Genetic Regulation of Pigment Epithelium-Derived Factor (PEDF): An Exome-Chip Association Analysis in Chinese Subjects With Type 2 Diabetes

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] and Karen S.L. Lam[1,2,4]

Diabetes 2019;68:198–206 | https://doi.org/10.2337/db18-0500

Elevated circulating levels of pigment epithelium-derived factor (PEDF) have been reported in patients with type 2 diabetes (T2D) and its associated microvascular complications. This study aimed to *1*) identify the genetic determinants influencing circulating PEDF levels in a clinical setting of T2D, *2*) examine the relationship between circulating PEDF and diabetes complications, and *3*) explore the causal relationship between PEDF and diabetes complications. An exome-chip association study on circulating PEDF levels was conducted in 5,385 Chinese subjects with T2D. A meta-analysis of the association results of the discovery stage (*n* = 2,936) and replication stage (*n* = 2,449) was performed. The strongest association was detected at *SERPINF1* (p.Met72Thr; $P_{combined} = 2.06 \times 10^{-57}$; β [SE] −0.33 [0.02]). Two missense variants of *SMYD4* (p.Arg131Ile; $P_{combined} = 7.56 \times 10^{-25}$; β [SE] 0.21 [0.02]) and *SERPINF2* (p.Arg33Trp; $P_{combined} = 8.22 \times 10^{-10}$; β [SE] −0.15 [0.02]) showed novel associations at genome-wide significance. Elevated circulating PEDF levels were associated with increased risks of diabetic nephropathy and sight-threatening diabetic retinopathy. Mendelian randomization analysis showed suggestive evidence of a protective role of PEDF on sight-threatening diabetic retinopathy (*P* = 0.085). Our study provided new insights into the genetic regulation of PEDF and further support for its potential application as a biomarker for diabetic nephropathy and sight-threatening diabetic retinopathy. Further studies to explore the causal relationship of PEDF with diabetes complications are warranted.

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 antiangiogenesis, retina protection, anti-inflammation, antifibrosis, stem cell renewal, neurogenesis, and neuroprotection (1,3–5). Previous studies reported that circulating PEDF levels were significantly higher in patients with type 2 diabetes (T2D) than in subjects without diabetes (6). Serum PEDF in T2D patients was also shown to be elevated after treatment with metformin, an antihyperglycemia agent (7). On the other hand, it has been suggested that PEDF plays a protective role against

[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 Centre 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, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China
[8]Department of Psychiatry, The University of Hong Kong, Hong Kong, China
[9]Centre for Genomic Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, China
[10]State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China

Corresponding authors: Karen S.L. Lam, ksllam@hku.hk; Pak-Chung Sham, pcsham@hku.hk; and Hung-Fat Tse, hftse@hkucc.hku.hk

Received 2 May 2018 and accepted 29 September 2018

This article contains Supplementary Data online at http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/db18-0500/-/DC1.

C.Y.Y.C. and C.-H.L. contributed equally to this work and should be considered co-first authors.

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 levels have been reported in patients with DR and DN in both type 1 diabetes and T2D (13–16). In subjects with T2D, we had also shown that increased serum PEDF levels 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* (Serpin Family F Member 1) gene, which encodes the PEDF protein, as genetic determinants of circulating PEDF levels (19–25). However, to date, no genome-wide or exome-wide association studies on circulating PEDF levels have been published. Furthermore, no other study has evaluated the causal role of PEDF on the risk of diabetes complications using a Mendelian randomization approach. Therefore, we performed the current study to identify the genetic determinants of circulating PEDF levels with an exome-chip association analysis using a custom Illumina HumanExome BeadChip (Asian Exome-chip) in Chinese subjects with T2D. We then examined the relationship between circulating PEDF and diabetes complications, including DN, sight-threatening DR (STDR), and CAD, and explored the causal effect of PEDF on diabetes complications.

RESEARCH DESIGN AND METHODS

This study consisted of an exome-chip association study evaluating the genetic determinants of the circulating PEDF level in 5,385 Chinese subjects with T2D. The discovery stage involved 2,936 subjects who had been examined in our previous exome-chip association studies (26–28), followed by a replication study in 2,449 independent subjects with T2D 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 diabetes complications were then examined. Mendelian randomization analyses were also performed to investigate the causal role of PEDF on diabetes complications in the Hong Kong West Diabetes Registry (HKWDR) cohort.

Subjects

All subjects were recruited from the HKWDR, consisting of unrelated Chinese subjects with T2D 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 levels are described in the Supplementary Data.

Genotyping and Data Quality Control

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 have been described previously (26–28). Sample-level quality control was carried out with regard to sex mismatch, duplication, biological relatedness, and possible sample contamination. A principal component (PC) analysis was performed using a panel of over 20,000 independent common SNPs (minor allele frequency [MAF] >0.05) for the detection of possible existence of non-Chinese samples, and outliers were removed. For SNP-level quality control, variants with MAF <0.1%, or >2% missingness, significantly deviated from Hardy-Weinberg equilibrium (HWE) with $P_{HWE} < 1 \times 10^{-5}$, or originally designed for quality control purpose were excluded from the analysis. After all quality control measures, a total of 2,936 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 ±500 kb from it.

