

Dynamic effects of genetic variation on gene expression revealed following hypoxic stress in cardiomyocytes

Michelle C. Ward^{1*#}, Nicholas E. Banovich^{2*+}, Abhishek Sarkar², Matthew Stephens^{2,3}, Yoav Gilad^{1,2#}

¹Department of Medicine, University of Chicago, Chicago, IL, USA

²Department of Human Genetics, University of Chicago, Chicago, IL, USA

³Department of Statistics, University of Chicago, Chicago, IL, USA

* These authors contributed equally to the work

^{\$}Present address: Department of Biochemistry and Molecular Biology, University of Texas Medical Branch, Galveston, TX, USA

⁺Present address: Integrated Cancer Genomics Division, Translational Genomics Research Institute, Phoenix, AZ, USA

[#]Correspondence should be addressed to M.C.W (miward@utmb.edu), and Y.G (gilad@uchicago.edu).

Abstract

One life-threatening outcome of cardiovascular disease is myocardial infarction, where cardiomyocytes are deprived of oxygen. To study inter-individual differences in response to hypoxia, we established an *in vitro* model of induced pluripotent stem cell-derived cardiomyocytes from 15 individuals. We measured gene expression levels, chromatin accessibility, and methylation levels in four culturing conditions that correspond to normoxia, hypoxia and short or long-term re-oxygenation. We characterized thousands of gene regulatory changes as the cells transition between conditions. Using available genotypes, we identified 1,573 genes with a *cis* expression quantitative locus (eQTL) in at least one condition, as well as 367 dynamic eQTLs, which are classified as eQTLs in at least one, but not in all conditions. A subset of genes with dynamic eQTLs is associated with complex traits and disease. Our data demonstrate how dynamic genetic effects on gene expression, which are likely relevant for disease, can be uncovered under stress.

Introduction

Cardiovascular disease (CVD), which ultimately damages heart muscle, is a leading cause of death worldwide (WHO, 2018). CVD encompasses a range of pathologies including myocardial infarction (MI), where ischemia or a lack of oxygen delivery to energy-demanding cardiomyocytes results in cellular stress, irreparable damage and cell death. Genome-wide association studies (GWAS) have identified hundreds of loci associated with coronary artery disease (Nikpay et al., 2015), MI, and heart failure (Shah, 2019), indicating the potential contribution of specific genetic variants to disease risk. Most disease-associated loci do not localize within coding regions of the genome, often making inference about the molecular mechanisms of disease challenging. That said, because most GWAS loci fall within non-coding regions, these variants are thought to have a role in regulating gene expression. One of the main goals of the Genotype-Tissue Expression (GTEx) project has been to bridge the gap between genotype and organismal level phenotypes by identifying associations between genetic variants and intermediate molecular level phenotypes such as gene expression levels (Consortium et al., 2017). The GTEx project has identified tens of thousands of expression quantitative trait loci (eQTLs); namely, variants that are associated with changes in gene expression levels, across dozens of tissues including ventricular and atrial samples from the heart. However, the GTEx project has reported that most eQTLs are shared across tissues, suggesting that they probably do not contribute to disease in a tissue-specific manner.

It is becoming increasingly evident that many genetic variants that are not associated with gene expression levels at steady state, may be found to impact dynamic programs of gene expression in specific contexts. This includes specific developmental stages (Strober et al., 2019), or specific exposure to an environmental stimulus such as endoplasmic reticulum stress (Dombroski et al., 2010), hormone treatment (Maranville et al., 2011), radiation-induced cell death (Smirnov et al., 2012), vitamin D exposure (Kariuki et al., 2016), drug-induced cardiotoxicity (Knowles et al., 2018), and response to infection (Alasoo et al., 2018; Barreiro et al., 2012; Caliskan et al., 2015; Kim-Hellmuth et al., 2017; Manry et al., 2017; Nedelec et al., 2016). The studies of context-specific dynamic eQTLs highlight the need to determine the effects of genetic variants in the relevant environment. Therefore, if we are to fully understand the effects of genetic variation on disease, we must assay disease-relevant cell types and disease-relevant perturbations. Most of the aforementioned studies were performed in whole blood or immune cells, which means that there are many cell types and disease-relevant states that have yet to be explored.

With advances in pluripotent stem cell technology, we can now generate otherwise largely inaccessible human cell types through directed differentiation of induced pluripotent stem cells (iPSCs) reprogrammed from easily accessible tissues such as fibroblasts or B-cells. One of the advantages of iPSC-derived cell types as a model system is that the environment can be controlled, and thus we can specifically test for genetic effects on molecular phenotypes in response to controlled perturbation. This is particularly useful for studies of complex diseases such as CVD, which result from a combination of both genetic and environmental factors.

The heart is a complex tissue consisting of multiple cell types, yet the bulk of the volume of the heart is comprised of cardiomyocytes (Donovan, 2019; Pinto et al., 2016), which are particularly susceptible to oxygen deprivation given their high metabolic activity. iPSC-derived cardiomyocytes (iPSC-CMs) have been shown to be a useful model for studying genetic effects on various cardiovascular traits and diseases, as well for studying gene regulation (Banovich et al., 2018; Benaglio et al., 2019; Brodehl et al., 2019; BurrIDGE et al., 2016; de la Roche et al., 2019; Ma et al., 2018; McDermott-Roe et al., 2019; Panopoulos et al., 2017; Pavlovic et al., 2018; Ward and Gilad, 2019).

In humans, coronary artery disease can lead to MI (Dzau et al., 2006) which results in ischemia and a lack of oxygen delivery to energy-demanding cardiomyocytes. Given the inability of cardiomyocytes to regenerate, this cellular stress ultimately leads to tissue damage. Advances in

treatment for MI, such as surgery to restore blood flow and oxygen to occluded arteries, have improved clinical outcomes. However, a rapid increase in oxygen levels post-MI can generate reactive oxygen species leading to ischemia-reperfusion (I/R) injury (Giordano, 2005). Both MI and I/R injury can thus ultimately influence the amount of damage in the heart. iPSC-CMs allow us to mimic the I/R injury process *in vitro* by manipulating the oxygen levels that cardiomyocytes are exposed to *in vivo*.

We thus designed a study aimed at developing an understanding of the genetic determinants of the response to a universal cellular stress, oxygen deprivation, in a disease-relevant cell type, mimicking a disease-relevant process. To do so, we established an *in vitro* model of oxygen deprivation (hypoxia) and re-oxygenation in a panel of iPSC-CMs from 15 genotyped individuals (Banovich et al., 2018). We collected data for three molecular level phenotypes: gene expression, chromatin accessibility and DNA methylation to understand both the genetic and regulatory responses to this cellular stress. This framework allowed us to identify eQTLs that are not evident at steady state, and assess their association with complex traits and disease.

Results

We differentiated iPSC-CMs from iPSCs of 15 Yoruba individuals that were part of the HapMap project (Banovich et al., 2018). To obtain a measure of variance associated with the differentiation process, and to more effectively account for batch effects, we replicated the iPSC-CM differentiation from three individuals three times, yielding 21 differentiation experiments in total. The proportion of iPSC-CMs in each cell culture was enriched by metabolic purification (see Methods). iPSC-CMs were matured by electrical pulsing and maintenance in cell culture for 30 days. On Day 30, the median cardiomyocyte purity was 81% (40-97% range), determined by flow cytometry as the proportion of cells that were positive for the cardiac-specific marker, TNNT2 (FigS1; STable; See methods).

We studied the response of the iPSC-CMs to hypoxia and re-oxygenation (Fig1A). To do so, we first cultured the iPSC-CMs at oxygen levels that are close to physiological oxygen levels (10% oxygen - Condition A) for seven days. We then subjected the iPSC-CMs to six hours of hypoxia (1% oxygen - Condition B), followed by re-oxygenation for 6 hours (10% oxygen - Condition C), or 24 hours (10% oxygen - Condition D) as previously described (Ward and Gilad, 2019). Oxygen levels were reproducibly controlled in cell culture (Fig1B, STable). In order to determine whether

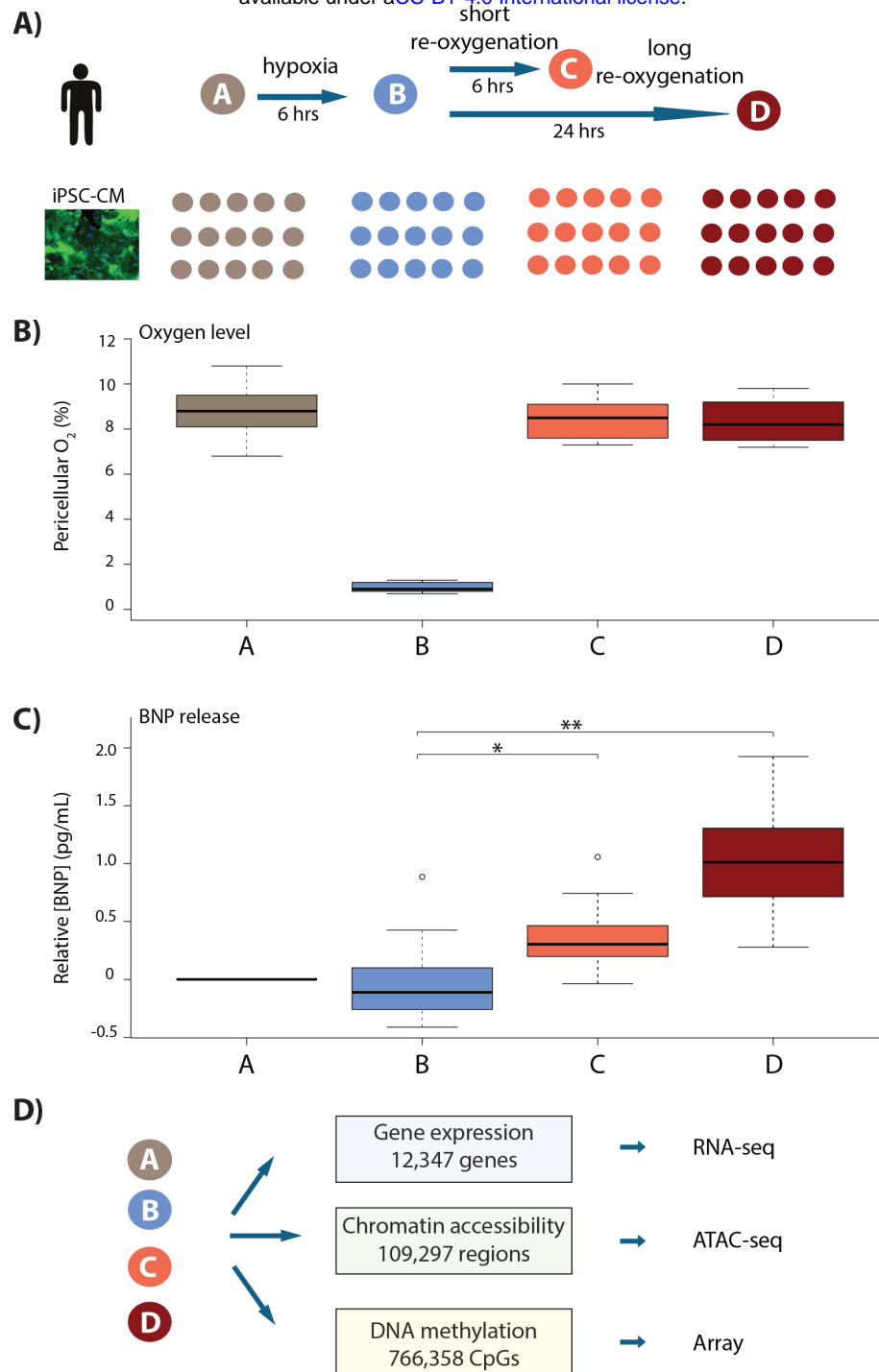


Figure 1: iPSC-derived cardiomyocytes elicit a cellular response to hypoxia. (A) Experimental design of the study. Cardiomyocytes differentiated from iPSCs (iPSC-CMs) from 15 Yoruba individuals were cultured in normoxic conditions (10% oxygen - condition A) and subjected to 6 hours of hypoxia (1% oxygen - condition B) followed by 6 and 24 hours of re-oxygenation (10% oxygen - conditions C and D). Immunocytochemistry of a representative cardiomyocyte culture where green: TNNT2; blue: nuclei. **(B)** Peri-cellular oxygen levels of each condition. Each point represents one individual undergoing the oxygen stress experiment. **(C)** Relative levels of BNP, a marker for cardiac stress, released into cell culture media. Asterisk denotes a statistically significant difference in BNP release (* $p < 0.05$, ** $p < 0.005$). **(D)** Molecular phenotypes collected from each individual in each condition.

the cardiomyocytes were affected by the changes in oxygen levels, we measured the enzymatic activity of released lactate dehydrogenase throughout the experiment, as a proxy for cytotoxicity. We also measured released BNP, a clinical marker of heart failure (Maeda et al., 1998). As expected, both cytotoxicity ($P = 0.01$, FigS2) and BNP ($P = 5 \times 10^{-6}$, Fig1C) levels increased following hypoxia and long-term re-oxygenation.

With this system established, we sought to understand the contribution of the global gene regulatory response to the molecular and cellular response to hypoxia and re-oxygenation. To do so, we collected global gene expression data (using RNA-seq; $n=15$), chromatin accessibility data (using ATAC-seq; $n = 14$), and DNA methylation data (using the EPIC arrays; $n = 13$; Fig1D) in each condition. With these data we studied both the gene regulatory response to oxygen perturbation, as well as the interaction of the response with the underlying genotype of the assayed individuals.

Gene expression changes in response to hypoxia and re-oxygenation

We first sought to identify those genes important for regulating the response by analyzing the gene expression (RNA-seq) data. We processed samples in batches as described in STable and mapped and filtered sequencing reads to prevent allelic mapping biases (FigS3; TableS1; See methods; (van de Geijn et al., 2015)). We observed that one sample (18852A) was a clear outlier when comparing read counts for 18,226 autosomal genes across all samples, and thus excluded it from further analysis (FigS4). We filtered out genes with low expression levels (See methods) to yield a final set with data from 12,347 expressed genes (See methods). We performed a number of correlation-based analyses using the data from the technical replicates (FigS5), and confirmed that the quality of the data is high and that, in line with our flow cytometry data, our iPSC-CMs express a range of cardiomyocyte marker genes including *MYH7* and *TNNT2* (FigS6).

