Characterization of caffeine response regulatory variants in vascular endothelial cells

Carly Boye [1], Cynthia Kalita [1], Anthony Findley [1], Adnan Alazizi [1], Julong Wei[1], Xiaoquan Wen[2], Roger Pique-Regi [1,3], Francesca Luca [1,3,4]

[1]Center for Molecular Medicine and Genetics, [2]Department of Biostatistics, University of Michigan, Ann Arbor, MI, [3]Department of Obstetrics and Gynecology, Wayne State University, Detroit, MI, [4]Department of Biology, University of Rome Tor Vergata, Rome, Italy

Abstract

Genetic variants in gene regulatory sequences can modify gene expression and mediate the molecular response to environmental stimuli. In addition, genotype-environment interactions (GxE) contribute to complex traits such as cardiovascular disease. Caffeine is the most widely consumed stimulant and is known to produce a vascular response. To investigate GxE for caffeine, we treated vascular endothelial cells with caffeine and used a massively parallel reporter assay to measure allelic effects on gene regulation for over 43,000 genetic variants. We identified 9,102 variants with allelic effects on gene regulation and 7,152 variants that regulate the gene expression response to caffeine (GxE, FDR<10%). When overlapping our GxE results with fine-mapped artery eQTLs, we dissected their regulatory mechanisms and showed a modulatory role for caffeine. Our results demonstrate that massively parallel reporter assay is a powerful approach to identify and molecularly characterize GxE in the specific context of caffeine consumption.

Introduction

Caffeine is the most widely consumed stimulant in the world (Planning Committee for a Workshop on Potential Health Hazards Associated with Consumption of Caffeine in Food and Dietary Supplements et al., 2014). Caffeine produces a vascular response in the endothelium, causing vaso-dilation. The vascular endothelium, the innermost layer of arteries, is involved in several important functions, including regulation of blood flow, angiogenesis, thrombosis, and coagulation (Hadi et al., 2005; Krüger-Genge et al., 2019). Endothelial dysfunction occurs in diseases such as atherosclerosis and hypertension (Xu et al., 2021), eventually leading to CAD (Matsuzawa and Lerman, 2014). Multiple studies have investigated the role of caffeine in cardiovascular disease (CVD), and more broadly, vascular health in general, with conflicting results (Chieng et al., 2022; Ding et al., 2014; Turnbull et al., 2017) on the role of caffeine in CVD risk. Ding *et al.* (2014) meta-analyzed 36 studies and found no association between heavy coffee consumption and increased risk of CVD (Ding et al., 2014). Similarly, Turnbull *et al.* (2017) observed that moderate caffeine consumption was not associated with an increased risk of CVD, or other cardiovascular

events such as heart failure (Turnbull et al., 2017). Multiple studies suggested that caffeine may be beneficial in reducing the risk of coronary artery disease (CAD) (Choi et al., 2015; Miranda et al., 2018; Voskoboinik et al., 2019), while others provided evidence that caffeine may reduce risk of heart failure, but had no significant effect on the risk of coronary heart disease or CVD (Stevens et al., 2021). Most recently, Chieng *et al.* (2022) found that decaffeinated, ground, and instant coffee significantly decreased CVD risk and mortality (Chieng et al., 2022). The conflicting results from these epidemiological studies may have several causes, including potential interactions between caffeine consumption and other environmental and genetic risk factors. Recent molecular studies investigated the consequences of caffeine exposure on chromatin accessibility and gene expression in vascular endothelial cells (Findley et al., 2019). This study identified response factor motifs for caffeine, defined as transcription factor motifs that are enriched in differentially accessible regions, and demonstrated that caffeine can induce changes in gene regulation in endothelial cells.

Analyzing the changes in gene expression upon exposure to environmental stimuli is a powerful approach to discover genotype-environment interactions (GxE). These molecular GxE result in a different response depending on genotype (Moyerbrailean et al., 2016a, Knowles et al., 2017, 2018;), potentially through allele-specific effects on response factor binding or other environmental-specific gene regulatory mechanisms. Yet regulatory sequences that are differentially bound in response to environmental perturbations are poorly annotated. Single nucleotide polymorphisms (SNPs) within caffeine response factor binding sites were enriched for artery expression quantitative trait loci (eQTLs) colocalized with CAD risk variants (Findley et al., 2019). The results of this study thus suggested that SNPs within regulatory elements active in the presence of caffeine may play a role in CAD risk and pointed to gene-environment interactions in gene regulation as a potential mechanism underlying caffeine modulation of genetic risk for CAD. However, only a limited number of molecular GxE for caffeine have been studied so far, thus the transcription factors and regulatory sequences involved in caffeine GxE remain uncharacterized. Furthermore, it is important to study GxE in the relevant cell type; i.e. endothelial cells which constitute the vascular endothelium. For these reasons, it is crucial to investigate and validate the mechanisms behind caffeine GxE in vascular endothelial cells.

Massively parallel reporter assays (MPRA) have allowed studies of non-coding genetic variants and their role in gene regulation, at unprecedented scale (Arnold et al., 2013; Gordon et al., 2020; Kalita et al., 2018; Melnikov et al., 2012; Patwardhan et al., 2012; Tewhey et al., 2016; Ulirsch et al., 2016; Vockley et al., 2015; Wang et al., 2018). Originally developed to study the gene regulatory potential of promoters and enhancer sequences, MPRA protocols have been further developed to study regulatory genetic variation and fine map association signals (Kalita et al., 2018; Tewhey et al., 2016; Ulirsch et al., 2016; Vockley et al., 2015). MPRAs with synthetic regulatory sequences can test allelic activity for candidate regulatory variants independently of their allele frequency in the population (Kalita et al., 2018; Tewhey et al., 2016; Ulirsch et al.,

2016).. In MPRAs, DNA sequences containing each allele are transfected into cells and RNA-Seq is used to quantify the transcripts for each allele. To directly test allelic effects of tens of thousands of candidate regulatory sequences predicted to affect transcription factor binding (CentiSNPs (Moyerbrailean et al. 2016)), we used an MPRA called Biallelic Targeted STARR-Seq (BiT-STARR-Seq) (Kalita et al., 2018). Only two previous studies have used MPRAs to investigate DNA sequences that regulate the transcriptional response to treatments (Johnson et al., 2018; Shlyueva et al., 2014). One study utilized STARR-Seq to characterize enhancer activity in Drosophila cells upon treatment with the hormone ecdysone, however it did not investigate genotype-environment interactions (GxE) (Shlyueva et al., 2014). The other study utilized STARR-Seq to investigate the response to glucocorticoid treatment in the human cell line A549 (Johnson et al., 2018). Although this study investigated GxE interactions, only a small number of variants were tested as this study was limited to pre-existing variation within the samples (as opposed to designed constructs), and only 2 variants had significant GxE (Johnson et al., 2018). We aim to identify and validate the DNA sequences that regulate the transcriptional response to caffeine in the vascular endothelium, and how genetic variation present in these regulatory elements may affect the transcriptional response to caffeine (figure 1 A).

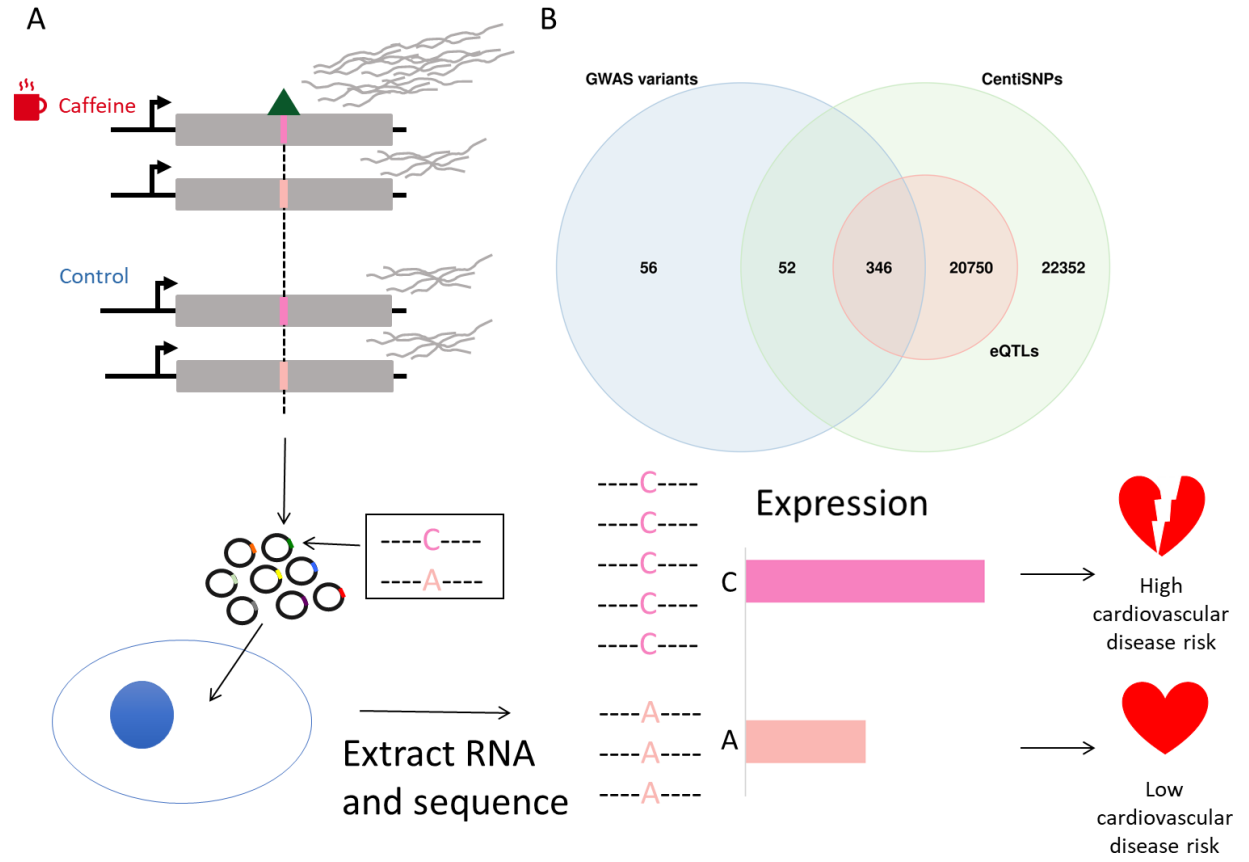


Figure 1. Study design. (A) Genetic variants modulate transcriptional response dependent on environmental conditions. The pink bars represent different alleles present in our constructs, and the green triangle represents a bound transcription factor. These constructs are transfected into

cells, RNA is extracted and sequenced, and then activity is measured for constructs with both alleles. (B) Library composition based on annotation category: SNPs predicted to alter transcription factor binding using a combination of sequence information and experimental data (centiSNPs) (Moyerbrailean et al., 2016b), SNPs associated with complex traits (GWAS), and eQTLs in GTEx.

