Drug specific transcriptomic signatures in healthy human subject iPSC derived cardiomyocytes

Jens Hansen[1,2], Yuguang Xiong[1,2], Priyanka Dhanan[3], Bin Hu[1,2], Gomathi Jayaraman[1,2], Rosa Tolentino[1,2], Yibang Chen[1,2], Kristin G. Beaumont[4], Robert Sebra[4], Dusica Vidovic[5], Stephan C. Schürer[5], Joseph Goldfarb[2], Joseph Gallo[6], Marc R. Birtwistle[7], Eric A. Sobie[2], Evren U. Azeloglu[2,8], Christoph Schaniel[3,9,*], Nicole C. Dubois[3,*], Ravi Iyengar[1,2,*]

[1] Mount Sinai Institute for Systems Biomedicine, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[2] Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[3] Department of Cell, Developmental and Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[4] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

[5] Institute for Data Science and Computing, University of Miami, Coral Gables, FL 33146, USA

[6] School of Pharmacy and Pharmaceutical Sciences ,University of Buffalo SUNY System, Buffalo NY 14260

[7] Chemical and Biomolecular Engineering, Clemson University, Clemson, SC, 29634, USA

[8] Department of Medicine, Division of Nephrology, Icahn School of Medicine at Mount Sinai, New, York, NY 10029, USA

[9] Department of Medicine, Division of Hematology and Medical Oncology, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

* Co-senior authors

Address correspondence to

Jens Hansen – jens.hansen@mssm.edu

or

Ravi Iyengar - ravi.iyengar@mssm.edu

Abstract

Drug-induced gene expression profiles are an important source for the characterization of drug-specific mechanisms of action that might also indicate potential mechanisms of drug toxicity. Unfortunately, drug-induced transcriptomic signatures are often the sum of multiple responses, such as cell line or cell-type specific responses, and can include false positive results. Both issues complicate the identification of drug-specific effects. To unmask drug-specific effects in drug-induced transcriptomic signatures, we use singular value decomposition to characterize shared transcriptomic responses induced in multiple cell lines treated with the same drug. Six different cardiomyocyte cell lines that were generated from induced pluripotent stem cells obtained from six healthy human subjects were stimulated with 25 protein kinase inhibitors, 4 monoclonal antibodies against protein kinases and 25 other cardiac- and non-cardiac acting drugs. Searching for drug-specific subspaces that characterize drug-specific responses in the original data, we could identify highly similar drug responses for multiple drugs. Pathway enrichment analysis of those responses predicts reasonable effects for multiple drugs. Revealing drug-specific responses from transcriptomic signatures might allow an easier characterization of drug-induced toxicity, such as cardiotoxic side effects induced by multiple protein kinase inhibitors and monoclonal antibodies against protein kinases.

Introduction

Unwanted side-effects of the therapeutically useful drugs continue to be a substantial problem (*1*). Even after a drug is approved and introduced into the market, pharmacovigilance studies often reveal adverse events that lead to warning label requirements mandated by the FDA (*2*). Practical reasons including time required to bring a potentially useful drug to market and costs of clinical trials which are rightfully focused on therapeutic efficacy often limit the patients with potential for side effects who can be studied during clinical trials to understand the details regarding adverse events. Hence, early indications of a potential of adverse events could be useful in the drug development process at various stages (*3, 4*). Adverse events can occur due to a variety of reasons due to mechanisms operative at the molecular and cellular levels (*5*). Although cellular mechanisms of adverse events are most often not well defined, preclinical studies at the molecular levels are often useful as has been demonstrated by HERG channel protein interacting drugs and the potential for arrythmias (*6, 7*). Hence, molecular studies in cell-based systems have the potential to be useful in predicting adverse event potential of drugs.

Drug related adverse events are often organ selective with kidney, liver and heart as major organs where drug toxicity leads to serious and often life-threating adverse events. In addition to drugs with the potential to produce arrythmias, many useful cancer drugs such as tyrosine kinase inhibitors which are used for targeted cancer therapy can often be associated with cardiac insufficiencies and development of heart failure. This has been recognized for nearly two decades (*8, 9*). However, a systematic understanding of the molecular pathways and potential cellular mechanisms associated with these toxicities are still not well understood.

Cell based assays using known cardiotoxic drugs have become a potentially useful approach for studies that can provide predictive understanding of why some drugs many be associated with cardiotoxicity. With the development of human iPSC-derived cardiomyocytes (*10*) several studies have demonstrated the potential for their use in understanding cardiotoxicity (*11, 12*). We have used human adult cardiomyocyte like cells to demonstrate the relationship between tyrosine kinase inhibitor induced transcriptional profiling and the cardiotoxicity risk as assessed from pharmacovigilance data (*13*). Hence, transcriptional profiles in human cardiomyocyte can be useful for studies focused on mechanism-based drug signatures that could be used for prediction of cardiotoxicity potential in some human subjects. In this study, we have used six healthy human subject cardiomyocyte cell lines to conduct a detailed study of 54 drugs to map drug selective signatures across the different human subject lines.

Very often drug-induced gene expression profiles are not only composed of drug-specific responses, but might also contain cell-type and cell line specific responses as well as false positives. Depending on the relative contribution of these effects, the drug-specific response might not be obviously detectable in the list of induced differentially expressed genes. Singular value decomposition has been applied to investigate multiple transcriptomic profiles. For example, it has been successfully used to uncover cycling gene expression profiles indicative of different yeast cell cycle phases that were hidden in the full gene expression dataset (*14*). Multiple applications of singular value decomposition (*15-17*) or other methods for the separation of omics data into subcomponents, such as independent component analysis (*18, 19*), followed.

Here, we use singular value decomposition to reveal drug-specific responses from the gene expression profiles induced by the 54 drugs in the six human cardiomyocyte cell lines. We document that singular value decomposition allows the identification of gene expression profiles that show high similarity among different cell lines treated with the same drug. Pathway enrichment analysis of the identified drug-specific gene expression profiles further underlines the high consistency between the results revealed from the different cell lines and predicts reasonable effects for multiple drugs.

Results

Six different cardiomyocyte cellines that were generated from induced pluripotent stems cells obtained from six healthy human subjects were stimulated with 25 protein kinase inhibitors, 4 monoclonal antibodies against protein kinases and 25 other cardiac- and non-cardiac acting drugs (Table 1). Transcriptomic analysis of gene expression profiles with and without drug treatment generated 266 lists of differentially expressed genes (DEGs) (Figure 1). To analyze if drug induced DEGs in the different cell lines document drug-specific gene expression profiles, we subjected the lists of DEGs to pairwise correlation analysis followed by hierarchical clustering (Figure 2A). If the lists of DEGs are representative of drug-specific responses, hierarchical clustering should group different samples of the same drug treatment within the same cluster. A few samples were indeed grouped together based on the treated drug, while most of the samples were grouped by the treated cell line. Additionally, the clustering outcome was also determined by the amplitude of the drug response that can be quantified by the number of significantly differentially expressed genes. To quantify the efficiency of drug-specific clustering for each drug, we calculated the F1 score (Figure 2B). The F1 score is the harmonic mean of precision and recall. Here, precision indicates the fraction of samples in a particular cluster that were treated with a particular drug, recall indicates the fraction of all samples treated with that drug that are in that particular cluster. For this analysis, any cluster emerging within the dendrogram was investigated, independently of the height at which the dendrogram needed to be cut to generate that cluster. For each drug, the maximal F1 score is documented. As expected, most drugs are associated with a very low F1 score. In summary, our analysis reveals that the documented gene expression profiles are mainly determined by the amplitude of the drug response as well as by cell line specific responses that both mask drug-specific responses.

Singular value decomposition

To dissolve drug-specific responses from our data we subjected the DEGs to singular value decomposition. Singular value decomposition is a factorization of a matrix into orthonormal left singular vectors, singular values and orthonormal right singular vectors (Supplemental figure 1A). Applied to gene expression data, the left singular vectors are often called eigenassays and the singular values eigenexpression values (*14*). Each gene expression profile is a linear combination of all eigenassays. The sample specific coefficients of this linear combination are described by the matrix of right singular vectors. The eigenexpression values document how much each eigenassay contributes to the total gene expression and need to be considered for the linear combination as well.