Replication Stage

In the replication stage, all eight index SNPs that achieved a $P_{discovery} < 5 \times 10^{-5}$ were selected for replication in another 2,449 subjects with T2D. Seven of the selected SNPs were genotyped by the MassARRAY Sequenome iPLEX Gold platform (San Diego, CA) at the Centre for Genomic Sciences of The University of Hong Kong. Genotyping of *SERPINF1* rs1136287 was conducted with the TaqMan predesigned genotyping assay (Assay ID:

C___1841779_20) according to the manufacturer's instructions. SNPs that 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, seven 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, sex, and the first two PCs (model 1). The first two 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 Fig. 1), appeared well calibrated. In order to assess for adiposity-independent associations, BMI was included as an additional covariate in the multiple linear regression analysis (model 2). As the previous study demonstrated that the use of metformin may affect circulating PEDF levels (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 $P < 5 \times 10^{-8}$ and $P < 6.53 \times 10^{-7}$ (= 0.05/76,951), respectively. In the replication stage, associations of the index SNPs and circulating PEDF levels were assessed by multiple linear regression analyses with adjustment for age and sex (model 1), 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 (Genome-Wide Association Meta-Analysis) software (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 *Q* test and I^2 index. The cross-sectional associations between the PEDF-associated SNPs with diabetes complications were examined by multiple logistic regression analyses, with adjustment for traditional risk factors including age, sex, duration of diabetes, glycated hemoglobin (HbA_{1c}), and presence of hypertension.

Gene-Based Association Analysis

To enhance the power for detecting low-frequency variants, a gene-based analysis was conducted using three tests implemented in RVTESTS (31) to evaluate their aggregate effect in each gene. These included *1*) the unweighted combined multivariate and collapsing (CMC)-Wald burden test (32), which collapsed and combined rare variants and then performed Wald test; *2*) the sequence kernel association test (SKAT) (33); and *3*) the variable threshold test (34). Only missense and loss-of-function (stop-gain 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, whereas only a MAF threshold of <5% was used for the variable threshold method. Only genes with at least five 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 $0.05/10{,}039 = 4.98 \times 10^{-6}$ for MAF <1% and $0.05/11{,}157 = 4.40 \times 10^{-6}$ for MAF <5%.

Cross-sectional Association Analyses of Circulating PEDF Levels With DN, STDR, and CAD

As 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 used for the Mendelian randomization analyses. The age- and sex-standardized PEDF level was first calculated. *SERPINF1* rs1136287, *SMYD4* (SET And MYND Domain Containing 4) rs7224496, and *SERPINF2* (Serpin Family F Member 2) rs2070863, which showed genome-wide significant associations with circulating PEDF in the current study, were selected as the instrumental variables. The inverse-variance weighted and weighted median methods implemented in the R package "MendelianRandomization" (version 0.3.0) (36), which allow for analyses with correlated variants, were used to calculate the Mendelian randomization estimates using summarized data. The MR-Egger method was then used to test for pleiotropy (36). Details of the Mendelian randomization analysis are described in the Supplementary Data.

RESULTS

Exome-Chip Association Study for Circulating PEDF Levels

In the discovery stage, which involved 2,936 Chinese subjects with T2D (Table 1), 76,951 polymorphic SNPs with MAF ≥0.1% were examined for their associations with circulating PEDF level in the single-variant association analysis (Fig. 1). Of these SNPs, 48.93% altered protein composition and 19.53% were Asian-specific

Table 1—Clinical characteristics of study subjects in the two stages

	Subjects in discovery stage	Subjects in replication stage
N	2,936	2,449
Age (years)	63.41 ± 12.98	63.77 ± 12.50
Duration of diabetes (years)	11.69 ± 8.74	11.89 ± 8.78
Sex (male %)	59.3	56.4
BMI (kg/m^2)	25.97 ± 4.26	26.09 ± 4.48
Fasting plasma glucose (mmol/L)	7.74 ± 2.55	7.71 ± 2.44
HbA_{1c} (%)*	7.40 (6.70–8.30)	7.20 (6.60–8.10)
eGFR (mL/min/1.73 m^2)*	77 (58–94)	80 (59–94)
Hypertension (%)	88.0	87.7
Dyslipidemia (%)	63.7	69.7
Ever smoke (%)	34.7	32.3
Antihyperglycemia treatment		
OAD (%)	94.5	93.4
Metformin (%)	74.9	74.3
Insulin (%)	30.7	28.8
Diet only (%)	2.0	3.1
DN (%)	55.2	54.2
STDR (%)	15.1	15.2
CAD (%)	37.1	14.5
PEDF (μg/mL)*	9.41 (8.07–11.22)	9.45 (7.95–11.25)

Data are mean ± SD, median (interquartile range), or %, unless otherwise stated. Hypertension was defined as blood pressure ≥140/90 mmHg or on antihypertension medications. Dyslipidemia was defined as fasting triglycerides ≥1.69 mmol/L, HDL cholesterol <1.04 mmol/L in men and < 1.29 mmol/L in women, LDL cholesterol ≥3.4 mmol/L, or on lipid-lowering agents. eGFR, estimated glomerular filtration rate; OAD, orally administered antidiabetes drugs. *Natural-log transformed before analysis.

