RESEARCH
Open Access

Genome-wide DNA methylation analysis of extreme phenotypes in the identification of novel epigenetic modifications in diabetic retinopathy

Shaopeng Yang††, Xiao Guo††, Weijing Cheng††, Ishith Seth2, Gabriella Bulloch2, Yifan Chen3, Xianwen Shang2, Zhuoting Zhu2, Wenyong Huang1 and Wei Wang1*

Abstract

Background: Aberrant epigenetic modifications such as DNA methylation may contribute to the pathogenesis of DR. We aimed at elucidating the role of novel DNA methylation modifications in diabetic retinopathy (DR) in patients with type 2 diabetes mellitus (T2DM) using an extreme phenotypic design.

Methods/results: Two consecutive studies were conducted. A cross-sectional study using an extreme phenotypic design was conducted to identify rare methylation modifications that might contribute to DR pathogenesis. A 2-year longitudinal nested case–control study was conducted to validate the results and assess whether these novel methylation modifications could be used as biomarkers for predicting DR onset. A large number of differentially methylated CpG sites were identified in the cross-sectional study, and two (cg12869254 and cg04026387) corresponding to known genes were replicated in the longitudinal study. Higher methylation of cg12869254 significantly correlated with macular RNFL thinning in the superior and nasal subregions, and that of cg04026387 correlated with reduced deep capillary plexus VD in the superior and inferior subregions after adjusting for covariates.

Conclusions: Cg12869254 and cg04026387 hypermethylation may complement the known risk factors that contribute to the pathogenesis of DR and as novel biomarkers for disease prediction.

Keywords: Diabetic retinopathy, Type 2 diabetes mellitus, DNA methylation, Extreme phenotypes, Biomarkers

Introduction

Diabetic retinopathy (DR) represents a microvascular end-organ complication of diabetes and is a leading cause of blindness in the working-age population [1]. DR has an insidious onset that delays early detection and treatment [2]. The traditional pathogenesis of DR is largely driven by risk factors such as hyperglycemia, hypertension, and dyslipidemia; however, they explain less than 12% of the risk of developing DR [3]. Growing evidence suggests a clear polygenic risk of developing DR, highlighting the need for new biomarkers for DR [4].

© The Author(s) 2022. 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/. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated in a credit line to the data.
Aberrant epigenetic modifications such as DNA methylation may contribute to the pathogenesis of DR [5, 6]. Maghbooli et al. [7] noted changes in peripheral blood DNA methylation in DR patients, suggesting the potential value of DNA methylation as a biomarker for DR. However, DNA methylation studies conducted on type 2 diabetes mellitus (T2DM) cohorts based on traditional epigenome-wide association study (EWAS) approaches have yielded unclear results, failing to identify epigenetic modifications reproducibly associated with DR [8–10]. The results of these studies were skewed due to the incomplete phenotypes of the patients included. Moreover, the traditional EWAS approaches used in previous studies are inherently limited to common epigenetic modifications, which means that rare changes with large effects will be missed. Extreme phenotypic design is an emerging study design with superior power for detecting rare variants and has been suggested for the screening and identification of DR markers [11–15]. Further, previous studies followed a cross-sectional design and lack of optical coherence tomography (OCT) and OCTA (OCT angiography) biomarkers, which limits causal inferences and interpretations of their results [8–10]. Lastly, existing studies did not collect patients’ systemic and ocular factors, specifically renal function and axial length (AL), which are known to be independently associated with DR onset and progression. [16–19]

In the present study, we hypothesized that the use of extreme phenotypic designs would greatly increase the power to identify candidate sites, thus providing the ability to search for rare DNA methylation modifications.

In the discovery stage, a large number of differentially methylated CpG sites (DMSs) in DR were identified from a cross-sectional study. A nested case–control study was conducted to validate the results and longitudinally assess whether aberrant methylation can be used as a biomarker to predict DR. To determine whether these candidate sites play a role in DR pathogenesis, the association between the methylation degree of these sites with macular retinal nerve fiber layer (RNFL) thickness and vessel density (VD) was analyzed. Since diabetic kidney disease often develops simultaneously with DR, we also assessed the potential association of these sites with renal function. [20]

**Methods**

**Study participants**

This cross-sectional and longitudinal nested case–control study was conducted at Zhongshan Ophthalmic Center (ZOC) in Guangzhou, China. The study protocol was approved by the institutional review board of ZOC, Sun Yat-sen University (2017KYPJ094). The study procedure followed the principles of the Declaration of Helsinki, and written informed consent was obtained from all the enrolled participants.

Forty-three T2DM patients with DR and 92 T2DM patients without DR at baseline were recruited from 2017 to 2019. All patients were diagnosed with T2DM according to the criteria of the American Diabetes Association [21]. DR was defined according to the revised Early Treatment Diabetic Retinopathy Study (ETDRS) grading scale based on 7-field retinal photography and OCT/A imaging. Patients were excluded from the study if they had (1) severe systemic diseases other than T2DM; (2) cognitive impairment, mental illness, or inability to cooperate with questionnaires or examinations; (3) best-corrected visual acuity (BCVA) less than 20/25; (4) intraocular pressure (IOP) greater than 21 mm Hg; (5) spherical equivalent greater than +3.0 diopters (D) or less than -6.0 D; (6) glaucoma, age-related macular degeneration, amblyopia, and other eye diseases; (7) history of previous intraocular surgery or any other ophthalmic treatment; (8) pterygium and severe refractive media clouding.

Among these patients, 10 were DR-free despite >20 years of T2DM. All 92 patients without DR at baseline were followed up for 2 years, and 24 of them had new-onset DR. Two consecutive studies were conducted. The cross-sectional study serves as the discovery stage, and the results were then validated in the longitudinal study. The cross-sectional study included (1) a control group (n = 10) diagnosed with T2DM for at least 20 years without any clinical signs or visual-related symptoms of DR, referred to as the NDR group, and (2) a case group (n = 10) of patients diagnosed with DR within 4 years of T2DM diagnosis matched 1:1 by age, sex, and glycated hemoglobin (HbA1c), referred to as the DR group. The longitudinal study enrolled patients without DR at baseline within 2 years of T2DM diagnosis and divided them into two groups according to DR outcome during the 2-year follow-up: (1) DR high-risk group (n = 10); patients with DR onset within 2 years, referred to as the incidence-DR group; and (2) DR stable group (n = 10); patients who remained NDR during the 2-year period matched similarly to the incidence-DR group, referred to as the stable-NDR group (Additional file 2: Fig. S1).

