

Single cell variant to enhancer to gene map for coronary artery disease

Junedh M. Amrute^{1,2#*}, Paul C. Lee^{1*}, Ittai Eres^{2*}, Chang Jie Mick Lee^{3,4}, Andrea Bredemeyer¹, Maya U. Sheth^{5,6,7}, Tracy Yamawaki², Rijan Gurung^{3,4}, Chukwuemeka Anene-Nzelu^{8,9}, Wei-Lin Qiu^{10,11}, Soumya Kundu⁵, Daniel Y. Li¹², Markus Ramste¹², Daniel Lu², Anthony Tan^{5,6}, Chul-Joo Kang¹, Ryan E. Wagoner¹, Arturo Alisio¹, Paul Cheng^{12,13}, Quanyi Zhao¹², Clint L. Miller¹⁴, Ira M. Hall^{15,16}, Rajat M. Gupta^{17,18}, Yi-Hsiang Hsu^{2,17,19}, Saptarsi M. Haldar², Kory J. Lavine^{1,20,21}, Simon Jackson², Robin Andersson^{10,11}, Jesse M. Engreitz^{5,6,11,13,22}, Roger S-Y Foo^{3,4}, Chi-Ming Li², Brandon Ason², Thomas Quertermous^{12,13}, Nathan O. Stitzel^{1,23#}

Affiliations:

¹Center for Cardiovascular Research, Division of Cardiology, Department of Medicine, Washington University School of Medicine, Saint Louis, MO, 63110, USA.

²Amgen Research, South San Francisco, CA, 94080, USA

³Cardiovascular Metabolic Disease Translational Research Programme, National University Health System, Centre for Translational Medicine, 14 Medical Drive, Singapore 117599, Singapore

⁴Institute of Molecular and Cell Biology, 61 Biopolis Drive, Singapore 138673, Singapore

⁵Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA

⁶Basic Sciences and Engineering Initiative, Betty Irene Moore Children's Heart Center, Lucile Packard Children's Hospital, Stanford, CA, USA

⁷Department of Bioengineering, Stanford University, Stanford, CA, USA

⁸Montreal Heart Institute, Montreal, 5000 Rue Belanger, QC, H1T 1C8, Canada

⁹Department of Medicine, Université de Montréal, 2900 Edouard Montpetit Blvd, Montréal, QC, H3T 1J4, Canada

¹⁰Department of Biology, University of Copenhagen, Copenhagen, Denmark

¹¹The Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute, Cambridge, MA, USA

¹²Department of Medicine, Division of Cardiovascular Medicine, Stanford University, Stanford, CA 94305

¹³Stanford Cardiovascular Institute, Stanford University, Stanford, CA 94305

¹⁴Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville

¹⁵Center for Genomic Health, Yale University, New Haven, CT, 06510, USA

¹⁶Department of Genetics, Yale University, New Haven, CT, 06510, USA

¹⁷Broad Institute of MIT and Harvard, Cambridge, MA, USA

¹⁸Divisions of Genetics and Cardiovascular Medicine, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA

¹⁹Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, 02215, USA

²⁰Department of Pathology and Immunology, Washington University School of Medicine, Saint Louis, MO, 63110, USA.

²¹Department of Developmental Biology, Washington University School of Medicine, Saint Louis, MO, 63110, USA.

²²Gene Regulation Observatory, Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA

²³Department of Genetics, Washington University School of Medicine, Saint Louis, MO, 63110, USA.

*These authors contributed equally.

#Co-corresponding authors.

52 Correspondence should be addressed to:

53

54 Junedh M. Amrute

55 Affiliation: Division of Cardiology, Department of Medicine, Washington University School of
56 Medicine; Amgen Global Research, South San Francisco, CA, 94080, USA

57 Address: 660 S Euclid, St. Louis, MO 63110.

58 Telephone: 626-487-7180

59 Email: jamrute@wustl.edu

60

61

62 Nathan O. Stitzel, MD, PhD,

63 Affiliation: Division of Cardiology, Department of Medicine, Washington University School of
64 Medicine.

65 Address: 660 S Euclid, Campus Box 8086, St. Louis, MO 63110.

66 Telephone: 314-362-1291

67 Email: nstitzel@wustl.edu

68

69

70

71

72

73

74

75

76

77

78

79

80

81

82

83

84

85

86

87

88

89

90

91

92

93

94

95

96

97

98

99

100

101

102

103 **Abstract**

104
105 Although genome wide association studies (GWAS) in large populations have identified hundreds
106 of variants associated with common diseases such as coronary artery disease (CAD), most
107 disease-associated variants lie within non-coding regions of the genome, rendering it difficult to
108 determine the downstream causal gene and cell type. Here, we performed paired single nucleus
109 gene expression and chromatin accessibility profiling from 44 human coronary arteries. To link
110 disease variants to molecular traits, we developed a meta-map of 88 samples and discovered
111 11,182 single-cell chromatin accessibility quantitative trait loci (caQTLs). Heritability enrichment
112 analysis and disease variant mapping demonstrated that smooth muscle cells (SMCs) harbor the
113 greatest genetic risk for CAD. To capture the continuum of SMC cell states in disease, we used
114 dynamic single cell caQTL modeling for the first time in tissue to uncover QTLs whose effects are
115 modified by cell state and expand our insight into genetic regulation of heterogenous cell
116 populations. Notably, we identified a variant in the *COL4A1/COL4A2* CAD GWAS locus which
117 becomes a caQTL as SMCs de-differentiate by changing a transcription factor binding site for
118 *EGR1/2*. To unbiasedly prioritize functional candidate genes, we built a genome-wide single cell
119 variant to enhancer to gene (scV2E2G) map for human CAD to link disease variants to causal
120 genes in cell types. Using this approach, we found several hundred genes predicted to be linked
121 to disease variants in different cell types. Next, we performed genome-wide Hi-C in 16 human
122 coronary arteries to build tissue specific maps of chromatin conformation and link disease variants
123 to integrated chromatin hubs and distal target genes. Using this approach, we show that
124 rs4887091 within the *ADAMTS7* CAD GWAS locus modulates function of a super chromatin
125 interactome through a change in a CTCF binding site. Finally, we used CRISPR interference to
126 validate a distal gene, *AMOTL2*, linked to a CAD GWAS locus. Collectively we provide a disease-
127 agnostic framework to translate human genetic findings to identify pathologic cell states and
128 genes driving disease, producing a comprehensive scV2E2G map with genetic and tissue level
129 convergence for future mechanistic and therapeutic studies.

130
131 **Keywords:** human genetics, coronary artery disease, genome wide association studies, single
132 cell RNA sequencing, single cell ATAC sequencing, Hi-C, quantitative trait loci (QTL), CRISPR
133 interference
134

135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153

154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204

Introduction

Common diseases are driven by a multitude of factors and are among the most difficult to treat given the lack of a unified disease mechanism^{1,2}. Coronary artery disease (CAD) is the most prevalent common heart disease and a leading cause of death worldwide³. Numerous genome-wide association studies (GWAS) have uncovered causal disease variants driving CAD risk and paved the way for new therapeutic target discovery⁴⁻⁹. Notably, much of the translation from GWAS to therapeutic targets in the CAD space has been restricted to lipid-lowering drugs^{10,11}. While classically considered a lipid driven disease, large-scale clinical outcome studies have shown a persistent risk of CAD independent of lipid levels and patients continue to suffer from atherosclerotic disease despite optimal lipid-lowering therapy¹²⁻¹⁴. Additionally, most of the GWAS loci in CAD are thought to regulate biological pathways outside lipid biology^{15,16}. Leveraging human genetics to uncover new lipid-independent biological mechanisms and drug targets in CAD presents a tremendous unmet need and would be a major milestone in translating population-scale GWAS findings to therapies for patients¹⁷⁻¹⁹.

While population scale GWAS studies offer the greatest opportunity for target discovery, most of the variants identified lie within non-coding regions of the genome, making it difficult to dissect the causal cells and molecular pathways^{20,21}. The advent of single cell sequencing technologies now allows for the profiling of human tissues in health and disease at unprecedented granularity²²⁻²⁵. Prior studies have utilized traditional single cell RNA-sequencing (scRNA-seq) for detailed characterization of cell states in human tissue that drive disease²⁶⁻³³; however, these studies do not profile the non-coding genome making it challenging to link GWAS variants to observed molecular changes. More recently, Turner *et al*³⁴ utilized single-nucleus assay for transposase-accessible chromatin with sequencing (snATAC-seq) in human coronary arteries to link CAD GWAS variants to regulatory elements. Although this represented a major advance in the field, the lack of complementary transcriptomic profiling rendered it challenging to nominate the causal genes in these specific cell types. Furthermore, isolated snATAC-seq precludes the discovery of distal gene regulation as a driver of disease,³⁵⁻³⁷ and there are currently no studies which profile the 3D chromatin architecture in human coronary arteries to enable the study of these distal regulatory connections. More recently, technological advances have paved the way for simultaneous multi-omic profiling of the transcriptome and epigenome at single-cell resolution,^{32,36,38-45} presenting a unique opportunity to link genetic variants to cell types and genes to nominate causal gene programs driving disease risk.

Herein, we performed Multiome (RNA + ATAC) sequencing in 44 human coronary arteries to build a multi-omic map of CAD. We utilized novel computational methods^{46,47} to link enhancers to putative target genes and built a single cell variant-to-enhancer-to-gene (V2E2G) map of human CAD. Furthermore, we performed single cell chromatin accessibility quantitative trait loci (caQTL) discovery to prioritize variants casually linked to changes in chromatin. Collectively, we found that the smooth muscle cells (SMCs) harbor the greatest genetic risk of CAD. To expand insight into gene regulation in functionally heterogeneous SMC cell states we utilized dynamic caQTL modeling⁴⁸ to uncover cell-state dependent pathogenicity of disease variants. To dissect the role of genetic variants in regulating larger scale gene networks, we performed genome-wide Hi-C⁴⁹ in 16 human coronary arteries to link variants to distal genes and validated our findings using cell-specific Hi-C with chromatin immunoprecipitation (HiChIP-seq)⁵⁰ and *in vitro* clustered regularly interspaced short palindromic repeats interference (CRISPRi). Collectively, we provide the first multi-omic map of human CAD and integrate transcriptomic, epigenetic, and chromatin architecture to link disease-associated genetic variants with causal cell types and genes.

205 Results

206

207 Multi-omic map of human coronary artery disease

208 We performed single nucleus Multiome (RNA + ATAC) sequencing in 44 human coronary arteries
209 from non-failing donors and patients with chronic heart failure isolated at the time of orthotopic
210 heart transplantation (**Fig. 1a, Extended Data Fig. 1a, Supplementary Table 1**). After quality
211 control⁵², doublet removal⁵², data integration⁵³, and clustering, we recovered 126,804 nuclei from
212 11 distinct cell types (**Fig. 1b, Extended Data Fig. 1b-d**) present in all patient samples (**Fig. 1c**).
213 To annotate cell types, we used paired RNA information from the nuclei for differential gene
214 expression analysis and used canonical marker genes to annotate cell types (**Fig. 1d, Extended**
215 **Data Fig. 2a-b, Supplementary Table 2**). We used MACS2 to call peaks from the ATAC-seq
216 data (**Fig. 1e**) and identified 143,743 cell type specific differentially accessible marker peaks (**Fig.**
217 **1f**). We found that there was a concordance between RNA and ATAC clustering; however, having
218 gene expression enabled annotation of rarer cell types missed by ATAC-based clustering
219 (**Extended Data Fig. 2c-e**). Using the RNA-based annotation and differentially-accessible peaks
220 across cells, we found cell-type-specific enrichment of transcriptional motifs (**Extended Data Fig.**
221 **2f**). Having paired RNA and ATAC information from the same nucleus offers a unique opportunity
222 to construct a comprehensive genome-wide enhancer gene map from tissue. We linked peaks to
223 genes using a new supervised single-cell model called scE2G (Wei-Lin Qiu, Maya Sheth, Robin
224 Andersson, and Jesse Engreitz, in preparation), which predicts enhancer-gene regulatory
225 interactions using features based on chromatin accessibility, distance, and peak-gene correlation
226 across single cells (**Fig. 1g, Extended Data Fig. 7a-f, Supplementary Table 13-13, Methods**).
227 We found that the non-immune cells (smooth muscle cells (SMCs), endothelial cells (ECs), and
228 fibroblasts) had the greatest number of shared enhancers relative to myeloid cells (**Fig. 1g**).

229

230 To test for enrichment of disease risk in the four major cell types (SMC, EC, Myeloid, Fibroblast),
231 we used stratified LD score regression⁵⁴ (LDSC) to partition heritability in cell-type marker peaks
232 for key cardiometabolic diseases (coronary artery disease, diabetes, abdominal aortic
233 aneurysms, diastolic/systolic blood pressure, and stroke)^{8,55-58}. Notably, we found that SMCs and
234 ECs harbor the greatest genetic risk across cardiometabolic disease variants (**Fig. 1h**). Next, we
235 performed genome-wide Hi-C in 16 human coronary arteries (8 African American and 8 European
236 individuals) (**Fig. 1a,i,j, Extended Data Fig. 3a-d**). We identified more loops at 10 kb resolution
237 than at 2 or 5 kb resolution that were also on average smaller in size (**Extended Data Fig. 3,f**).
238 To further enrich for high-confidence loops, we only retained loops identified in coronaries from
239 at least two individuals. To uncover higher-level chromatin structures that are not as dynamic as
240 loops, we also identified topologically associated domains (TADs) (**Extended Data Fig. 3**).

241

242 Single-nuclei caQTL discovery

243 Chromatin accessibility quantitative trait loci (caQTL) mapping can identify regulatory variants that
244 regulate accessibility of a regulatory element in the genome through mechanisms such as altered
245 transcription factor binding^{51,52}. Single cell caQTLs enable further identification of variants that
246 may regulate chromatin accessibility in a cell-type specific manner^{34,53}. To boost our power for
247 single-cell caQTL discovery, we built a chromatin accessibility meta-map by integrating our
248 Multiome ATAC-seq data with prior snATAC-seq from human coronary arteries³⁴ (**Extended Data**
249 **Fig. 4a**). Our integrated CAD snATAC-seq meta-map of genotyped samples included 88 samples
250 and 245,562 nuclei across 4 ancestries (**Extended Data Fig. 4a-c**). We then used this meta-map
251 and RASQUAL⁵⁹ to identify pseudobulk caQTLs in four major cell types (SMC, EC, Myeloid,
252 Fibroblast) (**Fig. 2a, Supplementary Table 14-21**). Notably, RASQUAL simultaneously models
253 allelic imbalance and QTL effect to improve fine-mapping of putative causal variants. A +/-10KB
254 *cis*-window around the peaks was tested for association between variant and chromatin
255 accessibility and age, sex, sequencing site, and the first four principal components of genotype

256 data were included as covariates (see Methods). At the 10% FDR threshold, we found 11,182
257 caQTLs in SMCs, ECs, Myeloid cells, and Fibroblasts. We found that the power for QTL discovery
258 was associated with number of nuclei (**Fig. 2b, Extended Data Fig. 4d**). As expected, we found
259 that caQTLs were strongly enriched near the peaks they were predicted to regulate (**Extended**
260 **Data Fig. 4e**). We used the scE2G map to link the caQTL peak to its predicted target gene. We
261 next overlapped our pseudobulk caQTLs with GTEx bulk arterial tissue eQTLs by cell type and
262 found several hundred shared QTL variants in SMCs, ECs, Myeloid cells, and Fibroblasts,
263 respectively (**Fig. 2c, Supplementary Tables 15, 17, 19, 21**). Furthermore, we found a strong
264 correlation in effect size between our SMC caQTLs and bulk coronary artery eQTLs from GTEx
265 (**Fig. 2d**).

266
267 To identify disease-associated caQTLs, we overlapped single-cell caQTLs for each cell type with
268 CAD GWAS variants (defined as variants associated with CAD at 1% FDR⁸ along with variants in
269 linkage disequilibrium with these at $R^2 \geq 0.8$ in the European population from 1000 Genomes
270 Project) which yielded numerous disease relevant cell-type-specific caQTLs (**Extended Data Fig.**
271 **5a-d**). For example, we found that rs7182567 was an SMC-specific caQTL that was also a GWAS
272 variant ($p = 5.4 \times 10^{-26}$) and a bulk GTEx eQTL for *ADAMTS7* in tibial artery (**Fig. 2e, Extended**
273 **Data Fig. 5a**). To map this molecular trait to cell state, we leveraged our paired RNA information
274 and sub-clustered the SMCs into transcriptionally distinct cell states, forming a continuum of
275 contractile SMCs (SMC1-3) to de-differentiated SMCs (fibromyocyte, FMC; chondromyocyte,
276 CMC) (**Fig. 2f**). Notably, we found that *ADAMTS7* was most highly expressed in FMCs (**Fig. 2f**)
277 suggesting a role in disease regulation.

278
279 Similarly, we found that rs8017642 is an EC-specific caQTL for *NEK9* (**Fig. 2g, Extended Data**
280 **Fig. 5b**). To identify functionally-distinct endothelium cell states, we mapped our data to a high-
281 resolution single-cell RNA-seq atherosclerosis atlas⁶⁰ and identified four transcriptionally-distinct
282 endothelium subsets (**Fig. 2h**). Notably, *NEK9* was expressed in endothelial cells enriched with
283 endothelium-to-mesenchymal transition (EndoMT) genes, such as *COL1A1* and *FN1*⁶¹ (**Fig. 2h,**
284 **Extended Data Fig. 5e**). Pathway enrichment for marker genes for the EndoMT population
285 showed increased EMT and stress response signals (**Extended Data Fig. 5f**). Next, we used the
286 41 EC-specific putative causal genes previously identified from in a V2G2P analysis of CAD
287 GWAS loci⁴⁷ to create a gene set score which showed maximal enrichment in intimal ECs
288 (**Extended Data Fig. 5g**). Additionally, as prior work has uncovered a gene program⁴⁷ driving
289 CAD pathogenesis in ECs (Program 8), we created a gene signature for all Program 8 genes.
290 Consistent with prior work,⁴⁷ we found an enrichment of this program in angiogenic/vaso vasorum
291 ECs (**Extended Data Fig. 5j**).

292
293 To illustrate how a cell type specific caQTL can impart functional transcriptional changes, we
294 examined the CAD GWAS variant rs658956. The variant is an SMC-specific caQTL and regulates
295 a putative enhancer element (chr1:59169569-59170069) that is linked to the *HSD52* gene (based
296 on scE2G) (**Fig. 2j,k**). Additionally, transcription factor binding site (TFBS) analysis found that
297 rs658956, which changes the reference (G) to alternative (T) allele, is predicted to disrupt a
298 BACH1/2 TF binding motif (**Fig. 2l**). Furthermore, we found that rs658956 is also a tibial artery
299 eQTL for *HSD52* expression ($p = 1.1 \times 10^{-41}$) wherein the switch from reference (G) to alternate (T)
300 is associated with decreased expression (**Fig. 2m**). Collectively, this would suggest that the
301 GWAS variant rs658956 may impact TF binding thereby decreasing chromatin accessibility of the
302 enhancer and subsequently reducing *HSD52* gene expression.

303

304 **Identifying caQTLs that are dynamic with SMC de-differentiation state**

305 Molecular QTL discovery using bulk or pseudobulk data is generally carried out in a cell state
306 agnostic-manner from tissues which collapses cells across a heterogeneous mix of cell states.
307 For example, in the development of atherosclerosis it is known that SMCs undergo a phenotypic
308 switch from a contractile to a fibroblast-like FMC phenotype³⁰ which is accompanied by
309 corresponding transcriptional changes. Paired multiomic RNA-ATAC sequencing offers the
310 opportunity to characterize individual cells within a global cell type across continuous
311 transcriptional cell states and to model this continuum of heterogeneous cell states to discover
312 peaks that may modify or be modified by QTL variant effects⁵⁴.