variants with MAF 0.1%–5%. Supplementary Table 1 shows the results of the single-variant association analysis. After LD-clumping, eight index SNPs were found to be significantly associated with circulating PEDF level at $P < 5 \times 10^{-5}$, including three missense variants that achieved genome-wide significance ($P < 5 \times 10^{-8}$) after adjustment for age, sex, and two PCs. The most significant association was observed at a missense variant of *SERPINF1*, rs1136287 (p.Met72Thr; $P_{discovery} = 3.93 \times 10^{-35}$; β [SE] −0.32 [0.03]). *SMYD4* rs7224496 (p.Arg131Ile; $P_{discovery} = 1.07 \times 10^{-18}$; β [SE] 0.25 [0.03]) and *SERPINF2* rs2070863 (p.Arg33Trp; $P_{discovery} = 3.73 \times 10^{-8}$; β [SE] −0.18 [0.03]) were also found to be strongly associated with circulating PEDF level at genome-wide significance (Table 2). Similar results were obtained when BMI (model 2) and the use of metformin (model 3) were further included in the multiple logistic regression analyses (Supplementary Table 1). The overall proportion of variance in circulating PEDF level explained by these three SNPs was 0.059. After conditioning on the top SNP rs1136287 at *SERPINF1*, the associations of *SMYD4* rs7224496 ($P_{conditioned} = 2.05 \times 10^{-4}$) and *SERPINF2* rs2070863 ($P_{conditioned} = 5.64 \times 10^{-4}$) with circulating PEDF levels were attenuated but remained highly suggestive of the presence of independent effects. No other variants reached exome-wide significance in the conditional analysis.

Among the seven SNPs that passed quality control in the replication stage, which involved another 2,449 subjects with T2D (Table 1), four SNPs showed consistent direction of effect as in the discovery stage (Table 2). The associations of *SERPINF1* rs1136287 ($P_{replication} = 9.32 \times 10^{-32}$; β [SE] −0.33 [0.03]) and *SMYD4* rs7224496 ($P_{replication} = 5.07 \times 10^{-10}$; β [SE] 0.19 [0.03]) were successfully replicated. *SERPINF2* rs2070863 showed marginal association with circulating PEDF level in the replication analysis ($P_{replication} = 0.021$; β [SE] −0.09 [0.04]) (Table 2). Similar results were obtained when further adjusted for BMI. In the meta-analysis of the two stages, these SNPs showed even stronger associations with circulating PEDF level: *SERPINF1* (rs1136287; $P_{combined} = 2.06 \times 10^{-57}$; β [SE] −0.33 [0.02]; $I^2 = 0$, $P_{heterogeneity} = 0.810$), *SMYD4* (rs7224496; $P_{combined} = 7.56 \times 10^{-25}$; β [SE] 0.22 [0.02]; $I^2 = 0.57$, $P_{heterogeneity} = 0.128$), and *SERPINF2* (rs2070863; $P_{combined} = 8.22 \times 10^{-10}$; β [SE] −0.15 [0.02]; $I^2 = 0.73$, $P_{heterogeneity} = 0.059$) (Table 2).

Gene-Based Association Analyses

To leverage the power for detecting novel PEDF-associated genes, a gene-based association test was conducted to assess the aggregate effect of low-frequency or rare variants across genes, as the single-variant association test tends to show limited power to detect the association of individual low-frequency or rare variants compared with common variants. The *ITGAV* gene ($P = 3.82 \times 10^{-5}$; β [SE] −0.46 [0.14]) exhibited the strongest effect among others in the gene-level test (SKAT <1%). A missense variant of this gene, rs61757099 (p.Asn769Ser), has also shown a suggestive association in the single-variant association test (Table 2). Supplementary Table 2 shows the genes demonstrating suggestive associations ($P < 5 \times 10^{-4}$) with circulating PEDF level in the gene-based association test, including *SERPINF1*, which showed the strongest association with circulating PEDF level in the single-variant association test. However, none of these genes reached the exome-wide significance level in the gene-based analysis.

Cross-sectional Associations of Circulating PEDF With DN, STDR, and CAD in the HKWDR Cohort

We further investigated the associations between circulating PEDF and both microvascular and macrovascular complications, including DN, STDR, and CAD, in the HKWDR cohort. Circulating PEDF levels showed significant associations with DN and STDR (both $P_{unadjusted} < 0.001$) but not with CAD ($P_{unadjusted} = 0.240$). Elevated PEDF levels were found to be independently associated with both DN

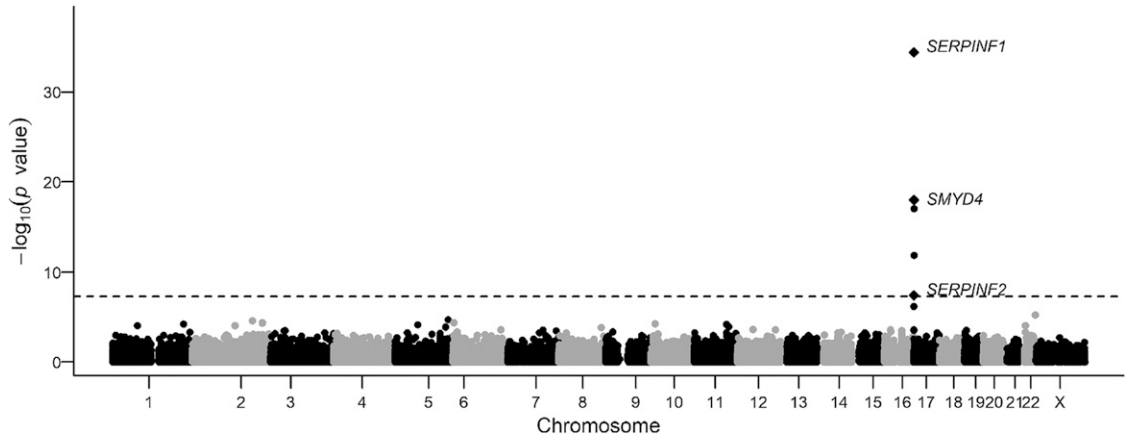

Figure 1— Manhattan plot of the discovery stage results. The *y*-axis represents the $-\log_{10} P$ value, and the *x*-axis represents the genomic position. The dots represent the 76,951 SNPs analyzed, relative to their position on each chromosome (alternating black and gray). The black horizontal dashed line indicates genome-wide significance ($P = 5 \times 10^{-8}$). The diamond symbols indicate the three variants that reached genome-wide significance, which included *SERPINF1* rs1136287 (p.Met72Thr), *SMYD4* rs7224496 (p.Arg131Ile), and *SERPINF2* rs2070863 (p.Arg33Trp).

