Supplementary Information

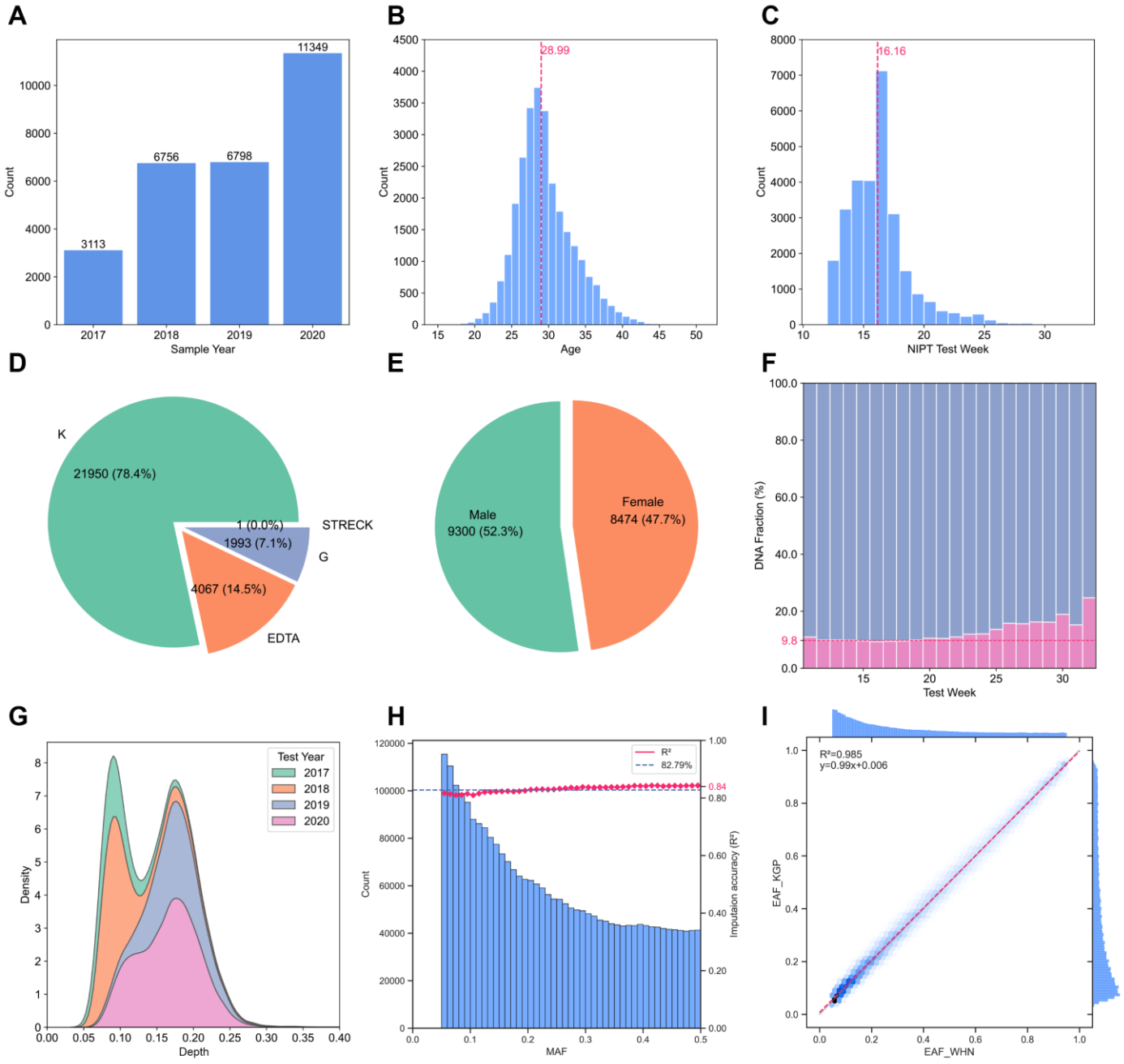
cfGWAS reveal genetic basis of cell-free DNA end motifs

Huanhuan Zhu[1,2]†, Yan Zhang[1,2]†, Linxuan Li[2,3]†, Shuang Zeng[4,5]†, Xinyi Zhang[2,3]†, Ying Lin[2,3]†, Rijing Ou[2]†, Lin Wang[2,6], Xiameizi Li[2,3], Yu Wang[2], Jingyu Zeng[2], Yu Lin[7], Chuang Xu[7,8], Guodan Zeng[2,3], Lingguo Li[2,3], Rongkang Zhao[2,3], Yongjian Jia[7], Fei Luo[4], Meng Yang[9], Yuxuan Hu[7], Han Xiao[9], Xun Xu[1], Jian Wang[10], Aifen Zhou[9,11]*, Haiqiang Zhang[2]*, Xin Jin[1,2,12,13,14]*

1. State Key Laboratory of Genome and Multi-omics Technologies, BGI Research, Shenzhen 518083, China
2. Shenzhen Key Laboratory of Transomics Biotechnologies, BGI Research, Shenzhen 518083, China
3. College of Life Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
4. BGI Research, Guiyang 550000, China
5. BGI Research, Chongqing 401329, China
6. School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, China
7. BGI Research, Shenzhen 518083, China
8. BGI Research, Wuhan 430074, China
9. Institute of Maternal and Child Health, Wuhan Children's Hospital (Wuhan Maternal and Child Health Care Hospital), Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430010, China
10. BGI-Shenzhen, Shenzhen 518083, China
11. Department of Obstetrics, Wuhan Children's Hospital (Wuhan Maternal and Child Health Care Hospital), Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430010, China
12. The Innovation Centre of Ministry of Education for Development and Diseases, School of Medicine, South China University of Technology, Guangzhou 510006, China
13. Shanxi Medical University-BGI Collaborative Center for Future Medicine, Shanxi Medical University, Taiyuan 030001, China
14. State Key Laboratory of Vascular Homeostasis and Remodeling, Peking University Health Science Center, Beijing 100191, China

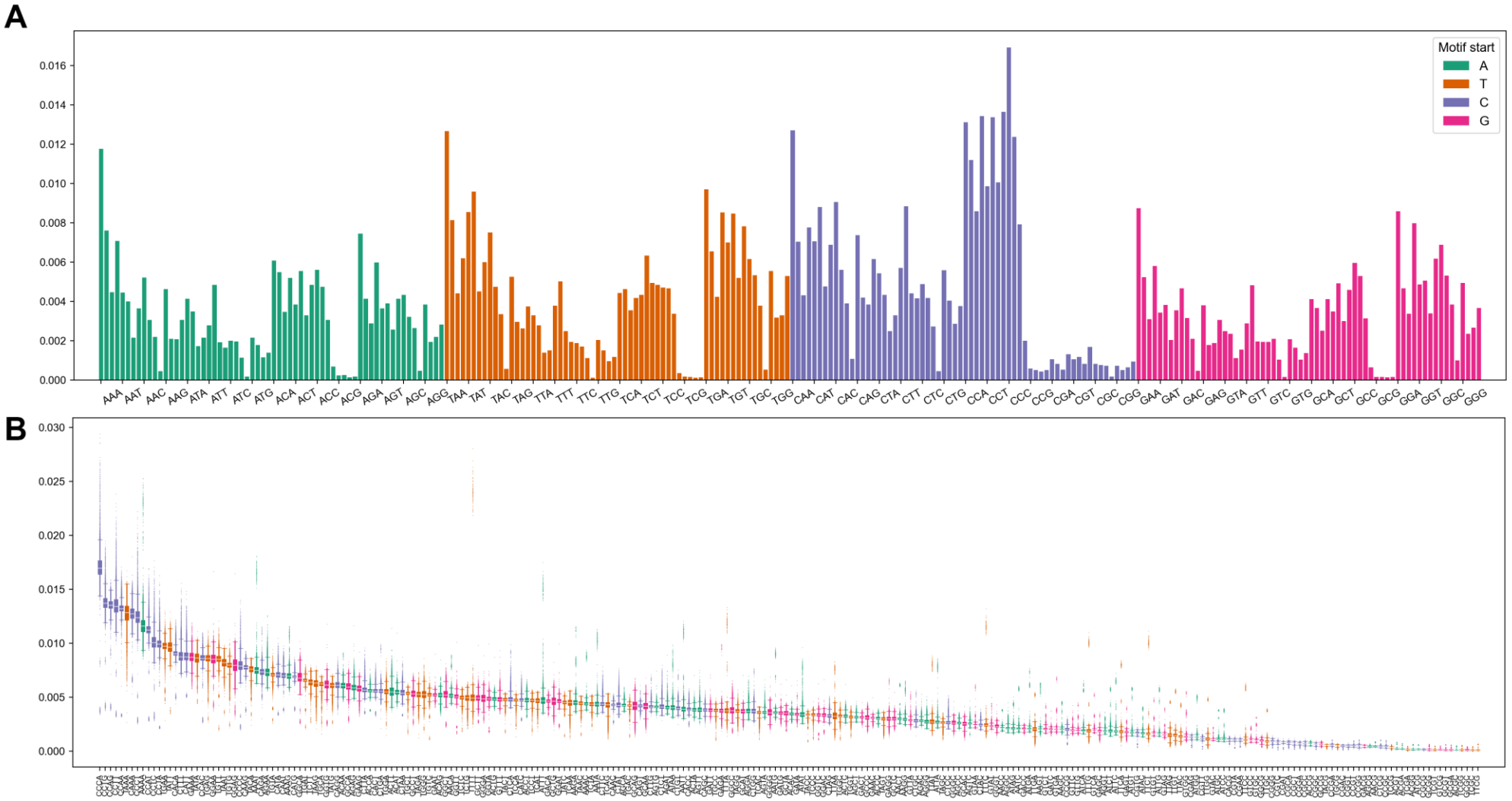
†These authors contributed equally.
*These authors are co-corresponding authors.

Correspondence:
Xin Jin (jinxin@genomics.cn)
Haiqiang Zhang (zhanghaiqiang@genomics.cn)
Aifen Zhou (april1972@163.com)



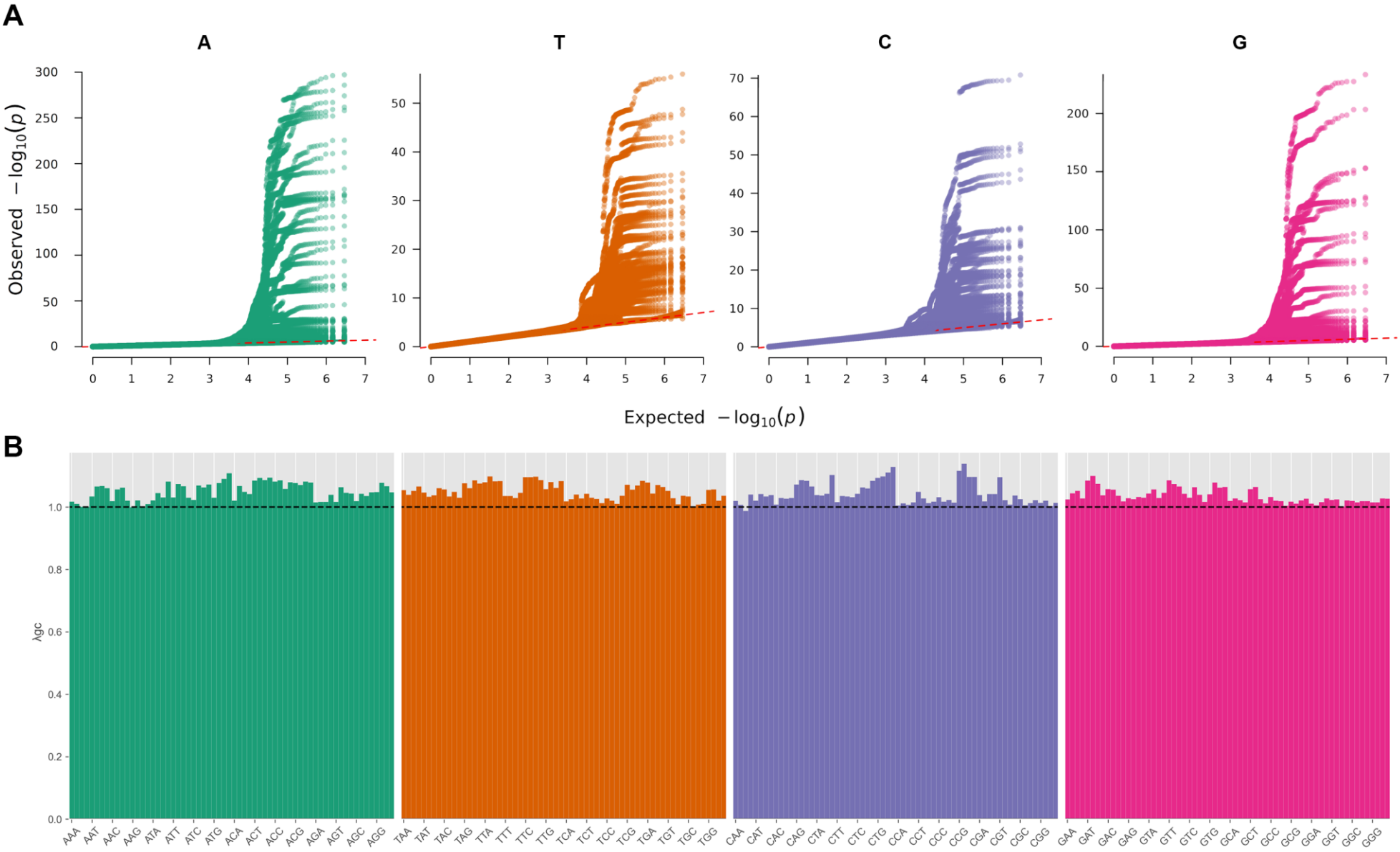
Supplementary Figure 1. The information of samples and variants in the discovery study

A) number of participants by sampling year, **B**) age distribution of participants, **C**) gestational weeks at the time of NIPT testing, **D**) types of blood collection tubes used, **E**) sex distribution of newborns, **F**) maternal and fetal DNA fractions by NIPT test week, **G**) sequencing depth by sampling year, **H**) imputation performance of imputed variants, and **I**) comparison of alternate allele frequencies between the discovery study and the 1000 Genomes Project East Asian population.



