



**HAL**  
open science

## **Laser capture microdissection of human pancreatic islets reveals novel eQTLs associated with type 2 diabetes**

Amna Khamis, Mickaël Canouil, Afshan Siddiq, Hutokshi Crouch, Mario Falchi, Manon Von Bulow, Florian Ehehalt, Lorella Marselli, Marius Distler, Daniela Richter, et al.

### ► To cite this version:

Amna Khamis, Mickaël Canouil, Afshan Siddiq, Hutokshi Crouch, Mario Falchi, et al.. Laser capture microdissection of human pancreatic islets reveals novel eQTLs associated with type 2 diabetes. *Molecular metabolism*, 2019, 24, pp.98 - 107. <10.1016/j.molmet.2019.03.004>. <hal-03485034>

**HAL Id: hal-03485034**

**<https://hal.science/hal-03485034v1>**

Submitted on 20 Dec 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

## **Laser capture microdissection of human pancreatic islets reveals novel eQTLs associated with type 2 diabetes**

Amna Khamis<sup>1,2</sup>, Mickaël Canouil<sup>2</sup>, Afshan Siddiq<sup>1</sup>, Hutokshi Crouch<sup>1</sup>, Mario Falchi<sup>1</sup>, Manon von Bulow<sup>3</sup>, Florian Eehalt<sup>4,5,6</sup>, Lorella Marselli<sup>7</sup>, Marius Distler<sup>4,5,6</sup>, Daniela Richter<sup>5,6</sup>, Jürgen Weitz<sup>4,5,6</sup>, Krister Bokvist<sup>8</sup>, Ioannis Xenarios<sup>9</sup>, Bernard Thorens<sup>10</sup>, Anke M. Schulte<sup>3</sup>, Mark Ibberson<sup>9</sup>, Amelie Bonnefond<sup>2</sup>, Piero Marchetti<sup>7</sup>, Michele Solimena<sup>5,6</sup> and Philippe Froguel<sup>1,2</sup>.

<sup>1</sup>*Imperial College London, Department of Genomics of Common Disease, London, UK.*

<sup>2</sup>*University of Lille, CNRS, Institute Pasteur de Lille, UMR 8199 - EGID, F-59000 Lille, France.*

<sup>3</sup>*Sanofi-Aventis Deutschland GmbH, Diabetes Research, Frankfurt, Germany*

<sup>4</sup>*Department of Visceral-Thoracic-Vascular Surgery, University Hospital Carl Gustav Carus and Faculty of Medicine, TU Dresden, 01307 Dresden, Germany*

<sup>5</sup>*Paul Langerhans Institute Dresden of the Helmholtz Center Munich at University Hospital Carl Gustav Carus and Faculty of Medicine, TU Dresden, 01307 Dresden, Germany.*

<sup>6</sup>*German Center for Diabetes Research (DZD e.V.), 85764 Neuherberg, Germany*

<sup>7</sup>*University of Pisa, Department of Clinical and Experimental Medicine, Pisa, Italy*

<sup>8</sup>*Lilly Research Laboratories, Eli Lilly, 46285-0001 Indianapolis, IN, USA*

<sup>9</sup>*Vital-IT Group, Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland.*

<sup>10</sup>*Center for Integrative Genomics, University of Lausanne, Genopode Building, Lausanne, 1015, Switzerland.*

**Corresponding author: Philippe Froguel, MD, PhD**

Professor of Genomic Medicine  
Genomics of Common Disease  
Imperial College London  
Burlington-Danes building  
Du Cane Road, London, W12 0NN  
United Kingdom  
Email: [p.froguel@imperial.ac.uk](mailto:p.froguel@imperial.ac.uk)  
Tel: 44 (0)207 594 65 20

## 1. Introduction

Genome-wide association studies (GWAS) meta-analyses have revealed >200 *loci* associated with type 2 diabetes (T2D) and associated traits, such as fasting glucose [1]–[5]. Recently, new statistical methods such as fine-mapping have identified further 40 *loci* [6]. However, deciphering the causal variants and making inferences from GWAS to physiology is still a challenge. Firstly, few GWAS single nucleotide polymorphisms (SNPs) are near insightful biological candidate genes. More importantly, most SNPs fall within the non-coding, regulatory regions of the genome [7], often far from coding regions or in regions where more than one gene lies. To overcome these complexities, combining GWAS with expression data to produce expression quantitative trait loci (eQTLs) has become a powerful tool to shed light into the causative mechanisms of these genetic associations [6], [8]–[10]. However, one limitation is that whilst there is a considerable overlap across human tissue transcriptomes, many eQTLs are cell type specific. Therefore, it is crucial to analyse eQTLs from appropriate tissues, e.g. those functionally involved in the pathology. With regard to diabetes, only a few studies have investigated eQTLs in pancreatic islet samples [11], [12]. However, an insufficient supply of human islets, and technical issues related to the origin of the samples (i.e. from brain-dead organ donors rather than from surgical living patients), have restricted the performance of such studies. Here, we tested a total of 203 islet samples from European subjects with the aim to identify islet cis-eQTLs related to T2D.

## 2. Materials and methods

### 2.1. Cohort

Blood and pancreatic tissue samples were collected from a total of 203 patients from two independent cohorts. The first cohort comprised of 100 islet samples isolated by limited proteolytic digestion of pancreas from brain-dead organ donors (OD), which included 19 subjects with T2D [13]. The second cohort consisted of 103 islet samples isolated by laser capture microdissection (LCM) from the healthy margins of surgical fragments of metabolically phenotyped pancreatectomized patients (PPP), who underwent surgery for different pancreatic diseases [13]. In the PPP cohort, 32 were normoglycaemic (fasting glycaemia <7.0 mmol/l; HbA1C ≤6.5%, glycaemia at 2 h after presurgical oral glucose tolerance test [OGTT] <7.8 mmol/l), 15 had impaired glucose tolerance (IGT, fasting glycaemia <7.0 mmol/l; HbA1C ≤6.5%, OGTT at 2 h of ≥7.8 to

<11.1 mmol/l), 36 had T2D (fasting glycaemia  $\geq 7.0$  mmol/l; HbA1C >6.5%, history of diabetes for >1 year) and 20 had type 3 diabetes (T3D), based on clinical history and pre-surgical laboratory tests according to the American Diabetes Association (ADA) guidelines. T3D was defined as a secondary form of diabetes with an onset not longer than 1 year prior to the appearance of the symptoms related to the primary pancreatic disease. Samples were collected as part of the IMIDIA consortium, and more information on the donors in these cohorts has been recently reported [13].

