

1 **A whole blood-based transcriptional risk score for nonobese type 2 diabetes**

2 **predicts dynamic changes in glucose metabolic traits**

3

4 **Running title:** Blood-based transcriptional risk score, diabetes

5

6 Yanan Hou,^{1,2*} Huajie Dai,^{1,2*} Na Chen,^{1,2*} Zhiyun Zhao,^{1,2*} Qi Wang,^{1,2} Tianzhichao

7 Hou,^{1,2} Jie Zheng,^{1,2} Tiange Wang,^{1,2} Mian Li,^{1,2} Hong Lin,^{1,2} Shuangyuan Wang,^{1,2}

8 Ruizhi Zheng,^{1,2} Jieli Lu,^{1,2} Yu Xu,^{1,2} Yuhong Chen,^{1,2} Ruixin Liu,^{1,2} Guang Ning,^{1,2}

9 Weiqing Wang,^{1,2} Yufang Bi,^{1,2†} Jiqiu Wang,^{1,2†} Min Xu^{1,2†}

10

11 **Author affiliations:**

12 1. Department of Endocrine and Metabolic Diseases, Shanghai Institute of Endocrine
13 and Metabolic Diseases, Ruijin Hospital, Shanghai Jiao Tong University School of
14 Medicine, Shanghai, China;

15 2. Shanghai National Clinical Research Center for metabolic Diseases, Key
16 Laboratory for Endocrine and Metabolic Diseases of the National Health Commission
17 of the PR China, Shanghai Key Laboratory for Endocrine Tumor, State Key
18 Laboratory of Medical Genomics, Ruijin Hospital, Shanghai Jiao Tong University
19 School of Medicine, Shanghai, China.

20

21 * These authors contributed equally to the work

22 † These authors co-supervised the work

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

23

24 †**Address of Correspondence:**

25 Min Xu, MD&PhD, Professor of medicine, Shanghai National Clinical Research

26 Center for Endocrine and Metabolic Diseases, Shanghai Institute of Endocrine and

27 Metabolic Diseases, Ruijin Hospital, Shanghai Jiao Tong University School of

28 Medicine, Shanghai, 200025, China. E-mail: della.xumin@shsmu.edu.cn;

29

30 Jiqu Wang, MD&PhD, Professor of medicine, Department of Endocrine and

31 Metabolic Diseases, Shanghai Institute of Endocrine and Metabolic Diseases, Ruijin

32 Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China,

33 200025. Email: wangjq@shsmu.edu.cn;

34

35 Yufang Bi, MD&PhD, Professor of medicine, Department of Endocrine and

36 Metabolic Diseases, Shanghai Institute of Endocrine and Metabolic Diseases, Ruijin

37 Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China.

38 Email: byf10784@rjh.com.cn.

39 **Abstract**

40 **Background:** The performance of peripheral blood transcriptional markers in
41 evaluating the risk of type 2 diabetes (T2D) with normal weight is unknown. We
42 developed a whole blood-based transcriptional risk score (wb-TRS) for nonobese T2D
43 and assessed its contributions to disease risk and dynamic changes in glucose
44 metabolism.

45 **Methods and findings:** We developed the wb-TRS in 1105 participants aged ≥ 40
46 years and in normal weight for up to 10 years from a well-defined community-based
47 cohort with blood transcriptome data and validated it in an external dataset (253
48 overweight/obese participants from a dietary intervention trial with 3 repeated
49 transcriptome data). Potential biology significance and causal inference were also
50 explored. The wb-TRS included 144 transcripts. Compared to the lowest tertile, wb-
51 TRS in tertile 3 associated with 8.68-folds (95% confidence interval [CI], 3.51-21.5),
52 and each 1-unit increment associated with 2.57-folds (95% CI, 1.86-3.56) higher risk
53 of nonobese T2D, after adjustments for traditional risk factors. Furthermore, baseline
54 wb-TRS was significantly associated with dynamic changes in average, daytime,
55 nighttime and 24h glucose and HbA1c, and area under the curve of glucose measured
56 in the continuous glucose monitoring during 6-month of intervention. The wb-TRS
57 improved the predicting performance for nonobese T2D in a model with fasting
58 glucose, triglycerides and demographic and anthropometric parameters. Mitch
59 analysis implicated oxidative phosphorylation, cholesterol metabolism and mTORC1
60 signaling involved in nonobese T2D pathogenesis. Transcriptome-wide Mendelian

61 randomization supported causal effects of gene transcripts such as RAB1A and

62 GCC1-PAX4 on nonobese T2D risk.

63 **Conclusions:** A whole blood based nonobese T2D associated TRS was validated to

64 predict dynamic changes in glucose metabolism. These findings also suggested

65 several genes and biological pathways that might involve in the pathogenesis of

66 nonobese T2D.

67

68 **Keywords:** nonobese type 2 diabetes, glucose metabolic traits, blood transcriptome,

69 transcriptional risk score, risk prediction, transcriptome-wide Mendelian

70 randomization

71 **Abbreviations:**

72 AUC – area under the curve

73 BMI – body mass index

74 CPM – count per million

75 CI – confidence interval

76 DBP – diastolic blood pressure

77 DPMH – Dietary Pattern and Metabolic Health

78 eQTL – expression quantitative trait locus

79 GEE – generalized estimating equation

80 HbA1c – hemoglobin A1c

81 HOMA-IR – homeostasis model assessment of insulin resistance

82 HOMA- β – homeostasis model assessment of beta cell function

83 HDL-C – high-density lipoprotein cholesterol

84 MET – metabolic equivalent of task

85 Mitch analysis – a multi-contrast gene set enrichment analysis

86 LDL-C – low-density lipoprotein cholesterol

87 LASSO – least absolute shrinkage and selection operator

88 OGTT – oral glucose tolerance test

89 OR – odds ratio

90 RCS – restricted cubic spline

91 RNA-seq – RNA sequencing

92 SBP – systolic blood pressure

- 93 T2D – Type 2 diabetes
- 94 TRS – transcriptional risk score
- 95 TWMR – Transcriptome-wide Mendelian randomization
- 96 TSGE – Tissue-specific gene expression

97 **Author summary**

98 **Why was this study done?**

- 99 ● Blood transcriptome has been emerging as a potential novel approach to
100 understand contribution of genes expression to biological functions and to
101 identify key genes that confer to complex disease susceptibility, such as
102 cardiovascular, psychiatry and chronic obstructive pulmonary disease.
- 103 ● Rare studies have investigated the performance of the peripheral blood
104 transcriptional markers in evaluating the risk and prediction of type 2 diabetes
105 (T2D).

106 **What did the researchers do and find?**

- 107 ● We developed a whole blood-based transcriptional risk score (wb-TRS) for T2D
108 risk in a longitudinal community-based cohort.
- 109 ● The wb-TRS was positively associated with risk of T2D and glucose metabolic
110 traits. Furthermore, the baseline wb-TRS as well associated with dynamic
111 changes of glucose metabolism measured in the continuous glucose monitoring in
112 response to 6-month dietary intervention.
- 113 ● The wb-TRS performed better and was complementary to fasting glucose,
114 triglycerides and the demographic and anthropometric factors in predicting T2D.
- 115 ● Transcriptome-wide Mendelian randomization and Mitch analysis and supported
116 several gene transcripts (RAB1A and GCC1-PAX4) and biological pathways
117 (oxidative phosphorylation, mTORC1 signaling, and cholesterol metabolism) that
118 might involve in the pathogenesis of T2D.

119 **What do these findings mean?**

- 120 ● Blood transcriptome highly related to and reflect risk of T2D and predict dynamic
121 changes in glucose metabolism. It may lend insight into discerning potential
122 causal genes and biological mechanisms in the pathogenesis of nonobese T2D.

123 **Introduction**

124 Type 2 diabetes (T2D) has been a worldwide epidemic in recent decades leading high
125 rates of early mortality and co-morbidity and possessing a considerable health and
126 economic burden for the patients [1]. Early prediction and detection of individuals
127 with glucose dysregulation and type 2 diabetes is extremely important for the
128 prevention and management of these disorders. Previous reliable risk prediction
129 models usually comprise the anthropometric measurements and traditional laboratory
130 data. Recent studies have provided evidence of the prediction effect by including
131 genetic factors and metabolic molecules in the models [2, 3]. However, as an essential
132 biological and pathogenesis process, the performance and effect of gene transcription
133 profile in evaluating T2D risk has not been well examined yet [4].

134 Tissue-specific gene expression (TSGE) is phenotypic manifestations integrating
135 both genetic and environmental variations, which confers further insights to the
136 intricate pathophysiology of diseases [5]. However, most tissues such as pancreas,
137 liver, etc. cannot be readily obtained without invasive procedures in humans. As an
138 alternative, blood transcriptome could better predict TSGE for ~60% of the genes on
139 average, with up to 81.7% in skeletal muscle, 71.5% in visceral adipose tissue, and
140 52.4% in pancreas [5]. Blood transcriptome has been emerging as a potential novel
141 approach for complex diseases to understand contribution of genes expression to
142 biological functions and to identify key genes that confer to complex disease
143 susceptibility [6-9], such as cardiovascular, psychiatry and chronic obstructive
144 pulmonary disease. However, rare studies have investigated the performance of the

145 peripheral blood transcriptional markers in evaluating the risk and prediction of T2D.

146 Though obesity is one of its major risk factors, the T2D risk is still high in normal
147 weight, especially in east Asians. One in three of diagnosed T2D had a body mass
148 index (BMI) of less than 25 kg/m², indicating the importance of metabolic health in
149 the pathogenesis regardless of adiposity [10, 11]. Studies focused on refining T2D
150 subtypes highlighting nonobese T2D, which characterized by a lower BMI, early
151 onset of age, severe insulin deficiency, and serious complications [12-15]. The east
152 Asians developed T2D risk at a relatively lower level of BMI [16]. The existing
153 evidence indicated that specific genes, pathways, and biological mechanisms might
154 underly the pathophysiology of nonobese T2D.

155 Thus, we developed a whole blood-based transcriptional risk score (wb-TRS) of
156 nonobese T2D in 1105 participants aged 40 and above in an ongoing longitudinal
157 Chinese cohort, in which the participants have maintained normal weight for up to 10
158 years. To validate its performance in a more general age and BMI range, we examined
159 its predictive value on dynamic change in glucose metabolism in an independent
160 external dataset (253 young overweight/obese participants from a dietary intervention
161 trial with 3 repeated whole blood-based transcriptome data). Furthermore, we
162 explored the potential genes and biological pathways that may conferred disease's
163 susceptibility and pathogenesis and performed transcriptome-wide Mendelian
164 randomization (TWMR) to investigate additional evidence supporting a causal links
165 between gene transcripts and T2D risk in a separate East Asian population and
166 European population. The wb-TRS would provide an additional layer of biological

167 interpretability for T2D.