General stress response

Singular value decomposition of our data identified 266 eigenassays that contribute with different degrees to the original gene expression profiles (Figure 3A). Since our initial clustering suggested that the amplitude of the response significantly influences the induced gene expression profiles, independently of the used drug or treated cell line, we correlated the sample specific coefficients of each eigenassay with the number of significantly differentially expressed genes in each sample (Supplemental figure 1B). We observed a correlation coefficient of -0.91 for the first eigenassay (Figure 3B), indicating that the first eigenassay captures the documented amplitude specific response. Enrichment analysis of the top 600 genes of the first eigenassay predicts multiple pathways related to muscle contractility (Figure 3C). We hypothesized that this amplitude specific response describes a general stress response that is induced by the magnitude of the cellular perturbation, independently of the pertubed cell line or pertubing drug. We removed the first eigenassay from the gene expression profiles. Hierarchical clustering of the obtained gene expression profiles documents a disruption of the initially observed grouping of samples by the number of significantly differentially expressed genes (Figure 3D). All following analyzes were generated with the data after removal of the first eigenassay.

Drug and cell line specific eigenassays

To identify eigenassays that capture drug-specific responses we used student's t-test to investigate if the coefficients related to each eigenassay significantly differ between the samples of a particular drug and the samples of all other drugs (Supplemental figure 1C). As a result, we obtained a p-value for each eigenassay and drug combination that was transformed into a $-\log_{10}$(p-value), allowing the assignment of each drug to a vector of eigenassay specific $-\log_{10}$(p-values). Similarly, we applied the same algorithm using cell lines instead of drugs, generating a vector of eigenassay specific $-\log_{10}$(p-values) for each cell line. We calculated all pairwise correlations between drug- and cell line-specific vectors, followed by hierarchical clustering (Figure 5A). Results revealed that the degrees to which each eigenassay captures drug-specific responses differ from the degrees to which each eigenassay captures cell line-specific responses, as documented by the grouping of the six different cell lines in one single cluster.

Drug specific subspaces capture drug-specific gene expression profiles

Any combination of eigenassays spans a subspace that contains a particular fraction of the full DEG profiles. For each drug, we ranked all eigenassays by significance and generated 264 subspaces spanned by the top three to 266 ranked eigenassays. We then projected all 266 full gene expression profiles into each of these subspaces, followed by pairwise correlation analysis and hierarchical clustering. The maximum F1 score for that particular drug in each subspace was calculated, as described above. To analyze how much of the initial information still lies in that subspace, we calculated the cosine similarity between the full gene expression vectors for that drug and their projections into each subspace. F1 scores and median cosine similarities were used to calculate a selection score for each subspace. Since we wanted to extract the similarities between the different gene expression profiles induced by the same drug, we used stringent focus on the F1 score and defined the selection score to be determined to 95% by the F1 score and to 5% by the median cosine similarity. That subspace that was associated with the highest selection score was selected as the specific subspace for that particular drug (Supplemental figure 2). Repeating the analysis for each drug lead to the identification of 54 drug-specific subspaces.

Projection of gene expression profiles into these drug-specific subspaces significantly improved the clustering efficiency (Figure 5A, Supplemental figure 3), as documented by the F1 scores (Figure 5B). The projected amount of initial information in the drug-specific subspaces as described by the median cosine similarities varies between the different drugs. Drug-specific gene expression profiles in their corresponding drug-specific subspaces were combined into a new matrix. Not surprisingly, clustering of this matrix identifies clusters that mostly contained samples that were treated with the same drug (Figure 5C).

Drug specific gene expression profiles lead to the prediction of similar mechanisms of action

Projection of gene expression profiles into drug-specific subspaces allowed us to identify common response patterns that were induced by treatment of different cell lines with the same drug. Focusing on drug-specific effects that are drug- but not cell line-dependent, we subjected the projected gene expression profiles to pathway enrichment analysis using the Molecular Biology of the Cell Ontology (*20*) (Figure 6, Supplemental figure 4). To investigate the effects of our decomposition pipeline on the predicted pathways, we additionally subjected the full (initial) gene expression profiles and the gene expression profiles after removal of the first eigenassay to pathway enrichment analysis (Supplemental figure 4). Predicted up- and downregulated subcellular processes (SCPs) were ranked by significance. Comparison of the top five predicted SCPs between the three datasets documents that our decomposition pipeline significantly increases the similarity between the predicted pathways for the same drug in different cell lines. In some cases the decomposition lead to the identification of SCPs that were not among the top five predictions of any cell line in the full datasets.

A common element in the predicted pathways for multiple EGFR inhibitors (erlotinib, afatinib, gefitinib, cetuximab, trastuzumab) is the downregulation of SCPs involved in DNA replication and chromosome segregation by the mitotic spindle (Figure 6, Supplemental figure 4). These pathways were also constantly downregulated by the kinase inhibitor trametenib, the anthracycline idarubicine and the glucocorticoid prednision. On the contrary, the kinase inhibitors nilotinib, imatinib, regorafenib, sorafenib and ponatinib as well as the antirarrhythmic flecainide consistently upregulate SCPs involved in centrosome dynamics, chromosomal separation and cell cycle progression. These findings are supported by morphological observations of centrosome aberrations in disease unrelated fibroblasts and cells from the oral mucosa obtained from patients treated with dasatinib, nilotinib, imatinib, sorafenib and sunitinib (*21*). The anthracycline idarubicine up-regulated DNA repair mechanisms and the SCP 'Cell cycle arrest due to DNA damage', in agreement with its stronger activity in inducing DNA damage, if compared to daunorubicin (*22*). The SCP 'JAK-STAT signaling pathway' was the top ranked downregulated SCP in four of six cell lines treated with the JAK-STAT inhibitor ruxolitinib. Investigation of the SCP genes documented that in all six cell lines the genes STAT3, STAT4, SOCS2 and SCOS3 were downregulated. The inhibitory components SOCS2 and SOCS3 are upregulated as part of a negative feedback loop in response to JAK-STAT signaling (*23*). Consequently, their downregulation might be the consequence of the inhibition of JAK-STAT signaling. The upregulation of proteasomal components and genes involved in poly-ubiquitination by the proteasome inhibitor bortezomib could be a compensatory mechanism to mitigate proteasome inhibition (Supplemental figure 4). The downregulation of the SCP 'Cellular iron storage' by the anthracyclines daunorubicin and doxyrubicine (Supplemental figure 4) is in agreement with the

interference of doxorubicine with cellular iron metabolism (*24*). Described pathways were predicted based on the decomposed datasets. In the full dataset they were either less consistently predicted, i.e. they were identified in less cell lines, or they were not among the top 5 predictions in any cell line.

Discussion

The identification of any drug-induced transcriptomic signatures including those that are indicative of cardiotoxicity depends on the separation of drug-specific transcriptomic responses from other responses, including cell type or cell line specific responses as well as general responses to perturbations of cells by extracellular ligands. These additional components of the overall response have to be removed before we can determine if we have drug specific signatures in a cell type of interest.

Here, we developed a pipeline for decomposition of transcriptomic profiles based on singular value decomposition that searches for the largest shared component in gene expression profiles induced by the same drug in different cell lines. Singular value decomposition of gene expression data results in the identification of orthonormal eigenassays. Our pipeline searches for drug-specific subspaces spanned by a subset of those eigenassays that contain most of the drug related gene expression response and exclude gene expression profiles involved in the other response types. To find such subspaces we document the clustering efficiency for each drug after projection of the full gene expression profiles into subspace candidates. As a second criterion for determining drug specific signatures across cell lines (i.e. that is different human subjects) we use the cosine similarity to quantify how much of the overall gene expression response for a drug of interest is still preserved in each potential subspace. In this study, we used a stringent focus on the similarity of drug responses, since our final decision score is determined to 95% by the clustering efficiency and to 5% by the cosine similarity. This might come with the cost of losing too much of the original information in some cases, so that other constellations are justifiable as well.

The identified drug related responses in the drug-specific subspaces are of high similarity and allow the prediction of similar and in some cases almost identical pathways induced by the same drugs in the different cell lines representing different human subjects. In contrast, pathways predicted from the full dataset show much larger variations, indicating that our decomposition pipeline enriches for drug-specific responses from the overall gene expression profiles. Literature research and biological interpretation of the identified pathways from these drug specific gene expression profiles show that there is reasonable agreement between the predicted pathways from mRNA expression profiles and pathways inferred from small-scale experimental studies. This congruence give us confidence that future studies can identify drug selective pathways associated with cellular mechanisms related to potential drug toxicity.

