A Roadmap to The Molecular Human Linking Multiomics with Population Traits and Diabetes Subtypes.

Anna Halama[1,2*], Shaza Zaghlool[1,2], Gaurav Thareja[1,2], Sara Kader[1,2], Wadha Al Muftha[3, 4], Marjonneke Mook-Kanamori[2], Hina Sarwath[5], Yasmin Ali Mohamoud[6], Nisha Stephan[1,2], Sabine Ameling[7,8], Maja Pucic Baković[9], Jan Krumsiek[2,10], Cornelia Prehn[11], Jerzy Adamski[12,13,14], Jochen M Schwenk[15], Nele Friedrich[7,16], Uwe Völker[7,8], Manfred Wuhrer[17], Gordan Lauc[9,18], S. Hani Najafi-Shoushtari[19,20], Joel A Malek[4,6], Johannes Graumann[21], Dennis Mook-Kanamori[22,23], Frank Schmidt[5,24], Karsten Suhre[1,2,10*].

[1] Bioinformatics Core, Weill Cornell Medicine-Qatar, Education City, Doha, Qatar
[2] Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, U.S.A.
[3] Qatar Genome Program, Qatar Foundation, Qatar Science and Technology Park, Innovation Center, Doha, Qatar.
[4] Department of Genetic Medicine, Weill Cornell Medicine, Doha, Qatar.
[5] Proteomics Core, Weill Cornell Medicine-Qatar, Education City, Doha, Qatar
[6] Genomics Core, Weill Cornell Medicine-Qatar, Education City, Doha, Qatar
[7] German Centre for Cardiovascular Research, Partner Site Greifswald, University Medicine Greifswald, Greifswald, Germany
[8] Department of Functional Genomics, Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany.
[9] Genos Glycoscience Research Laboratory, Zagreb, Croatia
[10] Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, New York, USA.
[11] Metabolomics and Proteomics Core, Helmholtz Zentrum München, Neuherberg, Germany
[12] Institute of Experimental Genetics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany.
[13] Department of Biochemistry, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore.
[14] Institute of Biochemistry, Faculty of Medicine, University of Ljubljana, Ljubljana, Slovenia.
[15] Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology and Health, KTH Royal Institute of Technology, 171 65 Solna, Sweden
[16] Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany
[17] Leiden University Medical Center, Center for Proteomics and Metabolomics, Postbus 9600, 2300 RC Leiden, The Netherlands.
[18] Faculty of Pharmacy and Biochemistry, University of Zagreb, Zagreb, Croatia.
[19] MicroRNA Core Laboratory, Division of Research, Weill Cornell Medicine-Qatar, Education City, Doha, Qatar
[20] Department of Cell and Developmental Biology, Weill Cornell Medicine, New York, NY, U.S.A.
[21] Institute of Translational Proteomics, Department of Medicine, Philipps-Universität Marburg, Marburg, Germany.
[22] Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, the Netherlands
[23] Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, the Netherlands
[24] Department of Biochemistry, Weill Cornell Medicine, New York, NY, U.S.A.

[*] Correspondence to K.S. (kas2049@qatar-med.cornell.edu) & A.H. (amh2025@qatar-med.cornell.edu)

Supplementary Information Content:

Supplementary Notes:

Supplementary Figures:

Supplementary Note 1

The largest number of MBHs was found between successive generations of the Metabolon platforms (PM ⇔ HDF), which is not surprising, given the technological similarity between them. Out of 369 identified hits, 291 paired identically annotated metabolites, 57 MBHs linked an unknown metabolite measured on the older platform generation to an annotated molecule measured on the more recent platform generation (e.g. X-18601 ⇔ androstenediol (3beta,17beta)-monosulfate), and 21 MBHs linked apparently differing molecules in related pathways (7 molecules; e.g. threonate ⇔ oxalate) or unknowns (14 molecules). As the Metabolon platforms differ with respect to the employed metabolite separation and detection methods (gas chromatography (GC) used for PM platform vs. liquid chromatography deployed for HDF (see methods)), this shows a robust concordance (79%) of platform performance and progressing component identification over time.

Detailed IgG glycosylation was determined by two independent glycomics platforms from two independent labs, referred to as IgG and IgA. Consequently, 29 out of 31 identified MBHs mapped to the same glycan structure, showing excellent agreement of 93% between both platforms.

Two affinity proteomics platforms were used, one based on the SOMAscan aptamer technology (SOMA, 1129 traits), and the other OLINK technology based on antibody pairs implementing the proximity extension assay (OLINK, 184 traits). Of 72 proteins that overlapped between both platforms, 52 were linked by MBH (72%). We found that overlapping proteins, which were not captured by a MBH, showed low correlation (**Supplementary Data 10**), further suggesting that those might be susceptible to different analytical parameters, hence should be validated with alternative technical platforms to ensure the correctness of the measurement.

Supplementary Note 2

We have investigated various GWAS associations to evaluate the performance of different platforms. For instance, SNP rs1047891 associated with glycine with a p-value = 7.4 $x10^{-18}$ in 320 samples on targeted lipidomics BM platform, with p-value = 7.4 $x10^{-15}$ in 291 samples on the HDF platform, and with p-value = 7.1 $x10^{-14}$ in 322 samples on the PM platform (**Supplementary Figure 6A**). In this example, the targeted assay appears to provide stronger signals, at least compared to the older PM platform, which was measured using an older generation of metabolic profiling described as HD2 by Metabolon. In another example however SNP rs1799958 associate with butyrylcarnitine (C4) with p-value = 7.9 $x10^{-9}$ in 323 samples on targeted lipidomics BM platform, and p-value = 2.3 $x10^{-15}$ in 294 on the HDF platform, and with p-value = 1.6 $x10^{-23}$ in 325 samples on the PM platform, suggesting that platform performance might depend on the molecule properties and how well a given molecule is captured by each platform

(**Supplementary Figure 6B**). We have observed similar trend across other platforms including measurements of urine metabolome with CM and UM (for SNP rs9922704 with 3-hydroxyisovalerate (**Supplementary Figure 6C**)) and with plasma proteome SOMA and OLINK (for rs3896287 with LILRB2 (**Supplementary Figure 6D**); for SNP rs8176693 with TIE1 (**Supplementary Figure 6E**); for SNP rs9892586 with CCL14 (**Supplementary Figure 6F**)).

Supplementary Note 3

The number of identified partial correlations varies according to the platform. For instance, out of 6,183 identified GGM's, the largest number of 2,689 lipid-lipid associate ions were identified for the LD platform. At the same time, only one glycan-glycan association was found for the IgG platform (**Supplementary Data 2**). The metabolite-metabolite associations in saliva, plasma, and urine were 56, 371, and 666, respectively. GGM's which we found can help explain biochemical processes by connecting chemical reactions as we find multiple substrate/product associations (e.g. cortisone/cortisol; fumarate/malate; glutamate/alpha-ketoglutarate), which shows how GGMs provide a simplified overview of the actual biological processes. GGM also help in the understanding of physiological processes such as the association between Luteinizing hormone (LHB) and follicle stimulating hormone (FSHB), which synergistically stimulate follicular growth and ovulation [1], as well as the metabolism and excretion of aspirin by indicating salicylate ⇔ salicilurate association in plasma and urine.

