Determining the genetic basis of anthracycline-cardiotoxicity by molecular response QTL mapping in induced cardiomyocytes

David A Knowles[1,2,3,*], Courtney K Burrows[4,*], John D Blischak[4], Kristen M Patterson[4], Carole Ober[4], Jonathan K Pritchard[1,5,6], Yoav Gilad[4,7]

[1] Department of Genetics, Stanford University, Stanford, CA

[2] Department of Computer Science, Stanford University, Stanford, CA

[3] Department of Radiology, Stanford University, Stanford, CA

[4] Department of Human Genetics, University of Chicago, Chicago, IL

[5] Department of Biology, Stanford University, Stanford, CA

[6] Howard Hughes Medical Institute, Stanford University, CA.

[7] Department of Medicine, University of Chicago, Chicago, IL

* These authors contributed equally to this work.

Correspondence should be addressed to D.A.K. (dak33@stanford.edu) or Y.G. (gilad@uchicago.edu).

Abstract

Anthracycline-induced cardiotoxicity (ACT) is a key limiting factor in setting optimal chemotherapy regimes for cancer patients, with almost half of patients expected to ultimately develop congestive heart failure given high drug doses. However, the genetic basis of sensitivity to anthracyclines such as doxorubicin remains unclear. To begin addressing this, we created a panel of iPSC-derived cardiomyocytes from 45 individuals and performed RNA-seq after 24h exposure to varying levels of doxorubicin. The transcriptomic response to doxorubicin is substantial, with the majority of genes being differentially expressed across treatments of different concentrations and over 6000 genes showing evidence of differential splicing. Overall, our observations indicate that splicing fidelity decreases in the presence of doxorubicin. We detect 376 response-expression QTLs and 42 response-splicing QTLs, i.e. genetic variants that modulate the individual transcriptomic response to doxorubicin in terms of expression and splicing changes respectively. We show that inter-individual variation in transcriptional response is predictive of cell damage measured *in vitro* using a cardiac troponin assay, which in turn is shown to be associated with *in vivo* ACT risk. Finally, the molecular QTLs we detected are enriched in lower ACT GWAS *p*-values, further supporting the *in vivo* relevance of our map of genetic regulation of cellular response to anthracyclines.

Introduction

Anthracyclines, including the prototypical doxorubicin, continue to be used as chemotherapeutic agents treating a wide range of cancers, particularly leukemia, lymphoma, multiple myeloma, breast cancer, and sarcoma. A well-known side-effect of doxorubicin treatment is anthracycline-induced cardiotoxicity (ACT). For some patients ACT manifests as an asymptomatic reduction in cardiac function, as measured by left ventricular ejection fraction (LVEF), but in more extreme cases ACT can lead to congestive heart failure (CHF). The risk of CHF is dosage-dependent: an early study[1] estimated 3% of patients at 400 mg/m2, 7% of patients at 550 mg/m2, and 18% of patients at 700 mg/m2 develop CHF, where a more recent study puts these numbers at 5%, 26% and 48% respectively[2]. Reduced LVEF shows a similar dosage-dependent pattern, but is not fully predictive of CHF. Perhaps most daunting for patients is that CHF can occur years after treatment: out of 1,807 cancer survivors followed for 7 years in a recent survey a third died of heart diseases compared to 51% of cancer recurrence[3].

Various candidate gene studies have attempted to find genetic determinants of ACT, but are plagued by small sample sizes and unclear endpoint definitions, resulting in limited replication between studies. Two ACT genome-wide association studies (GWAS) have been published[4,5]. While neither found genome-wide significant associations using their discovery cohorts, both found one variant that they were able to replicate in independent cohorts.

A nonsynonymous coding variant, rs2229774, in *RARG* (retinoic acid receptor γ) was found to be associated with pediatric ACT using a Canadian European discovery cohort of 280 patients[4], and replicated in both a European ($p = 0.004$) and non-European cohort ($p = 1 \times 10^{-4}$). Modest signal ($p = 0.076$) supporting rs2229774's association with ACT was also reported in a recent study primarily focused on trastuzumab-related cardiotoxicity[6]. *RARG* negative cell lines have reduced retinoic acid response element (RAREs) activity and reduced suppression of *Top2b*[4], which has been proposed as a mediator of ACT.

In a different study, a GWAS in 845 patients with European-ancestry from a large adjuvant breast cancer clinical trial, 51 of whom developed CHF, found no variants at genome-wide significance levels[5]. However, one of the most promising variants, rs28714259 ($p = 9 \times 10^{-6}$ in discovery cohort), was genotyped two further cohorts and showed modest replication ($p = 0.04, 0.018$). rs28714259 falls in a glucocorticoid receptor protein binding peak, which may play a role in cardiac development.

An exciting approach to studying complex phenotypes, including disease, in human is to use induced pluripotent stem cells (iPSC) and derived differentiated cells as *in vitro* model systems for disease. Work by us and others has demonstrated that iPSCs and iPSC-derived cell-types are powerful model systems for understanding cell-type specific genetic regulation of transcription[7,8,9,10,11], but it is less established whether these systems can be used to model the interplay of genetic and environmental factors in disease progression. Encouragingly, the response of iPSC-derived cardiomyocytes (ICs) to doxorubicin was recently extensively characterized[12]. ICs derived from four individuals who developed ACT after doxorubicin treatment ("DOXTOX" group) and four who did not ("DOX" group), showed clear differences in viability (via apoptosis), metabolism, DNA damage, oxidative stress and mitochondrial function when exposed to doxorubicin. These observations suggest that ICs recapitulate *in vivo* inter-individual differences in doxorubicin sensitivity. Gene expression response differences between the DOX and DOXTOX groups were found using RNA-sequencing data, but the sample size was insufficient (RNA-seq was generated for only 3 individuals in each group) to attempt mapping of genetic variants that might explain the observed functional differences between individuals.

Here we used a panel of iPSC-derived cardiomyocytes from 45 individuals, exposed to five different drug concentrations, to map the genetic basis of inter-individual differences in doxorubicin-sensitivity. We find hundreds of genetics variants that modulate the transcriptomic response, including 42 that act on alternative splicing. We show that the IC transcriptomic response predicts cardiac troponin levels in culture (indicative of cell lysis) in these cell-lines, and that troponin level is itself predictive of ACT. Finally we demonstrated that the mapped genetic variants show significant enrichment in lower ACT GWAS p-values.

Results

Measuring transcriptomic response to doxorubicin exposure

We generated iPSC-derived cardiomyocytes (ICs) for 45 Hutterite individuals (Figure 1a), and confirmed cardiomyocyte identity (see Methods). We exposed all 45 IC lines to doxorubicin at 5 different concentrations for 24 hours, after which samples were processed for RNA-sequencing. We obtained sufficient read depth (10M exonic reads) for downstream analysis for 217 of the $5 \times 45 = 225$ individual-concentration pairs, and confirmed sample identity by calling exonic SNPs (see Methods). We observed a strong gene regulatory response to doxorubicin across all concentrations, with 98% (12038 / 12317) of quantifiable genes (5% FDR) showing differential expression across the different treatment concentrations. Our data shows excellent concordance with an existing smaller RNA-seq dataset[12] (Supplementary Figure S1). Principal component analysis (PCA, Figure 1b) confirms that the main variation in the data is driven by doxorubicin concentration

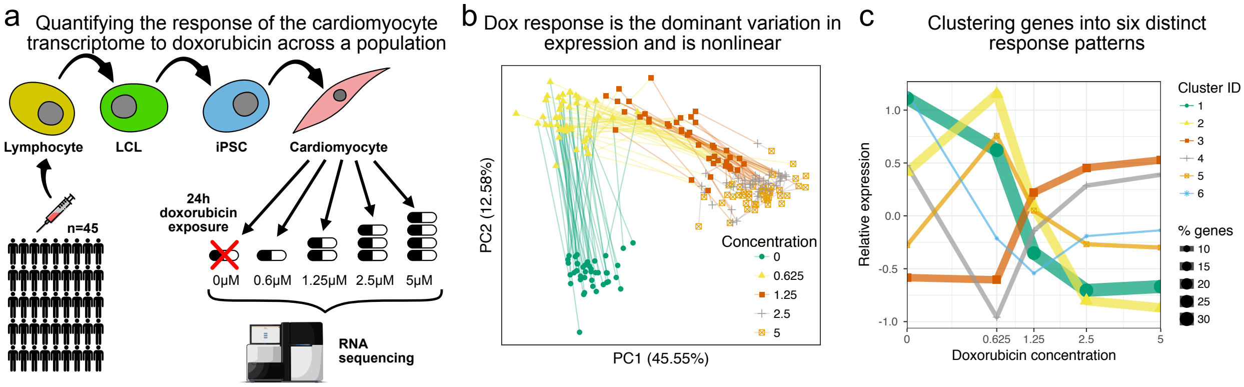


Figure 1: *The transcriptomic response of cardiomyocytes to doxorubicin is substantial.* ***a.*** *Cardiomyocytes were derived from lymphoblastoid cell lines (LCLs) of 45 Hutterite individuals, followed by exposure to differing concentrations of doxorubicin and RNA-sequencing.* ***b.*** *PCA of gene expression levels across samples reveals that doxorubicin concentration explains more variance than inter-individual differences, and that the response is non-linear with respect to concentration.* ***c.*** *A probabilistic mixture model uncovers six distinct patterss of response across genes.*

