Genome Biology

RESEARCH

Open Access

Genetic regulation of RNA splicing in human pancreatic islets

Goutham Atla[1,2,3], Silvia Bonàs-Guarch[1,2,3*], Mirabai Cuenca-Ardura[1,2], Anthony Beucher[1,2,3], Daniel J. M. Crouch[4], Javier Garcia-Hurtado[1,2], Ignasi Moran[3,5], the T2DSystems Consortium, Manuel Irimia[1], Rashmi B. Prasad[6,7], Anna L. Gloyn[8,9], Lorella Marselli[10], Mara Suleiman[10], Thierry Berney[11], Eelco J. P. de Koning[12,13], Julie Kerr-Conte[14], Francois Pattou[14], John A. Todd[4], Lorenzo Piemonti[15] and Jorge Ferrer[1,2,3*]

*Correspondence:
silvia.bonas@crg.eu; jorge.
ferrer@crg.eu

[1] Centre for Genomic Regulation, The Barcelona Institute of Science and Technology, Barcelona, Spain
Full list of author information is available at the end of the article

Abstract

Background: Non-coding genetic variants that influence gene transcription in pancreatic islets play a major role in the susceptibility to type 2 diabetes (T2D), and likely also contribute to type 1 diabetes (T1D) risk. For many loci, however, the mechanisms through which non-coding variants influence diabetes susceptibility are unknown.

Results: We examine splicing QTLs (sQTLs) in pancreatic islets from 399 human donors and observe that common genetic variation has a widespread influence on the splicing of genes with established roles in islet biology and diabetes. In parallel, we profile expression QTLs (eQTLs) and use transcriptome-wide association as well as genetic co-localization studies to assign islet sQTLs or eQTLs to T2D and T1D susceptibility signals, many of which lack candidate effector genes. This analysis reveals biologically plausible mechanisms, including the association of T2D with an sQTL that creates a nonsense isoform in *ERO1B*, a regulator of ER-stress and proinsulin biosynthesis. The expanded list of T2D risk effector genes reveals overrepresented pathways, including regulators of G-protein-mediated cAMP production. The analysis of sQTLs also reveals candidate effector genes for T1D susceptibility such as *DCLRE1B*, a senescence regulator, and lncRNA *MEG3*.

Conclusions: These data expose widespread effects of common genetic variants on RNA splicing in pancreatic islets. The results support a role for splicing variation in diabetes susceptibility, and offer a new set of genetic targets with potential therapeutic benefit.

Keywords: RNA splicing, Type 1 diabetes, Type 2 diabetes, TWAS, G-protein signaling, Pancreatic islets, Beta cells, Senescence, CTRB2, Pancreatic beta-cells, Diabetes pathophysiology, Quantitative trait loci

Atla *et al. Genome Biology*   (2022) 23:196

Page 2 of 28

Background

Genome-wide association studies have identified hundreds of genomic loci that carry genetic variants contributing to type 2 diabetes (T2D) and type 1 diabetes (T1D) susceptibility [1–5]. The vast majority of associated genetic variants are non-coding, and epigenomic studies have revealed that many are located in human pancreatic islet transcriptional cis-regulatory elements [6–9]. Numerous T2D risk loci have thus been assigned to effector transcripts through human islet expression quantitative trait loci (eQTLs), three-dimensional chromatin maps, and genome editing experiments [10–13]. Such studies have established that gene expression variation in pancreatic islets is critically important for T2D susceptibility. A subset of T1D susceptibility signals have also been proposed to act through expression variation in islet cells, sometimes involving shared genes with T2D [14]. A large fraction of T2D and T1D risk loci, however, cannot be ascribed to transcriptional regulatory mechanisms in pancreatic islets or other tissues, pointing to additional non-coding mechanisms that remain to be defined.

Alternative splicing of pre-mRNAs provides an additional mechanism whereby genetic variation can create functional diversity across human genomes. Rare mutations and common variants that influence pre-mRNA splicing have been linked to a broad range of human diseases [15–17]. The effects of genetic variants on pre-mRNA splicing in human pancreatic islets, however, are largely unexplored. GTEx, the most comprehensive catalogue of splicing QTLs (sQTLs) across human tissues, does not include human pancreatic islets [17]. Other studies have examined exon-level expression in islet RNAs [12, 18, 19], although this does not capture the complexity of alternative splicing and is confounded by total gene expression variation. We have now directly examined RNA splicing in a panel of 399 human islet samples and created an atlas of sQTLs. This showed that sQTLs impact key genes for islet biology and diabetes. We uncover sQTLs that show fine colocalization with genetic variants associated with T2D, discover new genetic associations, and expand the spectrum of putative gene effectors of disease susceptibility. We further reveal new candidate mediators of T1D susceptibility. These findings provide biological insights into genetic mechanisms underlying diabetes risk.

Results

To examine the impact of common genetic variation on RNA splicing in pancreatic islets, we aggregated RNA-seq and genotype data from 447 human pancreatic islet samples from four cohorts [11, 13, 18, 19] and processed them to yield 399 qualifying samples after applying genotype and RNA-seq quality controls (Fig. 1a). We corrected for known and unknown covariates (Additional file 1: Fig. S1) and performed QTL analysis on mRNA levels (eQTLs) and junction usage [20] (sQTLs), using 6.46 million high-quality common variants (Additional file 1: Fig. S2a). Focusing on genes expressed in >10% of samples in each cohort, we found *cis*-eQTLs in 3433 genes (eGenes) at FDR ≤ 1% (Fig. 1a, Additional file 2: Table S1 [21]). This analysis also revealed a widespread impact of genetic variants on splicing variation, with 4858 *cis*-sQTLs at FDR ≤ 1%, 25% of which showed >10% shift in splice site usage in reference vs. alternate alleles (Fig. 1a, b, Additional file 3: Table S2 [21]). The 4858 sQTL junctions included alternative usage of 5′ exons, 3′ exons, mutually exclusive or skipped exons, or influenced combinations of

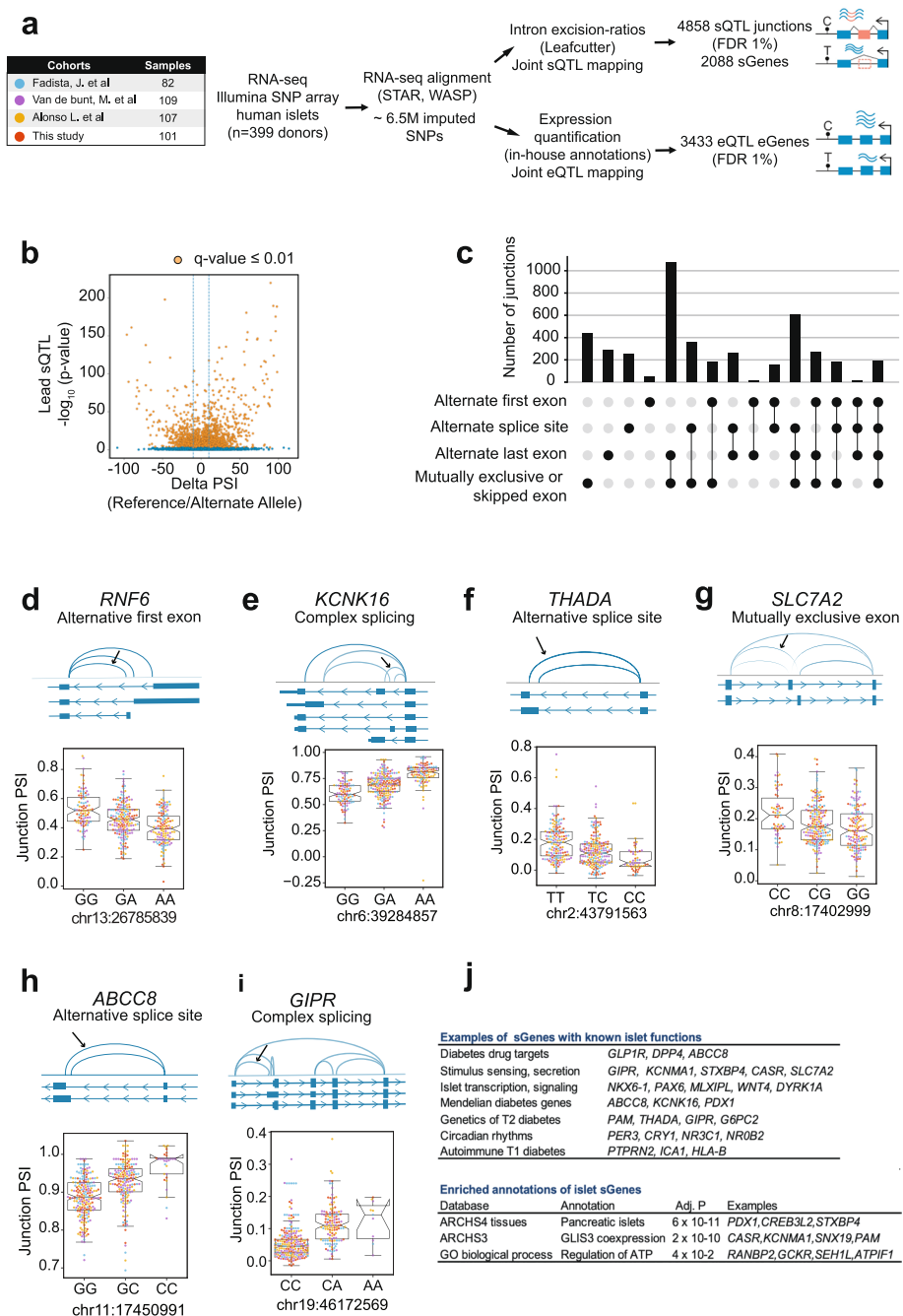Atla *et al. Genome Biology*   (2022) 23:196

Page 3 of 28

Fig. 1 Mapping sQTLs and eQTLs in human pancreatic islets. **a** Overview of the study design. **b** Volcano plot showing the reference to alternate allele change in percentage splice index (Delta-PSI) for junctions, and sQTLs -$\log_{10}$ *p*-values. Orange dots depict sQTLs junctions with $q \leq 0.01$. **c** Classification of sQTLs according to types of splicing events. **d–i** Selected examples of sGenes with different types of splicing events. An arrow signals the sQTL junction with best *p*-value, and adjacent boxplots show normalized, batch-corrected junction PSI values stratified by the lead sQTL genotype (IQR and 1.5 × IQR whiskers). Junction PSI values are colored according to the human islet dataset they belong to (see **a**). All boxplots show sQTLs with permutation *p*-values significant at FDR ≤ 1%, see Additional file 3: Table S2. **j** Functional annotations of sGenes. The top panel shows a manually curated list of examples with known functions in islet function and diabetes (see Additional file 4: Table S3); the bottom panel shows enriched annotations using EnrichR and Benjamini–Hochberg-adjusted *p*-values

such splice variants (Fig. 1c–i). The junctions mapped to 2088 distinct genes (sGenes), of which ~90% were known protein-coding genes, and ~7% were lncRNAs (Additional file 1: Fig. S2b).

We benchmarked human islet splice variants against GENCODE [22] and other available transcriptome maps and found that only 77% of the sQTL junctions were annotated. This overlap increased to 90% in comparisons with unpublished human islet transcript annotations built with long-reads (Atla, G., Beucher, A., et al., unpublished) (Additional file 1: Fig. S2c). This suggests that human islet transcripts are still incompletely annotated, but nevertheless well captured by our analysis of splice junctions.

We compared islet sQTLs to previously reported exon-QTLs [12], a possible proxy for sQTLs, and found that only 18% of sQTL junctions were flanked by exons from exon-QTLs. Furthermore, when sQTL and exon-QTLs affected the same gene, there was limited linkage disequilibrium ($r^2 < 0.6$) between the lead sQTL and exon-QTL variant for 45.2% of overlapping genes (Additional file 1: Fig. S2d, e). This indicates that sQTLs, which directly measure splice junction variation, and exon-QTLs, which measure exon expression levels and can thus be influenced by variables unrelated to RNA splicing, mostly capture fundamentally different events.

