratio of the average plasma membrane (PM) and average intracellular (IC) D4H-mVenus fluorescence (PM/IC ratio) for each cell on the captured images (see Methods 3.2.4.). Investigating the ratio as a function of elapsed time after drug treatment, we did not observe a decrease in loperamide and rosuvastatin-treated samples, while MβCD, chlorpromazine, and amiodarone treatment significantly decreased the ratio (linear model p values: $<10^{-200}$, $<10^{-200}$, $2.71*10^{-9}$, 0.25 and 0.0047 for MβCD, chlorpromazine, amiodarone, loperamide, rosuvastatin respectively, Figure 6C). This confirms that plasma membrane cholesterol was depleted by chlorpromazine and amiodarone, which were also identified as SREBF-activating drugs.

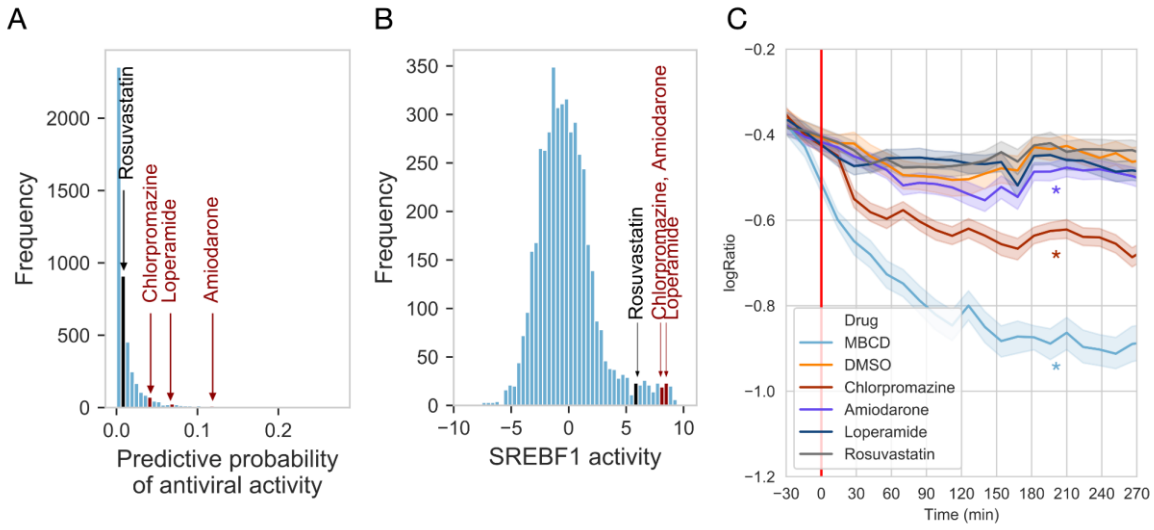

Figure 6. SREBF-activating drug PM cholesterol-depleting effects.

A) Predicted probability of *in vitro* antiviral activity in the Random Forest model and

B) Histogram of SREBF1 activity of the drugs, where SREBF-activating drugs and rosuvastatin, an HMG-CoA reductase inhibitor, are highlighted.

C) Time-dependent change of log2(PM/IC) ratio of average cholesterol sensor intensity in HEK293A cells treated with DMSO, MβCD, chlorpromazine, amiodarone, loperamide, or rosuvastatin. The red line marks drug treatment. *: significant ($p<0.001$) interaction between drug treatment and elapsed time in the linear model.

Cholesterol rescue experiments were performed by co-authors of (23) to test whether decreased plasma membrane cholesterol depletion affects SARS-CoV-2 infectivity. In these experiments, amiodarone significantly decreased the number of viral particles in Vero-E6 cells, while cholesterol replenishment significantly increased the number of viral particles in cells treated with amiodarone. This confirms that the cholesterol depletion caused by amiodarone treatment plays a crucial role in its antiviral effect (23).

4.6. Inference of receptor activities from ligand and receptor perturbation gene expression signatures

We established a model that infers receptor activity from the transcriptomic profiles, called RIDDEN, Receptor actIvity Data Driven inferENce. 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 required a curated database of ligand-receptor interactions obtained from OmniPath (55,56) and chemical (drug treatment, ligand stimulation) and genomic perturbation (CRISPR, shRNA, overexpression) signatures of these ligands and receptors from the LINCS L1000 database (3). To predict receptor activities, we created a dataset consisting of 38989 consensus perturbation signatures across five perturbation types applied to several cell lines for 599 receptors. Using this, we fitted linear models to describe the relationship between the perturbed receptor and each gene expression caused by the perturbation (see Methods 3.3.1.)

(Figure 7A). This method resulted in a parameter matrix of receptor-gene pairs that allows for inferring receptor activities from transcriptomic profiles of new samples, as it provides insight into what changes in the gene expression profiles are a consequence of altered receptor activity (Figure 7B). We use a permutation-based approach to estimate receptor activities in new samples (see Methods 3.3.2.). We applied receptor quality filtering to ensure the reliability of the prediction (see Methods 3.3.3.) Finally, the RIDDEN matrix consists of 14463 transcriptional signatures of 229 receptors that can be leveraged to infer receptor activities from bulk and single-cell transcriptomics. We also classified receptor activities into five confidence levels (A-E), where receptors with A confidence level had the best cross-validation performance (Methods 3.3.4.).

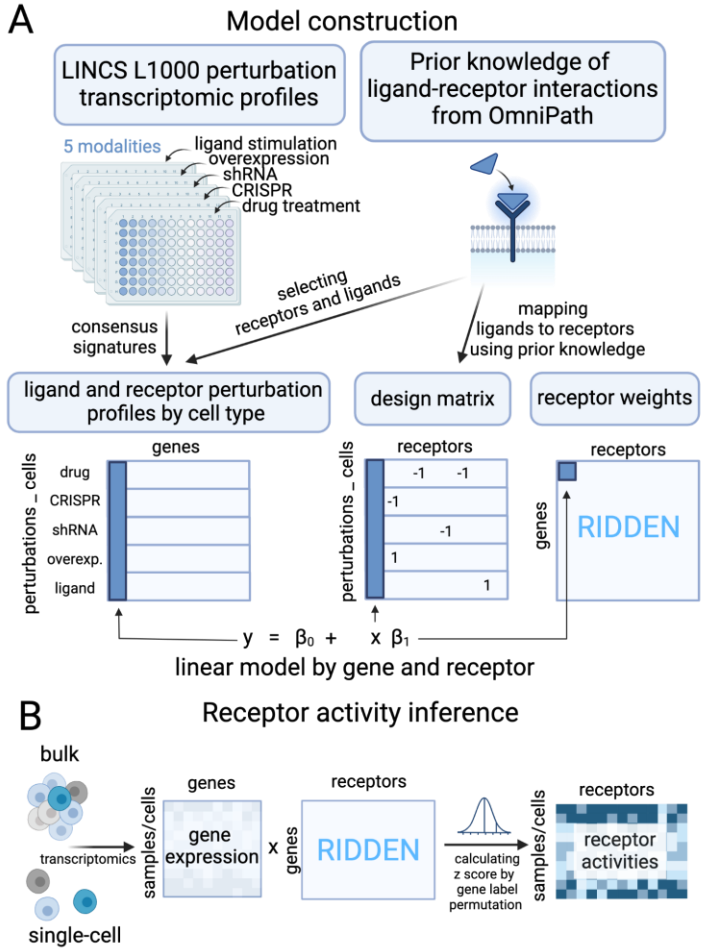

Figure 7. Establishment of the RIDDEN model

A) RIDDEN model construction. Ligand and receptor perturbation profiles are collected from the LINCS L1000 database, and their interaction are queried from the OmniPath. The ligands are mapped to corresponding receptors, known to be interacting according to prior knowledge, and linear models are fitted on the ligand-receptor perturbation profiles using the known receptor perturbations (+1 activated, -1 inhibited, 0 not perturbed) as input to create the RIDDEN.
B) RIDDEN estimates receptor activities from both bulk and single-cell transcriptomic data by calculating dot products of gene expression profiles and the RIDDEN weight matrix, followed by generating z-scores through random permutations of gene labels.

4.7. **Benchmarking the RIDDEN model**

We compared the RIDDEN to CytoSig (7), a model predicting cytokine (ligand) signaling activities from gene expression. Utilizing the Immune Dictionary (8), and the dataset from CytoSig containing cytokine stimulation signatures, we evaluated the performance of both models in predicting cytokine receptor and cytokine signaling activities based on cytokine and receptor perturbation signatures. This comparison assessed how accurately each model predicted the perturbations (see Methods 3.3.5.) on the *in vitro* cytokine and receptor perturbation datasets. The full RIDDEN yielded 0.61, while the CytoSig yielded 0.59 median ROC AUC. Moreover, the RIDDEN receptor models with confidence levels A-E reached 0.68, 0.64, 0.56, 0.51, and 0.52, respectively (Figure 8A). On the *in vivo* Immune Dictionary dataset, despite RIDDEN not explicitly modelling ligand perturbations, RIDDEN had a comparable ROC AUC score of 0.64 with CytoSig (0.67) (Figure 8B).

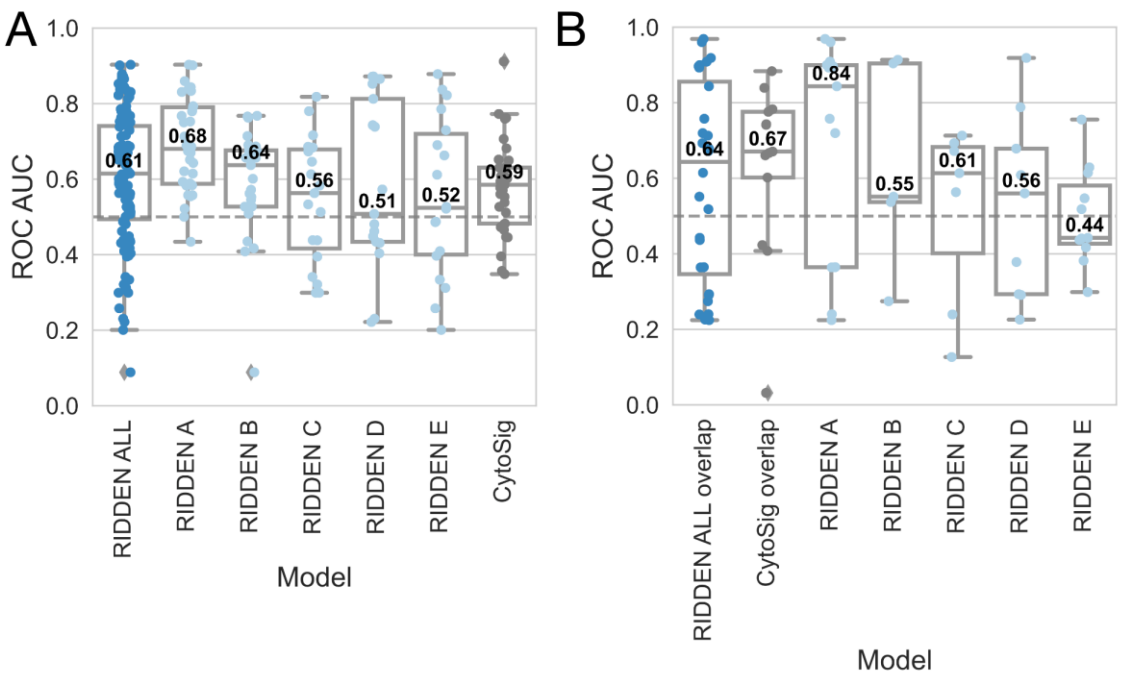

Figure 8. RIDDEN model benchmark

A) RIDDEN and CytoSig performance comparison. ROC AUC distributions were compared by applying RIDDEN and CytoSig to each other's training datasets. RIDDEN models were subsetted by receptor confidence scores, while CytoSig was evaluated on LINCS cytokine perturbation signatures (landmark + inferred genes). Boxplots show median ROC AUC (black numbers, central line), interquartile range (box), non-outlier range (whiskers), and outliers (diamonds).

B) RIDDEN and CytoSig evaluation on the Immune Dictionary dataset. The first two boxplots show performance on overlapping cytokines/receptors from the three resources. Subsequent boxes display RIDDEN's results across receptors grouped by confidence scores (A–E) on the ImmuneDictionary and RIDDEN receptor overlap. Boxplot features are as in (A).

4.8. **RIDDEN identifies biomarkers of cancer therapy response**

We assessed RIDDEN's ability to infer receptor activities of patients and how these are associated with cancer therapy response. We used clear cell carcinoma patients' gene expression profiles and response to immune-checkpoint blockade (ICB) therapy (see Methods 3.3.6.) (57), ICB targets ligands or receptors that are responsible for regulating immune responses, like the PD-1 receptor and its ligand, PD-L1 (80). To compare, the responses and transcriptomics of the mTOR inhibitor-treated patient cohort were utilized from the same study. From the patient's pretreatment samples, RIDDEN inferred receptor activities. We calculated the associations between receptor activity and gene expression and patient survival using the log-rank test and Cox regression. We found that PD-1 receptor activity was associated with overall survival (OS) in anti-PD-1 therapy (log-rank test p-value = $4*10^{-4}$, Cox regression $\beta = -0.36$, p-value = 0.012), nivolumab-treated samples but not in mTOR inhibitor, everolimus-treated patient samples (log-rank p = 0.092, Cox p = 0.106, Figure 9A). Neither the PD-1 nor the PD-L1 gene expression was associated with patient OS (log-rank test p-value = 0.465, 0.318, Cox p-value = 0.585, 0.630, Figure 9B). The negative association between PD-1 receptor activity and hazard rate suggests that higher PD-1 activity is associated with longer survival in nivolumab-treated patients. The nivolumab, by inhibiting the PD-L1/PD-1 interaction, is effective in patients with higher PD-1 activity. In nivolumab-treated patients, the chemokine receptors CXCR1 and CXCR2, as well as the serotonin receptor HTR2C and TGF-β receptors (TGFBR2 and TGFBR3), have significant negative coefficients in the Cox regression analysis, indicating a negative association between their activity and patient hazard. Conversely, BMPR1B, members of the ERBB receptor family, and EGFR showed the most significant positive coefficients, suggesting a positive association with the hazard (Figure 9C).

Furthermore, we predicted receptor activities based on the single-cell transcriptomic profiles of an independent cohort of renal cell carcinoma patients (Figure 9D).

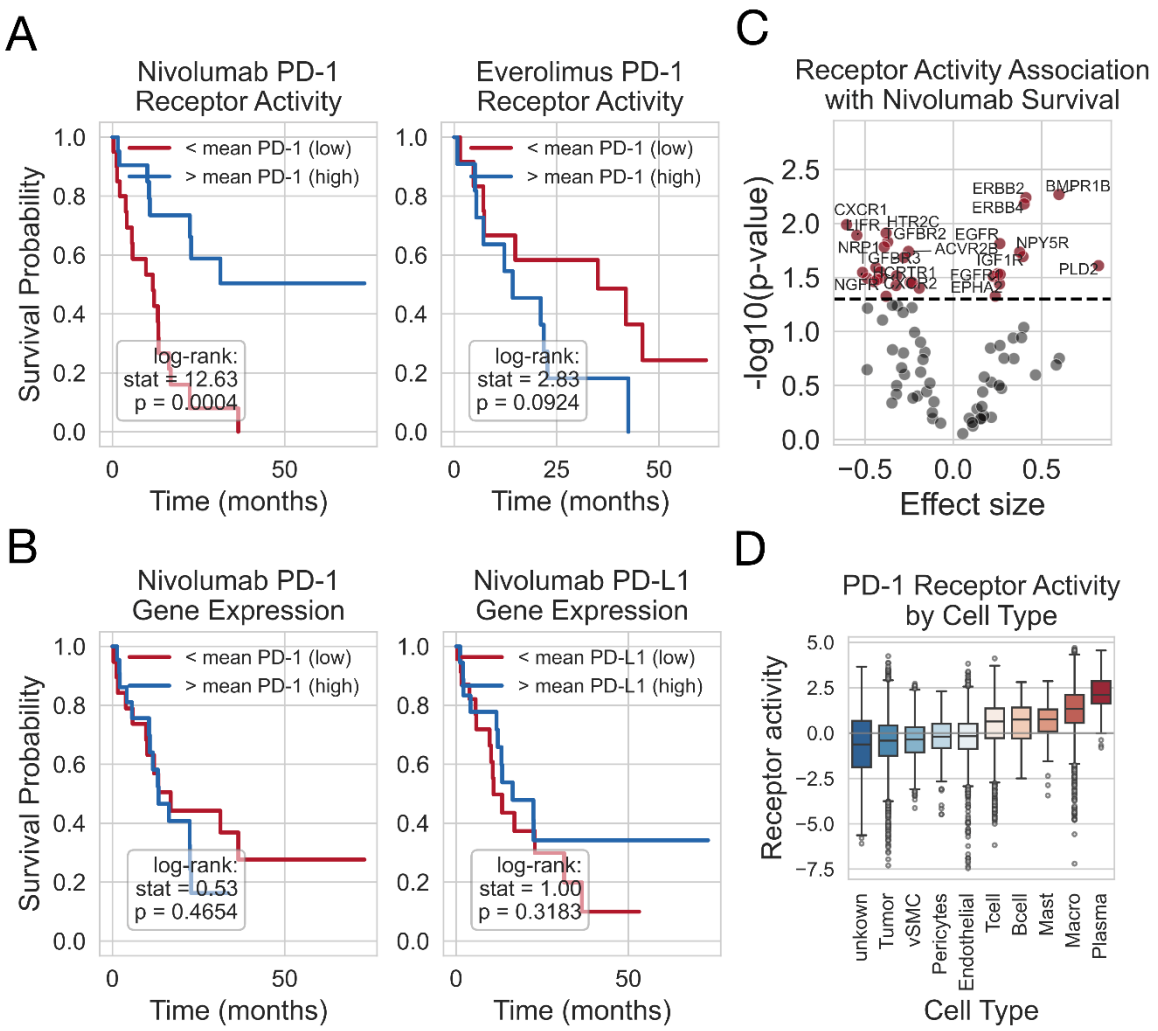

Figure 9. PD-1 expression and activity in tumor samples

A) PD-1 receptor activity is associated with improved overall survival in nivolumab-treated renal cell carcinoma patients (left), while no association is observed with everolimus treatment (right). Kaplan-Meier plots show survival probabilities stratified by PD-1 activity levels. Log-rank test results are indicated; blue and red curves represent high and low activity groups, respectively.

B) PD-1 and PD-L1 gene expression are not associated with patient survival in nivolumab-treated samples. Kaplan-Meier plots show survival based on PD-1 and PD-L1 gene expression levels using the mean as a threshold; no significant differences were observed.

C) Association of receptor activities with survival in nivolumab-treated patients. A volcano plot displays Cox regression coefficients (x-axis) and -log10 p-values (y-axis) for receptors with high-confidence scores. Significant receptors ($p < 0.03$) are labeled; the dashed line marks the $p = 0.05$ threshold.

D) Single-cell analysis of renal carcinoma samples reveals that PD-1 receptor activity is enriched in specific cell populations. Boxplots display PD-1 activity across major cell types, showing median activity values (central line), interquartile range (box), non-outlier range (whiskers), and outliers (diamonds). (The x-axis labels represent the following cell types: vMSC indicates vascular smooth muscle cells, Mast denotes mastocytes, Macro denotes macrophages, and Unknown refers to cells that cannot be characterized).

We found that PD-1 is active in macrophages and showed that the immune cells have high activity of PD-1, while no activity was observed in tumor cells, endothelial, perivascular, and smooth muscle cells.

5. DISCUSSION

Transcriptomic perturbation signatures, which quantify gene expression changes in a dose-dependent manner induced by genetic and chemical interventions (81), provide the possibility for measuring cellular responses that datasets measuring baseline gene expression, such as CCLE (82) or The Genotype-Tissue Expression (GTEx) (83), are not able to capture. While baseline signatures elucidate the normal, untreated state of cells, providing insights into physiological processes, disease states, and gene expression regulation (83), perturbation signatures enable a controlled investigation of how cells respond to various stimuli, such as drug treatments or genetic interventions like CRISPR knockout. By observing the transcriptional changes in cells resulting from the absence of functional proteins or altered activity, we can explore downstream effects and causal relationships. Previously, several studies have leveraged perturbation gene expression signatures for drug repurposing to identify novel biomarkers (84), resistance mechanisms (85), and therapeutic candidates (86). Although numerous studies have explored the use of perturbation gene expression signatures, our study (64) applies transcriptomic perturbation signatures to predict receptor activity from receptor-regulated transcriptional footprints, enabling the inference of more biologically meaningful receptor signaling activity from transcriptomic data compared to non-footprint-based methods.

In my research, I explored the application of perturbation signatures as a tool for drug repurposing against SARS-CoV-2 (23) and predicting receptor protein activity (64). The findings demonstrated that by utilizing genomic and chemical perturbation gene expression profiles, we can enhance our understanding of cellular responses and the underlying mechanisms of antiviral drug action. We demonstrated that perturbation signatures provide valuable insights into receptor activity. Furthermore, our results indicate that utilizing computational methods based on perturbation signatures can enhance the identification of effective treatments and improve the prediction of drug interactions and biological effects.

The growing number of studies in this field indicates that perturbation profiles offer valuable insights and are essential in this research field. While several perturbation profiles have been submitted to public databases, such as GEO (1) or ArrayExpress (2), they lack standardized experimental and processing workflows, and they often answer a specific

question, limiting their usability in a broader context. There is still a need for larger-scale databases with standardized measurements, like the LINCS L1000 (3) project. Current technologies enable the screening of perturbation profiles on a large scale, not only at the transcriptomic level but also at the protein level (87) and imaging-based approaches (88). Each method captures different aspects of the cellular responses, proteomics-based approaches capture the basic functional units of biological processes, or post-translational modifications (e.g., phosphorylation) (87), and image-based approaches can link morphological and phenotypic information to the cell response (88). The future of the perturbation-based approach may depend on integrating complementary modalities (11) to overcome the limitations of analyzing individual molecular layers by providing rich multi-omics information.

While measuring and integrating perturbation effects across multiple omics levels is beneficial, transcriptomics is currently the most widely used in drug discovery due to its broad accessibility. Several studies have shown that it can provide biologically relevant information about the state of the cell. This can be achieved by enrichment (29) of known genes of specific signaling pathways (27,89,90) or biological processes, molecular functions in differential expression signature (28,91), or by using methods that infer the activity of different proteins (33), such as transcription factors (62) or pathways (30,31). These methods are all well-suited for generating hypotheses; however, large-scale, well-annotated perturbation datasets are also crucial for benchmarking these methods and new computational approaches to ensure that they are robust across different biological contexts (11). For example, in this study, the Immune Dictionary (8) was used to compare the performance of RIDDEN and CytoSig (7) in predicting cytokine receptor and cytokine signaling activities. However, the lack of additional datasets limits the ability to benchmark our method across broader biological contexts.

There are several methods to gain functional insights from gene expression data. One of the most widely used is the GSEA, which links changes in gene expression to pathway or process annotations. However, this method assumes that the change in differential gene expression is directly associated with protein activity or abundance (11), although this is complicated by regulatory processes, post-translational modifications, or splicing (92,93).

An alternative approach, footprint-based methods, infers protein or pathway activity from the expression of their downstream targets, rather than relying on gene or protein abundance. (35). By combining perturbation signatures, these methods enable a context-specific assessment of regulatory activity. Studies have shown that this approach more accurately reflects biological phenotypes in multiple benchmarks (31,93,94). In my study, I utilized the latter approach to infer receptor protein activity. This method, as discussed in recent literature, analyzes the expression of genes regulated by the receptors rather than relying on receptor or ligand expression or co-expression. RIDDEN uses receptor-specific footprints learned from perturbation experiments, enabling the use of context-specific regulatory relationships (64).

I discussed the relevance of perturbation studies and how they can provide mechanistic insights. I also demonstrate their potential usability in the field of drug repurposing (23). Gene expression-based drug repurposing is a widely used method. Its applicability is demonstrated in several studies in cancer research (95,96) and other diseases (18,97). These studies apply a general approach focusing on compounds whose transcriptional perturbation profiles are inversely correlated with disease-associated gene expression signatures. In this work, I demonstrate that drugs similar to the SARS-CoV-2-induced, an infectious disease-induced gene expression signature, can also be effective. Whereas cancer and metabolic diseases often benefit from reversing pathological transcriptional states, viral infections like SARS-CoV-2 may require activating the host's antiviral response. We conducted a more unbiased analysis of drug repurposing based on gene signatures induced by SARS-CoV-2. By examining the gene expression patterns associated with known effective anti-SARS-CoV-2 drugs and comparing them to the signatures of SARS-CoV-2 infection, we found that the similarity between these signatures, rather than their dissimilarities, is predictive for antiviral efficacy. Our findings indicate that enhancing the antiviral responses of host cells may be a more effective approach than blocking the pathways activated by viral infection. It remains to be determined if this is unique to SARS-CoV-2, which is known for its ability to evade various antiviral defenses of host cells, or if it represents a broader mechanism applicable to other viral infections. By utilizing the "signature similarity" approach alongside network-based methods (23,70),

computational drug repurposing in the case of both established and emerging infectious diseases can be accelerated.

To gain mechanistic insight with the help of perturbation signatures, we performed functional genomic analysis on signatures of SARS-CoV-2-infected and drug-treated cells with anti-SARS-CoV-2. Similarly, in infected cells, especially in Calu-3 human airway epithelial cells and in drug-treated cells, we found NFkB and JAK-STAT pathway activation. This finding also supports the previously introduced concept of "signature similarity".

The analysis revealed significant findings, showing that SREBF1 and SREBF2 TFs were inhibited in samples infected with SARS-CoV-2. A comparison between machine learning approaches and a signature similarity-based method to predict antiviral activity based on perturbation signatures revealed that both methods can classify antiviral drugs with similar performance. Additionally, the machine learning approach identified SREBF1/2 as the most important features for predicting antiviral activity. These transcription factors are found to be activated in a large cluster of drug-treated cells, in contrast to the infected cells. SREBF1/2 are involved in regulating the expression of the key components of cholesterol synthesis (77). Previous research suggests that reducing plasma membrane cholesterol may inhibit SARS-CoV-2 infection (98,99). The downregulation of SREBFs and subsequent decrease in cholesterol synthesis can also be part of the physiological antiviral response (69,100,101). Additionally, previous studies showed that SREBFs and DNA interaction can lead to antiviral effect (102), and knocking-out components of the SREBF pathway leads to resistance against the virus (103). We observed an increase in SREBF activity in drug-treated cells, and this can be supported by studies showing the increased expression of lipid metabolism enzymes in antiviral drug-treated cells (22). Additionally, an increase in cholesterol synthesis has been shown to reduce SARS-CoV-2 infection (104).

Changes in gene expression may be a compensatory response rather than a cause. We propose that increased activity of SREBF1/2 could be a response to lower PM cholesterol levels following virus infection. To investigate this hypothesis, an experiment was designed and conducted by the co-authors of (23), where a fluorescent cholesterol sensor was used on antiviral drug-treated cells to monitor changes in PM cholesterol levels over time. Two *in vitro* effective antiviral drugs, amiodarone and chlorpromazine, were found to decrease PM

cholesterol levels. This was further supported by the finding of an *in vitro* SARS-CoV-2 infection assay, coupled with a cholesterol rescue experiment. This experiment demonstrated that replenishing cholesterol reduced the antiviral activity of amiodarone, thereby confirming the causal role of decreased plasma membrane cholesterol in the antiviral effect of amiodarone (23). This highlights importance of understanding how perturbation profiles can yield valuable insights into disease mechanisms and treatment strategies.

In addition to gaining valuable insights through analysis of perturbation signatures—as demonstrated in our study on SARS-CoV-2 and cholesterol depletion—these signatures have also proven valuable in building computational frameworks (7,30,48), where these signatures were systematically used to infer pathway or signaling activities and to model intercellular communication. Building on this foundation, we developed RIDDEN, a novel computational framework that utilizes the footprint-based approach and learns from receptor perturbation response signatures. This enables data-driven prediction of receptor activity from gene expression profiles. Unlike previous methods, RIDDEN leverages large-scale transcriptional response profiles, relies on curated pathway knowledge to map ligands to their receptors, and on the basis of these, predicts a large number of receptor activities.

