

**The American Journal of Human Genetics, Volume 99**

**Supplemental Data**

**Contrasting the Genetic Architecture  
of 30 Complex Traits from Summary Association Data**

**Huwenbo Shi, Gleb Kichaev, and Bogdan Pasaniuc**

# Decomposition of $\hat{\beta}^T \mathbf{V}^\dagger \hat{\beta}$ under in-sample and external reference LD

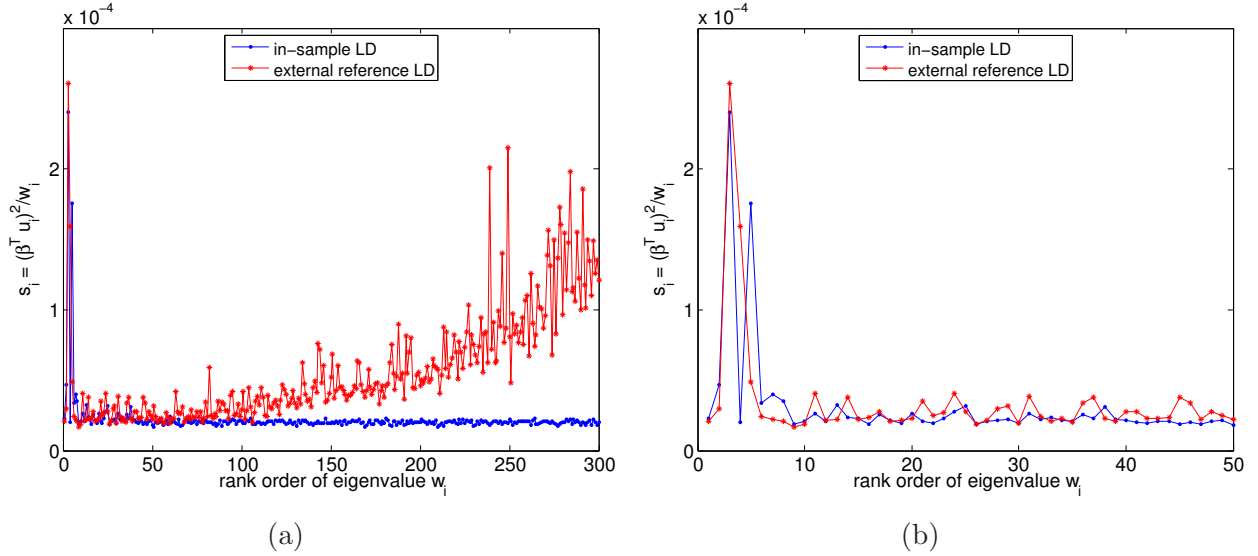


Figure S1:  $s_i = (\hat{\beta}^T \mathbf{u}_i)^2 / w_i$  as a function of the rank order of eigenvalue  $w_i$  obtained under in-sample LD (blue, rank=974) and external reference LD (red, rank=251) for a locus containing 1,377 SNPs. Each point represents the mean of  $s_i$  over 500 simulations. Figure S1a displays the first 300  $s_i$ . Figure S1b focuses on the first 50  $s_i$ .

## Comparison between HESS, LDSC, and Gamazon et al.'s method in simulations

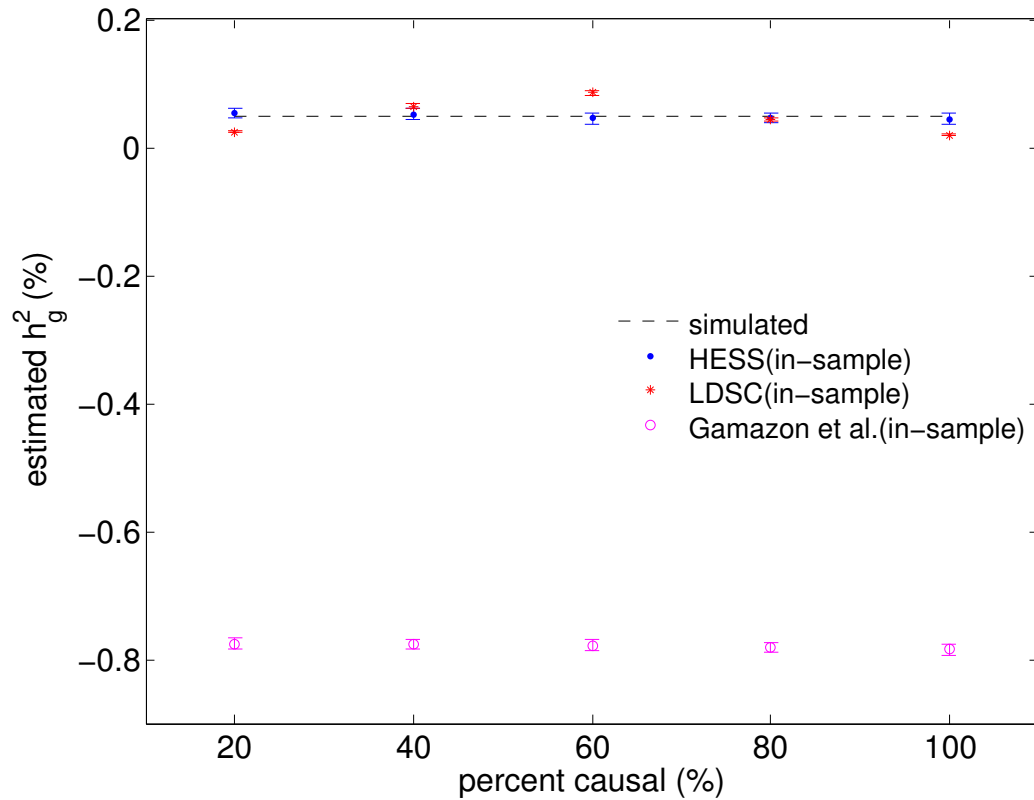


Figure S2: Gamazon et al.'s method severely underestimates simulated heritability due to overcorrection of bias by  $\frac{p}{n}$ . Summary statistics were computed based on simulated GWAS involving 50,000 individuals. Mean and standard errors were computed based on 500 simulations. Error bars represent 2 times the standard error.

# Comparison between HESS and LDSC in simulations

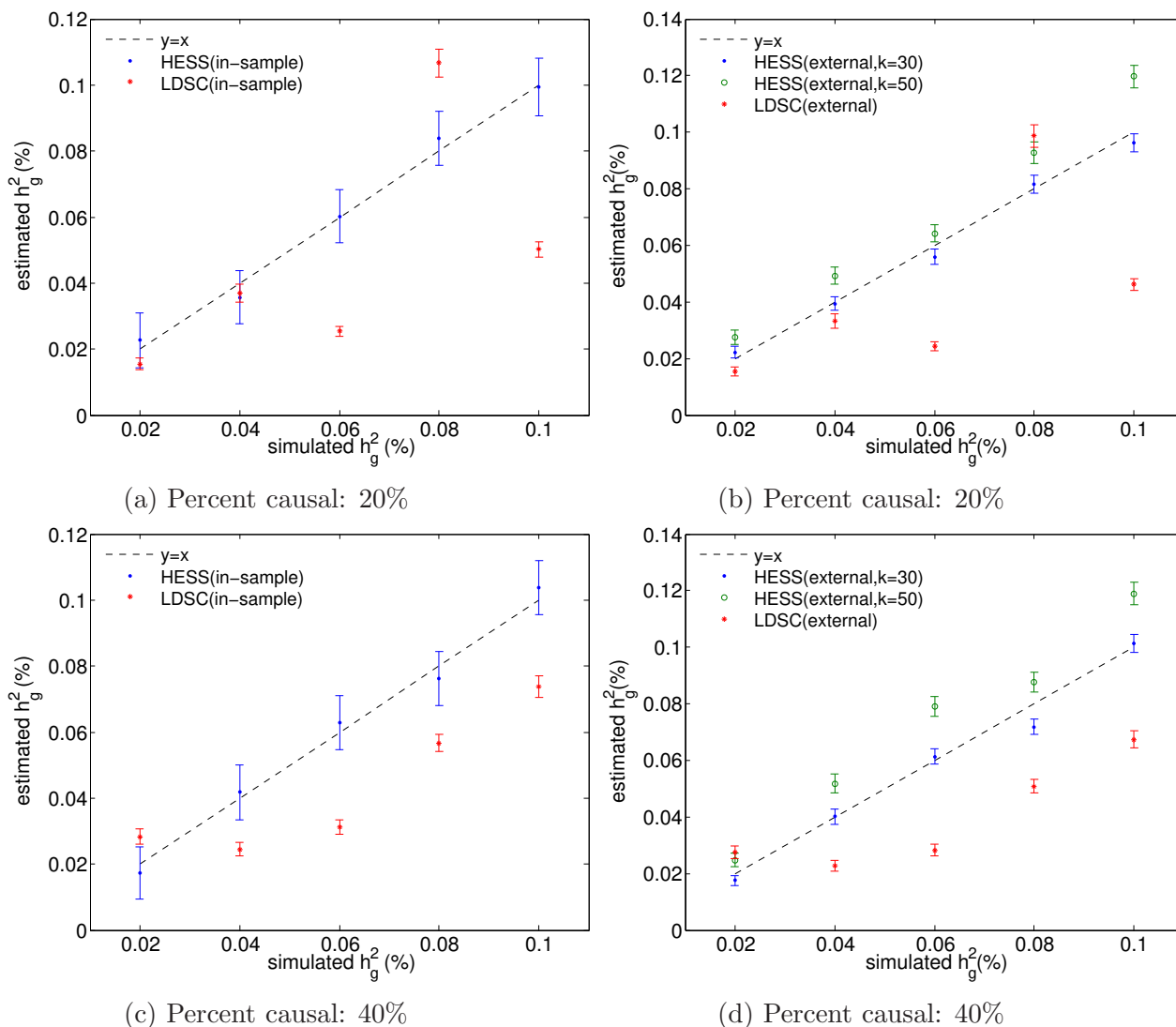
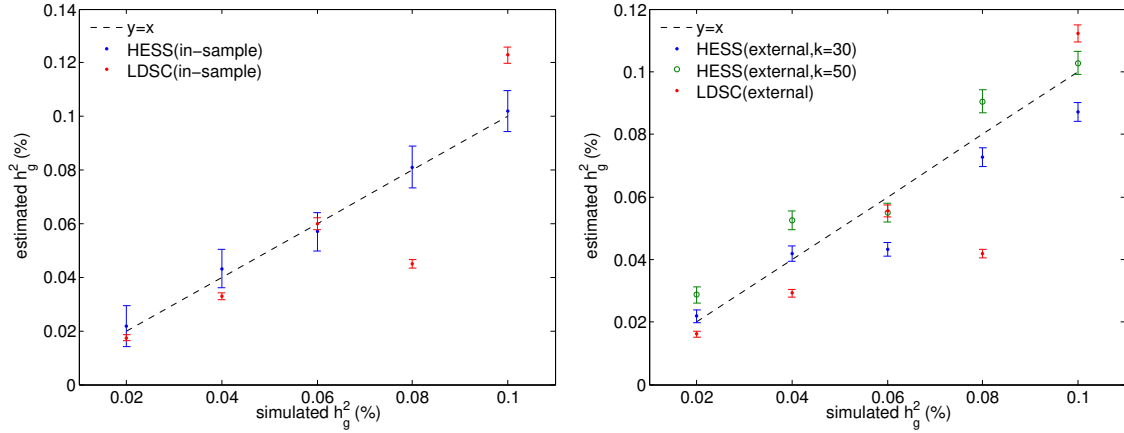
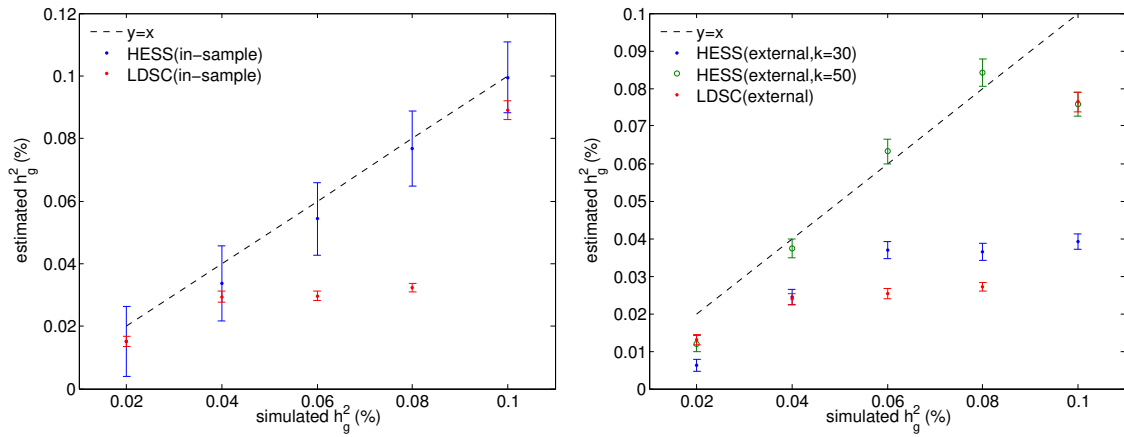