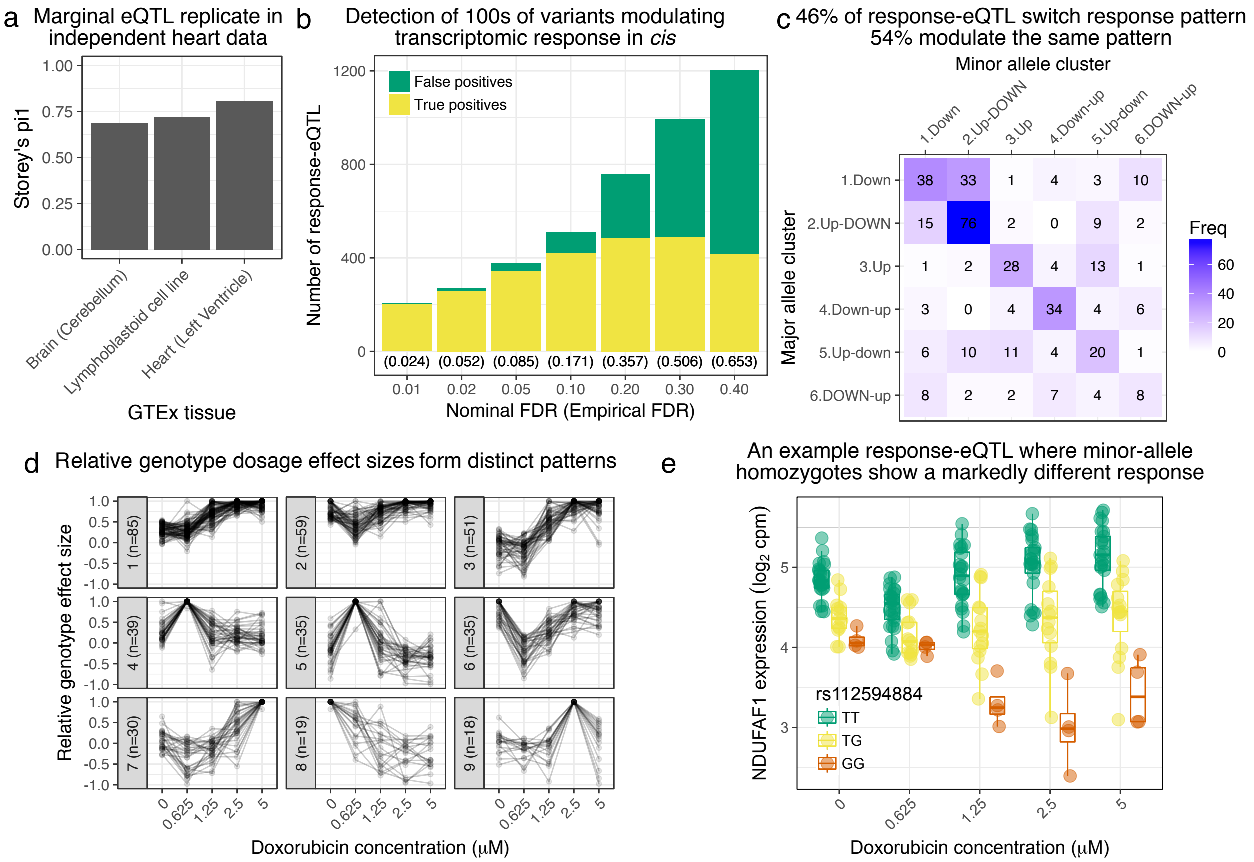
and that the effect of concentration on gene expression is nonlinear. For some individuals the expression data following doxorubicin treatment with $1.25\mu M$ is closer to the data from treatment with $0.625\mu M$, whereas for others it is closer to data from treatment with $2.5\mu M$. This general pattern provides the first indication in our data that that there is systematic variation in how different individuals respond to doxorubicin exposure. Since the majority of genes appear responsive to doxorubicin we clustered genes into six distinct response patterns using a mixture model approach (Figure 1c, see Methods). From largest to smallest, these clusters represent genes that, through the gradient from low to high concentration treatments, are 1) down regulated 2) initially up-regulated, then further down-regulated 3) up-regulated 4) transiently down regulated 5) transiently up-regulated 6) down-regulated then partially recover (Supplementary Table 1). Gene set enrichments (Supplementary Figure S2, Supplementary Table 2) for the up-regulated cluster include metabolic, mitochrondrial and extracellular processes, as well as known doxorubicin response genes in breast cancer cell lines[13]. The down-regulated cluster shares genes with those down-regulated in response to UV light, which, like doxorubicin, causes DNA-damage. Targets of p53, a transcription factor that responds to DNA damage, are overrepresented in clusters 2 and 5; these clusters involve up-regulation at low concentrations ($0.625\mu M$) but down-regulation at higher concentrations. Promoter analysis (Supplementary Figure S3, Supplementary Table 3) revealed 21, 45, and 6 significantly enriched transcription factor (TF) binding motifs for clusters 1, 2 and 3 respectively (and none for cluster 4-6). Examples include binding sites for *ZNF145*, a TF that promotes *GPX1* activity and protects cells from oxidative damage during mitochondrial respiratory dysfunction[14], which is enriched in cluster 1 (down regulation w/ dox); *RONIN*, a regulator of mitochrondrial development and function[15], which is enriched in clusters 1 and 2; and *MEF2*, myocyte enhancer factor 2, involved in regulating muscle development, stress-response and p38-mediated apoptosis[16], enriched in cluster 4 (although only at $q = 0.33$).

Mapping variants modulating doxorubicin response

We next sought to map single nucleotide polymorphisms (SNPs) that modulate the observed inter-individual transcriptomic response to doxorubicin, leveraging available genetic variation across the 45 individuals[17]. We developed a linear mixed model approach, called `suez` that extends the PANAMA framework[18] to account for relatedness amongst individuals, repeat measurements, multiple conditions and latent confounding. Testing SNPs within 1Mb of the transcription start site (TSS), 518 genes have a variant with a detectable marginal effect on expression (5% FDR, Supplementary Table 4). Reassuring, these eQTLs show strong replication (Storey's $\pi_1 = 0.80$) in GTEx heart tissue, and do so more strongly than in GTEx brain ($\pi_1 = 0.69$) or lymphoblastoid cell line ($\pi_1 = 0.72$) data (Figure 2a). Remarkably, even with our moderate number of individuals, we are able to detect many response-eQTLs (reQTLs), i.e. variants that modulate (directly or indirectly) transcriptomic response to doxorubicin. We found reQTLs for 376 genes at a nominal 5% FDR (Supplementary Table 5), which we estimate using a parametric bootstrap corresponds to a true FDR of 8.5% (Figure 2b).

To characterize the detected reQTLs we assigned the response of the major and minor allele to one of the six clusters previously learned (Figure 1c), with heterozygotes expected to display the average of the two homozygous responses. 172 (46%) of reQTLs result in a qualitatively distinct response as determined by the two alleles being assigned to different clusters. The most common transition, occurring for 33 reQTLs, is that the major allele is associated with simple down-regulation (cluster 1) in response to doxorubicin, whereas the minor allele shows a transient up-regulation at low concentration followed by down-regulation at higher concentration (cluster 2).

We further broke-down the significant reQTLs by considering the effect of genotype on expression at each concentration (β_c in Equation 6). We normalized the effect sizes relative to the β_c with the largest absolute value, i.e. we consider $\beta_c/\beta_{\arg\max|\beta_{c'}|}$, so that the largest genotype effect always corresponds to a normalized value of 1. The resulting normalized effect profiles were split into 9 clusters using k-means clustering (Figure 2d). The largest cluster (cluster 1, 85 reQTLs) represents reQTLs with a modest effect size at low concentrations $(0, 0.625\mu M)$ which is amplified at higher concentrations (Figure 2e shows a



	1.Down	2.Up-DOWN	3.Up	4.Down-up	5.Up-down	6.DOWN-up
1.Down	38	33	1	4	3	10
2.Up-DOWN	15	76	2	0	9	2
3.Up	1	2	28	4	13	1
4.Down-up	3	0	4	34	4	6
5.Up-down	6	10	11	4	20	1
6.DOWN-up	8	2	2	7	4	8

Figure 2: *Genetic variation regulates the transcriptomic response to doxorubicin exposure.* ***a.*** *Marginal eQTLs show strong replication in GTEx heart data, and lower replication in other tissues.* ***b.*** *We detect 100s of response-eQTLs (reQTLs): variants that modulate response to doxorubicin. The false positive rate (FPR) is estimated using a parametric bootstrap.* ***c.*** *We developed a statistical method to assign the major and minor allele response to one of the six clusters from Figure 1c. The strongest 46% of detected reQTLs result in a discretely different response, whereas the remainder only modulate the response.* ***d.*** *For significant reQTLs we calculated relative genotype effect sizes by dividing the fitted effect size at each concentration by the (signed) effect size with the largest absolute value. K-means clustering of these effect size profiles reveals distinct patterns, the most common being a small reduction in absolute effect size from* 0 *to* 0.625μM *followed by the largest effects being at the highest concentrations.* ***e.*** *An example response-eQTL where rs112594884 regulates the response of the mitochondrial complex I chaperone NDUFAF1. Under the major (T) allele we see moderate down-regulation at* 0.625μM *followed by up-regulation at higher concentrations. Under the minor (G) allele, there is little change at* 0.625μM *followed by substantial down-regulation. Since the genotype effects are reduced at* 0.625μM *and largest at high concentrations this reQTL is assigned to cluster 1 of panel* ***d****.*

highly significant example). Cluster 2 corresponds to reQTLs whose effect size is attenuated at the $0.625\mu M$ treatment: examples of reQTLs in this cluster tend to be associated with higher expression level at the $0.625\mu M$ treatment (e.g. rs16853200's association with ABCA12 response, Supplementary Figure S4).

Doxorubicin exposure reduces splicing fidelity

Oxidative stress, a major downstream consequence of doxorubicin exposure, disrupts splicing of individual genes including *HPRT*, *POLB*[19], and *SMA*[20]. We queried the extent to which doxorubicin exposure disrupts splicing patterns across the transcriptome using `LeafCutter`[21]. Across all samples `LeafCutter` detected 27769 alternative splicing "clusters" (referred to here as "ASCs" to avoid confusion with k-means clusters) , which correspond approximately to splicing events, with a median of 3.0 splice junctions per ASC. Of these, 10430 (59%) ASCs, corresponding to 6398 unique genes, showed an effect of doxorubicin exposure on splicing outcomes (5% FDR, Supplementary Tables 6-7). To characterize these changes we calculated the entropy of the splicing choices made for each significant ASC at each concentration and used k-means clusters patterns of change in entropy (Figure 3a). The largest cluster has 6166 ASCs (59%), and corresponds to the null of no clear change in entropy across concentrations. Clusters 2 ($n = 1136$) and 5 ($n = 475$) correspond to increasing entropy with concentration, and clusters 3, 4, 6, 8 and 9 correspond to the maximum entropy being at different concentrations and reaching different maximum levels. Interestingly, only the relatively small cluster 8 ($n = 304, 3\%$ of ACSs) corresponds to a reduction in entropy at higher concentrations, suggesting the dominant behavior is reduced splicing fidelity and increased alternative splicing in response to doxorubicin.

We further tested the hypothesis that splicing fidelity decreases in the presence of doxorubicin by comparing patterns of intronic percent excised (Ψ) with canonical vs cryptic (unannotated) splice site usage. We clustered the 7792 introns in significantly differentially spliced ASC, that have a change in percent excised ($\Delta\Psi$) > 0.1 for some pair of concentrations, into 8 response patterns based on their relative excision proportions across concentrations. For each cluster we calculated the proportion of member introns with neither end annotated, one end unannotated, or both ends annotated (Figure 3b). The clusters representing increased Ψ with concentration (clusters 2, 4, 6 and 7) all show enrichment for cryptic splice site usage. The two most populous clusters (1 and 2) correspond to Ψ decreasing and increasingly continuously with doxorubicin concentration, respectively, and the difference in levels of cryptic splicing is extremely apparent (hypergeometric $p < 2 \times 10^{-16}$, odds ratio for one annotated end vs two is 28.0).

