WORKING PAPER

Clinically-weighted transcriptomic signatures for protein kinase inhibitor associated cardiotoxicity

JGC van Hasselt[1,2], J Hansen[1], Y Xiong[1], J Shim[1], A Pickard[1], G Jayaraman[1], B Hu[1], M Mahajan[3], J. Gallo [1], EA Sobie[1], MR Birtwistle[1], EU Azeloglu[1], R Iyengar[1]

1. Department of Pharmacological Sciences and Systems Biology Center New York, Icahn School of Medicine at Mount Sinai, New York, NY, United States.
2. Department of Pharmacology, Leiden University, Leiden, Netherlands.
3. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States.

ABSTRACT

Cardiotoxicity (CT) involving diminished cardiac contractility and heart failure is a major adverse event associated with otherwise efficacious protein kinase inhibitors (KIs). Here, we sought to develop clinically-weighted transcriptomic signatures to predict risk of CT and to better understand the biological processes associated with CT risk. We obtained transcriptome-wide response profiles in four human primary cardiomyocyte cell lines that were treated with 22 different KIs using mRNA sequencing with 3' digital gene expression. The FDA Adverse Event Reporting System was used to derive relative risk scores for four types of CT for different KIs. We used elastic net regression to associate these transcriptomic profiles with KI-associated risk scores for CT subtypes to obtain clinically-weighted transcriptomic signatures, which showed good predictive properties (cross-validation $R^2 >0.87$). Our clinically-weighted transcriptomic signatures for KI-associated CT may be of relevance in early drug development for the prediction of KI-associated CT.

INTRODUCTION

Protein kinase inhibitors (KIs) are a clinically important class of anticancer agents[1,2]. There are currently more than 28 KIs[3] on the market, and more than 100 KIs in clinical development[4]. There are significant concerns regarding the safety profile of KIs. Cardiotoxicity (CT) is one clinically important adverse event associated with some KIs[5–7]. Typically, KI-associated CT manifests as loss of contractile function, which can subsequently lead to heart failure[8]. The human 'kinome' consists of more than 500 protein kinases[9]. Given that many KIs are not highly selective in the protein kinases they target[10], the inhibition of any of these kinases in healthy cell types,

1

including cardiomyocytes, may potentially lead to adverse drug effects such as CT[11]. For specific KIs, *in vitro* and *in vivo* studies have identified potential mechanisms for CT[12], including mitochondrial function [6,13,14], endoplasmic reticulum stress response[14] and AMPK inhibition[15]. Overall, however, the mechanisms of KI-induced CT are still poorly understood[12]. Furthermore, characterization of the link between clinically observed CT phenotypes and transcriptomic changes in cardiomyocytes could have tremendous prognostic value. Therefore, there is a need to better characterize the molecular changes of KI associated CT on a genome-wide basis.

The use of a systems pharmacology approach to understand and predict adverse events has received increasing interest[16,17], and several successful case examples have been described[18,19]. Such approaches typically rely on appropriate global characterization of drug-induced molecular changes in affected cells and tissues of interest. However, for KI-associated CT, a systematic assessment of molecular changes associated with KI treatment is lacking. Addressing this gap, as part of the NIH-funded Library of Integrated Network Based Cellular Signatures (LINCS), the Drug Toxicity Signature Generation Center (DToxS; www.dtoxs.org) is currently generating large-scale transcriptomics and proteomics profiling datasets in clinically-relevant, human-derived cell lines, which are perturbed with large numbers of different drugs, including cardiomyocytes and KIs. Here, we describe how transcriptomic profiles of 22 KIs in four human primary cardiomyocyte cell lines were analyzed in conjunction with clinical data from the FDA Adverse Event Reporting System, in order to derive clinically-weighted transcriptomic signatures that may allow ranking prospective KIs for their relative risk of CT.

RESULTS

Limited overlap of differentially expressed genes across different KIs in cardiomyocytes

We used mRNA sequencing with 3'-digital gene expression generated from four primary human cardiomyocyte cell lines that were treated with up to 22 KIs at their respective maximum therapeutic concentrations for 48 h (**Table 1, Table S3**) for our analyses. Mean differential gene expression fold-change values were computed across cell lines and replicates for all KIs. After ranking by absolute gene expression fold-change and keeping the top 500 genes, we computed the Jaccard index across these KI-associated top ranked gene expression profiles, which indicated limited similarity given the Jaccard index of <0.25 (**Figure 1A**). The genes that were most commonly present in the top 500 ranked gene list across drugs ranked by their frequency of occurrence is shown in **Figure 1B**, i.e. gene TC2N was present in these top ranking genes for all KIs except for cabozantinib. Principal component analysis showed variable gene expression patterns for 9 KIs, while for the remaining KIs little variation in gene expression was seen (**Figure 1C**), where PC3 appeared most discriminating.

Clinical risk profiles derived from FAERS show differences in risk profiles between cardiotoxicity subtypes

To obtain estimates of clinical risk of KI-associated CT, we analyzed data from the publicly accessible FDA Adverse Event Reporting System (FAERS) (**Figure 2A**). We distinguished between four relevant subtypes of CT that can be identified from the FAERS database: hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), ventricular dysfunction and heart failure (VDHF), and cardiotoxicity undefined (CTU). Subsequently, reporting odds ratios (RORs) and Z-scores were derived based on the relative frequencies of CT, for each KI present in the transcriptomics dataset (**Figure 2B**). For several KIs, there are observable differences in the risk for different subtypes of CT. The Z-scores were used as single metric to quantify CT risk, as it includes the confidence interval for the ROR. Here, the estimates for the ROR and associated Z-score indicate relative ranking of KI-associated toxicity, and not absolute risks of CT. Finally, using multivariate logistic regression analyses we estimated the effect of age and sex at 1.02 year^{-1} (2.6 CV%) and 1.47 (1.3 CV%), respectively. The low inter-KI variation of <2.6 CV% suggest that age and sex have negligible impact on individual KI-associated CT risk estimates.

Distinct and predictive signatures for 4 types of cardiotoxicity

Subsequently, we aimed to associate our KI-wide mean fold change gene expression profiles with the KI-specific clinical risk scores of four subtypes of CT (i.e. HCM, DCM, CTU, VDHF). Given the limited similarity between top ranking gene expression profiles across KIs (**Figure 1A-C**), the entirety of the gene expression profile across KIs were considered as potential predictors for KI-associated CT risk. As such, KI-specific expression profiles of 9,281 genes were available as potential predictors for 22 KI-specific CT risk scores. To identify genes most strongly associated with CT risk we used an elastic net penalized regression approach, which is suited to identify predictors whilst limiting the risk of overfitting[20]. Separate models were fit for all four derived subtypes of CT. A two-stage regression analysis was performed (**Figure 3**). In the first stage, we constructed bootstrap datasets obtained after random resampling of KI risk and associated gene expression profiles, which were subsequently fit as elastic net models. This first step was performed to identify gene-based predictors that could consistently predict CT risk and contributed significantly to the prediction of this risk. The bootstrap analysis resulted in stable selection of potential predictors after 250 bootstrap samples (**Figure S1**). Predictors to be included in the final elastic net regression model (**Supplemental material 4**) were selected based on their minimal root mean squared prediction error (RMSE, **Figure S2-3**) after cross validation. Since we had only a limited number of KIs available, we explicitly chose to not keep apart any data for a full external validation. Instead, we evaluated model predictions using repeated random 3-fold cross validation for each of the four subtypes of CT, which resulted in high R^2 prediction metrics (>0.87) (**Figure 4A**). The selected gene expression-based predictors in the final elastic net model consisted of 42, 44, 44 and 85 genes for HCM, CTU, VDHF and DCM, respectively (**Table S2, Figure 4B**). These sets of genes define our clinically-weighted transcriptomic signatures for cardiotoxicity; their

3

associated gene expression profiles are shown in **Figure 4C**. The importance of each of the selected genes for the prediction of CT risk was not equal (**Figure 4B**). CT subtype DCM required 85 genes for optimal predictive performance, whereas for all other subtypes <44 genes were sufficient. Potentially, this is because not enough KIs with a diverse CT profile were available for deriving an informative gene signature. When assessing the similarity between each of the 4 clinically weighted gene expression signatures, the largest overlap was present between DCM and HCM (4 genes), VDHF and DCM (4 genes), and VDHF and HCM (2 genes) (**Figure 4D**). The genes FASTKD2 (FAST kinase domain containing protein 2) and MED19 (Mediator Of RNA Polymerase II Transcription, Subunit 19) were present in the signatures for HCM, VDHF and DCM. FASTKD2 plays a role in energy balance regulation in mitochondria under stress[21]. This could indeed be of relevance for cardiotoxicity, because mitochondrial dysfunction has been associated with cardiotoxicity[13,22], and this finding may warrant further detailed analysis.

Signatures for the 4 types of cardiotoxicity are enriched for different biological processes

In order to identify the potential functional consequences of our clinically-weighted transcriptomic signatures, we performed enrichment analyses based on the KEGG database, estimating over-representation of signature genes using Fisher's exact test (**Figure 5**). Biological processes that appeared to be affected for all four types of CT included metabolic pathways, and common subcellular processes such as ribosome function, endocytosis, proteolysis, and splicing. However, similar to the differential clinical risks for the different types of cardiotoxicity, we observed distinct patterns in enriched subcellular processes for specific CT subtypes. Such differences include regulation of the actin cytoskeleton and oxidative phosphorylation for DCM, protein processing in the endoplasmic reticulum for HCM, VDHF), and apoptosis (DCM) amongst others (**Figure 5**).

DISCUSSION

The occurrence of drug treatment associated cardiotoxicity leading to decreased cardiac contractility lags behind the therapeutic effects of the drugs and may only be observed in a subset of the patients using the drug. These two factors raise the question of whether it would be possible to obtain early cell based indicators of potential for drug toxicity. This study was designed to address this question. The approach we have used is to determine we can associate drug treatment induced gene expression patterns and the clinical risk for the adverse events of interest. We estimated clinical risk from the FAERS database. Our use of FAERS is arguably a relevant and unbiased approach for the quantification of CT risk, because this data reflects unselected, real-life patients cohorts and may some of the potential bias associated with cohort selection associated with tightly controlled clinical trials. It is possible that risk metrics derived from controlled clinical trials can underestimate adverse event risks due to selective inclusion criteria of patients to demonstrate therapeutic efficacy, cohort size and duration of the trial.

Nevertheless there are limitations to the FAERS database as well that we have discussed and addressed in previous work[18], and similarly address in this study. In particular, we confirmed that demographics such as age and sex were not variable across different KIs. Moreover, as the KIs studied are cancer drugs, some similarity between populations is likely.

This project associated KI-associated transcriptomic response profiles generated from cultured human primary cardiomyocytes with clinical risk scores for different subtypes of CT in order to obtain a reduced set of genes that may predict the relative risk for KI-associated CT. The clinically weighted transcriptomic signatures consisted of <44 genes for HCM, CTU, and VDHF, whilst the signature for DCM consisted of 85 genes. The signatures showed good prediction of CT risk based on the cross validation analysis (**Figure 4A**). Using the clinically weighted signatures and the association regression coefficients identified in the elastic net model (**Supplemental material 4**), the relative risk for CT can be predicted. Here, the risk metric does not reflect the absolute risk for developing CT. Rather it reflects the relative risk for a subset of patients for which drug-associated adverse events were reported. As such, the risks predicted by our signatures and associated regression model can be used in drug development to rank the risk of potentially novel KIs with respect to the current ranking of existing KIs with better characterized clinical risks for CT. The issue of whether or not a KI can be considered as cardiotoxic has some underlying complexity. Some KIs may result in cardiotoxicity, but the severity and duration of CT may be different between patients and between different KIs. Therefore, we decided to develop a continuous risk metric in preference to a binary classification of drugs as CT or not. Secondly, it is unclear if all KIs induce CT through similar mechanisms and to what extent ultimate clinical pathologies are similar. Indeed, the FAERS database allows us to distinguish between different types of CT, rather than lump all potential types of CT together (**Figure 2B**). Nonetheless, the clinical classification of KI associated CT is an area that needs better resolution, and it is likely as EHR records become increasingly available that the data will become more detailed with respect to pathophysiology. Such improved clinical data can greatly enhance the value of analyses such as ours and aid in the drug development processes by contributing to early go/no-go decisions.