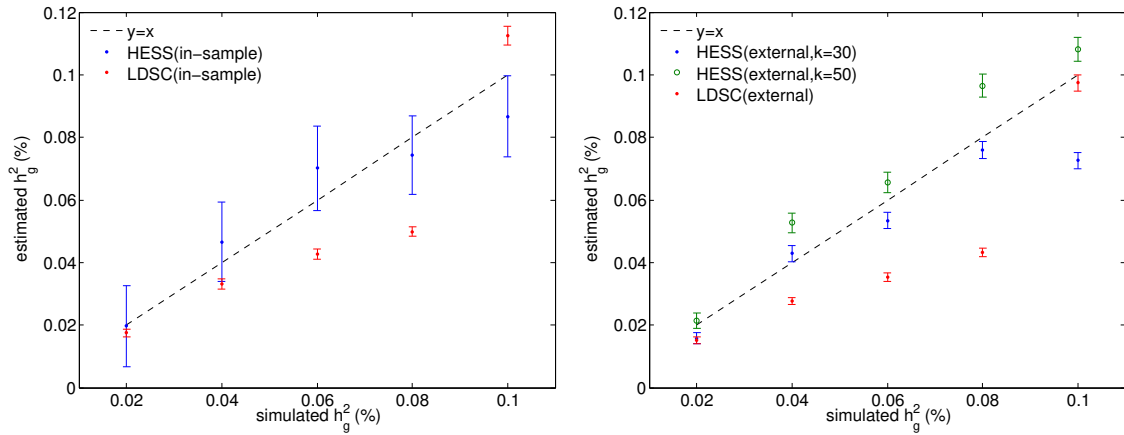
Figure S3: Comparison between HESS and LDSC at different percentages of causal SNPs and simulated heritability at a 1Mb locus (1,377 SNPs). Figures on the left show results obtained using in-sample LD. Figures on the right show results obtained using external reference LD. Mean and standard errors were computed based on 500 simulations. Error bars represent 2 times the standard error.



(a) Simulation results at a 0.75 Mb locus (809 SNPs)

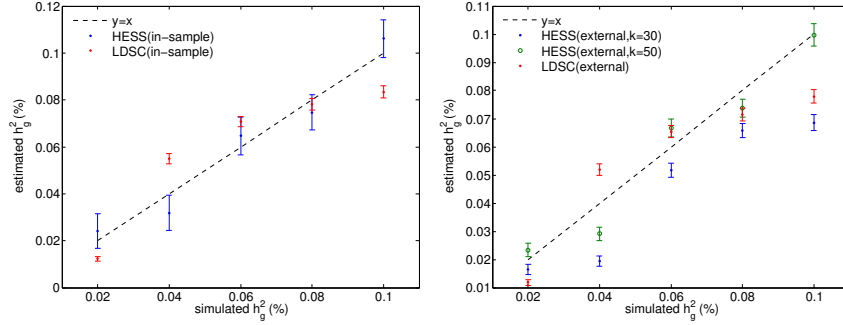


(b) Simulation results at a 1.25 Mb locus (1,980 SNPs)

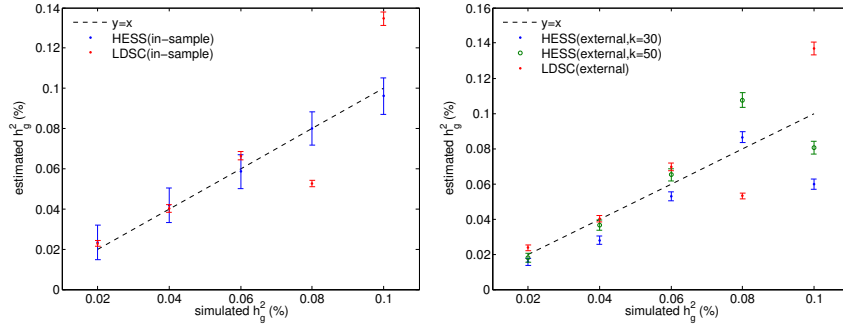


(c) Simulation results at a 1.5 Mb locus (2,350 SNPs)

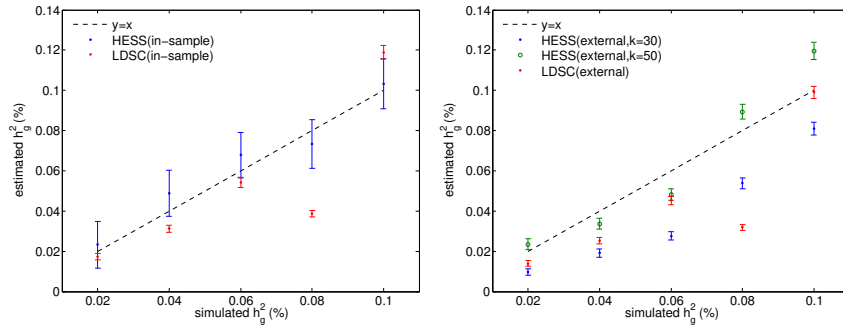
Figure S4: Comparison between HESS and LDSC under fixed-effect model at different simulated heritability at difference loci. In all simulations, 20% of the SNPs in each locus were randomly selected to be causal. Figures on the left show results obtained using in-sample LD. Figures on the right show results obtained using external reference LD. Mean and standard errors were computed based on 500 simulations. Error bars represent 2 times the standard error.



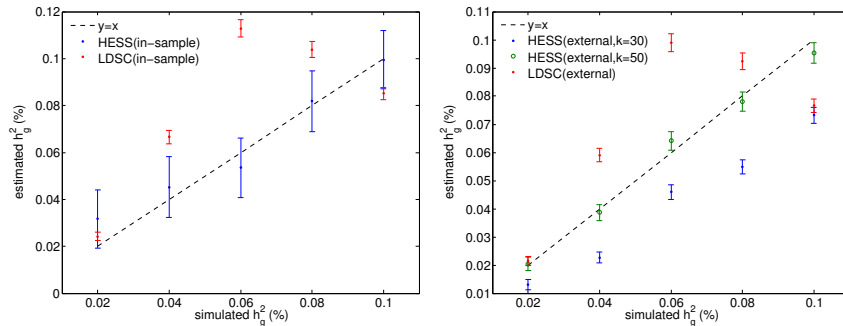
(a) Simulation results at a 0.75 Mb locus (809 SNPs)



(b) Simulation results at a 1 Mb locus (1,583 SNPs)



(c) Simulation results at a 1.25 Mb locus (1,980 SNPs)



(d) Simulation results at a 1.5 Mb locus (2,350 SNPs)

Figure S5: Comparison between HESS and LDSC under random-effect model at different simulated heritability at difference loci. In all simulations, 20% of the SNPs in each locus were randomly selected to be causal. Figures on the left show results obtained using in-sample LD. Figures on the right show results obtained using external reference LD. Mean and standard errors were computed based on 500 simulations. Error bars represent 2 times the standard error.