We additionally used `LeafCutter` quantification of percentage spliced in (PSI) for each splice junction to

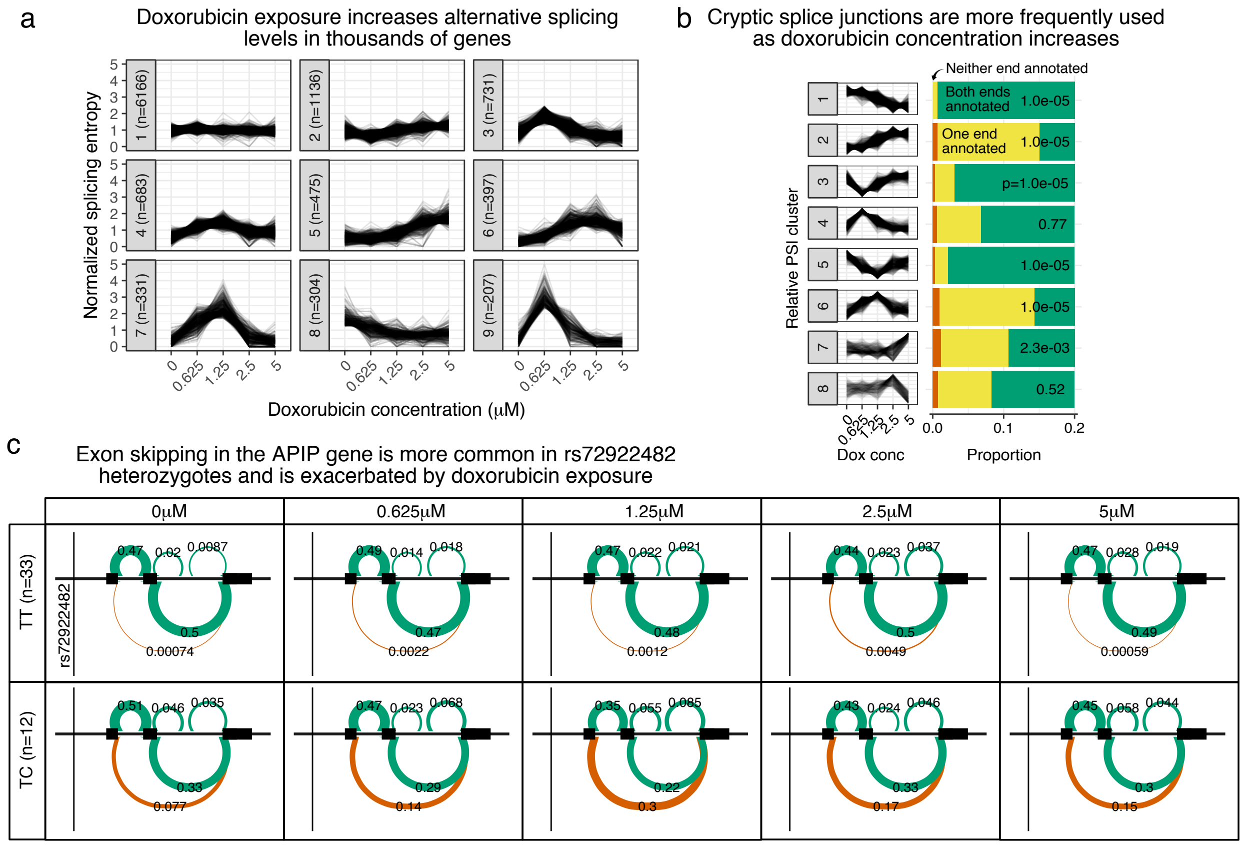


Figure 3: *Doxorubicin exposure significantly impacts alternative splicing.* ***a*** *The entropy of splicing choices increases in response to doxorubicin exposure. We measured splicing entropy at different concentrations within* `LeafCutter` *"Alternative Splicing Clusters" (ACSs) and clustered these into patterns of entropy change.* ***b*** *We separated introns differentially excised with* $\Delta\Psi > 0.1$ *into 8 clusters based on their relative excision level at each concentration. Introns in clusters corresponding to increased excision at higher doxorubicin concentrations (e.g. cluster 2) are far more likely to use a cryptic (unannotated) splice site at at least one end. p-values shown are for a hypergeometric test of that cluster against all others.* ***c*** *We mapped 42 ASCs with response splicing QTLs, variants that modulate the differential splicing response to doxorubicin.*

map splicing QTLs (sQTL) and response-splicing QTLs (rsQTL) using the same methodology as we employed for reQTL mapping. We tested SNPs within 100kb of either end of the splice junction. At 5% FDR we found 467 ASCs with a marginal effect sQTL (Supplementary Table 8) and 42 with a rsQTL (Supplementary Table 9). An example rsQTL is rs72922482's association with inclusion of exon 2 of APAF1 Interacting Protein (APIP). Under the major T allele exon skipping is extremely rare: the `LeafCutter` PSI for the spanning junction ranges from 0.00059 to 0.0049 across concentrations (Figure 3c). In rs72922482 heterozygotes, however, the exon is skipped in a significant proportion of transcripts, and this effect is most pronounced in the data collected after treatment at $1.25\mu M$, with approximately 50% exon inclusion, suggesting the minor C allele results in very low inclusion of the cassette exon. Another interesting example is *NDUFAF6*, another mitochrondrial Complex I protein, where doxorubicin exposure (particularly at $0.625\mu M$) results in increased use of an alternative downstream transcription start site (TSS) which unmasks the influence of rs896853 on a cassette exon between the two alternative TSS (Supplementary Figure S5).

Transcriptional response to doxorubicin is predictive of in-vitro cardiac-damage indicator troponin

We used the level of cardiac troponin released into the culture media by lysed cardiomyocytes (see Methods, Supplementary Table 10) to estimate damage occurring as a result of doxorubicin exposure at different concentrations. We observed significant variation in measurable damage caused by doxorubicin across individuals, with 13 of 45 cell lines having a significant correlation between doxorubicin dose and troponin measurement (Figure 4a). We first sought to determine whether the inter-individual variation in troponin in culture could be explained by variation in the overall gene expression response. Since we are interested in this case in inter-individual differences rather than differences between concentration we normalized the troponin measurements to have 0 mean and variance of 1 across samples at each doxorubicin treatment. We found 96.1% (95% credible interval [91.5%98.6%]) or 91.5% of the variance in this normalized troponin level could be explained using gene expression levels (we excluded the troponin genes *TNNT1-3* and *TNNI1-3* from the analysis) at the corresponding doxorubicin concentrations, using a GREML-analysis[22] or leave-out-one cross validated LASSO[23] respectively. The optimal LASSO model included 118 genes (Supplementary Table 11).

To further explore the relationship between transcriptomic response and troponin presence in culture, we analyzed differential expression (DE) with respect to troponin measurement at each doxorubicin concentration separately. We found 0, 7, 78, 2984 and 2863 differentially expressed genes (5% FDR, Supplementary Table 12) at the 5 concentrations respectively (Figure 4b). The most strongly DE gene (with respect to

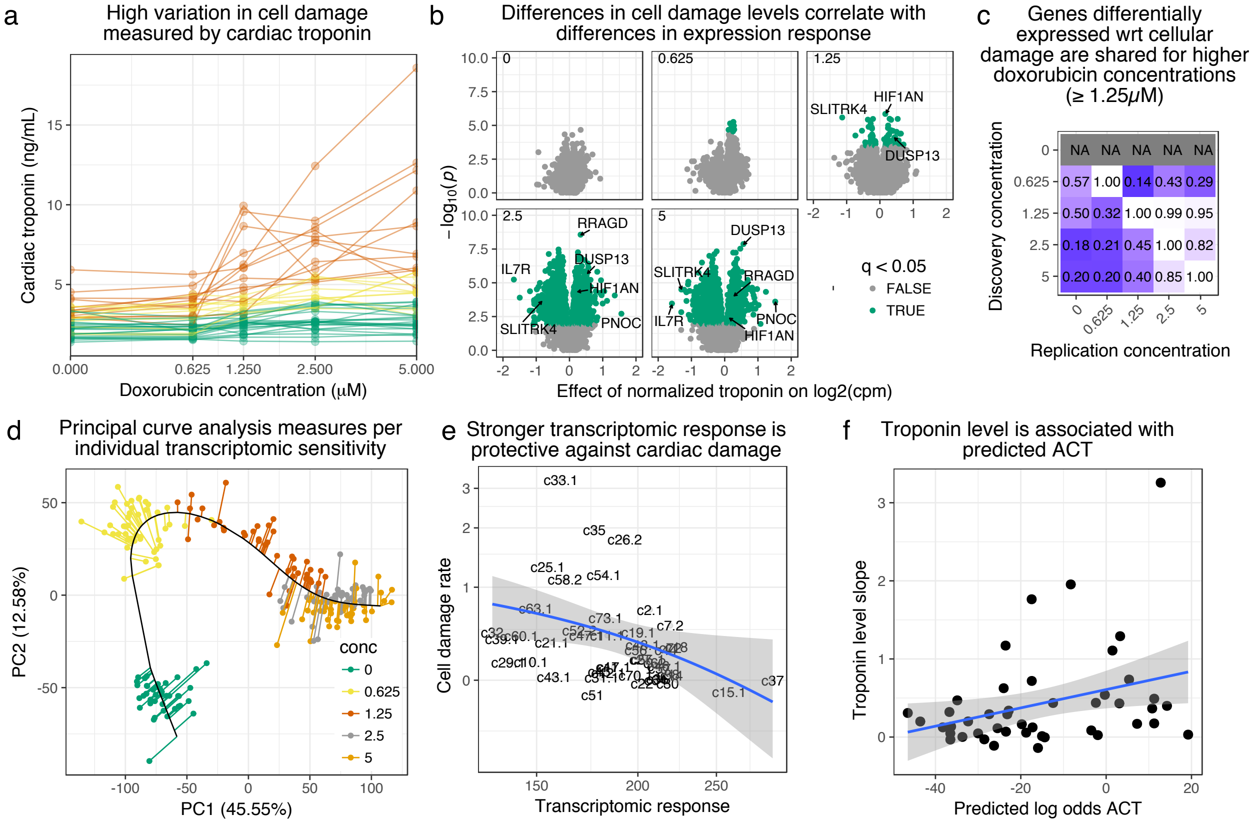


Figure 4: *Transcriptomic response is predictive of doxorubicin induced damage as measured by cardiac troponin.* ***a.*** *We measured cardiac troponin, a sensitive and specific test for myocardial cell damage, in response to doxorubicin, across all cell lines.* ***b.*** *We performed differential expression analyses with respect to troponin at each concentration separately, and observed more differentially expressed genes at higher concentrations corresponding to an increased dynamic range of troponin levels.* ***c.*** *We took differentially expressed genes (5% FDR) at each concentration and checked for "replication" (nominal $p < 0.05$) at the other concentrations. Note that no differentially expressed genes were discovered in control condition ($0\mu M$).* ***d.*** *We summarized gene expression response by first fitting a "principal curve" following increasing doxorubicin concentration, and then measuring the rate of progression along this curve for each individual.* ***e.*** *Increased transcriptomic response is associated with reduced cardiac troponin levels, suggesting that the bulk of expression changes we observe are in fact protective against cardiac damage.* ***f.*** *We trained a model to predict ACT risk from gene expression response using available 3 v. 3 case/control data*[12] *and applied this model to our data. Predicted ACT risk correlated significantly with the slope of troponin level (Spearman $\rho = 0.38, p = 0.01$), supporting the* in vivo *disease relevance of our IC system.*

effect size) at the $5\mu M$ treatment is *DUSP13*, a known regulator of *ASK1*-mediated apoptosis[24]. The large number of DE genes at the $2.5\mu M$ and $5.0\mu M$ treatments are broadly shared (nominal replication rate 82% to 85%), and DE genes at the $1.25\mu M$ treatment generally represent the most strongly DE genes at the higher concentrations (Figure 4c).