At the cellular level, commonly enriched subcellular processes associated with the clinically-weighted transcriptomic signatures included various metabolic processes, pathways related to oxidative phosphorylation, endoplasmic reticulum (ER) effects, and apoptosis. In addition, a number of subcellular processes related to transcription (ribosome, spliceosome), protein degradation and endocytosis were affected. Other recent studies into CT of anthracyclines and KIs agree with our findings that changes in metabolic processes[23,24], ER effects[14], oxidative phosphorylation[13,25] AMPK signaling[26], cell adhesion[27], cell survival pathways[12], and the FoxO-related transcription[28] are likely to be involved. It is possible that gene expression profiles associated with specific

5

subcellular processes could be co-expressed and hence highly correlated. However, due to the nature of the elastic net regression algorithm used, highly correlated genes are unlikely to be selected simultaneously, limiting over-representation of the identification of associated subcellular pathways.

Besides variation in drug exposure, pharmacodynamic inter-individual variation in severity of CT is likely. The current group of four cell lines are not of sufficient size to realistically capture such human variability. Therefore, in our analysis, we used mean fold change gene expression profiles across multiple cell lines. The resulting averaged gene expression profiles thus reflect relatively consistent changes in gene expression across cell lines, i.e. changes in gene expression that are less likely to be highly variable across individuals, yet may result in consistent predictors across the population. Given that the FAERS CT risk scores also reflect a population-level CT risk, the use of these mean values in fold-change gene expression values is a reasonable starting point for our analyses.

The experimental underpinning of the transcriptomic profiles generated in this study make them likely to be of value in selecting drug candidates for human use. Our analysis is based on primary human derived cardiomyocytes. Although these cell lines do have phenotypic limitations due to dedifferentiation[29], the signatures obtained from the cells could be relevant for prediction of clinical drug effects. These cell lines may be reflective of specific human cardiac pharmacology[30,31], even though further characterization and standardization is still needed. Our analyses used drug exposures at estimated therapeutic concentrations of the individual KIs, rather than using the same concentrations for all KIs. We used the reported maximum concentrations as reported in literature (Table S3). Here, we did not correct for protein binding, but, we consider that given the typical protein binding of >95% of these KIs, the concentration used may reflect patients with over-exposure to the KI. As such, the concentrations used are relevant to assess transcriptomic changes that are likely related to early changes in subcellular processes associated with the adverse event of interest.

We anticipate that clinically weighted transcriptomic signatures such as developed in this study may be of relevance to guide safety assessment in early drug development. Unlike the relatively well-established assessment of electrophysiological safety issues such as QT prolongation, the assessment of non-QT type of cardiotoxicity associated with KIs[12] and other novel drugs[32], lack reliable biomarkers. Transcriptomic signatures could help fill this gap. Transcriptomic analyses may also guide the discovery of subcellular pathway based biomarkers that can be measured in both preclinical model systems as well as in patients in surrogate cells. Data on biomarker levels could then be included in PK-PD models to personalize dose regimens that limit risk of CT[33–35], before the

appearance of imaging based indicators such as the left ventricular ejection fraction. The current study represents a starting point and many more studies both experimental and computational will be needed to utilize transcriptomic studies in drug development and clinical practice.

METHODS

Experiments to identify differentially expressed genes in human cardiomyocytes treated with KIs.

Adult human cardiomyoctes were purchased from PromoCell and grown in culture per manufacturer's instructions. Four different cell lines representing different human subjects were studied. Cells were treated with a single dose at the estimated therapeutic concentration used in humans for 48 hrs (Table S3). After drug treatment, the cells were lysed, RNA was collected using TRIzol, and gene expression profiles were measured using the 3' digital gene expression method[36,37]. Details of the experimental protocols have been described in another study[38] and step-by-step standard operating procedures for the various experiments are available on www.dtoxs.org.

Sequence alignment and processing of gene expression data

The raw sequences were demultiplexed. Combined standard RNA-seq sequencing files were aligned to the reference human genome provided by the Broad Institute[36], using the Burrows-Wheeler Alignment (BWA)[39] software. Details of these computational procedures are described elsewhere[38], and step-by-step protocols are available on ww.dtoxs.org. The resulting alignment files were parsed to identify the fragments with acceptable alignment quality, to remove duplicate fragments, and to assign accepted fragments to corresponding genes. The resulting read-count (i.e., transcript count) table was then subjected to correlation analysis at each treatment condition, to identify and remove outlier samples, determined by predefined thresholds. The gene read-count tables were then subjected to differential gene expression analysis using the R package EdgeR[40]. The resulting normalized and log-transformed fold-change gene expression values for each sample were deposited for public access to the DtoxS repository (www.dtoxs.org).

Processing and exploratory analysis or gene expression data

The mean log transformed gene expression fold-change value was calculated across all cell lines for each individual KI. The resulting matrix of gene fold changes values by KIs was used for the regression analysis. To obtain insight in the general patterns present in this KI-perturbed transcriptomics dataset, we generated rankings of the top 500 genes for each drug, by their absolute fold change value. For each of these KI-associated rankings we determined the frequency of these changes being also present in the ranking of other drugs, e.g. the similarity in top ranking gene expression. This was visualized using the Jaccard index, and by plotting the most highly drug-connected genes against the associated drugs. Finally we performed on the mean fold-change values for each drug

a principal component analysis for the first 3 principal components to further assess similarity between drugs in their gene expression values.

Calculation of clinical risk scores

The FDA Adverse Event Reporting Systems (FAERS) raw data files were downloaded from the FDA website (all report files up to 2015 Q3). Adverse events in the raw data files were mapped to the MEDDRA dictionary[41]. Subsequently we defined four types of CT, that were related to either hypertrophic cardiomyopathy, dilated cardiomyopathy, ventricular dysfunction and heart failure, and undefined cardiotoxicity. Electrophysiological forms of CT were excluded. The selection of the selected CT subtypes is provided in **Table S1**. Since drug names in FAERS are of low quality, i.e. they contain many spelling mistakes, mixed brand names, or combinations of drugs associated with a single adverse event, we performed extensive curation on the drug names converting the drug name records to generic drug names. This was done using dictionaries from the FDA and DrugBank. We also applied a conservative fuzzy match algorithm to detect drug names with spelling errors. Through iterative manual checking and automated string matching using R we derived a curated database of drug names associated with adverse events. Subsequently, the frequency of CT events in the 4 predefined subgroups, other non-CT adverse events were computed for individual KIs included in our experiments, and for all other drugs. Based on this we calculated a reported odds ratio (ROR) as follows, ROR = (a/b)/(c/d), where a, b, c, and d are the number of individuals with a drug-associated adverse event, where a is the number of individuals who received drug X, and had the adverse event of interest, b is the number of individuals who received drug X and did not have the adverse event of interest; c is the number of individuals who did not receive drug X but some other drug, and had the adverse event of interest; and d is the number of individuals who did not receive drug X and did not have the adverse event of interest. Subsequently, the standard error of the log transformed odds ratio (SE_{logROR}) was computed as follows: SE=sqrt(1/a+1/b+1/c+1/d). The Z-score was then computed as Z=log(ROR)/SE_{logROR}. To assess the impact of age and sex as confounders variable for different KIs, we performed multivariate logistic regression analyses for each of the individual KIs in combination with factors estimated for age and sex. For this we used the standard generalized linear model available in R.

Elastic net regression analysis

The FAERS-derived risk (Z-scores) for each of the four types of CT was regressed against the KI associated vectors of mean fold change values across the four cell lines. We chose to use Z-scores as metric reflecting CT-risk, as this metric takes into consideration the uncertainty of the reporting odds ratio. Prior to regression we filtered the gene lists by their presence in either the Gene Ontology Biological Processes, KEGG, Reactome, or WikiPathways databases, in order to exclude genes for which no biological function has been established. Subsequently we generated 250 bootstrap datasets with replacements. We did separate bootstrapping for each KI. Each of these bootstrap datasets was fit using an elastic net regression model (R package glmnet). The selected

genes that were selected as predictors (i.e. non-zero regression coefficient) and the scaled values of the gene-associated coefficients were saved for each bootstrap dataset. Across all bootstrap datasets the relative frequency of selection of gene-based predictors, and the mean scaled coefficient value of these coefficients was computed. We then calculated the product of the mean frequency and scaled coefficient value, rank predictors by their importance with respect to robustness (selection frequency) and their importance. Different upper percentiles of these rankings were regressed against each CT subtype and evaluated using 3-fold cross validation. The selection percentile resulting in optimal prediction errors (RMSE) was then used to select a subset of gene based predictors, and to select the model that generated the final gene expression signatures for each CT subtype. The selected predictor genes were then visualized ranked by their relative importance, and by their mean fold change values as clustered heatmaps.

Enrichment analyses

Enrichment analysis was performed based on a one-tailed Fisher's exact test using R, in order to identify enrichment of specific genes in predefined gene lists, as present in the KEGG database. We performed enrichment of the selected gene-based predictors for the four types of CT, as present in the final signatures. Diseases were excluded from the KEGG list of processes (e.g. diabetes, depression, cancer), in order to only evaluate general biological processes or pathways.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Cohen, P. The role of protein phosphorylation in human health and disease: Delivered on June 30th 2001 at the FEBS meeting in Lisbon. *Eur. J. Biochem.* **268**, 5001–5010 (2001).
2. Giamas, G. *et al.* Kinases as targets in the treatment of solid tumors. *Cell. Signal.* **22**, 984–1002 (2010).
3. Knapp, S. & Sundström, M. Recently targeted kinases and their inhibitors-the path to clinical trials. *Curr.*

Opin. Pharmacol. **17C**, 58–63 (2014).

4. Fabbro, D., Cowan-Jacob, S. W., Möbitz, H. & Martiny-Baron, G. Targeting cancer with small-molecular-weight kinase inhibitors. *Methods Mol. Biol.* **795**, 1–34 (2012).

5. Force, T. & Kolaja, K. L. Cardiotoxicity of kinase inhibitors: the prediction and translation of preclinical models to clinical outcomes. *Nat. Rev. Drug Discov.* **10**, 111–26 (2011).

6. Chu, T. F. *et al.* Cardiotoxicity associated with tyrosine kinase inhibitor sunitinib. *Lancet* **370**, 2011–2019 (2007).

7. Orphanos, G. S., Ioannidis, G. N. & Ardavanis, A. G. Cardiotoxicity induced by tyrosine kinase inhibitors. *Acta Oncol* **48**, 964–970 (2009).

8. Force, T. & Kerkelä, R. Cardiotoxicity of the new cancer therapeutics--mechanisms of, and approaches to, the problem. *Drug Discov. Today* **13**, 778–84 (2008).

9. Davis, M. I. *et al.* Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **29**, 1046–51 (2011).

10. Elkins, J. M. *et al.* Comprehensive characterization of the Published Kinase Inhibitor Set. *Nat. Biotechnol.* **34**, 95–103 (2016).

11. Hasinoff, B. B. & Patel, D. The lack of target specificity of small molecule anticancer kinase inhibitors is correlated with their ability to damage myocytes in vitro. *Toxicol. Appl. Pharmacol.* **249**, 132–139 (2010).

12. Force, T., Krause, D. S. & Etten, R. A. Van Molecular mechanisms of cardiotoxicity of tyrosine kinase inhibition. *Nat. Rev. Cancer* **7**, 332–344 (2007).

13. Will, Y. *et al.* Effect of the multitargeted tyrosine kinase inhibitors imatinib, dasatinib, sunitinib, and sorafenib on mitochondrial function in isolated rat heart mitochondria and H9c2 cells. *Toxicol. Sci.* **106**, 153–161 (2008).

14. Kerkelä, R. *et al.* Cardiotoxicity of the cancer therapeutic agent imatinib mesylate. *Nat. Med.* **12**, 908–916 (2006).

15. Doherty, K. R. *et al.* Multi-parameter in vitro toxicity testing of crizotinib, sunitinib, erlotinib, and nilotinib in human cardiomyocytes. *Toxicol. Appl. Pharmacol.* **272**, 245–55 (2013).