Supplementary Note 4

The composition of fatty acid (FA) side chains in complex lipids such as phosphatidylcholine or triacylglycerols play a role in a broad range of biological processes and it is thus critical to determine FA composition in measured lipids. This information is, however, not provided for phosphatidylcholines measured on the BM platform or triacylglycerols measured on the LD platform. We previously showed that the composition of phosphatidylcholines measured on the BM can be resolved by LD platform[2] . Here, we investigated whether MBH may support more complete determination of the structural composition of complex lipids. Indeed, after examining MBH between the two lipidomics platforms (LD ⇔ BM) and between the metabolomics and lipidomics platforms (HDF ⇔ BM) and (HDF ⇔ LD), we resolved the FA side chain composition for a number of complex lipids. The side chain composition of phosphatidylcholines measured on the BM platform, for instance, were delineated using MBH (e.g. PC_aa_C32:1 ⇔ PC(16:0/16:1), PC_aa_C40:6⇔ PC(18:0/22:6)) (**Supplementary Figure 7A**), in line with our previous study[2]. The characterization of fatty acid chains in triacylglycerol was similarly refined (TAG48:2-FA14:0 ⇔ myristoyl-linoleoyl-glycerol (14:0/18:2); TAG54:6-FA22:6 ⇔ palmitoyl-docosahexaenoyl-glycerl (16:0/22:6)) (**Supplementary Figure 7B**).

These examples indicate how combining two technologically similar platforms can add valuable biological information, making them complimentary rather than redundant.

Supplementary Note 5

The multiomics GWAS conducted with all omics phenotypes measured across all platforms resulted in identification of 768 omicQTLs at 586 independent genetic loci that reached a significance level of $p<5x10^{-8}$. We used PhenoScanner database of human genotype-phenotype associations [3] to check for the replications and potentially novel findings. Majority of our findings replicated previous reports (**Supplementary Data 7**) We were able to replicate multiple hits including GWAS associations in blood with metabolites at the *PYROXD2*, *NAT8*, *ACADS*, *NAT2*, *AGXT2*, *UGT1A4*, *CPS1*, *NAT16*, and *NAT16* loci (Gieger et al., 2008; Illig et al., 2010; Long et al., 2017; Shin et al., 2014; Suhre et al., 2011; Yu et al., 2014); with proteins including ACP1, ICAM1, FCGR2A, IL6R, ABO, SIGLEC9, ACP6, ENPP7, CCL15, PDGFRB (Di Narzo et al., 2017; Emilsson et al., 2018; Gilly et al., 2020; Suhre et al., 2017); with glycans including FUT6, ST6GAL1, MGAT5, C1GALT1 (Huffman et al., 2011; Kiryluk et al., 2017; Sharapov et al., 2019); and in the urine PYROXD2 and AGXT2 (Raffler et al., 2015; Schlosser et al., 2020); and in saliva one genetic loci *SLC2A9* reported in the first and so far, the only one GWAS metabolomics study in saliva (Nag et al., 2020).

We also replicated multiple EWAS hits including cg07839457 near *NLRC5*, cg08122652 near *PARP9* cg05575921 near *AHRR*, cg22910295 near *ICAM5*, cg13028630 near (*C4B*/*C4A*), and cg09488502 near *SIGLEC4* [4,5].

Almost all of our TWAS with gene-traits replicated previous findings.

Supplementary Note 6

The metabolomics GWAS hits which we have identified as previously unreported included 42 plasma metabolites (22 HDF and 20 PM platforms) and 36 urine metabolomics (21 with CM and 15 with PM). Some of the urine GWAS identified on the CM platform were prolific; these include e.g. galactose with 7 hits. The identified here urine GWAS hits including QTL near *ALMS1*, *ACADS*, *NAT2*, and *RNU6-675P* were previously reported in the blood [6-8] but not in the urine.

We found 96 previously unreported protein GWAS hits (including 76 pQTLs with SOMA and 20 pQTLs with OLINK). Out of the top 10 previously unreported significant associations 7 were proteins measured on OLINK panel, which could be explained by the relatively limited number of studies deploying this technology. The previously unreported pQTLs with the strongest association was between rs616114 (near *MEP1B* and *GAREM1* genes) and level of *MEP1B* (p-value = $2.1x10^{-45}$; beta = -0.88). The *MEP1B* is meprin β plasma membrane associated protein previously suggested to be involved in diabetic

nephropathy [9]. The *MEP1B* deficiency was shown to be associated with higher mortality rates and more severe diabetic kidney injury in mice with STZ-induced T1D [10], and was shown that *MEP1B* impacts complications of diabetes such as diabetic kidney injury by altering distinct metabolite profiles [11].

Out of 38 identified lipidomics-QTLs we have only replicated one previously reported namely association between rs1799958 and plasma butyrylcarnitine (p-value = $7.9x10^{-10}$; beta = 0.50) [12]. Among the remaining 37 associations some including 18 between rs4493662 (near *CTB-95D12.1/EEF1GP2*) and various different triacylglycerols (TAG55 – TAG58) and 6 between rs79659787 (near *RAB37*) and different triacylglycerols (TAG48 – TAG52) were prolific.

Supplementary Note 7

We have identified 108, 77, 22, and 7 significant oQTM's with IgA, OLINK, LD, and miRNA, respectively. The most prolific trait oQTL were among OLINK associations including TYMP (27 associations), ENPP7 (9 associations), FCGR3B (6 associations), and FCRL1 (6 associations). We found that two of CPG's associated with *FCGR3B* (cg26435281, and cg26561570) were on the gene locus of the associated protein (**Supplementary Figure 8A**) and out of nine CpG's associated with ENPP7 protein, seven were on the gene locus of associated protein and two in very near proximity on the same chromosome. The ENPP7 measured on OLINK showed MBH with ENPP7 measured on SOMA (p-value = $1.33x10^{-103}$, r = 0.83), which was associated with the same CpG's (**Supplementary Figure 8B**). Additionally, the ENPP7 protein associate with SNP rs3923265 near *ENPP7* (p-value = $2.8x10^{-24}$, beta = 0.67), which was previously reported (Sun et al., 2018). The *ENPP7* (Ectonucleotide pyrophosphatase/phosphodiesterase family member 7) is an alkaline sphingomyelinase that hydrolysis sphingomyelin (SM) to ceramide and was identified as implicated in T2D by insulin signaling inhibition [13]. Moreover, our recent study identified epigenetic modification of ENPP7 as factor contributing to T2D [14].

Supplementary Note 8

We have identified total of 1,381 (meQTL's) associations between gene SNPs and methylation levels, 15,991 (eQTM's) associations between methylation levels and mRNA and 17 (eQTL's) associations between gene SNPs and mRNA that reached a significance level of $p<5x10^{-8}$. The meQTL's consist of 203 unique SNPs and 308 unique methylation sites among which prolific SNPs (e.g. rs8283 near *ATF6B* associated with 33 methylation sites; rs4251552 near *IRAK4* associated with 21 methylation sites, or rs12924274 near *CNTNAP4* with 19 methylation sides) and prolific methylation sites (e.g. cg26831081 and cg05877118 near *CLYBL* with 75 and 73 SNPs, respectively; or cg03955321 near *NDUFAF1* with 63 SNPs)

were identified. Among eQTM's we identified 280 unique methylation sites and 239 unique mRNA's where we found many prolific methylation sites (e.g. cg04968013 near *CACNB2* associated with 216 mRNA's, cg08437802 near *KANK2* associated with 205 mRNA's) and mRNA's (e.g. *PRSS33* associates with 764 methylation sites, *SIGLEC8* associates with 673 methylation sites or *CLC* associated with 665 methylation sites). The 17 identified eQTL's included 12 unique SNP's and 7 unique mRNA's. We have identified two prolific mRNA's including *C4B* and *C4A* associated with 6 and 5 SNP's, respectively. We also identified association trios between the SNP-methylation, methylation-mRNA and SNP-mRNA (**Supplementary Figure 9**).

Supplementary Note 9

Among the protein and metabolite TWAS associations many reflect on the translation processes e.g. association between ENSG00000254415_*SIGLEC14* ⇔ SIGLEC14 (SOMA), and ENSG00000115523_*GNLY* ⇔GNLY (SOMA). Here the expressed gene serves as a guide for protein synthesis. Other associations, like the one found between the corticosteroid prednisolone and the FKBP5 gene transcript, were previously described as elevated under oral corticosteroid [15]. Thus, such interaction may indicate medication-triggered treatment responses.