# Evaluating the variance estimates for $h_{g,local}^2$ of HESS

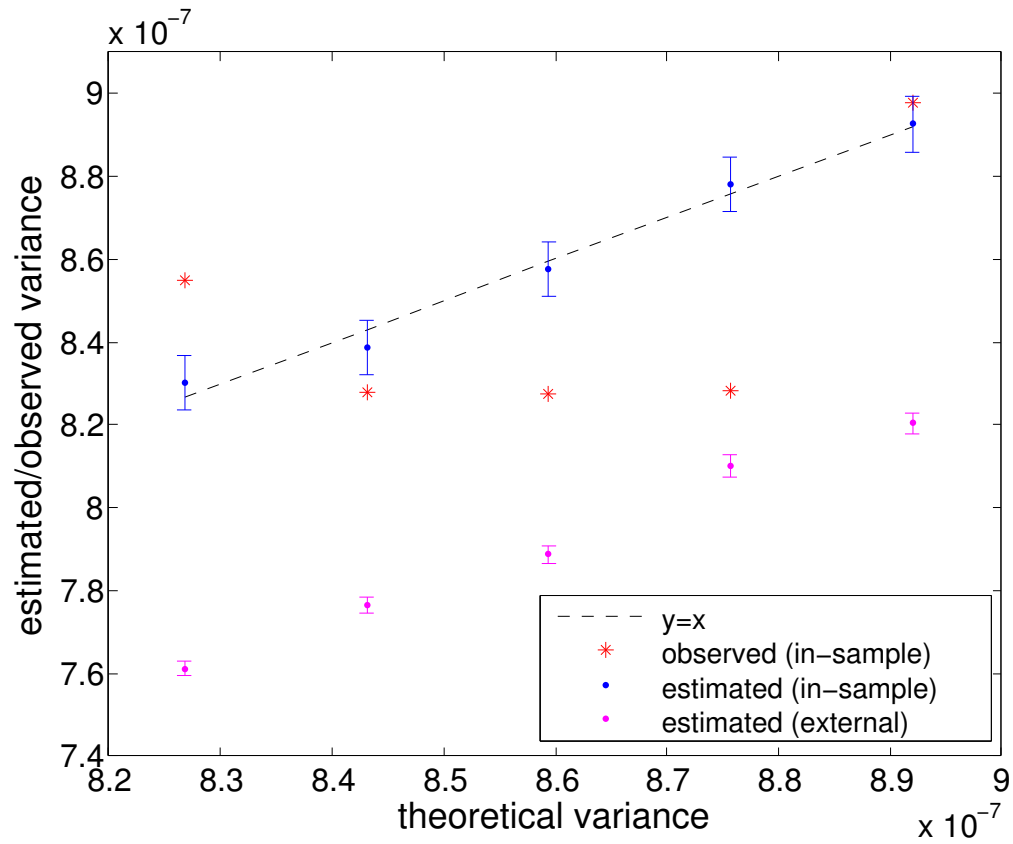


Figure S6: Estimates for  $\text{Var}[\hat{h}_g^2]$  based on in-sample and external reference LD. Mean and standard errors were computed based on 500 simulations. Error bars represent 2 times the standard error.

## Total SNP-heritability obtained by HESS and LDSC

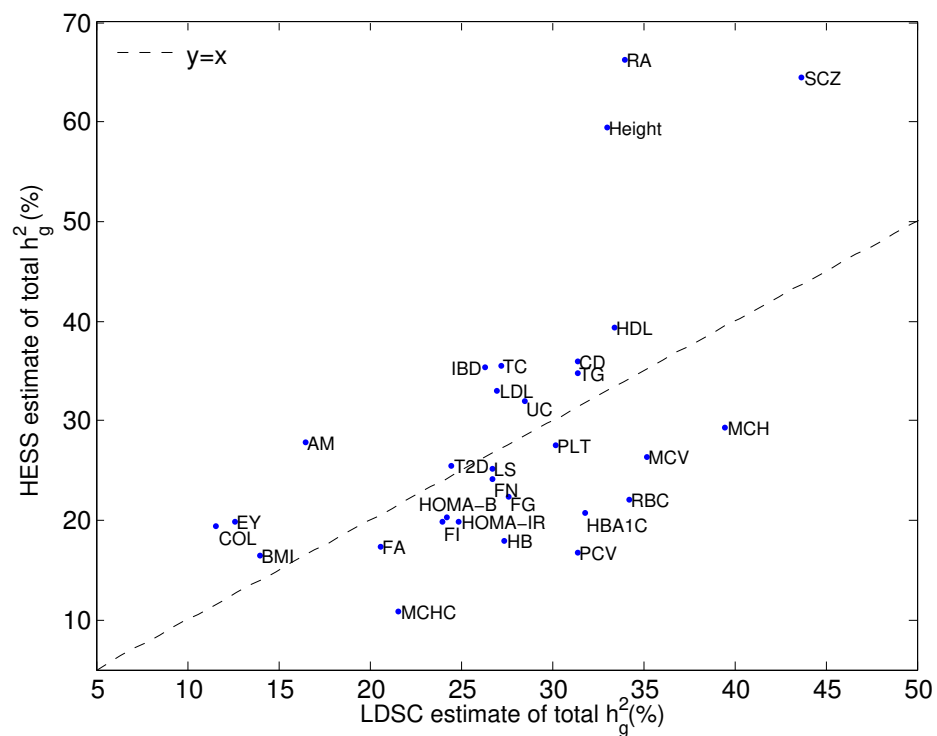


Figure S7: Estimates of total SNP-heritability obtained by HESS versus estimates obtained by LDSC. We see a strong correlation ( $R = 0.58$ ) between HESS and LDSC estimates.



# Comparing local SNP-heritability estimates obtained with or without known total $h_g^2$

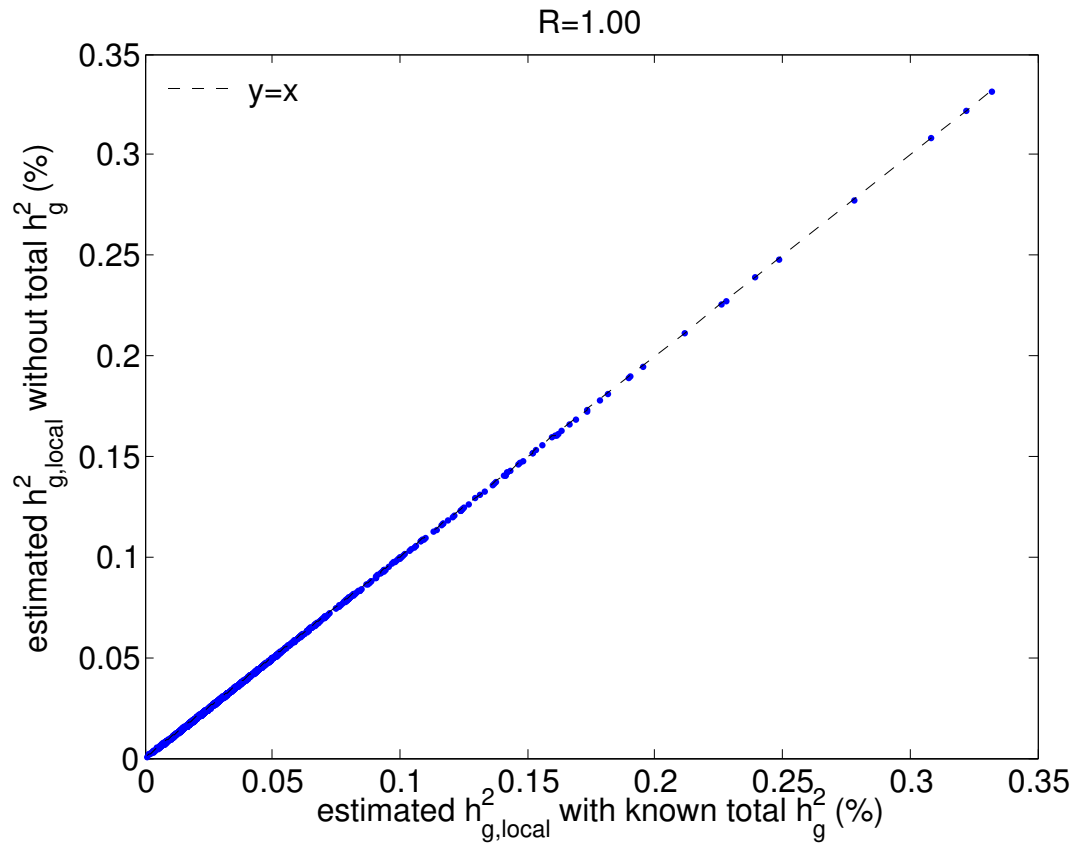


Figure S8: Local SNP-heritability estimates for height obtained without known total  $h_g^2$  versus estimates obtained using known total  $h_g^2$ .

# Percentage of total SNP-heritability attributable to different fractions of genome

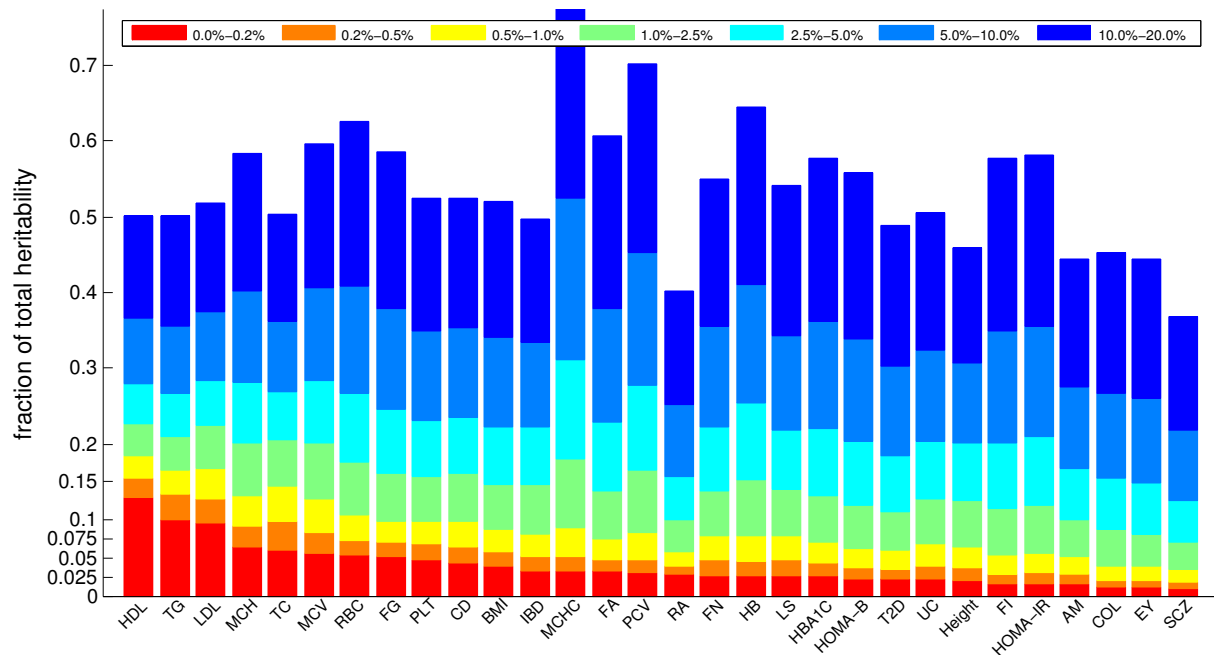


Figure S9: Stacked bar plot showing the percentage of total SNP-heritability (after removing local SNP-heritability estimates at loci overlapping the HLA region) attributable to different fractions of genome. We rank ordered all loci by their explained SNP-heritability and quantified the fraction of total SNP-heritability attributable to different percentile ranges. Traits with high polygenicity tend to have bars with height proportional to bin size, whereas less polygenic traits tend to have bars much larger than bin size.