16. Bai, J. P. F. & Abernethy, D. R. Systems pharmacology to predict drug toxicity: integration across levels of biological organization. *Annu. Rev. Pharmacol. Toxicol.* **53**, 451–73 (2013).

17. Berger, S. I. & Iyengar, R. Role of systems pharmacology in understanding drug adverse events. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **3**, 129–135 (2011).

18. Zhao, S. *et al.* Systems pharmacology of adverse event mitigation by drug combinations. *Sci. Transl. Med.* **5**, 206ra140 (2013).

19. Berger, S. I., Ma'ayan, A. & Iyengar, R. Systems pharmacology of arrhythmias. *Sci. Signal.* **3**, ra30 (2010).

20. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 301–320 (2005).

21. Simarro, M. *et al.* Fast kinase domain-containing protein 3 is a mitochondrial protein essential for cellular respiration. *Biochem. Biophys. Res. Commun.* **401**, 440–446 (2010).

22. Varga, Z. V, Ferdinandy, P., Liaudet, L. & Pacher, P. Drug-induced mitochondrial dysfunction and cardiotoxicity. *Am. J. Physiol. Heart Circ. Physiol.* **309**, H1453–67 (2015).

23. Harvey, P. A. & Leinwand, L. A. Oestrogen enhances cardiotoxicity induced by Sunitinib by regulation of drug transport and metabolism. *Cardiovasc. Res.* **107**, 66–77 (2015).

24. Balcarek, K. *et al.* Role of Pyrimidine Depletion in the Mitochondrial Cardiotoxicity of Nucleoside Analogue Reverse Transcriptase Inhibitors. *J Acquir Immune Defic Syndr* **55**, 550–557 (2010).

25. Berthiaume, J. M. & Wallace, K. B. Adriamycin-induced oxidative mitochondrial cardiotoxicity. In *Cell Biol. Toxicol.* **23**, 15–25 (2007).

26. Wang, S., Song, P. & Zou, M. H. Inhibition of AMP-activated protein kinase ?? (AMPK??) by doxorubicin accentuates genotoxic stress and cell death in mouse embryonic fibroblasts and cardiomyocytes: Role of p53 and SIRT1. *J. Biol. Chem.* **287**, 8001–8012 (2012).

27. Cheng, Z. *et al.* Focal adhesion kinase antagonizes doxorubicin cardiotoxicity via p21(Cip1.). *J. Mol. Cell. Cardiol.* **67**, 1–11 (2014).

28. Ronnebaum, S. M. & Patterson, C. The FoxO family in cardiac function and dysfunction. *Annu. Rev. Physiol.* **72**, 81–94 (2010).

29. Hansen, J. Disease neighborhoods PREPRINT. *bioRxiv* (2016).doi:bioRxiv DOI

30. Kawabata, M. *et al.* Downregulation of stanniocalcin 1 is responsible for sorafenib-induced cardiotoxicity. *Toxicol. Sci.* **143**, 374–84 (2015).

31. Seemann, I. *et al.* Radiation- and anthracycline-induced cardiac toxicity and the influence of ErbB2 blocking agents. *Breast Cancer Res. Treat.* **141**, 385–95 (2013).

32. Kopljar, I. *et al.* Functional and Transcriptional Characterization of Histone Deacetylase Inhibitor-Mediated Cardiac Adverse Effects in Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes. *Stem Cells Transl. Med.* **5**, 602–12 (2016).

33. Hasselt, J. G. C. van, Boekhout, A. H., Beijnen, J. H., Schellens, J. H. M. & Huitema, A. D. R. Population pharmacokinetic-pharmacodynamic analysis of trastuzumab-associated cardiotoxicity. *Clin. Pharmacol. Ther.* **90**, 126–32 (2011).

34. Hasselt, J. G. C. van, Schellens, J. H. M., Gillavry, M. R. Mac, Beijnen, J. H. & Huitema, A. D. R. Model-Based Evaluation and Optimization of Cardiac Monitoring Protocols for Adjuvant Treatment of Breast Cancer with Trastuzumab. *Pharm. Res.* **29**, 3499–3511 (2012).

35. Hasselt, J. G. C. van *et al.* Population pharmacokinetic-pharmacodynamic analysis for eribulin mesilate-associated neutropenia. *Br J Clin Pharmacol* **76**, 412–424 (2013).

36. Soumillon, M., Cacchiarelli, D., Semrau, S., Oudenaarden, A. van & Mikkelsen, T. S. Characterization of directed differentiation by high-throughput single-cell RNA-Seq. *bioRxiv* (2014).

37. Kivioja, T. *et al.* Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 (2011).

38. Xiong, Y. DGE vs Conventional RNAseq Comparison PREPRINT. *bioRxiv* (2016).doi:DOI bioRxiv

39. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

40. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–40 (2010).

41. Brown, E. G., Wood, L. & Wood, S. The Medical Dictionary for Regulatory Activities (MedDRA). *Drug Saf.* **20**, 109–117 (1999).

TABLES

Table 1. Overview of KIs included in this analysis.

Drug	Approval year[1]	Therapeutic targets	Therapeutic concentration (uM)	Cell lines
Afatinib	2013	ErbB2, EGFR	0.05	4
Axitinib	2012	VEGFR1/VEGFR2/VEGFR3/PDGFR	0.2	4
Bosutinib	2012	Bcr-Abl, SRC	0.1	2
Cabozantinib	2012	c-Met, VEGFR2	2	3
Ceritinib	2014	ALK	1	3
Crizotinib	2011	ALK, HGFR	0.25	4
Dabrafenib	2013	BRAF	2.5	4
Dasatinib	2006	ABL, ARG, KIT, PDGFRα/β, SRC	0.1	3
Erlotinib	2004	ErbB1	3	3
Gefitinib	2003	ErbB1	1	4
Imatinib	2001	Bcr-Abl	5	4
Nilotinib	2007	Bcr-Abl	3	4
Pazopanib	2009	VEGFR2, PDGFRα/β, KIT	10	3
Ponatinib	2012	Bcr-Abl, BEGFR, PDGFR, FGFR, EPH, SRC, c-KIT, RET, TIE2, FLT3	0.1	3
Regorafenib	2012	RET, VEGFR, PDGFR	1	2
Ruxolitinib	2011	JAK	1	3
Sorafenib	2005	B-RAF, VEGFRs, PDGFRα/β, FLT3,	0.5	2
Sunitinib	2006	VEGFR, PDGFR, CSF1R, FLT3, KIT	1	3
Trametinib	2013	MEK1, MEK2	0.1	3
Tofacitinib	2012	JAK	1	3
Vandetanib	2011	RET, VEGFR, EGFR	0.33	2
Vemurafenib	2011	BRAF	2	3

[1]US Approval date, first indication; See Table S3 for literature references to therapeutic concentrations.

FIGURES

Figure 1A. Transcriptomic profiling across KIs. For each drug, genes were ranked by mean fold-change gene expression value across replicates (>3) and cell lines, and the top 500 genes were kept. (A) Heatmap showing the Jaccard index which indicates the magnitude of similarity in top-ranking differentially expressed genes for all KI pairs.
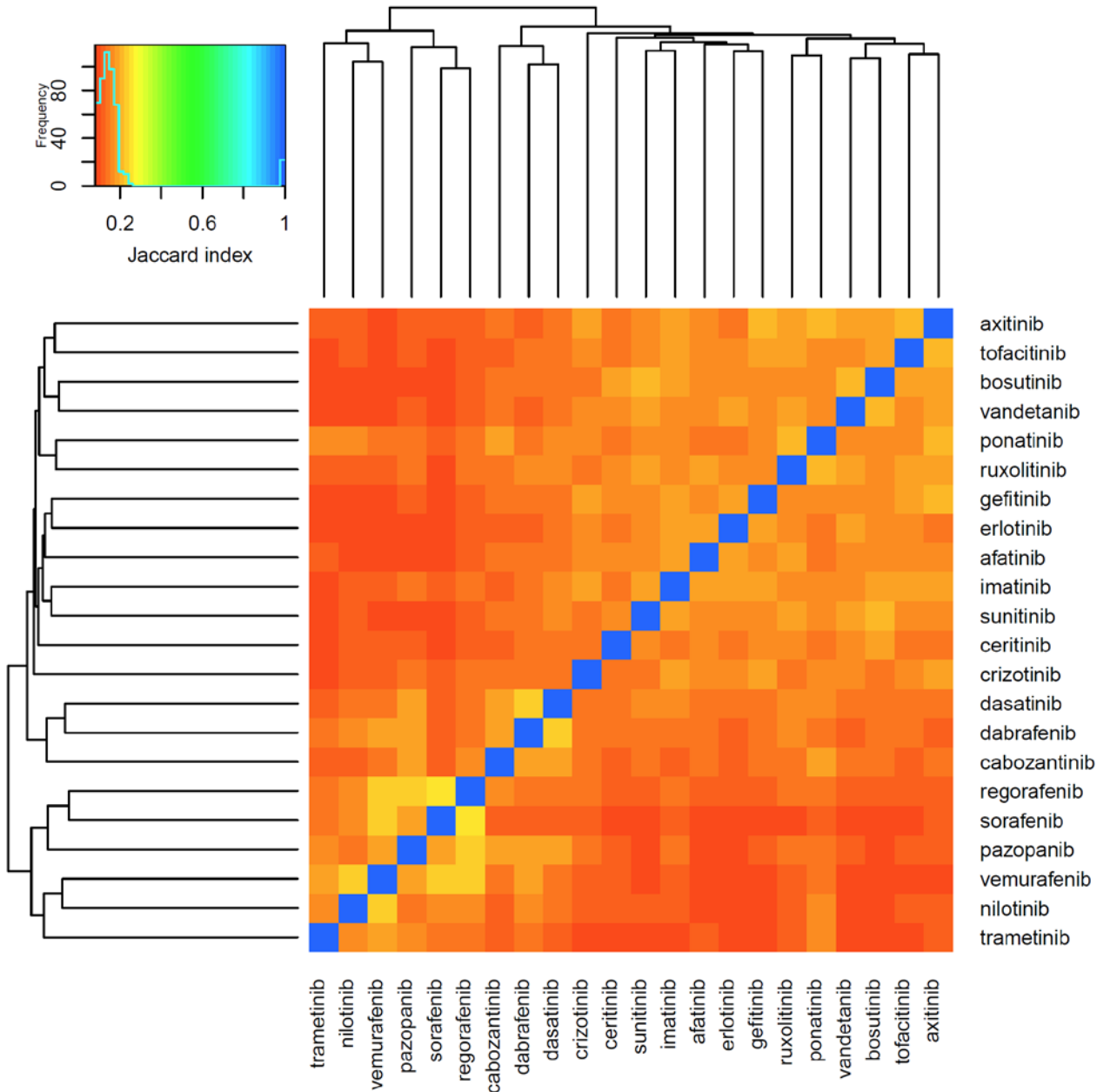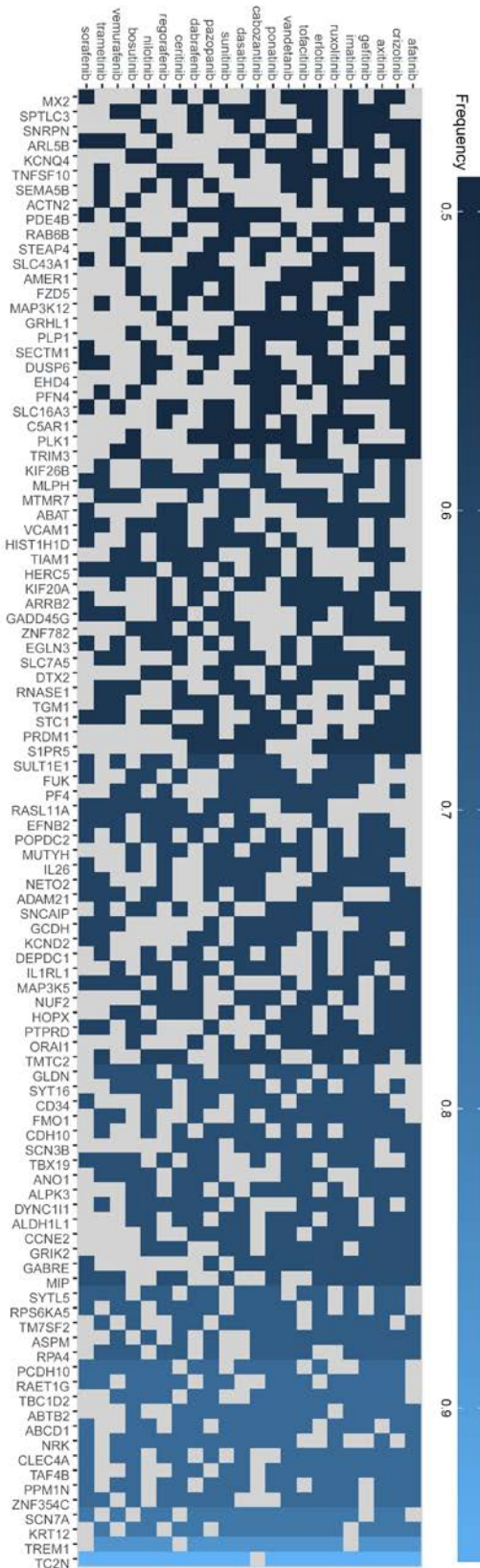
Figure 1B Transcriptomic profiling across KIs. For each drug, genes were ranked by mean fold-change gene expression value across replicates (>3) and cell lines, and the top 500 genes were kept. (B) Heatmap showing frequency of genes present for different KIs, with the genes and drugs ranked by frequency of occurrence