Across the multiomics TWAS, we have identified prolific associations between (e.g. protein, metabolite) and multiple gene transcripts. For instance, Mullerian-inhibiting factor (AMH) was associated with 195 different transcripts, P-selectin (SELP) with 18 transcripts, C-X-C Motif Chemokine Ligand 10 (CXCL10), and C-X-C Motif Chemokine Ligand 11 (CXCL11) with 16 and 6 transcripts, respectively. The CXCL10 and CXCL11 are inflammatory chemokines induced by the interferon (IFN)-γ, regulating the differentiation of naive T cells to T helper 1 (Th1) cells and migration of immune cells to their focal site [16]. All the gene transcripts associated with CXCL11 overlap with the one identified as significantly associated with CXC10 and were previously reported as molecules involved in responses to viral infection (**Supplementary Data 15**).

Supplementary Note 10

Overview of the MOD molecular network

In the MOD network, 26 molecules monitored across 10 platforms were identified (**Supplementary Data 18** and **Supplementary Figure 3**). We observed that 9 of those molecules were associated with leptin, forming the main molecular subcluster. The molecules showing association with leptin (insulin, glycerol, and sphingolipids, as well as clinical parameters such as body temperature) reflect on the previously known and biologically relevant leptin function. For instance, the role of leptin in insulin secretion [17], lipolysis [18], sphingomyelin and ceramide metabolism [19] as well as body temperature regulation [20] was

previously described. The leptin association with glycerol indicates its contribution to triglyceride turnover and lipolysis, as previously suggested [21]. Moreover, leptin is sexually dimorphic (lower in males), explaining its association with SEX phenotype. We have also identified the link between leptin and CXCL5, which was not previously defined. CXCL5 is a cytokine implicated in the chemotaxis of inflammatory cells. It was also identified as highly expressed in white adipose tissue (WAT), and high circulating levels were associated with human obesity and insulin resistance [22]. Additionally, CXCL5 was recently implicated in the browning of WAT [23]. Interestingly, leptin was indicated as a molecule enabling browning of WAT while injected subcutaneously [24], intracerebroventricular and intra-ARC [25]. Identifying the association between leptin and CXCL5, which we achieved by this integrated omics analysis, might suggest their interplay in the processes relevant for WAT remodeling. This could have significant relevance for MOD cluster characterized by high BMI and low insulin resistance.

Overview of the SIRD molecular network

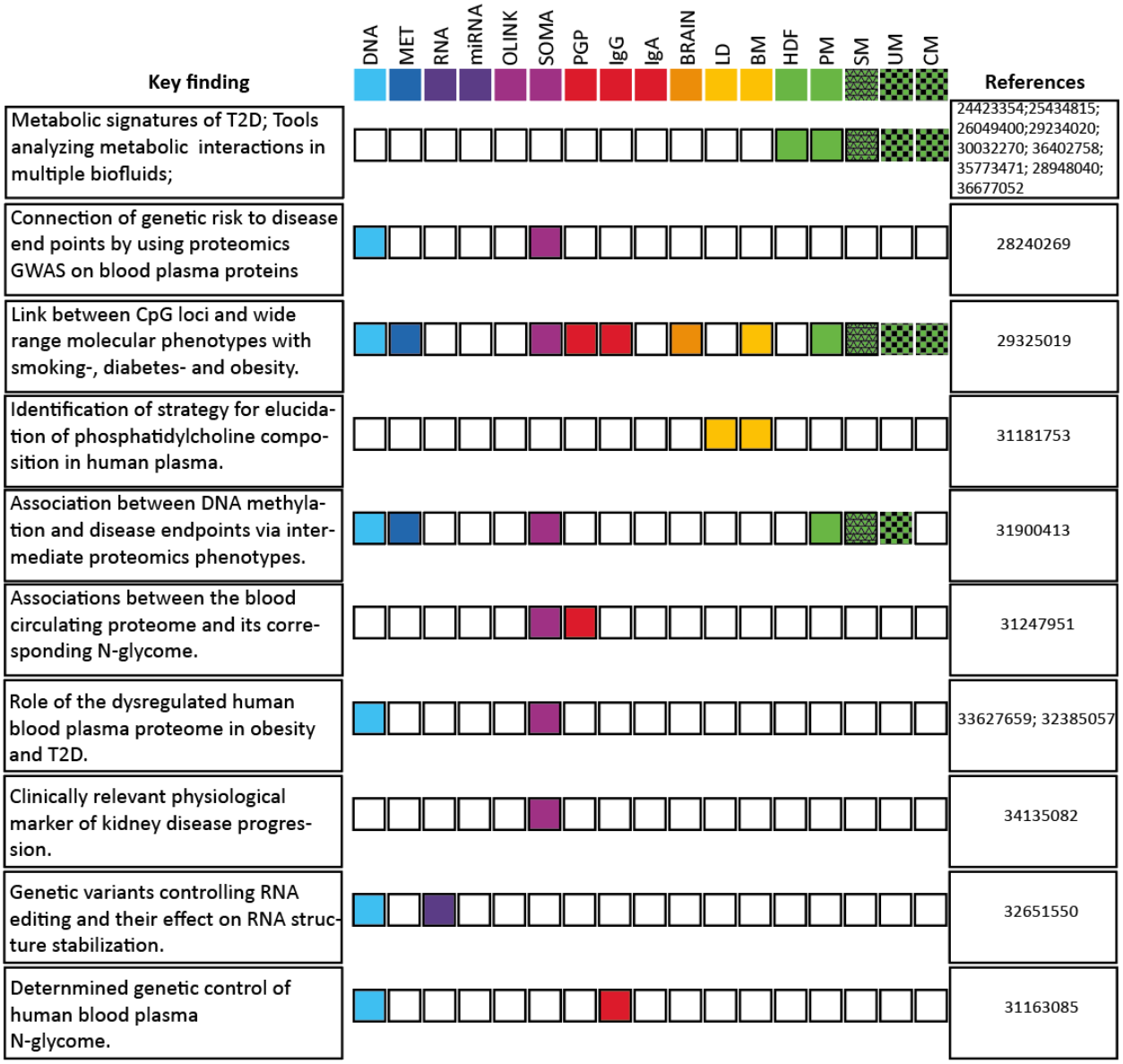
The molecular network of SIRD consists of 45 molecules measured on 9 platforms (**Supplementary Data 19** and **Supplementary Figure 4**). Those molecules form subclusters centered around insulin, insulin-like growth factor binding protein 1 (IGFBP1), phenylalanine, phosphate, 2-hydroxyarachidate, and linked pyrraline – DiHOME. Insulin was directly associated with leptin and pancreatic polypeptide (PPY), which is rapidly released in the postprandial state [26] and inhibits the glucagon secretion [27]. Interestingly, in this subcluster we also found 1-carboxyethylphenylanine and 1-carboxyethylvaline. The carboxyethylation of cysteine residues was recently attributed to gut microbiome and linked with generation of neoantigens in immune diseases [28]. Identified molecular network might potentially reflect on the secretory interplay between insulin, PPY and glucagon, which could be impacted by carboxyethylated branch chain and aromatic amino acids. Given that the SIRD subgroup is characterized by severe insulin resistance, the role of carboxyethylated branch chains and aromatic amino acids in PPY-glucagon interaction could provide further insight into the mechanisms related to insulin resistance. It would be necessary to further investigate these indications in a follow-up of this integrated omics analysis.

