Supplementary information for

IFAM: Improving Genomic Prediction Accuracy of Complex Traits by Integrating Massive Types of Functional Annotation Information

Zhenshuang Tang[1,2], Xiong Xiong[2,3], Haohao Zhang[4], Dong Yin[2], Yuhua Fu[2,3], Yunxia Zhao[1,2], Jingjin Li[2], Yuan Quan[5], Xiang Zhou[6,7], Xinyun Li[2,3], Lilin Yin[2,3*], Shuhong Zhao[1,2,3*], and Xiaolei Liu[2,3*]

[1] Yazhouwan National Laboratory, Sanya 572024, China;
[2] Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction, Ministry of Education, Frontiers Science Center for Animal Breeding and Sustainable Production, Key Laboratory of Swine Genetics and Breeding, Ministry of Agriculture, College of Animal Science and Technology, Huazhong Agricultural University, Wuhan 430070, PR China;
[3] Hubei Hongshan Laboratory, Wuhan 430070, PR China;
[4] School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, PR China;
[5] Hubei Key Laboratory of Agricultural Bioinformatics, College of Informatics, Huazhong Agricultural University, Wuhan 430070, PR China;
[6] Department of Biostatistics, University of Michigan, Ann Arbor, MI, USA;
[7] Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA;
**Correspondence: Lilin Yin (ylilin@mail.hzau.edu.cn), Shuhong Zhao (zhaoshuhong@yzwlab.cn), Xiaolei Liu (xiaoleiliu@mail.hzau.edu.cn)*

Supplementary Note

Genomic best linear unbiased prediction (GBLUP)

The GBLUP model is a statistical approach utilized for estimating the genetic merit of individuals in livestock and plant breeding and for predicating polygenic risk scores (PRS) for diseases in humans[1, 2]. GBLUP assumes that all genetic markers contribute equally to genetic variation, which implies the absence of major genes across the genome. There are two equivalent models for genomic BLUP prediction. The first model estimates individual marker effects, summing them to obtain the genomic estimated breeding values (GEBV), which is commonly known as ridge regression BLUP[3]. The second model, which we employed in our study, replaces the traditional

additive genetic relationship matrix that is derived from pedigree data with the genomic relationship matrix[4]. The latter, which is referred to as GBLUP[5], is represented by the equation:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Z\mu} + \boldsymbol{e}$$

where $\boldsymbol{y}$ is a vector of phenotypic values, which are binary values (case 1, control 0) for disease traits or continuous values for quantitative traits; $\boldsymbol{b}$ is a vector of the estimated effects of fixed covariates, $\boldsymbol{X}$ is the corresponding incidence matrix; there is only one random genetic effect ($\boldsymbol{\mu}$) in this model, which is called the GEBV, $\boldsymbol{\mu}$ follows a normal distribution $\boldsymbol{\mu} \sim \boldsymbol{N}\left(\boldsymbol{0}, \boldsymbol{G\sigma_g^2}\right)$, where $\boldsymbol{G}$ is the additive genomic relationship matrix constructed using all SNPs, and $\boldsymbol{\sigma_g^2}$ is the estimated genetic variance, $\boldsymbol{Z}$ is the incidence matrix for random genetic effect; $\boldsymbol{e}$ is a vector of residual effects, following a normal distribution $\boldsymbol{e} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{I\sigma_e^2})$ with $\boldsymbol{I}$ being an identity matrix, and $\boldsymbol{\sigma_e^2}$ is the residual variance.

BayesR and BayesRC

In addition to comparisons with the GBLUP model, we also assessed the performance of the IFAM model against the BayesR (Bayesian multiple regression) model and the BayesRC model using the WTCCC1 dataset[6]. BayesR belongs to a Bayesian mixture model and is typically employed to calculate risk predictors for diseases in humans or GEBV for economic traits in livestock. Compared with models within the BLUP framework, Bayesian models enhance predictive ability by setting different statistical

priors, albeit with relatively lower computational efficiency. The standard linear regression model that relates phenotype to markers is expressed as follows:

$$\boldsymbol{y} = \mathbf{1}_n\boldsymbol{\mu} + \boldsymbol{X\beta} + \boldsymbol{e}$$

where $\boldsymbol{y}$ is a vector of phenotypic values as mentioned previously; $\mathbf{1}_n$ is an n-dimensional vector of ones, $\boldsymbol{\mu}$ is the general mean; $\boldsymbol{X}$ is a genotypic matrix encoded as 0, 1, and 2 for the homozygote, heterozygote, and another homozygote, respectively, $\boldsymbol{\beta}$ is a vector of additive marker effects; and $\boldsymbol{e}$ is a vector of residuals, where $\boldsymbol{e}\sim\boldsymbol{N}(\mathbf{0}, \boldsymbol{I}\sigma_e^2)$ with $\boldsymbol{I}$ being an identity matrix. The prior assumption of the BayesR model is that SNP effects ($\boldsymbol{\beta}$) are derived from a mixture of four zero-mean normal distributions, where the relative variance for each mixture component is fixed:

$$\boldsymbol{p}(\boldsymbol{\beta}|\boldsymbol{\pi}, \sigma_g^2) = \boldsymbol{\pi_1} \times \boldsymbol{N}(\mathbf{0}, \mathbf{0} \times \sigma_g^2) + \boldsymbol{\pi_2} \times \boldsymbol{N}(\mathbf{0}, \mathbf{10^{-4}} \times \sigma_g^2) + \boldsymbol{\pi_3} \times \boldsymbol{N}(\mathbf{0}, \mathbf{10^{-3}} \times \sigma_g^2) + \boldsymbol{\pi_4} \times \boldsymbol{N}(\mathbf{0}, \mathbf{10^{-2}} \times \sigma_g^2)$$

where $\boldsymbol{\pi}$ is the mixture proportions that are constrained to sum to unity and $\boldsymbol{\sigma_g^2}$ is the additive genetic variance. BayesR software was used in this study, where the MCMC chain length was set to 10,000, and the burn-in step was set to 4,000.

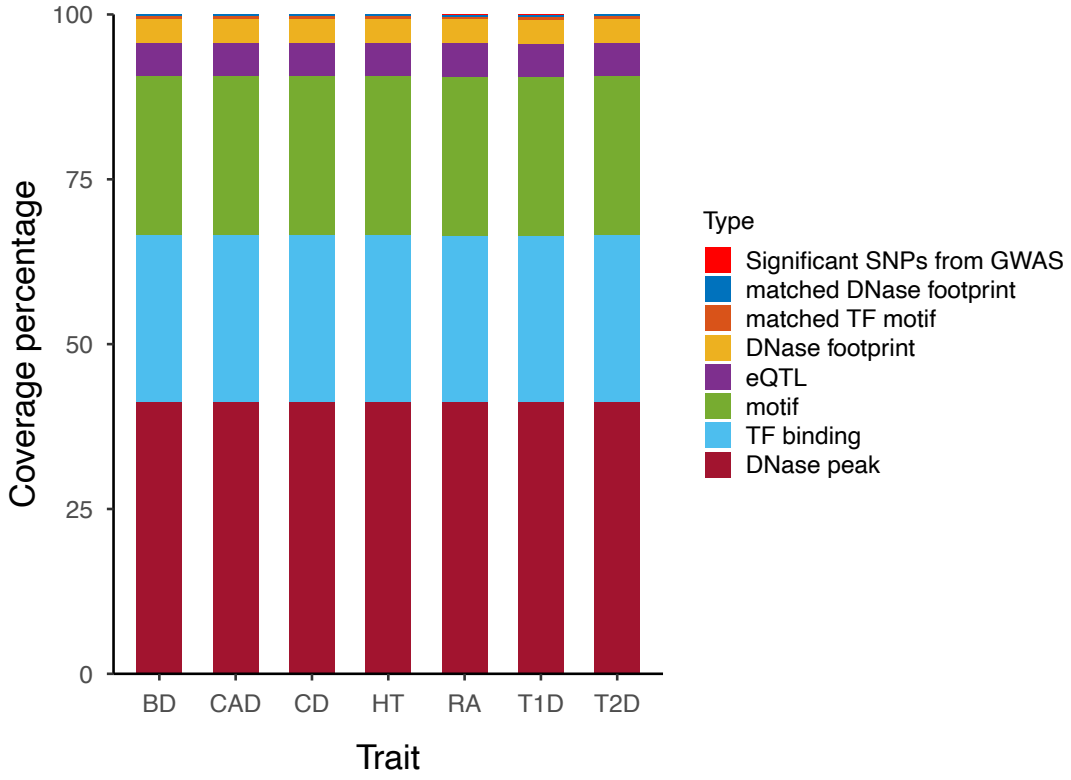
The BayesRC model incorporated prior biological information on the basis of BayesR model, for further details about BayesRC, see the study of MacLeod et al. Due the BayesRC is not adapted to cases where markers may have multiple annotations[7], we used the BayesRCO package to perferom the BayesRC, which was an extended version of BayesRC[8]. The MCMC chain length was set to 50,000, and the burn-in step was set to 20,000.

Calculation of SNP weights

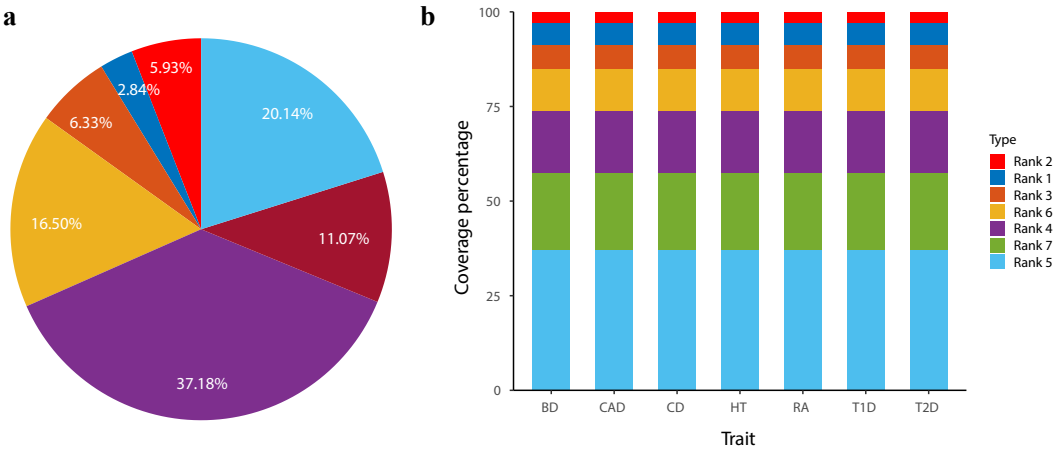
Following the approach of the KAML (Kinship-adjusted-multiple-loci) model for optimizing SNP weight strategy[9], the SNP weight $\boldsymbol{\xi_k}$ was determined using the following formula:

$$\xi_k|(\alpha,\beta) = \begin{cases} \mathbf{1}, & \mathbf{1-\beta} \\ \mathbf{1} + \boldsymbol{log_{\alpha}P_{m\beta}} - \boldsymbol{log_{\alpha}P_k}, & \boldsymbol{\beta} \end{cases}$$

