medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

CXCL12 drives natural variation in coronary artery anatomy across diverse populations

Pamela E. Rios Coronado[1]†, Daniela Zanetti[2,3,4]†, Jiayan Zhou[3,2]†, Jeffrey A. Naftaly[1], Pratima Prabala[1], Azalia M. Martínez Jaimes[5,1], Elie N. Farah[6], Xiaochen Fan[1], Soumya Kundu[7,8], Salil S. Deshpande[9], Ivy Evergreen[7], Pik Fang Kho[2,3], Austin T. Hilliard[3], Sarah Abramowitz[10,11,12], Saiju Pyarajan[13], Daniel Dochtermann[13], Million Veteran Program[14], Scott M. Damrauer[15,16,17], Kyong-Mi Chang[15,18], Michael G. Levin[10,15], Virginia D. Winn[19], Anca M. Paşca[20], Mary E. Plomondon[21,22], Stephen W. Waldo[21,22,23], Philip S. Tsao[3,24,25], Anshul Kundaje[7,8], Neil C. Chi[6], Shoa L. Clarke[2,3,26]‡, Kristy Red-Horse[1,27,28]*‡, Themistocles L. Assimes[2,3,25,29]*‡

[1]Department of Biology, Stanford University; Stanford, CA, USA
[2]Department of Medicine, Division of Cardiovascular Medicine, Stanford University School of Medicine; Stanford, CA, USA
[3]VA Palo Alto Health Care System; Palo Alto, CA, USA
[4]Institute of Genetic and Biomedical Research, National Research Council; Cagliari, Sardinia, Italy
[5]Department of Developmental Biology, Stanford University School of Medicine; Stanford, CA, USA
[6]Department of Medicine, Division of Cardiology, University of California San Diego; La Jolla, CA, USA
[7]Department of Genetics, Stanford University School of Medicine; Stanford, CA, USA
[8]Department of Computer Science, Stanford University; Stanford, CA, USA
[9]Institute for Computational and Mathematical Engineering, Stanford University School of Medicine; Stanford, CA, USA
[10]Department of Medicine, Division of Cardiovascular Medicine, University of Pennsylvania Perelman School of Medicine; Philadelphia, PA, USA
[11]Sarnoff Cardiovascular Research Foundation; McLean, VA, USA
[12]Donald and Barbara Zucker School of Medicine at Hofstra/Northwell; Hempstead, NY, USA
[13]Center for Data and Computational Sciences, VA Boston Healthcare System; Boston, MA, USA
[14]A list of consortium members can be found in the Supplementary File
[15]Corporal Michael J. Crescenz VA Medical Center; Philadelphia, PA, USA
[16]Department of Surgery, University of Pennsylvania Perelman School of Medicine; Philadelphia, PA, USA
[17]Department of Genetics, University of Pennsylvania Perelman School of Medicine; Philadelphia, PA, USA
[18]Department of Medicine, Division of Gastroenterology and Hepatology, University of Pennsylvania Perelman School of Medicine; Philadelphia, PA, USA
[19]Department of Obstetrics and Gynecology, Stanford University School of Medicine; Stanford, CA, USA
[20]Department of Pediatrics, Neonatology, Stanford University School of Medicine; Stanford, CA, USA

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

[21]Department of Medicine, Rocky Mountain Regional VA Medical Center; Aurora, CO, USA
[22]CART Program, VHA Office of Quality and Patient Safety; Washington, DC, USA
[23]Division of Cardiology, University of Colorado School of Medicine; Aurora, CO, USA
[24]Department of Medicine, Stanford University School of Medicine; Stanford, CA, USA
[25]Cardiovascular Institute, Stanford University School of Medicine; Stanford, CA, USA
[26]Department of Medicine, Stanford Prevention Research Center, Stanford University School of Medicine; Stanford, CA, USA
[27]Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine; Stanford, CA, USA
[28]Howard Hughes Medical Institute; Chevy Chase, MD, USA
[29]Department of Epidemiology and Population Health, Stanford University School of Medicine; Stanford, CA, USA

†Pamela E. Rios Coronado, Daniela Zanetti, and Jiayan Zhou contributed as co-lead junior authors.

‡Shoa L. Clarke, Kristy Red-Horse, and Themistocles L. Assimes contributed as co-supervising senior authors.

*Kristy Red-Horse. E-mail: kredhors@stanford.edu
*Themistocles L. Assimes. E-mail: tassimes@stanford.edu

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Abstract

To efficiently distribute blood flow to cardiac muscle, the coronary artery tree must follow a specific branching pattern over the heart. How this pattern arises in humans is unknown due to the limitations of studying human heart development. Here, we leveraged a natural variation of coronary artery anatomy, known as coronary dominance, in genetic association studies to identify the first known driver of human coronary developmental patterning. Coronary dominance refers to whether the right, left, or both coronary arteries branch over the posterior left ventricle, but whether this variability is heritable and how it would be genetically regulated was completely unknown. By conducting the first large-scale, multi-ancestry genome-wide association study (GWAS) of coronary dominance in 61,043 participants of the VA Million Veteran Program, we observed moderate heritability (27.7%) with ten loci reaching genome wide significance. An exceptionally strong association mapped DNA variants to a non-coding region near the chemokine *CXCL12* in both European and African ancestries, which overlapped with variants associated with coronary artery disease. Genomic analyses predicted these variants to impact *CXCL12* levels, and imaging revealed dominance to develop during fetal life coincident with *CXCL12* expression. Reducing *Cxcl12* in mice to model the human genetics altered septal artery dominance patterns and caused coronary branches to develop away from *Cxcl12* expression domains. *Cxcl12* heterozygosity did not compromise overall artery coverage as seen with full deletion, but instead changed artery patterning, reminiscent of the human scenario. Together, our data support *CXCL12* as a critical determinant of human coronary artery growth and patterning and lay a foundation for the utilization of developmental pathways to guide future precision 'medical revascularization' therapeutics.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Main Text

Introduction

The heart is a muscular organ that beats continuously throughout life. It requires a highly optimized network of coronary arteries to ensure efficient and constant perfusion with oxygenated blood[1,2]. Diseases of the coronary arteries that interrupt blood flow are a leading cause of death worldwide[3], yet we have little knowledge of how these important vessels are formed and patterned in human hearts.

Most of the major coronary artery branches follow a standard topological pattern in nearly all humans. The left main coronary artery (LCA) originates from the left side of the aorta and gives rise to the left anterior descending (LAD) and left circumflex (LCx) arteries, which provide blood flow to the anterior and lateral walls of the left ventricle, respectively. The right coronary artery (RCA) takes off from the right side of the aorta and follows a course around the right ventricle towards the back of the heart. Except for coronary anomalies, which can cause adverse cardiovascular outcomes, the origins and paths of the LMCA, LAD, LCx, and RCA, are the same throughout the population.

One exception to this standardized anatomy involves the posterior descending artery (PDA), which provides blood to the back of the heart, and variations in structure are termed coronary artery dominance[4]. In approximately 80% of the population, the PDA arises from the RCA (right dominant), but in ~10%, the RCA is diminutive, and the PDA instead arises from the LCx (left dominant). In the remaining ~10%, branches perfusing the inferior wall arise from both the RCA and LCx (co-dominant) (**Figure 1A**). An analogous normal variation in artery anatomy has been described in mice involving the origin of the septal artery, which can also be right, left, or co-dominant[5]. Crosses between different strains showed that mouse dominance is genetically regulated[5], and like humans, right dominance is by far the most common phenotype[5-7]. However, for both humans and mice, how and when dominance is determined is unknown.

Developmental biology research has demonstrated that artery development has conserved features across organs and species[8-10]. Work in model organisms has demonstrated that arteries, including coronary arteries, are generated through maturation of an immature capillary plexus. Specifically, endothelial cells (ECs) of the immature capillary plexus undergo artery differentiation, migrate against the direction of blood flow, and coalesce together to progressively lengthen artery branches[11-14]. During this process, the artery ECs induce surrounding pericytes or mesenchyme to differentiate into smooth muscle[15,16]. Thus, human coronary artery development and dominance establishment could involve some of the molecular pathways known to drive these processes in model organisms.

To study the mechanisms of human coronary artery developmental patterning, we took advantage of the natural variation in coronary dominance to conduct the first genome-wide association study (GWAS) of coronary anatomical phenotypes. We analyzed more than 61,000 genotyped participants of the Million Veteran Program (MVP) who had phenotypic data from invasive coronary angiograms. We found significant heritability and identified *CXCL12* - a gene implicated in coronary artery disease (CAD) over sixteen years ago[17] - as a potential regulator of coronary patterning. Genomics analyses nominated candidate causal variants in the *CXCL12* locus, and *in vivo* mouse experiments supported a functional role in influencing coronary artery localization and dominance during heart development.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Results

A large-scale, genetically diverse genome-wide association study of coronary artery dominance

We leveraged the Million Veteran Program (MVP) linked to the Veterans Health Administration (VA) Clinical Assessment Reporting and Tracking (CART) Program for cardiac catheterization laboratories to conduct a large-scale, multi-ancestry GWAS of dominance (**Figure 1B**). Within this biobank, we identified 61,043 participants with a dominance assessment and genotype data available imputed to a multi-ancestry TOPMed reference panel (**Figure 1B, Supplementary Table 1**). We used Scalable and Accurate Implementation of GEneralized mixed model (SAIGE)[18] to conduct each GWAS. Ancestry-specific GWAS were conducted in the three largest genetically inferred ancestry groups: European (EUR), African (AFR), and Admixed American (AMR). We also performed pooled analyses containing all participants (All ancestry), including those not assigned to the major ancestry groups. Because the genetics of dominance had never been explored, it was not known a priori whether left- and co-dominance were related and should be combined or not. Thus, to be thorough and unbiased, we ran two GWAS for each population: (1) right (reference) versus left plus co-dominance (RvsLCo) and (2) right (reference) versus left only (RvsL).

Routine statistical analyses indicated a robust genetic signature underlying coronary dominance. First, quantile-quantile (Q-Q) plots comparing P-values for genetic variants associated with dominance to those expected for no association (null hypothesis) demonstrated a strong genetic signal in the All ancestry (**Figure 2A**), EUR (**Figure 2B**), and AFR (**Figure 2C**) GWAS. There was no evidence of population stratification based on the calculated genomic control lambdas for common variants with minor allele frequency (MAF) > 0.01. These observations were bolstered by a point estimate of total SNP-based narrow-sense heritability (h^2) of 0.277 (95% confidence intervals, 0.150-0.404) among EUR using a linkage disequilibrium (LD) and MAF stratified multicomponent approach implemented in GCTA, GREML-LDMS-I, and assuming the observed prevalence of non-right dominance of 17% (**Supplementary Figure 1A-B,** and **Supplementary Table 2**). A large fraction of the additive genetic variance is derived from low frequency and rare variants (**Supplementary Figure 1C**). The genetic signal was weak to non-existent among AMR (**Figure 2D**), likely due to the lower numbers in this group (**Supplementary Table 1**). Thus, a clear genetic signal emerged for coronary artery dominance in datasets with greater than 2000 non-right individuals.

Ten loci across seven chromosomes reached genome wide significance (GWS) in one or more of the GWAS, as defined by an algorithm implemented in FUMA-GWAS[19] (**Figure 2E-H, Table 1, Supplementary Figure 2,** and **Supplementary Table 3**). Of these, we consider seven loci of highest confidence because they contained common lead variants in both the All ancestry GWAS and in one or more of the ancestry specific GWAS (**Table 1**). Two additional loci were defined by either one or a small number of rare variants in high LD with MAF < 0.005. One locus on chromosome 3 was defined by multiple low frequency variants reaching GWS only in AMR despite higher frequencies and better power to implicate the same region in the All ancestry, EUR, and AFR groups. Furthermore, the saddle point approximation (SPA) for P-values of SNPs found in this region for AMR did not converge in the other populations.

Variants were near genes known to play a role in cardiac development (e.g., *NKX2-5*) and the most significant with the highest effect size was downstream of *CXCL12*, which, remarkably, reached GWS in both EUR and AFR (**Figure 2E, F,** and **G**). Regional association and LD plots for GWS loci are included in **Supplementary Figure 3**. Statistical fine mapping with SuSiEx[20],

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

which is used to help predict potential phenotype-causing variants, identified a single credible set of SNPs in five of the loci, two credible sets within the *NKX2-5* locus on chromosome 5 and four credible sets within the *CXCL12* locus on chromosome 10. **Supplementary Table 4** lists the SNPs within each credible set with the highest posterior probability (PP) for being independently causal. **Supplementary Table 5** provides full SAIGE output for all lead and independently significant variants identified by FUMA, and all SuSiEx lead genetic variants. **Supplementary Table 6** is a list of mapped genes using GTEx v8 and chromatin folding maps as implemented in FUMA. Lastly, **Supplementary Table 7** provides annotation hyperlinks for the lead SNPs in Table 1 to five online variant-based portals. In total, these data identified several loci that merit further examination regarding their involvement in coronary artery patterning.

Lead genetic variants in both EUR and AFR populations are linked to* CXCL12 *expression

We next further assessed the *CXCL12* locus because it was the strongest signal observed in multiple GWAS (**Figure 2F** and **G**) and previous studies have demonstrated a role for *Cxcl12* during artery development in model organisms[11,21-25].

A detailed analysis of the *CXCL12* locus revealed three independent genomic regions containing highly significant variants in the EUR cohort, with two of those being significant in the AFR cohort (**Figure 3A$_{i}$** and **Supplementary Figure 4A, B**). A subgroup analysis of the locus restricted to a comparison of co-dominant individuals alone to right dominant individuals revealed a genome wide significant signal that was nearly identical in shape to the signal observed in the right versus left and the right versus left and co-dominant analyses (**Supplementary Figure 3**). To further validate this locus, we performed genetic association testing with coronary dominance in a small multi-ancestry cohort of 3,141 (75.1% EUR, 22.5% AFR, 2.4% Other) from the Penn Medicine Biobank (PMBB). In PMBB, the strongest common variant association with left/co-dominance at the *CXCL12* locus was at rs559580 in the significant genomic region closest to *CXCL12* (MAF: 0.18, OR: 1.46, $p=6.7x10^{-5}$) (**Figure 3A$_{ii}$**). Directionality was consistent with results for this variant in the MVP All ancestry GWAS (MAF 0.16, OR: 1.25, $p=1.3x10^{-28}$).

Significant variants within MVP were intergenic and overlapped with open chromatin regions in fetal heart cell types in which the CXCL12 promoter was also accessible (**Figure 3B**). Some of the variants also overlapped with a recently reported Cxcl12 enhancer that activates expression in mouse and zebrafish arterial endothelial cells (**Figure 3A$_{vi}$**)[26] We next performed CADD scoring to prioritize the non-coding variants and found multiple variants within each independent genomic region in both EUR and AFR with a score >10 consistent with a high likelihood of the variants being deleterious (**Supplementary Figure 4A** and **B**)[27,28]. We next identified variants associated with the expression of nearby genes (expression quantitative trait loci, eQTLs). In the EUR GWAS, many variants were eQTLs for *CXCL12* across multiple tissues including arteries, although a few were eQTLs for other genes nearby (**Supplementary Figure 4A** and **Supplementary Figure 5A**). In the AFR GWAS, variants were exclusively eQTLs for *CXCL12* and further identified coronary arteries as a relevant tissue (**Supplementary Figure 4B** and **Supplementary Figure 5B**). Databases utilizing HiC data revealed that dominance-associated eQTLs interacted with *CXCL12* and other nearby promoters among EUR, but only with the CXCL12 promoter in AFR (**Supplementary Figure 4C**). To investigate whether higher or lower expression might be correlated with right or left dominance, we chose eQTLs associated with arterial *CXCL12* expression and observed levels detected with the different alleles. This dataset suggested that the right dominance allele was associated with more

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

CXCL12 expression (**Figure 3A$_{iii}$** and **C**). Thus, at this locus, *CXCL12* is the strongest candidate for a causal gene influencing coronary artery patterning, and the variants identified here likely function through regulating gene expression.

Deep learning models predict transcription factor motifs affected by dominance SNPs

The above fine mapping tools are widely used to nominate candidate causal DNA variants associated with GWAS traits, but they primarily incorporate data from whole tissues and lack the cell type specific information required to glean detailed mechanistic information. Deep learning models such as convolutional neural networks (CNNs) trained to predict chromatin accessibility profiles from single cell ATAC-seq data are emerging as powerful methods for predicting allelic effects of GWAS variants[29-33]. When interpreted through the lens of the model, various functional predictions can be made such as the disruption of transcription factor motifs in individual cell types. The models can therefore provide specific hypotheses about a variant's predicted effect on transcription factor binding and chromatin accessibility and nominate putative cell types of action (**Figure 3D**). Indeed, a base-resolution CNN model trained on human fetal heart scATAC-seq data revealed an enrichment of predicted high effect size de-novo variants from congenital heart disease patients specifically in accessible chromatin in artery endothelial cells and identified key transcription factor binding sites disrupted by disease alleles[33].

We used these cell-type resolved fetal heart models along with similar models trained on chromatin accessibility profiles from cell types derived from adult hearts to score and identify dominance-associated variants predicted to affect chromatin accessibility in cardiac and vascular cell types. The highest-ranking variant at the *CXCL12* dominance locus, rs115213152, was predicted to disrupt motifs belonging to C2H2 domain transcription factors, particularly those in the KLF family, which play a central role in endothelial cell responses to blood flow[34] (**Figure 3E**). This variant overlapped with the dominance peak nearest gene *CXCL12* (**Figure 3A$_{iv}$**). This SNP was predicted to affect the accessibility of a regulatory element in various fetal cardiac cell types (**Figure 3F**). Interestingly, these cell types - cardiomyocytes, pericytes, fibroblasts, smooth muscle cells, vascular endothelial cells, and endocardial cells - also exhibit robust CXCL12 promoter accessibility (**Figure 3F**). This correlation implies that *CXCL12* could be the targeted gene of these regulatory elements in the developing and adult heart.

