Supplementary Note

Genome-wide meta-analysis of cognitive empathy: heritability, and correlates with sex, neuropsychiatric conditions and cognition

1. Participant information and phenotyping

All participants from the BLTS completed one of two versions of the short-version of the Eyes Test. 580 participants completed V1 (July 2008 to December 2009) of the short version of the Eyes Test, and 1141 participants completed V2 (July 2010 to November 2011) of the short version of the Eyes Test, totalling 1716 participants. Of these, 127 participants were not included in final analysis as they were not genotyped, 47 were not included as they were of non-Caucasian ancestry, and 45 were not included as they were missing data on the age covariate. 259 participants completed both V1 and V2 of the short version of the test. For these participants, we used scores from V1 of the test for the analysis to avoid a learning bias.

14 questions were common to both the versions, and so, the final short version of the Eyes Test had only these 14 questions (Table 1). In V1 of the test, participants had to choose the right answer from four different options describing various mental states. In V2, an additional 'don't know' option was provided as the fifth option. Both the images and the four options describing various mental states were the same across all three tests (complete Eyes Test, short Eyes Test V1, short Eyes Test V2). Scores on the short-version of the Eyes Test were unimodally and near-normally distributed. We visually inspected both the frequency histogram and the quantile-quantile plot (Figure 1) to determine the normalcy of the distribution. In addition, the measure of skewness (-0.44) and excess kurtosis (0.068) were within the acceptable range of ± 1 of a normal distribution.

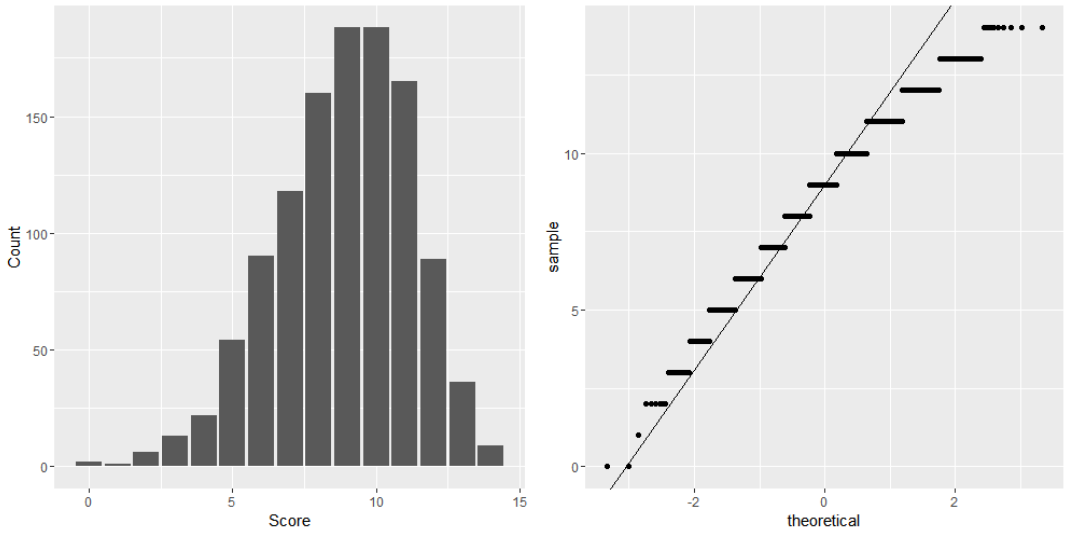
Table 1: Questions used in the three different versions of the Eyes Test

Question Number	Option 1	Option 2	Option 3	Option 4	Sex of the actor	Full version	Short V1	Short V2[#]	% correct[1]
example	jealous	panicked	arrogant	hateful	M	example	X		
1	**playful**	comforting	irritated	bored	M	X			
2	terrified	**upset**	arrogant	annoyed	M	X	X	X	71
3	joking	flustered	**desire**	convinced	F	X			
4	joking	**insisting**	amused	relaxed	M	X	X	X	69
5	irritated	sarcastic	**worried**	friendly	M	X			
6	aghast	**fantasizing**	impatient	alarmed	F	X	X		
7	apologetic	friendly	**uneasy**	dispirited	M	X			
8	**despondent**	relieved	shy	excited	M	X			
9	annoyed	hostile	horrified	**preoccupied**	F	X			
10	**cautious**	insisting	bored	aghast	M	X			
11	terrified	amused	**regretful**	flirtatious	M	X			
12	indifferent	embarrassed	**sceptical**	dispirited	M	X			
13	decisive	**anticipating**	threatening	shy	M	X			
14	irritated	disappointed	depressed	**accusing**	M	X	X	X	55
15	**contemplative**	flustered	encouraging	amused	F	X			
16	irritated	**thoughtful**	encouraging	sympathetic	M	X	X	X	83
17	**doubtful**	affectionate	playful	aghast	F	X			
18	**decisive**	amused	aghast	bored	F	X	X	X	61
19	arrogant	grateful	sarcastic	**tentative**	F	X			
20	dominant	**friendly**	guilty	horrified	M	X			
21	embarrassed	**fantasizing**	confused	panicked	F	X	X	X	78
22	**preoccupied**	grateful	insisting	imploring	F	X			
23	contented	apologetic	**defiant**	curious	M	X	X	X	39
24	**pensive**	irritated	excited	hostile	M	X			
25	panicked	incredulous	despondent	**interested**	F	X			
26	alarmed	shy	**hostile**	anxious	M	X	X	X	58
27	joking	**cautious**	arrogant	reassuring	F	X	X		
28*	**interested**	joking	affectionate	contented	F	X	X	X	57
29	impatient	aghast	irritated	**reflective**	F	X	X	X	68
30	grateful	**flirtatious**	hostile	disappointed	F	X			
31	ashamed	**confident**	joking	dispirited	F	X			
32	**serious**	ashamed	bewildered	alarmed	M	X	X	X	72
33	embarrassed	guilty	fantasizing	**concerned**	M	X			
34	aghast	baffled	**distrustful**	terrified	F	X	X	X	61
35	puzzled	**nervous**	insisting	contemplative	F	X	X	X	60
36	ashamed	nervous	**suspicious**	indecisive	M	X	X	X	75

[1] Number of participants who chose the correct option in the BLTS Cohort. Data unavailable for the 23andMe cohort.