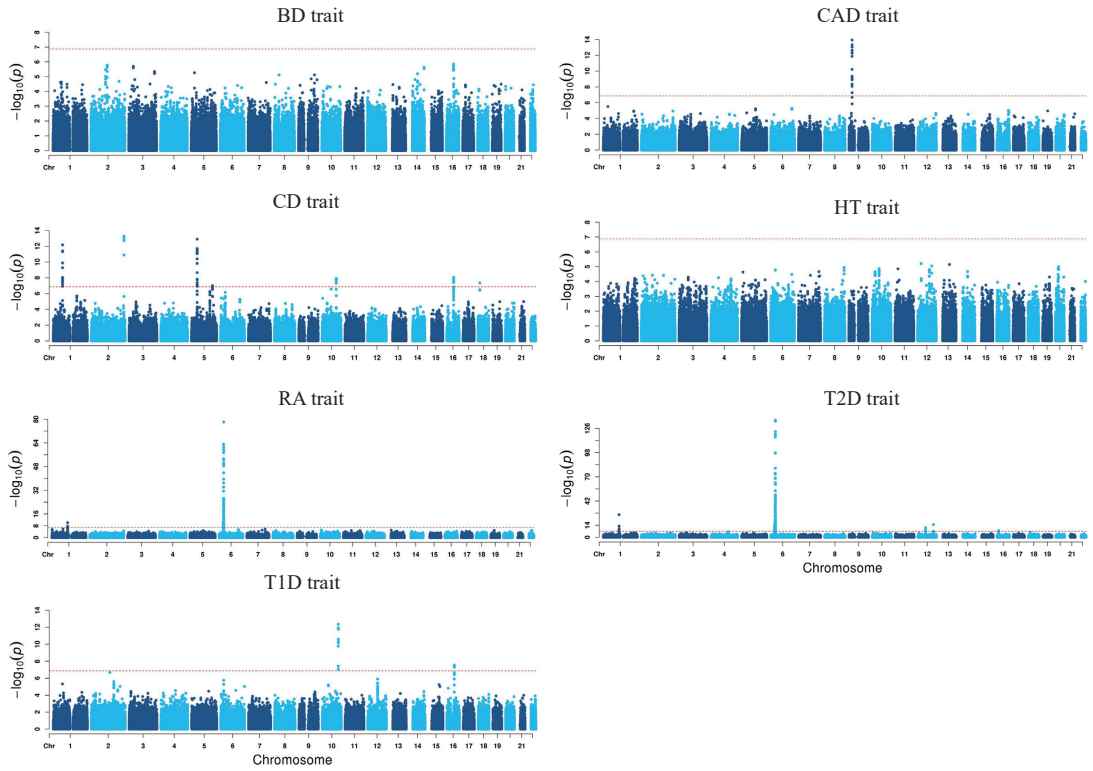
where $\boldsymbol{\alpha}$ is the base value of logarithmic function, $\boldsymbol{\beta}$ is the percentage of top significant SNPs to be weighted, which all were derived from the outputs of the KAML model that was estimated using the training set in cross-validation, $\boldsymbol{P}$ is the ordered *p* values of $\boldsymbol{k^{th}}$ SNP from the GWAS results, and $\boldsymbol{m}$ is the total number of markers.



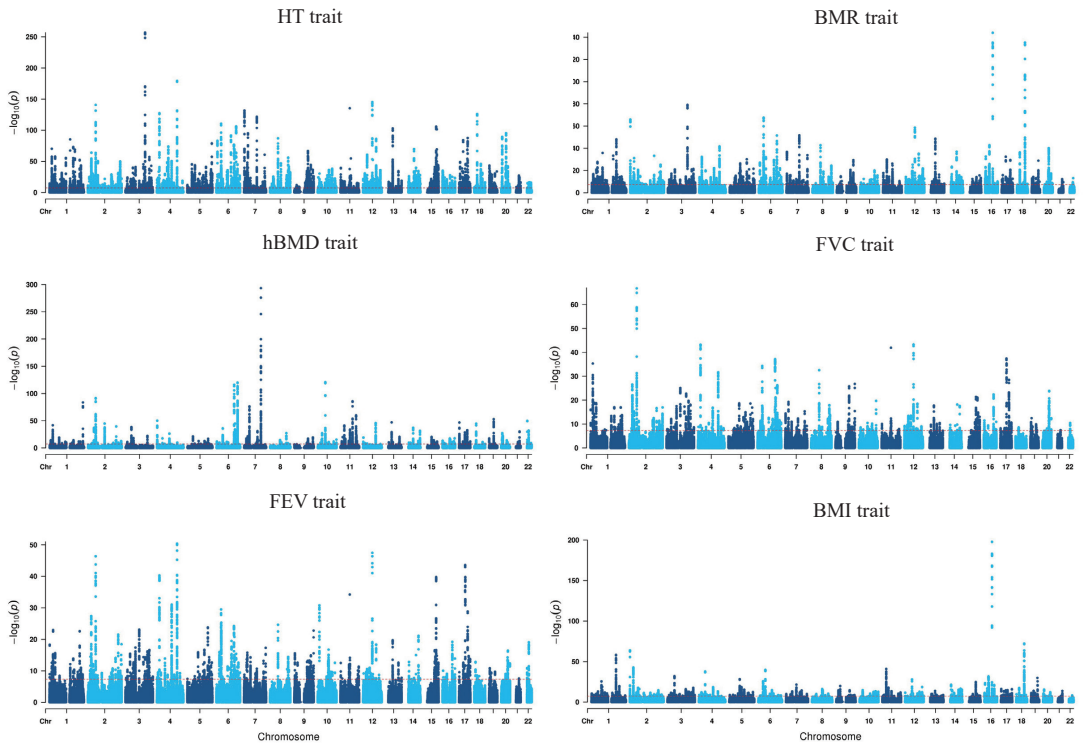
Supplementary Fig.1. The proportion of SNPs with distinct functional annotations for 7 traits. The annotation process utilized seven types of functional annotations from RegulomeDB database along with significant SNPs identified through genome-wide association studies (GWAS), see the Supplementary Table 8 for more details.



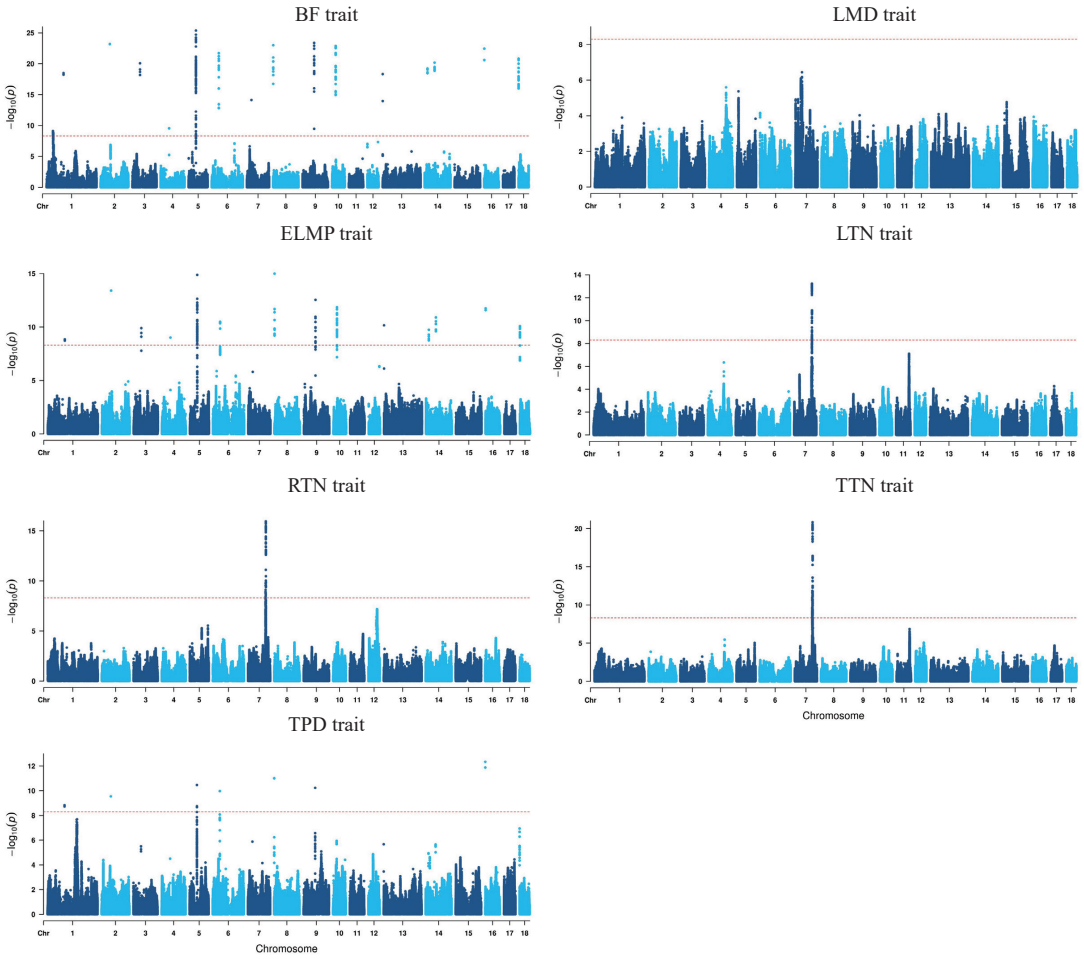
Supplementary Fig.2. **a**, The average proportion of SNPs with distinct functional annotations across 7 traits. **b**, The proportion of SNPs with distinct functional annotations for 7 traits. The annotation process utilized seven ranking functional annotations from RegulomeDB database. See the Supplementary Table 18 for more details.



