Supplementary Note

Type 2 diabetes patient subgroups

Ahlqvist et al. (26) have proposed to use a set of clinical traits (glutamic acid decarboxylase antibodies (GADA), age at diagnosis, BMI, HbA1c, HOMA-IR and HOMA-B) to cluster patients with newly diagnosed diabetes into subgroups. In the ANDIS cohort, five patient clusters are described (26), four of which are GADA negative and thus represent the type 2 diabetes population.

There are some important differences between the ANDIS and AGES-Reykjavik cohorts that should be noted. While the ANDIS cohort consists of newly diagnosed diabetes patients, the AGES-Reykjavik cohort is a population-based study of elderly people. In AGES-Reykjavik, of 588 patients with prevalent type 2 diabetes with complete data for clustering (excluding GADA measurements), 167 (28%) were diagnosed at the AGES-Reykjavik baseline visit. The implications are that the AGES-Reykjavik cohort includes cases that otherwise would have gone undiagnosed in the population, and that the average age at diagnosis is increased considerably (**Fig. S5a**). Second, the analysis in ANDIS is based on measurements taken shortly after the diagnosis of diabetes, whereas at the AGES-Reykjavik baseline visit the diabetes duration varies (**Fig. S5b**).

We found that using k=5 for the k-means clustering yielded most similar results to those of Ahlqvist et al., as here four of the five subgroups in the AGES-Reykjavik cohort corresponded well to the four GADA-negative subgroups in the ANDIS cohort with regard to clinical characteristics (**Table S13, Fig. S6a**). The fifth group, specific to the AGES-Reykjavik cohort, was characterized by having the lowest age of diabetes onset while also having good metabolic control. This group did not include any newly diagnosed diabetic patients (**Fig. S5c**). As the five groups did not differ substantially with regard to age at the AGES-Reykjavik baseline visit (**Fig. S5d**), this corresponded to subgroup 5 having the longest duration of disease (**Fig. S5e**).

When considering the proportions of each subgroup in AGES-Reykjavik, the largest difference compared to ANDIS was for subgroup 2, or the severely insulin deficient diabetes (SIDD) group, which is proportionally smaller in AGES-Reykjavik (**Table S13**). Thus, one could postulate that the individuals in the AGES-Reykjavik-specific subgroup 5 may have started out in another cluster (such as the SIDD group) at the onset of disease, but progressed to achieve good metabolic control and thus lost the defining pathophysiological characteristics of the original cluster.