**Clinical examinations and SS-OCT/A image acquisition protocol**

We obtained a 2 mL fasting venous blood sample anticoagulated with disodium ethylenediaminetetraacetate from each participant and stored at −80 °C without repeated freezing and thawing. Questionnaires and systemic and ocular examinations were administered to all participants. Details are as follows. After dilatation, seven fundus photographs were obtained according to the ETDRS-7 protocol, using a digital fundus camera (Canon...
on the fovea. The transverse section images were auto-
scan density of 320 A-scans
B-scans and Angio Macula 3
scan mode with a scan density of 512 A-scans
bisulfite conversion using the EZ DNA methylation Gold
of genomic DNA from each sample was used for sodium
ples were stored at -20 degrees. Approximately 500 ng
spectrophotometer (Thermo Scientific). Qualified sam-
were extracted from fasting venous
DNA were extracted from
Whole blood DNA was extracted from
Genome‑wide DNA methylation analysis
were present.
ual adjustments were performed if segmentation errors
segmentation were evaluated by an image expert, and the
DCP slabs were obtained. The results of the automated
lar systems were automatically segmented into four slabs,
were obtained. The en-face images of the retinal vascu-
ments of the RNFL thickness, including measurements
matically segmented into multiple layers, and measure-
ments of the superior, inferior, nasal, and temporal regions,
were obtained. The en-face images of the retinal microvascular
systems were automatically segmented into four slabs,
including the superficial capillary plexus (SCP) and deep
capillary plexus (DCP), and the VDs of each subregion
(superior, inferior, nasal, and temporal) in both SCP and
DCP slabs were obtained. The results of the automated
presentation of the results. Continuous variables
were presented as means ± standard deviations (SD),
and categorical variables were presented as counts (pro-
portions). Continuous variables were compared using
Student’s t test, and categorical variables were compared
using chi-square tests. To derive methylation levels, array
data were analyzed using the ChAMP package (V2.14.0).
Batch effects were corrected using the ComBat method.
The limma package (V3.40.6) was used to perform linear
regression and modified t tests, with age, sex, serum cre-
tinine, AL, and cellular composition included as covaria-
tes. Sites that met the significance level with $|\beta| \geq 0.1$
were used for further analyses. Gene Ontology (GO) and
Kyoto Encyclopedia of Genes and Genomes (KEGG)
enrichment analyses were performed to analyze the func-
tion of differentiated genes. The Benjamini–Hochberg

A commercial SS-OCT imaging system (Triton; Top-
con Inc., Tokyo, Japan) was used to obtain structural
OCT and OCTA images. The instrument has axial and
lateral resolutions of 8 and 20 µm, respectively, and uses
a wavelength of 1050 nm with a scan speed of 100,000
A-scans per second. Transverse section images of the
fundus and en-face images of the retinal microvascular
systems were obtained using a 3D Macula Cube 7×7 mm
scan mode with a scan density of 512 A-scans × 512
B-scans and Angio Macula 3×3 mm scan mode with a
scan density of 320 A-scans × 320 B-scans, both centered
on the fovea. The transverse section images were auto-
matically segmented into multiple layers, and measure-
ments of the superior, inferior, nasal, and temporal regions,
were obtained. The en-face images of the retinal vascu-
ar systems were automatically segmented into four slabs,
including the superficial capillary plexus (SCP) and deep
capillary plexus (DCP), and the VDs of each subregion
(superior, inferior, nasal, and temporal) in both SCP and
DCP slabs were obtained. The results of the automated
segmentation were evaluated by an image expert, and the
participants’ data were masked during processing. Man-
ual adjustments were performed if segmentation errors
were present.

Genome-wide DNA methylation analysis
Whole blood DNA was extracted from fasting venous
blood samples using the DNeasy Blood and Tissue Kit
(Qiagen, Hilden, Germany). The purity and concentra-
tion of the DNA were estimated using a NanoDrop 2000
spectrophotometer (Thermo Scientific). Qualified sam-
ple s were stored at -20 degrees. Approximately 500 ng
of genomic DNA from each sample was used for sodium
bisulfite conversion using the EZ DNA methylation Gold
Kit (Zymo Research, USA), following the manufactur-
er’s standard protocol. An Infinium MethylationEPIC
BeadChip Kit (Illumina Inc., USA) was used to perform
geno me-wide DNA methylation analyses. The distribu-
tion of cell types was inferred using the robust partial
correlation method based on DNA methylation signa-
tures of the constituent cell types.

Probe filtering was performed according to the manu-
facturer’s protocol to ensure valid analysis. Specifically,
all (1) probes with detection $p$ values ≥ 0.01, (2) probes
with a bead count less than 3 in more than 5% of sam-
ple s, (3) non-CpG probes, (4) multi-hit probes, (5) SNP-
related probes (within 5 bp of an SNP), and (6) probes
located in X and Y chromosomes were filtered out.

The experimental quality of each stage of the genome-
wide methylation analysis was carefully assessed using
internal reference samples. Staining controls were used
to examine the sensitivity and efficiency of the X-stain-
ing stage. The efficiency of single-base extension at the
X-staining stage was examined using extension controls.
Hybridization controls were used to examine the effi-
ciency of DNA hybridization. Target removal controls
were used to determine the efficiency of stripping off
DNA templates after the single-base extension phase.
Bisulfite conversion controls were used to evaluate the
conversion efficiency of sulfite to genomic DNA. Spe-
cific controls were used to detect non-specific exten-
sions of Infinium I and Infinium II probes. The overall
performance of the assay and the sample quality were
assessed using non-polymorphic controls. Finally, ran-
don sequences that did not hybridize with the target
DNA were used as negative controls. The $\beta$ value was
calculated and normalized for quantifying the degree of
methylation (Additional file 1: Methods).