Islet sGenes were enriched in islet-specific co-expression networks (Fig. 1j) and included numerous genes with well-established roles in islet biology and diabetes (Additional file 4: Table S3), including major drug targets (*GLP1R*, *DPP4*, *ABCC8*), regulators of hormone secretion (*SLC7A2*, *CASR*, *GIPR*), transcription (*NKX6-1*, *PAX6*, *MLXIPL/ ChREBP*), signaling (*DYRK1A*, *WNT4*), or circadian rhythms (*PER3*, *CRY1*, *NR3C1*, *NR0B2*). Importantly, sQTLs also affected genes that harbor variants that cause monogenic diabetes (*KCNK16*, *ABCC8*, *PDX1*) or influence T2D risk (*THADA*, *PAM*) as well as genes that play a role in T1D pathogenesis (*ICA1*, *PTPRN2*, *HLA-B*) (Fig. 1d–j, Additional file 1: Fig. S3, Additional files 3 and 4: Table S2-S3). Our results, therefore, disclosed a pervasive impact of common genetic variants on alternative splicing of human pancreatic islet transcripts, including numerous genes that are important for islet function and differentiation, diabetes treatment, or pathophysiology.

sQTLs and eQTLs are distinct

We next explored the extent to which genetic effects on splicing and expression of islet transcripts were distinct. Only 34% of genes that harbored sQTLs (715 sGenes) also harbored a significant eQTL (Fig. 2a). Furthermore, for those 715 common genes, the lead eQTL and sQTL were frequently not in linkage disequilibrium ($r^2 < 0.6$ for 57% of genes, <0.1 for 24% of genes, Fig. 2b). Thus, in most genes that harbored both eQTLs and sQTLs, these were driven by independent signals. This is illustrated by *RGS1*, encoding a regulator of G-protein signaling, which has independent variants affecting either mRNA expression or exon skipping (Fig. 2c). In keeping with these findings, eQTLs and sQTLs were enriched in different functional genomic annotations. sQTLs were predominantly enriched in 5′ or 3′ splice sites and exons, whereas eQTLs showed a predominant enrichment in active promoters and enhancers (Fig. 2d, Additional file 1: Fig. S2f, Additional files 5 and 6: Table S4-S5). We also found differences in the extent to which genetic effects on splicing or expression differed across tissues; ~60% of lead islet sQTLs showed significant sQTLs in <5 GTEx tissues [17], compared to ~30% of lead eQTLs, suggesting
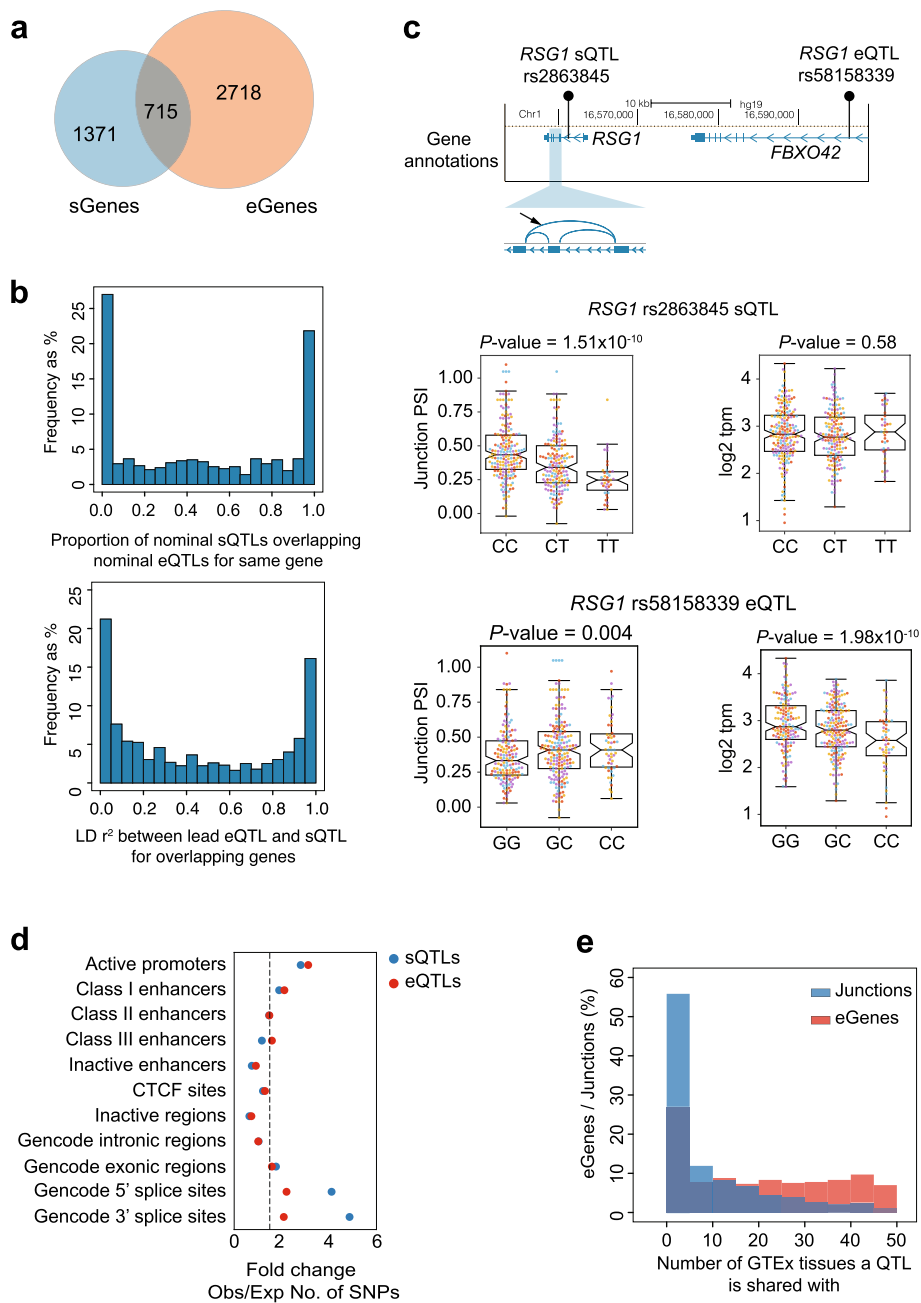
Atla *et al. Genome Biology*   (2022) 23:196

Page 5 of 28

Fig. 2. sQTLs and eQTLs are distinct genetic signals. **a** Overlap of sGenes and eGenes. **b** For 715 genes that have both eQTLs and sQTLs (overlapping genes in **a**), the top histogram shows the distribution of the percentage of variants shared between sets of nominally significant eQTLs and sQTLs. The bottom histogram shows the distribution of LD (r^2) values between the lead eQTL and sQTL. **c** *RSG1* has a distal eQTL, located in an intron of the *FBXO42* gene, and an intronic sQTL, both of which are in low LD (r^2=0.25). Boxplots represent *RSG1* expression and junction PSI values for both sQTL and eQTL, showing that the lead eQTL rs58158339 is not associated with *RSG1* splicing and the sQTL rs2863845 is not associated with expression. Boxplots show normalized, batch-corrected expression or junction PSI values stratified by the genotype of the lead QTL (IQR boxes and 1.5 × IQR whiskers). Individual samples are colored according to the human islet dataset they belong to (see color legend in Fig. 1a). Nominal QTL *p*-values are provided. **d** Enrichment of sQTL and eQTL variants in different functional genomic annotations using GREGOR. The *x*-axis represents GREGOR fold change of observed vs. expected number of SNPs at each functional annotation. The dotted line represents 1.5-fold change. **e** Percentage of eGenes and Junctions with eQTLs or sQTLs at FDR ≤ 1%, respectively, shared in different number of GTEx tissues. We used significant eQTLs and sQTLs from 47 distinct GTEx V8 release tissues

Atla *et al. Genome Biology*   (2022) 23:196

Page 6 of 28

a significant islet-specific component of sQTLs (Fig. 2e). Taken together, our results reveal two separable layers of genetic regulation of the human islet transcriptome.

Islet sQTLs provide new T2D and glycemic trait targets

Genetic susceptibility to T2D has been linked to sequence variants that influence gene transcription in human islets [10–13], but the relationship with islet splicing has not been systematically explored. To examine the potential contribution of islet sQTLs to T2D genetic associations [1], we first used quantile-quantile plots that compare the distribution of T2D association *p*-values of sQTL and eQTL variants against an expected null distribution (Fig. 3a). As anticipated, eQTLs showed strong inflation of more significant T2D association *p*-values. Remarkably, sQTLs also showed genomic inflation of T2D risk *p*-values (Fig. 3a). Importantly, this effect was maintained with islet-selective sQTLs (junctions with sQTLs in $\leq$ 5 GTEx tissues), or after omission of sQTLs that had high linkage disequilibrium with eQTLs (lead sQTL is in $r^2 \geq 0.6$ with the lead eQTL for the same gene) (Additional file 1: Fig. S4a, b). Furthermore, sQTLs also showed genomic inflation of low association *p*-values with T2D-related traits such as fasting glycemia (FG) or fasting insulin (FI) [23] (Additional file 1: Fig. S4c, d). These observations, therefore, suggest that splicing variation in human islets could contribute to T2D genetic susceptibility.

We reasoned that if splicing variation is linked to disease susceptibility at specific loci, a fraction of sQTLs and T2D associations should show high colocalization evidence (posterior probability of shared association between both phenotypes $\geq$ 0.8), and this could in turn point to specific candidate effector or causal transcripts underlying disease pathophysiology. To this end, we performed a systematic colocalization analysis between our islet sQTL or eQTLs, and independent GWAS signals for T2D ($n = 403$) [1] or glycemic traits (fasting glycemia, FG/fasting insulin, FI) ($n = 274$) [23]. We applied colocalization as implemented in *gwas-pw* [24], which draws upon the original *coloc* algorithm but does not rely on user-defined priors. This identified candidate effector transcripts with high colocalization support (posterior probability of shared association between both phenotypes $\geq$ 0.8) at 9 independent T2D GWAS signals using sQTLs, and 25 using eQTLs (Additional files 7 and 8: Table S6-S7, Additional file 1: Fig. S6a). At loci

(See figure on next page.)

Fig. 3 Role of human islet splicing variation in T2D susceptibility. **a** Quantile-quantile (QQ) plot showing observed T2D association *p*-values in human islet sQTLs (orange dots) and eQTLs (blue dots) against *p*-values under the null hypothesis. The grey-shaded region represents 1000 *p*-value distributions (in the -log$_{10}$ scale) of random sets of control sQTL variants (see the "Methods" section). Each set of control variants matches the number of islet sQTLs plotted. **b** Manhattan plot of splicing associations with T2D susceptibility (sTWAS). The *y*-axis shows -log$_{10}$ TWAS association *p*-values. Significant sTWAS associations in known T2D GWAS loci are colored in purple, and in previously unreported loci in orange. **c**, **d** Regional T2D GWAS signal plots for *PTPN9* and *PKLR* loci, two known T2D susceptibility regions, which showed significant sTWAS signals in *MAN2C1* (**c**) and *SCAMP3* (**d**) genes, respectively. Both splicing QTL effects in *MAN2C1* and *SCAMP3* are significant at FDR $\leq$ 1%; see Additional file 3: Table S2. LocusZoom plots show -log$_{10}$ T2D association *p*-values and locations in hg19 genome build. LocusCompare scatter plots show sQTL and T2D GWAS association *p*-values (-log$_{10}$ scale), illustrating co-localization of variants for both traits. Variants are colored according to the LD correlation (r^2) with the lead GWAS variant of the sTWAS association (purple diamond). Boxplots show normalized, batch-corrected junction PSI values on *y*-axis stratified by the genotype of the lead GWAS variant from the sTWAS association. Boxplots follow the color-to-batch legend from Fig. 1a. **e** Known T2D loci with target effector transcripts nominated by sQTLs and eQTLs from this study and/or eQTL maps from the InsPIRE consortium
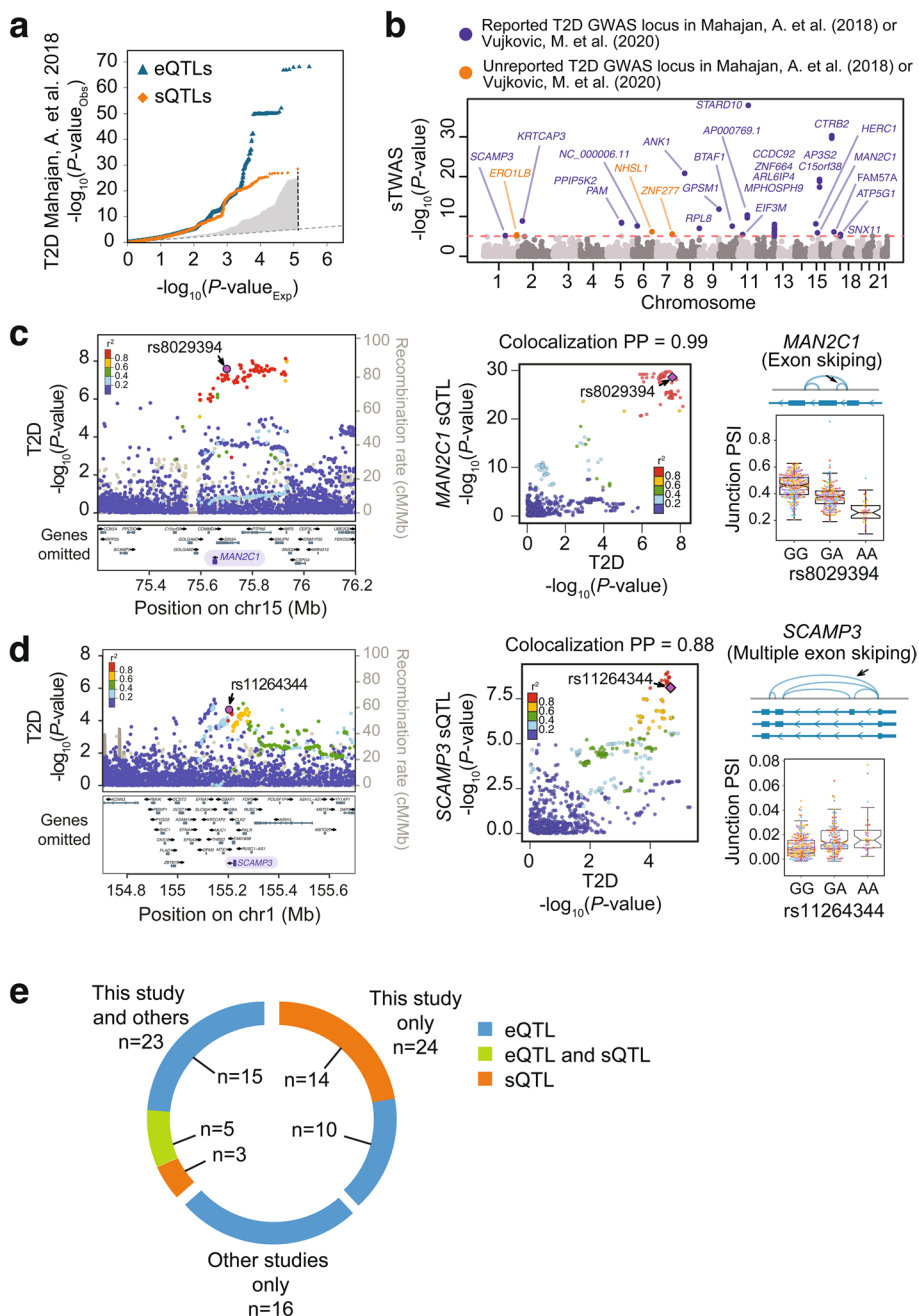
Fig. 3 (See legend on previous page.)

