Deconvolving specific from non-specific effects in differential gene expression experiments with Deep Generative Networks

Panagiotis Papasaikas[1,2], Dimos Gaidatzis[1,2], Charlotte Soneson[1,2] and Michael B. Stadler[1,2,3].*
[1] Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland.
[2] SIB Swiss Institute of Bioinformatics, 4058 Basel, Switzerland.
[3] University of Basel, 4058 Basel, Switzerland.
* Corresponding author. Email: panagiotis.papasaikas@fmi.ch;
ORCID: P.P.: 0000-0002-1640-7636; D.G: 0000-xxxx-xxxx-xxxx; C.S: 0000-0003-3833-2169; M.B.S.: 0000-0002-2269-4934.

Abstract

Understanding the mechanisms of action upon cellular perturbations is a fundamental endeavor in molecular and chemical biology. Differential expression analysis is a widely used approach for probing these mechanisms, yet it presents substantial interpretational challenges due to the presence of secondary effects and the complex impact of experimental treatments on gene expression. To address this, we introduce orthos, an approach that employs Deep Generative Networks to disentangle specific and non-specific effects of perturbations on gene expression. Trained on large collections of human and mouse gene expression contrasts compiled for this work, orthos isolates non-specific effects by learning the patterns of expression changes that manifest time and again in unrelated experiments. We demonstrate, in diverse experimental settings, that the specific component obtained from the decomposition is a more informative and robust experimental signature and a better proxy for the direct molecular effects of a treatment compared to the original contrast, thereby drastically enhancing the interpretability of differential expression results. In addition, orthos allows identification of experiments with similar specific effects, aiding in the mapping of new treatments to their mechanisms of action. In summary, orthos constitutes a novel strategy in the analysis and interpretation of gene expression data and offers a powerful platform for the study of genetic, physiological, and pharmacological treatments in basic and applied research.

Introduction

The elucidation of targets and mechanisms of action of cellular interventions is key in biology with ramifications extending from basic research on cell regulatory networks, to the study of molecular causes of diseases and drug discovery[1-3]. Differential Gene Expression (DGE) analysis of transcriptomic alterations, an eminent, accessible readout of cellular response to perturbations, frequently serves as the principal, and occasionally the sole, strategy for probing these questions.

Typically, the statistical pipelines employed for DGE analyses yield extensive lists of differentially expressed genes. These lists are difficult to interpret for a number of reasons. First, they are confounded by the multilayered repercussions of experimental treatments on mRNA production. For instance, modulating a transcription factor not only alters transcription of its direct targets but also triggers a cascade of up- or down-regulation of genes within the same or closely interlinked regulatory networks and pathways. These changes, in turn, can set off further responses downstream that ripple throughout the transcriptomic landscape potentially impacting core cellular processes[1, 2]. Second, genetic intervention technologies, treatment delivery agents and solvents often elicit a variety of unintended, systemic responses (such as immune, toxic, metabolic) that cannot be well-controlled for by the design of the study[4-11]. Finally, the contrasted RNA sequencing libraries can suffer from insidious technical systematic biases of varying degrees that impinge on DGE analyses and can be hard to detect and even harder to rectify[12-15]. The final experimentally measured gene expression changes are, therefore, a convolution of the specific effects and the non-specific secondary, lateral treatment and technical distortion effects. Crucially, these non-specific effects, by their nature, reverberate across DGE contrasts, spanning labs, projects, and treatments, opening a potential avenue for their detection and removal. To this end, we developed “orthos”, a Deep Generative Network (DGN) approach, that leverages information from a comprehensive collection of past experiments to effectively disentangle specific and non-specific effects on gene expression.

DGNs have swiftly taken root in the field of transcriptomics, on account of their capacity to accurately learn complex data-generating distributions characterized by high dimensionality and intricate non-

linear dependencies from noisy and biased samples. Applications run the gamut from dimensionality reduction, visualization, denoising and imputation to data harmonization, automated annotation, cell-type composition studies, gene network inference and the development of interpretable biological models[16-19]. These networks are typically trained on gene expression profiles, are tailored with different types and degrees of inductive bias, contingent on the application, and take advantage of the learned distribution to perform the various inference tasks. Our approach, while grounded in similar foundations, signifies a conceptual shift. We posit that, when operating on extensive collections of diverse expression contrasts, the learned generative distribution encapsulates regularities of the input signal that might not be of primary biological interest, namely the non-specific perturbation effects.