Overview of the SIDD molecular network

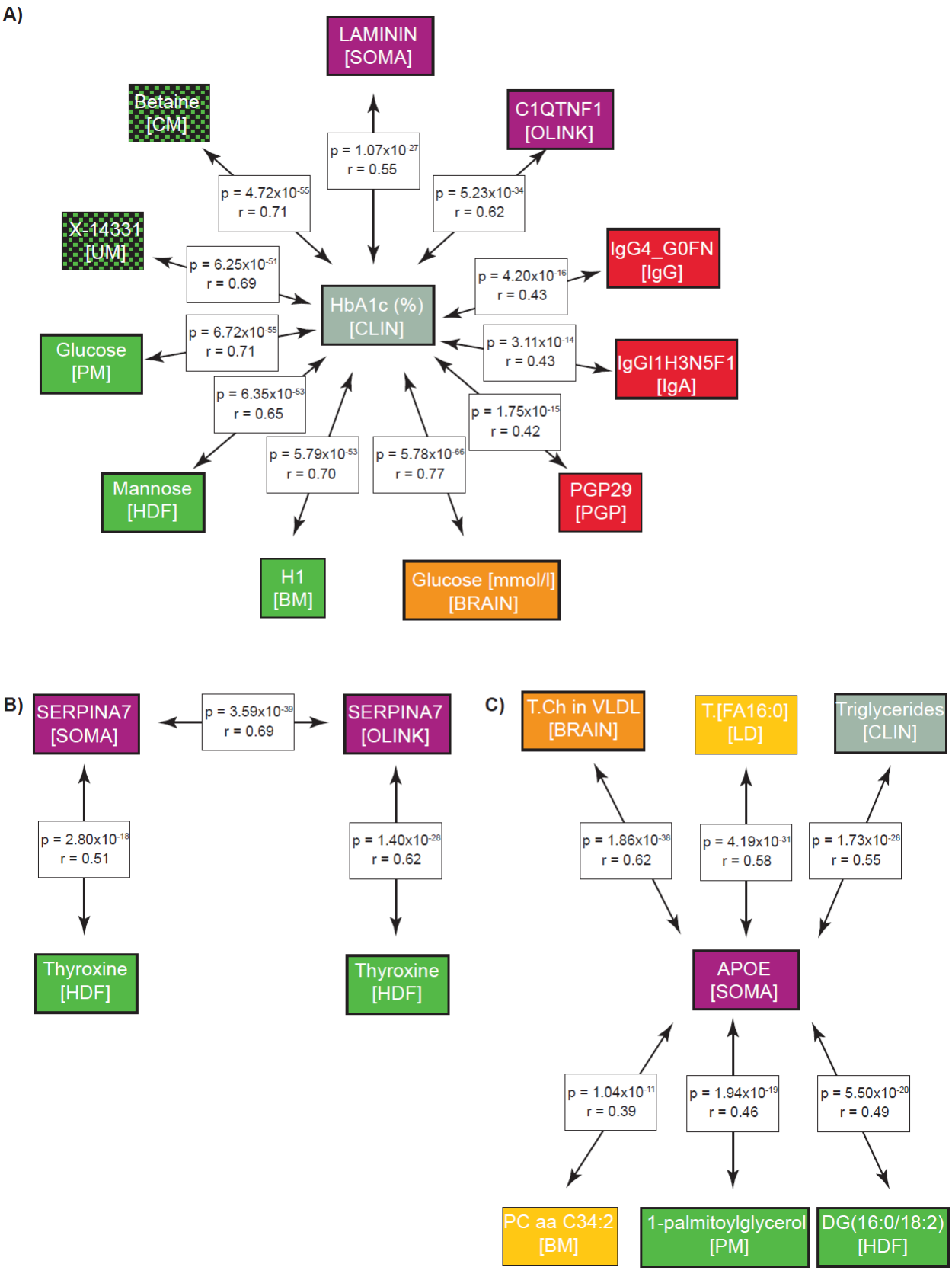
The molecular network generated for SIDD consists of 113 molecules measured across 14 platforms (**Supplementary Data 20** and **Supplementary Figure 5**). Those molecules are highly interconnected and display subclusters around alpha-ketobutyrate, gamma-glutamylphenylalanine, sphingomyeline (18:2/24:2), mannose-complement C2, coagulation Factor XI (F11), cortisol and glucose. We have

identified an association between F11 and kallikrein (KLKB1). Both proteins are components of the contact system recognized as potent procoagulant and proinflammatory plasma protease cascade [29]. Both proteins showed association with SNPs, where F11 associate with rs5030062 and KLKB1 with rs710446, located on chromosome 3 near to Kininogen-1 (*KNG1*). Interestingly, *KNG1* is part of the kinin-kallikrein system, which is also playing a role in inflammation, coagulation, and blood pressure control. Genetic variation in *KNG1* was shown to influence contact system protein levels [30]. Identified here, molecular interplay could indicate potential events promoting inflammation and endothelial dysfunction triggered by hyperglycemia-activated coagulation and prothrombotic state, relevant for SIDD cluster, which is characterized by poor glycemic control. We also found an association between KLKB1 and methylation site g01028142 near Cytidine monophosphate (UMP-CMP) kinase 2 (CMPK2). CMPK2 is a metabolic enzyme implicated in mitochondrial DNA synthesis, shown to be critical for NLRP3 inflammasome activation [31], which in turn is implicated in T2D and cardiovascular complications [32]. The interaction between genetic variants, protein, and methylation provides further insight into pathologies potentially related to SIDD subtypes. It suggests a susceptibility to cardiovascular events that could not be related without multiomics dataset.

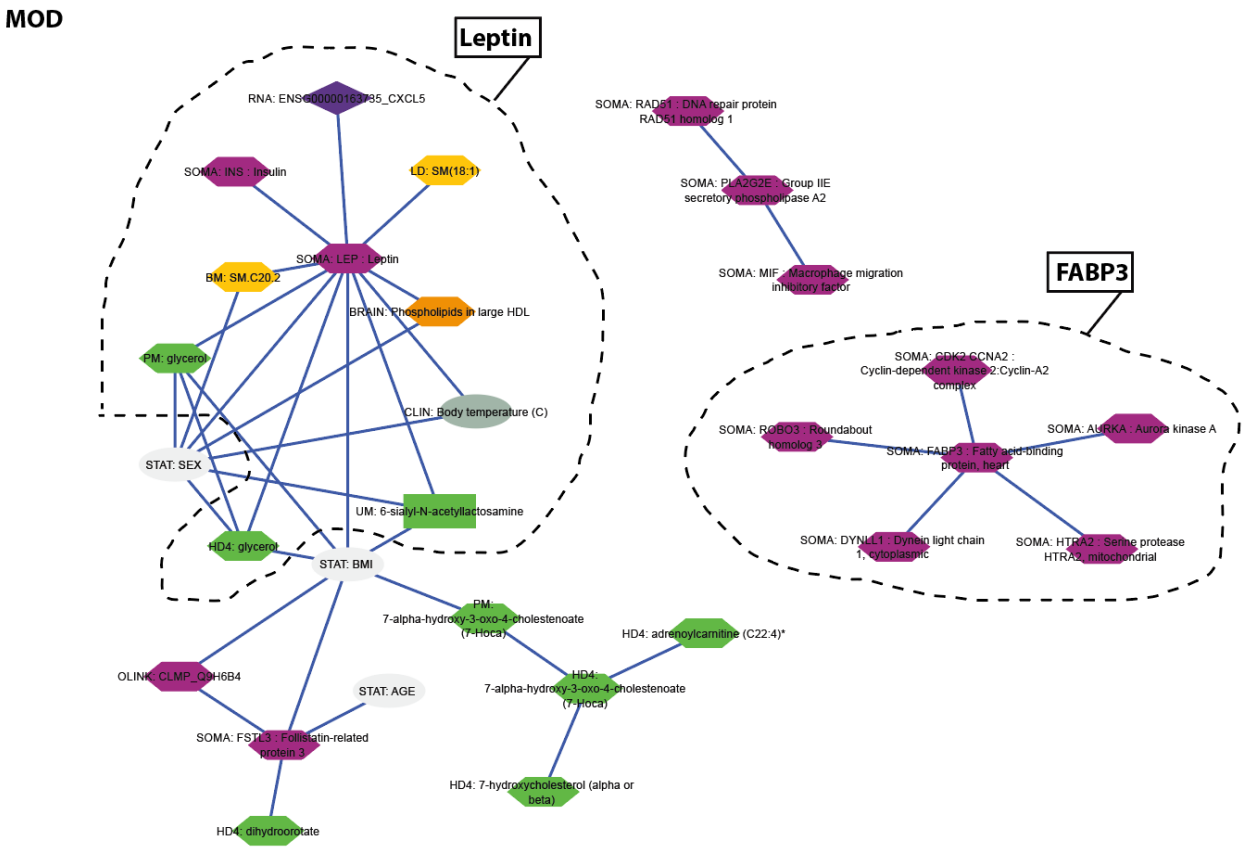
Supplementary Figures.