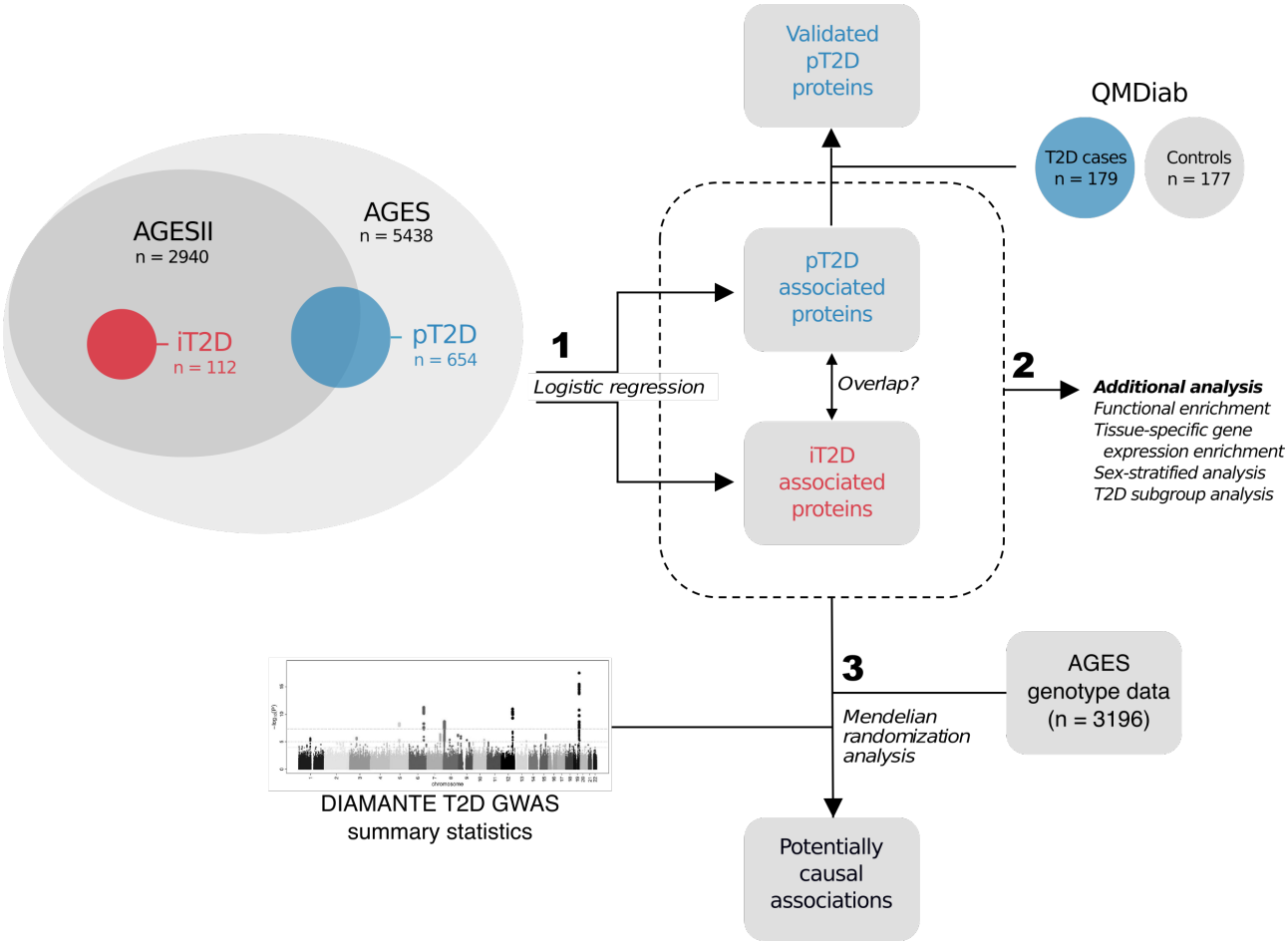

Fig. S1 Workflow of the current study. The top left Venn diagram provides an overview of the AGES cohort, stratified by type 2 diabetes status and follow-up visit participation. The workflow is divided into three major steps; 1) identifying proteins associated with prevalent or incident type 2 diabetes using logistic regression analysis, followed by validation of proteins associated with prevalent type 2 diabetes through similar analysis in the QMDiab cohort, 2) Further characterization of the diabetes-associated proteins, through enrichment and subgroup analyses, and 3) combining genetic data from AGES and summary statistics from the DIAMANTE type 2 diabetes GWAS to screen all diabetes-associated proteins for potential causal relationships using a bi-directional Mendelian randomization analysis. pT2D, prevalent type 2 diabetes; iT2D, incident type 2 diabetes in participants with AGESII follow-up visit.

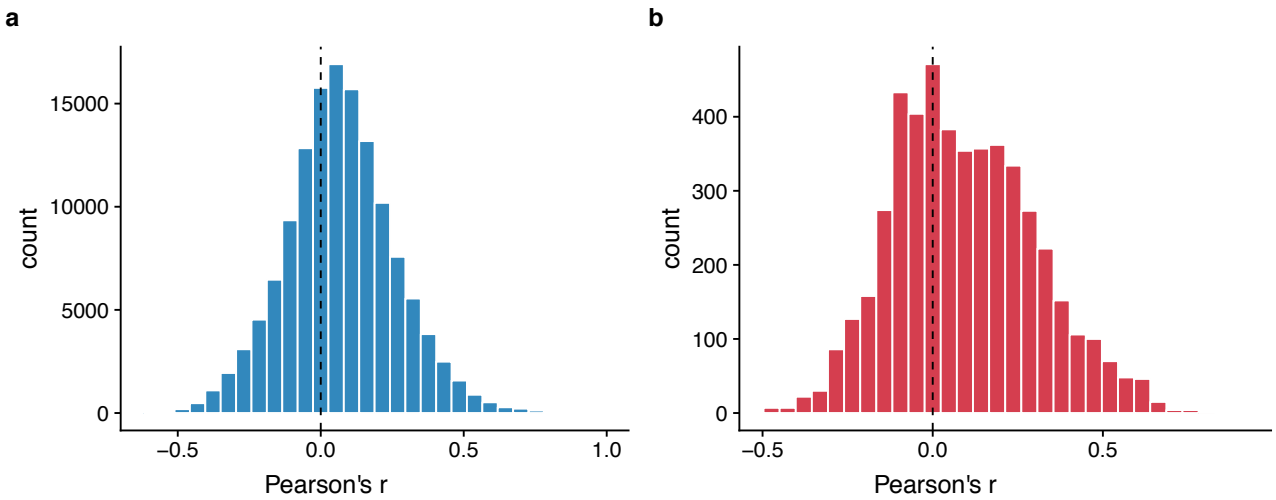

Fig. S2 Distribution of Pearson's correlation coefficients (r) for pairwise correlations between **a)** 520 proteins associated with prevalent type 2 diabetes and **b)** 99 proteins associated with incident type 2 diabetes.

