Multi-omics data integration reveals clinically relevant biomolecules associated with type 2 diabetes

Xiaoyuan Zhou[1*], Alexey Stukalov[1], Brittany Lee[1], Shadi Ferdosi[1], Amir Alavi[1], Guhan Venkataraman[1], Harendra Guturu[1], Anna Halama[2], Frank Schmidt[2], Karsten Suhre[2], and Daniel Hornburg[1]

[1]Seer, Inc., Redwood City, CA 94065, USA ; [2]Weill Cornell Medicine-Qatar, Doha, State of Qatar

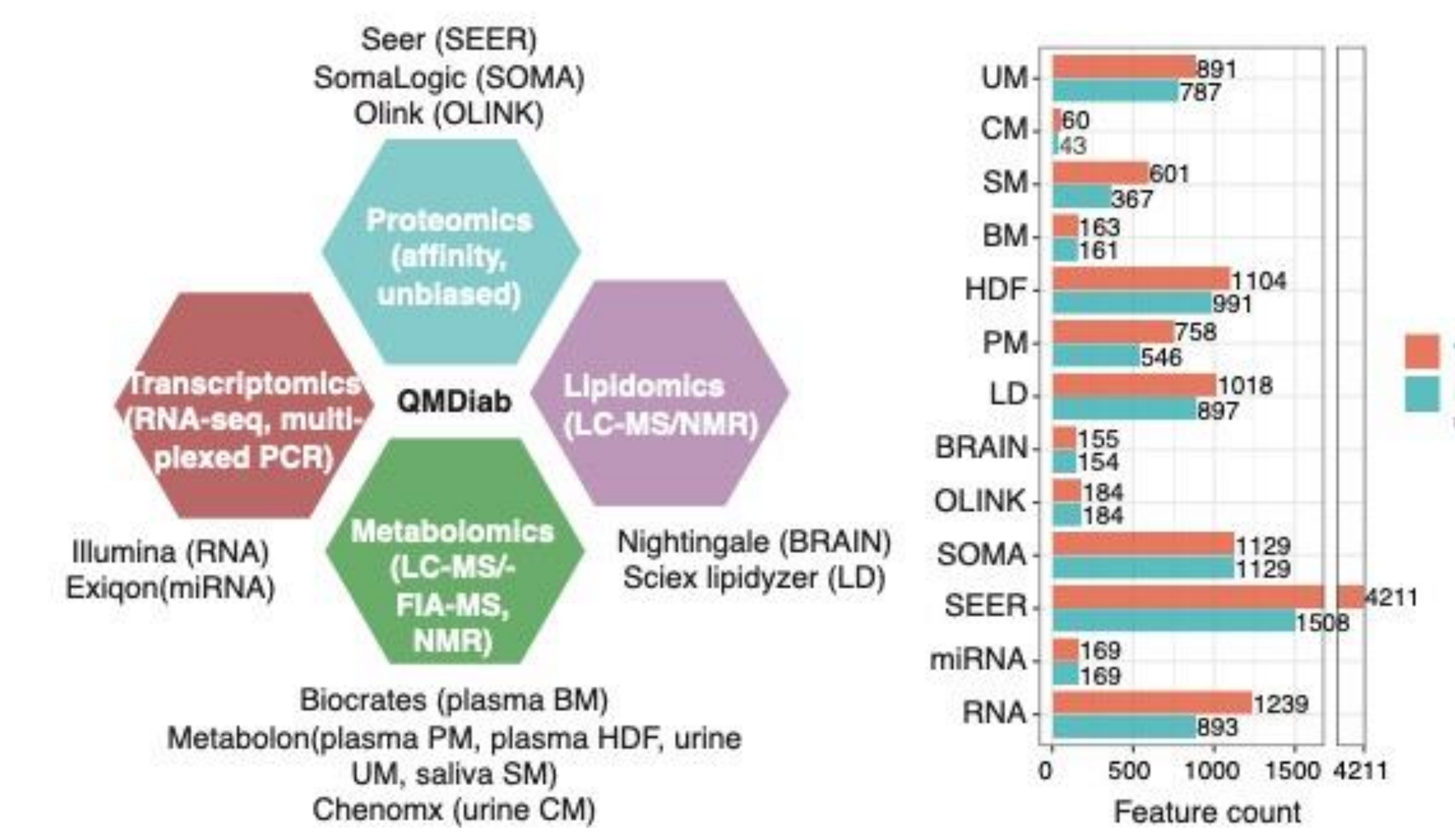
Deep and unbiased plasma proteomics for disease cohort studies at scale