Statistical analyses
R software (V4.1.0) was used for all data analyses and
the presentation of the results. Continuous variables
were presented as means ± standard deviations (SD),
and categorical variables were presented as counts (pro-
portions). Continuous variables were compared using
Student’s $t$ test, and categorical variables were compared
using chi-square tests. To derive methylation levels, array
data were analyzed using the ChAMP package (V2.14.0).
Batch effects were corrected using the ComBat method.
The limma package (V3.40.6) was used to perform linear
regression and modified $t$ tests, with age, sex, serum cre-
tinine, AL, and cellular composition included as covaria-
tes. Sites that met the significance level with $|\beta| \geq 0.1$
were used for further analyses. Gene Ontology (GO) and
Kyoto Encyclopedia of Genes and Genomes (KEGG)
enrichment analyses were performed to analyze the func-
tion of differentiated genes. The Benjamini–Hochberg
method was used to reduce the false discovery rate. The association between the degree of methylation of the identified sites and OCT/A measurements, renal function, and metabolic indexes was analyzed using multivariate linear regression. Two-sided p values less than 0.05 were considered statistically significant.

**Results**

**Characteristics of the study groups**

The demographics and characteristics of the subjects are summarized in Table 1. In the cross-sectional analysis, subjects from the DR and matched NDR groups shared similar characteristics, with the exception of BCVA ($P = 0.007$). Similarly, none of the systemic or ocular factors showed any significant differences between the incidence-DR and matched stable-DR groups in the prospective analysis (all $P > 0.05$). This indicated that the baseline characteristics of the control and case groups were well matched. The baseline characteristics of patients in the incidence-DR group in the prospective analysis were equally comparable to those of patients in the incidence-DR group in the cross-sectional analysis (all $P > 0.05$) (Additional file 2: Table S1).

**Table 1** Demographic and ocular characteristics of the study population

| Characteristics         | Cross-sectional study | P value* | Longitudinal study | P value |
|-------------------------|-----------------------|----------|--------------------|---------|
|                         | NDR group             | DR group | Stable-NDR group   | Incidence-DR group |
| Number of subjects      | 10                    | 10       | 10                 | 10      |
| Age, year               | 61.4 ± 3.2            | 64.2 ± 3.3 | 67.3 ± 0.3 | 66.3 ± 0.4 | 0.798 |
| Male, %                 | 5 (50%)               | 6 (60%)  | 3 (30%)            | 3 (30%) | 1.000 |
| BMI, kg/m²              | 26.0 ± 2.9            | 24.3 ± 2.5 | 24.2 ± 1.4 | 26.3 ± 4.2 | 0.152 |
| SBP, mm Hg              | 143.4 ± 15.0          | 139.4 ± 22.5 | 135.9 ± 10.9 | 139.1 ± 19.4 | 0.654 |
| DBP, mm Hg              | 73.2 ± 5.6            | 79.3 ± 13.9 | 72.9 ± 8.0 | 70.1 ± 9.5 | 0.486 |
| HbA1c, %                | 7.4 ± 1.5             | 6.8 ± 1.5 | 6.4 ± 0.7 | 6.9 ± 0.3 | 0.125 |
| TG, mmol/L              | 2.2 ± 1.0             | 2.7 ± 2.2 | 2.1 ± 1.1 | 2.1 ± 0.9 | 0.951 |
| TC, mmol/L              | 4.8 ± 0.7             | 4.7 ± 0.7 | 4.4 ± 1.1 | 4.8 ± 0.6 | 0.363 |
| LDL-c, mmol/L           | 3.0 ± 0.7             | 2.9 ± 0.8 | 2.9 ± 1.2 | 3.1 ± 0.8 | 0.611 |
| HDL-c, mmol/L           | 1.3 ± 0.4             | 1.2 ± 0.4 | 1.2 ± 0.3 | 1.5 ± 0.3 | 0.059 |
| Creatinine, µmol/L      | 69.5 ± 13.4           | 77.1 ± 21.6 | 66.4 ± 13.5 | 71.0 ± 14.3 | 0.480 |
| ACR, mg/g               | 6.7 ± 0.7             | 10.5 ± 2.1 | 8.7 ± 5.5 | 7.3 ± 1.2 | 0.494 |
| eGFR, mL/min †          | 92.5 ± 20.7           | 75.2 ± 19.2 | 74.8 ± 6.7 | 78.6 ± 6.2 | 0.686 |
| BCVA, logMAR            | 0.3 ± 0.1             | 0.2 ± 0.1 | 0.007             | 0.02 ± 0.1 | 0.883 |
| AL, mm                  | 23.7 ± 0.6            | 23.5 ± 1.5 | 23.4 ± 1.1 | 23.3 ± 0.7 | 0.803 |
| IOP, mmHg               | 16.2 ± 2.5            | 15.7 ± 2.7 | 14.1 ± 3.1 | 15.2 ± 2.3 | 0.397 |
| T2DM duration, year     | 220.0 ± 0.7           | 22.2 ± 0.4 | 2.0 ± 0.4 | 2.3 ± 0.5 | 0.930 |
| Insulin, %              | 3 (30%)               | 1 (10%) | 3 (10%) | 1 (10%) | 1.000 |
| Oral antihyperglycemic agents, % | 10 (100%) | 10 (100%) | 10 (100%) | 10 (100%) | 1.000 |

Data are expressed as the mean ± standard deviation or n (%). BMI = Body mass index; SBP = systolic blood pressure; DBP = diastolic blood pressure; HbA1c = glycated hemoglobin; TG = triglycerides; TC = total cholesterol; LDL-c = low-density lipoprotein cholesterol; HDL-c = high-density lipoprotein cholesterol; ACR = albumin creatinine ratio; eGFR = estimated glomerular filtration rate; BCVA = best-corrected visual acuity; AL = axial length; IOP = intraocular pressure; DM = diabetic mellitus; DR = diabetic retinopathy

*Bold indicates statistically significant
†The eGFR was calculated using the Cockcroft-Gault formula