# Correlations between traits across $h^2_{g,local}$ estimates



Figure S10: Correlation heatmap showing the correlations between all 30 studied traits across local SNP-heritability estimates. We group traits based on their corresponding GWAS consortia.

# Average local SNP-heritability across 30 traits at loci displaying pleiotropic effect

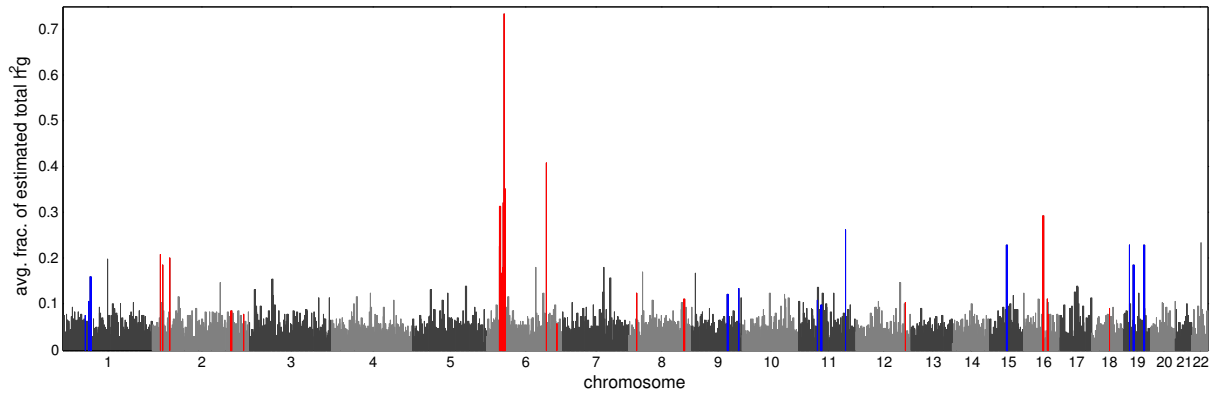


Figure S11: Manhattan-style plot displaying average local SNP-heritability across 30 traits. We mark the 36 selected loci displaying pleiotropic in colors.

## Cumulative fraction of total SNP-heritability versus fraction of genome

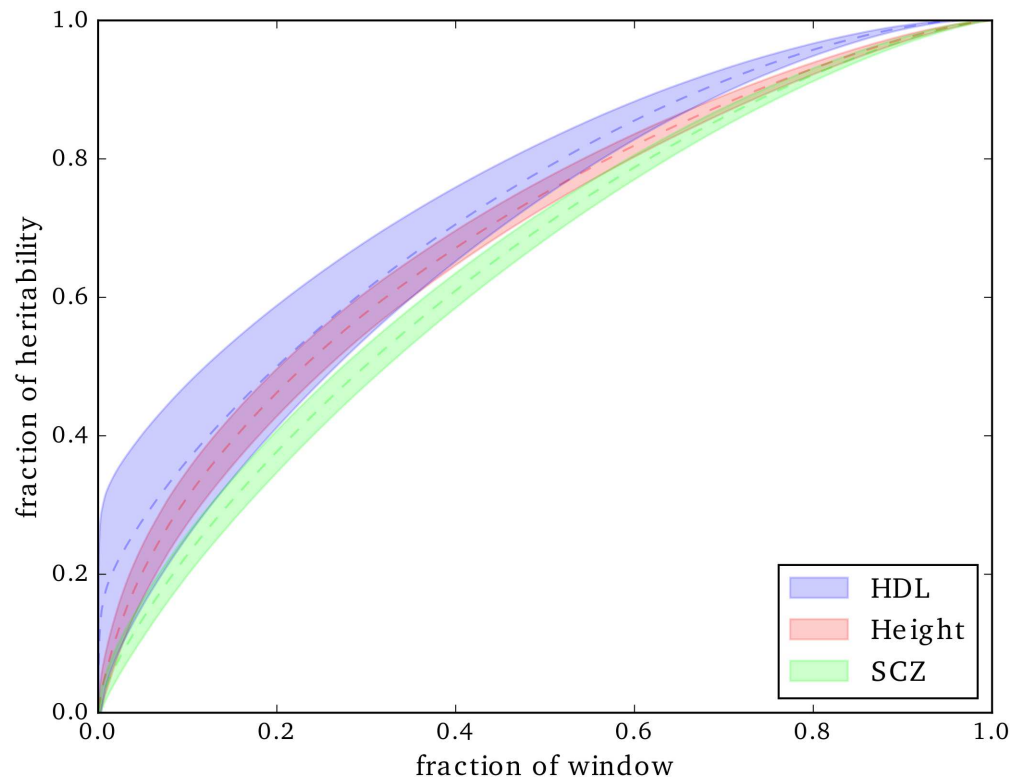


Figure S12: Dotted curves show the cumulative fraction of total SNP-heritability covered by loci with highest  $h_{g,local}^2$  versus fraction of genome covered by these loci. Shaded region shows 2 times the standard error obtained by jack-knifing over the loci.

# Chromosomal SNP-heritability versus chromosome size

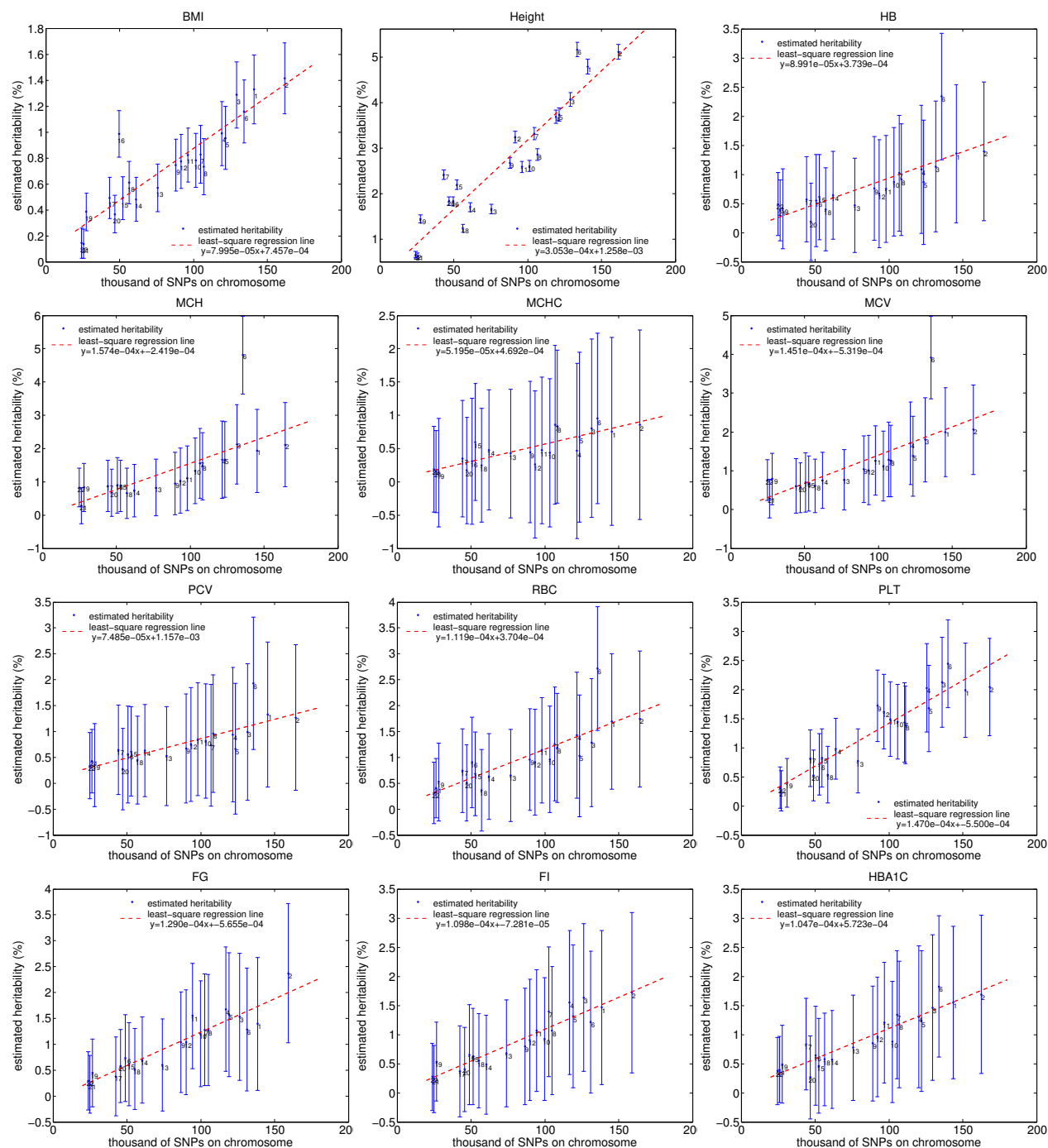


Figure S13: Chromosomal SNP-heritability estimates versus chromosome size. We compute chromosomal SNP-heritability by summing over local SNP-heritability within the chromosome. We obtain standard error estimates by summing variance estimates of local SNP-heritability and then take the square root. Error bars represent 2 times the standard error. High standard errors are likely due to relative small sample size of the GWAS.