associated with glycemic traits, we found colocalization with sQTLs for eight putative target genes, and eQTLs for 17 genes (Additional files 7 and 8: Table S6-S7).

We further harnessed Transcriptome-Wide Association Studies (TWAS) to expand the collection of candidate effector genes for T2D and related traits. In this approach, which so far has not been applied to human islet RNA-seq data, genetic effects on splicing or expression are used to impute transcript variation in cases versus controls from GWAS datasets. This allowed us to identify splicing or expression changes associated with T2D and related traits. More specifically, we used the FUSION algorithm [25] and

Atla *et al. Genome Biology*   (2022) 23:196

Page 8 of 28

GWAS summary statistics [1] to identify T2D associations with islet junction usage or gene expression (sTWAS and eTWAS, respectively). Because TWAS findings do not necessarily distinguish between shared genetic effects and linkage [26, 27], we focused on TWAS signals showing colocalization with the GWAS phenotype (PP4 $\geq$ 0.6), thus minimizing confounding effects from linkage. This identified 27 genes (42 splicing events) showing significant sTWAS with T2D risk, and 29 genes with eTWAS after multiple testing correction (Bonferroni $p= 8.6 \times 10^{-6}$ and 1.8×10^{-5} after correcting for 5804 splicing junctions and 2851 genes, respectively) (Fig. 3b, Additional file 1: Fig. S4e, S5, Additional files 9 and 10: Table S8-S9). For glycemic traits, we identified 22 candidate target genes (43 splicing junctions) of GWAS signals via sTWAS, and 16 candidate target genes via eTWAS (Additional file 1: Fig. S4f-i, Additional files 9 and 10: Table S8-S9).

As expected, most TWAS signals fell in loci showing significant T2D and glycemic trait associations in GWAS (Fig. 3b, Additional file 1: Fig. S4e-i, S6b, Additional files 9 and 10: Table S8-S9). This included a sTWAS association for *MAN2C1*, encoding α-Mannosidase that has been implicated in mitochondrial-induced apoptosis and tissue damage [28, 29] (Fig. 3c). However, sTWAS revealed six T2D associations that did not reach genome-wide significance in the reference GWAS [1], three of which (*SCAMP3*, *SNX11*, and *FAM57A*) were nevertheless significant in a recent trans-ancestral meta-analysis for T2D [3] (Additional file 9: Table S8). *SCAMP3* encodes a vesicular transport protein [30] with unknown function in islets (Fig. 3d). Other significant sTWAS for T2D did not reach significance in GWAS reported so far, namely those encoding *ERO1B*, *NHSL1*, and *ZNF277* (Fig. 3b, Additional file 9: Table S8; further details for *ERO1B* are shown in Fig. 4g). Thus, sTWAS nominated putative effector targets for T2D and glycemic trait genetic associations, and identified additional genetic associations.

Our studies also highlighted significant eTWAS for T2D, including *PCBD1*, which is mutated in a syndrome that includes monogenic diabetes [31, 32], and encodes a co-factor of HNF1A, an established monogenic diabetes gene [33] (Additional file 1: Fig. S6c, Additional file 10: Table S9). This locus was not significant in the reference GWAS [1] but another recent meta-analysis detected a significant association in this locus [3]. Our findings indicate that *PCBD1* is a strong candidate effector transcript for this T2D susceptibility locus. Previously unreported T2D genetic association signals were found through eTWAS for *CTC-228N24.2* and *VSNL1*, a calcium sensor that modulates cAMP and insulin secretion [34] (Additional file 1: Fig. S6d, Additional file 10: Table S9).

Taken together, the combination of our TWAS and QTL colocalization results identified candidate effector transcripts for 47 known T2D susceptibility loci, including 22 acting through sQTLs, and 30 through eQTLs. Out of these 47 loci, 24 lacked information about likely effector transcripts in a recent large-scale islet eQTL analysis [12]. In aggregate, our sQTL and eQTL analysis increases the number of T2D loci with candidate effector genes supported by molecular QTLs to 63, a 1.6-fold change relative to the largest islet eQTL analysis so far [12] (Fig. 3e, Additional file 11: Table S10). Likewise, the integration of our molecular QTLs and GWAS summary statistics data of glycemic traits from a recent trans-ancestral meta-analysis [23] allowed us to increase the number of GWAS loci with candidate effector genes up to 43, a 3.1-fold increase relative to the largest islet eQTL mapping study so far (Additional file 12: Table S11).
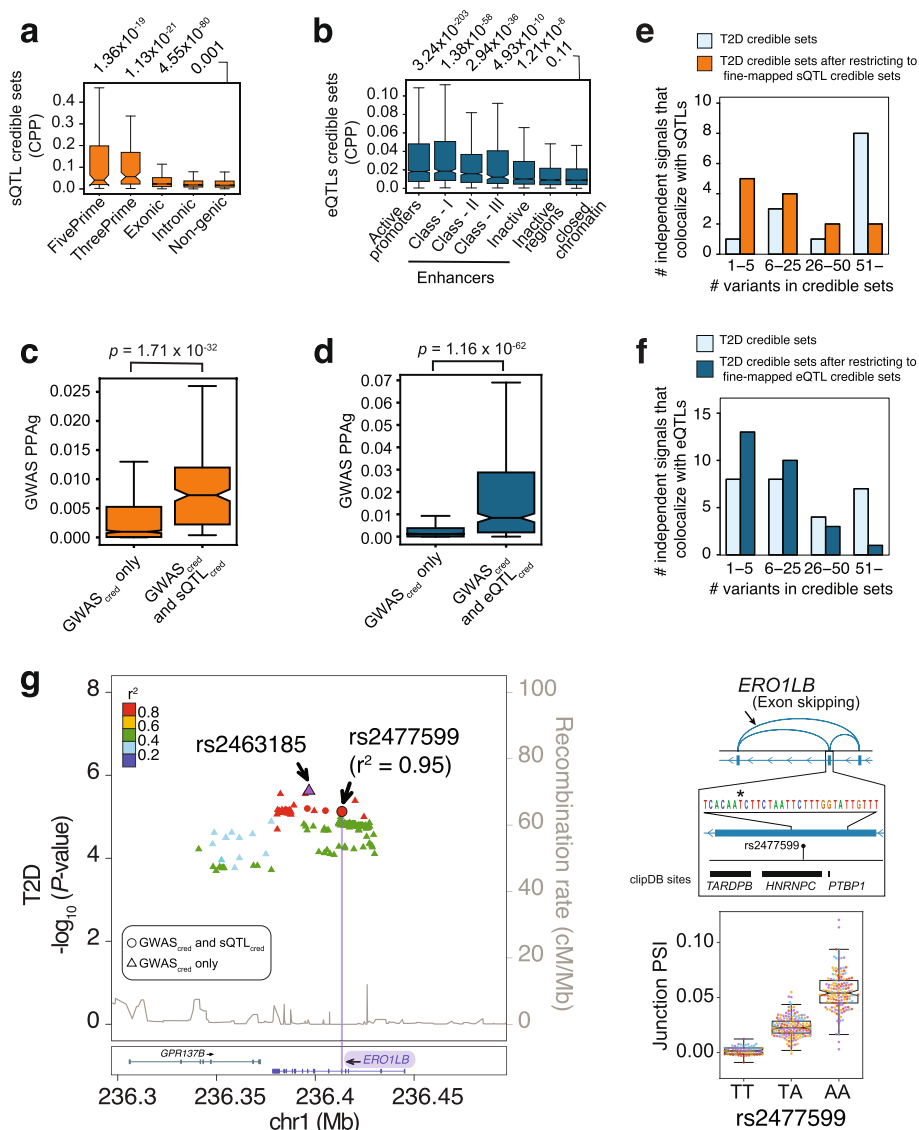
Atla *et al. Genome Biology* (2022) 23:196

Page 9 of 28

Fig. 4 Fine-mapping causal variants for known and novel T2D genetic associations. **a** Distribution of sQTL causal posterior probabilities (CPP) across different genic and non-genic regions. *P*-values on top correspond to Mann-Whitney comparisons with non-genic regions. **b** eQTL causal posterior probabilities across epigenomic annotations. *P*-values on top correspond to comparisons with credible set variants that fall outside islet epigenomic annotations (*closed* chomatin regions). **c**, **d** For all T2D-associated loci that colocalize with an islet QTL, we examined all fine-mapped variants (99% credible sets in GWAS, GWAS$_{cred}$) and compared the distribution of T2D causal posterior probabilities for variants that are also fine-mapped QTL variants (QTL$_{cred}$) vs. those that were not fine-mapped QTL variants. Mann-Whitney *p*-values are provided. Boxplots show IQR without outliers although *p*-values were calculated using all data points. **e, f** Integration of T2D GWAS credible set variants with credible sets from colocalizing sQTLs and eQTLs increases fine mapping resolution. Bar plots show the number of independent signals that fall into different bins of number of candidate causal variants before and after restricting for QTL variants. **G** Fine-mapping an sQTL and T2D association in *ERO1B*. The splicing QTL effect on *ERO1B* is significant at FDR ≤ 1%; see Additional file 3: Table S2. The LocusZoom shows T2D association -log$_{10}$ *p*-values; credible set variants for GWAS and sQTLs are shown as circles, and other GWAS credible set variants as triangles. The color of dots reflects r^2 with the lead GWAS variant (in purple) and includes the best fine-mapped candidate causal sQTL (rs2477599). The bottom inset depicts the alternative splicing event, along with the candidate causal sQTL variant and clipDB RBP binding sites. Boxplots are as described in Figs. 1, 2, and 3.

Atla *et al. Genome Biology* (2022) 23:196

Page 10 of 28

Splicing variants uncover candidate mediators of T1D risk

We next explored the role of islet splicing variation in T1D. Despite the fact that T1D is an autoimmune disease, several observations indicate that a fraction of T1D risk alleles act through β-cells or exocrine pancreatic cells, rather than directly through immune cells [5, 14, 35]. Furthermore, significant changes in RNA splicing have been reported upon cytokine-stimulation of human islets [36]. We observed that both islet eQTLs and sQTLs show inflation of more significant T1D association values [5] (Additional file 1: Fig. S7a), and identified eQTLs and sQTLs that exhibit significant associations with T1D (Additional file 1: Fig. S7b, see complete list in Additional files 7, 8, 9 and 10: Table S6-S9). For example, islet sQTLs at *MEG3*, *ESYT1*, and *DCLRE1B* co-localize with T1D association signals (Additional file 1: Fig. S7b-d, S8a, Additional file 9: Table S8). *ESYT1* is highly expressed in β-cells and encodes extended synaptotagmin 1 (E-Syt1), which has been implicated in insulin secretion regulation by clearing the plasma membrane of diacylglycerol [37], while *MEG3* and *DCLRE1B* have biological functions consistent with a role in autoimmune diabetes. *MEG3* is an imprinted lncRNA that is strongly expressed in human β-cells and regulates cell-specific cytokine responses [38, 39]. Downregulation of *MEG3* has been observed in islets of the murine model of T1D, the nonobese diabetic mouse (NOD) [40], while targeted inhibition increases the sensitivity of β-cells to cytokine-mediated oxidative stress [41]. Studies in T1D models and human islets have indicated that β-cell senescence contributes to autoimmune β-cell destruction, while *DCLRE1B* protects telomeres from replicative damage and prevents cellular senescence [42–44]. The *DCLRE1B* sQTL and sTWAS signal is in high local LD with one of the strongest T1D association signals at *PTPN22*, yet remains independent after conditional and joint (cojo-GCTA) variant analysis (Additional file 1: Fig. S8b-c). Taken together, the results nominate islet splicing and expression variants as potential molecular mediators of T1D predisposition.

