Title

High-throughput genetic clustering of type 2 diabetes loci reveals heterogeneous mechanistic pathways of metabolic disease.

Journal

Diabetologia, 66(3)

Authors

Kim, Hyunkyung

Westerman, Kenneth

Smith, Kirk

et al.

Publication Date

2023-03-01

DOI

10.1007/s00125-022-05848-6

Peer reviewed

HHS Public Access

Author manuscript

Diabetologia. Author manuscript; available in PMC 2024 March 01.

Published in final edited form as:

Diabetologia. 2023 March ; 66(3): 495–507. doi:10.1007/s00125-022-05848-6.

High-throughput genetic clustering of type 2 diabetes loci reveals heterogeneous mechanistic pathways of metabolic disease

Hyunkyung Kim[1,2,3], **Kenneth E. Westerman**[2,4], **Kirk Smith**[1,2,3], **Joshua Chiou**[5], **Joanne B. Cole**[2,3,6,7], **Timothy Majarian**[2], **Marcin von Grotthuss**[8], **Soo Heon Kwak**[9], **Jaegil Kim**[2,10], **Josep M. Mercader**[1,2,3,6], **Jose C. Florez**[1,2,3,6], **Kyle Gaulton**[5], **Alisa K. Manning**[2,4,6], **Miriam S. Udler**[1,2,3,6]

[1]Diabetes Unit, Massachusetts General Hospital, Boston, MA, USA

[2]Broad Institute of MIT and Harvard, Cambridge, MA, USA

[3]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

[4]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Boston, MA, USA

[5]Department of Pediatrics, University of California San Diego, San Diego, CA, USA

[6]Department of Medicine, Harvard Medical School, Boston, MA, USA

[7]Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA

[8]Takeda Pharmaceuticals, Cambridge, MA, USA

[9]Department of Internal Medicine, Seoul National University Hospital, Seoul, Republic of Korea

[10]Present address: GlaxoSmithKline, Cambridge, MA, USA

Abstract

Aims/hypothesis—Type 2 diabetes is highly polygenic and influenced by multiple biological pathways. Rapid expansion in the number of type 2 diabetes loci can be leveraged to identify such pathways.

Methods—We developed a high-throughput pipeline to enable clustering of type 2 diabetes loci based on variant–trait associations. Our pipeline extracted summary statistics from genome-wide association studies (GWAS) for type 2 diabetes and related traits to generate a matrix of 323 variant × 64 trait associations and applied Bayesian non-negative matrix factorisation (bNMF) to

Corresponding author: Miriam S. Udler, mudler@mgh.harvard.edu.

Contribution statement MSU, JCF and MG conceived the research question. MSU, JK, MG, KEW, HK, KS, JC and KG conceived the methodology, which included implementation of the clustering computational pipeline. HK, JC, MG, TM, JMM and MSU curated the data. HK, KS and JC conducted the analysis and visualised the results. HK, JC and MSU wrote the initial draft of the paper and incorporated co-author comments. KEW, JK, KS, JBC, TM, MG, JMM, SK, JCF, KG and AKM provided feedback on the analysis, and critically reviewed the manuscript. All co-authors approved the final version of the paper. MSU is the guarantor of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

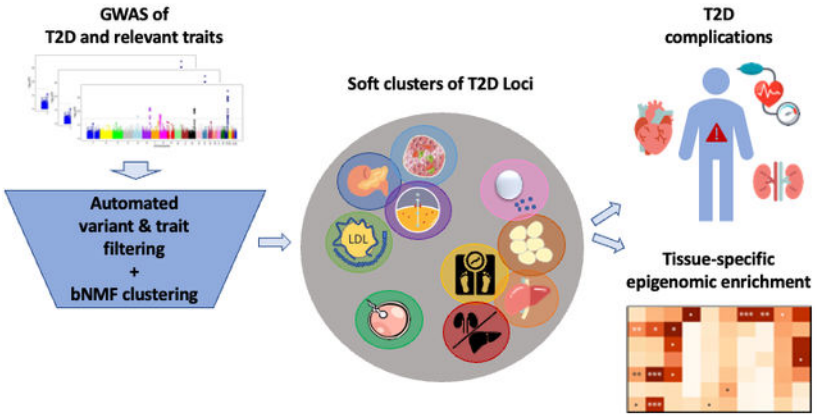
Authors' relationships and activities The authors declare that there are no relationships or activities that might bias, or be perceived to bias, their work.

identify genetic components of type 2 diabetes. Epigenomic enrichment analysis was performed in 28 cell types and single pancreatic cells. We generated cluster-specific polygenic scores and performed regression analysis in an independent cohort (N=25,419) to assess for clinical relevance.

Results—We identified ten clusters of genetic loci, recapturing the five from our prior analysis as well as novel clusters related to beta cell dysfunction, pronounced insulin secretion, and levels of alkaline phosphatase, lipoprotein A and sex hormone-binding globulin. Four clusters related to mechanisms of insulin deficiency, five to insulin resistance and one had an unclear mechanism. The clusters displayed tissue-specific epigenomic enrichment, notably with the two beta cell clusters differentially enriched in functional and stressed pancreatic beta cell states. Additionally, cluster-specific polygenic scores were differentially associated with patient clinical characteristics and outcomes. The pipeline was applied to coronary artery disease and chronic kidney disease, identifying multiple overlapping clusters with type 2 diabetes.

Conclusions/interpretation—Our approach stratifies type 2 diabetes loci into physiologically interpretable genetic clusters associated with distinct tissues and clinical outcomes. The pipeline allows for efficient updating as additional GWAS become available and can be readily applied to other conditions, facilitating clinical translation of GWAS findings. Software to perform this clustering pipeline is freely available.

Graphical Abstract



GWAS: genome-wide association studies; T2D: type 2 diabetes; bNMF: bayesian non-negative matrix factorization

Keywords

Bayesian non-negative matrix factorisation; bNMF; Clustering; Disease pathways; Genetics; GWAS; NMF; Polygenic risk scores; Subtypes; Type 2 diabetes

Introduction

Type 2 diabetes has variable contributions of insulin resistance and beta cell dysfunction, and is influenced by multiple risk factors, including genetics [1]. Untangling the heterogeneity of type 2 diabetes may improve the management of the condition and facilitate precision medicine.

Hundreds of loci associated with type 2 diabetes have been identified in large-scale genetic studies; however, translating these findings to improved understanding of disease pathophysiology has been challenging, largely owing to the abundance of non-protein coding lead variants [2]. Recent studies have leveraged genome-wide association study (GWAS) summary statistics to connect genetic loci to possible disease pathways by clustering loci based on shared patterns of associations across multiple traits [3–6]. In our previous work [5], Bayesian non-negative matrix factorisation (bNMF) soft clustering analysis was performed on 94 genome-wide significant type 2 diabetes variants manually curated from published studies and their associations with 47 diabetes-related traits. We identified five distinct genetic clusters, recognisable as relating to mechanisms of type 2 diabetes pathogenesis. Five similar clusters were independently identified by Mahajan et al, along with a sixth cluster of 'undetermined' physiological impact [4]. Of these five shared clusters, two related to beta cell dysfunction, and the other three to different mechanisms of insulin resistance: obesity-mediated, abnormal lipodystrophy-like fat distribution, and altered hepatic lipid metabolism [4, 5]. Clusters of SNPs can be used to generate partitioned polygenic scores (pPSs), which have been associated with distinct cellular and clinical features [7–10], supporting the notion that these clusters can point to genetic subtypes with specific disease mechanisms.

With new type 2 diabetes loci continuously being discovered and additional GWAS trait summary statistics becoming available, we sought to expand our prior work, which involved manual curation of type 2 diabetes loci. We developed a high-throughput pipeline to enable extraction of genetic variants and traits from multiple GWAS to be used for cluster analysis to identify new genetic pathways of disease.

