

1 Multi-omics characterization of type 2 diabetes associated genetic variation.

2
3 Ravi Mandla^{1,2,3*†}, Kim Lorenz^{4,5,6,7*}, Xianyong Yin^{8,9*}, Ozvan Bocher^{10*}, Alicia Huerta-
4 Chagoya^{1,3*}, Ana Luiza Arruda^{10,11*}, Anthony Piron^{12,13,14,15*}, Susanne Horn^{10*}, Ken Suzuki^{16,17,18},
5 Konstantinos Hatzikotoulas¹⁰, Lorraine Southam¹⁰, Henry Taylor^{19,20,21}, Kaiyuan Yang^{22,23}, Karin
6 Hrovatin^{24,25}, Yue Tong²⁶, Maria Lytrivi^{26,27}, Nigel W. Rayner¹⁰, James B. Meigs^{1,28,29}, Mark I.
7 McCarthy^{30,31,32}, Anubha Mahajan^{30,31}, Miriam S. Udler^{1,2,3,28}, Cassandra N. Spracklen³³, Michael
8 Boehnke⁹, Marijana Vujkovic^{4,34,35}, Jerome I. Rotter³⁶, Decio L. Eizirik²⁶, Miriam
9 Cnop^{26,27,37}#, Heiko Lickert^{22,23,38}#, Andrew P. Morris^{16,10,39}#†, Eleftheria Zeggini^{10,40}#†, Benjamin
10 F. Voight^{4,5,6,7}†#, Josep M. Mercader^{1,2,3,29}†#

11 Affiliations

12 ¹Programs in Metabolism and Medical and Population Genetics, Broad Institute of MIT and
13 Harvard, Cambridge, MA, USA. ²Diabetes Unit, Endocrine Division, Department of Medicine,
14 Massachusetts General Hospital, Boston, MA, USA. ³Center for Genomic Medicine,
15 Massachusetts General Hospital, Boston, MA, USA. ⁴Corporal Michael J. Crescenz VA Medical
16 Center, Philadelphia, PA, USA. ⁵Department of Systems Pharmacology and Translational
17 Therapeutics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA.
18 ⁶Department of Genetics, University of Pennsylvania Perelman School of Medicine,
19 Philadelphia, PA, USA. ⁷Institute for Translational Medicine and Therapeutics, University of
20 Pennsylvania - Perelman School of Medicine, Philadelphia PA. ⁸Department of Epidemiology,
21 School of Public Health, Nanjing Medical University, Nanjing, China. ⁹Department of
22 Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA.
23 ¹⁰Institute of Translational Genomics, Helmholtz Zentrum München, German Research Center
24 for Environmental Health, Neuherberg, Germany. ¹¹Graduate School of Experimental Medicine,
25 Technical University of Munich, Munich, Germany. ¹²ULB Center for Diabetes Research,
26 Medical Faculty, Université Libre de Bruxelles, Brussels, Belgium. ¹³Interuniversity Institute of
27 Bioinformatics in Brussels (IB2), Brussels, Belgium. ¹⁴Machine Learning Group, Université Libre
28 de Bruxelles, Brussels, Belgium. ¹⁵Diabetes and Inflammation Laboratory, Centre for Human
29 Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom.
30 ¹⁶Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research,
31 Division of Musculoskeletal and Dermatological Sciences, University of Manchester,
32 Manchester, UK. ¹⁷Department of Diabetes and Metabolic Diseases, Graduate School of
33 Medicine, University of Tokyo, Tokyo, Japan. ¹⁸Department of Statistical Genetics, Osaka
34 University Graduate School of Medicine, Suita, Japan. ¹⁹Center for Precision Health Research,
35 National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA.
36 ²⁰British Heart Foundation Cardiovascular Epidemiology Unit, Department of Public Health and
37 Primary Care, University of Cambridge, Cambridge, UK. ²¹Heart and Lung Research Institute,
38 University of Cambridge, Cambridge, UK. ²²Institute of Diabetes and Regeneration Research
39 (IDR), Helmholtz Munich, Neuherberg, Germany. ²³German Center for Diabetes Research
40 (DZD), Neuherberg, Germany. ²⁴Institute of Computational Biology (ICB), Helmholtz Munich,

41 Neuherberg, Germany. ²⁵School of Life Sciences Weihenstephan, Technical University of
42 Munich, Freising, Germany. ²⁶ULB Center for Diabetes Research, Medical Faculty, Université
43 Libre de Bruxelles, Brussels, Belgium. ²⁷Division of Endocrinology, Erasmus Hospital, Université
44 Libre de Bruxelles, Brussels, Belgium. ²⁸Department of Medicine, Harvard Medical School,
45 Boston, MA, USA. ²⁹Division of General Internal Medicine, Massachusetts General Hospital,
46 Boston, MA, USA. ³⁰Wellcome Centre for Human Genetics, Nuffield Department of Medicine,
47 University of Oxford, Oxford, UK. ³¹Oxford Centre for Diabetes, Endocrinology and Metabolism,
48 Radcliffe Department of Medicine, University of Oxford, Oxford, UK. ³²Oxford NIHR Biomedical
49 Research Centre, Churchill Hospital, Oxford University Hospitals NHS Foundation Trust,
50 Oxford, UK. ³³Department of Biostatistics and Epidemiology, University of Massachusetts
51 Amherst, Amherst, MA, USA. ³⁴Department of Medicine, University of Pennsylvania Perelman
52 School of Medicine, Philadelphia, PA, USA. ³⁵Department of Biostatistics, Epidemiology and
53 Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA.
54 ³⁶Institute for Translational Genomics and Population Sciences, Department of Pediatrics,
55 Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA,
56 USA. ³⁷WEL Research Institute, Wavre, Belgium. ³⁸School of Medicine, Technical University of
57 Munich, Munich, Germany. ³⁹Estonian Genome Centre, Institute of Genomics, University of
58 Tartu, Tartu, Estonia. ⁴⁰TUM School of Medicine and Health, Technical University of Munich and
59 Klinikum Rechts der Isar, Munich, Germany. *These authors contributed equally to this work.
60 #These authors supervised this work. †Correspondence to: rmandla@broadinstitute.org,
61 andrew.morris-5@manchester.ac.uk, eleftheria.zeggini@helmholtz-muenchen.de,
62 bvoight@upenn.edu, mercader@broadinstitute.org.

63

64

65 **Abstract**

66 Discerning the mechanisms driving type 2 diabetes (T2D) pathophysiology from genome-wide
67 association studies (GWAS) remains a challenge. To this end, we integrated omics information
68 from 16 multi-tissue and multi-ancestry expression, protein, and metabolite quantitative trait loci
69 (QTL) studies and 46 multi-ancestry GWAS for T2D-related traits with the largest, most
70 ancestrally diverse T2D GWAS to date.

71 Of the 1,289 T2D GWAS index variants, 716 (56%) demonstrated strong evidence of
72 colocalization with a molecular or T2D-related trait, implicating 657 *cis*-effector genes, 1,691
73 distal-effector genes, 731 metabolites, and 43 T2D-related traits. We identified 773 of these *cis*-
74 and distal-effector genes using either expression QTL data from understudied ancestry groups
75 or inclusion of T2D index variants enriched in underrepresented populations, emphasizing the
76 value of increasing population diversity in functional mapping. Linking these variants, genes,
77 metabolites, and traits into a network, we elucidated mechanisms through which T2D-
78 associated variation may impact disease risk. Finally, we showed that drugs targeting effector
79 proteins were enriched in those approved to treat T2D, highlighting the potential of these results
80 to prioritize drug targets for T2D.

81 These results represent a leap in the molecular characterization of T2D-associated genetic
82 variation and will aid in translating genetic findings into novel therapeutic strategies.

83

84 **Introduction**

85 Type 2 diabetes (T2D) and its associated complications are one of the biggest global health
86 problems of the 21st century^{1,2}. The biological mechanisms underlying T2D are not fully
87 understood, yet expanding upon pathways for T2D could inform therapeutic approaches. It has
88 been demonstrated that clinical trials of drugs that have genetic support are more likely to be
89 successful and to be granted expedited development and review by the FDA³⁻⁶. Therefore, a

90 logical approach to accelerate development of new T2D drugs is through linking T2D-associated
91 variants with their effector genes.

92
93 One strategy to nominate potential disease effector genes is to identify rare variants that alter
94 the protein coding sequence through putative loss-of-function (pLOF) or amino-acid altering
95 variants^{7,8}. While this approach has generated numerous leads⁹, discovery of rare coding
96 variants associated with T2D has been limited by small sample sizes from available exome or
97 genome sequencing studies, especially in underrepresented populations. In contrast, large-
98 scale multi-ancestry array-based genome-wide association studies (GWAS) have been
99 successful in identifying hundreds of genetic associations¹⁰⁻¹⁵. However, only 50 (4%) of the
100 1,289 T2D-associated genetic variants from the latest T2D GWAS are themselves a pLOF or
101 amino-acid altering variant¹⁵. Previous results have also suggested only 43% of T2D-associated
102 protein-altering variants show evidence of causality after fine-mapping¹⁶, complicating efforts
103 with this strategy alone to identify the effector genes underlying most T2D associations.
104 Additionally, inference of the tissue of action, effect direction, and physiological mechanisms of
105 T2D-associated variants are also important to translate these discoveries into mechanistic
106 insights and novel drug targets.

107
108 Many T2D variant-to-function efforts have attempted to determine these downstream targets
109 through the identification of shared causal signals underlying both T2D and a single omics layer
110 from a quantitative trait loci (QTL) analysis, with gene expression being the most common¹⁷⁻²⁰.
111 However, a recent GWAS for T2D could only map *cis*-effector genes for 21% of their signals
112 when using transcriptomic and proteomic data¹¹, motivating the pursuit of integrating larger and
113 more ancestrally diverse molecular data to identify additional mechanisms. To date, few studies
114 in T2D have attempted to jointly analyze transcriptomic, proteomic, and metabolomic data to link
115 genetic variations to T2D biological pathways. Furthermore, there have been limited attempts to

116 explore the overlap of these functional genomics datasets with cardiometabolic trait-associated
117 clusters for T2D^{15,21} to better characterize the biology underlying T2D heterogeneity. More
118 importantly, previous analyses have focused on populations genetically similar to European
119 ancestry individuals from the 1000 Genomes Project (1000G), which we henceforth refer to as
120 EUR-like, limiting our understanding of the biological downstream effects of T2D in other
121 population groups such as those genetically similar to African ancestry (AFR-like), admixed-
122 American (AMR-like), East Asian ancestry (EAS-like), and South asian ancestry (SAS-like).
123

124 Here, we tested for colocalization between the largest, recently reported multi-ancestry T2D
125 GWAS meta-analysis¹⁵ with association datasets for transcriptomic, proteomic, and
126 metabolomic traits to identify effector transcripts and downstream targets for T2D across
127 multiple tissues and ancestry groups. Additionally, we used GWAS data from T2D-related traits,
128 including glycemic, cardiometabolic, and anthropometric traits, to understand the physiological
129 mechanisms linked to T2D-associated variation and to further characterize the heterogeneity of
130 T2D loci. We found evidence of colocalization between T2D signals and molecular or T2D-
131 related traits for the majority of the T2D signals, expanding the catalog of variant-to-gene
132 mappings and physiologic correlates with disease. We combined these associations into an
133 interactive network to define relationships between genetic variants, molecular trait data, and
134 T2D-related traits, providing a resource to prioritize mechanisms and propose potential
135 therapeutic hypotheses.

136

137 **Results**

138 ***Overall analysis strategy***

139 To understand molecular mechanisms and physiological mechanisms of T2D-associated
140 genetic variation, we evaluated evidence of pairwise-colocalization between each of the 1,289
141 T2D-associated index variants identified in Suzuki et al.¹⁵ and *cis*-gene expression QTL from

142 diverse tissues and ancestries (eQTL, 10 datasets), protein QTL (*cis*- and *trans*-pQTL, 4
143 datasets), metabolite QTL (metabQTL, 2 datasets), and T2D-related cardiometabolic trait
144 GWAS (46 datasets; **Fig. 1a; Supplementary Tables 1 and 2; Methods**). We identified 12,180
145 colocalizations with a posterior probability of a shared causal variant (PP.H4) > 0.8,
146 corresponding to a total of 716 T2D index variants colocalizing with 657 *cis*-effector genes from
147 an eQTL or *cis*-pQTL, 1,691 distal-effector genes from *trans*-pQTL, 731 metabolites, and 43
148 T2D-related traits (**Fig. 1b, 1c; Supplementary Tables 3, 4, 5, and 6; Methods**).

149
150 We represented the 12,180 colocalizations as a multi-layer colocalization network, where nodes
151 represent T2D GWAS index variants, genes, metabolites, or traits, and edges represent a
152 colocalization between T2D and eQTLs, pQTLs, metabQTLs or trait GWAS data (**Fig. 1a;**
153 **Supplementary File**).

154

155 ***Identification of T2D cis-effector genes and their relevant tissues***

156 To identify candidate T2D effector genes and their tissues of action, we tested for evidence of
157 colocalization between T2D and eQTL datasets from blood^{22,23} and six T2D-related tissues:
158 pancreatic islets¹⁹, subcutaneous and visceral adipose tissue, liver, hypothalamus, and skeletal
159 muscle²³ (**Methods**). We observed 1,324 colocalizations, representing 632 candidate *cis*-
160 effector genes, and 349 (27.1%) of the 1,289 T2D index variants. Of the 349 index variants, 110
161 (32.0%) colocalized to different effector genes in different tissues, suggesting evidence of
162 varying effects by tissue (**Supplementary Table 7**).

163

164 To quantify the improvement in our list of T2D effector genes, we compared our results to
165 previous studies that performed colocalization analyses on different T2D GWAS^{11,14,19,24}. We
166 found that the present analysis doubles the number of candidate effector genes compared to
167 the largest previous study¹⁴, owing to the larger sample size and diversity of the T2D GWAS,

168 and the new and larger eQTL datasets we employed (**Fig. 2a**). Indeed, the number of
169 colocalizations per eQTL dataset was strongly correlated with the eQTL dataset sample size
170 (Pearson $r=0.79$, $P=6.8\times 10^{-3}$), highlighting the importance of statistical power in colocalization
171 analyses (**Extended Data Fig. 1**).

