UCSF
UC San Francisco Previously Published Works

Title
Analysis of putative cis-regulatory elements regulating blood pressure variation

Permalink
https://escholarship.org/uc/item/8s21n567

Journal
Human Molecular Genetics, 29(11)

ISSN
0964-6906

Authors
Nandakumar, Priyanka
Lee, Dongwon
Hoffmann, Thomas J
et al.

Publication Date
2020-07-21

DOI
10.1093/hmg/ddaa098

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

OXFORD

Human Molecular Genetics, 2020, Vol. 29, No. 11 1922–1932

doi: 10.1093/hmg/ddaa098
Advance Access Publication Date: 21 May 2020
Association Studies Article

ASSOCIATION STUDIES ARTICLE

Analysis of putative cis-regulatory elements regulating blood pressure variation

Priyanka Nandakumar[1], Dongwon Lee[1,2,3], Thomas J. Hoffmann[4,5], Georg B. Ehret[1,2,6], Dan Arking[1], Dilrini Ranatunga[7], Man Li[8], Megan L. Grove[9], Eric Boerwinkle[9], Catherine Schaefer[7], Pui-Yan Kwok[5], Carlos Iribarren[7], Neil Risch[4,5,7], Aravinda Chakravarti[1,2,*]

[1]Department of Genetic Medicine, McKusick-Nathans Institute, Baltimore, MD 21205, USA, [2]Center for Human Genetics and Genomics, NYU School of Medicine, New York, NY 10016, USA, [3]Division of Nephrology, Boston Children's Hospital, Boston, MA 02115, USA, [4]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA 94158, USA, [5]Institute for Human Genetics, University of California San Francisco, San Francisco, CA 94143, USA, [6]Cardiology, Department of Specialties of Internal Medicine, University of Geneva, Geneva 1211, Switzerland, [7]Kaiser Permanente Northern California Division of Research, Oakland, California 94612 USA, [8]Division of Nephrology, Department of Human Genetics, University of Utah, Salt Lake City, Utah 84132, USA and [9]Human Genetics Center, University of Texas Health Science Center, Houston, Texas 77030, USA

*To whom correspondence should be addressed. Tel: +(212) 2638029; Fax: +(646) 5014526; Email: aravinda.chakravarti@nyulangone.org

Abstract

Hundreds of loci have been associated with blood pressure (BP) traits from many genome-wide association studies. We identified an enrichment of these loci in aorta and tibial artery expression quantitative trait loci in our previous work in ∼100 000 Genetic Epidemiology Research on Aging study participants. In the present study, we sought to fine-map known loci and identify novel genes by determining putative regulatory regions for these and other tissues relevant to BP. We constructed maps of putative cis-regulatory elements (CREs) using publicly available open chromatin data for the heart, aorta and tibial arteries, and multiple kidney cell types. Variants within these regions may be evaluated quantitatively for their tissue- or cell-type-specific regulatory impact using deltaSVM functional scores, as described in our previous work. We aggregate variants within these putative CREs within 50 Kb of the start or end of 'expressed' genes in these tissues or cell types using public expression data and use deltaSVM scores as weights in the group-wise sequence kernel association test to identify candidates. We test for association with both BP traits and expression within these tissues or cell types of interest and identify the candidates *MTHFR*, *C10orf32*, *CSK*, *NOV*, *ULK4*, *SDCCAG8*, *SCAMP5*, *RPP25*, *HDGFRP3*, *VPS37B* and *PPCDC*. Additionally, we examined two known QT interval genes, *SCN5A* and *NOS1AP*, in the Atherosclerosis Risk in Communities Study, as a positive control, and observed the expected heart-specific effect. Thus, our method identifies variants and genes for further functional testing using tissue- or cell-type-specific putative regulatory information.

Received: December 23, 2019. **Revised:** March 29, 2020. **Accepted:** May 6, 2020

© The Author(s) 2020. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

Introduction

Genetic studies of complex disorders have identified hundreds to thousands of variants in the human non-coding genome. However, despite significant mapping progress, we do not yet know the identity of most of the underlying genes and variants, nor have a mechanistic understanding of how these genes, individually and together, contribute to a phenotype. Thus, we need to consider how such genomic studies can improve our knowledge of trait physiology. One approach would be to focus genetic analyses on organs and tissues of interest.

Pritchard and colleagues have hypothesized that the majority of genome-wide association study (GWAS) signals may be functionally spurious and arise from genes' peripheral to the core functions affected in a trait or disease (1). These false positives dominate because most genes in a cell type are connected by gene expression to one another through very shallow functional networks, a working hypothesis that fails to explain the stability of network perturbations (robustness) or their specificity (phenotypic effects) (2–4). To resolve this question, connecting genotypes to phenotypes through gene expression variation is of primary importance since expression quantitative trait loci (eQTL) are identifiable causal factors (5,6). However, utilizing gene expression in trait-related tissues is necessary (7), as genes exert their activities in the context of a core genetic network with intrinsic (cell autonomous) and extrinsic (non-autonomous) feedback (8).

Transcription within mammalian genomes is locally regulated within chromatin segments called topological associating domains (TADs), largely invariant across cell types and of variable sizes, but generally 200–1000 kilobases (kb), depending on callers (9,10). TADs contain numerous dispersed spatiotemporal expression cis-regulatory elements (CREs or enhancers) that allow binding of various transcription factors (TF) to enable gene expression control (11). Many enhancers are recognized by their DNaseI hypersensitivity (DHS), Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) assays (12) or adjacent histone (H3K4me1, H3K4me3, H3K27ac) modifications (13,14). Their phenotypic importance is evident from the fact that only 2.6% of the genome comprises DHS and histone marks (15) but explains ∼30% of the heritability of traits (16). Thus, trait variation is from sequence changes within TFs, their binding sites (TFBS) and CREs, all detectable through epigenomic marks in cell lines and tissues. In this study, we propose an approach wherein these types of epigenomic data are used to identify genes within a GWAS locus in tissues of interest.

The analyses we propose are enabled by numerous public genomic resources. The Encyclopedia of DNA Elements (ENCODE) Project (https://www.encodeproject.org/) has generated open chromatin, RNA and DNA sequencing and histone modification data, among other data types. The Genotype-Tissue Expression (GTEx) Project (https://www.gtexportal.org/) includes genotype and expression data across 53 tissues and is useful as a reference transcriptome and eQTL dataset. These public resources also enable the development of an annotation score, deltaSVM (17), in which the quantitative impact of a non-coding variant on tissue- or cell-type-specific gene regulation is predicted, based on a reference training set of regulatory regions. In this study, we exemplify this reverse genetic approach by focusing on blood pressure (BP) and QT interval variation.

Although the roles of the kidney and adrenal gland are well established in BP regulation and syndromes (18–20), our previous work in the Kaiser Permanente Research Program on Genes, Environment and Health (RPGEH) Genetic Epidemiology Research on Adult Health and Aging (GERA) (21,22) study demonstrated that associated variants at BP GWAS loci were enriched in eQTLs specific to the aorta and tibial arteries. Expanding on this work in this study, we aimed to connect groups of proximal putative regulatory variants within and around each gene to both the gene's expression and also to BP traits, inferring that the gene's expression in a potentially relevant tissue affected the regulation of BP. To accomplish this, we undertake this study with three main steps: (1) establish a putative CRE catalog from open chromatin experiments in tissues/cell types of interest, (2) score the predicted functional regulatory impact of the variants within these putative CREs ('putative CRE variants') and (3) take the putative CRE variants within a window around each gene and test them in aggregate using a gene-based test (Fig. 1). More specifically, within each artery dataset, we identified putative CREs and, by extension, putative CRE variants, for every gene, and tested these variants in aggregate for association with BP in the GERA study, as well as with expression in the GTEx study. We used the sequence kernel association test (SKAT) (23) for these association analyses, with each variant weighted by their deltaSVM score, to up-weight variants with greater predicted effects on gene regulatory activity. We supplemented our expression analyses with the software MetaXcan (24) to test whether the predicted expression of genes in each individual could be associated with BP. Prior to the novel BP gene discovery analyses of tissue involvement, and as a positive control, we first examined genes for the cardiac trait QT interval for which there is strong functional evidence of primarily heart involvement, using data from the Atherosclerosis Risk in Communities (ARIC) (25,26) study. Finally, we examined the effects of putative regulatory variation for monogenic BP syndrome genes, all known to be renal or adrenal disorders, in four available kidney cell types to test for a group effect on BP.