**Genome-wide DNA methylation analysis**

Comparisons revealed significant differences in cellular composition, both in cross-sectional and in longitudinal studies ($P < 0.05$) (Additional file 2: Figs. S2–S5). The DNA quality was adequate with values of 1.8–2.0 for optical density (OD) 260/OD280. The whole analysis was reliable given that all analytical stages passed rigorous quality control (Additional file 2: Figs. S6–S13). The methylation profiles were compared separately in two comparison groups: DR group versus NDR group, and incidence-DR group versus stable-NDR group. A total of 311 and 413 novel DMSs were identified in the cross-sectional and longitudinal studies, respectively (Fig. 1a, b). The DMSs identified in the cross-sectional study were contained in 141 genes, including $\text{CACNA1C}$, $\text{COL19A1}$, $\text{FLT3}$, $\text{S100A13}$, $\text{ZDHHC23}$, and $\text{SLC25A21}$. The ten most significantly differentiated genes between the NDR and DR groups are listed in Table 2. Compared to the NDR group, 217 sites (69.8%) were hypomethylated and 94 sites (30.2%) were hypermethylated (Fig. 2). Among them, 59 sites were located in the promoter region, 112 in the body region, 7 in the $3'$-untranslated region ($3'$-UTR), 1 in the ExonBnd region, and 32 in the intergenic
Concerning CpG positions, 43 sites were located in CpG islands, 183 in the open sea, 10 in the shelf regions, and 75 in the shore regions (Additional file 2: Fig. S14A, B).

The longitudinal nested case–control study identified 197 novel genes, including MICAL1, NLGN1, CACNG2, ZDHHC23, and SLC25A21. The ten most significantly differentiated genes between the stable-NDR and incidence-DR groups are listed in Table 3. Compared to the stable-NDR group, 248 sites (60.0%) were hypomethylated and 165 sites (40.0%) were hypermethylated (Fig. 3). Further analysis of the DNA functional domains revealed that 102 sites were located in the promoter region, 151 in the body region, 6 in the 3'-UTR, 1 in the ExonBnd region, and 153 in the IGR.

Replicated DMSs in the longitudinal study
The two novel DMSs corresponding to known genes identified in the cross-sectional analysis were replicated in the longitudinal analysis. Cg12869254 and cg04026387, contained in the ZDHHC23 and SLC25A21 genes, were the two novel sites that showed significant methylation changes in the same direction in both cross-sectional (cg12869254: $\beta$ difference = 0.11, $P = 0.012$;

Table 2 Top 10 significant genes harboring differentially methylated CpG positions (DMPs) between NDR group and DR group

| DMS          | Gene     | CHR | FGD     | Cgi   | $\beta$ value | $P$ value | NDR group | DR group | Difference |
|--------------|----------|-----|---------|-------|---------------|-----------|-----------|----------|------------|
| cg05129572   | NEDD4    | 15  | Body    | Open sea | 0.596         | 0.430     | −0.166    | 5.41 $\times 10^{-5}$ |
| cg12507788   | MRPL4    | 19  | TSS1500 | Shore  | 0.759         | 0.627     | −0.133    | 7.48 $\times 10^{-5}$ |
| cg07469467   | APAF1    | 12  | Body    | Open sea | 0.381         | 0.140     | −0.241    | 1.52 $\times 10^{-4}$ |
| cg27535148   | COL19A1  | 6   | Body    | Open sea | 0.381         | 0.265     | −0.116    | 1.60 $\times 10^{-4}$ |
| cg06773116   | KCNIP4   | 4   | Body    | Open sea | 0.503         | 0.631     | 0.128     | 2.23 $\times 10^{-4}$ |
| cg13563074   | OVCH2    | 11  | TSS1500 | Open sea | 0.498         | 0.397     | −0.101    | 4.22 $\times 10^{-4}$ |
| cg02294028   | MDFIC    | 7   | Body    | Shore  | 0.427         | 0.308     | −0.119    | 4.26 $\times 10^{-4}$ |
| cg24063962   | LOC100130274 | 8    | TSS200 | Island  | 0.607         | 0.721     | 0.114     | 5.02 $\times 10^{-4}$ |
| cg02738594   | CAMK1G   | 1   | Body    | Open sea | 0.456         | 0.596     | 0.140     | 5.86 $\times 10^{-4}$ |
| cg11686436   | FAM83E   | 19  | Body    | Shore  | 0.824         | 0.713     | −0.112    | 6.33 $\times 10^{-4}$ |

DMP = differentially methylated CpG position; CHR = chromosome; FGD = functional genomic distribution; Cgi = CpG island; NDR = without any diabetic retinopathy; DR = diabetic retinopathy;
Fig. 2. Heat map illustrating differentially methylated sites of DR group and NDR group. Red and blue represent levels of hypermethylation and hypomethylation. DR: diabetic retinopathy; NDR: non-DR.

Table 3. Top 10 significant genes harboring DMP between stable-NDR group and incidence-DR group

| DMS         | Gene      | CHR | FGD      | Cgi      | $\beta$ value Stable-NDR group | $\beta$ value Incident-DR group | Difference $\beta$ | $P$ value     |
|-------------|-----------|-----|----------|----------|-------------------------------|-------------------------------|--------------------|--------------|
| cg05119883  | ZMIZ1-AS1 | 10  | Body     | Open sea | 0.539                         | 0.324                         | -0.215             | 2.61 x 10^{-5}|
| cg04145948  | BLM       | 15  | Body     | Open sea | 0.357                         | 0.516                         | 0.159              | 3.63 x 10^{-5}|
| cg19522093  | PPP5D1    | 19  | Body     | Open sea | 0.111                         | 0.215                         | 0.104              | 5.48 x 10^{-5}|
| cg10725855  | PCDH12    | 5   | 1stExon  | Shore    | 0.679                         | 0.816                         | 0.137              | 7.14 x 10^{-5}|
| cg17682313  | FBXW7     | 4   | S’UTR    | Open sea | 0.768                         | 0.209                         | -0.559             | 1.09 x 10^{-4}|
| cg17457090  | NDRG4     | 16  | TSS200   | Shore    | 0.439                         | 0.250                         | -0.189             | 1.66 x 10^{-4}|
| cg17666981  | TRIM27    | 6   | Body     | Open sea | 0.639                         | 0.806                         | 0.167              | 1.82 x 10^{-4}|
| cg24093474  | NDRG4     | 16  | TSS1500  | Shore    | 0.715                         | 0.515                         | -0.200             | 1.82 x 10^{-4}|
| cg17107388  | NDRG4     | 16  | TSS1500  | Shore    | 0.710                         | 0.445                         | -0.264             | 1.86 x 10^{-4}|
| cg01877814  | POLG      | 15  | TSS1500  | Shore    | 0.327                         | 0.218                         | -0.109             | 2.45 x 10^{-4}|