To determine the potential transcription factor influencing this element, we scrutinized the expression of all candidate motif-binding factors, such as members of the KLF and SP families, in a publicly available fetal heart scRNAseq dataset[35]. *KLF2, KLF3, KLF4,* and *KLF6* showed both high expression and cell type specificity aligning closely with that of *CXCL12* (**Figure 3G**). Another high-ranking variant, rs5784627, was predicted to disrupt motifs belonging to bZIP domain transcription factors such as JUN and FOS (**Figure 3A$_{iv}$** and **Supplementary Figure 5C** and **D**), which were also expressed in cardiac cell types coincident with *CXCL12,* although with less cell type specificity (**Supplementary Figure 5E**). Taken together, these data further implicate dominance-associated variants as regulators of *CXCL12* expression in vascular cells of the heart and suggest potentially relevant transcription factors.

Colocalization of dominance* CXCL12 *variants with coronary artery disease variants

The *CXCL12* region identified in our dominance GWAS highly overlaps with a widely replicated signal for CAD[17]. We used the summary statistics from the Cardiogram*plus*C4D Consortium GWAS[36] and our GWAS of coronary dominance, analyzing the genomic window where the

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

CXCL12 locus is located (from 44210000 bp to 44810000 bp GRCh37) to perform colocalization analyses[37]. Colocalization predicted shared variants within the *CXCL12* locus for coronary dominance and CAD (**Figure 3A$_v$, H,** and **I**). When considering the full interval of the signal, COLOC calculated a global PP Bayesian factor H4 = 0.855. The SNP with the highest H4 PP was rs649192 but nine other highly correlated SNPs (r2>0.8) had similar H4 probabilities (**Supplementary Table 8A**).

The overlap of the signals for the two phenotypes suggested a possible second colocalization in the more proximal middle peak (**Figure 3H**). We thus masked the distal peak closest to *CXCL12* before running COLOC again and observed rs1870634 as a second shared variant within the middle peak. The SNPs highly correlated to rs1870634 also colocalized (**Figure 3I** and **Supplementary Table 8B**). Thus, there are potentially two shared genetic loci that contribute to both dominance status at birth and CAD development later in life.

In contrast, one of the three peaks furthest away from *CXCL12* associated with dominance in the All ancestry and EUR populations did not colocalize with known CAD loci (**Figure 3H** and **I,** grey shading). Interestingly, this peak does overlap open chromatin regions in developing heart cell types (**Figure 3A$_i$** and **B**). Thus, coronary artery development and the establishment of dominance may involve a development-specific regulator of *CXCL12* in addition to those also involved in CAD.

Coronary dominance is apparent at fetal stages when* CXCL12 *is expressed

When dominance is determined is unknown; we hypothesized it would be established during heart development with the formation of coronary arteries. Since angiograms and non-invasive imaging procedures are not performed during healthy pregnancies, we investigated whether dominance is present during fetal stages by performing whole-organ immunofluorescence that allowed visualization of the entire artery tree in 3D (**Figure 4A** and **B**)[38,39]. Tracing primary and secondary branches of the right and left coronary arteries of each heart revealed a pattern reminiscent of the adult anatomy. Left and right primary branches stemmed from the aorta and took a course typical of the LAD, LCx, and RCA in all hearts (**Supplementary Figure 6A-C**).

As in adults, there was variability in the extent to which the right and left branches covered the posterior and inferior aspects of the ventricles at gestational weeks (GW) 14-22. Six of the eight hearts (75%) were clearly right dominant in that the interventricular groove had a branch from the right where the adult PDA would be located, and additional right branches extended onto the inferior left ventricle (**Figure 4C**). We also traced a number of lower order arteries and arterioles (**Figure 4D**) and found that the septum was perfused by branches from the right coronary on the inferior side and the LAD on the anterior side (**Figure 4E**). Two of eight hearts (25%) had a different coverage that we scored as co-dominant because both right and left arteries branched into the inferior wall and septum (**Figure 4F-H** and **Supplementary Figure 6D-F**). For example, one heart showed an artery branching onto the interventricular groove from the right, but it only extended halfway down while the rest of the inferior aspect contained branches from the LCx and LAD (**Figure 4F** and **G**). This resulted in the inferior septum being perfused by both right and left (**Figure 4H**). The observation that dominance can be observed as early as GW14 at ratios reminiscent of those in adults (**Figure 4I**), supports a model where it is established during fetal development.

If dominance is established during fetal development and impacted by *CXCL12*, one would predict the expression of this chemokine and its receptor to be near developing coronary arteries during this time. Fluorescence *in situ* hybridization on fetal heart ventricles from GWs

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

13, 18, and 20 showed expression in coronary arteries, and, at the earliest stage, trabecular cardiomyocytes of the left ventricle (**Figure 4J** and **Supplementary Figure 6G**). Previously generated multiplexed error-robust fluorescence *in situ* hybridization (MERFISH)[40] data that interrogated the expression of 238 genes across all chambers of a GW13 heart[41] localized *CXCL12* to multiple regions and cell types. These included coronary arteries, outflow tract smooth muscle, left ventricle trabecular cardiomyocytes, valvular cardiomyocytes, epicardial-derived cells in the coronary sulcus, and others (**Figure 4K**). Single cell RNA sequencing (scRNAseq) from an additional heart at GW13 was integrated with the MERFISH data and showed expression in similar cell types (**Figure 4L**). The integrated datasets were also used to impute expression of *CXCR4*, the gene encoding the CXCL12 receptor. *CXCR4*-positive cells included those nearby *CXCL12* expression, for example, coronary artery endothelial cells and valve endocardial cells near valvular cardiomyocytes (**Figure 4M**). Blood cells and vascular mural cells (pericytes and smooth muscle) also had expression (**Figure 4L**). Coronary artery distribution in these sections is shown with *MYH11*, which marks artery smooth muscle cells (**Figure 4N**). Thus, *CXCL12* is expressed at the right time and place to influence human coronary artery dominance.

CXCL12 *levels regulate septal artery dominance in mice*

To provide functional evidence that changes in *CXCL12* levels could impact coronary branch variation, we analyzed dominance in mice heterozygous for a *Cxcl12* deletion allele. Heterozygous rather than full knockout mice were used to model the decreased expression predicted for human dominance variants, which would not be loss of function. Like humans, mouse coronary arteries have a branch where there is variability across the population. This variation is seen with the mouse-specific SpA, which like the human PDA, can originate from either the right, left, or both coronary arteries[6,7]. In rare instances, the SpA arises directly from the aorta, but still varies in whether it arises from the right, left, or both sides. SpA dominance in mice shares a key feature with human dominance in that the vast majority are right dominant[6,7]. Given these features, and because the cellular mechanisms of artery development are conserved across different organs and species[8], we considered SpA dominance a good model for studying regulators of human coronary variability.

We assessed SpA dominance in our experiments using either whole-organ immunofluorescence or *in vivo* labeling with fluorescently tagged Isolectin GS-IB$_4$ (**Figure 5A** and **Supplementary Figure 7A-C**). Our experiments recapitulated published SpA dominance ratios in both neonatal and adult hearts (**Figure 5B-D** and **Supplementary Figure 7D**), supporting the notion that, like humans, dominance is established early in life[6,7]. For *Cxcl12* haploinsufficiency, the *Cxcl12*^{DsRed} mouse strain was utilized. This strain is maintained on the C57Bl/6J background and has part of the *Cxcl12* coding region replaced with a red fluorescent protein, *DsRed*, so that it is both a deletion line and a gene expression reporter[42]. QPCR assessing *Cxcl12* mRNA levels in hearts from wildtype (*Cxcl12*$^{+/+}$), heterozygous (*Cxcl12*$^{DsRed/+}$), and homozygous (*Cxcl12*$^{DsRed/DsRed}$) mice demonstrated that expression in *Cxcl12*$^{DsRed/+}$ is approximately 50% lower than *Cxcl12*$^{+/+}$ (**Figure 5E**), making them appropriate for assessing SpA dominance in hearts with decreased *Cxcl12* expression.

Cxcl12$^{DsRed/+}$ mice had shifted SpA dominance ratios when compared with wild type littermates (**Figure 5F**). These heterozygous mice had a lower incidence of right dominance and a higher incidence of co-dominance and aortic-origin. Therefore, *Cxcl12* heterozygosity changed mouse SpA dominance patterns, supporting the hypothesis that genetic variants affecting

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

CXCL12 expression could influence coronary artery developmental patterning and dominance in humans.

To understand how *Cxcl12* levels influence SpA branching, we first investigated the developmental stages leading to SpA development and dominance establishment and then determined where *Cxcl12* was expressed during this time. Immunolabeling embryonic hearts with the artery endothelial cell marker Jagged1 revealed that Jagged+ vessels appeared in the septum around embryonic days (E)14.5 and 15.5 as a network of interconnected, immature artery precursors that were attached to both the right and left main branches (**Figure 5G** and **Supplementary Figure 8A-C**)[43]. (Note that Jagged1 also sparsely labeled trabecular cardiomyocytes.) This immature artery network is a typical early stage in artery development. Over the next two days, remodeling resulted in progressive development of a major artery within the septum **(Figure 5H).** Notably, most SpAs formed close to the right ventricular trabecular cardiomyocytes that line the chamber lumen—79% grew only alongside the right trabeculae and 18% alongside both while very few (3%) grew only near left trabeculae (**Supplementary Figure 8D, E,** and **F**). This pattern was also observed in neonatal WT hearts (**Figure 5I** and **Supplementary Figure 8I**).

Cxcl12 was expressed in multiple cell types, including the artery endothelial cells themselves and the right ventricle trabecular myocardium where we observed biased SpA formation (**Figure 5J, K,** and **Supplementary Figure 8F**)[25]. This expression in only one chamber's trabecular cardiomyocytes was also seen in human hearts, except the chambers were flipped between the species (i.e. human expression was specific to left and mouse to right) (compare **Figures 5J** with **Figure 4K** and **L**).

To test whether SpA placement was influenced by the localized expression in right ventricle trabeculae, we investigated whether SpA branching within the septum was altered in *Cxcl12* heterozygous neonatal mice. In *Cxcl12*$^{DsRed/+}$ mice, SpAs were shifted leftward away from the right ventricle trabeculae, and sometimes were only near the left (**Figure 5I** and **L**). These data support a role for increased *Cxcl12* levels in attracting SpAs during development.

Regarding dominance establishment, we frequently observed that one of the original connections, either to the left or right coronary artery, had grown larger than the other side during earlier developmental stages (**Supplementary Figure 8D**). We observed that mature SpAs without immature artery networks emerged at E16.5 and 17.5 in most hearts, suggesting the acquisition of left, right, or co-dominance (**Figure 5H** and **Supplementary Figure 8E** and **G**)[43]. In fact, when comparing SpA dominance at these stages (E16.5 and 17.5) between *Cxcl12*$^{+/+}$ and *Cxcl12*$^{DsRed/+}$ embryos, the distributions were similar to those quantified in neonates. Specifically, there was a trend away from right-dominance and towards co-dominance in *Cxcl12*$^{DsRed/+}$ embryos (compare **Supplementary Figure 8H** with **Figure 5F**). Together, these data show that *Cxcl12* expression is an attractive force dictating coronary artery placement, and that decreased levels result in altered coronary artery configurations.

Discussion

Here, we present a combined and stratified multi-ancestry GWAS of coronary artery dominance using genotypes from MVP participants who had undergone angiograms, including large numbers of participants of European, African, or Admixed American descent. State-of-the-art genotyping, imputation, and linear mixed-model analyses of genetic data enabled identification of ten loci with the strongest signal located downstream of *CXCL12*. Genomic analyses identified individual variants with the potential to alter transcriptional regulation. Whole-organ

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

immunofluorescence and spatial transcriptomics showed that dominance is apparent in humans during fetal development when *CXCL12* is expressed in the heart. Evaluation of heterozygous knockout mice demonstrated that *Cxcl12* expression levels influence mouse SpA dominance by imparting an attractive force on artery localization. Taken together, these data support the notion that *CXCL12* expression directly regulates patterning of human coronary dominance during development.

Prior studies in mouse and zebrafish models have demonstrated a role for *Cxcl12* in artery development. Initially discovered as a chemoattractant for hematopoietic cells[44,45], mouse knockouts deleting *Cxcl12* revealed a requirement for artery development[46]. Mechanistic studies in zebrafish showed that arteries form adjacent to cells expressing *Cxcl12*. In these locations, it acts as a paracrine signal attracting pre-artery ECs that express Cxcl12's receptor, Cxcr4. Upon activation, pre-artery cells migrate together to assemble into mature arteries or extend artery tips[21]. This same cellular mechanism functions during coronary artery development in mice with *Cxcl12* being expressed in the outflow tract, myocardium, and epicardium[13,22-24]. Mouse studies also demonstrated that *Cxcl12* selectively signals to pre-artery ECs, which are a minor subset of the ECs within the developing capillary plexus[25]. This causes pre-artery ECs within the network to specifically migrate against the direction of blood flow and join to extend arterial vessels, without compromising blood flow[12,47].

Animal models have also shown that *Cxcl12* stimulates coronary repair and regeneration. Neonatal mouse heart regeneration is enabled after myocardial infarction by capillary induction of *Cxcl12*, which signals to artery cells and guides their migration to a location where they can form natural collateral artery bypasses that support reinnervation[48-50]. In adult mouse hearts, which don't regenerate, a single localized injection of *CXCL12* stimulates revascularization myocardial recovery[48,51,52]. Similarly, *Cxcl12*-deficient zebrafish fail to fully regenerate their heart following apical resection due to deficient revascularization[22,53].

With this knowledge, we propose a model where *CXCL12* expression domains in human fetal hearts attract CXCR4+ pre-artery ECs to pattern artery branch growth. This activity could direct RCA development over the posterior left ventricle in most hearts. We also propose that variants associated with left dominance alter transcription factor binding and modestly decrease *CXCL12* transcription. This could stunt or redirect RCA growth so that the LCA eventually extends to compensate (**Figure 6**). In fact, experiments have shown artery compensation to be a prominent feature of the coronary vasculature[54,55]. Future work will aim to experimentally validate all the causal variants and identify the transcription factor activity they effect.

Our study stands out among other discovery GWASs in that it identified a susceptibility locus detected in both EUR and AFR populations. Identification of the *CXCL12* locus in both populations, despite EUR imbalance, underscores its importance in determining dominance patterns. A weakness here is the low number of female participants in MVP, although 35% of PMBB participants were female. We will at least in part correct this imbalance when angiogram data becomes available for other biobanks with higher female representation. However, we expect dominance patterning mechanisms to be the same in males and females since dominance ratios are conserved across sexes[56-58] and genetic data shows an overall genetic correlation of one between sexes for most human traits[59]. We expect to extend discovery in the future when genotyping of additional MVP and PMBB participants is made available, and as additional coronary artery evaluations are performed in participants without angiograms.

A somewhat unanticipated finding from the GWAS was the extensive overlap and colocalization of variants associated with dominance and CAD at the *CXCL12* locus. Whether

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

this finding represents a cascading effect of a single primary action of the gene (vertical pleiotropy) versus an independent effect on multiple traits (horizontal pleiotropy) remains to be determined[60]. In this context, we note that cardiologists have been fascinated by whether dominance impacts heart function or disease since the turn of the 20th century when it was first categorized in postmortem hearts[4]. Many studies have attempted to correlate dominance with cardiovascular events, but no consistent data has emerged[57,58,61-63]. While multiple reports suggest left dominance is associated with a poorer prognosis after a myocardial infarction or revascularization procedures[56-58,64-67], the clinical relationship between dominance and coronary plaque burden or incident myocardial infarction is less clear[68-70]. To answer these questions more definitively, large scale studies examining the relationship between dominance and an unbiased documentation of atherosclerotic burden need to be conducted, as well as robust and well-powered Mendelian Randomization studies[60,71].

In summary, our findings provide solid evidence that *CXCL12* patterns human coronary artery growth and branching. Given the ability of *CXCL12* to revascularize the heart through inducing collateral arteries in pre-clinical models, our data is a key piece of evidence increasing confidence that a therapeutic applying localized *CXCL12* could induce collateral arteries in humans. Our work represents an early, but critical, step toward developing a medical revascularization therapy that can one day obviate the need for open heart surgery to provide a coronary artery bypass.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Materials and Methods

Population genetic studies

The design of the MVP has been previously described[72]. Briefly, active users of the Veterans Health Administration (VA) of any age have been recruited from more than 60 VA Medical Centers nationwide since 2011 with current enrollment at >1,000,000. Informed consent is obtained from all participants to provide blood for genomic analysis and access to their full EHR within the VA prior to and after enrollment. We linked MVP participants to the Veterans Affairs Clinical Assessment, Reporting, and Tracking (CART) Program, a national quality and safety organization for invasive cardiac procedures, to identify participants who had undergone at least one coronary angiogram by June 2021[73]. Data were available retrospectively starting in 2004 in select sites and from all sites by 2010[74]. The study received ethical and study protocol approval from the VA Central Institutional Review Board.

Genetic Data and Quality Control

A total of 658,164 participants enrolled in MVP between 2011 and 2019 were genotyped with a customized Affymetrix Axiom array in three batches. A genotyping quality control (QC) procedure described in detail elsewhere[75] was applied to all three batches of data and further adapted to remove markers out of Hardy-Weinberg Equilibrium or with differential missingness between sexes. A total of 590,511 autosomal and 12,693 nonPAR X-chromosome markers passing QC were then phased using SHAPEIT4 v4.2.0[76]. Using Minimac4 (v1.0.2)[77], phased haplotypes were then imputed to a TOPMed panel (GRCh38 reference genome) that included ~194,000 phased haplotypes.

