Title

eQTL mapping in fetal-like pancreatic progenitor cells reveals early developmental insights into diabetes risk

Authors

Jennifer P. Nguyen[1,2], Timothy D. Arthur[2,3], Kyohei Fujita[6], Bianca M. Salgado[4], Margaret K.R. Donovan[1,2], iPSCORE Consortium[+], Hiroko Matsui[4], Ji Hyun Kim[5], Agnieszka D'Antonio-Chronowska[6], Matteo D'Antonio[2,4,6], Kelly A. Frazer[4,6,*]

Affiliations:

[1] Bioinformatics and Systems Biology Graduate Program, University of California, San Diego, La Jolla, CA, 92093, USA

[2] Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA, 92093, USA

[3] Biomedical Sciences Graduate Program, University of California, San Diego, La Jolla, CA 92093, USA

[4] Institute of Genomic Medicine, University of California San Diego, 9500 Gilman Dr, La Jolla, CA, 92093, USA

[5] Department of Pediatrics, Dongguk University Ilsan Hospital, Goyang, South Korea

[6] Department of Pediatrics, University of California, San Diego, La Jolla, CA, 92093, USA

+ A full list of Consortium members and their affiliations appears at the end of the manuscript

[*] Correspondence to: Kelly A. Frazer (kafrazer@health.ucsd.edu).

Abstract

The impact of genetic regulatory variation active in early pancreatic development on adult pancreatic disease and traits is not well understood. Here, we generate a panel of 107 fetal-like iPSC-derived pancreatic progenitor cells (iPSC-PPCs) from whole genome-sequenced individuals and identify 4,065 genes and 4,016 isoforms whose expression and/or alternative splicing are affected by regulatory variation. We integrate eQTLs identified in adult islets and whole pancreas samples, which reveal 1,805 eQTL associations that are unique to the fetal-like iPSC-PPCs and 1,043 eQTLs that exhibit regulatory plasticity across the fetal-like and adult pancreas tissues. Colocalization with GWAS risk loci for pancreatic diseases and traits show that some putative causal regulatory variants are active only in the fetal-like iPSC-PPCs and likely influence disease by modulating expression of disease-associated genes in early development, while others with regulatory plasticity likely exert their effects in both the fetal and adult pancreas by modulating expression of different disease genes in the two developmental stages.

Introduction

Genome-wide association studies (GWAS) have identified hundreds of genetic variants associated with adult pancreatic disease risk and phenotypes [1–4]. However, the majority of these associations map predominantly to non-coding regions of the genome, thereby hindering functional insights into disease processes [5–7]. Previous large-scale expression quantitative trait loci (eQTL) studies have made significant advancements toward understanding how genetic variation affects gene expression in various tissues and cell types, as well as their contribution to human traits and diseases [8–11]. However, these analyses were conducted in adult tissues and therefore the effects of regulatory variation on gene expression under fetal conditions remain unclear. Moreover, the integration of adult and fetal eQTL datasets would enable the investigation of regulatory plasticity of genetic variants, which refers to changes in variant function under different spatiotemporal contexts [9,12,13]. Understanding how genetic variation affects gene expression during early pancreas development, and how their function changes in adulthood, can expand our understanding of the biological mechanisms underlying adult pancreatic disease and GWAS complex trait loci.

Many lines of evidence from clinical and genomic studies indicate an important role of pancreas development in the health and onset of childhood and adult pancreatic diseases [14–17]. For example, mutations in genes critical to pancreatic development, such as *PDX1*, *HNF4A*, and *HNF1A*, are associated with childhood-onset diabetes [18–20]. Furthermore, type 2 diabetes (T2D)-risk variants map to transcription factors (TFs) that are crucial to pancreas development, including NEUROG3 and HNF1A, and are enriched in accessible pancreatic progenitor-specific enhancers [4,14]. To address the limited availability of fetal pancreatic tissues, protocols have been developed to efficiently guide the differentiation of human induced pluripotent stem cells (iPSCs) into pancreatic progenitor cells (iPSC-PPCs). This approach serves as a model system to study human pancreas development [21–28]. iPSC-PPCs demonstrate expression of key transcription factors associated with early pancreas development, including *PDX1*, *NKX6-1*, and *SOX9*, all pivotal for pancreas lineage specification and differentiation [22,29–33]. Additionally, iPSC-PPCs express developmental signaling pathway, including Notch, WNT, and Hedgehog, that are critical in pancreas development [23,25,28,34]. While iPSC-PPCs have provided extensive insights into pancreas developmental biology, they have not yet been utilized to examine the impact of genetic variation on gene expression in the fetal-like pancreas.

In this work, we conduct a large-scale eQTL analysis on 107 iPSC-PPC samples to map genetic loci associated with gene expression and isoform usage during early pancreas development. We integrate eQTLs from adult pancreatic tissues and identify eQTL loci that display temporal specificity in early pancreas development, as well as eQTL loci that are shared with adult but display regulatory plasticity. Annotation of GWAS risk loci using our spatiotemporally informed eQTL resource reveal causal regulatory variants with developmental-unique effects associated with complex pancreatic traits and disease.

Results

Overview of study

The goal of our study is to understand how regulatory variation active in early pancreas development influences adult pancreatic disease risk and phenotypes (Figure 1A). We differentiated 106 iPSC lines from the iPSCORE resource [35] derived from 106 whole-genome sequenced individuals to generate 107 iPSC-PPC samples (one iPSC line was differentiated twice) (Supplementary Figure 1, Supplementary Data 1, Supplementary Data 2). We characterized the fetal-like pancreatic transcriptome as well as the cellular composition using single-cell RNA-seq (scRNA-seq) of eight iPSC-PPC samples. Then, we conducted an eQTL analysis on bulk RNA-seq of all 107 samples to identify regulatory variants associated with fetal-like gene expression and isoform usage. To better understand the spatiotemporal context of genetic variants, we integrated eQTLs previously discovered in adult pancreatic islets and whole pancreas samples using colocalization and network analysis. Finally, using our eQTL resource of pancreas tissues (i.e., fetal-like iPSC-derived PPCs, adult islets, adult whole pancreas), we performed GWAS colocalization and fine-mapping to link developmental regulatory mechanisms and identify putative causal variants underlying pancreatic traits and disease associations.

Large-scale differentiation of fetal-like pancreatic progenitor cells

We derived 107 iPSC-PPC samples using iPSC lines reprogrammed from 106 individuals. Differentiation efficiency was assessed using flow cytometry analysis on PDX1 and NKX6-1, which are two markers routinely assayed for early pancreatic progenitor formation. PDX1 marks the specification of cells towards the pancreas lineage (referred to here as "early PPC"; PDX1^{+}/NKX6-1^{-}), while subsequent NKX6-1 expression marks the differentiation and maturation of pancreatic progenitor cells (referred to here as "late PPC"; PDX1^{+}/NKX6-1^{+}) [36]. We observed an 18.1% median percentage of early PPCs (PDX1^{+}/NKX6-1^{-}) across the 107 samples while the median percentage of late PPCs (PDX1^{+}/NKX6-1^{+}) was 74% (range: 9.4%-93.1%) (Figure 1B, Supplementary Figure 2, Supplementary Data 2). We further found that the median percentage of cells that expressed PDX1+ was more than 90%, confirming that the majority of cells have specified towards the pancreas lineage and that the differentiation procedure was highly efficient (Figure 1B, Supplementary Figure 2, Supplementary Data 2). Consistent with flow cytometry analysis, scRNA-seq of ten derived iPSC-PPCs confirmed the presence of both early and late PPCs and that the majority of the cells were late PPCs (Supplementary Figure 3-8; Supplementary Data 2-5; See Methods and Supplemental Note 1). Altogether, these results show that the majority of the cells in iPSC-PPCs were differentiated into late PPCs while a smaller fraction represented a primitive PPC state.

To examine the similarities between iPSC-PPC and adult pancreatic transcriptomes, we generated bulk RNA-seq for all 107 iPSC-PPC samples and inferred the pseudotime on each sample, along with 213 iPSCs [35,37], 87 pancreatic islets [38], and 176 whole pancreatic tissues [39]. Pseudotime analysis and comparative expression analysis of early developmental genes showed that the iPSC-PPC samples corresponded to an early timepoint of pancreas development (Supplementary Figure 9, Supplementary Data 6).

These analyses, combined with the results of previous studies [22,25,27], show that the 107 derived iPSC-PPCs represent a fetal-like state of pancreatic tissues.

Identification and characterization of gene and isoform eQTLs in fetal-like iPSC-PPCs

To characterize the effects of genetic variation on the fetal-like iPSC-PPC transcriptome, we performed an eQTL analysis to map the genetic associations with fetal-like gene expression (e_gQTL) and relative isoform usage (e_iQTL). Considering only autosomal chromosomes, we analyzed a total of 16,464 genes and 29,871 isoforms (corresponding to 9,624 autosomal genes) that were expressed in the fetal-like iPSC-PPCs. We identified 4,065 (24.7%) eGenes and 4,016 (13.0%) eIsoforms with an e_gQTL or e_iQTL, respectively (FDR < 0.01, Figure 1C-D, Supplementary Data 7). To identify additional independent eQTL signals (i.e., conditional eQTLs) [40], we performed a stepwise regression analysis for each eGene and eIsoform. This analysis yielded 368 e_gQTLs that mapped to 338 eGenes and 216 e_iQTLs that mapped to 198 eIsoforms, totaling to 4,433 independent e_gQTL associations and 4,232 independent e_iQTL associations (Figure 1C-D, Supplementary Data 7). We next predicted candidate causal variants underlying each eQTL (e_gQTL and e_iQTL) association using *coloc* genetic fine-mapping [41] (Supplementary Data 8) and tested their enrichments in transcribed regions and regulatory elements. We observed an enrichment of e_gQTLs in intergenic and promoter regions while e_iQTLs were enriched in splice sites and RNA-binding protein binding sites (Figure 1E). We additionally estimated the transcription factor (TF) binding score for each variant using the Genetic Variants Allelic TF Binding Database [42] and found that, at increasing posterior probability (PP, probability that the variant is causal for the association) thresholds, the candidate causal variants underlying e_gQTLs were more likely to affect TF binding compared to those underlying e_iQTLs (Figure 1F, Supplementary Data 9, Supplementary Data 10). These results corroborate similar findings from previous studies [10,12,43], showing that the genetic variants underlying e_gQTLs and e_iQTLs primarily affect gene regulation and coding regions or alternative splicing, respectively.

To further characterize the function of genetic variants associated with the fetal-like iPSC-PPC transcriptome, we examined the distributions of e_gQTLs and e_iQTLs per gene. Of the 5,619 genes whose phenotype was affected by genetic variation, 1,008 were impacted through both gene expression and isoform usage (i.e., had both e_gQTL and e_iQTLs, 17.9%) while 3,057 were impacted through only gene expression (i.e., had only e_gQTLs, 54.4%) and 1,554 through only isoform usage (i.e., had only e_iQTLs, 27.7%, Figure 1G, Supplementary Data 7). For the 1,008 genes with both e_gQTL and e_iQTLs, we performed colocalization with *coloc.abf* [41] to examine whether the same or different

121 genetic variants underpinned their associations. *Coloc.abf*[41] employs a Bayesian approach to estimate the PP that each
122 of the five colocalization models best explains the association between two genetic signals: H0) no associations
123 detected in either signal; H1) association detected in only signal 1; H2) association detected in only signal 2; H3)
124 associations detected in both signals but driven by different causal variants, and H4) associations detected in both
125 signals and driven by the same causal variant. We identified 410 (40.7%) genes that had at least one H4 (PP.H4,
126 posterior probability for H4 ≥ 80%) or H3 (PP.H3, posterior probability for H3 ≥ 80%) association between their
127 e_gQTL and e_iQTLs, of which the majority (333, 81.2%) had only overlapping signals (all H4), 38 (9.3%) had only non-
128 overlapping signals (all H3), and 39 (9.5%) had both overlapping and non-overlapping e_iQTLs (both H3 and H4; an
129 e_gQTL can colocalize with an e_iQTL corresponding to one isoform but not with another e_iQTL corresponding to a
130 second isoform) (Figure 1G, Supplementary Data 11). The remaining 598 genes had PP.H3 < 80% and PP.H4 < 80%
131 due to insufficient power (Figure 1G). These findings show that 19.5% (1,008 / 5,169) of genes had both e_gQTLs and
132 e_iQTLs and that their effects were commonly driven by the same causal variants (81.2%) while only a small fraction
133 was driven by different causal variants (9.3%). Enrichment analysis of overlapping e_gQTL and e_iQTLs showed that
134 these variants were enriched for stop codons as well as mechanisms affecting both gene expression and alternative
135 splicing (Supplementary Figure 11). Overall, our results show that the majority of genes had either only e_gQTLs or
136 e_iQTLs, indicating that the functional mechanisms underlying these associations are likely independent, where genetic
137 variants affecting alternative splicing do not affect the overall expression of the gene, and vice versa.

138 eQTL landscapes of fetal-like iPSC-PPC and adult pancreatic islets

139 Studies aimed at identifying and characterizing eGenes have been conducted in both adult human islets and whole
140 pancreatic tissues [8,10,11,38,44]; however, islet tissues have been more thoroughly studied because of their role in diabetes.
141 Therefore, we focused on understanding the similarities and differences between eGenes in the fetal-like iPSC-PPCs
142 and adult human islets.

143 We obtained eQTL summary statistics and intersected the 4,211 autosomal eGenes identified in 420 adult islet samples
144 [11] with the 4,065 eGenes in fetal-like iPSC-PPC. We found that only 1,501 (36.9% of 4,065) eGenes overlapped
145 between the fetal-like iPSC-PPC and adult islet tissues (Figure 2A). To determine whether the small overlap was due
146 to gene expression differences, we calculated how many of the eGenes were expressed in both the fetal-like iPSC-PPC
147 and adult islets. Of the 4,065 fetal-like iPSC-PPC eGenes, 88.7% (3,605) were also expressed in the adult islets;
148 likewise, of the 4,211 adult islet eGenes, 78.4% (3,301) were also expressed in the fetal-like iPSC-PPCs (Figure 2B).
149 These results suggest that most fetal-like iPSC-PPC eGenes were expressed but not associated with genetic variation
150 in the adult islet samples, and vice versa.

151 For eGenes that were present in both the fetal-like iPSC-PPC and adult islet samples, we next asked whether their
152 expressions were controlled by the same genetic variants. We performed colocalization between e_gQTLs for the 1,501

153 shared eGenes in the fetal-like iPSC-PPC and adult islets and found that 795 (52.3%) displayed strong evidence for
154 either H3 or H4 association (PP.H3 or PP.H4 ≥ 80%) (Supplementary Data 11). Of the 795 with an association, 701
155 (88.2%) had overlapping e_gQTL signals (PP.H4 ≥ 80%) while 94 (11.8%) had non-overlapping e_gQTL signals (PP.H3
156 ≥ 80%) (Figure 2C, Supplementary Data 11). These results indicate that most shared eGenes were associated with the
157 same genetic variants controlling their gene expressions in both fetal-like iPSC-PPC and adult islet tissues, while a
158 subset had non-overlapping genetic variants. For example, we identified *SNX29* as an eGene in both fetal-like iPSC-
159 PPC and adult islets but observed that its expression was associated with distinct eQTL signals approximately 520 kb
160 apart (Figure 2D). SNX29 is involved in various signaling pathways [45], including TGF-β, ErbB, and WNT signaling
161 pathways, and is predicted to be a causal gene for body-mass index (BMI) and T2D [46].

162 Taken together, our results show that a minor proportion of fetal-like iPSC-PPC eGenes (1,501, 37% of 4,065) was
163 shared with adult islets, whereas the majority (2,564 = 4,065–1,501, 63%) were fetal development-specific; and, while
164 most shared eGenes were associated with the same regulatory variants, ~12% were mediated by different eQTLs. These
165 findings indicate that regulatory variants tend to act in a developmental-specific manner, potentially by affecting the
166 binding of key regulatory TFs specific to fetal or adult pancreatic stages.

167 Developmental stage-unique and shared e_gQTLs

168 Above, we described eGenes that were unique to either fetal-like iPSC-PPCs or adult islets, or shared between both.
169 Here, we sought to identify eQTLs (i.e., regulatory variants) that specifically affect gene expression during the pancreas
170 development stage, in the adult stage, or both stages. Because fetal-like iPSC-PPCs give rise to both endocrine and
171 exocrine cell fates, we included eQTLs from both adult islets [11] and whole pancreas [39] tissues in our analyses. Due to
172 the many different types of eQTLs used in this study, we refer to all eQTLs as a collective unit as "eQTLs", eQTLs
173 that were associated with gene expression as "e_gQTLs" (as defined above), and eQTLs associated with changes in
174 alternative splicing (e_iQTLs, exon eQTLs, and sQTLs) as "e_{AS}QTLs". For simple interpretations, we only describe the
175 results for the analyses conducted on the e_gQTLs below, however, we identified unique and shared iPSC-PPC e_{AS}QTL
176 associations by conducting the same analyses (see Supplementary Note 2).

177 To identify e_gQTLs that shared the same regulatory variants, we performed pairwise colocalization using *coloc.abf*[41]
178 between e_gQTLs in fetal-like iPSC-PPC, adult islets [11], and adult whole pancreas samples [10]. We considered only
179 e_gQTLs that had at least one variant with causal PP ≥ 1% (from genetic fine-mapping [41]), were outside the MHC
180 region, and associated with genes annotated in GENCODE version 34 [47] (Supplementary Data 12). From
181 colocalization, we identified 7,893 pairs of e_gQTLs that displayed high evidence of colocalization with PP.H4 ≥ 80%,
182 and 8,570 e_gQTLs that did not colocalize with e_gQTLs (Supplementary Data 11). Hereafter, we refer to eQTLs that did
183 not colocalize with eQTLs as "singletons" and those that colocalized with another eQTL (PP.H4 ≥ 80%; same or
184 different tissue) as "combinatorial" (i.e., the 7,893 pairs of e_gQTLs).

185 We next sought to identify singleton and combinatorial e_gQTL signals that were unique to iPSC-PPC or shared between
186 iPSC-PPC and the adult pancreatic tissues. The singleton iPSC-PPC e_gQTLs were associated with a single eGene and
187 not active in the adult pancreatic samples, and hence tissue-unique. To ensure that there was no overlap of singleton
188 e_gQTLs with other e_gQTLs in either the fetal-like and adult pancreas tissues, we implemented an LD filter ($r^2 \geq 0.2$
189 with any e_gQTL within 500 Kb or within 500 Kb if LD could not be calculated; see Methods). We identified 3,517
190 tissue-unique singleton e_gQTLs (887 iPSC-PPC + 703 adult islet + 1,927 adult whole pancreas) that were not in LD
191 with any nearby e_gQTLs (Figure 3A, Supplementary Data 12).

192 To identify tissue-unique combinatorial e_gQTL signals, we created a network using the 7,893 pairs of colocalized
193 e_gQTL associations. We identified 1,852 e_gQTL modules that passed specific criteria for module identification and LD
194 filters (see Methods), of which 939 (50.7%) comprised two e_gQTLs while the remaining 913 (49.3%) had an average
195 of four e_gQTLs per module (range: 3-20 e_gQTLs) (Supplementary Data 12, Supplementary Data 13). In total, we
196 identified 199 (10.7% of 1,852) modules that were tissue-unique, of which 10 were iPSC-PPC-unique, 30 adult islet-
197 unique, and 159 adult whole pancreas-unique (Figure 3B), and altogether comprised 21, 62, and 354 e_gQTLs in
198 combinatorial associations, respectively (Figure 3A). In contrast, the remaining 1,653 (89.3% of 1,852) modules were
199 associated with multiple pancreatic tissues, of which 670 were shared between only adult islet and whole pancreas
200 tissues (referred to as "adult-shared"), 53 were shared between only iPSC-PPC and adult islets ("fetal-islet"), 278
201 between only iPSC-PPC and adult whole pancreas ("fetal-whole-pancreas"), and 652 between all three pancreatic
202 tissues ("fetal-adult") (Figure 3B). Together, the 983 (53 + 278 + 652) modules shared between iPSC-PPC and an adult
203 pancreatic tissue were composed of 1,122 iPSC-PPC, 870 adult islets, and 1,394 adult whole pancreas e_gQTLs
204 (Supplementary Data 12, Supplementary Data 13).

205 For e_{AS}QTLs, we observed similar trends in which the majority of fetal-like iPSC-PPC-unique e_{AS}QTLs were singletons
206 and that combinatorial e_{AS}QTLs were likely shared with the adult pancreas tissues (see Supplemental Note 2;
207 Supplementary Figure 11). Altogether, including e_{AS}QTLs, we identified 1,805 iPSC-PPC eQTLs that were unique to
208 fetal-like iPSC-PPC, of which 1,518 (887 e_gQTLs + 631 e_{AS}QTLs) functioned as singletons and 287 (21 e_gQTLs + 266
209 e_{AS}QTLs) in modules; while 1,977 (1,175 e_gQTLs + 802 e_{AS}QTLs) were shared with adult pancreatic tissues, and 4,326
210 (2,066 e_gQTLs + 2,260 e_{AS}QTLs) failed one or more the stringent filters and were marked as ambiguous
211 (Supplementary Data 12, Supplementary Data 13).

212 Functional enrichment and characterization of fetal-like iPSC-PPC-unique e_gQTLs

213 To functionally characterize the fetal-like iPSC-PPC tissue-unique singleton and combinatorial e_gQTLs, we calculated
214 their enrichments in chromatin state annotations from human ESC-derived PPCs[14]. At high PP thresholds, we observed
215 the strongest enrichment of singleton e_gQTLs in active promoter (TssA) regions, consistent with their role in regulating
216 the expression of a single gene (Figure 3C, Supplementary Data 14). For combinatorial e_gQTLs, we observed a strong

217 enrichment in PPC-specific stretch enhancer (PSSE) regions at high PP thresholds (p = 0.001, OR = 1,345, PP threshold
218 = 60%) (Figure 3C, Supplementary Data 14), consistent with their involvement in the transcriptional regulation of
219 multiple genes. We also evaluated the enrichment of fetal-like iPSC-PPC-specific singleton and combinatorial e_gQTLs
220 in adult islet chromatin states [48] (Figure 3C). No meaningful enrichments were observed for fetal-like iPSC-PPC-unique
221 combinatorial e_gQTLs, but iPSC-PPC-unique singleton e_gQTLs were enriched in adult promoter regions ($p = 4.1 \times 10^{-}$
222 ^{4}, OR = 28.4, PP threshold = 80%). For example, we observed that the iPSC-PPC-unique singleton e_gQTL in the
223 *CDC37L1-DT* locus overlapped an active promoter region in the adult islet, while in both adult islet and adult whole
224 pancreas, the variants in the same region are not active (Figure 3D). Overall, these results show that the e_gQTLs
225 annotated as iPSC-PPC tissue-unique were enriched in regulatory elements consistent with their proposed functions.

226 Next, we present three examples of tissue-unique e_gQTL modules that further illustrate context-specificity of regulatory
227 variants in the three pancreatic tissues. We identified the e_gQTL module GE_3_1 ("GE" means that this module is
228 associated with gene expression) as a fetal-unique e_gQTL locus (ch3:148903264-148983264) because the underlying
229 genetic variants were associated with *CP* and *HPS3* expression in only fetal-like iPSC-PPC while in adult islets and
230 whole pancreas, the variants were not detected as e_gQTLs (Figure 3E-F). Similarly, GE_15_13 was an adult islets-
231 unique e_gQTL locus (chr15:57746360-57916360) associated with *GCOM1, MYZAP,* and *POLR2M* expression, while
232 in the other two pancreatic tissues, the variants were inactive and not associated with gene expression (Figure 3G-H).
233 Finally, we discovered GE_5_32 as an adult whole pancreas-unique e_gQTL locus (chr5:146546063-146746063)
234 associated with *STK32A* and *STK32A-AS1* expression in only the adult whole pancreas (Figure 3I-J). Together, these
235 results show that gene regulation varies between fetal-like and adult pancreatic stages, as well as between the two adult
236 tissues, further demonstrating the importance of profiling multiple contexts of the pancreas to delineate molecular
237 mechanisms underlying pancreatic disease.

238 Regulatory plasticity in combinatorial e_gQTLs shared between fetal-like and adult pancreatic 239 tissues

240 Regulatory elements are known to have context-specific gene interactions [49]. To explore this further, we examined the
241 983 e_gQTL modules shared between fetal-like iPSC-PPC and adult pancreatic tissues and determined whether the
242 modules were associated with the same or different eGenes between the two stages. We characterized the eGene
243 overlap in five different ways (Figure 4A, Supplementary Data 13): A) 200 (20.3%) e_gQTL modules were associated
244 with same eGene(s) (range: 1-2) between fetal-like iPSC-PPC and only one of the two adult pancreatic tissues; B) 305
245 (31.0%) were associated with the expression of the same eGene(s) (range 1-2) in the fetal-like and both adult tissues;
246 C) 350 (35.6%) were associated with 2-12 eGenes, some of which were shared, but at least one eGene was different
247 between the fetal-like and at least one of the adult tissues (referred to as "partial overlap"); D) 88 (9.0%) were associated
248 with different eGenes (range: 2-5) between fetal-like iPSC-PPCs and one of the two adult pancreatic tissues; and E)

249 the remaining 40 (4.1%) were associated with different eGenes (range: 2-7) between the fetal-like and both adult islet
250 and whole pancreas tissues (i.e., there is no overlap of eGenes between the two developmental stages). These data show
251 that 51.3% (505, categories A and B) of the modules shared between the iPSC-PPCs and adult pancreatic tissues
252 regulated expression of the same genes, while 48.7% (478, categories C-E) displayed spatiotemporal regulatory
253 plasticity.

