



Computational approaches identify a transcriptomic fingerprint of drug-induced structural cardiotoxicity

Victoria P.W. Au Yeung · Olga Obrezanova · Jiarui Zhou · Hongbin Yang · Tara J. Bowen · Delyan Ivanov · Izzy Saffadi · Alfie S. Carter · Vigneshwari Subramanian · Inken Dillmann · Andrew Hall · Adam Corrigan · Mark R. Viant · Amy Pointon

Received: 19 October 2023 / Accepted: 15 May 2024 / Published online: 28 June 2024
© Crown 2024

Abstract Structural cardiotoxicity (SCT) presents a high-impact risk that is poorly tolerated in drug discovery unless significant benefit is anticipated. Therefore, we aimed to improve the mechanistic understanding of SCT. First, we combined machine learning methods with a modified calcium transient assay in human-induced pluripotent stem cell-derived cardiomyocytes to identify nine parameters that could predict SCT. Next, we applied transcriptomic profiling to human cardiac microtissues exposed to

structural and non-structural cardiotoxins. Fifty-two genes expressed across the three main cell types in the heart (cardiomyocytes, endothelial cells, and fibroblasts) were prioritised in differential expression and network clustering analyses and could be linked to known mechanisms of SCT. This transcriptomic fingerprint may prove useful for generating strategies to mitigate SCT risk in early drug discovery.

Keywords Structural cardiotoxicity · Calcium transients · Transcriptomics · Bioinformatics · Machine learning

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10565-024-09880-7>.

V. P. Au Yeung (✉) · I. Saffadi · A. S. Carter · A. Hall · A. Pointon
Safety Sciences, Clinical Pharmacology & Safety Sciences, R&D, AstraZeneca, Cambridge, UK
e-mail: victoria.ayeung.publications@gmail.com

I. Saffadi
e-mail: izzysaffadi@gmail.com

A. S. Carter
e-mail: alfie.carter@hertford.ox.ac.uk

A. Hall
e-mail: andrew.hall@astrazeneca.com

A. Pointon
e-mail: amy.pointon@astrazeneca.com

V. P. Au Yeung · A. Corrigan
Phenomics, Data Sciences & Quantitative Biology, R&D AstraZeneca, Cambridge, UK
e-mail: adam.corrigan@astrazeneca.com

O. Obrezanova
Imaging and Data Analytics, Clinical Pharmacology & Safety Sciences, R&D, AstraZeneca, Cambridge, UK
e-mail: olga.obrezanova@astrazeneca.com

J. Zhou · T. J. Bowen · M. R. Viant
School of Biosciences, University of Birmingham, Edgbaston, Birmingham, UK
e-mail: J.Zhou.3@bham.ac.uk

T. J. Bowen
e-mail: tara.bowen@md.catapult.org.uk

M. R. Viant
e-mail: m.viant@bham.ac.uk

H. Yang
Centre for Molecular Informatics, Department of Chemistry, University of Cambridge, Cambridge, UK
e-mail: hy353@cantab.ac.uk

Introduction

Drug-induced structural cardiotoxicity (SCT) is clinically-defined as changes in left ventricular ejection fraction which may lead to fibrosis, cardiomyopathy, heart failure, and death. SCT presents a high-impact risk that limits clinical utility of novel therapies. Consequently, it is important to detect and mitigate SCT during the early stages of drug discovery to enable the development of therapies without this liability. SCT is defined as morphological damage of cardiac tissue and changes to intracellular organelles which clinically result in a decrease in left ventricular ejection fraction and heart failure (Lavery et al. 2011; Pointon et al. 2013).

Structural cardiotoxins including anti-cancer compounds such as anthracyclines (Swain et al. 2003; von Hoff et al. 1979), chemotherapeutic agents (Sara et al. 2018), and tyrosine kinase inhibitors (Crisci et al. 2019; Yamaoka et al. 2018) (Crisci et al. 2019; Sara et al. 2018; Swain et al. 2003; von Hoff et al. 1979; Yamaoka et al. 2018), have been associated with multiple phenotypic mechanisms, including oxidative stress (Geisberg and Sawyer 2010; Ott et al. 2007; S. Zhang et al. 2012), inflammation, apoptosis (Geisberg and Sawyer 2010; Ott et al. 2007; Youle and Van Der Blik 2012), contractile dysfunction (Billingham et al. 1978; Geisberg and Sawyer 2010), extracellular matrix organisation and intercellular interactions (Hedin et al. 1997; Lorusso et al. 2020; Warn-Cramer and Lau 2004), and dysregulation of energy metabolism (Zhang et al. 2012) (Billingham et al. 1978; Geisberg and Sawyer 2010; Hedin et al. 1997;

Lorusso et al. 2020; Ott et al. 2007; Warn-Cramer and Lau 2004; Youle and Van Der Blik 2012; Zhang et al. 2012). These mechanisms have been linked to key organelles, for example mitochondria (Archer et al. 2018; Ott et al. 2007; Varga et al. 2015; Youle and Van Der Blik 2012) and endoplasmic reticulum (ER) (Archer et al. 2018), but the molecular mechanisms and cellular drivers underpinning these changes remain unknown (Archer et al. 2018; Ott et al. 2007; Varga et al. 2015; Youle and Van Der Blik 2012).

The development of humanised *in vitro* cardiac models provides an opportunity to study SCT by shedding light on the molecular mechanisms underlying phenotypic perturbations. Studies of human-induced pluripotent stem cell-derived cardiomyocytes (hiPS-CMs) and co-culture models (i.e. human cardiac microtissues) have shown that these cell models are sensitive to compound-induced changes in cardiomyocyte morphology and function (Karakikes et al. 2015; Pointon et al. 2013), and can be used to develop phenotypic assays for the detection of SCT (Archer et al. 2018) and molecular insights underlying these changes (Brandão et al. 2022; Chaudhari et al. 2016; Deidda et al. 2019; Glaab et al. 2021; Matsa et al. 2016; McSweeney et al. 2019; Palmer et al. 2020; Schmidt et al. 2023; van Hasselt et al. 2020; Yuan et al. 2020).

Phenotypic assays include assessment of key organelles including ER integrity and mitochondrial membrane potential (MMP) (Archer et al. 2018), previously reported as a predictive approach for SCT and physiological processes. For example, the calcium transient assay,² when combined with the software CardioWave (Yang et al. 2022), can derive up to 40 parameters. These parameters have been used to build machine learning models to study acute cardiotoxicity (Yang et al. 2022).

Molecular insights can be gained via ‘omics’ profiling to identify mechanistic markers of toxicity. To date, transcriptomic (Chaudhari et al. 2016; Glaab et al. 2021; Matsa et al. 2016; McSweeney et al. 2019; van Hasselt et al. 2020), metabolomic (Deidda et al. 2019; Palmer et al. 2020; Yuan et al. 2020), and proteomic (Brandão et al. 2022; Schmidt et al. 2023; Yuan et al. 2020) technologies, applied separately and *in tandem*, have identified several hallmark pathways of general cardiotoxicity. As highlighted in observational and clinical studies, pathways identified in these studies include, but

D. Ivanov
High-Throughput Screening, R&D, AstraZeneca,
Alderley Park, UK
e-mail: delyan.ivanov1@astrazeneca.com

V. Subramanian
Imaging and Data Analytics, Clinical Pharmacology &
Safety Sciences, R&D, AstraZeneca, Gothenburg, Sweden
e-mail: vigneshwari.subramanian@astrazeneca.com

I. Dillmann
Disease Molecular Profiling, Discovery Biology, R&D
AstraZeneca, Gothenburg, Sweden
e-mail: inken.dillmann@astrazeneca.com

M. R. Viant
Phenome Centre Birmingham, University of Birmingham,
Edgbaston, Birmingham, UK

are not limited to, cell adhesion, energy metabolism, oxidative stress, calcium homeostasis and contractility, protein homeostasis, apoptosis, and mitochondrial homeostasis (Brandão et al. 2022; Chaudhari et al. 2016; Deidda et al. 2019; Glaab et al. 2021; Matsa et al. 2016; McSweeney et al. 2019; Palmer et al. 2020; Schmidt et al. 2023; van Hasselt et al. 2020; Yuan et al. 2020). Despite the broad range of pathways and molecular players identified, these studies primarily focus on assessing a limited range of anthracyclines or tyrosine kinase inhibitors (TKIs) (Brandão et al. 2022; Chaudhari et al. 2016; Deidda et al. 2019; Glaab et al. 2021; Matsa et al. 2016; McSweeney et al. 2019; Palmer et al. 2020; Schmidt et al. 2023; van Hasselt et al. 2020; Yuan et al. 2020).

Furthermore, studies performed *in vitro* have focused on single cell types, predominantly cardiomyocytes, while ignoring other cellular constituents in the heart that represents 70% of the myocardium cell mass. Functional evidence exists to support the role of non-cardiomyocytes in SCT. For example, cardiac fibroblasts may regulate cardiomyocyte function by secreting paracrine signaling factors such as TGF-beta and interleukin-6, which in turn can lead to cardiac hypertrophy and electrophysiological changes (Cartledge et al., 2015). In SCT, these processes may be disrupted, leading to cardiac fibrosis. Another example is the crosstalk between endothelial cells and cardiomyocytes. Endothelial cells may secrete endothelin-1, which binds to cardiomyocyte endothelin-1 receptors to promote cardiomyocyte survival (Schorlemmer et al., 2008). With SCT, endothelial cell dysfunction may thus lead to cardiomyocyte apoptosis and cardiac degeneration. Other mechanisms are reviewed more comprehensively elsewhere (Guo et al., 2021), nevertheless, the study of additional cardiac cell types in a human-based *in vitro* model system may reveal physiologically-relevant and novel pathways of SCT.

Here, we use phenotypic assay data in cardiomyocytes and transcriptomic data in cardiac micro-tissues to i) identify phenotypic parameters associated with SCT; ii) characterise the gene expression changes in response to several known structural cardiotoxins, and iii) identify a molecular fingerprint of SCT (Fig. 1).