Assignment of coronary dominance

A unique feature of the MVP biobank is its linkage to the Veterans Health Administration (VA) Clinical Assessment, Reporting, and Tracking (CART) Program. This is a national quality and safety initiative that collects the results of all invasive cardiac procedures performed on Veterans within the nationally integrated VA healthcare system, including a structured variable defining coronary dominance. Coronary dominance is collected as a structured variable in the EHR template report used to document the results of coronary angiograms for the CART registry and is typically entered by a member of the team who performed the procedure. One of three dominance options can be selected including right, co-dominant, and left. We first excluded all coronary angiogram reports from individuals with a history of a heart transplant if they were performed on the day of or after the date of their cardiac transplant. For individuals with a single procedure, we assigned dominance based on that single procedure. Most individuals with more than one procedure in CART were fully concordant in their dominance assignments across all procedures. For the remaining, we required that at least 80% of the assignments were concordant to assign a dominance, otherwise the participant was excluded. A total of 61,044 MVP participants were found to have at least one coronary angiogram evaluation with a documentation of dominance.

Statistical Analysis

Genetically inferred ancestry assignment

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Population membership to all genotyped participants was assigned using a reference dataset of unrelated individuals from the 1000 Genomes Project (1KGP). This assignment was performed centrally as part of a core MVP project and made available as a core resource for all MVP investigators[78]. Specifically, the smartpca module in the EIGENSOFT package[79] was used to project the principal components (PCs) loadings from the 1KGP to genotyped MVP participants. To define the genetically similar population, we trained a random forest classifier using cross-population meta-data based on the top 10 PCs from the reference training data. We then used a random forest classifier trained on the predicted MVP PCA data to assign individuals to one of five 1000 Genomes major continental groups including White/Europeans (EUR), non-Hispanic Black/African American (AFR), Admixed American (AMR), East Asians (EAS), and South Asian (SAS). Individuals with random forest probability over 50% for a population were assigned to that population. Those who could not be assigned to a population were assigned to 'other'. PCs were then generated for all members within each of the five major continental groups.

Heritability analyses

We used GREML-LDMS as implemented in Genome-wide Complex Trait Analysis (GCTA) v1.94.1 Linux to estimate the multicomponent narrow sense heritability (h^2) of dominance. GREML-LDMS is one of the most accurate heritability estimation methods when considering MAF and LD structures that may bias such estimates[80]. Analyses were restricted to GIA groups with at least > 4,000 case subjects with non-right dominance when applying GREML-LDMS to binary traits to ensure an estimate of heritability with an acceptable standard error[81]. Restricted by computing memory requirements, we selected a random sample of right dominance subjects and all non-right dominance subjects at a ratio of 3:1 to run through GREML-LDMS[82,83]. First, SNPs with allele dosage information used for all GWAS were converted to hard-call genotypes using the default settings in PLINK v2.00a3LM. SNPs that were multi-allelic, had MAC < 3, or genotype call-rate < 95% were removed. Since dominance status is a binary trait, SNPs with P-value < 0.05 for Hardy-Weinberg equilibrium or differential missingness in cases (non-right) vs controls (right) were also removed[82,83]. LD scores were calculated separately for each autosome using default GCTA parameters with an r^2 cutoff of 0.01, and the genome-wide LD score distribution was utilized to allocate SNPs to 1 of 4 LD quartile groups, where groups 1-4 represented SNPs exhibiting increasing LD scores. Within each LD group, SNPs were further stratified into 6 MAF bins ([0.001, 0.01], [0.01, 0.1], [0.1, 0.2], [0.2, 0.3], [0.3, 0.4], [0.4, 0.5]) and a genetic relatedness matrix (GRM) was constructed from each bin, ultimately creating 24 GRMs. Finally, reml function in GCTA was used to fit a model of case status (non-right vs right) based on the 24 GRMs, with age and sex as covariates. Total observed heritability estimates were transformed to estimate disease liability across a range of presumed non-right dominance prevalence estimates in the general population including the observed prevalence in MVP.

Genome-wide association study in MVP

We used SAIGE v1.1.6.2[18] which uses the saddlepoint approximation (SPA) to calibrate unbalanced case-control ratios in score tests based on logistic mixed models and allows for model fitting with either full or sparse genetic relationship matrix (GRM) to conduct the coronary dominance GWASs. Each GWAS was adjusted for sex and 10 principal components where the PCs used for all subjects combined analysis were global and generated using genotype data and PLINK while the PCs used for the analyses stratified by major continental group were

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

generated within each population alone. All summary statistics underwent quality control using EasyQC[84]. Variants were excluded if minor allele count was less than or equal to 6, imputation quality was less than 0.7, or a variant was monomorphic. Sex chromosomes analyses were also excluded.

A locus was considered GWS if at least one lead genetic variant within it reached a $P < 5\times10^{-8}$ in any of the individual population GWAS or in the combined multi-population GWAS. Since no previous large-scale GWAS of coronary dominance has been reported, all loci reaching GWS were considered novel. We used Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA-GWAS)[76] to define genomic risk loci including independent, lead, and candidate variants. First, independent genetic variants were identified as variants with a *P* below a specific threshold and not in substantial linkage disequilibrium (LD) with each other ($r^2 < 0.6$). Second, variants in LD ($r^2 \geq 0.6$) with an independent variant and with $p < 0.05$ were retained as candidate variants to form an LD block. Third, LD blocks within 500kb of each other were merged into one locus. Lastly, a second clumping of the independent variants was performed to identify the subset of lead SNPs with LD $r^2 < 0.1$ within each locus. For our meta-analyses of Whites alone, we used a UK Biobank release 2b EUR reference panel of genotype data imputed to the UK10K/1000G SNPs created by FUMA including ~17 million SNPs. This panel includes a random subset of 10,000 unrelated subjects among all subjects with genotype data mapped to the 1000G populations based on the minimum Mahalanobis distance. We used the 1000G AFR reference panel of 661 subjects with ~43.7 million SNPs for our Black, and the AMR reference panel of 347 subjects with ~29.5 million SNPs for our Hispanics. We looked up select independently significant variants identified by FUMA in the GTExPortal to further characterize their expression[85]. Lastly, we annotated the lead SNP(s) at each novel locus by creating URL hyperlinks to five variant-base portals: OpenTargets, QTLbase, Common Metabolic Disease Knowledge, Open GWAS, and PhenoScanner.

Statistical fine mapping

Statistical fine-mapping was performed using SuSiEx, an accurate and computationally efficient method for cross-population fine-mapping using GWAS summary statistics, which builds on the single-population fine-mapping framework, Sum of Single Effects (SuSiE)[86]. SuSiEx improves the power and resolution of fine-mapping while producing well-calibrated false positive rates and retaining the ability to identify population specific causal variants[20]. We used the 1000 Genomes population phase 3 as reference panel to prioritize SNPs or sets of SNPs (credible set) driving each association signal, accounting for multiple causal variants in a genomic region. We calculated 95% credible sets for each identified signal representing the fewest number of variants whose posterior inclusion probabilities (PIP) for the signal summed to ≥ 0.95. We discarded credible sets in which the variants had a minimum $P > 1\times10^{-5}$ for the population group in which the signal was mapped.

Colocalization

We assessed the presence of colocalization of select genetic association signals for coronary dominance with another relevant trait using COLOC[37]. Evidence of colocalization at a locus with the same causal variant shared between two traits was defined as a posterior probability Bayesian factor H4 (PP.H4.abf) > 0.7 while evidence of colocalization with a different variant was defined as defined as a posterior probability Bayesian factor H3 (PP.H3.abf) > 0.7. We used LocusCompareR to visualize colocalization[87].

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Replication in the Penn Medicine Biobank

We used the multi-ancestry Penn Medicine Biobank (PMBB) to replicate select findings in the MVP[88]. Among 43,613 participants with available genomic data in PMBB, 3,141 had coronary angiography procedure data available for analysis. Coronary dominance was ascertained from coronary angiograms performed from 2007-2018 at the Hospital of the University of Pennsylvania (HUP) and Penn Presbyterian Medical Center (PPMC). Angiograms performed at PPMC have coronary dominance documented as a discrete field, which was directly extracted from the electronic health record. Coronary angiograms from HUP have coronary dominance documented as part of a semi-structured free-text report. To determine coronary dominance from HUP angiograms, natural language processing (NLP) was performed using the pre-trained "english-ewt" tree-bank model from the `udpipe` package in R[2]. Free-text coronary angiography reports underwent tokenization and lemmatization, and the adjectives (eg. "left", "right", "co-dom") referring to coronary dominance were then extracted using regular expressions. Assignment of coronary dominance was validated by manual review of 30 randomly sampled coronary angiography reports by a board-certified cardiologist. The NLP algorithm perfectly assigned coronary dominance, with positive- and negative-predictive values of 1. When multiple angiograms were available, we required 80% concordance across studies to assign a coronary dominance phenotype. Although most participants had a high degree of concordance across studies, individuals with <80% concordance across studies were excluded from further analysis. Of the 3,141 subjects, 2730 (87%) were classified by the NLP algorithm as right dominant, 152 (4.8%) as codominant, and 259 (8.2%) as left dominant. Genotyping, imputation, and quality control of PMBB to the TOPMed reference panel has been previously described[88]. GWAS of coronary dominance was performed using the same approach as in MVP adjusted for sex and five genetic principal components.

Deep learning models to score and interpret variant effects

Model Training

To model cell type-specific chromatin accessibility of human adult cardiac cell types, we trained ChromBPNet models (https://github.com/kundajelab/chrombpnet) on cell-type resolved pseudobulk ATAC-seq profiles. For each of the 87 single-nucleus ATAC-seq datasets in adult human heart from the ENCODE project (https://www.encodeproject.org/search/?searchTerm=snATAC-seq+in+heart+Homo+sapiens+adult+snyder), we subsetted the associated fragments file by cell type based on the ENCODE single cell project's barcode to cell type mapping. Next, we concatenated the subsetted fragment files across datasets and sorted them to create cell type-specific fragment files for each cell type present in the heart samples. We used the ENCODE peak calling pipeline to call peaks for each cell type based on the corresponding fragments file, and then used these peak and fragment files as inputs for ChromBPNet model training. The peaks and bigwigs used as input to the adult heart cell type ChromBPNet models as well as the ChromBPNet models themselves can be found on Synapse (https://www.synapse.org/Synapse:syn60940384/files/).

To train the ChromBPNet models, we used a 5-fold cross-validation scheme wherein each chromosome was present in the test set of at least one cross-validation fold. The table below shows which chromosomes were a part of the train, validation, and test splits in each fold.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Fold	Train Chromosomes	Validation Chromosomes	Test Chromosomes
Fold 0	Chr2, Chr4, Chr5, Chr7, Chr9, Chr10, Chr11, Chr12, Chr13, Chr14, Chr15, Chr16, Chr17, Chr18, Chr19, Chr21, Chr22, ChrX, ChrY	Chr8, Chr20	Chr1, Chr3, Chr6
Fold 1	Chr1, Chr3, Chr4, Chr5, Chr6, Chr7, Chr10, Chr11, Chr13, Chr14, Chr15, Chr18, Chr19, Chr20, Chr21, Chr22, ChrX, ChrY	Chr12, Chr17	Chr2, Chr8, Chr9, Chr16
Fold 2	Chr1, Chr2, Chr3, Chr5, Chr6, Chr8, Chr9, Chr10, Chr13, Chr14, Chr16, Chr17, Chr18, Chr19, Chr20, Chr21, ChrX	Chr7, Chr22	Chr4, Chr11, Chr12, Chr15, ChrY
Fold 3	Chr1, Chr2, Chr3, Chr4, Chr7, Chr8, Chr9, Chr11, Chr12, Chr13, Chr15, Chr16, Chr17, Chr19, ChrX, ChrY	Chr6, Chr21	Chr5, Chr10, Chr14, Chr18, Chr20, Chr22
Fold 4	Chr1, Chr2, Chr3, Chr4, Chr5, Chr6, Chr8, Chr9, Chr11, Chr12, Chr14, Chr15, Chr16, Chr20, Chr22, ChrY	Chr10, Chr18	Chr7, Chr13, Chr17, Chr19, Chr21, ChrX

We used the ChromBPNet framework to first train a Tn5 bias model on background regions (not overlapping peaks) from the cardiomyocyte pseudobulk ATAC-seq data. This bias model was used for bias correction while training ChromBPNet models on peaks and background regions from each of the cell types.

To model chromatin accessibility in fetal heart cell types, we trained ChromBPNet models on peak and fragment data from *Ameen et al.* (https://www.sciencedirect.com/science/article/pii/S0092867422015033). The peaks and bigwigs used as input to the fetal heart cell type ChromBPNet models as well as the ChromBPNet models themselves can be found on Synapse (https://www.synapse.org/Synapse:syn60940384/files/). The same 5-fold cross validation scheme was used for the fetal heart models. For Tn5 bias correction on the fetal models, a Tn5 bias model trained on a pseudobulk of ATAC-seq data from all cells (this is represented by the AllCells model). and schemes were used for the fetal heart models, as well.

For each of the fetal and adult heart cell types, we estimated base-resolution predictions of ATAC-seq coverage for each input sequence by averaging predictions from the 5 folds. ChromBPNet predicts two complementary outputs, the total coverage over the input sequence (log (total counts)) and the base-resolution positional probability profiles of reads across the sequence. Hence, we estimate the average total counts as the arithmetic mean of the exponentiated predicted log (total counts) over all 5 folds. We estimate the average of the logits

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

of the predicted profile probabilities per position over the 5 folds, followed by a softmax to obtain the aggregate probability profile. The coverage per position is simply the total coverage x the profile probability at that position.

We computed contribution scores of each base in any input sequence to predicted outputs (log(total counts) or profile probabilities) using the DeepSHAP (https://dl.acm.org/doi/10.5555/3295222.3295230) implementation of the DeepLIFT (https://proceedings.mlr.press/v70/shrikumar17a.html) algorithm, and then averaged scores across all 5 folds of the model, as well.

Variant Scoring

To score and identify dominance-associated variants predicted to affect chromatin accessibility, we used the ChromBPNet models for each cell type to predict base-resolution scATAC-seq coverage profiles for 1 kb genomic sequence windows containing reference and alternate alleles of all variants in the dominance GWAS loci. The effect size of each variant was estimated using two measures (1) the $\log_2$ fold change of the total predicted coverage (total counts) over each sequence window containing the reference and alternate allele (2) the Jensen Shannon distance (JSD) between the base-resolution predicted probability profiles for the ref and alt sequences (this measures a change in the profile shape).

Statistical significance was estimated for both scores on empirical null distributions of variant scores that were computed as follows. We shuffled the 2114 bp sequence centered at each variant multiple times while preserving the dinucleotide frequency, made 2 copies of each shuffled sequence, then inserted each allele of the variant at the center of the shuffled sequence to create a set of 1 million null variants in total. We scored each of these null variants with each model using the same procedure that we used to score each observed variant. Then, for each observed variant, we calculated the proportion of the null variants that had an equally high or higher (more extreme) score to generate an empirical P-value for both the log2 fold change and JSD scores.

This analysis produced a list of dominance-associated variants predicted to cause a statistically significant (absolute log2 fold change or JSD, false discovery rate < 0.01) change in allele-specific read counts. We defined a score called the active-allele-quantile, which is the quantile of the predicted read counts for the allele with the greatest predicted counts at each variant (the active allele), with respect to the distribution of predicted counts for all ATAC-seq peaks in the cell type where the model was trained. Using this score, we further filtered our prioritized variant list to those variants with an active-allele-quantile greater than 0.05, as these variants are likely to either reside in or create an accessible region and therefore are more likely to affect transcription factor binding. The code base for scoring variants is at https://github.com/kundajelab/variant-scorer

Animal Studies

All mouse colonies were housed and bred in the animal facility at Stanford University in accordance with institutional animal care and use committee (IACUC) guidance on 12□h/12□h□day and night cycle with water and food ad libitum.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

The following mouse strains were used: C57BL/6J, (Jackson Laboratory, strain code: 000664), CD1 (Charles River, strain 022) and *Cxcl12*^{DsRed} (Jackson Laboratory, strain code: 022458). All experiments were conducted in accordance with protocols approved by the Institutional Animal Care and Use (IACUC) Committee of Stanford University.
Cxcl12^{DsRed} mice were maintained on a C57BL/6J background. To assess SpA dominance, *Cxcl12*$^{DsRed/+}$ males were mated with wild type C57BL/6J females to produce litters with wild type (*Cxcl12*$^{+/+}$) and heterozygous (*Cxcl12*$^{DsRed/+}$) mice. All neonatal studies used mice between postnatal days (P)0-P6. For all embryonic studies, timed pregnancies were determined by defining the day on which a vaginal plug was found as E0.5.

Assessing SpA dominance

Dominance was assigned from whole-organ images of coronary arteries imaged using light sheet microscopy and obtained by one of two experimental methods, whole-organ immunofluorescence or *in vivo* labeling (see details for each below).
SpA dominance was determined in 3D using Imaris. First, the ventricular septum was identified using the Oblique & Ortho Slicer and Scissors Tools. Then, it was confirmed that the septum was principally irrigated by one or two SpAs. Lastly, using the Oblique & Ortho Slicer and Scissors tools, the SpA was followed until it merged with the most proximal RCA or LCA segments. On rare occasions, the SpA(s) were followed to an independent connection point (ostium) on the aorta. Based on these observations, SpA dominance was scored as right-, left-, or co-dominant, or as aorta. When scoring SpA dominance in *Cxcl12* mice, researcher was blinded to genotype.