254 Here, we illustrate examples of e_gQTL modules in three intervals that display regulatory plasticity between fetal-like
255 and adult states. In the chr1:201300813-201450813 locus, we identified a fetal-adult e_gQTL module (GE_1_163) that
256 comprised e_gQTL associations for different eGenes in iPSC-PPC and the two adult pancreatic tissues, specifically
257 *AC119427.1* in iPSC-PPC and *TNNI1* in the two adult tissues (Figure 4B). Likewise, the chr19:4213666-4433666 locus
258 corresponding to a fetal-adult e_gQTL module (GE_19_90) was associated with *MPND* expression in only iPSC-PPC
259 but in adult islets and whole pancreas, the underlying variants were associated with *STAP2* expression (Figure 4C).
260 Finally, the fetal-adult e_gQTL locus (GE_10_11) in chr10:1273918-1276118 affected *UROS* expression in all three
261 pancreatic tissues but in adult islets, the underlying variants also affected *BCCIP* expression (Figure 4D). Together,
262 these genomic loci illustrate examples of regulatory plasticity observed in genetic variants in which their genotypes
263 incur different impacts on transcriptional activity depending on the life stage of the pancreas.

264 Altogether, including e_{AS}QTLs, we discovered 655 (478 e_gQTL + 177 e_{AS}QTL modules, categories C-E) shared eQTL
265 loci that displayed regulatory plasticity in which the underlying regulatory variants were associated with one or more
266 different genes and could thereby affect different biological processes (see Supplemental Note 2, Supplementary Figure
267 11, Supplementary Data 13). These 655 shared eQTL loci comprise 1,043 iPSC-PPC, 934 adult islet, and 1,111 adult
268 whole pancreas eQTL associations (Supplementary Data 12).

269 Associations of spatiotemporal eQTLs with pancreatic traits and disease phenotypes

270 To better understand the role of regulatory variants associated with complex human traits and disease during early
271 development and adult pancreatic stages, we performed colocalization between GWAS signals and eQTLs (e_gQTL and
272 e_{AS}QTL) detected in fetal-like iPSC-PPC, adult islets, and adult whole pancreas tissues. For this analysis, we considered
273 GWAS data from ten different studies for two diseases involving the pancreas, including type 1 diabetes (T1D) [3] and
274 type 2 diabetes (T2D) [4], and seven biomarkers related to three traits: 1) glycemic control (HbA1c levels and fasting
275 glucose [FG]) [2,50]; 2) obesity (triglycerides, cholesterol, HDL level, and LDL direct) [50]; and 3) body mass index (BMI)
276 [50] (Supplementary Data 15).

277 *Singleton eQTLs*

278 Out of the 6,101 singleton eQTLs (3,517 e_gQTLs and 2,584 e_{AS}QTLs; see Supplemental Note 2) in the fetal-like iPSC-
279 PPC and two adult pancreatic tissues, we found 118 (1.9%) that displayed strong evidence for colocalization with at

280 least one GWAS signal, including 21 (of 1,518 total singleton eQTLs; 1.4%) fetal-like iPSC-PPC, 57 (of 2,225; 2.6%)
281 adult islets, and 40 (of 2,358; 1.7%) adult whole pancreas singleton eQTLs (Figure 5A, Supplementary Data 15). Given
282 that some traits were highly correlated with one another [51,52], we observed 38 singleton eQTLs that colocalized with
283 GWAS variants associated with more than one trait (range: 2-6 traits). In total, we identified 183 GWAS loci across
284 the ten traits that colocalized with fetal-like or adult pancreatic singleton eQTLs (each combination of colocalized
285 eQTL-GWAS trait variants was counted as a separate locus; Supplementary Figure 12, Supplementary Figure 13,
286 Supplementary Data 15). We next identified putative causal variants underlying both eQTL and trait associations using
287 *coloc.abf* [41] and constructed 99% credible sets (i.e., set of variants with a cumulative causal PP ≥ 99%; see Methods).
288 Of the total 183 colocalized GWAS loci, we resolved 21 to a single putative causal variant while 63 had between two
289 and ten variants and the remaining 99 had more than ten variants with an average of ~46 variants per locus (Figure 5B,
290 Supplementary Data 16).

291 *eQTL modules*

292 We next analyzed the combinatorial eQTLs for GWAS colocalization. We considered an eQTL module to overlap with
293 GWAS variants if more than 30% of the eQTLs in the module colocalized with PP.H4 ≥ 80% and the number of H4
294 associations was twice greater than the number of H3 associations (see Methods). Of the 2,832 (1,852 e_gQTL and 980
295 e_{AS}QTL) modules, 89 (57 e_gQTL + 32 e_{AS}QTL; 3.1%) colocalized with a total of 129 GWAS loci across the ten traits
296 (Supplementary Figure 12, Supplementary Figure 13, Supplementary Data 15). Of these 89 GWAS-colocalized
297 modules, 5 were iPSC-PPC-unique, 36 were shared between both iPSC-PPC and adult, (4 fetal-islet, 11 fetal-whole-
298 pancreas, and 21 fetal-adult modules), and 48 were associated with only adult (22 islet-unique, 5 whole pancreas-
299 unique, 21 adult-shared) (Figure 5A, Supplementary Figure 13). We observed that all 5 iPSC-PPC-unique eQTL
300 modules corresponded to e_{AS}QTL modules. This finding aligns with multiple studies showing that alternative splicing
301 is more dynamic and extensive in embryonic and fetal stages [53–55]. The 89 modules comprised 49 iPSC-PPC eQTLs
302 (41 genes), 98 adult islets eQTLs (75 genes), and 71 adult whole pancreas eQTLs (69 genes). To fine-map each of the
303 129 colocalized GWAS loci, we used the eQTL in the module that resulted in the least number of putative causal
304 variants (see Methods). 15 GWAS loci had a credible set size of one variant, 54 with two to ten variants, and the
305 remaining 60 had more than ten variants and an average of ~32 variants per set (Figure 5C, Supplementary Data 16).

306 Altogether, these results show complex pancreatic disease and trait GWAS variants colocalized with regulatory
307 variants that were uniquely active in either the fetal-like or adult developmental stages and with regulatory variants
308 shared across the life stage of the pancreas. Furthermore, our data show the utility of using spatiotemporally informed
309 eQTLs for fine-mapping causal variants in GWAS loci.

310 Spatiotemporally informed eQTL resource provides mechanistic insights into GWAS signals

311 To assess the utility of our spatiotemporally informed eQTL resource for interpreting GWAS signals, we initially
312 examined the role of regulatory plasticity in pancreatic disease and traits. We examined the 36 eQTL modules that
313 were shared between fetal-like iPSC-PPC and the adult pancreatic tissues (i.e., fetal-adult, fetal-islet, and fetal-whole-
314 pancreas) and colocalized with GWAS signals. Thirty of these modules were associated with the same genes (categories
315 A and B), while one was associated with partially overlapping eGenes (category C), and five were associated with
316 entirely different eGenes (category D and E) (see Figure 4A for category definitions) (Supplementary Data 15). These
317 results show that while the function of shared GWAS regulatory variants tended to be conserved across the fetal-like
318 and adult pancreatic stages, a subset (17%, n = 6 of 36 total eQTL modules) were associated with distinct genes between
319 the two stages

320 To further assess the utility of our spatiotemporally informed eQTL resource, we next examined GWAS signals that
321 could only be interpreted by including fetal pancreatic eQTLs. We calculated the fraction of GWAS loci that
322 colocalized with only iPSC-PPC eQTLs, only adult islet eQTLs, and only adult whole pancreas eQTLs. For fair
323 comparisons, we considered only e_gQTLs in this assessment. Of the 191 GWAS loci that colocalized with an e_gQTL
324 (Supplementary Data 15), we found that 13% (24 loci) colocalized with 16 iPSC-PPC-unique e_gQTLs, 25% (47) with
325 27 adult islet-unique e_gQTLs, and 28% (53) with 46 adult whole pancreas-unique e_gQTLs. The remaining 35% (67)
326 GWAS loci colocalized with 121 e_gQTLs shared between multiple tissues. We next determined how many of the 16
327 iPSC-PPC-unique e_gQTLs were active in 48 non-pancreatic tissues in the GTEx study [10]. We calculated LD (r2 > 0.2
328 within 500 Kb or within 500 Kb if LD information was not available) between the lead variants of each of the 16 iPSC-
329 PPC-unique e_gQTL and each e_gQTL for the non-pancreatic tissues in GTEx. We identified 8 (31% of 16, all singletons)
330 that were independent and exclusive to the fetal-like iPSC-PPC dataset (i.e., did not have LD; Supplementary Data 17).
331 One of these 8 iPSC-PPC e_gQTLs (*TPD52* e_gQTL) is described below in further detail. These results show that
332 integrating fetal-like iPSC-PPC eQTLs can help resolve certain GWAS loci that cannot be resolved using only adult
333 datasets.

334 Below, we demonstrate the application of our spatiotemporally informed eQTL resource by providing a detailed
335 description of eight GWAS loci. We propose potential causal mechanisms and offer insights into their spatiotemporal
336 contexts.

337 **Singleton e_gQTLs:** Here, we elucidate probable causal mechanisms for two GWAS loci associated with FG levels and
338 T1D-risk that colocalized with iPSC-PPC-unique singleton e_gQTLs.

339 *chr8:80998464-81093464 and TPD52 (iPSC-PPC-unique singleton)*

340 We found that in the chr8:80998464-81093464 locus, a GWAS signal associated with FG levels colocalized with a
341 fetal-like iPSC-PPC-unique singleton e_gQTL for *TPD52*, also known as tumor protein D52 (effect size = -0.99, PP.H4
342 = 91.7%) (Figure 6A, Supplementary Figure 14A, Supplementary Data 15). The reported causal variant underlying
343 this GWAS signal is rs12541643 [2]; however, colocalization with our eQTLs identified rs12549167

344 (chr8:81078464:C>T, PP = 33.9%, r^2 = 0.317 with rs12541643, Supplementary Data 16) as the most likely candidate
345 causal variant underlying both *TPD52* expression in fetal-like iPSC-PPC and FG association. TPD52 is a direct
346 interactor with the AMP-activated protein kinase (AMPK) and negatively affects AMPK signaling. AMPK controls a
347 wide range of metabolic processes and is responsible for maintaining cellular energy homeostasis particularly in tissues
348 associated with obesity, insulin resistance, T2D, and cancer such as muscle, liver, hypothalamus, and the pancreas [56–
349 59]. Dysregulation of AMPK has also been associated with developmental defects in which AMPK activation can lead
350 to fetal malformation [60]. Our findings suggest that decreased expression of *TPD52* during development may influence
351 changes in glucose metabolism and therefore fasting glucose levels during adult stage.

352 *chr9:4232083-4352083 and CDC37L1-DT (iPSC-PPC-unique singleton)*

353 We found that the well-known *GLIS3* GWAS locus associated with FG and T1D-risk [61,62] colocalized with a fetal-like
354 iPSC-PPC-unique singleton e_gQTL for the lncRNA CDC37L1 divergent transcript (*CDC37L1-DT*; effect size = 1.46;
355 PP.H4 for FG and T1D = 92.4% and 91.2%, respectively, Figure 6B, Supplementary Figure 14B, Supplementary Data
356 15). Consistent with previous studies [61,62], we identified rs10758593 (chr9:4292083:G>A, PP = 79.2%) as the lead
357 candidate causal variant underlying both eQTL and GWAS associations. Because *GLIS3* plays a critical role in
358 pancreatic beta cell development and function [58,59,60], it has often been reported as the susceptibility gene for this signal,
359 however it remains unclear what effects rs10758593 has on *GLIS3* expression. Our analysis suggests that another
360 potential gene target of rs10758593, specifically during pancreas development, is *CDC37L1-DT*. While the molecular
361 function of *CDC37L1-DT* is unknown, the gene has been associated with 9p duplication in neurodevelopmental
362 disorders [66]. Furthermore, a recent study observed a significant association between the rs10758593 risk allele and birth
363 weight, indicating a development role played by this locus [67]. Although additional studies are needed to understand the
364 function of *CDC37L1-DT* during pancreas development and in T1D pathology, our analysis indicates that *CDC37L1-*
365 *DT* may be another candidate susceptibility gene for the variants in the *GLIS3* locus. Assessment of *GLIS3* e_gQTLs in
366 the three pancreatic tissues showed that there was no overlap between the e_gQTLs and GWAS variants (Supplementary
367 Figure 15A).

368 **Combinatorial e_gQTLs:** Below, we describe two GWAS intervals associated with cholesterol, LDL direct levels, and
369 T1D. We show that the GWAS variants colocalized with combinatorial e_gQTLs, indicating that multiple genes, and
370 possibly multiple developmental stages of the pancreas, may be involved in trait predisposition.

371 *chr22:41049522-41449522 and ADSL and ST13 (fetal-adult combinatorial)*

372 We found that the GWAS signals associated with cholesterol and LDL direct levels in the chr22:41049522-41449522
373 locus colocalized with a "fetal-adult" e_gQTL module (module ID: GE_22_63, category E) (Figure 6C, Supplementary
374 Figure 14C-D, Supplementary Data 15). The module was associated with different eGenes between fetal-like iPSC-
375 PPC and both adult pancreatic tissues, in which the GWAS variants were associated with *ADSL* expression in iPSC-
376 PPC (effect size = 0.78) but *ST13* expression in adult whole pancreas (effect size = 0.27) and adult islets (effect size =

377 -0.15, weakly associated). Infants born with ADSL (adenylosuccinate lyase) deficiency suffer from impaired glucose
378 and lipid metabolism, while ST13, also known as Hsc70-interacting protein, is involved in lipid metabolism [68].
379 Overexpression of *ST13* was found to result in disordered lipid metabolism in chronic pancreatitis [68]. Although *ST13*
380 was reported to be the candidate causal gene for this locus [69], we determined that the underlying variants may also
381 affect *ADSL* expression but specifically during early pancreas development. Congruent with a previous study [69], our
382 colocalization identified rs138349 (chr22:41249522:A>G, PP = 21.9% for cholesterol and 20.9% for LDL) as the lead
383 candidate causal variant for the e_gQTLs and both cholesterol and LDL GWAS associations (Supplementary Data 16).
384 Altogether, annotation of the chr22:41049522-41449522 GWAS locus using our pancreatic eQTL resource suggests
385 that altered expression of *ADSL* during pancreas development and *ST13* in adult pancreatic tissues may contribute to
386 changes in cholesterol and LDL direct levels in adult. Additional studies are required to understand the degree to which
387 *ADSL* and *ST13* are causal for cholesterol and LDL direct levels.

388 *chr10:90001035-90066035 and PTEN and LIPJ (adult whole pancreas-unique combinatorial)*

389 We found a T1D-risk signal in the chr10:90001035-90066035 locus that colocalized with an "adult whole pancreas-
390 unique" e_gQTL module (module ID: GE_10_35) associated with *PTEN* and *LIPJ* expression in adult whole pancreas
391 (effect size = 0.48 and 0.49, respectively) (Supplementary Figure 14E, Supplementary Figure 15B, Supplementary
392 Data 15). Colocalization identified the distal regulatory variant rs7068821 (chr10:90051035:G>T; PP = 85.5%) as the
393 most likely candidate causal variant (Supplementary Data 16), which is in LD with the reported index SNP rs10509540
394 (r^2 = 0.876) in the GWAS catalogue. While *RNLS* was reported to be the susceptibility gene for this locus [70], our
395 analysis suggests that both *PTEN* and *LIPJ* may be candidate causal genes for this locus. Previous studies have shown
396 that knockout of pancreas-specific *PTEN* (PPKO) in mice resulted in enlarged pancreas and elevated proliferation of
397 acinar cells. PPKO mice also exhibited hypoglycemia, hypoinsulinemia, and altered amino metabolism [71]. LIPJ
398 encodes the lipase family member J and is involved in lipid metabolism [72]. Our findings provide additional biological
399 insight into the chr10:900001035-90066035 T1D locus and support previous studies suggesting a potential causal role
400 of the adult whole pancreas in T1D pathogenesis [3,67].

401 **Singleton and combinatorial e_{AS}QTLs:** Here, we illustrate three examples of putative causal variants involved in
402 alternative splicing uniquely in the fetal-like iPSC-PPC. Long-noncoding RNAs (lncRNAs) have previously been
403 shown to play important roles in pancreatic diseases [73]. One of our examples includes a lncRNA while two involve
404 protein-coding genes.

405 *chr14:101286447-101326447 and MEG3 (iPSC-PPC-unique singleton)*

406 The chr14:101286447-101326447 is a well-known GWAS locus associated with T1D and has been reported to affect
407 the lncRNA maternally expressed gene 3 (*MEG3*). While the role of *MEG3* in T1D and T2D pathogenesis has been
408 extensively studied [74–76], the genetic mechanism by which this locus affects *MEG3* expression and therefore, T1D-risk
409 is not well understood. Using our pancreatic eQTL resource, we found that the GWAS signal colocalized with a fetal-

like iPSC-PPC-unique singleton e_{AS}QTL for a *MEG3* isoform (ENST00000522618, PP.H4 = 98%, effect size = 1.3, Figure 7A, Supplementary Figure 16A, Supplementary Data 15). Colocalization with the *MEG3* e_{AS}QTL identified rs56994090 (chr14:101306447:T>C, PP = 100%) as the most likely candidate causal variant, which is concordant with the findings of a previous GWAS study [77] (Supplementary Data 16). Given that rs56994090 is located in the novel intron enhancer of *MEG3* [74], we hypothesize that alternative splicing of *MEG3* may alter the enhancer's regulatory function, as previously observed in other lncRNAs [78], and thereby, affect T1D-risk. Altogether, our findings describe a potential causal mechanism for the T1D-risk locus involving differential alternative splicing of *MEG3* specifically during pancreas development.

chr16:684685635-68855635 and CDH3 (iPSC-PPC-unique singleton)

We determined a known GWAS signal in the chr16:684685635-68855635 locus associated with HbA1c levels [79] colocalized with a fetal-like iPSC-PPC-unique singleton e_{AS}QTL for the P-cadherin 3 (*CDH3*) isoform ENST00000429102 (effect size = -1.6, PP.H4 = 83.1%) (Figure 7B, Supplementary Figure 16B, Supplementary Data 15). Colocalization using the e_{AS}QTL identified intronic variant rs72785165 (chr16:68755635:T>A, PP = 6.8%) as the most likely candidate causal variant (Supplementary Data 16), which is in high LD with the reported GWAS SNP (rs4783565, r^2 = 0.88) [79]. While it remains unclear how alternative splicing of *CDH3* affects HbA1c levels, studies have shown that chimeric proteins made of cadherin ectodomains, including the P-cadherin CDH3, are important for proper insulin secretion by pancreatic beta cells [80]. Based on our findings, we hypothesize that differential isoform usage of *CDH3* during pancreas development may influence glucose control and therefore, HbA1c levels, in adults.

chr13:30956642-31116642 and HMGB1 (iPSC-PPC-unique combinatorial)

The GWAS signals associated with T2D and BMI in the chr13:30956642-31116642 locus [81–84] colocalized with the iPSC-PPC-unique e_{AS}QTL module (module ID: AS_13_2) associated with three *HMGB1* isoforms: ENST00000326004, ENST00000399494 , ENST00000339872, and (effect size = 2.16, -2.26, and -0.85, respectively; HMGB1.1, HGMB1.2, and HMGB1.3, respectively) (Figure 7C, Supplementary Figure 16C-E, Supplementary Data 15). Our colocalization identified rs3742305 (chr13:31036642:C>G, PP = 49.3%) as a lead candidate causal variant underlying this locus, in which the risk allele (G) was associated with increased usage of ENST00000326004 and decreased usages of ENST0000339872 and ENST00000399494 (Supplementary Figure 16C-E, Supplementary Data 7, Supplementary Data 16). While a previous study [83] also reported *HMGB1* as the susceptibility gene, the precise mechanism by which rs3742305 affected *HMGB1* expression was unclear. HMGB1, also known as high-mobility group box 1, is an important mediator for regulating gene expression during both developmental and adult stages of life. Deletion of *HMGB1* disrupts cell growth and causes lethal hypoglycemia in mouse pups [85]. In T2D, *HMGB1* promotes obesity-induced adipose inflammation, insulin resistance, and islet dysfunction [84]. Our results suggest that differential usage of *HMGB1* isoforms during pancreas development may affect adult risk of developing obesity and/or T2D.

443 Altogether, our findings demonstrate the value of our pancreatic eQTL resource to annotate GWAS risk variants with
444 fetal-like and adult temporal and spatial regulatory information. We show that some causal regulatory variants
445 underlying disease-associated signals may influence adult traits by modulating the expression of genes in early
446 development, while in other cases, they may display regulatory plasticity and exert their effects by modulating the
447 expression of multiple different genes in fetal-like and adult pancreatic stages. Further, we identified an association
448 between whole pancreas and T1D, supporting a potential role of this tissue in diabetes pathogenesis [3].

449 **Discussion**

450 In this study, we leveraged one of the most well-characterized iPSC cohorts comprising >100 genotyped individuals
451 to derive pancreatic progenitor cells and generate a comprehensive eQTL resource for examining genetic associations
452 with gene expression and isoform usage in fetal-like pancreatic cells. We discovered 8,665 eQTLs in the fetal-like
453 iPSC-PPCs and showed that 60% of eGenes were associated with regulatory variation uniquely active during pancreas
454 development. For the eGenes that were shared with adult, ~12% were associated with different genomic loci, indicating
455 that different regulatory elements may modulate the same gene in fetal-like and adult pancreatic stages. We further
456 identified regulatory variants that displayed early pancreas development-unique function, of which 1,805 were
457 uniquely active in only iPSC-PPC and 1,043 were active in both developmental and adult contexts but exhibited
458 regulatory plasticity in the genes they regulate. These results concur with previous studies showing that the genetic
459 regulatory landscape changes between fetal tissues and their adult counterparts [87–89], and therefore, highlights the
460 importance of assessing variant function in both fetal and adult tissue contexts. Furthermore, it is widely known that
461 tight regulation of genes during development is essential [90], and our study reflects this in our findings that the majority
462 of developmental-unique eQTLs were restricted to a single eGene.

463 Finally, we highlighted examples of GWAS associations for which we utilized our spatiotemporally informed eQTL
464 resource to characterize causal risk mechanisms underlying adult pancreatic disease. We showed that some causal
465 regulatory variants underlying GWAS signals identified in the fetal-like iPSC-PPCs modulate the expression of genes
466 in early development, while others may exert their effects by modulating the expression of multiple different genes
467 across fetal-like and adult pancreatic stages. Of note, many of the fetal-unique regulatory variants underlying the
468 GWAS signals were e_{AS}QTLs, which is consistent with alternative splicing playing a key role in developing tissues [53–
469 55,91]. Hence, we believe that contribution of alternative splicing differences during fetal pancreas development to
470 complex traits warrants further investigation.

471 We offer limitations in our study and potential future directions for the field at large. We believe that studies using
472 larger sample sizes are needed to identify additional associations between genetic variation and gene expression in fetal
473 samples. Our eQTL mapping in fetal-like iPSC-PPC was conducted on much fewer samples compared to the two adult
474 studies that each used ~300-400 samples, rendering our dataset underpowered and not able to capture additional eQTL

475 associations that could be shared with the adult pancreatic tissues. Therefore, several eQTLs we annotated as adult
476 islet-unique or whole pancreas-unique may in reality be shared with fetal pancreas. Further, power differences between
477 the studies may also cause the observed results where there were many singleton eQTLs observed in a single tissue.
478 On the other hand, the eQTLs we annotated as iPSC-PPC-unique may less likely to be shared, as the signals in the
479 adult datasets are better powered and therefore sufficient for comparing against iPSC-PPC signals. Additionally, with
480 the rapid generation of eQTL datasets from different tissue contexts [1,2], the development and application of artificial
481 intelligence and machine learning as ways to identify shared eQTL associations between multiple tissues will be
482 extremely useful. While pairwise colocalization and network analysis can identify shared eQTL regulatory loci across
483 a handful of tissues, machine learning approaches could scale these analyses across spatiotemporal contexts of all
484 tissues, thereby providing valuable insights into regulatory elements that are exclusive to a specific context and also
485 those that exhibit regulatory plasticity across multiple contexts.

486 In summary, our study provides a valuable resource for discovering causal regulatory mechanisms underlying
487 pancreatic traits and disease across developmental and adult timepoints of the pancreas. We reveal that disease variants
488 may either display temporal-specificity in which they affect gene expression specifically in one timepoint, or regulatory
489 plasticity, in which they affect gene expression in multiple timepoints but affect different genes. Our findings lay the
490 groundwork for future employment of development contexts for the characterization of disease-associated variants.

491 # Methods

492 ## Subject Information

493 We used iPSC lines from 106 individuals recruited as part of the iPSCORE project (Supplementary Data 1). There
494 were 53 individuals belonging to 19 families composed of two or more subjects (range: 2-6). Each subject was assigned
495 an iPSCORE_ID (i.e., iPSCORE 4_1), where "4" indicates the family number and "1" indicates the individual number,
496 and a 128-bit universal unique identifier (UUID). The 106 individuals included 68 females and 38 males with ages
497 ranging from 15 to 88 years old at the time of enrollment. Recruitment of these individuals was approved by the
498 Institutional Review Boards of the University of California, San Diego, and The Salk Institute (project no. 110776ZF).