172
173 We observed many colocalizations in a single tissue only (**Fig. 2b**) with pancreatic islets having
174 the highest proportion of single-tissue observed colocalizations (0.50; **Fig. 2c**). One of the
175 colocalizations only observed in pancreatic islets was with *SCTR*, a G-protein-coupled receptor
176 in the same family as known T2D drug target *GLP1R* (**Fig. 2d**). *SCTR* is highly and specifically
177 expressed in pancreatic islets¹⁹ and is downregulated in pancreatic islets of T2D patients²⁵,
178 supporting the relevance of pancreatic islet expression of *SCTR* in T2D pathophysiology. *SCTR*
179 also harbors a missense variant associated with lower T2D risk^{12,13} that is enriched in EAS-like
180 populations.

181
182 To validate the candidate effector genes colocalizing with T2D in islets, we performed gene set
183 enrichment analyses between differentially expressed T2D genes in islets and these effector
184 genes stratified by the direction of effect for the T2D risk allele (**Supplementary Table 8**;
185 **Methods**). We found that genes from colocalizations where decreased expression is associated
186 with increased T2D risk are significantly enriched in genes downregulated in islets of T2D cases
187 (Normalized Enrichment Score [NES]=-1.7; $P=0.01$) (**Extended Data Fig. 2**). Enrichment only in
188 the colocalizations where the risk allele reduces expression may indicate a large abundance of
189 associations between diminished gene function and T2D risk.

190
191 ***pQTL colocalization identifies additional cis and trans effector proteins***

192 To expand our identification of effector genes beyond eQTLs, we tested for evidence of
193 colocalization between T2D and four pQTL datasets (plasma UKB, plasma deCODE, plasma

194 HELIC, and brain ROS/MAP). We identified 42 and 3,423 colocalizations with a *cis*-pQTL and
195 *trans*-pQTL respectively, encompassing 1,728 unique genes and 365 (28.3%) T2D index
196 variants.

197
198 We identified most of the colocalizations (2,451, 70.7%) with pQTLs from the deCODE dataset,
199 which had the second largest sample size but included more proteins, owing to the use of the
200 Somascan panels (N=35,559; N_{proteins}=4,907), compared with the OLINK panels employed by
201 the UKB data (N=54,219, N_{proteins}=1,472; **Fig. 3a**). We also observed 85 colocalizations with the
202 HELIC data (N=2,933) and 7 with the ROS/MAP data (N=330). Despite differences in sample
203 size and protein sets, we observed 37 identical colocalizations (all in *trans*) with deCODE and
204 UKB pQTL datasets, which is 7.5-fold higher than expected by chance considering the overlap
205 of proteins detected by both the Somascan and OLINK panels (Fisher's exact test $P=7.4 \times 10^{-20}$;
206 **Methods; Supplementary Table 9**). Among those 37, we found strong consistency of protein
207 effect size estimates between deCODE and UKB (Pearson $r=0.93$; $P=9.1 \times 10^{-17}$; **Fig. 3b**). One
208 example colocalization detected with both UKB and deCODE pQTL datasets was for CBLN4
209 with T2D index variant rs1415287 in *trans*, within the *LYPLAL1* locus defined in Suzuki et al.¹⁵,
210 where the risk allele associated with increased levels of CBLN4 (**Fig. 3c**). rs1415287
211 additionally colocalized with IGFBP-1 in *trans* (**Fig. 3d**), with decreased IGFBP-1 levels
212 associated with the T2D risk allele. A previous study found decreased CBLN4 and increased
213 IGFBP-1 plasma levels after administering SGLT2 inhibitors, a class of T2D drugs which
214 reduces glucose reabsorption within the proximal renal tubule, to participants of varying glucose
215 tolerance²⁶, mirroring our findings of increased CBLN4 and decreased IGFBP-1 levels with the
216 T2D risk allele.

217
218 We also identified 784 genes for which there was a pQTL colocalizing with more than one T2D
219 index variant (**Extended Data Fig. 3a**). We ranked these genes based on the association

220 between the pQTL and T2D GWAS effect sizes across all index variants that colocalized with a
221 given gene (**Supplementary Table 10, Methods**). Among the highest ranked genes, *IGFBP2*
222 had 14 colocalizations with overall negative associations between T2D-associated index variant
223 risk allele effect size and protein levels, supporting previously reported negative associations
224 between IGFBP-2 levels and T2D incidence²⁷ (**Extended Data Fig. 3b**). Seven of these index
225 variants (50%) have previously been classified as being involved in lipodystrophy-related
226 pathways¹⁵, suggesting IGFBP-2 may be involved lipodystrophy-like T2D. In agreement with
227 that, previous publications have found associations between increased IGFBP-2 and lower BMI,
228 lower waist-hip ratio, and increased triglycerides²⁸ and decreased NAFLD risk^{29,30}. Together,
229 these results highlight the value of highly powered pQTL datasets in identifying biologically
230 relevant effector genes.

231

232 ***Inclusion of understudied population datasets boosts effector gene identification***

233 In the latest multi-ancestry T2D meta-analysis, 289 index variants (22.4%) have a MAF half as
234 large in EUR-like populations compared to other populations¹⁵, in total mapping to 615 effector
235 genes from a previously-described T2D-relevant tissue eQTL colocalization or pQTL
236 colocalization. However, identifying effector genes for variants enriched in underrepresented
237 populations remains a challenge due to a lack of molecular data from the same population³¹. To
238 quantify the extent to which including molecular data from underrepresented populations
239 improves the discovery of effector genes, we tested for colocalization between blood eQTL
240 datasets collected in self-reported African American (AA), Mexican American (MX), and Puerto
241 Rican (PR) cohorts²² and the EUR-like GTEx dataset with both the multi-ancestry T2D GWAS
242 and with matched ancestry T2D GWAS containing only AFR-like participants within the USA or
243 AMR-like participants. We identified 633 colocalizations with one of these eQTL datasets
244 (47.8% of all eQTL colocalizations), representing 347 gene-to-variant mappings. We identified
245 204 (58.8%) of the 347 blood eQTL mappings in the AA, MX, or PR eQTL datasets only, and 37

246 (10.7%) only in the EUR-like GTEx dataset (**Fig. 4a; Methods; Supplementary Table 11**). In
247 total, adding the AA, MX, or PR eQTL data identified 158 additional T2D effector genes not
248 identified with any of the tested EUR-like eQTL or pQTL datasets, highlighting the importance of
249 collecting data from underrepresented populations, even when testing a tissue that is not
250 directly relevant for T2D pathophysiology. We found that T2D index variants with a
251 colocalization observed only with an AA, MX, or PR blood eQTL were more common in AMR-
252 like participants and less common in EUR-like participants relative to T2D index variants with a
253 colocalization observed only with the EUR-like GTEx blood eQTL dataset (**Fig. 4b; $P=0.02$**).
254 These results suggest differences in statistical power due to allele frequency heterogeneity
255 and/or varying linkage disequilibrium (LD) with the true causal variant may underly differential
256 detection of colocalizations across ancestry groups.

257
258 Two example colocalizations observed only in the MX blood eQTL data were with *LIN7A* and
259 *ACSS3*, both with the same lead variant from colocalization analyses, rs10128882 (**Fig. 4c**).
260 The LD between rs10128882 and the T2D GWAS index variant rs11114650 was different
261 between ancestry groups, with the highest being in EAS-like individuals ($r^2=0.89$) and the lowest
262 in AFR-like individuals ($r^2=0.02$), possibly reflecting differential LD to the true causal variant
263 across populations. The T2D risk allele of rs11114650 has a frequency of 1.5% in gnomAD-
264 defined European populations compared to 39% in gnomAD-defined African populations (**Fig.**
265 **4d**) and is associated with decreased blood expression levels of *LIN7A* and *ACSS3* (**Fig. 4e**).
266 *LIN7A*, mainly expressed in the brain, is involved in localizing, distributing, and maintaining
267 channels and receptors at polarized cell membranes³², while *ACSS3* is highly expressed in
268 brown adipose tissue and has been previously associated with insulin-resistant obesity-like
269 phenotypes in mice³³.

270

271 ***Identification of effector metabolites***

272 To understand the metabolic processes of T2D-associated genetic variation, we tested for
273 colocalization between T2D GWAS and plasma metabQTL data from UKB and the METSIM
274 study. We identified 5,221 metabQTL colocalizations, corresponding to 283 T2D index variants,
275 with a majority of these colocalizations (4,480, 85.8%) with the UKB metabQTL data. This is
276 expected given the larger sample size of the UKB data (N=114,999) compared to the METSIM
277 data (N=6,136). Of the 18 metabolites measured in the METSIM Metabolon panel that could be
278 mapped to metabolites in the UKB Nightingale panel, 7 colocalized with the T2D GWAS in both
279 METSIM and UKB, an overlap greater than expected by chance (Fisher's exact test OR=26.2,
280 $P=1.8 \times 10^{-7}$; **Methods; Supplementary Table 12**), suggesting good concordance between the
281 two platforms for shared metabolites.

282
283 To prioritize metabolites with the largest implication with T2D, we ranked metabolites by the
284 number of T2D index variants with which they colocalize, and by the association of genetic
285 effect sizes between T2D and metabolites across all the index variants that colocalized with
286 each metabolite (**Fig. 5a; Supplementary Table 13; Methods**). We found consistent
287 associations between the risk allele effects and expected metabolite levels, such as T2D risk
288 alleles associated with increased glucose levels (**Fig. 5a, 5b**). Among less well-established
289 metabolites, we found 10 colocalizations where the T2D risk alleles lowered phosphatidylcholine
290 levels (**Fig. 5c**), consistent with higher phosphatidylcholine intake previously associated with
291 lower risk of T2D^{34,35}. A subnetwork of all phosphatidylcholine-linked colocalizations was
292 enriched for metabolites known to be related to phosphatidylcholine metabolism, such as
293 sphingomyelin ($P_{adj.}=0.02$), cholines ($P_{adj.}=2.4 \times 10^{-17}$), and polyunsaturated fatty acids (PUFAs)
294 ($P_{adj.}=0.02$; **Fig. 5d; Supplementary Table 14; Methods**). Within this subnetwork, SMPD1
295 (sphingomyelin phosphodiesterase 1) protein levels also colocalized with 4 of the same 10 T2D
296 index variants, including two detected in two plasma pQTL studies (UKB pQTL and deCODE
297 pQTL for rs9987289 and UKB pQTL and HELIC pQTL for rs8107974), providing robustness to

298 these results. SMPD1 acts in the same biochemical pathways as phosphatidylcholine synthesis
299 and, along with PUFAs, has previously been associated with diabetic complications including
300 diabetic retinopathy³⁶. Combined, these results highlight the role of phosphatidylcholine
301 metabolic pathways in T2D.

302

303 ***Colocalization with T2D-related traits recapitulates previously described T2D genetic***
304 ***clusters***

305 We next performed colocalization analyses between T2D and 46 T2D-related traits to shed light
306 into distinct T2D pathophysiologic mechanisms. We identified 2,170 colocalizations, including
307 521 (40.4%) of the T2D index variants and 43 traits.

308

309 Previously, Suzuki et al. grouped T2D index variants based on their associations with a smaller
310 subset of 37 of these traits to define 8 distinct genetic T2D mechanistic clusters (Beta cell with
311 positive proinsulin, Beta cell with negative proinsulin, Residual glycaemic, Body fat, Metabolic
312 syndrome, Obesity, Lipodystrophy, Liver/lipid metabolism). Their analyses used a k-means
313 clustering approach on the Z-scores of the T2D index variants from each of the 37 T2D-related
314 trait GWAS¹⁵. We expanded on this work by assessing the proportion of T2D index variants per
315 cluster which could be mapped to a colocalization between T2D and at least one of the 37 traits
316 previously used for clustering. We found that the Obesity cluster had the highest number of
317 variants confirmed by colocalization, with 163 (70%) of the 223 index variants colocalizing with
318 at least one of the 37 traits, reflecting the large number of obesity-related traits considered
319 during colocalization testing. The colocalizing trait patterns also largely recapitulated the single-
320 variant association-derived clusters¹⁵ (**Fig. 6a**, left panel).

321

322 We also tested the enrichment of eQTL colocalization from different tissues and metabQTL
323 colocalizations in each of the clusters. We observed enrichment of subcutaneous adipose tissue

324 colocalizations with the Metabolic syndrome cluster ($P=2.2\times 10^{-5}$) and of the pancreatic islet
325 colocalizations with the Beta Cell +PI cluster ($P=1.6\times 10^{-5}$; **Extended Data Fig. 4**;
326 **Supplementary Table 15**). We also observed enrichment of fatty acid metabolites with the
327 Lipodystrophy clusters ($P=2.0\times 10^{-9}$), and of glycolysis related metabolites with the Beta Cell +PI
328 cluster ($P=4.7\times 10^{-9}$; **Extended Data Fig. 4**; **Supplementary Table 16**). When aligning effect
329 allele directions and effect sizes, we again recapitulated cluster patterns from Suzuki et al.
330 based on single variant lookups. The results mirror the effects between Obesity (largely same
331 direction) and Metabolic Syndrome/Lipodystrophy (largely opposite direction) cluster SNPs
332 across anthropometric and adipose traits¹⁵ (**Fig. 6a**, right panel). As we restricted the analyses
333 to loci that have evidence of colocalization, these results are more specific than those previously
334 presented.

335
336 Within each cluster, variants showed different patterns of colocalizations with different traits
337 (**Fig. 6b**). For example, the two Beta Cell clusters have the bulk of their colocalizing variants
338 linked to at least one glycaemic trait, while the variants in the Obesity cluster are linked to
339 anthropometric, adipose, and lipids traits (**Extended Data Fig. 5**). In contrast, the Lipodystrophy
340 cluster variants are more pleiotropic, with 19 of 27 variants linked to 4 or more trait groups,
341 suggestive of heterogeneity within lipodystrophy-related variants (**Fig. 6c**).

