



Lessons and Applications of Omics Research in Diabetes Epidemiology

Gechang Yu^{1,2,3} · Henry C. H. Tam^{1,2,3} · Chuiguo Huang^{1,2,3} · Mai Shi^{1,2,3} · Cadmon K. P. Lim^{1,2,3} · Juliana C. N. Chan^{1,2,3} · Ronald C. W. Ma^{1,2,3}

Accepted: 4 January 2024 / Published online: 31 January 2024
© The Author(s) 2024

Abstract

Purpose of Review Recent advances in genomic technology and molecular techniques have greatly facilitated the identification of disease biomarkers, advanced understanding of pathogenesis of different common diseases, and heralded the dawn of precision medicine. Much of these advances in the area of diabetes have been made possible through deep phenotyping of epidemiological cohorts, and analysis of the different omics data in relation to detailed clinical information. In this review, we aim to provide an overview on how omics research could be incorporated into the design of current and future epidemiological studies.

Recent Findings We provide an up-to-date review of the current understanding in the area of genetic, epigenetic, proteomic and metabolomic markers for diabetes and related outcomes, including polygenic risk scores. We have drawn on key examples from the literature, as well as our own experience of conducting omics research using the Hong Kong Diabetes Register and Hong Kong Diabetes Biobank, as well as other cohorts, to illustrate the potential of omics research in diabetes. Recent studies highlight the opportunity, as well as potential benefit, to incorporate molecular profiling in the design and set-up of diabetes epidemiology studies, which can also advance understanding on the heterogeneity of diabetes.

Summary Learnings from these examples should facilitate other researchers to consider incorporating research on omics technologies into their work to advance the field and our understanding of diabetes and its related co-morbidities. Insights from these studies would be important for future development of precision medicine in diabetes.

Keywords Diabetes · Epidemiology · Biomarkers · Genetics · Epigenetics · Biobanking · Phenotype · Precision medicine

Gechang Yu, Henry CH Tam, Chuiguo Huang and Mai Shi contributed equally to this work.

✉ Ronald C. W. Ma
rcwma@cuhk.edu.hk

¹ Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, HKSAR, China

² Chinese University of Hong Kong- Shanghai Jiao Tong University Joint Research Centre in Diabetes Genomics and Precision Medicine, Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Hong Kong, HKSAR, China

³ Laboratory for Molecular Epidemiology in Diabetes, Li Ka Shing Institute of Health Sciences, The Chinese University of Hong Kong, Hong Kong, HKSAR, China

Introduction

There has been a marked increase in the prevalence of diabetes with 420 million people affected globally [1]. Traditional population- and family-based epidemiological studies have provided important insights on distribution of disease in time, place and person as well as trends over time. These data have generated hypotheses regarding aetiologies and mechanisms. Recent advances in genotyping and the study of genetics and other molecular markers together with the availability of different high-throughput platforms used to investigate different omics layers have ushered an exciting era in omics research. When these technologies are applied to well-characterized epidemiological cohorts, they can provide deeper insights on aetiology, disease pathways, prognosis and causation to improve the precision of diagnosis and classification for personalized treatment [2]. Public health

workers and practising physicians with interests in epidemiology and precision medicine are in a prime position to set up new cohorts or leverage existing cohorts to discover diagnostic and risk stratification tools as well as drug targets aimed at preventing and improving clinical outcomes. In this review article, we provided a brief overview on how omics research could be incorporated into the design of current and future epidemiological studies. Throughout the review, we drew on examples from the literature, as well as our own experience of conducting omics research using the Hong Kong Diabetes Register and Hong Kong Diabetes Biobank, as well as other cohorts, to illustrate the value of adding omics research into epidemiological studies. Learnings from these examples should facilitate other researchers to consider incorporating research on omics technologies into their work to advance the field and our understanding of diabetes and its related co-morbidities.

Preparation for Omics Studies and Sample Considerations

Whilst there is a wide spectrum of omics that can be considered, ranging from information on the genome to epigenome, and the transcriptome, proteome, metabolome and others, the needs of the project need to be balanced against the costs and time required for collecting the required biospecimens and subsequent processing costs. Therefore, the best time to plan for omics studies is definitely before embarking on a new epidemiological study, so that the main focus of the project can be addressed, whilst also trying to “future proof” the cohort by considering any potentially necessary samples for future research. That said, it is possible to add the collection of specific samples after a project has commenced. For example, it is not uncommon for genetic data to be added during the follow-up phases of a longitudinal cohort. Whilst it is true that the DNA sequence cannot be changed and hence obtaining DNA at a later stage for genotyping or sequencing would not give rise to results that are different to results that would be generated if the DNA samples were collected at baseline, other non-coding changes in DNA, such as methylation or other epigenetic changes, including leukocyte telomere length (LTL), do change with time and environmental exposure, and hence would differ at follow-up compared to baseline, hence highlighting the additional benefit of having considered this upfront, and the potential benefit of serial collection of DNA to search for changes in these epigenetic marks. Table 1 summarizes some of the different types of omics research, and some general comments regarding the sample types and key considerations.

Another key issue that warrants some discussion is pre-analytic considerations. In addition to collecting the right types of specimens for subsequent omics profiling, it is

important to bear in mind the different characteristics or pitfalls of analysis using the different platforms. Whilst genetic variations represent stable and permanent changes in the DNA sequence, methylation changes are subject to aging, effects of environment and treatment. On the other hand, metabolomic markers are more dynamic and unstable, and subject to metabolic perturbations as well as degradation. Whilst most protein markers are relatively stable, peptide hormones such as insulin are easily degraded. Hence, sample collection, transport and storage need to take into account the need to preserve the most labile or unstable analytes, and the adoption of best practice for sample collection and biobanking can enhance the success of subsequent omics profiling. Minimal sample handling, ensuring minimal sample transport and ensuring low temperature during transit and temperature monitoring during transit before storage at -80°C are some of the necessary steps to ensure samples are suitable for metabolomics profiling [3]. Further details on biobanking are available from more specialized reviews on this topic, and the white papers from the UK Biobank represent another good source of information [4]. We should highlight that whilst these omics technologies can be applied to different sample types, we have focused our discussion on the analysis of samples that are most likely to be collected in the setting of epidemiological investigations, such as blood samples.

Genetic Studies, GWAS and Polygenic Risk Scores

Family studies have revealed notable genetic components for diabetes and diabetic complications [5–8]. Genetic studies of diabetes and its complications however underwent several phases as the molecular technology evolved [9]. As one of the earliest efforts, linkage analysis had limited discoveries of genetic loci with robust large effects. Candidate gene studies, another approach commonly adopted, were prone to false positives and confounders. Early genome-wide association studies (GWAS) provided an alternative, high-throughput solution, but suffered from insufficient sample sizes to identify loci with modest effect sizes, which in fact account for the majority of the genetic component underlying complex traits [10]. With the rapid development of sequencing technologies, establishment of large biobanks and collaborative meta-analyses aggregating a large number of diabetes cohorts over the last decade, major breakthroughs have been achieved in the understanding of the genomics and genetics related to diabetes. Numerous studies conducted in diverse populations have greatly advanced our understanding on the genomic architecture and genetic heritability of complex diseases including diabetes. Similar studies have been undertaken in the Hong Kong Diabetes

Table 1 Examples of the different types of omics research, the types of samples required and special considerations in the collection of samples and biobanking procedures in epidemiological studies. Abbreviations in the table: *CpG* region of DNA where a cytosine occurs next to a guanine nucleotide, *DNAase* deoxyribonuclease, *NMR* nuclear magnetic resonance, *PCR* polymerase chain reaction, *RRBS* reduced representation bisulfite sequencing, *SNP* single nucleotide polymorphism, *WGBS* whole genome bisulfite sequencing

Type of biomarkers	Sample type	Profiling methods or analytic approach	Typical sample sizes required	Special considerations
Genetics and genomes (e.g. genome wide association studies (GWAS))				
common variants/single genetic polymorphisms (SNPs)	DNA (commonly extracted from leukocytes or whole blood, but also potentially from other tissues)	Whole-genome genotyping arrays	Typically comparing thousands of cases versus controls	Quality of DNA less critical compared to other applications
Low frequency or rare variants / mutations	As above	Whole exome genotyping arrays or exome sequencing, targeted sequencing, whole genome sequencing	Family-based cohorts or large cohorts with sample size larger than GWAS due to low frequency of variants of interest	Population- or family-specific
All variants	As above	Whole genome sequencing	Variable depending on allele frequency and effect size	Limited by current high costs
SNPs in candidate genes	As above	Sanger sequencing, allele-specific PCR or real-time PCR	Variable depending on allele frequency and effect	Easier to perform but less informative

Register (HKDR) [11, 12] and Hong Kong Family Diabetes Study (HKFDS) [13, 14] using these approaches, ranging from linkage analyses [15], candidate gene studies [16], GWAS [17] and meta-analyses [18, 19], to more recent sequencing efforts.

Diabetes is classically divided into an earlier-onset autoimmune form (type 1 diabetes; T1D) and a later-onset non-autoimmune form (type 2 diabetes; T2D), which account for the two main types of diabetes. Genetic exploration of T1D has mainly focused on the human leukocyte antigen (HLA) region, which presents the strongest association with T1D [20–22]. In addition, more than 60 loci outside of the HLA region have also been linked to T1D, including common variants revealed by GWAS in genes involved in immune regulation such as *IL2RA* and *CTLA4* [23–25], as well as rare variants underscored by targeted sequencing of known loci, e.g. *PTPN22* [26]. Although sequencing studies are known to complement GWAS by providing a more comprehensive characterization of genetic variation in a locus,

there are currently limited large-scale sequencing efforts carried out in individuals with T1D.

The fast-growing epidemic of T2D and its devastating complications have spurred greater efforts in studying the genetics of T2D [27]. Earlier work in the genetic exploration of T2D has been extensively discussed elsewhere [28, 29]. To date, the largest GWAS on T2D is a multi-ancestry meta-analysis of more than 2.5 million individuals including over 0.4 million T2D cases [30]. This study identified 1289 distinct association signals mapped to 611 loci reaching genome-wide significance. Index genetic variants in the most recently identified disease loci tend to be predominantly common (minor allele frequency [MAF] over 5% in at least one ancestry group) and had small effect sizes (odds-ratios [OR] < 1.1). Apart from the expansion of known T2D loci, another contribution of this study was the detection of ancestry-correlated heterogeneity enriched at the T2D associations, which provided explanation to the varied allelic effects among populations of different ancestry groups.