168

169 **Methods**

170 **Study population**

171 The current analysis was based on one of the follow-up circles of our previous
172 community-based cohort of in Shanghai, China [17-19]. Briefly, all the participants
173 have taken part in a comprehensive examination that included a questionnaire and
174 biochemical measurements at baseline in 2008-2009, the first (5- years) in 2013 and
175 second (10-years) follow-up visits in 2018, respectively. The participants with known
176 T2D before baseline examination were excluded.

177 According to the information at baseline and the 5-year follow-up examination, at
178 the 10-year follow-up visit, we selected the participants whose BMI <25kg/m² in past
179 10 years to take part in the examination sequentially and collected the whole blood
180 samples for RNA-sequencing (RNA-seq). Finally, 1183 participants attended the
181 follow-up visit finished the blood sampling for RNA-seq and 1105 with high-quality
182 sequencing data were included in the present analysis. The flowchart of study
183 participants was shown in **Fig S1**.

184 The study protocol was approved by the Institutional Review Board of Rui-Jin
185 Hospital affiliated to Shanghai Jiao Tong University School of Medicine. All
186 participants gave the written consents for the study.

187

188 **Data collection**

189 A standard questionnaire was adopted to collect information on sociodemographic
190 characteristics, disease and medical history, and lifestyles. Current smoking or
191 drinking was defined as participants consumed cigarettes or alcohol regularly in the
192 past 6 months. We used Global Physical Activity Questionnaire (GPAQ) to obtain
193 frequency, duration, and intensity of physical activity. Physical activity level was
194 evaluated as Metabolic Equivalent of Task (MET) minutes per week. Body height,
195 weight and waist circumference were measured using a standard protocol. BMI was
196 calculated as body weight (kg) divided by squared height (m²). Three seated systolic-
197 and diastolic blood pressure (SBP and DBP) were measured (Omron HEM-752
198 Fuzzy; Omron, Dalian, China) at the non-dominant arm consecutively with 1 min
199 interval after at least 5-minutes rest. The three readings were averaged for analysis.

200

201 **Biochemical measurements**

202 Venous blood samples were obtained after an overnight fasting (>10 hours). All
203 participants underwent a 75 g oral glucose tolerance test (OGTT). Fasting and 2h-
204 OGTT glucose was measured using glucose oxidase method on an automated
205 analyzer (ADVIA-1650 Chemistry System; Bayer); fasting and 2h-OGTT insulin
206 using electrochemiluminescence assay on an automated analyzer (Roche Diagnostics,
207 Basel, Switzerland); and hemoglobin A1c (HbA1c) using high pressure liquid
208 chromatography on an automated analyzer (VARIANT II Hemoglobin Testing
209 System, Bio-Rad Laboratories). Fasting triglycerides, total cholesterol, low-density
210 and high-density lipoprotein cholesterol (LDL-C and HDL-C) were measured on the

211 automated analyzer (ADVIA-1650 Chemistry System; Bayer).

212

213 **Definition of T2D**

214 T2D was diagnosed as fasting glucose ≥ 7.0 mmol/l, 2h-OGTT glucose ≥ 11.1 mmol/l,

215 HbA1c $\geq 6.5\%$, and/or taking antidiabetic therapy [20]. The homeostasis model

216 assessment of insulin resistance (HOMA-IR) was calculated as fasting glucose

217 (mmol/l) \times fasting insulin (uIU/ml) / 22.5. The HOMA index for beta cell function

218 (HOMA- β) was calculated as $20 \times$ fasting insulin (uIU/ml) / (fasting glucose

219 [mmol/l]-3.5).

220

221 **RNA-seq**

222 The whole blood samples were collected by using the Paxgene RNA tubes according

223 to the manufacture's protocol, shipped and stored in -80°C before RNA extracting.

224 RNA was extracted by using the PAXgene Blood miRNA Kit Secondary BD

225 Hemogard Closures (Catalog no 763134, Qiagen, Germany). Nanodrop One

226 spectrometer (Thermo Fisher Scientific Inc., USA) and Qubit®3.0 Fluorometer (Life

227 Technologies, CA, USA) were used to determine RNA purity and concentration.

228 Isolated total RNA was analyzed on an Agilent Bioanalyzer 2100 (Agilent

229 technologies, Santa Clara, CA, US) for RNA integrity number quality check. We

230 excluded 78 samples with low RNA quality. 1ug RNA was used as input material for

231 the RNA-seq preparations.

232 RNA-seq strand-specific libraries were constructed using the VAHTS Total RNA-

233 seq (H/M/R) Library Prep Kit (Vazyme, China). Briefly, RNA was purified by
234 magnetic beads after removal of rRNA. Then RNA was fragmented into small pieces
235 using divalent cations for 8 minutes at 94°C. The cleaved RNA fragments were copied
236 into first strand cDNA using reverse transcriptase and random primers. Second strand
237 cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H.
238 These cDNA fragments then go through the end repair process, addition of a single
239 ‘A’ base, and ligation of the adapters. The products were purified and enriched by
240 PCR to create the final cDNA library. Purified libraries were quantified and validated
241 by Qubit® 3.0 Fluorometer and Agilent 2100 bioanalyzer to confirm the insert size
242 and calculate the mole concentration. Cluster was generated by cBot after the library
243 diluted to 10 pM. Cluster were then sequenced in 150-bp paired-end mode on the
244 Illumina NovaSeq 6000 platform (Illumina, USA) generating arounds 12 gigabases
245 (Gb) reads per samples.

246 The base call and sequence quality across all generated FASTQ data was evaluated
247 using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). We
248 removed adapters, low quality reads, ribosome RNA reads, and reads with a sequence
249 length less than 25 to obtain clean reads. Clean reads were mapped to human
250 GRCh38.102 with Hisat2 (Hierarchical Indexing for Spliced Alignment of
251 Transcripts, version 2.0.5) using default parameters. Overall, we obtained RNA-seq
252 data with base mass (Q30) $\geq 90\%$, average mapped ratio was 94%. Count data were
253 filtered to include transcripts with > 1 count per million (CPM) in 99% of T2D cases
254 or non-T2D controls.

255

256 **wb-TRS construction**

257 The 1105 participants with blood-based transcriptome data were randomly assigned to
258 training (75%) and testing (25%) dataset (R package caret) (**Fig 1**). To identify genes
259 that could better distinguish the diseases status, we included the transcripts with gene
260 expression level >1CPM in 99% of T2D cases or non-T2D controls and with a
261 nominal difference P value <0.01 between these two groups in the wb-TRS features
262 screening progress (n=757 gene transcripts). The wb-TRS construction was carried
263 out in the training dataset and validated in testing and an external dataset. A multi-
264 contrast gene set enrichment (Mits) analysis [21] and Reactome pathway analysis
265 [22] were performed to identify key genes and pathways.

266 In detail, we performed a penalized logistic regression framework using R package
267 glmnet to develop the wb-TRS for nonobese T2D. We adopted the least absolute
268 shrinkage and selection operator (LASSO) regression model to model training, which
269 provides shrinks coefficients toward zero and automated feature selection. We
270 adopted 10-fold cross-validation procedure to determine the lambda (0.006) that
271 maximized the area under curve (AUC) for predicting T2D (**Fig S2**). The regression
272 model with maximum AUC was used to calculate the wb-TRS.

273

274 **Statistical analysis**

275 Multivariable logistic regression analyses were used to assess the risk of nonobese
276 T2D in relation to wb-TRS. Odds ratio (OR) and the corresponding 95% confidence

277 interval (CI) were calculated. The linear regression model was to assess the
278 association of wb-TRS with the cardiometabolic traits. The restricted cubic spline
279 (RCS) analysis was used to test any nonlinear associations between wb-TRS and
280 glucose metabolism. The receiver operator characteristic curves (ROC) were used to
281 evaluate and compare the predictive values of 6 models for nonobese T2D (R package
282 pROC) including: 1. primary model (age, sex, BMI and family history of diabetes); 2.
283 wb-TRS model; 3. fasting glucose and triglycerides model; 4. primary model adding
284 fasting glucose and triglycerides (full); 5. primary model adding wb-TRS; 6. primary
285 model adding fasting glucose and triglycerides (full), and wb-TRS.

286

287 **Independent validation**

288 To validate the predictive value of wb-TRS for the dynamic changes in glucose
289 metabolism, we performed analysis using data of 3 repeated blood-based
290 transcriptome data and glucose metabolism in the Dietary Pattern and Metabolic
291 Health (DPMH) study, a parallel-arm randomized controlled feeding trial conducted
292 in 2019 (URL: <https://www.clinicaltrials.gov>; Unique identifier: NCT03856762)
293 [23]. A total of 253 overweight/obese participants (aged 25 to 60 years, BMI \geq
294 24.0 kg/m², fasting glucose \geq 5.6 mmol/l) with transcriptome data and glucose
295 metabolic traits at baseline, the first (3-months) and second (6-months) follow-up
296 visits were included in the validation analysis. The whole blood RNA was collected
297 and underwent RNA-seq following the same procedure with the above cohort study.
298 We adopted generalized estimating equations (GEE) models were to assess the

299 association of three-repeated measurements of wb-TRS with three-repeated
300 measurements of the glucose metabolism during the 6-months intervention. We
301 further fixed the generalized linear models to assess the effects of baseline wb-TRS
302 on dynamic changes of glucose metabolism at 3- and 6- months of intervention.

303

304 **The biological pathway and mechanism analysis**

305 To gain biological insight behind the wb-TRS, we performed several bioinformatics
306 analyses. We divided the participants in the initial cohort into 3 subgroups based on
307 wb-TRS tertiles. We performed a multi-contrast gene set enrichment analysis by using
308 R package mitch (v.1.4.0). The Mitch analysis uses a rank-MANOVA statistical
309 approach to identify sets of genes that exhibit joint enrichment or divergent responses
310 across multiple contrast. The input data was the \log_2 (fold change) of all detected
311 transcripts (n=54,458, including protein coding gene, lncRNA, and snRNA) for
312 comparison between the bottom and mid tertile, and the top and mid tertile,
313 respectively. The input genesets provided were 50 genesets: h.all.v7.5.1.symbols.gmt
314 in Molecular Signatures Database (MSigDB). Enrichment score (ES) was obtained
315 and ranged from -1 to 1. Meanwhile, we used the 144 gene transcripts included in the
316 wb-TRS as input profiles to perform Reactome pathway analysis on reactome website
317 (<https://reactome.org/PathwayBrowser/#TOOL=AT>).

