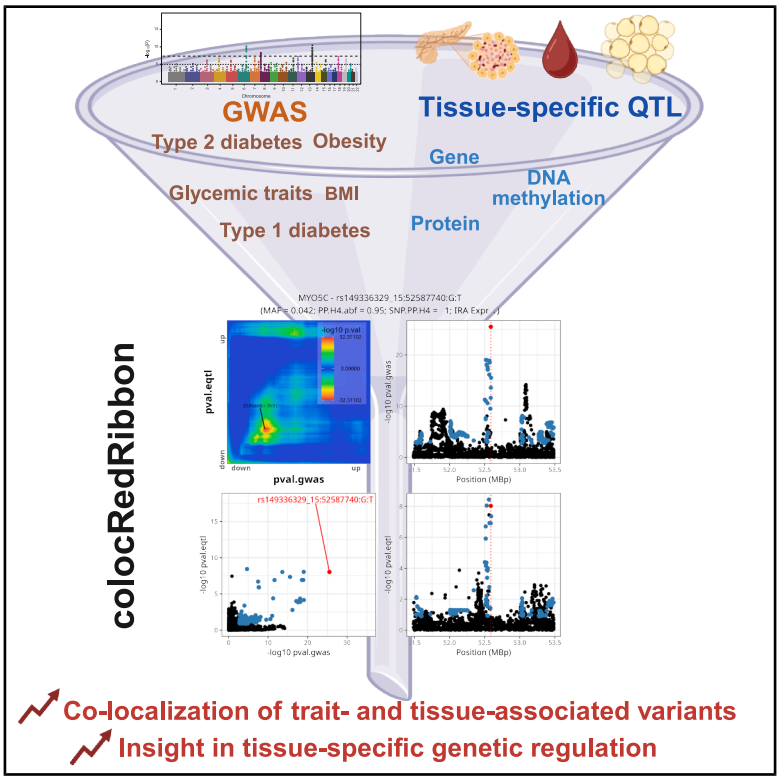
Cell Genomics

Article

Identification of novel type 1 and type 2 diabetes genes by co-localization of human islet eQTL and GWAS variants with *colocRedRibbon*

Graphical abstract



Highlights

- *colocRedRibbon* is a new tool to match genetic variants with gene expression changes
- The tool uncovers 292 islet gene regions tied to diabetes and glycemic traits
- It can co-localize genetic variants with transcripts, proteins, and DNA methylation
- These co-localizations provide insight into how GWAS variants confer disease risk

Authors

Anthony Piron, Florian Szymczak, Lise Folon, ..., Décio L. Eizirik, Josep M. Mercader, Miriam Cnop

Correspondence

anthony.piron@ulb.be (A.P.), miriam.cnop@ulb.be (M.C.)

In brief

Piron et al. present the tool *colocRedRibbon*, which identifies hundreds of gene regions where genetic variation influences diabetes risk and glycemic traits through effects on pancreatic islet gene expression. This approach enables the discovery of tissue-specific mechanisms and advances understanding of the genetic architecture of diabetes.

Piron et al., 2025, Cell Genomics *5*, 101004
November 12, 2025
https://doi.org/10.1016/j.xgen.2025.101004

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

Identification of novel type 1 and type 2 diabetes genes by co-localization of human islet eQTL and GWAS variants with *colocRedRibbon*

Anthony Piron,[1,2,3,4,16,*] Florian Szymczak,[1,3] Lise Folon,[1] Daniel J.M. Crouch,[4] Theodora Papadopoulou,[1] Maria Lytrivi,[1] Yue Tong,[1] Maria Inês Alvelos,[1] Maikel L. Colli,[1] Xiaoyan Yi,[1] Marcin L. Pekalski,[4,5] Konstantinos Hatzikotoulas,[6] Alicia Huerta-Chagoya,[7,8,9] Henry J. Taylor,[11,12,13] Type 2 Diabetes Global Genomics Initiative, Matthieu Defrance,[3] John A. Todd,[4] Décio L. Eizirik,[1] Josep M. Mercader,[7,8,9,10] and Miriam Cnop[1,14,15,*]

[1]ULB Center for Diabetes Research, Medical Faculty, Université Libre de Bruxelles, 1070 Brussels, Belgium
[2]Pharmacognosy, Bioanalysis and Drug Discovery, Pharmacy Faculty, Université Libre de Bruxelles, 1050 Brussels, Belgium
[3]Interuniversity Institute of Bioinformatics in Brussels (IB2), 1050 Brussels, Belgium
[4]JDRF/Wellcome Diabetes and Inflammation Laboratory, Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7BN, UK
[5]Military Institute of Medicine, National Research Institute Warsaw, Warsaw, Poland
[6]Institute of Translational Genomics, Helmholtz Zentrum München, German Research Center for Environmental Health, 85764 Neuherberg, Germany
[7]Diabetes Unit, Massachusetts General Hospital, Boston, MA 02144, USA
[8]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA 02114, USA
[9]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA
[10]Programs in Metabolism and Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[11]Center for Precision Health Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892-2152, USA
[12]British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB2 0BB, UK
[13]Heart and Lung Research Institute, University of Cambridge, Cambridge CB2 0BB, UK
[14]Division of Endocrinology, Erasmus Hospital, Université Libre de Bruxelles, 1070 Brussels, Belgium
[15]WEL Research Institute, 1300 Wavre, Belgium
[16]Lead contact
*Correspondence: anthony.piron@ulb.be (A.P.), miriam.cnop@ulb.be (M.C.)
https://doi.org/10.1016/j.xgen.2025.101004

SUMMARY

Over 1,000 genetic variants have been associated with diabetes by genome-wide association studies (GWASs), but for most, their functional impact is unknown; only 7% alter gene expression in pancreatic islets in expression quantitative trait locus (eQTL) studies. To fill this gap, we developed a co-localization pipeline, *colocRedRibbon*, that prefilters eQTLs by the direction of effect on gene expression and shortlists overlapping eQTL and GWAS variants prior to co-localization. Applying *colocRedRibbon* to recent diabetes and glycemic trait GWASs, we identified 292 co-localizing gene regions, including 24 co-localizations for type 1 diabetes and 268 for type 2 diabetes and glycemic traits, representing a 4-fold increase. A low-frequency type 2 diabetes protective variant increases islet *MYO5C* expression, and a type 1 diabetes protective variant increases *FUT2* expression. These novel co-localizations advance the understanding of diabetes genetics and its impact on human islet biology. *colocRedRibbon* has broad applicability to co-localize GWASs and various QTLs.

INTRODUCTION

Diabetes is a complex, multi-factorial disease characterized by elevated blood glucose levels. More than 589 million people are affected worldwide, and diabetes prevalence is forecasted to reach 1.31 billion by 2050.[1,2] Common forms are type 1 (10%–15%) and type 2 (>85%) diabetes. Environmental factors such as sedentary lifestyle, obesity, aging, viral infections, and gut microbiome play important roles in the development of diabetes and underlie the increasing prevalence.[3–5] Familial clustering and heritability studies have long provided evidence for the role of genetic factors, and type 1 and type 2 diabetes are now established as polygenic. Progressive pancreatic islet dysfunction, caused by environmental and genetic factors, is central in most forms of diabetes.[6–8] Therefore, the study of human islet pathophysiology is essential to better understand the disease.

Large genome-wide association studies (GWASs) and fine-mapping analyses have associated >700 genetic loci with the

Please cite this article in press as: Piron et al., Identification of novel type 1 and type 2 diabetes genes by co-localization of human islet eQTL and GWAS variants with *colocRedRibbon*, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101004

risk for type 1 and type 2 diabetes and glycemic traits,[9–15] but the mechanism of action for most remains poorly understood. 89% of fine-mapped type 2 diabetes GWAS variants lie in non-coding genomic regions[9]; these are typically named after the closest gene. *Cis*-expression quantitative trait locus (*cis*-eQTL) analyses link variants with gene expression. Such approaches have identified gene expression changes associated with variants lying farther away or in intronic regions of other genes in human tissues[16] and islets,[17,18] showing that the closest gene is not necessarily the effector. eQTLs further inform about the direction of effect, i.e., whether a variant is associated with lower or higher gene expression. Linking diabetes-associated variants to gene expression is therefore an essential step to understanding diabetes genetics and pathophysiology.

One of the methods to link disease variants with gene expression is to co-localize GWAS and *cis*-eQTL variants. The Genotype-Tissue Expression (GTEx) project generated eQTLs for 54 human tissues,[19] including pancreas. However, pancreatic islets—a key tissue in diabetes pathogenesis—are not part of the GTEx collection. Separate initiatives have identified human islet eQTLs and co-localized them with diabetes GWAS hits, advancing the domain but falling short of expectations: only around 50 are associated with gene expression variation.[17,18] We hypothesized that this missing link is, at least in part, caused by methodological limitations of the available co-localization approaches. Indeed, we observed that existing co-localization tools miss matching GWAS and eQTL signals when analyzing chromosomal regions that contain multiple strong GWAS signals. These tools also do not consider the direction of effect of variants in a region tested for co-localization.

To test this hypothesis and fill the co-localization gap, we developed a new co-localization pipeline, *colocRedRibbon*. The method uses novel prefiltering steps, shortlisting variants before running an existing co-localization tool. We use *colocRedRibbon* to co-localize human islet *cis*-eQTLs[17] and GWASs for type 1 diabetes,[20] type 2 diabetes,[15] and glycemic traits.[11] We significantly expand the number of co-localizations to 268 for type 2 diabetes and to 24 for type 1 diabetes, making this study the largest to date and improving knowledge about genetic regulation in human islets.

colocRedRibbon is a versatile tool designed to co-localize a wide range of datasets, extending beyond diabetes and islets. It enables the integration of GWAS data for any trait with various QTL types or with other GWAS datasets. To highlight its broad applicability, we applied *colocRedRibbon* to protein QTLs (pQTLs),[21] methylation QTLs (meQTLs),[22] obesity GWASs,[23] and BMI GWASs.[24]

RESULTS

Previous studies have linked diabetes GWAS signals to eQTL variants in human islets, a highly disease-relevant tissue that maintains glycemia. The yield of the 2 largest studies to date[17,18] has been limited, identifying co-localization with eQTL variants for only 7% of GWAS hits. To investigate the hypothesis that methodological limitations in co-localization approaches contributed to the "co-localization missing link," we developed a novel tool called *colocRedRibbon.* The pipeline employs a two-step approach for shortlisting variants and then runs the *coloc* co-localization package.[25] The novel shortlisting method first categorizes risk variants based on the direction of effect on gene expression and then uses *RedRibbon* rank-rank hypergeometric overlap.[26] In this second step, variants are ranked by *p* values for both GWASs or QTLs, and potential overlap between the ranked lists is statistically examined (see the STAR Methods section). In the last step, *coloc*[25] is used to assess the posterior probability (PP) that both dataset signals share the same causal variant, known as hypothesis H4; a PP.H4.abc > 0.8 is considered significant. Additionally, SNP.PP.H4 represents the PP that the lead SNP is the shared causal variant in both datasets. We applied *colocRedRibbon* to recent large-scale eQTL,[17] pQTL,[21] meQTL,[22] and GWAS[11,15,20,23,24] datasets (Figure 1). The eQTL dataset is from the translational human pancreatic islet genotype tissue-expression resource (TIGER), which is, with its meta-analysis of eQTLs in 4 cohorts comprising 404 islet samples, one of the largest eQTL analyses of human islets.[17] The European and multi-ancestry type 2 diabetes GWASs were from the Type 2 Diabetes Global Genomic Initiative (T2DGGI).[15] The Wellcome Center for Human Genetics provided the GWAS for type 1 diabetes.[20] Quantitative glycemic trait GWASs originate from MAGIC.[11] These GWASs are among the biggest to date, having included >2.5 million cases/controls for type 2 diabetes, >170,000 for type 1 diabetes, and up to 200,000 for glycemic traits (Table S1). The pQTL from deCODE includes data from 35,000 individuals on 4,907 aptamers.[21] The adipocyte meQTL from the Twins UK cohort comprises 538 participants.[22] The obesity GWAS from FinnGen analyzed 500,000 individuals, including 31,000 cases,[23] and the BMI GWAS from GIANT included 700,000 participants.[24]

The co-localization of variants identifies important diabetes genes

With this new *colocRedRibbon* pipeline, we identified a total of 434 co-localizations between islet eQTL and diabetes and glycemic trait GWAS variants. These co-localizations are shown by their chromosomal position in a circle plot, highlighting the regulated genes for GWAS SNPs for type 1 diabetes (n = 24), type 2 diabetes (multi-ancestry GWAS n = 216; European ancestry n = 143), and the glycemic traits fasting glucose (n = 22), post-challenge 2-h glucose (n = 4), HbA1c (n = 20), and fasting insulin (n = 2) (Figure 2; Tables S2 and S3).

We examined a total of 10,835,771 GWAS variants, including non-significant ones, within a 2 million base-pair region around the transcription start site of genes exhibiting at least one significant GWAS variant and one eQTL. The distribution of the 10,835,771 variants across traits revealed that type 2 diabetes accounts for the majority (73%), with multi-ancestry GWAS contributing 41% and European ancestry 32%. Type 1 diabetes represented 18% of the variants, and glycemic traits were represented to a lesser extent: HbA1c (5%), 2-h glucose (0.5%), fasting glucose (3%), and fasting insulin (0.5%) (Table S4).