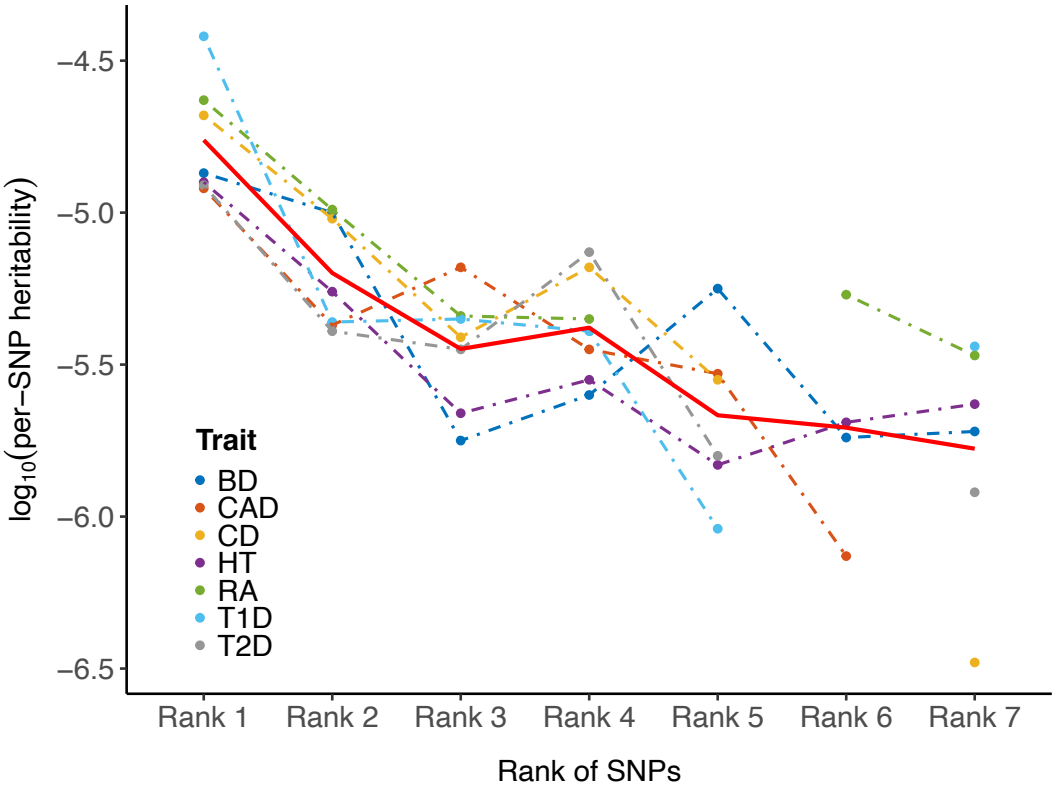
Supplementary Fig.3. The Manhattan plots for bipolar disorder (BD), coronary artery disease (CAD), Crohn's disease (CD), hypertension (HT), rheumatoid arthritis (RA), type 1 diabetes (T1D), and type 2 diabetes (T2D) trait from the WTCCC1 dataset. The *p*-values were computed using the general linear model of the rMVP software (v1.0.8). The red line represents the threshold of significance, the X axis is the genomic coordinates of markers, and the Y axis is the negative logarithm of the association *p*-value for each SNP.



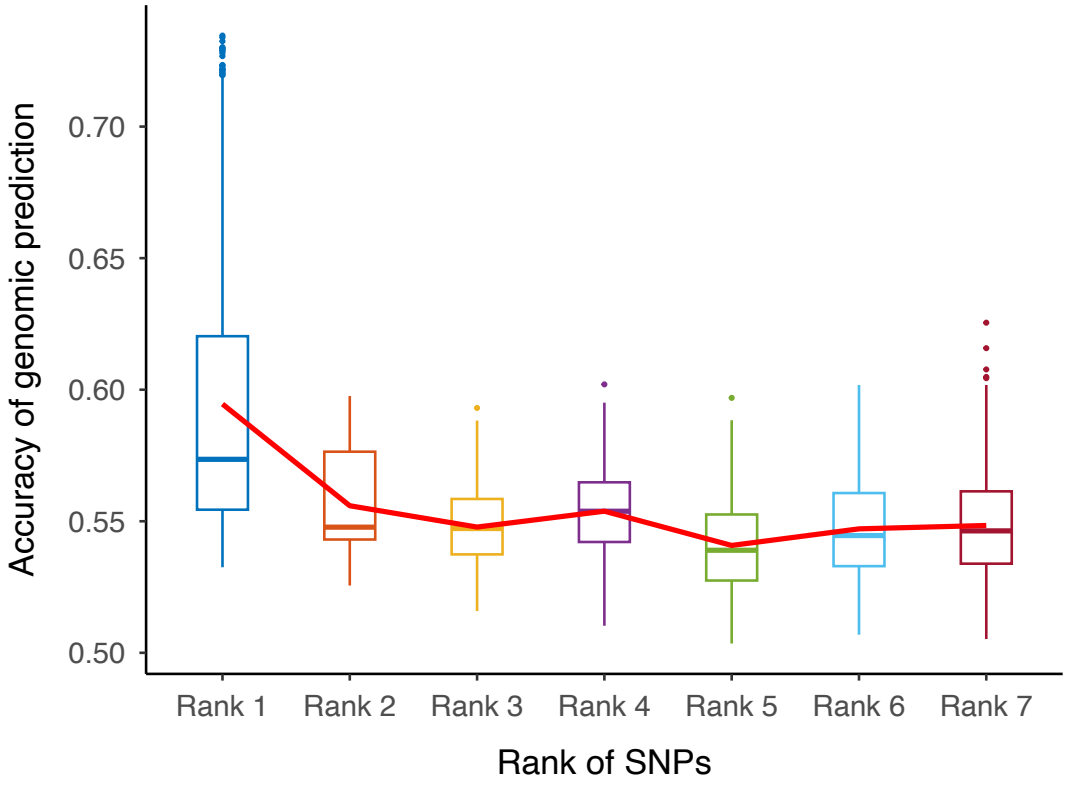
Supplementary Fig.4. The Manhattan plots for height (HT), basal metabolic rate (BMR), heel bone mineral density T-score (hBMD), forced vital capacity (FVC), forced expiratory volume in 1 s (FEV), and body mass index (BMI) trait from the UK BioBank HM3 dataset. The *p*-values were calculated using the linear regression method of PLINK software (v1.90). The red line represents the threshold of significance, the X axis is the genomic coordinates of markers, and the Y axis is the negative logarithm of the association *p*-value for each SNP.



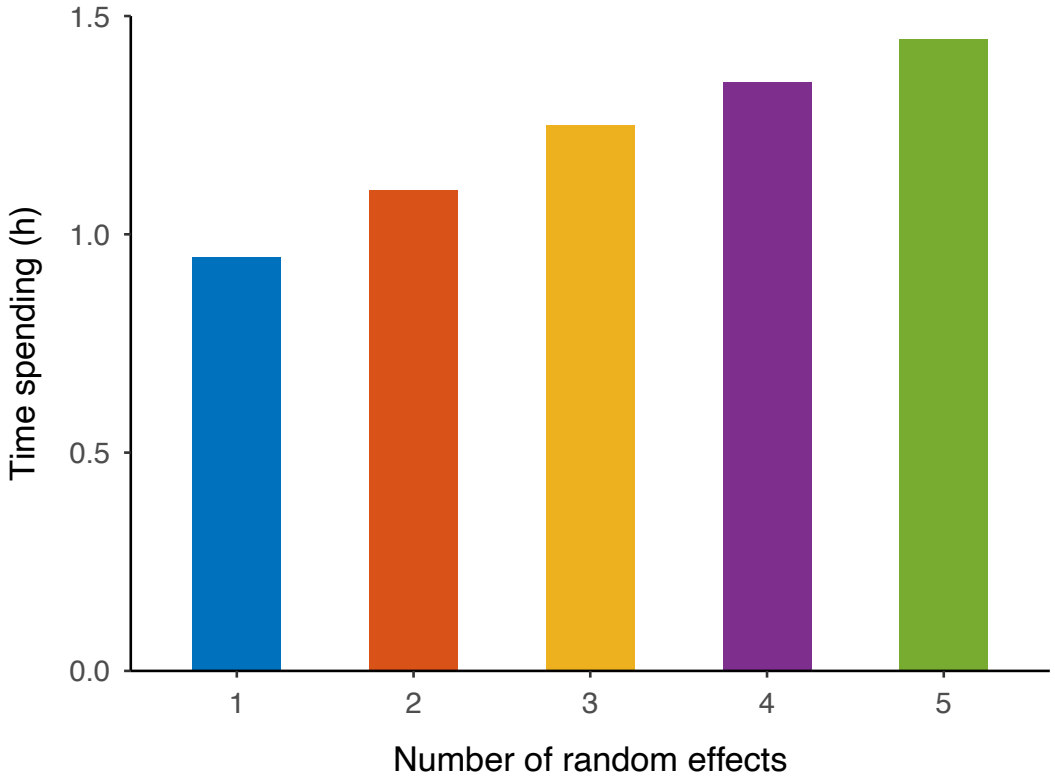
Supplementary Fig.5. The Manhattan plots for backfat thickness (BF), loin muscle depth (LMD), estimated lean meat percentage (ELMP), left teat number (LTN), right teat number (RTN), total teat number (TTN), and time spent eating per day (TPD) trait from the pig dataset. The *p*-values were estimated using the mixed linear model in rMVP software (v1.0.8). The red line represents the threshold of significance, the X axis is the genomic coordinates of markers, and the Y axis is the negative logarithm of the association *p*-value for each SNP.



Supplementary Fig.6. Examples of the importance assessment of functional annotation information. The seven ranking scores from the RegulomeDB database were viewed as the functional annotation information. The log10 scale of the per-SNP heritability (y-axis) was estimated from the IFAM model (no optimization of random effects) for each ranking score (x-axis). The average information restricted maximum likelihood (AI-REML) algorithm was chosen to estimate the heritability for WTCCC1 dataset. The dotted lines of different colors represent the values of different traits, and the solid red line is the mean value of all traits. See the Supplementary Table 19 for more details.



Supplementary Fig.7. The average prediction accuracy for seven disease traits in the WTCCC1 dataset. The annotations are the ranking scores from the RegulomeDB database. The analysis was carried out by the Genomic Best Linear Unbiased Prediction (GBLUP) model on the seven disease traits using the same number of SNPs that was selected randomly in the seven annotations. This process of random sampling was repeated 100 times. See the Supplementary Table 20 for more details.



Supplementary Fig.8. Comparisons of computational cost on time for different numbers of random effects within the IFAM model based on the "HE+PCG" strategy. The test was conducted using the heel bone mineral density T-score (hBMD) trait from the UK Biobank HM3 (UKB HM3) dataset, which involved around 200,000 individuals. Computational times were recorded for the estimation of the variance components using the Haseman–Elston (HE) regression algorithm and the additive genetics values using the pre-conditioned conjugate gradient (PCG) iterative algorithm. All results were computed on a Linux server with 48 CPUs and 1,440 GB memory. See the Supplementary Table 21 for more details.