Results

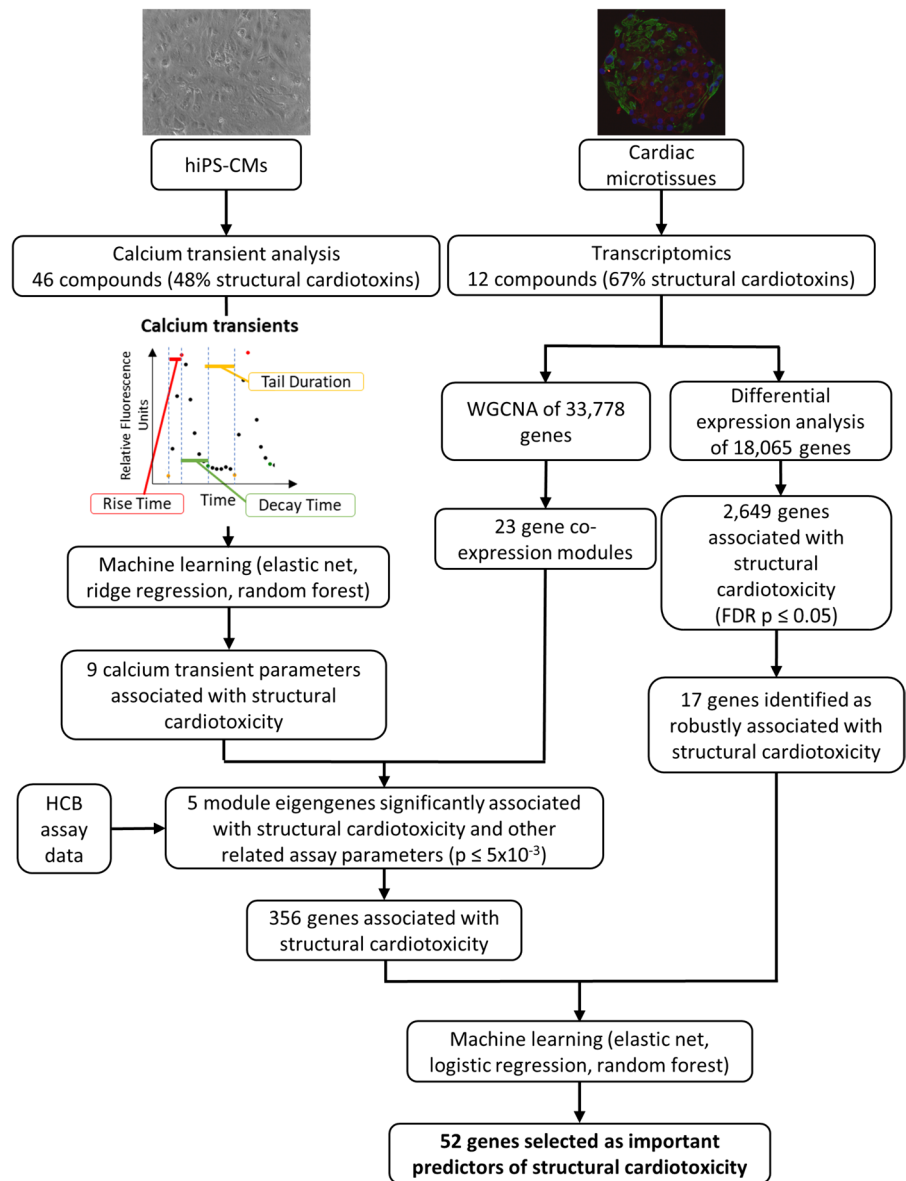
Identification and characterisation of calcium transient parameters associated with SCT

Calcium transient parameters were calculated from the calcium transients of hiPS-CMs exposed to 46 compounds (Supplementary Table 1; Supplementary Experimental Procedures). Of 40 parameters derived from CardioWave, 16 described the standard errors of other parameters and were not included due to limited interpretability compared to their corresponding mean parameter measures. The remaining 24 were unique parameters which described measures of peak count, amplitude, and other features of the calcium transient waveform including peak width, shoulder, tail, and valley (Supplementary Table 1).

Of the 24 parameters considered, we included 17 which were uncorrelated (Pearson $R^2 < 0.80$) as features to assess seven machine learning classifiers of SCT. Compared to a dummy classifier, all models performed better at predicting SCT by area under curve (AUC) (Supplementary Figure 1). The performance across machine learning models on predicting SCT was moderate across models (from AUC = 0.64 in Gaussian Naïve Bayes to AUC = 0.85 in logistic regression), in line with the expectation that calcium transients alone do not fully capture mechanisms of SCT.

To identify important features associated with SCT, we examined the top ten important features highlighted in two selected linear models (ridge regression and elastic net) and one non-linear model (random forest). Nine features (average inner lambda, average shoulder amplitude, average tail duration, average valley duration, maximum intensity, minimum intensity, mean peak width (10%), mean peak width (25%), peak count) were highlighted as important features of SCT in at least two of these models (Fig. 2). These features accounted for some compound-specific variation in the dataset (Fig. 2B). A logistic regression classifier of the 9 features performed better at distinguishing structural cardiotoxins from non-structural cardiotoxins than a baseline classifier including only peak count and average amplitude as features (Table 1). These results indicate the importance of leveraging multiple phenotypic assays and parameters to capture non-overlapping mechanisms of SCT.

Fig. 1 Study design. (Top left) Calcium transient assays were performed on hiPS-CMs derive parameters across 46 compounds (48% structural cardiotoxins). Machine learning was then used to select 9 calcium transient parameters as important putative predictors of SCT. (Top right) Transcriptomic analysis was performed on cardiac microtissues exposed to an overlapping set of 12 compounds (67% structural cardiotoxins). Differential expression and network clustering approaches were used to prioritise genes as features for machine learning classifiers of SCT, and feature selection was performed to identify a list of 52 gene predictors for SCT. HCB = High Content Biology



Characterisation of the effects of structural and non-structural cardiotoxins on the gene expression profiles of cardiac microtissues

Transcriptomic profiling was performed in cardiac microtissues using 12 compounds (67% structural cardiotoxins) at three timepoints (6 hr, 48 hr, 72 hr) and two concentrations (Supplementary Table 2; Experimental Procedures). High concentrations were determined for each compound based on the lowest concentration of either i) IC_{30} of ATP depletion in cardiac microtissues in a high content

biology assay associated with SCT; ii) $\leq 15\times$ total C_{max} (to maintain therapeutic relevance); or iii) lowest investigated concentration (Archer et al. 2018). Low concentrations were calculated as the half-log dilution of the high concentrations.

A total of 51,195 genes were measured across samples, of which 18,065 with sufficient gene counts and variance were assessed for differential gene expression analysis (Experimental Procedures). After adjusting for batch effects, we found that gene expression profiles were driven by compound-specific effects as well as smaller effects

Fig. 2 Identification of waveform parameters associated with SCT. **A)** Filled-in boxes and bolded parameter labels highlight parameters which were selected in two or more machine learning models of SCT. **B)** Principal component analysis of structural cardiotoxicity by the nine selected parameters. Some of the structural cardiotoxins emerged as distinct along PC2. avg_inner_lambda = average inner lambda; avg_shoulder_amp = average shoulder amplitude; avg_tail = average tail duration; avg_valley = average valley duration; max_intensity = maximum intensity; PW10_mean = mean peak width (10%); PW25_mean = mean peak width (25%); min_intensity = minimum intensity; n_peak = peak count

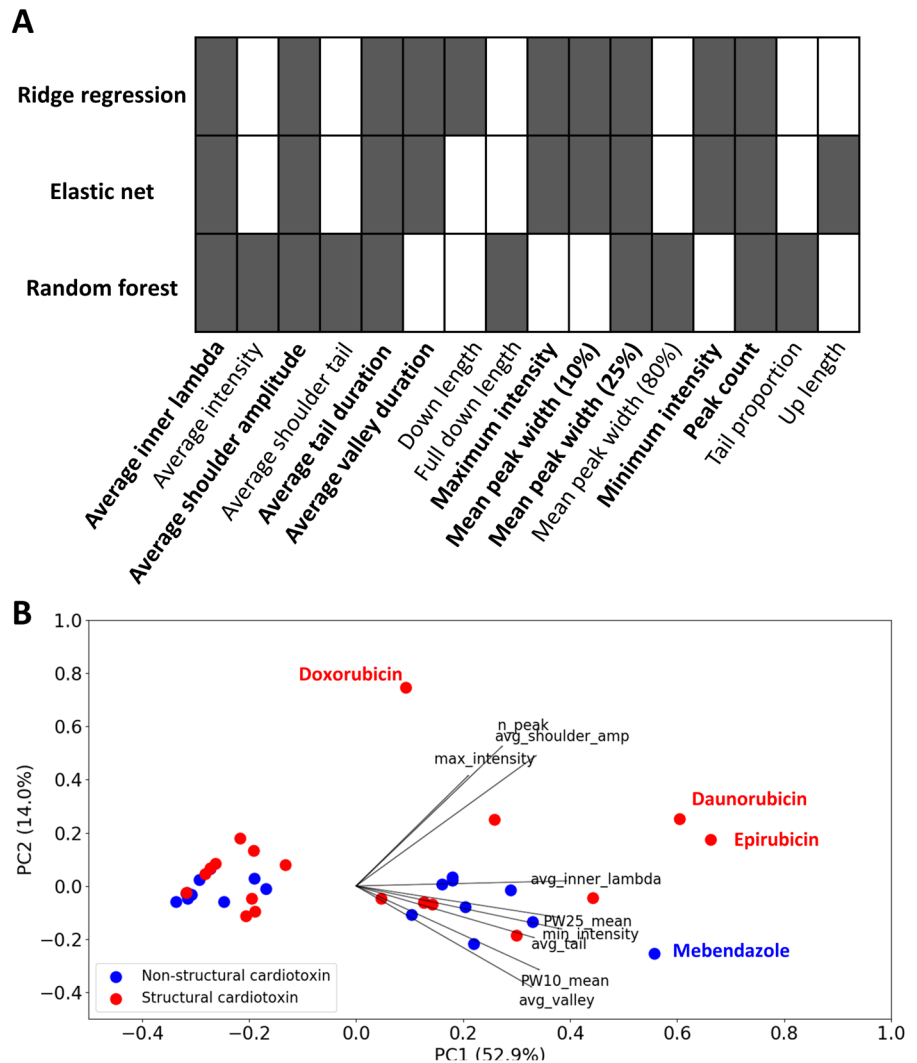


Table 1 A logistic regression classifier of nine selected calcium transient features performs better at distinguishing structural cardiotoxins from non-structural cardiotoxins than a

baseline classifier. Ten times repeated, five-fold internal cross-validation was performed on the 46 clinical compounds

Logistic regression classifier	Dataset	Area Under Curve (min-max)	Precision (min-max)	Recall (min-max)
Baseline	Cross-validation	0.61 (0.30 – 1)	0.60 (0.21 – 1)	0.61 (0.30 – 1)
Baseline	Test	0.65	0.69	0.50
9-parameter	Cross-validation	0.71 (0.30 – 1)	0.75 (0.21 – 1)	0.71 (0.30 – 1)
9-parameter	Test	0.80	0.88	0.68

induced by timepoint and concentration (Supplementary Figure 2).