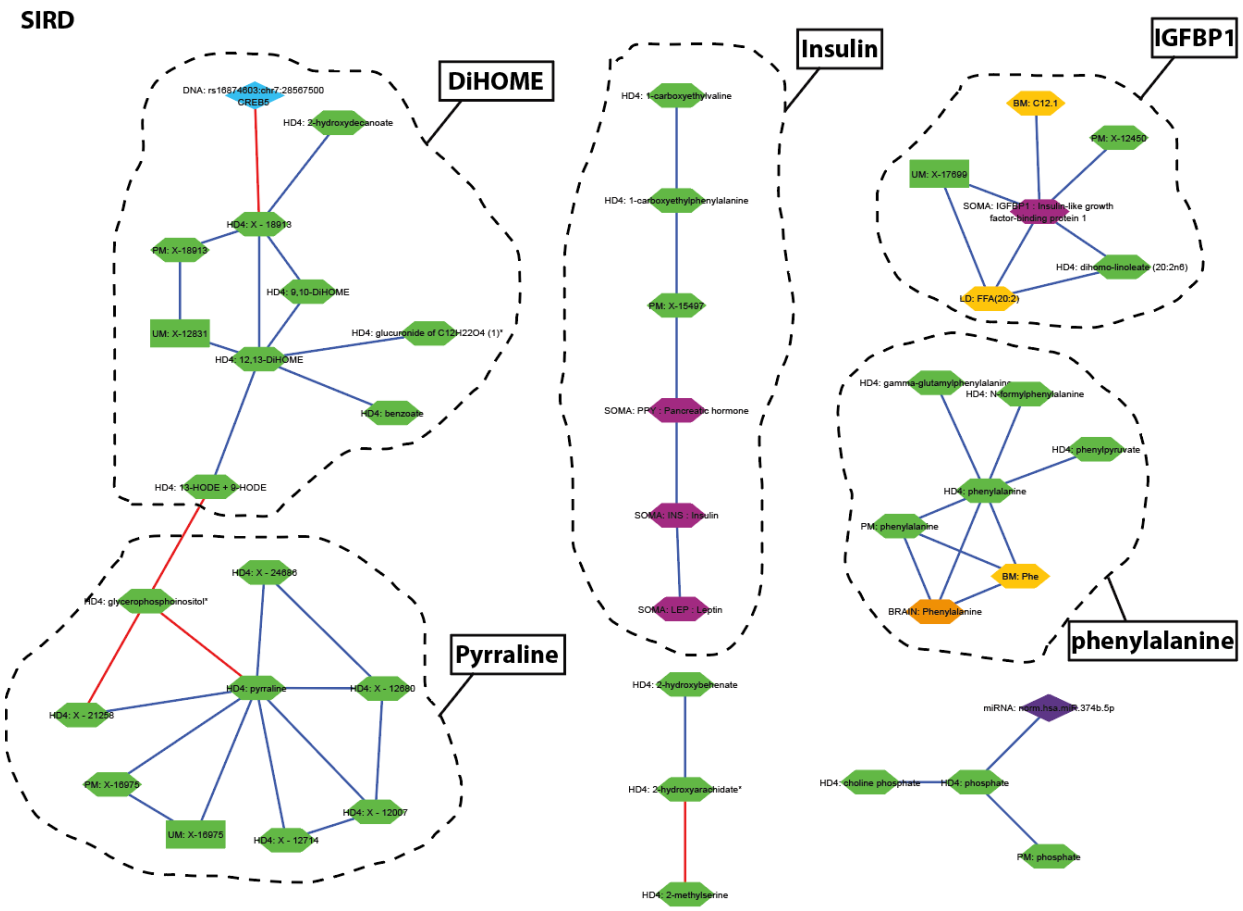
Supplementary Figure 1. Overview on the previous study using QmDiab datasets. The color code indicates each platform used in the given study. The letter codes above the colors reflect on each platform: DNA - Genotype; MET - DNA methylation; RNA - gene expression; miRNA - microRNA expression; OLINK - proteomics; SOMA - proteomics; PGP – glycomics (total plasma N-glycosylation); IgG - glycomics (IgG glycosylation); IgA – glycomics (IgA & IgG glycosylation); BRAIN - lipoproteomics; LD - lipidomics; BM - lipidomics; HDF – metabolomics from plasma with LC/MS method; PM – metabolomics from plasma with LC/MS & GC/MS method; SM – metabolomics in saliva with LC/MS & GC/MS method; UM – metabolomics from urine with LC/MS & GC/MS method; CM – metabolomics from urine with NMR method. The exact protein description is provided in Table 1 (main text).


A)
LAMININ [SOMA]
Betaine [CM]
C1QTNF1 [OLINK]
p = 1.07x10-27
r = 0.55
p = 4.72x10-55
r = 0.71
p = 5.23x10-34
r = 0.62
X-14331 [UM]
IgG4_G0FN [IgG]
p = 6.25x10-51
r = 0.69
p = 4.20x10-16
r = 0.43
HbA1c (%) [CLIN]
Glucose [PM]
p = 6.72x10-55
r = 0.71
p = 3.11x10-14
r = 0.43
IgGI1H3N5F1 [IgA]
p = 6.35x10-53
r = 0.65
p = 1.75x10-15
r = 0.42
p = 5.79x10-53
r = 0.70
p = 5.78x10-66
r = 0.77
Mannose [HDF]
PGP29 [PGP]
H1 [BM]
Glucose [mmol/l] [BRAIN]
B)
SERPINA7 [SOMA]
p = 3.59x10-39
r = 0.69
SERPINA7 [OLINK]
p = 2.80x10-18
r = 0.51
p = 1.40x10-28
r = 0.62
Thyroxine [HDF]
Thyroxine [HDF]
C)
T.Ch in VLDL [BRAIN]
T.[FA16:0] [LD]
Triglycerides [CLIN]
p = 1.86x10-38
r = 0.62
p = 4.19x10-31
r = 0.58
p = 1.73x10-28
r = 0.55
APOE [SOMA]
p = 1.04x10-11
r = 0.39
p = 1.94x10-19
r = 0.46
p = 5.50x10-20
r = 0.49
PC aa C34:2 [BM]
1-palmitoylglycerol [PM]
DG(16:0/18:2) [HDF]