Results

Active Regulatory Regions in Response to Caffeine

For this study we used a library of constructs that was designed to capture a large number of predicted gene regulatory variants in motifs for hundreds of transcription factor binding sites (figure 1 B). The constructs in our library consisted of self-transcribing enhancer regions containing a minimal promoter, a reporter gene, and the oligonucleotide containing the candidate regulatory SNP. These candidate regulatory SNPs belong to multiple categories, including SNPs predicted to alter transcription factor binding using a combination of sequence information and experimental data (centiSNPs) (Moyerbrailean et al., 2016b), SNPs associated with complex traits (GWAS), and eQTLs in GTEx. In addition, we included 1,676 negative control sequences. To test if these putative regulatory sequences mediate the response to caffeine, we used DESeq2 to test for differential activity of the constructs in cells treated with caffeine compared to cells in the control group (see materials and methods section for the specific model). The library contained motifs in both the forward and reverse orientations within separate constructs. Since these motifs may induce direction-specific effects, we performed the differential activity analysis per each direction separately (see supplemental figure 1), and considered any construct with FDR<10% in either direction as significant. We observed 772 significantly differentially active constructs: 546 upregulated constructs and 226 downregulated constructs (figure 2 A, supplemental table 1), showing that caffeine overall increases the activity of the regulatory elements.

We then focused on differentially active constructs containing a known caffeine response factor as determined based on chromatin accessibility data from endothelial cells treated with caffeine (Findley et al., 2019). We observed that these constructs had lower p-values, as observed in the QQ plot in figure 2 A (inset). To identify any additional transcription factors that may be important for the response to caffeine, we conducted a motif scan for 838 known transcription factor binding motifs using the JASPAR CORE Vertebrates 2022 database (Castro-Mondragon et al. 2022) (supplemental table 2). We found 19 motifs that were enriched for being within differentially active constructs (figure 2 B, supplemental table 3). We found the motif for ZNF423, one of the caffeine response factors, was enriched within the differentially active constructs. The 3 most enriched motifs were NFATC1, NFATC4, and NFATC2. The NFAT transcription factor family is known for their involvement in the Ca^{2+}/NFAT pathway. This signaling pathway plays an important role in maintaining the homeostasis of vascular endothelial cells (Wang et al., 2020) and

contributes to the mediation of proliferation and migration (Johnson et al., 2003; Wang et al., 2020). Thus, improper signaling of the Ca^{2+}/NFAT pathway can induce endothelial dysfunction (Garcia-Vaz et al., 2020; Wang et al., 2020). In diabetic mice, NFAT expression exacerbated atherosclerosis (Blanco et al., 2018; Zetterqvist et al., 2014) and increased foam cell formation (Du et al., 2021). In human coronary artery smooth muscle cells, NFAT signaling mediates vascular calcification (Goettsch et al., 2011). To better understand the regulatory response to caffeine, we then investigated which motifs were enriched for being within upregulated or downregulated constructs separately (supplemental table 3). We observed 19 motifs enriched for being within upregulated constructs (figure 2 C), and 23 motifs enriched for being within downregulated constructs (figure 2 D). Motifs enriched for being within upregulated constructs include the previously mentioned NFATC1 and ZNF423. Motifs enriched for being within downregulated constructs include the previously mentioned NFATC2 and ZNF423. We also observed that the motif for SREBF2, also called SREBP2, is enriched for being within downregulated constructs. In hepatocytes, caffeine is known to suppress SREBF2 activity, which reduces PCSK9 expression, and thus increases LDLR expression, which could be protective against cardiovascular disease (Lebeau et al., 2022). The corresponding transcription factors for these motifs could also play a role in mediating the response to caffeine in vascular endothelial cells. The motif for TEAD4 was also identified as enriched for being within downregulated constructs. Interestingly, a CAD GWAS risk variant disrupts binding of TEAD4 in smooth muscle cells, causing lower expression of p16, which could potentially contribute to the risk identified at this locus (Almontashiri et al., 2015).

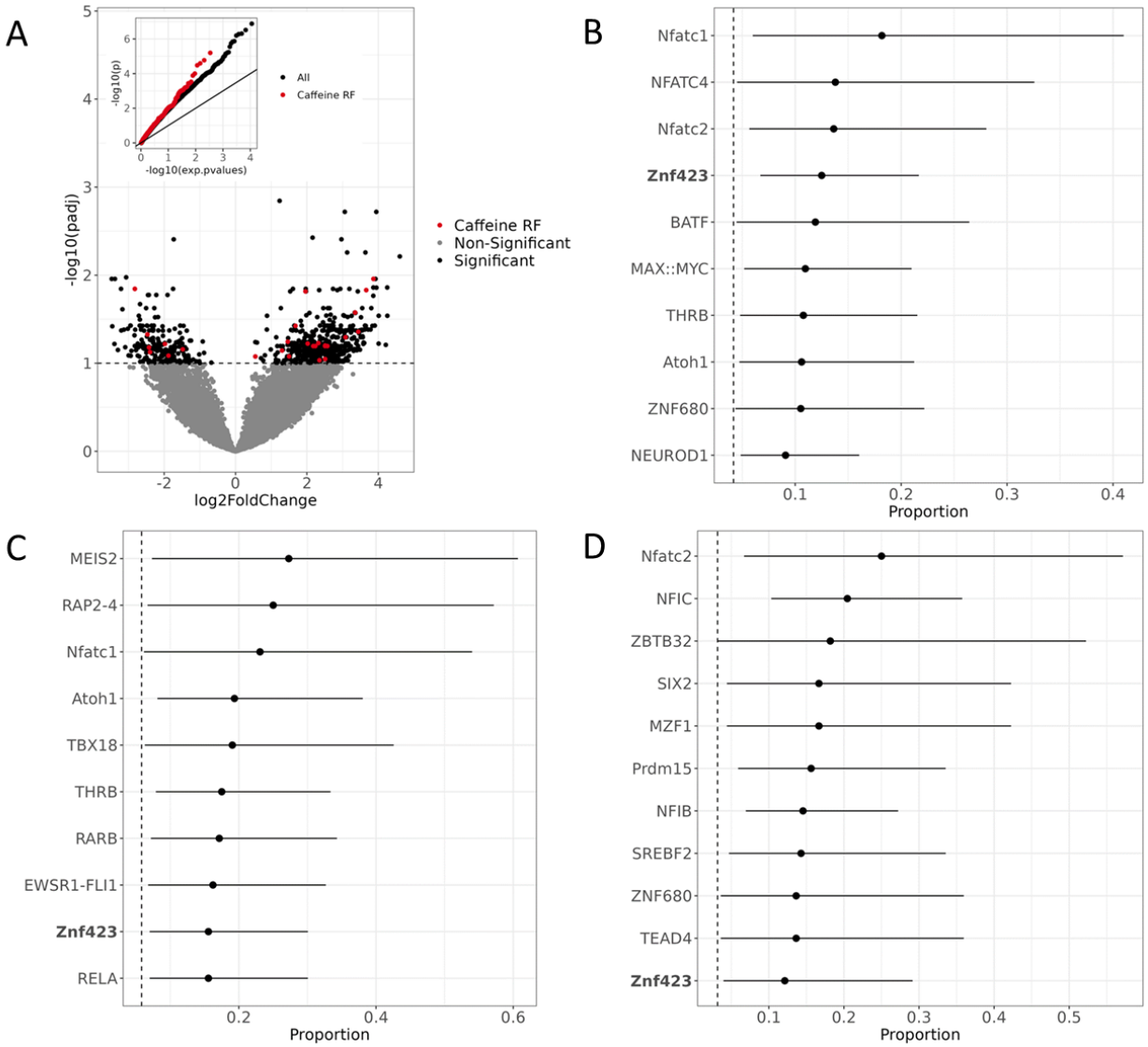


Figure 2. Active regulatory regions in caffeine response in vascular endothelial cells.(A) Volcano plot for DESeq2 results showing constructs differentially active in caffeine. The light red points are significant (FDR < 10%) constructs containing caffeine response factor binding sites, the black points are significant constructs not containing a caffeine response factor binding site and the gray points are non-significant constructs. The inset contains a QQ plot for constructs containing a caffeine response factor binding site (red), or no caffeine response factor binding site (black). (B) Motifs enriched via test of proportions ($p<0.05$) within differentially active constructs. Names of caffeine response factors are bolded. (C) Motifs enriched via test of proportions within upregulated constructs ($p<0.05$). (D) Motifs enriched via test of proportions within downregulated constructs ($p<0.05$).

Allelic Effects on Gene Regulation Within Conditions and in Response to Caffeine

To investigate how genetic variation affects regulatory sequences and their function in cells treated with caffeine and in the control samples, we tested for allele-specific effects (ASE). Since the library contained the same sequence in both the forward and reverse orientations in independent constructs and the regulatory effect may be direction-dependent, we tested for ASE in each SNP/direction pair separately (supplemental table 4, supplemental figure 2). In the control condition, we observed 5,040 SNP/direction pairs with significant ASE out of 30,096 tested (16.7%, FDR<0.1) (figure 3 A, blue). In the caffeine condition, we observed 14,398 SNP/direction pairs with significant ASE out of 49,441 tested (29.1%, FDR<0.1) (figure 3 A, red). We hypothesized that this difference is due to an increase in the number of constructs with regulatory activity in response to caffeine. Indeed we observed 2.42-fold more constructs with increased activity in response to caffeine (546 upregulated versus 226 downregulated), which is reflected in a 1.74-fold increase in the proportion of ASE in caffeine versus control. These results suggest that a substantial number of genetic effects on gene expression may be caffeine-specific and could thus indicate GxE.

To directly test for GxE in the molecular response to caffeine, we tested for conditional allele-specific effects (cASE), where ASE is only significant in one condition, or significantly different between the two conditions. When comparing the water and caffeine conditions, we observed 8,351 significant SNP/direction pairs out of 28,922 tested (28.9%) (figure 3 B, supplemental table 4, supplemental figure 3). We also observed that there are 1.40-fold more cASE variants with stronger allelic effects in the caffeine condition compared to the control condition. This again reflects the increase in regulatory activity in response to caffeine.

To investigate the regulatory architecture underlying these genetic effects on gene expression, we asked whether significant ASE (variants where an allele induces a significant difference in expression compared to the other allele) and cASE variants (variants where ASE is only significant in one condition, or significantly different between the two conditions) were enriched in open chromatin regions as annotated in Findley *et al.* (2019) (Findley et al., 2019). Both ASE (in caffeine or water) and cASE variants were enriched within open chromatin regions, with ASE variants being 4.1-fold enriched (Fisher's exact test p<2.2e-16) (figure 3 C, green), and cASE variants being 1.1-fold enriched (p=0.004) (figure 3 C, purple). This difference in enrichment could be due to the difference between the native chromatin context versus the reporter assay context. Environmental effects on gene regulatory sequences may have a more complex regulatory architecture influenced by the chromatin context that may explain the difference in the enrichment (supplemental table 3).

Genetic regulation of gene expression can be context-dependent, with factors such as cell type (Donovan et al., 2020; Kim-Hellmuth et al., 2020), developmental states (Cuomo et al., 2020; Strober et al., 2019), and environmental stimuli all contributing to gene-environment interactions (GxE-eQTLs, also known as response eQTLs, dynamic eQTLs, context-eQTLs) (see for example

Alasoo et al., 2019; Barreiro et al., 2012; Çalışkan et al., 2015; Findley et al., 2021; Kim-Hellmuth et al., 2017; Maranville et al., 2011; Moyerbrailean et al., 2016a). These context-specific effects can be captured without large cohorts if the appropriate experimental design is applied (Findley et al., 2021). Allele-specific expression experiments in two different conditions can detect GxE in small sample sizes compared to eQTL studies (Moyerbrailean et al., 2016a). To investigate the abundance of GxE missing from large databases such as GTEx (GTEx Consortium 2020), we tested if cASE variants were enriched for artery eQTLs. We conducted a Fisher's exact test and found that cASE variants were 1.92 times less likely to be artery eQTLs ($p<2.2e-16$, figure 3 C, purple). We also observed that ASE variants, which largely indicate genetic effects that are independent of a specific context, were instead 1.2 times enriched for artery eQTLs ($p=6.78e-14$, figure 3 C, green, supplemental table 3), as expected. This observation is consistent with previous findings (Findley et al., 2021), and we hypothesize this is because large eQTL studies such as those in GTEx do not account for different environmental contexts. Therefore, we expect a depletion, as we do not expect environmental context to be captured within large-scale eQTL databases.