Across the 12 compounds assessed, the number of differentially expressed genes (absolute

log₂(Fold-Change) > 1, FDR *p* ≤ 0.05) ranged from 0 to 2,838 (Supplementary Table 3). Compounds with similar gene expression profiles may indicate similar molecular mechanisms of cardiotoxicity; however,

compounds exhibit different potencies at different concentrations. To account for these differences, we selected the treatment with the largest number of differentially expressed genes for each compound for analysis. Hierarchical clustering based on the 4,292 genes which were significantly differentially expressed following exposure to one or more compounds identified four clusters (Supplementary Figure 3). Of these, two clusters had sufficient numbers of differentially expressed genes to provide evidence of pathway enrichment. One cluster, comprising erlotinib (48 hr high), idarubicin (48 hr high), and fluorouracil (48 hr high) (Supplementary Figure 3), had 41 over- or under-represented pathways. Enriched pathways included tRNA charging and Rho GTPase inhibition pathways, while depleted pathways included immune response and signalling pathways (Supplementary Table 4). Another cluster comprised of sorafenib (48 hr high), dasatinib (48 hr high), and sunitinib (48 hr high) (Supplementary Figure 3). Over-representation analysis of differentially expressed genes across these treatments revealed 70 over- or under-represented pathways (FDR p -value ≤ 0.05 , absolute(Z) ≥ 2). Pathways relating to DNA damage repair, cell cyclin regulation, and bile acid signalling and metabolism were up-regulated while those relating to apoptosis, IL-6/IL-8 mediated inflammation, and cell-cell

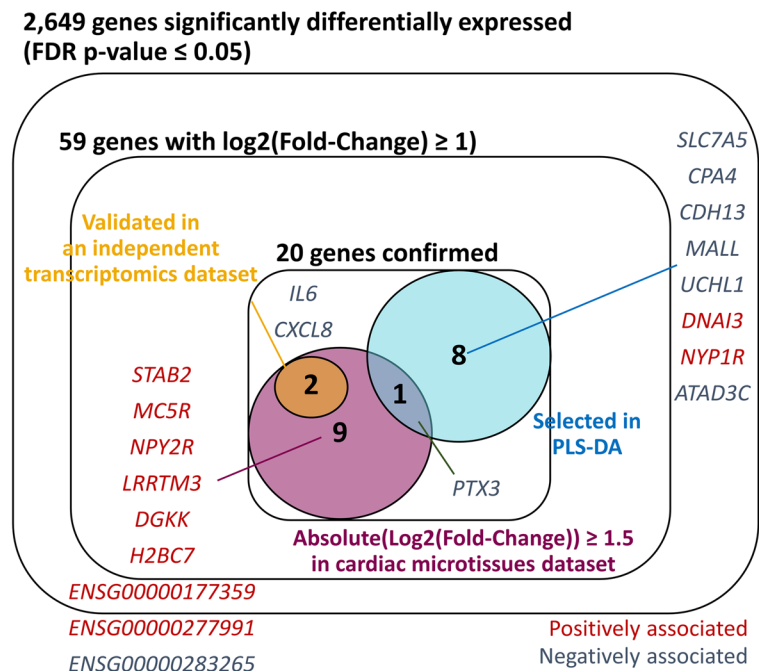
adhesion signalling were down-regulated (Supplementary Table 4). Taken together, these findings suggested that compounds exerted much larger effects on gene expression profiles irrespective of SCT class, and we therefore explored alternative approaches to identify mechanisms of SCT.

Differential expression analysis highlights 20 genes which are robustly associated with SCT

Of 18,065 assessed genes, 2,649 were significantly differentially expressed in structural cardiotoxins compared to non-structural cardiotoxins (FDR $p \leq 0.05$) (Fig. 3; Supplementary Table 5). Of the 2,649 genes, 1342 were down-regulated and 1307 were up-regulated. Over-representation analysis showed that the 2,649 genes were enriched in 222 pathways (Supplementary Table 6). The top 50 pathways related to cell-cell junctions and adhesion, extracellular matrix binding and organisation, endoplasmic reticulum stress and processes of protein folding and transport, and actin cytoskeleton remodelling (Supplementary Table 6).

Of the 2,649 genes which were differentially expressed in structural cardiotoxins compared to non-structural cardiotoxins, 59 also had an absolute effect size > 1 (Fig. 3; Supplementary Table 5). We

Fig. 3 Prioritised genes from differential gene expression analysis. Prioritisation was performed using additional methods: i) validation in an independent transcriptomic dataset (van Hasselt et al. 2020); ii) partial least squares determinant analysis, and iii) introducing a stricter $\log_2(\text{Fold-Change})$ threshold (≥ 1.5) on differential expression analysis results. See main text and experimental procedures for more detail



performed further assessment of these 59 genes using the following:

1. Replication in an independent transcriptomic profiling dataset: Differential expression analysis of 5 structural cardiotoxins compared to 16 non-structural cardiotoxins using gene expression data from van Hasselt et al. (2020). A total of 224 genes were significantly differentially expressed (FDR-adjusted $p \leq 0.05$), and 28 of these had an absolute effect size > 1 . Of the 28 genes, 2 (*IL6* and *CXCL8*) were also among the 59 prioritised genes in the cardiac microtissue dataset.
2. Partial least squares determinant analysis (PLS-DA): Hyperparameter fine-tuning showed that 2 principal components were sufficient for PLS-DA, and good separation of structural cardiotoxins and non-structural cardiotoxins was observed (Supplementary Figure 4A; B). When ranked by absolute weighting, the weights of significant genes identified in differential expression analysis showed a skewed distribution towards higher ranks compared to non-significant genes (Supplementary Figure 4C). Of the top 100 genes, 9 were also highlighted in differential expression analysis (Fig. 3).
3. Genes were also prioritised if they had absolute $\log_2(\text{Fold-Change})$ values ≥ 1.5 in the differential expression analysis performed in cardiac microtissues. Of the 59 prioritised genes, 12 had absolute effect sizes which met this threshold.

Altogether, a total of 20 genes were robustly associated with SCT, and the 17 protein-coding genes among this set were prioritised for downstream analysis (Fig. 3). Single cell expression data from GTEx v8.0 (Aguet et al. 2020) suggests that many of these genes are expressed predominantly in endothelial cells or fibroblasts, highlighting the importance of studying non-cardiomyocytes to understand mechanisms of SCT. A qRT-PCR analysis of six of the significant genes (*ATAD3C*, *CDH13*, *IL6*, *PTX3*, *SLC7A5*, *UCHL1*), as detailed in the Supplementary Experimental Procedures, showed that treated cardiac microtissues showed mainly consistent down-regulation of these genes compared to control at high concentrations after 48 hour exposures, which corroborated findings from next generation sequencing

(Supplementary Table 7; Supplementary Experimental Procedures).

Complementary WGCNA analysis identifies hub genes in biologically-relevant modules associated with SCT

Co-expressed genes are more likely to represent common biological pathways than genes which are not co-expressed. To test whether measured genes formed biologically-relevant clusters representing SCT, we performed WGCNA, a co-expression network clustering approach (Langfelder and Horvath 2008), and tested the association of module eigengenes (MEs) with SCT and related parameters from the calcium transient and HCB assays (Supplementary Experimental Procedures). At 6 hrs, we observed few differentially expressed genes across compounds (Supplementary Figure 2). Therefore, we excluded samples at 6 hrs from WGCNA analysis.

A total of 23 co-expressed gene modules were identified across 33,778 genes (Supplementary Table 8). We used random colours to name the modules. The median number of genes in a module was 433 (salmon module); the paleturquoise module had the least number of genes at 35, while the turquoise module had the maximum number of genes at 12,250. The eigengenes of these modules correlated strongly with those from modules of WGCNA networks which included vehicle controls (Supplementary Figure 5).

Of the 23 modules, 9 were associated with one or more tested phenotypes, and 5 of these (magenta, green, darkgrey, orange, and black) were associated with SCT at an adjusted p-value threshold of 5×10^{-3} (Fig. 4). The five modules associated with SCT showed strong and significant ($p \leq 0.05$) correlations of gene significance for SCT with module membership (Supplementary Figure 6).

The green ($n=2,511$), darkgrey ($n=207$), and orange ($n=731$) modules were positively-associated with SCT and with HCB phenotypes, and inversely-correlated with several calcium transient parameters (Fig. 4). Over-representation analysis of genes within each of these modules, as detailed in the Supplementary Experimental Procedures, showed that these modules represented related functional pathways (Supplementary Table 9). The green and orange modules were over-representative of genes

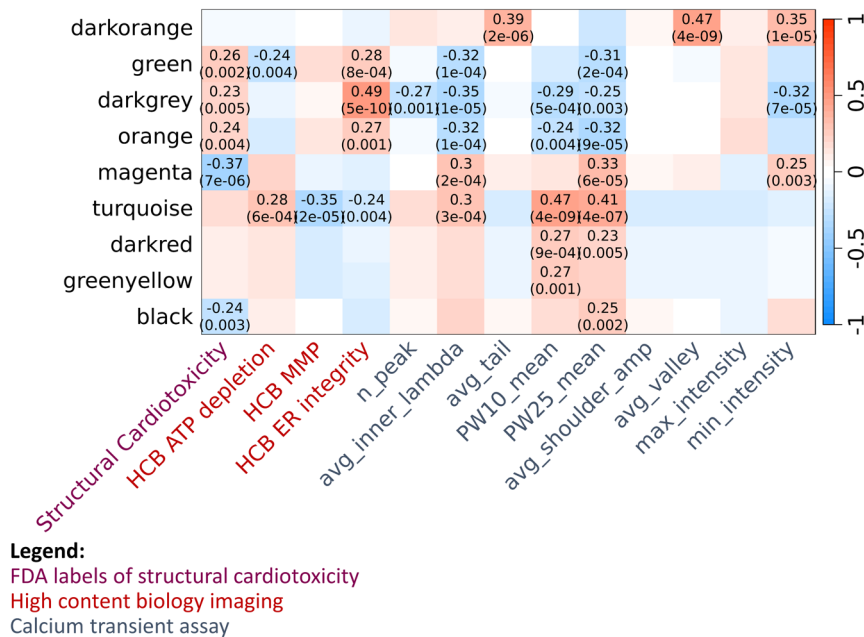


Fig. 4 ME-trait associations with SCT and related phenotypic assay parameters. Of the 23 modules identified, 9 had at least one association with SCT or a phenotypic assay parameter. Values reported in the cells are Pearson's R^2 correlations with unadjusted p-values reported in brackets; only cells reaching significance ($p \leq 5 \times 10^{-3}$) are labelled. N_peak = peak count;

avg_inner_lambda = average inner lambda; avg_tail = average tail duration; PW10_mean = mean peak width (10%); PW25_mean = mean peak width (25%); avg_shoulder_amp = average shoulder amplitude; avg_valley = average valley duration; max_intensity = maximum intensity; min_intensity = minimum intensity

associated with mitochondrial inner membrane and mitochondria-mediated energy metabolism, while the darkgrey module was over-representative of pathways relating to muscle contraction and sarcoplasmic reticulum (Supplementary Table 9). Among these modules, only the green module contained two differential expression analysis-prioritised genes (*MC5R*, *DGKK*) (Supplementary Figure 6; Supplementary Table 8).

