Functional dynamic genetic effects on gene regulation are specific to particular cell types and environmental conditions

Anthony S Findley[1], Alan Monziani[1], Allison L Richards[1],
Katie Rhodes[2], Michelle C Ward[3,10], Cynthia A Kalita[1], Adnan Alazizi[1],
Ali Pazokitoroudi[4], Sriram Sankararaman[4,5,6], Xiaoquan Wen[7], David E Lanfear[8],
Roger Pique-Regi[1,9,11], Yoav Gilad[2,3,11], Francesca Luca[1,9,11]

[1]Center for Molecular Medicine and Genetics, Wayne State University, Detroit, MI 48201, USA

[2]Department of Human Genetics, University of Chicago, Chicago, IL 60637, USA

[3]Department of Medicine, University of Chicago, Chicago, IL 60637, USA

[4]Department of Computer Science, UCLA, Los Angeles, CA, 90095, USA

[5]Department of Human Genetics, UCLA, Los Angeles, CA, 90095, USA

[6]Department of Computational Medicine, UCLA, Los Angeles, CA, 90095, USA

[7]Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

[8]Center for Individualized and Genomic Medicine Research, Henry Ford Hospital, Detroit, MI 48202, USA

[9]Department of Obstetrics and Gynecology, Wayne State University, Detroit, Michigan 48201, USA

[10]Present address: Department of Biochemistry and Molecular Biology,
University of Texas Medical Branch at Galveston, Galveston, TX, 77555, USA

[11]To whom correspondence should be addressed.
Email: fluca@wayne.edu, gilad@uchicago.edu, rpique@wayne.edu

January 29, 2021

Abstract

Genetic effects on gene expression and splicing can be modulated by cellular and environmental factors; yet interactions between genotypes, cell type and treatment have not been comprehensively studied together. We used an induced pluripotent stem cell system to study multiple cell types derived from the same individuals and exposed them to a large panel of treatments. Cellular responses involved different genes and pathways for gene expression and splicing processes, and were also highly variable across cell types and treatments. For thousands of genes, we identified variable allelic expression across contexts, and characterized different types of gene-environment interactions. Many of these G×E genes are associated with complex traits. We characterized promoter functional and evolutionary features that distinguish genes with elevated allelic imbalance mean and variance. More than 47% of the genes with dynamic regulatory interactions were missed by GTEx, but we identified them using a suitable allelic imbalance study design. This indicates the importance of exploring multiple treatments to reveal previously unrecognized regulatory loci that may be important for disease.

Introduction

Cells exist within complex environments, where levels of metabolites and signaling molecules can change rapidly. In order to thrive under such conditions, cells have evolved mechanisms to control gene expression in response to environmental perturbation. Variation in environmental exposure can explain variation in gene expression in human population samples [2, 31, 38, 49, 68, 110, 118]. For example, a highly correlated cluster of surfactant-related genes was found to be highly expressed in lung tissue from GTEx donors who died while on a ventilator [75]. Similarly, thousands of gene expression differences were identified between sun-exposed and non-sun-exposed skin [55]. In addition to changing gene expression, environmental exposures are also able to alter splicing processes [84]. This may be a direct effect of the environment altering normal splicing patterns, as well as the result of a compensatory effect initiated by the cell to overcome a stressful situation. Splicing is an important co-transcriptional and post-transcriptional process taking place within the nucleus, in which portions of pre-mRNAs called "introns" are removed, and others called "exons" are joined together to form a mature transcript [112]. All forms of splicing require a variable set of proteins and small nuclear RNAs to form small-nuclear ribonucleoproteins (snRNPs), ultimately giving rise to the spliceosome [116] which, together with a variable set of accessory proteins [39, 43] finely tune splicing. Expression of both spliceosomal and accessory factors may be influenced by the environment, which means the environment is eventually able to alter splicing processes [96].

It is widely accepted that gene expression varies across individuals and that such variation is under genetic control. Expression quantitative trait loci (eQTL) mapping in samples from a variety of tissues has elucidated the tissue specificity of the genetic control of gene expression [27, 34, 115]. Specifically, the GTEx consortium observed a U-shaped pattern for eQTL tissue specificity, with eQTLs tending to be either highly shared amongst tissues or highly tissue-specific [8]. These findings underscore the importance of evaluating the genetic control of gene expression across cell and tissue types.

To address the environment-specific control of gene expression, eQTL studies performed on *in vitro* treated cells from many donors are commonly used. The resulting genetic variants which influence gene expression response are known as response eQTL (reQTL) and represent gene-environment interactions (G×E) for molecular phenotypes. reQTL have been identified for infectious agents, drugs, and hormones, among other stimuli/perturbations [3, 4, 10, 19, 29, 48, 54, 57, 61, 70, 72, 73, 82, 91]. These studies have consistently shown that genes with G×E on gene expression are enriched for association with complex traits. However, given the large sample sizes needed to detect reQTLs, only a small number of conditions in a limited number of cell types can be tested at one time.

Allele-specific expression (ASE) is an alternative to eQTL mapping to identify the genetic control of gene expression. Rather than associating a genetic variant with gene expression, ASE uses heterozygous sites within coding regions to measure allelic imbalance in RNA-seq reads. A significant allelic imbalance within an individual implies *cis*-regulatory variation, as the *trans*-environment is constant. Similar to reQTL mapping, differences in ASE between environmental conditions is indicative of G×E, and we refer to such SNPs as conditional ASE (cASE) [81]. cASE has important implications for human health. For example, nearly 50% of genes with cASE were involved in complex traits by GWAS, which is significantly greater than ASE or eQTL genes [81]. Unlike reQTL mapping, ASE can identify G×E in small sample sizes, thereby allowing for interrogation of a broader spectrum of environmental exposures [81].

While gene expression is governed by both genetic and environmental factors, it is subject to fluctuations due to stochastic factors [93]. This noise in gene expression is gene-specific, dependent on promoter elements, and can be affected by genetic mutation [78, 93], indicating its importance in gene regulation. Indeed, there has been selection to minimize gene expression noise in yeast [62], and gene expression variation has been linked to differential gene expression in response to perturbation in flies and humans [106]. While the importance of variation in gene expression has been clearly established, much less is known on

variation in the genetic control of gene expression.

Taking advantage of the reduced sample size needed to measure genetic effects by ASE, we have established a system to measure the genetic effects on gene expression in three cell types from six individuals across a large number of environmental conditions. This study design allows us to simultaneously quantify context (cell type and environment) dependent and genetic effects on gene expression. In addition to identifying thousands of instances of cASE, we have partitioned the variance in gene and allelic expression into individual, treatment, and cell type components, illustrating how each of these components can influence genetic control.

Results

High-throughput scan of gene expression response in 84 different contexts

We have reprogrammed lymphoblastoid cell lines (LCLs) from six individuals into induced pluripotent stem cells (IPSCs), which were further differentiated into cardiomyocytes (CMs) (Figure 1A). Our study was performed in batches on 96-well plates. Each batch consisted of the same cell type from three individuals, with 28 treatments and two controls. Each experiment was performed in duplicate for a total of 12 batches. (Figure 1). In order to shift CM cellular metabolism from fetal-associated glycolysis to adult-associated aerobic respiration, we replaced the cell culture media with galactose-containing media on day 20 [92, 119]. We assessed the purity of CM cultures on days 25 and 27 using flow cytometry to measure the percentage of cells expressing cardiac troponin 2 (TNNT2). Purities ranged from 44.9% - 95% (Table S1). CMs were derived in a single differentiation experiment per individual, so all experiments from CMs in a single individual have identical purities. Principal components analysis on gene expression reveals three distinct clusters corresponding to the three cell types (Figure S1). Additionally, to further verify the identity of the three cell types, we evaluated the expression of LCL, IPSC, and CM marker genes (Figure S2), including TNNT2, which, as expected, showed high expression in CM samples and low expression in IPSC and LCL samples (Figure S2C).

We exposed all cell lines to 28 different treatments (Table S4), resulting in 84 cell type/treatment combinations of cellular contexts. Treatments included hormones, common drugs, vitamins, and environmental contaminants, among others. We used a two-step sequencing approach [80] to identify changes in gene expression, splicing, and allelic expression across cellular contexts. First, we performed an initial shallow RNA-sequencing step (median depth of 9.5M reads) to identify treatments which caused a significant response, indicated by changes in gene expression when comparing each treatment to its vehicle control with DESeq2 [66] (Figure 1B, Table S5). To characterize changes in gene expression, splicing, and allelic expression, we selected 12 of the 28 treatments which induced a substantial change in global gene expression for a total of 36 different contexts.(in bold in Figure 1B, see methods).

Gene expression and splicing response across treatments and cell types

From the 12 deep-sequenced treatments across all cell types (36 contexts and 6 controls, median depth/sample: 146M for LCLs, 148M for IPSCs, 273M for CMs, see Table S6), we identified differentially expressed genes between each treatment and its vehicle control in each cell type using DESeq2 (Figure 1C, Table S7). We found between 53 (insulin in IPSCs) and 21,593 (copper in LCLs) differentially expressed genes (DEGs; FDR < 10%) out of 58,300 genes. We confirmed that the treatments were having physiological effects by examining gene ontology enrichment of differentially expressed genes. For example, as expected glucose catabolic pathways were enriched (FDR = 0.40%) in genes upregulated in response to insulin in CMs relative to all genes expressed in CMs (Table S8). Also, divalent metal ion transport (FDR = 1.8%) and

3
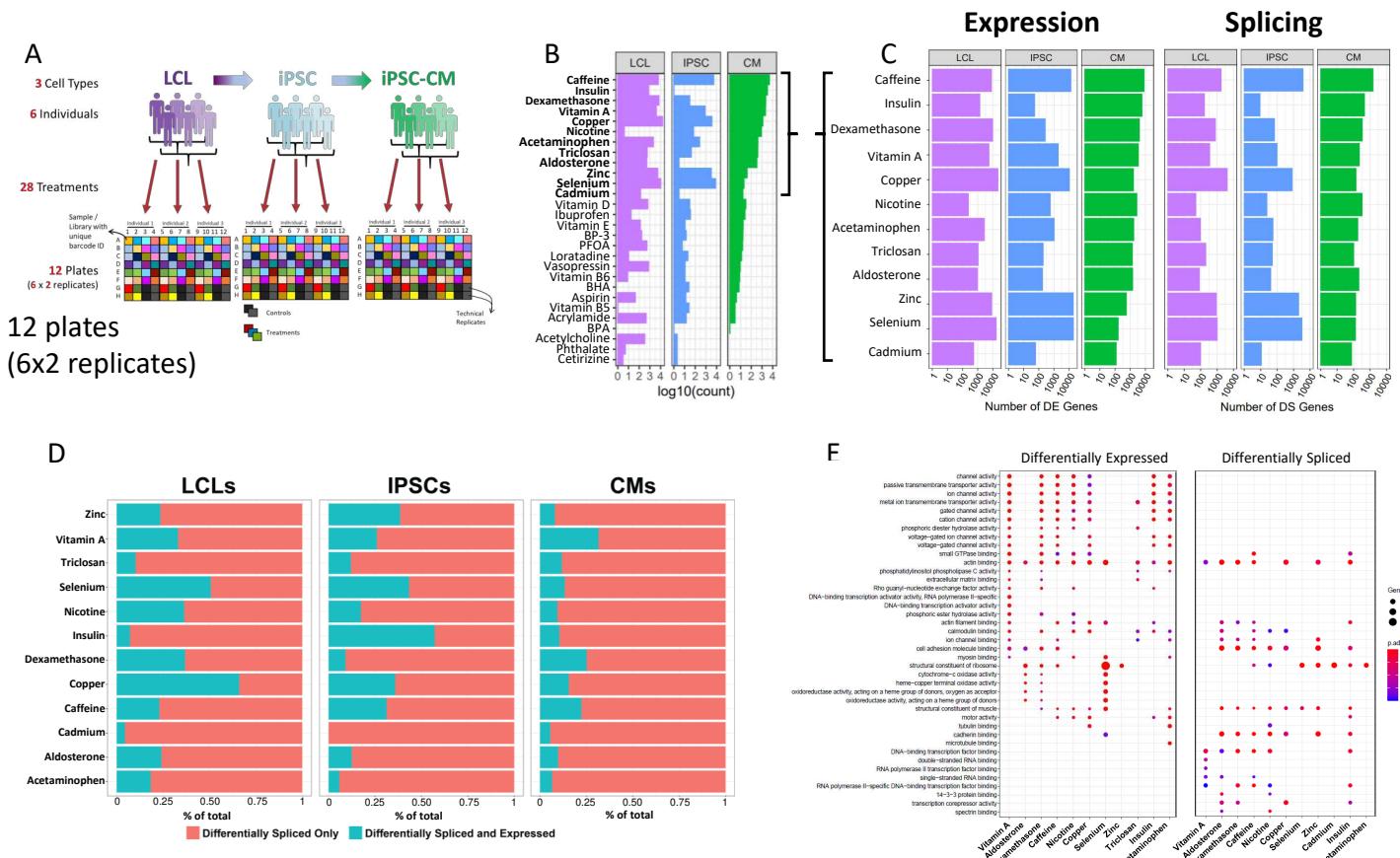
Figure 1: Gene expression and splicing changes in response to environmental perturbation (A) Study Design. We reprogramed LCLs from six donors into IPSCs, which were further differentiated into CMs. We treated all cell lines with 28 treatments and two vehicle controls, performing RNA-seq six hours after treatment. Three individuals per cell line were treated on one plate, and each plate was performed in duplicate, for a total of twelve plates. **(B)** Shallow Sequencing. Number of differentially expressed genes (DEGs) per treatment in each cell type from the shallow sequencing. Treatments in bold were selected for deep sequencing. **(C)** Deep Sequencing. Number of DEGs and differentially spliced genes (DSGs) per treatment in each cell type from the deep sequencing step. **(D)** Normalized stacked barplot of the number of genes that are only differentially spliced and those that are both differentially spliced and differentially expressed, for each treatment and cell type. **(E)** Dotplot generated by ClusterProfiler depicting the top enriched GO terms of CM DEGs and DSGs across environments.

transition metal ion transport (FDR = 2.8%) were enriched in genes upregulated in response to copper in LCLs. When we considered a model with all cell types together, we identified 4,835 genes that show evidence of cell type × treatment interaction effects on gene expression as measured by a likelihood ratio test in DESeq2 (FDR < 10%, Table S9, see Methods Eq. 2). These cell type × treatment interactions indicate that for these genes, the transcriptional response to the environment varies by cell type. Inter-individual variation within the regulatory sequences governing these cell type-specific responses could lead to G×E in a cell type-dependent manner. Interestingly, 96% of differentially expressed genes in CMs in response to treatment with insulin are not differentially expressed in LCLs or IPSCs. These unique insulin-CM genes are enriched for cholesterol and ADP metabolic processes. Together these results highlight the importance of studying environmental effects on gene expression in different cell types to avoid missing physiologically important cell type × treatment effects.

To comprehensively identify global shifts in splicing patterns across cell types, individuals, and environments we employed LeafCutter [63], an annotation-free splicing detection tool based on intron excision from pre-mRNAs. We identified 22,334,519 unique introns, corresponding to 101,450 different transcripts

and 14,066 different genes among all cell types and treatment exposures. We found between 9 (insulin in IPSCs) and 4,106 (copper in LCLs) differentially spliced genes (DSGs; FDR < 10%) between treatment and control in each cell type. As one would expect, we identified treatments triggering consistent splicing alterations across all three cell types, as well as treatments predominantly having an effect in just one or two of them (Figure 1C). For instance, selenium exposure leads to splicing alterations in all three cell types, particularly prominent in IPSCs (3,223 DSGs), while in LCLs and CMs the number of differentially spliced genes is much lower (1,006 and 130 DSGs, respectively). Intron excision can be directional, meaning a given intron can be more or less excised in any given condition, with respect to baseline levels [97]. This may result from a treatment inducing a concerted regulation of splicing towards, for example, intron retention across genes. To investigate this phenomenon, in each condition (i.e. treatment-cell type combination) we extracted the percent spliced in (Ψ) values and performed a two-sided binomial test to determine whether observed excision direction significantly shifts from a random 50:50 distribution. We identified 10 out of 36 environments ($\sim$29%) that show a consistent shift (FDR < 10%) in one direction of intron splicing (Figure S3). More specifically, 4 of them were identified in LCLs (insulin, selenium, zinc, copper), 2 in IPSCs (caffeine, zinc) and 4 in CMs (aldosterone, nicotine, dexamethasone, insulin). When considering the same treatment across different cell types, we found varying consistency in splicing direction. For instance, zinc treatment is associated with an increase in intron excision events in all the 3 cell types considered, although being significant only in LCLs and IPSCs. Response to insulin, instead, is characterized by an increase in intron retention in CMs and intron excision in LCLs. This would suggest that interaction between the environment and cellular contexts eventually determines whether cells tend to overall retain or excise introns.
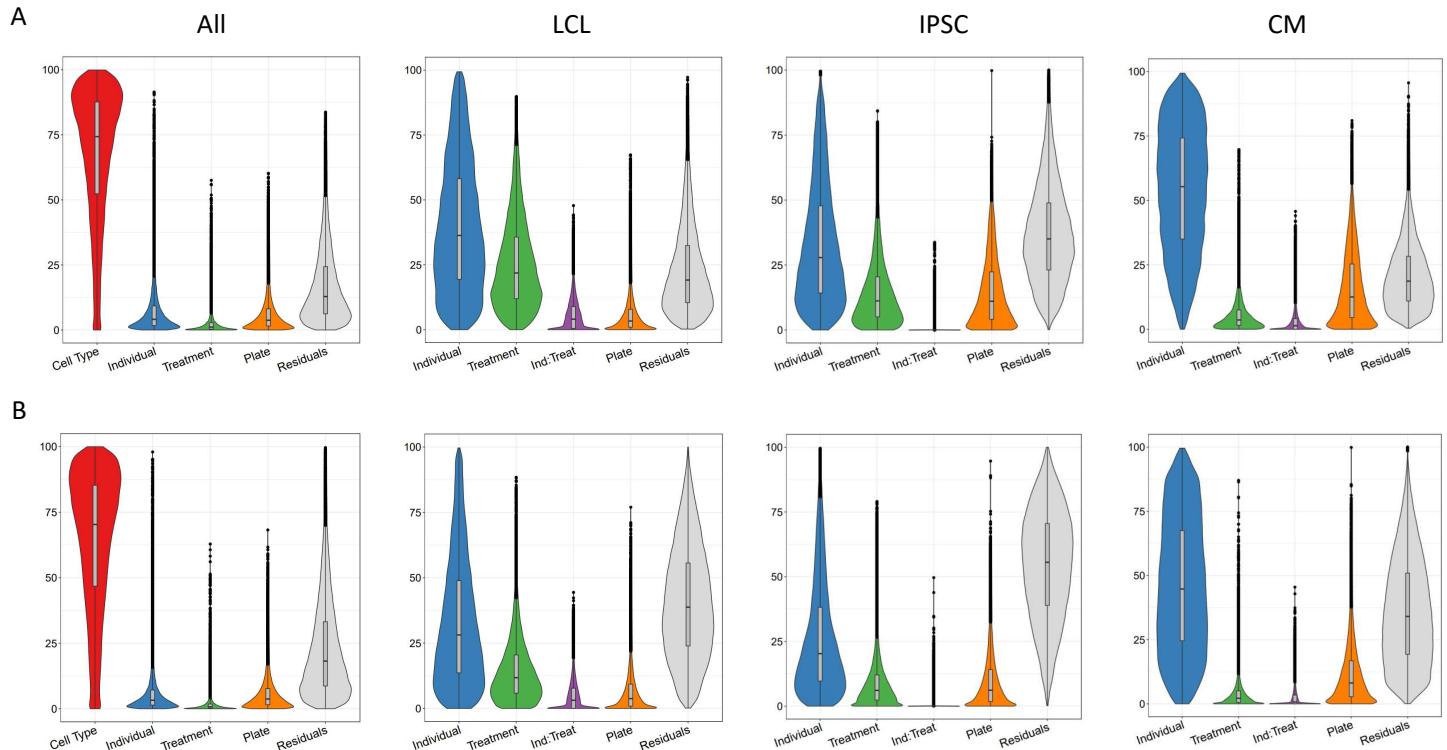

Figure 2: Variance partitioning of gene expression and splicing Sources of variation in gene expression and splicing. Variance for gene expression (**A**) and splicing (**B**) was partitioned into categories (horizontal axis).

Environment triggers distinct changes in gene expression and alternative splicing

Cells are continuously exposed to different stimuli and environments, even in physiological conditions. Responses to those exposures are multi-layered, involving changes at both the transcriptional and post-transcriptional level. Because of that, differentially expressed and differentially spliced genes may be involved in distinct processes, but ultimately contributing to restoring cellular homeostasis. To understand how gene expression changes and alternative splicing contribute to the cellular response, we investigated whether the same genes are DEG and DSG in a certain context. Overall, the number of DSGs is significantly correlated with the number of DEGs (Figure S4, $\rho = 0.91$, $p = 1.4 \times 10^{-14}$), thus indicating that the extent of the environmental effects on the transcriptome is similar for gene expression and splicing. DSGs are also generally enriched for DEGs, except for notable examples, such as insulin in CMs (Figure S5). On average 12.6% of the DSGs were found to be also DEGs in a given condition (Figure 1D). Overall this suggests that there are both co-transcriptional and also independent mechanisms controlling pre and post-transcriptional responses. Furthermore, gene regulatory mechanisms involved in different biological processes responding to the environment likely modify transcription by splicing or changes in expression independently. We identified different biological processes affected by differential gene expression and splicing in response to the treatments in the three cell types (Figure 1E, Figure S6 and Figure S7A and B). DEGs in CMs are enriched for biological processes related to ion channel activity and transmembrane signaling across seven treatments (vitamin A, dexamethasone, caffeine, nicotine, copper, insulin, and acetaminophen), whereas DSGs were enriched for cytoskeletal protein binding across seven treatments (aldosterone, dexamethasone, caffeine, nicotine, copper, zinc, and insulin). Similarly, IPSCs showed a difference, with DEGs being involved in ion channel activity in seven treatments (vitamin A, caffeine, nicotine, copper, selenium, zinc, and acetaminophen), whereas DSGs were enriched for DNA binding and RNA biology terms across all treatments. On the other hand, both DEGs and DSGs in LCLs were highly enriched for cancer and viral-related processes, pathways and diseases across most treatments, without showing any specific difference between DEGs and DSGs. We also considered disease-gene network annotations and found that DEGs and DSGs in CMs are both enriched for genes linked to cardiovascular diseases, including different forms of cardiomyopathies and vascular disorders across all treatments except insulin for DSGs (Figure S8B).

Sources of variation of gene expression and splicing

In order to identify the determinants of gene expression and splicing variation, we partitioned the variance in both gene expression and intron excision [46]. In particular, we were interested in uncovering the extent of the contribution of genetics, environmental and cell type effects to gene expression and intron excision. Variance which cannot be attributed to any of these defined components is counted as residual, which can be generally interpreted as due to stochasticity in the transcription process, technical variation, and/or unknown variables (e.g. cell cycle stage).

First, we considered the relative contribution of cell type, treatment, individual and batch effects (i.e. plate) to the overall variance. To this end, we partitioned the variance of all deep-sequenced samples. As expected, cell type identity has the strongest effect on both gene expression and splicing variance across samples for most genes (variance explained median value 74% and 70%, respectively, which is consistently bigger than the batch effect), with a relatively small contribution from the individual and treatment (Figure 2). Within each cell type, we considered treatment, individual, batch effects (i.e. plate), and we also considered the effect of any interaction between individual and treatment. These interactions should capture potential G×E effects, for a specific treatment. As expected, once we removed cell type effects, we observed a larger contribution from the other factors to both gene expression and splicing variation
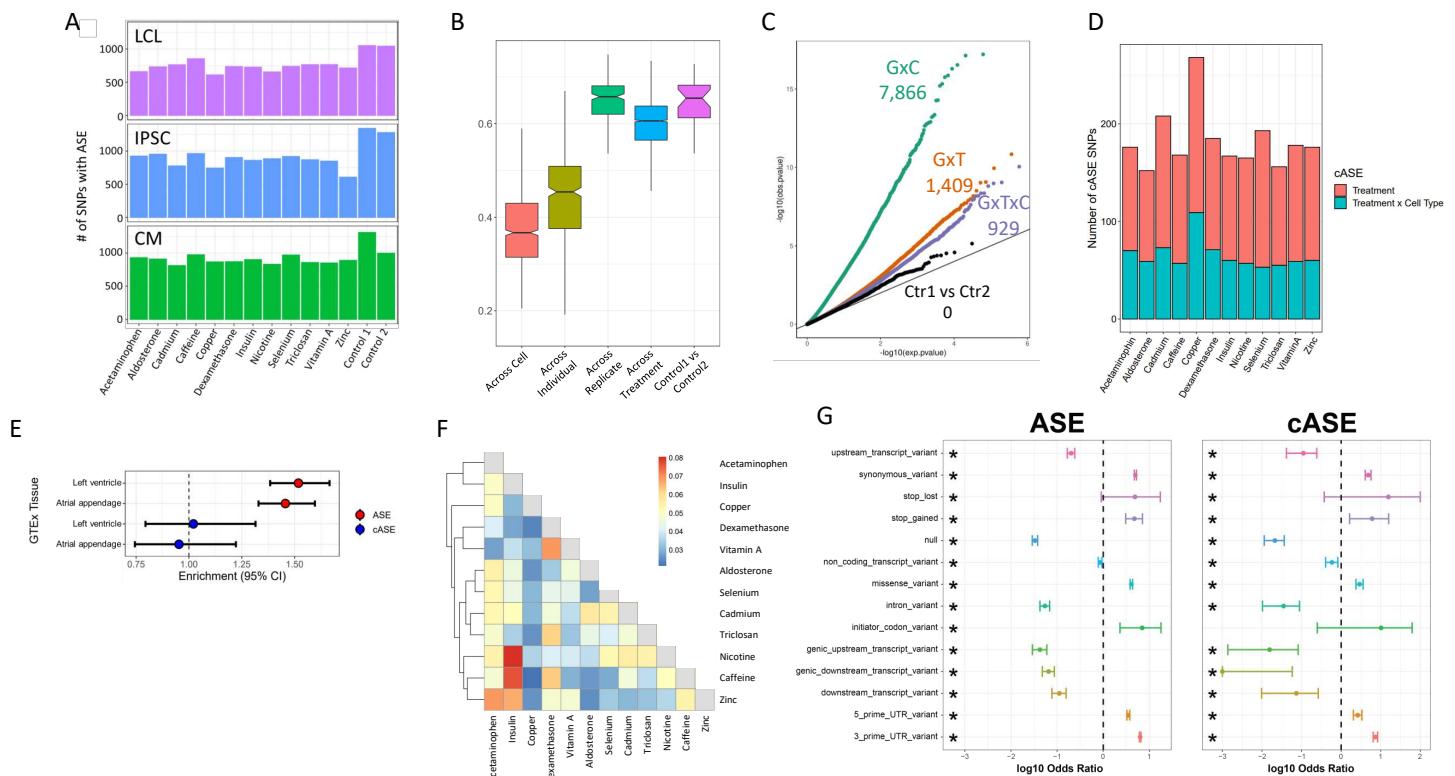
Figure 3: ASE and cASE (A) Number of SNPs with ASE in each treatment and cell type from QuASAR. **(B)** Spearman correlations between libraries across variables. "Across cell" indicates comparisons between libraries which are from the same individual and treatment, but different cell types. "Across individual" indicates comparisons between libraries from the same cell type, treatment, and replicate (i.e. plate), but different individuals. "Across replicate" indicates comparisons between libraries from the same cell type, treatment, and individual, but different replicates (i.e. plate). "Across treatment" indicates comparisons between libraries which are from the same cell type, individual, and replicate (i.e. plate), but different treatments. "Control1 vs Control2" indicates comparisons between the two vehicle controls from the same individual and replicate (i.e. plate). **(C)** QQ-plot depicting interaction *p*-values for SNPs tested for cASE. G×C indicates SNPs tested for GenexCell type effects, G×T indicates SNPs tested for Gene × treatment effects, G×C×T indicates genes tested for Gene × cell type × treatment effects. Ctr1 vs Ctr2 indicates the coefficient in the model that accounts for the vehicle control used for each treatment. **(D)** Number of treatment G×E. Gene × treatment and gene × treatment × cell type cASE is reported for each treatment, irrespective of the cell type in which it was identified. **E** Enrichment of genes with ASE and treatment cASE in GTEx eGenes. ASE and cASE are from CMs, while eGenes are from left ventricle and atrial appendage in GTEx. Odds ratios and 95% confidence intervals are plotted (Fisher's test). **(F)** Gene-by-treatment cASE shared between treatments. For each treatment pair we calculate the proportion of shared cASE over the union of cASE identified in each of the two treatments. **(G)** Forest plot depicting the significant (FDR < 10%) enrichment and depletion of ASE and cASE SNPs with respect to all SNPs tested for different genomic annotations.

