UNIVERSITY OF CALIFORNIA SAN DIEGO

Using Single Cell Epigenomic Analysis to Reveal Mechanisms of Complex Disease: From the Heart to the Whole Human Body

A Dissertation submitted in partial satisfaction of the requirements
for the degree Doctor of Philosophy

in

Biomedical Sciences

by

James D. Hocker

Committee in charge:

Professor Bing Ren, Chair
Professor Adam Engler
Professor Kelly Frazer
Professor Chris Glass
Professor Gene Yeo

2022

The Dissertation of James D. Hocker is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

I dedicate this dissertation to my family, friends, and mentors who have all graciously supported me on this journey.

Mom, thanks for teaching me to follow my heart. It's been a tough road, but I'm so glad that it ended up leading me back to your heart.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTAL FIGURES

LIST OF SUPPLEMENTAL FILES

Supplemental Table 1: Clinical metadata for heart samples.

Supplemental Table 2: Quality control and cell type composition data for each snATAC dataset.

Supplemental Table 3: Quality control, cell type composition, and integration with snATAC-seq results for snRNA-seq datasets.

Supplemental Table 4: snRNA-seq gene expression by major cluster and major cluster-specific genes

Supplemental Table 5: Union of 287,415 cCREs in the cell types of the human heart.

Supplemental Table 6: List of 19,447 cell type-specific cCREs.

Supplemental Table 7: GREAT analysis for cell type-specific cCREs.

Supplemental Table 8: ChromVAR motif enrichment results in snATAC-seq cell clusters.

Supplemental Table 9: HOMER motif enrichment results for cell type-specific cCREs.

Supplemental Table 10: Differentially accessible cCREs between heart chambers.

Supplemental Table 11: Co-accessible cCRE pairs from Cicero.

Supplemental Table 12: Lists of differentially accessible (DA) cCREs linked to differentially expressed genes.

Supplemental Table 13: GREAT analysis for differentially accessible (DA) cCREs between heart chambers in cardiomyocytes and fibroblasts.

Supplemental Table 14: HOMER motif enrichments for differentially accessible (DA) cCREs between heart chambers in cardiomyocytes and fibroblasts.

Supplemental Table 15: RPKM values and cluster membership for healthy and disease-associated candidate heart enhancers in different cell types.

LIST OF ABBREVIATIONS

AF	atrial fibrillation
ATAC-seq	assay for transposase accessible chromatin with sequencing
cCRE	candidate *cis*-regulatory element
ChIP-seq	chromatin immunoprecipitation with sequencing
DA	differentially accessible
DE	differentially expressed
DNase-seq	DNase I hypersensitive sites sequencing
GREAT	genomic regions enrichment of annotations tool
GWAS	genome wide association study
HF	heart failure
PPA	posterior probability of association
snATAC-seq	sci-ATAC-seq = single cell indexing/single nucleus ATAC-seq
snRNA-seq	single nucleus RNA-seq
TF	transcription factor
RPM	reads per million mapped reads
RPKM	reads per kilobase of transcript, per million mapped reads
SNP	single nucleotide polymorphism
UMAP	uniform manifold approximation and projection

ACKNOWLEDGEMENTS

I would like to acknowledge all of my incredible mentors and co-authors, without whom this work would absolutely not have been possible. I would especially like to thank Dr. Bing Ren for his constant support over the last three and a half years. He is in a category of his own in cultivating creativity and excitement about science, and under his guidance I have grown as a scientist and a person. I would also like to especially thank Dr. Sebastian Preissl, who became a close mentor and friend to me. From the biggest scientific picture to the little details of figure design, he took it upon himself to help me learn the ropes and I will always be grateful. I would also like to thank Dr. Neil Chi, for his invaluable guidance not only on the specific projects herein but also on the physician scientist path.

I thank the National Institute of General Medical Sciences and the National Human Genome Research Institute for their generous grant support. Finally, I would like to acknowledge the leaders and administrators of the UC San Diego Medical Scientist Training Program, the UC San Diego Biomedical Sciences Graduate Program, and the UC San Diego School of Medicine for their support and guidance.

The content of Chapter 2, in full, is a reprint of the material as it appears in Science Advances 2021. Hocker, James D.; Poirion, Olivier†; Zhu, Fugui†; Buchanan, Justin; Zhang, Kai; Chiou, Joshua; Wang, Tsui-min; Hou, Xiaomeng; Li, Yang, Zhang, Yanxiao; Farah, Elie N.; Wang, Allen; McCulloch, Andrew D.; Gaulton, Kyle J.; Ren, Bing*; Chi, Neil C.*; Preissl. Sebastian*. The dissertation author was the primary investigator and author of this paper.

The content of Chapter 3, in full, is a reprint of the material as it appears in Cell 2022. Zhang, Kai†; Hocker, James D.†; Miller, Michael; Hou, Xiaomeng; Chiou, Joshua; Poirion,

Olivier; Qiu, Yunjiang; Li, Yang; Gaulton, Kyle J.; Wang, Allen; Preissl. Sebastian; Chi, Ren, B. The dissertation author was a co-primary investigator and co-first author of this paper.

The content of Chapter 4, in part, is currently being prepared for submission for publication of the material. Hocker, James D.; Buchanan, Justin; Zhang, Kai; Miller, Michael; Shankar, Thirupura S.; Drakos, Stavros; Chi, Neil; Preissl. Sebastian; Chi, Ren, B. The dissertation author was the primary author of this material.

VITA

2010-2014	Research Assistant, University of Wisconsin, Madison
2014	Bachelor of Science, University of Wisconsin, Madison
2014-2016	Postbaccalaureate Fellow, National Institutes of Health
2016-2018	MD Student, University of California San Diego
2018-2022	PhD Student, University of California San Diego
2022	Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Kirkland NJ, Whitehead AW, Hocker JD, Beri P, Vogler G, Hum B, Wang M, Lakatta E, Ren B, Bodmer R, Engler AJ. Age-dependent Lamin remodeling induces cardiac dysfunction via dysregulation of cardiac transcriptional programs. Under review. Preprint: https://www.researchsquare.com/article/rs-1021378/v1 (2022).

Whitehead AW, Hocker JD, Ren B, Engler A. Methods for Cardiac Fibroblast Generation from iPSCs: Updates and Characterization. Journal of Molecular and Cellular Cardiology (2022).

Zhang K*, Hocker JD*, Miller M, Hou X, Chiou J, Poirion O, Qiu Y, Li Y, Zhang Y, Gaulton KJ, Wang A, Preissl S, Ren B. A single-cell atlas of chromatin accessibility in the human genome. Cell (2021) (*contributed equally) (PMID: 34774128).

Hocker JD, Poirion O, Zhu F, Buchanan J, Zhang K, Chiou J, Wang TW, Hou X, Li Y, Zhang Y, Farah EN, Wang A, McCulloch A, Gaulton KJ, Ren B*, Chi NC*, Preissl S*. Cardiac cell type-specific gene regulatory programs and disease risk association. Science Advances (2021). (PMID: 33990324)

Kubo N, Ishii H, Xiong X, Blanco S, Meitinger F, Hu R, Hocker JD, Conte M, Gorkin D, Yu M, Li B, Dixon JR, Hu M, Desai A, Nicodemi M, Zhao H, Ren B. Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. Nature Structural and Molecular Biology (2020) (PMID: 33398174).

Ji Y, Zhu W, Wang H, Wu T, Hu J, Fioravanti J, Lacey N, Gautam S, Le Gall J, Yang X, Hocker JD, Escobar TM, He S, Dell'Orso S, Hawk NV, Kapoor V, Telford WG, Di Croce L, Muljo SA, Zhang Y, Vittorio S, Gattinoni L. miR-155 harnesses Phf19 to potentiate cancer immunotherapy through epigenetic reprogramming of CD8+ T cell fate. Nature Communications (2019) (PMID: 31089138).

Gautam S, Fioravanti J, Zhu W, Le Gall JB, Brohawn P, Lacey NE, Hu J, Hocker JD, Voong N, Kapoor V, Telford WG, Gurusamy D, Yu Z, Bhandoola , Xue H, Roychoudhuri R, Higgs BW, Restifo NP, Bender TP, Ji Y, Gattinoni L. The transcription factor c-Myb enhances CD8+ T cell stemness and antitumor immunity. Nature Immunology (2019) (PMID: 30778251).

Sabatino M, Hu J, Sommariva M, Gautam S, Fellowes V, Hocker JD, Dougherty S, Qin H, Fry TJ, Gress RE, Kochenderfer JN, Stroncek DF, Gattinoni L. 2016. Generation of clinical-grade CD19-specific CAR-modified CD8+ T memory stem cells for the treatment of human B-cell malignancies. Blood (2016) (PMID: 27226436).

Ji Y, Hocker JD, Gattinoni L. Enhancing adoptive T cell immunotherapy with microRNA therapeutics. Sem Immunol (2016) (PMID: 26710685).

Hocker JD, Khan A, Singh M, Hook M, Simpson M, Malsch A, Vollbrecht M, Malone M. Can the Electronic Health Record Identify Vulnerable Older Adults in Need of a Palliative Care Assessment in the Hospital Setting? J Am Geriatr Soc (2015) (PMID: 26189857).

Wen B, Brewer M, Nadiarnykh O, Hocker JD, Singh V, Mackie T, Campagnola P. Texture analysis applied to second harmonic generation image data for ovarian cancer classification. J Biomed Opt (2014) (PMID: 26296156).

Tilbury K*, Hocker JD*, Wen B, Sandbo N, Singh V, Campagnola P. 2014. Second harmonic generation microscopy analysis of extracellular matrix changes in human idiopathic pulmonary fibrosis. J Biomed Opt (2014) (*contributed equally) (PMID: 25134793).

ABSTRACT OF THE DISSERTATION

Using Single Cell Epigenomic Analysis to Reveal Mechanisms of Complex Disease: From the Heart to the Whole Human Body

by

James D. Hocker

Doctor of Philosophy in Biomedical Sciences

University of California San Diego, 2022

Professor Bing Ren, Chair

Complex diseases, and cardiovascular diseases such as coronary artery disease, atrial fibrillation, and heart failure, are leading causes of morbidity and mortality worldwide. These diseases arise from interactions between lifestyle factors, environmental influences, and multiple disease associated genes. Efforts to identify the driving genes underlying complex diseases have culminated in genome-wide association studies (GWAS), which measure associations between common human sequence variants and disease phenotypes in large

population cohorts. To date, GWAS have identified tens of thousands of sequence variants associated with cardiovascular diseases and a spectrum of other complex diseases. However, the vast majority of these variants reside in the noncoding regions of the genome, and do not directly disrupt protein-coding sequences in genes.

Cis-regulatory elements (CREs) are noncoding sequences that regulate the expression levels of neighboring genes in a cell type-specific fashion. Observations that disease associated variants from GWAS are enriched in CREs led to the hypothesis that a major mechanism by which these variants influence disease is by disrupting the regulation of gene expression in specific cell types. However, we still lack comprehensive maps of CREs, not only in the cell types of the human heart, but also in the majority of tissues in the human body. The absence of such maps has posed a key challenge to discovery of the cell types through which disease-associated variants act and the interpretation of their detailed molecular mechanisms.

These challenges, reviewed in Chapter 1, led me to ask the following questions which form the backbone of my thesis research: 1) how do individual human cell types utilize CREs to regulate gene expression, and 2) how do disease-associated noncoding sequence variants from GWAS influence cell type-specific gene regulation to cause disease? In this dissertation, I set out to address these questions in projects of progressively more expansive scope.

First, in Chapter 2, I used single cell epigenomic and transcriptomic methods to define the regulation of gene expression by candidate CREs (cCREs) in nine cell types from the adult human heart. By localizing risk variants for cardiovascular diseases to these cCREs, I uncovered strong enrichments of variants associated with complex cardiovascular diseases in cCREs from individual cardiac cell types, such as atrial fibrillation (AF) variants in cardiomyocyte cCREs. Next, I examined the specific AF risk variants underlying these

enrichments, linked them to putative target genes, and tested their molecular mechanisms in human iPSC derived cardiomyocytes using luciferase reporter assays and CRISPR-Cas9 mediated genome editing. Results from these experiments showed that a cardiomyocyte-specific enhancer containing noncoding AF risk variants is necessary for *KCNH2* expression and regulation of action potential repolarization in cardiomyocytes.

Using this work as a foundation, in Chapter 3, I next applied single cell epigenomic methods to 30 different tissue types from across the entire adult human body. Integrating these datasets with corresponding data from 15 fetal tissue types revealed the cell type-specificity of over 1 million cCREs in 222 distinct human cell types. Moving beyond cardiovascular diseases and cardiac cell types, I next localized risk variants from the spectrum of complex human diseases and traits to body wide maps of cCREs in human cell types. This analysis resulted in thousands of significant enrichments of risk variants for complex diseases in cCREs of specific cell types. To link specific variants to putative molecular functions, I created a framework that incorporates statistical fine mapping, target gene linkage, and measurements of transcription factor binding site disruption to yield candidate molecular functions for hundreds of distinct noncoding risk variants. I lastly highlight examples of specific variants that may disrupt the activity of cell type-specific cCREs to contribute to complex diseases.

Finally, in Chapter 4 I summarize future directions of this research. First, I outline technological developments that will greatly enhance the utility of these data and frameworks for interpreting the functions of complex disease risk variants. Second, I describe ongoing work to use the healthy tissue datasets I generated as a springboard for uncovering cell type-specific gene regulatory programs in diseased human tissues, with a focus on ischemic heart failure.

Chapter 1: Introduction

1.1 Background

1.1.1 Introduction to complex diseases and GWAS

Complex diseases, and cardiovascular diseases such as coronary artery disease, atrial fibrillation, and heart failure, are leading causes of morbidity and mortality worldwide. The vast majority of human diseases are complex, meaning that they result from a combination of external and genetic influences (Craig, 2008). While lifestyle and environmental factors are critical contributors to disease and remain major targets of medical intervention (Arena et al., 2015), the interactions between numerous disease-associated genes play an important role in defining complex disease susceptibility, and mechanistic knowledge of these genes and interactions could lead to new treatments for complex diseases.

In contrast to monogenetic disorders, complex diseases arise from the interaction of multiple genes not only with one another but also with environmental factors. Historically, this complexity hindered the identification of target genes via traditional linkage studies of disease heritability within families (Bush and Moore, 2012; Cui et al., 2010). In the mid 2000's, the sequencing of the human genome and advances in microarray technology enabled novel strategies to detect sequence variants associated with complex human diseases. Specifically, the strategy of searching the whole genome for common single nucleotide polymorphisms (SNPs) that segregated with complex diseases and traits in large cohorts emerged as promising a method to identify disease associated genes and genomic regions (Wellcome Trust Case Control, 2007). Since then, these genome wide association studies (GWAS) have identified tens of thousands of unique sequence variants associated with cardiovascular diseases and a spectrum of other complex diseases and non-disease traits (Loos, 2020; Tam et al., 2019) – each of which could provide novel

insights into the biological mechanisms underlying complex diseases and traits. However, the vast majority of these variants reside in the noncoding regions of the genome, and do not directly disrupt protein-coding sequences in genes, posing a major challenge to interpretation of the mechanisms through which they may lead to disease (Claussnitzer et al., 2020).

1.1.2 Regulation of gene expression by *cis*-regulatory elements

Only around 1% of nucleotides in the human genome directly encode for proteins (Venter et al., 2001). However, much of the remaining noncoding sequences are active in controlling which genes are turned on, when they are turned on, and in which tissues and cell types. They do so in part through the action of *cis*-regulatory elements (CREs), which are noncoding regulatory sequences that dictate the expression patterns of target genes by recruiting sequence specific transcription factors (TFs) (Shlyueva et al., 2014). Upon binding of TFs, CREs frequently adopt conformational changes such that they are more accessible to endonucleases or transposases, enabling genome-wide discovery of CREs by combining assays incorporating these enzymes with high throughput sequencing (Buenrostro et al., 2013; John et al., 2013; Klemm et al., 2019). For example, assay for transposase accessible chromatin with sequencing (ATAC-seq) identifies cCREs by subjecting nuclear chromatin in its native context to the activity of a mutated, hyperactive transposase. At each genomic region this transposase is not sterically hindered from binding, it cleaves DNA and installs next generation sequencing adapters. The hyperactive transposase is unable to reach closed heterochromatin regions, and instead cuts and tags "accessible" or "open" chromatin regions. Purification of adapter-tagged DNA fragments from open chromatin regions followed by PCR amplification, sequencing, and alignment of the resulting

reads permits identification of open chromatin regions genome wide, each of which represents a candidate *cis*-regulatory element (cCRE).

Prior efforts to map cCREs in the human genome using conventional methods such as ATAC-seq, DNase-seq, and histone modification ChIP-seq have, in large part, used heterogeneous bulk tissues as input materials to produce population average measurements. However, human tissues are composed of diverse cell types, each with a unique complement of active CREs. As a result, the current catalogs of candidate regulatory sequences in the human genome (Andersson et al., 2014; Meuleman et al., 2020; Moore et al., 2020; Roadmap Epigenomics et al., 2015; Shen et al., 2012) still lack information about the cell type-specific activities of most elements.

1.1.3 Challenges in interpreting disease-associated variants

Initial observations that disease associated variants from GWAS were enriched inside of cCREs from human tissues and cell lines (Ernst et al., 2011; Maurano et al., 2012; Roadmap Epigenomics et al., 2015) led to the hypothesis that an important mechanism by which they influence disease is by disrupting TF binding sites within CREs, or otherwise disturbing the ability of CREs to regulate the expression of disease-associated genes. Further, the enrichment of complex disease and trait associated variants within human cCREs was shown to be highly tissue specific. Risk variants associated with specific phenotypes – such as cardiovascular electrophysiologic traits – were preferentially enriched in cCREs derived from related tissues - such as cardiac tissues (Maurano et al., 2012).

However, several major factors have limited our ability to interpret the molecular functions of trait and disease associated variants. Firstly, almost all GWAS have depended upon targeted genotyping of specific and pre-selected SNPs using microarrays (Uffelmann et al., 2021). These

pre-selected “index” or “lead” SNPs are likely to be inherited together with dozens of variants in neighboring genomic regions in a phenomenon known as linkage disequilibrium (LD) (Slatkin, 2008). Thus, the exact sequence variant responsible for the association at a given locus, referred to as the causal variant, is often not the SNP that has been directly genotyped and must be distinguished from other variants in LD (Broekema et al., 2020). Secondly, the vast majority of variants associated with complex diseases reside in noncoding regions of the genome, obscuring the target genes through which they may act (Claussnitzer et al., 2020). Thirdly, current catalogs of regulatory sequences in the human genome have focused on bulk tissues and lack cell type resolution (Andersson et al., 2014; Meuleman et al., 2020; Moore et al., 2020; Roadmap Epigenomics et al., 2015; Shen et al., 2012).

Innovative approaches have been developed to distinguish causal variants from local variants in linkage disequilibrium (LD) using fine mapping (Schaid et al., 2018; Wakefield, 2009) and to link variants to target genes using co-accessibility of open chromatin regions in single-cells (Pliner et al., 2018) or three-dimensional chromosomal contact-based linkage scores (Nasser et al., 2021). While these advances have made important strides toward the prioritization of causal variants and the prediction of their target genes respectively, we still lack comprehensive cell-type-resolved maps of cCREs from most primary tissues of the adult human body.

1.1.4 Advances in single cell genomics

Single-cell omics technologies, enabled by droplet-based, combinatorial barcoding or other approaches, have now enabled the profiling of transcriptome, epigenome, and chromatin organization from heterogenous tissues at single-cell resolution (Grosselin et al., 2019; Klein et al., 2015; Lake et al., 2018; Luo et al., 2017a; Macosko et al., 2015; Preissl et al., 2018). In

particular, combinatorial cellular barcoding-based assays such as single cell combinatorial-indexing assay for transposase accessible chromatin with sequencing (sci-ATAC-seq; also known as single nucleus ATAC-seq or snATAC-seq) (Cusanovich et al., 2015) have permitted the identification of cCREs in single nuclei from frozen, biobanked tissues without the need for physical purification of individual cell types (Preissl et al., 2018) . The resulting data can be used to deconvolute cell types from mixed cell populations and to dissect cell type-specific transcriptomic and epigenomic states in primary tissues. While sci-ATAC-seq and other single cell assays have been applied to mammalian tissues including murine biosamples (Cusanovich et al., 2018; Lareau et al., 2019; Li et al., 2021; Preissl et al., 2018; Sinnamon et al., 2019), human fetal tissues (Domcke et al., 2020; Trevino et al., 2021), and a few individual adult human organ systems (Chiou et al., 2021; Corces et al., 2020; Hocker et al., 2021; Wang et al., 2020), we still lack comprehensive cell-type-resolved maps of cCREs from most primary tissues of the adult human body.

1.2 Outline of the dissertation

To profile the activity of gene regulatory elements in diverse cell types and tissues in the human body and discover the specific mechanisms of noncoding sequence variants associated with complex diseases, I applied single cell epigenomic methods to a broad survey of human tissues with a focus on the human heart, and used these maps to interpret and test the functions of noncoding variants. First, in Chapter 2, I used single cell epigenomic and transcriptomic methods to define the regulation of gene expression by candidate CREs (cCREs) in nine cell types from the adult human heart. By localizing risk variants for cardiovascular diseases to these cCREs, I uncovered strong enrichments of disease associated variants in cCREs from individual cardiac cell

types, such as atrial fibrillation (AF) variants in cardiomyocyte cCREs. Next, I examined the specific AF risk variants underlying these enrichments, linked them to putative target genes, and tested their molecular mechanisms in human iPSC derived cardiomyocytes using luciferase reporter assays and CRISPR-Cas9 mediated genome editing. These studies revealed two of AF variants affecting a cardiomyocyte-specific cCRE controlling *KCNH2* expression and action potential repolarization.

Using this work as a foundation, in Chapter 3, I next applied single cell epigenomic methods to 30 different tissue types from across the entire adult human body. Integrating these datasets with corresponding data from 15 fetal tissue types revealed the cell type-specificity of over 1 million cCREs in 222 distinct human cell types. Moving beyond cardiovascular diseases and cardiac cell types, I next localized risk variants from the spectrum of complex human diseases and traits to body wide maps of cCREs in human cell types. This analysis resulted in thousands of significant enrichments of risk variants for complex diseases in cCREs of specific cell types. To link specific variants to putative molecular functions, I created a framework that incorporates statistical fine mapping, target gene linkage, and measurements of transcription factor binding site disruption to yield candidate molecular functions for hundreds of distinct noncoding risk variants. I lastly highlight examples of specific variants that may disrupt the activity of cell type-specific cCREs to contribute to complex diseases.

Finally, in Chapter 4 I summarize future directions of this research. First, I outline additional developments that will greatly enhance the utility of these data and frameworks for interpreting the functions of complex disease risk variants. Second, I describe ongoing work to use the healthy tissue datasets I generated as a foundation for uncovering cell type-specific gene regulatory programs in diseased human tissues, with a focus on ischemic heart failure.

Chapter 2: Cardiac cell type-specific gene regulatory programs and disease risk association

2.1 Abstract

Misregulated gene expression in human hearts can result in cardiovascular diseases that are leading causes of morbidity and mortality worldwide. However, the limited information on the genomic location of candidate *cis*-regulatory elements (cCREs) such as enhancers and promoters in distinct cardiac cell types has restricted the understanding of these diseases. Here, we defined >287,000 cCREs in the four chambers of the human heart at single-cell resolution, which revealed cCREs and candidate transcription factors associated with cardiac cell types in a region-dependent manner and during heart failure. We further discovered cardiovascular disease-associated genetic variants enriched within these cCREs including 38 candidate causal atrial fibrillation variants localized to cardiomyocyte cCREs. Additional functional studies revealed that two of these variants affect a cCRE controlling *KCNH2/HERG* expression and action potential repolarization. Overall, this comprehensive atlas of human cardiac cCREs provides the foundation for illuminating cell type-specific gene regulation in human hearts during health and disease.

2.2 Introduction

Disruption of gene regulation is an important contributor to cardiovascular disease, the leading cause of morbidity and mortality worldwide (WHO, 2017). *Cis*-regulatory elements such as enhancers and promoters are crucial for regulating gene expression (Consortium, 2012a; Consortium et al., 2020; Roadmap Epigenomics et al., 2015; Thurman et al., 2012). Mutations in transcription factors and chromatin regulators can result in heart disease (Smemo et al., 2012; Zaidi and Brueckner, 2017), and genetic variants associated with risk of cardiovascular disease are enriched within annotated candidate *cis*-regulatory elements (cCREs) in the human genome

(Maurano et al., 2012). However, a major barrier to understanding the genetic and molecular basis of cardiovascular diseases is the paucity of maps and tools to interrogate gene regulatory programs in the distinct cell types of the human heart. Recent single cell/nucleus RNA-seq (Cui et al., 2019; Litvinukova et al., 2020; Tucker et al., 2020) and spatial transcriptomic (Asp et al., 2019) studies have revealed gene expression patterns in distinct cardiac cell types across developmental and adulthood stages in the human heart, including some which display gene expression patterns that are cardiac chamber/region-specific (Litvinukova et al., 2020; Tucker et al., 2020). However, the transcriptional regulatory programs responsible for cell type-specific and chamber-specific gene expression, and their potential links to non-coding risk variants for cardiovascular diseases and traits, remain to be fully defined.

Candidate *cis*-regulatory elements (cCREs) have been annotated in the human genome with the use of ChIP-seq, DNase-Seq, ATAC-seq, GRO-seq, etc. in a broad spectrum of human tissues including in bulk heart tissues and in purified cardiomyocytes (Consortium, 2012a; Consortium et al., 2020; Dickel et al., 2016; Gilsbach et al., 2018; May et al., 2011; Roadmap Epigenomics et al., 2015; Spurrell et al., 2019; Tan et al., 2020; Thurman et al., 2012; Vierstra et al., 2020). These maps have provided important insights into dynamic gene regulation during heart failure (Gilsbach et al., 2018; Spurrell et al., 2019; Tan et al., 2020) and begun to shed light on the function of non-coding cardiovascular disease variants (Anene-Nzelu et al., 2020; Dickel et al., 2016; Gilsbach et al., 2018; Maurano et al., 2012; Tan et al., 2020). However, major limitations of these studies including their focus on particular chambers/regions of the heart and failure to interrogate *cis*-regulatory elements across all distinct cardiac cell types, have restricted their utility in understanding how specific gene regulatory mechanisms may impact distinct cell types and regions of human hearts in health and disease. Although recent single cell genomic tools provide

the opportunity to interrogate *cis*-regulatory elements at single cell resolution (Buenrostro et al., 2015; Cusanovich et al., 2015; Lareau et al., 2019; Satpathy et al., 2019), their application to mammalian hearts has been limited to one single cell ATAC-seq dataset from adult mouse heart (Cusanovich et al., 2018), fewer than 200 total cells from mouse fetal hearts (Jia et al., 2018), and fetal human heart (Domcke et al., 2020). Thus, to comprehensively investigate *cis*-regulatory elements in the specific cell types of the adult human heart, we profiled chromatin accessibility in ~80,000 heart cells using single nucleus ATAC-seq (snATAC-seq) (Cusanovich et al., 2015; Preissl et al., 2018) and created a comprehensive cardiac cell atlas of cCREs annotated by cell type and putative target genes. Integration of these data with single nucleus RNA-seq datasets from matched specimens revealed gene regulatory programs in nine major cardiac cell types. Using this human cardiac cCRE atlas, we further observed the remodeling of cell type-specific candidate enhancers during heart failure and the enrichment of cardiovascular disease-associated genetic variants in cCREs of specific cell types. Finally, we showed that a cardiomyocyte-specific enhancer harboring risk variants for atrial fibrillation is necessary for cardiomyocyte *KCNH2* expression and regulation of cardiac action potential repolarization.

2.3 Results

2.3.1 Single nucleus analysis of chromatin accessibility and transcriptome in adult human hearts

To assess the accessible chromatin landscape of distinct cardiovascular cell types, we performed snATAC-seq (Preissl et al., 2018), also known as sci-ATAC-seq (Cusanovich et al., 2015), on all cardiac chambers from four adult human hearts without known cardiovascular disease (Table S1). We obtained accessible chromatin profiles for 79,515 nuclei, with a median of 2,682

fragments mapped per nucleus (Fig. 1A, B, Fig. S1, Table S2).

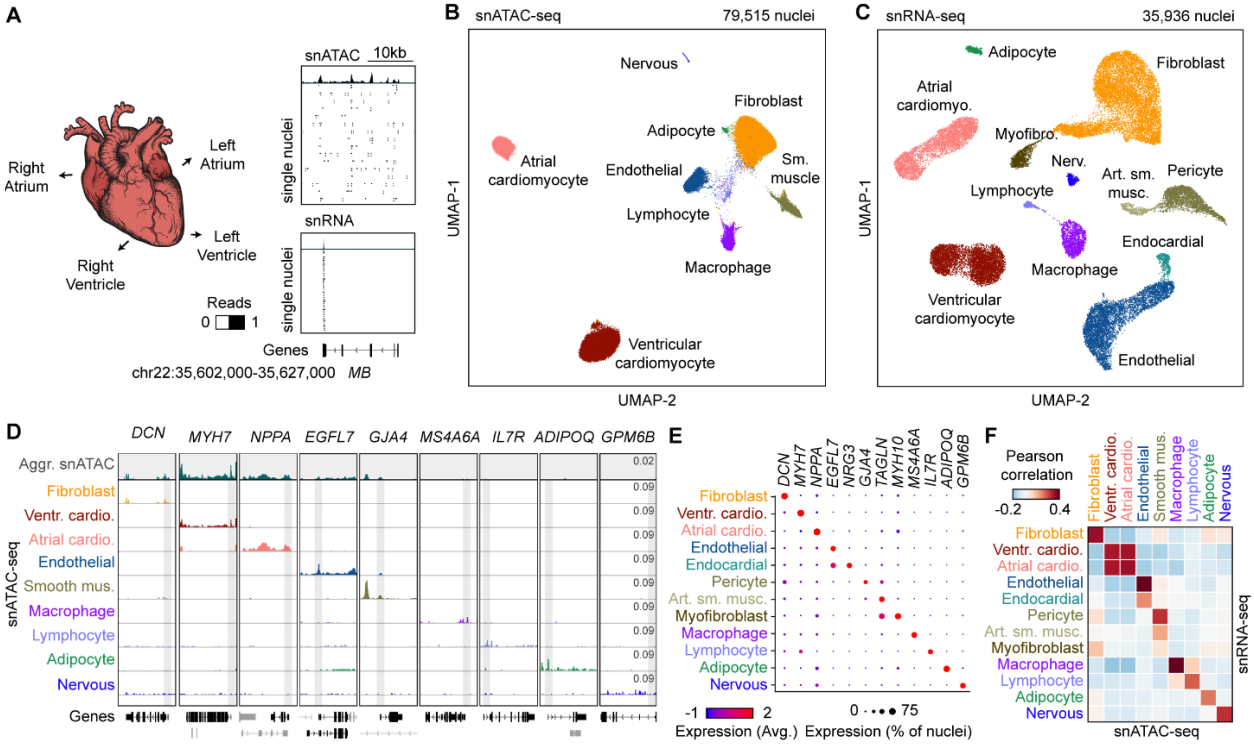


Figure 1: Single-nucleus chromatin accessibility and transcriptome profiling of human hearts. (A) snATAC-seq and snRNA-seq were performed on nuclei isolated from cardiac chambers from four human donors without cardiovascular pathology. snATAC-seq: n = 4 (left ventricle), n = 4 (right ventricle), n = 3 (left atrium), n = 2 (right atrium), snRNA-seq: n = 2 (left ventricle), n = 2 (right ventricle), n = 2 (left atrium), n = 1 (right atrium). (B) Uniform manifold approximation and projection (UMAP) (Leland McInnes, 2018) and clustering analysis of snATAC-seq data reveals nine clusters. Each dot represents a nucleus colored by cluster identity. (C) Uniform manifold approximation and projection (UMAP) (Leland McInnes, 2018) and clustering analysis of snRNA-seq data reveals 12 major clusters. Each dot represents a nucleus colored by cluster identity. Nerv. = Nervous. Art. sm. musc. = arterial smooth muscle. (D) Genome browser tracks (Robinson et al., 2011) of aggregate chromatin accessibility profiles (scale = reads per million; RPM) at selected representative marker gene examples for individual clusters and for all nuclei pooled together into an aggregated heart dataset (top track, grey). Black genes below tracks represent the indicated marker genes, non-marker genes are greyed. (E) Dot plot illustrating expression of representative marker gene examples in individual snRNA-seq clusters. (F) Heatmap illustrating the correlation between clusters defined by chromatin accessibility and transcriptomes. Pearson correlation coefficients were calculated between chromatin accessibility at cCREs within 2 kbp of annotated promoter regions (Harrow et al., 2006) and expression of the corresponding genes for each cluster.

We also performed single nucleus RNA-seq (snRNA-seq) for a subset of the above heart samples to complement the accessible chromatin data and obtained 35,936 nuclear transcriptomes, with a median of 2,184 unique molecular identifiers (UMIs) and 1,286 genes detected per nucleus

(Fig. 1A, C, Fig. S2A-F, Table S3). Using SnapATAC (Fang et al., 2021) and Seurat (Stuart et al., 2019), we identified nine clusters from snATAC-seq (Fig. 1B) and twelve major clusters from snRNA-seq (Fig. 1C, Fig. S2G, H), which were annotated based on chromatin accessibility at promoter regions or expression of known lineage-specific marker genes, respectively (Litvinukova et al., 2020; Tucker et al., 2020) (Fig. 1D, E, Table S4). For example, chromatin accessibility and gene expression of atrial and ventricular cardiomyocyte markers such as *NPPA* and *MYH7* (Ng et al., 2010) were used to classify these two cardiomyocyte subtypes (Fig. 1D, E). Although gene expression patterns of lineage markers strongly correlated with accessibility at promoter regions across annotated cell types (Fig. 1F) and single cell integration analysis (Stuart et al., 2019) revealed 93% concordance in annotation between snATAC-seq and snRNA-seq datasets (Fig. S3, Table S3), some cellular sub-types identified from snRNA-seq including endocardial cells and myofibroblasts were not detected by snATAC-seq (Fig. 1F). Notably, cluster correlation and integration analysis showed that these cell types are present within the snATAC-seq data as part of the endothelial and smooth muscle clusters, respectively (Fig. 1F, Fig. S3, Table S3). The discrepancy in clustering may be attributable to the conservative snATAC-seq clustering parameters or the sparse nature of snATAC-seq data (Chen et al., 2019). Additionally, atrial and ventricular cardiomyocyte nuclei from the left and right regions of the heart could be further clustered by transcriptome but not chromatin accessibility (Fig. S2I, J). We noted that cell type composition varied significantly between biospecimens and donors, highlighting the importance of single cell approaches to limit biases due to cell proportion differences in bulk assays (Fig. S4, Tables S2 and S3). In summary, we identified and annotated cardiac cell types using both chromatin accessibility and nuclear transcriptome profiles.

2.3.2 Identification of candidate cis-regulatory elements (cCREs) in distinct cell types of the human heart

To discover the cCREs in each cell type of the human heart, we aggregated snATAC-seq data from nuclei comprising each cell cluster individually and determined accessible chromatin regions with MACS2 (Zhang et al., 2008). We then merged the peaks from all nine cell clusters into a union of 287,415 cCREs, which covered 4.7% of the human genome (Fig. 2A, Table S5).

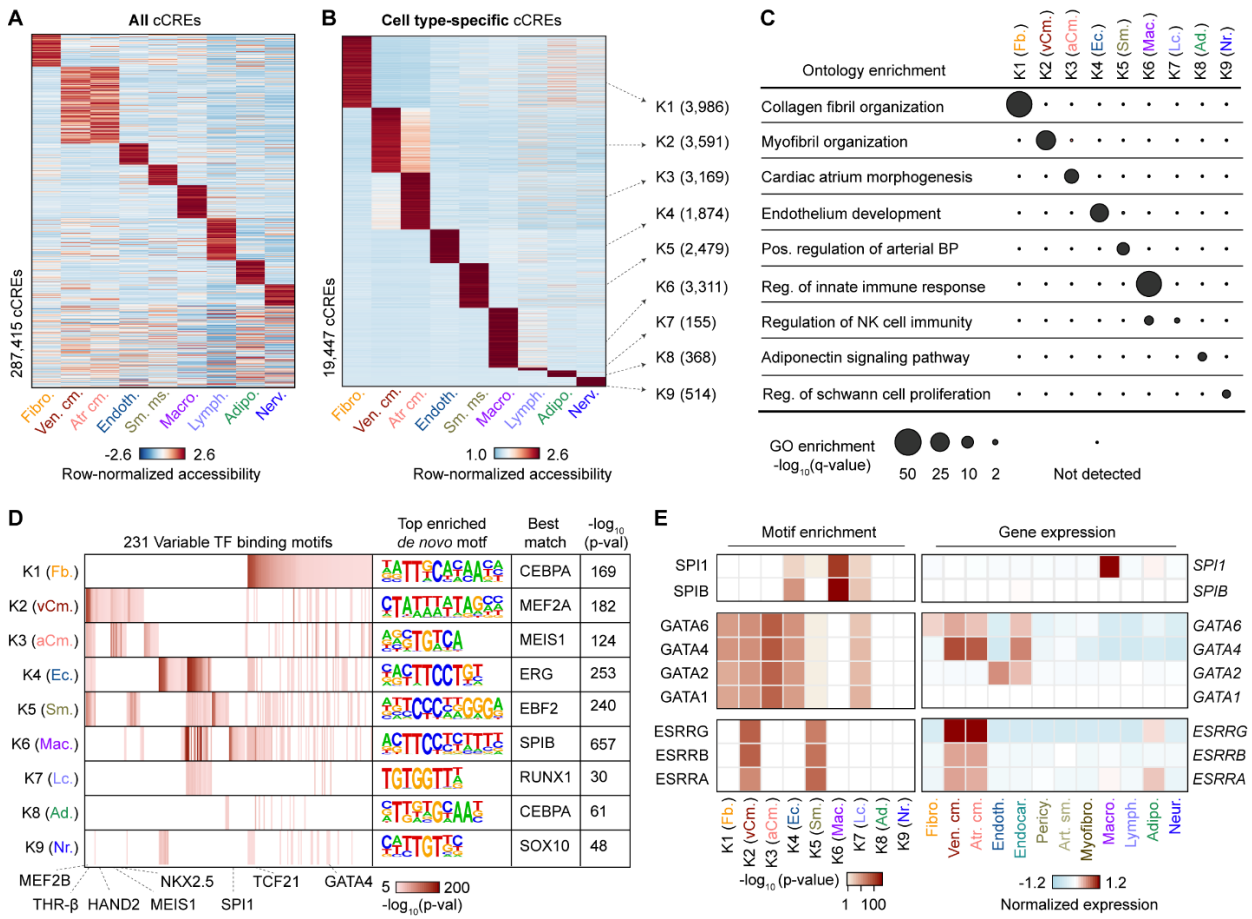


Figure 2: Characterization of gene regulatory programs in cardiac cell types. (A) Heatmap illustrating row-normalized chromatin accessibility values for the union of 287,415 cCREs. K-means clustering was performed to group cCREs based on relative accessibility patterns. (B) Heatmap showing row-normalized chromatin accessibility of 19,447 cell type-specific cCREs (FDR < 0.01 after Benjamini-Hochberg correction; fold change > 1.2). K-means clustering was performed to group cCREs based on relative accessibility patterns. Number of cCREs per K can be found in brackets. (C) GREAT ontology analysis (McLean et al., 2010) of cell type-specific cCREs. Q-value for enrichment indicates Bonferroni adjusted p-value. (D) Transcription factor motif enrichment (Heinz et al., 2010) for known and *de novo* motifs within cell type-specific cCREs. The heatmap in shows motifs with enrichment p-value $<10^{-5}$ in at least one cluster. For *de novo* transcription factor motifs the best matches for the top motifs are displayed. Statistical test for motif enrichment: hypergeometric test. P-values were not corrected for multiple testing. (E) Combination of transcription factor motif enrichment and gene expression shows cell type-specific roles for members of transcription factor families. Displayed are heatmaps for known motif enrichment in cell type-specific cCREs (left) and gene expression across clusters (right). (Fb. = Fibroblast, vCm. = Ventricular Cardiomyocyte, aCm. = Atrial Cardiomyocyte, Ec. = Endothelial, Sm. = Smooth Muscle, Mac. = Macrophage, Lc. = Lymphocyte, Ad. = Adipocyte, Nr. = Nervous).

67.0% of the cCREs identified in the current study overlapped previously annotated cCREs from a broad spectrum of human tissues and cell lines (Consortium et al., 2020) (Fig. S5A), and the union of heart cCREs captured 98.6% and 95.4% of candidate human heart enhancers reported in two previous bulk studies (Dickel et al., 2016; Spurrell et al., 2019) (Fig. S5B, C). Furthermore, 75% of cCREs in the union were at least 2 kbp away from annotated promoter regions, and 19,447 displayed high levels of cell type-specificity (FDR < 0.01, Fig. 2B, Table S6). Gene ontology analysis (McLean et al., 2010) revealed that these cell type-specific cCREs were proximal to genes involved in relevant biological processes, including collagen fibril organization for cardiac fibroblast-specific cCREs (K1), and myofibril organization for ventricular cardiomyocyte-specific cCREs (K2, Fig. 2C, Table S7). Employing chromVAR (Schep et al., 2017) (Table S8) and HOMER (Heinz et al., 2010) (Table S9), we detected cell type-dependent enrichment for 231 transcription factor binding signatures, such as MEF2A/B, NKX2.5, and THR-β sequence motifs in cardiomyocyte-specific cCREs and TCF21 motifs in cardiac fibroblast-specific cCREs (Fig. 2D). To discover the transcription factors that may bind to these sites, we combined corresponding snRNA-seq data with sequence motif enrichments to correlate expression of these transcription factors with motif enrichment patterns across cell types (Fig. 2E). As an example, we found strong enrichment of the binding motif for the macrophage transcription factor SPI1/PU.1 (Zhang et al., 1994) in macrophage-specific cCREs, and *SPI1* was exclusively expressed in macrophages (Fig. 2E, Tables S4). In addition, we observed that transcription factor family members were expressed in cell type-specific combinations. For instance, while GATA family members displayed similar motif enrichment patterns across sets of cell type-specific cCREs, we discovered that endothelial cells and cardiac fibroblasts expressed *GATA2* and *GATA6*, respectively, whereas cardiomyocytes

expressed both *GATA4* and *GATA6,* and endocardial cells expressed *GATA2*, *GATA4*, and *GATA6* (Fig. 2E, Tables S4). In summary, these results establish a resource of candidate *cis*-regulatory elements for interrogation of cardiac cell type-specific gene regulatory programs.

2.3.3 Cardiac cell type-specific gene regulatory programs implicated in chamber-specific structure and function

Each cardiac chamber performs a unique role that is crucial to system-level heart function (Moorman and Christoffels, 2003). To investigate the gene regulatory programs underlying chamber-specific gene expression and cellular functions in distinct cardiac cell types, we tested cCREs for differential accessibility across five of the most abundant cell types of the heart: cardiomyocytes, cardiac fibroblasts, endothelial cells, smooth muscle cells, and macrophages. We discovered 16,451 differentially accessible (DA) cCREs between pooled atria and ventricles, the majority of which were detected in cardiomyocytes (Fig. 3A-C, Table S10).

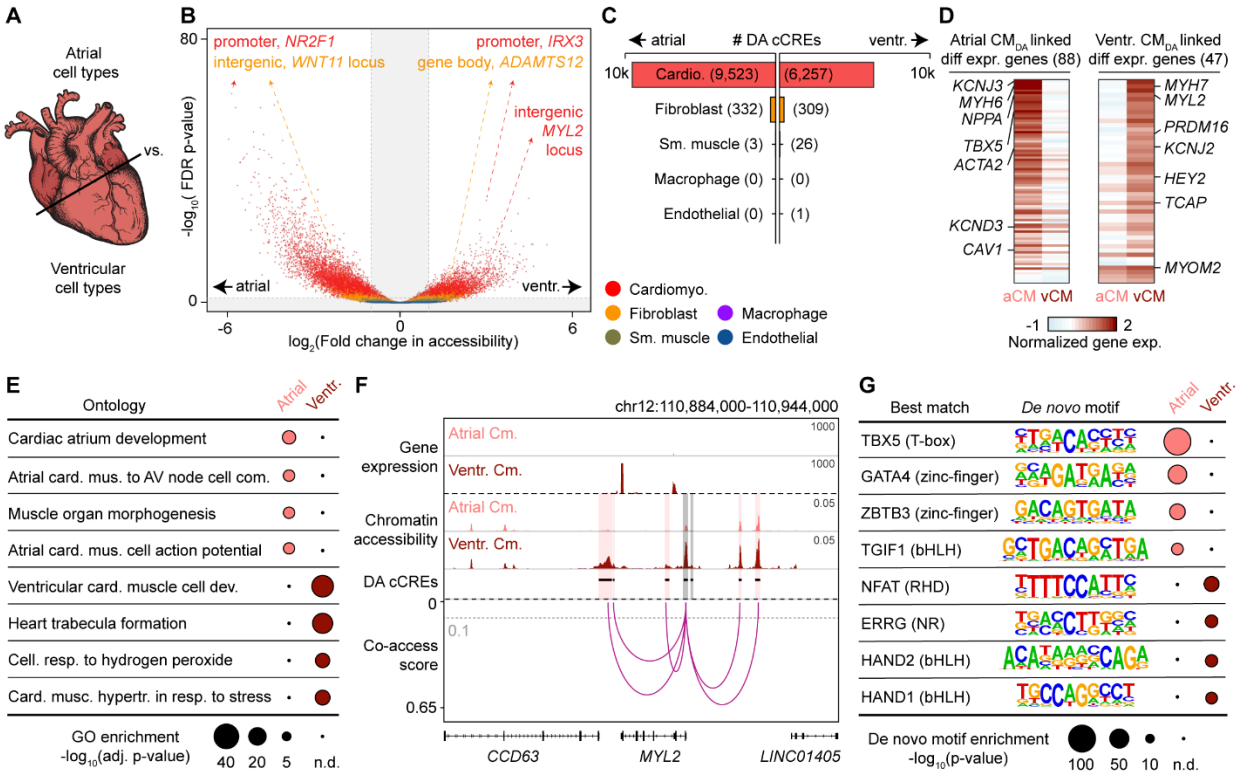


Figure 3: Cardiomyocyte cCREs display chamber-dependent differences in chromatin accessibility. (A) Scheme for comparison of major cell types across heart chambers. (B) Volcano plot showing differentially accessible (DA) candidate *cis*-regulatory elements (cCREs) in each cell type between atria and ventricles. Each dot represents a cCRE and the color indicates the cell type. cCREs with $\log_2$(fold change) > 1 and FDR < 0.05 after Benjamini-Hochberg correction (outside the shaded area) were considered as DA. (C) Number of DA cCREs between atria and ventricles by cell type. (D) Heatmaps showing normalized gene expression levels of differentially expressed genes between atrial (aCM) and ventricular cardiomyocytes (vCM) that were linked by co-accessibility to distal DA cCREs that were more accessible in atrial cardiomyocytes (Atrial CM_{DA}) or ventricular cardiomyocytes (Ventr. CM_{DA}), respectively. (E) GREAT ontology analysis (McLean et al., 2010) of DA cCREs between atrial and ventricular cardiomyocytes. P-values shown are Bonferroni adjusted (n.d.: not detected). (F) Genome browser tracks (Robinson et al., 2011) showing chromatin accessibility (scale = reads per million; RPM) and gene expression (scale = reads per kilobase million; RPKM) in atrial and ventricular cardiomyocytes as well as DA cCREs that were co-accessible with the promoter of *MYL2*. Grey dotted line indicates co-accessibility threshold (> 0.1). Red boxes: distal DA cCREs co-accessible with *MYL2* promoter. Gray boxes: DA cCREs overlapping the promoter region of *MYL2*. (G) Transcription factor motif enrichment analysis (Heinz et al., 2010) of DA cCREs between atrial and ventricular cardiomyocytes. The best matches for the top *de novo* motifs (score > 0.7) are shown. Statistical test for motif enrichment: hypergeometric test. P-values were not corrected for multiple testing (n.d.: not detected).

Specifically, 11,159 cCREs displayed differential accessibility between right atrium and right ventricle and 12,962 cCREs exhibited differential accessibility between left atrium and left ventricle (Fig. S6A-C, Table S10). Comparing the left and right sides of the heart, we identified

101 DA cCREs between the right and left ventricle (Fig. S6D), and 2,687 DA cCREs between left and right atria, which in contrast to comparisons between atria and ventricles were found primarily in cardiac fibroblasts (Fig. S6E, Table S10).

Utilizing co-accessibility analysis (Pliner et al., 2018) to link distal DA cCREs (~88% of all DA cCREs) to their putative target genes (Table S11, median distance: 88.7 kbp), we observed that distal DA cCREs in cardiomyocytes between atria and ventricles were associated with chamber-specific gene expression of their putative target genes (Fig. 3D, Figure S15B-E, Table S12), and genes near these DA cCREs were enriched for chamber-specific biological processes (Fig. 3E, Figure S15B-E, Table S13). Specifically, distal DA cCREs with higher accessibility in atrial cardiomyocytes were associated with genes such as *PITX2,* a transcriptional regulator of cardiac atrial development (Liu et al., 2001), as well as the ion channel subunit *SCN5A* which regulates cardiomyocyte action potential (Rivaud et al., 2020) (Fig. 3E, Table S13). Furthermore, we found distal DA cCREs with higher accessibility in atrial cardiomyocytes at the *HAMP* gene locus, which encodes a key regulator of ion homeostasis and was recently described as a potential novel cardiac gene in the right atrium by single nucleus transcriptomic analysis (Litvinukova et al., 2020; Tucker et al., 2020). Conversely, genes near distal DA cCREs with higher accessibility in ventricular cardiomyocytes were enriched for biological processes such as trabecula formation and ventricular cardiac muscle cell differentiation. For example, several distal DA cCREs with increased accessibility in ventricular cardiomyocytes compared to atrial cardiomyocytes were linked to the promoter region of *MYL2*, which encodes the ventricular isoform of myosin light chain 2 (Veevers et al., 2018) (Fig. 3F, Table S4), a regulator of ventricular cardiomyocyte sarcomere function.

Additionally, analysis of distal DA cCREs in cardiac fibroblasts revealed that putative target genes were involved in distinct biological processes between right and left atria. In particular, we found that DA cCREs with higher accessibility in right atrial cardiac fibroblasts were proximal to genes involved in heart development, heart growth, and tube development, whereas DA cCREs with higher accessibility in left atrial cardiac fibroblasts were adjacent to genes involved in biological processes such as wound healing and vasculature development (Fig. S6E, Table S13). We further found a cardiac fibroblast-specific DA cCRE with higher accessibility in left atria at the fibrinogen *FN1* gene locus, potentially indicating a more activated fibroblast state (Hortells et al., 2019; Litvinukova et al., 2020). Supporting these findings, we identified several other DA cCREs with higher accessibility in left atrial cardiac fibroblasts adjacent to genes involved in generation of extracellular matrix (ECM) such as *MMP2* and *FBLN2* (Table S13). These observations are consistent with previous findings that a higher fraction of ECM is produced in fibroblasts of the left atrium (Litvinukova et al., 2020).

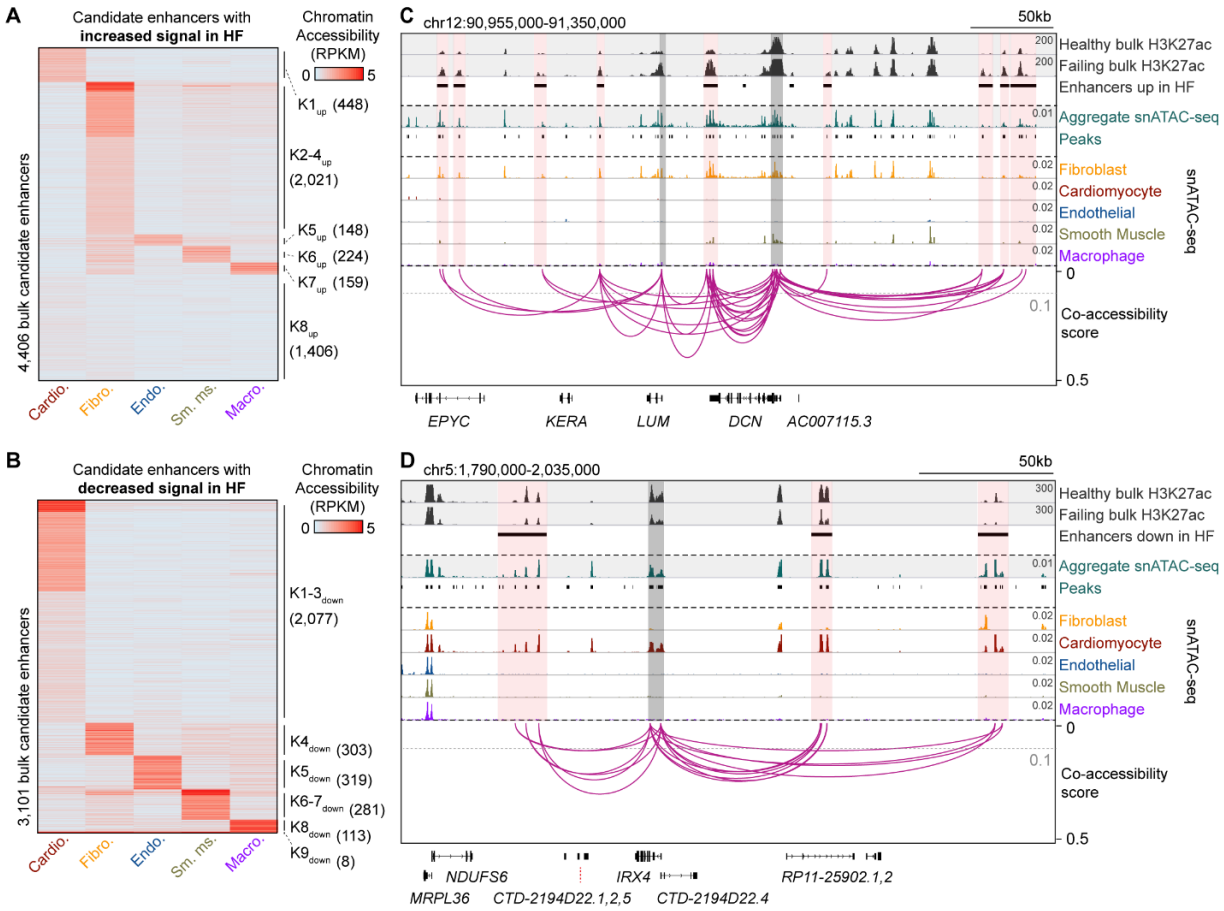
Using motif enrichment analysis, we inferred candidate transcriptional regulators involved in chamber-specific cellular specialization, including TBX5, GATA4, and TGIF1 for atrial cardiomyocytes, and NFAT, ERRG, HAND1, and HAND2 for ventricular cardiomyocytes (Fig. 3G, Table S14). While the TBX5 DNA binding motif was strongly enriched in both right and left atrial cardiomyocyte DA cCREs, the NFAT5 motif ranked highest in left ventricular cardiomyocyte DA cCREs and the TBX20 motif was strongly enriched in right ventricular cardiomyocyte DA cCREs (Fig. S6B, C, Table S14). Furthermore, cardiac fibroblast DA cCREs with higher accessibility in the right atrium were enriched for the binding motif of forkhead transcription factors (Fig. S6E), whereas cardiac fibroblast DA cCREs with higher accessibility in the left atrium were enriched for the homeobox transcription factor CUX1 motif (Fig. S6E, Table

S14). Altogether, we identified cCREs and candidate transcription factors associated with specific cardiac chambers, particularly within cardiomyocytes and cardiac fibroblasts.

2.3.4 Cell type specificity of candidate enhancers associated with heart failure

Recent large-scale studies profiling the H3K27ac histone modification in human hearts have uncovered candidate enhancers associated with heart failure (Spurrell et al., 2019; Tan et al., 2020). However, because these studies either examined heterogeneous bulk heart tissue (Spurrell et al., 2019; Tan et al., 2020) or focused solely on enriched cardiomyocytes (Gilsbach et al., 2018), it remains unclear what role, if any, additional cardiac cell types and cCREs may contribute to heart failure pathogenesis. Using our cell atlas of cardiac cCREs, we revealed the cell type specificity of candidate enhancers showing differential H3K27ac signal strength between human hearts from healthy donors and donors with dilated cardiomyopathy (heart failure) (Spurrell et al., 2019) (Fig. 4, Fig. S7A-E).

Figure 4: Cell type specificity of candidate enhancers associated with heart failure. A) Cell type-specificity of 4,406 candidate enhancers with increased H3K27ac signal in failing left ventricles (Spurrell et al., 2019). Heatmap displays cell type-resolved chromatin accessibility RPKM (reads per kilobase per million mapped reads) values for cell types from left ventricular snATAC-seq datasets. Candidate enhancers were grouped based on chromatin accessibility patterns across cell clusters using K-means. (B) Cell type-specificity of 3,101 candidate enhancers with decreased H3K27ac signal in failing left ventricles (Spurrell et al., 2019). (C) Genome browser tracks (Robinson et al., 2011) showing several candidate enhancers with increased activity during heart failure (HF) that were primarily accessible in fibroblasts and co-accessible with the promoters of *LUM* and/or *DCN*. For visualization, linkages between cCREs within candidate enhancers and all gene promoters are shown (co-accessibility > 0.1, grey dotted line). Candidate enhancers co-accessible with gene promoters are indicated by red shaded boxes and promoter regions are indicated by grey shaded boxes (scale = RPM) (D) Genome browser tracks (Robinson et al., 2011) showing several bulk candidate enhancers with decreased activity in heart failure that were primarily accessible in cardiomyocytes and co-accessible with the promoter of *IRX4* (scale = RPM). (E, F) Transcription factor motif enrichment (Heinz et al., 2010) in the candidate enhancers with (E) increased and (F) decreased activity in failing left ventricles. Analysis was performed on the indicated K cluster(s) from panels (A) and (B) respectively. The best matches for selected *de novo* motifs (score > 0.7) are shown. Statistical test for motif enrichment: hypergeometric test. P-values were not corrected for multiple testing.



E Candidate enhancers with **increased signal in HF**

	Enriched *de novo* motifs	Best match	-log$_{10}$ (*p* val)
		TEAD3 (TEF)	20
K1$_{up}$ (Cm.)		MYF6 (bHLH)	14
		NKX2-3 (NKX)	13
		AP4 (bHLH)	111
K2-4$_{up}$ (Fb.)		FOS:JUNB (bZIP)	98
		SMAD3 (SMAD)	18
K5$_{up}$ (Ec.)		EWS:ERG (ETS)	18
K6$_{up}$ (Sm.)		EBF (EBF)	19
K7$_{up}$ (Mac.)		SPIC (ETS)	33

F Candidate enhancers with **decreased signal in HF**

	Enriched *de novo* motifs	Best match	-log$_{10}$ (*p* val)
		MEF2A (MEF2)	127
K1-3$_{down}$ (Cm.)		GRE (NR)	72
		NF1-halfs. (CTF)	71
		CEBPA (bZIP)	59
K4$_{down}$ (Fb.)		AP1 (bZIP)	29
		PGR (NR)	21
K5$_{down}$ (Ec.)		ATF3 (bZIP)	54
K6-7$_{down}$ (Sm.)		PGR (NR)	39
K8$_{down}$ (Mac.)		ELF3 (ETS)	15

We observed that a large fraction of candidate enhancers that displayed increased activity (45%) during heart failure were accessible primarily in cardiac fibroblasts (Fig. 4A, K2-$4_{up,}$ Table S15), whereas a majority of those exhibiting decreased activity (67%) were accessible primarily in cardiomyocytes (Fig. 4B, K1-3_{down}, Table S15). Candidate enhancers with increased activity in cardiac fibroblasts were proximal to genes involved in extracellular matrix organization and connective tissue development (Fig. 4A, K2-4_{up}, Table S16), whereas those exhibiting decreased activity in cardiomyocytes were proximal to genes involved in regulation of heart contraction and cation transport (Fig. 4B, K1-3_{down}, Table S16). For example, several of these cardiac fibroblast candidate enhancers were present at loci encoding the extracellular matrix proteins lumican (*LUM*) and decorin (*DCN*) and co-accessible with the promoters of these genes (Fig. 4C). Consistent with these findings, both genes were primarily expressed in cardiac fibroblasts (Fig. S7F, Table S4), and *LUM* has been reported to exhibit increased expression in failing hearts compared to control hearts (Spurrell et al., 2019). On the other hand, several cardiomyocyte candidate enhancers displaying decreased activity in heart failure were co-accessible with the promoter region of *IRX4* (Fig. 4D), which encodes a ventricle-specific transcription factor (Bruneau et al., 2000) and is primarily expressed in cardiomyocytes of the left ventricle (Fig. S7G, Table S4).

To identify potential transcription factors regulating these pathologic responses during heart failure, we performed motif enrichment analysis in cell type-specific subsets of enhancers showing differential H3K27ac signal strength between healthy and failing hearts (Table S17). For candidate enhancers exhibiting increased activity in heart failure, we identified enrichment of not only bHLH motifs such as AP4 in cardiac fibroblast candidate enhancers which matched previous bulk analysis (Spurrell et al., 2019) (Fig. 4E, K2-4_{up}), but also TEAD3 and MYF6 motifs in cardiomyocyte candidate enhancers (Fig. 4E, $K1_{up}$). Conversely, for candidate enhancers

displaying decreased activity in heart failure, we observed enrichment of nuclear receptor motifs such as glucocorticoid response element (GRE) in cardiomyocyte candidate enhancers, which is consistent with previous findings (Spurrell et al., 2019) (Fig. 4F, K1-3_{down}), as well as other motifs which were not detected in bulk analyses, such as the bZIP transcription factor CEBPA for cardiac fibroblast candidate enhancers (Fig. 4F, $K4_{down}$). Thus, these results show that this cardiac cell atlas of cCREs may be used to assign disease-associated candidate enhancers from bulk assays to their affected cell types and infer transcriptional regulators involved in lineage-specific disease pathogenesis.

2.3.5 Interpreting non-coding risk variants of cardiac diseases and traits

Non-coding genetic variants contributing to risk of complex diseases are enriched within cCREs in a tissue and cell type-dependent manner (Chiou et al., 2021; Corces et al., 2020; Cusanovich et al., 2018; Hook and McCallion, 2020; Nott et al., 2019). To examine the enrichment of cardiovascular disease variants within cCREs active in cardiac cell types, we performed cell type-stratified LD (Linkage disequilibrium) score regression analysis (Bulik-Sullivan et al., 2015) using GWAS summary statistics for cardiovascular diseases (Arvanitis et al., 2020; Malik et al., 2018; Nelson et al., 2017; Nielsen et al., 2018; Shadrina et al., 2019) (Fig. 5A) and control traits (Fig. S8A, Table S18) by measuring the enrichment of disease-associated variants within all cCREs identified for each cell type. This analysis revealed significant enrichment of atrial fibrillation (AF)-associated variants in both atrial (Z = 5.61, FDR = 1.9e-6) and ventricular cardiomyocyte cCREs (Z = 6.80, FDR = 2.8e-9), varicose vein-associated variants in endothelial cell cCREs (Z = 4.36, FDR = 3.9e-4), and coronary artery disease-associated variants in cardiac fibroblast cCREs (Z = 3.29, FDR = 1.7e-2, Fig. 5A). Notably, except for atrial

fibrillation, these associations were not significant in a pseudobulk heart dataset created by combining chromatin accessibility profiles from all cardiac cell types (Fig. 5A). Furthermore, cardiovascular disease variants were not significantly enriched in accessible chromatin from non-cardiac tissues (Consortium, 2012a; Consortium et al., 2020; Vierstra et al., 2020) or human lung cell types (Wang et al., 2020), with the exception of a significant enrichment of varicose vein-associated variants in endothelial cells (Fig. S8B, C).

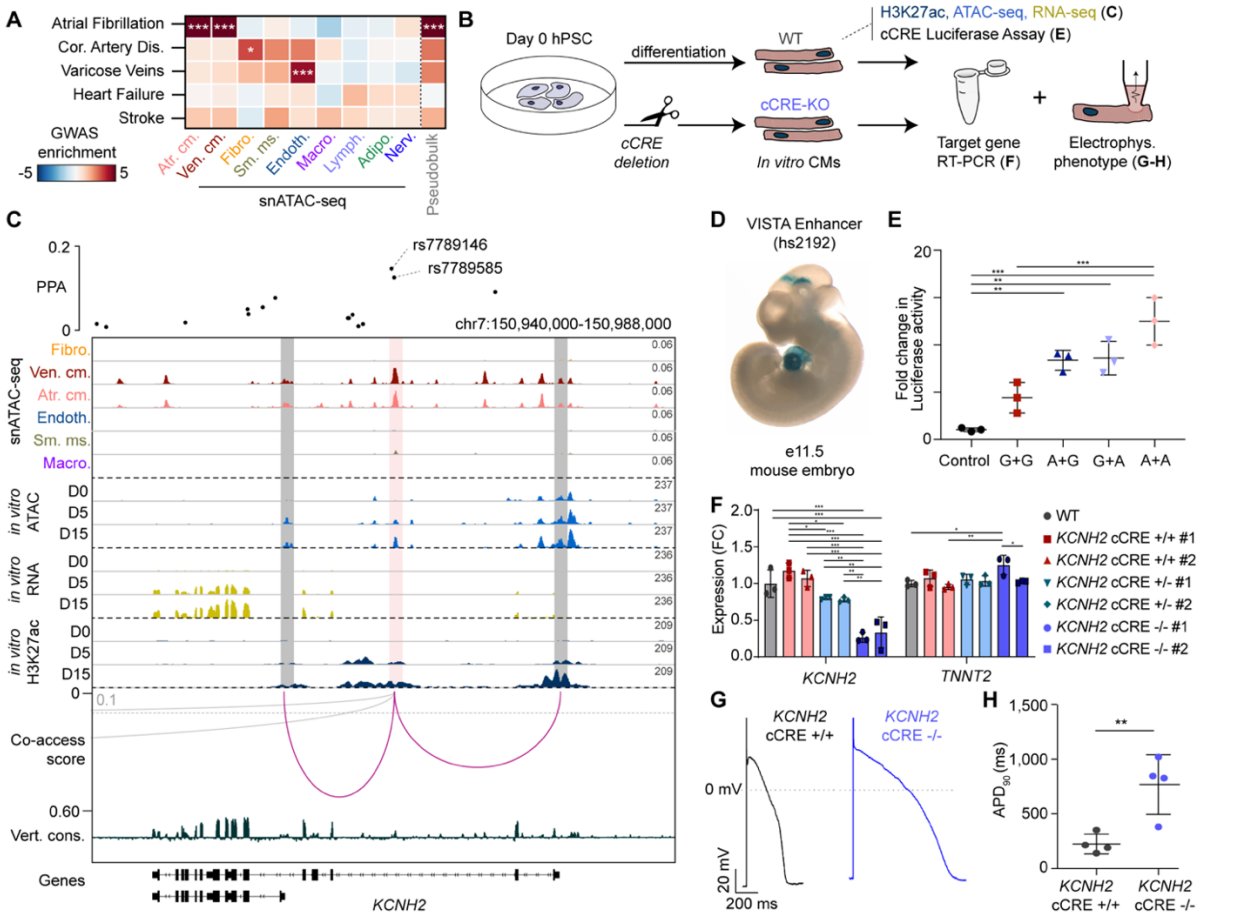


Figure 5: Identification and characterization of atrial fibrillation-associated variants at the KCNH2 locus. (A) Enrichment of cardiovascular disease variants within cardiac cell type cCREs. Z-scores are shown and were used to compute two-sided p-values for enrichments. * = FDR < 0.05, *** = FDR < 0.001. (B) Cardiomyocyte differentiation model schematic. hPSC = human pluripotent stem cell. (C) Fine mapping (Wakefield, 2009) and molecular characterization of two atrial fibrillation (AF) variants. Genome browser tracks (Robinson et al., 2011) for snATAC-seq (top; scale = RPM) and indicated molecular features during hPSC-cardiomyocyte differentiation (scale = RPKM). Co-accessibility track shows linkages between the AF variant-containing cCRE and promoters (cutoff > 0.1). PPA = Posterior probability of association (Wakefield, 2009). (D) Transgenic mouse embryo showing LacZ reporter expression under control of a genomic region (hs2192, Vista database (Visel et al., 2007)) overlapping the variant-cCRE pair. (E) Luciferase reporter assay for the AF variant-harboring cCRE in D15 cardiomyocytes. Genotypes for rs7789146 and rs7789585 were either both G (risk), both A (non-risk), or a combination. Each dot represents luciferase activity (average of two transfections) in independent replicates of D15 cardiomyocytes. Data are mean +/- SD. *** $p < 0.001$, ** $p < 0.01$ (one-way ANOVA and Tukey post hoc test). Control: minimal promoter. (F) Expression of *KCNH2* and *TNNT2* in D25 cardiomyocytes after CRIPSR/Cas9-mediated cCRE deletion. Each dot represents an independent cardiomyocyte differentiation. Data are mean +/- SD. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, (one-way ANOVA and Tukey post hoc test); WT = unperturbed control. (G) Action potential recordings in hPSC-derived cardiomyocytes with and without cCRE deletion at D25-35. (H) APD_{90} (action potential duration at 90% depolarization) at 1 Hz pacing for 4 independent hPSC derived cardiomyocytes with and without cCRE deletion at D25-35. Data are mean +/- SD. ** $p < 0.01$ (unpaired two-sided t-test).

Next, to identify likely causal AF risk variants in cardiomyocyte cCREs, we first determined the probability that variants were causal for AF (Posterior probability of association, PPA) at 111 known loci using Bayesian fine-mapping (Wakefield, 2009). We then intersected fine-mapped AF variants with cCREs detected in atrial and/or ventricular cardiomyocytes and identified 38 variants with PPA > 10% in cardiomyocyte cCREs, including previously reported variants at the *HCN4* (Dickel et al., 2016) and *SCN10A/SCN5A* (van den Boogaard et al., 2014) loci (Table S19). We further prioritized AF variants for molecular characterization based on their overlap with cCREs that were primarily accessible in cardiomyocytes, evolutionarily conserved, and co-accessible with promoters of genes expressed in cardiomyocytes. In order to experimentally validate the molecular functions of cCREs containing AF variants, we utilized a human pluripotent stem cell (hPSC)-derived cardiomyocyte differentiation model system (Zhang et al., 2019c) (Fig. 5B). From the variant prioritization analysis, we discovered a cCRE in the second intron of the potassium channel gene *KCNH2* (*HERG*) which was co-accessible with the *KCNH2* promoter (Fig. 5C) and harbored two variants, rs7789146 and rs7789585, with a combined PPA of 28% (Fig. 5C, Fig. S9A). *KCNH2* was primarily expressed in ventricular and atrial cardiomyocytes in the adult human heart (Fig S9B). The cCRE appeared to be activated during hPSC-cardiomyocyte differentiation as evidenced by an increase in H3K27ac signal that correlated with *KCNH2* expression (Fig. 5C). Supporting its *in vivo* role in regulating gene expression in mammalian hearts, a genomic region (hs2192) (Visel et al., 2007) containing this cCRE was previously shown to drive LacZ reporter expression in mouse embryonic hearts (Visel et al., 2007) (Fig. 5D).

2.3.6 A cardiomyocyte enhancer of *KCNH2* is affected by non-coding risk variants associated with atrial fibrillation

To investigate whether these AF variants may affect enhancer activity and thereby regulate *KCNH2* expression and cardiomyocyte electrophysiologic function, we initially carried out reporter assays using a hPSC cardiomyocyte model system. Results from these studies confirmed that in D15 hPSC-cardiomyocytes, the *KCNH2* enhancer carrying the rs7789146-G/rs7789585-G AF risk allele displayed significantly weaker enhancer activity than when containing the non-risk variants (Fig. 5E, Fig. S9C), thus supporting the functional significance of these AF variants. We next used CRISPR/Cas9 genome editing strategies to remove the enhancer and performed qPCR and electrophysiologic assays to examine its role in *KCNH2* expression and function. Supporting the aforementioned findings, CRISPR/Cas9 genome deletion of this cCRE in hPSC-cardiomyocytes resulted in decreased *KCNH2* expression in an enhancer dosage-dependent manner (Fig. 5F, Fig. S9D). Similar to human cardiomyocytes with loss of *KCNH2* function due to mutations in the *KCNH2* coding sequence (Curran et al., 1995) or gene knockdown (Jones et al., 2014), cellular electrophysiologic studies demonstrated that these cCRE-deleted hPSC cardiomyocytes displayed a significantly prolonged action potential duration (Fig. 5G, H), thus suggesting that cardiac repolarization abnormalities in atrial cardiomyocytes may lead to AF in an analogous manner to ventricular arrhythmias due to long QT syndrome (Curran et al., 1995). Taken together, these results highlight the utility of this single cell atlas for assigning non-coding cardiovascular disease risk variants to distinct cell types and affected cCREs, and functionally interrogating how these variants may contribute to cardiovascular disease risk.

2.4 Discussion

The limited ability to interrogate cell type-specific gene regulation in the human heart has been a major barrier for understanding molecular mechanisms of cardiovascular traits and diseases. Here, we report a cell type-resolved atlas of cCREs in the human heart, which was ascertained by profiling accessible chromatin in individual nuclei from all four chambers of multiple human hearts and includes both cell type-specific and heart chamber-specific cCREs. In particular, we observed chamber-specific differences in chromatin accessibility between ventricles and atria as well as left and right atria but notably detected few differences between left and right ventricles. This finding is consistent with recent single nucleus RNA-seq analysis in human hearts which found few differentially expressed genes between left and right ventricles (Tucker et al., 2020). We note that the power to detect chamber-specific differences in chromatin accessibility depends on the number of samples assayed and the total nuclei used as input for differential analysis. Thus, future studies with larger cohorts will likely reveal additional chamber-specific differences between cardiac cell types.

We further highlight the utility of this atlas of heart cCREs to provide new insight into aberrant gene regulation during cardiovascular pathology. To this end, we delineated the cell type-specificity of enhancers which were differentially active between healthy and failing heart tissue (Spurrell et al., 2019) and identified additional transcription factors that may be involved in the pathogenesis of specific cell types during heart failure. Such cell type-specific analysis is particularly important in the context of heart failure because cellular composition can differ between diseased and control hearts (Gilsbach et al., 2014; Gilsbach et al., 2018). This change in cellular composition may in part explain the cell type bias that we observed between candidate enhancers exhibiting increased and decreased activity during heart failure (i.e. cardiac fibroblasts

and cardiomyocytes, respectively). However, due to the large differences in H3K27ac signal, we suspect that measured changes in candidate enhancer activity could be due to a combination of both enhancer remodeling and shift in cell type composition. Thus, future studies profiling snATAC-seq and H3K27ac in parallel from the same cardiac sample or novel approaches to profile histone modifications in single nuclei (Kaya-Okur et al., 2019; Wang et al., 2019) will provide greater insight into the extent of changes in chromatin accessibility and enhancer activity in individual cardiac cell types from diseased hearts.

Finally, we show how this atlas can be used to not only assign non-coding genetic variants associated with cardiovascular disease risk to cCREs in specific cardiac cell types, but also illuminate their cellular and molecular consequences. In particular, we discovered significant enrichment of AF-associated variants within cardiomyocyte cCREs and functionally interrogated one of these cCREs by demonstrating its role in regulating *KCNH2* expression and cardiomyocyte repolarization. Similar to electrophysiologic phenotypes of human cardiomyocytes exhibiting *KCNH2* loss of function (Curran et al., 1995; Jones et al., 2014), hPSC-cardiomyocytes harboring deletions of this cCRE displayed action potential prolongation, suggesting that cardiac repolarization abnormalities may contribute to atrial fibrillation, possibly through similar mechanisms as to how they may contribute ventricular arrhythmias (Curran et al., 1995). On the other hand, we found no enrichment of variants associated with heart failure in any cardiac cell type. This finding may reflect the heterogeneous etiologies of cardiovascular diseases and the limited number of currently known risk loci for heart failure (Arvanitis et al., 2020). Future GWAS in large cohorts with detailed phenotyping, including biobanks such as the UK Biobank (Sudlow et al., 2015) and the BioBank Japan Project (Ishigaki et al., 2020) and whole genome sequencing efforts such as the NHLBI Trans-Omics for Precision Medicine (TOPMed) program (NHLBI,

2014), will help identify and refine disease association signals. Therefore, this atlas of cardiac cCREs will be a valuable resource for continued discovery of regulatory elements, target genes, and specific cell types that may be affected by non-coding cardiovascular genetic variants.

In summary, we created a human heart cell atlas of >287,000 cCREs, which may serve as a reference to further expand our knowledge of gene regulatory mechanisms underlying cardiovascular disease. To facilitate distribution of these data, we created a web portal at: http://catlas.org/humanheart. Integrating this resource with genomic and epigenomic clinical cardiac datasets, we built a systematic framework to interrogate how *cis*-regulatory elements and genetic variants might contribute to cardiovascular diseases such as heart failure or atrial fibrillation. Overall, such information will have great potential to provide new insight into the development of future cardiac therapies that are tailored to affected cell types and thus optimized for treating specific cardiovascular diseases.

2.5 Materials and Methods

2.5.1 Experimental Design

We performed single nucleus ATAC-seq to define a comprehensive catalogue of candidate *cis*-regulatory elements (cCREs) for the cell types in four regions of non-failing human hearts and generated in parallel snRNA-seq datasets for a subset to delineate gene expression patterns. We used the cCRE catalogue to computationally assign dynamic enhancers in failing hearts to cell types and to assign cardiovascular disease risk variants to cCREs in individual cardiac cell types. Finally, we applied reporter assays, genome editing and electrophysiogical measurements in *in vitro* differentiated human cardiomyocytes to validate the molecular mechanisms of cardiovascular disease risk variants.

2.5.2 Human Tissues

Adult human heart tissues were procured at the time of organ donation using an Institutional Review Board protocol (No. 101021) approved by the University of California, San Diego. Donated hearts were perfused with cold cardioplegia prior to cardiectomy and then explanted immediately into an ice-cold physiologic solution as we previously described (Smyth et al., 2010). Full-thickness samples from each chamber were obtained and epicardial fat rapidly removed before immediately flash freezing samples in liquid nitrogen. Samples were received from the United Network for Organ Sharing. Limited clinical data was obtained for each heart per approved Institutional Review Board protocol (Table S1). All samples were stored at -80˚C until processing.