Fine-mapping QTLs identifies candidate causal variants for T2D risk

Genetic fine-mapping of molecular QTLs has been shown to aid in the identification of effector genes as well as to identify the causal variants for complex traits [45]. To this end, we generated 95% credible sets of sQTLs and eQTLs. For sQTLs, the highest credible set posterior probabilities (CPP) of fine-mapped variants were observed for markers in 5′ and 3′ splices sites, followed by exonic and intronic regions (Mann-Whitney $p = 1.36 \times 10^{-19}$, 1.13×10^{-21}, 4.55×10^{-80}, and 1×10^{-3}, respectively, compared with intergenic variants) (Fig. 4a). Seqweaver, which provides variant effect predictions using RBP-based deep learning models [46], showed increased disease impact scores (DIS) of sQTLs in analogous annotations (Additional file 1: Fig. S9a). Furthermore, fine-mapped intronic and exonic sQTLs disrupted motifs of auxiliary splicing regulators (*SRSF3, SRSF9, HNRNPA1, HNRNPC)*, while those near 3′ splice sites showed recurrent disruption of branch point motifs and core splicing components [47, 48] (Additional file 1: Fig. S9c). These orthogonal analyses were consistent with known determinants of RNA splicing and highlight the potential of our fine-mapped sQTLs to prioritize causal variants.

Analogously, genetic fine-mapping of eQTLs revealed higher CPPs for variants in islet active promoters and enhancers (Mann-Whitney $p = 3.24 \times 10^{-203}$, 1.38×10^{-58}, in

promoters and mediator-enriched enhancers, respectively) (Fig. 4b). Similar enrichments were obtained with DeepSEA [49] (Additional file 1: Fig. S9b), and for disruption islet TF motifs [50–52] (Additional file 1: Fig. S9d). These results again supported that fine-mapped QTL variants have increased likelihood of driving splicing and expression variation in human islets.

Next, we hypothesized that if fine-mapped QTL variants are truly enriched in causal T2D variants, they should converge with variants that have highest CPPs in GWAS fine-mapping studies. To investigate this, we examined 99% credible set variants from T2D GWAS signals [1] with colocalizing QTLs (PP4>0.8, 16 loci for splicing, 28 loci for expression). We observed that GWAS credible set variants that were also in sQTL credible sets had higher GWAS CPP than GWAS credible variants that were not in the sQTL credible sets (Mann-Whitney $p= 1.71 \times 10^{-32}$) (Fig. 4c). Likewise, GWAS credible set variants that were also in eQTL credible sets showed higher GWAS CPP than other GWAS credible set variants (Mann-Whitney $p= 1.16 \times 10^{-62}$) (Fig. 4d). This convergence between GWAS and QTL fine-mapping provided further evidence that QTLs are likely to contain causal T2D risk variants. The integration of QTL and GWAS credible sets increased the number of associated loci with ≤ 5 putative causal variants from one to five loci by integrating fine-mapped sQTLs, and 8 to 13 loci with fine-mapped eQTLs (Fig. 4e, f; see also examples in Fig. 4g and Additional file 1: Fig. S9e). Thus, fine mapping QTLs has the potential to boost the genetic resolution of GWAS credible sets.

ERO1B represents an example in which we fine-mapped a putative causal variant for a candidate effector gene. *ERO1B* showed a significant sTWAS T2D association in our studies ($p= 6.5 \times 10^{-6}$, significant at FDR $\leq$ 1%). Fine-mapping highlighted an exonic rs2477599 variant that localizes to a splicing silencer *HNRNPC* motif, which causes an exon skipping event that results in premature truncation of the *ERO1B* open reading frame (Fig. 4g). Previous genetic loss-of-function studies have shown that *ERO1B* (endoplasmic reticulum oxidoreductase 1 beta, also known as *ERO1LB*) is an ER protein that plays a critical role in insulin biosynthesis and β-cell survival [53–55]. rs2477599 has not been previously reported in T2D GWAS studies, but shows suggestive associations in GWAS for T2D ($p= 8.2 \times 10^{-6}$ in [1], and $p= 3.42 \times 10^{-5}$ in FinnGen Biobank, data freeze 7) and random glucose ($p= 8.3 \times 10^{-5}$ UK Biobank Mendelian traits [56]). These findings point to a fine-mapped putative causal splicing variant and a plausible effector mechanism for T2D susceptibility.

sQTLs also shed interesting findings at the *BCAR1*/*CTRB2* locus. GWAS have fine-mapped an intergenic variant near *CTRB2* (rs72802342, CPP=0.66) that is associated with T2D [1], T1D [5], and pancreatic ductal adenocarcinoma (PDAC) [57] and colocalizes with an accessible chromatin region in human acinar cells and islets [5, 58]. *CTRB2* showed significant sTWAS associations with T2D and T1D (Additional file 9: Table S8), and our fine-mapped sQTLs pointed to the same lead candidate causal variant, rs72802342, which appeared to cause a complete exon-skipping event in CTRB2 transcripts (Additional file 1: Fig. S9e). Recent work, however, has shown that rs72802342 tags a 584 bp deletion that completely overlaps exon 6, and explains our observed sQTL [57]. Thus, an exon-skipping deletion at *CTRB2* (rather than an exon skipping splice variant) is likely to be the functional causal variant underlying T2D, as well as T1D and PDAC, susceptibility.

Islet QTLs provide insights into T2D pathways

The availability of an expanded list of candidate effector genes, as opposed to lists of genes that are located in the vicinity of associated SNPs, allowed us to explore the hypothesis that a subset of genetic signals could influence T2D susceptibility by acting on specific cellular pathways in pancreatic islet cells. We thus compiled 106 putative effector genes for T2D or glycemic traits (FG, FI) reported here as well as previously reported co-localizing islet eQTLs [12]. This list excluded those exclusively detected in non-endocrine cells after analysis of single cell RNA-seq datasets (Additional file 13: Table S12), as well as non-coding transcripts and genes without known function. Functional gene annotations revealed notable enriched pathways, including regulators of fatty acid biosynthesis (*SCD5*, *FADS1*, *GCDH*, *BDH2*, *PAM*; GO:0006636, GO:1901570, ENRICHR adjusted $p = 0.04$), and genes upregulated by hypoxia or mTORC1 activation (MsigDB Hallmark Hypoxia and mTORC1 signaling; $q < 0.05$ and < 0.001, respectively). We further used the list of 106 genes to identify networks using STRING [59] v11.5, using co-expression, experimental, and functional annotation databases. We allowed for minimal network inflation (≤ 5 interactions) and omitted text-mining to preclude bias arising from publications that name genes at significant GWAS loci. The resulting network exhibited 1.8-fold greater protein-protein interactions than random sets (PPI enrichment $p < 10^{-4}$) and contained two distinct subnetworks (Fig. 5a). One subnetwork contained components of the eIF3 translational initiation complex (FDR = 0.018). Several monogenic diabetes genes target translational initiation, in particular eIF2 complex (*EIF2B1*, *EIF2S3*, *EIF2AK3*), due to their impact on endoplasmic reticulum stress in β-cells [60, 61] (Fig. 5a). This process is also regulated by ERO1B [55]. Another notable sub-network was formed by molecular mediators of G-protein mediated enhancement of cAMP, a major insulinotropic pathway [62, 63]. Manual curation revealed eight well-characterized genes in this pathway, including *GLP1R*, *RGS17*, *PDE8B*, four members of the GNAI3 (G(i) subunit α3) protein-protein interaction complex (*GPSM1*, *RGS19*, *ADRA2A*, *ADCY5*), and *VSNL1*, a modulator of cAMP and insulin secretion [34] (Fig. 5a, b).

Discussion

This study adds splicing variation in pancreatic islets to the spectrum of molecular mechanisms that underlie T2D predisposition. Earlier studies had examined expression QTLs in human pancreatic islets [10–13, 18, 19]. The current study offers the first systematic analysis of how common genetic variants influence RNA splicing in human islets. Parallel profiling of splicing and expression QTLs in the same dataset demonstrated that these represent two distinct mechanisms through which genetic variation can influence islet biology and disease. We found islet sQTLs that impact genes that have major roles in islet cell function as well as in the pathophysiology or treatment of various forms of diabetes. Furthermore, we observed a selective inflation of T2D association p-values among sQTLs, some of which showed fine co-localization with T2D variants. Finally, we applied for the first time splicing and expression TWAS to nominate T2D target genes and identified novel T2D genetic association signals.
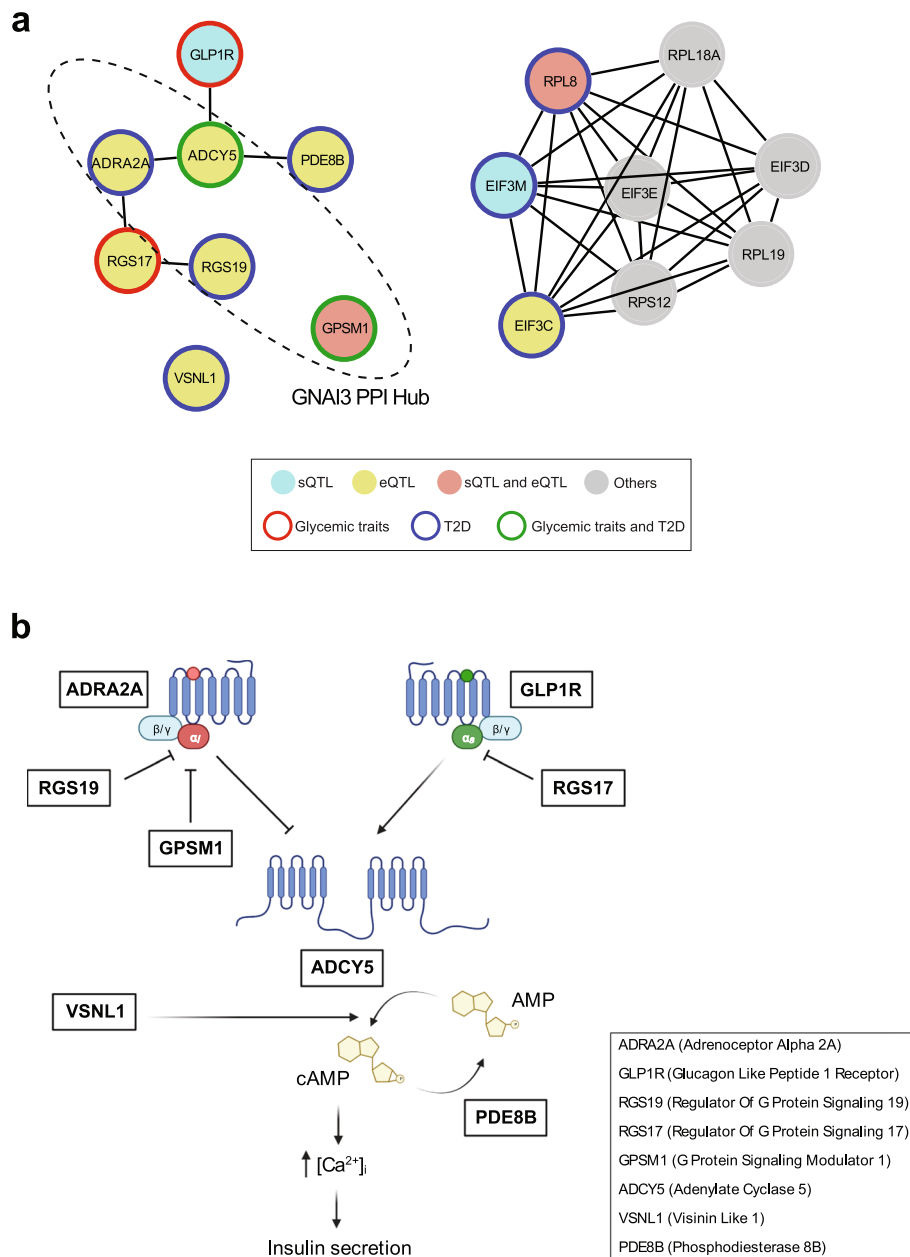
Fig. 5 Islet QTLs that co-localize with T2D signals target distinct pathways. **a** STRING v11.5 was used to analyze 106 genes with sQTLs or eQTLs co-localizing with T2D or glycemic traits, including previously reported eQTLs. We only considered genes with known or presumed protein-coding function, and not exclusively expressed in non-endocrine cells according to single cell RNA-seq. We allowed for inflation of ≤ 5 interactors, and used default (>0.4) confidence scores. Shown are two networks with >3 components, one populated by components of heterotrimeric G protein signaling, and another by genes involved in eIF3 translational initiation. Genes are colored as indicated in the legend; other interactors added through STRING analysis are shaded gray. **b** Manual curation was used to illustrate the relationship between components of the G protein-mediated insulinotropic pathway targeted by islet QTLs linked to T2D and related traits

Gene targets that are nominated based on human genetic evidence double their likelihood of success in drug development pipelines [64]. We have expanded the current list of putative effector genes that mediate diabetes risk at known and novel susceptibility regions, and include examples that provide insights into T2D pathophysiology.