References

1. E. N. de Vries, M. A. Ramrattan, S. M. Smorenburg, D. J. Gouma, M. A. Boermeester, The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care* **17**, 216-223 (2008).
2. K. Drozda, M. A. Pacanowski, C. Grimstein, I. Zineh, Pharmacogenetic Labeling of FDA-Approved Drugs: A Regulatory Retrospective. *JACC Basic Transl Sci* **3**, 545-549 (2018).
3. A. F. Smith, A. Klotz, I. M. Wormstone, Improving the drug development process by reducing the impact of adverse events: the case of cataracts considered. *Drug Discov Today* **21**, 510-516 (2016).
4. M. A. Dorato, L. A. Buckley, Toxicology in the drug discovery and development process. *Curr Protoc Pharmacol* **Chapter 10**, Unit10 13 (2006).
5. A. Ma'ayan, S. L. Jenkins, J. Goldfarb, R. Iyengar, Network analysis of FDA approved drugs and their targets. *Mt Sinai J Med* **74**, 27-32 (2007).
6. J. I. Vandenberg *et al.*, hERG K(+) channels: structure, function, and clinical significance. *Physiol Rev* **92**, 1393-1478 (2012).
7. E. Raschi, V. Vasina, E. Poluzzi, F. De Ponti, The hERG K+ channel: target and antitarget strategies in drug development. *Pharmacol Res* **57**, 181-195 (2008).
8. D. Jain, W. Aronow, Cardiotoxicity of cancer chemotherapy in clinical practice. *Hosp Pract (1995)* **47**, 6-15 (2019).
9. A. Garcia-Alvarez, X. Garcia-Albeniz, J. Esteve, M. Rovira, X. Bosch, Cardiotoxicity of tyrosine-kinase-targeting drugs. *Cardiovasc Hematol Agents Med Chem* **8**, 11-21 (2010).
10. Y. Yoshida, S. Yamanaka, Induced Pluripotent Stem Cells 10 Years Later: For Cardiac Applications. *Circ Res* **120**, 1958-1968 (2017).
11. A. Sharma *et al.*, High-throughput screening of tyrosine kinase inhibitor cardiotoxicity with human induced pluripotent stem cells. *Sci Transl Med* **9**, (2017).
12. H. Wang *et al.*, Adaptation of Human iPSC-Derived Cardiomyocytes to Tyrosine Kinase Inhibitors Reduces Acute Cardiotoxicity via Metabolic Reprogramming. *Cell Syst* **8**, 412-426 e417 (2019).
13. J. G. C. van Hasselt *et al.*, Transcriptomic profiling of human cardiac cells predicts protein kinase inhibitor-associated cardiotoxicity. *Nat Commun* **11**, 4809 (2020).
14. O. Alter, P. O. Brown, D. Botstein, Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci U S A* **97**, 10101-10106 (2000).
15. O. Alter, G. H. Golub, Singular value decomposition of genome-scale mRNA lengths distribution reveals asymmetry in RNA gel electrophoresis band broadening. *Proc Natl Acad Sci U S A* **103**, 11828-11833 (2006).
16. S. P. Ponnapalli, M. A. Saunders, C. F. Van Loan, O. Alter, A higher-order generalized singular value decomposition for comparison of global mRNA expression from multiple organisms. *PLoS One* **6**, e28072 (2011).
17. O. Alter, P. O. Brown, D. Botstein, Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms. *Proc Natl Acad Sci U S A* **100**, 3351-3356 (2003).
18. A. Biton *et al.*, Independent component analysis uncovers the landscape of the bladder tumor transcriptome and reveals insights into luminal and basal subtypes. *Cell Rep* **9**, 1235-1245 (2014).

19. W. Liu, S. H. Payne, S. Ma, D. Fenyo, Extracting Pathway-level Signatures from Proteogenomic Data in Breast Cancer Using Independent Component Analysis. *Mol Cell Proteomics* **18**, S169-S182 (2019).
20. J. Hansen, D. Meretzky, S. Woldesenbet, G. Stolovitzky, R. Iyengar, A flexible ontology for inference of emergent whole cell function from relationships between subcellular processes. *Sci Rep* **7**, 17689 (2017).
21. M. Giehl *et al.*, Detection of centrosome aberrations in disease-unrelated cells from patients with tumor treated with tyrosine kinase inhibitors. *Eur J Haematol* **85**, 139-148 (2010).
22. T. Fukushima, T. Ueda, M. Uchida, T. Nakamura, Action mechanism of idarubicin (4-demethoxydaunorubicin) as compared with daunorubicin in leukemic cells. *Int J Hematol* **57**, 121-130 (1993).
23. R. Morris, N. J. Kershaw, J. J. Babon, The molecular details of cytokine signaling via the JAK/STAT pathway. *Protein Sci* **27**, 1984-2009 (2018).
24. A. Ghigo, M. Li, E. Hirsch, New signal transduction paradigms in anthracycline-induced cardiotoxicity. *Biochim Biophys Acta* **1863**, 1916-1925 (2016).

Drug_class	Acronym	Full_drug_name	Drug_subclass
Kinase inhibitor	AFA	afatinib	Kinase inhibitor
Kinase inhibitor	AXI	axitinib	Kinase inhibitor
Kinase inhibitor	BOS	bosutinib	Kinase inhibitor
Kinase inhibitor	CAB	cabozantinib	Kinase inhibitor
Kinase inhibitor	CER	ceritinib	Kinase inhibitor
Kinase inhibitor	CRI	crizotinib	Kinase inhibitor
Kinase inhibitor	DAB	dabrafenib	Kinase inhibitor
Kinase inhibitor	DAS	dasatinib	Kinase inhibitor
Kinase inhibitor	ERL	erlotinib	Kinase inhibitor
Kinase inhibitor	GEF	gefitinib	Kinase inhibitor
Kinase inhibitor	IMA	imatinib	Kinase inhibitor
Kinase inhibitor	LAP	lapatinib	Kinase inhibitor
Kinase inhibitor	NIL	nilotinib	Kinase inhibitor
Kinase inhibitor	PAZ	pazopanib	Kinase inhibitor
Kinase inhibitor	PON	ponatinib	Kinase inhibitor
Kinase inhibitor	REG	regorafenib	Kinase inhibitor
Kinase inhibitor	RUX	ruxolitinib	Kinase inhibitor
Kinase inhibitor	SOR	sorafenib	Kinase inhibitor
Kinase inhibitor	SUN	sunitinib	Kinase inhibitor
Kinase inhibitor	TOF	tofacitinib	Kinase inhibitor
Kinase inhibitor	TRA	trametinib	Kinase inhibitor
Kinase inhibitor	VAN	vandetanib	Kinase inhibitor
Kinase inhibitor	VEM	vemurafenib	Kinase inhibitor
Monoclonal antibody	BEV	bevacizumab	Kinase inhibitor
Monoclonal antibody	CTX	cetuximab	Kinase inhibitor
Monoclonal antibody	RTX	rituximab	Kinase inhibitor
Monoclonal antibody	TRS	trastuzumab	Kinase inhibitor
Anthracycline	DAU	daunorubicin	Anthracycline
Anthracycline	DOX	doxorubicin	Anthracycline
Anthracycline	EPI	epirubicin	Anthracycline
Anthracycline	IDA	idarubicin	Anthracycline
Cardiac acting drug	AMI	amiodarone	Antiarrhythmic
Cardiac acting drug	DOB	dobutamine	Sympathomimetic
Cardiac acting drug	FLE	flecainide	Antiarrhythmic
Cardiac acting drug	ISO	isoprenaline	Sympathomimetic
Cardiac acting drug	MIL	milrinone	Vasodilatator
Cardiac acting drug	PHP	phenylephrine	Sympathomimetic
Cardiac acting drug	VRP	verapamil	Antiarrhythmic
Non cardiac acting drug	AZA	azacitidine	Other chemotherapy drug
Non cardiac acting drug	BOR	bortezomib	Proteasome inhibitor
Non cardiac acting drug	CAR	carfilzomib	Proteasome inhibitor
Non cardiac acting drug	CYC	cyclosporine	Immunosuppressant
Non cardiac acting drug	DEC	decitabine	Other chemotherapy drug
Non cardiac acting drug	DEL	delavirdine	Anti viral
Non cardiac acting drug	DIC	diclofenac	Nsaid
Non cardiac acting drug	EDN	endothelin-1	Endogenous vasoconstrictor
Non cardiac acting drug	EST	estradiol	Female sex hormone
Non cardiac acting drug	IGF	insulin-like growth factor 1	Hormone
Non cardiac acting drug	OLM	olmesartan	Antihypertensive
Non cardiac acting drug	PIO	pioglitazone	Antidiabetic
Non cardiac acting drug	PRE	prednisolone	Glucocorticoid
Non cardiac acting drug	ROS	rosiglitazone	Antidiabetic
Non cardiac acting drug	SAX	saxagliptin	Antidiabetic
Non cardiac acting drug	TNF	tnf-alpha	Cytokine