318

319 **Transcriptome-wide Mendelian randomization**

320 We performed a two-sample transcriptome-wide Mendelian randomization (TWMR)

321 to identify causal associations of gene transcripts included in wb-TRS with T2D risk
322 using the TwoSampleMR. The summary statistics of genome-wide association study
323 of T2D was from the Asian Genetic Epidemiology Network (AGEN) [16] and the
324 DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) [24], respectively.
325 We selected the significant expression quantitative trait loci (eQTLs with $P < 5 \times 10^{-8}$)
326 obtained from eQTLGen Consortium [25] and with linkage disequilibrium (LD) based
327 on $r^2 < 0.01$ as valid instruments variants (IVs) used in TWMR. When no shared
328 eQTLs were available between exposures and outcomes, proxies with LD $r^2 > 0.8$
329 were substituted. TWMR causality tests were performed using the inverse variance
330 weighted method by meta-analyzing the effect of each IV from Wald ratio.
331 Heterogeneity and pleiotropy of the results were assessed using Cochran's Q statistic
332 or MR-Egger intercept, respectively.

333 All analyses were performed using SAS (version 9.3; SAS Institute, Cary, NC,
334 USA) and R software (version 3.5.1; R Core Team). A two-sided P value < 0.05 was
335 considered statistically significant.

336

337 **Results**

338 **Clinical characteristics of study participants**

339 In the 1105 participants, 305 were defined as nonobese T2D (27.6%), 737 (66.7%)
340 were women, with an average age of 64.5 (standard deviation [SD], 7.0) years old.

341 The balanced clinical characteristics between the training and testing dataset indicated
342 a low risk of selection bias introduced by random sampling (**Table S1**).

343

344 **Construction and distribution of wb-TRS**

345 Finally, 144 important features (gene transcripts) were included in the wb-TRS, which
346 weights were displayed in **Table S2**. The wb-TRS was estimated as gene expression
347 level of transcript1 \times weight1 + gene expression level of transcript2 \times weight2 + ... +
348 gene expression level of transcript144 \times weight144.

349 The wb-TRS was positively associated with age (β 0.98, SE 0.15, $P < 0.0001$), and
350 was slightly higher in men (-1.7, SD 1.5) than that in women (-2.1, SD 1.3). The mean
351 wb-TRS was -0.5 (SD 0.9) in T2D and -2.5 (SD 1.1) in non-diabetes in training
352 dataset, and -0.9 (SD 1.4) in T2D and -2.4 (SD 1.3) in non-diabetes in testing dataset.

353 **Fig S3** displays the violin plot of the wb-TRS and gene transcripts most negatively
354 (FUOM) and positively (RPF1) associated with T2D in training dataset.

355

356 **Association of wb-TRS with nonobese T2D**

357 **Table 1** shows association of wb-TRS with nonobese T2D risk in testing dataset. Each
358 1-unit increment in wb-TRS (one score) was associated with a 2.49-folds higher risk
359 of nonobese T2D (95%CI 1.84-3.36), after adjustments of age, sex, BMI and family
360 history of diabetes. Additional adjustment for potential confounders including
361 lifestyles, education levels, waist circumference, SBP and DBP, and lipids profiles did
362 not change the results appreciably (OR 2.57, 95% CI 1.86-3.56). Specifically, each 1-
363 unit increment in wb-TRS was associated with 3.40-folds higher risk of newly
364 diagnosed nonobese T2D (95%CI 2.00-5.81), and 2.33-folds higher risk of diabetes

365 with duration ≥ 5 years (95%CI 1.56-3.48). Compared with the bottom tertile, the
366 middle or top tertile of wb-TRS was associated with 2.46- (95% CI 0.95-6.35) and
367 8.68-folds (95% CI 3.51-21.5) higher risk of nonobese T2D.

368 We then examined the associations of wb-TRS with nonobese T2D risk stratified
369 by several traditional risk factors of T2D (**Fig S4**). A stronger association was
370 identified in age <65 years and women (a marginal interaction between TRS and sex
371 was observed with P for interaction =0.051).

372

373 **Association of wb-TRS with glucose metabolism**

374 In both training and testing dataset, the wb-TRS was positively associated with fasting
375 glucose, 2h-OGTT glucose, and HbA1c; but inversely associated with HOMA- β , in
376 the fully adjusted model (**Table 2**). The RCS analysis showed similar results, except
377 for that it demonstrated a J shaped relationships between wb-TRS and HbA1c and
378 logHOMA_ β (all P for non-linear ≤ 0.001) (**Fig 2**).

379 The GEE analysis by using the 3 repeated measurements of glucose metabolism
380 traits and wb-TRS at baseline, 3 month- and 6 month-visits during the intervention
381 showed that each 1-unit increment in wb-TRS was associated with 0.11mmol/l (SE
382 0.05) and 0.15mmol/l (SE 0.02) increment of 1h- and 2h-OGTT glucose, respectively.
383 Interestingly, the baseline wb-TRS significantly affected the dynamic changes of
384 cardiometabolic traits at 6- months of dietary intervention (**Table 3**). Baseline wb-
385 TRS significantly associated with changes in average daytime, nighttime and 24h
386 glucose and HbA1c levels and the AUC of glucose measured in the continuous

387 glucose monitoring at 6 month of diet intervention (all $P < 0.03$).

388

389 **wb-TRS performed well in predicting T2D**

390 In both training and testing dataset, wb-TRS performed well in predicting nonobese
391 T2D, even when adding wb-TRS into a model with fasting glucose, triglycerides, and
392 the primary model (**Fig 3**). In total dataset, a model with wb-TRS along with fasting
393 glucose, triglycerides, and variables in primary model significantly improved the
394 AUC (0.93), as compared to the primary model (AUC =0.65), the fasting glucose and
395 triglycerides model (AUC=0.82), and the fasting glucose, triglycerides adding to the
396 primary model (full model, AUC =0.84) for predicting T2D (**Table S3**). The wb-TRS
397 improved predicting performance in a model with fasting glucose, triglycerides, and
398 the demographic and anthropometry parameters (AUC-TRS+full: 0.93 [0.91-0.95] vs.
399 AUC-full: 0.84 [95% CI 0.81-0.87]; $P < 0.0001$).

400

401 **The biological pathway and mechanism analysis**

402 The Mitch analysis compared the bottom versus middle and top versus middle tertiles
403 of wb-TRS (**Fig 4**). The top versus middle tertile group comparison demonstrated
404 downregulation of MYC target V2, and the bottom versus middle tertile group
405 comparison demonstrated upregulation of oxidative phosphorylation, mTORC1
406 signaling, cholesterol metabolism, peroxisome, fatty acid metabolism, etc. (**Fig 4A**).
407 The significant metabolic pathways were displayed in **Fig 4B**. The small GTPase
408 (RAB1A gene) enriched in mTORC1 signaling, fatty acid oxidation gene (CPT1A)

409 enriched in oxidative phosphorylation pathway, and cholesterol efflux gene (ABCA1)

410 enriched in cholesterol metabolism pathway.

411 The Reactome analyses suggested that the 144 gene transcripts included in wb-TRS

412 were involved in NR1H2 (LXR β) and NR1H3 (LXR α) regulate gene expression to

413 limit cholesterol uptake, regulation of lipid metabolism by PPAR alpha pathway, etc.

414 The cholesterol uptake suppression gene (MYLIP) enriched in LXR β and LXR α

415 regulate gene expression to limit cholesterol uptake pathway (**Table S4**).

416

417 **Causal relationships between gene transcripts included in wb_TRS and T2D**

418 Gene transcripts (n=19) with no validated IVs were removed from TWMR analysis.

419 Among the 125 remaining gene transcripts, 6 gene transcripts (GCC1-PAX4,

420 AKR1A1, AAMP, RAB1A, ITGA1, and ZNF639) were causally linked to T2D risk in

421 AGEN and 1 gene transcript (RAB1A) was successfully validated in DIAGRAM

422 (**Table S5**, $P < 0.0004 = 0.05/125$, where 125 corresponds to the number of gene

423 transcripts tested in TWMR). Specifically, RAB1A enriched in mTOCR1 signaling is

424 causally associated with 38% (95% CI: 1.18-1.62, $P=8.0 \times 10^{-5}$) and 47% (95% CI:

425 1.21-1.77, $P=5.4 \times 10^{-5}$) increase of T2D risk in AGEN and DIAGRAM, respectively.

426 GCC1-RAX4 converted 17% decrease of T2D risk (95% CI: 0.80-0.88, $P=5.2 \times 10^{-15}$)

427 in AGEN.

428

429 **Discussion**

430 In this study, we developed a wb-TRS for T2D risk in a longitudinal community-

431 based cohort in which the participants were middle aged and elderly and maintained
432 nonobese for up to 10 years. The wb-TRS was positively associated with risk of T2D
433 and could reflect dynamic change in glucose metabolism such as fasting and 2h-post
434 loading glucose, HbA1c, and beta cell dysfunction in the observational cohort.
435 Furthermore, the baseline wb-TRS well associated with the changes of glucose
436 metabolism measured in the continuous glucose monitoring in response to the 6-
437 month dietary intervention in a panel of younger and overweight participants. The
438 wb-TRS performed better and was a complementarity to fasting glucose, triglycerides
439 and the demographic and anthropometric factors in predicting T2D. The wb-TRS
440 supported several gene transcripts (RAB1A and GCC1-RAX4) and biological
441 pathways (oxidative phosphorylation, cholesterol metabolism, and mTORC1
442 signaling) that might involve in the etiology of nonobese T2D.

443 To the best of our knowledge, the present study is the first to examine the effects of
444 wb-TRS on nonobese T2D risk and dynamic change of related cardiometabolic traits
445 by using peripheral blood transcriptional markers. Recently, several studies proposed
446 the concept of TRS or polygenetic TRS, which used cumulative gene expression
447 information to provide an additional layer of biological interpretability and performed
448 better than both traditional risk factors and polygenic risk score for predicting Crohn's
449 disease, chronic obstructive pulmonary disease, and lung function [8, 26-28]. Our
450 study is in line with previous studies showing transcriptional variance is a reliable
451 intermediary mediating associations of genotypes with complex phenotypes [5, 6, 29]
452 and provides evidence of implement of blood transcriptional profile could yield a

453 promising better predictive effect on T2D and the dynamic changes in related
454 cardiometabolic traits.