Type 2 diabetes demonstrated the highest number of co-localizations (Tables S2, S3, and S5). The lowest number was found for fasting insulin, which reflects insulin resistance rather than β cell function. This aligns with the

Please cite this article in press as: Piron et al., Identification of novel type 1 and type 2 diabetes genes by co-localization of human islet eQTL and GWAS variants with *colocRedRibbon*, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101004

CellPress
OPEN ACCESS

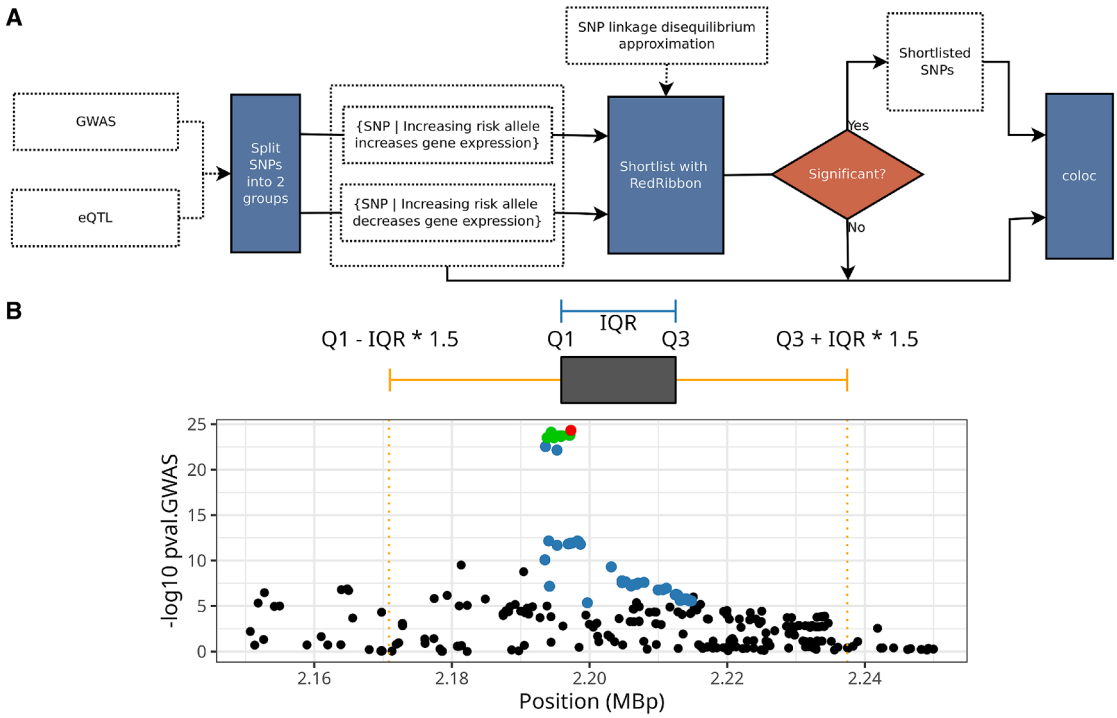


Figure 1. The *colocRedRibbon* pipeline

(A) The workflow of *colocRedRibbon*. GWAS and eQTL SNPs are split into two effect sets depending on the direction of effect on gene expression for the increasing risk allele. The *RedRibbon* rank-rank hypergeometric overlap method is applied on both SNP sets ranked by *p* values. If significant overlap is detected by *RedRibbon*, shortlisted SNPs are analyzed by *coloc*. Otherwise, *coloc* is applied to the two effect sets without overlap shortlisting.

(B) The interquartile range mode of *colocRedRibbon* adds a background of variants between dotted orange lines, where Q_1 is the 25th percentile (first quartile) of the chromosomal positions of the core set, Q_2 the 75th percentile (third quartile), and IQR the interquartile range.

expectation that insulin resistance would primarily manifest in skeletal muscle, liver, or adipose tissue rather than in pancreatic islets.

colocRedRibbon shortlisting steps solve the issue of multiple GWAS signals

Cis-eQTL signals for a gene typically manifest themselves as a single peak of variants in linkage disequilibrium, whereas GWAS signals often comprise multiple independent peaks. This multiplicity of signals can confound *coloc*, potentially leaving legitimate co-localizations undetected. A prime example is observed in a 2 million base-pair region surrounding the *RPL39L* gene. *coloc* identifies a lead variant in a strong type 2 diabetes GWAS signal (red dot at the top, Figure 3A) but overlooks an adjacent, highly significant—albeit weaker—signal (Figure 3A) that holds potential for a legitimate co-localization given the presence of a matching eQTL at the same chromosomal position (Figure 3A, bottom). *coloc* assumes under the H4 hypothesis that each trait has a single causal variant that is shared between traits. For *coloc* to be reliable, the analysis must be restricted to genomic regions where the single causal variant assumption holds. In the case of *RPL39L*, this could be achieved by narrowing the region and excluding the stronger GWAS signal.

To address this issue, we shortlisted SNPs using *RedRibbon* overlap. This approach involves ranking GWAS and eQTL variant lists from lowest to highest *p* value and searching for overlapping SNPs using a rank-rank hypergeometric overlap approach. The resulting overlap map reveals a strong and highly significant overlap among low *p* values (Figure 3B, top left). By selecting the variants with the most significant overlap (Figure 3B, blue dots at bottom left), we identified matching GWAS and eQTL peaks and the lead SNP (Figure 3B, red dot on the right). This shortlisting step eliminates spurious GWAS signals unrelated to eQTL peaks, enabling *coloc* to detect co-localization. The lead SNP rs3887925 is the only variant in the 99% credible set. The T effector allele that is associated with type 2 diabetes risk upregulates *RPL39L* transcript expression in human islets (Figure 3C). The SNP lies in regulatory genomic regions, and *RPL39L* is differentially expressed in islets from type 2 diabetic vs. non-diabetic donors, pointing to biological relevance (see below).

The *colocRedRibbon* method further refines the analysis through a direction-of-effect step, which splits variants into two disjoint subsets based on whether their risk alleles up- or downregulate a gene. This premise reduces the number of variants considered by *coloc*. For instance, in a region around the gene *TH* with multiple type 2 diabetes GWAS signals (Figure 3D, top), *coloc* fails to detect co-localization due to the confounding effect of the strongest GWAS signal. By only retaining variants for which the risk allele downregulates *TH* expression and excluding many highly significant GWAS

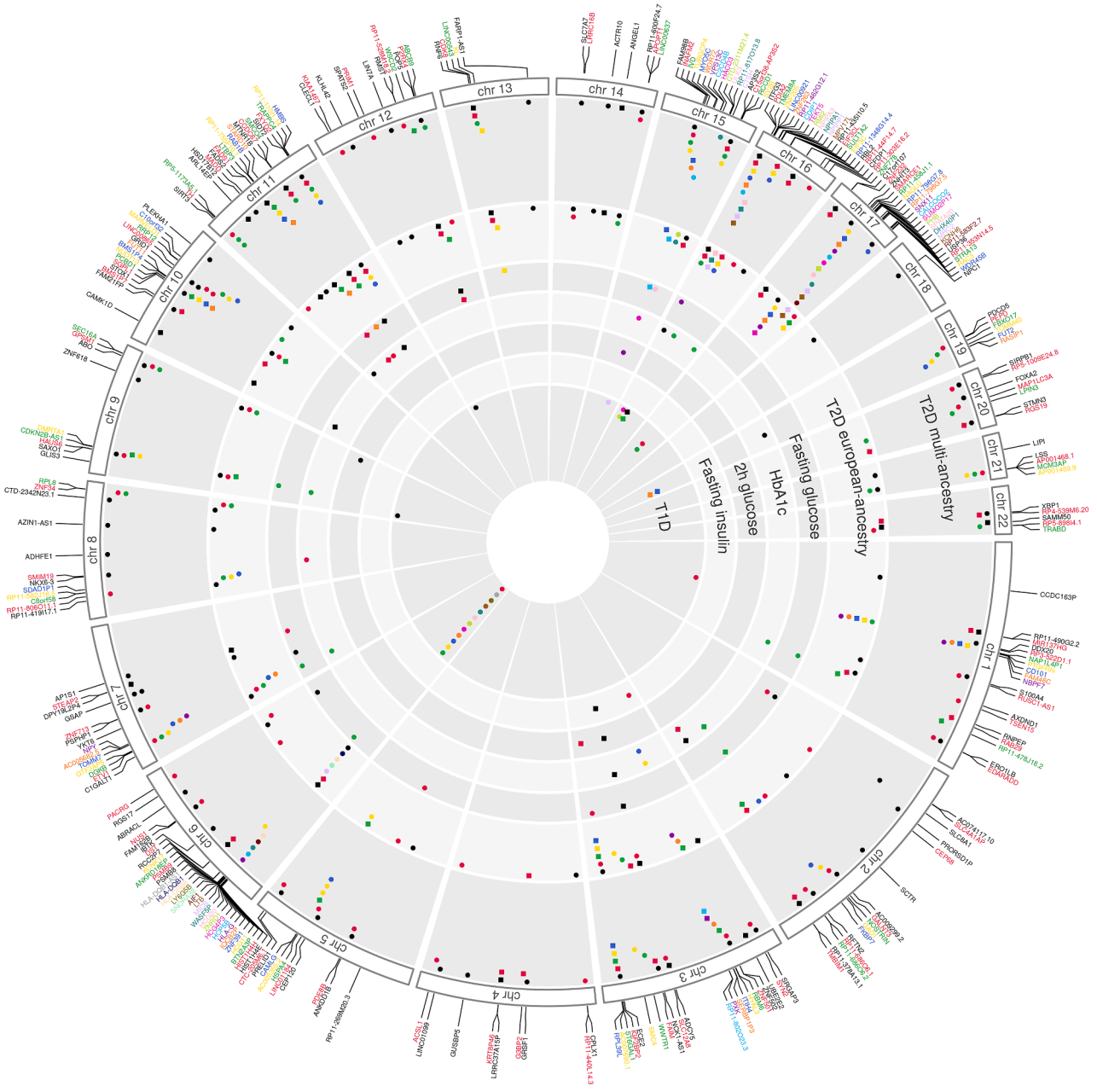


Figure 2. Circle plot of 434 *colocRedRibbon* co-localizations of islet eQTLs with diabetes and glycemic trait GWAS

eQTL genes are positioned at their starting position on the chromosome. Concentric circles correspond to GWASs, each dot to an SNP, and color matches eQTL gene color. Square dots indicate identical SNPs. T1D, type 1 diabetes; T2D, type 2 diabetes. See also Figures S1–S5 and Table S2.

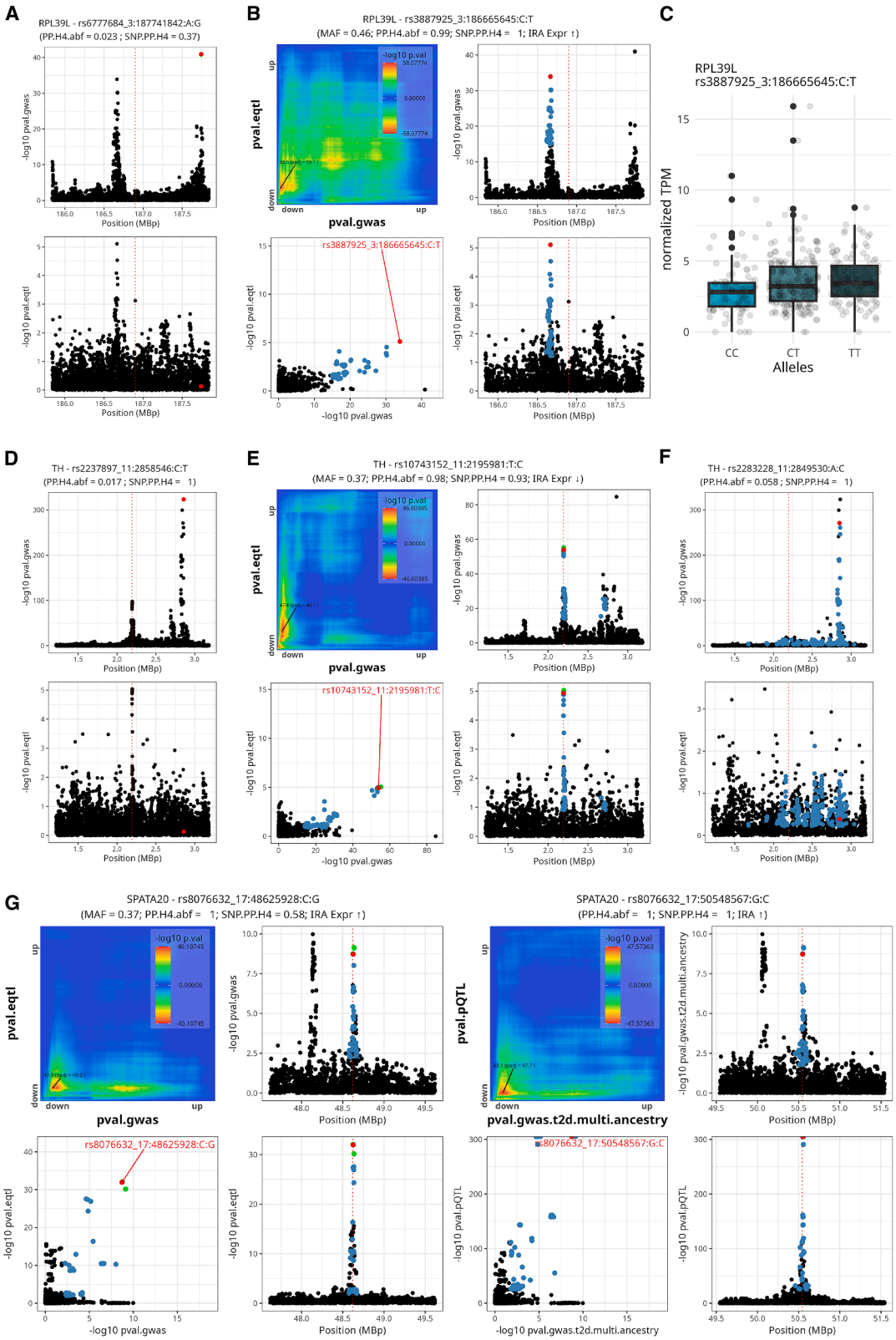
variants, *coloc* detects another significant signal (Figure 3E). Conversely, variants for which the risk allele upregulates *TH* expression do not exhibit co-localization (Figure 3F), aligning with the biologically plausible assumption that risk alleles only associate with one direction of effect.

Of note, five genes co-localize for risk allele expression in both downregulation and upregulation. Among these is gamma-secretase activating protein (*GSAP*). *colocRedRibbon* detects two distinct signals going in opposite gene expression directions, while both increase diabetes risk (Figure S1; Table S6). This gamma-secretase catalyzes intermembrane cleavage of membrane proteins. GSAP plays a role in lipid homeostasis and mitochondrial function.[27]

This refined approach demonstrates the power of *colocRedRibbon* in uncovering previously undetected co-localizations and providing more nuanced insights into the relationship between genetic variants, gene expression, and disease risk.