Table 1 (continued)

Structural variants (SV) or copy number variations (CNV)	As above	long-read sequencing, SNP genotyping or other methods	size Complements short read sequencing which discards many SVs and CNVs which may be significant	Limited by current high costs
Epigenomes (1) – DNA methylation (e.g. epigenome wide association studies (EWAS))				
Methylation markers	Methylation at CpG sites using peripheral blood	Pyrosequencing, EpiTYPER, methylation-specific PCR	Hundreds to thousands of cases and controls or within cohorts	Choice of tissue highly relevant
As above	As above	Pre-designed methylation arrays (e.g. Infinium arrays) as the most commonly used platform	Hundreds to thousands of cases and controls or within cohorts	Need to adjust for cell type composition in blood, as well as batch effects. Prospective cohorts are preferred since reverse causation due to methylation changes secondary to phenotype being examined can confound the analysis
As above	As above	Methyl-binding domain (MBD)-sequencing, Reduced representation bisulfite sequencing (RRBS) etc.	Tens to hundreds of samples, but larger sample sizes likely preferred	Allows more comprehensive investigation on genome-wide scale for EWAS
As above	As above	Whole-genome bisulfite sequencing	Tens to hundreds of samples, but	Currently the gold standard, providing

Protein-coding variants or variants with low frequency have long been considered one of the explanations for the missing heritability, whereby common variants cannot fully account for the heritability for complex diseases like T2D [31]. An exome-array association study in over 80,000 T2D cases and 370,000 controls from five population groups reported 40 distinct coding-variant associations across 38 loci at study-wide significance, among which 16 (40%) association signals were located beyond known T2D loci [32]. Population-specific analyses and trans-ethnic meta-analyses revealed similar associations at the reported loci in

European-descent populations and the overall study population, except for a variant in *PAX4* that was unique in East Asians [27]. Interestingly, although the study had increased power and sufficient coverage to capture low-frequency variants, 35 (87.5%) of the 40 independent coding-variant associations were common (MAF > 5%) and with modest effects (ORs 1.02–1.36). The remaining five signals also did not exert stronger effects than expected on the susceptibility to T2D, with ORs ranging from 1.09 to 1.29. Notably, fine-mapping analysis of the associated variants in the context of regional linkage disequilibrium (LD) by large-scale GWAS

Table 1 (continued)

		(WGBS)	larger sample sizes likely preferred	single CpG site resolution, but limited by high costs
Epigenomes (2) : Histone modifications and chromatin conformation				
Open Chromatin	Preparation of fresh samples for precipitation of chromatin	DNase I hypersensitivity sequencing (DNase seq), (formaldehyde-assisted isolation of regulatory elements sequencing (FAIRE-seq) or assay for transposase-accessible chromatin using sequencing (ATAC-seq)	Tens to hundreds of samples	
Chromatin structure	DNA from cells or tissues	Chromosome Conformation Capture (3C) analysis and other derived techniques	Tens to hundreds of samples	Requires living cells
Chromatin structure	DNA from cells or tissues	Hi-C, single-molecule long-read sequencing	Tens to hundreds of samples	detects genome-wide chromatin interactions
Epigenomes (3): non-coding RNA				
Small non-coding RNA e.g. Micro-RNA (miRNA)	Serum, plasma, urine or tissues for extraction of miRNA	Candidate miRNAs by real-time PCR	Tens to hundreds of samples	Lithium heparin tubes not suitable for miRNA profiling. Haemolysis may affect levels of certain miRNAs (e.g. miRNA-126)
		High-throughput profiling by miRNA arrays or small RNA sequencing		
Long non-coding RNA e.g. long intergenic ncRNAs (lincRNAs)		Known targets by Tiling array Novel discovery by RNA-Seq		

data only confirmed the causality of 16 (40%) signals, where one-third of the top coding-variant associations were likely to be driven by local non-coding variations nearby. Such “false leads” could result in an inappropriate inference on the disease etiology. Therefore, coding variants prioritized in genetic association studies, though functional, might not necessarily be causal. Careful investigation is required to correctly interpret the findings and the underlying biological significance.

As sequencing technologies become more high throughput and the costs continue to drop, whole-exome and whole

genome sequencing (WES/WGS) have been increasingly applied to the study of the genetics of complex diseases [33, 34]. So far, the largest sequencing study in diabetes reported exome sequencing analyses of over 20,000 individuals with T2D and ~24,000 non-diabetic controls [33]. Gene-level association analysis of rare variants (MAF < 0.5%) identified 4 genes at exome-wide significance: *MC4R*, *PAM*, *SLC30A8* and *UBE2NL*. Of note, the association of *SLC30A8* with T2D was driven by 90 missense variants, where the reduced protein activity was linked to decreased T2D risk. However, a comparison of rare and common

Table 1 (continued)

Transcriptomes				
mRNA/transcriptomics	Blood total RNA or tissues of interest	Expression arrays, RNA sequencing	Tens to hundreds of samples	Expression profiles are highly tissue-specific
Proteomes				
Protein markers	Blood or urine	Mass spectrometry (MS) or multiplex targeted platforms (e.g. Somalogic, Olink)	Hundreds to thousands of samples	
	Blood or urine	Measurement of a single biomarker, often using antibody-based assays such as ELISA	Hundreds to thousands of samples	Less informative but more accessible
Metabolomes				
Non-targeted metabolomics	Blood or urine	Nuclear magnetic resonance (NMR) spectroscopy, MS, Fourier transform infrared (FTIR) and high-performance liquid chromatography (HPLC) etc.	Hundreds to thousands of samples	Conditions during sample collection, transport, processing and storage very important. Need to standardize factors including time of day for sample collection, fasting status etc.
Targeted metabolomics	Blood or urine	Includes NMR-based panels (e.g. Bruker, Nightingale)	Hundreds to thousands of samples	
Gut microbiome	Usually would utilize stool samples	16s ribosomal RNA sequencing, shotgun metagenomics	Hundreds to thousands of samples	Highly variable depending on locations, diet, populations and use of medications
Other 'omes				
Exposome	Blood, urine or other samples for profiling of environmental pollutants	Mass spectrometry or others	Hundreds to thousands of samples	Variable depending on the parameter or chemical being evaluated
Phenomes Other phenotype data including data collected from wearables	Depends on the sensors used to collect data, ranging from diet, physical activity, sleep, to other traits	Clinical assessment and wearables	Variable	Often added dimension of different duration of data collection and variation during data collection

variant associations in subjects with both sequencing and genotyping data suggested that sequencing had limited power in detecting single-variant associations when compared with imputed array-based genotype data. In addition, even the strongest gene-level signals for rare variants resulted from sequencing data only explained 25% of the heritability of the strongest single-variant signals for common variants, arguing against conventional speculation that rare variants might have a large contribution to T2D heritability. Power analysis in the study further emphasized the need for large sample size in sequencing studies, where the authors estimated more than one million samples would be required for rare-variant signals in validated T2D therapeutic targets to reach exome-wide significance. These results highlight that sequencing study and array-based GWAS should be complementary in understanding the genetics of complex diseases like diabetes, with the former identifying informative alleles and the latter for locus discovery and fine-mapping. With the anticipated drop in sequencing costs, the integration of WES/WGS with available GWAS data in well-characterized cohorts will remain the main approach for identifying causal genes for diabetes in the near future. Nevertheless, it is also important to note that for complex diseases with partial penetrance, the concept of multicausality and the interaction of genetic predisposition with exposures over the life course are important considerations in these analyses [35, 36], and the availability of deep phenotyping, as well as the application of these principles in data interpretation, are important to improve understanding based on the genetic results generated.

Similar strategies have been applied to identify the genetic basis of diabetic complications and related outcomes, though the progress of these efforts lags significantly behind the studies on the genetics of T2D, mainly due to the limited sample size of studies so far conducted, and the heterogeneity of definitions being used for the outcomes of interests, such as diabetic kidney disease [37, 38], in part due to natural history of disease and modification by interventions. A detailed discussion on the genetics of diabetic complications is beyond the scope of this review, and is covered in more detail elsewhere [39, 40].

Polygenic Risk Scores for Diabetes

Recent advances in GWAS for diabetes have helped drive progress in the prediction of individual genetic susceptibility to diabetes. One such development is the polygenic risk score (PRS), which aggregates the effects of genetic variants into one summary score. There have been increasing studies highlighting the utility of PRS not only for risk stratification, but also for predicting diabetes progression, clinical outcomes, treatment response and disease subtyping [41–43].

For T1D, the majority of genetic risk is accounted for by the HLA class II (DR-DP loci), HLA class I and other non-HLA loci including *INS* and *PTNP22* [44]. Previous studies have demonstrated the inclusion of PRS into models with clinical parameters could improve T1D risk prediction [45–47] and aid discrimination of T1D from T2D [48–50], or from the general population [51]. Oram and colleagues [48] developed a 30-SNP PRS for T1D (denoted as *GRS-1*) that comprised HLA and non-HLA variants, which was shown to accurately distinguish between T1D and T2D (area under the curve [AUC]=0.88) and predict progression to insulin deficiency (AUC=0.87) in young adults with diabetes. The second version of Oram's T1D GRS (known as *GRS-2*) was extended to 67 HLA and non-HLA variants that comprehensively integrate DR-DQ risk, HLA interactions and additional non-HLA SNPs [49]. T1D *GRS-2* significantly improved the discrimination (AUC=0.93) of T1D from T2D and from controls. However, currently most T1D PRSs were developed and tested only in populations of European ancestry, which could restrict its generalization and transferability in non-European populations [43, 52]. Of note, non-European populations have ancestry-specific variants, different frequencies of specific alleles and haplotypes, and peculiar subtypes like fulminant T1D reported in China, Japan and India [53•].

In parallel, T2D PRSs have also been found to be of use for risk stratification in the general population, or help with prediction of disease progression in T2D patients [54]. A study by Mahajan and colleagues [55], utilizing data from UK Biobank, found that individuals in the top 2.5% of the PRS distribution were at 9.4-fold increased risk compared with the bottom 2.5%. Meanwhile, in a Chinese cohort of patients with T2D from the HKDR, a T2D PRS consisting of 123 T2D-associated SNPs predicted rapid progression to insulin therapy [56]. Further, increasing GWAS from non-European populations [57–59], large-scale trans-ancestry meta GWAS [54, 60] and novel cross-population/trans-ancestry PRS methods [61] has facilitated the transferability of T2D risk prediction in diverse populations. For instance, based on a Bayesian polygenic modeling method, PRS-CSx [62], Ge and colleagues [61] constructed a trans-ancestry T2D PRS by integrating T2D GWAS from European, African and East Asian populations and demonstrated that the top 2% of the trans-ancestry PRS distribution can identify individuals of European, African, Hispanic/Latino and East Asian ancestry with an approximately 2.5–4.5-fold of increase in T2D risk, which corresponds to a similar increased risk of T2D for first degree relatives of someone with T2D.