2.5.3 Single nucleus ATAC-seq

Combinatorial barcoding single nucleus ATAC-seq was performed as described previously (Cusanovich et al., 2015; Fang et al., 2021; Preissl et al., 2018) with slight modifications. Nuclei were isolated in gentleMACS M tubes (Miltenyi) on a gentleMACS Octo Dissociator (Miltenyi) using the "Protein_01_01" protocol in MACS buffer (5 mM $CaCl_2$, 2 mM EDTA, 1X protease inhibitor (Roche, 05-892-970-001), 300 mM MgAc, 10 mM Tris-HCL pH 8, 0.6 mM DTT). Nuclei were pelleted with a swinging bucket centrifuge (500 x g, 5 min, 4°C; 5920R, Eppendorf) and resuspended in 1 mL Nuclear Permeabilization Buffer (1X PBS, 5% Bovine Serum Albumin, 0.2% IGEPAL CA-630 (Sigma), 1 mM DTT, 1X Protease inhibitor). Nuclei were rotated at 4 ˚C for 5 minutes before being pelleted again with a swinging bucket centrifuge (500 x g, 5 min, 4°C; 5920R, Eppendorf). After centrifugation, permeabilized nuclei were resuspended in 500 µL high

salt tagmentation buffer (36.3 mM Tris-acetate (pH = 7.8), 72.6 mM potassium-acetate, 11 mM Mg-acetate, 17.6% DMF) and counted using a hemocytometer. Concentration was adjusted to 2,000 nuclei/9 µl, and 2,000 nuclei were dispensed into each well of a 96-well plate per sample (96 tagmentation wells/sample, samples were processed in batches of 2-4 samples). For tagmentation, 1 µL barcoded Tn5 transposomes (Table S20) were added using a BenchSmart™ 96 (Mettler Toledo), mixed five times, and incubated for 60 min at 37 °C with shaking (500 rpm). To inhibit the Tn5 reaction, 10 µL of 40 mM EDTA (final 20mM) were added to each well with a BenchSmart™ 96 (Mettler Toledo) and the plate was incubated at 37 °C for 15 min with shaking (500 rpm). Next, 20 µL of 2x sort buffer (2 % BSA, 2 mM EDTA in PBS) were added using a BenchSmart™ 96 (Mettler Toledo). All wells were combined into a separate FACS tube for each sample, and stained with Draq7 at 1:150 dilution (Cell Signaling). Using a SH800 (Sony), 20 nuclei per sample were sorted per well into eight 96-well plates (total of 768 wells) containing 10.5 µL EB (25 pmol primer i7, 25 pmol primer i5, 200 ng BSA (Sigma)). During the sort, nuclei with 2-8 copies of DNA (2-8n) were included since cardiomyocyte nuclei in human hearts are often polyploid (Gilsbach et al., 2018). Preparation of sort plates and all downstream pipetting steps were performed on a Biomek i7 Automated Workstation (Beckman Coulter). After addition of 1 µL 0.2% SDS, samples were incubated at 55 °C for 7 min with shaking (500 rpm). 1 µL 12.5% Triton-X was added to each well to quench the SDS. Next, 12.5 µL NEBNext High-Fidelity 2× PCR Master Mix (NEB) were added and samples were PCR-amplified (72 °C 5 min, 98 °C 30 s, (98 °C 10 s, 63 °C 30 s, 72°C 60 s) × 12 cycles, held at 12 °C). After PCR, all wells were combined. Libraries were purified according to the MinElute PCR Purification Kit manual (Qiagen) using a vacuum manifold (QIAvac 24 plus, Qiagen) and size selection was performed with SPRISelect reagent (Beckmann Coulter, 0.55x and 1.5x). Libraries were purified one more time with

SPRISelect reagent (Beckman Coulter, 1.5x). Libraries were quantified using a Qubit fluorimeter (Life technologies) and a nucleosomal pattern of fragment size distribution was verified using a Tapestation (High Sensitivity D1000, Agilent). Libraries were sequenced on a NextSeq500 sequencer (Illumina) using custom sequencing primers with following read lengths: 50 + 10 + 12 + 50 (Read1 + Index1 + Index2 + Read2). Primer and index sequences are listed in Table S20.

2.5.4 Single nucleus RNA-seq

Nuclei were isolated from heart tissue using a gentleMACS (Miltenyi) dissociator. ~40 mg of frozen heart tissue was suspended in 2 ml of MACS dissociation buffer (5 mM CaCl2 (G-Biosciences, R040), 2 mM EDTA (Invitrogen, 15575-038), 1X protease inhibitor (Roche, 05-892-970-001), 3 mM MgAc (Grow Cells, MRGF-B40), 10 mM Tris-HCl pH 8 (Invitrogen, 15568-075), 0.6 mM DTT (Sigma-Aldrich, D9779), and 0.2 U/µL of RNase inhibitor (Promega, N251B) in water (Corning, 46-000-CV)) and placed on wet ice. Next, samples were homogenized using gentleMACS dissociator (Miltenyi) with gentleMACS M tubes (Miltenyi, 130-096-335)) and the "Protein_01_01" protocol. Suspension was filtered through a 30 µM CellTrics filter (Sysmex, 04-0042-2316). M tube and filter were washed with 3 mL of MACS dissociation buffer and combined with the suspension. Suspension was centrifuged in a swinging bucket centrifuge (Eppendorf, 5920R) at 500 g for 5 minutes (4°C, ramp speed 3/3). Supernatant was carefully removed and pellet was resuspended in 500 µL of nuclei permeabilization buffer (0.1% Triton X-100 (Sigma-Aldrich, T8787), 1X protease inhibitor (Roche, 05-892-970-001), 1 mM DTT (Sigma-Aldrich, D9779), 0.2 U/µL RNase inhibitor (Promega, N251B), and 2% BSA (Sigma-Aldrich, SRE0036) in PBS). Sample was incubated on a rotator for 5 minutes at 4°C and then centrifuged at 500 g for 5 minutes (Eppendorf, 5920R; 4°C, ramp speed 3/3). Supernatant was removed and pellet was

resuspended in 600-1000 µl of sort buffer (1 mM EDTA and 0.2 U/µL RNase inhibitor in 2% BSA (Sigma-Aldrich, SRE0036) in PBS) and stained with DRAQ7 (1:100, Cell Signaling, 7406). 75,000 nuclei were sorted using a SH800 sorter (Sony) into 50 µL of collection buffer (1 U/ µL RNase inhibitor, 5% BSA (Sigma-Aldrich, SRE0036) in PBS); Sorted nuclei were then centrifuged at 1000 g for 15 minutes (Eppendorf, 5920R; 4°C, ramp speed 3/3) and supernatant was removed. Nuclei were resuspended in 18-25 ul of reaction buffer (0.2 U/µL RNase inhibitor, 1% BSA (Sigma-Aldrich, SRE0036) in PBS) and counted using a hemocytometer. 12,000 nuclei were loaded onto a Chromium controller (10x Genomics). Libraries were generated using the Chromium Single Cell 3′ Library Construction Kit v3 (10x Genomics, 1000078) according to manufacturer specifications. cDNA was amplified for 12 PCR cycles. SPRISelect reagent (Beckman Coulter) was used for size selection and clean-up steps. Final library concentration was assessed by Qubit dsDNA HS Assay Kit (Thermo-Fischer Scientific) and fragment size was checked using Tapestation High Sensitivity D1000 (Agilent) to ensure that fragment sizes were distributed normally around 500 bp. Libraries were sequenced using a NextSeq500 or HiSeq4000 (Illumina) using these read lengths: Read 1: 28 cycles, Read 2: 91 cycles, Index 1: 8 cycles.

2.5.5 Human pluripotent stem cell culture

An engineered H9-hTnnTZ-pGZ-D2 human pluripotent stem cell transgenic reporter line was purchased from WiCell and maintained on Geltrex (Gibco) pre-coated tissue culture plates in E8 medium (Chen et al., 2011) containing DMEM/F12, L-ascorbic acid-2-phosphate magnesium (64 mg/L), sodium selenium (14 µg/L), FGF2 (100 µg/L), insulin (19.4 mg/L), NaHCO3 (543 mg/L) transferrin (10.7 mg/L), and TGFβ1(2 µg/L). Cells were passaged every 3 to 5 days upon reaching ~80% confluency. For single cell passaging experiments, cells were incubated with pre-

warmed TrypLE™ Select Enzyme, no phenol red (1 mL per well of a 6-well plate) for 2-3 minutes in a 37°C, 5% CO2 incubator. Following incubation, cells were triturated to create a single cell suspension and cultured in E8 Medium supplied with Rock inhibitor (Watanabe et al., 2007) for 18-24 hours post-split, followed by daily feeding with E8 medium.

2.5.6 *In vitro* cardiomyocyte differentiation

The H9-hTnnTZ-pGZ-D2 cell line was differentiated into beating cardiomyocytes utilizing a previously reported Wnt-based monolayer differentiation protocol (Lian et al., 2013). Briefly, the H9-hTnnTZ-pGZ-D2 cell line was cultured in E8 medium for 3-10 passages. Prior to differentiation, human pluripotent stem cells were seeded at a density of 350,000-400,000 cells per well of a 12-well plate and cultured for two days. For direct differentiation, cells were treated with 10 μM CHIR99021 (Fisher, #442350) in RPMI/B-27 without insulin. Fresh RPMI/B-27 without insulin media was replaced at post 24hr and cells were then cultured two days. At day 3, cells were treated with 5 μM IWP2 (TOCRIS, #353310) in conditional medium and RPMI/B-27 without insulin 1:1 mix medium for another two days. At day 5, cells were exposed to fresh RPMI/B-27 without insulin media again for two days. Then, fresh RPMI/B-27 with insulin media was used and replenished every two days. Contracting cardiomyocytes were usually observed at day 7-8. D25 in vitro cardiomyocytes were purified utilizing PSC-derived cardiomyocyte isolation kit, human (Miltenyi Biotec, 130-110-188) and used for Real-time quantitative PCR (RT-qPCR).

2.5.7 Luciferase reporter assay

A genomic region harboring the *KCNH2* intronic enhancer (containing the risk allele: rs7789146-G / rs7789585-G) was amplified by nested-PCR (*KCNH2*-E-cF:

CTGGCTGAAGACACCTTACTTT; *KCNH2*-E-cR: ACGGAGCAGTCAAGGAAAC and *KCNH2*-In-cF: CGGGGTACCCCTCCGTAAATGAGGTGCTATC; *KCNH2*-In-cR: CCCTCGAGACGGAGCAGTCAAGGAAAC) using genomic DNA of H9-hTnnTZ-pGZ-D2 transgenic cells as a template and cloned into pGL4.23 [luc2/minP] (Promega, Cat#E8411) luciferase reporter vector. Synthetic DNA containing the *KCNH2* intronic 5'-half enhancer (rs7789045-rs7789690) with the non-risk/non-risk allele (rs7789146-A / rs7789585-A), non-risk/risk allele (rs7789146-A / rs7789585-G) and risk/non-risk allele (rs7789146-G / rs7789585-A) were purchased from integrated DNA technologies. *KCNH2* intronic 3'-half enhancer (rs7789654-rs7790480) was amplified by PCR (*KCNH2*-R-In-cF: GCTGTGCAGTGTCAGGTTAT; *KCNH2*-In-cR: CCCTCGAGACGGAGCAGTCAAGGAAAC). Then the whole *KCNH2* intronic enhancer with the non-risk/non-risk allele (rs7789146-A / rs7789585-A), non-risk/risk allele (rs7789146-A / rs7789585-G) and risk/non-risk allele (rs7789146-G / rs7789585-A) were generated by second PCR (KCNH2-In-cF: CGGGGTACCCCTCCGTAAATGAGGTGCTATC; KCNH2-In-cR: CCCTCGAGACGGAGCAGTCAAGGAAAC). One day prior to transfection, 3×10^5 of D15 *in vitro* differentiated cardiomyocytes were plated in a Geltrex-coated 24-well plate. Cardiomyocytes were transfected with 500 ng of pGL4.23 plasmid (either empty, *KCNH2* enhancer with G/G allele, A/A allele or mix) and 10 ng TK:Renilla-luc as internal control using Lipofectamine Stem Transfection Reagent (Invitrogen, #STEM00003). Media was replaced with fresh media at 24 hrs post-transfection. At 72 hrs post-transfection, media was removed and the cells were washed with PBS. Luminescence was measured using a Dual-Luciferase Reporter Assay System (Promega, #E2920) according to the manufacturer's protocol.

2.5.8 CRISPR mediated genome editing experiments

To interrogate the functional significance of the atrial fibrillation-associated risk variant-containing cCRE at the *KCNH2* locus, the cCRE sequence was genetically deleted in H9-hTnnTZ-pGZ-D2 transgenic hPSCs using an efficient CRISPR/Cas9-mediated knockout system (Teumer et al., 2016; Zhang et al., 2019c). Two adjacent gRNAs (KCNH2-enh gRNA-1, CTCATTTACGGAGGAGCGCA; KCNH2-enh gRNA-2, TACAGTGGCCTTCTAGACGA) targeting the cCRE were designed using a web-based software tool CRISPOR (Haeussler et al., 2016), based on targeting region of interest and minimizing potential off-target effects. The identified gRNAs were then synthesized in vitro using the GeneArt Precision gRNA Synthesis kit (Invitrogen) according to the manufacturer's protocol. One day prior to transfection, 1.5×10^5 H9-hTnnTZ-pGZ-D2 hPSCs were seeded in 12-well plates. A pair of RNP complexes containing 1.2 μg of Cas9 protein (NEB) and 400 ng of in vitro transcribed gRNA were then transfected (Kim et al., 2014; Zuris et al., 2015) using Lipofectamine stem transfection reagent (Invitrogen). 72 hours after the transfection, cells were diluted and clonally expanded another 7 days. Colonies were picked and lysates were prepared after the first passage for genotyping (Santos et al., 2016) (KCNH2-enh extended forward primer, ACACCTTACTTTGGGTGAGAAG; KCNH2-enh extended reverse primer, AGACAGAGCACAGACCTAGAA; KCNH2-enh internal forward primer, GCTGTGCAGTGTCAGGTTAT; KCNH2-enh internal reverse primer, TCTCCCTCCTTCTCTCTCATTC). After confirmation of genome-edited clones by Sanger sequencing, two transfected WT clones, two heterozygote clones, and two homozygote clones were selected for further functional analysis.

2.5.9 RT-qPCR

Total RNA was isolated from the cells using TRIzol reagent (Invitrogen). 1 μg of total RNA was reverse transcribed using the iScript Reverse Transcription Supermix kit (Bio-Rad) for RT-qPCR. RT-qPCR was performed using PowerUP™ SYBR™ Green Master Mix (Applied Biosystems) in the CFX Connect Real-Time System (Bio-Rad). The results were normalized to the *TBP* gene. The primers used for RT-qPCR are listed in Table S21.

2.5.10 Electrophysiology of cardiomyocytes

Both WT and *KCNH2* enhancer knockout D15 *in vitro* cardiomyocytes were purified using the PSC-derived cardiomyocyte isolation kit, human (Miltenyi Biotec, 130-110-188) and cultured for another 10-20 days in a low density prior to electrophysiological measurements. The single-pipette, whole-cell patch current-clamp technique was used for recordings. Action potentials were recorded with a patch clamp amplifier (Axopatch 200B, Axon) and experiments were performed at a temperature of 35 ± 0.5 °C. Current-clamp command pulses were generated by a digital-to-analog converter (DigiData 1440, Axon) which was controlled by the pCLAMP software (10.3, Axon). Pipettes (resistance 3-5 MΩ) were pulled using a micropipette puller (Model P-87, Sutter Instrument Co.). Several minutes after seal formation, the membrane was ruptured by gentle suction to establish the whole-cell configuration for voltage clamping. Subsequently, the amplifier was switched to the current-clamp mode. Cells were paced with 1 Hz, injected current stimuli from 3 to15 nA for 5 ms duration. Cells were superfused with extracellular solution containing (in mM): 140 NaCl, 5.4 KCl, 1.8 $CaCl_2$, 1.0 $MgCl_2$, 5.5 glucose and 5.0 HEPES (pH 7.4 adjusted with NaOH). Pipette solution contained (in mM): 120 K-gluconate, 10 KCl, 5 NaCl, 10 HEPES, 5 Phosphocreatine, 5 ATP-Mg_2 and Amphotericin 0.44 μM (pH 7.2 adjusted with KOH).

2.5.11 Demultiplexing of snATAC-seq reads

For each sequenced snATAC-Seq library, we obtained four FASTQ files, two for paired-end DNA reads as well as the combinatorial indexes for i5 (768 different PCR indices) and T7 (96 different tagmentation indices; Table S20). We selected all reads with <= 2 mistakes per individual index (Hamming distance between each pair of indices is 4) and subsequently integrated the full barcode at the beginning of the read name in the demultiplexed FASTQ files (https://gitlab.com/Grouumf/ATACdemultiplex/).

2.5.12 Filtering of snATAC-seq profiles by TSS enrichment and unique fragments

TSS (transcriptional start site) positions were obtained from the GENCODE database v31 (Harrow et al., 2006). Tn5-corrected insertions were aggregated ± 2000 bp around each TSS genome wide. Then, this profile was normalized to the mean accessibility ± (1900 to 2000) bp from the TSS and smoothed every 11 bp. The maximum value of the smoothed profile was taken as the TSS enrichment. We selected all nuclei that had at least 1,000 unique fragments and a TSS enrichment of at least 7 for all data sets.

2.5.13 Clustering strategy for snATAC-seq datasets

We utilized two rounds of clustering analysis to identify clusters. The first round of clustering analysis was performed on individual samples. We divided the genome into 5 kbp consecutive bins and then scored each nucleus for any insertions in these bins, generating a bin-by-cell binary matrix for each sample. We filtered out those bins that were generally accessible in all nuclei for each sample using a z-score threshold of 1.65 (equivalent to a one tailed p-value < 0.05). Based on the filtered matrix, we then carried out dimensionality reduction followed by

graph-based clustering to identify cell clusters. We called peaks using MACS2 (Zhang et al., 2008) for each cluster using the aggregated profile of accessibility and then merged the peaks from all clusters to generate a union peak list. Based on the peak list, we generated a cell-by-peak count matrix and used Scrublet (Wolock et al., 2019) to remove potential doublets with default parameters. Doublet scores returned by Scrublet (Wolock et al., 2019) were then used to fit a two-component Gaussian mixture model using the *BayesianGaussianMixture* function from the python package *scikit-learn* (Fabian Pedregosa, 2011). Nuclei in the component with the larger mean doublet score were removed from downstream analysis since they likely reflected doublets.

Next, to carry out the second round of clustering analysis, we merged peaks called from all samples to form a reference peak list. We generated a binary cell-by-peak matrix using nuclei from all samples and again performed the dimensionality reduction followed by graph-based clustering to obtain the final cell clusters across the entire dataset.

2.5.14 Dimensionality reduction and batch correction of snATAC-seq data

For processing of snATAC-seq data we adapted our previously published method, SnapATAC (Fang et al., 2021). To reduce the dimensionality of the peak by cell count matrix, SnapATAC utilizes spectral embedding for dimensionality reduction. To further increase the performance and scalability of spectral embedding, we applied the Nyström method (Bouneffouf, 2016) to enable handling of large datasets. Specifically, we first randomly sampled 35,000 nuclei as training data. . We computed the matrix $P = D^{-1}S$ S, where D is the diagonal matrix such that $D_{ii} = \sum_j S_{ij}$. The eigendecomposition was performed on P and the eigenvector with eigenvalue 1 was discarded. From the rest of the eigenvectors, we took k of them corresponding to the largest eigenvalues as the spectral embedding of the training data. We utilized the Nyström method

(Bouneffouf, 2016) to extend the embedding to the data outside the training set. Given a set of unseen samples, we computed the similarity matrix S' between the new samples and the training set. The embedding of the new samples is given by $' = S'U\Lambda^{-1}$, where U and Λ are the eigenvectors and eigenvalues of P obtained in the previous step. To correct for donor/batch specific effects, after dimensionality reduction we performed cell grouping on individual samples using k-mean clustering with k equal to 20. We then constructed k-NN graphs for each sample and used the MNN correction method to identify mutual nearest neighbors (Haghverdi et al., 2018).These mutual nearest neighbors were used as the anchors to match the cells between different samples and correct for donor/batch effects as described previously (Haghverdi et al., 2018).

2.5.15 Clustering of snATAC-seq data

We constructed the k-nearest neighbor graph (k-NNG) using low-dimensional embedding of the nuclei with k equal to 50. We then applied the Leiden algorithm (Traag et al., 2019) with constant Potts model (CPM) to find communities in the k-NNG corresponding to the cell clusters. The Leiden algorithm can be configured to use different quality functions. The modularity model is a popular choice but it is hampered by the resolution-limit, particularly when the network is large (Traag et al., 2011). Therefore, we used the modularity model only in the first round of clustering analysis to identify initial clusters. In the final round of clustering, we chose the constant Potts model as the quality function since it is resolution-limit-free and is better suited for identifying rare populations in a large dataset (Traag et al., 2011). Nuclei from two small clusters (280 and 254 nuclei) with low reproducibility and stability were discarded from downstream analysis. 34 nuclei that formed clusters of 1 and 2 nuclei were discarded as well.

2.5.16 Processing and clustering analysis of snRNA-seq datasets

Raw sequencing data was demultiplexed and preprocessed using the Cell Ranger software package v3.0.2 (10x Genomics). Raw sequencing files were first converted from Illumina BCL files to FASTQ files using *cellranger mkfastq*. Demultiplexed FASTQs were aligned to the GRCh38 reference genome (10x Genomics), and reads for exonic and intronic reads mapping to protein coding genes, long non-coding RNA, antisense RNA, and pseudogenes were used to generate a counts matrix using *cellranger count*; expect-cells parameter was set to 5,000. A separate counts matrix for each sample was also generated using only reads mapped to intronic regions.

Next, exon + intron count matrices for individual datasets were processed using the Seurat v3.1.4 R package (Stuart et al., 2019) (https://satijalab.org/seurat/) to assess dataset quality. Features represented in at least 3 cells and barcodes with between 500 and 4,000 genes were used for downstream processing; additionally, barcodes with mitochondrial read percentages greater than 5% were removed. Counts were log-normalized and scaled by a factor of 10,000 using *NormalizeData*. To identify variable genes, *FindVariableFeatures* was run with default parameters except for nfeatures = 3000 to return the top 3,000 variable genes. All genes were then scaled using *ScaleData*, which transforms the expression values for downstream analysis. Next, principal component analysis was performed using *RunPCA* with default parameters and the top 3,000 variable features as input. The first 20 principal components were used to run clustering using *FindNeighbors* and *FindClusters* (parameter res = 0.4). To generate UMAP coordinates *RunUMAP* was run using the first 20 principal components and with parameters umap.method = “umap-learn”, and metric = “correlation”. Doublet scores (pANN) were generated for cell

barcodes using DoubletFinder (McGinnis et al., 2019) (https://github.com/chris-mcginnis-ucsf/DoubletFinder) using the parameters pN =0.15 and pK = 0.005; the anticipated collision rate was set by specifying 2% collisions per thousand nuclei for individual datasets.

Individual datasets were merged together using the *merge* function in Seurat to combine the count matrices and designate unique barcodes. Cell barcodes with pANN scores greater than 0 were removed from downstream analysis. Metadata was also encoded for each barcode, and the merged dataset was processed in a similar manner as described above; clusters were identified using *FindNeighbors* and *FindClusters* (res = 0.8). To generate the UMAP coordinates, the first 14 principal components were used in *RunUMAP*; the UMAP algorithm for Seurat v3.1.4 uses the uwot R-package, and that setting was used to generate the coordinates here. To regress out donor specific effects, the Harmony R package (https://github.com/immunogenomics/harmony) (Korsunsky et al., 2019) was used, and the recomputed principal components were used to re-cluster the cells and rerun UMAP using the above parameters. For downstream analysis and comparison to snATAC-seq data we combined ventricular cardiomyocyte clusters, atrial cardiomyocyte clusters, fibroblast clusters, and endothelial cell clusters manually based on shared gene expression patterns (Fig S2G, H). Cluster-specific genes in the all-transcripts dataset were identified in a global differential gene expression test using *FindAllMarkers* with parameters logFC = 0.25, min.pct = 0.25, and only.pos = FALSE.

2.5.17 Integration of snRNA-seq and snATAC-seq data

The snRNA-seq and snATAC-seq datasets were used to perform label transfer from the RNA cells onto the snATAC-seq dataset using the Seurat v3.1.4 R package (https://satijalab.org/seurat/) (Stuart et al., 2019). Gene activity scores were calculated using

chromatin accessibility in regions from the promoter up to 2kb upstream for each ATAC nucleus. Activity scores were log-normalized and scaled using *NormalizeData* and *ScaleData*. To compare the snRNA and snATAC datasets and identify anchors, *FindTransferAnchors* was run considering the top 3,000 variable features from the snRNA-seq dataset. Anchor pairs were used to assign RNA-seq labels to the snATAC-seq cells using *TransferData*, with the weight.reduction parameter set to the principal components used in snATAC-seq clustering. The efficacy of integration was assessed by examining the distribution of the maximum prediction scores output by *TransferData* and the distribution of annotated snATAC-seq identities to the corresponding predicted label.

2.5.18 Creation of a consensus list of heart candidate *cis* regulatory elements

MACS2 (v2.1.2) (Zhang et al., 2008) was used to identify accessible chromatin sites for each cluster with the following parameters: *-q 0.01 --nomodel --shift -100 --extsize 200 -g 2789775646 --call-summits --keepdup-all.* Estimated genome size was determined to be 2789775646 bp and was indicated by the *-g* parameter. We next filtered out peaks overlapping with the ENCODE blacklist (Amemiya et al., 2019) (hg38, https://github.com/Boyle-Lab/Blacklist/).

To generate the union of heart cCREs, we merged the blacklist-filtered peaks obtained for each cluster using the BEDtools merge command with default settings (v2.25.0) (Quinlan and Hall, 2010).

2.5.19 Computing relative accessibility scores for candidate *cis* regulatory elements

To correct biases arising from differential read depth among cells and cell types, we derived a procedure that normalizes chromatin accessibility at cCREs identified by MACS2 peak calling

(v2.1.2) (Zhang et al., 2008). We define the set of accessible loci by *L* and we define a peak *p* as a subset of related loci *l* from *L.* Let a_l be the accessibility of accessible locus *l* and *P* the set of non-overlapping peaks used to define the loci. For a given cell type $S_i \in S$, we computed the median med_j number of reads sequenced per cells. For each feature $p_j \in P$, we computed m_{ij} the average number of reads sequenced from S_i and overlapping p_j. We then defined the activity $a_{ij}$of loci p_j in S_i as $a_{ij} = 10^6 . \frac{\sqrt[1/med_j]{(1-m_{ij})}}{\sum_{j \in P} \sqrt[1/med_j]{(1-m_{ij})}}$. We then define the relative accessibility score (RAS) $A_{ij} = \frac{a_{ij}}{\sum_{i \in S} a_{ij}}$.

2.5.20 K-means clustering of candidate *cis* regulatory elements

We clustered the union of 287,415 candidate *cis* regulatory elements (cCREs) using a K-means clustering procedure. We first created a sparse cell x peak matrix that was transformed into a RAS-normalized cell type x peak matrix. We then performed K-means on the normalized matrix with *K* from 2 to 12 and computed the Davies-Bouldin (DB) index for each K (Davies and Bouldin, 1979). Let $R_{xy} = \frac{(s_x + s_y)}{d_{xy}}$ with s_x the average distance of each cell of cluster *x* and d_{xy} the distance between the centroids of clusters *x* and *y*. The Davies-Bouldin index is defined as $DB = \frac{1}{K} \sum_{x,y \in K} \max_{x \neq y}(R_{xy})$. We selected K = 9 since it resulted in the lowest DB index which indicates the best partition. We used the python library scikit-learn (Fabian Pedregosa, 2011) to compute the K-means algorithm and the DB index (Davies and Bouldin, 1979).

2.5.21 Cell type annotation

We annotated snATAC-seq and snRNA-seq clusters based on chromatin accessibility at promoter regions or expression of known lineage marker genes, respectively. We annotated atrial and ventricular cardiomyocytes based on differential chromatin accessibility and gene expression at *NPPA, MYH6*, *KCNJ3*, *MYL7*, *MYH7, HEY2, MYL2* and other reported markers of atrial and ventricular cardiomyocytes (Kubalak et al., 1994; Ng et al., 2010; Sheikh et al., 2015). We used, for example, the gene *DCN* to annotate cardiac fibroblasts (Furtado et al., 2014); *VWF* and *EGFL7* for endothelial cells (Kalucka et al., 2020; Nichol and Stuhlmann, 2012); *GJA4* and *TAGLN* for smooth muscle cells (Schaum et al., 2018; Shanahan et al., 1993); *CD163* and *MS4A6A* for macrophages (Fabriek et al., 2005; Martinez et al., 2006); *IL7R* and *THEMIS* for lymphocytes (Lesourne et al., 2009; Peschon et al., 1994); *ADIPOQ* and *CIDEA* for adipocytes (Hu et al., 1996; Puri et al., 2008); *NRXN3* and *GPM6B* for a cluster of nervous cells with neuronal and Schwann-like gene expression and chromatin accessibility signatures (Litvinukova et al., 2020; Skelly et al., 2018; Tucker et al., 2020). From snRNA-seq, we identified a population of endothelial-like cells with specific expression of endocardial cell markers *NRG3* and *NPR3* (Tang et al., 2018; Zhao et al., 1998). We also identified subtypes of mesenchymal cells that included myofibroblasts with characteristic expression of embryonic smooth muscle actin *MYH10* (Baum and Duffy, 2011; Southern et al., 2016) as well as arterial smooth muscle cells with preferential expression of *ACTA2* and *TAGLN* relative to a larger cluster of pericytes (Vanlandewijck et al., 2018) (Table S4). snRNA-seq annotations were consistent with recent single cell transcriptomic analyses of adult human heart tissue (Litvinukova et al., 2020; Tucker et al., 2020).

2.5.22 Identification of cell type-specific candidate *cis*-regulatory elements

We used *edgeR* (version 3.24) in R (Robinson et al., 2010) to identify cell type-specific cCREs. For each cCRE, accessibility within a cell type was compared to average accessibility in all other clusters. For each cell type, we created a count table for each cCRE using the following strategy: each sample was described with a donor and a chamber ID. For each sample ID we reported read count within 1) the cell type and 2) the rest of the cell types in aggregate. We used this count matrix as input for edgeR analysis (Robinson et al., 2010) and performed a likelihood ratio test. We considered cCREs with fold change > 1.2 and FDR < 0.01 after Benjamini-Hochberg correction as cell type-specific.

2.5.23 Co-accessibility analysis using Cicero

We used the R package Cicero (Pliner et al., 2018) to infer co-accessible chromatin loci. For each chromosome, we used as input the corresponding peaks from our 287,415 cCRE union set and the coordinates of the snATAC-seq UMAP (Leland McInnes, 2018). We randomly subsampled 15,000 cells from our aggregate snATAC-seq dataset to construct input matrices for Cicero analysis. We used +/-250 kbp as cutoff for co-accessibility interactions. All other settings were default.

2.5.24 Correlation of gene expression and promoter accessibility

We defined promoter regions as transcriptional start sites (TSS) +/-2 kbp. Transcriptional start sites were extracted from annotation files from GENCODE release 33 (Harrow et al., 2006). We identified promoter-overlapping peaks using BEDtools (Quinlan and Hall, 2010) and a custom script (see Code availability). For each overlapping pair (peak, promoter) identified, we kept only the open chromatin site closest to the TSS in order to obtain a 1:1 correspondence between genes

and open chromatin peaks. We then used the relative accessibility score (RAS) and the cluster-scaled FPKM gene expression score to create feature x cell type matrices for RNA-seq and ATAC-seq datasets. We then used these matrices to create heatmaps and to perform ATAC-seq/RNA-seq cluster correlation analysis using the Pearson similarity metric. For each cell type, we computed the Pearson correlation score between the RAS vector of the 7,081 promoters and the scaled FPKM vector of the corresponding 7,081 genes identified via the 1:1 correspondence method described above.

2.5.25 Differentially accessible cCREs between heart chambers

Between-heart chamber differential accessibility analysis was performed for five cell types from our aggregated single nucleus ATAC-seq dataset. We considered only cell types which had a representation of at least 50 nuclei per dataset and at least 300 nuclei across each tested condition. The cell types that met these inclusion criteria included cardiomyocytes, fibroblasts, endothelial cells, smooth muscle cells, and macrophages. Within each cell type, a generalized linear model framework was employed using the R package edgeR (Robinson et al., 2010). All fragments for a given cell type were aggregated in the .bed format. MACS2(Zhang et al., 2008) was used to call peaks on the aggregate .bed file for each cell type with the parameters specified above. *NarrowPeak* output bed files were used for differential accessibility testing. The aggregate .bed file for each cell type was then partitioned based on dataset of origin using nuclear barcodes. The '*coverage*' option of the BEDtools package (Quinlan and Hall, 2010) was applied with default settings to count the total number of chromatin fragments from each dataset overlapping *narrowPeaks* called on the aggregate .bed file for the corresponding cell type. This yielded a raw count matrix in the format of single nucleus ATAC-seq datasets (columns) by *narrowPeaks* (rows)

for each cell type. The raw count matrix was used as input for edgeR analysis. To filter low-coverage peaks from our analysis, we used the '*filterByExpr*' command within edgeR with default settings. We applied an average prior count of one during fitting of the generalized linear model in order to avoid inflated fold changes in instances for which peaks lacked coverage for one but not both tested conditions. We modelled chromatin accessibility at each peak as a function of heart chamber (group) with sex as a covariate. The generalized linear model was expressed as follows in edgeR notation:

--

design <- model.matrix(~sex+group)

y <- estimateDisp(y, design, prior.count = 1)

glmFit(y, design)

--

Significance was tested using a likelihood ratio test. To account for testing multiple hypotheses, a Benjamini-Hochberg significance correction was applied for all cCREs tested within each considered cell type. Any cCRE with an absolute $\log_2$(fold change) > 1 and an FDR-corrected p value < 0.05 was considered significant.

2.5.26 Gene expression analysis of genes co-accessible with DA candidate *cis*-regulatory elements

To compare the expression of genes co-accessible with heart chamber-dependent distal DA cCREs (outside +/- 2 kb of TSS) in cardiomyocytes and fibroblasts, we performed differential expression testing for all genes between indicated heart chambers using Wilcoxon rank sum test in Seurat (Stuart et al., 2019). Genes with an absolute Fold Change > 1.5 and an FDR-adjusted P

value < 0.05 were considered differentially expressed. We then tested resulting genes for co-accessibility (Pliner et al., 2018) with distal DA cCREs at a co-accessibility score threshold of 0.1, and displayed scaled gene expression values from Seurat for the indicated differentially expressed genes linked to chamber-dependent distal DA cCREs.

2.5.27 GREAT ontology analysis

The Genomic Regions Enrichment of Annotations Tool (GREAT, http://great.stanford.edu/public/html/index.php) (McLean et al., 2010) was used with default settings for indicated cCREs or candidate enhancers in the .bed format. Biological process enrichments are reported. P-values shown for enrichment are Bonferroni-corrected binomial p-values.

2.5.28 Motif enrichment analysis

For *de novo* and known motif enrichment analysis of cluster-specific cCREs, the *findMotifsGenome.pl* utility of the HOMER package was used with default settings (Heinz et al., 2010). For display of enrichment patterns for motifs from the JASPAR (Mathelier et al., 2016) database with evidence of enrichment in at least one set of cell type-specific cCREs, motifs with an enrichment p-value < 10^{-5} in at least one set of cluster-specific cCREs were selected. For motif enrichment within differentially accessible cCREs, *narrowPeak* calls from MACS2 were used as input, with peaks called on the corresponding cell type (as described above) used as background. For enrichment of motifs within cell type-attributed bulk enhancers, snATAC-seq peaks from the union of snATAC-seq peaks were utilized. Summits were extracted from peaks that overlapped bulk enhancer annotations and extended by 250bp on either side to obtain fixed-width peaks. We

also computed motif enrichment scores at single-cell resolution using chromVAR (Schep et al., 2017).

For input to chromVAR, we used the summits of the 287,415 peaks in our consensus list extended by 250 base pairs in either direction, and a set of 870 non-redundant motifs as input (https://github.com/GreenleafLab/chromVARmotifs). To identify differentially enriched motifs in each cell type, we used the following strategy: for each cell type and each motif, we computed a Rank Sum test between the chromVAR Z-score distributions from cells within the cell type and outside of the cell type. Tests were run using a random sampling of 40,000 cells. Then, for each cell type we used 1e-8 as p-value cutoff. In addition, we applied a Bonferroni correction to account for multiple testing which resulted in selection of significant motifs with p-value < 1e-11.

2.5.29 Measuring single cell chromatin accessibility signal within bulk candidate heart enhancers

We obtained published candidate heart enhancers annotated by H3K27ac ChIP-seq from a recently reported bulk survey of healthy left ventricular tissue from 18 human donors (Spurrell et al., 2019). Candidate enhancers were defined per the study as H3K27ac ChIP-seq peaks that were at least 1kb away from a transcription start site and present in two or more donors. Because these reference annotations were derived from bulk profiling of healthy left ventricles, we selected only left ventricular nuclei from our aggregate dataset for comparison. We limited our analysis to cell types that comprised at least 5% of nuclei by proportion in our aggregate dataset. These included cardiomyocytes, fibroblasts, endothelial cells, smooth muscle cells, and macrophages. We first combined all fragments for each cell type from left ventricular datasets. The ‘coverage’ option of BEDtools (Quinlan and Hall, 2010) was applied with default settings to count the total number of

chromatin fragments from each ventricular cell type overlapping the candidate enhancer annotations. This yielded a raw count matrix in the format of snATAC-seq cell types (columns) by candidate enhancers (rows). The raw count matrix was normalized to RPKM (reads per kilobase per million mapped reads) for each candidate enhancer. We next used Cluster3.0 (de Hoon et al., 2004) to k-means cluster the 31,033 healthy heart candidate enhancers into K groups between 2 and 12 with the following settings (Method = *k-Means*, Similarity Metric = *Euclidian distance*, number of runs = *100*). We calculated the Davies-Bouldin (DB) index (Davies and Bouldin, 1979) as described above for each clustering using the *index.DB* function of the R package clusterSim (http://keii.ue.wroc.pl/clusterSim/). We selected a k-means of 8, which yielded the lowest DB index, indicating the best partitioning.

We repeated the above analysis for 4,406 candidate enhancers reported have increased bulk H3K27ac ChIP signal and 3,101 candidate enhancers reported to have decreased signal in 18 late stage idiopathic dilated cardiomyopathy (heart failure) left ventricles versus 18 healthy control left ventricles reported in the same study. We again clustered the candidate enhancers for both groups into k groups between 2 and 12 as above and selected the clustering that yielded the lowest DB index (Davies and Bouldin, 1979).

2.5.30 Genome-wide association study (GWAS) variant enrichment analysis

We used LD (linkage disequilibrium) score regression (Bulik-Sullivan et al., 2015; Finucane et al., 2015) to estimate genome-wide enrichment for GWAS traits using annotation sets from single cell chromatin accessibility from the heart or lung (Wang et al., 2020), or bulk DNase hypersensitivity sites for tissues from ENCODE (Consortium, 2012a; Consortium et al., 2020; Vierstra et al., 2020). For bulk DNase-seq datasets, peak annotations were merged across

biological replicates from the same tissue type. We compiled published GWAS summary statistics for cardiovascular diseases (Arvanitis et al., 2020; Malik et al., 2018; Nelson et al., 2017; Nielsen et al., 2018; Shadrina et al., 2019), other diseases (Aylward et al., 2018; Bentham et al., 2015; de Lange et al., 2017; Grove et al., 2019; Lambert et al., 2013; Mahajan et al., 2018; Okada et al., 2014; Ripke et al., 2014; Tachmazidou et al., 2019; Wiberg et al., 2019; Wray et al., 2018), and non-disease traits (Day et al., 2015; Day et al., 2017; den Hoed et al., 2013; Horikoshi et al., 2016; Jansen et al., 2019; Jiang et al., 2018; Locke et al., 2015; Okbay et al., 2016; Wittemans et al., 2019; Wood et al., 2014) using the European subset from transethnic studies where applicable. We created custom LD score files by using peaks from each cell type or tissue as a binary annotation. As background, we used baseline annotations included in the baseline-LD model v2.2. For each trait, we used LD score regression to estimate enrichment coefficient z-scores for each annotation relative to the background. Using these z-scores, we computed two-sided p-values for enrichment and used the Benjamini-Hochberg procedure to correct for multiple tests within each set of annotations.

2.5.31 Fine mapping for atrial fibrillation

We obtained published atrial fibrillation GWAS summary statistics and index variants for 111 disease-associated loci (Nielsen et al., 2018). To construct credible sets of variants for each locus, we first extracted all variants in linkage disequilibrium ($r^2 > 0.1$ using the EUR subset of 1000 Genomes Phase 3) (Auton et al., 2015a) in a large window (±2.5 Mb) around each index variant. We next calculated approximate Bayes factors (Wakefield, 2009) (ABF) for each variant using effect size and standard error estimates. We then calculated posterior probabilities of association (PPA) for each variant by dividing its ABF by the sum of ABF for all variants within

the locus. For each locus, we then defined 99% credible sets by sorting variants by descending PPA and retaining variants that added up to a cumulative PPA of > 0.99. This resulted in an output of 6,014 candidate causal variants.

2.5.32 Variant prioritization for functional validation

To prioritize variants for functional validation, we refined our list of candidate causal variants from fine mapping analysis to only those with a posterior probability of association (PPA) > 0.1 (216 remaining out of 6,014). We used BEDtools (Quinlan and Hall, 2010) to intersect these variants with ATAC-seq peaks called on an aggregate .bed file for atrial and ventricular cardiomyocyte snATAC-seq clusters (cardiomyocyte cCREs). This resulted in 40 fine-mapped variants that resided within 38 candidate cardiomyocyte cCREs (38 cCRE-variant pairs).

We assessed each remaining cCRE-variant pair via the following criteria:

- cCREs primarily accessible in cardiomyocytes
- presence of a corresponding ATAC-seq peak at a testable time point in the *in vitro* hPSC-cardiomyocyte differentiation model system
- sequence conservation in 100 vertebrates (genome browser track generated using phyloP of the PHAST5 package downloaded from UCSC genome browser (Siepel et al., 2005), http://hgdownload.soe.ucsc.edu/goldenPath/hg38/phyloP100way/)
- predicted co-accessibility of candidate enhancer with a gene promoter
- expression of putative target gene associated with cCRE appearance (chromatin accessibility and H3K27ac) during hPSC-cardiomyocyte differentiation (Zhang et al., 2019c)

A candidate cCRE-variant pair at the *KCNH2* locus was prioritized for functional experimentation.

2.5.33 ChIP-seq data processing

Reads were mapped to the human genome reference GRCh38 using Bowtie2 (version 2.2.6) (Langmead and Salzberg, 2012) and reads with MAPQ > 30 selected using SAMtools (version 1.3.1) (Li et al., 2009). PCR duplicates were removed using MarkDuplicates function of Picard tools (version 1.119) (Institute, 2019). RPKM normalized signal tracks were generated using BamCoverage function in deepTools (version 2.4.1) (Ramirez et al., 2014).

2.5.34 RNA-seq data processing

Reads were mapped to the human genome reference GRCh38 using STAR (version 020201) (Dobin et al., 2013) and reads with MAPQ > 30 selected using SAMtools (version 1.3.1) (Li et al., 2009). PCR duplicates were removed using MarkDuplicates function of Picard tools (version 1.1.19) (Institute, 2019). RPKM normalized signal tracks were generated using BamCoverage function in deepTools (version 2.4.1) (Ramirez et al., 2014).

2.5.35 ATAC-seq data processing

Reads were mapped to the human genome reference GRCh38 using Bowtie2 (version 2.2.6) (Langmead and Salzberg, 2012) and reads with MAPQ > 30 selected using SAMtools (version 1.3.1) (Li et al., 2009). PCR duplicates were removed using SAMtools (version 1.3.1) (Li et al., 2009). RPKM normalized signal tracks were generated using BamCoverage function in deepTools (version 2.4.1) (Ramirez et al., 2014).

2.5.36 Statistical analysis

No statistical methods were used to predetermine sample sizes. There was no randomization of the samples, and investigators were not blinded to the specimens being investigated. However, clustering of single nuclei based on chromatin accessibility was performed in an unbiased manner, and cell types were assigned after clustering. Low-quality nuclei and potential barcode collisions were excluded from downstream analysis as outlined above. Cluster-specificity at each cCRE was tested using edgeR (Robinson et al., 2010) as described above, with p-values corrected via the Benjamini Hochberg method. To identify differentially accessible sites between heart chambers and for each cell type, a likelihood ratio test was used and the resulting p-value was corrected using the Benjamini Hochberg method. For significance of ontology enrichments using GREAT, Bonferroni-corrected binomial p values were used (McLean et al., 2010). For significance testing of enrichment of *de novo* and known motifs, a hypergeometric test was used without correction for multiple testing (Heinz et al., 2010). For luciferase and qPCR data, we performed one-way ANOVA (ANalysis Of VAriance) analysis with post-hoc Tukey HSD (Honestly Significant Difference) using GraphPad Prism version 8.0.0 for Windows, GraphPad Software, San Diego, California USA, www.graphpad.com.

2.5.37 External datasets

Cardiomyocyte differentiation: RNA-Seq, H3K27ac day 0 (hPSC); day 5 (cardiac mesoderm); and day 15 (primitive cardiomyocytes) were downloaded from GSE116862 (Zhang et al., 2019c). Signal tracks for heart H3K27ac ChIP-seq data were downloaded from https://portal.nersc.gov/dna/RD/heart/. List of candidate enhancers was downloaded from

supplementary tables (Spurrell et al., 2019). H3K27ac ChIP-seq data for cardiomyocyte nuclei from non-failing donors (NF1) were downloaded from NCBI SRA BioProject ID PRJNA353755 (Deutsch et al., 2018). We acquired snATAC-seq data for human lung from GSE161383 (Wang et al., 2020) and bulk DNase-seq datasets for human tissues from ENCODE (Consortium, 2012a; Consortium et al., 2020; Vierstra et al., 2020) with the following identifiers: ENCSR053ZKP, ENCSR259GYP, ENCSR277KRY, ENCSR458AOS, ENCSR597NVK, ENCSR422IIZ, ENCSR968TPO, ENCSR788IZL, ENCSR859LTL, ENCSR060HPL, ENCSR783OCW, ENCSR171ADO, ENCSR171ETY, ENCSR520BAD, ENCSR686WJL, ENCSR791BHE, ENCSR856XLJ, ENCSR361DND, ENCSR579KDC, ENCSR693UHT, ENCSR365NDK, ENCSR712PYJ, ENCSR930PDT, ENCSR178JBL, ENCSR595HZQ, ENCSR261RWJ, ENCSR455GUW, ENCSR689DSM, ENCSR954AJK, ENCSR564FZH, ENCSR000ELO, ENCSR909HFI, ENCSR931UQB, ENCSR128GBN, ENCSR006IMH, ENCSR163PKT, ENCSR641ZPF, ENCSR782SSS, ENCSR101QXF, ENCSR709IYR, ENCSR866ODX, ENCSR195ONB, ENCSR450PWF, ENCSR549NRK, ENCSR749MUH, ENCSR080ISA, ENCSR090IDV, ENCSR102RSU, ENCSR401ESD, ENCSR484UAU, ENCSR508FVM, ENCSR340MRJ, ENCSR760QZM, ENCSR763AKE, ENCSR923JYH, ENCSR164WOF, ENCSR323UTX, ENCSR650FLQ, ENCSR702DPD, ENCSR129BZE, and ENCSR437AYW.

2.5.38 Code availability

The pipeline for processing snATAC-seq data is available as a part of the Taiji software: https://taiji-pipeline.github.io/

Custom code used for demultiplexing and downstream analysis for snATAC data is available here:

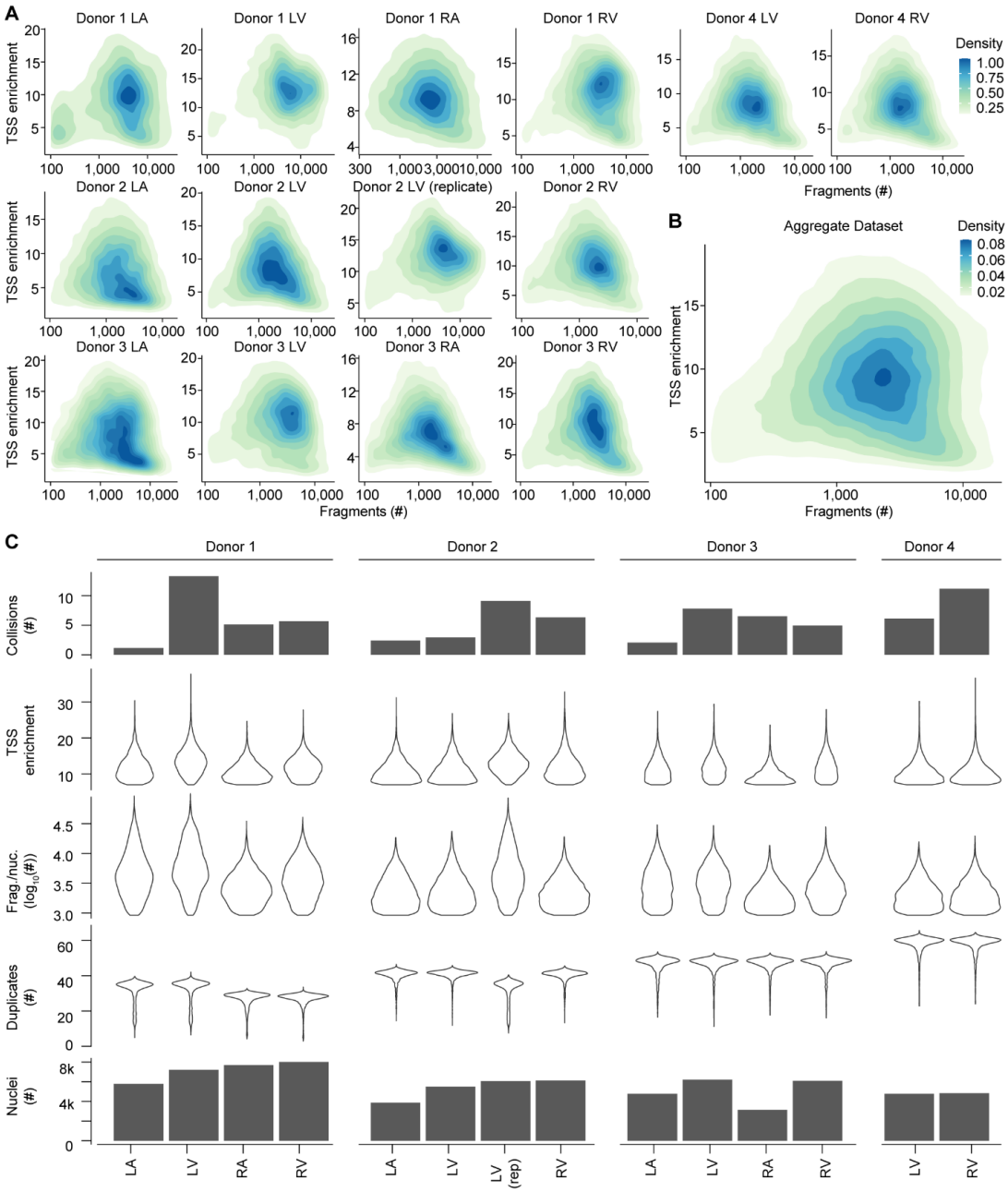
https://gitlab.com/Grouumf/ATACdemultiplex/-/tree/master/ATACdemultiplex

https://gitlab.com/Grouumf/ATACdemultiplex/-/blob/master/scripts/

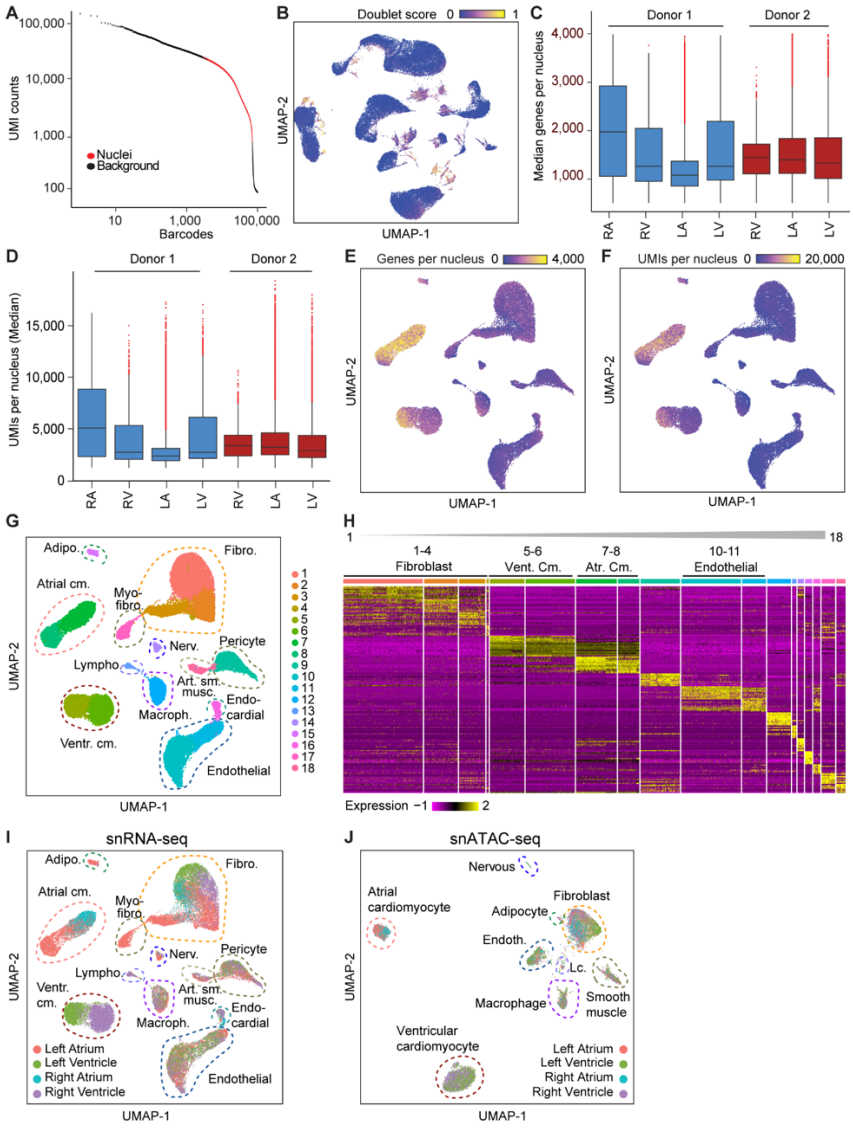
2.5.39 Data availability

Sequencing data are available from dbGaP (phs002204.v1.p1). Processed data are available from GEO (GSE165839) and can be explored using our publicly available web portal including a UCSC cell browser (https://github.com/maximilianh/cellBrowser) and genome browser track viewer (IGV.js: https://github.com/igvteam/igv.js#igvjs): http://catlas.org/humanheart. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

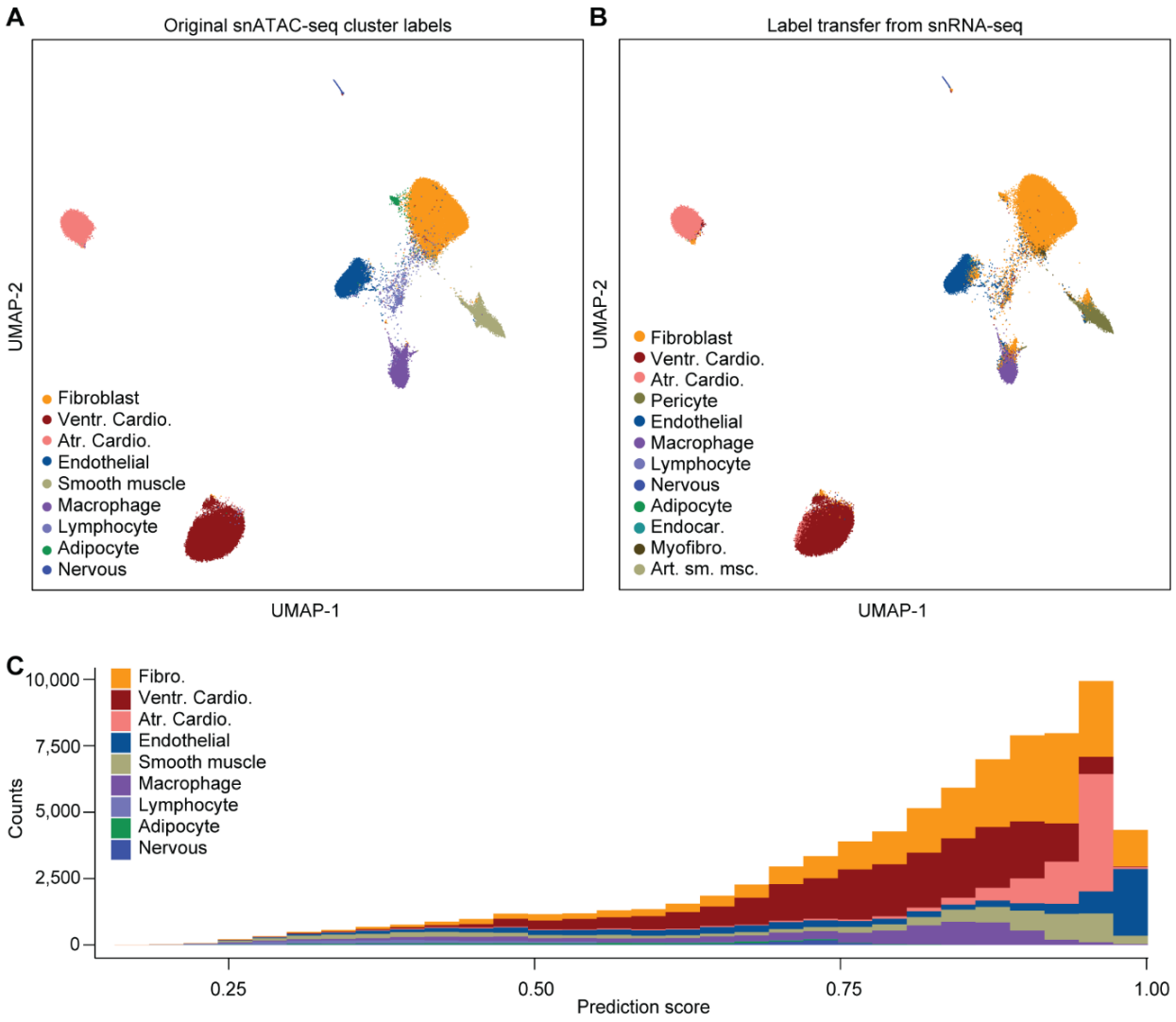
2.6 Supplemental Figures



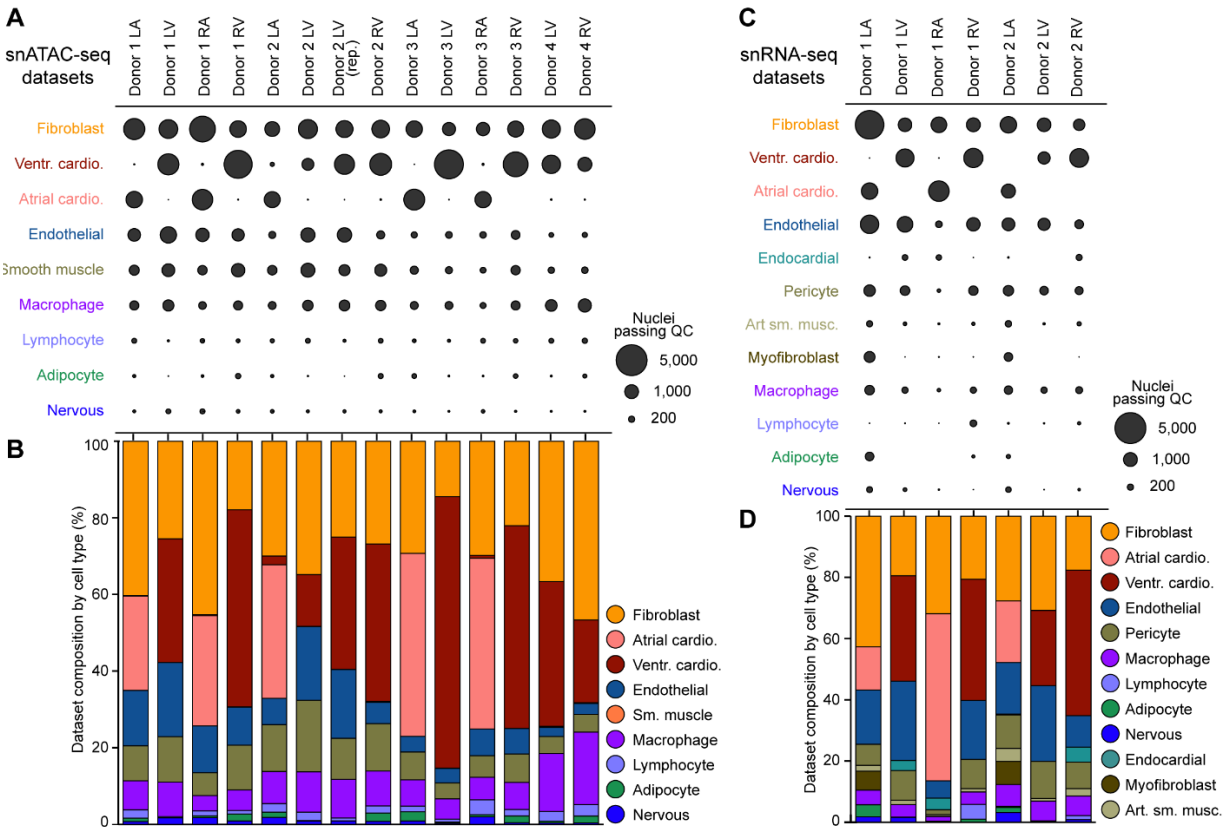
Supplemental Figure 1: Quality control for snATAC-seq datasets. (A) Density plots showing enrichment of fragments at transcription start sites (TSS enrichment) versus number of fragments per nucleus for each dataset. (B) Density plot of TSS enrichment versus number of fragments for all datasets combined. (C) Percentage of barcode collisions identified as heterotypic cell type collisions by Scrublet (Wolock et al., 2019) (top row), TSS enrichment (second row), fragments per nucleus (third row), duplicate read percentage (fourth row), and number of nuclei passing quality control (bottom row) for each snATAC-seq dataset.



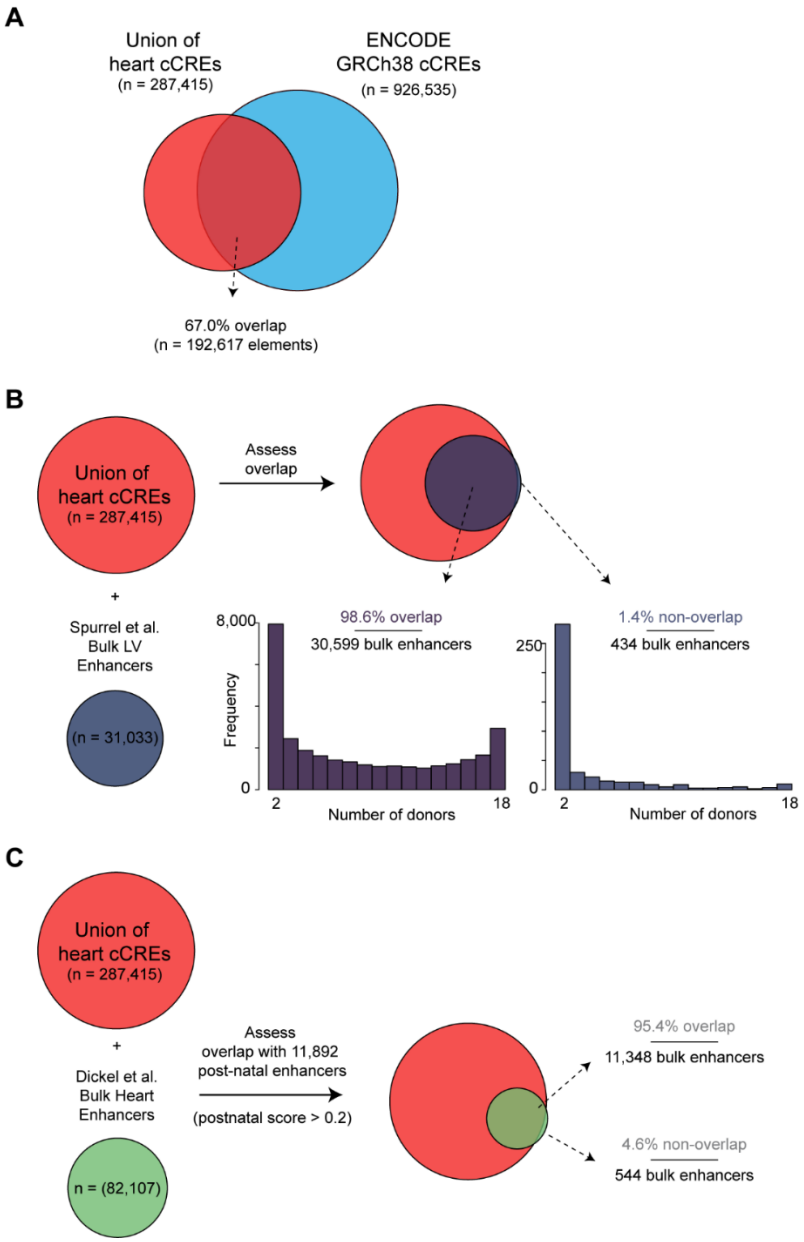
Supplemental Figure 2: Quality control for snRNA-seq datasets and annotation of snRNA-seq clusters. (A) Distribution of barcodes by unique molecular identifier (UMI) counts for nuclei (red; passing quality control) and background (black; not passing quality control) barcodes. (B) Distribution of doublet scores for all snRNA-seq nuclei that passed initial Cell Ranger (10x Genomics) and Seurat (Stuart et al., 2019) quality control. (C) Median genes detected per nucleus for each snRNA-seq dataset. (D) Median UMIs detected per nucleus for each snRNA-seq dataset. (E) Distribution of genes per nucleus on final snRNA-seq UMAP(Leland McInnes, 2018). (F) Distribution of UMIs per nucleus on final snRNA-seq UMAP(Leland McInnes, 2018). (G) Initial Seurat(Stuart et al., 2019) clustering result of snRNA-seq data showing 18 clusters, and dashed lines indicating final 12 major cell cluster annotations based on shared expression patterns (H). (H) Differential gene expression heatmap showing top 10 differentially expressed genes for each initial cluster by Seurat (Stuart et al., 2019). Initial clusters were merged into major cell clusters based on shared gene expression patterns as indicated above the heatmap. (I, J) UMAPs (Leland McInnes, 2018) showing chamber-of-origin for nuclei included in the final (I) snRNA-seq and (J) snATAC-seq datasets.



Supplemental Figure 3: Integration of snRNA-seq and snATAC-seq datasets. (A, B) Seurat (Stuart et al., 2019) was used to perform integration of chromatin accessibility and transcriptomes at the single cell level. (A) UMAP (Leland McInnes, 2018) showing nuclei colored based on original snATAC-seq cluster annotation (same as in Chapter 1 Fig. 1b). (B) UMAP (Leland McInnes, 2018) showing nuclei colored with cluster labels transferred from snRNA-seq. (C) 93% of nuclei showed a prediction score >0.5 indicating a match between chromatin accessibility and transcriptome. Histogram showing the prediction score distribution by original snATAC-seq cluster annotation.

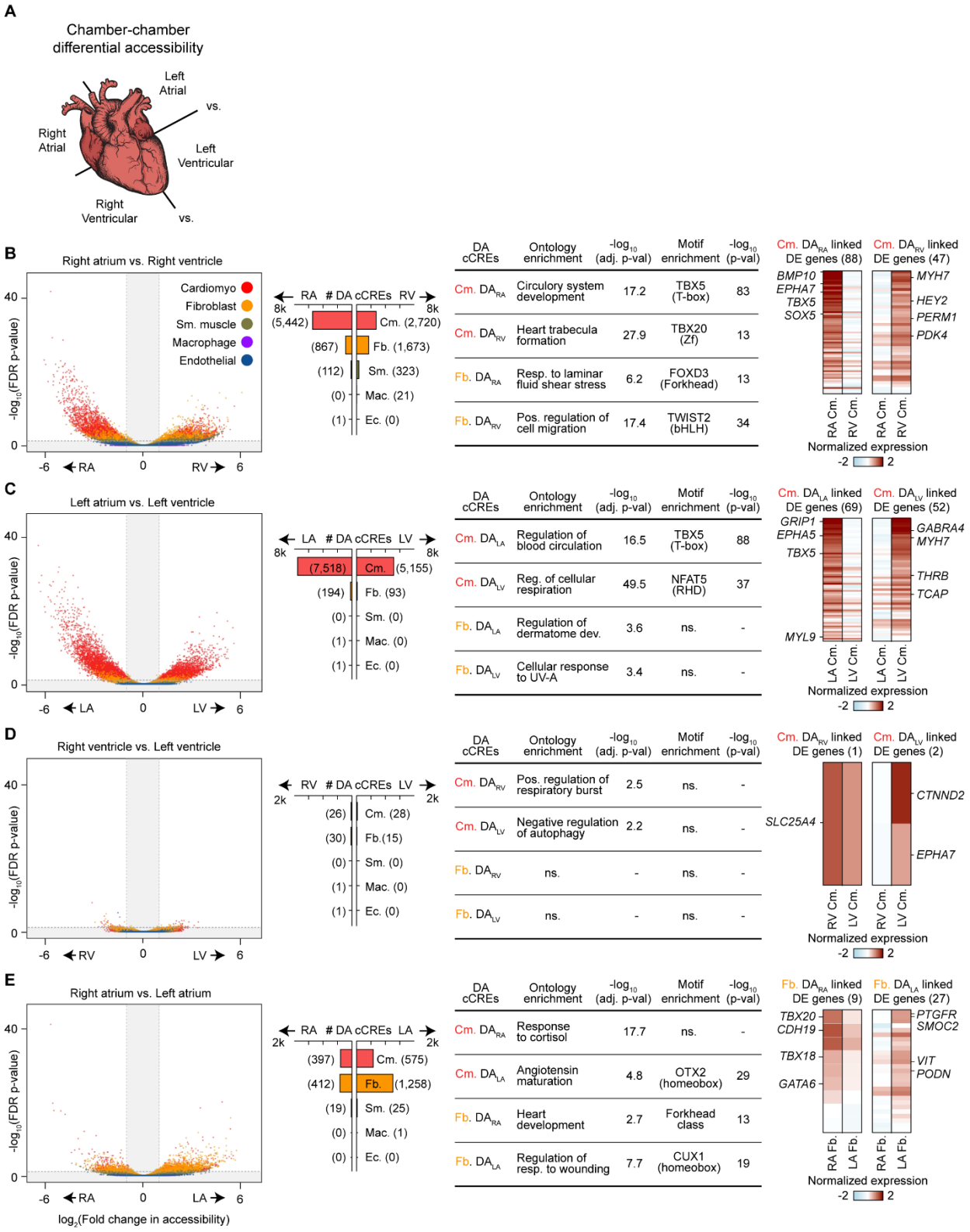


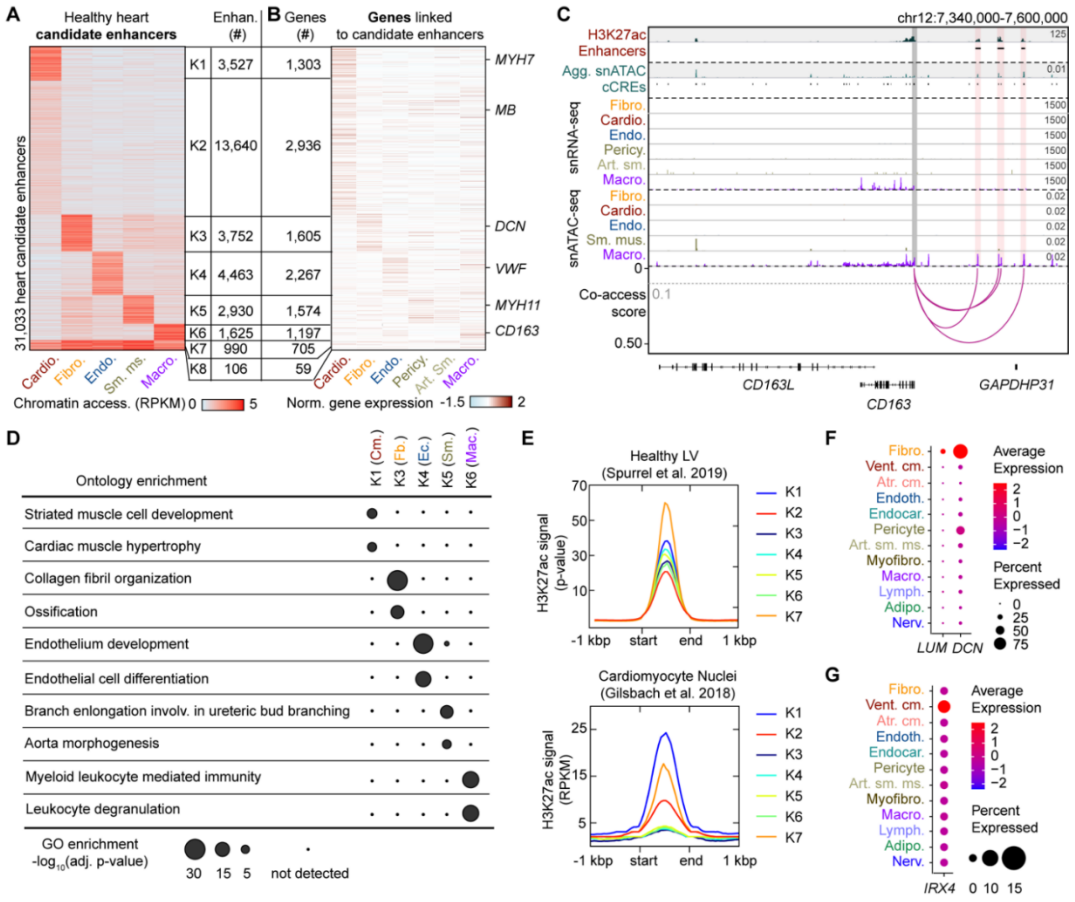
Supplemental Figure 4: Cellular composition of snATAC-seq and snRNA-seq datasets. (A) Dot plot showing number of nuclei passing quality control per cluster for each snATAC-seq dataset. (B) Bar plot showing cell type composition of each snATAC-seq dataset as percentage of cell types. (C) Dot plot showing number of nuclei passing quality control per cluster for each snRNA-seq dataset. (D) Bar plot showing cell type composition of each snRNA-seq dataset as percentage of cell types.



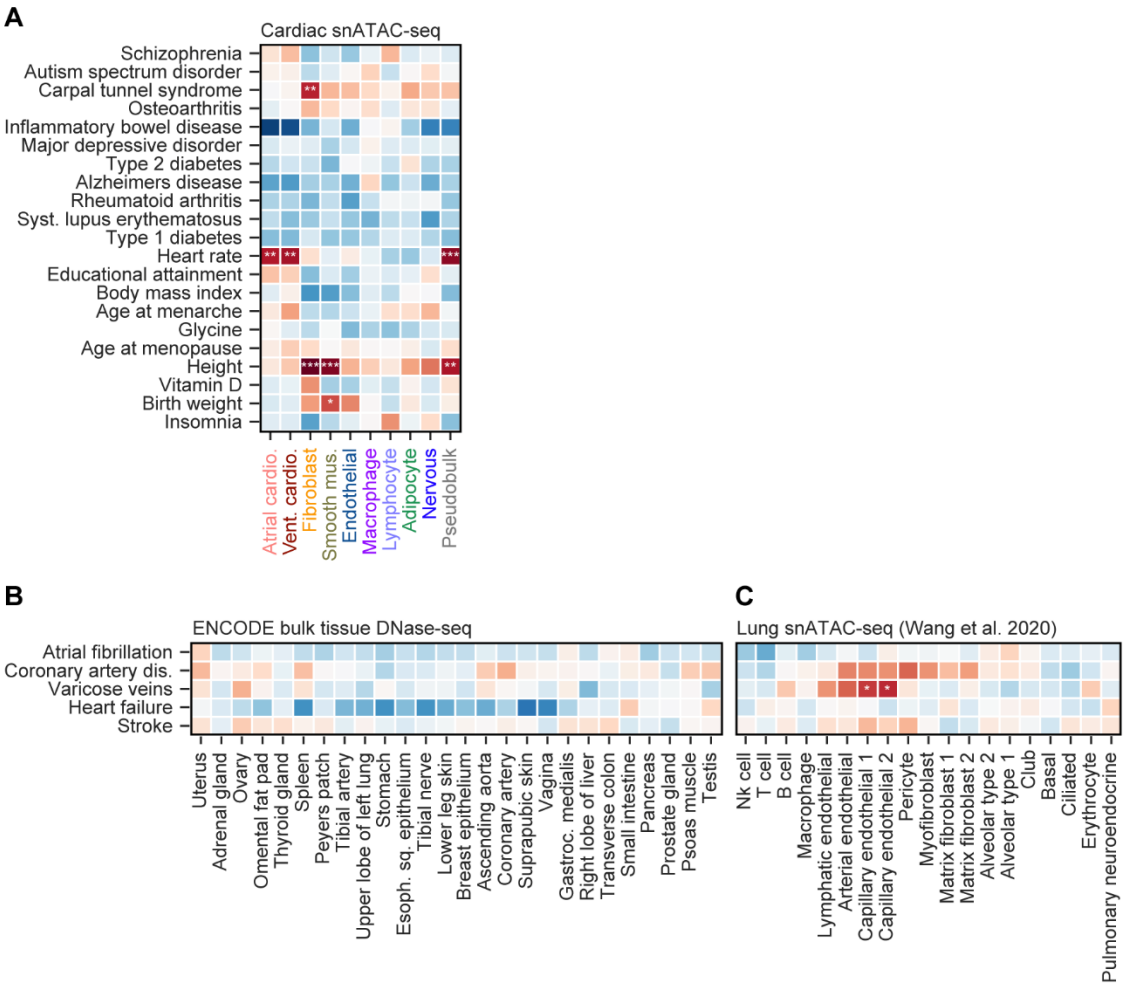
Supplemental Figure 5: Overlap of union of heart candidate cis regulatory elements (cCREs) with several reference datasets. (A) Overlap of the union of 287,415 heart cCREs from snATAC-seq with annotated cCREs in the human genome from the SCREEN database (Consortium, 2011, 2012b). (B) Overlap of union with healthy left ventricular candidate enhancers from 18 human donors (Spurrell et al., 2019). Arrows pointing from Venn diagram indicate number of overlapping (by at least one base pair) and non-overlapping genomic regions. Histograms display the number of donors harboring reported healthy heart enhancers (out of 18) for candidate enhancers that overlap union cCREs (left) and candidate enhancers that do not overlap union cCREs (right). (C) Overlap of heart cCREs with post-natal heart candidate enhancers (reported post-natal score > 0.2) from a meta-analysis of epigenomic data from human and mouse heart tissues (Dickel et al., 2016). Venn diagrams are not to scale.

Supplemental Figure 6: cCREs in cardiomyocytes and cardiac fibroblasts display chamber-dependent differences in accessibility. (A) Scheme for comparison of major cell types across individual heart chambers to identify differential accessible (DA) cCREs. (B-E) Comparisons were performed between (B) right atrium (RA) and right ventricle (RV), (C) left atrium (LA) and left ventricle (LV), (D) right ventricle (RV) and left ventricle (LV) and (E) right atrium (RA) and left atrium (LA). For each comparison the following data are displayed. Left: Volcano plots showing identification of differentially accessible (DA) cCREs in each cell type between indicated chambers. cCREs with $\log_2$(fold change) > 1 and FDR < 0.05 after Benjamini-Hochberg correction (outside the shaded area) were considered DA. Each dot represents a cCRE and the color indicates the cell type. Second from the left: Bar plots showing number of DA cCREs per cell type. Number of DA cCREs listed in brackets. Second from the right: GREAT ontology analysis (McLean et al., 2010) and transcription factor motif enrichment analysis result (Heinz et al., 2010) for the indicated DA cCREs. The best matches for selected *de novo* motifs (score > 0.7) are displayed. Statistical test for motif enrichment: hypergeometric test. P-values were not corrected for multiple testing. Ontology p-values were adjusted using Bonferroni correction. Right: Heatmaps showing normalized gene expression levels of differentially expressed genes linked to distal DA cCREs. Displayed are expression levels for putative target genes of distal DA cCREs for the cell type with most DA cCREs for the indicated chamber comparisons. Number of genes is shown in brackets. For lists of differentially expressed genes linked to distal DA cCREs for all comparisons in cardiomyocytes and fibroblasts see Chapter 1 Chapter 2, Table S12 (Cm. = cardiomyocyte, Fb. = fibroblast, ns. = no significant enrichment).