499 ## WGS data

500 Whole-genome sequencing data for the 106 iPSCORE individuals were downloaded from dbGaP (phs001325.v3) as a
501 VCF file [37]. We retained variants with MAF > 5% across all 273 individuals in the iPSCORE resource, that were in
502 Hardy-Weinberg equilibrium ($p > 10^{-6}$), and that were within 500 Kb of the expressed gene's body coordinates.
503 Specifically, we expanded the coordinates of each of the 16,464 expressed autosomal genes (500 Kb upstream and
504 downstream) and extracted all variants within these regions using *bcftools view* with parameters *--f PASS -q 0.05:minor*

505 [93]. Next, we normalized indels and split multi-allelic variants using *bcftools norm -m-* and removed variants that were
506 genotyped in fewer than 99% of samples using *bcftools filter -i 'F_PASS(GT!="mis") > 0.99* [93]. Finally, we converted
507 the resulting VCF files to text using *bcftools query* [93] and converted the genotypes from character strings (0/0, 0/1, and
508 1/1) to numeric (0, 0.5, and 1, respectively). This resulted in 6,593,484 total variants used for eQTL mapping.

509 **iPSC Generation**

510 Generation of the 106 iPSC lines has previously been described in detail [35]. Briefly, cultures of primary dermal
511 fibroblast cells were generated from a punch biopsy tissue [94], infected with the Cytotune Sendai virus (Life
512 Technologies) per manufacturer's protocol to initiate reprogramming. Emerging iPSC colonies were manually picked
513 after Day 21 and maintained on Matrigel (BD Corning) with mTeSR1 medium (Stem Cell Technologies). Multiple
514 independently established iPSC clones (i.e. referred to as lines) were derived from each individual. Many of the iPSC
515 lines were evaluated by flow cytometry for expression of two pluripotent markers: Tra-1-81 (Alexa Fluor 488 anti-
516 human, Biolegend) and SSEA-4 (PE anti-human, Biolegend) [35]. Pluripotency was also examined using PluriTest-
517 RNAseq [35]. This iPSCORE resource was established as part of the Next Generation Consortium of the National Heart,
518 Lung and Blood Institute and is available to researchers through the biorepository at WiCell Research Institute
519 (www.wicell.org; NHLBI Next Gen Collection). For-profit organizations can contact the corresponding author directly
520 to discuss line availability.

521 **Pancreatic Progenitor Differentiation**

522 We performed pancreatic progenitor cell (PPC) differentiation on each of the 106 iPSC lines. One iPSC line was
523 differentiated twice giving a total of 107 differentiations. Each differentiation was assigned a 128-bit universally unique
524 identifier (UUID), and a unique differentiation ID (UDID; "PPCXXX"), where "XXX" represents a numeric integer
525 (Supplementary Data 2).

526 *Differentiation Protocol*

527 The iPSC lines were differentiated into PPCs using the STEMdiff™ Pancreatic Progenitor Kit (StemCell Technologies)
528 protocol with minor modifications. Briefly, iPSC lines were thawed into mTeSR1 medium containing 10 µM Y-27632
529 ROCK Inhibitor (Selleckchem) and plated onto one well of a 6-well plate coated with Matrigel. iPSCs were grown
530 until they reached 80% confluency [95] and then passaged using 2mg/ml solution of Dispase II (ThermoFisher Scientific)
531 onto three wells of a 6-well plate (ratio 1:3). To expand the iPSC cells for differentiation, iPSCs were passaged a second
532 time onto six wells of a 6-well plate (ratio 1:2). When the iPSCs reached 80% confluency, cells were dissociated into
533 single cells using Accutase (Innovative Cell Technologies Inc.) and resuspended at a concentration of 1.85 x 10^6
534 cells/ml in mTeSR medium containing 10 µM Y-27632 ROCK inhibitor. Cells were then plated onto six wells of a 6-
535 well plate and grown for approximately 16 to 20 hours to achieve a uniform monolayer of 90-95% confluence (3.7 x

536 10^6 cells/well; about 3.9 x 10^5 cells/cm^2). Differentiation of the iPSC monolayers was initiated by the addition of the
537 STEMdiff™ Stage Endoderm Basal medium supplemented with Supplement MR and Supplement CJ (2 ml/well) (Day
538 1, D1). The following media changes were performed every 24 hours following initiation of differentiation (2 ml/well).
539 On D2 and D3, the medium was changed to fresh STEMdiff™ Stage Endoderm Basal medium supplemented with
540 Supplement CJ. On D4, the medium was changed to STEMdiff™ Pancreatic Stage 2-4 Basal medium supplemented
541 with Supplement 2A and Supplement 2B. On D5 and D6, the medium was changed to STEMdiff™ Pancreatic Stage
542 2-4 Basal medium supplemented with Supplement 2B. From D7 to D9, the medium was changed to STEMdiff™
543 Pancreatic Stage 2-4 Basal medium supplemented with Supplement 3. From D10 to D14, the medium was changed to
544 STEMdiff™ Pancreatic Stage 2-4 Basal medium supplemented with Supplement 4. On D15, cells were dissociated
545 with Accutase and then collected, counted, and processed for data generation. iPSC-PPC cells were cryopreserved in
546 CryoStor® CS10 (StemCell Technologies).

547 *iPSC-PPC Differentiation Efficiency*

548 To evaluate the efficiency of iPSC-PPC differentiation, we performed flow cytometry on two pancreatic precursor
549 markers, PDX1 and NKX6-1. Specifically, at least 2 x 10^6 cells were fixed and permeabilized using the
550 Fixation/Permeabilized Solution Kit with BD GolgiStop TM (BD Biosciences) following the manufacturer's
551 recommendations. Cells were resuspended in 1x BD Perm/Wash TM Buffer at a concentration of 1 x 10^7 cells/ml. For
552 each flow cytometry staining, 2.5 x 10^5 cells were stained for 75 minutes at room temperature with PE Mouse anti-
553 PDX1 Clone-658A5 (BD Biosciences; Catalog no. 562161; 1:10) and Alexa Fluor® 647 Mouse anti-NKX6.1 Clone
554 R11-560 (BD Bioscience; Catalog no. 563338; 1:10), or with the appropriate class control antibodies: PE Mouse anti-
555 IgG1 κ R-PE Clone MOPC-21 (BD Biosciences; Catalog no. 559320) and Alexa Fluor® 647 Mouse anti IgG1 κ
556 Isotype Clone MOPC-21 (BD Biosciences; Catalog no. 557732). PE Mouse anti-PDX1 Clone-658A5 and Alexa
557 Fluor® 647 Mouse anti-NKX6.1 Clone R11-560 were validated by the manufacturer to bind to mouse and human
558 PDX-1 and NKX6-1, respectively. Stained cells were washed three times, resuspended in PBS containing 1% BSA
559 and 1% formaldehyde, and immediately analyzed using FACS Canto II flow cytometer (BD Biosciences). The fraction
560 of PDX1- and NKX6-1-positive was calculated using FlowJo software version 10.4 (Supplementary Data 2).

561 **scRNA-seq**

562 To characterize the cellular composition of the fetal-like iPSC-PPC samples, we performed single-cell RNA-seq
563 (scRNA-seq) on one iPSC line (from differentiation PPC034) and ten iPSC-PPC samples with varying percentages of
564 double-positive PDX1+/NKX6-1+ cells based on flow cytometry (range: 9.4-91.7%) (Supplementary Figure 2,
565 Supplementary Figure 3, Supplementary Data 2). Because bulk RNA-seq was generated on cryopreserved cells, we
566 sought to also examine whether cell cryopreservation affects gene expression estimates using scRNA-seq. Therefore,
567 we included both freshly prepared (i.e., not frozen and processed immediately after differentiation) and cryopreserved

568 cells for four iPSC-PPC samples (PPC029, PPC027, PPC023, PPC034; Supplementary Data 2) for scRNA-seq
569 processing.

570 *Sample Collection*

571 Fresh cells from the iPSC line and seven iPSC-PPC samples were captured individually at D15. Cells from four of
572 these same iPSC-PPC samples that had been cryopreserved were pooled and captured immediately after thawing
573 (RNA_Pool_1). Cells from an additional three iPSC-PPC samples were captured only after cryopreservation
574 (RNA_Pool_2) (Supplementary Data 2).

575 *Library Preparation and Sequencing*

576 All single cells were captured using the 10X Chromium controller (10X Genomics) according to the manufacturer's
577 specifications and manual (Manual CG000183, Rev A). Cells from each scRNA-seq sample (one iPSC, seven fresh
578 iPSC-PPCs, RNA_Pool_1, and RNA_Pool_2) were loaded each onto an individual lane of a Chromium Single Cell
579 Chip B. Libraries were generated using Chromium Single Cell 3' Library Gel Bead Kit v3 (10X Genomics) following
580 manufacturer's manual with small modifications. Specifically, the purified cDNA was eluted in 24 µl of Buffer EB,
581 half of which was used for the subsequent step of the library construction. cDNA was amplified for 10 cycles and
582 libraries were amplified for 8 cycles. All libraries were sequenced on a HiSeq 4000 using custom programs (fresh: 28-
583 8-175 Pair End and cryopreserved: 28-8-98 Pair End). Specifically, eight libraries generated from fresh samples (one
584 iPSC and seven iPSC-PPC samples) were pooled together and loaded evenly onto eight lanes and sequenced to an
585 average depth of 163 million reads. The two libraries from seven cryopreserved lines (RNA_Pool_1 and RNA_Pool_2)
586 were each sequenced on an individual lane to an average depth of 265 million reads. In total, we captured 99,819 cells.
587 We observed highly correlated cell type proportions between fresh and cryopreserved iPSC-PPC samples
588 (Supplementary Figure 8).

589 *scRNA-seq Alignment*

590 We obtained FASTQ files for the ten scRNA-seq samples (one iPSC, seven fresh iPSC-PPCs, RNA_Pool_1, and
591 RNA_Pool_2) (Supplementary Data 2) and used CellRanger V6.0.1 (https://support.10xgenomics.com/) with default
592 parameters and GENCODE version 34 hg19 [96] gene annotations to generate single-cell gene counts and BAM files for
593 each of the ten samples.

594 *Dataset Integration and Quality Control*

595 We processed the single-cell gene counts by first aggregating the iPSC and seven fresh iPSC-PPC samples using the
596 *aggr* function on CellRanger V6.0.1 with normalization = F. Then, we integrated the aggregated dataset ("aggr") with
597 the two pools of cryopreserved samples (RNA_Pool_1 and RNA_Pool_2) using the standard integration workflow
598 described in Seurat (version 3.2; https://satijalab.org/seurat/archive/v3.2/integration.html). Specifically, for each

599 dataset (aggr, RNA_Pool_1, and RNA_Pool_2), we log-normalized the gene counts using *NormalizeData* (default
600 parameters) then used *FindVariableFeatures* with *selection.method = "vst", nfeatures = 2000,* and *dispersion.cutoff*
601 *= c(0.5, Inf)* to identify the top 2,000 most variable genes in each dataset. We then used *FindIntegrationAnchors* and
602 *IntegrateData* with *dims = 1:30* to integrate the three datasets. We scaled the integrated data with *ScaleData*, performed
603 principal component analysis with *RunPCA* for *npcs = 30*, and processed for UMAP visualization (*RunUMAP* with
604 *reduction = "pca"* and *dims = 1:30*). Clusters were identified using *FindClusters* with default parameters.

605 To remove low-quality cells, we examined the distribution of the number of genes per cell and the percentage of reads
606 mapping to the mitochondrial chromosome (chrM) in each cluster. We removed the cluster (11,677 cells) with fewer
607 than 500 genes per cell and more than 50% of the reads mapping to chrM. We re-processed the filtered data (*ScaleData*,
608 *RunPCA*, *FindClusters, RunUMAP*) and removed another cluster of cells that had the lowest median number of
609 expressed genes (723 versus 2,775) and highest median fraction of mitochondrial reads (34.0% versus 8.39%). After
610 this second filtering step, we retained 84,258 cells.

611 *Demultiplexing Sample Identity*

612 We used Demuxlet [97] to assign pooled cryopreserved cells in RNA_Pool_1 and RNA_Pool_2 (19,136 cells in total) to
613 the correct iPSC-PPC sample. Specifically, we provided CellRanger BAM files and a VCF file containing genotypes
614 for biallelic SNVs located at UTR and exon regions on autosomes as annotated by GENCODE version 34 hg19 [96]. We
615 excluded 33 cells that were incorrectly assigned to samples not associated with the pooled sample (i.e., cells from
616 RNA_POOL_1 were predicted to be from other samples not in RNA_Pool_1). 84,225 cells remained for downstream
617 analyses (Supplementary Data 3).

618 *Annotation of Cell Type Clusters*

619 We annotated the scRNA-seq clusters by first clustering with three different resolutions (0.5, 0.08, and 0.1)
620 (Supplementary Figure 4-6). We selected resolution = 0.08 because it best captured the expected iPSC-PPC cell types
621 based on each cluster's expression for the following gene markers: *POU5F1* (iPSC), *COL1A1*, *COL1A2*
622 (mesendoderm) *AFP, APOA* (early definitive endoderm), *GATA4*, *GATA6, PDX1* (early PPC), *PDX1*, *NKX6-1* (late
623 PPC), *PAX6, CHGA*, *INS*, *GCG*, *SST* (endocrine), and *FLT1* (early ductal). We validated our annotations by comparing
624 the iPSC-PPC clusters to those identified from scRNA-seq of ESC-PPC samples over 4 different stages of
625 differentiation [98] (GSE114412): Stage 3 (Day 6; 7,982 cells), Stage 4 (Day 13; 6,960 cells), Stage 5 (Day 18; 4,193
626 cells), and Stage 6 (Day 25; 5,186 cells). Specifically, we compared the expression patterns of the gene markers
627 between the clusters using z-normalized mean expression computed on cells expressing at least 1% of maximal
628 expression for the gene, as described in the reference study [98]. Metadata containing single cell annotations are reported
629 in Supplementary Data 3.

630 *Differentially Expressed Genes*

631 To identify differentially expressed genes for each iPSC-PPC cluster, we used the *FindAllMarkers* function in Seurat
632 [99] with *logfc.threshold = 0.01* and *min.pct = 0.01*. P-values were automatically adjusted by Seurat using Bonferroni
633 correction, and genes with adjusted p-values ≤ 0.05 were considered differentially expressed (Supplementary Data 4).

634 **Bulk RNA-seq**

635 *Library Preparation and Sequencing*

636 RNA was isolated from total-cell lysates using the Quick-RNA™ MiniPrep Kit (Zymo Research) with on-column
637 DNAse treatments. RNA was eluted in 48 µl RNAse-free water and analyzed on a TapeStation (Agilent) to determine
638 sample integrity. All iPSC-PPC samples had RNA integrity number (RIN) values over 9. Illumina TruSeq Stranded
639 mRNA libraries were prepared according to the manufacturer's instructions and sequenced on NovaSeq6000 for 101bp
640 paired-end sequencing. All samples except five were sequenced twice to obtain sufficient number of reads.

641 *Data Processing and Quality Control*

642 FASTQ files were obtained for all 107 iPSC-PPC samples and processed using a similar pipeline described in our
643 previous studies [37,100]. Specifically, RNA-seq reads were aligned with STAR (2.7.3) [101] to the hg19 reference using
644 GENCODE version 34 hg19[96] splice junctions with default alignment parameters and the following adjustments: -
645 *outFilterMultimapNmax 20, -outFilterMismatchNmax 999, -alignIntronMin 20, -alignIntronMax 1000000, -*
646 *alignMatesGapMax 1000000*. BAM files were sorted by coordinates, and duplicate reads were marked using Samtools
647 (1.9.0) [93]. RNA-seq QC metrics were calculated using Samtools (1.9.0) flagstat [93], Samtools (1.9.0) idxstats [93], and
648 Picard (2.20.1) CollectRnaSeqMetrics [102]. Across all 107 iPSC-PPC samples, the total read depth ranged from 32.3 M
649 to 160.4 M (mean = 70.7), the median percentage of intergenic bases was 3.31%, the median percentage of mRNA
650 bases was 92.1%, and the median percentage of duplicate reads was 22.2% (Supplementary Data 2).

651 *Sample Identity*

652 We obtained common bi-allelic and exonic variants from the 1000 Genomes Phase 3 panel [103] with minor allele
653 frequencies between 45% and 55% and predicted their genotypes in the 107 bulk RNA-seq samples using *mpileup* and
654 *call* functions in BCFtools (1.9.0) [104,105]. Then, we used the *genome* command in plink [102] to estimate the identity-by-
655 state (IBS) between each pair of bulk RNA-seq and WGS samples. All RNA-seq samples were correctly matched to
656 the subject with PI_HAT > 0.95 (Supplementary Data 2).

657 *Quantification of gene expression and relative isoform usage*

658 We calculated TPM and estimated relative isoform usage for each gene in each RNA-seq sample using RSEM (version
659 1.2.20) [106] with the following options *–seed 3272015 –estimate-rspd –paired-end –forward-prob*. To identify expressed
660 autosomal genes and isoforms to use for eQTL analyses, we used the same approach previously described [12]. Briefly,

661 autosomal genes were considered expressed if TPM ≥ 1 in at least 10% of samples. To identify expressed isoforms,
662 we required that isoforms had TPM ≥ 1 and usage ≥ 10% in at least 10% of samples and corresponded to expressed
663 genes with at least two expressed isoforms. In total, 16,464 autosomal genes were used for e_gQTL analysis, and 29,871
664 autosomal isoforms corresponding to 9,624 genes were used for e_iQTL analysis. We quantile-normalized TPM and
665 isoform usage across all 107 samples using the *normalize.quantiles* (preprocessCore) and *qnorm* functions in R (version
666 4.2.1) to obtain a mean expression = 0 and standard deviation = 1.

667 *Inferring pseudotime using Monocle*

668 We obtained FASTQ files for 213 iPSCs [35,37] (phs000924), 176 adult whole pancreas [8] (phs000424), and 87 adult
669 islets[38] (GSE50398), and processed the data using the same pipeline described above to obtain TPM counts for each
670 gene per sample. We then used Monocle (http://cole-trapnell-lab.github.io/monocle-release/docs/#constructing-single-
671 cell-trajectories) [107] to infer the pseudotime on all of the RNA-seq samples, including the 107 iPSC-PPCs. Following
672 the standard workflow under "Constructing Single Cell Trajectories" in the Monocle tutorial, we provided TPM counts
673 for all overlapping autosomal expressed genes in the four tissues as input. Then, we identified differentially expressed
674 genes using *differentialGeneTest*, ordered them (*setOrderingFilter*), and performed dimension reduction analysis using
675 *reduceDimension* with *max_components = 2* and *method = "DDRTree"*. Pseudotime was calculated by rooting time
676 (pseudotime = 0) in the 213 iPSC-PPCs using the *GM_state* and *orderCells* functions provided in the tutorial
677 (Supplementary Data 6).

678 *PCA analysis with iPSCs, adult whole pancreas, and adult islets*

679 We obtained TPM counts (described above) for the 213 iPSCs [37], 176 adult whole pancreas [8], 87 adult islets [38], and the
680 107 iPSC-PPCs and performed PCA analysis on the 2,000 most variable genes across the samples using *prcomp* in R
681 (version 4.2.1) with *scale = T* and *center = T*. We observed that the PC clusters corresponded to the iPSCs and each
682 of the three pancreatic tissue types: iPSC-PPC, adult islets, and adult whole pancreas (Supplementary Figure 9,
683 Supplementary Data 6).

684 *Cellular deconvolution*

685 For each of the eight cell types in scRNA-seq, we selected the top 200 most differentially expressed genes that were
686 unique to the cell type (i.e., not expressed in the other cell types). Replicating late PPCs and late PPCs had many
687 overlapping expressed genes so fewer (n = 16 and 164, respectively) were selected. We obtained the average expression
688 of the signature genes for each cell type using *AverageExpression* in Seurat and provided it as input into CIBERSORTx
689 [108] (https://cibersortx.stanford.edu/) along with bulk TPM matrix. Batch correction and quantile normalization were
690 both disabled. We ran CIBERSORTx [108] deconvolution on absolute mode with at least 100 permutations. The predicted
691 fraction of late PPCs and replicating late PPCs were compared to FACS measurements of double-positive PDX-
692 1^+/NKX6-1^+ cells (Supplementary Figure 7B; Supplementary Data 5).

693 **eQTL Analysis**

694 To investigate the effects of genetic variation on gene expression in iPSC-PPCs, we performed an expression
695 quantitative trait loci (eQTL) analysis on gene expression and isoform usage. The eQTLs associated with gene
696 expression were defined as e_gQTLs while those associated with relative isoform usage were defined as e_iQTLs.

697 *Covariates for eQTL Mapping*

698 We included the following as covariates for eQTL mapping of both gene expression and isoform usage: 1) sex; 2)
699 normalized number of RNA-seq reads; 3) percent of reads that mapped to autosome or sex chromosomes (labeled as
700 "uniquely_mapped_reads_canonical_chromosomes" in Supplementary Data 2); 4) percent of reads mapped to
701 mitochondrial chromosome; 5) 20 genotype principal components to account for global ancestry; 6) 20 PEER factors
702 to account for transcriptome variability; and 7) kinship matrix to account for genetic relatedness between samples.
703 Kinship matrix is provided in Supplementary Data 1, and the covariates are available in Supplementary Data 2.

704 Genotype Principal Component Analysis (PCA): Global ancestry was estimated using the genotypes of the 439,461
705 common variants with minor allele frequency (MAF) between 45 and 55% in the 1000 Genomes Phase 3 Panel [103]. We
706 merged the VCF files for the 106 iPSCORE subjects and the 2,504 subjects in the 1000 Genomes [103] and performed a
707 PCA analysis using *plink --pca* [102] (Supplementary Figure 1A). The top 20 principal components were used as
708 covariates in the eQTL model to account for global ancestry and can be found in Supplementary Data 1.

709 PEER Factors: We sought to determine the optimal number of PEER factors to use in the eQTL analysis that will result
710 in maximal eGene discovery. To this end, we initially calculated PEER factors on the 10,000 expressed genes with the
711 largest variance across all samples. To limit biases due to the expression levels of each gene, we divided the 16,464
712 expressed genes into ten deciles based on their average TPM, and selected 50 genes from each decile, for a total of 500
713 genes. We next performed eQTL analysis on each of the 500 genes using 10 to 60 PEER factors in increments of 10.
714 While 30 PEER factors resulted in the highest percentage of eGenes (14.0%), we opted for using 20 PEER factors
715 because the eQTL analysis had a comparable percentage of eGenes (11.8%) to GTEx tissues with similar sample sizes
716 [10] (Supplementary Figure 17). Although we observed variable fraction of double-positive PDX1+/NKX6-1+ cells in
717 the iPSC-PPC samples, we did not include this variable as a covariate because PEER factors 1 and 4 already accounted
718 for this variability (Supplementary Figure 18).

719 Kinship Matrix: The kinship matrix was included as a random effects term to account for the genetic relatedness
720 between individuals in our cohort. We constructed the kinship matrix using the same 439,461 variants employed above
721 using the *–make-rel square* function in plink [102]. The kinship matrix is available in Supplementary Data 1.

722 *eQTL Analysis*

723 We performed eQTL analysis using the same method described in our previous study [12]. For each expressed autosomal
724 gene and isoform, we tested variants that were within 500 Kb of the gene body coordinates using the *bcftools query*
725 function. To account for the genetic relatedness between the samples, we performed eQTL mapping using a linear
726 mixed model with the *scan* function in limix (version 3.0.4) [109] that incorporates the kinship matrix as a random effects
727 term. Specifically, eQTL mapping was implemented through the following model:

728 $$y_i = \beta_{ji} \cdot g_j + \sum_{n=1}^{N} \beta_n \cdot C_n + \mathrm{u} + \epsilon_{ij}$$

729 Where y_i is the normalized expression value for gene i, β_{ji} is the effect size of genotype of SNP j on gene i, g_j is the
730 genotype of SNP j, β_n is the effect size of covariate n, C_n is a vector of values for covariate n, u is the kinship matrix
731 as a random effect, and ϵ is the error term for the association between expression of gene i and genotype of SNP j. As
732 described above, we used the following as covariates: 1) sex, 2) normalized number of RNA-seq reads, 3) percent of
733 reads mapped to autosomal or sex chromosome, 4) percent of reads mapped to mitochondrial chromosome, 5) the top
734 20 genotype PCs (to account to global ancestry), and 6) the top 20 PEER factors (to account for confounders of
735 expression variability), and are available in Supplementary Data 1-2.

736 *FDR Correction*

737 To perform FDR correction, we used a two-step procedure described in Huang et al. [110], which first corrects at the
738 gene level and then at the genome-wide level. First, we performed FDR correction on the p-values of all variants tested
739 for each gene or isoform using eigenMT [109], which considers the LD structure of the variants. Then, we extracted the
740 lead eQTL for each gene or isoform based on the most significant FDR-corrected p-value. If more than one variant had
741 the same FDR-corrected p-value, we selected the one with the largest absolute effect size as the lead eQTL. For the
742 second correction, we performed an FDR-correction on all lead variants using the Benjamini-Hochberg method (q-
743 value) and considered only eQTLs with q-value ≤ 0.01 as significant (Supplementary Data 7).

