Title

Disease-associated regulatory variation often displays plasticity or temporal-specificity in fetal pancreas

Authors

Jennifer P. Nguyen[1,2], Timothy D. Arthur[2,3], Kyohei Fujita[4], Bianca M. Salgado[4], Margaret K.R. Donovan[1,2], iPSCORE Consortium[5], Hiroko Matsui[4], Agnieszka D'Antonio-Chronowska[6*], Matteo D'Antonio[2*], 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] A full list of Consortium members and their affiliations appears at the end of the manuscript

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

* Correspondence to: Agnieszka D'Antonio-Chronowska (adc055@health.ucsd.edu); Matteo D'Antonio (madantonio@health.ucsd.edu); Kelly A. Frazer (kafrazer@health.ucsd.edu).

Abstract

The role of genetic regulatory variation during fetal pancreas development is not well understood. 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 endocrine and exocrine eQTLs identified in adult pancreatic tissues, which reveals 2,683 eQTL associations that are unique to the fetal-like iPSC-PPCs and 1,139 eQTLs that exhibit regulatory plasticity across fetal-like and adult pancreas. Investigation of GWAS risk loci for pancreatic diseases shows that some putative causal regulatory variants are active 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 can 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 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 to disease processes [5–7]. Previous large-scale expression quantitative trait loci (eQTL) studies have made significant advancements towards 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 have been limited to adult tissues and therefore do not capture the effects of regulatory variation on gene expression under fetal conditions. In addition, integration of fetal and adult eQTL datasets will 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, could 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 pancreatic development to the health and childhood and adult onset pancreatic diseases[12,13,14,15]. 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 pancreatic development, including NEUROG3 and HNF1A, and are enriched in accessible pancreatic progenitor-specific enhancers [4,21]. To address the limited availability of fetal pancreatic tissues, protocols have been devised to efficiently guide the differentiation of human induced pluripotent stem cells (iPSCs) into pancreatic progenitor cells (iPSC-PPCs) as a model system to study the fetal pancreas [22–27]. While this model system has expanded our knowledge of pancreatic developmental biology, an eQTL study in this fetal-like pancreatic developmental stage, which requires hundreds of well-characterized iPSC-PPCs derived from different individuals, has not yet been conducted.

In this study, we derived and characterized a large resource of iPSC-PPCs and conducted an eQTL analysis to map genetic loci associated with gene expression and isoform usage in fetal-like pancreatic cells. We integrated eQTLs from adult pancreatic tissues and identified eQTL loci that displayed temporal-specificity in early pancreatic development, as well as loci that were shared with adult but displayed regulatory plasticity. Annotation of GWAS risk loci using our temporally informed eQTL resource revealed causal regulatory variants with developmental-specific 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 pancreatic development influences pancreatic disease risk and phenotypes (Figure 1A). We differentiated 106 iPSC lines from the iPSCORE resource derived from 106 whole-genome sequenced individuals to generate 107 iPSC-PPC samples (one iPSC line was differentiated twice) (Figure S1, Table S1, Table S2). We characterized the fetal-like pancreatic transcriptome as well as 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 understand the developmental-specificity and regulatory plasticity of genetic variants, we integrated eQTLs previously discovered in adult pancreatic endocrine and exocrine samples using colocalization and network analysis. Finally, using our eQTL resource of pancreatic tissues (i.e., fetal-like iPSC-derived PPCs, adult endocrine, adult exocrine), we performed GWAS colocalization and fine-mapping to link regulatory mechanisms and identify putative causal variants underlying pancreatic traits and disease associations.

Large-scale differentiation of fetal-like pancreatic progenitor cells

To assess differentiation efficiency of the 107 iPSC-PPCs, we performed flow cytometry analysis on each sample for the expression of PDX1 and NKX6-1, two markers routinely assayed for early pancreatic progenitor formation. Across the 107 samples, we observed a median percentage of PDX1+ cells of 92.1%, indicating that the majority of the cells in iPSC-PPCs had differentiated towards pancreas lineage (Figure 1B, Figure S2, Table S2). Pancreatic progenitor cells that express PDX1 further differentiate into pancreatic endoderm, which expresses both PDX1 and NKX6-1 and gives rise to both endocrine and exocrine pancreatic cell types [28]. Therefore, to determine the fraction of PPCs that differentiated into pancreatic endoderm (hereafter referred to as “late PPC”), we examined the percentage of cells expressing both PDX1 and NKX6-1 across the 107 iPSC-PPCs and found that the median percentage of $PDX1^{+}/NKX6\text{-}1^{+}$ cells was 74% (range: 9.4%-93.1%), whereas the median percentage of cells that expressed PDX1 but not NKX6-1 ($PDX1^{+}/NKX6\text{-}1^{-}$, hereafter referred to as “early PPC”) was 18.7% (range: 3.5-59.3%, Figure 1B, Figure S2, Table S2). 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 (Figure S3-8; Table S2-4; See Methods and Supplemental Note 1). Altogether, these results show that the majority of the cells in iPSC-PPCs have differentiated into pancreatic endoderm 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 [29,30], 87 pancreatic islets

[31], and 176 whole pancreatic tissues [32]. Because pancreatic islets consist primarily of endocrine cells, and whole pancreas samples consistent primarily of exocrine cells, we hereafter refer to these tissues as "adult pancreatic endocrine" and "adult pancreatic exocrine", respectively. Pseudotime analysis revealed that iPSC-PPC samples represented an intermediate stage between iPSCs and the adult pancreatic tissues, confirming that iPSC-PPC corresponded to an earlier developmental timepoint compared to the adult tissues (Figure S9, Table S5).

These analyses, combined with the results of previous studies [23,26,33], show that the 107 derived iPSC-PPCs represent a fetal-like state of pancreatic tissues, containing both pancreatic endocrine and exocrine progenitor cells.

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 mapping 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 (for genes ≥1 TPM in at least 10% of the samples or for isoforms ≥10% usage in at least 10% of samples). 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, Table S6). To detect additional independent eQTL signals [34], we performed a stepwise regression analysis to identify additional independent eQTLs (i.e., conditional eQTLs) for each eGene and eIsoform, and yielded 368 e_gQTLs (mapping to 338 eGenes) and 216 e_iQTLs (mapping to 198 eIsoforms), totaling to 4,433 independent e_gQTL associations and 4,232 independent e_iQTL associations (Figure 1C-D, Table S6). We next predicted candidate causal variants underlying each eQTL (e_gQTL and e_iQTL) association using genetic fine-mapping [35] (Table S7) 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 on each variant using the Genetic Variants Allelic TF Binding Database [36] and found that, at increasing posterior probability (PP) thresholds, the candidate causal variants underlying e_gQTLs were more likely to affect TF binding compared to those underlying e_iQTLs (Figure 1F, Table S7, Table S8). These results corroborate similar findings from previous studies [10,12,37] showing that the genetic variants underlying e_gQTLs primarily affect gene regulation and and e_iQTLs primarily affect coding regions or alternative splicing.

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 5,169 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, 53.6%) and 1,554 through only isoform usage (i.e., had only e_iQTLs, 27.7%, Figure 1G, Table S6). For the 1,008 genes with both e_gQTL and e_iQTLs, we examined whether the same or different genetic variants underpinned their associations using

colocalization. We identified 410 (40.7%) genes that had at least one H4 (model for shared causal variants; PP.H4 ≥ 80%) or H3 (model for distinct causal variants; PP.H3 ≥ 80%) association between their e_gQTL and e_iQTLs, of which the majority (333, 81.2%) had only overlapping signals (all H4), 38 (9.3%) had only non-overlapping signals (all H3), and 39 (9.5%) had both overlapping and non-overlapping e_iQTLs (both H3 and H4; an e_gQTL can overlap with an e_iQTL corresponding to one isoform but not with another e_iQTL corresponding to a second isoform) (Figure 1G, Table S9). The remaining 598 genes had PP.H3 < 80% and PP.H4 < 80% due to insufficient power (Figure 1G). Enrichment analysis of overlapping e_gQTL and e_iQTLs showed that these variants likely disrupt mechanisms affecting both gene expression and alternative splicing (Figure S10). These findings show that 17.9% of genes had both e_gQTLs, and e_iQTLs and that their effects were commonly drive by the same causal variants while a fraction were driven by different causal variants. Overall, our results show that the majority of genes had either only e_gQTLs or e_iQTLs, indicating that the functional mechanisms underlying these associations are likely independent where genetic variants affecting alternative splicing does not affect the overall expression of the gene, and vice versa.

Most fetal-like and adult endocrine eGenes show developmental stage specificity

Studies aimed at identifying and characterizing eGenes have been conducted in both adult pancreatic endocrine and exocrine tissues [8,10,11,31,38]; however, the endocrine tissue has been more thoroughly studied because of its role in diabetes. Therefore, we focused on understanding the similarities and differences between eGenes in the fetal-like iPSC-PPCs and adult pancreatic endocrine tissues.

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

For eGenes that were present in both the fetal-like iPSC-PPC and adult endocrine samples, we next asked whether their expression were controlled by the same genetic variants. We performed colocalization between e_gQTLs for the 1,501 shared eGenes in the fetal-like iPSC-PPC and adult endocrine, and found that 795 (52.3%) had either a H4 or H3 association (PP.H4 or PP.H3 ≥ 80%) (Table S9). Of the 795 with an association, 701 (88.2%) had overlapping e_gQTL signals (PP.H4 ≥ 80%) while 94 (11.8%) had non-overlapping e_gQTL signals (PP.H3 ≥ 80%) (Figure 2B, Table S9). These results indicate that most shared eGenes were associated with the same genetic variants controlling their gene expressions in both fetal-like and adult pancreatic endocrine tissues, while a subset had non-overlapping genetic

variants. For example, we identified *SNX29* as an eGene in both fetal-like iPSC-PPC and adult pancreatic endocrine but observed that its expression was associated with distinct eQTL signals approximately 520 kb apart (Figure 2C). SNX29 is involved in various signaling pathways [39], including TGF-β, ErbB, and WNT signaling pathway, and predicted to be a causal gene for body-mass index (BMI) and T2D [40].

Taken together, our results show that a minor proportion of fetal-like eGenes (1,501, 37%) were shared with adult endocrine tissues, whereas ~63% (2,564) were fetal development-specific; and, while most shared eGenes are associated with the same regulatory variants, ~12% are mediated by different eQTLs. These findings support previous observations that the chromatin landscape differs between fetal and adult involving developmental-specific enhancer-promoter interactions [41–43].

Identification of developmental-unique and shared e_gQTLs

Above, we described eGenes that were unique to fetal-like or adult endocrine, or shared between both pancreatic tissues. Here, we sought to identify eQTLs (i.e., regulatory variants) that specifically affect gene expression during pancreas development, in adult stage, or both stages. Because the iPSC-PPCs give rise to both endocrine and exocrine cell fates, we included eQTLs from both adult pancreatic endocrine [11] and pancreatic exocrine [32] tissues in our analyses. Due to the many different types of eQTLs used in this study, we refer to all eQTLs as a collective unit as "eQTLs", eQTLs that were associated with gene expression as "e_gQTLs", and eQTLs associated with changes in alternative splicing (e_iQTLs, exon eQTLs, and sQTLs) as "e_{AS}QTLs". For simple interpretations, we only describe the results for the analyses conducted on the e_gQTLs below, however, we identified unique and shared iPSC-PPC e_{AS}QTL associations by conducting the same analyses (see Supplementary Note 2).

To identify e_gQTLs that shared the same regulatory variants, we performed pairwise colocalization using *coloc* [35] between e_gQTLs in fetal-like iPSC-PPC, in adult endocrine [11], and in adult exocrine pancreatic samples [10]. We considered e_gQTLs that had at least one variant with causal PP $\geq$ 1%, outside the MHC region, and associated with genes annotated in GENCODE version 34 [44] (see Methods). We retained 4,149 fetal-like iPSC-PPC e_gQTLs, 3,948 adult endocrine e_gQTLs, and 8,312 adult exocrine e_gQTLs for downstream analyses (Table S10). We identified 7,893 total pairs of e_gQTLs that shared the same signal (PP.H4 $\geq$ 80%), which comprised 7,839 e_gQTLs (1,630 iPSC-PPC, 2,417 adult endocrine, and 3,792 adult exocrine; Figure S11A). Of the 7,893 pairs, 27.3% (2,157) were between pairs of e_gQTLs within the same pancreatic tissue associated with the expression of different eGenes and 72.7% (5,736) were between pairs of e_gQTLs active in two different pancreatic tissues (Figure S11B, Table S9). Of the 5,736 e_gQTL pairs, 43.5% (2,496) were associated with the expression of the same eGene in the two tissues while 56.5% (3,240) were associated with different eGenes (Figure S11B). Interestingly, we observed 1,301 iPSC-PPC, 902 adult endocrine, and 2,574 adult exocrine e_gQTLs that did not colocalize and were not in LD ($r^2 \geq 0.2$ and within 500 Kb) with nearby e_gQTLs, suggesting that the underlying genetic variants were associated with a single eGene and active only either

during early pancreas development or in a specific adult pancreatic tissue (Figure 3A, Table S10). Hereafter, we refer to eQTLs that did not colocalize with another eQTL and were not in LD with nearby eQTLs as "singletons" (i.e., such as the 1,301 iPSC-PPC, 902 adult endocrine, and 2,574 adult exocrine e_gQTLs described above) and those that colocalized with another eQTL (same or different tissue) as "combinatorial" (i.e., such as the 7,839 e_gQTLs described above). Given that singleton e_gQTLs were active in only the corresponding pancreatic tissue, singleton e_gQTLs were by-definition tissue-unique.