## **2.2. Islet collection and RNA extraction**

Specific details about the methodologies for islet retrieval and RNA extraction have been previously described [13]. Briefly, RNA was extracted from OD islet samples with the PicoPure RNA Isolation Kit, following the manufacturer's protocol. Islets were extracted using collagenase extraction methods [14]. Islets of PPP were retrieved from cryopreserved surgical samples by LCM with a Zeiss Palm MicroBeam system. RNA was extracted using the Arcturus PicoPure Isolation kit.

## **2.3. Expression microarray**

RNA was analysed for gene expression profiling using the Affymetrix Human Genome U133 Plus 2.0 Array (deposited under the accession number GSE76896), according to the manufacturer's protocol. Data from Gene Expression Omnibus (GEO) and annotation data have been imported using R package *affy*. The array data include 54,210 probe sets covering over 47,000 transcripts and splice variants. Expression data was normalised using the Robust Multichip Analysis (RMA) method as implemented in R package *affy* [15]. Possible batch effect was corrected for using the "ComBat" approach implement in R package *sva* [16], using only the batch information. Differential expression analysis to compare non-diabetic versus T2D samples was performed using *Limma* (Bioconductor R package) with age and gender as covariates. Multiple testing was accounting using Benjamini-Hochberg method to correct p-values. PPP and OD samples clustered separately (Supplementary Figure 4a), but there was no difference in clustering in other factors, such as gender (Supplementary Figure 4b). For PPP, as the HbA1C levels were similar between ND controls and IGT, both samples were categorised as controls and compared to T3D and T2D samples.

## 2.4. DNA extraction and genotyping

A total of 203 DNA samples were extracted from either blood or pancreatic tissue and isolated using the DNeasy Blood & Tissue kit (Qiagen, Germany). All samples were genotyped using the Illumina Omni2.5M array and run on the Illumina iScan platform at the Imperial College London Centre (Hammersmith Campus, Imperial College London, UK). Genotypes were called using the Genome-studio software. Quality-control was performed and SNPs were kept for further analysis according to the following thresholds: 1/ maf > 0.05, 2/ Hardy-Weinberg equilibrium > 0.001 and 3/ call rate > 0.95. Prephasing was performed with Shapelt (v2.r790) and imputation with Impute2 (v2.3.2) using 1,000 genomes panel (phase 3). This led to a dataset of 1,233,520 SNPs and imputation led to a further 7,574,416 SNPs, for a total of >8.7 million SNPs analysed. Principal component analysis (PCA) for ethnicity using genomics data from 1,000 genomes confirmed the vast majority of samples were of European descent (Figure 1A).

## 2.5. eQTL analysis

eQTL analysis was carried out on all 203 samples. The eQTL analysis was performed using FastQTL software [17], with gender and age as covariates. *P*-values were computed using an adaptive permutation pass approach with a permutation number set from 1,000 to 10,000. We performed a genome-wide genotyping and expression analysis in OD and PPP samples to identify eQTLs (Figure 1B). At first, a cis-window was defined as 1 Mb between SNPs and expression probes and then reduced to 500 Kb based on the observed *p*-values distribution (Supplementary Figure 2). All eQTLs with a FDR < 5% were considered to be significant. The eQTL regions were annotated to published binding sites of human islet transcription factors [18], which were subdivided into five classes: C1 (promoters), C2 (inactive enhancers), C3 (active enhancers), C4 (CTCF-bound sites) and C5 (others), which was not CTCF-bound and was not bound to a known histone modification.

## 2.6. Bioinformatics

The Database for Annotation, Visualisation and Integrated Discovery (DAVID) is a software tool that clusters proteins into functional annotations to facilitate understanding of complex datasets [19] (<https://david.ncifcrf.gov/>). The DAVID tool

was used to cluster genes to identify enriched gene ontological functions. *P*-values were calculated using Fisher's exact test to score the enriched gene ontology terms. In addition, the OD and PPP eQTL datasets were exported to Ingenuity Pathways Analysis (IPA) Software (Ingenuity Systems, Incorporated, California, USA; www.ingenuity.com). A core analysis was performed to identify pathways. The pathways were organised according to significance using a Fisher's exact test, with a cut-off of a  $-\log(p\text{-value}) < 1.3$ . A z-score was used to predict downstream regulator activity (activation or inhibition) patterns. For the upstream analysis, a *p*-value  $< 0.01$  was used for relationships between regulators.

### 3. Results

#### 3.1. Identification of cis-eQTLs in human islet samples from two distinct cohorts

Our analysis resulted in the identification of a total of 1,173 and 1,021 significant eQTLs in the samples of OD and PPP, respectively (Supplementary Table 1). We found that only 60% of eQTL genes were shared between the two cohorts and were consistent in the direction of the effect. These differences could conceivably be attributed to differences in the method for islet isolation between the two cohorts. Among the strongest eQTLs shared between OD and PPP were *CHURC1* (OD *p*-value= $1.71 \times 10^{-24}$ ; PPP *p*-value= $3.64 \times 10^{-24}$ ) and *PSPH* (OD *p*-value= $3.92 \times 10^{-26}$ ; PPP *p*-value= $3.64 \times 10^{-24}$ ) (Figure 2A). Churchill domain containing 1 (*CHURC1*) eQTLs have recently been reported to be associated with T2D and obesity in both adipose and muscle tissues [20], [21], suggesting a pleiotropic effect of this gene in T2D pathophysiology. Phosphoserine phosphatase (*PSPH*) catalyses the last step of serine biosynthesis from carbohydrates [22]. Other strong eQTLs unique in either dataset were *ACO1* in PPP (*p* =  $1.47 \times 10^{-25}$ ) and *ELP5* (*p*-value= $5.71 \times 10^{-24}$ ) in OD samples. Aconitase 1 (*ACO1*) is an essential enzyme involved in the TCA cycle and cellular iron homeostasis [23]. *ELP5* is a member of the RNA polymerase II elongator complex, which plays a role in chromatin remodelling and histone acetylation [24].

Of the identified eQTLs in OD and PPP, 42 were transcription factors. This was determined using the TFcheckpoint, a resource for transcription factors that provides a comprehensive list of transcription factor annotations with evidence from the literature [25]. In addition, a total of 46 in OD and 46 in PPP non-coding RNAs, including long non-coding and non-protein coding RNAs, were found to be significantly

affected by nearby SNPs. Of these, 26 were shared between the OD and PPP datasets.

