Similarity and diversity of genetic architecture for complex traits between East Asian and European populations

Jinhui Zhang[1#], Shuo Zhang[1#], Jiahao Qiao[1#], Ting Wang[1*], Ping Zeng[1-5*]

1. Department of Biostatistics, School of Public Health, Xuzhou Medical University, Xuzhou, Jiangsu, 221004, China

2. Center for Medical Statistics and Data Analysis, Xuzhou Medical University, Xuzhou, Jiangsu, 221004, China

3. Key Laboratory of Human Genetics and Environmental Medicine, Xuzhou Medical University, Xuzhou, Jiangsu, 221004, China

4. Key Laboratory of Environment and Health, Xuzhou Medical University, Xuzhou, Jiangsu, 221004, China

5. Engineering Research Innovation Center of Biological Data Mining and Healthcare Transformation, Xuzhou Medical University, Xuzhou, Jiangsu, 221004, China

Co-first authors

Correspondence to

Ting Wang, Department of Biostatistics, Xuzhou Medical University, Xuzhou, Jiangsu 221004, China. Tel: 86-516-8326 3736; Fax: 86-516-8326 3736; E-mail: 100002009025@xzhmu.edu.cn

Ping Zeng, Department of Biostatistics, Center for Medical Statistics and Data Analysis, Key Laboratory of Human Genetics and Environmental Medicine, Key Laboratory of Environment and Health, and Engineering Research Innovation Center of Biological Data Mining and Healthcare Transformation, Xuzhou Medical University, Xuzhou, Jiangsu 221004, China. Tel: 86-516-8326 3736; Fax: 86-516-8326 3736; E-mail: zpstat@xzhmu.edu.cn

29   **Abstract**

30   **Background:** Genome-wide association studies have detected a large number of

31   single-nucleotide polymorphisms (SNPs) associated with complex traits in diverse

32   ancestral groups. However, the trans-ethnic similarity and diversity of genetic

33   architecture is not well understood currently.

34   **Results:** By leveraging summary statistics of 37 traits from East Asian ($N_{\max}$=254,373)

35   or European ($N_{\max}$=693,529) populations, we first evaluated the trans-ethnic genetic

36   correlation (ρ_g) and found substantial evidence of shared genetic overlap underlying

37   these traits between the two populations, with $\hat{\rho}_g$ ranging from 0.53 (se=0.11) for

38   adult-onset asthma to 0.98 (se=0.17) for hemoglobin A1c. However, 88.9% of the

39   genetic correlation estimates were significantly less than one, indicating potential

40   heterogeneity in genetic effect across populations. We next identified common

41   associated SNPs using the conjunction conditional false discovery rate method and

42   observed 21.7% of trait-associated SNPs can be identified simultaneously in both

43   populations. Among these shared associated SNPs, 20.8% showed heterogeneous

44   influence on traits between the two ancestral populations. Moreover, we demonstrated

45   that population-common associated SNPs often exhibited more consistent linkage

46   disequilibrium and allele frequency pattern across ancestral groups compared to

47   population-specific or null ones. We also revealed population-specific associated

48   SNPs were much likely to undergo natural selection compared to population-common

49   associated SNPs.

50   **Conclusions:** Our study provides an in-depth understanding of similarity and

51   diversity regarding genetic architecture for complex traits across diverse populations,

52   and can assist in trans-ethnic association analysis, genetic risk prediction, and causal

53   variant fine mapping.

54   **Keywords:** Trans-ethnic analysis, genetic similarity and diversity, summary statistics,

55   genome-wide association study, conditional false discovery rate

Background

57 Over the past few years, large-scale genome-wide association studies (GWASs) have

58 convincingly detected a large number of genetic loci associated with a series of

59 complex traits by casting single-nucleotide polymorphisms (SNPs) across the entire

60 genome [1], generating novel biological knowledge and renovating diagnostic and

61 treatment tools for diseases [2, 3]. However, existing GWASs have been heavily

62 biased towards European (EUR) individuals, with 66.4% (=7,377/11,113, until 2023-

63 4-25) of studies conducted in EUR but only a few in other populations [4]. As medical

64 genomics studies have become increasingly large and diverse, acquiring insights into

65 similarity and diversity of trait-associated SNPs among distinct populations and

66 consequently the transferability of disease genetic risk is imperative in clinical

67 translation [5].

68 It has been revealed that many trait-associated SNPs identified in GWASs of EUR

69 ancestry can be replicated in other ethnic groups [2, 6-10], in the sense that they show

70 significant association with consistent direction of genetic effects in non-EUR

71 individuals, indicating complex traits enjoy common genetic components across

72 diverse continental populations. For example, the trans-ethnic genetic correlation is

73 0.79 (se=0.04) for ulcerative colitis and 0.76 (se=0.04) for Crohn's disease [11], 0.46

74 (se=0.06) for rheumatoid arthritis (RA) [12], 0.33 (se=0.03) for major depressive

75 disorder [13], and 0.39 (se=0.15) for attention-deficit hyperactivity disorder between

76 the EUR and East Asian (EAS) populations. However, these estimated trans-ethnic

77 genetic correlation are in general significantly less than one in spite of being larger

78 than zero [12, 14], which meanwhile suggests ancestral diversity.

79 Indeed, ancestry-relevant heterogeneity regarding varying allele frequency and

80 linkage disequilibrium (LD) patterns is observed for some causal variants such that a

81 significant association detected in one population is not readily identified in others [5,

82 15, 16]. Notable examples include a nonsense variant in *TBC1D4*, which confers

83 muscle insulin resistance and increases risk for type 2 diabetes (T2D) and is common

84 in Greenland but rare or absent in other populations [17], several common EAS-

85 specific coding variants that influence blood lipids by exerting a protective effect

86 against alcoholism [18, 19], and two loci associated with major depression that are

87   more common in the Chinese population than EUR (i.e., 45% vs. 2% for rs12415800,

88   and 28% vs. 6% for rs35936514) [20, 21]. As another example, multiple associated

89   loci located within *PCSK9*, *APOA*, *APOC*, and *ABCA1* all play key roles in lipid

90   genetics in both the EUR and African (AFR) Americans, yet the precise alleles within

91   each locus differ substantially between the two populations, supporting the

92   perspective that lipid-associated SNPs are largely shared across ancestral groups, but

93   the allelic structure within a locus may be shaped by population history and thus

94   exhibits considerable heterogeneity [22, 23].

95   Prior work has investigated the replicability of GWAS discoveries for some particular

96   traits, displaying the similarity and diversity of associated SNPs across ancestral

97   groups [7, 12, 24, 25]. However, those studies primarily focused on limited traits or a

98   small set of significant loci [7, 24]; it is unknown whether their conclusions can be

99   generalized to other traits or to genome-wide SNPs given the polygenic nature of

100   many complex phenotypes. In addition, some of previous studies focused mainly on

101   the trans-ethnic genetic correlation [12, 25, 26], which however only quantifies an

102   overall genetic similarity across the whole genome and cannot characterize detailed

103   association pattern for individual SNPs. A comprehensive genome-wide assessment of

104   trans-ethnic similarity and diversity of genetic components for traits is challenging

105   because our knowledge of genetic architecture within each population is not fully

106   understood.

107   To fill in this critical knowledge gap mentioned above, here we obtained GWAS

108   summary statistics of 37 complex traits from the EAS and EUR populations to

109   perform a complete comparison of genetic similarity and diversity across the two

110   populations. We first evaluated the trans-ethnic genetic correlation to quantify the

111   extent of common genetic basis to which these traits shared [12]. Then, we identified

112   population-specific and population-common trait-associated SNPs [27, 28].

113   Afterwards, we conducted the marginal genetic correlation analysis and heterogeneity

114   test for these associated SNPs [25]. Finally, we evaluated how the LD and minor

115   allele frequency (MAF) patterns varied among various types of SNPs and studied

116   whether the genetic differentiation between ancestry groups can be explained via

117   natural selection [7, 24]. We also assessed the genetic influence of associated SNPs on

118   traits across populations by calculating a genetic risk score (GRS) [24, 29].

Results

Overview of the used statistical methods

119 We here give a brief overview of some important statistical methods employed in our

122 analyses and showed more methodological descriptions in the Materials and Methods

123 Section. Technical details of all used methods could be found in respective original

124 papers. We here analyzed a total of 37 complex traits (10 binary and 27 continuous)

125 using summary statistics obtained from EAS-only or EUR-only GWASs (**Table 1**).

126 We first employed linkage disequilibrium score regression (LDSC) to estimate SNP-

127 based heritability for every trait in each population [30]. Next, to evaluate genetic

128 similarity and diversity of these traits across populations, we calculated the global

129 trans-ethnic genetic correlation (ρ_g) via popcorn [12]. We identified common trait-

130 associated SNPs via the conditional false discovery rate (cFDR) and conjunction

131 conditional false discovery rate (ccFDR) methods in the two populations [27, 28].

132 Relying on cFDR and ccFDR, for each trait we divided all analyzed SNPs into four

133 incompatible groups. For every trait we estimated the marginal genetic correlation (r_m)

134 of SNP effect sizes in each of the four groups using MAGIC [25]. For common trait-

135 associated SNPs, we examined the heterogeneity in genetic effect on the trait across

136 EAS and EUR populations via Cochran's Q test.

137 Then, we obtained the two LD scores and MAF for every SNP in the four groups

138 based on genotypes available from EAS or EUR individuals in the 1000 Genomes

139 Project, and calculated their coefficient of variation of LD scores (LDCV) and

140 coefficient of variation of MAF (MAFCV) across the populations to investigate

141 whether there exist different patterns of LD and MAF for trait-associated SNPs

142 compared to those null ones.

143 Finally, to further demonstrate the direction of genetic differentiation, for each trait we

144 conducted a GRS analysis [29]. We also investigated whether the sample size

145 difference could influence our findings with regards to genetic similarity and diversity

146 of traits between the EAS and EUR populations.

Heritability and trans-ethnic genetic correlation

Estimated heritability

We first present the estimated SNP-based heritability ($\hat{h}^2$) of these traits. It was shown some traits (e.g., height and schizophrenia [SCZ]) were more heritable, with a large heritability estimate; but other (e.g., ischemic stroke [IS] and atopic dermatitis [AD]) exhibited low heritability (**Table 1**). Although the estimates of heritability were highly correlated between the EAS and EUR populations (Pearson's correlation=0.738, P=1.83×10^{-7}), there still existed an obvious trans-ethnic distinction in heritability. For example, platelet count (PLT) showed the maximal difference in estimated heritability, with $\hat{h}^2$=1.1% (se=1.4%) in EAS but 20.1% (se=1.8%) in the EUR population; in contrast, the heritability of atrial fibrillation (AF) was much larger in the EAS population compared to that in the EUR population (9.5% (se=2.7%) vs. 2.0% (se=0.3%)). By carrying out an approximate normal test, we observed that 70.3% (=26/37) of heritability estimates were significantly different across the two populations (FDR<0.05) (**Table 1**), largely reflecting diversity in polygenic genetic architecture across traits and ancestral groups.

We did not find a substantial correlation between sample size and heritability in both populations (Pearson's correlation=0.112 with P=0.508 in the EAS population, and Pearson's correlation=-0.025 with P=0.884 in the EUR population). However, as expected, we indeed discovered suggestive evidence that larger sample size can lead to more accurate estimate of heritability, with Pearson's correlation=-0.266 in the EAS population and -0.102 in the EUR population between sample size and standard error of heritability. We further used the coefficient of variation of sample sizes (NCV) to measure the difference of sample sizes, but did not observe a significant correlation between NCV and the coefficient of variation of heritability (Pearson's correlation=0.029, P=0.867).