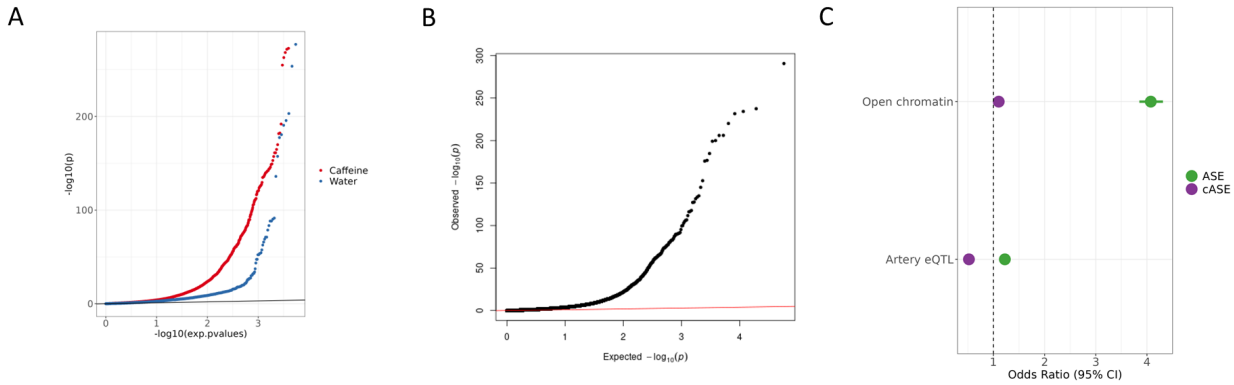


Figure 3. Allelic effects on gene regulation within conditions and in response to caffeine. (A) QQ plot depicting the p-value for ASE in the water (blue) and caffeine (red) conditions. (B) QQ plot depicting the p-value for cASE. (C) Enrichment via Fisher's exact test of ASE (green) and cASE (purple) variants in open chromatin regions and constructs containing artery eQTLs.

Characterizing Allele-Specific Effects and Conditional Allele-Specific Effects Across Transcription Factor Motifs

We hypothesized that the regulatory context defined by the transcription factor motifs present in each construct determines the effect of a genetic variant on expression in caffeine treated cells. We conducted a motif scan of the library of constructs for 838 known transcription factor binding motifs from JASPAR (Castro-Mondragon et al. 2022) (supplemental table 2). We then used a test of proportions to identify any motifs that were disproportionately within constructs containing significant ASE or cASE variants. For constructs containing ASE variants, we observed 71 enriched motifs (figure 4). For constructs containing cASE variants, we observed 8 enriched motifs (figure 4, supplemental table 3). Factors of interest for cardiovascular function include NRF1, which is enriched for constructs containing both ASE and cASE variants, is known to regulate lipid metabolism (Hirotsu et al., 2012; Huss and Kelly, 2004), and is annotated as part of the lipid

metabolism pathway in Reactome (Fabregat et al., 2018). KLF15 and KLF14 are enriched only in constructs with ASE. KLF15 is involved in cardiac lipid metabolism (Prosdocimo et al., 2014, 2015), and KLF14, has previously been associated with cardiovascular disease (Chen et al., 2012; Hu et al., 2018). We also wanted to investigate if ASE and cASE variants were disproportionately present in caffeine response factor binding sites, which may indicate that caffeine response factors' regulatory function may be modified by genetic variation. For this analysis, we used annotations from Findley *et al.* (2019) (Findley et al., 2019), which defines caffeine response factors as transcription factors with motifs that were significantly enriched or depleted in differentially accessible chromatin after treatment with caffeine. We observed constructs containing ASE variants were enriched for 2 out of 4 of the motifs (ZNF423 and PLAG1).

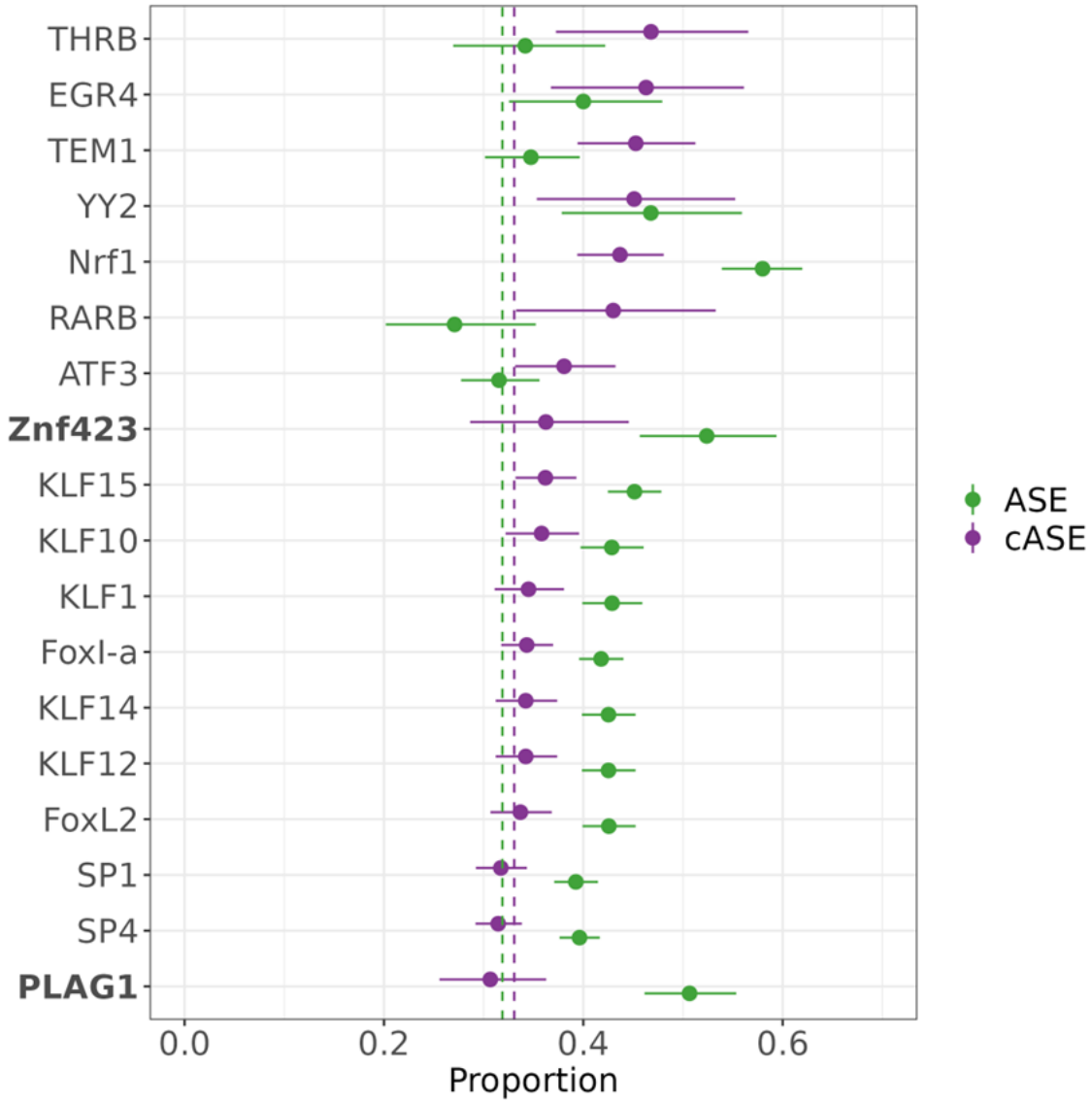


Figure 4. Transcription factors contributing to ASE and cASE. Motifs enriched via test of proportions for significant ASE (green), cASE (purple), or are motifs for caffeine response factors

(bold). The dotted lines represent the baseline proportion (mean number of significant variants within any motif) for ASE (green) and cASE (purple).

Validation of the Regulatory Mechanism for Fine-Mapped Artery eQTLs

Computational fine-mapping is a commonly used method to identify causal variants, often for complex traits, however further functional validation is usually needed to confirm the regulatory mechanism underlying fine-mapped causal variants. We previously showed that artery eQTLs are enriched in caffeine response factor motifs (also see supplemental figure 4) (Findley et al., 2019). We now leverage this finding to fine-map artery eQTLs using DAP-G and the caffeine response factor annotation from Findley et al. (2019) (Findley et al., 2019). In our library we tested 172 fine-mapped variants. We identified significant ASE for 103 SNPs (137 SNP/direction pairs) in either condition, thus validating the regulatory function of these fine mapped causal eQTLs.

Our experimental system also allows us to identify potentially hidden GxE in GTEx. To this end, we investigated whether fine-mapped artery eQTLs have cASE in response to caffeine. We identified 62 fine-mapped artery eQTLs with significant cASE, including 25 SNPs within a caffeine response factor binding site. We then determined which genes are regulated by these 25 eQTLs in GTEx v8 artery tissues and investigated the phenotypic consequences of each allele using pTWAS (supplemental table 5) (Zhang et al., 2020). rs228271 is a fine-mapped artery eQTL, significant cASE variant, and predicted to affect transcription factor binding of E2F1 and E2F4 via centiSNP (Moyerbrailean et al., 2016b) analysis. CentiSNP uses a combination of sequence information and experimental data to predict which allele will increase binding of transcription factors (Moyerbrailean et al., 2016b). Binding of transcription factors E2F1 and E2F4 is predicted to be stronger for the alternate allele (Moyerbrailean et al., 2016b). This binding, and thus the alternate allele, causes higher activity as measured in our BiT-STARR-seq experiments in the absence of caffeine, and confirmed in artery eQTL for PIP4K2B (figure 5 A). Higher expression of *PIP4K2B* is associated with hypertension (Zhang et al., 2020). In the presence of caffeine, the genes for E2F1 and E2F4 are downregulated (Moyerbrailean et al., 2016a), this results in overall lower regulatory activity of the construct containing rs228271 and lack of ASE (figure 5 B). As a consequence, caffeine counteracts the genetic risk of the alternate allele and should thus be protective for hypertension in individuals with the AA genotype.

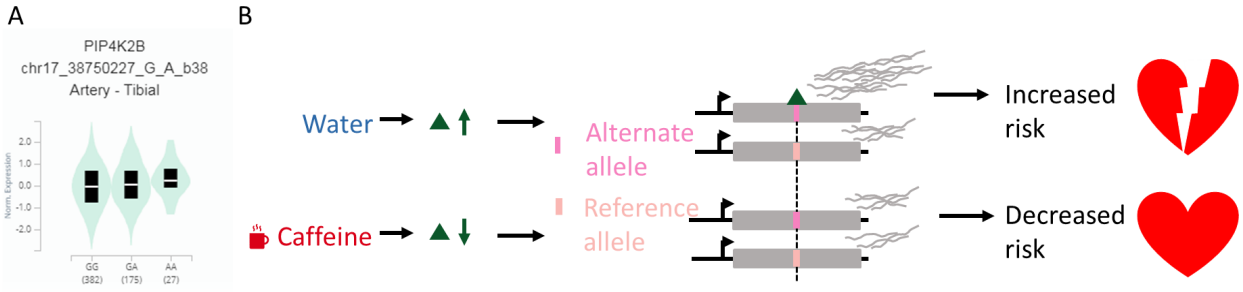


Figure 5: Proposed Example Mechanism. (A) GTEx eQTL violin plot for an eGene this SNP is an artery eQTL for. (B) Diagram of the proposed mechanism for rs228271.