Conventional CCC methods, like CellPhoneDB (46) or CellChat (45), rely on the prior knowledge networks of ligands and receptors and identify potential interacting ligands and receptors based on their gene expression. They do not capture whether these interactions lead to actual receptor activation or downstream signaling. In contrast, RIDDEN provides a more mechanistic perspective by inferring receptor activity from gene expression profiles, using perturbation transcriptional signatures (e.g., from receptor knockouts or activations) to assess whether a receptor is functionally active and what effect its activation has on the receiver cell. Furthermore, as a footprint-based method, RIDDEN derives more biological insights than gene set-based methods.

Benchmarking computational tools for activity inference and CCC is essential to assess their reliability and ensure biological relevance. However, the lack of large and harmonized ground-truth datasets limits this effort (105). Benchmarking the RIDDEN needs receptor perturbation datasets and prediction tools for receptor activities. Classical CCC methods are unsuitable for comparison with perturbation-based methods, as they depend on

the expression of endogenous ligands. Thus, they are unable to capture externally introduced ligands. CytoSig (7) provides a collection of cytokine perturbation signatures obtained from public databases and a tool for predicting cytokine signaling activities. Along with the Immune Dictionary, an *in vivo* dataset of cytokine perturbation responses can serve as the benchmark for RIDDEN. However, the comparison is limited to cytokines and their receptor signaling activities.

RIDDEN demonstrated comparable performance to CytoSig in cytokine signaling predictions while also providing confidence levels for receptor activity. In the evaluation of the top-performing receptors, the model outperformed CytoSig, demonstrating its robust performance in predicting cytokine receptor activity. RIDDEN has broader coverage of perturbation signatures, thus it can capture core signaling changes induced by receptor activity change across diverse cells.

The performance of the RIDDEN compared to CytoSig was assessed using the Immune Dictionary dataset. *In vivo* experiments model more complex responses to perturbations as they capture cellular responses within their microenvironment within a tissue. RIDDEN demonstrated a capability to predict receptor activities triggered by ligand stimulation in this complex environment. RIDDEN predicted which receptor-induced signaling was modulated by the immune ligand with performance comparable to CytoSig.

In our study (64), we extended the evaluation beyond predictive performance to biological relevance. RIDDEN also demonstrated value in analyzing patient therapy responses to an immune checkpoint inhibitor targeting the PD-1 receptor. While measurement of PD-L1 protein levels has been crucial for guiding therapy decisions and serves as an important marker of treatment response across various cancer types (106), targetable receptors can also be assessed directly from gene expression, providing additional and functional insights. Using pre-treatment gene expression profiles from patients treated with nivolumab, a PD-1 inhibitor RIDDEN was able to infer PD-1 receptor activity, which showed a significant association with overall survival. Notably, only the inferred receptor activity, and not the expression levels of PD-1 or its ligand, the PD-L1, were predictive of patient outcomes in Cox regression and effectively stratified patients based on survival in log-rank analysis. Additionally, we have observed associations between patient survival and

the activity of various other receptors that are already linked to immuno-oncology (107–112). These results highlight how RIDDEN can reveal important signaling activities that are not apparent from gene expression alone. Moreover, the PD-1 receptor activity can be accurately inferred in specific cell types likely to be present in the tissue microenvironment, such as T cells (113) and tumor-associated macrophages (114). In contrast, receptor activity was not predicted in tumor cells, likely due to the lack of a signature associated with receptor activation in these cells.

6. CONCLUSIONS

This study highlights importance and demonstrates potential applications of perturbation signatures in advancing drug repurposing strategies against virus infections and gaining mechanistic insights in transcriptional signatures over traditional tools, capturing baseline transcriptional states but not the causal consequences of regulation.

I systematically investigated the host transcriptional response to SARS-CoV-2 infection using computational functional genomics approaches. The analyses identified distinct pathways and transcription factor activities in infected cell lines, revealing an adaptive host immune response characterized by the activation of NFκB and JAK-STAT signaling, alongside suppression of sterol regulatory pathways. By comparing these infection-induced signatures with gene expression profiles of *in vitro* effective antiviral drugs, transcriptional similarities were observed that suggest effective drugs induce antiviral mechanisms in the host cell.

Notably, SREBF1/2 TFs were found to be active in several *in vitro* effective antiviral drugs. These factors are involved in cholesterol synthesis, opposing the virus-induced suppression of the same TFs. This observation led to the hypothesis that activation of cholesterol synthesis regulatory factors is the consequence of a decrease in cholesterol level. Thus, these drugs may exert antiviral effects via plasma membrane cholesterol depletion. I evaluated both the signature similarity approach and a machine learning model based on TF activities in classifying antiviral drugs. Both approaches effectively distinguished *in vitro* active from inactive antiviral compounds. Notably, the Random Forest model also identified SREBF1/2 as key features in predicting drug effectiveness, highlighting their relevance.

We validated the cholesterol depletion effect of SREBF-activating antiviral drugs through automated confocal microscopy imaging using a fluorescent cholesterol sensor. I performed the analysis of the detection of cholesterol levels in PM and cytoplasm on fluorescent images, confirming that the selected drugs indeed reduce plasma membrane cholesterol levels *in vitro.* These results support a mechanistic link between antiviral drug action and host lipid metabolism. Furthermore, amiodarone, one of the SREBF-activating drugs, was validated to exert its antiviral effect by cholesterol depletion in a cholesterol rescue experiment by the co-authors of our study (23).

I developed RIDDEN, a computational tool that infers receptor activity from gene expression signatures. It benefits from large-scale transcriptomic perturbation datasets and curated ligand-receptor interactions. While RIDDEN was developed as a systems biology application, it also demonstrates that perturbation signatures can provide mechanistic insights into receptor-driven signaling, highlighting broad applicability of these signatures.

I evaluated RIDDEN against the CytoSig model in predicting cytokine signaling and receptor activities, using an *in vitro* bulk and *in vivo* single-cell immune response dataset based on cytokine perturbations as ground truth. RIDDEN demonstrated comparable performance to CytoSig and outperformed it for the highest-confidence receptors, demonstrating its applicability in capturing real transcriptional responses also within complex immune environments.

RIDDEN also showed potential for finding associations between receptor activity inferred from patients' pretreatment transcriptomic profiles and their therapeutic outcomes. In patients with renal cell carcinoma undergoing treatment with a PD-1 inhibitor, RIDDEN identified activated receptors, including PD-1, that were associated with therapeutic outcomes, where this relationship was not captured by the gene expression of the receptor or its ligand.

7. SUMMARY

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 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 activated 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.

8. REFERENCES

1. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 2013 Jan;41(Database issue):D991–5.
2. Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, et al. ArrayExpress--a public database of microarray experiments and gene expression profiles. Nucleic Acids Res. 2007 Jan;35(Database issue):D747–50.
3. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell. 2017 Nov 30;171(6):1437–52.e17.
4. Han HW, Hahn S, Jeong HY, Jee JH, Nam MO, Kim HK, et al. LINCS L1000 dataset-based repositioning of CGP-60474 as a highly potent anti-endotoxemic agent. Sci Rep. 2018 Oct 8;8(1):14969.
5. Klein J, Caubet C, Camus M, Makridakis M, Denis C, Gilet M, et al. Connectivity mapping of glomerular proteins identifies dimethylaminoparthenolide as a new inhibitor of diabetic kidney disease. Sci Rep. 2020 Sep 10;10(1):14898.
6. Ganci F, Pulito C, Valsoni S, Sacconi A, Turco C, Vahabi M, et al. PI3K inhibitors curtail MYC-dependent mutant p53 gain-of-function in head and neck squamous cell carcinoma. Clin Cancer Res. 2020 Jun 15;26(12):2956–71.
7. Jiang P, Zhang Y, Ru B, Yang Y, Vu T, Paul R, et al. Systematic investigation of cytokine signaling activity at the tissue and single-cell levels. Nat Methods. 2021 Oct;18(10):1181–91.
8. Cui A, Huang T, Li S, Ma A, Pérez JL, Sander C, et al. Dictionary of immune responses to cytokines at single-cell resolution. Nature. 2024 Jan;625(7994):377–84.
9. Yao D, Binan L, Bezney J, Simonton B, Freedman J, Frangieh CJ, et al. Scalable genetic screening for regulatory circuits using compressed Perturb-seq. Nat Biotechnol. 2024 Aug;42(8):1282–95.
10. McFarland JM, Paolella BR, Warren A, Geiger-Schuller K, Shibue T, Rothberg M, et

al. Multiplexed single-cell transcriptional response profiling to define cancer vulnerabilities and therapeutic mechanism of action. Nat Commun. 2020 Aug 27;11(1):4296.

11. Szalai B, Veres DV. Application of perturbation gene expression profiles in drug discovery—From mechanism of action to quantitative modelling. Front Syst Biol. 2023 Feb 9;3:1126044.
12. Sertkaya A, Beleche T, Jessup A, Sommers BD. Costs of drug development and research and development intensity in the US, 2000-2018. JAMA Netw Open. 2024 Jun 3;7(6):e2415445.
13. Singh N, Vayer P, Tanwar S, Poyet JL, Tsaioun K, Villoutreix BO. Drug discovery and development: introduction to the general public and patient groups. Front Drug Discov (Lausanne). 2023 May 24;3:1201419.
14. Tobinick EL. The value of drug repositioning in the current pharmaceutical market. Drug News Perspect. 2009 Mar 1;22(2):119–25.
15. Hua Y, Dai X, Xu Y, Xing G, Liu H, Lu T, et al. Drug repositioning: Progress and challenges in drug discovery for various diseases. Eur J Med Chem. 2022 Apr 15;234(114239):114239.
16. Sirota M, Dudley JT, Kim J, Chiang AP, Morgan AA, Sweet-Cordero A, et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data. Sci Transl Med. 2011 Aug 17;3(96):96ra77.
17. Koudijs KKM, Böhringer S, Guchelaar HJ. Validation of transcriptome signature reversion for drug repurposing in oncology. Brief Bioinform. 2023 Jan 19;24(1):bbac490.
18. Malcomson B, Wilson H, Veglia E, Thillaiyampalam G, Barsden R, Donegan S, et al. Connectivity mapping (ssCMap) to predict A20-inducing drugs and their antiinflammatory action in cystic fibrosis. Proc Natl Acad Sci U S A. 2016 Jun 28;113(26):E3725–34.
19. Shukla R, Henkel ND, Alganem K, Hamoud AR, Reigle J, Alnafisah RS, et al. Signature-based approaches for informed drug repurposing: targeting CNS disorders. Neuropsychopharmacology. 2021 Jan;46(1):116–30.

20. Napolitano F, Gambardella G, Carrella D, Gao X, di Bernardo D. Computational drug repositioning and elucidation of mechanism of action of compounds against SARS-CoV-2 [Internet]. arXiv [q-bio.GN]. 2020 [cited 2024 Dec 11]. Available from: http://arxiv.org/abs/2004.07697
21. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov. 2020 Mar 16;6(1):14.
22. Hoagland DA, Clarke DJB, Møller R, Han Y, Yang L, Wojciechowicz ML, et al. Modulating the transcriptional landscape of SARS-CoV-2 as an effective method for developing antiviral compounds [Internet]. bioRxiv. bioRxiv; 2020 [cited 2024 Dec 11]. p. 2020.07.12.199687. Available from: https://www.biorxiv.org/content/10.1101/2020.07.12.199687v1.abstract
23. Barsi S, Papp H, Valdeolivas A, Tóth DJ, Kuczmog A, Madai M, et al. Computational drug repurposing against SARS-CoV-2 reveals plasma membrane cholesterol depletion as key factor of antiviral drug activity. PLoS Comput Biol. 2022 Apr;18(4):e1010021.
24. Kovács K, Szalai L, Szabó P, Gém JB, Barsi S, Szalai B, et al. An unexpected enzyme in vascular smooth muscle cells: Angiotensin II upregulates cholesterol-25-hydroxylase gene expression. Int J Mol Sci. 2023 Feb 1;24(4):3968.
25. Prabahar A, Zamora R, Barclay D, Yin J, Ramamoorthy M, Bagheri A, et al. Unraveling the complex relationship between mRNA and protein abundances: a machine learning-based approach for imputing protein levels from RNA-seq data. NAR Genom Bioinform. 2024 Mar 10;6(1):lqae019.
26. Hargrove JL, Hulsey MG, Beale EG. The kinetics of mammalian gene expression. Bioessays. 1991 Dec 1;13(12):667–74.
27. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015 Dec 23;1(6):417–25.
28. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000 May;25(1):25–9.
29. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al.

Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545–50.

30. Schubert M, Klinger B, Klünemann M, Sieber A, Uhlitz F, Sauer S, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. Nat Commun. 2018 Jan 2;9(1):20.
31. Holland CH, Tanevski J, Perales-Patón J, Gleixner J, Kumar MP, Mereu E, et al. Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. Genome Biol. 2020 Feb 12;21(1):36.
32. Badia-I-Mompel P, Vélez Santiago J, Braunger J, Geiss C, Dimitrov D, Müller-Dott S, et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. Bioinform Adv. 2022 Mar 8;2(1):vbac016.
33. Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH, et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. Nat Genet. 2016 Aug 20;48(8):838–47.
34. Teschendorff AE, Wang N. Improved detection of tumor suppressor events in single-cell RNA-Seq data. NPJ Genom Med. 2020 Oct 7;5(1):43.
35. Dugourd A, Saez-Rodriguez J. Footprint-based functional analysis of multiomic data. Curr Opin Syst Biol. 2019 Jun 1;15:82–90.
36. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020 May 1;20(5):533–4.
37. Ng YL, Salim CK, Chu JJH. Drug repurposing for COVID-19: Approaches, challenges and promising candidates. Pharmacol Ther. 2021 Dec 1;228(107930):107930.
38. Wyler E, Mösbauer K, Franke V, Diag A, Gottula LT, Arsiè R, et al. Transcriptomic profiling of SARS-CoV-2 infected human cell lines identifies HSP90 as target for COVID-19 therapy. iScience. 2021 Mar 19;24(3):102151.
39. Blanco-Melo D, Nilsson-Payant BE, Liu WC, Uhl S, Hoagland D, Møller R, et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. Cell. 2020 May 28;181(5):1036–45.e9.
40. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature. 2020 Jul

30;583(7816):459–68.

41. Zdrazil B, Felix E, Hunter F, Manners EJ, Blackshaw J, Corbett S, et al. The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. Nucleic Acids Res. 2024 Jan 5;52(D1):D1180–92.
42. Su J, Song Y, Zhu Z, Huang X, Fan J, Qiao J, et al. Cell-cell communication: new insights and clinical implications. Signal Transduct Target Ther. 2024 Aug 7;9(1):196.
43. Katze MG, Fornek JL, Palermo RE, Walters KA, Korth MJ. Innate immune modulation by RNA viruses: emerging insights from functional genomics. Nat Rev Immunol. 2008 Aug;8(8):644–54.
44. Altfeld M, Gale M Jr. Innate immunity against HIV-1 infection. Nat Immunol. 2015 Jun 19;16(6):554–62.
45. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, et al. Inference and analysis of cell-cell communication using CellChat. Nat Commun. 2021 Feb 17;12(1):1088.
46. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nat Protoc. 2020 Feb 26;15(4):1484–506.
47. Dimitrov D, Türei D, Garrido-Rodriguez M, Burmedi PL, Nagai JS, Boys C, et al. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. Nat Commun. 2022 Jun 9;13(1):3224.
48. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. Nat Methods. 2020 Feb;17(2):159–62.
49. Li C, Bankhead A 3rd, Eisfeld AJ, Hatta Y, Jeng S, Chang JH, et al. Host regulatory network response to infection with highly pathogenic H5N1 avian influenza virus. J Virol. 2011 Nov;85(21):10955–67.
50. Sims AC, Tilton SC, Menachery VD, Gralinski LE, Schäfer A, Matzke MM, et al. Release of severe acute respiratory syndrome coronavirus nuclear import block enhances host transcription in human lung cells. J Virol. 2013 Apr 1;87(7):3885–902.
51. Selinger C, Tisoncik-Go J, Menachery VD, Agnihothram S, Law GL, Chang J, et al. Cytokine systems approach demonstrates differences in innate and pro-inflammatory

host responses between genetically distinct MERS-CoV isolates. BMC Genomics. 2014 Dec 22;15(1):1161.

52. Josset L, Menachery VD, Gralinski LE, Agnihothram S, Sova P, Carter VS, et al. Cell host response to infection with novel human Coronavirus EMC predicts potential antivirals and important differences with SARS Coronavirus. MBio [Internet]. 2013 Jul 30 [cited 2024 Dec 13];4(3). Available from: https://journals.asm.org/doi/10.1128/mbio.00165-13
53. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47.
54. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014 Dec 5;15(12):550.
55. Türei D, Korcsmáros T, Saez-Rodriguez J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nat Methods. 2016 Nov 29;13(12):966–7.
56. Türei D, Valdeolivas A, Gul L, Palacio-Escat N, Klein M, Ivanova O, et al. Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. Mol Syst Biol. 2021 Mar;17(3):e9923.
57. Braun DA, Hou Y, Bakouny Z, Ficial M, Sant' Angelo M, Forman J, et al. Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma. Nat Med. 2020 Jun;26(6):909–18.
58. Zhang Y, Narayanan SP, Mannan R, Raskind G, Wang X, Vats P, et al. Single-cell analyses of renal cell cancers reveal insights into tumor microenvironment, cell of origin, and therapy response. Proc Natl Acad Sci U S A [Internet]. 2021 Jun 15;118(24). Available from: http://dx.doi.org/10.1073/pnas.2103240118
59. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018 Feb 6;19(1):15.
60. Wolock SL, Lopez R, Klein AM. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst. 2019 Apr 24;8(4):281–91.e9.
61. Polański K, Young MD, Miao Z, Meyer KB, Teichmann SA, Park JE. BBKNN: fast

batch alignment of single cell transcriptomes. Bioinformatics. 2020 Feb 1;36(3):964–5.

62. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. 2019 Aug;29(8):1363–75.
63. Stringer C, Wang T, Michaelos M, Pachitariu M. Cellpose: a generalist algorithm for cellular segmentation. Nat Methods. 2021 Jan;18(1):100–6.
64. Barsi S, Varga E, Dimitrov D, Saez-Rodriguez J, Hunyady L, Szalai B. RIDDEN: Data-driven inference of receptor activity from transcriptomic data [Internet]. bioRxiv. 2024 [cited 2025 Jun 10]. p. 2024.12.03.626558. Available from: https://www.biorxiv.org/content/10.1101/2024.12.03.626558v1.abstract
65. Dimitrov D, Schäfer PSL, Farr E, Rodriguez-Mier P, Lobentanzer S, Badia-I-Mompel P, et al. LIANA+ provides an all-in-one framework for cell-cell communication inference. Nat Cell Biol. 2024 Sep 2;26(9):1613–22.
66. Davidson-Pilon C. lifelines: survival analysis in Python. J Open Source Softw. 2019 Aug 4;4(40):1317.
67. Seth RB, Sun L, Chen ZJ. Antiviral innate immunity pathways. Cell Res. 2006 Feb 13;16(2):141–7.
68. Ramana CV, Grammatikakis N, Chernov M, Nguyen H, Goh KC, Williams BR, et al. Regulation of c-myc expression by IFN-gamma through Stat1-dependent and -independent pathways. EMBO J. 2000 Jan 17;19(2):263–72.
69. York AG, Williams KJ, Argus JP, Zhou QD, Brar G, Vergnes L, et al. Limiting cholesterol biosynthetic flux spontaneously engages type I IFN signaling. Cell. 2015 Dec 17;163(7):1716–29.
70. Liu A, Trairatphisan P, Gjerga E, Didangelos A, Barratt J, Saez-Rodriguez J. From expression footprints to causal pathways: contextualizing large signaling networks with CARNIVAL. NPJ Syst Biol Appl. 2019 Nov 11;5(1):40.
71. Rehwinkel J, Gack MU. RIG-I-like receptors: their regulation and roles in RNA sensing. Nat Rev Immunol. 2020 Sep 13;20(9):537–51.
72. Bouhaddou M, Memon D, Meyer B, White KM, Rezelj VV, Correa Marrero M, et al. The global phosphorylation landscape of SARS-CoV-2 infection. Cell. 2020 Aug

6;182(3):685–712.e19.

73. Kindrachuk J, Ork B, Hart BJ, Mazur S, Holbrook MR, Frieman MB, et al. Antiviral potential of ERK/MAPK and PI3K/AKT/mTOR signaling modulation for Middle East respiratory syndrome coronavirus infection as identified by temporal kinome analysis. Antimicrob Agents Chemother. 2015 Feb;59(2):1088–99.
74. Tian J, Zhang X, Wu H, Liu C, Li Z, Hu X, et al. Blocking the PI3K/AKT pathway enhances mammalian reovirus replication by repressing IFN-stimulated genes. Front Microbiol. 2015 Sep 2;6:886.
75. Ehrhardt C, Wolff T, Pleschka S, Planz O, Beermann W, Bode JG, et al. Influenza A virus NS1 protein activates the PI3K/Akt pathway to mediate antiapoptotic signaling responses. J Virol. 2007 Apr;81(7):3058–67.
76. Szalai B, Subramanian V, Holland CH, Alföldi R, Puskás LG, Saez-Rodriguez J. Signatures of cell death and proliferation in perturbation transcriptomics data-from confounding factor to effective prediction. Nucleic Acids Res. 2019 Nov 4;47(19):10010–26.
77. Horton JD, Goldstein JL, Brown MS. SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. J Clin Invest. 2002 May 1;109(9):1125–31.
78. Maekawa M, Fairn GD. Complementary probes reveal that phosphatidylserine is required for the proper transbilayer distribution of cholesterol. J Cell Sci. 2015 Apr 1;128(7):1422–33.
79. Maekawa M. Domain 4 (D4) of Perfringolysin O to visualize cholesterol in cellular membranes-the update. Sensors (Basel). 2017 Mar 3;17(3):504.
80. Ilie M, Hofman V, Dietel M, Soria JC, Hofman P. Assessment of the PD-L1 status by immunohistochemistry: challenges and perspectives for therapeutic strategies in lung cancer patients. Virchows Arch. 2016 May;468(5):511–25.
81. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006 Sep 29;313(5795):1929–35.
82. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The

Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012 Mar 28;483(7391):603–7.

83. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013 Jun 29;45(6):580–5.
84. Wu P, Feng Q, Kerchberger VE, Nelson SD, Chen Q, Li B, et al. Integrating gene expression and clinical data to identify drug repurposing candidates for hyperlipidemia and hypertension. Nat Commun. 2022 Jan 10;13(1):46.
85. Yu K, Basu A, Yau C, Wolf DM, Goodarzi H, Bandyopadhyay S, et al. Computational drug repositioning for the identification of new agents to sensitize drug-resistant breast tumors across treatments and receptor subtypes. Front Oncol. 2023 Jun 13;13:1192208.
86. Yang C, Zhang H, Chen M, Wang S, Qian R, Zhang L, et al. A survey of optimal strategy for signature-based drug repositioning and an application to liver cancer. Elife. 2022 Feb 22;11:e71880.
87. Zhao W, Li J, Chen MJM, Luo Y, Ju Z, Nesser NK, et al. Large-scale characterization of drug responses of clinically relevant proteins in cancer cell lines. Cancer Cell. 2020 Dec 14;38(6):829–43.e4.
88. Haghighi M, Caicedo JC, Cimini BA, Carpenter AE, Singh S. High-dimensional gene expression and morphology profiles of cells across 28,000 genetic and chemical perturbations. Nat Methods. 2022 Dec 7;19(12):1550–7.
89. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27–30.
90. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. Nucleic Acids Res. 2020 Jan 8;48(D1):D498–503.
91. Gene Ontology Consortium, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The Gene Ontology knowledgebase in 2023. Genetics. 2023 May 4;224(1):iyad031.
92. Armingol E, Officer A, Harismendy O, Lewis NE. Deciphering cell-cell interactions and communication from gene expression. Nat Rev Genet. 2021 Feb;22(2):71–88.
93. Szalai B, Saez-Rodriguez J. Why do pathway methods work better than they should? FEBS Lett. 2020 Dec;594(24):4189–200.

94. Douglass EF Jr, Allaway RJ, Szalai B, Wang W, Tian T, Fernández-Torras A, et al. A community challenge for a pancancer drug mechanism of action inference from perturbational profile data. Cell Rep Med. 2022 Jan 18;3(1):100492.
95. Chen B, Ma L, Paik H, Sirota M, Wei W, Chua MS, et al. Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. Nat Commun. 2017 Jul 12;8(1):16022.
96. Stathias V, Jermakowicz AM, Maloof ME, Forlin M, Walters W, Suter RK, et al. Drug and disease signature integration identifies synergistic combinations in glioblastoma. Nat Commun. 2018 Dec 14;9(1):5315.
97. Kunkel SD, Suneja M, Ebert SM, Bongers KS, Fox DK, Malmberg SE, et al. mRNA expression signatures of human skeletal muscle atrophy identify a natural compound that increases muscle mass. Cell Metab. 2011 Jun 8;13(6):627–38.
98. Wang S, Li W, Hui H, Tiwari SK, Zhang Q, Croker BA, et al. Cholesterol 25-Hydroxylase inhibits SARS-CoV-2 and other coronaviruses by depleting membrane cholesterol. EMBO J. 2020 Nov 2;39(21):e106057.
99. Sanders DW, Jumper CC, Ackerman PJ, Bracha D, Donlic A, Kim H, et al. SARS-CoV-2 requires cholesterol for viral entry and pathological syncytia formation. Elife [Internet]. 2021 Apr 23 [cited 2025 May 14];10. Available from: http://dx.doi.org/10.7554/eLife.65962
100. O'Neill LAJ. How Low Cholesterol Is Good for Anti-viral Immunity. Cell. 2015 Dec 17;163(7):1572–4.
101. Blanc M, Hsieh WY, Robertson KA, Watterson S, Shui G, Lacaze P, et al. Host defense against viral infection involves interferon mediated down-regulation of sterol biosynthesis. PLoS Biol. 2011 Mar;9(3):e1000598.
102. Yuan S, Chu H, Chan JFW, Ye ZW, Wen L, Yan B, et al. SREBP-dependent lipidomic reprogramming as a broad-spectrum antiviral target. Nat Commun. 2019 Jan 10;10(1):120.
103. Wang R, Simoneau CR, Kulsuptrakul J, Bouhaddou M, Travisano KA, Hayashi JM, et al. Genetic screens identify host factors for SARS-CoV-2 and common cold coronaviruses. Cell. 2021 Jan 7;184(1):106–19.e14.

104. Daniloski Z, Jordan TX, Wessels HH, Hoagland DA, Kasela S, Legut M, et al. Identification of required host factors for SARS-CoV-2 infection in human cells. Cell. 2021 Jan 7;184(1):92–105.e16.
105. Jin S, Ramos R. Computational exploration of cellular communication in skin from emerging single-cell and spatial transcriptomic data. Biochem Soc Trans. 2022 Feb 28;50(1):297–308.
106. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. Science. 2018 Mar 23;359(6382):1350–5.
107. Burger JA, Kipps TJ. CXCR4: a key receptor in the crosstalk between tumor cells and their microenvironment. Blood. 2006 Mar 1;107(5):1761–7.
108. Qin R, Ren W, Ya G, Wang B, He J, Ren S, et al. Role of chemokines in the crosstalk between tumor and tumor-associated macrophages. Clin Exp Med. 2023 Sep;23(5):1359–73.
109. Karmakar S, Lal G. Role of serotonin receptor signaling in cancer cells and anti-tumor immunity. Theranostics. 2021 Mar 11;11(11):5296–312.
110. de las Casas-Engel M, Domínguez-Soto A, Sierra-Filardi E, Bragado R, Nieto C, Puig-Kroger A, et al. Serotonin skews human macrophage polarization through HTR2B and HTR7. J Immunol. 2013 Mar 1;190(5):2301–10.
111. Yi M, Li T, Niu M, Wu Y, Zhao Z, Wu K. TGF-β: A novel predictor and target for anti-PD-1/PD-L1 therapy. Front Immunol. 2022 Dec 19;13:1061394.
112. Chen S, Crabill GA, Pritchard TS, McMiller TL, Wei P, Pardoll DM, et al. Mechanisms regulating PD-L1 expression on tumor and immune cells. J Immunother Cancer. 2019 Nov 15;7(1):305.
113. Arasanz H, Gato-Cañas M, Zuazo M, Ibañez-Vea M, Breckpot K, Kochan G, et al. PD1 signal transduction pathways in T cells. Oncotarget. 2017 Aug 1;8(31):51936–45.
114. Gordon SR, Maute RL, Dulken BW, Hutter G, George BM, McCracken MN, et al. PD-1 expression by tumour-associated macrophages inhibits phagocytosis and tumour immunity. Nature. 2017 May 25;545(7655):495–9.

