Contents

Supplementary Methods

Polygenic risk score (PRS) calculation

We calculated PRS in an additive genetic model with the formula below:

$$PRS_j = \sum_i S_i \times G_{ij}$$

j refers to the individual to be calculated. i refers to each variant selected in the calculation. S refers to the summary statistic for the effective allele (in our study we used Odds Ratio<OR> as the weight of each variant). G refers to the number of effective allele observed (0/1/2).

Besides the information of variant position and OR, the summary statistics also contain the p-value of variants in GWAS to represent the correlation between the variants and the disease. By setting the p-value threshold range (from 5e-08, with step size of 5e-05) within PRSice-2's command, it could find the best-fit model of certain SNP included.

The output file "[prefix].prsice" summarizes the p-value and R2 when different variant's p-value thresholds were selected. Here is an example of our output ".prsice" file:

Pheno	Set	Threshold	R2	P	Coefficient	Standard.Error	Num_SNP
-	Base	1.00E-15	0.00106923	0.349115	0.50458	0.538905	3
-	Base	5.00E-10	0.0076559	0.0124479	0.892064	0.356942	13
-	Base	4.00E-09	0.00722904	0.0151393	0.826793	0.34038	14
-	Base	6.00E-09	0.00824909	0.00950252	0.857054	0.330472	16
-	Base	1.10E-08	0.00875608	0.00755386	0.882742	0.330443	17
-	Base	2.45E-08	0.00847331	0.00858189	0.866742	0.329776	18
-	Base	2.55E-08	0.0119503	0.00182795	0.986313	0.316445	19
-	Base	3.60E-08	0.0130818	0.00111745	1.02463	0.314384	20

In the output, the variant's p-value threshold of 3.60E-08 showed the lowest P value (0.00111745) of the whole PRS model, which includes 20 SNPs. PRSice-2 also outputted a plot to visualize the correlation between model fit R2 and p-value

threshold (**Supplementary Figure 2**). Using "--print-snp" command, the variant information of the best-fit model could be output in another ".snp" file.

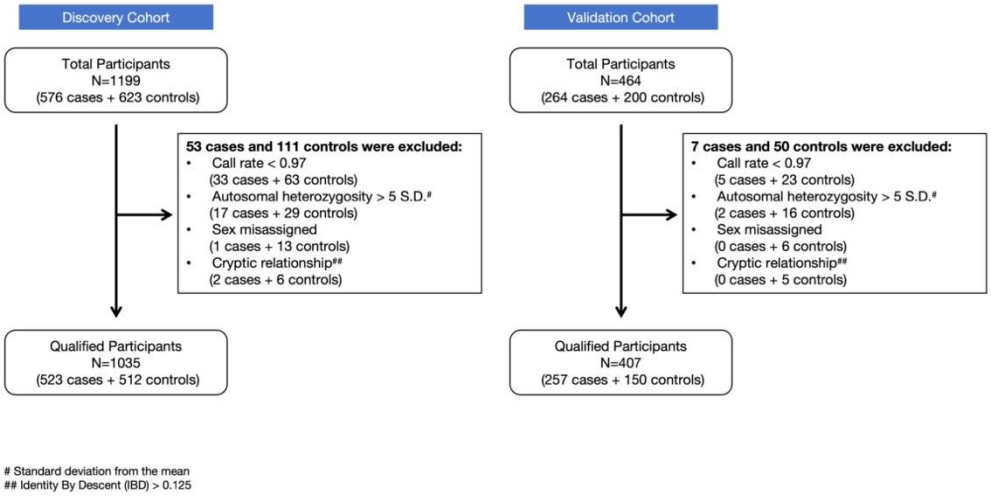
Polygenic risk score (PRS) adjustment

Given that the samples were recruited from over 54,000 women from six major geographical regions of the Chinese mainland (Northeast, North, East, South Central, Northwest and Southwest), inner population stratification might exist. Additionally, in the validation cohort, controls were genotyped using the Affymetrix Axiom Genome-Wide CHB 1 Array Plate (CHB-1), while cases were genotyped with a customized Illumina Infinium Asian Screening Array (ASA), which may introduce batch effects.

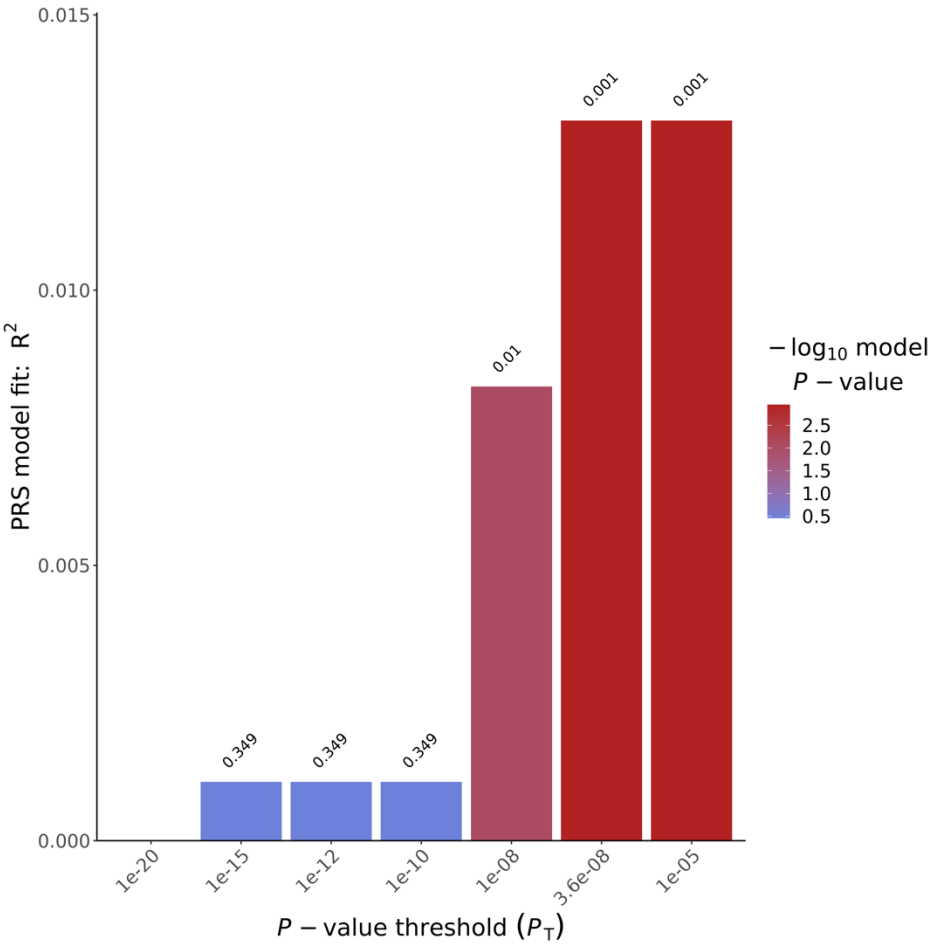
To verify the genetic homogeneity and to adjust for population stratification in subsequent analyses, we conducted principal component analysis (PCA) on cases and controls for both the discovery and validation cohorts (**Supplementary Figure 3**). For the discovery cohort, the PCA plot indicated good consistency and overlap between cases and controls, with no significant population stratification detected. For the validation cohort, though PCA did not show significant population stratification within the Chinese population, it did reveal a separation between cases and controls, which suggests potential batch effects due to different genotyping methods.

To mitigate these effects, we included the top three principal components (PC1-3) along with age as covariates to adjust PRS in both discovery and validation cohorts. This adjustment was performed using the "--cov" command in PRSice software, as detailed in its documentation.