Atla *et al. Genome Biology* (2022) 23:196

Page 14 of 28

This was illustrated by the association of T2D risk with a fine-mapped splice variant that creates a premature stop codon in *ERO1B*. This newly reported locus is supported by an independent T2D association signal observed in FinnGen, as well as prior experimental and genetic studies that point to a central role in ER homeostasis, insulin biosynthesis, and diabetes progression [53–55]. Our analysis of sQTLs revealed a spurious exon-skipping splicing event in *CTRB2* caused by a deletion. This exon deletion, which has been previously characterized in the context of PDAC susceptibility [57], provides a more compelling candidate molecular driver for T2D and T1D susceptibility at this locus than recently reported transcriptional variants [5, 58] and can be tested in models to determine pathogenic mechanisms.

T2D has a highly polygenic architecture, with a very large number of effector genes that individually exert small effects. It remains possible, however, that many such genes converge on a small number of biological pathways. The extended list of genes with islet QTLs that co-localize with GWAS signals allowed us to explore shared pathways underlying T2D susceptibility. We found supportive evidence for a role in translation, ER stress, and fatty acid metabolism, and, most clearly, eight out of 106 QTLs that co-localized with glycemic trait or T2D associations were linked to G protein-coupled receptor (GPCR) signaling through cAMP. This pathway has been extensively involved in transducing signals from a vast range of extracellular insulinotropic stimuli, including incretin hormones, neurotransmitters, nutrients such as fatty acids, or extracellular matrix components [62, 65, 66]. Effectors from this pathway included an sQTL for *GLP1R*, a major drug target for T2D [67], four members of the same protein interaction complex, and a novel T2D association for *VSNL1*, previously shown to stimulate cAMP production and insulin secretion [34]. These findings, therefore, suggest that abnormal production of cAMP plays a causal role in T2D susceptibility, plausibly through its impact on insulin secretion and islet gene transcription. This pathway is a known therapeutic target to stimulate insulin secretion. However, our findings raise the additional prospect that it is possible to target altered GPCR signaling in precision medicine strategies that aim to correct causal defects, and thereby modify disease progression in susceptible individuals.

Our resource of islet sQTLs is relevant for efforts to dissect multi-allele interactions. There is growing evidence that disease-associated haplotypes can contain more than one functionally interacting causal variant [68]. Islet splicing variants can thus act in *cis* with other functional variants to influence disease susceptibility. For example, sQTLs could modify the penetrance of causal coding or cis-acting variants in both polygenic or Mendelian settings. Likewise, we have shown that islet sQTLs can alter drug target genes and can thus be examined to understand how genetic variation alters the response to therapies.

Finally, this molecular QTL resource holds relevance for the biological interpretation of genetic variants that influence autoimmune T1D risk, in particular for efforts to link biological processes in β-cells into T1D predisposition. Of interest, our findings provide a genetic finding that aligns with recent proposals that intrinsic β-cell senescence processes contribute to autoimmune β-cell destruction [42, 43].

Our analysis of molecular QTLs entails some limitations. Splicing activity was estimated by junction usage, which does not directly inform isoform level regulation, and therefore provides a partial picture of how local splicing translates into gene function.

Furthermore, the analysis of molecular QTLs was limited to islets, where genetic variants are known to play a central role in diabetes mellitus, most clearly in T2D. Some disease associations are expected to be mediated through other cell types and the mediating mechanism is not captured by our studies. In addition, our study was performed on bulk pancreatic islet samples with heterogeneous cell-type compositions, after exposure to conventional culture conditions. This limits the power to discover new QTL signals that are confined to minor cell types, or conditioned by specific environmental contexts [69]. Furthermore, colocalization and TWAS evidence does not preclude horizontal pleiotropy, and we have thus not unequivocally linked non-coding variant disease associations with the causal molecular targets. Nevertheless, we have demonstrated genetic co-localization of disease risk variants and molecular variation in pancreatic islets, which play a central causal role in diabetes. This knowledge can thus be used to nominate plausible effector mechanisms and should warrant further orthogonal genetic and experimental studies to establish causality.

Conclusions

Our findings expose widespread effects of common genetic variants on RNA splicing in pancreatic islets, a tissue that is directly relevant to the pathophysiology of T1D and T2D. We highlight examples in which genetic and biological evidence supports a role in diabetes susceptibility. The results provide new avenues and resources to understand the genetic underpinnings of diabetes mellitus.

Methods

Human pancreatic islet datasets

We compiled 447 human islets samples, which comprised 89 samples from GEO accession number GSE50244 [19], 118 from EGA accession number EGAD00001001601 [18], 112 samples from T2DSystems EGA accession number EGAS00001005535 [13], and 128 samples from CRG cohort (EGAS00001006440) [70]. Out of 447 samples, 399 samples passed genotype and RNA-seq QC filters described below.

For the CRG cohort, human pancreatic islets from organ donors without a history of glucose intolerance were purified using established isolation procedures [71–74], shipped in culture medium and re-cultured at 37 °C in a humidified chamber with 5% CO_2 in glucose-free RPMI 1640 supplemented with 10% fetal calf serum, 100 U ml^{-1} penicillin, 100 U ml^{-1} streptomycin, and 11 mM glucose for 3 days before analysis. Islet isolation centers had permission to use islets for scientific research if they were insufficient for clinical transplantation following national regulations and ethical requirements and institutional approvals from University of Lille, University of Geneva, and Milano San Raffaele Hospital. Ethical approval for processing de-identified samples was granted by the Clinical Research Ethics Committee of Hospital Clinic de Barcelona and Parc de Salut Mar.

Genotype processing

The 128 CRG cohort samples were genotyped with Illumina Infinium OmniExpress 12 v1 and HumanOmni 2.5-8v1 on a total of 624K SNPs. The 89 samples from Fadista et al. (GSE50244) [19] were genotyped using Illumina HumanOmniExpress 12v1 C on

a total of 609K SNPs. The 118 samples from Van de Bunt et al. (EGAD00001001601) [18] were genotyped using Illumina Omni2.5+Exome array on a total of 2.6M SNPs. Finally, the 112 samples from T2DSystems (EGAS00001005535) [13] were genotyped using Illumina's Human Omni 2.5 exome array on a total of 2.6M SNPs.

A three-step quality control of genotype data, involving two stages of SNP removal and one intermediate stage of sample exclusion, was conducted in each cohort. Genotyped SNPs were filtered if (i) minor allele frequency (MAF) < 0.01, (ii) missing genotype rate $\geq$ 5%, and (iii) significantly deviated from Hardy-Weinberg equilibrium (HWE p-value $\leq 1\times10^{-6}$). Samples were excluded if (i) individual missing genotype rate $\geq$ 2%, (ii) cryptic relationships and sample duplicates (individuals with higher individual missingness genotype rate from pairs with pi $\geq$ 0.185), or (iii) showed >4 standard deviations from the mean according to the first four principal components in each given cohort.

After QC analysis for genotype data, (i) 113 individuals and 557,422 SNPs were retained for the CRG cohort, (ii) 112 individuals and 1,499,688 SNPs for the Van de Bunt et al. cohort [18], (iii) 89 individuals and 596,464 SNPs for the Fadista et al. [19] cohort, and (iv) 109 individuals and 1,543,968 SNPs for the T2DSystems cohort [13].

For each cohort, we generated per-chromosome VCF files after removing all strand ambiguous variants (AT or CG SNPs), and checking for strand alignment against the Haplotype Reference Consortium (HRC) and 1000 Genomes (1000G) reference SNP list. *HRC-1000G-check-bim.pl* script with the *-n* option (to turn off the removal of variants showing MAF differences between the reference panel and the study genotypes) was used. We submitted resulting VCF files to the Michigan Imputation Server (https://imputationserver.sph.umich.edu/index.html): EAGLE2 [75] was used for phasing, minimac3 [76] for genotype imputation with the HRC [77] r1.1 and the 1000G Phase 3 release reference panels [78], independently. For each dataset of imputed genotypes, we excluded variants with (i) MAF < 1%, (ii) imputation-quality R^2 < 0.7, and/or (iii) HWE p-value $\leq 1\times10^{-6}$. We extracted indels from the 1000G Phase3 imputed results, filtered them using the aforementioned criteria, and merged with the filtered HRC imputed dataset.

RNA-Seq data alignment and QC

Raw fastq files from all cohorts were aligned to the hg19 genome build using STAR [79] and the options, *--outFilterMultimapNmax 1 --outSAMstrandField intron-Motif --outSAMattributes All --twopassMode Basic*. WASP [80] pipeline was used to remove reads mapped with allelic bias. *VerifyBAMID* [81] was used to assess the concordance between genotypes and RNA-Seq, and samples with more than 2% contamination (CHIPMIX > 0.02 and FREEMIX > 0.02) were removed. This resulted in 101 samples from CRG cohort, 109 from samples Van de Bunt et al. cohort [18], 82 samples from Fadista et al. [19] cohort, and 107 samples from T2DSystems [13]. The resulting BAM files were used for gene expression quantification and to calculate splicing activity. In-case of EGAD00001001601 [18], only bam files were available; hence, the initial STAR alignment step was not performed.

Principal component analysis of genotypes

To identify individuals of divergent ancestry and to characterize population structure of 399 human islet samples, we first selected a subset of genotyped SNPs that were common in all 4 cohorts, that also passed all our QC filters (see the "Genotype processing" section), and with MAF $\geq$ 1% and missingness < 5% across all the samples. We also excluded SNPs in high LD (pairwise $r^2 \leq 0.1$ within 1 Mb window), C/G and A/T SNPs to avoid strand mismatches, and those located in previously reported regions with long-range LD. We aggregated the 1000 Genomes Phase3 reference dataset using the set of overlapping variants. flashPCA [82] tool was used to calculate genetic principal components (PCs) (Additional file 1: Fig. S1c-d). First four genetic PCs are used as covariates in subsequent sQTL and eQTL mapping analysis.

Gene expression quantification

In-house developed human pancreatic islet transcriptome annotations built with a combination of CAGE as well as short and long sequence reads were used to quantify total gene expression (Atla, G., Beucher, A. et al., unpublished). FeatureCounts [83] was used to obtain gene level qualifications using default parameters except using appropriate strandedness flag for each dataset. Genes with less than 5 raw reads mapped in less than 10% of the samples within each cohort were removed. CPM normalization was performed using edgeR [84] *cpm* function and then log2 transformed. Combat [85] was used to remove known sequencing batch effects. Principal components (PCs) were calculated on Combat corrected gene expression values using *prcomp* function in R (Additional file 1: Fig. S1a).

Splicing quantification

To quantify splicing activity, we used the annotation free method, *leafcutter* [20]. Briefly *bam2junc.sh* script from leafcutter was used to quantify junction spanning reads based on spliced alignments from bam files. We removed junctions that are not supported by at least 5 spliced reads in at least 10% of samples. We then used *leafcutter_cluster.py* script with options *-m 30 -l 500000* to cluster the junctions anchoring on shared splice sites. This identifies local splicing events which are a cluster of alternatively used junctions. *prepare_phenotype_table.py* script was used to get relative junction usage (ratios) across samples. The relative junction usage is also referred to as percent spliced in (junction PSI) in the manuscript. Combat [85] was used to remove known sequencing batch effects and principal components were generated using *prcomp* function in R (Additional file 1: Fig. S1b).

cis-eQTL mapping

cis-eQTL mapping was performed using QTLtools [86] for 399 samples with available QCed genotype and RNA-seq data using a cis-window of 500 kb up- and downstream of the transcription start site (TSS). Fifteen PCs derived from gene expression and 4 genetic PCs were used as covariates in the linear model. In order to identify best associated cis eQTL SNP-eGene pairs, QTLtools was run using the permutation pass mode using default parameters and *--permute 1000 --window 500000 --seed 123456.*

Beta approximated permutation *p*-values were adjusted for multiple testing correction using Storey *q*-values implemented in the *qvalue* R package. We set the significance threshold at FDR *q*-value ≤ 0.01. This resulted in 3433 genes (eGenes) with significant eQTLs (Additional file 2: Table S1). We also calculated nominal *p*-values for all cis-SNPs within a 500-kb window centered on the TSS of each gene (nominal pass mode from QTLtools, *--nominal 1 --window 500000 --seed 123456*). To identify all significant variant-gene pairs, we defined a genome-wide *p*-value threshold (*pt*), by considering the empirical *p*-value of the eGene closest to the 0.01 FDR threshold. A gene-based nominal *p*-value threshold was then calculated using *pt* and the beta distribution parameters from QTLtools. For each significant eGene, variants with a nominal *p*-value below the gene-level threshold were considered in subsequent analyses (significant nominal cis-eQTL variants) [21].

cis-sQTL mapping

We performed cis-sQTL mapping as described for cis-eQTL identification using intron excision ratios and a cis-window of 50 kb up- and downstream of the junction (*--window 50000*). In case of cis-sQTLs, 5 PCs derived from splicing ratios and 4 genetic PCs were used in the linear model. In order to identify best associated cis sQTL SNP-junction pairs, QTLtools was run using the permutation pass mode (1000 permutations), and beta approximated permutation *p*-values were adjusted for multiple test correction using Storey *q*-values implemented in the *qvalue* R package. We set the significance threshold at FDR *q*-value ≤ 0.01, resulting in 4858 junctions with a significant sQTL (Additional file 3: Table S2). We also calculated nominal *p*-values for all cis-SNPs within a 50-kb window around the junction (nominal pass mode from QTLtools). To identify all significant variant-junction pairs, we defined a genome-wide *p*-value threshold (*pt*) by considering the empirical *p*-value of the junctions closest to the 0.01 FDR threshold. A junction-based nominal *p*-value threshold was then calculated using *pt* and the beta distribution parameters from QTLtools. For each significant junction, variants with a nominal *p*-value below the junction-level threshold were considered in subsequent analyses (significant nominal cis-sQTL variants) [21].