9. BIBLIOGRAPHY OF THE CANDIDATE'S PUBLICATIONS

I. List of original publications within the topic of the PhD thesis:

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

Barsi S, Papp H, Valdeolivas A, Tóth DJ, Kuczmog A, Madai M, Hunyady L, Várnai P, Saez-Rodriguez J, Jakab F, Szalai B
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

II. List of original publications not relating to the topic of the PhD thesis:

Damouni A, Tóth DJ, **Barsi S**, Nagy DK, Kasbary A, Hunyady L, Cserző M, Várnai P
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

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

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, Balla András
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

Széky B, Mayer B, Gyöngy M, Hajdara A, **Barsi S**, Kárpáti S, Németh K
Tri-Lineage Differentiation of NTERA2 Clone D1 Cells towards Neural, Hepatic and Osteogenic Lineages in Vitro
FOLIA BIOLOGICA 67: 5-6 pp. 174-182. (2021)
IF: 1,167

Barsi S, Szalai B
Modeling in systems biology: Causal understanding before prediction?
PATTERNS 2: 6 Paper: 100280, 3 p. (2021)

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 László G, Mahmoud Al-Khrasan
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

Zádor F, Mohammadzadeh A, Balogh M, Zádori ZS, Király K, **Barsi S**, Galambos AR, László SB, Hutka B, Váradi A, Hosztafi S, Riba P, Benyhe S, Fürst S, Al-Khrasani M
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

Al-Khrasani Mahmoud, Mohammadzadeh Amir, Balogh Mihály, Király Kornél, **Barsi Szilvia**, Hajnal Benjámin, 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. (2019)
IF: 3,370

10. ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my supervisor, **Dr. Bence Szalai**, for his continuous support, guidance, and patience throughout my doctoral studies. His guidance has played an important role in developing my scientific thinking and research skills.

I am also truly grateful to **Prof. Dr. László Hunyady** for welcoming me to the Department of Physiology and supervising my doctoral studies. His scientific insights, valuable comments, and support have significantly contributed to the progress of my work. I also greatly appreciated his efforts in fostering collaborations and connecting us with other research teams.

I would like to acknowledge the members of the **Hunyady-Várnai Laboratory** for their valuable comments and support throughout my research. Special thanks to **Prof. Dr. Péter Várnai** and **Dr. Dániel Tóth** for their essential work supporting my first-author paper.

I am grateful to my colleagues in the **Department of Physiology** for their help in both the educational tasks I undertook and the scientific work. Their support and cooperation played an important role during my PhD.

Special thanks to **Prof. Julio Saez-Rodriguez** and **Dr. Daniel Dimitrov** for supervising and mentoring me during my four-month DAAD-funded research exchange in Germany. Their scientific guidance and support were crucial for the success of that experience and contibuted significantly to the publication of our paper.

Lastly, I would like to thank **my family** for their unconditional support through both the good and difficult times. Their encouragement and belief in me kept me going throughout this challenging journey.

Supplementary table 1: Table summarizing datasets used in this study.

Description of the data	Source	Number of samples	Data Type	Processing Steps
Perturbation signatures				
LINCS-L1000 perturbation signatures	CLUE Data Library, Expanded CMap LINCS Resource 2020 (3)	Drug repurposing: RIDDEN: 38989 consensus signatures (cell line-perturbation), 14463 considered high-quality	Level 5 gene expression (L1000)	Consensus signatures via moderated-z method, 5 modalities (shRNA, CRISPR, drug, ligand, overexpression)
Curated drug screens				
ChEMBL SARS-CoV-2 drugs	ChEMBL SARS-CoV-2 release (40)	47 drugs overlapping with LINCS	Drug activity data (list of effective antiviral drugs)	Effective antiviral drugs aligned with LINCS via drug names and SMILES
Virus infection signatures				
SARS-CoV-2 in Calu-3 A549 RSV in S3* A549, S8* A549 HPIV in S8* A549	GEO accession: GSE147507	2-3 replicates of 24 hpi infection and mock samples	RNA-Seq	Differential expression analysis infected vs mock: DESeq2
SARS-CoV-1 in Calu-3, SARS-CoV-2 in Calu-3	GEO accession: GSE148729	2 replicates of 24 hpi infection and mock samples	RNA-Seq	Differential expression analysis infected vs mock: DESeq2
H5N1 in Calu-3	GEO accession: GSE28166	3 replicates of 24 hpi infection and mock samples	microarray	Differential expression analysis infected vs mock: limma
Influenza in Calu-3	GEO accession: GSE37571	3 replicates of 24 hpi infection and mock samples	microarray	Differential expression analysis infected vs mock: limma
SARS-CoV-1 in Calu-3	GEO accession: GSE33267	3 replicates of 24 hpi infection and mock samples	microarray	Differential expression analysis infected vs mock: limma
MERS-CoV in Calu-3	GEO accession: GSE56677	3 replicates of 24 hpi infection and mock samples	microarray	Differential expression analysis infected vs mock: limma
MERS-CoV in Calu-3	GEO accession: GSE45042	2 replicates of 24 hpi infection and mock samples	microarray	Differential expression analysis infected vs mock: limma
Cytokine stimulatory signatures				
CytoSig cytokine signatures	(7), data from CytoSig publication (data from public databases)	2056 bulk samples, 112 cytokines	Bulk transcriptomics (cytokine stimulation)	Gene-wise z-scores, overlap with LINCS cytokines/receptors via OmniPath LRI
ImmuneDictionary	(8), data from ImmuneDictionary paper	17 immune cell types, 86 cytokines	Single-cell RNA-Seq (mouse lymph nodes)	Gene-wise z-scores; aggregated to average signatures per cell type; excluded low-DEG signatures; overlap with LINCS via OmniPath LRI
Transcriptomic profiles of the patients				
Patient transcriptional profiles (nivolumab/everolimus)	Data from supplementary of publication (57)	Pretreatment samples; OS data, 311 patients, 130 treated with Everolimus, 181 with Nivolumab	Bulk RNA-Seq	
Clear cell renal cell carcinoma scRNA-seq	Data from publication (58)	7 patients	Single-cell RNA-Seq	BBKNN batch correction, marker-based annotation
Abbreviations: DEG: differentially expressed genes, hpi: hours post-infection, LRI: ligand-receptor interactions, OS: overall survival, S: series				

Supplementary Table 2: Table summarizing the methods used in this study.

Method/Tool	Package	Purpose	Key Details	Versions
Differential expression analysis				
Limma	Limma R	Differential expression (microarray)	GEO virus data processing	3.50.0
DESeq2 (R package)	DESeq2 R	Differential expression (RNA-Seq)	GEO virus data processing	1.34.0
Functional genomic analysis				
PROGENy	Progeny R	Pathway activity inference	Z-scores via 1000 gene permutations	1.16.0
DoRothEA	Dorothea R	Transcription factor activity inference	High-confidence regulons (A-C); VIPER algorithm	Dorothea:1.6.0, viper: 1.28.0
Statistics, machine learning and model building				
Spearman's correlation	Scipy Python	Signature similarity and LINCS MODZ	Between virus/drug signatures on shared genes	SARS-CoV-2: 1.4.1 RIDDEN: 1.8.0
Random Forest	Scikit-learn Python	Predict antiviral drugs; feature importance	300 trees; 100 random subsamples (50 effective/50 non-effective); ROC AUC evaluation	0.22.1
OLS regression	Statsmodels Python	RIDDEN model construction	Receptor-gene matrix from perturbation data; ligand-receptor mapping via OmniPath	0.13.2
Survival analysis	Lifelines Python	Cox regression and log-rank test	Receptor/gene associations with overall survival	0.27.7
OmniPath	OmniPath Python	Ligand-receptor networks	Curated interactions for signature selection	3.2.0
Image segmentation				
CellPose	CellPose Python	Image segmentation	Confocal microscopy (cholesterol sensor); diameter >200px; dilation/erosion for boundaries	0.6.1

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Computational drug repurposing against SARS-CoV-2 reveals plasma membrane cholesterol depletion as key factor of antiviral drug activity

Szilvia Barsi[1], Henrietta Papp[2,3], Alberto Valdeolivas[4], Dániel J. Tóth[1], Anett Kuczmog[2,3], Mónika Madai[2,3], László Hunyady[1,5,6], Péter Várnai[1,5], Julio Saez-Rodriguez[4], Ferenc Jakab[2,3], Bence Szalai[1¤]*

1 Semmelweis University, Faculty of Medicine, Department of Physiology, Budapest, Hungary, **2** National Laboratory of Virology, University of Pécs, Pécs, Hungary, **3** Institute of Biology, Faculty of Sciences, University of Pécs, Pécs, Hungary, **4** Heidelberg University, Faculty of Medicine, and Heidelberg University Hospital, Institute for Computational Biomedicine, Bioquant, Heidelberg, Germany, **5** MTA-SE Laboratory of Molecular Physiology, Budapest, Hungary, **6** Institute of Enzymology, Research Centre for Natural Sciences, Budapest, Hungary

¤ Current address: Turbine Simulated Cell Technologies Ltd., Budapest, Hungary
* ben.szalai@gmail.com

Citation: Barsi S, Papp H, Valdeolivas A, Tóth DJ, Kuczmog A, Madai M, et al. (2022) Computational drug repurposing against SARS-CoV-2 reveals plasma membrane cholesterol depletion as key factor of antiviral drug activity. PLoS Comput Biol 18(4): e1010021. https://doi.org/10.1371/journal.pcbi.1010021

Editor: James Costello, University of Colorado Denver, UNITED STATES

Received: November 19, 2021

Accepted: March 15, 2022

Published: April 11, 2022

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pcbi.1010021

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files. All analysis code to reproduce the

Abstract

Comparing SARS-CoV-2 infection-induced gene expression signatures to drug treatment-induced gene expression signatures is a promising bioinformatic tool to repurpose existing drugs against SARS-CoV-2. The general hypothesis of signature-based drug repurposing is that drugs with inverse similarity to a disease signature can reverse disease phenotype and thus be effective against it. However, in the case of viral infection diseases, like SARS-CoV-2, infected cells also activate adaptive, antiviral pathways, so that the relationship between effective drug and disease signature can be more ambiguous. To address this question, we analysed gene expression data from *in vitro* SARS-CoV-2 infected cell lines, and gene expression signatures of drugs showing anti-SARS-CoV-2 activity. Our extensive functional genomic analysis showed that both infection and treatment with *in vitro* effective drugs leads to activation of antiviral pathways like NFkB and JAK-STAT. Based on the similarity—and not inverse similarity—between drug and infection-induced gene expression signatures, we were able to predict the *in vitro* antiviral activity of drugs. We also identified SREBF1/2, key regulators of lipid metabolising enzymes, as the most activated transcription factors by several *in vitro* effective antiviral drugs. Using a fluorescently labeled cholesterol sensor, we showed that these drugs decrease the cholesterol levels of plasma-membrane. Supplementing drug-treated cells with cholesterol reversed the *in vitro* antiviral effect, suggesting the depleting plasma-membrane cholesterol plays a key role in virus inhibitory mechanism. Our results can help to more effectively repurpose approved drugs against SARS-CoV-2, and also highlights key mechanisms behind their antiviral effect.

results of this manuscript is available at https://github.com/comp-sys-pharm/SARS-CoV-2-cholesterol.

Funding: BS was supported by the Premium Postdoctoral Fellowship Program of the Hungarian Academy of Sciences (460044). DJT and PV were supported by the Hungarian Scientific Research Fund (OTKA K134357). On behalf of Project DRUGSENSPRED we thank for the usage of ELKH Cloud (https://science-cloud.hu/) that significantly helped us achieve the results published in this paper. The in vitro SARS-CoV-2 experiments were funded by the Hungarian Scientific Research Fund (OTKA KH129599), by the European Union and the European Social Fund (EFOP-3.6.1.-16-2016-00004), and by the Ministry for Innovation and Technology of Hungary (TUDFO/47138/2019-ITM) to FJ. Also, project no. TKP2021-NVA-07 has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the TKP2021-NVA funding scheme to FJ. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: JSR reports funding from GSK and Sanofi and fees from Travere Therapeutics and Astex. BS is a full time employee of Turbine Ltd., Budapest, Hungary.

Author summary

Targeting the infected host cells is an effective strategy in infectious diseases, like COVID-19. Better understanding the virus and drug induced cellular mechanisms can help to identify new compounds with potential antiviral activity. We used computational methods to analyse gene expression data from *in vitro* SARS-CoV-2 infected cell lines, and gene expression signatures of drugs showing anti-SARS-CoV-2 activity. With the help of machine learning methods, we were able to predict *in vitro* effective antiviral drugs from gene expression based features. We found that effective drugs activate antiviral pathways like JAK-STAT and NFkB, and also the SREBF transcription factors, key regulators of cholesterol synthesis. Using microscopic measurements we validated that several antiviral drugs influence the cholesterol content of the plasma membrane. Finally, we showed that cholesterol rescue inhibited the *in vitro* antiviral effect of amiodarone, demonstrating the importance of drug induced cholesterol changes in the antiviral drug effect.

1. Introduction

The newly emerged Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), causing the coronavirus disease 2019 (COVID-19), has led to more than 420,000,000 reported infections and 5,500,000 reported deaths worldwide [1] until February 2022. Identification of new therapeutic compounds against SARS-CoV-2 / COVID-19 is an urgent need until effective vaccination is worldwide available and given the emergence of SARS-CoV-2 strains showing immune evasion [2]. The main therapeutic strategies include A) inhibiting key viral enzymes (like remdesivir [3]); B) modulating the infected cells to decrease viral replication [4,5] and C) modulating the over-activation of the immune system to treat late complications like "cytokine storm" [6–8]. Repurposing already approved drugs for these indications is especially important as it allows a shorter time of approval for anti-SARS-CoV-2 treatment.

Comparing gene expression signatures of drugs and diseases have been previously shown to be an effective strategy to repurpose drugs for new therapeutic indications [9]. The general principle of these studies is that a drug inducing an opposite gene expression signature to a disease signature can reverse the disease-related gene expression changes, thus the disease phenotype. This "signature reversal" principle has also been used to predict effective drugs against SARS-CoV-2 infection [10–12]. However, these predictions lack, in most cases, mechanistic insight and experimental validation. Moreover, as infected cells activate adaptive antiviral pathways (like interferon pathway), inhibiting these pathways does not necessarily decrease viral replication.

In this study, we analyzed transcriptomics data from *in vitro* SARS-CoV-2 infected cell lines (section 2.1) and from cell lines treated with drugs showing anti-SARS-CoV-2 activity (effective drugs, section 2.2). Functional genomic analysis revealed shared transcription factor and pathway activity changes (eg. increased activity NFkB and JAK-STAT pathways) in the infected and effective drug-treated cell lines. Similarity between infection signature and drug-induced signature was predictive for *in vitro* effective drugs, contradictory to the classical "signature reversal" principle (section 2.3). Machine learning-based prediction of effective drugs identified SREBF1 and SREBF2 transcription factors, key regulators of lipid metabolism, as important factors of antiviral drug effect. Using a fluorescently labeled cholesterol sensor, we showed the decreased level of plasma-membrane cholesterol in cells treated with effective drugs, like chlorpromazine, confirming the effect of these drugs on cholesterol metabolism (section 2.4). We also identified amiodarone, a drug decreasing plasma-membrane cholesterol

content, thus a potential *in vitro* effective drug. Using an *in vitro* SARS-CoV-2 infection assay, we demonstrated that the antiviral effect of amiodarone can be reversed by cholesterol supplement, underlying the relevance of decreased plasma-membrane cholesterol in the antiviral drug effect (section 2.5).

2. Results

2.1 Analysis of host pathway and transcription factor activities reveals adaptive response of SARS-CoV-2 infected cells

We analysed gene expression data from two recent studies (GSE147507 [13] and GSE148729 [14]), where lung epithelial cancer cell lines (Calu-3 and A549) were infected with SARS-CoV-2. To identify infection-induced pathway and transcription factor (TF) changes, we used the PROGENy [15,16] and DoRothEA [17,18] tools, respectively (more details in Methods).

PROGENy analysis showed increased activity of NFkB and TNFa pathways in both analysed cell lines, while the activity of JAK-STAT pathway increased more pronounced in infected Calu-3 cell lines (Fig 1A). DoRothEA analysis (Fig 1B) revealed strong activation of STAT, IRF and NFkB transcription factors, while cell growth-related transcription factors (E2Fs, Myc) showed decreased activity. Also SREBF1/2, key transcriptional regulators of cholesterol synthesis, showed decreased activity. STATs, IRFs and NFkB pathways / TFs play a key role in antiviral innate immunity [19]. Decreased activity of E2Fs and Myc [20] and decreased synthesis of cholesterol [21] are also part of the physiological antiviral / interferon response.

To further analyse which upstream signalling pathways regulate the inferred TF activity changes, we used CARNIVAL [22], a signaling network contextualisation tool, which connects transcription factor activities to perturbations in signaling networks via integer linear programming (more details in Methods). We performed CARNIVAL analysis using inferred transcription factor activities from a SARS-CoV-2 infected cell line (GSE147507, Calu-3), and used RIG-I like receptors (DDX58 and IFIH1), key receptors for foreign RNA sensing [23], as main perturbation target. CARNIVAL results showed (Fig 1C), that activation of RIG-I like receptors by the dsRNA of SARS-CoV-2 can directly lead to the observed transcription factor activity changes, including activation of NFkB, IRFs and STATs and inhibition of SREBF2 and E2F4. Key identified intermediate nodes AKT1 and MAPK1 were already connected to coronavirus infection [5,24] and other viral infections [25,26], also suggesting that the observed TF changes are initiated by the RIG-I like receptors, thus corresponding to the antiviral response of the host cell.

In summary, our functional analysis of the gene expression changes in SARS-CoV-2 infected cell lines suggests that a large part of the induced pathway / transcription factor activity changes are adaptive, i.e. part of the physiological antiviral response.

2.2 Analysis of *in vitro* anti-SARS-CoV-2 drug-induced pathway and transcription factor activities reveals similar changes to virus infection

To compare infection and drug-induced signatures, we used a large compendium of drug-induced gene expression signatures from the LINCS-L1000 project [27]. LINCS-L1000 contains drug-induced gene expression signatures from different cell lines, concentrations and time points. We calculated consensus gene signatures for each drug using our previous approach ([28], Methods), ending up with gene expression signatures for 4671 drugs. To select drugs effectively inhibiting SARS-CoV-2 replication *in vitro*, we used a curated database created by ChEMBL (http://chembl.blogspot.com/2020/05/chembl27-sars-cov-2-release.html). This dataset contains 133 drugs previously showing effective inhibition of viral replication in 8

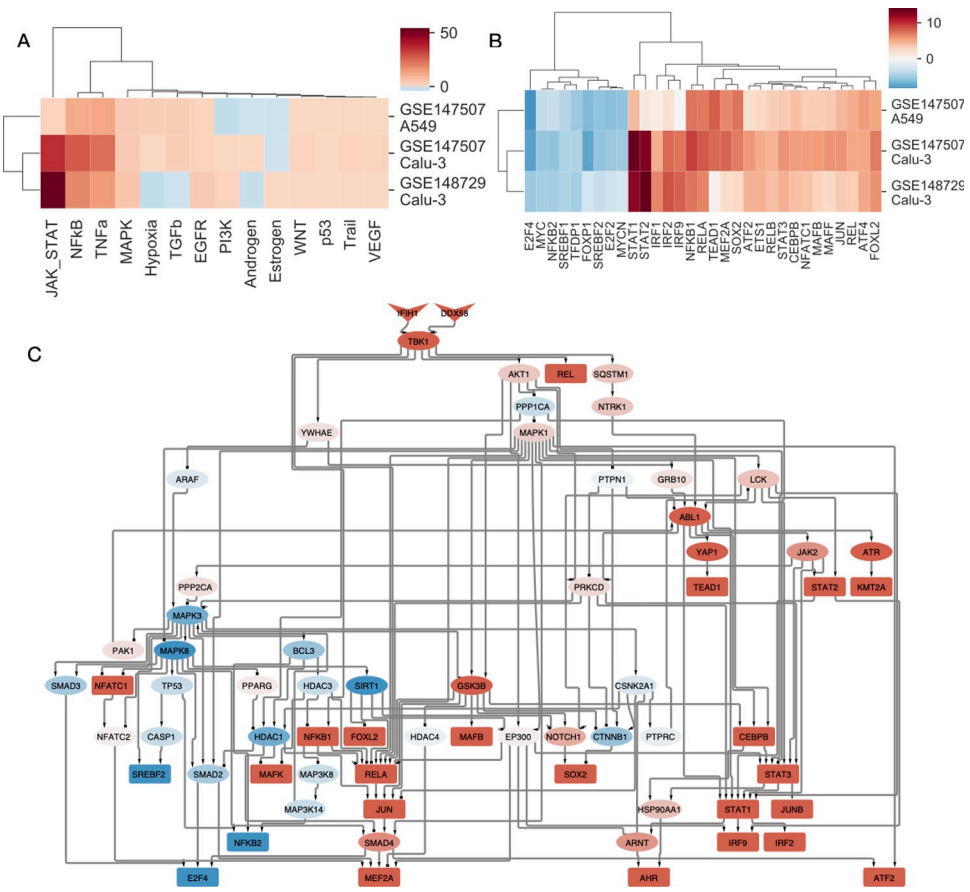

Fig 1. Functional genomic analysis of SARS-CoV-2 infected cell lines. (A) Inferred pathway and (B) TF activities of SARS-CoV-2 infected samples from lung epithelial cell lines (Calu-3 and A549). Activities were calculated from differential expression signatures (infected—control) using PROGENy and DoRothEA tools for pathway and TF activities, respectively. Only TFs with high absolute level of activity changes (absolute normalized enrichment score > 4) are shown. (C) Causal signalling network in SARS-CoV-2 infected Calu-3 cells (GSE147507) identified by CARNIVAL. RIG-I like receptors (DDX58 and IFIH1) as perturbation targets and DoRothEA inferred TF activities were used as the input of the CARNIVAL pipeline. Color code represents inferred activity of protein nodes (blue: inhibited, red: activated).

https://doi.org/10.1371/journal.pcbi.1010021.g001

studies [4,29–35]. We found an intersection of 47 drugs between LINCS-L1000 (available gene expression signatures) and ChEMBL dataset (*in vitro* effective drugs). To characterize drug-induced pathway and transcription activity changes, we analysed consensus drug-induced signatures using PROGENy and DoRothEA.

PROGENy analysis showed strong activation of NFkB and TNFa pathways by several drugs, including niclosamide, perhexiline and digoxin (Fig 2A). Several drugs also strongly activated the JAK-STAT pathway (RTK inhibitors osimertinib and regorafenib). In case of TF analysis, we found similar patterns (Fig 2B) to the infection-induced signatures: increased activity of NFkB and STAT transcription factors and decreased activity of Myc/E2Fs transcription factors. Interestingly, SREBF1/2 showed strongly increased activity for a large cluster of drugs, but (similar to the infection signatures) decreased in another cluster. To further analyse the TF activity changes in the different clusters of drugs, we calculated average TF activities for these clusters and plotted these values against the average TF activities of the 3 SARS-CoV-2 infection signatures (Fig 2C). One cluster (Fig 2C, upper left panel), showed high correlation (Spearman's rho = 0.64, p = 8.55e-35) across all TFs. Two other clusters (Fig 2C, upper middle

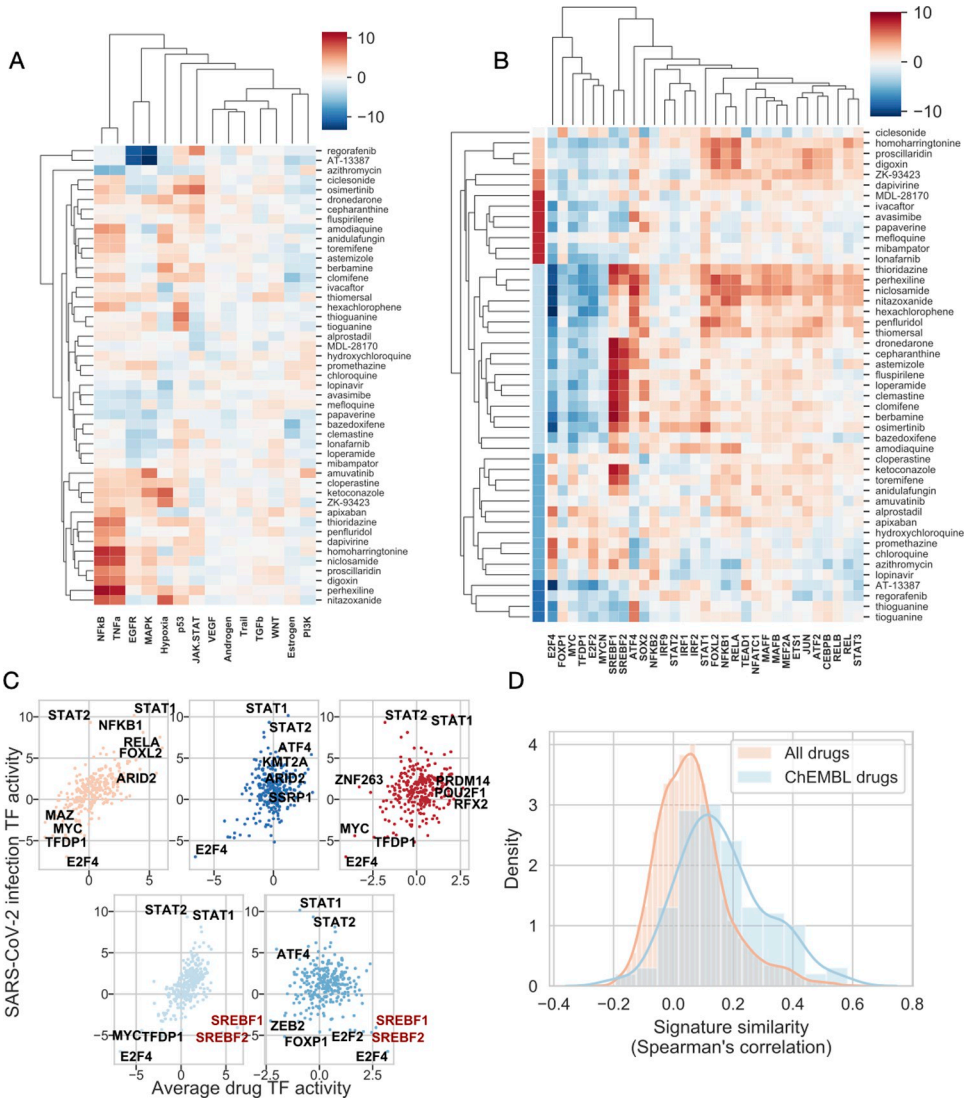

Fig 2. Functional genomic analysis of effective drugs treated cell lines. (A) Inferred pathway and (B) TF activities of anti-SARS-CoV-2 drug-treated cell lines. Activities were calculated from LINCS-L1000 consensus drug-induced signatures, using PROGENy and DoRothEA tools for pathway and TF activities, respectively. Drug clusters in (B) are color coded. Only selected transcription factors (corresponding to Fig 1B) are shown. (C) Relationship between average TF activities induced by drug treatment and SARS-CoV-2 infection for 5 different drug clusters (colors of clusters correspond to panel B). TFs with the highest/lowest average activities are text labeled. (D) Density plot of similarities between SARS-CoV-2- and drug-induced signatures for all LINCS-L1000 drugs and known anti-SARS-CoV-2 drugs (ChEMBL drugs).

https://doi.org/10.1371/journal.pcbi.1010021.g002

and upper right panels) showed lower, but still significant correlation with infection TF activity signature (Spearman's rho = 0.14 and 0.18, p = 0.0122 and 0.00174, respectively), with prominent increase of STATs and decrease of E2F4 transcription factor activity. For the remaining two large clusters, we found either negligible (Fig 2C, lower right panel) or high (Fig 2C, lower left panel) correlation with infection-induced TF activities (Spearman rho = 0.04 and 0.58, p = 0.484 and 3.14e-27, respectively), but we found high drug-induced activity of SREBF1/2 transcription factors in these clusters, opposite to the inhibition of these TFs by SARS-CoV-2 infection.