Starting from uniformly processed publicly available RNAseq experiments, we compiled large corpora of gene expression contrasts for human and mouse (~130K annotated, ~1M augmented) and trained organism-specific conditional variational models that learn and isolate non-specific effects that pervade across multiple treatments while accounting for tissue or cell line experimental context (Fig. 1a). We use these models to show, in multiple experimental settings, that the residual component derived from the removal of these effects is a highly discriminative and robust experimental signature that is more closely related to the direct molecular effects of the perturbation compared to the original contrast. We also show that, when utilized in the context of large screens, the models can be further fine-tuned to better account for salient within-study non-specific effects.

Beyond affording a more nuanced understanding of the effects of experimental treatments on gene expression, orthos offers a platform for researchers to query the contrast database with arbitrary DGE profiles and identify experiments with similar specific effects, highlighting its utility in mapping pharmacological, physiological, or genetic treatments to mechanisms of action.

Results

Training a model that deconvolves specific from non-specific experimental effects

We set out to design and train models able to decompose the variance of a given DGE experiment into a non-specific and an experiment-specific component. Our model architecture is based on the conditional variational autoencoders (cVAEs) architecture, that has been used successfully for transcriptomic analyses[16, 18]. The models encode DGE contrasts (gene expression log-fold-changes), conditioned on the context of the performed experiment (overall gene expression profile), to a concise latent representation (z_D) which retains their recurring and therefore compressible traits (see Methods). The compressed latent representation is then used to reconstruct a decoded version of the contrast. The decoded output subsumes gene variance that the model can account for because it has been repeatedly encountered during training. The residual obtained after removing the decoded contrast from the original one encompasses the gene variance that the model cannot account for, namely experiment-specific biological effects and experimental noise.

For training purposes, we compiled organism-specific human and mouse DGE contrast corpora, calculated from the ARCHS4 database of uniformly processed publicly available RNAseq datasets[20] using a combination of metadata semantic and quantitative analysis to determine the proper assignment of contrasted conditions (see Methods). After post-filtering, the derived contrast database is comprised of 74,731 human and 58,532 mouse contrasts from 7,866 and 7,474 Gene Expression Omnibus (GEO) studies respectively, spanning a variety of technologies, sequencing platforms, cellular and tissue contexts (Fig. S1).

Quantification of the contrast correlations between experimental series in the compiled contrast database confirmed that most contrasts display significant associations with unrelated experiments (Fig. 1b,c). Gene Set Enrichment Analysis (GSEA) of the compiled contrasts revealed that at least part of this redundancy can be attributed to commonly affected key cellular processes, with the same pathways being strongly (p-value < 1e-6) affected in large fractions (>5%) of the contrasts (Fig. 1d, Supplementary Fig. 2 and Supplementary table 1). The commonly affected gene sets mainly

correspond to core processes related to cell-cycle, immune response and metabolism, transcriptional, epigenetic and translational regulation as well as response to toxicity. These results are in line with our assertion that, irrespective of the specific intervention, transcriptional effects often converge to similar secondary effects that are largely non-informative with respect to the instigating mechanism of action.

In order to enhance generalization performance, we pretrained the cVAEs on an augmented collection of synthesized contrasts (over 500,000 contrasts per organism) and subsequently fine-tuned them to the actual contrast database (see Methods). The final orthos models retain on average 40% of the input contrast variance in the decoded output (median pearson's rho=0.68, Fig. 2b). As anticipated, the decoded contrasts show significant across-series correlations, whereas such correlations are mostly absent in the residuals, indicating that the model has effectively removed recurrent signals (Fig. 1b,c).

At the same time, analysis of independent biological replicates, extracted from the contrast database, demonstrates that reproducible biological variance is preserved in the residual fraction (Fig. 2c, 2d). On average ~30% of the residual variance (corresponding to ~27% of the input variance) is shared between replicates and corresponds to the non-random experiment-specific effects. The amount of residual reproducible variance is strongly correlated to the input reproducible variance indicating that it is largely a function of the experimental signal to noise ratio (Fig. 2c, 2d).

GSEA analysis of the decoded and residual contrasts confirmed that the variance that corresponds to commonly affected cellular processes winds-up almost exclusively in the decoded fraction while it has been expunged from the residual fraction (Fig. 1d, Fig Sx).

Overall, these results demonstrate that orthos efficiently deconvolves DGE signals to a decoded fraction that encompasses recurrent, predominantly non-specific effects and a residual fraction comprised partially of reproducible specific biological effects and partially of experimental noise.