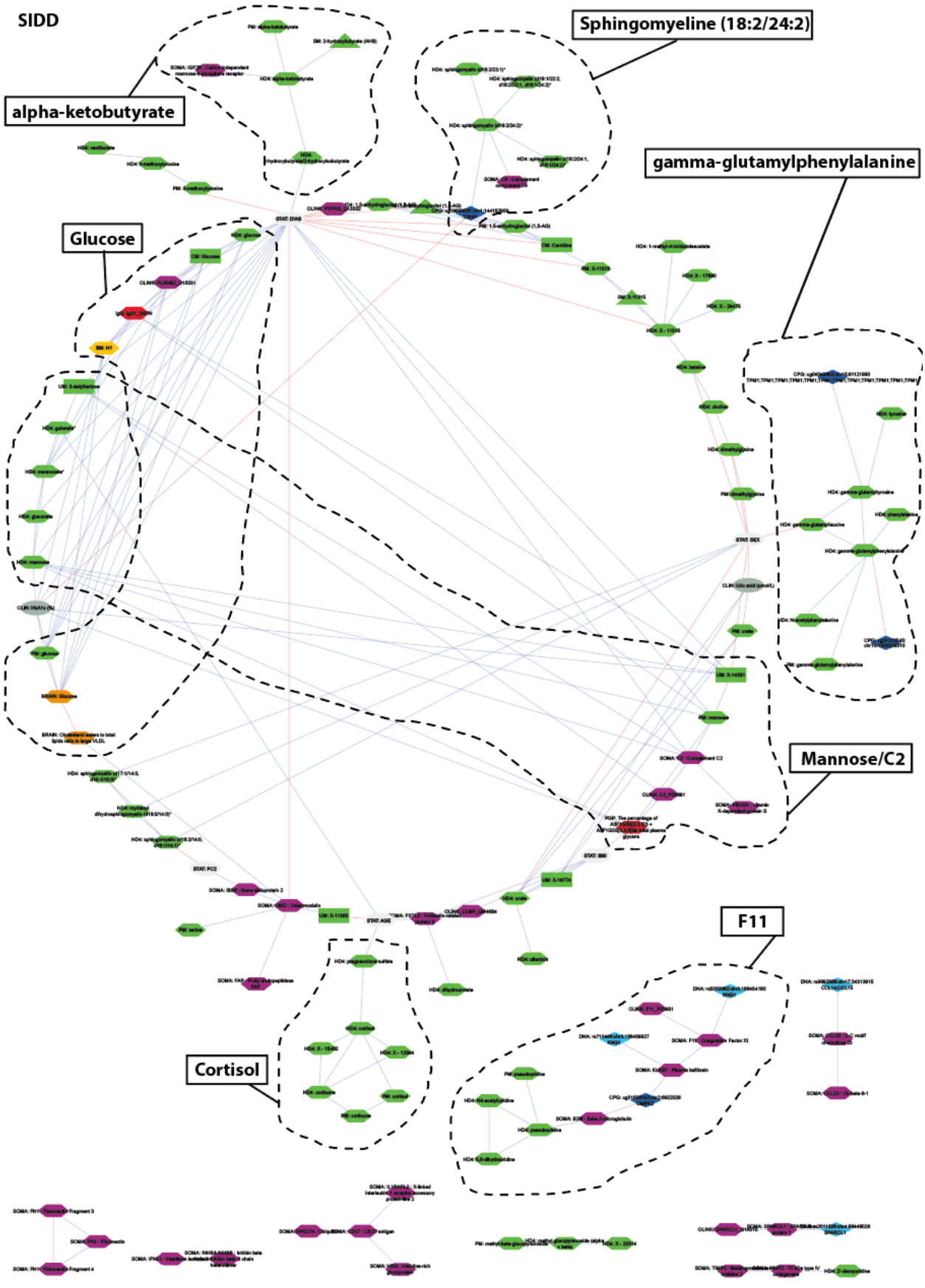
Supplementary Figure 2. A) MBH identifies clinically relevant association between HbA1C and molecules involved in pathology of T2D. **B)** MBH reflects on physiological processes such as thyroxine transport and **C)** lipid metabolism.



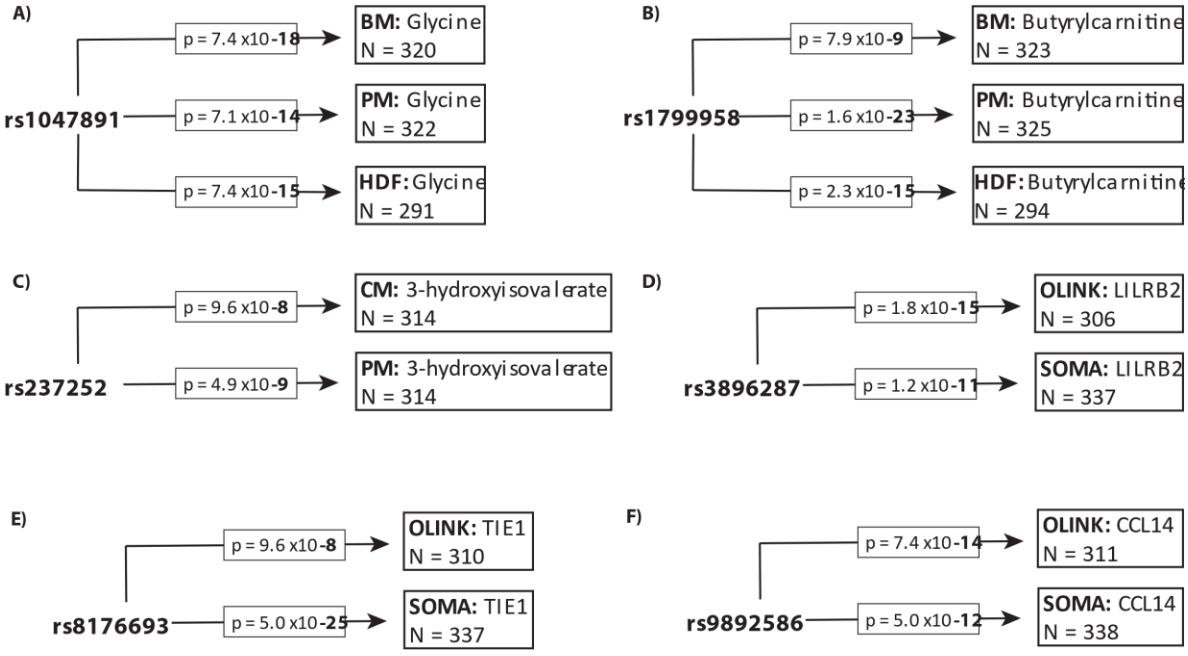
Supplementary Figure 3. Molecular network generated for mild obesity related (MOD) T2D subgroup characterized by high BMI with low insulin resistance.