As we found that, for several drug clusters, drug-induced TF activities showed positive correlation with SARS-CoV-2-induced TF activities, we were interested in the general similarity of drug and infection-induced gene expression signatures. To achieve this we calculated the signature similarity (Spearman's correlation coefficient, which has been previously shown to be an effective metric to analyse signature similarity for the LINCS-L1000 data [27,28]) between all the 4,671 drug-induced signatures from our LINCS-L1000 dataset and the infection signatures. We found that effective anti-SARS-CoV-2 drugs (ChEMBL dataset) have higher similarity to infection signatures, than ineffective drugs / drugs with unknown efficacy (Fig 2E, Mann-Whitney U test p-value = <1e-200).

In summary, we found that known *in vitro* effective anti-SARS-CoV-2 drugs induce similar pathway and TF activity patterns, and appropriately similar gene expression signatures to virus infection signatures. We also identified two large clusters of drugs inducing strong activation of SREBF1/2 transcription factors, key regulators of cholesterol / lipid metabolism.

2.3 Prediction of drugs with *in vitro* anti-SARS-CoV-2 activity

After identifying some general patterns in the gene expression signatures of *in vitro* effective anti-SARS-CoV-2 drugs, we investigated how well we can predict drug effectiveness using gene expression signatures.

As a first strategy, we simply used the previously calculated drug—infection signature similarity to predict effective drugs. Using these similarity values (predicted score) and the known *in vitro* effective drugs (ChEMBL dataset, true positive values) we performed ROC analysis (Fig 3A). We found that similarity to infection signatures is predictive for effective drugs, i.e. drugs with high similarity to infection signature are more frequently effective (ROC AUCs: 0.75, 0.74 and 0.64 for GSE147507 A549, GSE147507 Calu-3 and GSE148729 Calu-3, respectively). To test the specificity of this signature similarity-based approach for SARS-CoV-2 infection signature, we included several other virus infection-induced gene expression signatures for SARS-CoV (GSE33267 [36], GSE148729), MERS (GSE45042 [37], GSE56677 [38]), respiratory syncytial virus (RSV, GSE147507), influenza (GSE28166 [39], GSE37571) and human parainfluenza (HPIV, GSE147507) infected Calu-3 and/or A549 cell lines. Similarity to these infection signatures showed lower predictive performance for anti-SARS-CoV-2 drugs (ROC AUC values <0.7 except one SARS and RSV signature with ROC AUCs 0.70 and 0.71, respectively, Fig 3B), suggesting the relative SARS-CoV-2 specificity of the similarity-based methods.

Following this unsupervised prediction strategy, we also performed supervised, machine learning-based predictions. We used the drug-induced TF activities as features, and effective drugs from the ChEMBL dataset as positive examples, with Random Forest Classification as prediction algorithm. We set up a random subsampling based cross-validation scheme and evaluated the performance using ROC analysis (Methods). Our results showed a slightly improved performance compared to the unsupervised, similarity-based approach (mean ROC AUCs: 0.72 and 0.68, 0.66, 0.57, respectively for the machine learning and similarity-based methods, paired t-test p-values between machine learning and similarity-based methods: 3.02e-07, 2.76e-15, 4.89e-15 for GSE147507 A549, Calu-3 and GSE148729 Calu-3 signatures respectively, Fig 3C). To gain some more mechanistic insight from the prediction of machine learning models, we analysed feature importances (Gini importance, Fig 3D) of the Random Forest Regression models and found that SREBF1 and SREBF2 activity were the two most important features, followed by TFAP2A, HNF4A and TP63 transcription factors.

In summary, our two different prediction approaches showed reasonable performance (comparable to studies based on network medicine and chemical similarity [40,41]), to predict

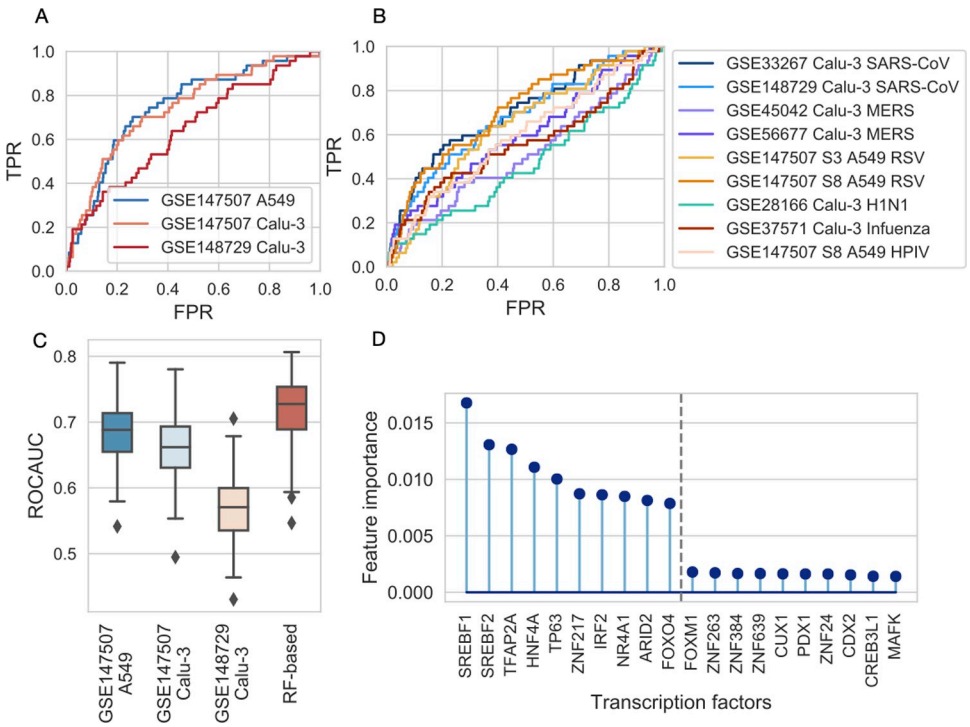

Fig 3. Evaluation of similarity-based and machine learning-based models in predicting *in vitro* effective drugs. (A, B) ROC analysis of similarity-based predictions of effective drugs against SARS-CoV-2. Drug—SARS-CoV-2 (A) or drug—other virus (B) infection signature similarity was used as prediction score, while known *in vitro* effective drugs (ChEMBL dataset) were used as true positives. (FPR: false positive rate, TPR: true positive rate) (C) Comparison of predictive performance (ROCAUCs) of similarity-based method (similarity to SARS-CoV-2 infection signature, x-axis) and random forest-based (RF-based, x-axis) prediction. Results of 100 random subsampling cross-validations. In case of similarity-based methods, ROC AUC curves were only calculated for the corresponding cross-validation sets. Boxplots represent the median (central line), first and third quartile (box), minimum and maximum non-outlier values (whiskers) and outliers (diamonds). (D) Feature importances (Gini importance) of the Random Forest model. Top and bottom 10 features (TFs) are shown according to importance.

https://doi.org/10.1371/journal.pcbi.1010021.g003

drugs with *in vitro* anti-SARS-CoV-2 activity, and also highlighted the importance of previously discussed SREBF1/2 transcription factors. Drug—SARS-CoV-2 signature similarities, and predicted probabilities of anti-SARS-CoV-2 activity is available in S1 Table.

2.4 Anti-SARS-CoV-2 drugs are increasing SREBF activity by depleting plasma membrane cholesterol

While in most cases we found similarity between the activity of SARS-CoV-2 infection and *in vitro* effective drug-induced transcription factor activities, in case of SREBF1/2 we found opposite changes: SARS-CoV-2 infection inhibited SREBF1/2, while a large cluster of effective drugs lead to increased activity of SREBFs. SREBFs are activated through the decreased cholesterol content of plasma membrane and endoplasmic reticulum, and activated SREBFs induce the expression of cholesterol, and other lipid synthesizing enzymes [42]. From this point of view, decreased SREBF activity during viral infection can lead to decreased cholesterol synthesis, which can inhibit the viral replication and/or viral entry [21], thus can be considered as an adaptive response of the host cell (Fig 4A). Interestingly, we observed a strongly increased SREBF activity in large clusters of effective drugs. To resolve this discordance, we hypothesized that these *in vitro* effective drugs directly decrease plasma membrane cholesterol (Fig 4A). In this case, drug-induced decrease of plasma membrane cholesterol can contribute to the

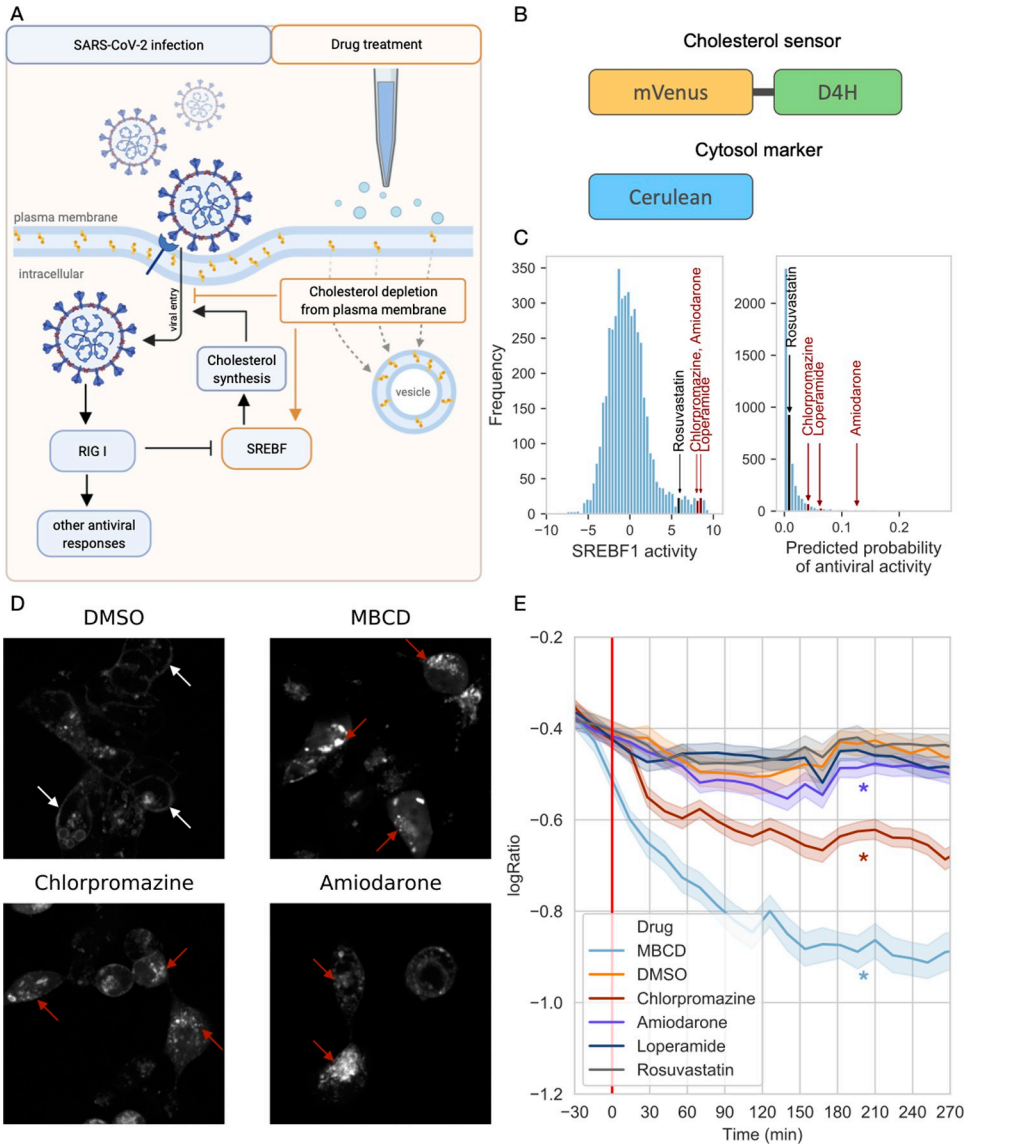

Fig 4. Cholesterol depleting effect of SREBF activating drugs. (A) Schematic figure of the hypothesis that antiviral drugs block virus entry into cells by cholesterol depletion from plasma membrane, and are leading to a compensatory increased SREBF1/2 activity. Effects induced by viral infection are marked with black arrows (left side), while orange arrows represent drug-induced changes (right side) The figure was created with BioRender.com. (B) Schematic representation of the used fluorescent constructs. (C) Histogram of SREBF1 activation (left panel) and histogram of predicted probabilities of *in vitro* antiviral activity of LINCS-L1000 drugs (right panel, according to the Random Forest model). Drugs selected for *in vitro* experiments are text labeled. (D) Representative confocal microscopy images of D4H-mVenus transfected HEK293A cells treated with DMSO, MβCD, chlorpromazine or amiodarone. White arrows mark plasma membrane, while red arrows show intracellular localised cholesterol sensors. (E) Time-dependent change of $\log_2$(PM/IC) ratio of average cholesterol sensor intensity in HEK293A cells treated with DMSO, MβCD, chlorpromazine, amiodarone, loperamide or rosuvastatin. Red line marks drug treatment. *: significant ($p<0.001$) interaction between drug treatment and elapsed time in linear model.

https://doi.org/10.1371/journal.pcbi.1010021.g004

antiviral effect, while decreased cholesterol levels can activate SREBFs, thus explaining the observed increased activity of these TFs in our bioinformatic analysis.

To confirm this hypothesis, we performed high-throughput, automatic confocal microscopy imaging using a fluorescently labeled cholesterol sensor domain, D4H-mVenus [43,44].

HEK293A cells were co-transfected with D4H-mVenus and cytoplasmic Cerulean as cytosolic marker (Fig 4B), and treated with dimethyl sulfoxide (DMSO, negative control), MβCD (methyl β-cyclodextrin, plasma membrane cholesterol depleting compound, as positive control) and 3 drugs from our computational drug repurposing pipeline, loperamide, amiodarone and chlorpromazine (all drugs were used in 10 μM final concentration). All these three drugs increased the activity of SREBF transcription factors (Fig 4C, left panel). Loperamide and chlorpromazine have been previously shown to be *in vitro* effective against SARS-CoV-2 (ChEMBL dataset), while amiodarone was one of the top predicted drugs of the Random Forest model (Fig 4C, right panel, ranked 36/4671 drugs, S1 Table). We also treated HEK293A cells with rosuvastatin, an inhibitor of cholesterol synthesis. Rosuvastatin also alters cellular cholesterol metabolism, however, it does not influence plasma membrane cholesterol directly, but inhibits HMG-CoA reductase, the rate limiting enzyme of *de novo* cholesterol synthesis. Rosuvastatin was not predicted as an effective anti-SARS-CoV-2 drug by the Random Forest model (Fig 4C left panel, ranked 1821/4671 drugs).

Cells were treated with the different drugs and serial confocal microscopy images were recorded for 4.5 hours. In untreated, or DMSO treated cells, we observed a predominantly plasma membrane localisation of the fluorescent protein labeled cholesterol sensor (Fig 4D, top left panel). Treatment with MβCD led to decreased plasma membrane cholesterol levels, while cholesterol accumulated in intracellular vesicles (Fig 4D, top right panel). We observed similar phenotypic changes in case of amiodarone and chlorpromazine (Fig 4D, bottom panels), while the localisation of cholesterol sensor in loperamide and rosuvastatin treated cells was more similar to control condition (S1 Fig).

For a more systematic and unbiased analysis of the changes in the localisation of cholesterol sensors, we performed quantitative image analysis (S2 Fig). For each cell in each image, we calculated the ratio of average plasma membrane (PM) and average intracellular (IC) D4H-mVenus fluorescence (PM/IC ratio). To segment cells in confocal microscopy images, we used *Cellpose* library ([45], Methods). Plotting the PM/IC ratio as a function of elapsed time after drug treatment (Fig 4E) revealed that PM/IC ratio did not decrease in loperamide and rosuvastatin treated samples, while MβCD, chlorpromazine and amiodarone treatment induced significant decrease of the ratio (linear model coefficients values for interaction between drug treatment and time: -0.002, -0.00086, -0.00017, -0.000032 and 0.000083 for MβCD, chlorpromazine, amiodarone, loperamide, rosuvastatin respectively, p values: $<$1e-200, $<$1e-200, 2.71e-09, 0.25 and 0.0047), confirming the plasma membrane cholesterol depleting effect of chlorpromazine and amiodarone, two SREBF activating drugs.

In summary, our high-throughput image acquisition and analysis pipeline confirmed that chlorpromazine and amiodarone decreased plasma membrane cholesterol content, which explains the increased activity of SREBF transcription factors in case of gene expression readout.

2.5 Supplementing cholesterol reverses anti-SARS-CoV-2 activity of amiodarone

As our experiments revealed that the selected drugs with *in vitro* anti-SARS-CoV-2 activity decreased the cholesterol content of plasma membrane, we were interested in whether decreased plasma membrane cholesterol levels could play a causal role in the antiviral effect, according to our assumptions (Fig 4A). To test this hypothesis, we performed *in vitro* SARS-CoV-2 viral infection assay with cholesterol rescue in Vero-E6 cells.

At first we tested whether the investigated drugs show anti-SARS-CoV-2 activity in our previously described experimental system [46]. Briefly, Vero-E6 cells were co-treated with

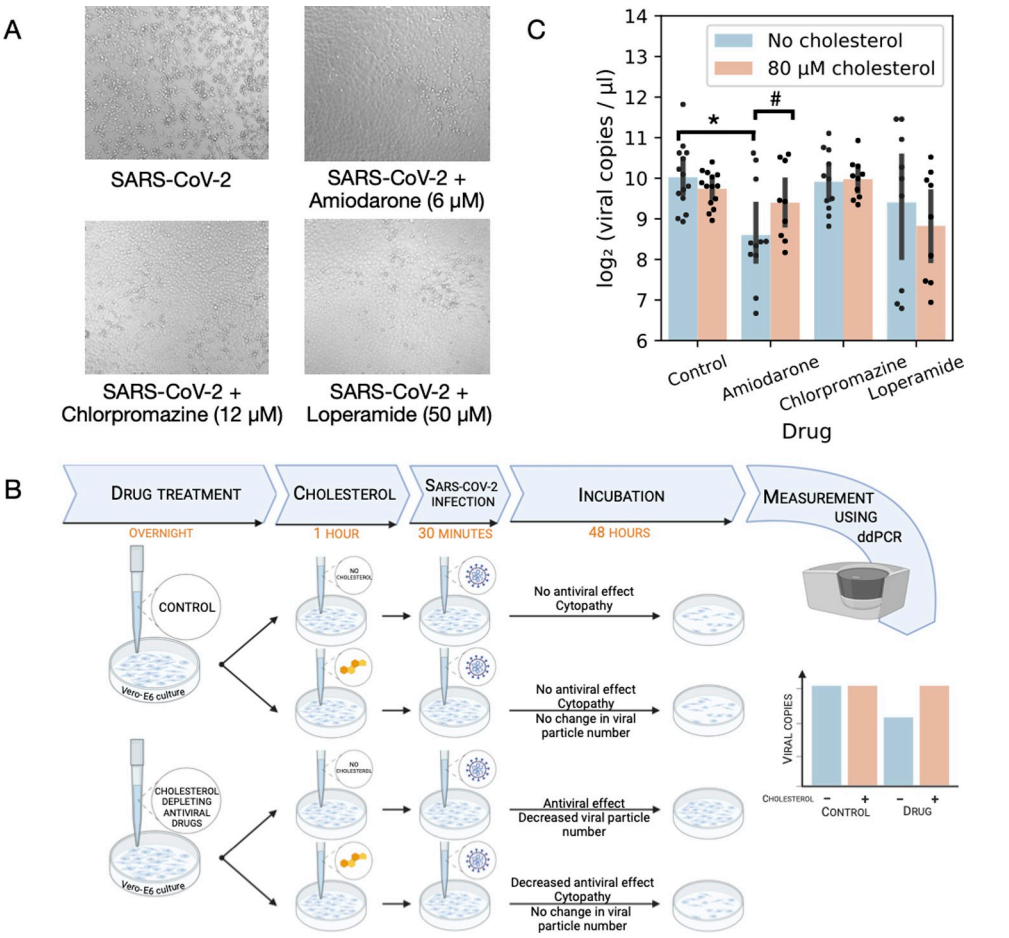

Fig 5. Cholesterol replenishment inhibits antiviral effect of amiodarone. (A) Predicted drugs inhibit SARS-CoV-2 replication in infected Vero-E6 cells. Vero-E6 cells were infected with SARS-CoV-2 (top left) and co-treated either with amiodarone (top right), chlorpromazine (bottom left) or loperamide (bottom right). Antiviral effect (reduced cytopathy) was evaluated by microscopic imaging (10x objective) 48 hours after infection. (B) Schematic figure of cholesterol rescue experiments. The figure was created with BioRender.com. (C) Effect of cholesterol rescue on antiviral drug effect. Vero-E6 cells were pretreated with drugs (x-axis), cholesterol was replenished (color code) and cells were infected with SARS-CoV-2. Antiviral effect of drugs was evaluated 48 hours after infection by droplet digital PCR (viral copies, y-axis). *, #: significant ($p<0.05$) effect of drug treatment and drug treatment-cholesterol interaction in linear model, respectively.

https://doi.org/10.1371/journal.pcbi.1010021.g005

SARS-CoV-2 and the selected drugs (6 μM amiodarone, 12 μM chlorpromazine or 50 μM loperamide, effective drug concentrations were selected based on preliminary experiments) for 30 minutes, then washed and incubated with the drugs for 48 hours. Infection efficacy was evaluated by microscopic examination of infection-induced cytopathic effect (CPE, more details in Methods). Untreated, SARS-CoV-2 infected cells showed strong cytopathy (Fig 5A, top left panel), while amiodarone, chlorpromazine and loperamide markedly reduced the infection-induced cytopathy, confirming the antiviral effect of these drugs (Fig 5A). The used compounds did not lead to cellular toxicity in the used concentrations (S3 Fig).

To test the effect of plasma membrane cholesterol depletion on SARS-CoV-2 infectivity, we performed cholesterol rescue experiments (Fig 5B). Vero-E6 cells were treated with drugs overnight, then the media was replaced with cholesterol (80 μM) containing media. After 1 hour of cholesterol treatment, the cells were infected for 30 min with SARS-CoV-2. Infection efficacy was evaluated 48 hours after infection by droplet digital PCR based viral RNA

quantification (Fig 5C). Chlorpromazine and loperamide did not have antiviral effect in the pretreatment setting (linear model p values: 0.62 and 0.18, respectively), while amiodarone decreased viral particle number significantly (linear model p-value: 1.47e-05). Cholesterol replenishment significantly increased viral particle number in amiodarone treated Vero-E6 cells (amiodarone: cholesterol interaction term p-value: 0.026), confirming the causal role of drug-induced cholesterol depletion in the antiviral effect of amiodarone.

Our *in vitro* SARS-CoV-2 infection assay confirmed the antiviral effects of chlorpromazine, loperamide and amiodarone, and cholesterol rescue experiments suggest that plasma membrane cholesterol depletion plays an important role in the antiviral effect of amiodarone.

3. Discussion

In this study, we analysed the gene expression signatures of *in vitro* SARS-CoV-2 infected cells and effective anti-SARS-CoV-2 drugs. Using functional genomic computational tools, we showed that both virus infection and drug treatment leads to similar changes of pathway and transcription factor activities, like activation of antiviral NFkB and JAK-STAT pathways. Signature similarity between infection and drug-induced signature was predictive for drugs with *in vitro* anti-SARS-CoV-2 activity, contrary to the classical "signature reversal" hypothesis. Using machine learning models we effectively predicted anti-SARS-CoV-2 drugs, and predicted amiodarone as an *in vitro* antiviral compound. More detailed functional genomic analysis of TF activities revealed that SREBF1/2 TFs are strongly activated by large clusters of effective drugs. Using a high-throughput confocal microscopy setup and quantitative image analysis we showed that two of the three investigated effective drugs influence cellular distribution of cholesterol, leading to decreased plasma membrane cholesterol content. Viral infection assay confirmed the already described *in vitro* antiviral activity of loperamide and chlorpromazine, and also the predicted antiviral activity of amiodarone. Cholesterol supplement reversed the antiviral effect of amiodarone, suggesting the causal role of decreased membrane cholesterol in the antiviral effect.

Gene expression-based computational drug repurposing is a promising field to find new disease indications of existing drugs [47]. Despite its simplicity, it has been used successfully to identify repurposable drugs for different diseases from cancer [48,49] through inflammatory [50] to metabolic [51] diseases. While most of the related works rely on the "signature reversal" hypothesis, in case of infection diseases, like COVID-19, it is less clear whether signature reversal (inhibiting the virus-hijacked signalisation) or signature similarity (promoting the antiviral response of infected host cells) can be more effective. While early studies at the beginning of the COVID-19 pandemic applied mostly the original signature reversal hypothesis, more recent works [52,53] also assumed that drugs with similarity to the SARS-CoV-2-induced gene expression signature can be effective. In our work, we performed a more unbiased analysis of signature-based drug repurposing against SARS-CoV-2. We compared the gene expression signatures of known effective drugs against SARS-CoV-2 infection signatures, and found that signature similarity, and not dissimilarity, is predictive for antiviral effect. These results suggest that increasing the antiviral response of host cells can be a more effective strategy than inhibiting viral infection-induced pathways. Whether this is specific for SARS-CoV-2, known for evading several antiviral systems of the host cell [54], or a general mechanism for (viral) infections, needs further analysis with large scale *in vitro* drug screenings against other viruses. Nevertheless, using our "signature similarity" principle instead of—or together with—the "signature reversal" hypothesis can accelerate computational drug repurposing against existing and emerging infectious diseases, complemented by network-based repurposing strategies [40,55].

While signature (dis)similarity-based computational drug repurposing has promising predictive performance, it gives no real mechanistic insight. To overcome this problem, we performed extensive functional genomic analysis of SARS-CoV-2 and drug-induced gene expression signatures. We found that both viral infection and effective drugs stimulate known antiviral pathways like NFkB and JAK-STAT. We observed lower induction of these pathways in virus-infected A549 cells, compared to Calu-3 cell lines, probably based on the lower expression of virus receptor ACE2 in A549 cells. The activation of antiviral pathways in virus-infected and effective drug-treated cells also supports the "signature similarity" principle.

Beside the activation of antiviral TFs and pathways, we also observed inhibition of (inferred) SREBF1/2 transcription factors in SARS-CoV-2 infected samples, while an activation of these TFs in a large cluster of antiviral drug-treated cells. SREBF1/2 regulate the expression of key members of cholesterol synthesis. Cholesterol depletion of plasma membrane can reduce SARS-CoV-2 infection [56,57], and decreasing SREBFs activity (and cholesterol synthesis) can be also part of the physiological, interferon-induced antiviral response of the host cell [21,58,59]. Recently, inhibitors of the SREBFs—DNA interaction were found to exert antiviral effects [60] and CRISPR based knock-out of the SREBF pathway members also led to SARS-CoV-2 resistant phenotype [61], suggesting that inhibition of SREBFs could be beneficial in case of SARS-CoV-2 infection. In contrast, we found increased activity of SREBFs in case of several effective drug-induced gene expression signatures. Previous works also showed the increased expression of lipid metabolic enzymes [12] in antiviral drug-treated cells, and a recent large scale CRISPR screen [62] also found that increased cholesterol synthesis can reduce SARS-CoV-2 infection. However, these two later conclusions were based on the analysis of gene expression changes of the cholesterol synthetic pathway. Gene expression changes are in several cases not the cause, but the (compensatory) consequence of perturbed cell states [63,64]. Based on this, we hypothesized that increased SREBF1/2 activity (based on transcriptional readout) can be a compensatory consequence of decreased plasma membrane cholesterol levels in case of several antiviral drugs. Using a fluorescent cholesterol sensor, we found that amiodarone and chlorpromazine, two effective *in vitro* antiviral drugs, indeed decreases the cholesterol content of plasma membranes, which can explain the (compensatory) increased SREBF1/2 activity. In an *in vitro* SARS-CoV-2 infection assay, coupled with cholesterol rescue, we also showed that cholesterol replenishment reduced the antiviral activity of amiodarone, thus confirmed the causal role of plasma membrane cholesterol decrease in the antiviral effect of amiodarone. While our computational analysis also predicted that PM cholesterol depletion plays a role in the antiviral effect of chlorpromazine and loperamide, we were not able to verify these predictions experimentally. Noteworthy, these two drugs had antiviral effect in case of co-treatment with virus infection, but not in the case of the pre-treatment setup used in cholesterol rescue experiments (probably due to pharmacokinetic factors). It is thus hard to draw conclusions about the role of cholesterol in the antiviral effect of these drugs.