342 343 ***Effector genes are enriched in drug targets for diabetes***

344 The prioritization of drugs targeting candidate genes can accelerate clinical translation^{5,37,38}. We
345 pulled known drugs which target one of the 2,311 genes identified from a *cis*-eQTL or a *cis*- or
346 *trans*-pQTL colocalization (**Methods**). Among 1,964 gene-to-drug mappings, consisting of 665
347 drugs listed as approved, not withdrawn and with information on disease indications, and 159 of
348 our colocalizing genes (**Supplementary Table 17**).

349

350 We tested if drugs targeting any of these genes were enriched with an approved indication for
351 diabetes. We observed a 3.5-fold enrichment for approved diabetes indication within drugs that
352 target our list of effector genes ($P=4.4\times 10^{-10}$; **Fig. 7a; Methods**) and a 2.0-fold enrichment for
353 any diabetes indication within drugs targeting the effector genes ($P=1.9\times 10^{-6}$; **Fig. 7b;**
354 **Methods**), compared to all other drugs. We then performed the same analysis stratified by the
355 source of colocalization, using lists of 1,681 genes identified only from a pQTLs (*cis* or *trans*),
356 1,654 only from *trans* pQTLs, or 583 only from eQTLs. Drugs targeting pQTL-only identified
357 genes were enriched for approved T2D drug targets (OR=2.4, $P=3.4\times 10^{-4}$). Drugs targeting
358 eQTL-only identified genes were the most enriched for approved drugs for diabetes (OR=3.9,
359 $P=3.7\times 10^{-7}$). These results suggest that drugs targeting any of our candidate effector genes are
360 potential candidates for repurposing to treat T2D. Permutation analyses similarly found
361 significant enrichment for diabetes drugs among our genes relative to random sets of genes
362 (**Extended Data Fig. 6; Methods**).

363
364 Among the genes observed only from an eQTL colocalization, we identified *FXVD2* (**Fig. 7c**)
365 which has been proposed as a pancreatic beta-cell specific biomarker³⁹. *FXVD2* modulates
366 Na,K-ATPase activity and cell proliferation⁴⁰ and is expressed at the early stages of the human
367 endocrine pancreas development (15 weeks), preceding insulin detection³⁹, suggesting a
368 putative role in islet development. We also detected *FXVD2* mRNA expression at the end of the
369 differentiation process of induced pluripotent stem cells into islet-like cells in both bulk and
370 single-cell datasets^{28,41-43} (**Extended Data Fig. 7**).

371
372 The risk allele of the T2D index variant rs529623 was associated with lower expression of
373 *FXVD2* in pancreatic islets¹⁹ (**Fig. 7d**; $P=3.4\times 10^{-7}$). We further explored expression differences
374 of *FXVD2* in existing bulk RNA-seq⁴⁴ and single-cell RNA-seq⁴⁵ (scRNA-seq) from pancreatic

375 islets of human donors with and without T2D (**Methods**). After adjusting for multiple hypothesis
376 testing, *FXVD2* was significantly downregulated in pancreatic islets of T2D patients (**Fig. 7e**)
377 specifically within Beta cells⁴⁵ (**Fig. 7f**). There are currently five approved drugs that target
378 *FXVD2* but are not approved for diabetes: delanoside, digoxin, digitoxin, acetyldigitoxin, and
379 lanatoside. These drugs potentially could be modified to treat T2D, particularly if further
380 preclinical investigations show indications that they can target and modulate beta cell function.

381
382 We also found multiple other drugs targeting proteins for one of our effector genes which had
383 clinical evidence to support a role in impacting glycemia and may represent possible targets for
384 drug repurposing (**Supplementary Table 18**). For example, amisulpride, an antagonist for the
385 Serotonin (5-HT) receptor encoded by *HTR6*, increases pancreatic insulin secretion in healthy
386 controls⁴⁶. We identified *HTR6* from a blood *trans*-pQTL colocalization where the T2D risk allele
387 is associated with increased protein levels, consistent with amisulpride, which antagonizes
388 *HTR6* and increases insulin secretion.

389 390 **Discussion**

391 T2D is a model for complex disease genetics owing to its high prevalence and polygenicity, with
392 the largest T2D multi-ancestry meta-analysis to date based on >2.5 million participants,
393 identifying 1,289 index variants¹⁵. However, understanding the function of associated variants
394 from a GWAS is extremely challenging, and only few examples have provided additional
395 insights into the molecular mechanisms underpinning signals⁴⁷⁻⁵¹. These comprehensive studies
396 take years to complete, and biological insights from previous studies were mostly described in
397 EUR-like populations^{14,18,19,24}, limiting findings for other populations in which T2D risk alleles are
398 now being increasingly reported^{11,15}.

399

400 Here, we aimed to create a catalog of effector genes, metabolites, and traits for the 1,289 T2D-
401 associated index variants. We identified >12,000 colocalizations with T2D across all omics
402 layers and T2D associated traits tested. We found *cis*-effector genes for 361 of the index
403 variants (28%) and identified a >2-fold increase in the number of *cis*-effector genes compared
404 with the previous largest T2D GWAS¹⁴. We observed many of the eQTL colocalizations in only
405 one of the tissues tested, emphasizing the importance of including multiple tissue types in
406 variant-to-function efforts for T2D. We also demonstrated the value of multi-ancestry eQTL data
407 through the identification of 158 additional effector genes (24% of all *cis*-effector genes) using
408 blood eQTL data from AA, PR, and MX populations²² that we did not observe in eQTL or pQTL
409 data from EUR-like populations alone. We demonstrate that even in blood, which is not directly
410 related to T2D progression, including eQTL data from multiple ancestry groups improves
411 effector gene discovery. Despite substantial progress in the inclusion of diverse ancestries in
412 GWAS meta-analysis, the availability of molecular data from diverse ancestries remains
413 limited³¹. We hope that the present findings will help motivate the generation of molecular QTL
414 resources from participants of diverse ancestries to unveil additional effector genes and
415 potential drug targets.

416
417 We compiled the pairwise colocalizations between T2D and each omics layer and additional
418 traits to create a network of T2D index variants mapped to genes, metabolites, and traits. This
419 provides the most comprehensive resource to date of the genetic relationships between genes,
420 traits, metabolites, and genetic variants associated with T2D to the research community. Such a
421 resource allows searches for genes, metabolites, or traits of interest and find the T2D index
422 variants underlying their shared causal signals with T2D to quickly generate hypotheses and
423 gain insights into T2D pathophysiology. We additionally include effect and variant information
424 from the molecular QTL, trait GWAS, and T2D GWAS summary statistics to further improve the
425 searchability of the network.

426

427 Identifying drug targets for complex diseases has traditionally focused on protein coding
428 variants, which involves designating the target gene and mimicking a loss-of-function or gain-of-
429 function effect^{52,53}. However, effector genes regulated by non-coding variants may also be
430 viable drug targets once the mechanism of action is understood. In this study, we identified
431 candidate effector genes, their tissue of action and direction of effect, for potential drug target
432 identification. We show that our list of effector genes is enriched for targets of approved drugs
433 for treatment of T2D, suggesting that drugs not currently used to treat diabetes but targeting one
434 of the effector genes are strong candidates for modification for the treatment of T2D. We
435 acknowledge that additional studies to test causality, druggability, and functional validation with
436 tissue and cell types of interest are needed.

437

438 Our study has limitations. First, we utilized the 1,289 index variants from Suzuki et al. to define
439 our tested colocalization regions, but we recognize that these variants were not assessed for
440 statistical independence applying formal fine-mapping approaches. Second, we leveraged the
441 multi-ancestry meta-analysis for T2D to maximize sample size. However, this approach may
442 underestimate the true number of colocalizations, as we could not perform fine-mapping to
443 differentiate multiple independent signals within the same locus due to current limitations in fine
444 mapping and conditional analyses utilizing only multi-population summary statistics data. Third,
445 confirmation of the tissue of action and the causality and directionality of the effect, especially
446 for distal-effector genes, remains to be determined. This necessitates careful investigation
447 through appropriate causal inference experiments, cellular models, and *in vivo* studies. To
448 facilitate these efforts, we provide here the most extensive list of effector genes, tissues, and
449 potential physiological pathways linked to T2D associated variation for initial prioritization
450 strategies.

451

452 In conclusion, we present a comprehensive functional follow-up of the largest T2D GWAS meta-
453 analysis and provide a framework model for functional annotation of other complex multi-
454 ancestry GWAS. We provide an interactive network to explore the relationships between T2D-
455 associated variants and genes, metabolites, and traits. We anticipate that this resource will
456 serve as a valuable tool for generating testable hypotheses that may lead to the discovery of
457 novel T2D drug targets and therapeutic strategies.

458

459 **Methods**

460 ***GWAS and Molecular Trait Associated Datasets***

461 We downloaded GWAS for T2D relevant traits downloaded from a variety of sources described
462 in **Supplementary Table 1**^{54–70}. We used T2D multi-group meta-analysis generated using
463 METAL⁷¹ from Suzuki et al.¹⁵ for all colocalization analyses. Additionally, single population
464 group meta-analyses using METAL from Suzuki et al. for AFR-like participants living within the
465 USA and for AMR-like participants were used for colocalization. We removed variants with a
466 MAF <0.5% across all the 1000G-continental similarity groups from the analyses.

467

468 We downloaded pancreatic Islet eQTL data from TIGER¹⁹. Since effect size and variance
469 information is not available in the TIGER eQTL data, we used p-value information from the
470 summary statistics and allele frequencies from UKB to calculate approximate bayes factors for
471 colocalization analyses. We downloaded eQTL data for subcutaneous adipose, visceral
472 adipose, skeletal muscle, hypothalamus, and liver tissue and blood from GTEx v8. We also
473 tested eQTL data for whole blood from self-reported African American (AA), Mexican American
474 (MX), and Puerto Rican (PR) populations for colocalization²². Additionally, we tested T2D
475 GWAS including only AFR-like participants from the USA for colocalization with the AA blood
476 eQTL data and tested T2D GWAS including only AMR-like participants for colocalization with
477 the MX and PR blood eQTL datasets. We restricted the colocalization analyses to just these

478 eQTL datasets as pancreatic islets, subcutaneous adipose, visceral adipose, skeletal muscle,
479 hypothalamus, and liver are known to be relevant to T2D. Despite blood representing multiple
480 tissues and not being a target of T2D, the availability of blood eQTL from diverse populations
481 motivated their inclusion.

482
483 We performed integration with pQTL data using summary statistics from deCODE, UKB, HELIC,
484 and ROS/MAP. The deCODE pQTL study contains 4,907 aptamers measured with the
485 SomaScan version 4 assay (SomaLogic) in the plasma samples of 35,559 Icelanders⁷². The
486 UKB Pharma Proteomics Project characterized 2,941 plasma protein analytes of 54,219 UKB
487 participants using the antibody-based Olink Explore 3072 proximity extension assay and tested
488 for pQTL with 2,923 proteins⁷³. pQTL for only 1,472 of these proteins were available at the time
489 of this study. Plasma pQTL from the Hellenic Isolated Cohorts (HELIC)⁷⁴ included 2,933
490 samples from two isolated Greek populations (Pomak and the Mylopotamos villages) and 543
491 proteins from six different OLINK panels: Cardiovascular II, Cardiovascular III, Metabolism,
492 Neurology, Neuro-exploratory and Cardiometabolic⁷⁵⁻⁷⁷. Genetic and proteomic data from post-
493 mortem samples of the dorsolateral prefrontal cortex of older adults was provided by the
494 Religious Orders Study (ROS) and Rush Memory and Aging Project (MAP)⁷⁸. The pQTL
495 analysis was conducted for every HapMap3 common variant (MAF >0.05) within a 100 Kb
496 window around protein-coding genes.

497
498 For the colocalization analyses with metabolites, we used summary statistics from the METSIM
499 and UKB cohorts. The METSIM metabolomics study measured 1,391 plasma metabolites in
500 6,136 Finnish men using the Metabolon Discovery HD4 mass spectrometry platform⁷⁹. In the
501 UKB cohort, a total of 249 metabolites (168 absolute levels and 81 derived ratios and
502 percentages) were measured from the Nightingale panel and tested for genome-wide
503 associations with 114,999 individuals of European ancestry⁸⁰.

504

505 **Colocalization Analyses**

506 We performed pairwise colocalizations between the multi-group T2D GWAS meta-analysis and
507 eQTL, pQTL, metabQTL, and relevant trait GWAS datasets using the *coloc.abf* function from
508 the *coloc*⁸¹ package in R. *coloc.abf* requires regions to be specified to test for colocalization. To
509 define these regions, index variants were sorted by p-value and a 500 Kb window was
510 constructed around each variant. If a less significant index variant was within the window of a
511 more significant variant, the less significant variant's window was removed from the list of
512 regions. To prevent additional overlap of regions, any regions overlapping by more than 200 Kb
513 were merged and regions overlapping by less than 200 Kb were split by half of their overlap
514 size. The final region boundaries used in the analyses can be found in **Supplementary Table 2**.
515 A colocalization analysis was performed between a T2D region and a QTL or cardiometabolic
516 trait if and only if there was evidence of an association (defined as $P < 1 \times 10^{-4}$) of the T2D index
517 variant or a proxy for the T2D index variant for that trait. Proxy variants were defined as a
518 variant in strong linkage disequilibrium (LD, $r^2 > 0.8$) in all continental population groups defined
519 by in 1000G (AFR, AMR, EAS, EUR, and SAS).

520

521 *coloc.abf* provides the posterior probability that a region shares a causal variant in both
522 association summary statistics tested and posterior probabilities that each variant present in
523 both summary statistics is the causal variant. A colocalization was deemed of interest if the
524 posterior probability or shared variation (PP.H4) was greater than 0.8. To determine if a T2D
525 GWAS index variant is the most likely shared causal variant for a region in two colocalizing
526 datasets, the variant with the highest posterior probability of being shared was mapped to a
527 index variant if it was in partial-to-strong LD with the index variant ($r^2 > 0.5$) in 1000G population
528 groups described above.