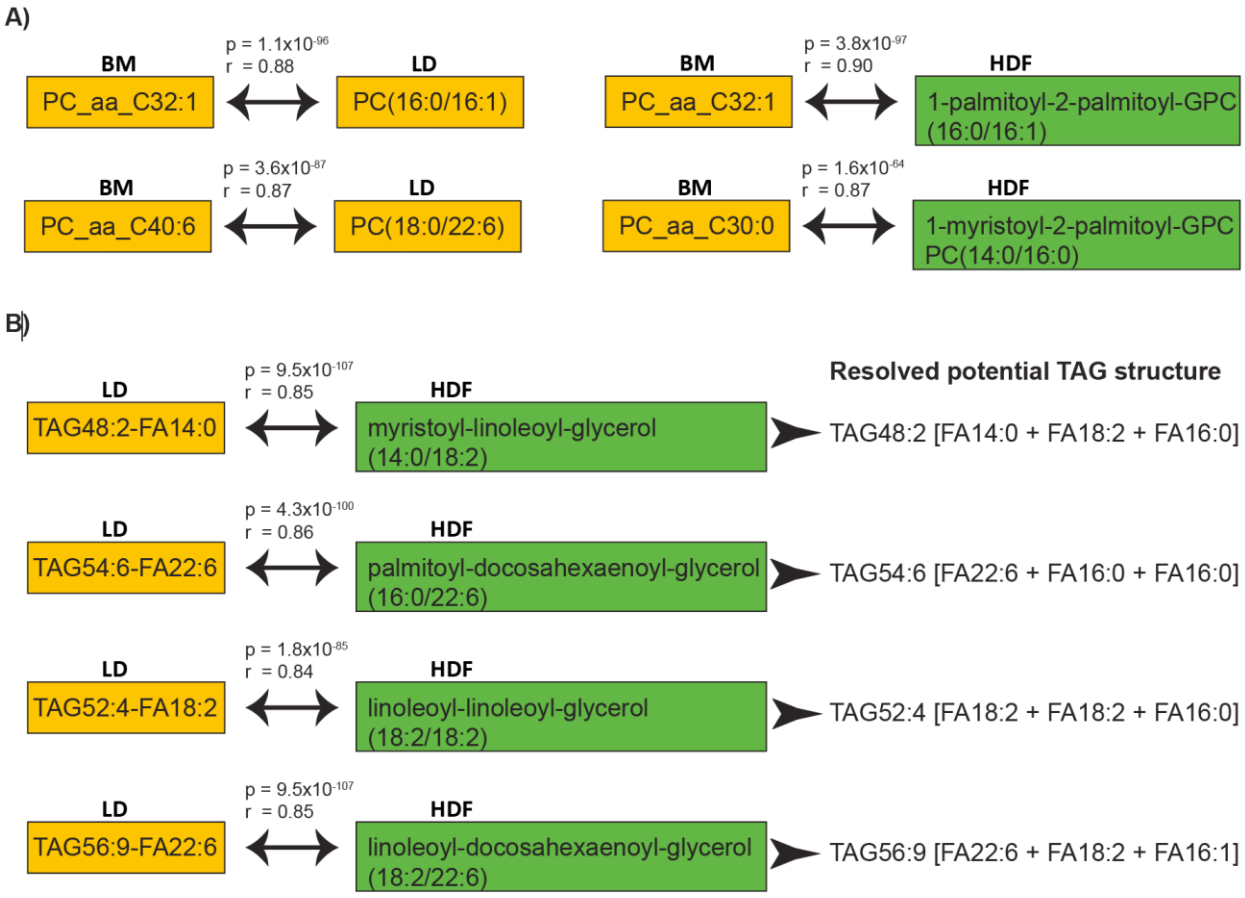
Supplementary Figure 4. Molecular network generated for severe insulin resistant (SIRD) T2D subgroup characterized by highest level of insulin resistance (HOMA2-IR) and high BMI.


SIDD
alpha-ketobutyrate
Sphingomyeline (18:2/24:2)
gamma-glutamylphenylalanine
Glucose
Mannose/C2
F11
Cortisol

Supplementary Figure 5. Molecular network generated for severe insulin deficient (SID) characterized by young age at onset, low BMI, low insulin secretion (HOMA2-B) and poor glycemic control (high HbA1c).

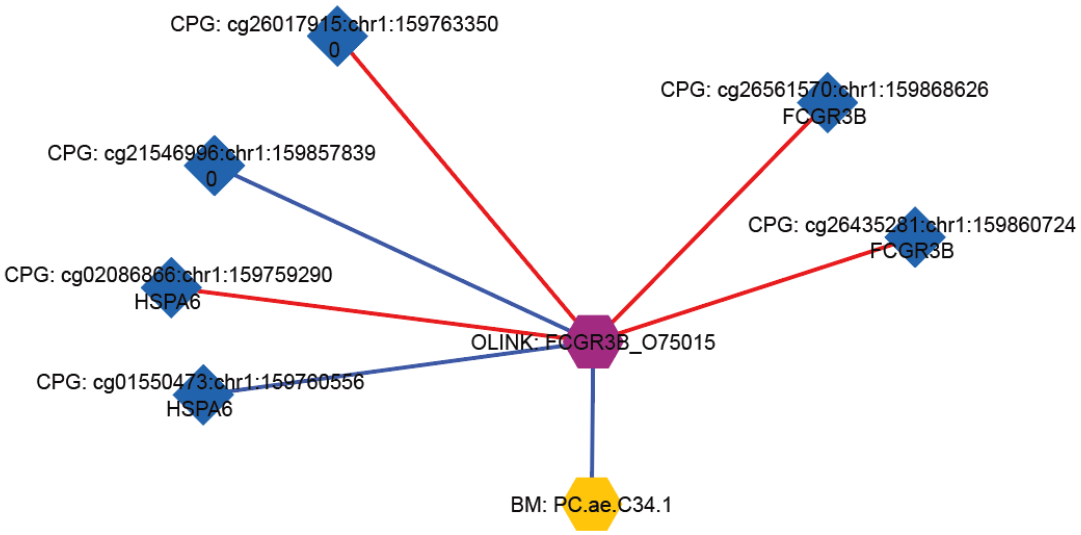


Supplementary Figure 6. Evaluation of platform performance through the strength of GWAS hits. A)-F) Multiomics GWAS hits.

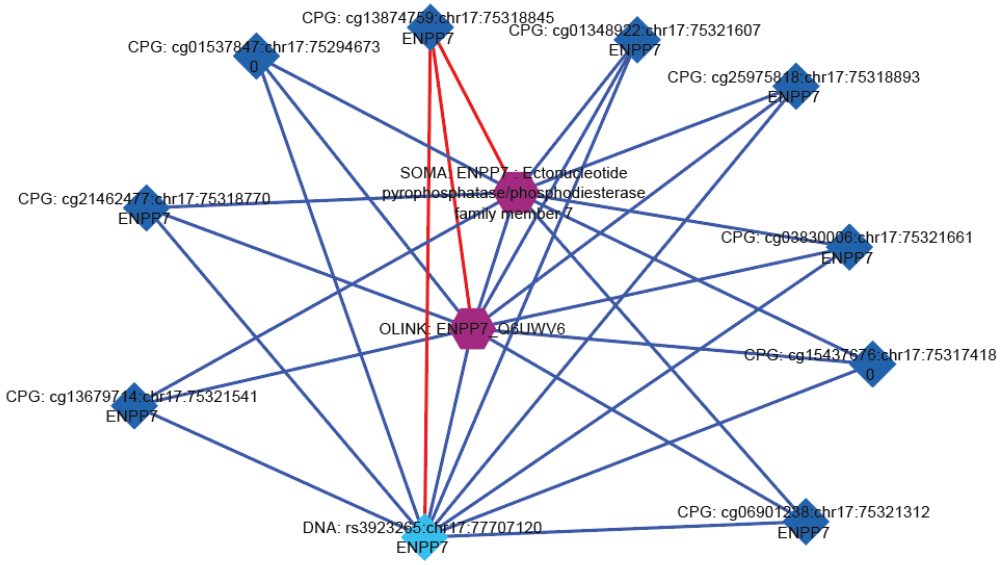


Supplementary Figure 7. Omics platforms overlap and complementarity. **A)** & **B)** The structure of complex lipids was revealed with by complementary platforms connected with MBH.