To identify combinatorial eQTL signals that were unique or shared between the three pancreatic tissues, we performed network analysis using the 7,893 pairs of colocalized e_gQTL associations to identify modules of e_gQTLs, which we defined as an eQTL signal that was either associated with multiple genes in a single tissue, or one or more genes in at least two of the three different pancreatic tissues. We identified 1,974 e_gQTL modules in total, of which 1,023 (51.8%) were composed of two e_gQTLs while the remaining 951 (48.2%) had an average of four e_gQTLs per module (range: 3-20 e_gQTLs) (Table S10, Table S11). We found that 237 (12.0% of 1,974) modules were tissue-unique (i.e., contained only e_gQTLs from one tissue and not in LD [$r^2 \geq 0.2$ and within 500 Kb] with nearby eQTLs in the other two tissues), of which 17 were fetal-like iPSC-PPC-unique, 37 adult endocrine-unique, and 183 adult exocrine-unique, and altogether comprised 35, 77, and 415 combinatorial e_gQTL associations, respectively (Figure 3A, Figure 3B, Table S10, Table S11). In contrast, the remaining 1,737 (88.0% of 1,974) modules were associated with multiple pancreatic tissues, of which 702 were shared between only adult pancreatic endocrine and exocrine tissues (referred to as "adult-shared"), 74 were shared between only iPSC-PPC and adult endocrine ("fetal-endocrine"), 309 between only iPSC-PPC and adult exocrine ("fetal-exocrine"), and 652 between all three pancreatic tissues ("fetal-adult") (Figure 3B, Table S10, Table S11). Together, the 1,035 (74 + 309 + 652) modules shared between iPSC-PPC and an adult pancreatic tissue were composed of 1,241 iPSC-PPC, 945 adult endocrine, and 1,440 adult exocrine e_gQTLs (Table S10, Table S11).

Altogether, we identified 1,336 (32.2% of 4,149) e_gQTLs that were unique to fetal-like iPSC-PPC, of which 1,301 functioned as singletons and 35 in modules, while 1,241 (29.9% of 4,149) e_gQTLs were shared with adult pancreatic tissues (Table S10, Table S11). The remaining iPSC-PPC e_gQTLs (1,572, 37.9% of 4,149) were annotated as "ambiguous" and excluded from downstream analyses due to potential associations with adult e_gQTLs based on LD and/or not meeting thresholds for module identification (see Methods, Table S10, Table S11). Our results show that the vast majority of iPSC-PPC-unique regulatory variants were singletons while combinatorial e_gQTLs tended to be shared across pancreatic issues and sometimes were associated with different eGenes. For e_{AS}QTLs, we observed similar trends in which the majority of iPSC-PPC-unique e_{AS}QTLs were singletons and that combinatorial e_{AS}QTLs were likely shared and potentially with different genes (see Supplemental Note 2; Figure S12, Table S9, Table S10).

Functional validation and characterization of tissue-unique e_gQTLs

To functionally characterize singleton and combinatorial tissue-unique e_gQTLs, we examined their enrichments in chromatin states defined in all three pancreatic tissues [7,21,45]. We observed that all three tissue-unique singleton e_gQTLs were enriched in active chromatin regions in their respective tissues, including enhancers, promoters, and the sequences corresponding to flanking promoters (Figure 3C, Figure S11C, Table S12). Adult endocrine-unique and exocrine-unique combinatorials were also enriched in active chromatin states but had stronger preferences for enhancers, which is consistent with the characteristic of enhancers in regulating multiple genes (Figure 3C, Figure S11C, Table S12). Of note, iPSC-PPC-unique combinatorial e_gQTLs were enriched in quiescent and genic enhancer regions in adult endocrine. Similarly, adult endocrine- and exocrine-unique singleton e_gQTLs were enriched in active regulatory regions in PPC but were not detected as e_gQTLs in iPSC-PPC. For example, in the chr2:198053627-198143627 locus overlapping an active PPC enhancer, we observed that the variants were associated with an e_gQTL signal only in adult endocrine (Figure S11D). For these eQTLs that overlap an active regulatory element in a different tissue but do not affect gene expression, it may be possible that these variants act through the disruption of tissue-unique TF binding [43]. Altogether, our results demonstrate that tissue-unique singleton and combinatorial e_gQTLs were strongly enriched for active chromatin regions with combinatorial e_gQTLs having the strongest preference for enhancers as observed in adult.

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

Regulatory plasticity in e_gQTL signals shared between fetal-like and adult pancreatic tissues

Above, we demonstrated that genetic variants can exhibit temporal-specificity between fetal-like and adult tissues. Next, we sought to examine regulatory variants that are shared between the two stages and understand how their function changes. Using the 1,035 e_gQTL modules shared between fetal-like iPSC-PPC and adult pancreatic tissues, we next sought to understand how genetic variant function changes between the two developmental stages. Specifically, we asked whether the underlying e_gQTL signals in the modules were associated with the same or different eGenes in the three pancreatic tissues. We identified the following five categories (Figure 4A, Table S11): A) 230

(22.2%) e_gQTL modules were associated with same eGene(s) (range: 1-2) between fetal-like iPSC-PPC and only one of the two adult pancreatic tissues; B) 305 (29.5%) were associated with the expression of the same eGene(s) (range 1-2) in the fetal-like and both adult tissues; C) 363 (35.1%) were associated with 2-9 eGenes, some of which were shared, but at least one eGene was different between the fetal-like and at least one of the adult tissues (referred to as "partial overlap"); D) 97 (9.4%) were associated with different eGenes (range: 2-5) between fetal-like iPSC-PPCs and one of the two adult pancreatic tissues; and E) the remaining 40 (3.9%) were associated with different eGenes (range: 2-8) between the fetal-like and both adult endocrine and exocrine tissues (i.e., there is no overlap of eGenes between the two developmental stages).

Here, we illustrate examples of e_gQTL modules in three intervals to highlight how eQTL associations varied between fetal-like and adult states. In the chr11:111505862-112155862 locus, we discovered a fetal-adult e_gQTL module (GE_11_69) that comprised e_gQTL associations with different eGenes in iPSC-PPC and the two adult pancreatic tissues, specifically *CRYAB* in iPSC-PPC and *C11orf1* in the two adult tissues (Figure 4B). Likewise, the chr19:4213666-4433666 locus corresponding to a fetal-adult e_gQTL module (GE_19_90) was associated with *MPND* expression in only iPSC-PPC but in adult pancreatic endocrine and exocrine, the underlying variants were associated with *STAP2* expression (Figure 4C). Finally, the fetal-adult e_gQTL locus (GE_10_11) in chr10:1273918-1276118 affected *UROS* expression in all three pancreatic tissues but in adult pancreatic endocrine, the underlying variants also affected *BCCIP* expression (Figure 4D). Together, these genomic loci illustrate examples of regulatory plasticity observed in genetic variants in which their genotypes incur different transcriptional phenotypes depending on the life stage of the pancreas.

Taken together, our findings reveal that 48.3% of shared e_gQTL loci (n = 500; categories C-E) comprising 691 iPSC-PPC, 578 adult endocrine, and 959 adult exocrine e_gQTL associations display regulatory plasticity in which the underlying regulatory variants are associated with one or more different eGenes and could thereby affect different biological processes. For e_{AS}QTLs, we found that 39.8% e_{AS}QTL loci (n = 208; categories C-E) are shared between fetal-like and adult pancreas and associated with multiple different genes, comprising 448 iPSC-PPC, 384 adult endocrine, and 217 adult exocrine e_{AS}QTLs (see Supplemental Note 2).

Associations of developmental stage-unique eQTLs with pancreatic traits and disease phenotypes

To better understand the role of regulatory variants associated with complex human traits and disease during early development and adult pancreatic stages, we performed colocalization between GWAS signals and eQTLs (e_gQTL and e_{AS}QTL) detected in fetal-like iPSC-PPC, adult endocrine, and adult exocrine tissues. For this analysis, we considered GWAS data from ten different studies for two diseases involving the pancreas, including type 1 diabetes (T1D) [3] and type 2 diabetes (T2D) [4], and seven biomarkers related to three traits: 1) glycemic control (HbA1c levels and fasting

277 glucose) [2,46]; 2) obesity (triglycerides, cholesterol, HDL level, and LDL direct) [46]; and 3) body mass index (BMI) [46]
278 (Table S13).

279 *Singleton eQTLs*

280 Out of the 8,137 singleton eQTLs (4,777 e_gQTLs and 3,360 e_{AS}QTLs see Supplemental Note 2) in the fetal-like iPSC-
281 PPC and two adult pancreatic tissues, we found 164 (2%) that displayed strong evidence for colocalization with at least
282 one GWAS signal, including 30 (of 2,205 total singleton eQTLs; 1.4%) fetal-like iPSC-PPC, 71 (of 2,705; 2.6%) adult
283 endocrine, and 63 (of 3,227; 2.0%) adult exocrine singleton eQTLs (Figure 5A, Figure S13, Table S13). Given that
284 some traits are highly correlated with one another [47,48], we observed 49 eQTLs that colocalized with GWAS variants
285 associated with more than one trait (average: 1.5 traits per singleton eQTL; range: 1-6 studies). In total, we identified
286 248 GWAS loci (across the ten GWAS studies) that displayed colocalization with fetal-like or adult pancreatic
287 singleton eQTLs (Table S13). We next identified putative causal variants underlying both eQTL and trait associations
288 and constructed 99% credible sets where the cumulative causal PP > 99% (see Methods). Of the total 248 GWAS loci,
289 we were able to resolve 34 loci to a single putative causal variant while 84 had between two and ten variants and the
290 remaining 130 had more than ten variants with an average of ~45 variants per locus (Figure 5B, Table S14).

291 *eQTL modules*

292 We next analyzed the combinatorial eQTLs (i.e., eQTLs that colocalize with one another) for GWAS colocalization.
293 We considered an eQTL module to overlap with GWAS variants if more than 30% of the eQTLs in the module
294 colocalized with PP.H4 > 80% and the number of H4 associations were twice greater than the number of H3
295 associations (see Methods). Of the 3,185 (1,974 e_gQTL and 1,211 e_{AS}QTL) modules, 105 (63 e_gQTL + 42 e_{AS}QTL;
296 3.3%) colocalized with a total of 149 GWAS signals (Table S13). Of these 105 GWAS-colocalized modules, 9 were
297 associated with only fetal-like iPSC-PPC eQTLs, 42 were shared between both iPSC-PPC and adult, (5 fetal-endocrine,
298 16 fetal-exocrine and 21 fetal-adult modules), and 54 were associated with only adult eQTLs (23 endocrine-unique, 8
299 exocrine-unique, 23 adult-unique) (Figure S13, Table S13). These 105 modules were composed of 49 iPSC-PPC
300 eQTLs, 84 adult endocrine eQTLs, and 49 adult exocrine eQTLs. Interestingly, we observed that all 9 of the fetal-
301 unique modules corresponded to e_{AS}QTL modules, which aligns with previous observations that alternative splicing is
302 overall more prominent in fetal compared with adult tissues and thus, tends to be highly developmental stage-specific
303 [49]. To obtain 99% credible sets for each of the 149 GWAS signals that colocalized with an eQTL module, we focused
304 on the eQTL association that resulted in the least number of putative causal variants (see Methods). 17 GWAS loci had
305 a credible set size of one variant, 58 with two to ten variants, and the remaining 74 had more than ten variants and an
306 average of ~34 variants per set (Figure 5C, Table S14).

307 In summary, we identified 79 eQTLs in iPSC-PPC (30 singleton + 49 combinatorial) that colocalized with GWAS
308 variants associated with complex pancreatic traits and disease, 30 of which we found to function as singleton eQTLs
309 (i.e., affect fetal-specific expression or alternative splicing of a single gene) while 49 were combinatorial eQTLs (i.e.,

affect expression or alternative splicing of multiple genes or isoforms). Of these combinatorial eQTLs, 16 were fetal-unique (i.e., only colocalized with iPSC-PPC eQTLs) while 33 were adult-shared (i.e., colocalized with an adult pancreatic eQTL). Furthermore, we observed that all 9 fetal-unique eQTL modules exclusively affected alternative splicing, indicating that fetal-unique regulatory variants associated with disease may likely affect splice mechanisms rather than gene transcription, consistent with previous studies demonstrating widespread alternative splicing during embryonic development [49–51].

Interpreting mechanisms of fine-mapped GWAS signals

Fetal pancreatic tissues are not typically assessed for GWAS annotation, so the role of developmental regulatory variants, and how their function changes in adult, is currently unclear. To better understand the function of disease-associated variants in both fetal-like and adult pancreatic contexts, we used our previous assessment of eGene overlap between the two stages to annotate GWAS loci. While GWAS loci that colocalized with fetal-adult shared e_gQTL modules were more likely to modulate the expression of the same genes (75%; n = 24), ~25% displayed different regulatory functions. Specifically, three loci were associated with partially overlapping eGenes (category C) and five were associated with entirely different eGenes (category D and E) (Table S13). Similarly, for GWAS loci that colocalized with fetal-adult shared e_{AS}QTL modules, 90% (9/10) were associated with splice changes of the same gene (categories A and B) while 10% (1/10) was associated with at least one different gene between fetal-like and adult (category C) (Table S13). These results show that while the function of shared GWAS regulatory variants is likely conserved across fetal-like and adult pancreatic stages, a subset (21.4%, n = 9) are associated with distinct genes between the two stages.

In total, we identified 397 GWAS loci (248 singleton and 149 module) colocalized with fetal-like and/or adult pancreas eQTLs. To demonstrate the power of the pancreas eQTL resource that we have generated, below we describe how our findings have contributed to biological insights in eight GWAS loci for pancreatic traits and diseases. Our examples show that our study provides putative causal mechanisms and temporal context underlying genetic associations with pancreatic complex traits and disease. We further demonstrate the regulatory plasticity of GWAS variants to produce different transcriptional effects on gene expression between fetal-like and adult pancreas.

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

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

We found that in the chr8:80998464-81093464 locus, a GWAS signal associated with FG levels colocalized with a fetal-like iPSC-PPC-unique singleton e_gQTL for *TPD52*, also known as tumor protein D52 (effect size = -0.99, PP.H4 = 91.7%) (Figure 6A, Figure S14A, Table S13). The reported causal variant underlying this GWAS signal is

rs12541643 [2]; however, colocalization with our eQTLs identified rs12549167 (chr8:81078464:C>T, PP = 33.9%, r^2 = 0.317 with rs12541643, Table S14) as the most likely candidate causal variant underlying both *TPD52* expression in fetal-like iPSC-PPC and FG association. TPD52 directly interacts with the AMP-activated protein kinase and negatively affects AMPK signaling. AMPK controls a wide range of metabolic processes and is responsible for maintaining cellular energy homeostasis particularly in tissues associated with obesity, insulin resistance, T2D, and cancer such as muscle, liver, hypothalamus, and the pancreas [52–55]. Dysregulation of AMPK has also been associated with developmental defects in which AMPK activation can lead to fetal malformation [56]. Our findings suggest that decreased expression of *TPD52* during development may influence changes in glucose metabolism and therefore fasting glucose levels in adult.