744 *Conditional eQTLs*

745 To identify additional independent eQTLs (i.e., conditional eQTLs) for each eGene and eIsoform, we performed a step-
746 wise regression analysis in which the genotype of the lead eQTL was included as a covariate in the model and the
747 eQTL mapping procedure (regression and multiple test correction) was re-performed. We repeated this analysis to
748 discover up to five additional associations for each eGene and eIsoform. Conditional eQTLs with q-values ≤ 0.01 were
749 considered significant (Supplementary Data 7).

750 **Functional characterization of iPSC-PPC eQTLs**

751 *Fine-mapping of eQTL Associations*

752 To define a credible set of candidate causal variants for each eQTL association, we performed genetic fine-mapping
753 using the *finemap.abf* function in *coloc* (version 5.1.0, R) [41]. This Bayesian method converts p-values of all variants
754 tested for a specific gene to posterior probabilities (PP) of association for being the causal variant. Variants with PP ≥
755 1% are available in Supplementary Data 8. The eQTLs not present in this table do not have variants with PP ≥ 1%
756 (i.e., all variants were estimated to have PP < 1%).

757 *Genomic enrichments of e_gQTLs and e_iQTLs*

758 For each independent eQTL association, we obtained candidate causal variants whose PP ≥ 5% (Supplementary Data
759 8) and determined their overlap with each of the following genomic annotations using *bedtools intersect*: short splice
760 acceptor sites (± 50bp), long splice acceptor sites (± 100bp), splice donor sites (± 50bp), UTR, intron, exon, intergenic,
761 promoters, and RNA-binding protein binding sites (RBP-BS). RBP-BS were downloaded from a published dataset that
762 utilized enhanced CLIP to identify binding sites of 73 RBPs [111]. We considered only binding sites with irreproducible
763 discovery rate (IDR) threshold of 0.01, indicating that these sites were reproducible across multiple biological samples.
764 Enrichment of candidate causal variants for genomic regions was calculated using a Fisher's Exact Test comparing the
765 proportion of SNPs that overlap each annotation between e_gQTLs and e_iQTLs. P-values were corrected using the
766 Benjamini-Hochberg method and were considered significant if their FDR-corrected p-value ≤ 0.05 (Figure 1E).

767 *Quantification of allele-specific binding of transcription factors using GVATdb*

768 To annotate each candidate causal variant by their effects on transcription factor (TF) binding, we used the Genetic
769 Variants Allelic TF Binding Database (GVATdb) to estimate the TF binding impact score associated with each variant
770 and each of the 58 iPSC-PPC-expressed TF available on the database and with a AUPRC > 0.75 indicating a high-
771 confidence deltaSVM model. We estimated the score using the instructions and reference files provided on the
772 GVATdb GitHub repository (https://github.com/ren-lab/deltaSVM). The software required a list of SNPs as input
773 along with hg19 reference files provided in the GVATdb repository. The output provides the deltaSVM score [112] for
774 each variant-TF pair (Supplementary Data 9), indicating whether the variant results in a promotion ("Gain"), disruption
775 ("Loss"), or no change ("None") in TF binding.

776 *Correlation between eQTL effect size and binding affinity of transcription factors*

777 To determine whether e_gQTLs were more likely to affect TF binding compared to e_iQTLs, we performed a Spearman
778 Correlation Analysis between deltaSVM score and eQTL effect size on candidate causal variants with PP ≥ 10%, 20%,
779 40%, 60% and 80%. We considered nominal p-value ≤ 0.05 as significant.

780 *Colocalization between iPSC-PPC gene and isoform eQTLs*

781 To determine the overlap of genetic variants between e_gQTLs and e_iQTLs for the same gene, we performed Bayesian
782 colocalization using the *coloc.abf* function in *coloc* (version 5.1.0, R) [41], where each pair of signals was given a
783 summary PP that each of the following five hypotheses was true: H0) no association was detected in both signals, H1)
784 an association was detected only in signal 1, H2) an association was detected only in signal 2, H3) an association was
785 detected in both signals but the underlying causal variants are different, and H4) an association was detected for both
786 signals and the underlying causal variants are the same. We filtered the results by requiring that each colocalization
787 used the number of overlapping variants (called "nsnps" in the *coloc.abf* output) ≥ 500. We considered two eQTL
788 signals to be shared if the PP for H4 (called "PP.H4.abf" in *coloc.abf* output; hereafter referred to as PP.H4) ≥ 80%.
789 Conversely, two signals were considered distinct if the PP for H3 (called "PP.H3.abf" in *coloc.abf* output; hereafter
790 referred to as PP.H3) ≥ 80%. eQTL associations with PP.H4 < 80% and PP.H3 < 80% were due to insufficient power
791 in one or both eQTL signals. As input into *coloc.abf*, we provided p-values, minor allele frequency, and sample size.
792 All associations with PP ≥ 80% for any model are available in Supplementary Data 11.

793 *Genomic enrichment of overlapping e_gQTL and e_iQTL signals compared to non-overlapping*

794 To test the enrichment of overlapping e_gQTLs and e_iQTLs in genomic regions compared to non-overlapping signals,
795 we determined the overlap of candidate causal variants with PP ≥ 1% in each genomic annotation using *bedtools*
796 *intersect* and compared the proportion of variants overlapping each annotation against a background set of 20,000
797 random variants using a Fisher's Exact Test as previously described [10]. For overlapping eQTLs, we used the candidate
798 causal variants predicted in the *coloc.abf* output. Enrichments with nominal p-value < 0.05 were considered significant
799 (Supplementary Figure 10).

800 **Downloading eQTL summary statistics for adult pancreatic tissues**

801 We downloaded complete eQTL summary statistics for gene and exon associations for 420 adult human islets from the
802 InSPIRE Consortium (https://zenodo.org/record/3408356) [11], and gene and splicing associations for 305 adult whole
803 pancreas from the GTEx Data Portal for GTEx Analysis version 8 [10]
804 (https://console.cloud.google.com/storage/browser/gtex-resources). All GTEx SNPs were converted to hg19 using the
805 UCSC liftOver Bioconductor package in R (https://www.bioconductor.org/help/workflows/liftOver/). Lead SNPs for
806 conditional associations in the adult islets and whole pancreas datasets were downloaded from their respective studies
807 (complete statistics were not readily available).

808 Due to the different types of eQTLs used in this study that are associated with changes in alternative splicing (e_iQTLs,
809 exon eQTLs, and sQTLs), hereafter we refer to this collective unit as "e_{AS}QTLs".

810 **Comparing eGenes between fetal-like iPSC-PPC and adult islets**

811 To identify eGenes that were shared between iPSC-PPC and adult islet tissues, we compared the 4,065 eGenes in iPSC-
812 PPC and the 4,211 eGenes in adult islets that complete summary statistics were available for. Specifically, we used the
813 *intersect* function in R to identify eGenes that overlapped between the two tissues and *setdiff* function in R to identify
814 eGenes that did not overlap. Similarly, using the *intersect* function in R, we compared the 22,266 expressed genes in
815 adult islet tissues with the 4,065 eGenes in iPSC-PPC to identify the proportion of iPSC-PPC eGenes that were
816 expressed in adult islets, and vice versa with the 17,098 expressed genes in iPSC-PPC and 4,211 eGenes in adult islets.
817 The 22,266 expressed genes in adult islet tissues were obtained from the complete summary statistics uploaded by the
818 previous study in https://zenodo.org/record/3408356.

819 **Comparing eQTLs present in fetal-like iPSC-PPC and adult pancreatic tissues**

820 *Colocalization between iPSC-PPC and adult eQTLs*

821 To identify eQTLs whose effects were driven by the same causal signals in iPSC-PPC and adult pancreatic tissues
822 (islets and whole pancreas), we performed Bayesian colocalization using the *coloc.abf* function in *coloc* (version 5.1.0,
823 R) [41]. Specifically, for each iPSC-PPC and adult eQTL, we tested its overlap with nearby eQTLs within 3 Mb from the
824 gene body coordinates. eQTLs with no overlapping variants would automatically not be tested. Then, we filtered the
825 results by requiring that each colocalization used the number of overlapping variants (called "nsnps" in the *coloc.abf*
826 output) ≥ 500. As described above, we considered two eQTL signals to be shared if PP.H4 ≥ 80% or distinct if PP.H3
827 ≥ 80%. eQTL associations with PP.H4 < 80% and PP.H3 < 80% were due to insufficient power in one or both eQTL
828 signals.

829 Because we, and others, have shown that e_gQTLs are functionally different from e_{AS}QTLs (e_iQTLs, exon eQTLs, and
830 splicing eQTLs), we performed colocalization for e_gQTLs and e_{AS}QTLs independently (i.e., colocalization of e_gQTL
831 was performed only with another e_gQTL and an e_{AS}QTL only with another e_{AS}QTL). All associations with PP ≥ 80%
832 for any model are reported in Supplementary Data 11.

833 *Fine-mapping of adult eQTL associations*

834 Similarly for iPSC-PPC eQTLs, we identified candidate causal variants using the *finemap.abf* function in *coloc* (version
835 5.1.0, R). This Bayesian method converts p-values of all variants tested for a specific gene to a PP value for being the
836 causal variant. Variants with PP ≥ 1% are available in Supplementary Data 8. The eQTLs not present in this table did
837 not have any variants with PP ≥ 1% (i.e., all variants were estimated to have PP < 1%).

838 **For all downstream analyses beyond this point, we used only iPSC-PPC, adult islets, and adult whole pancreas**
839 **eQTLs that had at least one candidate causal variant with PP ≥ 1%, were outside of the MHC region, and were**

840 **annotated in GENCODE version 34 hg19, to ensure that our analyses were sufficiently powered and the multiple**
841 **datasets were comparable.**

842 *Identifying tissue-unique singleton eQTLs*

843 To identify tissue-unique singleton eQTLs, we obtained all eQTLs that did not colocalize with another eQTL and
844 examined their LD with all other eQTLs of the same phenotype (e_gQTLs or e_{AS}QTLs) using their most likely candidate
845 causal variants based on the highest PP from fine-mapping (*finemap.abf*). If the candidate causal variant was not
846 genotyped in the 1000 Genomes Phase 3 panel, then we used the next top candidate causal variant. We repeated this
847 process until we found a variant that was in the 1000 Genomes or no more variants remained with causal PP ≥ 1%.
848 Because complete summary statistics were not available for the adult conditional eQTLs, we used their lead variants
849 publicly available from their respective studies to account for the presence of multiple causal variants in the genomic
850 region. LD was calculated using *plink --r2 square --keep-allele-order --make-bed* [102] and the 1000 Genomes Phase 3
851 panel [103]. We considered two eQTLs to be in LD if their candidate causal variants were within 500 Kb and had $r^2 \geq$
852 0.2. If LD could not be measured, because one of the variants was not genotyped in the 1000 Genomes, then we used
853 distance as a metric for LD, where if the variants were within 500 Kb of each other, we considered them to be in LD.
854 Singleton eQTLs that were found to be in LD with another eQTL (regardless of tissue) were re-annotated as
855 "ambiguous" and excluded from downstream analyses. Otherwise, we kept their annotations as tissue-unique
856 singletons. All annotations for singleton eQTLs are reported in Supplementary Data 12.

857 *Identifying eQTL modules*

858 eQTL modules were identified by first creating a network using the *graph_from_data_frame* function in igraph
859 (version 1.3.4, R) [113] where the input was a data frame containing all pairs of colocalized eQTLs (nsnps ≥ 500 and
860 PP.H4 ≥ 80%) as binary edges. We created networks for each chromosome and phenotype (gene expression and
861 alternatively splicing) independently, totaling to 44 networks (22 chromosomes x 2 phenotypes = 44 networks). Then,
862 we performed community detection analysis using the *cluster_leiden* function with *--objective_function =*
863 *"modularity"*, *n_iterations = 500, resolution = 0.3* to identify modules of eQTLs. Upon examining them in depth, we
864 observed that 5% of the modules contained at least one H3 association (PP.H3 ≥ 80%) between a pair of eQTLs,
865 indicating that signals within a module were predicted to have distinct genetic variants despite being assigned to the
866 same module. Therefore, to filter for modules that contained eQTLs likely to share the same causal variants, we
867 required that at least 30% of all eQTL pairs had a H4 association and that the number of H4 "edges" was twice the
868 number of H3 "edges" (number of H4 edges / number of H3 edges ≥ 2). For example, a module with four eQTLs would
869 have six possible pairwise combinations, and to be considered a validated module, we required at least two H4 edges
870 and no more than one H3 edge. Modules that did not pass these thresholds were annotated as "module_failed" and
871 excluded from downstream analyses. Summary of eQTL modules and their individual eQTL associations are reported
872 in Supplementary Data 13. Module IDs were assigned such that the first term indicates the phenotype the module was

873 associated with ("GE" for gene expression or "AS" for alternative splicing), the second term indicates the chromosome
874 number, and the third term indicates a unique integer. For example, "GE_1_32" indicates that this module is associated
875 with changes in gene expression, located in in chromosome 1, and assigned the number 32.

876 *Identifying tissue-unique and tissue-sharing eQTL modules*

877 Combinatorial eQTLs were defined in this study as an eQTL having at least one H4 association (PP.H4 ≥ 80%) with
878 another eQTL either in the same or different tissue. These combinatorial eQTLs were then connected to form a module,
879 which we identified using the network analysis described above. We then categorized each module based on the activity
880 of eQTLs in the three pancreatic tissues, having a total of seven module categories (Figure 3B):

881 1) iPSC-PPC-unique: contains eQTLs in **only** iPSC-PPC
882 2) Adult islet-unique: contains eQTLs in **only** adult islets
883 3) Adult whole pancreas-unique: contains eQTLs in **only** adult whole pancreas
884 4) Adult-shared: contains eQTLs in adult islets **and** adult whole pancreas
885 5) Fetal-islet: contains eQTLs in iPSC-PPC **and** adult islets
886 6) Fetal-whole-pancreas: contains eQTLs in iPSC-PPC **and** adult whole pancreas
887 7) Fetal-adult: contains eQTLs in **all** three pancreatic tissues

888 We next examined the eQTLs modules for LD with eQTLs in other tissues to confirm tissue specificity. Similar to the
889 analysis described above for identifying tissue-unique singletons, we calculated LD using *plink --r2 square --keep-*
890 *allele-order --make-bed* [102] and the 1000 Genomes Phase 3 panel [103] between the eQTLs' most likely candidate causal
891 variants (based on the highest PP; PP ≥ 1%). We considered two eQTLs to be in LD if they had an $r^2 \geq 0.2$ and were
892 within 500 Kb of each other. If LD could not be calculated because candidate causal variants were not genotyped in
893 the 1000 Genomes Phase 3 panel, then we used distance as a metric for LD and considered two eQTLs to be in LD if
894 their candidate causal variants were within 500 Kb. To account for the presence of multiple causal variants in the
895 genomic region, we included the lead variants from the adult islet and whole pancreas conditional eQTLs in the LD
896 comparisons to prevent misclassification of tissue-unique eQTLs.

897 For each of the module categories, we required that the following were true to be considered for downstream analyses:

898 1) iPSC-PPC-unique: contains eQTLs in only iPSC-PPC, and **all** eQTLs were not in LD with eQTLs in adult
899 islets **and** adult whole pancreas
900 2) Adult islet-unique: contains eQTLs in only adult islets, and **all** eQTLs were not in LD with eQTLs in adult
901 whole pancreas **and** iPSC-PPC
902 3) Adult whole pancreas-unique: contains eQTLs in only adult whole pancreas, and **all** eQTLs were not in LD
903 with eQTLs in adult islets **and** iPSC-PPC

904 4) Adult-shared: contains eQTLs in only adult islets and adult whole pancreas, and **all** eQTLs were not in LD
905 with eQTLs in iPSC-PPC
906 5) Fetal-islet: contains eQTLs in iPSC-PPC and adult islets, and **all** eQTLs were not in LD with eQTLs in adult
907 whole pancreas
908 6) Fetal-whole-pancreas: contains eQTLs in iPSC-PPC and adult whole pancreas, and **all** eQTLs were not in LD
909 with eQTLs in adult islets
910 7) Fetal-adult: contains eQTLs in **all** three pancreatic tissues.

911 For any module that did not meet the above requirements, we annotated the eQTLs in the module "ambiguous" and
912 excluded for downstream analysis. Hereafter, we refer the eQTL associations in tissue-unique modules (categories 1-
913 3) as tissue-unique combinatorial eQTLs and those in categories 5-7 as eQTLs shared between both fetal-like and adult
914 stages. All annotations for eQTL modules and their individual eQTLs are reported in Supplementary Data 12 and
915 Supplementary Data 13.

916 Enrichment of fetal-like iPSC-PPC-unique singleton and combinatorial eQTLs in chromatin states

917 We obtained chromatin state maps for human embryonic stem cell-derived pancreatic progenitor cells and adult islets
918 from previously published studies [14,114]. Because e_gQTLs were likely to affect non-coding regulatory elements (Figure
919 1E), we only considered them in this analysis and excluded e_{AS}QTLs. Enrichments for iPSC-PPC-unique singleton and
920 combinatorial e_gQTLs were calculated using a Fisher's Exact Test by comparing the proportion of fine-mapped variants
921 (from *finemap.abf*) of the e_gQTLs at different thresholds of PP from 0-0.8 at 0.1 intervals to a background set of 20,000
922 randomly selected variants. Enrichments were Benjamini-Hochberg-corrected. Corrected p-values $\leq$ 0.05 were
923 considered significant. Enrichment results are available in Supplementary Data 14, Figure 3C.

924 Overlap of eGenes in shared modules between fetal-like iPSC-PPC and adult pancreatic tissues

925 For the modules shared between both fetal-like and the two adult pancreatic tissues (categories 5-7; described above),
926 we compared the eGenes associated with: 1) iPSC-PPC e_gQTLs versus adult islet e_gQTLs; and 2) iPSC-PPC e_gQTLs
927 versus adult whole pancreas e_gQTLs. For e_{AS}QTLs, we compared the genes mapping to: 1) each isoform in iPSC-PPC
928 versus exon in adult islets; and 2) each isoform in iPSC-PPC versus splice interval in adult whole pancreas. From these
929 comparisons, we assigned each module an "islet_egene_overlap" label and an "whole_pancreas_egene_overlap" label
930 in Supplementary Data 13 (also shown in Figure 4A and Supplementary Figure 11C), where "zero" indicates that the
931 module did not contain an eQTL in the adult tissue, "same" indicates that the module contained eQTLs corresponding
932 to only the same eGenes in iPSC-PPC and adult, "partial" indicates that the module contained eQTLs corresponding
933 with partially overlapping eGenes between iPSC-PPC and adult, and "different" indicates that the module contained
934 eQTLs corresponding to only different genes. For example, if a module was annotated with "zero" for

935 islet_egene_overlap and "same" for whole_pancreas_egene_overlap, this meant that the module did not contain an
936 eQTL from adult islet and had only eQTLs associated with the same eGenes between iPSC-PPC and adult whole
937 pancreas. These annotations also meant that this module was in the "fetal-whole-pancreas" category (i.e, only contained
938 eQTLs from iPSC-PPC and adult whole pancreas).

939 **Complex Trait GWAS Associations**

940 *Colocalization of eQTLs with GWAS associations*

941 We obtained GWAS summary statistics from ten different studies: 1) type 1 diabetes [3], 2) type 2 diabetes [114], 3) body
942 mass index [50], 4) triglycerides [50], 5) HDL cholesterol [50], 6) LDL direct [50], 7) cholesterol [50], 8) glycated hemoglobin
943 A1C (HbA1c) levels from the MAGIC Consortium [115], 9) HbA1c levels from the Pan-UKBB Study [50], and 10) fasting
944 glucose [115]. All of the data, except for type 1 diabetes, were provided in hg19 coordinates, therefore we converted the
945 coordinates from hg38 to hg19 using the liftOver package in R [116]. We sorted and indexed each file using *tabix* [93]. For
946 each trait, we performed colocalization between GWAS variants and all filtered significant eQTLs (see bolded section
947 above) in the three pancreatic tissues with the *coloc.abf* function in *coloc* (version 5.1.0, R) [41] using p-values, MAF,
948 and sample size as inputs. Then, we filtered results based on whether the lead candidate causal variant underlying both
949 GWAS and eQTL association (from *coloc.abf* output) is genome-wide significant for GWAS association (p-value ≤
950 $5x10^{-8}$) and the number of overlapping variants used to test for colocalization (nsnps) ≥ 500. eQTLs were considered
951 to share a genetic signal with GWAS if PP.H4 ≥ 80% or have distinct signals with GWAS if PP.H3 ≥ 80%. For eQTL
952 modules, we required that at least 30% of the eQTLs in the module colocalized with GWAS (PP.H4 ≥ 80%) and that
953 the number of H4 associations is twice the number of H3 associations (number of H4 associations / number of H3
954 associations ≥ 2). Colocalization results for the 312 GWAS loci (183 singleton and 129 module) with PP.H4 ≥ 80%
955 are available in Supplementary Data 15.

956 *GWAS 99% Credible Sets*

957 For each GWAS locus (based on GWAS locus ID in Supplementary Data 15), we constructed 99% credible sets with
958 the predicted candidate causal variants underlying both eQTL and GWAS associations (from *coloc.abf* output). If the
959 GWAS locus colocalized with a singleton eQTL, the credible sets were constructed using the output of the eQTL's
960 colocalization with GWAS. If the GWAS locus colocalized with an eQTL module, we constructed credible sets for
961 each of the pairwise eQTL-GWAS colocalization and retained the eQTL that resulted in the least number of candidate
962 causal variants. If multiple eQTLs had the same number of variants in their credible set, we considered the eQTL with
963 the highest PP.H4 for GWAS colocalization. 99% credible sets were constructed by first sorting the variants by
964 descending order of causal PP and obtaining the least number of variants that resulted in a cumulative PP ≥ 99%. 99%
965 credible sets for each of the 312 GWAS loci (183 singleton and 129 module) are reported in Supplementary Data 16.

966 *LD with non-pancreatic GTEx tissues*

967 We downloaded the lead SNPs for all significant e_gQTLs (including their conditionals) for the 48 non-pancreatic tissues
968 in the GTEx dataset version 8 [10] and converted their genomic positions to hg19 using UCSC liftOver [116]. We then
969 calculated their LD with the lead SNPs of the 16 iPSC-PPC-unique e_gQTLs that colocalized with GWAS
970 (Supplementary Data 17) using *plink* [102] *--tag-kb 500 --tag-r2 0.2 --show-tags all* and the 1000 Genomes [103] as the
971 reference panel. We considered two eQTLs to be in LD if their lead SNPs were within 500 Kb and had $r^2 > 0.2$. If LD
972 could not be calculated because the SNP was not genotyped in the reference panel, we used distance as a metric in
973 which we considered two eQTLs to be in LD if their lead SNPs were within 500 Kb.

974 **Data Availability**

975 The iPSC-PPC scRNA-seq and bulk RNA-seq data generated in this study have been deposited in the GEO database
976 under accession codes GSE152610 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE152610] and
977 GSE182758 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE182758], respectively. The WGS data used in
978 this study for iPSCORE individuals were obtained as a VCF file from phs001325.v3. The reference gene annotation
979 file for aligning bulk RNA-seq data of iPSC-PPC were obtained from GENCODE release version 34 in GRCh37 as a
980 GTF file [https://www.gencodegenes.org/human/release_34.html]. The bulk RNA-seq data for iPSC, adult islet, and
981 adult whole pancreas samples used in PCA and pseudotime analyses were obtained from phs000924, GSE50398, and
982 phs000424, respectively. eQTL summary statistics for adult whole pancreas and islet samples were obtained from the
983 GTEx Data Repository [https://console.cloud.google.com/storage/browser/gtex-resources] and a previously published
984 study [11] [https://zenodo.org/record/3408356], respectively. GWAS summary statistics were obtained from the Pan UK
985 BioBank resource [https://pan.ukbb.broadinstitute.org/], the MAGIC (Meta-Analyses of Glucose and Insulin-related
986 traits) Consortium [https://magicinvestigators.org/downloads/; https://doi.org/10.1038/s41588-021-00852-9] the
987 DIAMANTE Consortium [https://diagram-consortium.org/downloads.html; http://doi.org/10.1038/s41588-018-0241-
988 6], and a previously published study [3] [http://ftp.ebi.ac.uk/pub/databases/gwas/summary_statistics/GCST90014001-
989 GCST90015000/GCST90014023/]. Full summary statistics for iPSC-PPC eQTLs, supplemental data, and processed
990 scRNA-seq have been deposited in Figshare: https://figshare.com/projects/Large-scale_eQTL_analysis_of_iPSC-
991 PPC/156987.

992 **Code Availability**

993 Scripts for processing RNA-seq and scRNA-seq data and performing downstream analyses are publicly available at
994 /https://github.com/jenniferngp/iPSC_PPC_eQTL_Project (version 1.0.0 of the release).