Please cite this article in press as: Piron et al., Identification of novel type 1 and type 2 diabetes genes by co-localization of human islet eQTL and GWAS variants with *colocRedRibbon*, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101004

CellPress
OPEN ACCESS



(legend on next page)

Please cite this article in press as: Piron et al., Identification of novel type 1 and type 2 diabetes genes by co-localization of human islet eQTL and GWAS variants with *colocRedRibbon*, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101004

colocRedRibbon doubles the number of co-localizations

The 434 co-localizations between human islet *cis*-eQTLs and GWASs for type 1 and type 2 diabetes and glycemic traits encompass 344 distinct lead variants, mapping to 289 unique genes (Figure 4; Table S3). Only 214 were detected by running *coloc* without shortlisting steps (Figure 4A); in other words, *colocRedRibbon* doubled the detection. Improvements were observed for most GWASs, and the largest was for type 1 diabetes. A total of 307 co-localizations were identified across 236 distinct regions for which no co-localization was previously reported.

Out of the 434 co-localizations, 48 had a unique lead variant in the 99% credible set (SNP.PP.H4 $\geq$ 0.99), and 37% had $\leq$5 candidate variants in the 99% credible set. This restricted number narrows the list of potential candidates for experimental validation. For instance, a unique co-localizing lead variant, rs10830963, was detected in *MTNR1B* (Table S2). This SNP has been extensively studied and is associated with increased type 2 diabetes risk.[30,31]

For type 2 diabetes co-localizations, risk alleles are significantly more associated with increased gene expression (Figure 4B). Half of the glycemic trait genes are also associated with type 2 diabetes (Figure 4C). Type 1 and type 2 diabetes share only 3 genes (FUT2, SULT1A2, and WASF5P; Figure 4D) with differing alleles and effects on gene expression (Table S2).

For type 2 diabetes, 35 genes are specific to European ancestry (Figure 4E; Table S7), and 106 are exclusively detected using multi-ancestry GWAS. This is not unexpected, as the multi-ancestry GWAS encompasses a larger number of samples, including those of European ancestry. For 10 of the 106 co-localizations, the lead variant GWAS *p* value is significant for non-European ancestry and non-significant for European ancestry (Table S8), suggesting potential ancestry specificity. For example, we found a novel co-localization for a variant that increases *PRELID1* expression in islets and is associated with reduced type 2 diabetes risk in East Asians. PRELID1 is a lipid-shuttling protein needed for mitochondrial structure and function. The minor-allele frequency for 50 of the 106 multi-ancestry co-localizations is at least 1.5 times higher in one or more non-European GWAS cohorts compared with the European cohort. This is illustrated by the novel co-localization for *GRSF1,* another protein needed for mitochondrial function, with an East Asian type 2 diabetes protective variant increasing islet *GRSF1* expression. This variant is more common in South Asian, Hispanic, East Asian, and African American ancestries (with respective minor-allele frequencies of 11%, 14%, 21%, and 49%) as compared with Europeans (minor-allele frequency of 4.5%). These findings underscore the importance of genetic diversity in enhancing the power to discover disease mechanisms.

Co-localizing variants largely lie in regulatory regions

Co-localizing lead variants showed a notable over-representation in regulatory regions of the genome (Figure 4F; Table S9), with 383 lead variants situated in such areas. Significant enrichment was observed in exonic (n = 71), promoter (n = 56), promoter-flanking (n = 65), and transcript (n = 310) regions. In this tissue-agnostic analysis, no enrichment was detected in open chromatin, enhancer, or CTCF-binding sites from Ensembl GRCh37 regulatory features.[32]

Given the tissue- or cell-type-specific nature of many regulatory regions, we further investigated the distribution of co-localizing lead variants within regions known to modulate human islet chromatin structure.[28] This revealed a significant enrichment in the human islet regulome (Figure 4G; Table S10), particularly in active promoters (n = 30), active enhancers (n = 50), inactive enhancers (n = 18), and strong CTCF-binding regions (n = 29). No enrichment was observed in open inactive chromatin regions. When applying an enhancer subclassification, the most significant enrichment of co-localizing lead variants was detected in active class I enhancers ($p = 2.16 \times 10^{-18}$; n = 25), characterized by higher H3K27ac marks and stronger Mediator protein occupancy, both key factors in gene expression regulation.[28]

In summary, 361 co-localizing lead variants were found to reside in exons, transcripts, and regulatory and islet-specific chromatin interaction regions that consist of active promoters and enhancers.[28,32] These findings underscore the functional importance of these variants in gene regulation and their relevance to disease susceptibility.

Co-localizing variants identify important pathways

Gprofiler2 pathway enrichment of pooled co-localizing genes for type 2 diabetes and glycemic traits revealed pathways related to cellular localization, vesicle dynamics, mitochondrial function, fatty acid metabolism, and synthesis (Figures 4H, S2A, and S2B), suggesting a role for these islet cell processes in type 2 diabetes pathogenesis. These pathway enrichments were corroborated by analyses using Metascape and EnrichR (Figures S2C–S2F). Genes downregulated by lead variants

Figure 3. *colocRedRibbon* shortlisting steps identify genuine co-localizations

(A) GWAS and eQTL Manhattan plots for the *RPL39L* region. Two signals are present for GWAS. *coloc* alone does not detect co-localization: the lead SNP (red dot) on the highest GWAS peak does not have an eQTL signal, while the matching eQTL and GWAS peaks are missed. The red dotted line is the gene transcription start site.

(B) *colocRedRibbon* shortlisting method applied to *RPL39L* detects a co-localization for the lower GWAS signal. Overlapping—shortlisted—SNPs are shown in blue and the lead SNP in red on the bottom.

(C) Islet *RPL39L* expression according to lead SNP C or T alleles.

(D) GWAS and eQTL Manhattan plots for the *TH* gene. *coloc* does not detect a significant co-localization for the eQTL peak and 6 GWAS peaks.

(E) *colocRedRibbon* shortlisting of SNPs that increase diabetes risk (increasing risk allele [IRA]) and lower *TH* expression unmasks co-localization. The most significant GWAS peak (see D) is dismantled by shortlisting the direction of effect of gene expression. The overlap map in the top left shows significant overlap between GWAS and eQTL SNPs. Shortlisted SNPs are shown in color (bottom left): overlapping SNPs in blue, 99% credible set in green, and lead SNP in red.

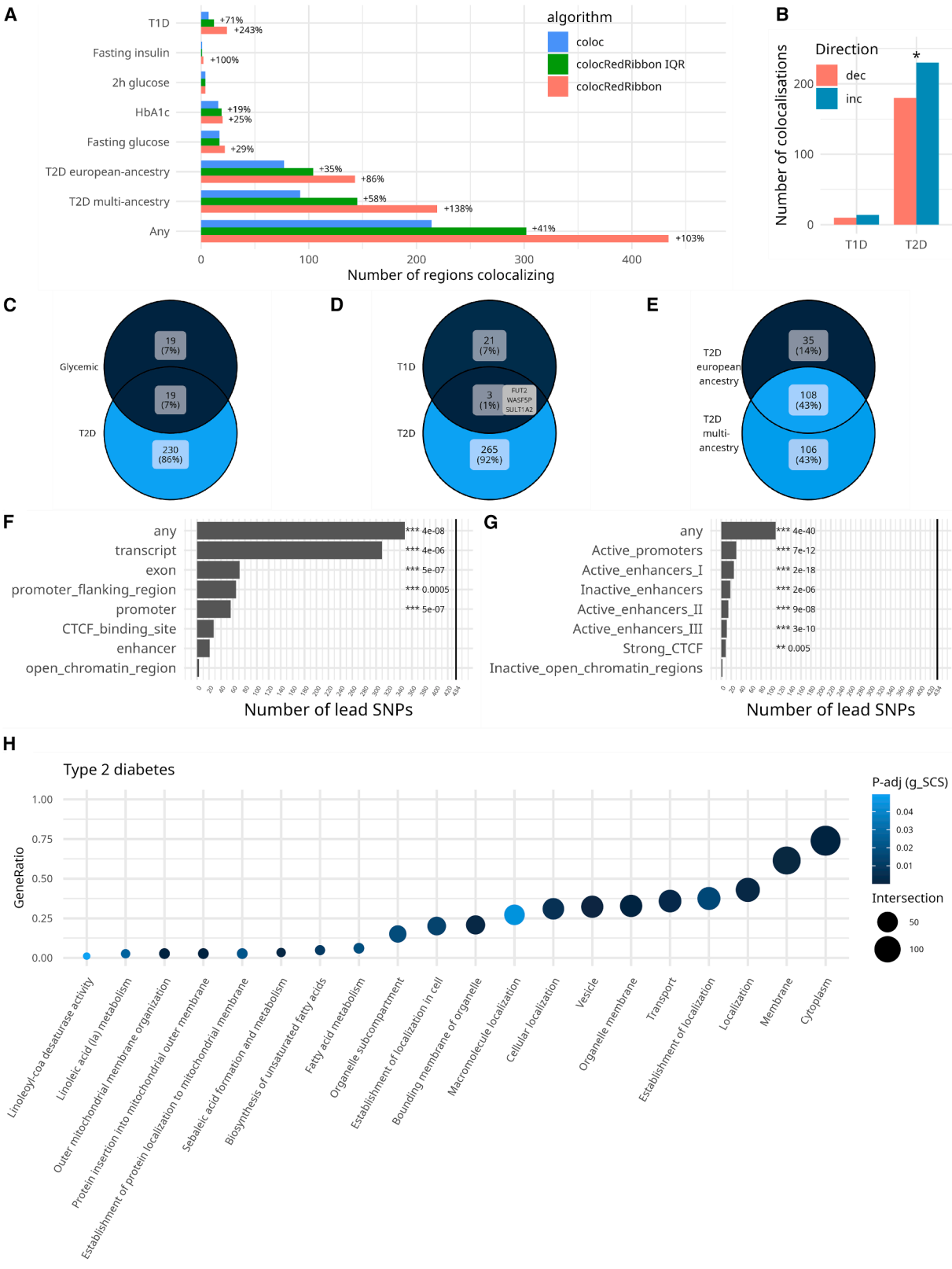
(F) There is no co-localization for risk alleles that decrease *TH* gene expression.

(G) *SPATA20* shows co-localization for type 2 diabetes GWAS variants with islet eQTLs (left) and plasma pQTLs (right), sharing lead variant rs8076632.

See also Figure S1.

Please cite this article in press as: Piron et al., Identification of novel type 1 and type 2 diabetes genes by co-localization of human islet eQTL and GWAS variants with *colocRedRibbon*, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101004

CellPress
OPEN ACCESS



(legend on next page)

were mainly involved in cellular localization and trafficking, whereas upregulated genes were primarily associated with cytoplasmic function (Figures S2A and S2B). The co-localizing gene set exhibited significant ($p = 0.00783$) overlap with genes differentially expressed in islets from type 2 diabetic vs. non-diabetic donors[33] (Table S2, column fGSEA.t2d). For example, we detected a signal associated with type 2 diabetes risk and decreased *PDE8B* islet expression. *PDE8B* was downregulated in islets from type 2 diabetic vs. non-diabetic donors and upon palmitate exposure of non-diabetic donor islets.[34] *PDE8B* encodes a high-affinity cAMP-specific phosphodiesterase that maintains low basal cAMP levels in β cells, with *PDE8B* knockdown disturbing glucose-stimulated sub-plasma-membrane cAMP oscillations and pulsatile insulin secretion.[35]

Among the type 2 diabetes genes in the vesicle pathway (Figures 4H and S2B) is *MYO5C*. The low minor-allele frequency (5%) T allele confers protection against type 2 diabetes and increases islet *MYO5C* gene expression (Figure S3). *MYO5C* encodes myosin type V, a molecular motor that moves along actin filaments and is involved in secretory granule trafficking.[36] The lead variant, rs149336329, is located in promoter and transcript regions, as well as in a human islet-specific active promoter (marked by Mediator, H3K27ac, and H3K4me3) (Tables S10 and S11) near the transcription start site (Figure S3A), making it a plausible causal variant. A different lead SNP (rs3825801) was reported by Mahajan et al.[10] after fine-mapping, but *coloc* co-localization was unsuccessful.

Another type 2 diabetes gene is *RPL39L*, with the T risk allele inducing islet *RPL39L* expression (Figures 3A–3C). This unique variant in the 99% credible set is noteworthy due to its presence in multiple regulatory regions, i.e., promoter flanking region, transcripts, and islet-specific active enhancers. RPL39L has been associated with type 2 diabetes in differential gene expression studies of islets from type 2 diabetic vs. non-diabetic donors,[33] a finding corroborated by type 2 diabetes fast gene set enrichment analysis (fGSEA) enrichment analysis (Table S2, column fGSEA.t2d). *RPL39L* encodes ribosomal protein L39-like; its function in islets is unknown. Interestingly, the same variant also affects *ST6GAL1* expression, although *ST6GAL1* is not part of the leading edge in our type 2 diabetes fGSEA enrichment for human islets (Table S12, column fGSEA.t2d). *RPL39L* and *ST6GAL1* are co-expressed in multiple tissues, including thyroid, prostate, ovary, artery, and fibroblasts.[37] Notably, *ST6GAL1* belongs to cellular pathways in our enrichment analysis, i.e., organelle subcompartment, bounding membrane of organelle, organelle membrane, membrane, and cytoplasm (Figure 4H; Table S12).

colocRedRibbon detected 8 co-localizations for type 1 diabetes in the HLA region of chromosome 6 (Figure 2). Pathway enrichment analysis for type 1 diabetes genes did not reach significance due to the relatively limited number ($n = 24$) of co-localizations.