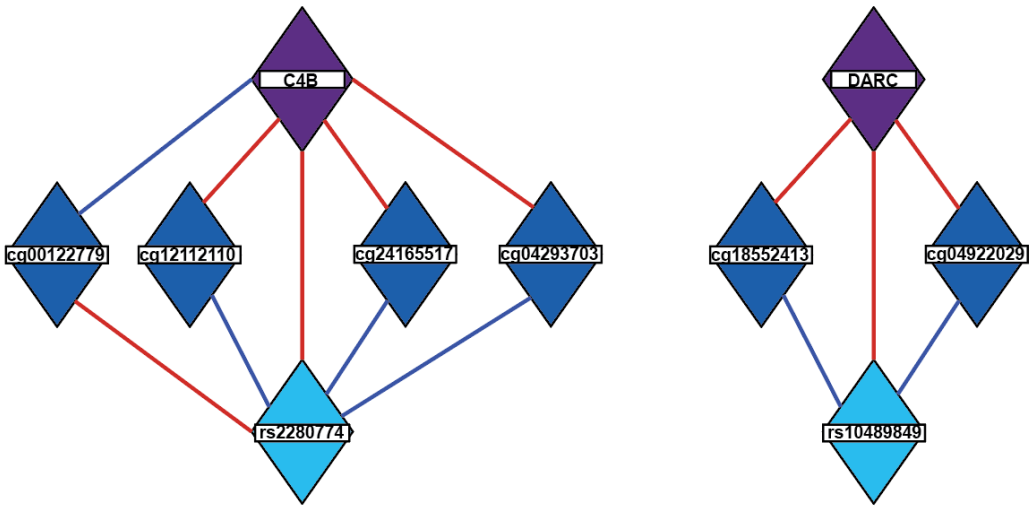
A)



B)



Supplementary Figure 8. Example of EWAS associations with proteins measured on OLINK. A) EWAS associations of Fc Gamma Receptor IIIb (FCGR3B) protein. B) EWAS associations of Ectonucleotide Pyrophosphatase/Phosphodiesterase 7 (ENPP7) protein.



Supplementary Figure 9. Example of association trios between the SNP-methylation, methylation-mRNA and SNP-mRNA.

Supplementary references

1. Filicori, M.*, et al.* Luteinizing hormone activity supplementation enhances follicle-stimulating hormone efficacy and improves ovulation induction outcome. *J Clin Endocrinol Metab* **84**, 2659-2663 (1999).
2. Quell, J.D.*, et al.* Characterization of Bulk Phosphatidylcholine Compositions in Human Plasma Using Side-Chain Resolving Lipidomics. *Metabolites* **9**, 109-109 (2019).
3. Kamat, M.A.*, et al.* PhenoScanner V2: An expanded tool for searching human genotype-phenotype associations. *Bioinformatics* **35**, 4851-4853 (2019).
4. Bonder, M.J.*, et al.* Disease variants alter transcription factor levels and methylation of their binding sites. *Nature genetics* **49**, 131-138 (2017).
5. Zaghlool, S.B.*, et al.* Epigenetics meets proteomics in an epigenome-wide association study with circulating blood plasma protein traits. *Nature Communications* **11**(2020).
6. Lotta, L.A.*, et al.* A cross-platform approach identifies genetic regulators of human metabolism and health. *Nature Genetics* **53**, 54-64 (2021).
7. Schlosser, P.*, et al.* Genetic studies of urinary metabolites illuminate mechanisms of detoxification and excretion in humans. *Nature Genetics* **52**, 167-176 (2020).
8. Shin, S.-Y.*, et al.* An atlas of genetic influences on human blood metabolites. *Nature Genetics* **46**, 543-550 (2014).
9. Imperatore, G., Knowler, W.C., Nelson, R.G. & Hanson, R.L. Genetics of diabetic nephropathy in the Pima Indians. Vol. 1 275-281 (Curr Diab Rep, 2001).
10. Bylander, J.E., Ahmed, F., Conley, S.M., Mwiza, J.M. & Ongeri, E.M. Meprin Metalloprotease Deficiency Associated with Higher Mortality Rates and More Severe Diabetic Kidney Injury in Mice with STZ-Induced Type 1 Diabetes. *Journal of Diabetes Research* **2017**(2017).
11. Gooding, J.*, et al.* Meprin β metalloproteases associated with differential metabolite profiles in the plasma and urine of mice with type 1 diabetes and diabetic nephropathy. *BMC Nephrology* **20**(2019).
12. Long, T.*, et al.* Whole-genome sequencing identifies common-to-rare variants associated with human blood metabolites. *Nature Genetics* **49**, 568-578 (2017).
13. Moore, A.F.*, et al.* The association of ENPP1 K121Q with diabetes incidence is abolished by lifestyle modification in the diabetes prevention program. *J Clin Endocrinol Metab* **94**, 449-455 (2009).
14. Gadd, D.A.*, et al.* Epigenetic scores for the circulating proteome as tools for disease prediction. *Elife* **11**(2022).
15. Bigler, J.*, et al.* A Severe Asthma Disease Signature from Gene Expression Profiling of Peripheral Blood from U-BIOPRED Cohorts. *American journal of respiratory and critical care medicine* **195**, 1311-1320 (2017).
16. Tannenbaum, C.S.*, et al.* The CXC chemokines IP-10 and Mig are necessary for IL-12-mediated regression of the mouse RENCA tumor. *Journal of immunology (Baltimore, Md. : 1950)* **161**, 927-932 (1998).
17. Zhao, A.Z., Bornfeldt, K.E. & Beavo, J.A. Leptin inhibits insulin secretion by activation of phosphodiesterase 3B. *J Clin Invest* **102**, 869-873 (1998).
18. Zeng, W.*, et al.* Sympathetic neuro-adipose connections mediate leptin-driven lipolysis. *Cell* **163**, 84-94 (2015).
19. Bonzon-Kulichenko, E.*, et al.* Central leptin regulates total ceramide content and sterol regulatory element binding protein-1C proteolytic maturation in rat white adipose tissue. *Endocrinology* **150**, 169-178 (2009).
20. Luheshi, G.N., Gardner, J.D., Rushforth, D.A., Loudon, A.S. & Rothwell, N.J. Leptin actions on food intake and body temperature are mediated by IL-1. *Proc Natl Acad Sci U S A* **96**, 7047-7052 (1999).

21. Siegrist-Kaiser, C.A.*, et al.* Direct effects of leptin on brown and white adipose tissue. *J Clin Invest* **100**, 2858-2864 (1997).
22. Chavey, C.*, et al.* CXC ligand 5 is an adipose-tissue derived factor that links obesity to insulin resistance. *Cell Metab* **9**, 339-349 (2009).
23. Lee, D.*, et al.* CXCL5 secreted from macrophages during cold exposure mediates white adipose tissue browning. *J Lipid Res* **62**, 100117 (2021).
24. Sarmiento, U.*, et al.* Morphologic and molecular changes induced by recombinant human leptin in the white and brown adipose tissues of C57BL/6 mice. *Lab Invest* **77**, 243-256 (1997).
25. Dodd, G.T.*, et al.* Leptin and insulin act on POMC neurons to promote the browning of white fat. *Cell* **160**, 88-104 (2015).
26. Adrian, T.E.*, et al.* Distribution and release of human pancreatic polypeptide. *Gut* **17**, 940-944 (1976).
27. Aragon, F.*, et al.* Pancreatic polypeptide regulates glucagon release through PPYR1 receptors expressed in mouse and human alpha-cells. *Biochim Biophys Acta* **1850**, 343-351 (2015).
28. Zhai, Y.*, et al.* Cysteine carboxyethylation generates neoantigens to induce HLA-restricted autoimmunity. *Science* **379**, eabg2482 (2023).
29. Weidmann, H.*, et al.* The plasma contact system, a protease cascade at the nexus of inflammation, coagulation and immunity. *Biochim Biophys Acta Mol Cell Res* **1864**, 2118-2127 (2017).
30. Rohmann, J.L.*, et al.* Genetic determinants of activity and antigen levels of contact system factors. *J Thromb Haemost* **17**, 157-168 (2019).
31. Zhong, Z.*, et al.* New mitochondrial DNA synthesis enables NLRP3 inflammasome activation. *Nature* **560**, 198-203 (2018).
32. Schroder, K., Zhou, R. & Tschopp, J. The NLRP3 inflammasome: a sensor for metabolic danger? *Science* **327**, 296-300 (2010).