We compared our analysis to published eQTLs in islets from OD: one study in 89 OD identified 616 eQTLs [11]; another study in 118 OD found 2,341 eQTLs [12]. While the reproducibility rate between the two studies was 43% [12], we found a 28% (OD) and 29% (PPP) shared genes with van de Bunt et al. [12] and 23% (OD) and 18% (PPP) with Fadista et al [11]. Of these, 24 genes were shared between our OD and PPP samples and those in these previous datasets, consistent in direction of effect (Table 1; Figure 2B). This low number is likely due to substantial differences in RNA extraction and transcriptomic methods among these studies. For instance, we used microarrays for transcriptomic analysis, whereas the other studies relied on RNA sequencing. Nevertheless, 42% of the shared genes were among the top 100 identified eQTLs in our study, including *CHURC1*, *ERAP2*, and *LDHC*.

### **3.2. Regulatory and biological interpretations of eQTL genes**

Consistent with these regions being regulatory, we annotated the eQTL datasets with active open chromatin maps in islets (promoters, active enhancers and CTCF regions) [15] and found 102 and 93 sites overlapping with open chromatin regions in OD and PPP, respectively (Figure 2C, Supplementary Table 2-3). Of these, 15 were shared between the two datasets, with the majority being in active enhancer sites and clusters. One shared gene relevant for beta-cell biology was *RPH3AL/NOC2*, which has been shown to be crucial for exocytosis of insulin in pancreatic beta cells [26].

The gene ontology (GO) tool was used to make inferences on the biological function of the eQTL genes. GO identified ontologies that relate to beta cell function included cell junction, positive regulation of GTPase activity and glutathione metabolism (Supplementary Figure 3). In addition, to make inferences about the identified eQTL genes, we utilised tools in the Ingenuity Pathway Analysis (IPA) software. None of the top pathways was shared between OD and PPP samples, pointing to differences in the donors and the procedure for islet isolation (Supplementary Figure 4). Interestingly, the transcriptional regulator of pancreatic beta-cell function and maturity-onset diabetes of the young (MODY4) gene *HNF4A* was found to be the most significant upstream regulator in the IPA analysis for both PPP ( $p$ -value=0.003; z-score=0.45) and OD ( $p$ -value=0.0001; z-score=0.98) (Figure 3). *HNF4A* modulates regulatory elements in the promoters and enhancers of genes involved in glucose, fatty acid and cholesterol

metabolism. It regulates gene expression in pancreatic beta cells to achieve glucose homeostasis and activates insulin genes both directly and indirectly. Our data show that many genes have a relevance for carbohydrate and lipid metabolism in addition to their association with T2D.

### 3.3. eQTL genes associated with T2D and associated traits

We then focused our analysis on GWAS *loci* and genes for T2D and associated traits. First, we found an enrichment for GWAS *loci* in eQTLs in MAGIC fasting glucose (OD  $p = 4.8 \times 10^{-4}$ ; PPP  $p = 4.1 \times 10^{-4}$ ), fasting insulin (OD  $p = 0.013$ ; PPP  $p = 0.059$ ), HOMA-IR (OD  $p = 0.019$ ; PPP  $p = 0.018$ ) and HOMA-B (OD  $p = 0.001$ ; PPP  $p = 0.0025$ ) and DIAGRAM (OD  $p = 0.0039$ ; PPP  $p = 3.7 \times 10^{-4}$ ) using a Fisher's exact test (Supplementary Table 4). We then investigated whether T2D genes were also identified in our eQTL study and found a total of eleven genes (Table 2). Seven of these overlapped GWAS genes associated with T2D alone (*AP3S2*, *GPSM1*, *MAEA*, *TLE4*, *TMEM163*, *SSR1*, and *UBE2Z*), three with T2D associated traits alone (*IGF1R* for fasting plasma glucose, *ABO* for disposition index and *UHRF1BP1* for fasting plasma insulin), and one for both T2D and fasting glucose levels (*ADCY5*). Of these genes, *AP3S2* and *IGF1R* were represented in both our OD and PPP datasets. *AP3S2*, encoding for an adaptor protein, was the only gene replicated in both our datasets in addition to reported eQTLs in islets [11], [12]. *ADCY5* is a member of the adenylate cyclase family, which has previously been shown in a knockout study to be involved in glucose stimulation and insulin secretion [27]. Interestingly, we have previously demonstrated that *SSR1* expression is enriched in human islets and isolated beta cells [28]. In addition, for both OD and PPP datasets, we found strong associations with *UBE2Z*, a ubiquitin-conjugating enzyme. Although this gene has not been previously been studied for its role in beta cells, *UBE2Z* was also reported to be associated with coronary artery disease in different populations [29]–[31]. *UHRF1BP1* was the only gene that was annotated to an active enhancer in pancreatic islets. We and others from the GIANT, MAGIC and ExomeBP consortia, recently identified 40 additional *loci* associated with T2D [6] and found that *TPCN2*, a gene encoding for a transmembrane ion channel with an effect on insulin action, also has an eQTL in both PPP and OD. *TPCN2*<sup>-/-</sup> mice display improved insulin sensitivity, which is in agreement with our

findings that an increased expression of *TPCN2* associated with the rs72928978 reported SNP [6].

We then assessed whether our eQTL SNPs were in linkage disequilibrium with GWAS reported SNPs for T2D and associated traits using the GWAS catalogue (<https://www.ebi.ac.uk/gwas/>) and a recently published GWAS [6]. The most significant *loci* in the PPP and OD datasets are shown in Table 3. Notably, approximately 60% did not implicate the nearest gene. This includes the *CENTD2* (*ARAP1*) locus, which a recent study confirmed to be associated with T2D risk through the reduced expression of the nearby *STARD10* gene [32]. In addition, one of the most significant eQTLs we found was the rs34569841 SNP, which is associated with decreased expression of the long non-coding RNA LOC101927636. This SNP is in strong linkage disequilibrium with the previously reported rs13134327 SNP ( $R^2=1$ ), which is associated with HbA1C and was mapped to the *FREM3* gene [33].