Discussion

This study utilized the MPRA BiT-STARR-Seq to identify gene regulatory activity in vascular endothelial cells exposed to caffeine. By utilizing BiT-STARR-Seq, we were able to identify a molecular response to caffeine, allele-specific effects, and conditional allele-specific effects (cASE). By combining our results with pre-existing annotations, we were able to characterize variants exhibiting cASE and identify potential mechanisms for some of these variants.

Heart disease is one of the leading causes of death in the United States according to the CDC (http://wonder.cdc.gov/ucd-icd10.html). The most common type of heart disease is coronary artery disease (CAD), which affects over 18 million adults over 20 (Fryar et al., 2012). Common risk factors of CAD include hypertension, high cholesterol levels, and family history (Brown et al., 2022; Hajar, 2017). CAD occurs when plaques form in the arteries (atherosclerosis), causing a narrowing of the artery, which reduces blood flow to the heart. The innermost layer of the artery is composed of endothelial cells. The endothelium is involved in several important functions, including regulation of blood flow, angiogenesis, thrombosis, and coagulation (Hadi et al., 2005; Krüger-Genge et al., 2019). Endothelial dysfunction occurs in diseases such as atherosclerosis and hypertension (Xu et al., 2021), eventually leading to CAD (Matsuzawa and Lerman, 2014). The molecular mechanisms behind endothelial dysfunction and the resulting diseases are largely unknown. Characterizing these molecular mechanisms is crucial in order to gain a more complete understanding of these disease phenotypes. Additionally, although caffeine is known to produce a vascular response, the current literature does not come to a consensus on the role of caffeine in CAD risk. Here we characterized the regulatory response of non-coding variants to caffeine in vascular endothelial cells using a massively parallel reporter assay (MPRA).

BiT-STARR-Seq, the MPRA used in this study, has several advantages over other methods used to detect genotype-environment interactions (GxE). One common method of detecting GxE is response eQTL mapping, which includes collecting samples from large cohorts and exposing those cells to environmental perturbations (see for example Alasoo et al., 2018, 2019; Barreiro et al., 2012; Çalışkan et al., 2015; Fairfax et al., 2014; Huang et al., 2020; Kim-Hellmuth et al., 2017; Knowles et al., 2018; Lee et al., 2014; Mangravite et al., 2013; Manry et al., 2017; Maranville et al., 2011; Nédélec et al., 2016; Quach et al., 2016). This method has several disadvantages, as it cannot easily interrogate rare variants, relies on variation existing in a cohort (instead of investigating variants of interest), and requires larger cohort sizes to have enough power to detect GxE. In contrast, because our method uses a designed library of constructs (Kalita et al., 2018), we are able to interrogate rare variants easily, as our constructs are synthesized. Similarly, we can design a library of specific variants to investigate (such as candidate regulatory variants) instead of relying on variation within a cohort (Kalita et al., 2018). BiT-STARR-Seq also allows us to directly compare 2 alleles within the same sequence context without requiring a large cohort.

Despite the advantages of BiT-STARR-Seq, unlike response eQTL mapping, we are not interrogating these SNPs in their native chromatin context. Future work may include using genome-editing tools such as CRISPR to directly insert the desired variants in their endogenous locations in the genome. For our study, we determined BiT-STARR-Seq to be the ideal assay to determine GxE for a large number of SNPs.

We observed a regulatory response to caffeine treatment, consistent with previous studies in the same cell type (Findley et al., 2019; Moyerbrailean et al., 2016a). These results suggest that caffeine exposure significantly changes the regulatory activity of vascular endothelial cells, which may have important implications regarding the impact of lifestyle in CAD. As caffeine may modulate gene regulatory activity, the resulting impact on gene expression may increase or decrease CAD risk. In addition, we identified novel transcription factors contributing to the regulatory response to caffeine including several NFAT transcription factors, and SREBF2. NFAT transcription factors are largely known for their role in the Ca^{2+}/NFAT signaling pathway, where Ca^{2+} binds to calmodulin, stimulating calcineurin, which then causes NFAT factors to localize in the nucleus (Crabtree and Olson, 2002; Klee et al., 1998). Caffeine is known to cause an increase in Ca^{2+} in human aortic endothelial cells (Corda et al., 1995), so it is understandable that we find these factors enriched for being within constructs that respond to caffeine exposure. SREBF2, also known as SREBP2, is involved in sterol homeostasis (Horton et al., 2003). This result implies these novel transcription factors important for the regulatory response to caffeine may also contribute to understanding the role of caffeine in CAD risk. This coincides with findings that caffeine exposure can alter expression of genes, including those for transcription factors in mouse cardiomyocytes (Fang et al., 2014). Another study aimed to uncover mechanisms relevant to cardiovascular disease upon caffeine exposure and found that caffeine inhibits the transcription factor SREBP2, which causes an overall protective effect against cardiovascular disease (Lebeau et al., 2022). These results coincide with our findings.

Non-coding regions of the genome contain regulatory variants that modulate gene expression, and thus are relevant to complex disease such as CAD. In this study we identify and characterize over 9,000 variants exhibiting allele-specific effects. Numerous non-coding variants have been implicated in CAD risk via GWAS (van der Harst and Verweij, 2018; Hartmann et al., 2022; Kessler and Schunkert, 2021; Koyama et al., 2020; Nikpay et al., 2015; Temprano-Sagrera et al., 2022), but they are generally uncharacterized. Only one lead non-coding variant has been thoroughly investigated, leading the authors to propose and validate a molecular mechanism connecting expression of the EDN1 gene to the phenotypic outcome (Gupta et al., 2017; Wang and Musunuru, 2018). These specific mechanisms that detail how these non-coding variants contribute to CAD will be critical in understanding CAD risk and ultimately developing clinical treatments. While understanding various genetic risk factors for CAD is important, genotype-environment interactions for these variants also have an impact on phenotype and have not been widely studied.

Since previous studies have shown that GxE-eQTL can modulate complex disease risk, we expect that GxE detected in our assay may be relevant to CAD (Alasoo et al., 2018, 2019; Barreiro et al., 2012; Çalışkan et al., 2015; Fairfax et al., 2014; Findley et al., 2021; Huang et al., 2020; Kim-Hellmuth et al., 2017; Knowles et al., 2018; Lee et al., 2014; Mangravite et al., 2013; Manry et al., 2017; Maranville et al., 2011; Nédélec et al., 2016; Quach et al., 2016). We tested for conditional allele-specific effects (cASE), which occurs when allele-specific effects are only significant in one condition, or significantly different between the two conditions. This analysis identifies genotype-environment interactions (GxE) which are important in understanding disease risk while accounting for genetic and environmental context. We identified 7,152 variants that regulate the gene expression response to caffeine. These cASE variants were also depleted for being eQTLs in artery tissue (OR=0.52). This observation is consistent with previous findings (Findley et al., 2021), and we hypothesize this is because eQTLs tend to be consistent across treatment conditions. To better understand GxE and further validate our results, we investigated specific cASE variants individually. By fine-mapping artery eQTLs and combining the data collected from our assay with pre-existing annotations, we were able to produce a potential mechanism for the cASE variant rs228271. Though no other studies validate this specific mechanism, there is evidence that the E2F family of transcription factors may contribute to hypertension. A recent study found E2F1/2/3a expression restores the function of dihydrofolate reductase, which prevents endothelial nitric oxide synthase 3 from uncoupling, thus lowering blood pressure (Li et al., 2019). This information improves the understanding of GxE. In addition, we were able to link this mechanism to a phenotype using TWAS, exhibiting the importance of accounting for environmental contexts and how they may contribute to phenotypic outcomes. By studying different environmental contexts, we can identify that, in this instance, caffeine can reduce risk of poor cardiovascular health outcomes. If environmental context was not considered and this work was conducted solely in the control condition, the decreased risk induced by caffeine would not have been observed.

Although we investigate GxE for caffeine in vascular endothelial cells, our experimental approach can be applied to various different complex diseases and their relevant cell types and treatments. To further validate our work, genome editing tools could be used to investigate the effect of these variants in their native chromatin context. Additional validation could include allele-specific and condition-specific transcription factor binding assays (such as electrophoretic mobility shift assays) for the fine-mapped variants.

Our study demonstrates the importance of considering environmental contexts when investigating gene regulatory activity, as we identify several thousand instances of GxE in our library of candidate regulatory variants. Our data, combined with pre-existing annotations, allowed us to identify transcription factors involved in GxE in caffeine, and describe specific potential molecular mechanisms for some of these GxE. Our results provide important insights into the molecular regulatory effect of caffeine exposure and GxE for caffeine in vascular endothelial cells.

Materials and Methods

Cell Culture

Human umbilical vein endothelial cells (HUVECs) were obtained from Lonza (catalog #: CC-2517-0000315288). Cells were cultured at 37°C with 5% CO2 and seeded at 5,000 cells/cm^2. EGM-2 growth medium (Lonza) was used to culture the cells.

Treatment

Treatment concentration was the same as used in previous studies (Findley et al., 2019; Moyerbrailean et al., 2016a). We used a caffeine concentration of 1.16×10−3 M. In addition, water was used as a vehicle control, as that was the solvent used to prepare the caffeine treatment.

BiT-STARR-Seq Library Design

Briefly, the library of constructs consisted of several categories of regulatory variants, including eQTLs(Innocenti et al., 2011; Wen et al., 2015), SNPs predicted to disrupt transcription factor binding (centiSNPs) (Moyerbrailean et al., 2016b), and SNPs associated with complex traits in GWAS (Pickrell, 2014). Negative controls that were not predicted to have a regulatory effect were also included in the library (Moyerbrailean et al., 2016b). We designed 43,556 constructs in total. Each construct had a length of 200 nucleotides with the test nucleotide in the middle of the construct. The library used is the same as reported in (Kalita et al., 2018).

BiT-STARR-Seq Experiments

Oligonucleotides were synthesized and used to create a DNA plasmid library, which was sequenced and used as a subsequent input for the ASE analysis. The DNA library was transfected into HUVECs using the Lonza Nucleofector X platform. Cells were electroporated using the DS-120 setting with primary cell solution P5. Caffeine was added at 1.16×10−3 M after transfection. Cells were incubated for 24 hours and lysed. We completed 6 replicates per treatment condition (caffeine and the water vehicle control), or 12 replicates in total.

Library Preparation and Sequencing

RNA was extracted using the RNeasy Plus Mini kit (Qiagen, catalog #: 74136). A cDNA library was prepared using the Superscript III First-Strand Synthesis kit (Invitrogen, catalog #: 18080-400). Sequencing was completed using the Illumina Nextseq 500 to generate 125 cycles for read 1, 30 cycles for read 2, 8 cycles for the fixed multiplexing index 2 and 10 cycles for index 1 (variable barcode). Average sequencing depth per library was 39,235,611 reads, for a total of 470,827,333 reads (supplemental figure 5, supplemental table 6).

Processing Sequence Data

To analyze the RNA-seq data, we began by demultiplexing our data using the bcl2fastq software to create demultiplexed FASTQ files. We then aligned to hg19 using HISAT2. Afterwards, we applied a filter to ensure the UMIs present match the expected UMI pattern (RDHBVDHBVD). Reads with short UMIs or those that do not match the expected sequence were removed. The resulting BAM files were then deduplicated using UMItools. We ran samtools mpileup followed by bcftools query to output read counts per each allele/direction combination.