A
Chamber-chamber differential accessibility
Left Atrial
vs.
Right Atrial
Left Ventricular
Right Ventricular
vs.
B
Right atrium vs. Right ventricle
Cardiomyo
Fibroblast
Sm. muscle
Macrophage
Endothelial
-log10(FDR p-value)
RA # DA cCREs RV
8k
(5,442)
Cm. (2,720)
(867)
Fb. (1,673)
(112)
Sm. (323)
(0)
Mac. (21)
(1)
Ec. (0)
-6 RA 0 RV 6
DA cCREs
Ontology enrichment
-log10 (adj. p-val)
Motif enrichment
-log10 (p-val)
Cm. DARA
Circulory system development
17.2
TBX5 (T-box)
83
Cm. DARV
Heart trabecula formation
27.9
TBX20 (Zf)
13
Fb. DARA
Resp. to laminar fluid shear stress
6.2
FOXD3 (Forkhead)
13
Fb. DARV
Pos. regulation of cell migration
17.4
TWIST2 (bHLH)
34
Cm. DARA linked DE genes (88)
Cm. DARV linked DE genes (47)
BMP10
EPHA7
TBX5
SOX5
MYH7
HEY2
PERM1
PDK4
RA Cm.
RV Cm.
Normalized expression
-2 2
C
Left atrium vs. Left ventricle
LA # DA cCREs LV
(7,518)
Cm.
(5,155)
(194)
Fb. (93)
Sm. (0)
Mac. (0)
Ec. (0)
-6 LA 0 LV 6
Cm. DALA
Regulation of blood circulation
16.5
TBX5 (T-box)
88
Cm. DALV
Reg. of cellular respiration
49.5
NFAT5 (RHD)
37
Fb. DALA
Regulation of dermatome dev.
3.6
ns.
-
Fb. DALV
Cellular response to UV-A
3.4
ns.
-
Cm. DALA linked DE genes (69)
Cm. DALV linked DE genes (52)
GRIP1
EPHA5
TBX5
MYL9
GABRA4
MYH7
THRB
TCAP
LA Cm.
LV Cm.
Normalized expression
-2 2
D
Right ventricle vs. Left ventricle
RV # DA cCREs LV
2k
(26)
Cm. (28)
(30)
Fb.(15)
(0)
Sm. (0)
(1)
Mac. (0)
(1)
Ec. (0)
-6 RV 0 LV 6
Cm. DARV
Pos. regulation of respiratory burst
2.5
ns.
-
Cm. DALV
Negative regulation of autophagy
2.2
ns.
-
Fb. DARV
ns.
-
ns.
-
Fb. DALV
ns.
-
ns.
-
Cm. DARV linked DE genes (1)
Cm. DALV linked DE genes (2)
SLC25A4
CTNND2
EPHA7
RV Cm.
LV Cm.
Normalized expression
-2 2
E
Right atrium vs. Left atrium
RA # DA cCREs LA
(397)
Cm. (575)
(412)
Fb.
(1,258)
(19)
Sm. (25)
(0)
Mac. (1)
(0)
Ec. (0)
-6 RA 0 LA 6
log2(Fold change in accessibility)
Cm. DARA
Response to cortisol
17.7
ns.
-
Cm. DALA
Angiotensin maturation
4.8
OTX2 (homeobox)
29
Fb. DARA
Heart development
2.7
Forkhead class
13
Fb. DALA
Regulation of resp. to wounding
7.7
CUX1 (homeobox)
19
Fb. DARA linked DE genes (9)
Fb. DALA linked DE genes (27)
TBX20
CDH19
TBX18
GATA6
PTGFR
SMOC2
VIT
PODN
RA Fb.
LA Fb.
Normalized expression
-2 2

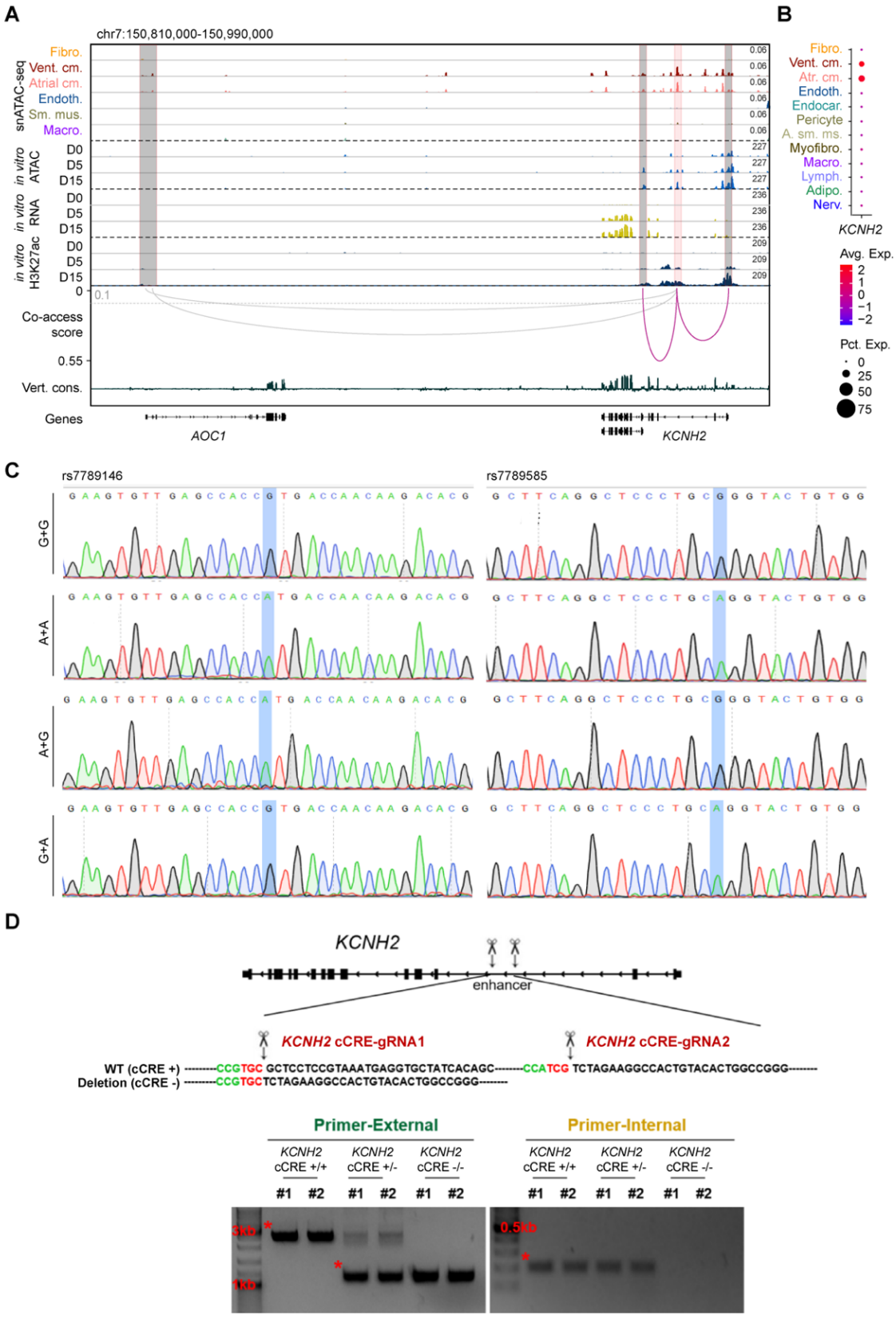


Supplemental Figure 7: Measurement of single cell chromatin accessibility signal within bulk candidate heart enhancers. (A) H3K27ac peaks from bulk healthy heart tissue samples (Spurrell et al., 2019) were attributed to major cardiac cell types using cell type-resolved chromatin accessibility data. Heatmap displays cell type-resolved chromatin accessibility RPKM (reads per kilobase per million mapped reads) values from left ventricular snATAC-seq datasets. Candidate enhancers were grouped based on chromatin accessibility pattern across cell clusters using K-means. (B) Heatmap displays cell-type resolved gene expression of putative enhancer target genes (co-accessibility score > 0.1) from left ventricular snRNA-seq datasets. (C) Genome browser tracks (Robinson et al., 2011) of H3K27ac in left ventricle tissue and cell type-resolved gene expression (snRNA-seq; scale = RPKM) and chromatin accessibility (snATAC-seq; scale = RPM) for several candidate heart enhancers (indicated by shaded red boxes) attributed to macrophages (K6 in panel (A)). The co-accessibility track shows linkages between the candidate enhancers and the promoter of *CD163* (cutoff > 0.1, grey dotted line). (D) GREAT analysis (McLean et al., 2010) of cell type-attributed candidate enhancers. Gene ontology enrichments are shown as Bonferroni-adjusted p-values. (E) Pileup tracks showing H3K27ac signal in bulk left ventricle datasets (Spurrell et al., 2019) (left) and from purified cardiomyocyte nuclei (Gilsbach et al., 2018) (right) from non-failing (NF) hearts in distinct groups of enhancers which were either associated with a cell type (K1-6 in panel (A)) or broadly accessible across cell types (K7 in panel (A)). H3K27ac signal in cardiomyocyte nuclei data was highest in the cardiomyocyte-attributed candidate enhancers as well as the widely accessible candidate enhancers (K1,2,7), whereas signal strength in left ventricular tissue was highest in widely accessible enhancers and comparable between groups of cell type-specific candidate enhancers. (F, G) Dot plots illustrating expression of (F) *LUM* and *DCN*, and (G) *IRX4* in individual snRNA-seq clusters in the adult human heart.



Supplemental Figure 8: Association of cardiac cell types, bulk tissues, and lung cell types with variants for cardiovascular and non-cardiovascular diseases and traits. Heatmap showing enrichment of variants associated with non-cardiovascular diseases and phenotypic traits from genome wide association studies in cardiac cell type-resolved cCREs and from all cardiac cell types combined (pseudobulk). (B) Heatmap showing enrichment of risk variants associated with cardiovascular diseases in cCREs identified by DNase-seq in non-heart tissues from ENCODE (Consortium, 2012a; Consortium et al., 2020; Vierstra et al., 2020). (C) Heatmap showing enrichment of risk variants associated with cardiovascular diseases in lung cell type-resolved cCREs from snATAC-seq of the human lung (Wang et al., 2020). Total cCREs identified independently in each cell type or dataset were used as input for analysis. Z-scores for enrichment are displayed, and were used to compute two-sided p-values for enrichments. P-values were corrected using the Benjamini-Hochberg procedure for multiple tests (*: FDR < 0.1; **: FDR < 0.01; ***: FDR < 0.001).

Supplemental Figure 9: Validation of KCNH2-associated candidate enhancer. (A) Genome browser tracks (Robinson et al., 2011) displaying cell type-resolved chromatin accessibility (scale = RPM) and co-accessibility from snATAC-seq as well as chromatin accessibility, H3K27ac signal, and gene expression during hPSC-cardiomyocyte differentiation (scale = RPKM). For illustration purposes, the co-accessibility track shows linkages between the AF variant-containing cCRE and annotated gene promoters (co-accessibility > 0.1, grey dotted line). The grey arc represents links to the promoter of *AOC1* which was not expressed. Fig. 5C shows a zoom into this locus. (B) Dot plot illustrating expression of *KCNH2* in individual snRNA-seq clusters in the adult human heart. (C) Representative Sanger sequencing peak map at *KCNH2* intronic cCRE showing the risk allele for AF (top row rs7789146-G / rs7789585-G), the non-risk allele for AF (second row rs7789146-A / rs7789585-A), and the constructs with a combination of risk and non-risk alleles (rs7789146-A / rs7789585-G; third tow and rs7789146-G / rs7789585-A bottom row) used for luciferase assay. Blue highlighted regions indicate positions of variants. (D) Schematic representation of the strategy for deletion of the *KCNH2* enhancer. The paired gRNAs (gRNA-1 and gRNA-2) were designed to target upstream and downstream of the *KCNH2* enhancer. Bottom panels show genomic DNA PCR verification of deletion in the H9-hTnnTZ-pGZ-D2 cell line. The red asterisk indicates specific bands.


A
chr7:150,810,000-150,990,000
snATAC-seq
Fibro.
Vent. cm.
Atrial cm.
Endoth.
Sm. mus.
Macro.
in vitro ATAC
D0
D5
D15
in vitro RNA
D0
D5
D15
in vitro H3K27ac
D0
D5
D15
Co-access score
0.55
Vert. cons.
Genes
AOC1
KCNH2
B
Fibro.
Vent. cm.
Atr. cm.
Endoth.
Endocar.
Pericyte
A. sm. ms.
Myofibro.
Macro.
Lymph.
Adipo.
Nerv.
KCNH2
Avg. Exp.
Pct. Exp.
C
rs7789146
rs7789585
G+G
A+A
A+G
G+A
D
KCNH2
enhancer
KCNH2 cCRE-gRNA1
KCNH2 cCRE-gRNA2
WT (cCRE +) --------CCGTGC GCTCCTCCGTAAATGAGGTGCTATCACAGC--------CCATCG TCTAGAAGGCCACTGTACACTGGCCGGG--------
Deletion (cCRE -) --------CCGTGC TCTAGAAGGCCACTGTACACTGGCCGGG--------
Primer-External
Primer-Internal
KCNH2 cCRE +/+
KCNH2 cCRE +/-
KCNH2 cCRE -/-
#1 #2 #1 #2 #1 #2
3kb
1kb
0.5kb

2.7 Supplemental Tables

Supplemental Table 1: Clinical metadata for heart samples.

Supplemental Table 2: Quality control and cell type composition data for each snATAC dataset.

Supplemental Table 3: Quality control, cell type composition, and integration with snATAC-seq results for snRNA-seq datasets.

Supplemental Table 4: snRNA-seq gene expression by major cluster and major cluster-specific genes.

Supplemental Table 5: Union of 287,415 cCREs in the cell types of the human heart.

Supplemental Table 6: List of 19,447 cell type-specific cCREs.

Supplemental Table 7: GREAT analysis for cell type-specific cCREs.

Supplemental Table 8: ChromVAR motif enrichment results in snATAC-seq cell clusters.

Supplemental Table 9: HOMER motif enrichment results for cell type-specific cCREs.

Supplemental Table 10: Differentially accessible cCREs between heart chambers.

Supplemental Table 11: Co-accessible cCRE pairs from Cicero.

Supplemental Table 12: Lists of differentially accessible (DA) cCREs linked to differentially expressed genes.

Supplemental Table 13: GREAT analysis for differentially accessible (DA) cCREs between heart chambers in cardiomyocytes and fibroblasts.

Supplemental Table 14: HOMER motif enrichments for differentially accessible (DA) cCREs between heart chambers in cardiomyocytes and fibroblasts.

Supplemental Table 15: RPKM values and cluster membership for healthy and disease-associated candidate heart enhancers in different cell types.

Supplemental Table 16: GREAT analysis for distinct groups of cell type-attributed candidate heart enhancer.

Supplemental Table 17: HOMER motif enrichment results for distinct groups of cell type-attributed candidate heart enhancers.

Supplemental Table 18: Studies for non-cardiovascular disease and non-disease trait GWAS used for LD score regression.

Supplemental Table 19: 38 fine mapped risk variants associated with atrial fibrillation within cardiomyocyte cCREs.

Supplemental Table 20: Primer sequences with indexes for snATAC-seq libraries.

Supplemental Table 21: Primer sets used in qPCR assays.

2.8 Competing Interests

B.R. is a shareholder and consultant of Arima Genomics, Inc. and co-founder of Epigenome Technologies. K.J.G is a consultant of Genentech, and shareholder in Vertex Pharmaceuticals. A.D.M. is a cofounder and Scientific Advisor to Insilicomed, Inc. and Vektor Medical, Inc. These relationships have been disclosed to and approved by the UCSD Independent Review Committee. The authors declare no other competing interests.

2.9 Acknowledgements

J.D.H., S.P., N.C.C., and B.R. conceived the project. J.D.H. and J.B. carried out snATAC-seq and snRNA-seq library preparation with help from X.H. F.Z. and T.W. performed validation experiments. J.D.H., O.P., J.B., K.Z., J.C., Y.L., and S.P. performed data analysis. O.P. created the web portal. X.H., E.F., Y.Z., A.W., A.D.M., K.J.G., and N.C.C. contributed to experimental design and computational analyses. J.D.H., S.P., N.C.C., and B.R. wrote the manuscript. All authors edited and approved the manuscript.

We thank B. Li for bioinformatics support. We thank K. Jepsen and the UCSD IGM Genomics Center for sequencing the snRNA-seq libraries. We thank the QB3 Macrolab at UC Berkeley for purification of the Tn5 transposase. We also thank members of the Ren, Chi, Gaulton, McCulloch labs and the Center for Epigenomics for feedback and discussions. This work was supported by the Ludwig Institute for Cancer Research (B.R.), and the National Institutes of Health (1UM1HL128773-01 to N.C., B.R., U01 HL126273 and R01 HL137100 to A.D.M.). J.D.H. was supported in part by a Ruth L. Kirschstein Institutional National Research Service Award T32 GM008666 from the National Institute of General Medical Sciences. Work at the Center for Epigenomics was supported in part by the UC San Diego School of Medicine.

The content of Chapter 2, in full, is a reprint of the material as it appears in *Science Advances* 2021. Hocker, James D.; Poirion, Olivier†; Zhu, Fugui†; Buchanan, Justin; Zhang, Kai; Chiou, Joshua; Wang, Tsui-min; Hou, Xiaomeng; Li, Yang, Zhang, Yanxiao; Farah, Elie N.; Wang, Allen; McCulloch, Andrew D.; Gaulton, Kyle J.; Ren, Bing*; Chi, Neil C.*; Preissl. Sebastian*. The dissertation author was the primary investigator and author of this paper.

Chapter 3: A single cell atlas of chromatin accessibility in the human genome

3.1 Abstract

Current catalogs of regulatory sequences in the human genome are still incomplete and lack cell type resolution. To profile the activity of gene regulatory elements in diverse cell types and tissues in the human body, we applied single-cell chromatin accessibility assays to 30 adult human tissue types from multiple donors. We integrated these datasets with single-cell chromatin accessibility data from 15 fetal tissue types to reveal the status of open chromatin for approximately 1.2 million candidate *cis*-regulatory elements (cCREs) in 222 distinct cell types comprised of >1.3 million nuclei. We used these chromatin accessibility maps to delineate cell type-specificity of fetal and adult human cCREs and to systematically interpret the noncoding variants associated with complex human traits and diseases. This rich resource provides a foundation for the analysis of gene regulatory programs in human cell types across tissues, life stages, and organ systems.

3.2 Introduction

The human body is comprised of various organs, tissues and cell types, each with highly specialized functions. The genes expressed in each tissue and cell type – and in turn their physiologic roles in the body – are regulated by *cis*-regulatory elements such as enhancers and promoters (Carter and Zhao, 2020). These sequences dictate the expression patterns of target genes by recruiting sequence specific transcription factors (TFs) in a cell-type specific manner (Shlyueva et al., 2014). Upon binding of TFs, *cis*-regulatory elements frequently adopt conformational changes such that they are more accessible to endonucleases or transposases, enabling genome-wide discovery of candidate *cis*-regulatory elements (cCREs) by combining assays incorporating

these enzymes with high throughput sequencing (Buenrostro et al., 2013; John et al., 2013; Klemm et al., 2019). However, conventional assays have, in large part, used heterogeneous tissues as input materials to produce population average measurements, and consequently, the current catalogs of candidate regulatory sequences in the human genome (Andersson et al., 2014; Meuleman et al., 2020; Moore et al., 2020; Roadmap Epigenomics et al., 2015; Shen et al., 2012) still lack information about the cell type-specific activities of most elements. This limitation has hampered our ability to study gene regulatory programs in distinct human cell types and to interpret the noncoding DNA in the human genome.

Genome wide association studies (GWAS) have identified hundreds of thousands of genetic variants associated with a broad spectrum of human traits and diseases. The large majority of these variants are noncoding (Claussnitzer et al., 2020). Observations that annotated *cis*-regulatory elements in disease-relevant tissues and cell types are enriched for noncoding disease risk variants (Ernst et al., 2011; Maurano et al., 2012; Roadmap Epigenomics et al., 2015) led to the hypothesis that a major mechanism by which noncoding variants influence disease risk is by affecting transcriptional regulatory elements in specific cell types. However, annotation of these noncoding risk variants has been hindered by a lack of cell type-resolved maps of regulatory elements in the human genome. While innovative approaches to distinguish causal variants from local variants in linkage disequilibrium (LD) using fine mapping (Wakefield, 2009), and to link variants to target genes using co-accessibility of open chromatin regions in single-cells (Pliner et al., 2018) or 3-dimensional chromosomal contact-based linkage scores (Nasser et al., 2021), have made important strides toward the prioritization of causal variants and the prediction of their target genes, functional interpretation of the noncoding variants continues to be challenging.

Single-cell omics technologies, enabled by droplet-based, combinatorial barcoding or other

approaches, have now enabled the profiling of transcriptome, epigenome and chromatin organization from complex tissues at single-cell resolution (Grosselin et al., 2019; Klein et al., 2015; Lake et al., 2018; Luo et al., 2017a; Macosko et al., 2015; Preissl et al., 2018). In particular, combinatorial cellular barcoding-based assays such as sci-ATAC-seq (Cusanovich et al., 2015) have permitted the identification of cCREs in single nuclei without the need for physical purification of individual cell types. The resulting data can be used to deconvolute cell types from mixed cell populations and to dissect cell type-specific transcriptomic and epigenomic states in primary tissues. While these tools have been applied to mammalian tissues including murine biosamples (Cusanovich et al., 2018; Lareau et al., 2019; Li et al., 2021; Preissl et al., 2018; Sinnamon et al., 2019), human fetal tissues (Domcke et al., 2020; Trevino et al., 2021), and a few individual adult human organ systems (Chiou et al., 2021; Corces et al., 2020; Hocker et al., 2021; Wang et al., 2020), we still lack comprehensive cell-type-resolved maps of cCREs from most primary tissues of the adult human body.

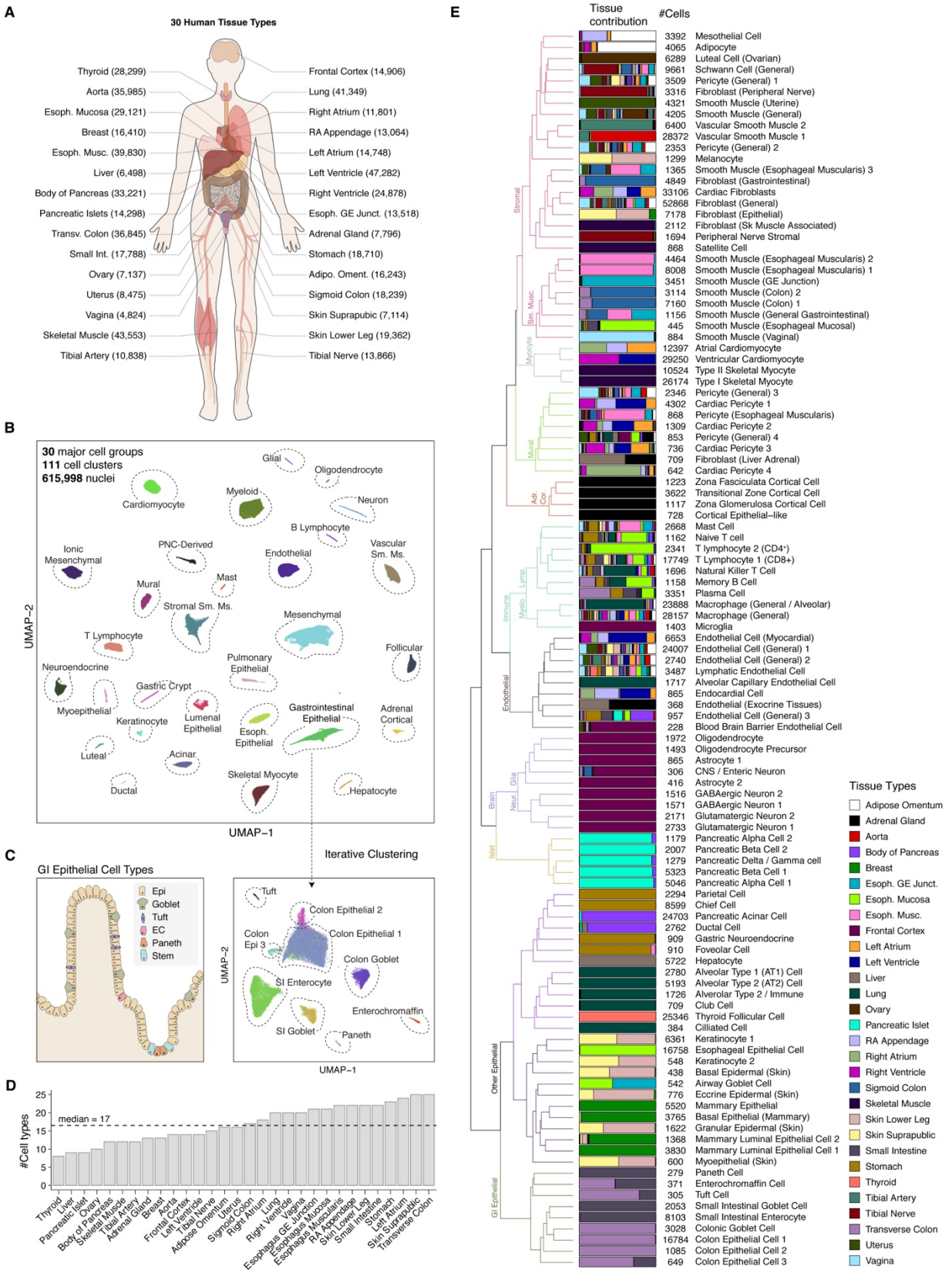
In the present study we used a modified sci-ATAC-seq protocol optimized for flash frozen primary tissues (Hocker et al., 2021; Preissl et al., 2018) to profile chromatin accessibility in 30 adult human tissue types from multiple donors. We profiled 615,998 nuclei from these tissues, grouped them into 111 distinct cell types based on similarity in chromatin landscapes, and identified a union of 890,130 open chromatin regions corresponding to cCREs from the resulting maps. We next integrated these data with a recent fetal cell atlas of chromatin accessibility (Domcke et al., 2020) to reveal open chromatin profiles for >1.3 million cells across the human lifespan, and chromatin accessibility maps at 1,154,611 cCREs covering 14.8% of the genome for 222 cell types. Finally, we used this cCRE atlas to interpret cell types and target genes for noncoding variants associated with 240 complex human traits and diseases, reveal cell type-disease

associations and suggest relevant therapeutic targets in human cell types. We created an interactive web atlas to disseminate this resource [CATLAS, Cis-element ATLAS] http://catlas.org/humanenhancer.

3.3 Results

To generate a cell atlas of cCREs in the adult human body, we performed sci-ATAC-seq (Cusanovich et al., 2015; Preissl et al., 2018) with primary tissue samples collected from 30 distinct anatomic sites in postmortem adult human donors (Figure 6A, Table S22).

Figure 6: Single-cell chromatin accessibility analysis of 30 adult human primary tissues. A total of 92 biosamples from 30 tissue types, were used for sci-ATAC-seq. The number of nuclei profiled per tissue is denoted in parentheses. B) Clustering of 615,998 nuclei revealed 30 major cell groups. Each dot represents a nucleus colored by cluster ID. Embedding was created by Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018). C) An example illustrating subclusters within the major cell group of gastrointestinal (GI) epithelial cells revealed by iterative clustering. D) Bar plot showing the number of cell types identified in each of the 30 human tissues, counting only cell types constituting >0.2% of all cells in the given tissue. E) Distribution of cell types across human tissues. The dendrogram on the left was created by hierarchical clustering of cell clusters based on chromatin accessibility. The bar chart represents relative contributions of tissues to cell clusters. Raw data are available on Mendeley Data: 10.17632/yv4fzv6cnm.1.


A
30 Human Tissue Types
Thyroid (28,299)
Aorta (35,985)
Esoph. Mucosa (29,121)
Breast (16,410)
Esoph. Musc. (39,830)
Liver (6,498)
Body of Pancreas (33,221)
Pancreatic Islets (14,298)
Transv. Colon (36,845)
Small Int. (17,788)
Ovary (7,137)
Uterus (8,475)
Vagina (4,824)
Skeletal Muscle (43,553)
Tibial Artery (10,838)
Frontal Cortex (14,906)
Lung (41,349)
Right Atrium (11,801)
RA Appendage (13,064)
Left Atrium (14,748)
Left Ventricle (47,282)
Right Ventricle (24,878)
Esoph. GE Junct. (13,518)
Adrenal Gland (7,796)
Stomach (18,710)
Adipo. Oment. (16,243)
Sigmoid Colon (18,239)
Skin Suprapubic (7,114)
Skin Lower Leg (19,362)
Tibial Nerve (13,866)
B
30 major cell groups
111 cell clusters
615,998 nuclei
Glial
Oligodendrocyte
Cardiomyocyte
Myeloid
Neuron
B Lymphocyte
Ionic Mesenchymal
PNC-Derived
Endothelial
Vascular Sm. Ms.
Mural
Mast
Stromal Sm. Ms.
Mesenchymal
T Lymphocyte
Follicular
Neuroendocrine
Pulmonary Epithelial
Gastric Crypt
Myoepithelial
Lumenal Epithelial
Gastrointestinal Epithelial
Adrenal Cortical
Keratinocyte
Esoph. Epithelial
Luteal
Acinar
Skeletal Myocyte
Ductal
Hepatocyte
UMAP–2
UMAP–1
Iterative Clustering
C
GI Epithelial Cell Types
Epi
Goblet
Tuft
EC
Paneth
Stem
Tuft
Colon Epithelial 2
Colon Epi 3
Colon Epithelial 1
Colon Goblet
SI Enterocyte
Enterochromaffin
SI Goblet
Paneth
UMAP–2
UMAP–1
D
#Cell types
median = 17
Thyroid
Liver
Pancreatic Islet
Ovary
Body of Pancreas
Skeletal Muscle
Tibial Artery
Adrenal Gland
Breast
Aorta
Frontal Cortex
Left Ventricle
Tibial Nerve
Adipose Omentum
Uterus
Sigmoid Colon
Right Atrium
Lung
Right Ventricle
Vagina
Esophagus GE Junction
Esophagus Mucosa
Esophagus Muscularis
RA Appendage
Skin Lower Leg
Small Intestine
Stomach
Left Atrium
Skin Suprapubic
Transverse Colon
E
Tissue contribution
#Cells
3392 Mesothelial Cell
4065 Adipocyte
6289 Luteal Cell (Ovarian)
9661 Schwann Cell (General)
3509 Pericyte (General) 1
3316 Fibroblast (Peripheral Nerve)
4321 Smooth Muscle (Uterine)
4205 Smooth Muscle (General)
6400 Vascular Smooth Muscle 2
28372 Vascular Smooth Muscle 1
2353 Pericyte (General) 2
1299 Melanocyte
1365 Smooth Muscle (Esophageal Muscularis) 3
4849 Fibroblast (Gastrointestinal)
33106 Cardiac Fibroblasts
52868 Fibroblast (General)
7178 Fibroblast (Epithelial)
2112 Fibroblast (Sk Muscle Associated)
1694 Peripheral Nerve Stromal
868 Satellite Cell
4464 Smooth Muscle (Esophageal Muscularis) 2
8008 Smooth Muscle (Esophageal Muscularis) 1
3451 Smooth Muscle (GE Junction)
3114 Smooth Muscle (Colon) 2
7160 Smooth Muscle (Colon) 1
1156 Smooth Muscle (General Gastrointestinal)
445 Smooth Muscle (Esophageal Mucosal)
884 Smooth Muscle (Vaginal)
12397 Atrial Cardiomyocyte
29250 Ventricular Cardiomyocyte
10524 Type II Skeletal Myocyte
26174 Type I Skeletal Myocyte
2346 Pericyte (General) 3
4302 Cardiac Pericyte 1
868 Pericyte (Esophageal Muscularis)
1309 Cardiac Pericyte 2
853 Pericyte (General) 4
736 Cardiac Pericyte 3
709 Fibroblast (Liver Adrenal)
642 Cardiac Pericyte 4
1223 Zona Fasciculata Cortical Cell
3622 Transitional Zone Cortical Cell
1117 Zona Glomerulosa Cortical Cell
728 Cortical Epithelial–like
2668 Mast Cell
1162 Naive T cell
2341 T lymphocyte 2 (CD4+)
17749 T Lymphocyte 1 (CD8+)
1696 Natural Killer T Cell
1158 Memory B Cell
3351 Plasma Cell
23888 Macrophage (General / Alveolar)
28157 Macrophage (General)
1403 Microglia
6653 Endothelial Cell (Myocardial)
24007 Endothelial Cell (General) 1
2740 Endothelial Cell (General) 2
3487 Lymphatic Endothelial Cell
1717 Alveolar Capillary Endothelial Cell
865 Endocardial Cell
368 Endothelial (Exocrine Tissues)
957 Endothelial Cell (General) 3
228 Blood Brain Barrier Endothelial Cell
1972 Oligodendrocyte
1493 Oligodendrocyte Precursor
865 Astrocyte 1
306 CNS / Enteric Neuron
416 Astrocyte 2
1516 GABAergic Neuron 2
1571 GABAergic Neuron 1
2171 Glutamatergic Neuron 2
2733 Glutamatergic Neuron 1
1179 Pancreatic Alpha Cell 2
2007 Pancreatic Beta Cell 2
1279 Pancreatic Delta / Gamma cell
5323 Pancreatic Beta Cell 1
5046 Pancreatic Alpha Cell 1
2294 Parietal Cell
8599 Chief Cell
24703 Pancreatic Acinar Cell
2762 Ductal Cell
909 Gastric Neuroendocrine
910 Foveolar Cell
5722 Hepatocyte
2780 Alveolar Type 1 (AT1) Cell
5193 Alveolar Type 2 (AT2) Cell
1726 Alverolar Type 2 / Immune
709 Club Cell
25346 Thyroid Follicular Cell
384 Cilliated Cell
6361 Keratinocyte 1
16758 Esophageal Epithelial Cell
548 Keratinocyte 2
438 Basal Epidermal (Skin)
542 Airway Goblet Cell
776 Eccrine Epidermal (Skin)
5520 Mammary Epithelial
3765 Basal Epithelial (Mammary)
1622 Granular Epidermal (Skin)
1368 Mammary Luminal Epithelial Cell 2
3830 Mammary Luminal Epithelial Cell 1
600 Myoepithelial (Skin)
279 Paneth Cell
371 Enterochromaffin Cell
305 Tuft Cell
2053 Small Intestinal Goblet Cell
8103 Small Intestinal Enterocyte
3028 Colonic Goblet Cell
16784 Colon Epithelial Cell 1
1085 Colon Epithelial Cell 2
649 Colon Epithelial Cell 3
Stromal
Sm. Musc.
Myocyte
Mural
Adr. Cor.
Immune
Lymp.
Myelo.
Endothelial
Brain
Neur.
Glia
Islet
Other Epithelial
GI Epithelial
Tissue Types
Adipose Omentum
Adrenal Gland
Aorta
Body of Pancreas
Breast
Esoph. GE Junct.
Esoph. Mucosa
Esoph. Musc.
Frontal Cortex
Left Atrium
Left Ventricle
Liver
Lung
Ovary
Pancreatic Islet
RA Appendage
Right Atrium
Right Ventricle
Sigmoid Colon
Skeletal Muscle
Skin Lower Leg
Skin Suprapublic
Small Intestine
Stomach
Thyroid
Tibial Artery
Tibial Nerve
Transverse Colon
Uterus
Vagina

Tissue samples were chosen to survey a breadth of human organ systems which differed in their nuclear compositions and sensitivities to mechanical dissociation, posing a technical challenge. We thus optimized nuclear isolation methods and buffer conditions for different tissue types (Table S22, see Methods). Subsequently, we generated sci-ATAC-seq datasets using a semi-automated workflow (Hocker et al., 2021; Preissl et al., 2018) and sequenced resulting libraries to 6,464 raw sequence reads per nucleus on average, with a median read duplication rate of 44.88% (Table S23). After filtering out lower quality nuclei and potential doublets, we finally obtained high quality open chromatin profiles for 615,998 nuclei, with a median of 2,822 unique open chromatin fragments per nucleus and an average transcription start site (TSS) enrichment score of 12.8 (±3.2) per nucleus (Figure 6B, Table S23, Figure S18, Figure S19, see Methods).

Analyzing large single-cell chromatin accessibility datasets has been challenging. In the latest development of SnapATAC (Fang et al., 2021), we further improved its scalability to handle millions of cells. Using this algorithm, we first identified 30 major cell groups (Figure 6B), 22 (73%) of which were found to consist of multiple subclusters during a second round of clustering analysis (see Methods and Figure S20). Altogether we uncovered a total of 111 distinct cell clusters (Figure 6B-E).

3.3.1 Single-cell chromatin accessibility analysis of adult human primary tissues

To generate a cell atlas of cCREs in the adult human body, we performed sci-ATAC-seq (Preissl et al., 2018) with primary tissue samples collected from 30 distinct anatomic sites in postmortem adult human donors (Figure 6A, Table S22). Tissue samples were chosen to survey a breadth of human organ systems which differed in their nuclear compositions and sensitivities to mechanical dissociation, posing a technical challenge. We thus optimized nuclear isolation

methods and buffer conditions for different tissue types (Table S22, see Methods). Subsequently, we generated sci-ATAC-seq datasets using a semi-automated workflow (Hocker et al., 2021; Preissl et al., 2018) and sequenced resulting libraries to 6,464 raw sequence reads per nucleus on average, with a median read duplication rate of 44.88% (Table S23). After filtering out lower quality nuclei and potential doublets, we finally obtained high quality open chromatin profiles for 615,998 nuclei, with a median of 2,822 unique open chromatin fragments per nucleus and an average transcription start site (TSS) enrichment score of 12.8 (±3.2) per nucleus (Figure 6B, Table S23, see Methods and Figure S18).

Analyzing large single-cell chromatin accessibility datasets has been challenging. In the latest development of SnapATAC (Fang et al., 2021), we further improved its scalability to handle millions of cells. Using this algorithm, we first identified 30 major cell groups (Figure 6B), 22 (73%) of which were found to consist of multiple subclusters during a second round of clustering analysis (see Methods and Figure S20). Altogether we uncovered a total of 111 distinct cell clusters (Figure 6B-E).

3.3.2 Annotation of major and sub-classes of human cell types

To annotate the resulting cell clusters, we first curated a set of marker genes from the PanglaoDB marker gene database (Franzén et al., 2019) corresponding to expected human cell types. We utilized chromatin accessibility at the promoter as a proxy for gene activity and computed cell-type enrichment scores for each of the 111 clusters to create initial cell cluster annotations (Figure S21, see Methods). We next manually reviewed these assignments based on focused consideration of marker gene accessibility (Figure S22, see Methods). Altogether, we annotated each of the 30 major cell groups and all 111 distinct clusters with a cell type label (Figure

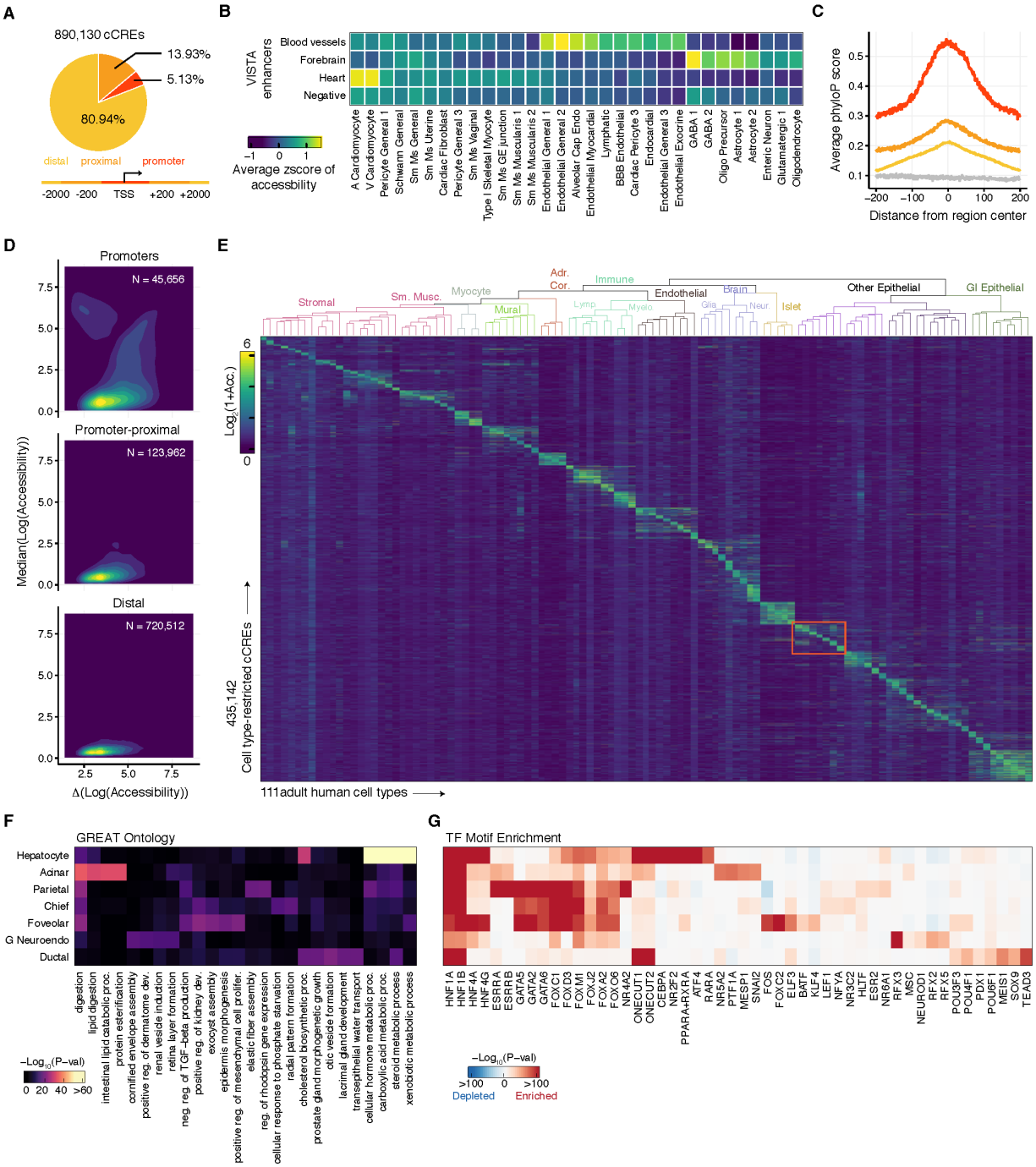
6E, Table S24). For example, within the major cell group of gastrointestinal epithelial cells, higher resolution subclustering and annotation revealed three clusters of colon epithelial cells, one cluster of enterocytes from the small intestine, two clusters of goblet cells from the colon and small intestine respectively, and three rare populations with distinct chromatin accessibility profiles including enterochromaffin cells (0.060% of total nuclei), tuft cells (0.050% of total nuclei), and Paneth cells (0.045% of total nuclei) (Figure 6B-C).

Encouragingly, several prevalent cell types detected in most tissue samples such as endothelial cells and myeloid cells clustered based on cell type rather than tissue of origin or individual (Figure 6E). On the other hand, tissue-resident fibroblasts clustered into seven subtypes with diverse tissues of origin for each (Figure 6E). Notably, the majority of the 111 cell types exhibited high tissue specificity. For example, highly specialized cell types such as follicular cells, pneumocytes, and hepatocytes were restricted to only one tissue type, reflecting their tissue-specific functions (Figure 6E). Finally, we observed that the cell types we identified by sci-ATAC-seq are highly concordant with those identified by single-cell RNA-seq experiments on corresponding tissues (See Methods and Tables 3-4).

3.3.3 An atlas of cCREs in adult human cell types

To identify accessible chromatin regions in each of the 111 cell types, we aggregated chromatin accessibility profiles from all nuclei comprising each cell cluster and applied a peak calling procedure optimized for single-cell data (Figure S23, see Methods). We then merged these accessible chromatin regions to obtain a list of 890,130 non-overlapping cCREs (Figure 7A).

Figure 7: An atlas of cCREs in adult human cell types. A) Classification of 890,130 cCREs across the human genome based on their distances to annotated TSSs. B) Heatmap showing the average chromatin accessibility for each of four groups (blood vessel, forebrain, heart, negative control) of validated tissue-specific enhancers from the VISTA database (Visel et al., 2007) across indicated cell types. Z-scores were calculated using all 111 cell types. The top 10 cell types in each validated enhancer group are shown. C) Average phyloP (Pollard et al., 2010) conservation scores of cCREs stratified by groups defined in A. Genomic background is indicated in gray. D) Two-dimensional density plot showing the median chromatin accessibility compared with the range (difference between maximum and minimum) of chromatin accessibility across 111 cell clusters for 890,130 cCREs, stratified by groups defined in A. E) Heatmap representation of 435,142 cCREs showing cell-type-restricted patterns in 111 cell types. Color represents $\log_2$-transformed chromatin accessibility. F,G) Heatmaps showing GO terms (F) and TF motifs (G) with maximal enrichment in cell-type-restricted cCREs of selected cell types. Only the most enriched TF motif in each of the previously identified motif archetypes (Vierstra et al., 2020) was selected as the representative and the top 10 motifs were selected for each cell type. Color represents $-\text{Log}_{10}\text{P}$. Full GO and motif enrichments are available on Mendeley Data: 10.17632/yv4fzv6cnm.1.


A
890,130 cCREs
13.93%
5.13%
80.94%
distal
proximal
promoter
-2000 -200 TSS +200 +2000
B
VISTA enhancers
Blood vessels
Forebrain
Heart
Negative
Average zscore of accessbility
A Cardiomyocyte
V Cardiomyocyte
Pericyte General 1
Schwann General
Sm Ms General
Sm Ms Uterine
Cardiac Fibroblast
Pericyte General 3
Sm Ms Vaginal
Type I Skeletal Myocyte
Sm Ms GE junction
Sm Ms Muscularis 1
Sm Ms Muscularis 2
Endothelial General 1
Endothelial General 2
Alveolar Cap Endo
Endothelial Myocardial
Lymphatic
BBB Endothelial
Cardiac Pericyte 3
Endocardial
Endothelial General 3
Endothelial Exocrine
GABA 1
GABA 2
Oligo Precursor
Astrocyte 1
Astrocyte 2
Enteric Neuron
Glutamatergic 1
Oligodendrocyte
C
Average phyloP score
Distance from region center
D
Promoters
N = 45,656
Promoter-proximal
N = 123,962
Distal
N = 720,512
Median(Log(Accessibility))
Δ(Log(Accessibility))
E
Stromal
Sm. Musc.
Myocyte
Mural
Adr. Cor.
Immune
Lymp.
Myelo.
Endothelial
Glia
Brain
Neur.
Islet
Other Epithelial
GI Epithelial
Log2(1+Acc.)
435,142 Cell type-restricted cCREs
111adult human cell types
F
GREAT Ontology
Hepatocyte
Acinar
Parietal
Chief
Foveolar
G Neuroendo
Ductal
digestion
lipid digestion
intestinal lipid catabolic proc.
protein esterification
cornified envelope assembly
positive reg. of dermatome dev.
renal vesicle induction
retina layer formation
neg. reg. of TGF-beta production
positive reg. of kidney dev.
exocyst assembly
epidermis morphogenesis
positive reg. of mesenchymal cell prolifer.
elastic fiber assembly
reg. of rhodopsin gene expression
cellular response to phosphate starvation
radial pattern formation
cholesterol biosynthetic proc.
prostate gland morphogenetic growth
otic vesicle formation
lacrimal gland development
transepithelial water transport
cellular hormone metabolic proc.
carboxylic acid metabolic proc.
steroid metabolic process
xenobiotic metabolic process
−Log10(P−val)
G
TF Motif Enrichment
HNF1A
HNF1B
HNF4A
HNF4G
ESRRA
ESRRB
GATA5
GATA2
GATA6
FOXC1
FOXD3
FOXM1
FOXJ2
FOXA2
FOXO6
NR4A2
ONECUT1
ONECUT2
CEBPA
NR2F2
PPARA-RXRA
ATF4
RARA
NR5A2
PTF1A
MESP1
SNAI2
FOS
FOXC2
ELF3
BATF
KLF4
LEF1
NFYA
NR3C2
HLTF
ESR2
NR6A1
RFX3
MSC
NEUROD1
RFX2
RFX5
POU3F3
POU4F1
PDX1
POU6F1
MEIS1
SOX9
TEAD3
−Log10(P−val)
Depleted
Enriched

These cCREs covered 58.9% of the elements in the registry of cCREs published by the ENCODE consortium (Moore et al., 2020), and also included 420,152 previously unannotated elements (Figure S10A). To benchmark these cCREs, we next compared chromatin accessibility profiles between biosamples profiled by bulk DNase-seq and cell types identified by sci-ATAC-seq in the current study. In aggregate, sci-ATAC-seq cell types resembled primary cell type biosamples more closely than bulk tissue or immortalized cell line biosamples (Figure S10B), and prevalent cell types with higher tissue abundance defined by sci-ATAC-seq showed closer similarities to bulk DNase-seq biosamples than rare cell types did (Figure S10C). Out of the 111 cell types profiled in the current study, 44 (40%) did not show statistically significant correlation with any bulk biosample profiled by the ENCODE consortium (Figure S10D). Many of these cell types were rare: their median maximal tissue abundance was only 3.2%, and 36 (81.8%) of them constituted fewer than 10% of all cells in any tissue. Taken together, these findings suggest that our dataset contributes previously underrepresented cCREs from in vivo human cell types to existing catalogues, particularly from cell types with low abundance in bulk tissues.

To assess the potential function of these cCREs, we next compared them with catalogs of transgenic reporter-validated mammalian enhancers (Visel et al., 2007) and found that validated tissue-specific enhancers exhibited much higher chromatin accessibility in cell types comprising a large proportion of nuclei identified in the corresponding tissue (Figure 7B). For example, validated enhancers in heart showed higher average chromatin accessibility in atrial cardiomyocytes (z-score: 1.41) and ventricular cardiomyocytes (z-score: 1.43) compared with other cell types (Figure 7B), suggesting a good correlation between cell type-specific chromatin accessibility and tissue-specific enhancer activity. We further found that eQTLs from 49 adult tissue types (Consortium, 2020) were most commonly accessible in prevalent cell types, such as

endothelial and smooth muscle cells. In addition, eQTLs from homogenous tissues, such as liver and thyroid, displayed strongest accessibility in the corresponding cell type which comprised a large proportion of nuclei identified in the tissue (Figure S11A-B). These results suggest that bulk tissue eQTLs best represent sequence variants associated with gene expression in abundant cell types and homogenous tissues, and may be less representative for rarer cell types within homogenous tissues or for unique cell types from heterogenous tissues.

We next categorized each cCRE based on distance to the nearest TSS as shown in Figure 7A. The majority (80.94%) of cCREs in the current catalogue resided more than 2,000 bp away from annotated TSSs. cCREs located directly over TSSs or near promoter regions displayed higher levels of sequence conservation and elevated chromatin accessibility (Figure 7C-D). By contrast, gene-distal cCREs were less accessible and showed larger variance relative to their accessibility (Figure 7D), suggesting the presence of shared programs of highly accessible promoter-proximal cCREs alongside variable programs of gene-distal cCREs across cell types and species. To further dissect cell-type specific chromatin signatures and regulatory programs, we applied an entropy-based strategy (Schug et al., 2005) to reveal 435,142 cCREs that demonstrated restricted accessibility in one or a few cell types (Figure 7E, see Methods). We next applied GREAT ontology enrichment analysis and motif enrichment analysis on cell-type restricted cCREs to reveal putative biological processes and TFs of each cell type, which largely correlated with expected cell type-specific functions (FDR < 0.01). For instance, cCREs restricted to hepatocytes yielded biological process ontology terms such as steroid metabolic process (Figure 7F), and were enriched for the binding sites of hepatocyte nuclear factor TF family members HNF1A/B, HNF4A/G, and ONECUT1/2 (Figure 7G) (Costa et al., 2003).

3.3.4 Integrative analysis of adult and fetal chromatin accessibility

To examine transcriptional regulators and cCRE remodeling between fetal and adult stages, we re-processed data from a recent cell atlas of chromatin accessibility in 15 human fetal tissue types (Domcke et al., 2020) using the same quality control, clustering, and annotation strategies described above, which lead to the discovery of 111 fetal cell types and 802,025 cCREs (Figure S12, Table S24). Combining these cCREs with those identified from the adult cell types, we mapped a total of 1,154,611 distinct cCREs spanning 14.8% of the human genome in 222 fetal and adult cell types (Mendeley Data: 10.17632/yv4fzv6cnm.1). These cCREs covered 58.5% and 69.7% of the elements in the EpiMap (Boix et al., 2021) and the ENCODE cCRE registry (Moore et al., 2020), respectively. In addition, 34.8% and 51.0% of our cCREs were not annotated by the EpiMap and the ENCODE cCRE registry, respectively.

To compare the 222 fetal and adult cell types across the two atlases of chromatin accessibility, we utilized SnapATAC followed by batch-correction to obtain a low dimensional representation of the 1,323,041 nuclei from both fetal and adult tissues (Figure 8AB, see Methods).

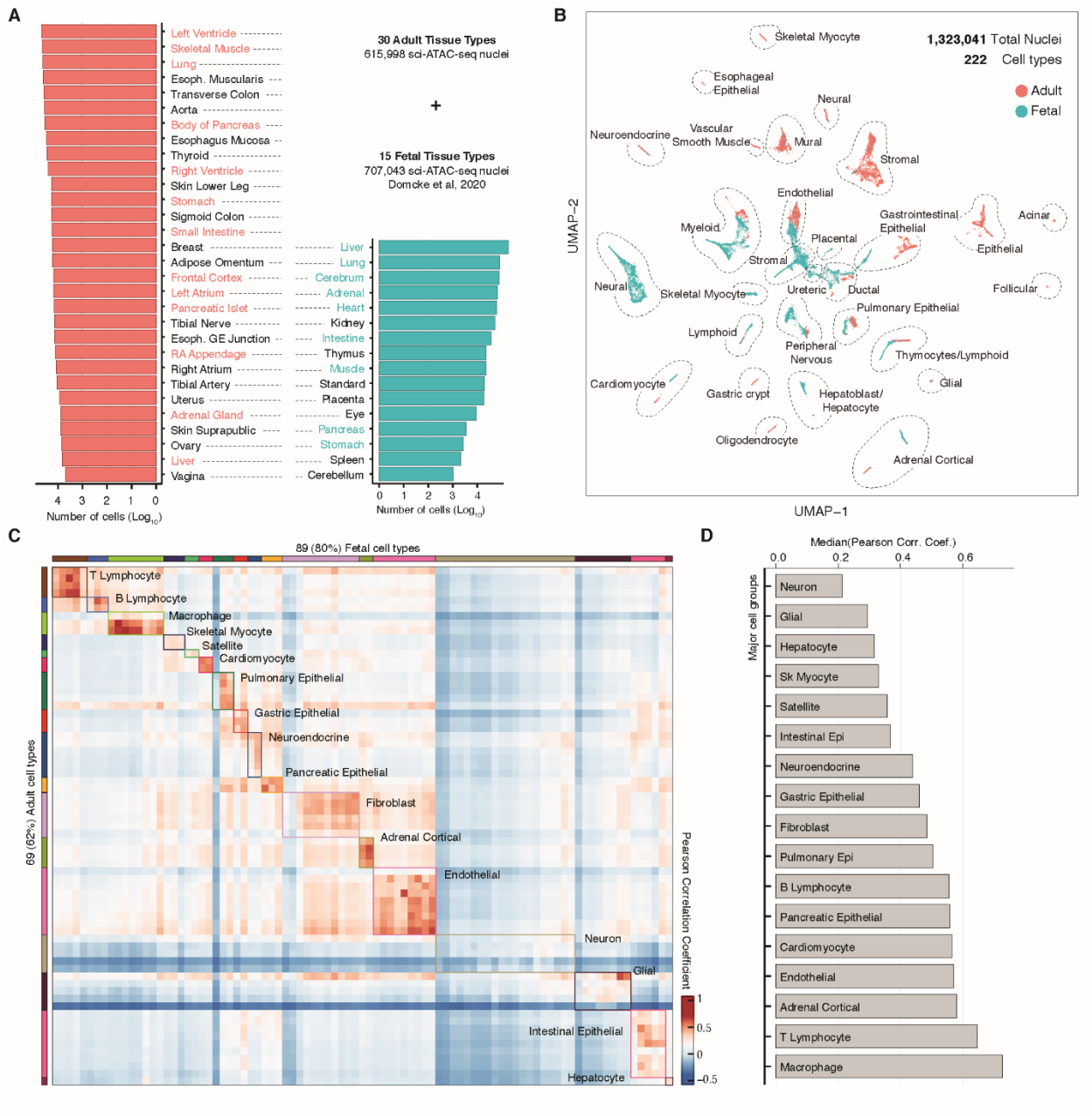


Figure 8: Integrative analysis of adult and fetal single-cell chromatin accessibility atlases. A) Number of sci-ATAC-seq cells per tissue type for 30 adult and 15 human fetal tissue types that were integrated. Matching tissue types between adult and fetal datasets are highlighted in red or blue respectively. Standard: sentinel tissue (trisomy 18 cerebrum). B) UMAP embedding of 1,323,041 nuclei from fetal and adult tissues. Each dot in the scatter plot represents a nucleus, colored by life stage. C) Heatmap showing Pearson correlation coefficient (PCC) between 69 adult cell types and 89 fetal cell types from 17 manually defined cell groups that are present in both adult and fetal tissues. A comprehensive heatmap is provided in Figure S14. D) Bar plot showing the median PCC for each major cell group indicated in C.

We next performed phylogenetic analysis to place the fetal and adult cell types into different groups based on the distance defined in the low dimensional space (see Methods, Figure S13A). In general, cell types belonging to different lineages separated into independent groups and harbored specific cCREs that were enriched for previously characterized lineage-specific TF motifs (Figure S13B). However, while many fetal cell types such as lymphoid, myeloid, and endothelial cells clustered near their adult counterparts in the tree, some cell types such as neurons and skeletal myocytes differed drastically between adult and fetal stages (Figure S13A), suggesting distinct cCRE usage by these cell types during development. To more systematically quantify differences in chromatin accessibility between adult and fetal cell types, we compared normalized accessibility across the list of 1,154,611 cCREs for each pair of fetal and adult cell types (Figure 8C-D, Figure S14). We found that fetal cell types such as immune and endothelial cells showed a relatively higher correlation with their adult counterparts than did other cell types such as neurons, glial cells, and skeletal myocytes (Figure 8D), consistent with the findings from our phylogenetic analysis. Together, these analyses suggest that the extent to which cCREs remodel to achieve developmental-stage specific functions varies greatly between human cell types.

To reveal the specific elements that may underlie fetal or adult-specific regulatory programs, we calculated life stage-specific cCREs for major cell groups which contained corresponding adult and fetal cell types (Figure 9A).

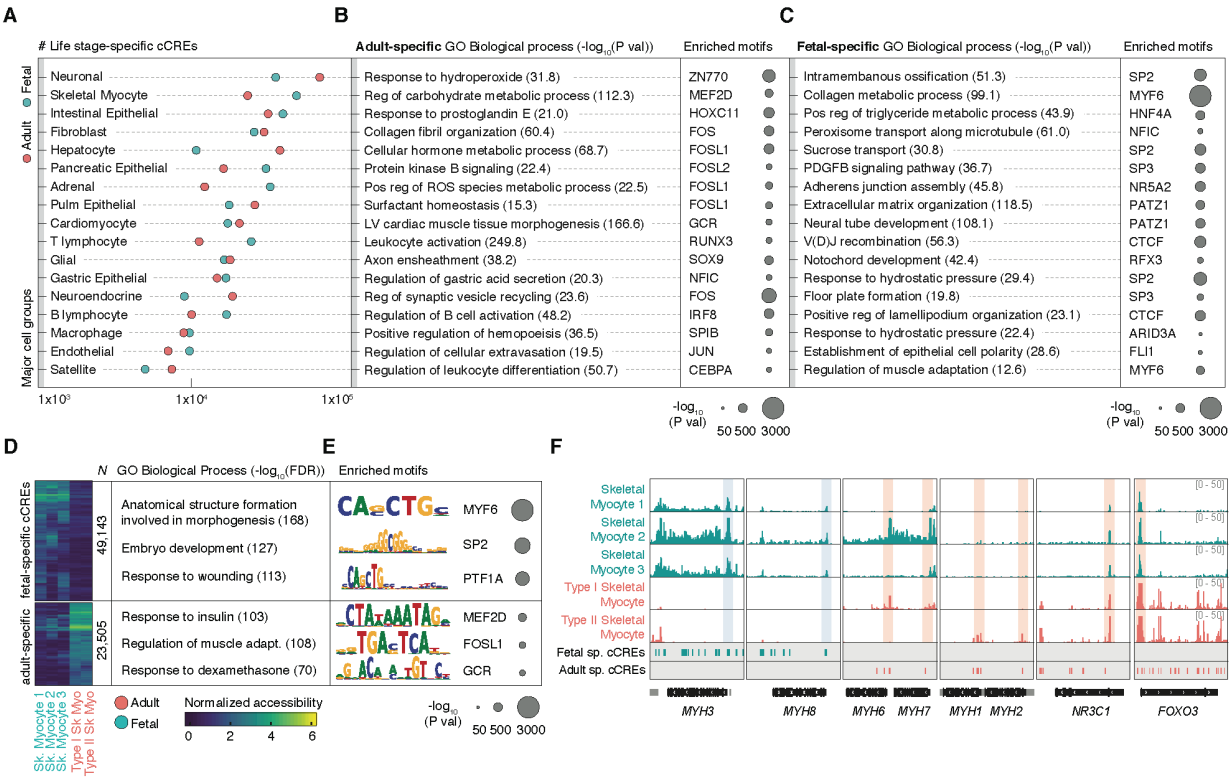


Figure 9: Differential chromatin accessibility landscapes in adult and fetal human cell types. A) Dot plot showing the number of adult and fetal specific cCREs detected for each major cell group indicated in C. B-C) Significant GO biological process ontology terms and transcription factor motif enrichments for adult-specific (B) and fetal-specific (C) cCREs. D) Heatmap representation of 72,648 differentially accessible (DA) cCREs between fetal and adult skeletal myocytes along with significant GREAT biological process ontology enrichments (McLean et al., 2010). Color represents log-transformed normalized signal. E) Significantly enriched TF motifs within fetal and adult skeletal myocyte DA cCREs. The most enriched motif within each motif archetype (Vierstra et al., 2020) was selected and the top three were displayed. F) Genome browser tracks showing chromatin accessibility for fetal and adult skeletal myocytes along with DA cCREs between the adult and fetal skeletal myocytes. Indicated genes are shown in black, other genes are shown in gray. TSSs of the indicated genes are shaded in red and blue.

Characterization of these elements revealed striking life stage-specific regulatory programs (Figure 9B-C). For example, skeletal myocytes differentiate substantially during pre and post-natal development (Chal and Pourquié, 2017) and showed lower global similarity between life stages than most other major cell types (Figure 8C-D). In total, we identified 72,648 differentially accessible (DA) cCREs between fetal and adult skeletal myocytes (Figure 9D). DA cCREs in fetal myocytes were associated with biological processes such as embryo development

and response to wounding, and were enriched for motifs of myogenic regulatory TFs (MRFs) which orchestrate normal myogenesis (Mary Elizabeth Pownall et al., 2002) (Figure 9E-F), highlighting the role of these elements in regulating myogenic properties of fetal myocytes. On the other hand, adult skeletal myocyte DA cCREs were associated with biological processes related to muscle adaptation to contractile activity as well as insulin and steroid hormone response, and were enriched for MEF family members (P = 1e-424) and AP-1 complex members including FOSL1 (P = 1e-274) (Figure 9D-E), suggesting a role for these elements in regulating transcriptional responses to hormonal exposures and load bearing in adult skeletal muscle. In line with these ontology results and with established patterns of myosin isoform expression across the human lifespan (Schiaffino and Reggiani, 2011; Schiaffino et al., 2015; Stuart et al., 2016), we discovered DA cCREs at loci encoding marker genes of pre-natal myocytes including *MYH3* and *MYH8*, the heavy chains of embryonic and neonatal myosin respectively, as well as markers of type I (slow) and type II (fast) twitch adult myocytes including *MYH6/MYH7* and *MYH1/MYH2* respectively (Figure 9F). Taken together, these findings reveal the regulatory elements that may underlie the proliferative capacity and mature functionality of fetal and adult skeletal myocytes, respectively, and emphasize the value of this dataset alongside emerging human cell atlases collected at different timepoints along the lifespan for determining life stage-specific gene regulatory programs at cell type resolution.

3.3.5 Delineation of cell-type specificity of human cCREs

To characterize the cell-type specificity of cCREs across fetal and adult cell types, we organized the 1,154,611 cCREs into 150 clusters, referred to as *cis*-regulatory modules (CRMs), based on their normalized accessibility across the 222 cell types. While several CRMs displayed shared accessibility patterns across all cell types, most CRMs were limited either to single fetal or

adult cell types or to groups of cell types that reflected shared cellular lineages (Figure 10A).

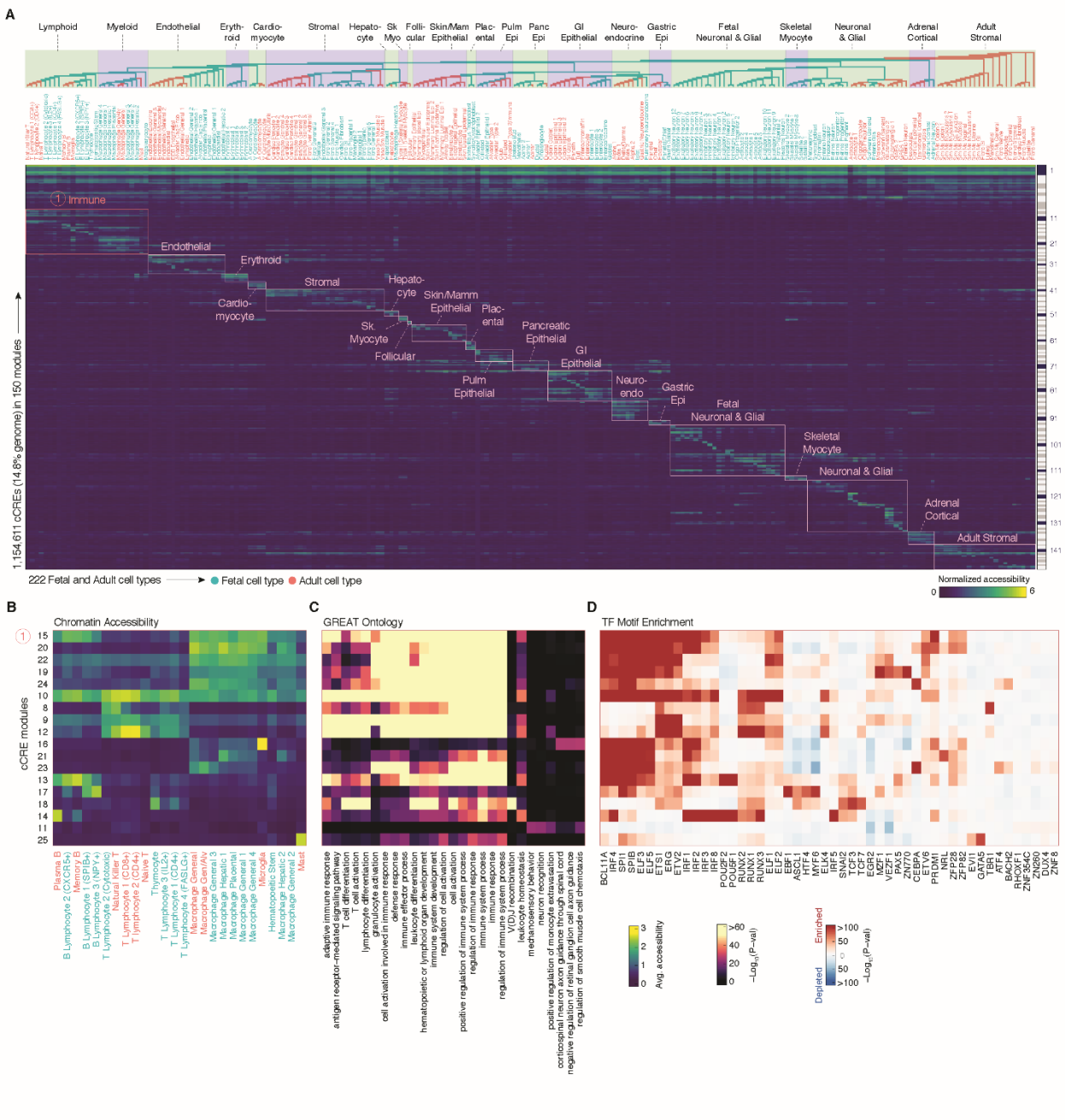


Figure 10: Delineation of CRE modules across 222 fetal and adult human cell types. A) Heatmap representation of chromatin accessibility for 1,154,611 cCREs across 222 fetal and adult cell types. Color represents normalized chromatin accessibility. cCREs were organized into 150 modules by K-means clustering, indicated by the color bars on the right. 20 groups of lineage-specific modules (colored boxes) are highlighted. B-D) Heatmaps showing chromatin accessibility (B), GO terms (C) and motifs (D) with maximal enrichment in a subset of CRE modules (rows) for immune cell types. The GO and motif heatmaps are colored by enrichment $-\log_{10}$P. Only the most enriched TF motif in each of the previously identified motif archetypes (Vierstra et al., 2020) was selected as the representative and the top 5 motifs were selected for each module. Full GO and motif enrichments are available on Mendeley Data: 10.17632/yv4fzv6cnm.1.

To annotate putative functions of CRMs, we applied GREAT ontology enrichment analysis (McLean et al., 2010). Broadly, CRMs showing preferential accessibility in specific fetal and adult cell types were enriched for biological process ontology terms related to both cell type and life stage-specific cellular processes (FDR < 0.01) (Figure 10B-C).

To identify sequence features underlying these CRMs, we next measured the enrichment of 1,565 human TF motifs across the 150 CRMs to reveal putative master regulators of fetal and adult human cell types. This analysis revealed a comprehensive catalogue of fetal and adult cell and lineage-specific TF motifs. For example, a module with strong accessibility in adult CD8+ T cells and natural killer T cells was distinguished by enrichment for TBR, EOMES, and TBX TF family motifs (Module 8, P < 1e-84; Figure 10B-D), modules with strong accessibility in B cells were distinguished by enrichment for EBF family TF motifs (Module 13, P = 1e-27; Module 17, P = 1e-197) and a module with strong accessibility in adult mast cells was distinguished by GATA family member motif enrichment (Module 25, P = 1e-84) (Figure 10B-D). Further, the module with the strongest accessibility across all identified cell types was characterized by enrichment of the SP1 motif (Module 1, P = 1e-9180), consistent with the original description of SP1 as a regulator of ubiquitously-expressed housekeeping genes (Black et al., 2001). In addition to these well-characterized associations, we also report previously undefined TF associations with human cell types that are challenging to study in their in vivo tissue contexts: for example, motifs of the ESRR (Module 92, P = 1e-357; Module 93, P = 0.1) and FOX (Module 92, P = 1e-36; Module 93, P = 1e-255) TF family were preferably enriched in modules accessible in fetal (Module 92) and adult (Module 93) gastric epithelial cells respectively (Figure 10A), and motifs of the FOS and JUN families were enriched in modules accessible in fetal and adult adrenal cortical cells (Modules 135-138, P < 1e-10; Figure 10A).

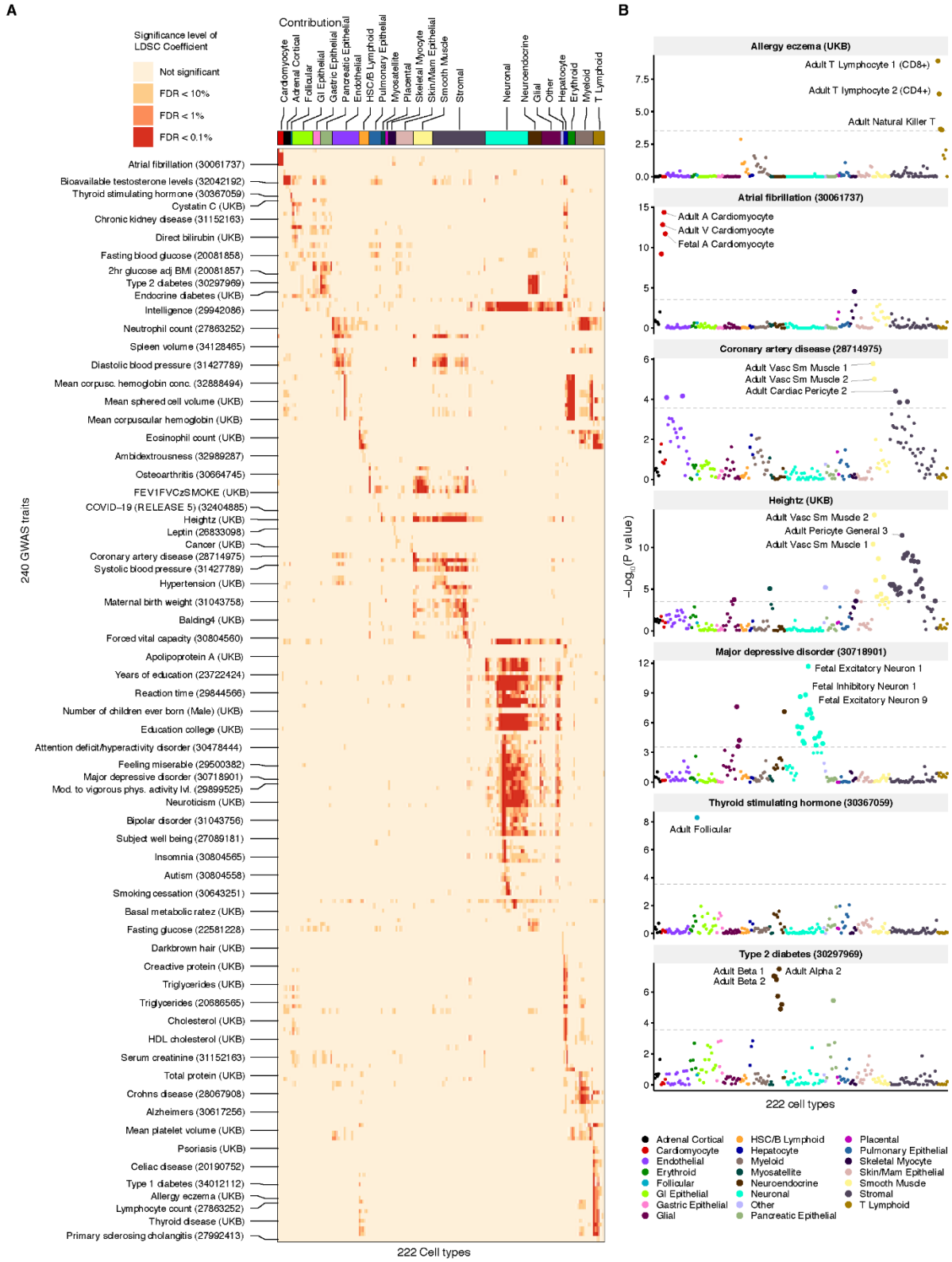
3.3.6 Association of human cell types with complex traits and diseases

We next sought to use our 1.2 million cell type-resolved cCREs to interpret genetic variants associated with complex traits and multigenic disease phenotypes. We downloaded the NHGRI-EBI GWAS catalogue (Buniello et al., 2019) and retained 1,123 well-powered GWAS with 10 or more significant SNPs and over 20,000 cases (14% of 8,219 GWAS publications). We then used a hypergeometric test to measure the enrichment of trait-associated variants within cCREs identified from the 222 fetal and adult cell types. GWAS variants of 450 traits/diseases were found to be enriched in cCREs from at least one cell type (FDR < 0.1%) (Figure S15). As a comparison, EpiMap, a comprehensive enhancer catalogue comprising 833 epigenomic maps from bulk human tissue samples, primary cells and *ex vivo* cell lines (Boix et al., 2021), captured 457 GWAS studies (FDR < 0.1%) (Figure S15). For the 290 traits shared by both this study and EpiMap, our data captured the strongest GWAS enrichment in 74.8% of cases (217 of 290) and provided improved resolution by linking complex traits to specific cell type(s) (Figure S15). Further, for 160 additional traits, we were able to identify enrichments that were not detected in previous analyses (Figure S15), highlighting the added value of cell type-resolved cCREs maps.

The GWAS enrichment analysis above considered only index variants, *i.e.*, SNPs in genome-wide significant loci. However, the index variants may not represent the specific causal variants due to linkage disequilibrium (Schaid et al., 2018) and much of the heritability lies in SNPs with associations that do not reach genome-wide significance (Yang et al., 2010). We thus curated 240 GWAS studies with publicly available summary statistics and examined the enrichment of their associated SNPs within cCREs annotated in fetal and adult cell types using stratified linkage disequilibrium score regression (LDSC), a method for identifying functional enrichment from GWAS summary statistics using genome-wide information from all SNPs and

explicitly modeling linkage disequilibrium (Finucane et al., 2015). This analysis revealed a total of 3,220 significant (FDR < 0.1) associations between fetal and adult cell types and human traits and disease phenotypes (Figure 11, Table S25).

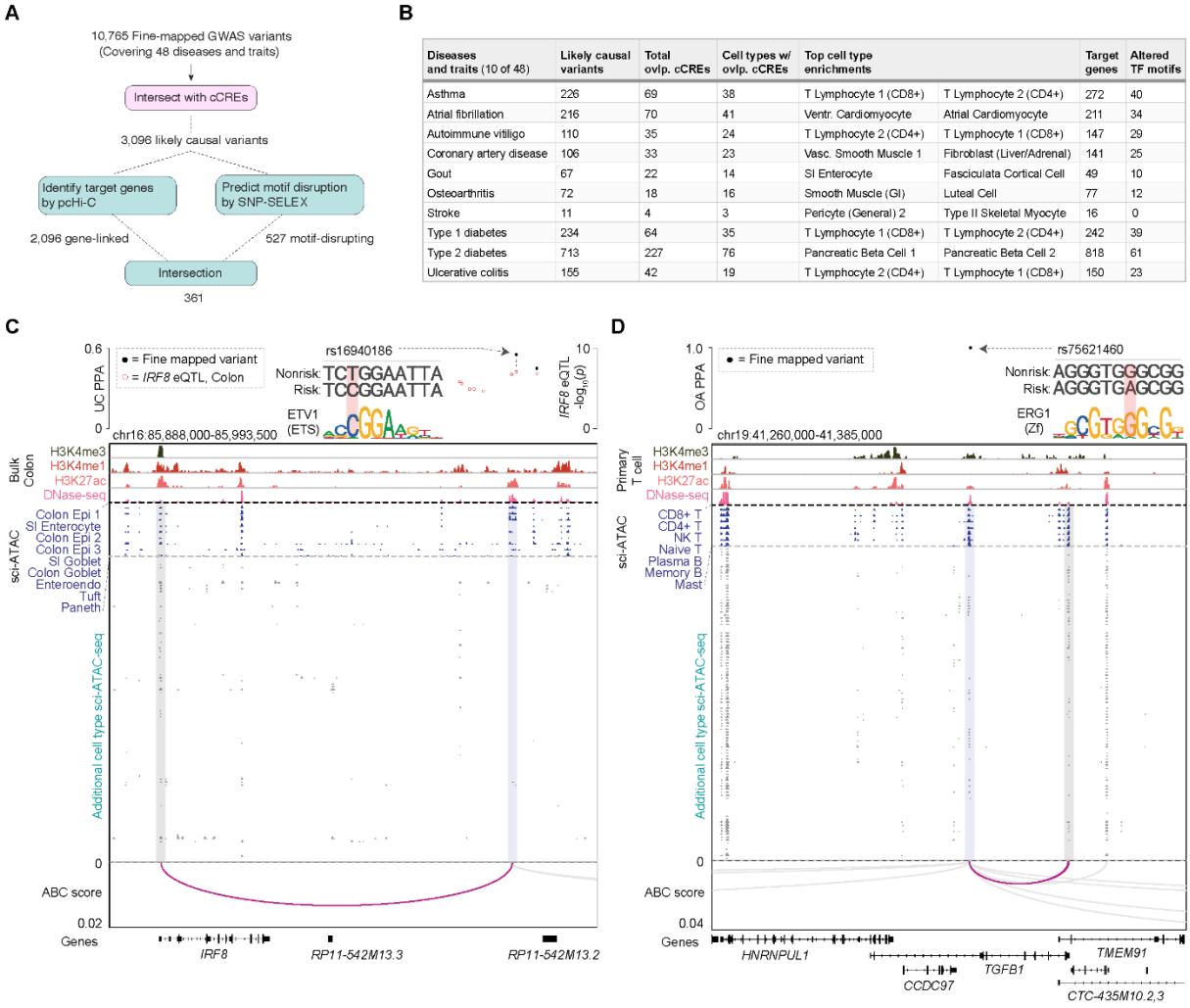
Figure 11: Association of fetal and adult human cell types with complex traits and diseases. A) Heatmap showing enrichment of risk variants associated with disease and non-disease traits from genome wide association studies in human cell type-resolved cCREs. Cell type-stratified linkage disequilibrium score regression (LDSC) analysis was performed using GWAS summary statistics for 240 phenotypes. Total cCREs identified independently from each fetal and adult cell type were used as input for analysis. P-values were corrected using the Benjamini Hochberg procedure for multiple tests. FDRs of LDSC coefficient are displayed. 66 selected traits were highlighted on the left, with PubMed identifiers (PMIDs) or "UKB", indicating summary statistics downloaded from the UK Biobank, enclosed in parentheses. Numerical results are reported in Table S25. B) Dot plots showing significance of enrichment for selected traits from panel A within cCREs from 222 fetal and adult cell types. Each circle represents a cell type. Large circles pass the cutoff of FDR < 1% at $-\log_{10}(P) = 3.55$. The top 3 most highly associated cell types are labeled for each trait. Comprehensive data are provided in Table S25.