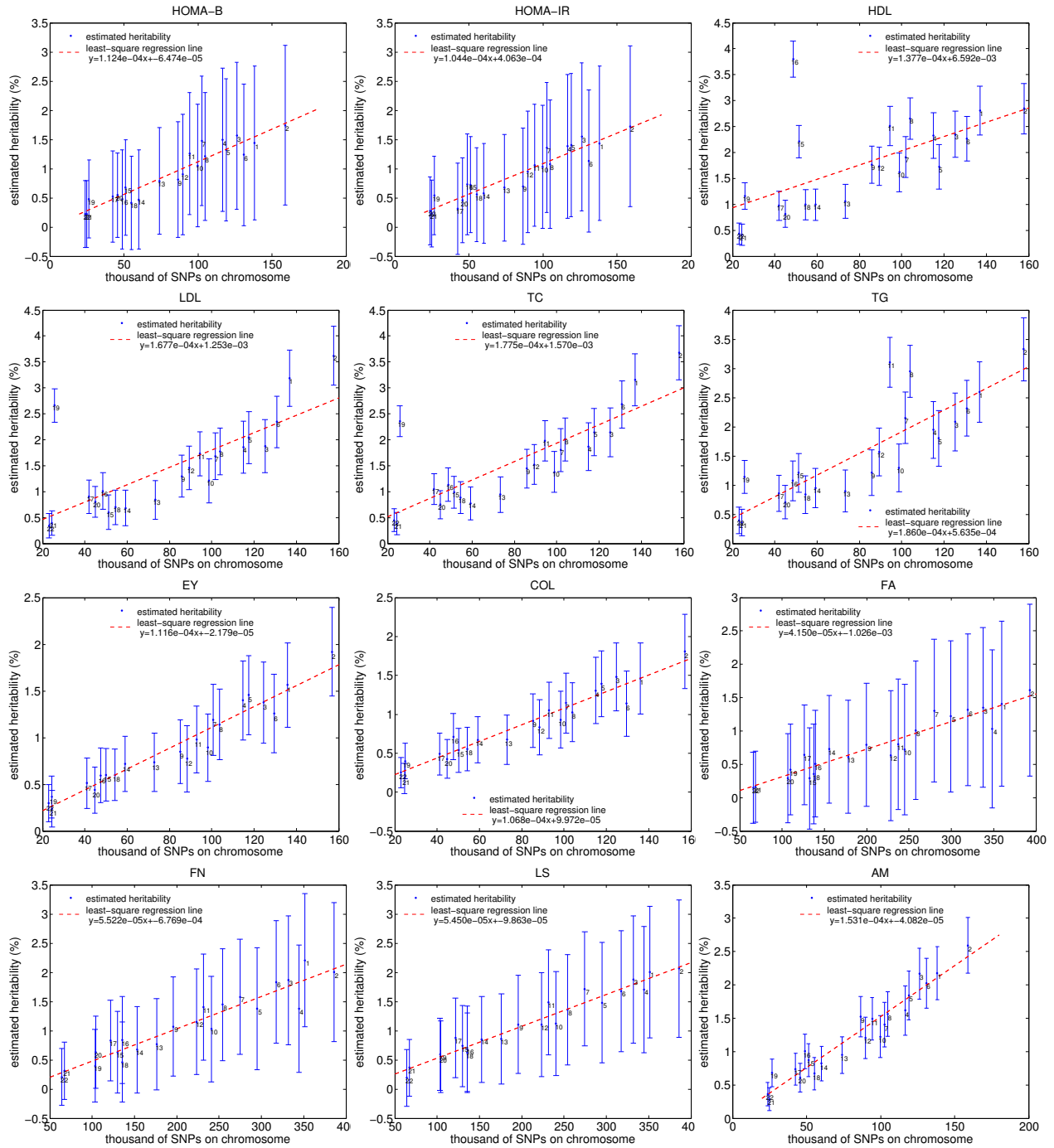


Figure S14: Chromosomal SNP-heritability estimates versus chromosome size (continued).

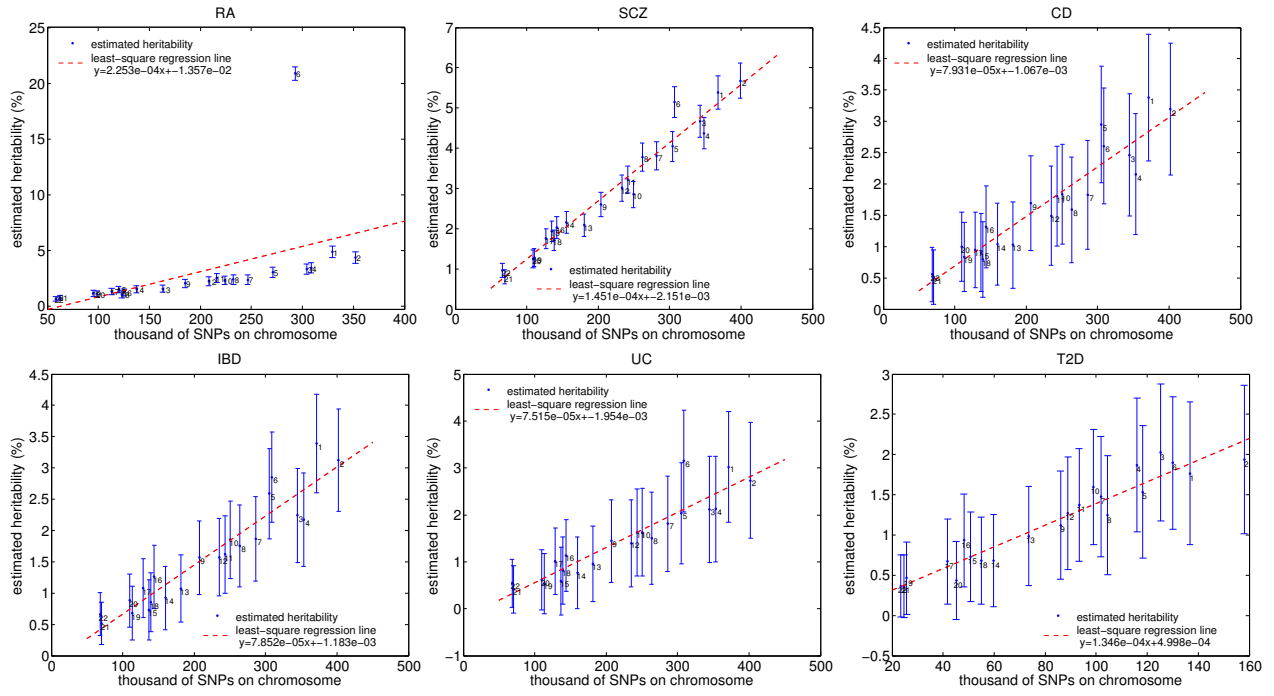


Figure S15: Chromosomal SNP-heritability estimates versus chromosome size (continued).



# Manhattan-style plots of $h_{g,local}^2$ estimates for 30 traits

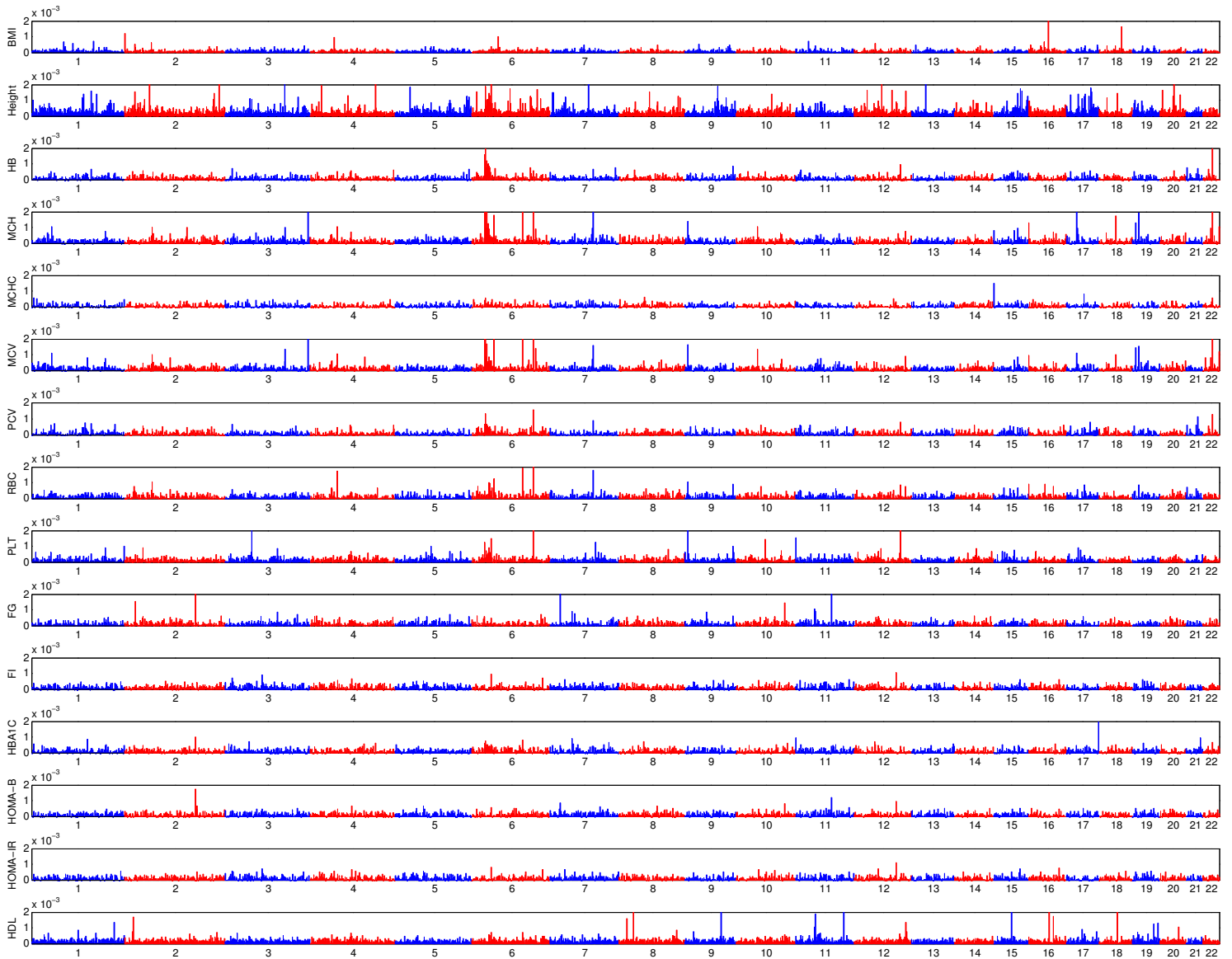


Figure S16: Manhattan-style plots of local SNP-heritability across the genome. We group traits based on their corresponding GWAS consortia.