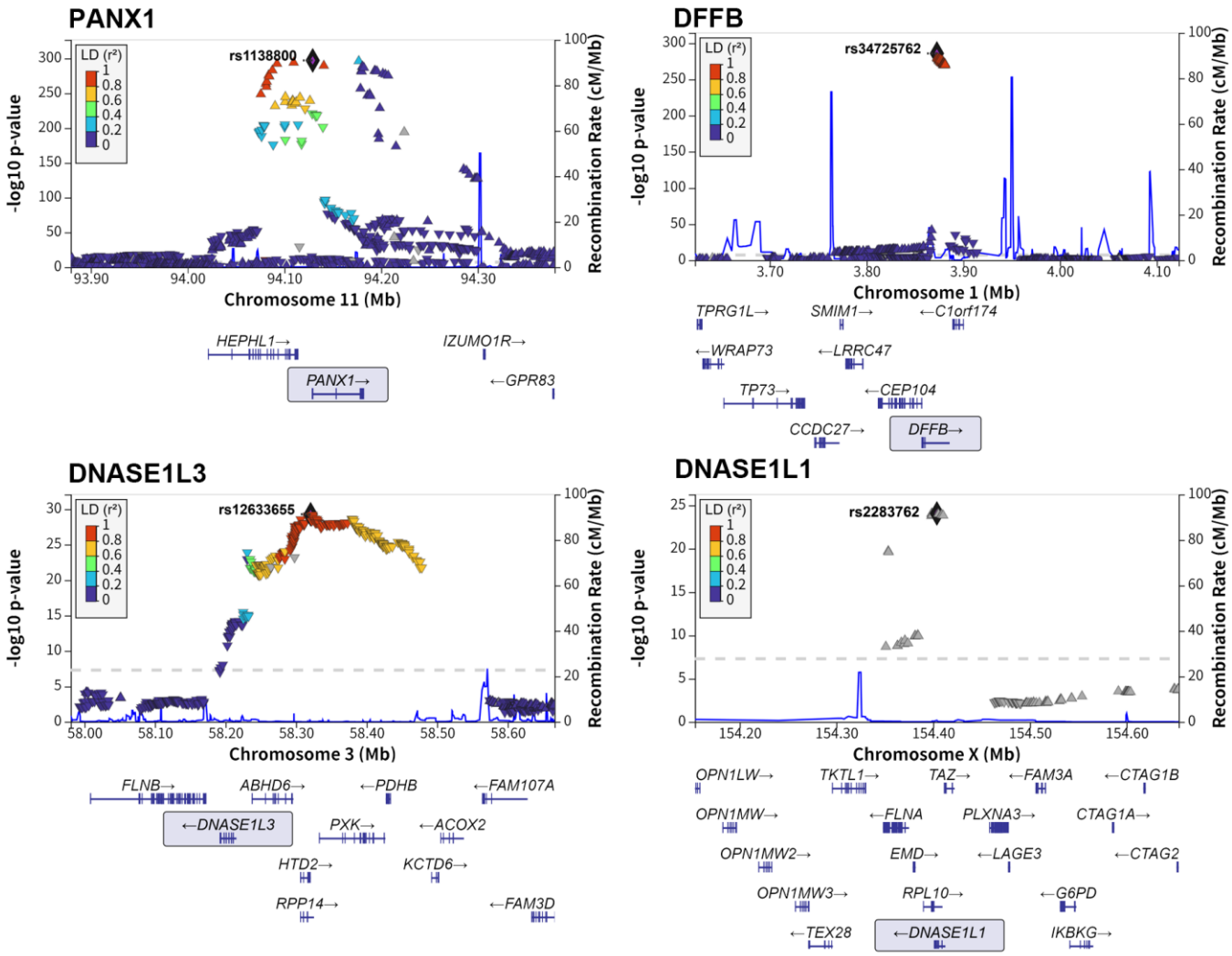
Supplementary Figure 2. Distributions of 256 cfDNA end motif frequencies

A) bar plots and **B**) box plots illustrating the frequency distribution of each 4-mer cfDNA end motif (n=28,016). In the box plots, the center line, box limits, and whiskers represent the median, 25th/75th percentiles, and the minimum/maximum values within 1.5× interquartile range (IQR) from the box, respectively.



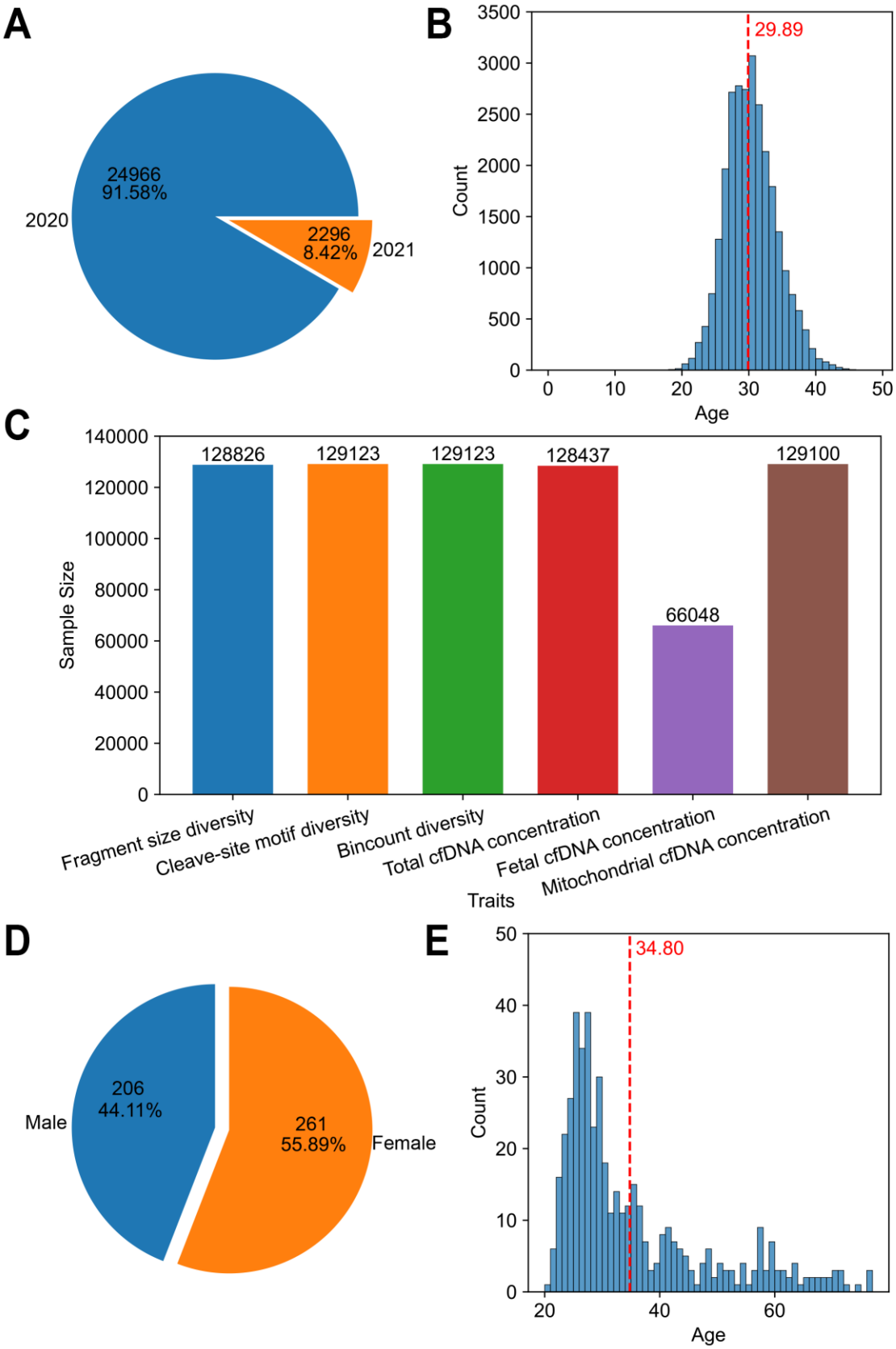
Supplementary Figure 3. The QQ-plots and genomic inflation factors of GWAS results for 256 cfDNA motif frequencies

A) the Quantile-Quantile plot of GWAS results for cfDNA motif frequencies grouped by motif start and **B**) the bar chart of the genomic inflation factor (λgc) from the GWAS on 256 motif frequencies, with the baseline (λgc = 1) indicated by the black dashed line.



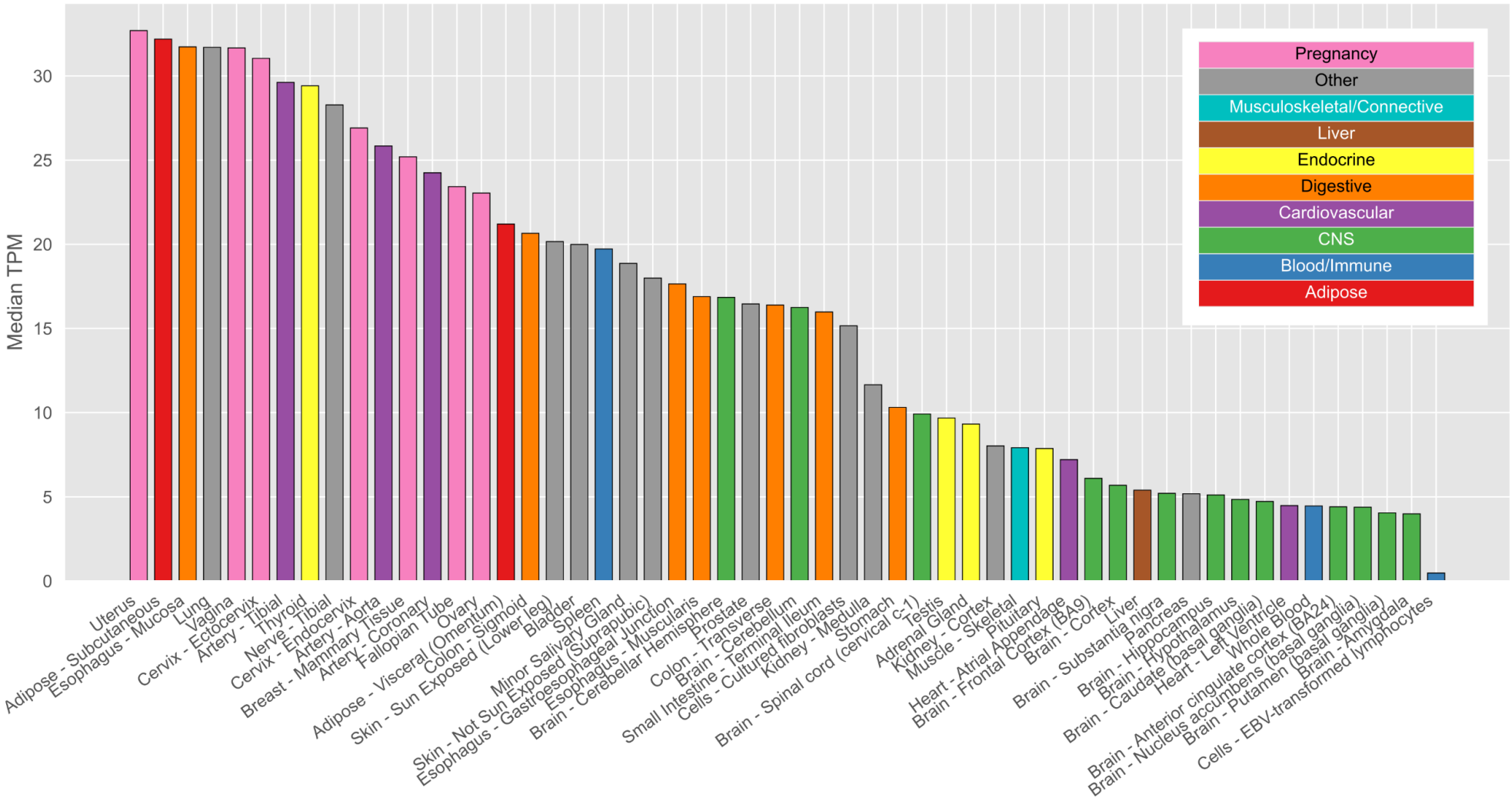
Supplementary Figure 4. The regional plots of the top 4 GWAS signals

Notes: the LocusZoom plot was created by locuszoom.org. In each locus, the lead SNP is marked by a diamond.



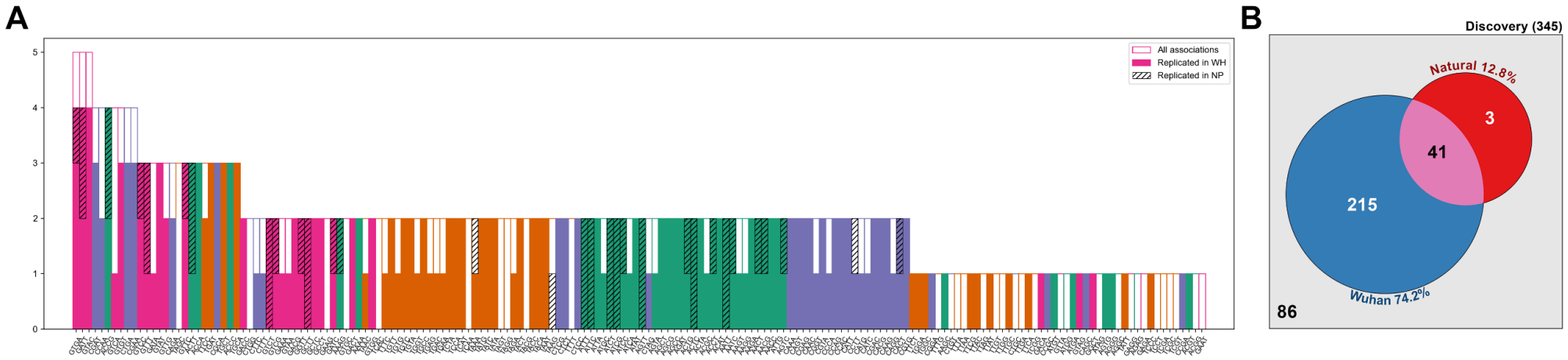
Supplementary Figure 5. The sample information in the three replication studies

A-B) number of samples collected in 2020 and 2021, and age distributions of participants in the Chinese NIPT replication study, **C**) sample sizes used for GWAS analysis of six cfDNA features in the Dutch replication study, **D-E**) sex and age distributions of participants in the natural replication study.



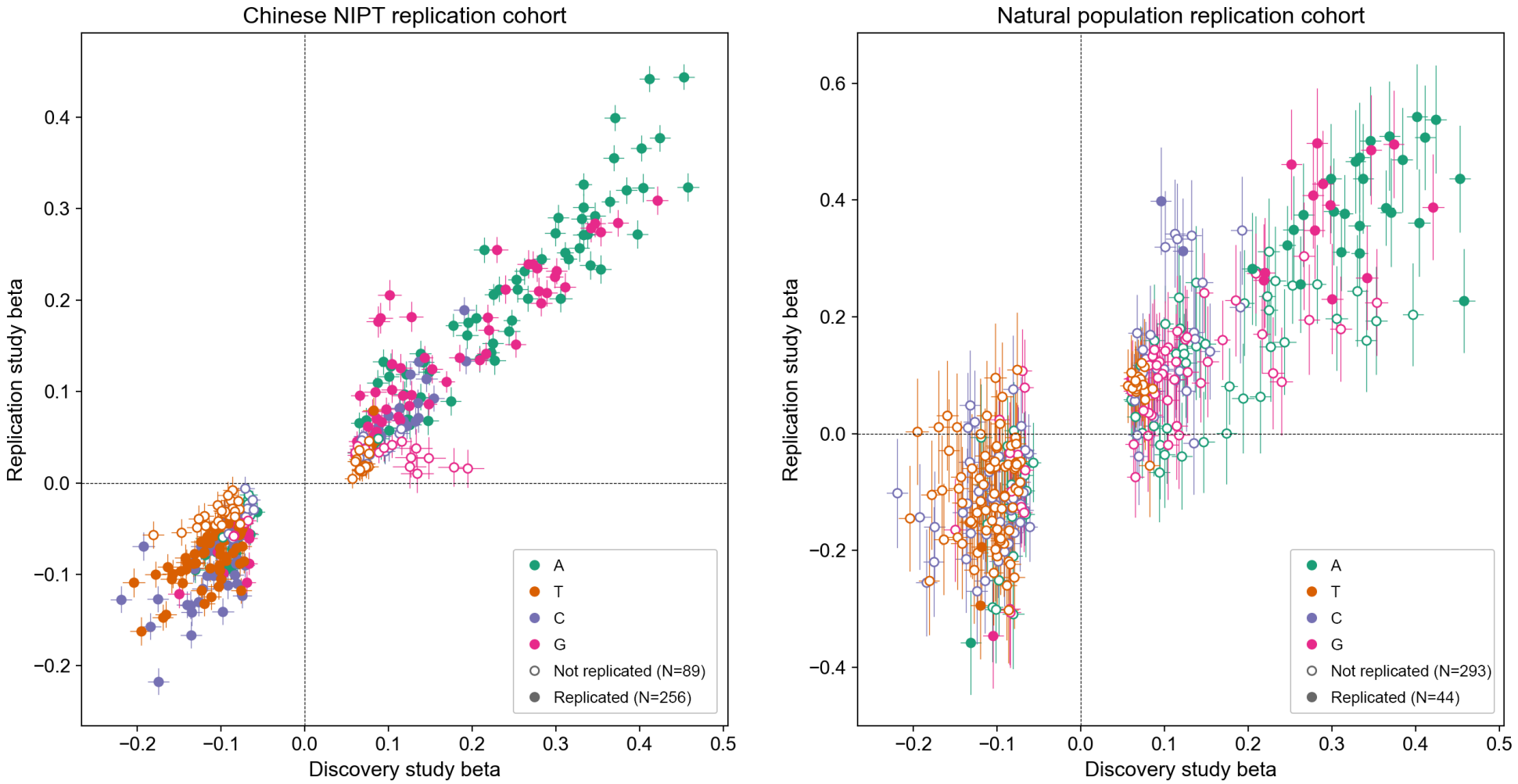
Supplementary Figure 6. The median value of transcripts per million (TPM) of gene *FCHO2* from the GTEx database

Notes: the bar chart displaying the median value of TPM of gene *FCHO2* from the GTEx database.



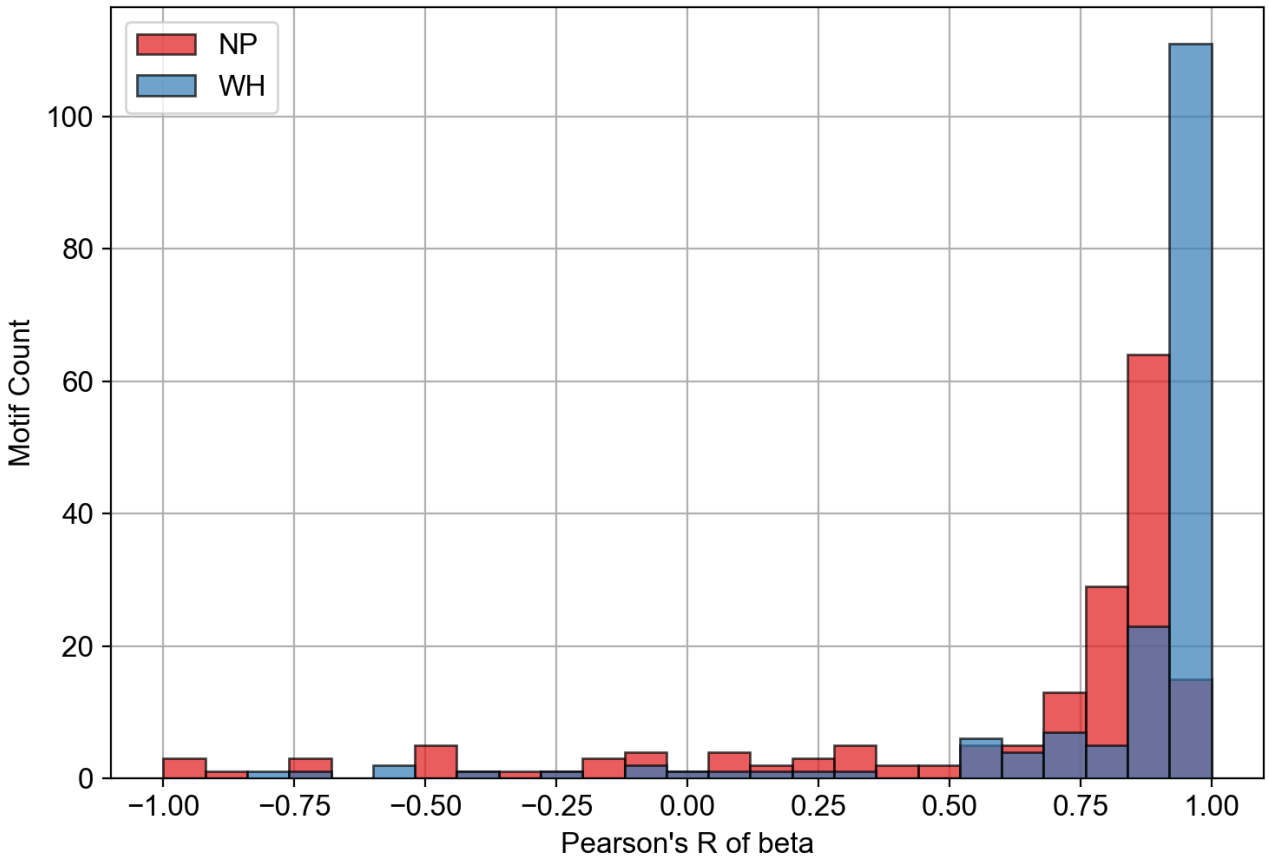
Supplementary Figure 7. Replication performance in Chinese NIPT and natural population cohorts illustrated by bar plots and Venn diagrams

A) Bar plots showing the replication performance for all 382 study-wide significant associations. Solid-colored bars represent associations replicated in the Chinese NIPT cohort, while hatched bars represent associations replicated in the natural population cohort. WH denotes the Chinese NIPT replication cohort, and NP denotes the natural population cohort and **B**) Venn diagrams illustrating the number of associations replicated by the Chinese NIPT cohort and the natural population cohort, respectively.