The residual signal is a more informative experimental signature and a better proxy for specific treatment effects compared to the raw contrast

We then sought to systematically assess the input and orthos-decomposed contrasts in terms of their specificity with respect to the applied treatment. We searched the contrast database for pairs of matching experiments, originating from different studies, that modulate the same gene targets either by gene inactivation or by overexpression. We calculated the pairwise similarity of those same-target, same-modulation-direction experiments when using their input or decomposed contrasts. We note that each experiment of a matched pair is typically performed in a different cell context and often using a different gene modulation technology (e.g. CRISPR vs RNAi for inactivation or different vectors for overexpression). This experimental variability adds a layer of complexity as it can introduce distinct secondary effects that can confound comparisons. We evaluated the computed pairwise similarity values of matching experiments for each contrast type relative to a corresponding background of random across-study experiment pairs. We find that the residual fraction is consistently more informative than the input contrast for identifying treatments that modulate the same molecular targets (Fig. 3a, 3b left panels). Conversely, the decoded fraction offers invariably lower separability compared to the input contrast for same-target experiment identification (Fig. 3a, 3b right panels).

To further probe the capacity of the input and decomposed contrasts to delineate treatments we turned to a high-throughput single-cell chemical screen that evaluated the transcriptomic response to hundreds of compounds at different doses and in different cancer cell lines[21]. We assessed the pairwise similarity of the DGE profiles of treatments, in pseudobulks, in two different cancer cell line contexts. Consistent with the results presented in the original study, we see clustering of the DGE profiles with respect to the molecular pathway targeted by the compound (Fig 3c, Input). At the same time, we observe extensive similarity in the effects of different compounds both within and across the groupings of targeted pathways. This suggests widespread overlaps in the secondary effects of distinct drugs, which is in line with the observation of the study authors that genes related to cell proliferation and cell-cycle arrest are ubiquitously affected across the screen. As a result, the identifiability of individual compounds solely based on their transcriptomic effects is poor. This is evidenced in our comparison of matching treatments across the two cell line contexts, where pairwise similarity of the corresponding DGE profiles is in many cases close to background levels (Fig. 3c, Fig. 3d). As anticipated, when using the decoded

fractions of the DGE profiles for pairwise comparisons, our ability to identify matching compound treatments across contexts is even more limited (Fig. 3c, Decoded, Fig. Sx). Importantly, compound identifiability is considerably improved in the residual fraction comparisons (Fig. 3c, Fig. 3d), emphasizing once more that the residual contrast constitutes a more distinctive probe for treatment-specific effects.

We further took advantage of the same study to assess the potential advantage conferred by fine-tuning the models to better capture pervasive unspecific effects present in large screens. The DGE profiles of the study are derived from single-cell transcriptomic libraries that feature particular technical biases. Crucially, all single-cell studies were excluded from the original contrast database and model training. In addition, as noted earlier, the study screen is afflicted with recurrent secondary biological effects specifically related to cell-cycle arrest and proliferation. Therefore, although the original model was clearly able to generalize, subsuming in the decoded contrasts a considerable portion of the input variance [[Fig. Sx]], we expected a boost in this portion after fine-tuning. We fine-tuned the human model using a select subset of the study DGE contrasts and a low learning rate / small number of epochs policy (see Methods). As anticipated, the fine-tuned (FT) model captures a higher proportion of variance of the input DGE profiles in the decoded fraction (Fig. Sx). More importantly, compound identifiability across contexts is improved with the FT model residual fraction (Fig. 3c, Fig. 3d), indicating that this increase did not come at the expense of expunging compound-specific effects from the residual fraction. We conclude that contrast decomposition can benefit from model fine-tuning, which increases the amount of nuisance variance captured in the decoded fraction and results in better deconvolution of treatment specific and non-specific effects. We note, however, that appropriate selection of the training policy and examples presented during fine-tuning can be critical in this respect (see Methods).

Together these results affirm that the residual contrast is a highly specific experimental signature, characteristic of the treatment and robust with respect to cellular context and secondary effects. As such, we postulate that it constitutes a better read-out for direct treatment effects and the underlying mechanisms of action.

Contrast decomposition of time-series experiments with orthos recapitulates the dynamics of direct and secondary treatment effects.

In the temporal evolution of experimental treatments, direct effects manifest first and are the earliest to be reflected in transcriptomic read-outs. Secondary effects appear with a delay and will typically become more prominent over time as they propagate to a broader range of cellular processes. Eventually, the cells will either reach a new steady-state or, in the case of transient treatment effects, they will revert to their original state. To test if orthos is able to trace these dynamics in actual data we applied it to various time course datasets.