995 ## Acknowledgements

996 This work was supported by the National Library Training Grant T15LM011271 (J.P.N., M.K.R.D., T.D.A.) and the
997 National Institute of Diabetes and Digestive and Kidney Disease (NIDDK) F31DK131867 (J.P.N.), U01DK105541,
998 DP3DK112155 and P30DK063491 (K.A.F.). Additional support was also received from the National Heart, Lung and
999 Blood Institute (NHLBI) F31HL158198 (T.D.A.) and the Dongguk University Research Fund of 2023 (J.H.K.). We
1000 thank Drs. Maike Sanders and Kyle Gaulton for their advice on experimental design and analyses. This publication
1001 includes data generated at the UC San Diego IGM Genomics Center utilizing an Illumina NovaSeq 6000 that was
1002 purchased with funding from a National Institutes of Health SIG grant S10OD026929.

1003 ## Author information

1004 ## Consortia

1005 **iPSCORE Consortium**

1006 Lana Ribeiro Aguiar[4], Angelo D. Arias[6], Timothy D. Arthur[2,3], Paola Benaglio[6], W. Travis Berggren[7], Juan Carlos
1007 Izpisua Belmonte[8], Victor Borja[4], Megan Cook[4], Matteo D'Antonio[2,4,6], Agnieszka D'Antonio-Chronowska[6],
1008 Christopher DeBoever[1], Kenneth E. Diffenderfer[7], Margaret K.R. Donovan[1,2], KathyJean Farnam[4], Kelly A. Frazer[4,6],
1009 Kyohei Fujita[6], Melvin Garcia[4], Olivier Harismendy[2], Benjamin A. Henson[4], David Jakubosky[2,3], Kristen Jepsen[4], He
1010 Li[6], Hiroko Matsui[4], Naoki Nariai[6], Jennifer P. Nguyen[1,2], Daniel T. O'Connor[9], Jonathan Okubo[4], Athanasia D.
1011 Panopoulos[8], Fengwen Rao[9], Joaquin Reyna[4], Bianca M. Salgado[4], Nayara Silva[6], Erin N. Smith[6], Josh Sohmer[4], Shawn
1012 Yost[1], William W. Young Greenwald[1]

1013 [7] Stem Cell Core, Salk Institute for Biological Studies, La Jolla, CA 92037, USA

1014 [8] Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, CA 92037, USA

1015 [9] Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA

1016 ## Contributions

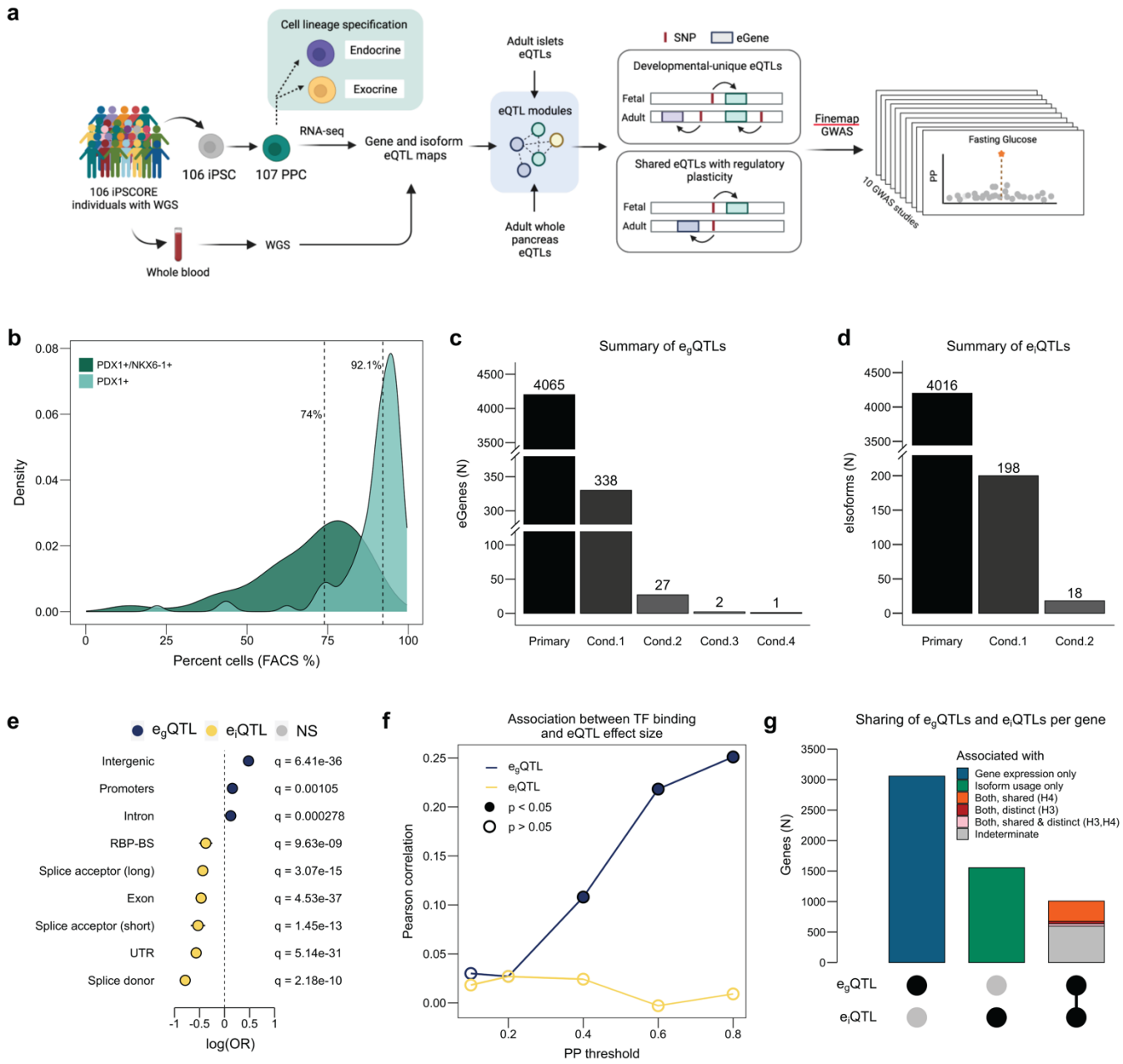
1017 K.A.F. and iPSCORE consortium members conceived the study. A.D.C., B.M.S., K.F., and iPSCORE consortium
1018 members performed the differentiations and generated molecular data. J.P.N., M.K.R.D. and H.M. performed quality
1019 check on scRNA-seq and RNA-seq samples. J.P.N., T.D.A., J.H.K. performed the computational analyses. K.A.F. and
1020 iPSCORE consortium members oversaw the study. J.P.N., M.D. and K.A.F. prepared the manuscript.

1021 # Ethics declaration

1022 ## Competing interests

1023 The authors declare no competing interests.

Figure 1. Discovery and Characterization of eQTLs in iPSC-PPC

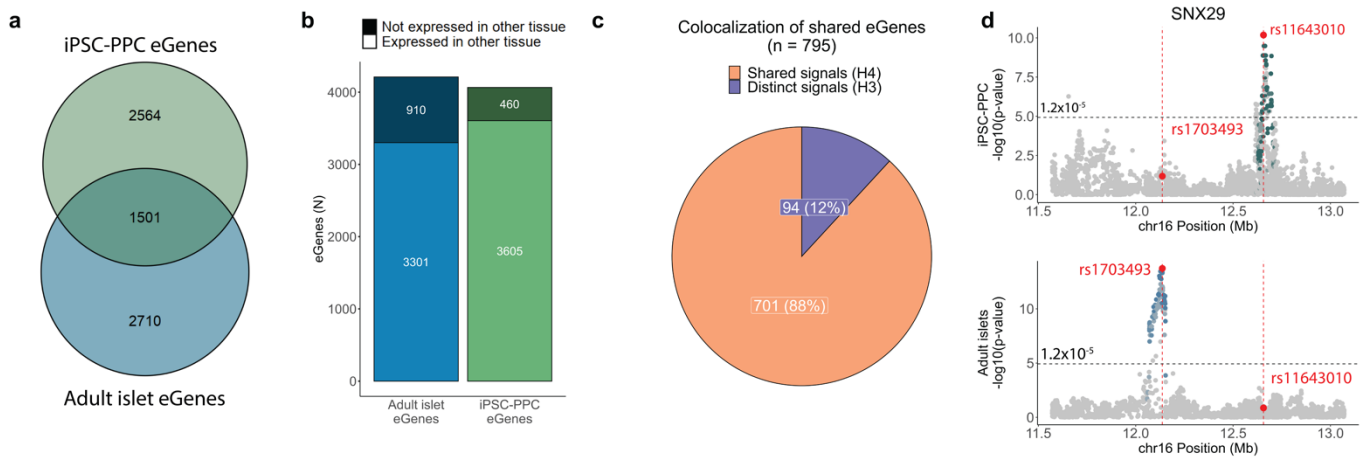


(**a**) Study overview created with Biorender.com. (**b**) Density plots showing the distribution of PDX1^{+} cells (%; regardless of NKX6-1 status; light green) and PDX1^{+}/NKX6-1^{+} cells (%; dark green) in the 107 iPSC-PPC samples. (**c**) Bar plot showing the number of eGenes with primary and conditional e_gQTLs. (**d**) Bar plot showing the number of eIsoforms with primary and conditional e_iQTLs. (**e**) Enrichment (odds ratio, X-axis) of eQTLs in genomic regions (Y-axis) using two-sided Fisher's Exact Tests comparing the proportion of variants with causal posterior probability (PP) ≥ 5% in the genomic regions between e_gQTLs (blue; n=8,763) and e_iQTLs (yellow; n=8,919). (**f**) Line plot showing Pearson correlations of TF binding score and eQTL effect size at different thresholds of causal PP for e_gQTLs (blue)

1033 and e_iQTLs (yellow) (Supplementary Data S10). Closed points indicate significant correlations (nominal $p<0.05$) while
1034 open points indicate non-significant correlations (nominal ($p>0.05$). (**g**) Bar plot showing the number of genes with
1035 only e_gQTLs (blue; n=3,057), only e_iQTLs (green; n=1,554), or both. Orange represents genes whose e_gQTLs
1036 colocalized with all their corresponding e_iQTLs (PP.H4 ≥ 80%; n=333). Red represents genes whose e_gQTLs did not
1037 colocalize with any of their corresponding e_iQTLs (PP.H3 ≥ 80%; n=38), and pink represents genes with both shared
1038 and distinct e_gQTLs and e_iQTLs (i.e., an eGene with two eIsoforms may colocalize with one eIsoform but not the other)
1039 (n=39). Gray represents genes whose eQTL signals were not sufficiently powered to test for colocalization (PP.H4 <
1040 80% and PP.H3 < 80%; n=598).

1041 **Figure 2. Comparison of the genetic architecture underlying gene expression between**

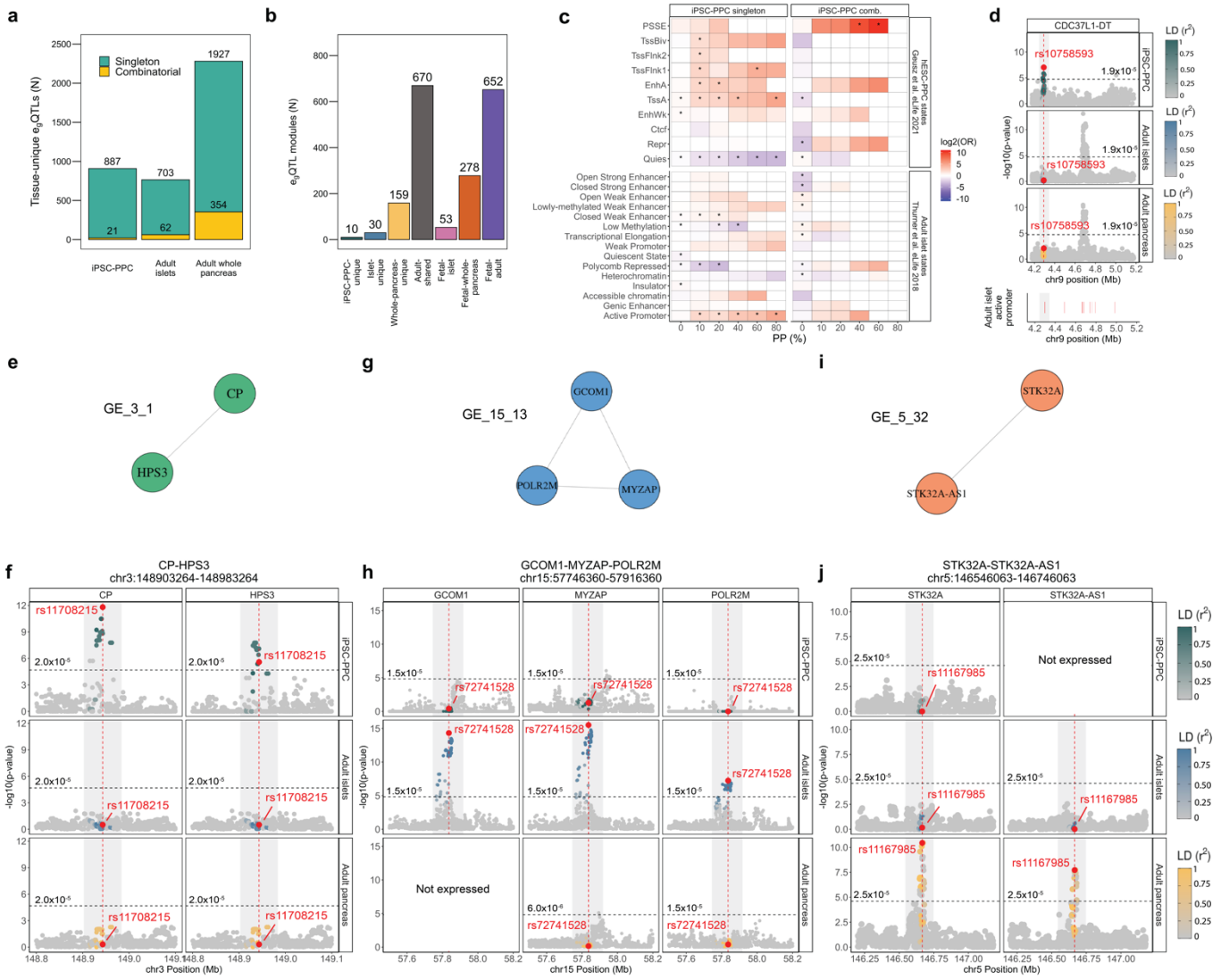
1042 **fetal-like and adult islets**



1043

1044 (**a**) Venn diagram showing the overlap of eGenes between fetal-like iPSC-PPC and adult islets. (**b**) Stacked bar plot
1045 showing the total number of eGenes detected in adult islets (blue; n=4,211 total) that were expressed in iPSC-PPC
1046 (light blue; n=3,301). Likewise, we show the total number of fetal-like iPSC-PPC eGenes (green; n=4,065 total) that
1047 were expressed in adult islets (light green; n=3,605). These results show that the majority of eGenes were expressed in
1048 both tissues, however, a large fraction was influenced by genetic variation in only one of the two tissues. Therefore,
1049 the small overlap of eGenes may be due to differences in the genetic regulatory landscape. (**c**) Pie chart showing the
1050 proportion of shared eGenes with distinct genetic loci (PP.H3 ≥ 80%, purple) or shared genetic loci (PP.H4 ≥ 80%,
1051 orange). These results show that 12% of the shared eGenes were associated with distinct regulatory variants between
1052 fetal-like and adult pancreatic stages. (**d**) Example of a shared eGene (*SNX29*) whose expression was associated with
1053 distinct e_gQTL signals (PP.H3 = 90.4%) in fetal-like iPSC-PPC (green, top panel) and adult islets (blue, bottom panel).
1054 The X-axis represents variant positions while the Y-axis shows the -log10(eQTL p-value) for the associations between
1055 the genotype of the tested variants and gene expression. For plotting purposes, we assign a single p-value for gene-
1056 level significance after Bonferroni-correction (0.05 / number of independent variants tested in fetal-like iPSC-PPC;
1057 horizontal line). Red vertical lines show the positions of the lead variants in fetal-like iPSC-PPC and adult islets
1058 (chr16:12656135:C>G and chr16:12136526:A>G, respectively).

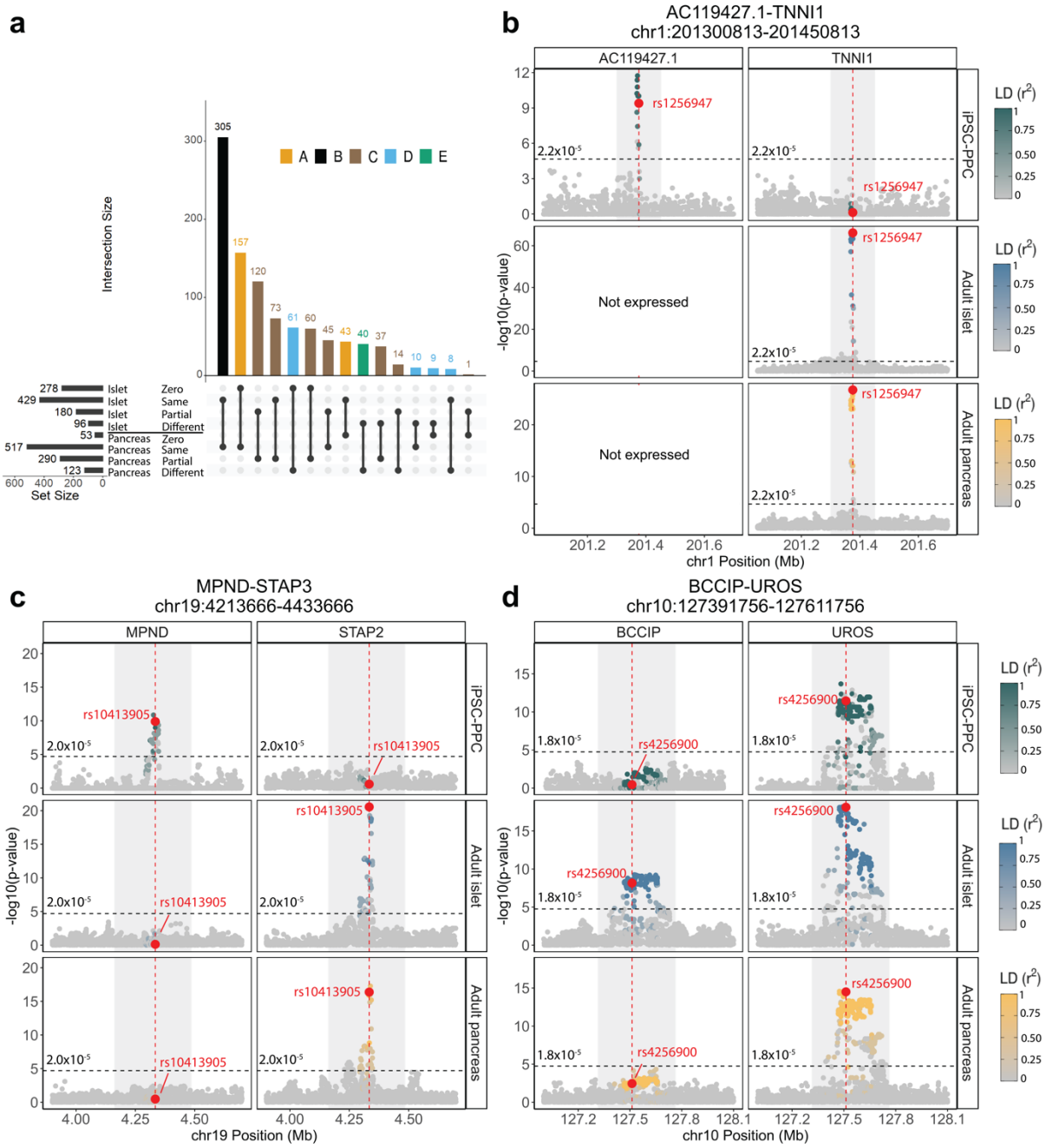
Figure 3. eQTL sharing between iPSC-PPC, adult islets, and adult whole pancreas



(**a**) Bar plot showing the number of tissue-unique e_gQTLs identified in fetal-like iPSC-PPC, adult islets and adult whole pancreas. (**b**) Bar plot showing the number of e_gQTL modules for each annotation. (**c**) Top panels: Enrichment (odds ratio) of iPSC-PPC singleton and combinatorial e_gQTLs in hESC-derived PPC chromatin states [14]. Bottom panels: Enrichment (odds ratio) of iPSC-PPC singleton and combinatorial e_gQTLs in adult islet chromatin states [48]. Enrichment was calculated using a two-sided Fisher's Exact Test comparing the proportion of candidate causal variants overlapping the chromatin states versus a background of randomly selected 20,000 variants at various PP thresholds. P-values were Benjamini-Hochberg-corrected and considered significant if the corrected p-values < 0.05 (indicated by asterisk, Supplementary Data 14). (**d**) *CDC37L1-DT* locus showing an iPSC-PPC-unique singleton e_gQTL overlapping an adult islet active promoter region. Lower panel shows the positions of active promoters in the adult islets. (**e-f**) The chr3:148903264-148983264 locus (gray rectangle) is an example of an "iPSC-PPC-unique" module (module ID:

1071 GE_3_1) associated with *CP* and *HPS3* expression. (**g-h**) The chr15:57746360-57916360 locus (gray rectangle) is as
1072 an example of an "adult islet-unique" module (module ID: GE_15_13) associated with *GCOM1*, *MYZAP*, and
1073 *POLR2M* expression. *GCOM1* was not expressed in adult whole pancreas and therefore, was not tested for e_gQTL
1074 association. (**i-j**) The chr5:146546063-146746063 locus (gray rectangle) is an "adult whole pancreas-unique" e_gQTL
1075 module (module ID: GE_5_32) associated with *STK32A* and *STK32A-AS1* expression. *STK32A-AS1* was not expressed
1076 in iPSC-PPC and therefore, was not tested for e_gQTL association. Panels **e, g, i** display the e_gQTL modules as networks
1077 in which the e_gQTL associations (nodes) are connected by edges due to colocalization (PP.H4 ≥ 80%). For panels **d,**
1078 **f, h, and j,** the X-axis represents variant positions while the Y-axis shows the -log10(eQTL p-value) for the associations
1079 between the genotype of the tested variants and gene expression. For plotting purposes, we assigned a single p-value
1080 for gene-level significance after Bonferroni-correction (0.05 / the number of independent variants tested in fetal-like
1081 iPSC-PPC; horizontal line). Red vertical lines indicate the positions of the lead candidate causal variants underlying
1082 the colocalization based on maximum PP.

1083 **Figure 4. Regulatory plasticity of e_gQTL loci**

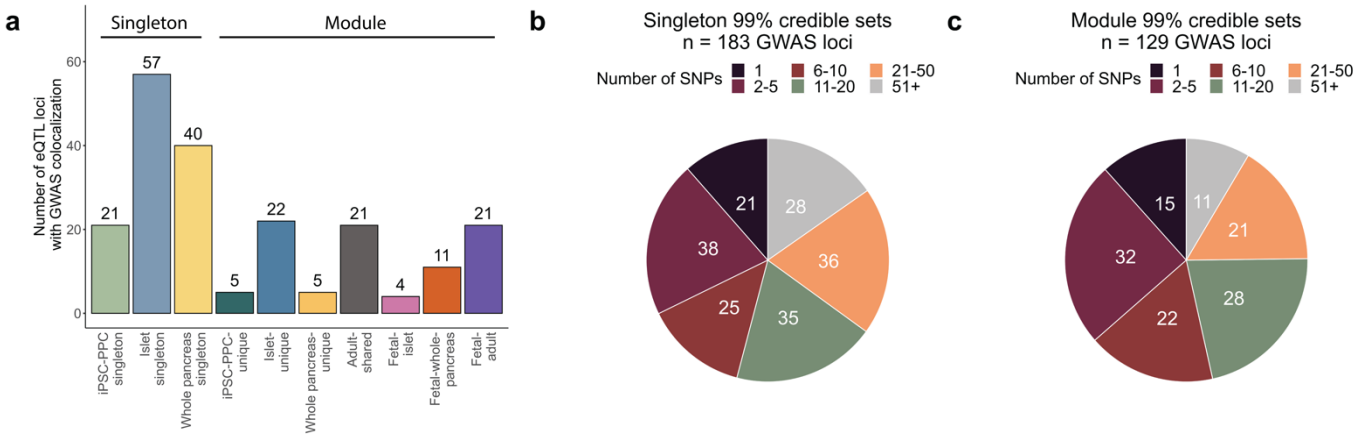


1084

1085 (**a**) Number of e_gQTL modules shared between iPSC-PPC and at least one adult pancreas tissue categorized by eGene
1086 overlap with adult. "Zero" indicates that the module did not contain an e_gQTL in the respective adult tissue. "Same"
1087 indicates that the module had e_gQTLs for only the same eGenes in iPSC-PPC and the adult tissue. "Partial" indicates
1088 that the module had e_gQTLs for partially overlapping eGenes between iPSC-PPC and the adult tissue. "Different"
1089 indicates that the module had e_gQTLs for only different eGenes between iPSC-PPC and the adult tissue. (**b-d**) Examples
1090 of e_gQTL loci demonstrating regulatory plasticity of genetic variation across fetal-like and adult pancreatic stages.