Among the type 1 diabetes co-localizations, variant rs11666792 emerged as the lead variant for both *FUT2* and *RASIP1* (Figure 5). These genes are co-expressed in the esophageal mucosa and skin.[37] *RASIP1* encodes Ras-interacting protein 1, which has GTPase-binding and protein homodimerization activities. *FUT2*, one of the three genes that also co-localize for type 2 diabetes (albeit with a distinct lead variant), encodes the enzyme alpha-1,2-fucosyltransferase. This enzyme is involved in the synthesis of the H antigen, the precursor of secreted ABO blood group antigens. *FUT2* has been associated with type 1 diabetes[38]: a common nonsense mutation in *FUT2* abolishes the secretion of ABO antigens in saliva and the gut, alters the gut microbiome, and confers susceptibility to type 1 diabetes while protecting from some viral infections.[38,39] In human islets, the protective G allele of rs11666792 increases *FUT2* expression (Figure 5B). FUT2 silencing in human EndoC-βH1 β cells increased coxsackievirus B1-induced apoptosis, suggesting that FUT2 might protect β cells against potentially diabetogenic viruses (Figures S4A and 5C). FUT2 silencing did not affect β cell function: insulin content and glucose- and glucose+forskolin-stimulated insulin secretion were comparable in FUT2-depleted and -competent EndoC-βH1 cells (Figures S4B and S4C).

Cathepsin H (encoded by *CTSH*), a type 1 diabetes candidate gene,[40,41] exhibits a co-localization for lead variant rs34593439, having a 0.10 minor-allele frequency and a small 99% credible set (5 variants, Figure S5A). The protective allele A (type 1 diabetes odds ratio = 0.84) lowers islet *CTSH* expression (Figure S5B). *CTSH* displays complex regulation in response to inflammatory stimuli. It is downregulated in EndoC-βH1 cells upon exposure to pro-inflammatory cytokines interleukin (IL)-1β + interferon (IFN)γ.[26] The risk allele T of variant rs3825932 has been shown to reduce islet *CTSH* expression[40]; this variant is in linkage disequilibrium with rs34593439. Lower methylation variability may make individuals with the protective allele potentially less sensitive to cytokines.[41] The gene is also associated with celiac disease[42]; type 1 narcolepsy,[43] where the rs34593439 A allele increases risk; and rheumatoid arthritis.[44] This association with multiple autoimmune diseases makes the gene and variant good candidates for further investigation.

Figure 4. *colocRedRibbon* doubles the number of co-localizations and detected variants that lie in islet-specific regulatory regions

(A) Number of co-localizing regions. *colocRedRibbon* (red) significantly improves detection of co-localizations compared with *coloc* (blue). The more conservative *colocRedRibbon* interquartile range (IQR) method (green) is also better than *coloc*. The percentage increase is shown for GWAS hits for European and multi-ancestry type 2 diabetes (T2D), type 1 diabetes (T1D), and glycemic traits HbA1c, fasting glucose, 2-h glucose post-challenge, and fasting insulin. "Any" includes all GWAS SNPs combined. Co-localizations overall increased by 103%.

(B) Direction of effect of the risk allele on gene expression. Inc, increased gene expression; dec, decreased gene expression.

(C–E) The number of regulated genes in common for glycemic traits and diabetes GWAS.

(F) Enrichment in regulatory regions for lead SNPs. "Any" means any of the regions.

(G) Human islet-specific regulome enrichment.[28,29] $^{***}p \leq 0.001$, $^{**}p \leq 0.01$, and $^{*}p \leq 0.05$.

(H) Gprofiler2 enrichment in co-localizing genes for type 2 diabetes and glycemic traits.

See also Figures S2 and S3.

Please cite this article in press as: Piron et al., Identification of novel type 1 and type 2 diabetes genes by co-localization of human islet eQTL and GWAS variants with *colocRedRibbon*, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101004

CellPress
OPEN ACCESS

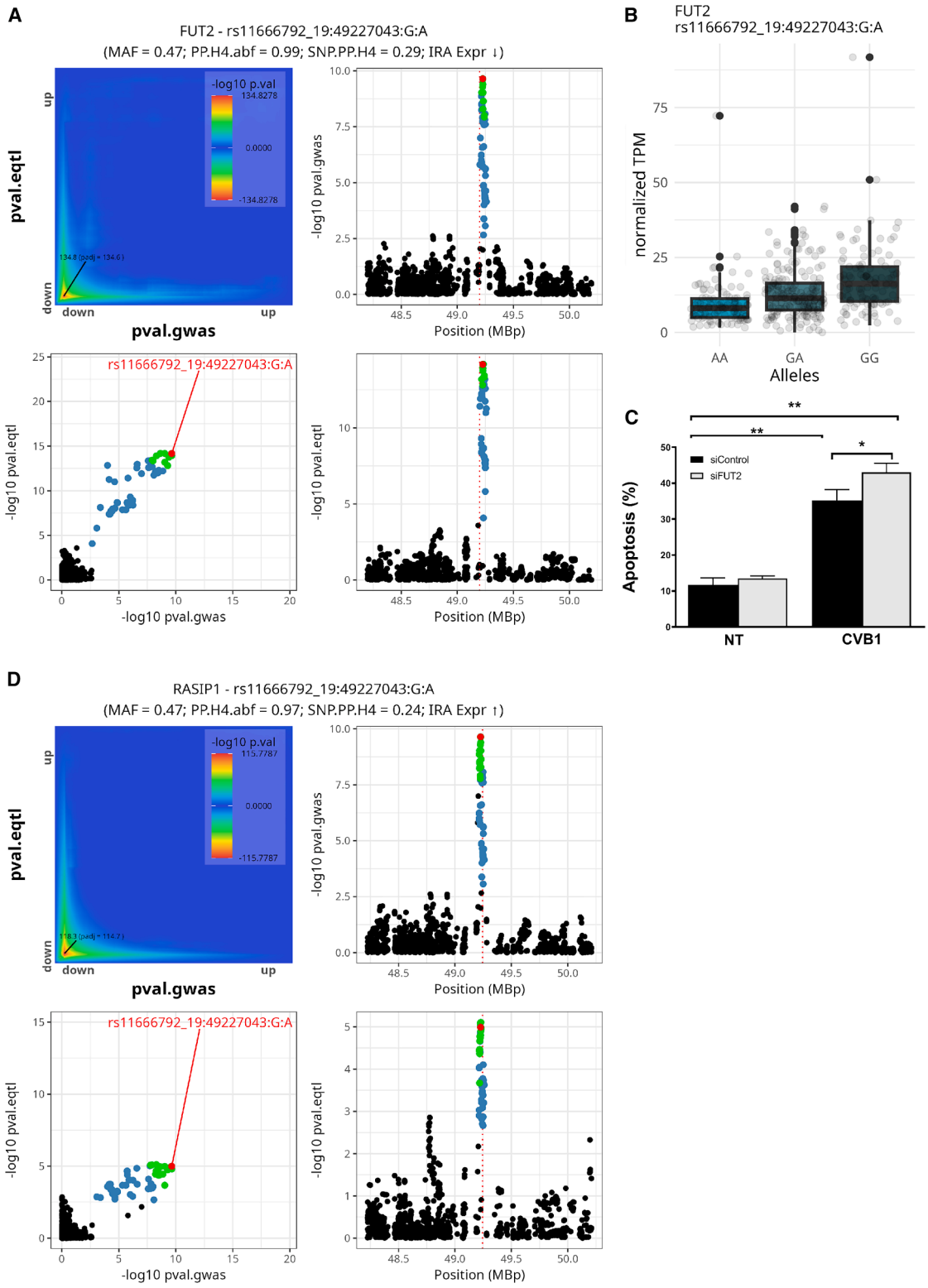


Figure 5. Co-localization for co-expressed *FUT2*/*RASIP1* eQTL and type 1 diabetes GWAS

(A) Co-localization plots for *FUT2*, showing lead SNP in red, 99% credible set in green, and *RedRibbon* overlap in blue. The red dotted line is the gene transcription start site.

(legend continued on next page)

Glycemic trait genes are shared with type 2, but not type 1, diabetes

Among the 48 co-localizations identified for glycemic traits, only *SULT1A2* was present among type 1 diabetes genes. In contrast, 19 (50%) glycemic trait genes were detected for type 2 diabetes, with 11 sharing an identical lead variant. A striking example of this overlap is single lead variant rs11708067 affecting *ADCY5* expression,[45] which co-localizes for type 2 diabetes and all glycemic traits (Figure 6). The type 2 diabetes risk allele A decreases human islet *ADCY5* expression (Figures 6A and 6B). Consistent with its effect on type 2 diabetes risk, the A allele is associated with increased HbA1c and fasting and 2-h glucose and decreased islet *ADCY5* expression (Figures 6C–6E). Conversely, the A allele is associated with increased fasting insulin and increased *ADCY5* expression (Figure 6F). *ADCY5* encodes adenylate cyclase 5, an enzyme that converts ATP to cAMP, generating an essential amplifier of insulin secretion. Decreased *ADCY5* expression impairs glucose-induced cAMP production and insulin secretion.[46] These congruent co-localization data strongly support rs11708067 as the causal variant that increases diabetes risk by decreasing *ADCY5* expression. It underscores the importance of ADCY5 in glucose homeostasis, highlighting how a single genetic variant can have pleiotropic effects on related traits.

Concordance between transcriptomic and proteomic co-localizations in diabetes

To determine whether co-localization signals at the transcriptome level are reflected in the proteome, we used *colocRedRibbon* to co-localize GWAS SNPs with plasma pQTLs.[21] Of 268 unique co-localizing genes for type 2 diabetes and glycemic traits and 24 for type 1 diabetes, 68 with at least one significant pQTL (Table S13) were detected by aptamers in plasma.[21] Of these, 21 regions showed co-localization with pQTLs, resulting in a total of 40 co-localizations across all traits (Tables S13 and 1). Variant rs8076632 co-localizes with the SPT20 protein and the corresponding spermatogenesis-associated 20 (*SPATA20*) gene (Figure 3G), with the type 2 diabetes risk allele located in an exon concordantly increasing mRNA and protein expression. While a previous study reported no co-localization in this region,[47] another did, albeit with an unconvincing lead variant.[48] The mechanism by which *SPATA20* alters type 2 diabetes risk remains to be elucidated.

CTSH protein contains the type 1 diabetes-associated variant rs2289702 within its 99% credible set. The risk allele again demonstrates a concordant direction of effects on gene and protein expression.

Concordance was also seen for PDCD5 and ABO, the latter in multiple GWASs (Table S13).

ColocRedRibbon enables co-localization for diverse diseases and traits

To demonstrate the broader applicability and versatility of *colocRedRibbon*, we used it for other combinations of GWASs and/or QTLs.

We first examined co-localizations between different GWASs using *colocRedRibbon*. We analyzed the region around *FTO*, known to be associated with type 2 diabetes and BMI (Table S14). Strong overlap was seen for variants from obesity[23] and type 2 diabetes GWASs, with rs11642015 being the lead SNP (Figure S6). The risk alleles are concordant, suggesting that this variant confers susceptibility to obesity and type 2 diabetes.

We next investigated the relationship between the obesity GWAS[23] and islet eQTLs,[17] expecting fewer co-localizations than for diabetes, given the central role of the brain and adipocytes in obesity. This identified 12 co-localizations (Table S15). Obesity-associated variants in a chromosome 16 locus increase islet *EIF3C* and *EIF3CL* expression and decrease lincRNA *ENSG00000251417* (Figure S7); these islet eQTLs also co-localize with type 2 diabetes (Table S2). The increased *EIF3C* transcript expression associated with the lead risk allele rs62036658 is concordant with obesity and type 2 diabetes. *EIF3C* encodes subunit c of eukaryotic translation initiation factor 3, a critical protein for mRNA start codon selection and ribosome assembly.[49]

Finally, we co-localized the BMI GWAS[24] with subcutaneous adipose tissue meQTLs for 28 regions targeted in McAllan et al.,[22] identifying 5 co-localizations (Table S16). The BMI-increasing rs12450225 allele augments methylation at the cg14884929 site (Figure S8). cg14884929 is located near *MIR10A*, encoding microRNA 10a that regulates adipose tissue inflammation and homeostasis.[50,51] These results suggest that genetic variants influencing BMI may, in part, exert their effects through the modulation of adipocyte DNA methylation.

Taking these results together, we show that *colocRedRibbon* has broad applicability for co-localization analyses across multiple omics and is versatile in uncovering tissue-specific or shared genetic mechanisms underlying disease phenotypes.

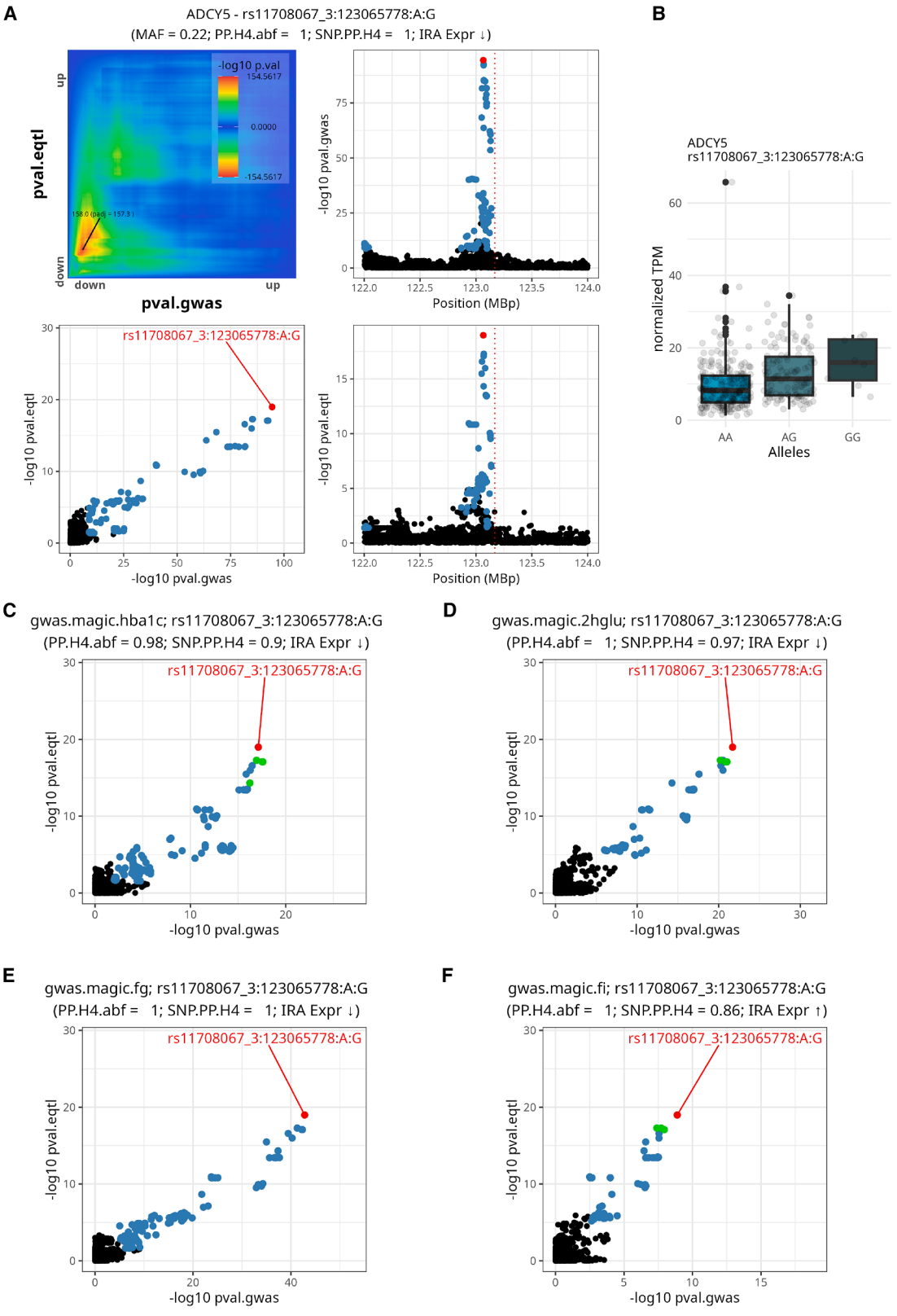
DISCUSSION

Over the past decades, a plethora of GWASs have linked common genetic variation to complex diseases and traits. Because 90% of all GWAS hits lie in non-coding regions, the identification of genes that mediate susceptibility is not straightforward. eQTL studies in disease-relevant tissues can connect SNPs to functional genes, but the yield of GWAS and eQTL variant co-localization studies tends to be low. This is illustrated by the >700 variants identified in the diabetes GWAS, of which only around