Table 1. Summary information of complex traits analyzed in the present study

trait	$\hat{h}_1^2$ (se_1)	$\hat{h}_2^2$ (se_2)	$P_{\Delta h^2}$	$\hat{\rho}_g$ (se, P_ρ)	trait	$\hat{h}_1^2$ (se_1)	$\hat{h}_2^2$ (se_2)	$P_{\Delta h^2}$	$\hat{\rho}_g$ (se, P_ρ)
SCZ	0.360 (0.020)	0.364 (0.015)	8.73×10^{-1}	**0.74 (0.10, 1.33×10^{-2})**	TG	0.117 (0.037)	0.209 (0.045)	1.14×10^{-1}	-
RA	0.143 (0.034)	0.115 (0.016)	4.56×10^{-1}	**0.70 (0.14, 2.65×10^{-2})**	HbA1c	0.073 (0.014)	0.039 (0.006)	**2.56×10^{-2}**	0.98 (0.17, 9.25×10^{-1})
AF	0.095 (0.027)	0.020 (0.003)	**5.77×10^{-3}**	-	eGFR	0.071 (0.007)	0.058 (0.004)	1.07×10^{-1}	**0.83 (0.05, 4.42×10^{-4})**
T2D	0.066 (0.004)	0.041 (0.002)	**2.27×10^{-8}**	0.93 (0.04, 5.89×10^{-2})	AAM	0.081 (0.008)	0.125 (0.006)	**1.08×10^{-5}**	**-**
COA	0.015 (0.003)	0.043 (0.006)	**3.00×10^{-5}**	**0.57 (0.09, 1.66×10^{-6})**	ANM	0.082 (0.013)	0.132 (0.013)	**6.54×10^{-3}**	**0.66 (0.09, 2.34×10^{-4})**
AOA	0.015 (0.003)	0.030 (0.003)	**4.07×10^{-4}**	**0.53 (0.11, 1.21×10^{-5})**	PLT	0.011 (0.014)	0.201 (0.018)	**7.95×10^{-17}**	**0.84 (0.07, 2.51×10^{-2})**
AD	0.008 (0.002)	0.040 (0.015)	**3.45×10^{-2}**	-	RBC	0.085 (0.010)	0.151 (0.017)	**8.19×10^{-4}**	0.92 (0.06, 1.39×10^{-1})
BRC	0.053 (0.030)	0.106 (0.010)	9.38×10^{-2}	-	MCV	0.132 (0.018)	0.227 (0.032)	**9.67×10^{-3}**	0.87 (0.07, 7.48×10^{-2})
IS	0.010 (0.002)	0.008 (0.001)	3.71×10^{-1}	0.89 (0.20, 5.69×10^{-1})	HCT	0.052 (0.006)	0.104 (0.010)	**8.24×10^{-6}**	0.88 (0.08, 1.38×10^{-1})
PCA	0.038 (0.007)	0.055 (0.007)	8.59×10^{-2}	0.83 (0.14, 2.13×10^{-1})	MCH	0.114 (0.017)	0.224 (0.036)	**5.73×10^{-3}**	0.87 (0.12, 2.52×10^{-1})
TL	0.075 (0.023)	0.062 (0.014)	6.29×10^{-1}	**0.61 (0.20, 4.63×10^{-2})**	MCHC	0.040 (0.007)	0.083 (0.014)	**6.01×10^{-3}**	0.84 (0.15, 2.75×10^{-1})
BMI	0.141 (0.008)	0.181 (0.007)	**1.68×10^{-4}**	**0.84 (0.04, 1.53×10^{-5})**	HGB	0.050 (0.006)	0.111 (0.013)	**2.04×10^{-5}**	**0.79 (0.10, 3.70×10^{-2})**
height	0.327 (0.018)	0.286 (0.015)	8.01×10^{-2}	**0.86 (0.04, 2.57×10^{-4})**	MONO	0.059 (0.010)	0.163 (0.019)	**1.27×10^{-6}**	**0.80 (0.09, 2.99×10^{-2})**
DBP	0.046 (0.005)	0.106 (0.005)	**2.16×10^{-17}**	**0.73 (0.06, 6.70×10^{-6})**	NEUT	0.087 (0.011)	0.026 (0.005)	**4.46×10^{-7}**	**0.77 (0.06, 2.42×10^{-4})**
SBP	0.056 (0.006)	0.109 (0.004)	**1.99×10^{-13}**	**0.71 (0.05, 5.62×10^{-9})**	EO	0.055 (0.011)	0.130 (0.013)	**1.06×10^{-5}**	**0.76 (0.09, 5.98×10^{-3})**
PP	0.037 (0.004)	0.094 (0.004)	**7.07×10^{-24}**	**0.73 (0.07, 8.80×10^{-5})**	BASO	0.039 (0.014)	0.058 (0.007)	2.25×10^{-1}	**0.63 (0.12, 2.07×10^{-3})**
HDL	0.154 (0.033)	0.241 (0.061)	2.10×10^{-1}	-	LYMPH	0.057 (0.010)	0.144 (0.011)	**4.85×10^{-9}**	0.84 (0.10, 8.26×10^{-2})
LDL	0.058 (0.012)	0.171 (0.029)	**3.18×10^{-4}**	0.77 (0.18, 1.97×10^{-1})	WBC	0.070 (0.008)	0.139 (0.013)	**6.17×10^{-6}**	**0.75 (0.06, 3.85×10^{-5})**
TC	0.054 (0.009)	0.183 (0.026)	**2.75×10^{-6}**	0.92 (0.11, 4.79×10^{-1})					

Note: the first ten traits are binary, and the remaining 27 traits are continuous; $\hat{h}_1^2$ and $\hat{h}_2^2$ are the estimated heritability of the trait in EAS and EUR GWASs using LDSC; se_1 and se_2 are the standard error of the trait in EAS and EUR GWASs. $P_{\Delta h^2}$ denotes the P values available from an approximate normal test for examining the difference between $\hat{h}_1^2$ and $\hat{h}_2^2$, with the significant difference after multiple-comparison correction showing in bold (FDR<0.05). The significant trans-ethnic genetic correlations (i.e., $\hat{\rho}_g$) compared with one are also displayed in bold. SCZ: schizophrenia; AF: atrial fibrillation; T2D: type 2 diabetes; COA: childhood-onset asthma; AOA: adult-onset asthma; AD: atopic dermatitis; BRC: breast cancer; IS: ischemic stroke; PCA: prostate cancer; RA: rheumatoid arthritis; BMI: body mass index; DBP: diastolic blood pressure; SBP: systolic blood pressure; PP: pulse pressure; HDL: high-density lipoprotein cholesterol; LDL: low-density lipoprotein cholesterol; TC: total cholesterol; TG: triglyceride; HbA1c: hemoglobin A1c; eGFR: estimated glomerular filtration rate; AAM: age at menarche, ANM: age at natural (non-surgical) menopause; PLT: platelet count; RBC: red blood cell count; MVC: mean corpuscular volume; HCT: hematocrit; MCH: mean corpuscular hemoglobin; MCHC: mean corpuscular hemoglobin concentration; HGB: hemoglobin concentration; MONO: monocyte count; NEUT: neutrophil count; EO: eosinophil count; BASO: basophil count; LYMPH: lymphocyte count; WBC: white blood cell count; TL: telomere length.

Estimated trans-ethnic genetic correlation

The trans-ethnic genetic correlation estimates of six traits (i.e., AF, AD, breast cancer [BRC], high-density lipoprotein cholesterol [HDL], age at menarche [AAM], and triglyceride [TG]) were larger than one and thus not included in the following descriptions. All the traits exhibited positive trans-ethnic genetic correlation between the EAS and EUR populations (**Table 1**), with $\hat{\rho}_g$ ranging from 0.53 (se=0.11) for adult-onset asthma (AOA) to 0.98 (se=0.17) for hemoglobin A1c (HbA1c).

All the trans-ethnic genetic correlation estimates were significantly larger than zero (H_0: ρ_g=0), but 61.3% (=19/31) were less than one (H_0: ρ_g=1) (FDR<0.05), indicating potential heterogeneity in genetic effects across populations. The average of $\hat{\rho}_g$ across traits was 0.79, and the average of $\hat{\rho}_g$ for binary phenotypes was 0.74 (se=0.15), which was slightly (although not significantly, P=0.217, possibly due to limited number of binary phenotypes under analysis) smaller than that for continuous ones (the average of $\hat{\rho}_g$=0.81 (se=0.09)).

Comparison for trans-ethnic genetic correlation, cross-trait trans-ethnic genetic correlation, and cross-trait genetic correlation

Furthermore, as an empirical comparison, we calculated cross-trait trans-ethnic genetic correlation and anticipated that the cross-trait correlation should be on average much smaller than that for the same traits in the two populations because of greatly distinct genetic foundations. As expected, it was found the same traits generally showed much greater genetic similarity compared to distinct traits although some of them, such as three lipids including low-density lipoprotein cholesterol [LDL], total cholesterol [TC], and TG, as well as three blood pressures including diastolic blood pressure [DBP], systolic blood pressure [SBP], and pulse pressure [PP], also exhibited relatively high cross-trait trans-ethnic genetic correlation (**Figure 1A** and **Table S1**). For instance, the cross-trait trans-ethnic genetic correlation ranged from -0.62 to 0.79, approximately 47.0% were negative, with an average of only 0.02 (**Figure 1B**).

We also included the cross-trait genetic correlation in the same population as a reference. Totally, it was seen that the cross-trait genetic correlations in the same

214   population were comparable to the cross-trait trans-ethnic genetic correlations across

215   different populations but were much lower than the trans-ethnic genetic correlation of

216   the same trait in diverse populations (**Figure 1B** and **Table S2**).
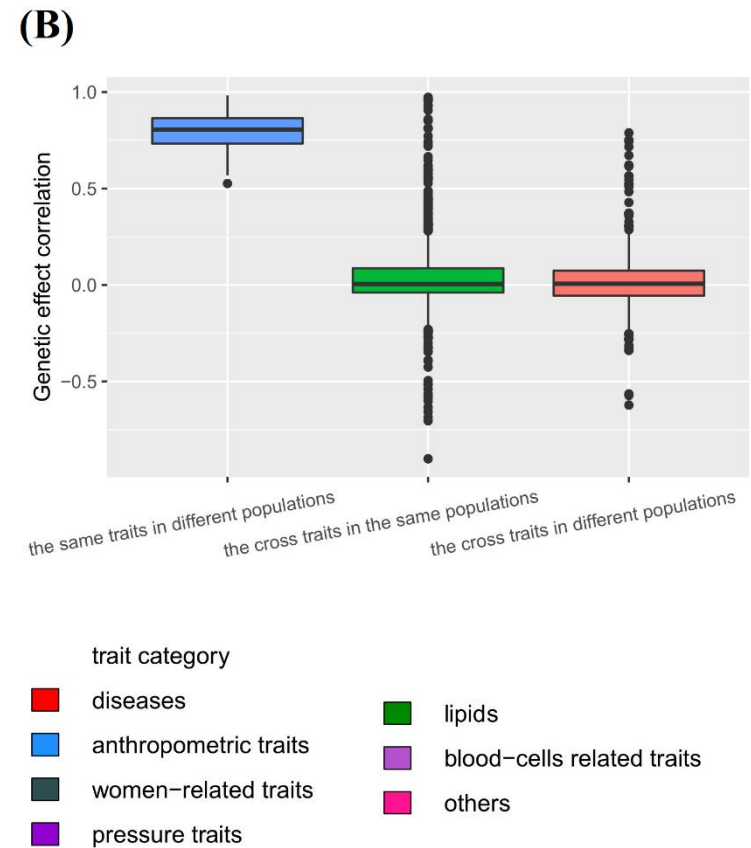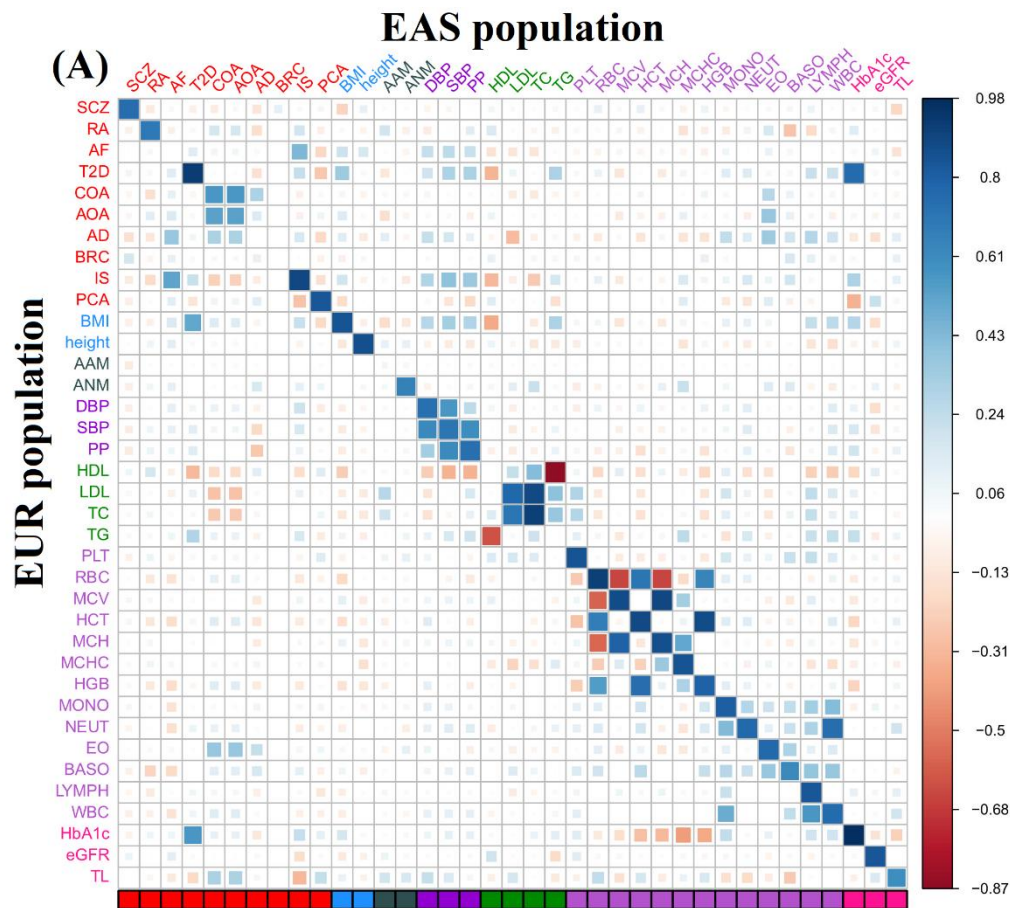
Figure 1. (**A**) Estimated trans-ethnic genetic correlation of 37 traits between the EAS and EUR populations. Elements in the diagonal represent the trans-ethnic genetic correlation for the same traits in the two populations, while elements in the off-diagonal represent the trans-ethnic genetic correlation for two different traits in the two populations. (**B**) Comparison of the estimated genetic correlation for the trait in the same population, two different traits in the same population, and two different traits in the EAS and EUR populations.