While we showed that PM cholesterol depletion can be an important factor in the *in vitro* antiviral effect of drugs, whether this can be translated to *in vivo* is still an open question. A recent large scale study [65] showed that several *in vitro* repurposable drugs exert their antiviral effect via altering the membrane composition of drug-treated cells, and this antiviral effect has low translation potential based on concerns regarding drug concentration and adverse effects. While the authors of this study concluded that *phospholipidosis* is the main drug-induced membrane component change, our results argue that altered cholesterol content can also be a causal factor in the antiviral effect of drugs. Whether altered lipid composition of cellular membranes is only a factor confounding drug repurposing studies, or this effect can be exploited towards effective therapy, needs further studies.

In summary, our study showed that *in vitro* SARS-CoV-2 infection and effective antiviral drugs lead to similar pathway and transcription activity changes. We found that gene expression signature similarity, and not the dissimilarity, predicts *in vitro* effective antiviral compounds, which can accelerate computational drug repurposing against infectious diseases, and we made the results of our predictions available for the research community (S1 Table). We also identified that plasma membrane cholesterol depletion plays an important role in the mechanism of action of several antiviral drugs, and that cholesterol replenishment inhibits the *in vitro* antiviral effect of amiodarone, thus our results also give mechanistic insight about the antiviral effect of repurposable drugs.

4. Methods

Virus infection-induced gene expression signatures

Microarray gene expression profiles of different virus-infected cell lines were downloaded from Gene Expression Omnibus (GEO) with accession numbers GSE28166 (H5N1), GSE37571 (Influenza), GSE33267 (SARS-CoV-1), GSE56677 and GSE45042 (MERS-CoV). Preprocessing and differential expression (DE) analysis was performed by using R package *limma* [66].

Total RNA-Seq profiles of SARS-CoV-2 and other virus-infected human cell lines were downloaded from GEO with accession numbers GSE147507 (SARS-CoV-2, RSV, IAV, HPIV) and GSE148729 (SARS-CoV-1 and 2). Differential expression (DE) analysis was performed using R library *DESeq2* [67].

In all gene expression datasets, we used (virus-infected—control) contrasts for differential expression calculation, where the control condition was mock infection. Where gene expression data after multiple time points were available, we used 24 h post-infection data. Shared genes across all datasets were selected and further analyzed.

Drug treatment-induced signatures

We used Level 5 gene expression profiles from the LINCS-L1000 dataset [27]. We calculated consensus expression signatures for each drug (across different cell lines, concentrations and time points) using the MODZ method [27,28]. We matched LINCS-L1000 drugs with ChEMBL effective drug dataset (http://chembl.blogspot.com/2020/05/chembl27-sars-cov-2-release.html) using drug names and simplified molecular-input line-entry system (SMILES). Only measured (landmark) genes were used in the further analysis.

Functional genomic analysis

From previously calculated SARS-CoV-2 infection and effective drug-induced signatures, we inferred pathway activities using PROGENy (R package *progeny* [15,16]) and transcription factor activities using DoRothEA (R package *dorothea* [18]).

PROGENy was applied to infer activities of 14 different pathways from expression and weight of their footprint gene sets. Z-scores of pathway activities were calculated using 10000 permutations of genes as background distribution. DoRothEA was applied to infer transcription factor activities using the *viper* algorithm [68]. DoRothEA is a collected, curated resource of signed TF-target interactions. Interactions are assigned a confidence level ranging from A (highest) to E (lowest) based on the number of supporting evidence. In this study interactions assigned A, B, C confidence levels were used. In transcription factor activity heatmaps (Figs 1B and 2B), only selected transcription factors (absolute normalised enrichment score > 4 in SARS-CoV–2 infection signatures) are shown.

We used the CARNIVAL tool [22] to contextualize our transcriptomics-based results into a mechanistic causal network. Briefly, CARNIVAL takes as input a prior knowledge network and a set of constraints and infers the most likely causal interactions by solving an integer linear programming problem. We assembled a curated prior knowledge signaling network from OmniPath resources [69]. As constraints, we selected the RIG-I like receptors (DDX58 and IFIH1) as upstream signaling perturbation and the top 25 most deregulated TFs (according to DoRothEA and *viper* results) upon SARS-CoV-2 infection as their downstream target. In addition, we used PROGENy pathway activity scores to weight the prior knowledge network and assist CARNIVAL in the discovery of optimal networks connecting the upstream perturbation (RIG-I like receptors) to the downstream targets (TFs).

Signature similarity and machine learning-based prediction

We calculated similarities using Spearman's correlation between each virus infection-induced and each drug treatment-induced signature after selecting shared genes.

TF activity scores from drug-treated cells were used to predict effective drugs against SARS-CoV-2 using Random Forest Classifier from scikit-learn Python library [70]. The model was trained using 300 trees, with default parameters otherwise and with 100 different training sets. Training sets consisted of a 50% random sampling of effective drugs and non-effective drugs as well. The average importance of features (TFs) was computed (sum of feature importances, divided by the number of models). Predicted probabilities of antiviral activity were also computed in each prediction and the mean of them was calculated for each drug (probabilities were summed for each drug and divided by the number of occurrences in validation sets).

We performed ROC analysis using scikit-learn Python library to evaluate similarity-based and machine learning-based predictions. Effective drugs against SARS-CoV-2 curated by ChEMBL and overlapping with drugs of the LINCS-L1000 dataset were used as the positive class. The negative class consisted of the part of drugs from the LINCS-L1000 dataset not considered as effective by ChEMBL. To compare machine learning-based and similarity-based methods ROC curves were computed for each different validation set (100) and signature similarity scores of the corresponding drugs were considered.

Fluorescent cholesterol sensor experiments

The cellular cholesterol sensor used in this study was the D4H domain [43,44] fluorescently labeled with monomer Venus (mVenus) on its N-terminus. To create the construct coding this sensor, we used a plasmid coding the bioluminescent version of the sensor (described in [71]), a kind gift from Tamas Balla (NICHD, NIH, Bethesda, USA). The D4H domain-coding sequence from this plasmid was subcloned into the pEYFP-C1 plasmid containing mVenus in place of EYFP, using BglII and BamHI restriction enzymes. Cytosolic Cerulean was expressed from a pEYFP-N1 plasmid where EYFP had been replaced with Cerulean.

For fluorescent imaging, HEK293A cells (ATCC, USA) were maintained in Dulbecco's Modified Eagle Medium (DMEM—Lonza, Switzerland) complemented with 10% fetal bovine serum (Biosera, France) and Penicillin/Streptomycin (100 U/ml and 100 μg/ml, respectively—Lonza, Switzerland). Cells were seeded on poly-L-lysine pretreated (0.001%, 1h) 24-well imaging plates (Eppendorf, Germany) at a density of 1e05 cells/well. On the next day, cells were co-transfected with plasmids coding cytoplasmic Cerulean and D4H-mVenus (0.25 μg/well each) using Lipofectamine 2000 (0.75 μl/well, Invitrogen, USA).

Image acquisition started 24h post-transfection, after the medium had been changed to 300 μl/well HEPES-buffered DMEM without phenol-red (Gibco, USA). Images were acquired automatically using the ImageXpress Micro Confocal High-Content Imaging System

(Molecular Devices, USA), with a 40x Plan Fluor objective. CFP-2432C and fluorescein isothiocyanate (FITC) filter sets were used for Cerulean and D4H-mVenus images, respectively, both with an exposure time of 300 ms. After acquiring control images (30 min), cells were treated with either DMSO (as control) or with the drugs indicated on Fig 4E in a volume of 100 μl/well (270 min). Measurements were performed at 30˚C. Three independent measurements were made, with duplicate wells for each condition and 5 images/well taken for each time point.

All chemicals used for treatment were purchased from Sigma-Aldrich Merck (Germany). Amiodarone HCl, chlorpromazine HCl, loperamid HCl and rosuvastatin calcium were dissolved in DMSO, stored at -20˚C as 10 mM stock solutions and diluted in cell medium promptly before cell treatment to a final concentration of 10 μM. MβCD was stored as powder at 4˚C and freshly dissolved in cell medium before treatment to a final concentration of 10 mM.

Image analysis pipeline

Images were segmented with *Cellpose* Python library [45], which is a generalist, deep learning-based segmentation method. To select high-quality images the cytoplasm marker channel was used with Laplace filtering. We used high-quality images (filtered according to an appropriate upper threshold of standard deviation of Laplace value in each experiment) as input of the Cellpose model, with parameter channel set to greyscale and cell diameter greater than 200 pixels.

After identifying cell boundaries, we applied binary erosion (scipy Python library [72]) with default structure and 10 iterations to determine cytoplasm boundary, or binary dilation with default structure and 5 iterations to determine PM outer boundary. The boundary of PM was determined by subtracting the cytoplasm boundary from the outer boundary. We calculated the $\log_2$ ratio of the mean PM and mean intracellular D4H fluorescence intensities for each cell in the D4H channel to examine the changes of plasma membrane cholesterol distribution. For statistical analysis, we used log2(PM/IC) ~ Time + Time: Drug + Exp linear model, where Time corresponds to elapsed time after drug treatment, Drug factor represents the used drug, using DMSO as reference level. Exp factor represents the (n = 3) individual experiments.

Viral infection and cholesterol rescue experiments

Amiodarone HCl (Sigma-Aldrich, Merck KGaA, Germany) was dissolved in DMSO (Sigma-Aldrich, Merck KGaA, Germany) and kept at -20˚C. Chlorpromazine (in house synthesized based on [73]) and loperamide HCl (Sigma-Aldrich, Merck KGaA, Germany) were freshly dissolved in water and filtered prior to the treatment. 10 mM stock solutions were made from the drugs. Vero-E6 cells were seeded in a 96-well plate on the day before the experiments. On the next day the cells were treated with 100 μl of 50 μM remdesivir or loperamide or 12 μM chlorpromazine or 6 μM amiodarone solution overnight. 1 hour prior to the infection the cell culture media containing the different drugs was replaced with media containing 80 μM cholesterol (Sigma-Aldrich, Merck KGaA, Germany). After the 1-hour-long cholesterol treatment the cells were infected with SARS-CoV-2 (GISAID accession ID: EPI_ISL_483637) at MOI:0.01 in a BSL-4 laboratory. Cells were incubated with the virus for 30 minutes then the media was replaced with fresh cell culture media. During the investigation (except cell seeding) DMEM (Lonza Group Ltd, Switzerland) supplemented with 1% Penicillin- Streptomycin (Lonza Group Ltd, Switzerland) and 2% heat-inactivated fetal bovine serum (Gibco, Thermo Fisher Scientific Inc., MA, USA) were used. 48 hours post infection (hpi) the cells were inspected under microscope and RNA was extracted from the supernatant (Zybio EXM 3000

Nucleic Acid Isolation System, Nucleic Acid Extraction Kit B200-32). Viral copy number was determined using droplet-digital PCR technology (Bio-Rad Laboratories Inc., CA, USA). SARS-CoV-2 RdRp gene specific primers and probe were utilized (Forward: GTGARATGGT CATGTGTGGCGG, reverse: CARATGTTAAASACACTATTAGCATA and the probe was: FAM-CAGGTGGAACCTCATCAGGAGATGC-BBQ). For statistical analysis, measured viral copy numbers were $\log_2$ transformed, and we used a log2(CV) ~ Drug * Cholesterol + Exp, where Drug factor represents the used drug (untreated as reference level), Cholesterol factor represents cholesterol replenishment treatment (no treatment as reference level). Exp factor corresponds to the (n = 4) individual experiments.

Supporting information

S1 Fig. Representative confocal microscopy images of D4H-mVenus transfected HEK293A cells treated with loperamide and rosuvastatin.
(TIFF)

S2 Fig. Determination of cytoplasm and membrane boundaries. Cells are detected on the cytoplasm marker channel, then boundaries of cytoplasm and membrane are determined for each cell. The D4H channel is used for the calculation of the PM/IC ratio.
(TIFF)

S3 Fig. Absence of marked toxic effects of used drugs in the tested concentrations.
(TIFF)

S1 Table. Drug signature similarities (Spearman's rho) to SARS-CoV-2 infection signatures (GSE147507 A549 SARS-CoV-2, GSE147507 Calu3 SARS-CoV-2 and GSE148729 Calu3 SARS-CoV-2 columns, respectively) and Random Forest Classifier based predicted probability of antiviral effect.
(CSV)

Acknowledgments

We thank Aurélien Dugourd, Gergő Gulyás, Kinga Kovács and András D. Tóth for the helpful discussions regarding the manuscript, and Péter Mátyus for his help in organising the collaborations between computational and experimental groups. We thank Kata Szabolcsi for technical assistance in the cholesterol sensor experiments, Katalin Gombos, Zsófia Lanszki and Balázs Somogyi for their help in the *in vitro* infection experiments, and Tamás Kálai for compound synthesis. On behalf of Project DRUGSENSPRED we thank for the usage of ELKH Cloud (https://science-cloud.hu/) that significantly helped us achieve the results published in this paper. Schematic figures were created with BioRender.com.

Author Contributions

Conceptualization: Szilvia Barsi, Bence Szalai.

Data curation: Szilvia Barsi, Bence Szalai.

Formal analysis: Szilvia Barsi, Alberto Valdeolivas, Bence Szalai.

Funding acquisition: Péter Várnai, Ferenc Jakab, Bence Szalai.

Investigation: Szilvia Barsi, Henrietta Papp, Dániel J. Tóth, Anett Kuczmog, Mónika Madai, Péter Várnai, Bence Szalai.

Methodology: Szilvia Barsi, Alberto Valdeolivas, Bence Szalai.

Project administration: László Hunyady.

Resources: Dániel J. Tóth, Péter Várnai, Ferenc Jakab, Bence Szalai.

Supervision: László Hunyady, Péter Várnai, Julio Saez-Rodriguez, Ferenc Jakab, Bence Szalai.

Validation: Henrietta Papp, Dániel J. Tóth, Anett Kuczmog, Mónika Madai, Péter Várnai.

Visualization: Szilvia Barsi, Alberto Valdeolivas.

Writing – original draft: Szilvia Barsi, Bence Szalai.

Writing – review & editing: Szilvia Barsi, Henrietta Papp, Alberto Valdeolivas, Dániel J. Tóth, Anett Kuczmog, Mónika Madai, László Hunyady, Péter Várnai, Julio Saez-Rodriguez, Ferenc Jakab, Bence Szalai.

References

1. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis. 2020; 20: 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1 PMID: 32087114
2. McCallum M, Bassi J, De Marco A, Chen A, Walls AC, Di Iulio J, et al. SARS-CoV-2 immune evasion by the B.1.427/B.1.429 variant of concern. Science. 2021; 373: 648–654. https://doi.org/10.1126/science.abi7994 PMID: 34210893
3. Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res. 2020; 30: 269–271. https://doi.org/10.1038/s41422-020-0282-0 PMID: 32020029
4. Gordon DE, Jang GM, Bouhaddou M, Xu J, Obernier K, White KM, et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature. 2020. https://doi.org/10.1038/s41586-020-2286-9 PMID: 32353859
5. Bouhaddou M, Memon D, Meyer B, White KM, Rezelj VV, Correa Marrero M, et al. The Global Phosphorylation Landscape of SARS-CoV-2 Infection. Cell. 2020. https://doi.org/10.1016/j.cell.2020.06.034 PMID: 32645325
6. Cuccarese MF, Earnshaw BA, Heiser K, Fogelson B, Davis CT, McLean PF, et al. Functional immune mapping with deep-learning enabled phenomics applied to immunomodulatory and COVID-19 drug discovery. 2020. p. 2020.08.02.233064. https://doi.org/10.1101/2020.08.02.233064
7. Ragab D, Salah Eldin H, Taeimah M, Khattab R, Salem R. The COVID-19 Cytokine Storm; What We Know So Far. Front Immunol. 2020; 11: 1446. https://doi.org/10.3389/fimmu.2020.01446 PMID: 32612617
8. Olbei M, Hautefort I, Modos D, Treveil A, Poletti M, Gul L, et al. SARS-CoV-2 Causes a Different Cytokine Response Compared to Other Cytokine Storm-Causing Respiratory Viruses in Severely Ill Patients. Front Immunol. 2021; 12: 629193. https://doi.org/10.3389/fimmu.2021.629193 PMID: 33732251
9. Sirota M, Dudley JT, Kim J, Chiang AP, Morgan AA, Sweet-Cordero A, et al. Discovery and preclinical validation of drug indications using compendia of public gene expression data. Sci Transl Med. 2011; 3: 96ra77. https://doi.org/10.1126/scitranslmed.3001318 PMID: 21849665
10. Napolitano F, Gambardella G, Carrella D, Gao X, di Bernardo D. Computational Drug Repositioning and Elucidation of Mechanism of Action of Compounds against SARS-CoV-2. arXiv [q-bio.GN]. 2020. Available: http://arxiv.org/abs/2004.07697
11. Zhou Y, Hou Y, Shen J, Huang Y, Martin W, Cheng F. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov. 2020; 6: 14. https://doi.org/10.1038/s41421-020-0153-3 PMID: 32194980
12. Hoagland DA, Clarke DJB, Møller R, Han Y, Yang L, Wojciechowicz ML, et al. Modulating the transcriptional landscape of SARS-CoV-2 as an effective method for developing antiviral compounds. bioRxiv. 2020. p. 2020.07.12.199687. https://doi.org/10.1101/2020.07.12.199687
13. Blanco-Melo D, Nilsson-Payant BE, Liu W-C, Uhl S, Hoagland D, Møller R, et al. Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. Cell. 2020; 181: 1036–1045.e9. https://doi.org/10.1016/j.cell.2020.04.026 PMID: 32416070

14. Wyler E, Mösbauer K, Franke V, Diag A, Gottula LT, Arsiè R, et al. Transcriptomic profiling of SARS-CoV-2 infected human cell lines identifies HSP90 as target for COVID-19 therapy. iScience. 2021; 24: 102151. https://doi.org/10.1016/j.isci.2021.102151 PMID: 33585804

15. Schubert M, Klinger B, Klünemann M, Sieber A, Uhlitz F, Sauer S, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. Nat Commun. 2018; 9: 20. https://doi.org/10.1038/s41467-017-02391-6 PMID: 29295995

16. Holland CH, Szalai B, Saez-Rodriguez J. Transfer of regulatory knowledge from human to mouse for functional genomics analysis. Biochim Biophys Acta Gene Regul Mech. 2019; 194431. https://doi.org/10.1016/j.bbagrm.2019.194431 PMID: 31525460

17. Garcia-Alonso L, Iorio F, Matchan A, Fonseca N, Jaaks P, Peat G, et al. Transcription Factor Activities Enhance Markers of Drug Sensitivity in Cancer. Cancer Res. 2018; 78: 769–780. https://doi.org/10.1158/0008-5472.CAN-17-1679 PMID: 29229604

18. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. 2019; 29: 1363–1375. https://doi.org/10.1101/gr.240663.118 PMID: 31340985

19. Seth RB, Sun L, Chen ZJ. Antiviral innate immunity pathways. Cell Res. 2006; 16: 141–147. https://doi.org/10.1038/sj.cr.7310019 PMID: 16474426

20. Ramana CV, Grammatikakis N, Chernov M, Nguyen H, Goh KC, Williams BR, et al. Regulation of c-myc expression by IFN-gamma through Stat1-dependent and -independent pathways. EMBO J. 2000; 19: 263–272. https://doi.org/10.1093/emboj/19.2.263 PMID: 10637230

21. York AG, Williams KJ, Argus JP, Zhou QD, Brar G, Vergnes L, et al. Limiting Cholesterol Biosynthetic Flux Spontaneously Engages Type I IFN Signaling. Cell. 2015; 163: 1716–1729. https://doi.org/10.1016/j.cell.2015.11.045 PMID: 26686653

22. Liu A, Trairatphisan P, Gjerga E, Didangelos A, Barratt J, Saez-Rodriguez J. From expression footprints to causal pathways: contextualizing large signaling networks with CARNIVAL. NPJ Syst Biol Appl. 2019; 5: 40. https://doi.org/10.1038/s41540-019-0118-z PMID: 31728204

23. Rehwinkel J, Gack MU. RIG-I-like receptors: their regulation and roles in RNA sensing. Nat Rev Immunol. 2020; 20: 537–551. https://doi.org/10.1038/s41577-020-0288-3 PMID: 32203325

24. Kindrachuk J, Ork B, Hart BJ, Mazur S, Holbrook MR, Frieman MB, et al. Antiviral potential of ERK/MAPK and PI3K/AKT/mTOR signaling modulation for Middle East respiratory syndrome coronavirus infection as identified by temporal kinome analysis. Antimicrob Agents Chemother. 2015; 59: 1088–1099. https://doi.org/10.1128/AAC.03659-14 PMID: 25487801

25. Tian J, Zhang X, Wu H, Liu C, Li Z, Hu X, et al. Blocking the PI3K/AKT pathway enhances mammalian reovirus replication by repressing IFN-stimulated genes. Front Microbiol. 2015; 6: 886. https://doi.org/10.3389/fmicb.2015.00886 PMID: 26388843

26. Ehrhardt C, Wolff T, Pleschka S, Planz O, Beermann W, Bode JG, et al. Influenza A virus NS1 protein activates the PI3K/Akt pathway to mediate antiapoptotic signaling responses. J Virol. 2007; 81: 3058–3067. https://doi.org/10.1128/JVI.02082-06 PMID: 17229704

27. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell. 2017; 171: 1437–1452.e17. https://doi.org/10.1016/j.cell.2017.10.049 PMID: 29195078

28. Szalai B, Subramanian V, Holland CH, Alföldi R, Puskás LG, Saez-Rodriguez J. Signatures of cell death and proliferation in perturbation transcriptomics data-from confounding factor to effective prediction. Nucleic Acids Res. 2019; 47: 10010–10026. https://doi.org/10.1093/nar/gkz805 PMID: 31552418

29. Riva L, Yuan S, Yin X, Martin-Sancho L, Matsunaga N, Burgstaller S, et al. A Large-scale Drug Repositioning Survey for SARS-CoV-2 Antivirals. bioRxiv. 2020. p. 2020.04.16.044016. https://doi.org/10.1101/2020.04.16.044016 PMID: 32511357

30. Weston S, Coleman CM, Haupt R, Logue J, Matthews K, Frieman MB. Broad anti-coronaviral activity of FDA approved drugs against SARS-CoV-2 in vitro and SARS-CoV in vivo. 2020. p. 2020.03.25.008482. https://doi.org/10.1101/2020.03.25.008482

31. Heiser K, McLean PF, Davis CT, Fogelson B, Gordon HB, Jacobson P, et al. Identification of potential treatments for COVID-19 through artificial intelligence-enabled phenomic analysis of human cells infected with SARS-CoV-2. bioRxiv. 2020. p. 2020.04.21.054387. https://doi.org/10.1101/2020.04.21.054387

32. Si L, Bai H, Rodas M, Cao W, Oh CY, Jiang A, et al. Human organ chip-enabled pipeline to rapidly repurpose therapeutics during viral pandemics. 2020. p. 2020.04.13.039917. https://doi.org/10.1101/2020.04.13.039917

33. Touret F, Gilles M, Barral K, Nougairède A, van Helden J, Decroly E, et al. In vitro screening of a FDA approved chemical library reveals potential inhibitors of SARS-CoV-2 replication. Sci Rep. 2020; 10: 13093. https://doi.org/10.1038/s41598-020-70143-6 PMID: 32753646

34. Ellinger B, Bojkova D, Zaliani A, Cinatl J, Claussen C, Westhaus S, et al. A SARS-CoV-2 cytopathicity dataset generated by high-content screening of a large drug repurposing collection. Sci Data. 2021; 8: 70. https://doi.org/10.1038/s41597-021-00848-4 PMID: 33637768

35. Jeon S, Ko M, Lee J, Choi I, Byun SY, Park S, et al. Identification of antiviral drug candidates against SARS-CoV-2 from FDA-approved drugs. https://doi.org/10.1101/2020.03.20.999730

36. Sims AC, Tilton SC, Menachery VD, Gralinski LE, Schäfer A, Matzke MM, et al. Release of severe acute respiratory syndrome coronavirus nuclear import block enhances host transcription in human lung cells. J Virol. 2013; 87: 3885–3902. https://doi.org/10.1128/JVI.02520-12 PMID: 23365422

37. Josset L, Menachery VD, Gralinski LE, Agnihothram S, Sova P, Carter VS, et al. Cell host response to infection with novel human coronavirus EMC predicts potential antivirals and important differences with SARS coronavirus. MBio. 2013; 4: e00165–13. https://doi.org/10.1128/mBio.00165-13 PMID: 23631916

38. Selinger C, Tisoncik-Go J, Menachery VD, Agnihothram S, Law GL, Chang J, et al. Cytokine systems approach demonstrates differences in innate and pro-inflammatory host responses between genetically distinct MERS-CoV isolates. BMC Genomics. 2014; 15: 1161. https://doi.org/10.1186/1471-2164-15-1161 PMID: 25534508

39. Li C, Bankhead A 3rd, Eisfeld AJ, Hatta Y, Jeng S, Chang JH, et al. Host regulatory network response to infection with highly pathogenic H5N1 avian influenza virus. J Virol. 2011; 85: 10955–10967. https://doi.org/10.1128/JVI.05792-11 PMID: 21865398

40. Morselli Gysi D, do Valle Í, Zitnik M, Ameli A, Gan X, Varol O, et al. Network medicine framework for identifying drug-repurposing opportunities for COVID-19. Proc Natl Acad Sci U S A. 2021;118. https://doi.org/10.1073/pnas.2025581118 PMID: 33906951

41. Liang H, Zhang L, Wang L, Gao M, Meng X, Li M, et al. Repositioning Drugs on Human Influenza A Viruses Based on a Novel Nuclear Norm Minimization Method. Front Physiol. 2020; 11: 597494. https://doi.org/10.3389/fphys.2020.597494 PMID: 33536933

42. Horton JD, Goldstein JL, Brown MS. SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. J Clin Invest. 2002; 109: 1125–1131. https://doi.org/10.1172/JCI15593 PMID: 11994399

43. Maekawa M, Fairn GD. Complementary probes reveal that phosphatidylserine is required for the proper transbilayer distribution of cholesterol. J Cell Sci. 2015; 128: 1422–1433. https://doi.org/10.1242/jcs.164715 PMID: 25663704

44. Maekawa M. Domain 4 (D4) of Perfringolysin O to Visualize Cholesterol in Cellular Membranes-The Update. Sensors. 2017;17. https://doi.org/10.3390/s17030504 PMID: 28273804

45. Stringer C, Wang T, Michaelos M, Pachitariu M. Cellpose: a generalist algorithm for cellular segmentation. Nat Methods. 2021; 18: 100–106. https://doi.org/10.1038/s41592-020-01018-x PMID: 33318659

46. Konrat R, Papp H, Szijártó V, Gesell T, Nagy G, Madai M, et al. The Anti-histamine Azelastine, Identified by Computational Drug Repurposing, Inhibits SARS-CoV-2 Infection in Reconstituted Human Nasal Tissue In Vitro. bioRxiv. 2020. p. 2020.09.15.296228. https://doi.org/10.1101/2020.09.15.296228

47. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. 2019; 18: 41–58. https://doi.org/10.1038/nrd.2018.168 PMID: 30310233

48. Chen B, Ma L, Paik H, Sirota M, Wei W, Chua M-S, et al. Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. Nat Commun. 2017; 8: 16022. https://doi.org/10.1038/ncomms16022 PMID: 28699633