Methods

Pipeline for input variant–trait association matrix for clustering

An overview of pre-processing steps for variants and traits used for generating the input matrix for variant–trait association clustering analysis is illustrated in ESM Fig. 1 with additional details in the ESM Methods. To obtain a comprehensive set of independent genetic variants associated with type 2 diabetes, we extracted variants reaching genome-wide significance ($p<5\times10^{-8}$) from large-scale type 2 diabetes studies [2, 4, 11–15] in the Accelerating Medicines Partnership-Common Metabolic Disease Knowledge Portal (AMP-CMDKP) [16] (ESM Table 1) and performed stringent LD-pruning of variants at $r^2<0.1$ (ESM Table 2) as well as filtering and replacing of multi-allelic, ambiguous (A/T or C/G), or poorly represented in trait GWAS.

For trait selection, we utilised summary statistics available for 75 GWAS of glycaemic, anthropometric traits, vital signs, and additional laboratory measures in the AMP-CMDKP or UK Biobank [17] (ESM Table 3). Our goal was to let the genetics guide trait-inclusion, and thus traits were used only if the minimum p value across the final set of variants was lower than a Bonferroni p value cut-off of $0.05/N\_final\_variants$ (N=323). We then removed highly correlated traits (with $|r|\geq0.85$) to reduce redundancy. We then used GWAS summary statistics to generate a matrix of standardised and scaled z scores, choosing the type 2 diabetes risk-increasing allele for each variant (ESM Methods). This pipeline was also used

for coronary artery disease (CAD) and chronic kidney disease (CKD) with six CAD GWAS [18–20] and 39 CKD-related GWAS [12, 17, 21–24] queried.

bNMF clustering

The variant–trait association matrix Z (m by n, m: no. of variants, n: no. of traits) was constructed as above. We then generated a non-negative input matrix X (2m by n) by concatenating two separate modifications of the original Z matrix: one containing all positive standardised *z* scores (zero otherwise) and the other all negative standardised *z* scores multiplied by −1. The bNMF procedure factorises X into two matrices, W (2m by K) and H^T (n by K), as X ~ WH with an optimal rank K, corresponding to the association matrix of variants and traits to the number of clusters (ESM Methods) [5]. The key features for each cluster are determined by the most strongly associated variants and traits, a natural output of the bNMF approach. To define a set of strongest-weighted variants and traits in each cluster, we employed a method to determine a weight cut-off that maximised the signal-to-noise ratio (ESM Fig. 2). For type 2 diabetes, the weight cut-off was 0.832.

Cluster associations with relevant phenotypes using GWAS summary statistics

We generated GWAS-partitioned polygenic scores (GWAS pPSs) for each cluster utilising inverse-variance weighted fixed effects meta-analysis of GWAS summary statistics including the set of strongest-weighted variants above the weight cut-off for each cluster using the dmetar package in R [25] (ESM Methods). For testing type 2 diabetes cluster associations with cardiometabolic outcomes, the significance threshold was set to 0.05/(7×K), representing a Bonferroni correction for K clusters and seven outcomes (ESM Table 4).

Functional annotation and enrichment analysis

At each locus, we calculated approximate Bayes factors (aBFs) for all variants within 500 kb with $r^2 \geq 0.1$ with the index variant (100% credible set) using the approach of Wakefield [26] (ESM Methods). We then calculated a posterior probability for each variant by dividing the aBF by the sum of all aBFs in the credible set. We obtained previously published 13-state ChromHMM [27] chromatin state calls for 28 cell types [28]. We also compiled candidate *cis*-regulatory elements (cCREs) for 14 cell types and subtypes from published single-cell chromatin accessibility datasets [29]. We assessed enrichment of annotations within clusters by overlapping 100% credible set variants for signals in each cluster with cell type epigenomic annotations (chromatin states and cCREs). We also assessed epigenomic enrichment in single-cell pancreatic tissue using a second method. As previously described [30], we subset loci from the Beta cell 1 and 2 clusters, annotated variants using cCREs from INS^{high} and INS^{low} beta cells, and applied fgwas [31] in the fine-mapping mode (ESM Methods).

pPS analysis in the Mass General Brigham Biobank

The Mass General Brigham (MGB) Biobank includes clinical and genetic data from patients across the MGB healthcare system [32]. Approval for data analysis was obtained by the MGB Institutional Review Board, study 2016P001018. Description of data quality control

is in the ESM Methods. We performed individual-level analyses on 25,419 participants of European ancestry based on self-reported ancestry and genetic principal components (PC's). Type 2 diabetes pPSs for each cluster were generated by multiplying a variant's genotype dosage by its cluster weight, with only the top-weighted variants included, as defined above. Logistic and linear regression were performed in R v3.6.2, adjusting for age, sex and ten genetic PC's.

Results

Ten type 2 diabetes genetic clusters identified by high-throughput approach

We employed a high-throughput pipeline to enable extraction of loci from GWAS summary statistics files and generate a variant–trait association input matrix for clustering analysis (ESM Methods, ESM Fig. 1). From 13 type 2 diabetes GWAS, we extracted 21,666 genome-wide significant variants and performed stringent LD-pruning and optimisation, resulting in 323 independent type 2 diabetes variants (ESM Methods). These variants guided selection of 64 traits, such that each trait was significantly associated with at least one type 2 diabetes variant. Soft clustering of the resulting 323 by 64 variant–trait association matrix was performed using bNMF.

The plurality of 1,000 bNMF iteration results converged on ten clusters (36.3%), which were the focus of downstream analyses (ESM Tables 5, 6). The clusters were named based on their top-weighted traits or similarity to clusters from our previous work. The remaining bNMF iterations converged on nested clusters (K=6: 0.3%, K=7: 1.1%, K=8: 8.3%, K=9: 26.6%, K=11: 22.6%, K=12: 4.4% and K=13: 0.4%). Six clusters (Beta cell 1, Beta cell 2, Proinsulin, Obesity, Lipodystrophy, Liver/Lipid, as described below) appeared to be captured in all iterations, based on inspection of shared top-weighted variants and traits.

To interpret the ten type 2 diabetes clusters, we examined their strongest-weighted loci and traits, as well as the aggregate associations of cluster loci with the traits via GWAS pPS, with the goal of relating the clusters to driving mechanisms of type 2 diabetes: insulin deficiency and insulin resistance (Fig. 1, ESM Table 7, ESM Fig. 3). Four of the clusters (Beta cell 1, Beta cell 2, Proinsulin, and Lipoprotein A) related to insulin deficiency, with type 2 diabetes risk-increasing alleles in each cluster collectively associated with reduced fasting insulin and HOMA-B (GWAS pPS p values<0.05). Another five clusters (Obesity, Lipodystrophy, Liver/Lipid, ALP [alkaline phosphatase] negative, Hyper Insulin Secretion) related to insulin resistance, with the type 2 diabetes risk alleles in these clusters associated with increased fasting insulin and HOMA-IR (GWAS pPS p values<0.05). The remaining cluster (SHBG [sex hormone-binding globulin]) was driven by one type 2 diabetes allele that was not significantly associated with fasting insulin, but had a positive direction of effect (p = 0.36; Fig. 1, ESM Fig. 3, ESM Table 7).

Of the four clusters related to insulin deficiency (Beta cell 1, Beta cell 2, Proinsulin, Lipoprotein A), Beta cell 1 and Beta cell 2 appeared to be a division of the single Beta cell cluster in our previous work [5], with each containing top traits/loci from that prior cluster (ESM Table 5), including several well-known loci related to beta cell function (e.g. [33]). In Beta cell 1, the top-weighted traits were decreased corrected insulin response (CIR)

and disposition index (DI), both indicators of reduced pancreatic beta cell function; the strongest-weighted loci included *MTNR1B*, *CDKAL1*, *HHEX*, *C2CD4A* and *SLC30A8*. Beta cell 2 cluster's top-weighted traits and loci included increased fasting proinsulin adjusted for fasting insulin (PI), reduced HOMA-B and fasting insulin, and *TCF7L2*, *ADCY5*, *GCK*, *DGKB* and *GLIS3* (Table 1, Fig. 2, ESM Tables 5, 6).