Table 1. Drug metadata

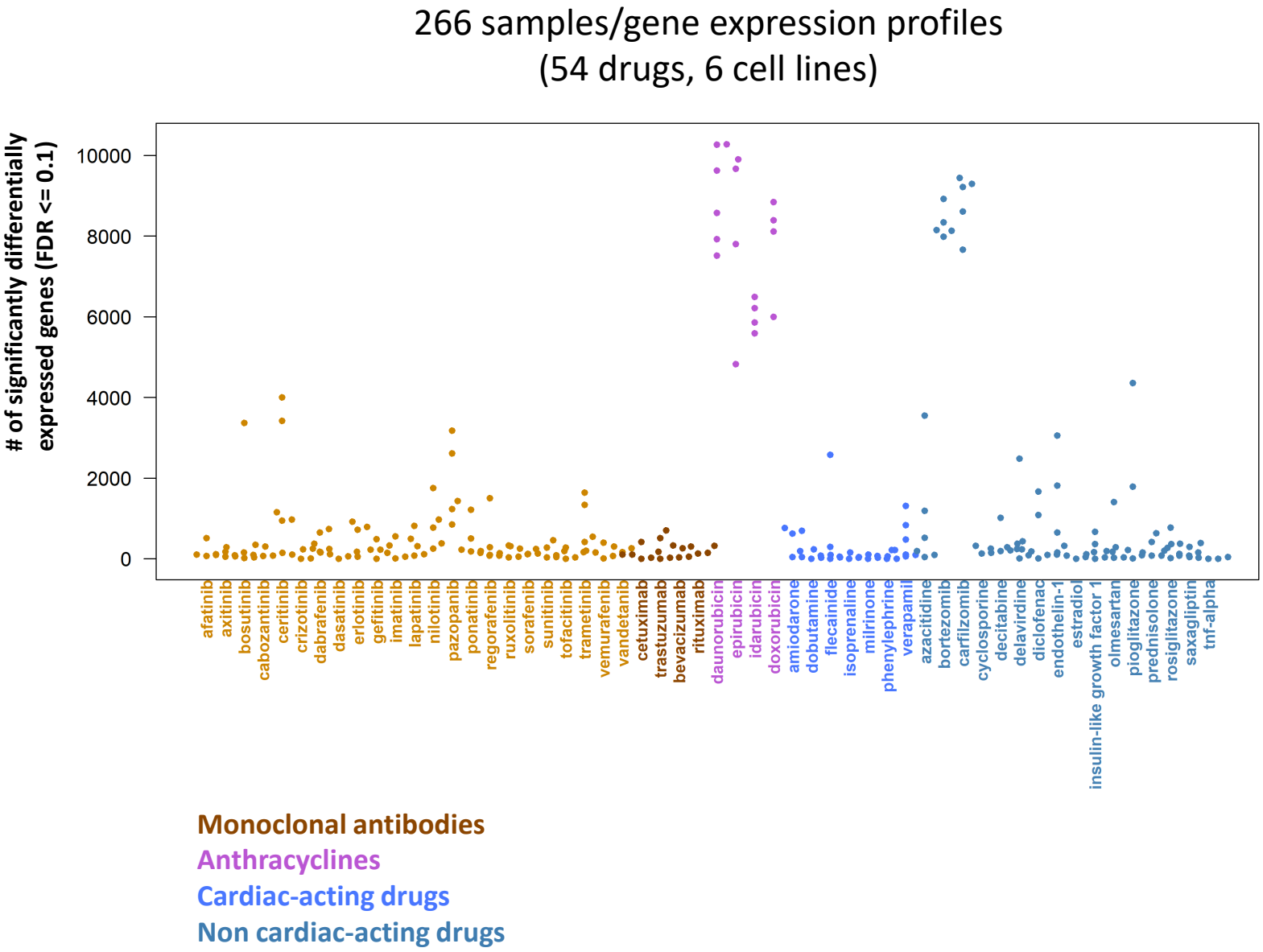
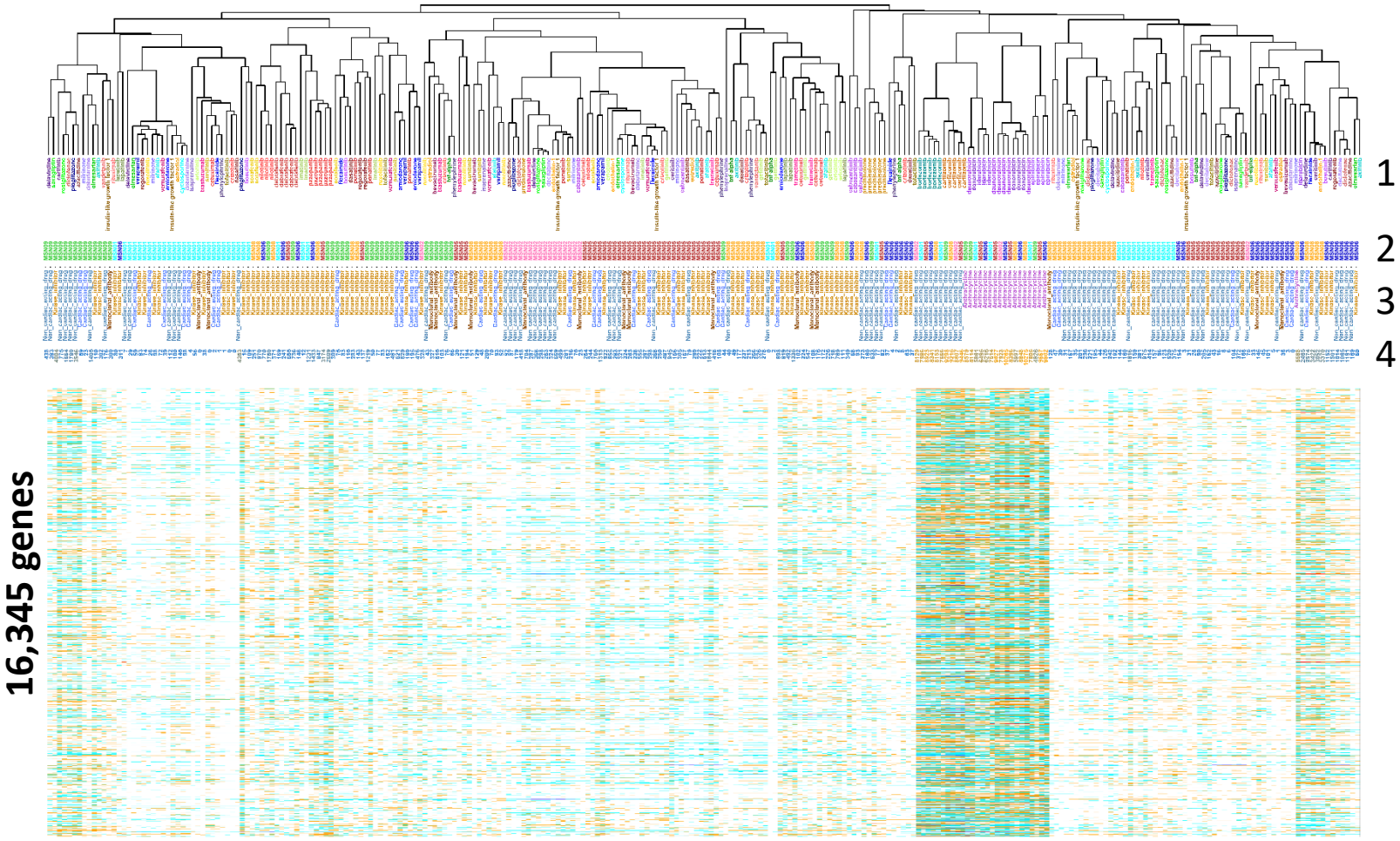

266 samples/gene expression profiles
(54 drugs, 6 cell lines)
# of significantly differentially expressed genes (FDR <= 0.1)
10000
8000
6000
4000
2000
0
afatinib
axitinib
bosutinib
cabozantinib
ceritinib
crizotinib
dabrafenib
dasatinib
erlotinib
gefitinib
imatinib
lapatinib
nilotinib
pazopanib
ponatinib
regorafenib
ruxolitinib
sorafenib
sunitinib
tofacitinib
trametinib
vemurafenib
vandetanib
cetuximab
trastuzumab
bevacizumab
rituximab
daunorubicin
epirubicin
idarubicin
doxorubicin
amiodarone
dobutamine
flecainide
isoprenaline
milrinone
phenylephrine
verapamil
azacitidine
bortezomib
carfilzomib
cyclosporine
decitabine
delavirdine
diclofenac
endothelin-1
estradiol
insulin-like growth factor 1
olmesartan
pioglitazone
prednisolone
rosiglitazone
saxagliptin
tnf-alpha
Monoclonal antibodies
Anthracyclines
Cardiac-acting drugs
Non cardiac-acting drugs