We took advantage of the fact that we have replicate experiments from three individuals to correct the data for unwanted variation (See methods; (Risso et al., 2014)). Following this procedure, our samples clustered both by oxygen level and individual (FigS7). To identify genes that respond to hypoxia and re-oxygenation, we first tested for differential expression between pairs of conditions using a linear model with a fixed effect for 'condition', a random effect for 'individual', and four unwanted factors of variation, learned from the data, as covariates. At an FDR of 10%, we identified thousands of genes that are differentially expressed between conditions (A vs. B = 4,983,

B vs. C = 6,311; B vs. D = 6,792; A vs. D = 2,835; FigS8A-B). We used Cormotif (Wei et al., 2015) to classify 2,113 genes (17% of all expressed genes) as responding to hypoxia (Fig2, FigS8C-D). Response genes are enriched for genes previously identified to respond to hypoxia in a Caucasian population of individuals (Chi-squared test; $P < 2.2 \times 10^{-16}$; (Ward and Gilad, 2019)), and are highly enriched for gene ontologies in transcription-related processes (See methods, modified Fisher's exact test; $P = 1 \times 10^{-19}$).

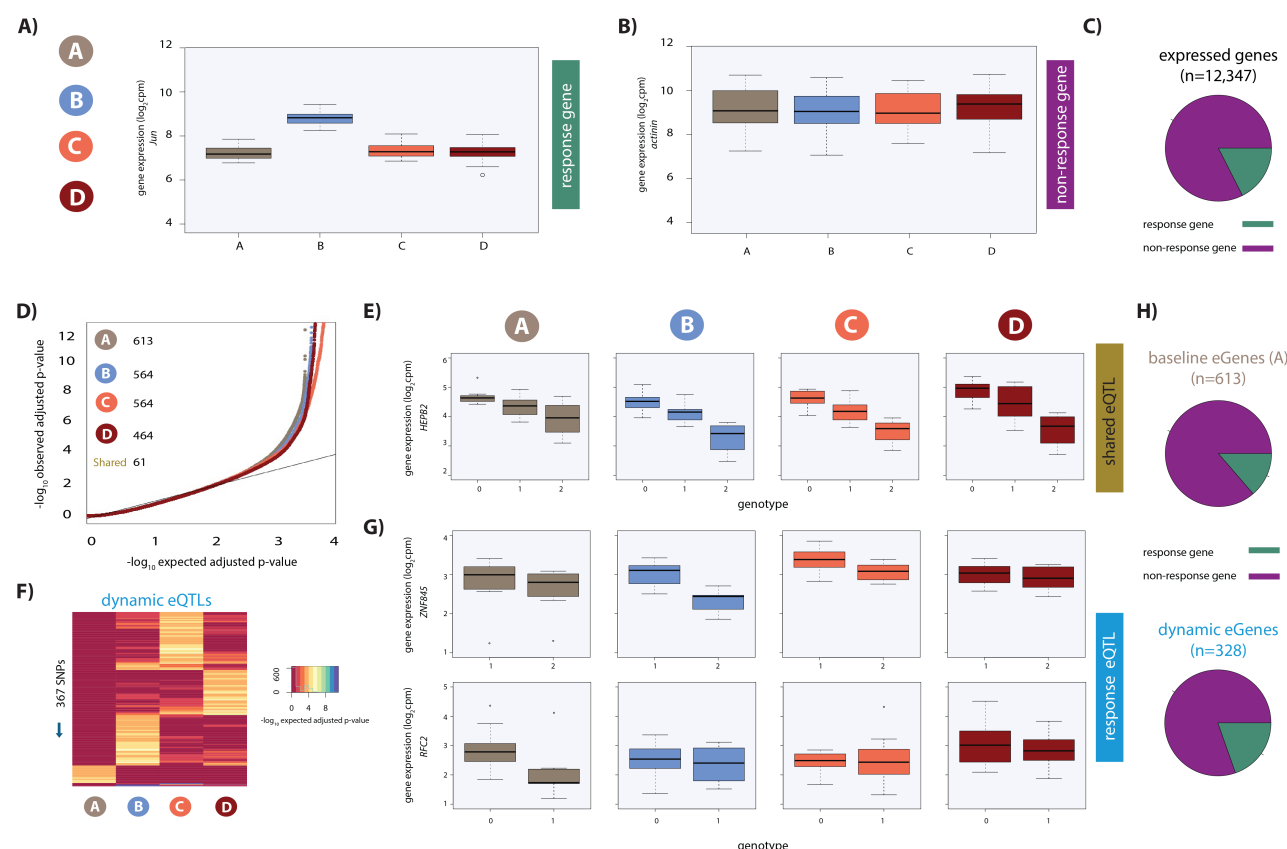


Figure 2: Hundreds of dynamic eQTLs are revealed following hypoxia and re-oxygenation. (A) Expression levels of *Jun*, a response gene, during the course of the experiment. (B) Expression levels of *actinin*, a non-response gene. (C) The proportion of all expressed genes that are classified as response genes (green), and non-response genes (magenta). (D) QQ plot illustrating an enrichment of associations between genetic variants and gene expression levels in each condition. Numbers represent the number of eGenes in each condition. (E) An example of a shared eQTL, *HEPB2*. (F) Heatmap illustrating the 367 SNPs that are classified as dynamic eQTLs. Each row represents a SNP that is an eQTL in at least one condition. Color represents the strength of the association p-value. (G) Examples of each of the two dynamic eQTL categories. Top panel: genes that become an eQTL following hypoxia e.g. *ZNF845*. Bottom panel: genes that are an eQTL at baseline but not following hypoxia e.g. *RFC2*. (H) The proportion of baseline eQTLs (all those identified in condition A), and dynamic eQTLs that are also response genes.

Dynamic eQTLs are revealed following hypoxia

Having established that oxygen stress initiates a transcriptional response affecting thousands of genes, we sought to identify eQTLs, either before or after oxygen stress. Using the combined haplotype test (CHT), an approach that leverages allele-specific information in small sample sizes (See methods; (van de Geijn et al., 2015)), we identified 1,573 genes with eQTLs (eGenes) in at least one condition ($q\text{-value} < 0.1$; A: 613; B: 564; C: 564 and D: 464; Fig2D). We refer to the 613 eGenes identified in condition A as baseline eGenes. We confirmed that genes whose expression levels are not significantly associated with a SNP in any condition show a largely uniform distribution of P-values in each condition (FigS9A), suggesting that, as expected given our study design, the overall power to detect eQTLs is similar across conditions.

Our goal was to identify dynamic eQTLs, which are either revealed or suppressed as the cells transition between conditions. Due to the small sample size of our study, we have incomplete power to detect eQTLs in any condition; thus, a naïve comparison of eQTLs classified as ‘significant’ across conditions will result in an over-estimation of the number of dynamic eQTLs. To address this challenge, we first considered eQTLs identified using a $q\text{-value} < 0.1$ in at least one condition, and visualized the P-value distributions of the corresponding eQTL associations in all other conditions. These P-value distributions are expected to be uniform if we had complete power to detect eQTLs in any condition (because in that theoretical case, even a naïve comparison of eQTLs classified as ‘significant’ across conditions will result in the identification of true condition-specific eQTLs). Due to incomplete power, this is clearly not the case (FigS9B); however, this distribution allowed us to choose a lenient secondary P-value-based cutoff, where values deviate from the uniform distribution, to classify dynamic QTLs (FigS9B).

We specifically focused on two dynamic scenarios. First, we defined suppressed eQTLs, as eQTLs that are identified in condition A at a $q\text{-value} < 0.1$ but not in any of the other conditions, with a P-value greater than 0.15 (37 instances; Fig2F). Second, we defined induced eQTLs as eQTLs identified in conditions B, C or D at a $q\text{-value} < 0.1$, but not in A, with a P-value greater than 0.15 (330 instances; Fig2G). This set of 367 dynamic eQTLs corresponds to 328 unique dynamic eGenes (See methods). While our choice of the particular statistical cutoffs is somewhat arbitrary, we can evaluate the false discovery rate associated with our chosen cutoff. Based on the P-value distributions of the corresponding eQTL associations in all other conditions, we estimate that our approach to classify dynamic eQTLs is associated with a false discovery rate of 48%. The

relatively high FDR associated with our choice of statistical cutoffs does not indicate that these loci are not eQTLs; rather it means that if we had a larger sample size, roughly half of our dynamic eQTLs should have been classified as eQTLs in more conditions, potentially in all of them.

We next wanted to determine whether the dynamic eQTLs we identified in iPSC-CMs are also eQTLs in primary heart tissue. To do so, we compared our list of eGenes to eGenes identified in tissue from two locations in the primary heart - left ventricle (LV) and atrial appendage (AA) - from hundreds of individuals in the GTEx study (Consortium et al., 2017). We found that the majority of eGenes present in all conditions ('shared', $n=61$) in our study are also eGenes in heart tissue (60% in LV, and 67% in AA). A smaller proportion of dynamic eGenes were also classified as primary heart tissue eGenes (38% in both LV and AA). Indeed, shared eGenes are more likely to be eGenes in heart tissue than dynamic eGenes (Chi-squared test; $P < 0.005$; FigS10A-B). Overall, 98% of dynamic eGenes are an eGene in at least one of 14 selected tissues assayed by the GTEx consortium, supporting our analysis indicating context-dependent inter-individual variation in expression (FigS10C).

Dynamic eGenes are enriched for response genes and transcription factors

To determine whether dynamic eGenes coincide with expression changes of the same genes following hypoxia, we integrated the results of our eQTL and differential expression analyses. In line with our previous findings, baseline eGenes, as well as LV and AA eGenes found in primary tissue (GTEx), are depleted for response genes (Chi-squared test; $P < 0.02$, (Ward and Gilad, 2019)). However, we found a significant enrichment in response genes amongst dynamic eGenes (61 of 328 genes; Chi-squared test; $P = 0.03$) when compared to baseline eGenes, suggesting that dynamic eQTLs often impact the regulation of genes that respond to hypoxic stress.

Given that thousands of genes are differentially expressed in response to hypoxia, and many of these genes correspond to transcription-related processes, we next investigated the role of transcription factors, which are likely to drive transcriptome differences in our system. We found a significant enrichment of annotated transcription factors amongst the genes responding to oxygen stress compared to non-response genes (327 of 1,639 annotated human TFs, chi-squared test; $P < 2 \times 10^{-16}$; FigS11). Given that stress affects transcription factor expression, we asked whether dynamic eGenes are also enriched for transcription factors. Indeed, transcription factors are enriched in dynamic eGenes compared to baseline eGenes (35 TFs; $P = 0.004$), including *MITF*

and *PPARA*, both of which are TFs that have been previously implicated in hypoxic response (Feige et al., 2011; Narravula and Colgan, 2001).

Chromatin accessibility changes following hypoxia and re-oxygenation

We next asked whether the hundreds of transcription factor expression changes following hypoxic stress are accompanied by global chromatin accessibility changes. To examine this, we performed ATAC-seq experiments to identify regions of open chromatin (we were only able to collect these data from 14 of the 15 individuals; STable). We filtered the ATAC-seq reads to include only those reads that map to the nuclear genome, and do not show allelic mapping biases (See methods, FigS12). We identified a set of open chromatin regions in each sample, and merged samples across individuals within each condition. Genomic regions identified as accessible in each condition were then merged to yield a set of 128,672 open chromatin regions across conditions (with a median length of 312 bp). Regions with low read counts were filtered out, resulting in a final set of 110,128 regions. Analysis of various metrics revealed the data to be of good quality (FigS13-14).

We sought to identify chromatin regions that are differentially accessible across pairs of conditions. Using a sensitive adaptive shrinkage based approach with a False Sign Rate of 10% (Stephens, 2017) we could not detect changes in accessibility between baseline and hypoxia; however, we identified 831 differentially accessible regions (DARs) between hypoxia and short-term re-oxygenation (BC-DARs; 429 regions with increased accessibility and 402 with decreased accessibility), and 71 DARs between hypoxia and long-term re-oxygenation (BD-DARs; Fig3A). There is a strong correlation in effect sizes between hypoxia and short-term re-oxygenation (BC-DARs), and hypoxia and long-term re-oxygenation (BD-DARs; Spearman correlation = 0.74), and 59 of the 71 BD-DARs are amongst the 831 BC-DARs, suggesting that most regions have returned to baseline levels of accessibility by the first re-oxygenation condition. We therefore considered the 831 BC-DARs, henceforth DARs, in further analysis. Manual inspection of accessibility levels at individual DARs, identified between hypoxia and re-oxygenation, shows that many of these regions appear to have small changes in accessibility between the baseline and hypoxic conditions, which is opposite to the direction of the effect between hypoxia and re-oxygenation. This includes a region within the intron of the *FOXO1* gene, a master regulator of the oxidative stress response (Fig3B). Global analysis reveals that there is a strong anti-correlation in the effect size between

these pairs of conditions across regions (Spearman correlation = -0.62; sign test $P = 4.6 \times 10^{-14}$; FigS15A-C).