The Beta cell 1 and Beta cell 2 clusters differed from each other with regard to the magnitude of their glycaemic trait associations. The Beta cell 1 GWAS pPS (63 loci) had a more marked association with reduced DI compared with Beta cell 2 ($\beta=-0.05$, $p=3.69\times10^{-61}$ vs $\beta=-0.03$, $p=9.02\times10^{-9}$), while Beta cell 2 (28 loci) had a more marked association with increased PI ($\beta=0.02$, $p=9.81\times10^{-43}$ vs $\beta=0.006$, $p=9.81\times10^{-7}$) (Fig. 1, ESM Table 7). Proinsulin is a prohormone precursor to insulin, and elevated PI levels indicate defective proinsulin processing, particularly related to beta cell stress [34]. The stronger association with increased PI for Beta cell 2 vs Beta cell 1 could therefore indicate that Beta cell 2 relates more specifically to beta cell stress.

The Proinsulin cluster, also captured in our previous work, had top-weighted traits of reduced PI and HOMA-B (Fig. 2, ESM Tables 5, 6). The top-weighted loci included two distinct signals in the *ARAP1/STARD10* region, which has previously been functionally connected to impaired beta cell function in mouse models where beta cell-selective deletion of *Stard10* impaired insulin secretion [35]. In contrast to the other insulin deficiency clusters, this cluster (18 loci) was significantly associated with decreased PI (GWAS pPS $p=3.51\times10^{-36}$) (ESM Table 7), potentially indicating a mechanism of lack of proinsulin substrate for insulin synthesis and secretion.

The Lipoprotein A cluster was novel to the present analysis and had the top-weighted trait, increased serum lipoprotein A [Lp(a)], and single top-weighted locus, *SLC22A3/LPA* (rs487152) (Fig. 2, ESM Tables 5, 6). *SLC22A3/LPA* contains the gene *LPA* encoding Lp(a), and the type 2 diabetes-risk-increasing allele of rs487152 was associated with increased Lp(a) levels ($p=4.06\times10^{-1586}$) (ESM Table 7), but the underlying mechanism relating to insulin deficiency is unknown.

Of the five clusters related to mechanisms of insulin resistance (Obesity, Lipodystrophy, Liver/Lipid, Hyper Insulin Secretion, ALP negative), three (Obesity, Lipodystrophy, and Liver/Lipid) were captured in our previous work, but gained additional loci (and traits) in this expanded analysis.

The Obesity cluster had most strongly weighted traits of increased BMI, waist circumference, per cent body fat, and C-reactive protein (CRP), and key genetic signals included the well-known obesity loci *FTO* and *MC4R* [36] (Fig. 2, ESM Table 5, 6). The GWAS pPS for the Obesity cluster (35 loci) identified significant associations with increased fasting insulin ($p=7.92\times10^{-22}$), HOMA-IR ($p=7.58\times10^{-19}$), BMI ($p=1.87\times10^{-1398}$), body fat ($p=6.94\times10^{-83}$), and CRP ($p=6.47\times10^{-260}$), supporting a mechanism of obesity-mediated insulin resistance.

The Lipodystrophy cluster had top-weighted traits and loci suggestive of 'lipodystrophy-like' or fat distribution-mediated insulin resistance as in our and other's prior work [5,

37, 38]; these included decreased adiponectin, HDL-cholesterol, and modified Stumvoll insulin sensitivity index (ISI; adjusted for age, sex and BMI), increased triglycerides and waist–hip ratio, and *IRS1*, *KLF14* and *PPARG* (Fig. 2, ESM Table 5, 6). The GWAS pPS for the Lipodystrophy cluster (54 loci) was associated with increased fasting insulin (p=3.16×10^{-43}), HOMA-IR (p=7.47×10^{-29}) and triglycerides (p=1.18×10^{-612}), decreased ISI (p=1.84×10^{-38}) and HDL (p=5.19×10^{-535}).

The Liver/Lipid cluster was defined by decreased triglycerides and γ-glutamyl transferase levels, and multiple loci previously connected to hepatic lipid or glycogen metabolism, including *GCKR, HNF1A, PPP1R3B, TOMM40/APOE,* and *PNPLA3* (Fig. 2, ESM Table 5, 6) [39]. The GWAS pPS for this cluster (11 loci) was associated with reduced triglycerides (p=3.64×10^{-181}) and interestingly also reduced CRP (p=7.75×10^{-106}) and white blood cell count (p=1.42×10^{-49}).

The two remaining insulin resistance clusters (ALP negative and Hyper Insulin Secretion) were novel, containing driving traits and loci not part of our prior clusters (Fig. 2, ESM Table 5, 6). The ALP negative cluster had decreased ALP level as its top-weighed trait, and the *ABO* locus as the top-weighted locus. The GWAS pPS in this cluster (4 loci) was associated with decreased ALP (p=1.97×10^{-1431}) and triglycerides (p=4.49×10^{-247}). The Hyper Insulin Secretion cluster included top-weighted traits of increased DI and CIR, and loci *PPP1R3B, CNTN2*, *DTNB*, *SREBF1* and *TNF*. More than 87% of the loci in this cluster were not part of our prior work. The Hyper Insulin Secretion GWAS pPS (32 loci) was associated with increased CIR (p=1.16×10^{-14}), DI (p=2.89×10^{-14}), BMI (p=1.01×10^{-26}), and reduced HDL (p=1.09×10^{-110}) and SHBG (p=1.07×10^{-100}).

The final cluster, SHBG, was novel to the current work and not significantly associated with fasting insulin (GWAS pPS p=0.36). The cluster was driven by a single trait and locus: decreased SHBG levels and the *SHBG* locus (ESM Table 5, 6). The GWAS pPS in this cluster (1 locus) was significantly associated with reduced SHBG (p=1.2×10^{-1784}) and IGF-1 (p=4.12×10^{-13}).

Type 2 diabetes clusters differ in tissue enrichment including single-cell islets

To acquire further evidence for the suspected disease mechanisms represented by clusters and assess biological differences between the clusters, we analysed the top-weighted loci in each type 2 diabetes cluster for enrichment of epigenomic annotations across 28 tissues (Fig. 3a, ESM Table 8a). In line with expected mechanisms, the Beta cell 1, Beta cell 2, and Proinsulin clusters were significantly enriched in pancreatic islet tissue (false discovery rate [FDR]<0.05). The Liver/Lipid and ALP negative clusters were significantly enriched in liver tissue (FDR<0.01). The Lipodystrophy cluster was strongly enriched in adipose tissue (FDR<0.01). Additionally, both Beta cell 1 and 2 had enrichment in adipose tissue and the brain hippocampus (FDR<0.01). The Obesity cluster was most transcriptionally enriched in human epidermal keratinocytes (NHEK) and hASC-t3 pre-adipose cells, both at nominal significance (p<0.05, FDR=0.11).