Differential Activity Analysis

To test for a molecular response to caffeine, we used the R package DESeq2 (Love et al., 2014). To determine which model would best test for a molecular response to caffeine, we completed principal component analysis to identify major sources of variation. We identified that the first PC clearly represented allelic effects (supplemental figure 6, also see supplemental figure 7), thus we included allele (reference or alternate) as part of our model. Our model tested the effect of treatment, correcting for allele (~allele + treatment), as we observed a strong allelic effect.We ran DESeq2 for each direction (see supplemental figure 1), as the library contained motifs in both the forward and reverse orientations within separate constructs. We considered constructs as significant with Benjamini-Hochberg FDR < 0.1.

ASE Analysis

To test for ASE, we utilized an R package created by our group called quantitative allele-specific analysis of reads (QuASAR-MPRA) (Kalita et al., 2017). QuASAR-MPRA is an expansion of the QuASAR package which allows for analysis of barcoded MPRA data. QuASAR-MPRA uses a beta-binomial model and accounts for uneven initial allelic proportions present in the DNA library. We use the fitQuasarMpra() function to test for ASE in each condition separately. We use unweighted betas from the QuASAR-MPRA output to create a z-score and subsequent p-values and FDR. Significant ASE is then defined as having an FDR of less than 0.1.

cASE Analysis

To test for cASE, we used a method previously developed by our lab called differential allele-specific test, or ΔAST. The calculation for this parameter ΔZ, is provided in (Moyerbrailean et al., 2016a). The QuASAR-MPRA package outputs betas for the treatment (βT) and the control (βC), as well as the standard error (se) for both groups, which are used to calculate ΔAST. From this ΔZ parameter, we use the pnorm() function in R to calculate a p-value for each variant, and then use the p.adjust() function to perform multiple test correction using the Benjamini-Hochberg procedure.

PWMScan and Test of Proportion

The motif scan was completed using the PWMScan tool, using all PWMs within the JASPAR (Castro-Mondragon et al. 2022) 2022 CORE database (838 motifs). A threshold of -t 10 (base 2) was used for the motif scan, which was restricted to the regions of our designed constructs. Once

the scan was complete, motifs that were present less than 100 times in the library were removed. For differentially active constructs in response to caffeine, 222 motifs passed this filter, for motifs containing ASE variants, 452 motifs passed the filter, and for motifs containing cASE variants, 346 motifs passed the filter. A test of proportion was performed per each transcription factor, where the null proportion was the total number of significant constructs/variants containing/within motifs (differentially active construct, ASE or cASE) divided by the total number of non-significant constructs/variants containing/within motifs. The test was done per motif, where the proportion being tested is the same as the null proportion, but conducted per motif rather than across all motifs. The related calculations are shown below:

Null proportion = n_a/n_b
Motif proportion = n_c/n_d

n_a = Number of constructs/variants of interest containing/within any motif
n_b = Number of constructs/variants containing/within any motif
n_c = Number of constructs/variants of interest containing/within specific motif
n_d = Number of constructs/variants containing/within specific motif

Open Chromatin Region Enrichment Analysis
To test if certain variants were within open chromatin regions, we obtained the list of differentially accessible regions tested in (Findley et al., 2019). We considered any accessible region (annotated as differentially accessible or not). Bedtools bed intersect tool was used to complete the overlap with the appropriate datasets (DESeq2, ASE, or cASE results), resulting in a list of constructs or SNPs that were within open chromatin regions. This annotation was then used to complete the Fisher's exact test.

Artery eQTL Enrichment Analysis
To determine if certain variants were artery eQTLs, we obtained GTEx v8 (Castro-Mondragon et al. 2022)data for aorta, coronary, and tibial artery tissues. We then intersected the list of variants of interest with the list of artery eQTLs. To determine if artery eQTLs were within differentially active constructs, bedtools intersect was used, resulting in a list of constructs that contained artery eQTLs. This annotation was then used to complete the Fisher's exact test.

DAP-G Analysis
Based on a previous study (Findley et al., 2019), we define caffeine response factors as transcription factors with motifs that were significantly enriched or depleted in differentially accessible chromatin after treatment with caffeine. We annotated genetic variants into two categories: (1) genetic variants in motifs for response factors (2) genetic variants in motifs for transcription factors that are not caffeine response factors. By integration of these genetic variants annotation, we estimated the probability of each SNP regulating gene expression in a Bayesian

hierarchical model using TORUS (Wen, 2016). These probabilities are then used in DAP-G (Zhang et al., 2020) to fine-map eQTLs from all 3 artery tissues in GTEx V8. 364,427,888 eQTLs were fine-mapped across 3 artery tissues. We filtered for a posterior inclusion probability of greater than 0.9 and retained. We fine mapped 4,735 SNPs, 465 of which were within our designed library. 317 SNP/direction pairs had significant ASE (corresponding to 172 SNPs).

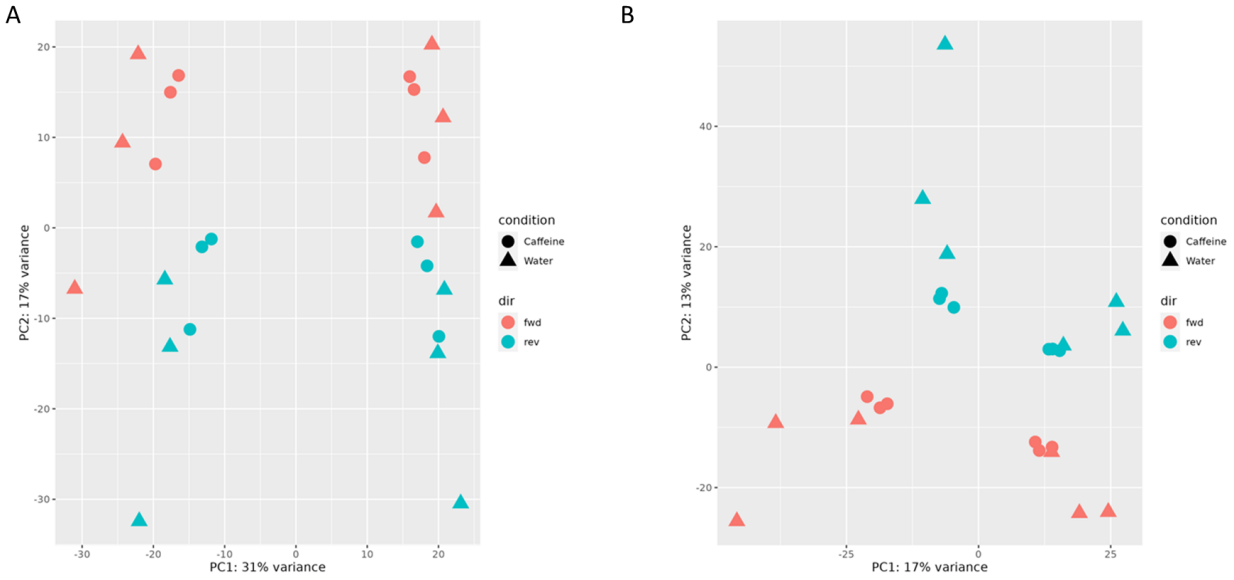
pTWAS overlap

The list of eGenes was overlapped with genes associated with complex traits by pTWAS (available at https://github.com/xqwen/ptwas, supplemental table 2 in (Zhang et al., 2020)).

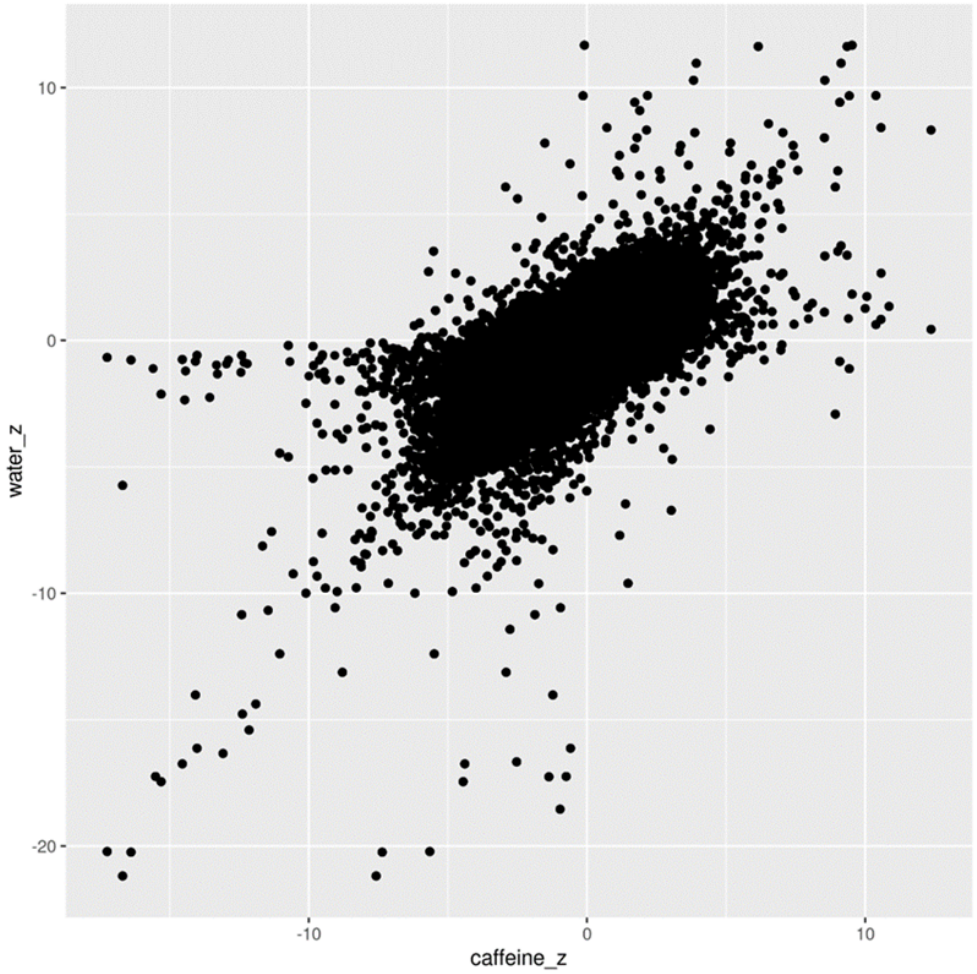
Acknowledgements

We thank the members of the Luca and Pique-Regi laboratories. This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health (R01GM109215).