529

530 We labelled an eQTL colocalization as being observed in one tissue or population group dataset
531 if the PP.H4 was greater than 0.8 in the dataset, but less than 0.3 in all other tissue or
532 population group datasets. A maximum of 0.3 was chosen to avoid selecting suggestive
533 colocalizations that may exist but not be sufficiently powered to be detected with our datasets.
534 For single tissue analyses, we combined subcutaneous adipose and visceral omentum adipose
535 results and combined blood results from each ancestry due to the larger similarities in gene
536 expression profiles between these datasets relative to the other tissues analyzed.

537
538 We characterized pQTL colocalizations as being in *cis* if the mapped index variant was within 1
539 Mb of the pQTL gene TSS; otherwise they were denoted as in *trans*. We tested for enrichment
540 of colocalizations identified between deCODE pQTL which uses the Somascan assay and UKB
541 pQTL which uses the Olink assay, using Fisher's exact test. We constructed contingency tables
542 from all gene to T2D index variant mappings possible for genes included in Somascan and/or
543 Olink respectively. We also used this approach to test for an enrichment of colocalizations
544 identified between METSIM metabQTL and UKB metabQTL, using all metabolite to T2D index
545 variant mappings possible for metabolites included in METSIM and/or UKB respectively.

546
547 We also tested for enrichment between T2D index variants mapped to an eQTL tissue
548 colocalization and T2D index variants within different T2D genetic clusters using Fisher's exact
549 test. We constructed contingency tables based on whether one of the 1,289 T2D index variants
550 is within a specific T2D genetic cluster and is also the lead colocalizing variant for a
551 colocalization with an eQTL dataset. The Liver/Lipid Metabolism cluster was excluded due to
552 only having 3 total T2D index variants. We then adjusted P-values using the Benjamini-
553 Hochberg procedure. We repeated these enrichments analyses for UKB metabQTL
554 colocalizations, grouped based on metabolite class.

555

556 ***Colocalization Network***

557 We constructed a network of all the colocalizations with the T2D meta-analyses using
558 cytoscape⁸² and measured network properties using cytoNCA.

559

560 ***Effect Direction Consistency***

561 We mapped many of the genes and metabolites identified from a pQTL or metabQTL dataset to
562 multiple T2D index variants. We then measured the consistency of effect estimates for T2D and
563 the pQTL or metabQTL for genes or metabolites with at least 3 mapped T2D index variants by
564 testing for an association between pQTL or metabQTL effect sizes with T2D effect sizes using
565 an ordinary least squares regression model without an intercept, adjusting for an FDR of 0.05
566 using the Benjamini-Hochberg procedure.

567

568 ***Expression Dataset Preparation***

569 We downloaded scRNA-seq data from HPAP⁴⁵ analyzed it with Scanpy⁸³ v1.9.3. The published
570 data from HPAP we used in this study was pre-QC'ed, clustered, and annotated, and a full
571 description of the HPAP data processing steps can be found in their publication⁴⁵. Briefly, cells
572 were filtered to have >500 expressed genes per cell and <15% mitochondrial reads. After,
573 ambient RNA was adjusted using SoupX, followed by batch correction with Harmony and
574 clustering using the Leiden algorithm with a resolution of 0.5. Cell types were defined by
575 pseudo-bulking each cell type and identifying marker genes identified with a Wald test. We
576 subset cells from healthy and T2D donors and created pseudobulks per sample by summing
577 raw counts across cells per cell-type. We only kept pseudobulks with at least 5×10^5 counts and
578 from adult individuals (age ≥ 20 years; N=21 for healthy and N=17 for T2D). We performed the
579 DE analysis with edgeR⁸⁴ v3.40.2. We removed lowly expressed genes per cell-type with
580 filterByExpr, which removes genes with less than 10 counts in 10 samples or 15 counts in all
581 samples, using diabetes status as a stratification covariate. After filtering, we tested a total of

582 9,995 genes in acinar cells, 8,150 genes in active stellate cells, 13,171 genes in alpha cells,
583 13,095 genes in beta cells, and 9,422 genes in ductal cells. We then fit a generalized linear
584 model with robust dispersion to the data, using as covariates diabetes status, age, sex, sex and
585 age interaction, self-reported race and ethnicity, and sequencing chemistry. We next used the
586 glmLRT function to identify T2D differentially expressed genes incorporating the dispersion
587 estimates and adjusted the p-values with the Benjamini-Hochberg procedure.

588

589 We pulled bulk expression dataset for human islets from Marselli et al.⁴⁴, including 28 T2D
590 cases and 58 ND controls. We performed transcript quantification using salmon 1.4.0⁸⁵, and
591 differential expression tests using DESeq2 1.38.3⁸⁶. The DESeq2 results from this study are
592 accessible on GEO under accession number GSE159984.

593

594 We downloaded bulk expression data of *FXRD* from different iPSC differentiation stages from
595 GEO under the accession number GSE190727. We also downloaded single-cell expression
596 data for stem-cell-derived endocrine cells from Krentz et al. 2018⁴², Xin et al. 2018²⁸, and
597 Balboa et al. 2022⁴³.

598

599 ***Open Targets Drug Identification***

600 To identify potential drug repurposing options from our candidate gene list, we queried data files
601 from Open Targets⁸⁷ on around 7k drug molecules and around 62k targets, where 4,930 drugs
602 were annotated with 1,538 gene targets. After also retrieving the reported indications of drugs,
603 we refined our search to the approved drugs, which were not withdrawn and did list indications.
604 Open Targets used drug annotations from ChEMBL to define drugs as being approved for an
605 indication if they come from a source of approved drug information (e.g. FDA, WHO ATC, EMA,
606 BNF). To test if diabetes drugs were identified more often than expected by chance, we
607 compared the drugs targeting our query genes and those targeting all documented non-query

608 genes using chi-squared tests. Additionally, we tested if permuted 100k random samples from
609 the same gene list would identify diabetes drug targets at the same rate, sampling 42,539
610 known genes with HGNC gene symbols and comparing the results from our query gene set and
611 the random samples.

612

613 **Acknowledgement**

614 A.H. is supported by the American Diabetes Association grant #11-23-PDF-35. O.B. has
615 received funding from the European Union's Horizon 2020 research and innovation programme
616 under Grant Agreement No 101017802 (OPTOMICS). A.P. is supported by the Wiener-Anspach
617 Foundation, and the Fonds National de la Recherche Scientifique (FNRS). J.B.M is supported
618 by NIDDK (U01DK078616 and R01DK078616). J.M.M. is supported by American Diabetes
619 Association Innovative and Clinical Translational Award 1-19-ICTS-068, American Diabetes
620 Association grant #11-22-ICTSPM-16, by NHGRI (U01HG011723), NIDDK (UM1-DK078616)
621 and Medical University of Bialystok (MUB) grant from the Ministry of Science and Higher
622 Education (Poland). This study also supported in part by the National Center for Advancing
623 Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and
624 Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the
625 Southern California Diabetes Endocrinology Research Center. Infrastructure for the CHARGE
626 Consortium is supported in part by the National Heart, Lung, and Blood Institute (NHLBI) grant
627 R01HL105756. This research was conducted in part using data and resources from the
628 Framingham Heart Study of the National Heart Lung and Blood Institute of the National
629 Institutes of Health and Boston University School of Medicine. The Framingham Heart Study
630 (FHS) acknowledges the support of contracts NO1-HC-25195, HHSN268201500001I and
631 75N92019D00031 from the National Heart, Lung and Blood Institute and grant supplement R01
632 HL092577-06S1 for this research. We also acknowledge the dedication of the FHS study
633 participants without whom this research would not be possible.

634

635 **Disclosures**

636 J.B.M is an academic associate for Quest Diagnostics Inc. Endocrine R&D. MIMcC is now an
637 employee of Genentech and a holder of Roche stock.

638

639 **Contributions**

640 R.M. Performed eQTL colocalization analyses, integrated all the colocalization results, built the
641 omics network, and characterized the network. Lead the writing of the manuscript and design of
642 figures. Lead the interpretation and discussion of the results, deep dives into specific locus and
643 wrote the paper.

644 K.M.L. Contributed to design the colocalization pipeline. Performed colocalization analyses with
645 GWAS traits. Integrated the colocalization results with the previously described T2D genetic
646 clusters. Contributed to interpretation and discussion of the results, deep dive into specific
647 locus, design of the figures, discussion of the narrative and writing of the paper.

648 X.Y. Performed colocalization analyses with deCODE pQLT, UKB pQLT, METSIM metabQTL.
649 Contributed to interpretation and discussion of the results, deep dive into specific locus, design
650 of the figures, discussion of the narrative and writing the paper.

651 O.B. Performed colocalization analyses with UKB metabQLT. Supervised the enrichment
652 analyses of drug targets using open targets data. Contributed to interpretation and discussion of
653 the results, deep dive into specific locus, design of the figures, discussion of the narrative and
654 writing of the paper.

655 A.H. Performed colocalization analyses with blood eQTLs from diverse ancestries. Identified
656 and evaluated the contribution of eQTL data from diverse ancestries in identifying effector
657 genes as well as the value of using ancestry-specific GWAS meta-analysis in improving the
658 identification of effector genes. Contributed to interpretation and discussion of the results, deep
659 dive into specific locus, design of the figures, discussion of the narrative and writing the paper.

660 A.L.A Performed colocalization analyses with ROSMAP and HELIC pQLT. Contributed to
661 interpretation and discussion of the results, deep dive into specific locus, design of the figures,
662 discussion of the narrative and writing the paper.

663 A.P. Performed the analysis of differential expression in pancreatic islets in T2D and controls
664 and islets in lipoglucotoxic conditions. Performed enrichment analyses of colocalization genes
665 with differentially expressed genes. Contributed to interpretation and discussion of the results,
666 deep dive into specific locus, design of the figures, discussion of the narrative and wrote the
667 paper.

668 S.H. Performed the curation and mapping of effector genes to drugs using open targets data.
669 Performed the enrichment analyses of effector genes with drug targets. Contributed to
670 interpretation and discussion of the results, design of the figures, discussion of the narrative and
671 writing of the paper.

672 K.Y, K.Hr, Y.T., M.L., H.L, M.C, D.L.E. provided and analyzed functional genomics data related
673 from single-cell and pancreatic islets with the FXVD2 locus.

674 L.S., K.S., K.Ha., H.J.T, N.W.R., J.B.M, MIMcC, A.M., M.U., C.N.S, M.B, M.V, J.I.R. and A.P.M
675 Provided feedback throughout the project and revised the paper.

676 R.M., A.P.M, B.V., E.Z and J.M.M conceived and planned the study, supervised and
677 coordinated the analyses, and wrote the manuscript.

678

679 **Figure Legends**

680 **Figure 1: Overview of Project. a)** Genomic regions containing the 1,289 T2D-associated index
681 variants from Suzuki et al.¹⁵ were tested for evidence of colocalization with 10 eQTL datasets, 4
682 pQTL datasets, 2 metabQTL datasets, and 46 related trait GWAS. Colocalizations were then
683 mapped to an interactive network to visualize genes (from an eQTL or pQTL, colored green),
684 metabolites (colored orange), or traits (colored purple) with evidence of sharing a causal variant
685 with T2D around a index variant (colored yellow). These results were then used in downstream

686 analyses to identify enrichment with expression datasets, better understand T2D pathways, and
687 identify drug targets for T2D. **b)** Upset plot of the 716 T2D index variants mapped to an
688 association in an eQTL, pQTL, metabQTL, or trait GWAS dataset (with a colocalization PP.H4
689 >0.8, Methods). **c)** Bar plot of the number of effector genes, metabolites, and traits identified
690 from colocalization analyses with T2D.

691 **Figure 2: Identification of putative effector genes for T2D.** **a)** Plotted are the number of
692 effector genes for T2D previously identified from colocalization analyses between various T2D
693 GWAS and eQTL datasets. Colors indicate the tissue type of the eQTL dataset and shape
694 indicates major self-reported population group of the eQTL dataset. Gray bars represent the
695 total number of unique transcripts across all colocalization analyses per GWAS. **b)** Upset plot of
696 the variant to gene mappings identified in each eQTL dataset analyzed. **c)** Proportion of
697 colocalizations with evidence in one tissue (PP.H4 >0.8) and no positive evidence observed in
698 other tissues (PP.H4 <0.3). **d)** Example of a colocalization observed only in Pancreatic Islets,
699 for the gene *SCTR*. Colors indicate LD in EUR populations from 1000G relative to rs2244214.

700 **Figure 3: pQTL colocalizations identified in multiple datasets.** **a)** Upset plot of variant to
701 gene mappings identified via colocalization analyses with four different pQTL datasets. **b)**
702 Correlation of pQTL effect sizes for colocalizations identified in both UKB and deCODE pQTL
703 datasets (Pearson R=0.93; P=9.1×10⁻¹⁷). **c)** Locus compare plot of *CBLN4* using deCODE
704 pQTL data. **d)** Colocalization subnetwork of rs1415287, containing the colocalizations with
705 *CBLN4* as well as a colocalization with *IGFBP-1*. Green nodes represent genes from an eQTL
706 or pQTL, orange nodes represent metabolites, purple nodes represent traits, and yellow nodes
707 represent T2D index variants. Size of the nodes indicate the number of colocalizations observed
708 in the full network. Gray edges represent colocalizations with a plasma/blood dataset, green
709 edges represent colocalizations with a subcutaneous adipose eQTL dataset, and pink edges
710 represent colocalizations with a trait. Dashed lines indicate colocalizations are in the opposite

711 direction as T2D risk and solid lines indicate colocalizations are in the same direction as T2D
712 risk.