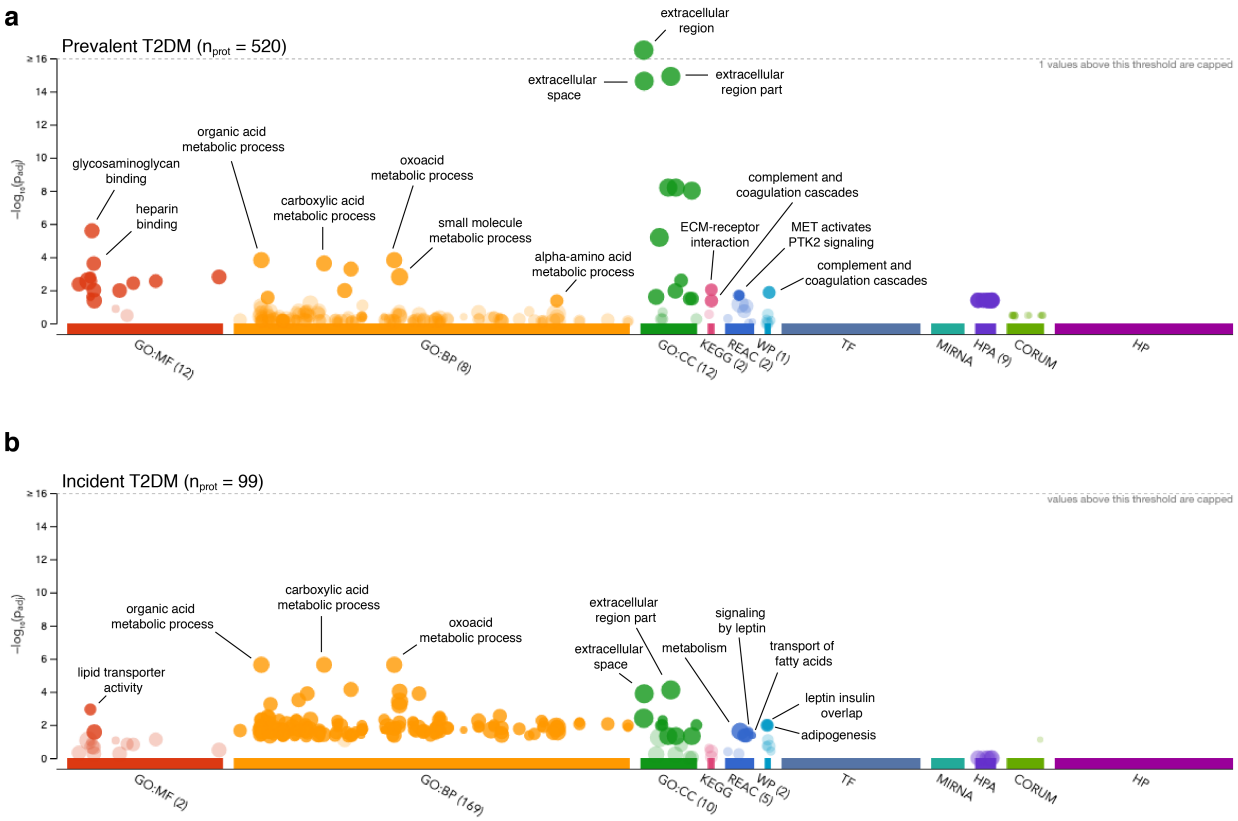

Fig. S3 Functional enrichment results from gProfiler for **a)** 520 proteins associated with prevalent type 2 diabetes and **b)** 99 proteins associated with incident type 2 diabetes, compared to a background of the full SOMAlogic panel.