By contrast, the magenta ($n=3,757$) and black ($n=1,120$) modules were inversely-correlated with SCT and with HCB phenotypes, and positively-correlated with several calcium transient phenotypes (Fig. 4). Over-representation analysis showed that these modules were over-representative of pathways relating to extracellular matrix organisation and structure, focal adhesion, ubiquitin-mediated protein degradation, and integrity and activity of other structural cellular components (Supplementary Table 9). The magenta module also contained nine genes prioritised in differential expression analysis (*ATAD3C*, *CDH13*,

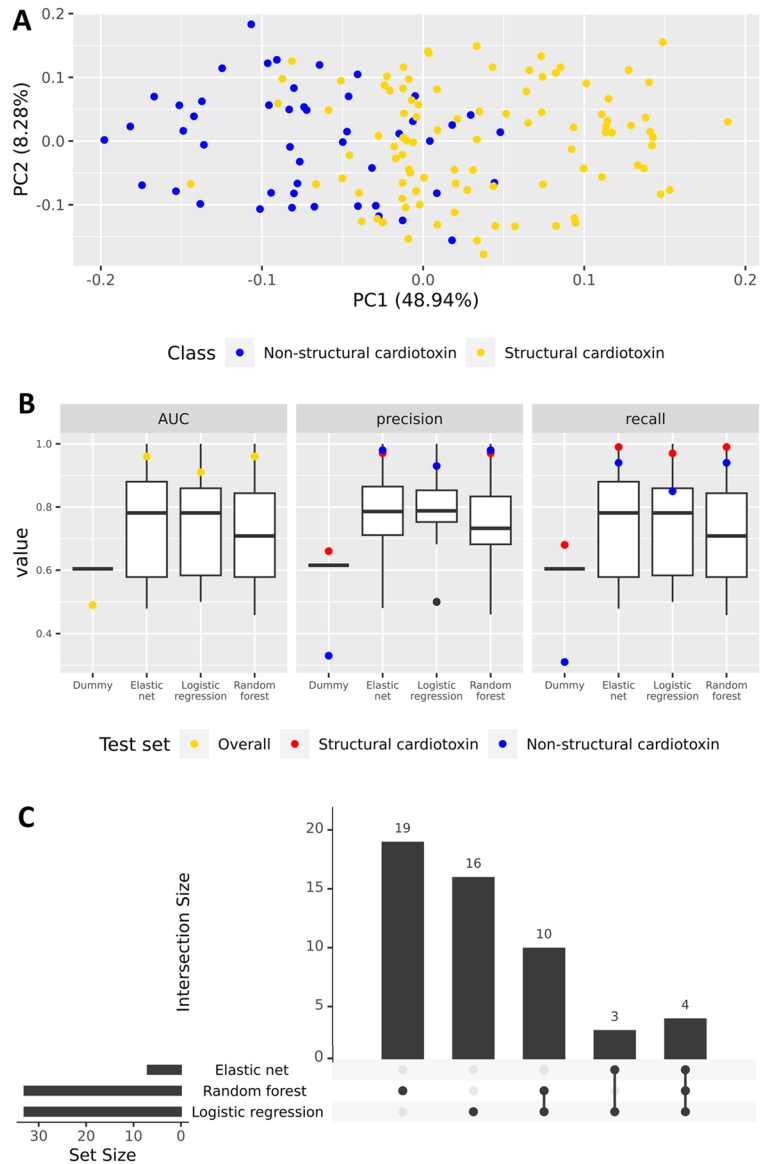
CPA4, *CXCL8*, *IL6*, *MALL*, *UCHL1*, *PTX3*, *SLC7A5*) (Supplementary Table 8).

Across the five modules implicated in SCT, we prioritised 353 genes with significant module membership (Bonferroni-corrected $p \leq 1.48 \times 10^{-6}$, adjusting for 33,778 genes) (Supplementary Figure 6), and a further 3 by intra-modular connectivity (Supplementary Figure 7). After including genes prioritised by DESeq2, we obtained a list of 367 unique genes for downstream analysis (Supplementary Table 8). Principal component analysis showed that these genes could partially separate structural cardiotoxins from non-structural cardiotoxins along PC1 (Fig. 5A).

Machine learning analysis prioritises 52 genes as important predictors of SCT

The 367 genes were uncorrelated (Pearson $R^2 < 0.95$) and had a median variance of 0.068 (minimum = 0.0066, maximum = 1.61). Of the 367 genes, 328 had high variance ($\sigma^2 \geq 0.02$) and were used as features

Fig. 5 SCT classification using 367 prioritised genes. **A)** Principal component analysis of 367 prioritised genes. **B)** Area under curve (AUC), precision, and recall of elastic net, logistic regression, and random forest models. Box-and-whisker plots represent cross-validation performance. Yellow points represent AUC performance in the full test dataset while red and blue points represent performance of structural cardiotoxins and non-structural cardiotoxins, respectively. **C)** Intersection plot summarising the number of important features which are shared and unique to each machine learning model



in machine learning analysis. All three machine learning classifiers performed better ($AUC \geq 0.91$, $precision_{\text{structural cardiotoxin}} \geq 0.93$, $recall_{\text{structural cardiotoxin}} \geq 0.97$) than a dummy classifier of SCT ($AUC = 0.49$, $precision_{\text{structural cardiotoxin}} \geq 0.66$, $recall_{\text{structural cardiotoxin}} \geq 0.68$) (Fig. 5B). Overall, the models had improved precision and recall for identifying structural cardiotoxins than non-structural cardiotoxins, which could be due to the class imbalance in the dataset (Fig. 5B).

Of the 328 gene predictors, 52 across SCT-associated modules were selected by one or more machine learning models as important features (Fig. 5C;

Table 2; Supplementary Table 10), of which five (*DNAI3*, *LRRTM3*, *NPY1R*, *NPY2R*, *SLC7A5*) were also prioritised by differential expression analysis. Single cell expression data from the Human Protein Atlas (Karlsson et al. 2021) showed that 32 genes were more highly expressed or almost exclusively expressed in non-cardiomyocytes compared to cardiomyocytes (Supplementary Table 10). Based on this evidence and on existing literature, genes were putatively assigned to one or more of 13 pathways: ‘Angiogenesis’, ‘Apoptosis’, ‘Cardiac development’, ‘Contractility’, ‘Drug metabolism’, ‘ECG’, ‘ECM/

Table 2 Summary of 52 prioritised genes and pathways relating to SCT

HGNC symbol	WGCNA module	Putatively assigned pathways	Primary cell expressed, as per Human Protein Atlas (Supplementary Table 10)
<i>ALDH1L2</i>	black	Apoptosis, Oxidative stress	Cardiac fibroblasts
<i>ASXL3</i>	green	Cardiac development, Energy metabolism	Cardiomyocytes
<i>B4GALT4</i>	black	Unknown	Cardiomyocytes
<i>BEND6</i>	black	Transcription	Cardiomyocytes
<i>CARS1</i>	magenta	Oxidative stress	Endothelial cells
<i>COL12A1</i>	magenta	ECM/Cell adhesion, Fibrosis	Cardiac fibroblasts
<i>COL15A1</i>	magenta	ECM/Cell adhesion, Fibrosis	Cardiac fibroblasts
<i>DEPP1</i>	magenta	Apoptosis, Oxidative stress	Cardiomyocytes
<i>DNAI3</i>	brown	ECM/Cell adhesion	Cardiomyocytes
<i>EPHX1</i>	magenta	Energy metabolism, Inflammation	Cardiac fibroblasts
<i>FAM13C</i>	green	Energy metabolism	Endothelial cells
<i>FNDC3B</i>	black	Angiogenesis, Fibrosis	Cardiac fibroblasts
<i>FNDC4</i>	magenta	Cardiac development, Inflammation	Cardiac fibroblasts
<i>FOS</i>	magenta	ER processes, Oxidative stress, Apoptosis	Cardiomyocytes
<i>GCLM</i>	black	Apoptosis, Oxidative stress	Endothelial cells
<i>GRB10</i>	magenta	ECM/Cell adhesion, Cardiac development, Angiogenesis	Endothelial cells
<i>HDAC4</i>	magenta	Cardiac development, Transcription, Apoptosis	Cardiomyocytes
<i>HSD11B1</i>	magenta	Angiogenesis, ER processes, Cardiac development	Cardiac fibroblasts
<i>INKA2</i>	orange	ER processes	Cardiomyocytes
<i>ITGA8</i>	magenta	ECM/Cell adhesion	Endothelial cells
<i>JDP2</i>	magenta	Transcription	Cardiac fibroblasts
<i>KIAA1217</i>	darkgrey	Cardiac development	Endothelial cells
<i>LINC00536</i>	magenta	Unknown	Not found in database
<i>LPAR5</i>	magenta	Transcription	Cardiomyocytes
<i>LPCAT1</i>	magenta	Energy metabolism	Endothelial cells
<i>LRRTM3</i>	turquoise	Unknown	Cardiomyocytes
<i>LUCAT1</i>	black	Oxidative stress, Apoptosis, Inflammation	Not found in database
<i>MGST1</i>	magenta	Oxidative stress	Cardiac fibroblasts
<i>NPY1R</i>	turquoise	Contractility, Cardiac development	Endothelial cells/Cardiac fibroblasts
<i>NPY2R</i>	turquoise	Contractility, Cardiac development	Not detected in experiments
<i>NQO1</i>	magenta	Oxidative stress	Endothelial cells
<i>NTN1</i>	magenta	ECM/Cell adhesion, Oxidative stress	Cardiomyocytes
<i>OPN3</i>	black	Unknown	Endothelial cells
<i>OSMR</i>	magenta	ECM/Cell adhesion, Angiogenesis, Inflammation	Endothelial cells
<i>PGD</i>	magenta	Energy metabolism	Cardiac fibroblasts
<i>PLSCR4</i>	black	Energy metabolism	Endothelial cells
<i>PRSS35</i>	green	ECG	Cardiomyocytes
<i>RNF24</i>	magenta	Contractility	Cardiac fibroblasts
<i>RPS6KA3</i>	black	Contractility	Endothelial cells
<i>SARS1</i>	magenta	Angiogenesis, Cardiac development	Cardiomyocytes
<i>SEL1L3</i>	magenta	Unknown	Endothelial cells
<i>SLC1A5</i>	magenta	Energy metabolism	Cardiac fibroblasts
<i>SLC7A11</i>	black	Oxidative stress	Cardiomyocytes

Table 2 (continued)

HGNC symbol	WGCNA module	Putatively assigned pathways	Primary cell expressed, as per Human Protein Atlas (Supplementary Table 10)
<i>SLC7A5</i>	magenta	Angiogenesis	Cardiac fibroblasts
<i>SRXN1</i>	magenta	Oxidative stress	Endothelial cells
<i>STAG3L3</i>	orange	Unknown	Not found in database
<i>SULT1C4</i>	green	Drug Metabolism	Endothelial cells
<i>THSD4</i>	magenta	Apoptosis, Oxidative stress, ECM/Cell adhesion	Cardiomyocytes
<i>TLL7</i>	black	Unknown	Cardiac fibroblasts
<i>TXNDC5</i>	magenta	ECM/Cell adhesion, Fibrosis, ER processes, Oxidative stress	Not detected in experiments
<i>UGCG</i>	black	Energy metabolism, Contractility	Endothelial cells
<i>WARS1</i>	magenta	Angiogenesis, Apoptosis	Endothelial cells

The module is the module assignment from WGCNA. Pathways are putatively assigned based on GO enrichment terms and literature review as detailed in Supplementary Table 10. ‘ER processes’ includes protein folding and protein synthesis, while ‘ECG’ refers to changes in intervals measured by electrocardiogram. ECG = Electrocardiography, ECM = Extracellular Matrix, ER = Endoplasmic Reticulum.