455 In our study, the wb-TRS could well predict dynamic changes in glucose
456 metabolism in response to 6 months of 25% calorie restriction dietary intervention in
457 overweight/obese subjects [23]. In our trial (DPMH), a 25% calorie restriction
458 produced an appreciable weight reduction and improved glucose homeostasis [23].
459 Calorie restriction improves cardiometabolic risk including atherogenic dyslipidemia
460 (elevated triglycerides and fasting glucose, reduced HDL-C, diabetes, hypertension,
461 and large waist circumference) [30, 31]. Previous findings from the Comprehensive
462 Assessment of Long-term Effects of Reducing Intake of Energy (CALERIE™)
463 study, a 24-month, multicenter, randomized controlled calorie restriction trial, showed
464 that in young to middle aged, normal to slightly overweight adults, two years of >12%
465 caloric restriction improved traditional cardiometabolic risk factors within ranges
466 considered normal [31]. Recent study suggested that in healthy adults without obesity,
467 a 12% calorie restriction led to improved reducing inflammation and branched chain
468 amino acids and shifting lipoproteins from atherogenic to cholesterol transporting
469 [32]. We suspected that the genes and pathways included in the wb-TRS played key
470 roles in cardiometabolic beneficial effects in response to the calorie restriction dietary
471 intervention, which might help to explore the underlying mechanisms for the
472 beneficial effects.

473 Our results indicated several potential pathways and gene transcripts that might
474 involve in the etiology process of nonobese T2D, which suggested that transcriptional

475 information in blood could be capable of reflecting diseases risk and could bridge the
476 knowledge gap on how gene expression affects nonobese T2D. We found mTORC1
477 signaling directly regulates β -cell functional maturation and identity maintenance
478 [33]. Of note, by using TWMR analysis, we also demonstrated that the RAB1A gene
479 encoding a small GTPase, which is enriched in mTORC1 signaling, was positively
480 and causally associated with T2D risk in both east Asians and Europeans. Another
481 T2D susceptibility gene (GCC1-PAX4) was also causally associated with T2D risk in
482 east Asians [34]. PAX4 implicated in a Japanese case of maturity onset diabetes of the
483 young that is only 30kb further away from GCC1 is an outstanding candidate for T2D
484 given its involvement in pancreatic islet development [34]. This molecular
485 mechanism of nonobese T2D was consistent with that the nonobese Asian Indians
486 with young-onset diabetes have lower genetically determined beta cell function [35].
487 We identified oxidative phosphorylation and cholesterol metabolism pathways
488 involved in nonobese T2D. Of note, fatty acid oxidation gene (CPT1A) enriched in
489 oxidative phosphorylation pathway might mediate mitochondrial overload increasing
490 insulin resistance in muscle, lipid accumulation in liver, and lipotoxicity in pancreatic
491 beta cells thus involved in homeostasis of glucose metabolism [36, 37] . Cholesterol
492 efflux gene (ABCA1) enriched in cholesterol metabolism pathway might via its
493 downregulation to inhibit cholesterol transportation to HDL and excretion to bile and
494 faeces, and then increase lipids profile and deteriorates cardio-metabolism [38].
495 In Reactome analysis, cholesterol uptake suppression gene (MYLIP) enriched in
496 LXR β and LXR α regulate gene expression to limit cholesterol uptake pathway.

497 Inhibition of MYLIP increases cellular cholesterol accumulation and leads to
498 pancreatic β -cell dysfunction and insulin resistance [39]. We suspected that
499 cholesterol metabolism in liver and pancreas and beta cell dysfunction involved in
500 nonobese T2D but warrant further study in the future. Besides, the consistency of
501 transcriptional levels of these genes in whole-blood and specific tissue is another
502 interesting issue.

503 The major strengths of our current study lie in our adopted analytical approaches as
504 well as the well-characterized study cohorts. We adopted high-throughput RNA-seq
505 technology, a penalized regression model with 10-fold cross validation method to
506 develop the wb-TRS, and successfully validated its performance in an independent
507 external cohort, which guaranteed the statistical power and robustness of our findings
508 and conclusions. In terms of the study cohorts, our initial cohort included T2D whose
509 BMI maintained normal for up to 10 years, which are expected to better capture
510 underlying pathophysiology mechanism of nonobese T2D.

511 Our study, however, also has several limitations. First, transcriptome data were
512 measured only at 10 years follow-up visit in initial cohort, which could not establish a
513 causal relationship between the wb-TRS and nonobese T2D risk. It would also be
514 greater interesting to see whether transcriptional change would continue to be
515 associated with incidence risk of nonobese T2D. Second, complex diseases are final
516 phenotypes of transcriptional dysregulation in multiple tissues. Though an appreciable
517 proportion of genes expressed in tissues is expressed in peripheral whole blood [40], it
518 still should be better to interpret the present wb-TRS as one of potential essential

519 complementary prediction tools for nonobese T2D, rather than presenting an exact
520 pathophysiological process of it. Nevertheless, the individual's blood transcriptome
521 performed well in predicting tissue-specific expression and performed as well as
522 measured tissue expression in predicting complex diseases [5]. Third, we did not
523 incorporate metabolic markers in the analysis, which restrained us from comparing
524 the predictive effect between them. Jointly considering the genetic and metabolic
525 profiling and transcriptional level of alterations would provide more precise and
526 accurate evaluation of T2D risk and prediction efficacy. Last, both the initial and
527 validation cohort consisted of participants from east Asia, cautions should be
528 exercised to generalize the finding to other ethnicity groups.

529 In conclusion, we developed a wb-TRS that comprising a panel of genes based on
530 their blood expression levels, and that may involve important biological function
531 pathways was positively associated with nonobese T2D risk and predicted dynamic
532 change in glucose metabolic traits. The study supported that blood transcriptome
533 highly related to and reflect impaired glucose metabolism. It may lend insight into
534 discerning potential causal genes and biological mechanisms in the pathogenesis of
535 nonobese T2D.

536 **Acknowledgments**

537 We thank all the study participants for their participation and contribution.

538

539 **Data availability**

540 The eQTLs data used in this study are available on the eQTLGen Consortium website

541 [<https://eqtngen.org/>]. The summary statistics of genome-wide association study of T2D

542 was available in the Asian Genetic Epidemiology Network (AGEN)

543 [<https://blog.nus.edu.sg/agen/>] and the DIAbetes Genetics Replication And Meta-

544 analysis (DIAGRAM) [<http://www.diagram-consortium.org/downloads.html/>]. RNA-

545 seq related additional data reported in this manuscript are available from the

546 corresponding author upon reasonable request.

547

548 **Funding statement**

549 This work was funded by the National Natural Science Foundation of China

550 (82270859, 91957124, 91857205, 81930021, 81970728, 82070880 and 82088102),

551 the Shanghai Municipal Education Commission–Gaofeng Clinical Medicine Grant

552 Support (20152508 Round 2), the Shanghai Shenkang Hospital Development Center

553 (SHDC12019101, SHDC2020CR1001A, and SHDC2020CR3064B). MX, JW, ML,

554 TW, ZZ, RL, YX, JL, YB, WW, and GN are members of innovative research team of

555 high-level local universities in Shanghai.

556

557 **Duality of interest**

558 No potential conflicts of interest relevant to this article were reported.

559

560 **Contribution statement**

561 M.X., Y.B., and J.W. contributed to the conception and design of the study. Y.H.,
562 H.D., and N.C. performed statistical analysis and interpreted data. Y.H., and M.X.
563 drafted the manuscript. G.N. and W.W. revised the manuscript critically for important
564 intellectual content. Z.Z., Q.W., T.H., J.Z., T.W., M.L., H.L., S.W., R.Z., J.L., Y.X.,
565 Y.C., and R.L. contributed to the acquisition of data. All authors were involving in
566 revising the manuscript and had final approval of the version to be published. M.X.,
567 Y.B. and J.W. are the guarantors of this work and, as such, had full access to all the
568 data in the study and take responsibility for the integrity of the data and the accuracy
569 of the data analysis.

570 **References**

- 571 1. IDF Diabetes Atlas 2021 – 10th edition. www.diabetesatlas.org.
- 572 2. Hahn SJ, Kim S, Choi YS, Lee J, Kang J. Prediction of type 2 diabetes using
573 genome-wide polygenic risk score and metabolic profiles: A machine learning analysis
574 of population-based 10-year prospective cohort study. *EBioMedicine*. 2022;86:104383.
575 Epub 20221130. doi: 10.1016/j.ebiom.2022.104383. PubMed PMID: 36462406;
576 PubMed Central PMCID: PMC9713286.
- 577 3. Wang S, Li M, Lin H, Wang G, Xu Y, Zhao X, et al. Amino acids, microbiota-
578 related metabolites, and the risk of incident diabetes among normoglycemic Chinese
579 adults: Findings from the 4C study. *Cell Rep Med*. 2022;3(9):100727. Epub 20220822.
580 doi: 10.1016/j.xcrm.2022.100727. PubMed PMID: 35998626; PubMed Central
581 PMCID: PMC9512668.
- 582 4. Padilla-Martinez F, Wojciechowska G, Szczerbinski L, Kretowski A. Circulating
583 Nucleic Acid-Based Biomarkers of Type 2 Diabetes. *Int J Mol Sci*. 2021;23(1). Epub
584 20211228. doi: 10.3390/ijms23010295. PubMed PMID: 35008723; PubMed Central
585 PMCID: PMC9745431.
- 586 5. Mahashweta B, Kun W, Eytan R, Sridhar H. Predicting tissue-specific gene
587 expression from whole blood transcriptome. 2021;7:eabd6991.
- 588 6. Yao S, Zhang X, Zou SC, Zhu Y, Li B, Kuang WP, et al. A transcriptome-wide
589 association study identifies susceptibility genes for Parkinson's disease. *NPJ Parkinsons*
590 *Dis*. 2021;7(1):79. Epub 20210909. doi: 10.1038/s41531-021-00221-7. PubMed PMID:
591 34504106; PubMed Central PMCID: PMC97429416.