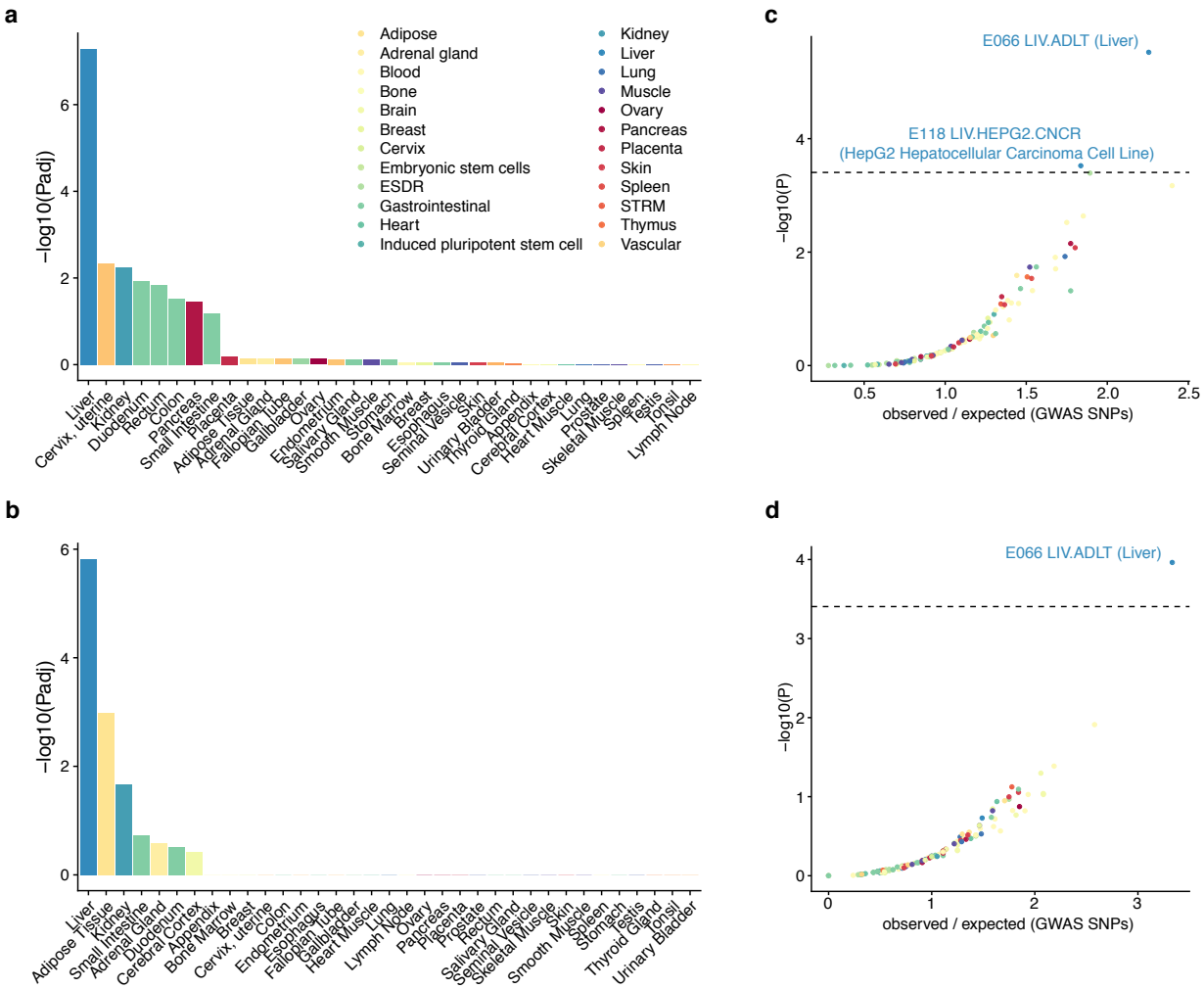

Fig. S4 **a)** Tissue-specific gene expression enrichment for the 520 proteins associated with prevalent type 2 diabetes compared to the full Somalogic panel, **b)** Tissue-specific gene expression enrichment for 99 proteins associated with incident type 2 diabetes compared to the full Somalogic panel, **c)** Cell-type specific enhancer enrichment of genetic variants regulating levels of proteins associated with prevalent type 2 diabetes compared to GWAS SNPs, **d)** Cell-type specific enhancer element enrichment of genetic variants regulating levels of proteins associated with incident type 2 diabetes compared to GWAS SNPs.

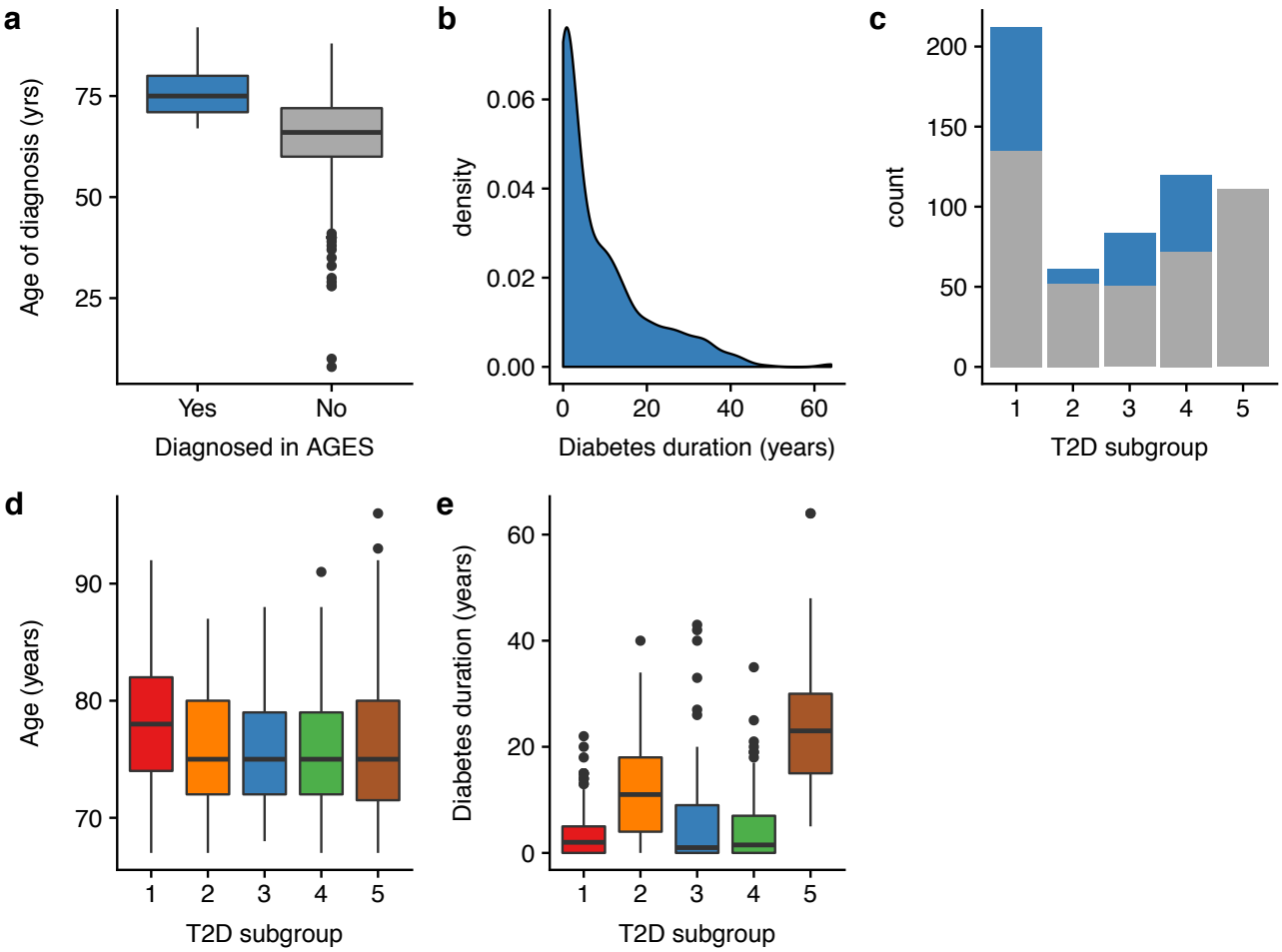

Fig. S5 All panels are restricted to the AGES-Reykjavik participants with prevalent type 2 diabetes and complete data for subgroup clustering (n = 592). **a)** Comparison of age at diabetes diagnosis in those who were diagnosed before (n = 421) or at (n = 171) the AGES-Reykjavik baseline visit, **b)** Density plot of diabetes duration (n = 592), **c)** Proportions of diabetes patients diagnosed before (grey) or at (blue) the AGES-Reykjavik baseline visit in each subgroup, **d)** Age at AGES-Reykjavik baseline visit by subgroup, **e)** Diabetes duration at AGES-Reykjavik baseline visit by subgroup. T2D, type 2 diabetes.