We also interrogated newly available chromatin profiles from 14.3k pancreatic islet cells, which Chiou et al subsetted based on their chromatin profiles [30]. In prior work, the

islets were found to have two epigenomic subsets, labelled Beta INS^{high} and Beta INS^{low}, indicating high or low insulin gene (*INS)* promoter accessibility; the Beta INS^{high} islet cells had enriched promoter accessibility for genes involved in insulin secretion, whereas the Beta INS^{low} cells had enrichment for stress-induced signalling response genes. When assessing enrichment of our clusters, we found that our Beta cell 1 cluster was enriched only in Beta INS^{high} cells (p=0.0001, FDR=0.0014), whereas our Beta cell 2 cluster was nominally enriched in both Beta INS^{high} and Beta INS^{low} cells (p=0.025, p=0.013, respectively, FDR=0.18 for both), (Fig. 3b, ESM Table 8b). The same trend was observed in fgwas enrichment analysis: Beta cell 1 was significantly enriched only in INS^{high} [$\log_e$(enrichment) (95% CI): INS^{high} 2.32 (1.31, 3.12); INS^{low} −0.36 (−1.79, 0.55)] whereas Beta cell 2 was significantly enriched in both single-cell subsets [$\log_e$(enrichment) (95% CI): INS^{high} 1.61 (0.22, 2.96); INS^{low} 2.11 (0.73, 3.46)] (Fig. 3c). Together these results support the likelihood that Beta cell 1 and Beta cell 2 clusters relate to distinct physiological mechanisms, with Beta cell 2 potentially connected to a stress-induced pancreatic state.

Also of interest within the pancreas single-cell data, the Liver/Lipid cluster was most enriched for alpha cells, (p=0.007, FDR=0.099); alpha cells secrete glucagon, which acts to release glucose from the glycogen stores in the liver, providing further connection between these type 2 diabetes loci and liver function (Fig. 3b).

Type 2 diabetes clusters are differentially associated with clinical traits and outcomes

To assess translation of the clusters to individuals, we generated cluster pPSs in the MGB Biobank (N=25,419). We first confirmed that cluster pPSs were associated with expected traits in this study population both in all individuals and in those with type 2 diabetes (ESM Table 9).

We next tested whether the cluster pPSs were associated cardiometabolic outcomes related to type 2 diabetes using GWAS summary statistics: CAD, CKD, eGFR, hypertension, ischaemic stroke and diabetic neuropathy (ESM Table 4, Fig. 4a, ESM Fig. 4a). All ten type 2 diabetes clusters were associated with at least one outcome. The GWAS pPS results for eGFR highlighted the utility of cluster-specific scores, with individual clusters having more significant associations than the full set of type 2 diabetes SNPs: increased pPSs for the Liver/Lipid, ALP negative and SHBG clusters were associated with reduced eGFR ($p<5\times10^{-4}$), whereas all cluster type 2 diabetes SNPs together did not reach Bonferroni-corrected significance (p=0.03, ESM Table 10). The most significant of these GWAS pPSs were replicated using individual-data from MGB Biobank: increased Obesity cluster pPS with increased risk of hypertension, increased Lipodystrophy cluster pPS with increased risk of CAD, and increased Liver/Lipid cluster pPS with reduced eGFR, in all individuals with and without adjustment for type 2 diabetes status (Fig. 4b, ESM Table 11).

Clusters from CAD and CKD share mechanistic pathways with type 2 diabetes

We applied our clustering pipeline to two other metabolic diseases, CAD and CKD, identifying five CAD clusters (219 loci): ALP negative, Lipoprotein A, HDL negative, Cholesterol and Blood markers increased), and five CKD clusters (70 loci): Blood markers increased, Urea increased, Reduced haematopoiesis, Beta cell opposite and Lipoprotein A

(ESM Tables 12–15, ESM Fig. 5). Based on inspection of constituent variants and traits in the clusters of type 2 diabetes, CAD and CKD, the Lipoprotein A cluster was shared by all three diseases. Similarly, the ALP negative cluster was shared between type 2 diabetes and CAD, and the Blood markers increased cluster between CAD and CKD.

Discussion

Novel approaches are needed to connect the currently identified hundreds of type 2 diabetes loci to disease pathophysiology and accommodate the rapid pace of new locus discovery. Here, we describe expanded clustering of type 2 diabetes variants, using a high-throughput pipeline for extracting and pre-processing variants from multiple GWAS datasets and generating a variant–trait association matrix. Applying bNMF soft clustering to this 323 by 64 type 2 diabetes variant–trait matrix, we identify ten type 2 diabetes genetic clusters, which we show have tissue epigenomic specificity and are associated with distinct metabolic outcomes.

Importantly, among the ten clusters, we again capture the five identified in our previous work of 94 type 2 diabetes variants (Beta cell, Proinsulin, Obesity, Lipodystrophy, Liver/Lipid) [5], with the Beta cell cluster subdivided into two distinct clusters, and also identified four novel clusters related to pronounced insulin secretion, levels of ALP, Lp(a) and SHBG. In contrast to our prior work, which involved manual curation of published GWAS loci to generate the input list of variants, the current approach allowed for use of uncurated GWAS datasets and included newly available datasets, more than tripling the number of input loci. Thus, rediscovery of the previously identified clusters provides strong validation of this high-throughput approach, with the newly identified clusters driven by traits or loci not available in the prior analysis.

Three of the ten type 2 diabetes clusters identified in this work (Beta cell 1, Beta cell 2, and Proinsulin) clearly relate to pancreatic beta cell function, differing in part due to the direction or magnitude of the PI association. All three clusters were enriched in pancreatic islet tissue in the epigenomics analysis. Additionally, loci in the Beta cell 2 cluster had a unique signal of enrichment for single beta cells predicted to be in a stressed state [30]. The functional distinctions between these clusters support our independent approach of phenotypically informed type 2 diabetes locus clustering.

Three other type 2 diabetes clusters related to pathways of insulin resistance (Obesity, Lipodystrophy, Liver/Lipid) were also captured in our prior work, but now gained additional loci and traits. Loci in these clusters were most enriched for enhancers in tissues for the suspected mechanisms: pre-adipocytes, adipocytes and liver tissue, respectively. Interestingly we also observed a significant association for the Liver/Lipid cluster with pancreatic alpha cells, which may relate to the liver activity of this cluster or could suggest a more direct role for glucagon in diabetes development [40]. The distinction between fat accumulation in the Obesity cluster and abnormal fat compartmentalisation in the Lipodystrophy cluster may be supported by the differential enhancer enrichment shown for different developmental stages of the adipocyte lineage.

We also identified four new type 2 diabetes clusters from this work, several of which were also captured in the clustering of CAD and CKD: ALP negative, Lipoprotein A, SHBG and Hyper Insulin Secretion.

The ALP negative cluster (seen for type 2 diabetes and CAD) was driven by reduced serum ALP levels and the *ABO* locus. Isoform levels of ALP have been shown to vary by blood group [41]. The *ABO* locus and blood type have previously been connected to type 2 diabetes [42] and CAD [43] risk, but the causal mechanisms are not fully understood.

The Lipoprotein A cluster (seen for type 2 diabetes, CAD and CKD) all included the top locus (*SLC22A3*/*LPA* tagged by rs487152) and top biomarker Lp(a), pointing to a genetic pathway leading to increased Lp(a) levels and increased risk of type 2 diabetes, CAD and CKD. The relationship between Lp(a) and cardiometabolic disease is complex, and genetic interrogation of *LPA* has been complicated by the fact that plasma concentration of Lp(a) is influenced by kringle IV type 2 repeats in addition to other genetic variation [44]. While epidemiological studies have connected elevated Lp(a) levels with increased risk of CAD and CKD [45, 46], an inverse association has been reported for type 2 diabetes [47]. Our genetic findings for type 2 diabetes therefore indicate that there are likely to be multiple pathways impacting Lp(a) level that may have differential effects on type 2 diabetes risk.

For the SHBG cluster (seen for type 2 diabetes), our results point to a genetic pathway whereby alteration of the *SHBG* locus leads to reduced SHBG levels and increased type 2 diabetes risk, which was consistent with previous epidemiological and genetic studies indicating that low circulating levels of SHBG were causally related to increased risk of type 2 diabetes in both sexes [48].