Imputation information score of lead eQTLs and sQTLs

We used QCTOOL v2.0.6 to calculate imputation scores (*-snp-stats)* for each lead eQTL and sQTL identified in 3433 genes and 4858 splicing junctions at FDR ≤ 1%, respectively.

Annotation of sQTL junctions

As leafcutter identifies junctions de-novo, we used transcriptome annotation GTF files as a base to annotate the junctions with respective genes. We used *gtf2leafcutter.pl* script from leafcutter's leafviz module (https://github.com/davidaknowles/leafcutter/tree/master/leafviz) to obtain the intron coordinates of all genes. The sQTL junctions were then mapped to the intron coordinates and annotated with respective gene names.

Magnitude of genetic effects on splicing

To quantify the magnitude of genetic effects on splicing, for each cluster, we chose a junction with the best *q*-value. Then, for each sQTL junction, we calculated the

Atla *et al. Genome Biology* (2022) 23:196

Page 19 of 28

difference in median junction usage (delta-psi) from samples with homozygous reference and homozygous alternate alleles and plotted as a function of the -$\log_{10}$ (q-value). If alternate homozygous samples were not available, we chose median junction usage from heterozygous samples.

Visualization of splicing events

The junctions identified from leafcutter were loaded into IGV [87] for visualization as arcs, and box plots were plotted using python.

Comparison of eQTLs and sQTLs

Our maps of genetic effects on splicing were compared to our eQTLs but also to previously reported exonQTLs and eQTLs in the largest study to date in human pancreatic islets by the InsPIRE consortium [12]. To estimate the gain of novel information uniquely provided by our sQTL analysis, we first determined the overlap between sGenes and eGenes in our study, and eQTLs and exon QTLs mapped by InsPIRE. For those genes that contain both sQTLs and either eQTLs or e/exon QTLs in InsPIRE, we calculated the LD (r^2 measure) between the lead sQTL and all other SNPs within 1Mb using PLINK [88] (v1.9 *--ld-window-kb 1000 --ld-window 99999 --ld-window-r2 0*). As a reference dataset for LD calculations, we used our high-quality genotypes in 399 samples. Then, for each gene, we plot the LD r^2 distribution between the lead sQTL and the lead QTL from the dataset being compared, respectively.

Genomic enrichment analysis of eQTLs and sQTLs

We used GREGOR [89] to perform enrichment analysis of lead sQTLs and eQTLs in different genomic annotations using the following options R2THRESHOLD=0.7, LDWINDOWSIZE=50000 (for sQTLs), LDWINDOWSIZE=100000 (for eQTLs), MIN_NEIGHBOR_NUM=200, and POPULATION=EUR.

Human islet regulatory annotations were described previously [11]. For genic annotations, we used Gencode [22] v34 GTF file. We used *gtf2leafcutter.pl* script to obtain all exon, intron, 5′ and 3′ coordinates from GTF file. 5′ and 3′ coordinates were extended to +3bp into exon and +2bp into intron. All annotations contain mutually exclusive genomic space.

In parallel, we used GARFIELD [90] for enrichment analysis among genomic annotations. We used QTLtools beta approximated permutation p-values of lead eQTL and sQTLs from all genes and junctions tested in cis-eQTL and sQTL analysis. If a variant was lead eQTL or sQTL for multiple genes or junctions, we kept the lowest p-value. We used the same annotations as in GREGOR. For each set of eQTLs and sQTLs, the method first removes variants with $r^2 \geq 0.1$ in a 1-Mb window from the most significantly associated variant, and then links each of the independent variants to genomic annotations if either the variant or a variant in LD ($r^2 \geq 0.8$) overlaps a given feature. Finally, the method tests for enrichment with a logistic regression model that controls for confounding factors such as distance to the TSS, number of LD proxies, and MAF that we binned in 4 quantiles. We used three different QTL p-value thresholds to test for enrichment, 5×10^{-3}, 5×10^{-5}, and 5×10^{-7}. Significant enrichments for eQTLs and sQTLs were identified after applying multiple-test correction for the effective number of

Atla *et al. Genome Biology*   (2022) 23:196

Page 20 of 28

annotations (estimated by GARFIELD), the number of QTL *p*-value thresholds, and the number of molecular QTL phenotypes.

Comparison with GTEx

We obtained summary statistics data for eQTLs and sQTLs from 49 GTEx tissues [17]. We first obtained eGenes and junctions with significant eQTLs and sQTLs, respectively, at FDR 0.01, consistent with our significance threshold. For the resulting significant variant-phenotype associations for each of the 49 tissues, variant position and junction coordinates (for sQTLs) were lifted down from hg38 to hg19 using liftOver [91]. For each GTEx tissue, we looked at the variant-phenotype (eGenes or Junctions) overlap with our islet eQTL and sQTLs, using nominal QTL variants. For example, for GTEx *x* eGene in *j* tissue, if any of the GTEx significant variants mapped any of our nominal eQTL variants for that *x* eGene, we considered that islet eQTL signal to be shared with that given *j* tissue. Same approach was implemented to sQTLs. We excluded from this analysis testis, given the pervasive number of eQTLs, and pancreas because it is a partial surrogate of pancreatic islets.

Quantile-quantile plots

We generated quantile-quantile (Q-Q) plots to estimate genomic inflation in sQTLs and eQTLs for T2D and T1D risk, and glycemic traits variation. For T2D, we used BMI-adjusted T2D summary statistics [1]. For glycemic traits, we leveraged summary statistics data from a trans-ancestral meta-analysis for fasting glucose (FG) and fasting insulin (FI) [23]. For T1D, we used summary statistics data from a recent large-scale meta-analysis [5]. We included variants with MAF $\geq$ 5% that were intersected with our nominal e- and sQTLs. For each trait comparison, we generated 1000 permutations of subsets of control sQTL variants to provide further support to the observed enrichment of e- and s-QTLs among GWAS variants. Each control set of sQTL-like variants was generated by first identifying independent LD blocks [92], that comprised nominal eQTL or sQTL variants, respectively. Then, we shuffled non-overlapping genomic regions that were created based on the size of the genomic ranges where our nominal sQTL variants were located, across the genome, but excluding those independent LD regions where either eQTL or sQTL variants were located, blacklisted regions (wgEncodeDacMapabilityConsensusExcludable.bed.gz and wgEncodeDukeMapabilityRegionsExcludable.bed.gz) and the MHC region. Among the set of shuffled independent LD regions, we randomly sampled the same number of nominal sQTL variants. This was done 1000 times.

For the specific purpose of addressing T2D susceptibility across shared or independent sQTL/eQTL effects, we considered sQTLs and eQTLs from junctions and eGenes whose corresponding lead sQTL and eQTL were in $r^2 \geq 0.6$ as shared. The rest were considered as sQTL or eQTL specific genetic effects, respectively. To estimate T2D risk inflation across islet-selective sQTLs and eQTLs, we grouped sQTLs and eQTLs that were shared in $\leq$ 6 GTEx tissues (see the "Comparison with GTEx" section) as islet-selective QTL effects. To assess the strength of T2D-risk inflation, we leveraged 1000 permutations of control-sets of sQTL-like variants as described above.

Colocalization analysis across T2D, T1D and FG/FI independent GWAS signals

We collected 403 independent lead variants for T2D [1] and 277 for FG and FI [23] and 136 for T1D [5]. For each trait, we performed colocalization as implemented in *gwas-pw* [24] at each independent GWAS signal. We only tested for colocalization with e- and sQTLs at those signals with at least one credible set variant (genetic posterior probability ≥ 0.01) in LD ($r^2 \geq 0.6$) with a lead eQTL or sQTL, respectively. Credible sets were not available for FG and FI GWAS. Thus, we tested for colocalization at FG/FI GWAS signals if at least one variant with GWAS $p \leq 5\times10^{-5}$ was in LD ($r^2 \geq 0.6$) with a given lead eQTL or sQTL. GWAS signal LD calculations were performed using the high-quality genotypes from our 399 islet donor samples. If for a given independent GWAS signal, any of the available proxy variants was not included in our imputed genotypes, we used 1000 Genomes Phase3 genotypes (with European descent) as the reference panel [78]. Colocalization was performed across 1Mb genomic interval centered on the reported lead independent variant for a given GWAS locus, and including variants with a GWAS $p \leq 5\times10^{-5}$. We nominated a region as a colocalized locus if the posterior probability for the model 3 (presence of the same genetic variant underlying the QTL and GWAS association, "colocalization") was ≥ 0.8. Colocalization signals were visualized using R-3.6.1 and LocusCompareR 1.0.0.

Visualization of regional association plots

Regional association plots for T2D and T1D susceptibility regions were created using LocusZoom [93] v1.4 in R-3.6.1.

TWAS analysis

FUSION [25] was used for TWAS analysis with T2D, T1D, and glycemic traits (FG and FI). For gene expression, first the weights were computed using *FUSION.compute_ weights.R* and options *--models top1, blup, bslmm, lasso, enet* on the same data that is used for eQTL analysis. Fifteen gene expression PCs and 4 genetic PCS were used as covariates and variants within a cis-window of 500kb from TSS were used. For splicing analysis, variants within 50kb from the junction boundaries were used and 5 splicing PCs and 4 genetic PCs were used to compute the weights.

After computing weights, *FUSION.assoc_test.R* script was used to test for association of the pre-computed weights and GWAS summary statistics data for T2D [1], T1D [5], and glycemic traits [23] (FG and FI). We only included GWAS data from variants that overlaid our ~6.5M high-quality imputed common genetic variants. Multiple test correction was applied to the resulting *p*-values using Bonferroni. We also excluded TWAS results with low colocalization posterior probabilities and with large confounding effects from linkage (gwas-pw PPA_3 or COLOC PP4 < 0.6).

We mapped TWAS results to T2D, glycemic, and T1D GWAS loci by identifying TWAS results whose best GWAS lead variant was in LD ($r^2 \geq 0.1$, calculated using the genotypes of individuals with European descent from the Phase 3 of the 1000 Genomes Project [78]) or less than 500kb away from lead GWAS signals, for each trait respectively. The rest of TWAS results were considered as novel GWAS loci. For T2D, we also considered as known GWAS loci if the TWAS signals are in LD ($r^2 \geq 0.1$) with lead independent signals identified in a recent large-scale meta-analysis [3].

Conditional analysis in the PTPN22 T1D locus

We performed conditional analysis on the *DCLRE1B* sQTL using COJO [94], GWAS summary statistics from a recent T1D meta-analysis [5] and the genotypes from 1000 Genomes Phase 3 as the LD reference panel. We first conducted COJO single-variant analysis (*gcta --cojo-cond*) that conditioned on the lead *PTPN22* missense variant (rs2476601) and the *DCLRE1B* sQTL (rs11102694), respectively. Additional joint SNP analysis (*--cojo-joint*) was performed with COJO considering both *PTPN22* missense and the *DCLRE1B* sQTL variant. For the joint SNP analysis, we used a fraction of 381,380 unrelated UK Biobank individuals as the LD reference panel (subjects related up to second-degree were filtered) of white European origin.

Single cell gene expression analysis

We quantified published single cell data sets [95–98] against in-house annotations (Atla, G., Beucher, A., et al., unpublished) using featureCounts [83]. Seurat V3 [99] was used to normalize and identify cell-types within each data set. A gene is annotated as expressed if it has normalized counts of ≥ 0.3 in at least 30 cells of a given cell-type in one of the data sets.