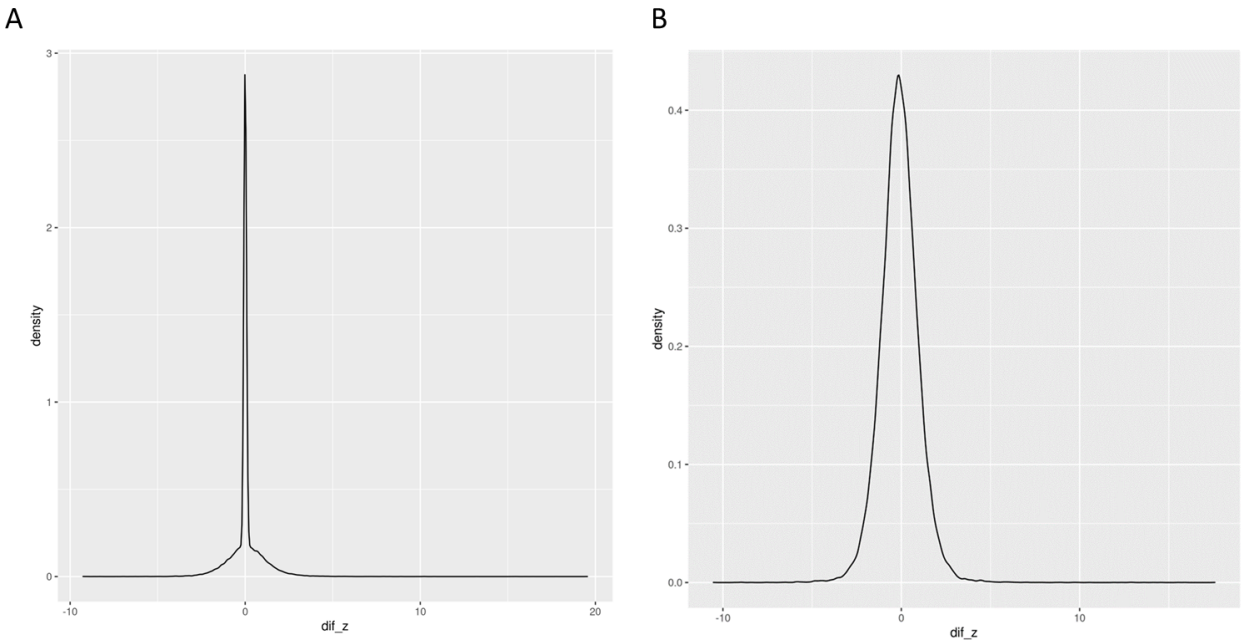
Supplemental Materials



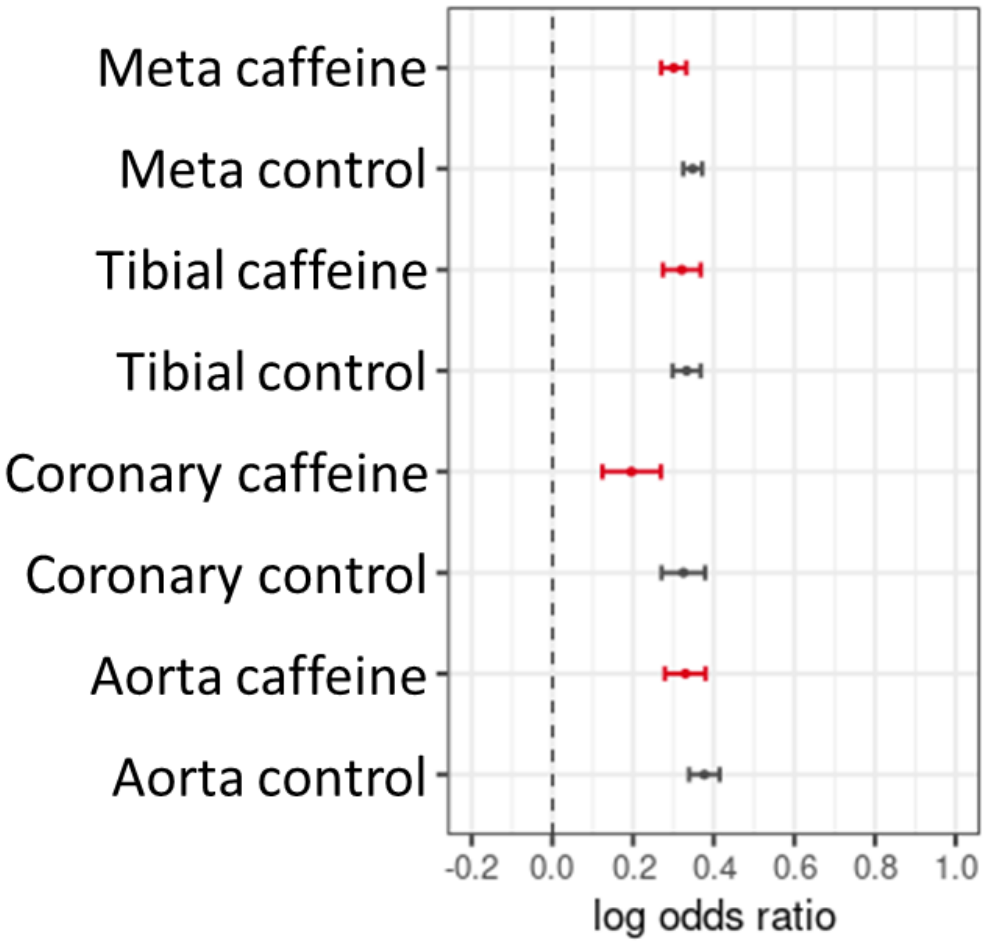
Supplemental figure 1: PCA plots from read count data in the first (A) and second (B) batch (experiment), annotated by direction.



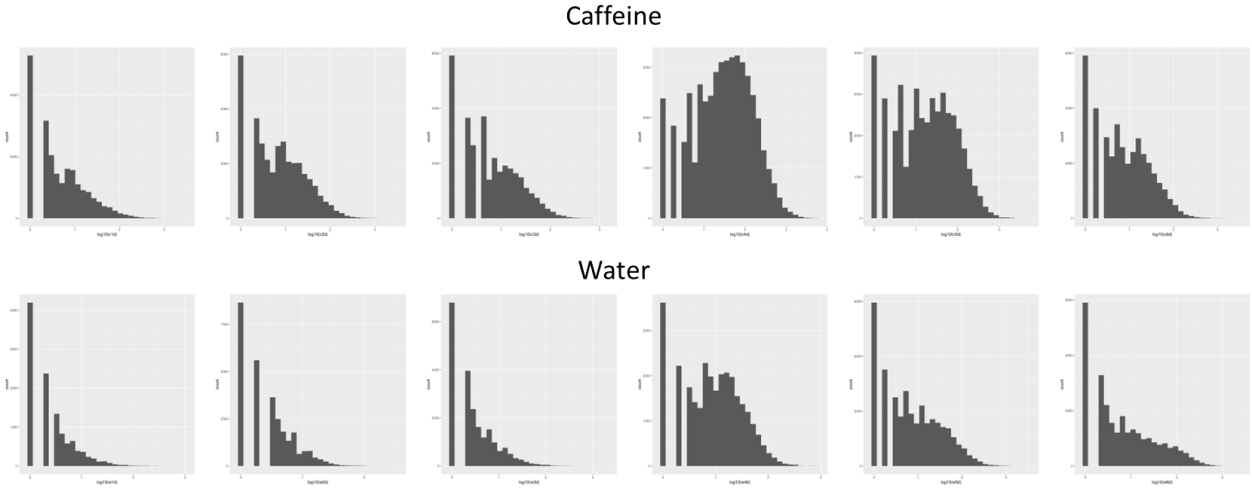
Supplemental figure 2: Z-score scatter plots from ASE analysis comparing water and ethanol (A) and caffeine and water (B).



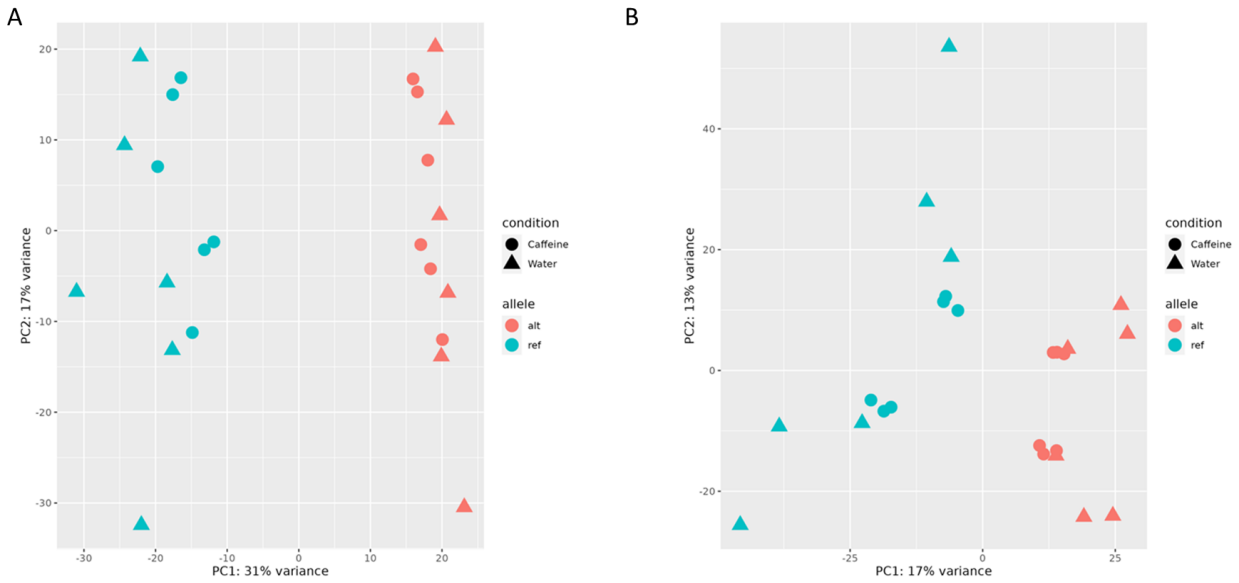
Supplemental figure 3: Distribution of delta Z scores from cASE analysis comparing caffeine and water.



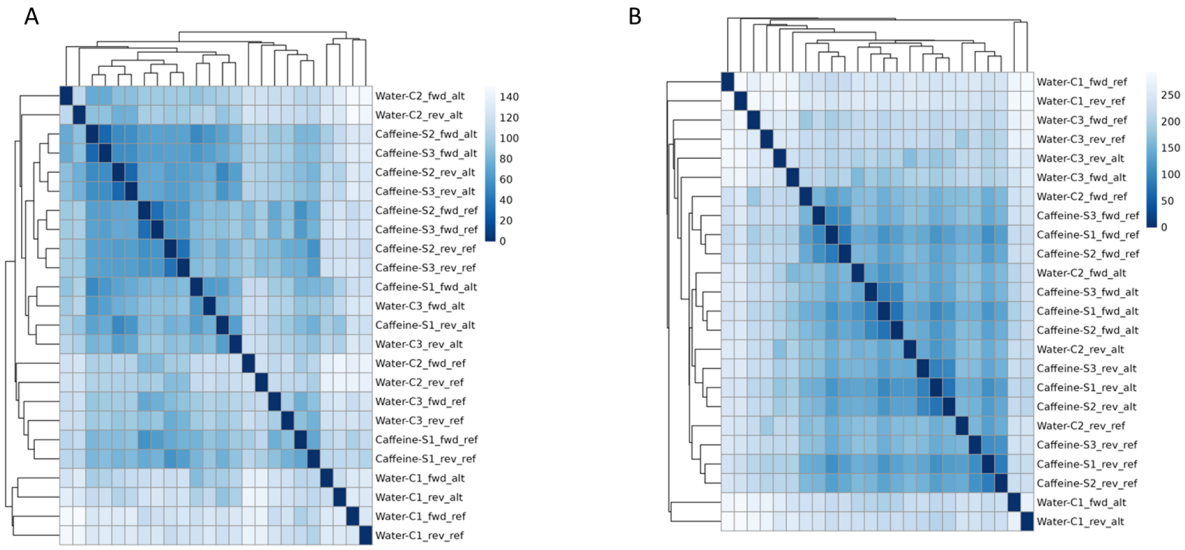
Supplemental figure 4: Artery eQTL enrichment for footprint SNPs (control, black) or caffeine response factors (caffeine, red). “Meta” is the combination of all 3 artery tissue types in GTEx.