222   We did not discover a substantial connection between NCV and $\hat{\rho}_g$ (Pearson's

223   correlation=-0.148, P=0.426), implying that the imbalanced sample sizes would not

224   significantly affect the estimate of trans-ethnic genetic correlation. However, we

225   observed that the difference of sample sizes in a pair of traits could affect the

226   significance of $\hat{\rho}_g$. More specifically, $\hat{\rho}_g$ trended to be statistically different from one

227   if larger imbalance of sample sizes was present (NCV=0.43 (se=0.25) for the 17 non-

228   significant traits vs. 0.69 (se=0.27) for the 14 significant ones, P=0.010 in terms of the

229   approximate normal test).
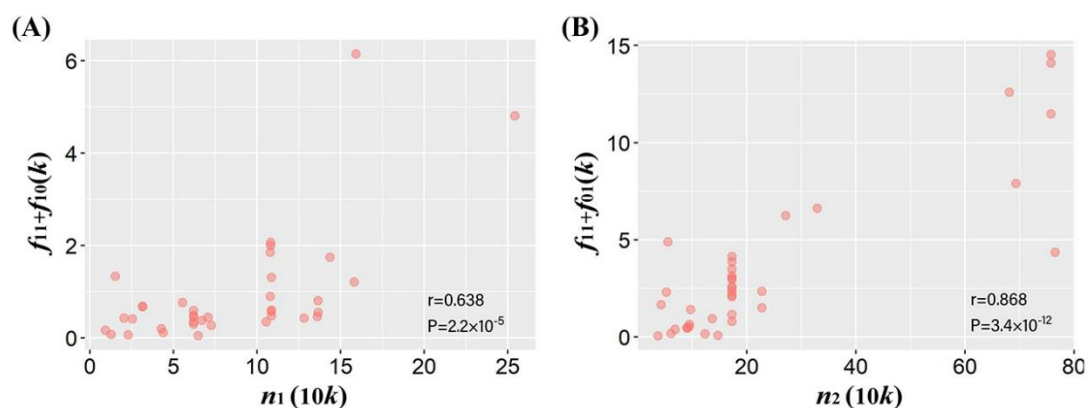
230
231   **Figure 2**. (**A**) Relationship between n_1 and $f_{11}+f_{10}$; (**B**) Relationship between n_2 and
232    $f_{11}+f_{01}$. Here, f_{10} and f_{01} are the number of identified associated SNPs for the trait in
233   the EAS or EUR population, respectively, and f_{11} is the number of shared trait-
234   associated SNPs; n_1 and n_2 stand for the effective sample size for traits in the EAS or
235   EUR population; r stands for the Pearson's correlation, with P the corresponding P
236   value; k on the x-axis means 1,000 for identified trait-associated SNPs and $10k$ on the
237   y-axis means 10,000 for sample size of traits.

238   **Associated SNPs for each trait in the EAS or EUR population**

239   To discover trait-associated SNPs shared across the EAS and EUR populations, we

240   carried out the cFDR analysis using a set of distinct SNPs [27, 28], with the number

241   of significant associations displayed in **Table 2**. In general, more associated SNPs

242   (cFDR<0.05) were discovered for most of the analyzed traits (94.6%=35/37, except

243   for telomere length [TL] and HbA1c) in the EUR population (denoted by $f_{01}+f_{11}$ in

244   **Table 2**) compared to the EAS population (denoted by $f_{10}+f_{11}$ in **Table 2**). This is a

245   direct consequence of higher power due to larger sample sizes of EUR traits (**Figure**

246   **2A-B**) and implies that additional trait-associated loci would be detected if increasing

247   samples especially in the EAS population. Only a few of SNPs were detected for

248   some traits such as AD and IS, but much more associated SNPs were identified for

249   others (e.g., BMI and height), partly indicating the polygenic nature of these traits [3,

250   31].

251   **Table 2**. Number of associated SNPs (cFDR/ccFDR<0.05) for traits in the EAS and EUR populations

trait	k	f_{10}	f_{01}	f_{11}	$\hat{\pi}_{11}(\%)$	P_{LRT}	trait	k	f_{10}	f_{01}	f_{11}	$\hat{\pi}_{11}(\%)$	P_{LRT}
SCZ	458,236	154	4,295	605	12.0	1.8×10^{-266}	TG	121,414	162	259	181	30.1	7.3×10^{-86}
RA	203,961	450	781	881	41.7	1.8×10^{-266}	HbA1c	124,962	88	50	106	43.4	3.3×10^{-58}
AF	236,293	166	1,262	242	14.5	1.2×10^{-51}	eGFR	236,693	805	3,413	933	18.1	1.8×10^{-266}
T2D	475,656	2,897	4,331	1,904	20.8	1.8×10^{-266}	AAM	319,439	35	6,271	340	5.1	7.6×10^{-174}
COA	250,069	259	1,877	421	16.5	6.2×10^{-259}	ANM	137,069	31	296	85	20.6	2.3×10^{-48}
AOA	249,871	239	970	432	26.3	1.1×10^{-217}	PLT	222,492	793	2,929	1,205	24.5	1.8×10^{-266}
AD	274,668	81	84	80	32.7	1.2×10^{-51}	RBC	222,589	413	2,258	890	25.0	1.8×10^{-266}
BRC	337,234	3	2,274	71	3.0	2.8×10^{-24}	MCV	222,587	790	2,597	1,273	27.3	1.8×10^{-266}
IS	246,666	10	30	39	49.4	1.6×10^{-25}	HCT	222,528	141	1,691	457	20.0	1.3×10^{-256}
PCA	247,201	134	641	293	27.4	2.4×10^{-171}	MCH	222,441	661	2,308	1,184	28.5	1.8×10^{-266}
TL	125,043	23	17	43	51.8	9.1×10^{-48}	MCHC	222,580	233	817	338	24.4	2.4×10^{-213}
BMI	118,899	438	11,818	768	5.9	1.8×10^{-134}	HGB	222,563	132	1,725	339	15.4	1.9×10^{-226}
height	139,126	3,528	5,292	2,616	22.9	1.8×10^{-266}	MONO	222,622	147	2,524	440	14.1	3.8×10^{-160}
DBP	214,705	84	13,624	468	3.3	2.1×10^{-128}	NEUT	222,435	170	2,020	309	12.4	4.3×10^{-201}
SBP	213,840	165	13,874	642	4.4	6.1×10^{-150}	EO	222,395	114	2,190	334	12.7	6.9×10^{-184}
PP	213,862	46	11,069	409	3.5	2.7×10^{-125}	BASO	222,525	143	598	200	21.3	2.7×10^{-90}
HDL	122,159	218	305	227	30.3	1.2×10^{-104}	LYMPH	222,490	32	2,307	263	10.1	1.3×10^{-160}
LDL	121,447	94	287	178	31.8	1.0×10^{-112}	WBC	222,550	343	2,451	555	16.6	6.7×10^{-272}
TC	122,106	151	367	275	34.7	3.9×10^{-170}							

252   **Note**: k is the total number of SNPs analyzed by cFDR; f_{10} and f_{01} are the number of identified SNPs that are only associated with the trait in the
253   EAS or EUR population, respectively, and f_{11} is the number of shared associated SNPs; $\hat{\pi}_{11}$ is the proportion of trait-associated SNPs that are
254   shared by both populations estimated in terms of the four-group model; P_{LRT} is the adjusted P value of genetic overlap obtained by the LRT
255   method in the four-group model.

Shared associated SNPs of traits between the EAS and EUR populations

Furthermore, we identified many significant SNPs shared across the EAS and EUR populations (ccFDR<0.05, denoted by f_{11} in **Table 2**). On average, 21.7% of associated SNPs were discovered simultaneously for the same traits in both populations. The LRT implemented under the four-group model framework also offered considerably strong evidence supporting common genetic foundation underlying each trait between the two population. The proportion of shared associated SNPs varied substantially across these traits, ranging from only 3.0% for BRC and 3.3% for DBP to 49.4% for IS and 51.8% for TL (**Table 2**).

On average, approximately 67.1% of trait-associated SNPs in the EUR population were also detected to be significant in the EAS population, but only 26.8% of trait-associated SNPs in the EAS population were replicated to be significant in the EUR population. Among these shared significant SNPs, 44.1% and 77.0% showed genome-wide significance ($P<5\times10^{-8}$) in the EAS or EUR population (**Table S3**). We observed that the number of population-common associated loci (i.e., f_{11}) was negatively correlated with NCV (Pearson's correlation=-0.095, with a marginally significant P value of 0.057), indicating that smaller difference of sample sizes in a pair of traits (e.g., increasing the sample size of traits in the EAS population) might lead to more discoveries of shared SNPs.