313
314 Using the top 100 marker genes defined in FMCs from Wirka *et al*³⁰, we created an FMC module
315 score across our integrated SMC map and identified a continuum of activation states (see
316 Methods, **Fig. 3a, Extended Data Fig. 6a-b**). We then binned nuclei into bottom, middle, and top
317 thirds based on FMC score for downstream modeling (**Fig. 3b**). To dissect caQTL dynamics in
318 the context of cell state, we utilized a Poisson mixed effect (PME) single cell model^{48,55} (**Fig. 3c,**
319 **Extended Data Fig. 6c**). This approach models cell-level fragment counts with random effect
320 covariates to account for relatedness shared by cells from the same sample and sequencing
321 conditions. Briefly, at baseline we model ATAC fragment counts with fixed effect covariates
322 including genotype, number of fragments per cell, TSS enrichment per cell, principal components
323 of single-cell ATAC counts, principal components of genotype data, random effect covariates
324 (batch and sequencing site), and intercept. To test for dynamic caQTLs where the genotype effect
325 is modified by SMC cell state, we then added a cell-level FMC score term and a genotype-FMC
326 score interaction term as covariates in the model (**Fig. 3c**). Finally, we binned the nuclei by
327 bottom, middle, and top FMC score buckets (**Fig. 3b,c**) and identified dynamic caQTLs which
328 differ according to cell-state.

329
330 We ran the PME single cell model on the lead variants of SMC-specific caQTLs identified through
331 the RASQUAL pseudobulk analysis and found 794 caQTLs which are dynamic with respect to
332 FMC score (**Fig. 3d, Supplementary Table 22**). The significant interaction term indicates a
333 modification of genotype effect on peak accessibility during cell-state transition or a genotype-
334 specific modification of cell-state dependent changes on peak accessibility, both of which are
335 plausible depending on the mechanism of the variant (e.g. variant modifies pioneer vs. non-
336 pioneer transcription factor binding). Notably, of the 794 ATAC peaks regulated by dynamic
337 caQTLs, 12 also contained CAD GWAS variants (**Fig. 3e**). We then used the scE2G map to link
338 the 12 dynamic disease-associated caQTLs to target genes. To connect the dynamic disease-
339 associated genes to gene programs, we used a large language model (ChatGPT) to group genes
340 by biological groups and found an enrichment of VSMC function/migration, transcriptional
341 regulation, and ECM remodeling cellular processes and signaling pathways (**Fig. 3e**).
342 Interestingly, we found that a gene set score of these genes (**Fig. 3e**) was enriched in FMCs
343 (**Extended Data Fig. 6d**) bolstering a role in disease regulation through cell state phenotypic
344 switching. For example, the CAD GWAS SNP rs11838776 is a dynamic caQTL variant for an
345 enhancer (chr13:110388102-110388602) that lies within *COL4A2/COL4A1* locus. The strongly
346 negative interaction term of the PME model suggests that the alternate allele of rs11838776 may
347 further decrease accessibility of this enhancer as the SMCs transition into a more FMC-like cell
348 state. We further confirmed this effect by visualizing minimal genotype-specific effect in ATAC
349 accessibility in “quiescent” SMCs (bottom third FMC score) versus larger negative genotype-
350 specific effect in “de-differentiated” FMCs (top third FMC score) (**Fig. 3f,g**). Interestingly, the bulk
351 GTEx data suggests that rs11838776 is not an eQTL for *COL4A1* in tibial arteries ($p = 0.69$) but
352 is an eQTL for *COL4A1* in cultured fibroblasts ($p = 2.2 \times 10^{-6}$) with the same directionality as the
353 interaction effect (**Fig. 3h**). Collectively, this would suggest that rs11838776 and its enhancer may
354 regulate *COL4A1* expression only as cells enter a fibrogenic fate. To dissect the molecular

355 mechanism driving this dynamic QTL effect, we explored TF binding motifs surrounding the
356 rs11838776 variant. We found that the reference allele (G) contributes to a consensus binding
357 motif for *EGR1/2*, while the alternate allele (A) is predicted to abrogate this motif (**Fig. 3i**).
358 Furthermore, we created a gene set score for *EGR1/2* and *COL4A1/2*, respectively, using RNA
359 data and found greatest expression in the CMC/FMC cell state suggesting that as SMCs de-
360 differentiate *EGR1/2* expression increases, driving *COL4A1/2* expression (**Fig. 3j**). We also found
361 that the CAD GWAS variant rs658956 was dynamic for chromatin accessibility at the SMC specific
362 enhancer and linked to *HSD52* (**Fig. 2j-k, Extended Data Fig. 6e**).

363

364 **Single cell disease variant-to-enhancer-to-gene map**

365 To map CAD disease variants to cell types and genes in an unbiased fashion, we built a disease
366 relevant variant-to-enhancer-to-gene (V2E2G) map as follows^{46,47,63} (**Fig. 4a,b**):

367

368 1) Select ATAC-seq peaks from the 4 major cell types which contain a CAD GWAS variant
369 (1,533)

370 2) Variant-to-enhancer (V2E): identify cell-type-specific peaks (as called by MACS2, see
371 Methods) overlapping disease variants, and H3K27ac peaks from bulk coronary
372 arteries^{56,57}. Using this approach, we found cell specific V2E pairs: SMC (490),
373 endothelium (424), fibroblast (431), and myeloid (367).

374 3) Variant-to-enhancer-to-gene (V2E2G): Link peaks from (2) to genes using a new
375 supervised single-cell model called scE2G (Wei-Lin Qiu, Maya Sheth, Robin Andersson,
376 and Jesse Engreitz, in preparation), which predicts enhancer-gene regulatory interactions
377 using features based on chromatin accessibility, distance, and peak-gene correlation
378 across single cells (**Fig. 4a, Extended Data Fig. 7a-f**, see Methods). Overlapping scE2G
379 predictions with our cell-type-specific peaks, we found the following (single-cell V2E2G
380 peaks, no. linked genes): SMC (276, 325), endothelium (239, 268), fibroblast (235, 279),
381 and myeloid (199, 236).

382 4) Finally, to nominate V2E2G links which also affect a molecular trait, we overlapped V2E2G
383 peaks with our single cell caQTLs and found the following (single-cell V2E2G peaks with
384 caQTLs, no. linked genes): SMC (49, 41), endothelium (5, 3), fibroblast (13, 11), and
385 myeloid (15, 12).

386

387 Notably, we found that SMCs were most enriched with V2E2G links and likely harbor the greatest
388 genetic risk of driving CAD. Overlapping V2E2G with single-cell caQTLs provides a high-fidelity
389 set of cell-type-specific hits which may be driving disease in SMCs through a direct change in a
390 molecular trait – however, this approach is challenging in more rare cell types where fewer nuclei
391 hinder QTL discovery (**Fig. 2b, 4b**).

392

393 **SMC V2E2Gs are disease relevant**

394 We utilized ChatGPT to group SMC V2E2G which are caQTLs (41 genes) into functional
395 biological programs and found enrichment across SMC processes including ECM remodeling,
396 cellular growth and proliferation, metabolism, cytoskeletal organization, and gene regulation (**Fig.**
397 **4c**). To identify SMC V2E2G targets which correlate with disease burden, we first used the top
398 100 FMC marker genes from Wirka *et al*³⁰ to generate a FMC marker gene score (**Fig. 4d**),
399 correlated expression of genes identified from V2E2G in SMCs with FMC score, and plotted the
400 correlation colored by adjusted p-value (**Fig. 4e**). We found that 278/325 genes from our V2E2G
401 showed a positive correlation with FMC score while no genes had a negative correlation,
402 demonstrating an enrichment for disease relevant linked genes (**Fig. 4e**). To identify those with
403 highest confidence, we restricted these 278/325 genes to those that were nominated from our
404 bulk arterial GTEx and SMC caQTL overlap analysis which resulted in a set of 5 genes (**Fig. 4e,f**).
405 A heatmap of these 5 genes (*MYO9B*, *COL4A1*, *COL4A2*, *FNDC3B*, and *LRRFIP1*) shows

406 increasing expression as SMCs transition from contractile (SMC1-3) towards de-differentiated
407 disease associated states (CMC, FMC) (**Fig. 4f**). Furthermore, we created a gene set score for
408 the 325 SMC V2E2Gs and found greatest enrichment in SMC3 and FMCs (**Fig. 4g**) – collectively,
409 these findings support the notion that FMCs harbor the greatest genetic risk of CAD. Among
410 these, we highlight rs7246865 as an example as we found this variant to be an SMC dynamic
411 caQTL, a pseudobulk caQTL, and a V2E2G link, thus underscoring its potential importance in
412 disease. We show that rs7246865 falls inside an SMC peak (snATAC-seq) and coronary artery
413 enhancer (ENCODE human coronary artery H3K27ac), and there is a E2G link to the *MYO9B*
414 and *HAUS8* transcription start site (**Fig. 4h**). Importantly, a prior study has used CRISPRi to
415 validate that disruption of rs7246865 is causally linked to *MYO9B/HAUS8* expression⁸.

416

417 **Disease variant chromatin networks**

418 To dissect the broader impact of disease variants on gene networks through chromatin looping,
419 we used genome-wide human coronary HiC to construct large scale chromatin networks (**Fig.**
420 **5a**). Briefly, we built interconnected chromatin networks with interacting anchors using a fast
421 greedy modularity optimization algorithm^{58,59}(**Fig. 5a**, see Methods). We found 5,331
422 interconnected networks (**Fig. 5b**). To identify disease associated networks and build a variant-
423 to-enhancer-to-network (V2E2N) map, we identified networks where any loop within the network
424 contains a V2E in either anchor (i.e. a snATAC peak that contains a GWAS variant and overlaps
425 with and ENCODE coronary artery H3K27ac) for SMCs, fibroblasts, ECs, and myeloid cells (**Fig.**
426 **5b**). Interestingly, we found that all networks contain an average of 18 HiC loops while disease
427 associated networks (V2E2N) contain an average of 59 HiC loops (in SMCs) (**Fig. 5c**). To identify
428 genes implicated in our V2E2N, we found all snATAC-seq peaks present in any HiC anchors in
429 the V2E2N (**Fig. 5a**) and identified linked genes using the scE2G map as before (**Fig. 4a,b**). Using
430 this interested approach, we mapped a variant to a catalogue of genes that may be regulated
431 locally or distally through a linked chromatin network (**Fig. 5d**). This strategy identified key disease
432 networks which form ‘super chromatin interactome’ hubs wherein a disease variant is part of
433 several hundred interconnected chromatin loops (**Fig. 5d,e**). Interestingly, we found that the CAD
434 GWAS variant rs4887091 (located within enhancer element chr15:78750919-78751419) was part
435 of a super chromatin interactome that consisted of numerous genes such as *CHRNA3*, *CHRNA4*,
436 *MORF4L1*, *CHRNA5*, *CTHS*, and *ADAMTS7* which were also V2E2G for rs4887091 – notably,
437 *CHRNA3*, *CHRNA4*, and *ADAMTS7* were also arterial eQTLs for rs4887091 (**Fig. 5e,f**). At the
438 disease variant rs4887091, a base pair change from the reference (T) to alternate (C) allele leads
439 to a change in a CTCF binding motif (**Fig. 5g**). We performed ChIP-seq for CTCF in cultured
440 human coronary artery smooth muscle cells (HCASMCs) and demonstrated that rs4887091
441 overlaps a CTCF binding site, suggesting that this variant may drive broader changes in chromatin
442 organization through regulation of CTCF binding (**Fig. 5h**).

443

444 **Cell type specific mapping of disease variants to distal genes**

445 Although our scE2G predictions are limited to enhancer-gene interactions within a *cis*-window,
446 leveraging our Hi-C based V2E2N analyses we can assign potential distal gene regulatory
447 function to disease variants. Using our V2E2N, we can map disease variants to distal genes. As
448 an example, we find that the CAD GWAS variant rs9876658 falls within a snATAC-seq peak in
449 SMCs, ECs, and fibroblasts (but not myeloid cells) highlighting stroma specific regulation (**Fig.**
450 **6a**). From the genome-wide coronary artery HiC we find a loop connecting the rs9876658
451 enhancer to the *AMOTL2* transcription start site – notably, we do not find a chromatin loop
452 between rs9876658 and the nearest gene *ANAPC13* (**Fig.6a**). Collectively, this analysis predicts
453 that rs9876658 regulates *AMOTL2* expression in stromal cells, however since our HiC is from
454 bulk coronary arteries we cannot implicate a specific cell type, and furthermore, snATAC-seq
455 alone does not identify functional enhancers.

456

457 Furthermore, published bulk coronary artery H3K27ac does not provide cell specificity for active
458 enhancers, so to map disease variants to cell type specific enhancers, we performed H3K27ac
459 CHIP-seq in human coronary artery smooth muscle cells (HCASMC) and human coronary
460 endothelial cells (HCAEC) and found that rs9876658 overlapped with a H3K27ac peak in SMCs
461 but not ECs (**Fig. 6b**), suggesting the ATAC-seq peak containing rs9876658 is a functional
462 enhancer in SMCs but not endothelial cells. Next, to address limitations of bulk tissue chromatin
463 capture, we performed HiChIP-seq in HCASMCs and HCAECs and found a direct loop between
464 the enhancer containing rs9876658 and the *AMOTL2* promoter in SMCs but not ECs (**Fig. 6b**).
465

466 **CRISPRi of rs9876658 affects *AMOTL2* expression in SMCs**

467 To causally link rs9876658 to cell specific regulation of *AMOTL2*, we performed CRISPRi
468 targeting the TSS (*AMOTL2*) and regulatory enhancer (containing rs9876658) with a Lenti-dCas9-
469 KRAB-blast vector (see Methods). We found that targeting the *AMOTL2* TSS led to a complete
470 knockdown of *AMOTL2* expression in HCASMCs and HCAECs validating our system (**Fig. 6c**).
471 Interestingly, targeting the *AMOTL2* TSS also led to a modest reduction in *ANAPC13* expression
472 in HCASMCs, consistent with our network analysis which identified *AMOTL2* and *ANAPC13* being
473 part of the same network (**Fig. 6c**). Importantly, targeting of the enhancer containing rs9876658
474 led to a reduction in *AMOTL2* expression in HCASMCs but not HCAECs bolstering our prediction
475 from human tissue mapping (**Fig. 6c**). Collectively, these analyses nominate rs9876658 as a
476 causal regulator of *AMOTL2* expression in SMCs but not ECs and implicate this gene as a novel
477 candidate for atherosclerosis.

478 **Discussion**

481 GWAS studies have identified hundreds of variants driving common diseases such as CAD^{1,7-9,60}.
482 A key challenge in going from human genetics to druggable therapies is our lack of understanding
483 of the causal cell types and gene programs driving disease¹⁵. The emergence of multi-omic single
484 cell technologies has given us a unique opportunity to profile human tissue at high resolution^{25,61}.
485 Numerous studies have leveraged these technologies to uncover cellular heterogeneity in human
486 tissue across many diseases contributing to a human cell atlas^{26-29,32,33,62,63}. Yet, there are few
487 studies which have mapped common disease variants to causal cell types and genes for target
488 discovery from human tissue. Further, a key limitation of prior studies is the use of uni-modal
489 sequencing technologies or reliance on cell-based systems which precludes discovery of disease
490 targets from the native tissue environment.

491 Our study utilized a new technique, single nucleus Multiome (RNA + ATAC)^{32,38,39}, in human
492 coronary arteries from 44 patients with paired genotyping for un-biased characterization of the
493 transcriptome and epigenome at single cell resolution. Capturing paired gene expression and
494 chromatin accessibility from the same nucleus allows for ground truth peak-gene linkage^{40,42,43,64,65}
495 and is an important milestone for gene regulation discovery in human tissue. We also utilized a
496 novel computational framework, scE2G^{46,47,66}, to build an un-biased genome-wide single cell
497 V2E2G map in human CAD. Notably, we found that SMCs were most enriched with V2E2G links
498 (325 linked genes) and likely harbor the greatest genetic risk of driving CAD. This systematic
499 approach can be applied across diseases and allows for an unbiased framework to prioritize
500 pathological genes and cell types driving human disease.

503 To validate our scV2E2G map using an orthogonal strategy in a broader patient cohort, we
504 integrated our data with published human coronary artery snATAC-seq³⁴ to build a meta-map of
505 88 samples for single cell caQTL discovery. Using this strategy, we discovered 11,182 single cell
506 caQTLs many of which overlapped with bulk arterial tissue GTEx eQTLs. Notably, this approach
507 allows for discovery of novel QTLs in rarer cell types that will be missed by traditional bulk

508 approaches^{54,67}. To enrich for disease associated QTLs, we overlapped our scV2E2G map with
509 caQTLs and to nominate a high confidence gene set with GWAS, scE2G, and molecular trait
510 evidence in regulating disease – we found 5 such genes (*MYO9B*, *COL4A1*, *COL4A2*, *FNDC3B*,
511 and *LRRFIP1*) whose expression was also correlated with the SMC fibromyocyte transition state.
512

513 Traditional pseudobulk QTL discovery is a cell state agnostic method^{34,68} and lacks the ability to
514 discriminate between molecular traits which are influenced by disease state⁴⁸. Here, we utilize
515 single cell genomics in a large population to incorporate cell state dynamics in one of the most
516 fluid cell states in the atherosclerotic plaque – SMCs. In atherosclerosis, contractile SMCs
517 undergo de-differentiation and acquire a fibromyocyte phenotype (FMC) with distinct
518 transcriptional signatures^{30,69,70}. We performed single cell caQTL discovery using the FMC de-
519 differentiation score as a continuous variable to identify QTLs which are cell state dependent.
520 Using this framework, we uncovered rs11838776 as a QTL for *COL4A1* which dynamically
521 becomes active as SMCs de-differentiate via a TF binding site change for *EGR1/2*. Interestingly,
522 *EGR1/2* are universal stripe factors which are known to orchestrate changes in chromatin
523 accessibility for recruitment of co-binding factors^{71,72}. This is the first application of single cell state
524 dependent QTL modeling in human tissue and can expand our understanding of gene regulation
525 in heterogenous cells from tissue. In cases where there are epigenetically or transcriptional
526 defined cell states and cell state trajectories that may be disease relevant, such approaches
527 present an efficient way to uncover molecular QTLs that may modify or be modified by disease-
528 associated differentiation process that may be missed while profiling a bulk tissue.
529

530 Given the genome is a dynamic 3D structure⁷³, chromatin looping through TADs is a crucial high-
531 level state governing downstream gene regulation^{73–76}. Currently, there are no studies which
532 profile chromatin looping in human CAD. Prior studies in cell-based systems and other tissues
533 have demonstrated the power of HiC technologies to uncover chromatin architecture in the
534 context of gene regulation and mapped disease variants to distal causal genes^{77–86}. Here, we
535 performed genome-wide HiC in 16 human coronary arteries and used network graphs to connect
536 chromatin communities. This approach allows for building tissue-specific gene regulatory
537 networks within TADs. Numerous studies have shown that large scale chromatin structures have
538 tissue and cell specificity^{78,80,87}, yet no study has profiled looping in diseased human arterial tissue.
539 We integrated our chromatin networks with the scV2E2G map to identify disease variants
540 enriched in a super chromatin interactome that may regulate a broader gene program. To our
541 surprise, we found that the disease variant rs4887091 drives a super chromatin interactome
542 through a modified CTCF binding site. Given the role of CTCF in regulating TAD insulation
543 boundaries, these results would suggest that this disease variant is orchestrating broad scale
544 chromatin changes. Furthermore, by integrating HiC tissue looping with multi-omic scV2E2G
545 mapping, we uncovered several novel candidate genes linked to CAD GWAS loci that were not
546 previously implicated. We found rs9876658 was linked to the distal candidate gene *AMOTL2* and
547 used cell specific H3K27ac and HiChIP-seq and CRISPRi to target the active enhancer containing
548 rs9876658 to show SMC specific regulation.
549

550 Our study is not without limitations. First, we rely on nuclear RNA from multi-omic capture which
551 precludes discovery of cellular transcripts as prior studies have shown that single cell RNA is
552 deeper for cell state annotation. We utilized high resolution published single cell data to perform
553 reference mapping for state imputation⁸⁸. Second, QTL associations can have pleiotropic effects
554 making it challenging to parse out relevant targets from our large caQTL discovery analysis – we
555 leveraged cell specific QTL discovery and single cell caQTL modeling to enrich for and prioritize
556 disease associated cell types and cell states. Third, our coronary artery HiC data is bulk
557 precluding inference of cell type specific looping – to prioritize cell types and increase confidence
558 for cell specific looping, we used HiChIP-seq in targeted cell types in vitro.