Whole-organ immunolabeling

Whole heart immunofluorescence was performed following the modified iDISCO+ protocol previously described[39,89]. For all following steps, tissue was always agitated unless noted otherwise. Briefly, animals were perfused with PBS through the dorsal vein, and fixed in 4% paraformaldehyde (PFA)(Electron Microscopy Science 15714) at 4°C for 1hr (embryonic and neonatal hearts) or 2hr (adult hearts), washed 3X in PBS and stored in PBS with 0.01% sodium azide (w/v, Sigma-Aldrich S8032) until ready to process. Hearts were dehydrated in increasing series of methanol/ddH-$_2$O dilutions (20%, 40%, 60%, 80%,100% 2X) for 1hr each, followed by overnight incubation in 66% dichloromethane (DCM, Sigma-Aldrich 34856) and 33% methanol. Next, tissue was washed 2X in 100% methanol for 4hrs and bleached overnight at 4°C in 5% hydrogen peroxide (Sigma-Aldrich 216763) in methanol. Next, the hearts are rehydrated in methanol/ddH-$_2$O dilutions (80%, 60%, 40%, 20%) for 1hr each, followed by PBS, 0.2% Triton X-100 PBS (2X) and incubated in 20% dimethyl sulfoxide (DMSO), 2.3% Glycine (w/v, Sigma G7126), and 0.2% Triton X-100 PBS at 37°C for 2 days.

For immunofluorescence, hearts were blocked in 10% DMSO, 6% Normal Donkey Serum (NDS, Jackson ImmunoResearch 017-000-121) in 0.2% Triton X-100 for 2 days at 37C. Primary antibodies were prepared in PBS with 5%DMSO, 3% NDS in 0.2% Tween-20 and 0.1% Heparin (w/v, Sigma-Aldrich H3393) and incubated at 37°C for 4-14 days. Antibodies included Cy3 conjugated-αSMA-Cy3 (1:300, Sigma C6198), Podocalyxin (1:1000, R&D Systems MAB1556), RFP (1:1000, Rockland 600-401-379), Jagged1 (Novus, AF599). Secondary antibodies conjugated to either Alexa 555 or 647 (Jackson ImmunoResearch) were matched 1:1 in concentration to their primary target and in prepared in PBS with 3% NDS in 0.2% Tween-20, 0.1% Heparin for the same primary incubation days at 37°C. Washes after each antibody

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

incubation in PBS with 0.2% Tween-20, 0.1% in Heparin were performed in 30min increment until the end of the day, followed by an overnight wash. Before clearing, samples were embedded in 1% low-melting agarose (Sigma-Aldrich A9414) in PBS and dehydrated in methanol/ddH-$_2$O dilutions (20%, 40%, 60%, 80%,100% 2X) for 1hr each and 100% overnight. Next, hearts were incubated in 66% DCM and 33% Methanol for 2.5hrs, followed 2X 30min 100% DCM. Finally, samples were cleared in ethyl cinnamate (ECi, Sigma Aldrich 112372), manually inverted a few times, and kept at RT in the dark until imaging.

In vivo vascular labeling

Neonate pups were gently wrapped in gauze and cooled on ice for 6 minutes to induce hypothermic circulatory arrest. Mice received an intravenous 10 µl retro-orbital injection of Isolectin GS-IB$_4$ DyLight™ 649 (1:5, Invitrogen, I32450, in sterile 1X PBS). Neonates were then allowed to recover at 37°C on a warm plate and, when conscious, returned to their mother's care for 30 min before euthanasia. Hearts were then dissected and fixed in 4% PFA (Electron Microscopy Science, 15714) at 4°C for 1 hour. Subsequently, hearts were washed three times in 1X PBS and stored in cold PBS with 0.01% sodium azide (w/v, Sigma-Aldrich, S8032) and covered from light until ready to process.

Hearts were then cleared following the protocol previously described[90]. For all following steps, tissue was always agitated unless noted otherwise. Before clearing, samples were embedded in 1% low-melting agarose (Sigma-Aldrich, A9414) in PBS and dehydrated in methanol/ddH-$_2$O dilutions (20%, 40%, 60%, 80%,100% 2X) for 1hr each and 100% overnight. Next, hearts were incubated in 66% DCM and 33% Methanol overnight, followed next day by 2X 100% DCM washes the next day. Finally, samples were cleared in ethyl cinnamate (ECi, Sigma Aldrich, 112372), manually inverted a few times, and kept at RT in the dark until and after imaging.

Light sheet imaging

Samples were imaged with Imspector Pro 7.0.98 software and LaVision BioTec Ultramicroscope II Light sheet microscope in a quartz cuvette filled with ECi. For imaging, we used a MVX10 zoom body (Olympus) with a 2x objective (pixel size of 3.25 µm / x,y) at magnification from 0.63x up to 1.6x. Up to 1400 images were taken for each heart using a z-step size of 3.5µm z step size, and light sheet numerical aperture to 0.111 NA. Band-pass emission filters (mean nm / spread) were used, depending on the excited fluorophores: 525/50 for autofluorescence; 595/40 for Cy3; and 680/30 for AF647. Exposure time was 10ms for single channel and 25ms for multichannel acquisition.

Image processing and visualization

Maximum projections were performed using NIH Fiji ImageJ 1.53 software. Representative images in figures were in some cases processed with Despeckle and/or Unsharp Mask features. Raw images are included in Supplement. Imaris 9.5.0 software (Bitplane) was used for 3D rendering. Pixel dimensions were updated from the non-reduced 16-bit image metadata. The Filament Object Tracer module in Imaris was used to generate a simple 3D outline of the main coronary arteries. A root point was manually placed where the RCA and LCA originate from the aorta, i.e. right and left ostium, and filaments were then created and drawn until the farthest length possible using the Autopath and AutoDepth features.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Quantitative PCR

Whole hearts were collected from E16.5 mice and lysed in 1 mL of TRIzol (Invitrogen 15596026) in tubes containing Lysing Matrix G (MP Biomedicals: 116916050) with homogenization (3X 30 sec at 30 Hz) in a Qiagen TissueLyser II followed by isolation of the aqueous phase according to the TRIzol reagent user guide. RNA was then extracted and isolated with a NucleoSpin RNA isolation kit (Takara 740955.50) according to the manufacturers' protocols. 100 ng of total RNA was then reverse transcribed with the SensiFAST cDNA synthesis kit (Meridian Bioscience BIO-65053). Total cDNA was then diluted 1:10 and qPCR was performed with the SensiFAST SYBR® No-ROX Kit (Meridian Bioscience BIO-98005) with 10 µL of reactants in a 384 well plate (ThermoFisher 4483285).

Primer pairs were designed such that they span exon/exon junctions to preclude amplification of genomic DNA and were specific for murine *Cxcl12*. Each well of the plate received 5µL of SensiFAST SYBR® No-ROX Mix, .04 µL of 100mM forward primer, .04 µL of 100mM reverse primer, 2.92µL of water and 2µL of cDNA. After plate preparation with gene specific primers combined in a well with a sample specific cDNA the plates were sealed with MicroAmp™ Optical Adhesive Film (ThermoFisher 4311971) and centrifuged for 5 min at 350g. Plates were run on a Quantstudio Flex 7 (ThermoFisher) with the following parameters: Initial dissociation (2 min at 50°C followed by 10 min at 95°C), 40x PCR stage with SYBR detection (15 sec 95°C, 1 min 60°C) with the final stage being a dissociation step to generate a melt curve for the PCR products to ensure a single product was amplified. During analysis the linear phase of the fluorescence was used to generate Ct (cycle threshold) values. The ddCt (delta-delta Ct) method was used to analyze the results. *Gapdh*, a housekeeping gene, was used as an internal control to normalize between samples which were then compared to the expression of the wild type to quantify *Cxcl12* expression between genotypes. For each well there were 3 technical replicates as well as a water negative control.

Embryonic heart analysis

Cxcl12$^{DsRed/+}$ males were mated with wild type C57BL/6J females to produce litters with wild type (*Cxcl12*$^{+/+}$) and heterozygous (*Cxcl12*$^{DsRed/+}$) embryos. Observance of a copulation plug was considered embryonic day 0.5, and litters were harvested at five embryonic stages (E13.5, 14.5, 15.5, 16.5, and 17.5). At each stage, embryos were fixed in 4% PFA for one hour at 4 degrees and washed in PBS before hearts were dissected. Hearts were next subjected to the whole-organ immunolabeling procedure described above for Jagged1 (arteries) and RFP (Cxcl12 reporter). Whole hearts were then imaged using confocal microscopy and then scored for the SpA anatomy categories reported in **Supplementary Figure 6** (artery network, mature SpA, proximity to ventricular lumen, SpA dominance). For scoring procedures, researchers were blinded to genotype. SpA lengthening in the interventricular septum was scored based on the SpA(s) proximity to the right, left, or both ventricular lumens. In E16.5-17.5 hearts, dominance was determined by tracing mature Jagged1 positive SpA(s) from their origin at the right, left, or both coronary arteries until their terminal end.

Cxcl12 expression analysis

Equal ROIs were placed at the border of each ventricular trabeculae and septum. The area covered by *Cxcl12* reporter signal was measured using Max Entropy threshold option in ImageJ Fiji, then normalized to the total area of ROI.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Statistical analysis

Statistical analysis and graphs were generated using Prism 9 (GraphPad). Graphs represent mean values obtained from multiple experiments and error bars represent standard deviation. For qPCR, two-tailed unpaired Student's t test was used to compare groups within an experiment and the level of significance were assigned to statistics in accordance with their p values (0.05 flagged as ∗, 0.01 flagged as ∗∗, less than 0.001 flagged as ∗∗∗, less than 0.0001 flagged as ∗∗∗∗). When analyzing SpA dominance in *Cxcl12* mice, a two-tailed Chi-square without Yates-correction test was performed in a 2x2 contingency table (right & non-right x wild type & heterozygotes).

Human Fetal Heart studies

Under IRB approved protocols, human fetal hearts were obtained for developmental analysis[91]. Gestational age was determined by standard dating criteria by last menstrual period and ultrasound[92]. Pregnancies complicated by multiple gestations and known fetal or chromosomal anomalies were excluded. Tissues were processed within 1hr following procedure at which time it was rinsed with cold, sterile PBS. For whole-organ immunolabeling, tissue was fixed in 4% PFA for 4hrs at 4°C before whole-organ immunolabeling. For *in situ* hybridization, hearts were fixed in 4% PFA for 24–48 hr at 4°C, followed by three 15 min washes in PBS.

Dominance analysis at fetal stages

3D reconstructions of fetal hearts subjected to whole-organ immunolabeling were performed by the method mentioned above under the heading Image processing and visualization. Briefly, root points were manually placed at the right and left ostia and filaments were then created and drawn until the farthest length possible using the Autopath and AutoDepth features in the Filament Object Tracer module in Imaris. Arterial landmarks and large diameter vessels were prioritized and traced first, followed by vessels extending into the septum on the inferior side of the heart. Right dominance was assigned when the filament trace of the PDA was connected to the RCA filament trace. Additionally, when looking at a central cross-section, the arteries dominating the volume were traced from the RCA and its ostium. Co-dominance was assigned when, looking at a central cross-section, arteries in this volume were equally occupied by those traced from the RCA and LCA to their ostia.

In situ hybridization

Hearts were sequentially dehydrated in 30%, 50%, 70%, 80%, 90%, and 100% ethanol, washed 3X for 30 min in xylene, washed several times in paraffin, and finally embedded in paraffin which was allowed to harden into a block. For each heart, the whole ventricle was cut into 10-µm-thick sections and captured on glass slides.

CXCL12 mRNA and α-SMA protein were simultaneously localized on sections using RNAscope® per manufacturers' instructions. Briefly, slides were baked in 60°C for 1hr in a dry oven. Next, slides were incubated 2X in xylene with slight agitation for 5 min in RT under fume hood. Immediately after, slides were incubated 2X in 100% ethanol for 2 min under the same conditions. Slides were then dried for 5 min in the drying oven at 60°C. Afterwards, ~5 drops of hydrogen peroxide were added to the deparaffinized slides and incubated at RT for 10 min. Slides were washed 2X with fresh distilled water. In a food steamer, target retrieval was

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

performed for 15 minutes with RNAscope Retrieval Buffer. The tissue was pretreated in mild conditions, with RNAscope Protease III for 15 min. *In situ* hybridization was continued with the RNAscope® Multiplex Fluorescent Reagent Kit v2 (ACD 323100) using the following probes: human *CXCL12* (ACD 422991), positive control (ACD 320861), and negative control (ACD 320871). The ACD RNAscope protocol was followed in Channel 1 for all probes with Opal 650 fluorophore (Akoya FP1496001KT). Prior to mounting, slides were fixed in 4% PFA for 30 min, followed by 3X 15 min washes in PBS-0.1% Tween. The slides were then immunostained with Cy3 conjugated-α-SMA-Cy3 (1:300, Sigma C6198) for 1hr at RT. Then, slides were washed for 10 min 3X in PBS-0.1% Tween. To remove blood cells and reduce autofluorescence, slides were washed in $CuSO_4$ for 30 min at RT in a nutating mixer. Slides were then washed with distilled water before adding DAPI and mounting in ProLong Gold Antifade reagent (Invitrogen P36934). Finally, slides were imaged within a week on a confocal microscope (Zeiss LSM 980).

Single cell and spatial transcriptomics

Detailed MERFISH and scRNAseq methods are reported elsewhere[41]. Briefly, heart samples were collected at the University of California, San Diego (UCSD) in strict observance of the legal and institutional ethical regulations under IRB-approved protocols.

For single-cell dissociation, tissue samples from eight hearts were enzymatically digested and subjected to FACS before before processing for scRNA-seq. Single-cell droplet libraries were prepared according to the manufacturer's instructions (10X Genomics) and sequenced using a HiSeq 4000 (Illumina). Raw sequencing reads were processed using the Cell Ranger v3.0.1 pipeline and processed using Seurat v4.0.1. Following standard quality controls, gene expression was normalized for genes expressed per cell and total expression by the NormalizeData function. A nearest neighbor graph was generated and used for graph-based, semi-unsupervised clustering and uniform manifold approximation and projection (UMAP) to project the cells into two dimensions. Cell identities were assigned to the clusters by cross-referencing their marker genes with known cardiac cell type markers from both human and mouse studies.

For MERFISH, the sample was embedded in OCT and sectioned such that all major cardiac structures were captured. MERFISH measurements were taken for 238 genes with 10 non-targeting blank controls using published methods[93,94]. The 238 genes were selected to spatially localize the 75 cell subpopulations present in the above-described scRNAseq dataset. This list was identified using a combination of differentially expressed genes in the 75 cell populations and manual selection of previously characterized cardiac cell type markers[41].

To integrate the scRNA-seq and MERFISH datasets, the 238 MERFISH genes were utilized with Scanpy's implementation of Harmony to project both datasets into a shared PCA space before visualization with UMAP. To impute a complete expression profile and cell label for each MERFISH profile, expression profiles and cell labels of the closest scRNA-seq cell were assigned to the MERFISH cells in the Harmony PCA space using the Euclidean distance metric. Performance was evaluated as described[41].

URLs

GCTA, http://cnsgenomics.com/software/gcta/#Overview;
SAIGE, https://saigegit.github.io/SAIGE-doc/

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

PLINK: https://www.cog-genomics.org/plink/;
EasyQC, https://www.uni-regensburg.de/medizin/epidemiologie-praeventivmedizin/genetische-epidemiologie/software/;
R statistical software, www.R-project.org;
FUMA, http://fuma.ctglab.nl/;
SuSiEx, https://github.com/getian107/SuSiEx
LocusZoom, https://my.locuszoom.org/
GTExPortal, https://www.gtexportal.org/home/
OpenTargets, https://genetics.opentargets.org/;
QTLbase, http://www.mulinlab.org/qtlbase;
Common Metabolic Disease Knowledge Portal, https://hugeamp.org/;
Open GWAS, https://gwas.mrcieu.ac.uk/;
PhenoScanner, http://www.phenoscanner.medschl.cam.ac.uk/;
CADD, https://cadd.gs.washington.edu/
COLOC, https://chr1swallace.github.io/coloc/articles/a01_intro.html
LDSC, https://github.com/bulik/ldsc;
CARDIoGRAMplusC4D, http://www.cardiogramplusc4d.org;

Acknowledgements

Support for title page creation and format was provided by AuthorArranger, a tool developed at the National Cancer Institute. 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/figures used for the analyses described in this manuscript were obtained from the GTExPortal on 10/18/21.

Consortia

The VA Million Veteran Program. Members of the consortium can be found in the Supplementary Text File.

Funding:

This research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, and was supported by Veterans Administration awards I01-01BX003362. The content of this manuscript does not represent the views of the Department of Veterans Affairs or the United States Government.

P.E.R.C was supported by NIH-TIMBS-T32HL094274.
E.N.F was supported by NIH-T32HL007444.
J.A.N was supported by NIH-T32GM007276.
S.L.C was supported by NIH-CTSA KL2TR003143.
K.R-H was supported by NIH-R01HL12850307 and the Howard Hughes Medical Institute.
M.G.L was supported by the Doris Duke Foundation award #2023-0224

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

The Penn Medicine Biobank is supported by the Perelman School of Medicine at University of Pennsylvania, a gift from the Smilow family, and the National Center for Advancing Translational Sciences of the National Institutes of Health under CTSA award number UL1TR001878.

Author Contributions

Concept and design: P.E.R.C., D.Z., S.L.C, K.R-H., T.L.A.