* The second option changed to 'insisting' in V1 and V2 of the short version of the test. #In addition, all questions in the V2 of the short version of the test had an additional 'Don't Know' option.

Figure 1: Frequency histogram (left) and Quantile-quantile plot of the scores on the short version (V2) of the Eyes Test.



We tested the properties of the short Eyes Test using data from 259 participants who had completed both versions of the Eyes Test. There was a significant increase in mean scores between V1 and V2 version of the test (8.84, sd = 2.06 and 9.55, sd = 2.07, $P < 0.001$; paired two-sided T-test; Cohen's d = 0.34 for V2 vs V1), indicating an advantageous effect of repeated testing on the scores. We cannot completely discount that the presence of the fifth option 'don't know' may have facilitated the test. We next performed tetrachoric correlation on individual item scores to look at test-retest reliability. For the overall short Eyes Test (14 common items), there was a modest correlation of 0.47. Three factors must be taken into consideration whilst interpreting this correlation. First, there was a gap of nearly two years between the two waves of testing. Second, the two versions of the Eyes Test are not identical and thereby do not facilitate direct comparison. Third, as mentioned above, the mean score on the V2 of the test was significantly higher than the mean score on V1 of the test.

We also investigated sex-difference in the short Eyes Test. On average, women scored significantly higher than men (8.99, sd = 2.30 and 8.66, sd = 2.41; P = 0.019; Cohen's d = 0.152). As this included participants on both versions of the short test, we also checked if there was a significant difference in the sex-ratio between V1 and V2 of the test, to account for the potential facilitation effect seen in V2. A chi-square test showed that there was no significant difference in the number of male and female participants between the two versions of the test (chi square = 0.73; two-tailed P = 0.39). The ratio of males to females was the same in both the versions of the test (0.66).

We next calculated if the valence of the items was significantly different between the two versions of the test (full version and short version). We divided all the items into three different valences : Positive, negative, and neutral[1]. In the full version of the test, there were 8 positive items, 12 negative items, and 16 neutral items. In the short version of the test, there were 3 positive items, 7 negative items, and 4 neutral items, indicating an excess of negative items and a deficit of neutral items. A chi-square test did not indicate that there was a significant difference in the valence of the two versions of the test. However, we cannot completely rule out that the difference in valences between the two traits can affect the genetic architecture between the two different GWAS datasets.

We next investigated the correlation between the short Eyes Test and the full Eyes Test (adult version). To do this, we used data from control participants from the Cambridge Autism Research Database (CARD). We identified individuals who had completed the full version of the Eyes Test, did not indicate that they had a psychiatric diagnosis, did not have anyone in the immediate family (parents, siblings, and children) with an autism diagnosis, and were above fifteen years of age. We excluded participants who had more than 3 missing answers (i.e. > 10% missing). In total, we had 855 participants who met our criteria (276 males and 579 females). Participant ages ranged from 16 – 81 years. For each participant, using data from the same test, calculated two sets of scores: a score using all 36 questions (full Eyes) and a score for 14 questions found in the short Eyes Test (short Eyes). There was a highly significant correlation between the two scores (r = 0.77; P < 0.001). Ideally, participants should complete two different versions of the Eyes Test, and the correlation must be calculated between the two different versions. However, we did not have access to these data. There was a unimodal and near-normal distribution for both the datasets as measured using visual inspection of frequency histograms and quantile-quantile plots.

We finally investigated the similarity between the two GWAS datasets (BLTS and 23andMe), by investigating the direction of effects for all independent nominally significant ($P < 0.05$) SNPs. We calculated the proportion of SNPs with concordant effect direction in the two datasets in the stratified and the non-stratified GWAS datasets, and quantified the significance using 1-sided binomial sign test. For the non-stratified analyses, 65% of the SNPs had a concordant effect direction, 66% for the males-only analyses, and 70% for the females-only analyses. All sign tests were significant ($P < 2.2 \times 10^{-16}$ for all three binomial sign tests)

2. Genetic correlations

We performed genetic correlation using non-stratified GWAS data from the 23andMe cohort, so as to keep the phenotype measure homogenous across the entire sample. For genetic correlations, we used summary GWAS data for schizophrenia, bipolar disorder[2], autism[2], anorexia[3], anxiety[4], and depression[2], that were downloaded from the Psychiatric Genomics Consortium website (http://www.med.unc.edu/pgc/downloads). Summary GWAS data for educational attainment measured through number of college years[5], educational attainment[6], and cognitive aptitude[7] were downloaded from the Social Science Genetic Association Consortium website (http://ssgac.org/Data.php). Cognitive aptitude is measured independent of knowledge of facts and words. Though the mental-state words provided in the Eyes Test are fairly common, we cannot completely discount the fact that word-knowledge may facilitate better performance on the test. Summary GWAS data for personality traits[8–10] were downloaded from the Genetics of Personality Consortium website: http://www.tweelingenregister.org/GPC/. Data for subcortical brain volumes[11] were downloaded from the ENIGMA consortium website (http://enigma.ini.usc.edu/download-enigma-gwas-results/). We did not include data for amygdala volume in the analysis due to non-significant heritability estimates using LDSC. For all the subcortical volume dataset, we first calculated the Z scores from the regression co-efficient and the standard errors, and used the Z scores to calculate genetic correlation. In addition, we used data for empathy measured using the Empathy Quotient, and data for Systemizing measured using the Systemizing Quotient from 23andMe, Inc. Data for the Borderline Personality Features GWAS were obtained from the authors of the paper[12].

3. Sex-difference enrichment analysis

3a. Methods

For sex-difference analysis, we ran MetaXcan on the sex-stratified analyses only for the cortical tissues. We focussed on the cortical tissue as it was relevant for the trait investigated and we had access to the list of sex-differentially expressed genes only from the cortex[13]. To check for overlap, we ran hypergeometric tests.

Overlap between sexes: First, to identify overlap between the sexes for the trait, we identified nominally significant genes ($P < 0.05$) in the two sexes separately and checked for overlap among these lists after pruning the background gene-lists to a common set of genes for both the sexes. We used a program available online to calculate both the overlap and the P-value of the overlap available here: http://nemates.org/MA/progs/overlap_stats.html. The test performed is a normal approximation of an exact hypergeometric test.