Figure 1C. Transcriptomic profiling across KIs. For each drug, genes were ranked by mean fold-change gene expression value across replicates (>3) and cell lines, and the top 500 genes were kept. (C) First three principal components based on full mean fold-change gene expression profiles across KIs.
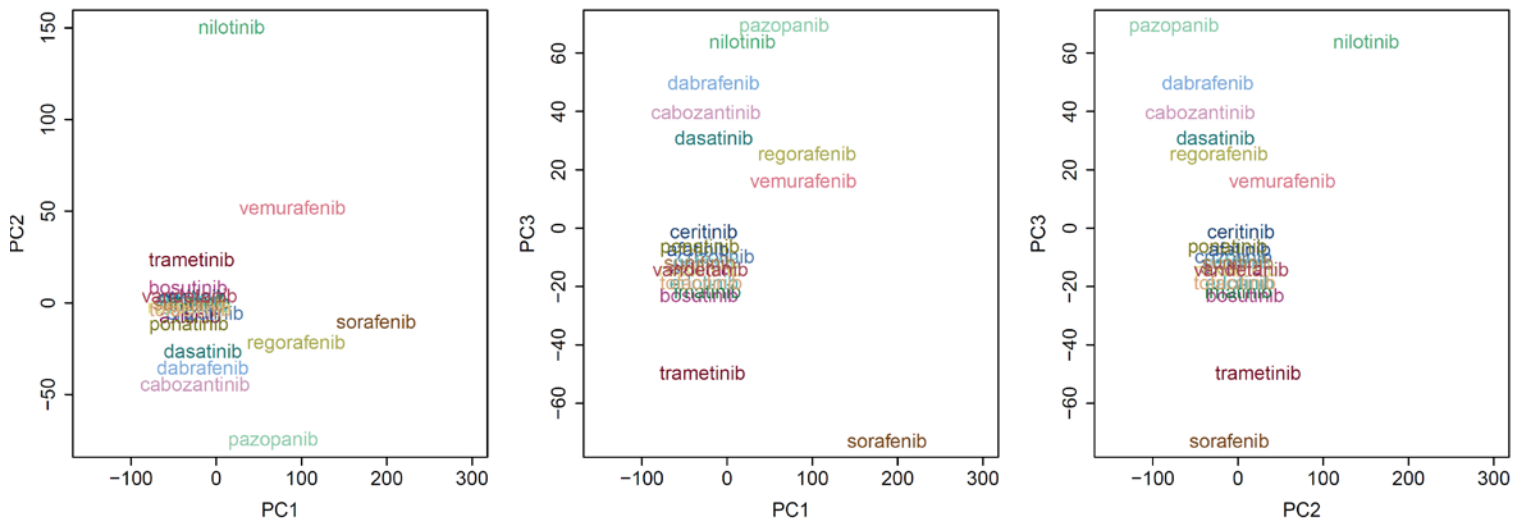
Figure 2. (A) Overview of data mining procedure to obtain FAERS based CT risk scores. (B). FDA Adverse Event Reporting System (FAERS) derived reporting odds ratios (top) and associated Z scores (bottom) for four different subtypes of cardiotoxicity: hypertrophic cardiomyopathy (HCM), dilated cardiomyopathy (DCM), ventricular dysfunction and heart failure (VDHF), and cardiotoxicity undefined (CTU), for different kinase inhibitors (KIs). A reporting odds ratio >1, or, a Z-score > 0 indicates above average risk for CT.
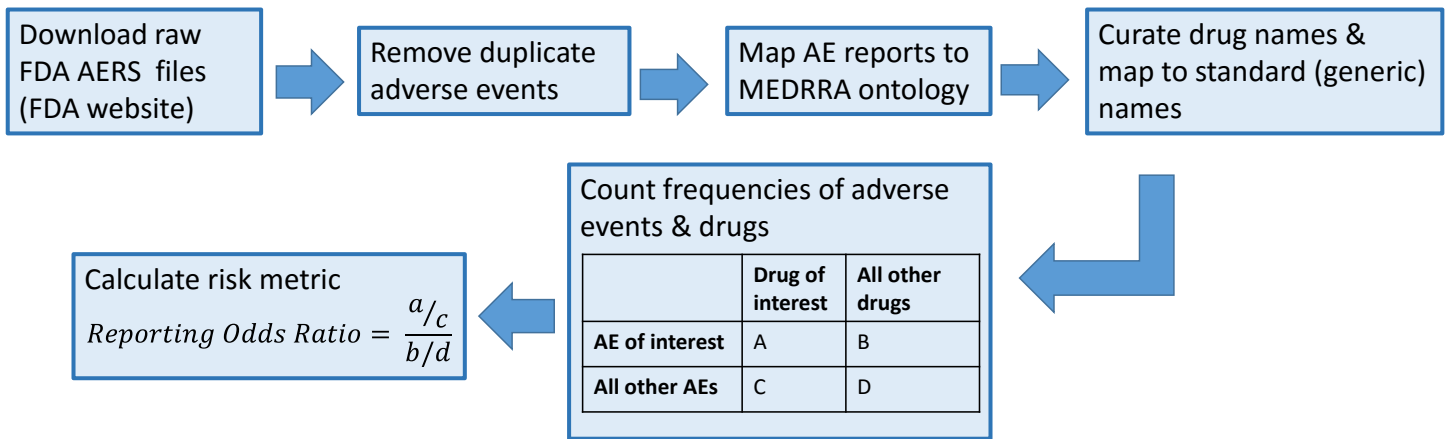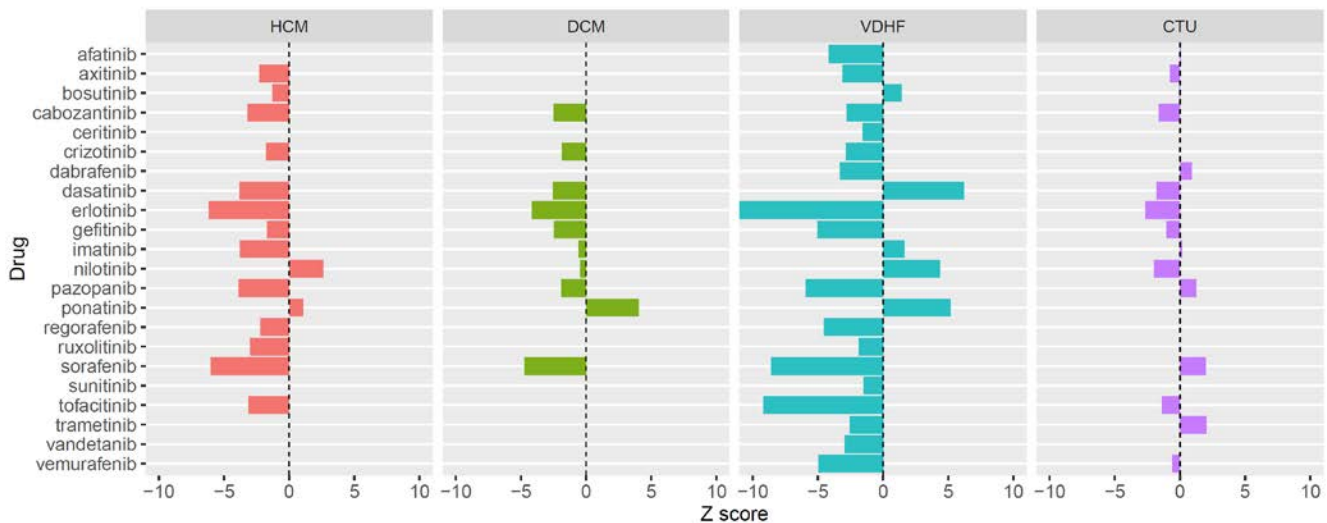