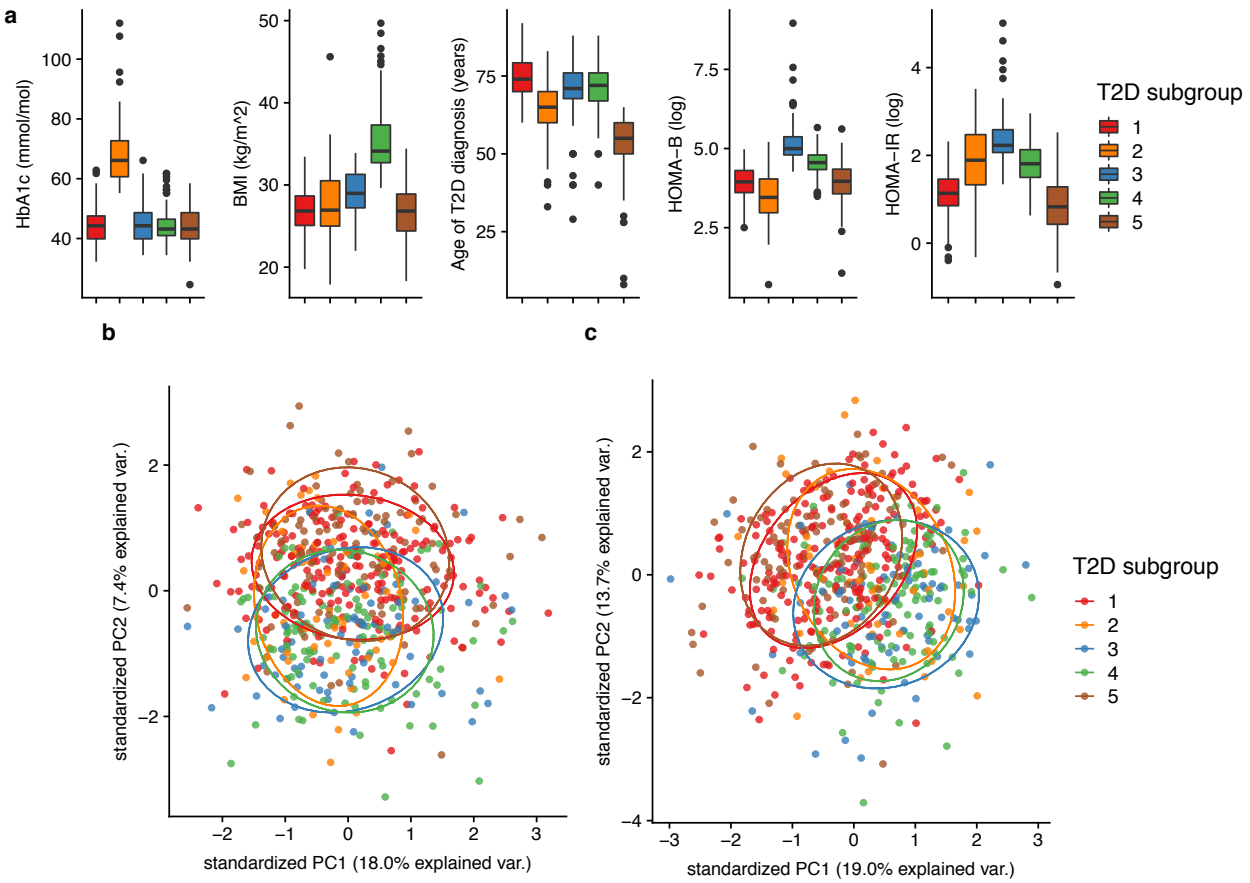

Fig. S6 a) Clinical characteristics of the five subgroups of type 2 diabetes patients in AGES (prevalent cases, n=592 with complete data for clustering), obtained by k-means clustering using the five clinical traits shown in the figure, as proposed by Ahlqvist et al. (26). **b-c)** PCA plot of the type 2 diabetes patients in AGES based on **b)** the 520 proteins associated with prevalent type 2 diabetes or **c)** the 99 proteins associated with incident type 2 diabetes. The colors indicate the type 2 diabetes subgroups obtained by clustering on clinical traits.