Acquisition, analysis, or interpretation of data: P.E.R.C., D.Z., J.Z., J.A.N, P.P., A.M.M., E.N.F., X.F., S.K., S.S.D., I.E., P.F.K., A.T.H., S.A, S.P., D.D., S.M.D., K-M.C., M.G.L., V.D.W., A.M.P., M.E.P., S.W.W., P.S.T, A.K., N.C.C., S.L.C, K.R-H., T.L.A.

Drafting of the manuscript: P.E.R.C, D.Z., J.Z., S.L.C., K.R-H., T.L.A.

Critical revision of the manuscript for important intellectual content: P.E.R.C, D.Z., M.G.L, S.L.C., K.R-H., T.L.A.

Competing interests: A.K. is on the scientific advisory board of SerImmune, TensorBio and OpenTargets, and a consultant with Arcadia Science and Inari Agriculture. A.K. was a scientific co-founder of RavelBio, a paid consultant with Illumina, was on the SAB of PatchBio and owns shares in DeepGenomics, Immunai, Freenome, and Illumina**.** All other authors declare they have no competing interests.

Data and materials availability: The full summary level association statistics from each genetically inferred ancestry population association analyses in MVP from this report will be available through dbGaP, with accession number phs001672 at the time of publication in a peer reviewed journal. All other data are available in the main text or the supplementary materials.

Supplementary Materials

Figs. S1 to S8

Supplementary Text

Tables S1 to S8

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

References

1. Baroldi, G., Scomazzoni, G., and United States. Surgeon-General's, O. (1967). Coronary circulation in the normal and the pathologic heart (Office of the Surgeon General, Dept. of the Army; for sale by the Superintendent of Documents, U.S. Govt. Print. Off.).
2. Libby, P., Bonow, R.O., Mann, D.L., Tomaselli, G.F., Bhattt, D.L., and Solomon, S. (2021). Braunwald's heart disease : a textbook of cardiovascular medicine, 12. Edition (Elsevier).
3. Martin, S.S., Aday, A.W., Almarzooq, Z.I., Anderson, C.A.M., Arora, P., Avery, C.L., Baker-Smith, C.M., Barone Gibbs, B., Beaton, A.Z., Boehme, A.K., et al. (2024). 2024 Heart Disease and Stroke Statistics: A Report of US and Global Data From the American Heart Association. Circulation *149*, e347-e913. 10.1161/CIR.0000000000001209.
4. Thiene, G., Frescura, C., Padalino, M., Basso, C., and Rizzo, S. (2021). Coronary Arteries: Normal Anatomy With Historical Notes and Embryology of Main Stems. Front Cardiovasc Med *8*, 649855. 10.3389/fcvm.2021.649855.
5. Bruell, J.H., Leininger, J.T., and Hellerstein, H.K. (1970). Inheritance of a cardiac arterial asymmetry in mice. Science (New York, N.Y *167*, 199-200. 10.1126/science.167.3915.199.
6. Fernandez, B., Duran, A.C., Fernandez, M.C., Fernandez-Gallego, T., Icardo, J.M., and Sans-Coma, V. (2008). The coronary arteries of the C57BL/6 mouse strains: implications for comparison with mutant models. J Anat *212*, 12-18. 10.1111/j.1469-7580.2007.00838.x.
7. Lopez-Garcia, A., Soto-Navarrete, M.T., Fernandez, M.C., Moncayo-Arlandi, J., Duran, A.C., Lopez-Unzu, M.A., Alonso-Briales, J.H., and Fernandez, B. (2016). Unusual anatomical origins of the coronary arteries in C57BL/6 mice. Are they strain-specific? J Anat *229*, 703-709. 10.1111/joa.12512.
8. Red-Horse, K., and Siekmann, A.F. (2019). Veins and Arteries Build Hierarchical Branching Patterns Differently: Bottom-Up versus Top-Down. BioEssays : news and reviews in molecular, cellular and developmental biology *41*, e1800198. 10.1002/bies.201800198.
9. Trimm, E., and Red-Horse, K. (2022). Vascular endothelial cell development and diversity. Nat Rev Cardiol, 1-14. 10.1038/s41569-022-00770-1.
10. Lee, H.W., Shin, J.H., and Simons, M. (2022). Flow goes forward and cells step backward: endothelial migration. Exp Mol Med *54*, 711-719. 10.1038/s12276-022-00785-1.
11. Xu, C., Hasan, S.S., Schmidt, I., Rocha, S.F., Pitulescu, M.E., Bussmann, J., Meyen, D., Raz, E., Adams, R.H., and Siekmann, A.F. (2014). Arteries are formed by vein-derived endothelial tip cells. Nature communications *5*, 5758. 10.1038/ncomms6758.
12. Pitulescu, M.E., Schmidt, I., Giaimo, B.D., Antoine, T., Berkenfeld, F., Ferrante, F., Park, H., Ehling, M., Biljes, D., Rocha, S.F., et al. (2017). Dll4 and Notch signalling couples sprouting angiogenesis and artery formation. Nature cell biology *19*, 915-927. 10.1038/ncb3555.
13. Su, T., Stanley, G., Sinha, R., D'Amato, G., Das, S., Rhee, S., Chang, A.H., Poduri, A., Raftrey, B., Dinh, T.T., et al. (2018). Single-cell analysis of early progenitor cells that build coronary arteries. Nature *559*, 356-362. 10.1038/s41586-018-0288-7.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

14. Geudens, I., Coxam, B., Alt, S., Gebala, V., Vion, A.C., Meier, K., Rosa, A., and Gerhardt, H. (2019). Artery-vein specification in the zebrafish trunk is pre-patterned by heterogeneous Notch activity and balanced by flow-mediated fine-tuning. Development *146*. 10.1242/dev.181024.
15. Volz, K.S., Jacobs, A.H., Chen, H.I., Poduri, A., McKay, A.S., Riordan, D.P., Kofler, N., Kitajewski, J., Weissman, I., and Red-Horse, K. (2015). Pericytes are progenitors for coronary artery smooth muscle. eLife *4*. 10.7554/eLife.10036.
16. Siekmann, A.F. (2023). Biology of vascular mural cells. Development *150*. 10.1242/dev.200271.
17. Samani, N.J., Erdmann, J., Hall, A.S., Hengstenberg, C., Mangino, M., Mayer, B., Dixon, R.J., Meitinger, T., Braund, P., Wichmann, H.E., et al. (2007). Genomewide association analysis of coronary artery disease. The New England journal of medicine *357*, 443-453. 10.1056/NEJMoa072366.
18. Zhou, W., Nielsen, J.B., Fritsche, L.G., Dey, R., Gabrielsen, M.E., Wolford, B.N., LeFaive, J., VandeHaar, P., Gagliano, S.A., Gifford, A., et al. (2018). Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nature genetics *50*, 1335-1341. 10.1038/s41588-018-0184-y.
19. Watanabe, K., Taskesen, E., van Bochoven, A., and Posthuma, D. (2017). Functional mapping and annotation of genetic associations with FUMA. Nature communications *8*, 1826. 10.1038/s41467-017-01261-5.
20. Yuan, K., Longchamps, R.J., Pardinas, A.F., Yu, M., Chen, T.T., Lin, S.C., Chen, Y., Lam, M., Liu, R., Xia, Y., et al. (2023). Fine-mapping across diverse ancestries drives the discovery of putative causal variants underlying human complex traits and diseases. medRxiv, 2023.2001.2007.23284293. 10.1101/2023.01.07.23284293.
21. Siekmann, A.F., Standley, C., Fogarty, K.E., Wolfe, S.A., and Lawson, N.D. (2009). Chemokine signaling guides regional patterning of the first embryonic artery. Genes & development *23*, 2272-2277. 10.1101/gad.1813509.
22. Harrison, M.R., Bussmann, J., Huang, Y., Zhao, L., Osorio, A., Burns, C.G., Burns, C.E., Sucov, H.M., Siekmann, A.F., and Lien, C.L. (2015). Chemokine-guided angiogenesis directs coronary vasculature formation in zebrafish. Developmental cell *33*, 442-454. 10.1016/j.devcel.2015.04.001.
23. Cavallero, S., Shen, H., Yi, C., Lien, C.L., Kumar, S.R., and Sucov, H.M. (2015). CXCL12 Signaling Is Essential for Maturation of the Ventricular Coronary Endothelial Plexus and Establishment of Functional Coronary Circulation. Developmental cell *33*, 469-477. 10.1016/j.devcel.2015.03.018.
24. Ivins, S., Chappell, J., Vernay, B., Suntharalingham, J., Martineau, A., Mohun, T.J., and Scambler, P.J. (2015). The CXCL12/CXCR4 Axis Plays a Critical Role in Coronary Artery Development. Developmental cell *33*, 455-468. 10.1016/j.devcel.2015.03.026.
25. D'Amato, G., Phansalkar, R., Naftaly, J.A., Fan, X., Amir, Z.A., Rios Coronado, P.E., Cowley, D.O., Quinn, K.E., Sharma, B., Caron, K.M., et al. (2022). Endocardium-to-coronary artery differentiation during heart development and regeneration involves sequential roles of Bmp2 and Cxcl12/Cxcr4. Developmental cell *57*, 2517-2532 e2516. 10.1016/j.devcel.2022.10.007.
26. Nornes, S., Bruche, S., Adak, N., McCracken, I., and Val, S.D. (2024). A catalogue of verified and characterized arterial enhancers for key arterial identity genes. bioRxiv, 2024.2004.2030.591717. 10.1101/2024.04.30.591717.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

27. Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nature genetics *46*, 310-315. 10.1038/ng.2892.
28. Rentzsch, P., Witten, D., Cooper, G.M., Shendure, J., and Kircher, M. (2019). CADD: predicting the deleteriousness of variants throughout the human genome. Nucleic Acids Res *47*, D886-D894. 10.1093/nar/gky1016.
29. Zhou, J., and Troyanskaya, O.G. (2015). Predicting effects of noncoding variants with deep learning-based sequence model. Nat Methods *12*, 931-934. 10.1038/nmeth.3547.
30. Kelley, D.R., Snoek, J., and Rinn, J.L. (2016). Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks. Genome Res *26*, 990-999. 10.1101/gr.200535.115.
31. Trevino, A.E., Muller, F., Andersen, J., Sundaram, L., Kathiria, A., Shcherbina, A., Farh, K., Chang, H.Y., Pasca, A.M., Kundaje, A., et al. (2021). Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. Cell *184*, 5053-5069 e5023. 10.1016/j.cell.2021.07.039.
32. Wang, S.K., Nair, S., Li, R., Kraft, K., Pampari, A., Patel, A., Kang, J.B., Luong, C., Kundaje, A., and Chang, H.Y. (2022). Single-cell multiome of the human retina and deep learning nominate causal variants in complex eye diseases. Cell Genom *2*. 10.1016/j.xgen.2022.100164.
33. Ameen, M., Sundaram, L., Shen, M., Banerjee, A., Kundu, S., Nair, S., Shcherbina, A., Gu, M., Wilson, K.D., Varadarajan, A., et al. (2022). Integrative single-cell analysis of cardiogenesis identifies developmental trajectories and non-coding mutations in congenital heart disease. Cell *185*, 4937-4953 e4923. 10.1016/j.cell.2022.11.028.
34. Fan, Y., Lu, H., Liang, W., Hu, W., Zhang, J., and Chen, Y.E. (2017). Kruppel-like factors and vascular wall homeostasis. J Mol Cell Biol *9*, 352-363. 10.1093/jmcb/mjx037.
35. Miao, Y., Tian, L., Martin, M., Paige, S.L., Galdos, F.X., Li, J., Klein, A., Zhang, H., Ma, N., Wei, Y., et al. (2020). Intrinsic Endocardial Defects Contribute to Hypoplastic Left Heart Syndrome. Cell Stem Cell *27*, 574-589 e578. 10.1016/j.stem.2020.07.015.
36. Aragam, K.G., Jiang, T., Goel, A., Kanoni, S., Wolford, B.N., Atri, D.S., Weeks, E.M., Wang, M., Hindy, G., Zhou, W., et al. (2022). Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. Nature genetics *54*, 1803-1815. 10.1038/s41588-022-01233-6.
37. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet *10*, e1004383. 10.1371/journal.pgen.1004383.
38. Renier, N., Wu, Z., Simon, D.J., Yang, J., Ariel, P., and Tessier-Lavigne, M. (2014). iDISCO: a simple, rapid method to immunolabel large tissue samples for volume imaging. Cell *159*, 896-910. 10.1016/j.cell.2014.10.010.
39. Rios Coronado, P.E., and Red-Horse, K. (2021). Enhancing cardiovascular research with whole-organ imaging. Curr Opin Hematol *28*, 214-220. 10.1097/MOH.0000000000000655.
40. Moffitt, J.R., Bambah-Mukku, D., Eichhorn, S.W., Vaughn, E., Shekhar, K., Perez, J.D., Rubinstein, N.D., Hao, J., Regev, A., Dulac, C., and Zhuang, X. (2018). Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. Science (New York, N.Y *362*. 10.1126/science.aau5324.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

41. Farah, E.N., Hu, R.K., Kern, C., Zhang, Q., Lu, T., Ma, Q., Tran, S., Zhang, B., Carlin, D., Monell, A., et al. (2024). Spatially-Organized Cellular Communities Form the Developing Human Heart. Nature.
42. Ding, L., and Morrison, S.J. (2013). Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. Nature *495*, 231-235. 10.1038/nature11885.
43. Kolesova, H., Bartos, M., Hsieh, W.C., Olejnickova, V., and Sedmera, D. (2018). Novel approaches to study coronary vasculature development in mice. Dev Dyn *247*, 1018-1027. 10.1002/dvdy.24637.
44. Nagasawa, T., Hirota, S., Tachibana, K., Takakura, N., Nishikawa, S., Kitamura, Y., Yoshida, N., Kikutani, H., and Kishimoto, T. (1996). Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. Nature *382*, 635-638. 10.1038/382635a0.
45. Nagasawa, T. (2000). A chemokine, SDF-1/PBSF, and its receptor, CXC chemokine receptor 4, as mediators of hematopoiesis. Int J Hematol *72*, 408-411.
46. Ara, T., Tokoyoda, K., Okamoto, R., Koni, P.A., and Nagasawa, T. (2005). The role of CXCL12 in the organ-specific process of artery formation. Blood *105*, 3155-3161. 10.1182/blood-2004-07-2563.
47. Chang, A.H., Raftrey, B.C., D'Amato, G., Surya, V.N., Poduri, A., Chen, H.I., Goldstone, A.B., Woo, J., Fuller, G.G., Dunn, A.R., and Red-Horse, K. (2017). DACH1 stimulates shear stress-guided endothelial cell migration and coronary artery growth through the CXCL12-CXCR4 signaling axis. Genes & development *31*, 1308-1324. 10.1101/gad.301549.117.
48. Das, S., Goldstone, A.B., Wang, H., Farry, J., D'Amato, G., Paulsen, M.J., Eskandari, A., Hironaka, C.E., Phansalkar, R., Sharma, B., et al. (2019). A Unique Collateral Artery Development Program Promotes Neonatal Heart Regeneration. Cell *176*, 1128-1142 e1118. 10.1016/j.cell.2018.12.023.
49. Salamon, R.J., Halbe, P., Kasberg, W., Bae, J., Audhya, A., and Mahmoud, A.I. (2023). Parasympathetic and sympathetic axons are bundled in the cardiac ventricles and undergo physiological reinnervation during heart regeneration. iScience *26*, 107709. 10.1016/j.isci.2023.107709.
50. Arolkar, G., Kumar, S.K., Wang, H., Gonzalez, K.M., Kumar, S., Bishnoi, B., Rios Coronado, P.E., Woo, Y.J., Red-Horse, K., and Das, S. (2023). Dedifferentiation and Proliferation of Artery Endothelial Cells Drive Coronary Collateral Development in Mice. Arterioscler Thromb Vasc Biol *43*, 1455-1477. 10.1161/ATVBAHA.123.319319.
51. Saxena, A., Fish, J.E., White, M.D., Yu, S., Smyth, J.W., Shaw, R.M., DiMaio, J.M., and Srivastava, D. (2008). Stromal cell-derived factor-1alpha is cardioprotective after myocardial infarction. Circulation *117*, 2224-2231. 10.1161/CIRCULATIONAHA.107.694992.
52. Hiesinger, W., Perez-Aguilar, J.M., Atluri, P., Marotta, N.A., Frederick, J.R., Fitzpatrick, J.R., 3rd, McCormick, R.C., Muenzer, J.R., Yang, E.C., Levit, R.D., et al. (2011). Computational protein design to reengineer stromal cell-derived factor-1alpha generates an effective and translatable angiogenic polypeptide analog. Circulation *124*, S18-26. 10.1161/CIRCULATIONAHA.110.009431.
53. Marin-Juez, R., El-Sammak, H., Helker, C.S.M., Kamezaki, A., Mullapuli, S.T., Bibli, S.I., Foglia, M.J., Fleming, I., Poss, K.D., and Stainier, D.Y.R. (2019). Coronary

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Revascularization During Heart Regeneration Is Regulated by Epicardial and Endocardial Cues and Forms a Scaffold for Cardiomyocyte Repopulation. Developmental cell *51*, 503-515 e504. 10.1016/j.devcel.2019.10.019.

54. Sharma, B., Ho, L., Ford, G.H., Chen, H.I., Goldstone, A.B., Woo, Y.J., Quertermous, T., Reversade, B., and Red-Horse, K. (2017). Alternative Progenitor Cells Compensate to Rebuild the Coronary Vasculature in Elabela- and Apj-Deficient Hearts. Developmental cell *42*, 655-666 e653. 10.1016/j.devcel.2017.08.008.

55. Sturny, R., Boulgakoff, L., Kelly, R.G., and Miquerol, L. (2024). The transient formation of collaterals contributes to the restoration of the arterial tree during cardiac regeneration in neonatal mice. bioRxiv, 2023.2012.2019.572474. 10.1101/2023.12.19.572474.