- 592 7. Wang B, Lunetta KL, Dupuis J, Lubitz SA, Trinquart L, Yao L, et al. Integrative
593 Omics Approach to Identifying Genes Associated With Atrial Fibrillation. *Circ Res.*
594 2020;126(3):350-60. Epub 20191205. doi: 10.1161/CIRCRESAHA.119.315179.
595 PubMed PMID: 31801406; PubMed Central PMCID: PMC7004281.
- 596 8. Moll M, Boueiz A, Ghosh AJ, Saferali A, Lee S, Xu Z, et al. Development of a
597 Blood-based Transcriptional Risk Score for Chronic Obstructive Pulmonary Disease.
598 *Am J Respir Crit Care Med.* 2022;205(2):161-70. doi: 10.1164/rccm.202107-1584OC.
599 PubMed PMID: 34739356; PubMed Central PMCID: PMC787248.
- 600 9. Li X, Su X, Liu J, Li H, Li M, andMe Research T, et al. Transcriptome-wide
601 association study identifies new susceptibility genes and pathways for depression.
602 *Transl Psychiatry.* 2021;11(1):306. Epub 20210521. doi: 10.1038/s41398-021-01411-
603 w. PubMed PMID: 34021117; PubMed Central PMCID: PMC78140098.
- 604 10. Teufel F, Seiglie JA, Geldsetzer P, Theilmann M, Marcus ME, Ebert C, et al. Body-
605 mass index and diabetes risk in 57 low-income and middle-income countries: a cross-
606 sectional study of nationally representative, individual-level data in 685 616 adults. *The*
607 *Lancet.* 2021;398(10296):238-48. doi: 10.1016/s0140-6736(21)00844-8.
- 608 11. Taylor R, Holman RR. Normal weight individuals who develop type 2 diabetes:
609 the personal fat threshold. *Clin Sci (Lond).* 2015;128(7):405-10. doi:
610 10.1042/CS20140553. PubMed PMID: 25515001.
- 611 12. Zou X, Zhou X, Zhu Z, Ji L. Novel subgroups of patients with adult-onset diabetes
612 in Chinese and US populations. *The Lancet Diabetes & Endocrinology.* 2019;7(1):9-11.
613 doi: 10.1016/s2213-8587(18)30316-4.

- 614 13. Zaharia OP, Strassburger K, Strom A, Bönhof GJ, Karusheva Y, Antoniou S, et al.
615 Risk of diabetes-associated diseases in subgroups of patients with recent-onset diabetes:
616 a 5-year follow-up study. *The Lancet Diabetes & Endocrinology*. 2019;7(9):684-94. doi:
617 10.1016/s2213-8587(19)30187-1.
- 618 14. Ahlqvist E, Storm P, Käräjämäki A, Martinell M, Dorkhan M, Carlsson A, et al.
619 Novel subgroups of adult-onset diabetes and their association with outcomes: a data-
620 driven cluster analysis of six variables. *The Lancet Diabetes & Endocrinology*.
621 2018;6(5):361-9. doi: 10.1016/s2213-8587(18)30051-2.
- 622 15. Wesolowska-Andersen A, Brorsson CA, Bizzotto R, Mari A, Tura A, Koivula R, et
623 al. Four groups of type 2 diabetes contribute to the etiological and clinical heterogeneity
624 in newly diagnosed individuals: An IMI DIRECT study. *Cell Rep Med*.
625 2022;3(1):100477. Epub 20220104. doi: 10.1016/j.xcrm.2021.100477. PubMed PMID:
626 35106505; PubMed Central PMCID: PMCPMC8784706.
- 627 16. Spracklen CN, Horikoshi M, Kim YJ, Lin K, Bragg F, Moon S, et al. Identification
628 of type 2 diabetes loci in 433,540 East Asian individuals. *Nature*. 2020;582(7811):240-
629 5. Epub 20200506. doi: 10.1038/s41586-020-2263-3. PubMed PMID: 32499647;
630 PubMed Central PMCID: PMCPMC7292783.
- 631 17. Wang B, Li M, Zhao Z, Lu J, Chen Y, Xu Y, et al. Urinary bisphenol A concentration
632 and glucose homeostasis in non-diabetic adults: a repeated-measures, longitudinal study.
633 *Diabetologia*. 2019;62(9):1591-600. doi: 10.1007/s00125-019-4898-x.
- 634 18. Hao M, Ding L, Xuan L, Wang T, Li M, Zhao Z, et al. Urinary bisphenol A
635 concentration and the risk of central obesity in Chinese adults: A prospective study. *J*

- 636 Diabetes. 2018;10(6):442-8. Epub 20170329. doi: 10.1111/1753-0407.12531. PubMed
637 PMID: 28097815.
- 638 19. Ning G, Bi Y, Wang T, Xu M, Xu Y, Huang Y, et al. Relationship of Urinary
639 Bisphenol A Concentration to Risk for Prevalent Type 2 Diabetes in Chinese Adults.
640 Annals of Internal Medicine. 2011 Sep 20;155(6):368-74.
- 641 20. American Diabetes A. Diagnosis and classification of diabetes mellitus. Diabetes
642 Care. 2010;33 Suppl 1:S62-9. doi: 10.2337/dc10-S062. PubMed PMID: 20042775;
643 PubMed Central PMCID: PMCPMC2797383.
- 644 21. Juergen C, Matthias M. 1D and 2D annotation enrichment: a statistical method
645 integrating quantitative proteomics with complementary high-throughput data. BMC
646 Bioinformatics. 2012;13(Suppl 16):S12.
- 647 22. Fabregat A, Sidiropoulos K, Viteri G, Forner O, Marin-Garcia P, Arnau V, et al.
648 Reactome pathway analysis: a high-performance in-memory approach. BMC
649 Bioinformatics. 2017;18(1):142. Epub 20170302. doi: 10.1186/s12859-017-1559-2.
650 PubMed PMID: 28249561; PubMed Central PMCID: PMCPMC5333408.
- 651 23. Luo Y, Wang J, Sun L, Gu W, Zong G, Song B, et al. Isocaloric-restricted
652 Mediterranean Diet and Chinese Diets High or Low in Plants in Adults With
653 Prediabetes. J Clin Endocrinol Metab. 2022;107(8):2216-27. doi:
654 10.1210/clinem/dgac303. PubMed PMID: 35579171; PubMed Central PMCID:
655 PMCPMC9282247.
- 656 24. Sladek R, Rocheleau G, Rung J, Dina C, Shen L, Serre D, et al. A genome-wide
657 association study identifies novel risk loci for type 2 diabetes. Nature.

- 658 2007;445(7130):881-5. Epub 20070211. doi: 10.1038/nature05616. PubMed PMID:
659 17293876.
- 660 25. Porcu E, Rueger S, Lepik K, e QC, Consortium B, Santoni FA, et al. Mendelian
661 randomization integrating GWAS and eQTL data reveals genetic determinants of
662 complex and clinical traits. *Nat Commun.* 2019;10(1):3300. Epub 20190724. doi:
663 10.1038/s41467-019-10936-0. PubMed PMID: 31341166; PubMed Central PMCID:
664 PMCPMC6656778.
- 665 26. Hu X, Qiao D, Kim W, Moll M, Balte PP, Lange LA, et al. Polygenic transcriptome
666 risk scores for COPD and lung function improve cross-ethnic portability of prediction
667 in the NHLBI TOPMed program. *Am J Hum Genet.* 2022;109(5):857-70. Epub
668 20220405. doi: 10.1016/j.ajhg.2022.03.007. PubMed PMID: 35385699; PubMed
669 Central PMCID: PMCPMC9118106.
- 670 27. Liang Y, Pividori M, Manichaikul A, Palmer AA, Cox NJ, Wheeler HE, et al.
671 Polygenic transcriptome risk scores (PTRS) can improve portability of polygenic risk
672 scores across ancestries. *Genome Biol.* 2022;23(1):23. Epub 20220113. doi:
673 10.1186/s13059-021-02591-w. PubMed PMID: 35027082; PubMed Central PMCID:
674 PMCPMC8759285.
- 675 28. Marigorta UM, Denson LA, Hyams JS, Mondal K, Prince J, Walters TD, et al.
676 Transcriptional risk scores link GWAS to eQTLs and predict complications in Crohn's
677 disease. *Nat Genet.* 2017;49(10):1517-21. Epub 20170814. doi: 10.1038/ng.3936.
678 PubMed PMID: 28805827; PubMed Central PMCID: PMCPMC5745037.
- 679 29. Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, et al.

- 680 Candidate causal regulatory effects by integration of expression QTLs with complex
681 trait genetic associations. *PLoS Genet.* 2010;6(4):e1000895. Epub 20100401. doi:
682 10.1371/journal.pgen.1000895. PubMed PMID: 20369022; PubMed Central PMCID:
683 PMCPMC2848550.
- 684 30. Luigi F, Timothy E MS, Klein., John O H. Long-term calorie restriction is highly
685 effective in reducing the risk for atherosclerosis in humans. *Proc Natl Acad Sci USA.*
686 2004;101 (17):6659–63.
- 687 31. Kraus WE, Bhapkar M, Huffman KM, Pieper CF, Krupa Das S, Redman LM, et al.
688 2 years of calorie restriction and cardiometabolic risk (CALERIE): exploratory
689 outcomes of a multicentre, phase 2, randomised controlled trial. *The Lancet Diabetes*
690 *& Endocrinology.* 2019;7(9):673-83. doi: 10.1016/s2213-8587(19)30151-2.
- 691 32. Huffman KM, Parker DC, Bhapkar M, Racette SB, Martin CK, Redman LM, et al.
692 Calorie restriction improves lipid-related emerging cardiometabolic risk factors in
693 healthy adults without obesity: Distinct influences of BMI and sex from CALERIE a
694 multicentre, phase 2, randomised controlled trial. *EClinicalMedicine.* 2022;43:101261.
695 Epub 20220103. doi: 10.1016/j.eclinm.2021.101261. PubMed PMID: 35028547;
696 PubMed Central PMCID: PMCPMC8741476.
- 697 33. Ni Q, Sun J, Wang Y, Wang Y, Liu J, Ning G, et al. mTORC1 is required for
698 epigenetic silencing during beta-cell functional maturation. *Mol Metab.*
699 2022;64:101559. Epub 20220805. doi: 10.1016/j.molmet.2022.101559. PubMed PMID:
700 35940555; PubMed Central PMCID: PMCPMC9418906.
- 701 34. Cho YS, Chen CH, Hu C, Long J, Ong RT, Sim X, et al. Meta-analysis of genome-