T2D is increasingly recognized as a collage of varied pathophysiological states with heterogeneous etiology [63]. In a similar vein, T2D PRS, as a unitary summary score, may not fully capture the complexity of the

heritable risk. According to the “palette” model [64••], a series of underlying pathophysiological processes contribute towards the development of T2D. Therefore, increasing attention has been drawn to partition PRS by genetic clustering or decomposition, in order to achieve more personalized risk prediction. Udler and colleagues have decomposed the genetic associations across diabetes-related biomarkers and identified five T2D genetic clusters using a “soft” clustering method [65], which was further extended into ten clusters in their recent work [66]. These clusters/partitioned PRSs showed distinct patterns of associations with metabolic traits and clinical outcomes related to T2D [65–67], suggesting the great potential of process-specific PRS to provide mechanistic insights into the heterogeneity of T2D and advance precision medicine in diabetes. Besides, fine-mapping, functional annotation and partition of T2D loci by integrating the transcriptome and epigenome information could help develop mechanism-driven tissue-specific PRS. Miguel-Escalada and colleague [68] constructed an islet-specific T2D PRS using islet enhancer hub variants involved in islet gene regulation and insulin secretion and showed its potential in delineating T2D risk profiles from the novel perspective of tissue-specific epigenomic mechanisms.

However, one should exercise caution when interpreting the results of risk stratification based on PRS because PRS alone may not perform so well on its own, but can be improved when PRS are considered in combination with existing risk factors [69]. Therefore, the clinical utility of PRS for diabetes should be assessed combining other clinical risk factors, such as BMI (body mass index), GADA (glutamic acid decarboxylase antibodies), and fasting glucose and insulin. Besides, diabetes is the product of the combination and interaction of genetic (G) and environmental (E) factors. It still remains unclear whether and how GxE interactions influence the effect of PRS. In other words, the effect of the T2D PRS would depend on the level of the environmental risk factors like lifestyle, socioeconomic status and environment pollutants. Merino and colleague [70] explored if T2D PRS and diet quality have a synergistic effect on the risk of T2D and found only independent associations between the two factors and incident T2D. Therefore, evaluating the interactions between T2D PRS and environmental factors in diverse populations could help improve the transferability of PRS and unravel the clinical heterogeneity of diabetes.

Whilst many epidemiological cohorts have contributed to the identification of genetic factors for diabetes, analyses increasingly require collaborative analyses across many cohorts involving participants from different ethnic populations taking into consideration the clinical contexts, e.g. ecological anthropology and

healthcare environment. Nevertheless, genetic studies have highlighted the potential of epidemiological cohorts to contribute towards biological understanding of common diseases. Discoveries from WGS/WES and GWAS will have different utilities and are complementary. PRS from GWAS can be used for prognostication and patient segmentation, and can be used as companion diagnostics whilst WGS/WES will provide drug targets for modifying disease pathways. Applying these 2 technologies to prospective cohorts with detailed phenotyping along with interventions provide an important strategy for precision medicine.

Epigenomics, miRNA and Other Epigenetic Biomarkers

Along with genetics, the regulation of gene expression can be achieved through epigenetic modifications without altering the DNA sequence [71]. There are multiple mechanisms involved in the epigenome including non-coding RNA regulation, DNA methylation and histone modifications. These mechanisms are critical in controlling gene expression, cell development and differentiation. As environmental exposures and external factors can modify the epigenome, dysregulation of the epigenome is thought to provide a picture integrating environmental as well as germ-line genetic variation which may contribute to the development of different chronic diseases, including diabetes and related co-morbidities [72, 73]. Furthermore, epigenetic changes have been implicated to mediate glycaemic or metabolic memory, whereby previous exposure to hyperglycaemia may lead to sustained effects on gene expression and increased risk of diabetes complications [74•, 75].

MicroRNAs (miRNAs) are a class of non-coding and single-stranded RNAs with lengths of 18–25 nucleotides [76, 77]. MiRNAs play an important role in the post-transcriptional regulation of gene expression by epigenetic modulation [78]. Capable of RNA interference, miRNAs can bind with target messenger RNAs (mRNAs) to inhibit translation or induce mRNA degradation [79]. Since miRNAs are involved in an extensive range of biological processes including cell development, differentiation, proliferation and apoptosis, they have been proposed as potential disease biomarkers [76, 80•]. With high stability and resistance to degradation in human biofluids, circulating miRNAs have emerged as promising biomarkers for diabetes [76] and diabetic complications [80•, 81].

In a recent sequencing-based analysis of islet miRNAs, human pancreatic islet samples were used for RNA sequencing, small-RNA sequencing and genotyping. Eighty-four miRNAs were found to be highly heritable, and

mainly regulated by trans-acting genetic effects. In addition, the study identified 14 T2D-related miRNAs including miR-21-5p and miR-187-3p [82]. High-throughput quantitative real-time PCR (qPCR) is another commonly used approach in miRNA profiling. OpenArray and Dynamic Array are high-throughput qPCR platforms that require small reaction volumes (nanoliters) for each reaction. They have been applied to validate a miRNA signature generated by small-RNA sequencing in diabetic retinopathy [83].

From the HKDR cohort, screening of miRNA markers for liver cancer in T2D patients was conducted using serum-extracted RNA. Applying microarray for discovery and qPCR validation, miR-122-5p and miR-455-3p were identified to be potential biomarkers to predict the development of liver cancer [84]. Circulating miRNAs could also be used to predict the development of T2D. Gestational diabetes mellitus (GDM) is a common pregnancy complication and is associated with sevenfold increased risk of T2D in later life [85]. Using postpartum plasma, an observational study explored the potential of circulating miRNAs to predict the future development of T2D. OpenArray and qPCR were applied for discovery and validation, with miR-369-3p identified to be significant after validation and multiple comparisons. It is suggested that the measurement of this miRNA could improve the subsequent prediction of T2D in women with GDM [86].

DNA methylation is another important epigenetic modification, occurring mainly in the CpG islands [87]. DNA methyltransferases (DNMTs) mediate DNA methylation by covalently adding a methyl group to the 5' position of the cytosine residue, leading to transcriptional silencing [74•]. In an early epigenome-wide association study (EWAS) using methylation arrays in the London Life Sciences Prospective Population (LOLIPOP) study, 5 methylation markers identified from methylation profiling of peripheral blood were found to be associated with incident T2D, including a site in the *TXNIP* gene, a methylation locus which has consistently been replicated since [88]. Interestingly, a methylation score constructed from the top 5 loci was strongly associated with incident diabetes across cohorts, independent of established risk factors [88]. Findings from such EWAS efforts complement findings from epigenomic profiling of diabetes-related tissues such as adipose tissue, which found differential methylation in some novel loci, in addition to established T2D-related genes, such as *PPARG*, *KCNQ1* and *TCF7L2* [89].

The Diabetes Control and Complications Trial (DCCT) and the long-term follow-up in the Epidemiology of Diabetes Interventions and Complications (EDIC) Study were landmark studies which highlighted the persistent impact of a period of suboptimal glucose control, subsequently named “metabolic memory”, on the progression of microvascular outcomes in people with T1D [74•, 90]. Profiling of DNA methylation in leukocyte and monocyte DNA

from participants who experienced metabolic memory and microvascular complications, compared to participants in the intensive control arm who were free of diabetes complications, identified differential methylation at a number of key loci, including hypomethylation at cg19693031 in the 3'-untranslated region (3'-UTR) of *TXNIP* [91]. This is particularly interesting given the established role of *TXNIP* in hyperglycaemia [88]. A study conducted in Native Americans with T2D identified methylation loci associated with baseline renal function and subsequent decline in renal function [92]. A study in the HKDR, with methylation profiling of 1271 patients with T2D, identified 40 CpG sites significantly associated with baseline eGFR, and 8 CpG sites associated with decline in eGFR. A prediction model was developed to estimate eGFR slope using methylation data, which was replicated in the Native American population. The model was also useful for improving prediction of end-stage renal failure among people with diabetes [93]. Interesting, the CpG sites were near genes enriched for functional roles in kidney disease, and several of the CpG sites identified showed association with renal fibrosis [93].

Histone modifications refer to the post-translational covalent addition of functional groups to histone proteins. These include histone H3 lysine 4 trimethylation ($H3K4me^3$) and H3/H4 lysine acetylation (K_{ac}) associated with active gene expression, as well as $H3K9me^{2/3}$, $H3K27me^3$ and/or DNA methylation that are usually associated with repressed gene expression [74•]. In participants from DCCT/EDIC who experienced metabolic memory and went on to develop microvascular complications had enrichment of the active chromatin mark $H3K9ac$, which was associated with the mean HbA1c during follow-up. Furthermore, the hyperacetylated promoters were enriched for genes involved in inflammatory pathways, highlighting the potential role of epigenetics in metabolic memory and diabetes complications [94].

Leukocyte telomere shortening is a biomarker of biological aging, and may represent another type of epigenetic marker in diabetes [95]. In the HKDR, relative leukocyte telomere length (rLTL) was found to be inversely associated with the risk of incident diabetic cardiovascular and renal complications [96, 97].

Summary of some applications of epigenetic biomarkers from the above studies are shown in Table 2.

These studies demonstrate the potential of epigenetic markers to improve diagnosis and outcome prediction for precision medicine in diabetes [2, 98]. Another area where epigenetic biomarkers are of particular interest is in the developmental origins of diabetes and the potential impact of maternal hyperglycaemia and nutrition on epigenetic changes [99]. A detailed discussion of this is beyond the scope of this review, and readers who are interested in this area are referred to review articles which provide more details [99–101].