49. Stathias V, Jermakowicz AM, Maloof ME, Forlin M, Walters W, Suter RK, et al. Drug and disease signature integration identifies synergistic combinations in glioblastoma. Nat Commun. 2018; 9: 5315. https://doi.org/10.1038/s41467-018-07659-z PMID: 30552330

50. Malcomson B, Wilson H, Veglia E, Thillaiyampalam G, Barsden R, Donegan S, et al. Connectivity mapping (ssCMap) to predict A20-inducing drugs and their antiinflammatory action in cystic fibrosis. Proc Natl Acad Sci U S A. 2016; 113: E3725–34. https://doi.org/10.1073/pnas.1520289113 PMID: 27286825

51. Kunkel SD, Suneja M, Ebert SM, Bongers KS, Fox DK, Malmberg SE, et al. mRNA expression signatures of human skeletal muscle atrophy identify a natural compound that increases muscle mass. Cell Metab. 2011; 13: 627–638. https://doi.org/10.1016/j.cmet.2011.03.020 PMID: 21641545

52. Laise P, Bosker G, Sun X, Shen Y, Douglass EF, Karan C, et al. The Host Cell ViroCheckpoint: Identification and Pharmacologic Targeting of Novel Mechanistic Determinants of Coronavirus-Mediated Hijacked Cell States. bioRxiv. 2020. p. 2020.05.12.091256. https://doi.org/10.1101/2020.05.12.091256 PMID: 32511361

53. Chen F, Shi Q, Pei F, Vogt A, Porritt RA, Garcia G Jr, et al. A systems-level study reveals host-targeted repurposable drugs against SARS-CoV-2 infection. Mol Syst Biol. 2021; 17: e10239. https://doi.org/10.15252/msb.202110239 PMID: 34339582

54. Lei X, Dong X, Ma R, Wang W, Xiao X, Tian Z, et al. Activation and evasion of type I interferon responses by SARS-CoV-2. Nat Commun. 2020; 11: 3810. https://doi.org/10.1038/s41467-020-17665-9 PMID: 32733001

55. Sadegh S, Matschinske J, Blumenthal DB, Galindez G, Kacprowski T, List M, et al. Exploring the SARS-CoV-2 virus-host-drug interactome for drug repurposing. Nat Commun. 2020; 11: 3518. https://doi.org/10.1038/s41467-020-17189-2 PMID: 32665542

56. Wang S, Li W, Hui H, Tiwari SK, Zhang Q, Croker BA, et al. Cholesterol 25-Hydroxylase inhibits SARS-CoV-2 and other coronaviruses by depleting membrane cholesterol. EMBO J. 2020; 39: e106057. https://doi.org/10.15252/embj.2020106057 PMID: 32944968

57. Sanders DW, Jumper CC, Ackerman PJ, Bracha D, Donlic A, Kim H, et al. SARS-CoV-2 requires cholesterol for viral entry and pathological syncytia formation. Elife. 2021;10. https://doi.org/10.7554/eLife.65962 PMID: 33890572

58. O'Neill LAJ. How Low Cholesterol Is Good for Anti-viral Immunity. Cell. 2015. pp. 1572–1574. https://doi.org/10.1016/j.cell.2015.12.004 PMID: 26687349

59. Blanc M, Hsieh WY, Robertson KA, Watterson S, Shui G, Lacaze P, et al. Host defense against viral infection involves interferon mediated down-regulation of sterol biosynthesis. PLoS Biol. 2011; 9: e1000598. https://doi.org/10.1371/journal.pbio.1000598 PMID: 21408089

60. Yuan S, Chu H, Chan JF-W, Ye Z-W, Wen L, Yan B, et al. SREBP-dependent lipidomic reprogramming as a broad-spectrum antiviral target. Nat Commun. 2019; 10: 120. https://doi.org/10.1038/s41467-018-08015-x PMID: 30631056

61. Wang R, Simoneau CR, Kulsuptrakul J, Bouhaddou M, Travisano KA, Hayashi JM, et al. Genetic Screens Identify Host Factors for SARS-CoV-2 and Common Cold Coronaviruses. Cell. 2021; 184: 106–119.e14. https://doi.org/10.1016/j.cell.2020.12.004 PMID: 33333024

62. Daniloski Z, Jordan TX, Wessels H-H, Hoagland DA, Kasela S, Legut M, et al. Identification of Required Host Factors for SARS-CoV-2 Infection in Human Cells. Cell. 2021; 184: 92–105.e16. https://doi.org/10.1016/j.cell.2020.10.030 PMID: 33147445

63. Dugourd A, Saez-Rodriguez J. Footprint-based functional analysis of multi-omic data. Current Opinion in Systems Biology. 2019. Available: https://www.sciencedirect.com/science/article/pii/S2452310019300149

64. Szalai B, Saez-Rodriguez J. Why do pathway methods work better than they should? FEBS Lett. 2020; 594: 4189–4200. https://doi.org/10.1002/1873-3468.14011 PMID: 33270910

65. Tummino TA, Rezelj VV, Fischer B, Fischer A, O'Meara MJ, Monel B, et al. Drug-induced phospholipidosis confounds drug repurposing for SARS-CoV-2. Science. 2021; 373: 541–547. https://doi.org/10.1126/science.abi4708 PMID: 34326236

66. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43: e47. https://doi.org/10.1093/nar/gkv007 PMID: 25605792

67. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15: 550. https://doi.org/10.1186/s13059-014-0550-8 PMID: 25516281

68. Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH, et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. Nat Genet. 2016; 48: 838–847. https://doi.org/10.1038/ng.3593 PMID: 27322546

69. Türei D, Valdeolivas A, Gul L, Palacio-Escat N, Klein M, Ivanova O, et al. Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. Mol Syst Biol. 2021; 17: e9923. https://doi.org/10.15252/msb.20209923 PMID: 33749993

70. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. J Mach Learn Res. 2011; 12: 2825–2830.

71. Sohn M, Korzeniowski M, Zewe JP, Wills RC, Hammond GRV, Humpolickova J, et al. PI(4,5)P2 controls plasma membrane PI4P and PS levels via ORP5/8 recruitment to ER-PM contact sites. J Cell Biol. 2018; 217: 1797–1813. https://doi.org/10.1083/jcb.201710095 PMID: 29472386

72. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020; 17: 261–272. https://doi.org/10.1038/s41592-019-0686-2 PMID: 32015543

73. Galons H, Miocque M, Combet-Farnoux C, Bensaid Y, Decodts G, Bram G. A Convenient Procedure for the Synthesis of Phenothiazine Drugs. Chem Pharm Bull. 1985; 33: 5108–5109.

PLOS Computational Biology

RESEARCH ARTICLE

RIDDEN: Data-driven inference of receptor activity from transcriptomic data

Szilvia Barsi[1,2], Eszter Varga[2], Daniel Dimitrov[3], Julio Saez-Rodriguez[3], László Hunyady[1,2]*, Bence Szalai[1,2¤]*

1 Institute of Molecular Life Sciences, Centre of Excellence of the Hungarian Academy of Sciences, HUN-REN Research Centre for Natural Sciences, Budapest, Hungary, **2** Department of Physiology, Faculty of Medicine, Semmelweis University, Budapest, Hungary, **3** Heidelberg University, Faculty of Medicine, and Heidelberg University Hospital, Institute for Computational Biomedicine, Heidelberg, Germany

¤ Current address: Turbine Ltd., Budapest, Hungary
* hunyady.laszlo@ttk.hu (LH); ben.szalai.cb@gmail.com (BS)

Citation: Barsi S, Varga E, Dimitrov D, Saez-Rodriguez J, Hunyady L, Szalai B (2025) RIDDEN: Data-driven inference of receptor activity from transcriptomic data. PLoS Comput Biol 21(6): e1013188. https://doi.org/10.1371/journal.pcbi.1013188

Editor: Joshua Welch, University of Michigan, UNITED STATES OF AMERICA

Received: December 10, 2024

Accepted: June 3, 2025

Published: June 16, 2025

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pcbi.1013188

Abstract

Intracellular signaling initiated from ligand-bound receptors plays a fundamental role in both physiological regulation and development of disease states, making receptors one of the most frequent drug targets. Systems level analysis of receptor activity can help to identify cell and disease type-specific receptor activity alterations. While several computational methods have been developed to analyze ligand-receptor interactions based on transcriptomics data, none of them focuses directly on the receptor side of these interactions. Also, most of the methods use directly the expression of ligands and receptors to infer active interaction, while co-expression of genes does not necessarily indicate functional interactions or activated state. To address these problems, we developed RIDDEN (Receptor actIvity Data Driven inferENce), a computational tool, which predicts receptor activities from the receptor-regulated gene expression profiles, and not from the expressions of ligand and receptor genes. We collected 14463 perturbation gene expression profiles for 229 different receptors. Using these data, we trained the RIDDEN model, which can effectively predict receptor activity for new bulk and single-cell transcriptomics datasets. We validated RIDDEN's performance on independent *in vitro* and *in vivo* receptor perturbation data, showing that RIDDEN's model weights correspond to known regulatory interactions between receptors and transcription factors, and that predicted receptor activities correlate with receptor and ligand expressions in *in vivo* datasets. We also show that RIDDEN can be used to identify mechanistic biomarkers in an immune checkpoint blockade-treated cancer patient cohort. RIDDEN, the largest transcriptomics-based receptor activity inference model, can be used to identify cell populations with altered receptor activity and, in turn, foster the study of cell-cell communication using transcriptomics data.

Data availability statement: The source code and analysis pipeline for RIDDEN are accessible at the following GitHub repository: https://github.com/basvaat/RIDDEN_analysis. The application is available at https://github.com/basvaat/RIDDEN_tool. The perturbation signatures of the receptors and the table of ligand-receptor interactions used to develop the RIDDEN model are available on Zenodo (DOI: https://doi.org/10.5281/zenodo.15127392).

Funding: This work was supported by the ÚNKP-22-3-II New National Excellence Program of the Ministry for Culture and Innovation from the source of National Research (SB), Development and Innovation Fund; the Hungarian National Research, Development and Innovation Fund ADVANCED 151284 (LH); National Research, Development and Innovation Office of Hungary (grant name: PharmaLab, RRF-2.3.1-21-2022-00015) (LH). BS was supported by the Premium Postdoctoral Fellowship Program of the Hungarian Academy of Sciences (460044). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: J.S.R. reports funding from GSK, Pfizer and Sanofi and fees/honoraria from Travere Therapeutics, Stadapharm, Astex, Pfizer, Grunenthal and Owkin. B.S. is a full time employee of Turbine Ltd. and reports fees from TheraMe! AG.

Author summary

Receptors play a crucial role in intercellular communication, thereby influencing essential physiological processes. By identifying which receptors are active and initiating signaling within cells we can gain insights into the processes triggered by this communication. Changes in gene expression patterns regulated by these receptors allow us to infer their activity, providing more insight than measuring the receptor or ligand gene expression levels, which often fail to accurately reflect the actual protein activity within the cells. To address this, we developed RIDDEN (Receptor actIvity Data Driven inferENce), a tool for predicting receptor activity by summarizing gene expression profiles regulated by receptors into interpretable activity profiles. We evaluated RIDDEN's performance and found that it reliably captures receptor activity and its biological implications. Thus, RIDDEN enhances our understanding of cellular processes related to communication between cells and helps identify the sources of signaling that can lead to various cellular phenotypes.

1. Introduction

Ligands and receptors are considered the key drivers of cell-cell communication (CCC), which process is essential for physiological functions. The ligand produced by the "sender-cell" reaches the "receiver-cell", where it binds to its receptor and changes its activity. This binding initiates downstream signaling from the receptor, and, as a result, it alters the receiver cell's state. This process is vital for cells to respond to their environment, which regulates essential processes, such as maintaining homeostasis, cell growth, differentiation or immune interactions. Dysregulation of the receptor activation, caused by altered ligand binding, mutation or overexpression of the receptor, can lead to various diseases, including changes in insulin, neurotransmitter, G protein-coupled or overexpression/overactivation of growth factor receptors leading to insulin resistance [1], neurological disorders [2], endocrine diseases [3] or cancer development [4], respectively.

Experimental investigation of ligand-receptor interactions on the systems biology scale remains challenging due to the complexity of cellular interactions [5,6]. The limited scope of studying a few isolated cells experimentally restricts the ability to obtain comprehensive information on communication [7]. A large number of computational methods have emerged to discover and analyze these interactions in the last decade [6]. Most methods use prior-knowledge-based lists of receptor-ligand interactions coupled with statistical methods to identify sender-receiver cell pairs with significant expression of corresponding ligand-receptor pairs [8–13]. Several new methods [14–16] focus on the ligand-induced expression changes of "receiver-cell" to infer the activity of ligand-induced signaling, however no current method is able to infer receptor activity directly.

Computational methods predominantly rely on transcriptomic data because it is broadly available, and RNA sequencing is a high-throughput and cost-effective

approach [17]. A limitation of transcriptomics data is that gene expression cannot be directly translated to protein levels due to regulatory mechanisms, splicing, or post-translational modifications [5,18]. However, if gene expression changes are considered a "footprint" (consequence) of altered protein activity, activity can be computationally inferred from gene expression changes. Methods using these ideas are generally called "footprint-based" tools [19]. These methods require information about the genes regulated by the proteins of interest. While these regulatory interactions are well-characterised in the case of transcription factors [20], in the case of other signaling proteins, like receptors, less information is available. Perturbation gene expression signatures, where gene expression is measured after the perturbation of a given protein, can help to identify such protein-gene regulatory interactions [20,21].

Building on this principle, a recent study [14] introduced a method (CytoSig), that predicts cytokine signaling activities from transcriptomic profiles. It is based on a collection of cytokine stimulation signatures accessible in public databases, like the Gene Expression Omnibus [22]. While CytoSig effectively predicts the activity changes for 43 cytokines, the authors describe that lack of data is a limitation of the research field. Some cytokines were trained on a low number of expression profiles. In addition, these experiments were not conducted based on a standardised method, and the use of different platforms for sequencing led to high variability between samples.

We developed RIDDEN (Receptor actIvity Data Driven inferENce), a statistical model that infers receptor activities from transcriptomic profiles to overcome the limitation of the low number of high-quality signatures in public databases and to extend the number of predictable signaling molecules. To enable the inference of receptor activities, we combined the advantages of the prior knowledge of ligand-receptor interactions from the OmniPath resource [23,24] and a large collection of uniformly conducted and processed ligand and receptor perturbation gene expression experiments from The Library of Integrated Network-Based Cellular Signatures (LINCS), which is the most extensive collection of gene expression profiles of perturbations applied in a wide range of time points, doses and several cell lines [25]. We developed a model to predict the activity of 229 receptors and demonstrated its performance by predicting cytokine signaling activities in comparison with the state-of-the-art method, the CytoSig [14]. The RIDDEN resource and console application are accessible at https://github.com/basvaat/RIDDEN_tool. We show how the collected resource correlates with fundamental biological processes and a case study where receptor activity can be used as a biomarker of patient survival in immunotherapy.

2. Results

2.1. Establishment of the RIDDEN model for receptor activity inference

As the first step of constructing the statistical model that infers receptor activity from transcriptomic profiles a comprehensive dataset containing receptor and ligand perturbation data was required. We obtained curated ligand-receptor interactions from the OmniPath database [23,24]. Based on this prior knowledge, we collected the chemical (drug treatment, ligand stimulation) and genomic (knock-out, knock-down, overexpression) perturbation profiles of all the receptors and ligands available in the LINCS L1000 database [25]. We used level 5, log-normalized, standardised gene expression profiles in our study. In total, our database consisted of 38989 consensus transcriptional profiles for 599 receptors and ligands in 5 different perturbation types, and these data serve as the basis of the statistical model.

For training RIDDEN, we fitted linear regression models on the receptor perturbation gene expression data, using the known receptor perturbations (+1 activated, -1 inhibited, 0 not perturbed) as input [26,27]. Linear models, despite their simplicity, remain among the most widely used model types for gene expression signature-based studies [14,26,28]. We fitted the models for each receptor-regulated gene pair and used the coefficients of the linear regression to define the RIDDEN model (Fig 1A). We use a permutation-based approach to estimate receptor activities in new samples (Methods, Fig 1B).

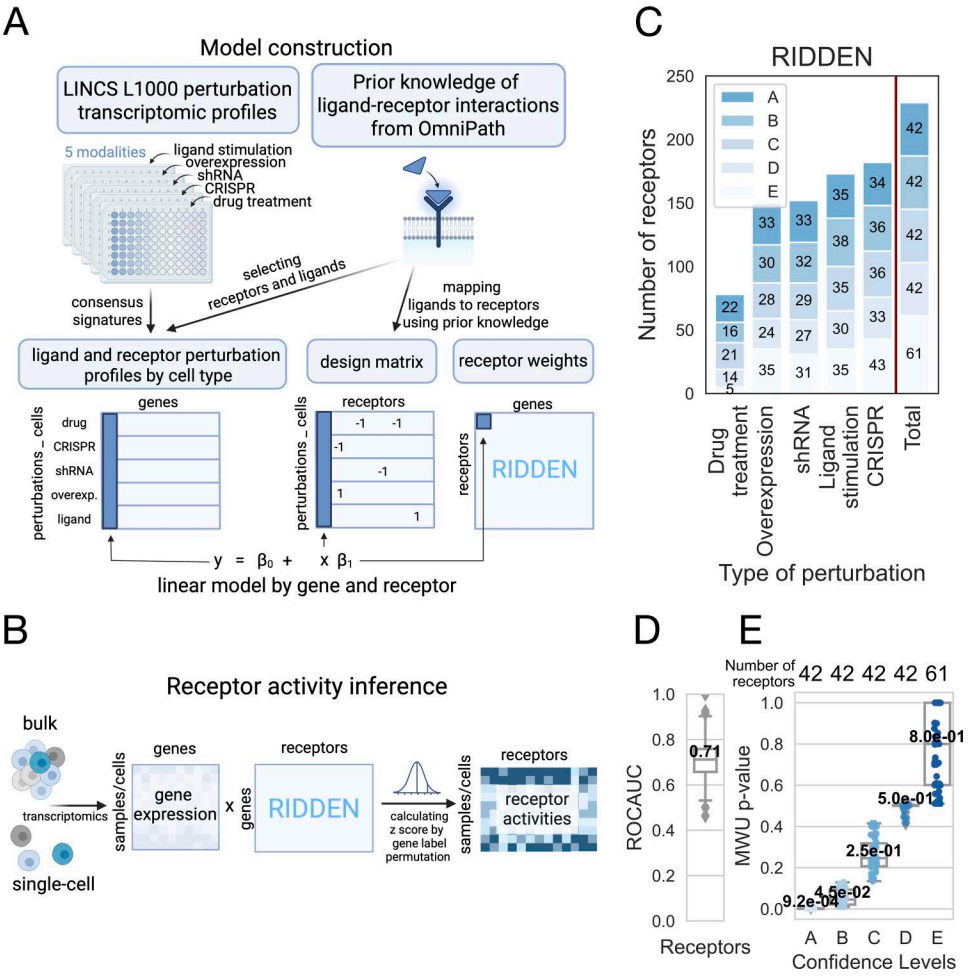

Fig 1. RIDDEN Construction of the model and workflow. (A) RIDDEN model construction. The perturbation profiles are collected from the LINCS L1000 database, and the ligand-receptor interactions are collected from OmniPath. The receptor and ligand perturbation signatures are then filtered from LINCS, and consensus gene expression signatures are calculated. Next, the ligands are mapped to corresponding receptors, known to be interacting according to prior knowledge, and linear models are fitted on the ligand-receptor perturbation profiles using the known receptor perturbations (+1 activated, -1 inhibited, 0 not perturbed) as input to create the RIDDEN. Created in BioRender. (B) Inference of receptor activities using RIDDEN. From bulk and single-cell transcriptomics, RIDDEN estimates receptor activities using dot products and computes z-scores using permutations of gene labels. Created in BioRender. (C) Number of receptors in RIDDEN. Barplots show receptor counts by perturbation types. Each bar shows the number of receptors corresponding to each confidence level. (D) The cross-validation performance was assessed using ROC AUC. The boxplot shows the median ROC AUC of 229 receptors, the first and third quartiles. The minimum and maximum values are shown as whiskers. (E) Receptor confidence levels. The boxplot shows the distribution of Mann-Whitney U (MWU) p-values for receptors across different confidence levels. The p-values were calculated using the mean of the splits applied during the evaluation process. The x-axis represents the confidence levels (A-E), and on the top, the number of receptors in each confidence group is shown. The median p-values are depicted in black, while the first and third quartiles are represented by the box, and the minimum and maximum values are shown as whiskers.

https://doi.org/10.1371/journal.pcbi.1013188.g001

To ensure that our statistical model only includes receptors demonstrating predictive capability, we kept only signatures passing certain criteria. For quality filtering, we first constructed 5 models for the 5 perturbation types (ligand, compound, CRISPR, etc.) separately, and we performed cross-validation to filter for receptors whose activity can be predicted at least in one other modality (Methods). Finally, the RIDDEN summarizes 14463 high-dimensional gene expression profiles for 229 receptors into differential receptor signatures that can be leveraged to infer receptor activities from bulk and single-cell transcriptomics. The distribution of expression values for each gene across all receptor

perturbations is shown in S1 Fig. Next, we conducted an additional cross-validation and calculated the predictive performance for all 229 receptors, for which we used the complete set of their perturbation signatures from every modality (Fig 1C, Methods). This model (Fig 1C, labeled with Total) was then applied for subsequent analyses. The cross-validation resulted in a median ROC AUC score of 0.71. (Fig 1D). Based on the cross-validation performance (Fig 1E, Methods), we further classified receptor signatures into five confidence levels (A-E, where A confidence level receptors had the best cross-validation performance). For further evaluation, we use the RIDDEN model fitted on all perturbation types.

2.2. Model benchmark

To assess the prediction performance of RIDDEN, we used a dataset of bulk cytokine perturbation transcriptomic profiles collected by the authors of CytoSig, and for further evaluation, we used the most recent Immune Dictionary's single-cell transcriptomic profiles in response to many different cytokines as ground truth. RIDDEN was used to predict the cytokine receptor activities of the samples, and we then compared its predictive performance with that of the CytoSig model in predicting cytokine signaling activities. We argued that if the model effectively predicts receptor activity or cytokine signaling activity from gene expression data, it can identify the perturbed receptor or cytokine. This was evaluated using the ROC AUC metric (Fig 2A).

First, to compare RIDDEN with CytoSig, we performed a cross-evaluation by testing each model on the dataset that was used to train the other model. Although RIDDEN was trained on LINCS landmark genes, we evaluated the prediction of CytoSig using both the landmark and the inferred genes in the LINCS signatures. The full RIDDEN yielded 0.61, while the CytoSig 0.59 median ROC AUC. Moreover, the RIDDEN receptor models with confidence levels A-E reached 0.68, 0.64, 0.56, 0.51 and 0.52, respectively (Fig 2B).

Next, we used the single-cell transcriptomic profiles of the Immune Dictionary [15] to determine if the RIDDEN and CytoSig methods could predict which cytokine was perturbed in the sample. We analyzed the cytokines and their receptors that were common in all three datasets. Here, we saw that despite not explicitly modelling ligand perturbations, the RIDDEN had a comparable ROC AUC score of 0.64 with CytoSig (0.67). We then assessed the prediction of cytokine receptors using the entire Immune Dictionary dataset that included all identifiable cytokines. The ROC AUC scores for receptor categories A-E are 0.84, 0.55, 0.61, 0.56, and 0.44, respectively (Fig 2C).

To compare the performance of RIDDEN to other ligand activity prediction methods, we also used NicheNet (NN) to predict ligand activity of Immune Dictionary cytokine perturbation data. The ligand activity predictions of the NN resulted in a median ROC AUC of 0.54, while evaluating 64 overlapping cytokines with Immune Dictionary (S2 Fig). RIDDEN achieved higher ROC AUC score on the three dataset overlap (Fig 2C), which included 28 cytokine receptors and their 58 interacting cytokines, 46 of which also overlapped with NN.

Taken together, RIDDEN achieved predictive performance similar to CytoSig in both the cross-validation setup and the independent perturbation dataset. Notably, for the receptors with the highest confidence, RIDDEN outperformed CytoSig (Fig 2B, C).

2.3. Model evaluation and correlation with biology

Besides benchmarking predictive performance, we also analyzed RIDDEN from a biological perspective. To achieve this, we examined whether receptors that share similar signaling pathways are clustered based on model weights. We also analyzed the relationship of the receptor with transcription factor (TF) activity. Finally, we calculated the predicted receptor activity correlation with baseline receptor and ligand expression of cell lines and patient samples.