Our results demonstrate the feasibility of identifying BP genes by tissue, which we expect will facilitate more comprehensive functional analyses of BP genes and BP control mechanisms.

Results

We conducted several tissue-specific analyses to identify tissues and genes of interest for BP regulation using the GERA study; an overview of our analysis scheme is shown in Figure 1. We initially focused on identifying tissues relevant to BP GWAS loci and subsequently expanded on this by using tissue-specific information to analyze putative CRE variation of genes in these tissues. The aim was to identify specific genes and variants of interest at these GWAS loci. We also studied putative regulatory variation at 20 monogenic syndromic hypertension and hypotension genes in several kidney cell types. To begin, our study also includes an analysis of QT interval as a positive control to demonstrate the identification of well-characterized genes for that trait.

Constructing CRE maps

Our previous eQTL and subsequent partitioned heritability analyses (see Supplementary Material, Methods and Results and Tables S1 and S2) indicated that regulatory elements in specific tissues are of interest in BP studies. With knowledge of tissues highly relevant to characterizing BP GWAS loci, our next

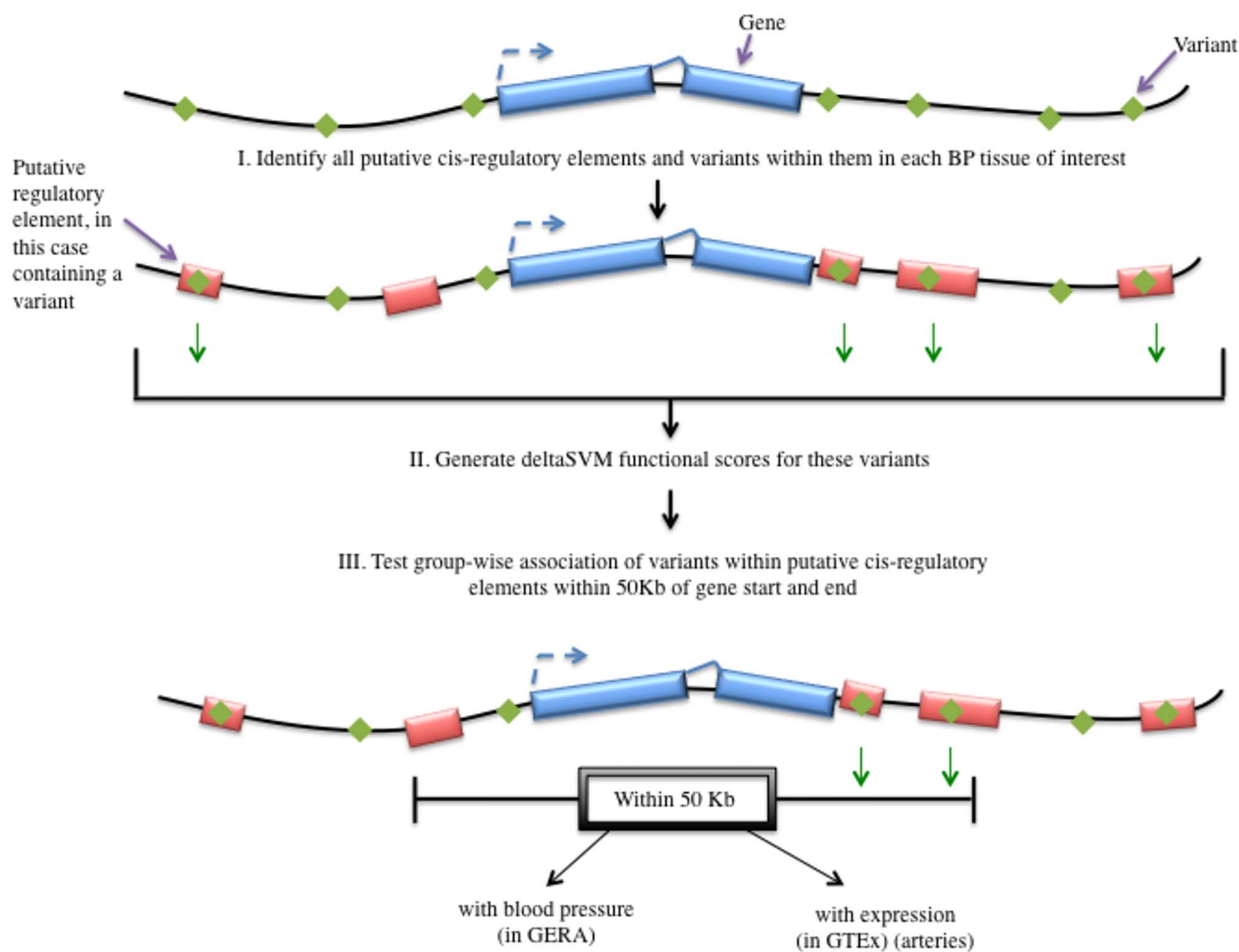

Figure 1. Overview of putative CRE identification and SKAT analysis.

aim was to test each gene's putative cis-regulatory variation for association with both gene expression and BP, in a tissue-specific context. This is expected to assist in identifying novel genes of interest, as well as provide tissue- or cell-type-specific information about known genes.

We first constructed CRE maps or a catalog of experimentally determined open chromatin, for the aorta and tibial arteries, as well as four kidney cell types (renal cortical epithelial cell, glomerular endothelial cell, epithelial cell of proximal tubule and glomerular visceral epithelial cell), because of the known involvement of the kidney in BP regulation (18,19), using ENCODE data (Supplementary Material, Table S3; Fig. 1, step I) (though many monogenic forms of BP disorders occur due to an effect of the adrenal gland on renal function (20)). These CRE maps were completed as an extension of the construction of our recent cardiac CRE map (27). We specifically focused on identifying putative enhancers for the aorta and tibial arteries (see Methods). We subsequently used these maps for training with the software gkm-SVM (28,29) in order to generate deltaSVM functional scores for all non-coding variants from the 1000 Genomes European ancestry sample, to be tested for association on a gene-level basis (Fig. 1, step II). The performance for each model is available in Supplementary Material, Table S4 (AUC range, 0.84–0.96), with the best performance in the renal cell types. A possible reason for the improved performance of the renal cell types is that the data were from individual cell types as opposed to a mixture of cell types comprising the arteries. The magnitude of the deltaSVM score for a variant reflects its predicted impact on regulatory functional activity, while its sign reflects the prediction with respect to the reference allele. Therefore, to represent the predicted impact of each variant irrespective of allele, we show the distributions of the absolute values of the deltaSVM scores for the arteries and kidney cell types in Supplementary Material, Figure S1.

Tissue-specific gene identification

As our emphasis in this section is to connect a gene's putative CRE variants to both a phenotype of interest and to its expression in relevant tissues, we first describe the overall analysis scheme as applied to a general phenotype of interest (Fig. 1, step III). We then describe how we applied these analyses, first to the QT interval in the ARIC study, as proof of principle to demonstrate the utility of these analyses, and then to our BP traits of interest in the GERA study.

We defined gene's 'cis'-regulatory variants in this analysis as those variants falling in putative CREs within 50 kb of the gene's start and end. We tested their aggregate effect for each gene using SKAT (23), for association tests with the phenotype(s) of interest in the relevant population, SBP and DBP in

the GERA study and QT interval in the ARIC study. SKAT is a test that has generally been used to study groups of variants together and is useful when variants can have bidirectional effects; rare variants are more highly weighted than common variants by default. In addition to the default weights, we ran the analysis using equal weights for all variants. We finally used the tissue- or cell-type-specific deltaSVM scores for the analyzed variants as weights for a customized SKAT test, the score scaling with the effect of the variant on functional regulatory activity.