559
560 In conclusion, we built the first multi-omic gene expression and chromatin accessibility map of
561 human CAD and provide a comprehensive framework to map CAD GWAS variants to cell types
562 and genes. Additionally, we use single cell QTL modeling to characterize state dependent
563 pathogenicity of disease variants in human tissue. We used tissue HiC to build large scale
564 chromatin networks and uncover how disease variants impact distal gene regulation. Finally, we
565 integrated tissue omics with cell based epigenetic profiling to prioritize and functionally test
566 candidate genes using enhancer TAP-seq and CRISPRi. Collectively we provide a disease
567 agnostic framework to translate human genetic findings to identify pathologic cell states and
568 genes driving disease – this study provides a comprehensive V2E2G map with genetic and tissue
569 level convergence for future mechanistic and therapeutic studies.

570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609

610 **Acknowledgments**

611
612 JMA was supported by the American Heart Association Predoctoral Fellowship (826325) and is
613 currently supported by Leducq Foundation Network Seed Grant (#20CVD02). JMA and PL are
614 supported by the Washington University School of Medicine Medical Scientist Training Program.
615 This work was supported by National Institutes of Health (NIH) grants UM1HG008853 (IMH and
616 NOS), R01HG013371 (IMH and NOS), R01HL159171 (NOS), R01HL171045 (TQ),
617 R01HL134817 (TQ), R01HL139478 (TQ), R01HL156846 (TQ), R01HL151535 (TQ),
618 R01HL158525 (TQ), UM1HG011972 (JME and TQ), U01HG011762 (TQ), R01HL159176 (JME),
619 R01HL164811 (RG and JME), American Heart Association 23SCISA1144703 (PC),
620 24SCEFIA1248386 (PC), 695 20CDA35310303 (PC), and the Novo Nordisk Foundation Center
621 for Genomic Mechanisms of Disease (NNF21SA0072102). MUS acknowledges the support of an
622 NSF Graduate Research Fellowship (DGE-1656518) and a graduate fellowship award from
623 Knight-Hennessy Scholars at Stanford University. NOS was also supported in part by the
624 Foundation for Barnes-Jewish Hospital. We thank Dr Matthew-Ackers Johnson from NUS
625 Cardiovascular Research Institute for the provision of the primary HCAEC and HCASMC cell
626 lines. RSYF is funded by Individual Research Grants from the National Medical Research Council
627 (NMRC) of Singapore (MOH-001480-00) and MOE Academic Research Fund (AcRF) Tier 3
628 (MOE-000333-00). The study was partially supported from an Amgen sponsored research
629 agreement. Study design schematics were created in BioRender.com. We thank the Genome
630 Technology Access Center at the McDonnell Genome Institute at Washington University School
631 of Medicine for help with genomic analysis. The Center is partially supported by NCI Cancer
632 Center Support Grant #P30 CA91842 to the Siteman Cancer Center. This publication is solely
633 the responsibility of the authors and does not necessarily represent the official view of the NIH.

634 **Author Contributions**

635
636
637 NOS and IMH conceived the study. JMA, TQ, and NOS drafted the manuscript with assistance
638 from all authors. JMA collected all coronary arteries. JMA and AB isolated nuclei from human
639 coronary arteries. JMA performed 10x Multiome cDNA construction for library preparations and
640 sequencing. JMA, PCL, IE, CJK, MS, WLQ, SK, DYL, DL, AT, PC, and QZ performed all
641 computational analysis. CLM assisted with QTL discovery. TY, KL, SJ, BA, CML, and YHH
642 assisted with the HiC study. CJML, RG, and RSYF performed all CRISPRi experiments. CAN,
643 QZ, MR, and CJML performed all in vitro HiChIP and H3K27ac experiments. RG and JME
644 assisted with EC specific analyses. REW and AA assisted with in vitro experiments. MUS, WQ,
645 RA, and JME developed and applied scE2G. TQ, RG, RSYF, SH, and SJ provided guidance on
646 GWAS mapping and data interpretation. All authors contributed to the experimental design, data
647 analysis and interpretation as well as manuscript production. NOS is responsible for all aspects
648 of this manuscript including experimental design, data analysis, and manuscript production. All
649 authors approved the final version of the manuscript.

650 **Competing Interests**

651
652
653 JA, IE, TY, DL, YHH, SH, SJ, CML, and BA were or are employed by Amgen.

654
655
656
657
658
659
660

661 **References**

- 662
- 663 1. Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify
664 individuals with risk equivalent to monogenic mutations. *Nature Genetics* 2018 50:9 50,
665 1219–1224 (2018).
 - 666 2. Uffelmann, E. *et al.* Genome-wide association studies. *Nature Reviews Methods Primers*
667 2021 1:1 1, 1–21 (2021).
 - 668 3. Martin, S. S. *et al.* 2024 Heart Disease and Stroke Statistics: A Report of US and Global
669 Data From the American Heart Association. *Circulation* 149, E347–E913 (2024).
 - 670 4. Kathiresan, S. & Srivastava, D. Leading Edge Review Genetics of Human Cardiovascular
671 Disease. (2012) doi:10.1016/j.cell.2012.03.001.
 - 672 5. Erdmann, J., Kessler, T., Munoz Venegas, L. & Schunkert, H. A decade of genome-wide
673 association studies for coronary artery disease: the challenges ahead. *Cardiovasc Res* 114,
674 1241–1257 (2018).
 - 675 6. Khera, A. V. & Kathiresan, S. Genetics of coronary artery disease: discovery, biology and
676 clinical translation. *Nat Rev Genet* 18, 331–344 (2017).
 - 677 7. Musunuru, K. & Kathiresan, S. Genetics of Common, Complex Coronary Artery Disease.
678 *Cell* 177, 132–145 (2019).
 - 679 8. Aragam, K. G. *et al.* Discovery and systematic characterization of risk variants and genes
680 for coronary artery disease in over a million participants. *Nature Genetics* 2022 54:12 54,
681 1803–1815 (2022).
 - 682 9. Tcheandjieu, C. *et al.* Large-scale genome-wide association study of coronary artery
683 disease in genetically diverse populations. *Nature Medicine* 2022 28:8 28, 1679–1692
684 (2022).
 - 685 10. Shapiro, M. D., Tavori, H. & Fazio, S. PCSK9: From Basic Science Discoveries to
686 Clinical Trials. *Circ Res* 122, 1420 (2018).
 - 687 11. Abifadel, M. *et al.* Mutations in PCSK9 cause autosomal dominant hypercholesterolemia.
688 *Nature Genetics* 2003 34:2 34, 154–156 (2003).
 - 689 12. Ridker, P. M. *et al.* Inflammation, Cholesterol, Lipoprotein(a), and 30-Year
690 Cardiovascular Outcomes in Women. *New England Journal of Medicine* (2024)
691 doi:10.1056/NEJMOA2405182/SUPPL_FILE/NEJMOA2405182_DATA-
692 SHARING.PDF.
 - 693 13. Karalis, D. G. Intensive Lowering of Low-Density Lipoprotein Cholesterol Levels for
694 Primary Prevention of Coronary Artery Disease. *Mayo Clin Proc* 84, 345 (2009).
 - 695 14. Bayturan, O. *et al.* Clinical Predictors of Plaque Progression Despite Very Low Levels of
696 Low-Density Lipoprotein Cholesterol. *J Am Coll Cardiol* 55, 2736–2742 (2010).
 - 697 15. Hartmann, K., Seweryn, M. & Sadee, W. Interpreting coronary artery disease GWAS
698 results: A functional genomics approach assessing biological significance. *PLoS One* 17,
699 (2022).
 - 700 16. Kessler, T. & Schunkert, H. Coronary Artery Disease Genetics Enlightened by Genome-
701 Wide Association Studies. *JACC Basic Transl Sci* 6, 610–623 (2021).
 - 702 17. Kamb, A., Harper, S. & Stefansson, K. Human genetics as a foundation for innovative
703 drug development. *Nat Biotechnol* 31, 975–978 (2013).
 - 704 18. Dowden, H. & Munro, J. Trends in clinical success rates and therapeutic focus. *Nat Rev*
705 *Drug Discov* 18, 495–496 (2019).
 - 706 19. Haldar, S. M. Keeping translational research grounded in human biology. *J Clin Invest*
707 134, (2024).

- 708 20. Wang, X., Savic, D., Hovig, E., Trynka, G. & Cano-Gamez, E. Making sense of GWAS:
709 using epigenomics and genome engineering to understand the functional relevance of
710 SNPs in non-coding regions of the human genome. *Epigenetics & Chromatin* 2015 8:1 8,
711 1–18 (2015).
- 712 21. Cano-Gamez, E. & Trynka, G. From GWAS to Function: Using Functional Genomics to
713 Identify the Mechanisms Underlying Complex Diseases. *Front Genet* 11, 505357 (2020).
- 714 22. Mardis, E. R. The impact of next-generation sequencing technology on genetics. *Trends*
715 *Genet.* 24, 133–141 (2008).
- 716 23. Schuster, S. C. Next-generation sequencing transforms today’s biology. *Nat. Methods* 5,
717 16–18 (2008).
- 718 24. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods*
719 2009 6:5 6, 377–382 (2009).
- 720 25. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The
721 Technology and Biology of Single-Cell RNA Sequencing. *Mol Cell* 58, 610–620 (2015).
- 722 26. Koenig, A. L. *et al.* Single-cell transcriptomics reveals cell-type-specific diversification in
723 human heart failure. *Nature Cardiovascular Research* 2022 1:3 1, 263–280 (2022).
- 724 27. Kuppe, C. *et al.* Spatial multi-omic map of human myocardial infarction. *Nature* (2022)
725 doi:10.1038/s41586-022-05060-x.
- 726 28. Tucker, N. R. *et al.* Transcriptional and Cellular Diversity of the Human Heart.
727 *Circulation* 142, 466–482 (2020).
- 728 29. Chaffin, M. *et al.* Single-nucleus profiling of human dilated and hypertrophic
729 cardiomyopathy. *Nature* (2022) doi:10.1038/S41586-022-04817-8.
- 730 30. RC, W. *et al.* Atheroprotective roles of smooth muscle cell phenotypic modulation and the
731 TCF21 disease gene as revealed by single-cell analysis. *Nat Med* 25, 1280–1289 (2019).
- 732 31. Zhao, Q. *et al.* A cell and transcriptome atlas of the human arterial vasculature. *bioRxiv*
733 2024.09.10.612293 (2024) doi:10.1101/2024.09.10.612293.
- 734 32. Amrute, J. M. *et al.* Targeting immune–fibroblast cell communication in heart failure.
735 *Nature* 2024 1–11 (2024) doi:10.1038/s41586-024-08008-5.
- 736 33. Amrute, J. M. *et al.* Defining cardiac functional recovery in end-stage heart failure at
737 single-cell resolution. *Nature cardiovascular research* 2, 399–416 (2023).
- 738 34. Turner, A. W. *et al.* Single-nucleus chromatin accessibility profiling highlights regulatory
739 mechanisms of coronary artery disease risk. *Nature Genetics* 2022 54:6 54, 804–816
740 (2022).
- 741 35. Smith, E. L., Mok, G. F. & Münsterberg, A. Investigating chromatin accessibility during
742 development and differentiation by ATAC-sequencing to guide the identification of cis-
743 regulatory elements. *Biochem Soc Trans* 50, 1167 (2022).
- 744 36. Sun, Y., Miao, N. & Sun, T. Detect accessible chromatin using ATAC-sequencing, from
745 principle to applications. *Hereditas* 156, 29 (2019).
- 746 37. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory
747 epigenome. *Nat Rev Genet* 20, 207–220 (2019).
- 748 38. Duren, Z. *et al.* Regulatory analysis of single cell multiome gene expression and
749 chromatin accessibility data with scREG. *Genome Biol* 23, 1–19 (2022).
- 750 39. Wang, S. K. *et al.* Single-cell multiome of the human retina and deep learning nominate
751 causal variants in complex eye diseases. *Cell Genomics* 2, 100164 (2022).

- 752 40. Mitra, S. *et al.* Single-cell multi-ome regression models identify functional and disease-
753 associated enhancers and enable chromatin potential analysis. *Nature Genetics* 2024 56:4
754 **56**, 627–636 (2024).
- 755 41. Zhu, K. *et al.* Multi-omic profiling of the developing human cerebral cortex at the single-
756 cell level. *Sci Adv* **9**, (2023).
- 757 42. Badia-i-Mompel, P. *et al.* Gene regulatory network inference in the era of single-cell
758 multi-omics. *Nat Rev Genet* **24**, 739–754 (2023).
- 759 43. Mathys, H. *et al.* Single-cell multiregion dissection of Alzheimer’s disease. *Nature* 2024
760 632:8026 **632**, 858–868 (2024).
- 761 44. Xiong, X. *et al.* Epigenomic dissection of Alzheimer’s disease pinpoints causal variants
762 and reveals epigenome erosion. *Cell* **186**, 4422–4437.e21 (2023).
- 763 45. Ma, S. *et al.* Chromatin Potential Identified by Shared Single-Cell Profiling of RNA and
764 Chromatin. *Cell* **183**, 1103–1116.e20 (2020).
- 765 46. Gschwind, A. R. *et al.* An encyclopedia of enhancer-gene regulatory interactions in the
766 human genome. *bioRxiv* (2023) doi:10.1101/2023.11.09.563812.
- 767 47. Schnitzler, G. R. *et al.* Convergence of coronary artery disease genes onto endothelial cell
768 programs. *Nature* **626**, 799–807 (2024).
- 769 48. Nathan, A. *et al.* Single-cell eQTL models reveal dynamic T cell state dependence of
770 disease loci. | *Nature* | **606**, (2022).
- 771 49. Rao, S. S. P. *et al.* A three-dimensional map of the human genome at kilobase resolution
772 reveals principles of chromatin looping. *Cell* **159**, 1665 (2014).
- 773 50. Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome
774 architecture. *Nat Methods* **13**, 919–922 (2016).
- 775 51. Gate, R. E. *et al.* Genetic determinants of co-accessible chromatin regions in activated T
776 cells across humans. *Nature Genetics* 2018 50:8 **50**, 1140–1150 (2018).
- 777 52. Degner, J. F. *et al.* DNase I sensitivity QTLs are a major determinant of human expression
778 variation. *Nature* 2012 482:7385 **482**, 390–394 (2012).
- 779 53. Xiong, X. *et al.* Epigenomic dissection of Alzheimer’s disease pinpoints causal variants
780 and reveals epigenome erosion. *Cell* **186**, 4422 (2023).
- 781 54. Cuomo, A. S. E., Nathan, A., Raychaudhuri, S., MacArthur, D. G. & Powell, J. E. Single-
782 cell genomics meets human genetics. *Nat Rev Genet* **24**, 535 (2023).
- 783 55. Fitzgerald, T., Jones, A. & Engelhardt, B. E. A Poisson reduced-rank regression model for
784 association mapping in sequencing data. *BMC Bioinformatics* **23**, (2022).
- 785 56. Luo, Y. *et al.* New developments on the Encyclopedia of DNA Elements (ENCODE) data
786 portal. *Nucleic Acids Res* **48**, D882–D889 (2020).
- 787 57. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome.
788 *Nature* **489**, 57–74 (2012).
- 789 58. Phanstiel, D. H. *et al.* Static and Dynamic DNA Loops form AP-1-Bound Activation Hubs
790 during Macrophage Development. *Mol Cell* **67**, 1037–1048.e6 (2017).
- 791 59. Rustamaji, H. C., Kusuma, W. A., Nurdianti, S. & Batubara, I. Community detection with
792 Greedy Modularity disassembly strategy. *Scientific Reports* 2024 14:1 **14**, 1–17 (2024).
- 793 60. Nikpay, M. *et al.* A comprehensive 1000 Genomes–based genome-wide association meta-
794 analysis of coronary artery disease. *Nature Genetics* 2015 47:10 **47**, 1121–1130 (2015).
- 795 61. Elmentaite, R., Domínguez Conde, C., Yang, L. & Teichmann, S. A. Single-cell atlases:
796 shared and tissue-specific cell types across human organs. *Nat. Rev. Genet.* **23**, 395–410
797 (2022).

- 798 62. Litviňuková, M. *et al.* Cells of the adult human heart. **588**, 466–472 (2020).
799 63. Reichart, D. *et al.* Pathogenic variants damage cell composition and single-cell
800 transcription in cardiomyopathies. *Science (1979)* **377**, (2022).
801 64. Kartha, V. K. *et al.* Functional inference of gene regulation using single-cell multi-omics.
802 *Cell Genomics* **2**, 100166 (2022).
803 65. Baysoy, A., Bai, Z., Satija, R. & Fan, R. The technological landscape and applications of
804 single-cell multi-omics. *Nature Reviews Molecular Cell Biology* **24**:10 **24**, 695–713
805 (2023).
806 66. GitHub - EngreitzLab/sc-E2G: Pipeline to run sc-E2G. [https://github.com/EngreitzLab/sc-](https://github.com/EngreitzLab/sc-E2G)
807 [E2G](https://github.com/EngreitzLab/sc-E2G).
808 67. Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human
809 tissues. *Science (1979)* **369**, 1318–1330 (2020).
810 68. Kumasaka, N., Knights, A. J. & Gaffney, D. J. Fine-mapping cellular QTLs with
811 RASQUAL and ATAC-seq. *Nat Genet* **48**, 206–213 (2016).
812 69. Alencar, G. F. *et al.* Stem Cell Pluripotency Genes Klf4 and Oct4 Regulate Complex SMC
813 Phenotypic Changes Critical in Late-Stage Atherosclerotic Lesion Pathogenesis.
814 *Circulation* **2045–2059** (2020) doi:10.1161/CIRCULATIONAHA.120.046672.
815 70. Allahverdian, S., Chaabane, C., Boukais, K., Francis, G. A. & Bochaton-Piallat, M. L.
816 Smooth muscle cell fate and plasticity in atherosclerosis. *Cardiovasc Res* **114**, 540 (2018).
817 71. Zhao, Y. *et al.* “Stripe” transcription factors provide accessibility to co-binding partners in
818 mammalian genomes. *Mol Cell* **82**, 3398–3411.e11 (2022).
819 72. Gehrke, A. R. *et al.* Acoel genome reveals the regulatory landscape of whole-body
820 regeneration. *Science (1979)* **363**, (2019).
821 73. Schleif, R. DNA looping. *Annu Rev Biochem* **61**, 199–223 (1992).
822 74. Matthews, K. S. DNA looping. *Microbiol Rev* **56**, 123 (1992).
823 75. Hansen, A. S., Cattoglio, C., Darzacq, X. & Tjian, R. Recent evidence that TADs and
824 chromatin loops are dynamic structures. *Nucleus* **9**, 20 (2018).
825 76. Grubert, F. *et al.* Landscape of cohesin-mediated chromatin loops in the human genome.
826 *Nature* **2020 583:7818** **583**, 737–743 (2020).
827 77. Xu, J. *et al.* Subtype-specific 3D genome alteration in acute myeloid leukaemia. *Nature*
828 **2022 611:7935** **611**, 387–398 (2022).
829 78. Heffel, M. G. *et al.* Temporally distinct 3D multi-omic dynamics in the developing human
830 brain. *Nature* **2024 17**, 1–9 (2024).
831 79. Lambuta, R. A. *et al.* Whole-genome doubling drives oncogenic loss of chromatin
832 segregation. *Nature* **2023 615:7954** **615**, 925–933 (2023).
833 80. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature*
834 **2021 593:7858** **593**, 238–243 (2021).
835 81. Flavahan, W. A. *et al.* Altered chromosomal topology drives oncogenic programs in SDH-
836 deficient GISTs. *Nature* **2019 575:7781** **575**, 229–233 (2019).
837 82. Erdmann-Pham, D. D. *et al.* Tracing cancer evolution and heterogeneity using Hi-C.
838 *Nature Communications* **2023 14:1** **14**, 1–17 (2023).
839 83. Yang, J. *et al.* Analysis of chromatin organization and gene expression in T cells identifies
840 functional genes for rheumatoid arthritis. *Nature Communications* **2020 11:1** **11**, 1–13
841 (2020).