across samples. Similar patterns were observed for both gene expression and splicing. For example the contribution of the individual is largest in CMs (55% gene expression, 45% splicing), followed by LCLs (36% gene expression, 28% splicing) and IPSCs (28% gene expression, 20% splicing), similar to what was shown previously for gene expression [9]. Treatment explained the greatest percentage of variance in LCLs (22% gene expression, 12% splicing), followed by IPSCs (11% gene expression, 6% splicing), then CMs (4% gene expression, 2% splicing). When we focused on the variance explained by measured variables (plate, treatment, individual and their interaction) we found that treatment (marginal and interaction with individual) explained at least half of the variance for 64% of the genes.
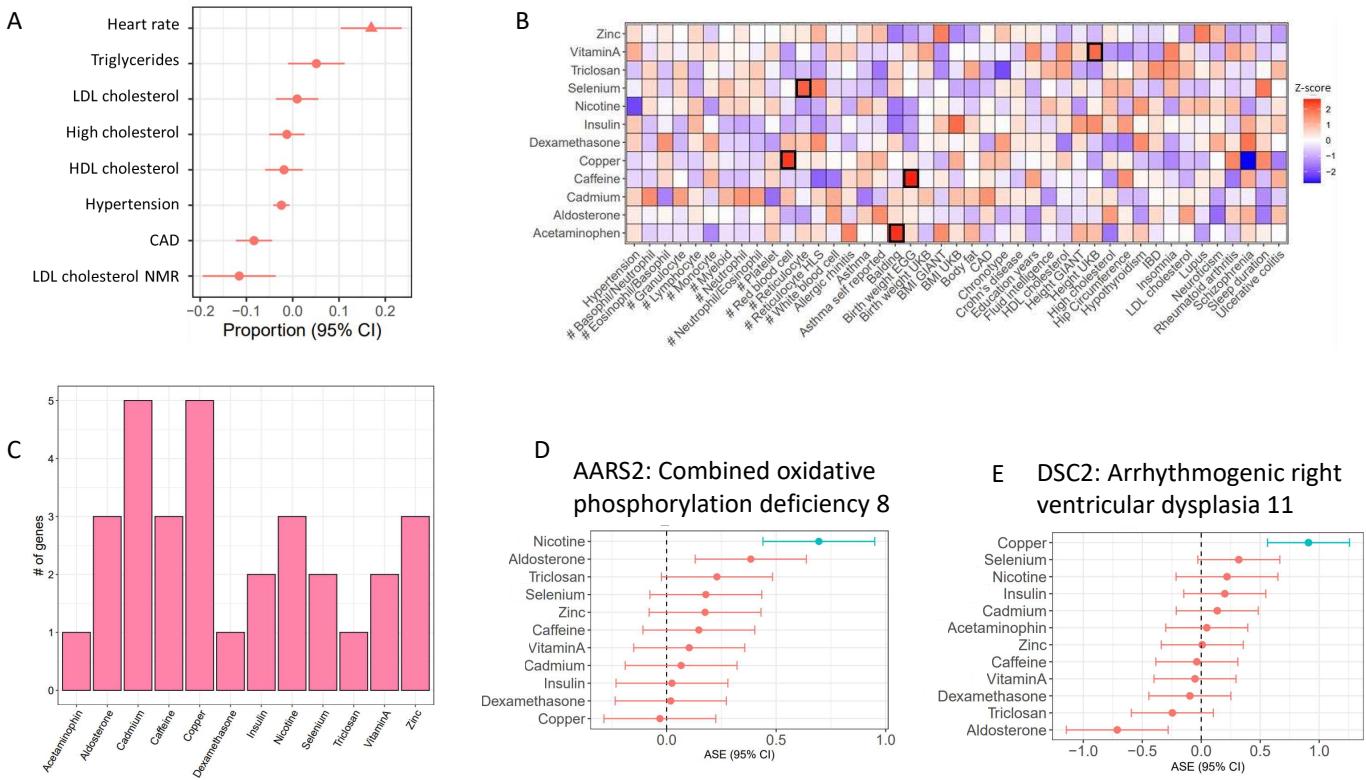
Figure 4: Treatment cASE genes in TWAS (A) CM cASE overlap for genes associated with complex traits in TWAS. For each trait, the forest plot shows the proportion of cASE genes overlapping TWAS genes, relative to the average proportion calculated across all traits. Overlap with genes associated with heart rate was significantly greater than other traits (p=0.01, FDR=0.08). **(B)** Overlap of gene-by-treatment cASE from all cell types with genes associated with complex traits in TWAS. The color of the box represents the Z-score of the one sample proportion test comparing the proportion of cASE-TWAS overlap for a treatment to the average cASE-TWAS overlap across treatments. **(C)** Gene-by-treatment cASE overlap with CARDIoGRAM TWAS. **(D,E)** Forest plots depicting change in allelic imbalance in CMs due to treatments for AARS2 and DSC2, two genes implicated in Medelian disease. Lines in blue indicate significant treatment cASE.

Genetic effects on gene expression across cell types and treatments

To investigate the genetic control of gene expression across cell types and treatments, we identified SNPs exhibiting allele-specific expression (ASE). ASE occurs when there is a transcriptional imbalance between the maternal and paternal copies of an allele. Because ASE is measured within each sample, *trans* factors are kept constant, so any differences in allelic expression is most likely due to *cis*-regulatory variants. We used QuASAR [45] to identify heterozygous genotypes and to provide an initial estimate of the ASE (Table S10). In total, we quantified ASE at 282,278 unique SNPs in 22,397 genes. The number of SNPs with ASE across treatments within each cell type ranges from 612 - 1052 for LCLs, 607-1342 for IPSCs, and 812-1310 for CMs (Figure 3A).

For each individual, we measured ASE in up to 84 experimental conditions (3 cell types, 14 treatments, 2 technical replicates). This gives us an unprecedented opportunity to determine the factors that modify allelic expression in an individual. To avoid excluding conditions where an allele may be lowly expressed, we tested for ASE using a new linear model (see Methods), which incorporated ASE measurements across all conditions for a SNP in a given individual, for a total of 69,205 SNPs (10,142 genes) that can be tested. By doing so, we directly infer the noise inherent in ASE measurements using the linear model (ANOVA).

We first identified 15,497 (~23%) SNPs that show evidence of ASE in any condition (ANOVA test eq. 3 and 4, FDR < 10%, Table S11), corresponding to 5,640 genes. Reassuringly, we observed high correlation of ASE between replicates (i.e. same individual, treatment, and cell type in different batches;

median=0.66), which confirms that technical effects have a limited impact and that ASE measurements are reproducible (Figure 3B). When we consider control conditions on the same plate, the median correlation is identical to the one between replicates, thus confirming that the 2 control conditions can be considered technical replicates of each other. Finally, we observe a significant decrease in the correlation between ASE measurements when we consider different treatments or different cell types, which indicates G×E effects.

We then used the linear model to identify significant cell type-specific effects, treatment-specific effects and cell type × treatment effects on ASE, which we refer to as conditional ASE (cASE)(ref. Methods, Eq. 5, Table S13). We identified 7,866 instances of cell type cASE (5,452 unique SNPs, FDR < 10%), corresponding to gene × cell type interactions; 1,409 instances of treatment cASE (1,102 unique SNPs, FDR 10%), corresponding to gene × treatment interactions; and 929 instances of cell type × treatment cASE (715 unique SNPs, FDR < 10%), corresponding to gene × cell type × treatment interactions (Figure 3C,D). When we considered genes with at least one SNP with ASE, we found 2,822 unique genes with gene × cell type interactions, 979 unique genes with gene × treatment interactions and 689 unique genes with gene × cell type × treatment interactions.

We next investigated whether these genetic effects on gene expression have been previously observed in GTEx, which identified eQTLs in hundreds of individuals and tissue types, but that largely ignored dynamic regulatory interactions. For this analysis we focused on the CMs and their tissue counterparts in GTEx (left ventricle and atrial appendage). In CMs, we identified 3,033 genes with ASE. Of these genes, 50% (1,519) and 52% (1,619) were eGenes in left ventricle and atrial appendage from GTEx, respectively. This translates to a 1.52 and 1.46-fold enrichment of ASE genes in GTEx genes in the left ventricle and atrial appendage, respectively ($p < 10^{-16}$, Fisher's exact test, Figure 3E). Of the 338 genes with treatment cASE in CMs, 170 and 178 are eGenes in left ventricle and atrial appendage in GTEx. Interestingly, treatment cASE genes in CMs were not significantly enriched in either tissue (odds ratio = 1.02, 0.95). These results indicate that investigating the control of gene expression across different environmental contexts, even in a small number of individuals, can identify new instances of genetic regulation that are missed in large eQTL mapping studies that do not explicitly sample different environmental contexts.

To investigate whether our G×E results replicate in other environments previously investigated, we calculated the overlap between genes with gene × treatment interactions in our dataset and genes with G×E identified in fourteen previous ASE and eQTL mapping studies spanning a range of cell types and treatments. Of 979 genes with treatment cASE, 850 (87%) replicated in at least one of these datasets (with $p < 0.05$ in the original study, Table S17). Knowles *et al* [57] exposed IPSC-derived CMs to doxorubicin, a chemotherapeutic agent, and performed response eQTL (reQTL) and response splicing QTL (rsQTL) mapping. When considering genes with a reQTL, 105 could be tested for cASE in CMs and 79 of these genes had nominally significant cASE (p < 0.05), while 12 displayed cASE after multiple test correction (FDR < 10%) spanning 9 different treatments in our study (Table S18). For example, for 3 of those genes (*FRAS1*, *PDGFC*, and *MPHOSPH6*) we can now annotate a G×E interaction with caffeine in addition to doxorubicin. Out of 740 genes with 1 or more sQTLs,182 and 88 have ASE and cASE in our datasets (p <0.05), respectively, whereas of the 62 genes with response sQTLs to doxorubicin 16 have ASE and 11 have cASE (p <0.05). Overall these results point to different regulatory mechanisms that can lead to G×E and that are context-specific.

The number of genes with gene × treatment interactions varies with a low of 96 (93 SNPs) for aldosterone and a high of 176 (159 SNPs) for copper (1.8 fold range between minimum and maximum number of cASE). This is in contrast with the spread in gene expression and splicing responses across treatments (two orders of magnitude lower than the range between minimum and maximum number of DEGs). When we considered cASE in each cell type separately, indeed we did not find a significant enrichment or depletion for cASE genes being differentially expressed or differentially spliced between treatment and control (Figure S9). However, despite the number of cASE being similar across treatments, the genes with G×E

are largely different. Most treatment cASE SNPs were specific to one treatment (80.5%, Figure S10), so there was little sharing between treatments (Figure 3F). While this could be due to a lack of power in detecting cASE with small effect sizes, it also suggests that genes with the strongest cASE are different across conditions.
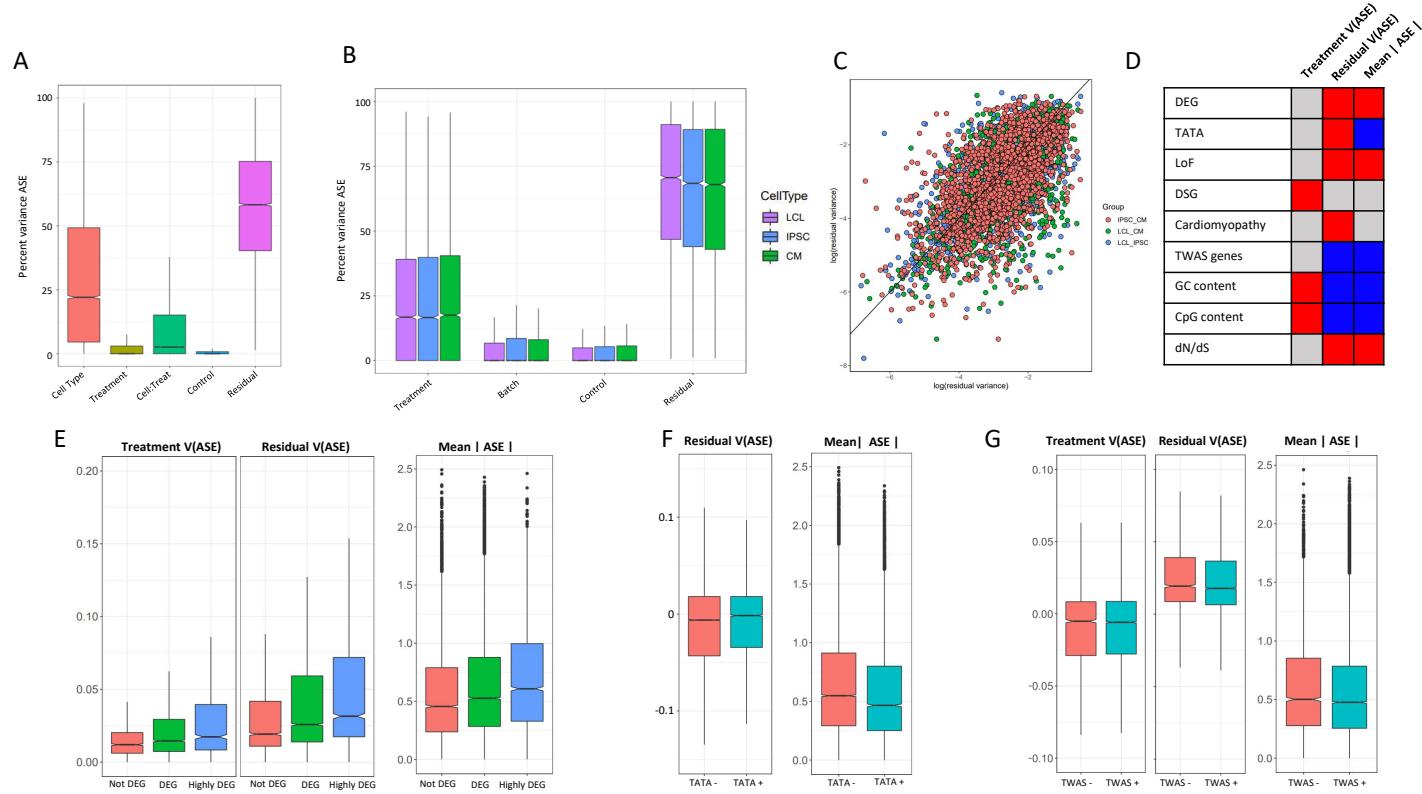

Figure 5: ASE residual variance (A) Variance partitioning of ASE across all cell types. **(B)** Variance partitioning of ASE within each cell type separately. **(C)** ASE residual variance across cell types. For SNPs with ASE variance measured in more than one cell type in the cell type-specific model, we have plotted the residual variance in each pair of cell types. **(D)** Relationships between 8 gene annotations and treatment ASE variance, residual ASE variance, and mean absolute ASE. Red indicates that the annotation is positively correlated to ASE variance or absolute ASE, while blue indicates a negative correlation. Gray indicates no significant relationship. **(E)** Treatment variance, residual variance, and mean absolute ASE by DEG. We have separated genes into 3 categories: 1) Not DEG: Not differentially expressed in response to any treatment. 2) DEG: Differentially expressed with log2 fold change > 0.25 in at least one treatment but not greater than 0.75 in any treatment. 3) Highly DEG: Differentially expressed with log2 fold change > 0.75 in at least one treatment. Each gene can only be in one category. **(F)** Residual ASE variance and mean absolute ASE for genes with or without a TATA box in the promoter. **(G)** Treatment variance, residual variance, and mean absolute ASE for genes significant or not significant by TWAS.

Biological and clinical significance of allele-specific expression across cell types and treatments

SNPs with ASE and cASE may have a direct effect on gene expression or more often may indirectly reflect the effect of a regulatory variant located in non-coding regions. Regulatory sites that could be located within the gene transcript itself include those in the 3' and 5' UTRs, in splicing junctions, premature stop codons, and possibly exonic enhancers. Thus SNPs with ASE and cASE located within these regulatory sites could be putative causal sites for the observed allelic imbalance [79].

10

We retrieved the genomic location of the SNPs associated with ASE and cASE, as well as their predicted functional impact from dbSNP (https://www.ncbi.nlm.nih.gov/snp, [105]) (Figure S11). When we considered all variants that could be tested for ASE, both ASE and cASE SNPs were found to be significantly enriched in 5' and 3' UTRs, and among variants classified as missense or altering a stop codon (FDR < 10%, Fisher's exact test; Figure 3G). On the contrary, SNPs located within upstream/downstream transcript and genic regions, within introns and in splicing sites were found to be significantly depleted in ASE/cASE SNPs. Taken together, these data suggest that functional SNPs with ASE or cASE most likely exert their function by changing transcriptional and/or RNA stability regulatory sequences because the 5' and 3' regions tend to affect these processes, rather than affecting splicing processes.

To investigate the role of genes with cASE in human phenotypes, we considered genes associated with complex traits. To this end, we used the results from Probabilistic Transcriptome Wide Association Studies (PTWAS) [124], which combined eQTL data from GTEx and GWAS data from several large scale studies to identify which genes are most likely to be in the causal pathway for complex traits. First, we considered eight cardiovascular traits, including blood lipids, hypertension, coronary artery disease, and heart rate. Heart rate was the only trait with a significant overlap with CM-specific cASE, compared to the average overlap with other cardiovascular traits (FDR < 10%). Interestingly, heart rate is the trait among those tested where cardiomyocytes have a direct physiological role (Figure 4A). We then investigated the role of G×E in a larger panel of 45 complex traits (Figure 4B). Five traits were enriched for treatment cASE genes in at least one condition (p < 0.05) compared to the other treatments. For example, genes associated with low birthweight are enriched for genes that have caffeine cASE (Figure 4B). For the 428 genes associated with coronary artery disease (CAD), cASE for metal treatments have the greatest overlap, with cadmium and copper each having five cASE genes associated with CAD, and zinc having three (Figure 4C). Interestingly, cadmium causes endothelial dysfunction and promotes atherosclerosis [77], while copper deficiency has been linked to the development of CAD [28]. *FAM213A* (also known as *PRXL2A*), which is involved in redox regulation [40, 120] is shared as treatment cASE for cadmium and copper, in addition to being treatment cASE for nicotine.

We also considered an atrial fibrillation GWAS which identified 151 genes near associated loci [83]. Ten of these genes displayed treatment cASE in at least one condition, for a total of 21 gene-cASE condition pairs (Figure S12, Table S19). The greatest overlap was with dexamethasone cASE. Corticosteroids are antiinflammatory drugs which share molecular pathways with inflammatory disease. Inflammation is a risk factor for atrial fibrillation [44, 47]. Additionally, in multiple population-based, case-control studies, corticosteroid use was associated with new-onset atrial fibrillation [23, 114]. Unlike atrial fibrillation, GWAS for heart failure have not identified many association signals, likely due to the highly heterogeneous nature of the disease. Two recent meta-analyses identified 3 and 13 genes, respectively [6, 102]. Three of the genes associated with heart failure had treatment cASE: *FAM241A* in 5 conditions (copper, dexamethasone, insulin, caffeine, and vitamin A); *BAG3* in 5 conditions (dexamethasone, caffeine, vitamin A, nicotine, and aldosterone); and *KLHL3* in triclosan.

While most causes of heart disease are not due to a mutation in a single gene, regulatory variation controlling the expression of Mendelian disease genes can affect complex trait risk [35]. In CMs, we found 3 genes in OMIM (Online Mendelian Inheritance in Man) [42] with gene × treatment interactions that are known to cause Mendelian forms of heart disease: *PSMA6*, *AARS2*, *DSC2*. *AARS2* displayed cASE in response to nicotine and mutations in this gene cause Combined Oxidative Phosphorylation Deficiency 8, which manifests as fatal infantile hypertrophic mitochondrial cardiomyopathy (Figure 4D). *DSC2*, instead, displayed cASE in response to copper and mutations in this gene cause Arrhythmogenic right ventricular dysplasia 11 (Figure 4E).

Much of complex trait heritability is not explained by genome-wide significant variants, but rather results from the contribution of many variants with smaller effect sizes [14, 71, 121]. Methods to partition

11

the heritability of complex traits can be used to determine the proportion of heritability attributable to genomic annotations [33, 122]. We used RHE-mc [87] to partition the heritability of complex traits from UK Biobank based on cASE annotations. First, we partitioned the heritability of 17 traits, using an annotation of genes with cell type cASE for CMs. CM cell type cASE was significantly enriched for diastolic blood pressure, autoimmune disease, and height. We then considered 22 traits, using treatment cASE as an annotation (Figure S13). Diastolic blood pressure was enriched for vitamin A, systolic blood pressure was enriched for vitamin A, copper, and zinc. Respiratory disease was enriched for caffeine, and smoking status for selenium.

Characterizing variation in genetic regulation of gene expression

The presence of cASE indicates that the genetic control of gene expression varies significantly with changes in the environment. To further investigate the role of the environment and cell type in controlling ASE, we used a mixed-effects model to partition the variance in ASE for each SNP-individual pair across all experimental conditions, similar to our analysis of gene expression and splicing variance (see methods). When considering all cell types together, the greatest amount of explainable variance in ASE was captured by the cell type (Figure 5A, Table S20), similar to the results explaining variability in gene expression (Figure 3A). However, when considering variance in ASE within each cell type, a much larger proportion of the explainable variance was attributed to the effect of a treatment (Figure 5B, Table S21). Specifically 17.5%, 16.6%, and 16.8% of variance (median) in ASE was attributed to treatment effects in CMs, IPSCs, and LCLs, respectively, compared to 3.6%, 11.2%, and 21.9% variance (median) in total gene expression.

There has been extensive investigation into gene expression stochasticity [106]. However, the role of stochasticity in the genetic control of gene expression has proven more difficult to study [100]. Because our experimental design measures ASE across many conditions, we can analyze the residual variance as a measure of genetic control stochasticity. We found that residual variance at a particular SNP is correlated ($r \in [0.55 - 0.59]$, $p < 10^{-16}$, 95% confidence intervals = [0.55, 0.63] for LCL vs IPSC, [0.50, 0.59] for LCL vs CM, and [0.55, 0.61] for IPSC vs CM) across cell types, indicating that unknown factors that contribute to the genetic control of gene expression are conserved across tissues (Figure 5C). In LCLs, genes with high residual variance were enriched in 22 GO terms, with the greatest enrichments being related to tissue morphogenesis and developmental processes, while genes with low residual variance were enriched in 82 GO terms, including RNA processing and immune activation. In IPSCs, low residual variance genes were enriched in 128 GO terms, including peptide metabolic pathways, which was also seen in CMs. We found that low residual variance genes in CMs were enriched in 123 GO terms, including processes related to cell signalling (Table S22, Figure S14).

To investigate the biological significance of ASE changes in magnitude and variance, we considered 9 genomic features: genes differentially expressed in response to any treatment, genes differentially spliced in response to a treatment, genes with a TATA-box promoter, CpG percentage in the promoter, GC percentage in the promoter, genes which cause mendelian cardiomyopathies, genes associated with complex traits by TWAS, gene tolerance to loss of function mutations, and dN/dS ratio (the ratio of nonsynonymous to synonymous mutations in a gene, which is a measure of selective pressure) (Figure 5D). We first considered the distribution of allelic expression for each of these categories. DEGs were characterized by high ASE compared to genes that do not respond to the treatments (Figure 5E). Allelic expression was also elevated for genes tolerant to loss of function mutations, while genes associated with complex traits tend to have lower allelic expression (Figure 5F,G), which reflects the difference in phenotypic relevance for these gene categories. Genes with lower ratio of nonsynonymous to synonymous mutations (dN/dS, i.e. genes under negative selection) are associated with lower mean ASE. When we considered regulatory features of the genes with ASE, we found lower ASE for genes with a TATA box, while ASE is generally

larger for genes with higher promoter CpG and GC content.

When focusing on variance in allelic expression, we found differential gene expression to be associated with increased residual variance, suggesting that for genes with more dramatic responses to the environment, gene expression is under a less stringent genetic control (Figure 5E). Differentially spliced genes, however, are associated with increased treatment variance, and not increased residual variance (Figure S15). We then considered promoter architecture. SNPs in genes with a TATA box had increased residual variance compared to SNPs in genes without a TATA box (Figure 5F). We found no significant relationship between treatment variance and whether the gene contains a TATA box. In addition, GC content and CpG islands in a gene's promoter was associated with increased treatment variance and decreased residual variance. Genes which are known to cause cardiomyopathies when mutated had a small increase in residual variance ($p = 0.0032$), while genes implicated in any of the 103 TWAS in [124] had decreased residual variance. Finally, we considered if intolerance to loss of function and dN/dS ratio could be related to ASE variance. As expected, genes that are less tolerant to loss of function mutations and genes under negative selection (i.e. low dN/dS) have significantly lower residual variance.

Discussion

A fundamental open question in the field of genomics is understanding the sources of variation in gene expression across different individuals, cell types and environmental contexts. Each of these components has been investigated separately in previous studies [2, 3, 4, 10, 19, 27, 29, 31, 32, 34, 38, 48, 49, 54, 57, 61, 68, 70, 72, 73, 81, 82, 91, 110, 115, 118]. Here, in order to investigate the transcriptome along the three different axes, we used a study design that combines IPSC technology, high-throughput screening and allele-specific expression analyses. The cell type context has the overall strongest effect on gene expression, splicing, and allelic expression. This is observed both in terms of differences of the mean gene expression levels across conditions using fixed effect models, or in analyzing the variance on gene expression using random effect models. Importantly, we demonstrate that a large number of context-specific genetic effects on gene expression are not captured by existing large cohort eQTL studies (e.g. GTEx), but can be discovered even with limited sample sizes, when using an allele-specific expression study design. Our unique approach, accounting for cell-type but also genetic and environmental influences has also revealed that environmental impact on gene expression is substantial and particularly important for genes that influence complex traits.