### **3.4. eQTL genes differentially expressed with T2D and HbA1C levels**

Our eQTL analysis further identified in the PPP dataset only 21 genes (Table 4), which were also differentially expressed, 11 of which also consistently in their directionality. Among them was *SCTR*, which encodes for the secretin receptor and is the most potent regulator for bicarbonate in the pancreas. Although early studies showed that *SCTR* is linked to insulin release in humans (Lerner and Porte 1970), this association has not since been re-investigated. eQTLs in the mitochondrial Acid Phosphatase 6 Lysophosphatidic *ACP6* have been associated with T2D and BMI [34]. Three *loci*, namely *KLHDC10*, *VAMP1*, and *CFLAR* lie within active islet enhancer regions. Among them, however, only *KLHDC10*, which is involved in the activation of oxidative stress [35], was consistent in its directional effect between eQTL and differential expression. In addition, of the differentially expressed loci, only two were transcription factors (*ZNF117* and *ONECUT2*). Lastly, to identify genes whose expression was altered by glycaemia, we correlated gene expression to HbA1C levels, which is a measure of long-term glycaemia in the PPP dataset. This revealed 170 differentially expressed genes (Supplementary Table 5), of which only four had an eQTL: *CYP4V2* (cytochrome P450 4V2), *TSEN2* (tRNA-splicing endonuclease subunit 2), *ZNF556* (zinc finger protein 556) and *ACVR1B* (activin A receptor type 1B), of which *CYP4V2* and *TSEN2* were consistent in directional effect.

#### 4. Discussion

Here we show the most comprehensive cis-eQTL analysis in relevant islet samples in two distinct cohorts, including samples from 100 OD and for the first time 103 PPP (e.g. from living surgical patients). To our knowledge, this is also the largest eQTL study using islet samples to date. Our thorough cohorts allowed us to identify eQTLs with a relevance to T2D and HbA1C levels. These genes have diverse roles within the cell and further studies need to be performed to understand their roles within the pancreatic beta cell (Figure 4). One main finding was the difference in results observed between our two cohorts. We found that the PPP and OD samples cluster separately, suggesting that the differences in transcriptomic signatures between the two cohorts is primarily due in part to the islet isolation methods. Indeed, we recently compared the transcriptomes of PPP and OD islets and found that those of islets extracted from the same ODs using either LCM or enzymatic digestion clustered separately, with LCM islet transcriptomes clustering with the transcriptomes of LCM islets from PPP [13]. These differences were also highlighted in our pathway analysis using IPA. In addition, these differences may be attributed to stressful conditions while OD patients were in intensive-care prior to brain-death, which may increase inflammatory markers. In contrast, for LCM samples, the surgical specimen are subject to immediate cryofixation, thereby limiting transcriptomic changes. As all previous eQTL and transcriptomic investigations have been performed using islets isolated by enzymatic digestion from OD [11], [12], [36], our study including LCM islets from PPP is unique. In addition to isolation methods, methodological differences in RNA expression analyses likely contribute to the low number of shared genes between our study and previous eQTL analyses in islets [11], [12]. The reason for the use of microarrays in our study is that at the time these studies were conceived and undertaken, RNA-seq procedures (based on next-generation sequencing) were still very expensive and not yet fully consolidated. However, one robust gene that survived these technical and methodological differences and which was replicated amongst all studies is *ZFP57*. Loss-of-function mutations in the *ZFP57* transcription factor is associated with an imprinting disorder, which includes transient neonatal diabetes [37], [38]. *ZFP57* is expressed in undifferentiated cells and downregulated during cell differentiation [39] and is a crucial regulator involved in DNA methylation during development [40]. It would be interesting to further investigate whether DNA methylation changes within this gene has a role in the cell biology of adult beta cells.

In our comparison of GWAS *loci* and eQTLs, we were able to provide an up-to-date analysis of approximately 300 identified genes linked to T2D and associated traits. An intriguing finding was that long non-coding RNAs were found to be the most significantly associated with GWAS SNPs, which includes the long non-coding RNA LOC101927636 *locus*, which in GWAS studies was previously assigned to *FREM3* [33]. This is in addition to approximately 66 non-coding RNAs identified between the OD and PPP datasets. Accumulating literature has shown that SNPs in long non-coding RNAs are associated with human diseases, including T2D, hence highlighting their role as master regulators [41], [42].

It is important to note that islets include a number of cell-types and although we have considered normalising to housekeeping genes identified from single-cell transcriptomic analysis in islets [43], [44], these genes were highly variable among our samples. While it is important to bear in mind that some of the eQTLs identified are not beta-cell specific, there is substantial data to gain from studying the islet as an intact microorgan.

In conclusion, we provide a catalogue of cis-eQTLs from the largest to-date sample size from two separate cohorts of non-diabetic and T2D subjects and which includes the first in-depth combined genetic and expression analysis of human islets isolated by LCM. Ultimately, the knowledge gained from eQTL approaches provides a preferable and more accurate 3-dimensional representation than those from DNA variants alone. *In vitro* and functional analyses are required to definitively prove the role of newly identified genes in relation to islet cell biology and T2D disease.

## Acknowledgements

This study was supported by non-profit organizations and public bodies for funding of scientific research conducted in France and within the European Union: “Centre National de la Recherche Scientifique”, “Université de Lille 2”, “Institut Pasteur de Lille”, “Société Francophone du Diabète”, “Contrat de Plan Etat-Région”, “Agence Nationale de la Recherche”, ANR-10-LABX-46, ANR EQUIPEX Ligan MP: ANR-10-EQPX-07-01, European Research Council GEPIDIAB – 294785; the German Center for Diabetes Research (DZD e.V.), which is supported by the German Ministry for Education and Research (BMBF); Innovative Medicines Initiative Joint Undertaking (grant no. 155005) for IMIDIA, which received financial contributions from the European Union’s Seventh Framework Program (FP7/2007–2013); companies belonging to the European Federation of Pharmaceutical Industries and Associations; Innovative Medicines Initiative 2 Joint Undertaking under grant agreements n° 155005 (IMIDIA), n° 115881 (RHAPSODY) and n° 115797 (INNODIA), which include financial contributions from European Union’s Seventh Framework Programme (FP7/2007-2013) and Horizon 2020 research and innovation programme and EFPIA. This work is also supported by

the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 16.0097.