(B) *FUT2* expression plot for co-localizing SNP.
(C) Human EndoC-βH1 β cell apoptosis 24 h after control or FUT2 small interfering RNA (siRNA) transfection and infection or not (NT) with coxsackievirus B1 ($n = 5$). $^{**}p \leq 0.01$and $^{*}p \leq 0.05$.
(D) Co-localization plots for *RASIP1*. SNPs are referenced by rsids, and PP.H4.abf and SNP PP.H4 are the posterior probabilities for *coloc* tool hypothesis 4. MAF, minor-allele frequency; IRA, increasing risk allele expression regulation direction. An arrow pointing up or down means increased or decreased gene expression for the risk allele, respectively.
See also Figure S4.

Please cite this article in press as: Piron et al., Identification of novel type 1 and type 2 diabetes genes by co-localization of human islet eQTL and GWAS variants with *colocRedRibbon*, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101004

CellPress
OPEN ACCESS



(legend on next page)

Please cite this article in press as: Piron et al., Identification of novel type 1 and type 2 diabetes genes by co-localization of human islet eQTL and GWAS variants with *colocRedRibbon*, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101004

50 have so far been linked to changes in pancreatic islet gene expression. Here, we significantly expanded the number of co-localizations for type 2 diabetes and identified many novel type 1 diabetes co-localizations using the novel *colocRedRibbon* pipeline. This pipeline detected 434 co-localizations corresponding to 289 genes and 344 distinct lead variants. This increase is attributed to a combination of new computational methods and larger, multi-ethnic, and multi-trait GWASs. Without the methodological advancements of *colocRedRibbon*, we would have identified 214 co-localizations (128 distinct gene regions), i.e., less than half, highlighting the importance of these improvements in uncovering novel associations. *colocRedRibbon* filters variants or delimits regions (interquartile range [IQR] mode, Figure 1B) to be further analyzed for co-localization. We used *coloc*, but any other tool may be used (e.g., *SuSiE + coloc*[52,53] or *smr*[54]).

Our analysis revealed that 88% of lead variants are localized in regulatory regions, including those specific to islets. For type 2 diabetes and glycemic traits, genes are involved in cellular localization, vesicle dynamics, mitochondrial function, fatty acid metabolism, and synthesis (Figure S1). The enrichment in regulatory regions and pathways crucial for β cell function/survival demonstrates the potential of *colocRedRibbon* to identify disease-relevant mechanisms. When comparing detected genes with those reported as differentially expressed in islets in type 2 diabetes,[33] 51 genes overlapped, pertaining primarily to cellular localization and vesicle pathways. This is substantial, considering the relatively modest transcriptome sample size (28 type 2 diabetic and 58 non-diabetic donors) compared to the large GWAS and eQTL datasets.

Co-localizing lead variants were also found in genes potentially involved in monogenic diabetes, such as *FOXA2* and *PCBD1*. The transcription factor FOXA2 contributes to pancreas development and islet gene expression. *FOXA2* mutations have been implicated in diabetes and congenital hyperinsulinism.[55] PCBD1 acts as a cofactor of transcription factor HNF1A, which is essential for pancreatic development and function.[56] These findings underscore parallels between monogenic and polygenic diabetes, as have been described for uncommon variants in monogenic diabetes genes that increase type 2 diabetes risk, on average, by 3-fold.[57]

The HLA region on chromosome 6 resulted in 8 co-localizations for type 1 diabetes (Figure 2). This region poses a significant challenge for co-localization analyses due to the complex structure of the linkage disequilibrium. It contains numerous immunity-related genes that are under strong selective pressure,[58] leading to highly intricate linkage disequilibrium patterns. For several co-localizations (e.g., *BTN2A3P* and *HCG20*), *colocRedRibbon* detected significant subsets of variants through the overlap prefiltering step. The visual representation of these co-localizations (see data and code availability) appears inconclusive, however, as the *colocRedRibbon*-identified variants lie within stronger GWAS signals. This could be a side effect of the simplification used to assess linkage disequilibrium in adjusted *p* value computation, overestimating the overlap significance or, alternatively, acting as a reflection of the method's ability to discern multiple overlapping GWAS signals, resulting in valid co-localization. Other co-localizations in the HLA region (e.g., *PSMB9* and *HLA-DQB1-AS1*) appear more convincing, with well-isolated GWAS signals and the lead variant close to the top. Future validation of HLA region co-localizations is required, probably using specialized tools tailored to handle the unique challenges posed by this genomic locus. Adapting and fine-tuning *colocRedRibbon* may help to address the complexities of the HLA region, but this will require a dedicated effort.

colocRedRibbon can now be applied to eQTL studies in other disease-relevant tissues, such as the immune system for type 1 diabetes,[59] as well as muscle, liver, and adipose tissue for type 2 diabetes and glycemic traits.[60–65] The use of the same pipeline across multiple tissues and QTLs will comprehensively characterize the tissue-specific genetic architecture of these diseases.

The tool is also applicable to QTLs for protein, open chromatin, and DNA methylation; the method only requires that GWAS and QTL data share common genomic positional features. A limitation is the current scarcity of pQTLs, open chromatin, and meQTLs in relevant tissues. We applied *colocRedRibbon* to plasma pQTLs,[21] meQTLs,[22] and obesity[23] and BMI[24] GWASs, showcasing its adaptability to different QTL types and its applicability across diseases or traits. The method is equally suitable for investigating chromosomal regions located farther away from the gene—such as those with *trans*-QTL signals. Users can specify any region of interest, including the union of regions around *trans*-QTL variants associated with the gene under study. This flexibility allows *colocRedRibbon* to be readily adapted for both *cis*- and *trans*-QTL co-localization analyses.

In summary, this study leveraged the novel *colocRedRibbon* pipeline to analyze a large dataset from human pancreatic islets, a tissue central in the pathogenesis of diabetes, allowing us to investigate the expression regulatory variation underlying this complex disease. *colocRedRibbon* significantly enhanced the detection of co-localizations of eQTL and GWAS variants compared with previous methods, refining the mapping of lead variants and uncovering numerous novel co-localizations. These results are a critical step toward understanding how diabetes-associated variants impact pancreatic islet cell function, elucidating the mechanisms linking genetic variation with diabetes and, potentially, guiding the translation of GWAS hits to (personalized) diabetes therapy.

Figure 6. Co-localization of *ADCY5* eQTL and type 2 diabetes and glycemic trait GWAS SNPs

Co-localization plots for *ADCY5* show the same lead SNP in red, the 99% credible set in green, and *RedRibbon* overlap in blue. The red dotted line is the gene transcription start site.

(A) Type 2 diabetes co-localization.

(B) Expression level for alleles regulating *ADCY5*.

(C–F) Co-localization for (C) glycated hemoglobin (HbA1c), (D) 2-h glucose (2hGlu), (E) fasting glucose (FG), and (F) fasting insulin (FI). SNPs are referenced by rsids, and PP.H4.abf and SNP PP.H4 are the posterior probabilities for *coloc* tool hypothesis 4. MAF, minor-allele frequency; IRA, increasing risk allele expression regulation direction. An arrow pointing up or down means increased or decreased gene expression for the risk allele, respectively.

Please cite this article in press as: Piron et al., Identification of novel type 1 and type 2 diabetes genes by co-localization of human islet eQTL and GWAS variants with *colocRedRibbon*, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101004

CellPress
OPEN ACCESS

Table 1. Selected co-localizations for diabetes and glycemic traits

Co-localization				PP.H4		Cross-analyses		
GWAS	Gene (pQTL)	Lead SNP	IRA	abf	SNP	Other GWASs	Previously reported	Regions
T2D	*CCDC67*	rs11600595	inc	86%	91%	T2D Eur.	INSPire	–
T2D	*TH*	rs10840495	dec	97%	98%	T2D Eur.	–	–
T2D	*GSAP*	rs3912067	inc	82%	21%	T2D Eur.	–	promoter flanking region
T2D	*GSAP*	rs9641219	dec	93%	92%	T2D Eur.	–	promoter, transcript, active promoters
T2D	*MTNR1B*	rs10830963	inc	99%	100%	FG, HbA1c, T2D Eur.	TIGER	transcript, Inactive enhancers
T2D	*FUT2*	rs681343	inc	99%	25%	T1D	–	exon, transcript
T2D	*WASF5P*	rs2905758	inc	97%	56%	T1D	–	transcript
T2D	*PRELID1*	rs28607106	dec	91%	62%	–	–	–
T2D	*GRSF1*	rs10433674	dec	97%	34%	–	GSEA	promoter flanking region, active enhancers
T2D	*PDE8B*	rs7732130	inc	96%	99%	FG, T2D Eur.	INSPire, GSEA	enhancer, transcript, active enhancers
T2D	*MYO5C*	rs149336329	dec	94%	99%	T2D Eur.	–	promoter, transcript, active promoters
T2D	*RPL39L*	rs3887925	inc	99%	98%	T2D Eur.	GSEA	promoter flanking region, transcript, active enhancers
T2D	*ST6GAL1*	rs3887925	inc	99%	99%	T2D Eur.	TIGER	promoter flanking region, transcript, active enhancers
T2D	*ADCY5*	rs11708067	dec	99%	99%	2hglu, FG, FI, HbA1c, T2D Eur.	TIGER, INSPire, GSEA	promoter flanking region, transcript, inactive enhancers
T2D	*FOXA2*	rs6048206	inc	89%	32%	–	GSEA	–
T2D	*PCBD1* (PHS)	rs827241	dec	98%	53%	T2D Eur.	TIGER, GSEA	promoter, exon, transcript, active promoters
T2D	*SPATA20* (SPT20)	rs8076632	inc	99%	58%	–	GSEA	exon, transcript
HbA1c	*SULT1A2*	rs4149406	inc	86%	22%	T1D	GSEA	promoter, transcript
T1D	*FUT2*	rs11666792	dec	98%	29%	T2D	–	transcript
T1D	*SULT1A2*	rs13331170	dec	81%	12%	HbA1c	–	–
T1D	*WASF5P*	rs3130297	inc	93%	99%	T2D	–	–
T1D	*RASIP1*	rs11666792	inc	96%	23%	–	–	transcript
T1D	*CTSH* (cathepsin H)	rs34593439	inc	99%	25%	–	–	transcript
T1D	*PSMB9*	rs2071465	inc	99%	99%	–	–	promoter, transcript
T1D	*HLA-DQB1-AS1*	rs17206350	inc	99%	100%	–	–	–

Gene shows gene symbol. If a protein quantitative trait locus (pQTL) also co-localizes, the protein name is displayed in parentheses. GWASs are for type 2 diabetes multi-ancestry (T2D), type 2 diabetes European ancestry (T2D Eur.), type 1 diabetes (T1D), fasting glucose (FG), fasting insulin (FI), 2-h glucose (2hglu), and HbA1c. IRA indicates the direction of gene expression associated with the risk-increasing allele, where "inc" denotes increased and "dec" decreased expression. GSEA refers to our fast gene set enrichment analysis (fGSEA) comparing type 2 diabetic vs. non-diabetic islet differential gene expression.[33] INSPire data are from Viñuela et al.[18] and TIGER data from Alonso et al.[17] Genomic regions are defined based on GENCODE genome annotation and chromatin regulation data.[28]

colocRedRibbon is a powerful new tool for integrating genetic and gene expression data to elucidate the complex interplay between genetic variation and disease risk. It is poised to expand knowledge of the genetic architecture of many complex diseases and highlight avenues for therapeutic development.

Limitations of the study

colocRedRibbon uses the *RedRibbon* approach that employs an evolutionary algorithm to identify the minimal hypergeometric *p* value within the overlap map. As a result, *colocRedRibbon* is subject to the inherent non-determinism

associated with such algorithms. This stochastic variability can be managed effectively by performing multiple independent runs of the overlap analysis to ensure the robustness of the results.