Cell adhesion’, ‘Energy metabolism’, ‘ER processes’ (including protein synthesis and folding), ‘Fibrosis’, ‘Inflammation’, ‘Oxidative stress’, and ‘Transcription’ (Table 2; Supplementary Table 10). Seven of the genes could not be assigned to a pathway due to lack of evidence and were labelled as ‘Unknown’ (Table 2; Supplementary Table 10).

We also tested whether the 5 MEs could be used to predict SCT. However, MEs could not differentiate between structural and non-structural cardiotoxins (Supplementary Figure 8A). Only a random forest classifier using four uncorrelated MEs as features could distinguish between structural and non-structural cardiotoxins. This model performed poorly in cross-validation compared to the dummy classifier (Supplementary Figure 8B).

Discussion

SCT presents a high-impact risk that is difficult to assess in early drug discovery. By applying machine learning methods to phenotypic and transcriptomic data in physiologically-relevant cardiac models across multiple cardiotoxic and non-cardiotoxic compounds, we identified additional phenotypic assay parameters and molecular markers which in concert capture diverse mechanisms of SCT.

In this study, we identified nine calcium transient parameters which were associated with SCT. Among them was peak count, which is commonly used to assess functional cardiotoxicity (Pointon et al. 2015). The other eight parameters comprised of intensity, mean peak width at 10% and 25% prominence, and average measures of the shoulder, tail, and valley durations of a peak. While these parameters performed better than peak count and average amplitude alone, we observed modest overall performance and large variations in model performance during cross-validation. This is likely due to the heterogeneous profile of structural cardiotoxins, where severe structural cardiotoxins may or may not act as inotropes (Pointon et al. 2013). A recent study also used CardioWave to derive 40 calcium transient parameters and build a random forest classifier of 48 inotropes and non-inotropes (Yang et al. 2023). The random forest model built in the previous study had a high AUC (0.90 – 0.94), though its specificity was lower (0.57 – 0.64) and only 2 of the 13 important features were shared with this study (mean peak width (10%) and mean peak width (25%)) (Yang et al. 2023). Differences in model performance and parameters identified may be explained by the different model endpoints used. While calcium transients alone are unlikely to fully explain mechanisms of SCT, they strongly

account for mechanisms of inotropy (Lee and Allen 1997). Indeed, other organelle-level assays like the HCB assay have been proven to be more predictive for SCT (Archer et al. 2018). We did not compare the predictivity of HCB-derived parameters with our calcium transient parameters in this study, however, future work may be performed to test whether parameters from these assays may be used in complement to gain a comprehensive assessment of SCT.

We also characterised the transcriptomic profiles of structural and non-structural cardiotoxins. The number of genes affected by structural cardiotoxins tended to increase with time and concentration; however, doxorubicin had a low impact on gene expression profiles (Supplementary Table 3). This could be due to doxorubicin's mechanism of action as an inhibitor of topoisomerase 2A, which is required for transcription (Hortobágyi 1997). We identified 52 genes which were expressed in one or more cardiac cell types that captured diverse mechanisms of SCT induced by one or more compounds. Based on existing literature, these genes were putatively mapped to biological mechanisms with plausible links to SCT, including extracellular matrix organisation and cell adhesion, oxidative stress, fibrosis, inflammation, apoptosis, and calcium signalling (Supplementary Table 10).

Calcium homeostasis plays an important role in cardiac contractility (Connell et al. 2020; McGregor et al. 2004) and is known to be affected by compound-induced cardiotoxicity (Pointon et al. 2013, 2015; Yang et al. 2022; Yuan et al. 2020). In this study, five genes (*NPY1R*, *NPY2R*, *RNF24*, *RPS6KA3*, and *UGCG*) were putatively linked to mechanisms of contractility. *NPY1R* and *NPY2R*, which regulate calcium channel abundance and cardiac left ventricle morphogenesis (Jacques et al. 2003; Safran et al. 2021), have been associated with left ventricular hypertrophy (Arnett et al. 2009). Of the other three genes, *RNF24* is an integral membrane protein which regulates calcium ion influx (Safran et al. 2021), *RPS6KA3* encodes a kinase which increases the beat amplitude of cardiomyocytes (Lamore et al. 2017), and *UGCG* is associated with contractility response to a known cardiotoxin, dobutamine (Andersson et al. 2021). In this study, *NPY1R* and *NPY2R* were significantly up-regulated while *RNF24*, *RPS6KA3*, and *UGCG*

were significantly down-regulated in structural cardiotoxins compared to non-structural cardiotoxins (Supplementary Table 10). These observed trends are in accordance with cardioprotective response to doxorubicin in mice (Eder and Molkentin 2011; Mattila et al. 2020; Norton et al. 2022) and indicate that they are responses to compound-induced SCT.

In heart failure, there is a shift in cardiomyocytes away from fatty acid beta-oxidation to glucose oxidation (Doenst et al. 2013). This shift has been detected in proteomic studies of compound-induced cardiotoxicity (Brandão et al. 2022), although directions of effect on proteins related to glycolysis and lipid metabolism depend heavily on the compound in question (Brandão et al. 2022). Among the 52 genes prioritised in this study, 6 (*ASXL3*, *EPHX1*, *FAM13C*, *LPCAT1*, *PLSCR4*, and *UGCG*) were linked to lipid metabolism, 1 (*PGD*) was linked to the pentose phosphate pathway, and 1 (*SLCIA5*) was linked to glutamine homeostasis (Supplementary Table 10). Among the lipid metabolism-associated genes, only *ASXL3* and *FAM13C* were up-regulated in structural cardiotoxins compared to non-structural cardiotoxins, though this is expected as *ASXL3* negatively regulates lipogenesis (Shin et al. 2014). In contrast, *PGD* and *SLCIA5* were down-regulated in structural cardiotoxins compared to non-structural cardiotoxins (Supplementary Table 5). These results indicate a greater reliance of cardiac cells on glycolysis and reduced flux through the lipid metabolism and pentose phosphate pathways (Cho et al. 2018).

Transcriptomics datasets generated using small sample sizes and bespoke study designs are often noisy and highly variable; therefore, bioinformatics methods applied to the same dataset may yield different results (Baik et al. 2020; Sánchez-Baizán et al. 2022). In this study, we therefore used complementary methods employing different assumptions and approaches to yield a list of high-confidence genes linked to SCT. Differential expression analysis models the variance and mean counts across genes to identify genes which significantly change in expression between conditions (Love et al. 2014). While this approach robustly modelled wide variations in expression across genes and provides intuitive hypothesis testing and interpretable results, it also suffered from small sample sizes which could introduce large variation or outliers and high uncertainty in estimates. By contrast, co-expression network approaches like

WGCNA constructs modules of genes which are co-expressed both at the gene level and the network topology level (Langfelder and Horvath 2008), with hub genes acting as highly-connected and representative nodes within the modules. While WGCNA therefore enables more robust functional annotation and correlation with traits of interest, it is still susceptible to small sample sizes, and identified hub genes may only represent genes correlated to the true mediating genes of SCT effects. Among the 17 protein-coding genes prioritised in differential expression analysis and included in WGCNA, 11 were members of SCT-correlated modules, thus indicating the value of employing orthogonal methods to prioritise genes.

Machine learning has recently gained traction in biomedical research as a method to identify important predictors of cardiotoxicity and cardiovascular disease (DeGroat et al. 2024; Grafton et al. 2021; Qian et al. 2023), though these methods are not without risks. Given our small dataset of twelve compounds, we used simple and more interpretable machine learning methods to identify 52 genes which could be potential biomarkers to assess and de-risk SCT. While we performed compound-specific cross-validation and fine-tuned the hyperparameters of each model to optimise model performance, the small size of the dataset our machine learning models were training on precluded the use of an independent test set. Therefore, our machine learning models may be overfit, i.e. demonstrate poor generalisability to other transcriptomics datasets. Validation presents an useful way to test the generalisability and utility of the 52 prioritised genes for SCT classification, however, at the time of writing we were not aware of other transcriptomics datasets derived from a comparable human-derived, *in vitro* cell model exposed to a diverse set of structural and non-structural cardiotoxins. As the goal of this work was to identify a putative shortlist of genes linked to SCT, future validation in independent transcriptomic studies is required to increase confidence for use in decision making during drug discovery.

We previously mentioned that existing SCT assays lack mechanistic interpretability beyond the organelle level. WGCNA analysis provided insight into observed trends of SCT-associated modules with relation to assay parameters, namely, that HCB parameters were correlated with module eigengenes in the same direction as SCT while calcium transient parameters were correlated in the opposite

direction. We also identified module eigengenes which were significantly correlated with HCB parameters and/or calcium transient parameters but not with SCT, demonstrating that these parameters may capture other compound-specific effects which could lend noise to the dataset. Although this was beyond the scope of our study, future work to understand the correlation between variations in these parameters with gene expression signatures will be valuable for improving the specificity of SCT assessment assays.

Core strengths of the study include the range of structural and non-structural cardiotoxins assessed, which allowed us to identify novel genes in pathways associated with SCT, and the use of multi-cell type, human cardiac *in vitro* models, as we identified several genes which were expressed in endothelial cells and/or fibroblasts, but not in cardiomyocytes (Supplementary Tables 51, 10). However, these strengths also presented some limitations. One limitation of this study was the hiPS-CMs we used. In particular, our hiPS-CMs presented a relatively immature phenotype which could have altered the calcium transient and transcriptomic findings we report. Nevertheless, these hiPS-CMs have previously been shown to capture transcriptomic changes linked to cardiotoxicity (Chaudhari et al. 2016; Matsa et al. 2016; McSweeney et al. 2019), and in the case of the calcium transients, present a contractile phenotype which more mature models like the PromoCell GmbH (Heidelberg, Germany) cardiomyocytes lack (van Hasselt et al. 2020, <https://promocell.com/product/human-cardiac-myocytes-hcm/#tab-description>). Finally, while the assigned pathways demonstrate putative functions of the 52 prioritised genes in SCT, we lacked the resolution to link pathways to specific cell types or identify biological pathways which might contribute more to SCT than others. Following validation of our genes, we highlight single-cell or spatial transcriptomics as one avenue of follow-up, as such studies have already successfully linked gene changes to individual cell types (Kanemaru et al. 2023; Walls et al. 2023).