- [1] B. F. Voight *et al.*, 'Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis', *Nat. Genet.*, vol. 42, no. 7, pp. 579–589, Jul. 2010.
- [2] DIABetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium *et al.*, 'Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility', *Nat. Genet.*, vol. 46, no. 3, pp. 234–244, Mar. 2014.
- [3] R. A. Scott *et al.*, 'An expanded genome-wide association study of type 2 diabetes in Europeans', *Diabetes*, vol. 66, no. 11, pp. 2888–2902, Nov. 2017.
- [4] J. Dupuis *et al.*, 'New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk', *Nat. Genet.*, vol. 42, no. 2, pp. 105–116, Feb. 2010.
- [5] E. Ingelsson and M. I. McCarthy, 'Human Genetics of Obesity and Type 2 Diabetes Mellitus: Past, Present, and Future', *Circ. Genomic Precis. Med.*, vol. 11, no. 6, p. e002090, Jun. 2018.
- [6] A. Mahajan *et al.*, 'Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes', *Nat. Genet.*, vol. 50, no. 4, pp. 559–571, Apr. 2018.
- [7] A. Varshney *et al.*, 'Genetic regulatory signatures underlying islet gene expression and type 2 diabetes', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 114, no. 9, pp. 2301–2306, 28 2017.
- [8] Q. Li *et al.*, 'A novel eQTL-based analysis reveals the biology of breast cancer risk loci', *Cell*, vol. 152, no. 3, pp. 633–641, Jan. 2013.
- [9] M. E. Adriaens and C. R. Bezzina, 'Genomic approaches for the elucidation of genes and gene networks underlying cardiovascular traits', *Biophys. Rev.*, vol. 10, no. 4, pp. 1053–1060, Aug. 2018.
- [10] D. C. Croteau-Chonka *et al.*, 'Expression Quantitative Trait Loci Information Improves Predictive Modeling of Disease Relevance of Non-Coding Genetic Variation', *PLoS One*, vol. 10, no. 10, p. e0140758, 2015.
- [11] J. Fadista *et al.*, 'Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism', *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 38, pp. 13924–13929, Sep. 2014.
- [12] M. van de Bunt *et al.*, 'Transcript Expression Data from Human Islets Links Regulatory Signals from Genome-Wide Association Studies for Type 2 Diabetes and Glycemic Traits to Their Downstream Effectors', *PLoS Genet.*, vol. 11, no. 12, Dec. 2015.
- [13] M. Solimena *et al.*, 'Systems biology of the IMIDIA biobank from organ donors and pancreatectomised patients defines a novel transcriptomic signature of islets from individuals with type 2 diabetes', *Diabetologia*, vol. 61, no. 3, pp. 641–657, 2018.
- [14] P. Marchetti *et al.*, 'The endoplasmic reticulum in pancreatic beta cells of type 2 diabetes patients', *Diabetologia*, vol. 50, no. 12, pp. 2486–2494, Dec. 2007.
- [15] L. Gautier, L. Cope, B. M. Bolstad, and R. A. Irizarry, 'affy--analysis of Affymetrix GeneChip data at the probe level', *Bioinforma. Oxf. Engl.*, vol. 20, no. 3, pp. 307–315, Feb. 2004.
- [16] J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, and J. D. Storey, 'The sva package for removing batch effects and other unwanted variation in high-throughput experiments', *Bioinforma. Oxf. Engl.*, vol. 28, no. 6, pp. 882–883, Mar. 2012.

- [17] H. Ongen, A. Buil, A. A. Brown, E. T. Dermitzakis, and O. Delaneau, 'Fast and efficient QTL mapper for thousands of molecular phenotypes', *Bioinformatics*, vol. 32, no. 10, pp. 1479–1485, May 2016.
- [18] L. Pasquali *et al.*, 'Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants', *Nat. Genet.*, vol. 46, no. 2, pp. 136–143, Feb. 2014.
- [19] D. W. Huang, B. T. Sherman, and R. A. Lempicki, 'Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources', *Nat. Protoc.*, vol. 4, no. 1, pp. 44–57, Jan. 2009.
- [20] S. P. Sajuthi *et al.*, 'Mapping Adipose and Muscle Tissue Expression Quantitative Trait Loci in African Americans to Identify Genes for Type 2 Diabetes and Obesity', *Hum. Genet.*, vol. 135, no. 8, pp. 869–880, Aug. 2016.
- [21] C. A. Glastonbury *et al.*, 'Adiposity-Dependent Regulatory Effects on Multi-tissue Transcriptomes', *Am. J. Hum. Genet.*, vol. 99, no. 3, pp. 567–579, Sep. 2016.
- [22] L. Sun *et al.*, 'cMyc-mediated activation of serine biosynthesis pathway is critical for cancer progression under nutrient deprivation conditions', *Cell Res.*, vol. 25, no. 4, pp. 429–444, Apr. 2015.
- [23] M. Moreno, F. Ortega, G. Xifra, W. Ricart, J. M. Fernández-Real, and J. M. Moreno-Navarrete, 'Cytosolic aconitase activity sustains adipogenic capacity of adipose tissue connecting iron metabolism and adipogenesis', *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.*, vol. 29, no. 4, pp. 1529–1539, Apr. 2015.
- [24] L. M. Belalcazar *et al.*, 'A common variant in the CLDN7/ELP5 locus predicts adiponectin change with lifestyle intervention and improved fitness in obese individuals with diabetes', *Physiol. Genomics*, vol. 47, no. 6, pp. 215–224, Jun. 2015.
- [25] K. Chawla, S. Tripathi, L. Thommesen, A. Lægreid, and M. Kuiper, 'TFcheckpoint: a curated compendium of specific DNA-binding RNA polymerase II transcription factors', *Bioinforma. Oxf. Engl.*, vol. 29, no. 19, pp. 2519–2520, Oct. 2013.
- [26] K. Matsunaga, M. Taoka, T. Isobe, and T. Izumi, 'Rab2a and Rab27a cooperatively regulate the transition from granule maturation to exocytosis through the dual effector Noc2', *J. Cell Sci.*, vol. 130, no. 3, pp. 541–550, 01 2017.
- [27] D. J. Hodson *et al.*, 'ADCY5 couples glucose to insulin secretion in human islets', *Diabetes*, vol. 63, no. 9, pp. 3009–3021, Sep. 2014.
- [28] F. K. Ndiaye *et al.*, 'Expression and functional assessment of candidate type 2 diabetes susceptibility genes identify four new genes contributing to human insulin secretion', *Mol. Metab.*, vol. 6, no. 6, pp. 459–470, Apr. 2017.
- [29] Y. Guo *et al.*, 'Genome-Wide Linkage Analysis of Large Multiple Multigenerational Families Identifies Novel Genetic Loci for Coronary Artery Disease', *Sci. Rep.*, vol. 7, no. 1, p. 5472, Jul. 2017.
- [30] M. Bastami *et al.*, 'Evaluating the association of common UBE2Z variants with coronary artery disease in an Iranian population', *Cell. Mol. Biol. Noisy--Gd. Fr.*, vol. 61, no. 7, pp. 50–54, Nov. 2015.
- [31] D. Lu *et al.*, 'Rs46522 in the Ubiquitin-Conjugating Enzyme E2Z Gene Is Associated with the Risk of Coronary Artery Disease in Individuals of Chinese Han Population with Type 2 Diabetes', *J. Diabetes Res.*, vol. 2017, 2017.
- [32] G. R. Carrat *et al.*, 'Decreased STARD10 Expression Is Associated with Defective Insulin Secretion in Humans and Mice', *Am. J. Hum. Genet.*, vol. 100, no. 2, pp. 238–256, Feb. 2017.