- 842 84. Ron, G., Globerson, Y., Moran, D. & Kaplan, T. Promoter-enhancer interactions
843 identified from Hi-C data using probabilistic models and hierarchical topological domains.
844 *Nature Communications* 2017 8:1 **8**, 1–12 (2017).
- 845 85. Schöpflin, R. *et al.* Integration of Hi-C with short and long-read genome sequencing
846 reveals the structure of germline rearranged genomes. *Nature Communications* 2022 13:1
847 **13**, 1–15 (2022).
- 848 86. Ing-Simmons, E. *et al.* Independence of chromatin conformation and gene regulation
849 during *Drosophila* dorsoventral patterning. *Nature Genetics* 2021 53:4 **53**, 487–499
850 (2021).
- 851 87. Schmitt, A. D. *et al.* A Compendium of Chromatin Contact Maps Reveal Spatially Active
852 Regions in the Human Genome. *Cell Rep* **17**, 2042 (2016).
- 853 88. Mosquera, J. V. *et al.* Integrative single-cell meta-analysis reveals disease-relevant
854 vascular cell states and markers in human atherosclerosis. (2023)
855 doi:10.1016/j.celrep.2023.113380.
- 856 89. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell
857 chromatin accessibility analysis. *Nature Genetics* 2021 53:3 **53**, 403–411 (2021).
- 858 90. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-
859 seq data using regularized negative binomial regression. *Genome Biol* **20**, 1–15 (2019).
- 860 91. Gaspar, J. M. Improved peak-calling with MACS2. doi:10.1101/496521.
- 861 92. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-
862 wide association summary statistics. *Nat Genet* **47**, 1228–1235 (2015).
- 863 93. Hinrichs, A. S. *et al.* The UCSC Genome Browser Database: update 2006. *Nucleic Acids*
864 *Res* **34**, (2006).
- 865 94. Suzuki, K. *et al.* Genetic drivers of heterogeneity in type 2 diabetes pathophysiology.
866 *Nature* **627**, 347–357 (2024).
- 867 95. Roychowdhury, T. *et al.* Genome-wide association meta-analysis identifies risk loci for
868 abdominal aortic aneurysm and highlights PCSK9 as a therapeutic target. *Nat Genet* **55**,
869 1831–1842 (2023).
- 870 96. Keaton, J. M. *et al.* Genome-wide analysis in over 1 million individuals of European
871 ancestry yields improved polygenic risk scores for blood pressure traits. *Nat Genet* **56**,
872 778–791 (2024).
- 873 97. Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects
874 identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet* **50**, 524–537
875 (2018).
- 876 98. Murphy, A. E., Schilder, B. M. & Skene, N. G. MungeSumstats: a Bioconductor package
877 for the standardization and quality control of many GWAS summary statistics.
878 *Bioinformatics* **37**, 4593–4596 (2021).
- 879 99. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**,
880 1284–1287 (2016).
- 881 100. Loh, P. R. *et al.* Reference-based phasing using the Haplotype Reference Consortium
882 panel. *Nat Genet* **48**, 1443–1448 (2016).
- 883 101. Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: a discriminative
884 modeling approach for rapid and robust local-ancestry inference. *Am J Hum Genet* **93**,
885 278–288 (2013).
- 886 102. Auton, A. *et al.* A global reference for human genetic variation. *Nature* **526**, 68–74
887 (2015).

- 888 103. Benaglio, P. *et al.* Mapping genetic effects on cell type-specific chromatin accessibility
889 and annotating complex immune trait variants using single nucleus ATAC-seq in
890 peripheral blood. *PLoS Genet* **19**, (2023).
- 891 104. Munz, M. *et al.* Qtlizer: comprehensive QTL annotation of GWAS results. *Sci Rep* **10**,
892 (2020).
- 893 105. Zhou, W. *et al.* Efficient and accurate mixed model association tool for single-cell eQTL
894 analysis. *medRxiv* (2024) doi:10.1101/2024.05.15.24307317.
- 895 106. Fulco, C. P. *et al.* Activity-by-contact model of enhancer–promoter regulation from
896 thousands of CRISPR perturbations. *Nature Genetics* **51**, 1664–1669 (2019).
- 897 107. Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics
898 pipelines. *Nat Biotechnol* **38**, 276–278 (2020).
- 899 108. Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing.
900 *Genome Biol* **16**, (2015).
- 901 109. Marchal, C., Singh, N., Corso-Díaz, X. & Swaroop, A. HiCRes: a computational method
902 to estimate and predict the genomic resolution of Hi-C libraries. *Nucleic Acids Res* **50**,
903 E35 (2022).
- 904 110. Abdennur, N. *et al.* Cooltools: Enabling high-resolution Hi-C analysis in Python. *PLoS*
905 *Comput Biol* **20**, (2024).
- 906 111. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals
907 principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
- 908 112. Kaul, A., Bhattacharyya, S. & Ay, F. Identifying statistically significant chromatin
909 contacts from Hi-C data with FitHiC2. *Nat Protoc* **15**, 991–1012 (2020).
- 910 113. Xu, W. *et al.* CoolBox: a flexible toolkit for visual analysis of genomics data. *BMC*
911 *Bioinformatics* **22**, (2021).
- 912 114. Clauset, A., Newman, M. E. J. & Moore, C. Finding community structure in very large
913 networks. (2004).
- 914 115. Bioconductor - RBGL. <https://bioconductor.org/packages/release/bioc/html/RBGL.html>.
- 915 116. Anene-Nzelu, C. G. *et al.* Assigning Distal Genomic Enhancers to Cardiac Disease-
916 Causing Genes. *Circulation* **142**, 910–912 (2020).
- 917 117. Wingett, S. *et al.* HiCUP: Pipeline for mapping and processing Hi-C data. *F1000Res* **4**,
918 (2015).
- 919 118. Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution
920 Hi-C Experiments. *Cell Syst* **3**, 95–98 (2016).
- 921 119. Bhattacharyya, S., Chandra, V., Vijayanand, P. & Ay, F. Identification of significant
922 chromatin contacts from HiChIP data by FitHiChIP. *Nat Commun* **10**, (2019).
- 923 120. Tan, W. L. W. *et al.* Epigenomes of Human Hearts Reveal New Genetic Variants
924 Relevant for Cardiac Disease and Phenotype. *Circ Res* **127**, 761–777 (2020).
- 925 121. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler
926 transform. *Bioinformatics* **25**, 1754–1760 (2009).
- 927 122. Kumar, V. *et al.* Uniform, optimal signal processing of mapped deep-sequencing data. *Nat*
928 *Biotechnol* **31**, 615–622 (2013).
- 929 123. Xie, S., Duan, J., Li, B., Zhou, P. & Hon, G. C. Multiplexed Engineering and Analysis of
930 Combinatorial Enhancer Activity in Single Cells. *Mol Cell* **66**, 285-299.e5 (2017).
- 931 124. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for
932 CRISPR screening. *Nat Methods* **11**, 783–784 (2014).
- 933
- 934

935 **Materials and Methods**

936

937 **Ethical Approval for Human Specimens**

938 The study is compliant with all relevant ethical regulations and has been approved by the
939 Washington University School of Medicine Institutional Review Board (IRB #201104172).
940 Informed consent was obtained from each patient prior to tissue collection by Washington
941 University School of Medicine and no compensation was provided in exchange for subject
942 participation in the study. All demographic and clinical data has been de-identified and provided
943 in Supplementary Table 1. Patients included in this study span diverse race, age, and sex to
944 provide an inclusive trans-ethnic study population.

945

946 **Inclusion Criteria**

947 Prior to tissue collection, specific inclusion criteria were employed to ensure well controlled study
948 groups. Any patients with HIV or hepatitis and known genetic cardiomyopathies were excluded
949 from this study. The left anterior descending coronary artery was isolated, and flash frozen from
950 donor hearts: patients with stable ejection fractions, no known history of cardiac disease and
951 experienced a non-cardiac cause of death/transplant and from patients with chronic heart failure.
952 For all samples the proximal left anterior descending coronary artery was used.

953

954 **Nuclei isolation for Multiome sequencing**

955 The left anterior descending coronary artery was dissected from explanted hearts, epicardial fat
956 removed, and arteries were flash frozen using liquid nitrogen. Identical regions from the proximal
957 left anterior descending artery were used from all patients. Single nuclei suspensions were
958 generated as previously described³². Nuclei were isolated according to 10x Genomics protocol
959 (CG00375; Nuclei Isolation Complex Sample for ATAC GEX Sequencing RevB) and flow
960 cytometry for 7-AAD (Sigma; SML1633-1ML) positive nuclei was used for sorting using a BD
961 FACS Melody (BD Biosciences) with a 100uM nozzle. Protocol CG000338 from 10x Genomics
962 was used for Chromium Next GEM Single Cell Multiome ATAC + Gene Expression. Briefly,
963 following nuclei isolation, permeabilization was performed, followed by transposition, GEM
964 generation and barcoding using ChipJ (10x Genomics; PN1000234), post-GEM clean up, pre-
965 amplification PCR, cDNA amplification, library construction, and sequencing. Gene expression
966 and ATAC libraries were sequenced to a read depth of 50,000 and 25,000 respectively using a
967 NovaSeq 6000 platform (Illumina) at the McDonnell Genome Institute.

968

969 **Multiome data processing**

970 Raw fastq files were aligned to the human GRCh38 reference genome (v) using CellRanger ARC
971 (10x Genomics, v6.1). ArchR⁸⁹ (<https://www.archrproject.com>) was used to process the ATAC
972 fragments and Seurat was used to process RNA. Quality control was performed to keep nuclei
973 with the following: TSS enrichment > 2, nFragments > 1000, 200 < nUMI GEX < 50,000, and percent
974 mito < 5%. Post-QC nuclei were used for doublet removal in ArchR (ATAC information) and then
975 using scrublet (RNA information). Raw RNA counts were normalized and scaled using
976 SCTransform⁹⁰ regressing out percent mitochondrial reads and nCount_RNA. Principal
977 component analysis, harmony batch integration (by sample), nearest neighbor clustering, and
978 UMAP embedding construction was then performed in Seurat. Cell types were annotated using
979 different expression and knowledge of canonical gene markers. The RNA annotations and
980 normalized gene expression matrix was added to the ArchR project onto the nuclei with the same
981 barcodes. ArchR was used to construct pseudobulk replicates across ATAC clusters and peaks
982 were called using MACS2⁹¹. The ArchR getMarkerFeatures function was used to identify peaks
983 that are unique to each cell type. The addPeak2GeneLinks function from ArchR was then used
984 to calculate peak to gene links using gene expression and accessibility from the same nucleus
985 with a correlation cut-off of 0.3. To visualize the correspondence between the p2g links, a heatmap

986 was constructed which shows ATAC and RNA z-scores with rows clustered using k-means
987 clustering through the ArchR package plotPeak2GeneHeatmap. We have computed the Pearson
988 correlation coefficient for all genes between the gene expression and accessibility vectors. The
989 ArchR peak 2 gene linkage function was used to identify putative CREs for each gene. In ArchR
990 the addIterativeLSI function was used on the RNA and ATAC modalities respectively with default
991 parameters followed by harmony integration to get modality specific clustering and dimensional
992 reductions. To generate a joint RNA/ATAC embedding, the ArchR addCombinedDims function
993 was used and this combined embedding was used for subsequent UMAP construction and
994 clustering. To compare how the ATAC derived cell clustering compares to the RNA and joint
995 embedding, we generated a confusion matrix.

996 **LDSC and GWAS Variant Overlap**

998 To test for enrichment of disease risk in the four major cell types, stratified LD score regression
999 was used to partition heritability in cell-type marker peaks⁹². First, cell-type specific marker peaks
1000 (FDR < 0.01, log₂FC > 1) were obtained using ArchR and converted to GRCh37 coordinates
1001 using liftOver⁹³. GWAS summary statistics for coronary artery disease⁸, diabetes⁹⁴, abdominal
1002 aortic aneurysms⁹⁵, diastolic/systolic blood pressure⁹⁶, and stroke⁹⁷ were obtained, and their
1003 formats were standardized using MungeSumStats package in R. Stratified LD score analysis was
1004 then carried out with --h2-cts flag for cell-type specific analyses. To examine overlap between
1005 CAD GWAS variants and ATAC peaks, lead variants from 241 genome-wide significant ($p < 5 \times$
1006 10^{-8}) loci and 897 conditionally independent variants meeting FDR cutoff (FDR < 0.01) were
1007 obtained from Aragam *et al.*^{8,98}. Additional variants that are in high LD ($r^2 > 0.80$ within 250KB)
1008 with these lead variants were obtained from the 1000 Genomes European panel using plink --tag-
1009 r2 flag. Additionally, these variants were combined with the functional fine-mapped credible set
1010 variants from Aragam *et al.* to create a superset of GWAS variants [cite Aragam *et al.*] to be tested
1011 for overlap. These variants were overlapped with ATAC peak regions using findOverlap function
1012 in IRanges R package.

1013 **Genotyping**

1015 To obtain high quality genotypes from patients, DNA from peripheral leukocytes of all individuals
1016 was genotyped using Illumina GSA-24-V3 SNP array. Following initial processing using Illumina
1017 GenomeStudio software, variants with both 1) minor allele frequency > 5% and call rate < 95% or
1018 2) minor allele frequency < 5% and call rate < 99% were excluded. No individuals were excluded
1019 by the call rate exclusion filter (<95%). Additional genotypes were further imputed from the
1020 TOPMED panel on the Michigan Imputation Server v. 1.7.1 using minimac4-1.0.2⁹⁹ and phasing
1021 with EAGLE¹⁰⁰. Following imputation, variants with imputation $R^2 < 0.3$ and MAF > 0.05 were
1022 further filtered for caQTL analysis for a final dataset containing 7,250,405 variants. To conduct
1023 combined caQTL mapping, the VCF files from Turner *et al.*³⁴ were merged with the imputed
1024 genotype dataset, and only shared SNPs in two datasets were kept for downstream analysis.
1025 Turner *et al.*³⁴ carried out low coverage whole genome-sequencing; further details are available in
1026 the original publication³⁴. Following the merge, 5,229,397 variants were available for QTL
1027 analysis. Local ancestries were inferred using RMMix2¹⁰¹ with default settings. YRI (n=186) and
1028 CEU (n=183) from the 1000 Genome Project¹⁰² were used as AFR / EUR reference populations,
1029 respectively.

1030 **Pseudobulk Chromatin Accessibility QTL discovery**

1032 Similar to previous studies for single cell caQTL discovery^{34,103}, we used RASQUAL⁶⁸ to identify
1033 pseudobulk caQTLs in four major cell types (SMC, Endothelial, Myeloid, Fibroblast). RASQUAL
1034 maximizes power for caQTL discovery by simultaneously modeling allelic imbalance and total
1035 read counts in each locus. To generate fragment counts within each ATAC peak regions, we
1036 loaded each fragment object and used Signac's FeatureMatrix function to generate a cell-by-

1037 fragment count matrix for all cells within each cell type⁶⁸. This matrix was subsequently
1038 aggregated at a sample pseudobulk level using rollup function from R slam package. For
1039 RASQUAL's allele-specific modeling, createASVCF.sh script in paired-read mode was used to
1040 generate allele-specific fragment counts at each site with additional BAM flag of $-F$ 1280 to
1041 exclude secondary or PCR duplicate reads. For each cell type, only samples with at least 20 cells
1042 of that cell type and only peaks with at least 5 reads on average across all samples were included
1043 for the final analysis.

1044
1045 For the caQTL association analysis, all variants within a \pm 10KB *cis*-window of the peaks were
1046 tested for association. Total library size for each sample was included as an offset. Age, sex,
1047 sequencing site, and first four principal components of genotype data (generated from plink $--pca$)
1048 were included as covariates. For multiple testing correction, we employed a two-step correction
1049 procedure recommended by RASQUAL. First, to correct for locus-wide multiple testing, a q-value
1050 corresponding to SNP level FDR was calculated using Benjamin-Hochberg method, and the SNP
1051 with the lowest q-value was selected for each locus. For multi-locus multiple testing correction,
1052 an empirical null distribution was calculated by running RASQUAL four times with $--random$ -
1053 permutation flag, which performs association analysis with randomly permuted genotypes. The
1054 four runs were averaged to estimate a null-distribution of locus-level q-values, which was
1055 subsequently used to obtain q-value cutoffs that would correspond to 1%, 5%, and 10% genome-
1056 wide FDR. The caQTL variants were subsequently queried in a database of bulk eQTL datasets
1057 using the Qtlizer package¹⁰⁴.

1058
1059 Pseudobulk caQTLs for the major cell types can be found in Supplementary Tables 14, 16, 18,
1060 and 20 at the FDR 5% cut-off. Pseudobulk caQTLs overlapping with GTEx arterial eQTLs at the
1061 FDR 5% cut-off can be found in Supplementary Tables 15, 17, 19, and 21.

1062 1063 **Single Cell Chromatin Accessibility QTL Discovery**

1064 In addition to the pseudobulk caQTL model, we utilized a Poisson mixed effect (PME) single cell
1065 model to map caQTLs in smooth muscle cells. This approach models cell-level fragment counts
1066 with random effect covariates to account for relatedness shared by cells from the same sample
1067 and batch. Such attempts to model discrete count distributions at a single cell-level has been
1068 shown to generally boost power in single-cell QTL discovery compared to pseudobulk and/or
1069 linear regression frameworks^{48,55,105}. Our PME model was adapted from the Poisson mixed effect
1070 model utilized by Nathan *et al*⁴⁸ for single cell eQTL mapping as follows:

$$1071$$
$$1072 \log(A) = \beta_G X_G + \beta_{age} X_{age} + \beta_{sex} X_{sex} + \beta_{nFragments} \log(X_{nFragments}) + \beta_{TSSEnrich} X_{TSSEnrich} +$$
$$1073 \sum_{k=1}^4 \beta_{accPCk} X_{accPCk} + \sum_{k=1}^4 \beta_{genoPCk} X_{genoPCk} + (\varphi_d | d) + (\epsilon_s | s) + \theta$$
$$1074$$

1075 Where ATAC fragment counts (A) was modeled with fixed effect covariates (G=genotype,
1076 nFragments=number of fragments per cell, TSSEnrich=TSS enrichment per cell, accPC=principal
1077 components of single-cell ATAC counts, genoPC=principal components of genotype data),
1078 random effect covariates (d=donor ID, s=sequencing site), and intercept (θ). All quantitative
1079 covariates were scaled and centered prior to the regression analysis. Significance of the genotype
1080 effect term (β_G) was tested by likelihood ratio test (LRT) of the full model containing the genotype
1081 term and a null model without the genotype term. Due to the high computational load of running
1082 Poisson regression analysis on all cells, this analysis was restricted to lead caQTLs variants from
1083 the pseudobulk RASQUAL analysis, which models the fragment counts using a negative binomial
1084 model. All Poisson models were fit using glmer function from the lme4 R package. We confirmed
1085 that our data is generally not over-dispersed by plotting the mean variance relationship between
1086 a random subset of peaks. Similarly, utilizing a zero-inflated Poisson regression model or a

1087 negative binomial model on a subset of our hits revealed little differences in overall regression
1088 estimates. Given that a small proportion of our pseudobulk hits likely represent null results, we
1089 did not conduct FDR correction due to our inability to accurately estimate the proportion of results
1090 from the null distribution. Instead, we compared Z-values between our pseudobulk and PME
1091 models.