1091 Panel **b** shows a locus strongly associated with *AC119427.1* expression in fetal-like iPSC-PPC and *TNNI1* expression
1092 in adult islet and whole pancreas. Panel **c** shows a locus associated with *MPND* expression in only fetal-like iPSC-PPC
1093 but *STAP2* expression in both the adult pancreatic tissues. Panel **d** shows a locus associated with partially overlapping
1094 eGenes between the two pancreatic stages (*UROS* in all three pancreatic tissues and *BCCIP* in only adult islets). The
1095 X-axis represents variant positions while the Y-axis shows the -log10(eQTL p-value) for the associations between the
1096 genotype of the tested variants and gene expression. For plotting purposes, we assigned a single p-value for gene-level
1097 significance after Bonferroni-correction (0.05 / the number of independent variants tested in fetal-like iPSC-PPC;
1098 horizontal line). Red vertical lines indicate the positions of the lead candidate causal variants underlying the
1099 colocalization based on maximum PP.

1100 **Figure 5. Summary of pancreatic GWAS associations**

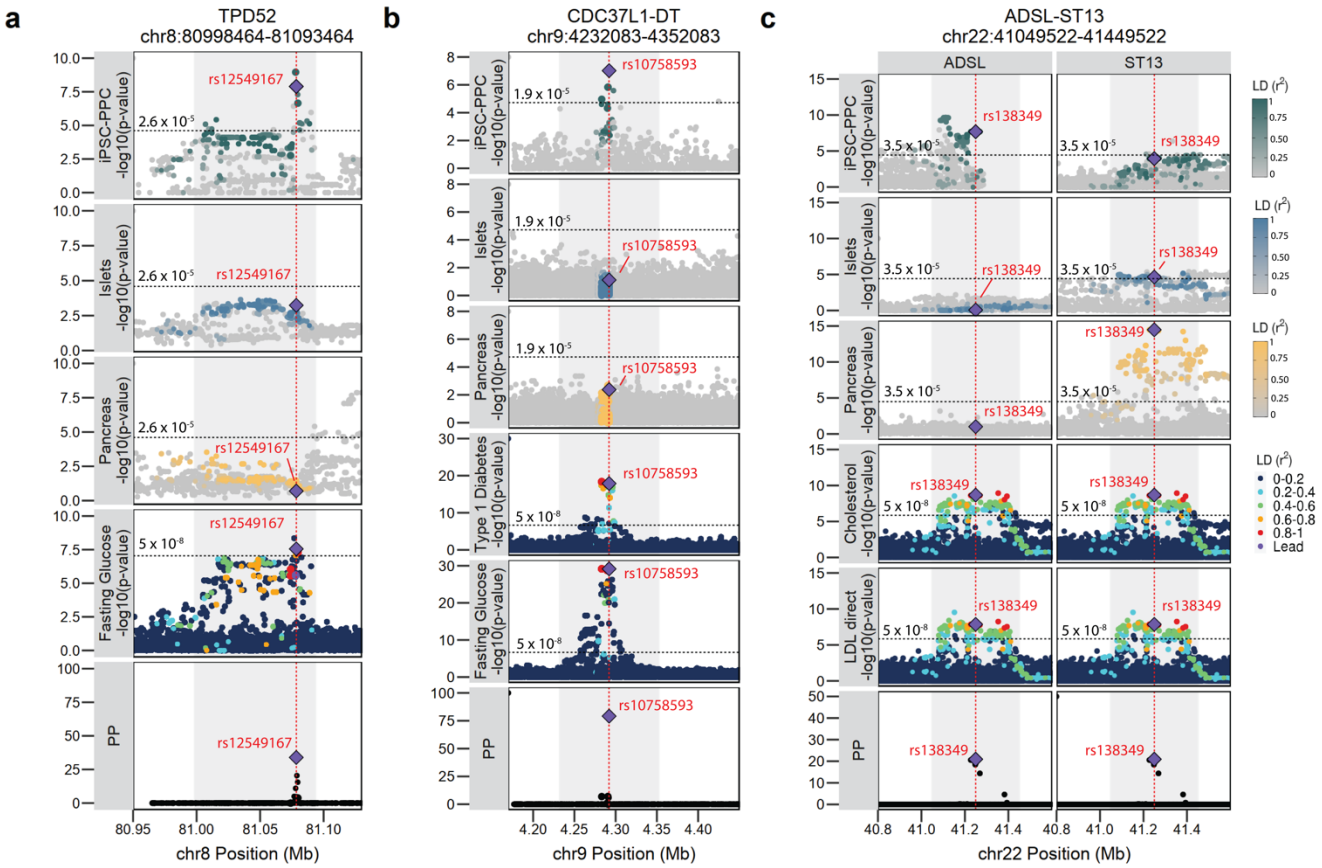


1101

1102 (**a**) Bar plot showing the number of eQTL loci that colocalized with GWAS variants (PP.H4 ≥ 80%) as a singleton or
1103 module. (**b**) Pie chart showing the number of singleton-colocalized GWAS loci (n=183) color-coded by the number of
1104 candidate causal variants identified in their 99% credible sets. (**c**) Pie chart showing the number of module-colocalized
1105 GWAS loci (n=129) color-coded by the number of candidate causal variants identified in their 99% credible sets.

1106 **Figure 6. Pancreatic GWAS associations with fetal-specific and adult-shared gene**

1107 **Expression**

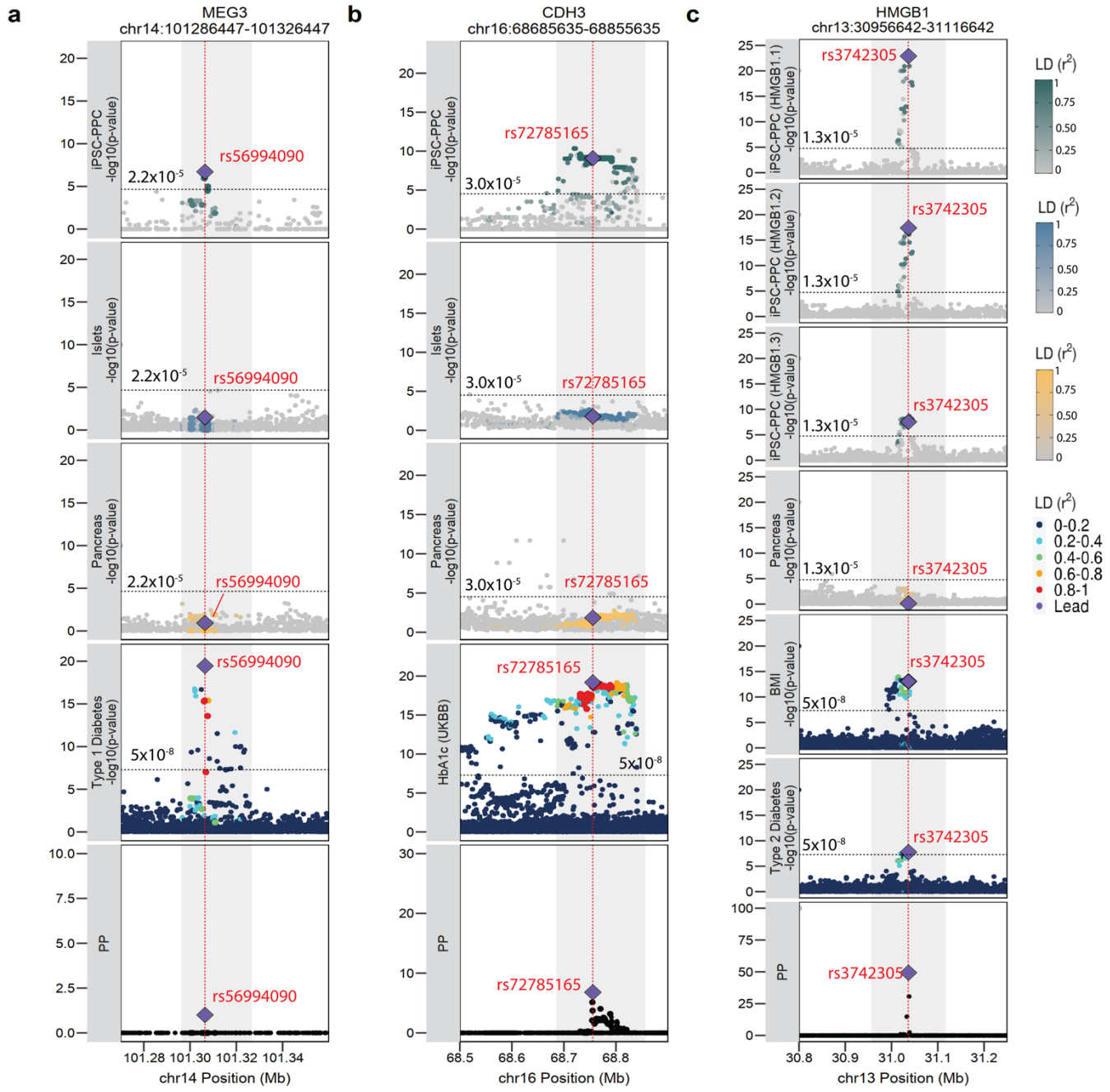


1108

1109 (**a**) The TPD52 locus is associated with fasting glucose levels and colocalized with a fetal-like iPSC-PPC-unique
1110 singleton e_gQTL with the predicted causal variant identified as rs12549167 (chr8:81078464:C>T, PP = 33.9%). (**b**)
1111 The CDC37L1-DT locus is associated with fasting glucose and type 1 diabetes and colocalized with an iPSC-PPC-
1112 unique singleton e_gQTL with the predicted causal variant identified as rs10758593 (chr9:4292083:G>A, PP = 79.2%).
1113 (**c**) Cholesterol and LDL direct GWAS loci colocalize with a fetal-adult e_gQTL module where the variants are strongly
1114 associated with *ADSL* expression in iPSC-PPC and *ST13* expression in the adult whole pancreas (also weakly
1115 associated with *ST13* expression in the adult islets). The predicted causal variant was identified as rs138349
1116 (chr22:41249522:A>G, PP = 21.9%). For plotting purposes, we assigned a single p-value for gene-level significance
1117 based on Bonferroni-correction (0.05 divided by the number of independent variants tested in fetal-like iPSC-PPC;
1118 horizontal line). We note that this p-value does not reflect the thresholds used to define a significant eQTL in the
1119 original adult studies [10,11]. Therefore, while the *ST13* eQTL in adult islets in panel **c** is below the horizontal line, it had
1120 an FDR < 1% in the original study [11]. In each panel, the X-axis represents variant positions while the Y-axis either
1121 shows the -log10(eQTL p-value) for the associations between the genotype of the tested variants and gene expression

1122 or the -log10(GWAS p-value) for the associations between the tested variants and the GWAS trait. For GWAS
1123 significance, we used -log10($5x10^{-8}$). Red vertical lines indicate the positions of the lead candidate causal variants
1124 underlying the colocalization based on maximum PP. For loci that colocalized with multiple GWAS traits, we used the
1125 credible set that yielded the smallest number of variants to plot the "PP" fine-mapping panel.

1126 **Figure 7. Pancreatic GWAS associations with fetal-specific alternative splicing**



1127

1128 (**a**) T1D-risk locus colocalized with a fetal-like iPSC-PPC-unique singleton e_{AS}QTL for *MEG3* with the predicted
1129 causal variant rs56994090 (chr14:101306447:T>C, PP = 100%). (**b**) GWAS locus associated with HbA1c colocalized
1130 with an iPSC-PPC-unique singleton e_{AS}QTL for *CDH3* with the predicted causal variant rs72785165
1131 (chr16:68755635:T>A, PP = 6.8%). (**c**) GWAS locus associated with T2D-risk and BMI colocalized with an iPSC-
1132 PPC-unique e_{AS}QTL module (AS_13_2) for differential usage of three *HMGB1* isoforms with a predicted causal variant
1133 rs3742305 (chr13:31036642:C>G, PP = 49.3%). In each panel, the X-axis represents variant positions while the Y-

1134 axis either shows the -log10(eQTL p-value) for the associations between the genotype of the tested variants and gene
1135 expression or the -log10(GWAS p-value) for the associations between the tested variants and the GWAS trait. For
1136 GWAS significance, we used -log10($5x10^{-8}$). For eQTL significance, we used a single p-value for gene-level
1137 significance after Bonferroni-correction (0.05 / the number of independent variants tested in fetal-like iPSC-PPC;
1138 horizontal line). Red vertical lines indicate the positions of the lead candidate causal variants underlying the
1139 colocalization based on maximum PP. For loci that colocalized with multiple GWAS traits, we used the credible set
1140 that yielded the smallest number of variants to plot the "PP" fine-mapping panel.

1141 # References

1142 1. Broad Genomics Platform, DiscovEHR Collaboration, CHARGE, et al. Exome sequencing of 20,791 cases of
1143 type 2 diabetes and 24,440 controls. *Nature*. 2019;570(7759):71-76. doi:10.1038/s41586-019-1231-2

1144 2. Chen J, Spracklen CN, Marenne G, et al. The trans-ancestral genomic architecture of glycemic traits. *Nat Genet*.
1145 2021;53(6):840-860. doi:10.1038/s41588-021-00852-9

1146 3. Chiou J, Geusz RJ, Okino ML, et al. Interpreting type 1 diabetes risk with genetics and single-cell epigenomics.
1147 *Nature*. 2021;594(7863):398-402. doi:10.1038/s41586-021-03552-w

1148 4. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-
1149 density imputation and islet-specific epigenome maps. *Nat Genet*. 2018;50(11):1505-1513. doi:10.1038/s41588-
1150 018-0241-6

1151 5. Ernst J, Kheradpour P, Mikkelsen TS, et al. Mapping and analysis of chromatin state dynamics in nine human
1152 cell types. *Nature*. 2011;473(7345):43-49. doi:10.1038/nature09906

1153 6. Maurano MT, Humbert R, Rynes E, et al. Systematic Localization of Common Disease-Associated Variation in
1154 Regulatory DNA. *Science*. 2012;337(6099):1190-1195. doi:10.1126/science.1222794

1155 7. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human
1156 epigenomes. *Nature*. 2015;518(7539):317-330. doi:10.1038/nature14248

1157 8. GTEx Consortium. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204-213.
1158 doi:10.1038/nature24277

1159 9. Kim-Hellmuth S, Aguet F, Oliva M, et al. Cell type–specific genetic regulation of gene expression across human
1160 tissues. *Science*. 2020;369(6509):eaaz8528. doi:10.1126/science.aaz8528

1161 10. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*.
1162 2020;369(6509):1318-1330. doi:10.1126/science.aaz1776

1163 11. Viñuela A, Varshney A, van de Bunt M, et al. Genetic variant effects on gene expression in human pancreatic
1164 islets and their implications for T2D. *Nat Commun*. 2020;11(1):4912. doi:10.1038/s41467-020-18581-8

1165 12. D'Antonio M, Nguyen JP, Arthur TD, et al. Fine mapping spatiotemporal mechanisms of genetic variants
1166 underlying cardiac traits and disease. *Nat Commun*. 2023;14(1):1132. doi:10.1038/s41467-023-36638-2

1167 13. Strober BJ, Elorbany R, Rhodes K, et al. Dynamic genetic regulation of gene expression during cellular
1168 differentiation. *Science*. 2019;364(6447):1287-1290. doi:10.1126/science.aaw0040

1169 14. Geusz RJ, Wang A, Chiou J, et al. Pancreatic progenitor epigenome maps prioritize type 2 diabetes risk genes
1170 with roles in development. *eLife*. 2021;10:e59067. doi:10.7554/eLife.59067

1171 15. Dabelea D, Pettitt DJ. Intrauterine Diabetic Environment Confers Risks for Type 2 Diabetes Mellitus and Obesity
1172 in the Offspring, in Addition to Genetic Susceptibility. *Journal of Pediatric Endocrinology and Metabolism*.
1173 2001;14(8). doi:10.1515/jpem-2001-0803

1174 16. Bhattacharya A, Freedman AN, Avula V, et al. Placental genomics mediates genetic associations with complex
1175 health traits and disease. *Nat Commun*. 2022;13(1):706. doi:10.1038/s41467-022-28365-x

1176 17. Petersen MBK, Gonçalves CAC, Kim YH, Grapin-Botton A. Recapitulating and Deciphering Human Pancreas
1177 Development From Human Pluripotent Stem Cells in a Dish. In: *Current Topics in Developmental Biology*. Vol
1178 129. Elsevier; 2018:143-190. doi:10.1016/bs.ctdb.2018.02.009

1179 18. Colclough K, Bellanne-Chantelot C, Saint-Martin C, Flanagan SE, Ellard S. Mutations in the Genes Encoding
1180 the Transcription Factors Hepatocyte Nuclear Factor 1 Alpha and 4 Alpha in Maturity-Onset Diabetes of the
1181 Young and Hyperinsulinemic Hypoglycemia. *Human Mutation*. 2013;34(5):669-685. doi:10.1002/humu.22279

1182 19. Hansen L, Urioste S, Petersen HV, et al. Missense Mutations in the Human Insulin Promoter Factor-1 Gene and
1183 Their Relation to Maturity-Onset Diabetes of the Young and Late-Onset Type 2 Diabetes Mellitus in
1184 Caucasians*. *The Journal of Clinical Endocrinology & Metabolism*. 2000;85(3):1323-1326.
1185 doi:10.1210/jcem.85.3.6421

1186 20. Sanyoura M, Philipson LH, Naylor R. Monogenic Diabetes in Children and Adolescents: Recognition and
1187 Treatment Options. *Curr Diab Rep*. 2018;18(8):58. doi:10.1007/s11892-018-1024-2

1188 21. Ameri J, Borup R, Prawiro C, et al. Efficient Generation of Glucose-Responsive Beta Cells from Isolated GP2 +
1189 Human Pancreatic Progenitors. *Cell Reports*. 2017;19(1):36-49. doi:10.1016/j.celrep.2017.03.032

1190 22. Gonçalves CA, Larsen M, Jung S, et al. A 3D system to model human pancreas development and its reference
1191 single-cell transcriptome atlas identify signaling pathways required for progenitor expansion. *Nat Commun*.
1192 2021;12(1):3144. doi:10.1038/s41467-021-23295-6

1193 23. Nostro MC, Sarangi F, Yang C, et al. Efficient Generation of NKX6-1+ Pancreatic Progenitors from Multiple
1194 Human Pluripotent Stem Cell Lines. *Stem Cell Reports*. 2015;4(4):591-604. doi:10.1016/j.stemcr.2015.02.017

1195 24. Pagliuca FW, Millman JR, Gürtler M, et al. Generation of Functional Human Pancreatic β Cells In Vitro. *Cell*.
1196 2014;159(2):428-439. doi:10.1016/j.cell.2014.09.040

1197 25. Rezania A, Bruin JE, Arora P, et al. Reversal of diabetes with insulin-producing cells derived in vitro from human
1198 pluripotent stem cells. *Nat Biotechnol*. 2014;32(11):1121-1133. doi:10.1038/nbt.3033

1199 26. Russ HA, Parent AV, Ringler JJ, et al. Controlled induction of human pancreatic progenitors produces functional
1200 beta-like cells *in vitro*. *EMBO J*. 2015;34(13):1759-1772. doi:10.15252/embj.201591058

1201 27. Veres A, Faust AL, Bushnell HL, et al. Charting cellular identity during human in vitro β-cell differentiation.
1202 *Nature*. 2019;569(7756):368-373. doi:10.1038/s41586-019-1168-5

1203 28. Sean DLO, Liu Z, Sun H, et al. *Single-Cell Multi-Omic Roadmap of Human Fetal Pancreatic Development*.
1204 Developmental Biology; 2022. doi:10.1101/2022.02.17.480942

1205 29. Seymour PA. Sox9: A Master Regulator of the Pancreatic Program. *Rev Diabet Stud*. 2014;11(1):51-83.
1206 doi:10.1900/RDS.2014.11.51

1207 30. Seymour PA, Freude KK, Tran MN, et al. SOX9 is required for maintenance of the pancreatic progenitor cell
1208 pool. *Proc Natl Acad Sci USA*. 2007;104(6):1865-1870. doi:10.1073/pnas.0609217104

1209 31. Aigha II, Abdelalim EM. NKX6.1 transcription factor: a crucial regulator of pancreatic β cell development,
1210 identity, and proliferation. *Stem Cell Res Ther*. 2020;11(1):459. doi:10.1186/s13287-020-01977-0

1211 32. Oliver-Krasinski JM, Stoffers DA. On the origin of the β cell. *Genes Dev*. 2008;22(15):1998-2021.
1212 doi:10.1101/gad.1670808

1213 33. Van Hoof D, D'Amour KA, German MS. Derivation of insulin-producing cells from human embryonic stem
1214 cells. *Stem Cell Research*. 2009;3(2-3):73-87. doi:10.1016/j.scr.2009.08.003

1215 34. Ramond C, Beydag-Tasöz BS, Azad A, et al. Understanding human fetal pancreas development using
1216 subpopulation sorting, RNA sequencing and single-cell profiling. *Development*. Published online January 1,
1217 2018:dev.165480. doi:10.1242/dev.165480

1218 35. Panopoulos AD, D'Antonio M, Benaglio P, et al. iPSCORE: A Resource of 222 iPSC Lines Enabling Functional
1219 Characterization of Genetic Variation across a Variety of Cell Types. *Stem Cell Reports*. 2017;8(4):1086-1100.
1220 doi:10.1016/j.stemcr.2017.03.012

1221 36. Jin W, Jiang W. Stepwise differentiation of functional pancreatic β cells from human pluripotent stem cells. *Cell*
1222 *Regen*. 2022;11(1):24. doi:10.1186/s13619-022-00125-8

1223 37. DeBoever C, Li H, Jakubosky D, et al. Large-Scale Profiling Reveals the Influence of Genetic Variation on Gene
1224 Expression in Human Induced Pluripotent Stem Cells. *Cell Stem Cell*. 2017;20(4):533-546.e7.
1225 doi:10.1016/j.stem.2017.03.009

1226 38. Fadista J, Vikman P, Laakso EO, et al. Global genomic and transcriptomic analysis of human pancreatic islets
1227 reveals novel genes influencing glucose metabolism. *Proc Natl Acad Sci USA*. 2014;111(38):13924-13929.
1228 doi:10.1073/pnas.1402665111

1229 39. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585.
1230 doi:10.1038/ng.2653

1231 40. Jansen R, Hottenga JJ, Nivard MG, et al. Conditional eQTL analysis reveals allelic heterogeneity of gene
1232 expression. *Hum Mol Genet*. 2017;26(8):1444-1451. doi:10.1093/hmg/ddx043

1233 41. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian Test for Colocalisation between Pairs of Genetic
1234 Association Studies Using Summary Statistics. Williams SM, ed. *PLoS Genet*. 2014;10(5):e1004383.
1235 doi:10.1371/journal.pgen.1004383

1236 42. Yan J, Qiu Y, Ribeiro dos Santos AM, et al. Systematic analysis of binding of transcription factors to noncoding
1237 variants. *Nature*. 2021;591(7848):147-151. doi:10.1038/s41586-021-03211-0

1238 43. Garrido-Martín D, Borsari B, Calvo M, Reverter F, Guigó R. Identification and analysis of splicing quantitative
1239 trait loci across multiple tissues in the human genome. *Nat Commun*. 2021;12(1):727. doi:10.1038/s41467-020-
1240 20578-2

1241 44. van de Bunt M, Manning Fox JE, Dai X, et al. Transcript Expression Data from Human Islets Links Regulatory
1242 Signals from Genome-Wide Association Studies for Type 2 Diabetes and Glycemic Traits to Their Downstream
1243 Effectors. Stranger BE, ed. *PLoS Genet*. 2015;11(12):e1005694. doi:10.1371/journal.pgen.1005694

1244 45. Chen JH, Zhao Y, Khan RAW, et al. SNX29, a new susceptibility gene shared with major mental disorders in
1245 Han Chinese population. *The World Journal of Biological Psychiatry*. 2021;22(7):526-534.
1246 doi:10.1080/15622975.2020.1845793

1247 46. Anderson D, Cordell HJ, Fakiola M, et al. First genome-wide association study in an Australian aboriginal
1248 population provides insights into genetic risk factors for body mass index and type 2 diabetes. *PLoS One*.
1249 2015;10(3):e0119333. doi:10.1371/journal.pone.0119333

1250 47. Frankish A, Diekhans M, Ferreira AM, et al. GENCODE reference annotation for the human and mouse genomes.
1251 *Nucleic Acids Research*. 2019;47(D1):D766-D773. doi:10.1093/nar/gky955

1252 48. Thurner M, van de Bunt M, Torres JM, et al. Integration of human pancreatic islet genomic data refines regulatory
1253 mechanisms at Type 2 Diabetes susceptibility loci. *eLife*. 2018;7:e31977. doi:10.7554/eLife.31977

1254 49. Bolt CC, Lopez-Delisle L, Hintermann A, et al. Context-dependent enhancer function revealed by targeted inter-
1255 TAD relocation. *Nat Commun*. 2022;13(1):3488. doi:10.1038/s41467-022-31241-3

1256 50. Pan-UKB team. Published online 2020. https://pan.ukbb.broadinstitute.org

1257 51. Dimas AS, Lagou V, Barker A, et al. Impact of Type 2 Diabetes Susceptibility Variants on Quantitative Glycemic
1258 Traits Reveals Mechanistic Heterogeneity. *Diabetes*. 2014;63(6):2158-2171. doi:10.2337/db13-0949

1259 52. Grarup N, Sandholt CH, Hansen T, Pedersen O. Genetic susceptibility to type 2 diabetes and obesity: from
1260 genome-wide association studies to rare variants and beyond. *Diabetologia*. 2014;57(8):1528-1541.
1261 doi:10.1007/s00125-014-3270-4