- [33] E. Wheeler *et al.*, 'Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis', *PLOS Med.*, vol. 14, no. 9, p. e1002383, Sep. 2017.
- [34] Sun *et al.*, 'Perturbed Mitochondrial Metabolism in Islets from Donors with Type-2 Diabetes | bioRxiv', 2017. [Online]. Available: <https://www.biorxiv.org/content/early/2017/11/01/212548>. [Accessed: 21-Aug-2018].
- [35] Y. Sekine *et al.*, 'The Kelch repeat protein KLHDC10 regulates oxidative stress-induced ASK1 activation by suppressing PP5', *Mol. Cell*, vol. 48, no. 5, pp. 692–704, Dec. 2012.
- [36] J. M. Locke, G. Hysenaj, A. R. Wood, M. N. Weedon, and L. W. Harries, 'Targeted allelic expression profiling in human islets identifies cis-regulatory effects for multiple variants identified by type 2 diabetes genome-wide association studies', *Diabetes*, vol. 64, no. 4, pp. 1484–1491, Apr. 2015.
- [37] S. E. Boonen *et al.*, 'Transient neonatal diabetes, ZFP57, and hypomethylation of multiple imprinted loci: a detailed follow-up', *Diabetes Care*, vol. 36, no. 3, pp. 505–512, Mar. 2013.
- [38] M. Bak *et al.*, 'Genome-wide DNA methylation analysis of transient neonatal diabetes type 1 patients with mutations in ZFP57', *BMC Med. Genet.*, vol. 17, p. 29, Apr. 2016.
- [39] T. Akagi *et al.*, 'Identification of Zfp-57 as a downstream molecule of STAT3 and Oct-3/4 in embryonic stem cells', *Biochem. Biophys. Res. Commun.*, vol. 331, no. 1, pp. 23–30, May 2005.
- [40] S. Quenneville *et al.*, 'In Embryonic Stem Cells, ZFP57/KAP1 Recognize a Methylated Hexanucleotide to Affect Chromatin and DNA Methylation of Imprinting Control Regions', *Mol. Cell*, vol. 44, no. 3, pp. 361–372, Nov. 2011.
- [41] S. Jain, N. Thakkar, J. Chhatai, M. Pal Bhadra, and U. Bhadra, 'Long non-coding RNA: Functional agent for disease traits', *RNA Biol.*, vol. 14, no. 5, pp. 522–535, 04 2017.
- [42] N. Goyal, D. Kesharwani, and M. Datta, 'Lnc-ing non-coding RNAs with metabolism and diabetes: roles of lncRNAs', *Cell. Mol. Life Sci.*, vol. 75, no. 10, pp. 1827–1837, May 2018.
- [43] Å. Segerstolpe *et al.*, 'Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes', *Cell Metab.*, vol. 24, no. 4, pp. 593–607, Oct. 2016.
- [44] M. J. Muraro *et al.*, 'A Single-Cell Transcriptome Atlas of the Human Pancreas', *Cell Syst.*, vol. 3, no. 4, pp. 385-394.e3, 26 2016.

Table 1: A list of 24 genes shared among OD, PPP Fadista et al. 2014 and Van de Bunt et al., 2015

Gene	Gene Name	OD p-value	PPP p value
CHURC1	Churchill Domain Containing 1	1.7E-24	3.6E-24
ERAP2	Endoplasmic Reticulum Aminopeptidase 2	5.8E-23	5.6E-12
FAM118A	Family With Sequence Similarity 118 Member A	3.9E-22	1.9E-23
POMZP3	POM121 And ZP3 Fusion	4.7E-21	6.4E-13
LDHC	Lactate Dehydrogenase C	8.6E-18	1.7E-10
C17orf97	Chromosome 17 Open Reading Frame 97	9.8E-17	4.0E-14
UBE2U	Ubiquitin Conjugating Enzyme E2 U (Putative)	8.5E-16	9.9E-10
ITGB3BP	Integrin Subunit Beta 3 Binding Protein	1.2E-13	5.2E-18
ZFP57	ZFP57 Zinc Finger Protein	7.6E-11	6.6E-17
DDX11	DEAD/H-Box Helicase 11	2.1E-10	4.5E-14
THNSL2	Threonine Synthase Like 2	1.5E-08	6.5E-11
TDRD5	Tudor Domain Containing 5	3.3E-08	6.0E-06

Table 2: GWAS T2D and glycaemic genes identified in eQTL study

Gene	Gene Name	Disease/Trait	PPP	OD	Fadista et al., 2014	Van de Bunt et al., 2015
ABO	ABO, Alpha 1-3-N-Acetylgalactosaminyltransferase And Alpha 1-3-Galactosyltransferase	Disposition index and soluble E-selectin levels	Y	Y	Y	
ADCY5	Adenylate Cyclase 5	Type 2 diabetes; fasting glucose-related traits	Y			Y
AP3S2	Adaptor Related Protein Complex 3 Sigma 2 Subunit	Type 2 diabetes	Y	Y	Y	Y
GPSM1	G Protein Signaling Modulator 1	Type 2 diabetes	Y			
IGF1R	Insulin Like Growth Factor 1 Receptor	Fasting plasma glucose	Y	Y		
MAEA	Macrophage Erythroblast Attacher	Type 2 diabetes	Y			
SSR1	Signal Sequence Receptor Subunit 1	Type 2 diabetes		Y		
TLE4	Transducin Like Enhancer Of Split 4	Type 2 diabetes		Y		
TPCN2	Two Pore Segment Channel 2	Type 2 diabetes	Y	Y		Y
TMEM163	Transmembrane Protein 163	Type 2 diabetes		Y		Y
UBE2Z	Ubiquitin Conjugating Enzyme E2 Z	Type 2 diabetes	Y	Y		
UHRF1BP1	UHRF1 Binding Protein 1	Fasting insulin-related traits		Y		