($P_{adjusted} < 0.001$; odds ratio [OR] [95% CI] 6.66 [4.70–9.42]) (Table 3) and STDR ($P_{adjusted} < 0.001$; OR [95% CI] 2.06 [1.52–2.78]) (Table 4) even after adjustment for the traditional risk factors.

Cross-sectional Associations of PEDF-Associated SNPs With DN and STDR

In view of the potential causal role of PEDF in DN and STDR, we also examined the associations between the PEDF-associated SNPs and these diseases. The *SMYD4* rs7224496 showed a trend for association with STDR in the discovery stage cohort ($P = 0.061$; OR [95% CI] 0.85 [0.72–1.01]) after adjustment for traditional risk factors, but this was not replicated in the replication stage cohort. None of the other SNPs showed significant association with DN or STDR (Table 5).

Mendelian Randomization Analysis on the Effect of PEDF on Risk of DN and STDR

We then proceeded to examine the causal role of PEDF on the risk of DN and STDR using the three PEDF-associated SNPs (*SERPINF1* rs1136287, *SMYD4* rs7224496, *SERPINF2* rs2070863) as instrumental variables in the HKWDR cohort (n = 5,385). All three SNPs were strongly associated with the age- and sex-standardized circulating PEDF levels, and *SERPINF1* rs1136287 showed the highest F-statistic and R^2 value among all SNPs (Supplementary Table 3). However, there was no association between these three SNPs and the risk of DN or STDR (Supplementary Table 4). All three SNPs did not show significant association with the potential confounders after correction for multiple testing (Supplementary Table 5). Mendelian randomization estimates of the associations between an SD increase in genetically determined age- and sex-standardized PEDF and risk of DN or STDR are presented in Table 6. Based on the Mendelian randomization estimates calculated using the weighted median method, there was suggestive evidence that the genetically determined age- and sex-standardized PEDF was associated with a reduced risk of STDR ($P = 0.085$; OR [95% CI] 0.80 [0.62–1.03]). However, there was no evidence supporting the causal role on the risk of DN. No substantial horizontal pleiotropy effects have been observed, as indicated by the insignificant *P* values of intercepts calculated by the MR-Egger method (Supplementary Table 6).

DISCUSSION

The current study reported the first exome-chip association analysis on circulating PEDF level in subjects with T2D. To our knowledge, no previous studies have reported an association of genetic variants with circulating PEDF level at genome-wide significance. We have demonstrated the independent associations of circulating PEDF levels with DN and STDR and undertaken a Mendelian randomization study to investigate the causal role of PEDF on the risk of DN and STDR.

The strongest association with circulating PEDF level was observed at *SERPINF1* rs1136287 (p.Met72Thr) in the single-variant association analysis, reaching genome-wide significance ($P_{combined} = 2.06 \times 10^{-57}$). Notably, in a previous study, this variant was shown to be associated with a lower level of plasma PEDF in healthy control subjects (n = 72; $P < 0.01$) but not in patients with colorectal cancer (n = 80) (25). Furthermore, carriers of *SERPINF1* pathogenic mutations have shown extremely low levels of circulating PEDF compared with noncarriers (20–24). Thus, using different approaches, the current study and previous studies (19–25) have demonstrated that the *SERPINF1* gene is the major genetic determinant of circulating PEDF levels. However, our observed associations of the *SMYD4* and *SERPINF2* variants with circulating PEDF have provided some novel insights into the genetic regulation of PEDF, although these variants only

showed suggestive associations with circulating PEDF levels in the conditional analysis. Further confirmation of these newly identified associations is required. The r^2 between the index SNPs of the *SERPINF1*, *SMYD4*, and *SERPINF2* gene regions were <0.2 in our population (Supplementary Fig. 2). *SERPINF1* is flanked by *SERPINF2* located ~6.5 kb upstream on the same DNA strand and *SMYD4* located ~1.5 kb downstream on the reverse DNA strand. As discussed in the following sections, *SERPINF2* and *SMYD4* may also be involved in the regulation of PEDF via different signaling pathways.

SERPINF2 rs2070863 (p.Arg33Trp) was found to be significantly associated with decreased circulating PEDF levels in the current study. *SERPINF2* encodes the α-2-antiplasmin (37). α-2-Antiplasmin specifically induces TGFβ1 production during the process of fibrosis (38). On the other hand, TGFβ1 has been reported to show direct positive regulation (39,40) and indirect negative regulation on the PEDF expression (41), depending on the cell environment and cell type used in the studies. It is possible that *SERPINF2* rs2070863 leads to a decreased expression of α-2-antiplasmin and hence a reduction in TGFβ1 production, thereby resulting in reduced TGFβ1-induced PEDF expression.

