Genetic regulation of pigment epithelium-derived factor (PEDF): an Exome-chip association analysis in Chinese subjects with type 2 diabetes

Running title: Genetic regulation of circulating PEDF

Chloe Y.Y. Cheung[1,2]*, Chi-Ho Lee[1]*, Clara S. Tang[3], Aimin Xu[1,2,4,5], Ka-Wing Au[1], Carol H.Y. Fong[1], Kelvin K.K. Ng[1], Kelvin H.M. Kwok[1], Wing-Sun Chow[1], Yu-Cho Woo[1], Michele M.A. Yuen[1], JoJo Hai[1], Kathryn C.B. Tan[1], Tai-Hing Lam[6], Hung-Fat Tse[1,7#], Pak-Chung Sham[8,9,10#], Karen S.L. Lam[1,2,4#]

[1]Department of Medicine, the University of Hong Kong, Hong Kong, China; [2]State Key Laboratory of Pharmaceutical Biotechnology, The University of Hong Kong, Hong Kong, China; [3]Department of Surgery, the University of Hong Kong, Hong Kong, China; [4]Research Center of Heart, Brain, Hormone and Healthy Aging, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; [5]Department of Pharmacology & Pharmacy, The University of Hong Kong, Hong Kong, China; [6]School of Public Health, the University of Hong Kong, Hong Kong, China; [7]Hong Kong-Guangdong Joint Laboratory on Stem Cell and Regenerative Medicine, the University of Hong Kong, Hong Kong, China; [8]Department of Psychiatry, the University of Hong Kong, Hong Kong, China; [9]Center for Genomic Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China; [10]State Key Laboratory in Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China.

*CYY Cheung and CH Lee contributed equally to this work and should be considered as co-first authors; #KSL Lam, PC Sham and HF Tse contributed equally to the supervision of this work and are co-corresponding authors.

Correspondence:

Prof Karen SL Lam: Department of Medicine, The University of Hong Kong, Queen Mary Hospital, 102 Pokfulam Road, Hong Kong, China; Phone: +852-2255-3348/ +852-2255-4783; Fax: +852-2816-2863; Email: ksllam@hku.hk

Prof Pak-Chung Sham: Centre for Genomic Sciences, The University of Hong Kong, 6/F, HKJC Building for Interdisciplinary Research, 5 Sassoon Road, Pokfulam Hong Kong, China; Phone: +852-2831-5425; Email: pcsham@hku.hk

Prof Hung-Fat Tse: Department of Medicine, The University of Hong Kong, Rm 1928, Block K, Queen Mary Hospital, 102 Pokfulam Road, Hong Kong, China; Phone: +852-2831-4694; Email: hftse@hkucc.hku.hk

Word count: 3955 words

Number of table: 6

Number of figure: 1

Number of supplementary table: 6

Number of supplementary figure: 3

Abstract (199 words)

Elevated circulating levels of pigment epithelium-derived factor (PEDF) have been reported in patients with type 2 diabetes (T2DM) and its associated microvascular complications. This study aimed to (i) identify the genetic determinants influencing circulating PEDF level in a clinical setting of T2DM; (ii) examine the relationship between circulating PEDF and diabetic complications; and (iii) explore the causal relationship between PEDF and diabetic complications. An exome-chip association study on circulating PEDF leve1 was conducted in 5385 Chinese subjects with T2DM. A meta-analysis of the association results of the discovery stage (n=2936) and replication stage (n=2449) was performed. The strongest association was detected at *SERPINF1* (p.Met72Thr; $P_{combined}$=2.06x10^{-57}; β[SE]:-0.33[0.02]). Two missense variants of *SMYD4* (p.Arg131Ile; $P_{combined}$=7.56x10^{-25}; β[SE]:0.21[0.02]) and *SERPINF2* (p.Arg33Trp; $P_{combined}$=8.22x10^{-10}; β[SE]:-0.15[0.02]) respectively, showed novel associations at genome-wide significance. Elevated circulating PEDF level was associated with increased risks of diabetic nephropathy (DN) and sight-threatening diabetic retinopathy (STDR). Mendelian randomization analysis showed suggestive evidence of a protective role of PEDF on STDR (*P*=0.085). Our study had provided new insights into the genetic regulation of PEDF and further support for its potential application as a biomarker for DN and STDR. Further studies to explore the causal relationship of PEDF with diabetic complications are warranted.