We then tested these groupings with expression data from GTEx v6p in the tissues of interest to link variants in the genes of interest to their gene expression. The groupings tested in the GTEx data with expression were not always identical to the groupings tested in the GERA or ARIC studies because of differences in imputation quality score filtering, missingness of genotypes from genotype probabilities to hard call conversion and variants present in the reference populations studied. However, this analysis still connects a given gene to its expression and to the phenotype via a highly overlapping set of CRE variants and was completed this way to test the most complete set of variants available meeting our criteria. In addition to testing putative regulatory variants with gene expression in GTEx, we used MetaXcan (24) to augment SKAT to identify any new associations by this method. This software predicts the association of gene expression with a phenotype, given genotypes for the population of interest based on training from reference genotypes and expression data.

Analysis of CREs in QT interval

As mentioned earlier, we considered the cardiac trait QT interval first to demonstrate proof of concept for tissue-specific gene identification. The QT interval is the time in ms between the onset of the Q wave and the end of the T wave in the surface 12-lead electrocardiogram (30), which has ∼30% heritability (31–34). In our recent work, we have demonstrated that a significant proportion of the heritability is explained by predicted cardiac regulatory variants (27). We analyzed the genes at previously published QT interval GWAS loci to determine whether or not a heart-specific effect could be observed. Two of the genes with major effects in a GWAS and functionally validated in QT interval heritability are *NOS1AP* (34–36) and *SCN5A* (36,37). The full results are presented in the Supplementary Material, Text S1 results, Table S5 and Figures S2 and S3; to summarize here briefly, we aimed to discover if a heart-specific effect could be revealed for each of these two genes. We observed a heart specificity for *SCN5A*; *NOS1AP* showed signal across all the cell types in the equal-weighted analyses, though considerably attenuated in some of the deltaSVM-weighted non-heart tissues. Considering both sets of effects, certainly variants with detectable signals present in open chromatin regions specific to the relevant tissue/cell types will allow the detection of a tissue-specific signal, as for *SCN5A*. It also appears, however, that gene-level signals may be captured by analyses in which all variants are weighted equally and when local open chromatin boundaries across tissues/cell types overlap considerably, especially when variants with strong signals are present within these shared regions. In this situation, we will not necessarily be able to differentiate between different tissue/cell types. Weighting with the tissue-specific deltaSVM scores introduces an additional tier of tissue specificity and is based on global open chromatin differences and is also not expected to be impacted by linkage disequilibrium (LD) in the ways that the other two weighting schemes are, as the generation of the scores are only dependent on sequence context. Finally, using the default weights shows least concordance with the other two sets of results, indicating that for this analysis, rare variants are not driving the signal as compared to common variants. This is as expected, as we prioritized noncoding variation for these analyses, and the rare variants with larger effects expected to make a detectable contribution are more likely to be in the exome.

Analysis of CREs at GWAS loci for BP regulation

We then applied these analyses to the tissues of interest for BP regulation, namely, aorta, tibial artery and four kidney cell types, in a subset of 71 404 unrelated GERA EUR individuals. We tested 14 548 genes expressed at reads per kilobase of transcript, per million mapped reads (RPKM) ≥0.3 in 197 aorta GTEx samples and 13 963 genes expressed at RPKM ≥0.3 in 285 tibial artery GTEx samples for the SKAT analyses. We used summary statistics available from 80 792 individuals (38) to maximize the sample size for which the MetaXcan analyses were run, for the aorta and tibial arteries. Results for each of the arteries are presented in Tables 1 and 2 and, more completely, in Supplementary Material, Tables S6–S9. In some cases, shared variants drive the positive signal for multiple genes at the same locus; expression in the relevant tissue or cell type may pinpoint a specific gene. However, it may be noted that the genes *CERS5*, *COX14* and *RP4-605O3.4* are all present at the same locus in the arteries (Tables 1 and 2), but evidence of expression association is present for many of these genes; this may be indicative of proximal variants affecting different genes or pleiotropy of single variants affecting expression of multiple genes.

On the whole, the 25 genes reported here across aorta and tibial artery genes have been identified at previous BP loci (39). While there are several genes in each analysis with interesting associations with BP traits, here we only highlight the genes that have statistical significance of $P < 1 \times 10^{-4}$ for both expression and BP in the aorta analyses (Tables 1 and 2). In our previous work, the aorta was demonstrated as the greatest outlier in an analysis of eQTL enrichment among GTEx tissues for BP traits (38). These genes include *MTHFR* (40,41) (SBP), *C10orf32* (40) (SBP), *CSK* (SBP), *NOV* (42) (DBP), *ULK4* (43) (DBP), *SDCCAG8* (DBP), *SCAMP5* (DBP), *RPP25* (DBP), *HDGFRP3* (DBP), *VPS37B* (DBP) and *PPCDC* (DBP). Most of these genes are present at or near previously replicated BP GWAS loci; *SDCCAG8* was identified as part of Hoffmann *et al.* (38). It is noteworthy that both *SCAMP5* and *PPCDC* are neighboring genes but have independent expression support in the same tissue.

Analysis of CREs for monogenic BP genes

We also studied the genes involved in monogenic forms of hypotension or hypertension in four kidney cell types available from the ENCODE project (see earlier). As the expression data available for kidney are insufficient, we studied each cell type individually and carried out only SKAT analyses for these genes; the results are in Supplementary Material, Tables S10 and S11. The most notable result is that of *CYP17A1* (Supplementary Material, Table 3), which shows an effect ($P \sim 10^{-5}$–10^{-7}) across all four cell types in the unweighted variants analyses for SBP only, and more specifically, only in the glomerular endothelial cell (p.SKAT.dsvm = 3.64×10^{-8}) in the deltaSVM-weighted results. However, as *C10orf32* is a gene of interest at the same

Table 1. Aorta and tibial arteries SBP SKAT and MetaXcan results

Tissue	Gene	p.sbp.dsvm	p.sbp.def	p.GTEx.dsvm	p.GTEx.def	p.MetX	Previous
Aorta	*NR3C1*	5.44×10^{-6}	0.064	0.002	0.529	–	FALSE
	WBP1L	6.66×10^{-6}	0.085	2.72×10^{-4}	0.137	1.35×10^{-5}	FALSE
	SBF2	9.27×10^{-6}	2.58×10^{-5}	0.003	0.022	2.01×10^{-5}	TRUE
	CLCN6	1.07×10^{-5}	2.32×10^{-6}	0.014	2.63×10^{-4}	0.028	TRUE
	MTHFR	1.08×10^{-5}	4.22×10^{-6}	1.85×10^{-5}	0.015	0.001	TRUE
	C10orf32	1.68×10^{-5}	0.026	4.63×10^{-15}	0.059	–	TRUE
	RP4-605O3.4	4.51×10^{-5}	0.069	0.016	0.425	–	FALSE
	COX14	4.51×10^{-5}	0.069	7.64×10^{-4}	0.338	0.034	TRUE
	CSK	6.34×10^{-5}	0.110	3.18×10^{-5}	0.579	1.24×10^{-4}	TRUE
	ULK3	6.34×10^{-5}	0.952	0.001	0.164	9.14×10^{-4}	TRUE
Tibial artery	*CLCN6*	9.65×10^{-9}	2.62×10^{-4}	5.07×10^{-8}	3.58×10^{-6}	2.67×10^{-9}	TRUE
	MTHFR	9.65×10^{-9}	2.62×10^{-4}	8.79×10^{-8}	0.161	0.076	TRUE
	C10orf32	6.98×10^{-8}	0.004	3.07×10^{-14}	7.30×10^{-8}	–	TRUE
	HOXC-AS1	4.00×10^{-5}	0.638	3.24×10^{-6}	0.194	–	FALSE
	CCDC6	4.49×10^{-5}	0.151	3.32×10^{-4}	0.035	1.08×10^{-4}	FALSE
	ATE1	5.67×10^{-5}	0.583	8.74×10^{-4}	0.086	0.816	FALSE
	SOX7	6.76×10^{-5}	0.036	0.007	0.121	–	FALSE
	AGT	8.29×10^{-5}	0.090	0.001	6.97×10^{-6}	0.937	TRUE
	NT5C2	8.75×10^{-5}	1.44×10^{-4}	0.004	0.163	0.006	TRUE
	DHX33	9.67×10^{-5}	0.372	3.67×10^{-7}	0.033	0.277	FALSE
	SFMBT1	1.15×10^{-4}	0.521	0.003	0.363	0.006	FALSE
	NPPA	1.16×10^{-4}	0.003	0.007	0.025	–	TRUE
	ERI1	1.33×10^{-4}	0.280	0.001	0.185	–	FALSE
	BCL2L2	1.59×10^{-4}	0.009	5.20×10^{-4}	0.356	–	FALSE
	BCL2L2-PABPN1	1.59×10^{-4}	0.009	0.013	0.717	–	FALSE
	NPPA-AS1	1.66×10^{-4}	0.003	1.89×10^{-23}	4.67×10^{-10}	–	FALSE
	C1orf132	1.74×10^{-4}	0.055	0.009	0.400	–	FALSE
	RPAIN	1.89×10^{-4}	0.627	0.003	0.479	0.546	FALSE
	CTC-524C5.2	1.89×10^{-4}	0.627	0.000162827	0.301	–	FALSE