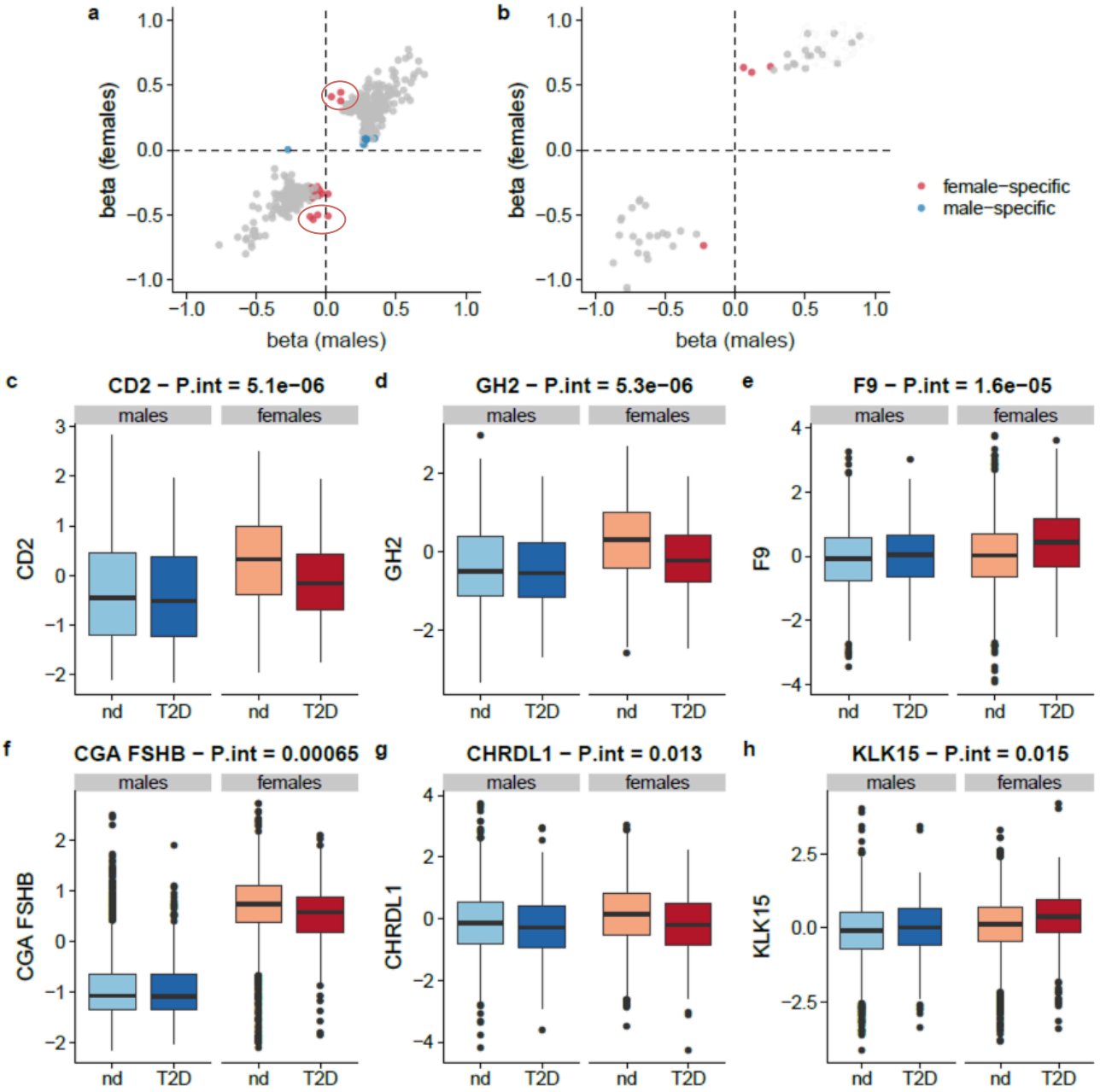

Fig. S7 Comparisons of beta coefficients in males and females for SOMAmers that were significantly (P<0.05/4782) associated with **a)** prevalent or **b)** incident type 2 diabetes in at least one sex in the sex-stratified analysis. Proteins defined as having sex-specific associations are highlighted in red (female-specific) and blue (male-specific) and circled proteins in panel a) are further shown in panels c-h). **c-h)** Boxplots comparing levels of the circled proteins in panel a), stratified by sex and prevalent type 2 diabetes status. The P-value for the sex*protein interaction term is shown above each plot.