Type 2 diabetes (T2D) is a complex metabolic disease and a major international health challenge. Discovering molecular signatures for accurate and early detection of T2D will be useful for disease prevention and to evaluate personalized risks. Despite the functional insights proteins can provide, the large dynamics range of the plasma proteome has historically required the trade-off between depth of coverage and study size.

To address this need, we introduced Proteograph™ Product Suite, a novel platform that leverages protein-coronas formed on the surface of functionalized nanoparticles (NPs) enabling deep and unbiased proteomics, detecting thousands of proteins at scale. We conducted a plasma proteome study using Proteograph™ Assay performed on 388 T2D cases and controls from Qatar Metabolomics Study of Diabetes (QMDiab). Here we integrated the comprehensive deep proteomics data with other omes collected from QMDiab cohort as previously described[1] to identify novel multi-omics signatures associated with T2D. Multi-omics data integration revealed multi-omics synergies, medication profiles, known and potential biomarker signatures involved in the dysregulation of lipase activity and lipid change in alignment with clinical measurements.

Multi-Omics Factor Analysis[1] (MOFA2) was applied to integrate proteomics data generated with Proteograph™ workflow[2] with other 12 data modalities from QMDiab cohort, and identify cross-ome variances related to phenotype

A QMDiab is a deep phenotyped cohort (195 T2D patients, 193 control) with molecular traits measured in multi-biofluids using distinct omics platform.



B MOFA decomposes M feature expression matrices (Y) into a matrix of factors (Z) for each sample and M weight matrices (W), one for each data modality.