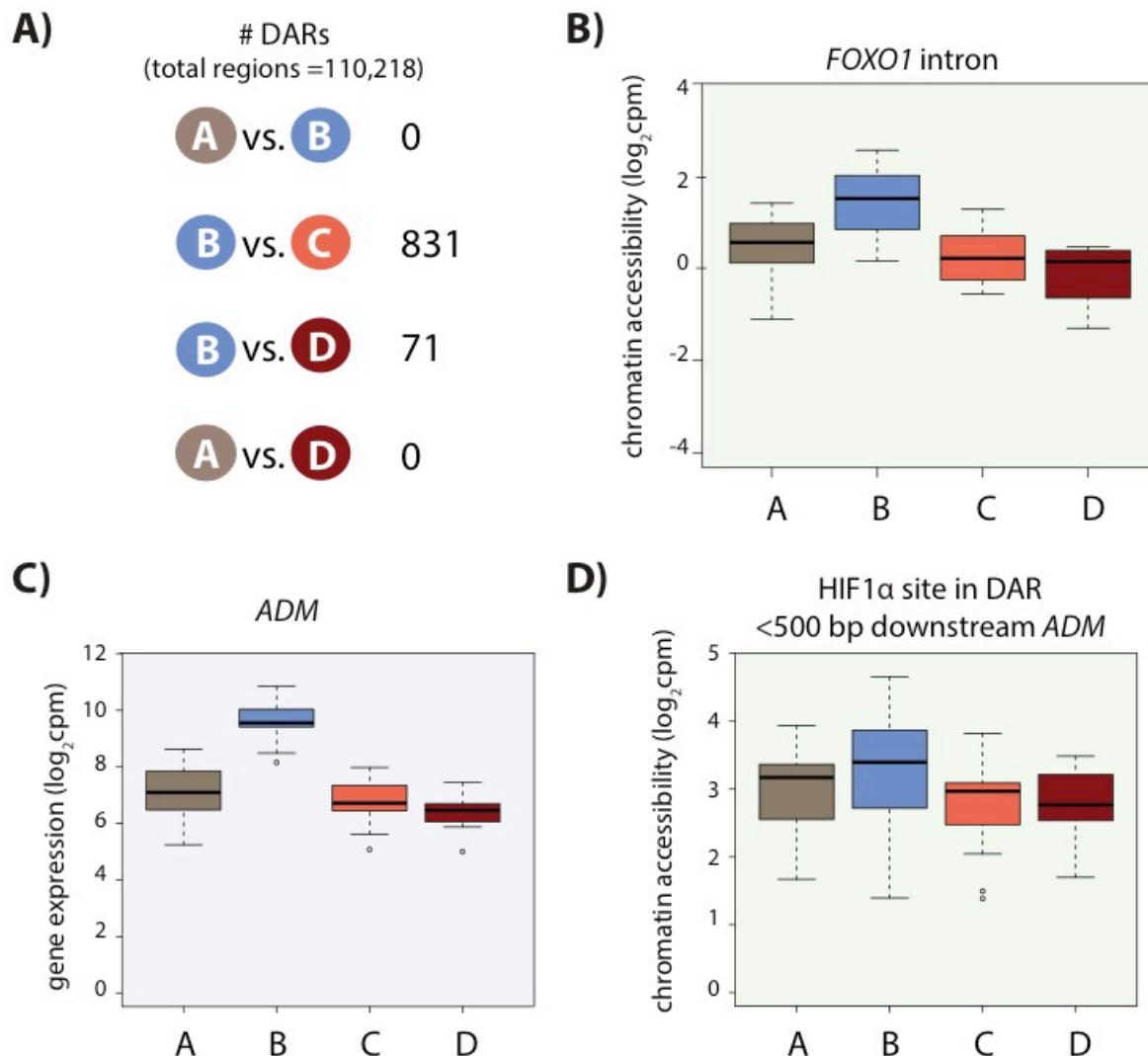


Figure 3: Chromatin accessibility changes following hypoxia and re-oxygenation. (A) Numbers of chromatin regions that are differentially accessible (DARs) between pairs of conditions. (B) Chromatin accessibility levels at a chromatin region within a *FOXO1* intron. (C) Expression levels of the hypoxia-responsive gene *ADM* following hypoxia. (D) Chromatin accessibility levels at a DAR, overlapping an induced HIF1 α -bound region, close to the *ADM* gene.

Linking chromatin accessibility changes with gene expression changes

Changes in chromatin accessibility likely cannot directly explain the thousands of gene expression changes that occur following hypoxia. However, we found that when considering a 50 kb window around the TSS of expressed genes, DARs are enriched near response genes compared to non-response genes (Chi-squared test; $P = 0.03$). 113 of 2,113 response genes have a DAR within 50 kb of the TSS. This set includes an accessible region, overlapping a HIF1 α site, within 500 bp of the 3' end of the classic hypoxia response gene, *ADM* (Fig3C-D).

We asked whether the changes in chromatin accessibility coincide with the appearance of dynamic eQTLs. We found that DARs are no more likely to be near dynamic eGenes than shared or baseline eGenes. In line with previous estimates of the proportion of eQTLs in open chromatin regions, 24 baseline eQTL SNPs (613 total SNPs) and 19 dynamic eQTL SNPs (367 total SNPs) overlap with accessible chromatin regions (Consortium et al., 2017). One dynamic eQTL SNP overlaps a DAR, near the actin filament binding protein gene, *FGD4*. This gene was also shown to be differentially expressed between children with congenital heart defects where the defect leads to a chronic hypoxic state (cyanotic disease), and children with a similar defect but where oxygen levels are not affected (acyanotic disease; (Ghorbel et al., 2010)).

To directly test whether there are genetic effects on chromatin accessibility, independent of gene expression, we sought to identify chromatin accessibility QTLs (caQTLs) i.e. genetic variants located within the 128,672 accessible regions, which coincide with different levels of accessibility based on genotype. We identified few caQTLs per condition (q-value < 0.1; A: 10, B: 1, C: 7, D: 6; FigS16A). Six of these caQTLs are classified as dynamic caQTLs i.e. induced or suppressed in response to hypoxia using the same definitions as used for the dynamic eQTLs, and include regions at the TSS of the mRNA decapping enzyme gene *DCPS*, and a region within 100 kb of the *C1Orf99* gene (FigS16B-C). These results suggest that gene expression changes, which respond to stress in a genotype-dependent or independent manner, occur largely in the absence of chromatin accessibility changes.

Genomic features associated with differentially accessible regions (DARs)

We next wanted to determine what distinguishes DARs from constitutively accessible regions. To do so, we investigated three classes of genomic features: 1) promoter- and enhancer-associated

marks, 2) transcription factor binding locations, and 3) underlying DNA sequence features. We found that DARs are more likely to overlap TSS than constitutively accessible regions (43% overlap vs. 11% overlap; $P < 2 \times 10^{-16}$; Fig4A) suggesting that DARs may be involved in the gene regulatory response. Indeed, DARs are more likely to coincide with active histone marks in left ventricle heart tissue than constitutively accessible regions (H3K4me3: 57% overlap DARs vs. 24% overlap constitutively accessible regions; chi-squared test; $P < 2.2 \times 10^{-16}$; H3K4me1: 86% overlap DARs vs. 52% overlap constitutively accessible regions; $P < 2.2 \times 10^{-16}$; Fig4B).

To determine whether sequence-specific hypoxia-responsive transcription factors associate with differentially accessible chromatin, we integrated DARs with published chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq) data for the well-studied hypoxia-inducible factors HIF1 α and HIF2 α (Schodel et al., 2011). Based on our inference, 234 of the 356 annotated HIF1 α sites (66%), and 150 of the 301 HIF2 α sites (50%) overlap with all accessible chromatin regions. We found that HIF1 α - and HIF2 α -bound regions are more likely to overlap the 831 differentially accessible regions than the 109,275 constitutively accessible regions (Fisher test; $P = 0.03$; Fig4C).

We next took an unbiased approach to identify transcription factor binding motifs that are enriched in DARs compared to all accessible regions (See methods). We found two motifs to be enriched in DARs compared to all regions (Fig4D). Motif 1 ($P = 2 \times 10^{-2}$) is recognized by *HTF4* and *TFE2*, both of which are non-response genes in our system. Motif 2 ($P = 6.2 \times 10^{-42}$) is posited to be recognized by ZN770, E2F3, and E2F4. Both ZN770 and E2F4 are response genes in our system. DARs arise between the hypoxia and re-oxygenation conditions, and E2F4 expression increases following re-oxygenation, suggesting that it may be involved in the response. To test this hypothesis, we obtained a published ChIP-seq data set for E2F4 (Lee et al., 2011), and overlapped the 16,245 E2F4-bound regions with DARs and constitutively accessible regions. E2F4-bound regions are significantly enriched in DARs compared to constitutively accessible regions (Chi-squared test; $P < 2.2 \times 10^{-16}$; Fig4D). E2F4 is important for survival following ischemia in neurons, and has been suggested to be an anti-apoptotic factor in cardiomyocytes (Dingar et al., 2012; Iyirhiaro et al., 2014).

To identify additional sequence features that associate with DARs, we asked whether transposable elements (TE), a potential source of regulatory sequence subjected to chromatin-level regulation, are enriched in these sites (Du et al., 2016; He et al., 2019). We found that while three of the main

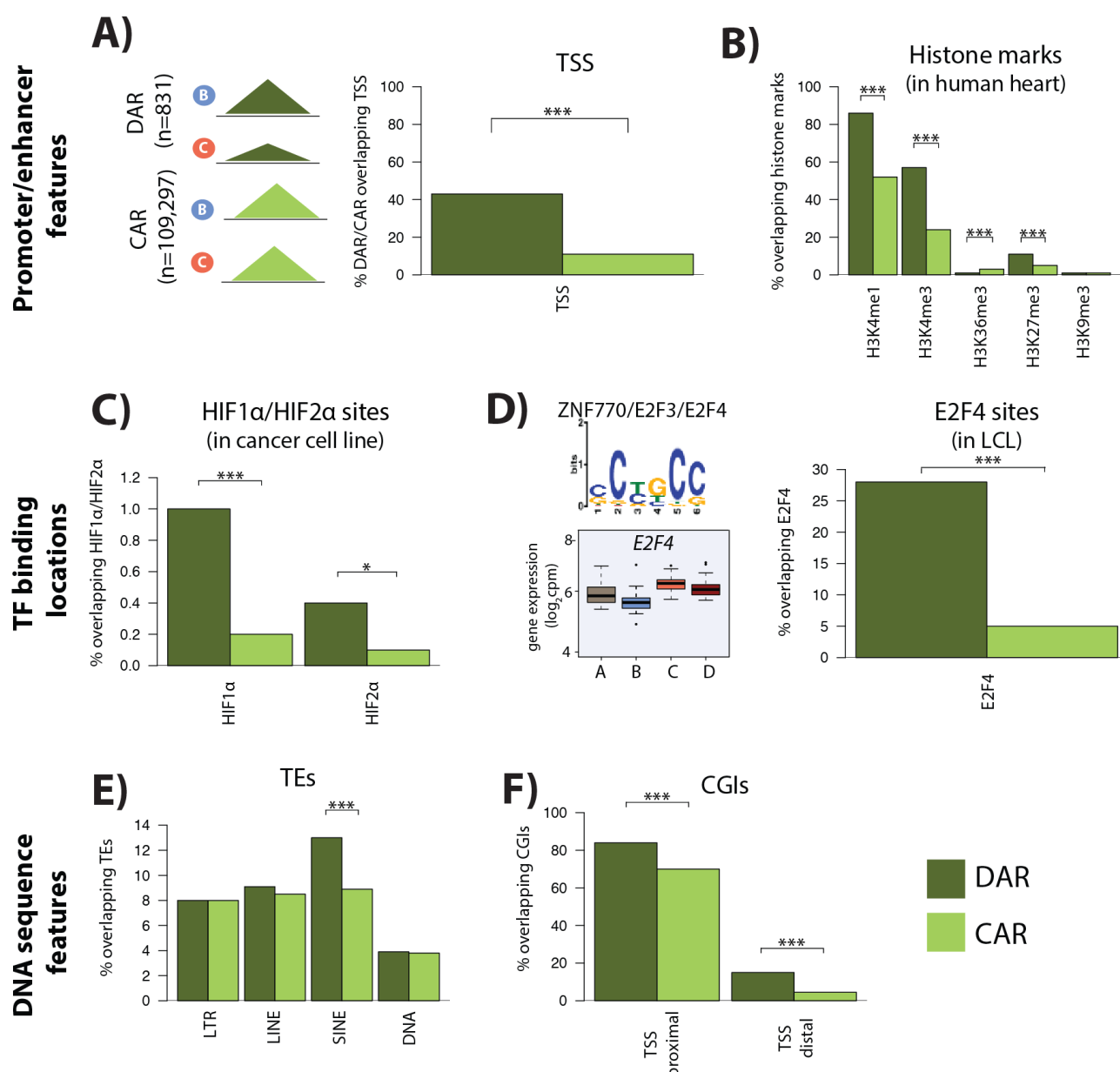


Figure 4: Differentially accessible regions are enriched for active chromatin features. (A) The proportion of differentially accessible regions (DARs) and constitutively accessible regions (CARs) that overlap with annotated TSS. **(B)** The proportion of DARs and CARs that overlap with the locations of histone marks determined by ChIP-seq in human heart tissue (Consortium, 2012). **(C)** The proportion of DARs and CARs that overlap with HIF1α and HIF2α binding locations determined by ChIP-seq in a breast cancer cell line (Schodel et al., 2011). **(D)** The most significant motif identified to be differentially enriched in DARs compared to all ARs that is putatively recognized by ZNF770, E2F3 and E2F4. We classify *E2F4* as a response gene and therefore determined the proportion of DARs and CARs that overlap with E2F4 ChIP-seq binding locations identified in a human LCL line (Lee et al., 2011). **(E)** The proportion of DARs and CARs that overlap four major transposable element (TE) classes – LTR, LINE, SINE, DNA. The proportion of DARs and CARs that overlap with CpG islands (CGIs) that is proximal to the TSS (+/- 2 kb from the TSS), and distal to the TSS. Asterisk denotes a statistically significant difference between DARs and CARs (* $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$).

TE classes, LINEs, LTRs, and DNA elements are similarly enriched in DARs compared to constitutively accessible regions; SINEs are specifically enriched in DARs ($P = 6.5 \times 10^{-6}$; Fig4E). There is an enrichment of both Alu and MIR SINE family members in DARs ($P = 3 \times 10^{-5}$ and $P = 0.006$). AluS elements, and the AluSq and AluSp sub-families, are particularly enriched within the Alu family ($P = 0.007$ and $P = 0.02$ respectively). A different cellular stress, heat shock, has previously been shown to remodel chromatin accessibility at Alu elements in cervical cancer cells (Kim et al., 2001).

As Alu and MIR TE sequences are notably CpG dense (Medstrand et al., 2002), we next asked about the enrichment of CpG-dense CpG islands (CGIs) in our differentially accessible regions. We found that CpG islands are enriched in DARs compared to constitutively accessible regions, whether these regions fall within 1 kb of TSS, which are typically enriched for CGIs, or not ($P < 2.2 \times 10^{-16}$; Fig4F).

DNA methylation state at stress-responsive genes and chromatin regions

Genes with CGI promoters are thought to allow flexibility in TSS choice compared to genes without CGI promoters (Carninci et al., 2006), and to allow for the rapid induction of gene expression in response to stimuli (Ramirez-Carrozzi et al., 2009). We therefore asked whether this promoter feature is enriched in the stress response genes. Indeed, we find that response genes are more likely to have CGI promoters than non-response genes (Chi-squared test; $P = 0.002$).