56. Goldberg, A., Southern, D.A., Galbraith, P.D., Traboulsi, M., Knudtson, M.L., Ghali, W.A., and Alberta Provincial Project for Outcome Assessment in Coronary Heart Disease, I. (2007). Coronary dominance and prognosis of patients with acute coronary syndrome. Am Heart J *154*, 1116-1122, 1116-22. 10.1016/j.ahj.2007.07.041.

57. Parikh, N.I., Honeycutt, E.F., Roe, M.T., Neely, M., Rosenthal, E.J., Mittleman, M.A., Carrozza, J.P., Jr., and Ho, K.K. (2012). Left and codominant coronary artery circulations are associated with higher in-hospital mortality among patients undergoing percutaneous coronary intervention for acute coronary syndromes: report From the National Cardiovascular Database Cath Percutaneous Coronary Intervention (CathPCI) Registry. Circulation. Cardiovascular quality and outcomes *5*, 775-782. 10.1161/CIRCOUTCOMES.111.964593.

58. Veltman, C.E., van der Hoeven, B.L., Hoogslag, G.E., Boden, H., Kharbanda, R.K., de Graaf, M.A., Delgado, V., van Zwet, E.W., Schalij, M.J., Bax, J.J., and Scholte, A.J. (2015). Influence of coronary vessel dominance on short- and long-term outcome in patients after ST-segment elevation myocardial infarction. Eur Heart J *36*, 1023-1030. 10.1093/eurheartj/ehu236.

59. Bernabeu, E., Canela-Xandri, O., Rawlik, K., Talenti, A., Prendergast, J., and Tenesa, A. (2021). Sex differences in genetic architecture in the UK Biobank. Nature genetics *53*, 1283-1289. 10.1038/s41588-021-00912-0.

60. Hemani, G., Bowden, J., and Davey Smith, G. (2018). Evaluating the potential role of pleiotropy in Mendelian randomization studies. Human molecular genetics *27*, R195-R208. 10.1093/hmg/ddy163.

61. Veltman, C.E., de Graaf, F.R., Schuijf, J.D., van Werkhoven, J.M., Jukema, J.W., Kaufmann, P.A., Pazhenkottil, A.P., Kroft, L.J., Boersma, E., Bax, J.J., et al. (2012). Prognostic value of coronary vessel dominance in relation to significant coronary artery disease determined with non-invasive computed tomography coronary angiography. Eur Heart J *33*, 1367-1377. 10.1093/eurheartj/ehs034.

62. Gebhard, C., Fuchs, T.A., Stehli, J., Gransar, H., Berman, D.S., Budoff, M.J., Achenbach, S., Al-Mallah, M., Andreini, D., Cademartiri, F., et al. (2015). Coronary dominance and prognosis in patients undergoing coronary computed tomographic angiography: results from the CONFIRM (COronary CT Angiography EvaluatioN For Clinical Outcomes: An InteRnational Multicenter) registry. Eur Heart J Cardiovasc Imaging *16*, 853-862. 10.1093/ehjci/jeu314.

63. Wang, L., Li, J., Gao, Y., Li, R., Zhang, J., Su, D., Wang, T., Yang, G., and Wang, X. (2019). Association between coronary dominance and acute inferior myocardial

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

infarction: a matched, case-control study. BMC cardiovascular disorders *19*, 35, 35. 10.1186/s12872-019-1007-5.

64. Kuno, T., Numasawa, Y., Miyata, H., Takahashi, T., Sueyoshi, K., Ohki, T., Negishi, K., Kawamura, A., Kohsaka, S., and Fukuda, K. (2013). Impact of coronary dominance on in-hospital outcomes after percutaneous coronary intervention in patients with acute coronary syndrome. PLoS One *8*, e72672, e72672. 10.1371/journal.pone.0072672.
65. Veltman, C.E., Hoogslag, G.E., Kharbanda, R.K., de Graaf, M.A., van Zwet, E.W., van der Hoeven, B.L., Delgado, V., Bax, J.J., and Scholte, A.J. (2014). Relation between coronary arterial dominance and left ventricular ejection fraction after ST-segment elevation acute myocardial infarction in patients having percutaneous coronary intervention. Am J Cardiol *114*, 1646-1650. 10.1016/j.amjcard.2014.08.035.
66. Mikaeilvand, A., Firuozi, A., Basiri, H., Varghaei, A., Izadpanah, P., Kojuri, J., Abdi-Ardekani, A., and Attar, A. (2020). Association of coronary artery dominance and mortality rate and complications in patients with ST-segment elevation myocardial infarction treated with primary percutaneous coronary intervention. J Res Med Sci *25*, 107, 107. 10.4103/jrms.JRMS_414_19.
67. Selcuk, E., Cevirme, D., and Bugra, O. (2020). Prognostic Value of Coronary Dominance in Patients Undergoing Elective Coronary Artery Bypass Surgery. Braz J Cardiovasc Surg *35*, 452-458. 10.21470/1678-9741-2019-0079.
68. Vasheghani-Farahani, A., Kassaian, S.E., Yaminisharif, A., Davoodi, G., Salarifar, M., Amirzadegan, A., Darabian, S., Fotouhi, A., Sadigh, G., Razavi, S.A., and Hakki, E. (2008). The association between coronary arterial dominancy and extent of coronary artery disease in angiography and paraclinical studies. Clin Anat *21*, 519-523. 10.1002/ca.20669.
69. Yan, B., Yang, J., Fan, Y., Zhao, B., Ma, Q., Yang, L., and Ma, X. (2018). Association of coronary dominance with the severity of coronary artery disease: a cross-sectional study in Shaanxi Province, China. BMJ Open *8*, e021292. 10.1136/bmjopen-2017-021292.
70. Azour, L., Steinberger, S., Toussie, D., Titano, R., Kukar, N., Babb, J., and Jacobi, A. (2021). Influence of coronary dominance on coronary artery calcification burden. Clin Imaging *77*, 283-286. 10.1016/j.clinimag.2021.06.008.
71. Verbanck, M., Chen, C.Y., Neale, B., and Do, R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nature genetics *50*, 693-698. 10.1038/s41588-018-0099-7.
72. Gaziano, J.M., Concato, J., Brophy, M., Fiore, L., Pyarajan, S., Breeling, J., Whitbourne, S., Deen, J., Shannon, C., Humphries, D., et al. (2016). Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J Clin Epidemiol *70*, 214-223. 10.1016/j.jclinepi.2015.09.016.
73. Byrd, J.B., Vigen, R., Plomondon, M.E., Rumsfeld, J.S., Box, T.L., Fihn, S.D., and Maddox, T.M. (2013). Data quality of an electronic health record tool to support VA cardiac catheterization laboratory quality improvement: the VA Clinical Assessment, Reporting, and Tracking System for Cath Labs (CART) program. Am Heart J *165*, 434-440. 10.1016/j.ahj.2012.12.009.
74. Maddox, T.M., Plomondon, M.E., Petrich, M., Tsai, T.T., Gethoffer, H., Noonan, G., Gillespie, B., Box, T., Fihn, S.D., Jesse, R.L., and Rumsfeld, J.S. (2014). A national clinical quality program for Veterans Affairs catheterization laboratories (from the

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Veterans Affairs clinical assessment, reporting, and tracking program). Am J Cardiol *114*, 1750-1757. 10.1016/j.amjcard.2014.08.045.

75. Hunter-Zinck, H., Shi, Y., Li, M., Gorman, B.R., Ji, S.G., Sun, N., Webster, T., Liem, A., Hsieh, P., Devineni, P., et al. (2020). Genotyping Array Design and Data Quality Control in the Million Veteran Program. Am J Hum Genet *106*, 535-548. 10.1016/j.ajhg.2020.03.004.
76. Delaneau, O., Zagury, J.F., Robinson, M.R., Marchini, J.L., and Dermitzakis, E.T. (2019). Accurate, scalable and integrative haplotype estimation. Nature communications *10*, 5436. 10.1038/s41467-019-13225-y.
77. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., Mcgue, M., et al. (2016). Next-generation genotype imputation service and methods. Nature genetics *48*, 1284-1287. 10.1038/ng.3656.
78. Verma, A., Huffman, J.E., Rodriguez, A., Conery, M., Liu, M., Ho, Y.L., Kim, Y., Heise, D.A., Guare, L., Panickan, V.A., et al. (2023). Diversity and Scale: Genetic Architecture of 2,068 Traits in the VA Million Veteran Program. medRxiv. 10.1101/2023.06.28.23291975.
79. Patterson, N., Price, A.L., and Reich, D. (2006). Population structure and eigenanalysis. PLoS Genet *2*, e190. 10.1371/journal.pgen.0020190.
80. Evans, L.M., Tahmasbi, R., Vrieze, S.I., Abecasis, G.R., Das, S., Gazal, S., Bjelland, D.W., de Candia, T.R., Haplotype Reference, C., Goddard, M.E., et al. (2018). Comparison of methods that use whole genome data to estimate the heritability and genetic architecture of complex traits. Nature genetics *50*, 737-745. 10.1038/s41588-018-0108-x.
81. Visscher, P.M., Hemani, G., Vinkhuyzen, A.A., Chen, G.B., Lee, S.H., Wray, N.R., Goddard, M.E., and Yang, J. (2014). Statistical power to detect genetic (co)variance of complex traits using SNP data in unrelated samples. PLoS Genet *10*, e1004269. 10.1371/journal.pgen.1004269.
82. Lee, S.H., Wray, N.R., Goddard, M.E., and Visscher, P.M. (2011). Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet *88*, 294-305. 10.1016/j.ajhg.2011.02.002.
83. Lee, S.H., DeCandia, T.R., Ripke, S., Yang, J., Sullivan, P.F., Goddard, M.E., Keller, M.C., Visscher, P.M., and Wray, N.R. (2012). Estimating the proportion of variation in susceptibility to schizophrenia captured by common SNPs. Nature genetics *44*, 247-250. 10.1038/ng.1108.
84. Winkler, T.W., Day, F.R., Croteau-Chonka, D.C., Wood, A.R., Locke, A.E., Magi, R., Ferreira, T., Fall, T., Graff, M., Justice, A.E., et al. (2014). Quality control and conduct of genome-wide association meta-analyses. Nature protocols *9*, 1192-1212. 10.1038/nprot.2014.071.
85. Consortium, G.T. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science (New York, N.Y *369*, 1318-1330. 10.1126/science.aaz1776.
86. Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A simple new approach to variable selection in regression, with application to genetic fine mapping. Journal of the Royal Statistical Society Series B-Statistical Methodology *82*, 1273-1300. 10.1111/rssb.12388.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

87. Liu, B., Gloudemans, M.J., Rao, A.S., Ingelsson, E., and Montgomery, S.B. (2019). Abundant associations with gene expression complicate GWAS follow-up. Nature genetics *51*, 768-769. 10.1038/s41588-019-0404-0.
88. Verma, A., Damrauer, S.M., Naseer, N., Weaver, J., Kripke, C.M., Guare, L., Sirugo, G., Kember, R.L., Drivas, T.G., Dudek, S.M., et al. (2022). The Penn Medicine BioBank: Towards a Genomics-Enabled Learning Healthcare System to Accelerate Precision Medicine in a Diverse Population. J Pers Med *12*. 10.3390/jpm12121974.
89. Renier, N., Adams, E.L., Kirst, C., Wu, Z., Azevedo, R., Kohl, J., Autry, A.E., Kadiri, L., Umadevi Venkataraju, K., Zhou, Y., et al. (2016). Mapping of Brain Activity by Automated Volume Analysis of Immediate Early Genes. Cell *165*, 1789-1802. 10.1016/j.cell.2016.05.007.
90. Roostalu, U., Thisted, L., Skytte, J.L., Salinas, C.G., Pedersen, P.J., Hecksher-Sorensen, J., Rolin, B., Hansen, H.H., MacKrell, J.G., Christie, R.M., et al. (2021). Effect of captopril on post-infarction remodelling visualized by light sheet microscopy and echocardiography. Scientific reports *11*, 5241. 10.1038/s41598-021-84812-7.
91. Cunningham, F.G., Leveno, K.J., Bloom, S.L., Dashe, J.S., Hoffman, B.L., Casey, B.M., and Spong, C.Y. (2018). Abortion. In Williams Obstetrics, 25e, (McGraw-Hill Education).
92. Cunningham, F.G., Leveno, K.J., Bloom, S.L., Dashe, J.S., Hoffman, B.L., Casey, B.M., and Spong, C.Y. (2018). Prenatal Care. In Williams Obstetrics, 25e, (McGraw-Hill Education).
93. Moffitt, J.R., Hao, J., Wang, G., Chen, K.H., Babcock, H.P., and Zhuang, X. (2016). High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. Proc Natl Acad Sci U S A *113*, 11046-11051. 10.1073/pnas.1612826113.
94. Huang, H., Zhu, Q., Jussila, A., Han, Y., Bintu, B., Kern, C., Conte, M., Zhang, Y., Bianco, S., Chiariello, A.M., et al. (2021). CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. Nature genetics *53*, 1064-1074. 10.1038/s41588-021-00863-6.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

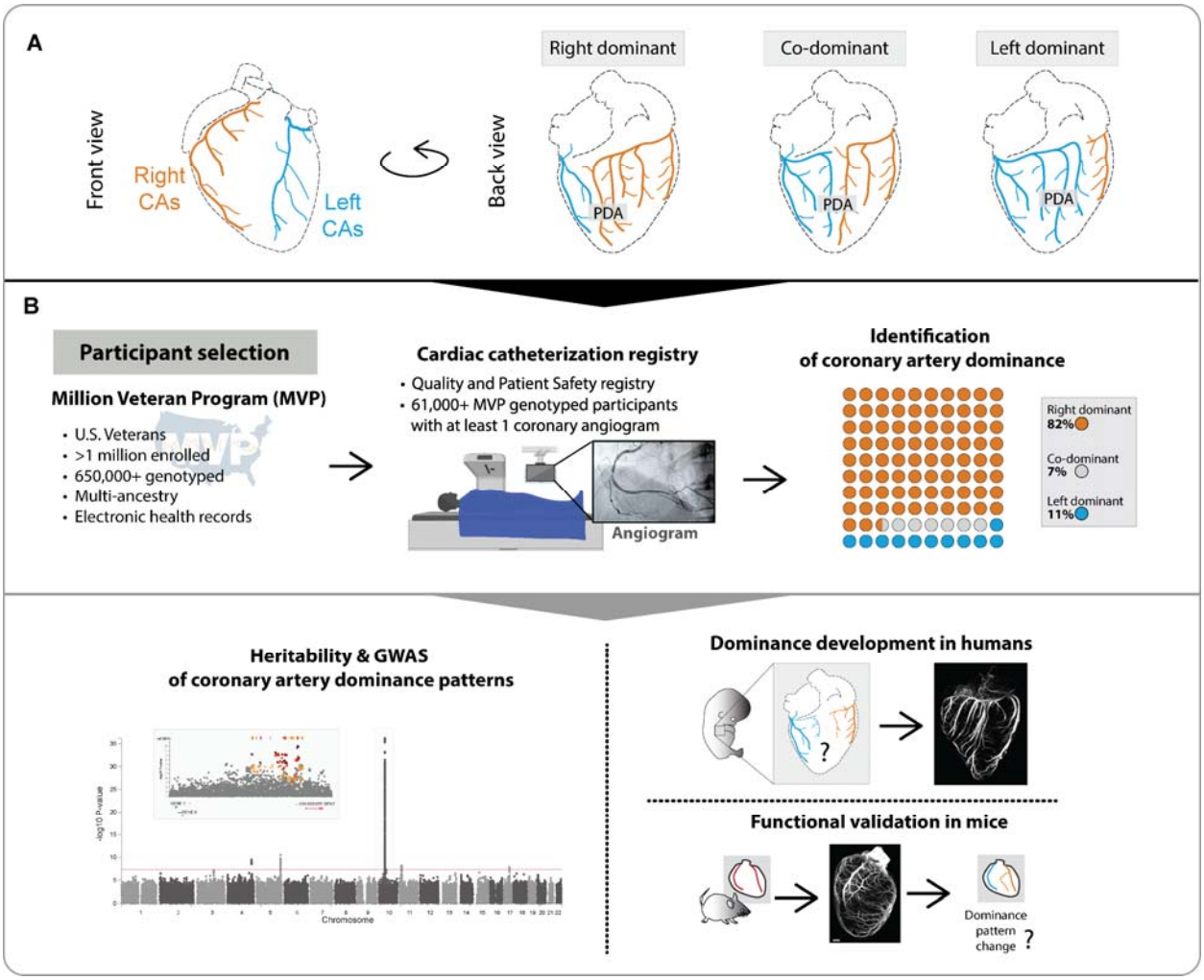
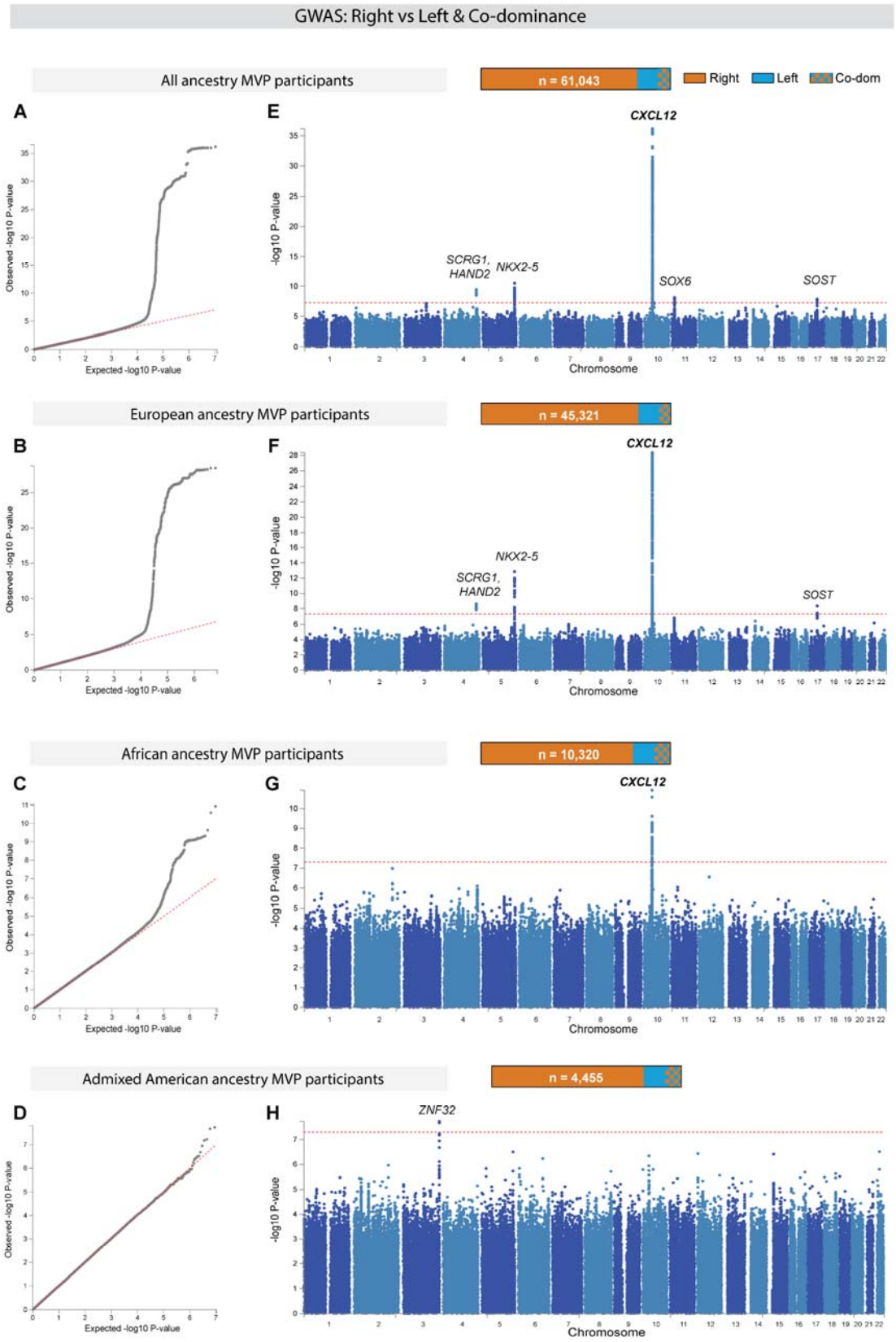