The missense variant of *SMYD4*, rs7224496 (p.Arg131-Ile), was associated with elevated circulating PEDF levels in the current study. This variant has been predicted to be functional by multiple prediction tools (SIFT score = 0.003; PROVEAN score = −3.02; PolypPhen2 HDIV score = 0.993; Polyphen2 HVAR score = 0.961). *SMYD4* is a member of the SMYD gene family that comprises both an MYND domain and a SET domain and often carries the lysine methyltransferase activity for histone modification (42). Histone lysine methylation and DNA methylation are highly interrelated, and these two systems rely on each other mechanistically for normal chromatin function (43,44). One previous study showed that PEDF promoter methylation status negatively correlated with the PEDF mRNA expression level in locally advanced rectal carcinoma tissues (45). Therefore, SMYD4 may influence PEDF expression via its potential effect on epigenetic modification. However, SMYD4 has been proposed to function as a tumor suppressor through inhibiting platelet-derived growth factor (PDGF) receptor α (PDGFRα) expression (46). Binding of PDGF-A, -B and -C chains to PDGFRα induces its dimerization and autophosphorylation (47), and PDGFRα signaling has been shown to be critical for the maximum induction of c-Jun N-terminal kinases (JNK)-1 by PDGF-BB (48). Furthermore, reduced PDGFRα activation in response to PDGF-BB resulted in decreased downstream p38 signaling (49). Interestingly, PEDF expression has also been shown to be downregulated by PDGF-BB, with JNK1/2 and p38 signaling being critical for this repression (41). Taken together, we postulate that *SMYD4* rs7224496 may lead to an increased SMYD4 expression, which results in the reduced expression of PDGFRα and hence decreased PDGF-BB activation of JNK-1 and p38,

Table 2—Association of SNPs with circulating PEDF level

				Discovery (*N* = 2,936)		Replication (*N* = 2,449)		Combined (*N* = 5,385)		
Nearest gene(s)	SNP	A1	A2	MAF	β (SE)	*P* value*	MAF	β (SE)	*P* value†	Dir	β (SE)	*P* value†
SERPINF1	rs1136287	C	T	0.452	−0.32 (0.03)	3.93×10^{-35}	0.463	−0.33 (0.03)	9.32×10^{-32}	− −	−0.33 (0.02)	2.06×10^{-57}
SMYD4	rs7224496	A	C	0.321	0.25 (0.03)	1.07×10^{-18}	0.323	0.19 (0.03)	5.07×10^{-10}	+ +	0.22 (0.02)	7.56×10^{-25}
SERPINF2	rs2070863	T	C	0.184	−0.18 (0.03)	3.73×10^{-8}	0.176	−0.09 (0.04)	0.021	− −	−0.15 (0.02)	8.22×10^{-10}
TENM2	rs9313307	A	C	0.118	0.17 (0.04)	1.90×10^{-5}	0.119	−0.03 (0.05)	0.540	+ −	0.08 (0.03)	5.63×10^{-3}
ITGAV	rs61757099	G	A	0.005	−0.75 (0.18)	2.44×10^{-5}	0.005	0.06 (0.20)	0.758	− +	−0.38 (0.13)	4.18×10^{-3}
NQO2	rs28383651	C	T	0.016	0.42 (0.10)	4.00×10^{-5}	0.019	−0.06 (0.10)	0.589	+ −	0.19 (0.07)	8.31×10^{-3}
TNP1/DIRC3	rs17778798	C	T	0.059	0.22 (0.05)	4.04×10^{-5}	0.062	0.08 (0.06)	0.176	+ +	0.16 (0.04)	2.32×10^{-5}

The βs are reported with respect to the minor allele. A1, minor allele; A2, major allele; Dir, direction of effect. *Adjusted for age, sex, and two PCs. †Adjusted for age and sex.

Table 3—Association analyses of circulating PEDF with DN in the HKWDR cohort

	With DN	Without DN	OR (95% CI)†	*P* value†
N	2,914	2,408	—	—
PEDF (μg/mL)*	10.48 (8.94–12.43)	8.46 (7.31–9.67)	6.66 (4.70–9.42)	<0.001
Age (years)	65.83 ± 12.49	63.05 ± 12.81	1.00 (0.99–1.01)	0.758
Sex (male)	57.6	58.4	0.84 (0.73–1.01)	0.999
Duration of diabetes (years)*	12.0 (6.0–19.0)	8.0 (4.0–15.0)	1.18 (1.08–1.28)	<0.001
Hypertension (%)	78.4	95.8	2.48 (1.95–3.16)	<0.001
HbA_{1c} (%)*	7.5 (6.7–8.4)	7.2 (6.6–8.0)	4.59 (2.96–7.12)	<0.001
BMI (kg/m^2)	26.49 ± 4.49	25.48 ± 4.16	1.05 (1.03–1.07)	<0.001
eGFR (mL/min/1.73 m^2)*	61.0 (44.0–84.0)	88.0 (77.0–98.0)	0.02 (0.01–0.03)	<0.001

Data are mean ± SD, median (interquartile range), or %, unless stated otherwise. eGFR, estimated glomerular filtration rate. *Natural-log transformed before analysis. †Adjusted for traditional risk factors.

leading to a reduction in PDGF-BB–mediated PEDF repression and thus elevated PEDF expression and circulating PEDF levels (Supplementary Fig. 3). Nonetheless, the roles of both α-2-antiplasmin (SERPINF2) and SMYD4 in the regulation of PEDF expression remains to be elucidated in further functional analyses.

Despite the strongly significant associations of the *SERPINF1*, *SMYD4*, and *SERPINF2* SNPs with circulating PEDF levels, we were unable to detect significant association of these SNPs with diabetes complications. Given the potential involvement of PEDF in diabetes complications, whether these PEDF-associated variants may be involved in the development of diabetes complications warrants further investigation in other independent cohorts. Similarly, further replications are also required to validate our findings on the gene-based association analysis.