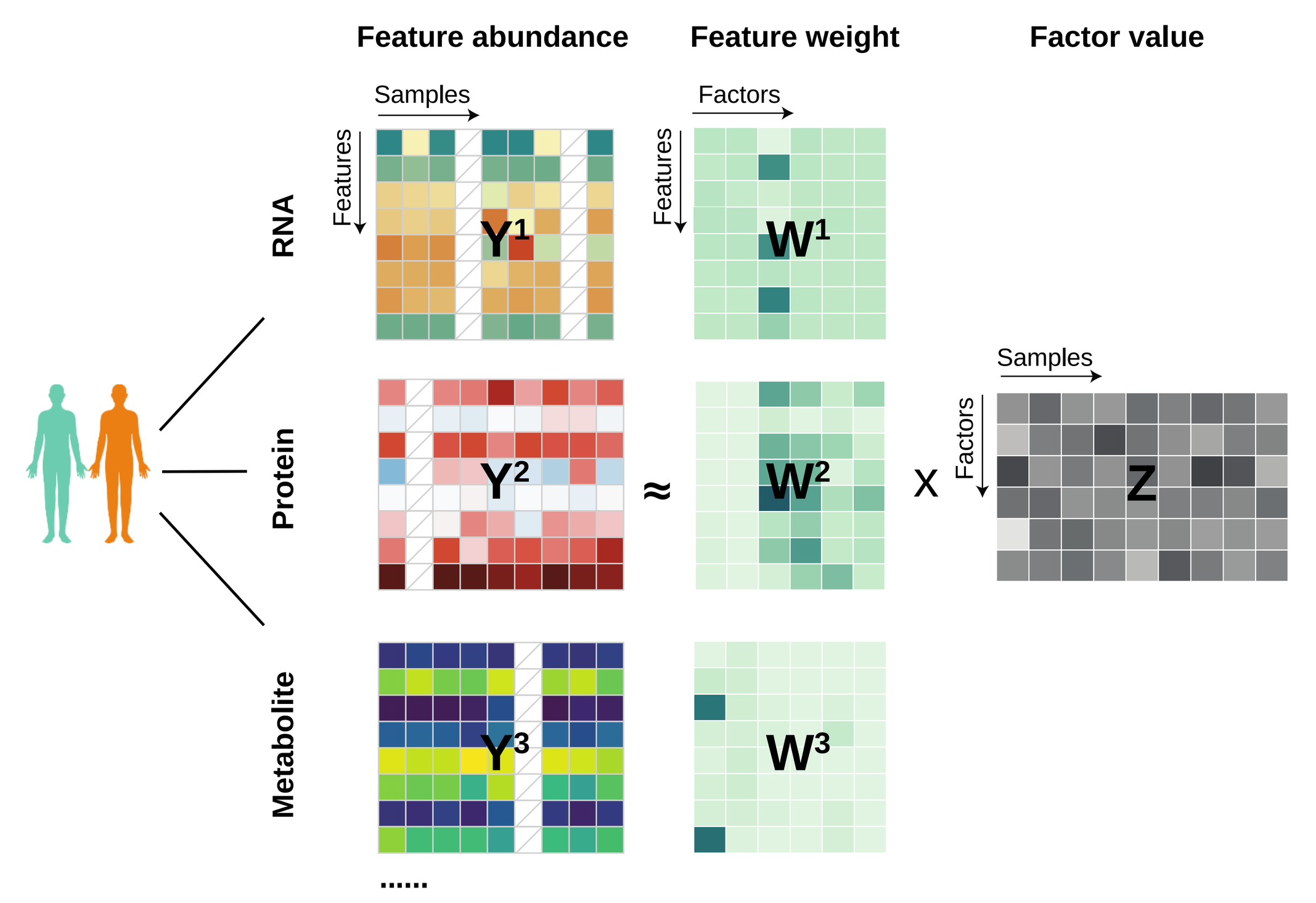


Table 1. QMDiab Subject Characteristics.

Feature	T2D	Control
Number	195	193
Female/Male	86/109	106/87
Age	53 (46.7,58.1)	40 (29.6,49.4)
BMI	30.6 (25.9,34.9)	28.8 (24.0,32)
HbA1c (%)	8.11 (6.8,9.1)	5.53 (5.3,5.8)

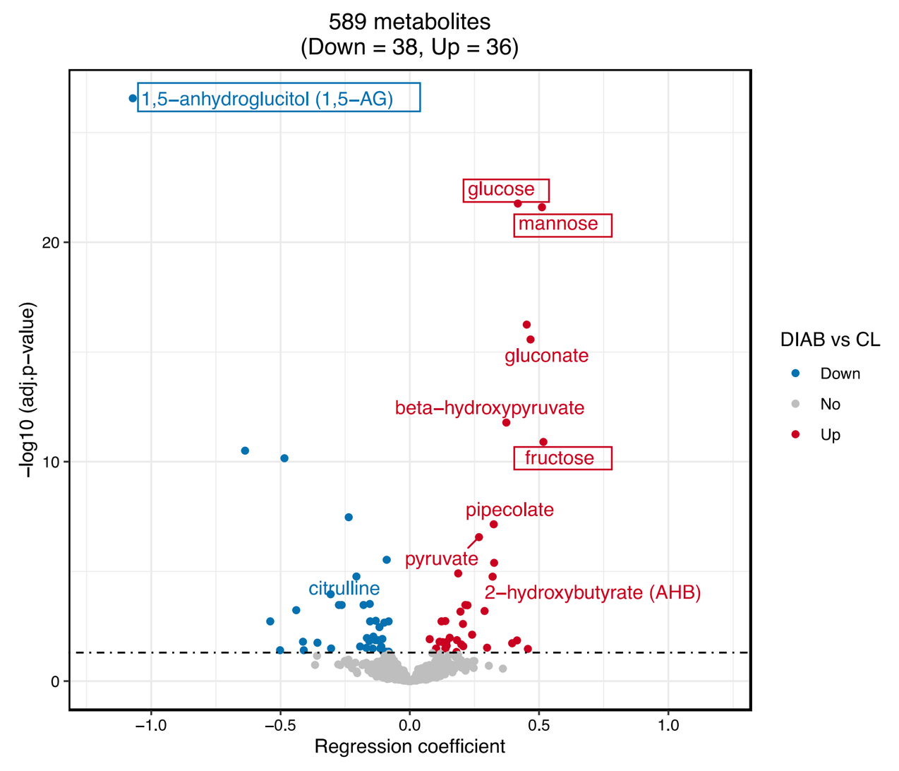
Integrate deep proteomics with other omes identified biomarkers related to type 2 diabetes