Given the enrichment of CpG islands in gene promoters and chromatin regions that are responsive to stress, we asked whether this feature corresponded to differences in CpG DNA methylation levels in these same regions. We measured global DNA methylation levels at 766,658 CpG sites in all conditions from 13 of our individuals (missing data from two of the individuals), together with 24 replicate samples from three individuals (See methods, STable). We found the expected bimodal distribution of DNA methylation Beta-values across CpGs (Beta-values represent the ratio of intensities between the methylated and unmethylated alleles; FigS17A). Additional analyses indicated the data to be of good quality (FigS17-18).

To determine whether steady-state DNA methylation levels mark genes or regions that will change their expression level in response to stress, we investigated baseline DNA methylation levels in the promoters of genes classified as response genes and non-response genes, as well as DARs and

constitutively accessible regions. To do so, we assessed the DNA methylation level at CpGs within 200 bp upstream of the TSS in the baseline condition. The majority of the assayed CpGs were hypomethylated with a median Beta-value of less than 0.2 across genes and regions (Fig5A). While there is no difference in median DNA methylation levels between response and non-response genes, we found that the median DNA methylation level is lower in DARs compared to all

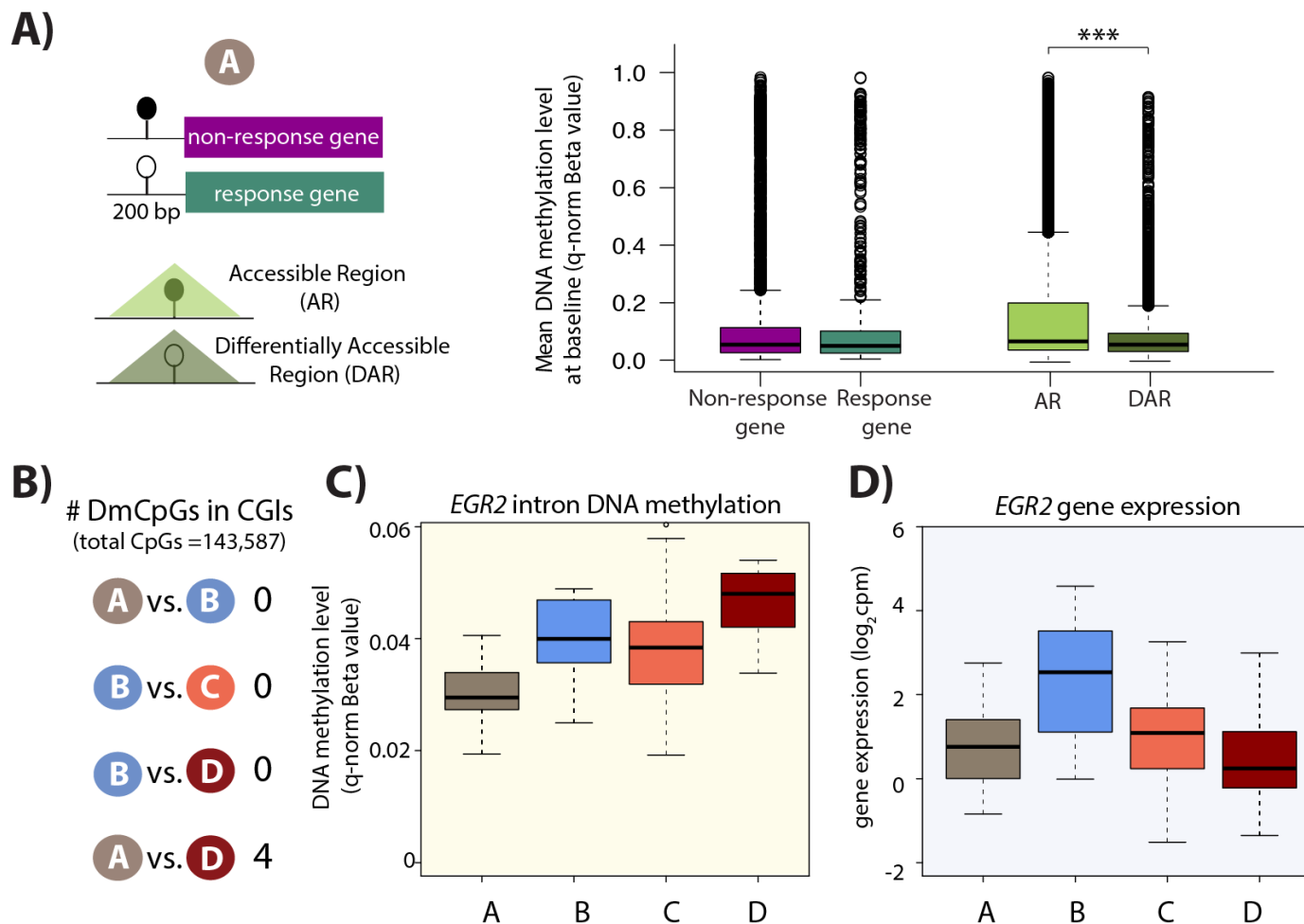


Figure 5: Minimal DNA methylation changes following hypoxia. (A) Mean DNA methylation levels (Beta-values) in the baseline condition (A) at CpGs within 200 bp upstream of the TSS of response genes and non-response genes, and within differentially accessible regions (DARs) and all accessible regions (ARs) (* $p < 0.05$, *** $p < 0.0005$). **(B)** Numbers of differentially methylated CpGs within CpG islands across pairs of conditions. **(C)** DNA methylation levels at a differentially methylated CpG within an intron of the *EGR2* response gene. **(D)** *EGR2* expression during the course of the experiment.

accessible regions ($P < 2.2 \times 10^{-16}$). These data suggest that the underlying DNA sequence features of these regions can affect their epigenetic profile, and that responsive chromatin regions may have specific epigenetic profiles which poise them for rapid response to stress.

DNA methylation levels are largely stable following hypoxia and re-oxygenation

Given that DNA methylation levels can associate with gene expression levels, we asked whether any CpGs are differentially methylated during the course of the oxygen perturbation experiment, which induces thousands of gene expression changes. When considering all 766,658 CpGs we did not find any differentially methylated CpGs (DMCpGs) across any pair of conditions (10% FDR; FigS19A), and all p-values are estimated to be true null p-values ($\pi_0=1$ when estimated by q-value across all pairs of conditions; FigS19B). We found this to be the case when considering two estimates of DNA methylation levels as input: Beta-values or M-values (\log_2 ratio of intensities of methylated versus unmethylated alleles), or when only considering the 32,794 CpGs that are located 200 bp upstream of the TSS (data not shown). Because of the CpG island enrichment in our response gene promoters, we then selected only those 143,587 CpGs present within CpG islands. Although we found no difference in CpG methylation between baseline and hypoxia, and the short-term re-oxygenation conditions, we identified four DMCpGs between the baseline and long-term re-oxygenation conditions. This set includes a CpG in the intron of the *EGR2* response gene, which shows increased DNA methylation levels over time (Fig4C-D). Methylation at CpG islands within the intron of *EGR2* has been shown to confer enhancer activity in cancer cells (Unoki and Nakamura, 2003). If we only select CpGs located within the promoters of the 2,113 response genes, we find one DMCpG within the promoter of the *FTSJ2* gene, a rRNA methyltransferase, that is differentially methylated between hypoxia and long-term re-oxygenation. Selecting CpGs located only within the 831 DARs reveals two DMCpGs between baseline and hypoxia, and one DMCpG between baseline and long-term re-oxygenation. Changes in DNA methylation are therefore not likely to be a major mechanism behind gene expression or chromatin accessibility changes following six hours of hypoxia.

Dynamic eQTLs associate with traits and disease

Finally, we wanted to determine whether any of the dynamic eQTL SNPs or genes that we identified are also associated with complex traits or disease. We first searched within a catalog of genetic variants associated with traits assayed in GWAS for overlap with our dynamic eQTL SNPs

(Buniello et al., 2019) and found an induced dynamic eQTL SNPs that is also associated with a measured phenotype - varicose veins (Table1A; Fig6A).

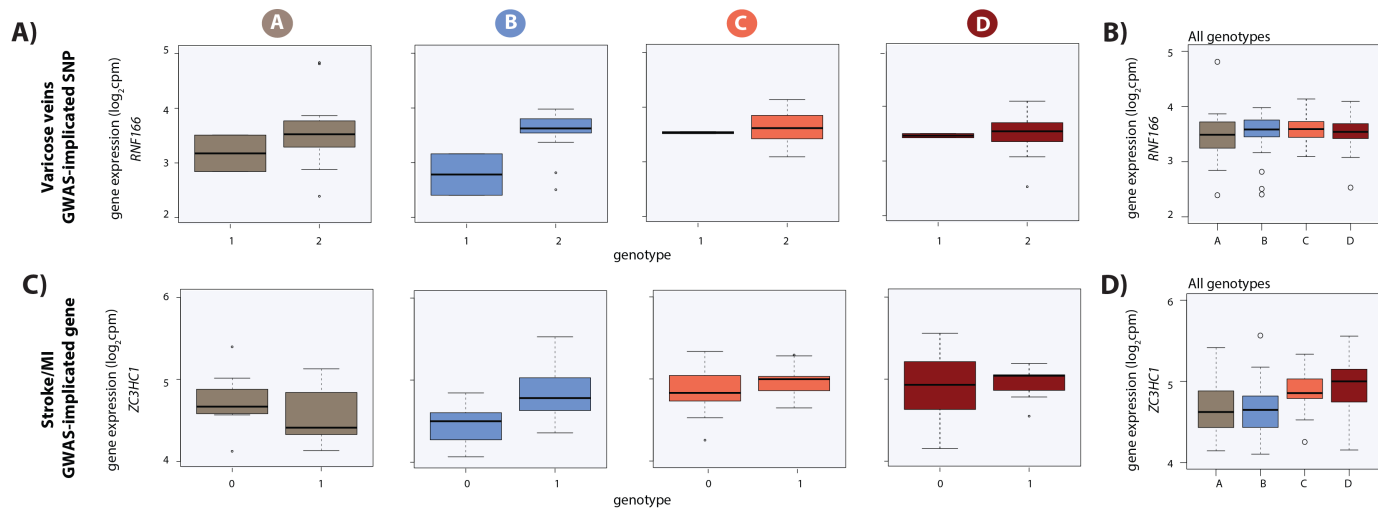


Figure 6: Dynamic eQTLs associate with SNPs and genes implicated in complex disease. (A) Example of a GWAS-implicated SNP that is also a dynamic eQTL SNP. *RNF166* is a dynamic eGene and the associated SNP is implicated in the presence of varicose veins. *RNF166* expression levels are stratified by genotype in each condition. (B) Expression levels of *RNF166* during the course of the experiment following aggregation of all individuals. (C) Example of a stroke and myocardial infarction (MI) GWAS-implicated gene that is also a dynamic eGene (*ZC3HC1*). (D) Expression levels of *ZC3HC1* during the course of the experiment following aggregation of all individuals.

We next took an orthogonal approach, using the same GWAS catalog, to specifically investigate three phenotypes that are associated with cardiovascular function or response to oxygen deprivation: MI, heart failure, and stroke (See methods). Six of our dynamic eGenes are also implicated in these disease states by GWAS (Table1B). This list includes the DNA damage and apoptosis factor *ZC3HC1*, which is implicated in MI and stroke (Fig6B). Importantly, *ZC3HC1* is not an eGene in LV or AA, but the SNP-gene pair is an eQTL in other tissues. These results suggest that perturbation studies in relevant cell types can give insight into the molecular basis for the genetic association with complex traits and disease, which might not be gleaned from the study of post-mortem tissues.

Table 1: Dynamic eQTL SNPs and genes associated with GWAS traits.

(A) Dynamic eQTL (deQTL) SNPs associated with GWAS traits. The eQTL SNP, associated gene, type of eQTL based on whether the eQTL is induced or suppressed in response to hypoxia, tested GWAS trait, whether the eQTL SNP is an eSNP in GTEx left ventricle or atrial appendage, and whether it is an eGene in GTEx ventricle or atrial appendage tissue. **(B)** Dynamic eGenes implicated in three relevant GWAS traits - heart failure, myocardial infarction (MI), and stroke. N.T: not tested.

A)	deQTL SNP	ENSGID	deQTL gene	type	trait	eSNP	eGene
	rs8053350	ENSG00000158717	<i>RNF166</i>	induced	Varicose veins	Yes	Yes

B)	deQTL SNP	ENSGID	deQTL gene	type	trait	eSNP	eGene
	rs60026782	ENSG00000134317	<i>GRHL1</i>	suppressed	Heart failure	No	Yes
	rs187463073	ENSG00000136111	<i>TBC1D4</i>	induced	Heart failure	No	Yes
	rs201280042	ENSG00000186197	<i>EDARADD</i>	induced	Heart failure	N.T	Yes
	rs10242432	ENSG00000091732	<i>ZC3HC1</i>	induced	MI/Stroke	No	No
	rs75401776	ENSG00000096401	<i>CDC5L</i>	induced	Stroke	No	Yes
	rs11612275	ENSG00000171840	<i>NINJ2</i>	induced	Stroke	No	Yes

Discussion

Studying gene expression across individuals in response to stress can reveal latent effects of genetic variation, which may contribute to higher-order phenotypes and disease. In order to understand the effects of genetic variation in a disease-relevant cell type and a disease-relevant process, we differentiated cardiomyocytes from a panel of genotyped individuals, and subjected them to hypoxia and re-oxygenation. We found hundreds of eQTLs that are revealed or suppressed following hypoxic stress (dynamic eQTLs), several of which have been associated with phenotypes measured in GWAS.