We demonstrated the independent associations of circulating PEDF levels with DN and STDR, but not with CAD, in the HKWDR cohort, in line with the findings in previous publications (13–16). Our results have provided further support for the potential use of circulating PEDF level as a biomarker for diabetic microvascular complications, such as DN and STDR. Our finding was in agreement with the notion that the increased circulating level of PEDF in disease status might represent a compensatory systemic response (1), thereby providing a protective effect against further damage. In the Mendelian randomization study, our results are not supportive of a causal role of PEDF on the risk of DN. However, we found suggestive evidence for a protective role of PEDF on STDR. Noteworthy, our Mendelian randomization analysis was obviously lacking in statistical power. Our sample sizes were only able to achieve ~25% and ~9% power, respectively, for detecting the causal effect of PEDF on the risk of STDR and DN, assuming a R^2 of 0.059 at a significance level of 0.05. Therefore, a definite conclusion on the role of PEDF on the risk of DN and STDR could not be drawn from the current study. Identification of more PEDF-associated variants to provide stronger instrumental variables will serve to improve the study power. Mendelian randomization studies with sufficient power in other populations to investigate further the causal role of PEDF on the risk of diabetes complications are warranted.

The current study has some limitations. First, this study lacks an external validation. Further replication study in independent cohorts in other populations would serve to validate our findings. Second, despite the use of the Asian exome chip, which had allowed us to examine more functional coding variants with a lower MAF compared with the conventional genotyping arrays, our study, with its relatively small sample size, has limited power to detect the association of rare or low-frequency variants. In view of this, we conducted the gene-based association analysis to assess the combined effect of these rare or low-frequency

Table 4—Association analyses of circulating PEDF with STDR in the HKWDR cohort

	With STDR	Without STDR	OR (95% CI)†	*P* value†
N	772	4,317	—	—
PEDF (μg/mL)*	9.97 (8.40–12.20)	9.33 (7.94–11.07)	2.06 (1.52–2.78)	<0.001
Age (years)	63.05 ± 12.81	65.83 ± 12.49	1.00 (0.99–1.01)	0.842
Sex (male)	53.9	59.3	0.84 (0.72–0.99)	0.034
Duration of diabetes (years)*	16.0 (9.0–23.0)	9.0 (4.0–15.0)	2.05 (1.81–2.33)	<0.001
Hypertension (%)	86.5	94.3	1.98 (1.41–2.78)	<0.001
HbA_{1c} (%)*	7.7 (6.9–8.7)	7.3 (6.6–8.1)	2.55 (1.58–4.14)	<0.001

Data are mean ± SD, median (interquartile range), or %, unless stated otherwise. *Natural-log transformed before analysis. †Adjusted for traditional risk factors.

Table 5—Association of PEDF-associated SNPs with DN and STDR with adjustment for traditional risk factors

				DN			
				Discovery		Replication	
Gene	SNP	A1	A2	OR (95% CI)	*P* value*	OR (95% CI)	*P* value†
SERPINF1	rs1136287	C	T	0.95 (0.85–1.07)	0.401	1.04 (0.92–1.18)	0.512
SMYD4	rs7224496	A	C	1.08 (0.96–1.22)	0.190	0.96 (0.84–1.09)	0.510
SERPINF2	rs2070863	T	C	0.95 (0.82–1.1)	0.484	1.15 (0.98–1.36)	0.093
				STDR			
				Discovery		Replication	
SERPINF1	rs1136287	C	T	1.04 (0.89–1.21)	0.633	1.06 (0.90–1.25)	0.511
SMYD4	rs7224496	A	C	0.85 (0.72–1.01)	0.061	0.98 (0.82–1.17)	0.853
SERPINF2	rs2070863	T	C	0.96 (0.79–1.17)	0.695	1.10 (0.88–1.36)	0.417

In the DN group, the discovery stage is 1,605 subjects with DN vs. 1,301 without; replication stage, 1,309 with vs. 1,107 without. In the STDR group, the discovery stage is 420 subjects with STDR vs. 2,359 without; replication stage, 352 with vs. 1,958 without. The ORs are reported with respect to the minor allele. A1, minor allele; A2, major allele. *Adjusted for age, sex, duration of diabetes, HbA_{1c}, presence of hypertension, and two PCs. †Adjusted for age, sex, duration of diabetes, HbA_{1c}, and presence of hypertension.

alleles. A whole-genome imputation using publicly available reference panels, such as the 1000 Genome Project, would increase the total evaluable SNPs in studies using array chips. However, the exome chip used in this study has been designed with enrichment of rare exonic variants identified primarily from the European population. This feature implies that a large proportion of variants are likely to be unique to the exome chip design and are therefore not present in the publicly available reference panels of non-European populations, such as Asians, and may in turn lead to reduced imputation accuracy. Future meta-analysis of genome-wide or exome-wide association studies would be useful to identify additional genetic variants influencing the circulating PEDF level. Moreover, the current study lacks functional analyses to validate the proposed mechanisms whereby the two newly identified genetic variants may regulate PEDF expression, which remain to be confirmed in future functional studies. Finally, we have limited power to investigate the cause–effect relationship between PEDF and the risk of diabetes complications. Further Mendelian randomization studies in larger cohorts may provide more insights into the causal role of PEDF on the risk of diabetes complications.

In summary, we have conducted an exome-chip association study and identified, for the first time, three missense variants of *SERPINF1*, *SMYD4,* and *SERPINF2* showing genome-wide significant associations with circulating PEDF level in subjects with T2D. Our findings have shed light on the genetic regulation of circulating PEDF level and provided further support for its potential application as a biomarker for DN and STDR. The causal role of PEDF on diabetes complications deserves further investigations.