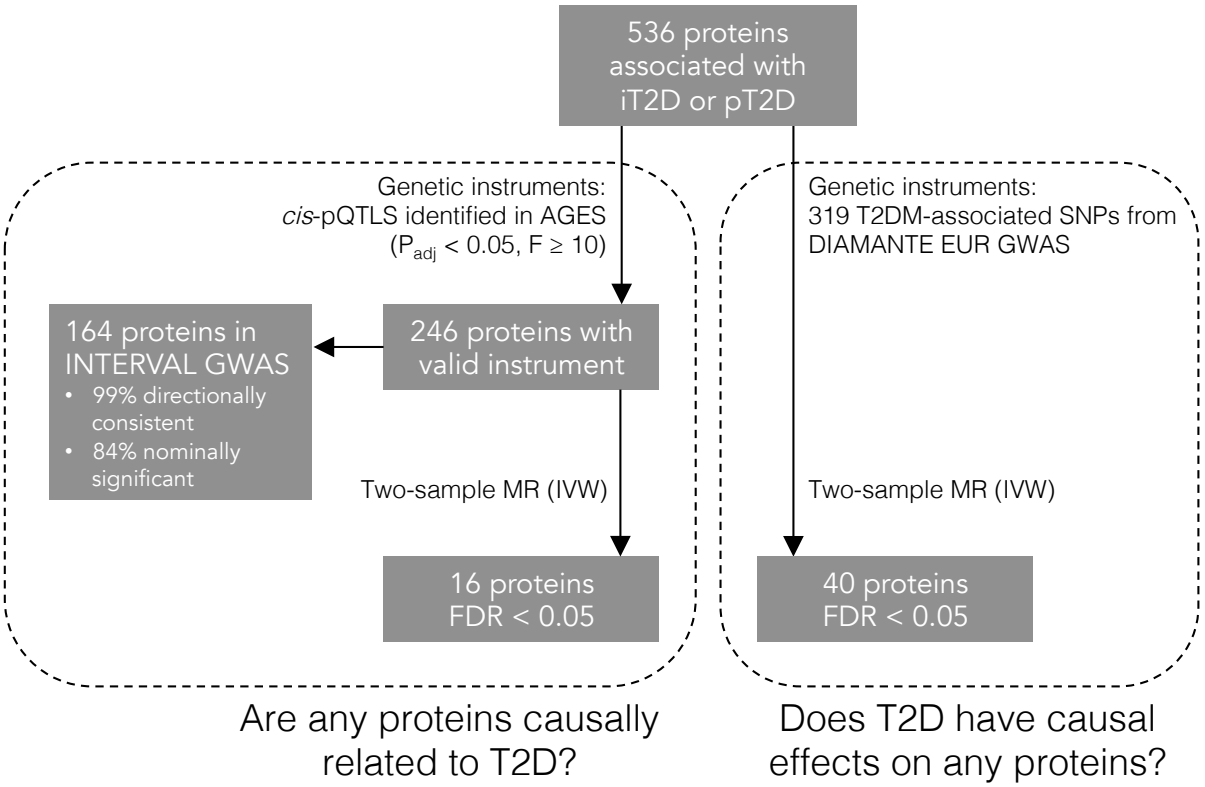

Fig. S8 Flowchart illustrating the main steps of the bi-directional two-sample Mendelian randomization analysis for 536 proteins associated with incident or prevalent type 2 diabetes in the AGES cohort. Here, genetic data (1000 Genomes imputation) for 3,219 AGES participants was used in combination with GWAS summary statistics for type 2 diabetes (DIAMANTE) and protein levels (INTERVAL study).

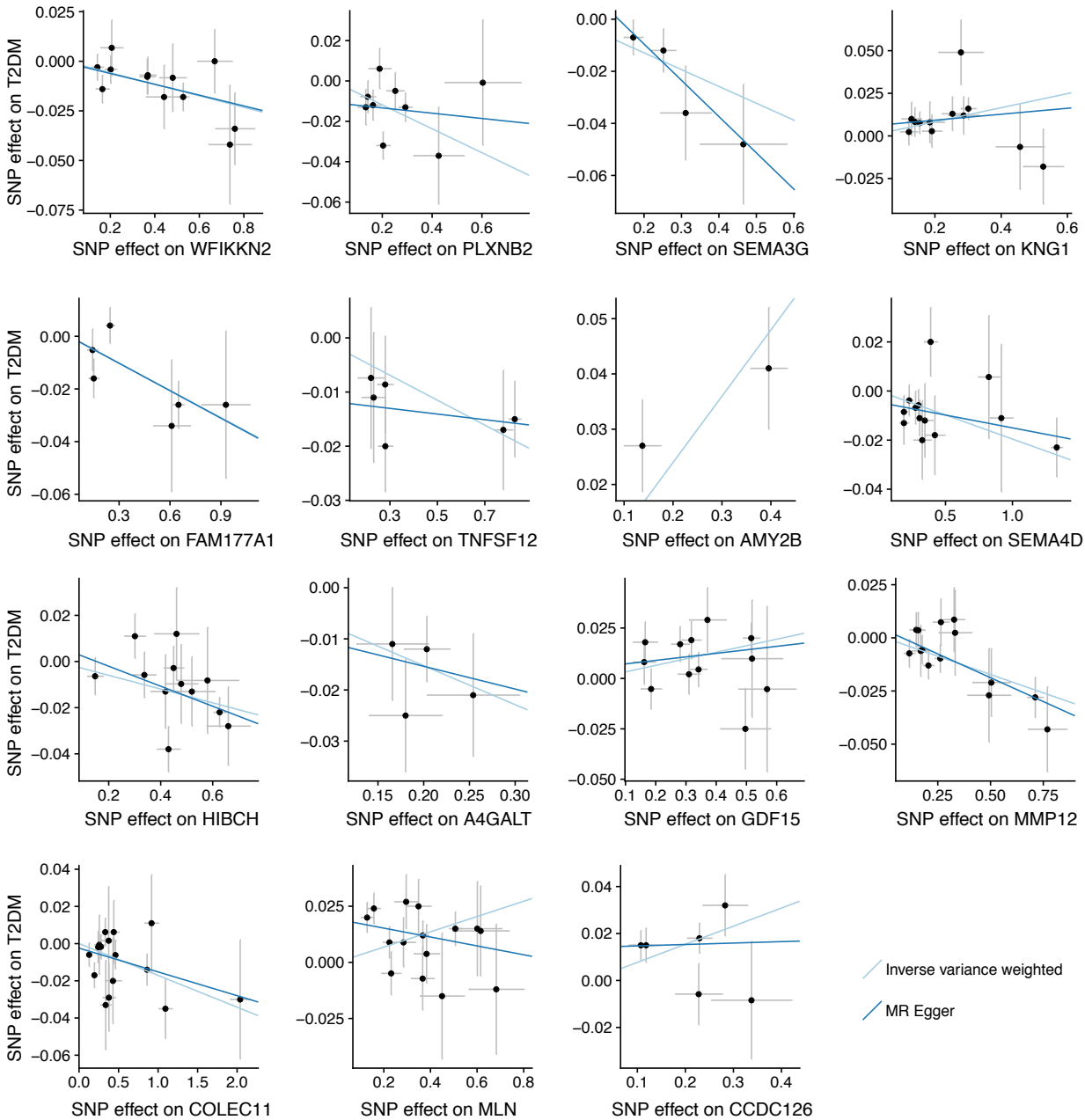

Fig. S9 Scatterplots for 15 proteins supported as having a causal effect on type 2 diabetes in the two-sample MR (FDR<0.05) and with >1 genetic instrument, demonstrating the estimated effects (with 95% confidence intervals) of their respective genetic instruments on the protein levels in AGES (x-axis) and the risk of type 2 diabetes in the DIAMANTE GWAS (y-axis). The blue lines indicate the inverse variance weighted and MR Egger estimates, which are directionally consistent for all except MLN.