Y=Yes

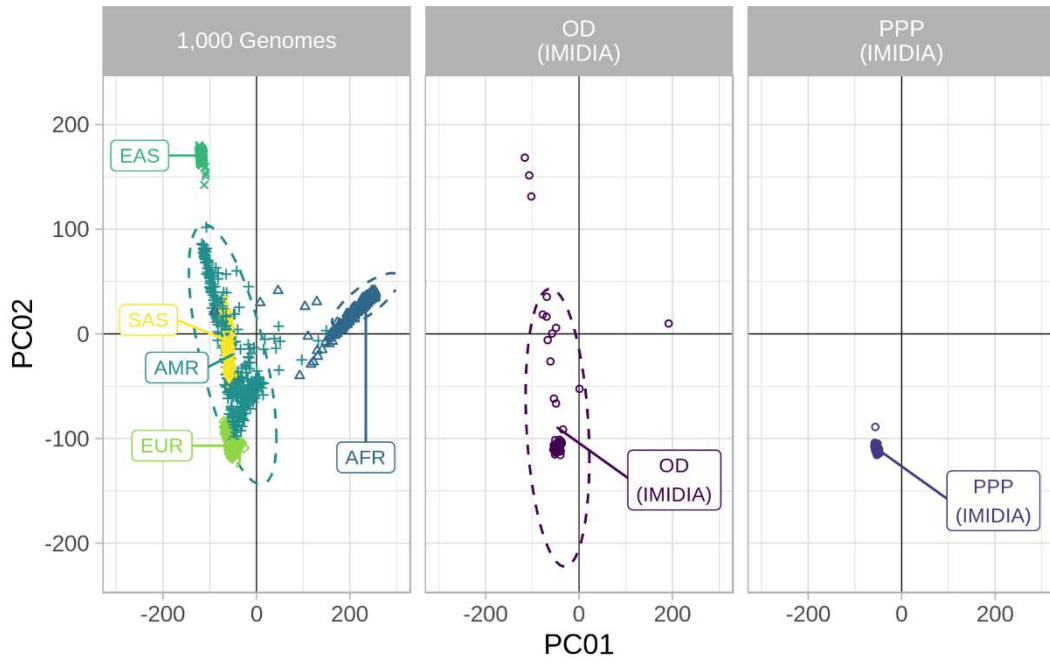
Table 3: Top GWAS T2D and associated traits loci co-localising with eQTLs

GWAS Catalog						
PPP						
GWAS gene	SNPS	eQTL SNP	eQTL distance	eQTL FDR	eQTL beta	Probe ID
FREM3	rs13134327	rs5015757	-174187	9.2E-15	-0.96	LOC101927636
UBE2Z	rs12453394	rs318092	1381	1.2E-14	-0.93	UBE2Z
HLA-DQA1	rs9271774	rs9271770	48612	6.7E-11	1.2	LOC100996809
OD						
SSR1	rs9505118	rs3087986	1363	2.6E-12	0.35	SSR1
UBE2Z	rs12453394	rs3744608	3813	4.1E-12	-0.77	UBE2Z
BRAF	rs9648716	rs28529157	81058	2E-08	-0.45	BRAF
Mahajan et al., 2018						
PPP						
GWAS gene	rs ID	eQTL SNP	eQTL distance	eQTL FDR	eQTL beta	Probe ID
TTL6	rs2032844	rs11657371	-145547	1.8E-05	-0.64	UBE2Z
MACF1	rs2296172	rs61779279	287263	0.0010	-0.2	MACF1
MLX	rs665268	rs646123	-114000	0.0014	0.24	CNTNAP1
OD						
KIF9	rs2276853	rs2276854	-47481	0.0009	-0.32	KLHL18
CENTD2	rs56200889	rs12575364	-56695	0.0009	0.29	STARD10
TTL6	rs2032844	rs11657371	-145547	0.0014	-0.41	UBE2Z

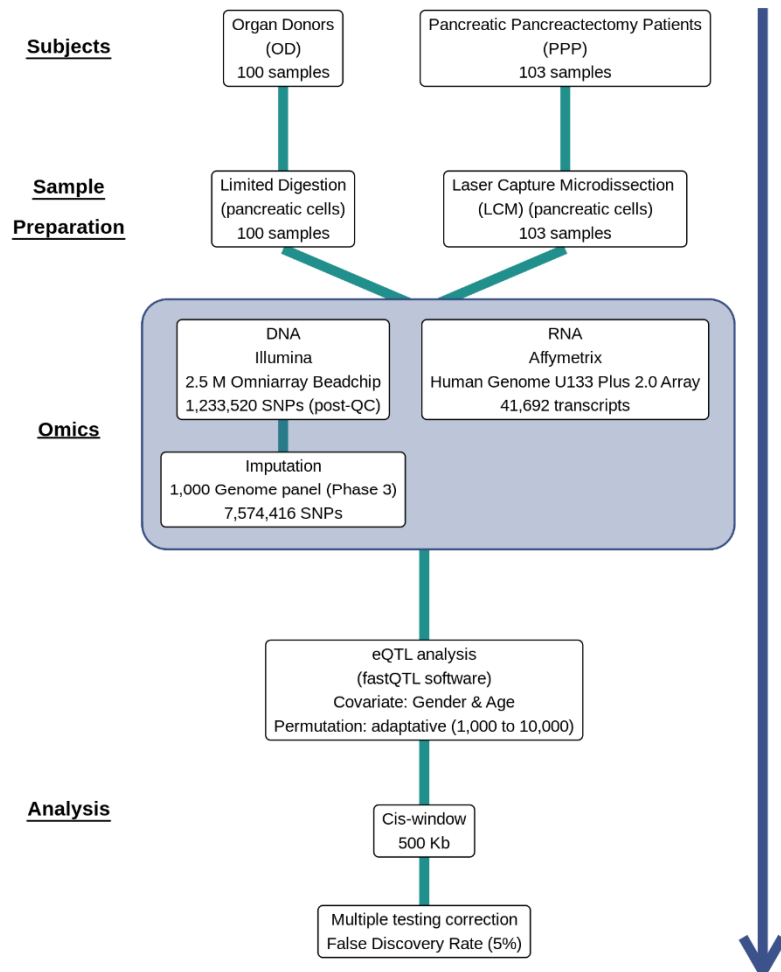
Table 4: Significant eQTL genes also differentially expressed in T2D compared to controls in PPP dataset