Table 6—Mendelian randomization estimates of the associations between an SD increase in genetically determined age- and sex-standardized PEDF and risk of DN or STDR

	Inverse-variance weighted method		Weighted median method	
	OR (95% CI)	*P* value	OR (95% CI)	*P* value
DN	1.04 (0.85–1.28)	0.671	1.07 (0.89–1.29)	0.442
STDR	0.82 (0.61–1.10)	0.187	0.80 (0.62–1.03)	0.085

Acknowledgments. The authors thank Rachel Wong (Department of Medicine, The University of Hong Kong, Hong Kong, China) for her technical assistance in the measurement of serum PEDF level and all the study subjects and clinical and research staffs of the HKWDR for their contribution to this research study.

Funding. This work was supported by the Hong Kong Research Grants Council Theme-based Research Scheme (T12-705/11).

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

Author Contributions. C.Y.Y.C. and C.-H.L. wrote the first draft of the manuscript. C.Y.Y.C. analyzed and interpreted the data. C.Y.Y.C., C.-H.L., A.X., K.-W.A., C.H.Y.F., K.K.K.N., K.H.M.K., W.-S.C., Y.-C.W., M.M.A.Y., J.H., K.C.B.T., and T.-H.L. were involved in the sample collection and selection and phenotype data preparation. C.S.T. and P.-C.S. provided useful comments to data analysis. H.-F.T., P-.C.S., and K.S.L.L. conceived the study, undertook project leadership, and contributed equally to the supervision of this work. All authors contributed to the drafting and critical revision of the manuscript and approved the final version of the manuscript. H.-F.T., P-.C.S., and K.S.L.L. are the guarantors of this work and, as such, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. He X, Cheng R, Benyajati S, Ma JX. PEDF and its roles in physiological and pathological conditions: implication in diabetic and hypoxia-induced angiogenic diseases. Clin Sci (Lond) 2015;128:805–823

2. Famulla S, Lamers D, Hartwig S, et al. Pigment epithelium-derived factor (PEDF) is one of the most abundant proteins secreted by human adipocytes and induces insulin resistance and inflammatory signaling in muscle and fat cells. Int J Obes 2011;35:762–772

3. Ek ET, Dass CR, Choong PF. PEDF: a potential molecular therapeutic target with multiple anti-cancer activities. Trends Mol Med 2006;12:497–502

4. Wang JJ, Zhang SX, Lu K, et al. Decreased expression of pigment epithelium-derived factor is involved in the pathogenesis of diabetic nephropathy. Diabetes 2005;54:243–250

5. Wang JJ, Zhang SX, Mott R, et al. Salutary effect of pigment epithelium-derived factor in diabetic nephropathy: evidence for antifibrogenic activities. Diabetes 2006;55:1678–1685