Network analysis

We selected all the nominated effector genes for T2D, FG, FI from the current study and InsPIRE [12], filtered for 106 that encode for proteins with known or presumed function, and used them to construct networks using StringDB [100]. StringApp [101] in cytoscape [102] was used to identify protein-protein interaction networks using default parameters, confidence score cutoff 0.4 and maximum additional interactors of 5. This yielded 1.8-fold higher protein-protein interactions than random sets, although only subnetworks with 3 or more components are shown.

Credible set analysis

We used fine-mapping approaches to identify candidate causal variants that underlie cis-eQTL and sQTL *loci* [21]. We identified 95% credible set variants using CAVIAR [103] software and allowing for one causal variant (*-c 1*). LD information between SNP pairs (i.e., the *r* matrix) was generated using PLINK [88] v1.9 *–matrix –r*, and as reference panel, we used the 399 human islet samples used in the eQTL and sQTL identification.

For each e/sQTL credible set, posterior probabilities were plotted with respect to underlying genomic annotations.

In silico functional scores

The potential impact on disease was assessed for credible set variants of both eQTLs and sQTLs on the basis of their predicted transcriptional and post-transcriptional regulatory effects. We used a deep-learning model that is trained on transcriptional regulatory features such as histone marks, DNAse I profiles, and transcription factors, a total of 2002 features, and is implemented as DeepSEA [49] as well as post-transcriptional regulatory features such as RNA-binding proteins binding data based

Atla *et al. Genome Biology*   (2022) 23:196

Page 23 of 28

on CLIP experiments on 82 unique RBPs (ENCODE and other CLIP datasets), which is implemented as Seqweaver [46]. We performed in silico mutagenesis on both eQTL and sQTL credible set variants using both DeepSEA and Seqweaver models and obtained Disease Impact Scores (DIS) values from each respective model [21]. Prior to in silico mutagenesis, the strand information was added to sQTL credible sets based on the orientation of the target gene. Credible sets were further stratified based on their location in the genome and DIS values from both models for each category of variants were shown as boxplots.

Integrating GWAS credible sets with QTL credible sets

We obtained pre-calculated genetic credible sets for T2D [1] (https://diagram-consortium.org/downloads.html). For each colocalized locus (colocalization PP >0.8), we then annotated credible set variants into two mutually exclusive categories: (i) GWAS credible set variant which is also a QTL credible set variant and (ii) GWAS credible set variant which is not a QTL credible set variant but in LD ($r^2 < 0.1$) with a lead QTL. We then plotted the distributions of CPPs of all GWAS credible sets stratified into above two categories. If a variant belongs to more than one credible set, we used the maximum CPP of that variant.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13059-022-02757-0.

Additional file 1: Supplementary Figures S1-S9 and figure legends.

Additional file 2: Table S1. Significant lead eQTLs.

Additional file 3: Table S2. Significant lead sQTLs.

Additional file 4: Table S3. Examples of sGenes with known functions in islet biology or diabetes.

Additional file 5: Table S4. GREGOR enrichments of functional annotations in e/sQTLs.

Additional file 6: Table S5. GARFIELD enrichments of functional annotations in e/sQTLs.

Additional file 7: Table S6. sQTL colocalizations with gwas-pw (T2D/FG/T1D).

Additional file 8: Table S7. eQTL colocalizations with gwas-pw (T2D/FG/FI/T1D).

Additional file 9: Table S8. sTWAS (T2D, FG, FI, T1D).

Additional file 10: Table S9. eTWAS (T2D, FG, FI, T1D).

Additional file 11: Table S10. T2D loci with target genes nominated in this study and comparison with InsPIRE.

Additional file 12: Table S11. FG/FI loci with target genes nominated in this study and comparison with InsPIRE.

Additional file 13: Table S12. scRNA expression of nominated genes for all traits.

Additional file 14: Review history.

Kingdom. Christian Thirion: Sirion Biotech, Planegg-Martinsried, Germany. David Torrents: Barcelona Supercomputing Center (BSC), Barcelona, Spain. Jorge Ferrer: Centre for Genomic Regulation, The Barcelona Institute of Science and Technology, Barcelona, Spain.

Peer review information

Kevin Pang was the primary editor of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Review history

The review history is available as Additional file 14.

Authors' contributions

G.A., S.B.G., and J.F. conceived, coordinated the study, interpreted the results, and wrote the manuscript with input from remaining authors. A.B. analyzed the transcript annotations. I.M. provided guidance on transcriptome analysis during initial phases. M.C.A provided insights into pathway analysis and M.I. on RNA splicing. A.L.G and L.G. provided the datasets. T.B., F.P, L.M., J-K-C., M.S., P.M, and L.P. procured human islets. J.G-H. processed the RNA and DNA samples. D.J.M.C. and J.A.T. performed the conditional analysis. G.A. and S.B.G. processed the genome data and designed and performed all genetic and statistical analysis. All authors read and approved the final manuscript.

Availability of data and materials

All processed data files for the analyses described here, including all nominal sQTL and eQTL results, and fine-mapped eQTL and sQTL data, are available at Zenodo DOI: 10.5281/zenodo.6546807 [21].
Raw RNA-seq and genotyped SNP array data are available from the European Genome–phenome Archive (EGAS00001006440) [70].

Declarations

Ethics approval and consent to participate

Islet isolation centers had permission to use islets from donors for scientific research if they were insufficient for clinical transplantation following national regulations and ethical requirements and institutional approvals from Leiden University Medical Center, Geneva University Hospitals, University of Lille, and Milano San Raffaele Hospital. Ethical approval for processing de-identified samples was granted by the Clinical Research Ethics Committee of Hospital Clinic de Barcelona (HCB/2014/0926 and HCB/2014/1151) and Clinical Research Ethics Committee from Parc Salut Mar (CBS20_007A). All studies complied with the Helsinki Declaration.

Consent for publication

Not applicable

Competing interests

ALG's spouse is an employee of Genentech and holds stock options in Roche. JAT is a member of the GSK Human Genetics Advisory Board.

Author details

[1]Centre for Genomic Regulation, The Barcelona Institute of Science and Technology, Barcelona, Spain. [2]Centro de Investigación Biomédica en red Diabetes y enfermedades metabólicas asociadas (CIBERDEM), Barcelona, Spain. [3]Department of Metabolism, Digestion and Reproduction, Imperial College London, London, UK. [4]JDRF/Wellcome Diabetes and Inflammation Laboratory, Wellcome Centre for Human Genetics, Nuffield Department of Medicine, NIHR Oxford Biomedical Research Centre, University of Oxford, Oxford, UK. [5]Present Address: Life Sciences Department, Barcelona Supercomputing Center (BSC), 08034 Barcelona, Spain. [6]Lund University Diabetes Centre, Clinical Research Center, Malmö, Sweden. [7]Department of Clinical Sciences in Malmö, Lund University, Malmö, Sweden. [8]Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [9]Department of Pediatrics, Division of Endocrinology, Stanford School of Medicine, Stanford, CA, USA. [10]Department of Clinical and Experimental Medicine, AOUP Cisanello University Hospital, University of Pisa, Pisa, Italy. [11]Cell Isolation and Transplantation Center, University of Geneva, Geneva, Switzerland. [12]Department of Medicine, Leiden University Medical

Atla *et al. Genome Biology*   (2022) 23:196

Page 25 of 28

Center, Leiden, the Netherlands. [13]Hubrecht Institute/KNAW, Utrecht, the Netherlands. [14]University of Lille, Institut National de la Santé et de la Recherche Médicale (INSERM), Centre Hospitalier Universitaire de Lille (CHU Lille), Institute Pasteur Lille, U1190 -European Genomic Institute for Diabetes (EGID), F59000 Lille, France. [15]Diabetes Research Institute, IRCCS Ospedale San Raffaele and Università Vita-Salute San Raffaele, Milan, Italy.

References

1. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet. 2018;50:1505–13.
2. Flannick J, Mercader JM, Fuchsberger C, Udler MS, Mahajan A, Wessel J, et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. Nature. 2019;570:71–6.
3. Vujkovic M, Keaton JM, Lynch JA, Miller DR, Zhou J, Tcheandjieu C, 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;52:680–91.
4. Robertson CC, Inshaw JRJ, Onengut-Gumuscu S, Chen WM, Santa Cruz DF, Yang H, et al. Fine-mapping, trans-ancestral and genomic analyses identify causal variants, cells, genes and drug targets for type 1 diabetes. Nat Genet. 2021;53:962–71.
5. Chiou J, Geusz RJ, Okino ML, Han JY, Miller M, Melton R, et al. Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. Nature. 2021;594:398–402.
6. Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, et al. A map of open chromatin in human pancreatic islets. Nat Genet. 2010;42:255–9.
7. Parker SCJ, Stitzel ML, Taylor DL, Orozco JM, Erdos MR, Akiyama JA. Bueren KLv, Chines PS, Narisu N, Program NCS, et al: Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc Natl Acad Sci. 2013;110:17921–6.
8. Pasquali L, Gaulton KJ, Rodríguez-Seguí SA, Mularoni L, Miguel-Escalada I, Akerman İ, et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat Genet. 2014;46:136–43.
9. Thurner M, van de Bunt M, Torres JM, Mahajan A, Nylander V, Bennett AJ, et al. Integration of human pancreatic islet genomic data refines regulatory mechanisms at type 2 diabetes susceptibility loci. Elife. 2018;7:e31977.
10. Varshney A, Scott LJ, Welch RP, Erdos MR, Chines PS, Narisu N, et al. Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. Proc Natl Acad Sci. 2017;114:201621192–21.
11. Miguel-Escalada I, Bonàs-Guarch S, Cebola I, Ponsa-Cobas J, Mendieta-Esteban J, Atla G, et al. Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. Nat Genet. 2019;51:1137–48.
12. Viñuela A, Varshney A, van de Bunt M, Prasad RB, Asplund O, Bennett A, et al. Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. Nat Commun. 2020;11:4912.
13. Alonso L, Piron A, Morán I, Guindo-Martínez M, Bonàs-Guarch S, Atla G, et al. TIGER: The gene expression regulatory variation landscape of human pancreatic islets. Cell Rep. 2021;37:109807.
14. Inshaw JRJ, Sidore C, Cucca F, Stefana MI, Crouch DJM, McCarthy MI, et al. Analysis of overlapping genetic association in type 1 and type 2 diabetes. Diabetologia. 2021;64:1342–7.
15. Li YI. Geijn Bvd, Raj A, Knowles DA, Petti AA, Golan D, Gilad Y, Pritchard JK. RNA splicing is a primary link between genetic variation and disease. Science. 2016;352:600–4.
16. Scotti MM, Swanson MS. RNA mis-splicing in disease. Nat Rev Genet. 2016;17:19–32.
17. GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020;369:1318–30.
18. Mvd B, Fox JEM, Dai X, Barrett A, Grey C, Li L, 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. PLOS Genet. 2015;11:e1005694.
19. Fadista J, Vikman P, Laakso EO, Mollet IG, Esguerra JL, Taneera J, et al. Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. Proc Natl Acad Sci. 2014;111:13924–9.
20. Li YI, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, Im HK, et al. Annotation-free quantification of RNA splicing using LeafCutter. Nat Genet. 2018;50:151–8.
21. Atla G, Guarch SB, Ferrer J: Genetic regulation of RNA splicing in human pancreatic islets. eQTL and sQTL summary statistics. Zenodo. https://zenodo.org/record/6546807.
22. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 2012;22:1760–74.
23. Chen J, Spracklen CN, Marenne G, Varshney A, Corbin LJ, Luan J, et al. The trans-ancestral genomic architecture of glycemic traits. Nat Genet. 2021;53:840–60.
24. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. Nat Genet. 2016;48:709–17.
25. Gusev A, Ko A, Shi H, Bhatia G, Chung W, Penninx BWJH, et al. Integrative approaches for large-scale transcriptome-wide association studies. Nat Genet. 2016;48:245–52.
26. Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, et al. Opportunities and challenges for transcriptome-wide association studies. Nat Genet. 2019;51:592–9.

Atla *et al. Genome Biology*   (2022) 23:196

Page 26 of 28

27. Ndungu A, Payne A, Torres JM. Bunt Mvd, McCarthy MI. A multi-tissue transcriptome analysis of human metabolites guides interpretability of associations based on multi-SNP models for gene expression. Am J Human Genet. 2020;106(2):188–201.

28. Paciotti S, Persichetti E, Klein K, Tasegian A, Duvet S, Hartmann D, et al. Accumulation of free oligosaccharides and tissue damage in cytosolic α-mannosidase (Man2c1)-deficient mice. J Biol Chem. 2014;289:9611–22.

29. Wang L, Suzuki T. Dual functions for cytosolic α-mannosidase (Man2C1): its down-regulation causes mitochondria-dependent apoptosis independently of its α-mannosidase activity. J Biol Chem. 2013;288:11887–96.