Methods

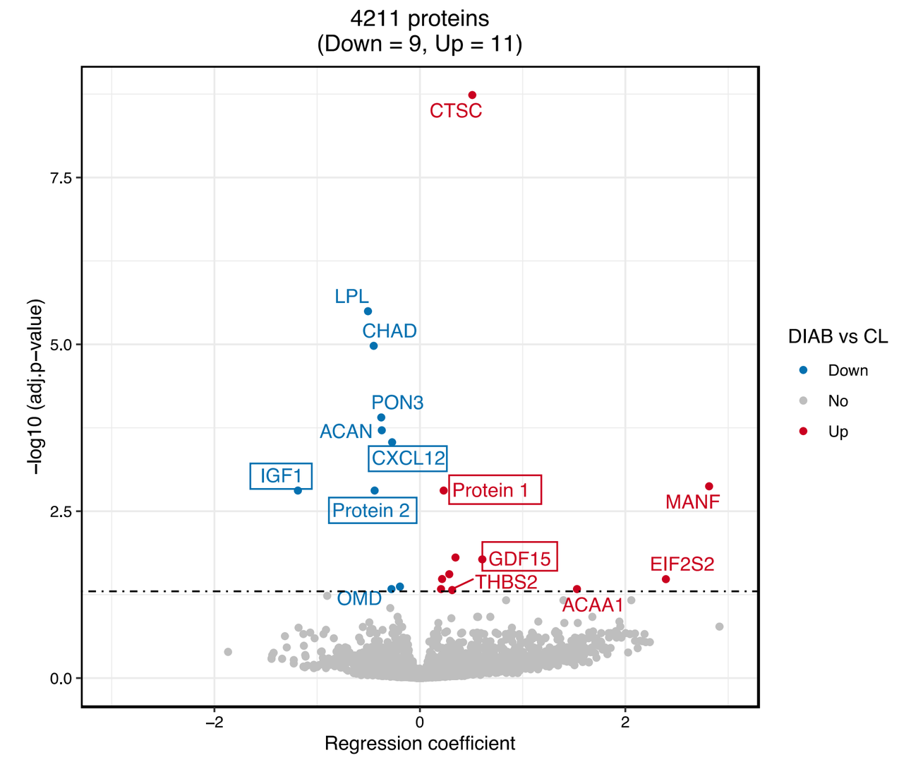
- Proteomics data was collected in DIA LC-MS mode using a 30-minute gradient on Bruker timsTOF Pro 2 mass spectrometer and data was analyzed with DIA-NN v1.8 in single group-run in library free mode.
- Processed transcriptomics, metabolomics and lipidomics data were collected from QMDiab cohort as described[3].
- Differential expression of single-ome molecular traits was computed using mixed linear regression models adjusted for age, sex and body mass index. P-value was corrected as FDR.
- Multi-Omics Factor Analysis[1] (MOFA2) was applied to identify latent factors that represent underlying data variance across multi-omics of QMDiab cohort, including proteomics, transcriptomics, metabolomics and lipidomics (Figure 1).
- MOFA factors were associated with phenotype and clinical measurements using Pearson correlation implied in MOFA, and further enriched in Gene Ontology terms based on ranked feature weights.

Results

A Known metabolite biomarkers[4] were identified in plasma (PM).