A
B
Significance level of LDSC Coefficient
Not significant
FDR < 10%
FDR < 1%
FDR < 0.1%
Contribution
Cardiomyocyte
Adrenal Cortical
Follicular
GI Epithelial
Gastric Epithelial
Pancreatic Epithelial
Endothelial
HSC/B Lymphoid
Pulmonary Epithelial
Myosatellite
Placental
Skeletal Myocyte
Skin/Mam Epithelial
Smooth Muscle
Stromal
Neuronal
Neuroendocrine
Glial
Other
Hepatocyte
Erythroid
Myeloid
T Lymphoid
240 GWAS traits
Atrial fibrillation (30061737)
Bioavailable testosterone levels (32042192)
Thyroid stimulating hormone (30367059)
Cystatin C (UKB)
Chronic kidney disease (31152163)
Direct bilirubin (UKB)
Fasting blood glucose (20081858)
2hr glucose adj BMI (20081857)
Type 2 diabetes (30297969)
Endocrine diabetes (UKB)
Intelligence (29942086)
Neutrophil count (27863252)
Spleen volume (34128465)
Diastolic blood pressure (31427789)
Mean corpusc. hemoglobin conc. (32888494)
Mean sphered cell volume (UKB)
Mean corpuscular hemoglobin (UKB)
Eosinophil count (UKB)
Ambidextrousness (32989287)
Osteoarthritis (30664745)
FEV1FVCzSMOKE (UKB)
COVID-19 (RELEASE 5) (32404885)
Heightz (UKB)
Leptin (26833098)
Cancer (UKB)
Coronary artery disease (28714975)
Systolic blood pressure (31427789)
Hypertension (UKB)
Maternal birth weight (31043758)
Balding4 (UKB)
Forced vital capacity (30804560)
Apolipoprotein A (UKB)
Years of education (23722424)
Reaction time (29844566)
Number of children ever born (Male) (UKB)
Education college (UKB)
Attention deficit/hyperactivity disorder (30478444)
Feeling miserable (29500382)
Major depressive disorder (30718901)
Mod. to vigorous phys. activity lvl. (29899525)
Neuroticism (UKB)
Bipolar disorder (31043756)
Subject well being (27089181)
Insomnia (30804565)
Autism (30804558)
Smoking cessation (30643251)
Basal metabolic ratez (UKB)
Fasting glucose (22581228)
Darkbrown hair (UKB)
Creactive protein (UKB)
Triglycerides (UKB)
Triglycerides (20686565)
Cholesterol (UKB)
HDL cholesterol (UKB)
Serum creatinine (31152163)
Total protein (UKB)
Crohns disease (28067908)
Alzheimers (30617256)
Mean platelet volume (UKB)
Psoriasis (UKB)
Celiac disease (20190752)
Type 1 diabetes (34012112)
Allergy eczema (UKB)
Lymphocyte count (27863252)
Thyroid disease (UKB)
Primary sclerosing cholangitis (27992413)
222 Cell types
Allergy eczema (UKB)
Adult T Lymphocyte 1 (CD8+)
Adult T lymphocyte 2 (CD4+)
Adult Natural Killer T
Atrial fibrillation (30061737)
Adult A Cardiomyocyte
Adult V Cardiomyocyte
Fetal A Cardiomyocyte
Coronary artery disease (28714975)
Adult Vasc Sm Muscle 1
Adult Vasc Sm Muscle 2
Adult Cardiac Pericyte 2
Heightz (UKB)
Adult Vasc Sm Muscle 2
Adult Pericyte General 3
Adult Vasc Sm Muscle 1
Major depressive disorder (30718901)
Fetal Excitatory Neuron 1
Fetal Inhibitory Neuron 1
Fetal Excitatory Neuron 9
Thyroid stimulating hormone (30367059)
Adult Follicular
Type 2 diabetes (30297969)
Adult Beta 1
Adult Beta 2
Adult Alpha 2
−Log10(P value)
222 cell types
Adrenal Cortical
Cardiomyocyte
Endothelial
Erythroid
Follicular
GI Epithelial
Gastric Epithelial
Glial
HSC/B Lymphoid
Hepatocyte
Myeloid
Myosatellite
Neuroendocrine
Neuronal
Other
Pancreatic Epithelial
Placental
Pulmonary Epithelial
Skeletal Myocyte
Skin/Mam Epithelial
Smooth Muscle
Stromal
T Lymphoid

These enrichments revealed many expected cell type-disease phenotype relationships - for example, eczema risk variants were strongly enriched in adult T lymphocyte cCREs, atrial fibrillation risk variants were strongly enriched in both adult and fetal atrial and ventricular cardiomyocyte cCREs (FDR < 0.001), and thyroid stimulating hormone variants were enriched in follicular cell cCREs (Figure 11, Table S25). In addition to expected relationships, our analysis also revealed GWAS enrichment for human cell types not presently annotated by bulk DNase-seq or ATAC-seq data. These included a strong enrichment of coronary artery disease variants in adult vascular smooth muscle cCREs (FDR < 0.001) in addition to fetal and adult fibroblast, pericyte, and endothelial cell cCREs (FDR < 0.01), COPD variants in several adult stromal smooth muscle cell types (FDR < 0.01), triglyceride and HDL cholesterol level-associated variants in adult adipocyte cCREs (FDR < 0.01), and a nominal enrichment of ulcerative colitis variants in colon epithelial cell cCREs (P < 0.02). Interestingly, we detected substantial differences in the enrichment of disease and trait associated noncoding variants in subtypes of adult and fetal fibroblasts. These included a significant enrichment of variants associated with birth weight in fetal fibroblasts (FDR < 0.01) but not in adult fibroblasts (Table S25). Further, we detected differences in the enrichment of disease and trait variants in subtypes of adult fibroblasts, each of which displayed unique regulatory elements in addition to comparable chromatin accessibility at a set of core fibroblast cCREs (Figure S16). While all adult fibroblast populations were enriched for variants associated with standing height to an equivalent degree (FDR < 0.001), adult epithelial fibroblasts displayed a striking enrichment for variants associated with balding (FDR < 0.001) and only adult cardiac fibroblasts showed any enrichment for variants associated with myocardial fractal dimensions (FDR < 0.1; Table S25).

3.3.7 Systematic interpretation of molecular functions for noncoding risk variants

Many noncoding genetic variants enriched within cCREs from the analysis above are hypothesized to alter the expression of disease-associated genes by disrupting TF binding to *cis*-regulatory elements (Claussnitzer et al., 2020). To systematically interpret molecular mechanisms for the specific genetic variants associated with complex traits, we first applied the Activity-by-Contact (ABC) model (Fulco et al., 2019) to link the cCREs identified in 111 adult cell types to their target genes using our previously published promoter capture Hi-C data from 15 adult human tissues (Jung et al., 2019) (See Methods). This analysis revealed 5,723,307 unique distal cCRE-to-gene linkages across the 111 adult cell types, with a median of 726,514 total linkages and 6,804 cell type-specific linkages per cell type (Figure S17). Second, we determined the probability that variants from 48 GWAS were causal for disease or trait association (Posterior probability of association, PPA) using Bayesian fine-mapping (Wakefield, 2009). We defined likely causal variants as those with a PPA > 0.1, and found that they were more likely to reside within cCREs than variants with low PPA (Figure S17A). Overall, we detected 3,096 likely causal variants residing within cCREs mapped in 111 adult human cell types (Figure 12A-B, Table S26), 2,096 of which were linked to putative target genes via the ABC model (Figure 12A, Table S26).



Diseases and traits (10 of 48)	Likely causal variants	Total ovlp. cCREs	Cell types w/ ovlp. cCREs	Top cell type enrichments		Target genes	Altered TF motifs
Asthma	226	69	38	T Lymphocyte 1 (CD8+)	T Lymphocyte 2 (CD4+)	272	40
Atrial fibrillation	216	70	41	Ventr. Cardiomyocyte	Atrial Cardiomyocyte	211	34
Autoimmune vitiligo	110	35	24	T Lymphocyte 2 (CD4+)	T Lymphocyte 1 (CD8+)	147	29
Coronary artery disease	106	33	23	Vasc. Smooth Muscle 1	Fibroblast (Liver/Adrenal)	141	25
Gout	67	22	14	SI Enterocyte	Fasciculata Cortical Cell	49	10
Osteoarthritis	72	18	16	Smooth Muscle (GI)	Luteal Cell	77	12
Stroke	11	4	3	Pericyte (General) 2	Type II Skeletal Myocyte	16	0
Type 1 diabetes	234	64	35	T Lymphocyte 1 (CD8+)	T Lymphocyte 2 (CD4+)	242	39
Type 2 diabetes	713	227	76	Pancreatic Beta Cell 1	Pancreatic Beta Cell 2	818	61
Ulcerative colitis	155	42	19	T Lymphocyte 2 (CD4+)	T Lymphocyte 1 (CD8+)	150	23

Figure 12: Systematic interpretation of molecular functions of noncoding risk variants. A) Schematic illustrating the workflow for annotating fine-mapped noncoding risk variants. B) Table showing the number of likely causal variants (PPA > 0.1), number of cCREs overlapping likely causal variants, number of cell types in which overlapping cCREs are accessible, top cell types variants are enriched in based on LD score regression (Bulik-Sullivan et al., 2015), number of predicted target genes for likely causal variants, and significantly altered motifs predicted by deltaSVM model trained using SNP-SELEX data for 10 examples out of 48 total fine-mapped diseases and traits. Comprehensive data are provided in Table S26. C,D) Fine mapping and molecular characterization of an ulcerative colitis (UC) risk variant (C) in a gastrointestinal (GI) epithelial cell cCRE and an osteoarthritis variant (D) in an immune cell cCRE. Genome browser tracks (GRCh38) display ChIP-seq and DNase-seq from ENCODE human colon datasets (C) and primary T cell datasets (D) as well as chromatin accessibility profiles for cell types from sci-ATAC-seq. Chromatin interaction tracks show linkages between the variant-containing cCREs and genes from promoter capture Hi-C data via Activity-by-Contact (ABC) (Fulco et al., 2019) analysis. All linkages shown have an ABC score > 0.015. PPA: Posterior probability of association.

Third, we applied our recently developed deltaSVM models for 94 TFs (Yan et al., 2021) to identify variants potentially disrupting binding by these regulators. This analysis revealed 527 TF binding sites predicted to be significantly altered by the likely causal variants (Figure 12A, Table S26). The intersection of these lists prioritized 361 likely causal variants that 1) resided within a human cell type cCRE, 2) significantly altered TF binding 3) and were linked to one or more target genes (Figure 12A-B, Table S26).

For example, one likely causal variant for ulcerative colitis (rs16940186) resided within an intergenic cCRE restricted to epithelial cells of the gastrointestinal tract, particularly colon epithelial cells, enterocytes, and goblet cells (Figure 12C). The cCRE containing rs16940186 was predicted to contact the TSS of *IRF8* (ABC score > 0.015), which encodes a TF involved in the regulation of immune cell maturation (Salem et al., 2020) and regulation of innate immunity in gastric epithelial cells (Yan et al., 2016). The rs16940186 risk allele is an eQTL associated with increased *IRF8* expression in human colon tissue and, consistent with these findings, deltaSVM models predicted this risk allele to create a binding site for the ETS family of activating TFs (Figure 12C), which are expressed in intestinal epithelia and have been suggested to regulate intestinal epithelial maturation (Jedlicka et al., 2009). One other prioritized likely causal risk variant for osteoarthritis (rs75621460) resided within a cCRE that was primarily accessible in immune cell types, was predicted to target the immunosuppressive cytokine gene *TGFB1,* and disrupted a binding site for the zinc-finger TF ERG1 (Figure 12D).

3.4 Discussion

Detailed knowledge of the regulatory programs that govern gene expression in the human body has key implications for understanding human development and disease pathogenesis. Here, we used single-cell ATAC-seq to profile chromatin accessibility in 615,998 cells across 30 adult

human tissues representing a wide range of human organ systems and integrated this dataset with single-cell chromatin accessibility data from human fetal tissues (Domcke et al., 2020). We mapped the state of activity for approximately 1.2 million cCREs across 222 fetal and adult cell types, bridging the key gap of cell type resolution in the annotation of candidate regulatory elements in the human genome. This work highlights the value of integrating human sci-ATAC-seq datasets from multiple sources and timepoints (Chiou et al., 2021; Domcke et al., 2020; Hocker et al., 2021; Wang et al., 2020) and, in the future, integration of these data along with new human single-cell datasets of increasing scale, breadth, and depth will enable a comprehensive understanding of gene regulatory features of human cell types throughout the lifespan.

While genome-wide association studies (GWAS) have been broadly used to enhance our understanding of polygenic human traits and reveal clinically-relevant therapeutic targets for complex diseases, to date the discovery of new variants has far outpaced our ability to interpret their molecular functions (Claussnitzer et al., 2020). Two central goals of the current study were thus to link individual human cell types to complex traits and to leverage cCRE maps to interpret the molecular functions of specific noncoding risk variants. By applying our datasets alongside cutting-edge methods to prioritize likely causal variants in LD, link distal cCREs to putative target genes, and predict motifs altered by risk variants, we revealed thousands of cell type-trait associations and created a framework to systematically interpret noncoding risk variants. For example, we highlight the likely causal ulcerative colitis-associated variant rs16940186. This risk variant may function to increase *IRF8* expression in gastrointestinal epithelial cells by creating a binding site for ETS family TFs in a GI epithelial-specific enhancer, and thereby alter the transcriptional responses of intestinal epithelial cells to inflammatory cytokines. Pending functional validation experiments, our results suggest that targeting *IRF8* in GI epithelial cells

could be a potential therapeutic target for ulcerative colitis. As future GWAS in large cohorts with detailed phenotyping, whole genome sequencing efforts, and additional association studies employing long read technologies to capture structural variants become available, we anticipate that this combined resource and framework will be of continued utility for the interpretation of molecular functions for noncoding genetic variants. This resource thus lays the foundation for the analysis of gene regulatory programs across human organ systems at cell type resolution, and accelerates the interpretation of noncoding sequence variants associated with complex human diseases and phenotypes. The datasets can be accessed and explored at http://catlas.org/humanenhancer.

The current study is still limited in several ways: firstly, we solely integrated data from two discrete life stages and in an incomplete sampling of organ systems. While we utilized tissue from anatomic sites corresponding directly to existing biosamples in large-scale databases (Carithers et al., 2015; Stranger et al., 2017), the size and diversity of adult human organ systems make it difficult to representatively sample them in their entirety. Additionally, our assay solely profiles chromatin accessibility in dissociated nuclei, and thus misses key orthogonal molecular and spatial information. Future assays that incorporate gene expression, chromatin accessibility, histone modifications, DNA methylation, chromosomal conformation, TF binding, and spatial information in the same single-cell will greatly enhance our understanding of gene regulation in human cell types (Zhu et al., 2020).

3.5 Supplemental Data

3.5.1 Comparing our dataset with tissue-level single-cell chromatin atlases

Compared with published scATAC-seq data previously generated in lung (Wang et al., 2020) and heart (Hocker et al., 2021) tissues, our newly generated datasets from the same tissues are of equal quality, indicated by a comparable number of nuclei passing QC, median TSS enrichment and median number of fragments between datasets (Table 1, Figure S18).

Table 1: Comparison of major QC metrics for sci-ATAC-seq data from the current study with previously published sci-ATAC-seq data for matching tissue types. Each column represents one sci-ATAC-seq dataset. QC = Quality Control. Median TSSe = median transcription start site enrichment per nucleus.

	New data				Wang et al., 2020		
Tissue type	upper lobe left lung	upper lobe left lung	upper lobe left lung	upper lobe left lung	lung	lung	lung
Nuclei passing QC	4878	5281	5153	6788	5460	8087	9318
Median TSSe	9.43	11.92	8.81	13.4	10.72	10.52	13.55
Median fragments	2559.5	2192	2496	2705	2418	4063	3593

	New data		Hocker et al., 2021				
Tissue type	heart left ventricle	heart left ventricle	heart left ventricle	heart left ventricle	heart left ventricle	heart left ventricle	heart left ventricle
Nuclei passing QC	9664	8806	5781	6476	5067	7334	6197
Median TSSe	15.1	9.65	10.99	12.73	10.3	15.67	14.7
Median fragments	3375	3771.5	2205	3052	2008	5004	4034

Regarding recovery of cell types from prior sci-ATAC-seq studies, all nine cardiac cell types in (Hocker et al., 2021) were also found in our clustering analysis. Our analysis additionally revealed several other cardiac cell types which were not resolved by Hocker et al, such as pericytes (~5%) and mesothelial cells (~1%). Similarly, 17 out of 19 lung cell types in (Wang et al., 2020) were also found in our clustering analysis. The two cell types missing in our dataset that were annotated by Wang et al. included pulmonary neuroendocrine cells (PNECs) and erythrocytes. According to (Wang et al., 2020), ~95% of PNECs were present only in donors with age less than 3-years-old, and ~90% of erythrocytes were present only in fetal lung samples. Since the youngest

adult lung dataset we analyzed is from a 37 year-old donor, the age difference likely underlies the absence of PNECs and erythrocytes from (Wang et al., 2020) in our datasets.

In another study we compared sci-ATAC-seq data generated from the same tissue type via our combinatorial indexing-based protocol and via the 10X Genomics single-cell ATAC-seq platform (Li et al., 2021). We found that both detected clusters and cell type proportions were comparable between the two methods. Side-by-side comparison using two biological replicates of mouse primary motor cortex showed comparable nuclei passing quality control for both methods (combinatorial barcoding: 15,939 nuclei, 10x: 16,314, Li et al., Extended Data Figure 8a). Co-embedding of these datasets showed that the chromatin accessibility profiles and cell clusters from both platforms were in excellent agreement across cell types (Li et al., Extended Data Figure 8a-c). This was further shown by a confusion matrix comparing the similarity between clusters derived from the combinatorial barcoding and the 10x platform, respectively (Li et al., Extended Data Figure 8d). Further, we did not observe a significant difference in cell type composition between the two platforms (Li et al., Extended Data Figure 8e), except for one small population of vascular cells (VLMC, 326 nuclei from 10x, 155 nuclei from sci-ATAC-seq).

3.5.2 Comparing our datasets with single-cell RNA-seq datasets from corresponding tissues

To benchmark our cell type composition against high resolution scRNA-seq datasets, we compared cell types identified in lung and heart tissues with high-resolution scRNA-seq data sets reported in lung (Wang et al., 2020) and heart (Litvinukova et al., 2020) respectively. Overall, we found that the cell types identified, and even the cell type fractions, estimated using sci-ATAC-seq are in good agreement with those reported by high-resolution scRNA-seq despite technical variations in sample preparation and nuclei isolation, and biological differences such as donor characteristics. For example, in lung tissues, we annotated 23 out of 29 cell types reported by

scRNA-seq in (Wang et al., 2020) (Table 2). Cell types that were not annotated included three blood cell types: monocytes, dendritic cells and enucleated erythrocytes, two stromal cell types: myofibroblasts and chondrocytes, and one goblet cell type. The Pearson's correlation of estimated fractions for matching major cell types between two technologies was 0.8897.

Table 2: Cell type proportions for 23 lung cell types annotated by sci-ATAC-seq in the current study and scRNA-seq in (Wang et al., 2020).

Cell type	scRNA-seq	scATAC-seq
Macrophage	27.38%	40.74%
AT2 cells	19.91%	16.51%
AT1 cells	10.51%	8.84%
T cells	5.73%	11.23%
Fibroblast	6.49%	2.95%
Ciliated cells	2.31%	0.69%
Club cells	2.21%	2.46%
Endothelial cells	1.23%	0.93%
Cap	2.56%	3.28%
NK cells	1.76%	3.10%
B cells	0.84%	1.62%
Pericytes	0.68%	0.47%
Basal cells	0.58%	0.12%
Mast cells	0.25%	0.52%
Smooth muscle	0.43%	0.62%
Correlation:	0.8897	

In heart tissue, we uncovered all 11 major cell types reported in (Litvinukova et al., 2020). Again, the cell type fractions estimated using sci-ATAC-seq correlated well with those estimated using scRNA-seq (Table 3). The average Pearson's correlation across all 4 heart chambers was 0.832.

Table 3: Cell type proportions for 11 cardiac cell types annotated by sci-ATAC-seq in the current study and scRNA-seq in (Litvinukova et al., 2020).

Cell type	Heart-LV		Heart-RV		Heart-LA		Heart-RA	
	scRNA-seq	scATAC-seq	scRNA-seq	scATAC-seq	scRNA-seq	scATAC-seq	scRNA-seq	scATAC-seq
Atrial cardiomyocyte	0.00%	0.00%	0.00%	0.13%	34.40%	35.01%	25.30%	33.12%
Ventricular Cardiomyocyte	51.20%	38.76%	51.90%	43.28%	0.00%	0.70%	0.00%	0.38%
Fibroblast	13.60%	28.01%	15.40%	28.28%	19.60%	33.74%	29.30%	41.80%
Pericyte	18.90%	8.08%	17.10%	6.18%	17.60%	6.31%	7.80%	4.64%
Endothelial cell	8.30%	13.95%	6.20%	7.72%	9.20%	9.73%	15.40%	11.07%
Myeloid cell	3.60%	8.68%	3.60%	9.76%	4.80%	7.39%	6.30%	4.34%
Smooth muscle cell	2.20%	0.11%	2.20%	0.26%	4.30%	0.27%	4.10%	0.53%
lymphoid cells	1.40%	0.77%	2.20%	1.02%	4.00%	0.92%	5.90%	1.20%
Neural cells	0.60%	1.05%	0.60%	0.71%	2.00%	1.30%	2.50%	2.17%
Adipocytes	0.20%	0.00%	0.80%	1.79%	3.40%	1.59%	2.80%	0.19%
Mesothelial cells	0.00%	0.00%	0.00%	0.44%	0.70%	0.90%	0.50%	0.25%
Correlation:	0.769034906		0.829081895		0.789409534		0.942373768	

3.6 Materials and Methods

3.6.1 Data and code availability

Single-nucleus ATAC-seq datasets generated in this study have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the key resources table and Table S23. This paper analyzes existing, publicly available data. These accession numbers for these datasets are listed in the key resources table. Raw data from Figures 1, 2, 4, 6, S4 and S6 were deposited on Mendeley at 10.17632/yv4fzv6cnm.1. All original code has been deposited at Github and is publicly available as of the date of publication. Links are listed in the key resources table. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

3.6.2 Human Subjects

Adult human tissue samples were acquired by the ENTEx collaborative project (Stranger et al., 2017) via the GTEx collection pipeline (Carithers et al., 2015). Donor characteristics including age and sex are provided in Table S22. All human donors were deceased, and informed consent was obtained via next-of-kin consent for the collection and banking of deidentified tissue samples for scientific research. Donor eligibility requirements were as described previously (Carithers et al., 2015), and excluded individuals with metastatic cancer and individuals who had received chemotherapy for cancer within the prior two years.

3.6.3 Tissue feasibility testing for sci-ATAC-seq

Frozen tissue samples were sectioned on dry ice into two aliquots of equivalent mass. For nuclear isolation, one aliquot was subjected to manual pulverization via mortar and pestle while

submerged in liquid nitrogen, and the other aliquot was homogenized in a gentleMACS M-tube (Miltenyi) on a gentleMACS Octo Dissociator (Miltenyi) using the "Protein_01_01" protocol in MACS buffer (5 mM CaCl2, 2 mM EDTA, 1X protease inhibitor (Roche, 05-892-970-001), 300 mM MgAc, 10 mM Tris-HCL pH 8, 0.6 mM DTT) and pelleted with a swinging bucket centrifuge (500 x g, 5 min, 4°C; 5920R, Eppendorf). Pulverized frozen tissue and pelleted nuclei from gentleMACS M-tubes were each split into two further aliquots. One aliquot from each of the two nuclear isolation conditions was then resuspended in 1 mL Nuclear Permeabilization Buffer (1X PBS, 5% Bovine Serum Albumin, 0.2% IGEPAL CA-630 (Sigma), 1 mM DTT, 1X Protease inhibitor), and the other aliquot from the same nuclear isolation condition was resuspended in 1 mL OMNI Buffer (10mM Tris-HCL (pH 7.5), 10mM NaCl, 3mM MgCl2, 0.1% Tween-20 (Sigma), 0.1% IGEPAL-CA630 (Sigma) and 0.01% Digitonin (Promega) in water), yielding a total of four nuclear isolation/nuclear permeabilization buffer conditions tested for each tissue type. Nuclei were rotated at 4 ˚C for 5 minutes before being pelleted again with a swinging bucket centrifuge (500 x g, 5 min, 4°C; 5920R, Eppendorf). After centrifugation, permeabilized nuclei were resuspended in 500 µL high salt tagmentation buffer (36.3 mM Tris-acetate (pH = 7.8), 72.6 mM potassium-acetate, 11 mM Mg-acetate, 17.6% DMF) and counted using a hemocytometer. Concentration was adjusted to 2,000 nuclei/9 µl, and 2,000 nuclei were dispensed 12 wells of a 96-well plate per nuclear isolation/permeabilization condition (samples were processed in batches of 4 nuclear isolation/permeabilization conditions per 2 different tissue samples). For tagmentation, 1 µL barcoded Tn5 transposomes (Table S27) were added using a BenchSmart™ 96 (Mettler Toledo), mixed five times, and incubated for 60 min at 37 °C with shaking (500 rpm). To inhibit the Tn5 reaction, 10 µL of 40 mM EDTA (final 20mM) were added to each well with a BenchSmart™ 96 (Mettler Toledo) and the plate was incubated at 37 °C for 15 min with shaking

(500 rpm). Next, 20 µL of 2x sort buffer (2 % BSA, 2 mM EDTA in PBS) were added using a BenchSmart™ 96 (Mettler Toledo). All 12 wells from each nuclear isolation/permeabilization condition were combined into a separate FACS tube, and stained with Draq7 at 1:150 dilution (Cell Signaling). For each nuclear isolation/permeabilization condition, we used a SH800 (Sony) to sort four wells containing 0 nuclei per well and four wells containing 80 nuclei per well into one 96-well plate (total of 768 wells) containing 10.5 µL EB (25 pmol primer i7, 25 pmol primer i5, 200 ng BSA (Sigma)). After addition of 1 µL 0.2% SDS using a BenchSmart™ 96 (Mettler Toledo), the 96 well plate was incubated at 55 °C for 7 min with shaking (500 rpm). 1 µL 12.5% Triton-X was added to each well to quench the SDS. Next, 12.5 µL NEBNext High-Fidelity 2× PCR Master Mix (NEB) were added to each well and samples were PCR-amplified (72 °C 5 min, 98 °C 30 s, (98 °C 10 s, 63 °C 30 s, 72°C 60 s) × 12 cycles, held at 12 °C). After PCR, all wells were assayed for DNA library concentration using the PerfeCTa NGS Quantification RT-qPCR Kit (Quanta Biosciecnces) according to manufacturer's protocols, and subsequently returned to the thermal cycler for a second round of PCR amplification (72 °C 5 min, 98 °C 30 s, (98 °C 10 s, 63 °C 30 s, 72°C 60 s) × 4 cycles, held at 12 °C). After the second PCR amplification, for each nuclear isolation/permeabilization condition, wells containing 0 nuclei were combined and wells containing 80 nuclei were combined. The resulting DNA libraries were purified according to the MinElute PCR Purification Kit manual (Qiagen) and size selection was performed with SPRISelect reagent (Beckmann Coulter, 0.55x and 1.5x). Final libraries were quantified using a Qubit fluorimeter (Life technologies) and a nucleosomal pattern of fragment size distribution was verified using a Tapestation (High Sensitivity D1000, Agilent). We calculated a signal to noise ratio for final feasibility test libraries using LightCycler® 480 SYBR Green I Master Mix (Roche) along with custom primers for the promoter of human GAPDH (5'-

CATCTCAGTCGTTCCCAAAGT-3', 5'-TTCCCAGGACTGGACTGT-3') and a heterochromatic gene desert region (5'-CCCAAACTCTGA GAGGCTTATT-3', 5'-GAGCCATCATCTAGACACCTTC-3'). For each tissue type, the nuclear isolation/permeabilization condition that resulted in optimized nuclear yield (nuclei/mg tissue), library concentrations > 50 pM per 80 sorted nuclei, nucleosomal distribution pattern of fragments, and a log2(signal to noise ratio) > 3.3 was selected for combinatorial indexing-assisted single nucleus ATAC-seq (Table S22).

3.6.4 Combinatorial indexing-assisted single nucleus ATAC-seq

Combinatorial indexing-assisted single nucleus ATAC-seq was performed as described previously (Preissl et al., 2018) with slight modifications (Hocker et al., 2021) and using new sets of oligos for tagmentation and PCR (Table S27). Nuclei were isolated and permeabilized according to the optimized conditions from feasibility testing (Table S22). After resuspension in permeabilization buffer, nuclei were rotated at 4 ˚C for 5 minutes before being pelleted again with a swinging bucket centrifuge (500 x g, 5 min, 4°C; 5920R, Eppendorf). After centrifugation, permeabilized nuclei were resuspended in 500 µL high salt tagmentation buffer (36.3 mM Tris-acetate (pH = 7.8), 72.6 mM potassium-acetate, 11 mM Mg-acetate, 17.6% DMF) and counted using a hemocytometer. Concentration was adjusted to 2,000 nuclei/9 µl, and 2,000 nuclei were dispensed into each well of a 96-well plate per sample (96 tagmentation wells/sample, samples were processed in batches of 2-4 samples). For tagmentation, 1 µL barcoded Tn5 transposomes (Table S27) were added using a BenchSmart™ 96 (Mettler Toledo), mixed five times, and incubated for 60 min at 37 °C with shaking (500 rpm). To inhibit the Tn5 reaction, 10 µL of 40 mM EDTA (final 20mM) were added to each well with a BenchSmart™ 96 (Mettler Toledo) and

the plate was incubated at 37 °C for 15 min with shaking (500 rpm). Next, 20 µL of 2x sort buffer (2 % BSA, 2 mM EDTA in PBS) were added using a BenchSmart™ 96 (Mettler Toledo). All wells were combined into a separate FACS tube for each sample, and stained with Draq7 at 1:150 dilution (Cell Signaling). Using a SH800 (Sony), 20 nuclei per sample were sorted per well into eight 96-well plates (total of 768 wells) containing 10.5 µL EB (25 pmol primer i7, 25 pmol primer i5, 200 ng BSA (Sigma)). Preparation of sort plates and all downstream pipetting steps were performed on a Biomek i7 Automated Workstation (Beckman Coulter). After addition of 1 µL 0.2% SDS, samples were incubated at 55 °C for 7 min with shaking (500 rpm). 1 µL 12.5% Triton-X was added to each well to quench the SDS. Next, 12.5 µL NEBNext High-Fidelity 2× PCR Master Mix (NEB) were added and samples were PCR-amplified (72 °C 5 min, 98 °C 30 s, (98 °C 10 s, 63 °C 30 s, 72°C 60 s) × 12 cycles, held at 12 °C). After PCR, all wells were combined. Libraries were purified according to the MinElute PCR Purification Kit manual (Qiagen) using a vacuum manifold (QIAvac 24 plus, Qiagen) and size selection was performed with SPRISelect reagent (Beckmann Coulter, 0.55x and 1.5x). Libraries were purified one more time with SPRISelect reagent (Beckman Coulter, 1.5x). Libraries were quantified using a Qubit fluorimeter (Life technologies) and a nucleosomal pattern of fragment size distribution was verified using a Tapestation (High Sensitivity D1000, Agilent). Libraries were sequenced on a NextSeq500 or HiSeq4000 sequencer (Illumina) using custom sequencing primers with following read lengths: 50 + 10 + 12 + 50 (Read1 + Index1 + Index2 + Read2). Primer and index sequences are listed in Table S27.

3.6.5 Demultiplexing of single nucleus ATAC-seq sequencing reads

For each sequenced single nucleus ATAC-Seq library, we obtained four FASTQ files, two

for paired end DNA reads and two for the combinatorial indexes for i5 and T7 (768 and 364 indices, respectively). We selected all reads with up to 2 mismatches per i5 and T7 index (Hamming distance between each pair of indices is 4) and integrated the concatenated barcode at the beginning of the read name in the demultiplexed FASTQ files. The customized scripts can be found at: https://gitlab.com/Grouumf/ATACdemultiplex/.

3.6.6 Quality control metrics: TSS enrichment and unique fragments

TSS positions were obtained from the GENCODE database v31 (Frankish et al., 2019). Tn5 corrected insertions were aggregated ±2000 bp relative (TSS strand-corrected) to each unique TSS genome wide. Then this profile was normalized to the mean accessibility ± (1900 to 2000) bp from the TSS and smoothed every 11 bp. The max of the smoothed profile was taken as the TSS enrichment. We then filtered out all single-cells that had fewer than 1,000 unique fragments and/or a TSS enrichment of less than 7 for all data sets.

3.6.7 Overall clustering strategy

We utilized multiple rounds of clustering analysis to identify cell clusters. The first round of clustering analysis was performed on individual samples. We divided the genome into 5kb consecutive windows and then scored each cell for any insertions in these windows, generating a window by cell binary matrix for each sample. We filtered out those windows that are generally accessible in all cells for each sample using z-score threshold 1.65. Based on the filtered matrix, we then carried out dimension reduction followed by graph-based clustering to identify cell clusters. We called peaks for each cluster using the aggregated profile of accessibility and then merged the peaks from all clusters to generate a union peak list. Based on the peak list, we

generated a cell-by-peak count matrix and used Scrublet (Wolock et al., 2019) to remove potential doublets. Next, to carry out the second round of clustering analysis, we merged peaks called from all samples to form a reference peak list. We then generated a single binary cell-by-peak matrix using cells from all samples and again performed the dimension reduction followed by graph-based clustering to obtain the major cell groups across the entire dataset. To further dissect cell-type heterogeneity within the major cell groups, we then performed another round of clustering analysis for each of the identified major cell group to identify subclusters.

3.6.8 Doublet removal

We applied Scrublet to the cell-by-peak count matrix with default parameters. Doublet scores returned by Scrublet were then used to fit a two-component Gaussian mixture model using the "BayesianGaussianMixture" function from the python package "scikit-learn". The component with larger mean doublet score is presumably formed by doublets and cells belonging to it were removed from downstream analysis.

3.6.9 Dimension reduction

To find the low-dimensional manifold of the single-cell data, we adapted our previously published method, SnapATAC (Fang et al., 2021), to reduce the dimensionality of the peak by cell count matrix. The previous iteration of SnapATAC utilized spectral embedding for dimension reduction. To increase scalability of spectral embedding, we applied the Nyström method (Bouneffouf, 2016) for handling large datasets. Specifically, we first randomly sampled 35,000 cells as the training data. We then computed the Jaccard index between each pair of cells in the training set and constructed the similarity matrix . We computed the matrix , where is the diagonal

matrix such that . The eigendecomposition was performed on and the eigenvector with eigenvalue 1 was discarded. From the rest of the eigenvectors, we took the first 30 of them corresponding to the largest eigenvalues as the spectral embedding of the training data. We utilized the Nyström method to extend the embedding to the data outside the training set. Given a set of unseen samples, we computed the similarity matrix between the new samples and the training set. The embedding of the new samples is given by , where and are the eigenvectors and eigenvalues of obtained in the previous step.

3.6.10 Correction of Batch Effects

We performed batch correction for each tissue separately. Inspired by the mutual nearest neighbor batch-effect-correction method (Haghverdi et al., 2018), we developed a variant using mutual nearest centroids to iteratively correct for batch effects in multiple donor samples. Specifically, after dimension reduction we performed k-means clustering on individual replicate or donor sample with k equal to 20. We choose this number because the number of major clusters in a given tissue sample is typically less than 20. We then computed the centroid for each cluster and identified pairs of mutual nearest centroids across different batches. These mutual nearest centroids were used as the anchors to match the cells between different batches and correct for batch effects as described previously (Haghverdi et al., 2018). We found that the result can be further improved by performing above steps iteratively. However, too many iterations may lead to over-correction. We therefore used two iterations in this study.

3.6.11 Graph-based clustering algorithm

We constructed the k-nearest neighbor graph (k-NNG) using low-dimensional embedding

of the cells with k equal to 50. We then applied the Leiden algorithm (Traag et al., 2019) to find communities in the k-NNG corresponding to cell clusters. The Leiden algorithm can be configured to use different quality functions. The modularity model is a popular choice but it suffers from the issue of resolution-limit, particularly when the network is large (Traag et al., 2011). Therefore, we used the modularity model only in the first round of clustering analysis to identify initial clusters. In the final round of clustering, we chose the constant Potts model as the quality function since it is resolution-limit-free and is better suited for identifying rare populations in a large dataset (Traag et al., 2011). To determine the optimal number of clusters, we varied the resolution parameter in the Leiden algorithm and computed the clustering stability and average silhouette score under each resolution. Cluster stability was defined as the consistency, measured by the average adjusted rand index, of results from five independent clustering analyses on perturbed inputs. The perturbation was introduced in a way that 2% of the edges were randomly selected and subjected to removal. We selected the resolution that leads to both high average silhouette score and high clustering stability as well as biological considerations, e.g., number of known cell types in the tissue, marker gene accessibility.

3.6.12 Peak calling and peak filtering

For each cell cluster, initial peak calling was performed on Tn5-corrected single-base insertions (each end of the Tn5-corrected fragments) using the MACS2 (Zhang et al., 2008) callpeak command with parameters "–shift -100 –extsize 200 –nomodel –call-summits –nolambda –keep-dup all", filtered by the ENCODE hg38 blacklist (accession: ENCFF356LFX). Due to the varying abundance of cell types in each tissue, single-cell assays typically profile different cell types at different sequencing depths. To account for these differences, we adapted peak calling

cutoffs to different sequencing depths. Specifically, we choose a cutoff of FDR less than 0.1, 0.05, 0.025, 0.01, and 0.001, corresponding to the situations when the number of reads is with then range of 0-5 million, 5-25 million, 25-50 million, 50-100 million, and 100 million and above. Using simulated datasets, we found that this procedure achieved good balance between the sensitivity and specificity for detecting peaks under different sequencing depths. Next, based on the chromatin accessibility at the single cell level, we developed a peak filtering procedure to further reduce the false positive rate by retaining only those peaks that were accessible in a significant fraction of the cells compared to background regions. To do so, we first randomly selected 1 million regions from the genome and for each of these regions we calculated the fraction of cells that are accessible. These were used to fit a beta distribution as the null model. We then computed the fraction of accessible cells and its significance level for each candidate peak identified by MACS2. Candidate peaks with FDR < 0.01 were included in the final peak list.

3.6.13 Generating the union peak set

To compile a union peak set, we combined peaks from all clusters and extended the peak summits by 200 bp on either side. Overlapping peaks were then handled using an iterative removal procedure. First, the most significant peak, *i.e.*, the peak with the smallest p-value, was kept and any peak that directly overlapped with it was removed. Then, this process was iterated to the next most significant peak and so on until all peaks were either kept or removed due to direct overlap with a more significant peak.

3.6.14 Computing relative accessibility scores

We define an accessible locus as the minimal genomic region that can be bound and cut by the Tn5 enzyme. We use $L \subset N$ to represent the set of all accessible loci. We further define a pseudo-locus as the set of accessible loci that relates to each other in certain meaningful way (for example, nearby loci, loci from different alleles). In this example, pseudo-loci correspond to peaks. We use $\{ d_i \mid d_i \subset L \}$ to represent the set of all pseudo-loci. Let a_l be the accessibility of accessible locus l, where $l \in L$. We define the accessibility of pseudo-locus d_i as $A_i = \sum_{k \in d_i} a_k$, *i.e.*, the sum of accessibility of accessible loci associated with di. Let C_j be the library complexity (the number of distinct molecules in the library) of cell j. Assuming unbiased PCR amplification, then the probability of being sequenced for any fragment in the library is: $s_j = 1 - (1 - \frac{1}{C_j})k_j$, where k_j is the total number of reads for cell j. If we assume that the probability of a fragment present in the library is proportional to its accessibility and the complexity of the library, then we can deduce that the probability of a given locus l in cell j being sequenced is: $p_{lj} \propto a_l C_j s_j$. For any pseudo-locus d_i, the number of reads in d_i for cell j follows a Poisson binomial distribution, and its mean is $m_{ij} = \sum_{k \in d_i} p_{kj} \propto C_j s_j \sum_{k \in d_i} a_k = C_j s_j A_i$. Given a pseudo-locus (or peak) by cell count matrix O, we have: $\sum_j O_{ij} = \sum_j m_{ij}$. Therefore, $A_i = Z \frac{\sum_j O_{ij}}{\sum_j C_j s_j}$, where Z is a normalization constant. When comparing across different samples the relative accessibility may be desirable as they sum up to a constant, *i.e.*, $\sum_i A_i = 1 \times 10^6$. In this case, we can derive $A_i = \frac{\sum_j O_{ij}}{\sum_{ij} O_{ij}} * 10^6$.

3.6.15 Assigning cell types to cell clusters

To annotate the cell clusters, we first curated a set of marker genes from the PanglaoDB (Franzén et al., 2019) corresponding to expected cell types. We aggregated open chromatin fragments from each cluster and utilized the promoter accessibility, defined as RPM of +/- 1kb

around TSS, as the proxy for gene activity. We then computed the raw cell type enrichment score as the logarithm of the geometric mean of marker genes' activity. The final enrichment scores were obtained by applying two rounds of z-score transformation, first across cell types and then across cell clusters, on raw enrichment scores. For each cluster, we picked the cell type that showed strongest enrichment to make initial assignments. Finally, we manually reviewed these assignments and made adjustments based on focused consideration of marker gene accessibility in conjunction with information about tissue(s) of origin.

3.6.16 Identification of cell type-restricted peaks

We used a Shannon entropy-based method (Schug et al., 2005) to identify cell type-specific peaks. Given the relative accessibility scores of a peak across clusters, we first converted the scores to probabilities: $p_i = q_i / \sum_i q_i$. The entropy was then calculated by: $H_p = -\sum_t p_t \log_2(p_t)$. The specificity score is $Q_{p|t} = H_p - \log_2(p_t)$. To estimate the statistical significance of specificity scores, we assumed that under the null hypothesis each peak has an average accessibility level across all cell types and that the log base 2 of the cell-type-dependent fold changes from the average level follow a normal distribution with mean equal to zero and standard deviation s. The value of s was estimated using the top 50% least variable peaks, and 500,000 samples were then drawn to form the empirical distribution of Q_p that are used to determine the p-values of specificity scores. The cell-type-restricted peaks were then identified using a p-value cutoff of 0.025.

3.6.17 Cell-type enrichment analysis of fine-mapped GTEx eQTLs

The fine-mapped eQTLs (GTEx Analysis V8) in each of the 49 tissues or cell lines were downloaded from the GTEx portal (https://gtexportal.org). For each tissue, we first identified the

overlapping cCREs with its eQTLs. We then calculated the average of log-transformed accessibility scores of these peaks in each of the 111 cell types. This yielded a tissue by cell-type table containing raw cell-type enrichment scores of eQTLs from each tissue. The raw enrichment scores were then normalized row-wise using z-score transformation. For each tissue, we define the maximum cell-type enrichment as the largest value of z-scores across 111 cell types. In general, we found that homogenous tissues tend to have higher maximum cell-type enrichment than tissues that are more heterogenous.

3.6.18 Differential peak analysis

To carry out differential peak analysis between foreground set and background set, we first removed all peaks with fold changes of relative accessibility less than 2. For each peak, we then built a full model and a reduced model.

$$\log \frac{P_{full}}{1 - P_{full}} = \beta_0 + \beta_1 r + \beta_2 c$$

$$\log \frac{P_{reduced}}{1 - P_{reduced}} = \beta_0 + \beta_1 r$$

$P_{reduced}$ and P_{full} represent the likelihood of the reduced model and full model respectively. r contains the logarithm of number of fragments. c is categorical variable indicating if the cell comes from foreground or background. We then used a likelihood ratio test framework to determine whether the full model provided a significantly better fit of the data than the reduced model. We selected the sites using a 5% FDR threshold (Benjamini-Hochberg method).

3.6.19 Identification of fibroblast core signature and subtype-specific signatures

We first performed pairwise differential peak analysis for the seven fibroblast subtypes.

We then defined fibroblast core signature as peaks that are shared by all subtypes and were not called as differentially accessible in any of the pairwise comparison. Likewise, we defined the specific signature for a subtype as peaks that are differentially more accessible in the given subtype for every pairwise comparison.

3.6.20 Measuring the similarity of chromatin accessibility profiles between cell types identified by sci-ATAC-seq and bulk biosamples

We downloaded bulk DNase-seq data from the ENCODE portal. We excluded samples collected at embryonic stage or originated from kidney, bladder or brain tissues, as we did not perform experiments on those tissues. As a result, 638 datasets were kept for downstream analysis. For each of the DNase-seq datasets, we calculated its Pearson correlation coefficient with 111 identified cell types based on RPKM values at identified cCREs. These correlation scores were then scaled using z-score transformation across 111 cell types. We used the maximum of scaled correlation scores to represent each biosample's overall similarity with sci-ATAC-seq cell types.

3.6.21 Identification of cCRE modules

A cCRE module is defined as co-accessible regions that share similar accessibility pattern across cell types. To identify cCRE modules, we first performed quantile normalization on the log2 transformed matrix containing accessibilities of 1,154,611 cCREs in 222 fetal and adult cell types. For each cCRE, we then divided its accessibility vector by the L2 norm, which allowed us to better extract the accessibility pattern from the data. Next we applied the k-mean algorithm to this matrix to identify clusters of cCREs. Using the "elbow" method, we determined the number of clusters to be 150.

3.6.22 Motif enrichment analysis

We measured the enrichment of 1565 human TF motifs consisting of the JASPAR (2018) core non-redundant vertebrate motifs, the HOCOMOCO v1156 human motif set and the SELEX motifs by Jolma et al.. We computed the enrichments for each of the 1565 motifs relative to a joint cCRE background and filtered the list using FDR cutoff 0.01. For each motif. We reported the motif with the highest enrichment for each of the 286 previously identified motif archetypes (Vierstra et al., 2020).

3.6.23 Identification of candidate driver TFs

We used the Taiji pipeline (Zhang et al., 2019b) to identify candidate driver TFs in each cell cluster. Briefly, for each cell type cluster, we constructed the TF regulatory network by scanning TF motifs at the accessible chromatin regions and linking them to the nearest genes. The network is directed with edges from TFs to target genes. The genes' weights in the network were determined based on the relative accessibility of their promoters. The weights of the edges were calculated by the relative accessibility of the promoters of the source TFs. We then used the personalized PageRank algorithm to rank the TFs in the network.

3.6.24 Integration of adult and fetal datasets

To integrate our dataset with the recent cell atlas of fetal chromatin accessibility (Domcke et al., 2020), we downloaded the fragment files for 63 fetal samples spanning 15 tissues and converted the genomic coordinates from GRC37 (hg19) to GRCh38 using the UCSC liftOver tool. We then performed the quality control, cell filtering and cell clustering using the same pipeline

described above and identified 111 fetal cell types. Next, we combined the QC passed cells from adult and fetal datasets and performed the joint embedding using the SnapATAC algorithm. We considered fetal or adult cells as belonging to different batches, and used a linear model to remove technical batch effects for each dimension in the reduced dimensional space. Using these batch-corrected lower-dimensional representations, we applied the UMAP algorithm to visualize the cells in a 2D space and used the FASTME algorithm (Guindon and Gascuel, 2003) to construct the phylogenetic tree for adult and fetal cell types.

3.6.25 Differential peak analysis between fetal and adult cells

To perform differential peak analysis between fetal and adult samples, we modified the likelihood-ratio test framework described above to account for technical batch effects between two datasets. We started with three set of cells. The first two sets of cells corresponded to foreground and background sets that are subject to the differential test. The third set was the auxiliary set corresponding to remaining cells that were not from the first two sets. The auxiliary set served as a proxy to estimate the batch effects. For instance, when performing differential test between two sub-trees of the phylogenetic tree of fetal and adult cell types, for each sub-tree we randomly sampled an equal number of cells for each cell type in the sub-tree. The cells sampled from one branch were considered as foreground and those from the other were considered as background. The remaining cells did not belong to the two sub-trees form the auxiliary set. For each peak, we then built a full model and a reduced model.

$$\log \frac{P_{full}}{1 - P_{full}} = \beta_0 + \beta_1 r + \beta_2 s + \beta_3 t + \beta_4 c$$

$$\log \frac{P_{reduced}}{1 - P_{reduced}} = \beta_0 + \beta_1 r + \beta_2 s + \beta_3 t$$

$P_{reduced}$ and P_{full} represent the likelihood of the reduced model and full model respectively. r contains the logarithm of number of fragments. s is a categorical variable indicating whether the cell comes from the fetal tissue or the adult tissue. t indicates whether the cell comes from the auxiliary set. c indicates whether the cell comes from foreground set. We then used a likelihood ratio test framework to determine whether the full model provided a significantly better fit of the data than the reduced model. We selected the sites using a 1% FDR threshold (Benjamini-Hochberg method).

3.6.26 Generation of bigwig tracks

Each Tn5-corrected insertion was extended in both directions by 100 bp to form a 200-bp fragment. We then counted the number of fragments overlapping with each base on the genome and generated a bedgraph file. The bedgraph file was converted to bigwig file using the "bedGraphToBigWig" tool.

3.6.27 Linking cCREs to target genes

We downloaded the chromosome interactions called from published promoter capture Hi-C data in 14 human tissues (Jung et al., 2019). In each tissue, we first filtered the chromosome interactions using a lenient p-value cutoff of 0.1. We then created the chromosome interaction matrix using the normalized interaction frequency. The interaction matrices from 14 tissues were then averaged to get the final interaction matrix. We applied the Activity-by-Contact (ABC) Model (Fulco et al., 2019) to compute the ABC Score for each cCRE-gene pair as the product of Activity (chromatin accessibility) and Contact (interaction frequency), normalized by the product of Activity and Contact for all other cCREs. We retained all distal cCRE-gene connections with an

ABC score greater than 0.015.

3.6.28 Stratified linkage disequilibrium (LD) score regression

We used LD score regression (Bulik-Sullivan et al., 2015) v1.0.1 to estimate genome-wide GWAS enrichment for disease and non-disease phenotypes within cell type resolved cCREs (peaks called on each cell cluster via MACS2 (Zhang et al., 2008) using the above parameters). We compiled published GWAS summary statistics for complex diseases (Bentham et al., 2015; Bronson et al., 2016; Consortium, 2019; Cordell et al., 2015; Jansen et al., 2019; Ji et al., 2017; Jin et al., 2016; Luo et al., 2017b; Mahajan et al., 2018; Malik et al., 2018; Michailidou et al., 2017; Nielsen et al., 2018; Nikpay et al., 2015; Okada et al., 2014; Paternoster et al., 2015; Pividori et al., 2019; Sakornsakolpat et al., 2019; Schafmayer et al., 2019; Shadrina et al., 2019; Tachmazidou et al., 2019; Tin et al., 2019; Watanabe et al., 2019; Wiberg et al., 2019; Wuttke et al., 2019) and endophenotypes (Astle et al., 2016; Hoffmann et al., 2018; Kemp et al., 2017; Kilpeläinen et al., 2016; Manning et al., 2012; Saxena et al., 2010; Shrine et al., 2019; Strawbridge et al., 2011; Teumer et al., 2018; Warrington et al., 2019) within European populations. Using cell type resolved cCREs as a binary annotation, we created custom partitioned LD score files by following the steps outlined in the LD score estimation tutorial. As background annotations, we included all baseline annotations in the baseline-LD model v1.2 as well as partitioned LD scores created from all merged cCREs. For each trait, we used LD score regression to then estimate coefficient p-value for each cell type relative to the background annotations and used the Benjamini-Hochberg procedure to correct for multiple tests.

3.6.29 GWAS enrichment analysis

We downloaded the NHGRI-EBI GWAS catalogue (Buniello et al., 2019) (downloaded from https:// www.ebi.ac.uk/gwas/docs/file-downloads on July 7, 2021) and pruned the catalogue using an approach described previously (Boix et al., 2021). Specifically, for each trait and PMID combination, we ranked associations by their significance (P value) and added SNPs iteratively if they were not within 5 kb of previously added SNPs. We then compiled a compendium of 1,123 well-powered GWAS with 10 or more significant SNPs and over 20,000 cases (14% of 8,219 GWAS publications) that capture over 81,057 GWAS loci.

For each cell type and trait combination, we computed the number of intersections between trait associated SNPs and cell-type associated cCREs. We compared this number with the number of intersections between SNPs and the entire set of cCREs from all cell types, using a hypergeometric test to evaluate the statistical significance of enrichments. To estimate the false discovery rate, we generated 1,000 null GWAS with the same lead SNP set size by randomly shuffling the trait associations across GWAS locations. We then computed the null association P values for each permuted GWAS and used the 0.1% top quantiles as the cut-off.

3.6.30 Fine mapping

We performed genetic fine mapping for GWAS of diseases and endophenotypes that had sufficient coverage (i.e., were at least imputed into 1000 Genomes). For GWAS with available fine mapping data, we took 99% credible sets directly from the supplemental tables. For GWAS without available fine mapping data, we calculated approximate Bayes factors (Wakefield, 2009) (ABF) for each variant assuming prior variance $\omega = 0.04$. For every trait, we obtained index variants for each locus from the supplemental tables of the respective study. We extracted all variants in at least low linkage disequilibrium ($r2 > 0.1$ using the European subset of 1000

Genomes Phase 3 (Auton et al., 2015b)) in a large window (±2.5 Mb) around each index variant. We calculated posterior probabilities of association (PPA) for each variant by dividing its ABF by the cumulative ABF for all variants within the locus. We then defined 99% credible sets for each locus by sorting variants by descending PPA and keeping variants adding up to a cumulative PPA of 0.99.

3.6.31 Predicting the effects of noncoding variants on TF binding

To identify SNPs that affect TF binding, we employed deltaSVM models as described previously (Yan et al., 2021). Briefly, 40 bp sequences centered on each SNP were used as input to 94 previously trained and validated TF models. For each SNP, we predicted the binding scores for both alleles by running "gkmpredict". A SNP was considered to be bound if the binding score passed the pre-defined threshold for either allele. Among those SNPs, deltaSVM scores were calculated using the "deltasvm.pl" script and SNPs with deltaSVM scores passing the threshold for the corresponding model are predicted to affect TF binding.

3.6.32 External genome browser track data

Genome browser tracks displaying ChIP-seq and DNase-seq signal from bulk transverse colon datasets and human primary T cell datasets were downloaded from ENCODE with the following identifiers: ENCSR340MRJ, ENCSR557OWY, ENCSR500QVK, ENCSR792VLP, ENCSR627UDJ, ENCSR902BOX, ENCSR218OEZ, ENCSR222QLW.

3.6.33 Quantification and statistical analysis

Statistical parameters were reported either in individual figures or corresponding figure

legends. Statistical details of experiments can be found in “METHOD DETAILS”. All statistical analyses were performed in either R or Python.

3.6.34 Additional Resources

The raw data and analyzed results are available at our interactive web portal: http://catlas.org/humanenhancer.

3.7 Supplemental Methods

3.7.1 Cell clustering

To fully dissect the heterogeneity within each tissue, we utilized an iterative clustering strategy. The first round of clustering analysis was performed on individual samples to identify initial clusters and candidate peaks. Using the peaks called in the first round of clustering analysis, we then generated a single binary cell-by-peak matrix using cells from all samples and again performed the dimension reduction followed by graph-based clustering to obtain the major cell groups across the entire dataset (Figure S19). We next performed sub-clustering analysis for each of the identified major cell group to identify subclusters (Figure S20).

The resolution parameter in the leiden algorithm was chosen according to the joint consideration of cluster separation (measured by the average silhouette width) and the stability of clustering results. Silhouette width is designed to assess the quality of clusters with a convex shape. Although single-cell clusters are generally non-convex, the spectral clustering technique employed by SnapATAC is a non-linear dimensionality reduction method and is able to transform the clusters into a convex shape suitable for applying silhouette width to measure cluster separation. Indeed, using benchmarking datasets we showed that the silhouette width is extremely useful for

selecting appropriate clustering parameters: silhouette width was highly correlated with the adjusted rand index (ARI), which measures the consistency between the clustering result and the ground truth. We also noticed that parameters that produced the optimal number of clusters were generally associated with high clustering stability. We therefore used a criterion of stability greater than 0.85 to filter clustering parameters. The parameters selected according to the criterion above were further tuned with biological considerations (Table 4). In total, we identified 30 highly stable major cell groups, and then subclustered each to arrive at a total of 111 adult cell types.

Table 4: Resolution parameters used in the sub-clustering analysis. Major cell groups that do not have subclusters were omitted from the table, including C10, C11, C14, C24, C26, C28, C29, and C30.

Major cluster ID	Resolution with highest silhouette (# of clusters)	Selected Resolution (# of clusters)	Comments
C1	0.005 (5)	0.01 (9)	The highest and second highest silhouette scores were very close, and UMAP embedding inidcated that the second highest score yielded more clusters. The resolution with the second highest silhouette was chosen.
C2	0.003 (2)	0.03 (3)	Original resolution missed Alveolar Macrophage population.
C3	0.04 (2)	0.04 (2)	
C4	0.0075 (2)	0.05 (9)	UMAP embedding inidcated that the resolution with the second highest silhouette was revealed more clusters.
C5	0.01 (2)	0.06 (11)	UMAP embedding inidcated that resolutions with lower silhouette revealed more clusters, so the resolution was adjusted to reveal these clusters.
C6	0.04 (1)	0.05 (2)	Original resolution failed to separate the type 1 and type 2 skeletal myocytes.
C7	0.03 (2)	0.03 (2)	
C8	0.01 (2)	0.01 (2)	
C9	0.07 (9)	0.07 (9)	
C12	0.04 (2)	0.06 (4)	Original resolution failed to reveal T cell subtypes. The resolution with the second highest sihouette was selected.
C13	0.32 (9)	0.32 (9)	
C14	N/A	N/A	
C15	0.01 (2)	0.32 (6)	Original resolution failed to identify Delta/Gamma cells.
C16	0.07 (2)	0.23 (5)	Original resolution failed to identify Alveolar Type 1 cells.
C17	0.003 (2)	0.16 (7)	UMAP embedding indicated that more clusters were present than revealed be highest silhouette value are. The resolution with the second highest silhouette was chosen.
C18	1 (8)	0.07 (3)	The original resolution appeared to lead to overclustering as some of the resulting clusters shared highly similar marker genes.
C19	0.05 (2)	0.05 (2)	
C20	0.25 (5)	0.25 (5)	
C21	0.1 (2)	0.1 (2)	
C22	0.64 (5)	0.16 (4)	The original resolution appeared to lead to overclustering as some of the resulting clusters shared highly similar marker genes.
C23	0.1 (2)	0.1 (2)	
C25	0.25 (2)	0.25 (2)	
C27	0.5 (3)	0.5 (3)	

3.7.2 Annotation of cell clusters

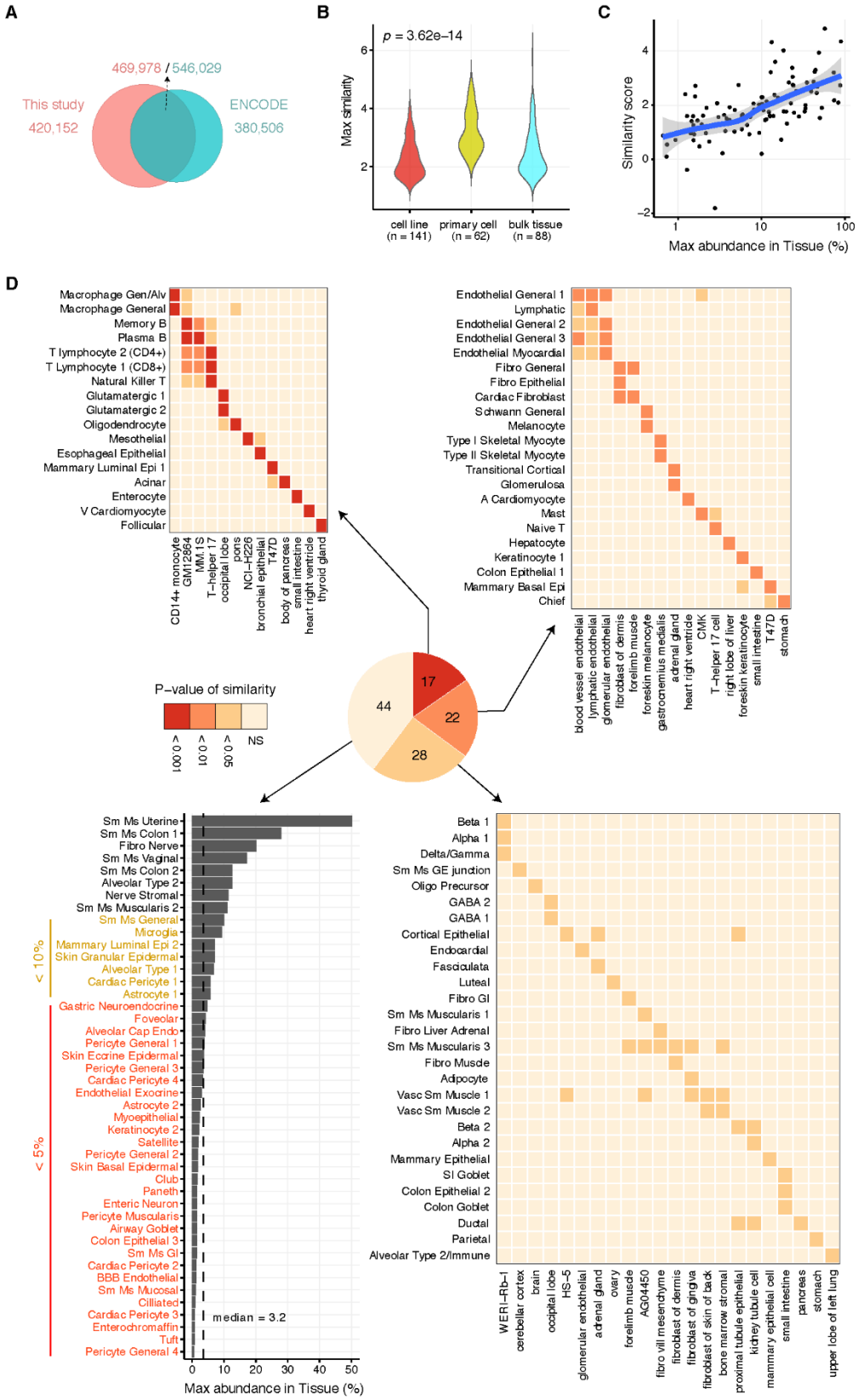
To annotate the cell clusters, we first curated a set of marker genes from the PanglaoDB (Franzén et al., 20x19) corresponding to expected cell types. We aggregated open chromatin fragments from each cluster and utilized the promoter accessibility, defined as RPM of +/- 1kb around TSS, as the proxy for gene activity. We then computed the raw cell type enrichment score as the logarithm of the geometric mean of marker genes' activity. The final enrichment scores were obtained by applying two rounds of z-score transformation, first across cell types and then across cell clusters, on raw enrichment scores (Figure S21). For each cluster, we picked the cell type that showed strongest enrichment to make initial assignments. Finally, we manually reviewed these assignments and made adjustments based on focused consideration of marker gene accessibility in conjunction with information about tissue(s) of origin. For example, the figures below show the marker gene accessibility for neuroendocrine cell types and non-neuroendocrine pancreatic cell types as controls at the genes encoding *GCG*, *INS-IGF2*, *SST*, and *GHRL* (Figure S22).

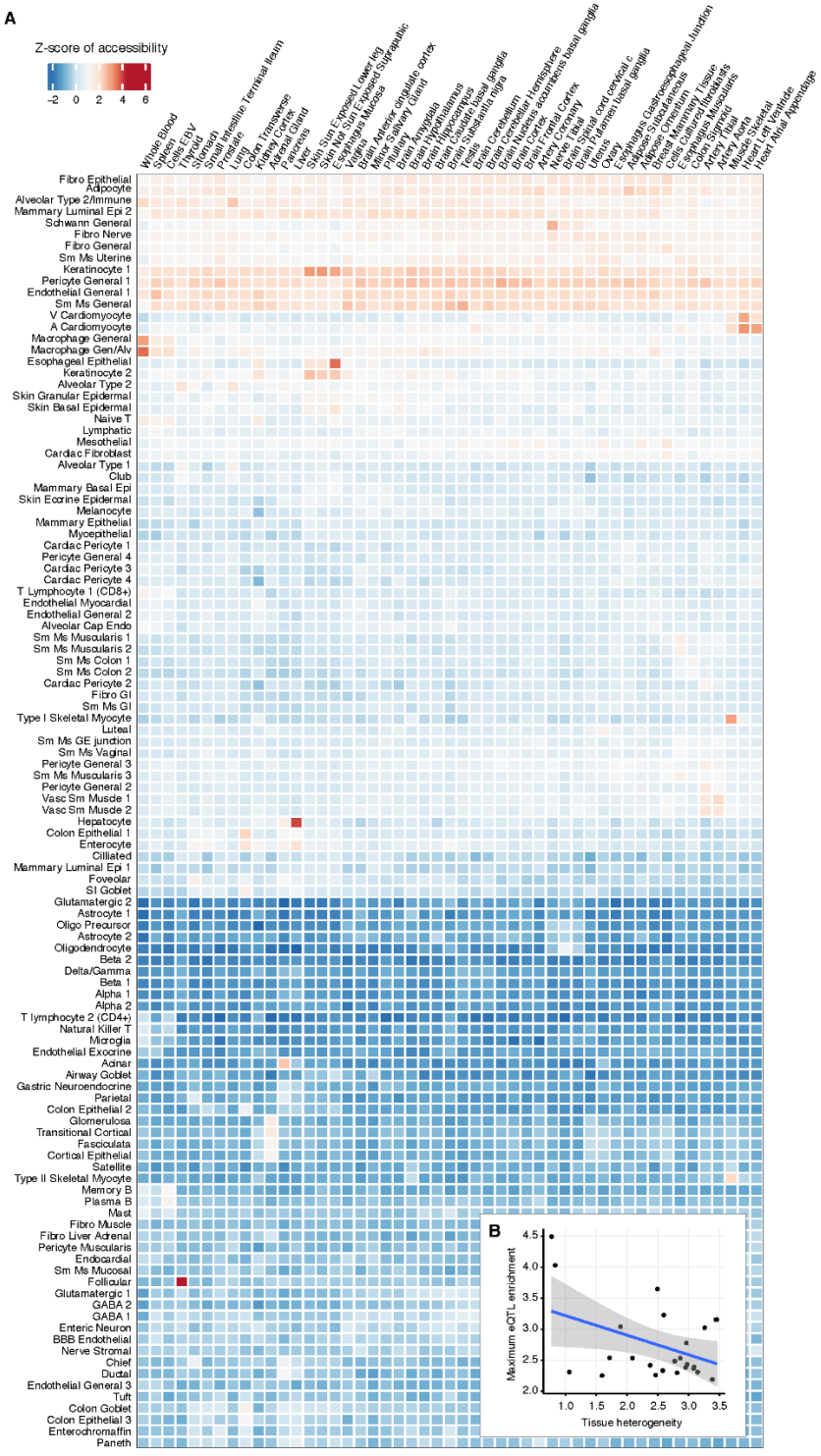
3.7.3 Optimizing peak calling procedure

Single-cell assays typically profile different cell types at different sequencing depths due to the varying abundance of cell types in each tissue. Using simulated datasets, we found that adapting peak calling cutoffs to each cell type's sequencing depth increased the sensitivity to detect legitimate peaks in rare cell types and decreased false discovery rate for calling peaks in cell types with high relative abundance. To further improve our overall confidence in identified peaks, we next adopted a peak-filtering protocol (Li et al., 2021) by removing peaks that were not significantly more accessible than background at the single-cell level.

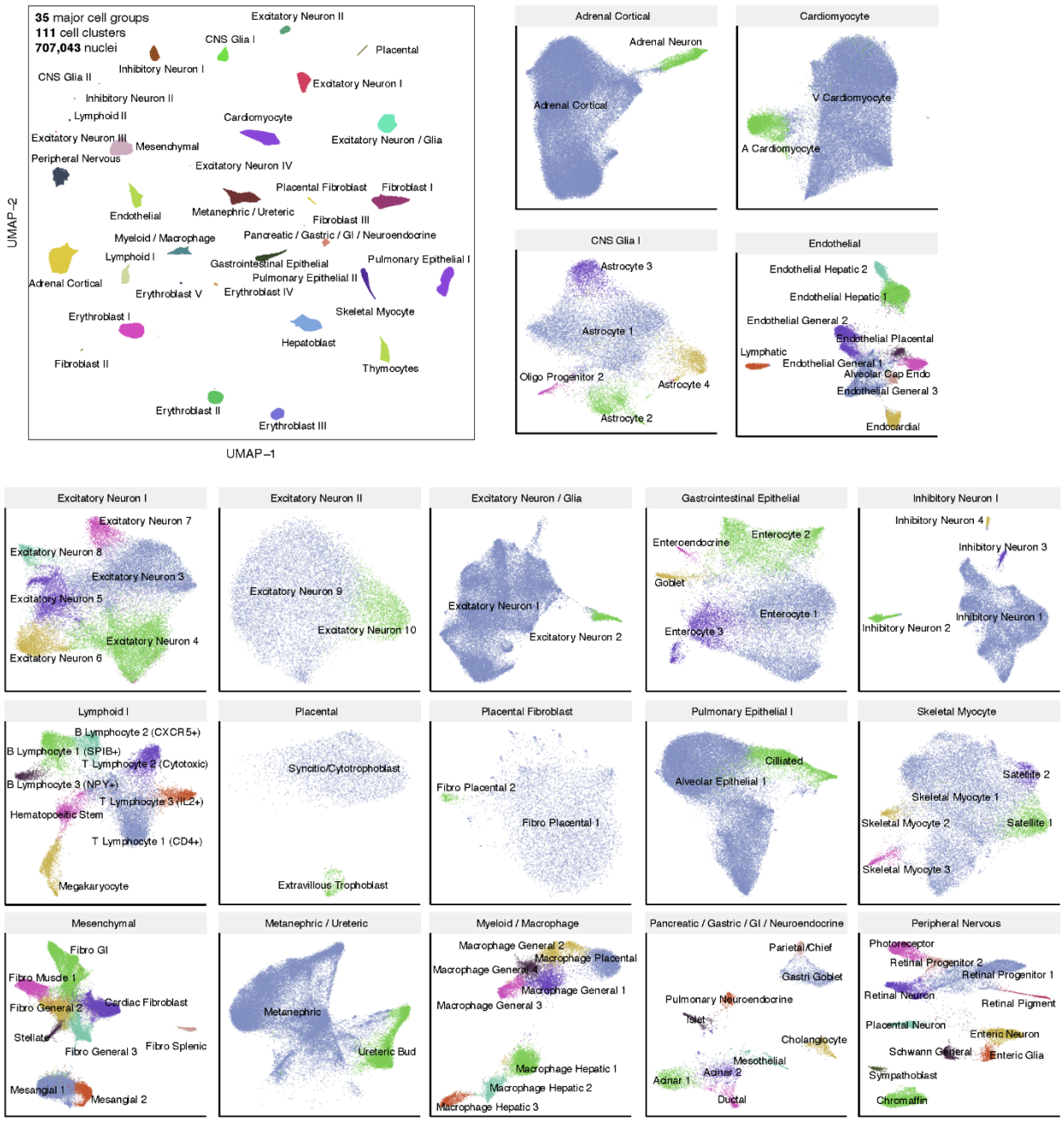
3.8 Supplemental Figures

Supplemental Figure 10: Comparison of open chromatin landscapes in adult human cell types from the current study with previous DNase-seq data from bulk biosamples. A) Venn diagram showing the intersection between cCREs identified in the current study with the registry of cCREs created by the ENCODE consortium. B) Distribution of similarity scores for 291 bulk DNase-seq samples stratified by sample classification. Similarity score is defined as the maximum of the standardized correlation scores of a bulk DNase-seq sample with 111 adult human cell types from sci-ATAC-seq. C) Scatter plot showing the similarity score of each cell type as function of its maximal abundance in tissues. Since we did not profile spleen tissue or peripheral blood mononuclear cells, the maximal abundances of immune cell types cannot be accurately estimated and immune cell types were thus excluded from this analysis. D) Pie chart showing similarity between cell types from sci-ATAC-seq and bulk DNase-seq seq samples based on significance level of similarity score. Heatmaps display p-values of similarity scores between sci-ATAC-seq cell types and bulk DNase-seq seq samples for each category, bar chart displays maximum abundance across all tissues for each cell type with no significant correlation to a bulk DNase-seq sample. NS = not significant.

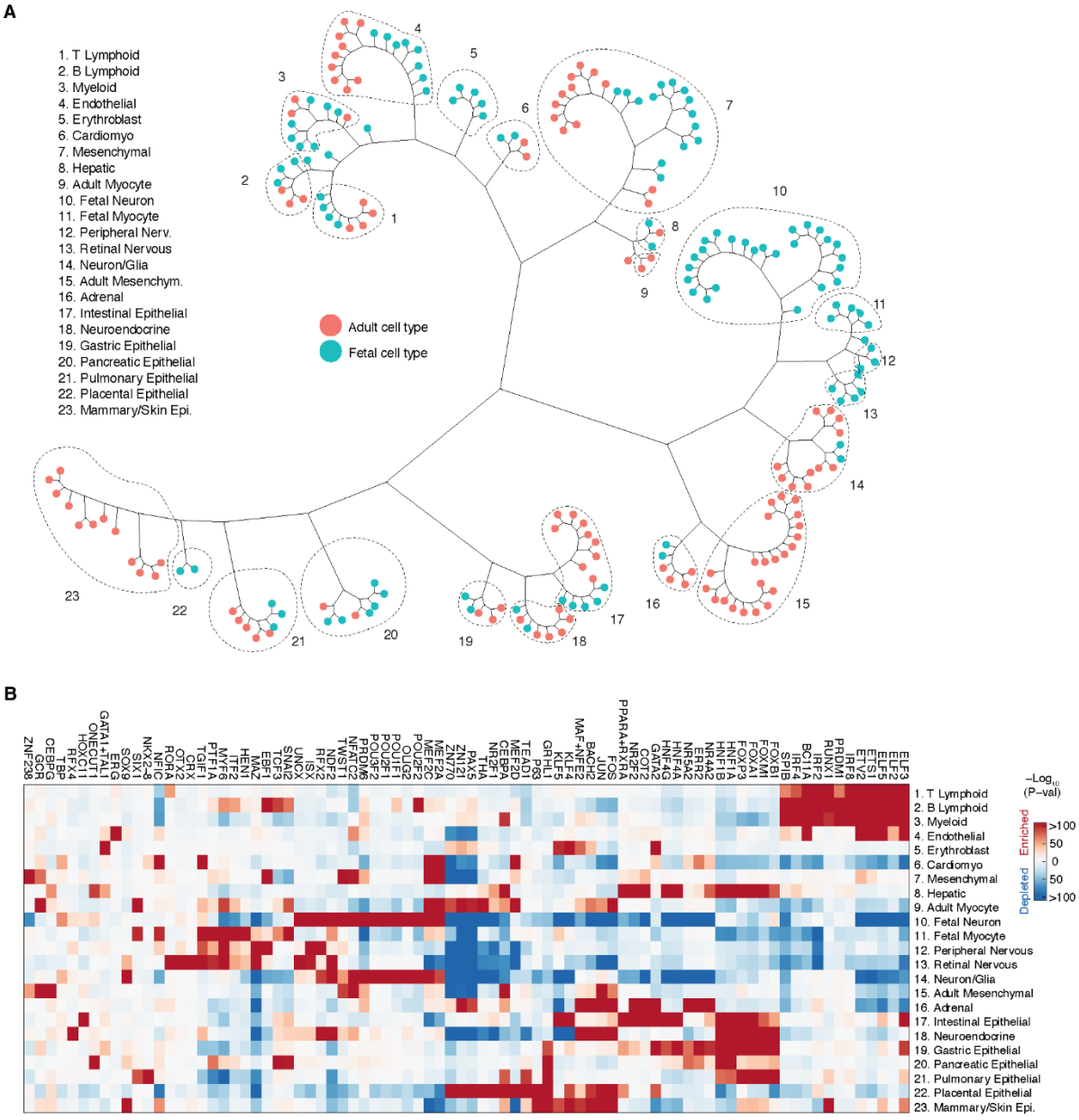

A
469,978 / 546,029
This study
420,152
ENCODE
380,506
B
p = 3.62e−14
Max similarity
cell line (n = 141)
primary cell (n = 62)
bulk tissue (n = 88)
C
Similarity score
Max abundance in Tissue (%)
D
Macrophage Gen/Alv
Macrophage General
Memory B
Plasma B
T lymphocyte 2 (CD4+)
T Lymphocyte 1 (CD8+)
Natural Killer T
Glutamatergic 1
Glutamatergic 2
Oligodendrocyte
Mesothelial
Esophageal Epithelial
Mammary Luminal Epi 1
Acinar
Enterocyte
V Cardiomyocyte
Follicular
CD14+ monocyte
GM12864
MM.1S
T−helper 17
occipital lobe
pons
NCI−H226
bronchial epithelial
T47D
body of pancreas
small intestine
heart right ventricle
thyroid gland
Endothelial General 1
Lymphatic
Endothelial General 2
Endothelial General 3
Endothelial Myocardial
Fibro General
Fibro Epithelial
Cardiac Fibroblast
Schwann General
Melanocyte
Type I Skeletal Myocyte
Type II Skeletal Myocyte
Transitional Cortical
Glomerulosa
A Cardiomyocyte
Mast
Naive T
Hepatocyte
Keratinocyte 1
Colon Epithelial 1
Mammary Basal Epi
Chief
blood vessel endothelial
lymphatic endothelial
glomerular endothelial
fibroblast of dermis
forelimb muscle
foreskin melanocyte
gastrocnemius medialis
adrenal gland
heart right ventricle
CMK
T−helper 17 cell
right lobe of liver
foreskin keratinocyte
small intestine
T47D
stomach
P−value of similarity
< 0.001
< 0.01
< 0.05
NS
17
22
28
44
Sm Ms Uterine
Sm Ms Colon 1
Fibro Nerve
Sm Ms Vaginal
Sm Ms Colon 2
Alveolar Type 2
Nerve Stromal
Sm Ms Muscularis 2
Sm Ms General
Microglia
Mammary Luminal Epi 2
Skin Granular Epidermal
Alveolar Type 1
Cardiac Pericyte 1
Astrocyte 1
Gastric Neuroendocrine
Foveolar
Alveolar Cap Endo
Pericyte General 1
Skin Eccrine Epidermal
Pericyte General 3
Cardiac Pericyte 4
Endothelial Exocrine
Astrocyte 2
Myoepithelial
Keratinocyte 2
Satellite
Pericyte General 2
Skin Basal Epidermal
Club
Paneth
Enteric Neuron
Pericyte Muscularis
Airway Goblet
Colon Epithelial 3
Sm Ms GI
Cardiac Pericyte 2
BBB Endothelial
Sm Ms Mucosal
Ciliated
Cardiac Pericyte 3
Enterochromaffin
Tuft
Pericyte General 4
< 10%
< 5%
median = 3.2
Max abundance in Tissue (%)
Beta 1
Alpha 1
Delta/Gamma
Sm Ms GE junction
Oligo Precursor
GABA 2
GABA 1
Cortical Epithelial
Endocardial
Fasciculata
Luteal
Fibro GI
Sm Ms Muscularis 1
Fibro Liver Adrenal
Sm Ms Muscularis 3
Fibro Musde
Adipocyte
Vasc Sm Muscle 1
Vasc Sm Muscle 2
Beta 2
Alpha 2
Mammary Epithelial
SI Goblet
Colon Epithelial 2
Colon Goblet
Ductal
Parietal
Alveolar Type 2/Immune
WERI−Rb−1
cerebellar cortex
brain
occipital lobe
HS−5
glomerular endothelial
adrenal gland
ovary
forelimb muscle
AG04450
fibro vil mesenchyme
fibroblast of dermis
fibroblast of gingiva
fibroblast of skin of back
bone marrow stromal
proximal tubule epithelial
kidney tubule cell
mammary epithelial cell
small intestine
pancreas
stomach
upper lobe of left lung



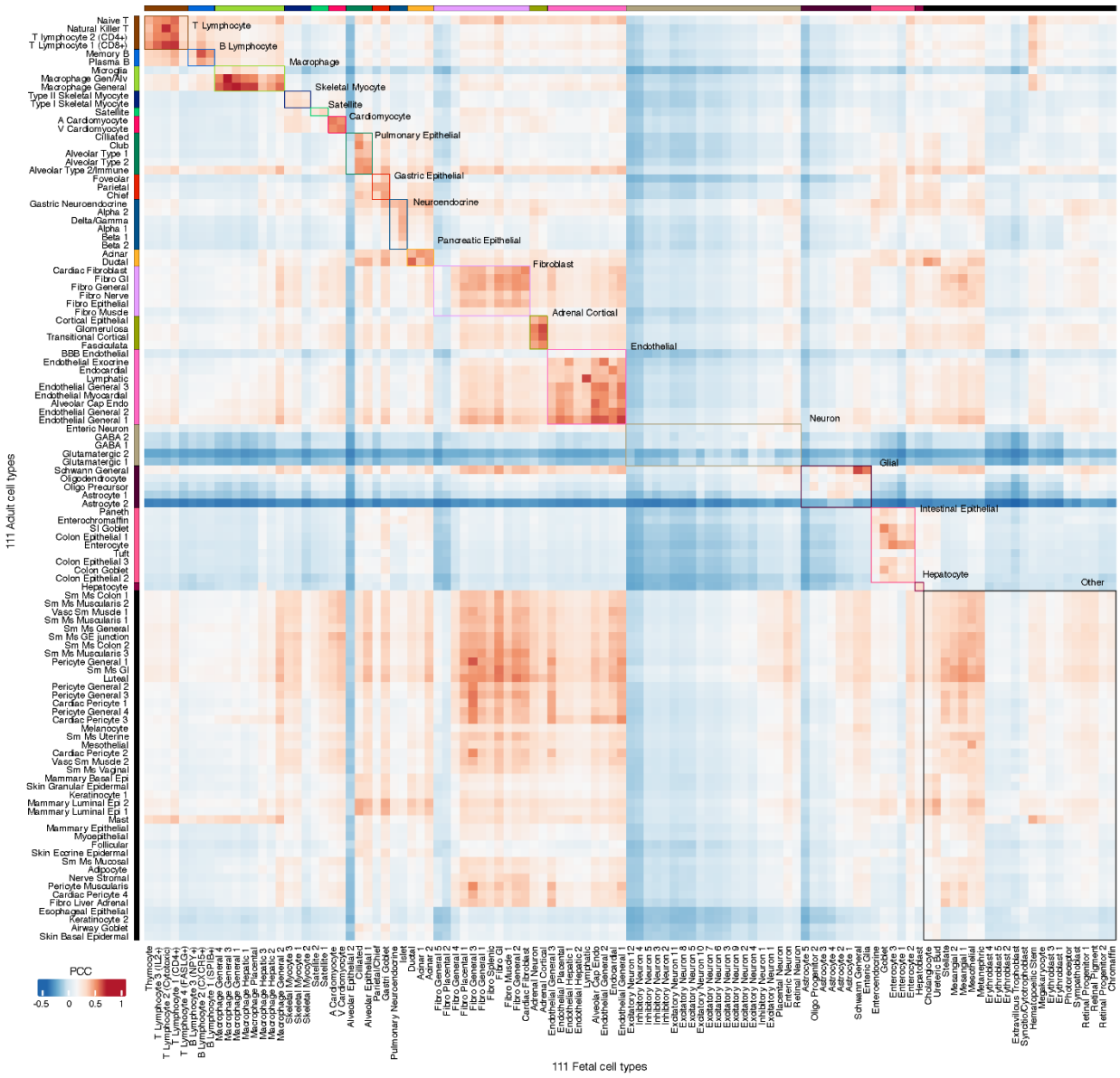
Supplemental Figure 11: Enrichment of GTEx tissue eQTLs in human cell type cCREs. A) Z-scores for enrichment of GTEX eQTLs from corresponding bulk tissues in cCREs from each sci-ATAC-seq cell type. B) The chromatin accessibility enrichment of GTEx tissue eQTLs in each tissue was computed as described in Methods, and the maximum value across the 25 matching tissue types was used for the plot.