30. Thomas P, Wohlford D, Aoh QL. SCAMP 3 is a novel regulator of endosomal morphology and composition. Biochem Biophys Res Commun. 2016;478:1028–34.

31. Simaite D, Kofent J, Gong M, Ruschendorf F, Jia S, Arn P, et al. Recessive mutations in PCBD1 cause a new type of early-onset diabetes. Diabetes. 2014;63:3557–64.

32. Ferrè S, de Baaij JH, Ferreira P, Germann R, de Klerk JB, Lavrijsen M, et al. Mutations in PCBD1 cause hypomagnesemia and renal magnesium wasting. J Am Soc Nephrol. 2014;25:574–86.

33. Mendel DB, Khavari PA, Conley PB, Graves MK, Hansen LP, Admon A, et al. Characterization of a cofactor that regulates dimerization of a mammalian homeodomain protein. Science. 1991;254:1762–7.

34. Dai FF, Zhang Y, Kang Y, Wang Q, Gaisano HY, Braunewell KH, et al. The neuronal Ca2+ sensor protein visinin-like protein-1 is expressed in pancreatic islets and regulates insulin secretion. J Biol Chem. 2006;281:21942–53.

35. Ramos-Rodríguez M, Raurell-Vila H, Colli ML, Alvelos MI, Subirana-Granés M, Juan-Mateu J, et al. The impact of pro-inflammatory cytokines on the β-cell regulatory landscape provides insights into the genetics of type 1 diabetes. Nat Genet. 2019;51:1588–95.

36. DcL E, Sammeth M, Bouckenooghe T, Bottu G, Sisino G, Igoillo-Esteve M, et al. The human pancreatic islet transcriptome: expression of candidate genes for type 1 diabetes and the impact of pro-inflammatory cytokines. PLoS Genet. 2012;8:e1002552.

37. Xie B, Nguyen PM, Idevall-Hagren O. Feedback regulation of insulin secretion by extended synaptotagmin-1. FASEB J. 2019;33:4716–28.

38. Iyer S, Modali SD, Agarwal SK. Long noncoding RNA MEG3 is an epigenetic determinant of oncogenic signaling in functional pancreatic neuroendocrine tumor cells. Mol Cell Biol. 2017;37(22):e00278–17.

39. Li R, Fang L, Pu Q, Bu H, Zhu P, Chen Z, et al. MEG3-4 is a miRNA decoy that regulates IL-1β abundance to initiate and then limit inflammation to prevent sepsis during lung infection. Sci Signal. 2018;11(536):eaao2387.

40. You L, Wang N, Yin D, Wang L, Jin F, Zhu Y, et al. Downregulation of long noncoding RNA Meg3 affects insulin synthesis and secretion in mouse pancreatic beta cells. J Cell Physiol. 2016;231:852–62.

41. Kameswaran V, Golson ML, Ramos-Rodríguez M, Ou K, Wang YJ, Zhang J, 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;67:1807–15.

42. Thompson PJ, Shah A, Ntranos V, Van Gool F, Atkinson M, Bhushan A. Targeted elimination of senescent beta cells prevents type 1 diabetes. Cell Metab. 2019;29:1045–1060.e1010.

43. Thompson PJ, Shah A, Apostolopolou H, Bhushan A. BET proteins are required for transcriptional activation of the senescent islet cell secretome in type 1 diabetes. Int J Mol Sci. 2019;20(19):4776.

44. Touzot F, Callebaut I, Soulier J, Gaillard L, Azerrad C, Durandy A, et al. Function of Apollo (SNM1B) at telomere highlighted by a splice variant identified in a patient with Hoyeraal-Hreidarsson syndrome. Proc Natl Acad Sci U S A. 2010;107:10097–102.

45. Kundu K, Tardaguila M, Mann AL, Watt S, Ponstingl H, Vasquez L, et al. Genetic associations at regulatory phenotypes improve fine-mapping of causal variants for 12 immune-mediated diseases. Nat Genet. 2022;54:251–62.

46. Zhou J, Park CY, Theesfeld CL, Wong AK, Yuan Y, Scheckel C, et al. Whole-genome deep-learning analysis identifies contribution of noncoding mutations to autism risk. Nat Genet. 2019;51:973–80.

47. Wahl MC, Will CL, Lührmann R. The spliceosome: design principles of a dynamic RNP machine. Cell. 2009;136:701–18.

48. Fu X-D, Ares M. Context-dependent control of alternative splicing by RNA-binding proteins. Nature Reviews Genetics. 2014;15(10):689–701.

49. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. Nat Methods. 2015;12:931–4.

50. Minami K. GATA transcription factors: New key regulators in pancreas organogenesis. J Diabetes Invest. 2013;4:426–7.

51. Shaw-Smith C, Franco ED, Allen HL, Batlle M, Flanagan SE, Borowiec M, et al. GATA4 mutations are a cause of neonatal and childhood-onset diabetes. Diabetes. 2014;63:2888–94.

52. Viger RS, SVM G, Anttonen M, Wilson DB, Heikinheimo M. Role of the GATA family of transcription factors in endocrine development, function, and disease. Mole Endocrinol (Baltimore, Md). 2008;22:781–98.

53. Zito E, Chin K-T, Blais J, Harding HP, Ron D. ERO1-β, a pancreas-specific disulfide oxidase, promotes insulin biogenesis and glucose homeostasisIslet dysfunction in Ero1b mutant mice. J Cell Biol. 2010;188:821–32.

54. Awazawa M, Futami T, Sakada M, Kaneko K, Ohsugi M, Nakaya K, et al. Deregulation of pancreas-specific oxidoreductin ERO1β in the pathogenesis of diabetes mellitus. Mole Cell Biol. 2014;34:1290–4.

55. Khoo C, Yang J, Rajpal G, Wang Y, Liu J, Arvan P, et al. Endoplasmic reticulum oxidoreductin-1-like β (ERO1Iβ) regulates susceptibility to endoplasmic reticulum stress and is induced by insulin flux in β-cells. Endocrinology. 2011;152:2599–608.

56. Common Metabolic Diseases Knowledge Portal (cmdkp.org). rs2477599 Variant page. 2022 Aug 6; [https://hugeamp.org/variant.html?variant=rs2477599]

57. Jermusyk A, Zhong J, Connelly KE, Gordon N, Perera S, Abdolalizadeh E, et al. A 584 bp deletion in CTRB2 inhibits chymotrypsin B2 activity and secretion and confers risk of pancreatic cancer. Am J Hum Genet. 2021;108:1852–65.

58. Mahajan A, Spracklen CN, Zhang W, Ng MCY, Petty LE, Kitajima H, et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet*. 2022;54:560–72.

Atla *et al. Genome Biology*   (2022) 23:196

Page 27 of 28

59. von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, Foglierini M, et al. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. Nucleic Acids Res. 2005;33:D433–7.

60. De Franco E, Caswell R, Johnson MB, Wakeling MN, Zung A, Dũng VC, et al. De novo mutations in EIF2B1 affecting eIF2 signaling cause neonatal/early onset diabetes and transient hepatic dysfunction. Diabetes. 2020;69:477–83.

61. Delépine M, Nicolino M, Barrett T, Golamaully M, Lathrop GM, Julier C. EIF2AK3, encoding translation initiation factor 2-alpha kinase 3, is mutated in patients with Wolcott-Rallison syndrome. Nat Genet. 2000;25:406–9.

62. Rorsman P, Ashcroft FM. Pancreatic β-cell electrical activity and insulin secretion: of mice and men. Physiol Rev. 2018;98:117–214.

63. Tengholm A, Gylfe E. cAMP signalling in insulin and glucagon secretion. Diabetes Obes Metab. 2017;19(Suppl 1):42–53.

64. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. Nat Genet. 2015;47:856–60.

65. Persaud SJ. Islet G-protein coupled receptors: therapeutic potential for diabetes. Curr Opin Pharmacol. 2017;37:24–8.

66. Ghislain J, Poitout V. Targeting lipid GPCRs to treat type 2 diabetes mellitus - progress and challenges. Nat Rev Endocrinol. 2021;17:162–75.

67. Nauck MA, Wefers J, Meier JJ. Treatment of type 2 diabetes: challenges, hopes, and anticipated successes. Lancet Diabetes Endocrinol. 2021;9:525–44.

68. Castel SE, Cervera A, Mohammadi P, Aguet F, Reverter F, Wolman A, et al. Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. Nat Genet. 2018;50:1327–34.

69. Kim-Hellmuth S, Aguet F, Oliva M, Muñoz-Aguirre M, Kasela S, Wucher V, et al. Cell type-specific genetic regulation of gene expression across human tissues. Science. 2020;369(6509):eaaz8528.

70. Atla G, Guarch SB, Ferrer J. Genetic regulation of RNA splicing in human pancreatic islets. RNA-Seq and Genotype data sets. The European Genome-phenome Archive (EGA). https://ega-archive.org/studies/EGAS00001006440.

71. Nano R, Bosco D, Kerr-Conte JA, Karlsson M, Charvier S, Melzi R, et al. Human islet distribution programme for basic research: activity over the last 5 years. Diabetologia. 2015;58:1138–40.

72. Kerr-Conte J, Vandewalle B, Moerman E, Lukowiak B, Gmyr V, Arnalsteen L, et al. Upgrading pretransplant human islet culture technology requires human serum combined with media renewal. Transplantation. 2010;89:1154–60.

73. Melzi R, Mercalli A, Sordi V, Cantarelli E, Nano R, Maffi P, et al. Role of CCL2/MCP-1 in islet transplantation. Cell Transplant. 2010;19:1031–46.

74. Bucher P, Mathe Z, Morel P, Bosco D, Andres A, Kurfuest M, et al. Assessment of a novel two-component enzyme preparation for human islet isolation and transplantation. Transplantation. 2005;79:91–7.

75. Loh PR, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, et al. Reference-based phasing using the Haplotype Reference Consortium panel. Nat Genet. 2016;48:1443–8.

76. Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. Bioinformatics. 2015;31:782–4.

77. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48:1279–83.

78. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015;526:68–74.

79. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.

80. Bvd G, McVicker G, Gilad Y, Pritchard JK. WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat Methods. 2015;12:1061–3.

81. Jun G, Flickinger M, Hetrick Kurt N, Romm Jane M, Doheny Kimberly F, Abecasis Galo R, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. Am J Human Genet. 2012;91:839–48.

82. Abraham G, Qiu Y, Inouye M. FlashPCA2: principal component analysis of Biobank-scale genotype datasets. Bioinformatics. 2017;33:2776–8.

83. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.

84. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.

85. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8:118–27.

86. Delaneau O, Ongen H, Brown AA, Fort A, Panousis NI, Dermitzakis ET. A complete tool set for molecular QTL discovery and analysis. Nat Commun. 2017;8:15452.

87. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24–6.

88. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. Bakker PIWd, Daly MJ, Sham PC: PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Human Genet. 2007;81:559–75.

89. Schmidt EM, Zhang J, Zhou W, Chen J, Mohlke KL, Chen YE, et al. GREGOR: evaluating global enrichment of trait-associated variants in epigenomic features using a systematic, data-driven approach. Bioinformatics. 2015;31:2601–6.

90. Iotchkova V, GRS R, Geihs M, Morganella S, Min JL, Walter K, et al. GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. Nat Genet. 2019;51:343–53.

91. Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, et al. The UCSC Genome Browser Database: update 2006. Nucleic Acids Res. 2006;34:D590–8.

92. Berisa T, Pickrell JK. Approximately independent linkage disequilibrium blocks in human populations. Bioinformatics. 2016;32:283–5.

Atla *et al. Genome Biology*   (2022) 23:196

Page 28 of 28

93. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics. 2010;26:2336–7.

94. Yang J, Ferreira T, Morris AP, Medland SE, Madden PA, Heath AC, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet. 2012;44(369-375):S361–3.

95. Lawlor N, George J, Bolisetty M, Kursawe R, Sun L, Sivakamasundari V, et al. Single-cell transcriptomes identify human islet cell signatures and reveal cell-type specific expression changes in type 2 diabetes. Genome Res. 2017;27(208):222.

96. Ãs S, Palasantza A, Eliasson P, Andersson E-M, Andreasson A-C, Sun X, et al. Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. Cell Metab. 2016;24(593):607.

97. Enge M, Arda HE, Mignardi M, Beausang J, Bottino R, Kim SK, et al. Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. Cell. 2017;171(321):330.e314.

98. Xin Y, Kim J, Okamoto H, Ni M, Wei Y, Adler C, et al. RNA sequencing of single human islet cells reveals type 2 diabetes genes. Cell Metab. 2016;24(608):615.

99. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive integration of single-cell data. Cell. 2019;177:1888–1902.e1821.

100. Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, et al. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 2013;41:D808–15.

101. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: network analysis and visualization of proteomics data. J Proteome Res. 2019;18:623–32.

102. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13:2498–504.

103. Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E. Identifying causal variants at loci with multiple signals of association. Genetics. 2014;198:genetics.114.167908.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.