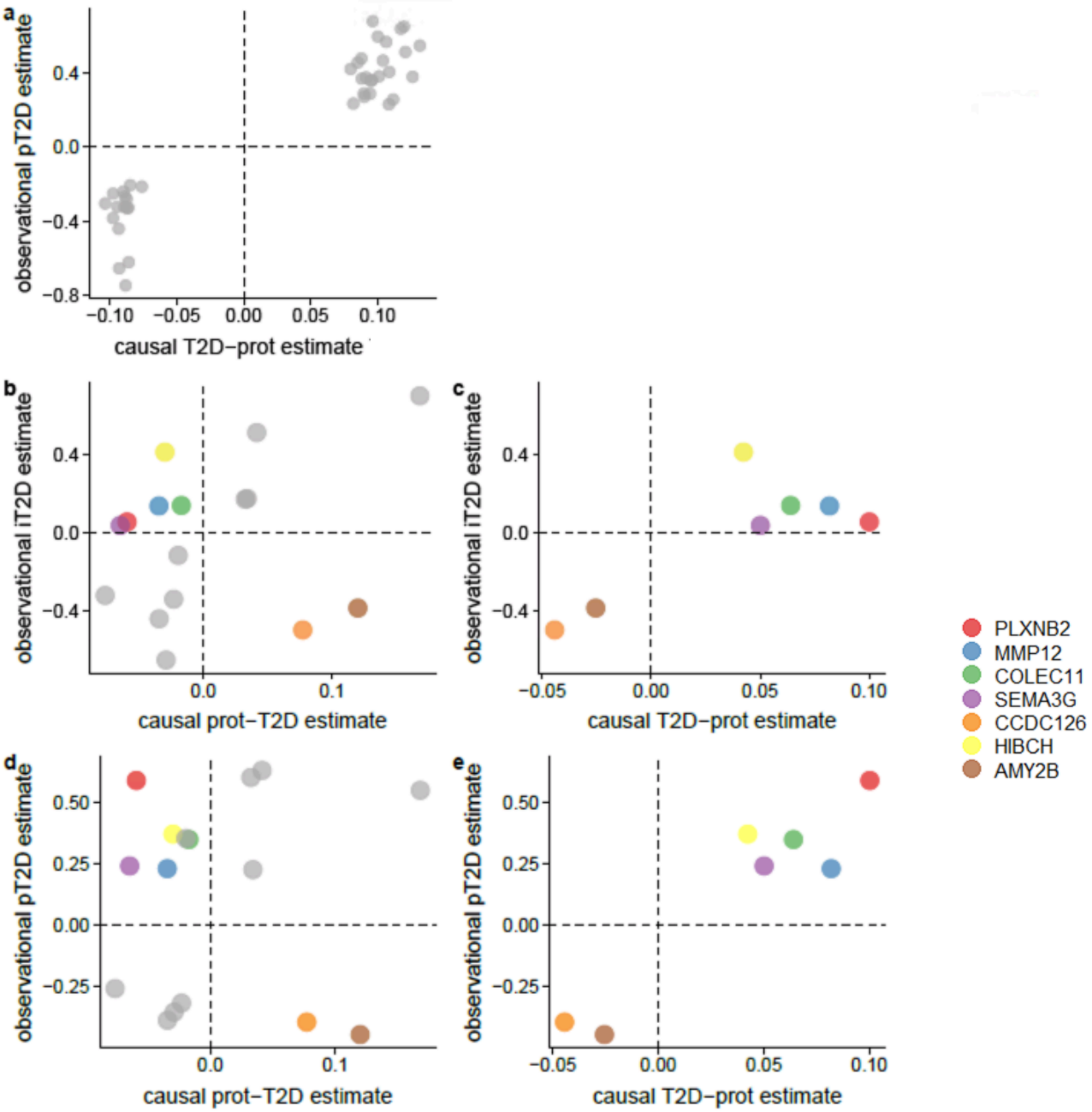

Fig. S10 a) Comparison of observational estimates for prevalent type 2 diabetes and causal estimates for the effect of type 2 diabetes on protein levels, for the 40 proteins that were significant in the type 2 diabetes-protein MR analysis. **b)** Comparison of observational estimates for incident type 2 diabetes and causal estimates for the effect of protein levels on type 2 diabetes, for the 16 proteins that were significant in the protein-type 2 diabetes MR analysis. Seven proteins that do not have consistent direction of effect between observational and causal estimates are colored as indicated in the side legend. **c)** The observational estimates for the seven proteins that were not directionally consistent with the protein-type 2 diabetes causal estimate in panel b) are instead directionally consistent with their type 2 diabetes-protein causal estimates. **d-e)** Same as b-c except using observational estimates for prevalent instead of incident type 2 diabetes.