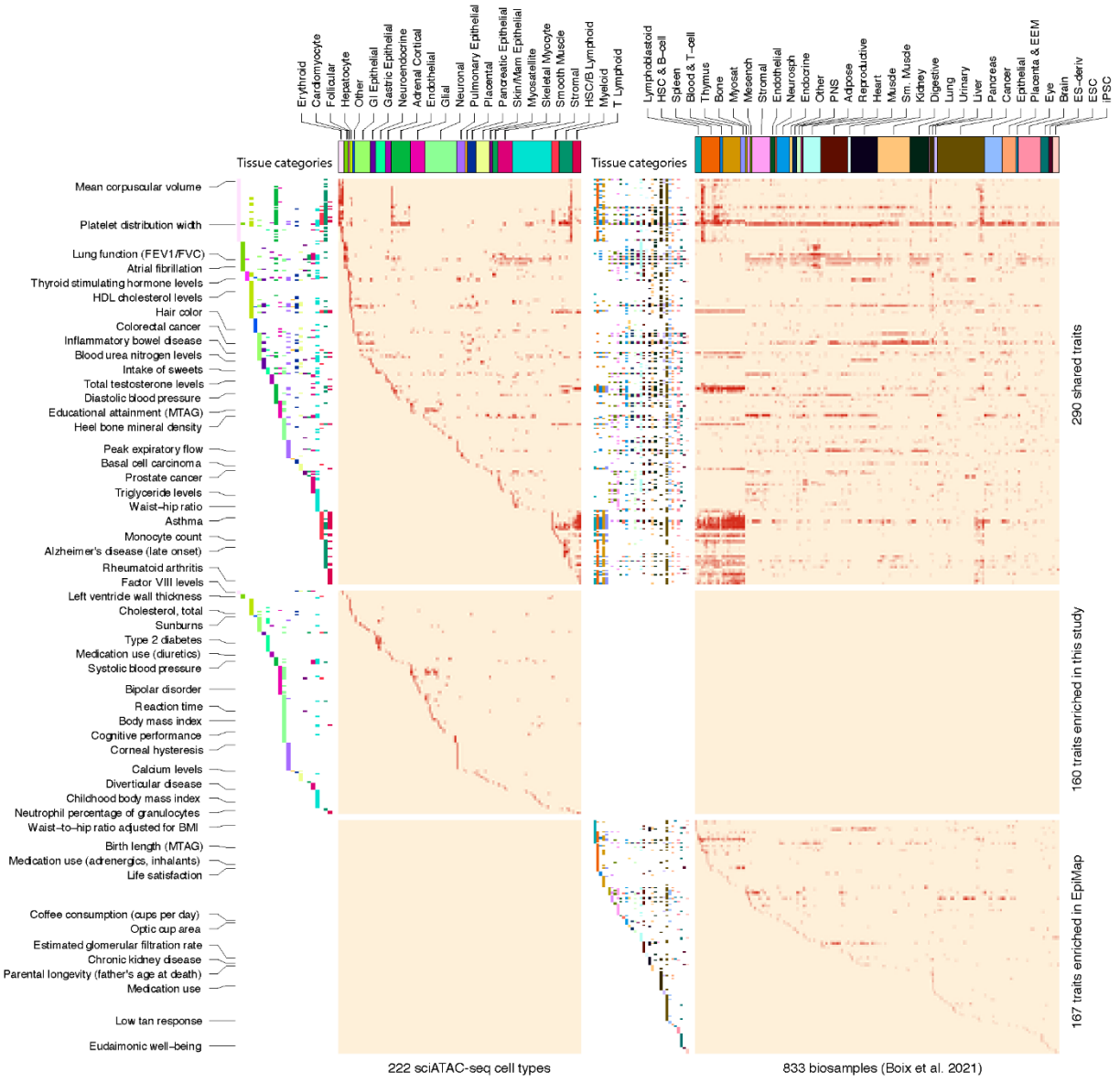
Supplemental Figure 12: Clustering analysis of sci-ATAC-seq data from 15 fetal tissues. Each scatter plot shows the UMAP embedding of nuclei from one of 35 major cell groups. Subclusters are indicated by different colors. Subclustering is displayed for 19 out of 35 major cell groups that had more than one subcluster.



Supplemental Figure 13: Phylogenetic analysis of fetal and adult cell types. Phylogenetic tree showing relationship between fetal and adult cell types. The tree was constructed using the FASTME algorithm. The distance between two cell types was calculated in the batch-corrected low-dimensional space, and defined as the average Euclidean distance of cell pairs from the two populations. Leaf nodes are colored by life stage. B) Heatmap showing the TF motif enrichments in cCREs that distinguish clades labeled in Panel A. Color represents $-\log_{10}P$. Only the most enriched TF motif in each of the previously identified motif archetypes (Vierstra et al., 2020) was selected as the representative and the top 5 motifs were selected for each group. Full GO and motif enrichments are available to download on Mendeley Data: 10.17632/yv4fzv6cnm.1.

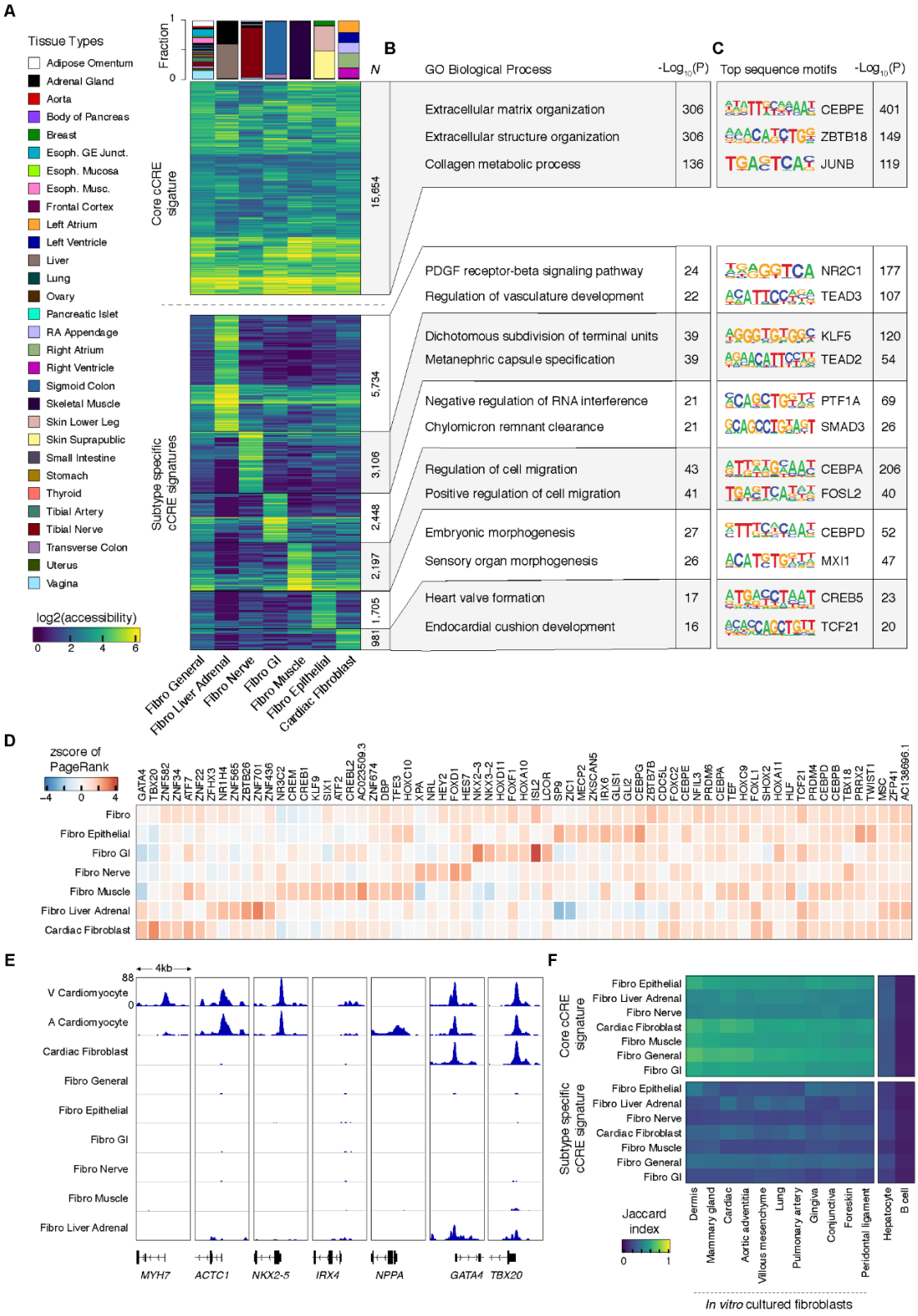


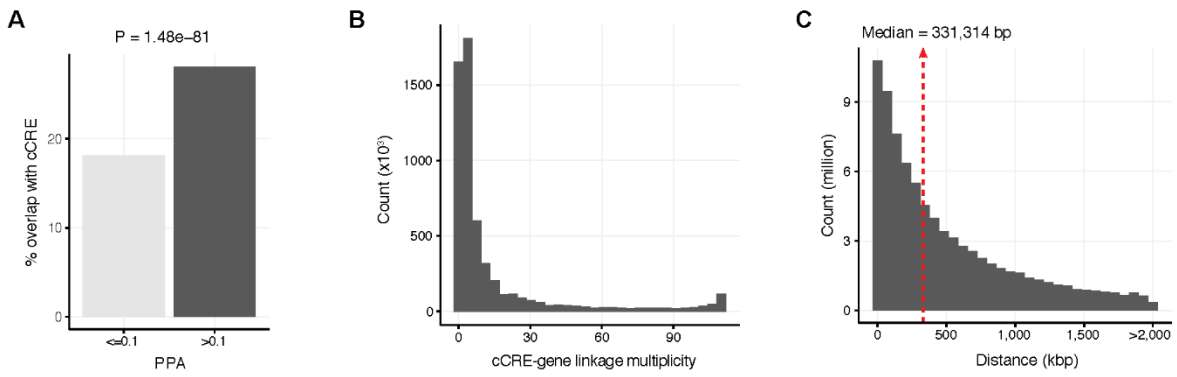
Supplemental Figure 14: Comparison between fetal and adult cell types. Heatmap showing Pearson correlation coefficients (PCC) between chromatin accessibility at the 1.2 million cCREs annotated in the current study in 111 adult cell types and 111 fetal cell types. Major cell groups are indicated in boxes.



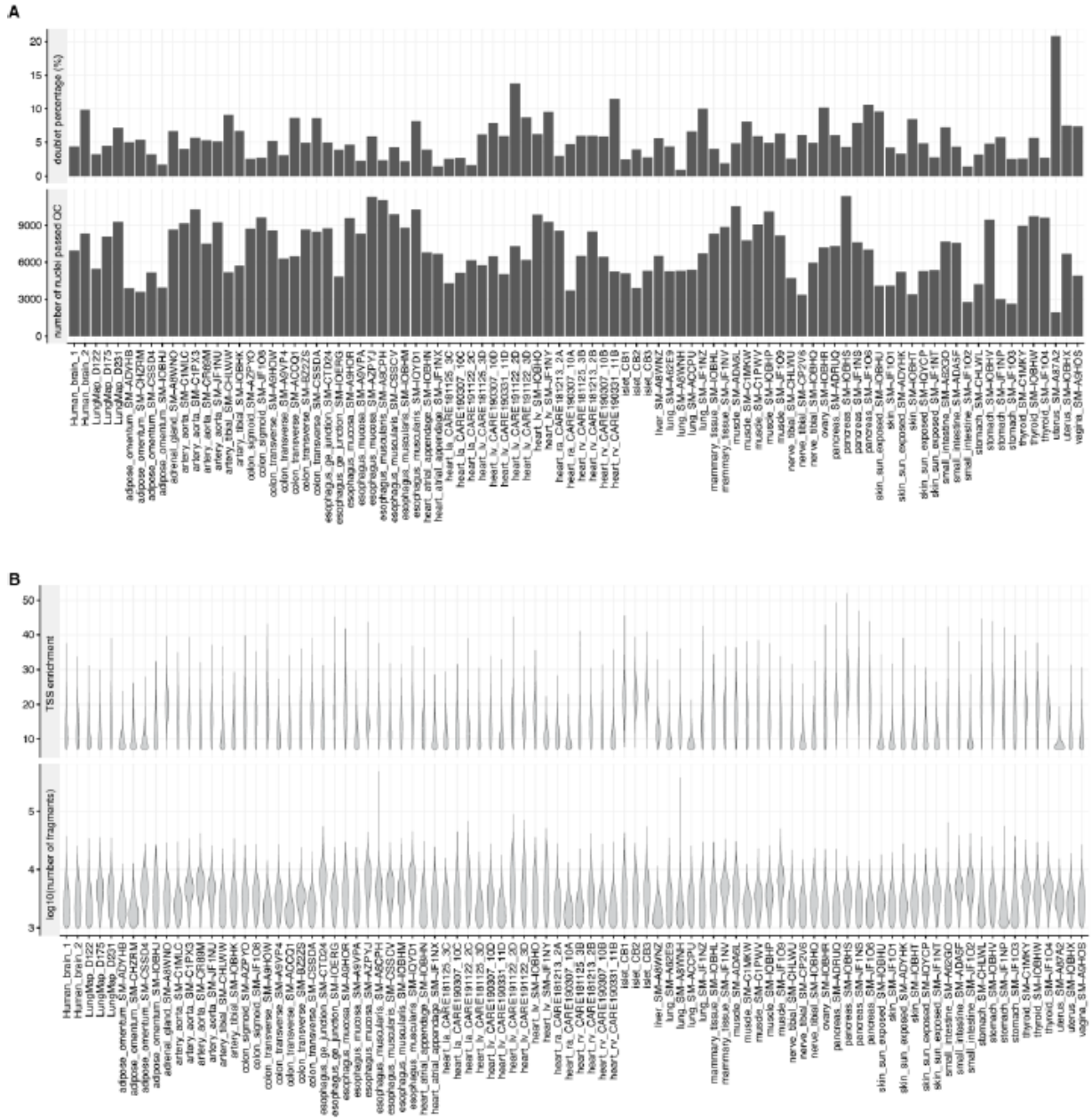
Supplemental Figure 15: GWAS variant enrichment analysis with single-cell cCRE atlases and EpiMaps. Heatmap showing significant hypergeometric enrichments (FDR < 0.1%) of index variants from 617 traits downloaded from the NHGRI-EBI catalogue (Buniello et al., 2019) (rows) within cCREs identified from 222 fetal and adult cell types from the current study (left column) and 833 aggregated epigenomic datasets from bulk tissues, cell lines, and primary cells compiled by (Boix et al., 2021) (right column). The data underlying the heatmap are available to download on Mendeley Data: 10.17632/yv4fzv6cnm.1.

Supplemental Figure 16: Chromatin features of fibroblasts in different tissue environments. A) Heatmap representation of core fibroblast cCREs and fibroblast subtype-specific elements. Color represents $\log_2$(accessibility). Bar plot on the top indicates tissues of origin by percentage for each fibroblast subtype. All subtypes showed comparable chromatin accessibility at a set of core fibroblast cCREs, each also showed subtype-specific chromatin accessibility patterns, which were enriched for biological process ontology terms that suggested potential subtype-specific functions. B) Top GREAT biological process ontology enrichments (McLean et al., 2010) for core fibroblast and fibroblast subtype-specific cCREs. C) *De novo* sequence motifs and their matched known TF motifs identified by HOMER (Heinz et al., 2010). D) Heatmap representation showing key TFs (row) in each fibroblast subtype (column) revealed using transcription regulatory network analysis. Color represents standardized PageRank scores. E) Genome browser tracks for cardiomyocytes and fibroblast subtypes from sci-ATAC-seq at several cardiomyocyte marker genes. Notably, cardiac fibroblasts were accessible at loci encoding cardiac developmental transcription factors *GATA4* and *TBX20* (Perrino and Rockman, 2006; Shen et al., 2011; Singh et al., 2005), but at other cardiomyocyte marker genes suggesting cardiogenic gene programs in cardiac fibroblasts (Furtado et al., 2014). F) Similarity indices between (top) core fibroblast cCREs and (bottom) subtype-specific cCREs with *in vitro* cultured fibroblast DNase-seq datasets, and non-fibroblast DNase-seq datasets. All fibroblast subtypes from the current study showed similarity to *in vitro* fibroblasts based on core fibroblast cCRE signatures, but minimal similarity based on subtype-specific fibroblast cCRE signatures suggesting that fibroblast subtype-specific signatures are environment dependent and may be lost during *in vitro* culturing.

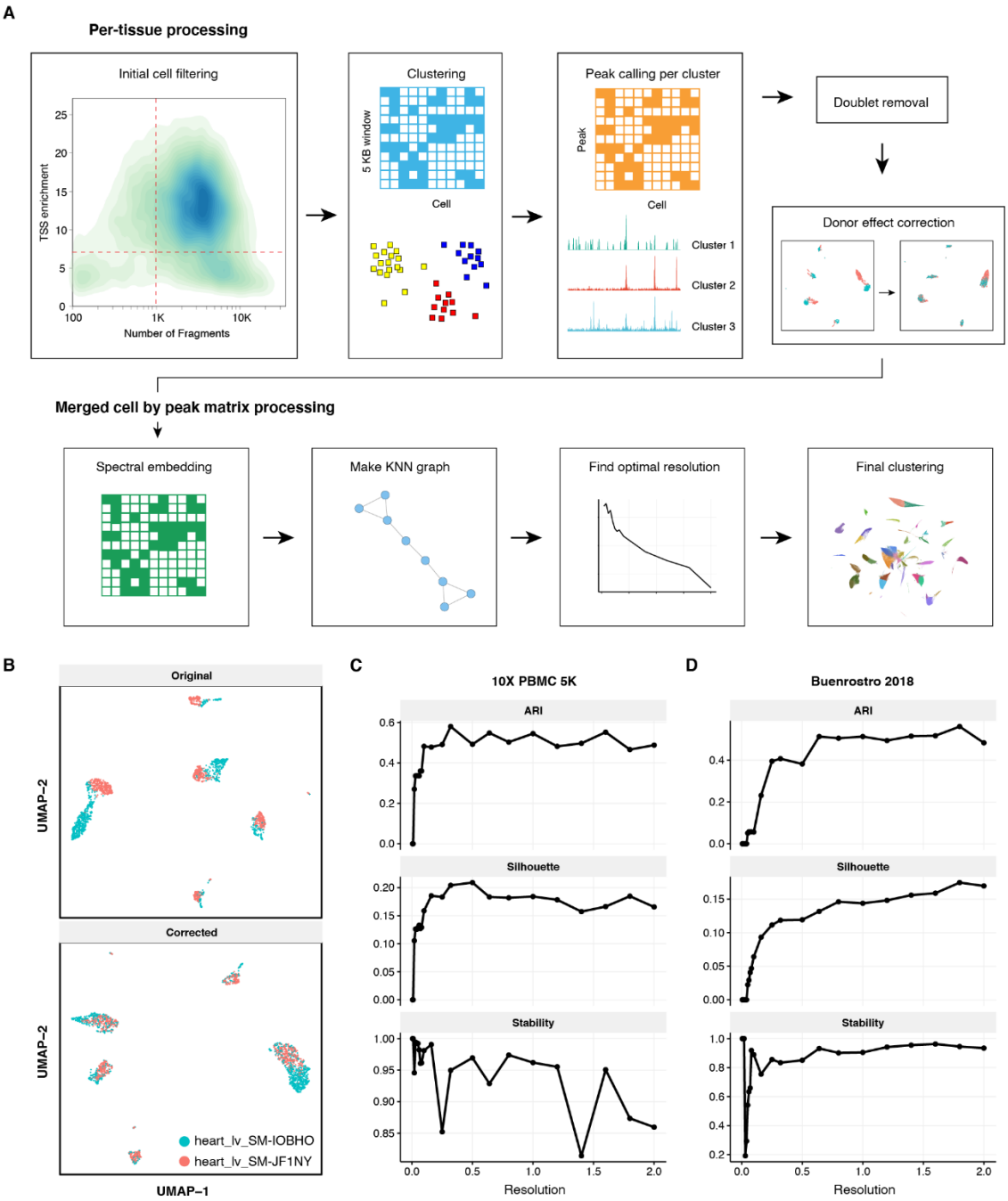

A
Tissue Types
Adipose Omentum
Adrenal Gland
Aorta
Body of Pancreas
Breast
Esoph. GE Junct.
Esoph. Mucosa
Esoph. Musc.
Frontal Cortex
Left Atrium
Left Ventricle
Liver
Lung
Ovary
Pancreatic Islet
RA Appendage
Right Atrium
Right Ventricle
Sigmoid Colon
Skeletal Muscle
Skin Lower Leg
Skin Suprapublic
Small Intestine
Stomach
Thyroid
Tibial Artery
Tibial Nerve
Transverse Colon
Uterus
Vagina
log2(accessibility)
Fraction
Core cCRE signature
Subtype specific cCRE signatures
15,654
5,734
3,106
2,448
2,197
1,705
981
Fibro General
Fibro Liver Adrenal
Fibro Nerve
Fibro GI
Fibro Muscle
Fibro Epithelial
Cardiac Fibroblast
B
N
GO Biological Process
-Log10(P)
Extracellular matrix organization 306
Extracellular structure organization 306
Collagen metabolic process 136
PDGF receptor-beta signaling pathway 24
Regulation of vasculature development 22
Dichotomous subdivision of terminal units 39
Metanephric capsule specification 39
Negative regulation of RNA interference 21
Chylomicron remnant clearance 21
Regulation of cell migration 43
Positive regulation of cell migration 41
Embryonic morphogenesis 27
Sensory organ morphogenesis 26
Heart valve formation 17
Endocardial cushion development 16
C
Top sequence motifs
-Log10(P)
CEBPE 401
ZBTB18 149
JUNB 119
NR2C1 177
TEAD3 107
KLF5 120
TEAD2 54
PTF1A 69
SMAD3 26
CEBPA 206
FOSL2 40
CEBPD 52
MXI1 47
CREB5 23
TCF21 20
D
zscore of PageRank
GATA4 TBX20 ZNF582 ZNF34 ATF7 ZNF22 ZFHX3 NR1H4 ZNF565 ZBTB26 ZNF701 ZNF436 NR3C2 CREM CREB1 KLF9 SIX1 ATF2 CREBL2 AC023509.3 ZNF674 DBP TFE3 HOXC10 XPA NRL HEY2 FOXD1 HES7 NKX2-3 NKX3-2 HOXD11 FOXF1 HOXA10 ISL2 LCOR SP9 ZIC1 MECP2 ZKSCAN5 IRX6 GLIS1 GLI2 CEBPG ZBTB7B CDC5L FOXC2 CEBPE NFIL3 PRDM6 CEBPA TEF HOXC9 FOXL1 SHOX2 HOXA11 HLF TCF21 PRDM4 CEBPD CEBPB TBX18 PRRX2 TWIST1 MSC ZFP41 AC138696.1
Fibro
Fibro Epithelial
Fibro GI
Fibro Nerve
Fibro Muscle
Fibro Liver Adrenal
Cardiac Fibroblast
E
4kb
88
0
V Cardiomyocyte
A Cardiomyocyte
Cardiac Fibroblast
Fibro General
Fibro Epithelial
Fibro GI
Fibro Nerve
Fibro Muscle
Fibro Liver Adrenal
MYH7
ACTC1
NKX2-5
IRX4
NPPA
GATA4
TBX20
F
Core cCRE signature
Subtype specific cCRE signature
Fibro Epithelial
Fibro Liver Adrenal
Fibro Nerve
Cardiac Fibroblast
Fibro Muscle
Fibro General
Fibro GI
Jaccard index
0
1
Dermis
Mammary gland
Cardiac
Aortic adventitia
Vilious mesenchyme
Lung
Pulmonary artery
Gingiva
Conjunctiva
Foreskin
Peridontal ligament
Hepatocyte
B cell
In vitro cultured fibroblasts



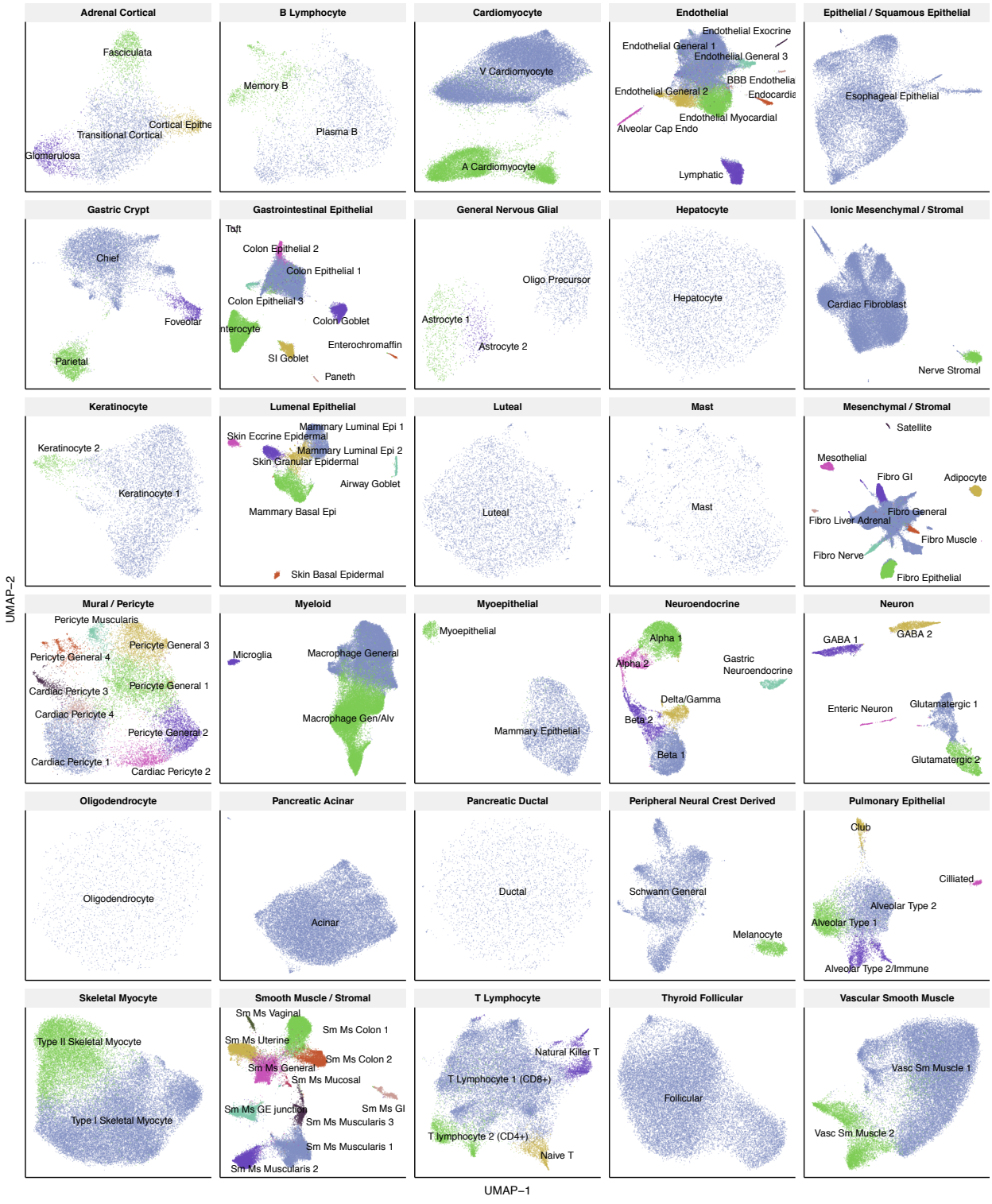
Supplemental Figure 17: Characterization of fine mapped risk variants. A) Bar graph showing the percentage of likely causal (Posterior Probability of Association; PPA > 0.1) fine mapped GWAS variants from 48 traits and diseases that overlap the union set of cCREs in adult cell types in the present study. Fisher's exact test was used to compute statistical significance. B) Histogram showing the multiplicities of cCRE-gene linkage (number of cell types having the linkage). C) Histogram showing distances in kilobase pairs (kbp) for distal cCRE-to-gene linkages from Activity by Contact (ABC) analysis (Fulco et al., 2019) (ABC score > 0.015).



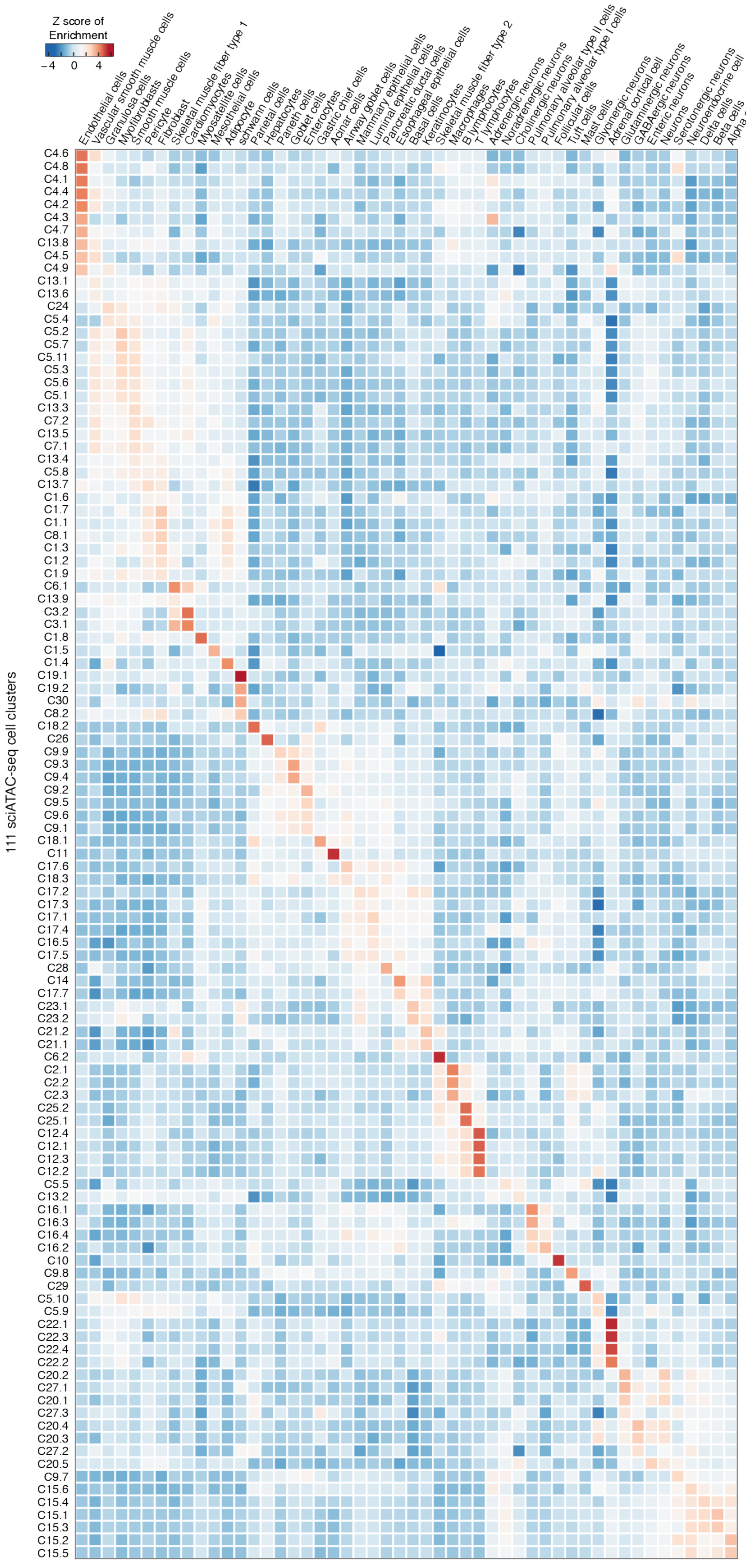
Supplemental Figure 18: Quality control for sci-ATAC-seq datasets. A) Upper bar plot shows the percentage of doublets detected in each dataset. Lower bar plot shows the number of nuclei that passed quality control in each experiment. Nuclei were first filtered by stringent quality control criteria (TSS enrichment greater than 7 and number of mapped fragments greater than 1000 per nucleus) and then subjected to doublet removal. B) Upper violin plot shows the distribution of TSS enrichments for nuclei that passed quality control in each experiment. Lower violin plot shows the distribution of number of fragments for nuclei that passed quality control in each dataset.



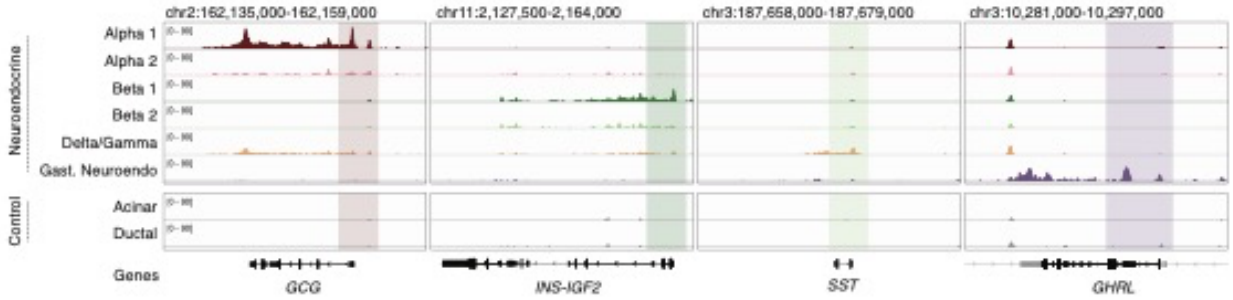
Supplemental Figure 19: Computational framework for analyzing sci-ATAC-seq data. A) Schematic illustrating the workflow of the analysis pipeline. B) Example scatter plots showing the UMAP embedding of nuclei before and after batch correction. Dots with the same color are derived from the same donor or batch. C,D) Line plots showing the Adjusted Rand Index (ARI), average silhouette width, and stability of clustering results as a function of resolution parameter in the Leiden algorithm. ARI was computed based the cell annotations from the previous study (Chen et al., 2019). To compute the stability under a particular resolution, five perturbations were conducted on the kNN graph. During each perturbation 2% of the edges were randomly selected and subjected to removal. The clustering was performed on the perturbed graph and the average ARI between different runs were taken as the stability.



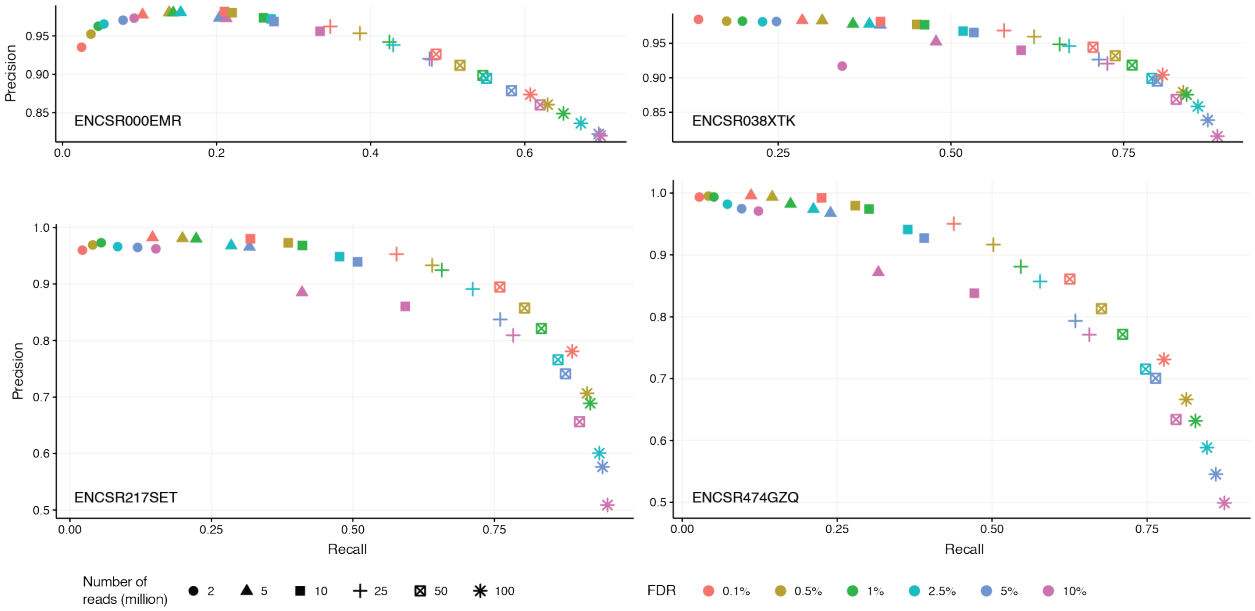
Supplemental Figure 20: Iterative clustering analysis of the 30 major cell groups. Each scatter plot shows the UMAP embedding of nuclei from one of 30 major cell groups. Subclusters are indicated by different colors. 22 out of 30 major cell groups had more than one subcluster.



Supplemental Figure 21: Evidence supporting the annotation of 111 cell clusters. Heatmap representation showing the marker gene enrichment of cell types. The marker genes were downloaded from the PanglaoDB (Franzén et al., 2019).



Supplemental Figure 22: Example of focused lineage marker gene consideration for cell type annotation. Genome browser tracks show chromatin accessibility for six neuroendocrine cell types subclustered from the neuroendocrine major cell group. Non-neuroendocrine pancreatic cell types are included as controls. Neuroendocrine cell marker genes encoding *GCG, INS-IGF2, SST,* and *GHRL* are indicated in black, neighboring genes are indicated in gray. Transcription start site(s) of the indicated genes are highlighted. Gast. Neuroendo = Gastric Neuroendocrine.



Supplemental Figure 23: Peak call benchmarking using different FDR cutoff and down-sampling rate. Each plot shows the precision (y-axis) and recall (x-axis) of peaks called by MACS2 (Zhang et al., 2008) under different combinations of FDR (color) and number of reads (shape). In all cases, the ground truth was taken as the peaks produced by the ENCODE consortium using all reads.

3.9 Supplemental Tables

Supplemental Table 22: Donor clinical characteristics and contributions to sci-ATAC-seq datasets.

Supplemental Table 23: Quality control data for sci-ATAC-seq datasets.

Supplemental Table 24: Adult and fetal cell cluster annotations and example marker genes.

Supplemental Table 25: P-values of LDSC coefficients for 240 complex traits.

Supplemental Table 26: Likely causal GWAS variant PPAs, overlapping cCREs, corresponding cell types, motifs altered, and candidate target genes.

Supplemental Table 27: Oligo and primer sequences for sci-ATAC-seq.

3.10 Competing Interests

B.R. is a shareholder and consultant of Arima Genomics, Inc., and a co-founder of Epigenome Technologies, Inc. K.J.G is a consultant of Genentech, and shareholder in Vertex Pharmaceuticals. These relationships have been disclosed to and approved by the UCSD Independent Review Committee.

3.11 Acknowledgements

Study was conceived by: J.D.H., S.P., A.W., and B.R. Study supervision: B.R. Supervision of data generation: S.P., A.W. and B.R. Contribution to data generation: J.D.H., X.H., M.M. Contribution to data analysis: K.Z., J.D.H., J.C., O.P. Y.E.L., Y.Q. Contribution to web portal: Y.E.L., K.Z. Contribution to data interpretation: K.Z., J.D.H., S.P., A.W., K.J.G. Contribution to writing the manuscript: K.Z., J.D.H., B.R. All authors edited and approved the manuscript.

We thank the ENCODE consortium, in particular Mike Pazin (NHGRI) and Idan Gadbank (Stanford), Kristin Ardlie (Broad Institute) and Ellen Gelfand (Broad Institute), for providing the human tissue samples for the present study. We thank B. Li for bioinformatics support. We thank S. Kuan for sequencing support. We thank B. Chen for valuable discussion and feedbacks in earlier versions of the manuscript. We thank the QB3 Macrolab at UC Berkeley for purification of the Tn5 transposase. This work was supported by the Ludwig Institute for Cancer Research (to B.R.), National Human Genome Research Institute (3U54HG006997-04S2 to B.R.) and Foundation for the National Institutes of Health (AMP T2D RFP 14 to K.J.G). J.D.H. was supported in part by a Ruth L. Kirschstein Institutional National Research Service Award T32 GM008666 from the National Institute of General Medical Sciences. Work at the UCSD Center for Epigenomics was supported by the UC San Diego School of Medicine.

The content of Chapter 3, in full, is a reprint of the material as it appears in Cell 2022. Zhang, Kai†; Hocker, James D.†; Miller, Michael; Hou, Xiaomeng; Chiou, Joshua; Poirion, Olivier; Qiu, Yunjiang; Li, Yang; Gaulton, Kyle J.; Wang, Allen; Preissl. Sebastian; Chi, Ren, B. The dissertation author was a co-primary investigator and co-first author of this paper.

Chapter 4: Future Directions and Preliminary Data

4.1 Summary

With this dissertation, I have shown the utility of single cell epigenomic analysis to reveal gene regulatory programs in distinct human cell types and decipher the association of cell type-resolved cCREs with noncoding risk variants. In chapter 2, I defined >287,000 cCREs in the four chambers of the human heart at single-cell resolution, which revealed cCREs and candidate transcription factors associated with cardiac cell types in a region-dependent manner and during heart failure. I next outlined the cardiovascular disease-associated genetic variants enriched within these cCREs, which included a strong enrichment of AF variants localized to cardiomyocyte cCREs. Further functional studies revealed that two noncoding AF variants affected a cardiomyocyte-specific cCRE controlling *KCNH2* expression and action potential repolarization.

In chapter 3, I built upon the framework from chapter 2 to apply single-cell chromatin accessibility assays to 30 adult human tissue types from multiple donors. Integration of these datasets with single-cell chromatin accessibility data from 15 fetal tissue types (Domcke et al., 2020) revealed the status of open chromatin for approximately 1.2 million candidate *cis*-regulatory elements (cCREs) in 222 distinct cell types composed of >1.3 million nuclei. I used these chromatin accessibility maps to delineate cell type-specificity of fetal and adult human cCREs and to systematically interpret the noncoding variants associated with complex human traits and diseases.

Together, these studies illustrate how knowledge of cell type-resolved human cCREs enables the interpretation of noncoding complex disease risk variants. Further, the datasets created herein provide a foundation for the analysis of gene regulatory programs in human cell types across tissues, life stages, and organ systems in health and disease.

4.2 Overview

In this final chapter, I summarize two major future directions. In section 4.3, I begin by summarizing technological limitations of current approaches to interpret mechanisms of risk variants from GWAS, along with future developments that will greatly enhance the utility of the data and frameworks developed in this dissertation. In section 4.4, I next describe the potential for the healthy reference datasets I generated to be used as a springboard for uncovering cell type-specific gene regulatory programs in diseased human tissues, with a focus on ischemic heart failure.

4.3 Improvements in risk variant association studies

4.3.1 Moving beyond common variants

Genome wide association studies have made major advances in our understanding of complex disease, and each of the tens of thousands of signals detected so far could reveal insight into the polygenic pathways underlying these diseases. However, because GWAS have traditionally depended upon targeted genotyping of specific and pre-selected SNPs using microarrays (Uffelmann et al., 2021), these studies have largely only captured information about common sequence variants. The common variants captured by traditional GWAS very often have low effect sizes - explaining only a portion of the heritability of complex diseases, which is defined as the portion of phenotypic variance in a population attributable to additive genetic factors. This discrepancy between the amount of phenotypic variance in complex diseases explained by GWAS and the measurable heritability of complex traits (Wray et al., 2018) means that traditional GWAS miss a significant portion of the genetic information underlying complex disease (Manolio et al.,

2009).

One likely possibility underlying this "missing heritability" is the failure of traditional GWAS to capture information about intermediate frequency variants, which are rarer than common variants but may have larger effect sizes. Thus, a major technological advance that would improve the findings of GWAS would be to forgo the use of pre-selected SNPs using microarrays in favor of whole genome sequencing (WGS) in future association studies. However, even at the level of whole genome sequencing, current technologies do not capture the spectrum of variants associated with complex disease. Cost effective approaches to perform WGS typically utilize "NextGen" sequencing/short read technologies – that is, DNA libraries amplified via PCR often from low starting quantities, fragmented or otherwise selected to appropriate fragment size (on the order of ~hundreds of base pairs), and sequenced in parallel on an Illumina flowcell. Currently, structural sequence variants such as large insertions, deletions, inversions, or translocations as well as copy number repeat variants such as short tandem repeats (STRs) are not identified well via these approaches due to problems with their amplification (polymerase is more likely to slip over repetitive regions), sequencing (repetitive regions are more difficult to sequence), and mapping to reference genomes (the predominant use of short read technology means that short reads may not span the full length of repetitive regions, and thus may not be accurately mapped). Thus, even if all GWAS transitioned to NextGen sequencing-based WGS, many putatively disease-associated variants such as STRs, which likely account for a significant proportion of variation in gene expression (Gymrek et al., 2016) and could underlie signals from index SNPs in GWAS (Fotsing et al., 2019), could still be missed.

By contrast, "third generation" sequencing approaches are able to sequence much longer reads (up to ~50 kilobases) than Next Gen sequencing technologies (van Dijk et al., 2018). This is

advantageous for detecting short tandem repeats (STRs), large structural variants, and other variants that are not identified well by NextGen sequencing. Although these methods are currently more expensive, have higher base call error rates, and are much lower throughput than NextGen sequencing – their continued development and ultimately their employment in GWAS could help close the "missing heritability" gap and explain the spectrum of sequence variants associated with complex traits and diseases. In the future, using the body-wide cCRE maps conceptual frameworks that I developed in this dissertation in parallel with the full spectrum of rare, intermediate, and common risk variants could yield additional cell type-resolved mechanisms of complex disease and ultimately to novel treatments.

4.3.2 Improving the design of current association studies

The majority of GWAS to date have been performed using genotype data from Caucasian individuals. Summary statistics from these types of studies may not be directly transferable to individuals from non-Caucasian populations, as linkage disequilibrium patterns and variant frequencies vary notably between populations (Duncan et al., 2019). Thus, even employing existing microarray-based technologies, future GWAS in larger, more diverse, and more finely-phenotyped cohorts, including biobanks such as the UK Biobank (Sudlow et al., 2015) and the BioBank Japan Project (Ishigaki et al., 2020), will be of improved utility for the discovery of common noncoding sequence variants associated with complex disease.

4.3.3 Improvements in single cell epigenomic technologies

In Chapters 2 and 3 of this dissertation, I mapped the state of activity for over one million cCREs in over 200 cell types in the four chambers of the adult heart and across the whole fetal and

adult human body. However, this work remains an initial and incomplete sampling of human tissues and organ systems, and permits a window into gene regulation only at concrete adult and fetal life stages. While I utilized tissue from anatomic sites corresponding directly to existing biosamples in large-scale databases (Carithers et al., 2015; Stranger et al., 2017), the size and diversity of adult human organ systems make it difficult to representatively sample them in their entirety. Additionally, the single cell ATAC-seq assay solely profiles chromatin accessibility in dissociated nuclei, and thus misses key orthogonal molecular information and spatial context for each cell. Future assays that incorporate gene expression, chromatin accessibility, histone modifications, DNA methylation, chromosomal conformation, TF binding, and spatial information in the same single-cell will greatly enhance our understanding of gene regulation in human cell types (Zhu et al., 2020) - leading to better body-wide maps of cCREs which can be employed through similar frameworks to those developed in this dissertation for the improved interpretation of noncoding complex disease variants.

4.4 Single cell epigenomic and transcriptomic analysis of human heart failure

4.4.1 Introduction

Cardiovascular diseases are the leading cause of morbidity and mortality worldwide (Benjamin et al., 2019), and comparison of the genes and cCREs that are differentially activated in diseased cardiac cell types has the potential to reveal disease mechanisms in individual heart cell types (Reilly and Bornfeldt, 2021). Heart failure (HF) in particular is a debilitating complex cardiovascular disease, with a median survival of fewer than 2 years from diagnosis, major impacts on quality of life for patients (Metra and Teerlink, 2017; Murphy et al., 2020), and an outsized financial burden on health care systems (Heidenreich et al., 2022). A hallmark of HF is the progressive remodeling and eventual malfunction of the left ventricle (LV), the cardiac chamber responsible for pumping oxygenated blood into circulation. Cellular modifications to the LV in LV remodeling include cardiomyocyte loss and progressive deposition of extracellular matrix by activated fibroblasts leading to cardiac fibrosis (Burchfield et al., 2013; Mann and Bristow, 2005). Due in part to the difficulty in regenerating damaged myocytes and in reversing LV remodeling, treatment options are currently limited and the optimal medial regimen for HF remains symptomatic therapy (McDonagh et al., 2021). Uncovering the molecular pathways underlying LV remodeling in HF, with the goal of identifying novel targets for new HF therapies, has thus been a major research focus for decades. Previous studies using bulk measurements of gene expression (Lowes et al., 2002; Margulies et al., 2005; Sweet et al., 2018; Tan et al., 2002) in healthy and failing heart tissue have established a molecular fingerprint for HF. However, these assays utilized heterogeneous myocardium as input material to produce population average measurements.

By contrast, the human heart is composed of diverse cell types - each of which engages a

unique gene expression program (Hocker et al., 2021; Litviňuková et al., 2020; Tucker et al., 2020). These gene expression programs are in turn governed by the action of highly-cell type specific noncoding DNA sequences called *cis*-regulatory elements (Carter and Zhao, 2020), which dictate the expression of distal genes by recruiting sequence specific transcription factors (TFs) (Shlyueva et al., 2014). Profiling of cCREs in bulk cardiac tissue in combination with next-generation sequencing has enabled genome-wide discovery of candidate cis-regulatory elements (cCREs) in both healthy (Dickel et al., 2016; Gilsbach et al., 2014; Gilsbach et al., 2018; Spurrell et al., 2019; Tan et al., 2020) and failing (Spurrell et al., 2019; Tan et al., 2020) hearts at the tissue level. Single-cell omics technologies including combinatorial cellular barcoding-based sci-ATAC-seq (Cusanovich et al., 2015) and droplet-based scRNA-seq (Zheng et al., 2017) have now enabled the profiling of the cardiac epigenome and transcriptome at cell type resolution without the need for marker-based purification of individual cell types (Hocker et al., 2021). Recently, single cell transcriptomic technologies have also been applied to reveal gene expression programs engaged by cardiac cell types in dilated cardiomyopathy (Koenig et al., 2021; Linna-Kuosmanen et al., 2021), but we still lack information about the cell type-specific activities of *cis*-regulatory elements in human HF.

In the following sections, I summarize preliminary findings from our application of single cell ATAC-seq (Cusanovich et al., 2015) and single cell RNA-seq to profile chromatin accessibility and gene expression in LV samples from 19 adult human donors with and without ischemic cardiomyopathy (hereafter referred to as HF). First, I describe our profiling of 157,091 and 159,784 nuclei via sci-ATAC-seq and snRNA-seq, respectively, and organized these nuclei into 13 major cardiac cell types that are shared between the two modalities. I next describe our use of cell type-stratified differential testing to identify 11,949 differentially accessible (DA) cCREs

and 2,876 differentially expressed (DE) genes between cardiac cell types from healthy and failing hearts, which are enriched for biological pathways related to HF pathogenesis. I lastly summarize possible future directions of this research including the validation of HF-specific cCREs and target genes in *in vitro* models of cardiac cell types and the integration of epigenomic and transcriptomic data to predict cell type-resolved transcriptional regulators in heart failure.

4.4.2 Preliminary Results

In order to generate a cell atlas of chromatin accessibility and gene expression in healthy and diseased human hearts, we performed sciATAC-seq (Cusanovich et al., 2015; Preissl et al., 2018) and snRNA-seq (Zheng et al., 2017) using primary myocardial tissue samples collected from the left ventricles of 9 healthy donors and 10 donors with ischemic cardiomyopathy (Figure 13A, Table S28).

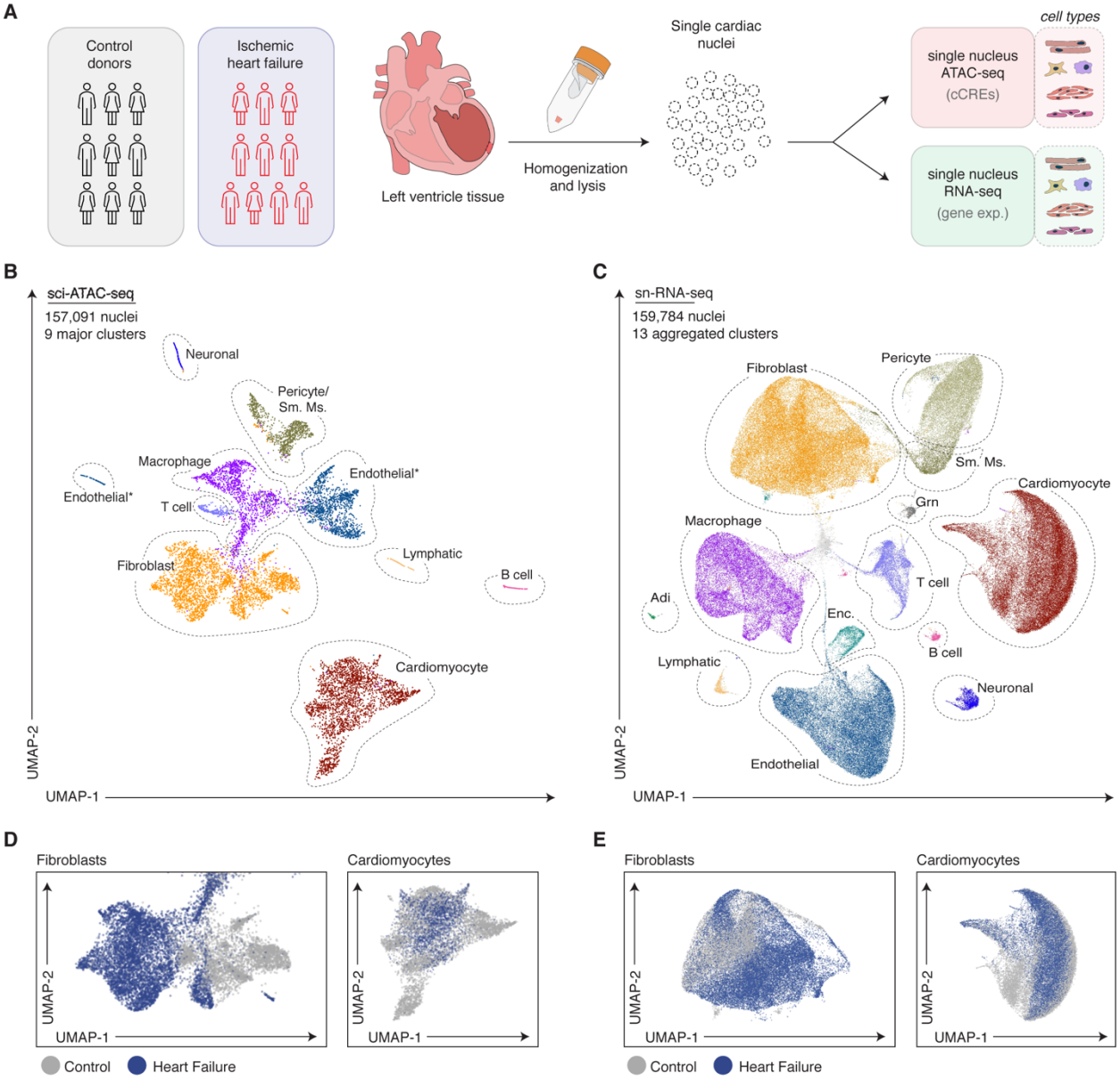


Figure 13: Single-cell chromatin accessibility and transcriptomic analysis of healthy and failing hearts. A) Experimental design schematic. Sci-ATAC-seq and snRNA-seq were performed on nuclei isolated from the left ventricle of control donors (n = 9) or ischemic heart failure donors (n = 10). (B) Uniform manifold approximation and projection (UMAP) (Leland McInnes, 2018) and clustering analysis of sci-ATAC-seq data revealed nine major clusters. * = both assigned to endothelial cluster. Left grouping of endothelial cluster was identified as endocardial cells in a subclustering analysis. Adipocytes, granulocytes, and smooth muscle were also detected by subclustering analysis leading to a total of 13 cell types identified by sciATAC-seq. C) Uniform manifold approximation and projection (UMAP) (Leland McInnes, 2018) and clustering analysis of snRNA-seq data revealed 13 major clusters. Grn. = Granulocyte. Sm. ms. = smooth muscle. D-E) UMAP clustering of fibroblasts and cardiomyocytes from (B-C) respectively, colored by control or heart failure status of the donor from which nuclei were derived.

Resulting sciATAC-seq libraries were sequenced to a median read duplication rate of 69.8% (Table S29). After filtering out low quality nuclei and potential doublets (Wolock et al., 2019), we finally obtained high quality open chromatin profiles for 157,091 nuclei, with a median of 7,061 unique open chromatin fragments per nucleus and an average transcription start site (TSS) enrichment score of 13.4 per nucleus (Figure S24). From single nucleus RNA-seq, after filtering out low quality nuclei, read contamination from highly-expressed cytoplasmic transcripts (Young and Behjati, 2020), and potential doublets (McGinnis et al., 2019), we obtained a total of 159,784 nuclear transcriptomes, with a median of 3,827 unique molecular identifiers (UMIs) and 2,010 genes detected per nucleus. We used SnapATAC (Fang et al., 2021) and Seurat (Stuart et al., 2019) to cluster nuclei based on chromatin accessibility and transcriptome profiles respectively. From sciATAC-seq data, we first identified 9 cardiac cell clusters, 4 of which were found to consist of multiple constituent cell types during a second round of clustering analysis (Figure 13B). From our transcriptomic datasets, we uncovered 23 cell groups, which we aggregated into 13 major clusters based on shared gene expression patterns (Figure 13C).

To annotate the resulting cell clusters, we examined chromatin accessibility at the transcription start sites of lineage marker genes or expression levels of the same genes for sci-ATAC-seq and scRNA-seq respectively. For instance, cardiac fibroblasts were annotated based on cluster-specific promoter accessibility and expression of *DCN,* and cardiomyocytes were annotated based on *MYH7* as exemplified in chapter 2. Altogether, we annotated each of the 13 corresponding cell clusters from sciATAC-seq and snRNA-seq with a cell type label (Figure 13B-C). We next examined our clustering results for disease-specific effects, which revealed major HF-specific differences in both chromatin accessibility and gene expression which were not explained by either donor or batch effects. HF-specific clustering differences were particularly pronounced

in cardiomyocytes and cardiac fibroblasts, which segregated within their respective clusters on the basis of HF vs. control status (Figure 1D-E). We next measured the proportion of nuclei comprising each cluster in HF and control LV samples. Consistent with previous reports, cardiomyocytes, cardiac fibroblasts, and immune cells showed disease-specific proportional differences according to HF status – with fewer cardiomyocytes, more cardiac fibroblasts, and more immune cells detected in failing left ventricles (Gilsbach et al., 2014; Gilsbach et al., 2018; Koenig et al., 2021) (Figure S25).

To examine cCREs and genes for differential expression between healthy and failing cardiac cell types, we began by identifying a master list of accessible chromatin regions in each of the 13 cardiac cell types from sciATAC-seq. We first aggregated chromatin accessibility profiles from all nuclei comprising each cell cluster then applied and a peak calling procedure optimized for single cell data (Zhang et al., 2021) to each cluster. We finally merged these accessible chromatin regions to obtain a list of 500,613 non-overlapping cCREs from both healthy and diseased cardiac cell types (Table S30). To benchmark these cCREs against previous catalogues, we compared our list with cCREs annotated by H3K27ac in healthy and failing human LVs from two previous studies (Spurrell et al., 2019; Tan et al., 2020). Our list of 500,613 cCREs captured >99% and 97.6% of cCREs annotated in bulk control and HF LVs by (Spurrell et al., 2019) and (Tan et al., 2020) respectively, and contained 336,109 additional cCREs that were not annotated by either bulk study (Figure S26). Compared to our previous atlas of cCREs defined by sci-ATAC-seq from healthy heart tissues from all four cardiac chambers (Hocker et al., 2021), our list of 500,613 HF and control LV cCREs captured 94.9% of previous elements, and contained an additional 277,697 previously unannotated elements (Figure S26). We finally compared our union of cCREs to the VISTA database of functionally validated tissue-specific mammalian enhancers

(Visel et al., 2007). Our list of cardiac cCREs captured was enriched for heart-specific enhancers, capturing 134 out of 141 (95.0%) validated cardiac enhancers compared to only 507 out of 900 (56.3%) validated non-heart enhancers (Figure S26).

We next tested for HF-specific genes and cCREs in each cardiac cell type. We first combined open chromatin fragments from sciATAC-seq or transcriptomic reads from snRNA-seq from each cell type into biological replicates, and used a generalized linear model controlling for age, BMI, and biological sex to test for significant differences in chromatin accessibility or gene expression between HF and control via the edgeR software package (Robinson et al., 2010) (Figure 14A-D).

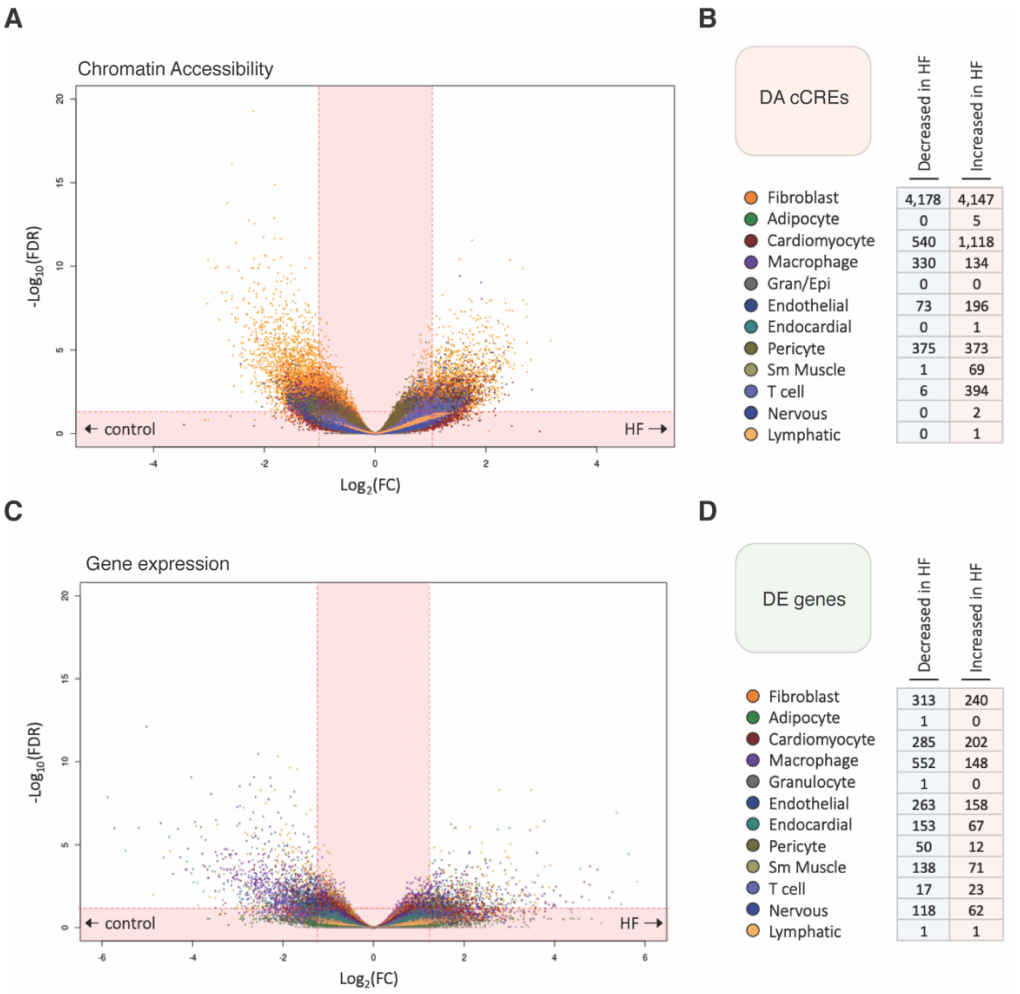


Figure 14: Differential chromatin accessibility and gene expression in healthy and failing cardiac cell types. A) Stacked volcano plots showing differentially accessible (DA) candidate *cis*-regulatory elements (cCREs) in each cell type between control (left) and heart failure left ventricles (right). Each dot represents a cCRE and the color indicates the cell type in which it was tested. cCREs with $\log_2$(fold change) > 1 and FDR < 0.05 after Benjamini-Hochberg correction (outside the shaded area) were considered as DA. (B) Number of DA cCREs between between control and heart failure left ventricles by cell type. C) Stacked volcano plot showing differentially expressed (DE) genes in each cell type between control (left) and heart failure left ventricles (right). Each dot represents a gene and the color indicates the cell type in which it was tested. Genes with $\log_2$(fold change) > 1 and FDR < 0.05 after Benjamini-Hochberg correction (outside the shaded area) were considered as DE. (D) Number of DE genes between control and heart failure left ventricles by cell type.

This strategy notably conserves information about biological replicates, a key feature for statistical tests utilizing single cell data, and demonstrated improved accuracy and specificity in a recent survey of single cell differential expression testing methods (Squair et al., 2021). In total, we detected 11,943 differentially accessible (DA) cCREs from sci-ATAC-seq and 2,876 differentially expressed (DE) genes from scRNA-seq between healthy and diseased cardiac cell types (Figure 14A-D). To benchmark this result, we repeated the differential accessibility analysis but randomly shuffled disease status labels for each biological replicate across three independent iterations. Randomly assigning disease statuses led to an average of only 64 DA cCREs detected, or a loss of >99% significant DA cCREs (Figure S27).

These initial results suggested promising disease-related differences detected between healthy and failing cardiac cell types. For example, HF-specific cCREs in cardiac fibroblasts neighbored genes involved in extracellular matrix reorganization, a defining feature of cardiac fibrosis and heart remodeling in HF, including genes such as *COL1A1* and *COL3A1* which encode protein components of type I and type III collagen isoforms respectively (Figure S28). Further, these DA cCREs were also enriched for the binding sites of transcription factors including *NFAT,* which functions downstream of calcineurin activation to drive myofibroblast differentiation, an important step leading to LV remodeling, both *in vivo* and *in vitro* (Berry et al., 2011; Davis et al., 2012; Lighthouse and Small, 2016; Molkentin et al., 1998; Wilkins and Molkentin, 2004).

4.4.3 Preliminary conclusions

There is currently no cure for heart failure, but investigation of the specific regulatory pathways engaged by individual cardiac cell types in the failing human heart could lead to novel cell type specific therapies (Reilly and Bornfeldt, 2021; Weldy and Ashley, 2021). In a preliminary

analysis of gene expression and chromatin accessibility in control and failing cardiac cell types, we identified thousands of differentially accessible cCREs and differentially expressed genes, providing a window in to the regulation of gene expression in failing human hearts at cell type resolution. The current study is still limited in several ways. Firstly, we profiled only the left ventricle from donors with and without ischemic heart failure. While the left ventricle is the site of the most severe and consequential pathology in left sided heart failure, it remains to be evaluated whether the differences detected here are chamber specific or global throughout the failing heart. Future studies with larger cohorts, additional etiologies of heart failure, and additional chambers will reveal a more comprehensive view of differences between healthy and diseased cardiac cell types in human heart failure. These limitations notwithstanding, I envision several exciting strategies to capitalize on these preliminary findings which I summarize below.

4.4.4 Future directions

4.4.4.1 Functional validation of HF-specific cCREs in cardiac fibroblasts

Cardiac fibroblasts remodel extensively in the failing heart leading to cardiac fibrosis and contributing to LV remodeling in end stage disease (Travers et al., 2016), and targeting myocardial fibrosis is an appealing possible avenue for future HF treatments (Aghajanian et al., 2019; Fan and Guan, 2016). We detected more HF-associated differences in chromatin accessibility in cardiac fibroblasts than any other cell type, along with hundreds of differentially expressed genes. However, because these differences may be a result of disease induced changes rather than causal regulators of maladaptive phenotypes (Porcu et al., 2021), functional validation of their roles in shaping HF-associated phenotypes in cardiac fibroblasts is a critical next step in their evaluation.

One possible method to link the preliminary findings described above to measurable

disease-associated phenotypes would be to apply high throughput CRISPRi screening methods (Chen et al., 2021) in *in vitro* model systems of human cardiac fibroblasts. One could begin by stimulating primary or iPSC-derived human cardiac fibroblasts (Bao et al., 2016; Whitehead et al., 2022; Zhang et al., 2019a) with recombinant human TGFβ to induce fibroblast activation and approximate chromatin accessibility and gene expression changes that occur in the failing heart (Alexanian et al., 2021). One could then perform Bulk ATAC-seq and RNA-seq on stimulated and unstimulated *in vitro* cardiac fibroblasts and compare the resulting data to control and HF fibroblasts from sci-ATAC-seq to assess 1) whether HF-specific chromatin accessibility and gene expression patterns can be induced *in vitro* and 2) the optimal *in vitro* fibroblast system, whether primary or iPSC-derived, to mimic *in vivo* HF-associated chromatin accessibility and gene expression changes.

Proceeding with these *in vitro* cardiac fibroblasts, one could apply recently-published high throughput CRISPR interference (CRISPRi) screening methods (Chen et al., 2021) to reveal pro-activation HF-specific cardiac fibroblast genes and enhancers. Briefly, one could design a CRISPR guide RNA library of redundant probes targeting 1) HF-specific cardiac fibroblast cCREs 2) HF-specific cardiac fibroblast gene promoters 3) the promoters of pro-growth genes such as *MYC* as a positive control, and 4) non-targeting controls. One could then transfect this library into iPSC-derived or primary CFs, activate them with TGFβ, and following several population doublings with CRISPRi, measure the redundant probes that are no longer present in activated fibroblast populations to reveal pro-growth enhancers. With cell type-resolved HF-specific, pro-growth enhancers in cardiac fibroblasts, one could link these enhancers to target genes using Activity by Contact (Fulco et al., 2019) with human left ventricle promoter capture HiC data (Jung et al., 2019) or measurements of co-accessibility from sci-ATAC-seq data (Pliner et al., 2018) as discussed in

Chapters 2 and 3. Deletion of HF-specific, pro-growth cCREs in *in vitro* fibroblasts followed by qPCR of target gene activity could validate accuracy of these predictions, as illustrated in Chapter 2. With target gene(s) of HF-specific pro-growth enhancers established, one could and perturb these targets using RNA interference in stimulated and unstimulated *in vitro* cardiac fibroblasts to measure whether they influence cardiac fibrosis related phenotypes such as the ability to induce contraction of a collagen gel (Alexanian et al., 2021).

These experiments illustrate one possible avenue to explore the functional mechanisms of the HF-specific cCREs we identified and link them to measurable phenotypes associated with human heart failure. They may also be applicable to other cell types and HF-associated phenotypes that can be modeled using primary cells or iPSC derived lines, such as cardiomyocytes and endothelial cells (Jang et al., 2019; Karakikes et al., 2015).

4.4.4.2 Transcrciptional regulators of HF-specific gene expression programs

A major advantage of having both differential chromatin accessibility and gene expression data together, instead of either modality separately, is that these data can be integrated to reveal transcriptional regulators of cardioprotective and/or maladaptive gene expression programs in specific cardiac cell types. For instance, using single cell RNA-seq data, one could subset all differentially expressed genes to only the transcriptional regulators in the human genome (Lambert et al., 2018) to suggest TFs with increased and decreased activity in HF for single cardiac cell types. Separately, using single cell ATAC-seq data, one could extract DA cCREs which contain a binding site for differentially expressed TFs (Grant et al., 2011). Using Activity by Contact (Fulco et al., 2019) or measurements of co-accessibility from sci-ATAC-seq data (Pliner et al., 2018) to link DA cCREs containing binding sites for these DE TFs genome wide, one could analyze the list

of target genes for biological processes and ontologies related to cardioprotective or maladaptive processes in single cell types. Transitioning to iPSC derived models of cardiac cell types such as cardiomyocytes, cardiac fibroblasts, or endothelial cells, one could perturb these TFs using RNA interference and perform RNA-seq to quantitate changes in predicted target gene expression, or functional assays such as collagen contraction following TGFβ stimulation in cardiac fibroblasts to assess the influence of these regulators on HF-associated phenotypes in the distinct cell types of the heart.

4.4.4.3 Intercellular communication in healthy and failing cardiac cell types

While the two strategies described above could yield additional insights into regulators of HF-associated phenotypes in individual cardiac cell types, they do not consider cellular communication between these cell types. Using a database of interactions among cellular receptors and ligands (Jin et al., 2021) and the single cell transcriptomic data from healthy and failing hearts we generated, one could examine differential intercellular communication networks between healthy and diseased cell types to peer into differential biological processes occurring between cell types in the failing heart.

4.4.4.4 Deciphering fetalization of gene expression and cCRE utilization in HF

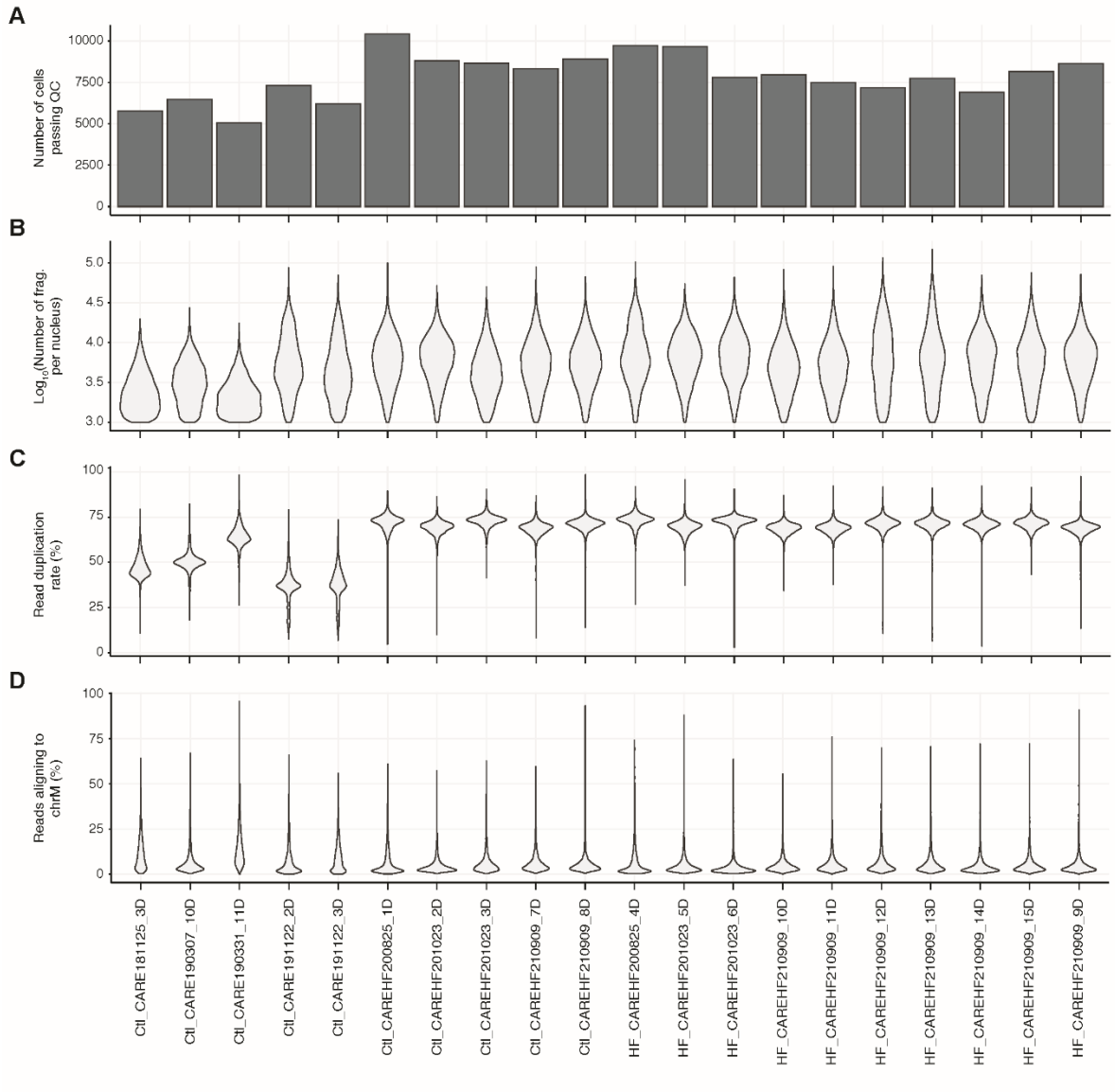
Previous studies have described a re-activation of the fetal of transcriptome (D'Antonio et al., 2021; Dirkx et al., 2013; Rajabi et al., 2007; Taegtmeyer et al., 2010) and enhancer architecture (Spurrell et al., 2019) in human heart failure. Using the single cell transcriptomic and epigenomic data we generated from healthy and failing hearts, particularly in combination with recently-published scRNA and sci-ATAC-seq datasets from fetal human heart tissues (Asp et al., 2019;

Domcke et al., 2020), could help to shed light onto the distinct cell types and cell type compositional shifts underlying these observations.