DMS = differentially methylated CpG position; CHR = chromosome; FGD = functional genomic distribution; Cgi = CpG island; NDR = without any diabetic retinopathy; DR = diabetic retinopathy.
Association of identified DMSs with OCT/A measurements

In the univariate linear model, the degree of methylation of cg12869254 was negatively correlated with macular RNFL thickness in the superior ($\beta = -2.6$, 95% CI $-5.1, -0.2$; $P = 0.039$) and nasal subregions ($\beta = -2.6$, 95% CI $-4.4, -0.7$; $P = 0.009$). The correlation remained stable in the multivariate model after adjusting for age, sex, T2DM duration, and HbA1c level ($\beta = -3.0$, 95% CI $-5.8, -0.2$; $P = 0.035$ and $\beta = -2.2$, 95% CI $-4.4, -0.1$; $P = 0.049$) (Table 4). The degree of methylation of cg04026387 was found to be negatively correlated with macular DCP VD in the superior ($\beta = -0.8$, 95% CI $-1.6, -0.1$; $P = 0.049$) and inferior subregions ($\beta = -1.4$, 95% CI $-2.4, -0.4$; $P = 0.008$) and positively correlated with SCP VD in the nasal subregion ($\beta = 1.9$, 95% CI 0.5, 3.3; $P = 0.011$) in the univariate linear model. The correlation remained stable in model adjusting for age, sex, T2DM duration, and HbA1c ($\beta = -0.8$, 95% CI $-1.7, -0.1$; $P = 0.039$, $\beta = -1.5$, 95% CI $-2.8, -0.2$; $P = 0.028$, and $\beta = 2.2$, 95% CI 1.1, 3.3; $P = 0.001$) (Table 5).

Association of identified DMSs with renal function and metabolic indexes

The degree of methylation of cg12869254 was significantly correlated with reduced eGFR ($\beta = -0.2$, 95% CI $-0.4, -0.01$; $P = 0.038$), and that of cg04026387 was significantly correlated with elevated ACR ($\beta = 0.7$, 95% CI 0.2, 1.1; $P = 0.014$), after adjusting for age, sex, T2DM duration, and HbA1c level (Additional file 2: Tables S2 and
Several metabolic indexes were also related to the methylation of these two sites (Additional file 2: Table S4, Additional file 1: Results).

**Discussion**

This study consisted of genome-wide DNA methylation analyses of extreme phenotypes within two consecutive studies that identified novel epigenetic modifications of DR. In the discovery stage, a large number of novel
candidate sites were identified. To validate the results and assess whether these methylation modifications could be used as biomarkers to predict DR, we analyzed DNA methylation in an independent longitudinal study. Our results suggest that aberrant methylation may predict DR, as hypermethylation at cg04026387 and cg12869254 identified in the cross-sectional study was successfully replicated in the longitudinal study. Finally, the methylation degree of these two sites negatively correlated with macular RNFL thickness and DCP VD, which further confirmed their potential for indicating DR [23, 24]. Therefore, methylation at these two sites may contribute to neural and vascular damage in the early stages of DR and serve as novel biomarkers for predicting the disease.

In this study, an extreme phenotypic design was adopted to identify the specific DMSs, which enabled us to identify novel methylation modifications, despite the relatively small yet sufficient sample size, as in studies of similar design [25–28]. Conversely, the results of previous studies on DR were skewed by the incomplete phenotypes of included participants, that is, healthy individuals and/or inclusion of individuals with diabetes diagnosis of <20 years as DR control groups, and/or DR patients with random DR-free T2DM duration as the DR case group. Current evidence suggests that an extreme phenotypic design is necessary for screening and identifying DR markers [29, 30]. The extreme phenotypic design is a recent concept that is used to increase homogeneity and better distinguish signals from noise in genome-wide analysis. Under this approach, the decreased variation in homogeneous samples results in a robust ability to identify signals, even in a relatively small number of samples. Concurrently, it enables the discovery of rare methylation alternations, which is another advantage of this design. Thus, we included patients diagnosed with T2DM for at least 20 years without any signs of DR as an extreme control group, and patients diagnosed with DR within 4 years of T2DM diagnosis were included as an extreme case group. As most patients typically develop DR within 17 years of T2DM diagnosis [11], while epigenetic changes typically initiate retinal impairment within a short period, the enrolled patients may represent extreme phenotypes enriched in “protective” and “risk” epigenetic alternations, respectively. Therefore, this model increased the candidate site identification in this study.

To our knowledge, this is the first study to longitudinally investigate the relationship between DNA methylation and T2DM in patients with DR. Thus far, all previous studies on T2DM patients have been cross-sectional, which limits causal inferences of their results. However, DNA methylation has been suggested to be dynamic rather than rigidly fixed in various diseases and aging process [31]. Therefore, it is important to clarify the time-ordered relationships between such aberrant methylation and DR pathogenesis. Our cross-sectional study revealed significant hypomethylation in sites within S100A13 in the gene-positive DR group, which is in agreement with the results obtained by Li et al. [8] Demethylation of this gene purportedly increases hyperglycemia-induced damage through calcium signaling and the RAGE pathway. However, these results were not replicated in the longitudinal study. Similarly, many other candidate sites identified in the cross-sectional discovery stage exhibited no significant differences in the longitudinal study. This highlights the limitations of previous cross-sectional studies in which the chronological relationship between exposure and outcome was unclear. The changes in methylation at these sites may be explained as concomitant consequences of DR development rather than as primary contributing factors to DR pathogenesis or progression. Cg12869254 and cg04026387 were successfully replicated in this longitudinal study, suggesting that methylation alterations at these two sites precede DR, supporting their role in the early pathogenesis of DR.