Hypergeometric tests are usually performed using 4 different lists. Let

a = list of genes in set a;

b = list of genes in set b;

x = list of overlapping genes in sets a and b, i.e. a intersection b, and

n = total list of background genes (note, this is different and usually larger than a union b).

Sets a, b, and x must be subsets of set n.

To identify overlap between sexes, a was the number of nominally significant genes in males-only Eyes Test GWAMA identified using MetaXcan, b was the number of nominally significant genes in the females-only Eyes Test GWAMA identified using MetaXcan, x was the overlapping genes between the two sets, and n was the set of common genes in the gene-based analyses of the males-only GWAMA and the gene based-analysis of the females-only GWAMA.

Sex-differentially expressed enrichment analyses: We performed hypergeometric tests to investigate if nominally significant genes for the Eyes Test in the sex-stratified GWAS are enriched for sex-differentially expressed genes. We wanted to check if genes that are nominally significant for the sex-stratified GWAMA of the Eyes Test are significantly enriched for genes that have sex-differential expression in the cortex.

To conduct this analysis, we first used MetaXcan to conduct gene-based association for the two sex-stratified GWAS using tissue weights from the cortex tissue in the GTEx project. This generated a list of 5951 genes with P-values for the males-only GWAS and 6071 genes with P-values for the females only GWAS. We also used a list of sex-differentially expressed genes identified in the Cortex from Werling et al., 2016. We included only autosomal genes with a fold-difference >1, regardless of the P-value. We identified a list of common genes that were identified by both MetaXcan and were investigated in Werling et al., 2016, *and this common set of genes were used as the background gene list n.* From this list of n, we defined sets a, b, and x.

For set a, we used genes with $P < 0.05$ in the gene-based association using MetaXcan.

For set b, using a list of sex-differentially expressed genes identified in the Cortex from Werling et al., 2016, we identified all genes with a fold-difference of greater than 1.

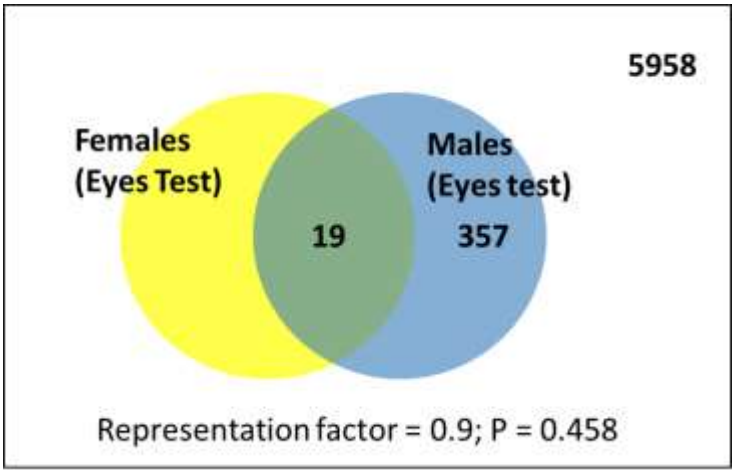
x was the intersection between sets a and b.

We performed four different enrichment analyses (Eyes Test-male: male-expressed; Eyes Test-female: male-expressed, Eyes Test-male: female-expressed; and Eyes Test-female: male-expressed), and used a P-value threshold of 0.025 to identify any significant enrichment.

3b. Results

Here, we present the result of the enrichment analysis. First, for the overlap in top genes between males and females for the Eyes Test, we identified all nominally significant genes ($P < 0.05$) in the cortex using MetaXcan[14]. To identify the background gene set, we overlapped all the genes for males and females after filtering out genes whose correlation with predicted models of expression was < 0.01 and where there were zero SNPs from our dataset. Results are shown in Figure 2.

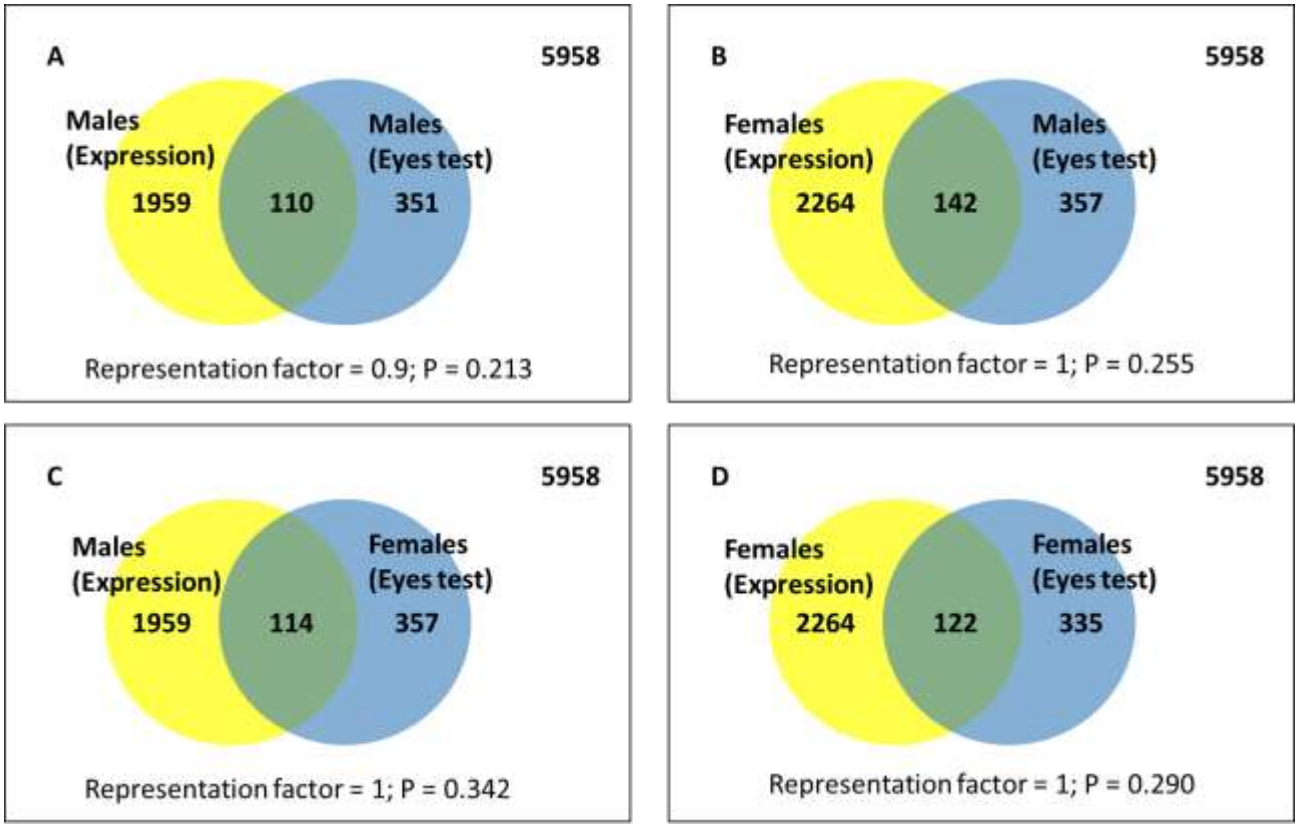
Figure 2: Overlap of top genes in males and females (Eyes Test)