To compare troponin measurements to transcriptomic response we determined an overall per-individual level of transcriptomic response with respect to doxorubicin concentration. To this end we fit a principal curve[25] through all gene expression samples, initializing the curve to pass sequentially through the successive doxorubicin concentrations (Figure 4d). Projecting every sample on the principal curve gives a single measure of "progression" through response to doxorubicin at increasing concentrations. We then regressed these values against concentration for each individual to obtain a progression rate. We found the troponin measurement slope is significantly negatively correlated (Spearman $\rho = -0.42, p = 0.004$, Figure 4e) with the transcriptomic response rate, suggesting that much of the gene expression program being activated in response to doxorubicin is in fact protective against cardiac damage.

Using previously published data[12], we built a predictive model of ACT risk trained on RNA-seq of ICs exposed to $1\mu M$ doxorubicin from doxorubicin-treated patients who did ("DOXTOX", $n = 3$) or did not ("DOX", $n = 3$) develop ACT. Using LASSO with fixed $\lambda = 10^{-5}$ the optimal model included 17 genes as features (Supplementary Table 13). We applied this model to our expression data from the $0.625\mu M$ treatment (since this concentration shows excellent concordance with the $1\mu M$ data of Burridge et al., see Supplementary Figure S1) to obtain predicted log-odds of ACT. While these log-odds are unlikely to be well-calibrated due to differences in the training and test datasets, they may still accurately represent relative risk of ACT across our 45 individuals. Indeed, the log-odds correlated significantly with the troponin measurement slope (Spearman correlation $p = 0.01$, Figure 4f), suggesting our troponin measurements, and by extension our expression response data, recapitulate *in vivo* cellular response to doxorubicin.

Cardiomyocyte molecular QTLs show enrichment in ACT GWAS

To determine the disease-relevance of our molecular QTLs we obtained summary statistics for the largest ACT GWAS to date[5]. While this GWAS was not sufficiently powered to find genome-wide significant associations, 11 variants representing 9 independent loci have $p < 10^{-5}$, with the most significant (rs2184559) at $p = 2.8 \times 10^{-6}$. Of the 8 GWAS variants with $p < 10^{-5}$ either tested in our eQTL mapping, or in high LD ($R^2 > 0.8$) with a tested SNP, 7 have a nominally significant marginal eQTL ($p < 0.05$, the 8th has $p = 0.07$) and four have a reQTL with $p < 0.1$. The one replicated variant in this GWAS, rs28714259, was

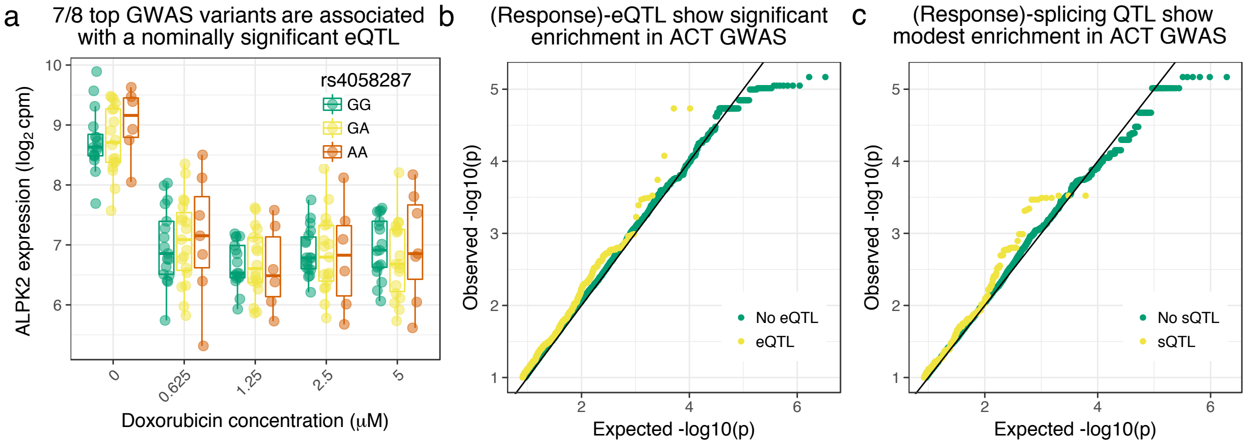


Figure 5: *Cardiomyocyte molecular QTLs are enriched in the largest available ACT GWAS*[5]*.* ***a.*** *rs4058287 has a GWAS p-value of* 9.68×10^{-6} *and is a nominally significant eQTL (*$p = 0.0016$*) for* ALPK2*, which is down-regulated in response to doxorubicin.* ***b.*** *SNPs that have either a marginal or response eQTL with* $p < 10^{-5}$ *are enriched in GWAS variants with* $p < 0.05$ *(hypergeometric test* $p = 3 \times 10^{-6}$*).* ***c.*** *SNPs with a marginal or response splicing QTL at* $p < 10^{-5}$ *show modest enrichment in GWAS* $p < 0.005$ *(hypergeometric* $p = 0.02$*).*

not genotyped in our data but is in high LD ($R^2 = 0.98$) with rs11855704 which is a nominally significant marginal eQTL for tubulin gamma complex associated protein 5 (*TUBGCP5*, Supplementary Figure S6). rs4058287 (GWAS p-value 9.68×10^{-6}) has a marginal effect on Alpha-Protein Kinase 2 (*ALPK2*, also known as "Heart Alpha-Protein Kinase" since it was discovered in mouse heart[26] and is expressed in few other tissues[27]) expression ($p = 0.0016$) as well as a weak interaction effect ($p = 0.06$, see Figure 5a). Interestingly, *ALPK2* has been shown to upregulate DNA repair genes and to enable caspase-3 cleavage and apoptosis in a colorectal cancer model[28]. The replicating variant from Aminkeng et al.[4], rs2229774 only occurs in two individuals in our cohort (who are heterozygous) making eQTL mapping infeasible. Additionally we find a marginal effect eQTL ($p = 0.0017$, Supplementary Figure S7) on SLC28A3 for rs885004, which has previously been associated with ACT in a candidate gene study[29]. rs885004 is intronic, falls in DNase I hypersensitivity and H3K27ac peaks and is in LD ($R^2 = 0.98$) with another ACT implicated variant, rs7853758[30].

Since the GWAS was overall underpowered, we additionally assessed whether there was detectable enrichment of low GWAS p-values for our regulatory QTLs. We considered SNPs with either a marginal or response eQTL with nominal $p < 10^{-5}$ (corresponding approximately to 5% FDR) and found a significant enrichment for GWAS p-values under 0.05 (one-sided hypergeometric $p = 3 \times 10^{-6}$, OR=1.2, Figure 5b). We observed a more modest enrichment for splicing QTLs: using the same criteria to define a set of sQTL variants we observe significant enrichment in GWAS p-values under 0.005 (one-sided hypergeometric $p = 0.02$, 5c) but not under 0.05 (one-sided hypergeometric $p = 0.7$).

Discussion

Human iPSC-derived somatic cells provide a powerful, renewable and reproducible tool for modeling cellular responses to external perturbation *in vitro*, especially for non-blood cell-types such as cardiomyocytes which are extremely challenging to collect and even then are typically only available post-mortem. We established a sufficiently large iPSC panel to effectively query the transcriptomic response of differentiated cardiomyocytes to doxorubicin. We were also able to characterize the role of genetic variation in modulating this response, both in terms of total expression and alternative splicing. There are, of course, caveats associated with using an *in vitro* system, which may not accurately represent certain aspects cardiac response to doxorubicin *in vivo*. That said, the replication of GTEx heart eQTLs, association of troponin levels with predicted ACT-risk[12], and the observed GWAS enrichment, all support the notion that the IC system recapitulates substantial elements of *in vivo* biology. It is challenging to quantify this agreement however, and there are *in vivo* factors that are certainly not represented: in particular, excessive fibrosis plays a role in ACT[31,32,33], although it is unclear whether fibroblasts are directly activated by doxorubicin exposure or simply respond indirectly to cardiomyocyte damage.

For many diseases such as ACT which involve an environmental perturbation it is reasonable to suppose that eQTLs detected at steady-state are only tangentially relevant when attempting to interpret disease variants. Such concerns motivated us to focus on response eQTLs, i.e. variants that that have functional consequences under specific cellular conditions because they interact, directly or indirectly, with the treatment. We used a statistical definition of reQTLs corresponding to cases where gene expression levels are significantly better explained using a model including an interaction term between genotype and treatment (represented as a categorical variable), compared to a model with only additive effects for genotype and treatment. Our characterization of the detected reQTL demonstrates that these variants are indeed candidate drivers of differences in individual transcriptomic response to doxorubicin. The strongest reQTL effects correspond to completely different response patterns for the major and minor alleles, while weaker effects correspond to more subtle modulation of the same response pattern. We note that it is not necessarily the case that such reQTLs are the only functionally relevant eQTLs. eSNPs with a marginal (additive) effect on expression of a gene responsive to doxorubicin (as most genes are) could still be important if the relationship between expression and ACT-risk is nonlinear, for example involving thresholding effects.

We observed a statistical enrichment of expression and (to a lesser extent) splicing QTLs in ACT GWAS.

However, with no genome-wide significant associations available, fine-mapping of causal variants remains fraught. We anticipate our findings will be increasingly valuable as larger-scale ACT GWAS become available.

We derived ICs from healthy individuals so we do not known which individuals would develop ACT if they required anthracycline treatment. Mapping molecular response QTLs in larger panels of ICs from patients treated with anthracyclines who do or do not develop ACT symptoms would allow stronger conclusions to be drawn about the contribution of the detected (r)eQTLs to disease etiology.

Finally, an interesting observation in our study is that splicing fidelity is reduced upon doxorubicin exposure. This is not completely unexpected since a key downstream side-effect of doxorubicin is increased oxidative stress, which has been previously associated with dysregulated splicing of specific genes[19,20]. Our finding that this effect is prevalent across the transcriptome poses further questions about what known effects of doxorubicin might, in fact, be mediated by changes in splicing.