IPSCs are a valuable model system for studying primary cell types which are difficult to obtain and/or culture [107]. Extensive work by our group and others have shown that despite the potential introduction of reprogramming effects, IPSC-derived CMs mimic primary CMs [86], recapitulate *in vivo* CM phenotypes [17, 22, 74], retain donor-specific expression signatures [18, 25, 85, 99], and can be used to assess the effects of environmental exposures on CM gene expression and cellular physiology [17, 56, 57, 74, 103, 104, 119]. This ability to study CMs themselves is especially important given that GWAS studies for heart failure have revealed little of the genetic architecture of the disease. While pedigree analysis suggests that the heritability of heart failure is 26% - 34% [65], two recent, large-scale GWAS identified only eleven [102] and three [6] genomic loci associated with heart failure. This could be due to a lack of quantitative measurements of heart failure, complex etiologies, or the significant environmental contributions to developing heart failure [94]. Our approach allows for interrogation of the effects of environmental exposures in a controlled environment on CM gene expression.

Some treatments had very dramatic effects on gene expression, such as copper and selenium in LCLs with more than 10,000 DEGs each. However, these large changes in gene expression were seen in our previous work [81] (7,869 DEGs in copper, 14,057 DEGs in selenium), and the increased sequencing depth and

greater number of individuals in this study increased our power to detect DEGs. On average, we show the number of DSGs is directly correlated with the number of DEGs, therefore suggesting that environmental perturbations have a similar degree of effect on both transcriptional and post-transcriptional processes. However, only a minor fraction of genes were found to be both differentially expressed and spliced, which suggests that changes in gene expression and splicing represent distinct regulatory mechanisms by which cells can respond to their environment. Indeed, this is supported by GO and KEGG enrichment analyses, showing DSGs in CMs are involved in cytoskeletal activities, whereas DEGs were found to play a role in ion channel processes.

Partitioning the variance in gene expression and intron excision across all cell types gives a broad overview of the relative contribution of cell type and treatment components to transcription and splicing processes. There were similar patterns for gene expression and splicing. Similar to previous work on partitioning the variance in gene expression across a panel of LCLs, IPSCs, and CMs by Banovich *et al* [9], variance due to the individual component was greatest in CMs, followed by LCLs and IPSCs. Yet herein, we were additionally able to consider the treatment component, which was greatest in LCLs and least in CMs. One possible explanation is that the chosen treatments may have more effect in LCLs; alternatively, this may be a consequence of greater inter-individual heterogeneity in CMs compared to LCLs. Regardless of cell type, residual variance was greater for splicing than gene expression. This is consistent with previous work showing that splicing is an error-prone process with a high degree of noise [76, 88, 117].

A key new feature of our study design is that it allowed us to analyze allelic expression across 12 environments (and 2 controls), 3 cell types, 6 individuals and 2 technical replicates each, for a total of 504 experimental samples. This is the largest single study of allelic expression comparing in parallel cell type effects, treatment conditions, and their interactions while controlling for technical variation. We used a new linear model to precisely estimate allelic expression and its variance directly from all the measurements of ASE for each individual (up to 84). As a result, we have a more complete view of the pervasiveness of G×E and we were able to directly investigate whether environmental effects on genetic regulation of gene expression differ across cell types. Our study design allowed us to specifically investigate interactions between genotype, cell type and treatment. Interestingly, many treatment effects on allelic expression vary across cell types, therefore it is important to consider jointly cellular and environmental contexts. As the IPSC technology allows investigators to study multiple cell types from the same individuals, future study designs should consider the importance of studying environmental effects across cell types to learn about pleiotropic effects of potential biomedical relevance.

Both eQTL and ASE studies have demonstrated that genetic effects on gene expression vary significantly across cellular [8] and environmental contexts [3, 4, 10, 19, 29, 48, 54, 57, 61, 70, 72, 73, 81, 82, 91]. However we don't know how stable the genetic control of gene expression is across technical and environmental contexts. This cannot be investigated with eQTL mapping because the genetic effects on gene expression are estimated across individuals, but can be explored with multiple measurements of allelic expression from the same individual. We observed that residual variance of allelic expression is conserved across cell types, which indicates that the intrinsic properties controlling variation in allelic expression are consistent in different tissues and have biological significance. For example, genes which had a strong expression response to at least one treatment had increased residual variance over genes which had no response or a smaller response, while genes with a splicing response had greater treatment variance only. We hypothesize that greater residual variance indicates a less stringent control of allelic expression. This may depend on the regulatory architecture of these genes and may similarly enable large fluctuations in gene expression across contexts. Prior work on gene expression variance has demonstrated that genes with a TATA box have increased noise [78]. This was also shown by [106] as well as in our data here (Figure S16). However, this finding has not been evaluated for the genetic control of allelic expression. Our results indicate that genes with ASE which are under the control of a TATA box promoter have increased residual variance

14

over ASE genes without a TATA box, and thus greater noise. TATA box promoters, as opposed to CpG island-associated promoters, have been associated with tissue-specific genes and high conservation across species [21]. Genes which are more tolerant to loss of function mutations showed greater allelic expression residual variance, indicating that redundancy in gene function allows for less stringent genetic control. This could allow for the evolution of new regulatory elements, resulting in new patterns of gene expression. Conversely, genes which are under negative selection (i.e. low dN/dS) have low residual variance, underscoring the importance of preserving stable expression of these genes. Tolerance to ASE variation could result in robustness against regulatory decoherence [60], which could be further explored in future studies. Cell type and treatment specific effects are the largest identifiable sources of variation in allelic expression for hundreds of genes, as we demonstrated using both variance decomposition and fixed effects as statistical models.

Splicing sites have very conserved sequences, which are recognized by spliceosomal and accessory proteins which ultimately determine splicing patterns. Moreover, splicing is finely tuned by regulatory *cis* RNA sequences within both exons and introns, which are recognized by several RNA-binding proteins (RBPs). Mutations within these consensus sequences therefore have catastrophic consequences in pre-mRNA splicing, eventually being associated with a plethora of different pathological conditions [5, 30, 67], including several forms of cystic fibrosis[12, 36] and in both Becker and Duchenne muscular dystrophies [41]. In addition to deleterious mutations, naturally-occurring polymorphisms such as SNPs may contribute to alter the strength of splicing signals, eventually changing splicing outcomes. However, SNPs linked either to ASE or cASE were found to be mainly depleted in both donor and acceptor splicing sites, eventually suggesting that differential intron splicing is not a major mechanism underlying ASE/cASE. In fact, those SNPs were found to be enriched within several genic regions, in particular within UTRs and, interestingly, causing either the loss or gain of stop codons. These events have dramatic effects on RNA stability, as improper stop codon localization can activate the nonsense mRNA decay (NMD) pathway, resulting in transcript degradation [16]. Moreover, synonymous polymorphisms may alter the codon optimization of a given mRNA, a process which has been linked to reduced half-life [90]. All these events may eventually lead to reduced abundance of the transcript carrying a particular allele, compared to the other copy, thus resulting in ASE.

We identified 3,198 genes with cASE, including 2,822 genes with cell type cASE, 979 genes with treatment cASE and 689 genes with treatment × cell type cASE. Many genes with treatment cASE overlap with genes that have context-specific genetic effects discovered by other studies; yet there is little sharing for any specific pair of conditions. In our study, since we compare ASE within the same genetic background, this limited sharing of ASE between treatments strongly suggests the existence of independent regulatory variants in context specific cis-regulatory modules. eQTL mapping consortia, like GTEx, identify eQTLs under one arbitrary condition, so genetic effects on gene expression which occur only transiently during development or under certain environmental conditions will be missed [24, 32, 81, 95, 108, 113]. A common approach to increasing identification of eQTLs is to increase the sample size. However, interrogating additional environmental conditions will be required to gain a more complete understanding of the genetic control of gene expression. This is reflected in the lack of enrichment for CM cASE genes in GTEx heart tissues. Even with a sample size of only six individuals, we identified many new genes in CMs which are genetically regulated under specific environmental conditions. Approximately half (∼1,500) of the ASE genes we identified in CMs were not present in GTEx heart tissues. They were identified as ASE genes in our study due to the wide range of compounds we exposed them to. More than 160 of these ASE genes unidentified by GTEx showed evidence of cASE, thus pinpointing the environmental condition which was responsible for altering regulation of gene expression. Indeed a large fraction of cASE genes (>47%) are not eGenes in GTEx. This approach could be further expanded by targeting individuals with large polygenic risk scores to discover additional G×E effects of clinical relevance, in future studies.

Our results show the importance of considering the relevant tissue type, as CM-specific cASE are enriched for heart rate, which is a phenotype with a direct physiological role for the heart muscle. The importance of the identification of G×E in a range of environments is illustrated by our findings that Mendelian cardiovascular disease genes and TWAS genes in a wide variety of traits display G×E. In total, across all cardiovascular related traits we identified 169 genes with cASE, spanning the 12 conditions tested. In cardiomyocytes, 52 genes with cASE were associated with cardiovascular disease traits. Among the treatment effects, metal ions (cadmium and copper) and corticosteroids (dexamethasone) are the most common environments interacting with genetic risk for CVD and atrial fibrillation, respectively.

For example, the aryl hydrocarbon receptor nuclear translocator 2 gene (ARNT2) was identified by GWAS for atrial fibrillation [83] and is a treatment cASE gene in CMs for insulin. ARNT2 is part of the hypoxia-inducible factor (HIF) pathway [69], which has a role in the progression of obesity and metabolic disease [37]. Indeed, insulin itself is known to induce the expression of HIF1A [111], which we confirm in our differential gene expression data. In turn, the HIF pathway is especially important for cardiovascular disease and the response to cardiac ischemia [101].

The polygenic nature of human complex traits provides a formidable challenge to tackle the genetic and molecular basis of interindividual variation. Our study demonstrates that interactions between genetic and environmental factors are common, but require specifically designed studies to be discovered. These results also have direct implications on issues related to the portability and interpretation of polygenic risk scores across individuals exposed to different environments. While gene-environment interactions further complicate the overall picture of human complex trait variation, they may represent an important contribution to the overall missing heritability that requires further study and careful consideration.

Methods

Cell culture and treatments

Experiments were conducted using three cell types derived from the same six individuals: Lymphoblastoid Cell Lines (LCL), Induced Pluripotent Stem Cells, (IPSC) and IPSC-derived Cardiomyocytes (CM). All cells were cultured at 37°C with 5% CO_2. Each batch consisted of the same cell type from three individuals, with 28 treatments and two controls. Each experiment was performed in duplicate for a total of 12 batches (Figure 1).

LCLs from six Yoruba individuals were purchased from Coriell: GM18858, GM18855, GM18505, GM18912, GM18520, and GM19209. LCLs were maintained at a density of 200,000 to 1 million cells/ml in supplemented RPMI media (500 mL RPMI-1640 with glutamine (Fisher Scientific, 15-040-CM), 75 mL FBS (Genemate S1200-500), 5 mL GlutaMAX (35050-061, ThermoFisher Scientific) and 5 mL penicillin/streptomycin). 50,000 cells were plated per well of a round-bottom 96-well plate in 100 ul supplemented RPMI media 48 hours before treatment. Cells from each individual were plated in 32 wells representing 28 treatments and 2 controls.

Each of the six LCLs were reprogrammed into iPSCs using episomal reprogramming vectors and expanded on a layer of MEF prior to conversion to feeder-independent growth as previously described [9]. iPSCs were seeded on plates coated with a 1:100 dilution of Matrigel hESC-qualified Matrix (354277, Corning, Bedford, MA, USA) and cultured in iE8 media (Invitrogen A1517001) supplemented with penicillin/streptomycin. Cells were passaged every 3-5 days using dissociation reagent (0.5 mM EDTA, 300 mM NaCl in PBS). During plating of cells, media was supplemented with 10 uM ROCK inhibitor (stemolecule Y27632, stemgent 04-0012) to aid in cell adherence. Media was changed every day thereafter. 50,000 cells were plated per well of a Matrigel-coated flat-bottom 96-well plate in supplemented E8 media

48 hours before treatment. Cells from each individual were plated in 32 wells representing 28 treatments and 2 controls performed in duplicate. Additional information on IPSC reprogramming, including number of passages and differentiation batch can be found in Table S2.

CMs were differentiated from iPSCs using small molecules as previously described [9]. Briefly, Wnt signaling was modulated by treating iPSCs with 12 uM of the GSK3 inhibitor CHIR99021 (4953, Tocris Bioscience, Bristol, UK) followed by 2 uM of the Wnt signaling inhibitor Wnt-C59 (5148, Tocris Bioscience). Cells start to spontaneously beat between days 7-10. The cardiomyocyte population was enriched through metabolic purification by culturing the cells in glucose-free, lactate-containing media from days 14-20. In order to promote aerobic respiration following purification, on Day 20, the cell culture media was replaced with galactose-containing media (500mL DMEM (without glucose, Life Tech A11430-01), 50 mL FBS (GeneMate S1200-500), 5 mL sodium pyruvate (Gibco 11360-070), 2.5 mL HEPES (Fisher SH3023701), 5 mL GlutaMAX (35050-061, ThermoFisher), 5 mL penicillin/streptomycin and 990 mg galactose (Sigma G5388). On Day 25, five days before treatment, cells were trypsinized and split to 114,000 cells/well in $100\mu L$ of galactose-containing media on a Matrigel-coated flat-bottom 96-well plate. Cells from each individual were plated in 32 wells representing 28 treatments and 2 controls. Purity of the cell cultures was determined by measuring the expression of the cardiomyocyte-specific marker cardiac troponin T (564767, BD Biosciences) by flow cytometry. Flow cytometry was performed on Day 25 for 18912, 18520, and 19209 and on Day 27 for 18858, 18855, and 18505. Purities ranged from 44% to 95% (Table S1). Additional information on CM differentiation, including cell counts and the day at which the cells spontaneously started beating can be found in Table S3.

Samples were treated on a total of twelve plates, and each plate was processed as a batch. Each plate contained samples from three individuals from a single cell type, exposed to 28 treatments plus 2 vehicle controls (water and ethanol, referred as Control 1 and Control 2). Additionally, each plate had a technical replicate performed the following day. Importantly all three cell types for the same group of individuals were treated and harvested in parallel on the same day, to avoid that cell-type effects are confounded with batch effects. Cells were treated for six hours at concentrations listed in Table S4. As in [81], concentrations were chosen based on the highest physiological concentrations as reported by the Mayo Clinic (http://www.mayomedicallaboratories.com) or the CDC (http://www.cdc.gov/biomonitoring/), as available.

RNA Library Preparation

For RNA library preparation, each plate was prepared as a batch. Treated cells were collected by centrifugation at 2000 rpm and washed 2x using ice cold PBS. Collected pellets were lysed on the plate, using Lysis/Binding Buffer (Invitrogen), and frozen at -80°. Poly-adenylated mRNAs were subsequently isolated from thawed lysates using the Dynabeads mRNA Direct Kit (Ambion) and following the manufacturer instructions. RNA-seq libraries were prepared using a protocol modified from the NEBNext Ultra II Directional RNA library preparation protocol to use 96 Barcodes from BIOOScientific added by ligation, as described in [80]. Libraries from the same plate were pooled together and quantified using the KAPA Library Quantification Kit, following the manufacturer instructions and using a custom-made series of standards obtained from serial dilutions of the phi-X DNA (Illumina). Library pools were sequenced to an average of 9.5M 75bp PE reads. Within each pool, individual library concentrations were normalized and repooled to achieve comparable sequencing depths. Twelve treatments were selected for deep sequencing on the basis of the strong transcriptional response they provoked in at least one cell type, in addition to both controls. Samples selected for deep sequencing were pooled within each plate and further sequenced on the NovaSeq 6000 using 300bp PE reads. Each plate of LCLs and IPSCs were sequenced once on one lane for an average of 147M reads, and each plate of CMs was sequenced twice on one lane for an average

17

of 273M reads. The number of reads per library can be seen in Table S6.

Alignment of RNA-sequencing

RNA-sequencing reads were aligned to the human reference genome using HISAT2 [53] (`https://ccb.jhu.edu/software/hisat2/index.shtml`, version `hisat2-2.0.4`) with the following options:

`HISAT2 -x <genome> -1 <fastq_R1.gz> -2 <fastq_R2.gz>`

where `<genome>` represents the location of the genome file (`genome_snp_tran`), and `<fastqs_R1.gz>` and `<fastqs_R2.gz>` represent that sample's fastq files.

Multiple sequencing runs were merged for each sample using samtools (version 2.25.0). We removed PCR duplicates and further removed reads with a quality score of < 10 (equating to reads mapped to multiple locations). Shallow and deep RNA-sequencing reads were aligned in an identical manner except for the reference used. GRCh38 was used for the deeply sequenced data and GRCh37 for the low coverage data which is not used after the initial screening.

Differential gene expression analysis

To identify differentially expressed (DE) genes in shallow and deep sequencing data, we used `DESeq2` [66] (`R` version `3.5.2`, `DESeq2` version `1.28.1`). `coverageBed` was utilized to count reads in transcripts from the Ensembl gene annotation with `-s` to account for strandedness and `-split` for BED12 input. The counts were then utilized in DESeq2 to determine changes in gene expression under the different treatment conditions.

`Eq(1) Model 1: CellType + Treatment.ID`

Each treatment was compared to its relevant vehicle control, except for plate CM1R2, where we identified a technical problem with control 1, so we only used control 2 for this plate (Table S6). Note that comparison between the two controls yielded the expected finding of fewer than six differentially expressed genes for all other plates, thus confirming that the 2 controls are essentially technical replicates of the untreated condition. Multiple test correction was performed using the Benjamini-Hochberg procedure [11] with a significance threshold of 10%. A gene was considered a DEG if at least one of its transcripts was differentially expressed and had an absolute $\log_2$ fold change > 0.25. Each cell type was run independently, and the model corrected for a composite variable of library preparation batch and individual. Full DESeq2 results from the shallow sequencing can be found in Table S5.

After identifying DEGs from all 28 treatments in the initial shallow sequencing, we selected twelve to sequence more deeply. Nine of these twelve treatments were selected because they resulted in at least 60 DEGs in CMs. Selenium and Zinc were added because they induced a strong response in both IPSCs and LCLs, and Cadmium was added to complete the set of metal treatments. Full DESeq2 results from the deep sequencing can be found in Table S7.

To identify genes that show evidence of treatment $\times$ cell type interactions, we analyzed the deep sequencing data across all cell types and treatments using the following likelihood ratio test to compare Model 1 and Model 2:

`(Eq 2) Model 2: CellType + Treatment.ID + CellType:Treatment.ID`

Differential intron excision analysis

To detect shifts in splicing patterns across cell types and environments we used LeafCutter[63], an intron-based splicing analysis tool. Briefly, LeafCutter uses short RNA-seq reads spanning exon-exon junctions (i.e. split reads) to estimate internal introns usage, hence being able to ultimately infer all splicing events which can be summarized with differential introns excisions. Overlapping introns sharing a splice sites are then identified, which are subsequently used to construct a graph where the connected components represent clusters. Lastly, for each cluster the counts of the composing introns are jointly modelled with a Dirichlet-multinomial generalized linear model. We used the provided `bam2junc.sh` script to convert HISAT2-generated .bam files from the deep sequencing data to .junc files, as well as `leafcutter_cluster.py` script to perform intron clustering with the `-s yes -m 50 -l 500000` options (keep strand information, >50 split reads supporting each cluster and accepting introns with length up to 500kb). We removed clusters localizing on sex, mitochondrial and scaffold chromosomes, and used `leafcutter_ds.R` to perform differential intron excision analysis by contrasting each treatment with the appropriate control, with individuals (i.e. cell lines) as confounders.

Determining significance of global directional shifts in RNA processing

LeafCutter produces an output file in which, for every intron within a successfully tested cluster it computes the percentage spliced in index (Ψ) in the control and treatment condition, as well as the corresponding $\Delta\Psi$. LeafCutter exploits that retention and alternative excision of introns [13, 15] act as proxies of several subtypes of alternative splicing events, including exon skipping (ES), intron retention (IR), mutually-exclusive exons (MxE), alternative first and last exons (AFE/ALE), alternative 5' and 3' splice sites (A5SS/A3SS), among others. For every condition, we extracted introns belonging to clusters considered to be differentially spliced (FDR < 10%), and we performed a two-sided binomial test defining n as the number of introns with $\Delta\Psi > 0$, and p equal to the average proportion of positive events among all conditions. We considered a shift to be significant if the corresponding FDR < 10%.

Gene Ontology, KEGG, Disease-Gene Network enrichment analyses

We used ClusterProfiler [123] R package to perform Gene Ontology [7, 20], KEGG [51], and Disease-Gene Network [89] enrichment analyses. We considered two separate approaches for enrichment analysis. First, we calculated an enrichment of DEGs within each treatment per cell type against a background of all expressed genes in that treatment. This identified the pathways that were most enriched in DEGs for each treatment (Table S8). Second, we calculated an enrichment within each cell type across all environments to compare GO, KEGG and DGN enrichments using the compareCluster function by submitting the lists of differentially spliced or expressed (FDR < 10%) genes in each condition, as well as one background gene list comprising all tested genes in that given cell type. We considered a process to be significantly enriched if FDR < 10%.

For performing GO analysis on ASE variance, we first identified the top 20% and bottom 20% of genes based on their variance (high variance or low variance genes, respectively). We tested for pathway enrichments using Gene Ontology annotations comparing the high variance against the low variance genes using ClusterProfiler with the same thresholds as described above (Table S22).

Variance partitioning of gene expression and intron excision

We used the Variance Partition package [46] to partition the total variance in gene expression. For each gene, the tool determines the fraction of the observed variance explained by each variable by implementing

a linear-mixed model, ultimately allowing multiple dimensions of variation to be analyzed simultaneously. Variance which cannot be attributed to any of the provided covariates is counted as residuals, which can be generally interpreted as due to noise that can have technical or biological origin. We included all transcripts with at least 1 count per million reads (cpm) in 50% of the libraries, and the resulting data was voom-normalized [59, 98]. All variables included in the model were categorical, and were therefore modeled as random effects. For the model including all cell types, we used the `fitExtractVarPartModel` function with the following formula: `(1|Individual) + (1|CellType) + (1|Plate) + (1|Treatment)`. When each cell type was considered independently, we dropped CellType and added a Individual:Treatment variable, accounting for the interaction between these two covariates. We restricted this analysis to genes tested for both splicing and gene expression so we can compare the origins of variation in these two transcriptional processes.

For splicing, we merged all LeafCutter `.junc` files across every cell type and condition, where each column represents a sample and each row a different intron, in order to recover the full list of detected introns. We applied the suggested cutoff to extract introns considered to be expressed (on average, at least 1 cpm in half of the libraries), and used the cell type covariate to design the matrix eventually used by voom to normalize expression counts. Finally, we used the same methods as for gene expression to partition the variance.

Genotyping and ASE quantification with QuASAR

Using the common SNPs from dbSNP Build 144, we calculated the number of RNA-seq reads mapping to each allele using samtools mpileup. We inferred genotypes from these pileups using all samples from a given individual across all treatments and each plate together using QuASAR [45]. Counts for each pair of controls on the same plate were combined. For all subsequent analyses we focused on heterozygous SNPs with a read coverage greater than five and located on autosomes. Initial testing for ASE was performed for each sample separately using QuASAR. The amount of ASE identified via QuASAR aggregated by treatment is shown in Figure 3A, and the full output from QuASAR for tested SNPs can be found in Table S10

ANOVA for identifying ASE

Previous efforts for detecting ASE have commonly used binomial or beta-binomial tests [26, 45, 58]. Given that we have a large number of replicates and experimental conditions and to avoid excluding conditions where an allele may be lowly expressed, we tested for ASE using a linear model which incorporated ASE measurements across all conditions for a SNP in a given individual. To this end, we re-calculated ASE by adding a pseudocount to both reference and alternate read counts, and then computed the natural logarithm of the number of reference reads divided by the number of alternate reads. By doing so, we directly infer the noise inherent in ASE measurements using the linear model, obviating the need for binomial models to quantify the technical variance. To detect ASE in the model with all cell types combined, we selected 69,683 SNPs for which ASE was measured in at least five conditions per cell type, including at least one measurement for each control. We used an ANOVA test to compare a full model including the control ID, cell type, treatment, and cell type × treatment interaction term against a reduced model with the intercept set to 0:

```
(Eq 3) Full: ASE ~ Control + CellType + Treatment + CellType:Treatment
(Eq 4) Reduced: ASE ~ 0
```

We only consider this test for SNPs with a sufficient number of observations given the number of parameters being estimated (i.e., at least 5 degrees of freedom). Full results for the ANOVA model can be found in Table S11. For SNPs with ASE detected by the ANOVA model, we performed spearman correlations for every pair of libraries using all SNPs shared between them. Pairs of libraries were then placed into four categories based on their differences. These were: "Across Cell" (n = 968 pairs), which included all pairs of libraries from the same individual and treatment, but different cell types; "Across Individual" (n = 490 pairs), which included all pairs of libraries from the same treatment, cell type, and plate, but different individuals; "Across Replicate" (n = 242 pairs), which included all pairs of libraries from the same individual, cell type, and treatment, but different plate; and "Across Treatment" (n = 3146 pairs), which included all pairs of libraries from the same individual, cell type, and plate, but different treatment. Finally, the category for "Control 1 vs Control 2" (n = 33 pairs) includes correlations between the two control libraries from the same individual, cell type, and plate.

Fixed effect linear model for identifying conditional ASE

To identify SNPs displaying cASE due to cell type, treatment, or the interaction between cell type and treatment, we used a fixed effect linear model. All SNPs with significant ASE detected via the ANOVA model were tested for cASE. Additionally, SNPs must have ASE measured in at least one of each control for each cell type to be tested. We modeled the ASE values for each SNP in an individual across all conditions as a function of cell type, treatment, cell type-treatment interaction, and a variable representing the vehicle used for that treatment (Control):

```
(Eq 5) ASE ~ Control + CellType + Treatment + CellType:Treatment
```

Results can be found in Table S13, and are summarized in Table S14.

We also investigate ASE, and cASE for each cell type separately which may recover situations in which a gene is only expressed in one cell type. First we detect ASE for each cell type across any condition:

```
(Eq 6) Full: ASE ~ Control + Plate + Treatment
(Eq 7) Reduced: ASE ~ 0
```

Then, we test for each treatment component in the Full model to detect cASE. For the model with the cell types analyzed separately, ANOVA results can be found in Table S12, and results from the Full, fixed effect model (Eq 6), can be found in Table S15 and are summarized in Table S16

Genomic locations of ASE and cASE SNPs

Using the genomic annotations from dbSNP build 153, we assigned SNPs tested for ASE to one of 18 categories on the basis of their location in relation to a gene and predicted function. For each category, we calculated the enrichment of ASE and cASE SNPs separately relative to all SNPs tested for ASE and performed Fisher's exact test to test for significance. Multiple test correction of p-values was performed using the Benjamini-Hochberg procedure.

cASE overlap with TWAS, OMIM, and other G×E studies

To identify traits which are enriched for CM cell type cASE, we focused on 8 TWAS for cardiovascular traits from [124]. For each trait, we intersected all cell type cASE with TWAS genes, and calculated the proportion of overlap which is attributable to CM cell type cASE only. We used the one-sample proportion

test to determine if the CM cell type cASE overlap for a given trait was significantly different from the average trait overlap for CM cell type cASE.