We identified cortical genes associated with the trait using MetaXcan for each sex using the sex-stratified GWAS and compared the number of nominally significant genes that are common to both the sexes. Total number of genes is 4738, yellow circle represents the nominally significant genes in Females, blue circle represents the nominally significant genes in males. The overlap is given in the shaded green portion in the middle. Fold difference = 1.2; P = 0.264.

For the overlap in sex-differentially expressed genes, we used the discovery dataset from Werling et al. (2016)[13], which was from the BrainSpan project. For the background gene set, we used all the genes identified in MetaXcan for the Eyes Test (males or females), as we reasoned that all known genes were covered in the RNA sequence analysis of human cortical tissues in the BrainSpan project. We used only cortical gene-expression from the MetaXcan results. In total, we conducted 4 separate enrichment analysis: Eyes Test-male: male-expressed; Eyes Test-female: male-expressed, Eyes Test-male: female-expressed; and Eyes Test-female: male-expressed. Figure 3 provides the results of the enrichment analyses. We used a P-value threshold of 0.025 (0.05/2) to account for two different tests performed for each Eyes Test dataset.

Figure 3: Sex-difference enrichment analyses.



A - Eyes Test-male: male-expressed; B - Eyes Test-male: female-expressed; C -Eyes Test-female: male-expressed, and D - Eyes Test-female: male-expressed. Background gene numbers are provided in the white box. Overlap is provided in the green overlapping space.

4. Twin heritability

We calculated twin heritability using twin pairs from the BLTS cohort. For this subsample of the BLTS, twin ages at the time of testing ranged from 18 to 31 years (M = 25.3, SD = 3.0). As described in Supplementary Note 1, some twins completed the Eyes Test twice, for these participants only their first attempt was included in analyses. The distribution of the Eyes Test data was normal; 3 univariate outliers (< -3 SD) were excluded, and there were no bivariate outliers within each zygosity group. In total, data were available for 749 twin individuals, including 122 complete monozygotic twin pairs (74 female, 48 male), and 176 complete dizygotic twin pairs (60 female, 33 male, and 83 opposite sex pairs) plus 149 unpaired individuals whose responses nevertheless strengthen estimates of mean and variances. MZ

correlation (r = .31) was more than twice the DZ correlation (r = .09), which suggests an ADE model would fit the data better than an ACE model. Structural equation models were fit to raw data using full information maximum likelihood estimation in OpenMx. A series of nested models indicated that means and variances could be equated across females and males, and MZ and DZ twins. Although age and sex could be dropped as covariates on the means without a significant loss of model fit, they were retained for consistency with the GWAS analyses and to reduce possible bias in parameter estimates. As ACE and ADE models are not nested, they were compared using Akaike's Information Criterion (AIC), and nested submodels (AC, AE, and E) were tested using the likelihood-ratio test (LRT). The AE model was the best fitting model, although the CE model for familial aggregation could not be formally rejected (p=0.07); model fitting statistics are reported in Table 2(a) and standardised parameter estimates for this model, and for the ACE and ADE models are shown in Table 2(b) along with their 95% confidence intervals. In this small sample, although there was ample power to detect genetic effects in a reduced model, there was low power in the full model to estimate C or D in the presence of A (or vice versa). Thus, while the total genetic variance is around 30% in all three models, the upper 95% confidence limit for C from the ACE model indicates that C could account for as much as 22% of variance.

Our results are similar to predicted heritability estimates based on a previous meta-analysis of twin studies of empathy and prosocial behaviour[15].

Table 2(a): Heritability analyses of the Eyes test short version using the BLTS cohort (14 items).

Model No.	Model Type	AIC	-2LL	df	Models Compared	LRT		
						Δ-2LL	Δdf	p
1	ADE	1702.025	3188.025	743	-	-	-	-
2	ACE	1702.632	3188.632	743	-	-	-	-
3	CE	1703.949	3191.949	744	3 vs. 2	3.317	1	0.069
4	AE	1700.632	3188.632	744	4 vs. 1	0.607	1	0.436
5	E	1711.615	3201.615	745	5 vs. 1	13.590	2	0.001