Another limitation lies in the limited availability of pQTL, open chromatin, and meQTL data in relevant tissues. This restricts the comprehensiveness of our analyses and, consequently, limits our ability to fully capture the regulatory landscape in all biologically relevant contexts.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Anthony Piron (anthony.piron@ulb.be).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The eQTL dataset is from Alonso et al.,[17] pQTLs from Ferkingstad et al.,[21] meQTLs from McAllan et al.,[22] and GWASs from Chen et al.,[11] Suzuki et al.,[15] Kurki et al.,[23] Yengo et al.,[24] and GCST90013791. The co-localization list, plots, and source code to generate these data are available on Zenodo via accession number https://doi.org/10.5281/zenodo.13987433.
- The R package code is open to the community with a permissive license (GPL3) and available for download from GitHub https://github.com/antpiron/colocRedRibbon (also available on Zenodo https://doi.org/10.5281/zenodo.13987433). This package depends on https://github.com/antpiron/RedRibbon (*colocRedRibbon* R package also contains this code).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

T2DGGI members and their affiliations can be found at https://diagram-consortium.org/T2DGGI.html. We thank Piero Marchetti, University of Pisa, and his team for their essential role in providing islet samples and data to TIGER. This work has been supported by the European Union's Horizon 2020 research and innovation programme T2DSystems under grant agreement no. 667191; the Fonds National de la Recherche Scientifique (FNRS); Walloon Region SPW-EER Win2Wal project BetaSource, Belgium; FWO and FRS-FNRS under the Excellence of Science (EOS) programme (Pandarome project 40007487); Walloon Region strategic axis FRFS-WELBIO; and the Innovative Medicines Initiative 2 Joint Undertaking under grant agreements 115797 (INNODIA) and 945268 (INNODIA HARVEST). This latter undertaking received support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations, JDRF, and the Leona M. and Harry B. Helmsley Charitable Trust. A.P. is supported by Fonds David et Alice Van Buuren, Fondation Jaumotte-Demoulin, Fondation Héger-Masson, Fondation Wiener-Anspach, and FNRS. F.S. is supported by a Research Fellow FNRS fellowship. D.L.E. is supported by grants from JDRF (now Breakthrough T1D) (3-SRA-2022-1201-S-B and 3-SRA-2022-1201-S-B), the National Institutes of Health Human Islet Research Network Consortium on Beta Cell Death & Survival from Pancreatic β-Cell Gene Networks to Therapy (HIRN-CBDS) (U01 DK127786), and National Institutes of Health NIDDK grants RO1DK126444 and RO1DK133881-01. J.M.M. is supported by American Diabetes Association grant #11-22-ICTSPM-16, NHGRI U01HG011723, the National Institute of Diabetes and Digestive and Kidney Diseases of the National Institutes of Health under award nos. R01DK137993 and U01 DK140757, and a Medical University of Bialystok (MUB) grant from the Ministry of Science and Higher Education (Poland). A.H.-C. is supported by American Diabetes Association grant 11-23-PDF-35.

AUTHOR CONTRIBUTIONS

Conceptualization, A.P., J.M.M., and M.C.; methodology, A.P., J.M.M., and M.C.; formal analysis, A.P.; data curation, A.P.; software, A.P.; software documentation, T.P.; visualization, A.P., L.F., and M.L.; investigation, A.P., F.S., L.F. M.L., and Y.T.; writing – original draft, A.P., F.S., L.F., and M.C.; writing – review & editing, D.J.M.C., T.P., M.L., Y.T., M.I.A., M.L.C., M.L.P., K.H., A.H.-C., H.J.T., M.D., D.L.E., J.M.M., and M.C.; resources, X.Y., D.J.M.C., J.A.T., and M.C.; supervision, M.C.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - colocRedRibbon workflow
  - colocRedRibbon shortlisting
  - colocRedRibbon R package
  - Pathway and regulatory elements enrichment
  - Normalization of gene expression
  - Viral infection and FUT2 knockdown in human beta cells
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Co-localization
  - Enrichment analyses
  - *FUT2* statistical analyses

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2025.101004.

Received: December 12, 2024
Revised: June 29, 2025
Accepted: August 18, 2025

REFERENCES

1. IDF (2025). IDF Diabetes Atlas (International Diabetes Federation). https://diabetesatlas.org/.
2. Ong, K.L., Stafford, L.K., McLaughlin, S.A., Boyko, E.J., Vollset, S.E., Smith, A.E., Dalton, B.E., Duprey, J., Cruz, J.A., Hagins, H., et al. (2023). Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet *402*, 203–234. https://doi.org/10.1016/S0140-6736(23)01301-6.
3. Ilonen, J., Lempainen, J., and Veijola, R. (2019). The heterogeneous pathogenesis of type 1 diabetes mellitus. Nat. Rev. Endocrinol. *15*, 635–650. https://doi.org/10.1038/s41574-019-0254-y.
4. Byndloss, M., Devkota, S., Duca, F., Niess, J.H., Nieuwdorp, M., Orho-Melander, M., Sanz, Y., Tremaroli, V., and Zhao, L. (2024). The gut microbiota and diabetes: research, translation, and clinical applications - 2023 Diabetes, Diabetes Care, and Diabetologia Expert Forum. Diabetologia *67*, 1760–1782. https://doi.org/10.1007/s00125-024-06198-1.
5. Xourafa, G., Korbmacher, M., and Roden, M. (2024). Inter-organ crosstalk during development and progression of type 2 diabetes mellitus. Nat. Rev. Endocrinol. *20*, 27–49. https://doi.org/10.1038/s41574-023-00898-1.
6. Cohrs, C.M., Panzer, J.K., Drotar, D.M., Enos, S.J., Kipke, N., Chen, C., Bozsak, R., Schöniger, E., Ehehalt, F., Distler, M., et al. (2020).

Dysfunction of Persisting β Cells Is a Key Feature of Early Type 2 Diabetes Pathogenesis. Cell Rep. *31*, 107469. https://doi.org/10.1016/j.celrep.2020.03.033.

7. Eizirik, D.L., Pasquali, L., and Cnop, M. (2020). Pancreatic β-cells in type 1 and type 2 diabetes mellitus: different pathways to failure. Nat. Rev. Endocrinol. *16*, 349–362. https://doi.org/10.1038/s41574-020-0355-7.
8. Suleiman, M., Marselli, L., Cnop, M., Eizirik, D.L., De Luca, C., Femia, F.R., Tesi, M., Del Guerra, S., and Marchetti, P. (2022). The Role of Beta Cell Recovery in Type 2 Diabetes Remission. Int. J. Mol. Sci. *23*, 7435.
9. Mahajan, A., Taliun, D., Thurner, M., Robertson, N.R., Torres, J.M., Rayner, N.W., Payne, A.J., Steinthorsdottir, V., Scott, R.A., Grarup, N., et al. (2018). Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat. Genet. *50*, 1505–1513. https://doi.org/10.1038/s41588-018-0241-6.
10. Mahajan, A., Spracklen, C.N., Zhang, W., Ng, M.C.Y., Petty, L.E., Kitajima, H., Yu, G.Z., Rüeger, S., Speidel, L., Kim, Y.J., et al. (2022). Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat. Genet. *54*, 560–572. https://doi.org/10.1038/s41588-022-01058-3.
11. Chen, J., Spracklen, C.N., Marenne, G., Varshney, A., Corbin, L.J., Luan, J., Willems, S.M., Wu, Y., Zhang, X., Horikoshi, M., et al. (2021). The trans-ancestral genomic architecture of glycemic traits. Nat. Genet. *53*, 840–860. https://doi.org/10.1038/s41588-021-00852-9.
12. Chiou, J., Geusz, R.J., Okino, M.-L., Han, J.Y., Miller, M., Melton, R., Beebe, E., Benaglio, P., Huang, S., Korgaonkar, K., et al. (2021). Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. Nature *594*, 398–402. https://doi.org/10.1038/s41586-021-03552-w.
13. Bonas-Guarch, S., Guindo-Martinez, M., Miguel-Escalada, I., Grarup, N., Sebastian, D., Rodriguez-Fos, E., Sanchez, F., Planas-Felix, M., Cortes-Sanchez, P., Gonzalez, S., et al. (2018). Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes. Nat. Commun. *9*, 321. https://doi.org/10.1038/s41467-017-02380-9.
14. Vujkovic, M., Keaton, J.M., Lynch, J.A., Miller, D.R., Zhou, J., Tcheandjieu, C., Huffman, J.E., Assimes, T.L., Lorenz, K., Zhu, X., et al. (2020). 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. *52*, 680–691. https://doi.org/10.1038/s41588-020-0637-y.
15. Suzuki, K., Hatzikotoulas, K., Southam, L., Taylor, H.J., Yin, X., Lorenz, K.M., Mandla, R., Huerta-Chagoya, A., Melloni, G.E.M., Kanoni, S., et al. (2024). Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. Nature *627*, 347–357. https://doi.org/10.1038/s41586-024-07019-6.
16. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330. https://doi.org/10.1126/science.aaz1776.
17. Alonso, L., Piron, A., Morán, I., Guindo-Martínez, M., Bonàs-Guarch, S., Atla, G., Miguel-Escalada, I., Royo, R., Puiggròs, M., Garcia-Hurtado, X., et al. (2021). TIGER: The gene expression regulatory variation landscape of human pancreatic islets. Cell Rep. *37*, 109807. https://doi.org/10.1016/j.celrep.2021.109807.
18. Viñuela, A., Varshney, A., van de Bunt, M., Prasad, R.B., Asplund, O., Bennett, A., Boehnke, M., Brown, A.A., Erdos, M.R., Fadista, J., et al. (2020). Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. Nat. Commun. *11*, 4912. https://doi.org/10.1038/s41467-020-18581-8.
19. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al. (2013). The Genotype-Tissue Expression (GTEx) project. Nature Genet. *45*, 580–585. https://doi.org/10.1038/ng.2653.
20. Crouch, D.J.M., Inshaw, J.R.J., Robertson, C.C., Zhang, J.-Y., Chen, W.-M., Onengut-Gumuscu, S., Cutler, A.J., Sidore, C., Cucca, F., Pociot, F., et al. (2022). Enhanced genetic analysis of type 1 diabetes by selecting variants on both effect size and significance, and by integration with autoimmune thyroid disease. Preprint at bioRxiv. https://doi.org/10.1101/2021.02.05.429962.
21. Ferkingstad, E., Sulem, P., Atlason, B.A., Sveinbjornsson, G., Magnusson, M.I., Styrmisdottir, E.L., Gunnarsdottir, K., Helgason, A., Oddsson, A., Halldorsson, B.V., et al. (2021). Large-scale integration of the plasma proteome with genetics and disease. Nat. Genet. *53*, 1712–1721. https://doi.org/10.1038/s41588-021-00978-w.
22. McAllan, L., Baranasic, D., Villicaña, S., Brown, S., Zhang, W., Lehne, B., Adamo, M., Jenkinson, A., Elkalaawy, M., Mohammadi, B., et al. (2023). Integrative genomic analyses in adipocytes implicate DNA methylation in human obesity and diabetes. Nat. Commun. *14*, 2784. https://doi.org/10.1038/s41467-023-38439-z.
23. Kurki, M.I., Karjalainen, J., Palta, P., Sipilä, T.P., Kristiansson, K., Donner, K.M., Reeve, M.P., Laivuori, H., Aavikko, M., Kaunisto, M.A., et al. (2023). FinnGen provides genetic insights from a well-phenotyped isolated population. Nature *613*, 508–518. https://doi.org/10.1038/s41586-022-05473-8.
24. Yengo, L., Sidorenko, J., Kemper, K.E., Zheng, Z., Wood, A.R., Weedon, M.N., Frayling, T.M., Hirschhorn, J., Yang, J., and Visscher, P.M.; GIANT Consortium (2018). Meta-analysis of genome-wide association studies for height and body mass index in ∼700000 individuals of European ancestry. Hum. Mol. Genet. *27*, 3641–3649. https://doi.org/10.1093/hmg/ddy271.
25. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A.D., Wallace, C., and Plagnol, V. (2014). Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. *10*, e1004383. https://doi.org/10.1371/journal.pgen.1004383.
26. Piron, A., Szymczak, F., Papadopoulou, T., Alvelos, M.I., Defrance, M., Lenaerts, T., Eizirik, D.L., and Cnop, M. (2024). RedRibbon: A new rank–rank hypergeometric overlap for gene and transcript expression signatures. Life Sci. Alliance *7*, e202302203. https://doi.org/10.26508/lsa.202302203.
27. Xu, P., Chang, J.C., Zhou, X., Wang, W., Bamkole, M., Wong, E., Bettayeb, K., Jiang, L.L., Huang, T., Luo, W., et al. (2021). GSAP regulates lipid homeostasis and mitochondrial function associated with Alzheimer's disease. J. Exp. Med. *218*, e20202446. https://doi.org/10.1084/jem.20202446.
28. Miguel-Escalada, I., Bonàs-Guarch, S., Cebola, I., Ponsa-Cobas, J., Mendieta-Esteban, J., Atla, G., Javierre, B.M., Rolando, D.M.Y., Farabella, I., Morgan, C.C., et al. (2019). Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. Nat. Genet. *51*, 1137–1148. https://doi.org/10.1038/s41588-019-0457-0.
29. Pasquali, L., Gaulton, K.J., Rodríguez-Seguí, S.A., Mularoni, L., Miguel-Escalada, I., Akerman, İ., Tena, J.J., Morán, I., Gómez-Marín, C., van de Bunt, M., et al. (2014). Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. Nat. Genet. *46*, 136–143. https://doi.org/10.1038/ng.2870.
30. Lyssenko, V., Nagorny, C.L.F., Erdos, M.R., Wierup, N., Jonsson, A., Spégel, P., Bugliani, M., Saxena, R., Fex, M., Pulizzi, N., et al. (2009). Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion. Nat. Genet. *41*, 82–88. https://doi.org/10.1038/ng.288.
31. Sparsø, T., Bonnefond, A., Andersson, E., Bouatia-Naji, N., Holmkvist, J., Wegner, L., Grarup, N., Gjesing, A.P., Banasik, K., Cavalcanti-Proença, C., et al. (2009). G-allele of intronic rs10830963 in MTNR1B confers increased risk of impaired fasting glycemia and type 2 diabetes through an impaired glucose-stimulated insulin release: studies involving 19,605 Europeans. Diabetes *58*, 1450–1456. https://doi.org/10.2337/db08-1660.
32. Zerbino, D.R., Wilder, S.P., Johnson, N., Juettemann, T., and Flicek, P.R. (2015). The ensembl regulatory build. Genome Biol. *16*, 56. https://doi.org/10.1186/s13059-015-0621-5.
33. Marselli, L., Piron, A., Suleiman, M., Colli, M.L., Yi, X., Khamis, A., Carrat, G.R., Rutter, G.A., Bugliani, M., Giusti, L., et al. (2020). Persistent or

Transient Human β Cell Dysfunction Induced by Metabolic Stress: Specific Signatures and Shared Gene Expression with Type 2 Diabetes. Cell Rep. *33*, 108466. https://doi.org/10.1016/j.celrep.2020.108466.