Figure 1. Study design to identify and understand the genetic determinants of coronary artery dominance. (**A**) Variation in the posterior descending artery (PDA) determines dominance in the human heart. (**B**) A multi-ancestry genome-wide association study (GWAS) was performed using genotype data from participants in MVP whose coronary arteries were imaged via angiogram. Analyses were conducted in all participants combined and in cohorts stratified by genetically inferred ancestry. Whole-organ immunolabeling of fetal hearts investigated when dominance is established, and mouse studies validated the most significant genetic association. CAs, coronary arteries.

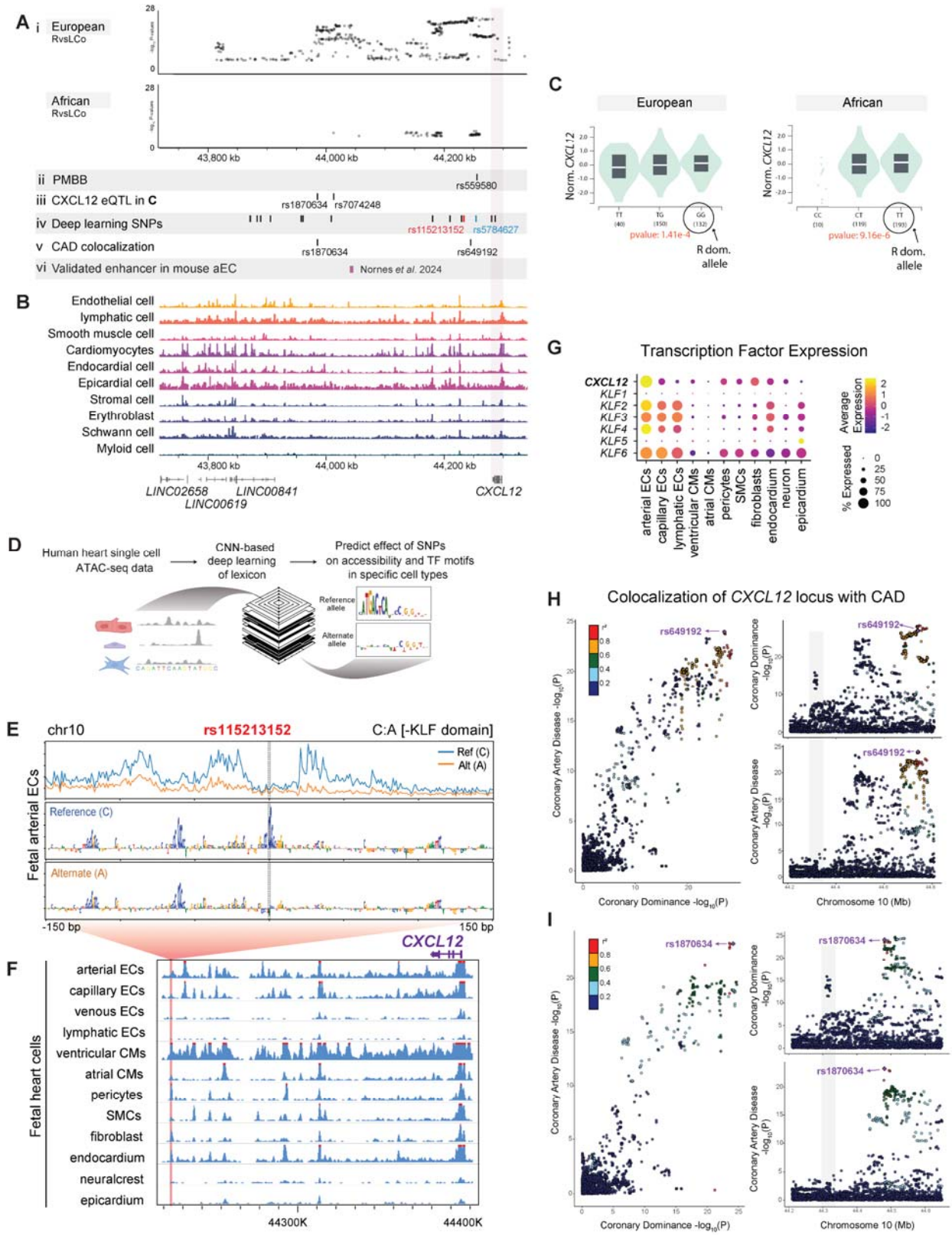
medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Figure 2. Genome-wide association study (GWAS) of right versus left/co-dominance in the Million Veteran Program (MVP). GWAS was conducted in all participants independent of ancestry (All ancestry) as well as in each of the three largest genetically inferred ancestry subgroups: European, African, and Admixed American. (**A-D**) Quantile-quantile plots (Q-Q) show the observed versus expected P-values from association for all genetic variants. Red line indicates null hypothesis. (**E-H**) Manhattan plots for the corresponding Q-Q plots with P-values of genetic variants. Gene(s) mapped to regions reaching genome wide significance (GWS) are labeled. Red line indicates GWS threshold (P-value$<5x10^{-8}$).

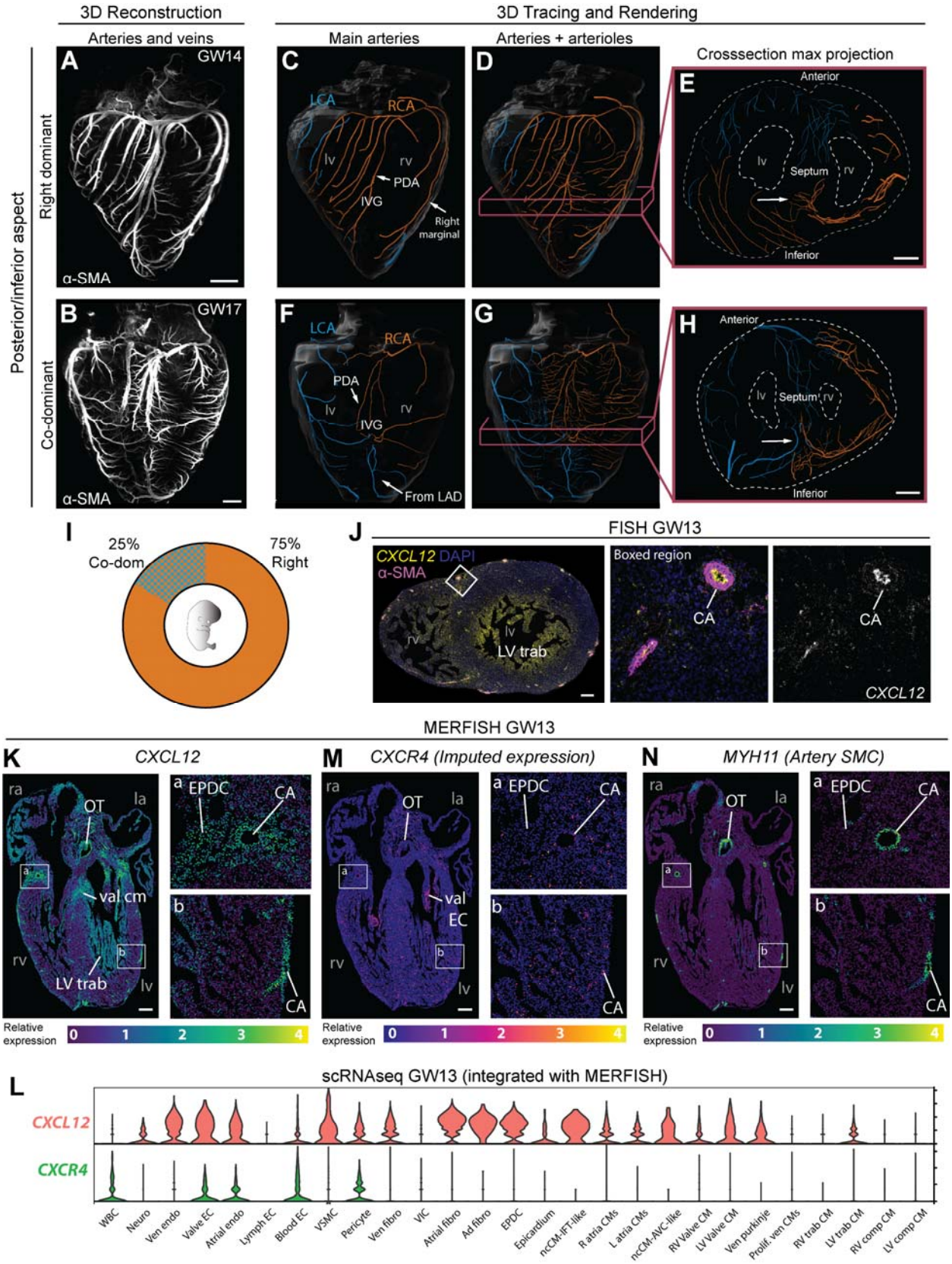
medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Figure 3. Dominance variants localize to putative *CXCL12* regulatory regions. (**A**) SNPs reaching genome wide significance in the indicated MVP populations aligned with variants from additional analyses: Penn Medicine Biobank (PMBB) replication study, Genotype-Tissue Expression (GTEx) eQTLs, deep learning-based predictions, and coronary artery disease (CAD) colocalization. An *in vivo* validated *Cxcl12* enhancer is also shown. (**B**) Location of accessible chromatin in human fetal heart cell types in relation to dominance SNPs in **A**. (**C**) Violin plots of two lead SNP eQTLs revealed that a higher prevalence of right dominance is associated with increased *CXCL12* in GTEx data. Left is rs1870634; right is rs7074248. Testis expression levels by genotype are shown, but directionality is the same in arterial and coronary tissues (see **Supplementary Figure 5A, B**). (**D**) Schematic of convolutional neural network (CNN) deep learning models used to predict the effect of alternate alleles. (**E**) Predicted per-base accessibility for reference and alternate alleles of rs115213152 suggest a disrupted KLF motif. (**F**) Sequencing tracks of chromatin accessibility near *CXCL12*. Red highlights rs115213152 region where open chromatin coincides with *CXCL12* promoter accessibility in select cardiac cell types, including coronary artery cells. (**G**) ScRNAseq from fetal hearts revealed co-expression of *CXCL12* and *KLFs*, supporting a potential role for these transcription factors in the activity predicted in **E**. (**H** and **I**) LocusCompare ouput displaying colocalization plot (left), and LocusZoom plots for each trait (right). The SNP with the highest H4 posterior probability using COLOC is labelled in purple with additional variants in linkage disequilibrium with labelled SNP colored according to R^2 value. **Top panel (H)** is unmodified analysis and **bottom panel (I)** shows result when the first colocalization signal at rs649192 (44650000-44850000) is masked.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Figure 4. Coronary dominance is apparent during fetal development when *CXCL12* is expressed. (**A** and **B**) Whole-organ immunolabeling of coronary artery smooth muscle (α-SMA) in hearts from the indicated gestational weeks (GW). (**C-H**) Representative artery tracings highlighted branches of the main arteries (**C, F**) or main and lower order arteries (**D, E, G, H**) that originated from the right (RCA, orange) or left (LCA, blue) coronary ostia. (**C-E**) Right dominance was indicated when the artery in the interventricular groove (IVG), i.e., the posterior descending artery (PDA), originated from the RCA (**C** and **D**) and when the inferior septum was occupied by right branches (arrow in **E**). (**F-H**) Co-dominance was indicated when arteries within the IVG stemmed from both the RCA and LCA (**F** and **G**) and the inferior septum was occupied by RCA and LCA branches (arrow in **H**). (**I**) Distribution in a cohort of n=8 human fetal hearts. (**J**) *CXCL12* fluorescence *in situ* hybridization (FISH) and α-SMA immunolabeling on a transverse section through a GW13 heart. Expression was in coronary arteries (CA) and trabecular cardiomyocytes (LV trab). DAPI labeled nuclei. (**K-N**) Spatial expression plots (**K, M, L**) and scRNAseq-derived violin plots (**L**) in GW13 human fetal hearts. MERFISH localized *CXCL12* and *MYH11* while *CXCR4* was imputed by integrating both datasets. ad, adventitial; AVC, atrioventricular; comp, compact; EC, endothelial cell; CM, cardiomyocyte; EPDC, epicardial-derived cells; Endo, endocardial cell; Fibro, fibroblast; IFT, inflow tract; Neuro, neuronal; la, left atrium; Lymph, lymphatic; lv, left ventricle; LV trab, left ventricle trabeculae; ncCM, non-chambered cardiomyocyte; OT, outflow tract; ra, right atrium; RV, right ventricle; trab, trabecular; val, valve; ven, ventricular; VIC, valve interstitial cell; VSMC, vascular smooth muscle; WBC, white blood cell. Scale bars: A, B, E, and J, 500μm; H, 1mm; J, 200μm; K-O, 250μm.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

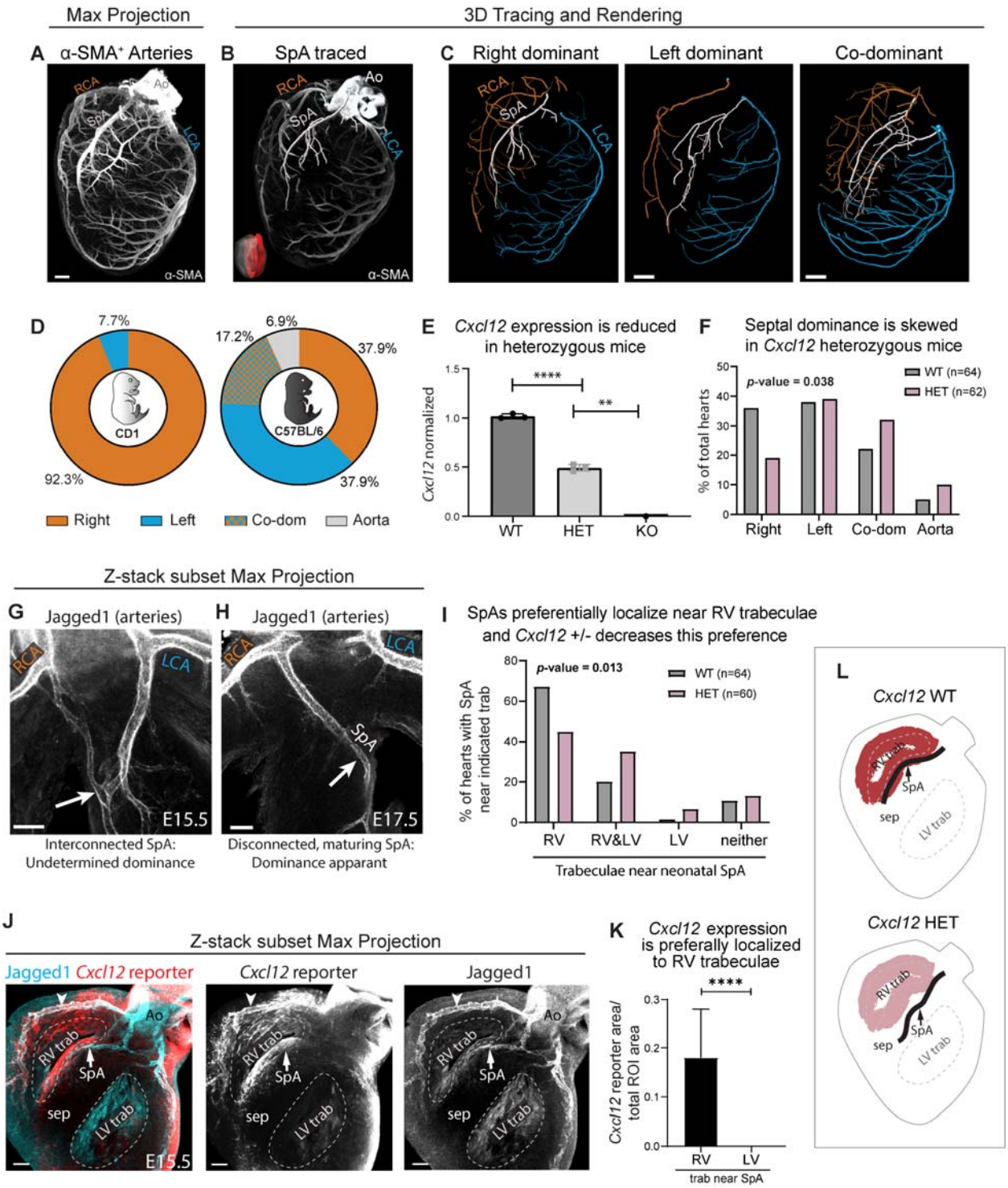


Figure 5. *Cxcl12* influences septal artery dominance in mice. (**A**) Whole-organ immunolabeling of coronary artery smooth muscle (α-SMA) in a postnatal day 6 mouse heart. (**B**) Septal artery (SpA) tracing overlayed onto a Z-stack subset max projection of internal coronary arteries (red region in insert). (**C**) Artery tracings in whole-organ images demonstrating various SpA connections. (**D**) SpA dominance ratios are strain specific. N=13 CD1 neonatal hearts and N=29 for C56BL/6J. (**E**) QPCR using whole hearts from embryonic day (E) 16.5.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