Supplemental figure 5: Coverage histograms for all 12 libraries used in this study. Caffeine libraries are on top, and water libraries are on bottom.



Supplemental figure 6: PCA plots from read count data in the first (A) and second (B) batch (experiment), annotated by allele.



Supplemental figure 7: Heatmap plots from read count data in the first (A) and second (B) batch (experiment).

Supplemental tables are available at: https://zenodo.org/record/7327509#.Y30U5XbMKUl

Supplemental table 1: Differential Activity Results. Output from DESeq2.

Supplemental table 2: PWMScan Results. Includes all variants in the designed library that are within a motif from the JASPAR 2022 CORE Vertebrates database determined by PWMScan.

Supplemental table 3: Characterization of Regulatory Regions and Variants. Includes contingency tables for all Fisher's exact tests reported in this study.

Supplemental table 4: Allele-Specific Effects and Conditional Allele-Specific Effects Results. Output from ASE/cASE analysis.

Supplemental table 5: Fine-Mapped Artery eQTLs with Significant cASE.

Supplemental table 6: Read Counts. Read counts for all libraries at different steps of the data processing pipeline.

References

Alasoo, K., Rodrigues, J., Mukhopadhyay, S., Knights, A.J., Mann, A.L., Kundu, K., HIPSCI Consortium, 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. https://doi.org/10.1038/s41588-018-0046-7.

Alasoo, K., Rodrigues, J., Danesh, J., Freitag, D.F., Paul, D.S., and Gaffney, D.J. (2019). Genetic effects on promoter usage are highly context-specific and contribute to complex traits. Elife *8*. https://doi.org/10.7554/eLife.41673.

Almontashiri, N.A.M., Antoine, D., Zhou, X., Vilmundarson, R.O., Zhang, S.X., Hao, K.N., Chen, H.-H., and Stewart, A.F.R. (2015). 9p21.3 Coronary Artery Disease Risk Variants Disrupt TEAD Transcription Factor-Dependent Transforming Growth Factor β Regulation of p16 Expression in Human Aortic Smooth Muscle Cells. Circulation *132*, 1969–1978. .

Arnold, C.D., Gerlach, D., Stelzer, C., Boryń, Ł.M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science *339*, 1074–1077. .

Barreiro, L.B., Tailleux, 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. .

Blanco, F., Heinonen, S.E., Gurzeler, E., Berglund, L.M., Dutius Andersson, A.-M., Kotova, O., Jönsson-Rylander, A.-C., Ylä-Herttuala, S., and Gomez, M.F. (2018). In vivo inhibition of nuclear factor of activated T-cells leads to atherosclerotic plaque regression in IGF-II/LDLR-/-ApoB100/100 mice. Diab. Vasc. Dis. Res. *15*, 302–313. .

Brown, J.C., Gerhardt, T.E., and Kwon, E. (2022). Risk Factors For Coronary Artery Disease. In StatPearls, (Treasure Island (FL): StatPearls Publishing),.

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

Chen, X., Li, S., Yang, Y., Yang, X., Liu, Y., Liu, Y., Hu, W., Jin, L., and Wang, X. (2012). Genome-wide association study validation identifies novel loci for atherosclerotic cardiovascular disease. J. Thromb. Haemost. *10*, 1508–1514. .

Chieng, D., Canovas, R., Segan, L., Sugumar, H., Voskoboinik, A., Prabhu, S., Ling, L.H., Lee, G., Morton, J.B., Kaye, D.M., et al. (2022). The impact of coffee subtypes on incident cardiovascular disease, arrhythmias, and mortality: long-term outcomes from the UK Biobank. Eur. J. Prev. Cardiol. https://doi.org/10.1093/eurjpc/zwac189.

Choi, Y., Chang, Y., Ryu, S., Cho, J., Rampal, S., Zhang, Y., Ahn, J., Lima, J.A.C., Shin, H., and Guallar, E. (2015). Coffee consumption and coronary artery calcium in young and middle-aged

asymptomatic adults. Heart *101*, 686–691. .

Corda, S., Spurgeon, H.A., Lakatta, E.G., Capogrossi, M.C., and Ziegelstein, R.C. (1995). Endoplasmic reticulum Ca2+ depletion unmasks a caffeine-induced Ca2+ influx in human aortic endothelial cells. Circ. Res. *77*, 927–935. .

Crabtree, G.R., and Olson, E.N. (2002). NFAT signaling: choreographing the social lives of cells. Cell *109 Suppl*, S67–S79. .

Cuomo, A.S.E., Seaton, D.D., McCarthy, D.J., Martinez, I., Bonder, M.J., Garcia-Bernardo, J., Amatya, S., Madrigal, P., Isaacson, A., Buettner, F., et al. (2020). Single-cell RNA-sequencing of differentiating iPS cells reveals dynamic genetic effects on gene expression. Nat. Commun. *11*, 810. .

Ding, M., Bhupathiraju, S.N., Satija, A., van Dam, R.M., and Hu, F.B. (2014). Long-term coffee consumption and risk of cardiovascular disease: a systematic review and a dose-response meta-analysis of prospective cohort studies. Circulation *129*, 643–659. .

Donovan, M.K.R., D'Antonio-Chronowska, A., D'Antonio, M., and Frazer, K.A. (2020). Cellular deconvolution of GTEx tissues powers discovery of disease and cell-type associated regulatory variants. Nat. Commun. *11*, 955. .

Du, M., Yang, L., Liu, B., Yang, L., Mao, X., Liang, M., and Huang, K. (2021). Inhibition of NFAT suppresses foam cell formation and the development of diet-induced atherosclerosis. FASEB J. *35*, e21951. .

Fabregat, A., Sidiropoulos, K., Viteri, G., Marin-Garcia, P., Ping, P., Stein, L., D'Eustachio, P., and Hermjakob, H. (2018). Reactome diagram viewer: data structures and strategies to boost performance. Bioinformatics *34*, 1208–1214. .

Fairfax, B.P., Humburg, P., Makino, S., Naranbhai, V., Wong, D., Lau, E., Jostins, L., Plant, K., Andrews, R., McGee, C., et al. (2014). Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. Science *343*, 1246949. .

Fang, X., Mei, W., Barbazuk, W.B., Rivkees, S.A., and Wendler, C.C. (2014). Caffeine exposure alters cardiac gene expression in embryonic cardiomyocytes. Am. J. Physiol. Regul. Integr. Comp. Physiol. *307*, R1471–R1487. .

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. .

Findley, A.S., Monziani, A., Richards, A.L., Rhodes, K., Ward, M.C., Kalita, C.A., Alazizi, A., Pazokitoroudi, A., Sankararaman, S., Wen, X., et al. (2021). Functional dynamic genetic effects on gene regulation are specific to particular cell types and environmental conditions. Elife *10*. https://doi.org/10.7554/eLife.67077.

Fryar, C.D., Chen, T.-C., and Li, X. (2012). Prevalence of uncontrolled risk factors for

cardiovascular disease: United States, 1999-2010. NCHS Data Brief 1–8. .

Garcia-Vaz, E., McNeilly, A.D., Berglund, L.M., Ahmad, A., Gallagher, J.R., Dutius Andersson, A.-M., McCrimmon, R.J., Zetterqvist, A.V., Gomez, M.F., and Khan, F. (2020). Inhibition of NFAT Signaling Restores Microvascular Endothelial Function in Diabetic Mice. Diabetes *69*, 424–435. .

Goettsch, C., Rauner, M., Hamann, C., Sinningen, K., Hempel, U., Bornstein, S.R., and Hofbauer, L.C. (2011). Nuclear factor of activated T cells mediates oxidised LDL-induced calcification of vascular smooth muscle cells. Diabetologia *54*, 2690–2701. .

Gordon, M.G., Inoue, F., Martin, B., Schubach, M., Agarwal, V., Whalen, S., Feng, S., Zhao, J., Ashuach, T., Ziffra, R., et al. (2020). lentiMPRA and MPRAflow for high-throughput functional characterization of gene regulatory elements. Nat. Protoc. *15*, 2387–2412. .

Gupta, R.M., Hadaya, J., Trehan, A., Zekavat, S.M., Roselli, C., Klarin, D., Emdin, C.A., Hilvering, C.R.E., Bianchi, V., Mueller, C., et al. (2017). A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression. Cell *170*, 522–533.e15. .

Hadi, H.A.R., Carr, C.S., and Al Suwaidi, J. (2005). Endothelial dysfunction: cardiovascular risk factors, therapy, and outcome. Vasc. Health Risk Manag. *1*, 183–198. .

Hajar, R. (2017). Risk Factors for Coronary Artery Disease: Historical Perspectives. Heart Views *18*, 109–114. .

van der Harst, P., and Verweij, N. (2018). Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. Circ. Res. *122*, 433–443. .

Hartmann, K., Seweryn, M., and Sadee, W. (2022). Interpreting coronary artery disease GWAS results: A functional genomics approach assessing biological significance. PLoS One *17*, e0244904. .

Hirotsu, Y., Hataya, N., Katsuoka, F., and Yamamoto, M. (2012). NF-E2-related factor 1 (Nrf1) serves as a novel regulator of hepatic lipid metabolism through regulation of the Lipin1 and PGC-1β genes. Mol. Cell. Biol. *32*, 2760–2770. .

Horton, J.D., Shah, N.A., Warrington, J.A., Anderson, N.N., Park, S.W., Brown, M.S., and Goldstein, J.L. (2003). Combined analysis of oligonucleotide microarray data from transgenic and knockout mice identifies direct SREBP target genes. Proc. Natl. Acad. Sci. U. S. A. *100*, 12027–12032. .

Hu, W., Lu, H., Zhang, J., Fan, Y., Chang, Z., Liang, W., Wang, H., Zhu, T., Garcia-Barrio, M.T., Peng, D., et al. (2018). Krüppel-like factor 14, a coronary artery disease associated transcription factor, inhibits endothelial inflammation via NF-κB signaling pathway. Atherosclerosis *278*, 39–48. .

Huang, Q.Q., Tang, H.H.F., Teo, S.M., Mok, D., Ritchie, S.C., Nath, A.P., Brozynska, M., Salim, A., Bakshi, A., Holt, B.J., et al. (2020). Neonatal genetics of gene expression reveal potential origins of autoimmune and allergic disease risk. Nat. Commun. *11*, 3761. .

Huss, J.M., and Kelly, D.P. (2004). Nuclear receptor signaling and cardiac energetics. Circ. Res. *95*, 568–578. .

Innocenti, F., Cooper, G.M., Stanaway, I.B., Gamazon, E.R., Smith, J.D., Mirkov, S., Ramirez, J., Liu, W., Lin, Y.S., Moloney, C., et al. (2011). Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. PLoS Genet. *7*, e1002078. .

Johnson, E.N., Lee, Y.M., Sander, T.L., Rabkin, E., Schoen, F.J., Kaushal, S., and Bischoff, J. (2003). NFATc1 mediates vascular endothelial growth factor-induced proliferation of human pulmonary valve endothelial cells. J. Biol. Chem. *278*, 1686–1692. .

Johnson, G.D., Barrera, A., McDowell, I.C., D'Ippolito, A.M., Majoros, W.H., Vockley, C.M., Wang, X., Allen, A.S., and Reddy, T.E. (2018). Human genome-wide measurement of drug-responsive regulatory activity. Nat. Commun. *9*, 5317. .

Kalita, C.A., Moyerbrailean, G.A., Brown, C., Wen, X., Luca, F., and Pique-Regi, R. (2017). QuASAR-MPRA: Accurate allele-specific analysis for massively parallel reporter assays. Bioinformatics https://doi.org/10.1093/bioinformatics/btx598.