713 **Figure 4: eQTL from understudied populations identify novel colocalizations with T2D. a)**

714 Upset plot of colocalizations identified with blood eQTL datasets from four different populations.

715 Orange bars represent colocalizations observed only in the Puerto Rican (PR), Mexican

716 American (MX), or African American (AA) datasets (PP.H4 >0.8) and not in the European (EUR)

717 dataset (PP.H4 <0.3). Blue bars represent colocalizations observed only in the EUR dataset

718 (PP.H4 >0.8) and not in PR, MX, or AA datasets (PP.H4 <0.3). Gray bars represent

719 colocalizations observed in any of the PR, MX, or AA datasets and showed a PP.H4 between

720 0.3 and 0.8 in the European dataset, or were observed in the European dataset but showed a

721 PP.H4 between 0.3 and 0.8 in any of the PR, MX, or AA datasets. **b)** Log-fold change of allele

722 frequencies between AMR-like and EUR-like populations for T2D index variants with a

723 colocalization observed in one population group. **c)** Locus compare plots of *LIN7A* and *ACSS3*

724 with MX data. Colors indicate LD in AMR continental ancestry from 1000G relative to variant

725 rs10128882. **d)** Allele frequencies of lead colocalizing variant rs10128882 (blue) and T2D index

726 variant rs11114650 (black) per continental ancestry and per inferred local ancestry among AMR

727 participants from gnomAD v4.0. **e)** Effect sizes of rs10128882 (blue) and rs11114650 (black) in

728 both the T2D GWAS and blood eQTL datasets, stratified by ancestry.

729 **Figure 5: Phosphatidylcholine has consistent negative effect directions with T2D risk. a)**

730 Joint scatterplot of metabolites comparing the number of T2D-associated index variants they are

731 mapped to compared to the change in T2D effect size per change in metabQTL effect size

732 across the variants. Points are colored and sized by Benjamoni-Hochberg adjusted p-value. **b)**

733 Increased glucose has consistent correlation with increased T2D risk at colocalizing index

734 variants. **c)** Decreased phosphatidylcholine has consistent correlation with decreased T2D risk

735 at index variants. **d)** Subnetwork of all colocalizations mapped to a Phosphatidylcholine-mapped

736 T2D-associated index variant. Green nodes represent genes from an eQTL or pQTL, orange

737 nodes represent metabolites, purple nodes represent traits, and yellow nodes represent T2D
738 index variants. Size of the nodes indicate the number of colocalizations observed in the full
739 network.

740 **Figure 6: Type 2 Diabetes Clusters Match Trait Colocalizations. a) left panel:** Color
741 indicates the percentage of total SNPs with a colocalization in each cluster; each row was a trait
742 used for clustering. Final rows list total number of SNPs in each cluster and the total % of SNPs
743 in the cluster with at least one colocalization. **right panel:** Color indicates the percentage of
744 colocalizing SNPs who have the same (red) or opposite (blue) effect direction as the linked T2D
745 SNP, with white indicating that SNPs are split 50/50 for direction, and gray indicating there were
746 no colocalizations between that trait and any SNPs in that cluster. Text in each square indicates
747 the total number of SNPs in that cluster that colocalize with that trait. **b)** Upset plot of all
748 colocalizing SNPs with 2 or more trait group colocalizations showing trait group overlaps, with
749 bars colored by cluster membership. **c)** Upset plot showing trait group overlaps for colocalizing
750 SNPs in the lipodystrophy cluster, with gold bars highlighting overlaps that include 4 or more
751 trait groups (n=19 out of 27 colocalizing SNPs).

752 **Figure 7: Identification of novel drug targets for T2D. a)** Forest plot of chi2 enrichment
753 results between drugs targeting colocalizing genes identified in at least one dataset (all), only
754 from an eQTL, only from a pQTL, and only from a *trans*-pQTL with drugs having an approved
755 indication of diabetes in Open Targets. **b)** Forest plot of enrichment results using drugs having
756 an indication of diabetes. **c)** Locus compare plot of *FXVD2*, identified from a colocalization
757 observed only with pancreatic islet eQTL data. **d)** Violin plot of *FXVD2* expression per rs529623
758 genotype. **e)** Violin plot of *FXVD2* expression among people with T2D and non-diabetes (ND)
759 stratified by weight status in bulk pancreatic islet data. **f)** Violin plot of *FXVD2* expression from
760 human donor single-cell pancreatic islet data stratified by cell-type. Adjusted P-values <0.001
761 are indicated with “***”, and adjusted P-values <0.05 with “**”. Due to the low number of cells per
762 sample, the differential expression test was not performed for some cell types (**Methods**).

763

764 **Extended Data Figure 1: eQTL colocalizations correlate with sample size.** Scatter plot of
765 the number of colocalizations between the T2D multi-ancestry meta-analysis and different eQTL
766 datasets compared to the sample size of the eQTL datasets.

767 **Extended Data Figure 2: Enrichment of pancreatic islet colocalizing effector transcripts**
768 **in differential gene expression data.** GSEA plots of colocalizing effector transcripts with
769 differentially expressed genes from the bulk human pancreatic islet dataset, stratified by
770 whether the T2D index variant risk allele is associated with increases (same direction) or
771 decreases (opposite direction) of gene expression.

772 **Extended Data Figure 3: IGFBP2 associations with T2D are highly consistent genome-**
773 **wide. a)** Joint scatterplot of proteins comparing the number of T2D-associated index variants
774 they are mapped to compared to the association of pQTL effect sizes to T2D effect sizes across
775 the variants. Points are colored and sized by Benjamoni-Hochberg adjusted p-value. **b)**
776 Decreased plasma protein levels of insulin growth factor binding protein 2 *IGFBP2* has
777 consistent correlation with increased T2D risk at colocalizing lead variants.

778 **Extended Data Figure 4: Overlap of T2D clusters with trait, eQTL, and metabQTL**
779 **colocalizations.** Heatmaps of the percent of index variants per cluster mapped to a
780 colocalization with a metabolite group, eQTL dataset, and additional related trait GWAS.

781 **Extended Data Figure 5: Colocalizations between index variants and trait groups per T2D**
782 **genetic cluster.** Upset plots of colocalizations between T2D index variants from different T2D
783 genetic clusters with related cardiometabolic trait groups.

784 **Extended Data Figure 6: Permutations of overlaps from T2D colocalizing genes with**
785 **drugs from Open Targets.** Ratios of approved diabetes drugs of detected drugs compared to
786 100k random samples (histogram) for a) genes from all T2D colocalizations, b) genes only
787 identified from a T2D colocalization with eQTL, c) genes only identified from a T2D colocalization
788 with pQTL, and d) genes only identified from a T2D colocalization with *trans*-pQTL. P-values

789 indicate difference in the approved proportion of diabetes indications per drug for colocalizing
790 query genes relative to random permutations.

791 **Extended Data Figure 7: *FXVD2* expression in islet progenitor cells. a)** Heatmap of *FXVD2*
792 expression throughout various stages of iPSC differentiation into beta-like cells, where colors
793 correspond to the $\log_2(\text{TPM}+1)$ expression level of an individual sample (1-5). The differentiation
794 stages include definitive endoderm (DE), pancreatic progenitor (PP), stem cell-derived islets at
795 stage 7 (SC-islets) and the islets following grafting into immune deficient mice for 4 months
796 (graft). **b)** Dot plot showing mean gene expression of *FXVD2* and *INS* in the endocrine cell
797 clusters in the integrated dataset of 46,261 stem-cell-derived endocrine cells and adult human
798 islet cells (Krentz et al. 2018, Xin et al. 2018, and Balboa et al. 2022). Dot size is relative to the
799 fraction of cells within a cluster expressing the gene. Endocrine Prog., Endocrine progenitors;
800 SC-EC, Stem-cell-derived enterochromaffin-like cells; SC-Beta, Stem-cell-derived beta cells;
801 SC-Alpha, Stem-cell-derived alpha cells.

802

- 803 1. Herman, W. H. & Zimmet, P. Type 2 Diabetes: An Epidemic Requiring Global Attention and
804 Urgent Action. *Diabetes Care* **35**, 943–944 (2012).
- 805 2. Jaacks, L. M., Siegel, K. R., Gujral, U. P. & Narayan, K. M. V. Type 2 diabetes: A 21st
806 century epidemic. *Best Pract. Res. Clin. Endocrinol. Metab.* **30**, 331–343 (2016).
- 807 3. Rusina, P. V. *et al.* Genetic support for FDA-approved drugs over the past decade. *Nat.*
808 *Rev. Drug Discov.* **22**, 864–864 (2023).
- 809 4. Minikel, E. V., Painter, J. L., Dong, C. C. & Nelson, M. R. Refining the impact of genetic
810 evidence on clinical success. *Nature* 1–6 (2024) doi:10.1038/s41586-024-07316-0.
- 811 5. Duffy, Á. *et al.* Development of a human genetics-guided priority score for 19,365 genes and
812 399 drug indications. *Nat. Genet.* **56**, 51–59 (2024).
- 813 6. Wang, X. *et al.* The impact on clinical success from the 23andMe cohort.
814 2024.06.17.24309059 Preprint at <https://doi.org/10.1101/2024.06.17.24309059> (2024).

- 815 7. Karczewski, K. J. *et al.* Systematic single-variant and gene-based association testing of
816 thousands of phenotypes in 394,841 UK Biobank exomes. *Cell Genomics* **2**, (2022).
- 817 8. Wang, Q. *et al.* Rare variant contribution to human disease in 281,104 UK Biobank exomes.
818 *Nature* **597**, 527–532 (2021).
- 819 9. Flannick, J. *et al.* Exome sequencing of 20,791 cases of type 2 diabetes and
820 24,440 controls. *Nature* **570**, 71–76 (2019).
- 821 10. Mahajan, A. *et al.* Fine-mapping type 2 diabetes loci to single-variant resolution using high-
822 density imputation and islet-specific epigenome maps. *Nat. Genet.* **50**, 1505–1513 (2018).
- 823 11. Mahajan, A. *et al.* Multi-ancestry genetic study of type 2 diabetes highlights the power of
824 diverse populations for discovery and translation. *Nat. Genet.* **54**, 560–572 (2022).
- 825 12. Spracklen, C. N. *et al.* Identification of type 2 diabetes loci in 433,540 East Asian individuals.
826 *Nature* **582**, 240–245 (2020).
- 827 13. Suzuki, K. *et al.* Identification of 28 new susceptibility loci for type 2 diabetes in the
828 Japanese population. *Nat. Genet.* **51**, 379–386 (2019).
- 829 14. Vujkovic, M. *et al.* Discovery of 318 new risk loci for type 2 diabetes and related vascular
830 outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat. Genet.* **52**,
831 680–691 (2020).
- 832 15. Suzuki, K. *et al.* Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature*
833 1–11 (2024) doi:10.1038/s41586-024-07019-6.
- 834 16. Mahajan, A. *et al.* Refining the accuracy of validated target identification through coding
835 variant fine-mapping in type 2 diabetes. *Nat. Genet.* **50**, 559–571 (2018).
- 836 17. Broadaway, K. A. *et al.* Loci for insulin processing and secretion provide insight into type 2
837 diabetes risk. *Am. J. Hum. Genet.* **110**, 284–299 (2023).
- 838 18. Viñuela, A. *et al.* Genetic variant effects on gene expression in human pancreatic islets and
839 their implications for T2D. *Nat. Commun.* **11**, 4912 (2020).

- 840 19. Alonso, L. *et al.* TIGER: The gene expression regulatory variation landscape of human
841 pancreatic islets. *Cell Rep.* **37**, (2021).
- 842 20. Nguyen, J. P. *et al.* eQTL mapping in fetal-like pancreatic progenitor cells reveals early
843 developmental insights into diabetes risk. *Nat. Commun.* **14**, 6928 (2023).
- 844 21. Smith, K. *et al.* Multi-ancestry polygenic mechanisms of type 2 diabetes. *Nat. Med.* 1–10
845 (2024) doi:10.1038/s41591-024-02865-3.
- 846 22. Kachuri, L. *et al.* Gene expression in African Americans, Puerto Ricans and Mexican
847 Americans reveals ancestry-specific patterns of genetic architecture. *Nat. Genet.* **55**, 952–
848 963 (2023).
- 849 23. THE GTEx CONSORTIUM. The GTEx Consortium atlas of genetic regulatory effects across
850 human tissues. *Science* **369**, 1318–1330 (2020).
- 851 24. Gloudemans, M. J. *et al.* Integration of genetic colocalizations with physiological and
852 pharmacological perturbations identifies cardiometabolic disease genes. *Genome Med.* **14**,
853 31 (2022).
- 854 25. Solimena, M. *et al.* Systems biology of the IMIDIA biobank from organ donors and
855 pancreatectomised patients defines a novel transcriptomic signature of islets from
856 individuals with type 2 diabetes. *Diabetologia* **61**, 641–657 (2018).
- 857 26. Ferrannini, E. *et al.* Mechanisms of Sodium–Glucose Cotransporter 2 Inhibition: Insights
858 From Large-Scale Proteomics. *Diabetes Care* **43**, 2183–2189 (2020).
- 859 27. Wittenbecher, C. *et al.* Insulin-Like Growth Factor Binding Protein 2 (IGFBP-2) and the Risk
860 of Developing Type 2 Diabetes. *Diabetes* **68**, 188–197 (2018).
- 861 28. Xin, Y. *et al.* Pseudotime Ordering of Single Human β -Cells Reveals States of Insulin
862 Production and Unfolded Protein Response. *Diabetes* **67**, 1783–1794 (2018).
- 863 29. Yang, J. *et al.* Circulating IGFBP-2 levels are inversely associated with the incidence of
864 nonalcoholic fatty liver disease: A cohort study. *J. Int. Med. Res.* **48**, 0300060520935219
865 (2020).