We assessed the impact of cluster pPSs in individuals, finding that individuals with increased cluster pPS had significant associations with clinical traits and disease outcomes, supporting prior findings for the original five clusters [8]. While the effect sizes of the pPSs on clinical outcomes were too small to be of clinical utility at the individual-level, the results point to marked heterogeneity in type 2 diabetes genetic associations with clinical features, suggesting important physiological implications. For example, we detect associations with CKD and eGFR for several of the insulin-resistance-related clusters, but not with the insulin-deficiency-related clusters. Our results are consistent with the phenotypic-based clustering finding in Ahlqvist et al of the ‘severe insulin-resistant diabetes’ group having the highest risk of developing CKD [49] and may have implications for preventing or treating diabetic kidney disease.

The strengths of this study include the high-throughput approach for pre-processing variants and traits from multiple GWAS datasets in a semi-automated way. This method can be readily applied to other diseases beyond type 2 diabetes to identify mechanisms of disease, and the code has been made publicly available. We included here application of the pipeline to CAD and CKD, demonstrating transferability of the approach and potential shared mechanisms of disease. Limitations include clustering of only available phenotypes from GWAS. It is possible that additional pathways exist that are not captured using the set of traits included in the analysis. Additionally, due to methodological limitations and data

availability we have focused on GWAS from populations of European ancestry, although we are actively pursuing application of this method in non-European populations through additional efforts. It is worth noting that bNMF generates weights for all included elements in the matrix, and it is not known how best to determine a cut-off threshold for cluster membership; we have applied a reasonable strategy to maximise signal-to-noise ratio. Future work would benefit from longitudinal analysis to assess the impact of clusters throughout the disease course as well as further validation that the genetic clusters map to predicted disease processes, as has been done for one of the original clusters using cellular characterisation [7].

In summary, we have identified ten robust genetic clusters pointing to mechanistic pathways of type 2 diabetes using a high-throughput clustering pipeline. These clusters displayed tissue-specific enrichment patterns even within single-cell pancreatic tissue subsets and could be used to generate pPSs that stratify patients genetically with distinct associations with clinical outcomes. We demonstrate that our approach can be applied to other complex diseases, with identification of overlapping clusters between type 2 diabetes, CAD and CKD. Thus, we contribute to further delineation of cardiometabolic disease genetic pathways using a data-driven approach informed by physiology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

This work was supported by FNIH RFP-13 and the MGH Transformative Scholars Award.

Data availability

Code for variant pre-processing, bNMF clustering, and basic visualisations is available at https://github.com/gwas-partitioning/bnmf-clustering.

Interactive results are viewable on the Common Metabolic Disease Knowledge Portal (https://hugeamp.org/).

Abbreviations

aBF	Approximate Bayes factor
AMP-CMDKP	Accelerating Medicines Partnership-Common Metabolic Disease Knowledge Portal
bNMF	Bayesian non-negative matrix factorisation
cCRE	*cis*-regulatory elements
ALP	Alkaline phosphatase
CAD	Coronary artery disease

CIR	Corrected insulin response
CKD	Chronic kidney disease
CRP	C-reactive protein
DI	Disposition index
FDR	False discovery rate
GWAS	Genome-wide association study
INS$^{high/low}$	High/low insulin gene (INS) promoter accessibility
ISI	Insulin sensitivity index
Lp(a)	Lipoprotein A
MGB	Mass General Brigham
PC	Principal component
pPS	Partitioned polygenic score
SHBG	Sex hormone-binding globulin

References

1. Redondo MJ, Hagopian WA, Oram R, et al. (2020) The clinical consequences of heterogeneity within and between different diabetes types. Diabetologia 63(10):2040–2048 [PubMed: 32894314]
2. Mahajan A, Taliun D, Thurner M, et al. (2018) Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. Nat Genet 50(11):1505–1513 [PubMed: 30297969]
3. Goodarzi MO, Palmer ND, Cui J, et al. (2020) Classification of type 2 diabetes genetic variants and a novel genetic risk score association with insulin clearance. J Clin Endocrinol Metab 105(4):1251–1260 [PubMed: 31714576]
4. Mahajan A, Wessel J, Willems SM, et al. (2018) Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. Nat Genet 50(4):559–571 [PubMed: 29632382]
5. Udler MS, Kim J, von Grotthuss M, et al. (2018) Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: a soft clustering analysis. PLoS Med 15(9):e1002654 [PubMed: 30240442]
6. Dimas AS, Lagou V, Barker A, et al. (2014) Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. Diabetes 63(6):2158–2171 [PubMed: 24296717]
7. Laber S, Strobel S, Mercader J-M, et al. (2021) Discovering cellular programs of intrinsic and extrinsic drivers of metabolic traits using LipocyteProfiler. bioRxiv
8. DiCorpo D, LeClair J, Cole JB, et al. (2022) Type 2 diabetes partitioned polygenic scores associate with disease outcomes in 454,193 individuals across 13 Cohorts. Diabetes Care 45(3):674–683 [PubMed: 35085396]
9. Wagner R, Heni M, Tabák AG, et al. (2021) Pathophysiology-based subphenotyping of individuals at elevated risk for type 2 diabetes. Nat Med 27(1):49–57 [PubMed: 33398163]
10. Nair ATN, Wesolowska-Andersen A, Brorsson C, et al. (2022) Heterogeneity in phenotype, disease progression and drug response in type 2 diabetes. Nat Med 28(5):982–988 [PubMed: 35534565]

11. Scott RA, Scott LJ, Mägi R, et al. (2017) An expanded genome-wide association study of type 2 diabetes in Europeans. Diabetes 66(11):2888–2902 [PubMed: 28566273]
12. Kurki MI, Karjalainen J, Palta P, et al. (2022) FinnGen: Unique genetic insights from combining isolated population and national health register data. bioRxiv
13. Bonàs-Guarch S, Guindo-Martínez M, Miguel-Escalada I, et al. (2018) Re-analysis of public genetic data reveals a rare X-chromosomal variant associated with type 2 diabetes. Nat Commun 9(1):321 [PubMed: 29358691]
14. Flannick J, Mercader JM, Fuchsberger C, et al. (2019) Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. Nature 570(7759):71–76 [PubMed: 31118516]
15. Guindo-Martínez M, Amela R, Bonàs-Guarch S, et al. (2021) The impact of non-additive genetic associations on age-related complex diseases. Nat Commun 12(1):2436 [PubMed: 33893285]
16. Accelerating Medicines Partnership. Common Metabolic Diseases Knowledge Portal. Available from https://hugeamp.org/. Accessed 19 Mar 2021
17. UK biobank. —. In: Neale lab. http://www.nealelab.is/uk-biobank. Accessed 18 Jul 2022
18. van der Harst P, Verweij N (2018) Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. Circ Res 122(3):433–443 [PubMed: 29212778]
19. Nikpay M, Goel A, Won H-H, et al. (2015) A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet 47(10):1121–1130 [PubMed: 26343387]
20. Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators, Stitziel NO, Stirrups KE, et al. (2016) Coding variation in ANGPTL4, LPL, and SVEP1 and the risk of coronary disease. N Engl J Med 374(12):1134–1144 [PubMed: 26934567]
21. Wuttke M, Li Y, Li M, et al. (2019) A catalog of genetic loci associated with kidney function from analyses of a million individuals. Nat Genet 51(6):957–972 [PubMed: 31152163]
22. Salem RM, Todd JN, Sandholm N, et al. (2019) Genome-wide association study of diabetic kidney disease highlights biology involved in glomerular basement membrane collagen. J Am Soc Nephrol 30(10):2000–2016 [PubMed: 31537649]
23. van Zuydam NR, Ahlqvist E, Sandholm N, et al. (2018) A genome-wide association study of diabetic kidney disease in subjects with type 2 diabetes. Diabetes 67(7):1414–1427 [PubMed: 29703844]
24. Locke AE, Steinberg KM, Chiang CWK, et al. (2019) Exome sequencing of Finnish isolates enhances rare-variant association power. Nature 572(7769):323–328 [PubMed: 31367044]
25. Companion R package for the guide Doing Meta-Analysis in R. http://dmetar.protectlab.org/. Accessed 18 Jul 2022
26. Wakefield J (2007) A Bayesian measure of the probability of false discovery in genetic epidemiology studies. Am J Hum Genet 81(2):208–227 [PubMed: 17668372]
27. Ernst J, Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. Nat Methods 9(3):215–216 [PubMed: 22373907]
28. Varshney A, Scott LJ, Welch RP, et al. (2017) Genetic regulatory signatures underlying islet gene expression and type 2 diabetes. Proc Natl Acad Sci U S A 114(9):2301–2306 [PubMed: 28193859]
29. Chiou J, Geusz RJ, Okino M-L, et al. (2021) Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. Nature 1–5
30. Chiou J, Zeng C, Cheng Z, et al. (2021) Single-cell chromatin accessibility identifies pancreatic islet cell type- and state-specific regulatory programs of diabetes risk. Nat Genet 53(4):455–466 [PubMed: 33795864]
31. Pickrell JK (2014) Joint analysis of functional genomic data and genome-wide association studies of 18 human traits. Am J Hum Genet 94(4):559 [PubMed: 24702953]
32. Smoller JW, Karlson EW, Green RC, et al. (2016) An eMERGE clinical center at Partners Personalized Medicine. Journal of Personalized Medicine 6(1):5 [PubMed: 26805891]