PPP						
Gene	SNP ID	distance	eQTL FDR	logFC	FDR	Direction of eQTL effect
SCTR	rs4344946	-15760	0.023	-0.78	0.03	-
CYP2U1	rs141713207	-22383	0.002	-0.55	0.01	+
CYP4V2	rs10866290	757	6.38E-11	-0.5	0.05	-
CASR	rs3749203	489	0.001	-0.45	0.04	+
ACP6	rs28700004	5788	7.94E-08	-0.41	0.01	-
LOC79160	rs3761133	653	0.015	-0.39	0.03	+
IGF1R	rs12591122	305650	3.08E-05	-0.34	0.04	+
LINC00667	rs34361006	49303	0.032	-0.31	0.02	-
ACVR1B	rs2854464	43389	0.008	-0.28	0.02	+
KLHDC10	rs4443587	68107	1.58E-10	-0.25	0.01	-
VAMP1	rs11613996	22444	0.016	-0.25	0.02	+
ELP5	rs222843	-9867	1.40E-15	-0.25	0.03	+
RASA3	rs9525230	80306	4.37E-07	-0.22	0.04	+
H2AFV	rs13245012	-669	0.002	-0.22	0.04	-
METTL15	rs4614434	-52250	0.027	-0.15	0.05	-
RNF213	rs35627722	15051	0.01	0.44	0.01	-
CAST	kgp11684924	12272	4.62E-07	0.46	0.02	+
ERAP1	kgp3909205	15984	3.09E-12	0.49	0.03	+
ZNF117	rs10262238	121749	0.015	0.5	0.02	+
CFLAR	rs10184098	9878	0.001	0.63	0.03	-
ONECUT2	rs514250	32118	0.008	0.71	0.04	+

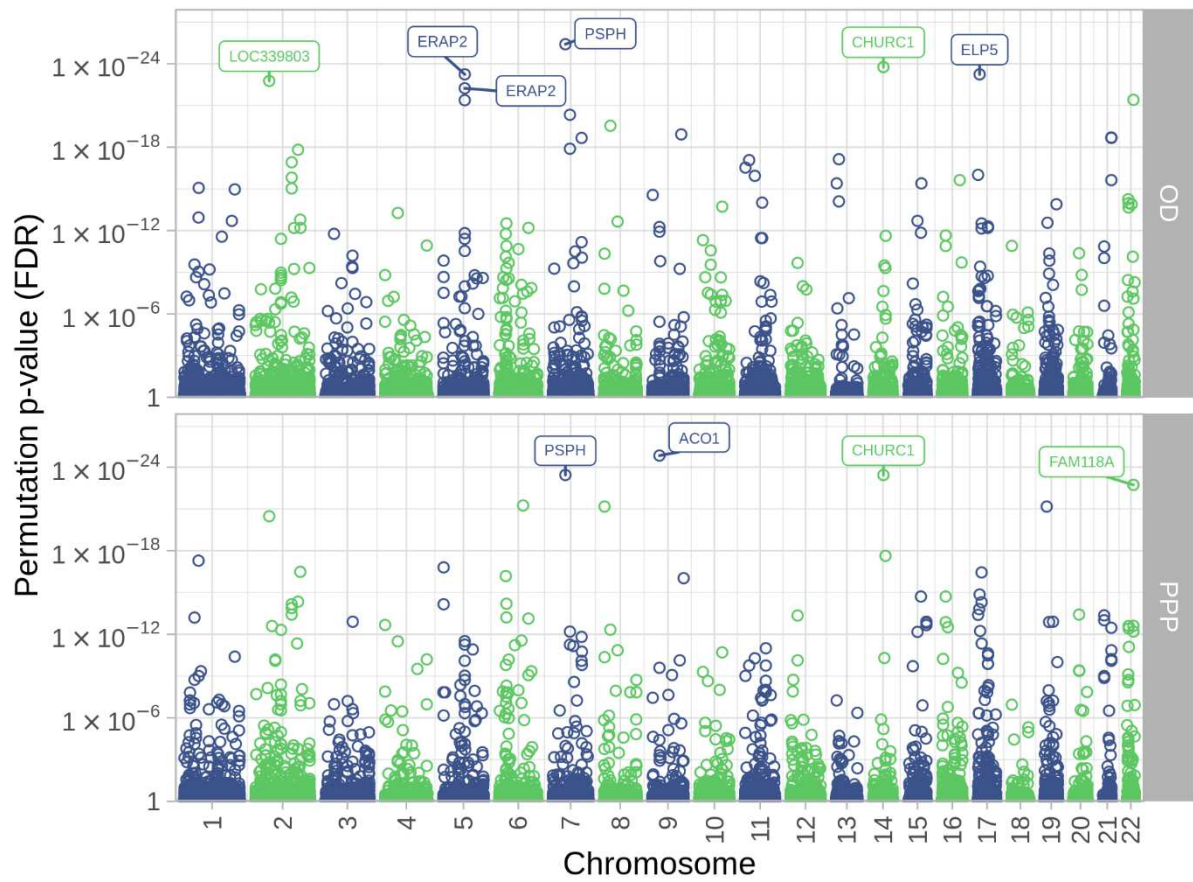
A.



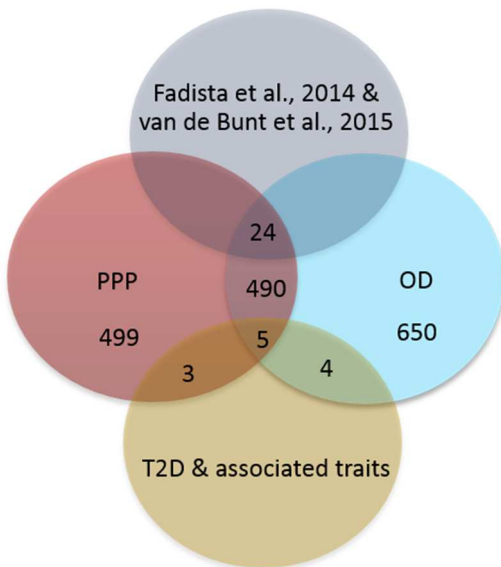
B.



A.



B.



C.