Methods

Sample collection and genotyping

Generation of lymphoblastoid cell lines (LCLs) and genome-wide genotyping of many individuals from a multi-generational pedigree were performed previously. Briefly, lymphocytes were isolated from whole blood samples using Ficoll-Paque and immortalized using Epstein Barr Virus[34,35]. Phased genotypes were obtained by combining pedigree information, genotypes from SNP arrays, and genotypes from whole genome sequencing of related individuals[17].

iPSC reprogramming and cardiomyocyte differentiation

We reprogrammed the 45 LCLs to iPSCs using episomal plasmid vectors, containing *OCT3/4*, *p53* shRNA, *SOX2*, *KLF4*, *L-MYC*, and *LIN28* which avoids integrating additional transgenes[36]. Initially, the lines were generated on mouse embryonic fibroblasts (MEF), which coated the well and served as feeder cells to create an environment supportive of pluripotent stem cells. The colony was then mechanically passaged on MEF and tested for expression of pluripotency-associated markers by immunofluorescence staining and RT-PCR. The lines were passaged for at least 10 weeks on MEF to ensure lines had stabilized. We characterized the iPSC lines using the embryoid body assay, karyotyping, and the PluriTest[37] classifier. iPSC lines were then transitioned to feeder-free conditions, which was necessary to prime the iPSCs for differentiation. Next we differentiated the iPSCs to cardiomyocytes[38,39]. iPSC lines were covered with a 1:60 dilution matrigel overlay for 24 hours. On day 0 iPSC lines were treated with 12μM of the *GSK3* inhibitor, CHIR99021, in RPMI+B27 medium (RPMI1640, 2nM L-glutamine and 1x B27 supplement minus insulin) for 24 hours at which time media was replaced with fresh RPMI+B27. 72 hours after the addition of CHIR99021 (Day 3), 2μM of the Wnt inhibitor Wnt C-59 was added for 48 hours. Fresh RPMI+B27 was added on Days 5, 7 and 10. Beating cells appeared between Days 8-10. These cardiomyocytes consisted of ventricular, atrial and pacemaker-like cells. The cells formed thick layers and contract throughout the well. Metabolic selection was used to purify the cardiomyocytes[40] from Day 14 to Day 20 when glucose-free RPMI media supplemented with the components essential for cardiomyocyte differentiation[39], ascorbic acid and human serum albumin, together with lactate, a substrate uniquely metabolized by cardiomyocytes, was added to cells. Because this lactate media can only be metabolized by cardiomyocytes, the non-cardiomyocytes in the culture were removed over the 6 day treatment. On day 20 the cardiomyocytes, now at a high cTnT purity, were replated for experiments in media that contains only galactose and fatty acids as an energy source. This galactose media forces the cardiomyocytes to undergo aerobic respiration, rather than anaerobic glycolysis common in cultured cells.

Doxorubicin exposure

We incubated the cardiomyocytes in 0, 0.625, 1.25, 2.5, or 5 μM doxorubicin. After 24 hours, we collected the serum and cells from each condition. From the serum, we measured cardiac Troponin T levels using the ABNOVA Troponin I (Human) ELISA kit (cat. no. KA0233). From the cells, we extracted RNA for sequencing. Each treatment batch contained 1 to 4 individuals. RNA quality was assessed with the Agilent Bioanalyzer.

RNA-sequencing

We prepared libraries using the Illumina TruSeq Library Kit and generated 50bp single-end reads on a HiSeq 4000 at the University of Chicago Functional Genomics Facility. We confirmed sequencing quality using FastQC and MultiQC[41]. We confirmed sample identity by 1) comparing allelic counts (quantified using `samtools mpileup`[42]) of exonic SNPs to the known genotypes and 2) running verifyBamID[43].

Expression quantification

We aligned RNA-seq reads using STAR version 2.5.2a[44] to GRCh38/GENCODE release 24. We counted reads using featureCounts[45] and calculated counts per million reads (cpm) using ‘cpm‘ from the ‘edgeR‘ R package (version 3.18.1)[46]. We discarded samples with $< 10^9$ reads and genes with median $\log_2(cpm)$ less than 0.

Differential expression analysis

We performed differential expression (DE) analysis across all 5 doxorubicin concentrations jointly, using either a linear model on quantile normalized cpm value or Spearman correlation, followed by Benjamini-Hochberg False Discovery Rate (FDR) control. Since the vast majority of genes showed differential expression we did not investigate better powered DE methods such as DESeq2.

We clustered genes into "response patterns" using a K-component mixture model

$$\begin{aligned} \pi &\sim \text{Dir}(1/K, \cdots, 1/K) \\ z_g|\pi &\sim \text{Discrete}(\pi) \\ y_{ngc}|z_g = k, \theta &\sim N(\theta_{ck}, \sigma^2) \end{aligned} \tag{1}$$

where π is a prior probability vector over cluster assignments, Dir is the Dirichlet distribution, z_g is cluster from which gene g is generated, y_{ngc} is the expression of gene g in individual n at concentration c, θ_{ck} is the mixture parameter (mean) across concentrations for cluster k, and σ^2 is a shared noise variance. We marginalize (sum) over z_g and optimize with respect to π, θ, σ using the `rstan` R package[47] (version 2.16.2). The hyperparameters of the Dirichlet distribution are set such that in the limit of large K the model approximates a Dirichlet process mixture[48] which automatically learns of an appropriate number of mixture components to use from data.

Gene set and promoter motif enrichment were performed using HOMER v4.9.1[49] using default parameters and without *de novo* motif search.

Response eQTL mapping

We developed an extension of the PANAMA[18] linear mixed model (LMM) framework to map eQTLs and response eQTLs while accounting for latent confounding, which we call `suez`. `suez` entails a two step procedure. Step one is used to learn latent factors from all genes, using the model

$$\begin{aligned} y_{ncg} &= \sum_k W_{kg} x_{nck} + u_{ng} + v_{cg} + \xi_{ncg} + \epsilon_{ncg} \\ W_{kg} &\sim N(0, \sigma_k^2) \qquad \textit{factor loadings/coefficients} \\ u_{ng} &\sim N(0, \sigma_u^2) \qquad \textit{individual random effects} \\ \xi &\sim MVN(0, \sigma_\xi^2 \Sigma) \qquad \textit{kinship random effect} \\ \epsilon &\sim MVN(0, \mathrm{diag}(\sigma_\epsilon^2)) \qquad \textit{noise} \end{aligned}$$

where x_{nck} are latent factors, v_{cg} are per gene, per concentration fixed effects. We integrate over W, u, ξ and ϵ, which results in a per gene multivariate normal,

$$y_{:g} \sim MVN\left(V v_{:g}, \sum_k \sigma_k^2 x_{:k} x_{k:}^T + \sigma_u^2 U + \sigma_\xi^2 \Sigma + \sigma_e^2 I \right), \tag{2}$$

where $y_{:g}$ refers to the vector of expression for gene g across all individuals and concentrations (i.e. all "samples" where a sample is an individual-concentration pair), V is a matrix mapping concentrations to samples (i.e. $V_{sc} = 1$ iff sample s is at concentration c) and U is a matrix of which samples are for the same individual (i.e. $U_{ss'} = 1$ if sample s and sample s' come from the same individual). We optimize x, v and the variances $\{\sigma_u^2, \sigma_k^2, \sigma_\xi^2, \sigma_\epsilon^2\}$ jointly across all genes g.

In step 2 we test individual gene-SNP pairs while accounting for confounding using the covariance matrix

$$\Sigma_\pi = \sum_k \sigma_k^2 x_{:k} x_{k:}^T + \sigma_u^2 U + \sigma_\xi^2 \Sigma \tag{3}$$

which includes both latent confounding, individual random effects and similarity due to kinship. We consider three LMMs, all with the same parameterization of the covariance $\sigma_\pi^2 \Sigma_\pi + \sigma_e^2 I$ where σ_π^2 and σ_e^2 are optimized along with the fixed effects to allow the extent to which each gene follows the global covariance pattern to be adapted. The simple structure of this covariance also allows pre-computation of the eigen-decomposition of Σ_π which enables linear (rather than cubic) time evaluation of the likelihood and its gradient.

Model 0 involves no effect of the SNP (and can therefore be fit once for a gene), a fixed effect for concentration. Model 1 adds a marginal effect of the SNP genotype dosage d. Finally model 2 adds an interaction effect between concentration and genotype, which is equivalent to a concentration-specific genotype effect.

In summary:

$$\text{Model 0: } \mathbb{E}[y_{ncg}] = v_{cg} \tag{4}$$

$$\text{Model 1: } \mathbb{E}[y_{ncg}] = v_{cg} + \beta d_n \tag{5}$$

$$\text{Model 2: } \mathbb{E}[y_{ncg}] = v_{cg} + \beta_c d_n \tag{6}$$

We optimize σ_π^2, σ_e^2 and the regression coefficients for each of the three models separately, and use likelihood ratio tests (LRT) to compare the models. Comparing Model 1 vs 0 (one degree of freedom) tests whether there is a marginal effect of the variant. Comparing Model 2 vs 1 ($C - 1 = 4$ degrees of freedom, where C is the number of conditions/concentrations) tests whether there is an interaction effect, i.e. whether the genetic effect on expression is different at different concentrations (or equivalently whether the response to doxorubicin is different for different genotypes). Finally Model 2 vs 0 ($C = 5$ degrees of freedom) tests whether there is any effect of genotype on expression, either in terms of a marginal or concentration-specific effect. We use the conservative approach of using Bonferroni correction across SNPs for a gene, followed by Benjamini-Hochberg FDR control.

We quantile normalize the expression levels across all samples for each gene to a standard normal distribution so that the distributional assumptions of our linear mixed model are reasonable. However, optimizing the variance parameters σ_π^2 and σ_e^2 means that the χ^2 distribution for the LRT will only hold asymptotically and p-values for finite sample sizes will tend to be somewhat anti-conservative. To account for this for response-eQTLs, we use a parametric bootstrap since there is no fully valid permutation strategy for testing interaction effects. This involves first fitting Model 1 and then simulating new expression data under the fitted model. Models 1 and 2 are then (re)fit to this data and compared using an LRT. We then perform Bonferroni correction across SNPs for each gene to obtain an empirical null distribution of per gene p-values which we use to estimate the true FDR for our response-eQTL results.

For significant reQTLs we assigned the response of the minor allele and major allele to the previously determined clusters using the model

$$y_{nc} | z_A, z_a, \theta \sim N\left(\frac{1}{2} d_n \theta_{cz_A} + \frac{1}{2}(2 - d_n)\theta_{cz_a}, \sigma^2\right),$$

where y_{nc} is the expression for individual n at concentration c, z_A and z_a are the cluster assignments for the major and minor allele respectively, $d_n \in \{0, 1, 2\}$ is the genotype dosage, and θ and σ^2 are fixed at the values learned in Equation 1. For each reQTL separately we calculate the likelihood of y given all possible pairs of assignments (z_A, z_a) and choose the maximum likelihood solution.

As for all k-means clustering in the paper, we used `KMeans_rcpp` function of the R package `ClusterR` v1.0.6, taking the best of 10 initializations using the `k-means++` option, to cluster the normalized genotype effect profiles of the significant associations. The choice of 9 clusters was determined manually.

Splicing analysis

We ran `LeafCutter` v0.2.6_dev (using default settings) which allows joint differential intron excision testing across more than two conditions. For each Alternative Splicing Cluster (ASC) `LeafCutter` fits a set of $PercentSplicedIn$ probability vectors ψ_c, across detected splice junctions i, at each concentration c. For ASCs determined to be significantly (5% FDR) differential spliced across concentrations, we calculated the entropy $h_c = -\sum_i \psi_{ci} \log \psi_{ci}$ at each concentration c. We normalized these profiles as $\tilde{h_c} = h_c / \bar{h_c}$ and clustered these profiles, using `KMeans_rcpp` as above.