1092 A key benefit of modeling individual cells instead of sample-level pseudobulk is our ability to
1093 simultaneously model cell-state covariates (e.g. differentiation trajectories, gene/protein
1094 expression) as well as potential interaction between genotypes and cell-state covariates. We
1095 defined fibromyocyte (FMC) identity score by calculating the gene activity of the top 100
1096 fibromyocyte (modulated SMC) marker genes defined by Wirka *et al*³⁰ from our integrated
1097 snATAC-seq meta map using ArchR's addGeneScoreMatrix() function. To test for "dynamic"
1098 caQTL where the genotype effect is modified by SMC cell state, we then added the FMC score
1099 and a genotype-FMC score interaction term as covariates in the following model:

$$\begin{aligned} 1100 & \\ 1101 & \log(A) = \beta_G X_G + \beta_{age} X_{age} + \beta_{sex} X_{sex} + \beta_{nFrag} \log(X_{nFrag}) + \beta_{TSSE} X_{TSSE} + \\ 1102 & \sum_{k=1}^4 \beta_{accPCK} X_{accPCK} + \sum_{k=1}^4 \beta_{genoPCK} X_{genoPCK} + (\varphi_d | d) + (\epsilon_s | s) + \beta_{FMC} X_{FMC} + \\ 1103 & \beta_{FMC \times G} X_{FMC \times G} + \theta \end{aligned}$$

1104
1105 Again, this analysis was carried out just on the significant (FDR<5%) lead variants from the
1106 pseudobulk caQTL analyses, though it is entirely possible that there are non-significant
1107 pseudobulk caQTLs (i.e. genotype effect is near zero for average FMC score) that may have
1108 significant interaction terms. The significance of the interaction term was tested by likelihood ratio
1109 test of PME models with and without the interaction term. The LRT p-values were then corrected
1110 for multiple testing using the Benjamin-Hochberg method.

1111
1112 Dynamic caQTLs can be found in Supplementary Tables 22.

1113 **scE2G model predictions**

1114 We used the scE2G^{Multiome} model⁶⁶ to predict enhancer–gene connections in 11 cell types from
1115 10x Multiome data (Maya Sheth, Wei-Lin Qiu, Robin Andersson, and Jesse Engreitz, in
1116 preparation). In each cell type, scE2G scores every candidate element–gene pair (where
1117 candidate elements are ATAC peaks within 5 Mb of the transcription start site of the gene) by
1118 integrating several features. Regulatory enhancer–gene interactions were defined as element–
1119 gene pairs with a score greater than 0.171.

1120
1121 The scE2G^{Multiome} model is a supervised classifier adapted for multiomic data from ENCODE-
1122 rE2G⁴⁶. It integrates eight features, including 1) a pseudobulk Activity-By-Contact (ABC) score¹⁰⁶,
1123 where 3D contact is estimated by an inverse function of genomic distance; 2) the Kendall
1124 correlation across single cells between element accessibility and gene expression; 3) whether the
1125 gene is "ubiquitously expressed," and 4) several other measures of genomic distance and
1126 chromatin accessibility around the element and promoter. The score threshold of 0.171 was
1127 determined as the score yielding 70% recall when evaluating predictions in K562 cells against
1128 CRISPRi-validated enhancer-gene pairs⁴⁶.

1129
1130 scE2G predictions for all cell types can be found in Supplementary Tables 3-13.

1131 **Human coronary artery Hi-C Data Collection and Processing**

1132 Bulk *in-situ* Hi-C libraries for 16 different coronary artery samples across a multi-ethnic cohort
1133 were generated with the Arima-HiC kit, according to the manufacturer's protocols. Specifically,
1134 one Hi-C library was created from each biological sample, with an input of ~800k cells each.
1135
1136

1137 Restriction enzymes cutting at \wedge GATC and G \wedge ANTC were utilized in library construction, and
1138 approximately 600 million reads were sequenced on an Illumina HiSeq 4000 per library.

1139
1140 To process the data, we used version 2.0.0 of the nf-core¹⁰⁷ Hi-C pipeline (Zenodo:
1141 <https://doi.org/10.5281/zenodo.7556794>). While we largely used default parameters, we did
1142 modify the pipeline by following a recommendation from Arima to trim the first 5 bases from the
1143 5' end of each adapter sequence, which we accomplished with Cutadapt version 3.4
1144 (<https://doi.org/10.14806/ej.17.1.200>). Arima recommended this change to help increase the
1145 mapping rate when using pipelines such as HiC-Pro¹⁰⁸, and we did observe an increase in
1146 percentage of successfully mapped reads upon implementation. We also set the nf-core pipeline
1147 parameter `min_mapq` to 0, to maximize our retained reads prior to further filtering. Briefly, the
1148 pipeline aligned paired-end sequencing reads to human genome hg38, and trimmed unaligned
1149 reads with a ligation junction before attempted re-alignment (“rescue” of chimeric reads).
1150 Ultimately, ICE-normalized contact maps at 2kb, 5kb, 10kb, 50kb, 100kb, 500kb, and 1Mb
1151 resolutions were generated. We assessed maximum library resolution using HiCRes version
1152 2.0¹⁰⁹, and ultimately decided to focus our analyses on the 10kb maps based on the range of
1153 inferred resolutions across samples.

1154
1155 To assess significant loops within each Hi-C library, we used two different algorithms: the
1156 cooltools version 0.7.1¹¹⁰ “dots” function (based on HiCCUPS¹¹¹), and FitHiC version 2.0.8¹¹². For
1157 cooltools, we analyzed loops at 2kb, 5kb, and 10kb resolutions, using default parameters for each
1158 and donut-based kernels of $w=(7, 7, 5)$ and $p=(4, 4, 2)$, respectively. For FitHiC, we analyzed the
1159 same resolutions, and performed a single spline pass per sample, allowing for bias values down
1160 to a minimum of 0 based on our observed biases across samples. In both methods, we retained
1161 identified loops with q -values ≤ 0.05 as significant. To interrogate topologically associating
1162 domains (TADs), we employed cooltools insulation score algorithm, with window sizes of 150kb,
1163 250kb, and 500kb. We kept the boundaries identified as “strong” at each window size to define
1164 genomic intervals representing TAD bodies. To visualize genome-wide Hi-C contact maps, we
1165 also used cooltools. For visualization of individual loci in concert with gene, loop, and TAD
1166 annotations, we used coolbox version 0.3.8¹¹³. In figures with TAD annotations, TADs across all
1167 tested window sizes are shown, whereas significant loop annotations focus on 10kb loop calls
1168 only. Intra-chromosomal loop and TAD sizes within each sample were calculated by subtracting
1169 the lower bin’s start coordinate from the higher bin’s start coordinate. These data were then
1170 visualized using ggplot2.

1171
1172 **HiC network analysis**
1173 We kept loops detected in at least 2 patient coronaries for this analysis. To construct an un-biased
1174 connected chromatin network⁵⁸, we detected communities of interacting HiC loop anchors using
1175 a fast greedy modularity optimization algorithm^{59,114}. Using the matrix of connected DNA loops,
1176 we used the RBGL package (R interface for Boost graph library algorithms)¹¹⁵ – specifically, we
1177 used the `ftM2graphNEL` function to create a graph from the connected matrix and then the
1178 `connectedComp` function to build a connected network graph where each node is a HiC loop and
1179 the edges are linking loops to other loops whose anchors overlap. Visualization of the graphs was
1180 done using the `Rgraphviz` package in R. To construct a variant-to-enhancer-to-network (V2E2N)
1181 map, we identified loops where any anchor in the network contains a V2E + ENCODE coronary
1182 artery H3K27ac peak. Finally, to build a catalogue of all genes predicted to be regulated within a
1183 chromatin network, we overlapped all HiC anchors in a network with the E2G to construct a final
1184 V2E2N with implicated genes.

1185

1186 **H3K27ac HCASMC and HCAEC HiChIP**

1187 HiChIP was adapted from previously published protocols with some modifications^{50,116}. Briefly, at
1188 least one million human healthy coronary artery smooth muscle cells (HCASMC) (Thermo
1189 Scientific, Product code C-017-5C, Lot. 1689414) and human healthy coronary artery endothelial
1190 cells (HCAEC) (ATCC, Product code PCS-100-020, Lot 62382179) were crosslinked with 1%
1191 Formaldehyde. The initial Hi-C portion followed the Arima HiC protocol described in the Arima
1192 High Coverage Hi-C Kit (Arima Genomics, Material Part Number: A410110). For the
1193 immunoprecipitation step, sheared chromatin after the Hi-C portion of the protocol was
1194 immunoprecipitated with 5 µg of H3K27ac antibody (Abcam, ab4729) with 50 µl of protein G
1195 Dynabeads (Invitrogen 10004D) and allowed to incubate overnight at 4°C. The protein G
1196 Dynabeads were washed and the immunoprecipitated DNA was de-crosslinked at 67°C for 2
1197 hours and eluted (50 mM sodium bicarbonate, 1% SDS). Subsequently, Dynabeads MyOne
1198 Streptavidin C1 beads (Invitrogen 65002) were used to enrich ligated junctions as per Hi-C
1199 protocol. Library preparation was performed using the NEB Ultra II library preparation kit (NEB,
1200 E7645L), according to the manufacturer's protocol. 12 PCR cycles were performed using indexed
1201 primers and subsequently DNA fragments between 300 to 500 bp were size selected using the
1202 Omega MagBind NGS cleanup Magnetic beads (Omega M1378). The libraries were sequenced
1203 by paired-end sequencing with at least 300 million read pairs at 2 x 151 bp read length on illumina
1204 HiSeq 4000 platforms. Raw paired-end reads in Fastq format for respective samples were
1205 processed using the HiCuP pipeline¹¹⁷ (Version 0.7.4). The hg38 human genome was used as
1206 reference to generate the HiCuP Digest file using two restriction motifs GATC and GANTC via
1207 the hicup_digester command for running the HiCuP pipeline. Paired contacts were extracted from
1208 aligned filtered bam files using Samtools (Version 1.13) and used as input for the pre function in
1209 juicer package to generate .hic files¹¹⁸. Paired interaction counts at 5kb resolution were extracted
1210 from the matrix, and High-confidence chromatin loops were identified using the FitHiChIP tool as
1211 previously described using default parameters¹¹⁹. Using HCASMC and HCAEC ChIP-Seq peak
1212 files as input, and initial paired interaction files, high-confidence peak-to-all interactions with a 5kb
1213 bin size were processed using the bash FitHiChIP_HiCPro.sh script ([https://ay-
1214 lab.github.io/FitHiChIP/](https://ay-lab.github.io/FitHiChIP/)). High-confidence chromatin interactions were sorted and indexed in bed
1215 format and converted to long-range format for visualization on the WashU browser. The overlay
1216 of HiChIP, ChIP-seq, and GWAS data was plotted using the locuszoomr package (version 0.3.5)
1217 and R (version 4.3.1). Raw Fastq files and processed HiChIP data will be uploaded to NCBI SRA
1218 upon publication of the manuscript.

1219

1220 **H3K27ac Chromatin Immunoprecipitation (ChIP-Seq)**

1221 Chromatin immunoprecipitation with sequencing (ChIP-seq) was performed as previously
1222 described with minor optimisations¹²⁰. Briefly, one million human coronary artery smooth muscle
1223 cells (HCASMC) and human coronary endothelial cells (HCAEC) from the same lot used for
1224 HiChIP experiments were crosslinked with 1% Formaldehyde for 10 min at room temperature and
1225 quenched with glycine (125mM) for 5 min. The cells were rinsed twice with cold 1XPBS and
1226 sonicated in nuclei lysis buffer (50mM HEPES-KOH, pH 7.5, 150mM NaCl, 1mM EDTA, 1% Triton
1227 X, 0.1% Sodium deoxycholate, 1% SDS, Takara, 1x protease inhibitor) using the Bioruptor
1228 sonicator to obtain chromatin fragments between 200 to 500 bp. The fragmented chromatin was
1229 immunoprecipitated overnight at 4°C, with 5 µg of H3K27ac antibody (Abcam, ab4729) and 50 µl
1230 of protein G beads (Invitrogen, 10004D). Subsequently, the beads were washed and
1231 immunoprecipitated DNA was de-crosslinked and eluted in elution buffer (50 mM Tris-HCl, pH7.5,
1232 and 10 mM EDTA). Eluted ChIP fragments were isolated using Phenol-Chloroform extraction
1233 method, and DNA was purified by ethanol precipitation. Library preparation was performed on the
1234 eluted ChIP DNA using the NEB Ultra II library preparation kit, according to the manufacturer's
1235 protocol. 10 PCR cycles were performed using indexed primers and the library of DNA fragments
1236 with sizes between 300 to 500 bp was selected using the Omega MagBind NGS cleanup Magnetic

1237 beads (Omega M1378). The libraries were sent for paired end sequencing with at least 50 million
1238 read pairs at 2 x 151 bp read length on the illumina HiSeq 4000 platform. Paired-end sequencing
1239 reads were aligned to human genome (hg38) using BWA mem version 0.7.5¹²¹, and reads were
1240 deduplicated using Picard MarkDuplicates function. Aligned mapped reads in BAM file were used
1241 as input for peak-calling using Dfilter¹²² to identify significant H3K27ac peaks using the following
1242 setting: -ks=60, -bs=100, -lpvalue=8. Bigwig tracks were normalized for sequencing depths and
1243 the sum of normalized binned tag-count represents the peak height. Raw Fastq files and
1244 processed ChIP-seq data will be uploaded to NCBI SRA upon manuscript publication.

1245
1246 **Culture of HCAEC and HCASMC primary cells**
1247 human healthy coronary artery smooth muscle cells (HCASMC) (Thermo Scientific, Product code
1248 C-017-5C, Lot. 1689414) are cultured in human vascular smooth muscle (VSMC) basal medium
1249 (Gibco, M231500), in Smooth Muscle Growth Supplement (SMGS) (Gibco, S00725) and 1%
1250 penicillin/streptomycin. human healthy coronary artery endothelial cells (HCAEC) (ATCC, Product
1251 code PCS-100-020, Lot 62382179) are cultured in Endothelial Cell Growth medium MV2
1252 (PromoCell, C-22022).

1253
1254 **Construction of lentiviral sgRNA plasmids and Lentivirus production**
1255 The Lentiguide-Puro plasmid was obtained as a gift from Feng Zhang (Addgene #52963,
1256 LentiGuide-Puro) in which the Non-Targeting Control sgRNAs and sgRNAs targeting the TSS and
1257 Regulatory enhancer were annealed and ligated into the Esp3I cut-site of the Lentiguide-Puro
1258 vector using T4 Ligase (NEB, M0202) as per manufacturer's instructions. Ligation products were
1259 transformed into One Shot Stbl3 Chemically competent E.coli (Invitrogen, C737303), and
1260 plasmids of positive clones were isolated using FavorPrep Plasmid DNA Extract Kit (Favorgen,
1261 FAPDE001-1). The CRISPR interference expression vector for Lenti-dCas9-KRAB-blast
1262 (Addgene, #89567) was a gift from Gary Hon. For lentivirus production, HEK293T cells were
1263 cultured in DMEM+10% FBS until 70% confluency prior to transfection. 10ug of individual
1264 plasmids, 7.5ug of pMDLg/RRE, 2.5ug of pRSV-Rev and 2.5ug of pMD2.G lentivirus packaging
1265 plasmids were co-transfected with 50ul of Polyethylenimine (PEI) diluted in 3ml of Opti-MEM™ I
1266 Reduced Serum Medium (#31985070, ThermoFisher Scientific). Medium was refreshed overnight
1267 with reduced 5% FBS in DMEM medium, and the expended medium was collected twice after
1268 24hr and 48 hr respectively. Pooled supernatant was filtered through a PES 0.45uM filter and viral
1269 particles were concentrated using Viro-PEG Lentivirus Concentrator (Ozbiosciences, #LVG100)
1270 as per manufacturer's instructions.

1271
1272 The sgRNAs are listed below:
1273 Non-Targeting Control_sgRNA_F: CACCGAACGTGCTGACGATGCGGGC
1274 Non-Targeting Control_sgRNA_R: AAACGCCCGCATCGTCAGCACGTTC
1275 AMOTL2_TSS_sgRNA_F: 5'-CACCGGCGCGAACAGCCAGAGCGT-3'
1276 AMOTL2_TSS_sgRNA_R: 5'- AAACACGCTCTGGCTGTTCGCGCC-3'
1277 AMOTL2_Enhancer_sgRNA1_F: 5'- CACCGTATTCATAGACATCACTAA-3'
1278 AMOTL2_Enhancer_sgRNA1_R: 5'- AAACCTTAGTGATGTCTATGAATAC-3'
1279 AMOTL2_Enhancer_sgRNA2_F: 5'- CACCGATCCCTATGGAATCCTTGG-3'
1280 AMOTL2_Enhancer_sgRNA2_R: 5'- AAACCCAAGGATTCCATAGGGATC-3'

1281
1282 **Transduction of primary cells for CRISPR interference and gene expression analysis**
1283 At least two independent sgRNAs targeting a single regulatory element were tested against the
1284 non-targeting control. sgRNA targeting the transcription start site (TSS) was used as positive
1285 control. Both human healthy coronary artery smooth muscle cells (HCASMC) (Thermo Scientific,
1286 Product code C-017-5C, Lot. 1689414) and human healthy coronary artery endothelial cells
1287 (HCAEC) (ATCC, Product code PCS-100-020, Lot 62382179) were co-transduced in 5ug/ml

1288 polybrene with lentiviral particles packaged from Lenti-dCas9-KRAB-blast (Addgene, #89567)¹²³,
1289 and individual sgRNAs cloned into Lentiguide-Puro backbone (Addgene, #52963)¹²⁴ at multiplicity
1290 of infection of 3. Respective HCASMC and HCAEC growth media were refreshed after 24h of
1291 infection, and co-selection with 10ug/ml blasticidin (Gibco, A1113903), and 1ug/ml puromycin
1292 (Gibco, A1113803) was performed after 72h post-infection. Selected cells were allowed to recover
1293 and expanded for 2 weeks before RNA isolation using Trizol Reagent (Gibco, 15596026) and
1294 Direct-Zol RNA miniprep (Zymo Research, R2052) as per manufacturer's instructions. cDNA
1295 conversion was performed with at least 200ng RNA using HiScript III RT SuperMix (Vazyme
1296 Biotech, R323-01) and RT-PCR was performed using gene-specific primers to assess for gene
1297 expression.

1298
1299 Below are the following RT-PCR primers used:

1300
1301 AMOTL2_F: AGTGAGCGACAAACAGCAGACG
1302 AMOTL2_R: ATCTCTGCTCCCGTGTTTGGCA
1303 ANAPC13_F: GATTGATGATGCTTGGCG
1304 ANAPC13_R: GTAAGGCTAAGTCTGTCC
1305 CEP63_F: TGGGAAGGACGTACACATGC
1306 CEP63_R: ACATCCAACACTGACTCCTAAGACT
1307 GAPDH_F: GTGGACCTGACCTGCCGTCT
1308 GAPDH_R: GGAGGAGTGGGTGTCGCTGT
1309 PPIA_F: CACCGTGTTCTTCGACATTG
1310 PPIA_R: TTCTGCTGTCTTTGGGACCT

1311
1312 **Data Availability**

1313 Raw and processed sequencing files can be found on the Gene Expression Omnibus super series
1314 ([https://www.ncbi.nlm.nih.gov/geo/](#)).

1315
1316 **Code Availability**

1317 Scripts used for analysis in this manuscript can be found at (<https://github.com/jamrute>).

1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338

1339 **Figure 1.** Multiomic profiling of human coronary arteries. (a) Study design. (b) UMAP embedding
1340 plot of integrated data clustered by RNA. (c) Cell type composition stack plot by sample. (d)
1341 Dotplot of canonical RNA marker genes by cell type; colored by average expression and dot size
1342 indicates percent of cells in the cluster which express the gene. (e) Multiome peaks grouped by
1343 distal, exonic, intronic, and promoter from macs2 peak calls split by cell type. (f) Differentially
1344 accessible marker peaks heatmap by cell type; statistically significant peaks ($FDR < 0.1$ and
1345 $\log_2FC > 0.5$). (g) Number of shared enhancers heatmap from single cell enhancer-gene map;
1346 defined as the number of enhancers where more than 50% of the peak width overlaps with
1347 another cell types/number of total enhancers. (h) Linkage disequilibrium score regression for
1348 cardiometabolic traits by cell type. (i) HiC heatmap connectivity matrix genome-wide and zoomed
1349 in on chromosome 20.