Supplementary Table 1. The data information for the WTCCC1 dataset. This dataset had seven disease traits, and quality control was carried out using PLINK software (v1.90).

Trait	Ori data	Data after quality control
Bipolar disorder (BD)	458,868 SNPs, 4,806 samples, 1,868 cases	373,369 SNPs, 4,806 samples, 1,868 cases
Boronary artery disease (CAD)	458,868 SNPs, 4,864 samples, 1,926 cases	372,541 SNPs, 4,864 samples, 1,926 cases
Crohn's disease (CD)	458,868 SNPs, 4,686 samples, 1,748 cases	374,113 SNPs, 4,686 samples, 1,748 cases
Hypertension (HT)	458,868 SNPs, 4,890 samples, 1,952 cases	373,338 SNPs, 4,890 samples, 1,952 cases
Rheumatoid arthritis (RA)	458,868 SNPs, 4,798 samples, 1,860 cases	373,056 SNPs, 4,798 samples, 1,860 cases
Type 1 diabetes (T1D)	458,868 SNPs, 4,901 samples, 1,963 cases	372,964 SNPs, 4,901 samples, 1,963 cases
Type 2 diabetes (T2D)	458,868 SNPs, 4,862 samples, 1,924 cases	373,149 SNPs, 4,862 samples, 1,924 cases

Supplementary Table 2. The genomic functional annotations for human from the website of the LD Scores Regression (LDSC) model, there are 74 types and see the study about stratified LDSC model for more details[10].

Numbe	Type	Number	Type
1	Ancient_Sequence_Age_Human_Enhancer	38	H3K4me3_Trynka.extend.500
2	Ancient_Sequence_Age_Human_Enhancer.extend.500	39	H3K9ac_peaks_Trynka
3	Ancient_Sequence_Age_Human_Promoter	40	H3K9ac_Trynka
4	Ancient_Sequence_Age_Human_Promoter.extend.500	41	H3K9ac_Trynka.extend.500
5	BivFlnk	42	Human_Enhancer_Villar
6	BivFlnk.extend.500	43	Human_Enhancer_Villar.extend.500
7	Coding_UCSC	44	Human_Promoter_Villar_ExAC
8	Coding_UCSC.extend.500	45	Human_Promoter_Villar_ExAC.extend.5
9	Conserved_LindbladToh	46	00 Human_Promoter_Villar
10	Conserved_LindbladToh.extend.500	47	Human_Promoter_Villar.extend.500
11	Conserved_Mammal_phastCons46way	48	Intron_UCSC
12	Conserved_Mammal_phastCons46way.extend.500	49	Intron_UCSC.extend.500
13	Conserved_Primate_phastCons46way	50	Promoter_UCSC
14	Conserved_Primate_phastCons46way.extend.500	51	Promoter_UCSC.extend.500
15	Conserved_Vertebrate_phastCons46way	52	PromoterFlanking_Hoffman
16	Conserved_Vertebrate_phastCons46way.extend.500	53	PromoterFlanking_Hoffman.extend.500
17	CTCF_Hoffman	54	Repressed_Hoffman
18	CTCF_Hoffman.extend.500	55	Repressed_Hoffman.extend.500
19	DGF_ENCODE	56	Repressed_Hoffman.extend.500
20	DGF_ENCODE.extend.500	57	SuperEnhancer_Hnisz
21	DHS_Trynka	58	SuperEnhancer_Hnisz.extend.500
22	DHS_Trynka.extend.500	59	TFBS_ENCODE
23	Enhancer_Andersson	60	TFBS_ENCODE.extend.500
24	Enhancer_Andersson.extend.500	61	Transcribed_Hoffman
25	Enhancer_Hoffman	62	Transcribed_Hoffman.extend.500
26	Enhancer_Hoffman.extend.500	63	TSS_Hoffman
27	FetalDHS_Trynka	64	TSS_Hoffman.extend.500
28	FetalDHS_Trynka.extend.500	65	UTR_3_UCSC
29	H3K27ac_Hnisz	66	UTR_3_UCSC.extend.500
30	H3K27ac_Hnisz.extend.500	67	UTR_5_UCSC
31	H3K27ac_PGC2	68	UTR_5_UCSC.extend.500
32	H3K27ac_PGC2.extend.500	69	Vahedi_Tcell_SE_500bp
33	H3K4me1_peaks_Trynka	70	Vahedi_Tcell_SE
34	H3K4me1_Trynka	71	Vahedi_Tcell_TE_500bp
35	H3K4me1_Trynka.extend.500	72	Vahedi_Tcell_TE
36	H3K4me3_peaks_Trynka	73	WeakEnhancer_Hoffman
37	H3K4me3_Trynka	74	WeakEnhancer_Hoffman.extend.500

Supplementary Table 3. The functional annotations for human from the RegulomeDB database which have seven types.

Number	Type
1	eQTL
2	TF binding
3	matched TF motif
4	motif
5	matched DNase footprint
6	DNase footprint
7	DNase peak

Supplementary Table 4. The RegulomeDB variant classification scheme. There are seven ranks.

Category	SNPs	Description
Rank 1	75,189	Linked to expression of gene parget and likely to affect binding eQTL + TF binding + matched TF motif + any motif + matched DNase footprint + DNase footprint + DNase peak
Rank 2	449,374	Likely to affect binding TF binding + matched TF motif + any motif + matched DNase footprint + DNase footprint + DNase peak
Rank 3	847,825	Less likely to affect binding TF binding + any motif + matched TF motif + DNase peak
Rank 4	1,831,986	Minimal binding evidence TF binding + DNase peak
Rank 5	4,880,805	TF binding / DNase peak
Rank 6	2,469,270	Motif hit
Rank 7	2,720,574	Other SNPs on genome

Supplementary Table 5. The functional annotations for pig from the IFmut database. There were 10 types.

Number	Type
1	open chromatin region
2	nucleosome free region
3	footprint
4	matched_motif
5	intersect_motif
6	any_motif
7	active_promoter_narrowPeak
8	enhancer_narrowPeak
9	active_promoter
10	enhancer

Supplementary Table 6. The number of overlapping SNPs between functional annotations from the RegulomeDB database. Taking the bipolar disorder (BD) trait from the WTCCC1 dataset as an example, the number of annotated SNPs are shown on the diagonal and the overlap in SNPs between annotations are shown above the diagonal.

Annotations	Matched Dnase Footprint	Matched TF motif	Dnase Footprint	eQTL	motif	TF binding	DNase peak
Matched Dnase Footprint	974	943	974	153	953	974	953
Matched TF motif		2,627	1,268	342	2,627	2,627	2,372
Dnase Footprint			19,882	2,354	16,967	13,789	17,655
eQTL				27,918	9,923	14,361	19,526
motif					13,1832	49,994	78,880
TF binding						138,825	107,280
DNase peak							225,259

Supplementary Table 7. The variance component and genomic prediction accuracy were estimated by a genomic prediction model that contains multiple random effects, GBLUP and IFAM model

Trait[1]	R1-VC[2]	R2-VC	R3-VC	R4-VC	R5-VC	R6-VC	R7-VC	MR-VC[3]	GBLUP-VC[3]	IFAM-VC[3]	MR-Accuracy[4]	GBLUP-Accuracy[4]	IFAM-Accuracy[4]
BD	0.0012	0.0058	0.0150	0.0073	0.0293	0.0430	0.0690	0.1706	0.1720	0.1626	0.6456	0.6474	0.6463
CAD	0.0015	0.0051	0.0063	0.0072	0.0220	0.0437	0.0456	0.1314	0.1162	0.1147	0.5895	0.5845	0.6045
CD	0.0076	0.0197	0.0167	0.0227	0.0000	0.0553	0.0512	0.1731	0.1521	0.1419	0.6393	0.6322	0.6758
HT	0.0030	0.0024	0.0066	0.0106	0.0297	0.0271	0.0449	0.1243	0.1181	0.1187	0.5921	0.5947	0.5951
RA	0.0121	0.0188	0.0215	0.0359	0.0337	0.0435	0.0000	0.1655	0.1405	0.0993	0.6378	0.6113	0.7215
T1D	0.0117	0.0198	0.0240	0.0449	0.0000	0.0327	0.0427	0.1759	0.1721	0.2069	0.6975	0.6542	0.8699
T2D	0.0016	0.0130	0.0235	0.0133	0.0027	0.0334	0.0642	0.1516	0.1200	0.1204	0.5983	0.5968	0.6156

[1] BD: bipolar disorder; CAD: coronary artery disease; CD: Crohn's disease; HT: hypertension; RA: rheumatoid arthritis; T1D: type 1 diabetes; T2D: type 2 diabetes;

[2] The variance component estimate for each genomic annotation from the RegulomeDB database which has seven annotations using a genomic prediction model that contains multiple random effects;

[3] The estimates of variance component were calculated by a genomic prediction model that contains multiple random effects (MR), GBLUP and IFAM model;

4 The genomic prediction accuracy were estimated by a genomic prediction model that contains multiple random effects (MR), GBLUP and IFAM model;

Supplementary Table 8. The number and proportion of SNPs with distinct functional annotations for 7 traits in the WTCCC1 dataset. The annotation process utilized seven types of functional annotations from RegulomeDB database along with significant SNPs identified through genome-wide association studies (GWAS) (Supplementary Table 22). The numbers outside the parentheses represent the number of annotated SNPs and the numbers inside the parentheses represent the proportion of annotated SNPs across the genome.

Number	Type of annotation	Bipolar disorder (BD)	Coronary artery disease (CAD)	Crohn's disease (CD)	Hypertension (HT)	Rheumatoid arthritis (RA)	Type 1 diabetes (T1D)	Type 2 diabetes (T2D)
1	Significant SNPs from GWAS analysis	1(0.00%)	50(0.01%)	178(0.05%)	23(0.01%)	586(0.16%)	825(0.22%)	40(0.01%)
2	eQTL	27,918(7.48%)	27,914(7.49%)	27,935(7.47%)	27,872(7.47%)	27,855(7.47%)	27,652(7.41%)	27,786(7.45%)
3	TF binding	138,825(37.18%)	138,306(37.13%)	139,239(37.22%)	138,878(37.20%)	138,820(37.21%)	138,706(37.19%)	138,797(37.20%)
4	matched TF motif	2,627(0.70%)	2,631(0.71%)	2,628(0.70%)	2,647(0.71%)	2,620(0.70%)	2,635(0.71%)	2,647(0.71%)
5	motif	131,832(35.31%)	131,544(35.31%)	132,320(35.37%)	131,745(35.29%)	131,730(35.31%)	131,690(35.31%)	131,701(35.29%)
6	matched DNase footprint	974(0.26%)	981(0.26%)	979(0.26%)	987(0.26%)	972(0.26%)	974(0.26%)	984(0.26%)
7	DNase footprint	19,882(5.33%)	19,776(5.31%)	19,938(5.33%)	19,869(5.32%)	19,837(5.32%)	19,846(5.32%)	19,874(5.33%)
8	DNase peak	225,259(60.33%)	224,668(60.31%)	225,832(60.36%)	225,308(60.35%)	225,179(60.36%)	225,068(60.35%)	225,205(60.35%)

Supplementary Table 9. The average number and proportion of SNPs with distinct functional annotations in the UK BioBank (UKB) HM3 dataset. The annotation process utilized seven types of functional annotations from RegulomeDB database along with significant SNPs identified through genome-wide association studies (GWAS) (Supplementary Table 23). The numbers outside the parentheses represent the number of annotated SNPs, and the numbers inside the parentheses represent the proportion of annotated SNPs across the genome.

Number	Type of annotation	UKB HM3
1	Significant SNPs from GWAS analysis	6,013(0.55%)
2	eQTL	56,262(5.14%)
3	TF binding	439,789(40.17%)
4	matched TF motif	9,324(0.85%)
5	motif	415,621(37.96%)
6	matched DNase footprint	3,990(0.36%)
7	DNase footprint	71,470(6.53%)
8	DNase peak	692,277(63.23%)

Supplementary Table 10. The average number and proportion of SNPs with distinct functional annotations in the pig dataset. The annotation process utilized ten types of functional annotations from IFmut database along with significant SNPs identified through genome-wide association studies (GWAS) (Supplementary Table 24). The numbers outside the parentheses represent the number of annotated SNPs, and the numbers inside the parentheses represent the proportion of annotated SNPs across the genome.

Number	Type of annotation	Pig dataset
1	Significant SNPs from GWAS analysis	2,597(0.03%)
2	Open Chromatin Region	1,252,515(12.62%)
3	Nucleosome Free Region	372,221(3.75%)
4	footprint	74,769(0.75%)
5	matched_motif	4,178(0.04%)
6	intersect_motif	11,303(0.11%)
7	any_motif	477,510(4.81%)
8	active_promoter_narrowPeak	1,357,705(13.68%)
9	enhancer_narrowPeak	4,410,240(44.45%)
10	active_promoter	631,481(6.36%)
11	enhancer	2,858,528(28.81%)

Supplementary Table 11. The number and proportion of SNPs with distinct functional annotations for 7 traits in the WTCCC1 dataset. The annotation process utilized 74 types of functional annotations from the website of the LD Scores Regression (LDSC) model along with significant SNPs identified through genome-wide association studies (GWAS) (Supplementary Table 22). The numbers outside the parentheses represent the number of annotated SNPs and the numbers inside the parentheses represent the proportion of annotated SNPs across the genome.

Number	Type of annotation	Bipolar disorder (BD)	Coronary artery disease (CAD)	Crohn's disease (CD)	Hypertension (HT)	Rheumatoid arthritis (RA)	Type 1 diabetes (T1D)	Type 2 diabetes (T2D)
1	Significant SNPs from GWAS analysis	1(0.00%)	50(0.01%)	178(0.05%)	23(0.01%)	586(0.16%)	825(0.22%)	40(0.01%)
2	Ancient_Sequence_Age_Human_Enhancer	2,588(0.69%)	2,566(0.69%)	2,580(0.69%)	2,577(0.69%)	2,585(0.69%)	2,594(0.70%)	2,575(0.69%)
3	Ancient_Sequence_Age_Human_Enhancer.extend.500	6,286(1.68%)	6,249(1.68%)	6,247(1.67%)	6,280(1.68%)	6,280(1.68%)	6,272(1.68%)	6,209(1.66%)
4	Ancient_Sequence_Age_Human_Promoter	1,839(0.49%)	1,841(0.49%)	1,848(0.49%)	1,847(0.49%)	1,839(0.49%)	1,842(0.49%)	1,818(0.49%)
5	Ancient_Sequence_Age_Human_Promoter.extend.500	3,879(1.04%)	3,868(1.04%)	3,873(1.04%)	3,888(1.04%)	3,878(1.04%)	3,890(1.04%)	3,854(1.03%)
6	BivFlnk	4,823(1.29%)	4,824(1.29%)	4,822(1.29%)	4,820(1.29%)	4,784(1.28%)	4,824(1.29%)	4,853(1.30%)
7	BivFlnk.extend.500	10,589(2.84%)	10,549(2.83%)	10,612(2.84%)	10,603(2.84%)	10,520(2.82%)	10,609(2.84%)	10,600(2.84%)
8	Coding_UCSC	7,605(2.04%)	7,608(2.04%)	7,606(2.03%)	7,591(2.03%)	7,593(2.04%)	7,617(2.04%)	7,622(2.04%)
9	Coding_UCSC.extend.500	26,831(7.19%)	26,787(7.19%)	26,866(7.18%)	26,828(7.19%)	26,784(7.18%)	26,880(7.21%)	26,772(7.17%)
10	Conserved_LindbladToh	16,786(4.50%)	16,787(4.51%)	16,820(4.50%)	16,787(4.50%)	16,773(4.50%)	16,791(4.50%)	16,786(4.50%)
11	Conserved_LindbladToh.extend.500	139,987(37.49%)	139,821(37.53%)	140,354(37.52%)	140,154(37.54%)	139,849(37.49%)	140,002(37.54%)	139,750(37.45%)
12	Conserved_Mammal_phastCons46way	15,817(4.24%)	15,805(4.24%)	15,856(4.24%)	15,803(4.23%)	15,825(4.24%)	15,823(4.24%)	15,825(4.24%)
13	Conserved_Mammal_phastCons46way.extend.500	143,853(38.53%)	143,713(38.58%)	144,237(38.55%)	144,002(38.57%)	143,697(38.52%)	143,877(38.58%)	143,654(38.50%)
14	Conserved_Primate_phastCons46way	14,140(3.79%)	14,136(3.79%)	14,148(3.78%)	14,134(3.79%)	14,109(3.78%)	14,134(3.79%)	14,110(3.78%)
15	Conserved_Primate_phastCons46way.extend.500	79,829(21.38%)	79,798(21.42%)	80,149(21.42%)	79,969(21.42%)	79,794(21.39%)	79,881(21.42%)	79,700(21.36%)
16	Conserved_Vertebrate_phastCons46way	18,964(5.08%)	18,952(5.09%)	19,026(5.09%)	18,940(5.07%)	18,945(5.08%)	18,961(5.08%)	18,964(5.08%)
17	Conserved_Vertebrate_phastCons46way.extend.500	167,772(44.93%)	167,566(44.98%)	168,217(44.96%)	167,969(44.99%)	167,635(44.94%)	167,834(45.00%)	167,550(44.90%)
18	CTCF_Hoffman	8,851(2.37%)	8,807(2.36%)	8,835(2.36%)	8,844(2.37%)	8,822(2.36%)	8,845(2.37%)	8,849(2.37%)
19	CTCF_Hoffman.extend.500	25,652(6.87%)	25,606(6.87%)	25,655(6.86%)	25,667(6.88%)	25,641(6.87%)	25,642(6.88%)	25,682(6.88%)
20	DGF_ENCODE	53,055(14.21%)	52,930(14.21%)	53,283(14.24%)	53,072(14.22%)	53,100(14.23%)	53,022(14.22%)	53,016(14.21%)
21	DGF_ENCODE.extend.500	204,797(54.85%)	204,194(54.81%)	205,457(54.92%)	204,845(54.87%)	204,691(54.87%)	204,648(54.87%)	204,659(54.85%)
22	DHS_peaks_Trynka	45,162(12.10%)	45,014(12.08%)	45,289(12.11%)	45,166(12.10%)	45,144(12.10%)	45,097(12.09%)	45,137(12.10%)
23	DHS_Trynka	67,999(18.21%)	67,849(18.21%)	68,213(18.23%)	68,031(18.22%)	68,001(18.23%)	67,924(18.21%)	67,874(18.19%)
24	DHS_Trynka.extend.500	193,025(51.70%)	192,351(51.63%)	193,477(51.72%)	192,930(51.68%)	192,817(51.69%)	192,707(51.67%)	192,761(51.66%)
25	Enhancer_Andersson	1,676(0.45%)	1,665(0.45%)	1,684(0.45%)	1,675(0.45%)	1,658(0.44%)	1,679(0.45%)	1,668(0.45%)
26	Enhancer_Andersson.extend.500	7,206(1.93%)	7,204(1.93%)	7,232(1.93%)	7,220(1.93%)	7,188(1.93%)	7,174(1.92%)	7,168(1.92%)
27	Enhancer_Hoffman	16,725(4.48%)	16,685(4.48%)	16,803(4.49%)	16,750(4.49%)	16,692(4.47%)	16,717(4.48%)	16,756(4.49%)
28	Enhancer_Hoffman.extend.500	34,490(9.24%)	34,409(9.24%)	34,671(9.27%)	34,523(9.25%)	34,491(9.25%)	34,440(9.23%)	34,463(9.24%)

29	FetalDHS_Trynka	35,202(9.43%)	35,055(9.41%)	35,303(9.44%)	35,164(9.42%)	35,157(9.42%)	35,158(9.43%)	35,157(9.42%)
30	FetalDHS_Trynka.extend.500	112,451(30.12%)	112,078(30.08%)	112,819(30.16%)	112,372(30.10%)	112,429(30.14%)	112,395(30.14%)	112,327(30.10%)
31	H3K27ac_Hnisz	150,403(40.28%)	150,207(40.32%)	150,963(40.35%)	150,478(40.31%)	150,326(40.30%)	150,305(40.30%)	150,326(40.29%)
32	H3K27ac_Hnisz.extend.500	161,880(43.36%)	161,626(43.38%)	162,525(43.44%)	162,003(43.39%)	161,862(43.39%)	161,807(43.38%)	161,834(43.37%)
33	H3K27ac_PGC2	105,723(28.32%)	105,516(28.32%)	106,120(28.37%)	105,865(28.36%)	105,698(28.33%)	105,654(28.33%)	105,575(28.29%)
34	H3K27ac_PGC2.extend.500	130,593(34.98%)	130,364(34.99%)	131,059(35.03%)	130,723(35.01%)	130,528(34.99%)	130,557(35.01%)	130,472(34.97%)
35	H3K4me1_peaks_Trynka	68,965(18.47%)	68,742(18.45%)	69,184(18.49%)	68,902(18.46%)	68,908(18.47%)	68,873(18.47%)	68,835(18.45%)
36	H3K4me1_Trynka	168,064(45.01%)	167,789(45.04%)	168,664(45.08%)	168,229(45.06%)	168,031(45.04%)	168,109(45.07%)	167,895(44.99%)
37	H3K4me1_Trynka.extend.500	233,792(62.62%)	233,377(62.64%)	234,533(62.69%)	233,863(62.64%)	233,647(62.63%)	233,709(62.66%)	233,696(62.63%)
38	H3K4me3_peaks_Trynka	16,030(4.29%)	16,020(4.30%)	16,126(4.31%)	16,077(4.31%)	15,986(4.29%)	15,998(4.29%)	16,077(4.31%)
39	H3K4me3_Trynka	50,118(13.42%)	50,025(13.43%)	50,315(13.45%)	50,135(13.43%)	50,117(13.43%)	50,055(13.42%)	50,111(13.43%)
40	H3K4me3_Trynka.extend.500	94,026(25.18%)	93,806(25.18%)	94,404(25.23%)	94,092(25.20%)	94,038(25.21%)	93,897(25.18%)	93,909(25.17%)
41	H3K9ac_peaks_Trynka	14,957(4.01%)	14,946(4.01%)	15,032(4.02%)	15,004(4.02%)	14,953(4.01%)	14,962(4.01%)	14,951(4.01%)
42	H3K9ac_Trynka	48,333(12.95%)	48,237(12.95%)	48,522(12.97%)	48,376(12.96%)	48,339(12.96%)	48,341(12.96%)	48,339(12.95%)
43	H3K9ac_Trynka.extend.500	86,293(23.11%)	86,055(23.10%)	86,652(23.16%)	86,349(23.13%)	86,338(23.14%)	86,325(23.15%)	86,267(23.12%)
44	Human_Enhancer_Villar	13,138(3.52%)	13,072(3.51%)	13,100(3.50%)	13,047(3.49%)	13,118(3.52%)	13,083(3.51%)	12,998(3.48%)
45	Human_Enhancer_Villar.extend.500	16,720(4.48%)	16,615(4.46%)	16,663(4.45%)	16,619(4.45%)	16,646(4.46%)	16,607(4.45%)	16,511(4.42%)
46	Human_Promoter_Villar_ExAC	1,261(0.34%)	1,251(0.34%)	1,252(0.33%)	1,240(0.33%)	1,261(0.34%)	1,254(0.34%)	1,247(0.33%)
47	Human_Promoter_Villar_ExAC.extend.500	1,485(0.40%)	1,473(0.40%)	1,473(0.39%)	1,461(0.39%)	1,479(0.40%)	1,475(0.40%)	1,465(0.39%)
48	Human_Promoter_Villar	5,985(1.60%)	5,952(1.60%)	6,002(1.60%)	5,976(1.60%)	5,995(1.61%)	6,003(1.61%)	5,980(1.60%)
49	Human_Promoter_Villar.extend.500	7,276(1.95%)	7,228(1.94%)	7,292(1.95%)	7,273(1.95%)	7,278(1.95%)	7,279(1.95%)	7,259(1.95%)
50	Intron_UCSC	154,652(41.42%)	154,369(41.44%)	154,900(41.40%)	154,705(41.44%)	154,488(41.41%)	154,634(41.46%)	154,404(41.38%)
51	Intron_UCSC.extend.500	159,383(42.69%)	159,121(42.71%)	159,648(42.67%)	159,472(42.72%)	159,226(42.68%)	159,381(42.73%)	159,177(42.66%)
52	Promoter_UCSC	17,476(4.68%)	17,480(4.69%)	17,539(4.69%)	17,465(4.68%)	17,442(4.68%)	17,496(4.69%)	17,526(4.70%)
53	Promoter_UCSC.extend.500	21,347(5.72%)	21,378(5.74%)	21,444(5.73%)	21,354(5.72%)	21,310(5.71%)	21,389(5.73%)	21,398(5.73%)
54	PromoterFlanking_Hoffman	3,472(0.93%)	3,454(0.93%)	3,478(0.93%)	3,440(0.92%)	3,448(0.92%)	3,473(0.93%)	3,466(0.93%)
55	PromoterFlanking_Hoffman.extend.500	12,713(3.40%)	12,702(3.41%)	12,770(3.41%)	12,719(3.41%)	12,704(3.41%)	12,690(3.40%)	12,719(3.41%)
56	Repressed_Hoffman	168,511(45.13%)	168,172(45.14%)	168,707(45.10%)	168,404(45.11%)	168,341(45.12%)	168,359(45.14%)	168,394(45.13%)
57	Repressed_Hoffman.extend.500	262,903(70.41%)	262,320(70.41%)	263,237(70.36%)	262,793(70.39%)	262,617(70.40%)	262,547(70.39%)	262,576(70.37%)
58	SuperEnhancer_Hnisz	64,510(17.28%)	64,400(17.29%)	64,886(17.34%)	64,565(17.29%)	64,448(17.28%)	64,511(17.30%)	64,561(17.30%)
59	SuperEnhancer_Hnisz.extend.500	65,699(17.60%)	65,578(17.60%)	66,073(17.66%)	65,739(17.61%)	65,637(17.59%)	65,711(17.62%)	65,737(17.62%)
60	TFBS_ENCODE	50,482(13.52%)	50,343(13.51%)	50,706(13.55%)	50,578(13.55%)	50,472(13.53%)	50,518(13.55%)	50,515(13.54%)
61	TFBS_ENCODE.extend.500	129,432(34.67%)	128,920(34.61%)	129,767(34.69%)	129,452(34.67%)	129,237(34.64%)	129,225(34.65%)	129,295(34.65%)
62	Transcribed_Hoffman	132,766(35.56%)	132,595(35.59%)	133,134(35.59%)	132,880(35.59%)	132,834(35.61%)	132,644(35.56%)	132,765(35.58%)
63	Transcribed_Hoffman.extend.500	284,250(76.13%)	283,603(76.13%)	284,762(76.12%)	284,251(76.14%)	284,051(76.14%)	283,872(76.11%)	284,052(76.12%)
64	TSS_Hoffman	6,635(1.78%)	6,602(1.77%)	6,647(1.78%)	6,640(1.78%)	6,606(1.77%)	6,604(1.77%)	6,616(1.77%)
65	TSS_Hoffman.extend.500	12,645(3.39%)	12,595(3.38%)	12,657(3.38%)	12,621(3.38%)	12,591(3.38%)	12,595(3.38%)	12,606(3.38%)
66	UTR_3_UCSC	4,990(1.34%)	5,007(1.34%)	5,002(1.34%)	4,991(1.34%)	4,975(1.33%)	5,015(1.34%)	4,964(1.33%)
67	UTR_3_UCSC.extend.500	11,015(2.95%)	11,024(2.96%)	11,030(2.95%)	11,000(2.95%)	10,969(2.94%)	11,082(2.97%)	10,977(2.94%)
68	UTR_5_UCSC	2,315(0.62%)	2,311(0.62%)	2,328(0.62%)	2,323(0.62%)	2,296(0.62%)	2,327(0.62%)	2,335(0.63%)
69	UTR_5_UCSC.extend.500	10,388(2.78%)	10,359(2.78%)	10,434(2.79%)	10,399(2.79%)	10,373(2.78%)	10,418(2.79%)	10,407(2.79%)

70	Vahedi_Tcell_SE_500bp	8,520(2.28%)	8,526(2.29%)	8,547(2.28%)	8,503(2.28%)	8,545(2.29%)	8,478(2.27%)	8,474(2.27%)
71	Vahedi_Tcell_SE	8,370(2.24%)	8,372(2.25%)	8,399(2.25%)	8,357(2.24%)	8,398(2.25%)	8,326(2.23%)	8,324(2.23%)
72	Vahedi_Tcell_TE_500bp	9,855(2.64%)	9,828(2.64%)	9,881(2.64%)	9,864(2.64%)	9,852(2.64%)	9,758(2.62%)	9,824(2.63%)
73	Vahedi_Tcell_TE	8,266(2.21%)	8,262(2.22%)	8,295(2.22%)	8,289(2.22%)	8,274(2.22%)	8,190(2.20%)	8,242(2.21%)
74	WeakEnhancer_Hoffman	8,034(2.15%)	8,022(2.15%)	8,070(2.16%)	8,033(2.15%)	8,028(2.15%)	8,025(2.15%)	8,023(2.15%)
75	WeakEnhancer_Hoffman.extend.500	33,327(8.93%)	33,159(8.90%)	33,456(8.94%)	33,313(8.92%)	33,337(8.94%)	33,242(8.91%)	33,291(8.92%)

Supplementary Table 12. Examples of the importance assessment of functional annotation information in the WTCCC1 dataset. The seven ranking scores from the RegulomeDB database were viewed as the functional annotation information. The average information restricted maximum likelihood (AI-REML) algorithm was chosen to estimate the variance component.

Category	BD[1]	CAD	CD	HT	RA	T1D	T2D	Mean
Rank 1	3.21E-06	2.90E-06	4.82E-06	3.04E-06	5.45E-06	8.62E-06	2.89E-06	4.42E-06
Rank 2	2.38E-06	1.01E-06	2.21E-06	1.33E-06	2.42E-06	9.96E-07	9.55E-07	1.61E-06
Rank 3	4.24E-07	1.59E-06	8.96E-07	5.28E-07	1.07E-06	1.01E-06	8.38E-07	9.07E-07
Rank 4	5.96E-07	8.48E-07	1.51E-06	6.72E-07	1.04E-06	9.29E-07	1.74E-06	1.05E-06
Rank 5	1.34E-06	6.98E-07	6.45E-07	3.53E-07	--	2.04E-07	3.76E-07	5.16E-07
Rank 6	4.32E-07	1.77E-07	--	4.94E-07	1.27E-06	--	--	3.39E-07
Rank 7	4.53E-07	--	7.56E-08	5.57E-07	7.96E-07	8.21E-07	2.87E-07	4.27E-07

[1] BD: bipolar disorder; CAD: coronary artery disease; CD: Crohn's disease; HT: hypertension; RA: rheumatoid arthritis; T1D: type 1 diabetes; T2D: type 2 diabetes

Supplementary Table 13. Examples of the importance assessment of functional annotation information in the UKB dataset. The seven ranking scores from the RegulomeDB database were viewed as the functional annotation information. The Haseman–Elston (HE) regression algorithm was chosen to estimate the variance component and heritability.