Here, we showed that phenotypic assays applied to *in vitro* model systems can be used *in tandem* to capture the multi-faceted mechanisms of SCT. We also show that bioinformatic and machine learning approaches can identify a transcriptomic fingerprint of SCT which may be used to de-risk compounds in early drug discovery. Future work may expand upon

these findings by increasing the number of structural cardiotoxins studied, integrating other omics technologies to increase confidence in these findings, and assessing the genes highlighted here as potential markers in targeted assays of SCT.

Experimental Procedures

Details about cell culture, RNA-seq library preparation and sequencing, and data acquisition using the calcium transient and high content biology assays can be found in the Supplementary Experimental Procedures.

Compound selection and annotation

A total of 46 clinical compounds (48% structural cardiotoxins) representing chemically-diverse structural and non-structural cardiotoxins were included for calcium transient analysis (Supplementary Table 2). Compounds were considered structural cardiotoxins if the corresponding FDA label referenced heart failure or a decrease in left ventricular ejection fraction. Other compounds were labelled as non-structural cardiotoxins, even if the corresponding FDA label referenced non SCT-related cardiac pathologies.

A more selective approach was performed to select compounds for transcriptomic profiling. Eight structural cardiotoxins which represented a range of chemical classes contributing to SCT and which had in-house data available for the HCB assay were selected. In addition, four chemically-diverse non-structural cardiotoxins which were correctly predicted in the HCB assay were chosen to enable clear separation of structural cardiotoxins from non-structural cardiotoxins and represent a full diversity of responses.

Feature selection for calcium transient parameters associated with SCT

For feature selection, waveform parameters which were uncorrelated (Pearson $R^2 < 0.95$) and had high variance ($\sigma^2 > 0.02$) were used to build classification models of SCT. Ten times repeated, five-fold cross-validated models of SCT were built on 46 clinical compounds (48% structural cardiotoxins) with the Python scikit-learn v1.0.2 package (Pedregosa et al. 2011). To select the best models

for selecting features, seven models were compared: logistic regression, elastic net, ridge regression, random forest, support vector machine, K nearest neighbours, and Gaussian (Naïve Bayes) models. A randomised search grid algorithm was used to tune and select hyperparameters for each model. Area under the curve (AUC), precision, and recall were assessed, and models were tested on the 46 clinical compounds (48% structural cardiotoxins) used to train the model. Model performances were compared to performance of a dummy classifier which makes stratified predictions using the Python scikit-learn v1.0.2 package (Pedregosa et al. 2011). Stratified prediction accounts for potential differences in cross-fold validation performance by accounting for class imbalance in the dataset (Pedregosa et al. 2011).

The ten features with the largest absolute weightings were obtained for ridge regression and elastic net, and the ten features with the largest mean decrease in GINI were obtained for random forest. Features were considered as associated with SCT if they were selected in at least two of these selected models. To test whether selected features could discriminate between structural cardiotoxins and non-structural cardiotoxins, a logistic regression classifier of selected features was compared with a baseline classifier containing only peak count and average amplitude as features.

Transcriptomic profiling

In short, transcriptomic profiles were generated from cardiac microtissues by sequencing mRNA-seq libraries on an Illumina sequencing platform.

Cardiac microtissues, which were cultured and prepared for sequencing as detailed in the Supplementary Experimental Procedures, were exposed to 12 compounds (8 structural cardiotoxins, 4 non-structural cardiotoxins) at two concentrations (low and high) and three timepoints (6, 48, and 72 hrs) with three biological replicates and one technical replicate. High concentrations were determined for each compound based on the lowest concentration of either i) IC_{30} of ATP depletion (i.e. cell death) in cardiac microtissues in a high content biology assay associated with SCT; ii) $\leq 15\times$ total C_{max} (to maintain therapeutic relevance); or iii) lowest investigated

concentration (Archer et al. 2018). The half-log dilution of each high concentration was then used as the low concentration. ATP depletion was not seen in cardiac microtissues for any of the compounds at the low concentration. A summary of the compounds, annotations, and concentrations tested can be found in Supplementary Table 2.

Differential gene expression analysis

Raw count data was batch-corrected using ComBat-seq in the R package *sva* v3.40.0 (Zhang et al. 2020), then normalised and variance stabilisation transformed (vst-transformed) using the R package DESeq2 v1.32.0 (Love et al. 2014).

To characterise the gene expression changes induced by individual compounds, compound samples at each treatment were compared to batch- and timepoint-matched 0.1% (v/v) DMSO vehicle control samples using DESeq2 v1.32.0. Hierarchical clustering of treatments with the largest number of significant differentially expressed genes for each compound ($\text{abs}(\log_2(\text{Fold-Change})) > 1$, FDR p -value ≤ 0.05) was performed to identify similar gene expression profiles across compounds, and over-representation analysis of clustering gene expression profiles was performed using Ingenuity Pathway Analysis v90348151 (Qiagen Inc., <https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis>) (Kr amer et al. 2014) ($\text{abs}(\text{activation Z-score}) \geq 2$, FDR p -value ≤ 0.05). Differential expression analysis was also performed using DESeq2 v1.32.0 (Love et al. 2014) to characterise the gene expression changes in structural cardiotoxins compared to non-structural cardiotoxins, and over-representation analysis of significant genes was performed using the R package *clusterProfiler* v4.0.5 (Wu et al. 2021).

We also compared the gene expression profiles of structural cardiotoxins to non-structural cardiotoxins in an independent transcriptomic dataset comprising of 342 samples across 21 kinase inhibitors (5 structural cardiotoxins, 16 non-structural cardiotoxins). Samples were obtained from adult human cardiomyocytes from four healthy volunteers (two male, two female) that were exposed to kinase inhibitors at maximal concentrations for 48 hours (van Hasselt et al. 2020). Cell line-adjusted, normalised, and vst-transformed gene counts were compared between

structural cardiotoxins and non-structural cardiotoxins using DESeq2 v1.32.0 (Love et al. 2014). Significance was assessed at an FDR-adjusted p -value ≤ 0.05 .

Weighted gene correlation network analysis

Weighted gene correlation network analysis (WGCNA) (Langfelder and Horvath 2008) was performed on batch-corrected, normalised, and vst-transformed gene counts from structural cardiotoxin and non-structural cardiotoxin samples at 48 hrs and 72 hrs. Spearman's correlations were calculated for 33,778 genes with Ensembl gene IDs to create a signed network. After setting the seed to 42, a soft thresholding power of 5 was determined using the 'pickSoftThreshold' command, as it was the minimum value which exceeded an independence degree of 0.80 (Supplementary Figure 9). Distinct clusters were identified based on hierarchical clustering using Euclidean distance metrics with a distance threshold ≥ 0.25 . MEs were tested for significant correlation (adjusted $p \leq 5 \times 10^{-3}$) with the following:

1. FDA labels of SCT
2. pIC50 values of ATP depletion, ER integrity, and MMP as measured by the HCB imaging assay (Archer et al. 2018)
3. pIC50 values of calcium transient assay parameters which were associated with SCT in this study

Important genes were identified either by identifying important module members which i) had expression profiles correlated with SCT ($p \leq 1.48 \times 10^{-6}$, adjusting for 33,778 genes), or ii) were among the top 5% most connected genes within a module as measured by intramodular connectivity.

To determine whether the inclusion of vehicle controls affected gene clustering, additional WGCNA networks were constructed. These networks included i) all vehicle control, structural cardiotoxin, and non-structural cardiotoxin samples at 48 hrs and 72 hrs, and ii) vehicle control and structural cardiotoxin samples at 48 hrs and 72 hrs. To assess model robustness, MEs were correlated between networks.

WGCNA was performed using the R package WGCNA v1.72-1 (Langfelder and Horvath 2008).

Machine learning models of SCT using prioritised genes and MEs

Classifiers of SCT were built using the combined set of genes prioritised in differential expression and WGCNA analysis in scikit-learn 1.0.2 (Pedregosa et al. 2011). Batch-corrected and normalised counts of genes which were uncorrelated (Pearson $R^2 < 0.95$) and had high variance ($\sigma^2 \geq 0.02$) were included as features, and a standard scaler was applied to normalise values across the dataset. Samples at all concentrations after 6 hours were used ($n=144$, 67% structural cardiotoxins). The dataset was split into 80% training and 20% test sets, ensuring that all samples belonging to the same compound were only in the training or only in the test set, and that class balance reflected that in the full dataset after splitting. Hyperparameters to build the logistic regression, elastic net, and random forest classifiers were tuned using a randomised search grid algorithm, and three-times repeated, stratified four-fold cross-validation was performed on a per-compound basis to obtain robust results. Classifier models were compared for performance based on AUC, precision, and recall with performance from a dummy classifier built using stratified predictions in scikit-learn v1.0.2 (Pedregosa et al. 2011) to account for class imbalance. Assigned weights were used to identify important features in logistic regression and elastic net classifiers, and mean GINI decrease was used for the random forest classifier. Genes with absolute weight values in the top 90th percentile were prioritised for each model.

Classifiers of SCT were also built and assessed using MEs associated with SCT as features. The analysis performed was largely similar to that for the gene-feature classifiers except that MEs were not filtered by variance, and significant features were assessed using weight values in the top 80th percentile.

Annotation of prioritised genes

Prioritised genes from machine learning models were annotated for function based on multiple sources of evidence: i) gene ontology (GO) terms based on over-representation analysis of WGCNA modules performed with clusterProfiler v4.0.5 (Wu et al. 2021); ii) gene function, as reported in GeneCards (Safran et al. 2021); iii) differential expression analysis of structural cardiotoxins compared to non-structural

cardiotoxins in this study, cross-referenced against results from individual compounds compared to timepoint- and batch-matched controls; iv) tissue heart expression from the Gene-Tissue Expression Atlas consortium version 8 (GTEx v8.0) (Aguet et al. 2020), and v) single cell expression in cardiomyocytes, endothelial cells, and cardiac fibroblasts obtained from clusters c-2, c-6, and c-7 in the Human Protein Atlas (Karlsson et al. 2021), respectively, as queried on May 25, 2023. A literature review of the 52 genes was also performed to assess whether they were previously identified in omics studies of cardiac development or cardiotoxicity. After assembling this evidence, pathways were assigned to genes in an expansive manner based on observations.

Software

All analyses and graphics were performed and generated using R 4.1.0 (R Core Team 2021) and Python 3.8.12 (Van Rossum and Drake 2009).

Acknowledgements The authors would like to acknowledge Marta Potapo for her assistance with the RNA extractions, library preparation and sequencing, as well as Jennifer Tan, Eleanor Williams, Praveen Anand, and Abel Souza for their advice on transcriptomic data analysis.

Author contributions Conceptualization, A.P.; Methodology, VPW.AY., O.O., J.Z., H.Y., D.I., V.S., and A.H.; Software, VPW.AY. and H.Y.; Validation, A.S.C.; Formal Analysis, VPW.AY.; Investigation, VPW.AY., T.J.B., I.S., A.S.C., A.H., and I.D.; Resources, H.Y., I.D., and A.H.; Data Curation, VPW.AY.; Writing – Original Draft, VPW.AY., A.H., I.D.; Writing – Review & Editing, VPW.AY., O.O., A.H., I.D., A.Co., and A.P.; Visualization, VPW.AY.; Supervision, O.O., J.Z., D.I., V.S., A.H., and A.Co., M.R.V., and A.P.; Project Administration, A.P.; Funding Acquisition, A.P.