1350
1351 **Figure 2.** Single-cell chromatin accessibility QT discovery. (a) Schematic of sc caQTL discovery
1352 with disease enrichment. (b) Number of pseudobulk caQTL by number of nuclei colored by cell
1353 type at various FDR cut-offs. (c) sc caQTL overlap with GTEx arterial eQTL by cell type. (d) GTEx
1354 (v8 coronary artery) versus caQTL effect size for FDR < 10% list in SMCs. Rasquial effect sizes
1355 are estimated by pi parameter which is centered at 0.5 (less than 0.5 is negative; greater than 0.5
1356 is positive) (e) Box plot of pseudobulk accessibility for peak chr15:78752467-78752967 in SMCs
1357 containing disease SNP rs7182567 by genotype. (f) UMAP embedding plot of SMC cell state
1358 annotations (left) and dotplot of *ADAMTS7* expression by SMC cell state. (g) Box plot of
1359 pseudobulk accessibility for peak chr14:75123953-75124453 in endothelial cells containing
1360 disease SNP rs8017642 by genotype. (h) UMAP embedding plot of endothelial cell state
1361 annotations (left) and dotplot of *NEK* expression by endothelial cell state. (i) Number of
1362 overlapping sc caQTLs with CAD GWAS loci by cell type. (j) CAD GWAS loci and caQTL
1363 Manhattan plot. (k) snATAC-seq tracts with E2G link centered at the rs658956 locus showing
1364 SMC specific enrichment of the enhancer with a E2G link to *HSD52*. (l) Ref (G) to Alt (T) at
1365 rs658956 disrupts a TF binding site for *BACH1* in SMCs. (m) Violin plot showing normalized
1366 expression by genotype for rs658956 and *HSD52* in tibial arteries from GTEx bulk data (p-value
1367 = 1.1×10^{-41}).

1368
1369 **Figure 3.** Dynamic QTL discovery in SMC de-differentiation in atherosclerosis. (a) FMC gene set
1370 score (top 100 genes from Wirka *et al*³⁰ for FMC marker gene) in integrated SMC meta-map. (b)
1371 Nuclei binned into tertiles based on FMC score. (c) Dynamic single cell caQTL modeling strategy
1372 using the PME model for cell-state dependent single nuclei caQTL analysis. Covariate adjustment
1373 and interaction between a continuous FMC cell state score (binned into low, medium and high)
1374 and genotype (see Methods). (d) SMC single cell dynamic caQTLs volcano plot showing -
1375 $\log_{10}(\text{adjusted p-value})$ by interaction coefficient ($\beta_{\text{total}} = \beta_G + \text{FMC score} \times \beta_{\text{FMC}}$); dots are colored
1376 by: black = not significant, blue/red = adjusted p-value < 0.05, where blue and red represent β_{total}
1377 < 0 and $\beta_{\text{total}} > 0$ respectively. (e) Genes linked to dynamic caQTLs which are CAD GWAS loci
1378 using E2G grouped by biological category using ChatGPT. (f) Interaction coefficient of the
1379 rs11838776 caQTL for *COL4A1* in SMCs with FMC score. (g) Box plots show the caQTL effect
1380 for SMCs in the bottom (left), middle (center), and top (right) thirds of FMC scores. (h) rs11838776
1381 eQTL in tibial artery (left) and cultured fibroblasts (right) for *COL4A1* showing no correlation in
1382 tibial arteries but a negative correlation in fibroblasts by genotype. (i) rs11838776 ref (G) to alt (A)
1383 removes a *EGR1/2* binding site. (j) *EGR1 + 2* and *COL4A1 + 2* gene set score dotplot by SMC
1384 cell state showing greatest expression in CMC/FMC cell states.

1385
1386 **Figure 4.** Building a single cell variant to enhancer to gene (scV2E2G) map of coronary artery
1387 disease. (a) Model schematic for single cell V2E2G mapping using a supervised classifier. (b)
1388 Number of CAD GWAS variants mapping to snATAC-peaks and linked genes across cell types
1389 with overlapping caQTLs. (c) 41 SMC scV2E2G and caQTL grouped by biological category using

1390 ChatGPT; genes highlighted in blue are also genes linked to a dynamic caQTL peak. (d) FMC
1391 gene set score (top 100 genes from Wirka *et al*³⁰ for FMC marker gene) in SMC UMAP embedding
1392 plot. (e) Correlation between FMC score and combined gene-set score for 325 genes identified
1393 from V2E2G map in SMCs colored by p-value. (f) Heatmap for SMC V2E2G links positively
1394 correlated with FMC which are also single cell caQTLs by grouped by SMC cell state showing
1395 increasing expression in diseased SMC cell states; blue indicates genes linked to a dynamic
1396 caQTL peak. (g) Gene Set score for the 325 genes identified from V2E2G map in SMCs grouped
1397 by SMC cell state. (h) ENCODE coronary artery H3K27ac and snATAC-seq SMC tract at
1398 rs7246865 SNP with scE2G link to *MYO9B* and *HAUS8*.
1399

1400 **Figure 5.** Building a variant to enhancer to network (scV2E2G) map of coronary artery disease.
1401 (a) Analysis framework for constructing interconnected graph networks using scV2E and genome
1402 wide tissue HiC. (b) Number of distinct network graphs identified from HiC loop connectivity (left)
1403 and filtered for networks which contain a cell specific V2E. (c) Number of loops in all networks
1404 and disease enriched networks (networks where any HiC anchor contains a CAD GWAS variant
1405 inside a coronary artery enhancer) by cell type. (d) SMC V2E2N: number of linked genes in
1406 networks which contain a disease variant (V2E) in SMCs. (e) Disease variant rs4887091 super
1407 chromatin interactome showing SMC snATAC-seq and coronary HiC heatmap with looping
1408 between chr15:71374954-79178639. (f) Genes in the rs4887091 super chromatin interactome;
1409 heatmap showing expression by cell type, caQTL/GTex eQTL/scE2G for rs4887091, and CMD
1410 GWAS related genes. (g) rs4887091 ref (T) to alt leads to increased CTCF binding. (h) CTCF
1411 ChIP-seq in HCASMCs overlaid with SMC snATAC-seq and scE2G at the rs4887091 locus.
1412

1413 **Figure 6.** rs9876658 distally linked to *AMOTL2* in SMCs. (a) snATAC-seq tracts and genome
1414 wide coronary HiC loops showing rs9876658 lies within a non-immune peak linked distally to the
1415 *AMOTL2* transcription start site. (b) Manhattan plot for rs9876658 overlaid on HiChIP-seq and
1416 H3K27ac from primary SMC and endothelial cells showing rs9876658 lies within a SMC specific
1417 enhancer and is linked to the *AMOTL2* TSS in SMCs but not endothelial cells. (c) CRISPRi for
1418 enhancer containing the rs9876658 variant in SMCs and endothelial cells shows that rs9876658
1419 regulates expression of *AMOTL2* and *ANAPC13* in SMCs but not endothelial cells. CRISPRi for
1420 a non-targeting (negative control) and TSS (positive control) shown as reference.
1421

1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440

1441 **Extended Data Figure 1.** Multiome quality control. (a) Experimental workflow to isolate nuclei
1442 from flash frozen coronary arteries with flow cytometry to sort for 7-AAD+ nuclei with final
1443 morphology. (b) Scrublet score after doublet removal. (c) Quality control metrics for RNA (number
1444 of UMI counts, number of genes, mitochondrial RNA ratio, and ribosomal RNA ratio per nucleus)
1445 and ATAC (TSS enrichment and number of fragments per nucleus) grouped by sample. (d)
1446 snATAC-seq fragment size distribution per sample.

1447
1448 **Extended Data Figure 2.** Multimodal clustering of Multiome data. (a) Gene set scores for marker
1449 genes (adjusted p-value < 0.05 and log₂FC > 0.58) in each cell type. (b) DotPlot of marker gene
1450 set score form (a) grouped by cell type. (c) UMAP embedding and clustering using ATAC LSI,
1451 RNA LSI, RNA/ATAC integrated with harmony batch correction, and RNA/ATAC integrated with
1452 harmony batch correction with annotations from RNA based clustering. Confusion matrix for
1453 cluster annotations in (d) RNA and ATAC clustering and (e) RNA/ATAC integrated with harmony
1454 batch correction and RNA. (f) Motif enrichment using chromVar by major cell types using
1455 differentially accessible peaks (FDR < 0.1 and log₂FC > 0.5).

1456
1457 **Extended Data Figure 3.** HiC quality control. Number of significant (a) loops called b FitHiC and
1458 (b) TADs identified at different resolutions by sample. (c) Loop and (d) TAD size distributions by
1459 sample. (e) Number of loops called using FitHiC at various anchor resolutions. (f) Average loop
1460 size at various anchor resolutions.

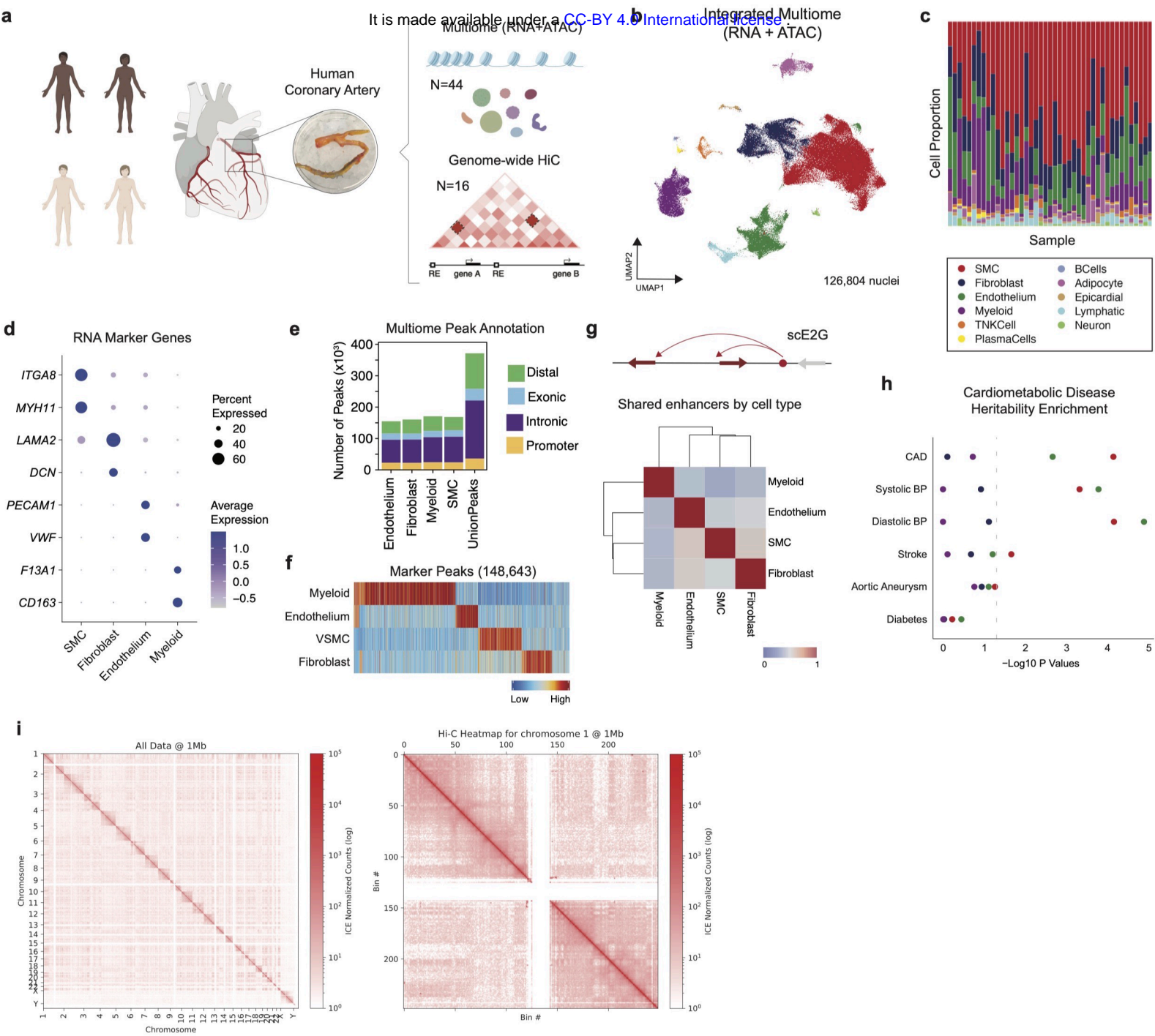
1461
1462 **Extended Data Figure 4.** Genotyping and pseudobulk caQTL discovery. (a) PC1 and 2 for
1463 genotyping data (post imputation from the TOPMED panel and filtering for variants with imputation
1464 R² < 0.3 and MAF > 0.05) panel for each sample. (b) Local ancestry inference (YRI (n=186) and
1465 CEU (n=183) from the 1000 Genome Project¹⁰² were used as AFR / EUR reference populations,
1466 respectively.) genetic proportion stack plot by sample. (c) Correlation heatmap for ancestry PCs
1467 and ATAC count PCs. (d) Multi-trait pseudobulk Manhattan plot for caQTL colored by cell type
1468 genome wide. (e) Distance from caQTL to the regulated peak colored by cell type.

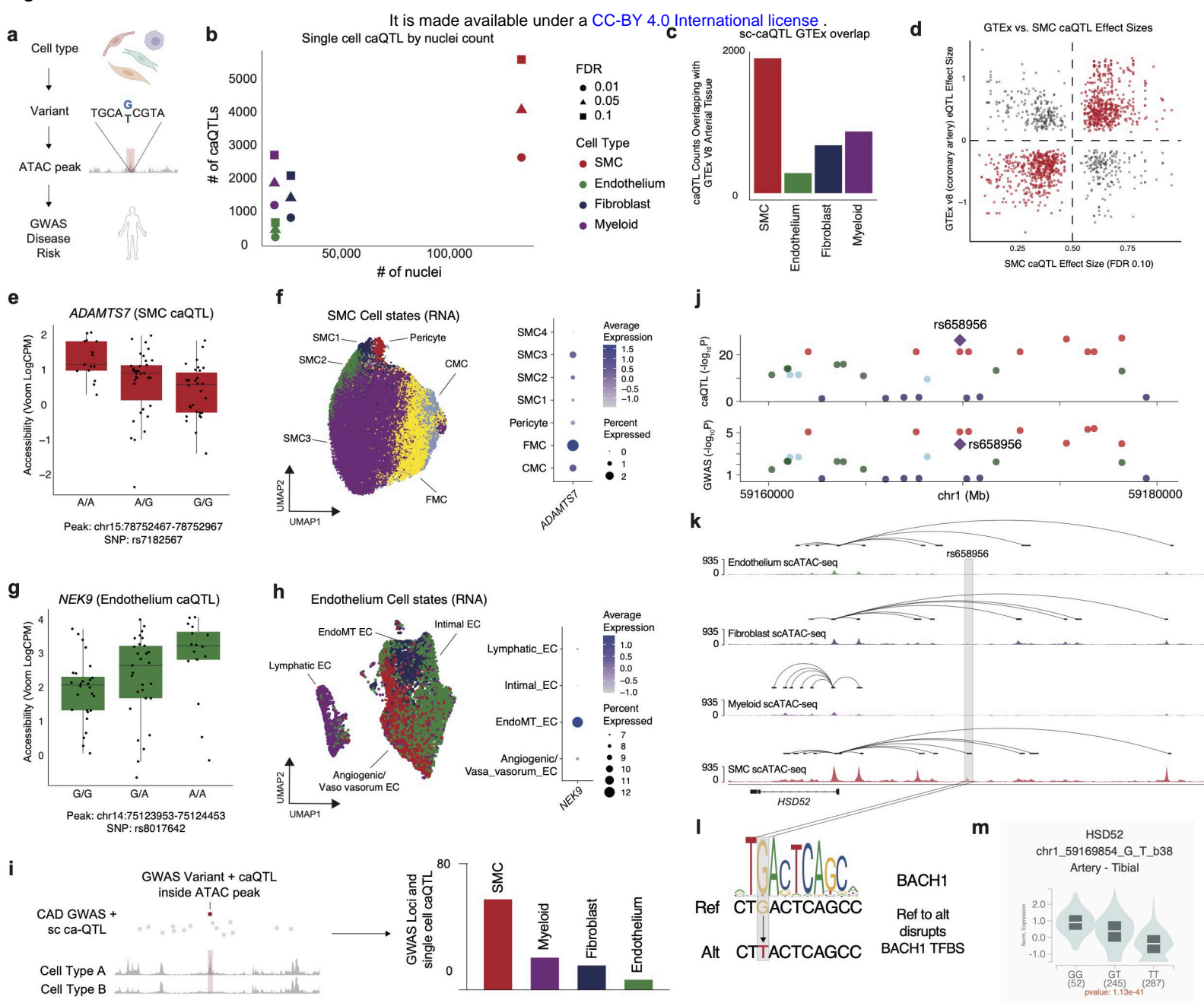
1469
1470 **Extended Data Figure 5.** Cell type and state regulation of molecular traits. Pseudobulk
1471 accessibility by genotype plots for (a) *ADAMTS7*, (b) *NEK9*, (c) *SMAD7*, and (d) *FDX* are cell type
1472 specific caQTLs for SMCs, endothelium, fibroblasts and myeloid cells respectively. (e) DotPlot of
1473 canonical marker genes for endothelium cell biology grouped by mapped cell state annotation. (f)
1474 Pathway enrichment for statistically significant genes upregulated in EndoMT (adjusted p-value <
1475 0.05 and log₂FC > 0.58). (g) Endothelium acting loci from V2G2P in Schnitzler *et al*⁴⁷ gene set
1476 score grouped by EC state. (h) Program 8 gene set score from Schnitzler *et al*⁴⁷ in UMAP
1477 embedding plot (left) and DotPlot grouped by EC state (right).