A.
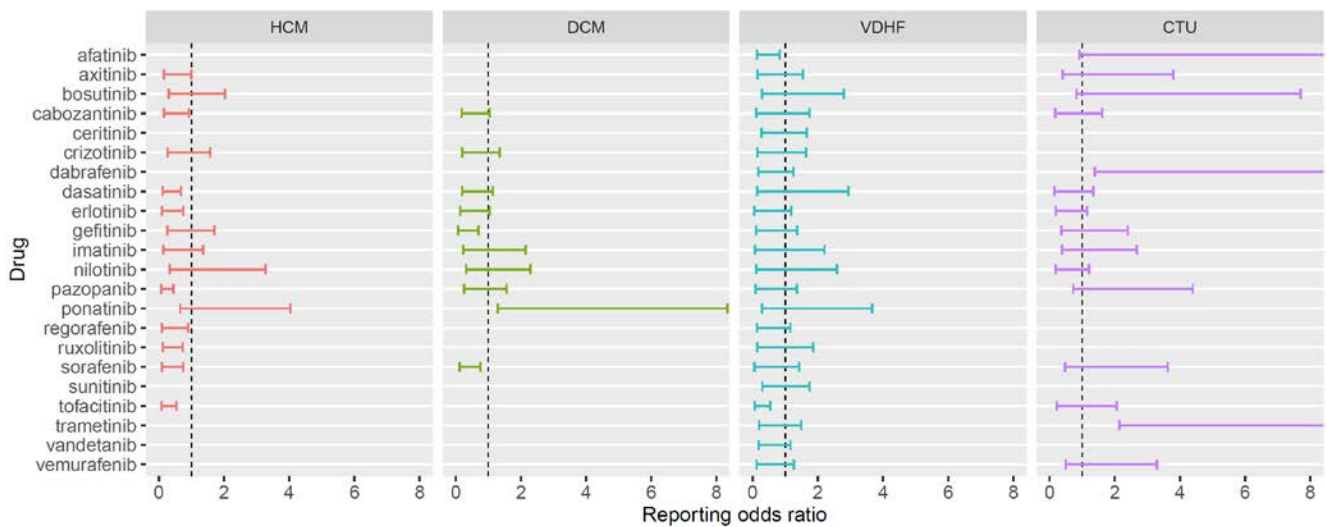
B.

Figure 3. Schematic overview of the methods used for data generation and analysis.
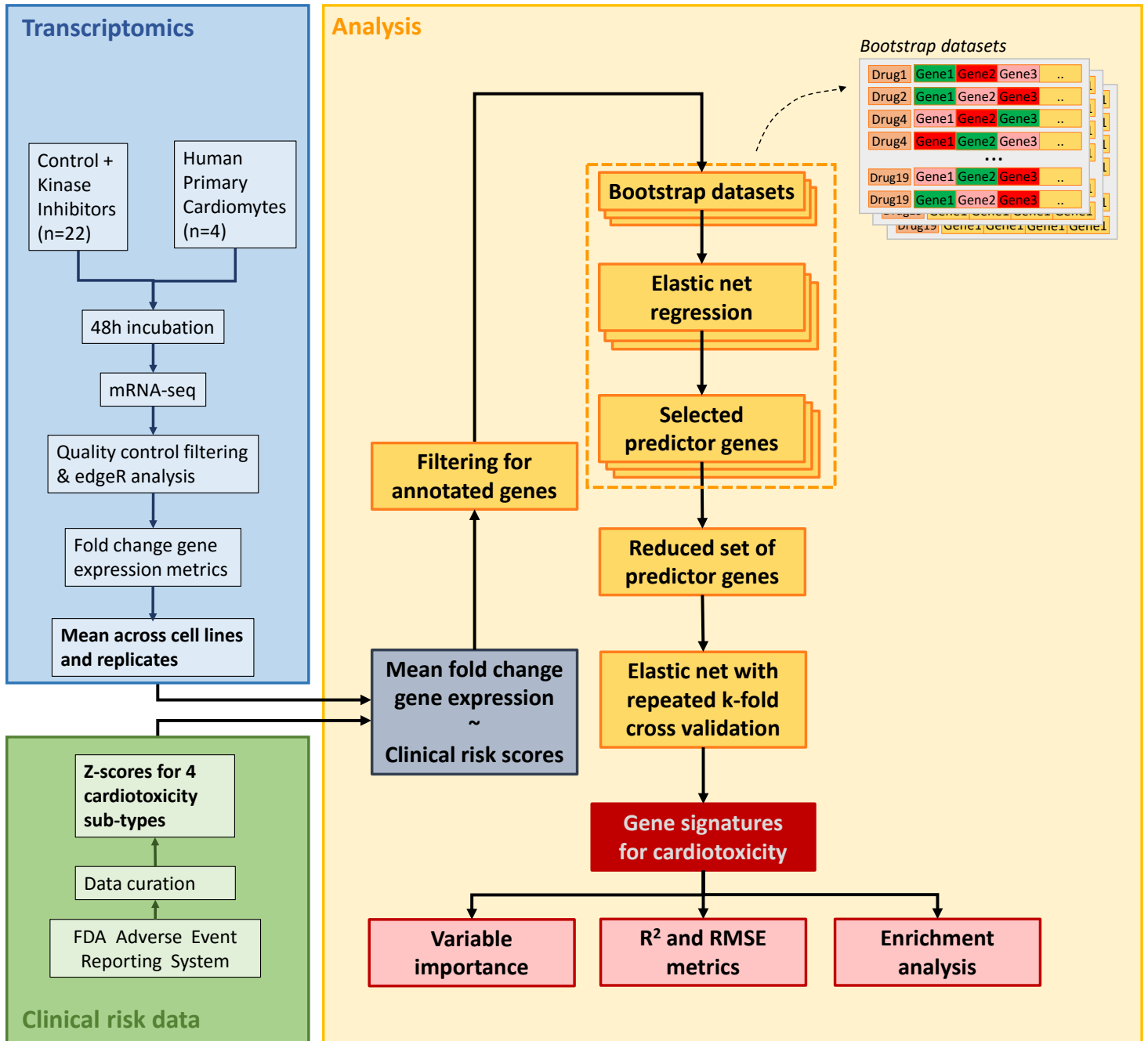
Figure 4A. Regression analysis and signatures derived from clinical risk scores and transcriptomics profiling. (A) 3-fold cross validation based mean and standard deviation of predicted and observed clinical Z-score metrics, and their associated mean R^2 and RMSE metrics for four subtypes of cardiotoxicity.
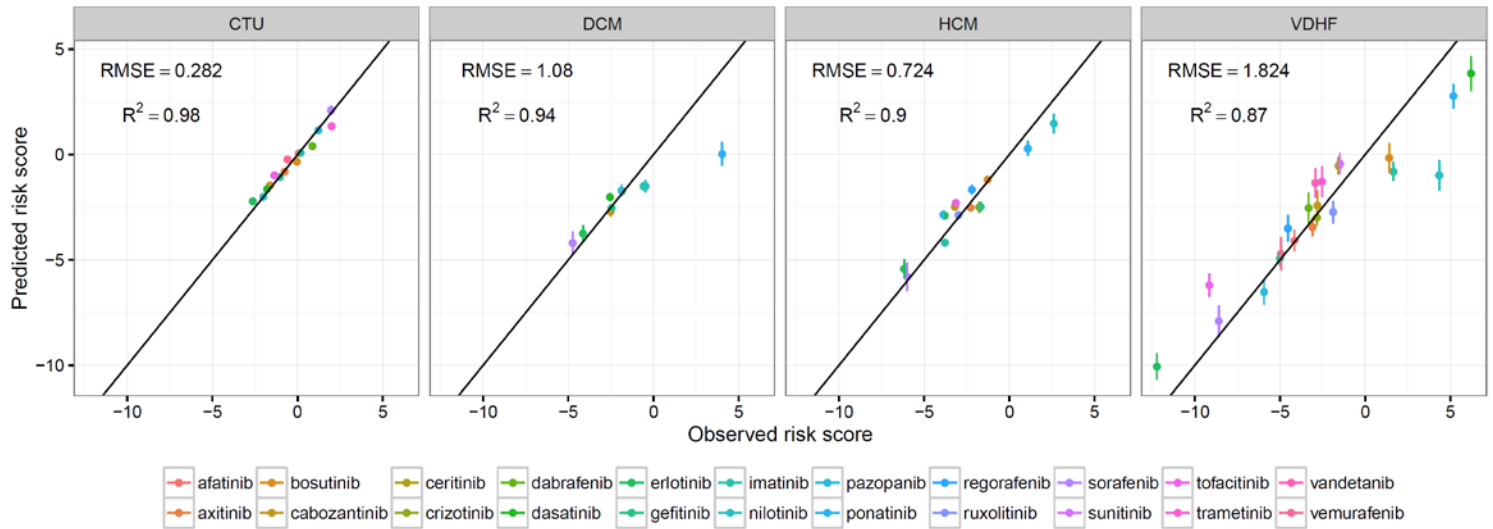
Figure 4B. Regression analysis and signatures derived from clinical risk scores and transcriptomics profiling. (A) 3-fold cross validation based mean and standard deviation of predicted and observed clinical Z-score metrics, and their associated mean R2 and RMSE metrics, for four subtypes of cardiotoxicity. (B) Variable importance ranking of final genes signatures, with number of total genes in signature.
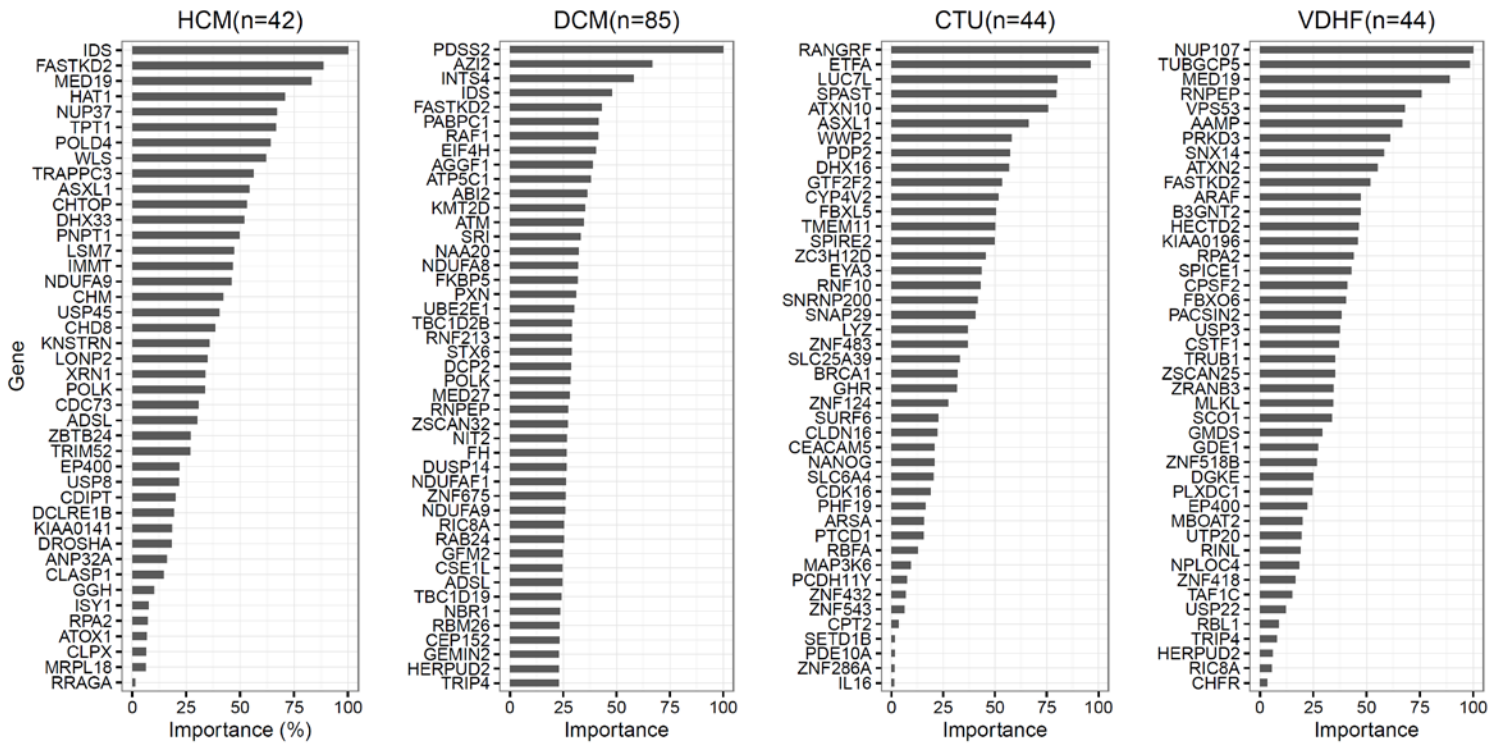
Figure 4C. Clustered heatmaps of selected predictors and associated mean fold change gene expression values for different KIs. In plots B and C, for DCM, the lowest ranking genes were not included. These were: ZKSCAN7, ANAPC4, COPS4, SLC25A3, DAB2, PQLC2, ING5, MRPL18, NUP37, CLPX, MRPL51, RBFA, ZDHHC20, ERBB2IP, MBOAT2, MALSU1, PLK2, GMDS, NADSYN1, WLS, ATG12, ENY2, ZNF180, NR2C2, MED19, FAM175B, DLD, ZNF702P, SUCLG1, TIA1, ZNF28, PML, CBLB, ZMYM4, ZSWIM7, KMT2E, NDUFA4, DST, ATG7, PCDHB7.
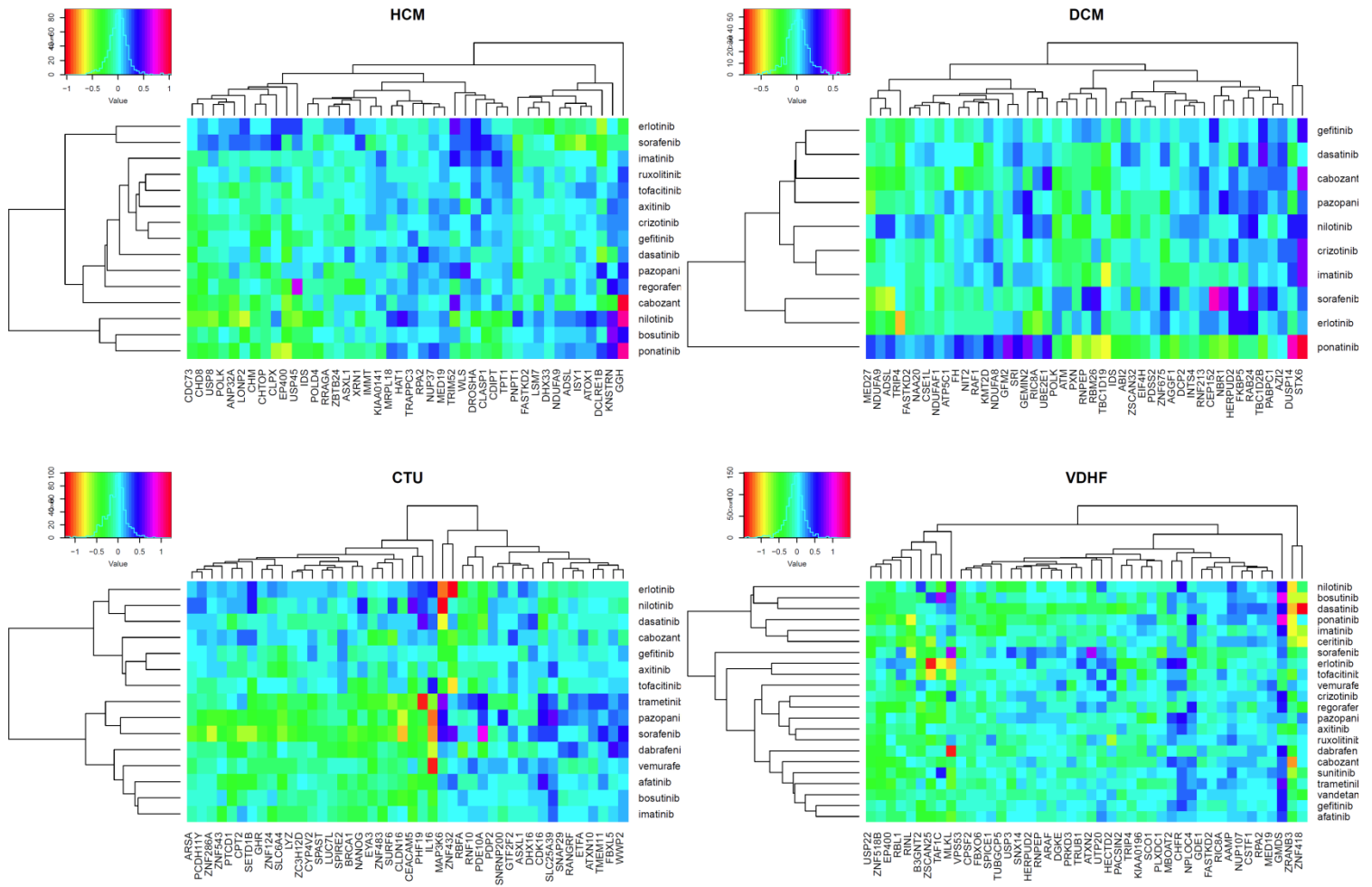
Figure 4D. Venn diagrams of gene selected in the final gene signature. Overlapping genes have been indicated.
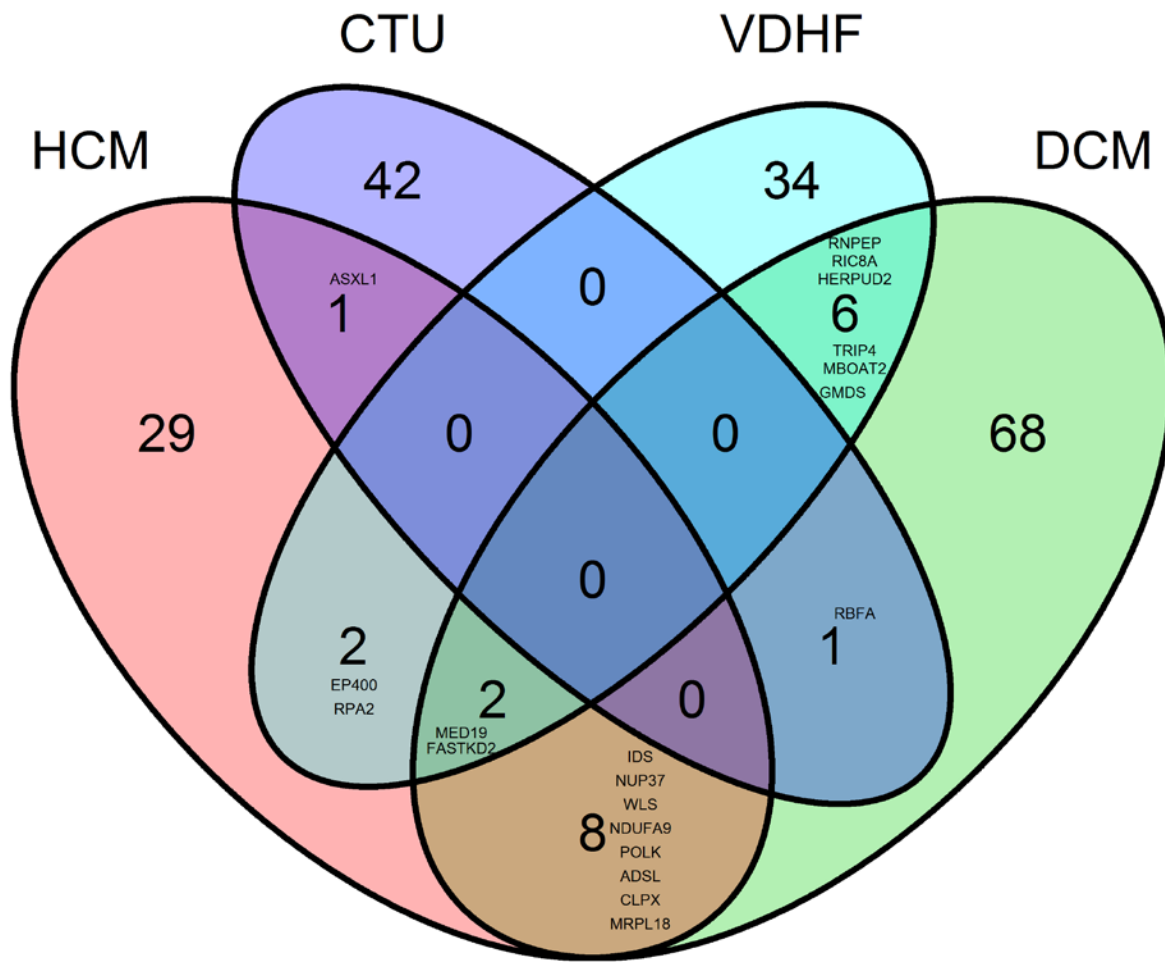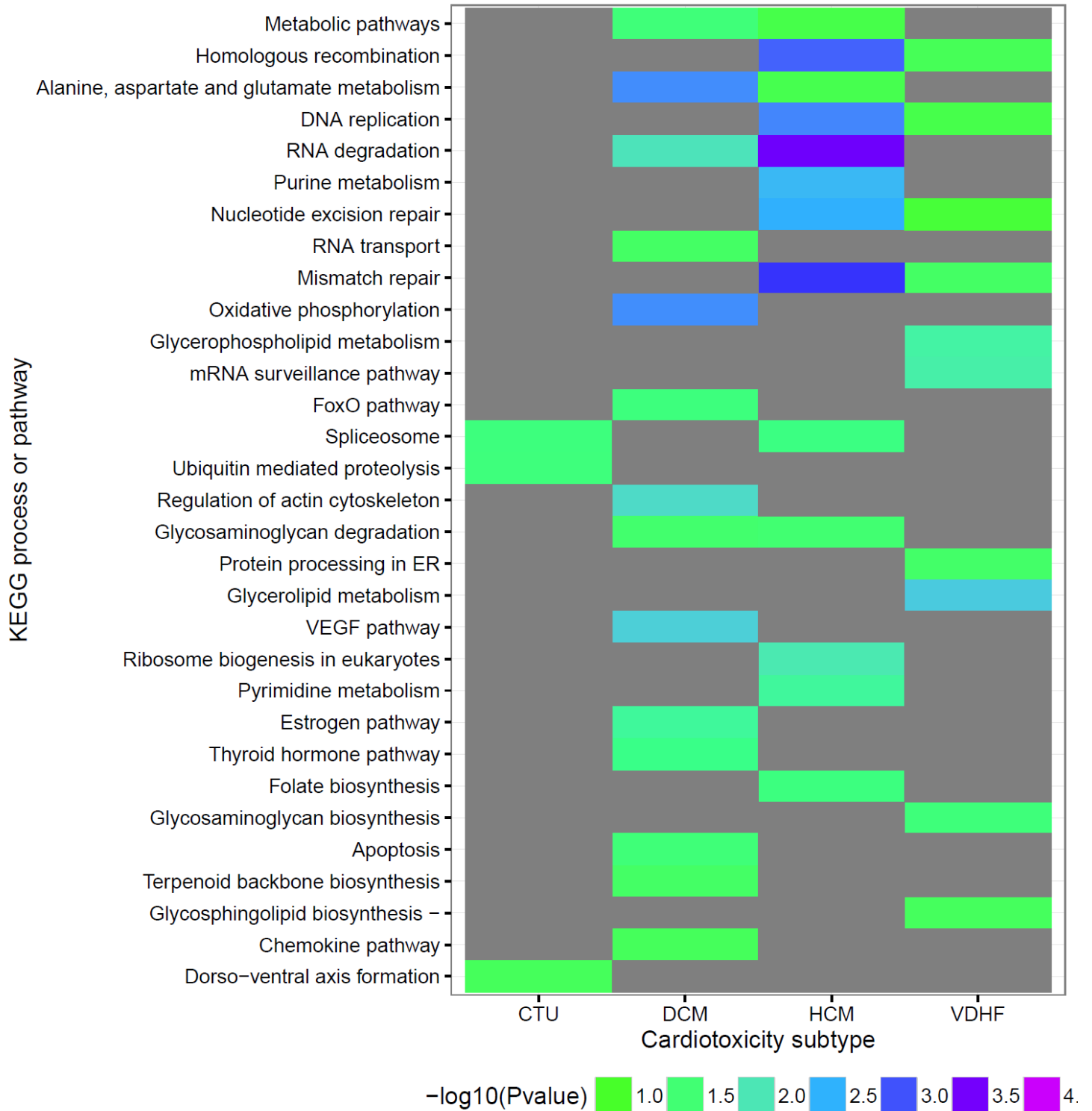
Figure 5. Enrichment analysis based on the KEGG database of clinically weighted gene expression signatures for cardiotoxicity subtypes ranked by their combined significance.