Funding This research was funded and supported by Astra-Zeneca PLC.

Data availability The data underlying this manuscript is available on reasonable request.

Code availability The code to perform analyses is available at <https://github.com/vauyeung38/sct-transcriptomics-waveforms>.

Declarations

Ethical approval Ethical approval was not required for data collection for this study.

Conflicts of interest VPW.AY., O.O., D.I., V.S., I.D., A.H., A.Co., and A.P. are employees and shareholders of AstraZeneca PLC. At the time this research was conducted, I.S. and A.S.C. were also employees and shareholders of AstraZeneca PLC.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aguet F, Anand S, Ardlie KG, Gabriel S, Getz GA, Graubert A, Hadley K, Handsaker RE, Huang KH, Kashin S, Li X, MacArthur DG, Meier SR, Nedzel JL, Nguyen DT, Segrè AV, Todres E, Balliu B, Barbeira AN, et al. The GTEx consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509):1318–30. <https://doi.org/10.1126/science.aaz1776>.
- Andersson L, Cinato M, Mardani I, Miljanovic A, Arif M, Koh A, Lindbom M, Laudette M, Bollano E, Omerovic E, Klevstig M, Henricsson M, Fogelstrand P, Swärd K, Ekstrand M, Levin M, Wikström J, Doran S, Hyötyläinen T, et al. Glucosylceramide synthase deficiency in the heart compromises β 1-adrenergic receptor trafficking. *Eur Heart J*. 2021;42(43):4481–92. <https://doi.org/10.1093/EURHEARTJ/EHAB412>.
- Archer CR, Sargeant R, Basak J, Pilling J, Barnes JR, Pointon A. Characterization and validation of a human 3D cardiac microtissue for the assessment of changes in cardiac pathology. *Sci Rep*. 2018;8(1):1–15. <https://doi.org/10.1038/s41598-018-28393-y>.
- Arnett DK, Devereux RB, Rao DC, Li N, Tang W, Kraemer R, Claas SA, Leon JM, Broeckel U. Novel genetic variants contributing to left ventricular hypertrophy: the HyperGEN study. *J Hypertens*. 2009;27(8):1585–93. <https://doi.org/10.1097/HJH.0b013e32832be612>.
- Baik B, Yoon S, Nam D. Benchmarking RNA-seq differential expression analysis methods using spike-in and simulation data. *PLoS One*. 2020;15(4) <https://doi.org/10.1371/JOURNAL.PONE.0232271>.
- Billingham ME, Mason JW, Bristow MR, Daniels JR. Anthracycline cardiomyopathy monitored by morphologic changes. *Cancer Treat Rep*. 1978;62(6):865–72.
- Brandão SR, Carvalho F, Amado F, Ferreira R, Costa VM. Insights on the molecular targets of cardiotoxicity induced by anticancer drugs: a systematic review based on proteomic findings. *Metab Clin Exp*. 2022;134. <https://doi.org/10.1016/J.METABOL.2022.155250>.
- Cartledge JE, Kane C, Dias P, Tesfom M, Clarke L, Mckee B, Al Ayoubi S, Chester A, Yacoub MH, Camelliti P, Terracciano CM. Functional crosstalk between cardiac fibroblasts and adult cardiomyocytes by soluble mediators. *Cardiovascular Research*. 2015;105(3):260–70. <https://doi.org/10.1093/CVR/CVU264>.
- Chaudhari U, Nemade H, Wagh V, Gaspar JA, Ellis JK, Srinivasan SP, Spitkovski D, Nguemo F, Louise J, Bremer S, Hescheler J, Keun HC, Hengstler JG, Sachinidis A. Identification of genomic biomarkers for anthracycline-induced cardiotoxicity in human iPSC-derived cardiomyocytes: an in vitro repeated exposure toxicity approach for safety assessment. *Arch Toxicol*. 2016;90(11):2763–77. <https://doi.org/10.1007/S00204-015-1623-5>.
- Cho ES, Cha YH, Kim HS, Kim NH, Yook JI. The pentose phosphate pathway as a potential target for cancer therapy. *Biomol Ther*. 2018;26(1):29–38. <https://doi.org/10.4062/biomolther.2017.179>.
- Connell P, Word TA, Wehrens XHT. Targeting pathological leak of ryanodine receptors: preclinical progress and the potential impact on treatments for cardiac arrhythmias and heart failure. *Expert Opin Ther Targets*. 2020;24(1):25–36. <https://doi.org/10.1080/14728222.2020.1708326>.
- Crisci S, Amitrano F, Saggese M, Muto T, Sarno S, Mele S, Vitale P, Ronga G, Berretta M, Di Francia R. Overview of current targeted anti-cancer drugs for therapy in oncohematology. *Medicina (Kaunas, Lithuania)*. 2019;55(8) <https://doi.org/10.3390/MEDICINA55080414>.
- DeGroat W, Abdelhalim H, Patel K, Mendhe D, Zeeshan S, Ahmed Z. Discovering biomarkers associated and predicting cardiovascular disease with high accuracy using a novel nexus of machine learning techniques for precision medicine. *Sci Rep*. 2024;14(1):1–13. <https://doi.org/10.1038/s41598-023-50600-8>.
- Deidda M, Mercurio V, Cuomo A, Noto A, Mercurio G, Dessalvi CC. Metabolomic perspectives in antitubercular cardiotoxicity and cardioprotection. *Int J Mol Sci*. 2019;20(19) <https://doi.org/10.3390/IJMS20194928>.
- Doenst T, Nguyen TD, Abel ED. Cardiac metabolism in heart failure: implications beyond ATP production. *Circ Res*. 2013;113(6):709–24. <https://doi.org/10.1161/CIRCRESAHA.113.300376>.
- Eder P, Molkentin JD. TRPC channels as effectors of cardiac hypertrophy. *Circ Res*. 2011;108(2):265–72. <https://doi.org/10.1161/CIRCRESAHA.110.225888>.
- Geisberg CA, Sawyer DB. Mechanisms of anthracycline cardiotoxicity and strategies to decrease cardiac damage. *Curr Hypertens Rep*. 2010;12(6):404–10. <https://doi.org/10.1007/S11906-010-0146-Y/METRICS>.
- Glaab WE, Holder D, He YD, Bailey WJ, Gerhold DL, Beare C, Erdos Z, Lane P, Michna L, Muniappa N, Lawrence JW, Tanis KQ, Sina JF, Skopek TR, Sistare FD. Universal toxicity gene signatures for early identification of drug-induced tissue injuries in rats. *Toxicol Sci*. 2021;181(2):148–59. <https://doi.org/10.1093/TOXSCI/KFAB038>.
- Grafton F, Ho J, Ranjbarvaziri S, Farshidfar F, Budan A, Steltzer S, Maddah M, Loewke KE, Green K, Patel S, Hoey T, Mandegar MA. Deep learning detects cardiotoxicity in a high-content screen with induced pluripotent stem

- cell-derived cardiomyocytes. *ELife*. 2021;10. <https://doi.org/10.7554/ELIFE.68714>.
- Guo F, Hall AR, Tape CJ, Ling S, Pointon A. Intra- and inter-cellular signaling pathways associated with drug-induced cardiac pathophysiology. *Trends in Pharmacological Sciences*. 2021;42(8):675–87. <https://doi.org/10.1016/J.TIPS.2021.05.004>.
- Hedin U, Thyberg J, Roy J, Dumitrescu A, Tran PK. Role of tyrosine kinases in extracellular matrix-mediated modulation of arterial smooth muscle cell phenotype. *Arterioscler Thromb Vasc Biol*. 1997;17(10):1977–84. <https://doi.org/10.1161/01.ATV.17.10.1977>.
- Hortobágyi GN. Anthracyclines in the treatment of cancer. An overview. *Drugs*. 1997;54(SUPPL. 4):1–7. <https://doi.org/10.2165/00003495-199700544-00003/METRICS>.
- Jacques D, Sader S, Perreault C, Fournier A, Pelletier G, Beck-Sickinger AG, Descorbeth M. Presence of neuropeptide Y and the Y1 receptor in the plasma membrane and nuclear envelope of human endocardial endothelial cells: modulation of intracellular calcium. *Can J Physiol Pharmacol*. 2003;81(3):288–300. <https://doi.org/10.1139/y02-165>.
- Kanemaru K, Cranley J, Muraro D, Miranda AMA, Ho SY, Wilbrey-Clark A, Patrick Pett J, Polanski K, Richardson L, Litvinukova M, Kumasaka N, Qin Y, Jablonska Z, Semprich CI, Mach L, Dabrowska M, Richoz N, Bolt L, Mamanova L, et al. Spatially resolved multiomics of human cardiac niches. *Nature*. 2023;619(7971):801–10. <https://doi.org/10.1038/s41586-023-06311-1>.
- Karakikes I, Ameen M, Termglinchan V, Wu JC. Human induced pluripotent stem cell-derived cardiomyocytes: insights into molecular, cellular, and functional phenotypes. *Circ Res*. 2015;117(1):80–8. <https://doi.org/10.1161/CIRCRESAHA.117.305365>.
- Karlsson M, Zhang C, Méar L, Zhong W, Digre A, Katona B, Sjöstedt E, Butler L, Odeberg J, Dusart P, Edfors F, Oksvold P, von Feilitzen K, Zwahlen M, Arif M, Altay O, Li X, Ozcan M, Mardinoglu A, et al. A single-cell type transcriptomics map of human tissues. *Sci Adv*. 2021;7(31) <https://doi.org/10.1126/sciadv.abh2169>.
- Krämer A, Green J, Pollard J, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*. 2014;30(4):523–30. <https://doi.org/10.1093/BIOINFORMATICS/BTT703>.
- Lamore SD, Ahlberg E, Boyer S, Lamb ML, Hortigon-Vinagre MP, Rodriguez V, Smith GL, Sagemark J, Carlsson L, Bates SM, Choy AL, Stålring J, Scott CW, Peters MF. Deconvoluting Kinase Inhibitor Induced Cardiotoxicity. *Toxicol Sci*. 2017;158(1):213–26. <https://doi.org/10.1093/TOXSCI/KFX082>.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9(1):559. <https://doi.org/10.1186/1471-2105-9-559>.
- Laverty HG, Benson C, Cartwright EJ, Cross MJ, Garland C, Hammond T, Holloway C, McMahon N, Milligan J, Park BK, Pirmohamed M, Pollard C, Radford J, Roome N, Sager P, Singh S, Suter T, Suter W, Trafford A, et al. How can we improve our understanding of cardiovascular safety liabilities to develop safer medicines? *Br J Pharmacol*. 2011;163(4):675. <https://doi.org/10.1111/J.1476-5381.2011.01255.X>.
- Lee JA, Allen DG. Calcium sensitisers: mechanisms of action and potential usefulness as inotropes. *Cardiovasc Res*. 1997;36(1):10–20. [https://doi.org/10.1016/S0008-6363\(97\)00174-0](https://doi.org/10.1016/S0008-6363(97)00174-0).
- Lorusso G, Rüegg C, Kuonen F. Targeting the extra-cellular matrix—tumor cell crosstalk for anti-cancer therapy: emerging alternatives to integrin inhibitors. *Front Oncol*. 2020;10:525701. <https://doi.org/10.3389/FONC.2020.01231/BIBTEX>.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):1–21. <https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9>.
- Matsa E, Burrige PW, Yu KH, Ahrens JH, Termglinchan V, Wu H, Liu C, Shukla P, Sayed N, Churko JM, Shao N, Woo NA, Chao AS, Gold JD, Karakikes I, Snyder MP, Wu JC. Transcriptome profiling of patient-specific human iPSC-cardiomyocytes predicts individual drug safety and efficacy responses in vitro. *Cell Stem Cell*. 2016;19(3):311–25. <https://doi.org/10.1016/J.STEM.2016.07.006>.
- Mattila M, Söderström M, Ailanen L, Savontaus E, Savontaus M. The Effects of Neuropeptide Y Overexpression on the Mouse Model of Doxorubicin-Induced Cardiotoxicity. *Cardiovasc Toxicol*. 2020;20(3):328–38. <https://doi.org/10.1007/s12012-019-09557-2>.
- McGregor E, Kempster L, Wait R, Gosling M, Dunn MJ, Powell JT. F-actin capping (CapZ) and other contractile saphenous vein smooth muscle proteins are altered by hemodynamic stress: a proteomic approach. *Mol Cell Proteomics*. 2004;3(2):115–24. <https://doi.org/10.1074/MCP.M300046-MCP200>.
- McSweeney KM, Bozza WP, Alterovitz WL, Zhang B. Transcriptomic profiling reveals p53 as a key regulator of doxorubicin-induced cardiotoxicity. *Cell Death Dis*. 2019;5(1) <https://doi.org/10.1038/S41420-019-0182-6>.
- Norton N, Bruno KA, Di Florio DN, Whelan ER, Hill AR, Morales-Lara AC, Mease AA, Sousou JM, Malavet JA, Dorn LE, Salomon GR, Macomb LP, Khatib S, Anastasiadis ZP, Necela BM, McGuire MM, Giresi PG, Kotha A, Beetler DJ, et al. Trpc6 promotes doxorubicin-induced cardiomyopathy in male mice with pleiotropic differences between males and females. *Front Cardiovasc Med*. 2022;8:757784. <https://doi.org/10.3389/fcvm.2021.757784>.
- Ott M, Gogvadze V, Orrenius S, Zhivotovsky B. Mitochondria, oxidative stress and cell death. *Apoptosis*. 2007;12(5):913–22. <https://doi.org/10.1007/S10495-007-0756-2>.
- Palmer JA, Smith AM, Gryshkova V, Donley ELR, Valentin J-P, Burrier RE. A targeted metabolomics-based assay using human induced pluripotent stem cell-derived cardiomyocytes identifies structural and functional cardiotoxicity potential. *Toxicol Sci*. 2020;174(2):218–40. <https://doi.org/10.1093/toxsci/kfaa015>.
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay É. Scikit-learn: machine learning in python. *J Mach Learn Res*. 2011;12:2825–30. <http://scikit-learn.sourceforge.net>
- Pointon A, Abi-gerges N, Cross MJ, Sidaway JE. Phenotypic profiling of structural cardiotoxins in vitro reveals