- 866 30. Fahlbusch, P. *et al.* Physiological Disturbance in Fatty Liver Energy Metabolism Converges
867 on IGFBP2 Abundance and Regulation in Mice and Men. *Int. J. Mol. Sci.* **21**, 4144 (2020).
- 868 31. Arruda, A. L., Morris, A. P. & Zeggini, E. Advancing equity in human genomics through
869 tissue-specific multi-ancestry molecular data. *Cell Genomics* **4**, 100485 (2024).
- 870 32. Olsen, O., Wade, J. B., Morin, N., Bredt, D. S. & Welling, P. A. Differential localization of
871 mammalian Lin-7 (MALS/Veli) PDZ proteins in the kidney. *Am. J. Physiol. Renal Physiol.*
872 **288**, F345-352 (2005).
- 873 33. Jia, Z. *et al.* ACSS3 in brown fat drives propionate catabolism and its deficiency leads to
874 autophagy and systemic metabolic dysfunction. *Clin. Transl. Med.* **12**, e665 (2022).
- 875 34. Virtanen, J. K., Tuomainen, T.-P. & Voutilainen, S. Dietary intake of choline and
876 phosphatidylcholine and risk of type 2 diabetes in men: The Kuopio Ischaemic Heart
877 Disease Risk Factor Study. *Eur. J. Nutr.* **59**, 3857–3861 (2020).
- 878 35. Wong, T. H. T. *et al.* A two-sample Mendelian randomization study explores metabolic
879 profiling of different glycemic traits. *Commun. Biol.* **7**, 1–9 (2024).
- 880 36. Opreanu, M. *et al.* The Unconventional Role of Acid Sphingomyelinase in Regulation of
881 Retinal Microangiopathy in Diabetic Human and Animal Models. *Diabetes* **60**, 2370–2378
882 (2011).
- 883 37. Minikel, E. V., Painter, J. L., Dong, C. C. & Nelson, M. R. Refining the impact of genetic
884 evidence on clinical success. 2023.06.23.23291765 Preprint at
885 <https://doi.org/10.1101/2023.06.23.23291765> (2023).
- 886 38. Sadler, M. C., Auwerx, C., Deelen, P. & Kutalik, Z. Multi-layered genetic approaches to
887 identify approved drug targets. *Cell Genomics* **3**, 100341 (2023).
- 888 39. Flamez, D. *et al.* A genomic-based approach identifies FXYP domain containing ion
889 transport regulator 2 (FXYP2)gammaa as a pancreatic beta cell-specific biomarker.
890 *Diabetologia* **53**, 1372–1383 (2010).

- 891 40. Wetzel, R. K., Pascoa, J. L. & Arystarkhova, E. Stress-induced Expression of the γ Subunit
892 (FXVD2) Modulates Na,K-ATPase Activity and Cell Growth*. *J. Biol. Chem.* **279**, 41750–
893 41757 (2004).
- 894 41. Chandra, V. *et al.* The type 1 diabetes gene TYK2 regulates β -cell development and its
895 responses to interferon- α . *Nat. Commun.* **13**, 6363 (2022).
- 896 42. Krentz, N. A. J. *et al.* Single-Cell Transcriptome Profiling of Mouse and hESC-Derived
897 Pancreatic Progenitors. *Stem Cell Rep.* **11**, 1551–1564 (2018).
- 898 43. Balboa, D. *et al.* Functional, metabolic and transcriptional maturation of human pancreatic
899 islets derived from stem cells. *Nat. Biotechnol.* **40**, 1042–1055 (2022).
- 900 44. Marselli, L. *et al.* Persistent or Transient Human β Cell Dysfunction Induced by Metabolic
901 Stress: Specific Signatures and Shared Gene Expression with Type 2 Diabetes. *Cell Rep.*
902 **33**, (2020).
- 903 45. Elgamal, R. M. *et al.* An Integrated Map of Cell Type–Specific Gene Expression in
904 Pancreatic Islets. *Diabetes* **72**, 1719–1728 (2023).
- 905 46. Kopf, D. *et al.* Insulin Secretion and Sensitivity after Single-Dose Amisulpride, Olanzapine or
906 Placebo in Young Male Subjects: Double Blind, Cross-Over Glucose Clamp Study.
907 *Pharmacopsychiatry* **45**, 223–228 (2012).
- 908 47. Glunk, V. *et al.* A non-coding variant linked to metabolic obesity with normal weight affects
909 actin remodelling in subcutaneous adipocytes. *Nat. Metab.* **5**, 861–879 (2023).
- 910 48. Rusu, V. *et al.* Type 2 Diabetes Variants Disrupt Function of SLC16A11 through Two
911 Distinct Mechanisms. *Cell* **170**, 199-212.e20 (2017).
- 912 49. Rottner, A. K. *et al.* A genome-wide CRISPR screen identifies CALCOCO2 as a regulator of
913 beta cell function influencing type 2 diabetes risk. *Nat. Genet.* **55**, 54–65 (2023).
- 914 50. Dwivedi, O. P. *et al.* Loss of ZnT8 function protects against diabetes by enhanced insulin
915 secretion. *Nat. Genet.* **51**, 1596–1606 (2019).

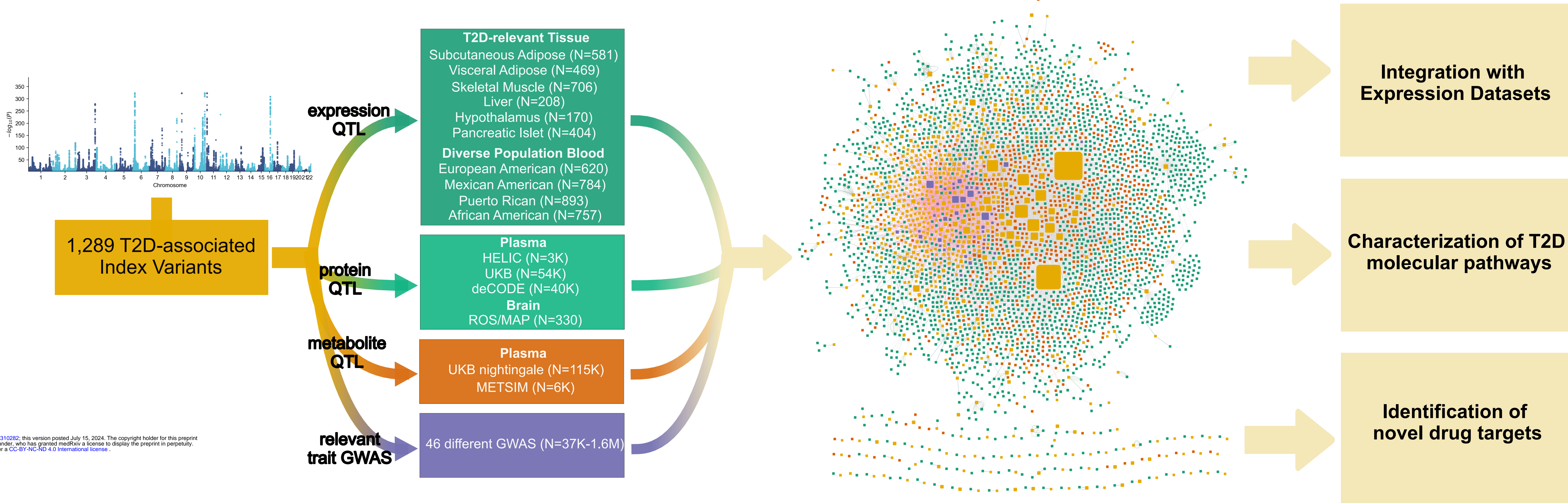
- 916 51. Thomsen, S. K. *et al.* Type 2 diabetes risk alleles in PAM impact insulin release from human
917 pancreatic β -cells. *Nat. Genet.* **50**, 1122–1131 (2018).
- 918 52. Abifadel, M. *et al.* Mutations in PCSK9 cause autosomal dominant hypercholesterolemia.
919 *Nat. Genet.* **34**, 154–156 (2003).
- 920 53. Santer, R. & Calado, J. Familial renal glucosuria and SGLT2: from a mendelian trait to a
921 therapeutic target. *Clin. J. Am. Soc. Nephrol. CJASN* **5**, 133–141 (2010).
- 922 54. Chen, J. *et al.* The trans-ancestral genomic architecture of glycemic traits. *Nat. Genet.* **53**,
923 840–860 (2021).
- 924 55. Graham, S. E. *et al.* The power of genetic diversity in genome-wide association studies of
925 lipids. *Nature* **600**, 675–679 (2021).
- 926 56. Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human
927 phenotypes. *Nat. Genet.* **53**, 1415–1424 (2021).
- 928 57. Sulc, J. *et al.* Composite trait Mendelian randomization reveals distinct metabolic and
929 lifestyle consequences of differences in body shape. *Commun. Biol.* **4**, 1–13 (2021).
- 930 58. Liu, Y. *et al.* Genetic architecture of 11 organ traits derived from abdominal MRI using deep
931 learning. *eLife* **10**, e65554 (2021).
- 932 59. Warrington, N. M. *et al.* Maternal and fetal genetic effects on birth weight and their
933 relevance to cardio-metabolic risk factors. *Nat. Genet.* **51**, 804–814 (2019).
- 934 60. Agrawal, S. *et al.* Inherited basis of visceral, abdominal subcutaneous and gluteofemoral fat
935 depots. *Nat. Commun.* **13**, 3771 (2022).
- 936 61. Evangelou, E. *et al.* Genetic analysis of over 1 million people identifies 535 new loci
937 associated with blood pressure traits. *Nat. Genet.* **50**, 1412–1425 (2018).
- 938 62. Stanzick, K. J. *et al.* Discovery and prioritization of variants and genes for kidney function in
939 >1.2 million individuals. *Nat. Commun.* **12**, 4350 (2021).
- 940 63. Teumer, A. *et al.* Genome-wide association meta-analyses and fine-mapping elucidate
941 pathways influencing albuminuria. *Nat. Commun.* **10**, 4130 (2019).

- 942 64. Hartiala, J. A. *et al.* Genome-wide analysis identifies novel susceptibility loci for myocardial
943 infarction. *Eur. Heart J.* **42**, 919–933 (2021).
- 944 65. Jiang, L., Zheng, Z., Fang, H. & Yang, J. A generalized linear mixed model association tool
945 for biobank-scale data. *Nat. Genet.* **53**, 1616–1621 (2021).
- 946 66. Sandholm, N. *et al.* Genome-wide meta-analysis and omics integration identifies novel
947 genes associated with diabetic kidney disease. *Diabetologia* **65**, 1495–1509 (2022).
- 948 67. van Zuydam, N. R. *et al.* Genome-Wide Association Study of Peripheral Artery Disease.
949 *Circ. Genomic Precis. Med.* **14**, e002862 (2021).
- 950 68. Wuttke, M. *et al.* A catalog of genetic loci associated with kidney function from analyses of a
951 million individuals. *Nat. Genet.* **51**, 957–972 (2019).
- 952 69. Liu, H. *et al.* Epigenomic and transcriptomic analyses define core cell types, genes and
953 targetable mechanisms for kidney disease. *Nat. Genet.* **54**, 950–962 (2022).
- 954 70. Malik, R. *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies
955 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
- 956 71. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of
957 genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
- 958 72. Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics and
959 disease. *Nat. Genet.* **53**, 1712–1721 (2021).
- 960 73. Sun, B. B. *et al.* Plasma proteomic associations with genetics and health in the UK Biobank.
961 *Nature* **622**, 329–338 (2023).
- 962 74. Panoutsopoulou, K. *et al.* Genetic characterization of Greek population isolates reveals
963 strong genetic drift at missense and trait-associated variants. *Nat. Commun.* **5**, 5345 (2014).
- 964 75. Png, G. *et al.* Identifying causal serum protein-cardiometabolic trait relationships using
965 whole genome sequencing. *Hum. Mol. Genet.* **32**, 1266–1275 (2023).
- 966 76. Png, G. *et al.* Mapping the serum proteome to neurological diseases using whole genome
967 sequencing. *Nat. Commun.* **12**, 7042 (2021).

- 968 77. Gilly, A. *et al.* Genome-wide meta-analysis of 92 cardiometabolic protein serum levels. *Mol.*
969 *Metab.* **78**, 101810 (2023).
- 970 78. Robins, C. *et al.* Genetic control of the human brain proteome. *Am. J. Hum. Genet.* **108**,
971 400–410 (2021).
- 972 79. Yin, X. *et al.* Genome-wide association studies of metabolites in Finnish men identify
973 disease-relevant loci. *Nat. Commun.* **13**, 1644 (2022).
- 974 80. Borges, M. C. *et al.* Role of circulating polyunsaturated fatty acids on cardiovascular
975 diseases risk: analysis using Mendelian randomization and fatty acid genetic association
976 data from over 114,000 UK Biobank participants. *BMC Med.* **20**, 210 (2022).
- 977 81. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic
978 Association Studies Using Summary Statistics. *PLOS Genet.* **10**, e1004383 (2014).
- 979 82. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular
980 interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
- 981 83. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression
982 data analysis. *Genome Biol.* **19**, 15 (2018).
- 983 84. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for
984 differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140
985 (2010).
- 986 85. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and
987 bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
- 988 86. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for
989 RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
- 990 87. Ochoa, D. *et al.* Open Targets Platform: supporting systematic drug-target identification and
991 prioritisation. *Nucleic Acids Res.* **49**, D1302–D1310 (2021).
- 992