34. American Diabetes Association (2009). Diagnosis and classification of diabetes mellitus. Diabetes Care *32*, S62–S67. https://doi.org/10.2337/dc09-S062.
35. Tian, G., Sågetorp, J., Xu, Y., Shuai, H., Degerman, E., and Tengholm, A. (2012). Role of phosphodiesterases in the shaping of sub-plasma-membrane cAMP oscillations and pulsatile insulin secretion. J. Cell Sci. *125*, 5084–5095. https://doi.org/10.1242/jcs.107201.
36. Jacobs, D.T., Weigert, R., Grode, K.D., Donaldson, J.G., and Cheney, R.E. (2009). Myosin Vc is a molecular motor that functions in secretory granule trafficking. Mol. Biol. Cell *20*, 4471–4488. https://doi.org/10.1091/mbc.e08-08-0865.
37. M Ribeiro, D., Ziyani, C., and Delaneau, O. (2022). Shared regulation and functional relevance of local gene co-expression revealed by single cell analysis. Commun. Biol. *5*, 876. https://doi.org/10.1038/s42003-022-03831-w.
38. Smyth, D.J., Cooper, J.D., Howson, J.M.M., Clarke, P., Downes, K., Mistry, T., Stevens, H., Walker, N.M., and Todd, J.A. (2011). FUT2 Nonsecretor Status Links Type 1 Diabetes Susceptibility and Resistance to Infection. Diabetes *60*, 3081–3084. https://doi.org/10.2337/db11-0638.
39. Yang, P., Li, H.-L., and Wang, C.-Y. (2011). FUT2 Nonfunctional Variant: A "Missing Link" Between Genes and Environment in Type 1 Diabetes? Diabetes *60*, 2685–2687. https://doi.org/10.2337/db11-1104.
40. Fløyel, T., Brorsson, C., Nielsen, L.B., Miani, M., Bang-Berthelsen, C.H., Friedrichsen, M., Overgaard, A.J., Berchtold, L.A., Wiberg, A., Poulsen, P., et al. (2014). CTSH regulates β-cell function and disease progression in newly diagnosed type 1 diabetes patients. Proc. Natl. Acad. Sci. USA *111*, 10305–10310. https://doi.org/10.1073/pnas.1402571111.
41. Ye, J., Stefan-Lifshitz, M., and Tomer, Y. (2021). Genetic and environmental factors regulate the type 1 diabetes gene CTSH via differential DNA methylation. J. Biol. Chem. *296*, 100774. https://doi.org/10.1016/j.jbc.2021.100774.
42. Márquez, A., Kerick, M., Zhernakova, A., Gutierrez-Achury, J., Chen, W.M., Onengut-Gumuscu, S., González-Álvaro, I., Rodriguez-Rodriguez, L., Rios-Fernández, R., González-Gay, M.A., et al. (2018). Meta-analysis of Immunochip data of four autoimmune diseases reveals novel single-disease and cross-phenotype associations. Genome Med. *10*, 97. https://doi.org/10.1186/s13073-018-0604-8.
43. Miyagawa, T., and Tokunaga, K. (2019). Genetics of narcolepsy. Hum. Genome Var. *6*, 4. https://doi.org/10.1038/s41439-018-0033-7.
44. Szymczak, F., Colli, M.L., Mamula, M.J., Evans-Molina, C., and Eizirik, D.L. (2021). Gene expression signatures of target tissues in type 1 diabetes, lupus erythematosus, multiple sclerosis, and rheumatoid arthritis. Sci. Adv. *7*, eabd7600. https://doi.org/10.1126/sciadv.abd7600.
45. Xue, D., Narisu, N., Taylor, D.L., Zhang, M., Grenko, C., Taylor, H.J., Yan, T., Tang, X., Sinha, N., Zhu, J., et al. (2023). Functional interrogation of twenty type 2 diabetes-associated genes using isogenic human embryonic stem cell-derived β-like cells. Cell Metab. *35*, 1897–1914.e11. https://doi.org/10.1016/j.cmet.2023.09.013.
46. Hodson, D.J., Mitchell, R.K., Marselli, L., Pullen, T.J., Gimeno Brias, S., Semplici, F., Everett, K.L., Cooper, D.M.F., Bugliani, M., Marchetti, P., et al. (2014). ADCY5 couples glucose to insulin secretion in human islets. Diabetes *63*, 3009–3021. https://doi.org/10.2337/db13-1607.
47. Li, Y., Miao, Y., Feng, Q., Zhu, W., Chen, Y., Kang, Q., Wang, Z., Lu, F., and Zhang, Q. (2024). Mitochondrial dysfunction and onset of type 2 diabetes along with its complications: a multi-omics Mendelian randomization and colocalization study. Front. Endocrinol. *15*, 1401531. https://doi.org/10.3389/fendo.2024.1401531.
48. Ghanbari, F., Yazdanpanah, N., Yazdanpanah, M., Richards, J.B., and Manousaki, D. (2022). Connecting Genomics and Proteomics to Identify Protein Biomarkers for Adult and Youth-Onset Type 2 Diabetes: A Two-Sample Mendelian Randomization Study. Diabetes *71*, 1324–1337. https://doi.org/10.2337/db21-1046.
49. Petrychenko, V., Yi, S.H., Liedtke, D., Peng, B.Z., Rodnina, M.V., and Fischer, N. (2025). Structural basis for translational control by the human 48S initiation complex. Nat. Struct. Mol. Biol. *32*, 62–72. https://doi.org/10.1038/s41594-024-01378-4.
50. Kiran, S., Mandal, M., Rakib, A., Bajwa, A., and Singh, U.P. (2023). miR-10a-3p modulates adiposity and suppresses adipose inflammation through TGF-β1/Smad3 signaling pathway. Front. Immunol. *14*, 1213415. https://doi.org/10.3389/fimmu.2023.1213415.
51. Lee, S., Cho, Y.K., Kim, H., Choi, C., Kim, S., and Lee, Y.-H. (2024). miR-10a regulates cell death and inflammation in adipose tissue of male mice with diet-induced obesity. Mol. Metab. *90*, 102039. https://doi.org/10.1016/j.molmet.2024.102039.
52. Wallace, C. (2021). A more accurate method for colocalisation analysis allowing for multiple causal variants. PLoS Genet. *17*, e1009440. https://doi.org/10.1371/journal.pgen.1009440.
53. Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A simple new approach to variable selection in regression, with application to genetic fine mapping. J. R. Stat. Soc. Series B Stat. Methodol. *82*, 1273–1300. https://doi.org/10.1111/rssb.12388.
54. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., and Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. *48*, 481–487. https://doi.org/10.1038/ng.3538.
55. Stekelenburg, C., Gerster, K., Blouin, J.L., Lang-Muritano, M., Guipponi, M., Santoni, F., and Schwitzgebel, V.M. (2019). Exome sequencing identifies a de novo FOXA2 variant in a patient with syndromic diabetes. Pediatr. Diabetes *20*, 366–369. https://doi.org/10.1111/pedi.12814.
56. Simaite, D., Kofent, J., Gong, M., Rüschendorf, F., Jia, S., Arn, P., Bentler, K., Ellaway, C., Kühnen, P., Hoffmann, G.F., et al. (2014). Recessive mutations in PCBD1 cause a new type of early-onset diabetes. Diabetes *63*, 3557–3564. https://doi.org/10.2337/db13-1784.
57. Huerta-Chagoya, A., Schroeder, P., Mandla, R., Li, J., Morris, L., Vora, M., Alkanaq, A., Nagy, D., Szczerbinski, L., Madsen, J.G.S., et al. (2024). Rare variant analyses in 51,256 type 2 diabetes cases and 370,487 controls reveal the pathogenicity spectrum of monogenic diabetes genes. Nat. Genet. *56*, 2370–2379. https://doi.org/10.1038/s41588-024-01947-9.
58. Evseeva, I., Nicodemus, K.K., Bonilla, C., Tonks, S., and Bodmer, W.F. (2010). Linkage disequilibrium and age of HLA region SNPs in relation to classic HLA gene alleles within Europe. Eur. J. Hum. Genet. *18*, 924–932. https://doi.org/10.1038/ejhg.2010.32.
59. Robertson, C.C., Inshaw, J.R.J., Onengut-Gumuscu, S., Chen, W.M., Santa Cruz, D.F., Yang, H., Cutler, A.J., Crouch, D.J.M., Farber, E., Bridges, S.L., Jr., et al. (2021). Fine-mapping, trans-ancestral and genomic analyses identify causal variants, cells, genes and drug targets for type 1 diabetes. Nat. Genet. *53*, 962–971. https://doi.org/10.1038/s41588-021-00880-5.
60. Greenawalt, D.M., Dobrin, R., Chudin, E., Hatoum, I.J., Suver, C., Beaulaurier, J., Zhang, B., Castro, V., Zhu, J., Sieberts, S.K., et al. (2011). A survey of the genetics of stomach, liver, and adipose gene expression from a morbidly obese cohort. Genome Res. *21*, 1008–1016. https://doi.org/10.1101/gr.112821.110.
61. Chen, J., Meng, Y., Zhou, J., Zhuo, M., Ling, F., Zhang, Y., Du, H., and Wang, X. (2013). Identifying candidate genes for Type 2 Diabetes Mellitus and obesity through gene expression profiling in multiple tissues or cells. J. Diabetes Res. *2013*, 970435. https://doi.org/10.1155/2013/970435.
62. Sajuthi, S.P., Sharma, N.K., Chou, J.W., Palmer, N.D., McWilliams, D.R., Beal, J., Comeau, M.E., Ma, L., Calles-Escandon, J., Demons, J., et al. (2016). Mapping adipose and muscle tissue expression quantitative trait loci in African Americans to identify genes for type 2 diabetes and obesity. Hum. Genet. *135*, 869–880. https://doi.org/10.1007/s00439-016-1680-8.

63. Taylor, D.L., Jackson, A.U., Narisu, N., Hemani, G., Erdos, M.R., Chines, P.S., Swift, A., Idol, J., Didion, J.P., Welch, R.P., et al. (2019). Integrative analysis of gene expression, DNA methylation, physiological traits, and genetic variation in human skeletal muscle. Proc. Natl. Acad. Sci. USA *116*, 10883–10888. https://doi.org/10.1073/pnas.1814263116.

64. Raulerson, C.K., Ko, A., Kidd, J.C., Currin, K.W., Brotman, S.M., Cannon, M.E., Wu, Y., Spracklen, C.N., Jackson, A.U., Stringham, H.M., et al. (2019). Adipose Tissue Gene Expression Associations Reveal Hundreds of Candidate Genes for Cardiometabolic Traits. Am. J. Hum. Genet. *105*, 773–787. https://doi.org/10.1016/j.ajhg.2019.09.001.

65. Torres, J.M., Abdalla, M., Payne, A., Fernandez-Tajes, J., Thurner, M., Nylander, V., Gloyn, A.L., Mahajan, A., and McCarthy, M.I. (2020). A Multi-omic Integrative Scheme Characterizes Tissues of Action at Loci Associated with Type 2 Diabetes. Am. J. Hum. Genet. *107*, 1011–1028. https://doi.org/10.1016/j.ajhg.2020.10.009.

66. Ravassard, P., Hazhouz, Y., Pechberty, S., Bricout-Neveu, E., Armanet, M., Czernichow, P., and Scharfmann, R. (2011). A genetically engineered human pancreatic β cell line exhibiting glucose-inducible insulin secretion. J. Clin. Investig. *121*, 3589–3597. https://doi.org/10.1172/JCI58447.

67. Costa-Junior, J.M., Coomans de Brachène, A., Musuaya, A.E., Zimath, P.L., Martin-Vazquez, E., Oliveira, J.G., Carpentier, J., Faoro, V., Klass, M., Cnop, M., and Eizirik, D.L. (2025). Exercise-induced meteorin-like protein protects human pancreatic beta cells from cytokine-induced apoptosis. Diabetologia *68*, 1524–1529. https://doi.org/10.1007/s00125-025-06426-2.

68. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., and Peterson, H. (2020). gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. F1000Res. *9*, ELIXIR-709. https://doi.org/10.12688/f1000research.24956.2.

69. Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M.N., and Sergushichev, A. (2021). Fast gene set enrichment analysis. Preprint at bioRxiv. https://doi.org/10.1101/060012.

70. Alvelos, M.I., Szymczak, F., Castela, Â., Marín-Cañas, S., de Souza, B.M., Gkantounas, I., Colli, M., Fantuzzi, F., Cosentino, C., Igoillo-Esteve, M., et al. (2021). A functional genomic approach to identify reference genes for human pancreatic beta cell real-time quantitative RT-PCR analysis. Islets *13*, 51–65. https://doi.org/10.1080/19382014.2021.1948282.

71. Moore, F., Naamane, N., Colli, M.L., Bouckenooghe, T., Ortis, F., Gurzov, E.N., Igoillo-Esteve, M., Mathieu, C., Bontempi, G., Thykjaer, T., et al. (2011). STAT1 is a master regulator of pancreatic {beta}-cell apoptosis and islet inflammation. J. Biol. Chem. *286*, 929–941. https://doi.org/10.1074/jbc.M110.162131.