- dependency on multiple mechanisms of toxicity. *Toxicol Sci.* 2013;132(2):317–26. <https://doi.org/10.1093/TOXSCI/KFT005>.
- Pointon A, Harmer AR, Dale IL, Abi-Gerges N, Bowes J, Pollard C, Garside H. Assessment of cardiomyocyte contraction in human-induced pluripotent stem cell-derived cardiomyocytes. *Toxicol Sci.* 2015;144(2):227–37. <https://doi.org/10.1093/TOXSCI/KFU312>.
- Qian H, Qian Y, Liu Y, Cao J, Wang Y, Yang A, Zhao W, Lu Y, Liu H, Zhu W. Identification of novel biomarkers involved in doxorubicin-induced acute and chronic cardiotoxicity, respectively, by integrated bioinformatics. *Front Cardiovasc Med.* 2023;9 <https://doi.org/10.3389/FCVM.2022.996809/BIBTEX>.
- R Core Team. (2021). R: A language and environment for statistical computing (<https://www.R-project.org/>). R Foundation for Statistical Computing.
- Safran M, Rosen N, Twik M, BarShir R, Stein TI, Dahary D, Fishilevich S, Lancet D. The GeneCards Suite. In: *Practical Guide to Life Science Databases* (pp. 27–56). Springer Nature Singapore; 2021. https://doi.org/10.1007/978-981-16-5812-9_2.
- Sánchez-Baizán N, Ribas L, Piferrer F. Improved biomarker discovery through a plot twist in transcriptomic data analysis. *BMC Biol.* 2022;20(1):1–26. <https://doi.org/10.1186/S12915-022-01398-W/METRICS>.
- Sara JD, Kaur J, Khodadadi R, Rehman M, Lobo R, Chakrabarti S, Herrmann J, Lerman A, Grothey A. 5-fluorouracil and cardiotoxicity: a review. *Ther Adv Med Oncol.* 2018;10 <https://doi.org/10.1177/1758835918780140>.
- Schmidt AF, Bourfiss M, Alasiri A, Puyol-Anton E, Chopade S, van Vugt M, van der Laan SW, Gross C, Clarkson C, Henry A, Lumbers TR, van der Harst P, Franceschini N, Bis JC, Velthuis BK, Te Riele ASJM, Hingorani AD, Ruijsink B, Asselbergs FW, et al. Druggable proteins influencing cardiac structure and function: implications for heart failure therapies and cancer cardiotoxicity. *Sci Adv.* 2023;9(17) <https://doi.org/10.1126/SCIADV.ADD4984>.
- Schorlemmer A, Matter ML, Shohet RV. Cardioprotective Signaling by Endothelin. *Trends in Cardiovascular Medicine.* 2008;18(7):233–9. <https://doi.org/10.1016/J.TCM.2008.11.005>.
- Shin N, Lee Y-K, Park U-H, Jeong J-C, Um S-J. Repression of LXR α by a novel member of additional sex comb-like family, ASXL3. *Biochem Biophys Res Commun.* 2014;454(4):479–85. <https://doi.org/10.1016/j.bbrc.2014.10.074>.
- Swain SM, Whaley FS, Ewer MS. Congestive heart failure in patients treated with doxorubicin: a retrospective analysis of three trials. *Cancer.* 2003;97(11):2869–79. <https://doi.org/10.1002/CNCR.11407>.
- van Hasselt JGC, Rahman R, Hansen J, Stern A, Shim JV, Xiong Y, Pickard A, Jayaraman G, Hu B, Mahajan M, Gallo JM, Goldfarb J, Sobie EA, Birtwistle MR, Schlessinger A, Azeloglu EU, Iyengar R. Transcriptomic profiling of human cardiac cells predicts protein kinase inhibitor-associated cardiotoxicity. *Nat Commun.* 2020;11(1) <https://doi.org/10.1038/S41467-020-18396-7>.
- Van Rossum G, Drake FL. *Python 3 Reference Manual*. CreateSpace; 2009.
- Varga ZV, Ferdinandy P, Liaudet L, Pacher P. Drug-induced mitochondrial dysfunction and cardiotoxicity. *Am J Physiol Heart Circ Physiol.* 2015;309(9):H1467. <https://doi.org/10.1152/AJPHEART.00554.2015>.
- von Hoff DD, Layard MW, Basa P, Davis HL, Von Hoff AL, Rozencweig M, Muggia FM. Risk factors for doxorubicin-induced congestive heart failure. *Ann Intern Med.* 1979;91(5):710–7. <https://doi.org/10.7326/0003-4819-91-5-710>.
- Walls GM, Ghita M, Queen R, Edgar KS, Gill EK, Kuburas R, Grieve DJ, Watson CJ, McWilliam A, Van Herk M, Williams KJ, Cole AJ, Jain S, Butterworth KT. Spatial gene expression changes in the mouse heart after base-targeted irradiation. *Int J Radiat Oncol Biol Phys.* 2023;115(2):453–63. <https://doi.org/10.1016/J.IJROBP.2022.08.031>.
- Warn-Cramer BJ, Lau AF. Regulation of gap junctions by tyrosine protein kinases. *Biochim Biophys Acta.* 2004;1662(1–2):95. <https://doi.org/10.1016/J.BBAMEM.2003.10.018>.
- Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, Feng T, Zhou L, Tang W, Zhan L, Fu X, Liu S, Bo X, Yu G. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Cambridge (Mass)).* 2021;2(3) <https://doi.org/10.1016/J.XINN.2021.100141>.
- Yamaoka T, Kusumoto S, Ando K, Ohba M, Ohmori T. Receptor tyrosine kinase-targeted cancer therapy. *Int J Mol Sci.* 2018;19(11) <https://doi.org/10.3390/IJMS19113491>.
- Yang H, Obrezanova O, Pointon A, Stebbeds W, Francis J, Beattie KA, Clements P, Harvey JS, Smith GF, Bender A. Prediction of inotropic effect based on calcium transients in human iPSC-derived cardiomyocytes and machine learning. *Toxicol Appl Pharmacol.* 2023;459:116342. <https://doi.org/10.1016/J.TAAP.2022.116342>.
- Yang H, Stebbeds W, Francis J, Pointon A, Obrezanova O, Beattie KA, Clements P, Harvey JS, Smith GF, Bender A. Deriving waveform parameters from calcium transients in human iPSC-derived cardiomyocytes to predict cardiac activity with machine learning. *Stem Cell Rep.* 2022;17(3):556–68. <https://doi.org/10.1016/J.STEMCR.2022.01.009>.
- Youle RJ, Van Der Blik AM. Mitochondrial fission, fusion, and stress. *Science.* 2012;337(6098):1065. <https://doi.org/10.1126/SCIENCE.1219855>.
- Yuan Y, Fan S, Shu L, Huang W, Xie L, Bi C, Yu H, Wang Y, Li Y. Exploration the mechanism of doxorubicin-induced heart failure in rats by integration of proteomics and metabolomics data. *Front Pharmacol.* 2020;11. <https://doi.org/10.3389/FPHAR.2020.600561>.
- Zhang S, Liu X, Bawa-Khalife T, Lu LS, Lyu YL, Liu LF, Yeh ETH. Identification of the molecular basis of doxorubicin-induced cardiotoxicity. *Nat Med.* 2012;18(11):1639–42. <https://doi.org/10.1038/NM.2919>.
- Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform.* 2020;2(3) <https://doi.org/10.1093/NARGAB/LQAA078>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.