Introduction

Pigment epithelium-derived factor (PEDF) is a multifunctional glycoprotein that belongs to the serine protease inhibitor (serpin) superfamily (1). PEDF is widely distributed in multiple organs and tissues, such as kidney, eye, liver, lung and adipose tissues (1; 2). PEDF possesses diverse biological functions in different tissues, including anti-angiogenesis, retina protection, anti-inflammation, anti-fibrosis, stem cell renewal, neurogenesis and neuroprotection (1; 3-5). Previous studies reported that circulating PEDF level was significantly higher in patients with type 2 diabetes (T2DM) than in subjects without diabetes (6). Serum PEDF in T2DM patients was also shown to be elevated after treatment with metformin, an anti-hyperglycemic agent (7). On the other hand, PEDF has been suggested to play a protective role against diabetic microvascular damages (4; 8). Dysregulation of PEDF expression was reported to be involved in the pathogenesis of microvascular complications of diabetes (4; 8). Downregulation of PEDF was observed in ocular tissues of patients with diabetic retinopathy (DR) (9; 10) and the role of PEDF in diabetic nephropathy (DN) was suggested by the finding of reduced PEDF expression in kidneys of diabetic animals (4). PEDF has also been shown to suppress the expression of angiogenic, fibrogenic and proinflammatory factors (5; 11; 12), thereby contributing to the pathological changes in early DN.

Adipose tissue and liver have been considered the predominant sources of circulating PEDF (2). In contrast to the observations in tissues and organs, elevated circulating PEDF level has been reported in patients with DR and DN, in both type 1 diabetes (T1DM) and T2DM (13-16). In subjects with T2DM, we had also shown that increased serum PEDF level independently predicted nephropathy progression, in particular the development of

albuminuria among subjects with relatively well preserved renal function at baseline (16). It has been suggested that the increased circulating level of PEDF might represent a compensatory systemic response to the reduced PEDF expression in the diseased tissues and organs (1). Given the multifunctional properties of PEDF, genetic polymorphisms at or near the *PEDF* gene have been reported to be associated with various diseases, such as age-related macular degeneration (17), coronary artery disease (CAD) (18), overall adiposity and obesity-related insulin resistance (19). Furthermore, several studies have previously identified pathogenic mutations or common variants of the *SERPINF1* gene, which encodes the PEDF protein, as genetic determinants of circulating PEDF level (19-25). However, to date, no genome-wide or exome-wide association studies on circulating PEDF level have been published. Furthermore, no other study has evaluated the causal role of PEDF on the risk of diabetic complications using a Mendelian randomization approach. Therefore, we performed the current study to identify the genetic determinants of circulating PEDF level with an exome-chip association analysis using a custom Illumina HumanExome BeadChip (Asian Exome-chip) in Chinese subjects with T2DM. We then examined the relationship between circulating PEDF and diabetic complications, including DN, sight-threatening DR (STDR) and CAD; and explored the causal effect of PEDF on diabetic complications.

Research Design and Methods

This study consisted of an exome-chip association study evaluating the genetic determinants of circulating PEDF level in 5385 Chinese subjects with T2DM. The discovery stage involved 2936 subjects who had been examined in our previous exome-chip association studies (26-28), followed by a replication study in 2449 independent subjects with T2DM who had not been genotyped with the exome-chip. The cross-sectional associations of circulating PEDF levels and the PEDF-associated single nucleotide polymorphisms (SNPs)

with diabetic complications were then examined. Mendelian randomization analyses were also performed to investigate the causal role of PEDF on diabetic complications in the HKWDR cohort.

Subjects

All subjects were recruited from the Hong Kong West Diabetes Registry (HKWDR) consisted of unrelated Chinese subjects with T2DM regularly followed-up at the medical specialist clinics of the Hong Kong West Cluster. The study protocol was approved by the Institutional Review Board of the University of Hong Kong/Hospital Authority Hong Kong West Cluster. Written informed consent was obtained from all recruited subjects prior to any study related procedures. Details of the study cohort, definitions of DN, STDR and CAD, and measurement of serum PEDF level are described in the Supplementary Data.

Genotyping and data quality control

(i) Discovery stage:

All subjects were genotyped using a specially designed Asian Exome-chip at the Centre for Genomic Sciences of the University of Hong Kong. Details of the Asian Exome-chip had been described previously (26-28). Sample-level QC was carried out with regard to gender mismatch, duplication, biological relatedness, as well as possible sample contamination. A principal component (PC) analysis was performed using a panel of over 20,000 independent common single nucleotide polymorphisms (SNPs) (minor allele frequency [MAF] >0.05) for the detection of possible existence of non-Chinese samples, and outliers were removed. For SNP-level QC, variants with MAF <0.1%; or >2% missingness; or significantly deviated from Hardy-Weinberg Equilibrium (HWE) with P_{HWE} <1x10^{-5}; or originally designed for QC

purpose, were excluded from the analysis. After all QC measures, a total of 2936 subjects and 76,951 polymorphic SNPs (MAF ≥0.1%) remained in the exome-chip association analysis. In order to account for the between-SNP linkage disequilibrium (LD), the *P*-value informed LD-based clumping approach with the "--clump" command implemented in PLINK version 1.9 (29) was performed. Each index SNP had the strongest association *P*-value within each clumped region. The other variants within the same clumped region were in LD ($r^2 \geq 0.2$) with the index SNP and within ± 500kb from it.