Tissue, artery tissue analyzed; p.sbp.*, *P*-values from SKAT analysis of BP in GERA with deltaSVM (dsvm) or default (def) weights; p.GTex.*, *P*-values from SKAT analysis of expression in GTEx with deltaSVM (dsvm) or default (def) weights; p.MetX, MetaXcan *P*-value; previous, TRUE if found at a previously identified GWAS locus from the UKB list, FALSE otherwise.

locus, based on the artery results above, we examined and noted that the results are somewhat similar for this gene, although not as striking, due to variant set sharing in the SKAT analyses for these genes; the breakdown of individual variants analyzed for these two genes is in Supplementary Material, Table S12. The variant rs3824754, with an SBP association $P = 1.40 \times 10^{-11}$, appears in the groupings of both genes for all four cell types, but has the highest deltaSVM magnitude in the endothelial cell. Additionally, there is a set of four variants with SBP association ($P < 1 \times 10^{-4}$; rs284853, rs284854, rs284855, rs284856), which only appear in the glomerular endothelial cell type. We observed that while *CYP17A1* was similarly associated with, or demonstrated evidence of association with, SBP in the deltaSVM and unweighted variants analysis for the aorta and tibial arteries (Table 1), the analysis of variants in GTEx for the same tissues did not reflect any significant association ($P > 0.01$). In contrast, *C10orf32* demonstrated significant association with SBP and with expression in GTEx for the aorta and tibial arteries (Table 1). The same four variants unique to the glomerular endothelial cell type above with strong associations with SBP are also present in the artery groupings. Three of these variants (rs284854, rs284855, rs284856) are eQTLs for *C10orf32* in the aorta and tibial arteries; these variants, however, do not show association with *CYP17A1* expression in these tissues (all $P > 0.03$ for aorta, all $P > 0.21$ for tibial artery, from eQTL data available from the GTEx portal (https://www.gtexportal.org/), accessed 09/08/17). Additionally, as the *CYP17A1* gene primarily demonstrates an adrenal effect in the monogenic disorder (44), we also examined the associations of these three variants in the GTEx portal with adrenal gland expression data for both genes; all have $P > 0.26$ for *CYP17A1* and $P > 0.04$ for *C10orf32*. This may reflect an endothelial-cell-specific effect for *C10orf32* rather than a tissue-type effect, especially as this locus has been identified in several previous BP GWAS studies (40,43,45–47); it may also not be very informative for the kidney, though suitable expression data for kidney would be required to assess this (Table 3).

Quantile–quantile (QQ) plots are shown in Supplementary Material, Figures S4–S15 for each of the six tissues and two BP phenotypes and each of three delta SVM weighting schemes. Although the deltaSVM weighting scheme demonstrates a greater enrichment of genes than the default weighting scheme, the equal weighting scheme marginally presents the greatest enrichment. In many cases, deltaSVM discriminates between different tissue/cell types while equal-weighted results do not; this is especially clear with the QT interval results.

We analyzed the union of genes that met significance in the association analysis with BP, regardless of association of expression, to maximize our gene list for annotation, using DAVID 6.8 (48,49). The results are shown in Supplementary Material, Table S13.

Table 2. Aorta and tibial arteries DBP SKAT and MetaXcan results

Tissue	Gene	p.dbp.dsvm	p.dbp.def	p.GTEx.dsvm	p.GTEx.def	p.MetX	Previous
Aorta	*NOV*	1.40×10^{-8}	0.664	4.26×10^{-5}	1	0.102	TRUE
	ULK4	1.75×10^{-8}	0.001	2.49×10^{-22}	0.094	2.95×10^{-10}	TRUE
	COX14	1.35×10^{-7}	0.138	7.64×10^{-4}	0.338	0.003	TRUE
	IGFBP3	3.41×10^{-7}	0.006	0.011	0.842	–	FALSE
	SDCCAG8	6.92×10^{-7}	0.359	5.57×10^{-6}	0.481	2.95×10^{-9}	TRUE
	CEP170	7.06×10^{-7}	8.44×10^{-7}	2.52×10^{-4}	3.59×10^{-4}	0.002	TRUE
	CSK	9.93×10^{-7}	0.068	3.18×10^{-5}	0.579	7.93×10^{-5}	TRUE
	ULK3	9.93×10^{-7}	0.826	0.001	0.164	2.09×10^{-4}	TRUE
	SCAMP5	3.98×10^{-6}	0.009	6.70×10^{-8}	4.15×10^{-10}	1.93×10^{-5}	FALSE
	RPP25	4.67×10^{-6}	0.928	3.58×10^{-17}	4.79×10^{-9}	–	FALSE[a]
	HDGFRP3	1.52×10^{-5}	0.578	3.90×10^{-7}	0.042	0.001	TRUE
	COX4I2	2.12×10^{-5}	0.003	0.002	0.729	–	FALSE
	SBF2	2.27×10^{-5}	0.004	0.003	0.022	7.41×10^{-4}	TRUE
	RNF40	2.39×10^{-5}	0.242	2.74×10^{-4}	0.338	–	TRUE
	RP11-382A20.2	4.82×10^{-5}	0.627	0.005	0.066	–	FALSE
	RNASEH2C	7.50×10^{-5}	7.50×10^{-5}	0.002	0.002	0.690	TRUE
	SLC25A37	1.16×10^{-4}	0.121	4.01×10^{-4}	0.537	0.169	FALSE
	SENP2	1.36×10^{-4}	0.711	0.004	0.202	0.050	TRUE
	VPS37B	1.46×10^{-4}	7.73×10^{-5}	8.50×10^{-6}	2.03×10^{-4}	3.43×10^{-5}	FALSE
	ZNF652	1.69×10^{-4}	1.69×10^{-4}	8.73×10^{-4}	8.73×10^{-4}	1.53×10^{-5}	TRUE
	NR3C1	2.21×10^{-4}	0.018	0.002	0.529	–	FALSE
	PPCDC	2.25×10^{-4}	0.009	6.68×10^{-8}	4.82×10^{-8}	0.002	FALSE
Tibial artery	*NOV*	1.70×10^{-8}	0.731	7.00×10^{-6}	0.943	1.44×10^{-6}	TRUE
	CERS5	1.66×10^{-6}	0.004	3.41×10^{-4}	3.55×10^{-4}	1.17×10^{-4}	TRUE
	COX14	1.66×10^{-6}	0.004	4.26×10^{-5}	6.54×10^{-4}	0.004	TRUE
	RP4-605O3.4	1.66×10^{-6}	0.004	1.57×10^{-7}	2.36×10^{-7}	–	FALSE
	JAG1	4.59×10^{-6}	0.032	0.018	0.304	–	TRUE
	ULK4	5.69×10^{-6}	0.008	2.78×10^{-11}	0.158	5.40×10^{-12}	TRUE
	IPO9	6.08×10^{-6}	0.416	0.009	8.59×10^{-4}	0.952	FALSE
	LIMA1	1.86×10^{-5}	0.638	0.008	0.926	0.120	TRUE
	NAV1	3.83×10^{-5}	0.554	2.64×10^{-4}	0.010	–	FALSE
	COX4I2	3.92×10^{-5}	0.360	1.94×10^{-4}	0.656	–	FALSE
	UBN1	4.23×10^{-5}	0.417	3.19×10^{-7}	0.299	1.51×10^{-5}	TRUE
	SCAMP5	5.29×10^{-5}	0.331	4.47×10^{-11}	5.09×10^{-5}	8.04×10^{-8}	FALSE
	RNASEH2C	5.88×10^{-5}	0.056	0.006	0.006	0.831	TRUE
	CEP120	6.36×10^{-5}	0.003	4.43×10^{-5}	0.123	1.63×10^{-5}	TRUE
	CLCN6	7.30×10^{-5}	0.001	5.07×10^{-8}	3.58×10^{-6}	1.16×10^{-5}	TRUE
	MTHFR	7.30×10^{-5}	0.001	8.79×10^{-8}	0.161	0.086	TRUE
	SDCCAG8	9.04×10^{-5}	0.130	4.55×10^{-5}	0.273	8.08×10^{-8}	TRUE
	ACSF3	1.05×10^{-4}	5.00×10^{-4}	0.003	0.548	0.328	FALSE
	RPP25	1.11×10^{-4}	0.559	4.36×10^{-12}	0.001	–	FALSE[a]
	COX5A	1.11×10^{-4}	0.559	0.004	0.053	–	TRUE
	MKL2	1.12×10^{-4}	0.031	0.017	0.020	–	FALSE
	VPS37B	1.40×10^{-4}	5.38×10^{-4}	2.08×10^{-15}	1.12×10^{-14}	0.007	FALSE
	FAM20B	1.91×10^{-4}	0.690	3.35×10^{-5}	0.337	0.030	FALSE
	PLA2G4B	2.10×10^{-4}	1	3.41×10^{-7}	0.320	0.005	FALSE
	ALDH2	2.15×10^{-4}	0.658	5.44×10^{-8}	1.26×10^{-6}	0.002	FALSE
	DUSP15	2.39×10^{-4}	0.008	0.018	0.923	0.028	FALSE
	RP11-65J21.3	2.40×10^{-4}	0.660	0.015	0.382	–	FALSE
	MAPKBP1	2.52×10^{-4}	0.077	7.04×10^{-5}	0.056	0.019	FALSE
	JMJD7	2.54×10^{-4}	0.077	1.47×10^{-9}	0.203	3.34×10^{-4}	FALSE
	CENPW	2.79×10^{-4}	0.075	0.004	0.066	0.002	FALSE
	ATF1	2.81×10^{-4}	0.639	4.80×10^{-19}	9.06×10^{-5}	2.13×10^{-5}	FALSE