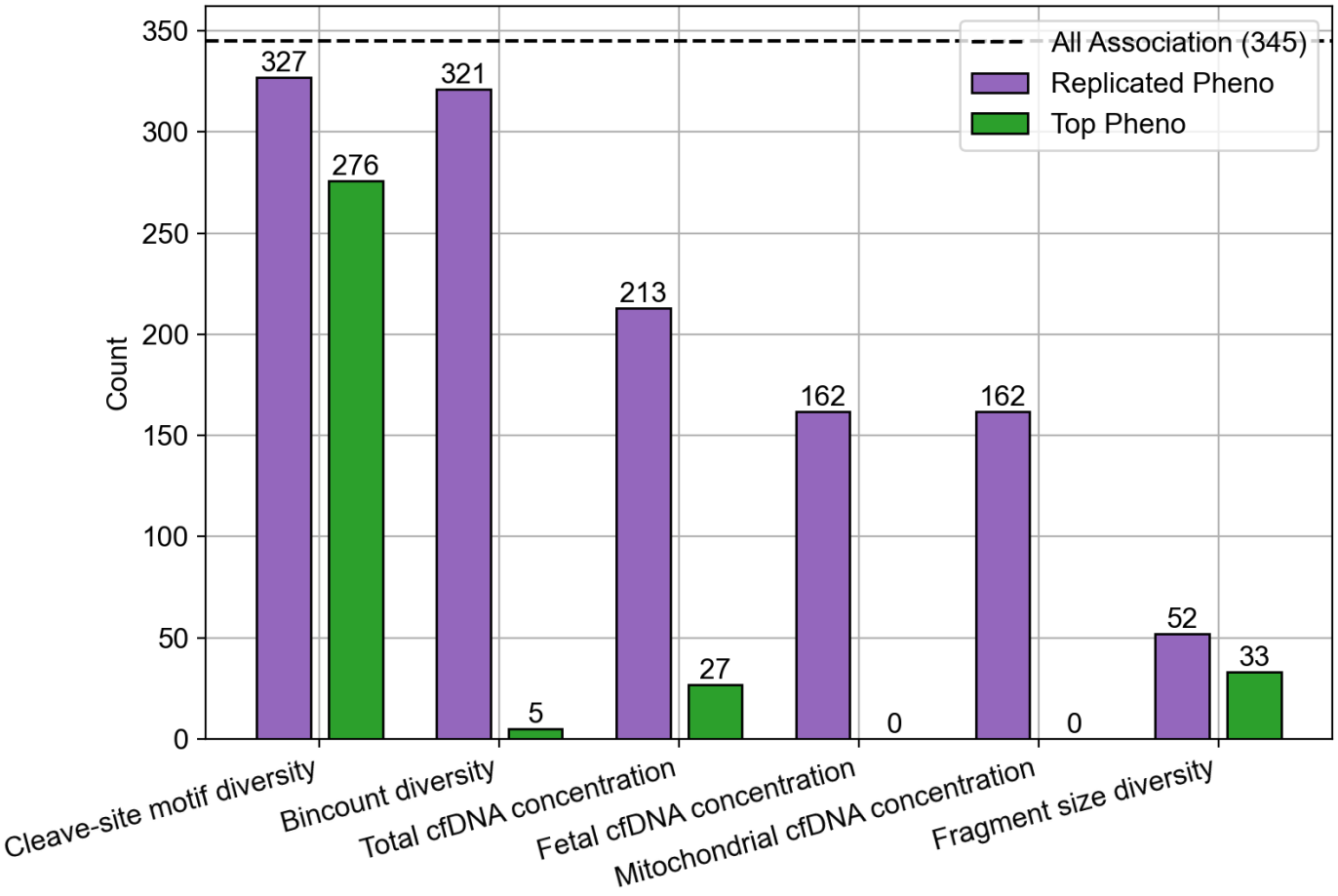
Supplementary Figure 8. Scatter plots of genetic effect sizes for the top SNPs in the discovery dataset and those in replication cohorts (NIPT and natural population)

Notes: The scatter plots comparing the effect sizes of top SNPs for a total of 345 motif-locus associations in the discovery study with those of the corresponding SNPs in the two replication studies, and the error bars on the points represent the standard errors (SE) of the effect sizes. Notably, since 8 top SNPs from the discovery study were not present in the natural population study, the number of points displayed is 337 rather than 345.

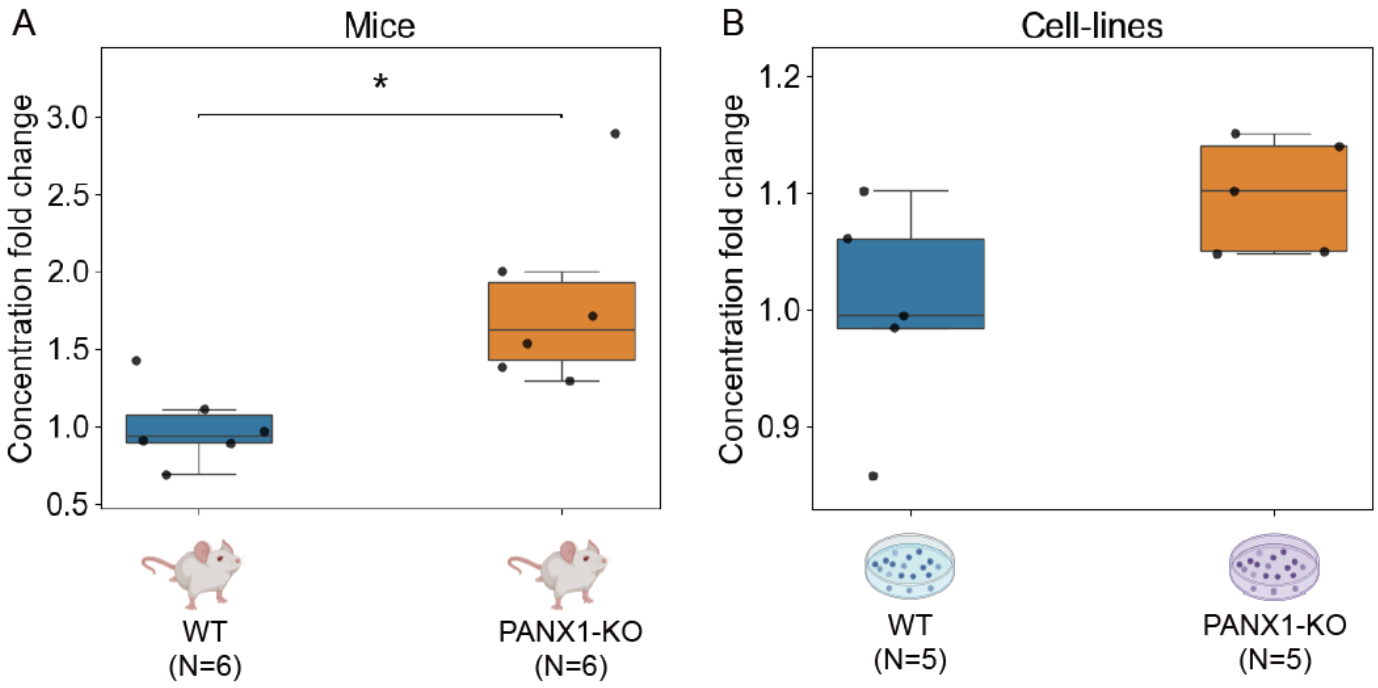


Supplementary Figure 9. Bar plots of Pearson correlation coefficients between genetic effect sizes in the discovery dataset and replication cohorts (NIPT and natural population)

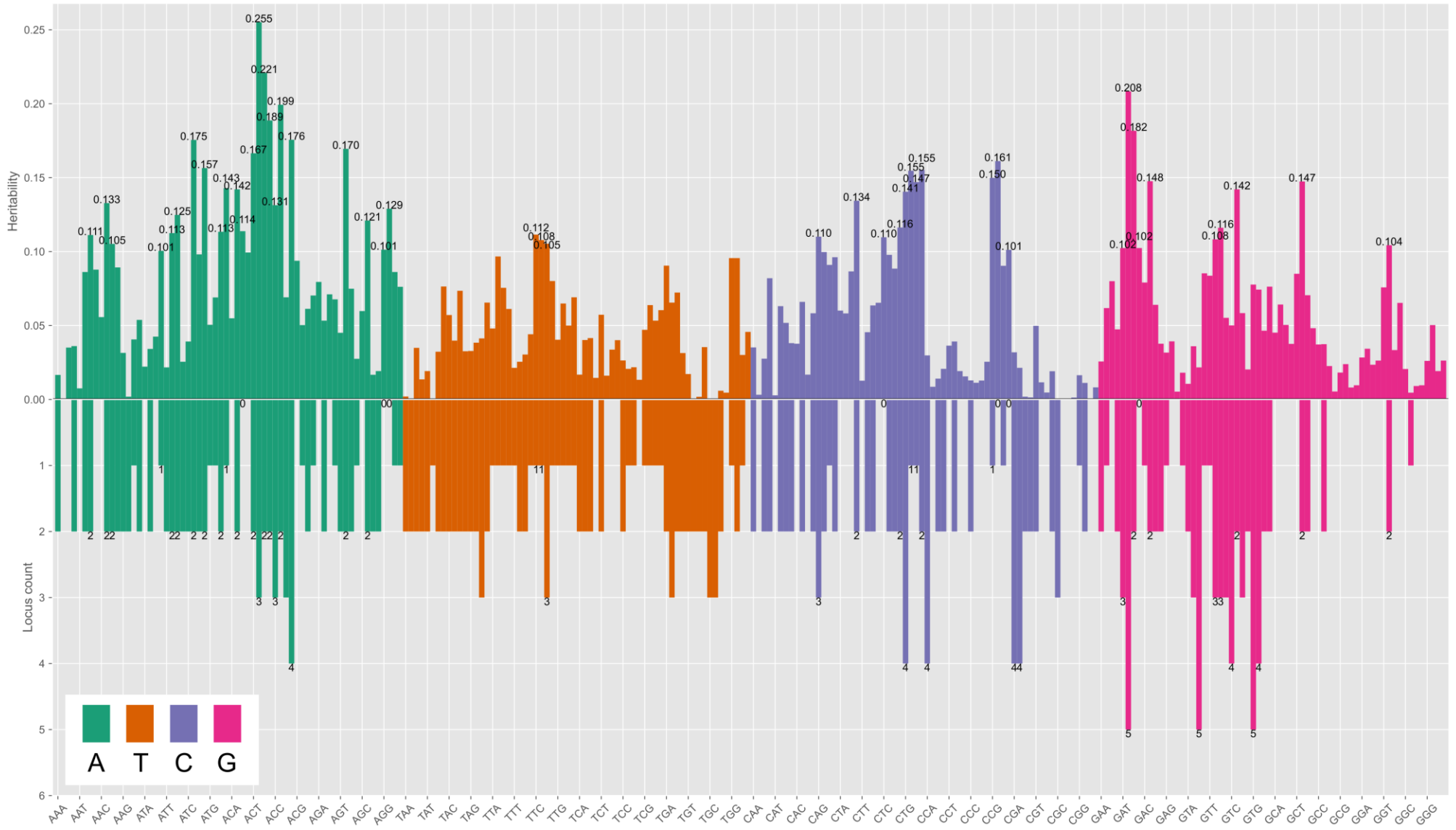
Notes: WH denotes the Chinese NIPT replication cohort, and NP denotes the natural population cohort.



Supplementary Figure 10. Bar plots of the number of motif-locus associations each Dutch phenotype was involved in and how often each phenotype served as the top phenotype

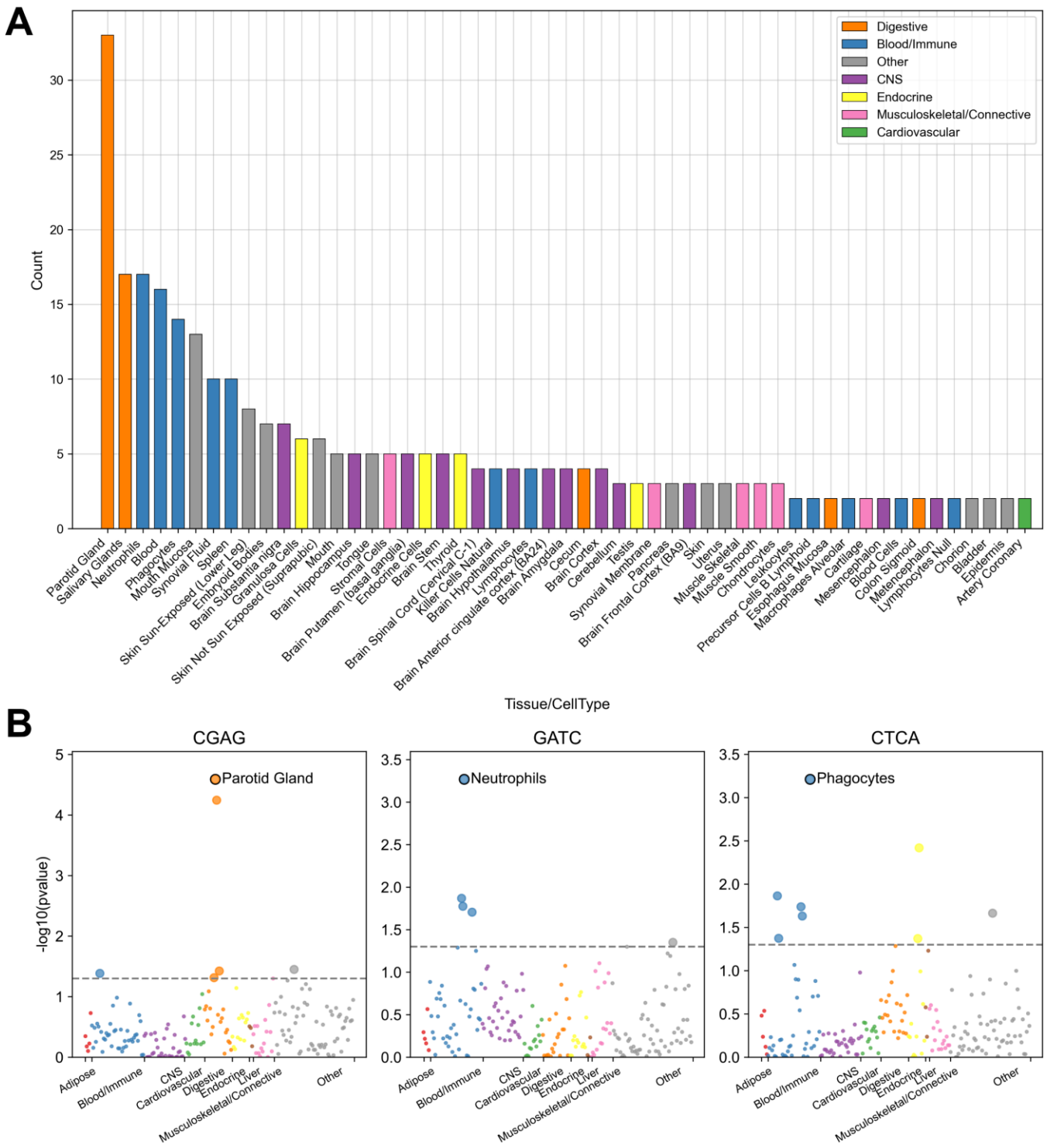


Supplementary Figure 11. The fold change of cfDNA concentration in mouse and cell line experiments **A**) box plot showing the concentration fold change of cfDNA for plasma samples from wild-type (WT) and *Panx1* KO mice and **B**) box plot showing the concentration fold change of cfDNA for cell culture supernatant from WT and *PANX1* KO cell lines. These box plots are defined as follows: the center line represents the median, the box limits indicate the 25th and 75th percentiles, and the whiskers extend to the minimum and maximum values. Statistical significance, determined by a two-sided t-test, is indicated by asterisks with the following conventions: *p < 0.05, **p < 0.01, and ***p < 0.001. Created in BioRender. Su, S. (2025) https://BioRender.com/m1tkhdf.



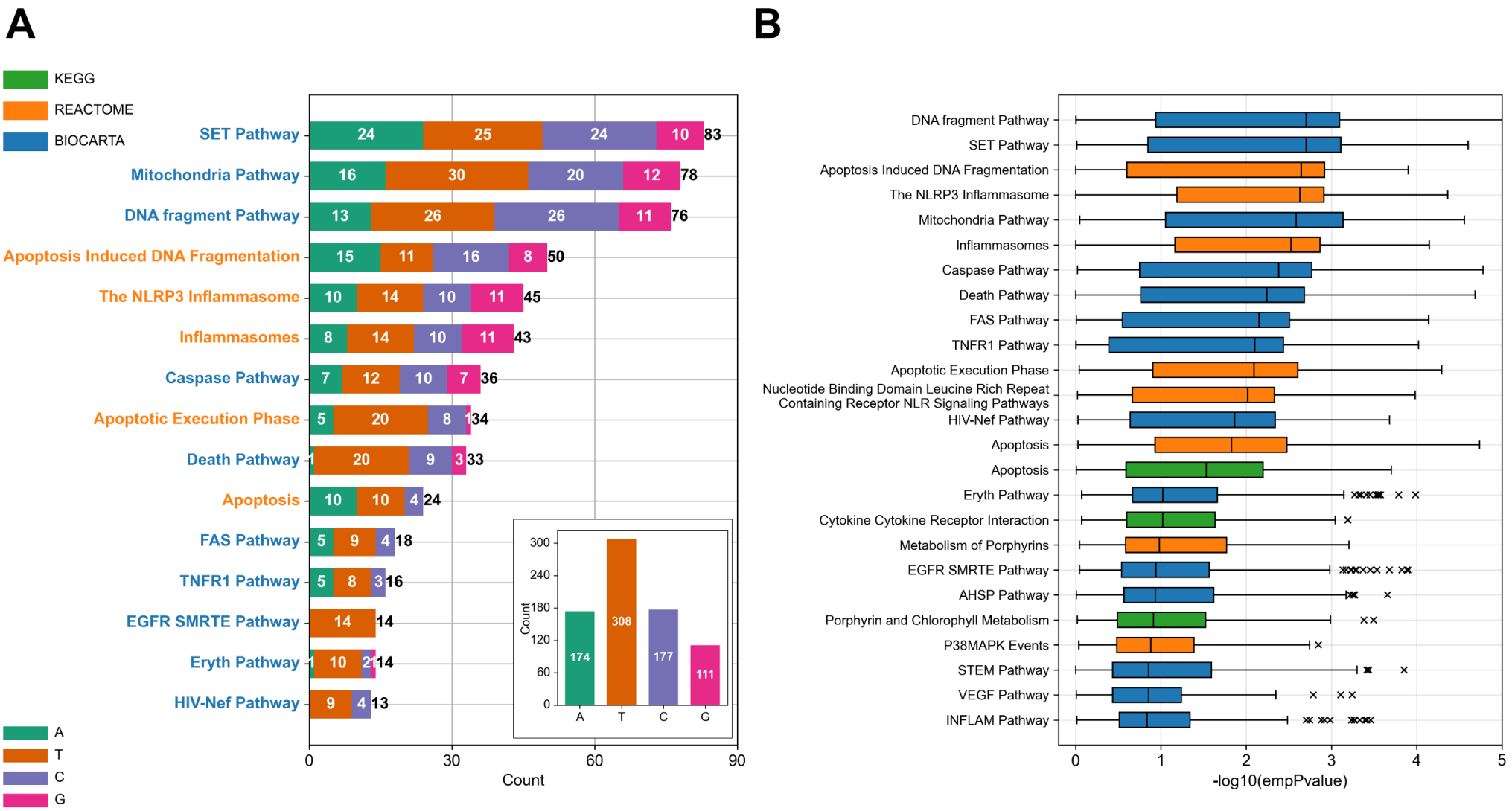
Supplementary Figure 12. The heritability and number of study-wide significant loci of 256 motifs

Notes: The top axis shows the heritability estimates for each motif, with values annotated for those exceeding 10%. The bottom axis shows the count of independent motif-locus associations from the GWAS, highlighting the number of loci linked to motifs with heritability >10%.



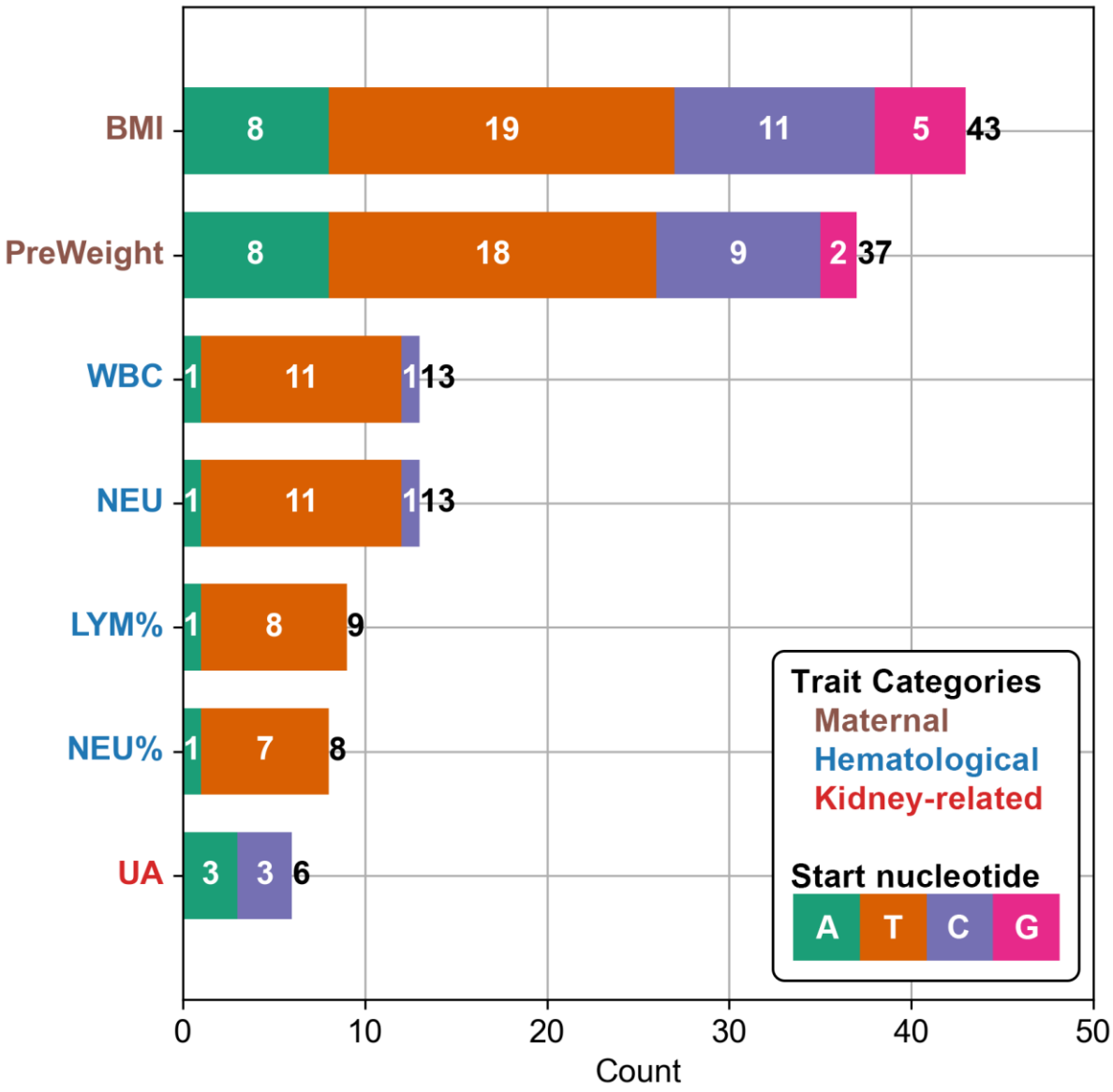
Supplementary Figure 13. Results of heritability partition analysis

A) number of associated cfDNA motifs for tissues/cell types with significant associations and **B**) heritability partition results for three cfDNA motifs, with the most significantly associated cell types labeled.



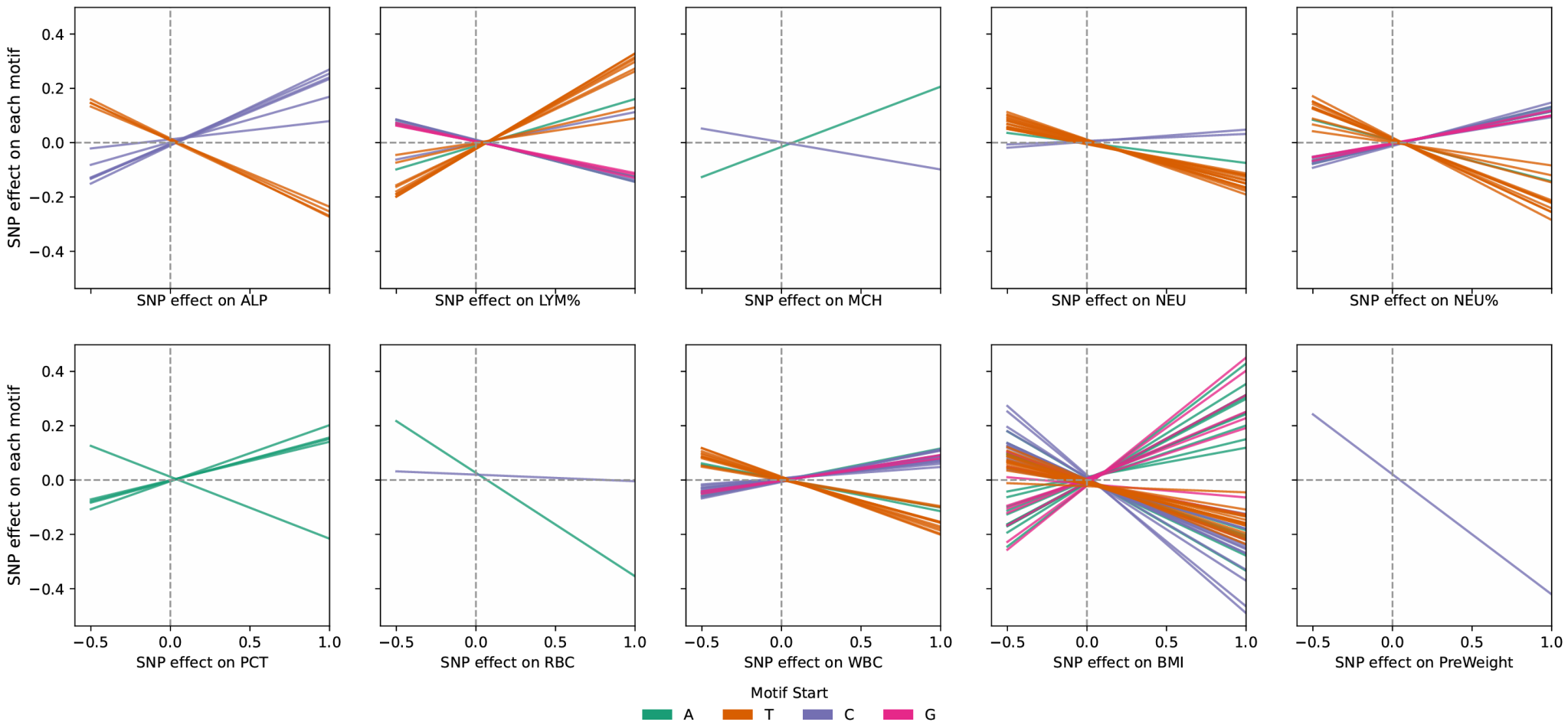
Supplementary Figure 14. Significant count and p-value distribution in pathway-based analysis

A) stacked bar chart of the top 15 enriched pathways with the highest number of associated motifs, and the numbers of pathways enriched by each motif start are presented in the bottom right corner and **B**) box plots showing the distribution (n=256) of -log10(empirical p-values) in the top 25 pathways (ranked by median empirical p-value). Box plot elements are defined as follows: center line, median; box limits, 25th/75th percentiles; whiskers, minimum/maximum values in 1.5×IQR range.



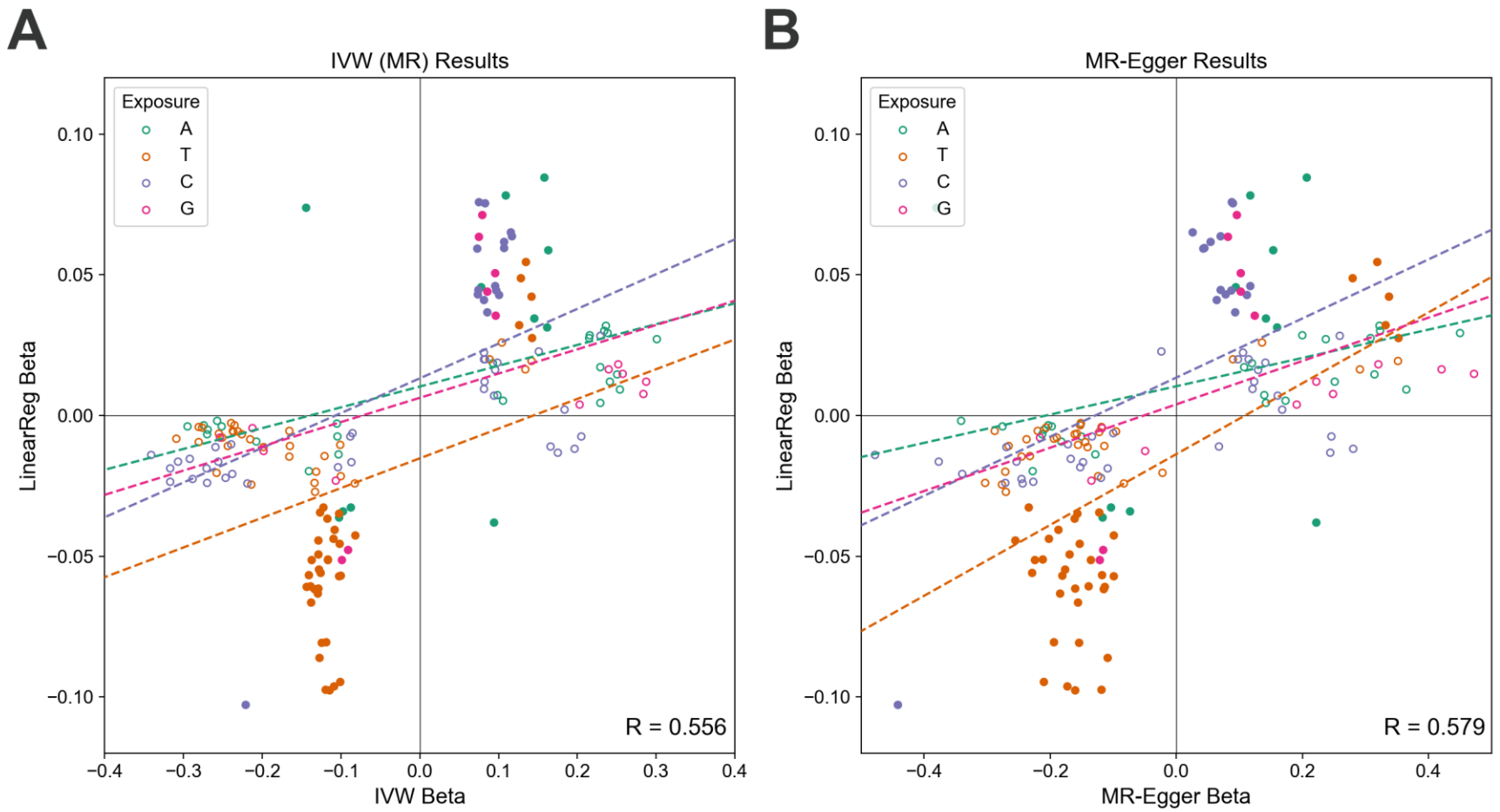
Supplementary Figure 15. Count of significantly genetically correlated motifs for each phenotype

Notes: BMI represents maternal body mass index, PreWeight represents maternal pre-pregnancy weight, WBC represents white blood cell counts, NEU represents neutrophil count, UA indicates uric acid, and NEU% represents the percentage of neutrophils.



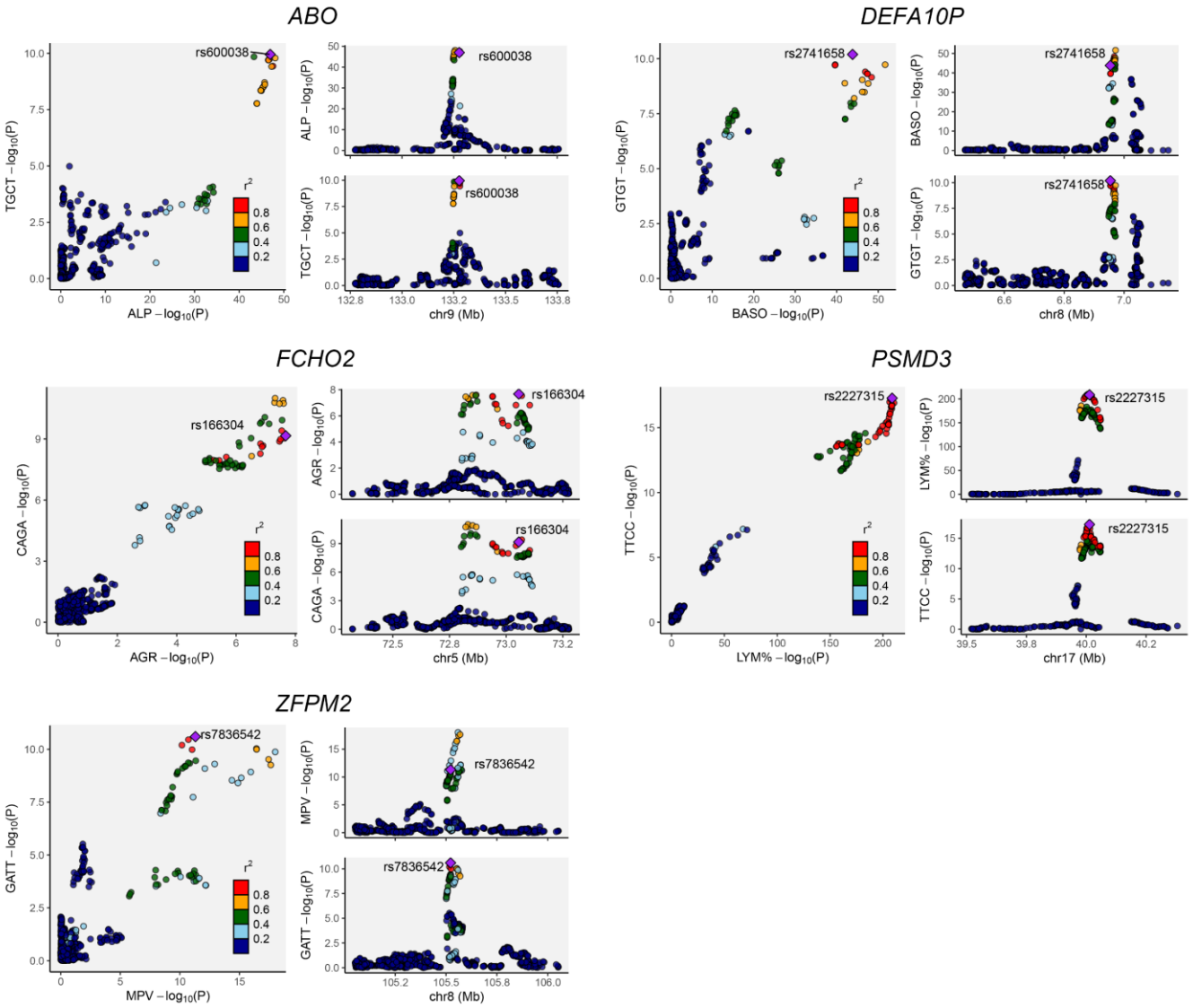
Supplementary Figure 16. The Mendelian randomization regression lines of the 10 causally related phenotypes

Notes: ALP represents alkaline phosphatase, LYM% represents percentage of lymphocytes, MCH represents mean corpuscular hemoglobin, NEU represents neutrophil counts, NEU% represents percentage of neutrophils, PCT represents plateletcrit, RBC represents red blood cell count, WBC represents white blood cell count, BMI represents body mass index, and PreWeight represents Pre-pregnancy weight.



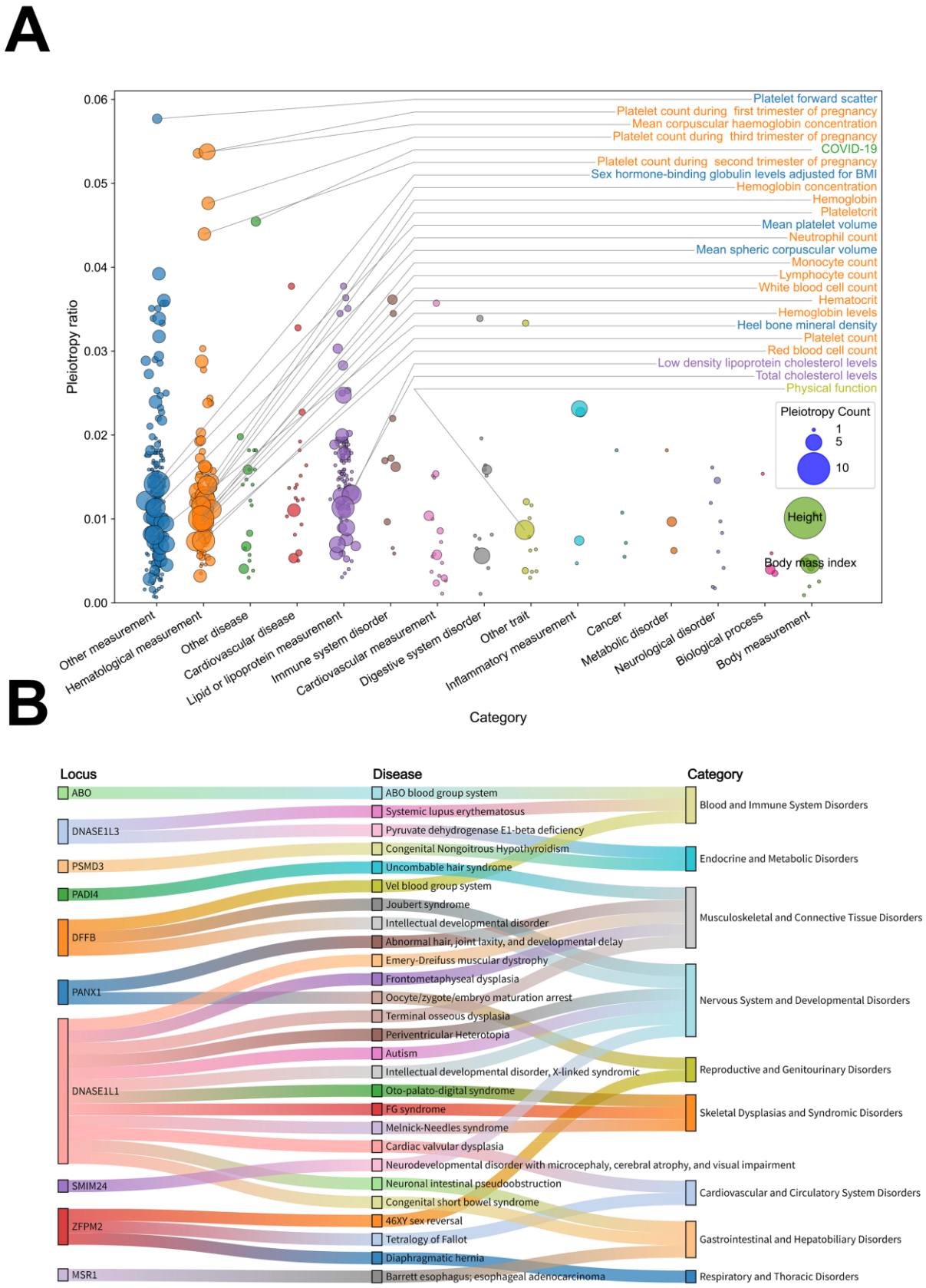
Supplementary Figure 17. The comparison of MR and linear regression effect estimates between motifs and phenotypes

A) the scatter plot comparing the beta values of the IVW (inverse variance weighted) method and linear regression displays 168 significant motif-phenotype pairs for IVW. Among these, pairs that also show a significant correlation in linear regression are marked with solid dots, while the others are indicated with hollow dots and **B**) the scatter plot comparing the beta values of the MR-Egger method and linear regression. The dots also represent the 168 significant motif-phenotype pairs in IVW, regardless of their significance in the MR-Egger method.

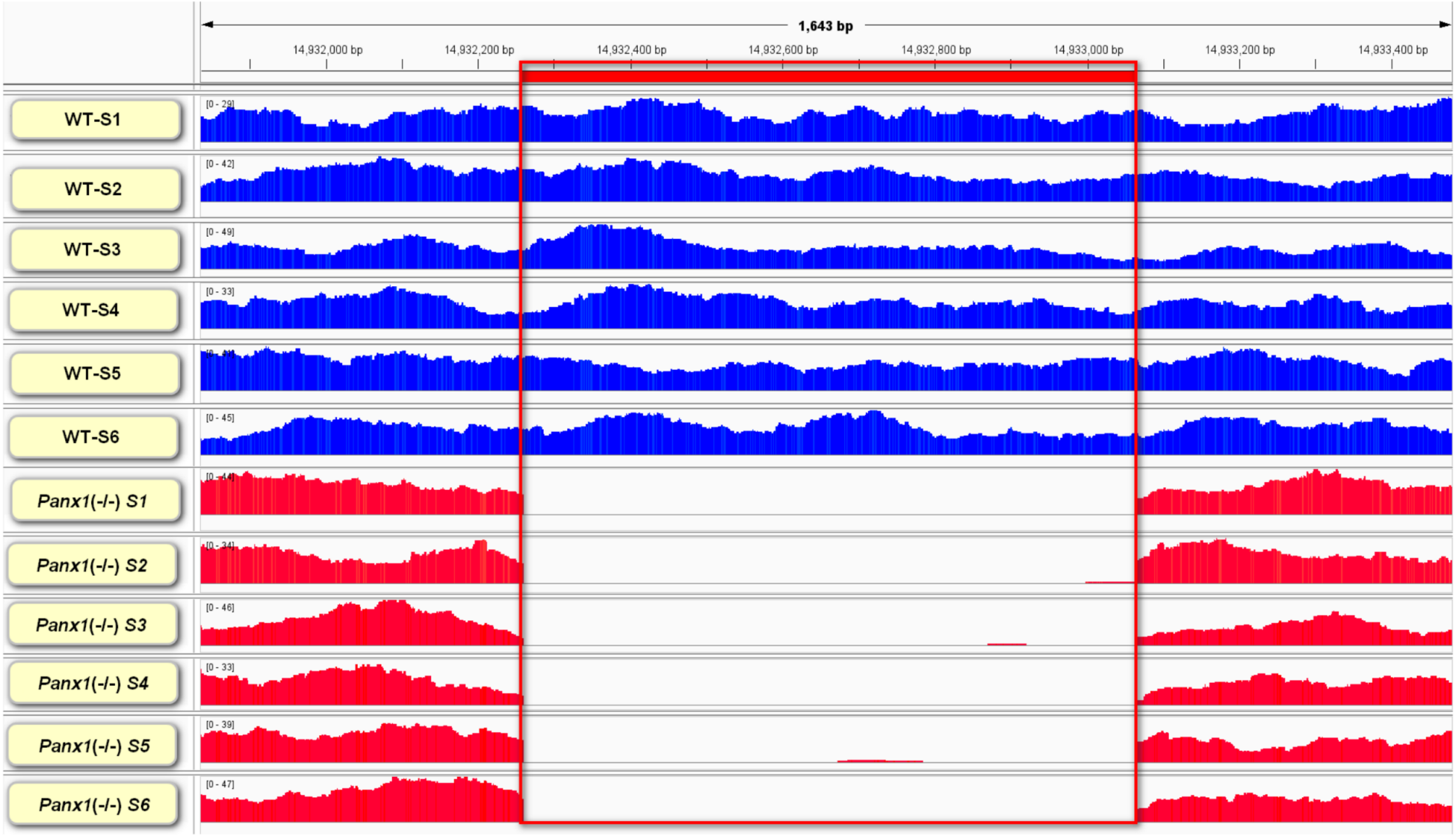


Supplementary Figure 18. Colocalization regional plot for five loci between cfDNA motifs and pregnancy phenotypes

Notes: ALP represents alkaline phosphatase, BASO represents basophil count, AGR represents albumin/globulin ratio, LYM% represents percentage of lymphocytes, and MPV represents mean platelet volume.



Supplementary Figure 19. The pleiotropy analysis results with GWAS catalog and OMIM database **A)** the bubble plot illustrating the summary of the motif associated loci that overlap with reported association in the GWAS catalog; all phenotypes presented have more than 50 variant records in the GWAS catalog and **B)** the Sankey diagram, from left to right, sequentially displays the study-wide significant loci associated with cfDNA motif frequencies, the associated disease in OMIM database, and the disease categories.



Supplementary Figure 20. The plasma cfDNA sequencing and alignment results verified the *Panx1*-deficient mice model

Supplementary Table 1. Number of associated motifs for tissues/cell types with significant associations in heritability partition analysis

Name	Category	Frequency	min_P	Top motif
A03.556.500.760.464.Parotid.Gland	Digestive	33	2.57E-05	CGAG
A15.382.490.315.583.Neutrophils	Blood/Immune	17	2.27E-04	GTTA
A03.556.500.760.Salivary.Glands	Digestive	17	5.68E-05	CGAG
A15.145.Blood	Blood/Immune	16	1.12E-04	CACC
A15.382.680.Phagocytes	Blood/Immune	14	6.14E-04	CTCA
A10.615.550.599.Mouth.Mucosa	Other	13	7.07E-04	CATA
A02.835.583.443.800.800.Synovial.Fluid	Blood/Immune	10	2.98E-04	GTAC
A15.382.520.604.700.Spleen	Blood/Immune	10	2.79E-03	CCTA
Skin_Sun_Exposed_(Lower_leg)	Other	8	1.57E-03	ACGC
A11.872.190.260.Embryoid.Bodies	Other	7	2.05E-04	GTCC
Brain_Substantia_nigra	CNS	7	6.34E-04	CATG
Skin_Not_Sun_Exposed_(Suprapubic)	Other	6	2.14E-03	ACGC
A11.436.329.Granulosa.Cells	Endocrine	6	6.16E-04	CTTG
Thyroid	Endocrine	5	2.36E-03	AGAG
Brain_Hippocampus	CNS	5	6.23E-04	CATG
A14.549.Mouth	Other	5	6.93E-04	CGGC
A14.549.885.Tongue	Other	5	1.96E-03	ACGT
A08.186.211.132.Brain.Stem	CNS	5	4.23E-04	AGCT
A11.329.830.Stromal.Cells	Musculoskeletal/Connective	5	3.92E-03	GGCC
Brain_Putamen_(basal_ganglia)	CNS	5	5.66E-03	TTGG
A11.382.Endocrine.Cells	Endocrine	5	1.93E-03	CTTC
Brain_Anterior_cingulate_cortex_(BA24)	CNS	4	3.59E-03	CATG
Brain_Cortex	CNS	4	3.44E-03	TCGG
A15.382.490.555.567.Lymphocytes	Blood/Immune	4	2.96E-03	CCTG
A03.556.249.249.209.Cecum	Digestive	4	4.84E-03	TTCT
Brain_Hypothalamus	CNS	4	2.78E-03	CATG
Brain_Spinal_cord_(cervical_c-1)	CNS	4	6.46E-03	CGCA
Brain_Amygdala	CNS	4	9.17E-04	CATG
A15.382.490.555.567.537.Killer.Cells..Natural	Blood/Immune	4	1.08E-03	GTCG
A17.815.Skin	Other	3	3.18E-03	CCCG
Muscle_Skeletal	Musculoskeletal/Connective	3	5.31E-03	GTTT
A10.690.467.Muscle..Smooth	Musculoskeletal/Connective	3	2.26E-04	GGCC
A02.835.583.443.800.Synovial.Membrane	Musculoskeletal/Connective	3	4.64E-03	GTCC
Testis	Endocrine	3	3.76E-03	ACAA
Brain_Frontal_Cortex_(BA9)	CNS	3	1.16E-03	CATG
Pancreas	Other	3	3.26E-03	CAGT
A11.329.171.Chondrocytes	Musculoskeletal/Connective	3	3.28E-04	GTCG
A08.186.211.132.810.428.200.Cerebellum	CNS	3	8.57E-03	TGTA
Uterus	Other	3	2.82E-03	TCCT
A10.272.497.Epidermis	Other	2	1.50E-04	ACGC
A15.145.229.Blood.Cells	Blood/Immune	2	5.17E-03	CGTT
A03.556.249.249.356.668.Colon..Sigmoid	Digestive	2	5.16E-03	GAAA
Bladder	Other	2	3.22E-03	CGCC
A02.165.Cartilage	Musculoskeletal/Connective	2	5.54E-03	GAAG
A11.118.637.555.567.562.440.Precursor.Cells..B.Lymphoid	Blood/Immune	2	5.56E-03	TGTT
A11.329.372.600.Macrophages..Alveolar	Blood/Immune	2	7.09E-03	CACA
A10.615.284.473.Chorion	Other	2	5.71E-03	CCGA
A15.382.490.555.567.622.Lymphocytes..Null	Blood/Immune	2	2.15E-03	CATA
A11.118.637.Leukocytes	Blood/Immune	2	1.89E-03	GAAC
Esophagus_Mucosa	Digestive	2	8.19E-03	TGCT
A08.186.211.865.428.Metencephalon	CNS	2	5.87E-03	AGCT
Artery_Coronary	Cardiovascular	2	1.18E-03	TGCC
A08.186.211.653.Mesencephalon	CNS	2	5.29E-03	TTAC
A11.627.624.249.Monocyte.Macrophage.Precursor.Cells	Blood/Immune	1	7.00E-03	CATG
Liver	Liver	1	6.76E-03	CTCT
A03.556.875.875.Stomach	Digestive	1	8.16E-03	CGAA
A11.872.653.Neural.Stem.Cells	CNS	1	7.29E-03	TACC
Cells_EBV-transformed_lymphocytes	Blood/Immune	1	9.30E-03	ATCG
A15.145.229.637.555.567.562.725.Plasma.Cells	Blood/Immune	1	7.33E-03	CGTT
Pituitary	Endocrine	1	7.34E-03	ATTG
A08.186.211.Brain	CNS	1	7.87E-03	TGTA
A11.436.294.064.Glucagon.Secreting.Cells	Other	1	7.94E-03	CGCG
A03.556.875.500.Esophagus	Digestive	1	6.76E-03	CGCC
Artery_Tibial	Cardiovascular	1	9.66E-03	TGCG
A03.556.Gastrointestinal.Tract	Digestive	1	8.92E-03	CCAA
A10.165.450.300.Cicatrix	Musculoskeletal/Connective	1	3.44E-03	TGAA
A15.378.316.Bone.Marrow.Cells	Blood/Immune	1	6.63E-03	GACT
A14.724.Pharynx	Other	1	6.37E-03	GTAC
A10.615.789.Serous.Membrane	Other	1	6.25E-03	AACG
A11.872.190.Embryonic.Stem.Cells	Other	1	5.59E-03	AAGC
A11.627.340.360.Granulocyte.Precursor.Cells	Blood/Immune	1	4.64E-03	AACG
A05.810.453.324.Kidney.Cortex	Other	1	4.43E-03	GGAG
Brain_Cerebellar_Hemisphere	CNS	1	4.05E-03	ATCG
Kidney_Cortex	Other	1	3.93E-03	CGCA
A11.497.497.600.Oocytes	Other	1	3.62E-03	GCAC
A11.329.114.Adipocytes	Adipose	1	3.20E-03	TGTA
A07.541.510.110.Aortic.Valve	Cardiovascular	1	2.62E-03	TCTT
A15.382.Immune.System	Blood/Immune	1	2.28E-03	CGTT
A11.118.637.555.567.569.T.Lymphocytes	Blood/Immune	1	2.08E-03	GAGG
Brain_Nucleus_accumbens_(basal_ganglia)	CNS	1	1.83E-03	CATG
A03.556.875.Upper.Gastrointestinal.Tract	Digestive	1	1.60E-03	AGCA
A11.329.Connective.Tissue.Cells	Musculoskeletal/Connective	1	9.80E-03	ACTC