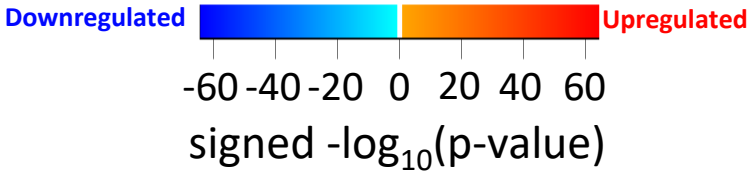
Figure 1

Figure 1. Drug-induced transcriptomic signatures. Six cardiomyocyte cell lines that were derived from induced pluripotent stem cells obtained from six healthy human subjects were treated with one of 54 drugs or vehicle for 48 hours. Drug- and vehicle treated cell lines were subjected to bulk RNAseq, followed by the identification of differentially expressed genes in drug-treated versus vehicle-treated control cell lines. Numbers of significantly differentially expressed genes induced by the different drugs in each cell line are shown (based on an FDR cutoff of 0.1). Each dot visualizes one cell line drug combination.

A

266 Samples/gene expression profiles





1: drugs
2: cell lines
3: drug classes (see figure 1 for color codes)
4: # of sign. DEGs (FDR <= 0.1) (high ←→ low)

B

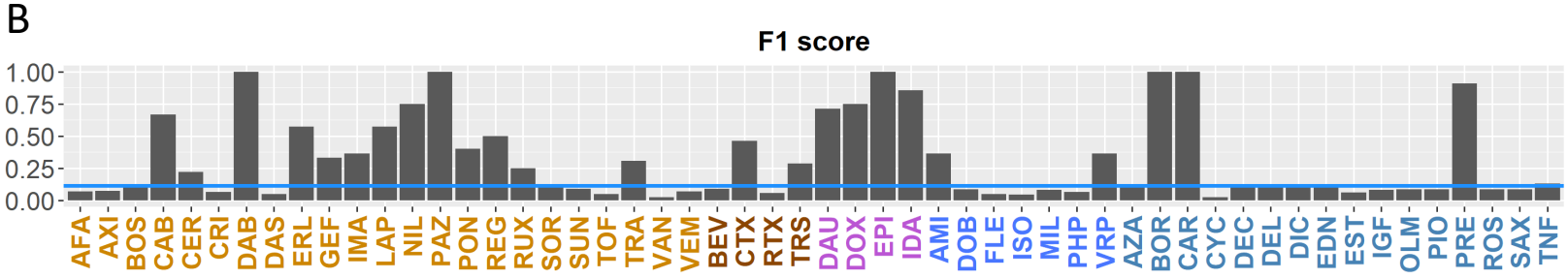
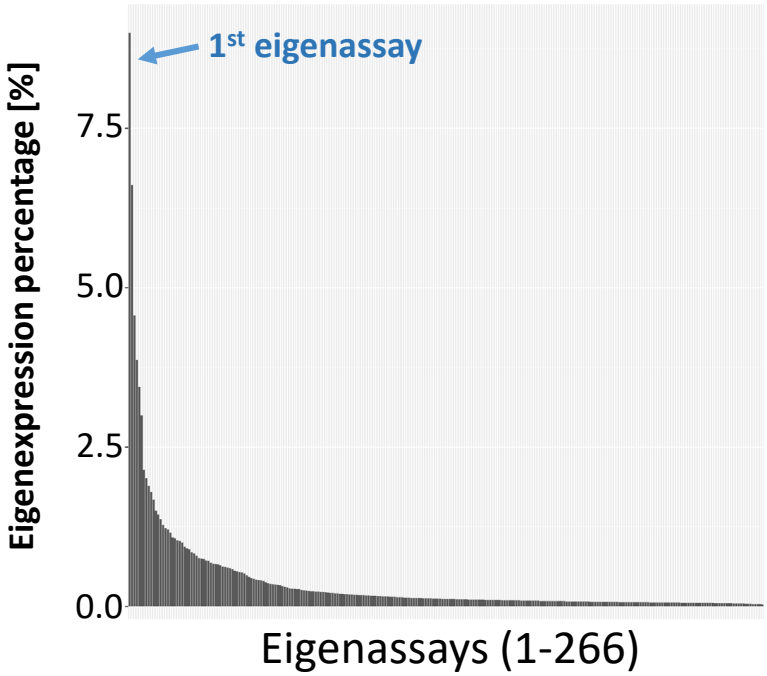


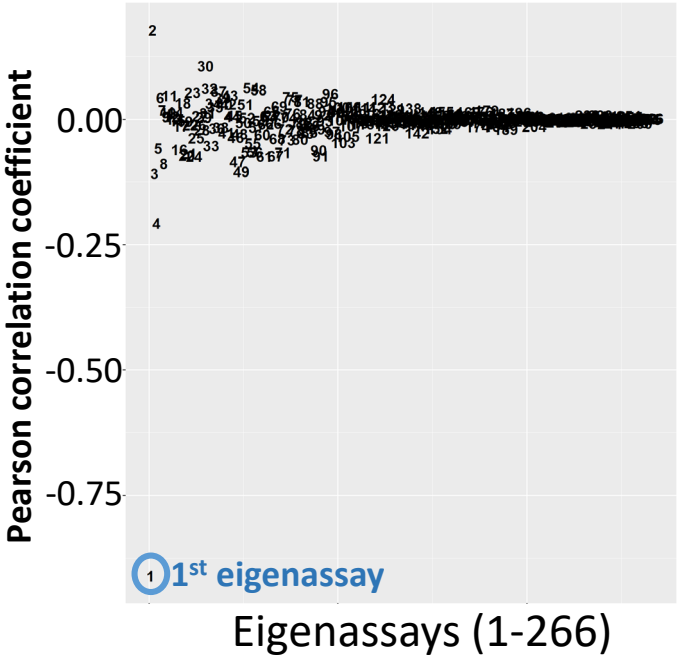
Figure 2

Figure 2. Drug-induced gene expression profiles in iPSC derived cardiomyocytes are mostly cell line specific. (A) Minus $\log_{10}$(p-values) of up- or downregulated genes were defined as positive or negative, respectively. Pairwise Pearson correlation coefficients were calculated based on the generated signed minus $\log_{10}$(p-values) and used to hierarchically cluster the samples. Heatmap of signed minus $\log_{10}$(p-values) was rearranged according to the clustering results. Dendrogram labels show the treated cell lines, drugs used for treatment and the number of significantly differentially expressed genes in each sample as defined by an FDR <= 0.1. **(B)** If the transcriptomic responses are mainly determined by the drug used for treatment, hierarchical clustering should group those samples that were treated by the same drug into the same cluster. To document the efficiency of drug specific clustering, we calculated the F1 score for each drug and cluster. The F1 score is the harmonic mean of precision and recall, i.e. of how many samples in a particular cluster belong to a particular drug and how many samples treated with that drug are within that cluster, respectively. For each drug, we identified the highest F1 score within any cluster that can be obtained by cutting the dendrogram at any height. Results document low maximum F1 scores for most of the drugs, documenting that the used drugs are not the main determinant of the observed gene expression responses.