(ii) Replication stage:

In the replication stage, all 8 index SNPs which achieved a $P_{discovery}<5x10^{-5}$ were selected for replication in another 2449 subjects with T2DM. Seven of the selected SNPs were genotyped by the MassARRAY Sequenome iPLEX Gold platform (San Diego, CA, USA) at the Center for Genomic Sciences of the University of Hong Kong. Genotyping of *SERPINF1* rs1136287 was conducted with the TaqMan pre-designed genotyping assay (Assay ID: C___1841779_20) according to the manufacturer's instructions. SNPs which showed a low genotyping call rate of less than 90%; or significantly deviated from HWE (P_{HWE} <0.006 [=0.05/8]), were excluded from further analysis. *TUBGCP6* rs76062207 was excluded from further analysis due to low genotyping call rate. Consequently, 7 SNPs were included in the replication analysis. The average genotyping call rate and concordance rate of these SNPs were 99.17% and 99.63%, respectively.

Statistical analysis

Exome-chip association analysis for single markers

All statistical analyses in the discovery and replication stages were performed with PLINK version 1.9 (29). Serum PEDF level was natural-logarithmically transformed before analysis.

In the discovery stage, single variant association analysis was carried out on the standardized residuals of PEDF level generated from the multiple linear regression analysis, under the additive genetic model, with adjustment for age, gender and the first 2 principal components (PCs) (Model 1). The first 2 PCs showed *P*-values <0.05 in the Tracy-Widom test and were therefore included in the adjustment model to control for population stratification. The test statistics, as visualized in a Quantile-Quantile plot (Supplementary Figure 1), appeared well-calibrated. In order to assess for adiposity independent associations, body mass index (BMI) was included as an additional covariate in the multiple linear regression analysis (Model 2). Since previous study has demonstrated that use of metformin may affect circulating PEDF level (7), the use of metformin was also included as a covariate in the multiple logistic regression analysis (Model 3). Genome- and exome-wide significance were defined as $P<5\text{x}10^{-8}$ and $P<6.53\text{x}10^{-7}$ (=0.05/76,951), respectively. In the replication stage, associations of the index SNPs and circulating PEDF level were assessed by multiple linear regression analyses with adjustment for age and gender (Model 1), and with BMI as an additional covariate (Model 2). A Bonferroni corrected *P*-value<0.007 (=0.05/7) was used as the threshold for successful replication. Meta-analysis of the association results of the discovery and replication stages was conducted using GWAMA (30). The inverse variance fixed-effect method was used to meta-analyze the summary statistics of the two stages. The heterogeneity of effect was assessed using Cochran's Q-test and I^2 index. The cross-sectional associations between the PEDF-associated SNPs with diabetic complications were examined by multiple logistic regression analyses, with adjustment for traditional risk factors including age, gender, duration of diabetes, glycated hemoglobin (HbA1c) and presence of hypertension.

Gene-based association analysis

To enhance the power for detecting low frequency variants, a gene-based analysis was conducted using 3 tests implemented in RVTESTS (31) to evaluate their aggregate effect in each gene. These included (i) the unweighted combined multivariate and collapsing (CMC)-Wald burden test (32), which collapsed and combined rare variants and then performed Wald test; (ii) the sequence kernel association test (SKAT) (33); and (iii) the variable threshold (VT) test (34). Only missense and loss-of-function (stopgain and splicing) variants as predicted to be damaging by KGGSeq (35) were grouped into the gene sets. Two MAF thresholds (MAF<1% and MAF<5%) were used for the CMC-Wald and SKAT methods, while only a MAF threshold of <5% was employed for the VT method. Only genes with at least 5 copies of rare alleles were considered (n=10,039 for MAF<1%; and n=11,157 for MAF<5%). The gene-based significance thresholds were defined as $0.05/10{,}039=4.98\times10^{-6}$ for MAF<1% and $0.05/11{,}157=4.40\times10^{-6}$ for MAF<5%.

Cross-sectional association analyses of circulating PEDF level with DN, STDR and CAD

Since all subjects from the discovery and replication stages were recruited from the HKWDR cohort and measurement of circulating PEDF was conducted using the same assay kit in the same laboratory, data from the two stages were combined for analyses. Binary logistic regression analyses were used to examine for the associations of circulating PEDF with DN, STDR and CAD. Multiple logistic regression analyses with adjustment for traditional risk factors were used to investigate the independent association of circulating PEDF.

Mendelian randomization

The IBM SPSS Statistics 25 and R version 3.4.3 (available at www.r.project.org) were employed for the Mendelian randomization analyses. The age and gender-standardized PEDF level was first calculated. *SERPINF1* rs1136287, *SMYD4* rs7224496 and *SERPINF2*