Table 2 Examples of epigenetic biomarker applications in diabetes

Epigenetic biomarker	Study subjects	Sample types	Detection Methods	Findings	References
miRNA	Human individual	Pancreatic islet	RNA-sequencing, small-RNA sequencing and genotyping	14 T2D-related miRNAs	[82]
miRNA	Individuals with or without diabetes	Serum	Next-generation sequencing and qPCR technologies (ViiA 7, OpenArray, Dynamic Array)	High-throughput qPCR can be applied for quantitative assessment of miRNA signature	[83]
miRNA	T2D patients from HKDR cohort	Serum	Microarray and qPCR	miR-122-5p and miR-455-3p are potential biomarkers for development of liver cancer in T2D	[84]
miRNA	Women with GDM	Postpartum plasma	OpenArray and qPCR	miR-369-3p is a potential biomarker to predict T2D in women with GDM	[86]
DNA methylation	Incident T2D patients in LOLIPOP study	Peripheral blood	Methylation arrays, next-generation sequencing and pyrosequencing	5 methylation markers including a site in TXNIP gene are associated with incident T2D	[88]
DNA methylation	Individuals with or without diabetes complications	Genomic DNA of whole blood and blood monocytes	Methylation arrays and pyrosequencing	Differential methylation at a number of loci	[91]
DNA methylation	Individuals with T2D and with chronic kidney disease	Peripheral blood	Methylation arrays	Methylation loci associated with baseline and decline in renal function	[92]
DNA methylation	T2D patients from HKDR cohort	Genomic DNA of whole blood	Methylation arrays	CpG sites associated with baseline eGFR and eGFR slope respectively	[93]
Histone modification	DCCT/EDIC participants	Peripheral blood	Chromatin immunoprecipitation (ChIP)—qPCR	Association between HbA1c and H3K9ac, and potential role of epigenetics in metabolic memory	[94]
Relative leukocyte telomere length (rLTL)	T2D patients from HKDR cohort	DNA from whole blood	qPCR for rLTL measurements	rLTL is inversely associated with incident diabetic cardiovascular and renal complications	[96, 97]

Applications of Proteomics in Diabetes

Proteomics, an analytical discipline dedicated to exploring the dynamic fluctuations in protein composition, expression and post-translational modifications, has been instrumental in elucidating the pathophysiological mechanisms underlying metabolic disorders such as diabetes mellitus [102]. Affinity-based assays, which commonly employ either monoclonal or polyclonal antibodies tethered to a reporter molecule via luminescence, fluorescence, radioactivity or enzymatic activity in enzyme-linked immunosorbent assays (ELISA), remain the most commonly used approach for protein identification and quantification [103]. Nevertheless, these assays are inherently constrained by their selectivity for a single analyte of interest, and therefore cannot provide an unbiased measure of all proteins in the sample. Moreover, the low abundance of most serum proteins poses significant challenges to their detectability via traditional assays [104]. Mass spectrometry (MS) has emerged as a robust alternative, effectively employed in the identification and quantification of proteins such as galectin-1 and apolipoprotein A-1, both of which exhibit altered expression profiles in diabetic individuals [105]. MS methodologies can be executed in a targeted or non-targeted modality, affording either high specificity in protein identification or simultaneous quantification of multiple analytes, even those present in low concentrations. Nonetheless, MS entails a laborious and temporally extensive workflow, necessitating the depletion of high-abundance plasma proteins, mechanical protein separation, trypsin digestion and subsequent verification via immunoassays or other confirmatory protocols [106, 107].

Recent advancements in proteomic profiling technologies have appreciably augmented the efficacy and scope of detectable circulating proteins. Affinity-based techniques that incorporate antibody multiplexing or innovative affinity reagents have substantially broadened the quantitative capabilities of these assays. High-throughput methodologies such as nucleic acid affinity reagents (aptamers) or nucleotide-labeled antibodies have become increasingly prevalent. The SomaLogic platform employs aptamers, capitalizing on the structural versatility of oligonucleotides to specifically bind protein epitopes, thereby facilitating protein quantification [108]. Conversely, the Olink platform employs nucleic acid-labeled antibodies, enabling the utilization of polymerase chain reaction (PCR) technology for protein amplification, detection and quantification [108]. Nonetheless, the specificity of binding remains an inherent limitation, which can be mitigated through corroborative studies employing traditional immunoassays, MS and integrative genomic analyses [109, 110].

A Swedish study utilized nucleic acid-labeled antibodies and proximity extension assay to identify seven

circulating proteins associated with the homeostatic model assessment of insulin resistance (HOMA-IR). These included the novel association of cathepsin D, as well as previously reported proteins such as leptin, renin, IL-1ra (interleukin-1 receptor antagonist), hepatocyte growth factor, FABP4 (FA-binding protein 4) and tPA (tissue-type plasminogen activator). However, the associations of IL-1ra and tPA with incident diabetes were completely attenuated after adjustments for fasting glucose [111]. Mendelian randomization analyses also suggested that insulin resistance had a causal effect on tPA levels. In a more recent, larger cross-sectional population-based study using the EpiHealth study from Sweden, 29 proteins were found to be associated with prevalent diabetes mellitus at a false discovery rate of less than 5%. Of these, 14 of the reported protein associations with T2D were novel [112]. However, Mendelian randomization analyses did not find any causal relationship between these proteins and diabetes, suggesting they may be more useful as biomarkers. Yazdanpanah et al. used proteome-wide Mendelian randomization to identify signal regulatory protein γ (SIRPG), interleukin-27 Epstein-Barr virus-induced 3 (IL27.EBI3) and chymotrypsinogen B1 (CTRB1) as potential drug targets for T1D treatment [113]. In a similar vein, the group utilized available proteomics datasets to identify C-type mannose receptor 2 (MRC2), sodium/potassium-transporting ATPase subunit β 2 (ATP1B2), spermatogenesis-associated protein 20 (SPATA20), HP, MANSC domain containing 4 (MANSC4) and α 1-3-galactosyltransferase (ABO) as causal proteins for T2D [114].

Overall, proteomics studies have not only led to identification of novel protein biomarkers, but given the pivotal role of proteins in disease pathogenesis, advanced the understanding of diabetes. Advancements in protein profiling techniques have allowed for the measurement of a greater number of circulating proteins with higher specificity. These developments have the potential to improve our understanding of diabetes and other diseases, leading to better diagnosis and potential novel treatment options in the future.

Application of Metabolomics in Diabetes

Metabolomics, an expanding field of scientific research, has its roots in early metabolite analysis and is now primarily used to identify disease biomarkers, including in diabetes [115, 116]. Diabetes is associated with metabolic disturbances of sugar, protein, fat, water and electrolytes that negatively impact organs such as the liver, skeletal muscle and adipose tissue [117]. Metabolomics offers a

comprehensive view, or “snapshot”, of the metabolic landscape, capable of tracking thousands of metabolites across cells, tissues or entire organisms [118, 119]. The technology can detect the composition of metabolites and their changing trends over time, or after specific perturbations. The combination of metabolic and biochemical information can also highlight the interaction between relevant metabolic and signaling pathways, made possible due to improved analytical techniques and data handling systems [118, 120, 121]. Various platforms for metabolomic profiling are available, such as nuclear magnetic resonance (NMR) spectroscopy, mass spectrometry (MS), Fourier transform infrared (FTIR) spectroscopy and high-performance liquid chromatography (HPLC), and these have all been utilized to advance diabetes research [122, 123].

In the human metabolome, there is a wealth of information regarding low molecular weight metabolites that originate from diet, such as nutrient intermediates, lipids, hormones and other signaling molecules [116]. In a nested case–control study of 503 baseline plasma samples from the Swedish prospective Västerbotten Intervention Programme cohort, taken at a median time of 7 years prior to the diagnosis of diabetes, untargeted liquid chromatography-MS metabolomics led to identification of 46 metabolites associated with incident diabetes, including some novel findings, such as phosphatidylcholines (PCs) containing odd-chained lecithins. Many of these metabolites exhibited temporal shifts correlated with the progression toward diabetes, and 42 of the 46 remained significant after adjustments for baseline BMI, fasting glucose and lifestyle factors [124]. Metabolomic studies have also highlighted the potential role of gut microbiota in the development of metabolic diseases. For instance, higher levels of indole propionic acid, produced by intestinal microbes, correlated with improved insulin sensitivity and a lowered risk of T2D onset in the Finnish Diabetes Prevention Study [125]. Large-scale NMR metabolomic profiling has also been conducted in the large China Kadoorie Biobank. The study identified 163 metabolites related to the risk of developing T2D, and 147 of these remained statistically significant after controlling for baseline glucose levels. Elevated levels of specific factors, such as the ratio of apolipoprotein B to apolipoprotein A-1, triglycerides, and the branched-chain amino acids leucine and isoleucine, were all linked to an increased risk of incident T2D [126].

Lipidomics, a specialized subfield of metabolomics, employs high-throughput methodologies to elucidate changes in lipid composition and expression. A study using HPLC-multiple reaction monitoring (MRM) measured 667 serum lipids in subjects with incident diabetes and their matched controls, revealing 38 lipids significantly correlated with T2D risk. These included triacylglycerols

(TAGs), lyso-phosphatidyl inositols, phosphatidylcholines, polyunsaturated fatty acid (PUFA)–plasmalogen phosphatidylethanolamines (PUFA-PEps) and cholesteryl esters. This lipidomic profile enhanced predictive accuracy beyond traditional clinical risk factors [127].

In summary, metabolomics represents a powerful tool for identifying metabolic disturbances in diabetes and potential biomarkers for early diagnosis and targeted therapy [103]. As the technology continues to advance, it has the potential to considerably expand our understanding of the pathogenesis of diabetes and pave the way for precision medicine.