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

We found that the well-known *GLIS3* GWAS locus associated with FG and T1D-risk [57,58] colocalized with a fetal-like iPSC-PPC-unique singleton e_gQTL for the lncRNA CDC37L1 divergent transcript (*CDC37L1-DT*; effect size = 1.46; PP.H4 for FG and T1D = 92.4% and 91.2%, respectively, Figure 6B, Figure S14B, Table S13). Consistent with previous studies [57,58], we identified rs10758593 (chr9:4292083:G>A, PP = 79.2%) as the lead candidate causal variant underlying both eQTL and GWAS associations. Because *GLIS3* plays a critical role in pancreatic beta cell development and function [58,59,60], it has often been reported as the susceptibility gene for this signal, however it remains unclear what effects rs10758593 has on *GLIS3* expression. Our analysis suggests that another potential gene target of rs10758593 during pancreas development is *CDC37L1-DT*. While the molecular function of *CDC37L1-DT* is unknown, the gene has been associated with 9p duplication in neurodevelopmental disorders [62]. Furthermore, a recent study observed a significant association between the rs10758593 risk allele and birth weight, indicating a development role played by this locus [63]. Although additional studies are needed to understand the function of *CDC37L1-DT* during pancreas development and in T1D pathology, our analysis indicates that *CDC37L1-DT* may be another candidate susceptibility gene for the variants in the *GLIS3* locus. Assessment of *GLIS3* e_gQTLs in the three pancreatic tissues showed that that is no overlap between the e_gQTLs and GWAS variants (Figure S15A).

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

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

We found that the GWAS signals associated with cholesterol and LDL direct levels in the chr22:41049522-41449522 locus [64] colocalized with a "fetal-adult" e_gQTL module (module ID: GE_22_63, category E) (Figure 6C, Figure S14C-D, Table S13). The module was associated with different eGenes between fetal-like iPSC-PPC and both adult pancreatic tissues, in which the GWAS variants were associated with *ADSL* expression in iPSC-PPC (effect size = 0.78) but *ST13* expression in both adult pancreatic tissues (effect size = -0.15 in adult endocrine and 0.27 in adult

exocrine). Infants born with ADSL (adenylosuccinate lyase) deficiency suffer from impaired glucose and lipid metabolism while ST13, also known as Hsc70-interacting protein, is involved in lipid metabolism [65]. Overexpression of *ST13* was found to result in disordered lipid metabolism in chronic pancreatitis [65]. Although *ST13* was reported to be the candidate causal gene for this locus [64], we determined that the underlying variants may also affect *ADSL* expression but specifically during early pancreas development. Congruent with the previous study [64], our colocalization identified rs138349 (chr22:41249522:A>G, PP = 21.9%) as the lead candidate causal variant for the e_gQTLs and both cholesterol and LDL GWAS associations (Table S14). Altogether, annotation of the chr22:41049522-41449522 GWAS locus using our pancreatic eQTL resource suggests that altered expression of *ADSL* during pancreas development and *ST13* in adult tissues may contribute to changes in cholesterol and LDL direct levels in adult. Additional studies are required to understand the degree to which *ADSL* and *ST13* are causal for cholesterol and LDL direct levels.

chr10:90001035-90066035 and PTEN and LIPJ (adult exocrine-unique combinatorial)

We found a T1D-risk signal in the chr10:90001035-90066035 locus that colocalized with an "adult exocrine-unique" e_gQTL module (module ID: GE_10_35) associated with *PTEN* and *LIPJ* expression in adult pancreatic exocrine (effect size = 0.48 and 0.49, respectively) (Figure S15B, Figure S14E, Table S13). Colocalization identified the distal regulatory variant rs7068821 (chr10:90051035:G>T; PP = 85.5%) as the most likely candidate causal variant (Table S14), which is in LD with the reported index SNP rs10509540 (r^2 = 0.876) in the GWAS catalogue. While *RNLS* was reported to be the susceptibility gene for this locus [66], our analysis suggests that *PTEN* and *LIPJ* may be candidate causal genes for this locus. Previous studies have shown that pancreas-specific *PTEN* knockout (PPKO) mice resulted in enlarged pancreas and elevated proliferation of acinar cells. PPKO mice also exhibited hypoglycemia, hypoinsulinemia, and altered amino metabolism [67]. LIPJ encodes the lipase family member J and is involved in lipid metabolism [68]. Our findings provide additional biological insight into the chr10:900001035-90066035 T1D locus and support previous studies suggesting a potential causal role of the adult exocrine pancreas in T1D pathogenesis [3,63].

Singleton e_{AS}QTLs: Here, we illustrate three examples of putative causal variants involved in alternative splicing in the fetal-like pancreas. Long-noncoding RNAs (lncRNAs) have previously been shown to play important roles in pancreatic diseases [69]. Two of our examples include lncRNAs while one is a protein-coding gene.

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

The chr14:101286447-101326447 is a well-known GWAS locus associated with T1D and has been reported to affect the lncRNA maternally expressed gene 3 (*MEG3*). While the role of *MEG3* in T1D and T2D pathogenesis has been extensively studied [70–72], the genetic mechanism by which this locus affects *MEG3* expression and therefore, T1D risk 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, Figure S16A, Table S13). 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 [73] (Table S14). 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 [74], 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 [75] colocalized with a fetal-like iPSC-PPC-unique singleton e_iQTL for the P-cadherin 3 (*CDH3*) isoform ENST00000429102 (effect size = -1.6, PP.H4 = 83.1%) (Figure 7B, Figure S16B, Table S13). Colocalization using the e_{AS}QTL identified intronic variant rs72785165 (chr16:68755635:T>A, PP = 6.8%) as the most likely candidate causal variant (Table S14), which is in high LD with the reported GWAS SNP (rs4783565, $r^2 = 0.88$) [75]. While no studies have examined 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 [76]. 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.

Combinatorial e_{AS}QTLs: Here, we present potential causal mechanisms during pancreas development that involve alternative splicing and are associated with T2D-risk and BMI.

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

The GWAS signals associated with T2D and BMI in the chr13:30956642-31116642 locus [77–80] colocalized with the iPSC-PPC-unique e_{AS}QTL module (module ID: AS_13_2) associated with three *HMGB1* isoforms: ENST00000326004, ENST00000339872, and ENST00000399494 (effect size = 2.16, -0.85, and -2.26, respectively) (Figure 7C, Figure S16C-E, Table S13). Our colocalization identified rs3742305 (chr13:31036642:C>G, PP = 49.3%) as the 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 (Figure S16C-E, Table S6, Table S14). While a previous study [79] 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 [81]. 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.

Altogether, our findings demonstrate the value of our pancreatic eQTL resource to annotate GWAS risk variants with fetal-like and adult temporal and regulatory information. We show that some causal regulatory variants underlying

disease-associated signals may influence adult traits by modulating the expression of genes in early development, while in other cases, they may display regulatory plasticity and exert their effects by modulating the expression of multiple different genes in fetal-like and adult pancreatic stages. Further, we identified an association between exocrine pancreas and T1D, supporting a potential role of this tissue in diabetes pathogenesis [3].

Discussion

In this study, we leveraged one of the most well-characterized iPSC cohorts comprising >100 genotyped individuals to derive pancreatic progenitor cells and generate a comprehensive eQTL resource for examining genetic associations with gene expression and isoform usage in fetal-like pancreatic cells. We discovered 8,665 eQTLs in the fetal-like iPSC-PPCs and showed that 60% of eGenes were associated with regulatory variation specifically active during pancreas development. For the eGenes that were shared with adult, ~12% were regulated by different genomic loci, indicating that different regulatory elements modulate the same gene in fetal-like and adult pancreas. We further identified regulatory variants that displayed developmental-specific function, 70% of which were uniquely active in only iPSC-PPC while in other cases, the variants were active in both developmental and adult contexts but exhibited regulatory plasticity in the genes they regulate. These results concur with previous studies showing that the genetic regulatory landscape changes between fetal tissues and their adult counterparts [41,43,83], and therefore, highlights the importance of assessing variant function in both fetal and adult tissue contexts. Furthermore, it is widely known that tight regulation of genes during development is essential [84], and our study reflects this in our findings that the majority of developmental-unique eQTLs were restricted to a single eGene. Because conditional associations were not readily available for the adult pancreatic tissues, additional analyses are required to recapitulate our findings.

Finally, we highlighted examples of GWAS associations for which we utilized our temporally informed eQTL resource to characterize novel causal risk mechanisms underlying adult pancreatic disease. We showed that some causal regulatory variants underlying GWAS signals identified in the fetal-like iPSC-PPCs modulate the expression of genes in early development, while others may exert their effects by modulating the expression of multiple different genes across fetal-like and adult pancreatic stages. Of note, many of the fetal-unique regulatory variants underlying the GWAS signals were e_{AS}QTLs, which is consistent with alternative splicing playing a key role in developing tissues [49–51,85]. Hence, we believe that contribution of alternative splicing differences during fetal pancreas development to complex traits warrants further investigation given the novel biology presented in our results.

We offer limitations in our study and potential future directions for the field at large. We believe that studies using larger sample sizes are needed to identify additional associations between genetic variation and gene expression in fetal samples. Our eQTL mapping in iPSC-PPC was conducted on much fewer samples compared to the other two studies that used ~400 samples, rendering our dataset underpowered and not being able to capture weaker eQTL associations that could be shared with the adult pancreatic tissues. Therefore, some of the eQTLs we annotated as adult endocrine-

unique or exocrine-unique may in reality be shared with fetal pancreas. On the other hand, the eQTLs we annotated as iPSC-PPC-unique are less likely to be shared, as the signals in the adult datasets are better powered and therefore sufficient for comparing against iPSC-PPC signals. Additionally, with the rapid generation of eQTL datasets from different tissue contexts [1,2], the development and application of artificial intelligence and machine learning as ways to identify shared eQTL associations between multiple tissues will be extremely useful. While pairwise colocalization and network analysis was able to identify shared eQTL regulatory loci across the fetal-like and two adult pancreatic tissues in our study, machine learning approaches would enable these analyses to scale across spatiotemporal contexts of all tissues and thereby, provide insights into regulatory elements that are unique to a specific context, as well as those that display regulatory plasticity across multiple contexts.

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

Methods

Subject Information

We used iPSC lines from 106 individuals recruited as part of the iPSCORE project (Table S1). There were 53 individuals belonging to 19 families composed of two or more subjects (range: 2-6). Each subject was assigned an iPSCORE_ID (i.e., iPSCORE 4_1), where "4" indicates the family number and "1" indicates the individual number, and a 128-bit universal unique identifier (UUID). The 106 individuals included 68 females and 38 males with ages ranging from 15 to 88 years old at the time of enrollment. Recruitment of these individuals was approved by the Institutional Review Boards of the University of California, San Diego, and The Salk Institute (project no. 110776ZF). Each of the subjects provided consent to publish information for this study.

iPSC Generation

Generation of the 106 iPSC lines has previously been described in detail [30]. Briefly, cultures of primary dermal fibroblast cells were generated from a punch biopsy tissue [87], infected with the Cytotune Sendai virus (Life Technologies) per manufacturer's protocol to initiate reprogramming. Emerging iPSC colonies were manually picked after Day 21 and maintained on Matrigel (BD Corning) with mTeSR1 medium (Stem Cell Technologies). Multiple independently established iPSC clones (i.e. referred to as lines) were derived from each individual. Many of the iPSC

lines were evaluated by flow cytometry for expression of two pluripotent markers: Tra-1-81 (Alexa Fluor 488 anti-human, Biolegend) and SSEA-4 (PE anti-human, Biolegend) [30]. Pluripotency was also examined using PluriTest-RNAseq [30]. This iPSCORE resource was established as part of the Next Generation Consortium of the National Heart, Lung and Blood Institute and is available to researchers through the biorepository at WiCell Research Institute (www.wicell.org; NHLBI Next Gen Collection). For-profit organizations can contact the corresponding author directly to discuss line availability.

Pancreatic Progenitor Differentiation

We performed pancreatic progenitor cell (PPC) differentiation on each of the 106 iPSC lines. One iPSC line was differentiated twice giving a total of 107 differentiations. Each differentiation was assigned a 128-bit universally unique identifier (UUID), and a unique differentiation ID (UDID; "PPCXXX"), where "XXX" represents a numeric integer (Table S2).

Differentiation Protocol

The iPSC lines were differentiated into PPCs using the STEMdiff™ Pancreatic Progenitor Kit (StemCell Technologies) protocol with minor modifications. Briefly, iPSC lines were thawed into mTeSR1 medium containing 10 µM Y-27632 ROCK Inhibitor (Selleckchem) and plated onto one well of a 6-well plate coated with Matrigel. iPSCs were grown until they reached 80% confluency [88] and then passaged using 2mg/ml solution of Dispase II (ThermoFisher Scientific) onto three wells of a 6-well plate (ratio 1:3). To expand the iPSC cells for differentiation, iPSCs were passaged a second time onto six wells of a 6-well plate (ratio 1:2). When the iPSCs reached 80% confluency, cells were dissociated into single cells using Accutase (Innovative Cell Technologies Inc.) and resuspended at a concentration of 1.85 x 10^6 cells/ml in mTeSR medium containing 10 µM Y-27632 ROCK inhibitor. Cells were then plated onto six wells of a 6-well plate and grown for approximately 16 to 20 hours to achieve a uniform monolayer of 90-95% confluence (3.7 x 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 STEMdiff™ Stage Endoderm Basal medium supplemented with Supplement MR and Supplement CJ (2 ml/well) (Day 1, D1). The following media changes were performed every 24 hours following initiation of differentiation (2 ml/well). On D2 and D3, the medium was changed to fresh STEMdiff™ Stage Endoderm Basal medium supplemented with Supplement CJ. On D4, the medium was changed to STEMdiff™ Pancreatic Stage 2-4 Basal medium supplemented with Supplement 2A and Supplement 2B. On D5 and D6, the medium was changed to STEMdiff™ Pancreatic Stage 2-4 Basal medium supplemented with Supplement 2B. From D7 to D9, the medium was changed to STEMdiff™ Pancreatic Stage 2-4 Basal medium supplemented with Supplement 3. From D10 to D14, the medium was changed to STEMdiff™ Pancreatic Stage 2-4 Basal medium supplemented with Supplement 4. On D15, cells were dissociated with Accutase and then collected, counted, and processed for data generation. iPSC-PPC cells were cryopreserved in CryoStor® CS10 (StemCell Technologies).

iPSC-PPC Differentiation Efficiency

To evaluate the efficiency of iPSC-PPC differentiation, we performed flow cytometry on two pancreatic precursor markers, PDX1 and NKX6-1. Specifically, at least 2 x 10^6 cells were fixed and permeabilized using the Fixation/Permeabilized Solution Kit with BD GolgiStop TM (BD Biosciences) following the manufacturer's recommendations. Cells were resuspended in 1x BD Perm/Wash TM Buffer at a concentration of 1 x 10^7 cells/ml. For each flow cytometry staining, 2.5 x 10^5 cells were stained for 75 minutes at room temperature with PE Mouse anti-PDX1 Clone-658A5 (BD Biosciences; 1:10) and Alexa Fluor® 647 Mouse anti-NKX6.1 Clone R11-560 (BD Bioscience; 1:10), or with the appropriate class control antibodies: PE Mouse anti-IgG1 κ R-PE Clone MOPC-21 (BD Biosciences) and Alexa Fluor® 647 Mouse anti IgG1 κ Isotype Clone MOPC-21 (BD Biosciences). Stained cells were washed three times, resuspended in PBS containing 1% BSA and 1% formaldehyde, and immediately analyzed using FACS Canto II flow cytometer (BD Biosciences). The fraction of PDX1- and NKX6-1-positive was calculated using FlowJo software version 10.4 (Table S2).