To identify genes with treatment cASE that could influence complex traits, we intersected our cASE genes from the model with all cell types combined with genes whose expression has been implicated by transcriptome-wide association study (TWAS) to influence complex traits in 103 TWAS [124]. 45 TWAS had at least 20 cASE gene overlaps across all treatments, and those were selected for further analysis. As above for CM cell type cASE, we used the one-sample proportion test to determine if the treatment cASE overlap for a given trait was significantly different from the average overlap for that trait across all treatments.

$$P(t,c) = \frac{O(t,c)}{N(t)}$$

$$Z(t,c) = \frac{P(t,c) - P_0(c)}{\sqrt{P0(c) * (1 - P_0(c))/N(t)}}$$

Where t is the trait, c is the treatment, $O(t,c)$ is the number of genes which are significant for both the trait and the treatment, $N(t)$ is the sum of all genes for the trait which are treatment cASE in any treatment, and $P_0(c)$ is the average $P(t,c)$ across all traits.

For plotting purposes, we abbreviated the TWAS study names from Zhang *et al* [124] according to Table S23.

To identify genes responsible for Mendelian traits, we downloaded OMIM's Synopsis of the Human Gene Map (morbidmap.txt) on December 19, 2019 and intersected treatment cASE genes from the model containing only the CM data with OMIM genes. This resulted in 95 OMIM traits, which were manually curated to identify traits relevant to heart disease.

To investigate whether our G×E results replicate in other environments previously investigated, we calculated the overlap between genes with gene × treatment interactions in our dataset and genes with G×E (p < 0.05) identified in fourteen previous ASE and eQTL mapping studies [10, 19, 48, 50, 57, 58, 61, 70, 73, 81, 82, 91, 109, 125]. The full table of treatment cASE genes which replicated in these other studies can be found in Table S17.

Enrichment of ASE and cASE genes in GTEx eGenes

We used Fisher's exact test to test for an enrichment in ASE and cASE genes in GTEx eGenes. ASE and cASE genes were detected in CMs using the model that examined each cell type separately. We considered GTEx eGenes detected in heart tissues (left ventricle and atrial appendage), which are the relevant tissues for CMs. We restricted our analysis to include genes which had been tested by GTEx for being eGenes and had also been tested for ASE or cASE. GTEx v8 data was downloaded from `https://storage.googleapis.com/gtex_analysis_v8/single_tissue_qtl_data/GTEx_Analysis_v8_eQTL.tar`.

Heritability explained by cASE

To estimate the heritability explained by G×E, we used RHE-mc [87]. We performed two separate analyses using the UK Biobank data. First, we quantified the heritability of complex traits using an annotation of CM gene × cell type genes. Second, we quantified the heritability of complex traits using an annotation of gene × treatment genes. In both cases, SNPs were annotated to G×E genes within 100 Kb. For the gene × treatment analysis, we included both gene × treatment and gene × treatment × cell type genes, with each treatment forming a separate annotation.

The heritability partitioning was performed as described in [87]. Briefly, we excluded SNPs with greater than 1% missingness and minor allele frequency smaller than 0.1%. Further, we excluded SNPs that fail the Hardy-Weinberg test at significance threshold 10^{-7} as well as SNPs that lie within the MHC region (Chr6: 25–35 Mb) to obtain $7,774,235$ SNPs. We included age, sex, and the top 20 genetic principal components (PCs) as covariates in our analysis for all traits. We used PCs precomputed by the UK Biobank from a superset of $488,295$ individuals. Additional covariates were used for waist-to-hip ratio (adjusted for BMI) and diastolic/systolic blood pressure (adjusted for cholesterol-lowering medication, blood pressure medication,insulin, hormone replacement therapy, and oral contraceptives). For the CM gene × cell type annotation, we partitioned the heritability of 17 complex traits, while for the gene × treatment analysis, we expanded to 22 traits in the same sample from the UK Biobank. For each annotation, we computed the heritability enrichment as the ratio of the percentage of heritability explained to the percentage of SNPs in that annotation.

Mixed effect linear model for quantifying ASE variance

To quantify ASE variance, we used the same linear model as for identifying cASE, with the exception of modeling the cell type, treatment, cell type-treatment interaction, and batch variables as random effects:

(Eq 8) ASE ~ (1 | Control) + (1 | CellType) + (1 | Treatment) + (1 | CellType:Treatment)

As before, we only tested SNPs with significant ASE as determined by the ANOVA. To analyze each cell type independently, we used the following model:

(Eq 9) ASE ~ (1 | Control) + (1 | Batch) + (1 | Treatment)

Biological features of ASE variance

Residual variance is partially a function of sequencing depth, so for all analyses of variance components for ASE, at each SNP, we have adjusted for the total number of reads covering that position. To determine the biological significance of ASE variance, we identified biological features which contributed to differences in treatment or residual variance. Given the overwhelming contribution of cell type to ASE variance, we used the ASE variance calculated within each cell type. We tested for a relationship between residual and treatment variance and six annotations using the following model:

(Eq 10) Variance ~ Annotation + SNP expression

We assigned SNPs tested for ASE to the genes in which they reside, and considered the following sevengene annotations:

1. Differentially expressed genes: We categorized genes as 1) not differentially expressed in any condition; 2) differentially expressed in at least one condition (absolute value of logFC > 0.25 but < 0.75; or 3) highly differentially expressed (absolute value of logFC > 0.75) in at least one condition.

2. Differentially spliced genes: Genes were classified as being differentially spliced as described in the "Differential intron excision analysis" section.

3. Promoter architecture characteristics: Information on promoter characteristics, including presence of a TATA box, GC percentage, and CpG percentage were downloaded from refTSS [1]: `http://reftss.clst.riken.jp/datafiles/current/human/tata_box_annotations/hg38_tata_annotation_v3.txt.gz`

23

4. Loss of function tolerance: Loss of function statistics by gene were downloaded from Gnomad, version 2.1.1 (`https://storage.googleapis.com/gnomad-public/release/2.1.1/constraint/gnomad.v2.1.1.lof_metrics.by_gene.txt.bgz`). Specifically, we use the observed over expected ratio for predicted loss of function variants for a transcript (oe_lof). Lower scores indicate less tolerance to loss of function mutations. For example, haploinsufficient genes tend to have low oe_lof scores [52].

5. dN/dS ratio: dN/dS ratios were downloaded from Lindblad-Toh *et al* [64]. Ratios were derived from an analysis of 29 mammalian genomes. (`http://www.broadinstitute.org/ftp/pub/assemblies/mammals/29mammals/PositivelySelectedCodons.tar.gz`)

6. Cardiomyopathy genes: The 55 genes tested by the Mayo Clinic Laboratory Comprehensive Cardiomyopathy Multi-Gene Panel were considered as cardiomyopathy genes. For this category, we only tested ASE variance in CMs.

7. TWAS genes: All genes significant for at least one trait in the PTWAS analysis by [124] were considered TWAS genes.

Data and code availability

Sequencing files have been uploaded to the Sequence Read Archive (SRA) under Bioproject PRJNA694697. Code is available at `https://github.com/piquelab/GxExC`.

Additional Files

See supplementary tables and figures for additional results.

Acknowledgments

Competing Interests

The authors declare that they have no competing interests.

References

[1] Imad Abugessaisa, Shuhei Noguchi, Akira Hasegawa, Atsushi Kondo, Hideya Kawaji, Piero Carninci, and Takeya Kasukawa. refTSS: A Reference Data Set for Human and Mouse Transcription Start Sites. *Journal of Molecular Biology*, 431(13):2407–2422, jun 2019.

[2] Raul Aguirre-Gamboa, Irma Joosten, Paulo C.M. Urbano, Renate G. van der Molen, Esther van Rijssen, Bram van Cranenbroek, Marije Oosting, Sanne Smeekens, Martin Jaeger, Maria Zorro, Sebo Withoff, Antonius E. van Herwaarden, Fred C.G.J. Sweep, Romana T. Netea, Morris A. Swertz, Lude Franke, Ramnik J. Xavier, Leo A.B. Joosten, Mihai G. Netea, Cisca Wijmenga, Vinod Kumar, Yang Li, and Hans J.P.M. Koenen. Differential Effects of Environmental and Genetic Factors on T and B Cell Immune Traits. *Cell Reports*, 17(9):2474–2487, nov 2016.

[3] Kaur Alasoo, Julia Rodrigues, John Danesh, Daniel F. Freitag, Dirk S. Paul, and Daniel J. Gaffney. Genetic effects on promoter usage are highly context-specific and contribute to complex traits. *eLife*, 8, jan 2019.

[4] Kaur Alasoo, Julia Rodrigues, Subhankar Mukhopadhyay, Andrew J. Knights, Alice L. Mann, Kousik Kundu, Christine Hale, Gordon Dougan, and Daniel J. Gaffney. Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nature Genetics*, page 1, jan 2018.

[5] Abramowicz Anna and Gos Monika. Splicing mutations in human genetic disorders: examples, detection, and confirmation, aug 2018.

[6] Marios Arvanitis, Emmanouil Tampakakis, Yanxiao Zhang, Wei Wang, Adam Auton, Michelle Agee, Stella Aslibekyan, Robert K. Bell, Katarzyna Bryc, Sarah K. Clark, Sarah L. Elson, Kipper Fletez-Brant, Pierre Fontanillas, Nicholas A. Furlotte, Pooja M. Gandhi, Karl Heilbron, Barry Hicks, David A. Hinds, Karen E. Huber, Ethan M. Jewett, Yunxuan Jiang, Aaron Kleinman, Keng Han Lin, Nadia K. Litterman, Jennifer C. McCreight, Matthew H. McIntyre, Kimberly F. McManus, Joanna L. Mountain, Sahar V. Mozaffari, Priyanka Nandakumar, Elizabeth S. Noblin, Carrie A.M. Northover, Jared O'Connell, Steven J. Pitts, G. David Poznik, J. Fah Sathirapongsasuti, Anjali J. Shastri, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Robert J. Tunney, Vladimir Vacic, Xin Wang, Amir S. Zare, Diptavo Dutta, Stephanie Glavaris, Ali Keramati, Nilanjan Chatterjee, Neil C. Chi, Bing Ren, Wendy S. Post, and Alexis Battle. Genome-wide association and multi-omic analyses reveal ACTN2 as a gene linked to heart failure. *Nature Communications*, 11(1), dec 2020.

[7] Michael Ashburner, Catherine A. Ball, Judith A. Blake, David Botstein, Heather Butler, J. Michael Cherry, Allan P. Davis, Kara Dolinski, Selina S. Dwight, Janan T. Eppig, Midori A. Harris, David P. Hill, Laurie Issel-Tarver, Andrew Kasarskis, Suzanna Lewis, John C. Matese, Joel E. Richardson, Martin Ringwald, Gerald M. Rubin, and Gavin Sherlock. Gene ontology: Tool for the unification of biology, may 2000.

[8] No authors listed. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, 369(6509):1318–1330, 09 2020.

[9] Nicholas E Banovich, Yang I Li, Anil Raj, Michelle C Ward, Peyton Greenside, Diego Calderon, Po Yuan Tung, Jonathan E Burnett, Marsha Myrthil, Samantha M Thomas, Courtney K Burrows, Irene Gallego Romero, Bryan J Pavlovic, Anshul Kundaje, Jonathan K Pritchard, and Yoav Gilad. Impact of regulatory variation across human iPSCs and differentiated cells. *Genome Research*, 28(1):122–131, jan 2018.

[10] Luis B Barreiro, Ludovic Tailleux, Athma A Pai, Brigitte Gicquel, John C Marioni, and Yoav Gilad. Deciphering the genetic architecture of variation in the immune response to Mycobacterium

tuberculosis infection. *Proceedings of the National Academy of Sciences of the United States of America*, 109(4):1204–9, jan 2012.

[11] Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing, 1995.

[12] Joseph L. Bobadilla, Milan Macek, Jason P. Fine, and Philip M. Farrell. Cystic fibrosis: A worldwide analysis of CFTR mutations - Correlation with incidence data and application to screening, 2002.

[13] Paul L. Boutz, Arjun Bhutkar, and Phillip A. Sharp. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes and Development*, 29(1):63–80, jan 2015.

[14] Evan A. Boyle, Yang I. Li, and Jonathan K. Pritchard. An Expanded View of Complex Traits: From Polygenic to Omnigenic, jun 2017.

[15] Ulrich Braunschweig, Nuno L. Barbosa-Morais, Qun Pan, Emil N. Nachman, Babak Alipanahi, Thomas Gonatopoulos-Pournatzis, Brendan Frey, Manuel Irimia, and Benjamin J. Blencowe. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Research*, 24(11):1774–1786, nov 2014.

[16] Saverio Brogna and Jikai Wen. Nonsense-mediated mRNA decay (NMD) mechanisms, feb 2009.

[17] Paul W Burridge, Yong Fuga Li, Elena Matsa, Haodi Wu, Sang-Ging Ong, Arun Sharma, Alexandra Holmström, Alex C Chang, Michael J Coronado, Antje D Ebert, Joshua W Knowles, Melinda L Telli, Ronald M Witteles, Helen M Blau, Daniel Bernstein, Russ B Altman, and Joseph C Wu. Human induced pluripotent stem cell-derived cardiomyocytes recapitulate the predilection of breast cancer patients to doxorubicin-induced cardiotoxicity. *Nature Medicine*, 22(5):547–556, 2016.

[18] Courtney K. Burrows, Nicholas E. Banovich, Bryan J. Pavlovic, Kristen Patterson, Irene Gallego Romero, Jonathan K. Pritchard, and Yoav Gilad. Genetic Variation, Not Cell Type of Origin, Underlies the Majority of Identifiable Regulatory Differences in iPSCs. *PLOS Genetics*, 12(1):e1005793, jan 2016.

[19] Minal Çalışkan, Samuel W. Baker, Yoav Gilad, and Carole Ober. Host Genetic Variation Influences Gene Expression Response to Rhinovirus Infection. *PLOS Genetics*, 11(4):e1005111, apr 2015.

[20] S. Carbon, E. Douglass, N. Dunn, B. Good, N. L. Harris, S. E. Lewis, C. J. Mungall, S. Basu, R. L. Chisholm, R. J. Dodson, E. Hartline, P. Fey, P. D. Thomas, L. P. Albou, D. Ebert, M. J. Kesling, H. Mi, A. Muruganujan, X. Huang, S. Poudel, T. Mushayahama, J. C. Hu, S. A. LaBonte, D. A. Siegele, G. Antonazzo, H. Attrill, N. H. Brown, S. Fexova, P. Garapati, T. E.M. Jones, S. J. Marygold, G. H. Millburn, A. J. Rey, V. Trovisco, G. Dos Santos, D. B. Emmert, K. Falls, P. Zhou, J. L. Goodman, V. B. Strelets, J. Thurmond, M. Courtot, D. S. Osumi, H. Parkinson, P. Roncaglia, M. L. Acencio, M. Kuiper, A. Lreid, C. Logie, R. C. Lovering, R. P. Huntley, P. Denny, N. H. Campbell, B. Kramarz, V. Acquaah, S. H. Ahmad, H. Chen, J. H. Rawson, M. C. Chibucos, M. Giglio, S. Nadendla, R. Tauber, M. J. Duesbury, N. T. Del, B. H.M. Meldal, L. Perfetto, P. Porras, S. Orchard, A. Shrivastava, Z. Xie, H. Y. Chang, R. D. Finn, A. L. Mitchell, N. D. Rawlings, L. Richardson, A. Sangrador-Vegas, J. A. Blake, K. R. Christie, M. E. Dolan, H. J. Drabkin, D. P. Hill, L. Ni, D. Sitnikov, M. A. Harris, S. G. Oliver, K. Rutherford, V. Wood, J. Hayles, J. Bahler, A. Lock, E. R. Bolton, J. De Pons, M. Dwinell, G. T. Hayman, S. J.F. Laulederkind, M. Shimoyama, M. Tutaj, S. J. Wang, P. D'Eustachio, L. Matthews, J. P. Balhoff, S. A. Aleksander, G. Binkley,

B. L. Dunn, J. M. Cherry, S. R. Engel, F. Gondwe, K. Karra, K. A. MacPherson, S. R. Miyasato, R. S. Nash, P. C. Ng, T. K. Sheppard, A. Shrivatsav Vp, M. Simison, M. S. Skrzypek, S. Weng, E. D. Wong, M. Feuermann, P. Gaudet, E. Bakker, T. Z. Berardini, L. Reiser, S. Subramaniam, E. Huala, C. Arighi, A. Auchincloss, K. Axelsen, G. P. Argoud, A. Bateman, B. Bely, M. C. Blatter, E. Boutet, L. Breuza, A. Bridge, R. Britto, H. Bye-A-Jee, C. Casals-Casas, E. Coudert, A. Estreicher, L. Famiglietti, P. Garmiri, G. Georghiou, A. Gos, N. Gruaz-Gumowski, E. Hatton-Ellis, U. Hinz, C. Hulo, A. Ignatchenko, F. Jungo, G. Keller, K. Laiho, P. Lemercier, D. Lieberherr, Y. Lussi, A. Mac-Dougall, M. Magrane, M. J. Martin, P. Masson, D. A. Natale, N. N. Hyka, I. Pedruzzi, K. Pichler, S. Poux, C. Rivoire, M. Rodriguez-Lopez, T. Sawford, E. Speretta, A. Shypitsyna, A. Stutz, S. Sundaram, M. Tognolli, N. Tyagi, K. Warner, R. Zaru, C. Wu, J. Chan, J. Cho, S. Gao, C. Grove, M. C. Harrison, K. Howe, R. Lee, J. Mendel, H. M. Muller, D. Raciti, K. Van Auken, M. Berriman, L. Stein, P. W. Sternberg, D. Howe, S. Toro, and M. Westerfield. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, 47(D1):D330–D338, jan 2019.

[21] Piero Carninci, Albin Sandelin, Boris Lenhard, Shintaro Katayama, Kazuro Shimokawa, Jasmina Ponjavic, Colin A.M. Semple, Martin S. Taylor, Pär G. Engström, Martin C. Frith, Alistair R.R. Forrest, Wynand B. Alkema, Sin Lam Tan, Charles Plessy, Rimantas Kodzius, Timothy Ravasi, Takeya Kasukawa, Shiro Fukuda, Mutsumi Kanamori-Katayama, Yayoi Kitazume, Hideya Kawaji, Chikatoshi Kai, Mari Nakamura, Hideaki Konno, Kenji Nakano, Salim Mottagui-Tabar, Peter Arner, Alessandra Chesi, Stefano Gustincich, Francesca Persichetti, Harukazu Suzuki, Sean M. Grimmond, Christine A. Wells, Valerio Orlando, Claes Wahlestedt, Edison T. Liu, Matthias Harbers, Jun Kawai, Vladimir B. Bajic, David A. Hume, and Yoshihide Hayashizaki. Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics*, 38(6):626–635, jun 2006.

[22] Xonia Carvajal-Vergara, Ana Sevilla, Sunita L. Dsouza, Yen Sin Ang, Christoph Schaniel, Dung Fang Lee, Lei Yang, Aaron D. Kaplan, Eric D. Adler, Roye Rozov, Yongchao Ge, Ninette Cohen, Lisa J. Edelmann, Betty Chang, Avinash Waghray, Jie Su, Sherly Pardo, Klaske D. Lichtenbelt, Marco Tartaglia, Bruce D. Gelb, and Ihor R. Lemischka. Patient-specific induced pluripotent stem-cell-derived models of LEOPARD syndrome. *Nature*, 465(7299):808–812, jun 2010.

[23] Christian Fynbo Christiansen, Steffen Christensen, Frank Mehnert, Steven R. Cummings, Roland D. Chapurlat, and Henrik Toft Sørensen. Glucocorticoid use and risk of atrial fibrillation or flutter: A population-based, case-control study. *Archives of Internal Medicine*, 169(18):1677–1683, oct 2009.

[24] Anna S.E. Cuomo, Daniel D. Seaton, Davis J. McCarthy, Iker Martinez, Marc Jan Bonder, Jose Garcia-Bernardo, Shradha Amatya, Pedro Madrigal, Abigail Isaacson, Florian Buettner, Andrew Knights, Kedar Nath Natarajan, Chukwuma A. Agu, Alex Alderton, Petr Danecek, Rachel Denton, Richard Durbin, Daniel J. Gaffney, Angela Goncalves, Reena Halai, Sarah Harper, Christopher M. Kirton, Anja Kolb-Kokocinski, Andreas Leha, Shane A. McCarthy, Yasin Memari, Minal Patel, Ewan Birney, Francesco Paolo Casale, Laura Clarke, Peter W. Harrison, Helena Kilpinen, Ian Streeter, Davide Denovi, Ruta Meleckyte, Natalie Moens, Fiona M. Watt, Willem H. Ouwehand, Angus I. Lamond, Dalila Bensaddek, Philip Beales, Ludovic Vallier, John C. Marioni, Mariya Chhatriwala, and Oliver Stegle. Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. *Nature Communications*, 11(1):810, dec 2020.

[25] Christopher DeBoever, He Li, David Jakubosky, Paola Benaglio, Joaquin Reyna, Katrina M. Olson, Hui Huang, William Biggs, Efren Sandoval, Matteo D'Antonio, Kristen Jepsen, Hiroko Matsui, Angelo Arias, Bing Ren, Naoki Nariai, Erin N. Smith, Agnieszka D'Antonio-Chronowska, Emma K.

Farley, and Kelly A. Frazer. Large-Scale Profiling Reveals the Influence of Genetic Variation on Gene Expression in Human Induced Pluripotent Stem Cells. *Cell Stem Cell*, 20(4):533–546.e7, apr 2017.

[26] Jacob F. Degner, Athma A. Pai, Roger Pique-Regi, Jean-Baptiste Veyrieras, Daniel J. Gaffney, Joseph K. Pickrell, Sherryl De Leon, Katelyn Michelini, Noah Lewellen, Gregory E. Crawford, Matthew Stephens, Yoav Gilad, and Jonathan K. Pritchard. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature*, 482(7385):390–394, feb 2012.

[27] Antigone S Dimas, Samuel Deutsch, Barbara E Stranger, Stephen B Montgomery, Christelle Borel, Homa Attar-Cohen, Catherine Ingle, Claude Beazley, Maria Gutierrez Arcelus, Magdalena Sekowska, Marilyne Gagnebin, James Nisbett, Panos Deloukas, Emmanouil T Dermitzakis, and Stylianos E Antonarakis. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science (New York, N.Y.)*, 325:1246–1250, 2009.

[28] James J. Dinicolantonio, Dennis Mangan, and James H. O'Keefe. Copper deficiency may be a leading cause of ischaemic heart disease. *Open Heart*, 5(2):784, oct 2018.

[29] Benjamin P Fairfax, Peter Humburg, Seiko Makino, Vivek Naranbhai, Daniel Wong, Evelyn Lau, Luke Jostins, Katharine Plant, Robert Andrews, Chris McGee, and Julian C Knight. Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science (New York, N.Y.)*, 343(6175):1246949, mar 2014.

[30] Nuno André Faustino and Thomas A. Cooper. Pre-mRNA splicing and human disease, feb 2003.

[31] Marie Julie Favé, Fabien C. Lamaze, David Soave, Alan Hodgkinson, Héloïse Gauvin, Vanessa Bruat, Jean Christophe Grenier, Elias Gbeha, Kimberly Skead, Audrey Smargiassi, Markey Johnson, Youssef Idaghdour, and Philip Awadalla. Gene-by-environment interactions in urban populations modulate risk phenotypes. *Nature Communications*, 9(1):1–12, dec 2018.

[32] Anthony S. Findley, Allison L. Richards, Cristiano Petrini, Adnan Alazizi, Elizabeth Doman, Alexander G. Shanku, Gordon O. Davis, Nancy Hauff, Yoram Sorokin, Xiaoquan Wen, Roger Pique-Regi, and Francesca Luca. Interpreting coronary artery disease risk through gene-environment interactions in gene regulation. *Genetics*, 213(2):651–663, oct 2019.

[33] Hilary K. Finucane, Brendan Bulik-Sullivan, Alexander Gusev, Gosia Trynka, Yakir Reshef, Po Ru Loh, Verneri Anttila, Han Xu, Chongzhi Zang, Kyle Farh, Stephan Ripke, Felix R. Day, Shaun Purcell, Eli Stahl, Sara Lindstrom, John R.B. Perry, Yukinori Okada, Soumya Raychaudhuri, Mark J. Daly, Nick Patterson, Benjamin M. Neale, and Alkes L. Price. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, 47(11):1228–1235, nov 2015.

[34] Timothée Flutre, Xiaoquan Wen, Jonathan Pritchard, and Matthew Stephens. A Statistical Framework for Joint eQTL Analysis in Multiple Tissues. *PLoS Genetics*, 9(5):e1003486, may 2013.

[35] Malika Kumar Freund, Kathryn S. Burch, Huwenbo Shi, Nicholas Mancuso, Gleb Kichaev, Kristina M. Garske, David Z. Pan, Zong Miao, Karen L. Mohlke, Markku Laakso, Päivi Pajukanta, Bogdan Pasaniuc, and Valerie A. Arboleda. Phenotype-Specific Enrichment of Mendelian Disorder Genes near GWAS Regions across 62 Complex Traits. *American Journal of Human Genetics*, 103(4):535–552, oct 2018.

[36] Kenneth J. Friedman, Jolanta Kole, Jonathan A. Cohn, Michael R. Knowles, Lawrence M. Silverman, and Ryszard Kole. Correction of aberrant splicing of the cystic fibrosis transmembrane conductance regulator (CFTR) gene by antisense oligonucleotides. *Journal of Biological Chemistry*, 274(51):36193–36199, dec 1999.

[37] Joana M. Gaspar and Lício A. Velloso. Hypoxia inducible factor as a central regulator of metabolism - implications for the development of obesity, nov 2018.

[38] Greg Gibson. The environmental contribution to gene expression profiles, aug 2008.

[39] Thomas Gonatopoulos-Pournatzis, Mingkun Wu, Ulrich Braunschweig, Jonathan Roth, Hong Han, Andrew J. Best, Bushra Raj, Michael Aregger, Dave O'Hanlon, Jonathan D. Ellis, John A. Calarco, Jason Moffat, Anne Claude Gingras, and Benjamin J. Blencowe. Genome-wide CRISPR-Cas9 Interrogation of Splicing Networks Reveals a Mechanism for Recognition of Autism-Misregulated Neuronal Microexons. *Molecular Cell*, 72(3):510–524.e12, nov 2018.

[40] Fang Guo, Hui He, Zhi Chao Fu, Shengping Huang, Tingtao Chen, Christopher J. Papasian, Leslie R. Morse, Yan Xu, Ricardo A. Battaglino, Xiao Feng Yang, Zhisheng Jiang, Hong Bo Xin, and Mingui Fu. Adipocyte-derived PAMM suppresses macrophage inflammation by inhibiting MAPK signalling. *Biochemical Journal*, 472(3):309–318, dec 2015.

[41] Y. Habara, Y. Takeshima, H. Awano, Y. Okizuka, Z. Zhang, K. Saiki, M. Yagi, and M. Matsuo. In vitro splicing analysis showed that availability of a cryptic splice site is not a determinant for alternative splicing patterns caused by +1G–¿A mutations in introns of the dystrophin gene. *Journal of Medical Genetics*, 46(8):542–547, aug 2009.

[42] Ada Hamosh, Alan F. Scott, Joanna S. Amberger, Carol A. Bocchini, and Victor A. McKusick. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Research*, 33(DATABASE ISS.), jan 2005.

[43] Hong Han, Ulrich Braunschweig, Thomas Gonatopoulos-Pournatzis, Robert J. Weatheritt, Calley L. Hirsch, Kevin C.H. Ha, Ernest Radovani, Syed Nabeel-Shah, Tim Sterne-Weiler, Juli Wang, Dave O'Hanlon, Qun Pan, Debashish Ray, Hong Zheng, Frederick Vizeacoumar, Alessandro Datti, Lilia Magomedova, Carolyn L. Cummins, Timothy R. Hughes, Jack F. Greenblatt, Jeffrey L. Wrana, Jason Moffat, and Benjamin J. Blencowe. Multilayered Control of Alternative Splicing Regulatory Networks by Transcription Factors. *Molecular Cell*, 65(3):539–553.e7, feb 2017.