Characteristics of shared associated SNPs

Similarity and heterogeneity of associated SNPs across populations

It needs to highlight that we can divide all analyzed SNPs into four groups based on the associations identified above: (i) null SNPs; (ii) EAS-specific associated SNPs; (iii) EUR-specific associated SNPs; (iv) population-common associated SNPs. For these SNPs, it is shown that population-common associated SNPs often exhibited a maximal positive correlation in effect sizes compared to population-specific associated SNPs and null ones (**Figure 3A** and **Table S4**). For example, the marginal genetic correlation of effect sizes for shared trait-associated SNPs was 0.92 (se=0.04) for TL, followed by 0.81 (se=0.02) for white blood cell count (WBC) and 0.80 (se=0.03) for NEUT, with an average estimate of 0.61 (se=0.04) across these traits,

286   which was much higher than that for EAS-specific ($\hat{r}_m$=0.29 (se=0.05)) or EUR-

287   specific ($\hat{r}_m$=0.33 (se=0.05)) trait-associated SNPs or null SNPs ($\hat{r}_m$=0.09 (se=0.02)).
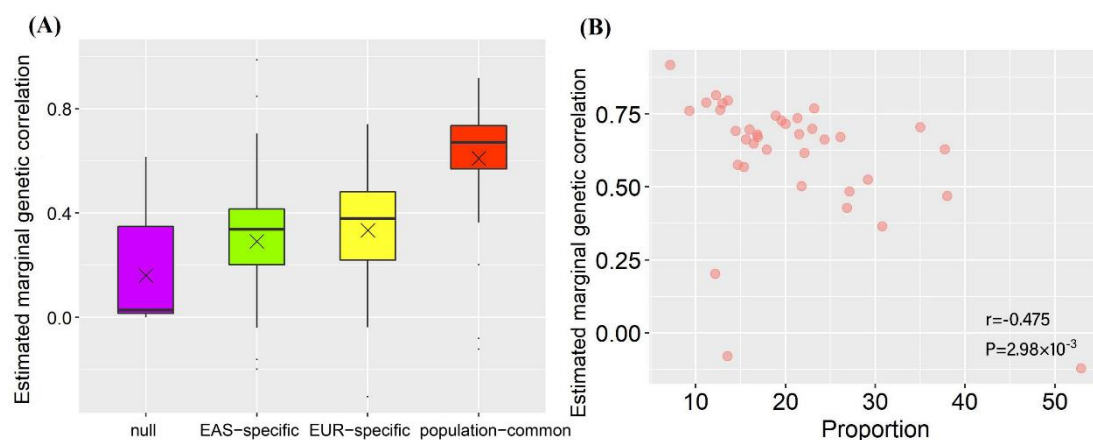
288
289   **Figure 3**. (**A**) Estimated marginal genetic correlation of effect sizes for SNPs in the
290   four incompatible groups; (**B**) Relationship between proportions of genetic effect
291   heterogeneity of shared associated SNPs and the cross-population marginal genetic
292   correlations of SNP effect sizes; × means the median.

293   We found that the effect size slopes of population-common associated SNPs of eight

294   traits (i.e., AF, AD, BRC, TL, SBP, PP, age at natural non-surgical menopause [ANM],

295   and basophil count [BASO]) were not significantly different from one (FDR>0.05)

296   (**Supplement File** and **Figure S1**), suggesting that effect sizes of shared associated

297   SNPs are considerably consistent in magnitude for these traits; whereas great

298   deviations of the estimated slopes from one were observed for the remaining 29 traits

299   (FDR<0.05) (e.g., HbA1c, T2D, estimated glomerular filtration rate (eGFR), SCZ and

300   AAM), indicating substantial trans-ethnic diversity of SNP effect sizes for these traits.

301   In addition, we did not detect substantial evidence supporting the influence of

302   imbalanced sample sizes (i.e., NCV) on the estimated slope (*P*=0.053).

303   We next performed a heterogeneity test using Cochran's Q test and again focused only

304   on population-common associated SNPs because the population-specific associated

305   SNPs can be naturally considered to be heterogeneous (i.e., theoretically, their SNP

306   effect sizes were non-zero in one population but zero in another population). On

307   average, approximately 20.8% of the common trait-associated SNPs, ranging from 7.1%

308   for TL to 52.9% for AAM, showed heterogeneous genetic effect on traits between the

309   two ancestral populations after Bonferroni's correction for multiple comparisons

310   (**Table S5**). We here used Bonferroni's method to take the LD among local SNPs into

311   account as it was much more stringent compared to FDR. The high heterogeneity in

312   SNP effect sizes for AAM was in agreement with a prior finding that AAM-associated

313   SNPs often exhibited distinct effect sizes across populations [32]. It can be expected

314   that greater proportion of shared trait-associated SNPs having heterogeneous effects

315   would lead to weaker trans-ethic marginal genetic correlation (Pearson's correlation=-

316   0.48, P=2.98×10^{-3}) (**Figure 3B**). However, we found little evidence supporting the

317   influence of the sample size difference in a pair of traits (i.e., NCV) on the proportion

318   of population-common SNPs with heterogeneous effects (P=0.771).

319   **Difference in LD, MAF, and F_{st} for trait-associated SNPs across populations**

320   First, we observed that, except for three lipid traits (i.e., TL, HDL, and LDL), all the

321   remaining traits showed substantial different variations in LD for SNPs in various

322   groups between the EAS and EUR populations (FDR<0.05) (**Figure 4A** and **Figure**

323   **S2**). The average coefficient of variation of LD scores (LDCV) for population-

324   common SNPs was smaller compared to that for null SNPs (0.22 (se=0.02) vs. 0.30

325   (se=0.02), P=1.04×10^{-17}), for these population-specific loci in EAS (0.22 (se=0.02) vs.

326   0.28 (se=0.04), P=1.82×10^{-9}) or EUR (0.22 (se=0.02) vs. 0.28 (se=0.03), P=1.52×10^{-12})

327   ).
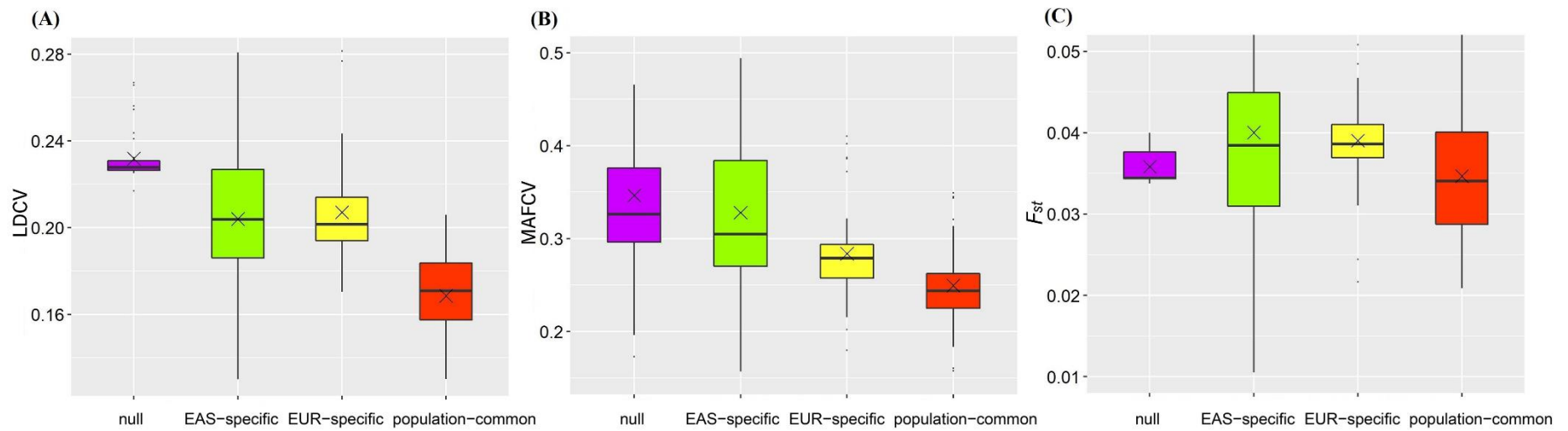
Figure 4. (**A**) Distribution for the average LDCV across all analyzed traits in the four groups of SNPs; (**B**) Distribution for the average MAFCV across all analyzed traits in the four groups of SNPs; (**C**) Distribution for the average F_{st} across all analyzed traits in the four groups of SNPs; × means the median.

332 On the other hand, we observed that, except for three traits (i.e., HDL, TG, and ANM),

333 all the remaining traits showed substantially different MAFCV for SNPs in distinct

334 groups between the EAS and EUR populations (FDR<0.05) (**Figure 4B** and **Figure**

335 **S3**). On average, the MAFCV for population-common loci were much smaller

336 compared to that for null SNPs (0.33 (se=0.07 vs. 0.46 (se=0.10), $P=1.54\times10^{-12}$), for

337 these population-specific loci in EAS (0.33 (se=0.07) vs. 0.43 (se=0.11), $P=7.30\times10^{-8}$)

338 or EUR (0.33 (se=0.07) vs. 0.40 (se=0.07), $P=2.22\times10^{-8}$).

339 We found that SNPs suffered from natural selection for more than half of traits

340 (62.2%=23/37) (FDR<0.05) (**Figure S4**). For all analyzed traits, we observed that

341 population-common loci tended to have a smaller mean F_{st} compared with population-

342 specific associated SNPs in EAS (0.056 (se=0.01) vs. 0.062 (se=0.02), $P=0.032$) or

343 EUR (0.056 (se=0.01) vs. 0.063 (se=0.006), $P=6.77\times10^{-4}$), and they also showed a

344 lower mean F_{st} relative to null ones (0.056 (se=0.01) vs. 0.060 (se=0.001), with a

345 marginally significant $P=0.086$) (**Figure 4C**). Furthermore, we did not observe

346 significant correlation between F_{st} and LDCV ($P=0.781$) as well as between F_{st} and

347 MAFCV ($P=0.602$) (**Figure S5**), indicating the between-population diversity of LD

348 and MAF patterns is not possibly confounded by the differentiation in F_{st}.

Overall and partial GRS of trait-associated SNPs

350 First, we considered the GRS calculated from all associated SNPs (i.e., $f_{10}+f_{11}$ for

351 EAS or $f_{01}+f_{11}$ for EUR) as an overall measurement of genetic effect on a given trait in

352 each population. Among these, most of the traits had a substantially different GRS

353 between the two populations (**Figure S6**). For example, six traits (i.e., RA, T2D,

354 childhood-onset asthma [COA], AOA, BRC and prostate cancer [PCA]) showed a

355 higher overall genetic effect on EAS, while some (i.e., SCZ, AF, AD and IS)

356 displayed a larger overall genetic effect on EUR.

357 We further generated the GRS computed with only population-common associated

358 SNPs (i.e., f_{11}) or only population-specific associated SNPs (i.e., f_{10} for EAS or f_{01} for

359 EUR), respectively. We viewed these new GRSs to quantify a measurement of partial

360 genetic effect because only part of associated SNPs is employed. Interestingly, 16

361 traits (i.e., RA, T2D, AD, PCA, DBP, SBP, LDL, TC, TG, eGFR, ANM, mean

362 corpuscular hemoglobin [MCV], hematocrit [HCT], monocyte count [MONO],

363   eosinophil count [EO] and WBC) had consistent partial genetic effect compared to the

364   overall one. Specifically, 15 traits (e.g., AF, COA, AOA, BRC, BMI, height, PP,

365   HDL, HbA1c, RBC, mean corpuscular hemoglobin [MCH], mean corpuscular

366   hemoglobin concentration [MCHC], hemoglobin concentration [HGB], NEUT, and

367   BASO) showed consistent partial genetic effect in terms of GRS calculated with EAS-

368   or EUR-specific associated SNPs, whereas only 4 traits (i.e., SCZ, AAM, PLT, and

369   lymphocyte count [LYMPH]) exhibited consistent partial genetic effect in terms of

370   GRS calculated with shared associated SNPs. Particularly, IS and TL exhibited a

371   completely opposite partial genetic effect in terms of the EAS/EUR-specific

372   associated SNPs or population-common associated SNPs compared with the overall

373   genetic impact measured with all trait-associated SNPs.

Discussion

Summary of results in the present study

374   The present study has analyzed a total of 37 complex traits and sought to compare

375

376   The present study has analyzed a total of 37 complex traits and sought to compare

377   shared and distinct genetic components underlying them between the EAS and EUR

378   populations. We discovered there existed pervasive consistence in heritability and

379   trans-ethnic genetic correlation for these traits. Additionally, it needs to highlight that

380   the trans-ethnic genetic correlation of continuous traits was on average higher

381   compared to those binary traits, which may be due to the loss of information when

382   converting some continuous phenotypes into binary ones (e.g., using HbA1c to define

383   T2D) or in part reflects the discrepancy of classification and diagnosis of diseases in

384   distinct populations.