Steady-state and dynamic eQTLs may help understand CVD

Attempts have been made to identify genetic variants that associate with gene expression levels and CVD phenotypes in easily accessible biological samples such as blood. However, less than half of CVD/MI GWAS loci are associated with an eQTL in whole blood when thousands of individuals are tested (Joehanes et al., 2017). To determine the effects of genetic variation on gene expression specifically in the heart, more targeted studies have taken advantage of left ventricle tissue (Consortium et al., 2017; Koopmann et al., 2014), left atrium tissue (Lin et al., 2014; Sigurdsson et al., 2017), and right atrial appendage tissue (Consortium et al., 2017) obtained during cardiac surgeries or post-mortem. Using fewer than a hundred individuals, a handful of identified eQTL SNPs correspond to SNPs associated with cardiac traits, thus linking specific

genes to organismal-level phenotypes. A compelling example is the association between *MYOZ1* expression and atrial fibrillation (Lin et al., 2014; Sigurdsson et al., 2017). Across tissues, the GTEx consortium reported that ~50% of eQTLs are also associated with variation in other measured complex traits (Consortium et al., 2017), and Heinig *et al.* have shown that 20% of left ventricle eQTLs relate to heart-associated loci (Heinig et al., 2017). However, these variants, identified in healthy individuals, are unlikely to represent all genetic variants that have consequences on disease. Indeed, Heinig *et al.* identified 100 dilated cardiomyopathy-specific eQTLs (not seen in healthy individuals) in a case-control study of 97 individuals with dilated cardiomyopathy and 108 healthy donors (Heinig et al., 2017). Similarly, by collecting samples pre- and post-surgically-induced ischemia, Stone *et al.* identified genetic associations that are only detected under stress (Stone et al., 2019). While these studies provide a set of gene targets for further investigation, there are many loci that remain unexplained.

The heart is a complex tissue consisting of multiple cell types. The effects of some genetic variants in specific cell types might well be masked when considering heterogeneous tissue samples. As we are now able to direct iPSCs towards a cardiac fate, we can test for genetic effects on specific cell types such as cardiomyocytes (Panopoulos et al., 2017). As one would expect, iPSC-CMs are better suited to study cardiovascular traits than the immortalized B-cells or iPSCs from which they are derived (Banovich et al., 2018). However, given the high degree of eQTL sharing across diverse tissues (GTEx), identifying eQTLs in the disease-relevant terminal cell type at steady state may not give substantial insight into disease biology. A significant advantage of using iPSC-CMs is that these cells provide a system to interrogate gene expression dynamics.

Cellular stressors that perturb gene expression levels and the cell state can unmask additional layers of regulatory variation (Alasoo et al., 2018; Barreiro et al., 2012; Caliskan et al., 2015; Kim-Hellmuth et al., 2017; Knowles et al., 2018; Manry et al., 2017; Nedelec et al., 2016). Intermediate developmental cell states can similarly provide insight into GWAS loci where eQTL analysis in terminal cell types cannot (Strober et al., 2019). Further evidence for the notion that steady state eQTLs may have limited applicability to disease states comes from our previous work, where we used a comparative evolutionary approach to investigate the response to stress. We showed that genes that have a conserved response to oxygen stress in iPSC-CMs from both humans and chimpanzees, and are therefore likely relevant for disease, are depleted for eQTLs identified in heart tissue (Ward and Gilad, 2019).

In the current study, by subjecting iPSC-CMs from a panel of individuals to perturbation (oxygen deprivation), we were able to identify a dynamic eQTL SNP (rs8053350) that is associated with varicose veins, and the level of *RNF166* expression (Fukaya et al., 2018). This SNP falls within an intron of the *PIEZO1* gene. Varicose veins are associated with a risk for developing deep vein thrombosis and other vascular diseases (Chang et al., 2018). When we performed an analogous analysis focused on genes previously associated with three relevant traits – MI, heart failure and stroke, we identified a novel heart eGene, *ZC3HC1*, encoding the NIPA protein, which is implicated in MI, coronary artery disease and ischemic stroke (Consortium, 2011; Nikpay et al., 2015; Schunkert et al., 2011). This dynamic eQTL SNP is also associated with bronchodilator responsiveness in chronic obstructive pulmonary disease (Hardin et al., 2016).

Mechanisms behind response genes and dynamic eQTLs

Changes in gene expression can associate with other molecular-level phenotypes. The response to hypoxia is mediated by the HIF1 α transcription factor (Samanta and Semenza, 2017), but given that there are hundreds of HIF1 α binding locations and thousands of differentially expressed genes, regulation by this factor alone cannot directly explain all the transcriptional changes. We explored two additional molecular phenotypes in the context of oxygen deprivation - the locations and level of accessibility of open chromatin regions, and DNA methylation levels. We did not find either to contribute substantially to the gene expression response we observed. There are minimal changes in accessibility following hypoxia, which is in contrast to observations of studies that considered stimulation of immune cell types (Alasoo et al., 2018; Calderon et al., 2019; Pacis et al., 2015). This could reflect cell type specificity in response to stress, or the specificity of the cellular response to different stressors. Indeed, despite large gene expression changes in response to various stimulants in endothelial cells, there are a relatively small number of differentially accessible regions (Findley et al., 2019). We speculate that the transcriptional response to oxygen stress could result in the induction of transcription factors, which bind already accessible regions of open chromatin, and that cells are primed for a quick response to this universal cellular stress. Indeed, it has been shown that chromatin contacts exist between HIF1 α binding sites and hypoxia-inducible genes in the normoxic state (Platt et al., 2016). Conversely, it has been suggested that hypoxia results in the induction of HIF1 α , and significant changes in histone methylation (Batie et al., 2019). As we did not measure histone marks in our system, these changes may occur in the absence of chromatin accessibility changes, but we also cannot rule out the possibility that the

choice of a single timepoint following six hours of hypoxia, or insufficient statistical power in our sample size, contributed to the minimal differences in accessibility that we observed.

Using an approach designed to measure small effect sizes between conditions, we did identify a set of 831 DARs between hypoxia and short-term re-oxygenation that are enriched for marks of active chromatin, CpG islands, and TEs. These regions do not appear to explain many of the gene expression differences we observed. Hypoxia and oxidative damage are likely to also affect the genome in ways that do not directly impact gene expression. Indeed, the distribution of oxidative DNA damage sites varies across the genome following stress such that TEs and active chromatin regions are enriched for DNA damage, while promoters are depleted (Poetsch et al., 2018). We found enrichment for TEs, specifically Alu SINE elements, in DARs. Interestingly, TEs, and DNA transposons in particular, are also enriched in regions that become accessible in macrophages in response to bacterial infection; suggesting sequence-specific effects of TEs in response to different cellular stressors (Bogdan, 2019). Alu elements have previously been found to associate with the response to stress in other contexts. Serum starvation induces binding of TFIIIC, which recruits RNA polymerase III, to Alu elements (Ferrari et al., 2019), and heat shock increases chromatin accessibility around Alu elements (Kim et al., 2001).

There are several studies, which suggest that DNA methylation levels are dynamic and change in response to stressors such as hypoxia. We did not find any notable differences in DNA methylation levels pre- and post-hypoxia and re-oxygenation, which suggests that like chromatin accessibility, DNA methylation levels do not make large contributions to changes in gene expression levels or the appearance of dynamic eQTLs in our system. Many of the DNA methylation changes that have been described in response to hypoxia occur in chronic and intermittent hypoxia, and not acute hypoxia as investigated in our study (Hartley et al., 2013; Robinson et al., 2012; Watson et al., 2014). DNA methylation levels are also altered in response to other stressors such as bacterial infection (Pacis et al., 2015); however, the importance of timing is highlighted by the fact that, in this system, gene expression responses precede DNA methylation changes (Pacis et al., 2019). It is also important to note that our study considers baseline oxygen levels to be 10% oxygen, which is closer to physiological oxygen levels (5-13%) than atmospheric oxygen levels (21%; (Brahimi-Horn and Pouyssegur, 2007; Carreau et al., 2011; Jagannathan et al., 2016). Most studies define normoxia as 21% oxygen saturation, and while this likely leads to larger effect size differences in known hypoxia response genes following hypoxia, these comparisons may not give meaningful insight into the *in vivo* state.

One can speculate about different mechanisms that might lead to the appearance or disappearance of dynamic eQTLs. In the context of the immune response, it has been shown that the same response variants affect both gene expression and chromatin accessibility (Alasoo et al., 2018). This is in line with the general notion that changing cellular environments results in differences in chromatin accessibility at transcription factor binding sites, which leads to gene expression changes. We found that this does not appear to be a major mechanism in our system as there are minimal changes in accessibility following hypoxia. We observed that there is an enrichment of response genes amongst dynamic eQTLs suggesting that the change in environment results in a change in expression levels that is dependent on the associated genotype. We also find enrichment for TFs amongst response genes and dynamic eQTLs, suggesting that dynamic eQTLs can appear through secondary *trans* effects.

Potential limitations of our model

To understand the effects of genetic variation on human heart tissue, and how this variation might contribute to the MI and I/R injury etiologies of CVD, we carefully perturbed oxygen levels that cardiomyocytes in culture are exposed to. This *in vitro* approach is by design a model system, and therefore will likely not fully recapitulate the *in vivo* state. However, we previously found that out of 2,549 genes that respond to hypoxia in iPSC-CMs from humans and chimpanzees, only 16% are differentially expressed between iPSC-CMs and heart tissue (Pavlovic et al., 2018; Ward and Gilad, 2019). This suggests that our *in vitro* system is applicable to heart tissue. There is still a possibility that the dynamic eQTLs that we identify in our *in vitro* system are not physiologically relevant.

Our study comprised a small number of individuals (15), far fewer than what is typical for identifying eQTLs. Our work is therefore a first step towards understanding the effects of genetic variation on gene expression in response to stress. Nevertheless, with a small number of individuals we were able to identify a couple of hundred of dynamic eQTLs that are revealed or suppressed under stress, suggesting that this paradigm is worth exploring further in larger cohorts. Under the simplifying assumption of a single causal variant, we determined that we have ~6% power to detect an effect which explains ~38% of the heritability, and an equal false positive rate to call it a dynamic eQTL (See methods; FigS20). This suggests that the impact of stress on genotype-dependent effects on gene expression will be even greater in studies which have higher power to detect smaller effects of genotype. For perspective, early eQTL studies were similarly

powered to our study, using 70 individuals; yet these studies still led to important insights opening an avenue of research focused on assaying the consequences of genetic variation by RNA-seq (Pickrell et al., 2010).

In summary, there have been few studies assessing the effects of genetic variation in response to CVD-relevant perturbations in cardiomyocytes. Here we profiled the response to oxygen deprivation in cardiomyocytes from a panel of genotyped individuals. We find that eQTLs can appear and disappear in response to oxygen deprivation, and that some of these eQTLs have effects on relevant complex traits and disease.

Materials and Methods

Cardiomyocyte differentiation from iPSCs

We used fifteen individuals from the Yoruba YRI HapMap population. iPSCs were reprogrammed from lymphoblastoid cell lines (Banovich et al., 2018). iPSCs were maintained in a feeder-independent state in Essential 8 Medium (A1517001, ThermoFisher Scientific, Waltham, MA, USA) with Penicillin/Streptomycin (30002, Corning, NY, USA) on Matrigel hESC-qualified Matrix (354277, Corning, Bedford, MA, USA) at a 1:100 dilution. Cells were passaged at ~70% confluence every 3-4 days with dissociation reagent (0.5 mM EDTA, 300 mM NaCl in PBS), and seeded with ROCK inhibitor Y-27632 (ab12019, Abcam, Cambridge, MA, USA).

Cardiomyocyte differentiations were performed largely as previously described (Ward and Gilad, 2019), except the duration and concentration of the Wnt agonist and antagonist differed for this panel of individuals, which included only human samples. Briefly, on Day 0, iPSC lines at 70-100% confluence in 100 mm plates were treated with 12 μ M GSK3 inhibitor CHIR99021 trihydrochloride (4953, Tocris Bioscience, Bristol, UK) in 12 ml Cardiomyocyte Differentiation Media [500 mL RPMI1640 (15-040-CM ThermoFisher Scientific), 10 mL B-27 Minus Insulin (A1895601, ThermoFisher Scientific), 5 mL Glutamax (35050-061, ThermoFisher Scientific), and 5 mL Penicillin/Streptomycin], and a 1:100 dilution of Matrigel. 24 hours later, on Day 1, the media was replaced with Cardiomyocyte Differentiation Media. 48 hours later, on Day 3, 2 μ M of the Wnt inhibitor Wnt-C59 (5148, Tocris Bioscience), diluted in Cardiomyocyte Differentiation Media, was added to the cultures. Cardiomyocyte Differentiation Media was replaced on Days 5,7,10, and 12. Cardiomyocytes were purified by metabolic purification by the addition of glucose-free, lactate-

containing media (Purification Media) [500 mL RPMI without glucose (11879, ThermoFisher Scientific), 106.5 mg L-Ascorbic acid 2-phosphate sesquimagnesium salt (sc228390, Santa Cruz Biotechnology, Santa Cruz, CA, USA), 3.33 mL 75 mg/mL Human Recombinant Albumin (A0237, Sigma-Aldrich, St Louis, MO, USA), 2.5 mL 1 M lactate in 1 M HEPES (L(+))Lactic acid sodium (L7022, Sigma-Aldrich)), and 5 mL Penicillin/Streptomycin] on Days 14, 16 and 18. 1.5 million cardiomyocytes were re-plated per well of a 6-well plate on Day 20 in Cardiomyocyte Maintenance Media [500 mL DMEM without glucose (A14430-01, ThermoFisher Scientific), 50 mL FBS (S1200-500, Genemate), 990 mg Galactose (G5388, Sigma-Aldrich), 5 mL 100 mM sodium pyruvate (11360-070, ThermoFisher Scientific), 2.5 mL 1 M HEPES (SH3023701, ThermoFisher Scientific), 5 mL Glutamax (35050-061, ThermoFisher Scientific), 5 mL Penicillin/Streptomycin]. iPSC-CMs were matured in culture for a further 10 days with Cardiomyocyte Maintenance Media replaced on Days 23, 25, 27, 28 and 30.

On Day 25, iPSC-CMs were transferred to a 10% oxygen environment (representative of *in vivo* levels) in an oxygen-controlled incubator (HERAcell 150i CO₂ incubator, ThermoFisher Scientific). From Day 27 onwards, iPSC-CMs were pulsed at a voltage of 6.6 V/cm, frequency of 1 Hz, and pulse frequency of 2 ms using an IonOptix C-Dish & C-Pace EP Culture Pacer to further mature the cells and synchronize beating.