- 702 wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat*
703 *Genet.* 2011;44(1):67-72. Epub 20111211. doi: 10.1038/ng.1019. PubMed PMID:
704 22158537; PubMed Central PMCID: PMCPMC3582398.
- 705 35. Siddiqui MK, Anjana RM, Dawed AY, Martoeau C, Srinivasan S, Saravanan J, et
706 al. Young-onset diabetes in Asian Indians is associated with lower measured and
707 genetically determined beta cell function. *Diabetologia.* 2022;65(6):973-83. Epub
708 20220305. doi: 10.1007/s00125-022-05671-z. PubMed PMID: 35247066; PubMed
709 Central PMCID: PMCPMC9076730.
- 710 36. Lai CQ, Parnell LD, Smith CE, Guo T, Sayols-Baixeras S, Aslibekyan S, et al.
711 Carbohydrate and fat intake associated with risk of metabolic diseases through
712 epigenetics of CPT1A. *Am J Clin Nutr.* 2020;112(5):1200-11. doi:
713 10.1093/ajcn/nqaa233. PubMed PMID: 32930325; PubMed Central PMCID:
714 PMCPMC7657341.
- 715 37. Meng ZX, Yin Y, Lv JH, Sha M, Lin Y, Gao L, et al. Aberrant activation of liver X
716 receptors impairs pancreatic beta cell function through upregulation of sterol regulatory
717 element-binding protein 1c in mouse islets and rodent cell lines. *Diabetologia.*
718 2012;55(6):1733-44. Epub 20120314. doi: 10.1007/s00125-012-2516-2. PubMed
719 PMID: 22415588.
- 720 38. Wang JQ, Li LL, Hu A, Deng G, Wei J, Li YF, et al. Inhibition of ASGR1 decreases
721 lipid levels by promoting cholesterol excretion. *Nature.* 2022;608(7922):413-20. Epub
722 20220803. doi: 10.1038/s41586-022-05006-3. PubMed PMID: 35922515.
- 723 39. Hao M, Head WS, Gunawardana SC, Hasty AH, Piston DW. Direct effect of

724 cholesterol on insulin secretion: a novel mechanism for pancreatic beta-cell dysfunction.
725 Diabetes. 2007;56(9):2328-38. Epub 20070615. doi: 10.2337/db07-0056. PubMed
726 PMID: 17575085.

727 40. Radzikowska U, Ding M, Tan G, Zhakparov D, Peng Y, Wawrzyniak P, et al.
728 Distribution of ACE2, CD147, CD26, and other SARS-CoV-2 associated molecules in
729 tissues and immune cells in health and in asthma, COPD, obesity, hypertension, and
730 COVID-19 risk factors. Allergy. 2020;75(11):2829-45. Epub 20200824. doi:
731 10.1111/all.14429. PubMed PMID: 32496587; PubMed Central PMCID:
732 PMCPMC7300910.

733

734 **Supporting information**

735 **Table S1.** Clinical characteristics of study participants.

736 **Table S2.** Weights of 144 gene transcripts included in wb-TRS, ordered from most
737 negative to most positive associated with nonobese type 2 diabetes.

738 **Table S3.** Receiver operator characteristic curves for prediction of nonobese type 2
739 diabetes.

740 **Table S4.** Top30 Reactome pathways associated with genes in wb-TRS.

741 **Table S5.** Transcriptome wide Mendelian randomization shows the causal effect of
742 gene transcripts included in wb_TRS on type 2 diabetes.

743 **Fig S1.** Flowchart of study participants.

744 **Fig S2.** Cross-validation results for least absolute shrinkage and selection operator
745 (LASSO) logistics-model for nonobese type 2 diabetes in training dataset.

746 **Fig S3.** Violin plot of wb-TRS and gene transcripts most negatively (FUOM) and
747 positively (RPF1) associated with nonobese type 2 diabetes.

748 **Fig S4.** Stratified analysis. Data are odds ratio (OR) and 95% confidence interval (CI)
749 calculated using multivariable logistic regression models. The results were adjusted for
750 age, sex, body mass index, family history of diabetes, waist circumference, smoking
751 and drinking status, education level, physical activity, systolic and diastolic blood
752 pressure, total cholesterol, triglycerides, high-density and low-density lipoprotein
753 cholesterol.

754 **Figure Legends**

755 **Fig 1. The schematic study design.** At the final follow-up visit (10-year's), the
756 whole blood samples of the participants with BMI <25kg/m² during the past 10 years
757 were collected and underwent RNA sequencing (RNA-seq). The 1105 nonobese
758 participants with RNA-seq data were randomly split into training (75%) and testing
759 (25%) set. Transcripts with gene expression level >1 CPM in 99% of T2D cases or
760 non-T2D controls and with a nominal difference P value <0.01 between these two
761 groups in the wb-TRS features screening progress. In training dataset, we trained the
762 LASSO logistic-model and adopted 10-fold cross-validation to conduct the
763 transcriptional risk score (wb-TRS). In testing and external dataset, we tested the
764 performance of wb-TRS. Finally, Mitch analysis and Reactome pathway analysis
765 were performed to identify key genes and pathways. Abbreviations: T2D, type 2
766 diabetes; BMI, body mass index; CPM, counts per million; LASSO, least absolute
767 shrinkage and selection operator; TRS, transcriptional risk score; Mitch analysis,
768 multi-contrast gene set enrichment analysis.

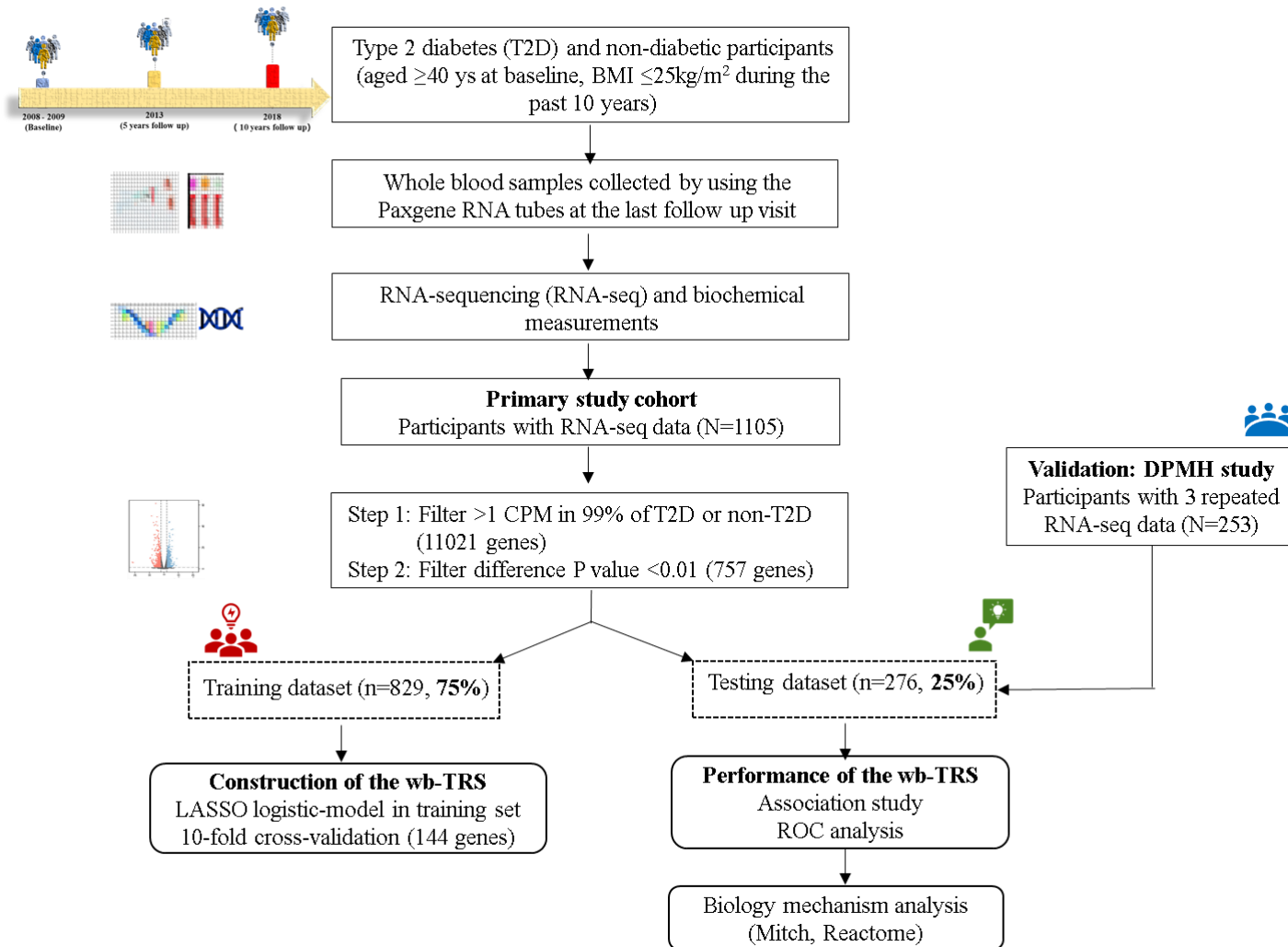
769 **Fig 2. Restricted cubic spline analysis for associations between the wb-TRS and**
770 **cardiometabolic traits in testing dataset.** Data were adjusted for age, sex, body
771 mass index, family history of diabetes, waist circumference, smoking and drinking
772 status, education level, physical activity, systolic and diastolic blood pressure,
773 triglycerides, low-density lipoprotein cholesterol and anti-diabetic therapy.

774 **Fig 3. Receiver operator characteristic curves for prediction of type 2 diabetes.**
775 Primary model comprised age, sex, BMI and family history of diabetes.

776 Abbreviations: wb-TRS, whole blood-based transcriptional risk score; BMI, body
777 mass index; FBG, fasting glucose index; TG, triglycerides; and AUC, area under the
778 curve.