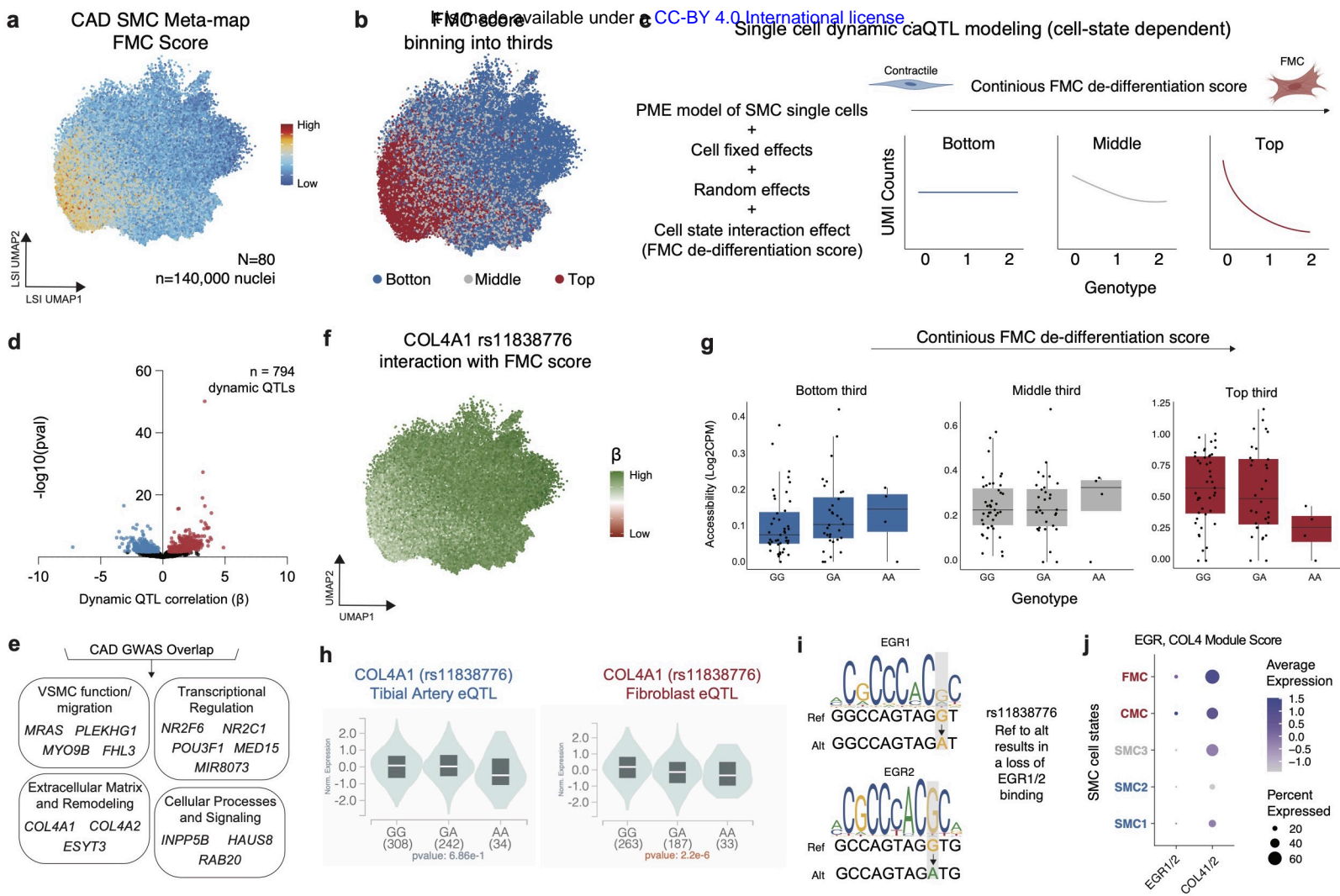
1478
1479 **Extended Data Figure 6.** Dynamic caQTL modeling. (a) SMCs integrated from Turner *et al*³⁴ and
1480 Multiome snATAC-seq to construct a meta-map for dynamic single cell QTL modeling grouped
1481 by dataset. (b) FMC marker genes gene activity score split by dataset. (c) Framework for single
1482 cell dynamic caQTL modeling with a Poisson Mixed Effect model showing an FMC state
1483 dependent effect of genotype on chromatin accessibility. (d) Gene set score for dynamic caQTLs
1484 overlapped with CAD GWAS variants linked to genes with scE2G grouped by SMC state. (e)
1485 *HSD52* betas for dynamic caQTL at rs658956 in UMAP embedding (left) and pseudobulk box
1486 plots by genotype and FMC score tertile (right).

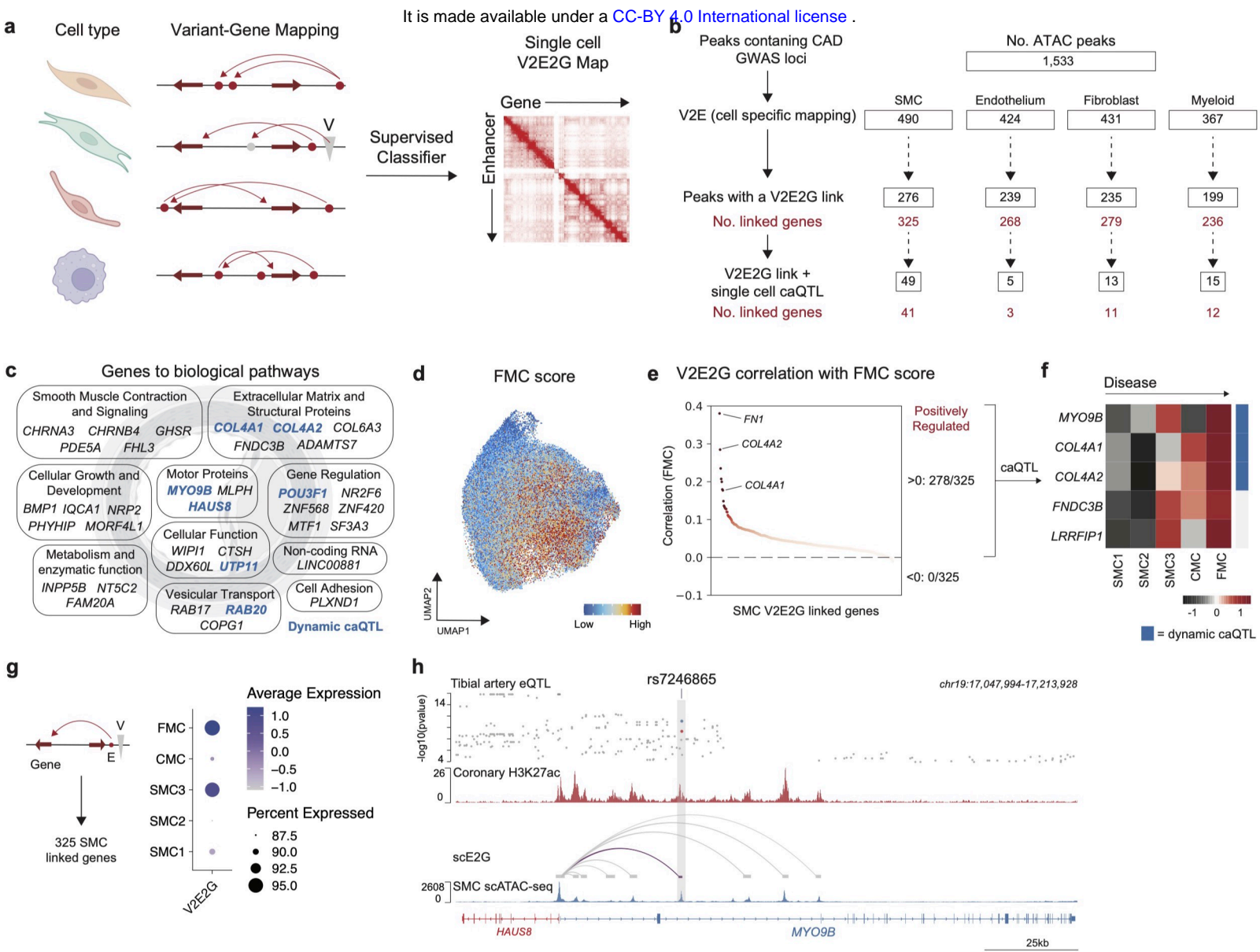
1487
1488 **Extended Data Figure 7.** Each point represents enhancer-gene predictions by scE2G^{Multiome} in
1489 one cell cluster (N = 11). All properties were computed on binarized predictions omitting promoter
1490 elements. (a) Average number of unique ATAC fragments per cell by number of cells per
1491 cluster. (b) Number of enhancer-gene links by number of unique ATAC fragments per cluster. T

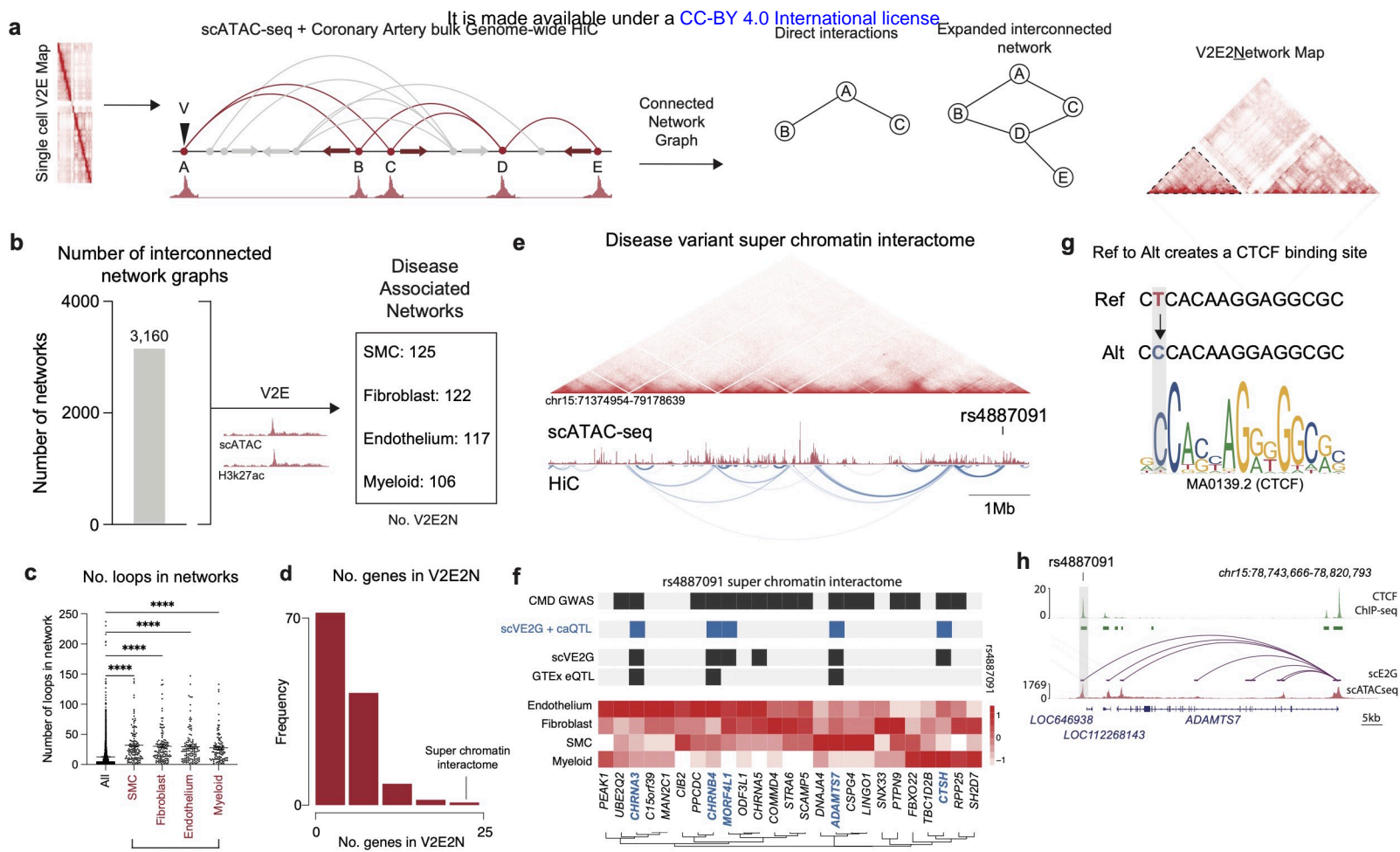
1492 (c) Mean enhancer width by number of unique enhancer elements regulating at least one distal
1493 gene. (d) Mean distance between enhancer element and target gene transcription start site. (e)
1494 Mean number of enhancers per gene by mean number of genes per enhancer. (f) Number of
1495 genes with at least one non-promoter enhancer by number of unique ATAC fragments.





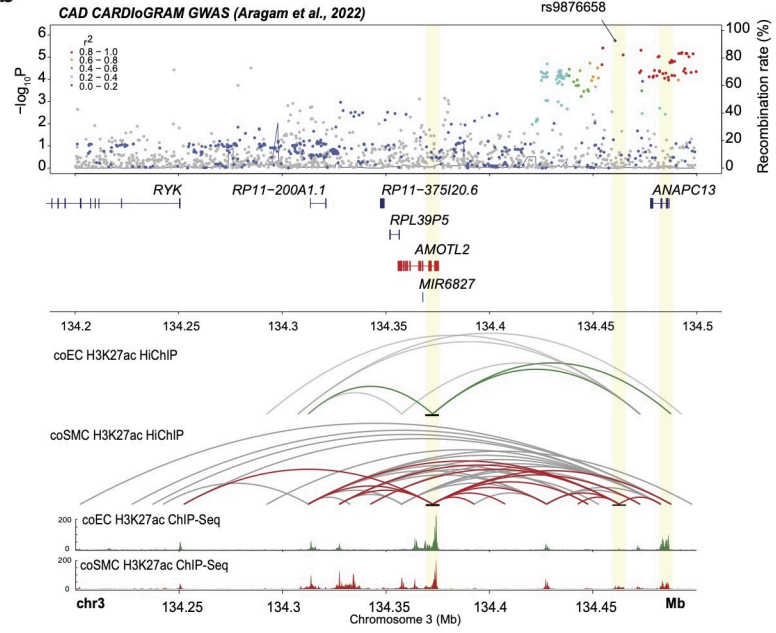
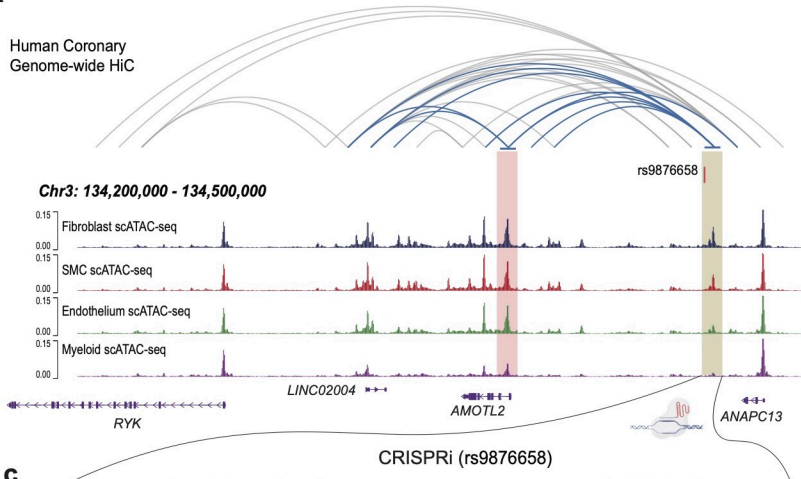