Flow cytometry

Purity of the cardiomyocyte cultures was assessed ~Day 30 as previously described (Ward and Gilad, 2019). Briefly, cells were stained with Zombie Violet Fixable Viability Kit (423113, BioLegend), and PE Mouse Anti-Cardiac Troponin T antibody (564767, clone 13-11, BD Biosciences, San Jose, CA, USA), and analyzed on a BD LSRFortessa Cell Analyzer together with negative control samples of iPSCs, and iPSC-CMs that are incubated without the troponin antibody, or without either the troponin antibody or viability stain.

Hypoxia experiment

On Day 31/32 iPSC-CMs were subjected to the hypoxia experiment. At time = 0, condition A samples remained at 10% O₂ (normoxia), while samples for conditions B, C and D were transferred to an incubator set at 1% O₂ (hypoxia). After 6 hours, conditions A and B were harvested while plates C and D were returned to normoxic oxygen conditions. Plate C was harvested 6 hours

following the hypoxic treatment, and Plate D was harvested 24 hours following the hypoxic treatment. Oxygen levels, experienced by the cells in culture, were measured in cultures from each experimental batch using an oxygen-sensitive sensor (SP-PSt3-NAU-D5-YOP, PreSens Precision Sensing GmbH, Regensburg, Germany), optical fiber (NWDV29, Coy, Grass Lake, MI, USA), and oxygen meter (Fibox 3 Transmitter NWDV16, Coy).

Material collection

Cell culture media for ELISA and cytotoxicity assays

Aliquots of cell culture media from each experiment were centrifuged at 10 000 rpm for 10 min at 4°C to remove cellular debris. The supernatant was stored at -80°C until further use.

Nuclei for ATAC-seq

Cardiomyocytes from each well of a 6-well plate were washed twice with cold PBS on ice before collection by manual scraping in 1.5 ml PBS. 200 µl of cells were pelleted by centrifugation at 500 g for 5 min. Cell pellets were re-suspended in 50 µl cold ATAC-seq lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% Igepal CA630, dH₂O). Nuclei were pelleted by centrifugation at 500 g for 5 min at 4°C. Nuclei were re-suspended in 50 µl transposition mix (25 µl 2xTD buffer, 2.5 µl Tn5 transposase, 22.5 µl nuclease-free dH₂O) from the Nextera DNA sample kit (FC-121-1031, Illumina). The transposition reaction was performed at 37°C for 30 min. Transposed DNA was purified with Qiagen MinElute Kit (28004, Qiagen, MD, USA), re-suspended in 12 µl elution buffer, and stored at -20°C.

Cell pellets for RNA-seq & DNA methylation arrays

Cells from each well of a 6-well plate were washed twice with cold PBS on ice before collection by manual scraping in 1.5 ml PBS. Cells were pelleted by centrifugation at 7 000 rpm for 8 min at 4°C, flash-frozen and stored at -80°C.

RNA/DNA extraction

RNA and DNA were extracted from the same frozen cell pellets using the ZR-Duet DNA/RNA MiniPrep kit (D7001, Zymo, CA, USA) according to the manufacturer's instructions. All four conditions from three or four individuals were extracted in the same batch. RNA samples had a median RIN score of 8.5 (STable).

RNA-seq library prep

500 ng of RNA were used to prepare sequencing libraries using the Illumina TruSeq RNA Sample Preparation Kit v2 (RS-122-2001 & -2002, Illumina). Libraries were pooled into five master mixes containing 12 or 16 samples. Each pool was sequenced 50 bp, single-end on the HiSeq2500 or HiSeq4000 according to the manufacturer's instructions.

DNA methylation array

9 chips (8 samples per chip) with 60-1000 ng DNA were bisulfite-converted and processed on an Illumina Infinium MethylationEPIC array at the University of Chicago Functional Genomics facility.

ATAC-seq

We performed ATAC-seq in 14 of the 15 individuals we had gene expression data for (STable for details). ATAC-seq libraries were prepared using the Illumina Nextera DNA sample kit. Libraries were amplified for 10-16 cycles depending on the amplification rate of each library. Each library was amplified in a PCR reaction containing 10 µl DNA, 10 µl dH₂O, 15 µl NMP (PCR master mix), 5 µl PPC (PCR primer cocktail), and 5 µl index N5, 5 µl index N7). PCR conditions were set at 72°C for 5 min, 98°C for 30 sec, 98°C for 10 sec, 63°C for 30 sec, 72°C for 1 min, repeat steps 3-5 4x and hold at 4°C. The number of cycles per library was determined using a qPCR side reaction as described in Buenrostro *et al.* (Buenrostro et al., 2013). Libraries were purified using Agencourt AMPure XP beads (A63880, Beckman Coulter, IN, USA), and bioanalyzed to determine library quality. 12 or 16 samples were pooled together to generate four master mixes. Each master mix was sequenced 50 bp paired-end on the HiSeq4000 according to the manufacturer's instructions.

Lactate dehydrogenase activity assay

Lactate dehydrogenase activity (LDH) was measured in 5 μ l cell culture media using the Lactate Dehydrogenase Activity Assay Kit (MAK066, MilliporeSigma, MO, USA) according to the manufacturer's instructions. Each sample was assayed in triplicate. LDH activity was measured as the difference in absorbance prior to the addition of the substrate, and 10 min after the initiation of the enzymatic reaction, calculated relative to a standard curve. Measurements are standardized relative to A, and reported as A (A-A), B (B-A), C (C-B) and D (D-B).

BNP ELISA

125 μ l of cell culture media was assayed to quantify the level of secreted BNP using the Brain Natriuretic Peptide EIA kit (RAB0386, MilliporeSigma). Each sample was assayed in duplicate on two 96-well plates. BNP levels were quantified relative to a standard curve using 4- and 5-parameter logistic models using the R package drc. Measurements are standardized relative to A, and reported as A (A-A), B (B-A), C (C-B) and D (D-B).

RNA-seq analysis

Reads were aligned to hg19 using subread align (Liao et al., 2013). The mapped reads were then reprocessed to reduce reference bias for downstream analyses using the WASP pipeline (van de Geijn et al., 2015). Briefly, reads overlapping polymorphisms segregating in our population were remapped to the genome using the true read, and a version of the read with the alternative allele. Only reads that mapped uniquely to the same locations with both possible alleles were kept. The median number of reads across conditions was similar (A: 34,353,716; B: 33,493,298; C: 33,883,532; D: 38,147,083). The number of filtered reads mapping to genes was quantified using featureCounts within subread (Liao et al., 2014). We obtained measurements for 19,081 genes. Sample 18852A was an outlier when considering read count correlations between pairs of samples, and was therefore removed prior to subsequent analyses.

Differential expression analysis

We selected autosomal genes for downstream analysis (18,226). Log₂-transformed counts per million were calculated (Robinson, 2010), and genes with a mean log₂cpm < 0 were excluded. We

used the fact that we have replicate data from three individuals to remove unwanted variation in our data. We used the RUVs function in the RUVSeq package in R (Risso et al., 2014) to identify such factors. By manual inspection, our data segregated by individual or condition after correction with four factors. For the differential expression analysis, we excluded sample replicate one to avoid the outlier sample and randomly selected replicate two, instead of replicate three, for individuals with replicate samples. We used the RUV factors as covariates in our differential expression analysis using the TMM-voom-limma pipeline (Law et al., 2014; Robinson et al., 2010; Smyth, 2004). We used fixed effects for each condition (A, B, C, D), the RUVs factors as covariates, and a random effect for individual, which was implemented using duplicateCorrelation. Genes with a Benjamini and Hochberg FDR < 0.1 are classified as differentially expressed (Benjamini and Hochberg, 1995).

-Gene expression trajectory analysis

To identify response genes, we used the Cormotif package in R (Wei et al., 2015) to jointly model pairs of tests. We used TMM-normalized log₂cpm values as input and considered the following pairs of tests: A vs B, B vs. C and B vs. D to determine which genes are changing their expression during the course of the experiment. The best fit was determined to correspond to two correlation motifs or clusters using BIC and AIC. We classified genes as response genes if the probability of differential expression between conditions was > 0.5 in all pairs of tests.

eQTL identification

To map eQTLs, we analyzed the same samples considered in the differential expression analysis. Given the sample size in this study, we utilized the combined haplotype test (CHT) to identify eQTLs (van de Geijn et al., 2015). This test models both allelic imbalance and total read depth at a region to identify QTLs. We require 50 total counts and 10 ten allele-specific counts for each gene, and tested variants 25 kb upstream and 25 kb downstream of the TSS, resulting in 1,040,874 shared tests (A: 1,215,476; B: 1,211,099, C: 1,224,612, D: 1,201,078). As previously reported, we found that null p-values for the CHT were not calibrated in our data. To calibrate the p-values, we estimated the null distribution of the CHT by permuting the data 100 times and fitting a Beta distribution to the permuted p-values for each SNP-gene pair (previously proposed by (Delaneau et al., 2017; Ongen et al., 2016)). We then computed an adjusted p-value for each SNP-gene pair by taking the CDF of the fitted Beta distribution, evaluated at the reported CHT p-value.

To call significant eQTLs, we estimated q-values for the set of adjusted p-values for each phenotype, and took tests with $q < 0.1$. The number of eGenes in each condition was determined by taking the most significant SNP-gene association in each condition (i.e. the top SNP). We defined dynamic eQTLs as either: 1) significant only in A ($q < 0.1$ in A and permutation-adjusted $p > 0.15$ in B and C and D; suppressed eQTL); 2) significant in at least one of B, C, or D ($q < 0.1$) and not nominally significant in A (adjusted $p > 0.15$; induced eQTL).

Power analysis

For QTL mapping, we assume a linear model

$$\begin{aligned} y_i &= x_i\beta + \epsilon_i \\ \epsilon_i &\sim \mathcal{N}(0, \sigma^2) \end{aligned}$$

where y_i denotes the phenotype of individual i and x_i denotes the genotype of individual i at a single SNP of interest. We estimate an effect size $\hat{\beta}$

$$\hat{\beta} \sim \mathcal{N}\left(\beta, \frac{\sigma^2}{n}\right)$$

where n is the sample size. Let $\lambda = \beta/\sigma$ be the standardized effect size. Then,

$$\hat{\lambda} \sim \mathcal{N}\left(\lambda, \frac{1}{n}\right)$$

and

$$\text{Power}(\lambda, \alpha, n) = \Phi\left(\Phi^{-1}(\alpha/2) + \lambda\sqrt{n}\right)$$

where α denotes the significance level and Φ denotes the standard Gaussian CDF. To simplify the analysis, we consider $\alpha = 0.05/20000 = 2.5 \times 10^{-6}$ (i.e., Bonferroni correction; this is equivalent to controlling the FDR when all tests are null, and is conservative otherwise). Assume there is a single causal variant. Then, the phenotypic variance explained is:

$$h^2 = \frac{\lambda^2}{\lambda^2 + 1}$$

We defined a dynamic eQTL as either significant only in A, or significant (after Bonferroni correction, in this analysis) in one of B, C, or D and not significant in A. To estimate the false

positive rate of dynamic eQTL calling, we asked what was the probability of a SNP passing this definition, assuming the standardized effect size λ was *identical* in all four conditions. We then computed phenotypic variance explained, power to detect an eQTL, and false positive rate to call a dynamic eQTL for every choice of standardized effect size λ .

Overlapping response genes and eGenes with existing gene sets

-Gene ontology analysis

Gene set enrichment analysis was performed on response genes, and a background set of all expressed genes using the DAVID genomic annotation tool (Huang et al., 2009a, b). GO Terms related to Biological Processes were selected, and those with a Benjamini-Hochberg controlled FDR < 0.05 were designated as significantly enriched. Each of the five significantly enriched processes relates to transcription (“DNA-templated transcription”, “DNA-templated regulation of transcription”, “DNA-templated negative regulation of transcription”, “negative regulation of transcription from RNA polymerase II promoter”, “positive regulation of transcription from RNA polymerase II promoter”). The most significantly enriched GO terms related to Molecular Functions include “transcription factor activity, sequence-specific DNA binding”, “nucleic acid binding” and “DNA binding”.

-Transcription factors

A list of 1,637 annotated human TFs was obtained from (Lambert et al., 2018), and intersected with our gene sets.

-GTEx eQTLs

eQTLs in LV and AA, and twelve other randomly selected tissues (adipose, brain cortex, colon, lung, liver, muscle, pancreas, pituitary, skin, spleen, thyroid, whole blood) were downloaded from v7 in the GTEx portal (www.gtexportal.org). eGenes were selected at 5% FDR in each tissue, and intersected with our gene categories.

ATAC-seq analysis

Paired-end sequencing reads were aligned to hg19 using bowtie2 with default settings (Langmead and Salzberg, 2012). Reads were filtered using Picard Tools (<https://broadinstitute.github.io/picard/>) to remove duplicate reads, and reads mapping to the

mitochondrial genome. Reads were then remapped using the WASP pipeline as described above. We retained a similar median number of reads across conditions (A: 28,998,060; B: 33,662,261; C: 30,161,640; D: 34,534,416). Across conditions, there is no significant difference in the number of mapped reads, number of regions identified, or fraction of reads mapped to open chromatin regions (FigS13A-C). All libraries, across conditions, show the expected fragment size distribution, enrichment of reads at transcription start sites (TSS), and footprints at well-defined CTCF motifs (FigS13D-F). Correlation analysis of read counts between pairs of samples revealed clustering by individual and condition (FigS14). As expected, the correlation of read counts between samples at the 10,633 regions overlapping the TSS is higher than the correlation across all regions (median $\rho = 0.83$ vs. 0.56). Pairs of samples from the same condition are marginally more correlated in their accessibility profiles than pairs of samples across all conditions (median $\rho = 0.84$ vs. 0.83 at the TSS).

Identification of accessible chromatin regions

To generate a unified list of regions with accessible chromatin across conditions and samples, we first used MACS2 (Zhang et al., 2008) to identify peaks within each sample independently. Next, we used BEDtools (Quinlan and Hall, 2010) with the `multiIntersectBed` function to identify overlapping peaks within each condition separately. Within each condition, we retained peaks with support from more than three individuals and used the `mergeBed` function to create a condition-specific consensus. We then combined and merged the bed files across the four conditions to make a final consensus file containing all the filtered accessible regions. The number of reads mapping to accessible chromatin regions was quantified using `featureCounts` within subread (Liao et al., 2014).