Integrating Omics Research for Novel Discoveries and Precision Medicine

Whilst the earlier sections have highlighted recent advances to use epidemiological cohorts to identify different biomarkers for diabetes and related complications, it is the integrative analysis of multiple omics that are likely to provide the most useful novel insights towards diabetes and related complications. In addition to the integration of different layers of omics datasets from different cohorts, there should be considerable advantage in leveraging the measurement of multiple omics in the same individuals (Fig. 1). The UK Biobank, in which the large-scale collection of detailed information, prospective follow-up, whole genome genotyping and exome sequencing, proteomics and metabolomics, and the availability of access to the data for *bona fide* researchers, has been transformative in its global impact on biomedical research [128, 129]. Earlier multi-omics studies have utilized the genetic data to integrate the proteomics data for protein quantitative trait loci (pQTL) analyses to identify genetic variants that regulate protein expression and provide genetic instruments to explore causality between protein biomarkers and different diseases using the MR framework [112], and likewise for methylation quantitative trait loci (meQTL) analyses for methylation markers [88]. In addition, multi-omics data has been generated in an increasing number of cohorts, including the INTERVAL study [110, 130], the Fenland cohort [131], China Kadoorie Biobank [126], FinnDiane [132] and other cohorts. Similar work is ongoing in the Hong Kong Diabetes Register, Hong Kong Diabetes Biobank and the TRansomics ANalysis of Complications and ENdpoints in Diabetes (TRANSCEND) Consortium [93, 133–135]. This generation of deep phenotyping with multi-omics data in large epidemiological cohorts, preferably with prospective follow-up, represents another important dimension in “big data” analytics. The integration of whole exome sequencing data with metabolomics data has facilitated the identification of novel

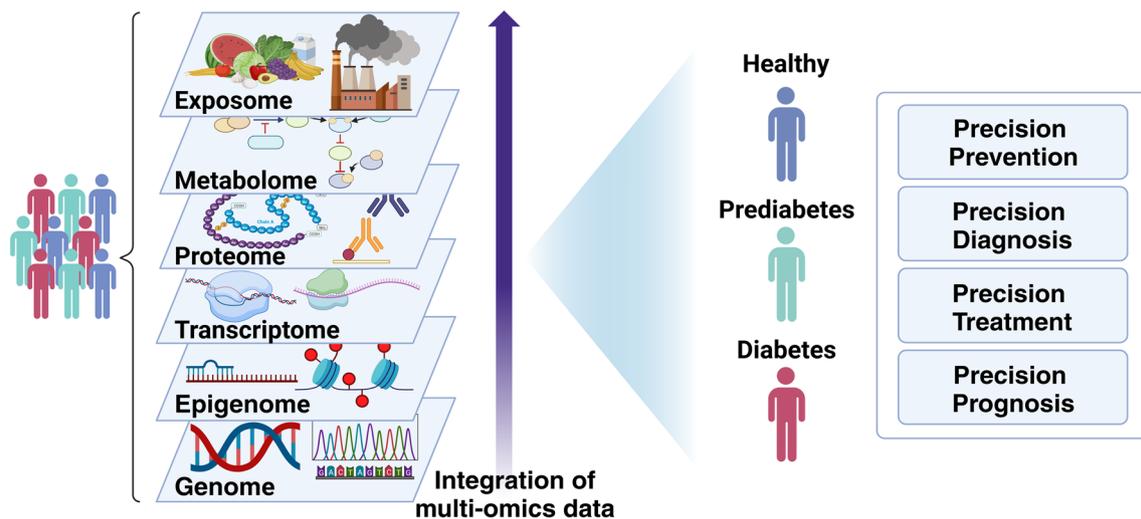


Fig. 1 The integration of multi-omics analyses in individuals from cohorts can help to drive the development of precision medicine in diabetes. Legend: each layer represents increasing complexity which have arisen from the genome, epigenome, transcriptome, proteome, metabolome and exposome. These information, representing deep

phenotyping of individuals, may provide information that can help inform and guide disease classification and treatment selection, as well as predict future risk of complications. Precision Medicine in Diabetes reflects the overall efforts to utilize these as well as clinical information to guide treatment selection and clinical decisions

associations implicating rare coding variants, and can advance future drug development [136]. Based on first principle, genome is the most upstream causal factor (the concept underlying MR) with epigenomes, proteome and metabolome, being regulators and mediators expressed as phenomes further modified by age, sex, exposome, demographic microbiome and intervention including pharmacological and non-pharmacological. Whilst the discovery of biomarkers can be epidemiology-focused, unravelling the meaning and applications is very much a clinical science, further complicated by the importance of health beliefs and behaviour for a chronic disease such as diabetes. There are also numerous successful examples where the incorporation of omics research into clinical trials have helped to identify novel biomarkers for clinical outcomes as well as biomarkers related to treatment response [137–141], though there are different limitations including the sample sizes of clinical trials being powered to detect differences in the primary outcomes in relation to the interventions being examined, and relatively short duration of the intervention. Given the complexity of the subject and the inter-disciplinary nature of this kind of work, there are significant advantages for such work to be led by physician researchers, epidemiologists or researchers with knowledge in epidemiology, human biology and medical treatment, in close collaboration with and supported by allied health professionals (e.g. diet, exercise, behavioural), molecular biologists, geneticists, data scientists and members with complementary expertise (e.g.

chemistry and drug development) to tackle this complex subject and advance the field.

In this review, we have provided some representative examples of omic studies as applied to the epidemiological investigation of diabetes and related outcomes. It is only possible to provide a snapshot of the most commonly employed technologies, and there are numerous other omics platforms that we did not have space to discuss in more detail, in particular radiomics or imaging data, as well as gut microbiome and exposome [142, 143]. For investigators planning to embark on epidemiological analyses with biobanking for future omics analyses, it is important to be forward-looking, and not be limited by currently available technologies. Other challenges investigators are likely to encounter in this area of work include the capacity and costs for data storage, capabilities for data linkage, methodologies for data analysis, integration of data across the different omics platforms and clinical phenotypes, and the need for multidisciplinary collaboration given the different domains of expertise required. Whilst the different omics are tools with the potential utility to elucidate disease mechanisms, classify disease subtypes, stratify risk and inform targeted treatment, when analyzed in an integrated manner, may provide new insights for predicting, preventing, classifying and personalizing care in diabetes. Advances in the development of methods, including those involving machine learning and artificial intelligence, to integrate data from these multi-omics datasets, would also be key to future development in this area for precision medicine in diabetes [144, 145, 146••].

Author Contribution Conception of article: RCWM.

Literature review and data analysis: GY, HCHT, CH, MS, RCWM. Drafting of manuscript and critical revision: GY, HCHT, CH, MS, CKPL, JCNC, RCWM.

All authors contributed meaningfully to this manuscript and approved the final version. RCWM is the guarantor of this work.

Funding JCNC acknowledges support from the Hong Kong Government Health and Medical Research Fund (CFS-CUHK2), Hong Kong Genome Institute, Chinese University of Hong Kong Research Committee Postdoctoral Fellowship Scheme and Innovation and Technology Commission Research Talent Hub. RCWM acknowledges support from the Research Grants Council of the Hong Kong Special Administrative Region (CU R4012-18), the Innovation and Technology Fund Mid-stream Research Programme for Universities (MRP/054/20X), Innovation and Technology Commission Research Talent Hub, the University Grants Committee Research Grants Matching Scheme, the Research Committee Postdoctoral Fellowship Scheme and the Provost's Scheme for PhD scholarship of the Chinese University of Hong Kong.

Data Availability Not applicable.

Declarations

Ethical Approval Not applicable for this review article.

Competing Interests JCNC, CKPL, RCWM are co-founders of GemV-Care, a technology start-up initiated with support from the Hong Kong Government Innovation and Technology Commission and its Technology Start-up Support Scheme for Universities (TSSSU).

Human and Animal Rights and Informed Consent All reported studies/experiments with human or animal subjects performed by the authors were performed in accordance with all applicable ethical standards including the Helsinki declaration and its amendments, institutional/national research committee standards, and international/national/institutional guidelines.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

Papers of particular interest, published recently, have been highlighted as:

- Of importance
- Of major importance

1. Magliano DJ, Boyko EJ, IDF Diabetes Atlas Committee. IDF Diabetes Atlas. IDF diabetes atlas. Brussels: International Diabetes Federation © International Diabetes Federation, 2021.

2. Chung WK, Erion K, Florez JC, Hattersley AT, Hivert MF, Lee CG, et al. Precision medicine in diabetes: a consensus report from the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care*. 2020;43(7):1617–35.
3. Kirwan JA, Brennan L, Broadhurst D, Fiehn O, Cascante M, Dunn WB, et al. Preanalytical Processing and biobanking procedures of biological samples for metabolomics research: a white paper, Community Perspective (for “Precision Medicine and Pharmacometabolomics Task Group”-The Metabolomics Society Initiative). *Clin Chem*. 2018;64(8):1158–82.
4. Peakman TC, Elliott P. The UK Biobank sample handling and storage validation studies. *Int J Epidemiol*. 2008;37(Suppl 1):i2–6.
5. Altobelli E, Chiarelli F, Valenti M, Verrotti A, Blasetti A, Di Orio F. Family history and risk of insulin-dependent diabetes mellitus: a population-based case-control study. *Acta Diabetol*. 1998;35(1):57–60.
6. Seaquist ER, Goetz FC, Rich S, Barbosa J. Familial clustering of diabetic kidney disease. Evidence for genetic susceptibility to diabetic nephropathy. *N Engl J Med*. 1989;320(18):1161–5.
7. Earle K, Walker J, Hill C, Viberti G. Familial clustering of cardiovascular disease in patients with insulin-dependent diabetes and nephropathy. *N Engl J Med*. 1992;326(10):673–7.
8. Tuomilehto J, Borch-Johnsen K, Molarius A, Forsen T, Rastenyte D, Sarti C, et al. Incidence of cardiovascular disease in Type 1 (insulin-dependent) diabetic subjects with and without diabetic nephropathy in Finland. *Diabetologia*. 1998;41(7):784–90.
9. Cole JB, Florez JC. Genetics of diabetes mellitus and diabetes complications. *Nat Rev Nephrol*. 2020;16(7):377–90.
10. Hara K, Shojima N, Hosoe J, Kadowaki T. Genetic architecture of type 2 diabetes. *Biochem Biophys Res Commun*. 2014;452(2):213–20.
11. Chan JC, So W, Ma RC, Tong PC, Wong R, Yang X. The complexity of vascular and non-vascular complications of diabetes: the Hong Kong Diabetes Registry. *Curr Cardiovasc Risk Rep*. 2011;5(3):230–9.
12. Chan JCN, Lim LL, Luk AOY, Ozaki R, Kong APS, Ma RCW, et al. From Hong Kong Diabetes Register to JADE Program to RAMP-DM for Data-Driven Actions. *Diabetes Care*. 2019;42(11):2022–31.
13. Li JK, Ng MC, So WY, Chiu CK, Ozaki R, Tong PC, et al. Phenotypic and genetic clustering of diabetes and metabolic syndrome in Chinese families with type 2 diabetes mellitus. *Diabetes Metab Res Rev*. 2006;22(1):46–52.
14. Zhang Y, Luk AOY, Chow E, Ko GTC, Chan MHM, Ng M, et al. High risk of conversion to diabetes in first-degree relatives of individuals with young-onset type 2 diabetes: a 12-year follow-up analysis. *Diabet Med*. 2017;34(12):1701–9.
15. Ng MC, So WY, Cox NJ, Lam VK, Cockram CS, Critchley JA, et al. Genome-wide scan for type 2 diabetes loci in Hong Kong Chinese and confirmation of a susceptibility locus on chromosome 1q21-q25. *Diabetes*. 2004;53(6):1609–13.
16. Ng MC, Lam VK, Tam CH, Chan AW, So WY, Ma RC, et al. Association of the POU class 2 homeobox 1 gene (POU2F1) with susceptibility to Type 2 diabetes in Chinese populations. *Diabet Med*. 2010;27(12):1443–9.
17. Ma RC, Hu C, Tam CH, Zhang R, Kwan P, Leung TF, et al. Genome-wide association study in a Chinese population identifies a susceptibility locus for type 2 diabetes at 7q32 near PAX4. *Diabetologia*. 2013;56(6):1291–305.
18. Cho YS, Chen CH, Hu C, Long J, Ong RT, Sim X, et al. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat Genet*. 2011;44(1):67–72.