WGS data

Whole-genome sequencing data for the 106 iPSCORE individuals were downloaded from dbGaP (phs001325) as a VCF file [29]. We retained variants with MAF > 5% across all 273 individuals in the iPSCORE resource, that were in Hardy-Weinberg equilibrium ($p > 10^{-6}$), and that were within 500 Kb of the expressed gene's body coordinates. Specifically, we expanded the coordinates of each of the 16,464 expressed autosomal genes (500 Kb upstream and downstream) and extracted all variants within these regions using *bcftools view* with parameters *--f PASS -q 0.05:minor* [89]. Next, we normalized indels and split multi-allelic variants using *bcftools norm -m-* and removed variants that were genotyped in fewer than 99% of samples using *bcftools filter -i 'F_PASS(GT!="mis") > 0.99* [89]. Finally, we converted the resulting VCF files to text using *bcftools query* [89] and converted the genotypes from character strings (0/0, 0/1, and 1/1) to numeric (0, 0.5, and 1, respectively). This resulted in 6,593,484 total variants used for eQTL mapping.

Bulk RNA-seq

Library Preparation and Sequencing

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

Data Processing and Quality Control

FASTQ files were obtained for all 107 iPSC-PPC samples and processed using a similar pipeline described in our previous studies [29,90]. Specifically, RNA-seq reads were aligned with STAR (2.7.3) [91] to the hg19 reference using GENCODE version 34 hg19[92] splice junctions with default alignment parameters and the following adjustments: *-outFilterMultimapNmax 20, -outFilterMismatchNmax 999, -alignIntronMin 20, -alignIntronMax 1000000, -alignMatesGapMax 1000000*. BAM files were sorted by coordinates, and duplicate reads were marked using Samtools (1.9.0) [89]. RNA-seq QC metrics were calculated using Samtools (1.9.0) flagstat [89], Samtools (1.9.0) idxstats [89], and Picard (2.20.1) CollectRnaSeqMetrics [93]. Across all 107 iPSC-PPC samples, the total read depth ranged from 32.3 M to 160.4 M (mean = 70.7), the median percentage of intergenic bases was 3.31%, the median percentage of mRNA bases was 92.1%, and the median percentage of duplicate reads was 22.2% (Table S2).

Sample Identity

We obtained common bi-allelic variants from the 1000 Genomes Phase 3 panel [94] with minor allele frequencies between 45% and 55% and predicted their genotypes in the 107 bulk RNA-seq samples using *mpileup* and *call* functions in BCFtools (1.9.0) [95,96]. Then, we used the *genome* command in plink [93] to estimate the identity-by-state (IBS) between each pair of bulk RNA-seq and WGS samples. All RNA-seq samples were correctly matched to the subject with PI_HAT > 0.95 (Table S2).

Quantification of gene expression and relative isoform usage

We calculated TPM and estimated relative isoform usage for each gene in each RNA-seq sample using RSEM (version 1.2.20) [97] with the following options *–seed 3272015 –estimate-rspd –paired-end –forward-prob*. To identify expressed autosomal genes and isoforms to use for eQTL analyses, we used the same approach previously described [12]. Briefly, autosomal genes were considered expressed if TPM ≥ 1 in at least 10% of samples. To identify expressed isoforms, we required that isoforms had TPM ≥ 1 and usage ≥ 10% in at least 10% of samples and corresponded to expressed genes with at least two expressed isoforms. In total, 16,464 autosomal genes were used for e_gQTL analysis, and 29,871 autosomal isoforms corresponding to 9,624 genes were used for e_iQTL analysis. We quantile-normalized TPM and isoform usage across all 107 samples using the *normalize.quantiles* (preprocessCore) and *qnorm* functions in R (version 4.2.1) to obtain a mean expression = 0 and standard deviation = 1.

Inferring pseudotime using Monocle

We obtained FASTQ files for 213 iPSCs [29,30] (phs000924), 176 adult pancreatic exocrine [8] (phs000424), and 87 adult pancreatic endocrine [31] (GSE50398), and processed the data using the same pipeline described above to obtain TPM counts for each gene per sample. We then used Monocle (http://cole-trapnell-lab.github.io/monocle-release/docs/#constructing-single-cell-trajectories) [98] to infer the pseudotime on all of

the RNA-seq samples, including the 107 iPSC-PPCs. Following the standard workflow under "Constructing Single Cell Trajectories" in the Monocle tutorial, we provided TPM counts for all overlapping autosomal expressed genes in the four tissues as input. Then, we identified differentially expressed genes using *differentialGeneTest*, ordered them (*setOrderingFilter*), and performed dimension reduction analysis using *reduceDimension* with *max_components = 2* and *method = "DDRTree"*. Pseudotime was calculated by rooting time (pseudotime = 0) in the 213 iPSC-PPCs using the *GM_state* and *orderCells* functions provided in the tutorial (Table S5).

PCA analysis with iPSCs, adult pancreatic exocrine, and adult pancreatic endocrine

We obtained TPM counts (described above) for the 213 iPSCs [29], 176 adult exocrine [8], 87 adult endocrine [31], and the 107 iPSC-PPCs and performed PCA analysis on the 2,000 most variable genes across the samples using *prcomp* in R (version 4.2.1) with *scale = T* and *center = T*. We observed that the PC clusters corresponded to the iPSCs and each of the three pancreatic tissue types: iPSC-PPC, adult endocrine, and adult exocrine (Figure S9, Table S5).

scRNA-seq

To characterize the cellular composition of the fetal-like iPSC-PPC samples, we performed single-cell RNA-seq (scRNA-seq) on one iPSC line (from differentiation PPC034) and ten iPSC-PPC samples with varying percentages of double-positive PDX1+/NKX6-1+ cells based on flow cytometry (range: 9.4-91.7%) (Figure S2, Figure S3, Table S2). Because bulk RNA-seq was generated on cryopreserved cells, we sought to also examine whether cell cryopreservation affects gene expression estimates using scRNA-seq. Therefore, we included both freshly prepared (i.e., not frozen and processed immediately after differentiation) and cryopreserved cells for four iPSC-PPC samples (PPC029, PPC027, PPC023, PPC034; Table S2) for scRNA-seq processing.

Sample Collection

Fresh cells from the iPSC line and seven iPSC-PPC samples were captured individually at D15. Cells from four of these same iPSC-PPC samples that had been cryopreserved were pooled and captured immediately after thawing (RNA_Pool_1). Cells from an additional three iPSC-PPC samples were captured only after cryopreservation (RNA_Pool_2) (Table S2).

Library Preparation and Sequencing

All single cells were captured using the 10X Chromium controller (10X Genomics) according to the manufacturer's specifications and manual (Manual CG000183, Rev A). Cells from each scRNA-seq sample (one iPSC, seven fresh iPSC-PPCs, RNA_Pool_1, and RNA_Pool_2) were loaded each onto an individual lane of a Chromium Single Cell Chip B. Libraries were generated using Chromium Single Cell 3' Library Gel Bead Kit v3 (10X Genomics) following manufacturer's manual with small modifications. Specifically, the purified cDNA was eluted in 24 µl of Buffer EB,

half of which was used for the subsequent step of the library construction. cDNA was amplified for 10 cycles and libraries were amplified for 8 cycles. All libraries were sequenced on a HiSeq 4000 using custom programs (fresh: 28-8-175 Pair End and cryopreserved: 28-8-98 Pair End). Specifically, eight libraries generated from fresh samples (one iPSC and seven iPSC-PPC samples) were pooled together and loaded evenly onto eight lanes and sequenced to an average depth of 163 million reads. The two libraries from seven cryopreserved lines (RNA_Pool_1 and RNA_Pool_2) were each sequenced on an individual lane to an average depth of 265 million reads. In total, we captured 99,819 cells. We observed highly correlated cell type proportions between fresh and cryopreserved iPSC-PPC samples (Figure S8).

scRNA-seq Alignment

We obtained FASTQ files for the ten scRNA-seq samples (one iPSC, seven fresh iPSC-PPCs, RNA_Pool_1, and RNA_Pool_2) (Table S2) and used CellRanger V6.0.1 (https://support.10xgenomics.com/) with default parameters and GENCODE version 34 hg19 [92] gene annotations to generate single-cell gene counts and BAM files for each of the ten samples.

Dataset Integration and Quality Control

We processed the single-cell gene counts by first aggregating the iPSC and seven fresh iPSC-PPC samples using the *aggr* function on CellRanger V6.0.1 with normalization = F. Then, we integrated the aggregated dataset ("aggr") with the two pools of cryopreserved samples (RNA_Pool_1 and RNA_Pool_2) using the standard integration workflow described in Seurat (version 3.2; https://satijalab.org/seurat/archive/v3.2/integration.html). Specifically, for each dataset (aggr, RNA_Pool_1, and RNA_Pool_2), we log-normalized the gene counts using *NormalizeData* (default parameters) then used *FindVariableFeatures* with *selection.method = "vst", nfeatures = 2000,* and *dispersion.cutoff = c(0.5, Inf)* to identify the top 2,000 most variable genes in each dataset. We then used *FindIntegrationAnchors* and *IntegrateData* with *dims = 1:30* to integrate the three datasets. We scaled the integrated data with *ScaleData*, performed principal component analysis with *RunPCA* for *npcs = 30*, and processed for UMAP visualization (*RunUMAP* with *reduction = "pca"* and *dims = 1:30*). Clusters were identified using *FindClusters* with default parameters.

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

Demultiplexing Sample Identity

We used Demuxlet [99] to assign pooled cryopreserved cells in RNA_Pool_1 and RNA_Pool_2 (19,136 cells in total) to the correct iPSC-PPC sample. Specifically, we provided CellRanger Bam files and a VCF file containing genotypes for biallelic SNVs located at UTR and exon regions on autosomes as annotated by GENCODE version 34 hg19 [92]. We excluded 33 cells that were incorrectly assigned to samples not associated with the pooled sample (i.e., cells from RNA_POOL_1 were predicted to be from other samples not in RNA_Pool_1). 84,225 cells remained for downstream analyses (Table S3).

Annotation of Cell Type Clusters

We annotated the scRNA-seq clusters by first clustering at three different resolutions (0.5, 0.08, and 0.1) (Figure S4-6). We selected resolution = 0.08 because it best captured the expected iPSC-PPC cell types based on each cluster's expression for the following gene markers: *POU5F1* (iPSC), *COL1A1*, *COL1A2* (mesendoderm) *AFP, APOA* (early definitive endoderm), *GATA4*, *GATA6, PDX1* (early PPC), *PDX1*, *NKX6-1* (late PPC), *PAX6, CHGA*, *INS*, *GCG*, *SST* (endocrine), and *FLT1* (early ductal). We validated our annotations by comparing the iPSC-PPC clusters to those identified from scRNA-seq of ESC-PPC samples over 4 different stages of differentiation [100] (GSE114412): Stage 3 (Day 6; 7,982 cells), Stage 4 (Day 13; 6,960 cells), Stage 5 (Day 18; 4,193 cells), and Stage 6 (Day 25; 5,186 cells). Specifically, we compared the expression patterns of the gene markers between the clusters using z-normalized mean expression computed on cells expressing at least 1% of maximal expression for the gene, as described in the reference study [100]. Metadata containing single cell annotations are reported in Table S3.

Differentially Expressed Genes

To identify differentially expressed genes for each iPSC-PPC cluster, we used the *FindAllMarkers* function in Seurat [101] with *logfc.threshold = 0.01* and *min.pct = 0.01*. P-values were automatically adjusted using Bonferroni correction, and genes with adjusted p-values ≤ 0.05 were considered differentially expressed (Table S4).

eQTL Analysis

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

Covariates for eQTL Mapping

We included the following as covariates for eQTL mapping of both gene expression and isoform usage: 1) sex; 2) normalized number of RNA-seq reads; 3) percent of reads that mapped to autosome or sex chromosomes (labeled as "pct_uniquely_mapped_to_canonical_chromosomes" in Table S2); 4) percent of reads mapped to mitochondrial chromosome; 5) 20 genotype principal components to account for global ancestry; 6) 20 PEER factors to account for

683 transcriptome variability; and 7) kinship matrix to account for genetic relatedness between samples. All covariates are
684 available in Table S1-3.

685 Genotype Principal Component Analysis (PCA): Global ancestry was estimated using the genotypes of the 439,461
686 common variants with minor allele frequency (MAF) between 45 and 55% in the 1000 Genomes Phase 3 Panel [94]. We
687 merged the VCF files for the 106 iPSCORE subjects and the 2,504 subjects in the 1000 Genomes [94] and performed a
688 PCA analysis using *plink --pca* [93] (Figure S1A). The top 20 principal components were used as covariates in the eQTL
689 model to account for global ancestry and can be found in Table S1.

690 PEER Factors: We sought to determine the optimal number of PEER factors to use in the eQTL analysis that will result
691 in maximal eGene discovery. To this end, we initially calculated PEER factors on the 10,000 expressed genes with the
692 largest variance across all samples. To limit biases due to the expression levels of each gene, we divided the 16,464
693 expressed genes into ten deciles based on their average TPM, and selected 50 genes from each decile, for a total of 500
694 genes. We next performed eQTL analysis on each of the 500 genes using 10 to 60 PEER factors in increments of 10.
695 While 30 PEER factors resulted in the highest percentage of eGenes (14.0%), we opted for using 20 PEER factors
696 because the eQTL analysis had a comparable percentage of eGenes (11.8%) to GTEx tissues with similar sample sizes
697 [10] (Figure S18). Although we observed variable fraction of double-positive PDX1+/NKX6-1+ cells in the iPSC-PPC
698 samples, we did not include this variable as a covariate because PEER factors 1 and 4 already accounted for this
699 variability (Figure S19).