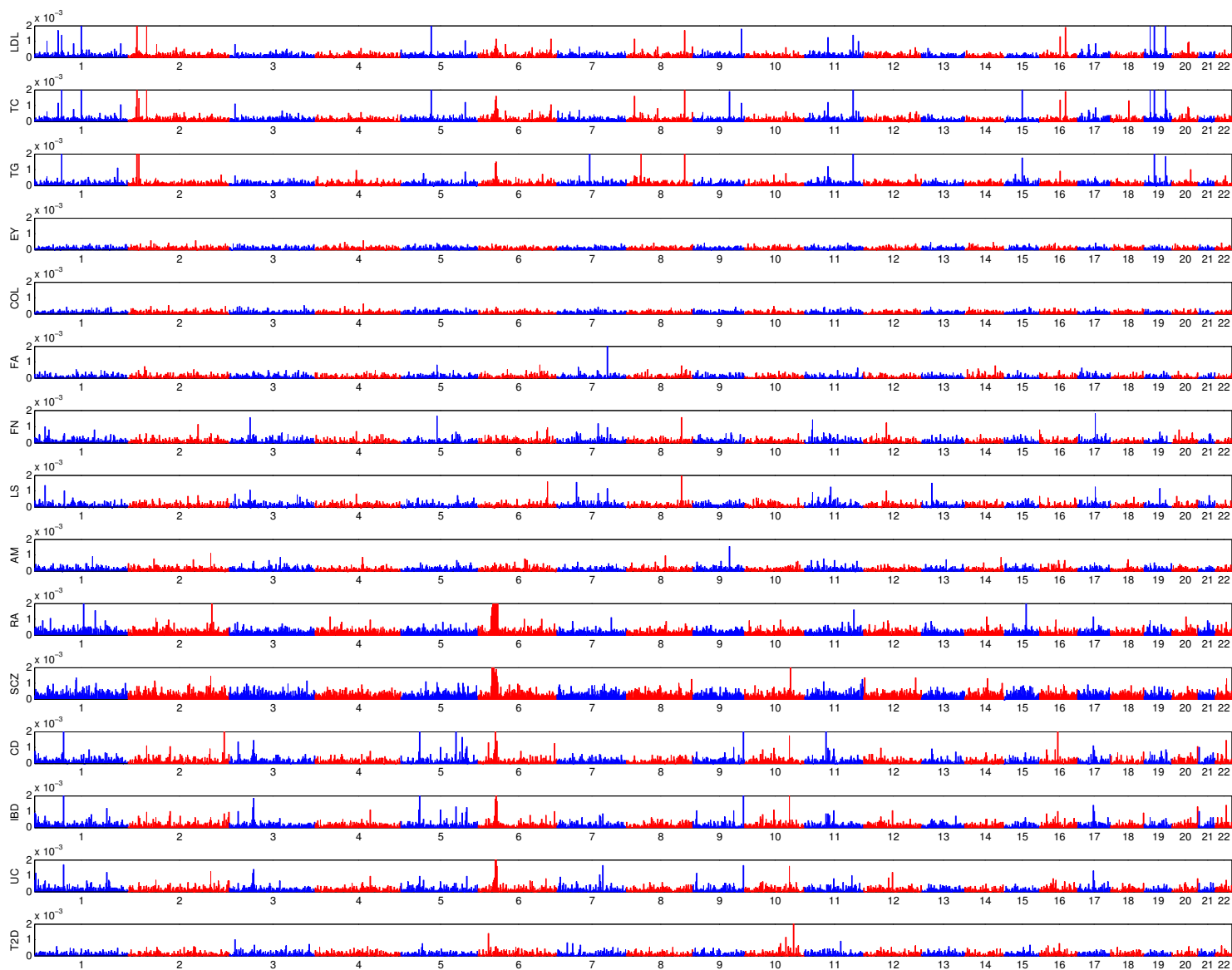


Figure S17: Manhattan-style plots of local SNP-heritability across the genome (continued).

# Distribution of $h^2_{g,local}$ at loci with/without GWAS hits

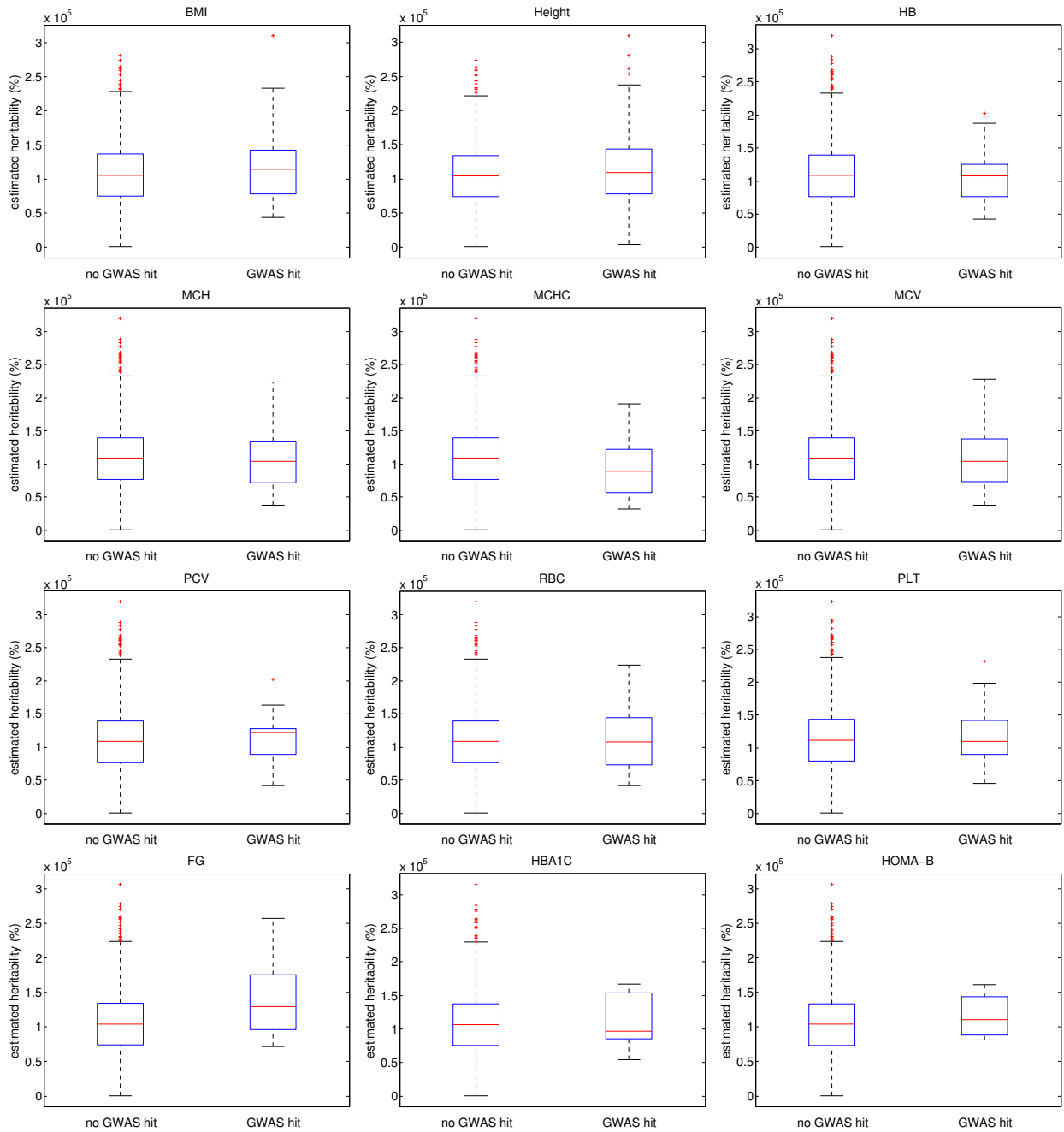


Figure S18: Box plots of  $h^2_{g,local}$  at loci with and without GWAS hits.