First, we investigated the similarity of receptors in the model by hierarchical clustering of receptor weight vectors. We can identify receptor family members clustered together on the dendrogram due to similar downstream signaling (Fig 3A). We performed Gene Ontology biological process [29] enrichment analysis on gene weights of the RIDDEN receptors

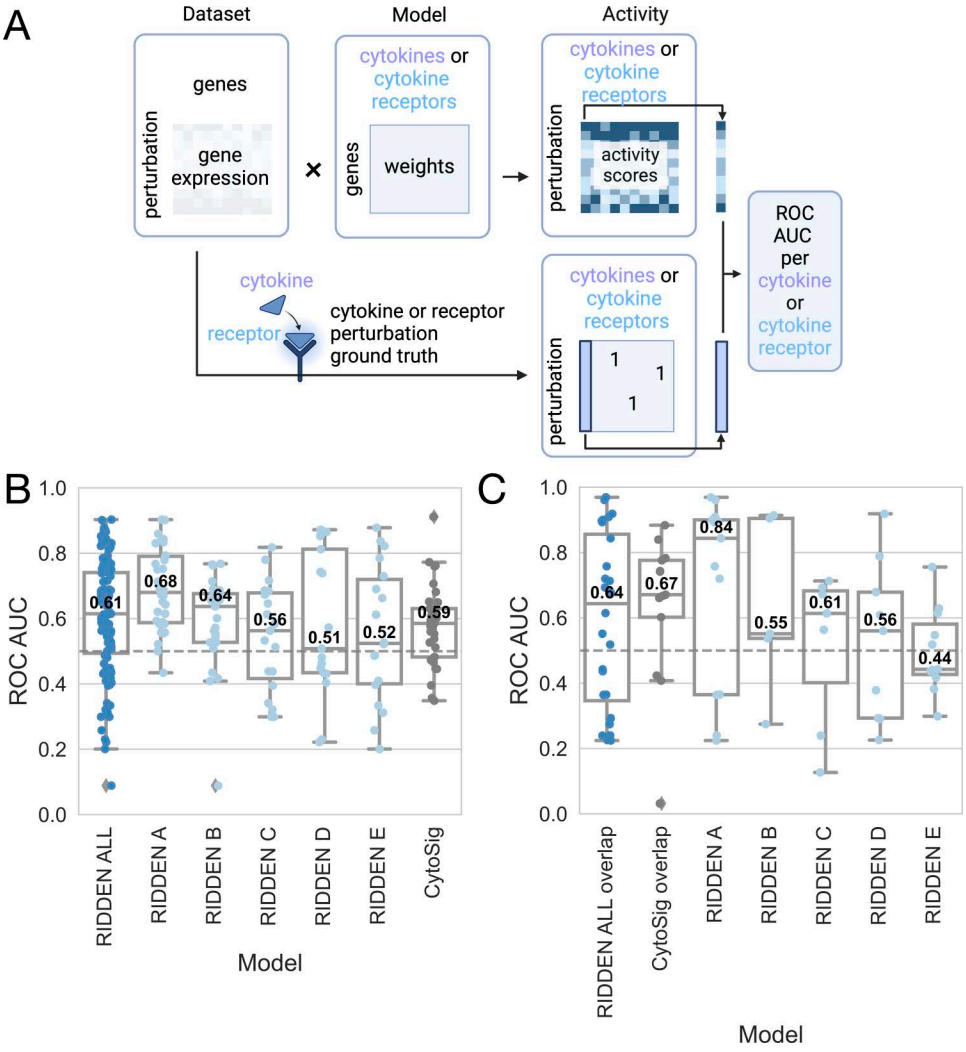

Fig 2. Model benchmark. (A) The receptor activities are inferred from cytokine stimulation and cytokine receptor perturbation gene expression using the CytoSig and RIDDEN models. The ground truth describing the perturbed cytokine or receptor in the sample is used to calculate the ROC AUC per cytokine in the case of CytoSig prediction or cytokine receptor in the case of RIDDEN prediction. Created in BioRender. (B) Comparison of the predictive performance of the RIDDEN and the CytoSig model. The boxplot shows the distribution of ROC AUC values derived from predictions with the CytoSig and RIDDEN models on the other model's training dataset. The RIDDEN model matrix was split into subsets based on confidence scores, and these models were tested against CytoSig. In the case of the CytoSig model, the evaluation dataset is the LINCS cytokine perturbation signatures consisting of landmark genes + inferred genes. The median ROC AUC values (central line and white numbers), first and third quartile (box), minimum and maximum non-outlier values (whiskers) and outliers (diamonds) are shown on the boxplot. (C) Benchmarking of the RIDDEN and the CytoSig models on the Immune Dictionary dataset. The boxplot shows the distribution of ROC AUC values. The first two boxes (RIDDEN ALL and CytoSig ALL overlap) represent the evaluation of models on the overlapping cytokines and their receptors of the three resources. The cytokines of Immune Dictionary were mapped to receptors of RIDDEN, thus resulting in a different number of data points in this comparison. The following boxes show the ROC AUC values of RIDDEN models predicting the activities of all overlapping Immune Dictionary cytokine receptors. These models contain receptors with different confidence scores (A-E). The boxplot features are as described in B.

https://doi.org/10.1371/journal.pcbi.1013188.g002

with distinct functions, to assess whether the model captures biologically meaningful patterns and to validate that genes with high absolute weights reflect relevant downstream processes of each receptor. As some representative examples, the top enriched terms aligned with known receptor functions, such as CXCR4 (C-X-C chemokine receptor type 4, a G protein-coupled chemokine receptor), which is associated with lymphocyte/leukocyte homeostasis [30]; IFNGR1

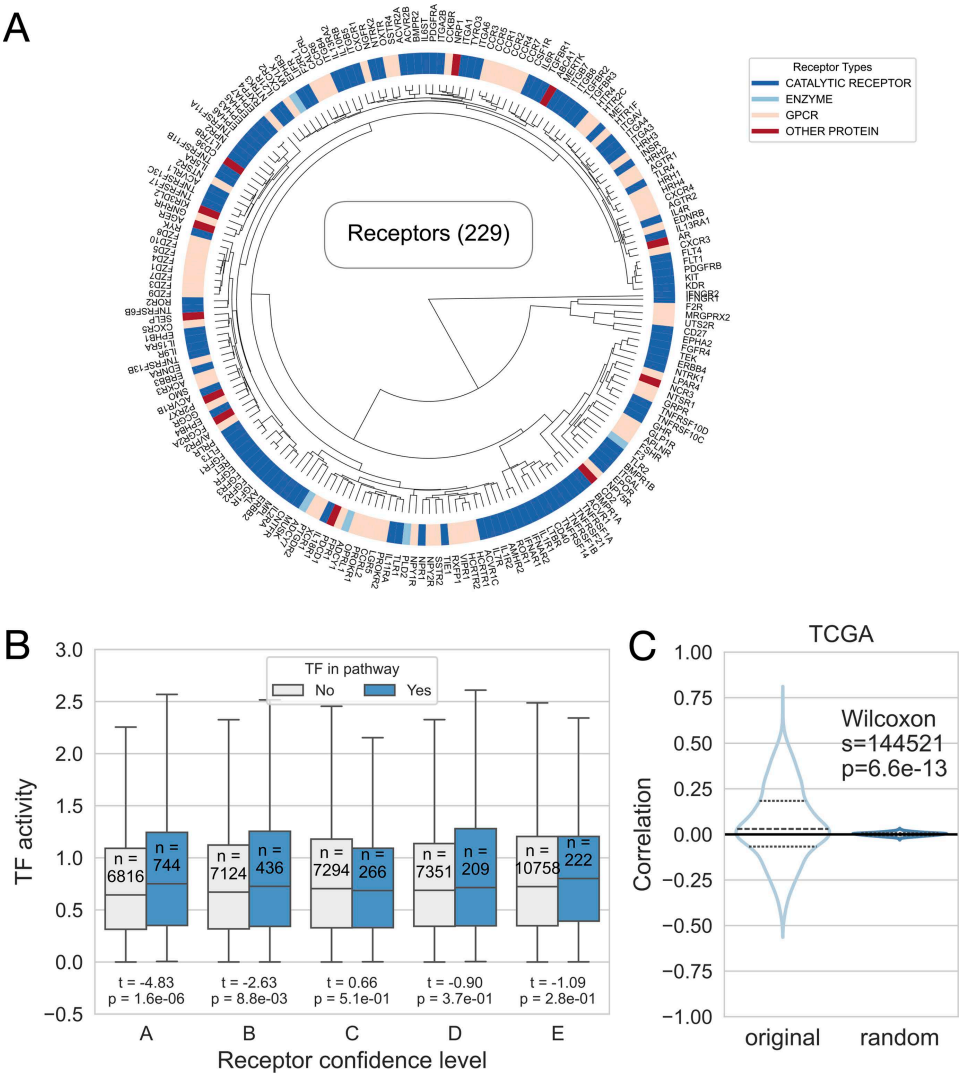

Fig 3. Evaluation of the model. (A) Hierarchical clustering of the receptor weights of the RIDDEN model. The colours indicate different receptor families: catalytic receptors, enzymes, G-protein coupled receptors and other receptors, such as transporters or ion channels, from red to blue. (B) Relationship between TF activities and receptors in the same pathways. The pairs of boxes show the calculated TF activities from the RIDDEN receptor weight vectors when the TF and the receptor have a shared KEGG pathway (light box) or not (dark box) at different confidence levels. Each pair of boxes represents the receptors with different confidence levels. The number of pathways and median TF activities, first and third quartile (box), and minimum and maximum non-outlier values (whiskers) are shown on the boxplot. (C) Correlation analysis of the TCGA samples receptor * ligand TPM values and their receptor activities. The distribution of correlation values in the dataset (original - left) and the random distribution (1000 gene label permutation - right) are represented in a violin plot. A dashed line indicates the quartiles of the correlation values.

https://doi.org/10.1371/journal.pcbi.1013188.g003

(interferon gamma receptor 1, a type II cytokine receptor) has a role in antigen processing and presentation [31]; IL6R (interleukin 6 receptor) os crucial for regulating immune responses as a cytokine receptor [32], and also plays a role in regulating autophagy [33]; TNFRSF1B (tumor necrosis factor receptor superfamily member 1B, a TNF receptor) is linked to T cell activation in immune response [34,35]; and TGFBR1 (transforming growth factor beta receptor 1, a TGF-beta receptor) is essential for all phases of wound healing [36,37], and is directly involved in hemostasis [38] (S1 Table). While enriched terms reflect receptor biology, complete overlap is not expected because methods like RIDDEN capture functional changes based on downstream effects rather than relying on predefined gene set enrichment.

RIDDEN infers receptor activity by capturing the gene expression changes induced by receptor perturbation. However, the gene expression does not necessarily reflect the activity of the proteins, therefore, the receptor-induced gene expression signature is expected to correlate more with the activity of downstream effectors, such as TFs, that directly regulate target genes in response to receptor signaling, rather than with the gene expression levels of those effectors themselves.

Active receptors initiate signaling cascades that modulate the activity of downstream TFs. Thus, we anticipated that the absolute activity of TFs, which are downstream to the active receptors is increased compared to TFs, which are not affected in the signaling pathway. Therefore, we estimated transcription factor activities of receptor model weights using decoupleR [28] with TF regulons from DoRothEA [20]. The RIDDEN model's weights represent the receptor-induced gene expression signature. We compared the activities of transcription factors included or not in the same KEGG signaling pathways [39] as the receptors. We found that TF-receptor pairs sharing KEGG pathways had significantly higher activity than those that do not share KEGG pathways (t-test p-value = 1.6x10-5). We also saw significant differences in receptors with the highest confidence scores, p-values of $1.6x10^{-6}$ for A-level receptors and $8.8x10^{-3}$ for B-level receptors. (Fig 3B). Additionally, we correlated receptor activity with both absolute TF activity and TF gene expression levels calculated from the CytoSig dataset. We found that the difference in correlation distribution between TF–receptor pairs within the same KEGG pathway versus unrelated pairs was substantially greater when using TF activity (t-test p-value = 9.3x10-54) compared to TF gene expression (t-test p-value = 1.5x10-22) (S3 Fig).

To demonstrate the significance of predicting receptor activities, we performed a correlation analysis based on the assumption that, if the ligand and its receptor are upregulated, their increased accessibility indicates the potential for receptor activation. For this analysis, we used patient tumor samples from The Cancer Genome Atlas (TCGA) [40], where cancer cells are bulk sequenced together with their microenvironment thus, paracrine signaling can be assessed. We also performed our analysis on the more homogenous cancer cell line data from the Cancer Cell Line Encyclopedia (CCLE) [41]. We investigated the distribution of Pearson's correlations between receptor activity and the product of receptor and its ligands expression. Although the average correlation was relatively low, the distribution significantly differs from the random distribution with p-value = 6.6x10-13 and 1.7x10-2 for TCGA (Fig 3C) and CCLE (S4 Fig) data, respectively.

To summarize, these findings suggest that the integrated receptor perturbation signatures capture biological processes and that the model can be leveraged for receptor activity prediction.

2.4. RIDDEN identifies biomarkers for cancer therapy response

After evaluating RIDDEN's prediction performance and biological validity, we analyzed its potential to predict patient response to cancer therapy. Here, we used an immune checkpoint therapy dataset, as immune-oncology therapies rely on disturbing the interactions between cancer and immune cells, thus cell-cell communication methods, in this case, can be especially relevant. Immune checkpoint blockade (ICB) therapies have emerged as a promising advancement in cancer treatment. They target immune checkpoint molecules, which are pairs of ligands and receptors responsible for regulating immune responses, such as the PD-1 and PD-L1. PD-1 is expressed on the surface of various immune cells, and activated PD-1 inhibits the anti-cancer activity of these cells. PD-L1 is prominently expressed on the surface of tumor cells and activates PD-1, thus facilitating the immune escape of cancer cells. While ICB is effective in a group of patients, it is important to identify biomarkers to determine which patients can benefit from the therapy [42].

We analyzed the data from a study, where clear renal cell carcinoma patients were treated with either nivolumab, a PD-1 blockade therapy, or everolimus, an mTOR inhibitor [43]. We focused on the pretreatment samples, where gene expression was measured less than one year prior to the initiation of the therapy. We analyzed the associations between PD-1 receptor expression/ PD-L1 ligand expression/ PD-1 receptor activity and patient overall survival (OS). There was no association between OS and PD-1 or PD-L1 gene expression in the nivolumab-treated patient cohort (log-rank test p-value = 0.465, 0.318, respectively) (Fig 4A). Although ligand or receptor mRNA levels were not predictive of patient survival, the RIDDEN-estimated PD-1 receptor activity was associated with the survival of nivolumab-treated

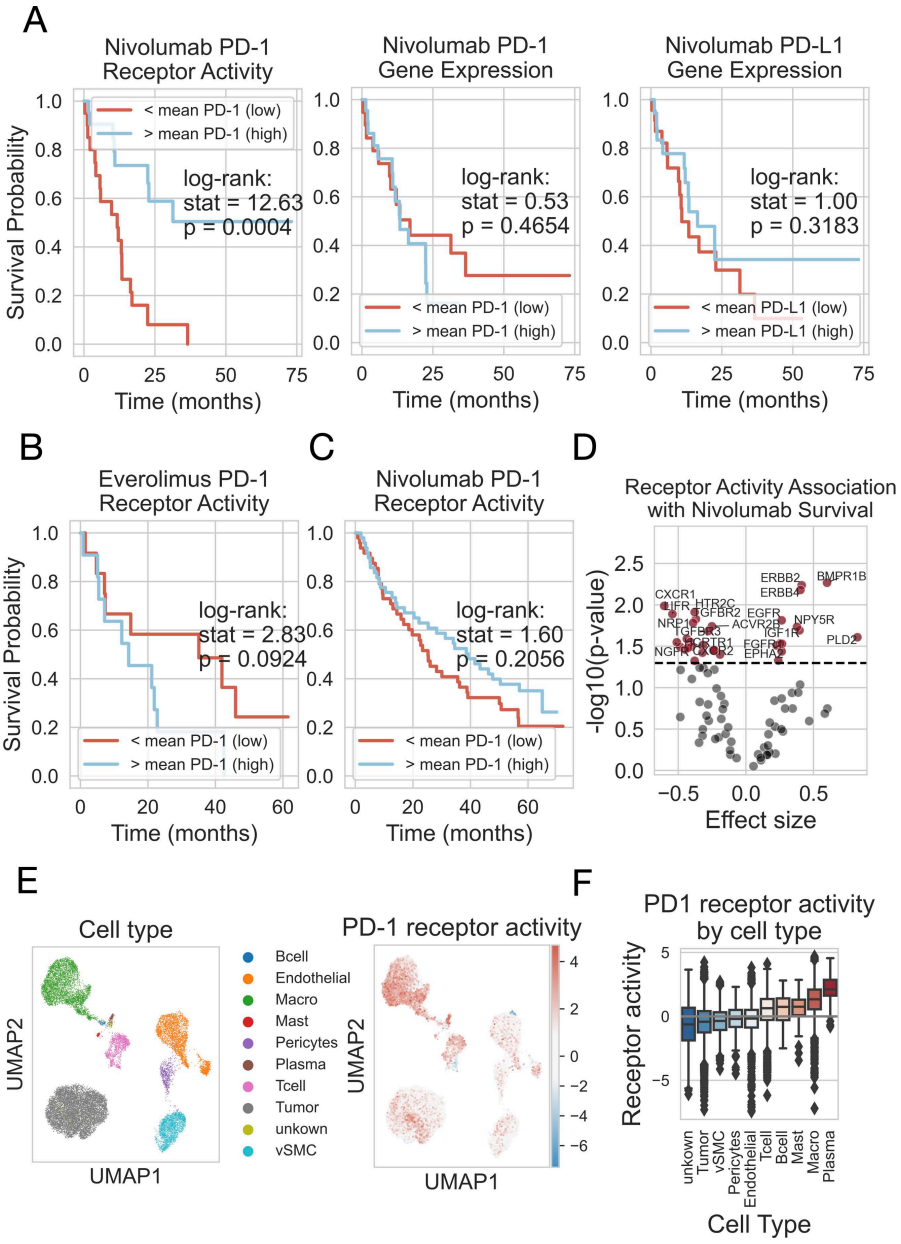

Fig 4. PD-1 expression and activity in tumor samples. (A) PD-1 receptor activity is associated with the survival of patients with renal cell carcinoma, whereas PD-1 and PD-L1 gene expression are not associated. Kaplan-Meier plots represent the overall survival in months (x-axis) and the survival probability (y-axis) of samples with high or low activity of PD-1 receptor (left), expression of PD-1 receptor (middle) and expression of PD-L1 ligand (right), where the mean value threshold is used. The result of the log-rank test is shown. The red line indicates a low; the blue indicates a high expression/activity. (B) The activity of the PD-1 receptor was not associated with patient survival in the case of patients treated with everolimus, an mTOR inhibitor. (C) The activity of the PD-1 receptor is not associated with survival in samples that were taken more than a year before treatment with nivolumab, a PD-1 inhibitor. (D) Association of receptor activity with patient survival in response to nivolumab. The volcano plot shows the Cox regression coefficients (effect size) (x-axis) and the -$\log_{10}$(p-values) (y-axis) for the receptors with A and B confidence (dots). The black dashed line indicates the p-value<0.05 significance level. Receptors with a p-value < 0.03 are shown on the plot. (E) A 2-dimensional representation of the high-dimensional single-cell transcriptomic profile of RCC patients. The UMAP shows the different cell type clusters with different colors (left) and the PD-1 receptor activity of the single cells (right), where high activity is marked by red and low activity is marked by blue. (F) The PD-1 receptor activity of different cell types in samples from 7 patients with renal cell carcinoma. The x-axis represents cell types, while the y-axis represents PD-1 receptor activity. The median activity values (central line), first and third quartile (box), minimum and maximum non-outlier values (whiskers) and outliers (diamonds) are shown on the boxplot.

https://doi.org/10.1371/journal.pcbi.1013188.g004

(p-value = 4*10-4), but not with the everolimus-treated patients (p-value = 9.2*10-2) (Fig 4A, B). The effect of PD-1 receptor activity on the overall hazard in the nivolumab-treated patient cohort was found to be negative (Cox regression β = -0.36, p-value = 0.012) (Table 1). This means that higher PD-1 receptor activity was associated with a decrease in the hazard rate, leading to a longer survival time in patients treated with nivolumab. However, this effect was not significant in patients treated with everolimus, and the association was in the opposite direction (β = 0.33, p-value = 0.106) (Table 1). As nivolumab inhibits PD-L1/ PD-1 interaction, the observed increased effectiveness of the drug in high PD-1 activity corresponds to its mechanism of action. The gene expression of the ligand or the receptor showed no association with survival in everolimus-treated patients (Table 1). Importantly, in the patient cohort, where the samples were taken more than one year before the treatment, the PD-1 receptor activity did not show an association with patient survival in any of the examined cases (Fig 4C). In the samples of nivolumab-treated patients, chemokine receptors (CXCR1, CXCR2), serotonin receptor HTR2C, TGF-β receptors (TGFBR2, TGFBR3), or LIFR exhibited significant negative effects, and BMPR1B, members of the ERBB receptor family, and EGFR had the most significant positive effect on the overall hazard in Cox regression analysis. (Fig 4D and S2 Table). The Cox regression results for gene expression of receptor genes among nivolumab-treated patients and receptor activities in the everolimus-treated patient cohort are available in S3 and S4 Tables, respectively.

Furthermore, we investigated whether RIDDEN receptor activity signatures reflect the potential effects of PD-1 activation in different cell types. For this purpose, we analyzed a dataset containing 7 renal cell carcinoma patients' single-cell gene expression profiles [44]. Each cell shown in the UMAP is clustered by expression after batch correction (Fig 4E). We calculated the PD-1 receptor activities in these cells and showed that the immune cells have high activity of PD-1, while no activity was observed on tumor cells, endothelial, perivascular and smooth muscle cells. (Fig 4E and F).

3. Discussion

RIDDEN (Receptor Activity Data-Driven Inference) is a computational tool developed to infer receptor activities by summarizing thousands of ligand and receptor perturbation gene expression profiles into interpretable receptor activity states. Receptors play a crucial role in both health and disease by initiating and mediating signaling during cell-to-cell communication. By providing insights into the initiators of signaling within cells, RIDDEN enhances our understanding of cellular processes and supports research hypotheses, including assessing the effects of drugs on receptor activity or identifying sources of the abnormal signaling associated with disease phenotypes.

RIDDEN combines an extensive collection of receptor and ligand perturbation transcriptomic profiles [25] and prior knowledge of ligand-receptor interactions [23,24]. The collected dataset enables us to summarize key changes upon receptor activity in different samples. Current state-of-the-art methods in predicting cell communication are typically based on the prior knowledge networks of ligands and receptors. While these methods can give valuable insight into the potential receptor-ligand interactions and communicating cell types, RIDDEN directly infers the receptor level effect of these

Table 1. The effect of PD-1 receptor activity and PD-1 or PD-L1 gene expression on the overall survival of patients diagnosed with renal cell carcinoma determined by Cox regression analysis.

Gene name	Type	Treatment	Effect size	p-value
PD-1	gene expression	nivolumab	−0.0513	0.5849
PD-L1	gene expression	nivolumab	−0.0694	0.6299
PD-1	receptor activity	nivolumab	−0.3622	0.0124
PD-1	gene expression	everolimus	−0.0256	0.8604
PD-L1	gene expression	everolimus	0.1422	0.4779
PD-1	receptor activity	everolimus	0.3320	0.1061

https://doi.org/10.1371/journal.pcbi.1013188.t001

interactions on the receiver cell. It contains high-quality receptors that are validated with models constructed from other modalities. In addition, as a footprint-based method [45], it derives more biological insights than gene set-based methods.

While the RIDDEN model is based on high-throughput measurements, limitations need to be noted. The perturbation profiles consist of 978 landmark genes and do not use the inferred set of more than 10,000 genes [25]. Although whole transcriptome measurements could help distinguish signals between receptors with similar downstream signaling, they are not available at this scale. However, a promising approach has been introduced, which measures single-cell transcriptomic profiles of immune cell types on cytokine perturbation [15], yet there is a need for data on other ligands and receptors that are crucial in intercellular communication.

The increasing number of computational models developed to study the CCC brings attention to the importance of benchmarking and evaluating such models. In computational biology, one of the major challenges is the common lack of ground truth to evaluate predictive models [46]. In terms of receptor activities, cytokine perturbation gene expression data collected by CytoSig [14] or NicheNet [10] can be a reasonable basis with regard to cytokine receptors. The NicheNet model is trained on a small coverage of public experiments, and more than half of its data overlaps with CytoSig. Additionally, the data available in public databases is often specific to a particular experimental design and only measures a pair of cell lines or conditions. The existing studies mainly evaluate the model performance using collected experimental data [9] and bulk or single-cell tissue measurements [10,14,47]. This may change with new studies, such as the Immune Dictionary [15], which relies on *in vivo* data. Furthermore, comparing RIDDEN with classical CCC methods on perturbation datasets would not yield meaningful insights, as these methods rely on endogenous ligand expression and are unable to capture the externally introduced ligands, making them unsuitable for estimating interactions with receptors in perturbation experiments.

The CytoSig model is suitable for benchmarking against our model as it predicts the activity of ligand-stimulated signaling using cytokine perturbation bulk datasets, similar to how RIDDEN employs cytokine receptor perturbation datasets. RIDDEN performed comparably to CytoSig in cytokine signaling or cytokine receptor activity predictions, moreover, RIDDEN provides confidence levels for receptors. In the evaluation of the top-performing receptors, the model outperformed CytoSig, demonstrating its robust performance in predicting cytokine receptor activity. RIDDEN has an advantage over this model due to its broader coverage of perturbation signatures within the same experimental design, encompassing a greater number of receptor and ligand perturbations. This allows the model to capture the core signaling changes induced by receptor activity that occur similarly across different cells.

We compared our model to the published CytoSig model and evaluated its performance using *in vivo* ligand perturbation dataset from the Immune Dictionary. *In vivo* experiments model more complex responses to perturbations as they capture cellular responses within their microenvironment within a tissue. RIDDEN's ability to predict the receptor activities induced by ligand stimulation in a complex environment suggests that by learning from *in vitro* data, RIDDEN can capture the fundamental responses of receptor activation. RIDDEN predicted which receptor-induced signaling was modulated by the immune ligand with performance comparable to CytoSig.

Besides benchmarking the prediction performance, we also analyzed RIDDEN's performance from a biological perspective. First, we show that in the RIDDEN model receptor family members have highly similar signatures consistent with biological expectations because they may form heterodimers during activation, bind the same ligands or have shared effectors. Then, we used two large transcriptomic datasets, TCGA and CCLE, for evaluation and testing of further fundamental biological concepts, such as the correlation of receptor-ligand expression with receptor activity. Another strategy is using pathway gene sets to validate the connection between the activity of transcription factors and the activated receptors that could potentially affect them. Although we observed low correlation in these comparative analyses, we saw that our model performed better than random. We anticipate such a low correlation given the expected disagreement between expression and activity [48]. This observation highlights the importance of predicting activities. The results of these analyses suggest that we can provide a good approximation of these processes.

In addition to physiological processes, the model's applicability extends to examining cell lines carrying a disease phenotype or patient samples. We present a case where the receptor plays a vital role in cancer therapy. Immune checkpoint inhibitors, especially the PD-1 receptor inhibitor, are widely used in cancer therapy [49]. One of the criteria for patients receiving anti-PD-1 treatment is whether they carry high PD-L1 expression measured by immunohistochemical staining of the tumor and immune cells [50]. This suggests if inhibiting the receptor can be effective in the patient. We cannot determine whether the patient will benefit from the therapy based on the gene expression of the PD-L1 ligand or even the PD-1 receptor. However, estimating the activity of receptor proteins can serve as a reliable predictor for the effectiveness of therapy. We have found that the patient's overall survival on anti-PD-1 therapy is associated with increased PD-1 receptor activity before the treatment. Additionally, we have observed associations between patient survival and the activity of several other receptors already connected to immune oncology. These include chemokine receptors that play a crucial role in communication between cells in the tumor microenvironment (TME) [51,52], serotonin receptors that have been reported to impact cancer progression by influencing immunological processes [53], such as promoting immunosuppressive M2-like macrophage polarization [54], or and TGFB receptors whose signaling promotes tumor immune evasion in TME [55], These receptors are activated in the TME where the PD-1 signaling can be enhanced [56].

Importantly, PD-1 activity is associated with increased survival only in the nivolumab-treated cohort and not in the everolimus-treated patients, underlying the specificity of the prediction method.

When investigating the association between mutations of patient samples and patient survival, the time of sample collection is less crucial than examining changes in gene expression or receptor activity and their association with patient survival. In this regard, the timing of sample collection can have a significant impact, as various factors, regulatory mechanisms, and cellular processes can influence gene expression. The activity of the PD-1 receptor in samples taken less than one year before anti-PD-1 therapy initiation is associated with the overall survival of patients, but this association is not observed in samples taken more than 1 year before therapy initiation.

In addition, we have also shown that we can precisely detect the activity of the receptor in the cell types that may be present, such as T cells [57] or tumor-associated macrophages [58]. On the other hand, we do not predict receptor activity in tumor cells due to the lack of a signature induced by the active receptor.

In summary, the RIDDEN method is a reliable and easy-to-use tool for inferring receptor activities across biological contexts, which can be used to obtain a comprehensive overview of active and inactive receptors from transcriptomics data.

4. Methods

4.1. Collection of ligand and receptor perturbation signatures

We queried the OmniPath database [23,24] for curated ligand-receptor interactions using the OmniPath R package to obtain the most reliable collection of possible interactions. From there, we retained all of the receptor and ligand perturbation signatures from the LINCS database from 5 modalities: the genetic (shRNA, CRISPR, overexpression) and the chemical (ligand stimulation, drug treatment) perturbations. For model construction, we used the level 5 LINCS L1000 signatures and landmark gene set. To calculate consensus signatures for all cell and perturbation pairs, we aggregated the perturbation signatures and computed moderated-Z weighted averages for every perturbation in each cell line, following the method outlined in the LINCS publication [25,59].

Finally, the final dataset comprises 14463 perturbation profiles, where each signature corresponds to a receptor perturbation transcriptomic profile in a cell line, referred to as a sample. This dataset includes 229 different perturbed receptors, 228 distinct cell lines, and 747 unique perturbations derived from the different data modalities.

4.2. Construction of the model

After collecting and calculating the consensus gene expression profiles of all possible ligand and receptor perturbations, we constructed linear regression models for the expression of all genes and receptor perturbations.

We used ordinary least squares (OLS) regression models (statsmodels Python package) to estimate the relationship between receptor perturbation and the expression of a gene. In the following linear model equation: $g_i = \beta_0 + \beta_j r_j + \in$, where the predictor r_j is a vector that represents samples (cell-receptor perturbation pairs), where each value denotes the perturbation of receptor *j* in the sample. The values are encoded as follows: 1 denotes stimulation, -1 denotes inhibition, and 0 is no perturbation. If a ligand is perturbed, curated ligand-receptor interactions are used to translate this ligand into corresponding receptor perturbations. The response variable g_i is a vector of gene expression values, with each value representing the expression level of a specific gene in each sample. The β_0 denotes the intercept term, β_j is the coefficient for the receptor *j* (the parameter of interest), and $\in$ is the error term.

Each linear model generated a coefficient (β_j) for each receptor-gene pair, indicating the strength and direction of the receptor perturbation's influence on gene expression. The coefficients from each regression model were organized into a receptor-gene parameter matrix, reflecting the relationship between receptor perturbations and gene expression under different conditions (perturbation type, direction, and cell line). This parameter matrix summarizes the high-dimensional gene expression profile associated with variations in receptor activities.

4.3. Inference of receptor activities from gene expression

To estimate the receptor activity, we calculate the dot product of the gene expression profile of the samples and the RIDDEN matrix (described in 4.2.). This method aligns with a previously described approach, according to pathway activities that can be inferred from the sample's transcriptional profiles [26]. An additional step involves generating a background distribution of receptor activities by performing 1000 permutations of gene labels. Subsequently, we calculate the standardised score (z-score) of the value relative to the background.

4.4. Improving the quality by signature filtering based on cross-validation

We aimed to capture the conserved changes of receptor or ligand perturbations in cell lines by calculating consensus signatures from all cell lines and perturbation pairs in LINCS L1000. However, some consensus signatures may still have poor quality due to insufficient or ineffective genetic perturbation and drug dosage. We employ receptor filtering based on cross-validation. Briefly, we used OLS to fit linear models for each of the five modalities separately, and then, using the parameter matrices of each modality, we inferred receptor activities from the perturbation signatures of the other modalities with the methods described in sections 4.2 and 4.3.

To assess performance, we calculated the ROC AUC values based on receptor activity vectors and the true value vector, which indicates whether the receptor was perturbed in the sample. Inhibiting and activating permutations were handled separately. First, for inhibition perturbation, all the activation perturbations were not considered (set to 0), and the -1 values were reversed to 1. Then, for activation perturbation, all the receptor inhibitions (-1) were not considered in the true vector. The ROC AUC values were calculated separately, resulting in 2 values for receptors with perturbations in both directions. For negative values, we considered 1-ROC AUC. We kept receptors that are predictive in at least one case with a minimum ROC AUC of 0.6, either in the positive or negative direction. Non-predictive receptors were defined as those where the ROC AUC was below the threshold of 0.6 in both the positive and negative directions. We excluded cases where the CRISPR model predicted shRNA perturbation and conversely because if the receptor is not present in the cell, the receptor knock-out or knock-down will give us incorrect information about the potential gene expression change upon receptor perturbation. Finally, after filtering the non-predictive receptors, we fitted OLS linear models using all the remaining signatures, including all modalities. This resulted in a model containing 229 different receptors.

4.5. Assigning confidence levels to receptors

The dataset was randomly split into training and test sets five times, with each set comprising half of the signatures. A model was trained on each training set and then used to predict receptor activity on the corresponding test set. The

Mann-Whitney U (MWU) test was conducted on each split to determine whether there is a significant difference in receptor activities between the perturbed and non-perturbed conditions, based on the predicted receptor activities and the prior knowledge of receptor perturbations. Inhibitory (-1) and stimulatory (1) perturbations were analyzed separately, as was done for ROC AUC calculations in section 4.4. Receptors that had perturbations in both directions had their statistics averaged to obtain a single score for each receptor per split. An overall mean value was calculated for each receptor across splits. ROC AUC values were calculated in the same workflow. Confidence levels for the receptors were assigned based on the aggregated MWU p-values, with confidence levels A-D determined by categorizing p-values into quartiles. Receptors with fewer than 8 signatures in the evaluation were given an E confidence level.

4.6. Benchmarking: comparing performance with the CytoSig model

We compared the performance of CytoSig [14] with the RIDDEN model in two settings, where the prediction of the cytokine receptor activities or cytokine signaling activities are evaluated.

To compare the models directly, we assessed the predictive performance using the ground truth dataset of the other model containing perturbation profiles of cytokines and cytokine receptors. First, we inferred the activities of the cytokine receptors using RIDDEN (with the method described in section 4.3), the receptors that have been perturbed by their ligands (cytokines) in the CytoSig dataset. Then, we used the CytoSig model to calculate the cytokine signaling activities from RIDDEN's cytokine and receptor perturbation signatures, which consist of not only the landmark genes but the inferred genes as well to ensure the comparison is equitable. When evaluating RIDDEN on the CytoSig dataset, we mapped receptors to their interacting cytokines based on the ligand-receptor interaction table from OmniPath [21,22], allowing for both one-to-one and one-to-many interactions. When evaluating CytoSig on the RIDDEN dataset, we mapped cytokines to their corresponding receptors while accounting for the type of perturbation applied, which can be inhibitory or activatory. We evaluated the models' predictive performance using the ROC AUC metric (using the scikit-learn Python package), with separate ROC curves for each receptor. Perturbed ligands or receptors were denoted by 1 in the true values vector, while non-perturbed ones were denoted by 0. We handled different perturbation directions separately (similarly as described in section 4.4.), calculating AUC values for stimulatory and 1-AUC values for inhibitory perturbations. The maximum ROC AUC values were selected for a cytokine with both types of perturbations.

As a ground truth dataset, we used the perturbational single-cell RNA-sequencing profiles of Immune Dictionary [15]. After obtaining the cytokine response profiles from the Immune Dictionary's web portal, the perturbation signatures were normalised using gene-wise z-scores for each immune cell type, and then the cell signatures were aggregated to obtain the average cytokine perturbation signatures of the immune cell types. Signatures containing fewer than ten differentially expressed genes (DEG) were filtered out, as they do not capture the translational changes of the perturbation. The number of DEG of the cytokine perturbations was obtained from the study [15].

We inferred the cytokine signaling activities with CytoSig and the cytokine receptor activities using RIDDEN (as described in section 4.3) of the aggregated cytokine perturbation profiles of immune cell types. We assessed the model performances in inferring the perturbed cytokines. We filtered for the overlapping cytokines and their receptors between the CytoSig, RIDDEN and the dataset and evaluated the performance of the models on this subset using the ROC AUC metric, where the perturbed cytokines indicate the ground truth labels. We used prior knowledge of ligand-receptor interactions to map the RIDDEN receptors to their interacting cytokines. We calculated the ROC AUC for RIDDEN receptors with different confidence scores using all possible cytokine receptors whose ligand was perturbed in the dataset.

4.7. Benchmarking: NicheNet model ligand activity prediction evaluation on immune dictionary

We used NicheNet [10] to calculate ligand activity scores based on the transcriptomic response of cells to cytokine perturbations from the Immune Dictionary [13,15]. We applied a sender-agnostic approach to evaluate all LR interactions where the cognate receptors are expressed in the receiver cells. For each of the cytokine perturbations in the Immune Dictionary,

we calculated the ligand (cytokine) activities for each single-cell type (acting as the receiver). We excluded signatures with fewer than ten DEGs. We assessed the NicheNet model performance using the ROC AUC metric, with the perturbed cytokines as true labels. We used AUPR (Area Under the Precision-Recall Curve) as a metric for ligand activity, previously suggested as the most informative measure for defining ligand activity [10]. In total, we evaluated 64 cytokine perturbations for which activity can be estimated using NicheNet.

4.8. Evaluation of the model assuming activities of transcription factors can be affected by active receptors

We examined whether the model reflects biological processes according to upstream activated/inhibited receptors that can influence the activity of the downstream transcription factors (TFs). To achieve this, we used the decoupleR [28] Python package to estimate TF activities from the dataset using the TF regulons from DoRothEA [20]. We used KEGG pathways [39] to identify TFs and receptors with shared pathways. For each group of receptors, classified by confidence levels, we compared the activities of TFs included or not in the same KEGG pathways as the receptors. We evaluated the difference by Student's t-test.

We further examined whether receptor–TF pairs within the same KEGG signaling pathway exhibit higher correlations than unrelated pairs, and whether this relationship is better captured by TF activity than by TF gene expression. We used TF expression and TF activity profiles from the CytoSig dataset. of 204 TFs. We computed Pearson correlations between RIDDEN-inferred receptor activities and both (1) TF activity scores [20] and (2) TF gene expression levels. We then compared the distribution of correlation values between the within the KEGG pathway and out-of-pathway groups using Student's t-test.

4.9. Receptor clustering

To investigate how receptors with similar mechanisms of action or different receptor families are represented in our model, we used hierarchical clustering and dendrogram algorithms (from the Scipy Python package [60]) on receptor model coefficients to visualise the similarities between receptor vectors. We obtained receptor family classifications from The IUPHAR/BPS Guide to Pharmacology [61].

4.10. Gene Ontology enrichment based on the RIDDEN's receptor weights

We performed Gene Ontology (GO) Biological Process gene set enrichment analysis (GSEA) on absolute gene weight vectors of the receptors using the decoupleR package [28]. P-values were corrected using the false discovery rate method. We show the significantly enriched terms for the selected receptors with the five highest normalized enrichment scores.

4.11. Biological relevance evaluation through correlation analysis using cell line and patient data

We leveraged patients' gene expression profiles of The Cancer Genome Atlas and cell lines' baseline gene expression profiles of the Cancer Cell Line Encyclopedia to demonstrate that the model can predict potentially activated and inhibited receptors in a large cohort of untreated samples. We inferred receptor activities from the gene expression of the samples (Methods 4.3.) We calculated random distribution by gene label permutation performed 100 times. Subsequently, we calculated Pearson's correlation (with the Scipy Python package [60]) between receptor transcript per million (TPM) values multiplied by ligand TPM values and the inferred receptor activities. We compared the random with the original receptor activity distribution using the Wilcoxon test.

4.12. Receptor activity and patient survival association using pretreated samples

We leveraged gene expression of pretreatment samples and the overall survival of patients from publicly available data [43]. We inferred receptor activities. Then, we investigated the relationship between the activities and the patient response

to therapy using the log-rank test and the Cox regression analysis (with the lifelines Python package [62]). In the log-rank test, we separated patient groups by the mean of the gene expression or receptor activity. In the Cox regression, we used the following equation: $h(t|x) = h_0(t) * \exp(\beta * X)$, where $h(t|x)$ denotes the hazard function, $h_0(t)$ is a baseline hazard, β is the coefficient and X denotes the receptor activity or the gene expression.

4.13. Receptor activity in patients' single-cell transcriptomic profiles

To show the cell-wise resolution of receptor activity, we used single-cell transcriptomic profiles of tumor patients' samples from a recent publication [44]. We used only the patients' data with clear cell renal cell carcinoma (ccRCC). Analysis was performed by Scanpy [63]. First, cells containing fewer than 300 genes expressed and genes expressed in fewer than 5 cells were removed from the dataset. Then, potential cell doublets identified by scrublet [64], as well as mitochondrial, ribosomal, and sex genes were removed. Cells with more than 25% mitochondrial content were excluded, and cells where the percentage of ribosomal gene counts exceeded 5% were kept. The total number of counts was normalised to 5000 per cell. Then, the data was log-transformed. Highly variable genes were identified, and batch correction was performed based on highly variable genes using the BBKNN algorithm [65]. Receptor activities were calculated on log-normalized counts before batch correction. We visualised the cells using UMAP 2-D projection. We used the cell line annotations provided in the publication, with an extension of grouping cells that contain different marker genes but are classified into the same cell type, like different endothelial cells.

Supporting information

S1 Fig. Distribution of z-scored expression values for all landmark genes across the receptor perturbations.
(TIFF)

S2 Fig. NicheNet model performance on Immune Dictionary. Boxplot showing ROC AUC values for ligand activity prediction on the Immune Dictionary across 64 cytokine perturbations. The median ROC AUC value (central line and value), first and third quartile (box), minimum and maximum non-outlier values (whiskers) are shown on the boxplot.
(TIFF)

S3 Fig. Pearson correlation between receptor and TF activities or TF expression, informed by whether the pairs intersect with the same KEGG pathway or not. Correlations were calculated using absolute values, and the analysis was limited to TFs with measured expression in CytoSig and TF activity profiles. TF–receptor pairs within the same pathway exhibit significantly higher correlations than unrelated pairs, the difference between within-pathway and out-of-pathway correlations is greater when using TF activity compared to TF gene expression.
(TIFF)

S4 Fig. Correlation analysis of the CCLE samples receptor * ligand TPM values and their receptor activities. The distribution of correlation values in the dataset (original. left) and the random distribution (1000 gene label permutation. right) are represented in a violin plot. A dashed line indicates the quartiles of the correlation values.
(TIFF)

S1 Table. Table of Gene Ontology enrichment analysis for five receptors (CXCR4, IFNGR1, IL6R, TNFRSF1B, TGFBR1), based on the absolute gene weights derived from the RIDDEN model. The top five significantly (FDR < 0.05) enriched biological process terms are listed for each receptor.
(CSV)

S2 Table. Table of Cox regression analysis results of nivolumab-treated patient overall survival and receptor activities with sample collection maximum 1 year before therapy.
(CSV)

S3 Table. Table of Cox regression analysis results of nivolumab-treated patient overall survival and gene expression of receptor genes with sample collection maximum 1 year before therapy.
(CSV)

S4 Table. Table of Cox regression analysis results of everolimus-treated patient overall survival and receptor activities with sample collection maximum 1 year before therapy.
(CSV)

S1 File. Graphical abstract.
(TIFF)

Acknowledgments

The graphical abstract, Figs 1A, B and 2A were created with BioRender.com.

Author contributions

Conceptualization: Szilvia Barsi, Bence Szalai.

Data curation: Szilvia Barsi, Eszter Varga.

Formal analysis: Szilvia Barsi, Eszter Varga.

Funding acquisition: László Hunyady, Bence Szalai.

Investigation: Szilvia Barsi.

Methodology: Szilvia Barsi, Daniel Dimitrov.

Project administration: Bence Szalai.

Software: Szilvia Barsi.

Supervision: Julio Saez-Rodriguez, László Hunyady, Bence Szalai.

Visualization: Szilvia Barsi.

Writing – original draft: Szilvia Barsi.

Writing – review & editing: Szilvia Barsi, Eszter Varga, Daniel Dimitrov, Julio Saez-Rodriguez, László Hunyady, Bence Szalai.

References

1. Li M, Chi X, Wang Y, Setrerrahmane S, Xie W, Xu H. Trends in insulin resistance: insights into mechanisms and therapeutic strategy. Signal Transduct Target Ther. 2022;7(1):216. https://doi.org/10.1038/s41392-022-01073-0 PMID: 35794109
2. Xu Y, Yan J, Zhou P, Li J, Gao H, Xia Y, et al. Neurotransmitter receptors and cognitive dysfunction in Alzheimer's disease and Parkinson's disease. Prog Neurobiol. 2012;97(1):1–13. https://doi.org/10.1016/j.pneurobio.2012.02.002 PMID: 22387368
3. Schöneberg T, Liebscher I. Mutations in G Protein-Coupled Receptors: Mechanisms, Pathophysiology and Potential Therapeutic Approaches. Pharmacol Rev. 2021;73(1):89–119. https://doi.org/10.1124/pharmrev.120.000011 PMID: 33219147
4. Dricu A. Oncogenic Signalling of Growth Factor Receptors in Cancer: Mechanisms and Therapeutic Opportunities. Int J Mol Sci. 2022;23(13):7376. https://doi.org/10.3390/ijms23137376 PMID: 35806381
5. Armingol E, Officer A, Harismendy O, Lewis NE. Deciphering cell-cell interactions and communication from gene expression. Nat Rev Genet. 2021;22(2):71–88. https://doi.org/10.1038/s41576-020-00292-x PMID: 33168968

6. Armingol E, Baghdassarian HM, Lewis NE. The diversification of methods for studying cell-cell interactions and communication. Nat Rev Genet. 2024;25(6):381–400. https://doi.org/10.1038/s41576-023-00685-8 PMID: 38238518
7. Almet AA, Cang Z, Jin S, Nie Q. The landscape of cell-cell communication through single-cell transcriptomics. Curr Opin Syst Biol. 2021;26:12–23. https://doi.org/10.1016/j.coisb.2021.03.007 PMID: 33969247
8. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat Protoc. 2020;15(4):1484–506. https://doi.org/10.1038/s41596-020-0292-x PMID: 32103204
9. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH. Nat Commun. 2021;12:1088. https://doi.org/10.1038/s41467-021-21390-0
10. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. Nat Methods. 2020;17(2):159–62. https://doi.org/10.1038/s41592-019-0667-5 PMID: 31819264
11. Noël F, Massenet-Regad L, Carmi-Levy I, Cappuccio A, Grandclaudon M, Trichot C, et al. Dissection of intercellular communication using the transcriptome-based framework ICELLNET. Nat Commun. 2021;12(1):1089. https://doi.org/10.1038/s41467-021-21244-x PMID: 33597528
12. Dimitrov D, Türei D, Garrido-Rodriguez M, Burmedi PL, Nagai JS, Boys C, et al. Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data. Nat Commun. 2022;13(1):3224. https://doi.org/10.1038/s41467-022-30755-0 PMID: 35680885
13. Dimitrov D, Schäfer PSL, Farr E, Rodriguez-Mier P, Lobentanzer S, Badia-I-Mompel P, et al. LIANA+ provides an all-in-one framework for cell-cell communication inference. Nat Cell Biol. 2024;26(9):1613–22. https://doi.org/10.1038/s41556-024-01469-w PMID: 39223377
14. Jiang P, Zhang Y, Ru B, Yang Y, Vu T, Paul R, et al. Systematic investigation of cytokine signaling activity at the tissue and single-cell levels. Nat Methods. 2021;18(10):1181–91. https://doi.org/10.1038/s41592-021-01274-5 PMID: 34594031
15. Cui A, Huang T, Li S, Ma A, Pérez JL, Sander C, et al. Dictionary of immune responses to cytokines at single-cell resolution. Nature. 2024;625(7994):377–84. https://doi.org/10.1038/s41586-023-06816-9 PMID: 38057668
16. Innes BT, Bader GD. Transcriptional signatures of cell-cell interactions are dependent on cellular context. bioRxiv. 2021. p. 2021.09.06.459134. https://doi.org/10.1101/2021.09.06.459134
17. D'Agostino N, Li W, Wang D. High-throughput transcriptomics. Sci Rep. 2022;12(1):20313. https://doi.org/10.1038/s41598-022-23985-1 PMID: 36446824
18. Szalai B, Saez-Rodriguez J. Why do pathway methods work better than they should?. FEBS Lett. 2020;594(24):4189–200. https://doi.org/10.1002/1873-3468.14011 PMID: 33270910
19. Dugourd A, Saez-Rodriguez J. Footprint-based functional analysis of multi-omic data. Current Opinion in Systems Biology. 2019. Available from: https://www.sciencedirect.com/science/article/pii/S2452310019300149
20. Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. 2019;29(8):1363–75. https://doi.org/10.1101/gr.240663.118 PMID: 31340985
21. Müller-Dott S, Tsirvouli E, Vazquez M, Ramirez Flores RO, Badia-I-Mompel P, Fallegger R, et al. Expanding the coverage of regulons from high-confidence prior knowledge for accurate estimation of transcription factor activities. Nucleic Acids Res. 2023;51(20):10934–49. https://doi.org/10.1093/nar/gkad841 PMID: 37843125
22. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res. 2013;41(Database issue):D991-5. https://doi.org/10.1093/nar/gks1193 PMID: 23193258
23. Türei D, Korcsmáros T, Saez-Rodriguez J. OmniPath: guidelines and gateway for literature-curated signaling pathway resources. Nat Methods. 2016;13(12):966–7. https://doi.org/10.1038/nmeth.4077 PMID: 27898060
24. Türei D, Valdeolivas A, Gul L, Palacio-Escat N, Klein M, Ivanova O, et al. Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. Mol Syst Biol. 2021;17(3):e9923. https://doi.org/10.15252/msb.20209923 PMID: 33749993
25. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell. 2017;171:1437-1452.e17.
26. Schubert M, Klinger B, Klünemann M, Sieber A, Uhlitz F, Sauer S, et al. Perturbation-response genes reveal signaling footprints in cancer gene expression. Nat Commun. 2018;9(1):20. https://doi.org/10.1038/s41467-017-02391-6 PMID: 29295995
27. Holland CH, Szalai B, Saez-Rodriguez J. Transfer of regulatory knowledge from human to mouse for functional genomics analysis. Biochim Biophys Acta Gene Regul Mech. 2019;194431.
28. Badia-I-Mompel P, Vélez Santiago J, Braunger J, Geiss C, Dimitrov D, Müller-Dott S, et al. decoupleR: ensemble of computational methods to infer biological activities from omics data. Bioinform Adv. 2022;2(1):vbac016. https://doi.org/10.1093/bioadv/vbac016 PMID: 36699385
29. Carbon S, Ireland A, Mungall CJ, Shu S, Marshall B, Lewis S, et al. AmiGO: online access to ontology and annotation data. Bioinformatics. 2009;25(2):288–9. https://doi.org/10.1093/bioinformatics/btn615 PMID: 19033274
30. Bianchi ME, Mezzapelle R. The chemokine receptor CXCR4 in cell proliferation and tissue regeneration. Front Immunol. 2020;11:2109.
31. Zhou F. Molecular mechanisms of IFN-gamma to up-regulate MHC class I antigen processing and presentation. Int Rev Immunol. 2009;28(3–4):239–60. https://doi.org/10.1080/08830180902978120 PMID: 19811323
32. Wolf J, Rose-John S, Garbers C. Interleukin-6 and its receptors: a highly regulated and dynamic system. Cytokine. 2014;70: 11–20.

33. Hu F, Song D, Yan Y, Huang C, Shen C, Lan J, et al. IL-6 regulates autophagy and chemotherapy resistance by promoting BECN1 phosphorylation. Nat Commun. 2021;12(1):3651. https://doi.org/10.1038/s41467-021-23923-1 PMID: 34131122
34. Croft M. The role of TNF superfamily members in T-cell function and diseases. Nat Rev Immunol. 2009;9:271–85.
35. Alam MS, Otsuka S, Wong N, Abbasi A, Gaida MM, Fan Y. TNF plays a crucial role in inflammation by signaling via T cell TNFR2. Proceedings of the National Academy of Sciences of the United States of America. 2021;118:e2109972118.
36. Penn JW, Grobbelaar AO, Rolfe KJ. The role of the TGF-β family in wound healing, burns and scarring: a review. Int J Burns Trauma. 2012;2:18–28.
37. Ramirez H, Patel SB, Pastar I. The Role of TGFβ Signaling in Wound Epithelialization. Adv Wound Care (New Rochelle). 2014;3(7):482–91. https://doi.org/10.1089/wound.2013.0466 PMID: 25032068
38. Jablonska E, Markart P, Zakrzewicz D, Preissner KT, Wygrecka M. Transforming growth factor-β1 induces expression of human coagulation factor XII via Smad3 and JNK signaling pathways in human lung fibroblasts. J Biol Chem. 2010;285(15):11638–51. https://doi.org/10.1074/jbc.M109.045963 PMID: 20142324
39. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30. https://doi.org/10.1093/nar/28.1.27 PMID: 10592173
40. Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, et al.; Cancer Genome Atlas Research Network. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet. 2013;45(10):1113–20. https://doi.org/10.1038/ng.2764 PMID: 24071849
41. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S. The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature. 2012;483:603–7.
42. Ilie M, Hofman V, Dietel M, Soria J-C, Hofman P. Assessment of the PD-L1 status by immunohistochemistry: challenges and perspectives for therapeutic strategies in lung cancer patients. Virchows Arch. 2016;468(5):511–25. https://doi.org/10.1007/s00428-016-1910-4 PMID: 26915032
43. Braun DA, Hou Y, Bakouny Z, Ficial M, Sant' Angelo M, Forman J, et al. Interplay of somatic alterations and immune infiltration modulates response to PD-1 blockade in advanced clear cell renal cell carcinoma. Nat Med. 2020;26(6):909–18. https://doi.org/10.1038/s41591-020-0839-y PMID: 32472114
44. Zhang Y, Narayanan SP, Mannan R, Raskind G, Wang X, Vats P, et al. Single-cell analyses of renal cell cancers reveal insights into tumor microenvironment, cell of origin, and therapy response. Proc Natl Acad Sci U S A. 2021;118(24):e2103240118. https://doi.org/10.1073/pnas.2103240118 PMID: 34099557
45. Holland CH, Tanevski J, Perales-Patón J, Gleixner J, Kumar MP, Mereu E, et al. Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data. Genome Biol. 2020;21(1):36. https://doi.org/10.1186/s13059-020-1949-z PMID: 32051003
46. Jin S, Ramos R. Computational exploration of cellular communication in skin from emerging single-cell and spatial transcriptomic data. Biochem Soc Trans. 2022;50(1):297–308. https://doi.org/10.1042/BST20210863 PMID: 35191953
47. Cabello-Aguilar S, Alame M, Kon-Sun-Tack F, Fau C, Lacroix M, Colinge J. SingleCellSignalR: inference of intercellular networks from single-cell transcriptomics. Nucleic Acids Res. 2020;48(10):e55. https://doi.org/10.1093/nar/gkaa183 PMID: 32196115
48. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 2012;13(4):227–32. https://doi.org/10.1038/nrg3185 PMID: 22411467
49. Ribas A, Wolchok JD. Cancer immunotherapy using checkpoint blockade. Science. 2018;359(6382):1350–5. https://doi.org/10.1126/science.aar4060 PMID: 29567705
50. Akhtar M, Rashid S, Al-Bozom IA. PD-L1 immunostaining: what pathologists need to know. Diagn Pathol. 2021;16(1):94. https://doi.org/10.1186/s13000-021-01151-x PMID: 34689789
51. Burger JA, Kipps TJ. CXCR4: a key receptor in the crosstalk between tumor cells and their microenvironment. Blood. 2006;107(5):1761–7. https://doi.org/10.1182/blood-2005-08-3182 PMID: 16269611
52. Qin R, Ren W, Ya G, Wang B, He J, Ren S, et al. Role of chemokines in the crosstalk between tumor and tumor-associated macrophages. Clin Exp Med. 2023;23(5):1359–73. https://doi.org/10.1007/s10238-022-00888-z PMID: 36173487
53. Karmakar S, Lal G. Role of serotonin receptor signaling in cancer cells and anti-tumor immunity. Theranostics. 2021;11(11):5296–312. https://doi.org/10.7150/thno.55986 PMID: 33859748
54. de las Casas-Engel M, Domínguez-Soto A, Sierra-Filardi E, Bragado R, Nieto C, Puig-Kroger A. Serotonin skews human macrophage polarization through HTR2B and HTR7. J Immunol. 2013;190:2301–10.
55. Yi M, Li T, Niu M, Wu Y, Zhao Z, Wu K. TGF-β: A novel predictor and target for anti-PD-1/PD-L1 therapy. Front Immunol. 2022;13:1061394. https://doi.org/10.3389/fimmu.2022.1061394 PMID: 36601124
56. Chen S, Crabill GA, Pritchard TS, McMiller TL, Wei P, Pardoll DM, et al. Mechanisms regulating PD-L1 expression on tumor and immune cells. J Immunother Cancer. 2019;7(1):305. https://doi.org/10.1186/s40425-019-0770-2 PMID: 31730010
57. Arasanz H, Gato-Cañas M, Zuazo M, Ibañez-Vea M, Breckpot K, Kochan G, et al. PD1 signal transduction pathways in T cells. Oncotarget. 2017;8(31):51936–45. https://doi.org/10.18632/oncotarget.17232 PMID: 28881701
58. Gordon SR, Maute RL, Dulken BW, Hutter G, George BM, McCracken MN, et al. PD-1 expression by tumour-associated macrophages inhibits phagocytosis and tumour immunity. Nature. 2017;545(7655):495–9. https://doi.org/10.1038/nature22396 PMID: 28514441

59. Szalai B, Subramanian V, Holland CH, Alföldi R, Puskás LG, Saez-Rodriguez J. Signatures of cell death and proliferation in perturbation transcriptomics data-from confounding factor to effective prediction. Nucleic Acids Res. 2019;47(19):10010–26. https://doi.org/10.1093/nar/gkz805 PMID: 31552418

60. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. 2020;17(3):261–72. https://doi.org/10.1038/s41592-019-0686-2 PMID: 32015543

61. Harding SD, Armstrong JF, Faccenda E, Southan C, Alexander SPH, Davenport AP. The IUPHAR/BPS Guide to PHARMACOLOGY in 2024. Nucleic Acids Res. 2024;52:D1438–49.

62. Davidson-Pilon C. lifelines: survival analysis in Python. JOSS. 2019;4(40):1317. https://doi.org/10.21105/joss.01317

63. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19(1):15. https://doi.org/10.1186/s13059-017-1382-0 PMID: 29409532

64. Wolock SL, Lopez R, Klein AM. Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst. 2019;8(4):281-291.e9. https://doi.org/10.1016/j.cels.2018.11.005 PMID: 30954476

65. Polański K, Young MD, Miao Z, Meyer KB, Teichmann SA, Park J-E. BBKNN: fast batch alignment of single cell transcriptomes. Bioinformatics. 2020;36(3):964–5. https://doi.org/10.1093/bioinformatics/btz625 PMID: 31400197