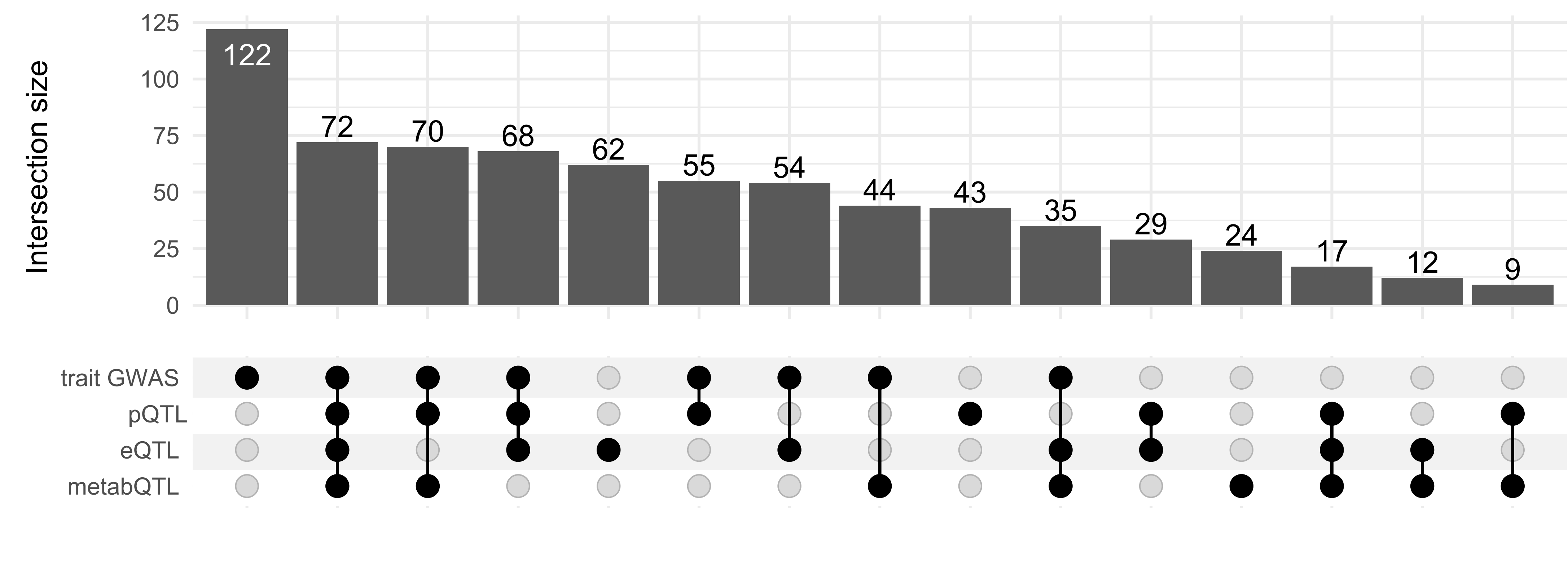
Figure 1

a



medRxiv preprint doi: <https://doi.org/10.1101/2024.07.15.24310282>; this version posted July 15, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/).

b



c

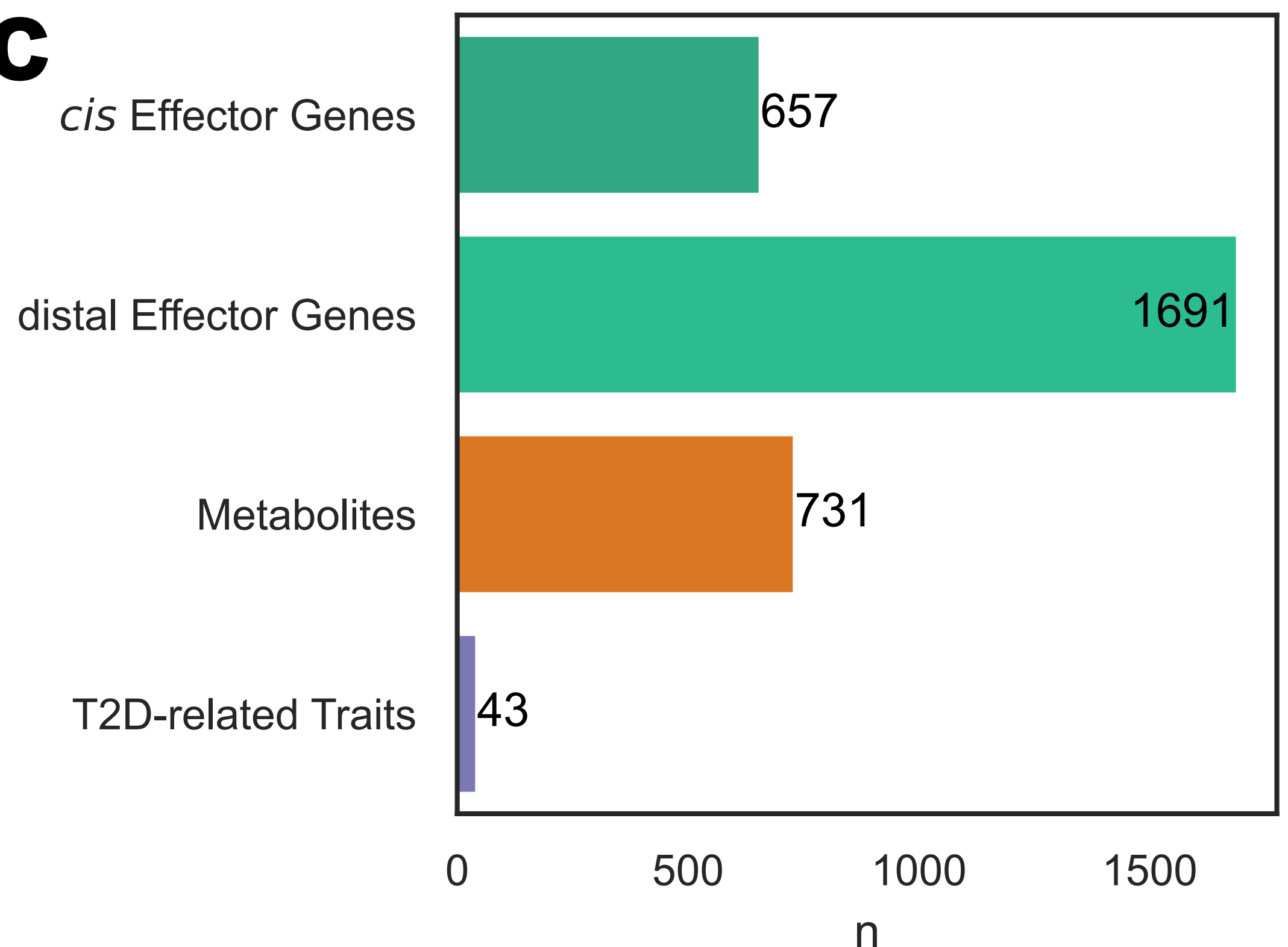
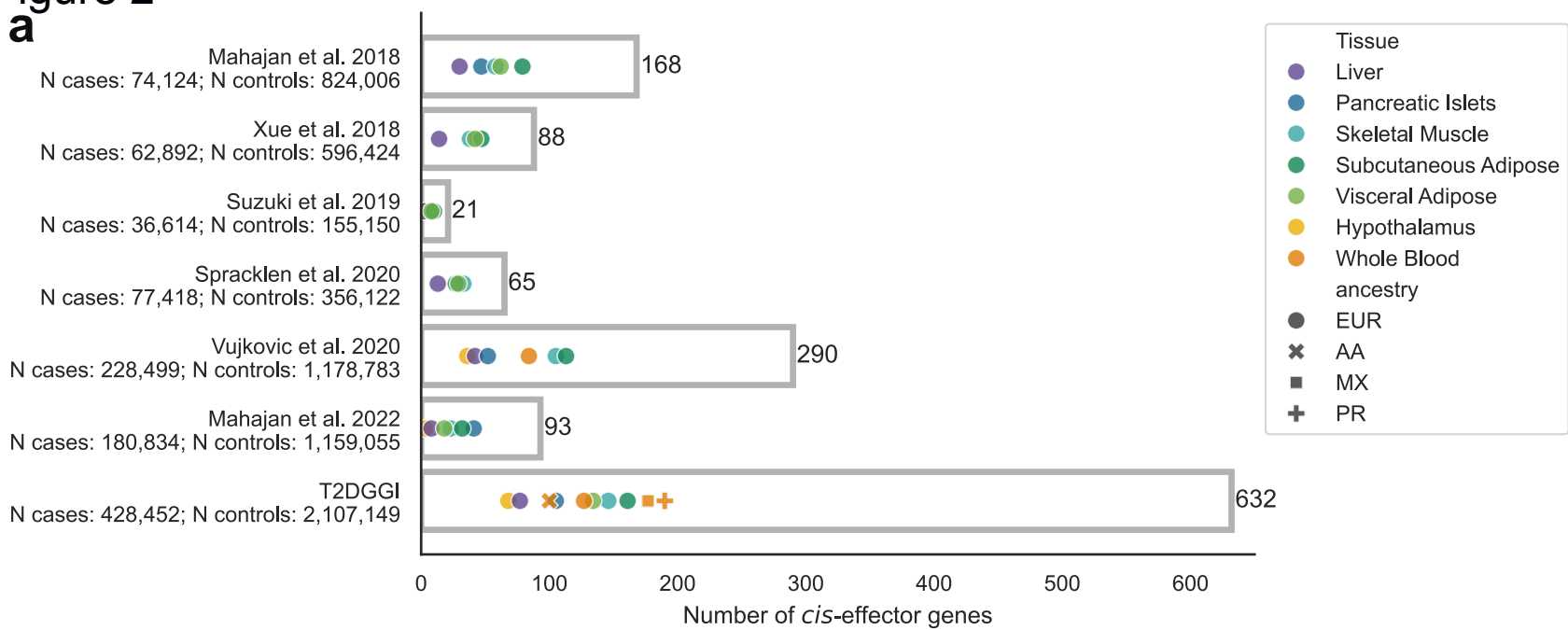
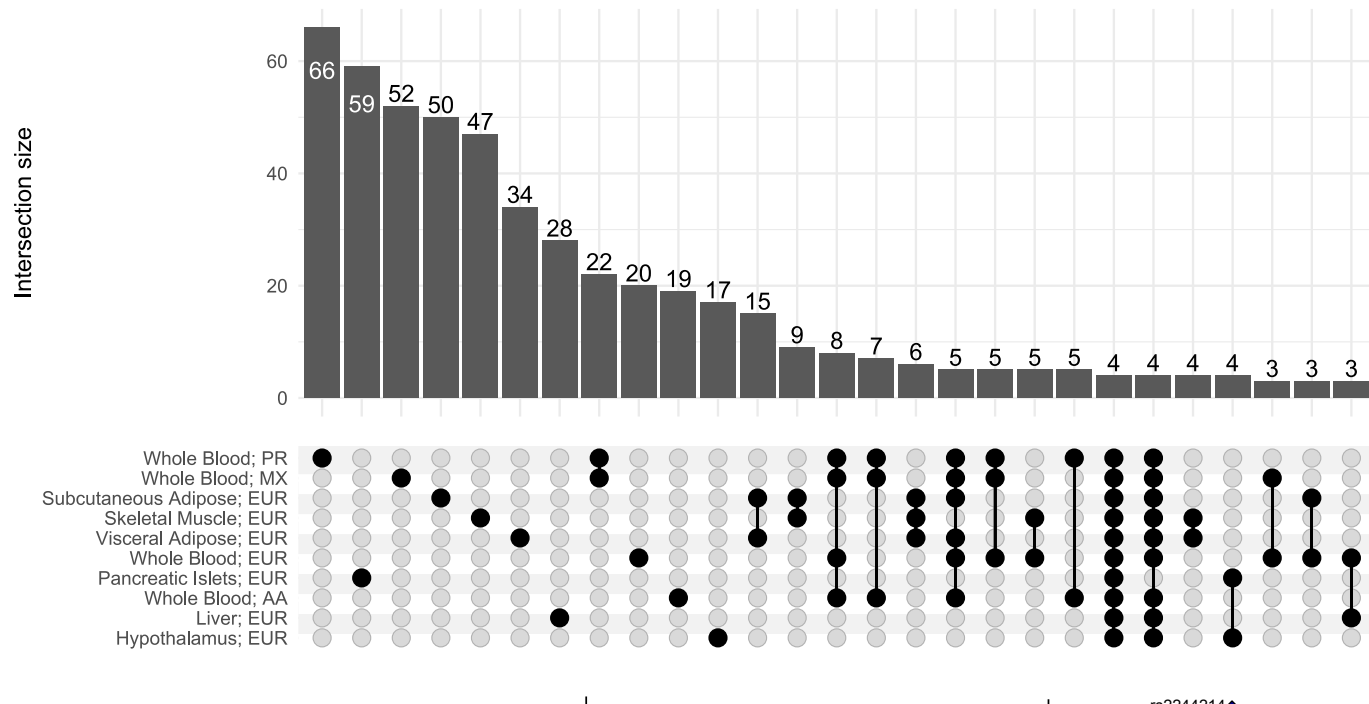