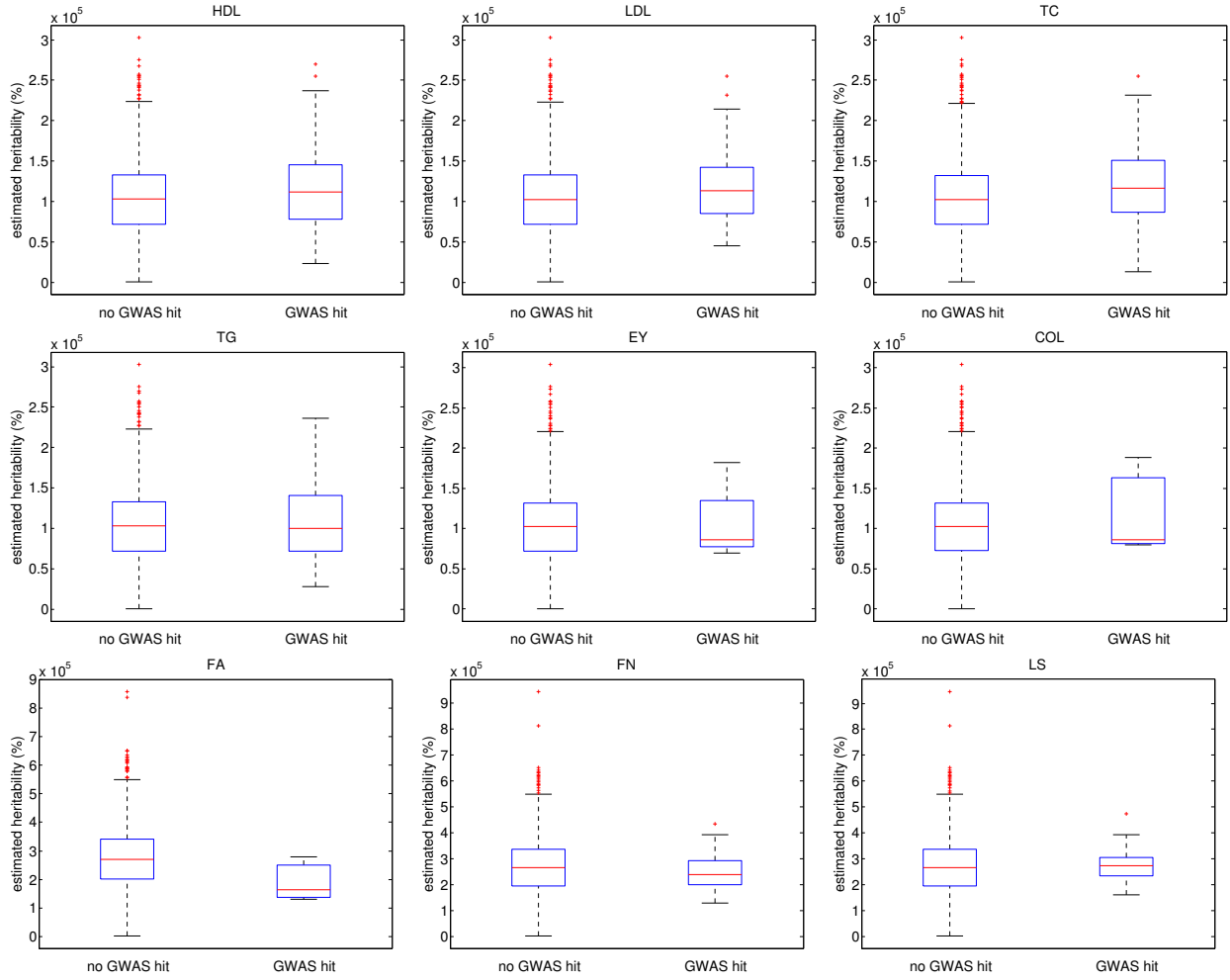


Figure S19: Box plots of  $h_{g,local}^2$  at loci with and without GWAS hits.

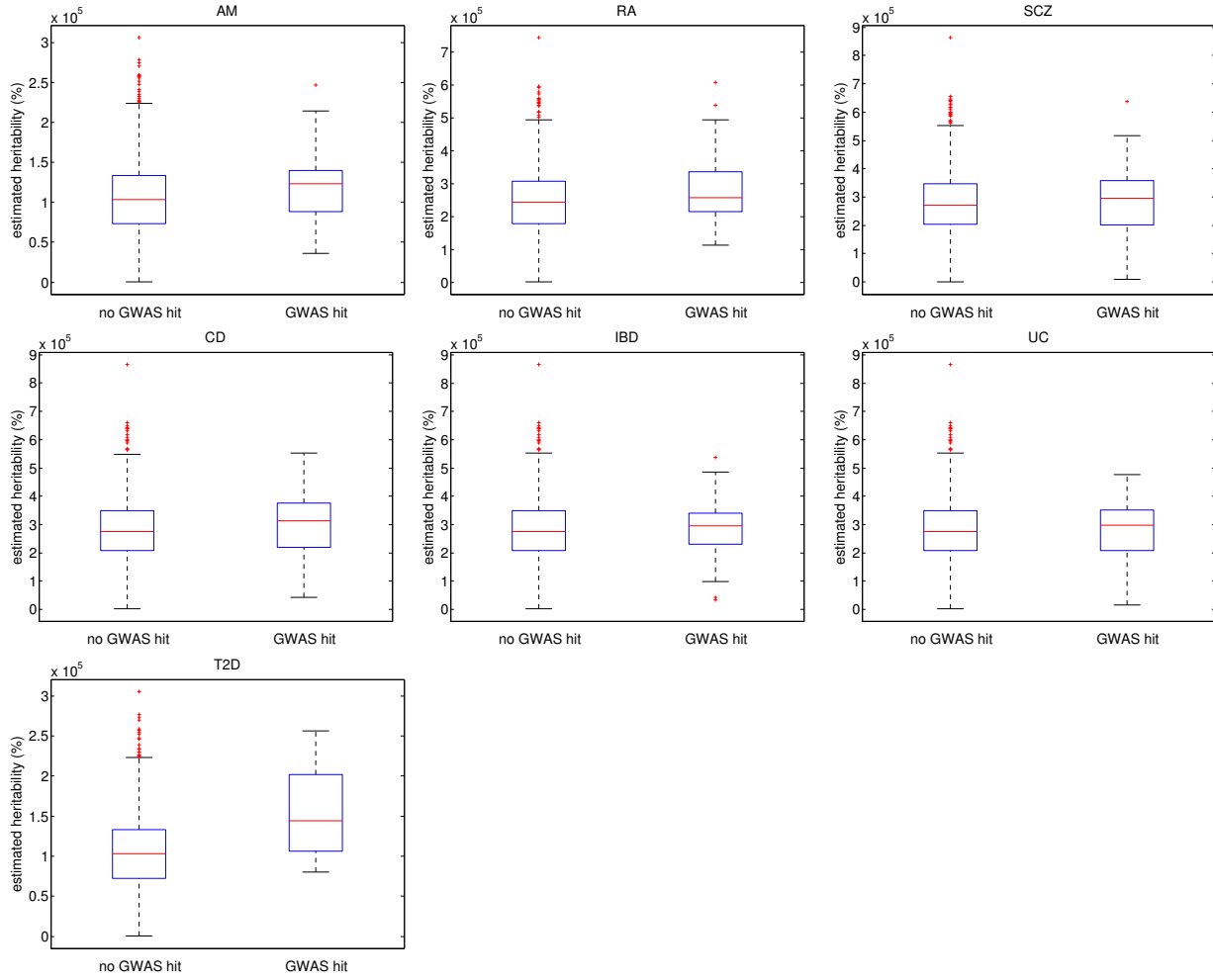


Figure S20: Box plots of  $h_{g,local}^2$  at loci with and without GWAS hits.

# Details of summary GWAS data for the 30 traits

Trait	Sample size	No. SNPs	No. GWAS hits	No. index SNPs	Fraction <sup>a</sup>
BMI (Body Mass Index) <sup>1</sup>	229269	1859666	1851	79	5.31
Height (Height) <sup>2</sup>	244015	1854761	26374	476	31.15
HB (Haemoglobin) <sup>3</sup>	52666	1894024	459	24	1.38
MCH (Mean Cell Haemoglobin) <sup>3</sup>	44658	1892019	1585	37	2.25
MCHC (MCH Concentration) <sup>3</sup>	48252	1893281	223	15	0.9
MCV (Mean Cell Volume) <sup>3</sup>	49808	1893769	1602	46	3.08
PCV (Packed Cell Volume) <sup>3</sup>	46169	1893412	288	14	0.92
RBC (Red Blood Cell Count) <sup>3</sup>	46465	1892553	1132	31	2.1
PLT (Number of Platelets) <sup>4</sup>	66867	1954590	954	40	2.54
FG (Fasting Glucose) <sup>5</sup>	46186	1824182	290	12	0.97
FI (Fasting Insulin) <sup>5</sup>	46186	1822388	–	–	–
HBA1C (HBA1C) <sup>6</sup>	46368	1870395	187	11	0.6
HOMA-B (HOMA-B) <sup>5</sup>	46186	1820938	119	4	0.24
HOMA-IR (HOMA-IR) <sup>5</sup>	46186	1821061	–	–	–
HDL (High Density Lipoprotein) <sup>7</sup>	96335	1805617	3445	92	6.28
LDL (Low Density Lipoprotein) <sup>7</sup>	91529	1803637	2971	76	4.87
TC (Total Cholesterol) <sup>7</sup>	96596	1805676	4039	91	5.98
TG (Triglycerides) <sup>7</sup>	92768	1803908	3149	91	3.95
EY (Education Years) <sup>8</sup>	126559	1788888	11	4	0.25
FA (Forearm BMD) <sup>9</sup>	53236	4725343	152	3	0.18
FN (Femoral Neck BMD) <sup>9</sup>	53236	4637340	867	21	1.21
LS (Lumbar Spine) <sup>9</sup>	53236	4636561	1077	24	1.39
AM (Age at Menarche) <sup>10</sup>	132989	1821879	2391	73	4.61
COL (College) <sup>8</sup>	126559	1792881	61	3	0.2
RA (Rheumatoid Arthritis) <sup>11</sup>	14361/43923	4265540	19575	51	3.06
SCZ (Schizophrenia) <sup>12</sup>	32405/42221	4772186	8113	103	6.9
CD (Crohn's Disease) <sup>13</sup>	17897/33977	4822932	5179	54	3.48
IBD <sup>b</sup> (Inflammatory Bowel Disease) <sup>13</sup>	13769/33977	4823603	9243	70	4.17
UC (Ulcerative Colitis) <sup>13</sup>	31666/33977	4823578	5114	42	2.45
T2D (Type 2 Diabetes) <sup>14</sup>	12171/56862	1806359	236	13	1.0

Table S1: Details of the summary GWSA data for the 30 analyzed traits. <sup>a</sup>Fraction refers to the fraction of genome with GWAS hits. <sup>b</sup>IBD refers to the union of CD and UC. For case-control traits, we list sample size as No. cases / No. controls.