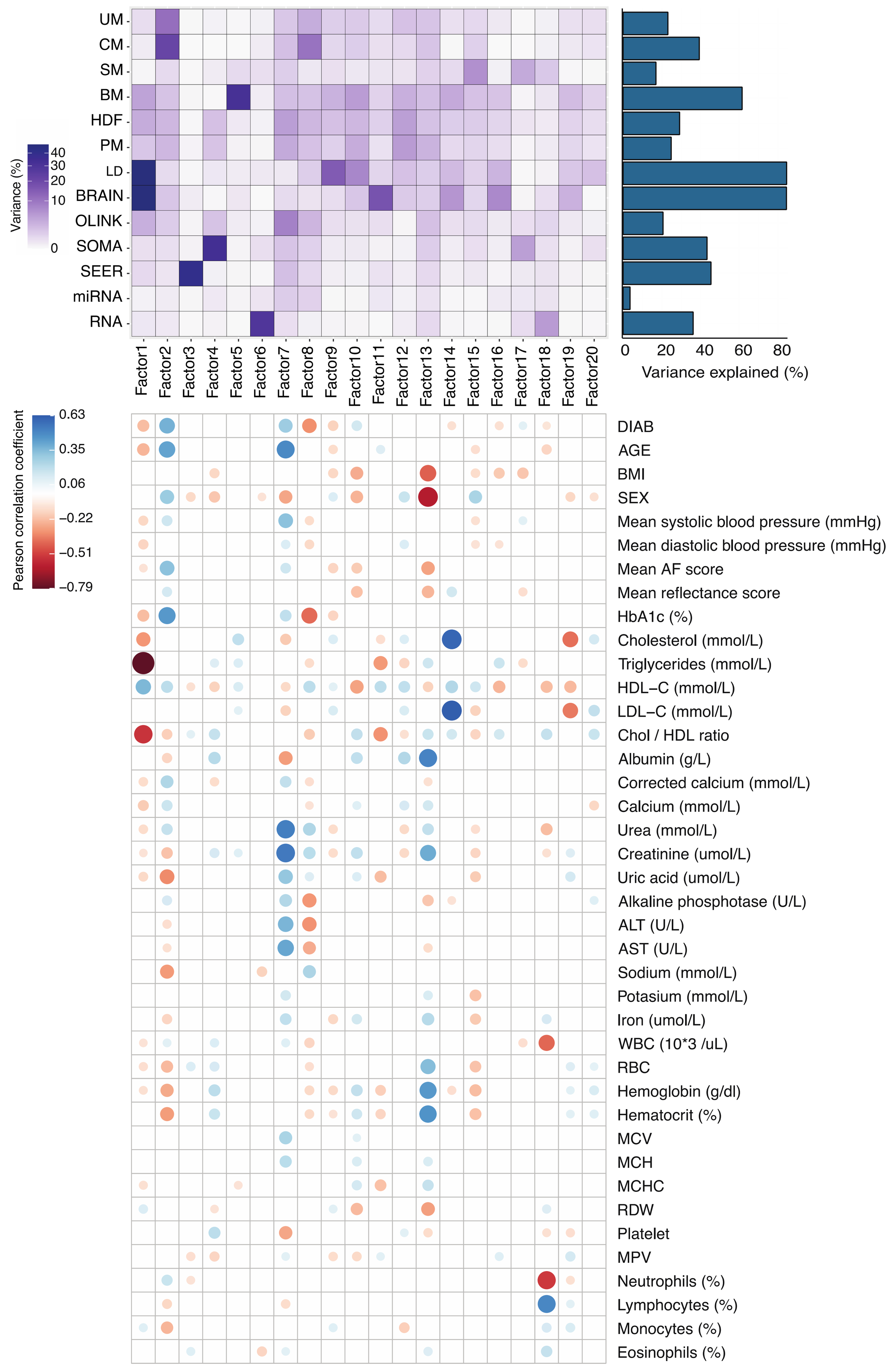


B The Proteograph™ workflow successfully detected known (IGF1,CXCL12) and potential low abundant biomarkers (e.g protein 1,2), and the increase of GDF15 may be regulated by medicines.