33. Rosengren AH, Braun M, Mahdi T, et al. (2012) Reduced insulin exocytosis in human pancreatic β-cells with gene variants linked to type 2 diabetes. Diabetes 61(7):1726–1733 [PubMed: 22492527]
34. Mezza T, Ferraro PM, Sun VA, et al. (2018) Increased β-cell workload modulates proinsulin-to-insulin ratio in humans. Diabetes 67(11):2389–2396 [PubMed: 30131390]
35. Carrat GR, Hu M, Nguyen-Tu M-S, et al. (2017) Decreased STARD10 expression is associated with defective insulin secretion in humans and mice. Am J Hum Genet 100(2):238–256 [PubMed: 28132686]
36. Choquet H, Meyre D (2011) Genetics of Obesity: what have we learned? Curr Genomics 12(3):169–179 [PubMed: 22043165]
37. Yaghootkar H, Scott RA, White CC, et al. (2014) Genetic evidence for a normal-weight "metabolically obese" phenotype linking insulin resistance, hypertension, coronary artery disease, and type 2 diabetes. Diabetes 63(12):4369–4377 [PubMed: 25048195]
38. Lotta LA, EPIC-InterAct Consortium, Gulati P, et al. (2017) Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. Nat Genet 49(1):17–26 [PubMed: 27841877]
39. Tavaglione F, Targher G, Valenti L, Romeo S (2020) Human and molecular genetics shed lights on fatty liver disease and diabetes conundrum. Endocrinol Diabetes Metab 3(4):e00179 [PubMed: 33102799]
40. Finan B, Capozzi ME, Campbell JE (2020) Repositioning glucagon action in the physiology and pharmacology of diabetes. Diabetes 69(4):532–541 [PubMed: 31178432]
41. Domar U, Hirano K, Stigbrand T (1991) Serum levels of human alkaline phosphatase isozymes in relation to blood groups. Clin Chim Acta 203(2–3):305–313 [PubMed: 1777990]
42. Li-Gao R, Carlotti F, de Mutsert R, et al. (2019) Genome-wide association study on the early-phase insulin response to a liquid mixed meal: results from the NEO Study. Diabetes 68(12):2327–2336 [PubMed: 31537524]
43. Chen Z, Yang S-H, Xu H, Li J-J (2016) ABO blood group system and the coronary artery disease: an updated systematic review and meta-analysis. Sci Rep 6(1):23250 [PubMed: 26988722]
44. Tolbus A, Mortensen MB, Nielsen SF, Kamstrup PR, Bojesen SE, Nordestgaard BG (2017) Kringle IV type 2, not low lipoprotein(a), as a cause of diabetes: a novel genetic approach using SNPs associated selectively with lipoprotein(a) concentrations or with Kringle IV type 2 repeats. Clin Chem 63(12):1866–1876 [PubMed: 28971985]
45. Tipping RW, Ford CE, Simpson LM, et al. (2009) Lipoprotein(a) concentration and the risk of coronary heart disease, stroke, and nonvascular mortality. JAMA 302:412–423 [PubMed: 19622820]
46. Xuan L, Wang T, Dai H, et al. (2020) Serum lipoprotein (a) associates with a higher risk of reduced renal function: a prospective investigation. J Lipid Res 61(10):1320–1327 [PubMed: 32703886]
47. Mora S, Kamstrup PR, Rifai N, Nordestgaard BG, Buring JE, Ridker PM (2010) Lipoprotein(a) and risk of type 2 diabetes. Clin Chem 56(8):1252–1260 [PubMed: 20511445]
48. Perry JRB, Weedon MN, Langenberg C, et al. (2010) Genetic evidence that raised sex hormone binding globulin (SHBG) levels reduce the risk of type 2 diabetes. Hum Mol Genet 19(3):535–544 [PubMed: 19933169]
49. Ahlqvist E, Storm P, Käräjämäki A, et al. (2018) Novel subgroups of adult-onset diabetes and their association with outcomes: a data-driven cluster analysis of six variables. Lancet Diabetes Endocrinol 6(5):361–369 [PubMed: 29503172]

Research in context

What is already known about this subject?

- Type 2 diabetes is highly polygenic and influenced by multiple biological pathways
- Five genetic clusters have been identified in previous work studying 94 type 2 diabetes variants

What is the key question?

- Can we identify additional genetic clusters by expanding the number of phenotypes and variants in the clustering and, if so, will the clusters point to mechanisms of type 2 diabetes pathogenesis with epigenomic tissue specificity and distinct clinical characteristics?

What are the new findings?

- Ten clusters are identified, including those from our prior analysis as well as novel clusters related to pronounced insulin secretion, and levels of alkaline phosphatase, lipoprotein A and sex hormone-binding globulin
- The clusters displayed tissue-specific epigenomic enrichment. Two beta cell clusters (splitting the ‘Beta cell’ cluster from our prior work) were differentially enriched in functional and stressed pancreatic beta cell states
- Cluster-specific polygenic scores were associated with clinical outcomes across genome-wide association studies and in participants in an independent hospital-based biobank. Multiple type 2 diabetes clusters overlapped with coronary artery disease and chronic kidney disease

How might this impact on clinical practice in the foreseeable future?