SUPPLEMENTAL MATERIALS. TABLES.

Table S1. Definition of MEDDRA subgroups for different cardiotoxicity subtypes used.

Meddra term	Classification
acute right ventricular failure	VDHF
right ventricular failure	VDHF
cardiac asthma	VDHF
ventricular failure	VDHF
cardiac failure congestive	VDHF
cardiac failure	VDHF
cardiac failure acute	VDHF
cardiac failure chronic	VDHF
cardiac ventricular disorder	VDHF
oedema due to cardiac disease	VDHF
cardiovascular insufficiency	VDHF
ventricular dyskinesia	VDHF
ventricular hypokinesia	VDHF
systolic dysfunction	VDHF
myocardial oedema	VDHF
diastolic dysfunction	VDHF
myocardial fibrosis	VDHF
ventricular dysfunction	VDHF
left ventricular dysfunction	VDHF
right ventricular dysfunction	VDHF
ischaemic cardiomyopathy	VDHF
cytotoxic cardiomyopathy	VDHF
cardiomyopathy	VDHF
congestive cardiomyopathy	VDHF
cardiomyopathy acute	VDHF
restrictive cardiomyopathy	VDHF
metabolic cardiomyopathy	VDHF
non-obstructive cardiomyopathy	VDHF
left ventricular heave	HCM
left atrial hypertrophy	HCM
right atrial hypertrophy	HCM
cardiac septal hypertrophy	HCM
ventricular hypertrophy	HCM
left ventricular hypertrophy	HCM
cardiac hypertrophy	HCM
right ventricular hypertrophy	HCM
atrial hypertrophy	HCM
hypertrophic cardiomyopathy	HCM
left atrial dilatation	DCM
dilatation ventricular	DCM
right atrial dilatation	DCM
cardiotoxicity	CTU
heart injury	CTU

Table S2. Final gene signatures for four subtypes of CT.