Tissue, artery tissue analyzed; p.dbp.*, *P*-values from SKAT analysis of BP in GERA with deltasvm (dsvm) or default (def) weights; p.GTEx.*, *P*-values from SKAT analysis of expression in GTEx with deltaSVM (dsvm) or default (def) weights; p.MetX, MetaXcan *P*-value; previous, TRUE if found at a previously identified GWAS locus from the UKB list, FALSE otherwise

[a]*RPP25* has been previously identified in GWAS, but our method to identify genes at GWAS loci was conservative and missed this gene; see text

Discussion

Our previous genetic analyses identified the aorta and tibial arteries as relevant to BP regulation (38). In this study we have now identified several genes with regulatory variants linking significantly to both BP traits and to expression data in these tissues, most at previously replicated BP loci. Although the involvement of the kidney is well established in BP regulation through physiological evidence, we sought to identify genes at any of

Table 3. Kidney SBP and DBP SKAT results for *CYP17A1*

Experiment	Cell-type	SBP		DBP	
		p.bp.dsvm	p.bp.def	p.bp.dsvm	p.bp.def
ENCSR000EOK	Renal cortical epithelial cell	0.069	0.018	0.505	0.337
ENCSR000EOM	Glomerular endothelial cell	3.64×10^{-8}	0.013	0.053	0.187
ENCSR000EPW	Epithelial cell of proximal tubule	0.006	0.045	0.360	0.262
ENCSR785BDQ	Glomerular visceral epithelial cell	0.008	0.020	0.551	0.355

the hundreds of BP GWAS loci in a broader set of tissues. We examined groupings of multiple proximal and putatively causal regulatory variants defined around genes within a single tissue in order to identify specific genes of interest. We also examined QT interval genes at previous GWAS loci to highlight the identification of functionally characterized genes for this trait.

We identified several genes of potential interest to the aorta/arteries for BP, mostly at previously identified GWAS loci: *MTHFR*, *C10orf32*, *CSK*, *NOV*, *ULK4*, *SDCCAG8*, *SCAMP5*, *RPP25*, *HDGFRP3*, *VPS37B* and *PPCDC*. We note here that our method of identifying genes at previous loci was conservative: *RPP25* was not present in this list but is present just outside the TAD boundary used. In addition to its role in the progression of various cancers, the *NOV* gene has been identified as a player in angiogenesis (50,51) and vascular homeostasis (52). The *ULK4* gene has been previously associated with DBP (43), and variation in this gene has also been associated with aortic disease and acute aortic dissections (53). The association of a homozygous variant (C677T) in its neighboring gene, *MTHFR*, has long been associated with BP and vascular disease (54–57); more generally, this locus has been identified in large BP GWAS (40,41). The locus including *C10orf32* has been identified previously (40) and neighbors the well-studied *CYP17A1* gene. Though we initially examined only the latter among kidney cell types, because of its known role in monogenic hypertension, we note that both genes show BP association in endothelial contexts as well, but it is *C10orf32* that has strong expression support in the artery datasets in our study, while *CYP17A1* does not (58). The gene *SDCCAG8* is a centrosomal protein linked with nephronophthisis-related ciliopathies (OMIM: Senior–Loken syndrome-7, 613 615; Bardet–Biedl syndrome-16, 615 993, and Airik *et al*. (59)) and is expressed in the kidney and lung epithelia (59). The *CSK* gene, encoding a tyrosine kinase, is at a previous BP GWAS locus (60) and has been found to be associated with SBP in young children (61); there is also prior evidence through experiments in mouse aortas that this gene regulates BP through Src (62). Finally, the *SCAMP5* and *PPCDC* genes (within the same locus) (40), and *RPP25* (43), are previously identified BP genes.

As mentioned above, one major limitation in our study is the statistical power of the SKAT eQTL analysis, with small sample sizes available for each of the GTEx tissues. The power of implicating effects for a given tissue also depends on its total contribution and the numbers of eQTLs identified. The requirement in our study that a gene meet significance for both BP and expression therefore produced a more conservative list. However, the QT interval results, especially for the *SCN5A* gene, still illustrate the utility of this method. The availability of additional samples in the future will contribute to the success of this method in identifying genes of interest with greater statistical power. The gene annotation analyses revealed no clear BP-specific pathways or annotation, so these will also benefit from producing more specific and possibly larger gene sets. Additionally, we used hard genotype calls for analysis, necessitating some missing genotype data; the power of our methods could be improved by using imputed probabilities of genotypes.

Our attempts to expand findings beyond the known pathogenic coding variation with respect to the 20 genes involved in monogenic forms of hypertension or hypotension were inconclusive. We attribute this to the dearth of publicly available data for the kidney at this time and expect that the availability of more extensive data will resolve some of the issues in further studies. Additionally, though it is beyond the scope of this study, as the effects of many of these monogenic disorders are likely through the adrenal gland, a full analysis of adrenal gland data will be necessary to assess them.

The MetaXcan software has supported most of the genes highlighted here and identified novel associations, although there were some limitations with the availability of the models for all genes. Additionally, our results indicated that deltaSVM weighting might be validly discriminatory between cell types; this is most evident with several QT interval genes, such as *NOS1AP* and *SCN5A*. It is also suggestive of cell-type specificity with the results for *CYP17A1* in the kidney cell types. It may be informative moving forward to characterize these BP genes at the individual cell-type level in the arteries as well.