WT, *Cxcl12*$^{+/+}$; HET, *Cxcl12*$^{DsRed/+}$; KO, *Cxcl12*$^{DsRed/DsRed}$. Each dot is the average value for genotypes within single litters normalized to WT littermates. Error bars, mean±st dev: **, p=0.0027, p=0.0016, p=0.0011; ****, p<0.0001 by two-sided Student's t-test. (**F**) Blinded scoring of septal artery dominance in WT and HET neonatal mice. P-value by two-tailed Chi-square test. (**G** and **H**) Whole-organ immunolabeling of WT embryonic hearts. First, an immature artery network is attached to both sides (**G**) followed by maturation into a SpA with dominance shortly before birth (**H**). (**I**) Blinded scoring of SpA location as related to trabeculae (see **Supplementary Figure 8I**). P-value by two-tailed Chi-square test. (**J**) *Cxcl12* reporter allele shows that SpA develops in regions of high expression. Arrowhead, expression in artery endothelial cells. (**K**) Reporter allele fluorescence in standardized regions of interest (ROI) along right or left trabeculae. N=14 hearts. Error bars, mean±st dev: ****, p<0.0001 by two-sided Student's t-test. (**L**) Schematic of *Cxcl12* bias in trabeculae, and HET's effect on SpA localization. Ao, aorta; LCA, left coronary artery; LV, left ventricle; RCA, right coronary artery; RV, right ventricle; sep, septum; trab, trabecular cardiomyocytes. Scale bars: A, C 400μm; G - I, 100μm.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

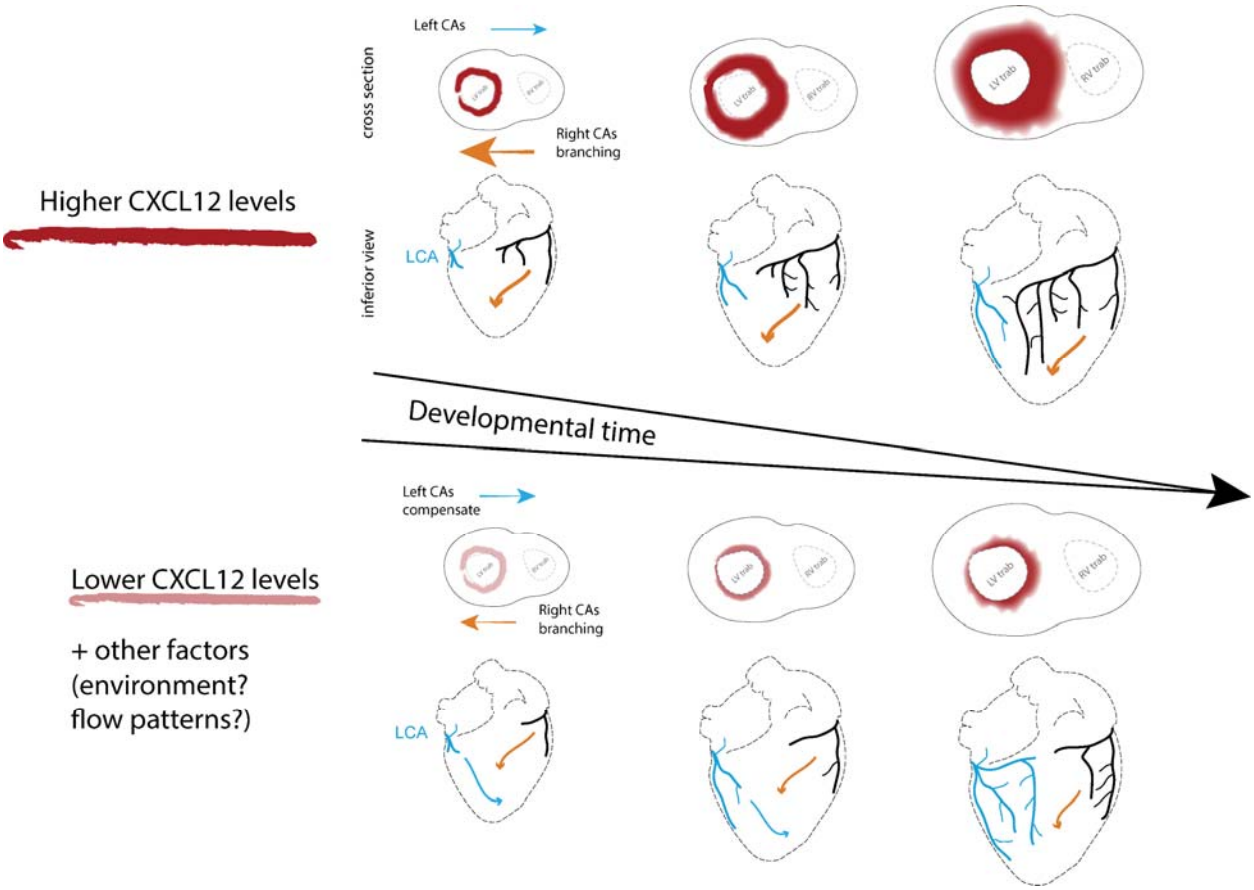


Figure 6. Working model of how *CXCL12* genetic variants influence human coronary artery development and result in different dominance configurations. Genomic data suggested that dominance associated SNPs influence *CXCL12* gene expression with right CA dominance being associated with higher CXCL12 levels. Mouse developmental genetics showed that *Cxcl12* levels influence SpA dominance and indicated that arteries are attracted to form near expression domains. We propose that arterial endothelial cells are attracted to *CXCL12* expression in the epicardium, myocardium, and artery endothelial cells to preferentially extend the right coronary artery on the back of the heart. When *CXCL12* levels are lower, extension of right branches is delayed or misplaced, and the left grows to compensate. Colored arrows indicate the extension path(s) that right (orange) and left (blue) CAs will take based on CXCL12 levels.

Table 1. Lead genetic variants, effect sizes, allele frequencies, and mapped genes at loci reaching genome wide significance for coronary artery dominance status

Genomic Locus No	GIA	GWAS Analysis	chr	snp rsID	bp hg38/GRCh38	bp hg19/GRCh37	Allele 1	Allele 2	AF Allele 2	BETA	SE	OR (95%CIs)	p.value	AF case	AF ctrl	N_case	N_ctrl	N	Mapped Genes
1	AFR	RvsL	3	rs544905040	2757792	2841160	A	G	0.005	1.33	0.25	3.8 (2.33-6.2)	4.71E-08	0.013	0.004	1167	8280	9447	CNTN4, CNTN6
2	AMR	RvsLCo	3	rs9873160	197417938	197144809	G	C	0.073	0.56	0.10	1.75 (1.43-2.14)	1.82E-08	0.108	0.066	738	3717	4455	DLG1
3	EUR	RvsLCo	4	rs56410855	173452142	174373293	G	T	0.094	0.17	0.03	1.19 (1.12-1.26)	2.32E-09	0.107	0.091	7708	37613	45321	SCRG1, HMGB2, FBXO8, CEP44
3	EUR	RvsL	4	rs56410855	173452142	174373293	G	T	0.094	0.21	0.04	1.24 (1.15-1.33)	1.03E-09	0.111	0.091	4944	37613	42557	SCRG1, HMGB2, FBXO8, CEP44
3	ALL	RvsL	4	rs78510568/rs74882405 5	173475861	174397012	C	T	0.078	0.21	0.03	1.24 (1.16-1.32)	1.95E-10	0.092	0.076	6640	50405	57045	SCRG1, HMGB2, FBXO8, CEP44
3	ALL	RvsLCo	4	rs112133513	173487900	174409051	G	A	0.081	0.17	0.03	1.19 (1.12-1.25)	3.40E-10	0.090	0.079	10638	50405	61043	SCRG1, HMGB2, FBXO8, CEP44
4	ALL	RvsLCo	5	rs567007743	99585888	98921592	A	G	0.002	0.67	0.15	1.95 (1.45-2.63)	5.05E-06	0.003	0.002	10638	50405	61043	ELL2, PCSK1, ERAP1, ERAP2, LNPEP, LIX1, RIOK2, RGMB, CHD1, FAM174A, ST8SIA4, SLCO4C1
5	EUR	RvsLCo	5	rs954635	173220844	172647847	A	T	0.282	0.15	0.02	1.16 (1.11-1.2)	1.46E-13	0.306	0.277	7708	37613	45321	SH3PXD2B, NEURL1B, DUSP1, ERGIC1, ATP6V0E1, CREBRF, BNIP1, **NKX2-5**, C5orf47, MSX2
5	ALL	RvsLCo	5	rs954635	173220844	172647847	A	T	0.362	0.11	0.02	1.12 (1.08-1.16)	3.03E-11	0.389	0.357	10638	50405	61043	SH3PXD2B, NEURL1B, DUSP1, ATP6V0E1, CREBRF, BNIP1, **NKX2-5**, C5orf47, MSX2
5	EUR	RvsL	5	rs62385088	173222031	172649034	G	A	0.164	0.21	0.03	1.23 (1.17-1.3)	1.09E-13	0.190	0.161	4944	37613	42557	SH3PXD2B, NEURL1B, DUSP1, ATP6V0E1, CREBRF, BNIP1, **NKX2-5**, C5orf47, MSX2
5	ALL	RvsL	5	rs55893552	173223617	172650620	T	A	0.158	0.17	0.03	1.18 (1.12-1.24)	3.79E-11	0.175	0.156	6640	50405	57045	SH3PXD2B, NEURL1B, DUSP1, ATP6V0E1, CREBRF, BNIP1, **NKX2-5**, C5orf47, MSX2
6	ALL	RvsL	10	rs2902339	44018658	44514106	G	A	0.784	-0.24	0.02	0.79 (0.75-0.82)	5.14E-28	0.746	0.788	6640	50405	57045	HNRNPF,ZNF239, ZNF485, ZNF32, **CXCL12,** TMEM72, RASSF4, C10orf10, C10orf25, ZNF22, ALOX5, NPY4R
6	EUR	RvsL	10	rs7917534	44041975	44537423	G	C	0.798	-0.25	0.03	0.78 (0.74-0.82)	1.29E-21	0.762	0.803	4944	37613	42557	HNRNPF,ZNF239, ZNF485, ZNF32, **CXCL12,** TMEM72, RASSF4, C10orf10, C10orf25, ZNF22, ALOX5, NPY4R
6	EUR	RvsLCo	10	rs589655	44249867	44745315	C	G	0.100	0.31	0.03	1.37 (1.3-1.45)	4.37E-29	0.126	0.095	7708	37613	45321	HNRNPF, ZNF239, ZNF485, ZNF239, ZNF485, ZNF32, **CXCL12**, TMEM72, RASSF4, C10orf10, C10orf25, ZNF22, ALOX5, NPY4R
6	AFR	RvsL	10	rs2576354	44254945	44750393	G	A	0.259	0.32	0.05	1.37 (1.25-1.51)	9.21E-11	0.317	0.251	1167	8280	9447	ZNF485, ZNF32, CXCL12, TMEM72, RASSF4, C10orf10, C10orf25, ZNF22, ALOX5, NPY4R
6	AFR	RvsLCo	10	rs2576354	44254945	44750393	G	A	0.262	0.27	0.04	1.3 (1.21-1.41)	1.17E-11	0.305	0.251	2040	8280	10320	HNRNPF, ZNF485, ZNF32, CXCL12, TMEM72, RASSF4, C10orf10, C10orf25, ZNF22, NPY4R
6	ALL	RvsLCo	10	rs606314	44255031	44750479	C	G	0.101	0.30	0.02	1.36 (1.29-1.42)	6.58E-37	0.126	0.096	10638	50405	61043	RASGEF1A, HNRNPF, ZNF239, ZNF485, ZNF32, **CXCL12**, TMEM72, RASSF4, C10orf10, C10orf25, ZNF22, ALOX5, NPY4R
7	ALL	RvsL	10	rs1194743	52452837	54212597	T	C	0.779	0.15	0.02	1.17 (1.11-1.22)	2.44E-10	0.802	0.776	6640	50405	57045	A1CF, PRKG1, DKK1
7	EUR	RvsL	10	rs1194743	52452837	54212597	T	C	0.747	0.15	0.03	1.16 (1.1-1.23)	7.51E-09	0.770	0.744	4944	37613	42557	A1CF, PRKG1, DKK1
8	ALL	RvsL	11	rs35342212/rs67428213	15874288	15895834	G	A	0.229	-0.13	0.02	0.87 (0.83-0.92)	1.05E-08	0.206	0.232	6640	50405	57045	INSC, SOX6, C11orf58
8	ALL	RvsLCo	11	rs35342212/rs67428213	15874288	15895834	G	A	0.228	-0.11	0.02	0.9 (0.86-0.93)	7.43E-09	0.208	0.232	10638	50405	61043	INSC, SOX6, C11orf58
9	ALL	RvsL	15	rs58473469	174592722	174885063	C	T	0.200	-0.16	0.03	0.85 (0.81-0.9)	9.75E-10	0.183	0.202	6640	50405	57045	ADPGK, C15orf60, LOXL1, STOML1, PML, STRA6, CCDC33, CYP11A1, **SEMA7A**, **UBL7**, **ARID3B**, **CLK3**, **CYP1A1**, CYP1A2, CSK, LMAN1L, ULK3, SCAMP2, FAM219B, COX5A, RPP25, SCAMP5, GOLGA6D, COMMD4, NEIL1, SNUPN
10	ALL	RvsLCo	17	rs851058	43760351	41837719	G	A	0.391	0.09	0.02	1.09 (1.06-1.13)	1.40E-08	0.405	0.388	10638	50405	61043	AOC2, ARL4D, DHX8, ETV4, SOST, **DUSP3**, C17orf105, MPP3, CD300LG, MPP2, NAGS, G6PC3
10	EUR	RvsLCo	17	rs851058	43760351	41837719	G	A	0.417	0.11	0.02	1.11 (1.07-1.15)	4.54E-09	0.438	0.413	7708	37613	45321	AOC2, ARL4D, DHX8, ETV4, SOST, **DUSP3**, C17orf105, MPP3, CD300LG, MPP2, NAGS

Table 1. **Lead genetic variants, effect sizes, allele frequencies, and mapped genes at loci reach genome wide significance for coronary artery dominance status.** Genomic locus No: Index of genomic risk loci, GIA: Genetically inferred ancestry (ALL; all ancestral groups combined, EUR; genetically inferred European subjects, AFR; genetically inferred African subjects, AMR; genetically inferred Admixed-Americans), GWAS: Genome-wide assocociation analysis comparison (RvsLCo; right dominance as reference versus left and co-dominant subjects combined, RvsL; right dominance (reference) versus left dominance subjects excluding subjects with mixed dominance, chr: chromosome, SNP rsID: variant ID, bp hg38/GRCh38: base position on human build 38, bp hg19/GRCh37: base position on human build 37, Allele1: allele 1, Allele2: allele 2, AF_Allele2: allele frequency of allele 2, BETA:

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license.

effect size of allele 2, SE: standard error of BETA, OR (95%CIs): Odds ratio and 95% confidence intervals of non-right dominance to right dominance for allele 2 compared to allele1, p.value: p value with SaddlePoint Approximation applied, AF_case: allele frequency of allele 2 in cases, AF_ctrl: allele frequency of allele 2 in controls, N_case: sample size in cases, N_ctrl: sample size in controls, N: sample size cases and controls combined. Mapped genes include either by position (ANNOVAR), gene expression (eQTL), or chromatin interaction. Genes mapped by all three methods are bolded.

medRxiv preprint doi: https://doi.org/10.1101/2023.10.27.23297507; this version posted July 5, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .