SUPPLEMENTARY DATA

Supplementary notes

The Hong Kong West Diabetes Registry (HKWDR) was established since 2008. All subjects, at enrollment to the registry, were invited to participate in a prospective cohort study that aimed to identify risk factors, including genetic and protein biomarkers, which predisposed them to the development of diabetic complications. At regular visits, subjects underwent comprehensive clinical assessments and laboratory investigations to determine their glycemic control as well as the presence of diabetic complications. 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. Fasting venous blood was drawn for the measurement of plasma glucose, lipids and glycated hemoglobin (HbA1c). At the baseline visit, demographic data, including age, gender, occupation, smoking, alcohol consumption and physical activity were obtained. Detailed family, medical and medication histories were ascertained using a standardized questionnaire. Anthropometric parameters, including body weight, height, body mass index (BMI), waist circumference and blood pressure were measured. Blood samples were also obtained at the baseline visit and stored for genetic and biomarker analysis.

Definitions of DN, STDR and CAD

DN cases (n=2914) was defined as subjects with moderately increased urine albumin-to-creatinine ratio (ARC) (albuminuria category A2), or severely increased urine ARC (albuminuria category A3), or having an estimated glomerular filtration rate (eGFR) < 60mL/min/1.73m^2 (GFR categories G3a-G5) (1). Those subjects with normal to mildly increased urine ARC (albuminuria category A1) and an eGFR ≥ 60 mL/min/1.73m^2 were classified as non-DN controls (n=2408). The eGFR values were calculated based on serum creatinine in mg/dL and age as at the reference date of serum creatinine measurement, using the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation (2).

The severity of DR was graded according to the English National Screening Program guidelines (3). STDR cases (n=772) were defined as subjects with either pre-proliferative DR (pre-PDR; graded R2), proliferative DR (PDR; graded R3), or showing features of maculopathy (graded M1) (4; 5). Subjects without DR (graded R0) or with background DR (graded R1) were considered as non-STDR controls (n=4317).

CAD cases (n=1406) were defined as subjects who had coronary artery revascularization interventions (including percutaneous coronary intervention and/or coronary artery bypass graft surgery); or those who had been diagnosed with MI. Subjects who had no documented history or angiogram evidence of CAD were defined as non-CAD controls (n=3832).

Subjects with undetermined disease status were excluded from the corresponding association analyses.

Measurement of circulating PEDF level

Serum PEDF level was measured from the stored serum samples collected at the baseline assessment, using a human PEDF enzyme-linked immunosorbent assay (ELISA) kit according to the manufacturer's protocol (BioVendor Laboratory Medicine, Inc., Modrice, Czech Republic), with calibration and quality control (QC) performed as previously described (6-8). The intra- and inter-assay coefficients of variation of the PEDF ELISA were 2.9 to 4.1% and 5.3 to 6.6%, respectively.

Mendelian randomization analysis

To assess the strength of the instrumental variables, linear regression analysis was used to calculate the F-statistic and R^2 values between the SNPs and PEDF level. A F-statistic value >10 indicates that the SNP is a valid instrumental variable (9). To test for the assumption that the instrumental variables were not associated with potential confounders for the association of circulating PEDF with DN and STDR, linear or logistic regression analyses were conducted to estimate the associations between the SNPs and the major risk factors for DN or STDR. The summarized data of coefficients and standard errors (SE)

from linear regression of SNPs on the PEDF level, as well as the coefficients (Ln[OR]) and SE from logistic regression of SNPs on risk of DN or STDR in the HKWDR cohort were generated to calculate the Mendelian randomization estimates. The inverse-variance weighted (IVW) and weighted median methods implemented in the R package "MendelianRandomization" (version 0.3.0) (10) , which allow for analyses with correlated variants, were used to calculate the Mendelian randomization estimates using the summarized data. The MR-Egger method was then used to test for pleiotropy (10). The power of the Mendelian randomization study was estimated by the "mRnd" online tool (available at http://cnsgenomics.com/shiny/mRnd/) (11).

References

1. Levin A, Stevens PE: Summary of KDIGO 2012 CKD Guideline: behind the scenes, need for guidance, and a framework for moving forward. Kidney Int 2014;85:49-61
2. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF, 3rd, Feldman HI, Kusek JW, Eggers P, Van Lente F, Greene T, Coresh J, Ckd EPI: A new equation to estimate glomerular filtration rate. Ann Intern Med 2009;150:604-612
3. UK National Screening Committee.: Essential Elements in Developing a Diabetic Retinopathy Screening Programme. Workbook version 4.4. UK: UK NAtional Screening Committee; 2012 http://diabeticeye.screening.nhs.uk/qualityassurance (assessed 1 July 2018)
4. Lian JX, Gangwani RA, McGhee SM, Chan CK, Lam CL, Wong DS: Systematic screening for diabetic retinopathy (DR) in Hong Kong: prevalence of DR and visual impairment among diabetic population. Br J Ophthalmol 2016;100:151-155
5. Cheung CY, Hui EY, Lee CH, Kwok KH, Gangwani RA, Li KK, Chan JC, Woo YC, Chow WS, Yuen MM, Wong RL, Fong CH, Xu A, Wong DS, Sham PC, Lam KS: Impact of Genetic Loci Identified in Genome-Wide Association Studies on Diabetic Retinopathy in Chinese Patients With Type 2 Diabetes. Invest Ophthalmol Vis Sci 2016;57:5518-5524
6. Hui E, Yeung CY, Lee PC, Woo YC, Fong CH, Chow WS, Xu A, Lam KS: Elevated circulating pigment epithelium-derived factor predicts the progression of diabetic nephropathy in patients with type 2 diabetes. J Clin Endocrinol Metab 2014;99:E2169-2177
7. Chen C, Tso AW, Law LS, Cheung BM, Ong KL, Wat NM, Janus ED, Xu A, Lam KS: Plasma level of pigment epithelium-derived factor is independently associated with the development of the metabolic syndrome in Chinese men: a 10-year prospective study. J Clin Endocrinol Metab 2010;95:5074-5081
8. Chen C, Tso AW, Cheung BM, Law LS, Ong KL, Wat NM, Janus ED, Xu A, Lam KS: Plasma concentration of pigment epithelium-derived factor is closely associated with blood pressure and predicts incident hypertension in Chinese: a 10-year prospective study. Clin Endocrinol (Oxf) 2012;76:506-513
9. Lawlor DA, Harbord RM, Sterne JA, Timpson N, Davey Smith G: Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. Stat Med 2008;27:1133-1163
10. Yavorska OO, Burgess S: MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. Int J Epidemiol 2017;46:1734-1739

Brion MJ, Shakhbazov K, Visscher PM: Calculating statistical power in Mendelian randomization studies. Int J Epidemiol 2013;42:1497-1501

SUPPLEMENTARY DATA

Supplementary Table 1. Association results of SNPs reaching $P<1x10^{-4}$ in the discovery stage.

Nearest gene(s)	SNP	Position	Feature	MAF	A1	A2	β(SE)	*P-value**	*P-value*†	*P-value*‡
SERPINF1	rs1136287	17:1673276	p.Met72Thr	0.452	C	T	-0.32(0.03)	$3.93x10^{-35}$	$2.95x10^{-38}$	$8.28x10^{-41}$
SMYD4	rs7224496	17:1704296	p.Arg131Ile	0.321	A	C	0.25(0.03)	$1.07x10^{-18}$	$9.63x10^{-20}$	$1.18x10^{-19}$
SERPINF2	rs2070863	17:1648502	p.Arg33Trp	0.184	T	C	-0.18(0.03)	$3.73x10^{-8}$	$2.01x10^{-7}$	$8.87x10^{-7}$
TUBGCP6	rs760622076	22:50682370	p.Ser173Arg	0.002	T	G	-1.37(0.30)	$5.75x10^{-6}$	$1.81x10^{-5}$	$5.17x10^{-6}$
TENM2	rs9313307	5:165553143	intergenic	0.118	A	C	0.17(0.04)	$1.90x10^{-5}$	$2.42x10^{-4}$	$3.20x10^{-4}$
ITGAV	rs61757099	2:187533499	p.Asn769Ser	0.005	G	A	-0.75(0.18)	$2.44x10^{-5}$	$1.30x10^{-4}$	$2.52x10^{-4}$
NQO2	rs28383651	6:3019744	p.Val184Ala	0.016	C	T	0.42(0.10)	$4.00x10^{-5}$	$9.41x10^{-4}$	$1.80x10^{-3}$
TNP1/DIRC3	rs17778798	2:217927421	intergenic	0.059	C	T	0.22(0.05)	$4.04x10^{-5}$	$6.98x10^{-5}$	$4.72x10^{-5}$
FRMD4A	rs1541010	10:13755544	intronic	0.140	T	C	0.15(0.04)	$5.33x10^{-5}$	$9.07x10^{-5}$	$6.50x10^{-5}$
C1orf140/DUSP10	rs143420369	1:221568756	intergenic	0.009	G	C	0.54(0.13)	$5.74x10^{-5}$	$7.00x10^{-5}$	$2.10x10^{-4}$
ARHGAP42	rs4469852	11:100633536	intronic	0.014	C	T	0.45(0.11)	$6.19x10^{-5}$	$4.72x10^{-5}$	$8.07x10^{-5}$
MRPS27	rs3209157	5:71519664	p.Gly284Asp	0.033	T	C	0.29(0.07)	$6.99x10^{-5}$	$1.04x10^{-4}$	$5.93x10^{-5}$
USP33	rs149399157	1:78194235	p.Ser325Gly	0.002	C	T	1.24(0.32)	$8.45x10^{-5}$	$5.47x10^{-4}$	$9.73x10^{-4}$
ARVCF	rs114991369	22:19960536	p.Ala821Val	0.012	A	G	0.46(0.12)	$8.73x10^{-5}$	$1.81x10^{-5}$	$5.63x10^{-5}$
LYPD1	rs772102691	2:133425974	p.Ala11Ala	0.001	C	G	1.75(0.45)	$8.93x10^{-5}$	$3.50x10^{-4}$	$2.07x10^{-3}$
CWF19L2	rs372255196	11:107328524	p.Ala7Thr	0.006	T	C	-0.65(0.17)	$1.19x10^{-4}$	$1.84x10^{-4}$	$5.83x10^{-4}$
EBF1	rs140882557	5:158285031	intronic	0.012	C	T	0.46(0.12)	$1.23x10^{-4}$	$4.77x10^{-4}$	$9.98x10^{-4}$

SUPPLEMENTARY DATA

OC90	rs148636047	8:133067228	p.Gly16Arg	0.001	T	C	1.55(0.41)	$1.43x10^{-4}$	$9.84x10^{-4}$	$2.19x10^{-3}$
H1FNT	rs150155042	12:48723108	p.Arg12Trp	0.012	T	C	-0.45(0.12)	$2.25x10^{-4}$	$3.89x10^{-4}$	$1.08x10^{-3}$
SERPINF1	rs150899084	17:1673212	p.Val51Met	0.024	A	G	-0.32(0.09)	$2.49x10^{-4}$	$2.03x10^{-4}$	$2.23x10^{-4}$
SMG1	rs118055386	16:18864949	p.Asn1575Ser	0.012	C	T	0.43(0.12)	$2.57x10^{-4}$	$5.65x10^{-3}$	$6.70x10^{-3}$
UST/TAB2	rs141544488	6:149537947	intergenic	0.021	G	C	0.34(0.09)	$2.57x10^{-4}$	$6.98x10^{-4}$	$1.83x10^{-3}$
KSR2	rs502394	12:118257893	intronic	0.249	T	A	-0.11(0.03)	$2.57x10^{-4}$	$2.67x10^{-4}$	$2.34x10^{-4}$
EIF3IP1/IMMP2L	rs2940382	7:110122615	intergenic	0.304	A	T	0.10(0.03)	$2.61x10^{-4}$	$2.17x10^{-4}$	$2.57x10^{-4}$
SAFB	rs768659441	19:5641953	p.H181Arg	0.002	G	A	1.15(0.32)	$2.82x10^{-4}$	$5.62x10^{-4}$	$7.01x10^{-4}$
SMYD4	rs58337165	17:1684605	p.Pro797His	0.106	T	G	0.15(0.04)	$2.91x10^{-4}$	$7.27x10^{-5}$	$8.96x10^{-5}$
CCRL2	rs145501356	3:46450429	p.Leu299Ile	0.004	A	C	0.71(0.20)	$3.02x10^{-4}$	$3.35x10^{-3}$	$5.45x10^{-3}$
FTO	rs17817964	16:53828066	intronic	0.165	T	C	0.13(0.04)	$3.02x10^{-4}$	$7.97x10^{-3}$	$1.11x10^{-2}$
KLHL40	rs201856772	3:42733381	p.Glu588Lys	0.001	A	G	1.36(0.38)	$3.30x10^{-4}$	$4.21x10^{-4}$	$2.05x10^{-3}$
WDR86	rs4726024	7:151102867	intronic	0.459	G	A	-0.09(0.03)	$3.34x10^{-4}$	$1.42x10^{-3}$	$2.90x10^{-3}$
TMC5	rs141710338	16:19483429	p.Arg355His	0.003	A	G	0.92(0.26)	$3.97x10^{-4}$	$1.24x10^{-4}$	$9.98x10^{-5}$
PHACTR1	rs145126821	6:12934291	intronic	0.006	A	G	0.58(0.17)	$4.48x10^{-4}$	$1.18x10^{-3}$	$1.17x10^{-3}$
CDKN2B-AS1	rs180698367	9:22105256	ncRNA	0.002	A	G	1.06(0.30)	$4.51x10^{-4}$	$8.63x10^{-4}$	$1.04x10^{-3}$
TLE2	rs118030930	19:3011146	p.Ala310Thr	0.053	T	C	0.20(0.06)	$4.56x10^{-4}$	$1.27x10^{-3}$	$1.16x10^{-3}$
HEATR4	rs769705232	14:73962022	p.Val899Leu	0.003	G	C	0.83(0.24)	$4.59x10^{-4}$	$3.27x10^{-3}$	$7.39x10^{-3}$
WDR64	rs12095445	1:241929542	p.Arg647Gln	0.002	A	G	1.17(0.33)	$4.60x10^{-4}$	$8.95x10^{-4}$	$3.86x10^{-4}$

ZDHHC8	rs175174	22:20127554	intronic	0.293	A	G	0.10(0.03)	$4.76x10^{-4}$	$8.69x10^{-4}$	$1.12x10^{-3}$

A1: Minor allele; A2: Major allele; MAF: Minor allele frequency; β: beta; BMI: body mass index; ncRNA: non-coding RNA; PC: principal component. The βs are reported with respect to the minor allele. *Adjusted for age, gender, PC1 and PC2. †Adjusted for age, gender, BMI, PC1 and PC2. ‡Adjusted for age, gender, BMI, use of metformin, PC1 and PC2. SNPs with $P_{discovery}<5x10^{-5}$ in adjustment Model 1 and thus selected for replication are underlined.

SUPPLEMENTARY DATA

Supplementary Table 2. Genes demonstrating suggestive associations ($P<5x10^{-4}$) with circulating PEDF level in the gene-based association test.

Gene	No. of variants	Beta(SE)	*P-value*	Best gene-based test*
ITGAV	3	-0.46(0.14)	$3.83x10^{-5}$	SKAT <1%
RAB22A	1	1.79(0.45)	$6.09x10^{-5}$	Burden <1%
NQO2	4	0.22(0.08)	$6.80x10^{-5}$	SKAT < 5%
PTPRM	2	-0.02(0.19)	$1.00x10^{-4}$	VT < 5%
ARVCF	4	0.32(0.10)	$1.14x10^{-4}$	SKAT < 5%
GPSM1	6	-0.26(0.07)	$1.14x10^{-4}$	Burden < 5%
USP33	2	0.85(0.24)	$1.43x10^{-4}$	SKAT <1%
OC90	2	0.62(0.25)	$1.56x10^{-4}$	SKAT < 5%
SERPINF1	3	-0.32(0.08)	$1.58x10^{-4}$	Burden < 5%
PI4K2B	2	-0.67(0.18)	$1.90x10^{-4}$	Burden <1%
TUBGCP6	6	-0.25(0.10)	$2.00x10^{-4}$	SKAT <5%
COL3A1	6	0.15(0.05)	$2.00x10^{-4}$	SKAT < 5%
MRPS27	6	0.21(0.06)	$2.19x10^{-4}$	Burden < 5%
SMAD9	2	0.54(0.15)	$2.37x10^{-4}$	Burden <1%
CAD	5	-0.42(0.11)	$2.43x10^{-4}$	Burden <1%
WDFY3	5	-0.28(0.08)	$2.56x10^{-4}$	Burden < 5%
SAFB	1	1.15(0.32)	$2.76x10^{-4}$	Burden <1%
KRT82	1	0.78(0.45)	$3.00x10^{-4}$	SKAT < 5%
TLE6	3	-0.76(0.21)	$3.46x10^{-4}$	Burden <1%
KLK6	2	0.62(0.29)	$4.00x10^{-4}$	VT < 5%
CFAP221	6	-0.25(0.07)	$4.04x10^{-4}$	Burden < 5%

Gene-based tests (Burden [CMC-Wald], SKAT or VT) for damaging and missense variants with <1% or <5% MAF. Exome-wide significance thresholds were defined as 0.05/10,039 = $4.98x10^{-6}$ and 0.05/11,157 = $4.40x10^{-6}$, respectively.

Supplementary Table 3. F-statistics and R^2-values between age and sex-standardized PEDF level and the PEDF-associated SNPs in the HKWDR cohort (n=5385).

SNP	Effect allele*	Other allele	F-statistic	R^2	β(SE)	*P-value*
rs1136287	T	C	301.028	0.053	0.33(0.02)	<0.001
rs7224496	A	C	113.92	0.021	0.22(0.02)	<0.001
rs2070863	C	T	34.763	0.006	0.16(0.03)	<0.001

*Allele associated with increasing PEDF level. β: Beta; SE: Standard error. The βs are reported with respect to the effect allele.

SUPPLEMENTARY DATA

Supplementary Table 4. Association of PEDF-associated SNPs with DN and STDR.

	Effect allele*	Other allele	DN 2914 cases VS 2408 controls		STDR 772 cases VS 4317 controls	
			OR(95%CI)	$P_{unadjusted}$	OR(95%CI)	$P_{unadjusted}$
rs1136287	T	C	1.02(0.95-1.10)	0.606	0.96(0.86-1.08)	0.509
rs7224496	A	C	1.03(0.95-1.11)	0.539	0.92(0.82-1.03)	0.155
rs2070863	C	T	0.96(0.86-1.07)	0.435	0.95(0.82-1.10)	0.462

*Allele associated with increasing PEDF level. OR: odds ratio; 95%CI: 95% confidence interval. The ORs are reported with respect to the effect allele.

Supplementary table 5. Association of the PEDF-associated SNPs with the potential confounders for DN and STDR.

	rs1136287 (T)		rs7224496 (A)		rs2070863 (C)	
	β(SE)	*P-value*	β(SE)	*P-value*	β(SE)	*P-value*
Age	0.140(0.247)	0.572	0.472(0.263)	0.072	0.285(0.338)	0.399
Duration of diabetes*	-0.013(0.016)	0.442	0.002(0.017)	0.886	-0.046(0.022)	0.041
eGFR*	-0.001(0.009)	0.949	-0.012(0.01)	0.193	-0.0003(0.012)	0.979
HbA1c*	0.002(0.003)	0.526	0.001(0.003)	0.884	-0.005(0.005)	0.302
BMI	-0.067(0.085)	0.431	-0.118(0.09)	0.191	0.108(0.116)	0.350
	OR(95%CI)	***P-value***	**OR(95%CI)**	***P-value***	**OR(95%CI)**	***P-value***
HT	1.05(0.93-1.17)	0.451	1.04(0.92-1.18)	0.509	0.91(0.77-1.07)	0.244
Gender	1.01(0.93-1.09)	0.907	0.98(0.90-1.06)	0.576	1.00(0.90-1.12)	0.940

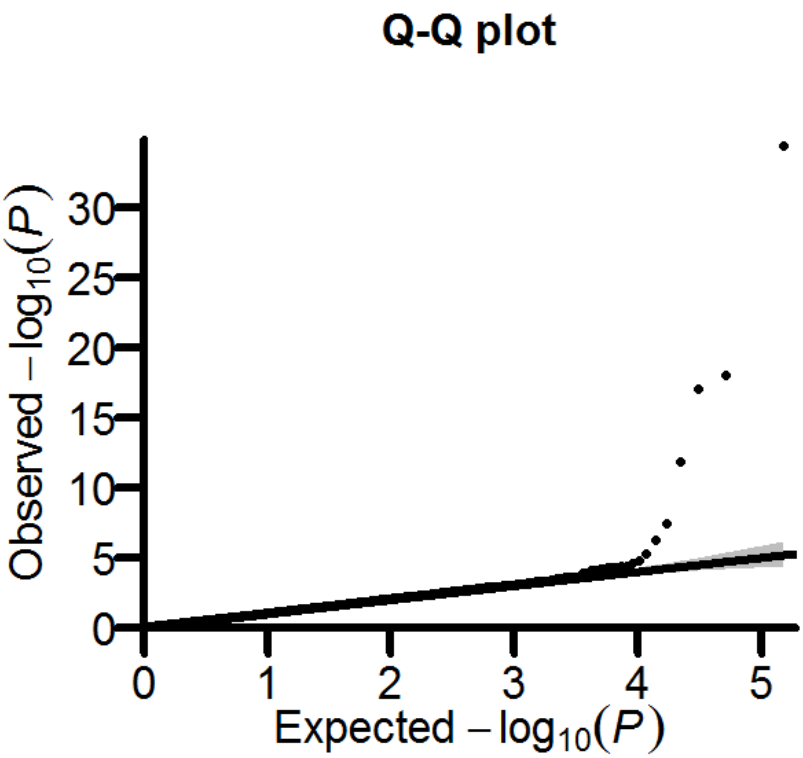
*Natural-log-transformed before analysis. eGFR: Estimated glomerular filtration rate; HbA1c: Glycated hemoglobin; BMI: Body mass index; HT: hypertension. OR: odds ratio; 95%CI: 95% confidence interval; β: Beta; SE: Standard error. The βs/ORs are reported with respect to the effect allele.

SUPPLEMENTARY DATA

Supplementary Table 6. MR-Egger pleiotropy test of associations between a SD increase in genetically determined age and sex-standardized PEDF level and risk of DN and STDR in the HKWDR cohort.

	Intercept(95%CI)	*P-value*
DN	0.22(-0.59, 1.02)	0.600
STDR	0.21(-1.84, 2.26)	0.841

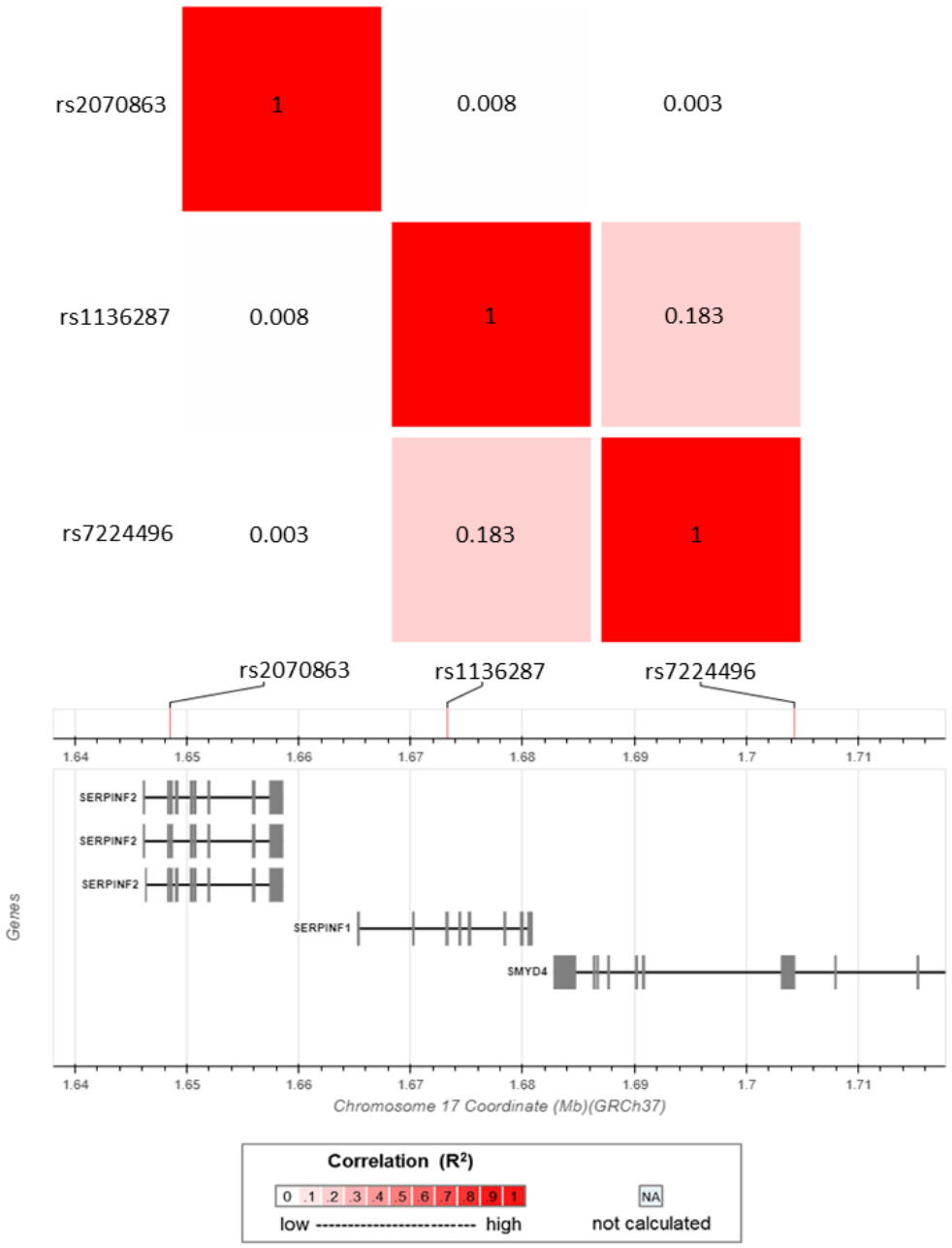
Supplementary Figure 1. Quantile-Quantile plot