Identification of differentially accessible regions (DARs)

The 128,673 open chromatin regions associated with count data were filtered to include only those regions on the autosomes, and those which had mean $\log_2\text{cpm}$ values > 0 for each region. First, to identify differentially accessible regions we used the same limma framework described above for the RNA-seq data. To test for differences between conditions, a linear model with a fixed effect for condition was used together with a random effect for individual. We did not identify any significantly differentially accessible regions with a Benjamini and Hochberg FDR < 0.1 . To identify regions with small effect size differences between conditions we used an adaptive shrinkage method

implemented in the *ashr* package in R (Stephens, 2017). We used the regression estimates (regression coefficients, posterior standard errors, and posterior degrees of freedom) generated by *limma* to calculate a posterior mean (shrunken regression coefficients), FDR, and False Sign Rate (FSR, probability that the sign of the effect size is wrong). We considered regions to be differentially accessible at $\text{FSR} < 0.1$. We denote regions that are not differentially accessible as constitutively accessible regions.

Overlap of DARs with genomic features

-TSS

Transcription start sites were obtained from the UCSC Table Browser (<http://genome.ucsc.edu/cgi-bin/hgTables>) using 'txStart' from Ensembl genes (Karolchik et al., 2004). TSS were defined based on the TSS of the 5' most transcript on the sense strand and 3' most transcript on the anti-sense strand. TSS regions, and subsequent genomic features, were intersected with DARs and constitutively accessible regions requiring a 1 bp overlap using *bedtools intersect* (Quinlan and Hall, 2010).

-Histone marks

We obtained histone mark data (.bed files) for human heart tissue from the ENCODE consortium (Consortium, 2012; Davis et al., 2018) ENCODE portal, (<https://www.encodeproject.org>). We selected H3K4me3 (Experiment ENCSR181ATL), H3K4me1 (Experiment ENCSR449FRQ), H3K36me3 (Experiment ENCSR799KLF), H3K27me3 (Experiment: ENCSR613PPL), and H3K9me3 (Experiment ENCSR803MVC) ChIP-seq data from heart left ventricle tissue from a 51-year-old female individual (Biosample ENCBS684IAD).

-Transcription factor binding locations

We obtained ChIP-seq data for the hypoxia-responsive factors HIF1 α and HIF2 α assayed in the MCF-7 breast cancer cell line (Schodel et al., 2011), and E2F4 in the GM06990 lymphoblastoid cell line (Lee et al., 2011). Co-ordinates of the 356 HIF1 α , 301 HIF2 α and 16,245 E2F4 bound regions were converted from hg18 to hg19 using the *liftOver* tool in the Galaxy platform (<http://galaxyproject.org/>; (Afgan et al., 2018)).

-Motif enrichment analysis in DARs

We obtained sequences for all accessible regions and differentially accessible regions using the Galaxy platform (Afgan et al., 2018). We used the MEME-ChIP tool within The MEME Suite (Bailey et al., 2009; Machanick and Bailey, 2011) in Differential Enrichment mode to identify motifs differentially enriched in DARs compared to all accessible regions.

-TEs

We obtained repeat annotations from the RepeatMasker track (Jurka, 2000; Smith, 2010) from the UCSC Table browser (Karolchik et al., 2004). We intersected the Repeatmasker track with our accessible regions and reported those elements where 50% of their length overlaps a DAR or constitutively accessible region. We stratified TEs by TE class: LINE, SINE, DNA and LTR, and then by TE family and type within the SINE class.

-CpG islands

We obtained CpG island annotations from the UCSC Table Browser, and overlapped these regions with DARs and constitutively accessible regions.

caQTL identification

The caQTLs were identified in the same manner as described for the eQTLs. However, in the caQTL analyses, we limited tested SNPs to those falling within the peak regions from our consensus file, as opposed to testing variants within 25 kb of the region. Dynamic caQTLs were identified as for dynamic eQTLs.

DNA methylation analysis

To allow for accurate quantification of DNA methylation levels we removed probes overlapping SNPs with a minor allele frequency of > 0.1 , and only retained probes with a detection p-value of > 0.75 across samples. Beta-values (ratio of methylated probe intensity and overall probe intensity, and bounded between 0-1) were quantile normalized using lumiN, and, when appropriate, converted to M-values (\log_2 ratio of intensities of methylated probe versus unmethylated probe) using lumi (Du et al., 2008).

The methylation level of CpGs coincides with the expected distribution based on their annotated genomic location i.e. low levels of DNA methylation in CpG islands, and higher levels in CpG island shores, and CpG island shelves respectively (FigS17B). Correlation analysis across all pairs of samples, including replicate samples, reveals clustering primarily by individual rather than condition (FigS18).

To measure the DNA methylation level at gene set promoters, we selected CpGs 200 bp upstream of the TSS (TSS200 defined on the array). We considered all CpGs when overlapping with DARs.

Identification of differentially methylated CpGs (DMCpGs)

Differentially methylated CpGs were identified using the same limma framework as described for the RNA-seq data. Analysis was run using both Beta-values and M-values.

Integration with GWAS-implicated genes

We intersected the Reference SNP cluster ID of our dynamic QTLs with the 158,654 SNPs in the NHGRI-EBI GWAS Catalog available from the UCSC Table Browser (Buniello et al., 2019) in August 2019.

We also considered the ‘mapped genes’ results from GWAS from three relevant traits: myocardial infarction (EFO_0000612, 89 genes), heart failure (EFO_0003144, 164 genes) and stroke (EFO_0000712, 255 genes), downloaded from the NHGRI-EBI GWAS Catalog in August 2019. Gene lists were intersected with our response eGenes.

Supplemental material

Supplemental material: Document containing Supplemental Figures 1-20, and Table S1.

Supplemental tables: Document containing Supplemental Tables listing experimental batches, oxygen levels, RIN scores, response genes, eGenes in A, B, C, D, dynamic eGenes, DARs, caQTLs in A, B, C, D, dynamic caQTLs and DMCpGs.

Data Access

All RNA-seq, ATAC-seq and DNA methylation data have been deposited in the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo/) under accession number GSE144426.

Acknowledgements

We thank Kristen Patterson and Amy Mitrano for experimental assistance, David Knowles and Benjamin Strober for preliminary data exploration, all members of the Gilad lab and Luis Barreiro for helpful discussions, and Natalia Gonzales for comments on the manuscript. We thank the Genomics Core Facility at the University of Chicago for sequencing the libraries and processing the DNA methylation arrays. We thank The Genotype-Tissue Expression (GTEx) Project, supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH and NINDS, for providing data. The data used for the analyses described in this manuscript were obtained from the GTEx portal v7 on May 24th 2018. We thank the ENCODE Consortium and the Bernstein Lab at Broad for generating and making the histone mark ChIP-seq data available. M.C.W was supported by an EMBO Long-Term Fellowship (ALTF 751-2014), and the European Commission Marie Curie Actions. N.E.B was supported by NIH grant AG044948. This work was funded by an NIH grant from NHLBI (HL092206) to Y.G.

Contributions

M.C.W, N.E.B and Y.G conceived and designed the study. M.C.W and N.E.B performed experiments. M.C.W, N.E.B, and A.S analyzed the data. M.C.W and Y.G wrote the paper. M.S and Y.G supervised the work.

Disclosure Declaration

The authors declare no competing interests.

References

- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., Chilton, J., Clements, D., Coraor, N., Gruning, B.A., *et al.* (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* 46, W537-W544.
- Alasoo, K., Rodrigues, J., Mukhopadhyay, S., Knights, A.J., Mann, A.L., Kundu, K., Consortium, H., Hale, C., Dougan, G., and Gaffney, D.J. (2018). Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat Genet* 50, 424-431.
- Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., and Noble, W.S. (2009). MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37, W202-208.
- Banovich, N.E., Li, Y.I., Raj, A., Ward, M.C., Greenside, P., Calderon, D., Tung, P.Y., Burnett, J.E., Myrthil, M., Thomas, S.M., *et al.* (2018). Impact of regulatory variation across human iPSCs and differentiated cells. *Genome Res* 28, 122-131.
- Barreiro, L.B., Talleux, L., Pai, A.A., Gicquel, B., Marioni, J.C., and Gilad, Y. (2012). Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. *Proc Natl Acad Sci U S A* 109, 1204-1209.
- Batie, M., Frost, J., Frost, M., Wilson, J.W., Schofield, P., and Rocha, S. (2019). Hypoxia induces rapid changes to histone methylation and reprograms chromatin. *Science* 363, 1222-1226.
- Benaglio, P., D'Antonio-Chronowska, A., Ma, W., Yang, F., Young Greenwald, W.W., Donovan, M.K.R., DeBoever, C., Li, H., Drees, F., Singhal, S., *et al.* (2019). Allele-specific NKX2-5 binding underlies multiple genetic associations with human electrocardiographic traits. *Nat Genet* 51, 1506-1517.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* 57, 289-300.

Bogdan, L.B., L.; Bourque, G. (2019). Transposable elements have contributed human regulatory regions that are activated upon bacterial infection. *bioRxiv*.

Brahimi-Horn, M.C., and Pouyssegur, J. (2007). Oxygen, a source of life and stress. *FEBS Lett* 581, 3582-3591.

Brodehl, A., Ebbinghaus, H., Deutsch, M.A., Gummert, J., Gartner, A., Ratnavadivel, S., and Milting, H. (2019). Human Induced Pluripotent Stem-Cell-Derived Cardiomyocytes as Models for Genetic Cardiomyopathies. *Int J Mol Sci* 20.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10, 1213-1218.

Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., *et al.* (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 47, D1005-D1012.

Burridge, P.W., Li, Y.F., Matsa, E., Wu, H., Ong, S.G., Sharma, A., Holmstrom, A., Chang, A.C., Coronado, M.J., Ebert, A.D., *et al.* (2016). Human induced pluripotent stem cell-derived cardiomyocytes recapitulate the predilection of breast cancer patients to doxorubicin-induced cardiotoxicity. *Nat Med* 22, 547-556.

Calderon, D., Nguyen, M.L.T., Mezger, A., Kathiria, A., Muller, F., Nguyen, V., Lescano, N., Wu, B., Trombetta, J., Ribado, J.V., *et al.* (2019). Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat Genet* 51, 1494-1505.

Caliskan, M., Baker, S.W., Gilad, Y., and Ober, C. (2015). Host genetic variation influences gene expression response to rhinovirus infection. *PLoS Genet* 11, e1005111.

Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C.A., Taylor, M.S., Engstrom, P.G., Frith, M.C., *et al.* (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* 38, 626-635.

Carreau, A., El Hafny-Rahbi, B., Matejuk, A., Grillon, C., and Kieda, C. (2011). Why is the partial oxygen pressure of human tissues a crucial parameter? Small molecules and hypoxia. *J Cell Mol Med* 15, 1239-1253.

Chang, S.L., Huang, Y.L., Lee, M.C., Hu, S., Hsiao, Y.C., Chang, S.W., Chang, C.J., and Chen, P.C. (2018). Association of Varicose Veins With Incident Venous Thromboembolism and Peripheral Artery Disease. *JAMA* 319, 807-817.

Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57-74.

Consortium, G.T., Laboratory, D.A., Coordinating Center -Analysis Working, G., Statistical Methods groups-Analysis Working, G., Enhancing, G.g., Fund, N.I.H.C., Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida, *et al.* (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204-213.

Consortium, I.K.C. (2011). Large-scale gene-centric analysis identifies novel variants for coronary artery disease. *PLoS Genet* 7, e1002260.

Davis, C.A., Hitz, B.C., Sloan, C.A., Chan, E.T., Davidson, J.M., Gabdank, I., Hilton, J.A., Jain, K., Baymuradov, U.K., Narayanan, A.K., *et al.* (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* 46, D794-D801.

de la Roche, J., Angsutararux, P., Kempf, H., Janan, M., Bolesani, E., Thiemann, S., Wojciechowski, D., Coffee, M., Franke, A., Schwanke, K., *et al.* (2019). Comparing human iPSC-cardiomyocytes versus HEK293T cells unveils disease-causing effects of Brugada mutation A735V of NaV1.5 sodium channels. *Sci Rep* 9, 11173.