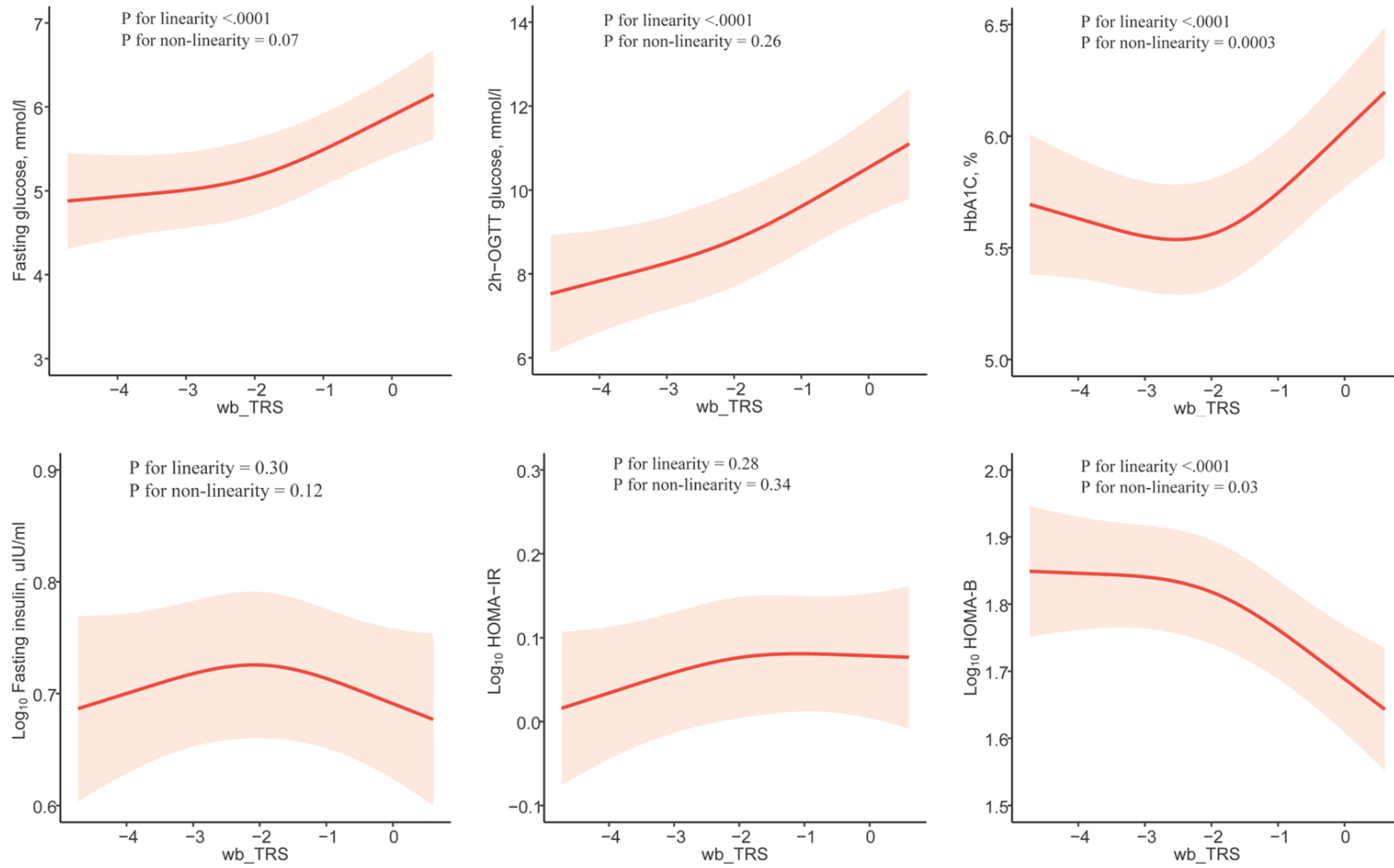
779 **Fig 4. Mitch analysis.** The gene expression level in the top and bottom tertile of the
780 wb-TRS were compared with those in the middle tertile. **A**, a heatmap of enrichment
781 scores for gene expression contrasts between these two comparisons with hierarchical
782 clustering. Cells in red indicate a positive enrichment score, whereas blue indicated a
783 negative enrichment score. **B**, the significant metabolic pathways. Abbreviations:
784 botVmid, bottom versus middle tertile; topVmid, top versus middle tertile.
785

786 **Fig 1.**



787

788 **Fig 2.**

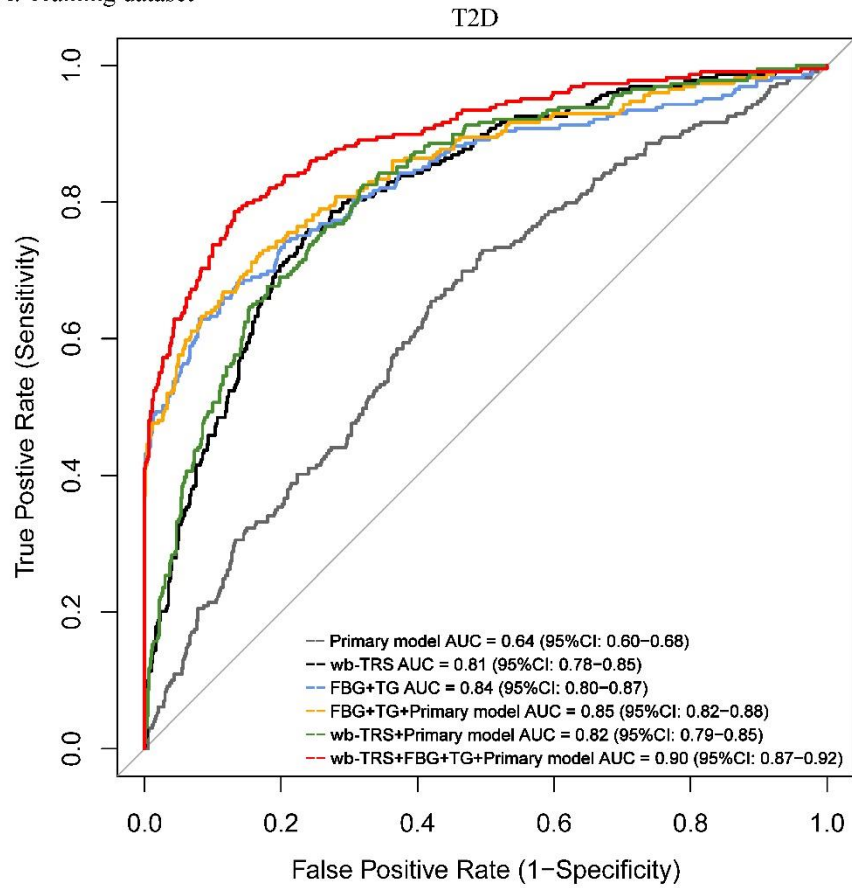


789

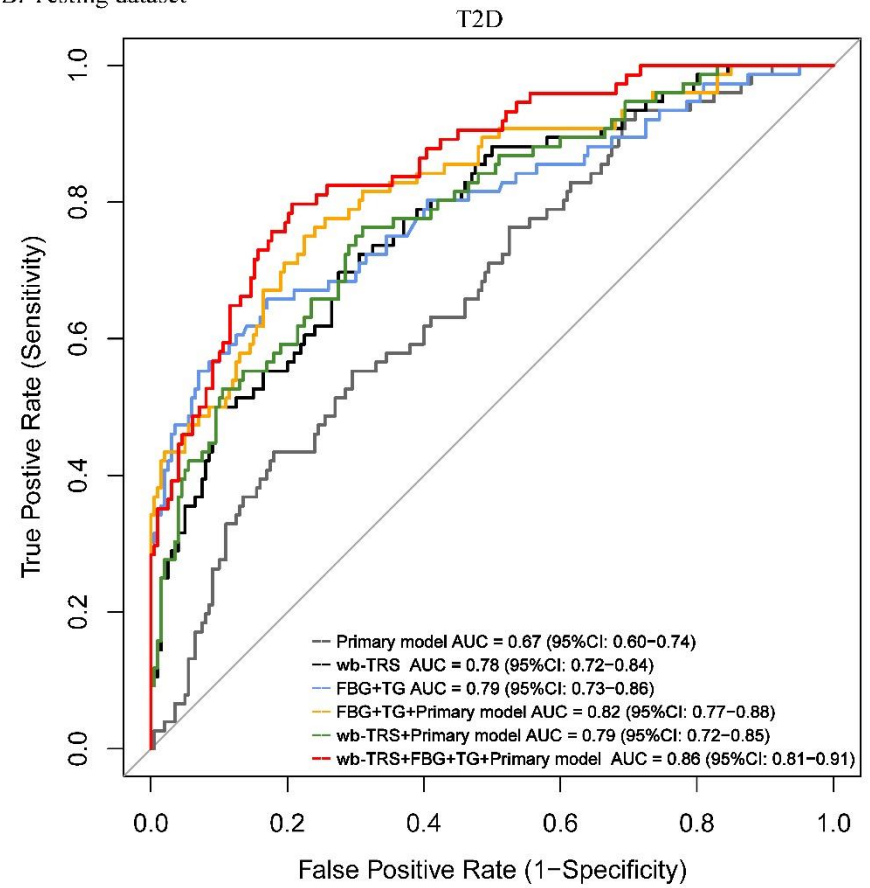
790

Fig 3.