To investigate the relative usage of cryptic splice sites we first determined the set of 7792 splice junctions that a) fell in ASCs determined to be significantly differentially spliced (5% FDR) and b) had $\max_c \psi_{ci} - \min_c \psi_{ci} > 0.1$. We obtained normalized intron excision rates by subtracting the per intron mean and dividing by the per intron standard deviation. These ψ profiles were clustered using `KMeans_rcpp`. Cryptic splice site usage was determined by considering all exons in Gencode v26 and ignoring transcript structure (i.e. a junction spanning two splice sites used but only in different transcripts would still be considered "annotated").

For (response) splicing QTL we calculated within ASC intron excision ψ with pseudocount of 0.5, and set entries with 0 denominator (no reads for that ASC in that sample) to the mean across all other samples. These values were then 1) z-score normalized across samples and 2) quantile normalized to a normal across introns. QTL mapping was then performed using `suez` considering each intron as a "gene".

Modeling cardiac troponin level

We assessed the proportion of variance in cardiac troponin explained by gene expression response. Let y_{ci} represent the troponin level measured in individual i at doxorubicin concentration c, normalized to have 0 mean and variance 1 across individuals at each concentration. Let x_{cig} be the expression of gene g (in individual i at concentration c), z-score normalized across samples. We consider the linear model

$$y_{ci} = \sum_g \beta_g x_{cig} + \epsilon_{ci} \tag{7}$$

where $\epsilon_{ci} \sim N(0, \sigma_\epsilon^2)$ is noise and the coefficients β_g are given a prior $N(0, \sigma_\beta^2/G)$ where $G = 12,317$ is the number of genes in the analysis. Integrating over β_g we have

$$y_{:} \sim N\left(0, \sigma_\beta^2 \frac{1}{G} \sum_g x_{:g} x_{:g}^T + \sigma_\epsilon^2 I\right) \tag{8}$$

We optimize this model wrt σ_β and σ_ϵ to obtain an estimate $\sigma_\beta^2/(\sigma_\beta^2 + \sigma_\epsilon^2)$ of the percent variance of y explained by x.

Code and data availability

All the custom analysis scripts used for this project are available at `https://github.com/davidaknowles/dox`. The `suez` response eQTL mapping R package is available at `https://github.com/davidaknowles/suez`. The following data are available as Supplementary Data: 1) differential expression cluster assignments, 2) significant (5% FDR) eQTLs and sQTLs, 3) differential splicing results, 4) levels of cardiac troponin and the predicted transcriptomic response. In addition to the Supplementary Data included with this paper full results are hosted at `http://web.stanford.edu/~dak33/dox/` including 1) gene-by-sample matrix of RNA-seq quantification (log counts per million), 2) LeafCutter intron excision quantification 3) p-values for all tested eQTLs, reQTLs, sQTLs, and rsQTLs. The RNA-seq FASTQ files will be added to the dbGaP database[50] under dbGaP accession phs000185 (`https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000185`). The genotype data files cannot be shared because releasing genotype data from a subset of individuals in the pedigree would enable the reconstruction of genotypes of other members of the pedigree, which would violate the original protocol approved by the research ethics board[17]. The summary statistics for the ACT GWAS were given to us by the authors of the study[5].

Acknowledgements

The idea for the study was developed through conversations with Julian Solway (University of Chicago). We gratefully acknowledge Bryan Schneider for providing GWAS summary statistics. This work was supported by the Howard Hughes Medical Institute, and the US National Institutes of Health (NIH grants

HL092206, HG008140, HG009431). CKB was supported by a NIH (http://www.ncats.nih.gov/ctsa) Clinical and Translational Science Award 5 TL1 TR 432-7 pre-doctoral fellowship.

Author Contributions

YG, CKB and CO conceived the project and provided source material. CKB performed the experiments with assistance from KMP. DAK developed algorithms, and analyzed the data with assistance from JDB. JKP, CO and YG supervised the project. DAK wrote the paper with input from all authors.

References

[1] Von Hoff, D. D. *et al.* Risk factors for doxorubicin-induced congestive heart failure. *Ann Intern Med* **91**, 710–717 (1979).

[2] Swain, S. M., Whaley, F. S. & Ewer, M. S. Congestive heart failure in patients treated with doxorubicin. *Cancer* **97**, 2869–2879 (2003).

[3] Vejpongsa, P. & Yeh, E. T. Prevention of anthracycline-induced cardiotoxicity: challenges and opportunities. *Journal of the American College of Cardiology* **64**, 938–945 (2014).

[4] Aminkeng, F. *et al.* A coding variant in RARG confers susceptibility to anthracycline-induced cardiotoxicity in childhood cancer. *Nature Genetics* **47**, 1079–1084 (2015).

[5] Schneider, B. P. *et al.* Genome-Wide Association Study for Anthracycline-Induced Congestive Heart Failure. *Clinical Cancer Research* (2016).

[6] Serie, D. J. *et al.* Genome-wide association study of cardiotoxicity in the NCCTG N9831 (Alliance) adjuvant trastuzumab trial **9831**, 1–8 (2017).

[7] Thomas, S. M. *et al.* Reprogramming LCLs to iPSCs Results in Recovery of Donor-Specific Gene Expression Signature. *PLoS genetics* **11**, e1005216 (2015).

[8] Burrows, C. K. *et al.* Genetic Variation, Not Cell Type of Origin, Underlies the Majority of Identifiable Regulatory Differences in iPSCs. *PLoS Genetics* **12**, 1–18 (2016).

[9] Banovich, N. E. *et al.* Impact of regulatory variation across human ipscs and differentiated cells. *bioRxiv* (2016). arXiv:`https://www.biorxiv.org/content/early/2016/12/05/091660.full.pdf`.

[10] Kilpinen, H. *et al.* Common genetic variation drives molecular heterogeneity in human iPSCs. *Nature* **546**, 370–375 (2017). arXiv:`NIHMS150003`.

[11] Alasoo, K. *et al.* Shared genetic effects on chromatin and gene expression reveal widespread enhancer priming in immune response. *bioRxiv* (2017). arXiv:`https://www.biorxiv.org/content/early/2017/05/18/102392.full.pdf`.

[12] Burridge, P. W. *et al.* Human induced pluripotent stem cell-derived cardiomyocytes recapitulate the predilection of breast cancer patients to doxorubicin-induced cardiotoxicity.-supple. *Nature medicine* **22**, 547–56 (2016). arXiv:`15334406`.

[13] Graessmann, M., Berg, B., Fuchs, B., Klein, A. & Graessmann, A. Chemotherapy resistance of mouse wap-svt/t breast cancer cells is mediated by osteopontin, inhibiting apoptosis downstream of caspase-3. *Oncogene* **26**, 2840 (2007).

[14] Lu, W. *et al.* Znf143 transcription factor mediates cell survival through upregulation of the gpx1 activity in the mitochondrial respiratory dysfunction. *Cell death & disease* **3**, e422 (2012).

[15] Poché, R. A. *et al.* RONIN is an essential transcriptional regulator of genes required for mitochondrial function in the developing retina. *Cell Rep.* **14**, 1684–1697 (2016).

[16] Zarubin, T. & Han, J. Activation and signaling of the p38 map kinase pathway. *Cell research* **15**, 11 (2005).

[17] Livne, O. E. *et al.* PRIMAL: Fast and accurate pedigree-based imputation from sequence data in a founder population. *PLoS computational biology* **11**, e1004139 (2015).

[18] Fusi, N., Stegle, O. & Lawrence, N. D. Joint modelling of confounding factors and prominent genetic regulators provides increased accuracy in genetical genomics studies. *PLoS Computational Biology* **8** (2012).

[19] Disher, K. & Skandalis, A. Evidence of the modulation of mrna splicing fidelity in humans by oxidative stress and p53. *Genome* **50**, 946–953 (2007).

[20] Seo, J. *et al.* Oxidative stress triggers body-wide skipping of multiple exons of the spinal muscular atrophy gene. *PloS one* **11**, e0154390 (2016).

[21] Li, Y. I. *et al.* LeafCutter: Annotation-free quantification of RNA splicing. *Nature Genetics (to appear)* 044107 (2017).

[22] Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).

[23] Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* 267–288 (1996).

[24] Park, J. E. *et al.* Positive regulation of apoptosis signal-regulating kinase 1 by dual-specificity phosphatase 13a. *Cellular and molecular life sciences* **67**, 2619–2629 (2010).

[25] Hastie, T. & Stuetzle, W. Principal curves. *Journal of the American Statistical Association* **84**, 502–516 (1989).

[26] Ryazanov, A. G., Pavur, K. S. & Dorovkov, M. V. Alpha-kinases: a new class of protein kinases with a novel catalytic domain. *Current Biology* **9**, R43–R45 (1999).

[27] Melé, M. *et al.* The human transcriptome across tissues and individuals. *Science* **348**, 660–665 (2015).

[28] Yoshida, Y. *et al.* Alpk2 is crucial for luminal apoptosis and dna repair-related gene expression in a three-dimensional colonic-crypt model. *Anticancer research* **32**, 2301–2308 (2012).

[29] Visscher, H. *et al.* Validation of variants in SLC28A3 and UGT1A6 as genetic markers predictive of anthracycline-induced cardiotoxicity in children. *Pediatric & Blood Cancer* **60**, 1375–1381 (2013).

[30] Visscher, H. *et al.* Pharmacogenomic prediction of anthracycline-induced cardiotoxicity in children. *J. Clin. Oncol.* **30**, 1422–1428 (2012).

[31] Cascales, A. *et al.* Association of anthracycline-related cardiac histological lesions with nadph oxidase functional polymorphisms. *The Oncologist* **18**, 446–453 (2013).

[32] Zhan, H. *et al.* Ataxia telangiectasia mutated in cardiac fibroblasts regulates doxorubicin-induced cardiotoxicity. *Cardiovascular Research* **110**, 85–95 (2016).

[33] Farhad, H. *et al.* Characterization of the changes in cardiac structure and function in mice treated with anthracyclines using serial cardiac magnetic resonance imagingclinical perspective. *Circulation: Cardiovascular Imaging* **9**, e003584 (2016).

[34] Cusanovich, D. a. *et al.* The combination of a genome-wide association study of lymphocyte count and analysis of gene expression data reveals novel asthma candidate genes. *Human molecular genetics* **21**, 2111–23 (2012).

[35] Cusanovich, D. a. *et al.* Integrated analyses of gene expression and genetic association studies in a founder population. *Human molecular genetics* **25**, 2104–2112 (2016).

[36] Okita, K. *et al.* A more efficient method to generate integration-free human iPS cells. *Nature Methods* **8**, 409–412 (2011).

[37] Müller, F.-J. *et al.* A bioinformatic assay for pluripotency in human cells. *Nature Methods* **8**, 315–317 (2011).

[38] Lian, X. *et al.* Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/β-catenin signaling under fully defined conditions. *Nature Protoc.* **8**, 162–175 (2013).

[39] Burridge, P. W. *et al.* Chemically defined generation of human cardiomyocytes. *Nat. Methods* **11**, 855–860 (2014).

[40] Tohyama, S. *et al.* Distinct metabolic flow enables large-scale purification of mouse and human pluripotent stem cell-derived cardiomyocytes. *Cell Stem Cell* **12**, 127–137 (2013).

[41] Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics (Oxford, England)* btw354 (2016).