The two novel sites identified in this study are included in the ZDHHC23 and SLC25A21 genes, which have not been identified in previous studies. A recent animal study demonstrated the neurotoxic effects mediated by the downregulation of the ZDHHC23 gene [32]. Decreased ZDHHC23 levels dysregulated interactions between palmitoyl-protein thioesterase-1 and acyl protein thioesterase-1, leading to increased plasma membrane H-Ras and subsequent microglial proliferation in mouse brains. Activation of microglia leads to increased secretion of tumor necrosis factor-alpha, interleukin-6, monocyte chemoattractant protein-1, and complement component C1q, promoting the transformation of astrocytes to the neurotoxic A1 phenotype [33]. The astrocytes of the A1 phenotype not only lose their original neurotrophic role, but also actively produce neurotoxins that degrade neurons. Retinal neurodegeneration is considered an early component of DR, which can precede visible vasculopathy [23, 24]. In this study, the degree of methylation of cg12869254 negatively correlated with macular RNFL thickness in the superior and nasal subregions, while the RNFL thickness represented the axons of retinal ganglion cells. Our teams and other groups have demonstrated that reduced RNFL thickness was significantly associated with a higher risk of development and progression of DR [34]. Therefore, we hypothesize that ZDHHC23 gene-mediated neurotoxicity caused by hypermethylation of cg12869254 may also contribute to the retinal neurodegeneration in DR. Additional studies are warranted to further elucidate the potential role of ZDHHC23 in DR.
SLC25A21 is involved in intracellular organic acid catabolism and metabolism [35]. Within mitochondria, it transports 2-oxoadipate and 2-oxoglutarate to be converted into acetyl-CoA. Boczonadi et al. [36] found that the dysfunction of SLC25A21 may lead to impaired transport of 2-oxoadipate, which in turn disrupts the degradation of lysine and tryptophan, leading to the accumulation of 2-oxoadipate, piperidinic acid, and quinolinic acid. Overexposure to these cytotoxic substances leads to irreversible mitochondrial and tissue damage mediated by free radicals. Consistently, the methylation degree of cg04026387 locus negatively correlated with macular DCP VD in our study, suggesting that it is a more sensitive indicator than VD in other plexus [37]. Therefore, we hypothesize that similar mechanisms of mitochondrial damage and vascular damage may also occur in the pathogenesis of early DR. However, SCP VD in the nasal subregion appears to be increased with higher methylation of cg04026387. Similar results have been reciprocated in studies on rats and humans [38, 39]. One possible explanation for this is that the deeper capillary layer is more sensitive to mitochondrial damage. In the early stages of DR, the above-mentioned factors may act mainly on the DCP, whereas the SCP exhibits a compensatory flow increase secondary to the reduced DCP VD. However, further studies investigating these genes are required to validate our hypothesis.

DKD is another major microvascular complication of diabetes mellitus [40]. Given the commonalities in pathogenesis and frequent co-development of DR and DKD, we further analyzed the relationship between these two sites and renal function indexes [20, 41–43]. Hypermethylation at these two identified sites was significantly associated with elevated ACR and reduced eGFR, further increasing our confidence in the involvement of these two sites in diabetic microvascular complications. We hypothesize that inflammatory responses, mitochondrial dysfunction, and oxidative stress discussed above may also damage the endothelial cells of glomerular capillaries [41, 43–45], thus leading to urinary protein leakage and decreased glomerular filtration function.

To our knowledge, few studies have collected systemic and ocular factors, such as renal function and AL, while analyzing the molecular etiology of DR. These factors have been independently associated with DR onset and progression [16–19]. Thus, the differentiated sites identified in previous studies might have been confounded by the presence of these factors. This study obtained various systemic and ocular factors from the participants, and most characteristics were well matched in each comparison group. This increased the reliability of our results.

This study had certain limitations. First, since most patients typically develop DR within 17 years of T2DM diagnosis [11], patients diagnosed with T2DM for 20 years without DR were included in the gene-negative control (NDR) group. However, even if the odds are low, these patients may still develop DR in the future. Future longitudinal studies are needed to determine whether these protective factors are still effective when the diabetes duration is over 20 years. Second, the sample size of this study was relatively small, and therefore the results should be treated with care. However, the adopted extreme phenotypic design greatly increases genetic homogeneity, which provides a robust ability to distinguish signals from noise. Despite the relatively small sample size, two sites were successfully replicated in the longitudinal study and were correlated with OCT/A parameters.

Conclusion
In conclusion, using extreme phenotypes, this study identified a large number of novel methylation modifications of DR in the discovery stage from a cross-sectional study and further validated these results via a longitudinal study. Cg12869254 and cg04026387 were the two sites corresponding to known genes that were successfully replicated, and the degree of methylation at these sites was associated with macular RNFL and VD. These two novel sites may serve as complementary to the known risk factors that contribute to the pathogenesis of DR and as biomarkers to identify diabetic patients at high risk of incident DR. Monitoring these sites may facilitate risk stratification of the diabetic population and offer a potential tool for screening and therapeutic interventions.

Abbreviations
DR: Diabetic retinopathy; T2DM: Type 2 diabetes mellitus; EWAS: Epigenome-wide association study; OCT/A: Optical coherence tomography/angiography; AL: Axial length; DMS: Differentially methylated Cpg site; RNFL: Retinal nerve fiber layer; VD: Vessel density; HbA1c: Glycated hemoglobin; BMI: Body mass index; SCP: Superficial capillary plexus; DCP: Deep capillary plexus; GO: Gene Ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; OD: Optical density; 3′-UTR: 3′-Untranslated region; IGR: Intergenic region.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s13148-022-01354-2.

Additional file 1. Supplementary methods and results.
Additional file 2. Supplementary tables and figures.

Acknowledgements
The authors thank all the participants and staff in the ZOC preventive department.

Author contributions
ZZ, WH, and WW contributed to study concept and design; all authors performed acquisition, analyses, or interpretation; WW and SY drafted the manuscript; SY, IS, YC, and GB performed critical revision of the manuscript for
important intellectual content; WW and SYH performed statistical analyses; WW, WYH, and ZZ provided the funding; ZZ and WW provided the administrative, technical, or material support; WYH supervised the study. All authors read and approved the final manuscript.