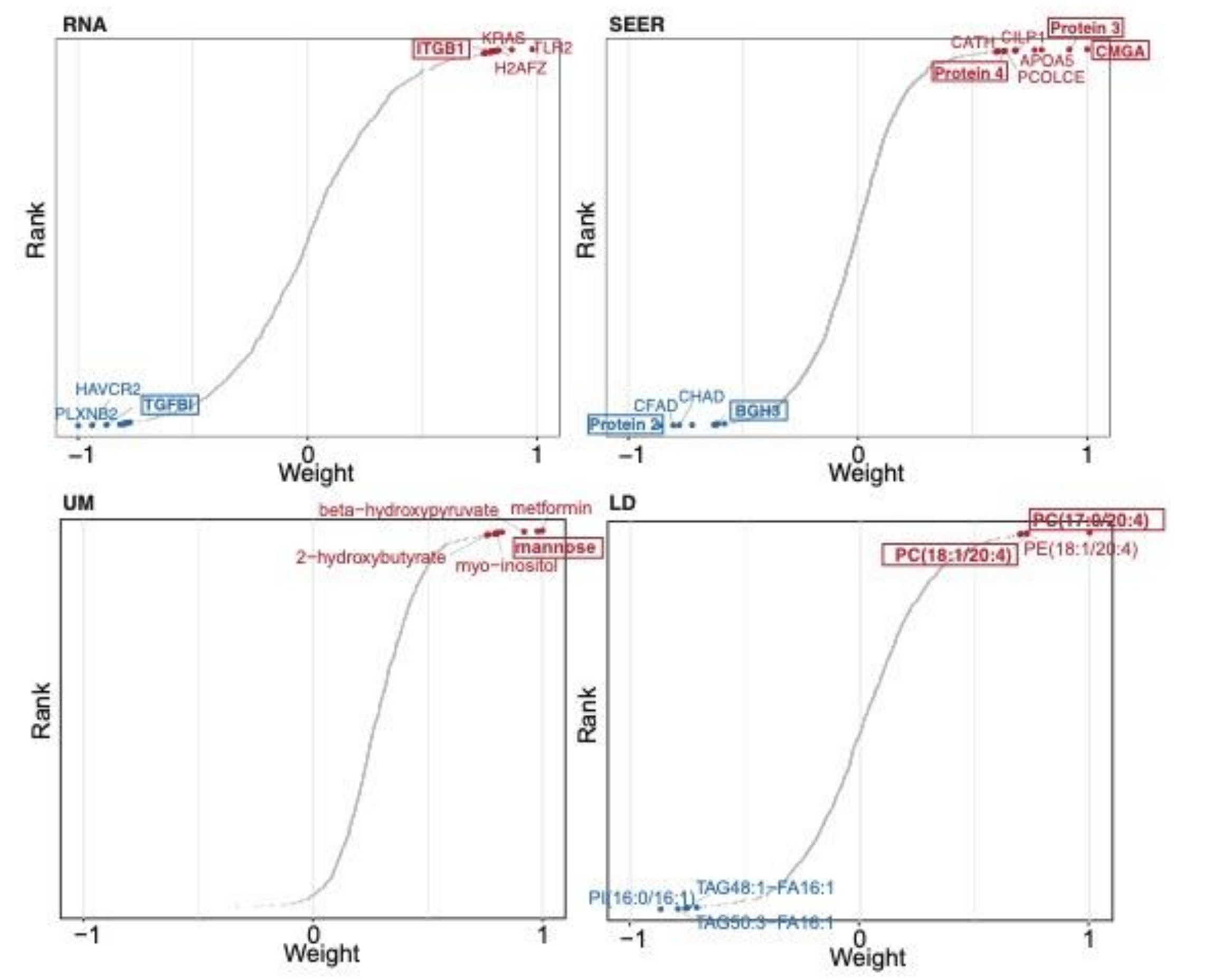
MOFA2 captured T2D associated biological variances.

Factor 1, 2 ,7 and 8 captured cross-ome variance, and had significant associations with diabetes and clinical traits such as triglycerides, Hb1Ac. These variances are more likely to be involved in T2D specifically.