[42] Li, H. *et al.* The sequence alignment/map format and samtools. *Bioinformatics* **25**, 2078–2079 (2009).

[43] Jun, G. *et al.* Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *American journal of human genetics* **91**, 839–48 (2012).

[44] Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

[45] Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)* **30**, 923–30 (2014).

[46] Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

[47] Carpenter, B. *et al.* Stan: A probabilistic programming language. *Journal of Statistical Software* **20**, 1–37 (2016).

[48] MacEachern, S. N. & Müller, P. Estimating mixture of dirichlet process models. *Journal of Computational and Graphical Statistics* **7**, 223–238 (1998).

[49] Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. *Molecular cell* **38**, 576–589 (2010).

[50] Tryka, K. A. *et al.* NCBI's database of genotypes and phenotypes: dbGaP. *Nucleic acids research* **42**, D975–9 (2014).

Supplementary material

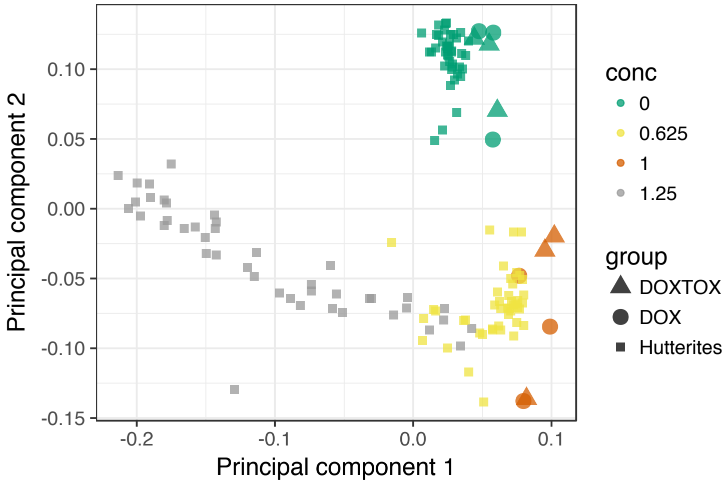
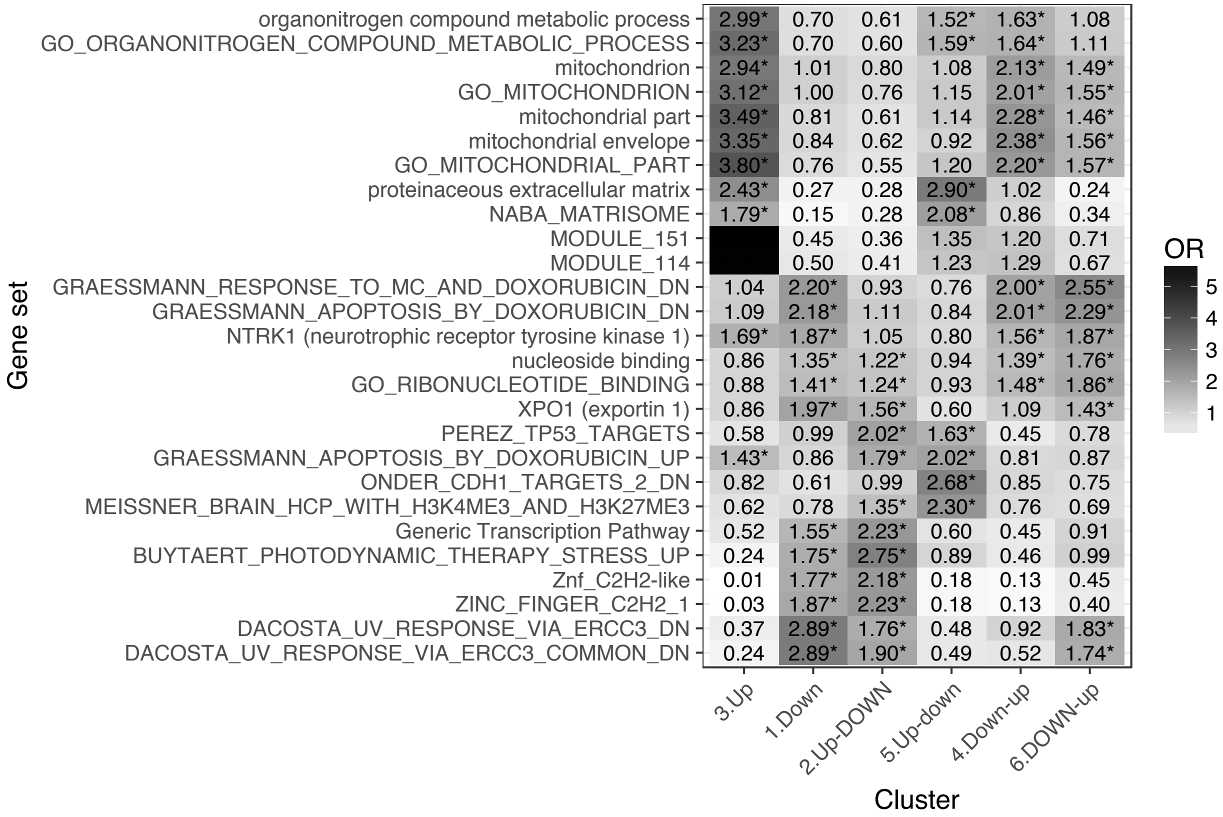


Figure S1: *Our expression data is concordant with an existing small RNA-seq dataset*[12]. *DOXTOX and DOX correspond to samples from patients that did and did not develop ACT after doxorubicin chemotherapy respectively. We additionally see that the transcriptional response at higher concentrations cannot be extrapolated from that at lower concentrations. Higher concentrations are not shown since including these compresses the first PC obscuring the relevant variation.*



Gene set	3.Up	1.Down	2.Up-DOWN	5.Up-down	4.Down-up	6.DOWN-up
organonitrogen compound metabolic process	2.99*	0.70	0.61	1.52*	1.63*	1.08
GO_ORGANONITROGEN_COMPOUND_METABOLIC_PROCESS	3.23*	0.70	0.60	1.59*	1.64*	1.11
mitochondrion	2.94*	1.01	0.80	1.08	2.13*	1.49*
GO_MITOCHONDRION	3.12*	1.00	0.76	1.15	2.01*	1.55*
mitochondrial part	3.49*	0.81	0.61	1.14	2.28*	1.46*
mitochondrial envelope	3.35*	0.84	0.62	0.92	2.38*	1.56*
GO_MITOCHONDRIAL_PART	3.80*	0.76	0.55	1.20	2.20*	1.57*
proteinaceous extracellular matrix	2.43*	0.27	0.28	2.90*	1.02	0.24
NABA_MATRISOME	1.79*	0.15	0.28	2.08*	0.86	0.34
MODULE_151		0.45	0.36	1.35	1.20	0.71
MODULE_114		0.50	0.41	1.23	1.29	0.67
GRAESSMANN_RESPONSE_TO_MC_AND_DOXORUBICIN_DN	1.04	2.20*	0.93	0.76	2.00*	2.55*
GRAESSMANN_APOPTOSIS_BY_DOXORUBICIN_DN	1.09	2.18*	1.11	0.84	2.01*	2.29*
NTRK1 (neurotrophic receptor tyrosine kinase 1)	1.69*	1.87*	1.05	0.80	1.56*	1.87*
nucleoside binding	0.86	1.35*	1.22*	0.94	1.39*	1.76*
GO_RIBONUCLEOTIDE_BINDING	0.88	1.41*	1.24*	0.93	1.48*	1.86*
XPO1 (exportin 1)	0.86	1.97*	1.56*	0.60	1.09	1.43*
PEREZ_TP53_TARGETS	0.58	0.99	2.02*	1.63*	0.45	0.78
GRAESSMANN_APOPTOSIS_BY_DOXORUBICIN_UP	1.43*	0.86	1.79*	2.02*	0.81	0.87
ONDER_CDH1_TARGETS_2_DN	0.82	0.61	0.99	2.68*	0.85	0.75
MEISSNER_BRAIN_HCP_WITH_H3K4ME3_AND_H3K27ME3	0.62	0.78	1.35*	2.30*	0.76	0.69
Generic Transcription Pathway	0.52	1.55*	2.23*	0.60	0.45	0.91
BUYTAERT_PHOTODYNAMIC_THERAPY_STRESS_UP	0.24	1.75*	2.75*	0.89	0.46	0.99
Znf_C2H2-like	0.01	1.77*	2.18*	0.18	0.13	0.45
ZINC_FINGER_C2H2_1	0.03	1.87*	2.23*	0.18	0.13	0.40
DACOSTA_UV_RESPONSE_VIA_ERCC3_DN	0.37	2.89*	1.76*	0.48	0.92	1.83*
DACOSTA_UV_RESPONSE_VIA_ERCC3_COMMON_DN	0.24	2.89*	1.90*	0.49	0.52	1.74*

Figure S2: *Gene set enrichment analysis of genes in each response cluster confirms expected patterns: metabolic, mitochrondrial and DNA damage processes, as well as existing doxorubicin response genes.*

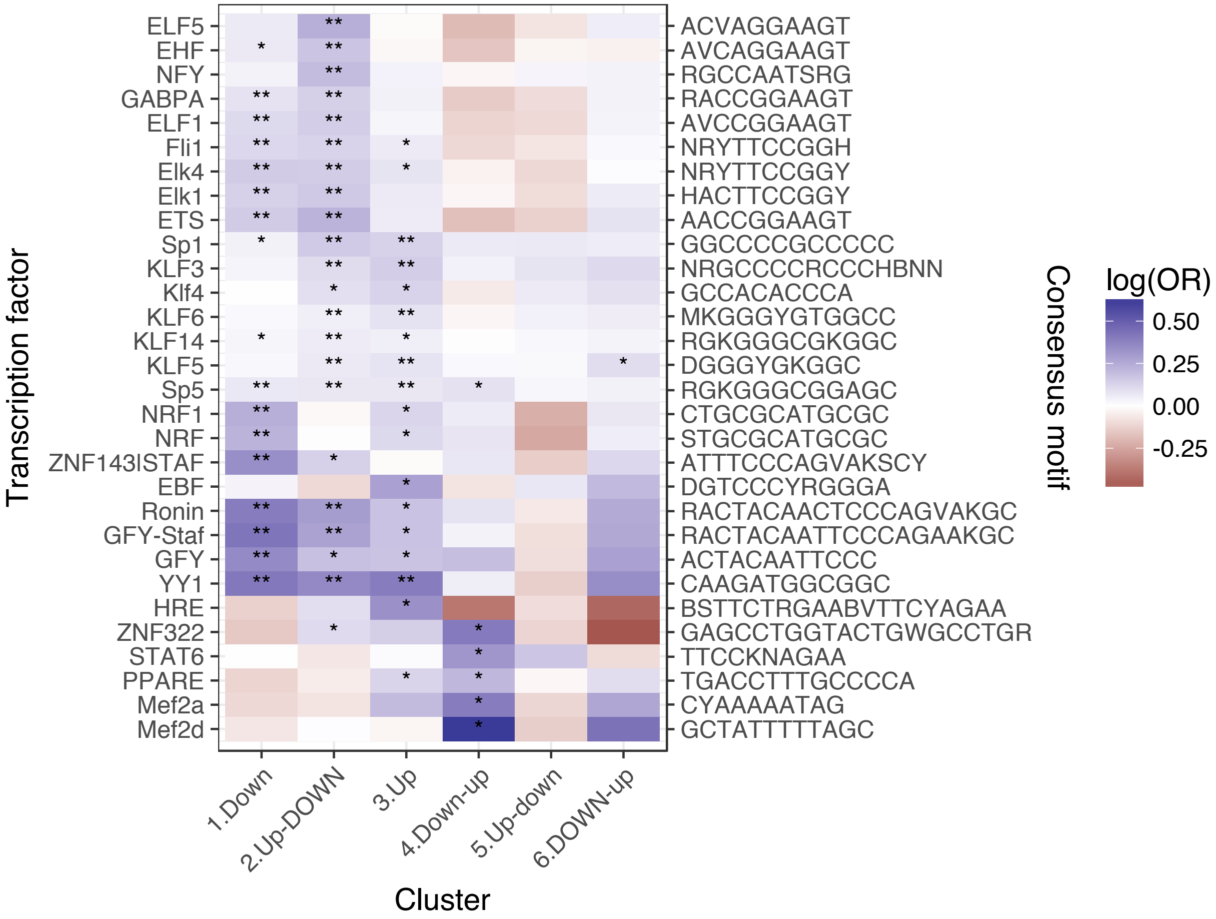


Figure S3: *Enrichment of transcription factor binding motifs for each response pattern, using HOMER. ** denotes $q < 0.05$, * denotes $q < 0.5$.*

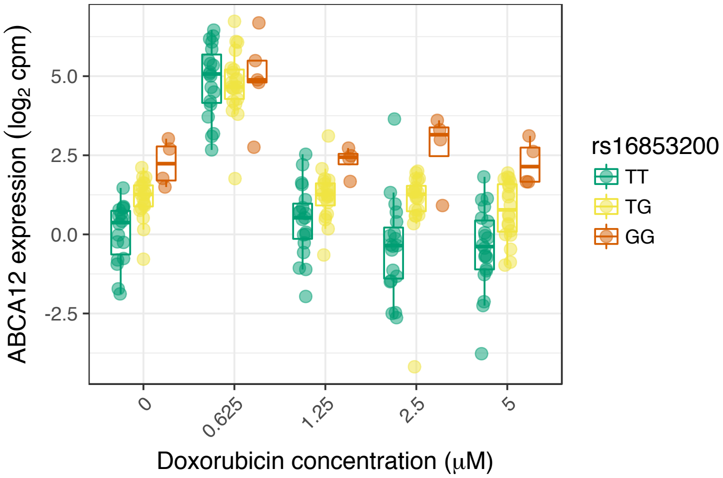


Figure S4: *An example response expression QTL that may act through buffering at high expression levels.*

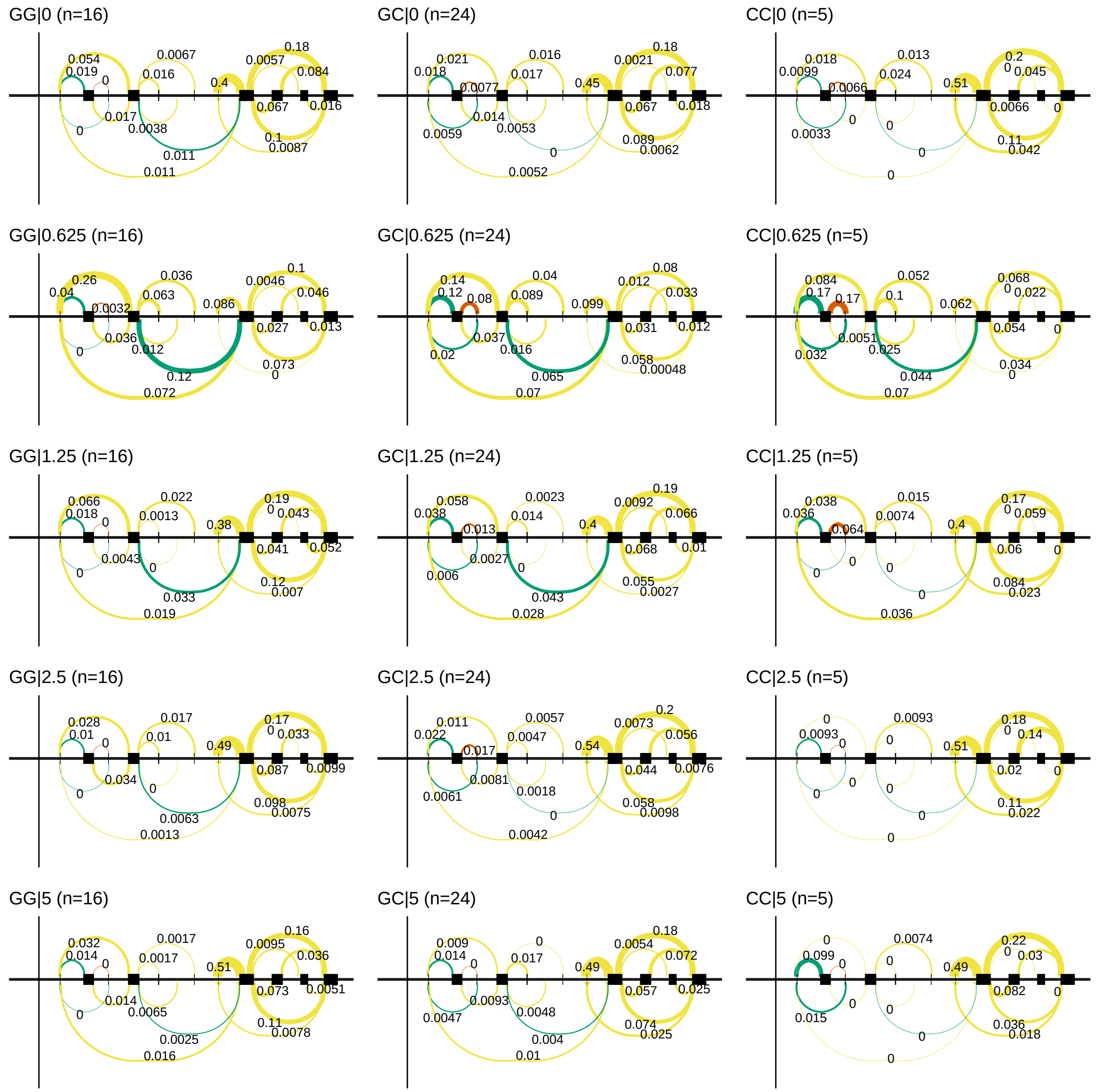


Figure S5: *Doxorubicin exposure results in the use of a downstream alternative TSS which uncovers an association between rs896853 genotype and inclusion of the exon at chr8:94975318-94975415.*

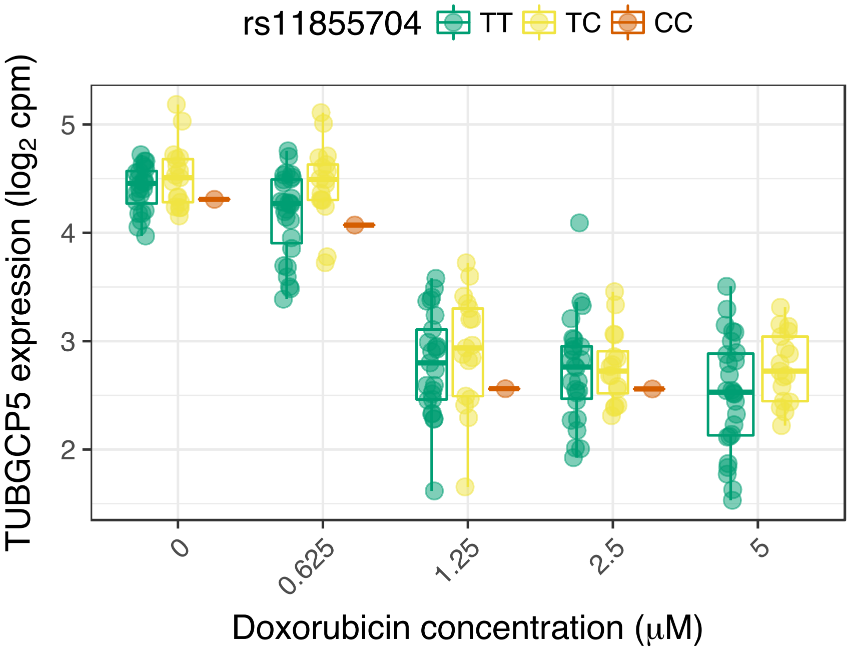


Figure S6: *rs28714259, the one replicated variant from Schneider et al.*[5] *is in high LD* ($R^2 = 0.98$) *with rs11855704, which is a nominally significant* ($p = 0.036$) *eQTL for* TUBGCP5*.* TUBGCP5 *is strongly down-regulated in the presence of doxorubicin.*

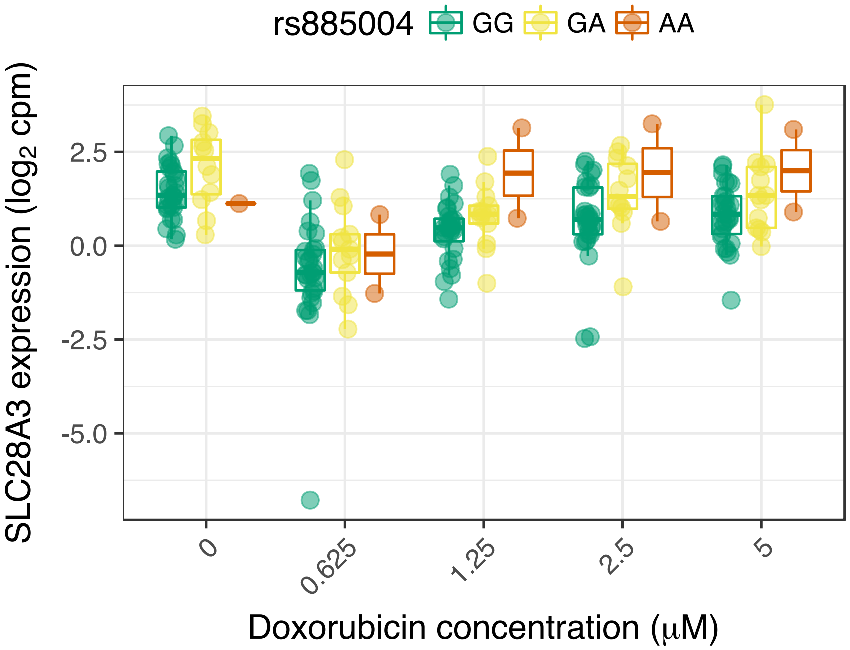


Figure S7: *The intronic (to SLC28A3) rs885004, and the closely linked (R^2 = 0.98) synonymous variant rs7853758, have been associated with ACT by candidate gene studies*[30,29]. *We find rs885004, but not rs7853758, has a significant marginal effect on SLC28A3 expression.*