DCM	HCM	CTU	VDHF
PDSS2	IDS	RANGRF	NUP107
AZI2	FASTKD2	ETFA	TUBGCP5
INTS4	MED19	LUC7L	MED19
IDS	HAT1	SPAST	RNPEP
FASTKD2	NUP37	ATXN10	VPS53
PABPC1	TPT1	ASXL1	AAMP
RAF1	POLD4	WWP2	PRKD3
EIF4H	WLS	PDP2	SNX14
AGGF1	TRAPPC3	DHX16	ATXN2
ATP5C1	ASXL1	GTF2F2	FASTKD2
ABI2	CHTOP	CYP4V2	ARAF
KMT2D	DHX33	FBXL5	B3GNT2
ATM	PNPT1	TMEM11	HECTD2
SRI	LSM7	SPIRE2	KIAA0196
NAA20	IMMT	ZC3H12D	RPA2
NDUFA8	NDUFA9	EYA3	SPICE1
FKBP5	CHM	RNF10	CPSF2
PXN	USP45	SNRNP200	FBXO6
UBE2E1	CHD8	SNAP29	PACSIN2
TBC1D2B	KNSTRN	LYZ	USP3
RNF213	LONP2	ZNF483	CSTF1
STX6	XRN1	SLC25A39	TRUB1
DCP2	POLK	BRCA1	ZSCAN25
POLK	CDC73	GHR	ZRANB3
MED27	ADSL	ZNF124	MLKL
RNPEP	ZBTB24	SURF6	SCO1
ZSCAN32	TRIM52	CLDN16	GMDS
NIT2	EP400	CEACAM5	GDE1
FH	USP8	NANOG	ZNF518B
DUSP14	CDIPT	SLC6A4	DGKE
NDUFAF1	DCLRE1B	CDK16	PLXDC1
ZNF675	KIAA0141	PHF19	EP400
NDUFA9	DROSHA	ARSA	MBOAT2
RIC8A	ANP32A	PTCD1	UTP20
RAB24	CLASP1	RBFA	RINL
GFM2	GGH	MAP3K6	NPLOC4
CSE1L	ISY1	PCDH11Y	ZNF418
ADSL	RPA2	ZNF432	TAF1C
TBC1D19	ATOX1	ZNF543	USP22
NBR1	CLPX	CPT2	RBL1
RBM26	MRPL18	SETD1B	TRIP4
CEP152	RRAGA	PDE10A	HERPUD2
GEMIN2		ZNF286A	RIC8A
HERPUD2		IL16	CHFR
TRIP4			
ZKSCAN7			
ANAPC4			
COPS4			
SLC25A3			
DAB2			
PQLC2			
ING5			
MRPL18			
NUP37			
CLPX			
MRPL51			
RBFA			
ZDHHC20			
ERBB2IP			
MBOAT2			
MALSU1			
PLK2			
GMDS			
NADSYN1			
WLS			
ATG12			
ENY2			
ZNF180			
NR2C2			
MED19			
FAM175B			
DLD			
ZNF702P			
SUCLG1			
TIA1			
ZNF28			
PML			
CBLB			
ZMYM4			
ZSWIM7			
KMT2E			
NDUFA4			
DST			
ATG7			
PCDHB7			

Table S3. Typical concentrations used for experiments (exposure for 48h). Concentrations were selected based on the mean maximum drug concentrations reported in clinical studies. For dose escalation studies the highest concentration was used.

Drug	In Vitro (uM)	Reference
Sorafenib	0.5	[1]
Sunitinib	1	[2][3]
Crizotinib	0.25	[4]
Erlotinib	3	[5][6]
Imatinib	5	[7]
Dasatinib	0.1	[8][9]
Tofacitinib	1	[10]
Gefitinib	1	[11]
Bosutinib	0.1	[12]
Vandetanib	0.333	[13]
Lapatinib	2	[14]
Nilotinib	3	[15][16]
Axitinib	0.2	[17][18]
Pazopanib	10	[19]
Ruxolitinib	1	[20][21]
Afatinib	0.05	[22][23]
Regorafenib	1	[24]
Ponatinib	0.1	[25]
Dabrafenib	2.5	[26][27]
Vemurafenib	2	[28]
Cabozantinib	2	[29]
Trametinib	0.1	[30]
Ceritinib	1	[31]

1. Abou-Alfa GK, Schwartz L, Ricci S, et al (2006) Phase II study of sorafenib in patients with advanced hepatocellular carcinoma. J Clin Oncol 24:4293–300. doi: 10.1200/JCO.2005.01.3441

2. Minkin P, Zhao M, Chen Z, et al (2008) Quantification of sunitinib in human plasma by high-performance liquid chromatography-tandem mass spectrometry. J Chromatogr B Anal Technol Biomed Life Sci 874:84–88. doi: 10.1016/j.jchromb.2008.09.007

3. Faivre S, Delbaldo C, Vera K, et al (2006) Safety, pharmacokinetic, and antitumor activity of SU11248, a novel oral multitarget tyrosine kinase inhibitor, in patients with cancer. J Clin Oncol 24:25–35. doi: 10.1200/JCO.2005.02.2194

4. Kwak E, Camidge D, Clark J, et al (2009) Clinical activity observed in a phase I dose escalation trial of an oral c-MET and ALK inhibitor. J Clin Oncol 27:

5. Herbst RS, Prager D, Hermann R, et al (2005) TRIBUTE: A phase III trial of erlotinib hydrochloride (OSI-774) combined with carboplatin and paclitaxel chemotherapy in advanced non-small-cell lung cancer. J Clin Oncol 23:5892–5899. doi: 10.1200/JCO.2005.02.840

6. Soulieres D, Senzer NN, Vokes EE, et al (2004) Multicenter phase II study of erlotinib, an oral epidermal growth factor receptor tyrosine kinase inhibitor, in patients with recurrent or metastatic squamous cell cancer of the head and neck. J Clin Oncol 22:77–85. doi: 10.1200/JCO.2004.06.075

7. Demetri GD, Wang Y, Wehrle E, et al (2009) Imatinib plasma levels are correlated with clinical benefit in patients with unresectable/metastatic gastrointestinal stromal tumors. J Clin Oncol 27:3141–3147. doi: 10.1200/JCO.2008.20.4818

8. Dai CL, Tiwari AK, Wu CP, et al (2008) Lapatinib (Tykerb, GW572016) reverses multidrug resistance in cancer cells by inhibiting the activity of ATP-binding cassette subfamily B member 1 and G member 2. Cancer Res 68:7905–7914. doi: 10.1158/0008-5472.CAN-08-0499

9. McCaig AM, Cosimo E, Leach MT, Michie AM (2012) Dasatinib inhibits CXCR4 signaling in chronic lymphocytic leukaemia cells and impairs migration towards CXCL12. PLoS One 7:e48929. doi: 10.1371/journal.pone.0048929

10. Dowty ME, Lin J, Ryder TF, et al (2014) The pharmacokinetics, metabolism, and clearance mechanisms of tofacitinib, a janus kinase inhibitor, in humans. Drug Metab Dispos 42:759–73. doi: 10.1124/dmd.113.054940

11. Hirano S, Sano K, Takeda Y, et al (2012) The pharmacokinetics and long-term therapeutic effects of gefitinib in patients with lung adenocarcinoma harboring the epidermal growth factor receptor(EGFR)mutation. Gan To Kagaku Ryoho 39:1501–6.

12. Abbas R, Hug BA, Leister C, et al (2012) A phase I ascending single-dose study of the safety, tolerability, and pharmacokinetics of bosutinib (SKI-606) in

healthy adult subjects. Cancer Chemother Pharmacol 69:221–7. doi: 10.1007/s00280-011-1688-7

13. Martin P, Oliver S, Robertson J, et al (2011) Pharmacokinetic drug interactions with vandetanib during coadministration with rifampicin or itraconazole. Drugs R D 11:37–51. doi: 10.2165/11586980-000000000-00000

14. LoRusso PM, Jones SF, Koch KM, et al (2008) Phase I and pharmacokinetic study of lapatinib and docetaxel in patients with advanced cancer. J Clin Oncol 26:3051–6. doi: 10.1200/JCO.2007.14.9633

15. Trent J, Molimard M (2011) Pharmacokinetics and pharmacodynamics of nilotinib in gastrointestinal stromal tumors. Semin Oncol 38 Suppl 1:S28–33. doi: 10.1053/j.seminoncol.2011.01.014

16. Tanaka C, Yin OQP, Sethuraman V, et al (2010) Clinical pharmacokinetics of the BCR-ABL tyrosine kinase inhibitor nilotinib. Clin Pharmacol Ther 87:197–203. doi: 10.1038/clpt.2009.208

17. Pithavala YK, Chen Y, Toh M, et al (2012) Evaluation of the effect of food on the pharmacokinetics of axitinib in healthy volunteers. Cancer Chemother Pharmacol 70:103–12. doi: 10.1007/s00280-012-1888-9

18. Rugo HS, Herbst RS, Liu G, et al (2005) Phase I trial of the oral antiangiogenesis agent AG-013736 in patients with advanced solid tumors: pharmacokinetic and clinical results. J Clin Oncol 23:5474–83. doi: 10.1200/JCO.2005.04.192

19. Hurwitz HI, Dowlati A, Saini S, et al (2009) Phase I Trial of Pazopanib in Patients with Advanced Cancer rial of Pazopanib in Patients with Advanced Cancer. 4220–4227. doi: 10.1158/1078-0432.CCR-08-2740

20. Shi JG, Chen X, McGee RF, et al (2011) The pharmacokinetics, pharmacodynamics, and safety of orally dosed INCB018424 phosphate in healthy volunteers. J Clin Pharmacol 51:1644–54. doi: 10.1177/0091270010389469

21. Chen X, Shi JG, Emm T, et al (2014) Pharmacokinetics and pharmacodynamics of orally administered ruxolitinib (INCB018424 phosphate) in renal and hepatic impairment patients. Clin Pharmacol drug Dev 3:34–42. doi: 10.1002/cpdd.77

22. Schnell D, Buschke S, Fuchs H, et al (2014) Pharmacokinetics of afatinib in subjects with mild or moderate hepatic impairment. Cancer Chemother Pharmacol 74:267–75. doi: 10.1007/s00280-014-2484-y

23. Stopfer P, Marzin K, Narjes H, et al (2012) Afatinib pharmacokinetics and metabolism after oral administration to healthy male volunteers. Cancer Chemother Pharmacol 69:1051–61. doi: 10.1007/s00280-011-1803-9

24. Rey J-B, Launay-Vacher V, Tournigand C (2015) Regorafenib as a single-agent in the treatment of patients with gastrointestinal tumors: an overview for pharmacists. Target Oncol 10:199–213. doi: 10.1007/s11523-014-0333-x

25. Narasimhan NI, Dorer DJ, Niland K, et al (2013) Effects of ketoconazole on the pharmacokinetics of ponatinib in healthy subjects. J Clin Pharmacol 53:974–81. doi: 10.1002/jcph.109

26. Ouellet D, Gibiansky E, Leonowens C, et al (2014) Population pharmacokinetics of dabrafenib, a BRAF inhibitor: effect of dose, time, covariates, and relationship with its metabolites. J Clin Pharmacol 54:696–706. doi: 10.1002/jcph.263

27. Falchook GS, Long G V, Kurzrock R, et al (2012) Dabrafenib in patients with melanoma, untreated brain metastases, and other solid tumours: a phase 1 dose-escalation trial. Lancet (London, England) 379:1893–901. doi: 10.1016/S0140-6736(12)60398-5

28. Grippo JF, Zhang W, Heinzmann D, et al (2014) A phase I, randomized, open-label study of the multiple-dose pharmacokinetics of vemurafenib in patients with BRAF V600E mutation-positive metastatic melanoma. Cancer Chemother Pharmacol 73:103–11. doi: 10.1007/s00280-013-2324-5

29. Kurzrock R, Sherman SI, Ball DW, et al (2011) Activity of XL184 (Cabozantinib), an oral tyrosine kinase inhibitor, in patients with medullary thyroid cancer. J Clin Oncol 29:2660–6. doi: 10.1200/JCO.2010.32.4145

30. Kim KB, Kefford R, Pavlick AC, et al (2013) Phase II study of the MEK1/MEK2 inhibitor Trametinib in patients with metastatic BRAF-mutant cutaneous melanoma previously treated with or without a BRAF inhibitor. J Clin Oncol 31:482–9. doi: 10.1200/JCO.2012.43.5966

31. Shaw AT, Kim D-W, Mehra R, et al (2014) Ceritinib in ALK-rearranged non-small-cell lung cancer. N Engl J Med 370:1189–97. doi: 10.1056/NEJMoa1311107