6. Jenkins A, Zhang SX, Gosmanova A, et al. Increased serum pigment epithelium derived factor levels in type 2 diabetes patients. Diabetes Res Clin Pract 2008;82:e5–e7
7. Akın S, Aksoy DY, Cınar N, et al. Pigment epithelium-derived factor increases in type 2 diabetes after treatment with metformin. Clin Endocrinol (Oxf) 2012;77:852–856
8. Boehm BO, Lang G, Volpert O, et al. Low content of the natural ocular anti-angiogenic agent pigment epithelium-derived factor (PEDF) in aqueous humor predicts progression of diabetic retinopathy. Diabetologia 2003;46:394–400
9. Spranger J, Osterhoff M, Reimann M, et al. Loss of the antiangiogenic pigment epithelium-derived factor in patients with angiogenic eye disease. Diabetes 2001;50:2641–2645
10. Ogata N, Nishikawa M, Nishimura T, Mitsuma Y, Matsumura M. Unbalanced vitreous levels of pigment epithelium-derived factor and vascular endothelial growth factor in diabetic retinopathy. Am J Ophthalmol 2002;134:348–353
11. Awad AS, Gao T, Gvritishvili A, et al. Protective role of small pigment epithelium-derived factor (PEDF) peptide in diabetic renal injury. Am J Physiol Renal Physiol 2013;305:F891–F900
12. Wang JJ, Zhang SX, Mott R, et al. Anti-inflammatory effects of pigment epithelium-derived factor in diabetic nephropathy. Am J Physiol Renal Physiol 2008;294:F1166–F1173
13. Jenkins AJ, Zhang SX, Rowley KG, et al. Increased serum pigment epithelium-derived factor is associated with microvascular complications, vascular stiffness and inflammation in type 1 diabetes. Diabet Med 2007;24:1345–1351
14. Ogata N, Matsuoka M, Matsuyama K, et al. Plasma concentration of pigment epithelium-derived factor in patients with diabetic retinopathy. J Clin Endocrinol Metab 2007;92:1176–1179
15. Matsuyama K, Ogata N, Matsuoka M, et al. Relationship between pigment epithelium-derived factor (PEDF) and renal function in patients with diabetic retinopathy. Mol Vis 2008;14:992–996
16. Hui E, Yeung CY, Lee PC, et al. Elevated circulating pigment epithelium-derived factor predicts the progression of diabetic nephropathy in patients with type 2 diabetes. J Clin Endocrinol Metab 2014;99:E2169–E2177
17. Lin JM, Wan L, Tsai YY, et al. Pigment epithelium-derived factor gene Met72Thr polymorphism is associated with increased risk of wet age-related macular degeneration. Am J Ophthalmol 2008;145:716–721
18. Ma SY, Guo YY, Wang SX, et al. The T allele of rs8075977 in the 5′-flanking region of the pedf gene is associated with reduced risk of coronary artery disease in elderly chinese men. Tohoku J Exp Med 2017;241:297–308
19. Böhm A, Ordelheide AM, Machann J, et al. Common genetic variation in the SERPINF1 locus determines overall adiposity, obesity-related insulin resistance, and circulating leptin levels. PLoS One 2012;7:e34035
20. Homan EP, Rauch F, Grafe I, et al. Mutations in SERPINF1 cause osteogenesis imperfecta type VI. J Bone Miner Res 2011;26:2798–2803
21. Venturi G, Gandini A, Monti E, et al. Lack of expression of SERPINF1, the gene coding for pigment epithelium-derived factor, causes progressively deforming osteogenesis imperfecta with normal type I collagen. J Bone Miner Res 2012;27:723–728
22. Al-Jallad H, Palomo T, Moffatt P, Roughley P, Glorieux FH, Rauch F. Normal bone density and fat mass in heterozygous SERPINF1 mutation carriers. J Clin Endocrinol Metab 2014;99:E2446–E2450
23. Wang JY, Liu Y, Song LJ, et al. Novel mutations in SERPINF1 result in rare osteogenesis imperfecta type VI. Calcif Tissue Int 2017;100:55–66
24. Wang JY, Li LJ, Zhang Q, et al. Extremely low level of serum pigment epithelium-derived factor is a special biomarker of Chinese osteogenesis imperfecta patients with SERPINF1 mutations. Clin Chim Acta 2018;478:216–221
25. Wågsäter D, Löfgren S, Zar N, Hugander A, Dimberg J. Pigment epithelium-derived factor expression in colorectal cancer patients. Cancer Invest 2010;28:872–877
26. Tang CS, Zhang H, Cheung CY, et al. Exome-wide association analysis reveals novel coding sequence variants associated with lipid traits in Chinese. Nat Commun 2015;6:10206
27. Cheung CY, Tang CS, Xu A, et al. Exome-chip association analysis reveals an Asian-specific missense variant in PAX4 associated with type 2 diabetes in Chinese individuals. Diabetologia 2017;60:107–115
28. Cheung CYY, Tang CS, Xu A, et al. An exome-chip association analysis in Chinese subjects reveals a functional missense variant of *GCKR* that regulates FGF21 levels. Diabetes 2017;66:1723–1728
29. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575
30. Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics 2010;11:288
31. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics 2016;32:1423–1426
32. Li B, Leal SM. Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet 2008;83:311–321
33. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 2011;89:82–93
34. Price AL, Kryukov GV, de Bakker PI, et al. Pooled association tests for rare variants in exon-resequencing studies. Am J Hum Genet 2010;86:832–838
35. Li MX, Gui HS, Kwan JS, Bao SY, Sham PC. A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. Nucleic Acids Res 2012;40:e53
36. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol 2017;46:1734–1739
37. Coughlin PB. Antiplasmin: the forgotten serpin? FEBS J 2005;272:4852–4857
38. Kanno Y, Kuroki A, Okada K, et al. Alpha2-antiplasmin is involved in the production of transforming growth factor beta1 and fibrosis. J Thromb Haemost 2007;5:2266–2273
39. Cosgrove GP, Brown KK, Schiemann WP, et al. Pigment epithelium-derived factor in idiopathic pulmonary fibrosis: a role in aberrant angiogenesis. Am J Respir Crit Care Med 2004;170:242–251
40. Liakouli V, Elies J, El-Sherbiny YM, et al. Scleroderma fibroblasts suppress angiogenesis via TGF-β/caveolin-1 dependent secretion of pigment epithelium-derived factor. Ann Rheum Dis 2018;77:431–440
41. Nwani NG, Deguiz ML, Jimenez B, et al. Melanoma cells block PEDF production in fibroblasts to induce the tumor-promoting phenotype of cancer-associated fibroblasts. Cancer Res 2016;76:2265–2276
42. Calpena E, Palau F, Espinós C, Galindo MI. Evolutionary history of the Smyd gene family in Metazoans: a framework to identify the orthologs of human Smyd genes in drosophila and other animal species. PLoS One 2015;10:e0134106
43. Rose NR, Klose RJ. Understanding the relationship between DNA methylation and histone lysine methylation. Biochim Biophys Acta 2014;1839:1362–1372
44. Cedar H, Bergman Y. Linking DNA methylation and histone modification: patterns and paradigms. Nat Rev Genet 2009;10:295–304
45. Yi H, Ji D, Zhan T, et al. Prognostic value of pigment epithelium-derived factor for neoadjuvant radiation therapy in patients with locally advanced rectal carcinoma. Int J Oncol 2016;49:1415–1426
46. Hu L, Zhu YT, Qi C, Zhu YJ. Identification of Smyd4 as a potential tumor suppressor gene involved in breast cancer development. Cancer Res 2009;69:4067–4072
47. Andrae J, Gallini R, Betsholtz C. Role of platelet-derived growth factors in physiology and medicine. Genes Dev 2008;22:1276–1312
48. Yu J, Deuel TF, Kim HR. Platelet-derived growth factor (PDGF) receptor-alpha activates c-Jun NH2-terminal kinase-1 and antagonizes PDGF receptor-beta -induced phenotypic transformation. J Biol Chem 2000;275:19076–19082
49. Kikuchi A, Pradhan-Sundd T, Singh S, Nagarajan S, Loizos N, Monga SP. Platelet-derived growth factor receptor α contributes to human hepatic stellate cell proliferation and migration. Am J Pathol 2017;187:2273–2287