Funding
This study was supported from the grants of the Guangzhou Science & Technology Plan of Guangdong Pearl River Talents Program (202102010162), Fundamental Research Funds of the State Key Laboratory of Ophthalmology (303060202400201209), and National Natural Science Foundation of China (82000091, 82171084).

Availability of data and materials
Data are available on reasonable request. Any requests for data can be made to the corresponding author and are subject to ethics approval.

Declarations

Ethics approval and consent to participate
This study was designed approved by the Institutional Review Board of Zhongshan Ophthalmic Centre (ZOC), Sun Yat-sen University, and is adhere to the tenets of the Declaration of Helsinki. Written informed consent was obtained from all subjects.

Competing interests
The authors declare that they have no competing interests.

Author details
1 State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangdong Provincial Key Laboratory of Ophthalmology and Visual Science, Guangdong Provincial Clinical Research Center for Ocular Diseases, Guangzhou, China. 2Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, University of Melbourne, Level 7, 32 Gisborne Street, East Melbourne, VIC 3002, Australia. 3John Radcliffe Hospital, Oxford University Hospitals NHS Foundation Trust, Oxford, UK.

Received: 21 August 2022   Accepted: 11 October 2022
Published online: 31 October 2022

References
1. Zheng Y, LEY SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nat Rev Endocrinol. 2018;14(2):88–98. https://doi.org/10.1038/nrendo.2017.151. 2. Vujosevic S, Aldington SJ, Silva P, et al. Screening for diabetic retinopathy: new perspectives and challenges. Lancet Diabetes Endocrinol. 2020;8(4):337–47. https://doi.org/10.1016/S2213-8587(19)30411-5.
3. Jampol LM, Glassman AR, Sun J. Evaluation and care of patients with diabetic retinopathy. N Engl J Med. 2020;382(17):1629–37. https://doi.org/10.1056/NEJMA1909637.
4. Cho H, Sobrin L. Genetics of diabetic retinopathy. Curr Diabetes Rep. 2014;14(8):515. https://doi.org/10.1007/s11892-014-0515-z.
5. Reddy MA, Zhang E, Natarajan R. Epigenetic mechanisms in diabetic complications and metabolic memory. Diabetologia. 2015;58(3):443–55. https://doi.org/10.1007/s00125-014-3462-y.
6. Corso-Diaz X, Jaeger C, Chaitankar V, Swaroop A. Epigenetic control of gene regulation during development and disease: a view from the retina. Prog Retin Eye Res. 2018;65:1–27. https://doi.org/10.1016/j.preteyeres.2018.03.002.
7. Maghbbooli Z, Hossein-Nezhad A, Larjani B, Amini M, Keshkpar A. Global DNA methylation as a possible biomarker for diabetic retinopathy. Diabetes Metab Res Rev. 2015;31(2):183–9. https://doi.org/10.1002/dmrr.2584.
8. Li T, Xu Y, Shi Y, et al. Genome-wide analysis of DNA methylation identifies 5100A13 as an epigenetic biomarker in individuals with chronic (≥30 year) type 2 diabetes without diabetic retinopathy. Clin Epigenetics. 2020;12(1):77. https://doi.org/10.1186/s13148-020-00871-z.
9. Chen H, Zhang X, Liao N, et al. Identification of NLRP3 inflammation-related gene promoter hypomethylation in diabetic retinopathy. Invest Ophthalmol Vis Sci. 2020;61(13):12. https://doi.org/10.1167/iovs.20-26132.
10. Miao A, Lu J, Wang Y, et al. Identification of the aberrantly methylated differentially expressed genes in proliferative diabetic retinopathy. Exp Eye Res. 2020;199:108141. https://doi.org/10.1016/j.exer.2020.108141.
11. Cabrera AP, Monickaraj F, Rangasamy S, Hobbs S, Mcguire P, Das A. Do genomic factors play a role in diabetic retinopathy? J Clin Med. 2020;9(1):216. https://doi.org/10.3390/jcm9010216.
12. Cabrero AP, Mankad RN, Marek L, et al. Genotypes and phenotypes: a search for influential genes in diabetic retinopathy. Int J Mol Sci. 2020;21(8):2712. https://doi.org/10.3390/ijms21082712.
13. Sharma A, Valle ML, Beveidge C, Liu Y, Sharma S. Unraveling the role of genetics in the pathogenesis of diabetic retinopathy. Eye (Lond). 2019;33(4):534–41. https://doi.org/10.1038/s41433-019-0337-y.
14. Kuo JZ, Wong TY, Rotter JI. Challenges in elucidating the genetics of diabetic retinopathy. JAMA Ophthalmol. 2014;132(1):96–107. https://doi.org/10.1001/jamaophthalmol.2013.5024.
15. Shitr C, Aldahmesh MA, Al-Dahmah S, et al. Exome-based case-control association study using extreme phenotype design reveals novel candidates with protective effect in diabetic retinopathy. Hum Genet. 2016;135(2):193–200. https://doi.org/10.1007/s00439-015-1624-8.
16. Wang W, He M, Gong X, et al. Association of renal function with retinal vessel density in patients with type 2 diabetes by using swept-source optical coherence tomographic angiography. Br J Ophthalmol. 2020;104(12):1768–73. https://doi.org/10.1136/bjophthalmol-2019-315450.
17. Hsieh YT, Tsai MJ, Tu ST, Hsieh MC. Association of abnormal renal profiles and proliferative diabetic retinopathy and diabetic macular Edema in an Asian population with type 2 diabetes. Jama Ophthalmol. 2018;136(1):68–74. https://doi.org/10.1001/jamaophthalmology.2017.5202.
18. Hainsworth DP, Gao X, Belbu I, et al. Refractive error and retinopathy outcomes in type 1 diabetes: the diabetes control and complications trial/epidemiology of diabetes interventions and complications study. Ophthalmology. 2021;128(4):554–60. https://doi.org/10.1016/j. jophtha.2020.09.014.
19. Man R, Gan A, Gupta P, et al. Is myopia associated with the incidence and progression of diabetic retinopathy? Am J Ophthalmol. 2019;208:226–33. https://doi.org/10.1016/j.ajo.2019.05.012.
20. Friedman EA, L’Esperance FJ. Diabetic renal-retinal syndrome: the prognosis improves. Arch Intern Med. 1980;140(9):1149–50.
21. Diagnosis and classification of diabetes mellitus. Diabetes Care. 2010;33 Suppl 1:S62-9. https://doi.org/10.2337/dc10-062.
22. Cockcroft DW, Gault MH. Prediction of creatinine clearance from serum creatinine. Nephron. 1976;16(1):31–41. https://doi.org/10.1159/000181580.
23. Jonsson KB, Frydkjær-Olsen U, Grauslund J. Vascular changes and neurodegeneration in the early stages of diabetic retinopathy: Which comes first? Ophthalmic Res. 2016;56(1):1–9. https://doi.org/10.1155/2004444998.
24. Carpineto P, Toto L, Aloia R, et al. Neuroretinal alterations in the early stages of diabetic retinopathy in patients with type 2 diabetes mellitus. Eye (Lond). 2016;30(5):673–9. https://doi.org/10.1038/eye.2016.13.
25. Das A, Rangasamy S, Naymik M, et al. Novel genetic variants in extreme phenotypes of diabetic retinopathy: DRGens study. Invest Ophthalm Vis Sci. 2018;59(9):1911–1911.
26. Ung C, Sanchez AV, Shen L, et al. Whole exome sequencing identification of novel candidate genes in patients with proliferative diabetic retinopathy. Vis Res. 2017;139:168–76. https://doi.org/10.1016/j.visres.2017.03.007.
27. Emond MJ, Louie T, Emerson J, et al. Exome sequencing of extreme phenotypes identifies DCTN4 as a modifier of chronic Pseudomonas aeruginosa infection in cystic fibrosis. Nat Genet. 2012;44(8):886–9. https://doi.org/10.1038/ng.2344.
28. Seyres D, Cabassi A, Lambourne JJ, et al. Transcriptional, epigenetic and metabolic signatures in cardiometabolic syndrome defined by extreme phenotypes. Clin Epigenetics. 2022;14(1):39. https://doi.org/10.1186/s13148-022-02157-z.
29. Shir C, Aldahmesh MA, Al-Dahmah S, et al. Exome-based case–control association study using extreme phenotype design reveals novel candidates with protective effect in diabetic retinopathy. Hum Genet. 2015;135(2):193–200. https://doi.org/10.1007/s00439-015-1624-8.
30. Li D, Lewinger JP, Gauderman WJ, Murray CE, Conti DV. Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. Genet Epidemiol. 2011;35(8):769–80. https://doi.org/10.1002/gepi.20628.
31. Luo C, Hajkova P, Ecker JR. Dynamic DNA methylation: in the right place at the right time. Science. 2018;361(6409):1336–40. https://doi.org/10.1126/science.aat6806.