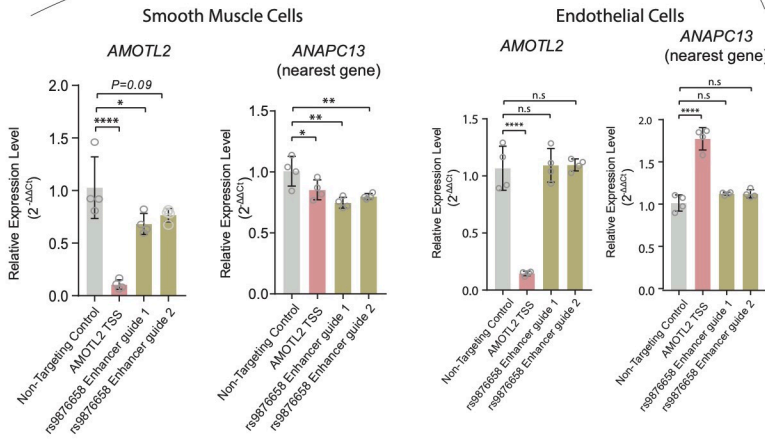


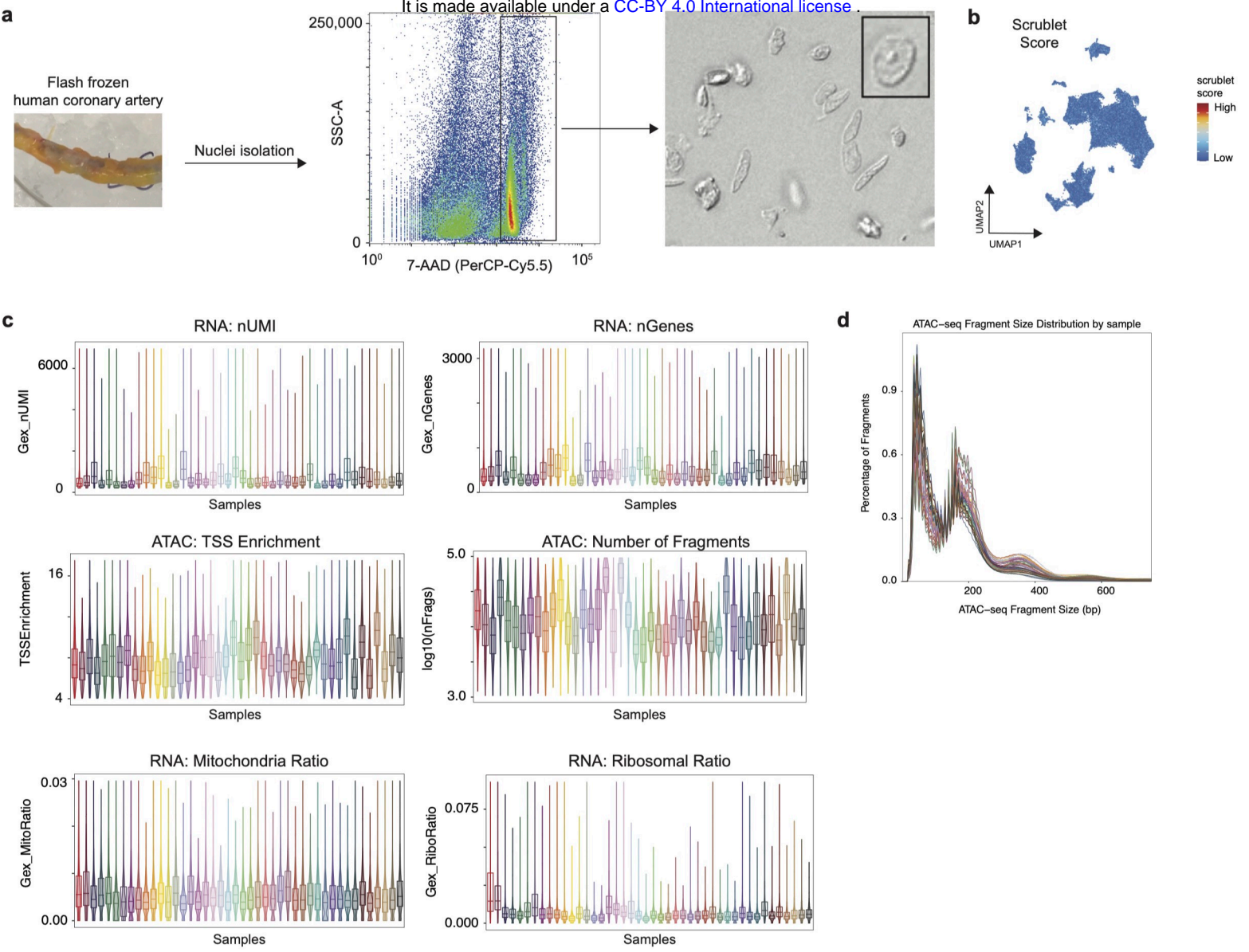
a

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

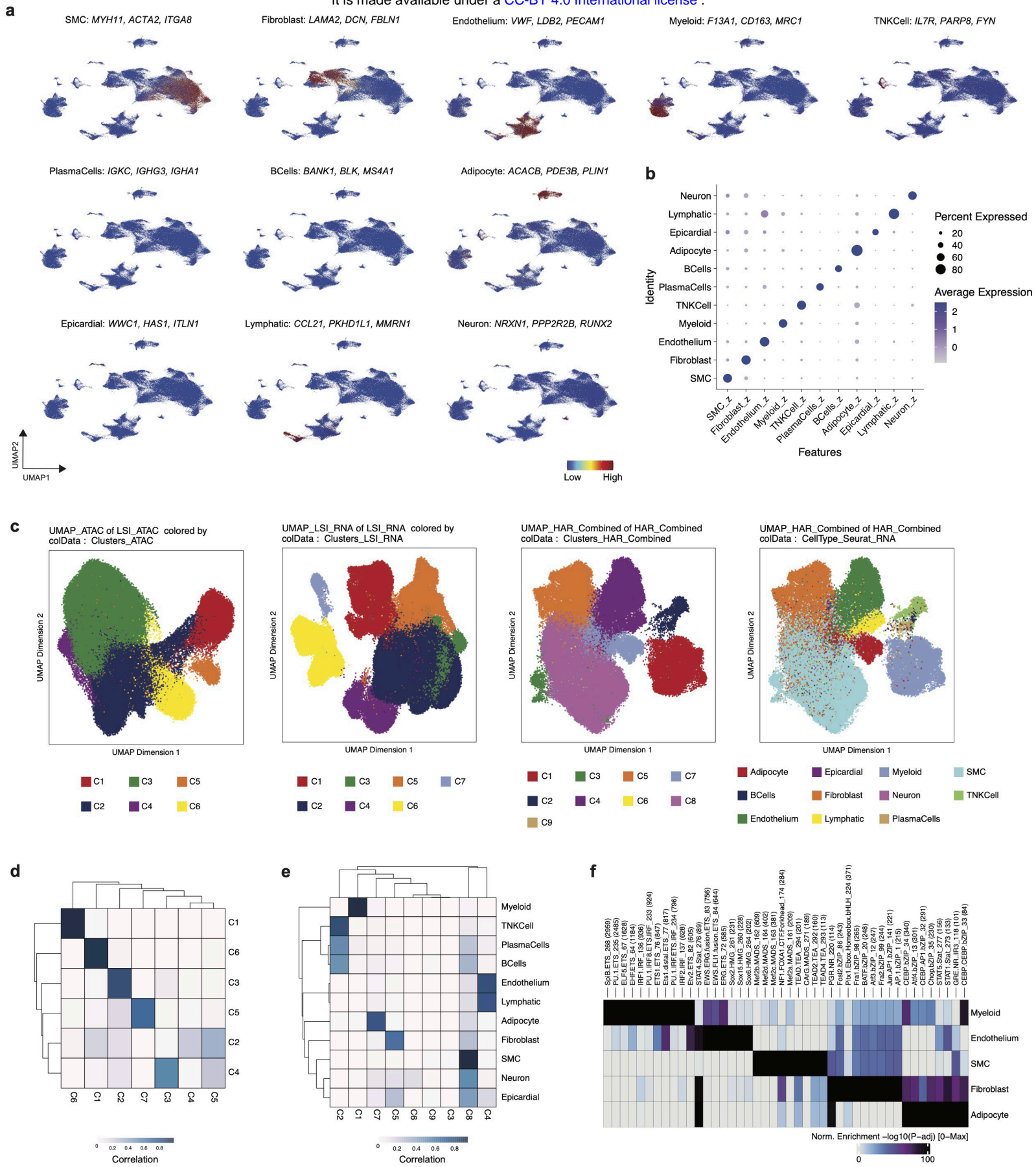


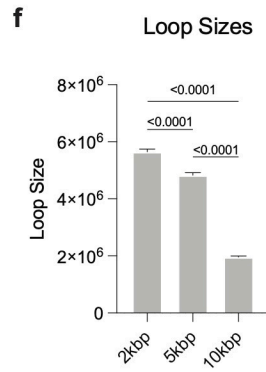
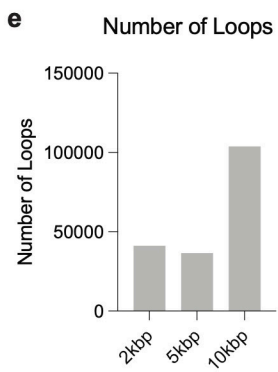
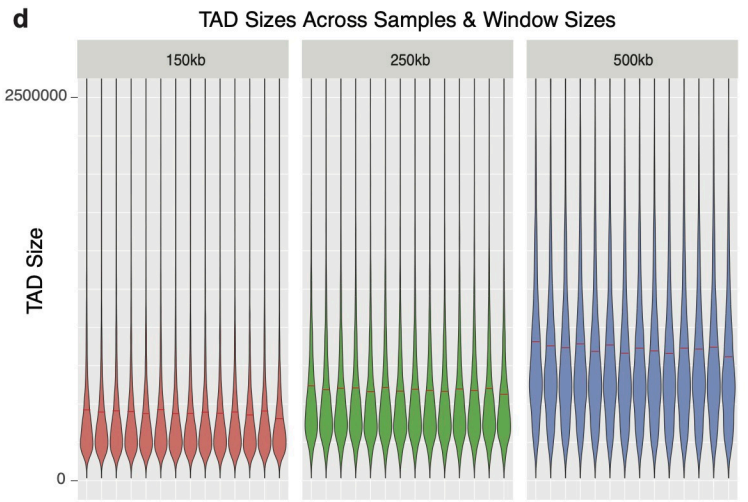
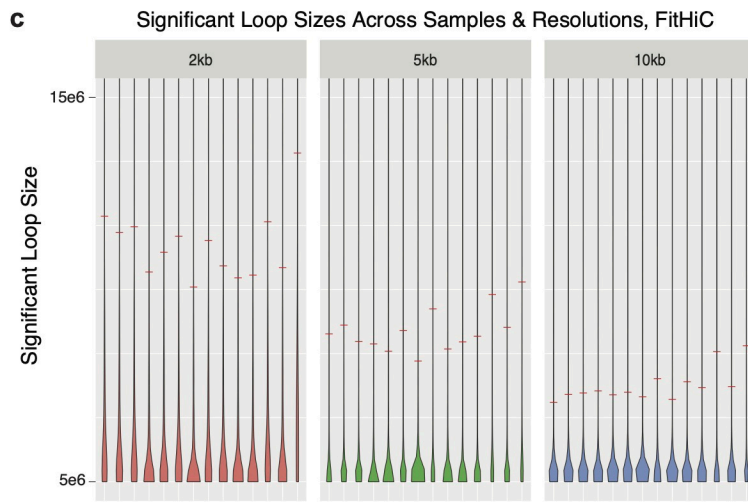
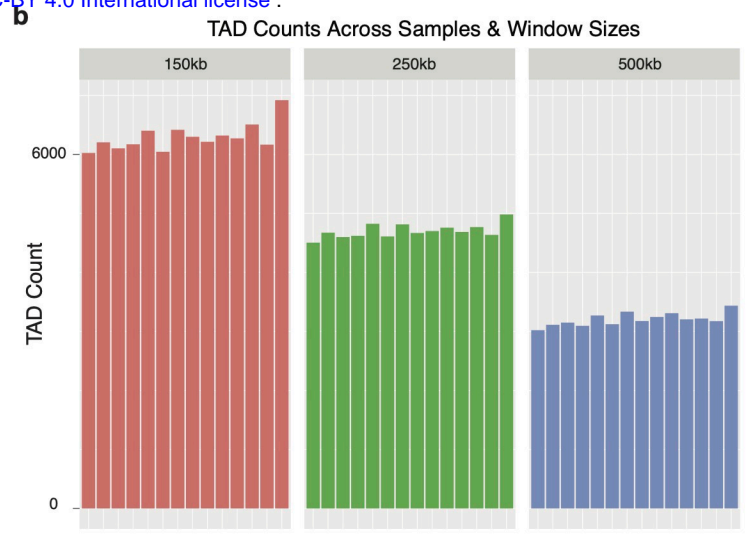
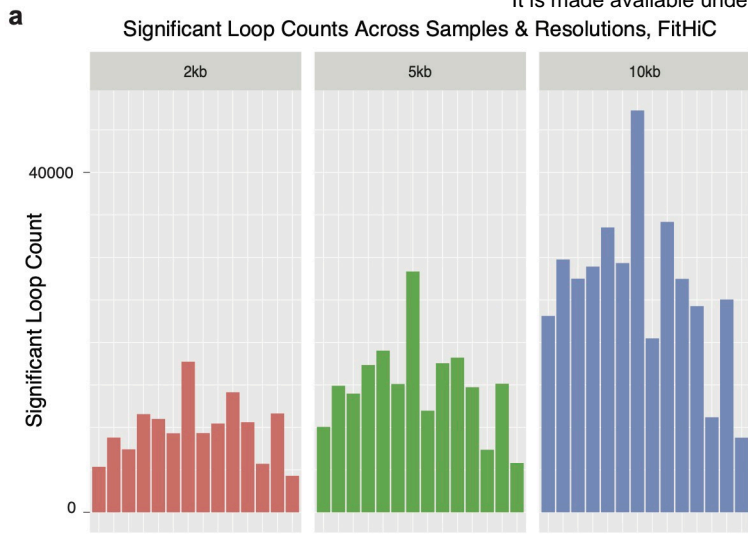
c





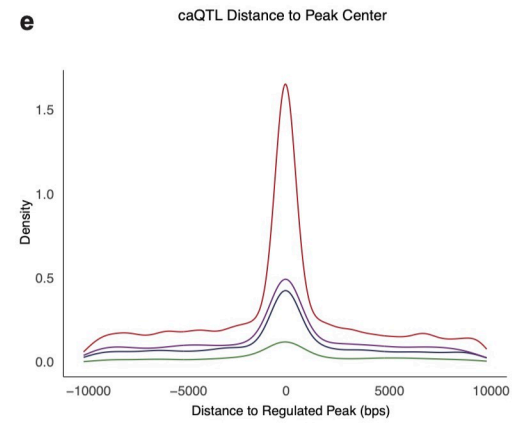
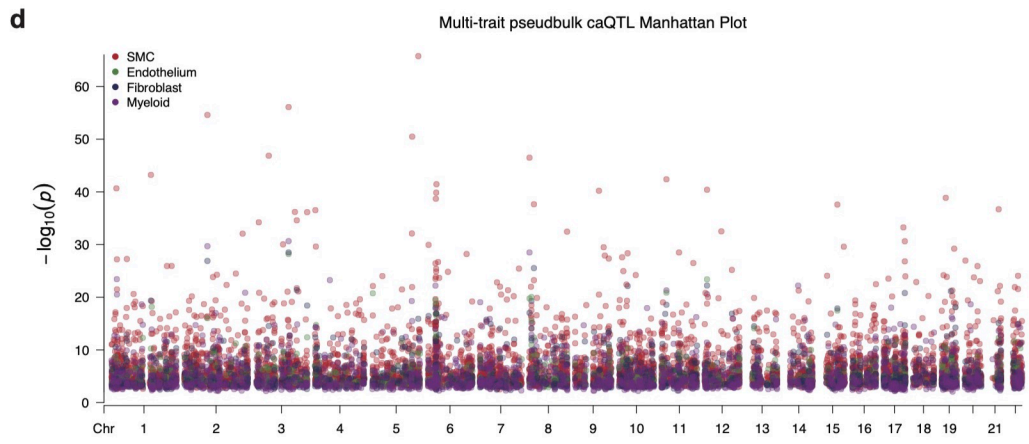
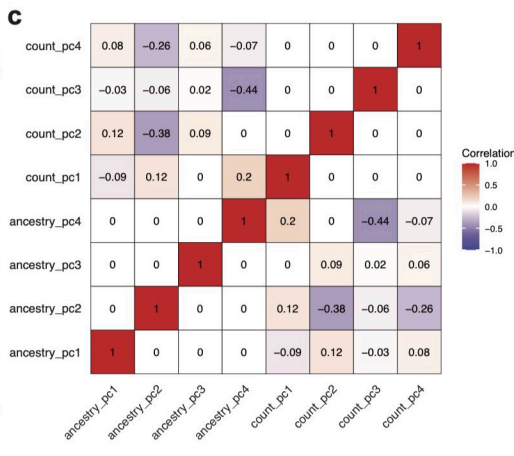
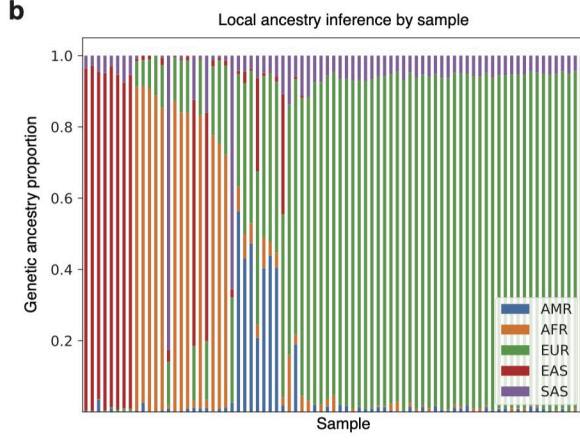
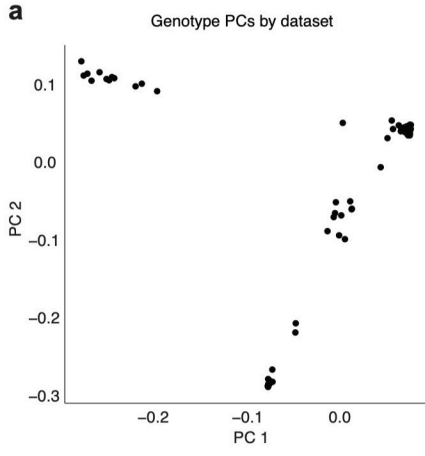
It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



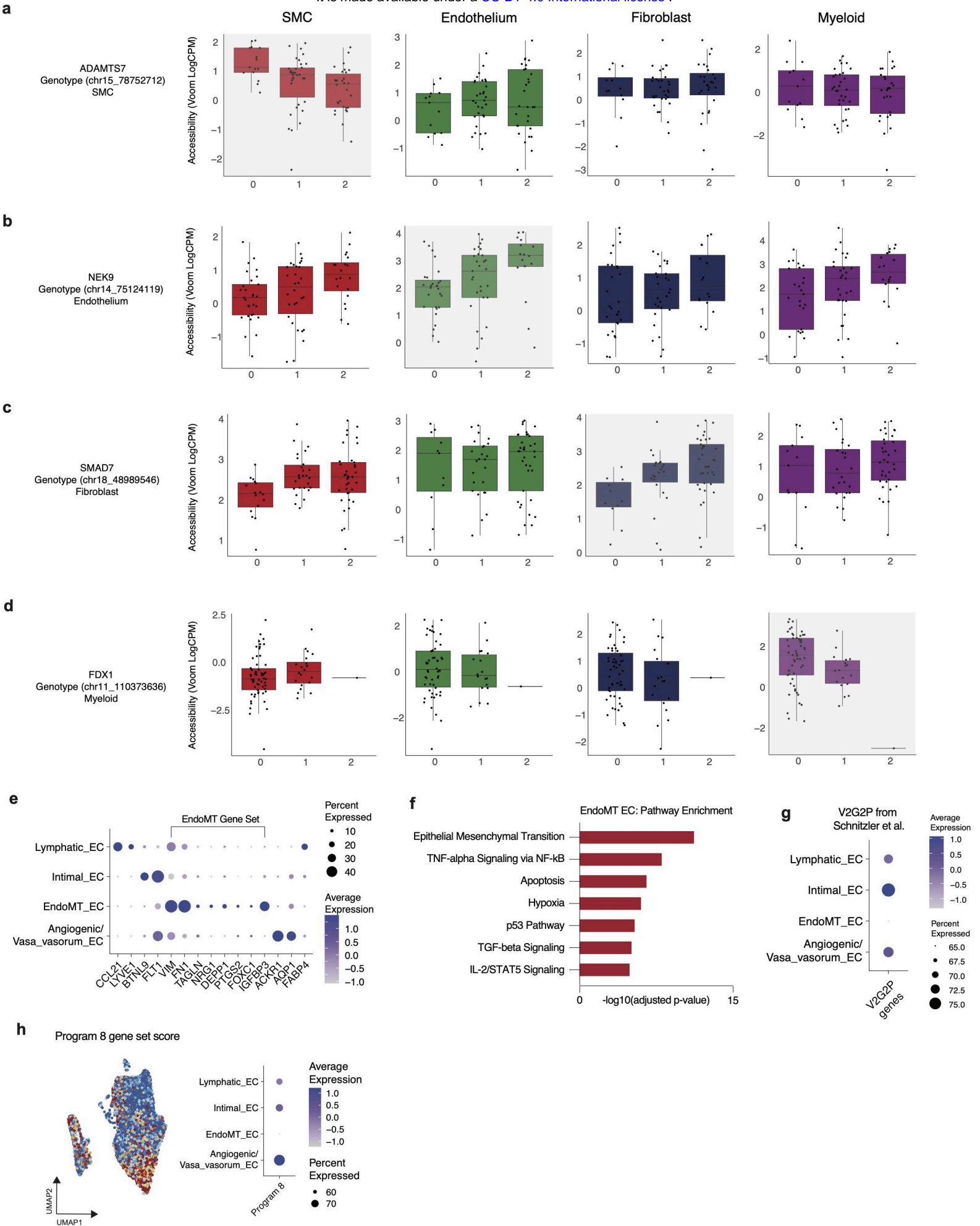


Extended Data Figure 4

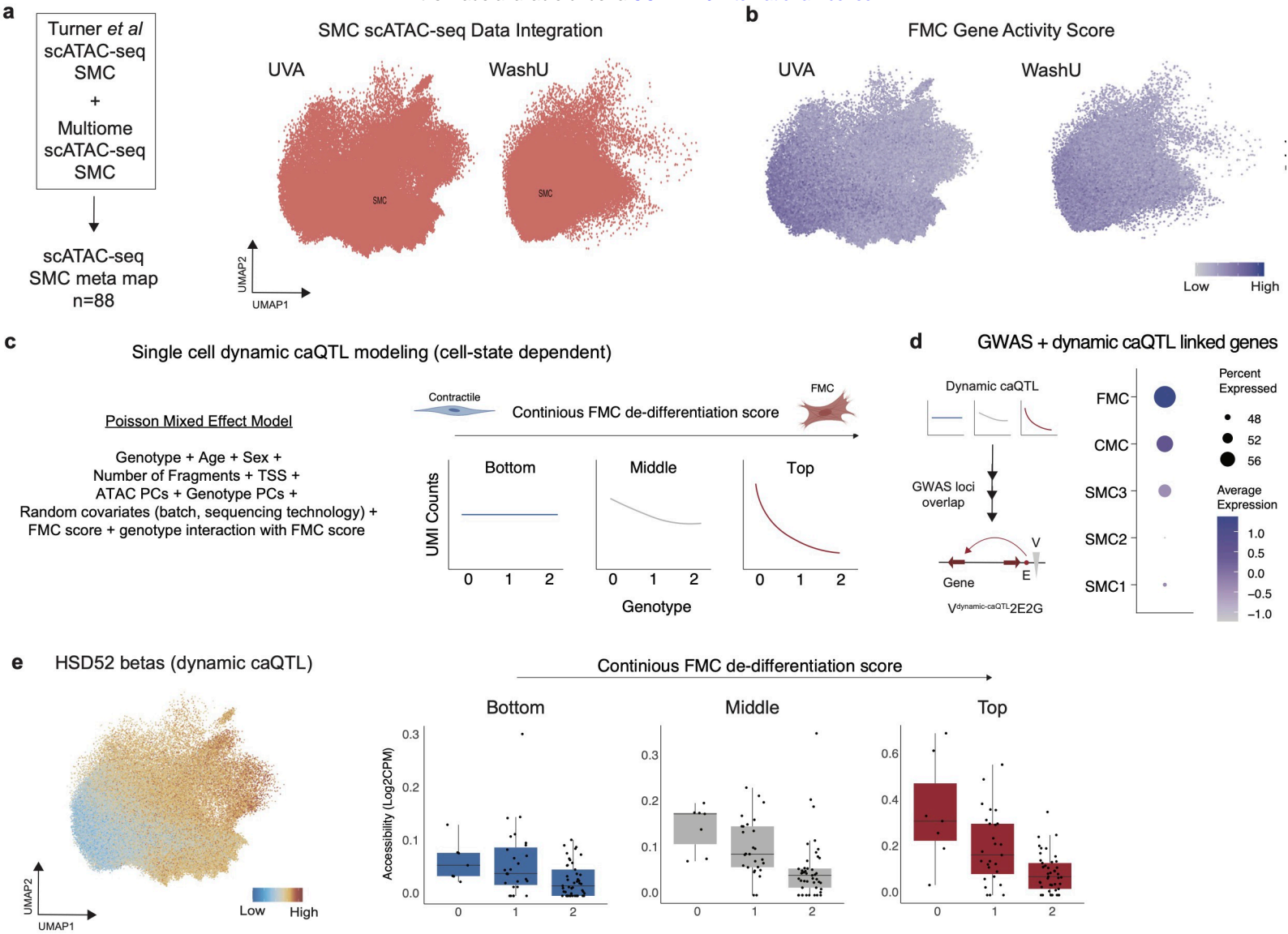
It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



Extended Data Figure 7

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