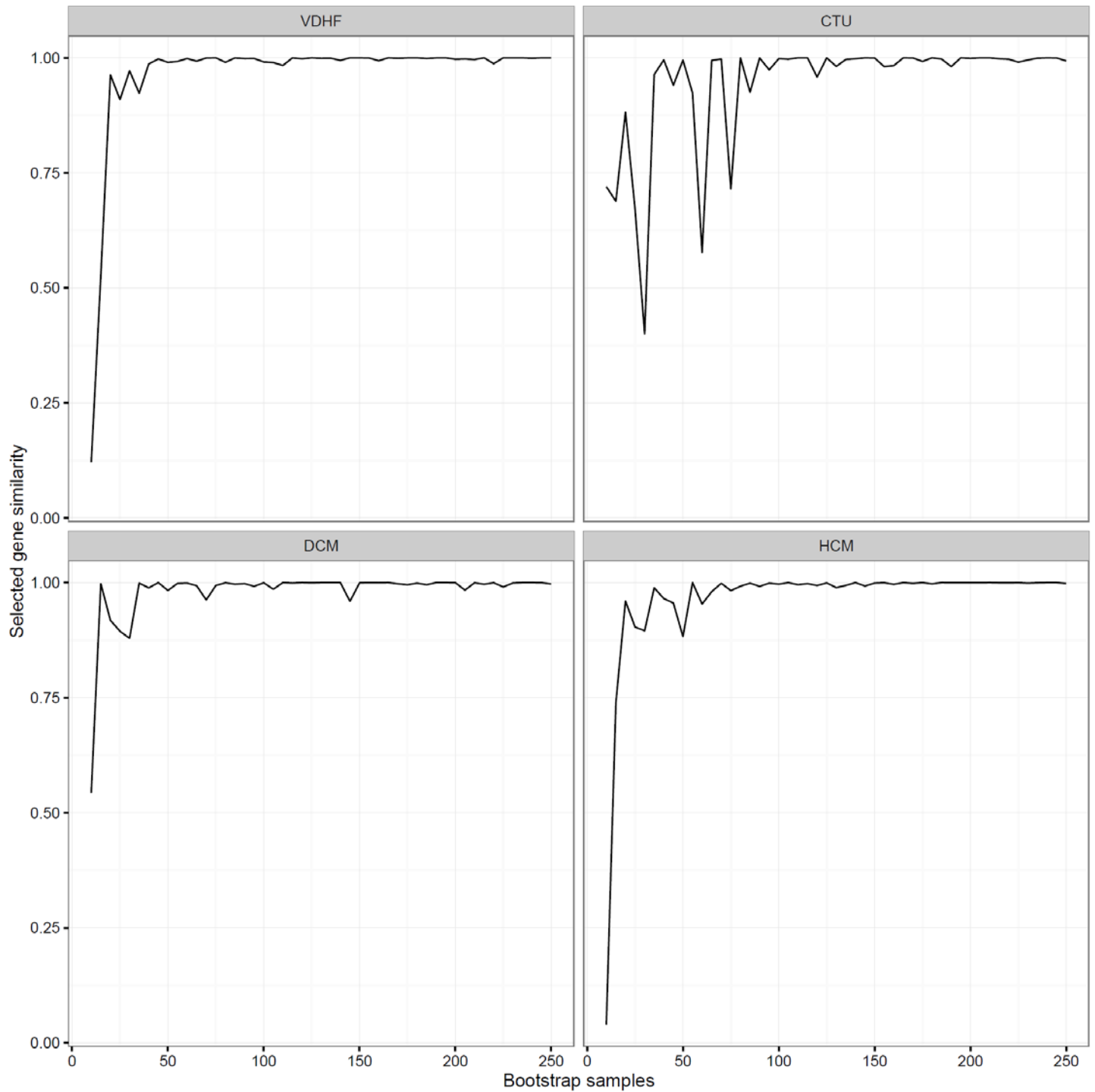## SUPPLEMENTAL MATERIAL 3. FIGURES.

Figure S1. Selected gene similarity versus number of bootstrap samples for 4 sub-types of cardiotoxicity.
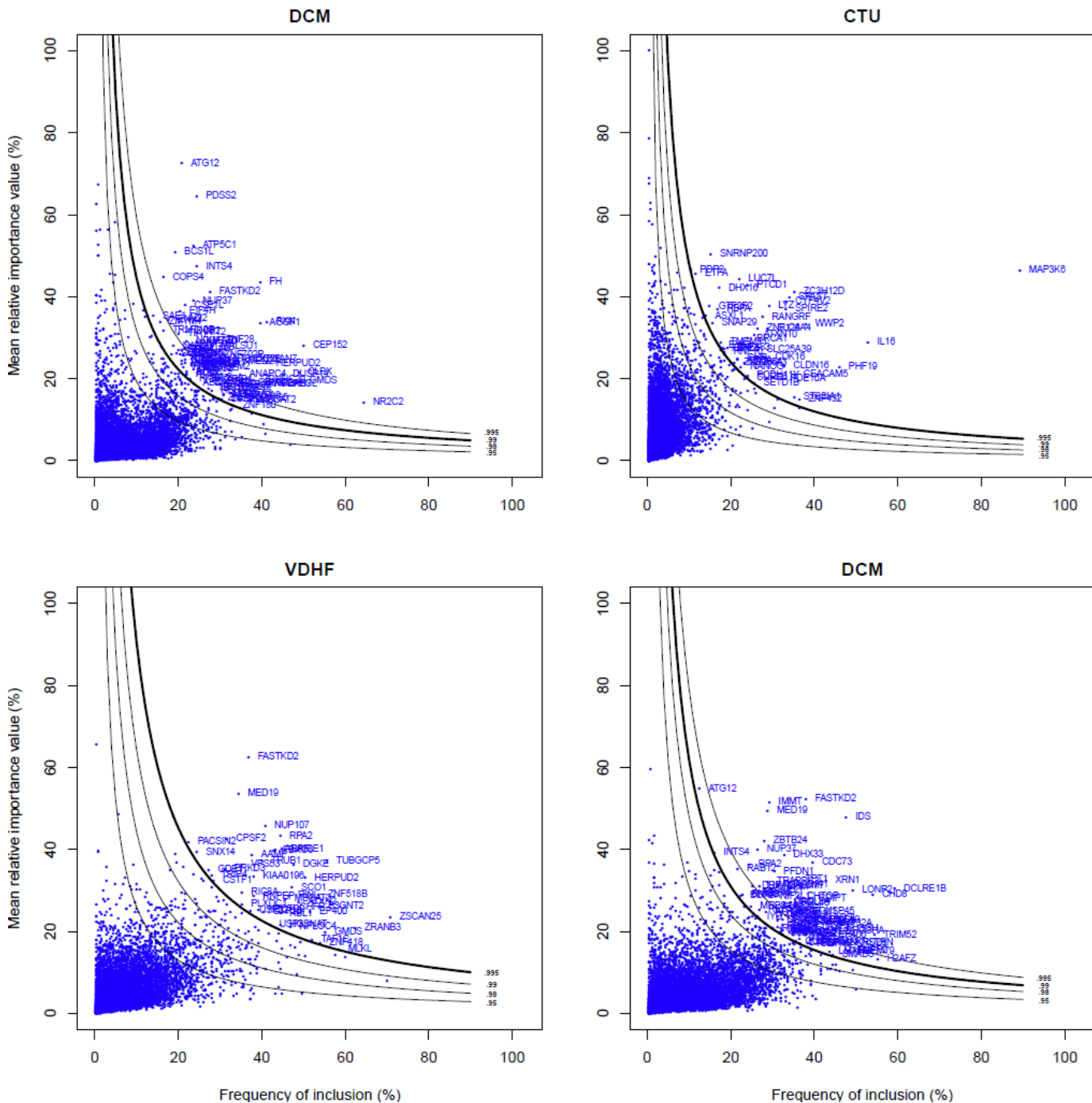
Figure S2. Selection of predictors from the bootstrap analysis for different CT subtypes. The blue circles indicate individual genes that were selected with their associated frequency of inclusion value and mean relative importance value, computed from the bootstrap analysis (n=250). The solid lines indicate different percentiles for the product of the frequency and the mean relative importance. The solid line with increased with indicates the optimal percentile used to fit the final gene signature regression models. The names of selected genes above the selected percentile are depicted.
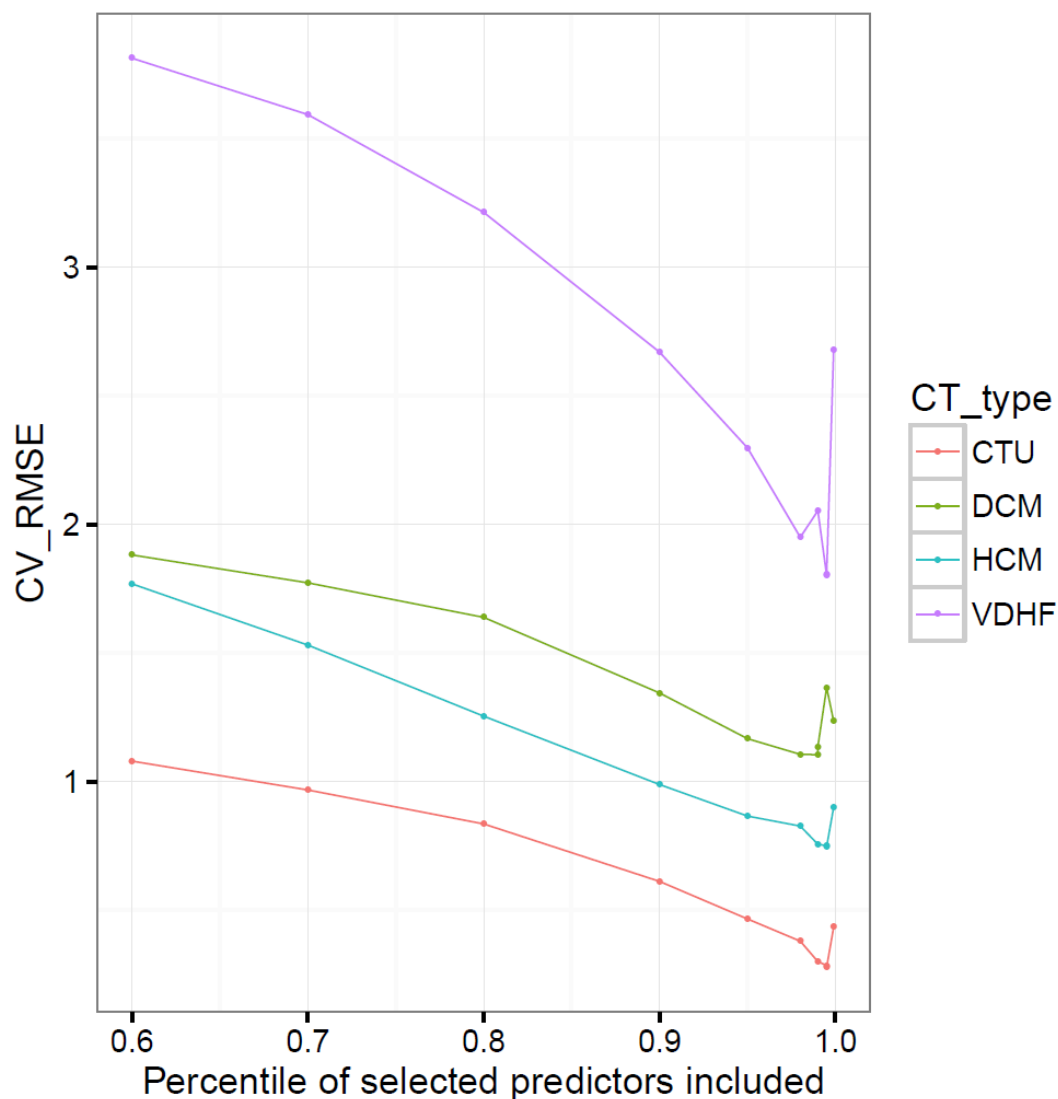
Figure S3. Selection of optimal threshold percentile for inclusion of genes in the final gene signature regression analysis for different subtypes of cardiotoxicity, based on the cross-validation RMSE obtained. The percentile was selected for the minimum RMSE.

SUPPLEMENTAL 4. Elastic net model implemented in R.