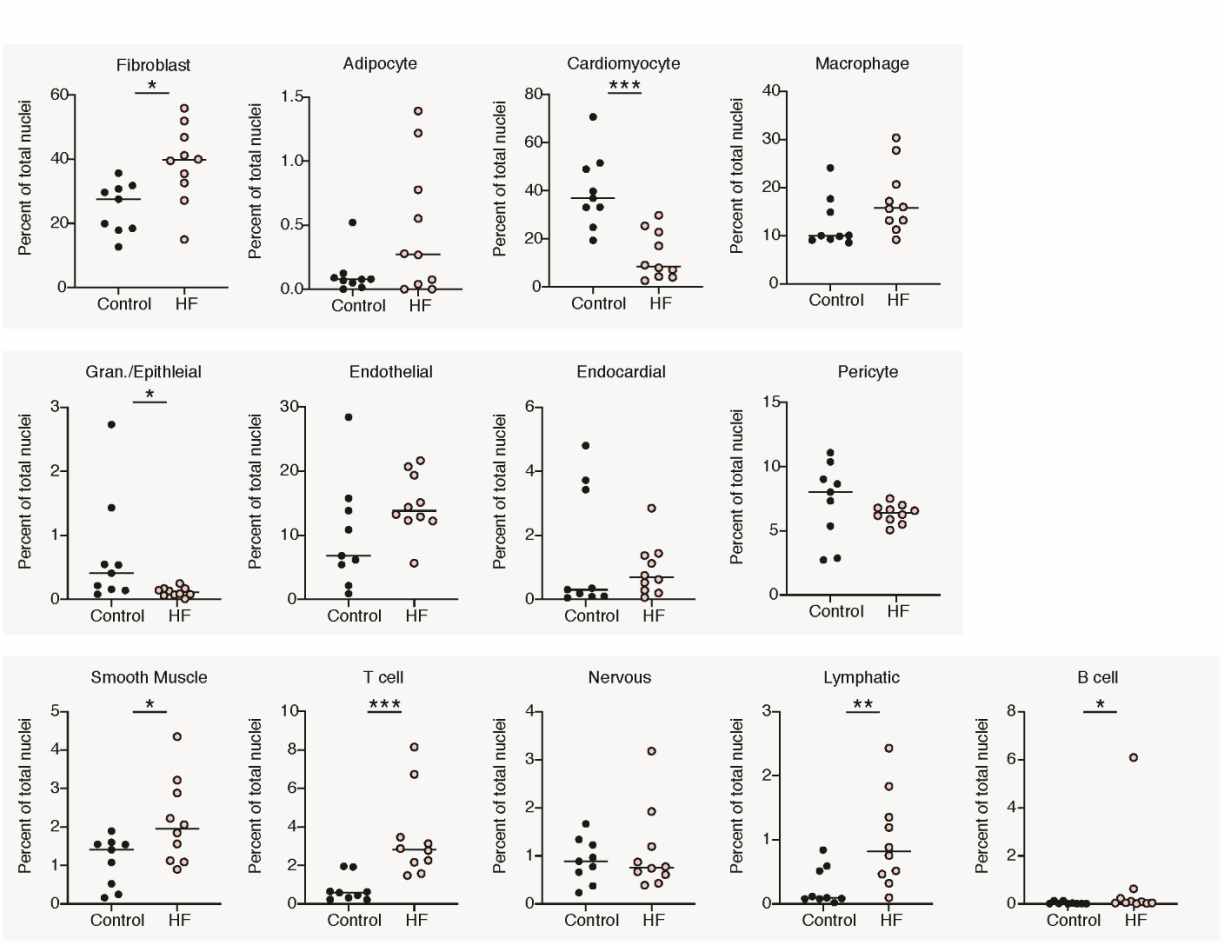
4.4.4.5 Deciphering fetalization of gene expression and cCRE utilization in HF

Chapters 2 and 3 of this dissertation focused on deciphering the association of noncoding risk variants with cell type-resolved cCREs from healthy human tissues. However, it remains to be determined whether these enrichments change in cell types from diseased human tissues. Examination of whether common variants associated with cardiovascular diseases and phenotypes – including heart failure (Arvanitis et al., 2020; Shah et al., 2020), hypertrophic cardiomyopathy (Harper et al., 2021), and cardiac morphologic measurements (Tadros et al., 2021) – are differentially enriched in cCREs from healthy vs. failing heart cell types could reveal novel associations of these complex traits with diseased cell types, or define specific mechanisms by which common variants function through cCREs that remodel in cardiac disease.

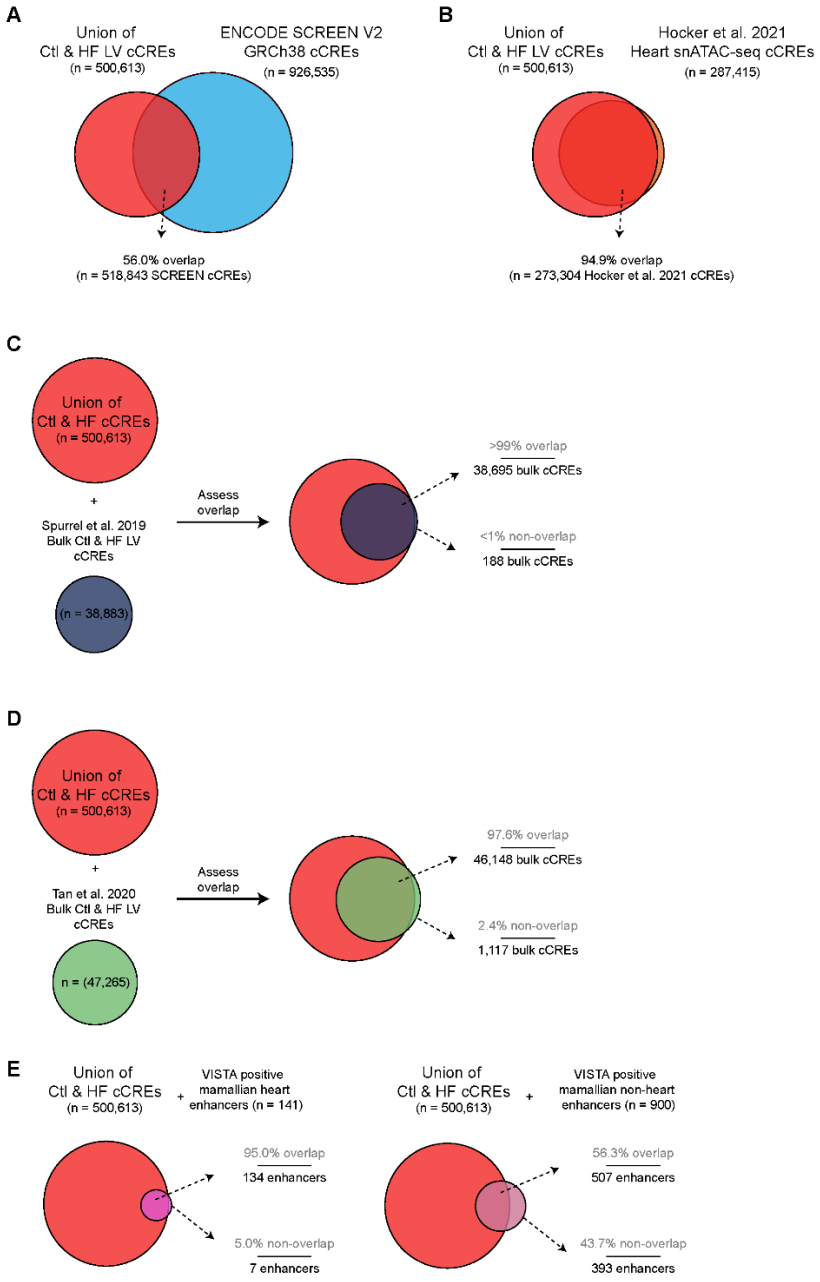
4.4.5 Supplemental Figures



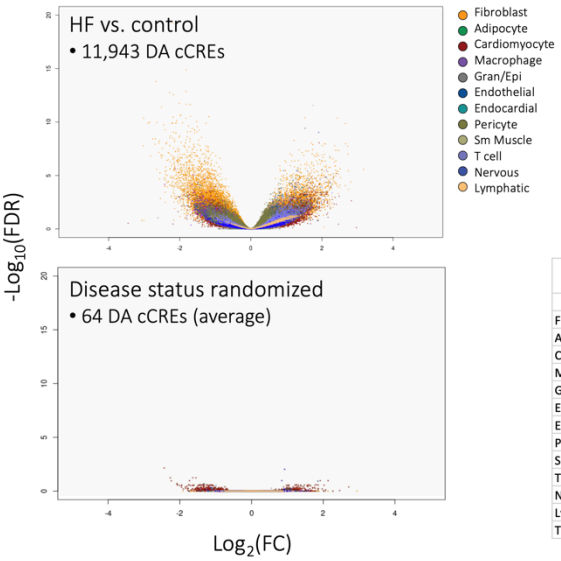
Supplemental Figure 24: Quality control for sciATAC-seq datasets. A) Upper bar plot shows the number of nuclei passing quality control for each dataset. Nuclei were first filtered by stringent quality control criteria (TSS enrichment greater than 7 and number of mapped fragments greater than 1000 per nucleus) and then subjected to doublet removal using scrublet (Wolock et al., 2019). B-D) violin plots show the distribution of fragments, read duplication rate, and mitochondrially-aligning reads per nucleus respectively stratified by dataset.



Supplemental Figure 25: Comparison of cell type composition in healthy and failing left ventricles. Dot plots displaying percent of total nuclei per sample comprising each cell type, based on the aggregated sciATAC-seq dataset. Each data point represents one human left ventricle dataset. Lines = mean. Error bars = standard deviation. Statistical significance was evaluated via Mann-Whitney-U test, * = $p<0.05$, ** $p<0.01$, *** $p<0.001$.



Supplemental Figure 26: Overlap of cCREs from healthy and diseased left ventricle with prior catalogues. A) Overlap of the union of 500,613 healthy and diseased left ventricle cCREs identified from sciATAC-seq with annotated cCREs in the human genome from the SCREEN database (Moore et al., 2020). (B) Overlap of union with healthy heart cCREs defined by sci-ATAC-seq in (Hocker et al., 2021). (C-D) Overlap of union with catalogue of healthy and diseased heart cCREs defined by bulk H3K27ac ChIP-seq in (Spurrell et al., 2019) and (Tan et al., 2020) respectively. Arrows pointing from Venn diagram indicate the number of overlapping (by at least one base pair) and non-overlapping genomic regions. (E) Overlap of union with murine *in vivo* LacZ reporter assay-validated heart enhancers (left) and non-heart enhancers (right) from the VISTA enhancer database (Visel et al., 2007). Venn diagrams are not drawn to scale.

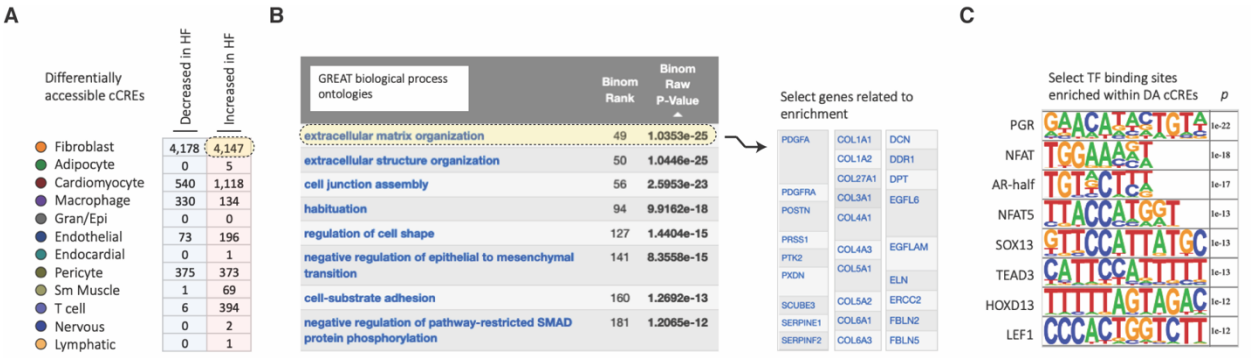


Disease status randomized over three independent iterations

	True disease status	
	Ctl_hi	HF_hi
Fibroblast	4,178	4,147
Adipocyte	0	5
Cardiomyocyte	540	1,118
Macrophage	330	134
Gran/Epi	0	0
Endothelial	73	196
Endocardial	0	1
Pericyte	375	373
Sm Muscle	1	69
T cell	6	394
Nervous	0	2
Lymphatic	0	1
Total	*11,943*	

Randomized disease staus (it1)		Randomized disease staus (it2)		Randomized disease staus (it3)	
Ctl_hi	HF_hi	Ctl_hi	HF_hi	Ctl_hi	HF_hi
38	7	25	14	0	0
0	0	0	0	0	0
63	21	10	11	1	0
0	0	0	0	0	0
1	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
1	0	0	1	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
0	0	0	0	0	0
131		*61*		*1*	

Supplemental Figure 27: Random assignment of disease status in differential accessibility test. (Left) Disease or control status was randomly shuffled among the 19 donors over three independent iterations, and the differentially accessibility test repeated. Top volcano plot shows result of true DA test (see Figure 14). Bottom volcano plot shows a representative result following random HF status assignment. (Right) Table shows number of DA cCREs detected in each cell type ($Log_2FC > 1$, FDR > 0.05) for the true DA test as compared to three independent iterations of random HF status assignment.



Supplemental Figure 28: Ontologies and motif enrichments for HF-specific cCREs in fibroblasts. A) Differentially accessible cCREs detected between HF and control cardiac cell types ($Log_2FC > 1$, FDR > 0.05; see Figure 14). HF-specific cCREs in cardiac fibroblasts are highlighted and were used as input for analyses represented in the following panels. B) Select significant GREAT biological process ontology terms for HF-specific cCREs in cardiac fibroblasts, ranked by significance. Arrow indicates genes neighboring HF-specific cCREs in cardiac fibroblasts, underlying the "extracellular matrix organization" enrichment. C) *De novo* transcription factor motif enrichment (Heinz et al., 2010) in HF-specific cCREs from cardiac fibroblasts. Statistical test for motif enrichment: hypergeometric test. P-values were not corrected for multiple testing.

4.4.6 Supplemental Tables

Supplemental Table 28: Clinical metadata for healthy and diseased heart donors.

Supplemental Table 29: Quality control data for each sciATAC-seq dataset.

Supplemental Table 30: Union of cCREs in healthy and failing cardiac cell types.

4.4.7 Acknowledgements

This work was supported by the Ludwig Institute for Cancer Research (B.R.). J.D.H. was supported in part by a Ruth L. Kirschstein Institutional National Research Service Award T32 GM008666 from the National Institute of General Medical Sciences. Work at the Center for Epigenomics was supported in part by the UC San Diego School of Medicine.

The content of Chapter 4, in part, is currently being prepared for submission for publication of the material. Hocker, James D.; Buchanan, Justin; Zhang, Kai; Miller, Michael; Shankar, Thirupura S.; Drakos, Stavros; Chi, Neil; Preissl. Sebastian; Chi, Ren, B. The dissertation author is the primary author of this material.

4.5 Conclusion

With this dissertation, I have shown the utility of single cell epigenomic analysis to reveal gene regulatory programs in distinct human cell types, decipher the association of cell type-resolved cCREs with noncoding risk variants from GWAS, and reveal testable hypothesis about the function of specific noncoding risk variants. It is my sincere hope that the datasets and frameworks outlined here provide a foundation for the discovery of new target genes and treatments for complex disease, and ultimately the improvement of human health.

REFERENCES

Aghajanian, H., Kimura, T., Rurik, J.G., Hancock, A.S., Leibowitz, M.S., Li, L., Scholler, J., Monslow, J., Lo, A., Han, W., Wang, T., Bedi, K., Morley, M.P., Linares Saldana, R.A., Bolar, N.A., McDaid, K., Assenmacher, C.-A., Smith, C.L., Wirth, D., June, C.H., Margulies, K.B., Jain, R., Puré, E., Albelda, S.M., and Epstein, J.A. (2019). Targeting cardiac fibrosis with engineered T cells. Nature *573*, 430-433.

Alexanian, M., Przytycki, P.F., Micheletti, R., Padmanabhan, A., Ye, L., Travers, J.G., Gonzalez-Teran, B., Silva, A.C., Duan, Q., Ranade, S.S., Felix, F., Linares-Saldana, R., Li, L., Lee, C.Y., Sadagopan, N., Pelonero, A., Huang, Y., Andreoletti, G., Jain, R., McKinsey, T.A., Rosenfeld, M.G., Gifford, C.A., Pollard, K.S., Haldar, S.M., and Srivastava, D. (2021). A transcriptional switch governs fibroblast activation in heart disease. Nature *595*, 438-443.

Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. Scientific Reports *9*, 9354.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., Ntini, E., Arner, E., Valen, E., Li, K., Schwarzfischer, L., Glatz, D., Raithel, J., Lilje, B., Rapin, N., Bagger, F.O., Jorgensen, M., Andersen, P.R., Bertin, N., Rackham, O., Burroughs, A.M., Baillie, J.K., Ishizu, Y., Shimizu, Y., Furuhata, E., Maeda, S., Negishi, Y., Mungall, C.J., Meehan, T.F., Lassmann, T., Itoh, M., Kawaji, H., Kondo, N., Kawai, J., Lennartsson, A., Daub, C.O., Heutink, P., Hume, D.A., Jensen, T.H., Suzuki, H., Hayashizaki, Y., Muller, F., Forrest, A.R.R., Carninci, P., Rehli, M., and Sandelin, A. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455-461.

Anene-Nzelu, C.G., Tan, W.L.W., Lee, C.J.M., Wenhao, Z., Perrin, A., Dashi, A., Tiang, Z., Autio, M.I., Lim, B., Wong, E., Tan, H.S., Pan, B., Morley, M.P., Margulies, K.B., Cappola, T.P., and Foo, R.S. (2020). Assigning Distal Genomic Enhancers to Cardiac Disease-Causing Genes. Circulation *142*, 910-912.

Arena, R., Guazzi, M., Lianov, L., Whitsel, L., Berra, K., Lavie, C.J., Kaminsky, L., Williams, M., Hivert, M.-F., Cherie Franklin, N., Myers, J., Dengel, D., Lloyd-Jones, D.M., Pinto, F.J., Cosentino, F., Halle, M., Gielen, S., Dendale, P., Niebauer, J., Pelliccia, A., Giannuzzi, P., Corra, U., Piepoli, M.F., Guthrie, G., Shurney, D., Group, A.W., Arena, R., Berra, K., Dengel, D., Franklin, N.C., Hivert, M.-F., Kaminsky, L., Lavie, C.J., Lloyd-Jones, D.M., Myers, J., Whitsel, L., Williams, M., Group, E.E.W., Corra, U., Cosentino, F., Dendale, P., Giannuzzi, P., Gielen, S., Guazzi, M., Halle, M., Niebauer, J., Pelliccia, A., Piepoli, M.F., Pinto, F.J., Group, A.W., Guthrie, G., Lianov, L., and Shurney, D. (2015). Healthy lifestyle interventions to combat noncommunicable disease—a novel nonhierarchical connectivity model for key stakeholders: a policy statement from the American Heart Association, European Society of Cardiology, European Association for Cardiovascular Prevention and Rehabilitation, and American College of Preventive Medicine. European Heart Journal *36*, 2097-2109.

Arvanitis, M., Tampakakis, E., Zhang, Y., Wang, W., Auton, A., Agee, M., Aslibekyan, S., Bell, R.K., Bryc, K., Clark, S.K., Elson, S.L., Fletez-Brant, K., Fontanillas, P., Furlotte, N.A., Gandhi, P.M., Heilbron, K., Hicks, B., Hinds, D.A., Huber, K.E., Jewett, E.M., Jiang, Y., Kleinman, A.,

Lin, K.-H., Litterman, N.K., McCreight, J.C., McIntyre, M.H., McManus, K.F., Mountain, J.L., Mozaffari, S.V., Nandakumar, P., Noblin, E.S., Northover, C.A.M., O'Connell, J., Pitts, S.J., Poznik, G.D., Sathirapongsasuti, J.F., Shastri, A.J., Shelton, J.F., Shringarpure, S., Tian, C., Tung, J.Y., Tunney, R.J., Vacic, V., Wang, X., Zare, A.S., Dutta, D., Glavaris, S., Keramati, A., Chatterjee, N., Chi, N.C., Ren, B., Post, W.S., Battle, A., and andMe Research, T. (2020). Genome-wide association and multi-omic analyses reveal ACTN2 as a gene linked to heart failure. Nature Communications *11*, 1122.

Asp, M., Giacomello, S., Larsson, L., Wu, C., Furth, D., Qian, X., Wardell, E., Custodio, J., Reimegard, J., Salmen, F., Osterholm, C., Stahl, P.L., Sundstrom, E., Akesson, E., Bergmann, O., Bienko, M., Mansson-Broberg, A., Nilsson, M., Sylven, C., and Lundeberg, J. (2019). A Spatiotemporal Organ-Wide Gene Expression and Cell Atlas of the Developing Human Heart. Cell *179*, 1647-1660 e1619.

Astle, W.J., Elding, H., Jiang, T., Allen, D., Ruklisa, D., Mann, A.L., Mead, D., Bouman, H., Riveros-Mckay, F., Kostadima, M.A., Lambourne, J.J., Sivapalaratnam, S., Downes, K., Kundu, K., Bomba, L., Berentsen, K., Bradley, J.R., Daugherty, L.C., Delaneau, O., Freson, K., Garner, S.F., Grassi, L., Guerrero, J., Haimel, M., Janssen-Megens, E.M., Kaan, A., Kamat, M., Kim, B., Mandoli, A., Marchini, J., Martens, J.H.A., Meacham, S., Megy, K., O'Connell, J., Petersen, R., Sharifi, N., Sheard, S.M., Staley, J.R., Tuna, S., van der Ent, M., Walter, K., Wang, S.Y., Wheeler, E., Wilder, S.P., Iotchkova, V., Moore, C., Sambrook, J., Stunnenberg, H.G., Di Angelantonio, E., Kaptoge, S., Kuijpers, T.W., Carrillo-de-Santa-Pau, E., Juan, D., Rico, D., Valencia, A., Chen, L., Ge, B., Vasquez, L., Kwan, T., Garrido-Martín, D., Watt, S., Yang, Y., Guigo, R., Beck, S., Paul, D.S., Pastinen, T., Bujold, D., Bourque, G., Frontini, M., Danesh, J., Roberts, D.J., Ouwehand, W.H., Butterworth, A.S., and Soranzo, N. (2016). The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. Cell *167*, 1415-1429.e1419.

Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E., Flicek, P., Gabriel, S.B., Gibbs, R.A., Green, E.D., Hurles, M.E., Knoppers, B.M., Korbel, J.O., Lander, E.S., Lee, C., Lehrach, H., Mardis, E.R., Marth, G.T., McVean, G.A., Nickerson, D.A., Schmidt, J.P., Sherry, S.T., Wang, J., Wilson, R.K., Gibbs, R.A., Boerwinkle, E., Doddapaneni, H., Han, Y., Korchina, V., Kovar, C., Lee, S., Muzny, D., Reid, J.G., Zhu, Y., Wang, J., Chang, Y., Feng, Q., Fang, X., Guo, X., Jian, M., Jiang, H., Jin, X., Lan, T., Li, G., Li, J., Li, Y., Liu, S., Liu, X., Lu, Y., Ma, X., Tang, M., Wang, B., Wang, G., Wu, H., Wu, R., Xu, X., Yin, Y., Zhang, D., Zhang, W., Zhao, J., Zhao, M., Zheng, X., Lander, E.S., Altshuler, D.M., Gabriel, S.B., Gupta, N., Gharani, N., Toji, L.H., Gerry, N.P., Resch, A.M., Flicek, P., Barker, J., Clarke, L., Gil, L., Hunt, S.E., Kelman, G., Kulesha, E., Leinonen, R., McLaren, W.M., Radhakrishnan, R., Roa, A., Smirnov, D., Smith, R.E., Streeter, I., Thormann, A., Toneva, I., Vaughan, B., Zheng-Bradley, X., Bentley, D.R., Grocock, R., Humphray, S., James, T., Kingsbury, Z., Lehrach, H., Sudbrak, R., Albrecht, M.W., Amstislavskiy, V.S., Borodina, T.A., Lienhard, M., Mertes, F., Sultan, M., Timmermann, B., Yaspo, M.-L., Mardis, E.R., Wilson, R.K., Fulton, L., Fulton, R., Sherry, S.T., Ananiev, V., Belaia, Z., Beloslyudtsev, D., Bouk, N., Chen, C., Church, D., Cohen, R., Cook, C., Garner, J., Hefferon, T., Kimelman, M., Liu, C., Lopez, J., Meric, P., O'Sullivan, C., Ostapchuk, Y., Phan, L., Ponomarov, S., Schneider, V., Shekhtman, E., Sirotkin, K., Slotta, D., Zhang, H., McVean, G.A., Durbin, R.M., Balasubramaniam, S., Burton, J., Danecek, P., Keane, T.M., Kolb-Kokocinski, A., McCarthy, S., Stalker, J., Quail, M., Schmidt, J.P., Davies, C.J., Gollub, J., Webster, T., Wong,

B., Zhan, Y., Auton, A., Campbell, C.L., Kong, Y., Marcketta, A., Gibbs, R.A., Yu, F., Antunes, L., Bainbridge, M., Muzny, D., Sabo, A., Huang, Z., Wang, J., Coin, L.J.M., Fang, L., Guo, X., Jin, X., Li, G., Li, Q., Li, Y., Li, Z., Lin, H., Liu, B., Luo, R., Shao, H., Xie, Y., Ye, C., Yu, C., Zhang, F., Zheng, H., Zhu, H., Alkan, C., Dal, E., Kahveci, F., Marth, G.T., Garrison, E.P., Kural, D., Lee, W.-P., Fung Leong, W., Stromberg, M., Ward, A.N., Wu, J., Zhang, M., Daly, M.J., DePristo, M.A., Handsaker, R.E., Altshuler, D.M., Banks, E., Bhatia, G., del Angel, G., Gabriel, S.B., Genovese, G., Gupta, N., Li, H., Kashin, S., Lander, E.S., McCarroll, S.A., Nemesh, J.C., Poplin, R.E., Yoon, S.C., Lihm, J., Makarov, V., Clark, A.G., Gottipati, S., Keinan, A., Rodriguez-Flores, J.L., Korbel, J.O., Rausch, T., Fritz, M.H., Stütz, A.M., Flicek, P., Beal, K., Clarke, L., Datta, A., Herrero, J., McLaren, W.M., Ritchie, G.R.S., Smith, R.E., Zerbino, D., Zheng-Bradley, X., Sabeti, P.C., Shlyakhter, I., Schaffner, S.F., Vitti, J., Cooper, D.N., Ball, E.V., Stenson, P.D., Bentley, D.R., Barnes, B., Bauer, M., Keira Cheetham, R., Cox, A., Eberle, M., Humphray, S., Kahn, S., Murray, L., Peden, J., Shaw, R., Kenny, E.E., Batzer, M.A., Konkel, M.K., Walker, J.A., MacArthur, D.G., Lek, M., Sudbrak, R., Amstislavskiy, V.S., Herwig, R., Mardis, E.R., Ding, L., Koboldt, D.C., Larson, D., Ye, K., Gravel, S., The Genomes Project, C., Corresponding, a., Steering, c., Production, g., Baylor College of, M., Shenzhen, B.G.I., Broad Institute of, M.I.T., Harvard, Coriell Institute for Medical, R., European Molecular Biology Laboratory, E.B.I., Illumina, Max Planck Institute for Molecular, G., McDonnell Genome Institute at Washington, U., Health, U.S.N.I.o., University of, O., Wellcome Trust Sanger, I., Analysis, g., Affymetrix, Albert Einstein College of, M., Bilkent, U., Boston, C., Cold Spring Harbor, L., Cornell, U., European Molecular Biology, L., Harvard, U., Human Gene Mutation, D., Icahn School of Medicine at Mount, S., Louisiana State, U., Massachusetts General, H., McGill, U., and National Eye Institute, N.I.H. (2015a). A global reference for human genetic variation. Nature *526*, 68-74.

Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015b). A global reference for human genetic variation. Nature *526*, 68-74.

Aylward, A., Chiou, J., Okino, M.L., Kadakia, N., and Gaulton, K.J. (2018). Shared genetic risk contributes to type 1 and type 2 diabetes etiology. Hum Mol Genet.

Bao, X., Lian, X., Hacker, T.A., Schmuck, E.G., Qian, T., Bhute, V.J., Han, T., Shi, M., Drowley, L., Plowright, A.T., Wang, Q.-D., Goumans, M.-J., and Palecek, S.P. (2016). Long-term self-renewing human epicardial cells generated from pluripotent stem cells under defined xeno-free conditions. Nature Biomedical Engineering *1*, 0003.

Baum, J., and Duffy, H.S. (2011). Fibroblasts and myofibroblasts: what are we talking about? J Cardiovasc Pharmacol *57*, 376-379.

Benjamin, E.J., Muntner, P., Alonso, A., Bittencourt, M.S., Callaway, C.W., Carson, A.P., Chamberlain, A.M., Chang, A.R., Cheng, S., Das, S.R., Delling, F.N., Djousse, L., Elkind, M.S.V., Ferguson, J.F., Fornage, M., Jordan, L.C., Khan, S.S., Kissela, B.M., Knutson, K.L., Kwan, T.W., Lackland, D.T., Lewis, T.T., Lichtman, J.H., Longenecker, C.T., Loop, M.S., Lutsey, P.L., Martin, S.S., Matsushita, K., Moran, A.E., Mussolino, M.E., O'Flaherty, M., Pandey, A., Perak, A.M., Rosamond, W.D., Roth, G.A., Sampson, U.K.A., Satou, G.M., Schroeder, E.B., Shah, S.H., Spartano, N.L., Stokes, A., Tirschwell, D.L., Tsao, C.W., Turakhia, M.P., VanWagner, L.B., Wilkins, J.T., Wong, S.S., Virani, S.S., American Heart Association Council on, E., Prevention

Statistics, C., and Stroke Statistics, S. (2019). Heart Disease and Stroke Statistics-2019 Update: A Report From the American Heart Association. Circulation *139*, e56-e528.

Bentham, J., Morris, D.L., Cunninghame Graham, D.S., Pinder, C.L., Tombleson, P., Behrens, T.W., Martín, J., Fairfax, B.P., Knight, J.C., Chen, L., Replogle, J., Syvänen, A.-C., Rönnblom, L., Graham, R.R., Wither, J.E., Rioux, J.D., Alarcón-Riquelme, M.E., and Vyse, T.J. (2015). Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nature Genetics *47*, 1457-1464.

Berry, J.M., Le, V., Rotter, D., Battiprolu, P.K., Grinsfelder, B., Tannous, P., Burchfield, J.S., Czubryt, M., Backs, J., Olson, E.N., Rothermel, B.A., and Hill, J.A. (2011). Reversibility of adverse, calcineurin-dependent cardiac remodeling. Circ Res *109*, 407-417.

Black, A.R., Black, J.D., and Azizkhan-Clifford, J. (2001). Sp1 and krüppel-like factor family of transcription factors in cell growth regulation and cancer. Journal of Cellular Physiology *188*, 143-160.

Boix, C.A., James, B.T., Park, Y.P., Meuleman, W., and Kellis, M. (2021). Regulatory genomic circuitry of human disease loci by integrative epigenomics. Nature *590*, 300-307.

Bouneffouf, D.B., I (2016). Theoretical analysis of the Minimum Sum of Squared Similarities sampling for Nyström-based spectral clustering. . 2016 International Joint Conference on Neural Networks (IJCNN), 3856–3862.

Broekema, R.V., Bakker, O.B., and Jonkers, I.H. (2020). A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. Open Biology *10*, 190221.

Bronson, P.G., Chang, D., Bhangale, T., Seldin, M.F., Ortmann, W., Ferreira, R.C., Urcelay, E., Pereira, L.F., Martin, J., Plebani, A., Lougaris, V., Friman, V., Freiberger, T., Litzman, J., Thon, V., Pan-Hammarström, Q., Hammarström, L., Graham, R.R., and Behrens, T.W. (2016). Common variants at PVT1, ATG13-AMBRA1, AHI1 and CLEC16A are associated with selective IgA deficiency. Nat Genet *48*, 1425-1429.

Bruneau, B.G., Bao, Z.Z., Tanaka, M., Schott, J.J., Izumo, S., Cepko, C.L., Seidman, J.G., and Seidman, C.E. (2000). Cardiac expression of the ventricle-specific homeobox gene Irx4 is modulated by Nkx2-5 and dHand. Developmental biology *217*, 266-277.

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

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486-490.

Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., Neale, B.M., and Schizophrenia Working Group of the Psychiatric Genomics, C.

(2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature Genetics *47*, 291-295.

Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., Suveges, D., Vrousgou, O., Whetzel, P.L., Amode, R., Guillen, J.A., Riat, H.S., Trevanion, S.J., Hall, P., Junkins, H., Flicek, P., Burdett, T., Hindorff, L.A., Cunningham, F., and Parkinson, H. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res *47*, D1005-D1012.

Burchfield, J.S., Xie, M., and Hill, J.A. (2013). Pathological Ventricular Remodeling. Circulation *128*, 388-400.

Bush, W.S., and Moore, J.H. (2012). Chapter 11: Genome-Wide Association Studies. PLOS Computational Biology *8*, e1002822.

Carithers, L.J., Ardlie, K., Barcus, M., Branton, P.A., Britton, A., Buia, S.A., Compton, C.C., DeLuca, D.S., Peter-Demchok, J., Gelfand, E.T., Guan, P., Korzeniewski, G.E., Lockhart, N.C., Rabiner, C.A., Rao, A.K., Robinson, K.L., Roche, N.V., Sawyer, S.J., Segrè, A.V., Shive, C.E., Smith, A.M., Sobin, L.H., Undale, A.H., Valentino, K.M., Vaught, J., Young, T.R., and Moore, H.M. (2015). A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. Biopreserv Biobank *13*, 311-319.

Carter, B., and Zhao, K. (2020). The epigenetic basis of cellular heterogeneity. Nature Reviews Genetics.

Chal, J., and Pourquié, O. (2017). Making muscle: skeletal myogenesis in vivo and in vitro. Development *144*, 2104-2122.

Chen, G., Gulbranson, D.R., Hou, Z., Bolin, J.M., Ruotti, V., Probasco, M.D., Smuga-Otto, K., Howden, S.E., Diol, N.R., Propson, N.E., Wagner, R., Lee, G.O., Antosiewicz-Bourget, J., Teng, J.M., and Thomson, J.A. (2011). Chemically defined conditions for human iPSC derivation and culture. Nat Methods *8*, 424-429.

Chen, H., Lareau, C., Andreani, T., Vinyard, M.E., Garcia, S.P., Clement, K., Andrade-Navarro, M.A., Buenrostro, J.D., and Pinello, L. (2019). Assessment of computational methods for the analysis of single-cell ATAC-seq data. Genome Biology *20*, 241.

Chen, P.B., Fiaux, P.C., Li, B., Zhang, K., Kubo, N., Jiang, S., Hu, R., Wu, S., Wang, M., Wang, W., McVicker, G., Mischel, P.S., and Ren, B. (2021). Discovery and Functional Characterization of Pro-growth Enhancers in Human Cancer Cells. bioRxiv, 2021.2002.2004.429675.

Chiou, J., Zeng, C., Cheng, Z., Han, J.Y., Schlichting, M., Huang, S., Wang, J., Sui, Y., Deogaygay, A., Okino, M.-L., Qiu, Y., Sun, Y., Kudtarkar, P., Fang, R., Preissl, S., Sander, M., Gorkin, D., and Gaulton, K.J. (2021). Single-cell chromatin accessibility identifies pancreatic islet cell type- and state-specific regulatory programs of diabetes risk. Nature Genetics *53*, 455-466.

Claussnitzer, M., Cho, J.H., Collins, R., Cox, N.J., Dermitzakis, E.T., Hurles, M.E., Kathiresan, S., Kenny, E.E., Lindgren, C.M., MacArthur, D.G., North, K.N., Plon, S.E., Rehm, H.L., Risch, N., Rotimi, C.N., Shendure, J., Soranzo, N., and McCarthy, M.I. (2020). A brief history of human disease genetics. Nature *577*, 179-189.

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

Consortium, E.P., Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., Kaul, R., Halow, J., Van Nostrand, E.L., Freese, P., Gorkin, D.U., Shen, Y., He, Y., Mackiewicz, M., Pauli-Behn, F., Williams, B.A., Mortazavi, A., Keller, C.A., Zhang, X.O., Elhajjajy, S.I., Huey, J., Dickel, D.E., Snetkova, V., Wei, X., Wang, X., Rivera-Mulia, J.C., Rozowsky, J., Zhang, J., Chhetri, S.B., Zhang, J., Victorsen, A., White, K.P., Visel, A., Yeo, G.W., Burge, C.B., Lecuyer, E., Gilbert, D.M., Dekker, J., Rinn, J., Mendenhall, E.M., Ecker, J.R., Kellis, M., Klein, R.J., Noble, W.S., Kundaje, A., Guigo, R., Farnham, P.J., Cherry, J.M., Myers, R.M., Ren, B., Graveley, B.R., Gerstein, M.B., Pennacchio, L.A., Snyder, M.P., Bernstein, B.E., Wold, B., Hardison, R.C., Gingeras, T.R., Stamatoyannopoulos, J.A., and Weng, Z. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature *583*, 699-710.

Consortium, G. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318.

Consortium, I.M.S.G. (2019). Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility. Science *365*.

Consortium, T.E.P. (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol *9*, e1001046.

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

Corces, M.R., Shcherbina, A., Kundu, S., Gloudemans, M.J., Fresard, L., Granja, J.M., Louie, B.H., Eulalio, T., Shams, S., Bagdatli, S.T., Mumbach, M.R., Liu, B., Montine, K.S., Greenleaf, W.J., Kundaje, A., Montgomery, S.B., Chang, H.Y., and Montine, T.J. (2020). Single-cell epigenomic analyses implicate candidate causal variants at inherited risk loci for Alzheimer's and Parkinson's diseases. Nat Genet *52*, 1158-1168.

Cordell, H.J., Han, Y., Mells, G.F., Li, Y., Hirschfield, G.M., Greene, C.S., Xie, G., Juran, B.D., Zhu, D., Qian, D.C., Floyd, J.A., Morley, K.I., Prati, D., Lleo, A., Cusi, D., Gershwin, M.E., Anderson, C.A., Lazaridis, K.N., Invernizzi, P., Seldin, M.F., Sandford, R.N., Amos, C.I., and Siminovitch, K.A. (2015). International genome-wide meta-analysis identifies new primary biliary cirrhosis risk loci and targetable pathogenic pathways. Nat Commun *6*, 8019.

Costa, R.H., Kalinichenko, V.V., Holterman, A.X., and Wang, X. (2003). Transcription factors in liver development, differentiation, and regeneration. Hepatology *38*, 1331-1347.

Craig, J. (2008). Complex diseases: Research and applications. Nature Education *1*, 184.

Cui, Y., Li, G., Li, S., and Wu, R. (2010). Designs for linkage analysis and association studies of complex diseases. Methods Mol Biol *620*, 219-242.

Cui, Y., Zheng, Y., Liu, X., Yan, L., Fan, X., Yong, J., Hu, Y., Dong, J., Li, Q., Wu, X., Gao, S., Li, J., Wen, L., Qiao, J., and Tang, F. (2019). Single-Cell Transcriptome Analysis Maps the Developmental Track of the Human Heart. Cell Rep *26*, 1934-1950 e1935.

Curran, M.E., Splawski, I., Timothy, K.W., Vincent, G.M., Green, E.D., and Keating, M.T. (1995). A molecular basis for cardiac arrhythmia: HERG mutations cause long QT syndrome. Cell *80*, 795-803.

Cusanovich, D.A., Daza, R., Adey, A., Pliner, H.A., Christiansen, L., Gunderson, K.L., Steemers, F.J., Trapnell, C., and Shendure, J. (2015). Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. Science *348*, 910-914.

Cusanovich, D.A., Hill, A.J., Aghamirzaie, D., Daza, R.M., Pliner, H.A., Berletch, J.B., Filippova, G.N., Huang, X., Christiansen, L., DeWitt, W.S., Lee, C., Regalado, S.G., Read, D.F., Steemers, F.J., Disteche, C.M., Trapnell, C., and Shendure, J. (2018). A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. Cell *174*, 1309-1324.e1318.

D'Antonio, M., Nguyen, J.P., Arthur, T.D., Matsui, H., Donovan, M.K.R., D'Antonio-Chronowska, A., and Frazer, K.A. (2021). In heart failure reactivation of RNA-binding proteins drives the transcriptome into a fetal state. bioRxiv, 2021.2004.2030.442191.

Davies, D.L., and Bouldin, D.W. (1979). A Cluster Separation Measure. IEEE Transactions on Pattern Analysis and Machine Intelligence *PAMI-1*, 224-227.

Davis, J., Burr, A.R., Davis, G.F., Birnbaumer, L., and Molkentin, J.D. (2012). A TRPC6-dependent pathway for myofibroblast transdifferentiation and wound healing in vivo. Dev Cell *23*, 705-715.

Day, F.R., Ruth, K.S., Thompson, D.J., Lunetta, K.L., Pervjakova, N., Chasman, D.I., Stolk, L., Finucane, H.K., Sulem, P., Bulik-Sullivan, B., Esko, T., Johnson, A.D., Elks, C.E., Franceschini, N., He, C., Altmaier, E., Brody, J.A., Franke, L.L., Huffman, J.E., Keller, M.F., McArdle, P.F., Nutile, T., Porcu, E., Robino, A., Rose, L.M., Schick, U.M., Smith, J.A., Teumer, A., Traglia, M., Vuckovic, D., Yao, J., Zhao, W., Albrecht, E., Amin, N., Corre, T., Hottenga, J.J., Mangino, M., Smith, A.V., Tanaka, T., Abecasis, G., Andrulis, I.L., Anton-Culver, H., Antoniou, A.C., Arndt, V., Arnold, A.M., Barbieri, C., Beckmann, M.W., Beeghly-Fadiel, A., Benitez, J., Bernstein, L., Bielinski, S.J., Blomqvist, C., Boerwinkle, E., Bogdanova, N.V., Bojesen, S.E., Bolla, M.K., Borresen-Dale, A.L., Boutin, T.S., Brauch, H., Brenner, H., Bruning, T., Burwinkel, B., Campbell, A., Campbell, H., Chanock, S.J., Chapman, J.R., Chen, Y.I., Chenevix-Trench, G., Couch, F.J., Coviello, A.D., Cox, A., Czene, K., Darabi, H., De Vivo, I., Demerath, E.W., Dennis, J., Devilee, P., Dork, T., Dos-Santos-Silva, I., Dunning, A.M., Eicher, J.D., Fasching, P.A., Faul, J.D., Figueroa, J., Flesch-Janys, D., Gandin, I., Garcia, M.E., Garcia-Closas, M., Giles, G.G., Girotto, G.G., Goldberg, M.S., Gonzalez-Neira, A., Goodarzi, M.O., Grove, M.L., Gudbjartsson, D.F., Guenel, P., Guo, X., Haiman, C.A., Hall, P., Hamann, U., Henderson, B.E., Hocking, L.J., Hofman, A., Homuth, G., Hooning, M.J., Hopper, J.L., Hu, F.B., Huang, J., Humphreys, K., Hunter, D.J., Jakubowska, A., Jones, S.E., Kabisch, M., Karasik, D., Knight, J.A., Kolcic, I.,

Kooperberg, C., Kosma, V.M., Kriebel, J., Kristensen, V., Lambrechts, D., Langenberg, C., Li, J., Li, X., Lindstrom, S., Liu, Y., Luan, J., Lubinski, J., Magi, R., Mannermaa, A., Manz, J., Margolin, S., Marten, J., Martin, N.G., Masciullo, C., Meindl, A., Michailidou, K., Mihailov, E., Milani, L., Milne, R.L., Muller-Nurasyid, M., Nalls, M., Neale, B.M., Nevanlinna, H., Neven, P., Newman, A.B., Nordestgaard, B.G., Olson, J.E., Padmanabhan, S., Peterlongo, P., Peters, U., Petersmann, A., Peto, J., Pharoah, P.D.P., Pirastu, N.N., Pirie, A., Pistis, G., Polasek, O., Porteous, D., Psaty, B.M., Pylkas, K., Radice, P., Raffel, L.J., Rivadeneira, F., Rudan, I., Rudolph, A., Ruggiero, D., Sala, C.F., Sanna, S., Sawyer, E.J., Schlessinger, D., Schmidt, M.K., Schmidt, F., Schmutzler, R.K., Schoemaker, M.J., Scott, R.A., Seynaeve, C.M., Simard, J., Sorice, R., Southey, M.C., Stockl, D., Strauch, K., Swerdlow, A., Taylor, K.D., Thorsteinsdottir, U., Toland, A.E., Tomlinson, I., Truong, T., Tryggvadottir, L., Turner, S.T., Vozzi, D., Wang, Q., Wellons, M., Willemsen, G., Wilson, J.F., Winqvist, R., Wolffenbuttel, B., Wright, A.F., Yannoukakos, D., Zemunik, T., Zheng, W., Zygmunt, M., Bergmann, S., Boomsma, D.I., Buring, J.E., Ferrucci, L., Montgomery, G.W., Gudnason, V., Spector, T.D., van Duijn, C.M., Alizadeh, B.Z., Ciullo, M., Crisponi, L., Easton, D.F., Gasparini, P.P., Gieger, C., Harris, T.B., Hayward, C., Kardia, S.L.R., Kraft, P., McKnight, B., Metspalu, A., Morrison, A.C., Reiner, A.P., Ridker, P.M., Rotter, J.I., Toniolo, D., Uitterlinden, A.G., Ulivi, S., Volzke, H., Wareham, N.J., Weir, D.R., Yerges-Armstrong, L.M., Price, A.L., Stefansson, K., Visser, J.A., Ong, K.K., Chang-Claude, J., Murabito, J.M., Perry, J.R.B., and Murray, A. (2015). Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. Nat Genet *47*, 1294-1303.

Day, F.R., Thompson, D.J., Helgason, H., Chasman, D.I., Finucane, H., Sulem, P., Ruth, K.S., Whalen, S., Sarkar, A.K., Albrecht, E., Altmaier, E., Amini, M., Barbieri, C.M., Boutin, T., Campbell, A., Demerath, E., Giri, A., He, C., Hottenga, J.J., Karlsson, R., Kolcic, I., Loh, P.-R., Lunetta, K.L., Mangino, M., Marco, B., McMahon, G., Medland, S.E., Nolte, I.M., Noordam, R., Nutile, T., Paternoster, L., Perjakova, N., Porcu, E., Rose, L.M., Schraut, K.E., Segrè, A.V., Smith, A.V., Stolk, L., Teumer, A., Andrulis, I.L., Bandinelli, S., Beckmann, M.W., Benitez, J., Bergmann, S., Bochud, M., Boerwinkle, E., Bojesen, S.E., Bolla, M.K., Brand, J.S., Brauch, H., Brenner, H., Broer, L., Brüning, T., Buring, J.E., Campbell, H., Catamo, E., Chanock, S., Chenevix-Trench, G., Corre, T., Couch, F.J., Cousminer, D.L., Cox, A., Crisponi, L., Czene, K., Davey Smith, G., de Geus, E.J.C.N., de Mutsert, R., De Vivo, I., Dennis, J., Devilee, P., dos-Santos-Silva, I., Dunning, A.M., Eriksson, J.G., Fasching, P.A., Fernández-Rhodes, L., Ferrucci, L., Flesch-Janys, D., Franke, L., Gabrielson, M., Gandin, I., Giles, G.G., Grallert, H., Gudbjartsson, D.F., Guénel, P., Hall, P., Hallberg, E., Hamann, U., Harris, T.B., Hartman, C.A., Heiss, G., Hooning, M.J., Hopper, J.L., Hu, F., Hunter, D.J., Ikram, M.A., Im, H.K., Järvelin, M.-R., Joshi, P.K., Karasik, D., Kellis, M., Kutalik, Z., LaChance, G., Lambrechts, D., Langenberg, C., Launer, L.J., Laven, J.S.E., Lenarduzzi, S., Li, J., Lind, P.A., Lindstrom, S., Liu, Y., Luan, J.a., Mägi, R., Mannermaa, A., Mbarek, H., McCarthy, M.I., Meisinger, C., Meitinger, T., Menni, C., Metspalu, A., Michailidou, K., Milani, L., Milne, R.L., Montgomery, G.W., Mulligan, A.M., Nalls, M.A., Navarro, P., Nevanlinna, H., Nyholt, D.R., Oldehinkel, A.J., O'Mara, T.A., Padmanabhan, S., Palotie, A., Pedersen, N., Peters, A., Peto, J., Pharoah, P.D.P., Pouta, A., Radice, P., Rahman, I., Ring, S.M., Robino, A., Rosendaal, F.R., Rudan, I., Rueedi, R., Ruggiero, D., Sala, C.F., Schmidt, M.K., Scott, R.A., Shah, M., Sorice, R., Southey, M.C., Sovio, U., Stampfer, M., Steri, M., Strauch, K., Tanaka, T., Tikkanen, E., Timpson, N.J., Traglia, M., Truong, T., Tyrer, J.P., Uitterlinden, A.G., Edwards, D.R.V., Vitart, V., Völker, U., Vollenweider, P., Wang, Q., Widen, E., van Dijk, K.W., Willemsen, G., Winqvist, R., Wolffenbuttel, B.H.R., Zhao, J.H.,

Zoledziewska, M., Zygmunt, M., Alizadeh, B.Z., Boomsma, D.I., Ciullo, M., Cucca, F., Esko, T., Franceschini, N., Gieger, C., Gudnason, V., Hayward, C., Kraft, P., Lawlor, D.A., Magnusson, P.K.E., Martin, N.G., Mook-Kanamori, D.O., Nohr, E.A., Polasek, O., Porteous, D., Price, A.L., Ridker, P.M., Snieder, H., Spector, T.D., Stöckl, D., Toniolo, D., Ulivi, S., Visser, J.A., Völzke, H., Wareham, N.J., Wilson, J.F., Spurdle, A.B., Thorsteindottir, U., Pollard, K.S., Easton, D.F., Tung, J.Y., Chang-Claude, J., Hinds, D., Murray, A., Murabito, J.M., Stefansson, K., Ong, K.K., Perry, J.R.B., The LifeLines Cohort, S., The InterAct, C., kConFab, A.I., Endometrial Cancer Association, C., Ovarian Cancer Association, C., and consortium, P. (2017). Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. Nature Genetics *49*, 834-841.

de Hoon, M.J.L., Imoto, S., Nolan, J., and Miyano, S. (2004). Open source clustering software. Bioinformatics *20*, 1453-1454.

de Lange, K.M., Moutsianas, L., Lee, J.C., Lamb, C.A., Luo, Y., Kennedy, N.A., Jostins, L., Rice, D.L., Gutierrez-Achury, J., Ji, S.-G., Heap, G., Nimmo, E.R., Edwards, C., Henderson, P., Mowat, C., Sanderson, J., Satsangi, J., Simmons, A., Wilson, D.C., Tremelling, M., Hart, A., Mathew, C.G., Newman, W.G., Parkes, M., Lees, C.W., Uhlig, H., Hawkey, C., Prescott, N.J., Ahmad, T., Mansfield, J.C., Anderson, C.A., and Barrett, J.C. (2017). Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nature Genetics *49*, 256-261.

den Hoed, M., Eijgelsheim, M., Esko, T., Brundel, B.J.J.M., Peal, D.S., Evans, D.M., Nolte, I.M., Segrè, A.V., Holm, H., Handsaker, R.E., Westra, H.-J., Johnson, T., Isaacs, A., Yang, J., Lundby, A., Zhao, J.H., Kim, Y.J., Go, M.J., Almgren, P., Bochud, M., Boucher, G., Cornelis, M.C., Gudbjartsson, D., Hadley, D., van der Harst, P., Hayward, C., den Heijer, M., Igl, W., Jackson, A.U., Kutalik, Z., Luan, J.a., Kemp, J.P., Kristiansson, K., Ladenvall, C., Lorentzon, M., Montasser, M.E., Njajou, O.T., O'Reilly, P.F., Padmanabhan, S., St. Pourcain, B., Rankinen, T., Salo, P., Tanaka, T., Timpson, N.J., Vitart, V., Waite, L., Wheeler, W., Zhang, W., Draisma, H.H.M., Feitosa, M.F., Kerr, K.F., Lind, P.A., Mihailov, E., Onland-Moret, N.C., Song, C., Weedon, M.N., Xie, W., Yengo, L., Absher, D., Albert, C.M., Alonso, A., Arking, D.E., de Bakker, P.I.W., Balkau, B., Barlassina, C., Benaglio, P., Bis, J.C., Bouatia-Naji, N., Brage, S., Chanock, S.J., Chines, P.S., Chung, M., Darbar, D., Dina, C., Dörr, M., Elliott, P., Felix, S.B., Fischer, K., Fuchsberger, C., de Geus, E.J.C., Goyette, P., Gudnason, V., Harris, T.B., Hartikainen, A.-L., Havulinna, A.S., Heckbert, S.R., Hicks, A.A., Hofman, A., Holewijn, S., Hoogstra-Berends, F., Hottenga, J.-J., Jensen, M.K., Johansson, Å., Junttila, J., Kääb, S., Kanon, B., Ketkar, S., Khaw, K.-T., Knowles, J.W., Kooner, A.S., Kors, J.A., Kumari, M., Milani, L., Laiho, P., Lakatta, E.G., Langenberg, C., Leusink, M., Liu, Y., Luben, R.N., Lunetta, K.L., Lynch, S.N., Markus, M.R.P., Marques-Vidal, P., Leach, I.M., McArdle, W.L., McCarroll, S.A., Medland, S.E., Miller, K.A., Montgomery, G.W., Morrison, A.C., Müller-Nurasyid, M., Navarro, P., Nelis, M., O'Connell, J.R., O'Donnell, C.J., Ong, K.K., Newman, A.B., Peters, A., Polasek, O., Pouta, A., Pramstaller, P.P., Psaty, B.M., Rao, D.C., Ring, S.M., Rossin, E.J., Rudan, D., Sanna, S., Scott, R.A., Sehmi, J.S., Sharp, S., Shin, J.T., Singleton, A.B., Smith, A.V., Soranzo, N., Spector, T.D., Stewart, C., Stringham, H.M., Tarasov, K.V., Uitterlinden, A.G., Vandenput, L., Hwang, S.-J., Whitfield, J.B., Wijmenga, C., Wild, S.H., Willemsen, G., Wilson, J.F., Witteman, J.C.M., Wong, A., Wong, Q., Jamshidi, Y., Zitting, P., Boer, J.M.A., Boomsma, D.I., Borecki, I.B., van Duijn, C.M., Ekelund, U., Forouhi, N.G., Froguel, P., Hingorani, A., Ingelsson, E., Kivimaki, M., Kronmal, R.A., Kuh,

D., Lind, L., Martin, N.G., Oostra, B.A., Pedersen, N.L., Quertermous, T., Rotter, J.I., van der Schouw, Y.T., Verschuren, W.M.M., Walker, M., Albanes, D., Arnar, D.O., Assimes, T.L., Bandinelli, S., Boehnke, M., de Boer, R.A., Bouchard, C., Caulfield, W.L.M., Chambers, J.C., Curhan, G., Cusi, D., Eriksson, J., Ferrucci, L., van Gilst, W.H., Glorioso, N., de Graaf, J., Groop, L., Gyllensten, U., Hsueh, W.-C., Hu, F.B., Huikuri, H.V., Hunter, D.J., Iribarren, C., Isomaa, B., Jarvelin, M.-R., Jula, A., Kähönen, M., Kiemeney, L.A., van der Klauw, M.M., Kooner, J.S., Kraft, P., Iacoviello, L., Lehtimäki, T., Lokki, M.-L.L., Mitchell, B.D., Navis, G., Nieminen, M.S., Ohlsson, C., Poulter, N.R., Qi, L., Raitakari, O.T., Rimm, E.B., Rioux, J.D., Rizzi, F., Rudan, I., Salomaa, V., Sever, P.S., Shields, D.C., Shuldiner, A.R., Sinisalo, J., Stanton, A.V., Stolk, R.P., Strachan, D.P., Tardif, J.-C., Thorsteinsdottir, U., Tuomilehto, J., van Veldhuisen, D.J., Virtamo, J., Viikari, J., Vollenweider, P., Waeber, G., Widen, E., Cho, Y.S., Olsen, J.V., Visscher, P.M., Willer, C., Franke, L., Erdmann, J., Thompson, J.R., Pfeufer, A., Sotoodehnia, N., Newton-Cheh, C., Ellinor, P.T., Stricker, B.H.C., Metspalu, A., Perola, M., Beckmann, J.S., Smith, G.D., Stefansson, K., Wareham, N.J., Munroe, P.B., Sibon, O.C.M., Milan, D.J., Snieder, H., Samani, N.J., Loos, R.J.F., Global, B.C., Consortium, C.A., Consortium, P.G., Consortium, Q.G., Consortium, Q.-I., and Consortium, C.-A. (2013). Identification of heart rate–associated loci and their effects on cardiac conduction and rhythm disorders. Nature Genetics *45*, 621-631.

Deutsch, M.A., Doppler, S.A., Li, X., Lahm, H., Santamaria, G., Cuda, G., Eichhorn, S., Ratschiller, T., Dzilic, E., Dressen, M., Eckart, A., Stark, K., Massberg, S., Bartels, A., Rischpler, C., Gilsbach, R., Hein, L., Fleischmann, B.K., Wu, S.M., Lange, R., and Krane, M. (2018). Reactivation of the Nkx2.5 cardiac enhancer after myocardial infarction does not presage myogenesis. Cardiovasc Res *114*, 1098-1114.

Dickel, D.E., Barozzi, I., Zhu, Y., Fukuda-Yuzawa, Y., Osterwalder, M., Mannion, B.J., May, D., Spurrell, C.H., Plajzer-Frick, I., Pickle, C.S., Lee, E., Garvin, T.H., Kato, M., Akiyama, J.A., Afzal, V., Lee, A.Y., Gorkin, D.U., Ren, B., Rubin, E.M., Visel, A., and Pennacchio, L.A. (2016). Genome-wide compendium and functional assessment of in vivo heart enhancers. Nature communications *7*, 12923-12923.

Dirkx, E., da Costa Martins, P.A., and De Windt, L.J. (2013). Regulation of fetal gene expression in heart failure. Biochim Biophys Acta *1832*, 2414-2424.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15-21.

Domcke, S., Hill, A.J., Daza, R.M., Cao, J., O'Day, D.R., Pliner, H.A., Aldinger, K.A., Pokholok, D., Zhang, F., Milbank, J.H., Zager, M.A., Glass, I.A., Steemers, F.J., Doherty, D., Trapnell, C., Cusanovich, D.A., and Shendure, J. (2020). A human cell atlas of fetal chromatin accessibility. Science *370*, eaba7612.

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., Peterson, R., and Domingue, B. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. Nature Communications *10*, 3328.

Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., Ku, M., Durham, T., Kellis, M., and Bernstein, B.E. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43-49.

Fabian Pedregosa, G.V., Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research *12(85)*, 2825−2830.

Fabriek, B.O., Dijkstra, C.D., and van den Berg, T.K. (2005). The macrophage scavenger receptor CD163. Immunobiology *210*, 153-160.

Fan, Z., and Guan, J. (2016). Antifibrotic therapies to control cardiac fibrosis. Biomater Res *20*, 13.

Fang, R., Preissl, S., Li, Y., Hou, X., Lucero, J., Wang, X., Motamedi, A., Shiau, A.K., Zhou, X., Xie, F., Mukamel, E.A., Zhang, K., Zhang, Y., Behrens, M.M., Ecker, J.R., and Ren, B. (2021). Comprehensive analysis of single cell ATAC-seq data with SnapATAC. Nat Commun *12*, 1337.

Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., Ripke, S., Day, F.R., Purcell, S., Stahl, E., Lindstrom, S., Perry, J.R.B., Okada, Y., Raychaudhuri, S., Daly, M.J., Patterson, N., Neale, B.M., Price, A.L., ReproGen, C., Schizophrenia Working Group of the Psychiatric Genomics, C., and The, R.C. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nature Genetics *47*, 1228-1235.

Fotsing, S.F., Margoliash, J., Wang, C., Saini, S., Yanicky, R., Shleizer-Burko, S., Goren, A., and Gymrek, M. (2019). The impact of short tandem repeat variation on gene expression. Nature Genetics *51*, 1652-1659.

Frankish, A., Diekhans, M., Ferreira, A.M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I.T., Garcia Giron, C., Gonzalez, J.M., Grego, T., Hardy, M., Hourlier, T., Hunt, T., Izuogu, O.G., Lagarde, J., Martin, F.J., Martinez, L., Mohanan, S., Muir, P., Navarro, F.C.P., Parker, A., Pei, B., Pozo, F., Ruffier, M., Schmitt, B.M., Stapleton, E., Suner, M.M., Sycheva, I., Uszczynska-Ratajczak, B., Xu, J., Yates, A., Zerbino, D., Zhang, Y., Aken, B., Choudhary, J.S., Gerstein, M., Guigo, R., Hubbard, T.J.P., Kellis, M., Paten, B., Reymond, A., Tress, M.L., and Flicek, P. (2019). GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res *47*, D766-D773.

Franzén, O., Gan, L.-M., and Björkegren, J. (2019). PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. Database The Journal of Biological Databases and Curation *2019*, 46.

Fulco, C.P., Nasser, J., Jones, T.R., Munson, G., Bergman, D.T., Subramanian, V., Grossman, S.R., Anyoha, R., Doughty, B.R., Patwardhan, T.A., Nguyen, T.H., Kane, M., Perez, E.M., Durand, N.C., Lareau, C.A., Stamenova, E.K., Aiden, E.L., Lander, E.S., and Engreitz, J.M.

(2019). Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. Nature Genetics *51*, 1664-1669.

Furtado, M.B., Costa, M.W., Pranoto, E.A., Salimova, E., Pinto, A.R., Lam, N.T., Park, A., Snider, P., Chandran, A., Harvey, R.P., Boyd, R., Conway, S.J., Pearson, J., Kaye, D.M., and Rosenthal, N.A. (2014). Cardiogenic genes expressed in cardiac fibroblasts contribute to heart development and repair. Circulation research *114*, 1422-1434.

Gilsbach, R., Preissl, S., Gruning, B.A., Schnick, T., Burger, L., Benes, V., Wurch, A., Bonisch, U., Gunther, S., Backofen, R., Fleischmann, B.K., Schubeler, D., and Hein, L. (2014). Dynamic DNA methylation orchestrates cardiomyocyte development, maturation and disease. Nat Commun *5*, 5288.

Gilsbach, R., Schwaderer, M., Preissl, S., Gruning, B.A., Kranzhofer, D., Schneider, P., Nuhrenberg, T.G., Mulero-Navarro, S., Weichenhan, D., Braun, C., Dressen, M., Jacobs, A.R., Lahm, H., Doenst, T., Backofen, R., Krane, M., Gelb, B.D., and Hein, L. (2018). Distinct epigenetic programs regulate cardiac myocyte development and disease in the human heart in vivo. Nat Commun *9*, 391.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics *27*, 1017-1018.

Grosselin, K., Durand, A., Marsolier, J., Poitou, A., Marangoni, E., Nemati, F., Dahmani, A., Lameiras, S., Reyal, F., Frenoy, O., Pousse, Y., Reichen, M., Woolfe, A., Brenan, C., Griffiths, A.D., Vallot, C., and Gerard, A. (2019). High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. Nat Genet *51*, 1060-1066.

Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R., Awashti, S., Belliveau, R., Bettella, F., Buxbaum, J.D., Bybjerg-Grauholm, J., Bækvad-Hansen, M., Cerrato, F., Chambert, K., Christensen, J.H., Churchhouse, C., Dellenvall, K., Demontis, D., De Rubeis, S., Devlin, B., Djurovic, S., Dumont, A.L., Goldstein, J.I., Hansen, C.S., Hauberg, M.E., Hollegaard, M.V., Hope, S., Howrigan, D.P., Huang, H., Hultman, C.M., Klei, L., Maller, J., Martin, J., Martin, A.R., Moran, J.L., Nyegaard, M., Nærland, T., Palmer, D.S., Palotie, A., Pedersen, C.B., Pedersen, M.G., dPoterba, T., Poulsen, J.B., Pourcain, B.S., Qvist, P., Rehnström, K., Reichenberg, A., Reichert, J., Robinson, E.B., Roeder, K., Roussos, P., Saemundsen, E., Sandin, S., Satterstrom, F.K., Davey Smith, G., Stefansson, H., Steinberg, S., Stevens, C.R., Sullivan, P.F., Turley, P., Walters, G.B., Xu, X., Wray, N.R., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Air, T.M., Andlauer, T.F.M., Bacanu, S.-A., Beekman, A.T.F., Bigdeli, T.B., Binder, E.B., Blackwood, D.H.R., Bryois, J., Buttenschøn, H.N., Cai, N., Castelao, E., Clarke, T.-K., Coleman, J.R.I., Colodro-Conde, L., Couvy-Duchesne, B., Craddock, N., Crawford, G.E., Davies, G., Deary, I.J., Degenhardt, F., Derks, E.M., Direk, N., Dolan, C.V., Dunn, E.C., Eley, T.C., Escott-Price, V., Kiadeh, F.F.H., Finucane, H.K., Forstner, A.J., Frank, J., Gaspar, H.A., Gill, M., Goes, F.S., Gordon, S.D., Hall, L.S., Hansen, T.F., Herms, S., Hickie, I.B., Hoffmann, P., Homuth, G., Horn, C., Hottenga, J.-J., Ising, M., Jansen, R., Jorgenson, E., Knowles, J.A., Kohane, I.S., Kraft, J., Kretzschmar, W.W., Krogh, J., Kutalik, Z., Li, Y., Lind, P.A., MacIntyre, D.J., MacKinnon, D.F., Maier, R.M., Maier, W., Marchini, J., Mbarek, H., McGrath, P., McGuffin, P., Medland, S.E., Mehta, D., Middeldorp,

C.M., Mihailov, E., Milaneschi, Y., Milani, L., Mondimore, F.M., Montgomery, G.W., Mostafavi, S., Mullins, N., Nauck, M., Ng, B., Nivard, M.G., Nyholt, D.R., O'Reilly, P.F., Oskarsson, H., Owen, M.J., Painter, J.N., Peterson, R.E., Pettersson, E., Peyrot, W.J., Pistis, G., Posthuma, D., Quiroz, J.A., Rice, J.P., Riley, B.P., Rivera, M., Mirza, S.S., Schoevers, R., Schulte, E.C., Shen, L., Shi, J., Shyn, S.I., Sigurdsson, E., Sinnamon, G.C.B., Smit, J.H., Smith, D.J., Streit, F., Strohmaier, J., Tansey, K.E., Teismann, H., Teumer, A., Thompson, W., Thomson, P.A., Thorgeirsson, T.E., Traylor, M., Treutlein, J., Trubetskoy, V., Uitterlinden, A.G., Umbricht, D., Van der Auwera, S., van Hemert, A.M., Viktorin, A., Visscher, P.M., Wang, Y., Webb, B.T., Weinsheimer, S.M., Wellmann, J., Willemsen, G., Witt, S.H., Wu, Y., Xi, H.S., Yang, J., Zhang, F., Arolt, V., Baune, B.T., Berger, K., Boomsma, D.I., Cichon, S., Dannlowski, U., de Geus, E.J.C., DePaulo, J.R., Domenici, E., Domschke, K., Esko, T., Grabe, H.J., Hamilton, S.P., Hayward, C., Heath, A.C., Kendler, K.S., Kloiber, S., Lewis, G., Li, Q.S., Lucae, S., Madden, P.A.F., Magnusson, P.K., Martin, N.G., McIntosh, A.M., Metspalu, A., Müller-Myhsok, B., Nöthen, M.M., O'Donovan, M.C., Paciga, S.A., Pedersen, N.L., Penninx, B.W.J.H., Perlis, R.H., Porteous, D.J., Potash, J.B., Preisig, M., Rietschel, M., Schaefer, C., Schulze, T.G., Smoller, J.W., Tiemeier, H., Uher, R., Völzke, H., Weissman, M.M., Lewis, C.M., Levinson, D.F., Breen, G., Agee, M., Alipanahi, B., Auton, A., Bell, R.K., Bryc, K., Elson, S.L., Fontanillas, P., Furlotte, N.A., Hromatka, B.S., Huber, K.E., Kleinman, A., Litterman, N.K., McIntyre, M.H., Mountain, J.L., Noblin, E.S., Northover, C.A.M., Pitts, S.J., Sathirapongsasuti, J.F., Sazonova, O.V., Shelton, J.F., Shringarpure, S., Tung, J.Y., Vacic, V., Wilson, C.H., Stefansson, K., Geschwind, D.H., Nordentoft, M., Hougaard, D.M., Werge, T., Mors, O., Mortensen, P.B., Neale, B.M., Daly, M.J., Børglum, A.D., Autism Spectrum Disorder Working Group of the Psychiatric Genomics, C., Bupgen, Major Depressive Disorder Working Group of the Psychiatric Genomics, C., and andMe Research, T. (2019). Identification of common genetic risk variants for autism spectrum disorder. Nature Genetics *51*, 431-444.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol *52*, 696-704.

Gymrek, M., Willems, T., Guilmatre, A., Zeng, H., Markus, B., Georgiev, S., Daly, M.J., Price, A.L., Pritchard, J.K., Sharp, A.J., and Erlich, Y. (2016). Abundant contribution of short tandem repeats to gene expression variation in humans. Nature Genetics *48*, 22-29.

Haeussler, M., Schönig, K., Eckert, H., Eschstruth, A., Mianné, J., Renaud, J.-B., Schneider-Maunoury, S., Shkumatava, A., Teboul, L., Kent, J., Joly, J.-S., and Concordet, J.-P. (2016). Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. Genome Biology *17*, 148.

Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol *36*, 421-427.

Harper, A.R., Goel, A., Grace, C., Thomson, K.L., Petersen, S.E., Xu, X., Waring, A., Ormondroyd, E., Kramer, C.M., Ho, C.Y., Neubauer, S., Tadros, R., Ware, J.S., Bezzina, C.R., Farrall, M., and Watkins, H. (2021). Common genetic variants and modifiable risk factors underpin hypertrophic cardiomyopathy susceptibility and expressivity. Nat Genet *53*, 135-142.

Harrow, J., Denoeud, F., Frankish, A., Reymond, A., Chen, C.K., Chrast, J., Lagarde, J., Gilbert, J.G., Storey, R., Swarbreck, D., Rossier, C., Ucla, C., Hubbard, T., Antonarakis, S.E., and Guigo, R. (2006). GENCODE: producing a reference annotation for ENCODE. Genome Biol *7 Suppl 1*, S4 1-9.

Heidenreich, P.A., Fonarow, G.C., Opsha, Y., Sandhu, A.T., Sweitzer, N.K., and Warraich, H.J. (2022). Economic Issues in Heart Failure in the United States. Journal of Cardiac Failure *0*.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.

Hocker, J.D., Poirion, O.B., Zhu, F., Buchanan, J., Zhang, K., Chiou, J., Wang, T.-M., Zhang, Q., Hou, X., Li, Y.E., Zhang, Y., Farah, E.N., Wang, A., McCulloch, A.D., Gaulton, K.J., Ren, B., Chi, N.C., and Preissl, S. (2021). Cardiac Cell Type-Specific Gene Regulatory Programs and Disease Risk Association. Science Advances *7*.

Hoffmann, T.J., Theusch, E., Haldar, T., Ranatunga, D.K., Jorgenson, E., Medina, M.W., Kvale, M.N., Kwok, P.Y., Schaefer, C., Krauss, R.M., Iribarren, C., and Risch, N. (2018). A large electronic-health-record-based genome-wide study of serum lipids. Nat Genet *50*, 401-413.

Hook, P.W., and McCallion, A.S. (2020). Leveraging mouse chromatin data for heritability enrichment informs common disease architecture and reveals cortical layer contributions to schizophrenia. Genome research.

Horikoshi, M., Beaumont, R.N., Day, F.R., Warrington, N.M., Kooijman, M.N., Fernandez-Tajes, J., Feenstra, B., van Zuydam, N.R., Gaulton, K.J., Grarup, N., Bradfield, J.P., Strachan, D.P., Li-Gao, R., Ahluwalia, T.S., Kreiner, E., Rueedi, R., Lyytikainen, L.P., Cousminer, D.L., Wu, Y., Thiering, E., Wang, C.A., Have, C.T., Hottenga, J.J., Vilor-Tejedor, N., Joshi, P.K., Boh, E.T.H., Ntalla, I., Pitkanen, N., Mahajan, A., van Leeuwen, E.M., Joro, R., Lagou, V., Nodzenski, M., Diver, L.A., Zondervan, K.T., Bustamante, M., Marques-Vidal, P., Mercader, J.M., Bennett, A.J., Rahmioglu, N., Nyholt, D.R., Ma, R.C.W., Tam, C.H.T., Tam, W.H., Ganesh, S.K., van Rooij, F.J., Jones, S.E., Loh, P.R., Ruth, K.S., Tuke, M.A., Tyrrell, J., Wood, A.R., Yaghootkar, H., Scholtens, D.M., Paternoster, L., Prokopenko, I., Kovacs, P., Atalay, M., Willems, S.M., Panoutsopoulou, K., Wang, X., Carstensen, L., Geller, F., Schraut, K.E., Murcia, M., van Beijsterveldt, C.E., Willemsen, G., Appel, E.V.R., Fonvig, C.E., Trier, C., Tiesler, C.M., Standl, M., Kutalik, Z., Bonas-Guarch, S., Hougaard, D.M., Sanchez, F., Torrents, D., Waage, J., Hollegaard, M.V., de Haan, H.G., Rosendaal, F.R., Medina-Gomez, C., Ring, S.M., Hemani, G., McMahon, G., Robertson, N.R., Groves, C.J., Langenberg, C., Luan, J., Scott, R.A., Zhao, J.H., Mentch, F.D., MacKenzie, S.M., Reynolds, R.M., Lowe, W.L., Jr., Tonjes, A., Stumvoll, M., Lindi, V., Lakka, T.A., van Duijn, C.M., Kiess, W., Korner, A., Sorensen, T.I., Niinikoski, H., Pahkala, K., Raitakari, O.T., Zeggini, E., Dedoussis, G.V., Teo, Y.Y., Saw, S.M., Melbye, M., Campbell, H., Wilson, J.F., Vrijheid, M., de Geus, E.J., Boomsma, D.I., Kadarmideen, H.N., Holm, J.C., Hansen, T., Sebert, S., Hattersley, A.T., Beilin, L.J., Newnham, J.P., Pennell, C.E., Heinrich, J., Adair, L.S., Borja, J.B., Mohlke, K.L., Eriksson, J.G., Widen, E.E., Kahonen, M., Viikari, J.S., Lehtimaki, T., Vollenweider, P., Bonnelykke, K., Bisgaard, H., Mook-Kanamori,

D.O., Hofman, A., Rivadeneira, F., Uitterlinden, A.G., Pisinger, C., Pedersen, O., Power, C., Hypponen, E., Wareham, N.J., Hakonarson, H., Davies, E., Walker, B.R., Jaddoe, V.W., Jarvelin, M.R., Grant, S.F., Vaag, A.A., Lawlor, D.A., Frayling, T.M., Davey Smith, G., Morris, A.P., Ong, K.K., Felix, J.F., Timpson, N.J., Perry, J.R., Evans, D.M., McCarthy, M.I., and Freathy, R.M. (2016). Genome-wide associations for birth weight and correlations with adult disease. Nature *538*, 248-252.

Hortells, L., Johansen, A.K.Z., and Yutzey, K.E. (2019). Cardiac Fibroblasts and the Extracellular Matrix in Regenerative and Nonregenerative Hearts. J Cardiovasc Dev Dis *6*.

Hu, E., Liang, P., and Spiegelman, B.M. (1996). AdipoQ Is a Novel Adipose-specific Gene Dysregulated in Obesity. Journal of Biological Chemistry *271*, 10697-10703.

Institute, B. (2019). Picard Toolkit. Broad Institute, GitHub Repository.

Ishigaki, K., Akiyama, M., Kanai, M., Takahashi, A., Kawakami, E., Sugishita, H., Sakaue, S., Matoba, N., Low, S.K., Okada, Y., Terao, C., Amariuta, T., Gazal, S., Kochi, Y., Horikoshi, M., Suzuki, K., Ito, K., Koyama, S., Ozaki, K., Niida, S., Sakata, Y., Sakata, Y., Kohno, T., Shiraishi, K., Momozawa, Y., Hirata, M., Matsuda, K., Ikeda, M., Iwata, N., Ikegawa, S., Kou, I., Tanaka, T., Nakagawa, H., Suzuki, A., Hirota, T., Tamari, M., Chayama, K., Miki, D., Mori, M., Nagayama, S., Daigo, Y., Miki, Y., Katagiri, T., Ogawa, O., Obara, W., Ito, H., Yoshida, T., Imoto, I., Takahashi, T., Tanikawa, C., Suzuki, T., Sinozaki, N., Minami, S., Yamaguchi, H., Asai, S., Takahashi, Y., Yamaji, K., Takahashi, K., Fujioka, T., Takata, R., Yanai, H., Masumoto, A., Koretsune, Y., Kutsumi, H., Higashiyama, M., Murayama, S., Minegishi, N., Suzuki, K., Tanno, K., Shimizu, A., Yamaji, T., Iwasaki, M., Sawada, N., Uemura, H., Tanaka, K., Naito, M., Sasaki, M., Wakai, K., Tsugane, S., Yamamoto, M., Yamamoto, K., Murakami, Y., Nakamura, Y., Raychaudhuri, S., Inazawa, J., Yamauchi, T., Kadowaki, T., Kubo, M., and Kamatani, Y. (2020). Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. Nat Genet.

Jang, S., Collin de l'Hortet, A., and Soto-Gutierrez, A. (2019). Induced Pluripotent Stem Cell-Derived Endothelial Cells: Overview, Current Advances, Applications, and Future Directions. Am J Pathol *189*, 502-512.

Jansen, I.E., Savage, J.E., Watanabe, K., Bryois, J., Williams, D.M., Steinberg, S., Sealock, J., Karlsson, I.K., Hägg, S., Athanasiu, L., Voyle, N., Proitsi, P., Witoelar, A., Stringer, S., Aarsland, D., Almdahl, I.S., Andersen, F., Bergh, S., Bettella, F., Bjornsson, S., Brækhus, A., Bråthen, G., de Leeuw, C., Desikan, R.S., Djurovic, S., Dumitrescu, L., Fladby, T., Hohman, T.J., Jonsson, P.V., Kiddle, S.J., Rongve, A., Saltvedt, I., Sando, S.B., Selbæk, G., Shoai, M., Skene, N.G., Snaedal, J., Stordal, E., Ulstein, I.D., Wang, Y., White, L.R., Hardy, J., Hjerling-Leffler, J., Sullivan, P.F., van der Flier, W.M., Dobson, R., Davis, L.K., Stefansson, H., Stefansson, K., Pedersen, N.L., Ripke, S., Andreassen, O.A., and Posthuma, D. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat Genet *51*, 404-413.

Jedlicka, P., Sui, X., Sussel, L., and Gutierrez-Hartmann, A. (2009). Ets transcription factors control epithelial maturation and transit and crypt-villus morphogenesis in the mammalian intestine. Am J Pathol *174*, 1280-1290.