In a first setting, we reanalyzed nine publicly available time-series studies of various genetic, drug and physiological treatments spanning a wide range of cell line contexts and timescales. Using orthos, we decomposed the expression changes over time to their decoded and residual components. We computed the portion of variance explained by each component and also kept track of the total treatment variance (Fig. 4a). We observe a consistent trend whereby both the total variance and the fraction explained by the decoded contrast component gradually increase resulting in a concomitant decrease of the residual-contrast-explained variance. This profile is consistent with the dynamics outlined earlier whereby total variance increases as secondary effects compound and cells move away from their initial steady state. It is also in line with the asserted roles of the decomposed contrast components, with the decoded contrast tracking the gradually more prominent secondary effects and the residual contrast mirroring the subsiding relative importance of direct treatment effects and experimental noise. Interestingly, in many experiments, a plateau is reached for the total variance and the decoded-explained variance portion but, in several cases, this is attained first in the latter. This is precisely what one would anticipate if secondary effects become dominant before a new steady state is reached, with the two plateaus being hallmarks of these two milestones.

We then moved to a high-throughput single-cell study that interrogated the effects of specific compounds in multiple cell lines, at two distinct time-points (6 hours and 24 hours) after treatment [[MIX-Seq]]. After decomposition of the DGE profiles of condition-grouped pseudobulks we calculated the fraction of variance explained by each component in the two evaluated time-points (Fig. 4b). We find that, with the exception of a small number of drug treatments that produce no effects, the fraction of

variance explained by the decoded fraction is consistently higher at the later time point and conversely for the residual fraction. This is true irrespective of drug, cell line context and treatment effect size in any of the two time points (Fig. 4b, Fig Sx). As in the previous analysis, this showcases a robust association of the decoded and residual contrasts with delayed and early treatment effects respectively. We interpret this as a direct outcome of the time dynamics of direct and secondary treatment effects, with the former being reflected in the residual contrasts and the latter in the decoded contrasts.

Discussion

We present here orthos, a novel approach for dissecting the effects of experimental treatments on gene expression, by decomposing their specific and non-specific effects. This addresses a common problem in differential gene expression studies that is key to their interpretation and has important repercussions in both basic and applied research. By providing a more discerning picture of how gene expression changes in response to genetic perturbations, it can offer new insights into gene function and the biological processes they are involved in. In addition, by aiding the disambiguation of primary and secondary effects upon molecular or physiological treatments it can advance the study of complex regulatory networks that govern cellular function. In chemical biology and drug discovery we envision orthos as a new tool for the identification of the molecular targets of compounds and elucidation of their mechanisms of action, which are critical yet notoriously laborious aspects in drug research and development.

The utility of orthos is further enhanced by its integration with the compiled contrast database, which encompasses data from over 100,000 differential gene expression experiments. This corpus of data provides a rich resource for researchers seeking to compare their results with existing experiments. We release the trained models, the complete input and decomposed contrast database, including extensive sample and feature annotation and programmatic bindings for their retrieval as a standalone resource in a companion data package:

https://bioconductor.org/packages/devel/data/experiment/html/orthosData.html.

We provide functions for new contrast decomposition, querying against the contrast database and results visualization in the main orthos analysis package:

https://bioconductor.org/packages/devel/bioc/html/orthos.html.

Acknowledgements

We thank Sebastien Smallwood, Juan Valcarcel and Jeffrey Chao for their valuable comment and feedback

Author Contributions

P.P. and D.G conceived the study. P.P. designed model architecture, compiled the training data, and performed model evaluation. P.P and D.G performed analyses of experimental datasets. P.P, C.S and M.S developed the orthos and orthosData packages, P.P. wrote the manuscript with input from all authors. M.S supervised the project.

Competing interest statement.
The authors declare that they have no conflict of interest.

Supplementary Information is available for this paper.

Correspondence and requests for materials should be addressed to P. Papasaikas.