72. Cunha, D.A., Hekerman, P., Ladrière, L., Bazarra-Castro, A., Ortis, F., Wakeham, M.C., Moore, F., Rasschaert, J., Cardozo, A.K., Bellomo, E., et al. (2008). Initiation and execution of lipotoxic ER stress in pancreatic beta-cells. J. Cell Sci. *121*, 2308–2318. https://doi.org/10.1242/jcs.026062.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Prototype strain of enterovirus, CVB1/Conn-5	ATCC	ATCC VR-28
Chemicals, peptides, and recombinant proteins		
Lipofectamine RNAiMAX Transfection Reagent	Invitrogen	Cat #13778150
Propidium iodide	Sigma-Aldrich	Cat#P4170; CAS: 25535-16-4
Hoechst 33342 (bisBenzimide H 33342 trihydrochloride)	Sigma-Aldrich	Cat#14533; CAS: 875756-97-1
Forskolin	Sigma-Aldrich	Cat#F6886; CAS: **66575-29-9**
Beta Krebs	Human Cell Design Group	Cat#BK-100
SsoAdvanced Universal SYBR Green Supermix	Bio-Rad	CAT#1725274
Critical commercial assays		
Human Insulin ELISA	Mercodia	Cat #10-1113-10
Invitrogen™ PureLink™ RNA Micro Scale Kit	Invitrogen	Cat#12183016
Applied Biosystems™ TaqMan™ Reverse Transcription Reagents	Applied Biosystems	Cat#4304134
Deposited data		
eQTL	Alonso et al.[17]	http://tiger.bsc.es/
pQTL	Ferkingstad et al.[21]	N/A
meQTL	McAllan et al.[22]	N/A
GWAS	Chen et al.,[11] Suzuki et al.,[15] Kurki et al.,[23] Yengo et al.[24]	N/A
Co-localization list, plots and source code	This paper	https://doi.org/10.5281/zenodo.13987433
Experimental models: Cell lines		
EndoC-βH1	Ravassard et al.[66]	RRID:CVCL_L909
Oligonucleotides		
AllStars Negative Control siRNA	Qiagen	Cat #1027281
Human siFUT2, Stealth siRNA	Thermofisher Scientific	ID:HSS103862
Human SiFUT2#2, Silencer Select siRNA	Thermofisher Scientific	ID:S533826
Primer for *FUT2* Forward: AGCGGCTAGCGAA GATTCA	This paper	N/A
Primer for *FUT2* Reverse: GGTCCCAGTGCCT TTGATGT	This paper	N/A
Primer for *ACTB* Forward: CTGTACGCCAACA CAGTGCT	Costa-Junior et al.[67]	N/A
Primer for *ACTB* Reverse: GCTCAGGAGGAGC AATGATC	Costa-Junior et al.[67]	N/A
Primer for *VAPA* Forward: TACCGAAACAAGGA AACTAATGGAA	Costa-Junior et al.[67]	N/A
Primer for *VAPA* Reverse: GCCTTAAACCTTCA TCTCTCAGGT	Costa-Junior et al.[67]	N/A
Software and algorithms		
colocRedRibbon	This paper	https://github.com/antpiron/colocRedRibbon and https://doi.org/10.5281/zenodo.13987433 (version used for this paper)
RedRibbon	Piron et al.[26]	https://github.com/antpiron/RedRibbon

(*Continued on next page*)

Please cite this article in press as: Piron et al., Identification of novel type 1 and type 2 diabetes genes by co-localization of human islet eQTL and GWAS variants with *colocRedRibbon*, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.101004

CellPress
OPEN ACCESS

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
GProfiler2	Kolberg et al.[68]	RRID:SCR_018190 https://cran.r-project.org/web/packages/gprofiler2/index.html
Coloc	Wallace[52]	RRID:SCR_026041 https://cran.r-project.org/web/packages/coloc/index.html
fGSEA	Korotkevitch et al.[69]	RRID:SCR_020938 https://bioconductor.org/packages/release/bioc/html/fgsea.html
GraphPad Prism 9	GraphPad Software	RRID:SCR_002798

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

All human sample data and associated metadata are available from the respective data owners, as detailed in the key resources table section.

METHOD DETAILS

The novel *colocRedRibbon* R package integrates the *coloc* co-localization package[25] with a variant shortlisting method. The shortlisting method is based on *RedRibbon* rank-rank hypergeometric overlap[26] and uses the direction of effect of *GWAS* risk alleles in the eQTL analysis and is described in more detail below.

colocRedRibbon workflow

colocRedRibbon is a tool designed to identify common causal candidates by co-localizing GWAS and eQTL (Figure 1A). This approach aims to pinpoint variants associated both with diabetes risk and gene expression. The method employs a two-step approach for shortlisting variants: a risk allele effect step and a *RedRibbon* overlap step. The risk allele effect step categorizes variants into two disjoint sets based on the direction of effect on gene expression of the disease-risk increasing allele, i.e., down- or upregulating. The upregulating variant set comprises the variants whose risk alleles increase gene expression, and the downregulating set risk alleles that decrease gene expression. The subsequent steps are applied independently on both variant sets.

In the shortlisting step, the *RedRibbon* rank-rank hypergeometric overlap method[26] is applied on GWAS and eQTL variants. This step involves the ranking of variants by *p*-values for both GWAS or eQTLs and statistically examining potential overlaps between the ranked lists. If significant overlap is detected, the shortlisted SNPs are further analyzed by the *coloc* package. In cases where *RedRibbon* overlap is not significant, *coloc* analysis is performed without prior shortlisting.

colocRedRibbon shortlisting

Risk allele effect shortlisting

The shortlisting step leverages a known biological principle to categorize variants into two disjoint subsets. This principle is based on the expectation that risk alleles of multiple variants within a gene region that are in linkage disequilibrium should produce concordant expression effects, either upregulating or downregulating gene expression. Risk alleles of variants are hence split into two sets, one upregulating and the other downregulating a gene. Only one of these two sets is likely to yield meaningful co-localization results, as relevant associations are expected to have consistent directional effects. Compared to analyses that do not include this step, this division typically halves the number of candidate loci, streamlining subsequent analyses.

RedRibbon overlap shortlisting

The *RedRibbon* method[26] ranks GWAS and eQTL variant lists by increasing *p*-value, prioritizing variants with the strongest statical significance. *RedRibbon* aims to uncover the best overlap among variants with the lowest *p*-values in both lists. For each coordinate (i,j) in the two lists, where i represents the GWAS rank and j the eQTL rank, an overlap *p*-value is computed using a one-sided hypergeometric statistic. Let c be the number of variants common to both lists at the coordinate (i,j) in a *RedRibbon* overlap map of size $n \times n$, where n is the number of elements in each list, the overlap *p*-value is calculated as:

$$pval = 1 - CDF_{hyper}(c - 1, i, j, n)$$

where CDF_{hyper} represents the hypergeometric cumulative distribution function. This *p*-value quantifies the probability of observing c or more overlapping variants by chance, given the positions in the ranked lists. *RedRibbon* overlap results in a list of variants corresponding to the minimal *p*-value overlap. This approach allows for a statistical assessment of the overlap between GWAS and eQTL SNPs, facilitating the identification of potential causal variants that influence both disease risk and gene expression.

RedRibbon overlaps shortlisting in interquartile range mode

The list of shortlisted variants can be directly fed into the *coloc* tool. Or, alternatively, the interquartile range (IQR) method can be employed to add background, providing a more comprehensive analysis (Figure 1B). The RedRibbon-shortlisted variants

serve as the core set. The core set is expanded with additional variants based on their chromosomal positions, using as lower bound, $Q_1 - 1.5 * IQR$ and upper bound $Q_2 + 1.5 * IQR$ where Q_1 is the 25th percentile (first quartile) of the chromosomal positions of the core set, Q_2 the 75th percentile (third quartile), and *IQR* is the interquartile range $(Q3 - Q1)$. This method adds a background of variants while excluding potential outliers, providing a broader genomic context and ensuring a more robust set for *coloc* analysis.

Adjusted P-value computation

The *colocRedRibbon* adjusted minimal *p*-value is computed, leveraging the *p*-value adjustment capabilities of the *RedRibbon* package.[26] This approach utilizes a hybrid permutation-prediction method that splits the variants into two sets, the permuted and predicted set. The permuted set undergoes, as the name indicates, random permutation. The predicted set is computed from the permuted set values with the following:

$$P_y = |r|P_x + (1 - |r|)\widehat{P}$$

where the *p*-value P_y for variant *y* is predicted from the *p*-value P_x for variant x, the coefficient of correlation *r* representing the linkage disequilibrium, and a random *p*-value $\widehat{P}$ generated from the distribution of all values in the original list.

The hybrid permutation-prediction of *colocRedRibbon* uses a linkage disequilibrium approximation to compute the adjusted *p*-value. This approximation is given by *r* function, an approximation of the coefficient of correlation between two variants:

$$r(SNP_1, SNP_2) = \frac{half}{half + distance(SNP_1, SNP_2)}$$

where *distance* is the distance in base pairs between the variants under consideration, and *half* the distance between the variants where *r* equals 0.5.

The rationale behind the approximation is to trade-off some of the adjusted *p*-value accuracy for performance, because using real linkage disequilibrium would be computationally expensive as it needs to be computed for all pairs of variants. Subsequent *coloc* analysis is expected to filter out the remaining false positives.

colocRedRibbon R package

colocRedRibbon is available as a user-friendly R package (see data availability section), designed to streamline the process of co-localization analysis. This package seamlessly combines risk allele effect and *RedRibbon* shortlisting steps with the *coloc* package,[25] providing a comprehensive analysis pipeline. It leverages the fast and accurate rank-rank overlap method from *RedRibbon* R package.[26] The speed of *RedRibbon* allows for thousands of co-localizations to be run within hours on standard computer hardware.

Pathway and regulatory elements enrichment

Pathway enrichment was done with *gprofiler2* (v0.1.9) and regulatory element enrichment with the hypergeometric model using Ensembl Regulatory Build GRCh37[32] and human islet chromatin regions from.[28,29] We used *fGSEA* R package[69] to overlap the co-localizing genes with type 2 diabetes differential gene expression from.[33]

Normalization of gene expression

In the gene expression plots, expression was normalized to the geometric mean of stable genes *VAPA*, *ACTB* and *ARF1* (in transcripts per million.[70] For each sample numbered *j*, an adjustment factor β_j is computed with

$$\beta_j = \sqrt[n]{\prod_{i=1}^{n} Expr_{ref_{i,j}}}$$

The expression adjustment is computed for each sample *j* and gene *i* as $AdjExpr_{i,j} = \beta_j * Exp\ r_i$. This normalization was used only for the purpose of visualization of gene expression.

Viral infection and FUT2 knockdown in human beta cells

Cell culture

The human EndoC-βH1 beta cell line,[66] kindly provided by R.Scharfmann (Institut Cochin, Paris), was cultured on Matrigel-fibronectin-coated plates in DMEM with 5.6 mM glucose, 2% fatty acid-free BSA, 50 μM 2-mercaptoethanol, 10 mM nicotinamide, 5.5 μg/mL transferrin, 6.7 ng/mL sodium selenite, and antibiotics.

Viral infection

EndoC-βH1 cells were infected with prototype strain of enterovirus CVB1/Conn-5 (ATCC) at 0.1 multiplicity of infection using medium without FBS. After 2-h adsorption at 37°C, the viral inoculum was removed, cells were washed twice and FBS-containing medium was added to allow viral replication. Cytopathogenic effect was assessed 24 h post-infection.

siRNA transfection

EndoC-βH1 were transfected overnight with 60 nM Allstars Negative Control siRNA (Qiagen, Venlo, The Netherlands) or human FUT2 siRNA (siFUT2, Stealth siRNA, ID:HSS103862, or siFUT2 #2, Silencer Select siRNA, ID:S533826, both from Thermofisher Scientific) using Lipofectamine RNAiMAX (Invitrogen) according to the manufacturer's protocol.[71] After 48–72-h recovery, cells were infected

with CVB1 coxsackieviruses as described above. RNA extraction (Invitrogen) and reverse transcription (Applied Biosystems) were performed following the manufacturer's protocols, and qPCR was carried out as previously described.[67]

Cell viability

Cells were stained with the DNA-binding dyes propidium iodide (5 μg/mL) and Hoechst 33342 (5 μg/mL, Sigma). The number of viable, apoptotic and necrotic cells was assessed in at least 600 cells by two observers, one of whom was blinded to sample identity. Observer agreement exceeded 90%. This method has been validated by comparison with electron microscopy, caspase-3 cleavage and DNA laddering.[72]

Insulin secretion

48 h after siRNA transfection, EndoC-βH1 cells were preincubated for 24 h in DMEM containing 2.8 mM glucose, followed by incubation in glucose-free Krebs solution for 1 h. Cells were exposed in parallel to Krebs solution containing 0 or 20 mM glucose or 20 mM glucose +10 μM forskolin for 1 h. Insulin was measured in cell-free supernatants and acid ethanol–extracted cell lysates by ELISA (Mercodia) and normalized to protein content, determined by Bradford assay.

QUANTIFICATION AND STATISTICAL ANALYSIS

Co-localization

ColocRedRibbon method is fully detailed in the method details section. Overlap significance in *ColocRedRibbon* is defined as an adjusted *p*-value $\leq$0.05, while significance for *coloc* tool is defined as a posterior probability PP.H4.abc $\geq$0.8. To address the non-deterministic nature of RedRibbon[26] (see limitations of the study section), *colocRedRibbon* analyses were repeated up to three times in cases where *coloc* returned a posterior probability 0.8 > PP.H4.abf $\geq$0.6 to ensure result robustness. When minor allele frequencies were not available from QTL analyses, we used the corresponding GWAS minor allele frequencies as proxies. We ran the co-localization over the regions containing at least one significant GWAS (*P*-value $\leq 10^{-8}$) and one significant QTL, using significance thresholds as defined by the data owner or, if not provided, derived from a false discovery rate criterion.

Enrichment analyses

Gene enrichment analysis for type 2 diabetes was performed using the *fGSEA* software with parameters minSize = 15 and maxSize = 1000.[69] Pathway enrichment analysis with *gProfiler2* was carried out using the following data sources: GO, KEGG, REAC, MIRNA, HP, HPA, CORUM, and WP.[68] GProfiler2 vizualisation plots were generated with a custom plotting function, which is available in our zenodo repository number (https://doi.org/10.5281/zenodo.13987433, file src/r-lib/enrichment.R). Variant enrichment in regulatory regions was assessed using a one-sided hypergeometric test, with multiple testing correction applied by controlling for false discovery rate (FDR $\leq$0.05, zenodo file src/r-lib/regulatory.R).

FUT2 statistical analyses

Statistical analyses for *FUT2* (Figure S4) were performed using one-way ANOVA for panels A–B and two-way ANOVA for panel C, followed by Dunnett's post hoc test for pairwise comparisons using GraphPad software. Data are presented as mean ± SEM, with individual points indicating independent experiments.