The question of identification of core genes networks may be facilitated by our approach in this study, which includes using eQTL information from tissues or cell types of interest and genotypes to identify potentially relevant genes for a trait. As the expansion of publicly available resources continues, more information may be used for these purposes. Our analysis implicates specific variants that can be functionally tested for their effect on both gene expression and the phenotype.

Materials and Methods

Study participants and summary of genotypes, phenotypes and association results used in this study

The full descriptions of the prior underlying studies, phenotypes and association results for the GERA cohort are in Hoffmann *et al*. (38) and are briefly recapitulated here. The Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort, part of the RPGEH, consists of individuals from five ethnic backgrounds; the majority is non-Hispanic white (EUR), with the remainder including Latino, East Asians, African Americans and South Asians. A total of 99 785 individuals were analyzed, of which 80 792 were EUR individuals. The populations were each genotyped on custom population-specific Affymetrix Axiom SNP genotyping arrays (63,64) and imputed to the 1000 Genomes Phase I Integrated Release Version 3 haplotype panel. Analyses of GERA alone, with the results of the International Consortium for Blood Pressure (ICBP, $n = 69\,396$) study (65), and with the ICBP and the UK Biobank (UKB, $n = 152\,081$) study (66), identified 316 novel BP loci. Combined with the set of replicated BP GWAS loci available at that time, there were a total of 390 BP loci

we considered to be of interest. Of these, 367 had minor allele frequency (MAF) > 0.001 in the GERA EUR study, which was used as the reference population for the eQTL analyses described below.

For the purpose of several of the analyses described in this paper, we used these association results, as well as summary statistics available from 80 792 GERA EUR individuals from the Hoffmann *et al*. (38) study and genotypes from a subset of 71 404 GERA EUR 'unrelated' individuals (third degree or beyond, pruned by the KING software for relationship inference) (67). We converted genotypes prepared in the Hoffmann *et al*. (38) study after imputation from IMPUTE2 genotype probability format to PLINK 'hard' calls (the most likely genotype), setting genotypes with uncertainty greater than 0.25 to missing and retaining variants with <10% missing data, a Hardy Weinberg equilibrium test $P < 1 \times 10^{-6}$, and imputation quality score ≥0.3. In order to report univariate summary statistics within the 71 404 individuals, we used the —association option for analysis of a quantitative trait (Wald test) with PLINK v1.9 (68). We analyzed covariate-adjusted longitudinal systolic (SBP) and diastolic (DBP) blood pressure in this study, as also described in Hoffmann *et al*. (38).

The ARIC genotypes, phenotypes and association methods are described in the Supplementary Material, Methods.

Ethics statement

The Kaiser Foundation Research Institute and University of California San Francisco Institutional Review Boards approved the study (using GERA study data). Approval was granted from the relevant institutional review boards for the participating study centers (University of North Carolina, University of Minnesota, University of Mississippi Medical Center, and Johns Hopkins University) of the ARIC study. Informed, written consent was obtained from all study participants.

GTEx genotypes and expressions

We analyzed genotypes and expression data from the Genotype-Tissue Expression (GTEx; phs000424.v6.p1) Project (69) v6p for the SKAT analysis (see below) from the aorta, tibial artery, heart left ventricle and heart atrial appendage tissues. Normalized expression was analyzed for these tissues, with the top three principal components, available PEER factors (15–35, depending on sample size), genotyping array platform and sex used as covariates, all available from the GTEx portal. We used SNP-gene associations from the associated *.v6p.all_snpgene_pairs.txt.gz files from the authors' eQTL analyses.

Partitioned heritability and generation of regulatory element maps

Partitioned heritability analyses and generation of putative regulatory element maps and deltaSVM scores are described in the Supplementary Material, Methods.

Gene-based testing with SKAT

We used the sequence-kernel association test (SKAT) (23,70) to test genes with median RPKM ≥0.3 in GTEx samples for the aorta ($n = 197$) and tibial ($n = 285$) arteries with their respective variant sets. For each gene, we tested all variants within 50 kb of the gene start or end, inclusive of the entire gene body, per GENCODE v19 annotations (https://www.gencodegenes.org/releases/19.html). The weights used were taken as the absolute value of the deltaSVM score for each variant to reflect its predicted impact; for comparison, we also ran SKAT using default weights with beta density parameters [weights.beta = c(1,25), which up-weights rare variants as compared to common variants], as well as equal weights to all variants [weights.beta = c(1,1)]. We tested association of each gene with adjusted SBP and DBP phenotype residuals (see above), as well as the GTEx normalized expression data with covariates (release v6p, https://www.gtexportal.org/), from the aorta and tibial arteries. We restricted our primary analyses in each of the kidney cell types to the 20 monogenic hypertension and hypotension genes. We additionally tested tissue- or cell-type-specific groupings in the ARIC dataset with the adjusted QT interval phenotype using the sets for the heart and heart tissues from GTEx, arteries and kidney cell types, as described above.

Predicted gene expression association with BP and gene annotation

The predicted gene expression association and gene annotation methods are described in the Supplementary Material, Methods.

Statistical significance

Statistical significance was determined using the Benjamini–Hochberg (71) (BH) method for multiple test correction to adjust for the number of genes within each analysis. We made no additional adjustments for the number of tissues, in part due to the correlation of specific subsets (the arteries and individual kidney cell types), and as we examined genes across multiple analyses, for phenotype and for expression.

Data availability

The full GERA data are available through application to the KP Research Bank Portal, http://researchbank.kaiserpermanente.org/for-researchers/. A subset (~78%) of dbGaP-consented GERA participants has genotype and hypertension status phenotype data available under the dbGaP accession code phs000674.v1.p1. The full ARIC data cannot be shared publicly because participants did not consent to this. ARIC data are available upon request from the ARIC Data Coordinating Center at the University of North Carolina, Chapel Hill (https://sites.cscc.unc.edu/aric/distributionagreements). The GTEx data used for the analyses described in this manuscript were obtained from the GTEx Portal and dbGaP accession number phs000424.v6.p1.

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

The authors thank the staff and participants of the ARIC study for their important contributions. The Genotype-Tissue Expression (GTEx) Project was 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. The data used for

the analyses described in this manuscript were obtained from the GTEx Portal and dbGaP accession number phs000424.v6.p1. We are grateful to the Kaiser Permanente Northern California members who have generously agreed to participate in the Kaiser Permanente Research Program on Genes, Environment and Health.

Funding

National Heart, Lung, and Blood Institute contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C and HHSN268201100012C), R01HL087641, R01HL59367, R01HL086694; the National Human Genome Research Institute (contract U01HG004402); the National Institutes of Health (contract HHSN268200625226C) (Atherosclerosis Risk in Communities Study, which is carried out as a collaborative study); the National Institutes of Health and National Institutes of Health Roadmap for Medical Research (UL1RR025005); the National Institutes of Health (HL128782, HL086694 to A.C.); the Robert Wood Johnson Foundation; the Wayne and Gladys Valley Foundation; the Ellison Medical Foundation; the Kaiser Permanente Community Benefit Programs; the National Institute on Aging, National Institute of Mental Health and the National Institute of Health Common Fund (RC2 AG036607 to C.S. and N.R.); Geneva University Hospitals; and the Foundation of Medical Researchers, Geneva (to G.E.).

Conflict of Interest statement. The authors declare no conflicts of interest.