- Delaneau, O., Ongen, H., Brown, A.A., Fort, A., Panousis, N.I., and Dermitzakis, E.T. (2017). A complete tool set for molecular QTL discovery and analysis. *Nat Commun* 8, 15452.
- Dingar, D., Konecny, F., Zou, J., Sun, X., and von Harsdorf, R. (2012). Anti-apoptotic function of the E2F transcription factor 4 (E2F4)/p130, a member of retinoblastoma gene family in cardiac myocytes. *J Mol Cell Cardiol* 53, 820-828.
- Dombroski, B.A., Nayak, R.R., Ewens, K.G., Ankener, W., Cheung, V.G., and Spielman, R.S. (2010). Gene expression and genetic variation in response to endoplasmic reticulum stress in human cells. *Am J Hum Genet* 86, 719-729.
- Donovan, M.K.R.D.A.-C., A.; D'Antonio, M.; Frazer, K.A. (2019). Cellular deconvolution of GTEx tissues powers eQTL studies to discover thousands of novel disease and cell-type associated regulatory variants. *bioRxiv*.
- Du, J., Leung, A., Trac, C., Lee, M., Parks, B.W., Lusi, A.J., Natarajan, R., and Schones, D.E. (2016). Chromatin variation associated with liver metabolism is mediated by transposable elements. *Epigenetics Chromatin* 9, 28.
- Du, P., Kibbe, W.A., and Lin, S.M. (2008). lumi: a pipeline for processing Illumina microarray. *Bioinformatics* 24, 1547-1548.
- Dzau, V.J., Antman, E.M., Black, H.R., Hayes, D.L., Manson, J.E., Plutzky, J., Popma, J.J., and Stevenson, W. (2006). The cardiovascular disease continuum validated: clinical evidence of improved patient outcomes: part I: Pathophysiology and clinical trial evidence (risk factors through stable coronary artery disease). *Circulation* 114, 2850-2870.
- Feige, E., Yokoyama, S., Levy, C., Khaled, M., Igras, V., Lin, R.J., Lee, S., Widlund, H.R., Granter, S.R., Kung, A.L., *et al.* (2011). Hypoxia-induced transcriptional repression of the melanoma-associated oncogene MITF. *Proc Natl Acad Sci U S A* 108, E924-933.
- Ferrari, R., de Llobet Cucalon, L.I., Di Vona, C., Le Dilly, F., Vidal, E., Lioutas, A., Oliete, J.Q., Jochem, L., Cutts, E., Dieci, G., *et al.* (2019). TFIIIC Binding to Alu Elements Controls Gene Expression via Chromatin Looping and Histone Acetylation. *Mol Cell*.
- Findley, A.S., Richards, A.L., Petrini, C., Alazizi, A., Doman, E., Shanku, A.G., Davis, G.O., Hauff, N., Sorokin, Y., Wen, X., *et al.* (2019). Interpreting Coronary Artery Disease Risk Through Gene-Environment Interactions in Gene Regulation. *Genetics* 213, 651-663.
- Fukaya, E., Flores, A.M., Lindholm, D., Gustafsson, S., Zanetti, D., Ingelsson, E., and Leeper, N.J. (2018). Clinical and Genetic Determinants of Varicose Veins. *Circulation* 138, 2869-2880.
- Ghorbel, M.T., Cherif, M., Jenkins, E., Mokhtari, A., Kenny, D., Angelini, G.D., and Caputo, M. (2010). Transcriptomic analysis of patients with tetralogy of Fallot reveals the effect of chronic hypoxia on myocardial gene expression. *J Thorac Cardiovasc Surg* 140, 337-345 e326.
- Giordano, F.J. (2005). Oxygen, oxidative stress, hypoxia, and heart failure. *J Clin Invest* 115, 500-508.
- Hardin, M., Cho, M.H., McDonald, M.L., Wan, E., Lomas, D.A., Coxson, H.O., MacNee, W., Vestbo, J., Yates, J.C., Agusti, A., *et al.* (2016). A genome-wide analysis of the response to inhaled beta2-agonists in chronic obstructive pulmonary disease. *Pharmacogenomics J* 16, 326-335.
- Hartley, I., Elkhoury, F.F., Heon Shin, J., Xie, B., Gu, X., Gao, Y., Zhou, D., and Haddad, G.G. (2013). Long-lasting changes in DNA methylation following short-term hypoxic exposure in primary hippocampal neuronal cultures. *PLoS One* 8, e77859.
- He, J., Fu, X., Zhang, M., He, F., Li, W., Abdul, M.M., Zhou, J., Sun, L., Chang, C., Li, Y., *et al.* (2019). Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. *Nat Commun* 10, 34.
- Heinig, M., Adriaens, M.E., Schafer, S., van Deutekom, H.W.M., Lodder, E.M., Ware, J.S., Schneider, V., Felkin, L.E., Creemers, E.E., Meder, B., *et al.* (2017). Natural genetic variation of the cardiac transcriptome in non-diseased donors and patients with dilated cardiomyopathy. *Genome Biol* 18, 170.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009a). Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 37, 1-13.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009b). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4, 44-57.

Iyirhiaro, G.O., Zhang, Y., Estey, C., O'Hare, M.J., Safarpour, F., Parsanejad, M., Wang, S., Abdel-Messih, E., Callaghan, S.M., During, M.J., *et al.* (2014). Regulation of ischemic neuronal death by E2F4-p130 protein complexes. *J Biol Chem* 289, 18202-18213.

Jagannathan, L., Cuddapah, S., and Costa, M. (2016). Oxidative stress under ambient and physiological oxygen tension in tissue culture. *Curr Pharmacol Rep* 2, 64-72.

Joehanes, R., Zhang, X., Huan, T., Yao, C., Ying, S.X., Nguyen, Q.T., Demirkale, C.Y., Feolo, M.L., Sharopova, N.R., Sturcke, A., *et al.* (2017). Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies. *Genome Biol* 18, 16.

Jurka, J. (2000). Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet* 16, 418-420.

Kariuki, S.N., Maranville, J.C., Baxter, S.S., Jeong, C., Nakagome, S., Hrusch, C.L., Witonsky, D.B., Sperling, A.I., and Di Rienzo, A. (2016). Mapping Variation in Cellular and Transcriptional Response to 1,25-Dihydroxyvitamin D3 in Peripheral Blood Mononuclear Cells. *PLoS One* 11, e0159779.

Karolchik, D., Hinrichs, A.S., Furey, T.S., Roskin, K.M., Sugnet, C.W., Haussler, D., and Kent, W.J. (2004). The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* 32, D493-496.

Kim, C., Rubin, C.M., and Schmid, C.W. (2001). Genome-wide chromatin remodeling modulates the Alu heat shock response. *Gene* 276, 127-133.

Kim-Hellmuth, S., Bechheim, M., Putz, B., Mohammadi, P., Nedelec, Y., Giangreco, N., Becker, J., Kaiser, V., Fricker, N., Beier, E., *et al.* (2017). Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat Commun* 8, 266.

Knowles, D.A., Burrows, C.K., Blischak, J.D., Patterson, K.M., Serie, D.J., Norton, N., Ober, C., Pritchard, J.K., and Gilad, Y. (2018). Determining the genetic basis of anthracycline-cardiotoxicity by molecular response QTL mapping in induced cardiomyocytes. *Elife* 7.

Koopmann, T.T., Adriaens, M.E., Moerland, P.D., Marsman, R.F., Westerveld, M.L., Lal, S., Zhang, T., Simmons, C.Q., Bacsko, I., dos Remedios, C., *et al.* (2014). Genome-wide identification of expression quantitative trait loci (eQTLs) in human heart. *PLoS One* 9, e97380.

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. *Cell* 175, 598-599.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357-359.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 15, R29.

Lee, B.K., Bhinge, A.A., and Iyer, V.R. (2011). Wide-ranging functions of E2F4 in transcriptional activation and repression revealed by genome-wide analysis. *Nucleic Acids Res* 39, 3558-3573.

Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res* 41, e108.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923-930.

Lin, H., Dolmatova, E.V., Morley, M.P., Lunetta, K.L., McManus, D.D., Magnani, J.W., Margulies, K.B., Hakonarson, H., del Monte, F., Benjamin, E.J., *et al.* (2014). Gene expression and genetic variation in human atria. *Heart Rhythm* 11, 266-271.

Ma, N., Zhang, J.Z., Itzhaki, I., Zhang, S.L., Chen, H., Haddad, F., Kitani, T., Wilson, K.D., Tian, L., Shrestha, R., *et al.* (2018). Determining the Pathogenicity of a Genomic Variant of

Uncertain Significance Using CRISPR/Cas9 and Human-Induced Pluripotent Stem Cells. *Circulation* 138, 2666-2681.

Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* 27, 1696-1697.

Maeda, K., Tsutamoto, T., Wada, A., Hisanaga, T., and Kinoshita, M. (1998). Plasma brain natriuretic peptide as a biochemical marker of high left ventricular end-diastolic pressure in patients with symptomatic left ventricular dysfunction. *Am Heart J* 135, 825-832.

Manry, J., Nedelec, Y., Fava, V.M., Cobat, A., Orlova, M., Thuc, N.V., Thai, V.H., Laval, G., Barreiro, L.B., and Schurr, E. (2017). Deciphering the genetic control of gene expression following *Mycobacterium leprae* antigen stimulation. *PLoS Genet* 13, e1006952.

Maranville, J.C., Luca, F., Richards, A.L., Wen, X., Witonsky, D.B., Baxter, S., Stephens, M., and Di Rienzo, A. (2011). Interactions between glucocorticoid treatment and cis-regulatory polymorphisms contribute to cellular response phenotypes. *PLoS Genet* 7, e1002162.

McDermott-Roe, C., Lv, W., Maximova, T., Wada, S., Bukowy, J., Marquez, M., Lai, S., Shehu, A., Benjamin, I., Geurts, A., *et al.* (2019). Investigation of a dilated cardiomyopathy-associated variant in BAG3 using genome-edited iPSC-derived cardiomyocytes. *JCI Insight* 4.

Medstrand, P., van de Lagemaat, L.N., and Mager, D.L. (2002). Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res* 12, 1483-1495.

Narravula, S., and Colgan, S.P. (2001). Hypoxia-inducible factor 1-mediated inhibition of peroxisome proliferator-activated receptor alpha expression during hypoxia. *J Immunol* 166, 7543-7548.

Nedelec, Y., Sanz, J., Baharian, G., Szpiech, Z.A., Pacis, A., Dumaine, A., Grenier, J.C., Freiman, A., Sams, A.J., Hebert, S., *et al.* (2016). Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell* 167, 657-669 e621.

Nikpay, M., Goel, A., Won, H.H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C.P., Hopewell, J.C., *et al.* (2015). A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* 47, 1121-1130.

Ongen, H., Buil, A., Brown, A.A., Dermitzakis, E.T., and Delaneau, O. (2016). Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* 32, 1479-1485.

Pacis, A., Mailhot-Leonard, F., Tailleux, L., Randolph, H.E., Yotova, V., Dumaine, A., Grenier, J.C., and Barreiro, L.B. (2019). Gene activation precedes DNA demethylation in response to infection in human dendritic cells. *Proc Natl Acad Sci U S A* 116, 6938-6943.

Pacis, A., Tailleux, L., Morin, A.M., Lambourne, J., MacIsaac, J.L., Yotova, V., Dumaine, A., Danckaert, A., Luca, F., Grenier, J.C., *et al.* (2015). Bacterial infection remodels the DNA methylation landscape of human dendritic cells. *Genome Res* 25, 1801-1811.

Panopoulos, A.D., D'Antonio, M., Benaglio, P., Williams, R., Hashem, S.I., Schuldt, B.M., DeBoever, C., Arias, A.D., Garcia, M., Nelson, B.C., *et al.* (2017). iPSCORE: A Resource of 222 iPSC Lines Enabling Functional Characterization of Genetic Variation across a Variety of Cell Types. *Stem Cell Reports* 8, 1086-1100.

Pavlovic, B.J., Blake, L.E., Roux, J., Chavarria, C.I., and Gilad, Y. (2018). A Comparative Assessment of Human and Chimpanzee iPSC-derived Cardiomyocytes with Primary Heart Tissues. *bioRxiv*.

Pickrell, J.K., Marioni, J.C., Pai, A.A., Degner, J.F., Engelhardt, B.E., Nkadori, E., Veyrieras, J.B., Stephens, M., Gilad, Y., and Pritchard, J.K. (2010). Understanding mechanisms underlying human gene expression variation with RNA sequencing. *Nature* 464, 768-772.

Pinto, A.R., Illykh, A., Ivey, M.J., Kuwabara, J.T., D'Antoni, M.L., Debuque, R., Chandran, A., Wang, L., Arora, K., Rosenthal, N.A., *et al.* (2016). Revisiting Cardiac Cellular Composition. *Circ Res* 118, 400-409.

Platt, J.L., Salama, R., Smythies, J., Choudhry, H., Davies, J.O., Hughes, J.R., Ratcliffe, P.J., and Mole, D.R. (2016). Capture-C reveals preformed chromatin interactions between HIF-binding sites and distant promoters. *EMBO Rep* 17, 1410-1421.

Poetsch, A.R., Boulton, S.J., and Luscombe, N.M. (2018). Genomic landscape of oxidative DNA damage and repair reveals regioselective protection from mutagenesis. *Genome Biol* 19, 215.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841-842.

Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* 138, 114-128.

Risso, D., Ngai, J., Speed, T.P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat Biotechnol* 32, 896-902.

Robinson, C.M., Neary, R., Levendale, A., Watson, C.J., and Baugh, J.A. (2012). Hypoxia-induced DNA hypermethylation in human pulmonary fibroblasts is associated with Thy-1 promoter methylation and the development of a pro-fibrotic phenotype. *Respir Res* 13, 74.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140.

Samanta, D., and Semenza, G.L. (2017). Maintenance of redox homeostasis by hypoxia-inducible factors. *Redox Biol* 13, 331-335.

Schodel, J., Oikonomopoulos, S., Ragoussis, J., Pugh, C.W., Ratcliffe, P.J., and Mole, D.R. (2011). High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq. *Blood* 117, e207-217.

Schunkert, H., König, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., Preuss, M., Stewart, A.F., Barbalic, M., Gieger, C., *et al.* (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* 43, 333-338.

Shah, S.H., A.; Roselli, C.; Honghuang, L.; Sveinbjornsson, G.; Fatemifar, G. ... (2019). Genome-wide association study provides new insights into the genetic architecture and pathogenesis of heart failure. *bioRxiv*.

Sigurdsson, M.I., Saddic, L., Heydarpour, M., Chang, T.W., Shekar, P., Aranki, S., Couper, G.S., Shernan, S.K., Muehlschlegel, J.D., and Body, S.C. (2017). Post-operative atrial fibrillation examined using whole-genome RNA sequencing in human left atrial tissue. *BMC Med Genomics* 10, 25.

Smirnov, D.A., Brady, L., Halasa, K., Morley, M., Solomon, S., and Cheung, V.G. (2012). Genetic variation in radiation-induced cell death. *Genome Res* 22, 332-339.

Smith, A., Hubley, R. & Green, P. (2010). RepeatMasker.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 3, Article3.

Stephens, M. (2017). False discovery rates: a new deal. *Biostatistics* 18, 275-294.

Stone, G., Choi, A., Meritxell, O., Gorham, J., Heydarpour, M., Seidman, C.E., Seidman, J.G., Aranki, S.F., Body, S.C., Carey, V.J., *et al.* (2019). Sex differences in gene expression in response to ischemia in the human left ventricular myocardium. *Hum Mol Genet* 28, 1682-1693.

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

Unoki, M., and Nakamura, Y. (2003). Methylation at CpG islands in intron 1 of EGR2 confers enhancer-like activity. *FEBS Lett* 554, 67-72.

van de Geijn, B., McVicker, G., Gilad, Y., and Pritchard, J.K. (2015). WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat Methods* 12, 1061-1063.

Ward, M.C., and Gilad, Y. (2019). A generally conserved response to hypoxia in iPSC-derived cardiomyocytes from humans and chimpanzees. *Elife* 8.

Watson, C.J., Collier, P., Tea, I., Neary, R., Watson, J.A., Robinson, C., Phelan, D., Ledwidge, M.T., McDonald, K.M., McCann, A., *et al.* (2014). Hypoxia-induced epigenetic modifications are associated with cardiac tissue fibrosis and the development of a myofibroblast-like phenotype. *Hum Mol Genet* 23, 2176-2188.

Wei, Y., Tenzen, T., and Ji, H. (2015). Joint analysis of differential gene expression in multiple studies using correlation motifs. *Biostatistics* 16, 31-46.

WHO (2018). World Health Statistics 2018:Monitoring health for the SDGs.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137.