700 Kinship Matrix: The kinship matrix was included as a random effects term to account for the genetic relatedness
701 between individuals in our cohort. We constructed the kinship matrix using the same 439,461 variants employed above
702 using the *–make-rel square* function in plink [93]. The kinship matrix is available in Table S2.

703 *eQTL Analysis*

704 We performed eQTL analysis using the same method described in our previous study [12]. For each expressed autosomal
705 gene and isoform, we tested variants that were within 500 Kb of the gene body coordinates using the *bcftools query*
706 function. To account for the genetic relatedness between the samples, we performed eQTL mapping using a linear
707 mixed model with the *scan* function in limix v.3.0.4 [102] that incorporates the kinship matrix as a random effects term.
708 Specifically, eQTL mapping was implemented through the following model:

709 $$y_i = \beta_{ji} \cdot g_j + \sum_{n=1}^{N} \beta_n \cdot C_n + \mathrm{u} + \epsilon_{ij}$$

710 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
711 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
712 as a random effect, and ϵ is the error term for the association between expression of gene i and genotype of SNP j. As

described above, we used the following as covariates: 1) sex, 2) normalized number of RNA-seq reads, 3) percent of reads mapped to autosomal or sex chromosome, 4) percent of reads mapped to mitochondrial chromosome, 5) the top 20 genotype PCs (to account to global ancestry), and 6) the top 20 PEER factors (to account for confounders of expression variability), and are available in Tables S1-2.

FDR Correction

To perform FDR correction, we used a two-step procedure described in Huang et al. [103], which first corrects at the gene level and then at the genome-wide level. First, we performed FDR correction on the p-values of all variants tested for each gene or isoform using eigenMT [102], which considers the LD structure of the variants. Then, we extracted the lead eQTL for each gene or isoform based on the most significant FDR-corrected p-value. If more than one variant had the same FDR-corrected p-value, we selected the one with the largest absolute effect size as the lead eQTL. For the second correction, we performed an FDR-correction on all lead variants using the Benjamini-Hochberg method (q-value) and considered only eQTLs with q-value ≤ 0.01 as significant (Table S6).

Conditional eQTLs

To identify additional independent eQTLs (i.e., conditional eQTLs) for each eGene and eIsoform, we performed a step-wise regression analysis in which the genotype of the lead eQTL was included as a covariate in the model and the eQTL mapping procedure (regression and multiple test correction) was re-performed. We repeated this analysis to discover up to five additional associations for each eGene and eIsoform. Conditional eQTLs with q-values ≤ 0.01 were considered significant (Table S6).

Functional characterization of iPSC-PPC eQTLs

Fine-mapping of eQTL Associations

To define a credible set of candidate causal variants for each eQTL association, we performed genetic fine-mapping using the *finemap.abf* function in *coloc* (version 5.1.0, R) [35]. This Bayesian method converts p-values of all variants tested for a specific gene to posterior probabilities (PP) of association for being the causal variant. Variants with PP ≥ 1% are available in Table S7. The eQTLs not present in this table do not having any variants with PP ≥ 1% (i.e., all variants were estimated to have PP < 1%).

Genomic enrichments of e_gQTLs and e_iQTLs

For each independent eQTL association, we obtained candidate causal variants whose PP ≥ 5% (Table S7) and determined their overlap with each of the following genomic annotations using *bedtools intersect*: short splice acceptor sites (± 50bp), long splice acceptor sites (± 100bp), splice donor sites (± 50bp), UTR, intron, exon, intergenic, promoters, and RNA-binding protein binding sites (RBP-BS). RBP-BS were downloaded from a published dataset that

utilized enhanced CLIP to identify binding sites of 73 RBPs [104]. We considered only binding sites with irreproducible discovery rate (IDR) threshold of 0.01, indicating that these sites were reproducible across multiple biological samples. Enrichment of candidate causal variants for genomic regions was calculated using a Fisher's Exact Test comparing the proportion of SNPs that overlap each annotation between e_gQTLs and e_iQTLs. P-values were corrected using the Benjamini-Hochberg method and were considered significant if their FDR-corrected p-value ≤ 0.05 (Figure 1E).

Quantification of allele-specific binding of transcription factors using GVATdb

To annotate each candidate causal variant by their effects on transcription factor (TF) binding, we used the Genetic Variants Allelic TF Binding Database (GVATdb) to estimate the TF binding impact score associated with each variant and each of the 58 iPSC-PPC-expressed TF available on the database and with a AUPRC > 0.75 indicating a high-confidence deltaSVM model. We estimated the score using the instructions and reference files provided on the GVATdb GitHub repository (https://github.com/ren-lab/deltaSVM). The software required a list of SNPs as input along with hg19 reference files provided in the GVATdb repository. The output provides the deltaSVM score [105] for each variant-TF pair (Table S8), indicating whether the variant results in a promotion ("Gain"), disruption ("Loss"), or no change ("None") in TF binding.

Correlation between eQTL effect size and binding affinity of transcription factors

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

Colocalization between iPSC-PPC gene and isoform eQTLs

To determine the overlap of genetic variants between e_gQTLs and e_iQTLs for the same gene, we performed Bayesian colocalization using the *coloc.abf* function in *coloc* (version 5.1.0, R) [35], where each pair of signals was given a summary PP that each of the following five hypotheses was true: H0) no association was detected in both signals, H1) an association was detected in signal 1, H2) an association was detected in signal 2, H3) an association was detected in both signals but the underlying causal variants are different, and H4) an association was detected for both signals and the underlying causal variants are the same. We considered two eQTL signals to be shared if the number of overlapping variants used to test for colocalization (called "nsnps" in *coloc.abf* output) ≥ 500 and the PP for H4 (called "PP.H4.abf" in *coloc.abf* output; hereafter referred to as PP.H4) ≥ 80%. Conversely, two signals were considered distinct if nsnps ≥ 500 and PP for H3 (called "PP.H3.abf" in *coloc.abf* output; hereafter referred to as PP.H3) ≥ 80%. eQTL associations with PP.H4 < 80% and PP.H3 < 80% were due to insufficient power in one or both eQTL signals. As input into *coloc.abf*, we provided p-values, minor allele frequency, and sample size. All associations with PP ≥ 80% for any model are available in Table S9.

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

To test the enrichment of overlapping e_gQTLs and e_iQTLs in genomic regions compared to non-overlapping signals, we used a similar approach described in a previous study [10]. We determined the overlap of candidate causal variants with PP ≥ 1% in each genomic annotation using *bedtools intersect* and compared the proportion of variants overlapping each annotation against a background set of 20,000 random variants using a Fisher's Exact Test. For overlapping eQTLs, we used the candidate causal variants predicted in the *coloc.abf* output. Enrichments with nominal p-value < 0.05 were considered significant (Figure S10).

Downloading eQTL summary statistics for adult pancreatic tissues

We downloaded complete eQTL summary statistics for gene and exon associations for 420 adult pancreatic endocrine from the InSPIRE Consortium (https://zenodo.org/record/3408356) [11], and gene and splicing associations for 305 adult pancreatic exocrine from the GTEx Data Portal for GTEx Analysis version 8 [10] (https://console.cloud.google.com/storage/browser/gtex-resources). All GTEx SNPs were converted to hg19 using the UCSC liftOver Bioconductor package in R (https://www.bioconductor.org/help/workflows/liftOver/). Complete statistics for conditional associations in the adult endocrine and exocrine datasets were not readily available and therefore, not included in our analyses.

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

Comparing eGenes between iPSC-PPC and adult endocrine

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

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

Colocalization between iPSC-PPC and adult eQTLs

To identify eQTLs whose effects were driven by the same causal signals in iPSC-PPC and adult pancreatic tissues (endocrine and exocrine), we performed Bayesian colocalization using the *coloc.abf* function in *coloc* (version 5.1.0, R) [35]. Specifically, for each iPSC-PPC and adult eQTL, we tested its overlap with nearby eQTLs within 3 Mb from the gene body coordinates. eQTLs with no overlapping variants would automatically not be tested. Then, we filtered the results by requiring that each colocalization used the number of overlapping variants (called "nsnps" in the *coloc.abf* output) ≥ 500. As described above, we considered two eQTL signals to be shared if PP.H4 ≥ 80% or distinct if PP.H3 ≥ 80%. eQTL associations with PP.H4 < 80% and PP.H3 < 80% were due to insufficient power in one or both eQTL signals.

Because we, and others, have shown that e_gQTLs are functionally different from e_{AS}QTLs (e_iQTLs, exon eQTLs, and splicing eQTLs), we performed colocalization for e_gQTLs and e_{AS}QTLs independently (i.e., colocalization of e_gQTL was performed only with another e_gQTL and an e_{AS}QTL only with another e_{AS}QTL). All associations with PP ≥ 80% for any model are reported in Table S9.

Fine-mapping of adult eQTL associations

Similarly for iPSC-PPC eQTLs, we identified candidate causal variants using the *finemap.abf* function in *coloc* (version 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 causal variant. Variants with PP ≥ 1% are available in Table S7. The eQTLs not present in this table do not having any variants with PP ≥ 1% (i.e., all variants were estimated to have PP < 1%).

For all downstream analyses beyond this point, we used only iPSC-PPC, adult pancreatic endocrine, and adult pancreatic exocrine eQTLs with at least one candidate causal variant with PP ≥ 1%, outside of the MHC region, and are annotated in GENCODE version 34 hg19, to ensure that our analyses were powered sufficiently and the multiple datasets were comparable.

Identifying tissue-unique singleton eQTLs

Singleton eQTLs were defined in this study as an eQTL not colocalizing or in LD (r^2 ≥ 0.2 and within 500 Kb) with another eQTL in the same or different pancreatic tissue. Singleton eQTLs were also considered tissue-unique as they were functional in only the tested tissue. For each eQTL that did not display a H4 association with another eQTL, we examined their LD with nearby eQTLs of the same phenotype (gene expression or alternative splicing) in all three pancreatic tissues using their most likely candidate causal variants based on the highest PP (from *finemap.abf* output). LD was calculated using *plink --r2 square --keep-allele-order --make-bed* [93] and the 1000 Genomes Phase 3 panel [94]. A singleton eQTL was considered in LD with another eQTL if the singleton's candidate causal variant was within 500 Kb and in LD (r^2 ≥ 0.2) with another eQTL's candidate causal variant. If the candidate causal variant was not genotyped

in the 1000 Genomes Phase 3 panel, then we used the next top candidate causal variant and repeat the process, if needed, until no more variant was remaining with causal PP ≥ 1%. If none of the candidate causal variants with PP ≥ 1% were genotyped in 1000 Genomes, then we used distance as a metric for determining potential associations, where if the singleton candidate causal variant was within 500 Kb with another eQTL's candidate causal variant, we considered them to be potentially associated. A singleton eQTL in LD or potentially associated based on distance was annotated as "ambiguous" and excluded from further analysis, otherwise we annotated the eQTL as a tissue-unique singleton. All annotated tissue-unique singleton eQTLs are reported in Table S10.

Identifying eQTL modules

eQTL modules were identified by first creating a network using the *graph_from_data_frame* function in igraph (version 1.3.4, R) [106] where the input was a data frame containing all pairs of colocalized eQTLs (nsnps ≥ 500 and PP.H4 ≥ 80%) as binary edges. We created networks for each chromosome and phenotype (gene expression and alternatively splicing) independently, totaling to 44 networks (22 chromosomes x 2 phenotypes = 44 networks). Then, we performed community detection analysis using the *cluster_leiden* function with *--objective_function = "modularity"*, *n_iterations = 500, resolution = 0.3* to identify modules of eQTLs. Upon examining them in depth, we observed that 5% of the modules contained at least one H3 association (PP.H3 ≥ 80%) between a pair of eQTLs, indicating that signals within a module were predicted to have distinct genetic variants despite being assigned to the same module. Therefore, to filter for modules that contained eQTLs likely to share the same genetic variants, we required that at least 30% of all eQTL pairs had a H4 association and that the number of H4 "edges" was twice the number of H3 "edges" (number of H4 edges / number of H3 edges ≥ 2). For example, a module with four eQTLs would have six possible pairwise combinations, and to be considered a validated module, we required at least two H4 edges and no more than one H3 edge. Modules that did not pass these thresholds were annotated as "module_failed" and excluded from downstream analyses. Summary of eQTL modules and their individual eQTL associations are reported in Table S11. Module IDs were assigned such that the first term indicates the phenotype the module was associated with ("GE" for gene expression or "AS" for alternative splicing), the second term indicates the chromosome number, and the third term indicates a unique integer. For example, "GE_1_32" indicates that this module is associated with changes in gene expression, located in in chromosome 1, and assigned the number 32.

Identifying tissue-unique and tissue-sharing eQTL modules

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

1) Fetal-unique: contains eQTLs in **only** iPSC-PPC

2) Adult endocrine-unique: contains eQTLs in **only** adult endocrine
3) Adult exocrine-unique: contains eQTLs in **only** adult exocrine
4) Adult-shared: contains eQTLs in adult endocrine **and** adult exocrine
5) Fetal-endocrine: contains eQTLs in iPSC-PPC **and** adult endocrine
6) Fetal-exocrine: contains eQTLs in iPSC-PPC **and** adult exocrine
7) Fetal-adult: contains eQTLs in **all** three pancreatic tissues

We next filtered the eQTL modules based on their LD relationships with other tissues to confirm the module's tissue specificity. For example, we required fetal-endocrine modules to contain eQTLs specific to only iPSC-PPC and adult endocrine and not be in LD with an eQTL from adult exocrine. Similar to the analysis described above for identifying tissue-unique singletons, we calculated LD between each pair of eQTLs' most likely candidate causal variants (based on the highest PP; PP $\geq$ 1%) using *plink --r2 square --keep-allele-order --make-bed* [93] and the 1000 Genomes Phase 3 panel [94]. For each of the module categories, we required that the following were true to be considered for downstream analyses:

1) Fetal-unique: contains eQTLs in only iPSC-PPC, and **all** eQTLs were not in LD with eQTLs in adult endocrine **and** adult exocrine
2) Adult endocrine-unique: contains eQTLs in only adult endocrine, and **all** eQTLs were not in LD with eQTLs in adult exocrine **and** iPSC-PPC
3) Adult exocrine-unique: contains eQTLs in only adult exocrine, and **all** eQTLs were not in LD with eQTLs in adult endocrine **and** iPSC-PPC
4) Adult-shared: contains eQTLs in only adult endocrine and adult exocrine, and **all** eQTLs were not in LD with eQTLs in iPSC-PPC
5) Fetal-endocrine: contains eQTLs in iPSC-PPC and adult endocrine, and **all** eQTLs were not in LD with eQTLs in adult exocrine
6) Fetal-exocrine: contains eQTLs in iPSC-PPC and adult exocrine, and **all** eQTLs were not in LD with eQTLs in adult endocrine
7) Fetal-adult: contains eQTLs in **all** three pancreatic tissues.