# Total SNP-heritability estimation obtained by HESS and LDSC

Trait	$h_g^2$ (HESS)	k	Estimated $\lambda_{gc}$	$h_{g,local,gwas}^{2\ddagger}$	Enrichment <sup>a</sup>	$h_g^2$ (LDSC)
BMI (Body Mass Index) <sup>1</sup>	16.5(0.5)	50	1.33	2.45(0.11)	3.22(0.27)	14.0(0.9)
Height (Height) <sup>2</sup>	59.4(0.3)	50	1.00	23.86(0.20)	1.73(0.05)	33.0(1.7)
HB (Haemoglobin) <sup>3</sup>	17.9(2.1)	16	1.29	1.40(0.28)	6.19(1.38)	27.4(1.4)
MCH (Mean Cell Haemoglobin) <sup>3</sup>	29.3(2.2)	14	1.32	3.16(0.39)	6.71(1.28)	39.5(2.6)
MCHC (MCH Concentration) <sup>3</sup>	10.9(2.5)	15	1.30	0.40(0.25)	5.41(1.70)	21.6(0.9)
MCV (Mean Cell Volume) <sup>3</sup>	26.3(2.0)	15	1.31	3.08(0.39)	5.66(0.91)	35.2(2.1)
PCV (Packed Cell Volume) <sup>3</sup>	16.7(2.5)	14	1.31	0.64(0.25)	4.71(1.26)	31.4(1.5)
RBC (Red Blood Cell Count) <sup>3</sup>	22.0(2.3)	14	1.32	1.61(0.35)	4.48(0.82)	34.2(1.7)
PLT (Number of Platelets) <sup>4</sup>	27.5(1.5)	20	1.26	2.41(0.25)	4.04(0.44)	30.2(1.4)
FG (Fasting Glucose) <sup>5</sup>	22.3(2.3)	14	1.20	0.66(0.21)	3.58(1.11)	27.6(1.6)
FI (Fasting Insulin) <sup>5</sup>	19.9(2.4)	14	1.19	0.10(0.06)	15.41(0.00)	24.0(1.0)
HBA1C (HBA1C) <sup>6</sup>	20.8(2.3)	14	1.24	0.69(0.20)	5.31(1.89)	31.8(1.2)
HOMA-B (HOMA-B) <sup>5</sup>	20.3(2.4)	14	1.19	0.06(0.12)	1.26(0.68)	24.2(1.1)
HOMA-IR (HOMA-IR) <sup>5</sup>	19.9(2.4)	14	1.20	0.11(0.06)	16.17(0.00)	24.9(1.1)
HDL (High Density Lipoprotein) <sup>7</sup>	39.4(0.9)	29	1.13	4.33(0.24)	2.78(0.26)	33.4(7.5)
LDL (Low Density Lipoprotein) <sup>7</sup>	33.0(1.0)	27	1.16	3.97(0.24)	3.34(0.33)	27.0(4.5)
TC (Total Cholesterol) <sup>7</sup>	35.5(0.9)	29	1.16	5.27(0.25)	3.19(0.28)	27.2(3.8)
TG (Triglycerides) <sup>7</sup>	34.8(0.9)	28	1.18	3.76(0.21)	3.69(0.47)	31.4(5.2)
EY (Education Years) <sup>8</sup>	19.9(0.8)	38	1.05	0.15(0.04)	3.20(1.45)	12.6(0.5)
FA (Forearm BMD) <sup>9</sup>	17.4(2.2)	16	1.18	0.19(0.10)	9.90(4.33)	20.6(0.9)
FN (Femoral Neck BMD) <sup>9</sup>	24.1(2.1)	16	1.17	1.43(0.25)	5.39(0.81)	26.7(1.2)
LS (Lumbar Spine) <sup>9</sup>	25.1(2.0)	16	1.17	1.61(0.26)	4.70(0.61)	26.7(1.1)
AM (Age at Menarche) <sup>10</sup>	27.8(0.7)	40	1.05	3.18(0.17)	2.60(0.16)	16.5(0.7)
COL (College) <sup>8</sup>	19.4(0.8)	38	1.08	0.13(0.04)	3.34(0.98)	11.6(0.5)
RA (Rheumatoid Arthritis) <sup>11</sup>	66.3(0.9)	18	1.20	5.98(0.32)	5.82(1.29)	34.0(8.7)
SCZ (Schizophrenia) <sup>12</sup>	64.5(0.7)	22	1.00	8.36(0.21)	2.20(0.15)	43.7(1.4)
CD (Crohn's Disease) <sup>13</sup>	35.9(1.8)	16	1.12	3.64(0.37)	3.47(0.43)	31.4(2.1)
IBD <sup>c</sup> (Inflammatory Bowel Disease) <sup>13</sup>	35.3(1.4)	20	1.09	4.45(0.33)	3.66(0.39)	26.3(1.5)
UC (Ulcerative Colitis) <sup>13</sup>	31.9(2.1)	15	1.11	2.96(0.36)	4.35(0.73)	28.5(1.3)
T2D (Type 2 Diabetes) <sup>14</sup>	25.4(1.6)	19	1.19	0.71(0.16)	2.63(0.49)	24.5(1.1)

Table S2: Total SNP-heritability for the 30 traits obtained by HESS and LDSC. To obtain LDSC estimate, we compute LD scores for all SNPs with MAF greater than 5% using the same reference panel as used by HESS. Since HESS does not account for population stratification, we obtain LDSC estimate without the intercept.  $h_{g,local,gwas}^{2\ddagger}$  refers to the estimated SNP-heritability attributable to loci containing GWAS hit after all GWAS hits are removed. <sup>a</sup>We define enrichment as the ratio between the fraction of  $h_g^2$  attributable to  $h_{g,local,gwas}^{2\ddagger}$  and the fraction of genome covered by these loci. We obtain standard errors by jack-knife over the loci.

## References

- [1] Adam E Locke, Bratati Kahali, Sonja I Berndt, Anne E Justice, Tune H Pers, Felix R Day, Corey Powell, Sailaja Vedantam, Martin L Buchkovich, Jian Yang, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*, 518(7538): 197–206, 2015.
- [2] Andrew R Wood, Tonu Esko, Jian Yang, Sailaja Vedantam, Tune H Pers, Stefan Gustafsson, Audrey Y Chu, Karol Estrada, Jian'an Luan, Zoltán Kutalik, et al. Defin-

- ing the role of common variation in the genomic and biological architecture of adult human height. *Nature genetics*, 46(11):1173–1186, 2014.
- [3] Pim van der Harst, Weihua Zhang, Irene Mateo Leach, Augusto Rendon, Niek Verweij, Joban Sehmi, Dirk S Paul, Ulrich Elling, Hooman Allayee, Xinzhong Li, et al. Seventy-five genetic loci influencing the human red blood cell. *Nature*, 492(7429):369–375, 2012.
- [4] Christian Gieger, Aparna Radhakrishnan, Ana Cvejic, Weihong Tang, Eleonora Porcu, Giorgio Pistis, Jovana Serbanovic-Canic, Ulrich Elling, Alison H Goodall, Yann Labrune, et al. New gene functions in megakaryopoiesis and platelet formation. *Nature*, 480(7376):201–208, 2011.
- [5] José Dupuis, Claudia Langenberg, Inga Prokopenko, Richa Saxena, Nicole Soranzo, Anne U Jackson, Eleanor Wheeler, Nicole L Glazer, Nabila Bouatia-Naji, Anna L Gloyn, et al. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature genetics*, 42(2):105–116, 2010.
- [6] Nicole Soranzo, Serena Sanna, Eleanor Wheeler, Christian Gieger, Dörte Radke, José Dupuis, Nabila Bouatia-Naji, Claudia Langenberg, Inga Prokopenko, Elliot Stolerman, et al. Common variants at 10 genomic loci influence hemoglobin a1c levels via glycemic and nonglycemic pathways. *Diabetes*, 59(12):3229–3239, 2010.
- [7] Global Lipids Genetics Consortium et al. Discovery and refinement of loci associated with lipid levels. *Nature genetics*, 45(11):1274–1283, 2013.
- [8] Cornelius A Rietveld, Sarah E Medland, Jaime Derringer, Jian Yang, Tonu Esko, Nicolas W Martin, Harm-Jan Westra, Konstantin Shakhbazov, Abdel Abdellaoui, Arpana Agrawal, et al. Gwas of 126,559 individuals identifies genetic variants associated with educational attainment. *science*, 340(6139):1467–1471, 2013.
- [9] Hou-Feng Zheng, Vincenzo Forgetta, Yi-Hsiang Hsu, Karol Estrada, Alberto Rosello-Diez, Paul J Leo, Chitra L Dahia, Kyung Hyun Park-Min, Jonathan H Tobias, Charles Kooperberg, et al. Whole-genome sequencing identifies *en1* as a determinant of bone density and fracture. *Nature*, 526(7571):112–117, 2015.
- [10] John RB Perry, Felix Day, Cathy E Elks, Patrick Sulem, Deborah J Thompson, Teresa Ferreira, Chunyan He, Daniel I Chasman, Tõnu Esko, Gudmar Thorleifsson, et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature*, 514(7520):92–97, 2014.
- [11] Yukinori Okada, Di Wu, Gosia Trynka, Towfique Raj, Chikashi Terao, Katsunori Ikari, Yuta Kochi, Koichiro Ohmura, Akari Suzuki, Shinji Yoshida, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, 506(7488):376–381, 2014.
- [12] Schizophrenia Working Group of the Psychiatric Genomics Consortium et al. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, 511(7510):421–427, 2014.
- [13] Jimmy Z Liu, Suzanne van Sommeren, Hailiang Huang, Siew C Ng, Rudi Alberts, Atsushi Takahashi, Stephan Ripke, James C Lee, Luke Jostins, Tejas Shah, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature genetics*, 47(9):979–986, 2015.
- [14] Andrew P Morris, Benjamin F Voight, Tanya M Teslovich, Teresa Ferreira, Ayellet V Segre, Valgerdur Steinthorsdottir, Rona J Strawbridge, Hassan Khan, Harald Grallert, Anubha Mahajan, et al. Large-scale association analysis provides insights into the



genetic architecture and pathophysiology of type 2 diabetes. *Nature genetics*, 44(9): 981, 2012.