Quantile-Quantile plot of association *P*-values of all tested variants for circulating PEDF level.

SUPPLEMENTARY DATA

Supplementary Figure 2. Heatmap matrix of pairwise linkage disequilibrium (LD) statistics of the 3 PEDF-associated SNPs that achieved genome-wide significant.



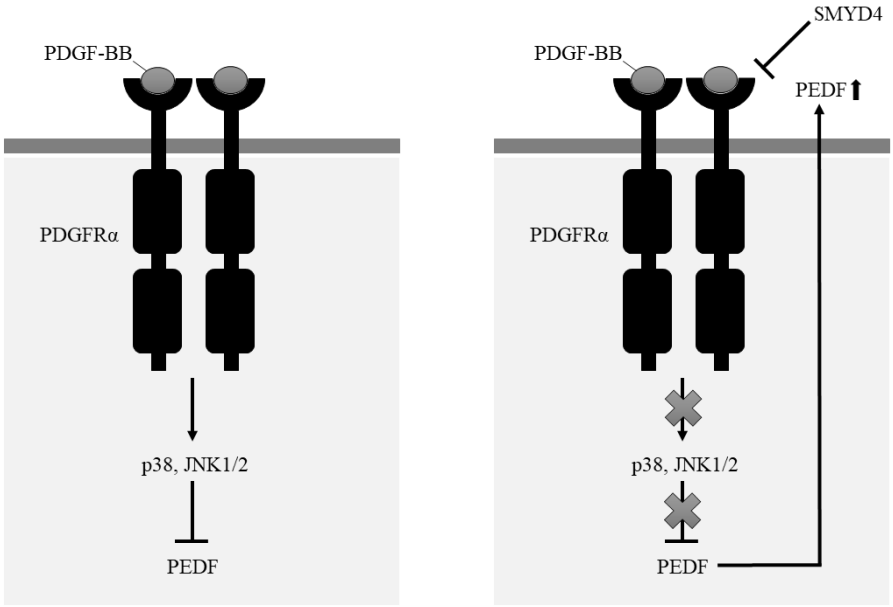
The heatmap matrix was generated by LDlink (1) using the CHS population of the 1000 Genome Project as reference population. Value within each block represents the pairwise LD statistics between the SNPs.

Reference:

1. Machiela MJ, Chanock SJ: LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015; 31(21):3555-7.

SUPPLEMENTARY DATA

Supplementary Figure 3. Illustration of possible SMYD4 regulation on PEDF expression



PEDF: pigment epithelium-derived factor; PDGFRα: platelet-derived growth factor receptor alpha; PDGF, platelet-derived growth factor; JNK: c-Jun N-terminal kinase.
Left panel: PEDF expression has been shown to be repressed by PDGF-BB signaling, with p38 and JNK 1/2 found to be critical in this PDGF-BB mediated repression.
Right panel: SMYD4 has been proposed to function as a tumor suppressor gene through inhibition of PDGFRα. Increased SMYD4 might led to a decreased expression of PDGFRα, and hence decreases PDGF-BB activation of JNK-1 and p38, causing a reduction in PDGF-BB mediated repression of PEDF, thereby increases PEDF expression and secretion