Kalita, C.A., Brown, C.D., Freiman, A., Isherwood, J., Wen, X., Pique-Regi, R., and Luca, F. (2018). High throughput characterization of genetic effects on DNA:protein binding and gene transcription. Genome Res. https://doi.org/10.1101/gr.237354.118.

Kessler, T., and Schunkert, H. (2021). Coronary Artery Disease Genetics Enlightened by Genome-Wide Association Studies. JACC Basic Transl Sci *6*, 610–623. .

Kim-Hellmuth, S., Bechheim, M., Pütz, B., Mohammadi, P., Nédélec, 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. .

Kim-Hellmuth, S., Aguet, F., Oliva, M., Muñoz-Aguirre, M., Kasela, S., Wucher, V., Castel, S.E., Hamel, A.R., Viñuela, A., Roberts, A.L., et al. (2020). Cell type-specific genetic regulation of gene expression across human tissues. Science *369*. https://doi.org/10.1126/science.aaz8528.

Klee, C.B., Ren, H., and Wang, X. (1998). Regulation of the calmodulin-stimulated protein phosphatase, calcineurin. J. Biol. Chem. *273*, 13367–13370. .

Knowles, D.A., Davis, J.R., Edgington, H., Raj, A., Favé, M.-J., Zhu, X., Potash, J.B., Weissman, M.M., Shi, J., Levinson, D.F., et al. (2017). Allele-specific expression reveals interactions between genetic variation and environment. Nat. Methods https://doi.org/10.1038/nmeth.4298.

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*. https://doi.org/10.7554/eLife.33480.

Koyama, S., Ito, K., Terao, C., Akiyama, M., Horikoshi, M., Momozawa, Y., Matsunaga, H., Ieki, H., Ozaki, K., Onouchi, Y., et al. (2020). Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. Nat. Genet. *52*, 1169–1177. .

Krüger-Genge, A., Blocki, A., Franke, R.-P., and Jung, F. (2019). Vascular Endothelial Cell Biology: An Update. Int. J. Mol. Sci. *20*. https://doi.org/10.3390/ijms20184411.

Lebeau, P.F., Byun, J.H., Platko, K., Saliba, P., Sguazzin, M., MacDonald, M.E., Paré, G., Steinberg, G.R., Janssen, L.J., Igdoura, S.A., et al. (2022). Caffeine blocks SREBP2-induced hepatic PCSK9 expression to enhance LDLR-mediated cholesterol clearance. Nat. Commun. *13*, 770. .

Lee, M.N., Ye, C., Villani, A.-C., Raj, T., Li, W., Eisenhaure, T.M., Imboywa, S.H., Chipendo, P.I., Ran, F.A., Slowikowski, K., et al. (2014). Common genetic variants modulate pathogen-sensing responses in human dendritic cells. Science *343*, 1246980. .

Li, H., Li, Q., Zhang, Y., Liu, W., Gu, B., Narumi, T., Siu, K.L., Youn, J.Y., Liu, P., Yang, X., et al. (2019). Novel Treatment of Hypertension by Specifically Targeting E2F for Restoration of Endothelial Dihydrofolate Reductase and eNOS Function Under Oxidative Stress. Hypertension *73*, 179–189. .

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. .

Mangravite, L.M., Engelhardt, B.E., Medina, M.W., Smith, J.D., Brown, C.D., Chasman, D.I., Mecham, B.H., Howie, B., Shim, H., Naidoo, D., et al. (2013). A statin-dependent QTL for GATM expression is associated with statin-induced myopathy. Nature *502*, 377–380. .

Manry, J., Nédélec, Y., Fava, V.M., Cobat, A., Orlova, M., Van Thuc, N., 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 Rienzo, A. (2011). Interactions between glucocorticoid treatment and Cis-regulatory polymorphisms contribute to cellular response phenotypes. PLoS Genet. *7*. https://doi.org/10.1371/journal.pgen.1002162.

Matsuzawa, Y., and Lerman, A. (2014). Endothelial dysfunction and coronary artery disease: assessment, prognosis, and treatment. Coron. Artery Dis. *25*, 713–724. .

Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C.G., Jr, Kinney, J.B., et al. (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat. Biotechnol. *30*, 271–

277. .

Miranda, A.M., Steluti, J., Goulart, A.C., Benseñor, I.M., Lotufo, P.A., and Marchioni, D.M. (2018). Coffee Consumption and Coronary Artery Calcium Score: Cross-Sectional Results of ELSA-Brasil (Brazilian Longitudinal Study of Adult Health). J. Am. Heart Assoc. *7*. https://doi.org/10.1161/JAHA.117.007155.

Moyerbrailean, G.A., Richards, A.L., Kurtz, D., Kalita, C.A., Davis, G.O., Harvey, C.T., Alazizi, A., Watza, D., Sorokin, Y., Hauff, N., et al. (2016a). High-throughput allele-specific expression across 250 environmental conditions. Genome Res. *26*, 1627–1638. .

Moyerbrailean, G.A., Kalita, C.A., Harvey, C.T., Wen, X., Luca, F., and Pique-Regi, R. (2016b). Which Genetics Variants in DNase-Seq Footprints Are More Likely to Alter Binding? PLoS Genet. *12*, e1005875. .

Nédélec, 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.e21. .

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. .

Patwardhan, R.P., Hiatt, J.B., Witten, D.M., Kim, M.J., Smith, R.P., May, D., Lee, C., Andrie, J.M., Lee, S.-I., Cooper, G.M., et al. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. Nat. Biotechnol. *30*, 265–270. .

Pickrell, J.K. (2014). Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Am. J. Hum. Genet. *94*, 559–573. .

Planning Committee for a Workshop on Potential Health Hazards Associated with Consumption of Caffeine in Food and Dietary Supplements, Food and Nutrition Board, Board on Health Sciences Policy, and Institute of Medicine (2014). Caffeine in Food and Dietary Supplements: Examining Safety: Workshop Summary (Washington (DC): National Academies Press (US)).

Prosdocimo, D.A., Anand, P., Liao, X., Zhu, H., Shelkay, S., Artero-Calderon, P., Zhang, L., Kirsh, J., Moore, D. 'vesharronne, Rosca, M.G., et al. (2014). Kruppel-like factor 15 is a critical regulator of cardiac lipid metabolism. J. Biol. Chem. *289*, 5914–5924. .

Prosdocimo, D.A., John, J.E., Zhang, L., Efraim, E.S., Zhang, R., Liao, X., and Jain, M.K. (2015). KLF15 and PPARα Cooperate to Regulate Cardiomyocyte Lipid Gene Expression and Oxidation. PPAR Res. *2015*, 201625. .

Quach, H., Rotival, M., Pothlichet, J., Loh, Y.-H.E., Dannemann, M., Zidane, N., Laval, G., Patin, E., Harmant, C., Lopez, M., et al. (2016). Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human Populations. Cell *167*, 643–656.e17. .

Shlyueva, D., Stelzer, C., Gerlach, D., Yáñez-Cuna, J.O., Rath, M., Boryń, Ł.M., Arnold, C.D., and Stark, A. (2014). Hormone-responsive enhancer-activity maps reveal predictive motifs, indirect repression, and targeting of closed chromatin. Mol. Cell *54*, 180–192. .

Stevens, L.M., Linstead, E., Hall, J.L., and Kao, D.P. (2021). Association Between Coffee Intake and Incident Heart Failure Risk: A Machine Learning Analysis of the FHS, the ARIC Study, and the CHS. Circ. Heart Fail. *14*, e006799. .

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. .

Temprano-Sagrera, G., Sitlani, C.M., Bone, W.P., Martin-Bornez, M., Voight, B.F., Morrison, A.C., Damrauer, S.M., de Vries, P.S., Smith, N.L., and Sabater-Lleal, M. (2022). Multi-phenotype analyses of hemostatic traits with cardiovascular events reveal novel genetic associations. J. Thromb. Haemost. *20*, 1331–1349. .

Tewhey, R., Kotliar, D., Park, D.S., Liu, B., Winnicki, S., Reilly, S.K., Andersen, K.G., Mikkelsen, T.S., Lander, E.S., Schaffner, S.F., et al. (2016). Direct Identification of Hundreds of Expression-Modulating Variants using a Multiplexed Reporter Assay. Cell *165*, 1519–1529. .

Turnbull, D., Rodricks, J.V., Mariano, G.F., and Chowdhury, F. (2017). Caffeine and cardiovascular health. Regul. Toxicol. Pharmacol. *89*, 165–185. .

Ulirsch, J.C., Nandakumar, S.K., Wang, L., Giani, F.C., Zhang, X., Rogov, P., Melnikov, A., McDonel, P., Do, R., Mikkelsen, T.S., et al. (2016). Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. Cell *165*, 1530–1545. .

Vockley, C.M., Guo, C., Majoros, W.H., Nodzenski, M., Scholtens, D.M., Hayes, M.G., Lowe, W.L., Jr, and Reddy, T.E. (2015). Massively parallel quantification of the regulatory effects of noncoding genetic variation in a human cohort. Genome Res. *25*, 1206–1214. .

Voskoboinik, A., Koh, Y., and Kistler, P.M. (2019). Cardiovascular effects of caffeinated beverages. Trends Cardiovasc. Med. *29*, 345–350. .

Wang, X., and Musunuru, K. (2018). Confirmation of Causal rs9349379- PHACTR1 Expression Quantitative Trait Locus in Human-Induced Pluripotent Stem Cell Endothelial Cells. Circ Genom Precis Med *11*, e002327. .

Wang, X., He, L., Goggin, S.M., Saadat, A., Wang, L., Sinnott-Armstrong, N., Claussnitzer, M., and Kellis, M. (2018). High-resolution genome-wide functional dissection of transcriptional regulatory regions and nucleotides in human. Nat. Commun. *9*, 5380. .

Wang, Y., Hu, J., Liu, J., Geng, Z., Tao, Y., Zheng, F., Wang, Y., Fu, S., Wang, W., Xie, C., et al. (2020). The role of Ca2+/NFAT in Dysfunction and Inflammation of Human Coronary Endothelial Cells induced by Sera from patients with Kawasaki disease. Sci. Rep. *10*, 4706. .

Wen, X. (2016). Molecular QTL discovery incorporating genomic annotations using Bayesian

false discovery rate control. Aoas *10*, 1619–1638. .

Wen, X., Luca, F., and Pique-Regi, R. (2015). Cross-population joint analysis of eQTLs: fine mapping and functional annotation. PLoS Genet. *11*, e1005176. .

Xu, S., Ilyas, I., Little, P.J., Li, H., Kamato, D., Zheng, X., Luo, S., Li, Z., Liu, P., Han, J., et al. (2021). Endothelial Dysfunction in Atherosclerotic Cardiovascular Diseases and Beyond: From Mechanism to Pharmacotherapies. Pharmacol. Rev. *73*, 924–967. .

Zetterqvist, A.V., Berglund, L.M., Blanco, F., Garcia-Vaz, E., Wigren, M., Dunér, P., Andersson, A.-M.D., To, F., Spegel, P., Nilsson, J., et al. (2014). Inhibition of nuclear factor of activated T-cells (NFAT) suppresses accelerated atherosclerosis in diabetic mice. PLoS One *8*, e65020. .

Zhang, Y., Quick, C., Yu, K., Barbeira, A., GTEx Consortium, Luca, F., Pique-Regi, R., Kyung Im, H., and Wen, X. (2020). PTWAS: investigating tissue-relevant causal molecular mechanisms of complex traits using probabilistic TWAS analysis. Genome Biol. *21*, 232. .