385   Using cFDR [27, 28] we detected a great number of population-common association

386   signals as well as many population-specific associated SNPs. A further exploration

387   demonstrated these shared trait-associated SNPs generally showed the maximal

388   positive correlation in effect sizes compared to population-specific trait-associated

389   SNPs and null ones [7, 25]. Interestingly, we observed that even the shared

390   association signals also exhibited a considerable degree of heterogeneity in genetic

391   influence on traits across the EAS and EUR populations.

392   Furthermore, we revealed that population-specific associated SNPs were often more

393   possible to suffer from natural selection compared with population-common

394   associations, whereas population-common associated SNPs often displayed more

395   consistent patterns in LD and MAF across continental populations.

396   Especially, the GRS analysis showed that population-common and population-specific

397   associated SNPs have potentially different genetic influence on phenotypic variation

398   and that the genetic differentiation from associated SNPs may at least partly explain

399   the observed phenotypic variation across diverse ancestral groups. For example, the

400   average GRS of SCZ in the EUR population was on average higher than that in the

401   EAS population, partly contributing to the observation that SCZ was more prevalent

402   in individuals of EUR ancestry than those of EAS ancestry [33]. The mean GRS of

403   T2D was higher in the EAS population than that in the EUR population, supported by

404   the observed incidence difference between the populations [34, 35]. It was reported

405   that the absolute risk of T2D tended to be higher among Asians compared with

406   Caucasians for any given level of body mass index [BMI] and waist-hip ratio [34]. As

407   another example, for PCA higher mean GRS was observed in EAS compared to EUR,

408   in line with a previous study which indicated that more than half of SNPs showed

409   larger effect sizes in EAS than EUR [36].

410   However, we also observed patterns that seemed to be opposite against prior findings.

411   For example, it was shown the mean GRS of AD in EUR was higher than that in EAS,

412   in contrast to previous observation that Asians and Pacific Islanders were seven-fold

413   more likely than whites to be diagnosed with AD because stronger Th17/Th22

414   polarization and mutations in immune-related genes such as *DEFB1* [37-39]. These

415   inconsistent results can be expected because of the complicated nature of these traits,

416   and can be explained by gene-gene/gene-environment interaction, ethnic difference

417   and genetic factors that are largely underappreciated in our current study.

418   **Comparison our discoveries to prior studies**

419   Like the findings obtained from prior studies that however often focused only on a

420   single trait [11, 40, 41], our work, which considered much more traits and diseases,

421   further confirmed extensive genetic overlap and identified a large number of common

422   associated genetic loci across different populations. From a biological perspective,

423   there is no doubt about the widespread existence of population-shared risk variants

424   because the risk variants targeted by GWASs are often common genetic loci that are

425   believed to be of ancient origin and largely shared among different populations.

426   As revealed in our study, while some of associated SNPs vary substantially across

427   populations, common associated SNPs in the EAS and EUR populations nevertheless

428   often show much more similar effect sizes and effect directions; therefore, at least part

429   of trait-associated SNP mapping results discovered in one population can be

430   transferred to the other populations [42]. It needs to highlight that we may

431   underestimate the degree of genetic sharing between various ancestry groups because

432   the much smaller number of individuals in the EAS GWASs reduces power to detect

433   homogeneity of effect compared to the EUR GWASs. These findings are largely

434   consistent with some recent discoveries that most common causal SNPs were shared

435   across the EAS and EUR populations, high-posterior SNPs identified by fine-mapping

436   often showed highly correlated effects, and population-specific genetic regions likely

437   harbored common trait-associated SNPs which however failed to be detected in the

438   other GWASs due to differences in LD, MAF, and/or sample size [43].

439   On the other hand, despite highly shared genetic architecture, we indeed found

440   evidence of heterogeneity at trait-associated variants, which meanwhile challenges in

441   assessing the transferability of risk variants across different ethnic populations based

442   on associations discovered in EUR GWASs [44, 45]. For example, although the trans-

443   ethnic marginal genetic correlation for population-common associated SNPs of

444   HbA1c between EAS and EUR was as high as 0.63, heterogeneity was detected at

445   37.7% shared associated loci. This diversity may point to the difference in clinical

446   definitions and phenotype measurements [46]; and it can be in part explained by

447   interaction between gene-gene and gene-environment [47].

448   The genetic difference may also underlie the well-known trans-ethnic dissimilarities

449   in prevalence or characteristics of the traits [48-52]. In our analysis, these population-

450   specific association signals largely indicate the significant trans-ethnic difference,

451   which exist in distinct LD and allele frequency [7, 53-55]. In addition, many studies

452   have revealed that, unlike in most European ancestry populations, the population

453   genetic history of non-European ancestry groups has undergone selective pressure due

454   to the effects of malaria and other infectious diseases on erythrocytes [40, 56].

455   Another possible factor for the genetic inconsistency of complex traits between

456   ancestry groups may be due to sample size difference and thus different statistical

457   power between EAS and EUR studies.

458   **Important scientific implications of our findings**

459   Our findings regarding the degree of common or diverse genetic components of the

460   traits across ancestral groups have important implication in practice. For example,

461   theoretically, genetic correlation offers the maximal boundary of trans-ethnic genetic

462   prediction power [5, 12]; however, both overall trans-ethnic correlation and marginal

463   trans-ethnic correlation imply low accuracy when implementing genetic prediction in

464   one population of interest on the basis of associated loci discovered in other

465   populations [25], indicating the need to carry out GWASs with more ancestral groups.

466   In addition, it is helpful for aggregating multiple study cohorts across ethnicities to

467   conduct trans-ethnic GWAS analysis [57-60], developing trans-ethnic genetic risk

468   prediction [25, 61], and fine-mapping causal genetic variants in minority populations

469   [62, 63]. It also holds the key to benefit more ethnic groups from current medical

470   genomics researches [64-66]. All of these offer promising avenues in post-GWAS era

471   by integrating trans-ethnic information.

472   **Potential limitations**

473   The present study is not without limitation. First, as mentioned before, the studies in

474   EAS are in generally underpowered due to much smaller sample size compared to that

475   in EUR and hence our results may be affected by power issues. The small sample size

476   may also lead to unstable effect estimation for these SNPs. Moreover, as shown before,

477   we cannot completely rule out the possibility that the imbalanced sample sizes can

478   also partly interpret the observed trans-ethnic genetic difference in traits. For example,

479   we found that sample size was significantly positively correlated with the number of

480   trait-associated loci in both populations, and that the estimated trans-ethnic genetic

481   correlation would become less significantly different from one when the difference of

482   sample sizes in a pair of traits reduced. Thus, the external validation of our results

483   with larger sample size especially for EAS GWASs is warranted.

484   Second, besides the difference in sample sizes, other discrepancies in study designs

485   such as phenotypic definition, statistical methods, and covariate adjustment can be

486   also contributable to the observed trans-ethnic genetic similarity and diversity.

487   Examining and quantifying the relative contributions are imperative for understanding

488   genetic heterogeneity across populations. However, compared to the difference in

489   sample sizes which are already reported in summary statistics, the design

490   discrepancies in GWAS are difficult to handle with only summary statistics. It needs

491   large-scale individual-level data of phenotypes and genotypes, and is thus challenged

492   by privacy concerns when sharing data [67].

493   Third, individuals of EAS and EUR ancestries are to a great extent genetically similar,

494   whereas more major genetic differences are expected to be found between AFR and

495   non-AFR populations [68]. Therefore, it is not clear whether our findings can be

496   generalized to comparison in other populations such as EUR and AFR. Unfortunately,

497   the number of GWASs performed in individuals of AFR descent is still too limited to

498   enable comparative studies.

499   Fourth, our analysis only considered common SNPs (MAF>0.01) and ignored rare

500   variants, which usually have a recent origin compared with common ones from

501   ancient origin. Theoretically, rare risk variants might be more likely to be population-

502   specific and could possibly carry a greater risk effect, which probably leads us to

503   underestimate the genetic heterogeneity across populations.

504   Fifth, to our knowledge, this is the first time that the cFDR method has been

505   employed in detecting trans-ethnic genetic overlap for a large range of complex traits.

506   However, it implements association mapping at a fixed FDR level rather than the

507   standard error measure such as family-wise error rate (FWER) or type I error rate

508   which is more widely-used in a typical GWAS. FDR is more liberal compared to

509   FWER; thus, we can discover a much larger number of trait-associated SNPs.

510   Although the cFDR method has been well-established under the context of pleiotropy-

511   informed association mapping in ancestry-matched populations [27, 28] and also

512   demonstrated to be well-calibrated in our study (**Figures S7-S8**), its ability of

513   controlling FWER is less understood. Consequently, we just considered cFDR as a

514   powerful tool for discovering evidence of trans-ethnic genetic overlap in our

515   application, and by no means attempted to replace the standard GWAS analysis

516   strategy with cFDR nor the cFDR level (e.g., 0.05 used here) with the genome-wide

517   significance level (e.g., 5×10^{-8}).

Conclusions

519   Our work provides an in-depth understanding of similarity and diversity regarding

520   genetic architecture for complex traits across diverse populations, and can assist in

521   trans-ethnic association analysis, genetic risk prediction, and causal variant fine

522   mapping.

523 **Materials and Methods**

524 **Summary statistics of 37 complex traits**

525 We obtained summary statistics (e.g., marginal effect size and standard error) of 37

526 complex traits (10 binary and 27 continuous) analyzed on EAS-only or EUR-only

527 individuals (**Table 1** and **Tables S6-S7**). These traits included lipids (e.g., TG), blood

528 cell phenotypes (e.g., neutrophil [NEUT] and MONO), diseases (e.g., BRC, T2D, and

529 prostate cancer [PCA]), and anthropometric phenotypes (e.g., BMI and height).

530 For each analyzed trait, we carried out stringent quality control in both populations by

531 following previous work [12, 30, 69]: (i) filtered out SNPs without rs label; (ii)

532 deleted non-biallelic SNPs and those with strand-ambiguous alleles; (iii) removed

533 SNPs whose alleles did not match with those in the 1000 Genomes Project; (iv)

534 excluded duplicated SNPs; (v) filtered out palindromic SNPs containing the same

535 bases; (vi) kept only common SNPs (MAF>0.01 in each population) which were also

536 included within the 1000 Genomes Project and shared between the EAS and EUR

537 GWASs.

538 Here, MAF was calculated with genotypes of EAS (N=504) or EUR (N=503)

539 individuals in the 1000 Genomes Project if missing in the original summary statistics

540 data; the threshold value of 0.01 for MAF was selected as it was widely used in

541 summary statistics-based studies [12, 30, 69]. We further aligned the effect allele of

542 all remaining SNPs for each trait between the two populations.

543 **Estimation of heritability and cross-trait genetic correlation in the same**
544 **population**

545 We first employed LDSC to estimate SNP-based heritability for each trait separately

546 in the two populations [30]. The LDSC software (version v1.0.1) was downloaded

547 from https://github.com/bulik/ldsc and the analysis was conducted with default

548 parameter settings. The LD score was calculated with genotypes of SNPs (MAF>0.01

549 and the P value of Hardy Weinberg equilibrium test>1.0×10^{-5}) with a 10Mb window

550 on EAS or EUR individuals in the 1000 Genomes Project. Then, the LD score of SNP

551 was regressed on the square of Z-statistic of the analyzed trait, with the regression

552 slope offering an unbiased estimate for heritability. Besides quality control procedures

553   described above, we here further removed SNPs located within the major

554   histocompatibility complex region (chr6: 28.5~33.5Mb) because of its complicated

555   structure which was often difficult to estimate accurately from an external reference

556   panel [30, 70]. Relying only on summary statistics and LD scores, LDSC can be also

557   applied to calculate the cross-trait genetic correlation in the same population [69].

Estimation of trans-ethnic genetic correlation across populations

559   To evaluate genetic similarity and diversity for these traits across populations, we

560   calculate the global trans-ethnic genetic correlation (ρ_g) via popcorn [12].