For any module that did not meet the above requirements, we annotated the eQTLs in the module "ambiguous" and excluded for downstream analysis. Hereafter, we refer the eQTL associations in tissue-unique modules (categories 1-3) as tissue-unique combinatorial eQTLs and those in categories 5-7 as eQTLs shared between both fetal-like and adult stages. All annotations for eQTL modules and their individual eQTLs are reported in Table S10 and Table S11.

Enrichment of tissue-unique eQTLs in pancreatic chromatin states

We obtained chromatin state maps for adult endocrine and human embryonic stem cell-derived pancreatic progenitor cells from previously published studies [21,107] and adult pancreatic exocrine from the Roadmap Epigenome Project (epigenome ID: E098) [7]. Because e_gQTLs were likely to affect non-coding regulatory elements (Figure 1E), we examined their enrichments in chromatin states to better understand, and validate, their functional mechanisms. Enrichments were calculated using a Fisher's Exact Test by comparing the proportion of candidate causal variants (from *finemap.abf*, see above sections; PP $\geq$ 10%) of tissue-unique singleton and combinatorial e_gQTLs in each chromatin state to a background set of 20,000 randomly selected variants. Enrichments with Benjamini-Hochberg-corrected p-values $\leq$ 0.05 were considered significant. Enrichment results are available in Table S12, Figure 3I, and Figure S11C.Functional plasticity of eQTLs in fetal-like and adult pancreatic tissues

For the modules shared between both fetal-like and adult pancreatic tissue (categories 5-7; described above), we compared the eGenes associated with 1) iPSC-PPC eQTLs versus adult endocrine eQTLs and 2) iPSC-PPC eQTLs versus adult exocrine eQTLs. For e_{AS}QTLs, we compared the genes mapping to 1) each isoform in iPSC-PPC versus exon in adult endocrine and 2) each isoform in iPSC-PPC versus splice interval in adult exocrine. From these comparisons, we assign each module an "endocrine_egene_overlap" label and an "exocrine_egene_ overlap" label in Table S11 (also shown in Figure 4A and Figure S12D), where "zero" indicates that the module does not contain an eQTL in the adult tissue, "same" indicates that the module contains eQTLs associated with the same gene in iPSC-PPC and adult, "partial" indicates that the module contains eQTLs associated with partially overlapping genes between iPSC-PPC and adult, and "different" indicates that the module contains eQTLs associated with entirely different genes. For example, if a module was annotated with "zero" for endocrine_egene_overlap and "same" for exocrine_egene_overlap, this indicates that the module was shared between only fetal-like and adult exocrine (i.e, "fetal-exocrine" or category 6 as described above; does not contain an adult endocrine eQTL) and the genes associated with this locus were the same in both tissues.

Complex Trait GWAS Associations

Colocalization of eQTLs with GWAS associations

We obtained GWAS summary statistics from ten different studies: 1) type 1 diabetes [3], 2) type 2 diabetes [107], 3) body mass index [46], 4) triglycerides [46], 5) HDL cholesterol [46], 6) LDL direct [46], 7) cholesterol [46], 8) glycated hemoglobin A1C (HbA1c) levels from the MAGIC Consortium [108], 9) HbA1c levels from the Pan-UKBB Study [46], and 10) fasting glucose [108]. All of the data, except for type 1 diabetes, were provided in hg19 coordinates, therefore we converted the coordinates from hg38 to hg19 using the liftOver package in R [109]. We sorted and indexed each file using *tabix* [89]. For each trait, we performed colocalization between GWAS variants and all filtered significant eQTLs (see bolded section above) in the three pancreatic tissues with the *coloc.abf* function in *coloc* (version 5.1.0, R) [35] using p-values, MAF, and sample size as inputs. Then, we filtered results based on whether the lead candidate causal variant underlying both GWAS and eQTL association (from *coloc.abf* output) is genome-wide significant for GWAS association (p-value $\leq$

$5x10^{-8}$) and the number of overlapping variants used to test for colocalization (nsnps) ≥ 500. eQTLs were considered to share a genetic signal with GWAS if PP.H4 ≥ 80% or have distinct signals with GWAS if PP.H3 ≥ 80%. For eQTL modules, we required that at least 30% of the eQTLs in the module colocalized with GWAS (PP.H4 ≥ 80%) and that the number of H4 associations is twice the number of H3 associations (number of H4 associations / number of H3 associations ≥ 2). Colocalization results for the 397 GWAS loci with PP.H4 ≥ 80% are available in Table S13.

GWAS 99% Credible Sets

For each GWAS locus (based on GWAS locus ID in Table S13), we constructed 99% credible sets with the predicted candidate causal variants underlying both eQTL and GWAS associations (from *coloc.abf* output). If the GWAS locus colocalized with a singleton eQTL, the credible sets were constructed using the output of the eQTL's colocalization with GWAS. If the GWAS locus colocalized with an eQTL module, we constructed credible sets for each of the pairwise eQTL-GWAS colocalization and retained the eQTL that resulted in the least number of candidate causal variants. If multiple eQTLs had the same number of variants in their credible set, we considered the eQTL with the highest PP.H4 for GWAS colocalization. 99% credible sets were constructed by first sorting the variants by descending order of causal PP and obtaining the least number of variants that resulted in a cumulative PP ≥ 99%. 99% credible sets for each of the 397 GWAS loci (248 singleton and 149 module) are reported in Table S14.

Data Availability

FASTQ sequencing data for iPSC-PPC scRNA-seq and bulk RNA-seq have been deposited into GSE152610 and GSE182758, respectively. RNA-seq for iPSC, adult endocrine, and adult exocrine samples used in PCA and pseudotime analyses were downloaded from phs000924, GSE50398, and phs000424, respectively. eQTL summary statistics for adult endocrine and exocrine samples were obtained from the GTEx Data Repository (https://console.cloud.google.com/storage/browser/gtex-resources) and a previously published study [11] (https://zenodo.org/record/3408356), respectively. WGS data for iPSCORE subjects were downloaded as a VCF file from phs001325. GWAS summary statistics were obtained from the Pan UK BioBank resource (https://pan.ukbb.broadinstitute.org/), the MAGIC (Meta-Analyses of Glucose and Insulin-related traits) Consortium (https://magicinvestigators.org/downloads/; https://doi.org/10.1038/s41588-021-00852-9), the DIAMANTE Consortium (https://diagram-consortium.org/downloads.html; http://doi.org/10.1038/s41588-018-0241-6), and a previously published study [3]. Full eQTL summary statistics for iPSC-PPC, supplemental tables, and processed scRNA-seq data have been deposited in Figshare: https://figshare.com/projects/Large-scale_eQTL_analysis_of_iPSC-PPC/156987.

Author information

iPSCORE Consortium, University of California, San Diego, La Jolla, CA, 92093, US

Angelo D. Arias, Timothy D. Arthur, Paola Benaglio, Victor Borja, Megan Cook, Matteo D'Antonio, Agnieszka D'Antonio-Chronowska, Christopher DeBoever, Margaret K.R. Donovan, KathyJean Farnam, Kelly A. Frazer, Kyohei Fujita, Melvin Garcia, Olivier Harismendy, David Jakubosky, Kristen Jepsen, Isaac Joshua, He Li, Hiroko Matsui, Naoki Nariai, Jennifer P. Nguyen, Daniel T. O'Connor, Jonathan Okubo, Fengwen Rao, Joaquin Reyna, Lana Ribeiro Aguiar, Bianca Salgado, Nayara Silva, Erin N. Smith, Josh Sohmer, Shawn Yost, William W. Young Greenwald

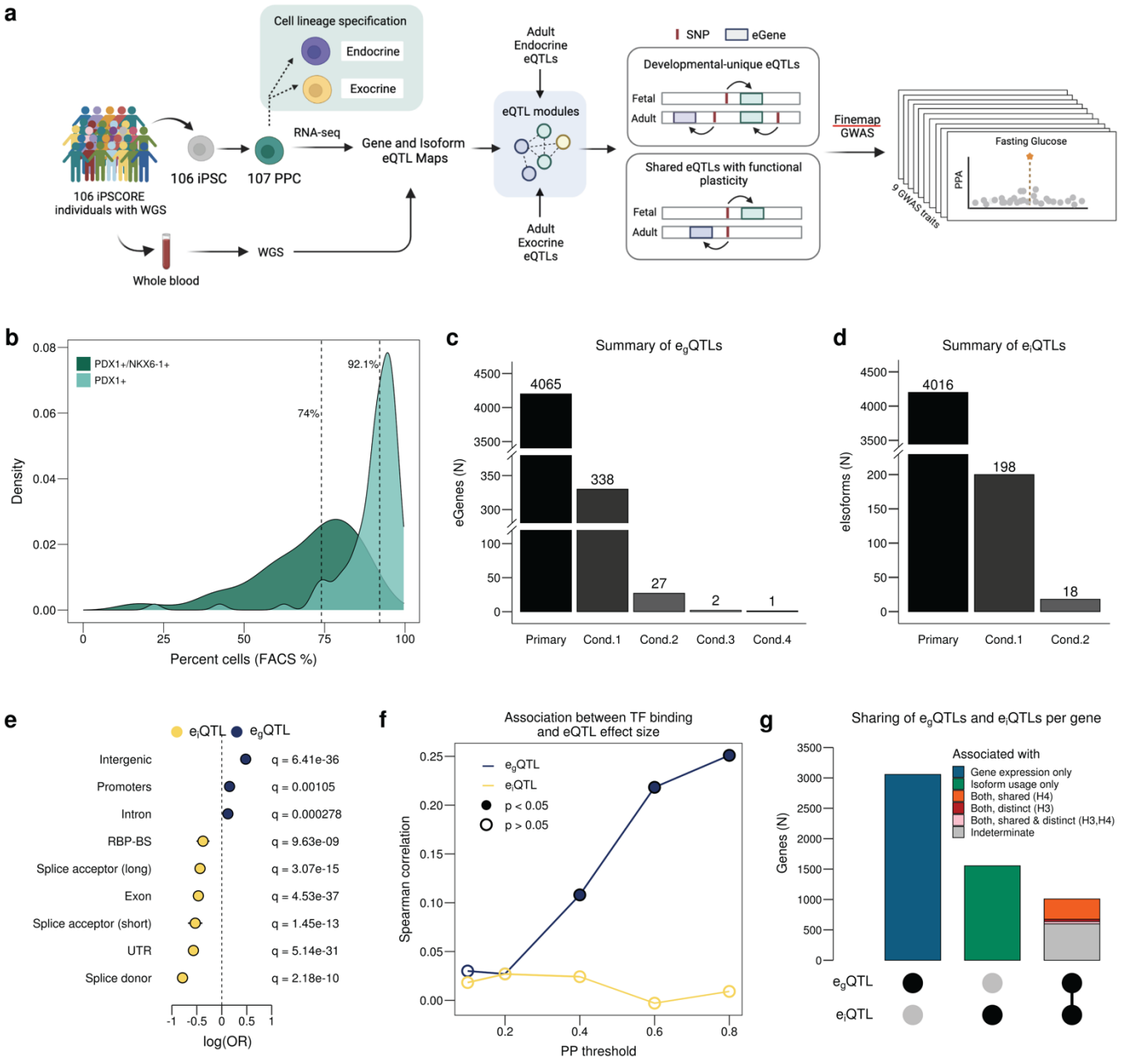
Contributions

KAF conceived the study. ADC, BS, and KF performed the differentiations and generated molecular data. JPN, MKRD and HM performed quality check on scRNA-seq and RNA-seq samples. JPN and TDA performed the computational analyses. KAF, ADC, MD oversaw the study. JPN, MD and KAF prepared the manuscript.

Acknowledgements

This work was supported by the National Library Training Grant T15LM011271 and the National Institute of Diabetes and Digestive and Kidney Disease (NIDDK) F31DK131867, U01DK105541, DP3DK112155 and P30DK063491. Additional support was also received from the National Heart, Lung and Blood Institute (NHLBI) F31HL158198. This publication includes data generated at the UC San Diego IGM Genomics Center utilizing an Illumina NovaSeq 6000 that was purchased with funding from a National Institutes of Health SIG grant S10OD026929.