1262 53. D'Antonio M, Nguyen JP, Arthur TD, et al. In heart failure reactivation of RNA-binding proteins is associated
1263 with the expression of 1,523 fetal-specific isoforms. Zhang Z, ed. *PLoS Comput Biol*. 2022;18(2):e1009918.
1264 doi:10.1371/journal.pcbi.1009918

1265 54. Baralle FE, Giudice J. Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell*
1266 *Biol*. 2017;18(7):437-451. doi:10.1038/nrm.2017.27

1267 55. Mazin PV, Khaitovich P, Cardoso-Moreira M, Kaessmann H. Alternative splicing during mammalian organ
1268 development. *Nat Genet*. 2021;53(6):925-934. doi:10.1038/s41588-021-00851-w

1269 56. Brun T, Jiménez-Sánchez C, Madsen JGS, et al. AMPK Profiling in Rodent and Human Pancreatic Beta-Cells
1270 under Nutrient-Rich Metabolic Stress. *IJMS*. 2020;21(11):3982. doi:10.3390/ijms21113982

1271 57. Minokoshi Y, Alquier T, Furukawa N, et al. AMP-kinase regulates food intake by responding to hormonal and
1272 nutrient signals in the hypothalamus. *Nature*. 2004;428(6982):569-574. doi:10.1038/nature02440

1273 58. Shaw RJ, Lamia KA, Vasquez D, et al. The Kinase LKB1 Mediates Glucose Homeostasis in Liver and
1274 Therapeutic Effects of Metformin. *Science*. 2005;310(5754):1642-1646. doi:10.1126/science.1120781

1275 59. Yamauchi T, Kamon J, Minokoshi Y, et al. Adiponectin stimulates glucose utilization and fatty-acid oxidation
1276 by activating AMP-activated protein kinase. *Nat Med*. 2002;8(11):1288-1295. doi:10.1038/nm788

1277 60. Wu Y, Viana M, Thirumangalathu S, Loeken MR. AMP-activated protein kinase mediates effects of oxidative
1278 stress on embryo gene expression in a mouse model of diabetic embryopathy. *Diabetologia*. 2012;55(1):245-254.
1279 doi:10.1007/s00125-011-2326-y

1280 61. Grant SFA, Qu HQ, Bradfield JP, et al. Follow-Up Analysis of Genome-Wide Association Data Identifies Novel
1281 Loci for Type 1 Diabetes. *Diabetes*. 2009;58(1):290-295. doi:10.2337/db08-1022

1282 62. the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, Morris AP, Voight BF, et al.
1283 Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2
1284 diabetes. *Nat Genet*. 2012;44(9):981-990. doi:10.1038/ng.2383

1285 63. Kang HS, Kim YS, ZeRuth G, et al. Transcription Factor Glis3, a Novel Critical Player in the Regulation of
1286 Pancreatic β-Cell Development and Insulin Gene Expression. *Mol Cell Biol*. 2009;29(24):6366-6379.
1287 doi:10.1128/MCB.01259-09

1288 64. Kang HS, Takeda Y, Jeon K, Jetten AM. The Spatiotemporal Pattern of Glis3 Expression Indicates a Regulatory
1289 Function in Bipotent and Endocrine Progenitors during Early Pancreatic Development and in Beta, PP and Ductal
1290 Cells. Blondeau B, ed. *PLoS ONE*. 2016;11(6):e0157138. doi:10.1371/journal.pone.0157138

1291 65. Yang Y, Chang BH, Chan L. Sustained expression of the transcription factor GLIS3 is required for normal beta
1292 cell function in adults. *EMBO Mol Med*. 2013;5(1):92-104. doi:10.1002/emmm.201201398

1293 66. Sams EI, Ng JK, Tate V, et al. From karyotypes to precision genomics in 9p deletion and duplication syndromes.
1294 *Human Genetics and Genomics Advances*. 2022;3(1):100081. doi:10.1016/j.xhgg.2021.100081

1295 67. Aylward A, Chiou J, Okino ML, Kadakia N, Gaulton KJ. Shared genetic risk contributes to type 1 and type 2
1296 diabetes etiology. *Human Molecular Genetics*. Published online November 7, 2018. doi:10.1093/hmg/ddy314

1297 68. Cao R chang, Yang W jun, Xiao W, et al. St13 protects against disordered acinar cell arachidonic acid pathway
1298 in chronic pancreatitis. *J Transl Med*. 2022;20(1):218. doi:10.1186/s12967-022-03413-8

1299 69. Graham SE, Clarke SL, Wu KHH, et al. The power of genetic diversity in genome-wide association studies of
1300 lipids. *Nature*. 2021;600(7890):675-679. doi:10.1038/s41586-021-04064-3

1301 70. Barrett JC, Clayton DG, Concannon P, et al. Genome-wide association study and meta-analysis find that over 40
1302 loci affect risk of type 1 diabetes. *Nat Genet*. 2009;41(6):703-707. doi:10.1038/ng.381

1303 71. Tong Z, Fan Y, Zhang W, et al. Pancreas-specific Pten deficiency causes partial resistance to diabetes and
1304 elevated hepatic AKT signaling. *Cell Res*. 2009;19(6):710-719. doi:10.1038/cr.2009.42

1305 72. Wong H, Schotz MC. The lipase gene family. *Journal of Lipid Research*. 2002;43(7):993-999.
1306 doi:10.1194/jlr.R200007-JLR200

1307 73. Gaertner B, van Heesch S, Schneider-Lunitz V, et al. A human ESC-based screen identifies a role for the
1308 translated lncRNA LINC00261 in pancreatic endocrine differentiation. *eLife*. 2020;9:e58659.
1309 doi:10.7554/eLife.58659

1310 74. Chang W wei, Zhang L, Yao X ming, et al. Upregulation of long non-coding RNA MEG3 in type 2 diabetes
1311 mellitus complicated with vascular disease: a case–control study. *Mol Cell Biochem*. 2020;473(1-2):93-99.
1312 doi:10.1007/s11010-020-03810-x

1313 75. Kameswaran V, Bramswig NC, McKenna LB, et al. Epigenetic Regulation of the DLK1-MEG3 MicroRNA
1314 Cluster in Human Type 2 Diabetic Islets. *Cell Metabolism*. 2014;19(1):135-145. doi:10.1016/j.cmet.2013.11.016

1315 76. Kameswaran V, Golson ML, Ramos-Rodríguez M, et al. The Dysregulation of the *DLK1 - MEG3* Locus in Islets
1316 From Patients With Type 2 Diabetes Is Mimicked by Targeted Epimutation of Its Promoter With TALE-DNMT
1317 Constructs. *Diabetes*. 2018;67(9):1807-1815. doi:10.2337/db17-0682

1318 77. Onengut-Gumuscu S, Chen WM, Burren O, et al. Fine mapping of type 1 diabetes susceptibility loci and evidence
1319 for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet*. 2015;47(4):381-386.
1320 doi:10.1038/ng.3245

1321 78. Chen J, Liu Y, Min J, et al. Alternative splicing of lncRNAs in human diseases. *Am J Cancer Res*.
1322 2021;11(3):624-639.

1323 79. Wheeler E, Leong A, Liu CT, et al. Impact of common genetic determinants of Hemoglobin A1c on type 2
1324 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLoS*
1325 *Med*. 2017;14(9):e1002383. doi:10.1371/journal.pmed.1002383

1326 80. Parnaud G, Lavallard V, Bedat B, et al. Cadherin engagement improves insulin secretion of single human β-cells.
1327 *Diabetes*. 2015;64(3):887-896. doi:10.2337/db14-0257

1328 81. Pulit SL, Stoneman C, Morris AP, et al. Meta-analysis of genome-wide association studies for body fat
1329 distribution in 694 649 individuals of European ancestry. *Hum Mol Genet*. 2019;28(1):166-174.
1330 doi:10.1093/hmg/ddy327

1331 82. Sakaue S, Kanai M, Tanigawa Y, et al. A cross-population atlas of genetic associations for 220 human
1332 phenotypes. *Nat Genet*. 2021;53(10):1415-1424. doi:10.1038/s41588-021-00931-x

1333 83. Vujkovic M, Keaton JM, Lynch JA, et al. Discovery of 318 new risk loci for type 2 diabetes and related vascular
1334 outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat Genet*. 2020;52(7):680-691.
1335 doi:10.1038/s41588-020-0637-y

1336 84. Zhu Z, Guo Y, Shi H, et al. Shared genetic and experimental links between obesity-related traits and asthma
1337 subtypes in UK Biobank. *J Allergy Clin Immunol*. 2020;145(2):537-549. doi:10.1016/j.jaci.2019.09.035

1338 85. Calogero S, Grassi F, Aguzzi A, et al. The lack of chromosomal protein Hmg1 does not disrupt cell growth but
1339 causes lethal hypoglycaemia in newborn mice. *Nat Genet*. 1999;22(3):276-280. doi:10.1038/10338

1340 86. Wang Y, Zhong J, Zhang X, et al. The Role of HMGB1 in the Pathogenesis of Type 2 Diabetes. *J Diabetes Res*.
1341 2016;2016:2543268. doi:10.1155/2016/2543268

1342 87. Chen C, Yu W, Tober J, et al. Spatial Genome Re-organization between Fetal and Adult Hematopoietic Stem
1343 Cells. *Cell Reports*. 2019;29(12):4200-4211.e7. doi:10.1016/j.celrep.2019.11.065

1344 88. Gorkin DU, Barozzi I, Zhao Y, et al. An atlas of dynamic chromatin landscapes in mouse fetal development.
1345 *Nature*. 2020;583(7818):744-751. doi:10.1038/s41586-020-2093-3

1346 89. Zhang K, Hocker JD, Miller M, et al. A single-cell atlas of chromatin accessibility in the human genome. *Cell*.
1347 2021;184(24):5985-6001.e19. doi:10.1016/j.cell.2021.10.024

1348 90. Ong C, Corces VG. Enhancers: emerging roles in cell fate specification. *EMBO Rep*. 2012;13(5):423-430.
1349 doi:10.1038/embor.2012.52

1350 91. Su CH, D D, Tarn WY. Alternative Splicing in Neurogenesis and Brain Development. *Front Mol Biosci*.
1351 2018;5:12. doi:10.3389/fmolb.2018.00012

1352 92. Urbut SM, Wang G, Carbonetto P, Stephens M. Flexible statistical methods for estimating and testing effects in
1353 genomic studies with multiple conditions. *Nat Genet*. 2019;51(1):187-195. doi:10.1038/s41588-018-0268-8

1354 93. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. *GigaScience*. 2021;10(2):1-4.
1355 doi:10.1093/gigascience/giab008

1356 94. Israel MA, Yuan SH, Bardy C, et al. Probing sporadic and familial Alzheimer's disease using induced pluripotent
1357 stem cells. *Nature*. 2012;482(7384):216-220. doi:10.1038/nature10821

1358 95. D'Antonio-Chronowska A, D'Antonio M, Frazer K. In vitro Differentiation of Human iPSC-derived
1359 Cardiovascular Progenitor Cells (iPSC-CVPCs). *Bio-Protocol*. 2020;10(18):1-43. doi:10.21769/bioprotoc.3755

1360 96. Harrow J, Frankish A, Gonzalez JM, et al. GENCODE: The reference human genome annotation for the
1361 ENCODE project. *Genome Research*. 2012;22(9):1760-1774. doi:10.1101/gr.135350.111

1362 97. Kang HM, Subramaniam M, Targ S, et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic
1363 variation. *Nat Biotechnol*. 2018;36(1):89-94. doi:10.1038/nbt.4042

1364 98. Veres A, Faust AL, Bushnell HL, et al. Charting cellular identity during human in vitro β-cell differentiation.
1365 *Nature*. 2019;569(7756):368-373. doi:10.1038/s41586-019-1168-5

1366 99. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different
1367 conditions, technologies, and species. *Nat Biotechnol*. 2018;36(5):411-420. doi:10.1038/nbt.4096

1368 100. D'Antonio-Chronowska A, Donovan MKR, Young Greenwald WW, et al. Association of Human iPSC Gene
1369 Signatures and X Chromosome Dosage with Two Distinct Cardiac Differentiation Trajectories. *Stem Cell*
1370 *Reports*. 2019;13(5):924-938. doi:10.1016/j.stemcr.2019.09.011

1371 101. Dobin A, Davis CA, Schlesinger F, et al. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*.
1372 2013;29(1):15-21. doi:10.1093/bioinformatics/bts635

1373 102. Shaun Purcell CC. PLINK 1.9.0.

1374 103. The 1000 Genomes Project Consortium, Corresponding authors, Auton A, et al. A global reference for human
1375 genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

1376 104. Danecek P, McCarthy SA, HipSci Consortium, Durbin R. A Method for Checking Genomic Integrity in Cultured
1377 Cell Lines from SNP Genotyping Data. *PLoS One*. 2016;11(5):e0155014. doi:10.1371/journal.pone.0155014

1378 105. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical
1379 parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987-2993.
1380 doi:10.1093/bioinformatics/btr509

1381 106. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference
1382 genome. *BMC Bioinformatics*. Published online 2011. doi:10.1201/b16589

1383 107. Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by
1384 pseudotemporal ordering of single cells. *Nat Biotechnol*. 2014;32(4):381-386. doi:10.1038/nbt.2859

1385 108. Newman AM, Steen CB, Liu CL, et al. Determining cell type abundance and expression from bulk tissues with
1386 digital cytometry. *Nat Biotechnol*. 2019;37(7):773-782. doi:10.1038/s41587-019-0114-2

1387 109. Casale FP, Rakitsch B, Lippert C, Stegle O. Efficient set tests for the genetic analysis of correlated traits. *Nat*
1388 *Methods*. 2015;12(8):755-758. doi:10.1038/nmeth.3439

1389 110. Huang QQ, Ritchie SC, Brozynska M, Inouye M. Power, false discovery rate and Winner's Curse in eQTL
1390 studies. *Nucleic Acids Research*. 2018;46(22):e133-e133. doi:10.1093/nar/gky780

1391 111. Van Nostrand EL, Pratt GA, Shishkin AA, et al. Robust transcriptome-wide discovery of RNA-binding protein
1392 binding sites with enhanced CLIP (eCLIP). *Nat Methods*. 2016;13(6):508-514. doi:10.1038/nmeth.3810

1393 112. Lee D, Gorkin DU, Baker M, et al. A method to predict the impact of regulatory variants from DNA sequence.
1394 *Nat Genet*. 2015;47(8):955-961. doi:10.1038/ng.3331

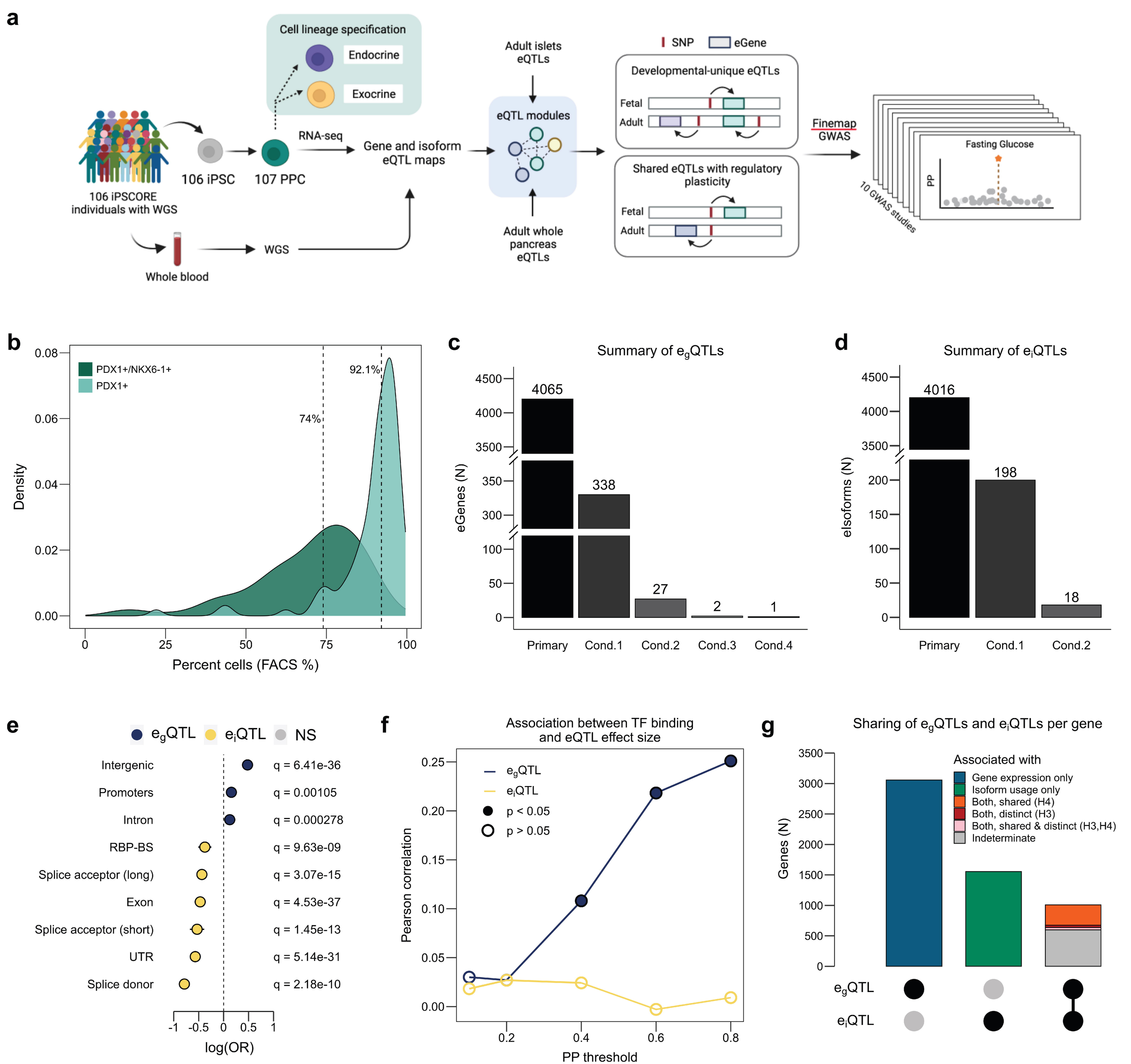
1395 113. Csardi, Gabor N Tamas. The igraph software package for complex network research. *InterJounal*. 2006;Complex
1396 Systems:1695.

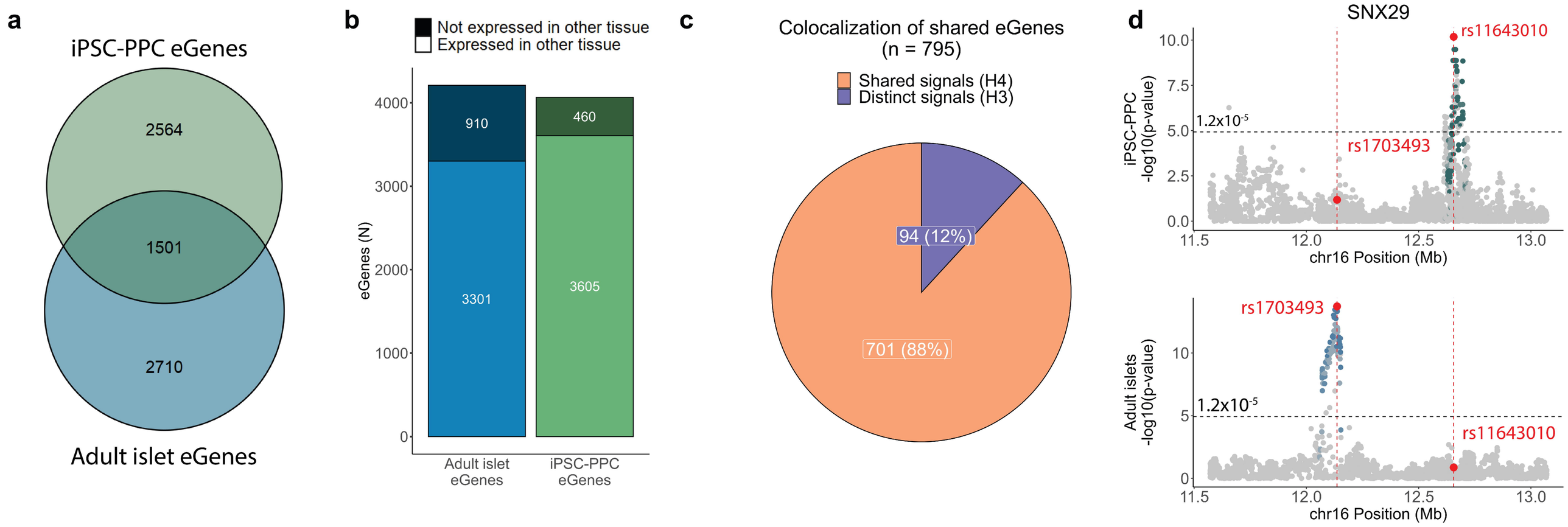
1397 114. Mahajan A, Taliun D, Thurner M, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-
1398 density imputation and islet-specific epigenome maps. *Nature Genetics*. 2018;50(11):1505-1513.
1399 doi:10.1038/s41588-018-0241-6

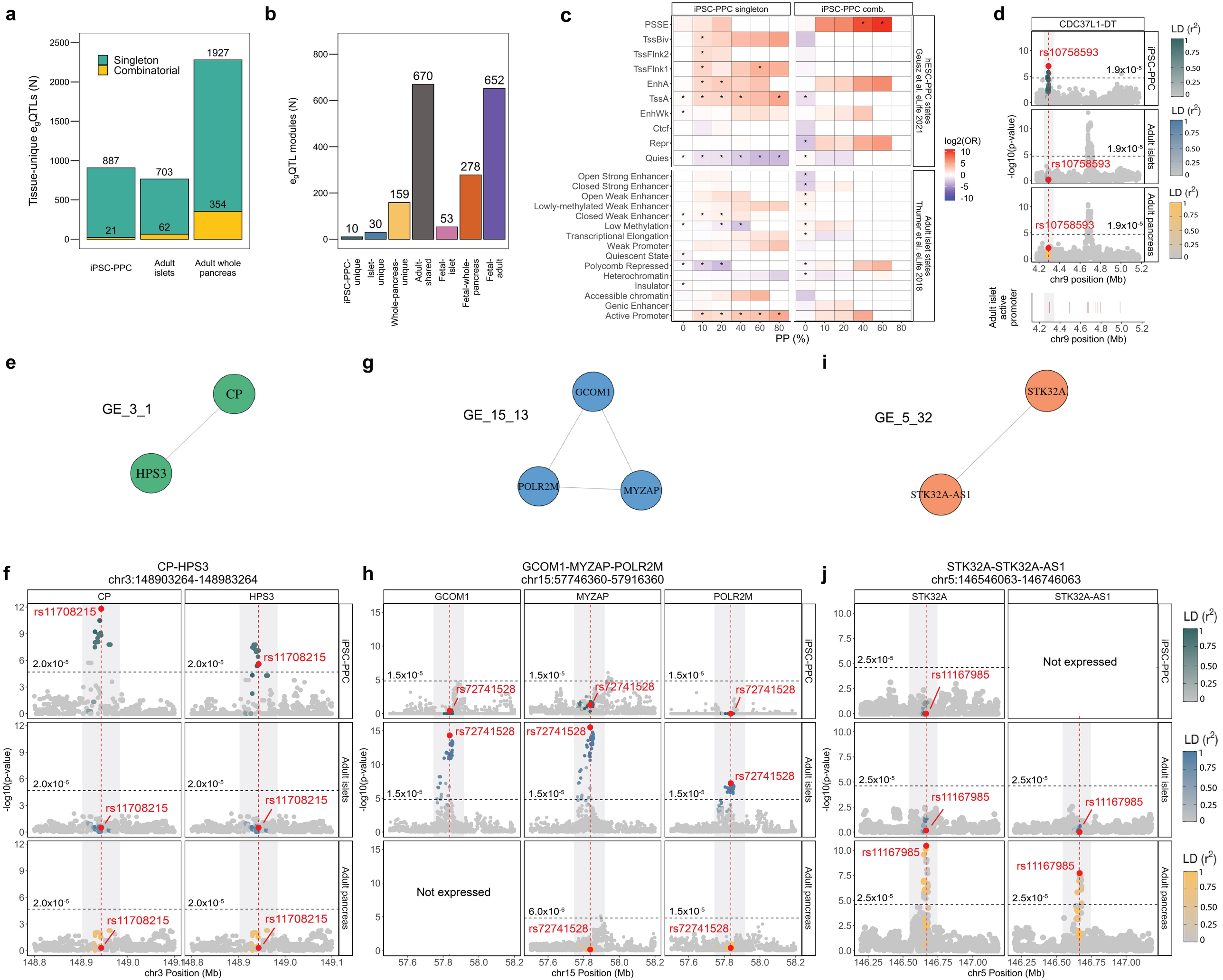
1400 115. Chen J, Spracklen CN, Marenne G, et al. The trans-ancestral genomic architecture of glycemic traits. *Nature*
1401 *Genetics*. 2021;53(6):840-860. doi:10.1038/s41588-021-00852-9

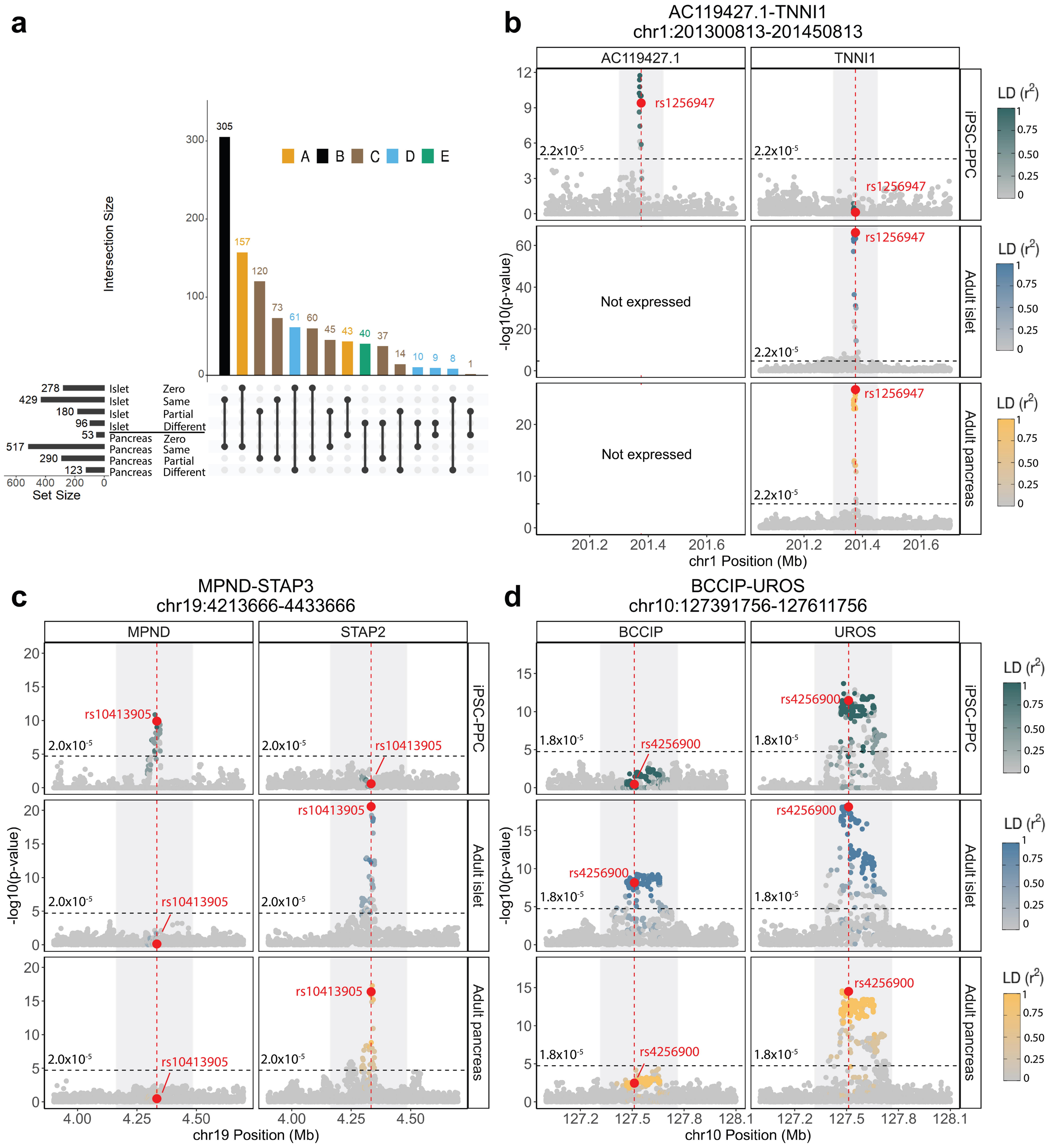
1402 116. Bioconductor Package Maintainer. liftOver: Changing genomic coordinate systems with rtracklayer::liftOver.
1403 Published online 2022.