561   Conceptually, ρ_g is defined as the correlation between SNP effect sizes of the trait in

562   various ancestral groups to measure the extent to which the same SNP exhibits the

563   same or similar impact on phenotypic variation [12, 25, 26]. Methodologically,

564   popcorn was proposed from the Bayesian perspective by assuming effect sizes of

565   SNPs following an infinitesimal model [71], and can be considered as a natural trans-

566   ethnic extension of LDSC. The trans-ethnic LD score of each SNP was downloaded

567   from https://github.com/brielin/popcorn, which was calculated with genotypes of EAS

568   or EUR individuals in the 1000 Genomes Project between the focal SNP and all the

569   flanking ones within a 10Mb window. To obtain an unbiased estimate of trans-ethnic

570   genetic correlation, we implement an unbounded estimation algorithm in popcorn,

571   which likely leads to an estimate less than -1 or greater than 1.

Identification of associated SNPs and shared genetic overlap across populations

Conditional false discovery rate and conjunction conditional false discovery rate

574   From a statistical perspective, we observe that the identification of trans-ethnic

575   genetic overlap can be handled by applying the similar principle of detecting

576   pleiotropic associations for genetically correlated traits in a single ancestral group.

577   Over the past few years many methods have been developed for detecting pleiotropy

578   [72-74]. Among those, cFDR is a novel pleiotropy-informed method to discover

579   genetic overlap and can be viewed a novel extension of the classical FDR for a single

580   trait in one population to the same trait in trans-ethnic cases [27, 28]. By integrating

581   association results from multiple traits, this method could offer important sights into

582   trans-ethnic genetic overlap and increased power to identify less significant

583   association signals.

584   In our application context, the null hypothesis of FDR is no association between a
585   SNP and the trait of focus in one population. Based on this definition and the principle
586   of FDR, cFDR is logically defined as the posterior probability that a random SNP is
587   null for the trait in one population given that the observed P values for the trait in both
588   populations are less than a predetermined threshold. Formally, with two sets of P
589   values as input, cFDR is calculated as $cFDR_{1|2}=Prob(H_0^{(1)}|P_1<p_1, P_2<p_2)$, where p_1
590   and p_2 are the observed P values of a particular SNP for the trait in the two
591   populations, respectively; and $H_0^{(1)}$ denotes the null hypothesis indicating there does
592   not exist an association between the SNP and the trait in the first population. cFDR is
593   efficiently estimated with an empirical Bayesian manner that was proposed for
594   computing the local FDR [75].

595   As the principal and conditional positions for the trait in cFDR described above are
596   exchangeable between the populations, $cFDR_{2|1}$ is defined in a similar way. Therefore,
597   ccFDR for identifying shared SNPs is simply expressed as $ccFDR_{1|2}=max(cFDR_{1|2},$
598    $cFDR_{2|1})$, which is defined as the probability that a particular SNP has a false positive
599   association with the trait in the two populations given the observed P values. Finally,
600   SNPs with ccFDR less than a given significance threshold can be prioritized to be
601   population-common SNPs. Although the traditional meta-analysis is also often
602   applied in trans-ethic association studies [76-81]; however, the association discovered
603   by meta-analysis cannot certainly suggest trans-ethic genetic overlap because such
604   association might be only present in one special population.

605   Because cFDR and ccFDR are constructed for relatively independent SNPs, to
606   generate uncorrelated SNPs, following prior work [82] we applied the LD pruning
607   (the width of SNP window=50 and r^2=0.1) in PLINK with genotypes of EAS or EUR
608   individuals separately in the 1000 Genomes Project as the reference panel for LD
609   calculation, and then combined the two sets of SNPs available from both populations
610   (**Table S6**). In addition, based on findings observed in other studies [73, 74],
611   genetic/phenotypic correlation between traits can result in inflated test statistics.
612   Therefore, to minimize false discovery in our cFDR analysis, we further generated
613   uncorrelated Z-statistics for every trait across populations by multiplying them by the

614   inverse of a correlation matrix, which can be easily calculated in terms of the two

615   statistics of null SNPs (e.g., SNPs with $P>10^{-4}$) [72, 83]. This decorrelation strategy

616   maximizes the transformed test statistics and the original ones [84]; therefore, it has

617   the minimal influence on association identification. These uncorrelated Z-statistics

618   were ultimately transformed into two-sided P values for the cFDR analysis based on

619   normal approximation.

620   Our simulation studies already demonstrated that the cFDR method could maintain

621   well-calibrated control of FDR at the given level and behaved better when identifying

622   population-common trait-associated SNPs compared to the naïve minimum P-value

623   method (**Supplement notes** and **Figures S7-S8**).

624   **Various types of SNPs and four-group model**

625   Relying on cFDR and ccFDR, for each trait we could divide all the analyzed SNPs

626   into four incompatible groups: (i) not associated with the trait in neither population

627   (i.e., null SNPs); (ii) only associated with the trait in the EAS population but not in

628   the EUR population (i.e., EAS-specific associated SNPs); (iii) only associated with

629   the trait in the EUR population but not in the EAS population (i.e., EUR-specific

630   associated SNPs); (iv) associated with the trait in both the two populations (i.e.,

631   population-common associated SNPs).

632   To measure the degree of genetic components shared by the same trait across the

633   populations, we further applied the four-group model [85] which examined SNPs in

634   the four groups mentioned above. This model aims to estimate the proportions of

635   SNPs in each group, and employs the LRT method to assess the statistical significance

636   for overall trans-ethnic genetic overlap. Statistically, the four-group model assumes

637   that P values of null SNPs (not associated with the trait in neither populations) follow

638   the uniform distribution and P values of non-null SNPs (associated with the trait at

639   least in one population) follow the Beta distribution.

Genetic correlation and heterogeneity of SNPs between the two populations

Marginal genetic correlation across populations

For every trait we calculated the marginal genetic correlation (r_m) of SNP effect sizes in each of the four groups using a recently proposed method called MAGIC [25]. Compared to the traditional Pearson's correlation, which often underestimates the marginal genetic correlation thus leads to the so-called correlation attenuation because of failing to take the estimation error of effect sizes into account [25, 86], MAGIC has the advantage of generating unbiased correlation estimation by accounting for the uncertainty under the framework of measurement error model [87].

Linear regression for SNP effect sizes

We also carried out a simple linear model without the intercept term for only population-common associated SNPs of each trait by regressing effect sizes of SNPs in one population on those in another population. The slope of the linear regression model provides an indicator for the relative magnitude of effect sizes for shared trait-associated SNPs between the two populations. For common associated SNPs, we examined the heterogeneity in genetic effect of SNPs on the trait across EAS and EUR populations via Cochran's Q test in the R metafor package.

Characteristics of associated SNPs between EAS and EUR populations

Patterns for LD, MAF, and Wright's fixation index

After detecting common associations, we wondered whether there exist different patterns of LD and MAF for trait-associated SNPs compared to those null ones. To this aim, we first obtained the two LD scores for every SNP in the four groups based on genotypes available from EAS or EUR individuals in the 1000 Genomes Project, and then calculated their LDCV across the populations. Here, coefficient of variation, rather than variance, was utilized because SNPs with higher LD scores tended to have greater variation between populations [7, 24]. In a similar way, we calculated MAFCV for each SNP between the two populations to evaluate how MAF varies between populations.

We further evaluated whether an identified trait-associated SNP had been under

669 natural selection. If this was the case, then a substantial between-population

670 differentiation would be observed in the allele frequency [24, 88]. To this aim, we

671 applied the Wright's fixation index (F_{st}) to evaluate such an allele frequency diversity

672 across populations under natural selection [24, 88, 89], and calculated F_{st} of each SNP

673 in the four groups for every trait with genotypes available from EAS and EUR

674 individuals in the 1000 Genomes Project using PLINK.

675 To examine LDCV, MAFCV, and F_{st} in the four groups, we carried out the Kruskal

676 test for each trait, with the resulting P values being FDR corrected to account for

677 multiple testing using the Benjamini-Hochberg procedure. We also performed a paired

678 test to compare the average of LDCV, MAFCV, and F_{st} across the traits while simply

679 ignoring the uncertainty of the estimated average in each group.

680 **Partial and overall genetic risk score analysis**

681 To further demonstrate the direction of genetic differentiation, for each trait we

682 conducted a GRS analysis [29]. The calculated GRS in part measures the stratification

683 of the whole population based on estimates of individual's genetic susceptibility. In

684 our analysis the GRS of a given individual was simply computed as GRS=$\sum_j^J G_j \hat{\beta}_j$ /J,

685 where J was the number of SNPs used and G represented the genotype (coded as 0, 1,

686 or 2) available from the EAS or EUR individuals in the 1000 Genomes Project. For

687 binary phenotypes, $\hat{\beta}$ was the original marginal SNP effect size; whereas, for the

688 compatibility across populations, for continuous phenotypes we re-scaled $\hat{\beta}$ based on

689 Z-statistic [90]; that is, $\hat{\beta}_{new} = Z/\sqrt{2 \times \text{MAF} \times (1 - \text{MAF}) \times \text{sample size} + Z^2}$.

690 For each pair of traits under analysis, three types of GRS were generated using three

691 distinct sets of SNPs (**Table 2**), including population-specific loci ($J=f_{10}$ or f_{01}),

692 population-common loci ($J=f_{11}$) and population-associated loci ($J=f_{10}+f_{11}$ or $f_{01}+f_{11}$).

693 For the convenience of description, we referred to the first two GRSs as partial GRS,

694 while the third one as overall GRS.

695 **Exploring the influence of sample size difference**

696 We finally attempted to investigate whether the sample size difference could influence

697 our findings with regards to genetic similarity and diversity of traits between the EAS

698   and EUR populations. First, after obtaining the heritability estimate and its standard

699   error for a given trait in the two populations (denoted by $\hat{h}_1^2$ and $\hat{h}_2^2$,, se_1 and se_2,

700   respectively), we studied the influence of sample sizes of traits on the estimated

701   heritability and its standard error.

702   Second, to assess the impact of sample size on the trans-ethnic diversity of estimated

703   heritability, we employed NCV to measure the difference of sample sizes, and

704   examined its relation with the coefficient of variation of heritability, the estimated

705   slope of effect sizes for population-common SNPs and the proportion of population-

706   common SNPs with heterogeneous effects.

707   Third, we implemented an approximation normal test to examine the significance of

708   the difference in the estimated heritability between the two populations by calculating

709   $u = (\hat{h}_1^2 - \hat{h}_2^2)/(se_1^2 + se_2^2)^{-1/2}$. We obtained the P value of u using the standard normal

710   distribution as the null distribution, with the issue of multiple testing corrected via by

711   the Benjamini-Hochberg method. By doing this, we explicitly took the difference of

712   sample sizes in traits into account by modeling its standard error when comparing

713   heritability.

Declarations

Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations (declaration of Helsinki).

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information file.

Competing interests

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgements

Funding

741   University (TD202008). The research of Ting Wang was supported in part by the

742   Social Development Project of Xuzhou City (KC20062).

743   **Authors' contributions**

744   PZ conceived the idea for the study. PZ and JZ obtained and cleared the datasets; TW,

745   PZ, JQ, SZ and JZ performed the data analyses. TW, PZ, JQ, SZ, and JZ interpreted

746   the results of the data analyses. TW, PZ and JZ wrote the manuscript with the help

747   from other authors.

748   **Abbreviations**

749   SNP: single-nucleotide polymorphism

750   GWAS: genome-wide association study

751   EUR: European

752   EAS: East Asian

753   LD: linkage disequilibrium

754   cFDR: conditional false discovery rate

755   ccFDR: conjunction conditional false discovery rate

756   MAF: minor allele frequency

757   GRS: genetic risk score

758   LDSC: linkage disequilibrium score regression

759   LDCV: coefficient of variation of LD scores

760   MAFCV: coefficient of variation of MAF

761   F_{st}: Wright's fixation index

References

1. Tam V, Patel N, Turcotte M, Bosse Y, Pare G, Meyre D: **Benefits and limitations of genome-wide association studies**. *Nat Rev Genet* 2019, **20**(8):467-484.

2. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA: **Potential etiologic and functional implications of genome-wide association loci for human diseases and traits**. *Proc Natl Acad Sci U S A* 2009, **106**(23):9362-9367.

3. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, Yang J: **10 Years of GWAS Discovery: Biology, Function, and Translation**. *Am J Hum Genet* 2017, **101**(1):5-22.

4. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E *et al*: **The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019**. *Nucleic Acids Res* 2019, **47**(D1):D1005-D1012.

5. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD, Kenny EE: **Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations**. *Am J Hum Genet* 2017, **100**(4):635-649.

6. Carlson CS, Matise TC, North KE, Haiman CA, Fesinmeyer MD, Buyske S, Schumacher FR, Peters U, Franceschini N, Ritchie MD *et al*: **Generalization and Dilution of Association Results from European GWAS in Populations of Non-European Ancestry: The PAGE Study**. *PLoS Biol* 2013, **11**(9):e1001661.

7. Marigorta UM, Navarro A: **High Trans-ethnic Replicability of GWAS Results Implies Common Causal Variants**. *PLoS Genet* 2013, **9**(6):e1003566.

8. Kraft P, Zeggini E, Ioannidis JPA: **Replication in Genome-Wide Association Studies**. *Statistical Science* 2009, **24**(4):561-573.

9. Li YH, Lan L, Wang YL, Yang C, Tang WG, Cui GQ, Luo SQ, Cheng YB, Liu YC, Liu JY *et al*: **Extremely cold and hot temperatures increase the risk of diabetes mortality in metropolitan areas of two Chinese cities**. *Environ Res* 2014, **134**:91-97.

10. Waters KM, Stram DO, Hassanein MT, Le Marchand L, Wilkens LR, Maskarinec G, Monroe KR, Kolonel LN, Altshuler D, Henderson BE *et al*: **Consistent Association of Type 2 Diabetes Risk Variants Found in Europeans in Diverse Racial and Ethnic Groups**. *PLoS Genet* 2010, **6**(8):e1001078.

11. Liu JZ, van Sommeren S, Huang HL, Ng SC, Alberts R, Takahashi A, Ripke S, Lee JC, Jostins L, Shah T *et al*: **Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations**. *Nat Genet* 2015, **47**(9):979-986.

12. Brown BC, Ye CJ, Price AL, Zaitlen N, Network AGE: **Transethnic Genetic-Correlation Estimates from Summary Statistics**. *Am J Hum Genet* 2016,

806   **99**(1):76-88.

807   13.   Bigdeli TB, Ripke S, Peterson RE, Trzaskowski M, Bacanu SA, Abdellaoui A,
808   Andlauer TFM, Beekman ATF, Berger K, Blackwood DHR *et al*: **Genetic**
809   **effects influencing risk for major depressive disorder in China and**
810   **Europe**. *Translational Psychiatry* 2017, **7**:e1074 e1074.

811   14.   Veturi Y, de los Campos G, Yi NJ, Huang W, Vazquez AI, Kuhnel B:
812   **Modeling Heterogeneity in the Genetic Architecture of Ethnically Diverse**
813   **Groups Using Random Effect Interaction Models**. *Genetics* 2019,
814   **211**(4):1395-1407.

815   15.   Sirugo G, Williams SM, Tishkoff SA: **The Missing Diversity in Human**
816   **Genetic Studies**. *Cell* 2019, **177**(1):26-31.

817   16.   Popejoy AB, Fullerton SM: **Genomics is failing on diversity**. *Nature* 2016,
818   **538**(7624):161-164.

819   17.   Moltke I, Grarup N, Jorgensen ME, Bjerregaard P, Treebak JT, Fumagalli M,
820   Korneliussen TS, Andersen MA, Nielsen TS, Krarup NT *et al*: **A common**
821   **Greenlandic TBC1D4 variant confers muscle insulin resistance and type 2**
822   **diabetes**. *Nature* 2014, **512**(7513):190-193.

823   18.   Tang CS, Zhang H, Cheung CYY, Xu M, Ho JCY, Zhou W, Cherny SS, Zhang
824   Y, Holmen O, Au KW *et al*: **Exome-wide association analysis reveals novel**
825   **coding sequence variants associated with lipid traits in Chinese**. *Nat*
826   *Commun* 2015, **6**:10206.

827   19.   Clarke TK, Adams MJ, Davies G, Howard DM, Hall LS, Padmanabhan S,
828   Murray AD, Smith BH, Campbell A, Hayward C *et al*: **Genome-wide**
829   **association study of alcohol consumption and genetic overlap with other**
830   **health-related traits in UK Biobank (N=112117)**. *Mol Psychiatry* 2017,
831   **22**(10):1376-1384.

832   20.   Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A,
833   Adams MJ, Agerbo E, Air TM, Andlauer TMF *et al*: **Genome-wide**
834   **association analyses identify 44 risk variants and refine the genetic**
835   **architecture of major depression**. *Nat Genet* 2018, **50**(5):668-681.

836   21.   Cai N, Bigdeli TB, Kretzschmar W, Li YH, Liang JQ, Song L, Hu JC, Li QB,
837   Jin W, Hu ZF *et al*: **Sparse whole-genome sequencing identifies two loci for**
838   **major depressive disorder**. *Nature* 2015, **523**(7562):588-591.

839   22.   Wu Y, Waite LL, Jackson AU, Sheu WHH, Absher SBD, Absher D, Arnett DK,
840   Boerwinkle E, Bonnycastle LL, Carty CL *et al*: **Trans-Ethnic Fine-Mapping**
841   **of Lipid Loci Identifies Population-Specific Signals and Allelic**
842   **Heterogeneity That Increases the Trait Variance Explained**. *PLoS Genet*
843   2013, **9**(3):e1003379.

844   23.   Coram MA, Duan Q, Hoffmann TJ, Thornton T, Knowles JW, Johnson NA,
845   Ochs-Balcom HM, Donlon TA, Martin LW, Eaton CB *et al*: **Genome-wide**
846   **Characterization of Shared and Distinct Genetic Components that**
847   **Influence Blood Lipid Levels in Ethnically Diverse Human Populations**.
848   *Am J Hum Genet* 2013, **92**(6):904-916.

849   24.   Guo J, Wu Y, Zhu ZH, Zheng ZL, Trzaskowski M, Zeng J, Robinson MR,
850   Visscher PM, Yang J: **Global genetic differentiation of complex traits**

851   **shaped by natural selection in humans**. *Nat Commun* 2018, **9**:1865.

852   25.   Lu HJ, Wang T, Zhang JH, Zhang SO, Huang SP, Zeng P: **Evaluating**
853   **marginal genetic correlation of associated loci for complex diseases and**
854   **traits between European and East Asian populations**. *Hum Genet* 2021,
855   **140**(9):1285-1297.

856   26.   Galinsky KJ, Reshef YA, Finucane HK, Loh PR, Zaitlen N, Patterson NJ,
857   Brown BC, Price AL: **Estimating cross-population genetic correlations of**
858   **causal effect sizes**. *Genet Epidemiol* 2019, **43**(2):180-188.

859   27.   Andreassen OA, Djurovic S, Thompson WK, Schork AJ, Kendler KS,
860   O'Donovan MC, Rujescu D, Werge T, van de Bunt M, Morris AP *et al*:
861   **Improved Detection of Common Variants Associated with Schizophrenia**
862   **by Leveraging Pleiotropy with Cardiovascular-Disease Risk Factors**. *Am J*
863   *Hum Genet* 2013, **92**(2):197-209.

864   28.   Smeland OB, Frei O, Shadrin A, O'Connell K, Fan CC, Bahrami S, Holland D,
865   Djurovic S, Thompson WK, Dale AM *et al*: **Discovery of shared genomic**
866   **loci using the conditional false discovery rate approach**. *Hum Genet* 2020,
867   **139**(1):85-94.

868   29.   Choi SW, Mak TSH, O'Reilly PF: **Tutorial: a guide to performing polygenic**
869   **risk score analyses**. *Nat Protoc* 2020, **15**(9):2759-2772.

870   30.   Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, Daly
871   MJ, Price AL, Neale BM, Grp SW: **LD Score regression distinguishes**
872   **confounding from polygenicity in genome-wide association studies**. *Nat*
873   *Genet* 2015, **47**(3):291-295.

874   31.   Boyle EA, Li YI, Pritchard JK: **An Expanded View of Complex Traits:**
875   **From Polygenic to Omnigenic**. *Cell* 2017, **169**(7):1177-1186.

876   32.   Pathak PK, Tripathi N, Subramanian SV: **Secular Trends in Menarcheal Age**
877   **in India-Evidence from the Indian Human Development Survey**. *PLoS*
878   *ONE* 2014, **9**(11):e111027.

879   33.   Ikeda M, Takahashi A, Kamatani Y, Okahisa Y, Kunugi H, Mori N, Sasaki T,
880   Ohmori T, Okamoto Y, Kawasaki H *et al*: **A genome-wide association study**
881   **identifies two novel susceptibility loci and trans population polygenicity**
882   **associated with bipolar disorder**. *Mol Psychiatry* 2018, **23**(3):639-647.

883   34.   Huxley R, James WPT, Barzi F, Patel JV, Lear SA, Suriyawongpaisal P, Janus
884   E, Caterson I, Zimmet P, Prabhakaran D *et al*: **Ethnic comparisons of the**
885   **cross-sectional relationships between measures of body size with diabetes**
886   **and hypertension**. *Obes Rev* 2008, **9**:53-61.

887   35.   Spracklen CN, Horikoshi M, Kim YJ, Lin K, Bragg F, Moon S, Suzuki K, Tam
888   CHT, Tabara Y, Kwak SH *et al*: **Identification of type 2 diabetes loci in**
889   **433,540 East Asian individuals**. *Nature* 2020, **582**(7811):240-245.

890   36.   Virlogeux V, Graff RE, Hoffmann TJ, Witte JS: **Replication and Heritability**
891   **of Prostate Cancer Risk Variants: Impact of Population-Specific Factors**.
892   *Cancer Epidemiol Biomark Prev* 2015, **24**(6):938-943.

893   37.   Kaufman BP, Guttman-Yassky E, Alexis AF: **Atopic dermatitis in diverse**
894   **racial and ethnic groups-Variations in epidemiology, genetics, clinical**

895   **presentation and treatment**. *Exp Dermatol* 2018, **27**(4):340-357.

896   38.   Janumpally SR, Feldman SR, Gupta AK, Fleischer AB: **In the United States,**
897   **blacks and Asian/Pacific Islanders are more likely than whites to seek**
898   **medical care for atopic dermatitis**. *Arch Dermatol* 2002, **138**(5):634-637.

899   39.   Kim E, Lee JE, Namkung JH, Kim PS, Kim S, Shin ES, Cho EY, Yang JM:
900   **Single nucleotide polymorphisms and the haplotype in the DEFB1 gene**
901   **are associated with atopic dermatitis in a Korean population**. *J Dermatol*
902   *Sci* 2009, **54**(1):25-30.

903   40.   Wheeler E, Leong A, Liu CT, Hivert MF, Strawbridge RJ, Podmore C, Li M,
904   Yao J, Sim X, Hong JY *et al*: **Impact of common genetic determinants of**
905   **Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally**
906   **diverse populations: A transethnic genome-wide meta-analysis**. *PLoS Med*
907   2017, **14**(9):e1002383.

908   41.   Dastani Z, Johnson T, Kronenberg F, Nelson CP, Assimes TL, Marz W,
909   Richards JB, Consortium C, Consortium A: **The shared allelic architecture**
910   **of adiponectin levels and coronary artery disease**. *Atherosclerosis* 2013,
911   **229**(1):145-148.

912   42.   Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE,
913   Sekowska M, Smith GD, Evans D, Gutierrez-Arcelus M *et al*: **Patterns of Cis**
914   **Regulatory Variation in Diverse Human Populations**. *PLoS Genet* 2012,
915   **8**(4):272-284.

916   43.   Shi H, Burch KS, Johnson R, Freund MK, Kichaev G, Mancuso N, Manuel
917   AM, Dong N, Pasaniuc B: **Localizing Components of Shared Transethnic**
918   **Genetic Architecture of Complex Traits from GWAS Summary Data**.
919   *American journal of human genetics* 2020, **106**(6):805-817.