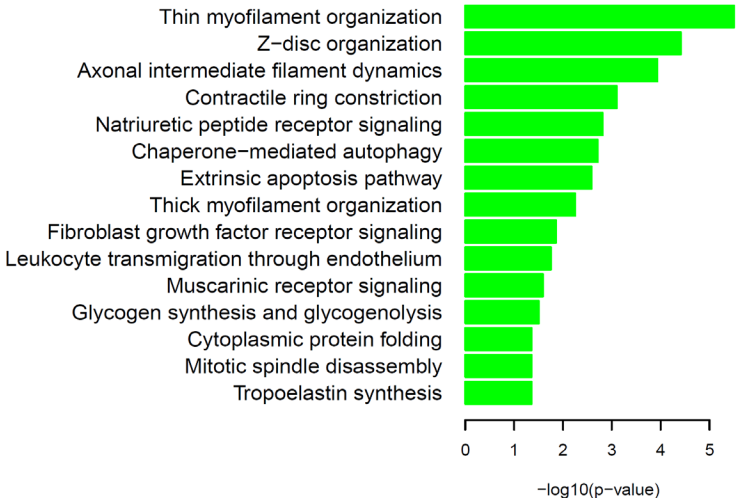
A



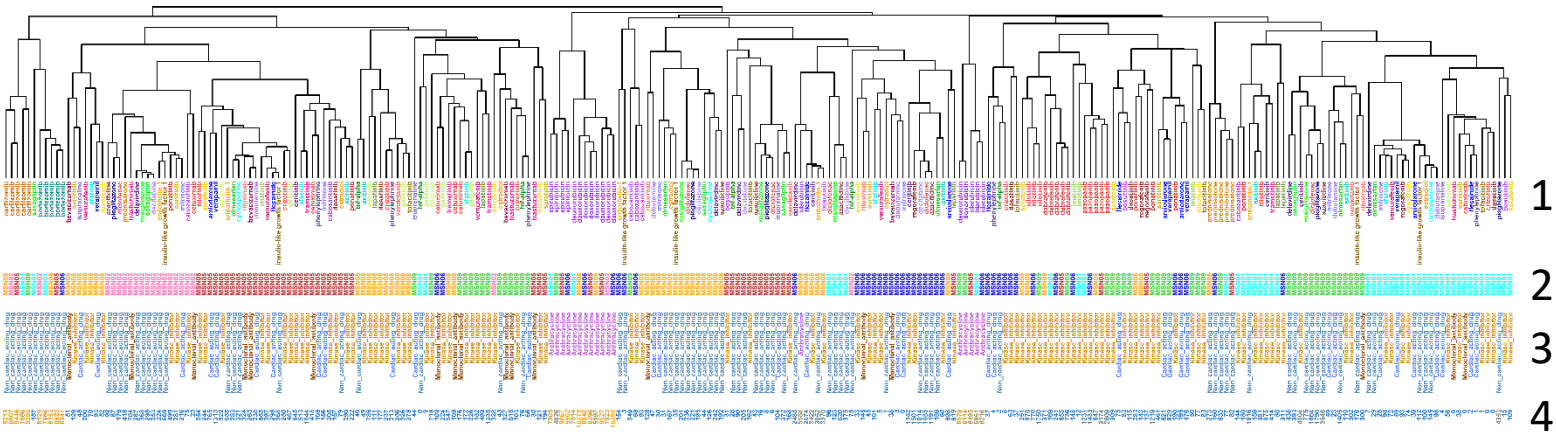
B



C



D



1: drugs
2: cell lines
3: drug classes (see figure 1 for color codes)
4: # of sign. DEGs in full dataset (FDR <= 0.1)
(high ←→ low)

Figure 3

Figure 3. Singular value decomposition identifies an unspecific stress response that is proportional to the magnitude of disturbance. (A) Investigation of the relative contribution of each eigenassay documents that the first eigenassay explains ~9% of the total variance. **(B)** The gene expression profile of each sample is a linear combination of all 266 eigenassays, based on sample specific coefficients for each eigenassay. We correlated the sample specific coefficients associated with each eigenassay with the number of significantly induced differentially expressed genes (FDR <= 10%). Calculated Pearson correlation coefficient of the first eigenassay was -0.91, suggesting that the contribution of the first eigenassay to the gene expression profiles increases with the magnitude of disturbance and is independently of the drug used for treatment. **(C)** To investigate the biology captured by the first eigenassay, we subjected the top 600 genes of this eigenassay to pathway enrichment analysis. Results revealed an enrichment of genes involved in muscle contraction, supporting our hypothesis of a cell type, but not drug specific response that is captured by this eigenassay.

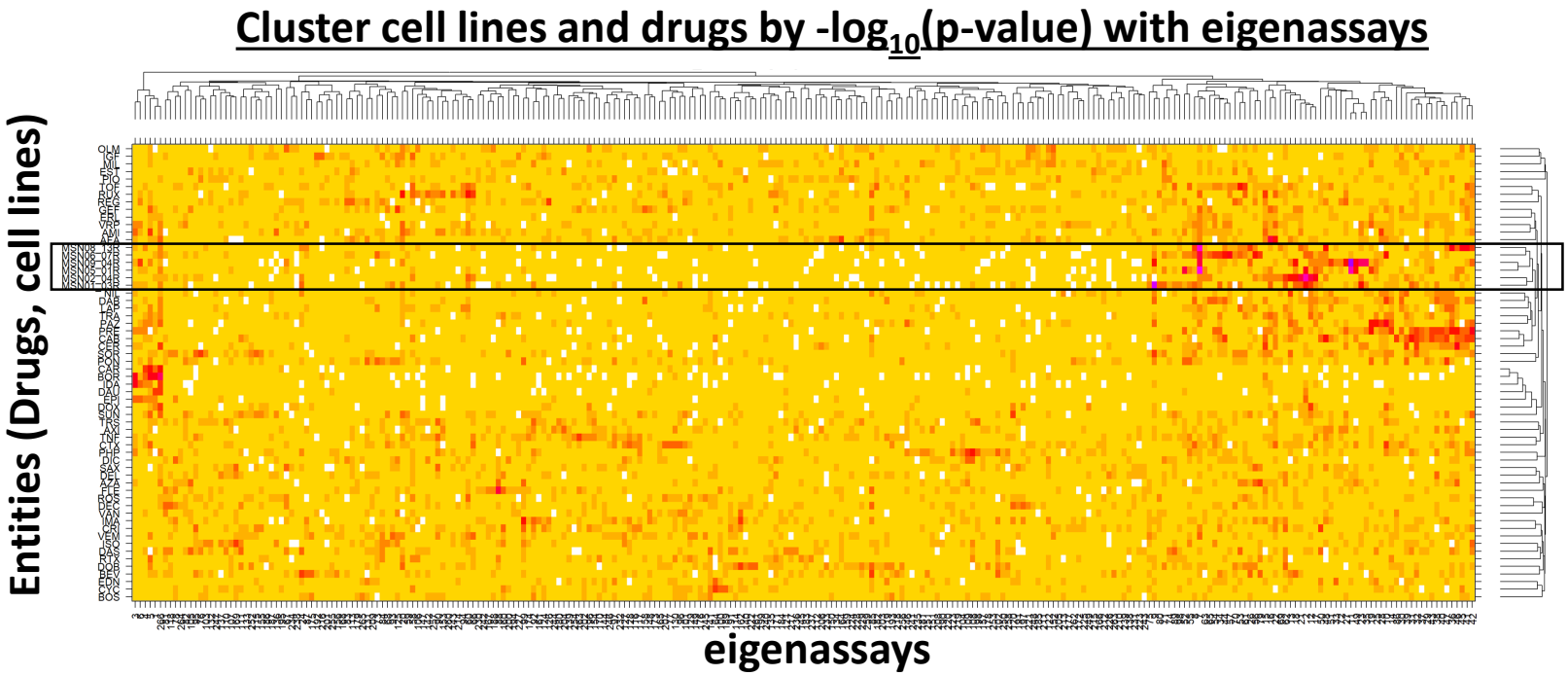
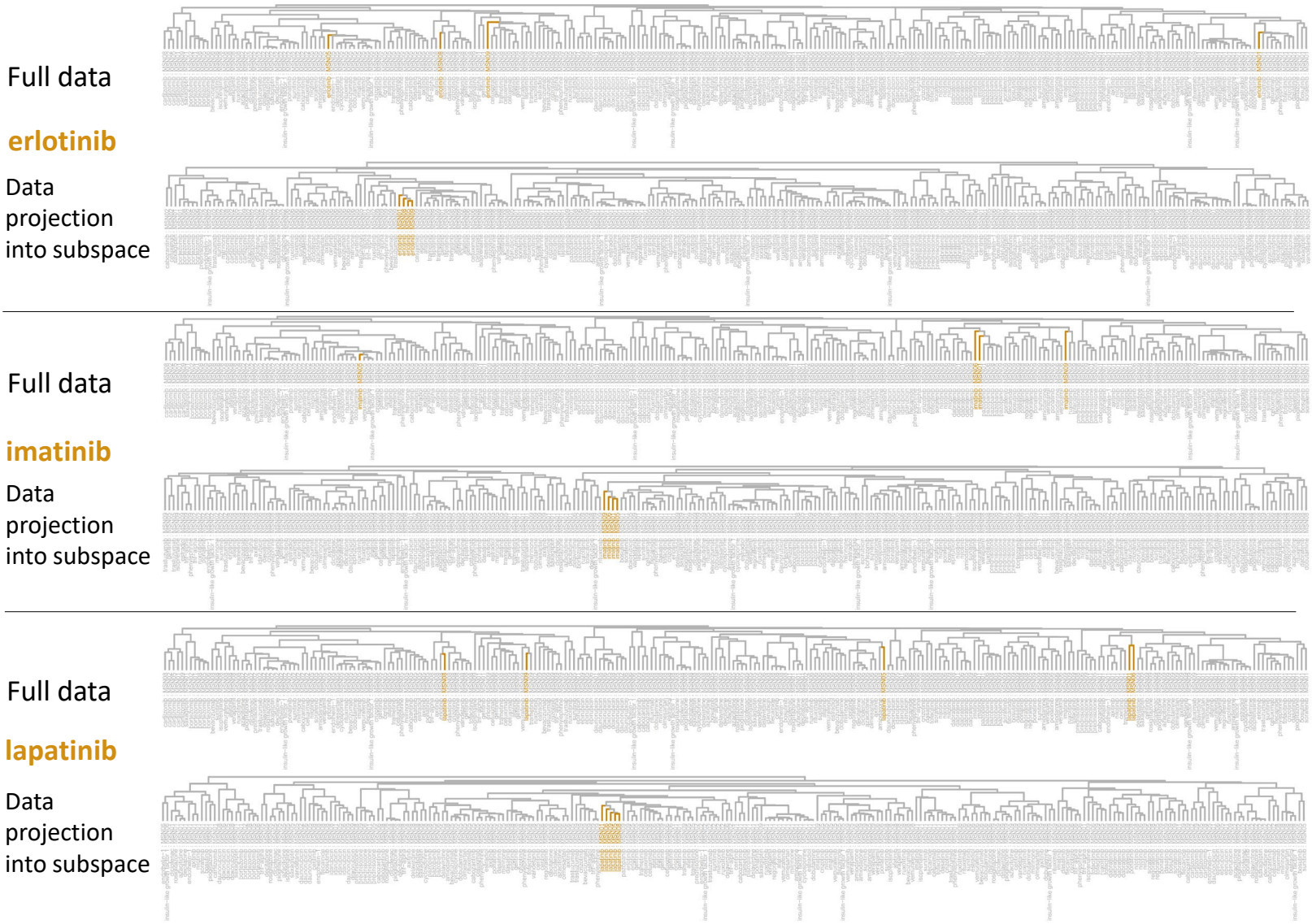