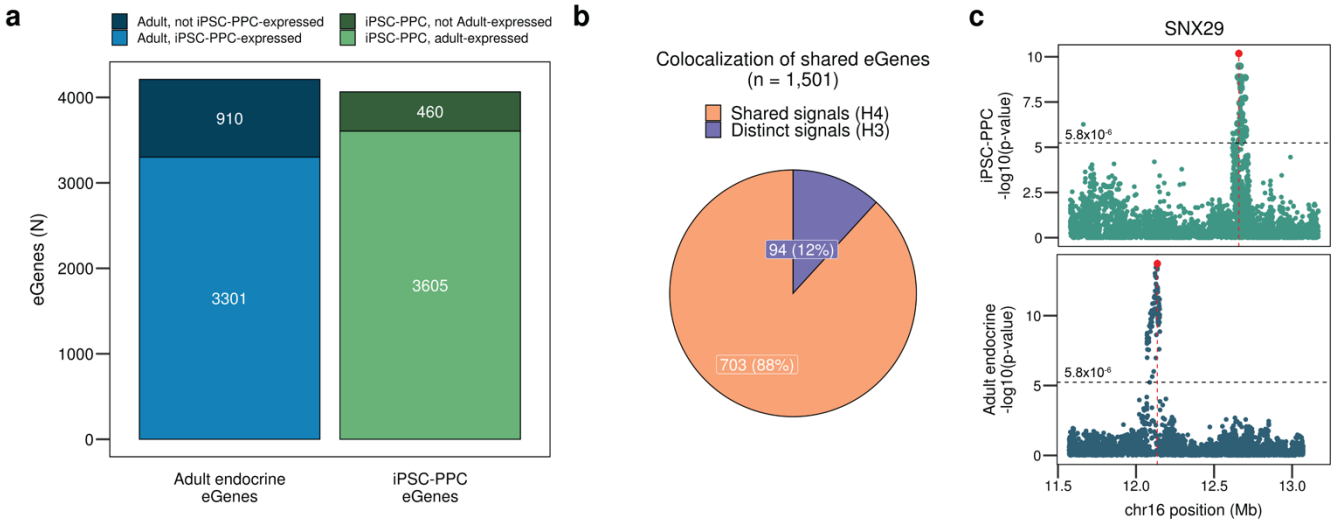
Figure 1. Discovery and Characterization of eQTLs in iPSC-PPC



(**a**) Study overview. (**b**) Density plots showing the distribution of PDX1$^+$ cells (%; regardless of NKX6-1 status; light green) and PDX1$^+$/NKX6-1$^+$ cells (%; dark green). (**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) of eQTLs for functional genomic annotations using a two-sided Fisher's Exact Test comparing the proportion of SNPs with causal PP ≥ 5% between e_gQTLs (blue; n = 8,763) and e_iQTLs (yellow; n = 8,919). (**f**) Line plot comparing the spearman correlation between TF binding score and eQTL effect size at different thresholds of PP for e_gQTLs (blue) and e_iQTLs (yellow). Closed points indicate significance of correlation based on nominal p <

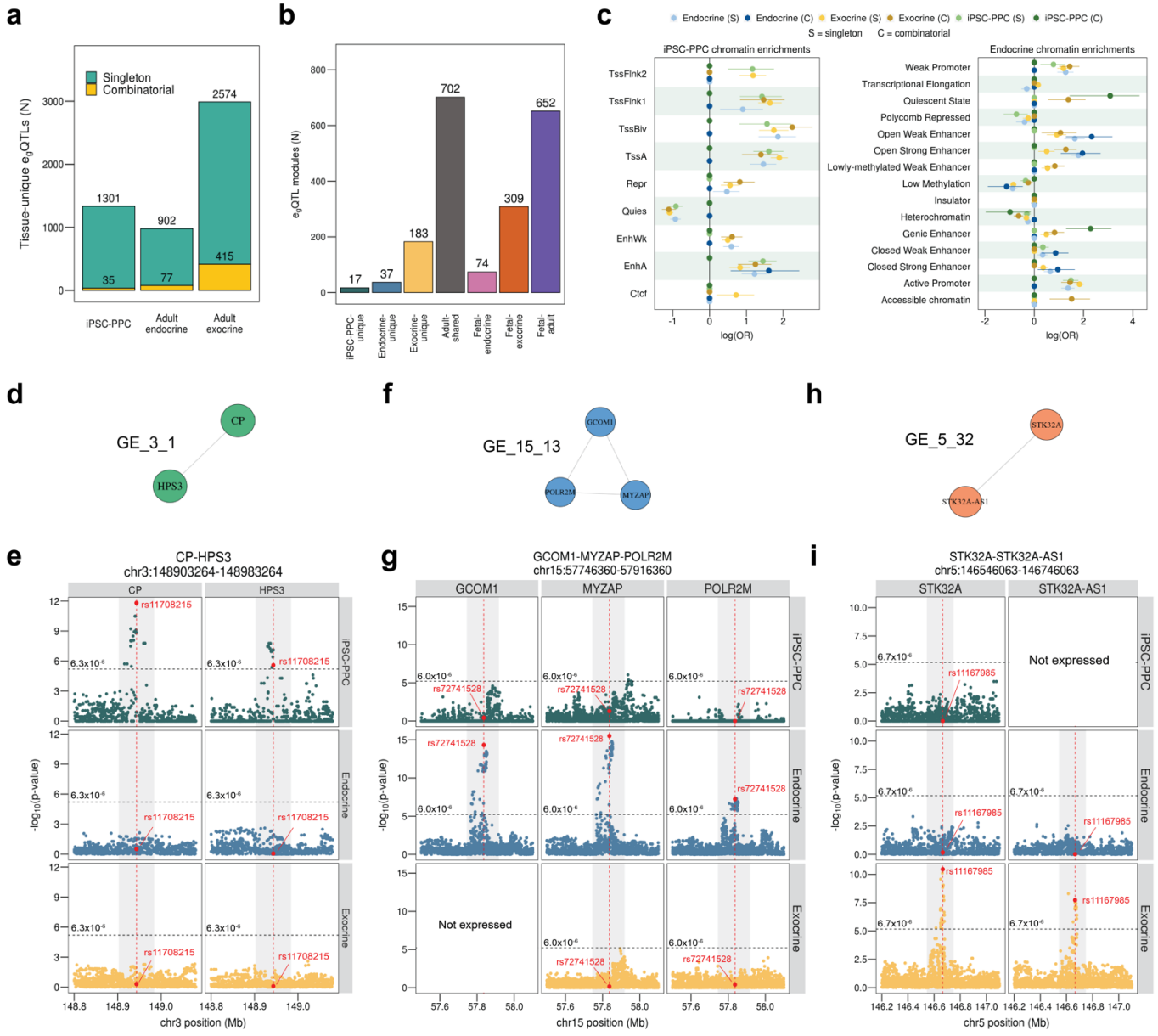
0.05. (**g**) Bar plot showing the number of genes that have only e_gQTLs (blue; n = 3,057), only e_iQTLs (green; n = 1,554), or both. Orange represents genes with only overlapping e_gQTLs and e_iQTLs (PP.H4 ≥ 80%; n = 333) based on colocalization. Red represents genes with only distinct e_gQTLs and e_iQTLs (PP.H3 ≥ 80%; n = 38), and pink represents genes with both shared and distinct e_gQTLs and e_iQTLs (i.e., an eGene with two eIsoforms may colocalize with one eIsoform but not the other) (n = 39). Gray represents genes whose eQTL signals were not sufficiently powered to test for colocalization (PP.H4 < 80% and PP.H3 < 80%; n = 598).

Figure 2. Comparison of the genetic architecture underlying gene expression between fetal-like and adult endocrine



(**a**) Stacked bar plot showing the number of eGenes detected in adult endocrine (blue; n = 4,211 total) that are expressed in iPSC-PPC. Likewise, we show the number of iPSC-PPC eGenes (green; n = 4,065 total) that are expressed in adult endocrine. Darker shades represent eGenes that are expressed in the other tissue while lighter shades represent those that were expressed. These results show that the majority of iPSC-PPC and adult endocrine eGenes were expressed in the other tissue. Therefore, the small overlap of eGenes between the two tissues were not due to expression differences but instead due to differences in the genetic regulatory landscape. (**b**) Pie chart showing that 12% of the shared eGenes between iPSC-PPC and adult endocrine were associated with distinct genetic loci (PP.H3 ≥ 80%), indicating that different regulatory mechanisms facilitate the expression of the same gene in iPSC-PPC and adult endocrine. (**c**) Example of a shared eGene (*SNX29*) whose expression was associated with different e_gQTL signals in iPSC-PPC (green, top panel) and adult endocrine (blue, bottom panel). For plotting purposes, we assigned a single p-value for gene-level significance based on Bonferroni-correction (0.05 divided by the number of variants tested for the gene; horizontal line). Red vertical lines indicate the positions of the lead variants in the adult endocrine and fetal-like iPSC-PPC based on p-value (chr16:12656135 and chr16:12136526, respectively).

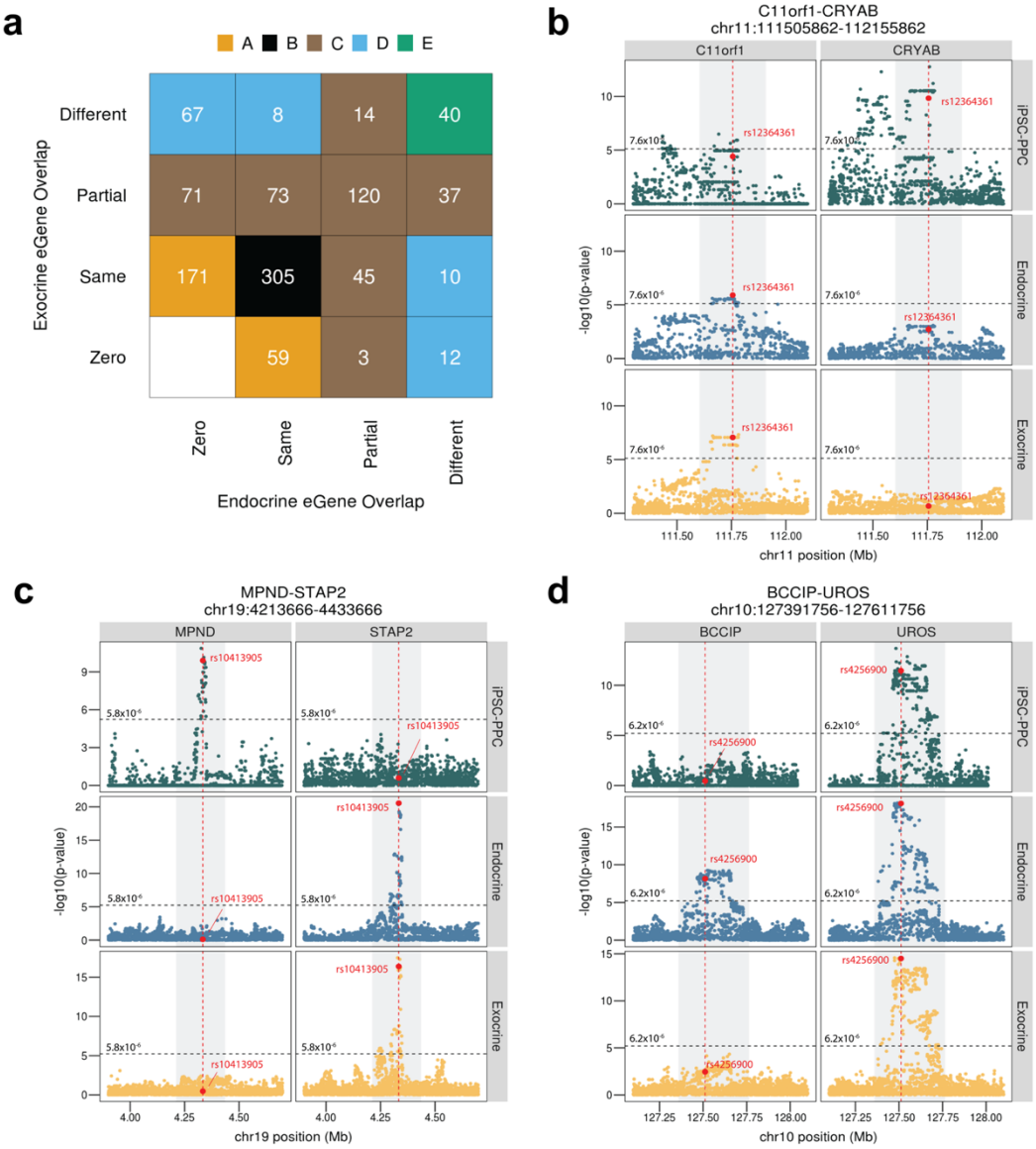
Figure 3. eQTL sharing between iPSC-PPC and adult pancreas



(**a**) Bar plot showing the number of tissue-unique e_gQTLs identified in fetal-like iPSC-PPC, adult pancreatic endocrine, and adult pancreatic exocrine. (**b**) Bar plot showing the number of e_gQTL modules for each annotation. (**c**) Plot showing the enrichment (odds ratio) of tissue-unique singleton (S) and combinatorial (C) e_gQTLs in PPC [21] (left) and endocrine [45] (right) chromatin states. Nomenclature for the chromatin states used in the previously studies was maintained. Enrichment was tested using a two-sided Fisher's Exact Test comparing the proportion of candidate causal variants with causal PP ≥ 20% overlapping the chromatin states between the e_gQTLs in question versus a background of randomly selected 20,000 variants. P-values were Benjamini-Hochberg-corrected and considered significant if the corrected p-values < 0.05. Non-significant results are set to log(odds ratio) = 0. Error bars represent 95% confidence

intervals for the odds ratios. (**d-e**) The chr3:148903264-148983264 locus (gray rectangle) was an example of an "iPSC-PPC-unique" e_gQTL locus (module ID: GE_3_1) associated with *CP* and *HPS3* expression. (**f-g**) The chr15:57746360-57916360 locus (gray rectangle) was an example of an "adult endocrine-unique" e_gQTL locus (module ID: GE_15_13) associated with *GCOM1*, *MYZAP*, and *POLR2M* expression. We show that the e_gQTL locus was unique to adult endocrine and not active in iPSC-PPC and adult exocrine. *GCOM1* was not expressed in adult exocrine and therefore, was not tested for e_gQTL discovery. (**h-i**) The chr5:146546063-146746063 locus (gray box) is an example of an "adult exocrine-unique" e_gQTL locus (module ID: GE_5_32) associated with *STK32A* and *STK32A-AS1* expression only in adult endocrine. *STK32A-AS1* was not expressed in iPSC-PPC and therefore, was not tested for e_gQTL discovery. Panel **d, f, h** display the e_gQTL modules as networks in which the e_gQTL associations (nodes) are connected by edges based on colocalization (PP.H4 $\geq$ 80%). For plotting purposes, we assigned a single p-value for gene-level significance based on Bonferroni-correction (0.05 divided by the number of variants tested for the gene; horizontal line). Red vertical lines indicate the positions of the lead candidate causal variants underlying the colocalization based on maximum PP.