920   44.   Marigorta UM, Lao O, Casals F, Calafell F, Morcillo-Suarez C, Faria R, Bosch
921   E, Serra F, Bertranpetit J, Dopazo H *et al*: **Recent human evolution has**
922   **shaped geographical differences in susceptibility to disease**. *BMC*
923   *Genomics* 2011, **12**:55.

924   45.   Fu JY, Festen EAM, Wijmenga C: **Multi-ethnic studies in complex traits**.
925   *Hum Mol Genet* 2011, **20**:R206-R213.

926   46.   Evangelou E, Fellay J, Colombo S, Martinez-Picado J, Obel N, Goldstein DB,
927   Telenti A, Ioannidis JPA: **Impact of Phenotype Definition on Genome-Wide**
928   **Association Signals: Empirical Evaluation in Human Immunodeficiency**
929   **Virus Type 1 Infection**. *Am J Epidemiol* 2011, **173**(11):1336-1342.

930   47.   Baye TM, Abebe T, Wilke RA: **Genotype-environment interactions and**
931   **their translational implications**. *Personalized Medicine* 2011, **8**(1):59-70.

932   48.   HermanGiddens ME, Slora EJ, Wasserman RC, Bourdony CJ, Bhapkar MV,
933   Koch GG, Hasemeier CM: **Secondary sexual characteristics and menses in**
934   **young girls seen in office practice: A study from the pediatric research in**
935   **office settings network**. *Pediatrics* 1997, **99**(4):505-512.

936   49.   Anderson SE, Dallal GE, Must A: **Relative weight and race influence**
937   **average age at menarche: Results from two nationally representative**
938   **surveys of US girls studied 25 years apart**. *Pediatrics* 2003, **111**(4):844-850.

939 50. Chumlea WC, Schubert CM, Roche AF, Kulin HE, Lee PA, Himes JH, Sun SS:
940   **Age at menarche and racial comparisons in US girls**. *Pediatrics* 2003,
941   **111**(1):110-113.

942 51. Sarnowski C, Cousminer DL, Franceschini N, Raffield LM, Jia G, Fernández-
943   Rhodes L, Grant SFA, Hakonarson H, Lange LA, Long J *et al*: **Large trans-**
944   **ethnic meta-analysis identifies AKR1C4 as a novel gene associated with**
945   **age at menarche**. *Hum Reprod* 2021, **36**(7):1999-2010.

946 52. Dvornyk V, Waqar-ul-Haq: **Genetics of age at menarche: a systematic**
947   **review**. *Hum Reprod Update* 2012, **18**(2):198-210.

948 53. Beavan S, Prentice A, Dibba B, Yan L, Cooper C, Ralston SH: **Polymorphism**
949   **of the collagen type Ialpha1 gene and ethnic differences in hip-fracture**
950   **rates**. *N Engl J Med* 1998, **339**(5):351-352.

951 54. Dvornyk V, Liu XH, Shen H, Lei SF, Zhao LJ, Huang QR, Qin YJ, Jiang DK,
952   Long JR, Zhang YY *et al*: **Differentiation of Caucasians and Chinese at**
953   **bone mass candidate genes: Implication for ethnic difference of bone mass**.
954   *Ann Hum Genet* 2003, **67**:216-227.

955 55. Lei SF, Deng FY, Liu XH, Huang QR, Qin YJ, Zhou Q, Jiang DK, Li YM, Mo
956   XY, Liu MY *et al*: **Polymorphisms of four bone mineral density candidate**
957   **genes in Chinese populations and comparison with other populations of**
958   **different ethnicity**. *J Bone Miner Metab* 2003, **21**(1):34-42.

959 56. Luzzatto L, Usanga FA, Reddy S: **Glucose-6-phosphate dehydrogenase**
960   **deficient red cells: resistance to infection by malarial parasites**. *Science*
961   1969, **164**(3881):839-842.

962 57. Wojcik GL, Graff M, Nishimura KK, Tao R, Haessler J, Gignoux CR,
963   Highland HM, Patel YM, Sorokin EP, Avery CL *et al*: **Genetic analyses of**
964   **diverse populations improves discovery for complex traits**. *Nature* 2019,
965   **570**(7762):514-518.

966 58. Guenther CA, Tasic B, Luo L, Bedell MA, Kingsley DM: **A molecular basis**
967   **for classic blond hair color in Europeans**. *Nature Genetics* 2014, **46**(7):748-
968   752.

969 59. Zaitlen N, Pasaniuc B, Gur T, Ziv E, Halperin E: **Leveraging Genetic**
970   **Variability across Populations for the Identification of Causal Variants**.
971   *Am J Hum Genet* 2010, **86**(1):23-33.

972 60. Morris AP: **Transethnic meta-analysis of genomewide association studies**.
973   *Genet Epidemiol* 2011, **35**(8):809-822.

974 61. Coram MA, Fang HY, Candille SI, Assimes TL, Tang H: **Leveraging Multi-**
975   **ethnic Evidence for Risk Assessment of Quantitative Traits in Minority**
976   **Populations**. *Am J Hum Genet* 2017, **101**(2):218-226.

977 62. Spracklen CN, Chen P, Kim YJ, Wang X, Cai H, Li SX, Long JR, Wu Y, Wang
978   YX, Takeuchi F *et al*: **Association analyses of East Asian individuals and**
979   **trans-ancestry analyses with European individuals reveal new loci**
980   **associated with cholesterol and triglyceride levels**. *Hum Mol Genet* 2017,
981   **26**(9):1770-1784.

982 63. Coram MA, Candille SI, Duan Q, Chan KHK, Li Y, Kooperberg C, Reiner AP,

983   Tang H: **Leveraging Multi-ethnic Evidence for Mapping Complex Traits in**
984   **Minority Populations: An Empirical Bayes Approach**. *Am J Hum Genet*
985   2015, **96**(5):740-752.

986 64. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA, Jankovic I, Boehnke M:
987   **Genome-wide association studies in diverse populations**. *Nat Rev Genet*
988   2010, **11**(5):356-366.

989 65. Bustamante CD, Burchard EG, De La Vega FM: **Genomics for the world**.
990   *Nature* 2011, **475**(7355):163-165.

991 66. Teo YY, Small KS, Kwiatkowski DP: **Methodological challenges of genome-**
992   **wide association analysis in Africa**. *Nat Rev Genet* 2010, **11**(2):149-160.

993 67. Stein CM: **Challenges of Genetic Data Sharing in African Studies**. *Trends*
994   *Genet* 2020, **36**(12):895-896.

995 68. Li H, Durbin R: **Inference of human population history from individual**
996   **whole-genome sequences**. *Nature* 2011, **475**(7357):493-U484.

997 69. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, Duncan
998   L, Perry JRB, Patterson N, Robinson EB *et al*: **An atlas of genetic**
999   **correlations across human diseases and traits**. *Nat Genet* 2015,
1000   **47**(11):1236-1241.

1001 70. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR,
1002   Anttila V, Xu H, Zang CZ, Farh K *et al*: **Partitioning heritability by**
1003   **functional annotation using genome-wide association summary statistics**.
1004   *Nat Genet* 2015, **47**(11):1228-1235.

1005 71. Falconer D, Mackay TJULG: **Introduction to quantitative genetics. Essex**.
1006   1996.

1007 72. Zeng P, Hao XJ, Zhou X: **Pleiotropic mapping and annotation selection in**
1008   **genome-wide association studies with penalized Gaussian mixture models**.
1009   *Bioinformatics* 2018, **34**(16):2797-2807.

1010 73. Ray D, Chatterjee N: **A powerful method for pleiotropic analysis under**
1011   **composite null hypothesis identifies novel shared loci between Type 2**
1012   **Diabetes and Prostate Cancer**. *PLoS Genet* 2020, **16**(12):e1009218.

1013 74. Wang T, Lu H, Zeng P: **Identifying pleiotropic genes for complex**
1014   **phenotypes with summary statistics from a perspective of composite null**
1015   **hypothesis testing**. *Brief Bioinform* 2021.

1016 75. Efron B: **Size, power and false discovery rates**. *Annals of Statistics* 2007,
1017   **35**(4):1351-1377.

1018 76. Wang X, Chua HX, Chen P, Ong RTH, Sim XL, Zhang WH, Takeuchi F, Liu
1019   XY, Khor CC, Tay WT *et al*: **Comparing methods for performing trans-**
1020   **ethnic meta-analysis of genome-wide association studies**. *Hum Mol Genet*
1021   2013, **22**(11):2303-2311.

1022 77. Keller MF, Reiner AP, Okada Y, van Rooij FJA, Johnson AD, Chen MH,
1023   Smith AV, Morris AP, Tanaka T, Ferrucci L *et al*: **Trans-ethnic meta-analysis**
1024   **of white blood cell phenotypes**. *Hum Mol Genet* 2014, **23**(25):6944-6960.

1025 78. Iyengar SK, Sedor JR, Freedman BI, Kao WHL, Kretzler M, Keller BJ,
1026   Abboud HE, Adler SG, Best LG, Bowden DW *et al*: **Genome-Wide**

Association and Trans-ethnic Meta-Analysis for Advanced Diabetic Kidney Disease: Family Investigation of Nephropathy and Diabetes (FIND). *PLoS Genet* 2015, **11**(8):e1005352.

79. Tang CSM, Gui HS, Kapoor A, Kim JH, Luzon-Toro B, Pelet A, Burzynski G, Lantieri F, So MT, Berrios C *et al*: **Trans-ethnic meta-analysis of genome-wide association studies for Hirschsprung disease**. *Hum Mol Genet* 2016, **25**(23):5265-5275.

80. Magi R, Horikoshi M, Sofer T, Mahajan A, Kitajima H, Franceschini N, McCarthy MI, Morris AP, Consortium C-K, Consortium TD-G: **Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution**. *Hum Mol Genet* 2017, **26**(18):3639-3650.

81. van Rooij FJA, Qayyum R, Smith AV, Zhou Y, Trompet S, Tanaka T, Keller MF, Chang LC, Schmidt H, Yang ML *et al*: **Genome-wide Trans-ethnic Meta-analysis Identifies Seven Genetic Loci Influencing Erythrocyte Traits and a Role for RBPMS in Erythropoiesis**. *Am J Hum Genet* 2017, **100**(1):51-63.

82. Li CY, Yang TM, Ou RW, Wei QQ, Shang HF: **Genome-wide genetic links between amyotrophic lateral sclerosis and autoimmune diseases**. *BMC Med* 2021, **19**(1):27.

83. Lin DY, Sullivan PF: **Meta-Analysis of Genome-wide Association Studies with Overlapping Subjects**. *Am J Hum Genet* 2009, **85**(6):862-872.

84. LeBlanc M, Zuber V, Thompson WK, Andreassen OA, Frigessi A, Andreassen BK, Consortium PG: **A correction for sample overlap in genome-wide association studies in a polygenic pleiotropy-informed framework**. *BMC Genomics* 2018, **19**:494.

85. Chung DJ, Yang C, Li C, Gelernter J, Zhao HY: **GPA: A Statistical Approach to Prioritizing GWAS Results by Integrating Pleiotropy and Annotation**. *PLoS Genet* 2014, **10**(11):e1004787.

86. Charles EP: **The correction for attenuation due to measurement error: clarifying concepts and creating confidence sets**. *Psychological methods* 2005, **10**(2):206-226.

87. Buonaccorsi JP: **Measurement Error: Models, Methods and Applications**: Measurement Error: Models, Methods and Applications; 2010.

88. Lewontin RC, Krakauer J: **Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms**. *Genetics* 1973, **74**(1):175-195.

89. Holsinger KE, Weir BS: **Genetics in geographically structured populations: defining, estimating and interpreting F(ST)**. *Nat Rev Genet* 2009, **10**(9):639-650.

90. Zhu ZH, Zhang FT, Hu H, Bakshi A, Robinson MR, Powell JE, Montgomery GW, Goddard ME, Wray NR, Visscher PM *et al*: **Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets**. *Nat Genet* 2016, **48**(5):481-487.

1071

1072