References

1. Boyle, E.A., Li, Y.I. and Pritchard, J.K. (2017) An expanded view of complex traits: from polygenic to omnigenic. *Cell*, **169**, 1177–1186.
2. Kirschner, M.W. (2005) The meaning of systems biology. *Cell*, **121**, 503–504.
3. Alon, U. (2006, 2006) An introduction to systems biology: design principles of biological circuits. In *An Introduction to Systems Biology: Design Principles of Biological Circuits*, 1st edn. Chapman and Hall/CRC, Boca Raton, FL.
4. Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U. and Gaul, U. (2008) Predicting expression patterns from regulatory sequence in Drosophila segmentation. *Nature*, **451**, 535–540.
5. Emilsson, V., Thorleifsson, G., Zhang, B., Leonardson, A.S., Zink, F., Zhu, J., Carlson, S., Helgason, A., Walters, G.B., Gunnarsdottir, S. *et al*. (2008) Genetics of gene expression and its effect on disease. *Nature*, **452**, 423–428.
6. Zhong, H., Yang, X., Kaplan, L.M., Molony, C. and Schadt, E.E. (2010) Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am. J. Hum. Genet*., **86**, 581–591.
7. 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.
8. Davidson, E.H. (2010) Emerging properties of animal gene regulatory networks. *Nature*, **468**, 911–920.
9. Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S. *et al*. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, **159**, 1665–1680.
10. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S. and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, **485**, 376–380.
11. Phillips-Cremins, J.E. (2014) Unraveling architecture of the pluripotent genome. *Curr. Opin. Cell Biol*., **28**, 96–104.
12. 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.
13. ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, **489**, 57–74.
14. Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J. *et al*. (2015) Integrative analysis of 111 reference human epigenomes. *Nature*, **518**, 317–330.
15. Maurano, M.T., Haugen, E., Sandstrom, R., Vierstra, J., Shafer, A., Kaul, R. and Stamatoyannopoulos, J.A. (2015) Large-scale identification of sequence variants influencing human transcription factor occupancy in vivo. *Nat. Genet*., **47**, 1393–1401.
16. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K. *et al*. (2015) Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet*., **47**, 1228–1235.
17. Lee, D., Gorkin, D.U., Baker, M., Strober, B.J., Asoni, A.L., McCallion, A.S. and Beer, M.A. (2015) A method to predict the impact of regulatory variants from DNA sequence. *Nat. Genet*., **47**, 955–961.
18. Lifton, R.P. (1996) Molecular genetics of human blood pressure variation. *Science*, **272**, 676–680.
19. Lifton, R.P., Gharavi, A.G. and Geller, D.S. (2001) Molecular mechanisms of human hypertension. *Cell*, **104**, 545–556.
20. Choi, M., Scholl, U.I., Yue, P., Björklund, P., Zhao, B., Nelson-Williams, C., Ji, W., Cho, Y., Patel, A., Men, C.J. *et al*. (2011) K+ channel mutations in adrenal aldosterone-producing adenomas and hereditary hypertension. *Science*, **331**, 768–772.
21. Banda, Y., Kvale, M.N., Hoffmann, T.J., Hesselson, S.E., Ranatunga, D., Tang, H., Sabatti, C., Croen, L.A., Dispensa, B.P., Henderson, M. *et al*. (2015) Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. *Genetics*, **200**, 1285–1295.
22. Kvale, M.N., Hesselson, S., Hoffmann, T.J., Cao, Y., Chan, D., Connell, S., Croen, L.A., Dispensa, B.P., Eshragh, J., Finn, A. *et al*. (2015) Genotyping informatics and quality control for 100,000 subjects in the genetic epidemiology research on adult health and aging (GERA) cohort. *Genetics*, **200**, 1051–1060.
23. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M. and Lin, X. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet*., **89**, 82–93.
24. Barbeira, A.N., Dickinson, S.P., Bonazzola, R., Zheng, J., Wheeler, H.E., Torres, J.M., Torstenson, E.S., Shah, K.P., Garcia, T., Edwards, T.L. *et al*. (2018) Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat. Commun*., **9**, 1825.

25. Atherosclerosis Risk in Communities [Internet]. https://www2.cscc.unc.edu/aric [Accessed: 23 October 2015]
26. The Atherosclerosis Risk in Communities (ARIC) Study: Design and Objectives (1989) The ARIC investigators. *Am. J. Epidemiol.*, **129**, 687–702.
27. Lee, D., Kapoor, A., Safi, A., Song, L., Halushka, M.K., Crawford, G.E. and Chakravarti, A. (2018) Human cardiac cis-regulatory elements, their cognate transcription factors, and regulatory DNA sequence variants. *Genome Res.*, **28**, 1577–1588.
28. Ghandi, M., Lee, D., Mohammad-Noori, M. and Beer, M.A. (2014) Enhanced regulatory sequence prediction using gapped k-mer features. *PLoS Comput. Biol.*, **10**, e1003711.
29. Lee, D. (2016) LS-GKM: a new gkm-SVM for large-scale datasets. *Bioinforma. Oxf. Engl.*, **32**, 2196–2198.
30. Postema, P.G. and Wilde, A.A.M. (2014) The measurement of the QT interval. *Curr. Cardiol. Rev.*, **10**, 287–294.
31. Dekker, J.M., Crow, R.S., Hannan, P.J., Schouten, E.G., Folsom, A.R. and ARIC Study (2004) Heart rate-corrected QT interval prolongation predicts risk of coronary heart disease in black and white middle-aged men and women: the ARIC study. *J. Am. Coll. Cardiol.*, **43**, 565–571.
32. Newton-Cheh, C., Larson, M.G., Corey, D.C., Benjamin, E.J., Herbert, A.G., Levy, D., D'Agostino, R.B. and O'Donnell, C.J. (2005) QT interval is a heritable quantitative trait with evidence of linkage to chromosome 3 in a genome-wide linkage analysis: the Framingham heart study. *Heart Rhythm.*, **2**, 277–284.
33. Busjahn, A., Knoblauch, H., Faulhaber, H.D., Boeckel, T., Rosenthal, M., Uhlmann, R., Hoehe, M., Schuster, H. and Luft, F.C. (1999) QT interval is linked to 2 long-QT syndrome loci in normal subjects. *Circulation*, **99**, 3161–3164.
34. Arking, D.E., Pfeufer, A., Post, W., Kao, W.H.L., Newton-Cheh, C., Ikeda, M., West, K., Kashuk, C., Akyol, M., Perz, S. *et al.* (2006) A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization. *Nat. Genet.*, **38**, 644–651.
35. Tomás, M., Napolitano, C., De Giuli, L., Bloise, R., Subirana, I., Malovini, A., Bellazzi, R., Arking, D.E., Marban, E., Chakravarti, A. *et al.* (2010) Polymorphisms in the NOS1AP gene modulate QT interval duration and risk of arrhythmias in the long QT syndrome. *J. Am. Coll. Cardiol.*, **55**, 2745–2752.
36. Avery, C.L., Wassel, C.L., Richard, M.A., Highland, H.M., Bien, S., Zubair, N., Soliman, E.Z., Fornage, M., Bielinski, S.J., Tao, R. *et al.* (2017) Fine mapping of QT interval regions in global populations refines previously identified QT interval loci and identifies signals unique to African and Hispanic descent populations. *Heart Rhythm.*, **14**, 572–580.
37. Arking, D.E., Pulit, S.L., Crotti, L., van der Harst, P., Munroe, P.B., Koopmann, T.T., Sotoodehnia, N., Rossin, E.J., Morley, M., Wang, X. *et al.* (2014) Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat. Genet.*, **46**, 826–836.
38. Hoffmann, T.J., Ehret, G.B., Nandakumar, P., Ranatunga, D., Schaefer, C., Kwok, P.-Y., Iribarren, C., Chakravarti, A. and Risch, N. (2017) Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat. Genet.*, **49**, 54–64.
39. Evangelou, E., Warren, H.R., Mosen-Ansorena, D., Mifsud, B., Pazoki, R., Gao, H., Ntritsos, G., Dimou, N., Cabrera, C.P., Karaman, I. *et al.* (2018) Genetic analysis of over one million people identifies 535 novel loci for blood pressure. *Nat. Genet.*, **50**, 1412–1425.
40. Newton-Cheh, C., Johnson, T., Gateva, V., Tobin, M.D., Bochud, M., Coin, L., Najjar, S.S., Zhao, J.H., Heath, S.C., Eyheramendy, S. *et al.* (2009) Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.*, **41**, 666–676.
41. Johnson, T., Gaunt, T.R., Newhouse, S.J., Padmanabhan, S., Tomaszewski, M., Kumari, M., Morris, R.W., Tzoulaki, I., O'Brien, E.T., Poulter, N.R. *et al.* (2011) Blood pressure loci identified with a gene-centric array. *Am. J. Hum. Genet.*, **89**, 688–700.
42. Wain, L.V., Verwoert, G.C., O'Reilly, P.F., Shi, G., Johnson, T., Johnson, A.D., Bochud, M., Rice, K.M., Henneman, P., Smith, A.V. *et al.* (2011) Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nat. Genet.*, **43**, 1005–1011.
43. Levy, D., Ehret, G.B., Rice, K., Verwoert, G.C., Launer, L.J., Dehghan, A., Glazer, N.L., Morrison, A.C., Johnson, A.D., Aspelund, T. *et al.* (2009) Genome-wide association study of blood pressure and hypertension. *Nat. Genet.*, **41**, 677–687.
44. Goldsmith, O., Solomon, D.H. and Horton, R. (1967) Hypogonadism and mineralocorticoid excess. The 17-hydroxylase deficiency syndrome. *N. Engl. J. Med.*, **277**, 673–677.
45. Li, Q., Gao, T., Yuan, Y., Wu, Y., Huang, Q., Xie, F., Ran, P., Sun, L. and Xiao, C. (2017) Association of CYP17A1 genetic polymorphisms and susceptibility to essential hypertension in the southwest Han Chinese population. *Med. Sci. Monit. Int. Med. J. Exp. Clin. Res.*, **23**, 2488–2499.
46. Li, C., Kim, Y.K., Dorajoo, R., Li, H., Lee, I.-T., Cheng, C.-Y., He, M., Sheu, W.H.-H., Guo, X., Ganesh, S.K. *et al.* (2017) Genome-wide association study meta-analysis of long-term average blood pressure in east Asians. *Circ. Cardiovasc. Genet.*, **10**, e001527.
47. Nguyen, K.-D.H., Pihur, V., Ganesh, S.K., Rakha, A., Cooper, R.S., Hunt, S.C., Freedman, B.I., Coresh, J., Kao, W.H.L., Morrison, A.C. *et al.* (2013) Effects of rare and common blood pressure gene variants on essential hypertension: results from the family blood pressure program, CLUE, and atherosclerosis risk in communities studies. *Circ. Res.*, **112**, 318–326.
48. Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
49. Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.*, **37**, 1–13.
50. Lin, C.G., Chen, C.-C., Leu, S.-J., Grzeszkiewicz, T.M. and Lau, L.F. (2005) Integrin-dependent functions of the angiogenic inducer NOV (CCN3): implication in wound healing. *J. Biol. Chem.*, **280**, 8229–8237.
51. Lin, C.G., Leu, S.-J., Chen, N., Tebeau, C.M., Lin, S.-X., Yeung, C.-Y. and Lau, L.F. (2003) CCN3 (NOV) is a novel angiogenic regulator of the CCN protein family. *J. Biol. Chem.*, **278**, 24200–24208.
52. Shimoyama, T., Hiraoka, S., Takemoto, M., Koshizaka, M., Tokuyama, H., Tokuyama, T., Watanabe, A., Fujimoto, M., Kawamura, H., Sato, S. *et al.* (2010) CCN3 inhibits neointimal hyperplasia through modulation of smooth muscle cell growth and migration. *Arterioscler. Thromb. Vasc. Biol.*, **30**, 675–682.
53. Guo, D.-C., Grove, M.L., Prakash, S.K., Eriksson, P., Hostetler, E.M., LeMaire, S.A., Body, S.C., Shalhub, S., Estrera, A.L., Safi, H.J. *et al.* (2016) Genetic variants in LRP1 and ULK4 are associated with acute aortic dissections. *Am. J. Hum. Genet.*, **99**, 762–769.