32. Sadhukhan T, Bagh MB, Appu AP, et al. In a mouse model of INCL reduced S-palmitoylation of cytosolic thioesterase APT1 contributes to microglia proliferation and neuroinflammation. J Inherit Metab Dis. 2021;44(4):1051–69. https://doi.org/10.1002/jimd.12379.

33. Xu S, Lu J, Shao A, Zhang JH, Zhang J. Glial cells: role of the immune response in ischemic stroke. Front Immunol. 2020;11:1294. https://doi.org/10.3389/fimmu.2020.00294.

34. Gong X, Wang W, Xiong K, et al. Associations between peripapillary retinal nerve fiber layer and choroidal thickness with the development and progression of diabetic retinopathy. Invest Ophthalmol Vis Sci. 2022;63(2):7. https://doi.org/10.1167/iovs.63.2.7.

35. Fiorentino G, Dolce V, Palmieri L, et al. Identification of the human mitochondrial oxodicarboxylate carrier: bacterial expression, reconstitution, functional characterization, tissue distribution, and chromosomal location. J Biol Chem. 2001;276(11):8225–30. https://doi.org/10.1074/jbc.M009607200.

36. Boczonadi V, King MS, Smith AC, et al. Mitochondrial oxodicarboxylate carrier deficiency is associated with mitochondrial DNA depletion and spinal muscular atrophy-like disease. Genet Med. 2018;20(10):1224–35. https://doi.org/10.1038/gem.2017.251.

37. Ashraf M, Sampani K, Clermont A, et al. Vascular density of deep, intermediate and superficial vascularplexuses are differentially affected by diabetic retinopathy severity. Invest Ophthalmol Vis Sci. 2020;61(10):53. https://doi.org/10.1167/iovs.61.10.53.

38. Grunwald JE, Dupont J, Riva C. Retinal haemodynamics in patients with early diabetes mellitus. Br J Ophthalmol. 1996;80(4):327–31. https://doi.org/10.1136/bjo.80.4.327.

39. Cringle SJ, YuDY, Alder VA, Su EN. Retinal blood flow by hydrogen clearance polarography in the streptozotocin-induced diabetic rat. Invest Ophthalmol Vis Sci. 1993;34(5):1716–21.

40. Thomas MC, Brownlee M, Susztak K, et al. Diabetic kidney disease. Nat Rev Dis Primers. 2015;1:15018. https://doi.org/10.1038/nrdp.2015.18.

41. Brownlee M. Biochemistry and molecular cell biology of diabetic complications. Nature. 2001;414(6865):813–20. https://doi.org/10.1038/3514813a.

42. Hui Z, Chen YM, Gong WK, et al. Shared and specific biological signaling pathways for diabetic retinopathy, peripheral neuropathy and nephropathy by high-throughput sequencing analysis. Diab Vasc Dis Res. 2022;19(4):14791641221122918. https://doi.org/10.1177/14791641221122918.

43. Barrett EJ, Liu Z, Khamaisi M, et al. Diabetic microvascular disease: an endocrine society scientific statement. J Clin Endocrinol Metab. 2017;102(12):4343–410. https://doi.org/10.1210/jc.2017-01922.

44. Forbes JM, Thorburn DR. Mitochondrial dysfunction in diabetic kidney disease. Nat Rev Nephrol. 2018;14(5):291–312. https://doi.org/10.1038/nrneph.2018.9.

45. Reidy K, Kang HM, Hostetter T, Susztak K. Molecular mechanisms of diabetic kidney disease. J Clin Invest. 2014;124(6):2333–40. https://doi.org/10.1172/JCI72271.

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