Category	HT[1]	BMR	hBMD	FVC	BMI	FEV	Mean
Rank 1	4.90E-06	1.78E-06	2.15E-06	1.53E-06	1.55E-06	1.58E-06	2.25E-06
Rank 2	2.58E-06	1.17E-06	6.67E-07	8.54E-07	8.12E-07	8.60E-07	1.16E-06
Rank 3	1.77E-06	6.65E-07	1.37E-06	5.42E-07	3.69E-07	5.06E-07	8.70E-07
Rank 4	3.14E-06	1.68E-06	1.32E-06	1.03E-06	8.94E-07	1.06E-06	1.52E-06
Rank 5	7.38E-07	5.83E-07	9.18E-07	4.34E-07	7.45E-07	5.21E-07	6.56E-07
Rank 6	8.05E-08	3.60E-07	--	1.15E-07	5.65E-07	1.20E-07	1.53E-07
Rank 7	3.79E-08	2.60E-07	--	1.66E-07	3.88E-07	2.23E-07	1.78E-07

[1] HT: height; BMR: basal metabolic rate; hBMD: heel bone mineral density T-score; FVC: forced vital capacity; FEV: forced expiratory volume in 1s; BMI: body mass index

Supplementary Table 14. Prediction performance of the IFAM model with two sets of functional annotations in the WTCCC1 dataset. We tested the predictive accuracy of the IFAM model using functional annotation information from both the website of the linkage disequilibrium score regression (LDSC) model and the RegulomeDB database; the former included 74 types of functional annotations, and the latter included seven types. The accuracy of genomic prediction was assessed by the area under the curve (AUC) statistic using 20-fold cross-validation. The mean AUC values and standard errors of each trait are shown.

Trait[1]	IFAM_LDSC	IFAM_RegulomeDB	GBLUP	Adaptive MultiBLUP	BayesR	BayesRC
BD	0.6463(0.0030)	0.6465(0.0029)	0.6474(0.0029)	0.6671(0.0027)	0.6503(0.0037)	0.5261(0.0043)
CAD	0.6045(0.0030)	0.6021(0.0026)	0.5845(0.0028)	0.6075(0.0105)	0.5952(0.0027)	0.5178(0.0042)
CD	0.6758(0.0039)	0.6756(0.0035)	0.6322(0.0032)	0.6179(0.0119)	0.6705(0.0026)	0.5371(0.0047)
HT	0.5951(0.0031)	0.5945(0.0030)	0.5947(0.0031)	0.6142(0.0040)	0.5927(0.0033)	0.5160(0.0031)
RA	0.7215(0.0031)	0.7214(0.0031)	0.6113(0.0032)	0.6043(0.0138)	0.7137(0.0034)	0.5291(0.0041)
T1D	0.8699(0.0024)	0.8699(0.0024)	0.6542(0.0037)	0.6490(0.0256)	0.8612(0.0020)	0.5510(0.0043)
T2D	0.6156(0.0028)	0.6143(0.0027)	0.5968(0.0024)	0.6137(0.0087)	0.6164(0.0034)	0.5256(0.0036)
Mean	0.6755(0.0365)	0.6749(0.0367)	0.6173(0.0104)	0.6248(0.0237)	0.6714(0.0356)	0.5290(0.0120)

[1] BD: bipolar disorder; CAD: coronary artery disease; CD: Crohn's disease; HT: hypertension; RA: rheumatoid arthritis; T1D: type 1 diabetes; T2D: type 2 diabetes

Supplementary Table 15. Difference on predictive accuracy for four other models relative to the IFAM model were calculated using the WTCCC1 dataset. We tested the predictive accuracy of the IFAM model using functional annotation information from the RegulomeDB database which have seven types. The "IFAM_Rank" refers to the analysis that used ranking scores as annotation information from the RegulomeDB database. The "IFAM_KAML" involved optimizing the construction of the genetic relationship matrices in the IFAM model using the strategy from the KAML model. The "IFAM_pseudo annotation" represents running the IFAM model using pseudo-annotation information, which was selected randomly from the genome based on the scale of annotations from the RegulomeDB database. The accuracy of genomic prediction was assessed by the area under the curve (AUC) statistic using 20-fold cross-validation. The mean AUC values and standard errors of each trait are shown.

Trait[1]	IFAM	IFAM_Rank	IFAM_KAML	GBLUP	IFAM_pseudo annotation
BD	0.6463(0.0030)	0.6462(0.0030)	0.6451(0.0033)	0.6474(0.0029)	0.6455(0.0033)
CAD	0.6045(0.0030)	0.6032(0.0025)	0.6025(0.0027)	0.5845(0.0028)	0.5836(0.0032)
CD	0.6758(0.0039)	0.6770(0.0037)	0.6709(0.0036)	0.6322(0.0032)	0.6305(0.0034)
HT	0.5951(0.0031)	0.5935(0.0029)	0.5914(0.0031)	0.5947(0.0031)	0.5930(0.0032)
RA	0.7215(0.0031)	0.7212(0.0031)	0.7246(0.0030)	0.6113(0.0032)	0.6079(0.0041)
T1D	0.8699(0.0024)	0.8668(0.0033)	0.8650(0.0023)	0.6542(0.0037)	0.6550(0.0033)
T2D	0.6156(0.0028)	0.6132(0.0029)	0.6174(0.0028)	0.5968(0.0024)	0.5948(0.0024)
Mean	0.6755(0.0365)	0.6743(0.0363)	0.6738(0.0362)	0.6173(0.0104)	0.6158(0.0106)

[1] BD: bipolar disorder; CAD: coronary artery disease; CD: Crohn's disease; HT: hypertension; RA: rheumatoid arthritis; T1D: type 1 diabetes; T2D: type 2 diabetes

Supplementary Table 16. Prediction performance of the IFAM model on the UKB dataset. The functional information used by IFAM model came from the RegulomeDB database, which included seven types. The accuracy of genomic prediction was assessed by the correlation between estimated additive genetic values and phenotypic values using 10-fold cross-validation. The mean correlation values and standard errors of each trait are shown.

Trait[1]	IFAM	GBLUP
HT	0.5943(0.0012)	0.5629(0.0008)
BMR	0.4064(0.0007)	0.3930(0.0009)
hBMD	0.3904(0.0013)	0.3293(0.0012)
FVC	0.2983(0.0025)	0.2854(0.0023)
BMI	0.3448(0.0012)	0.3407(0.0010)
FEV	0.3029(0.0010)	0.2903(0.0011)
Mean	0.3895(0.0447)	0.3669(0.0423)

[1] HT: height; BMR: basal metabolic rate; hBMD: heel bone mineral density T-score; FVC: forced vital capacity; FEV: forced expiratory volume in 1s; BMI: body mass index

Supplementary Table 17. Prediction performance of four models using the pig dataset. The functional information used by IFAM model came from the IFmut database, which included 10 types of annotations. The "IFAM_KAML" represents running the IFAM model with a weighted genetic relationship matrix following the strategy in the KAML model. The "IFAM_pseudo annotation" represents running the IFAM model using pseudo-annotation information, which was selected randomly from the genome based on the scale of annotations from the IFmut database. The accuracy of genomic prediction was assessed by the correlation between estimated additive genetic values and phenotypic values using 20-fold cross-validation. The mean correlation values and standard errors of each trait are shown.

Trait[1]	IFAM	GBLUP	IFAM_KAML	IFAM_pseudo annotation
BF	0.4259(0.0079)	0.3966(0.0084)	0.4361(0.0082)	0.3887(0.0081)
LMD	0.3762(0.0052)	0.3745(0.0050)	0.3741(0.0053)	0.3749(0.0048)
ELMP	0.4482(0.0076)	0.4380(0.0075)	0.4526(0.0075)	0.4365(0.0078)
LTN	0.3393(0.0068)	0.3200(0.0070)	0.3402(0.0077)	0.3141(0.0081)
RTN	0.2921(0.0067)	0.2634(0.0077)	0.2912(0.0070)	0.2575(0.0080)
TTN	0.4173(0.0070)	0.3891(0.0072)	0.4197(0.0073)	0.3885(0.0073)
TPD	0.4663(0.0068)	0.4618(0.0068)	0.4727(0.0072)	0.4555(0.0069)
Mean	0.3950(0.0236)	0.3776(0.0256)	0.3981(0.0248)	0.3737(0.0259)

[1] BF: backfat thickness; LMD: loin muscle depth; ELMP: estimated lean meat percentage; LTN: left teat number; RTN: right teat number; TTN: total teat number; TPD: time spent eating per day

Supplementary Table 18. The number and proportion of SNPs with distinct functional annotations for 7 traits in the WTCCC1 dataset. The annotation process utilized seven rankings of functional annotations from RegulomeDB database. The numbers outside the parentheses represent the number of annotated SNPs, and the numbers inside the parentheses represent the proportion of annotated SNPs across the genome.

Type of annotation	Bipolar disorder (BD)	Coronary artery disease (CAD)	Crohn's disease (CD)	Hypertension (HT)	Rheumatoid arthritis (RA)	Type 1 diabetes (T1D)	Type 2 diabetes (T2D)
Rank 1	22184(5.94%)	22170(5.95%)	22202(5.93%)	22145(5.93%)	22148(5.94%)	22001(5.90%)	22086(5.92%)
Rank 2	10611(2.84%)	10569(2.84%)	10673(2.85%)	10627(2.85%)	10595(2.84%)	10616(2.85%)	10603(2.84%)
Rank 3	23652(6.33%)	23532(6.32%)	23772(6.35%)	23570(6.31%)	23640(6.34%)	23627(6.33%)	23571(6.32%)
Rank 4	61545(16.48%)	61315(16.46%)	61738(16.50%)	61684(16.52%)	61643(16.52%)	61557(16.50%)	61645(16.52%)
Rank 5	138812(37.18%)	138502(37.18%)	139008(37.16%)	138802(37.18%)	138657(37.17%)	138758(37.20%)	138827(37.20%)
Rank 6	41345(11.07%)	41299(11.09%)	41439(11.08%)	41292(11.06%)	41292(11.07%)	41303(11.07%)	41310(11.07%)
Rank 7	75220(20.15%)	75154(20.17%)	75281(20.12%)	75218(20.15%)	75081(20.13%)	75102(20.14%)	75107(20.13%)

Supplementary Table 19. Examples of the importance assessment of functional annotation information in the WTCCC1 dataset. The seven ranking scores from the RegulomeDB database were viewed as the functional annotation information. The average information restricted maximum likelihood (AI-REML) algorithm was chosen to estimate the heritability.