Cluster cell lines and drugs by -log10(p-value) with eigenassays
Entities (Drugs, cell lines)
eigenassays

Figure 4

Figure 4. Identification of drug-specific eigenassays. For each drug and eigenassay we analyzed, if the coefficients assigned to all samples that were treated with this drug are significantly different from the coefficients assigned to all other samples. Consequently, we calculated one p-value for each drug-eigenassay combination. Similarly, we analyzed if the coefficients assigned to each eigenassay and all samples of the same cell line are significantly different from those assigned to the same eigenassay and all other samples. All p-values were transformed into $-\log_{10}$(p-values) and used to calculate pairwise correlation coefficients between all drugs and cell lines, followed by hierarchical clustering of drugs and cell lines based on the coefficients. Initial heatmap of -log10(p-values) was rearranged according to the clustering results. Results document that the six cell lines are grouped into a single cluster, suggesting that drug-specific eigenassays are different from cell line or cell type specific eigenassays. Significance increases from white to red.

A


Full data
erlotinib
Data projection into subspace
Full data
imatinib
Data projection into subspace
Full data
lapatinib
Data projection into subspace

B

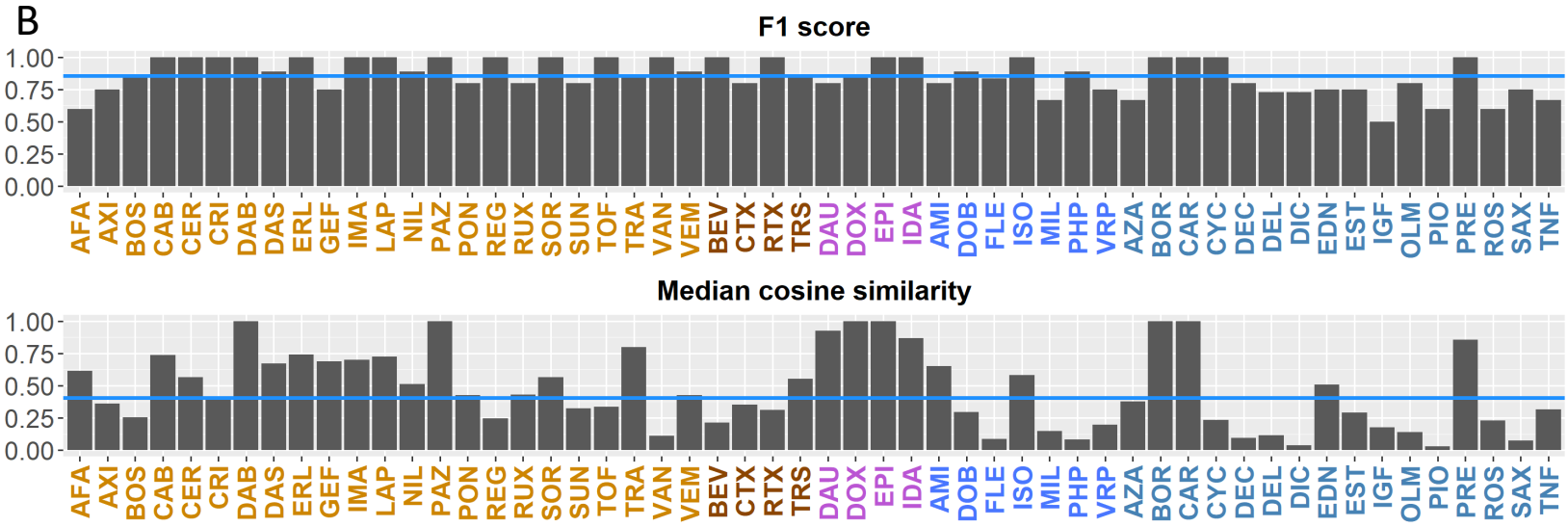
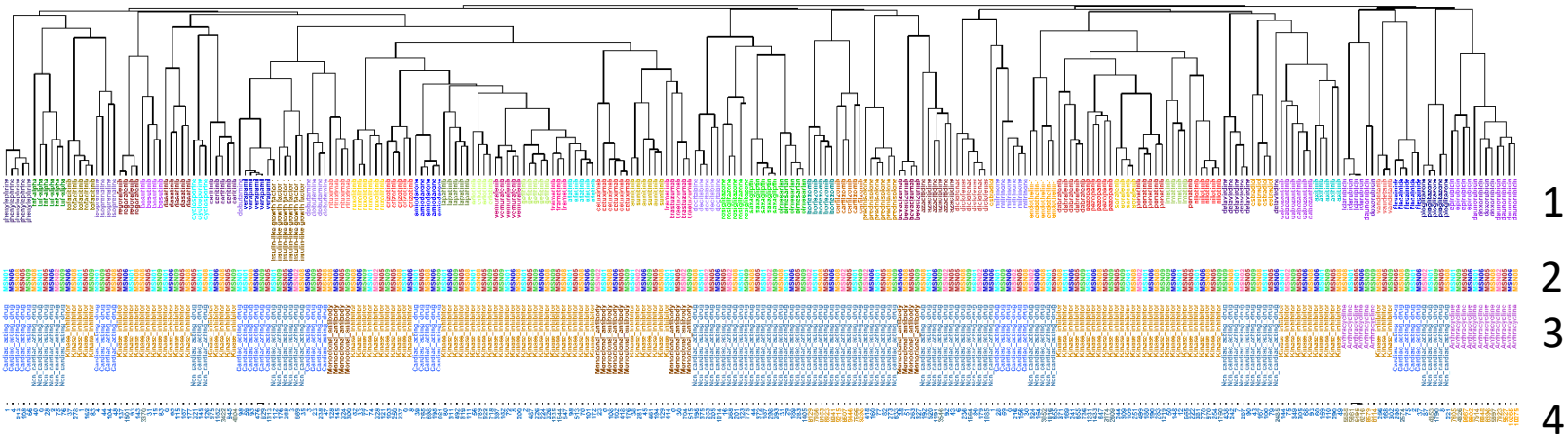

F1 score
1.00
0.75
0.50
0.25
0.00
AFA AXI BOS CAB CER CRI DAB DAS ERL GEF IMA LAP NIL PAZ PON REG RUX SOR SUN TOF TRA VAN VEM BEV CTX RTX TRS DAU DOX EPI IDA AMI DOB FLE ISO MIL PHP VRP AZA BOR CAR CYC DEC DEL DIC EDN EST IGF OLM PIO PRE ROS SAX TNF
Median cosine similarity
1.00
0.75
0.50
0.25
0.00
AFA AXI BOS CAB CER CRI DAB DAS ERL GEF IMA LAP NIL PAZ PON REG RUX SOR SUN TOF TRA VAN VEM BEV CTX RTX TRS DAU DOX EPI IDA AMI DOB FLE ISO MIL PHP VRP AZA BOR CAR CYC DEC DEL DIC EDN EST IGF OLM PIO PRE ROS SAX TNF

Figure 5

C



1: drugs
2: cell lines
3: drug classes (see figure 1 for color codes)
4: # of sign. DEGs in full dataset (FDR <= 0.1)
(**high** ←→ **low**)

Figure 5

Figure 5. Identification of drug-specific subspaces. (A) After removal of the first eigenassay, we aimed at identifying drug-specific subspaces that allow characterization of drug-specific gene expression responses. For this, we ranked all eigenassays by decreasing significance for each drug, generating 54 ranked lists. Any combination of eigenassays defines a subspace that captures a particular amount of the original data. For each drug, we generated 264 subspaces that were spanned by the top three to 266 ranked eigenassays for that particular drug. We projected the full data into each of these subspaces, followed by pairwise correlation analysis, hierarchical clustering and identification of the largest F1 score for that drug, as described in figure caption 2. Calculation of the median cosine similarity between the full gene expression profiles of that drug and their projections into each subspace documented how much of the initial full response is preserved within each subspace. We combined both values to a selection score (0.95 x F1 score + 0.05 median cosine similarity) and defined that subspace with the highest selection score as the drug-specific subspace. Shown is the clustering behavior of three example drugs in the full dataset (after removal of the first eigenassay) and in the identified drug-specific subspaces. **(B)** The figure shows the identified F1 scores and median cosine similarities for each drug in each drug-specific subspace. Blue lines indicate median values of bar heights. **(C)** Drug-specific gene expression profiles within each of the drug-specific subspaces were merged and subjected to pairwise correlation analysis and hierarchical clustering, as described in figure caption 2.

Upregulated

Downregulated

erlotinib

	MSN01	MSN05	MSN08	MSN09
Serine and glycine metabolism	2	1	2	1
ER unfolded protein response pathway	1	2	1	2
Glutamate and glutamine metabolism	9	5	24	7
Thin myofilament organization	7	16	20	4
Aspartate and arginine metabolism	8	4	–	5
Protein polyubiquitination	–	3	8	8
Inhibition of apoptosis	–	18	4	29
Non–vesicular phospholipid transport	21	–	–	4
Fibrillar collagen core structure organization	28	–	3	–
Cholesterol synthesis	3	–	–	–
Serotonin inactivation	4	–	–	–
Transmembrane glucose transport	5	–	–	–

	MSN01	MSN05	MSN08	MSN09
DNA replication initiation	2	1	2	1
Centrosome separation	3	2	1	3
Eukaryotic kinetochore dynamics	1	4	3	2
DNA replication elongation	9	3	4	4
Metaphase to anaphase checkpoint	4	7	5	5
Mitotic spindle assembly	6	5	6	6
Collagen fibril organization by fibril–associated bridges	5	24	10	12

nilotinib

	MSN01	MSN05	MSN06	MSN08	MSN09
Centrosome separation	1	1	1	1	1
Mitotic spindle assembly	4	3	2	3	3
Serine and glycine metabolism	8	4	5	2	2
Eukaryotic kinetochore dynamics	3	8	4	6	4
Metaphase to anaphase checkpoint	2	6	7	5	6
Contractile ring constriction	12	5	6	13	5
Sister chromatid segregation	5	15	8	7	17
ER unfolded protein response pathway	62	10	3	4	9
Cholesterol synthesis	34	2	11	–	10

	MSN01	MSN05	MSN06	MSN08	MSN09
Collagen fibril organization by fibril–associated bridges	3	2	4	34	1
Retinol metabolism	10	5	12	10	8
WNT–Beta–catenin signaling pathway	39	4	1	1	9
Epithelial intermediate filament dynamics	5	1	41	6	4
Extracellular matrix breakdown and membrane shedding by adamalysins	24	3	6	23	3
Z–disc organization	11	51	2	2	2
Fibrillar collagen core structure organization	1	42	7	7	48
Collagen fiber crosslinking	34	28	30	3	31
Extracellular matrix breakdown by serine proteases	4	35	37	37	38
Inhibition of amyloid aggregation, amyloid degradation and uptake	2	–	20	19	18
Desmosome organization	64	–	3	4	14
Vascular endothelial growth factor receptor signaling	42	–	5	5	38

idarubicin

	MSN01	MSN05	MSN06	MSN08
RNA editing	2	2	2	1
Extrinsic apoptosis pathway	1	1	1	6
Nucleotide excision repair	3	3	3	2
Intrinsic apoptosis pathway	5	5	5	3
Cell cycle arrest in response to DNA damage	4	4	4	7
Desmosome organization	6	6	6	4
Bicarbonate transmembrane transport	6	6	6	4

	MSN01	MSN05	MSN06	MSN08
Chaperone–mediated autophagy	5	4	5	4
DNA replication elongation	4	2	4	10
Centrosome separation	2	1	2	20
Thin myofilament organization	13	13	14	2
Fibulin receptor signaling	28	26	28	4
Lipid droplet mitochondria interaction	34	32	34	5
Eukaryotic kinetochore dynamics	1	3	1	>
Metaphase to anaphase checkpoint	3	5	3	>
Amyloid plaque organization	–	68	–	1

ruxolitinib

	MSN01	MSN02	MSN05	MSN06	MSN08	MSN09
Fibrillar collagen core structure organization	1	1	1	2	1	2
Osteonectin receptor signaling	3	24	20	4	4	4
Extracellular matrix breakdown and membrane shedding by adamalysins	2	6	18	16	17	6
Retinol metabolism	30	61	3	3	10	3
Extracellular matrix breakdown by matrix metalloproteases	4	2	6	40	–	44
Macrophage migration inhibitory factor signaling	–	30	23	5	23	23
Cholesterol synthesis	–	–	2	1	2	1
Collagen fibril organization by fibril–associated bridges	20	38	–	27	5	31
Axonal intermediate filament dynamics	–	4	34	33	32	38
Cardiomyocyte pacemaker current generation	–	5	38	–	6	7
Cellular iron uptake and export	–	61	46	10	3	53
Protein polyubiquitination	–	13	4	55	46	–
Desaturation of fatty acids	–	–	34	6	–	5
Extracellular matrix breakdown by serine proteases	24	–	5	–	–	38
Proteasomal regulatory particle organization	–	4	–	–	–	–

	MSN01	MSN02	MSN05	MSN06	MSN08	MSN09
JAK–STAT signaling pathway	6	1	1	1	3	1
Leptin receptor signaling	10	4	6	6	8	4
Microtubule crosslinking and bundling	5	10	12	10	2	8
Z–disc organization	2	8	10	27	1	28
Inhibition of amyloid aggregation, amyloid degradation and uptake	7	30	36	4	6	2
Serine and glycine metabolism	3	20	5	21	86	3
Notch receptor signaling	92	12	3	14	19	13
Antigen presentation via MHC class I molecules	8	18	23	3	81	20
Glycolysis and Gluconeogenesis	1	3	19	92	57	16
Aspartate and arginine metabolism	22	18	4	70	25	73
Hepatocyte growth factor receptor signaling	–	–	2	2	4	26
Inhibition of matrix metalloproteinase and adamalysin activity	–	30	–	4	6	30
CCN intercellular signaling protein family receptor signaling	4	6	–	42	–	43
Epithelial intermediate filament dynamics	18	65	75	62	5	–
Semaphorin signaling	38	2	81	>	–	12

Figure 6

Figure 6. Drug-specific gene expression responses describe reasonable drug effects. Drug-specific gene expression profiles that were obtained by projecting the full gene expression profiles into the drug-specific subspaces were ranked by decreasing absolute -log10(p-values). Up- and downregulated genes among the top 600 genes were separately subjected to pathway enrichment analysis using Fisher's Exact Test and the Molecular Biology of the Cell Ontology. Predicted up- and downregulated subcellular processes (SCPs) were ranked by significance. Top five ranked up- and downregulated SCPs for three kinase inhibitors and one anthracycline are shown. If a pathway was identified in at least one cell line for the same drug, pathway ranks are shown for all cell lines, even if they exceeded five. Fields are colored based on pathway ranks, any ranks above 15 are colored gray. Minus signs indicate that the pathway was not identified at all.