19. Ng MC, Park KS, Oh B, Tam CH, Cho YM, Shin HD, et al. Implication of genetic variants near TCF7L2, SLC30A8, HHEX, CDKAL1, CDKN2A/B, IGF2BP2, and FTO in type 2 diabetes and obesity in 6,719 Asians. *Diabetes*. 2008;57(8):2226–33.
20. Noble JA, Erlich HA. Genetics of type 1 diabetes. *Cold Spring Harb Perspect Med*. 2012;2(1):a007732.
21. Robertson CC, Rich SS. Genetics of type 1 diabetes. *Curr Opin Genet Dev*. 2018;50:7–16.
22. Quattrin T, Mastrandrea LD, Walker LSK. Type 1 diabetes. *Lancet*. 2023;401(10394):2149–62.
23. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet*. 2009;41(6):703–7.
24. Onengut-Gumuscu S, Chen WM, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, et al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat Genet*. 2015;47(4):381–6.
25. Robertson CC, Inshaw JRJ, Onengut-Gumuscu S, Chen WM, Santa Cruz DF, Yang H, et al. Fine-mapping, trans-ancestral and genomic analyses identify causal variants, cells, genes and drug targets for type 1 diabetes. *Nat Genet*. 2021;53(7):962–71.
26. Ge Y, Onengut-Gumuscu S, Quinlan AR, Mackey AJ, Wright JA, Buckner JH, et al. Targeted deep sequencing in multiple-affected sibships of European Ancestry identifies rare deleterious variants in PTPN22 that confer risk for type 1 diabetes. *Diabetes*. 2016;65(3):794–802.
27. Fuchsberger C, Flannick J, Teslovich TM, Mahajan A, Agarwala V, Gaulton KJ, et al. The genetic architecture of type 2 diabetes. *Nature*. 2016;536(7614):41–7.
28. Ingelsson E, McCarthy MI. Human genetics of obesity and type 2 diabetes mellitus: past, present, and future. *Circ Genom Precis Med*. 2018;11(6):e002090.
29. Prasad RB, Groop L. Genetics of type 2 diabetes-pitfalls and possibilities. *Genes (Basel)*. 2015;6(1):87–123.
30. Suzuki K, Hatzikotoulas K, Southam L, Taylor HJ, Yin X, Lorenz KM, et al. Multi-ancestry genome-wide study in >2.5 million individuals reveals heterogeneity in mechanistic pathways of type 2 diabetes and complications. *medRxiv*. 2023.
31. Wainschein P, Jain D, Zheng Z, Group TOAW Consortium NT-OfPM Cupples LA, et al. Assessing the contribution of rare variants to complex trait heritability from whole-genome sequence data. *Nat Genet*. 2022;54(3):263–73.
32. Mahajan A, Wessel J, Willems SM, Zhao W, Robertson NR, Chu AY, et al. Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes. *Nat Genet*. 2018;50(4):559–71.
33. Flannick J, Mercader JM, Fuchsberger C, Udler MS, Mahajan A, Wessel J, et al. Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature*. 2019;570(7759):71–6.
34. DiCorpo D, Gaynor SM, Russell EM, Westerman KE, Raffield LM, Majarian TD, et al. Whole genome sequence association analysis of fasting glucose and fasting insulin levels in diverse cohorts from the NHLBI TOPMed program. *Commun Biol*. 2022;5(1):756.
35. Rothman KJ, Greenland S. Causation and causal inference in epidemiology. *Am J Public Health*. 2005;95(Suppl 1):S144–50.
36. Hill AB. The environment and disease: association or causation? *Proc R Soc Med*. 1965;58(5):295–300.
37. van Zuydam NR, Ahlqvist E, Sandholm N, Deshmukh H, Rayner NW, Abdalla M, et al. A genome-wide association study of diabetic kidney disease in subjects with type 2 diabetes. *Diabetes*. 2018;67(7):1414–27.
38. Salem RM, Todd JN, Sandholm N, Cole JB, Chen WM, Andrews D, et al. Genome-wide association study of diabetic kidney disease highlights biology involved in glomerular basement membrane collagen. *J Am Soc Nephrol*. 2019;30(10):2000–16.
39. Ma RC. Genetics of cardiovascular and renal complications in diabetes. *J Diabetes Investig*. 2016;7(2):139–54.
40. Sandholm N, Dahlstrom EH, Groop PH. Genetic and epigenetic background of diabetic kidney disease. *Front Endocrinol (Lausanne)*. 2023;14:1163001.
41. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med*. 2020;12(1):44.
42. Wang Y, Tsuo K, Kanai M, Neale BM, Martin AR. Challenges and opportunities for developing more generalizable polygenic risk scores. *Annu Rev Biomed Data Sci*. 2022;5:293–320.
43. Kullo IJ, Lewis CM, Inouye M, Martin AR, Ripatti S, Chatterjee N. Polygenic scores in biomedical research. *Nat Rev Genet*. 2022;23(9):524–32.
44. Pociot F, Lernmark A. Genetic risk factors for type 1 diabetes. *Lancet*. 2016;387(10035):2331–9.
45. Winkler C, Krumsiek J, Buettner F, Angermuller C, Giannopoulou EZ, Theis FJ, et al. Feature ranking of type 1 diabetes susceptibility genes improves prediction of type 1 diabetes. *Diabetologia*. 2014;57(12):2521–9.
46. Redondo MJ, Geyer S, Steck AK, Sharp S, Wentworth JM, Weedon MN, et al. A Type 1 Diabetes genetic risk score predicts progression of islet autoimmunity and development of type 1 diabetes in individuals at risk. *Diabetes Care*. 2018;41(9):1887–94.
47. Ferrat LA, Vehik K, Sharp SA, Lernmark A, Rewers MJ, She JX, et al. A combined risk score enhances prediction of type 1 diabetes among susceptible children. *Nat Med*. 2020;26(8):1247–55.
48. Oram RA, Patel K, Hill A, Shields B, McDonald TJ, Jones A, et al. A Type 1 diabetes genetic risk score can aid discrimination between type 1 and type 2 diabetes in young adults. *Diabetes Care*. 2016;39(3):337–44.
49. Lynam A, McDonald T, Hill A, Dennis J, Oram R, Pearson E, et al. Development and validation of multivariable clinical diagnostic models to identify type 1 diabetes requiring rapid insulin therapy in adults aged 18–50 years. *BMJ Open*. 2019;9(9):e031586.
50. Carr ALJ, Perry DJ, Lynam AL, Chamala S, Flaxman CS, Sharp SA, et al. Histological validation of a type 1 diabetes clinical diagnostic model for classification of diabetes. *Diabet Med*. 2020;37(12):2160–8.
51. Qu HQ, Qu J, Glessner J, Liu Y, Mentch F, Chang X, et al. Improved genetic risk scoring algorithm for type 1 diabetes prediction. *Pediatr Diabetes*. 2022;23(3):320–3.
52. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51(4):584–91.
53. Redondo MJ, Gignoux CR, Dabelea D, Hagopian WA, Onengut-Gumuscu S, Oram RA, et al. Type 1 diabetes in diverse ancestries and the use of genetic risk scores. *Lancet Diabetes Endocrinol*. 2022;10(8):597–608. **This review article provides a state-of-the-art review on the use of genetic risk scores in risk stratification of type 1 diabetes**
54. Vujkovic M, Keaton JM, Lynch JA, Miller DR, Zhou J, Tcheandjieu C, et al. Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat Genet*. 2020;52(7):680–91.
55. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet*. 2018;50(11):1505–13.
56. Jiang G, Luk AO, Tam CHT, Lau ES, Ozaki R, Chow EYK, et al. Obesity, clinical, and genetic predictors for glycemic progression in Chinese patients with type 2 diabetes: a cohort study using the Hong Kong Diabetes Register and Hong Kong Diabetes Biobank. *PLoS Med*. 2020;17(7):e1003209.