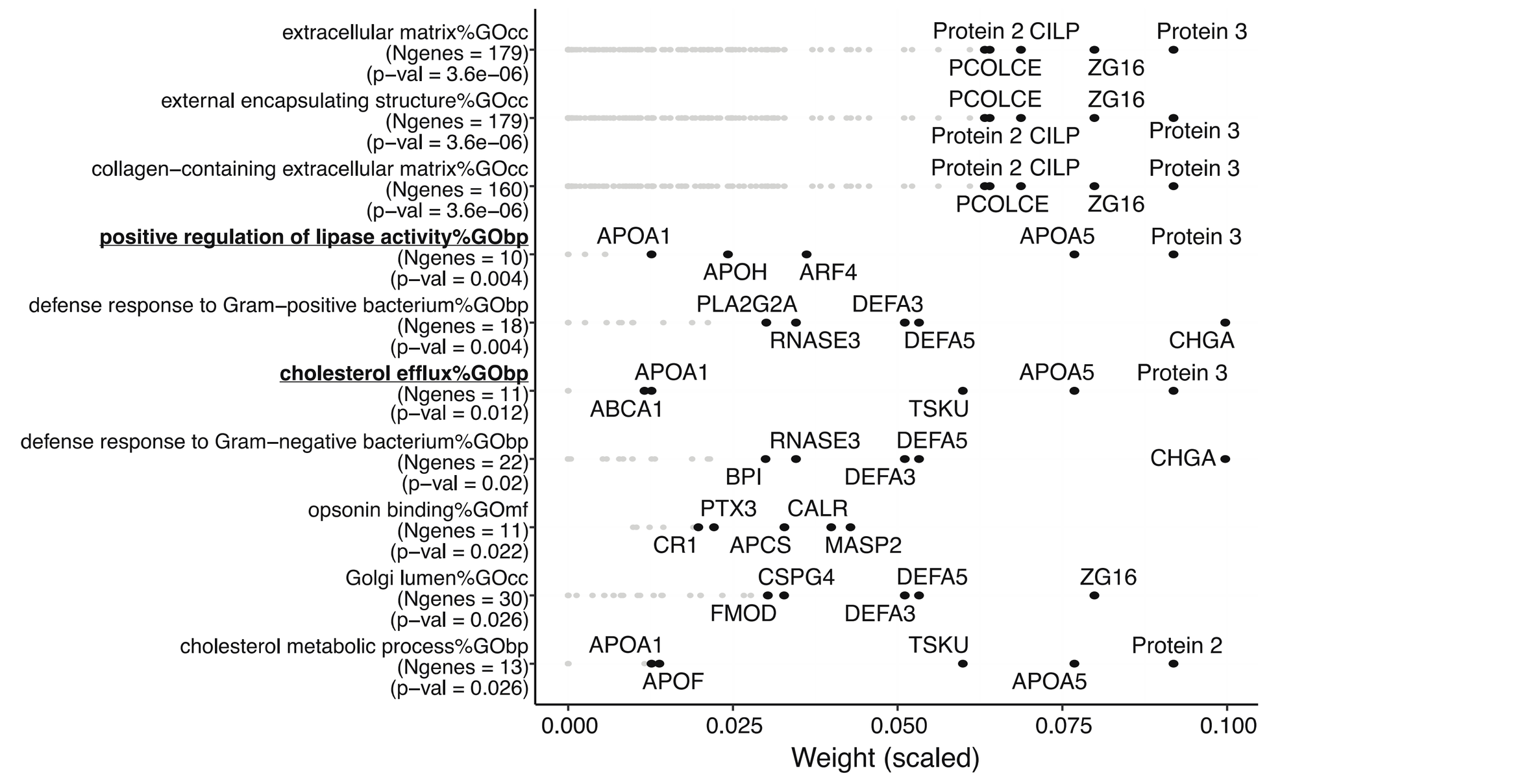


Multi-omics data integration revealed potential biomarker signatures, medication profiles, and multi-omics synergies.

Factor 2 captured **i)** difference of gene ITGB1(integrin Subunit Beta 1), protein Chromogranin A (CMGA) and mannose among the known T2D associated molecules[3],and potential biomarker signatures; **ii)** metformin intake in T2D; **iii)** synergies of TGF-β1 (transforming growth factor-β1) gene expression and encoded protein BGH3 level.



The proteins identified with Proteograph workflow weighted in factor 2 were significantly enriched in functional pathways related to T2D.



Conclusion

- Single ome captures biological variance, such as known T2D biomarkers 1,5-AG, sugar metabolites, IGF1 protein and low abundant proteins, but to different degree.
- Multi-omics integration using MOFA identified biological variances, highlighting the cross-ome features related to T2D phenotype and clinical measurement.
- Multi-omics integration revealed medication profiles, potential biomarker signatures and multi-omics synergies.

References

[1]Argelaguet, R. et al. Mol Syst Biol (2018)
[2]Suhre, K. et al. bioRxiv (2023)
[3]Halama, A. et al. Medrxiv (2022)
[4] Mook-Kanamori, D. O. et al. J Clin Endocrinol Metabolism (2014)

Publications