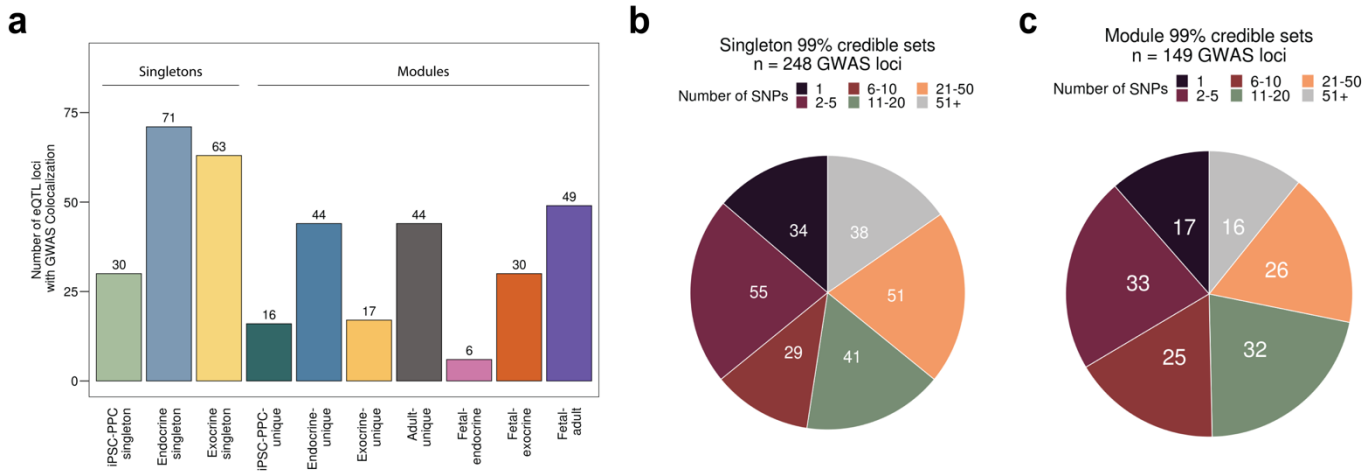
Figure 4. Regulatory plasticity of e_gQTL loci



(**a**) Number of e_gQTL modules based on eGene overlap between iPSC-PPC and the two adult pancreatic tissues. "Zero" indicates that the module does not contain an e_gQTL in the respective adult tissue. "Same" indicates that the module contains e_gQTLs for only the same eGenes in iPSC-PPC and the adult tissue. "Partial" indicates that the module contains e_gQTLs for partially overlapping eGenes between iPSC-PPC and the adult tissue. "Different" indicates that the module contains e_gQTLs for only different eGenes between iPSC-PPC and the adult tissue. For example, the 171 e_gQTL modules in category A (orange) contain e_gQTLs from only iPSC-PPC and adult exocrine (zero e_gQTLs from adult endocrine) and are associated with the same eGenes between the two tissues. (**b-d**) Examples of e_gQTL loci demonstrating regulatory plasticity of genetic variation across fetal-like and adult pancreatic stages. Panel **b** shows a locus associated with different eGenes in iPSC-PPC (*CRYAB*) and both the adult tissues (*C11orf1*). Panel **c** shows a

locus associated with *MPND* expression in only iPSC-PPC but *STAP2* expression in both the adult tissues. Panel **d** shows a locus associated with partially overlapping eGenes between the two pancreatic stages (*UROS* in all three pancreatic tissues and *BCCIP* in only adult endocrine). For plotting purposes, we assigned a single p-value for gene-level significance based on Bonferroni-correction (0.05 divided by the number of variants tested for the gene; horizontal line). Red vertical lines indicate the positions of the lead candidate causal variants underlying the colocalization based on maximum PP.

1045 **Figure 5. Summary of Pancreatic GWAS Associations**

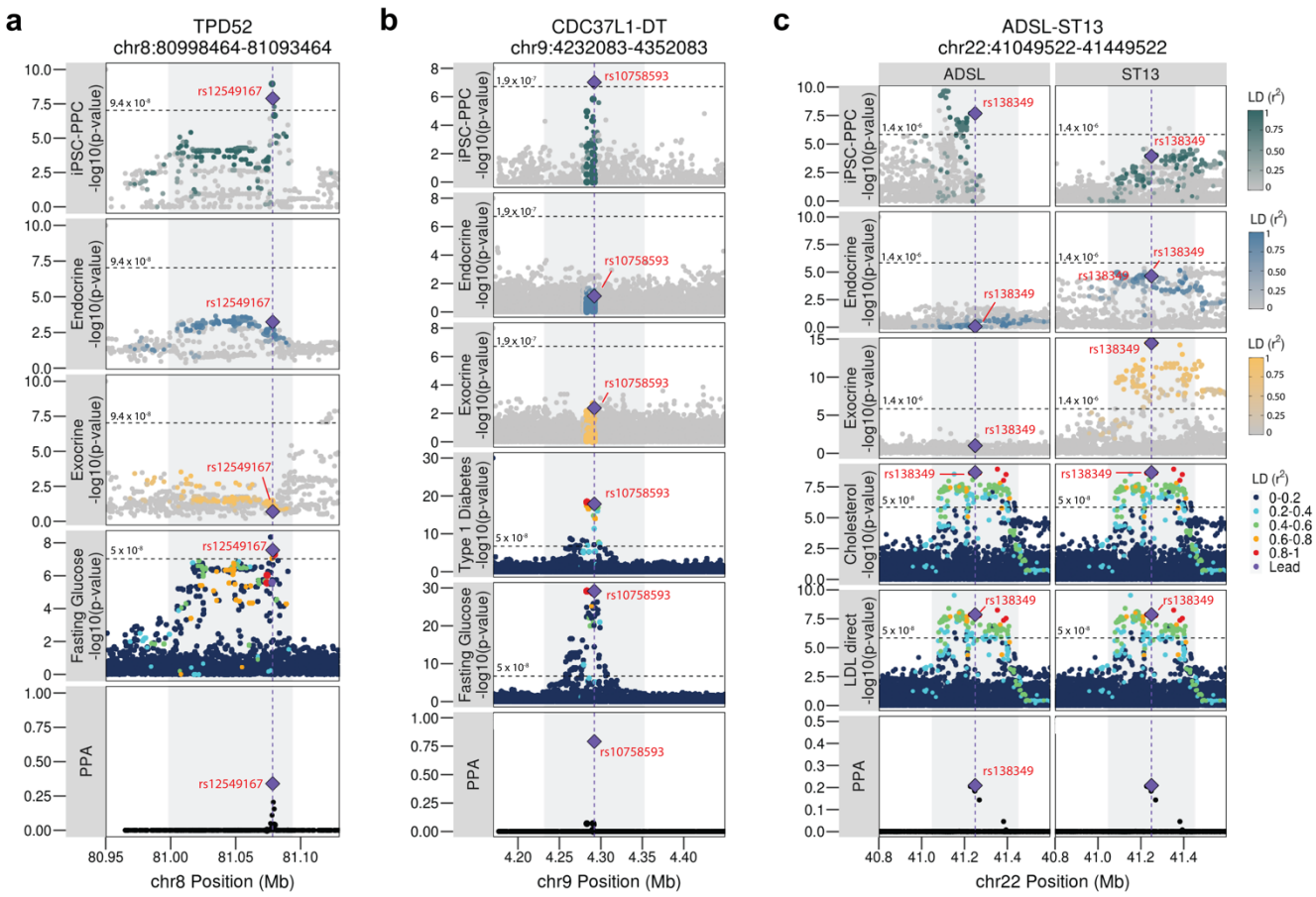


1046

1047 (**a**) Bar plot showing the number of eQTL loci that colocalized with GWAS variants (PP.H4 ≥ 80%) as a singleton or
1048 module. (**b**) Pie chart showing the number of singleton-colocalized GWAS loci (n = 248) color-coded by the number
1049 of candidate causal variants identified in their 99% credible sets. (**c**) Pie chart showing the number of module-
1050 colocalized GWAS loci (n = 149) color-coded by the number of candidate causal variants identified in their 99%
1051 credible sets.

1052 **Figure 6. Pancreatic GWAS Associations with Fetal-specific and Adult-shared Gene**

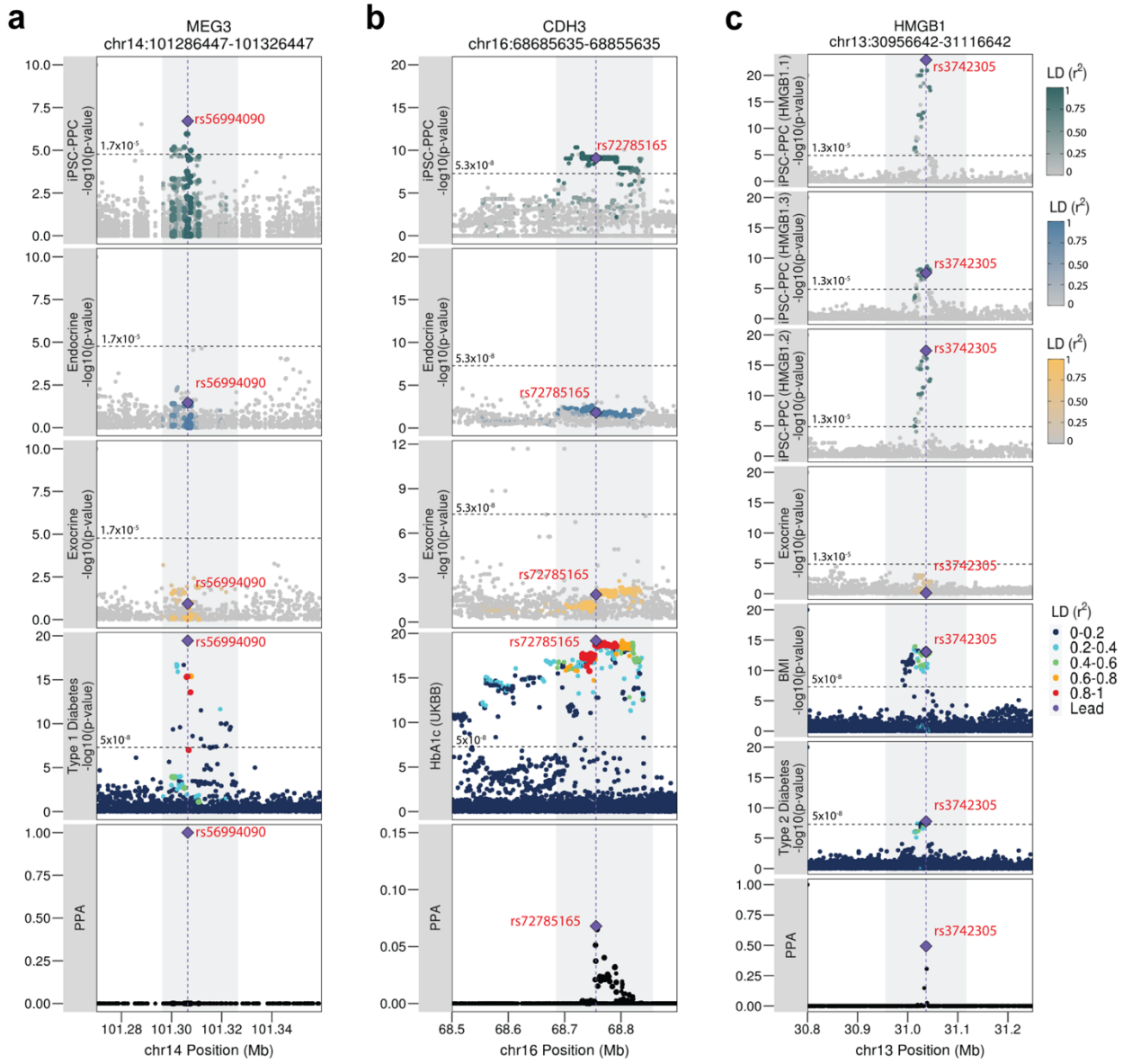
1053 **Expression**



1054

1055 (**a**) The TPD52 locus is associated with fasting glucose levels and colocalized with an iPSC-PPC-unique singleton
1056 e_gQTL with the predicted causal variant identified as rs12549167 (chr8:81078464:C>T, PP = 33.9%). (**b**) The
1057 CDC37L1-DT locus is associated with fasting glucose and type 1 diabetes and colocalized with an iPSC-PPC-unique
1058 singleton e_gQTL with the predicted causal variant identified as rs10758593 (chr9:4292083:G>A, PP = 79.2%). (**c**)
1059 Cholesterol and LDL direct GWAS loci colocalize with a fetal-adult e_gQTL module where the variants are associated
1060 with *ADSL* expression in iPSC-PPC and *ST13* expression in the adult tissues. The predicted causal variant was
1061 identified as rs138349 (chr22:41249522:A>G, PP = 21.9%). For plotting purposes, we assigned a single p-value for
1062 gene-level significance based on Bonferroni-correction (0.05 divided by the number of variants tested for the gene;
1063 horizontal line). Red vertical lines indicate the positions of the lead candidate causal variants underlying the
1064 colocalization based on maximum PP.

Figure 7. Pancreatic GWAS Associations with Fetal-specific Alternative Splicing



(**a**) T1D-risk locus colocalized with an iPSC-PPC-unique singleton e_{AS}QTL for *MEG3* with the predicted causal variant identified as rs56994090 (chr14:101306447:T>C, PP = 100%). (**b**) GWAS locus associated with HbA1c colocalized with an iPSC-PPC-unique singleton e_{AS}QTL for *CDH3* with the predicted causal variant identified as rs72785165 (chr16:68755635:T>A, PP = 6.8%). (**c**) *HMGB1* locus was associated with T2D-risk and BMI and colocalized with an iPSC-PPC-unique e_{AS}QTL module for differential usage of three *HMGB1* isoforms with the predicted causal variant identified as rs3742305 (chr13:31036642:C>G, PP = 49.3%). For plotting purposes, we assigned a single p-value for gene-level significance based on Bonferroni-correction (0.05 divided by the number of variants tested for the gene; horizontal line). Red vertical lines indicate the positions of the lead candidate causal variants underlying the colocalization based on maximum PP.

References

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

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

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

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

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

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

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

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

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

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

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

12. D'Antonio M, Arthur TD, Nguyen JP, Matsui H, D'Antonio-Chronowska A, Frazer KA. *Fine Mapping Spatiotemporal Mechanisms of Genetic Variants Underlying Cardiac Traits and Disease*. Genetics; 2022. doi:10.1101/2021.09.01.458619

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

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

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

16. Travers ME, Mackay DJG, Dekker Nitert M, et al. Insights Into the Molecular Mechanism for Type 2 Diabetes Susceptibility at the *KCNQ1* Locus From Temporal Changes in Imprinting Status in Human Islets. *Diabetes*. 2013;62(3):987-992. doi:10.2337/db12-0819

17. Zhao J, Bradfield JP, Zhang H, et al. Examination of All Type 2 Diabetes GWAS Loci Reveals *HHEX-IDE* as a Locus Influencing Pediatric BMI. *Diabetes*. 2010;59(3):751-755. doi:10.2337/db09-0972

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

42. Dixon JR, Jung I, Selvaraj S, et al. Chromatin architecture reorganization during stem cell differentiation. *Nature*. 2015;518(7539):331-336. doi:10.1038/nature14222

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

93. Shaun Purcell CC. PLINK 1.9.0.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

109. Bioconductor Package Maintainer. liftOver: Changing genomic coordinate systems with rtracklayer::liftOver. Published online 2022.

1326