- Delineation of genetic pathways of type 2 diabetes may improve understanding of disease pathogenesis and identify genetic type 2 diabetes subtypes, both potentially leading to improved management of the condition among people with diabetes

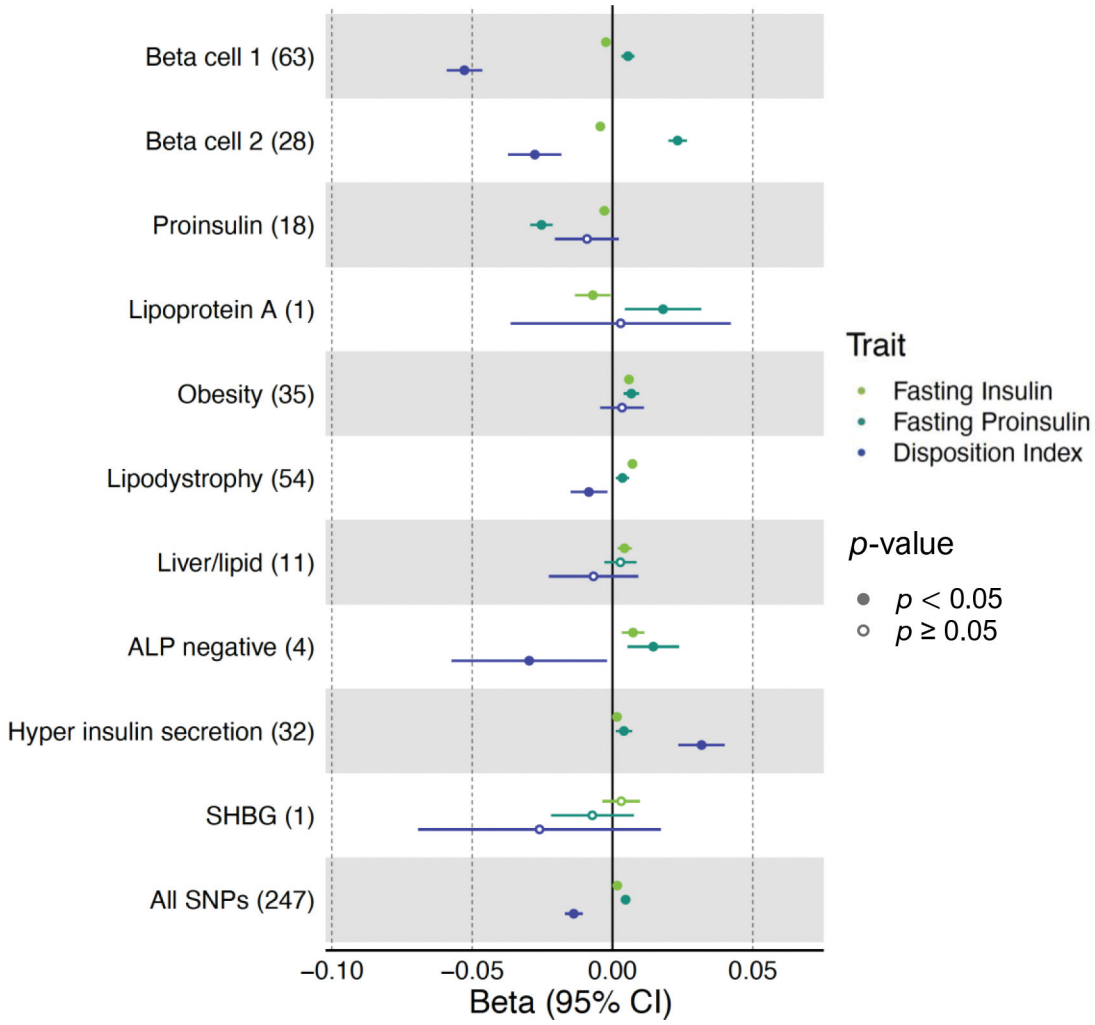


Fig. 1.

Cluster associations with metabolic traits using GWAS. Forest plot showing standardised effect sizes with 95% CI of cluster pPS–trait associations derived from GWAS summary statistics. Three metabolic traits (fasting insulin, fasting proinsulin adjusted for fasting insulin, and DI) that help discriminate clusters are displayed. The numbers in parentheses next to cluster names indicate the number of variants included in the analysis in each cluster. 'All SNPs' includes all the variants that are top-weighted in at least one cluster. Filled points indicate *p* values <0.05

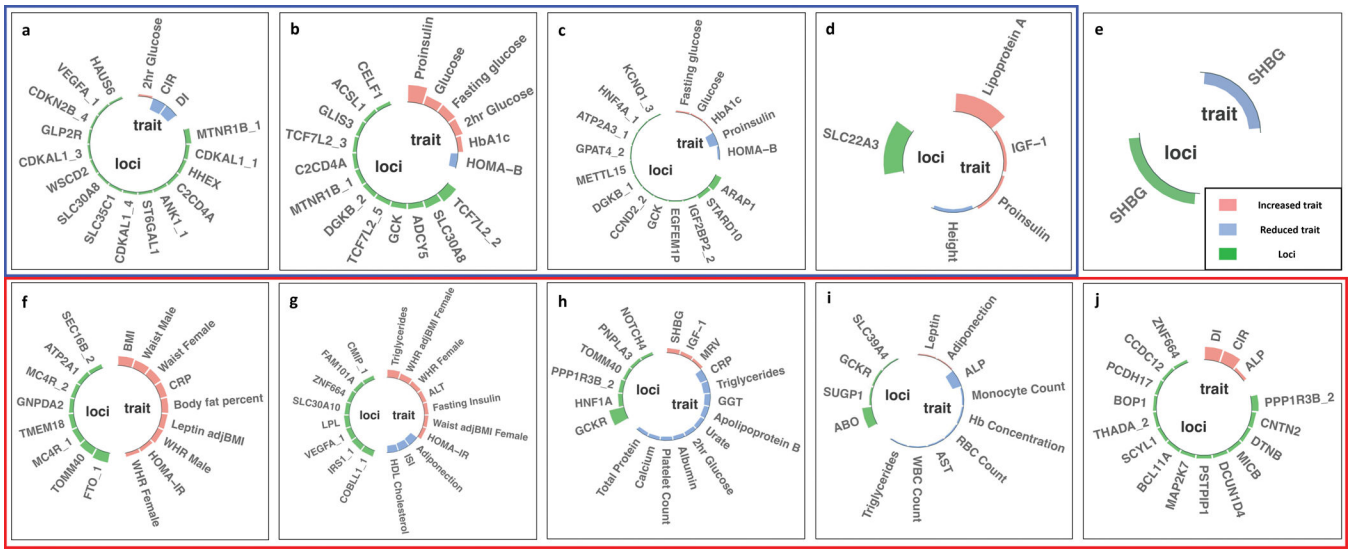


Fig. 2.

Clusters of type 2 diabetes loci. Top-weighted loci and traits in each of the ten clusters are represented in circular plots: (**a**) Beta cell 1, (**b**) Beta cell 2, (**c**) Proinsulin, (**d**) Lipoprotein A, (**e**) SHBG, (**f**) Obesity, (**g**) Lipodystrophy, (**h**) Liver/lipid, (**i**) ALP negative, (**j**) Hyper insulin secretion. The length of the bars shows the weights. Green bars represent top-weighted loci, red bars represent increased trait association, and blue bars represent decreased trait association with each cluster. A maximum of 35 elements (loci and traits) based on highest weights are displayed in each cluster. The blue outline indicates clusters associated with decreased fasting insulin levels, and the red outline indicates clusters associated with increased fasting insulin levels