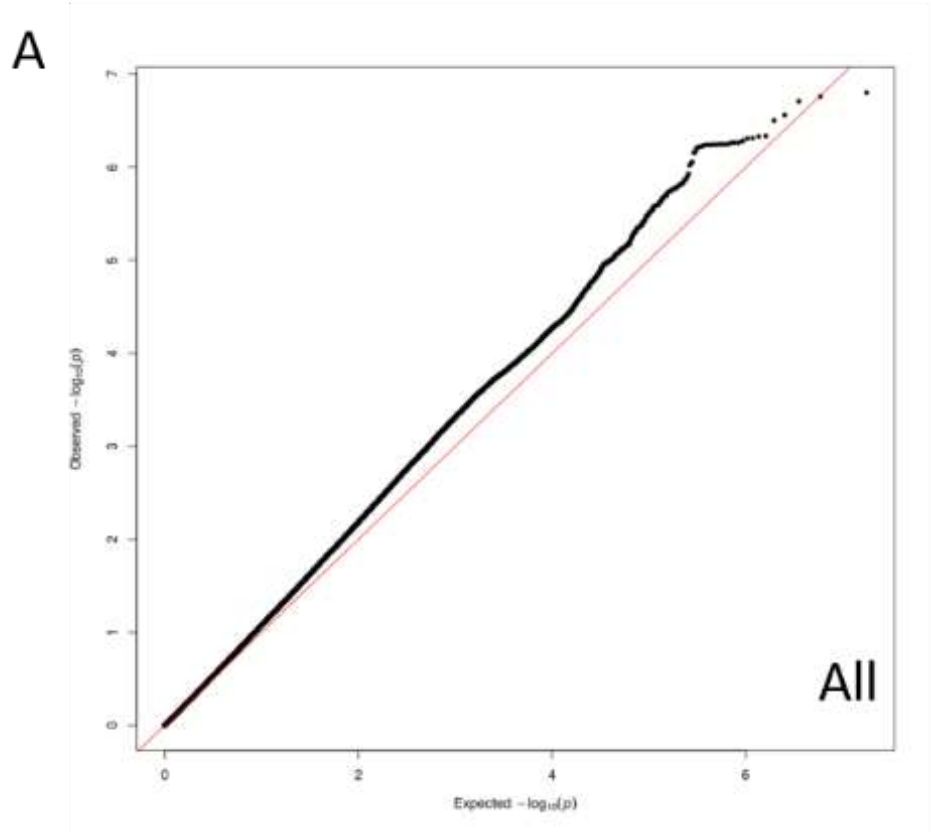
Table 2(b): Standardised variance components with 95% CIs for the ACE, ADE, and AE models

Model No.	Model Type	A Variance (95% CI)	D Variance (95% CI)	C Variance (95% CI)	E Variance (95% CI)
1	ACE	0.28 (0 - 0.42)	NA	0 (0 - 0.22)	0.72 (0.58 - 0.87)
2	ADE	0.05 (0 - 0.41)	0.26 (0 - 0.46)	NA	0.69 (0.54 - 0.86)
3	AE	0.28 (0.13 - 0.42)	NA	NA	0.72 (0.58 - 0.87)

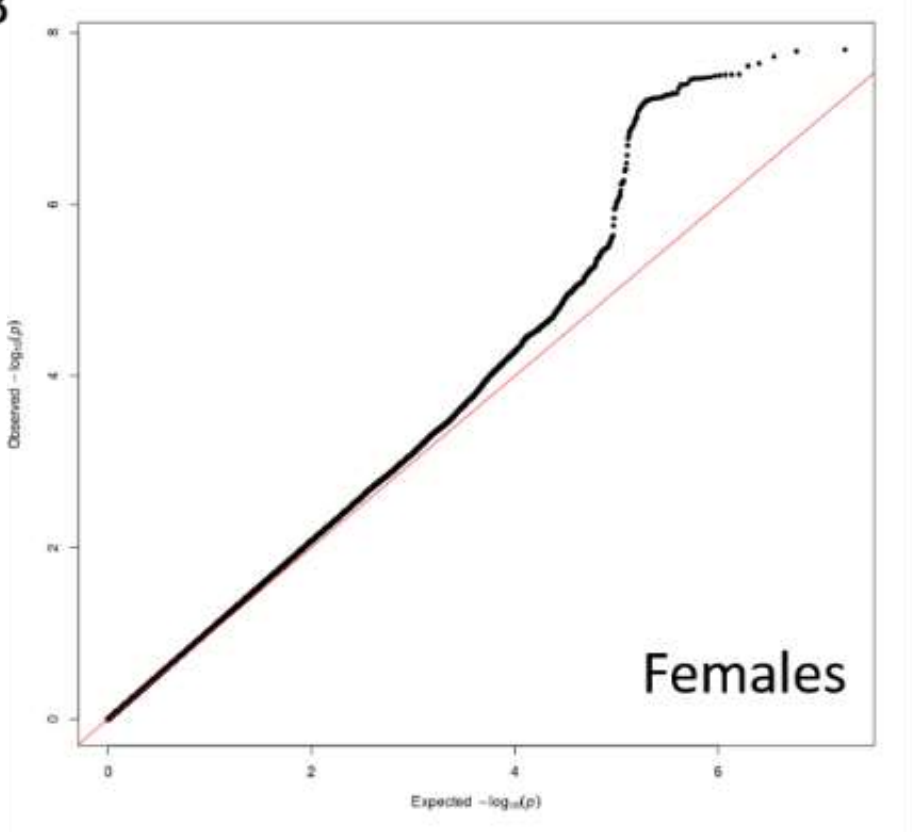
ADE and ACE are not nested models and were compared using AIC, where a lower value indicates a better fitting model. AE, CE, and E models are nested within the ADE and/or ACE models and were compared to the fuller model with the likelihood ratio test (LRT). A non-significant p-value from the LRT indicates the submodel is an acceptable fit to the data.. Although D could be dropped from the ADE model without a significant loss of fit, A and D could not, indicating significant genetic effects on variation in the Eyes Test.

4. Manhattan and QQ-plots of the GWAS

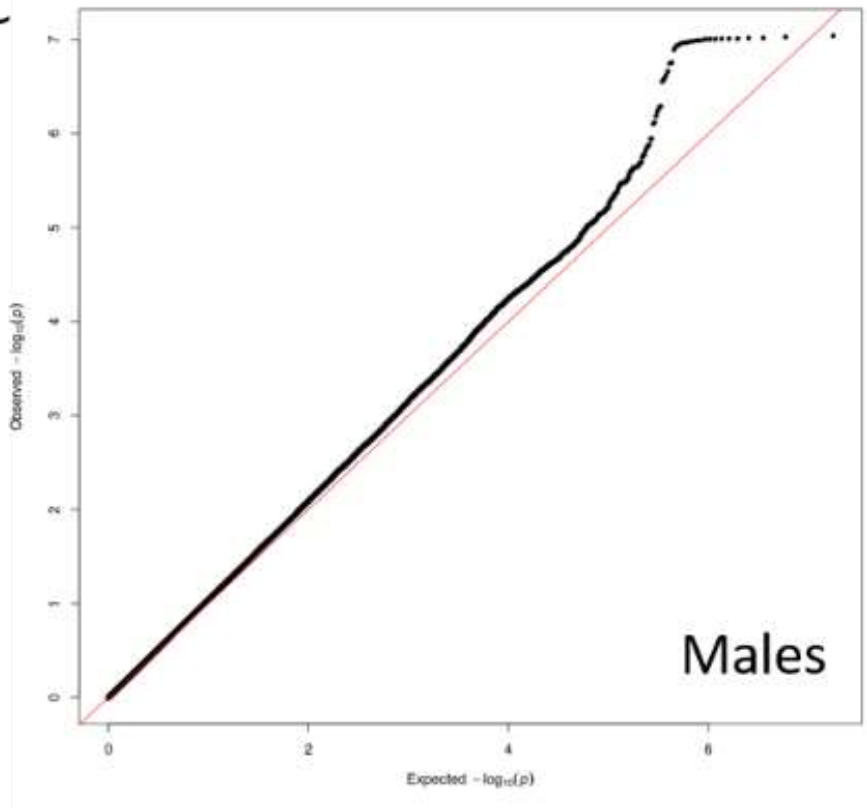
Figure 4: Quantile-quantile plot for the Eyes Test meta-analysis