[44] Masahide Harada, David R. Van Wagoner, and Stanley Nattel. Role of inflammation in Atrial fibrillation pathophysiology and management, 2015.

[45] C.T. Chris T Harvey, G.A. Gregory A Moyerbrailean, Gordon O G.O. Davis, Xiaoquan Wen, Francesca Luca, and Roger Pique-Regi. QuASAR: quantitative allele-specific analysis of reads. *Bioinformatics (Oxford, England)*, 31(8):1235–42, apr 2015.

[46] Gabriel E Hoffman and Eric E Schadt. variancePartition: interpreting drivers of variation in complex gene expression studies. *BMC Bioinformatics*, 17(1):483, dec 2016.

[47] Yu Feng Hu, Yi Jen Chen, Yenn Jiang Lin, and Shih Ann Chen. Inflammation and the pathogenesis of atrial fibrillation, apr 2015.

[48] Qin Qin Huang, Howard Tang, Shu Mei Teo, Scott Ritchie, Artika Nath, Marta Brozynska, Agus Salim, Andrew Bakshi, Barbara Holt, Danny Mok, Chiea Chuen Khor, Peter Sly, Patrick Holt, Kathryn Holt, and Michael Inouye. Neonatal genetics of gene expression reveal the origins of autoimmune and allergic disease risk. *bioRxiv*, page 683086, jun 2019.

[49] Youssef Idaghdour, Wendy Czika, Kevin V. Shianna, Sang H. Lee, Peter M. Visscher, Hilary C. Martin, Kelci Miclaus, Sami J. Jadallah, David B. Goldstein, Russell D. Wolfinger, and Greg Gibson. Geographical genomics of human leukocyte gene expression variation in southern Morocco. *Nature Genetics*, 42(1):62–67, jan 2010.

[50] Youssef Idaghdour, Jacklyn Quinlan, Jean Philippe Goulet, Joanne Berghout, Elias Gbeha, Vanessa Bruat, Thibault De Malliard, Jean Christophe Grenier, Selma Gomez, Philippe Gros, Mohamed Cheŕif Rahimy, Ambaliou Sanni, and Philip Awadalla. Evidence for additive and interaction effects of host genotype and infection in malaria. *Proceedings of the National Academy of Sciences of the United States of America*, 109(42):16786–16793, oct 2012.

[51] Wixon J and Kell D. The Kyoto Encyclopedia of Genes and Genomes—KEGG. *Yeast*, 1(1):48–55, 2000.

[52] Konrad J. Karczewski, Laurent C. Francioli, Grace Tiao, Beryl B. Cummings, Jessica Alföldi, Qingbo Wang, Ryan L. Collins, Kristen M. Laricchia, Andrea Ganna, Daniel P. Birnbaum, Laura D. Gauthier, Harrison Brand, Matthew Solomonson, Nicholas A. Watts, Daniel Rhodes, Moriel Singer-Berk, Eleina M. England, Eleanor G. Seaby, Jack A. Kosmicki, Raymond K. Walters, Katherine Tashman, Yossi Farjoun, Eric Banks, Timothy Poterba, Arcturus Wang, Cotton Seed, Nicola Whiffin, Jessica X. Chong, Kaitlin E. Samocha, Emma Pierce-Hoffman, Zachary Zappala, Anne H. O'Donnell-Luria, Eric Vallabh Minikel, Ben Weisburd, Monkol Lek, James S. Ware, Christopher Vittal, Irina M. Armean, Louis Bergelson, Kristian Cibulskis, Kristen M. Connolly, Miguel Covarrubias, Stacey Donnelly, Steven Ferriera, Stacey Gabriel, Jeff Gentry, Namrata Gupta, Thibault Jeandet, Diane Kaplan, Christopher Llanwarne, Ruchi Munshi, Sam Novod, Nikelle Petrillo, David Roazen, Valentin Ruano-Rubio, Andrea Saltzman, Molly Schleicher, Jose Soto, Kathleen Tibbetts, Charlotte Tolonen, Gordon Wade, Michael E. Talkowski, Carlos A. Aguilar Salinas, Tariq Ahmad, Christine M. Albert, Diego Ardissino, Gil Atzmon, John Barnard, Laurent Beaugerie, Emelia J. Benjamin, Michael Boehnke, Lori L. Bonnycastle, Erwin P. Bottinger, Donald W. Bowden, Matthew J. Bown, John C. Chambers, Juliana C. Chan, Daniel Chasman, Judy Cho, Mina K. Chung, Bruce Cohen, Adolfo Correa, Dana Dabelea, Mark J. Daly, Dawood Darbar, Ravindranath Duggirala, Josée Dupuis, Patrick T. Ellinor, Roberto Elosua, Jeanette Erdmann, Tõnu Esko, Martti Färkkilä, Jose Florez, Andre Franke, Gad Getz, Benjamin Glaser, Stephen J. Glatt, David Goldstein, Clicerio Gonzalez, Leif Groop, Christopher Haiman, Craig Hanis, Matthew Harms, Mikko Hiltunen, Matti M. Holi, Christina M. Hultman, Mikko Kallela, Jaakko Kaprio, Sekar Kathiresan, Bong Jo Kim, Young Jin Kim, George Kirov, Jaspal Kooner, Seppo Koskinen, Harlan M. Krumholz, Subra Kugathasan, Soo Heon Kwak, Markku Laakso, Terho Lehtimäki, Ruth J.F. Loos, Steven A. Lubitz, Ronald C.W. Ma, Daniel G. MacArthur, Jaume Marrugat, Kari M. Mattila, Steven McCarroll, Mark I. McCarthy, Dermot McGovern, Ruth McPherson, James B. Meigs, Olle Melander, Andres Metspalu, Benjamin M. Neale, Peter M. Nilsson, Michael C. O'Donovan, Dost Ongur, Lorena Orozco, Michael J. Owen, Colin N.A. Palmer, Aarno Palotie, Kyong Soo Park, Carlos Pato, Ann E. Pulver, Nazneen Rahman, Anne M. Remes, John D. Rioux, Samuli Ripatti, Dan M. Roden, Danish Saleheen, Veikko Salomaa, Nilesh J. Samani, Jeremiah Scharf, Heribert Schunkert, Moore B. Shoemaker, Pamela Sklar, Hilkka Soininen, Harry Sokol, Tim Spector, Patrick F. Sullivan, Jaana Suvisaari, E. Shyong Tai, Yik Ying Teo, Tuomi

Tiinamaija, Ming Tsuang, Dan Turner, Teresa Tusie-Luna, Erkki Vartiainen, James S. Ware, Hugh Watkins, Rinse K. Weersma, Maija Wessman, James G. Wilson, Ramnik J. Xavier, Benjamin M. Neale, Mark J. Daly, and Daniel G. MacArthur. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809):434–443, may 2020.

[53] Daehwan Kim, Ben Langmead, and Steven L Salzberg. HISAT: a fast spliced aligner with low memory requirements. *Nature methods*, 12(4):357–60, apr 2015.

[54] Sarah Kim-Hellmuth, Matthias Bechheim, Benno Pütz, Pejman Mohammadi, Yohann Nédélec, Nicholas Giangreco, Jessica Becker, Vera Kaiser, Nadine Fricker, Esther Beier, Peter Boor, Stephane E. Castel, Markus M. Nöthen, Luis B. Barreiro, Joseph K. Pickrell, Bertram Müller-Myhsok, Tuuli Lappalainen, Johannes Schumacher, and Veit Hornung. Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nature Communications*, 8(1):1–10, dec 2017.

[55] Ryosuke Kita and Hunter B. Fraser. Local Adaptation of Sun-Exposure-Dependent Gene Expression Regulation in Human Skin. *PLoS Genetics*, 12(10):1–18, 2016.

[56] Tomoya Kitani, Sang-Ging Ong, Chi Keung Lam, June-Wha Rhee, Joe Z. Zhang, Angelos Oikonomopoulos, Ning Ma, Lei Tian, Jaecheol Lee, Melinda L. Telli, Ronald M. Witteles, Arun Sharma, Nazish Sayed, and Joseph C. Wu. Human-Induced Pluripotent Stem Cell Model of Trastuzumab-Induced Cardiac Dysfunction in Patients With Breast Cancer. *Circulation*, 139(21):2451–2465, may 2019.

[57] David A Knowles, Courtney K Burrows, John D Blischak, Kristen M Patterson, Daniel J Serie, Nadine Norton, Carole Ober, Jonathan K Pritchard, and Yoav Gilad. Determining the genetic basis of anthracycline-cardiotoxicity by molecular response QTL mapping in induced cardiomyocytes. *eLife*, 7, may 2018.

[58] David A Knowles, Joe R Davis, Hilary Edgington, Anil Raj, Marie-Julie Favé, Xiaowei Zhu, James B Potash, Myrna M Weissman, Jianxin Shi, Douglas F Levinson, Philip Awadalla, Sara Mostafavi, Stephen B Montgomery, and Alexis Battle. Allele-specific expression reveals interactions between genetic variation and environment. *Nature methods*, 14(7):699–702, jul 2017.

[59] Charity W Law, Yunshun Chen, Wei Shi, and Gordon K Smyth. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology*, 15(2):R29, feb 2014.

[60] Amanda Lea, Meena Subramaniam, Arthur Ko, Terho Lehtimäki, Emma Raitoharju, Mika Kähönen, Ilkka Seppälä, Nina Mononen, Olli T. Raitakari, Mika Ala-Korpela, Päivi Pajukanta, Noah Zaitlen, and Julien F. Ayroles. Genetic and environmental perturbations lead to regulatory decoherence. *eLife*, 8, mar 2019.

[61] Michelle N. Lee, Chun Ye, A.-C. Villani, Towfique Raj, Weibo Li, Thomas M Eisenhaure, Selina H Imboywa, Portia I Chipendo, F Ann Ran, Kamil Slowikowski, Lucas D Ward, Khadir Raddassi, Cristin McCabe, Michelle H Lee, Irene Y Frohlich, David a Hafler, Manolis Kellis, Soumya Raychaudhuri, Feng Zhang, Barbara E Stranger, Christophe O Benoist, Philip L De Jager, Aviv Regev, and Nir Hacohen. Common Genetic Variants Modulate Pathogen-Sensing Responses in Human Dendritic Cells. *Science*, 343(6175):1246980–1246980, mar 2014.

[62] Ben Lehner. Selection to minimise noise in living systems and its implications for the evolution of gene expression. *Molecular Systems Biology*, 4(1):170, jan 2008.

[63] Yang I. Li, David A. Knowles, Jack Humphrey, Alvaro N. Barbeira, Scott P. Dickinson, Hae Kyung Im, and Jonathan K. Pritchard. Annotation-free quantification of RNA splicing using LeafCutter. *Nature Genetics*, 50(1):151–158, jan 2018.

[64] Kerstin Lindblad-Toh, Manuel Garber, Or Zuk, Michael F. Lin, Brian J. Parker, Stefan Washietl, Pouya Kheradpour, Jason Ernst, Gregory Jordan, Evan Mauceli, Lucas D. Ward, Craig B. Lowe, Alisha K. Holloway, Michele Clamp, Sante Gnerre, Jessica Alföldi, Kathryn Beal, Jean Chang, Hiram Clawson, James Cuff, Federica Di Palma, Stephen Fitzgerald, Paul Flicek, Mitchell Guttman, Melissa J. Hubisz, David B. Jaffe, Irwin Jungreis, W. James Kent, Dennis Kostka, Marcia Lara, Andre L. Martins, Tim Massingham, Ida Moltke, Brian J. Raney, Matthew D. Rasmussen, Jim Robinson, Alexander Stark, Albert J. Vilella, Jiayu Wen, Xiaohui Xie, Michael C. Zody, Kim C. Worley, Christie L. Kovar, Donna M. Muzny, Richard A. Gibbs, Wesley C. Warren, Elaine R. Mardis, George M. Weinstock, Richard K. Wilson, Ewan Birney, Elliott H. Margulies, Javier Herrero, Eric D. Green, David Haussler, Adam Siepel, Nick Goldman, Katherine S. Pollard, Jakob S. Pedersen, Eric S. Lander, Manolis Kellis, Jen Baldwin, Toby Bloom, Chee Whye Chin, Dave Heiman, Robert Nicol, Chad Nusbaum, Sarah Young, Jane Wilkinson, Andrew Cree, Huyen H. Dihn, Gerald Fowler, Shalili Jhangiani, Vandita Joshi, Sandra Lee, Lora R. Lewis, Lynne V. Nazareth, Geoffrey Okwuonu, Jireh Santibanez, Kim Delehaunty, David Dooling, Catrina Fronik, Lucinda Fulton, Bob Fulton, Tina Graves, Patrick Minx, and Erica Sodergren. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature*, 478(7370):476–482, oct 2011.

[65] Magnus P. Lindgren, Mir Nabi Pirouzi Fard, J. Gustav Smith, Jan Sundquist, Kristina Sundquist, and Bengt Zöller. A swedish nationwide adoption study of the heritability of heart failure. *JAMA Cardiology*, 3(8):703–710, aug 2018.

[66] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550, dec 2014.

[67] Kiven E. Lukong, Kai wei Chang, Edouard W. Khandjian, and Stéphane Richard. RNA-binding proteins in human genetic disease, aug 2008.

[68] Zhila Maghbooli, Arash Hossein-nezhad, Elham Adabi, Effat Asadollah-pour, Mahsa Sadeghi, Sara Mohammad-nabi, Leila Zakeri Rad, Ali-asghar Malek Hosseini, Mehrnaz Radmehr, Fatemeh Faghihi, Atoosa Aghaei, Abolfazl Omidifar, Yasaman Aghababaei, and Hadis Behzadi. Air pollution during pregnancy and placental adaptation in the levels of global DNA methylation. *PLOS ONE*, 13(7):e0199772, jul 2018.

[69] M. Mandl, M. K. Lieberum, and R. Depping. A HIF-1α-driven feed-forward loop augments HIF signalling in Hep3B cells by upregulation of ARNT. *Cell Death and Disease*, 7(6):e2284–e2284, jun 2016.

[70] Lara M Mangravite, Barbara E Engelhardt, Marisa W Medina, Joshua D Smith, Christopher D Brown, Daniel I Chasman, Brigham H Mecham, Bryan Howie, Heejung Shim, Devesh Naidoo, QiPing Feng, Mark J Rieder, Yii-Der I Chen, Jerome I Rotter, Paul M Ridker, Jemma C Hopewell, Sarah Parish, Jane Armitage, Rory Collins, Russell a Wilke, Deborah a Nickerson, Matthew Stephens, and Ronald M Krauss. A statin-dependent QTL for GATM expression is associated with statin-induced myopathy. *Nature*, 502(7471):377–80, 2013.

[71] Teri A. Manolio, Francis S. Collins, Nancy J. Cox, David B. Goldstein, Lucia A. Hindorff, David J. Hunter, Mark I. McCarthy, Erin M. Ramos, Lon R. Cardon, Aravinda Chakravarti, Judy H. Cho,

Alan E. Guttmacher, Augustine Kong, Leonid Kruglyak, Elaine Mardis, Charles N. Rotimi, Montgomery Slatkin, David Valle, Alice S. Whittemore, Michael Boehnke, Andrew G. Clark, Evan E. Eichler, Greg Gibson, Jonathan L. Haines, Trudy F.C. MacKay, Steven A. McCarroll, and Peter M. Visscher. Finding the missing heritability of complex diseases, oct 2009.

[72] Jérémy Manry, Yohann Nédélec, Vinicius M. Fava, Aurélie Cobat, Marianna Orlova, Nguyen Van Thuc, Vu Hong Thai, Guillaume Laval, Luis B. Barreiro, and Erwin Schurr. Deciphering the genetic control of gene expression following Mycobacterium leprae antigen stimulation. *PLOS Genetics*, 13(8):e1006952, aug 2017.

[73] Joseph C Maranville, Francesca Luca, Allison L Richards, Xiaoquan Wen, David B Witonsky, Shaneen Baxter, Matthew Stephens, and Anna Di Rienzo. Interactions between glucocorticoid treatment and cis-regulatory polymorphisms contribute to cellular response phenotypes. *PLoS genetics*, 7:e1002162, 2011.

[74] Elena Matsa, Paul W. Burridge, Kun Hsing Yu, John H. Ahrens, Vittavat Termglinchan, Haodi Wu, Chun Liu, Praveen Shukla, Nazish Sayed, Jared M. Churko, Ningyi Shao, Nicole A. Woo, Alexander S. Chao, Joseph D. Gold, Ioannis Karakikes, Michael P. Snyder, and Joseph C. Wu. Transcriptome Profiling of Patient-Specific Human iPSC-Cardiomyocytes Predicts Individual Drug Safety and Efficacy Responses In-Vitro. *Cell Stem Cell*, 19(3):311–325, sep 2016.

[75] Matthew N N. McCall, Peter B B Illei, and Marc K K Halushka. Complex Sources of Variation in Tissue Expression Data: Analysis of the GTEx Lung Transcriptome. *American Journal of Human Genetics*, 99(3):624–635, 2016.

[76] Eugene Melamud and John Moult. Stochastic noise in splicing machinery. *Nucleic Acids Research*, 37(14):4873–4886, aug 2009.

[77] Barbara Messner, Michael Knoflach, Andreas Seubert, Andreas Ritsch, Kristian Pfaller, Blair Henderson, Ying H. Shen, Iris Zeller, Johann Willeit, Günther Laufer, Georg Wick, Stefan Kiechl, and David Bernhard. Cadmium is a novel and independent risk factor for early atherosclerosis mechanisms and in vivo relevance. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 29(9):1392–1398, sep 2009.

[78] Ilaria Mogno, Francesco Vallania, Robi D. Mitra, and Barak A. Cohen. TATA is a modular component of synthetic promoters. *Genome Research*, 20(10):1391–1397, oct 2010.

[79] Pejman Mohammadi, Stephane E. Castel, Beryl B. Cummings, Jonah Einson, Christina Sousa, Paul Hoffman, Sandra Donkervoort, Zhuoxun Jiang, Payam Mohassel, A. Reghan Foley, Heather E. Wheeler, Hae Kyung Im, Carsten G. Bonnemann, Daniel G. MacArthur, and Tuuli Lappalainen. Genetic regulatory variation in populations informs transcriptome analysis in rare disease. *Science*, 366(6463):351–356, oct 2019.

[80] G.A. A Moyerbrailean, G.O. O Davis, C.T. T Harvey, D. Watza, X. Wen, R. Pique-Regi, and F. Luca. A high-throughput RNA-seq approach to profile transcriptional responses. *Scientific reports*, 5:14976, oct 2015.

[81] Gregory A. Moyerbrailean, Allison L. Richards, Daniel Kurtz, Cynthia A. Kalita, Gordon O. Davis, Chris T. Harvey, Adnan Alazizi, Donovan Watza, Yoram Sorokin, Nancy Hauff, Xiang Zhou, Xiaoquan Wen, Roger Pique-Regi, and Francesca Luca. High-throughput allele-specific expression across 250 environmental conditions. *Genome research*, 26(12):1627–1638, dec 2016.

[82] Yohann Nédélec, Joaquín Sanz, Golshid Baharian, Zachary A. Z.A. Szpiech, Alain Pacis, Anne Dumaine, J.-C. Jean-Christophe Grenier, Andrew Freiman, A.J. Aaron J. Sams, Steven Hebert, Ariane Pagé Sabourin, Francesca Luca, Ran Blekhman, Ryan D. R.D. Hernandez, Roger Pique-Regi, Jenny Tung, Vania Yotova, and Luis B. L.B. Barreiro. Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell*, 167(3):657–669.e21, oct 2016.

[83] Jonas B. Nielsen, Rosa B. Thorolfsdottir, Lars G. Fritsche, Wei Zhou, Morten W. Skov, Sarah E. Graham, Todd J. Herron, Shane McCarthy, Ellen M. Schmidt, Gardar Sveinbjornsson, Ida Surakka, Michael R. Mathis, Masatoshi Yamazaki, Ryan D. Crawford, Maiken E. Gabrielsen, Anne Heidi Skogholt, Oddgeir L. Holmen, Maoxuan Lin, Brooke N. Wolford, Rounak Dey, Håvard Dalen, Patrick Sulem, Jonathan H. Chung, Joshua D. Backman, David O. Arnar, Unnur Thorsteinsdottir, Aris Baras, Colm O'Dushlaine, Anders G. Holst, Xiaoquan Wen, Whitney Hornsby, Frederick E. Dewey, Michael Boehnke, Sachin Kheterpal, Bhramar Mukherjee, Seunggeun Lee, Hyun M. Kang, Hilma Holm, Jacob Kitzman, Jordan A. Shavit, José Jalife, Chad M. Brummett, Tanya M. Teslovich, David J. Carey, Daniel F. Gudbjartsson, Kari Stefansson, Gonçalo R. Abecasis, Kristian Hveem, and Cristen J. Willer. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nature genetics*, 50(9):1234–1239, sep 2018.

[84] Athma A. Pai and Francesca Luca. Environmental influences on RNA processing: Biochemical, molecular and genetic regulators of cellular response, jan 2019.

[85] Athanasia D. Panopoulos, Matteo D'Antonio, Paola Benaglio, Roy Williams, Sherin I. Hashem, Bernhard M. Schuldt, Christopher DeBoever, Angelo D. Arias, Melvin Garcia, Bradley C. Nelson, Olivier Harismendy, David A. Jakubosky, Margaret K.R. Donovan, William W. Greenwald, Kathy Jean Farnam, Megan Cook, Victor Borja, Carl A. Miller, Jonathan D. Grinstein, Frauke Drees, Jonathan Okubo, Kenneth E. Diffenderfer, Yuriko Hishida, Veronica Modesto, Carl T. Dargitz, Rachel Feiring, Chang Zhao, Aitor Aguirre, Thomas J. McGarry, Hiroko Matsui, He Li, Joaquin Reyna, Fangwen Rao, Daniel T. O'Connor, Gene W. Yeo, Sylvia M. Evans, Neil C. Chi, Kristen Jepsen, Naoki Nariai, Franz Josef Müller, Lawrence S.B. Goldstein, Juan Carlos Izpisua Belmonte, Eric Adler, Jeanne F. Loring, W. Travis Berggren, Agnieszka D'Antonio-Chronowska, Erin N. Smith, and Kelly A. Frazer. iPSCORE: A Resource of 222 iPSC Lines Enabling Functional Characterization of Genetic Variation across a Variety of Cell Types. *Stem Cell Reports*, 8(4):1086–1100, apr 2017.

[86] Bryan J. Pavlovic, Lauren E. Blake, Julien Roux, Claudia Chavarria, and Yoav Gilad. A Comparative Assessment of Human and Chimpanzee iPSC-derived Cardiomyocytes with Primary Heart Tissues. *Scientific Reports*, 8(1):1–14, dec 2018.

[87] Ali Pazokitoroudi, Yue Wu, Kathryn S. Burch, Kangcheng Hou, Aaron Zhou, Bogdan Pasaniuc, and Sriram Sankararaman. Efficient variance components analysis across millions of genomes. *Nature Communications*, 11(1):1–10, dec 2020.

[88] Joseph K. Pickrell, Athma A. Pai, Yoav Gilad, and Jonathan K. Pritchard. Noisy Splicing Drives mRNA Isoform Diversity in Human Cells. *PLoS Genetics*, 6(12):e1001236, dec 2010.

[89] Janet Piñero, Juan Manuel Ramírez-Anguita, Josep Saüch-Pitarch, Francesco Ronzano, Emilio Centeno, Ferran Sanz, and Laura I. Furlong. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research*, 48(D1):D845–D855, jan 2020.

[90] Vladimir Presnyak, Najwa Alhusaini, Ying Hsin Chen, Sophie Martin, Nathan Morris, Nicholas Kline, Sara Olson, David Weinberg, Kristian E. Baker, Brenton R. Graveley, and Jeff Coller. Codon optimality is a major determinant of mRNA stability. *Cell*, 160(6):1111–1124, mar 2015.

[91] Hélène Quach, Maxime Rotival, Julien Pothlichet, Yong Hwee Eddie Loh, Michael Dannemann, Nora Zidane, Guillaume Laval, Etienne Patin, Christine Harmant, Marie Lopez, Matthieu Deschamps, Nadia Naffakh, Darragh Duffy, Anja Coen, Geert Leroux-Roels, Frederic Clément, Anne Boland, Jean François Deleuze, Janet Kelso, Matthew L. Albert, and Lluis Quintana-Murci. Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. *Cell*, 167(3):643–656.e17, oct 2016.

[92] Payal Rana, Blake Anson, Sandra Engle, and Yvonne Will. Characterization of human-induced pluripotent stem cell-derived cardiomyocytes: Bioenergetics and utilization in safety screening. *Toxicological Sciences*, 130(1):117–131, nov 2012.

[93] Jonathan M. Raser and Erin K. O'Shea. Control of stochasticity in eukaryotic gene expression. *Science*, 304(5678):1811–1814, jun 2004.

[94] Christoph D. Rau, Aldons J. Lusis, and Yibin Wang. Genetics of common forms of heart failure: Challenges and potential solutions, may 2015.

[95] Justyna Resztak, Allison Farrell, Henriette Mair-Meijers, Adnan Alazizi, Xiaoquan Wen, Derek Wildman, Samuele Zilioli, Richard Slatcher, Roger Pique-Regi, and Francesca Luca. Psychosocial experiences modulate asthma-associated genes through gene-environment interactions. *bioRxiv*, page 2020.07.16.206912, jul 2020.

[96] Allison L. Richards, Donovan Watza, Anthony Findley, Adnan Alazizi, Xiaoquan Wen, Athma A. Pai, Roger Pique-Regi, and Francesca Luca. Environmental perturbations lead to extensive directional shifts in RNA processing. *PLoS Genetics*, 13(10), oct 2017.

[97] Allison L. Richards, Donovan Watza, Anthony Findley, Adnan Alazizi, Xiaoquan Wen, Athma A. Pai, Roger Pique-Regi, and Francesca Luca. Environmental perturbations lead to extensive directional shifts in RNA processing. *PLoS Genetics*, 13(10):e1006995, oct 2017.

[98] Matthew E Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research*, 43(7):e47, apr 2015.

[99] Foad Rouhani, Natsuhiko Kumasaka, Miguel Cardoso de Brito, Allan Bradley, Ludovic Vallier, and Daniel Gaffney. Genetic Background Drives Transcriptional Variation in Human Induced Pluripotent Stem Cells. *PLoS Genetics*, 10(6):e1004432, jun 2014.

[100] Abhishek K Sarkar, Po Yuan Tung, John D. Blischak, Jonathan E. Burnett, Yang I. Li, Matthew Stephens, and Yoav Gilad. Discovery and characterization of variance QTLs in human induced pluripotent stem cells. *PLoS genetics*, 15(4):e1008045, sep 2019.

[101] Gregg L. Semenza. Hypoxia-inducible factor 1 and cardiovascular disease. *Annual Review of Physiology*, 76:39–56, feb 2014.

[102] Sonia Shah, Albert Henry, Carolina Roselli, Honghuang Lin, Gardar Sveinbjörnsson, Ghazaleh Fatemifar, Åsa K. Hedman, Jemma B. Wilk, Michael P. Morley, Mark D. Chaffin, Anna Helgadottir, Niek Verweij, Abbas Dehghan, Peter Almgren, Charlotte Andersson, Krishna G. Aragam, Johan Ärnlöv, Joshua D. Backman, Mary L. Biggs, Heather L. Bloom, Jeffrey Brandimarto, Michael R. Brown, Leonard Buckbinder, David J. Carey, Daniel I. Chasman, Xing Chen, Xu Chen, Jonathan Chung, William Chutkow, James P. Cook, Graciela E. Delgado, Spiros Denaxas, Alexander S. Doney, Marcus Dörr, Samuel C. Dudley, Michael E. Dunn, Gunnar Engström, Tõnu Esko, Stephan B. Felix, Chris Finan, Ian Ford, Mohsen Ghanbari, Sahar Ghasemi, Vilmantas Giedraitis, Franco Giulianini, John S. Gottdiener, Stefan Gross, Daníel F. Gudbjartsson, Rebecca Gutmann, Christopher M. Haggerty, Pim van der Harst, Craig L. Hyde, Erik Ingelsson, J. Wouter Jukema, Maryam Kavousi, Kay Tee Khaw, Marcus E. Kleber, Lars Køber, Andrea Koekemoer, Claudia Langenberg, Lars Lind, Cecilia M. Lindgren, Barry London, Luca A. Lotta, Ruth C. Lovering, Jianan Luan, Patrik Magnusson, Anubha Mahajan, Kenneth B. Margulies, Winfried März, Olle Melander, Ify R. Mordi, Thomas Morgan, Andrew D. Morris, Andrew P. Morris, Alanna C. Morrison, Michael W. Nagle, Christopher P. Nelson, Alexander Niessner, Teemu Niiranen, Michelle L. O'Donoghue, Anjali T. Owens, Colin N.A. Palmer, Helen M. Parry, Markus Perola, Eliana Portilla-Fernandez, Bruce M. Psaty, Goncalo Abecasis, Joshua Backman, Xiaodong Bai, Suganthi Balasubramanian, Nilanjana Banerjee, Aris Baras, Leland Barnard, Christina Beechert, Andrew Blumenfeld, Michael Cantor, Yating Chai, Jonathan Chung, Giovanni Coppola, Amy Damask, Frederick Dewey, Aris Economides, Gisu Eom, Caitlin Forsythe, Erin D. Fuller, Zhenhua Gu, Lauren Gurski, Paloma M. Guzzardo, Lukas Habegger, Young Hahn, Alicia Hawes, Cristopher van Hout, Marcus B. Jones, Shareef Khalid, Michael Lattari, Alexander Li, Nan Lin, Daren Liu, Alexander Lopez, Kia Manoochehri, Jonathan Marchini, Anthony Marcketta, Evan K. Maxwell, Shane McCarthy, Lyndon J. Mitnaul, Colm O'Dushlaine, John D. Overton, Maria Sotiropoulos Padilla, Charles Paulding, John Penn, Manasi Pradhan, Jeffrey G. Reid, Thomas D. Schleicher, Claudia Schurmann, Alan Shuldiner, Jeffrey C. Staples, Dylan Sun, Karina Toledo, Ricardo H. Ulloa, Louis Widom, Sarah E. Wolf, Ashish Yadav, Bin Ye, Kenneth M. Rice, Paul M. Ridker, Simon P.R. Romaine, Jerome I. Rotter, Perttu Salo, Veikko Salomaa, Jessica van Setten, Alaa A. Shalaby, Diane T. Smelser, Nicholas L. Smith, Steen Stender, David J. Stott, Per Svensson, Mari Liis Tammesoo, Kent D. Taylor, Maris Teder-Laving, Alexander Teumer, Gudmundur Thorgeirsson, Unnur Thorsteinsdottir, Christian Torp-Pedersen, Stella Trompet, Benoit Tyl, Andre G. Uitterlinden, Abirami Veluchamy, Uwe Völker, Adriaan A. Voors, Xiaosong Wang, Nicholas J. Wareham, Dawn Waterworth, Peter E. Weeke, Raul Weiss, Kerri L. Wiggins, Heming Xing, Laura M. Yerges-Armstrong, Bing Yu, Faiez Zannad, Jing Hua Zhao, Harry Hemingway, Nilesh J. Samani, John J.V. McMurray, Jian Yang, Peter M. Visscher, Christopher Newton-Cheh, Anders Malarstig, Hilma Holm, Steven A. Lubitz, Naveed Sattar, Michael V. Holmes, Thomas P. Cappola, Folkert W. Asselbergs, Aroon D. Hingorani, Karoline Kuchenbaecker, Patrick T. Ellinor, Chim C. Lang, Kari Stefansson, J. Gustav Smith, Ramachandran S. Vasan, Daniel I. Swerdlow, and R. Thomas Lumbers. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nature Communications*, 11(1), dec 2020.

[103] Arun Sharma, Paul W. Burridge, Wesley L. McKeithan, Ricardo Serrano, Praveen Shukla, Nazish Sayed, Jared M. Churko, Tomoya Kitani, Haodi Wu, Alexandra Holmström, Elena Matsa, Yuan Zhang, Anusha Kumar, Alice C. Fan, Juan C. Del Álamo, Sean M. Wu, Javid J. Moslehi, Mark Mercola, and Joseph C. Wu. High-throughput screening of tyrosine kinase inhibitor cardiotoxicity with human induced pluripotent stem cells. *Science Translational Medicine*, 9(377), feb 2017.

[104] Arun Sharma, Caleb Marceau, Ryoko Hamaguchi, Paul W. Burridge, Kuppusamy Rajarajan, Jared M. Churko, Haodi Wu, Karim I. Sallam, Elena Matsa, Anthony C. Sturzu, Yonglu Che,

Antje Ebert, Sebastian Diecke, Ping Liang, Kristy Red-Horse, Jan E. Carette, Sean M. Wu, and Joseph C. Wu. Human induced pluripotent stem cell-derived cardiomyocytes as an in vitro model for coxsackievirus B3-induced myocarditis and antiviral drug screening platform. *Circulation Research*, 115(6):556–566, aug 2014.

[105] S. T. Sherry, M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, and K. Sirotkin. DbSNP: The NCBI database of genetic variation. *Nucleic Acids Research*, 29(1):308–311, jan 2001.

[106] Olga M Sigalova, Amirreza Shaeiri, Mattia Forneris, Eileen EM Furlong, and Judith B Zaugg. Predictive features of gene expression variation reveal mechanistic link with differential expression. *Molecular Systems Biology*, 16(8), aug 2020.

[107] Jared L. Sterneckert, Peter Reinhardt, and Hans R. Schöler. Investigating human disease using stem cell models, jul 2014.

[108] B. J. Strober, R. Elorbany, K. Rhodes, N. Krishnan, K. Tayeb, A. Battle, and Y. Gilad. Dynamic genetic regulation of gene expression during cellular differentiation. *Science*, 364(6447):1287–1290, 2019.

[109] D. Leland Taylor, David A. Knowles, Laura J. Scott, Andrea H. Ramirez, Francesco Paolo Casale, Brooke N. Wolford, Li Guan, Arushi Varshney, Ricardo D'Oliveira Albanus, Stephen C. J. Parker, Narisu Narisu, Peter S. Chines, Michael R. Erdos, Ryan P. Welch, Leena Kinnunen, Jouko Saramies, Jouko Sundvall, Timo A. Lakka, Markku Laakso, Jaakko Tuomilehto, Heikki A. Koistinen, Oliver Stegle, Michael Boehnke, Ewan Birney, and Francis S. Collins. Interactions between genetic variation and cellular environment in skeletal muscle gene expression. *PLOS ONE*, 13(4):e0195788, apr 2018.

[110] Rob ter Horst, Martin Jaeger, Sanne P. Smeekens, Marije Oosting, Morris A. Swertz, Yang Li, Vinod Kumar, Dimitri A. Diavatopoulos, Anne F.M. Jansen, Heidi Lemmers, Helga Toenhake-Dijkstra, Antonius E. van Herwaarden, Matthijs Janssen, Renate G. van der Molen, Irma Joosten, Fred C.G.J. Sweep, Johannes W. Smit, Romana T. Netea-Maier, Mieke M.J.F. Koenders, Ramnik J. Xavier, Jos W.M. van der Meer, Charles A. Dinarello, Norman Pavelka, Cisca Wijmenga, Richard A. Notebaart, Leo A.B. Joosten, and Mihai G. Netea. Host and Environmental Factors Influencing Individual Human Cytokine Responses. *Cell*, 167(4):1111–1124.e13, nov 2016.

[111] Caroline Treins, Sophie Giorgetti-Peraldi, Joseph Murdaca, Gregg L. Semenza, and Emmanuel Van Obberghen. Insulin stimulates hypoxia-inducible factor 1 through a phosphatidylinositol 3-kinase/target of rapamycin-dependent signaling pathway. *Journal of Biological Chemistry*, 277(31):27975–27981, aug 2002.

[112] Jernej Ule and Benjamin J. Blencowe. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution, oct 2019.

[113] Benjamin D. Umans, Alexis Battle, and Yoav Gilad. Where Are the Disease-Associated eQTLs?, 2020.

[114] Cornelis S. Van Der Hooft, Jan Heeringa, Guy G. Brusselle, Albert Hofman, Jacqueline C.M. Witteman, J. Herre Kingma, Miriam C.J.M. Sturkenboom, and Bruno H.Ch Stricker. Corticosteroids and the risk of atrial fibrillation. *Archives of Internal Medicine*, 166(9):1016–1020, may 2006.

[115] Monique G. P. van der Wijst, Harm Brugge, Dylan H. de Vries, Patrick Deelen, Morris A. Swertz, and Lude Franke. Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nature Genetics*, 50(4):493–497, apr 2018.

[116] Markus C. Wahl, Cindy L. Will, and Reinhard Lührmann. The Spliceosome: Design Principles of a Dynamic RNP Machine, feb 2009.

[117] Yihan Wan and Daniel R. Larson. Splicing heterogeneity: Separating signal from noise. *Genome Biology*, 19(1):1–10, jul 2018.

[118] Teresa W. Wang, Roel C.H. Vermeulen, Wei Hu, Gang Liu, Xiaohui Xiao, Yuriy Alekseyev, Jun Xu, Boris Reiss, Katrina Steiling, George S. Downward, Debra T. Silverman, Fusheng Wei, Guoping Wu, Jihua Li, Marc E. Lenburg, Nathaniel Rothman, Avrum Spira, and Qing Lan. Gene-expression profiling of buccal epithelium among non-smoking women exposed to household air pollution from smoky coal. *Carcinogenesis*, 36(12):1494–1501, dec 2015.

[119] Michelle C. Ward and Yoav Gilad. A generally conserved response to hypoxia in iPSC-derived cardiomyocytes from humans and chimpanzees. *eLife*, 8, apr 2019.

[120] Yan Xu, Leslie R. Morse, Raquel Assed Bezerra Da Silva, Paul R. Odgren, Hajime Sasaki, Philip Stashenko, and Ricardo A. Battaglino. PAMM: A redox regulatory protein that modulates osteoclast differentiation. *Antioxidants and Redox Signaling*, 13(1):27–37, jul 2010.

[121] Jian Yang, Beben Benyamin, Brian P. McEvoy, Scott Gordon, Anjali K. Henders, Dale R. Nyholt, Pamela A. Madden, Andrew C. Heath, Nicholas G. Martin, Grant W. Montgomery, Michael E. Goddard, and Peter M. Visscher. Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics*, 42(7):565–569, jul 2010.

[122] Jian Yang, S. Hong Lee, Michael E. Goddard, and Peter M. Visscher. GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, 88(1):76–82, jan 2011.

[123] Guangchuang Yu, Li Gen Wang, Yanyan Han, and Qing Yu He. ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS A Journal of Integrative Biology*, 16(5):284–287, may 2012.

[124] Yuhua Zhang, Corbin Quick, Ketian Yu, Alvaro Barbeira, Francesca Luca, Roger Pique-Regi, Hae Kyung Im, and Xiaoquan Wen. PTWAS: Investigating tissue-relevant causal molecular mechanisms of complex traits using probabilistic TWAS analysis. *Genome Biology*, 21(1):232, sep 2020.

[125] Daria V Zhernakova, Patrick Deelen, Martijn Vermaat, Maarten van Iterson, Michiel van Galen, Wibowo Arindrarto, Peter van 't Hof, Hailiang Mei, Freerk van Dijk, Harm-Jan Westra, Marc Jan Bonder, Jeroen van Rooij, Marijn Verkerk, P Mila Jhamai, Matthijs Moed, Szymon M Kielbasa, Jan Bot, Irene Nooren, René Pool, Jenny van Dongen, Jouke J Hottenga, Coen D A Stehouwer, Carla J H van der Kallen, Casper G Schalkwijk, Alexandra Zhernakova, Yang Li, Ettje F Tigchelaar, Niek de Klein, Marian Beekman, Joris Deelen, Diana van Heemst, Leonard H van den Berg, Albert Hofman, André G Uitterlinden, Marleen M J van Greevenbroek, Jan H Veldink, Dorret I Boomsma, Cornelia M van Duijn, Cisca Wijmenga, P Eline Slagboom, Morris A Swertz, Aaron Isaacs, Joyce B J van Meurs, Rick Jansen, Bastiaan T Heijmans, Peter A C 't Hoen, and Lude Franke. Identification of context-dependent expression quantitative trait loci in whole blood. *Nature genetics*, 49(1):139–145, jan 2017.

Supplementary Materials

Supplementary Tables

Table S1: Flow cytometry results from CMs. Purity of the cell cultures was determined by measuring the expression of the cardiac troponin T in each CM cell line, plus one IPSC line to serve as a negative control. Columns 1-5 are: 1) Individual; 2) Cell Type; 3) Differentiation Day; 4) Percentage of live cells; 5) Percentage of cells expressing cardiac troponin T

Individual	CellType	Differentiation day	% live	% cTnT positive
GM18858	CM	27	60.3	95
GM18855	CM	27	37.6	74.1
GM18505	CM	27	58.1	90.7
GM18912	CM	25	17.2	84.9
GM18520	CM	25	11.2	72.3
GM19209	CM	25	23.8	44.9
GM18912	IPSC	0	89.9	0

Table S2: IPSC differentiation metadata. Columns 1-6 are: 1) Individual; 2) Number of passages on mouse embryonic fibroblasts (MEF); 3) Number of passages without MEF; 4) Reprogramming batch; 5) Replicate; 6) Day at which cells reached confluency

Line	Feeder passage	FF passage	Batch	Replicate	Confluence day
GM18505	23	41	1	1	90
GM18855	14	51	1	1	85
GM18858	18	47	1	1	97
GM18505	23	42	1	2	40
GM18855	14	51	1	2	95
GM18858	18	47	1	2	98
GM19209	23	51	2	1	85
GM18520	15	42	2	1	80
GM18912	16	36	2	1	95
GM19209	23	52	2	2	80
GM18520	15	43	2	2	95
GM18912	16	37	2	2	70

Table S3: CM reprogramming metadata.Columns 1-8 are: 1) Individual; 2) Average percent confluence on day 0; 3) Number of passages on mouse embryonic fibroblasts (MEF); 4) Number of passages without MEF; 5) Differentiation batch; 6) Day at which cells started beating; 7) Cells/ml on day 20; 8) Total number of cells from 4 plates

Individual	% confluence	Feeder Passage	FF Passage	Batch	Day beating	Counts day20	Total cells
GM18858	76.25	18	41	1	7.5	1820000	54600000
GM18855	72.5	11	45	1	7	1440000	43200000
GM18505	73.75	23	35	1	8	1630000	48900000
GM19209	83.75	23	44	2	9.75	2096000	20960000
GM18912	95	16	50	2	7	4700000	70500000
GM18520	90	15	46	2	8	2630000	78900000

Table S4: Treatment concentrations. Concentrations for all treatments used in this study. Columns 1-4 are: 1) Treatment ID; 2) Treatment Name; 3) Final treatment concentration in the culture media; 4) Control ID

Treatment.ID	Common Name	Final Conc	Control.ID	Category
CO1	Media	–	–	Control
CO2	Ethanol	1uL in 10mL	–	Control
T12C1	Dexamethasone	10-6 M	CO2	Steroid Hormones
T13C1	Caffeine	225 ug/mL	CO1	Dietary components
T14C1	Nicotine	100 ug/mL	CO1	Environmental contaminants and common chemic
T15C1	Copper	60 uM	CO1	Metal Ions
T19C1	Selenium	10 uM	CO1	Metal Ions
T20C1	Zinc	80 uM	CO1	Metal Ions
T23C1	Vitamin D	0.1uM	CO2	Dietary components
T24C1	Acrylamide	2ng/mL	CO1	Environmental contaminants and common chemic
T25C1	BP-3	1000ng/mL	CO2	Environmental contaminants and common chemic
T26C1	BPA	20ng/mL	CO1	Environmental contaminants and common chemic
T27C1	Cadmium	2ng/mL	CO1	Metal Ions
T29C1	PFOA	15ng/mL	CO1	Environmental contaminants and common chemic
T30C1	Triclosan	1200ng/mL	CO2	Environmental contaminants and common chemic
T33C1	Insulin	1004.6ng/mL	CO1	Peptide Hormones and neurotransmitters
T36C1	Vasopressin	0.017ng/mL	CO1	Peptide Hormones and neurotransmitters
T37C1	Acetylcholine	293.9ng/mL	CO1	Peptide Hormones and neurotransmitters
T40C1	BHA	0.2ug/mL	CO2	Environmental contaminants and common chemic
T41C1	Ibuprofen	5ug/mL	CO2	Common drugs
T42C1	Acetaminophen	50ug/mL	CO1	Common drugs
T43C1	Aspirin	10ug/mL	CO2	Common drugs
T44C1	Loratadine	50ng/mL	CO2	Common drugs
T45C1	Cetirizine	10ng/mL	CO1	Common drugs
T47C1	Phthalate	100ng/mL	CO2	Environmental contaminants and common chemic
T4C1	Vitamin B5	100 nM	CO1	Dietary components
T5C1	Vitamin B6	10-6M	CO1	Dietary components
T6C1	Vitamin A	10 nM	CO2	Dietary components
T7C1	Vitamin E	50 uM	CO2	Dietary components
T9C1	Aldosterone	10-6 M	CO2	Steroid Hormones

Table S5: Shallow DEG results. Results from DESeq2 using shallow sequencing data (eq. 1). Columns 1-8 are: 1) Transcript ID; 2) FDR; 3) P-value; 4) Log2 fold change; 5) Ensembl gene ID; 6) Gene ID; 7) Cell type; 8) Treatment ID

http://genome.grid.wayne.edu/GxExC/Supp_Tables/DESeq_Shallow.tab.gz

Table S6: Sample information and sequencing depth for deep-sequenced libraries. Each row represents a sequencing library. Columns 1-12 are: 1) Library ID; 2) Plate; 3) Sequencing barcode; 4) Treatment ID; 5) Cell type; 6) Individual; 7) Plate well, 8) Control ID, 9) Treatment Name, 10) Number of sequencing reads after quality control and removal of PCR duplicates, 11) Total number of raw sequencing reads, 12) Percentage of raw sequencing reads remaining after quality control and deduplication

`http://genome.grid.wayne.edu/GxExC/Supp_Tables/cv_reads.tab`

Table S7: Deep DEG results. Results from DESeq2 using deep sequencing data (eq. 1). Columns 1-8 are: 1) Transcript ID; 2) FDR; 3) P-value; 4) Log2 fold change; 5) Ensembl gene ID; 6) Gene ID; 7) Cell type; 8) Treatment ID

`http://genome.grid.wayne.edu/GxExC/Supp_Tables/DESeq_Deep.tab.gz`

Table S8: DEG GO terms. Significant gene ontology enrichments for DEGs. Tables are organized in subfolders based on cell type, and files are named according to treatment and which group of genes were used for the enrichment (upregulated, downregulated, or all DEGs). The naming convention is: *Treatment ID*_enrichment_summary_*Gene group*.txt, where *Gene group* is either *upDEGs*, *downDEGs*, or *allDEGs* for up-regulated, downregulated, and all DEGs, respectively. Columns 1-9 are: 1) GO term ID; 2) GO description; 3) Gene ratio; 4) Background ratio; 5) P-value; 6) FDR; 7) Q-value; 8) Gene IDs for DEGs in GO term; 9) Number of DEGs in GO term

`http://genome.grid.wayne.edu/GxExC/Supp_Tables/DEG_GO_terms/`

Table S9: Cell type × treatment interaction gene expression results . Results from DESeq2 using deep sequencing data to identify genes with Cell type × treatment interactions (Eq. 2). Columns 1-6 are: 1) Transcript ID; 2) FDR; 3) P-value; 4) Log2 fold change; 5) Ensembl gene ID; 6) Gene ID

`http://genome.grid.wayne.edu/GxExC/Supp_Tables/DESeq_CellSpecific.tab.gz`

Table S10: QuASAR output. Results from QuASAR for all SNPs tested for ASE. SNPs must be covered by at least 5 reads. Each row represents a SNP measured in a particular library. Columns 1-17 are: 1) Chromosome; 2) 0-based coordinate; 3) 1-based coordinate; 4) Reference allele; 5) Alternate allele; 6) rsID; 7) Allele frequency; 8) Number of RNA-seq reads mapping to reference allele; 9) Number of RNA-seq reads mapping to alternate allele; 10) Beta: measure of ASE from QuASAR, where positive indicates increased expression of reference allele; 11) Standard error of ASE measurement from QuASAR; 12) P-value for ASE; 13) Q-value for ASE; Columns 14-17 describe the library from where the ASE measurement comes from: 14) Plate ID; 15) Cell type; 16) Individual; 17) Name of treatment; 18) Treatment ID

`http://genome.grid.wayne.edu/GxExC/Supp_Tables/Quasar_output.tab.gz`

Table S11: ANOVA results with all cell types analyzed together. Results from ANOVA model with all cell types to identify SNPs with ASE (eq. 3 and 4). Columns 1-8 are: 1) SNP by individual identifier; 2) Residual degrees of freedom; 3) Residual sum of squares; 4) Degrees of freedom; 5) Sum of squares; 6) F-value; 7) P-value; 8) FDR.

`http://genome.grid.wayne.edu/GxExC/Supp_Tables/ANOVA_allCell.tab`

Table S12: ANOVA results with cell types analyzed separately. Results from ANOVA model with each cell type analyzed separately to identify SNPs with ASE (eq. 6 and 7). Columns 1-9 are: 1) SNP by individual identifier; 2) Residual degrees of freedom; 3) Residual sum of squares; 4) Degrees of freedom; 5) Sum of squares; 6) F-value; 7) P-value; 8) FDR; 9) Cell type.

`http://genome.grid.wayne.edu/GxExC/Supp_Tables/ANOVA_cellSep.tab`

Table S13: Fixed effects model output with all cell types analyzed together. Results from the linear model with all cell types analyzed together to identify cASE. Columns 1-8 are: 1) Variable in the linear model being tested (in Eq. 5; e.g. cell type, treatment, etc.); 2) Estimate; 3) Standard error; 4) Test statistic; 5) P-value; 6) Degrees of freedom; 7) Rank; 8) SNP by individual identifier.

`http://genome.grid.wayne.edu/GxExC/Supp_Tables/FixedEffect_allCell.tab`

Table S14: Number of significant tests from fixed effect model for all cell types analyzed together. Summary of results from Table S13, including the number of significant tests (FDR < 10%) per variable. Columns 1-3 are: 1) Variable in the linear model being tested (in Eq. 5; e.g. treatment, plate,etc); 2) Number of tests performed; 3) Number of tests with FDR < 10%.

Term	# tests	# Sig.
IPSC	13009	4084
CM	13009	3782
Acetaminophen	12903	106
Aldosterone	12941	93
Cadmium	12941	135
Caffeine	12912	111
Copper	12932	159
Dexamethasone	12926	114
Insulin	12950	107
Nicotine	12928	108
Selenium	12891	140
Triclosan	12927	101
VitaminA	12929	119
Zinc	12841	116
Acetaminophen x Cell Type	12189	70
Aldosterone x Cell Type	12370	59
Cadmium x Cell Type	12454	73
Caffeine x Cell Type	12308	57
Copper x Cell Type	12301	109
Dexamethasone x Cell Type	12345	71
Insulin x Cell Type	12369	60
Nicotine x Cell Type	12277	57
Selenium x Cell Type	12089	53
Triclosan x Cell Type	12371	55
VitaminA x Cell Type	12372	59
Zinc x Cell Type	12040	60

Table S15: Fixed effects model output with cell types analyzed separately. Results from the fixed effects model with cell types analyzed separately to identify cASE. Columns 1-9 are: 1) Variable in the linear model being tested (in Eq. 6; e.g. treatment, plate, etc.); 2) Estimate; 3) Standard error; 4) Test statistic; 5) P-value; 6) Degrees of freedom; 7) Rank; 8) SNP by individual identifier; 9) Cell type.

http://genome.grid.wayne.edu/GxExC/Supp_Tables/FixedEffect_cellSep.tab

Table S16: Number of significant tests from fixed effect model for cell types analyzed separately. Summary of results from Table S15, including the number of significant tests (FDR < 10%) per variable. Columns 1-4 are: 1) Cell type; 2) Variable in the linear model being tested (in Eq. 6; e.g. treatment, plate,etc); 3) Number of tests performed; 4) Number of tests with FDR < 10%.

Cell	Term	Num tests	Num sig
CM	Acetaminophen	6524	35
CM	Aldosterone	6590	38
CM	Cadmium	6573	53
CM	Caffeine	6556	45
CM	Copper	6598	46
CM	Dexamethasone	6524	39
CM	Insulin	6561	37
CM	Nicotine	6372	56
CM	Selenium	6517	31
CM	Triclosan	6551	39
CM	VitaminA	6549	37
CM	Zinc	6566	36
IPSC	Acetaminophen	6573	38
IPSC	Aldosterone	6589	40
IPSC	Cadmium	6577	44
IPSC	Caffeine	6535	35
IPSC	Copper	6544	47
IPSC	Dexamethasone	6578	42
IPSC	Insulin	6580	39
IPSC	Nicotine	6587	46
IPSC	Selenium	6442	49
IPSC	Triclosan	6601	41
IPSC	VitaminA	6578	40
IPSC	Zinc	6040	50
LCL	Acetaminophen	6682	16
LCL	Aldosterone	6820	13
LCL	Cadmium	6897	21
LCL	Caffeine	6863	22
LCL	Copper	6546	31
LCL	Dexamethasone	6770	22
LCL	Insulin	6861	14
LCL	Nicotine	6760	16
LCL	Selenium	6741	24
LCL	Triclosan	6903	13
LCL	VitaminA	6877	13
LCL	Zinc	6776	25

Table S17: cASE gene overlap in other G×E studies. Information from significant cASE genes which replicate (p < 0.05) in another G×E study. Columns 1-2 describe the gene: 1) Ensembl gene ID; 2) Gene ID; Columns 3-10 describe the cASE SNP: 3) cASE treatment; 4) SNP identifier; 5) Measurement of cASE effect from linear model; 6) Standard error of cASE measurement; 7) Test statistic; 8) P-value; 9) FDR; 10) Degrees of freedom; Columns 11-20 describe the G×E SNP in the replication datasets: 11) rsID; 12) Treatment condition; 13) Cell type; 14) Effect size; 15) P-value; 16) FDR; 17) Significance as reported by study (TRUE or FALSE); 18) Author and year; 19) DOI; 20) Supplemental table from which the results for the replication analysis were obtained.

`http://genome.grid.wayne.edu/GxExC/Supp_Tables/cASE_GxE_rep.tab`

Table S18: cASE overlap with doxorubicin reQTLs. Information on treatment cASE genes in CMs which overlap with doxorubicin reQTLs from [57]. Columns 1-10 describe the cASE gene: 1) SNP identifier; 2) Treatment; 3) Treatment ID; 4) Measurement of cASE effect from linear model; 5) Standard error of cASE measurement; 6) Test statistic; 7) P-value; 8) FDR; 9) Ensembl gene ID; 10) Gene ID.

SNP_Individual	term	Treatment.ID	Estimate	S.E.	Statistic	p-value	FDR	ENSG	Gene ID
rs72873318_GM18858	Acetaminophen	T42C1	0.57	0.094	6	0.00019	0.088	ENSG00000138759	FRAS1
rs3751862_GM18520	Caffeine	T13C1	0.98	0.14	7.2	0.00037	0.093	ENSG00000135698	MPHOSPH6
rs1425486_GM18858	Caffeine	T13C1	0.58	0.11	5.5	0.00019	0.088	ENSG00000145431	PDGFC
rs55815425_GM18520	Caffeine	T13C1	0.9	0.18	5.1	0.00021	0.09	ENSG00000198130	HIBCH
rs7670138_GM18912	Copper	T15C1	-1.5	0.3	-5.2	0.00039	0.094	ENSG00000153012	LGI2
rs57348345_GM19209	Copper	T15C1	2.3	0.43	5.2	0.00053	0.098	ENSG00000136267	DGKB
rs10070440_GM18912	Dexamethasone	T12C1	1.1	0.24	4.8	0.00056	0.099	ENSG00000122012	SV2C
rs3760_GM18912	Dexamethasone	T12C1	0.68	0.11	6.2	0.00015	0.086	ENSG00000078674	PCM1
rs1425486_GM18858	Dexamethasone	T12C1	0.58	0.1	5.5	0.00019	0.088	ENSG00000145431	PDGFC
rs77438708_GM18912	Insulin	T33C1	-1.2	0.21	-5.6	0.00011	0.085	ENSG00000125124	BBS2
rs72873318_GM18858	Nicotine	T14C1	0.67	0.12	5.6	0.00034	0.093	ENSG00000138759	FRAS1
rs2303650_GM18855	Nicotine	T14C1	0.59	0.11	5.4	0.00061	0.1	ENSG00000198677	TTC37
rs1331260_GM18855	Selenium	T19C1	-1.1	0.23	-4.8	0.00041	0.096	ENSG00000107249	GLIS3
rs3760_GM18912	Triclosan	T30C1	0.7	0.088	8	2.2e-05	0.064	ENSG00000078674	PCM1
rs9644500_GM19209	Triclosan	T30C1	2.2	0.44	5	0.00038	0.093	ENSG00000169436	COL22A1
rs1425486_GM18858	VitaminA	T6C1	0.69	0.1	6.6	4.1e-05	0.069	ENSG00000145431	PDGFC

Table S19: Treatment cASE gene overlap in atrial fibrillation GWAS. Significant treatment cASE genes which were associated with atrial fibrillation risk in GWAS [83]. Columns 1-2 describe the gene: 1) Ensembl gene ID; 2) Gene ID; Columns 3-10 describe the cASE SNP: 3) SNP by individual identifier; 4) cASE treatment name; 5) Treatment ID; 6) Measurement of cASE effect from linear model; 7) Standard error of cASE measurement; 8) Test statistic; 9) P-value; 10) FDR; 11) Degrees of freedom.

`http://genome.grid.wayne.edu/GxExC/Supp_Tables/cASE_AFib.tab`

Table S20: Mixed effects model output with all cell types analyzed together. Results from the mixed effects model with all cell types analyzed together to quantify ASE variance. Columns 1-6 are: 1) SNP by individual identifier; 2) Variance due to treatment; 3) Variance due to cell type; 4) Variance due to treatment × cell type interaction; 5) Variance due to control; 6) Variance due to residual.

`http://genome.grid.wayne.edu/GxExC/Supp_Tables/MixedEffect_allCell.tab`

46

Table S21: Mixed effects model output with cell types analyzed separately. Results from the mixed effects model with cell types analyzed separately to quantify ASE variance. Columns 1-9 are: 1) SNP by individual identifier; 2) Ensembl gene ID; 3) Gene ID; 4) Variance due to treatment; 5) Variance due to plate; 6) Variance due to control; 7) Variance due to residual; 8) Cell type.

`http://genome.grid.wayne.edu/GxExC/Supp_Tables/MixedEffect_cellSep.tab`

Table S22: ASE GO terms. Gene ontology enrichments for genes with high or low ASE variance. Files are named according to cell type, variance type (treatment, residual, or plate), and high/low variance using the following naming convention: *cell type_variance type_high/low variance_*`GO.txt` Columns 1-9 are: 1) GO term ID; 2) GO description; 3) Gene ratio; 4) Background ratio; 5) P-value; 6) FDR; 7) Q-value; 8) Gene IDs for high or low variance genes in GO term; 9) Number of high or low variance genes in GO term.

`http://genome.grid.wayne.edu/GxExC/Supp_Tables/ASE_GO/`

Table S23: TWAS abbreviations. Abbreviations for TWAS found in [124]. Additional details on TWAS studies can be found in [124]. Columns 1-2 are: 1) TWAS name as seen in Zhang et al; 2) TWAS abbreviation used in Figure 4A,B.

`http://genome.grid.wayne.edu/GxExC/Supp_Tables/TWAS_abbrev.tab`

Supplementary Figures
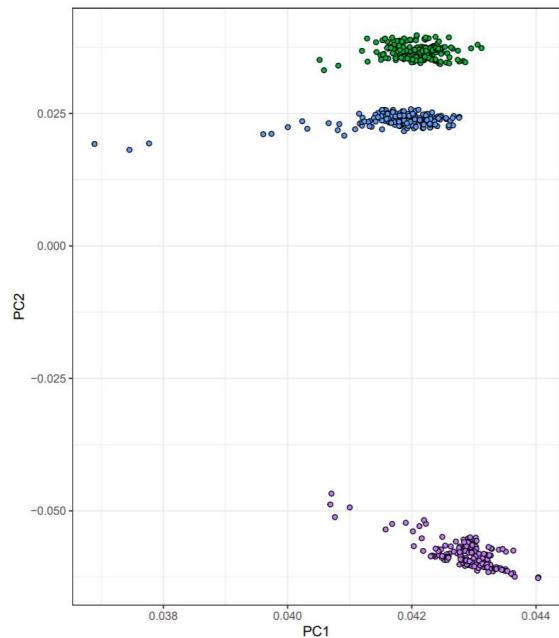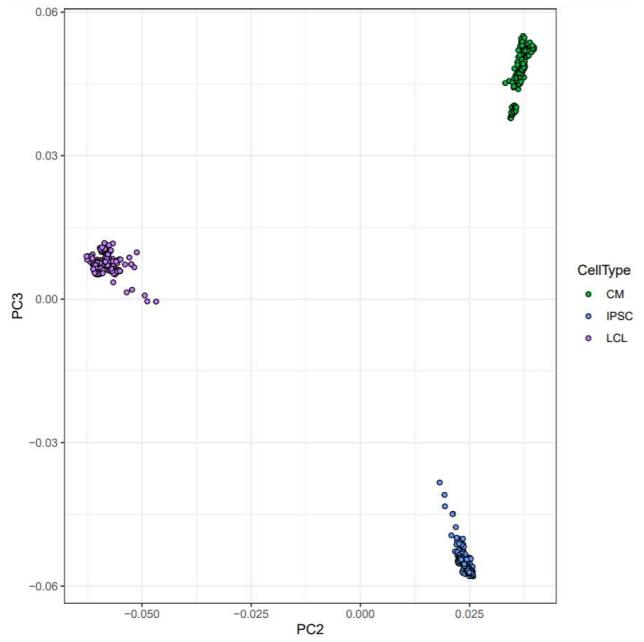
Figure S1: Principal component analysis of gene expression Principal components analysis was performed on normalized gene expression reads. Each point represents a deep-sequenced library, colored by cell type, for **A** PC1 vs PC2 and **B** PC2 vs PC3.
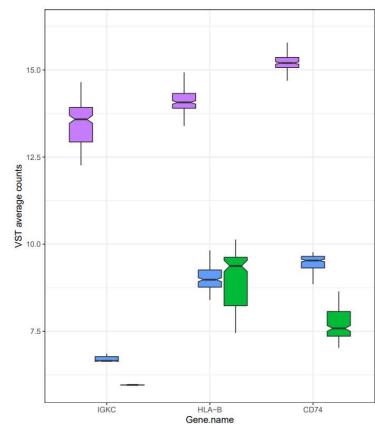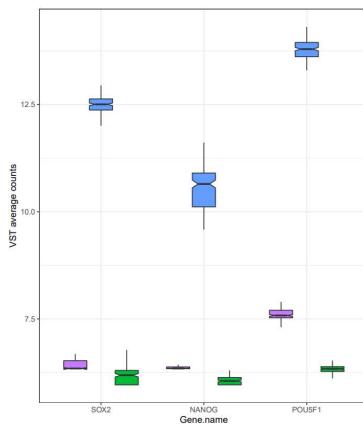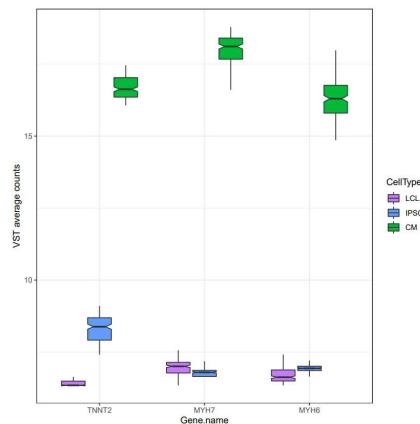
Figure S2: Expression of marker genes Normalized expression of three marker genes for each cell type from all deep-sequenced libraries are plotted by cell type.

48
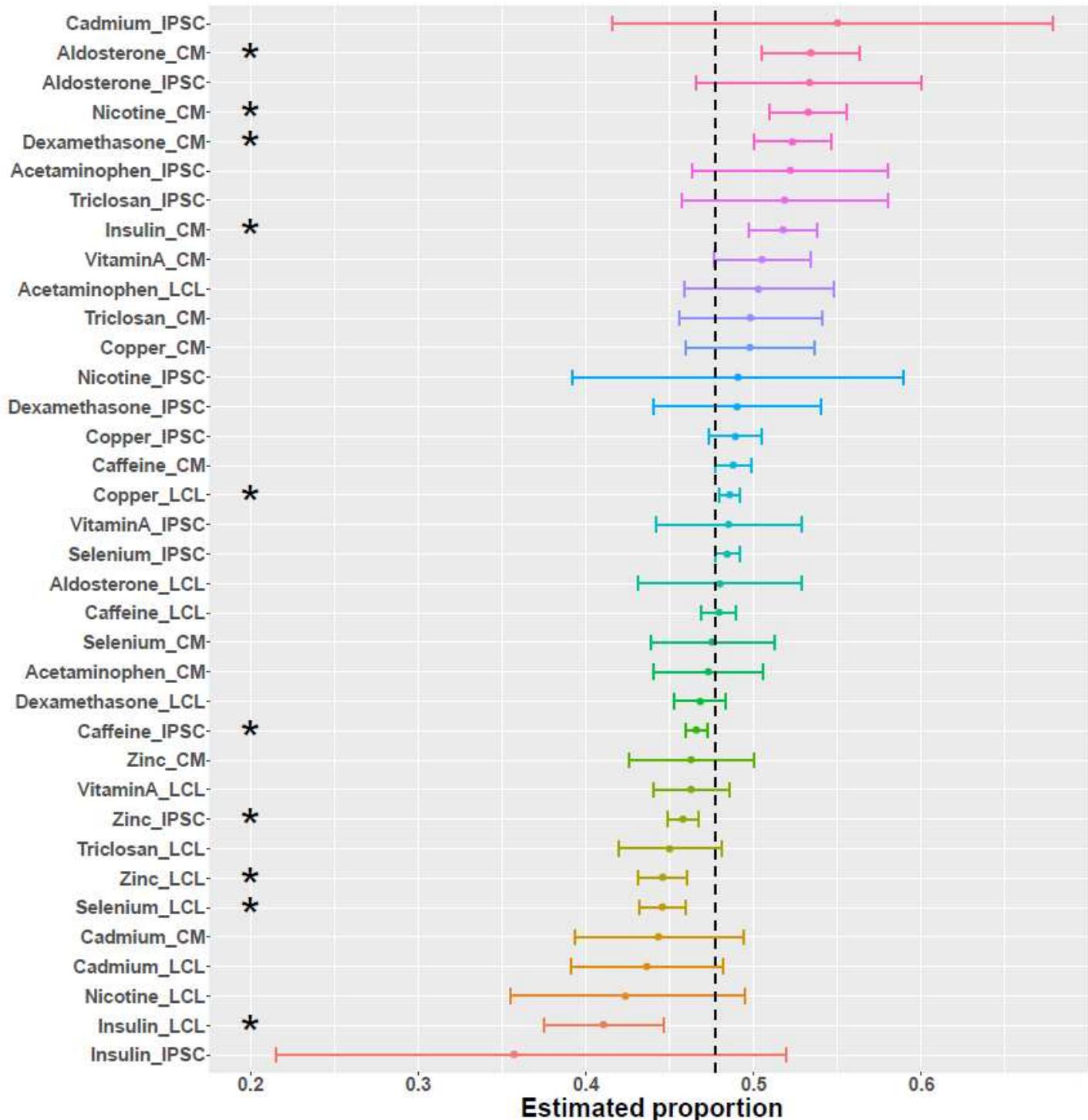
Figure S3: Direction of differential splicing Forest plot of the proportion of differential splicing events which result in intron retention as opposed to intron excision. An asterisk indicates a significance at FDR < 10%.
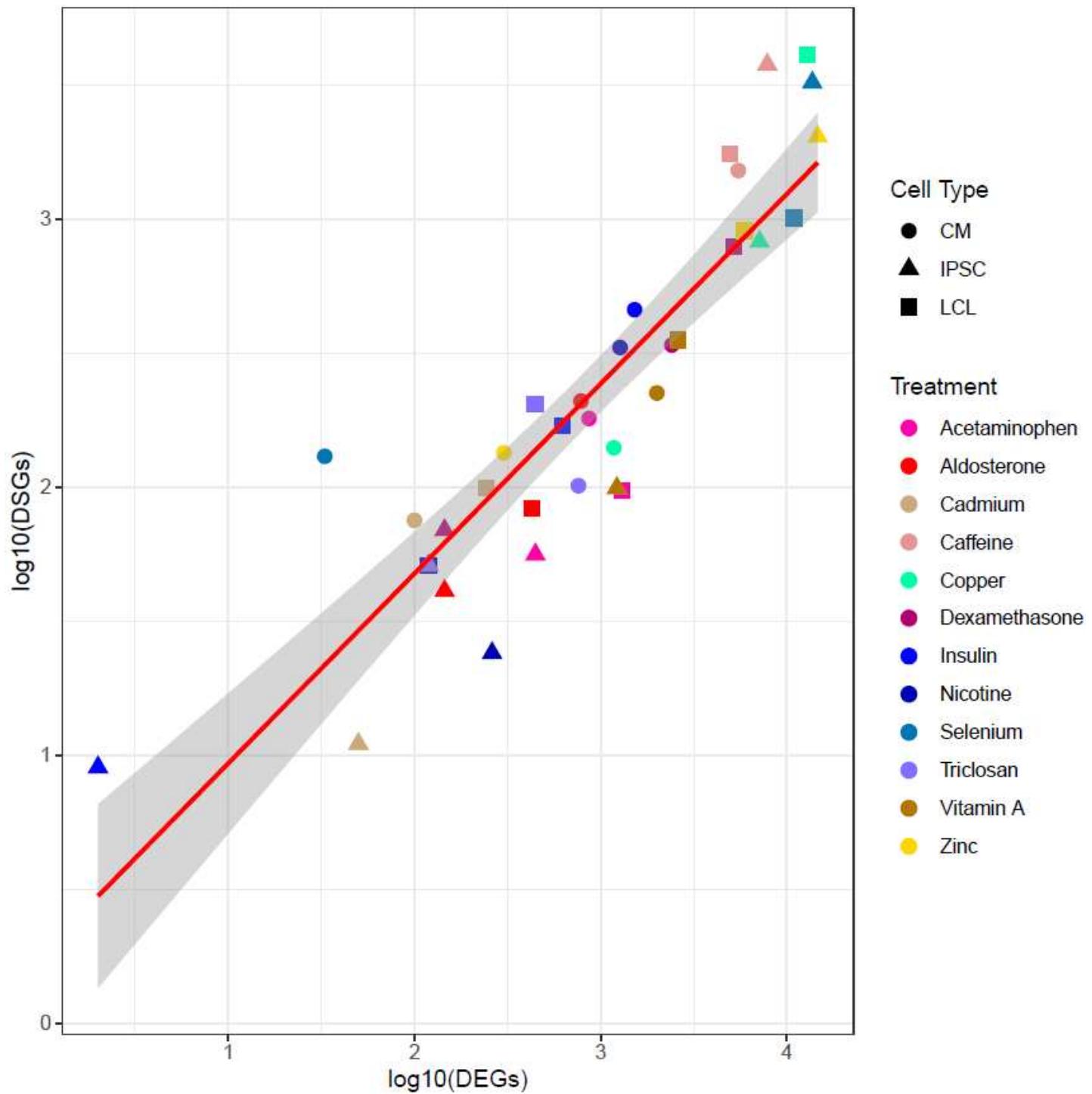
Figure S4: Correlation between number of DEGs and DSGs Scatterplot between the number of differentially expressed (DEGs) and differentially spliced (DSGs) genes. Each dot represents a unique cell type-environment, with shape and color representing cell type and treatment, respectively.
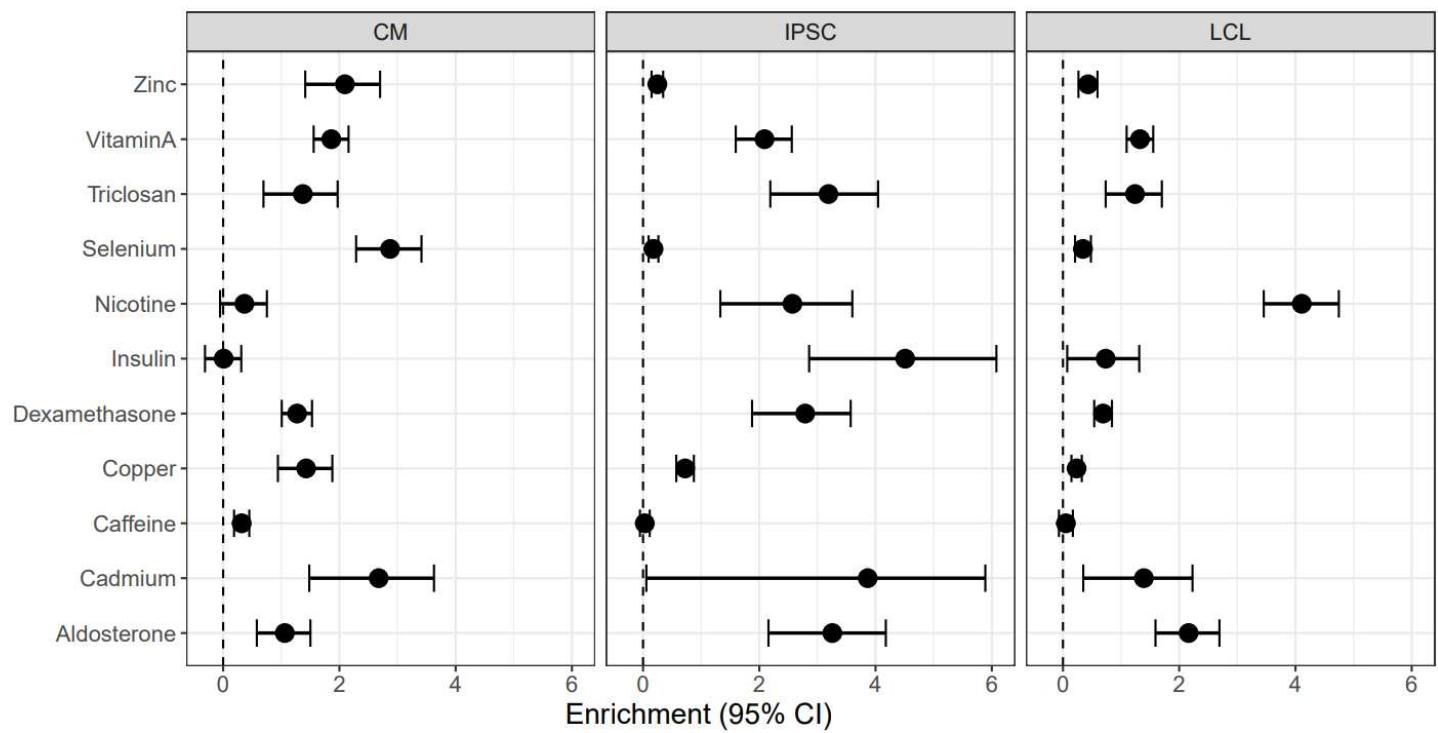
Figure S5: Enrichment of DSGs in DEGs Enrichment odds ratios and 95% confidence intervals have been log-transformed.
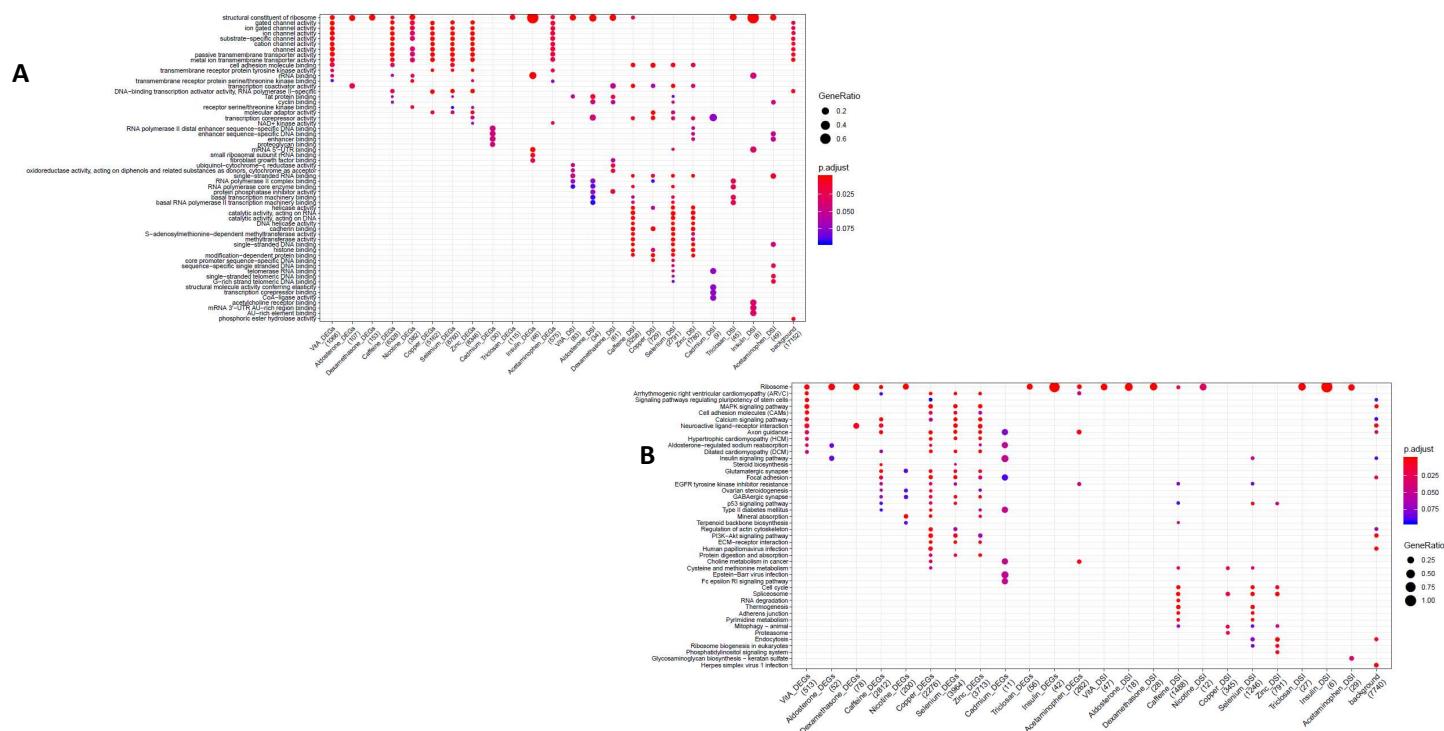
Figure S6: GO and KEGG enrichment dotplot for IPSCs Dotplot generated by ClusterProfiler depicting the top enriched GO (A) and KEGG (B) terms of CM DEGs and DSGs across environments, as well as within a background list composed by all tested genes.
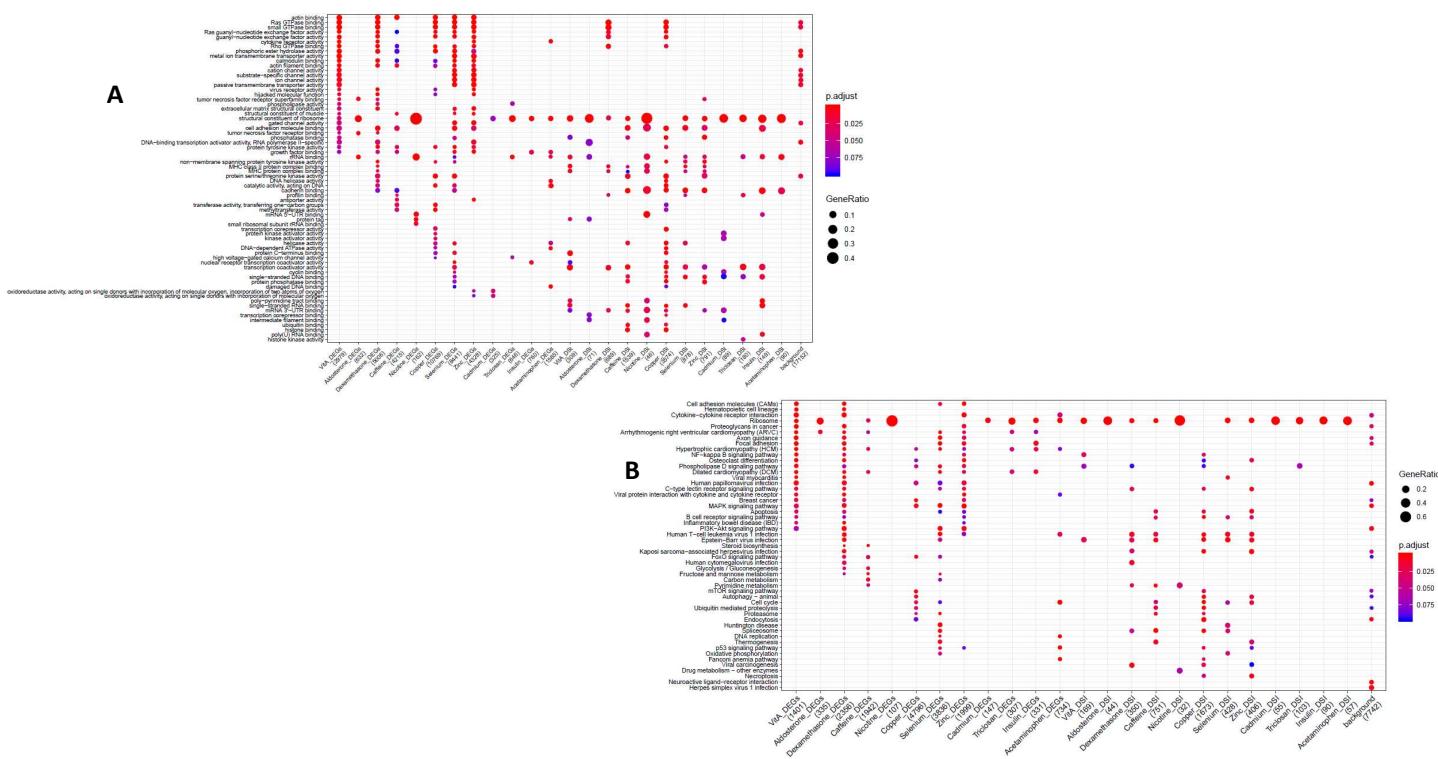
Figure S7: GO and KEGG enrichment dotplot for LCLs Dotplot generated by ClusterProfiler depicting the top enriched GO (A) and KEGG (B) terms of CM DEGs and DSGs across environments, as well as within a background list composed by all tested genes.

53
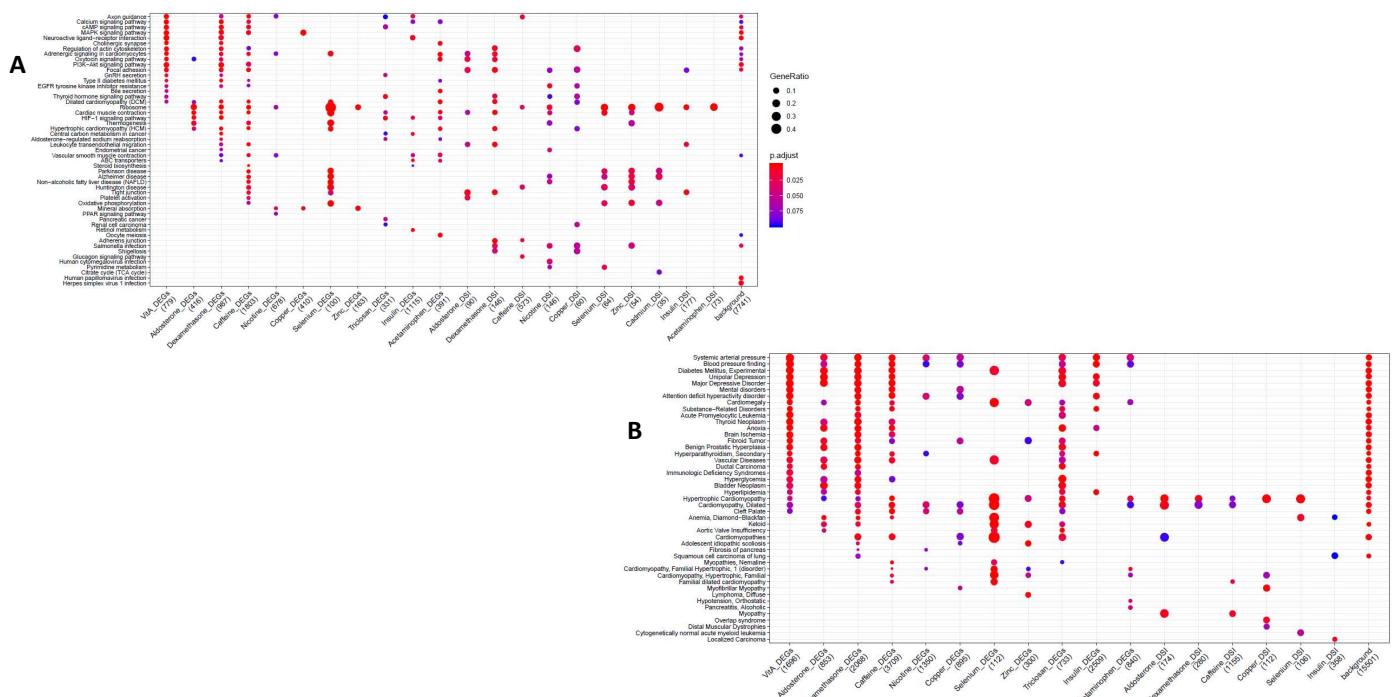
Figure S8: KEGG and DGN enrichment dotplot for CMs Dotplot generated by ClusterProfiler depicting the top enriched KEGG (A) and DGN (B) terms of CM DEGs and DSGs across environments, as well as within a background list composed by all tested genes.
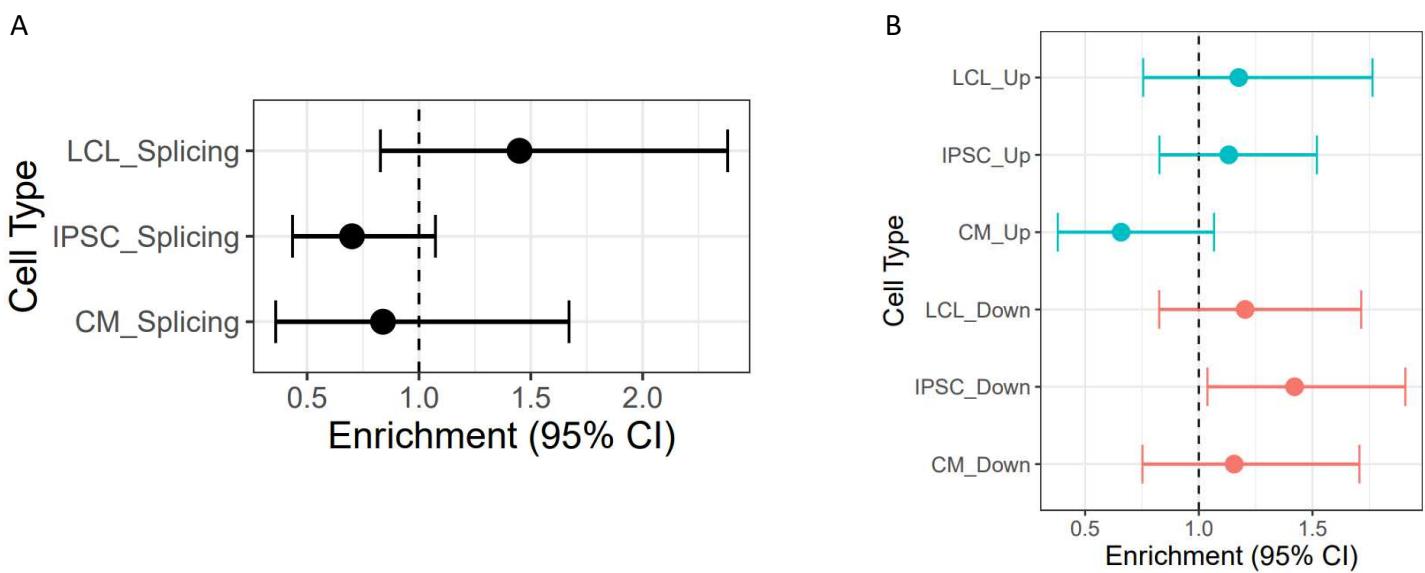
Figure S9: Enrichment of cASE in DSGs and DEGs cASE is matched by treatment to DSGs and DEGs.

54
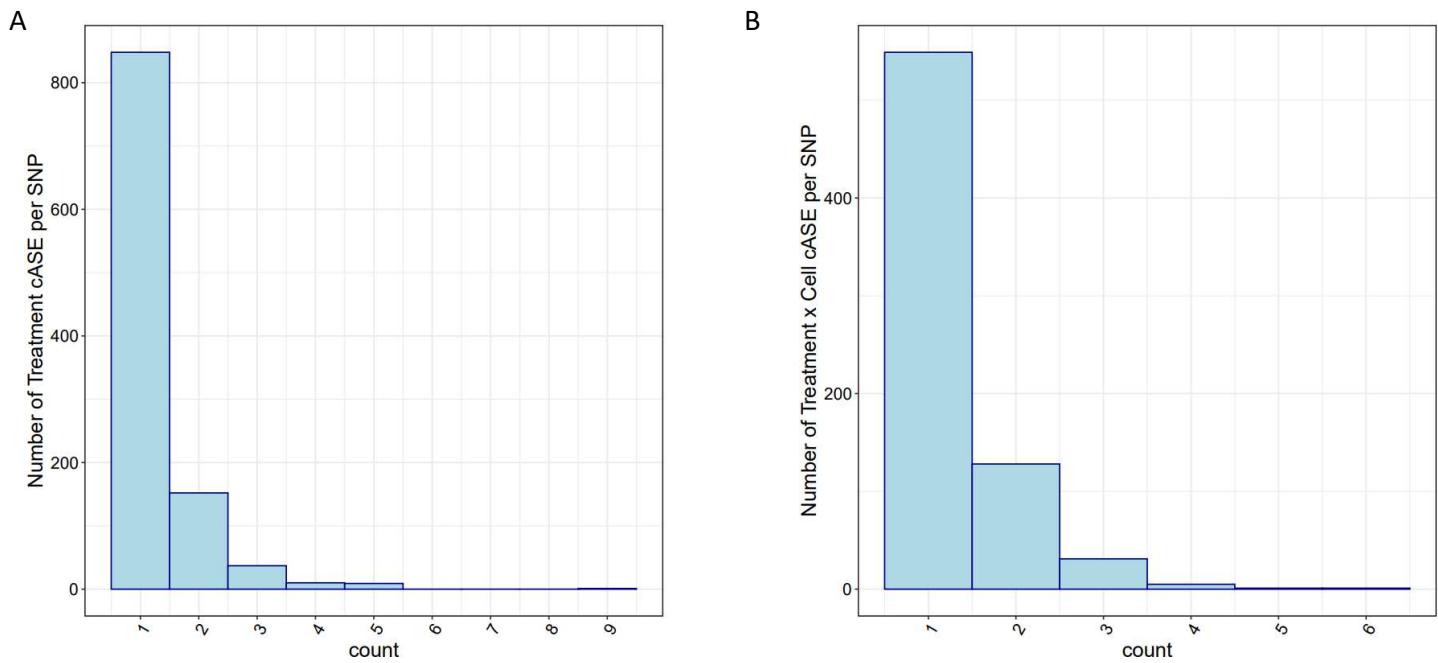
Figure S10: cASE per SNP For SNPs with significant cASE, barplots depict the number of **(A)** treatment cASE and **(B)** treatment by cell type cASE per SNP.
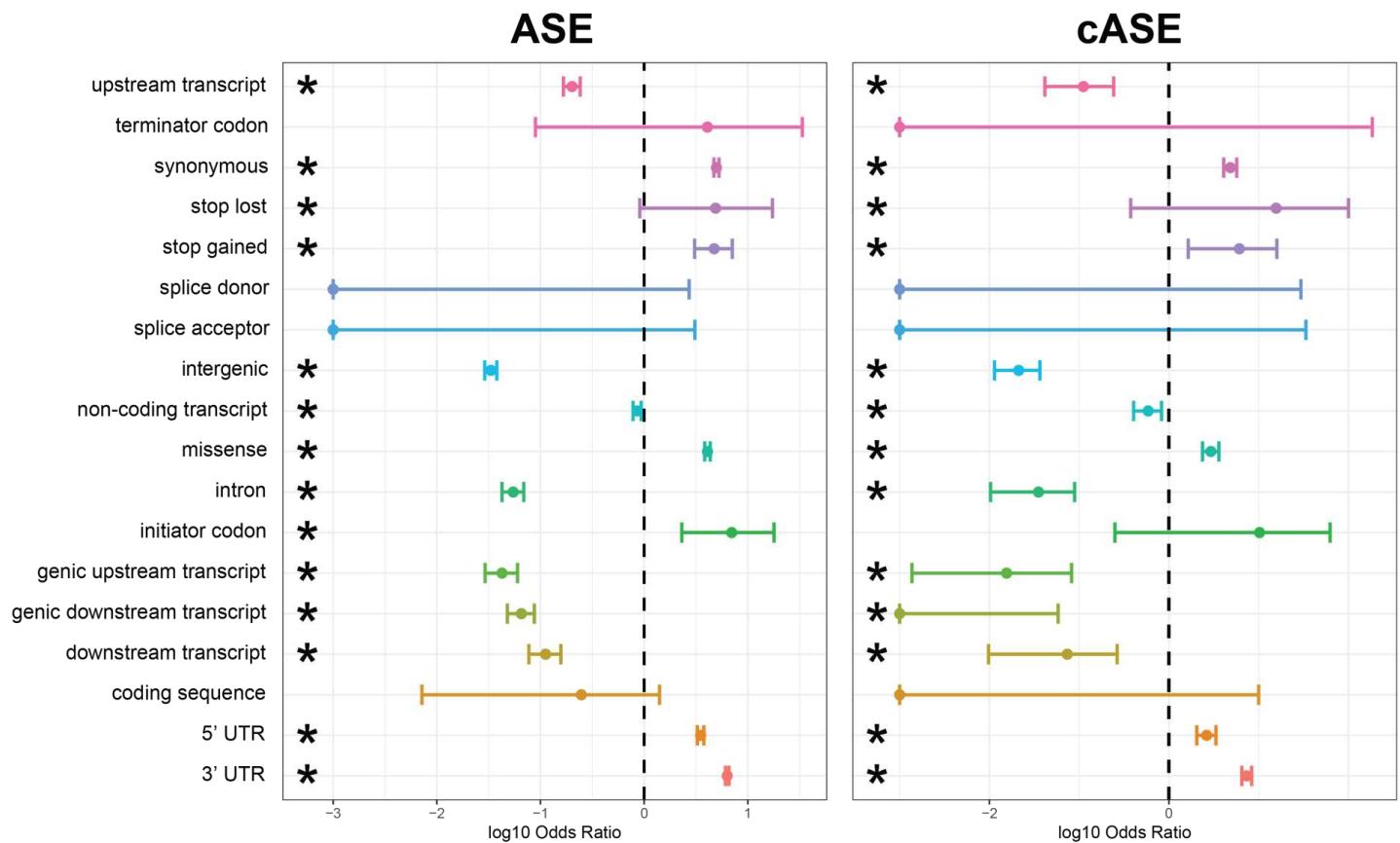
Figure S11: Enrichment and depletion of ASE/cASE SNPs within different genomic regions Full forest plot from Figure 3 with all genomic categories which were tested for an enrichment and depletion of ASE and cASE SNPs with respect to all SNPs tested.
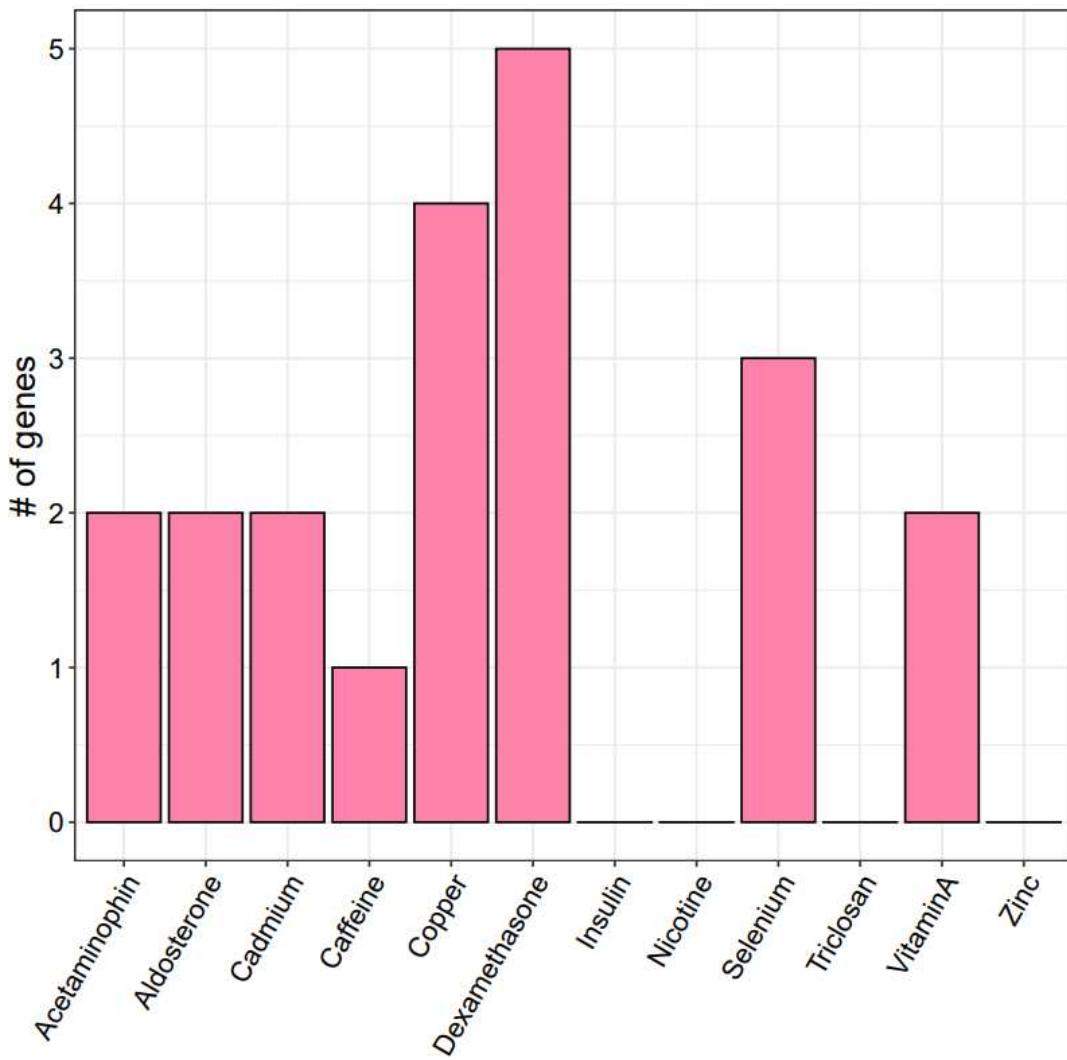
Figure S12: Overlap between atrial fibrillation GWAS prioritized genes and treatment cASE Number of treatment cASE genes which overlap with 151 genes prioritized from an atrial fibrillation GWAS.
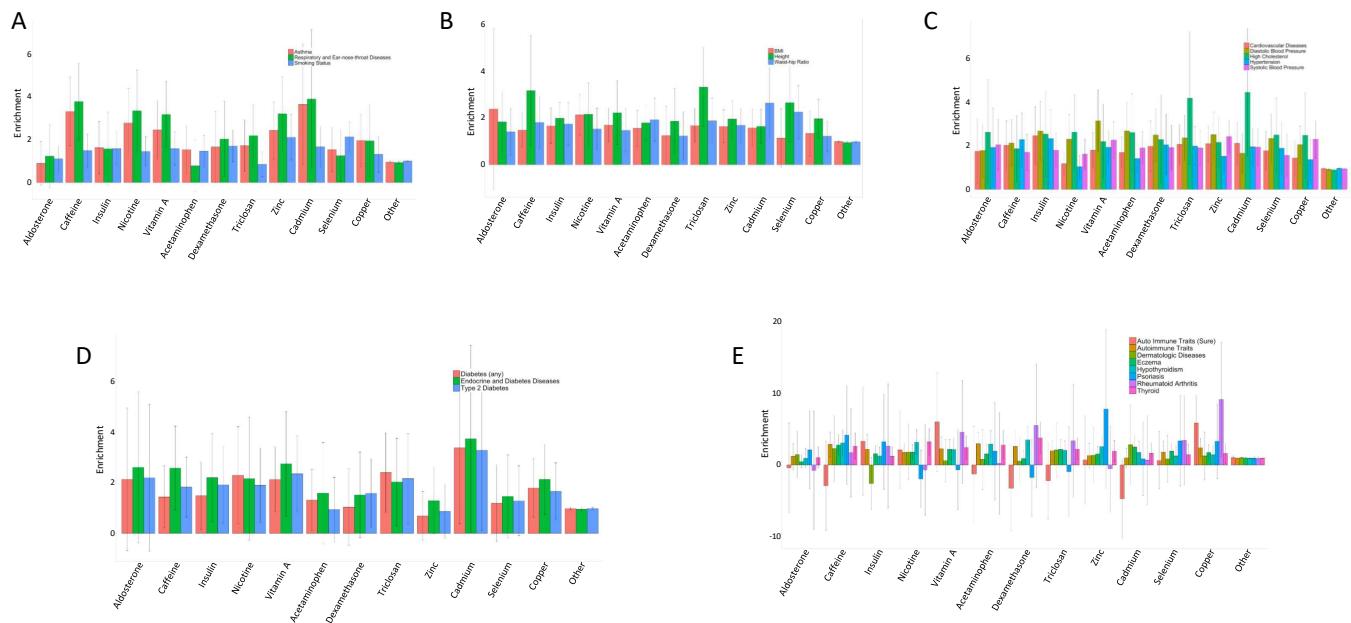
Figure S13: Enrichment of treatment cASE annotations in 22 GWAS Enrichment and 95% confidence intervals for treatment cASE annotations in **A** respiratory diseases, **B** body size traits, **C** cardiovascular traits, **D** endocrine diseases, and **E** autoimmune diseases.
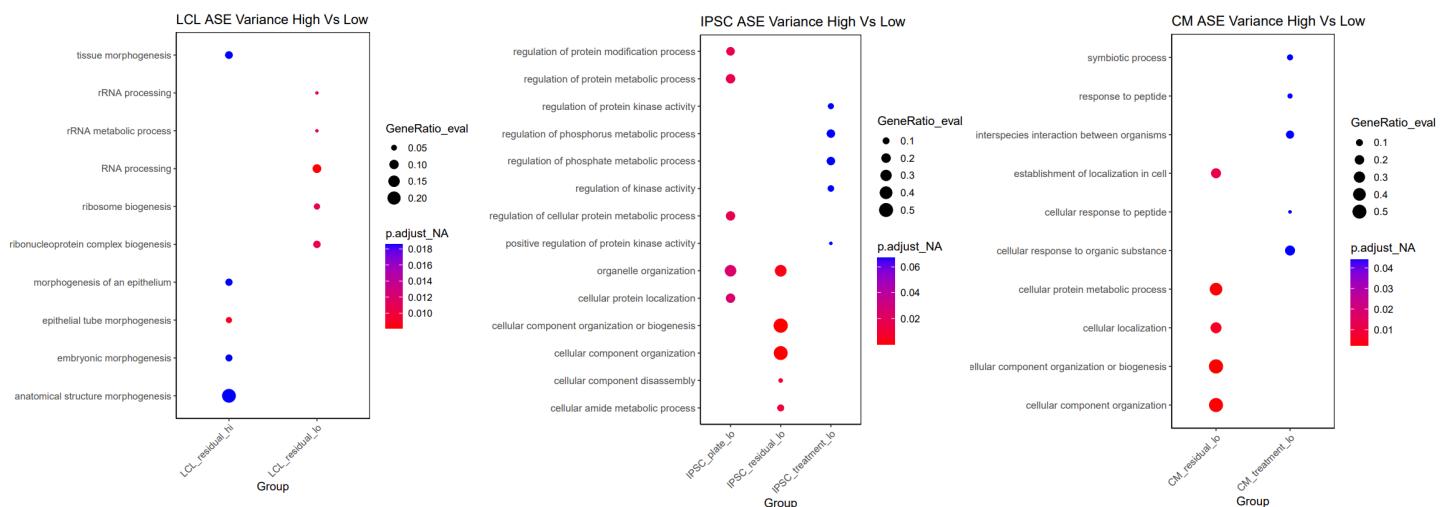

Figure S14: Gene ontology for ASE variance For each cell type, the 5 most significant GO terms are plotted for each group.
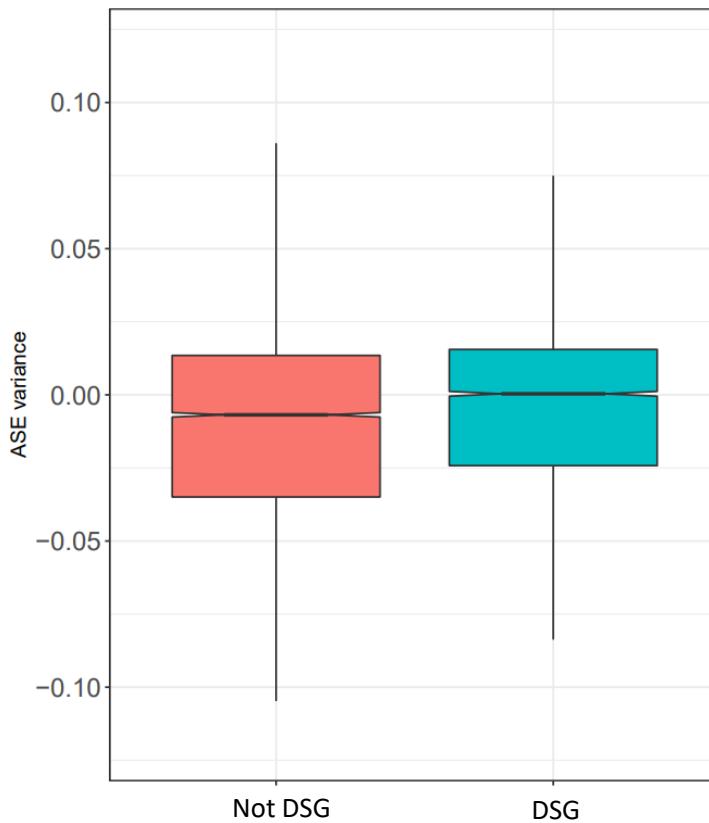
Figure S15: Treatment variance of DSGs Treatment variance of ASE based on whether the gene is differentially spliced in response to any condition.
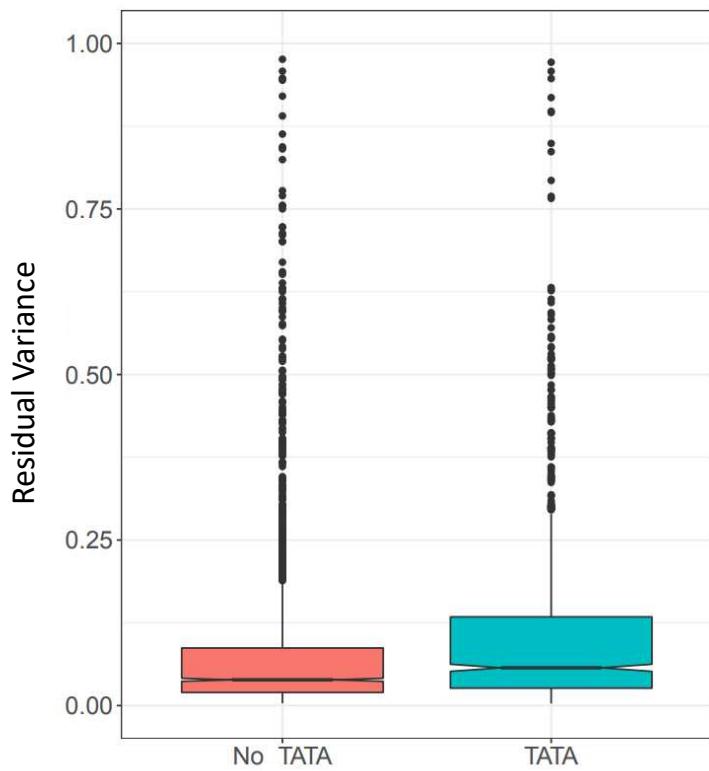
Figure S16: **Gene expression residual variance by presence or absence of TATA box**