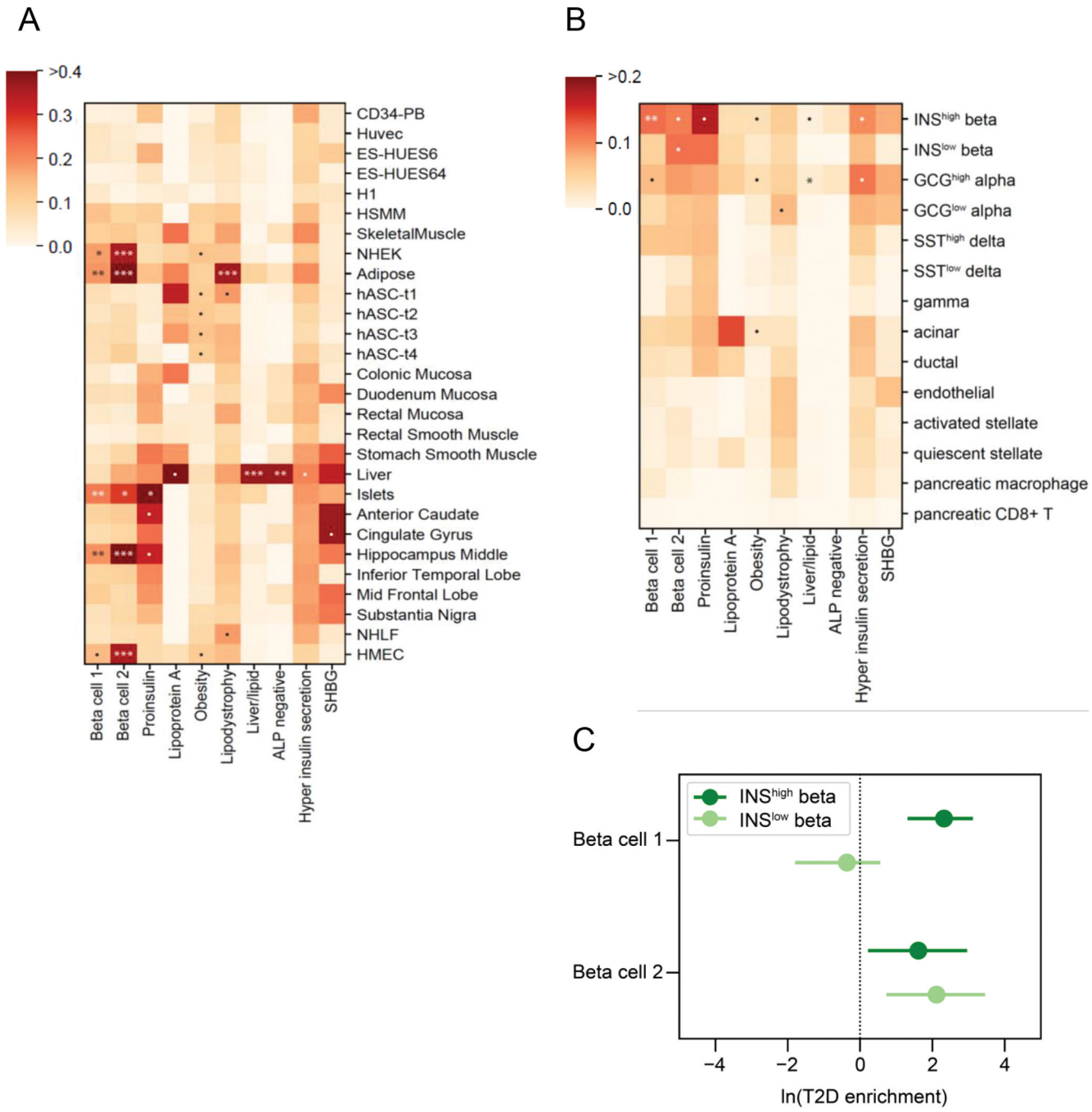


Fig. 3.

Enrichment for tissue-specific enhancers in type 2 diabetes clusters. (**a**) Heatmap of tissue enhancer/promoter enrichment analysis result. (**b**) Heatmap of pancreatic islet cell enrichment analysis result. Significance was indicated as follows: *** FDR < 0.001, ** FDR < 0.01, * FDR < 0.1, • $p < 0.05$. (**c**) Forest plot of comparison of Beta cell 1 and Beta cell 2 clusters in fgwas enrichment analysis in functional and stressed beta cell states

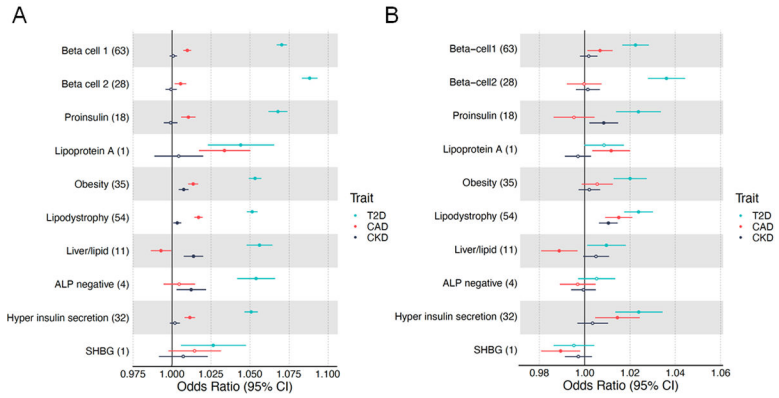


Fig. 4.

Forest plot of cluster associations with outcomes using (**a**) GWAS and (**b**) individual-level data from MGB Biobank. (**a**) Forest plot showing standardised effect sizes with 95% CI of cluster pPS–outcome associations derived from GWAS summary statistics. Three metabolic outcomes (type 2 diabetes, CAD and CKD, all unadjusted for type 2 diabetes) are displayed. The numbers in parentheses next to cluster names indicate the number of variants included in the analysis in each cluster. Filled points indicate p values <0.05. (**b**) Forest plot of associations of pPSs in individuals in the MGB Biobank with clinical outcomes. Three outcomes including type 2 diabetes are displayed. T2D, type 2 diabetes

Table 1

Overview of type 2 diabetes genetic clusters

Cluster	Expected physiological impact	Key top-weighted traits	Key top-weighted loci	Suspected mechanism	Note
Beta cell 1 (63)	Insulin deficiency	CIR (−), DI (−)	*MTNR1B, CDKAL1, HHEX, C2CD4A, ANK1, ST6GAL1, SLC35C1, SLC30A8*	Beta cell function, glucose homeostasis	Beta Cell cluster from Udler et al 2018 [5] divided into 2 clusters
Beta cell 2 (28)	Insulin deficiency	Fasting proinsulin adj. fasting insulin (+), HOMA-B (−), fasting insulin (−)	*TCF7L2, SLC30A8, ADCY5, GCK, DGKB, MTNR1B, C2CD4A*	Beta cell function, insulin processing	Beta Cell cluster from Udler et al 2018 [5] divided into 2 clusters
Proinsulin (18)	Insulin deficiency	Fasting proinsulin adj. fasting insulin (−), HOMA-B (−)	*ARAP1, STARD10*	Insulin synthesis	Recaptures Proinsulin cluster from Udler et al 2018 [5]
Lipoprotein A (1)	Insulin deficiency	Lp(a) (+)	*SLC22A3/LPA*	Lp(a) metabolism	New cluster in this study
Obesity (35)	Insulin resistance	BMI (+), waistC (+), % body fat (+), CRP (+)	*FTO, MC4R*	Obesity-mediated insulin resistance	Recaptures Obesity cluster from Udler et al 2018 [5]
Lipodystrophy (54)	Insulin resistance	Adiponectin (−), ISI (−), HDL (−)	*IRS, PPARG, KLF14*	Fat distribution-mediated insulin resistance	Recaptures Lipodystrophy cluster from Udler et al 2018 [5]
Liver/Lipid (11)	Insulin resistance	CRP (−), TG (−), GGT (−)	*GCKR, HNF1A, PPP1R3B, TOMM40, PNPLA3*	Liver/lipid metabolism	Recaptures Liver/Lipid cluster from Udler et al 2018 [5]
ALP negative (4)	Insulin resistance	ALP (−)	*ABO*	ALP activity levels	New cluster in this study
Hyper Insulin Secretion (32)	Insulin resistance	DI (+), CIR (+)	*PPP1R3B, CNTN2, DTNB, TNF, SREBF1*	Insulin secretion, inflammation	New cluster in this study
SHBG (1)	Unclear	SHBG (−)	*SHBG*	SHBG metabolism	New cluster in this study

The numbers in parentheses next to cluster names indicate the numbers of top-weighted variants in each of the clusters

Decrease and increase are indicated by (−) and (+), respectively adj., adjusted; GGT, γ-glutamyl transferase; waistC, waist circumference

Diabetologia. Author manuscript; available in PMC 2024 March 01.