B

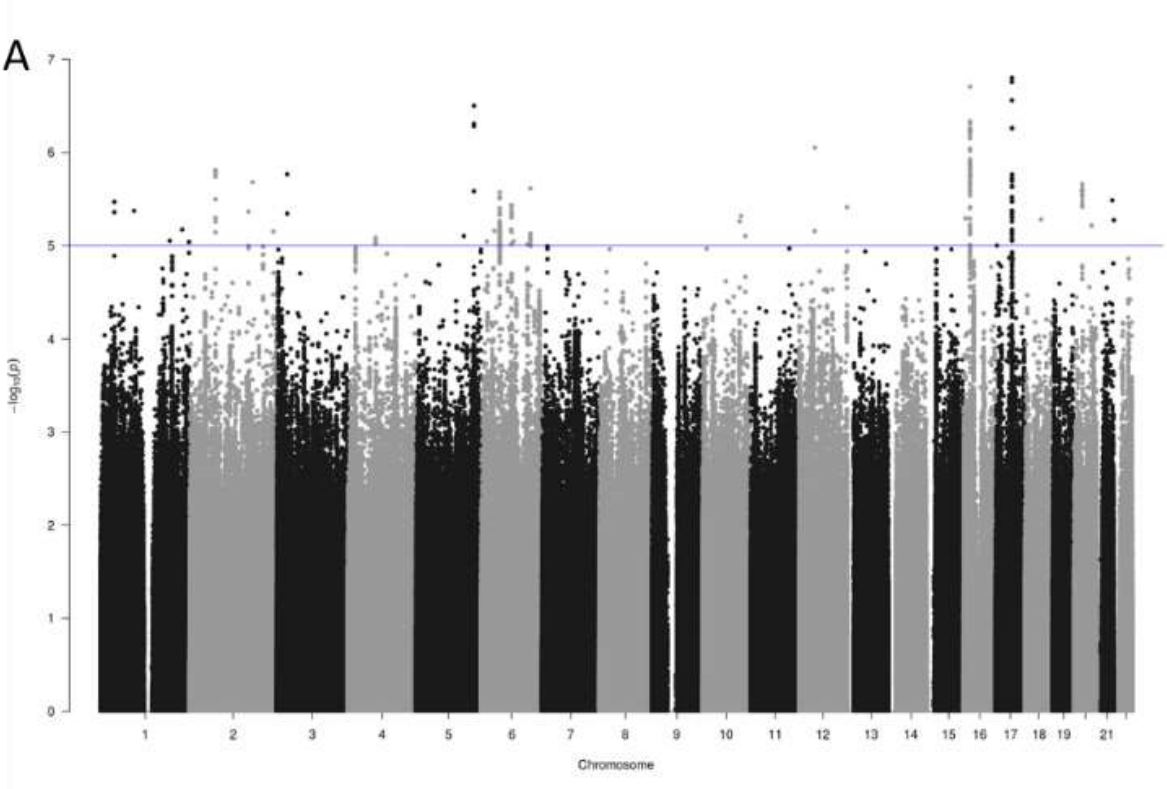

Females
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
0
2
4
6
8

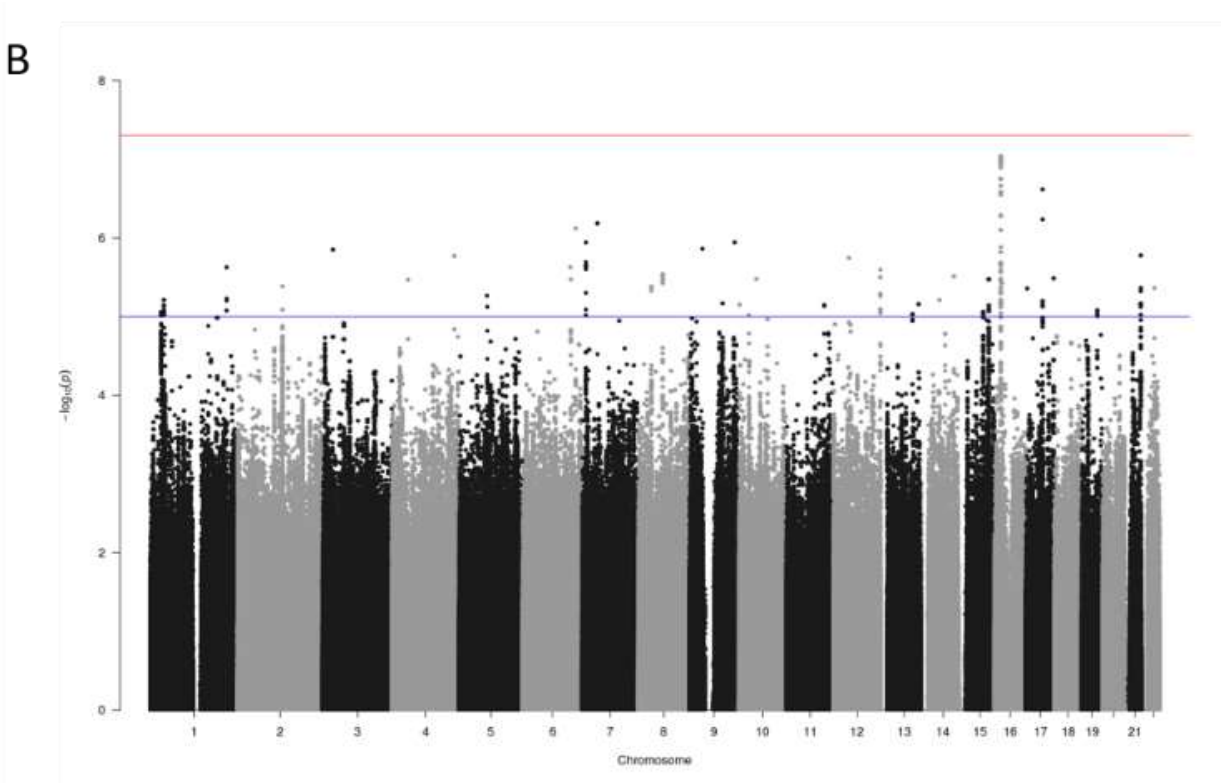
C


Males
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$
0
1
2
3
4
5
6
7

Quantile-quantile plots for the non-stratified GWAS (A), the females-only GWAS (B), and the males-only GWAS (C). n = 89553, λ_{gc} = 1.089, LDS intercept = 1.01 and for the non-stratified GWAS. n = 44,574, λ_{gc} = 1.05, LDS intercept = 1.005 for the females only gwas. n = 44088, λ_{gc} = 1.06, LDS intercept = 1.006 for the males-only GWAS.

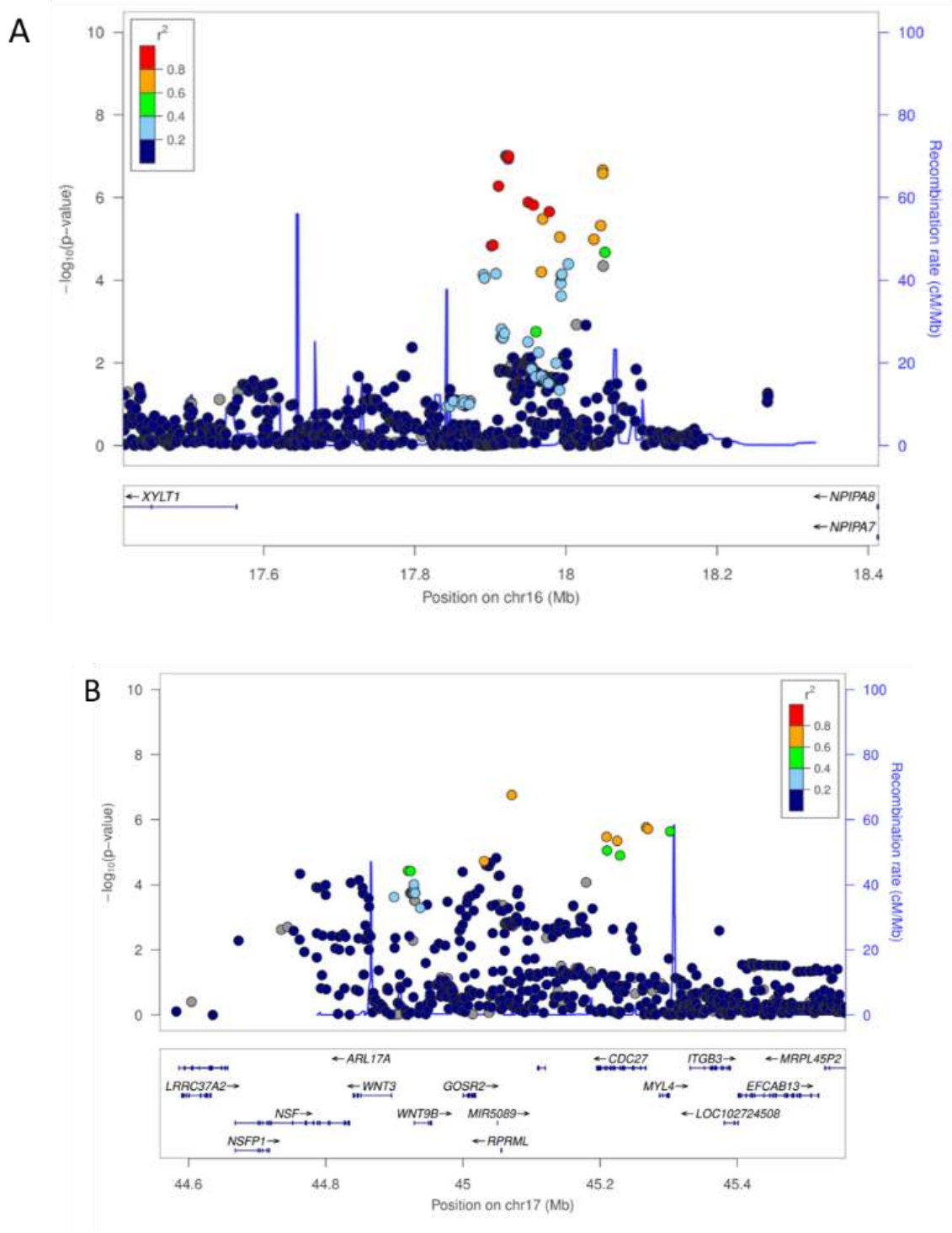
Figure 5: Manhattan plot of Eyes Test (all) meta-analysis