References

1. Lelli KM, Slattery M, Mann RS. Disentangling the many layers of eukaryotic transcriptional regulation. Annu Rev Genet. 2012;46:43-68.
2. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. Cell. 2013;152(6):1237-51.
3. Schenone M, Dancik V, Wagner BK, Clemons PA. Target identification and mechanism of action in chemical biology and drug discovery. Nat Chem Biol. 2013;9(4):232-40.
4. Fedorov Y, Anderson EM, Birmingham A, Reynolds A, Karpilow J, Robinson K, et al. Off-target effects by siRNA can induce toxic phenotype. RNA. 2006;12(7):1188-96.
5. Kanasty RL, Whitehead KA, Vegas AJ, Anderson DG. Action and reaction: the biological response to siRNA and its delivery vehicles. Mol Ther. 2012;20(3):513-24.
6. Meng Z, Lu M. RNA Interference-Induced Innate Immunity, Off-Target Effect, or Immune Adjuvant? Front Immunol. 2017;8:331.
7. Kim S, Koo T, Jee HG, Cho HY, Lee G, Lim DG, et al. CRISPR RNAs trigger innate immune responses in human cells. Genome Res. 2018;28(3):367-73.
8. Goel K, Ploski JE. RISC-y Business: Limitations of Short Hairpin RNA-Mediated Gene Silencing in the Brain and a Discussion of CRISPR/Cas-Based Alternatives. Front Mol Neurosci. 2022;15:914430.
9. Annoni A, Gregori S, Naldini L, Cantore A. Modulation of immune responses in lentiviral vector-mediated gene transfer. Cell Immunol. 2019;342:103802.
10. Timm M, Saaby L, Moesby L, Hansen EW. Considerations regarding use of solvents in in vitro cell based assays. Cytotechnology. 2013;65(5):887-94.
11. Adler S, Pellizzer C, Paparella M, Hartung T, Bremer S. The effects of solvents on embryonic stem cell differentiation. Toxicol In Vitro. 2006;20(3):265-71.
12. Mandelboum S, Manber Z, Elroy-Stein O, Elkon R. Recurrent functional misinterpretation of RNA-seq data caused by sample-specific gene length bias. PLoS Biol. 2019;17(11):e3000481.
13. Parekh S, Ziegenhain C, Vieth B, Enard W, Hellmann I. The impact of amplification on differential expression analyses by RNA-seq. Sci Rep. 2016;6:25533.
14. Zheng W, Chung LM, Zhao H. Bias detection and correction in RNA-Sequencing data. BMC Bioinformatics. 2011;12:290.
15. Love MI, Hogenesch JB, Irizarry RA. Modeling of RNA-seq fragment sequence bias reduces systematic errors in transcript abundance estimation. Nat Biotechnol. 2016;34(12):1287-91.
16. Gayoso A, Lopez R, Xing G, Boyeau P, Valiollah Pour Amiri V, Hong J, et al. A Python library for probabilistic analysis of single-cell omics data. Nat Biotechnol. 2022;40(2):163-6.
17. Lotfollahi M, Klimovskaia Susmelj A, De Donno C, Hetzel L, Ji Y, Ibarra IL, et al. Predicting cellular responses to complex perturbations in high-throughput screens. Mol Syst Biol. 2023;19(6):e11517.
18. Lotfollahi M, Naghipourfar M, Luecken MD, Khajavi M, Buttner M, Wagenstetter M, et al. Mapping single-cell data to reference atlases by transfer learning. Nat Biotechnol. 2022;40(1):121-30.

19. Lotfollahi M, Rybakov S, Hrovatin K, Hediyeh-Zadeh S, Talavera-Lopez C, Misharin AV, et al. Biologically informed deep learning to query gene programs in single-cell atlases. Nat Cell Biol. 2023;25(2):337-50.
20. Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, et al. Massive mining of publicly available RNA-seq data from human and mouse. Nat Commun. 2018;9(1):1366.
21. Srivatsan SR, McFaline-Figueroa JL, Ramani V, Saunders L, Cao J, Packer J, et al. Massively multiplex chemical transcriptomics at single-cell resolution. Science. 2020;367(6473):45-51.

Figure 1.