Notes: The categorization of tissue and cell types followed Finucane et al [1]. For each cell/tissue type, the minimum observed p-value and its corresponding motif is reported. These are unadjusted p-values from the heritability partition LD score regression, based on two-sided Wald tests.

Supplementary Table 2. Number of associated cfDNA motifs for each pathway in the pathway-based analysis

Pathway	Frequency
BIOCARTA_SET_PATHWAY	83
BIOCARTA_MITOCHONDRIA_PATHWAY	78
BIOCARTA_DNAFRAGMENT_PATHWAY	76
REACTOME_APOPTOSIS_INDUCED_DNA_FRAGMENTATION	50
REACTOME_THE_NLRP3_INFLAMMASOME	45
REACTOME_INFLAMMASOMES	43
BIOCARTA_CASPASE_PATHWAY	36
REACTOME_APOPTOTIC_EXECUTION_PHASE	34
BIOCARTA_DEATH_PATHWAY	33
REACTOME_APOPTOSIS	24
BIOCARTA_FAS_PATHWAY	18
BIOCARTA_TNFR1_PATHWAY	16
BIOCARTA_EGFR_SMRTE_PATHWAY	14
BIOCARTA_ERYTH_PATHWAY	14
BIOCARTA_HIVNEF_PATHWAY	13
REACTOME_NUCLEOTIDE_BINDING_DOMAIN_LEUCINE_RICH_REPEAT_CONTAINING_RECEPTOR_NLR_SIGNALING_PATHWAYS	12
REACTOME_NUCLEAR_RECEPTOR_TRANSCRIPTION_PATHWAY	11
KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_BIOSYNTHESIS	11
KEGG_APOPTOSIS	11
BIOCARTA_IL17_PATHWAY	10
BIOCARTA_GRANULOCYTES_PATHWAY	10
BIOCARTA_AHSP_PATHWAY	8
BIOCARTA_STEM_PATHWAY	8
BIOCARTA_INFLAM_PATHWAY	7
KEGG_JAK_STAT_SIGNALING_PATHWAY	7
REACTOME_REGULATION_OF_PYRUVATE_DEHYDROGENASE_PDH_COMPLEX	5
KEGG_CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	5
REACTOME_VEGF_LIGAND_RECEPTOR_INTERACTIONS	5
REACTOME_CDT1_ASSOCIATION_WITH_THE_CDC6_ORC_ORIGIN_COMPLEX	4
REACTOME_METABOLISM_OF_PORPHYRINS	4
REACTOME_ACTIVATION_OF_NF_KAPPAB_IN_B_CELLS	3
REACTOME_REGULATION_OF_INSULIN_SECRETION	3
REACTOME_INCRETIN_SYNTHESIS_SECRETION_AND_INACTIVATION	3
REACTOME_SYNTHESIS_SECRETION_AND_INACTIVATION_OF_GLP1	3
REACTOME_MICRORNA_MIRNA_BIOGENESIS	3
REACTOME_BIOLOGICAL_OXIDATIONS	2
REACTOME_PTM_GAMMA_CARBOXYLATION_HYPUSINE_FORMATION_AND_ARYLSULFATASE_ACTIVATION	2
REACTOME_PYRUVATE_METABOLISM	2
REACTOME_PYRUVATE_METABOLISM_AND_CITRIC_ACID_TCA_CYCLE	2
REACTOME_DOWNSTREAM_SIGNALING_EVENTS_OF_B_CELL_RECEPTOR_BCR	2
REACTOME_CROSS_PRESENTATION_OF_SOLUBLE_EXOGENOUS_ANTIGENS_ENDOSOMES	2
REACTOME_IMMUNE_SYSTEM	2
BIOCARTA_VEGF_PATHWAY	2
BIOCARTA_INSULIN_PATHWAY	2
REACTOME_REGULATORY_RNA_PATHWAYS	2
REACTOME_SIGNALING_BY_WNT	2
REACTOME_ANTIGEN_PROCESSING_CROSS_PRESENTATION	2
REACTOME_SYNTHESIS_SECRETION_AND_DEACYLATION_OF_GHRELIN	2
KEGG_STEROID_HORMONE_BIOSYNTHESIS	2
KEGG_PORPHYRIN_AND_CHLOROPHYLL_METABOLISM	2
KEGG_GAP_JUNCTION	2
REACTOME_REGULATION_OF_ORNITHINE_DECARBOXYLASE_ODC	1
BIOCARTA_IGF1MTOR_PATHWAY	1
REACTOME_REGULATION_OF_APOPTOSIS	1
REACTOME_SCF_BETA_TRCP_MEDIATED_DEGRADATION_OF_EMI1	1
BIOCARTA_HSP27_PATHWAY	1
REACTOME_P53_INDEPENDENT_G1_S_DNA_DAMAGE_CHECKPOINT	1
REACTOME_VIF_MEDIATED_DEGRADATION_OF_APOBEC3G	1
REACTOME_P53_DEPENDENT_G1_DNA_DAMAGE_RESPONSE	1
REACTOME_FATTY_ACYL_COA_BIOSYNTHESIS	1
BIOCARTA_IL4_PATHWAY	1
BIOCARTA_MTOR_PATHWAY	1
KEGG_BASAL_CELL_CARCINOMA	1
KEGG_DRUG_METABOLISM_OTHER_ENZYMES	1
KEGG_GLYCOLYSIS_GLUCONEOGENESIS	1
KEGG_GLYCOSPHINGOLIPID_BIOSYNTHESIS_LACTO_AND_NEOLACTO_SERIES	1
BIOCARTA_RAC1_PATHWAY	1
REACTOME_APC_C_CDC20_MEDIATED_DEGRADATION_OF_MITOTIC_PROTEINS	1
BIOCARTA_PTDINS_PATHWAY	1
REACTOME_ASSEMBLY_OF_THE_PRE_REPLICATIVE_COMPLEX	1
REACTOME_MTORC1_MEDIATED_SIGNALLING	1
REACTOME_AUTODEGRADATION_OF_THE_E3_UBIQUITIN_LIGASE_COP1	1
REACTOME_CDK_MEDIATED_PHOSPHORYLATION_AND_REMOVAL_OF_CDC6	1
REACTOME_DESTABILIZATION_OF_MRNA_BY_AUF1_HNRNP_D0	1
REACTOME_E2F_ENABLED_INHIBITION_OF_PRE_REPLICATION_COMPLEX_FORMATION	1
REACTOME_ER_PHAGOSOME_PATHWAY	1
REACTOME_GASTRIN_CREB_SIGNALLING_PATHWAY_VIA_PKC_AND_MAPK	1
REACTOME_INSULIN_RECEPTOR_SIGNALLING_CASCADE	1
REACTOME_XENOBIOTICS	1

Notes: The database of pathways include REACTOME [2], KEGG [3], and BIOCARTA [4].

Supplementary Table 3. Number of associated motifs for seven pregnancy phenotypes in genetic correlation analysis

Phenotype	Freq
BMI	43
PreWeight	37
WBC	13
NEU	13
LYM%	9
NEU%	8
UA	6

Notes: BMI represents maternal body mass index, PreWeight represents maternal pre-pregnancy weight, WBC represents white blood cell counts, NEU represents neutrophil count, UA indicates uric acid, and NEU% represents the percentage of neutrophils.

Supplementary Table 4. Genes associated with both cfDNA motif frequencies and diseases from OMIM database

Defined_locus_name	Gene	Diseases	System Affected
ABO	*ABO*	ABO blood group system	Blood and Immune System Disorders
DFFB	*CEP104*	Joubert syndrome	Nervous System and Developmental Disorders
		Intellectual developmental disorder	Nervous System and Developmental Disorders
	SMIM1	Vel blood group system	Blood and Immune System Disorders
DNASE1L1	*EMD*	Emery-Dreifuss muscular dystrophy	Musculoskeletal and Connective Tissue Disorders
	FLNA	Oto-palato-digital syndrome	Skeletal Dysplasias and Syndromic Disorders
		Neuronal intestinal pseudoobstruction	Gastrointestinal and Hepatobiliary Disorders
		Congenital short bowel syndrome	Gastrointestinal and Hepatobiliary Disorders
		FG syndrome	Skeletal Dysplasias and Syndromic Disorders
		Cardiac valvular dysplasia	Cardiovascular and Circulatory System Disorders
		Frontometaphyseal dysplasia	Musculoskeletal and Connective Tissue Disorders
		Melnick-Needles syndrome	Skeletal Dysplasias and Syndromic Disorders
		Periventricular Heterotopia	Nervous System and Developmental Disorders
		Terminal osseous dysplasia	Musculoskeletal and Connective Tissue Disorders
	RPL10	Autism	Nervous System and Developmental Disorders
		Intellectual developmental disorder, X-linked syndromic	Nervous System and Developmental Disorders
DNASE1L3	*DNASE1L3*	Systemic lupus erythematosus	Blood and Immune System Disorders
	PDHB	Pyruvate dehydrogenase E1-beta deficiency	Endocrine and Metabolic Disorders
MSR1	*MSR1*	Barrett esophagus; esophageal adenocarcinoma	Gastrointestinal and Hepatobiliary Disorders
PADI4	*PADI3*	Uncombable hair syndrome	Musculoskeletal and Connective Tissue Disorders
PANX1	*HEPHL1*	Abnormal hair, joint laxity, and developmental delay	Musculoskeletal and Connective Tissue Disorders
	PANX1	Oocyte/zygote/embryo maturation arrest	Reproductive and Genitourinary Disorders
PSMD3	*THRA*	Congenital Nongoitrous Hypothyroidism	Endocrine and Metabolic Disorders
SMIM24	*DOHH*	Neurodevelopmental disorder with microcephaly, cerebral atrophy, and visual impairment	Nervous System and Developmental Disorders
ZFPM2	*ZFPM2*	Diaphragmatic hernia	Respiratory and Thoracic Disorders
		Tetralogy of Fallot	Cardiovascular and Circulatory System Disorders
		46XY sex reversal	Reproductive and Genitourinary Disorders

Supplementary References

1 Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nature genetics* **50**, 621-629, doi:10.1038/s41588-018-0081-4 (2018).

2 Milacic, M. *et al.* The Reactome Pathway Knowledgebase 2024. *Nucleic acids research* **52**, D672-d678, doi:10.1093/nar/gkad1025 (2024).

3 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**, 27-30, doi:10.1093/nar/28.1.27 (2000).

4 Nishimura, D. BioCarta. *Biotech Software & Internet Report* **2**, 117-120, doi:10.1089/152791601750294344 (2001).