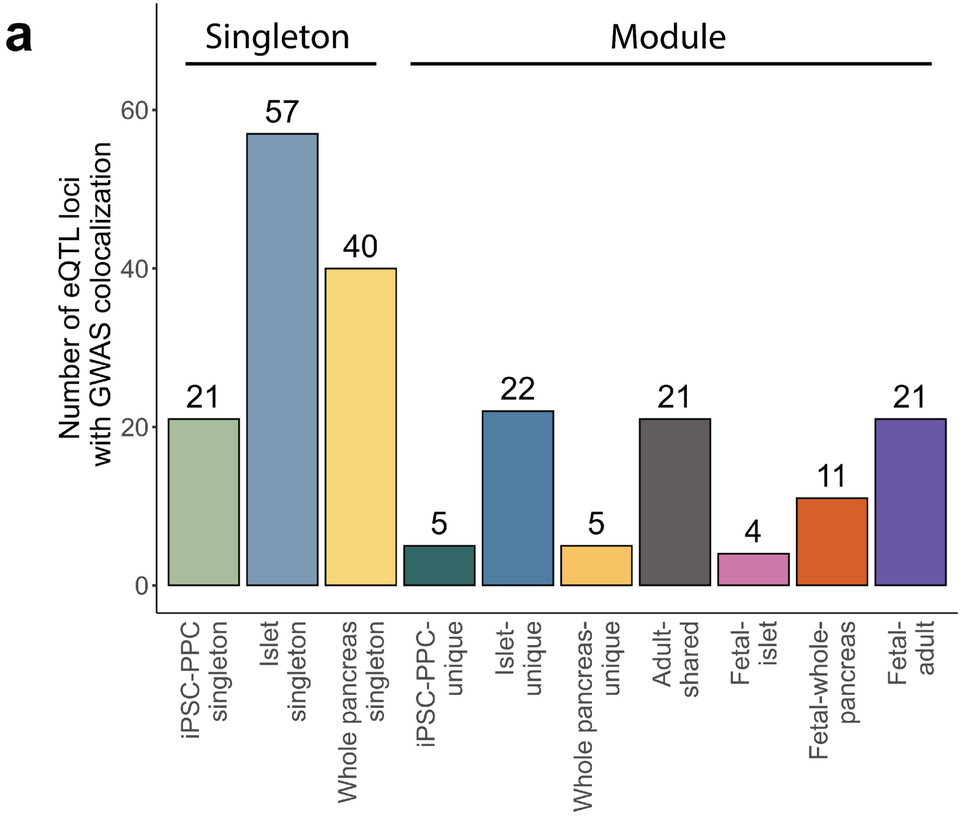
1404

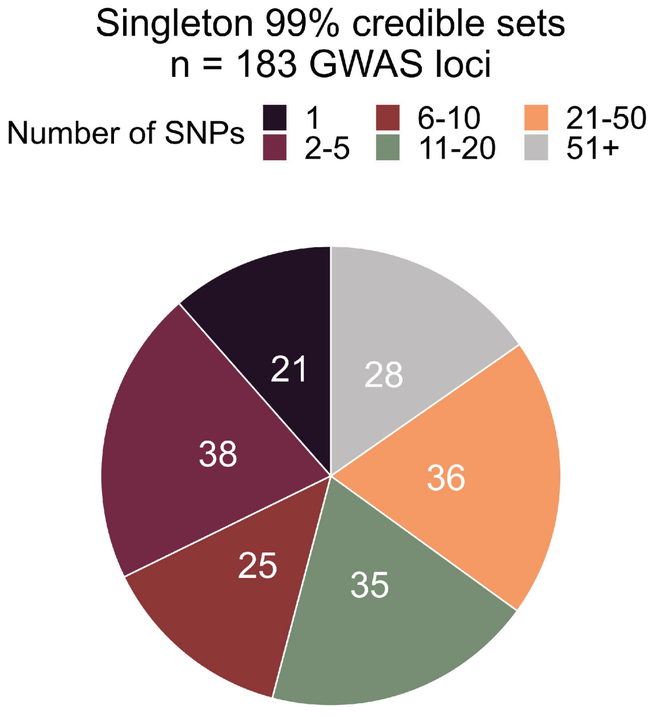

a
Cell lineage specification
Endocrine
Exocrine
106 iPSCORE individuals with WGS
106 iPSC
107 PPC
RNA-seq
Gene and isoform eQTL maps
Whole blood
WGS
Adult islets eQTLs
eQTL modules
Adult whole pancreas eQTLs
SNP
eGene
Developmental-unique eQTLs
Fetal
Adult
Shared eQTLs with regulatory plasticity
Finemap
GWAS
10 GWAS studies
Fasting Glucose
PP
b
PDX1+/NKX6-1+
PDX1+
92.1%
74%
Density
Percent cells (FACS %)
c
Summary of e_gQTLs
4065
338
27
2
1
eGenes (N)
Primary
Cond.1
Cond.2
Cond.3
Cond.4
d
Summary of e_iQTLs
4016
198
18
elsoforms (N)
Primary
Cond.1
Cond.2
e
e_gQTL
e_iQTL
NS
Intergenic
Promoters
Intron
RBP-BS
Splice acceptor (long)
Exon
Splice acceptor (short)
UTR
Splice donor
q = 6.41e-36
q = 0.00105
q = 0.000278
q = 9.63e-09
q = 3.07e-15
q = 4.53e-37
q = 1.45e-13
q = 5.14e-31
q = 2.18e-10
log(OR)
f
Association between TF binding and eQTL effect size
e_gQTL
e_iQTL
p < 0.05
p > 0.05
Pearson correlation
PP threshold
g
Sharing of e_gQTLs and e_iQTLs per gene
Associated with
Gene expression only
Isoform usage only
Both, shared (H4)
Both, distinct (H3)
Both, shared & distinct (H3,H4)
Indeterminate
Genes (N)
e_gQTL
e_iQTL

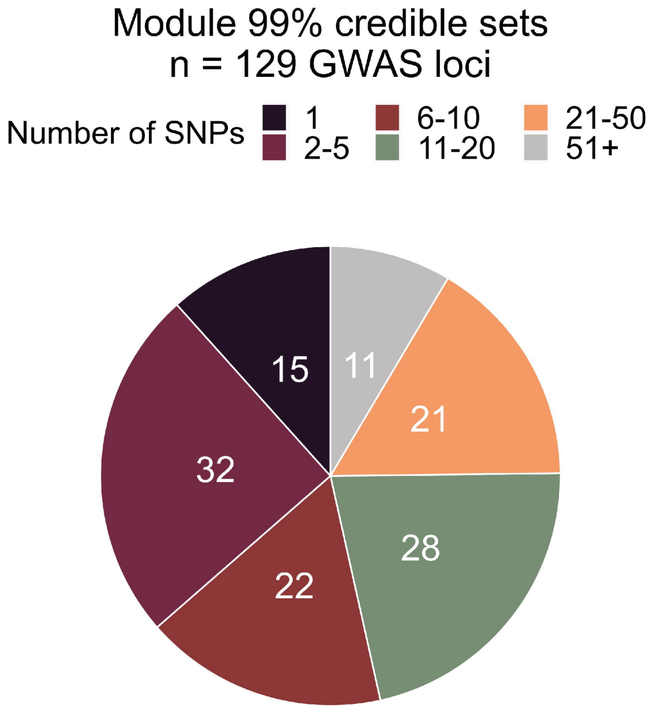

a
iPSC-PPC eGenes
2564
1501
2710
Adult islet eGenes
b
Not expressed in other tissue
Expressed in other tissue
eGenes (N)
4000
3000
2000
1000
0
910
3301
460
3605
Adult islet eGenes
iPSC-PPC eGenes
c
Colocalization of shared eGenes
(n = 795)
Shared signals (H4)
Distinct signals (H3)
94 (12%)
701 (88%)
d
SNX29
rs11643010
rs1703493
1.2x10-5
iPSC-PPC -log10(p-value)
10.0
7.5
5.0
2.5
0.0
11.5
12.0
12.5
13.0
chr16 Position (Mb)
rs1703493
rs11643010
1.2x10-5
Adult islets -log10(p-value)
10
5
0
11.5
12.0
12.5
13.0
chr16 Position (Mb)

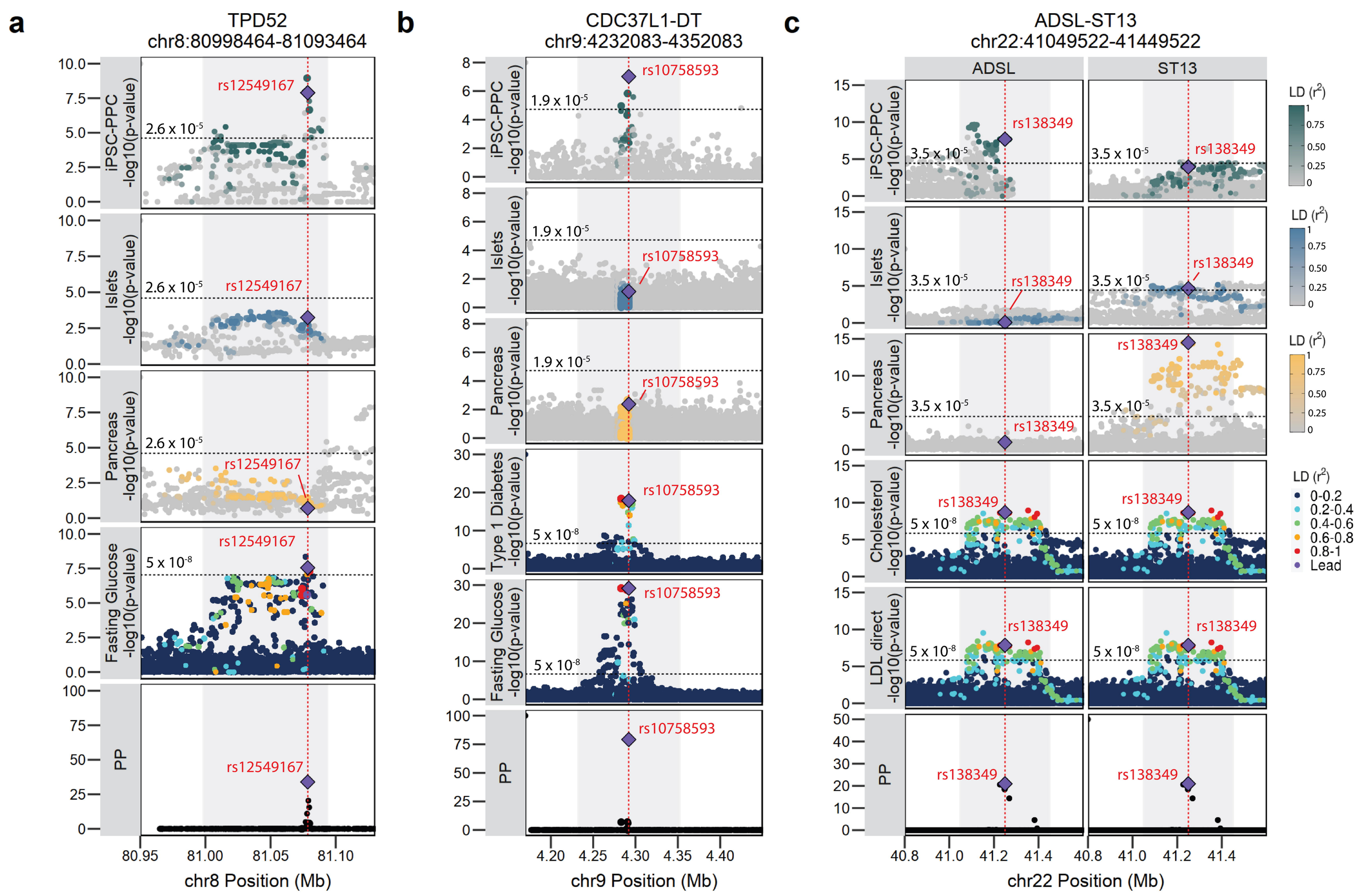

a
Tissue-unique e_gQTLs (N)
Singleton
Combinatorial
887
21
703
62
1927
354
iPSC-PPC
Adult islets
Adult whole pancreas
b
e_gQTL modules (N)
10
30
159
670
53
278
652
iPSC-PPC-unique
Islet-unique
Whole-pancreas-unique
Adult-shared
Fetal-islet
Fetal-whole-pancreas
Fetal-adult
c
iPSC-PPC singleton
iPSC-PPC comb.
PSSE
TssBiv
TssFlnk2
TssFlnk1
EnhA
TssA
EnhWk
Ctcf
Repr
Quies
hESC-PPC states Geusz et al. eLife 2021
Open Strong Enhancer
Closed Strong Enhancer
Open Weak Enhancer
Lowly-methylated Weak Enhancer
Closed Weak Enhancer
Low Methylation
Transcriptional Elongation
Weak Promoter
Quiescent State
Polycomb Repressed
Heterochromatin
Insulator
Accessible chromatin
Genic Enhancer
Active Promoter
Adult islet states Thurner et al. eLife 2018
log2(OR)
PP (%)
d
CDC37L1-DT
rs10758593
1.9x10-5
iPSC-PPC
Adult islets
Adult pancreas
LD (r2)
-log10(p-value)
Adult islet active promoter
chr9 position (Mb)
e
GE_3_1
CP
HPS3
g
GE_15_13
GCOM1
POLR2M
MYZAP
i
GE_5_32
STK32A
STK32A-AS1
f
CP-HPS3
chr3:148903264-148983264
rs11708215
2.0x10-5
chr3 Position (Mb)
h
GCOM1-MYZAP-POLR2M
chr15:57746360-57916360
rs72741528
1.5x10-5
6.0x10-6
Not expressed
chr15 Position (Mb)
j
STK32A-STK32A-AS1
chr5:146546063-146746063
rs11167985
2.5x10-5
Not expressed
chr5 Position (Mb)

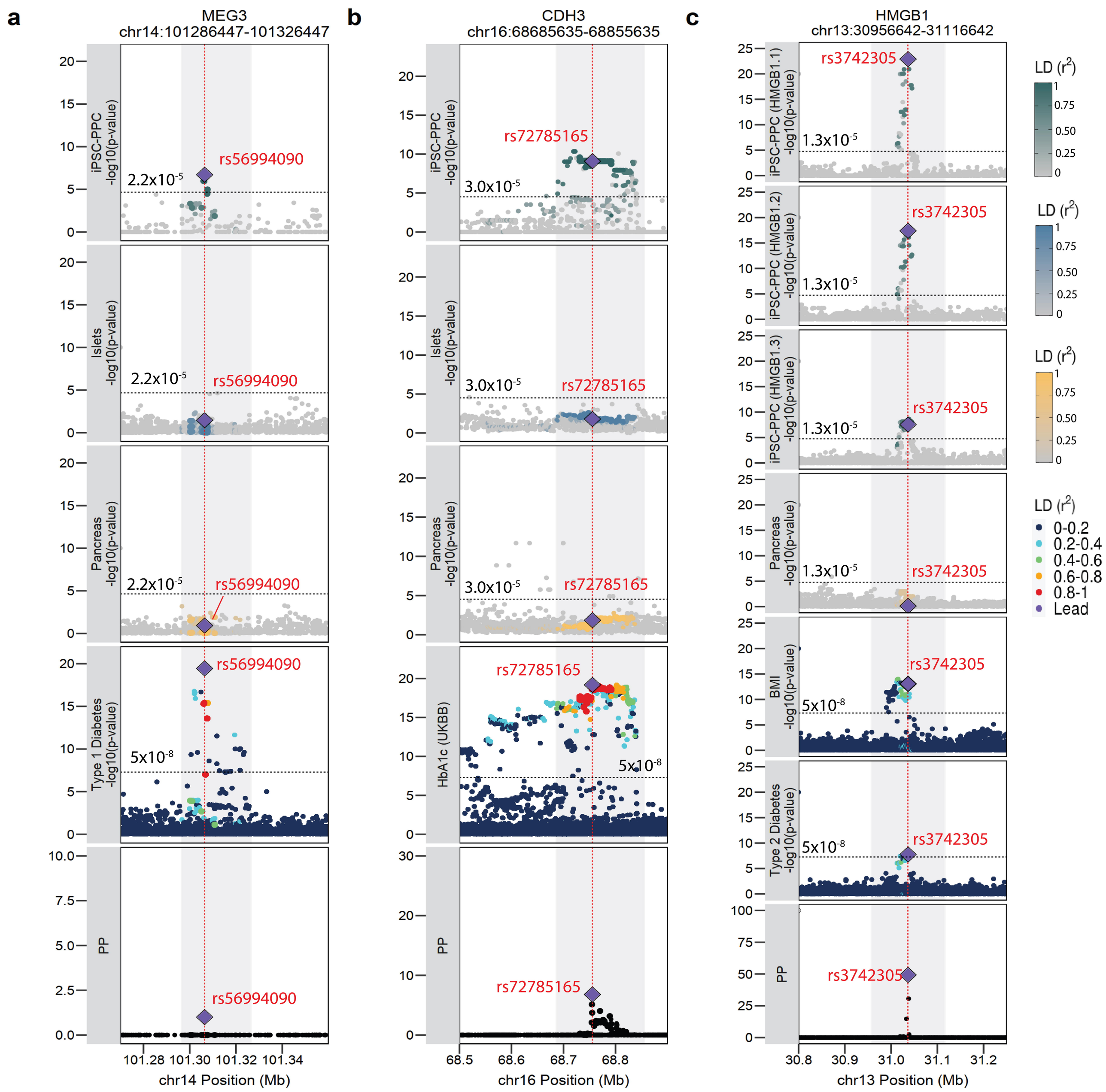

a
Intersection Size
305
157
120
73
61
60
45
43
40
37
14
10
9
8
1
A B C D E
278 Islet Zero
429 Islet Same
180 Islet Partial
96 Islet Different
53 Pancreas Zero
517 Pancreas Same
290 Pancreas Partial
123 Pancreas Different
Set Size
b
AC119427.1-TNNI1
chr1:201300813-201450813
AC119427.1
TNNI1
rs1256947
2.2x10-5
iPSC-PPC
Adult islet
Adult pancreas
Not expressed
-log10(p-value)
LD (r2)
chr1 Position (Mb)
c
MPND-STAP3
chr19:4213666-4433666
MPND
STAP2
rs10413905
2.0x10-5
iPSC-PPC
Adult islet
Adult pancreas
-log10(p-value)
chr19 Position (Mb)
d
BCCIP-UROS
chr10:127391756-127611756
BCCIP
UROS
rs4256900
1.8x10-5
iPSC-PPC
Adult islet
Adult pancreas
-log10(p-value)
chr10 Position (Mb)


a
Singleton
Module
Number of eQTL loci with GWAS colocalization
0
20
40
60
21
57
40
5
22
5
21
4
11
21
iPSC-PPC singleton
Islet singleton
Whole pancreas singleton
iPSC-PPC-unique
Islet-unique
Whole pancreas-unique
Adult-shared
Fetal-islet
Fetal-whole-pancreas
Fetal-adult


b
Singleton 99% credible sets
n = 183 GWAS loci
Number of SNPs
1
2-5
6-10
11-20
21-50
51+
21
38
25
35
36
28


c
Module 99% credible sets
n = 129 GWAS loci
Number of SNPs
1
2-5
6-10
11-20
21-50
51+
15
32
22
28
21
11


a
TPD52
chr8:80998464-81093464
iPSC-PPC -log10(p-value)
Islets -log10(p-value)
Pancreas -log10(p-value)
Fasting Glucose -log10(p-value)
PP
rs12549167
2.6 x 10-5
5 x 10-8
chr8 Position (Mb)
b
CDC37L1-DT
chr9:4232083-4352083
iPSC-PPC -log10(p-value)
Islets -log10(p-value)
Pancreas -log10(p-value)
Type 1 Diabetes -log10(p-value)
Fasting Glucose -log10(p-value)
PP
rs10758593
1.9 x 10-5
5 x 10-8
chr9 Position (Mb)
c
ADSL-ST13
chr22:41049522-41449522
ADSL
ST13
iPSC-PPC -log10(p-value)
Islets -log10(p-value)
Pancreas -log10(p-value)
Cholesterol -log10(p-value)
LDL direct -log10(p-value)
PP
rs138349
3.5 x 10-5
5 x 10-8
LD (r2)
0-0.2
0.2-0.4
0.4-0.6
0.6-0.8
0.8-1
Lead
chr22 Position (Mb)


a
MEG3
chr14:101286447-101326447
iPSC-PPC -log10(p-value)
Islets -log10(p-value)
Pancreas -log10(p-value)
Type 1 Diabetes -log10(p-value)
PP
rs56994090
2.2x10-5
5x10-8
101.28 101.30 101.32 101.34
chr14 Position (Mb)
b
CDH3
chr16:68685635-68855635
iPSC-PPC -log10(p-value)
Islets -log10(p-value)
Pancreas -log10(p-value)
HbA1c (UKBB)
PP
rs72785165
3.0x10-5
5x10-8
68.5 68.6 68.7 68.8
chr16 Position (Mb)
c
HMGB1
chr13:30956642-31116642
iPSC-PPC (HMGB1.1) -log10(p-value)
iPSC-PPC (HMGB1.2) -log10(p-value)
iPSC-PPC (HMGB1.3) -log10(p-value)
Pancreas -log10(p-value)
BMI -log10(p-value)
Type 2 Diabetes -log10(p-value)
PP
rs3742305
1.3x10-5
5x10-8
30.8 30.9 31.0 31.1 31.2
chr13 Position (Mb)
LD (r2)
0-0.2
0.2-0.4
0.4-0.6
0.6-0.8
0.8-1
Lead