57. Spracklen CN, Horikoshi M, Kim YJ, Lin K, Bragg F, Moon S, et al. Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature*. 2020;582(7811):240–5.
58. Chen J, Sun M, Adeyemo A, Pirie F, Carstensen T, Pomilla C, et al. Genome-wide association study of type 2 diabetes in Africa. *Diabetologia*. 2019;62(7):1204–11.
59. Loh M, Zhang W, Ng HK, Schmid K, Lamri A, Tong L, et al. Identification of genetic effects underlying type 2 diabetes in South Asian and European populations. *Commun Biol*. 2022;5(1):329.
60. Mahajan A, Spracklen CN, Zhang W, Ng MCY, Petty LE, Kitajima H, et al. Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet*. 2022;54(5):560–72.
61. Ge T, Irvin MR, Patki A, Srinivasainagendra V, Lin YF, Tiwari HK, et al. Development and validation of a trans-ancestry polygenic risk score for type 2 diabetes in diverse populations. *Genome Med*. 2022;14(1):70.
62. Ruan Y, Lin YF, Feng YA, Chen CY, Lam M, Guo Z, et al. Improving polygenic prediction in ancestrally diverse populations. *Nat Genet*. 2022;54(5):573–80.
63. Leslie RD, Ma RC, Franks PW, Nadeau KN, Pearson ER, Redondo MJ. Understanding diabetes heterogeneity: key steps towards precision medicine in diabetes *Lancet Diabetes Endocrinology*. 2023;S2213-8587(23):00159–6. [https://doi.org/10.1016/S2213-8587\(23\)00159-6](https://doi.org/10.1016/S2213-8587(23)00159-6).
64. ●● McCarthy MI. Painting a new picture of personalised medicine for diabetes. *Diabetologia*. 2017;60(5):793–9. **This important review article by one of the leaders in the field of genetics of diabetes provides an important conceptual advance in the understanding of diabetes heterogeneity and underlying genetic risk, and its clinical implications for personalized medicine in diabetes**
65. Udler MS, Kim J, von Grothuss M, Bonas-Guarch S, Cole JB, Chiou J, et al. Type 2 diabetes genetic loci informed by multi-trait associations point to disease mechanisms and subtypes: A soft clustering analysis. *PLoS Med*. 2018;15(9):e1002654.
66. Kim H, Westerman KE, Smith K, Chiou J, Cole JB, Majarian T, et al. High-throughput genetic clustering of type 2 diabetes loci reveals heterogeneous mechanistic pathways of metabolic disease. *Diabetologia*. 2023;66(3):495–507.
67. DiCorpo D, LeClair J, Cole JB, Sarnowski C, Ahmadizar F, Bielak LF, et al. Type 2 diabetes partitioned polygenic scores associate with disease outcomes in 454,193 individuals across 13 cohorts. *Diabetes Care*. 2022;45(3):674–83.
68. Miguel-Escalada I, Bonàs-Guarch S, Cebola I, Ponsa-Cobas J, Mendieta-Esteban J, Atla G, et al. Human pancreatic islet three-dimensional chromatin architecture provides insights into the genetics of type 2 diabetes. *Nat Genet*. 2019;51(7):1137–48.
69. Lewis CM, Vassos E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med*. 2020;12(1):1–11.
70. Merino J, Guasch-Ferre M, Li J, Chung W, Hu Y, Ma B, et al. Polygenic scores, diet quality, and type 2 diabetes risk: an observational study among 35,759 adults from 3 US cohorts. *PLoS Med*. 2022;19(4):e1003972.
71. Rosen ED, Kaestner KH, Natarajan R, Patti ME, Sallari R, Sander M, et al. Epigenetics and epigenomics: implications for diabetes and obesity. *Diabetes*. 2018;67(10):1923–31.
72. Relton CL, Davey SG. Is epidemiology ready for epigenetics? *Int J Epidemiol*. 2012;41(1):5–9.
73. Ling C, Ronn T. Epigenetics in human obesity and type 2 diabetes. *Cell Metab*. 2019;29(5):1028–44.
74. ● Natarajan R. Epigenetic mechanisms in diabetic vascular complications and metabolic memory: the 2020 Edwin Bierman Award Lecture. *Diabetes*. 2021;70(2):328–37. **This review article provides an update on different epigenetic mechanisms for diabetic complications, in particular drawing on studies based on epidemiological cohorts in type 1 diabetes**
75. Rodriguez H, El-Osta A. Epigenetic contribution to the development and progression of vascular diabetic complications. *Antioxid Redox Signal*. 2018;29(11):1074–91.
76. Pordzik J, Jakubik D, Jarosz-Popek J, Wicik Z, Eyileten C, De Rosa S, et al. Significance of circulating microRNAs in diabetes mellitus type 2 and platelet reactivity: bioinformatic analysis and review. *Cardiovasc Diabetol*. 2019;18(1):113.
77. Feng J, Xing W, Xie L. Regulatory roles of microRNAs in diabetes. *Int J Mol Sci*. 2016;17(10).
78. Yao Q, Chen Y, Zhou X. The roles of microRNAs in epigenetic regulation. *Curr Opin Chem Biol*. 2019;51:11–7.
79. Hashimoto N, Tanaka T. Role of miRNAs in the pathogenesis and susceptibility of diabetes mellitus. *J Hum Genet*. 2017;62(2):141–50.
80. ● Fan B, Luk AOY, Chan JCN, Ma RCW. MicroRNA and diabetic complications: a clinical perspective. *Antioxid Redox Signal*. 2018;29(11):1041–63. **This article provides a comprehensive overview of microRNAs in diabetes and diabetic complications based on epidemiological studies and cohorts.**
81. Angelescu MA, Andronic O, Dima SO, Popescu I, Meivar-Levy I, Ferber S, et al. miRNAs as biomarkers in diabetes: moving towards precision medicine. *Int J Mol Sci*. 2022;23(21).
82. Taylor HJ, Hung YH, Narisu N, Erdos MR, Kanke M, Yan T, et al. Human pancreatic islet microRNAs implicated in diabetes and related traits by large-scale genetic analysis. *Proc Natl Acad Sci U S A*. 2023;120(7):e2206797120.
83. Farr RJ, Januszewski AS, Joglekar MV, Liang H, McAulley AK, Hewitt AW, et al. A comparative analysis of high-throughput platforms for validation of a circulating microRNA signature in diabetic retinopathy. *Sci Rep*. 2015;5:10375.
84. Lee HM, Wong WKK, Fan B, Lau ES, Hou Y, O CK, et al. 2021 Detection of increased serum miR-122–5p and miR-455–3p levels before the clinical diagnosis of liver cancer in people with type 2 diabetes. *Sci Rep* 11(1):23756.
85. Song C, Lyu Y, Li C, Liu P, Li J, Ma RC, et al. Long-term risk of diabetes in women at varying durations after gestational diabetes: a systematic review and meta-analysis with more than 2 million women. *Obes Rev*. 2018;19(3):421–9.
86. Joglekar MV, Wong WKM, Ema FK, Georgiou HM, Shub A, Hardikar AA, et al. Postpartum circulating microRNA enhances prediction of future type 2 diabetes in women with previous gestational diabetes. *Diabetologia*. 2021;64(7):1516–26.
87. Lu J, Huang Y, Zhang X, Xu Y, Nie S. Noncoding RNAs involved in DNA methylation and histone methylation, and acetylation in diabetic vascular complications. *Pharmacol Res*. 2021;170:105520.
88. Chambers JC, Loh M, Lehne B, Drong A, Kriebel J, Motta V, et al. Epigenome-wide association of DNA methylation markers in peripheral blood from Indian Asians and Europeans with incident type 2 diabetes: a nested case-control study. *Lancet Diabetes Endocrinol*. 2015;3(7):526–34.
89. Nilsson E, Jansson PA, Perfiljev A, Volkov P, Pedersen M, Svensson MK, et al. Altered DNA methylation and differential expression of genes influencing metabolism and inflammation in adipose tissue from subjects with type 2 diabetes. *Diabetes*. 2014;63(9):2962–76.
90. Writing Team for the Diabetes Control and Complications Trial/ Epidemiology of Diabetes Interventions and Complications Research Group. Sustained effect of intensive treatment of type 1 diabetes mellitus on development and progression of diabetic nephropathy: the Epidemiology of Diabetes Interventions and Complications (EDIC) study. *JAMA*. 2003;290(16):2159–67.
91. Chen Z, Miao F, Paterson AD, Lachin JM, Zhang L, Schones DE, et al. Epigenomic profiling reveals an association between