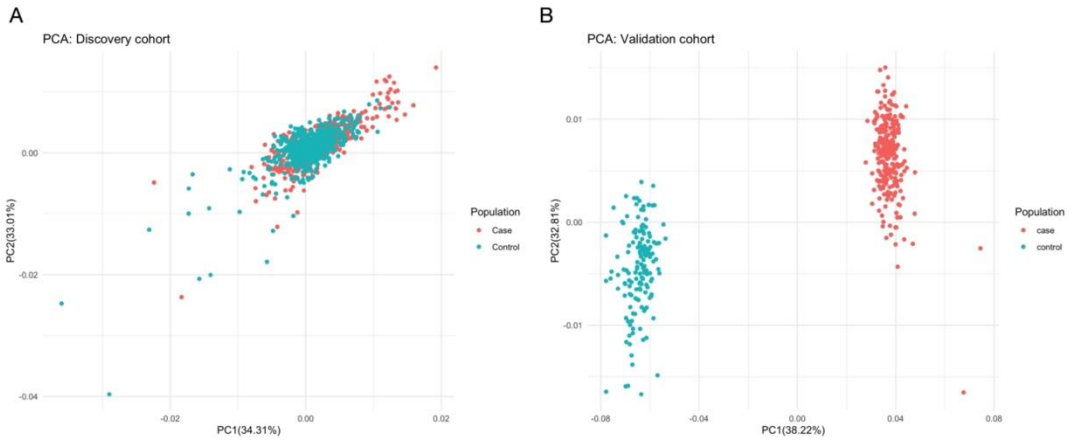
Supplementary figures



Supplementary figure 1. Flowchart of sample-level quality control. The flowchart illustrates the exclusion criteria and the number of exclusions in the discovery cohort (left panel) and the validation cohort (right panel).



Supplementary figure 2. Best-fit PRS model selection by PRSice-2. The barplot illustrated the correlation of PRS model fit with a range of p-value thresholds. The numbers above the plot show the whole PRS model's p-value.



Supplementary figure 3. Principal component analysis (PCA) of polygenic risk score (PRS) distribution. The scattered plots illustrate the distribution of the first two principal components (PC1 and PC2) for cases and controls in both the discovery cohort (Panel A) and the validation cohort (Panel B).

Supplementary Table 1. Genetic variants selected for PRS construction

CHR	POS	Effect allele	rsID	Allele Frequency (gnomAD_NFE)	Allele Frequency (gnomAD_EAS)	Allele Frequency (NyuWa_EAS)	Prioritized gene(s)	Function
2	20878105	G	rs9306894	0.37	0.4718	0.4508	*GDF7, LDAH*	GDF7: ligand of the transforming growth factor-beta superfamily; LDAH: lipid droplet associated hydrolase
2	56111309	T	rs3791675	0.2498	0.756	0.7896	*EFEMP1*	Member of the fibulin family of extracellular matrix glycoproteins
4	126924684	C	rs28403275	0.1964	0.2502	0.2302	*FAT4*	Member of the protocadherin family, role in regulating planar cell polarity, epithelial-to-mesenchymal transition, cancer suppressor[2]
12	114673421	A	rs1247943	0.5306	0.7381	0.7284	*TBX5*	Member of a phylogenetically conserved family, role in heart development and specification of limb identity[3]
3	127721333	A	rs58170120	0.154	0.4436	0.4243	*SEC61A1*	sec61 translocon alpha 1 subunit, the main polypeptide-conducting channel in the endoplasmic reticulum membrane. Mutations in SEC61A1 were reported to be pathogenic in common variable immunodeficiency and glomerulocystic kidney disease[4]

4	66666895	T	rs201194999	0.1322	0.4593	NA	*EPHA5*	A receptor tyrosine kinase, role in tumorigenesis and developmental events, particularly in the nervous system[5]
4	127076188	G	rs10013769	0.6869	0.8911	0.8783	*FAT4*	Member of the protocadherin family, role in regulating planar cell polarity, epithelial-to-mesenchymal transition, cancer suppressor[2]
5	4978695	G	rs42400	0.3528	0.4418	0.4238	*ADAMTS16*	Role in extracellular matrix maintenance
5	127512064	G	rs251217	0.5545	0.2119	0.1787	*SLC12A2, FBN2*	SLC12A2: membrane protein mediating sodium and chloride transport and reabsorption, maintaining proper ionic balance and cell volume, role in corticogenesis and neurodevelopmental[6] FBN2: fibrillin 2, component of connective tissue microfibrils
8	71905587	G	rs1493202	0.5053	0.3376	0.3418	*LACTB2-AS1*	LACTB2 antisense RNA 1, LACTB2 is involved in the turnover of mitochondrial RNA, and is essential for mitochondrial function[7]

11	74392514	C	rs4944936	0.7934	0.6049	0.5559	*CHRDL2*	Secreted proteins that share a cysteine-rich pro-collagen repeat domain and associate with members of the transforming growth factor beta superfamily, bone morphogenetic protein antagonist, preventing from interacting with their homologous cell surface receptors, role in adult cartilage regeneration[8]
11	10308991	T	rs6484161	0.3553	0.2433	0.2439	*SBF2, ADM*	SBF2: member of the myotubularin-related protein family[9] ADM: a preprohormone which is cleaved to form two biologically active peptides, adrenomedullin and proadrenomedullin N-terminal 20 peptide[10]
11	32479807	A	rs11031796	0.3717	0.7475	0.7366	*WT1*	Urogenital system development, cardiac development and disease
11	32346397	C	rs35166569	0.04123	0.6816	0.6754	*WT1*	Urogenital system development, cardiac development and disease
12	12668410	T	rs12314243	0.11	0.4304	0.4528	*DUSP16*	Inhibiting JNK and p38 activation[11]

15	33023486	A	rs12915554	0.3583	0.02921	0.0295	*GREM1*	A member of the BMP (bone morphogenic protein) antagonist family, role in regulating organogenesis, body patterning, and tissue differentiation[12]
15	31636424	G	rs4779517	0.5088	0.4867	0.5195	*KLF13*	Transcription factor, member of the Krüppel-like family of zinc finger protein, genetically defective KLF13 predisposes to familial congenital heart defect[13]
16	84940479	C	rs1874008	0.7494	0.7916	0.7931	*CRISPLD2*	Involved in face morphogenesis[14]
21	28636684	C	rs235929	0.4274	0.3181	0.3301	*ADAMTS5*	Mediating aggrecan cleavage, role in extracellular matrix maintenance[15]
22	38598234	G	rs2267372	0.6002	0.7438	0.7437	*MAFF,PLA2G6*	MAFF: cellular stress response, promoting antioxidant responses, HIF-1 target gene, increasing the invasive and metastatic behavior of tumor cells[16, 17] PLA2G6: A2 phospholipase, role in phospholipid remodelling[18]

Abbreviations: CHR, chromosome; POS, position(hg19); rsID, reference SNP cluster ID; gnomAD_NFE, effect allele frequency of non-Finnish European in gnomAD database (v3.1.2); gnomAD_EAS, effect allele frequency of East Asian in gnomAD database; NyuWa_EAS, effect allele frequency of East Asian in NyuWa database; NA, not applicable.

References

1. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Aguilar Salinas CA, Ahmad T, Albert CM, Ardissino D, Atzmon G, Barnard J, Beaugerie L, Benjamin EJ, Boehnke M, Bonnycastle LL, Bottinger EP, Bowden DW, Bown MJ, Chambers JC, Chan JC, Chasman D, Cho J, Chung MK, Cohen B, Correa A, Dabelea D, Daly MJ, Darbar D, Duggirala R, Dupuis J, Ellinor PT, Elosua R, Erdmann J, Esko T, Färkkilä M, Florez J, Franke A, Getz G, Glaser B, Glatt SJ, Goldstein D, Gonzalez C, Groop L, Haiman C, Hanis C, Harms M, Hiltunen M, Holi MM, Hultman CM, Kallela M, Kaprio J, Kathiresan S, Kim B-J, Kim YJ, Kirov G, Kooner J, Koskinen S, Krumholz HM, Kugathasan S, Kwak SH, Laakso M, Lehtimäki T, Loos RJF, Lubitz SA, Ma RCW, MacArthur DG, Marrugat J, Mattila KM, McCarroll S, McCarthy MI, McGovern D, McPherson R, Meigs JB, Melander O, Metspalu A, Neale BM, Nilsson PM, O'Donovan MC, Ongur D, Orozco L, Owen MJ, Palmer CNA, Palotie A, Park KS, Pato C, Pulver AE, Rahman N, Remes AM, Rioux JD, Ripatti S, Roden DM, Saleheen D, Salomaa V, Samani NJ, Scharf J, Schunkert H, Shoemaker MB, Sklar P, Soininen H, Sokol H, Spector T, Sullivan PF, Suvisaari J, Tai ES, Teo YY, Tiinamaija T, Tsuang M, Turner D, Tusie-Luna T, Vartiainen E, Vawter MP, Ware JS, Watkins H, Weersma RK, Wessman M, Wilson JG, Xavier RJ, Neale BM, Daly MJ, MacArthur DG, Genome Aggregation Database C. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-43.
2. Wei R, Xiao Y, Song Y, Yuan H, Luo J, Xu W. FAT4 regulates the EMT and autophagy in colorectal cancer cells in part via the PI3K-AKT signaling axis. J Exp Clin Cancer Res. 2019;38(1):112.
3. van Ouwerkerk AF, Bosada FM, van Duijvenboden K, Houweling AC, Scholman KT, Wakker V, Allaart CP, Uhm JS, Mathijssen IB, Baartscheer T, Postma AV, Barnett P, Verkerk AO, Boukens BJ, Christoffels VM. Patient-Specific TBX5-G125R Variant Induces Profound Transcriptional Deregulation and Atrial Dysfunction. Circulation. 2022;145(8):606-19.
4. Van Nieuwenhove E, Barber JS, Neumann J, Smeets E, Willemsen M, Pasciuto E, Prezzemolo T, Lagou V, Seldeslachts L, Malengier-Devlies B, Metzemaekers M, Hassdenteufel S, Kerstens A, van der Kant R, Rousseau F, Schymkowitz J, Di Marino D, Lang S, Zimmermann R, Schlenner S, Munck S, Proost P, Matthys P, Devalck C, Boeckx N, Claessens F, Wouters C, Humblet-Baron S, Meyts I, Liston A. Defective Sec61alpha1 underlies a novel cause of autosomal dominant severe congenital neutropenia. J Allergy Clin Immunol. 2020;146(5):1180-93.
5. Li Y, Chu J, Feng W, Yang M, Zhang Y, Zhang Y, Qin Y, Xu J, Li J, Vasilatos SN, Fu Z, Huang Y, Yin Y. EPHA5 mediates trastuzumab resistance in HER2-positive breast cancers through regulating cancer stem cell-like properties. FASEB J. 2019;33(4):4851-65.

6. McNeill A, Iovino E, Mansard L, Vache C, Baux D, Bedoukian E, Cox H, Dean J, Goudie D, Kumar A, Newbury-Ecob R, Fallerini C, Renieri A, Lopergolo D, Mari F, Blanchet C, Willems M, Roux AF, Pippucci T, Delpire E. SLC12A2 variants cause a neurodevelopmental disorder or cochleovestibular defect. Brain. 2020;143(8):2380-7.
7. Levy S, Allerston CK, Liveanu V, Habib MR, Gileadi O, Schuster G. Identification of LACTB2, a metallo-beta-lactamase protein, as a human mitochondrial endoribonuclease. Nucleic Acids Res. 2016;44(4):1813-32.
8. Wang L, Xu W, Mei Y, Wang X, Liu W, Zhu Z, Ni Z. CHRDL2 promotes cell proliferation by activating the YAP/TAZ signaling pathway in gastric cancer. Free Radic Biol Med. 2022;193(Pt 1):158-70.
9. Berger P, Berger I, Schaffitzel C, Tersar K, Volkmer B, Suter U. Multi-level regulation of myotubularin-related protein-2 phosphatase activity by myotubularin-related protein-13/set-binding factor-2. Hum Mol Genet. 2006;15(4):569-79.
10. Voors AA, Kremer D, Geven C, Ter Maaten JM, Struck J, Bergmann A, Pickkers P, Metra M, Mebazaa A, Dungen HD, Butler J. Adrenomedullin in heart failure: pathophysiology and therapeutic application. Eur J Heart Fail. 2019;21(2):163-71.
11. Low HB, Wong ZL, Wu B, Kong LR, Png CW, Cho YL, Li CW, Xiao F, Xin X, Yang H, Loo JM, Lee FYX, Tan IBH, DasGupta R, Shen HM, Schwarz H, Gascoigne NRJ, Goh BC, Xu X, Zhang Y. DUSP16 promotes cancer chemoresistance through regulation of mitochondria-mediated cell death. Nat Commun. 2021;12(1):2284.
12. Lan L, Evan T, Li H, Hussain A, Ruiz EJ, Zaw Thin M, Ferreira RMM, Ps H, Riising EM, Zen Y, Almagro J, Ng KW, Soro-Barrio P, Nelson J, Koifman G, Carvalho J, Nye EL, He Y, Zhang C, Sadanandam A, Behrens A. GREM1 is required to maintain cellular heterogeneity in pancreatic cancer. Nature. 2022;607(7917):163-8.
13. Wang SS, Wang TM, Qiao XH, Huang RT, Xue S, Dong BB, Xu YJ, Liu XY, Yang YQ. KLF13 loss-of-function variation contributes to familial congenital heart defects. Eur Rev Med Pharmacol Sci. 2020;24(21):11273-85.
14. Gaikwad AS, Hu J, Chapple DG, O'Bryan MK. The functions of CAP superfamily proteins in mammalian fertility and disease. Hum Reprod Update. 2020;26(5):689-723.
15. Jiang L, Lin J, Zhao S, Wu J, Jin Y, Yu L, Wu N, Wu Z, Wang Y, Lin M. ADAMTS5 in Osteoarthritis: Biological Functions, Regulatory Network, and Potential Targeting Therapies. Front Mol Biosci. 2021;8:703110.
16. Kannan MB, Solovieva V, Blank V. The small MAF transcription factors MAFF, MAFG and MAFK: current knowledge and perspectives. Biochim Biophys Acta. 2012;1823(10):1841-6.
17. Moon EJ, Mello SS, Li CG, Chi JT, Thakkar K, Kirkland JG, Lagory EL, Lee IJ, Diep AN, Miao Y, Rafat M, Vilalta M, Castellini L, Krieg AJ, Graves EE, Attardi LD, Giaccia AJ. The HIF target MAFF promotes tumor invasion and metastasis through IL11 and STAT3 signaling. Nat Commun. 2021;12(1):4308.
18. Beharier O, Tyurin VA, Goff JP, Guerrero-Santoro J, Kajiwara K, Chu T, Tyurina YY, St Croix CM, Wallace CT, Parry S, Parks WT, Kagan VE, Sadovsky Y. PLA2G6 guards placental trophoblasts against ferroptotic injury. Proc Natl Acad Sci U S A. 2020;117(44):27319-28.