A. Training dataset



B. Testing dataset



791

792

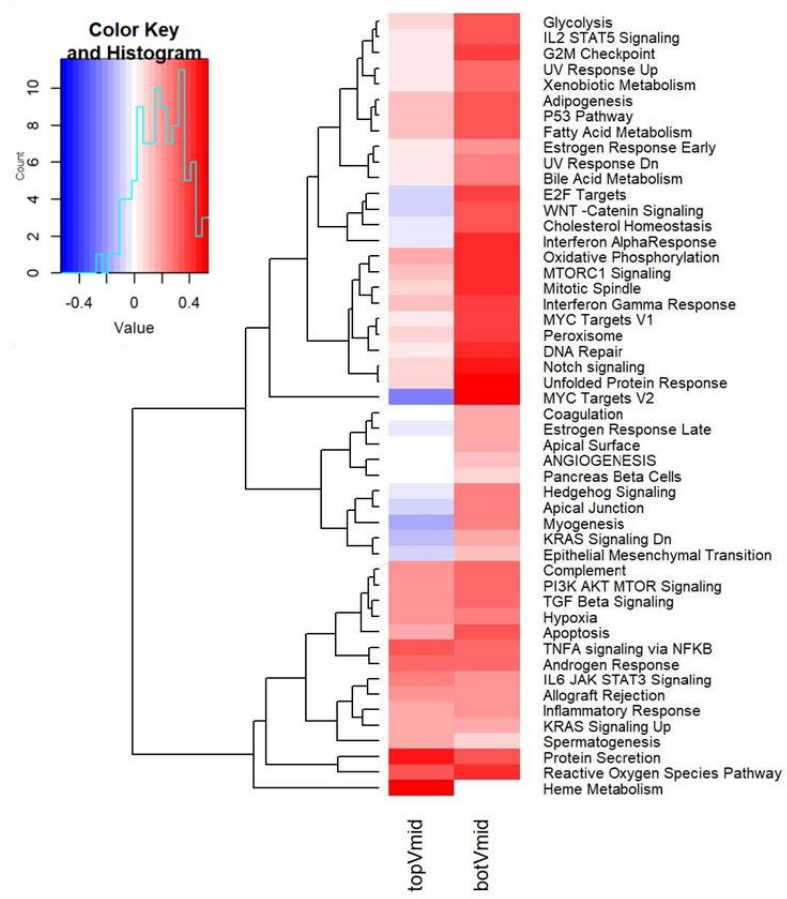
793

794

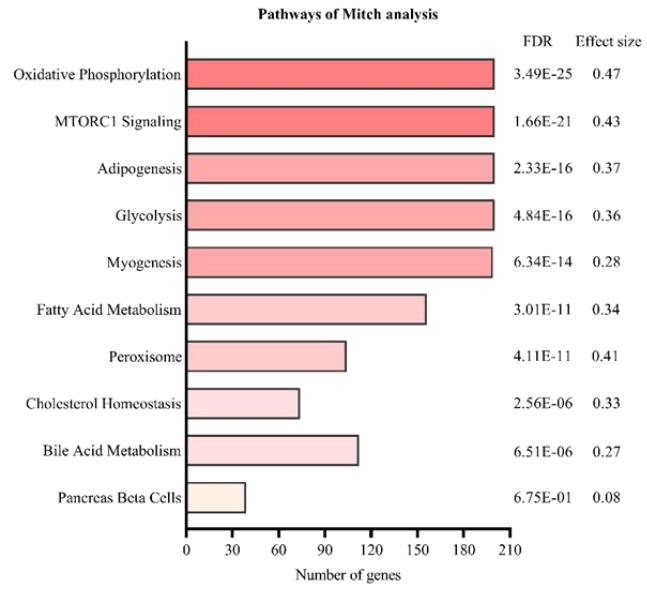
795

Fig 4.

A.



B.



796

797

Table 1. Adjusted ORs (95% CI) of type 2 diabetes associated with whole blood-based transcriptional risk score (wb-TRS) in testing dataset.

	Tertile 1(-7.0 ~ -2.5)	Tertile 2 (-2.5 ~ -1.4)		Tertile 3 (-1.4 ~ -3.9)		P _{for trend}	Each 1-unit	
	/	OR (95% CI)	P value	OR (95% CI)	P value		OR (95% CI)	P value
Type 2 diabetes								
Cases, n (%)	8 (9.0)	21 (23.3)		47 (48.5)			76 (27.5)	
Model 1	1.00 (Ref.)	2.57 (1.05-6.30)	0.04	8.36 (3.56-19.7)	<.0001	<.0001	2.49 (1.84-3.36)	<.0001
Model 2	1.00 (Ref.)	2.46 (0.95-6.35)	0.06	8.68 (3.51-21.5)	<.0001	<.0001	2.57 (1.86-3.56)	<.0001
Newly diagnosed type 2 diabetes								
Cases, n (%)	2 (2.3)	13 (14.4)		19 (19.6)			34 (12.3)	
Model 1	1.00 (Ref.)	6.41 (1.38-29.9)	0.01	17.2 (3.72-79.8)	0.0003	<.0001	2.87 (1.84-4.48)	<.0001
Model 2	1.00 (Ref.)	6.77 (1.30-35.3)	0.02	20.6 (3.94-108)	0.0003	0.001	3.40 (2.00-5.81)	<.0001
Type 2 diabetes with duration ≥5 years								
Cases, n (%)	6 (6.7)	7 (7.8)		27 (27.8)			40 (14.5)	
Model 1	1.00 (Ref.)	1.08 (0.33-3.58)	0.90	5.74 (2.09-15.8)	0.001	0.0002	2.32 (1.62-3.34)	<.0001
Model 2	1.00 (Ref.)	1.17 (0.33-4.24)	0.81	6.12 (2.00-18.7)	0.002	0.0001	2.33 (1.56-3.48)	<.0001

Data were odds ratio (OR), 95% confidence interval (CI) calculated from the multivariable logistic regression models. Model 1: adjusted for age, sex, body mass index, and family history of diabetes; Model 2: further adjusted for waist circumference, smoking and drinking status, education

level, physical activity, systolic and diastolic blood pressure, total cholesterol, triglycerides, HDL-C, and LDL-C. Abbreviation: HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol.

Table 2. Association of each 1-unit increment in the wb_TRS with cardiometabolic traits (n=1105)

	Training dataset		Testing dataset	
	β (SE)	<i>P</i> value	β (SE)	<i>P</i> value
Body mass index, kg/m ²				
Model 1	0.09 (0.06)	0.12	0.08 (0.10)	0.43
Model 2	-0.05 (0.04)	0.22	-0.003 (0.08)	0.97
Waist circumference, cm				
Model 1	0.38 (0.14)	0.01	0.14 (0.23)	0.53
Model 2	0.35 (0.14)	0.01	0.20 (0.24)	0.41
Systolic blood pressure, mmHg				
Model 1	1.52 (0.49)	0.002	0.80 (0.82)	0.33
Model 2	1.00 (0.39)	0.01	0.45 (0.65)	0.49
Diastolic blood pressure, mmHg				
Model 1	0.41 (0.28)	0.14	0.02 (0.47)	0.97
Model 2	-0.07 (0.22)	0.77	-0.35 (0.37)	0.36
Total cholesterol, mmol/l				
Model 1	-0.06 (0.03)	0.03	0.11 (0.05)	0.04
Model 2	-0.01 (0.01)	0.22	-0.002 (0.01)	0.86
Triglycerides, mmol/l				
Model 1	0.09 (0.03)	0.001	0.06 (0.03)	0.08
Model 2	0.09 (0.03)	0.001	0.04 (0.03)	0.29
HDL-C, mmol/l				
Model 1	-0.02 (0.01)	0.01	0.02 (0.01)	0.25
Model 2	-0.01 (0.01)	0.43	0.01 (0.01)	0.43
LDL-C, mmol/l				
Model 1	-0.04 (0.02)	0.03	0.08 (0.04)	0.04
Model 2	-0.06 (0.02)	0.01	0.06 (0.04)	0.12
Fasting glucose, mmol/l				
Model 1	0.42 (0.03)	<.0001	0.34 (0.06)	<.0001
Model 2	0.25 (0.03)	<.0001	0.23 (0.06)	<.0001
2h-OGTT glucose, mmol/l				

Model 1	1.16 (0.08)	<.0001	0.84 (0.13)	<.0001
Model 2	1.01 (0.08)	<.0001	0.68 (0.14)	<.0001
HbA1c, %				
Model 1	0.21 (0.02)	<.0001	0.19 (0.04)	<.0001
Model 2	0.07 (0.02)	<.0001	0.10 (0.03)	0.002
Log Fasting insulin, uIU/ml				
Model 1	0.03 (0.005)	<.0001	0.0003 (0.01)	0.97
Model 2	0.01 (0.01)	0.02	-0.004 (0.01)	0.61
Log HOMA-IR				
Model 1	0.05 (0.01)	<.0001	0.02 (0.01)	0.02
Model 2	0.03 (0.01)	<.0001	0.01 (0.01)	0.24
Log HOMA- β				
Model 1	-0.04 (0.01)	<.0001	-0.05 (0.01)	<.0001
Model 2	-0.03 (0.01)	<.0001	-0.04 (0.01)	<.0001

Data were β and standard error (SE). Linear regression model was used to evaluate associations of the wb_TRS (independent) with cardiometabolic traits (dependent variables). Model 1, adjusted for age, sex, body mass index, and family history of diabetes; Model 2, further adjusted for waist circumference, smoking and drinking status, education level, physical activity, systolic and diastolic blood pressure, triglycerides, LDL-C and anti-diabetic therapy. Abbreviation: HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; OGTT, oral glucose tolerance test; HbA1c, glycated hemoglobin A1c. HOMA-IR, homeostasis model assessment of insulin resistance; and HOMA- β , homeostasis model assessment of β cell function.

Table 3. The effects of baseline wb-TRS on changes of glucose metabolic traits during the 6-month dietary intervention (n=253)

Glucose metabolic traits	At 3 months		At 6 months	
	β (SE)	P value	β (SE)	P value
OGTT				
Δ Fasting glucose, mmol/l	-0.01 (0.02)	0.72	0.01 (0.02)	0.64
Δ 0.5h-OGTT glucose, mmol/l	0.03 (0.05)	0.51	-0.0005 (0.06)	0.99
Δ 1h-OGTT glucose, mmol/l	0.03 (0.08)	0.74	0.12 (0.08)	0.15
Δ 2h-OGTT glucose, mmol/l	0.04 (0.07)	0.59	0.10 (0.07)	0.16
Continuous glucose monitoring (CGM)				
Δ Average glucose, mmol/l	-0.21 (0.45)	0.64	1.42 (0.55)	0.01
Δ Average daytime glucose, mg/dl	-0.26 (0.48)	0.59	1.55 (0.60)	0.01
Δ Average nighttime glucose, mg/dl	-0.16 (0.42)	0.71	1.09 (0.48)	0.03
Δ AUC glucose, mmol*h/l	0.21 (0.45)	0.65	1.41 (0.55)	0.01
Δ AUC daytime glucose, mmol*h/l	0.26 (0.48)	0.59	1.56 (0.60)	0.01
Δ AUC nighttime glucose, mmol*h/l	0.16 (0.42)	0.70	0.21 (0.45)	0.65
Δ Estimated HbA1c	-0.004 (0.02)	0.81	0.05 (0.02)	0.01
Δ Estimated daytime HbA1c	-0.01 (0.02)	0.55	0.06 (0.02)	0.01
Δ Estimated nighttime HbA1c	0.01 (0.01)	0.58	0.04 (0.02)	0.03

Data were change (β) and standard error (SE). P values were from the general linear models after adjusted for sex, baseline age, body mass index, waist circumference, smoking and drinking status, education level, physical activity, systolic and diastolic blood pressure, triglycerides, LDL-C and baseline values for the respective outcome trait. Change (Δ) level was calculated as the numeric value at follow-up visits minus that at baseline. Estimated HbA1c was calculated using the formula: (average

glucose+46.7)/28.7.