Figure 2

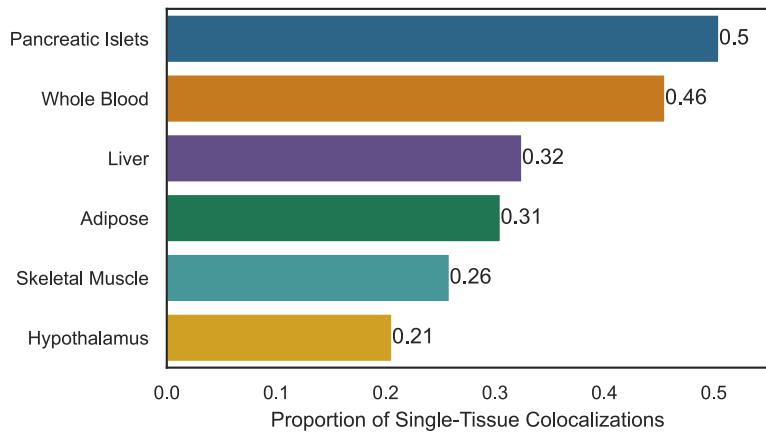
a



b



c



d

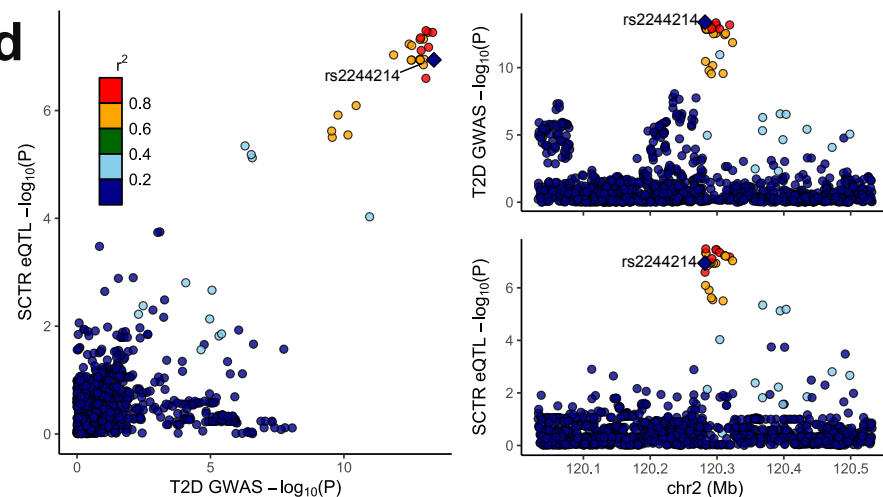
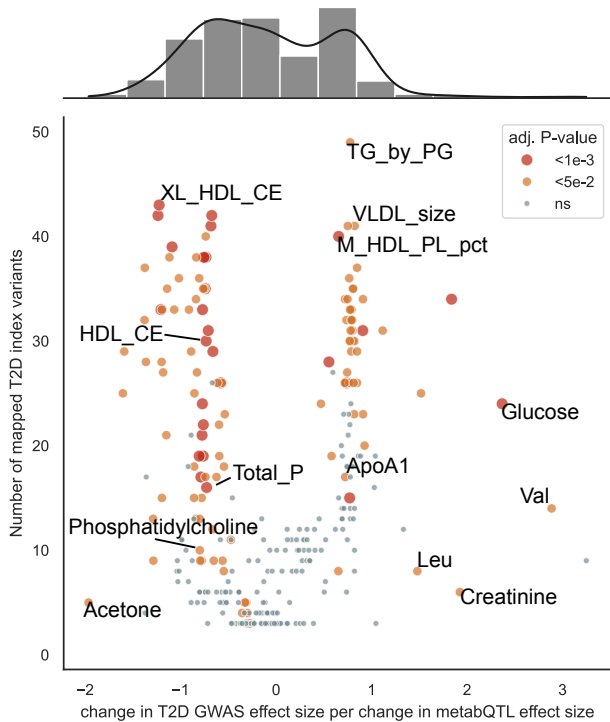
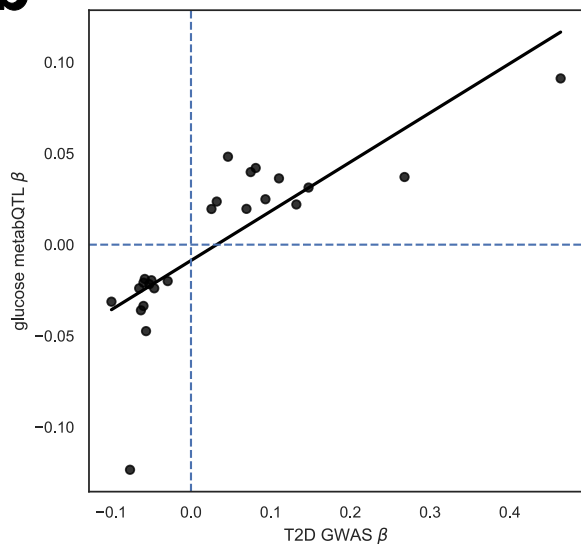


Figure 5

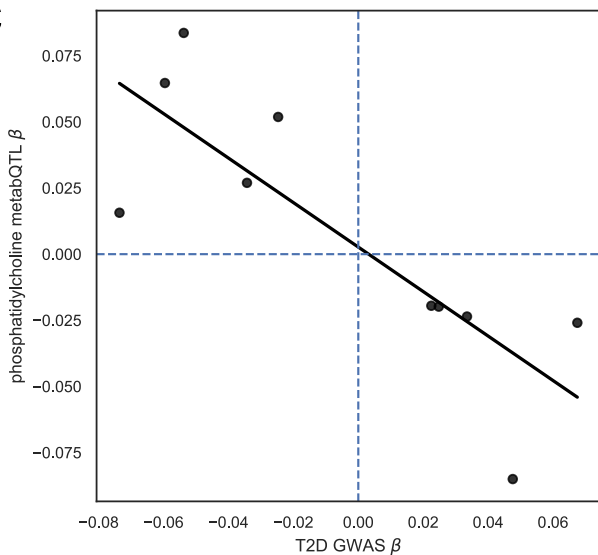
a



b



c



d

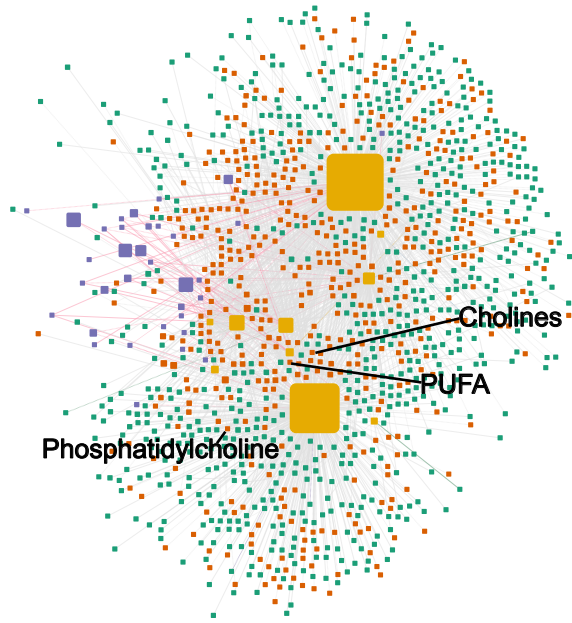
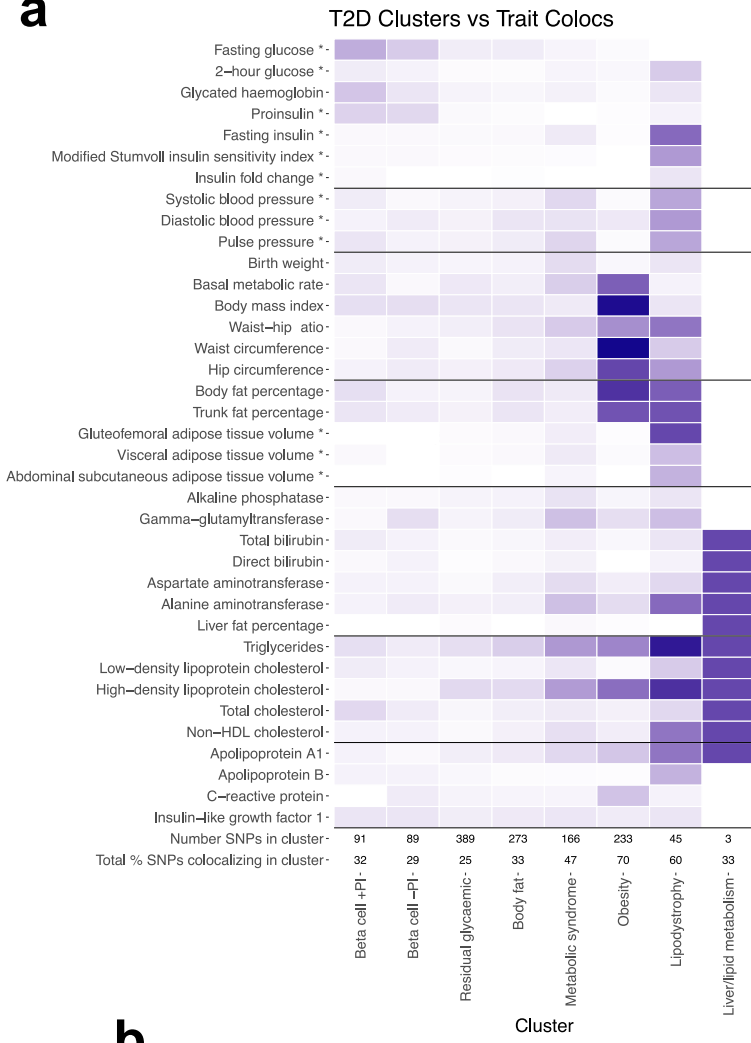
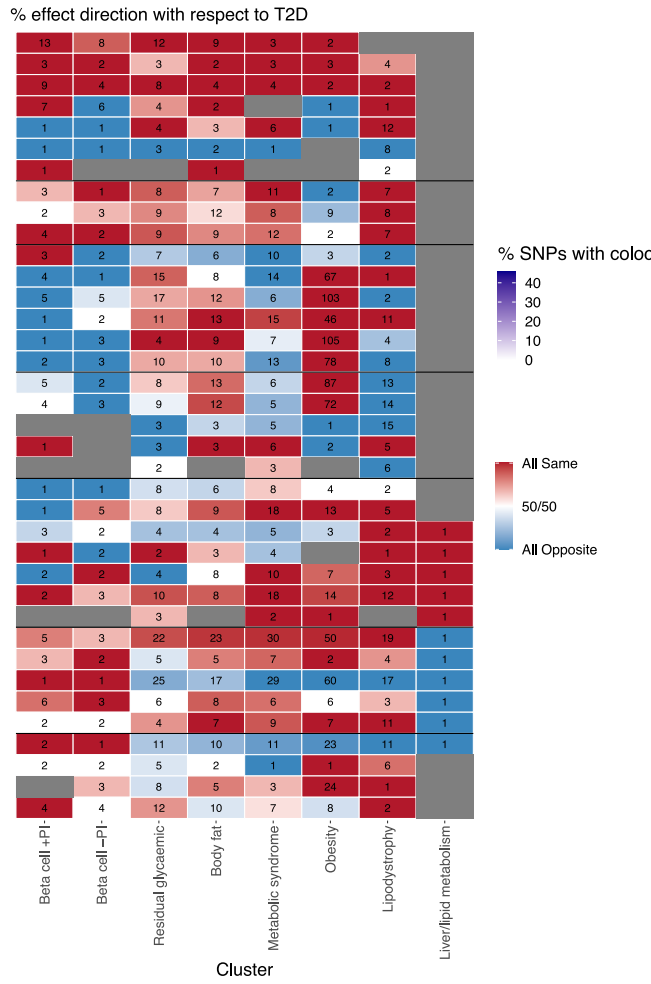


Figure 6

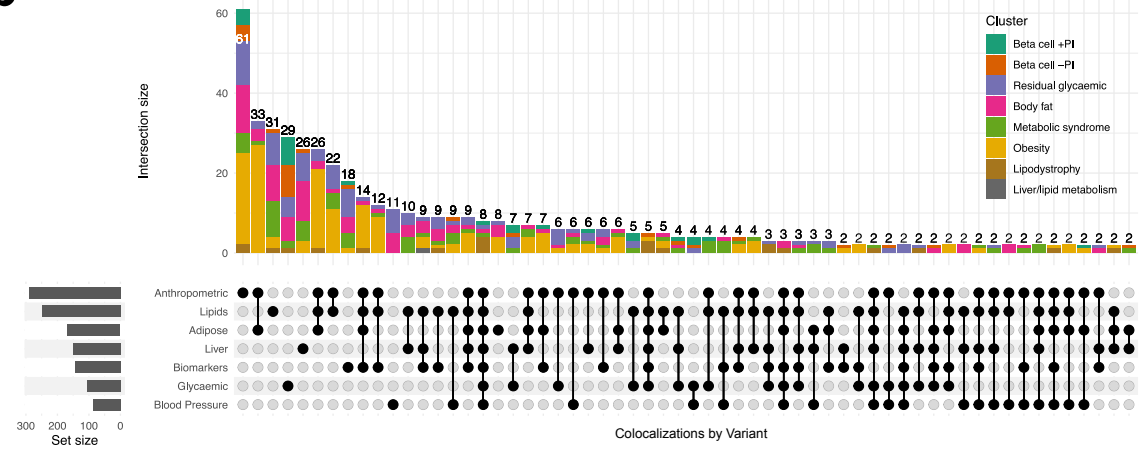
a



T2D Clusters vs Trait Colocs



b



c

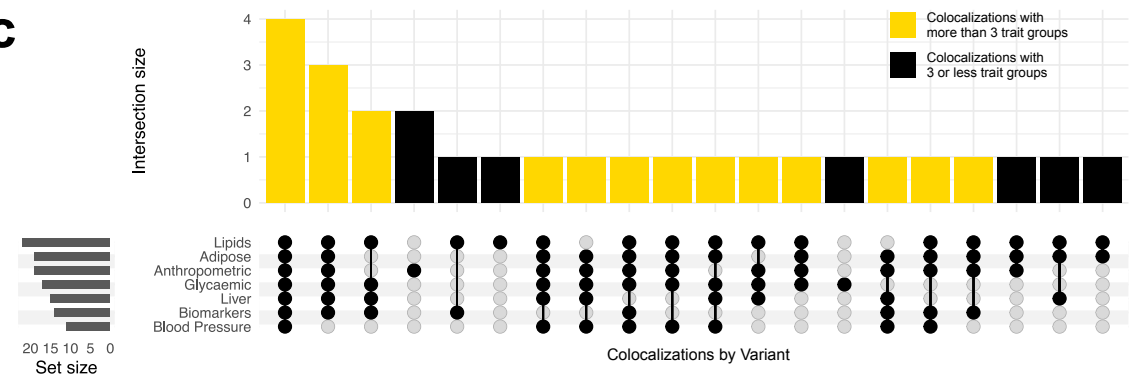
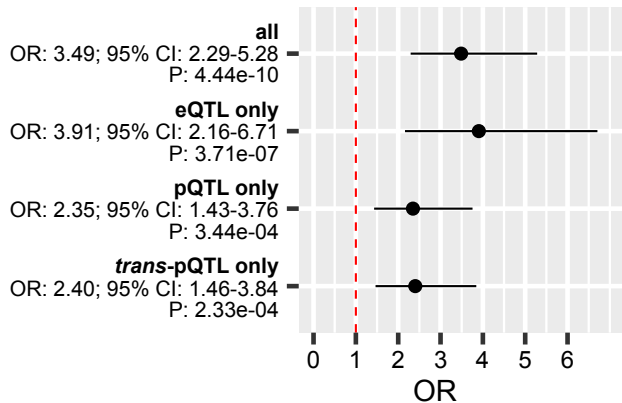


Figure 7

a Approved Diabetes Indication



b Diabetes Indication