Ji, S.G., Juran, B.D., Mucha, S., Folseraas, T., Jostins, L., Melum, E., Kumasaka, N., Atkinson, E.J., Schlicht, E.M., Liu, J.Z., Shah, T., Gutierrez-Achury, J., Boberg, K.M., Bergquist, A., Vermeire, S., Eksteen, B., Durie, P.R., Farkkila, M., Müller, T., Schramm, C., Sterneck, M., Weismüller, T.J., Gotthardt, D.N., Ellinghaus, D., Braun, F., Teufel, A., Laudes, M., Lieb, W., Jacobs, G., Beuers, U., Weersma, R.K., Wijmenga, C., Marschall, H.U., Milkiewicz, P., Pares, A., Kontula, K., Chazouillères, O., Invernizzi, P., Goode, E., Spiess, K., Moore, C., Sambrook, J., Ouwehand, W.H., Roberts, D.J., Danesh, J., Floreani, A., Gulamhusein, A.F., Eaton, J.E., Schreiber, S., Coltescu, C., Bowlus, C.L., Luketic, V.A., Odin, J.A., Chopra, K.B., Kowdley, K.V., Chalasani, N., Manns, M.P., Srivastava, B., Mells, G., Sandford, R.N., Alexander, G., Gaffney, D.J., Chapman, R.W., Hirschfield, G.M., de Andrade, M., Rushbrook, S.M., Franke, A., Karlsen, T.H., Lazaridis, K.N., and Anderson, C.A. (2017). Genome-wide association study of primary sclerosing cholangitis identifies new risk loci and quantifies the genetic relationship with inflammatory bowel disease. Nat Genet *49*, 269-273.

Jia, G., Preussner, J., Chen, X., Guenther, S., Yuan, X., Yekelchyk, M., Kuenne, C., Looso, M., Zhou, Y., Teichmann, S., and Braun, T. (2018). Single cell RNA-seq and ATAC-seq analysis of cardiac progenitor cell transition states and lineage settlement. Nat Commun *9*, 4877.

Jiang, X., O'Reilly, P.F., Aschard, H., Hsu, Y.H., Richards, J.B., Dupuis, J., Ingelsson, E., Karasik, D., Pilz, S., Berry, D., Kestenbaum, B., Zheng, J., Luan, J., Sofianopoulou, E., Streeten, E.A., Albanes, D., Lutsey, P.L., Yao, L., Tang, W., Econs, M.J., Wallaschofski, H., Volzke, H., Zhou, A., Power, C., McCarthy, M.I., Michos, E.D., Boerwinkle, E., Weinstein, S.J., Freedman, N.D., Huang, W.Y., Van Schoor, N.M., van der Velde, N., Groot, L., Enneman, A., Cupples, L.A., Booth, S.L., Vasan, R.S., Liu, C.T., Zhou, Y., Ripatti, S., Ohlsson, C., Vandenput, L., Lorentzon, M., Eriksson, J.G., Shea, M.K., Houston, D.K., Kritchevsky, S.B., Liu, Y., Lohman, K.K., Ferrucci, L., Peacock, M., Gieger, C., Beekman, M., Slagboom, E., Deelen, J., Heemst, D.V., Kleber, M.E., Marz, W., de Boer, I.H., Wood, A.C., Rotter, J.I., Rich, S.S., Robinson-Cohen, C., den Heijer, M., Jarvelin, M.R., Cavadino, A., Joshi, P.K., Wilson, J.F., Hayward, C., Lind, L., Michaelsson, K., Trompet, S., Zillikens, M.C., Uitterlinden, A.G., Rivadeneira, F., Broer, L., Zgaga, L., Campbell, H., Theodoratou, E., Farrington, S.M., Timofeeva, M., Dunlop, M.G., Valdes, A.M., Tikkanen, E., Lehtimaki, T., Lyytikainen, L.P., Kahonen, M., Raitakari, O.T., Mikkila, V., Ikram, M.A., Sattar, N., Jukema, J.W., Wareham, N.J., Langenberg, C., Forouhi, N.G., Gundersen, T.E., Khaw, K.T., Butterworth, A.S., Danesh, J., Spector, T., Wang, T.J., Hypponen, E., Kraft, P., and Kiel, D.P. (2018). Genome-wide association study in 79,366 European-ancestry individuals informs the genetic architecture of 25-hydroxyvitamin D levels. Nat Commun *9*, 260.

Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nat Commun *12*, 1088.

Jin, Y., Andersen, G., Yorgov, D., Ferrara, T.M., Ben, S., Brownson, K.M., Holland, P.J., Birlea, S.A., Siebert, J., Hartmann, A., Lienert, A., van Geel, N., Lambert, J., Luiten, R.M., Wolkerstorfer,

A., Wietze van der Veen, J.P., Bennett, D.C., Taïeb, A., Ezzedine, K., Kemp, E.H., Gawkrodger, D.J., Weetman, A.P., Kõks, S., Prans, E., Kingo, K., Karelson, M., Wallace, M.R., McCormack, W.T., Overbeck, A., Moretti, S., Colucci, R., Picardo, M., Silverberg, N.B., Olsson, M., Valle, Y., Korobko, I., Böhm, M., Lim, H.W., Hamzavi, I., Zhou, L., Mi, Q.S., Fain, P.R., Santorico, S.A., and Spritz, R.A. (2016). Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. Nat Genet *48*, 1418-1424.

John, S., Sabo, P.J., Canfield, T.K., Lee, K., Vong, S., Weaver, M., Wang, H., Vierstra, J., Reynolds, A.P., Thurman, R.E., and Stamatoyannopoulos, J.A. (2013). Genome-Scale Mapping of DNase I Hypersensitivity. Current Protocols in Molecular Biology *103*, 21.27.21-21.27.20.

Jones, D.K., Liu, F., Vaidyanathan, R., Eckhardt, L.L., Trudeau, M.C., and Robertson, G.A. (2014). hERG 1b is critical for human cardiac repolarization. Proc Natl Acad Sci U S A *111*, 18073-18077.

Jung, I., Schmitt, A., Diao, Y., Lee, A.J., Liu, T., Yang, D., Tan, C., Eom, J., Chan, M., Chee, S., Chiang, Z., Kim, C., Masliah, E., Barr, C.L., Li, B., Kuan, S., Kim, D., and Ren, B. (2019). A compendium of promoter-centered long-range chromatin interactions in the human genome. Nature genetics *51*, 1442-1449.

Kalucka, J., de Rooij, L.P.M.H., Goveia, J., Rohlenova, K., Dumas, S.J., Meta, E., Conchinha, N.V., Taverna, F., Teuwen, L.-A., Veys, K., García-Caballero, M., Khan, S., Geldhof, V., Sokol, L., Chen, R., Treps, L., Borri, M., de Zeeuw, P., Dubois, C., Karakach, T.K., Falkenberg, K.D., Parys, M., Yin, X., Vinckier, S., Du, Y., Fenton, R.A., Schoonjans, L., Dewerchin, M., Eelen, G., Thienpont, B., Lin, L., Bolund, L., Li, X., Luo, Y., and Carmeliet, P. (2020). Single-Cell Transcriptome Atlas of Murine Endothelial Cells. Cell *180*, 764-779.e720.

Karakikes, I., Ameen, M., Termglinchan, V., and Wu, J.C. (2015). Human induced pluripotent stem cell-derived cardiomyocytes: insights into molecular, cellular, and functional phenotypes. Circ Res *117*, 80-88.

Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun *10*, 1930.

Kemp, J.P., Morris, J.A., Medina-Gomez, C., Forgetta, V., Warrington, N.M., Youlten, S.E., Zheng, J., Gregson, C.L., Grundberg, E., Trajanoska, K., Logan, J.G., Pollard, A.S., Sparkes, P.C., Ghirardello, E.J., Allen, R., Leitch, V.D., Butterfield, N.C., Komla-Ebri, D., Adoum, A.T., Curry, K.F., White, J.K., Kussy, F., Greenlaw, K.M., Xu, C., Harvey, N.C., Cooper, C., Adams, D.J., Greenwood, C.M.T., Maurano, M.T., Kaptoge, S., Rivadeneira, F., Tobias, J.H., Croucher, P.I., Ackert-Bicknell, C.L., Bassett, J.H.D., Williams, G.R., Richards, J.B., and Evans, D.M. (2017). Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. Nat Genet *49*, 1468-1475.

Kilpeläinen, T.O., Carli, J.F., Skowronski, A.A., Sun, Q., Kriebel, J., Feitosa, M.F., Hedman Å, K., Drong, A.W., Hayes, J.E., Zhao, J., Pers, T.H., Schick, U., Grarup, N., Kutalik, Z., Trompet, S., Mangino, M., Kristiansson, K., Beekman, M., Lyytikäinen, L.P., Eriksson, J., Henneman, P., Lahti, J., Tanaka, T., Luan, J., Del Greco, M.F., Pasko, D., Renström, F., Willems, S.M., Mahajan,

A., Rose, L.M., Guo, X., Liu, Y., Kleber, M.E., Pérusse, L., Gaunt, T., Ahluwalia, T.S., Ju Sung, Y., Ramos, Y.F., Amin, N., Amuzu, A., Barroso, I., Bellis, C., Blangero, J., Buckley, B.M., Böhringer, S., YD, I.C., de Craen, A.J., Crosslin, D.R., Dale, C.E., Dastani, Z., Day, F.R., Deelen, J., Delgado, G.E., Demirkan, A., Finucane, F.M., Ford, I., Garcia, M.E., Gieger, C., Gustafsson, S., Hallmans, G., Hankinson, S.E., Havulinna, A.S., Herder, C., Hernandez, D., Hicks, A.A., Hunter, D.J., Illig, T., Ingelsson, E., Ioan-Facsinay, A., Jansson, J.O., Jenny, N.S., Jørgensen, M.E., Jørgensen, T., Karlsson, M., Koenig, W., Kraft, P., Kwekkeboom, J., Laatikainen, T., Ladwig, K.H., LeDuc, C.A., Lowe, G., Lu, Y., Marques-Vidal, P., Meisinger, C., Menni, C., Morris, A.P., Myers, R.H., Männistö, S., Nalls, M.A., Paternoster, L., Peters, A., Pradhan, A.D., Rankinen, T., Rasmussen-Torvik, L.J., Rathmann, W., Rice, T.K., Brent Richards, J., Ridker, P.M., Sattar, N., Savage, D.B., Söderberg, S., Timpson, N.J., Vandenput, L., van Heemst, D., Uh, H.W., Vohl, M.C., Walker, M., Wichmann, H.E., Widén, E., Wood, A.R., Yao, J., Zeller, T., Zhang, Y., Meulenbelt, I., Kloppenburg, M., Astrup, A., Sørensen, T.I., Sarzynski, M.A., Rao, D.C., Jousilahti, P., Vartiainen, E., Hofman, A., Rivadeneira, F., Uitterlinden, A.G., Kajantie, E., Osmond, C., Palotie, A., Eriksson, J.G., Heliövaara, M., Knekt, P.B., Koskinen, S., Jula, A., Perola, M., Huupponen, R.K., Viikari, J.S., Kähönen, M., Lehtimäki, T., Raitakari, O.T., Mellström, D., Lorentzon, M., Casas, J.P., Bandinelli, S., März, W., Isaacs, A., van Dijk, K.W., van Duijn, C.M., Harris, T.B., Bouchard, C., Allison, M.A., Chasman, D.I., Ohlsson, C., Lind, L., Scott, R.A., Langenberg, C., Wareham, N.J., Ferrucci, L., Frayling, T.M., Pramstaller, P.P., Borecki, I.B., Waterworth, D.M., Bergmann, S., Waeber, G., Vollenweider, P., Vestergaard, H., Hansen, T., Pedersen, O., Hu, F.B., Eline Slagboom, P., Grallert, H., Spector, T.D., Jukema, J.W., Klein, R.J., Schadt, E.E., Franks, P.W., Lindgren, C.M., Leibel, R.L., and Loos, R.J. (2016). Genome-wide meta-analysis uncovers novel loci influencing circulating leptin levels. Nat Commun *7*, 10494.

Kim, S., Kim, D., Cho, S.W., Kim, J., and Kim, J.S. (2014). Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome research *24*, 1012-1019.

Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell *161*, 1187-1201.

Klemm, S.L., Shipony, Z., and Greenleaf, W.J. (2019). Chromatin accessibility and the regulatory epigenome. Nature Reviews Genetics *20*, 207-220.

Koenig, A.L., Shchukina, I., Andhey, P.S., Zaitsev, K., Lai, L., Amrute, J., Bajpai, G., Bredemeyer, A., Smith, G., Jones, C., Terrebonne, E., Rentschler, S.L., Artyomov, M.N., and Lavine, K.J. (2021). Single Cell Transcriptomics Reveals Cell Type Specific Diversification in Human Heart Failure. bioRxiv, 2021.2007.2006.451312.

Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods *16*, 1289-1296.

Kubalak, S.W., Miller-Hance, W.C., O'Brien, T.X., Dyson, E., and Chien, K.R. (1994). Chamber specification of atrial myosin light chain-2 expression precedes septation during murine cardiogenesis. Journal of Biological Chemistry *269*, 16961-16970.

Lake, B.B., Chen, S., Sos, B.C., Fan, J., Kaeser, G.E., Yung, Y.C., Duong, T.E., Gao, D., Chun, J., Kharchenko, P.V., and Zhang, K. (2018). Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. Nat Biotechnol *36*, 70-80.

Lambert, J.-C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., Jun, G., DeStefano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., Russo, G., Thornton-Wells, T.A., Jones, N., Smith, A.V., Chouraki, V., Thomas, C., Ikram, M.A., Zelenika, D., Vardarajan, B.N., Kamatani, Y., Lin, C.-F., Gerrish, A., Schmidt, H., Kunkle, B., Dunstan, M.L., Ruiz, A., Bihoreau, M.-T., Choi, S.-H., Reitz, C., Pasquier, F., Hollingworth, P., Ramirez, A., Hanon, O., Fitzpatrick, A.L., Buxbaum, J.D., Campion, D., Crane, P.K., Baldwin, C., Becker, T., Gudnason, V., Cruchaga, C., Craig, D., Amin, N., Berr, C., Lopez, O.L., De Jager, P.L., Deramecourt, V., Johnston, J.A., Evans, D., Lovestone, S., Letenneur, L., Morón, F.J., Rubinsztein, D.C., Eiriksdottir, G., Sleegers, K., Goate, A.M., Fiévet, N., Huentelman, M.J., Gill, M., Brown, K., Kamboh, M.I., Keller, L., Barberger-Gateau, P., McGuinness, B., Larson, E.B., Green, R., Myers, A.J., Dufouil, C., Todd, S., Wallon, D., Love, S., Rogaeva, E., Gallacher, J., St George-Hyslop, P., Clarimon, J., Lleo, A., Bayer, A., Tsuang, D.W., Yu, L., Tsolaki, M., Bossù, P., Spalletta, G., Proitsi, P., Collinge, J., Sorbi, S., Sanchez-Garcia, F., Fox, N.C., Hardy, J., Naranjo, M.C.D., Bosco, P., Clarke, R., Brayne, C., Galimberti, D., Mancuso, M., Matthews, F., Moebus, S., Mecocci, P., Del Zompo, M., Maier, W., Hampel, H., Pilotto, A., Bullido, M., Panza, F., Caffarra, P., Nacmias, B., Gilbert, J.R., Mayhaus, M., Lannfelt, L., Hakonarson, H., Pichler, S., Carrasquillo, M.M., Ingelsson, M., Beekly, D., Alvarez, V., Zou, F., Valladares, O., Younkin, S.G., Coto, E., Hamilton-Nelson, K.L., Gu, W., Razquin, C., Pastor, P., Mateo, I., Owen, M.J., Faber, K.M., Jonsson, P.V., Combarros, O., O'Donovan, M.C., Cantwell, L.B., Soininen, H., Blacker, D., Mead, S., Mosley, T.H., Bennett, D.A., Harris, T.B., Fratiglioni, L., Holmes, C., de Bruijn, R.F.A.G., Passmore, P., Montine, T.J., Bettens, K., Rotter, J.I., Brice, A., Morgan, K., Foroud, T.M., Kukull, W.A., Hannequin, D., Powell, J.F., Nalls, M.A., Ritchie, K., Lunetta, K.L., Kauwe, J.S.K., Boerwinkle, E., Riemenschneider, M., Boada, M., Hiltunen, M., Martin, E.R., Schmidt, R., Rujescu, D., Wang, L.-S., Dartigues, J.-F., Mayeux, R., Tzourio, C., Hofman, A., Nöthen, M.M., Graff, C., Psaty, B.M., Jones, L., Haines, J.L., Holmans, P.A., Lathrop, M., Pericak-Vance, M.A., Launer, L.J., Farrer, L.A., van Duijn, C.M., Van Broeckhoven, C., Moskvina, V., Seshadri, S., Williams, J., Schellenberg, G.D., Amouyel, P., European Alzheimer's Disease, I., Genetic, Environmental Risk in Alzheimer's, D., Alzheimer's Disease Genetic, C., Cohorts for, H., and Aging Research in Genomic, E. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nature Genetics *45*, 1452-1458.

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650-665.

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

Lareau, C.A., Duarte, F.M., Chew, J.G., Kartha, V.K., Burkett, Z.D., Kohlway, A.S., Pokholok, D., Aryee, M.J., Steemers, F.J., Lebofsky, R., and Buenrostro, J.D. (2019). Droplet-based

combinatorial indexing for massive-scale single-cell chromatin accessibility. Nat Biotechnol *37*, 916-924.

Leland McInnes, J.H., Nathaniel Saul, Lukas Großberger (2018). UMAP: Uniform Manifold Approximation and Projection. Journal of Open Source Software *3(29)*, 861.

Lesourne, R., Uehara, S., Lee, J., Song, K.-D., Li, L., Pinkhasov, J., Zhang, Y., Weng, N.-P., Wildt, K.F., Wang, L., Bosselut, R., and Love, P.E. (2009). Themis, a T cell-specific protein important for late thymocyte development. Nat Immunol *10*, 840-847.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Li, Y.E., Preissl, S., Hou, X., Zhang, Z., Zhang, K., Qiu, Y., Poirion, O.B., Li, B., Chiou, J., Liu, H., Pinto-Duarte, A., Kubo, N., Yang, X., Fang, R., Wang, X., Han, J.Y., Lucero, J., Yan, Y., Miller, M., Kuan, S., Gorkin, D., Gaulton, K.J., Shen, Y., Nunn, M., Mukamel, E.A., Behrens, M.M., Ecker, J.R., and Ren, B. (2021). An atlas of gene regulatory elements in adult mouse cerebrum. Nature *598*, 129-136.

Lian, X., Zhang, J., Azarin, S.M., Zhu, K., Hazeltine, L.B., Bao, X., Hsiao, C., Kamp, T.J., and Palecek, S.P. (2013). Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/β-catenin signaling under fully defined conditions. Nature protocols *8*, 162-175.

Lighthouse, J.K., and Small, E.M. (2016). Transcriptional control of cardiac fibroblast plasticity. J Mol Cell Cardiol *91*, 52-60.

Linna-Kuosmanen, S., Schmauch, E., Galani, K., Boix, C.A., Hou, L., Örd, T., Toropainen, A., Stolze, L.K., Meibalan, E., Mantero, J.C., Renfro, A., Ojanen, J., Agudelo, L.Z., Hollmen, M., Jalkanen, J., Gunn, J., Tavi, P., Romanoski, C.E., MacRae, C.A., Kaikkonen, M.U., Garcia-Cardena, G., Kiviniemi, T., and Kellis, M. (2021). Single-cell dissection of live human hearts in ischemic heart disease and heart failure reveals cell-type-specific driver genes and pathways. bioRxiv, 2021.2006.2023.449672.

Litvinukova, M., Talavera-Lopez, C., Maatz, H., Reichart, D., Worth, C.L., Lindberg, E.L., Kanda, M., Polanski, K., Heinig, M., Lee, M., Nadelmann, E.R., Roberts, K., Tuck, L., Fasouli, E.S., DeLaughter, D.M., McDonough, B., Wakimoto, H., Gorham, J.M., Samari, S., Mahbubani, K.T., Saeb-Parsy, K., Patone, G., Boyle, J.J., Zhang, H., Zhang, H., Viveiros, A., Oudit, G.Y., Bayraktar, O., Seidman, J.G., Seidman, C.E., Noseda, M., Hubner, N., and Teichmann, S.A. (2020). Cells of the adult human heart. Nature.

Litviňuková, M., Talavera-López, C., Maatz, H., Reichart, D., Worth, C.L., Lindberg, E.L., Kanda, M., Polanski, K., Heinig, M., Lee, M., Nadelmann, E.R., Roberts, K., Tuck, L., Fasouli, E.S., DeLaughter, D.M., McDonough, B., Wakimoto, H., Gorham, J.M., Samari, S., Mahbubani, K.T., Saeb-Parsy, K., Patone, G., Boyle, J.J., Zhang, H., Zhang, H., Viveiros, A., Oudit, G.Y., Bayraktar, O.A., Seidman, J.G., Seidman, C.E., Noseda, M., Hubner, N., and Teichmann, S.A. (2020). Cells of the adult human heart. Nature *588*, 466-472.

Liu, C., Liu, W., Lu, M.F., Brown, N.A., and Martin, J.F. (2001). Regulation of left-right asymmetry by thresholds of Pitx2c activity. Development *128*, 2039-2048.

Locke, A.E., Kahali, B., Berndt, S.I., Justice, A.E., Pers, T.H., Day, F.R., Powell, C., Vedantam, S., Buchkovich, M.L., Yang, J., Croteau-Chonka, D.C., Esko, T., Fall, T., Ferreira, T., Gustafsson, S., Kutalik, Z., Luan, J.a., Mägi, R., Randall, J.C., Winkler, T.W., Wood, A.R., Workalemahu, T., Faul, J.D., Smith, J.A., Hua Zhao, J., Zhao, W., Chen, J., Fehrmann, R., Hedman, Å.K., Karjalainen, J., Schmidt, E.M., Absher, D., Amin, N., Anderson, D., Beekman, M., Bolton, J.L., Bragg-Gresham, J.L., Buyske, S., Demirkan, A., Deng, G., Ehret, G.B., Feenstra, B., Feitosa, M.F., Fischer, K., Goel, A., Gong, J., Jackson, A.U., Kanoni, S., Kleber, M.E., Kristiansson, K., Lim, U., Lotay, V., Mangino, M., Mateo Leach, I., Medina-Gomez, C., Medland, S.E., Nalls, M.A., Palmer, C.D., Pasko, D., Pechlivanis, S., Peters, M.J., Prokopenko, I., Shungin, D., Stančáková, A., Strawbridge, R.J., Ju Sung, Y., Tanaka, T., Teumer, A., Trompet, S., van der Laan, S.W., van Setten, J., Van Vliet-Ostaptchouk, J.V., Wang, Z., Yengo, L., Zhang, W., Isaacs, A., Albrecht, E., Ärnlöv, J., Arscott, G.M., Attwood, A.P., Bandinelli, S., Barrett, A., Bas, I.N., Bellis, C., Bennett, A.J., Berne, C., Blagieva, R., Blüher, M., Böhringer, S., Bonnycastle, L.L., Böttcher, Y., Boyd, H.A., Bruinenberg, M., Caspersen, I.H., Ida Chen, Y.-D., Clarke, R., Warwick Daw, E., de Craen, A.J.M., Delgado, G., Dimitriou, M., Doney, A.S.F., Eklund, N., Estrada, K., Eury, E., Folkersen, L., Fraser, R.M., Garcia, M.E., Geller, F., Giedraitis, V., Gigante, B., Go, A.S., Golay, A., Goodall, A.H., Gordon, S.D., Gorski, M., Grabe, H.-J., Grallert, H., Grammer, T.B., Gräßler, J., Grönberg, H., Groves, C.J., Gusto, G., Haessler, J., Hall, P., Haller, T., Hallmans, G., Hartman, C.A., Hassinen, M., Hayward, C., Heard-Costa, N.L., Helmer, Q., Hengstenberg, C., Holmen, O., Hottenga, J.-J., James, A.L., Jeff, J.M., Johansson, Å., Jolley, J., Juliusdottir, T., Kinnunen, L., Koenig, W., Koskenvuo, M., Kratzer, W., Laitinen, J., Lamina, C., Leander, K., Lee, N.R., Lichtner, P., Lind, L., Lindström, J., Sin Lo, K., Lobbens, S., Lorbeer, R., Lu, Y., Mach, F., Magnusson, P.K.E., Mahajan, A., McArdle, W.L., McLachlan, S., Menni, C., Merger, S., Mihailov, E., Milani, L., Moayyeri, A., Monda, K.L., Morken, M.A., Mulas, A., Müller, G., Müller-Nurasyid, M., Musk, A.W., Nagaraja, R., Nöthen, M.M., Nolte, I.M., Pilz, S., Rayner, N.W., Renstrom, F., Rettig, R., Ried, J.S., Ripke, S., Robertson, N.R., Rose, L.M., Sanna, S., Scharnagl, H., Scholtens, S., Schumacher, F.R., Scott, W.R., Seufferlein, T., Shi, J., Vernon Smith, A., Smolonska, J., Stanton, A.V., Steinthorsdottir, V., Stirrups, K., Stringham, H.M., Sundström, J., Swertz, M.A., Swift, A.J., Syvänen, A.-C., Tan, S.-T., Tayo, B.O., Thorand, B., Thorleifsson, G., Tyrer, J.P., Uh, H.-W., Vandenput, L., Verhulst, F.C., Vermeulen, S.H., Verweij, N., Vonk, J.M., Waite, L.L., Warren, H.R., Waterworth, D., Weedon, M.N., Wilkens, L.R., Willenborg, C., Wilsgaard, T., Wojczynski, M.K., Wong, A., Wright, A.F., Zhang, Q., Brennan, E.P., Choi, M., Dastani, Z., Drong, A.W., Eriksson, P., Franco-Cereceda, A., Gådin, J.R., Gharavi, A.G., Goddard, M.E., Handsaker, R.E., Huang, J., Karpe, F., Kathiresan, S., Keildson, S., Kiryluk, K., Kubo, M., Lee, J.-Y., Liang, L., Lifton, R.P., Ma, B., McCarroll, S.A., McKnight, A.J., Min, J.L., Moffatt, M.F., Montgomery, G.W., Murabito, J.M., Nicholson, G., Nyholt, D.R., Okada, Y., Perry, J.R.B., Dorajoo, R., Reinmaa, E., Salem, R.M., Sandholm, N., Scott, R.A., Stolk, L., Takahashi, A., Tanaka, T., van't Hooft, F.M., Vinkhuyzen, A.A.E., Westra, H.-J., Zheng, W., Zondervan, K.T., Heath, A.C., Arveiler, D., Bakker, S.J.L., Beilby, J., Bergman, R.N., Blangero, J., Bovet, P., Campbell, H., Caulfield, M.J., Cesana, G., Chakravarti, A., Chasman, D.I., Chines, P.S., Collins, F.S., Crawford, D.C., Adrienne Cupples, L., Cusi, D., Danesh, J., de Faire, U., den Ruijter, H.M., Dominiczak, A.F., Erbel, R., Erdmann, J., The LifeLines Cohort, S., The, A.C., The, A.-B.M.I.W.G., The, C.D.C., The, C.C., The, G., The, I., The, M.I., The Mu, T.C., The, M.C., The,

P.C., The ReproGen, C., The, G.C., and The International Endogene, C. (2015). Genetic studies of body mass index yield new insights for obesity biology. Nature *518*, 197-206.

Loos, R.J.F. (2020). 15 years of genome-wide association studies and no signs of slowing down. Nature Communications *11*, 5900.

Lowes, B.D., Gilbert, E.M., Abraham, W.T., Minobe, W.A., Larrabee, P., Ferguson, D., Wolfel, E.E., Lindenfeld, J., Tsvetkova, T., Robertson, A.D., Quaife, R.A., and Bristow, M.R. (2002). Myocardial Gene Expression in Dilated Cardiomyopathy Treated with Beta-Blocking Agents. New England Journal of Medicine *346*, 1357-1365.

Luo, C., Keown, C.L., Kurihara, L., Zhou, J., He, Y., Li, J., Castanon, R., Lucero, J., Nery, J.R., Sandoval, J.P., Bui, B., Sejnowski, T.J., Harkins, T.T., Mukamel, E.A., Behrens, M.M., and Ecker, J.R. (2017a). Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. Science *357*, 600-604.

Luo, Y., de Lange, K.M., Jostins, L., Moutsianas, L., Randall, J., Kennedy, N.A., Lamb, C.A., McCarthy, S., Ahmad, T., Edwards, C., Serra, E.G., Hart, A., Hawkey, C., Mansfield, J.C., Mowat, C., Newman, W.G., Nichols, S., Pollard, M., Satsangi, J., Simmons, A., Tremelling, M., Uhlig, H., Wilson, D.C., Lee, J.C., Prescott, N.J., Lees, C.W., Mathew, C.G., Parkes, M., Barrett, J.C., and Anderson, C.A. (2017b). Exploring the genetic architecture of inflammatory bowel disease by whole-genome sequencing identifies association at ADCY7. Nat Genet *49*, 186-192.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., Trombetta, J.J., Weitz, D.A., Sanes, J.R., Shalek, A.K., Regev, A., and McCarroll, S.A. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202-1214.

Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., Payne, A.J., Steinthorsdottir, V., Scott, R.A., Grarup, N., Cook, J.P., Schmidt, E.M., Wuttke, M., Sarnowski, C., Mägi, R., Nano, J., Gieger, C., Trompet, S., Lecoeur, C., Preuss, M.H., Prins, B.P., Guo, X., Bielak, L.F., Below, J.E., Bowden, D.W., Chambers, J.C., Kim, Y.J., Ng, M.C.Y., Petty, L.E., Sim, X., Zhang, W., Bennett, A.J., Bork-Jensen, J., Brummett, C.M., Canouil, M., Ec kardt, K.-U., Fischer, K., Kardia, S.L.R., Kronenberg, F., Läll, K., Liu, C.-T., Locke, A.E., Luan, J.a., Ntalla, I., Nylander, V., Schönherr, S., Schurmann, C., Yengo, L., Bottinger, E.P., Brandslund, I., Christensen, C., Dedoussis, G., Florez, J.C., Ford, I., Franco, O.H., Frayling, T.M., Giedraitis, V., Hackinger, S., Hattersley, A.T., Herder, C., Ikram, M.A., Ingelsson, M., Jørgensen, M.E., Jørgensen, T., Kriebel, J., Kuusisto, J., Ligthart, S., Lindgren, C.M., Linneberg, A., Lyssenko, V., Mamakou, V., Meitinger, T., Mohlke, K.L., Morris, A.D., Nadkarni, G., Pankow, J.S., Peters, A., Sattar, N., Stančáková, A., Strauch, K., Taylor, K.D., Thorand, B., Thorleifsson, G., Thorsteinsdottir, U., Tuomilehto, J., Witte, D.R., Dupuis, J., Peyser, P.A., Zeggini, E., Loos, R.J.F., Froguel, P., Ingelsson, E., Lind, L., Groop, L., Laakso, M., Collins, F.S., Jukema, J.W., Palmer, C.N.A., Grallert, H., Metspalu, A., Dehghan, A., Köttgen, A., Abecasis, G.R., Meigs, J.B., Rotter, J.I., Marchini, J., Pedersen, O., Hansen, T., Langenberg, C., Wareham, N.J., Stefansson, K., Gloyn, A.L., Morris, A.P., Boehnke, M., and McCarthy, M.I. (2018). Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nature Genetics *50*, 1505-1513.

Malik, R., Chauhan, G., Traylor, M., Sargurupremraj, M., Okada, Y., Mishra, A., Rutten-Jacobs, L., Giese, A.-K., van der Laan, S.W., Gretarsdottir, S., Anderson, C.D., Chong, M., Adams, H.H.H., Ago, T., Almgren, P., Amouyel, P., Ay, H., Bartz, T.M., Benavente, O.R., Bevan, S., Boncoraglio, G.B., Brown, R.D., Butterworth, A.S., Carrera, C., Carty, C.L., Chasman, D.I., Chen, W.-M., Cole, J.W., Correa, A., Cotlarciuc, I., Cruchaga, C., Danesh, J., de Bakker, P.I.W., DeStefano, A.L., den Hoed, M., Duan, Q., Engelter, S.T., Falcone, G.J., Gottesman, R.F., Grewal, R.P., Gudnason, V., Gustafsson, S., Haessler, J., Harris, T.B., Hassan, A., Havulinna, A.S., Heckbert, S.R., Holliday, E.G., Howard, G., Hsu, F.-C., Hyacinth, H.I., Ikram, M.A., Ingelsson, E., Irvin, M.R., Jian, X., Jiménez-Conde, J., Johnson, J.A., Jukema, J.W., Kanai, M., Keene, K.L., Kissela, B.M., Kleindorfer, D.O., Kooperberg, C., Kubo, M., Lange, L.A., Langefeld, C.D., Langenberg, C., Launer, L.J., Lee, J.-M., Lemmens, R., Leys, D., Lewis, C.M., Lin, W.-Y., Lindgren, A.G., Lorentzen, E., Magnusson, P.K., Maguire, J., Manichaikul, A., McArdle, P.F., Meschia, J.F., Mitchell, B.D., Mosley, T.H., Nalls, M.A., Ninomiya, T., O'Donnell, M.J., Psaty, B.M., Pulit, S.L., Rannikmäe, K., Reiner, A.P., Rexrode, K.M., Rice, K., Rich, S.S., Ridker, P.M., Rost, N.S., Rothwell, P.M., Rotter, J.I., Rundek, T., Sacco, R.L., Sakaue, S., Sale, M.M., Salomaa, V., Sapkota, B.R., Schmidt, R., Schmidt, C.O., Schminke, U., Sharma, P., Slowik, A., Sudlow, C.L.M., Tanislav, C., Tatlisumak, T., Taylor, K.D., Thijs, V.N.S., Thorleifsson, G., Thorsteinsdottir, U., Tiedt, S., Trompet, S., Tzourio, C., van Duijn, C.M., Walters, M., Wareham, N.J., Wassertheil-Smoller, S., Wilson, J.G., Wiggins, K.L., Yang, Q., Yusuf, S., Bis, J.C., Pastinen, T., Ruusalepp, A., Schadt, E.E., Koplev, S., Björkegren, J.L.M., Codoni, V., Civelek, M., Smith, N.L., Trégouët, D.A., Christophersen, I.E., Roselli, C., Lubitz, S.A., Ellinor, P.T., Tai, E.S., Kooner, J.S., Kato, N., He, J., van der Harst, P., Elliott, P., Chambers, J.C., Takeuchi, F., Johnson, A.D., Malik, R., Chauhan, G., Traylor, M., Sargurupremraj, M., Okada, Y., Mishra, A., Rutten-Jacobs, L., Giese, A.-K., van der Laan, S.W., Gretarsdottir, S., Anderson, C.D., Chong, M., Adams, H.H.H., Ago, T., Almgren, P., Amouyel, P., Ay, H., Bartz, T.M., Benavente, O.R., Bevan, S., Boncoraglio, G.B., Brown, R.D., Butterworth, A.S., Carrera, C., Carty, C.L., Chasman, D.I., Chen, W.-M., Cole, J.W., Correa, A., Cotlarciuc, I., Cruchaga, C., Danesh, J., de Bakker, P.I.W., DeStefano, A.L., Hoed, M.d., Duan, Q., Engelter, S.T., Falcone, G.J., Gottesman, R.F., Grewal, R.P., Gudnason, V., Gustafsson, S., Haessler, J., Harris, T.B., Hassan, A., Havulinna, A.S., Heckbert, S.R., Holliday, E.G., Howard, G., Hsu, F.-C., Hyacinth, H.I., Ikram, M.A., Ingelsson, E., Irvin, M.R., Jian, X., Jiménez-Conde, J., Johnson, J.A., Jukema, J.W., Kanai, M., Keene, K.L., Kissela, B.M., Kleindorfer, D.O., Kooperberg, C., Kubo, M., Lange, L.A., Langefeld, C.D., Langenberg, C., Launer, L.J., Lee, J.-M., Lemmens, R., Leys, D., Lewis, C.M., Lin, W.-Y., Lindgren, A.G., Lorentzen, E., Magnusson, P.K., Maguire, J., Manichaikul, A., McArdle, P.F., Meschia, J.F., Mitchell, B.D., Mosley, T.H., Nalls, M.A., Ninomiya, T., O'Donnell, M.J., Psaty, B.M., Pulit, S.L., Rannikmäe, K., Reiner, A.P., Rexrode, K.M., Rice, K., Rich, S.S., Ridker, P.M., Rost, N.S., Rothwell, P.M., Rotter, J.I., Rundek, T., Sacco, R.L., Sakaue, S., Sale, M.M., Salomaa, V., Sapkota, B.R., Schmidt, R., Schmidt, C.O., Schminke, U., Sharma, P., Slowik, A., Sudlow, C.L.M., Tanislav, C., Tatlisumak, T., Taylor, K.D., Thijs, V.N.S., Thorleifsson, G., Thorsteinsdottir, U., Tiedt, S., Trompet, S., Tzourio, C., van Duijn, C.M., Walters, M., Wareham, N.J., Wassertheil-Smoller, S., Wilson, J.G., Wiggins, K.L., Yang, Q., Yusuf, S., Amin, N., Aparicio, H.S., Arnett, D.K., Attia, J., Beiser, A.S., Berr, C., Buring, J.E., Bustamante, M., Caso, V., Cheng, Y.-C., Choi, S.H., Chowhan, A., Consortium, A.F., Cohorts for, H., Aging Research in Genomic Epidemiology, C., International Genomics of Blood Pressure, C., Consortium, I., Starnet, BioBank Japan Cooperative Hospital, G., Consortium, C., Consortium, E.-C., Consortium, E.P.-I., International Stroke Genetics, C., Consortium, M., Neurology Working Group of the, C.C.,

Network, N.S.G., Study, U.K.Y.L.D., and Consortium, M. (2018). Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. Nature Genetics *50*, 524-537.

Mann, D.L., and Bristow, M.R. (2005). Mechanisms and Models in Heart Failure. Circulation *111*, 2837-2849.

Manning, A.K., Hivert, M.F., Scott, R.A., Grimsby, J.L., Bouatia-Naji, N., Chen, H., Rybin, D., Liu, C.T., Bielak, L.F., Prokopenko, I., Amin, N., Barnes, D., Cadby, G., Hottenga, J.J., Ingelsson, E., Jackson, A.U., Johnson, T., Kanoni, S., Ladenvall, C., Lagou, V., Lahti, J., Lecoeur, C., Liu, Y., Martinez-Larrad, M.T., Montasser, M.E., Navarro, P., Perry, J.R., Rasmussen-Torvik, L.J., Salo, P., Sattar, N., Shungin, D., Strawbridge, R.J., Tanaka, T., van Duijn, C.M., An, P., de Andrade, M., Andrews, J.S., Aspelund, T., Atalay, M., Aulchenko, Y., Balkau, B., Bandinelli, S., Beckmann, J.S., Beilby, J.P., Bellis, C., Bergman, R.N., Blangero, J., Boban, M., Boehnke, M., Boerwinkle, E., Bonnycastle, L.L., Boomsma, D.I., Borecki, I.B., Böttcher, Y., Bouchard, C., Brunner, E., Budimir, D., Campbell, H., Carlson, O., Chines, P.S., Clarke, R., Collins, F.S., Corbatón-Anchuelo, A., Couper, D., de Faire, U., Dedoussis, G.V., Deloukas, P., Dimitriou, M., Egan, J.M., Eiriksdottir, G., Erdos, M.R., Eriksson, J.G., Eury, E., Ferrucci, L., Ford, I., Forouhi, N.G., Fox, C.S., Franzosi, M.G., Franks, P.W., Frayling, T.M., Froguel, P., Galan, P., de Geus, E., Gigante, B., Glazer, N.L., Goel, A., Groop, L., Gudnason, V., Hallmans, G., Hamsten, A., Hansson, O., Harris, T.B., Hayward, C., Heath, S., Hercberg, S., Hicks, A.A., Hingorani, A., Hofman, A., Hui, J., Hung, J., Jarvelin, M.R., Jhun, M.A., Johnson, P.C., Jukema, J.W., Jula, A., Kao, W.H., Kaprio, J., Kardia, S.L., Keinanen-Kiukaanniemi, S., Kivimaki, M., Kolcic, I., Kovacs, P., Kumari, M., Kuusisto, J., Kyvik, K.O., Laakso, M., Lakka, T., Lannfelt, L., Lathrop, G.M., Launer, L.J., Leander, K., Li, G., Lind, L., Lindstrom, J., Lobbens, S., Loos, R.J., Luan, J., Lyssenko, V., Mägi, R., Magnusson, P.K., Marmot, M., Meneton, P., Mohlke, K.L., Mooser, V., Morken, M.A., Miljkovic, I., Narisu, N., O'Connell, J., Ong, K.K., Oostra, B.A., Palmer, L.J., Palotie, A., Pankow, J.S., Peden, J.F., Pedersen, N.L., Pehlic, M., Peltonen, L., Penninx, B., Pericic, M., Perola, M., Perusse, L., Peyser, P.A., Polasek, O., Pramstaller, P.P., Province, M.A., Räikkönen, K., Rauramaa, R., Rehnberg, E., Rice, K., Rotter, J.I., Rudan, I., Ruokonen, A., Saaristo, T., Sabater-Lleal, M., Salomaa, V., Savage, D.B., Saxena, R., Schwarz, P., Seedorf, U., Sennblad, B., Serrano-Rios, M., Shuldiner, A.R., Sijbrands, E.J., Siscovick, D.S., Smit, J.H., Small, K.S., Smith, N.L., Smith, A.V., Stančáková, A., Stirrups, K., Stumvoll, M., Sun, Y.V., Swift, A.J., Tönjes, A., Tuomilehto, J., Trompet, S., Uitterlinden, A.G., Uusitupa, M., Vikström, M., Vitart, V., Vohl, M.C., Voight, B.F., Vollenweider, P., Waeber, G., Waterworth, D.M., Watkins, H., Wheeler, E., Widen, E., Wild, S.H., Willems, S.M., Willemsen, G., Wilson, J.F., Witteman, J.C., Wright, A.F., Yaghootkar, H., Zelenika, D., Zemunik, T., Zgaga, L., Wareham, N.J., McCarthy, M.I., Barroso, I., Watanabe, R.M., Florez, J.C., Dupuis, J., Meigs, J.B., and Langenberg, C. (2012). A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. Nat Genet *44*, 659-669.

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S., Boehnke, M., Clark, A.G., Eichler, E.E., Gibson, G., Haines, J.L., Mackay, T.F.C., McCarroll, S.A., and Visscher, P.M. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747-753.

Margulies, K.B., Matiwala, S., Cornejo, C., Olsen, H., Craven, W.A., and Bednarik, D. (2005). Mixed Messages. Circulation Research *96*, 592-599.

Martinez, F.O., Gordon, S., Locati, M., and Mantovani, A. (2006). Transcriptional Profiling of the Human Monocyte-to-Macrophage Differentiation and Polarization: New Molecules and Patterns of Gene Expression. The Journal of Immunology *177*, 7303-7311.

Mary Elizabeth Pownall, Marcus K. Gustafsson, and Charles P. Emerson, J. (2002). Myogenic Regulatory Factors and the Specification of Muscle Progenitors in Vertebrate Embryos. Annual Review of Cell and Developmental Biology *18*, 747-783.

Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.Y., Denay, G., Lee, J., Shi, W., Shyr, C., Tan, G., Worsley-Hunt, R., Zhang, A.W., Parcy, F., Lenhard, B., Sandelin, A., and Wasserman, W.W. (2016). JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. Nucleic acids research *44*, D110-115.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutyavin, T., Stehling-Sun, S., Johnson, A.K., Canfield, T.K., Giste, E., Diegel, M., Bates, D., Hansen, R.S., Neph, S., Sabo, P.J., Heimfeld, S., Raubitschek, A., Ziegler, S., Cotsapas, C., Sotoodehnia, N., Glass, I., Sunyaev, S.R., Kaul, R., and Stamatoyannopoulos, J.A. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science (New York, NY) *337*, 1190-1195.

May, D., Blow, M.J., Kaplan, T., McCulley, D.J., Jensen, B.C., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Afzal, V., Simpson, P.C., Rubin, E.M., Black, B.L., Bristow, J., Pennacchio, L.A., and Visel, A. (2011). Large-scale discovery of enhancers from human heart tissue. Nature genetics *44*, 89-93.

McDonagh, T.A., Metra, M., Adamo, M., Gardner, R.S., Baumbach, A., Böhm, M., Burri, H., Butler, J., Čelutkienė, J., Chioncel, O., Cleland, J.G.F., Coats, A.J.S., Crespo-Leiro, M.G., Farmakis, D., Gilard, M., Heymans, S., Hoes, A.W., Jaarsma, T., Jankowska, E.A., Lainscak, M., Lam, C.S.P., Lyon, A.R., McMurray, J.J.V., Mebazaa, A., Mindham, R., Muneretto, C., Francesco Piepoli, M., Price, S., Rosano, G.M.C., Ruschitzka, F., Kathrine Skibelund, A., and Group, E.S.D. (2021). 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: Developed by the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC) With the special contribution of the Heart Failure Association (HFA) of the ESC. European Heart Journal *42*, 3599-3726.

McGinnis, C.S., Murrow, L.M., and Gartner, Z.J. (2019). DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. Cell Systems *8*, 329-337.e324.

McInnes, L., Healy, J., and Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:180203426.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nature Biotechnology *28*, 495-501.

Metra, M., and Teerlink, J.R. (2017). Heart failure. The Lancet *390*, 1981-1995.

Meuleman, W., Muratov, A., Rynes, E., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Teodosiadis, A., Reynolds, A., Haugen, E., Nelson, J., Johnson, A., Frerker, M., Buckley, M., Sandstrom, R., Vierstra, J., Kaul, R., and Stamatoyannopoulos, J. (2020). Index and biological spectrum of human DNase I hypersensitive sites. Nature *584*, 244-251.

Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., Lemaçon, A., Soucy, P., Glubb, D., Rostamianfar, A., Bolla, M.K., Wang, Q., Tyrer, J., Dicks, E., Lee, A., Wang, Z., Allen, J., Keeman, R., Eilber, U., French, J.D., Qing Chen, X., Fachal, L., McCue, K., McCart Reed, A.E., Ghoussaini, M., Carroll, J.S., Jiang, X., Finucane, H., Adams, M., Adank, M.A., Ahsan, H., Aittomäki, K., Anton-Culver, H., Antonenkova, N.N., Arndt, V., Aronson, K.J., Arun, B., Auer, P.L., Bacot, F., Barrdahl, M., Baynes, C., Beckmann, M.W., Behrens, S., Benitez, J., Bermisheva, M., Bernstein, L., Blomqvist, C., Bogdanova, N.V., Bojesen, S.E., Bonanni, B., Børresen-Dale, A.L., Brand, J.S., Brauch, H., Brennan, P., Brenner, H., Brinton, L., Broberg, P., Brock, I.W., Broeks, A., Brooks-Wilson, A., Brucker, S.Y., Brüning, T., Burwinkel, B., Butterbach, K., Cai, Q., Cai, H., Caldés, T., Canzian, F., Carracedo, A., Carter, B.D., Castelao, J.E., Chan, T.L., David Cheng, T.Y., Seng Chia, K., Choi, J.Y., Christiansen, H., Clarke, C.L., Collée, M., Conroy, D.M., Cordina-Duverger, E., Cornelissen, S., Cox, D.G., Cox, A., Cross, S.S., Cunningham, J.M., Czene, K., Daly, M.B., Devilee, P., Doheny, K.F., Dörk, T., Dos-Santos-Silva, I., Dumont, M., Durcan, L., Dwek, M., Eccles, D.M., Ekici, A.B., Eliassen, A.H., Ellberg, C., Elvira, M., Engel, C., Eriksson, M., Fasching, P.A., Figueroa, J., Flesch-Janys, D., Fletcher, O., Flyger, H., Fritschi, L., Gaborieau, V., Gabrielson, M., Gago-Dominguez, M., Gao, Y.T., Gapstur, S.M., García-Sáenz, J.A., Gaudet, M.M., Georgoulias, V., Giles, G.G., Glendon, G., Goldberg, M.S., Goldgar, D.E., González-Neira, A., Grenaker Alnæs, G.I., Grip, M., Gronwald, J., Grundy, A., Guénel, P., Haeberle, L., Hahnen, E., Haiman, C.A., Håkansson, N., Hamann, U., Hamel, N., Hankinson, S., Harrington, P., Hart, S.N., Hartikainen, J.M., Hartman, M., Hein, A., Heyworth, J., Hicks, B., Hillemanns, P., Ho, D.N., Hollestelle, A., Hooning, M.J., Hoover, R.N., Hopper, J.L., Hou, M.F., Hsiung, C.N., Huang, G., Humphreys, K., Ishiguro, J., Ito, H., Iwasaki, M., Iwata, H., Jakubowska, A., Janni, W., John, E.M., Johnson, N., Jones, K., Jones, M., Jukkola-Vuorinen, A., Kaaks, R., Kabisch, M., Kaczmarek, K., Kang, D., Kasuga, Y., Kerin, M.J., Khan, S., Khusnutdinova, E., Kiiski, J.I., Kim, S.W., Knight, J.A., Kosma, V.M., Kristensen, V.N., Krüger, U., Kwong, A., Lambrechts, D., Le Marchand, L., Lee, E., Lee, M.H., Lee, J.W., Neng Lee, C., Lejbkowicz, F., Li, J., Lilyquist, J., Lindblom, A., Lissowska, J., Lo, W.Y., Loibl, S., Long, J., Lophatananon, A., Lubinski, J., Luccarini, C., Lux, M.P., Ma, E.S.K., MacInnis, R.J., Maishman, T., Makalic, E., Malone, K.E., Kostovska, I.M., Mannermaa, A., Manoukian, S., Manson, J.E., Margolin, S., Mariapun, S., Martinez, M.E., Matsuo, K., Mavroudis, D., McKay, J., McLean, C., Meijers-Heijboer, H., Meindl, A., Menéndez, P., Menon, U., Meyer, J., Miao, H., Miller, N., Taib, N.A.M., Muir, K., Mulligan, A.M., Mulot, C., Neuhausen, S.L., Nevanlinna, H., Neven, P., Nielsen, S.F., Noh, D.Y., Nordestgaard, B.G., Norman, A., Olopade, O.I., Olson, J.E., Olsson, H., Olswold, C., Orr, N., Pankratz, V.S., Park, S.K., Park-Simon, T.W., Lloyd, R., Perez, J.I.A., Peterlongo, P., Peto, J., Phillips, K.A., Pinchev, M., Plaseska-Karanfilska, D., Prentice, R., Presneau, N., Prokofyeva, D., Pugh, E., Pylkäs, K., Rack, B., Radice, P., Rahman, N., Rennert, G., Rennert, H.S., Rhenius, V., Romero, A., Romm, J., Ruddy, K.J., Rüdiger, T., Rudolph, A., Ruebner, M., Rutgers, E.J.T., Saloustros, E., Sandler, D.P., Sangrajrang, S., Sawyer, E.J., Schmidt, D.F., Schmutzler, R.K., Schneeweiss, A., Schoemaker, M.J., Schumacher, F., Schürmann, P., Scott, R.J., Scott, C., Seal, S., Seynaeve, C., Shah, M., Sharma, P., Shen, C.Y., Sheng, G., Sherman,

M.E., Shrubsole, M.J., Shu, X.O., Smeets, A., Sohn, C., Southey, M.C., Spinelli, J.J., Stegmaier, C., Stewart-Brown, S., Stone, J., Stram, D.O., Surowy, H., Swerdlow, A., Tamimi, R., Taylor, J.A., Tengström, M., Teo, S.H., Beth Terry, M., Tessier, D.C., Thanasitthichai, S., Thöne, K., Tollenaar, R., Tomlinson, I., Tong, L., Torres, D., Truong, T., Tseng, C.C., Tsugane, S., Ulmer, H.U., Ursin, G., Untch, M., Vachon, C., van Asperen, C.J., Van Den Berg, D., van den Ouweland, A.M.W., van der Kolk, L., van der Luijt, R.B., Vincent, D., Vollenweider, J., Waisfisz, Q., Wang-Gohrke, S., Weinberg, C.R., Wendt, C., Whittemore, A.S., Wildiers, H., Willett, W., Winqvist, R., Wolk, A., Wu, A.H., Xia, L., Yamaji, T., Yang, X.R., Har Yip, C., Yoo, K.Y., Yu, J.C., Zheng, W., Zheng, Y., Zhu, B., Ziogas, A., Ziv, E., Lakhani, S.R., Antoniou, A.C., Droit, A., Andrulis, I.L., Amos, C.I., Couch, F.J., Pharoah, P.D.P., Chang-Claude, J., Hall, P., Hunter, D.J., Milne, R.L., García-Closas, M., Schmidt, M.K., Chanock, S.J., Dunning, A.M., Edwards, S.L., Bader, G.D., Chenevix-Trench, G., Simard, J., Kraft, P., and Easton, D.F. (2017). Association analysis identifies 65 new breast cancer risk loci. Nature *551*, 92-94.

Molkentin, J.D., Lu, J.R., Antos, C.L., Markham, B., Richardson, J., Robbins, J., Grant, S.R., and Olson, E.N. (1998). A calcineurin-dependent transcriptional pathway for cardiac hypertrophy. Cell *93*, 215-228.

Moore, J.E., Purcaro, M.J., Pratt, H.E., Epstein, C.B., Shoresh, N., Adrian, J., Kawli, T., Davis, C.A., Dobin, A., Kaul, R., Halow, J., Van Nostrand, E.L., Freese, P., Gorkin, D.U., Shen, Y., He, Y., Mackiewicz, M., Pauli-Behn, F., Williams, B.A., Mortazavi, A., Keller, C.A., Zhang, X.-O., Elhajjajy, S.I., Huey, J., Dickel, D.E., Snetkova, V., Wei, X., Wang, X., Rivera-Mulia, J.C., Rozowsky, J., Zhang, J., Chhetri, S.B., Zhang, J., Victorsen, A., White, K.P., Visel, A., Yeo, G.W., Burge, C.B., Lécuyer, E., Gilbert, D.M., Dekker, J., Rinn, J., Mendenhall, E.M., Ecker, J.R., Kellis, M., Klein, R.J., Noble, W.S., Kundaje, A., Guigó, R., Farnham, P.J., Cherry, J.M., Myers, R.M., Ren, B., Graveley, B.R., Gerstein, M.B., Pennacchio, L.A., Snyder, M.P., Bernstein, B.E., Wold, B., Hardison, R.C., Gingeras, T.R., Stamatoyannopoulos, J.A., Weng, Z., and The, E.P.C. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature *583*, 699-710.

Moorman, A.F.M., and Christoffels, V.M. (2003). Cardiac Chamber Formation: Development, Genes, and Evolution. Physiological Reviews *83*, 1223-1267.

Murphy, S.P., Ibrahim, N.E., and Januzzi, J.L., Jr (2020). Heart Failure With Reduced Ejection Fraction: A Review. JAMA *324*, 488-504.

Nasser, J., Bergman, D.T., Fulco, C.P., Guckelberger, P., Doughty, B.R., Patwardhan, T.A., Jones, T.R., Nguyen, T.H., Ulirsch, J.C., Lekschas, F., Mualim, K., Natri, H.M., Weeks, E.M., Munson, G., Kane, M., Kang, H.Y., Cui, A., Ray, J.P., Eisenhaure, T.M., Collins, R.L., Dey, K., Pfister, H., Price, A.L., Epstein, C.B., Kundaje, A., Xavier, R.J., Daly, M.J., Huang, H., Finucane, H.K., Hacohen, N., Lander, E.S., and Engreitz, J.M. (2021). Genome-wide enhancer maps link risk variants to disease genes. Nature *593*, 238-243.

Nelson, C.P., Goel, A., Butterworth, A.S., Kanoni, S., Webb, T.R., Marouli, E., Zeng, L., Ntalla, I., Lai, F.Y., Hopewell, J.C., Giannakopoulou, O., Jiang, T., Hamby, S.E., Di Angelantonio, E., Assimes, T.L., Bottinger, E.P., Chambers, J.C., Clarke, R., Palmer, C.N.A., Cubbon, R.M., Ellinor, P., Ermel, R., Evangelou, E., Franks, P.W., Grace, C., Gu, D., Hingorani, A.D., Howson, J.M.M.,

Ingelsson, E., Kastrati, A., Kessler, T., Kyriakou, T., Lehtimäki, T., Lu, X., Lu, Y., März, W., McPherson, R., Metspalu, A., Pujades-Rodriguez, M., Ruusalepp, A., Schadt, E.E., Schmidt, A.F., Sweeting, M.J., Zalloua, P.A., AlGhalayini, K., Keavney, B.D., Kooner, J.S., Loos, R.J.F., Patel, R.S., Rutter, M.K., Tomaszewski, M., Tzoulaki, I., Zeggini, E., Erdmann, J., Dedoussis, G., Björkegren, J.L.M., Schunkert, H., Farrall, M., Danesh, J., Samani, N.J., Watkins, H., Deloukas, P., Consortium, E.-C., CardioGramplusC4D, and The, U.K.B.C.C.C.H.D.w.g. (2017). Association analyses based on false discovery rate implicate new loci for coronary artery disease. Nature Genetics *49*, 1385-1391.

Ng, S.Y., Wong, C.K., and Tsang, S.Y. (2010). Differential gene expressions in atrial and ventricular myocytes: insights into the road of applying embryonic stem cell-derived cardiomyocytes for future therapies. Am J Physiol Cell Physiol *299*, C1234-1249.

NHLBI (2014). Trans-Omics for Precision Medicine (TOPMed) Program.

Nichol, D., and Stuhlmann, H. (2012). EGFL7: a unique angiogenic signaling factor in vascular development and disease. Blood *119*, 1345-1352.

Nielsen, J.B., Thorolfsdottir, R.B., Fritsche, L.G., Zhou, W., Skov, M.W., Graham, S.E., Herron, T.J., McCarthy, S., Schmidt, E.M., Sveinbjornsson, G., Surakka, I., Mathis, M.R., Yamazaki, M., Crawford, R.D., Gabrielsen, M.E., Skogholt, A.H., Holmen, O.L., Lin, M., Wolford, B.N., Dey, R., Dalen, H., Sulem, P., Chung, J.H., Backman, J.D., Arnar, D.O., Thorsteinsdottir, U., Baras, A., O'Dushlaine, C., Holst, A.G., Wen, X., Hornsby, W., Dewey, F.E., Boehnke, M., Kheterpal, S., Mukherjee, B., Lee, S., Kang, H.M., Holm, H., Kitzman, J., Shavit, J.A., Jalife, J., Brummett, C.M., Teslovich, T.M., Carey, D.J., Gudbjartsson, D.F., Stefansson, K., Abecasis, G.R., Hveem, K., and Willer, C.J. (2018). Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. Nat Genet *50*, 1234-1239.

Nikpay, M., Goel, A., Won, H.H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C.P., Hopewell, J.C., Webb, T.R., Zeng, L., Dehghan, A., Alver, M., Armasu, S.M., Auro, K., Bjonnes, A., Chasman, D.I., Chen, S., Ford, I., Franceschini, N., Gieger, C., Grace, C., Gustafsson, S., Huang, J., Hwang, S.J., Kim, Y.K., Kleber, M.E., Lau, K.W., Lu, X., Lu, Y., Lyytikäinen, L.P., Mihailov, E., Morrison, A.C., Pervjakova, N., Qu, L., Rose, L.M., Salfati, E., Saxena, R., Scholz, M., Smith, A.V., Tikkanen, E., Uitterlinden, A., Yang, X., Zhang, W., Zhao, W., de Andrade, M., de Vries, P.S., van Zuydam, N.R., Anand, S.S., Bertram, L., Beutner, F., Dedoussis, G., Frossard, P., Gauguier, D., Goodall, A.H., Gottesman, O., Haber, M., Han, B.G., Huang, J., Jalilzadeh, S., Kessler, T., König, I.R., Lannfelt, L., Lieb, W., Lind, L., Lindgren, C.M., Lokki, M.L., Magnusson, P.K., Mallick, N.H., Mehra, N., Meitinger, T., Memon, F.U., Morris, A.P., Nieminen, M.S., Pedersen, N.L., Peters, A., Rallidis, L.S., Rasheed, A., Samuel, M., Shah, S.H., Sinisalo, J., Stirrups, K.E., Trompet, S., Wang, L., Zaman, K.S., Ardissino, D., Boerwinkle, E., Borecki, I.B., Bottinger, E.P., Buring, J.E., Chambers, J.C., Collins, R., Cupples, L.A., Danesh, J., Demuth, I., Elosua, R., Epstein, S.E., Esko, T., Feitosa, M.F., Franco, O.H., Franzosi, M.G., Granger, C.B., Gu, D., Gudnason, V., Hall, A.S., Hamsten, A., Harris, T.B., Hazen, S.L., Hengstenberg, C., Hofman, A., Ingelsson, E., Iribarren, C., Jukema, J.W., Karhunen, P.J., Kim, B.J., Kooner, J.S., Kullo, I.J., Lehtimäki, T., Loos, R.J.F., Melander, O., Metspalu, A., März, W., Palmer, C.N., Perola, M., Quertermous, T., Rader, D.J., Ridker, P.M., Ripatti, S., Roberts, R., Salomaa, V., Sanghera, D.K., Schwartz, S.M., Seedorf, U., Stewart, A.F., Stott, D.J., Thiery, J.,

Zalloua, P.A., O'Donnell, C.J., Reilly, M.P., Assimes, T.L., Thompson, J.R., Erdmann, J., Clarke, R., Watkins, H., Kathiresan, S., McPherson, R., Deloukas, P., Schunkert, H., Samani, N.J., and Farrall, M. (2015). A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet *47*, 1121-1130.

Nott, A., Holtman, I.R., Coufal, N.G., Schlachetzki, J.C.M., Yu, M., Hu, R., Han, C.Z., Pena, M., Xiao, J., Wu, Y., Keulen, Z., Pasillas, M.P., O'Connor, C., Nickl, C.K., Schafer, S.T., Shen, Z., Rissman, R.A., Brewer, J.B., Gosselin, D., Gonda, D.D., Levy, M.L., Rosenfeld, M.G., McVicker, G., Gage, F.H., Ren, B., and Glass, C.K. (2019). Brain cell type–specific enhancer–promoter interactome maps and disease-risk association. Science *366*, 1134-1139.

Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., Graham, R.R., Manoharan, A., Ortmann, W., Bhangale, T., Denny, J.C., Carroll, R.J., Eyler, A.E., Greenberg, J.D., Kremer, J.M., Pappas, D.A., Jiang, L., Yin, J., Ye, L., Su, D.-F., Yang, J., Xie, G., Keystone, E., Westra, H.-J., Esko, T., Metspalu, A., Zhou, X., Gupta, N., Mirel, D., Stahl, E.A., Diogo, D., Cui, J., Liao, K., Guo, M.H., Myouzen, K., Kawaguchi, T., Coenen, M.J.H., van Riel, P.L.C.M., van de Laar, M.A.F.J., Guchelaar, H.-J., Huizinga, T.W.J., Dieudé, P., Mariette, X., Louis Bridges Jr, S., Zhernakova, A., Toes, R.E.M., Tak, P.P., Miceli-Richard, C., Bang, S.-Y., Lee, H.-S., Martin, J., Gonzalez-Gay, M.A., Rodriguez-Rodriguez, L., Rantapää-Dahlqvist, S., Ärlestig, L., Choi, H.K., Kamatani, Y., Galan, P., Lathrop, M., Eyre, S., Bowes, J., Barton, A., de Vries, N., Moreland, L.W., Criswell, L.A., Karlson, E.W., Taniguchi, A., Yamada, R., Kubo, M., Liu, J.S., Bae, S.-C., Worthington, J., Padyukov, L., Klareskog, L., Gregersen, P.K., Raychaudhuri, S., Stranger, B.E., De Jager, P.L., Franke, L., Visscher, P.M., Brown, M.A., Yamanaka, H., Mimori, T., Takahashi, A., Xu, H., Behrens, T.W., Siminovitch, K.A., Momohara, S., Matsuda, F., Yamamoto, K., Plenge, R.M., the, R.c., and the, G.c. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature *506*, 376-381.

Okbay, A., Beauchamp, J.P., Fontana, M.A., Lee, J.J., Pers, T.H., Rietveld, C.A., Turley, P., Chen, G.-B., Emilsson, V., Meddens, S.F.W., Oskarsson, S., Pickrell, J.K., Thom, K., Timshel, P., de Vlaming, R., Abdellaoui, A., Ahluwalia, T.S., Bacelis, J., Baumbach, C., Bjornsdottir, G., Brandsma, J.H., Pina Concas, M., Derringer, J., Furlotte, N.A., Galesloot, T.E., Girotto, G., Gupta, R., Hall, L.M., Harris, S.E., Hofer, E., Horikoshi, M., Huffman, J.E., Kaasik, K., Kalafati, I.P., Karlsson, R., Kong, A., Lahti, J., Lee, S.J.v.d., deLeeuw, C., Lind, P.A., Lindgren, K.-O., Liu, T., Mangino, M., Marten, J., Mihailov, E., Miller, M.B., van der Most, P.J., Oldmeadow, C., Payton, A., Pervjakova, N., Peyrot, W.J., Qian, Y., Raitakari, O., Rueedi, R., Salvi, E., Schmidt, B., Schraut, K.E., Shi, J., Smith, A.V., Poot, R.A., St Pourcain, B., Teumer, A., Thorleifsson, G., Verweij, N., Vuckovic, D., Wellmann, J., Westra, H.-J., Yang, J., Zhao, W., Zhu, Z., Alizadeh, B.Z., Amin, N., Bakshi, A., Baumeister, S.E., Biino, G., Bønnelykke, K., Boyle, P.A., Campbell, H., Cappuccio, F.P., Davies, G., De Neve, J.-E., Deloukas, P., Demuth, I., Ding, J., Eibich, P., Eisele, L., Eklund, N., Evans, D.M., Faul, J.D., Feitosa, M.F., Forstner, A.J., Gandin, I., Gunnarsson, B., Halldórsson, B.V., Harris, T.B., Heath, A.C., Hocking, L.J., Holliday, E.G., Homuth, G., Horan, M.A., Hottenga, J.-J., de Jager, P.L., Joshi, P.K., Jugessur, A., Kaakinen, M.A., Kähönen, M., Kanoni, S., Keltigangas-Järvinen, L., Kiemeney, L.A.L.M., Kolcic, I., Koskinen, S., Kraja, A.T., Kroh, M., Kutalik, Z., Latvala, A., Launer, L.J., Lebreton, M.P., Levinson, D.F., Lichtenstein, P., Lichtner, P., Liewald, D.C.M., Cohort Study, L., Loukola, A., Madden, P.A., Mägi, R., Mäki-Opas, T., Marioni, R.E., Marques-Vidal, P., Meddens, G.A., McMahon, G., Meisinger, C., Meitinger, T., Milaneschi, Y., Milani, L., Montgomery, G.W.,

Myhre, R., Nelson, C.P., Nyholt, D.R., Ollier, W.E.R., Palotie, A., Paternoster, L., Pedersen, N.L., Petrovic, K.E., Porteous, D.J., Räikkönen, K., Ring, S.M., Robino, A., Rostapshova, O., Rudan, I., Rustichini, A., Salomaa, V., Sanders, A.R., Sarin, A.-P., Schmidt, H., Scott, R.J., Smith, B.H., Smith, J.A., Staessen, J.A., Steinhagen-Thiessen, E., Strauch, K., Terracciano, A., Tobin, M.D., Ulivi, S., Vaccargiu, S., Quaye, L., van Rooij, F.J.A., Venturini, C., Vinkhuyzen, A.A.E., Völker, U., Völzke, H., Vonk, J.M., Vozzi, D., Waage, J., Ware, E.B., Willemsen, G., Attia, J.R., Bennett, D.A., Berger, K., Bertram, L., Bisgaard, H., Boomsma, D.I., Borecki, I.B., Bültmann, U., Chabris, C.F., Cucca, F., Cusi, D., Deary, I.J., Dedoussis, G.V., van Duijn, C.M., Eriksson, J.G., Franke, B., Franke, L., Gasparini, P., Gejman, P.V., Gieger, C., Grabe, H.-J., Gratten, J., Groenen, P.J.F., Gudnason, V., van der Harst, P., Hayward, C., Hinds, D.A., Hoffmann, W., Hyppönen, E., Iacono, W.G., Jacobsson, B., Järvelin, M.-R., Jöckel, K.-H., Kaprio, J., Kardia, S.L.R., Lehtimäki, T., Lehrer, S.F., Magnusson, P.K.E., Martin, N.G., McGue, M., Metspalu, A., Pendleton, N., Penninx, B.W.J.H., Perola, M., Pirastu, N., Pirastu, M., Polasek, O., Posthuma, D., Power, C., Province, M.A., Samani, N.J., Schlessinger, D., Schmidt, R., Sørensen, T.I.A., Spector, T.D., Stefansson, K., Thorsteinsdottir, U., Thurik, A.R., Timpson, N.J., Tiemeier, H., Tung, J.Y., Uitterlinden, A.G., Vitart, V., Vollenweider, P., Weir, D.R., Wilson, J.F., Wright, A.F., Conley, D.C., Krueger, R.F., Davey Smith, G., Hofman, A., Laibson, D.I., Medland, S.E., Meyer, M.N., Yang, J., Johannesson, M., Visscher, P.M., Esko, T., Koellinger, P.D., Cesarini, D., and Benjamin, D.J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. Nature *533*, 539-542.

Paternoster, L., Standl, M., Waage, J., Baurecht, H., Hotze, M., Strachan, D.P., Curtin, J.A., Bønnelykke, K., Tian, C., Takahashi, A., Esparza-Gordillo, J., Alves, A.C., Thyssen, J.P., den Dekker, H.T., Ferreira, M.A., Altmaier, E., Sleiman, P.M., Xiao, F.L., Gonzalez, J.R., Marenholz, I., Kalb, B., Yanes, M.P., Xu, C.J., Carstensen, L., Groen-Blokhuis, M.M., Venturini, C., Pennell, C.E., Barton, S.J., Levin, A.M., Curjuric, I., Bustamante, M., Kreiner-Møller, E., Lockett, G.A., Bacelis, J., Bunyavanich, S., Myers, R.A., Matanovic, A., Kumar, A., Tung, J.Y., Hirota, T., Kubo, M., McArdle, W.L., Henderson, A.J., Kemp, J.P., Zheng, J., Smith, G.D., Rüschendorf, F., Bauerfeind, A., Lee-Kirsch, M.A., Arnold, A., Homuth, G., Schmidt, C.O., Mangold, E., Cichon, S., Keil, T., Rodríguez, E., Peters, A., Franke, A., Lieb, W., Novak, N., Fölster-Holst, R., Horikoshi, M., Pekkanen, J., Sebert, S., Husemoen, L.L., Grarup, N., de Jongste, J.C., Rivadeneira, F., Hofman, A., Jaddoe, V.W., Pasmans, S.G., Elbert, N.J., Uitterlinden, A.G., Marks, G.B., Thompson, P.J., Matheson, M.C., Robertson, C.F., Ried, J.S., Li, J., Zuo, X.B., Zheng, X.D., Yin, X.Y., Sun, L.D., McAleer, M.A., O'Regan, G.M., Fahy, C.M., Campbell, L.E., Macek, M., Kurek, M., Hu, D., Eng, C., Postma, D.S., Feenstra, B., Geller, F., Hottenga, J.J., Middeldorp, C.M., Hysi, P., Bataille, V., Spector, T., Tiesler, C.M., Thiering, E., Pahukasahasram, B., Yang, J.J., Imboden, M., Huntsman, S., Vilor-Tejedor, N., Relton, C.L., Myhre, R., Nystad, W., Custovic, A., Weiss, S.T., Meyers, D.A., Söderhäll, C., Melén, E., Ober, C., Raby, B.A., Simpson, A., Jacobsson, B., Holloway, J.W., Bisgaard, H., Sunyer, J., Hensch, N.M.P., Williams, L.K., Godfrey, K.M., Wang, C.A., Boomsma, D.I., Melbye, M., Koppelman, G.H., Jarvis, D., McLean, W.I., Irvine, A.D., Zhang, X.J., Hakonarson, H., Gieger, C., Burchard, E.G., Martin, N.G., Duijts, L., Linneberg, A., Jarvelin, M.R., Noethen, M.M., Lau, S., Hübner, N., Lee, Y.A., Tamari, M., Hinds, D.A., Glass, D., Brown, S.J., Heinrich, J., Evans, D.M., and Weidinger, S. (2015). Multi-ancestry genome-wide association study of 21,000 cases and 95,000 controls identifies new risk loci for atopic dermatitis. Nat Genet *47*, 1449-1456.

Perrino, C., and Rockman, H.A. (2006). GATA4 and the Two Sides of Gene Expression Reprogramming. Circulation Research *98*, 715-716.

Peschon, J.J., Morrissey, P.J., Grabstein, K.H., Ramsdell, F.J., Maraskovsky, E., Gliniak, B.C., Park, L.S., Ziegler, S.F., Williams, D.E., Ware, C.B., Meyer, J.D., and Davison, B.L. (1994). Early lymphocyte expansion is severely impaired in interleukin 7 receptor-deficient mice. J Exp Med *180*, 1955-1960.

Pividori, M., Schoettler, N., Nicolae, D.L., Ober, C., and Im, H.K. (2019). Shared and distinct genetic risk factors for childhood-onset and adult-onset asthma: genome-wide and transcriptome-wide studies. Lancet Respir Med *7*, 509-522.

Pliner, H.A., Packer, J.S., McFaline-Figueroa, J.L., Cusanovich, D.A., Daza, R.M., Aghamirzaie, D., Srivatsan, S., Qiu, X., Jackson, D., Minkina, A., Adey, A.C., Steemers, F.J., Shendure, J., and Trapnell, C. (2018). Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. Molecular cell *71*, 858-871.e858.

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res *20*, 110-121.

Porcu, E., Sadler, M.C., Lepik, K., Auwerx, C., Wood, A.R., Weihs, A., Sleiman, M.S.B., Ribeiro, D.M., Bandinelli, S., Tanaka, T., Nauck, M., Völker, U., Delaneau, O., Metspalu, A., Teumer, A., Frayling, T., Santoni, F.A., Reymond, A., and Kutalik, Z. (2021). Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome. Nature Communications *12*, 5647.

Preissl, S., Fang, R., Huang, H., Zhao, Y., Raviram, R., Gorkin, D.U., Zhang, Y., Sos, B.C., Afzal, V., Dickel, D.E., Kuan, S., Visel, A., Pennacchio, L.A., Zhang, K., and Ren, B. (2018). Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. Nat Neurosci *21*, 432-439.

Puri, V., Ranjit, S., Konda, S., Nicoloro, S.M.C., Straubhaar, J., Chawla, A., Chouinard, M., Lin, C., Burkart, A., Corvera, S., Perugini, R.A., and Czech, M.P. (2008). Cidea is associated with lipid droplets and insulin sensitivity in humans. Proceedings of the National Academy of Sciences *105*, 7833-7838.

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

Rajabi, M., Kassiotis, C., Razeghi, P., and Taegtmeyer, H. (2007). Return to the fetal gene program protects the stressed heart: a strong hypothesis. Heart Failure Reviews *12*, 331-343.

Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic acids research *42*, W187-191.

Reilly, M.P., and Bornfeldt, K.E. (2021). Integrative Multiomics Approaches for Discovery of New Drug Targets for Cardiovascular Disease. Circulation *143*, 2471-2474.

Ripke, S., Neale, B.M., Corvin, A., Walters, J.T.R., Farh, K.-H., Holmans, P.A., Lee, P., Bulik-Sullivan, B., Collier, D.A., Huang, H., Pers, T.H., Agartz, I., Agerbo, E., Albus, M., Alexander, M., Amin, F., Bacanu, S.A., Begemann, M., Belliveau Jr, R.A., Bene, J., Bergen, S.E., Bevilacqua, E., Bigdeli, T.B., Black, D.W., Bruggeman, R., Buccola, N.G., Buckner, R.L., Byerley, W., Cahn, W., Cai, G., Campion, D., Cantor, R.M., Carr, V.J., Carrera, N., Catts, S.V., Chambert, K.D., Chan, R.C.K., Chen, R.Y.L., Chen, E.Y.H., Cheng, W., Cheung, E.F.C., Ann Chong, S., Robert Cloninger, C., Cohen, D., Cohen, N., Cormican, P., Craddock, N., Crowley, J.J., Curtis, D., Davidson, M., Davis, K.L., Degenhardt, F., Del Favero, J., Demontis, D., Dikeos, D., Dinan, T., Djurovic, S., Donohoe, G., Drapeau, E., Duan, J., Dudbridge, F., Durmishi, N., Eichhammer, P., Eriksson, J., Escott-Price, V., Essioux, L., Fanous, A.H., Farrell, M.S., Frank, J., Franke, L., Freedman, R., Freimer, N.B., Friedl, M., Friedman, J.I., Fromer, M., Genovese, G., Georgieva, L., Giegling, I., Giusti-Rodríguez, P., Godard, S., Goldstein, J.I., Golimbet, V., Gopal, S., Gratten, J., de Haan, L., Hammer, C., Hamshere, M.L., Hansen, M., Hansen, T., Haroutunian, V., Hartmann, A.M., Henskens, F.A., Herms, S., Hirschhorn, J.N., Hoffmann, P., Hofman, A., Hollegaard, M.V., Hougaard, D.M., Ikeda, M., Joa, I., Julià, A., Kahn, R.S., Kalaydjieva, L., Karachanak-Yankova, S., Karjalainen, J., Kavanagh, D., Keller, M.C., Kennedy, J.L., Khrunin, A., Kim, Y., Klovins, J., Knowles, J.A., Konte, B., Kucinskas, V., Ausrele Kucinskiene, Z., Kuzelova-Ptackova, H., Kähler, A.K., Laurent, C., Lee Chee Keong, J., Hong Lee, S., Legge, S.E., Lerer, B., Li, M., Li, T., Liang, K.-Y., Lieberman, J., Limborska, S., Loughland, C.M., Lubinski, J., Lönnqvist, J., Macek Jr, M., Magnusson, P.K.E., Maher, B.S., Maier, W., Mallet, J., Marsal, S., Mattheisen, M., Mattingsdal, M., McCarley, R.W., McDonald, C., McIntosh, A.M., Meier, S., Meijer, C.J., Melegh, B., Melle, I., Mesholam-Gately, R.I., Metspalu, A., Michie, P.T., Milani, L., Milanova, V., Mokrab, Y., Morris, D.W., Mors, O., Murphy, K.C., Murray, R.M., Myin-Germeys, I., Müller-Myhsok, B., Nelis, M., Nenadic, I., Nertney, D.A., Nestadt, G., Nicodemus, K.K., Nikitina-Zake, L., Nisenbaum, L., Nordin, A., O'Callaghan, E., O'Dushlaine, C., O'Neill, F.A., Oh, S.-Y., Olincy, A., Olsen, L., Van Os, J., Pantelis, C., Papadimitriou, G.N., Papiol, S., Parkhomenko, E., Pato, M.T., Paunio, T., Pejovic-Milovancevic, M., Perkins, D.O., Pietiläinen, O., Pimm, J., Pocklington, A.J., Powell, J., Price, A., Pulver, A.E., Purcell, S.M., Quested, D., Rasmussen, H.B., Reichenberg, A., Reimers, M.A., Richards, A.L., Roffman, J.L., Roussos, P., Ruderfer, D.M., Salomaa, V., Sanders, A.R., Schall, U., Schubert, C.R., Schulze, T.G., Schwab, S.G., Scolnick, E.M., Scott, R.J., Seidman, L.J., Shi, J., Sigurdsson, E., Silagadze, T., Silverman, J.M., Sim, K., Slominsky, P., Smoller, J.W., So, H.-C., Spencer, C.A., Stahl, E.A., Stefansson, H., Steinberg, S., Stogmann, E., Straub, R.E., Strengman, E., Strohmaier, J., Scott Stroup, T., Subramaniam, M., Suvisaari, J., Svrakic, D.M., Szatkiewicz, J.P., Söderman, E., Thirumalai, S., Toncheva, D., Tosato, S., Veijola, J., Waddington, J., Walsh, D., Wang, D., Wang, Q., Webb, B.T., Weiser, M., Wildenauer, D.B., Williams, N.M., Williams, S., Witt, S.H., Wolen, A.R., Wong, E.H.M., Wormley, B.K., Simon Xi, H., Zai, C.C., Zheng, X., Zimprich, F., Wray, N.R., Stefansson, K., Visscher, P.M., Trust Case-Control Consortium, W., Adolfsson, R., Andreassen, O.A., Blackwood, D.H.R., Bramon, E., Buxbaum, J.D., Børglum, A.D., Cichon, S., Darvasi, A., Domenici, E., Ehrenreich, H., Esko, T., Gejman, P.V., Gill, M., Gurling, H., Hultman, C.M., Iwata, N., Jablensky, A.V., Jönsson, E.G., Kendler, K.S., Kirov, G., Knight, J., Lencz, T., Levinson, D.F., Li, Q.S., Liu, J., Malhotra, A.K., McCarroll, S.A., McQuillin, A., Moran, J.L., Mortensen, P.B., Mowry, B.J., Nöthen, M.M., Ophoff, R.A., Owen, M.J., Palotie, A., Pato, C.N., Petryshen, T.L., Posthuma, D., Rietschel, M., Riley, B.P., Rujescu, D., Sham, P.C., Sklar, P., St Clair, D., Weinberger, D.R., Wendland, J.R., Werge, T., Schizophrenia Working Group of the Psychiatric Genomics, C., and Psychosis

Endophenotypes International, C. (2014). Biological insights from 108 schizophrenia-associated genetic loci. Nature *511*, 421-427.

Rivaud, M.R., Delmar, M., and Remme, C.A. (2020). Heritable arrhythmia syndromes associated with abnormal cardiac sodium channel function: ionic and non-ionic mechanisms. Cardiovasc Res *116*, 1557-1570.

Roadmap Epigenomics, C., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., Amin, V., Whitaker, J.W., Schultz, M.D., Ward, L.D., Sarkar, A., Quon, G., Sandstrom, R.S., Eaton, M.L., Wu, Y.C., Pfenning, A.R., Wang, X., Claussnitzer, M., Liu, Y., Coarfa, C., Harris, R.A., Shoresh, N., Epstein, C.B., Gjoneska, E., Leung, D., Xie, W., Hawkins, R.D., Lister, R., Hong, C., Gascard, P., Mungall, A.J., Moore, R., Chuah, E., Tam, A., Canfield, T.K., Hansen, R.S., Kaul, R., Sabo, P.J., Bansal, M.S., Carles, A., Dixon, J.R., Farh, K.H., Feizi, S., Karlic, R., Kim, A.R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T.R., Neph, S.J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R.C., Siebenthall, K.T., Sinnott-Armstrong, N.A., Stevens, M., Thurman, R.E., Wu, J., Zhang, B., Zhou, X., Beaudet, A.E., Boyer, L.A., De Jager, P.L., Farnham, P.J., Fisher, S.J., Haussler, D., Jones, S.J., Li, W., Marra, M.A., McManus, M.T., Sunyaev, S., Thomson, J.A., Tlsty, T.D., Tsai, L.H., Wang, W., Waterland, R.A., Zhang, M.Q., Chadwick, L.H., Bernstein, B.E., Costello, J.F., Ecker, J.R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J.A., Wang, T., and Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317-330.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24-26.

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

Sakornsakolpat, P., Prokopenko, D., Lamontagne, M., Reeve, N.F., Guyatt, A.L., Jackson, V.E., Shrine, N., Qiao, D., Bartz, T.M., Kim, D.K., Lee, M.K., Latourelle, J.C., Li, X., Morrow, J.D., Obeidat, M., Wyss, A.B., Bakke, P., Barr, R.G., Beaty, T.H., Belinsky, S.A., Brusselle, G.G., Crapo, J.D., de Jong, K., DeMeo, D.L., Fingerlin, T.E., Gharib, S.A., Gulsvik, A., Hall, I.P., Hokanson, J.E., Kim, W.J., Lomas, D.A., London, S.J., Meyers, D.A., O'Connor, G.T., Rennard, S.I., Schwartz, D.A., Sliwinski, P., Sparrow, D., Strachan, D.P., Tal-Singer, R., Tesfaigzi, Y., Vestbo, J., Vonk, J.M., Yim, J.J., Zhou, X., Bosse, Y., Manichaikul, A., Lahousse, L., Silverman, E.K., Boezen, H.M., Wain, L.V., Tobin, M.D., Hobbs, B.D., and Cho, M.H. (2019). Genetic landscape of chronic obstructive pulmonary disease identifies heterogeneous cell-type and phenotype associations. Nat Genet *51*, 494-505.

Salem, S., Salem, D., and Gros, P. (2020). Role of IRF8 in immune cells functions, protection against infections, and susceptibility to inflammatory diseases. Human Genetics *139*, 707-721.

Santos, D.P., Kiskinis, E., Eggan, K., and Merkle, F.T. (2016). Comprehensive Protocols for CRISPR/Cas9-based Gene Editing in Human Pluripotent Stem Cells. Curr Protoc Stem Cell Biol *38*, 5b.6.1-5b.6.60.

Satpathy, A.T., Granja, J.M., Yost, K.E., Qi, Y., Meschi, F., McDermott, G.P., Olsen, B.N., Mumbach, M.R., Pierce, S.E., Corces, M.R., Shah, P., Bell, J.C., Jhutty, D., Nemec, C.M., Wang, J., Wang, L., Yin, Y., Giresi, P.G., Chang, A.L.S., Zheng, G.X.Y., Greenleaf, W.J., and Chang, H.Y. (2019). Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat Biotechnol *37*, 925-936.

Saxena, R., Hivert, M.F., Langenberg, C., Tanaka, T., Pankow, J.S., Vollenweider, P., Lyssenko, V., Bouatia-Naji, N., Dupuis, J., Jackson, A.U., Kao, W.H., Li, M., Glazer, N.L., Manning, A.K., Luan, J., Stringham, H.M., Prokopenko, I., Johnson, T., Grarup, N., Boesgaard, T.W., Lecoeur, C., Shrader, P., O'Connell, J., Ingelsson, E., Couper, D.J., Rice, K., Song, K., Andreasen, C.H., Dina, C., Köttgen, A., Le Bacquer, O., Pattou, F., Taneera, J., Steinthorsdottir, V., Rybin, D., Ardlie, K., Sampson, M., Qi, L., van Hoek, M., Weedon, M.N., Aulchenko, Y.S., Voight, B.F., Grallert, H., Balkau, B., Bergman, R.N., Bielinski, S.J., Bonnefond, A., Bonnycastle, L.L., Borch-Johnsen, K., Böttcher, Y., Brunner, E., Buchanan, T.A., Bumpstead, S.J., Cavalcanti-Proença, C., Charpentier, G., Chen, Y.D., Chines, P.S., Collins, F.S., Cornelis, M., G, J.C., Delplanque, J., Doney, A., Egan, J.M., Erdos, M.R., Firmann, M., Forouhi, N.G., Fox, C.S., Goodarzi, M.O., Graessler, J., Hingorani, A., Isomaa, B., Jørgensen, T., Kivimaki, M., Kovacs, P., Krohn, K., Kumari, M., Lauritzen, T., Lévy-Marchal, C., Mayor, V., McAteer, J.B., Meyre, D., Mitchell, B.D., Mohlke, K.L., Morken, M.A., Narisu, N., Palmer, C.N., Pakyz, R., Pascoe, L., Payne, F., Pearson, D., Rathmann, W., Sandbaek, A., Sayer, A.A., Scott, L.J., Sharp, S.J., Sijbrands, E., Singleton, A., Siscovick, D.S., Smith, N.L., Sparsø, T., Swift, A.J., Syddall, H., Thorleifsson, G., Tönjes, A., Tuomi, T., Tuomilehto, J., Valle, T.T., Waeber, G., Walley, A., Waterworth, D.M., Zeggini, E., Zhao, J.H., Illig, T., Wichmann, H.E., Wilson, J.F., van Duijn, C., Hu, F.B., Morris, A.D., Frayling, T.M., Hattersley, A.T., Thorsteinsdottir, U., Stefansson, K., Nilsson, P., Syvänen, A.C., Shuldiner, A.R., Walker, M., Bornstein, S.R., Schwarz, P., Williams, G.H., Nathan, D.M., Kuusisto, J., Laakso, M., Cooper, C., Marmot, M., Ferrucci, L., Mooser, V., Stumvoll, M., Loos, R.J., Altshuler, D., Psaty, B.M., Rotter, J.I., Boerwinkle, E., Hansen, T., Pedersen, O., Florez, J.C., McCarthy, M.I., Boehnke, M., Barroso, I., Sladek, R., Froguel, P., Meigs, J.B., Groop, L., Wareham, N.J., and Watanabe, R.M. (2010). Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. Nat Genet *42*, 142-148.

Schafmayer, C., Harrison, J.W., Buch, S., Lange, C., Reichert, M.C., Hofer, P., Cossais, F., Kupcinskas, J., von Schönfels, W., Schniewind, B., Kruis, W., Tepel, J., Zobel, M., Rosendahl, J., Jacobi, T., Walther-Berends, A., Schroeder, M., Vogel, I., Sergeev, P., Boedeker, H., Hinrichsen, H., Volk, A., Erk, J.U., Burmeister, G., Hendricks, A., Hinz, S., Wolff, S., Böttner, M., Wood, A.R., Tyrrell, J., Beaumont, R.N., Langheinrich, M., Kucharzik, T., Brezina, S., Huber-Schönauer, U., Pietsch, L., Noack, L.S., Brosch, M., Herrmann, A., Thangapandi, R.V., Schimming, H.W., Zeissig, S., Palm, S., Focke, G., Andreasson, A., Schmidt, P.T., Weitz, J., Krawczak, M., Völzke, H., Leeb, G., Michl, P., Lieb, W., Grützmann, R., Franke, A., Lammert, F., Becker, T., Kupcinskas, L., D'Amato, M., Wedel, T., Datz, C., Gsur, A., Weedon, M.N., and Hampe, J. (2019). Genome-wide association analysis of diverticular disease points towards neuromuscular, connective tissue and epithelial pathomechanisms. Gut *68*, 854-865.

Schaid, D.J., Chen, W., and Larson, N.B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet *19*, 491-504.

Schaum, N., Karkanias, J., Neff, N.F., May, A.P., Quake, S.R., Wyss-Coray, T., Darmanis, S., Batson, J., Botvinnik, O., Chen, M.B., Chen, S., Green, F., Jones, R.C., Maynard, A., Penland, L., Pisco, A.O., Sit, R.V., Stanley, G.M., Webber, J.T., Zanini, F., Baghel, A.S., Bakerman, I., Bansal, I., Berdnik, D., Bilen, B., Brownfield, D., Cain, C., Chen, M.B., Chen, S., Cho, M., Cirolia, G., Conley, S.D., Darmanis, S., Demers, A., Demir, K., de Morree, A., Divita, T., du Bois, H., Dulgeroff, L.B.T., Ebadi, H., Espinoza, F.H., Fish, M., Gan, Q., George, B.M., Gillich, A., Green, F., Genetiano, G., Gu, X., Gulati, G.S., Hang, Y., Hosseinzadeh, S., Huang, A., Iram, T., Isobe, T., Ives, F., Jones, R.C., Kao, K.S., Karnam, G., Kershner, A.M., Kiss, B.M., Kong, W., Kumar, M.E., Lam, J.Y., Lee, D.P., Lee, S.E., Li, G., Li, Q., Liu, L., Lo, A., Lu, W.-J., Manjunath, A., May, A.P., May, K.L., May, O.L., Maynard, A., McKay, M., Metzger, R.J., Mignardi, M., Min, D., Nabhan, A.N., Neff, N.F., Ng, K.M., Noh, J., Patkar, R., Peng, W.C., Penland, L., Puccinelli, R., Rulifson, E.J., Schaum, N., Sikandar, S.S., Sinha, R., Sit, R.V., Szade, K., Tan, W., Tato, C., Tellez, K., Travaglini, K.J., Tropini, C., Waldburger, L., van Weele, L.J., Wosczyna, M.N., Xiang, J., Xue, S., Youngyunpipatkul, J., Zanini, F., Zardeneta, M.E., Zhang, F., Zhou, L., Bansal, I., Chen, S., Cho, M., Cirolia, G., Darmanis, S., Demers, A., Divita, T., Ebadi, H., Genetiano, G., Green, F., Hosseinzadeh, S., Ives, F., Lo, A., May, A.P., Maynard, A., McKay, M., Neff, N.F., Penland, L., Sit, R.V., Tan, W., Waldburger, L., Youngyunpipatkul, J., Batson, J., Botvinnik, O., Castro, P., Croote, D., Darmanis, S., DeRisi, J.L., Karkanias, J., Pisco, A.O., Stanley, G.M., Webber, J.T., Zanini, F., Baghel, A.S., Bakerman, I., Batson, J., Bilen, B., Botvinnik, O., Brownfield, D., Chen, M.B., Darmanis, S., Demir, K., de Morree, A., Ebadi, H., Espinoza, F.H., Fish, M., Gan, Q., George, B.M., Gillich, A., Gu, X., Gulati, G.S., Hang, Y., Huang, A., Iram, T., Isobe, T., Karnam, G., Kershner, A.M., Kiss, B.M., Kong, W., Kuo, C.S., Lam, J.Y., Lehallier, B., Li, G., Li, Q., Liu, L., Lu, W.-J., Min, D., Nabhan, A.N., Ng, K.M., Nguyen, P.K., Patkar, R., Peng, W.C., Penland, L., Rulifson, E.J., Schaum, N., Sikandar, S.S., Sinha, R., Szade, K., Tan, S.Y., Tellez, K., Travaglini, K.J., Tropini, C., van Weele, L.J., Wang, B.M., Wosczyna, M.N., Xiang, J., Yousef, H., Zhou, L., Batson, J., Botvinnik, O., Chen, S., Darmanis, S., Green, F., May, A.P., Maynard, A., Pisco, A.O., Quake, S.R., Schaum, N., Stanley, G.M., Webber, J.T., Wyss-Coray, T., Zanini, F., Beachy, P.A., Chan, C.K.F., de Morree, A., George, B.M., Gulati, G.S., Hang, Y., Huang, K.C., Iram, T., Isobe, T., Kershner, A.M., Kiss, B.M., Kong, W., Li, G., Li, Q., Liu, L., Lu, W.-J., Nabhan, A.N., Ng, K.M., Nguyen, P.K., Peng, W.C., Rulifson, E.J., Schaum, N., Sikandar, S.S., Sinha, R., Szade, K., Travaglini, K.J., Tropini, C., Wang, B.M., Weinberg, K., Wosczyna, M.N., Wu, S.M., Yousef, H., Barres, B.A., Beachy, P.A., Chan, C.K.F., Clarke, M.F., Darmanis, S., Huang, K.C., Karkanias, J., Kim, S.K., Krasnow, M.A., Kumar, M.E., Kuo, C.S., May, A.P., Metzger, R.J., Neff, N.F., Nusse, R., Nguyen, P.K., Rando, T.A., Sonnenburg, J., Wang, B.M., Weinberg, K., Weissman, I.L., Wu, S.M., Quake, S.R., Wyss-Coray, T., The Tabula Muris, C., Overall, c., Logistical, c., Organ, c., processing, Library, p., sequencing, Computational data, a., Cell type, a., Writing, g., Supplemental text writing, g., and Principal, i. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature *562*, 367-372.

Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat Methods *14*, 975-978.

Schiaffino, S., and Reggiani, C. (2011). Fiber types in mammalian skeletal muscles. Physiol Rev *91*, 1447-1531.

Schiaffino, S., Rossi, A.C., Smerdu, V., Leinwand, L.A., and Reggiani, C. (2015). Developmental myosins: expression patterns and functional significance. Skeletal muscle *5*, 22-22.

Schug, J., Schuller, W.P., Kappen, C., Salbaum, J.M., Bucan, M., and Stoeckert, C.J., Jr. (2005). Promoter features related to tissue specificity as measured by Shannon entropy. Genome Biol *6*, R33.

Shadrina, A.S., Sharapov, S.Z., Shashkova, T.I., and Tsepilov, Y.A. (2019). Varicose veins of lower extremities: Insights from the first large-scale genetic study. PLoS Genet *15*, e1008110-e1008110.

Shah, S., Henry, A., Roselli, C., Lin, H., Sveinbjörnsson, G., Fatemifar, G., Hedman, Å.K., Wilk, J.B., Morley, M.P., Chaffin, M.D., Helgadottir, A., Verweij, N., Dehghan, A., Almgren, P., Andersson, C., Aragam, K.G., Ärnlöv, J., Backman, J.D., Biggs, M.L., Bloom, H.L., Brandimarto, J., Brown, M.R., Buckbinder, L., Carey, D.J., Chasman, D.I., Chen, X., Chen, X., Chung, J., Chutkow, W., Cook, J.P., Delgado, G.E., Denaxas, S., Doney, A.S., Dörr, M., Dudley, S.C., Dunn, M.E., Engström, G., Esko, T., Felix, S.B., Finan, C., Ford, I., Ghanbari, M., Ghasemi, S., Giedraitis, V., Giulianini, F., Gottdiener, J.S., Gross, S., Guðbjartsson, D.F., Gutmann, R., Haggerty, C.M., van der Harst, P., Hyde, C.L., Ingelsson, E., Jukema, J.W., Kavousi, M., Khaw, K.-T., Kleber, M.E., Køber, L., Koekemoer, A., Langenberg, C., Lind, L., Lindgren, C.M., London, B., Lotta, L.A., Lovering, R.C., Luan, J.a., Magnusson, P., Mahajan, A., Margulies, K.B., März, W., Melander, O., Mordi, I.R., Morgan, T., Morris, A.D., Morris, A.P., Morrison, A.C., Nagle, M.W., Nelson, C.P., Niessner, A., Niiranen, T., O'Donoghue, M.L., Owens, A.T., Palmer, C.N.A., Parry, H.M., Perola, M., Portilla-Fernandez, E., Psaty, B.M., Abecasis, G., Backman, J., Bai, X., Balasubramanian, S., Banerjee, N., Baras, A., Barnard, L., Beechert, C., Blumenfeld, A., Cantor, M., Chai, Y., Chung, J., Coppola, G., Damask, A., Dewey, F., Economides, A., Eom, G., Forsythe, C., Fuller, E.D., Gu, Z., Gurski, L., Guzzardo, P.M., Habegger, L., Hahn, Y., Hawes, A., van Hout, C., Jones, M.B., Khalid, S., Lattari, M., Li, A., Lin, N., Liu, D., Lopez, A., Manoochehri, K., Marchini, J., Marcketta, A., Maxwell, E.K., McCarthy, S., Mitnaul, L.J., O'Dushlaine, C., Overton, J.D., Padilla, M.S., Paulding, C., Penn, J., Pradhan, M., Reid, J.G., Schleicher, T.D., Schurmann, C., Shuldiner, A., Staples, J.C., Sun, D., Toledo, K., Ulloa, R.H., Widom, L., Wolf, S.E., Yadav, A., Ye, B., Rice, K.M., Ridker, P.M., Romaine, S.P.R., Rotter, J.I., Salo, P., Salomaa, V., van Setten, J., Shalaby, A.A., Smelser, D.T., Smith, N.L., Stender, S., Stott, D.J., Svensson, P., Tammesoo, M.-L., Taylor, K.D., Teder-Laving, M., Teumer, A., Thorgeirsson, G., Thorsteinsdottir, U., Torp-Pedersen, C., Trompet, S., Tyl, B., Uitterlinden, A.G., Veluchamy, A., Völker, U., Voors, A.A., Wang, X., Wareham, N.J., Waterworth, D., Weeke, P.E., Weiss, R., Wiggins, K.L., Xing, H., Yerges-Armstrong, L.M., Yu, B., Zannad, F., Zhao, J.H., Hemingway, H., Samani, N.J., McMurray, J.J.V., Yang, J., Visscher, P.M., Newton-Cheh, C., Malarstig, A., Holm, H., Lubitz, S.A., Sattar, N., Holmes, M.V., Cappola, T.P., Asselbergs, F.W., Hingorani, A.D., Kuchenbaecker, K., Ellinor, P.T., Lang, C.C., Stefansson, K., Smith, J.G., Vasan, R.S., Swerdlow, D.I., Lumbers, R.T., and Regeneron Genetics, C. (2020). Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. Nature Communications *11*, 163.

Shanahan, C.M., Weissberg, P.L., and Metcalfe, J.C. (1993). Isolation of gene markers of differentiated and proliferating vascular smooth muscle cells. Circulation Research *73*, 193-204.

Sheikh, F., Lyon, R.C., and Chen, J. (2015). Functions of myosin light chain-2 (MYL2) in cardiac muscle and disease. Gene *569*, 14-20.

Shen, T., Aneas, I., Sakabe, N., Dirschinger, R.J., Wang, G., Smemo, S., Westlund, J.M., Cheng, H., Dalton, N., Gu, Y., Boogerd, C.J., Cai, C.-l., Peterson, K., Chen, J., Nobrega, M.A., and Evans, S.M. (2011). Tbx20 regulates a genetic program essential to adult mouse cardiomyocyte function. The Journal of Clinical Investigation *121*, 4640-4654.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., and Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116-120.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet *15*, 272-286.

Shrine, N., Guyatt, A.L., Erzurumluoglu, A.M., Jackson, V.E., Hobbs, B.D., Melbourne, C.A., Batini, C., Fawcett, K.A., Song, K., Sakornsakolpat, P., Li, X., Boxall, R., Reeve, N.F., Obeidat, M., Zhao, J.H., Wielscher, M., Weiss, S., Kentistou, K.A., Cook, J.P., Sun, B.B., Zhou, J., Hui, J., Karrasch, S., Imboden, M., Harris, S.E., Marten, J., Enroth, S., Kerr, S.M., Surakka, I., Vitart, V., Lehtimäki, T., Allen, R.J., Bakke, P.S., Beaty, T.H., Bleecker, E.R., Bossé, Y., Brandsma, C.A., Chen, Z., Crapo, J.D., Danesh, J., DeMeo, D.L., Dudbridge, F., Ewert, R., Gieger, C., Gulsvik, A., Hansell, A.L., Hao, K., Hoffman, J.D., Hokanson, J.E., Homuth, G., Joshi, P.K., Joubert, P., Langenberg, C., Li, X., Li, L., Lin, K., Lind, L., Locantore, N., Luan, J., Mahajan, A., Maranville, J.C., Murray, A., Nickle, D.C., Packer, R., Parker, M.M., Paynton, M.L., Porteous, D.J., Prokopenko, D., Qiao, D., Rawal, R., Runz, H., Sayers, I., Sin, D.D., Smith, B.H., Soler Artigas, M., Sparrow, D., Tal-Singer, R., Timmers, P., Van den Berge, M., Whittaker, J.C., Woodruff, P.G., Yerges-Armstrong, L.M., Troyanskaya, O.G., Raitakari, O.T., Kähönen, M., Polašek, O., Gyllensten, U., Rudan, I., Deary, I.J., Probst-Hensch, N.M., Schulz, H., James, A.L., Wilson, J.F., Stubbe, B., Zeggini, E., Jarvelin, M.R., Wareham, N., Silverman, E.K., Hayward, C., Morris, A.P., Butterworth, A.S., Scott, R.A., Walters, R.G., Meyers, D.A., Cho, M.H., Strachan, D.P., Hall, I.P., Tobin, M.D., and Wain, L.V. (2019). New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. Nat Genet *51*, 481-493.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., Weinstock, G.M., Wilson, R.K., Gibbs, R.A., Kent, W.J., Miller, W., and Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome research *15*, 1034-1050.

Singh, M.K., Christoffels, V.M., Dias, J.M., Trowe, M.-O., Petry, M., Schuster-Gossler, K., Bürger, A., Ericson, J., and Kispert, A. (2005). Tbx20 is essential for cardiac chamber differentiation and repression of Tbx2. Development *132*, 2697-2707.

Sinnamon, J.R., Torkenczy, K.A., Linhoff, M.W., Vitak, S.A., Mulqueen, R.M., Pliner, H.A., Trapnell, C., Steemers, F.J., Mandel, G., and Adey, A.C. (2019). The accessible chromatin landscape of the murine hippocampus at single-cell resolution. Genome Research *29*, 857-869.

Skelly, D.A., Squiers, G.T., McLellan, M.A., Bolisetty, M.T., Robson, P., Rosenthal, N.A., and Pinto, A.R. (2018). Single-Cell Transcriptional Profiling Reveals Cellular Diversity and Intercommunication in the Mouse Heart. Cell Rep *22*, 600-610.

Slatkin, M. (2008). Linkage disequilibrium — understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics *9*, 477-485.

Smemo, S., Campos, L.C., Moskowitz, I.P., Krieger, J.E., Pereira, A.C., and Nobrega, M.A. (2012). Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. Hum Mol Genet *21*, 3255-3263.

Smyth, J.W., Hong, T.T., Gao, D., Vogan, J.M., Jensen, B.C., Fong, T.S., Simpson, P.C., Stainier, D.Y., Chi, N.C., and Shaw, R.M. (2010). Limited forward trafficking of connexin 43 reduces cell-cell coupling in stressed human and mouse myocardium. J Clin Invest *120*, 266-279.

Southern, B.D., Grove, L.M., Rahaman, S.O., Abraham, S., Scheraga, R.G., Niese, K.A., Sun, H., Herzog, E.L., Liu, F., Tschumperlin, D.J., Egelhoff, T.T., Rosenfeld, S.S., and Olman, M.A. (2016). Matrix-driven Myosin II Mediates the Pro-fibrotic Fibroblast Phenotype. The Journal of biological chemistry *291*, 6083-6095.

Spurrell, C.H., Barozzi, I., Mannion, B.J., Blow, M.J., Fukuda-Yuzawa, Y., Afzal, S.Y., Akiyama, J.A., Afzal, V., Tran, S., Plajzer-Frick, I., Novak, C.S., Kato, M., Lee, E., Garvin, T.H., Pham, Q.T., Harrington, A.N., Lisgo, S., Bristow, J., Cappola, T.P., Morley, M.P., Margulies, K.B., Pennacchio, L.A., Dickel, D.E., and Visel, A. (2019). Genome-Wide Fetalization of Enhancer Architecture in Heart Disease. bioRxiv, 591362.

Squair, J.W., Gautier, M., Kathe, C., Anderson, M.A., James, N.D., Hutson, T.H., Hudelle, R., Qaiser, T., Matson, K.J.E., Barraud, Q., Levine, A.J., La Manno, G., Skinnider, M.A., and Courtine, G. (2021). Confronting false discoveries in single-cell differential expression. Nature Communications *12*, 5692.

Stranger, B.E., Brigham, L.E., Hasz, R., Hunter, M., Johns, C., Johnson, M., Kopen, G., Leinweber, W.F., Lonsdale, J.T., McDonald, A., Mestichelli, B., Myer, K., Roe, B., Salvatore, M., Shad, S., Thomas, J.A., Walters, G., Washington, M., Wheeler, J., Bridge, J., Foster, B.A., Gillard, B.M., Karasik, E., Kumar, R., Miklos, M., Moser, M.T., Jewell, S.D., Montroy, R.G., Rohrer, D.C., Valley, D.R., Davis, D.A., Mash, D.C., Gould, S.E., Guan, P., Koester, S., Little, A.R., Martin, C., Moore, H.M., Rao, A., Struewing, J.P., Volpi, S., Hansen, K.D., Hickey, P.F., Rizzardi, L.F., Hou, L., Liu, Y., Molinie, B., Park, Y., Rinaldi, N., Wang, L., Van Wittenberghe, N., Claussnitzer, M., Gelfand, E.T., Li, Q., Linder, S., Zhang, R., Smith, K.S., Tsang, E.K., Chen, L.S., Demanelis, K., Doherty, J.A., Jasmine, F., Kibriya, M.G., Jiang, L., Lin, S., Wang, M., Jian, R., Li, X., Chan, J., Bates, D., Diegel, M., Halow, J., Haugen, E., Johnson, A., Kaul, R., Lee, K., Maurano, M.T., Nelson, J., Neri, F.J., Sandstrom, R., Fernando, M.S., Linke, C., Oliva, M., Skol, A., Wu, F., Akey, J.M., Feinberg, A.P., Li, J.B., Pierce, B.L., Stamatoyannopoulos, J.A., Tang, H., Ardlie, K.G., Kellis, M., Snyder, M.P., Montgomery, S.B., and e, G.P. (2017). Enhancing GTEx by bridging the gaps between genotype, gene expression, and disease. Nature Genetics *49*, 1664-1670.

Strawbridge, R.J., Dupuis, J., Prokopenko, I., Barker, A., Ahlqvist, E., Rybin, D., Petrie, J.R., Travers, M.E., Bouatia-Naji, N., Dimas, A.S., Nica, A., Wheeler, E., Chen, H., Voight, B.F., Taneera, J., Kanoni, S., Peden, J.F., Turrini, F., Gustafsson, S., Zabena, C., Almgren, P., Barker, D.J., Barnes, D., Dennison, E.M., Eriksson, J.G., Eriksson, P., Eury, E., Folkersen, L., Fox, C.S., Frayling, T.M., Goel, A., Gu, H.F., Horikoshi, M., Isomaa, B., Jackson, A.U., Jameson, K.A., Kajantie, E., Kerr-Conte, J., Kuulasmaa, T., Kuusisto, J., Loos, R.J., Luan, J., Makrilakis, K., Manning, A.K., Martínez-Larrad, M.T., Narisu, N., Nastase Mannila, M., Ohrvik, J., Osmond, C., Pascoe, L., Payne, F., Sayer, A.A., Sennblad, B., Silveira, A., Stancáková, A., Stirrups, K., Swift, A.J., Syvänen, A.C., Tuomi, T., van 't Hooft, F.M., Walker, M., Weedon, M.N., Xie, W., Zethelius, B., Ongen, H., Mälarstig, A., Hopewell, J.C., Saleheen, D., Chambers, J., Parish, S., Danesh, J., Kooner, J., Ostenson, C.G., Lind, L., Cooper, C.C., Serrano-Ríos, M., Ferrannini, E., Forsen, T.J., Clarke, R., Franzosi, M.G., Seedorf, U., Watkins, H., Froguel, P., Johnson, P., Deloukas, P., Collins, F.S., Laakso, M., Dermitzakis, E.T., Boehnke, M., McCarthy, M.I., Wareham, N.J., Groop, L., Pattou, F., Gloyn, A.L., Dedoussis, G.V., Lyssenko, V., Meigs, J.B., Barroso, I., Watanabe, R.M., Ingelsson, E., Langenberg, C., Hamsten, A., and Florez, J.C. (2011). Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. Diabetes *60*, 2624-2634.

Stuart, C.A., Stone, W.L., Howell, M.E., Brannon, M.F., Hall, H.K., Gibson, A.L., and Stone, M.H. (2016). Myosin content of individual human muscle fibers isolated by laser capture microdissection. Am J Physiol Cell Physiol *310*, C381-389.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902 e1821.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., and Collins, R. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med *12*, e1001779.

Sweet, M.E., Cocciolo, A., Slavov, D., Jones, K.L., Sweet, J.R., Graw, S.L., Reece, T.B., Ambardekar, A.V., Bristow, M.R., Mestroni, L., and Taylor, M.R.G. (2018). Transcriptome analysis of human heart failure reveals dysregulated cell adhesion in dilated cardiomyopathy and activated immune pathways in ischemic heart failure. BMC Genomics *19*, 812.

Tachmazidou, I., Hatzikotoulas, K., Southam, L., Esparza-Gordillo, J., Haberland, V., Zheng, J., Johnson, T., Koprulu, M., Zengini, E., Steinberg, J., Wilkinson, J.M., Bhatnagar, S., Hoffman, J.D., Buchan, N., Süveges, D., Yerges-Armstrong, L., Smith, G.D., Gaunt, T.R., Scott, R.A., McCarthy, L.C., Zeggini, E., and arc, O.C. (2019). Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. Nature Genetics *51*, 230-236.

Tadros, R., Francis, C., Xu, X., Vermeer, A.M.C., Harper, A.R., Huurman, R., Kelu Bisabu, K., Walsh, R., Hoorntje, E.T., te Rijdt, W.P., Buchan, R.J., van Velzen, H.G., van Slegtenhorst, M.A., Vermeulen, J.M., Offerhaus, J.A., Bai, W., de Marvao, A., Lahrouchi, N., Beekman, L., Karper, J.C., Veldink, J.H., Kayvanpour, E., Pantazis, A., Baksi, A.J., Whiffin, N., Mazzarotto, F., Sloane, G., Suzuki, H., Schneider-Luftman, D., Elliott, P., Richard, P., Ader, F., Villard, E., Lichtner, P.,

Meitinger, T., Tanck, M.W.T., van Tintelen, J.P., Thain, A., McCarty, D., Hegele, R.A., Roberts, J.D., Amyot, J., Dubé, M.-P., Cadrin-Tourigny, J., Giraldeau, G., L'Allier, P.L., Garceau, P., Tardif, J.-C., Boekholdt, S.M., Lumbers, R.T., Asselbergs, F.W., Barton, P.J.R., Cook, S.A., Prasad, S.K., O'Regan, D.P., van der Velden, J., Verweij, K.J.H., Talajic, M., Lettre, G., Pinto, Y.M., Meder, B., Charron, P., de Boer, R.A., Christiaans, I., Michels, M., Wilde, A.A.M., Watkins, H., Matthews, P.M., Ware, J.S., and Bezzina, C.R. (2021). Shared genetic pathways contribute to risk of hypertrophic and dilated cardiomyopathies with opposite directions of effect. Nature Genetics *53*, 128-134.

Taegtmeyer, H., Sen, S., and Vela, D. (2010). Return to the fetal gene program: a suggested metabolic link to gene expression in the heart. Ann N Y Acad Sci *1188*, 191-198.

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. Nature Reviews Genetics *20*, 467-484.

Tan, F.-L., Moravec, C.S., Li, J., Apperson-Hansen, C., McCarthy, P.M., Young, J.B., and Bond, M. (2002). The gene expression fingerprint of human heart failure. Proceedings of the National Academy of Sciences *99*, 11387.

Tan, W.L.W., Anene-Nzelu, C.G., Wong, E., Lee, M., Tan, H.S., Tang, S.J., Perrin, A.M.Y., Wu, K.X., Zheng, W.H., Ashburn, R.J., Pan, B., Lee, M.Y., Autio, M.I., Morley, M., Tam, W.L., Cheung, C., Margulies, K.B., Chen, L., Cappola, T.P., Loh, M., Chambers, J., Prabhakar, S., and Foo, R.S. (2020). Epigenomes of Human Hearts Reveal New Genetic Variants Relevant for Cardiac Disease and Phenotype. Circ Res.

Tang, J., Zhang, H., He, L., Huang, X., Li, Y., Pu, W., Yu, W., Zhang, L., Cai, D., Lui, K.O., and Zhou, B. (2018). Genetic Fate Mapping Defines the Vascular Potential of Endocardial Cells in the Adult Heart. Circ Res *122*, 984-993.

Teumer, A., Chaker, L., Groeneweg, S., Li, Y., Di Munno, C., Barbieri, C., Schultheiss, U.T., Traglia, M., Ahluwalia, T.S., Akiyama, M., Appel, E.V.R., Arking, D.E., Arnold, A., Astrup, A., Beekman, M., Beilby, J.P., Bekaert, S., Boerwinkle, E., Brown, S.J., De Buyzere, M., Campbell, P.J., Ceresini, G., Cerqueira, C., Cucca, F., Deary, I.J., Deelen, J., Eckardt, K.U., Ekici, A.B., Eriksson, J.G., Ferrrucci, L., Fiers, T., Fiorillo, E., Ford, I., Fox, C.S., Fuchsberger, C., Galesloot, T.E., Gieger, C., Gögele, M., De Grandi, A., Grarup, N., Greiser, K.H., Haljas, K., Hansen, T., Harris, S.E., van Heemst, D., den Heijer, M., Hicks, A.A., den Hollander, W., Homuth, G., Hui, J., Ikram, M.A., Ittermann, T., Jensen, R.A., Jing, J., Jukema, J.W., Kajantie, E., Kamatani, Y., Kasbohm, E., Kaufman, J.M., Kiemeney, L.A., Kloppenburg, M., Kronenberg, F., Kubo, M., Lahti, J., Lapauw, B., Li, S., Liewald, D.C.M., Lim, E.M., Linneberg, A., Marina, M., Mascalzoni, D., Matsuda, K., Medenwald, D., Meisinger, C., Meulenbelt, I., De Meyer, T., Meyer Zu Schwabedissen, H.E., Mikolajczyk, R., Moed, M., Netea-Maier, R.T., Nolte, I.M., Okada, Y., Pala, M., Pattaro, C., Pedersen, O., Petersmann, A., Porcu, E., Postmus, I., Pramstaller, P.P., Psaty, B.M., Ramos, Y.F.M., Rawal, R., Redmond, P., Richards, J.B., Rietzschel, E.R., Rivadeneira, F., Roef, G., Rotter, J.I., Sala, C.F., Schlessinger, D., Selvin, E., Slagboom, P.E., Soranzo, N., Sørensen, T.I.A., Spector, T.D., Starr, J.M., Stott, D.J., Taes, Y., Taliun, D., Tanaka, T., Thuesen, B., Tiller, D., Toniolo, D., Uitterlinden, A.G., Visser, W.E., Walsh, J.P., Wilson, S.G., Wolffenbuttel, B.H.R., Yang, Q., Zheng, H.F., Cappola, A., Peeters, R.P., Naitza, S., Völzke, H.,

Sanna, S., Köttgen, A., Visser, T.J., and Medici, M. (2018). Genome-wide analyses identify a role for SLC17A4 and AADAT in thyroid hormone regulation. Nat Commun *9*, 4455.

Teumer, A., Tin, A., Sorice, R., Gorski, M., Yeo, N.C., Chu, A.Y., Li, M., Li, Y., Mijatovic, V., Ko, Y.A., Taliun, D., Luciani, A., Chen, M.H., Yang, Q., Foster, M.C., Olden, M., Hiraki, L.T., Tayo, B.O., Fuchsberger, C., Dieffenbach, A.K., Shuldiner, A.R., Smith, A.V., Zappa, A.M., Lupo, A., Kollerits, B., Ponte, B., Stengel, B., Kramer, B.K., Paulweber, B., Mitchell, B.D., Hayward, C., Helmer, C., Meisinger, C., Gieger, C., Shaffer, C.M., Muller, C., Langenberg, C., Ackermann, D., Siscovick, D., Dcct/Edic, Boerwinkle, E., Kronenberg, F., Ehret, G.B., Homuth, G., Waeber, G., Navis, G., Gambaro, G., Malerba, G., Eiriksdottir, G., Li, G., Wichmann, H.E., Grallert, H., Wallaschofski, H., Volzke, H., Brenner, H., Kramer, H., Mateo Leach, I., Rudan, I., Hillege, H.L., Beckmann, J.S., Lambert, J.C., Luan, J., Zhao, J.H., Chalmers, J., Coresh, J., Denny, J.C., Butterbach, K., Launer, L.J., Ferrucci, L., Kedenko, L., Haun, M., Metzger, M., Woodward, M., Hoffman, M.J., Nauck, M., Waldenberger, M., Pruijm, M., Bochud, M., Rheinberger, M., Verweij, N., Wareham, N.J., Endlich, N., Soranzo, N., Polasek, O., van der Harst, P., Pramstaller, P.P., Vollenweider, P., Wild, P.S., Gansevoort, R.T., Rettig, R., Biffar, R., Carroll, R.J., Katz, R., Loos, R.J., Hwang, S.J., Coassin, S., Bergmann, S., Rosas, S.E., Stracke, S., Harris, T.B., Corre, T., Zeller, T., Illig, T., Aspelund, T., Tanaka, T., Lendeckel, U., Volker, U., Gudnason, V., Chouraki, V., Koenig, W., Kutalik, Z., O'Connell, J.R., Parsa, A., Heid, I.M., Paterson, A.D., de Boer, I.H., Devuyst, O., Lazar, J., Endlich, K., Susztak, K., Tremblay, J., Hamet, P., Jacob, H.J., Boger, C.A., Fox, C.S., Pattaro, C., and Kottgen, A. (2016). Genome-wide Association Studies Identify Genetic Loci Associated With Albuminuria in Diabetes. Diabetes *65*, 803-817.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., Garg, K., John, S., Sandstrom, R., Bates, D., Boatman, L., Canfield, T.K., Diegel, M., Dunn, D., Ebersol, A.K., Frum, T., Giste, E., Johnson, A.K., Johnson, E.M., Kutyavin, T., Lajoie, B., Lee, B.K., Lee, K., London, D., Lotakis, D., Neph, S., Neri, F., Nguyen, E.D., Qu, H., Reynolds, A.P., Roach, V., Safi, A., Sanchez, M.E., Sanyal, A., Shafer, A., Simon, J.M., Song, L., Vong, S., Weaver, M., Yan, Y., Zhang, Z., Zhang, Z., Lenhard, B., Tewari, M., Dorschner, M.O., Hansen, R.S., Navas, P.A., Stamatoyannopoulos, G., Iyer, V.R., Lieb, J.D., Sunyaev, S.R., Akey, J.M., Sabo, P.J., Kaul, R., Furey, T.S., Dekker, J., Crawford, G.E., and Stamatoyannopoulos, J.A. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75-82.

Tin, A., Marten, J., Halperin Kuhns, V.L., Li, Y., Wuttke, M., Kirsten, H., Sieber, K.B., Qiu, C., Gorski, M., Yu, Z., Giri, A., Sveinbjornsson, G., Li, M., Chu, A.Y., Hoppmann, A., O'Connor, L.J., Prins, B., Nutile, T., Noce, D., Akiyama, M., Cocca, M., Ghasemi, S., van der Most, P.J., Horn, K., Xu, Y., Fuchsberger, C., Sedaghat, S., Afaq, S., Amin, N., Ärnlöv, J., Bakker, S.J.L., Bansal, N., Baptista, D., Bergmann, S., Biggs, M.L., Biino, G., Boerwinkle, E., Bottinger, E.P., Boutin, T.S., Brumat, M., Burkhardt, R., Campana, E., Campbell, A., Campbell, H., Carroll, R.J., Catamo, E., Chambers, J.C., Ciullo, M., Concas, M.P., Coresh, J., Corre, T., Cusi, D., Felicita, S.C., de Borst, M.H., De Grandi, A., de Mutsert, R., de Vries, A.P.J., Delgado, G., Demirkan, A., Devuyst, O., Dittrich, K., Eckardt, K.U., Ehret, G., Endlich, K., Evans, M.K., Gansevoort, R.T., Gasparini, P., Giedraitis, V., Gieger, C., Girotto, G., Gögele, M., Gordon, S.D., Gudbjartsson, D.F., Gudnason, V., Haller, T., Hamet, P., Harris, T.B., Hayward, C., Hicks, A.A., Hofer, E., Holm, H., Huang, W., Hutri-Kähönen, N., Hwang, S.J., Ikram, M.A., Lewis, R.M., Ingelsson, E., Jakobsdottir, J., Jonsdottir, I., Jonsson, H., Joshi, P.K., Josyula, N.S., Jung, B., Kähönen, M.,

Kamatani, Y., Kanai, M., Kerr, S.M., Kiess, W., Kleber, M.E., Koenig, W., Kooner, J.S., Körner, A., Kovacs, P., Krämer, B.K., Kronenberg, F., Kubo, M., Kühnel, B., La Bianca, M., Lange, L.A., Lehne, B., Lehtimäki, T., Liu, J., Loeffler, M., Loos, R.J.F., Lyytikäinen, L.P., Magi, R., Mahajan, A., Martin, N.G., März, W., Mascalzoni, D., Matsuda, K., Meisinger, C., Meitinger, T., Metspalu, A., Milaneschi, Y., O'Donnell, C.J., Wilson, O.D., Gaziano, J.M., Mishra, P.P., Mohlke, K.L., Mononen, N., Montgomery, G.W., Mook-Kanamori, D.O., Müller-Nurasyid, M., Nadkarni, G.N., Nalls, M.A., Nauck, M., Nikus, K., Ning, B., Nolte, I.M., Noordam, R., O'Connell, J.R., Olafsson, I., Padmanabhan, S., Penninx, B., Perls, T., Peters, A., Pirastu, M., Pirastu, N., Pistis, G., Polasek, O., Ponte, B., Porteous, D.J., Poulain, T., Preuss, M.H., Rabelink, T.J., Raffield, L.M., Raitakari, O.T., Rettig, R., Rheinberger, M., Rice, K.M., Rizzi, F., Robino, A., Rudan, I., Krajcoviechova, A., Cifkova, R., Rueedi, R., Ruggiero, D., Ryan, K.A., Saba, Y., Salvi, E., Schmidt, H., Schmidt, R., Shaffer, C.M., Smith, A.V., Smith, B.H., Spracklen, C.N., Strauch, K., Stumvoll, M., Sulem, P., Tajuddin, S.M., Teren, A., Thiery, J., Thio, C.H.L., Thorsteinsdottir, U., Toniolo, D., Tönjes, A., Tremblay, J., Uitterlinden, A.G., Vaccargiu, S., van der Harst, P., van Duijn, C.M., Verweij, N., Völker, U., Vollenweider, P., Waeber, G., Waldenberger, M., Whitfield, J.B., Wild, S.H., Wilson, J.F., Yang, Q., Zhang, W., Zonderman, A.B., Bochud, M., Wilson, J.G., Pendergrass, S.A., Ho, K., Parsa, A., Pramstaller, P.P., Psaty, B.M., Böger, C.A., Snieder, H., Butterworth, A.S., Okada, Y., Edwards, T.L., Stefansson, K., Susztak, K., Scholz, M., Heid, I.M., Hung, A.M., Teumer, A., Pattaro, C., Woodward, O.M., Vitart, V., and Köttgen, A. (2019). Target genes, variants, tissues and transcriptional pathways influencing human serum urate levels. Nat Genet *51*, 1459-1474.

Traag, V.A., Van Dooren, P., and Nesterov, Y. (2011). Narrow scope for resolution-limit-free community detection. Phys Rev E Stat Nonlin Soft Matter Phys *84*, 016114.

Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep *9*, 5233.

Travers, J.G., Kamal, F.A., Robbins, J., Yutzey, K.E., and Blaxall, B.C. (2016). Cardiac Fibrosis: The Fibroblast Awakens. Circulation research *118*, 1021-1040.

Trevino, A.E., Müller, F., Andersen, J., Sundaram, L., Kathiria, A., Shcherbina, A., Farh, K., Chang, H.Y., Pașca, A.M., Kundaje, A., Pașca, S.P., and Greenleaf, W.J. (2021). Chromatin and gene-regulatory dynamics of the developing human cerebral cortex at single-cell resolution. Cell.

Tucker, N.R., Chaffin, M., Fleming, S.J., Hall, A.W., Parsons, V.A., Bedi, K.C., Jr., Akkad, A.D., Herndon, C.N., Arduini, A., Papangeli, I., Roselli, C., Aguet, F., Choi, S.H., Ardlie, K.G., Babadi, M., Margulies, K.B., Stegmann, C.M., and Ellinor, P.T. (2020). Transcriptional and Cellular Diversity of the Human Heart. Circulation.

Uffelmann, E., Huang, Q.Q., Munung, N.S., de Vries, J., Okada, Y., Martin, A.R., Martin, H.C., Lappalainen, T., and Posthuma, D. (2021). Genome-wide association studies. Nature Reviews Methods Primers *1*, 59.

van den Boogaard, M., Smemo, S., Burnicka-Turek, O., Arnolds, D.E., van de Werken, H.J.G., Klous, P., McKean, D., Muehlschlegel, J.D., Moosmann, J., Toka, O., Yang, X.H., Koopmann, T.T., Adriaens, M.E., Bezzina, C.R., de Laat, W., Seidman, C., Seidman, J.G., Christoffels, V.M.,

Nobrega, M.A., Barnett, P., and Moskowitz, I.P. (2014). A common genetic variant within SCN10A modulates cardiac SCN5A expression. J Clin Invest *124*, 1844-1852.

van Dijk, E.L., Jaszczyszyn, Y., Naquin, D., and Thermes, C. (2018). The Third Revolution in Sequencing Technology. Trends Genet *34*, 666-681.

Vanlandewijck, M., He, L., Mäe, M.A., Andrae, J., Ando, K., Del Gaudio, F., Nahar, K., Lebouvier, T., Laviña, B., Gouveia, L., Sun, Y., Raschperger, E., Räsänen, M., Zarb, Y., Mochizuki, N., Keller, A., Lendahl, U., and Betsholtz, C. (2018). A molecular atlas of cell types and zonation in the brain vasculature. Nature *554*, 475-480.

Veevers, J., Farah, E.N., Corselli, M., Witty, A.D., Palomares, K., Vidal, J.G., Emre, N., Carson, C.T., Ouyang, K., Liu, C., van Vliet, P., Zhu, M., Hegarty, J.M., Deacon, D.C., Grinstein, J.D., Dirschinger, R.J., Frazer, K.A., Adler, E.D., Knowlton, K.U., Chi, N.C., Martin, J.C., Chen, J., and Evans, S.M. (2018). Cell-Surface Marker Signature for Enrichment of Ventricular Cardiomyocytes Derived from Human Embryonic Stem Cells. Stem Cell Reports *11*, 828-841.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., Gocayne, J.D., Amanatides, P., Ballew, R.M., Huson, D.H., Wortman, J.R., Zhang, Q., Kodira, C.D., Zheng, X.H., Chen, L., Skupski, M., Subramanian, G., Thomas, P.D., Zhang, J., Miklos, G.L.G., Nelson, C., Broder, S., Clark, A.G., Nadeau, J., McKusick, V.A., Zinder, N., Levine, A.J., Roberts, R.J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Francesco, V.D., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A.E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T.J., Higgins, M.E., Ji, R.-R., Ke, Z., Ketchum, K.A., Lai, Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G.V., Milshina, N., Moore, H.M., Naik, A.K., Narayan, V.A., Neelam, B., Nusskern, D., Rusch, D.B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z.Y., Wang, A., Wang, X., Wang, J., Wei, M.-H., Wides, R., Xiao, C., Yan, C., Yao, A., Ye, J., Zhan, M., Zhang, W., Zhang, H., Zhao, Q., Zheng, L., Zhong, F., Zhong, W., Zhu, S.C., Zhao, S., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M.L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y.-H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N.N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J.F., Guigó, R., Campbell, M.J., Sjolander, K.V., Karlak, B., Kejariwal, A., Mi, H., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y.-H., Coyne, M., Dahlke, C., Mays, A.D., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M.,

Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., Nodell, M., Pan, S., Peck, J., Peterson, M., Rowe, W., Sanders, R., Scott, J., Simpson, M., Smith, T., Sprague, A., Stockwell, T., Turner, R., Venter, E., Wang, M., Wen, M., Wu, D., Wu, M., Xia, A., Zandieh, A., and Zhu, X. (2001). The Sequence of the Human Genome. Science *291*, 1304-1351.

Vierstra, J., Lazar, J., Sandstrom, R., Halow, J., Lee, K., Bates, D., Diegel, M., Dunn, D., Neri, F., Haugen, E., Rynes, E., Reynolds, A., Nelson, J., Johnson, A., Frerker, M., Buckley, M., Kaul, R., Meuleman, W., and Stamatoyannopoulos, J.A. (2020). Global reference mapping of human transcription factor footprints. Nature *583*, 729-736.

Visel, A., Minovitsky, S., Dubchak, I., and Pennacchio, L.A. (2007). VISTA Enhancer Browser--a database of tissue-specific human enhancers. Nucleic acids research *35*, D88-92.

Wakefield, J. (2009). Bayes factors for genome-wide association studies: comparison with P-values. Genetic epidemiology *33*, 79-86.

Wang, A., Chiou, J., Poirion, O.B., Buchanan, J., Valdez, M.J., Verheyden, J.M., Hou, X., Kudtarkar, P., Narendra, S., Newsome, J.M., Guo, M., Faddah, D.A., Zhang, K., Young, R.E., Barr, J., Sajti, E., Misra, R., Huyck, H., Rogers, L., Poole, C., Whitsett, J.A., Pryhuber, G., Xu, Y., Gaulton, K.J., Preissl, S., Sun, X., and Consortium, N.L. (2020). Single-cell multiomic profiling of human lungs reveals cell-type-specific and age-dynamic control of SARS-CoV2 host genes. Elife *9*.

Wang, Q., Xiong, H., Ai, S., Yu, X., Liu, Y., Zhang, J., and He, A. (2019). CoBATCH for High-Throughput Single-Cell Epigenomic Profiling. Mol Cell *76*, 206-216 e207.

Warrington, N.M., Beaumont, R.N., Horikoshi, M., Day, F.R., Helgeland, Ø., Laurin, C., Bacelis, J., Peng, S., Hao, K., Feenstra, B., Wood, A.R., Mahajan, A., Tyrrell, J., Robertson, N.R., Rayner, N.W., Qiao, Z., Moen, G.H., Vaudel, M., Marsit, C.J., Chen, J., Nodzenski, M., Schnurr, T.M., Zafarmand, M.H., Bradfield, J.P., Grarup, N., Kooijman, M.N., Li-Gao, R., Geller, F., Ahluwalia, T.S., Paternoster, L., Rueedi, R., Huikari, V., Hottenga, J.J., Lyytikäinen, L.P., Cavadino, A., Metrustry, S., Cousminer, D.L., Wu, Y., Thiering, E., Wang, C.A., Have, C.T., Vilor-Tejedor, N., Joshi, P.K., Painter, J.N., Ntalla, I., Myhre, R., Pitkänen, N., van Leeuwen, E.M., Joro, R., Lagou, V., Richmond, R.C., Espinosa, A., Barton, S.J., Inskip, H.M., Holloway, J.W., Santa-Marina, L., Estivill, X., Ang, W., Marsh, J.A., Reichetzeder, C., Marullo, L., Hocher, B., Lunetta, K.L., Murabito, J.M., Relton, C.L., Kogevinas, M., Chatzi, L., Allard, C., Bouchard, L., Hivert, M.F., Zhang, G., Muglia, L.J., Heikkinen, J., Morgen, C.S., van Kampen, A.H.C., van Schaik, B.D.C., Mentch, F.D., Langenberg, C., Luan, J., Scott, R.A., Zhao, J.H., Hemani, G., Ring, S.M., Bennett, A.J., Gaulton, K.J., Fernandez-Tajes, J., van Zuydam, N.R., Medina-Gomez, C., de Haan, H.G., Rosendaal, F.R., Kutalik, Z., Marques-Vidal, P., Das, S., Willemsen, G., Mbarek, H., Müller-Nurasyid, M., Standl, M., Appel, E.V.R., Fonvig, C.E., Trier, C., van Beijsterveldt, C.E.M., Murcia, M., Bustamante, M., Bonas-Guarch, S., Hougaard, D.M., Mercader, J.M., Linneberg, A., Schraut, K.E., Lind, P.A., Medland, S.E., Shields, B.M., Knight, B.A., Chai, J.F., Panoutsopoulou, K., Bartels, M., Sánchez, F., Stokholm, J., Torrents, D., Vinding, R.K., Willems, S.M., Atalay, M., Chawes, B.L., Kovacs, P., Prokopenko, I., Tuke, M.A., Yaghootkar, H., Ruth, K.S., Jones, S.E.,

Loh, P.R., Murray, A., Weedon, M.N., Tönjes, A., Stumvoll, M., Michaelsen, K.F., Eloranta, A.M., Lakka, T.A., van Duijn, C.M., Kiess, W., Körner, A., Niinikoski, H., Pahkala, K., Raitakari, O.T., Jacobsson, B., Zeggini, E., Dedoussis, G.V., Teo, Y.Y., Saw, S.M., Montgomery, G.W., Campbell, H., Wilson, J.F., Vrijkotte, T.G.M., Vrijheid, M., de Geus, E., Hayes, M.G., Kadarmideen, H.N., Holm, J.C., Beilin, L.J., Pennell, C.E., Heinrich, J., Adair, L.S., Borja, J.B., Mohlke, K.L., Eriksson, J.G., Widén, E.E., Hattersley, A.T., Spector, T.D., Kähönen, M., Viikari, J.S., Lehtimäki, T., Boomsma, D.I., Sebert, S., Vollenweider, P., Sørensen, T.I.A., Bisgaard, H., Bønnelykke, K., Murray, J.C., Melbye, M., Nohr, E.A., Mook-Kanamori, D.O., Rivadeneira, F., Hofman, A., Felix, J.F., Jaddoe, V.W.V., Hansen, T., Pisinger, C., Vaag, A.A., Pedersen, O., Uitterlinden, A.G., Järvelin, M.R., Power, C., Hyppönen, E., Scholtens, D.M., Lowe, W.L., Jr., Davey Smith, G., Timpson, N.J., Morris, A.P., Wareham, N.J., Hakonarson, H., Grant, S.F.A., Frayling, T.M., Lawlor, D.A., Njølstad, P.R., Johansson, S., Ong, K.K., McCarthy, M.I., Perry, J.R.B., Evans, D.M., and Freathy, R.M. (2019). Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. Nat Genet *51*, 804-814.

Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., de Leeuw, C., Polderman, T.J.C., van der Sluis, S., Andreassen, O.A., Neale, B.M., and Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. Nat Genet *51*, 1339-1348.

Watanabe, K., Ueno, M., Kamiya, D., Nishiyama, A., Matsumura, M., Wataya, T., Takahashi, J.B., Nishikawa, S., Nishikawa, S., Muguruma, K., and Sasai, Y. (2007). A ROCK inhibitor permits survival of dissociated human embryonic stem cells. Nat Biotechnol *25*, 681-686.

Weldy, C.S., and Ashley, E.A. (2021). Towards precision medicine in heart failure. Nature Reviews Cardiology *18*, 745-762.

Wellcome Trust Case Control, C. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature *447*, 661-678.

Whitehead, A.J., Hocker, J.D., Ren, B., and Engler, A.J. (2022). Improved epicardial cardiac fibroblast generation from iPSCs. J Mol Cell Cardiol *164*, 58-68.

WHO (2017). Cardiovascular Diseases (CVDs).

Wiberg, A., Ng, M., Schmid, A.B., Smillie, R.W., Baskozos, G., Holmes, M.V., Künnapuu, K., Mägi, R., Bennett, D.L., and Furniss, D. (2019). A genome-wide association analysis identifies 16 novel susceptibility loci for carpal tunnel syndrome. Nature Communications *10*, 1030.

Wilkins, B.J., and Molkentin, J.D. (2004). Calcium-calcineurin signaling in the regulation of cardiac hypertrophy. Biochem Biophys Res Commun *322*, 1178-1191.

Wittemans, L.B.L., Lotta, L.A., Oliver-Williams, C., Stewart, I.D., Surendran, P., Karthikeyan, S., Day, F.R., Koulman, A., Imamura, F., Zeng, L., Erdmann, J., Schunkert, H., Khaw, K.-T., Griffin, J.L., Forouhi, N.G., Scott, R.A., Wood, A.M., Burgess, S., Howson, J.M.M., Danesh, J., Wareham, N.J., Butterworth, A.S., and Langenberg, C. (2019). Assessing the causal association of glycine with risk of cardio-metabolic diseases. Nature Communications *10*, 1060.

Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Systems *8*, 281-291.e289.

Wood, A.R., Esko, T., Yang, J., Vedantam, S., Pers, T.H., Gustafsson, S., Chu, A.Y., Estrada, K., Luan, J.a., Kutalik, Z., Amin, N., Buchkovich, M.L., Croteau-Chonka, D.C., Day, F.R., Duan, Y., Fall, T., Fehrmann, R., Ferreira, T., Jackson, A.U., Karjalainen, J., Lo, K.S., Locke, A.E., Mägi, R., Mihailov, E., Porcu, E., Randall, J.C., Scherag, A., Vinkhuyzen, A.A.E., Westra, H.-J., Winkler, T.W., Workalemahu, T., Zhao, J.H., Absher, D., Albrecht, E., Anderson, D., Baron, J., Beekman, M., Demirkan, A., Ehret, G.B., Feenstra, B., Feitosa, M.F., Fischer, K., Fraser, R.M., Goel, A., Gong, J., Justice, A.E., Kanoni, S., Kleber, M.E., Kristiansson, K., Lim, U., Lotay, V., Lui, J.C., Mangino, M., Leach, I.M., Medina-Gomez, C., Nalls, M.A., Nyholt, D.R., Palmer, C.D., Pasko, D., Pechlivanis, S., Prokopenko, I., Ried, J.S., Ripke, S., Shungin, D., Stancáková, A., Strawbridge, R.J., Sung, Y.J., Tanaka, T., Teumer, A., Trompet, S., van der Laan, S.W., van Setten, J., Van Vliet-Ostaptchouk, J.V., Wang, Z., Yengo, L., Zhang, W., Afzal, U., Ärnlöv, J., Arscott, G.M., Bandinelli, S., Barrett, A., Bellis, C., Bennett, A.J., Berne, C., Blüher, M., Bolton, J.L., Böttcher, Y., Boyd, H.A., Bruinenberg, M., Buckley, B.M., Buyske, S., Caspersen, I.H., Chines, P.S., Clarke, R., Claudi-Boehm, S., Cooper, M., Daw, E.W., De Jong, P.A., Deelen, J., Delgado, G., Denny, J.C., Dhonukshe-Rutten, R., Dimitriou, M., Doney, A.S.F., Dörr, M., Eklund, N., Eury, E., Folkersen, L., Garcia, M.E., Geller, F., Giedraitis, V., Go, A.S., Grallert, H., Grammer, T.B., Gräßler, J., Grönberg, H., de Groot, L.C.P.G.M., Groves, C.J., Haessler, J., Hall, P., Haller, T., Hallmans, G., Hannemann, A., Hartman, C.A., Hassinen, M., Hayward, C., Heard-Costa, N.L., Helmer, Q., Hemani, G., Henders, A.K., Hillege, H.L., Hlatky, M.A., Hoffmann, W., Hoffmann, P., Holmen, O., Houwing-Duistermaat, J.J., Illig, T., Isaacs, A., James, A.L., Jeff, J., Johansen, B., Johansson, Å., Jolley, J., Juliusdottir, T., Junttila, J., Kho, A.N., Kinnunen, L., Klopp, N., Kocher, T., Kratzer, W., Lichtner, P., Lind, L., Lindström, J., Lobbens, S., Lorentzon, M., Lu, Y., Lyssenko, V., Magnusson, P.K.E., Mahajan, A., Maillard, M., McArdle, W.L., McKenzie, C.A., McLachlan, S., McLaren, P.J., Menni, C., Merger, S., Milani, L., Moayyeri, A., Monda, K.L., Morken, M.A., Müller, G., Müller-Nurasyid, M., Musk, A.W., Narisu, N., Nauck, M., Nolte, I.M., Nöthen, M.M., Oozageer, L., Pilz, S., Rayner, N.W., Renstrom, F., Robertson, N.R., Rose, L.M., Roussel, R., Sanna, S., Scharnagl, H., Scholtens, S., Schumacher, F.R., Schunkert, H., Scott, R.A., Sehmi, J., Seufferlein, T., Shi, J., Silventoinen, K., Smit, J.H., Smith, A.V., Smolonska, J., Stanton, A.V., Stirrups, K., Stott, D.J., Stringham, H.M., Sundström, J., Swertz, M.A., Syvänen, A.-C., Tayo, B.O., Thorleifsson, G., Tyrer, J.P., van Dijk, S., van Schoor, N.M., van der Velde, N., van Heemst, D., van Oort, F.V.A., Vermeulen, S.H., Verweij, N., Vonk, J.M., Waite, L.L., Waldenberger, M., Wennauer, R., Wilkens, L.R., Willenborg, C., Wilsgaard, T., Wojczynski, M.K., Wong, A., Wright, A.F., Zhang, Q., Arveiler, D., Bakker, S.J.L., Beilby, J., Bergman, R.N., Bergmann, S., Biffar, R., Blangero, J., Boomsma, D.I., Bornstein, S.R., Bovet, P., Brambilla, P., Brown, M.J., Campbell, H., Caulfield, M.J., Chakravarti, A., Collins, R., Collins, F.S., Crawford, D.C., Cupples, L.A., Danesh, J., de Faire, U., den Ruijter, H.M., Erbel, R., Erdmann, J., Eriksson, J.G., Farrall, M., Ferrannini, E., Ferrières, J., Ford, I., Forouhi, N.G., Forrester, T., Gansevoort, R.T., Gejman, P.V., Gieger, C., Golay, A., Gottesman, O., Gudnason, V., Gyllensten, U., Haas, D.W., Hall, A.S., Harris, T.B., Hattersley, A.T., Heath, A.C., Hengstenberg, C., Hicks, A.A., Hindorff, L.A., Hingorani, A.D., Hofman, A., Hovingh, G.K., Humphries, S.E., Hunt, S.C., Hypponen, E., Jacobs, K.B., Jarvelin, M.-R., Jousilahti, P., Jula, A.M., Kaprio, J., Kastelein, J.J.P., Kayser, M., Kee, F., Keinanen-Kiukaanniemi, S.M., Kiemeney, L.A., Kooner, J.S., Kooperberg, C., Koskinen, S., Kovacs, P., Kraja, A.T., Kumari, M., Kuusisto, J., Lakka, T.A., Langenberg, C., Le Marchand, L., Lehtimäki, T., Lupoli, S., and Madden, P.A.F. (2014). Defining the role of

common variation in the genomic and biological architecture of adult human height. Nature Genetics *46*, 1173-1186.

Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., Bacanu, S.-A., Bækvad-Hansen, M., Beekman, A.F.T., Bigdeli, T.B., Binder, E.B., Blackwood, D.R.H., Bryois, J., Buttenschøn, H.N., Bybjerg-Grauholm, J., Cai, N., Castelao, E., Christensen, J.H., Clarke, T.-K., Coleman, J.I.R., Colodro-Conde, L., Couvy-Duchesne, B., Craddock, N., Crawford, G.E., Crowley, C.A., Dashti, H.S., Davies, G., Deary, I.J., Degenhardt, F., Derks, E.M., Direk, N., Dolan, C.V., Dunn, E.C., Eley, T.C., Eriksson, N., Escott-Price, V., Kiadeh, F.H.F., Finucane, H.K., Forstner, A.J., Frank, J., Gaspar, H.A., Gill, M., Giusti-Rodríguez, P., Goes, F.S., Gordon, S.D., Grove, J., Hall, L.S., Hannon, E., Hansen, C.S., Hansen, T.F., Herms, S., Hickie, I.B., Hoffmann, P., Homuth, G., Horn, C., Hottenga, J.-J., Hougaard, D.M., Hu, M., Hyde, C.L., Ising, M., Jansen, R., Jin, F., Jorgenson, E., Knowles, J.A., Kohane, I.S., Kraft, J., Kretzschmar, W.W., Krogh, J., Kutalik, Z., Lane, J.M., Li, Y., Li, Y., Lind, P.A., Liu, X., Lu, L., MacIntyre, D.J., MacKinnon, D.F., Maier, R.M., Maier, W., Marchini, J., Mbarek, H., McGrath, P., McGuffin, P., Medland, S.E., Mehta, D., Middeldorp, C.M., Mihailov, E., Milaneschi, Y., Milani, L., Mill, J., Mondimore, F.M., Montgomery, G.W., Mostafavi, S., Mullins, N., Nauck, M., Ng, B., Nivard, M.G., Nyholt, D.R., O'Reilly, P.F., Oskarsson, H., Owen, M.J., Painter, J.N., Pedersen, C.B., Pedersen, M.G., Peterson, R.E., Pettersson, E., Peyrot, W.J., Pistis, G., Posthuma, D., Purcell, S.M., Quiroz, J.A., Qvist, P., Rice, J.P., Riley, B.P., Rivera, M., Saeed Mirza, S., Saxena, R., Schoevers, R., Schulte, E.C., Shen, L., Shi, J., Shyn, S.I., Sigurdsson, E., Sinnamon, G.B.C., Smit, J.H., Smith, D.J., Stefansson, H., Steinberg, S., Stockmeier, C.A., Streit, F., Strohmaier, J., Tansey, K.E., Teismann, H., Teumer, A., Thompson, W., Thomson, P.A., Thorgeirsson, T.E., Tian, C., Traylor, M., Treutlein, J., Trubetskoy, V., Uitterlinden, A.G., Umbricht, D., Van der Auwera, S., van Hemert, A.M., Viktorin, A., Visscher, P.M., Wang, Y., Webb, B.T., Weinsheimer, S.M., Wellmann, J., Willemsen, G., Witt, S.H., Wu, Y., Xi, H.S., Yang, J., Zhang, F., Arolt, V., Baune, B.T., Berger, K., Boomsma, D.I., Cichon, S., Dannlowski, U., de Geus, E.C.J., DePaulo, J.R., Domenici, E., Domschke, K., Esko, T., Grabe, H.J., Hamilton, S.P., Hayward, C., Heath, A.C., Hinds, D.A., Kendler, K.S., Kloiber, S., Lewis, G., Li, Q.S., Lucae, S., Madden, P.F.A., Magnusson, P.K., Martin, N.G., McIntosh, A.M., Metspalu, A., Mors, O., Mortensen, P.B., Müller-Myhsok, B., Nordentoft, M., Nöthen, M.M., O'Donovan, M.C., Paciga, S.A., Pedersen, N.L., Penninx, B.W.J.H., Perlis, R.H., Porteous, D.J., Potash, J.B., Preisig, M., Rietschel, M., Schaefer, C., Schulze, T.G., Smoller, J.W., Stefansson, K., Tiemeier, H., Uher, R., Völzke, H., Weissman, M.M., Werge, T., Winslow, A.R., Lewis, C.M., Levinson, D.F., Breen, G., Børglum, A.D., Sullivan, P.F., eQtlgen, andMe, and the Major Depressive Disorder Working Group of the Psychiatric Genomics, C. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nature Genetics *50*, 668-681.

Wuttke, M., Li, Y., Li, M., Sieber, K.B., Feitosa, M.F., Gorski, M., Tin, A., Wang, L., Chu, A.Y., Hoppmann, A., Kirsten, H., Giri, A., Chai, J.F., Sveinbjornsson, G., Tayo, B.O., Nutile, T., Fuchsberger, C., Marten, J., Cocca, M., Ghasemi, S., Xu, Y., Horn, K., Noce, D., van der Most, P.J., Sedaghat, S., Yu, Z., Akiyama, M., Afaq, S., Ahluwalia, T.S., Almgren, P., Amin, N., Ärnlöv, J., Bakker, S.J.L., Bansal, N., Baptista, D., Bergmann, S., Biggs, M.L., Biino, G., Boehnke, M., Boerwinkle, E., Boissel, M., Bottinger, E.P., Boutin, T.S., Brenner, H., Brumat, M., Burkhardt, R., Butterworth, A.S., Campana, E., Campbell, A., Campbell, H., Canouil, M., Carroll, R.J., Catamo, E., Chambers, J.C., Chee, M.L., Chee, M.L., Chen, X., Cheng, C.Y., Cheng, Y., Christensen, K.,

Cifkova, R., Ciullo, M., Concas, M.P., Cook, J.P., Coresh, J., Corre, T., Sala, C.F., Cusi, D., Danesh, J., Daw, E.W., de Borst, M.H., De Grandi, A., de Mutsert, R., de Vries, A.P.J., Degenhardt, F., Delgado, G., Demirkan, A., Di Angelantonio, E., Dittrich, K., Divers, J., Dorajoo, R., Eckardt, K.U., Ehret, G., Elliott, P., Endlich, K., Evans, M.K., Felix, J.F., Foo, V.H.X., Franco, O.H., Franke, A., Freedman, B.I., Freitag-Wolf, S., Friedlander, Y., Froguel, P., Gansevoort, R.T., Gao, H., Gasparini, P., Gaziano, J.M., Giedraitis, V., Gieger, C., Girotto, G., Giulianini, F., Gögele, M., Gordon, S.D., Gudbjartsson, D.F., Gudnason, V., Haller, T., Hamet, P., Harris, T.B., Hartman, C.A., Hayward, C., Hellwege, J.N., Heng, C.K., Hicks, A.A., Hofer, E., Huang, W., Hutri-Kähönen, N., Hwang, S.J., Ikram, M.A., Indridason, O.S., Ingelsson, E., Ising, M., Jaddoe, V.W.V., Jakobsdottir, J., Jonas, J.B., Joshi, P.K., Josyula, N.S., Jung, B., Kähönen, M., Kamatani, Y., Kammerer, C.M., Kanai, M., Kastarinen, M., Kerr, S.M., Khor, C.C., Kiess, W., Kleber, M.E., Koenig, W., Kooner, J.S., Körner, A., Kovacs, P., Kraja, A.T., Krajcoviechova, A., Kramer, H., Krämer, B.K., Kronenberg, F., Kubo, M., Kühnel, B., Kuokkanen, M., Kuusisto, J., La Bianca, M., Laakso, M., Lange, L.A., Langefeld, C.D., Lee, J.J., Lehne, B., Lehtimäki, T., Lieb, W., Lim, S.C., Lind, L., Lindgren, C.M., Liu, J., Liu, J., Loeffler, M., Loos, R.J.F., Lucae, S., Lukas, M.A., Lyytikäinen, L.P., Mägi, R., Magnusson, P.K.E., Mahajan, A., Martin, N.G., Martins, J., März, W., Mascalzoni, D., Matsuda, K., Meisinger, C., Meitinger, T., Melander, O., Metspalu, A., Mikaelsdottir, E.K., Milaneschi, Y., Miliku, K., Mishra, P.P., Mohlke, K.L., Mononen, N., Montgomery, G.W., Mook-Kanamori, D.O., Mychaleckyj, J.C., Nadkarni, G.N., Nalls, M.A., Nauck, M., Nikus, K., Ning, B., Nolte, I.M., Noordam, R., O'Connell, J., O'Donoghue, M.L., Olafsson, I., Oldehinkel, A.J., Orho-Melander, M., Ouwehand, W.H., Padmanabhan, S., Palmer, N.D., Palsson, R., Penninx, B., Perls, T., Perola, M., Pirastu, M., Pirastu, N., Pistis, G., Podgornaia, A.I., Polasek, O., Ponte, B., Porteous, D.J., Poulain, T., Pramstaller, P.P., Preuss, M.H., Prins, B.P., Province, M.A., Rabelink, T.J., Raffield, L.M., Raitakari, O.T., Reilly, D.F., Rettig, R., Rheinberger, M., Rice, K.M., Ridker, P.M., Rivadeneira, F., Rizzi, F., Roberts, D.J., Robino, A., Rossing, P., Rudan, I., Rueedi, R., Ruggiero, D., Ryan, K.A., Saba, Y., Sabanayagam, C., Salomaa, V., Salvi, E., Saum, K.U., Schmidt, H., Schmidt, R., Schöttker, B., Schulz, C.A., Schupf, N., Shaffer, C.M., Shi, Y., Smith, A.V., Smith, B.H., Soranzo, N., Spracklen, C.N., Strauch, K., Stringham, H.M., Stumvoll, M., Svensson, P.O., Szymczak, S., Tai, E.S., Tajuddin, S.M., Tan, N.Y.Q., Taylor, K.D., Teren, A., Tham, Y.C., Thiery, J., Thio, C.H.L., Thomsen, H., Thorleifsson, G., Toniolo, D., Tönjes, A., Tremblay, J., Tzoulaki, I., Uitterlinden, A.G., Vaccargiu, S., van Dam, R.M., van der Harst, P., van Duijn, C.M., Velez Edward, D.R., Verweij, N., Vogelezang, S., Völker, U., Vollenweider, P., Waeber, G., Waldenberger, M., Wallentin, L., Wang, Y.X., Wang, C., Waterworth, D.M., Bin Wei, W., White, H., Whitfield, J.B., Wild, S.H., Wilson, J.F., Wojczynski, M.K., Wong, C., Wong, T.Y., Xu, L., Yang, Q., Yasuda, M., Yerges-Armstrong, L.M., Zhang, W., Zonderman, A.B., Rotter, J.I., Bochud, M., Psaty, B.M., Vitart, V., Wilson, J.G., Dehghan, A., Parsa, A., Chasman, D.I., Ho, K., Morris, A.P., Devuyst, O., Akilesh, S., Pendergrass, S.A., Sim, X., Böger, C.A., Okada, Y., Edwards, T.L., Snieder, H., Stefansson, K., Hung, A.M., Heid, I.M., Scholz, M., Teumer, A., Köttgen, A., and Pattaro, C. (2019). A catalog of genetic loci associated with kidney function from analyses of a million individuals. Nat Genet *51*, 957-972.

Yan, J., Qiu, Y., Santos, A.M.R.d., Yin, Y., Li, Y.E., Vinckier, N., Nariai, N., Benaglio, P., Raman, A., Li, X., Fan, S., Chiou, J., Chen, F., Frazer, K.A., Gaulton, K.J., Sander, M., Taipale, J., and Ren, B. (2021). Systematic Analysis of Transcription Factor Binding to Noncoding Variants in the Human Genome. Nature *in press*.

Yan, M., Wang, H., Sun, J., Liao, W., Li, P., Zhu, Y., Xu, C., Joo, J., Sun, Y., Abbasi, S., Kovalchuk, A., Lv, N., Leonard, W.J., and Morse, H.C., 3rd (2016). Cutting Edge: Expression of IRF8 in Gastric Epithelial Cells Confers Protective Innate Immunity against Helicobacter pylori Infection. J Immunol *196*, 1999-2003.

Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., Goddard, M.E., and Visscher, P.M. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat Genet *42*, 565-569.

Young, M.D., and Behjati, S. (2020). SoupX removes ambient RNA contamination from droplet-based single-cell RNA sequencing data. Gigascience *9*.

Zaidi, S., and Brueckner, M. (2017). Genetics and Genomics of Congenital Heart Disease. Circ Res *120*, 923-940.

Zhang, D.E., Hetherington, C.J., Chen, H.M., and Tenen, D.G. (1994). The macrophage transcription factor PU.1 directs tissue-specific expression of the macrophage colony-stimulating factor receptor. Molecular and cellular biology *14*, 373-381.

Zhang, H., Tian, L., Shen, M., Tu, C., Wu, H., Gu, M., Paik, D.T., and Wu, J.C. (2019a). Generation of Quiescent Cardiac Fibroblasts From Human Induced Pluripotent Stem Cells for In Vitro Modeling of Cardiac Fibrosis. Circulation Research *125*, 552-566.

Zhang, K., Hocker, J.D., Miller, M., Hou, X., Chiou, J., Poirion, O.B., Qiu, Y., Li, Y.E., Gaulton, K.J., Wang, A., Preissl, S., and Ren, B. (2021). A single-cell atlas of chromatin accessibility in the human genome. Cell *184*, 5985-6001.

Zhang, K., Wang, M., Zhao, Y., and Wang, W. (2019b). Taiji: System-level identification of key transcription factors reveals transcriptional waves in mouse embryonic development. Science Advances *5*, eaav3262.

Zhang, Y., Li, T., Preissl, S., Grinstein, J., Farah, E.N., Destici, E., Lee, A.Y., Chee, S., Qiu, Y., Ma, K., Ye, Z., Hu, R., Zhu, Q., Huang, H., Fang, R., Yu, L., Izpisua Belmonte, J.C., Wu, J., Evans, S.M., Chi, N.C., and Ren, B. (2019c). 3D Chromatin Architecture Remodeling during Human Cardiomyocyte Differentiation Reveals A Role Of HERV-H In Demarcating Chromatin Domains. bioRxiv.

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

Zhao, Y.-y., Sawyer, D.R., Baliga, R.R., Opel, D.J., Han, X., Marchionni, M.A., and Kelly, R.A. (1998). Neuregulins Promote Survival and Growth of Cardiac Myocytes: PERSISTENCE OF ErbB2 AND ErbB4 EXPRESSION IN NEONATAL AND ADULT VENTRICULAR MYOCYTES. Journal of Biological Chemistry *273*, 10261-10269.

Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., Gregory, M.T., Shuga, J., Montesclaros, L., Underwood, J.G., Masquelier, D.A., Nishimura, S.Y., Schnall-Levin, M., Wyatt, P.W., Hindson, C.M., Bharadwaj, R., Wong, A., Ness, K.D., Beppu, L.W., Deeg, H.J., McFarland, C., Loeb, K.R., Valente, W.J., Ericson, N.G., Stevens, E.A., Radich, J.P., Mikkelsen, T.S., Hindson, B.J., and Bielas, J.H. (2017). Massively parallel digital transcriptional profiling of single cells. Nat Commun *8*, 14049.

Zhu, C., Preissl, S., and Ren, B. (2020). Single-cell multimodal omics: the power of many. Nature Methods *17*, 11-14.

Zuris, J.A., Thompson, D.B., Shu, Y., Guilinger, J.P., Bessen, J.L., Hu, J.H., Maeder, M.L., Joung, J.K., Chen, Z.Y., and Liu, D.R. (2015). Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. Nat Biotechnol *33*, 73-80.