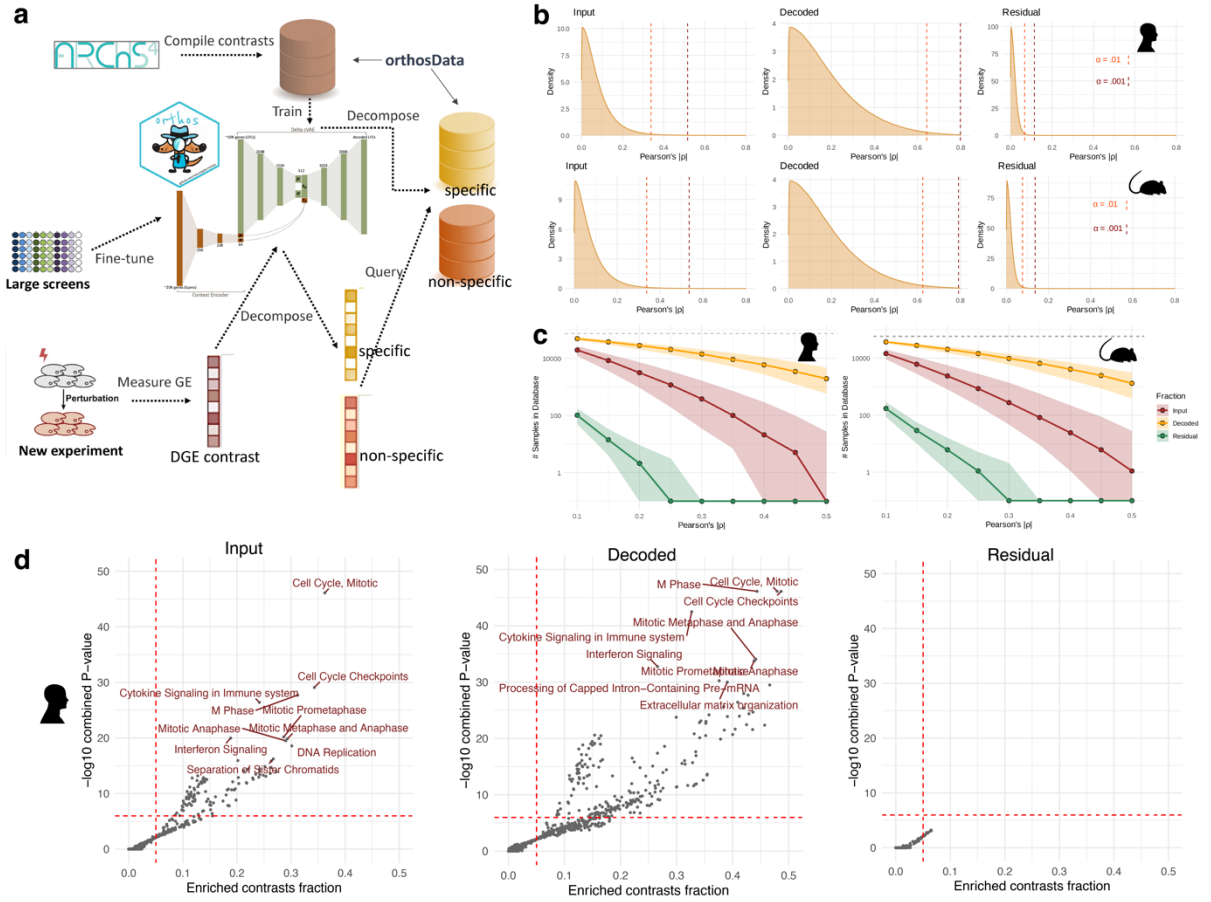


Figure 2.

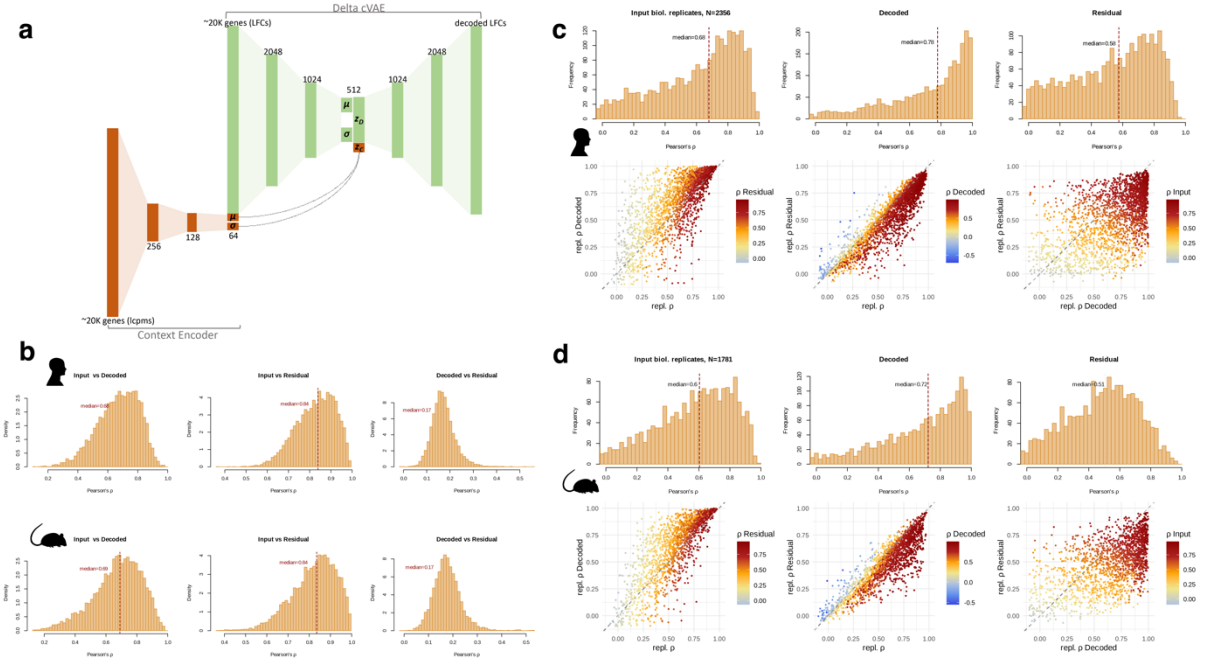


Figure 3.

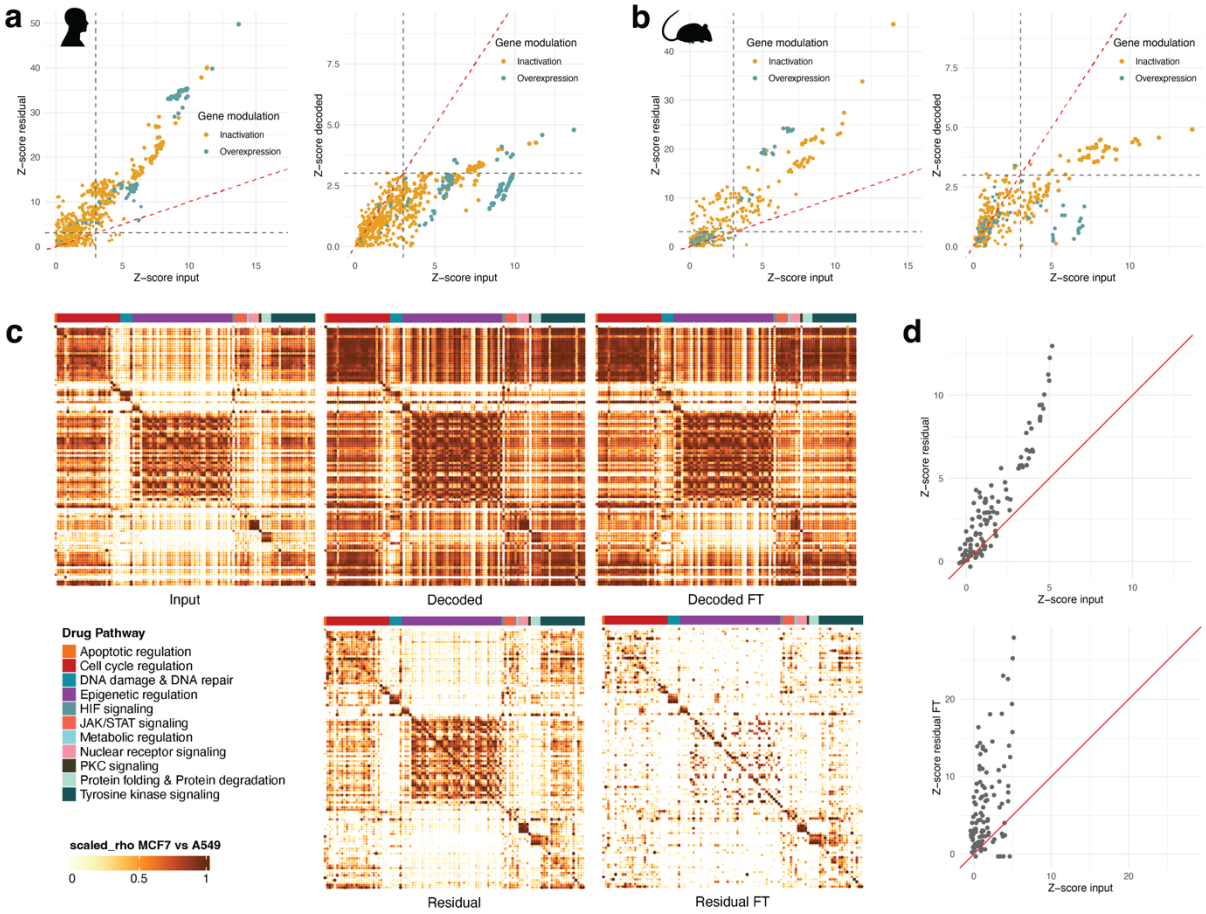
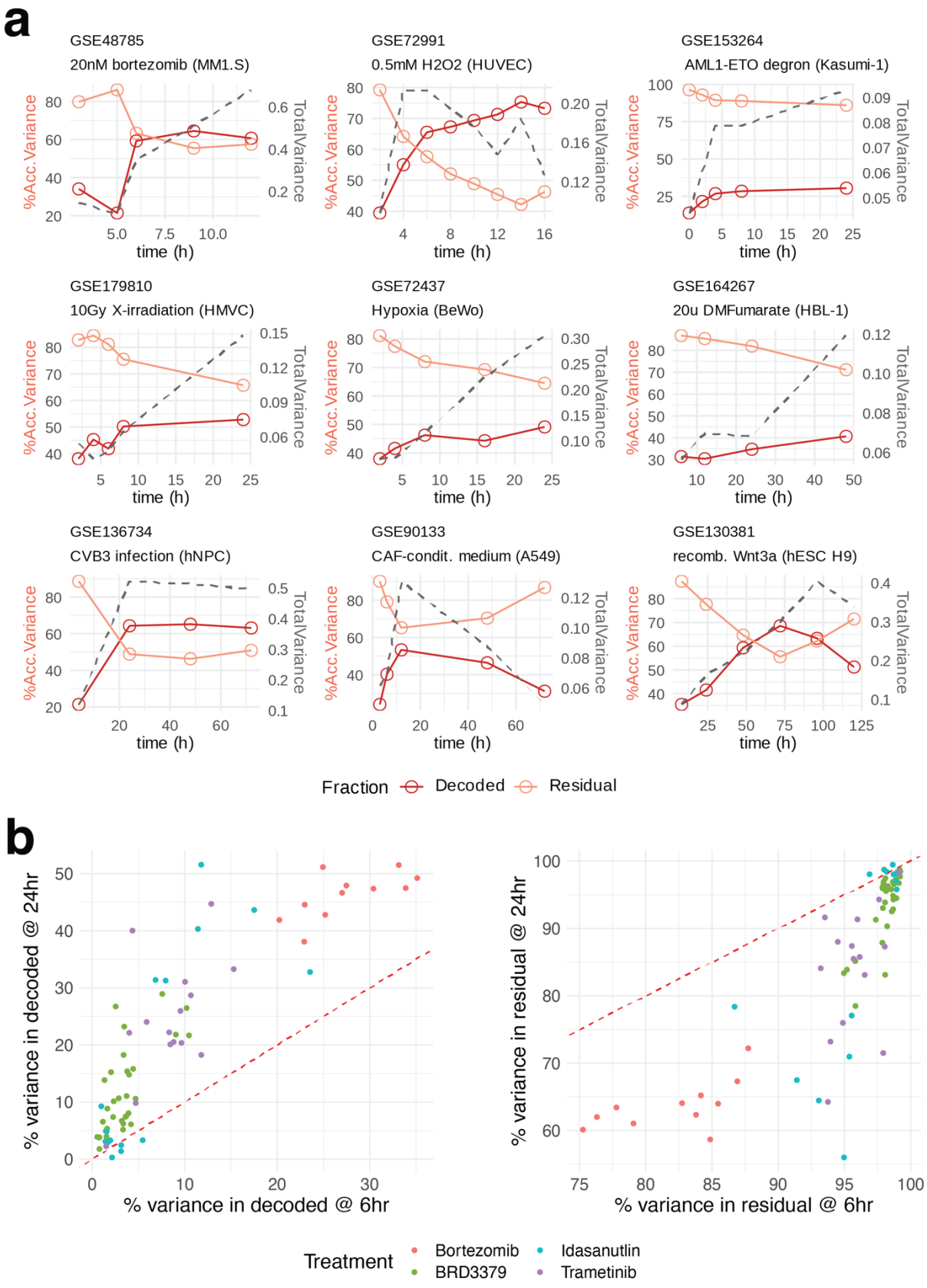