Category	BD[1]	CAD	CD	HT	RA	T1D	T2D	Mean
Rank 1	1.36E-05	1.21E-05	2.09E-05	1.27E-05	2.33E-05	3.81E-05	1.22E-05	1.90E-05
Rank 2	1.01E-05	4.22E-06	9.58E-06	5.54E-06	1.03E-05	4.41E-06	4.03E-06	6.89E-06
Rank 3	1.80E-06	6.66E-06	3.89E-06	2.20E-06	4.55E-06	4.46E-06	3.54E-06	3.87E-06
Rank 4	2.52E-06	3.55E-06	6.54E-06	2.80E-06	4.45E-06	4.11E-06	7.36E-06	4.48E-06
Rank 5	5.66E-06	2.92E-06	2.80E-06	1.47E-06	--	9.05E-07	1.59E-06	2.19E-06
Rank 6	1.83E-06	7.43E-07	--	2.06E-06	5.43E-06	--	--	1.44E-06
Rank 7	1.92E-06	--	3.28E-07	2.32E-06	3.40E-06	3.64E-06	1.21E-06	1.83E-06

[1] BD: bipolar disorder; CAD: coronary artery disease; CD: Crohn's disease; HT: hypertension; RA: rheumatoid arthritis; T1D: type 1 diabetes; T2D: type 2 diabetes

Supplementary Table 20. The average prediction accuracy for seven disease traits in the WTCCC1 dataset. The annotations are the ranking scores from the RegulomeDB database. The analysis was carried out by the Genomic Best Linear Unbiased Prediction (GBLUP) model on the seven disease traits using the same number of SNPs that was selected randomly in the seven annotations. This process of random sampling was repeated 100 times. The mean correlation values and standard errors of each trait are shown.

Category	BD[1]	CAD	CD	HT	RA	T1D	T2D
Rank 1	0.5735(0.0005)	0.5459(0.0005)	0.5955(0.0005)	0.5559(0.0005)	0.6267(0.0009)	0.7089(0.0013)	0.5554(0.0006)
Rank 2	0.5800(0.0005)	0.5417(0.0005)	0.5875(0.0004)	0.5449(0.0005)	0.5448(0.0005)	0.5456(0.0006)	0.5468(0.0005)
Rank 3	0.5634(0.0008)	0.5458(0.0008)	0.5624(0.0009)	0.5372(0.0010)	0.5317(0.0007)	0.5509(0.0010)	0.5432(0.0010)
Rank 4	0.5614(0.0009)	0.5403(0.0010)	0.5726(0.0012)	0.5375(0.0010)	0.5453(0.0010)	0.5607(0.0011)	0.5590(0.0010)
Rank 5	0.5675(0.0011)	0.5309(0.0010)	0.5555(0.0009)	0.5367(0.0011)	0.5234(0.0011)	0.5343(0.0012)	0.5375(0.0010)
Rank 6	0.5675(0.0010)	0.5263(0.0008)	0.5469(0.0011)	0.5402(0.0009)	0.5631(0.0023)	0.5525(0.0014)	0.5333(0.0009)
Rank 7	0.5697(0.0010)	0.5307(0.0009)	0.5557(0.0009)	0.5419(0.0010)	0.5399(0.0015)	0.5664(0.0020)	0.5343(0.0010)

[1] BD: bipolar disorder; CAD: coronary artery disease; CD: Crohn's disease; HT: hypertension; RA: rheumatoid arthritis; T1D: type 1 diabetes; T2D: type 2 diabetes

Supplementary Table 21. Comparisons of computational cost on time for different numbers of random effects within the IFAM model based on the "HE+PCG" strategy. The test was conducted using the heel bone mineral density T-score (hBMD) trait from the UK Biobank HM3 (UKB HM3) dataset, which involved around 200,000 individuals. Computational times were recorded for the estimation of the variance components using the Haseman–Elston (HE) regression algorithm and the additive genetics values using the pre-conditioned conjugate gradient (PCG) iterative algorithm. All results were computed on a Linux server with 48 CPUs and 1,440 GB memory.

Number of random effects	Time (h)
1	0.9480
2	1.1000
3	1.2505
4	1.3479
5	1.4457

Supplementary Table 22. The significant SNPs for seven disease traits in the WTCCC1 dataset. The general linear model was used to carry out GWAS analysis. The numbers before the slash (/) represent the significant SNPs that exceed the threshold line, and the numbers after the slash represent all markers that are located within a 50kb interval upstream and downstream of the significant SNPs.

Trait[1] / Repetition	BD	CAD	CD	HT	RA	T1D	T2D
1	--/--	13/46	35/175	--/--	107/561	168/853	9/30
2	--/--	15/47	38/223	--/--	127/648	168/827	10/32
3	--/--	13/46	32/171	1/23	114/741	176/811	12/33
4	--/--	9/43	25/132	--/--	120/538	166/834	13/69
5	--/--	12/46	33/215	--/--	121/636	169/820	10/32
6	--/--	10/46	25/158	--/--	104/577	161/814	10/32
7	--/--	13/46	27/167	--/--	116/606	176/803	7/100
8	--/--	18/47	38/199	--/--	101/448	160/869	10/32
9	--/--	7/41	20/133	--/--	115/603	167/834	5/23
10	--/--	15/47	26/154	--/--	100/524	143/726	12/33
11	--/--	15/47	34/171	--/--	104/590	182/839	10/32
12	--/--	12/46	47/194	--/--	114/487	156/812	15/94
13	1/1	14/46	27/207	--/--	123/604	161/809	9/32
14	--/--	11/46	27/154	--/--	126/674	169/817	10/32
15	--/--	14/46	22/204	--/--	115/584	169/786	8/31
16	--/--	12/137	36/181	--/--	100/509	165/858	10/32
17	--/--	7/41	28/158	--/--	118/596	164/862	10/32
18	--/--	10/45	25/154	--/--	123/715	176/818	10/32
19	--/--	8/43	36/273	--/--	115/557	173/803	9/32
20	--/--	13/46	25/130	--/--	101/524	170/895	10/32

[1] BD: bipolar disorder; CAD: coronary artery disease; CD: Crohn's disease; HT: hypertension; RA: rheumatoid arthritis; T1D: type 1 diabetes; T2D: type 2 diabetes

Supplementary Table 23. The significant SNPs for six traits in the UKB dataset. The linear regression model was used to carry out GWAS analysis.

Repetition \ Trait[1]	HT	BMR	hBMD	FVC	BMI	FEV
1	18,302	5,787	3,488	2,376	3,287	2,883
2	18,240	5,759	3,515	2,621	3,186	3,094
3	18,583	5,528	3,488	2,545	3,122	2,864
4	18,201	5,481	3,393	2,722	3,198	2,938
5	18,234	5,749	3,619	2,496	3,122	2,904
6	18,261	5,734	3,516	2,481	3,302	2,835
7	18,250	5,738	3,540	2,515	3,238	2,740
8	18,116	5,523	3,470	2,587	3,127	2,834
9	18,151	5,803	3,536	2,533	3,042	2,747
10	18,459	5,578	3,486	2,791	3,217	2,877

[1] HT: height; BMR: basal metabolic rate; hBMD: heel bone mineral density T-score; FVC: forced vital capacity; FEV: forced expiratory volume in 1s; BMI: body mass index

Supplementary Table 24. The significant SNPs for seven agricultural economic traits in the pig dataset. The mixed linear model was used to carry out GWAS analysis. The numbers before the slash (/) represent the significant SNPs that exceed the threshold line, and the numbers after the slash represent all markers that are located within a 50kb interval upstream and downstream of the significant SNPs.

Trait[1] / Repetition	BF	LMD	ELMP	LTN	RTN	TTN	TPD
1	198/4,918	--/--	80/4,819	257/833	500/1,361	925/4,327	11/994
2	193/4,918	--/--	109/3,309	259/895	520/1,814	635/2,602	9/992
3	193/4,918	--/--	78/3,329	255/833	518/2,468	671/2,602	4/821
4	195/5,264	--/--	46/4,205	678/2,625	498/1,361	699/2,821	8/982
5	201/4,918	--/--	61/3,215	556/1,976	502/1,756	627/2,531	11/994
6	201/4,918	--/--	96/4,409	505/1,631	510/2,283	689/2,602	5/907
7	196/5,080	--/--	17/1,548	503/1,423	515/1,973	686/2,603	4/821
8	214/5,267	--/--	84/4,557	256/833	516/2,044	588/2,602	7/981
9	198/5,264	--/--	122/4,463	256/833	499/1,361	730/3,359	4/757
10	195/4,918	--/--	26/2,310	499/1,361	509/2,283	614/2,560	7/981
11	197/4,921	--/--	103/4,458	549/1,759	511/2,342	663/2,602	1/603
12	208/6,662	--/--	57/2,970	268/1,236	513/1,973	683/2,604	6/823
13	188/4,918	--/--	41/3,126	560/1,759	503/1,432	669/2,606	2/603
14	198/4,918	--/--	139/4,493	663/2,004	511/1,759	646/2,067	5/894
15	196/4,918	--/--	12/1,953	256/833	516/2,556	679/2,604	8/918
16	194/4,918	--/--	25/1,867	256/833	510/1,759	689/2,602	14/996
17	199/4,918	--/--	47/2,219	528/1,641	511/1,830	709/2,670	9/992
18	204/5,267	--/--	81/4,378	656/1,780	566/2,568	687/2,611	1/603
19	215/5,267	--/--	108/4,798	620/1,779	511/1,759	515/2,028	15/996
20	202/5,264	--/--	65/3,227	510/1,665	537/3,219	690/2,604	--/--

[1] BF: backfat thickness; LMD: loin muscle depth; ELMP: estimated lean meat percentage; LTN: left teat number; RTN: right teat number; TTN: total teat number; TPD: time spent eating per day

References

1. Goddard ME, Hayes BJ. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat Rev Genet* **10**, 381-391 (2009).

2. Georges M, Charlier C, Hayes B. Harnessing genomic information for livestock improvement. *Nature Reviews Genetics* **20**, 135-156 (2019).

3. Piepho H-P. Ridge regression and extensions for genomewide selection in maize. *Crop Science* **49**, 1165-1176 (2009).

4. Henderson CR. Best linear unbiased estimation and prediction under a selection model. *Biometrics* **31**, 423-447 (1975).

5. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci* **91**, 4414-4423 (2008).

6. Moser G, Lee SH, Hayes BJ, Goddard ME, Wray NR, Visscher PM. Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model. *Plos Genetics* **11**, e1004969 (2015).

7. MacLeod IM*, et al.* Exploiting biological priors and sequence variants enhances QTL discovery and genomic prediction of complex traits. *BMC Genomics* **17**, 144 (2016).

8. Mollandin F, Gilbert H, Croiseau P, Rau A. Accounting for overlapping annotations in genomic prediction models of complex traits. *BMC Bioinformatics* **23**, 365 (2022).

9. Yin L*, et al.* KAML: improving genomic prediction accuracy of complex traits using machine learning determined parameters. *Genome Biology* **21**, 146 (2020).

10. Finucane HK*, et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics* **47**, 1228-1235 (2015).