54. Niu, W.-Q., You, Y.-G. and Qi, Y. (2012) Strong association of methylenetetrahydrofolate reductase gene C677T polymorphism with hypertension and hypertension-in-pregnancy in Chinese: a meta-analysis. *J. Hum. Hypertens.*, **26**, 259–267.
55. Frosst, P., Blom, H.J., Milos, R., Goyette, P., Sheppard, C.A., Matthews, R.G., Boers, G.J., den Heijer, M., Kluijtmans, L.A. and van den Heuvel, L.P. (1995) A candidate genetic risk factor for vascular disease: a common mutation in methylenetetrahydrofolate reductase. *Nat. Genet.*, **10**, 111–113.
56. Nakata, Y., Katsuya, T., Takami, S., Sato, N., Fu, Y., Ishikawa, K., Takiuchi, S., Rakugi, H., Miki, T., Higaki, J. *et al.* (1998) Methylenetetrahydrofolate reductase gene polymorphism: relation to blood pressure and cerebrovascular disease. *Am. J. Hypertens.*, **11**, 1019–1023.
57. Kosmas, I.P., Tatsioni, A. and Ioannidis, J.P.A. (2004) Association of C677T polymorphism in the methylenetetrahydrofolate reductase gene with hypertension in pregnancy and pre-eclampsia: a meta-analysis. *J. Hypertens.*, **22**, 1655–1662.
58. Van Woudenberg, M., Shin, J., Bernard, M., Syme, C., Abrahamowicz, M., Leonard, G., Perron, M., Richer, L., Veillette, S., Gaudet, D. *et al.* (2015) CYP17A1 and blood pressure reactivity to stress in adolescence. *Int. J. Hypertens.*, **2015**, 734586.
59. Airik, R., Slaats, G.G., Guo, Z., Weiss, A.-C., Khan, N., Ghosh, A., Hurd, T.W., Bekker-Jensen, S., Schrøder, J.M., Elledge, S.J. *et al.* (2014) Renal-retinal ciliopathy gene Sdccag8 regulates DNA damage response signaling. *J. Am. Soc. Nephrol. JASN*, **25**, 2573–2583.
60. Ganesh, S.K., Tragante, V., Guo, W., Guo, Y., Lanktree, M.B., Smith, E.N., Johnson, T., Castillo, B.A., Barnard, J., Baumert, J. *et al.* (2013) Loci influencing blood pressure identified using a cardiovascular gene-centric array. *Hum. Mol. Genet.*, **22**, 1663–1678.
61. Ahn, S.-Y. and Gupta, C. (2017) Genetic programming of hypertension. *Front. Pediatr.*, **5**, 285.
62. Lee, H.-J., Kang, J.-O., Kim, S.-M., Ji, S.-M., Park, S.-Y., Kim, M.E., Jigden, B., Lim, J.E., Hwang, S.-Y., Lee, Y.-H. *et al.* (2016) Gene silencing and Haploinsufficiency of Csk increase blood pressure. *PLoS One*, **11**, e0146841.
63. Hoffmann, T.J., Kvale, M.N., Hesselson, S.E., Zhan, Y., Aquino, C., Cao, Y., Cawley, S., Chung, E., Connell, S., Eshragh, J. *et al.* (2011) Next generation genome-wide association tool: design and coverage of a high-throughput European-optimized SNP array. *Genomics*, **98**, 79–89.
64. Hoffmann, T.J., Zhan, Y., Kvale, M.N., Hesselson, S.E., Gollub, J., Iribarren, C., Lu, Y., Mei, G., Purdy, M.M., Quesenberry, C. *et al.* (2011) Design and coverage of high throughput genotyping arrays optimized for individuals of East Asian, African American, and Latino race/ethnicity using imputation and a novel hybrid SNP selection algorithm. *Genomics*, **98**, 422–430.
65. International Consortium for Blood Pressure Genome-Wide Association Studies, Ehret, G.B., Munroe, P.B., Rice, K.M., Bochud, M., Johnson, A.D., Chasman, D.I., Smith, A.V., Tobin, M.D., Verwoert, G.C. *et al.* (2011) Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*, **478**, 103–109.
66. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M. *et al.* (2015) UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.*, **12**, e1001779.
67. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M. and Chen, W.-M. (2010) Robust relationship inference in genome-wide association studies. *Bioinforma. Oxf. Engl.*, **26**, 2867–2873.
68. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
69. GTEx Consortium, Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) Groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI *et al.* (2017) Genetic effects on gene expression across human tissues. *Nature*, **550**, 204–213.
70. Seunggeun Lee, with contributions from Larisa Miropolsky and Michael Wu (2015). *SKAT: SNP-Set (Sequence) Kernel Association Test*. R package version 1.1.2. https://CRAN.R-project.org/package=SKAT.
71. Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.*, **57**, 289–300.