a
b
c
d
Z−score residual
Z−score decoded
Z−score input
Gene modulation
Inactivation
Overexpression
Input
Decoded
Decoded FT
Residual
Residual FT
Drug Pathway
Apoptotic regulation
Cell cycle regulation
DNA damage & DNA repair
Epigenetic regulation
HIF signaling
JAK/STAT signaling
Metabolic regulation
Nuclear receptor signaling
PKC signaling
Protein folding & Protein degradation
Tyrosine kinase signaling
scaled_rho MCF7 vs A549
0
0.5
1
Z−score residual FT

Figure 4.


a
GSE48785
20nM bortezomib (MM1.S)
GSE72991
0.5mM H2O2 (HUVEC)
GSE153264
AML1-ETO degron (Kasumi-1)
GSE179810
10Gy X-irradiation (HMVC)
GSE72437
Hypoxia (BeWo)
GSE164267
20u DMFumarate (HBL-1)
GSE136734
CVB3 infection (hNPC)
GSE90133
CAF-condit. medium (A549)
GSE130381
recomb. Wnt3a (hESC H9)
%Acc.Variance
TotalVariance
time (h)
Fraction
Decoded
Residual
b
% variance in decoded @ 24hr
% variance in decoded @ 6hr
% variance in residual @ 24hr
% variance in residual @ 6hr
Treatment
Bortezomib
BRD3379
Idasanutlin
Trametinib