- persistence of DNA methylation and metabolic memory in the DCCT/EDIC type 1 diabetes cohort. *Proc Natl Acad Sci U S A*. 2016;113(21):E3002–11.
92. Qiu C, Hanson RL, Fufaa G, Kobes S, Gluck C, Huang J, et al. Cytosine methylation predicts renal function decline in American Indians. *Kidney Int*. 2018;93(6):1417–31.
 93. Li KY, Tam CHT, Liu H, Day S, Lim CKP, So WY, et al. DNA methylation markers for kidney function and progression of diabetic kidney disease. *Nat Commun*. 2023;14(1):2543.
 94. Miao F, Chen Z, Genuth S, Paterson A, Zhang L, Wu X, et al. Evaluating the role of epigenetic histone modifications in the metabolic memory of type 1 diabetes. *Diabetes*. 2014;63(5):1748–62.
 95. Cheng F, Carroll L, Joglekar MV, Januszewski AS, Wong KK, Hardikar AA, et al. Diabetes, metabolic disease, and telomere length. *Lancet Diabetes Endocrinol*. 2021;9(2):117–26.
 96. Cheng F, Luk AO, Tam CHT, Fan B, Wu H, Yang A, et al. Shortened relative leukocyte telomere length is associated with prevalent and incident cardiovascular complications in type 2 diabetes: analysis from the Hong Kong Diabetes Register. *Diabetes Care*. 2020;43(9):2257–65.
 97. Cheng F, Luk AO, Wu H, Tam CHT, Lim CKP, Fan B, et al. Relative leukocyte telomere length is associated with incident end-stage kidney disease and rapid decline of kidney function in type 2 diabetes: analysis from the Hong Kong Diabetes Register. *Diabetologia*. 2022;65(2):375–86.
 98. Farr RJ, Joglekar MV, Hardikar AA. Circulating microRNAs in diabetes progression: discovery, validation, and research translation. *Exp Suppl*. 2015;106:215–44.
 99. Gluckman PD, Hanson MA, Buklijas T, Low FM, Beedle AS. Epigenetic mechanisms that underpin metabolic and cardiovascular diseases. *Nat Rev Endocrinol*. 2009;5(7):401–8.
 100. Ma RC, Tutino GE, Lillycrop KA, Hanson MA, Tam WH. Maternal diabetes, gestational diabetes and the role of epigenetics in their long term effects on offspring. *Prog Biophys Mol Biol*. 2015;118(1–2):55–68.
 101. Block T, El-Osta A. Epigenetic programming, early life nutrition and the risk of metabolic disease. *Atherosclerosis*. 2017;266:31–40.
 102. Wang N, Zhang S, Zhu F, Yang Y, Chen L, Lu P, et al. Proteomic study on the new potential mechanism and biomarkers of diabetes. *Proteomics Clin Appl*. 2019;13(3):e1800043.
 103. Chen ZZ, Gerszten RE. Metabolomics and proteomics in type 2 diabetes. *Circ Res*. 2020;126(11):1613–27.
 104. Smith JG, Gerszten RE. Emerging affinity-based proteomic technologies for large-scale plasma profiling in cardiovascular disease. *Circulation*. 2017;135(17):1651–64.
 105. Liu X, Feng Q, Chen Y, Zuo J, Gupta N, Chang Y, et al. Proteomics-based identification of differentially-expressed proteins including galectin-1 in the blood plasma of type 2 diabetic patients. *J Proteome Res*. 2009;8(3):1255–62.
 106. Jensen TM, Witte DR, Pieragostino D, McGuire JN, Schjerning ED, Nardi C, et al. Association between protein signals and type 2 diabetes incidence. *Acta Diabetol*. 2013;50(5):697–704.
 107. O'Rourke MB, Januszewski AS, Sullivan DR, Lengyel I, Stewart AJ, Arya S, et al. Optimised plasma sample preparation and LC-MS analysis to support large-scale proteomic analysis of clinical trial specimens: application to the Fenofibrate Intervention and Event Lowering in Diabetes (FIELD) trial. *Proteomics Clin Appl*. 2023;17(3):e2200106.
 108. Rohloff JC, Gelinas AD, Jarvis TC, Ochsner UA, Schneider DJ, Gold L, et al. Nucleic Acid ligands with protein-like side chains: modified aptamers and their use as diagnostic and therapeutic agents. *Mol Ther Nucleic Acids*. 2014;3(10):e201.
 109. Ngo D, Sinha S, Shen D, Kuhn EW, Keyes MJ, Shi X, et al. Aptamer-based proteomic profiling reveals novel candidate biomarkers and pathways in cardiovascular disease. *Circulation*. 2016;134(4):270–85.
 110. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. *Nature*. 2018;558(7708):73–9.
 111. Nowak C, Sundstrom J, Gustafsson S, Giedraitis V, Lind L, Ingelsson E, et al. Protein biomarkers for insulin resistance and type 2 diabetes risk in two large community cohorts. *Diabetes*. 2016;65(1):276–84.
 112. Beijer K, Nowak C, Sundstrom J, Arnlov J, Fall T, Lind L. In search of causal pathways in diabetes: a study using proteomics and genotyping data from a cross-sectional study. *Diabetologia*. 2019;62(11):1998–2006.
 113. Yazdanpanah N, Yazdanpanah M, Wang Y, Forgetta V, Polak M, Polychronakos C, et al. Clinically relevant circulating protein biomarkers for type 1 diabetes: evidence from a two-sample Mendelian randomization study. *Diabetes Care*. 2022;45(1):169–77.
 114. Ghanbari F, Yazdanpanah N, Yazdanpanah M, Richards JB, Manousaki D. Connecting genomics and proteomics to identify protein biomarkers for adult and youth-onset type 2 diabetes: a two-sample Mendelian randomization study. *Diabetes*. 2022;71(6):1324–37.
 115. Ahola-Olli AV, Mustelin L, Kalimeri M, Kettunen J, Jokelainen J, Auvinen J, et al. Circulating metabolites and the risk of type 2 diabetes: a prospective study of 11,896 young adults from four Finnish cohorts. *Diabetologia*. 2019;62(12):2298–309.
 116. Wang TJ, Larson MG, Vasani RS, Cheng S, Rhee EP, McCabe E, et al. Metabolite profiles and the risk of developing diabetes. *Nat Med*. 2011;17(4):448–53.
 117. Malik VS, Popkin BM, Bray GA, Despres JP, Willett WC, Hu FB. Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes: a meta-analysis. *Diabetes Care*. 2010;33(11):2477–83.
 118. Yang J, Xu G, Hong Q, Liebich HM, Lutz K, Schmulling RM, et al. Discrimination of type 2 diabetic patients from healthy controls by using metabolomics method based on their serum fatty acid profiles. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2004;813(1–2):53–8.
 119. Jin Q, Ma RCW. Metabolomics in diabetes and diabetic complications: insights from epidemiological studies. *Cells*. 2021;10(11). **This article provides a comprehensive overview of metabolomics studies in diabetes conducted in epidemiological cohorts, and also include more detailed discussions of the merits of different platforms of technology for metabolomic profiling**
 120. Yuan K, Kong H, Guan Y, Yang J, Xu G. A GC-based metabolomics investigation of type 2 diabetes by organic acids metabolic profile. *J Chromatogr B Analyt Technol Biomed Life Sci*. 2007;850(1–2):236–40.
 121. Zhang J, Yan L, Chen W, Lin L, Song X, Yan X, et al. Metabolomics research of diabetic nephropathy and type 2 diabetes mellitus based on UPLC-oeTOF-MS system. *Anal Chim Acta*. 2009;650(1):16–22.
 122. Aboulizadeh E, Ranji M, Sorenson CM, Sepehr R, Sheibani N, Hirschmugl CJ. Retinal oxidative stress at the onset of diabetes determined by synchrotron FTIR widefield imaging: towards diabetes pathogenesis. *Analyst*. 2017;142(7):1061–72.
 123. Perez de Souza L, Alseekh S, Scossa F, Fernie AR. Ultra-high-performance liquid chromatography high-resolution mass spectrometry variants for metabolomics research. *Nat Methods*. 2021;18(7):733–46.
 124. Shi L, Brunius C, Lehtonen M, Auriola S, Bergdahl IA, Rolandsen O, et al. Plasma metabolites associated with type 2 diabetes in a Swedish population: a case-control study nested in a prospective cohort. *Diabetologia*. 2018;61(4):849–61.

125. de Mello VD, Paananen J, Lindstrom J, Lankinen MA, Shi L, Kuusisto J, et al. Indolepropionic acid and novel lipid metabolites are associated with a lower risk of type 2 diabetes in the Finnish Diabetes Prevention Study. *Sci Rep.* 2017;7:46337.
126. Bragg F, Kartsonaki C, Guo Y, Holmes M, Du H, Yu C, et al. Circulating metabolites and the development of type 2 diabetes in Chinese adults. *Diabetes Care.* 2022;45(2):477–80.
127. Lu J, Lam SM, Wan Q, Shi L, Huo Y, Chen L, et al. High-coverage targeted lipidomics reveals novel serum lipid predictors and lipid pathway dysregulation antecedent to type 2 diabetes onset in normoglycemic Chinese adults. *Diabetes Care.* 2019;42(11):2117–26.
128. Glynn P, Greenland P. Contributions of the UK biobank high impact papers in the era of precision medicine. *Eur J Epidemiol.* 2020;35(1):5–10.
129. Backman JD, Li AH, Marcketta A, Sun D, Mbatchou J, Kessler MD, et al. Exome sequencing and analysis of 454,787 UK Biobank participants. *Nature.* 2021;599(7886):628–34.
130. Ritchie SC, Lambert SA, Arnold M, Teo SM, Lim S, Scepanovic P, et al. Integrative analysis of the plasma proteome and polygenic risk of cardiometabolic diseases. *Nat Metab.* 2021;3(11):1476–83.
131. Carrasco-Zanini J, Pietzner M, Lindbohm JV, Wheeler E, Oerton E, Kerrison N, et al. Proteomic signatures for identification of impaired glucose tolerance. *Nat Med.* 2022;28(11):2293–300.
132. Makinen VP, Tynkkynen T, Soinen P, Peltola T, Kangas AJ, Forsblom C, et al. Metabolic diversity of progressive kidney disease in 325 patients with type 1 diabetes (the FinnDiane Study). *J Proteome Res.* 2012;11(3):1782–90.
133. McKay J. Nursing Aid Fertility control. *Nurs Times.* 1987;83(34):46–7.
134. Jin Q, Lau ESH, Luk AO, Tam CHT, Ozaki R, Lim CKP, et al. High-density lipoprotein subclasses and cardiovascular disease and mortality in type 2 diabetes: analysis from the Hong Kong Diabetes Biobank. *Cardiovasc Diabetol.* 2022;21(1):293.
135. Tam CHT, Lim CKP, Luk AOY, Shi M, Man Cheung H, Ng ACW, et al. Identification of a common variant for coronary heart disease at PDE1A contributes to individualized treatment goals and risk stratification of cardiovascular complications in Chinese patients with type 2 diabetes. *Diabetes Care.* 2023;46(6):1271–81.
136. Bomba L, Walter K, Guo Q, Surendran P, Kundu K, Non-gmaithem S, et al. Whole-exome sequencing identifies rare genetic variants associated with human plasma metabolites. *Am J Hum Genet.* 2022;109(6):1038–54.
137. Ong KL, Wu L, Januszewski A, O'Connell R, Xu A, Scott RS, et al. The relationship of neutrophil elastase and proteinase 3 with risk factors, and chronic complications in type 2 diabetes: a Fenofibrate Intervention and Event Lowering in Diabetes (FIELD) sub-study. *Diab Vasc Dis Res.* 2021;18(4):14791641211032548.
138. Morieri ML, Gao H, Pigeire M, Shah HS, Sjaarda J, Mendonca C, et al. Genetic tools for coronary risk assessment in type 2 diabetes: a cohort study from the ACCORD clinical trial. *Diabetes Care.* 2018;41(11):2404–13.
139. Shah HS, Gao H, Morieri ML, Skupien J, Marvel S, Pare G, et al. Genetic predictors of cardiovascular mortality during intensive glycemic control in type 2 diabetes: findings from the ACCORD clinical trial. *Diabetes Care.* 2016;39(11):1915–24.
140. Tremblay J, Haloui M, Attaoua R, Tahir R, Hishmih C, Harvey F, et al. Polygenic risk scores predict diabetes complications and their response to intensive blood pressure and glucose control. *Diabetologia.* 2021;64(9):2012–25.
141. Zelniker TA, Wiviott S, Mosenzou O, Goodrich EL, Jarolim P, Cahn A, et al. Association of cardiac biomarkers with major adverse cardiovascular events in high-risk patients with diabetes: a secondary analysis of the DECLARE-TIMI 58 trial. *JAMA Cardiol.* 2023;8(5):503–9.
142. Chu N, Chan JCN, Chow E. Pharmacomicrobiomics in Western medicine and traditional Chinese medicine in type 2 diabetes. *Front Endocrinol (Lausanne).* 2022;13:857090.
143. Misra BB, Misra A. The chemical exposome of type 2 diabetes mellitus: opportunities and challenges in the omics era. *Diabetes Metab Syndr.* 2020;14(1):23–38.
144. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: a review. *Biotechnol Adv.* 2021;49:107739.
145. Hu C, Jia W. Multi-omics profiling: the way towards precision medicine in metabolic diseases. *J Mol Cell Biol.* 2021;13(8):576–93.
- 146.●● Deirdre KT, Jordi M, Abrar A, Catherine A, Jamie LB, Dhana-sekaran B, et al. Second international consensus report on gaps and opportunities for the clinical translation of precision diabetes medicine. *Nat Med.* 2023;29(10):2438–57. <https://doi.org/10.1038/s41591-023-02502-5>. **This international consensus report from the American Diabetes Association/European Association for the Study of Diabetes Precision Medicine in Diabetes Initiative (ADA/EASD PMDI) provides an up-to-date overview of the current state of knowledge for precision medicine in diabetes, and is based on the conclusions from 15 systematic reviews addressing different aspects of precision medicine in type 1 diabetes, type 2 diabetes, and gestational diabetes.**

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This review article is partly based on an invited presentation delivered by RCWM at the International Diabetes Epidemiology Group (IDEG) 2022 conference at Porto, Portugal in December 2022.