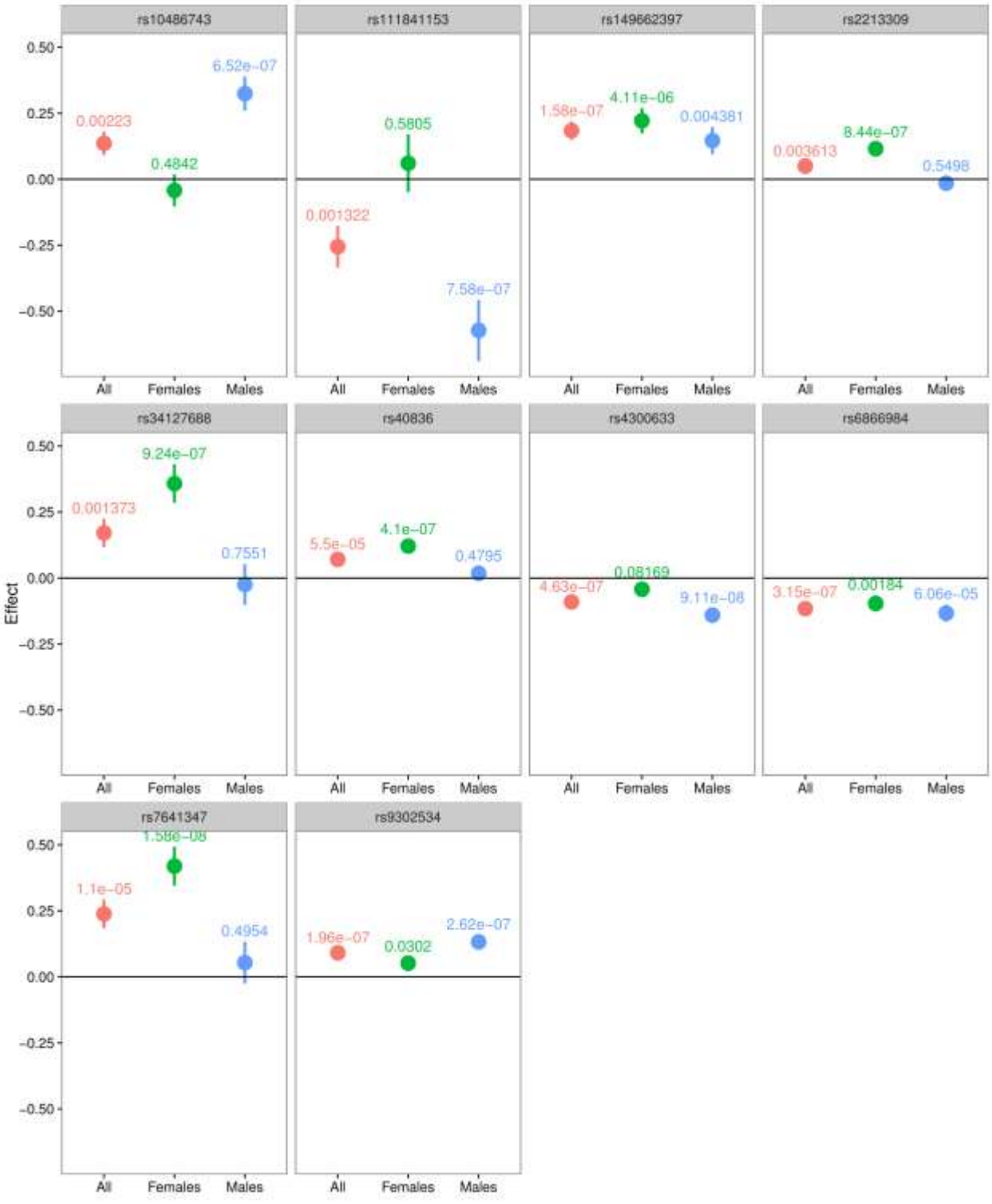
Manhattan plots for the non-stratified GWAS (A), and the males-only GWAS (B). n = 89553 and λ_{gc} = 1.089 and for the non-stratified GWAS. n = 44088 and λ_{gc} = 1.06 for the males-only GWAS.

Figure 6: Locus zoom plots for the most significant loci



Locus zoom plots for the most significant SNP in the males-only GWAS (A) (rs4300633, P = 9.11x10^{-8}), and the non-stratified GWAS (B) (rs149662397, P = 1.58x10^{-7}).

Figure 7: Effect direction for independent suggestive SNPs (P < $1x10^{-6}$) in the meta-analysis (23andMe+BLTS cohorts)



Point estimates are effect sizes (uncorrected for winner's curse) and bars represent standard errors. P-values provided for each SNP

6. References

1. Harkness, K., Sabbagh, M., Jacobson, J., Chowdrey, N. & Chen, T. Enhanced accuracy of mental state decoding in dysphoric college students. *Cogn. Emot.* **19,** 999–1025 (2005).

2. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. *Lancet* **381,** 1371–9 (2013).

3. Duncan, L. *et al.* Genome-Wide Association Study Reveals First Locus for Anorexia Nervosa and Metabolic Correlations. *bioRxiv* (2016).

4. Otowa, T. *et al.* Meta-analysis of genome-wide association studies of anxiety disorders. *Mol. Psychiatry* (2016). doi:10.1038/mp.2015.197

5. Rietveld, C. A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340,** 1467–71 (2013).

6. Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533,** 539–542 (2016).

7. Rietveld, C. A. *et al.* Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proc. Natl. Acad. Sci.* **111,** 13790–13794 (2014).

8. de Moor, M. H. M. *et al.* Meta-analysis of Genome-wide Association Studies for Neuroticism, and the Polygenic Association With Major Depressive Disorder. *JAMA psychiatry* **72,** 642–50 (2015).

9. van den Berg, S. M. *et al.* Meta-analysis of Genome-Wide Association Studies for Extraversion: Findings from the Genetics of Personality Consortium. *Behav. Genet.* **46,** 170–82 (2016).

10. de Moor, M. H. M. *et al.* Meta-analysis of genome-wide association studies for personality. *Mol. Psychiatry* **17,** 337–49 (2012).

11. Hibar, D. P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **520,** 224–9 (2015).

12. Lubke, G. H. *et al.* Genome-wide analyses of borderline personality features. *Mol. Psychiatry* **19,** 923–929 (2014).

13. Werling, D. M., Parikshak, N. N. & Geschwind, D. H. Gene expression in human brain implicates sexually dimorphic pathways in autism spectrum disorders. *Nat. Commun.* **7,** 10717 (2016).

14. Barbeira, A. *et al. MetaXcan: Summary Statistics Based Gene-Level Association Method Infers Accurate PrediXcan Results*. *bioRxiv* (2016).

15. Knafo, A. & Uzefovsky, F. in *The infant mind: Origins of the social Brain* (eds. Legerstee, M., Haley, D. W. & Bornstein, M. H.) 97 – 121 (The Guilford Press, 2013).