Supplementary Figures

Contents

Figure S1a-b: Quantile-quantile (QQ) plot based on (a) lambda$_{GC}$ corrected input files, and (b) based on uncorrected input files 2

Figure S2a-i. Manhattan plots of the individual samples in the discovery meta-analysis (lamdba corrected) 3

Figure S3a-i. Quantile-quantile plots of the individual samples in the discovery meta-analysis (lambda corrected). 12

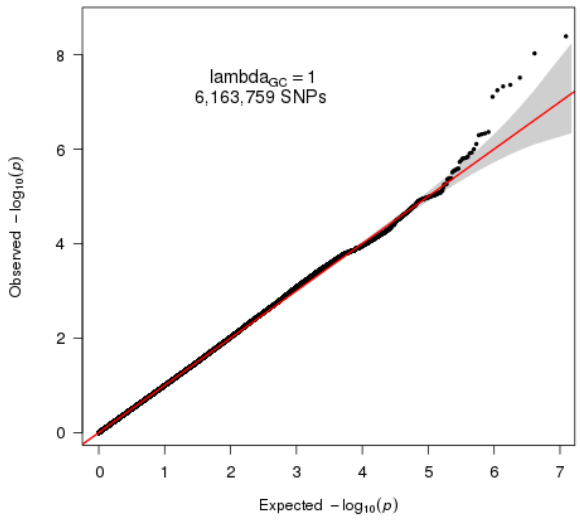
Figure S4a-l. Regional association plots showing signal around top SNPs 21

Figure S5 a-k. Forest plots top SNPs 33

Figure S6. The quantile-quantile (QQ) plot of the gene-based test 44

Figure S1a-b: Quantile-quantile (QQ) plot based on (a) $lambda_{GC}$ corrected input files, and (b) based on uncorrected input files

a.



b.

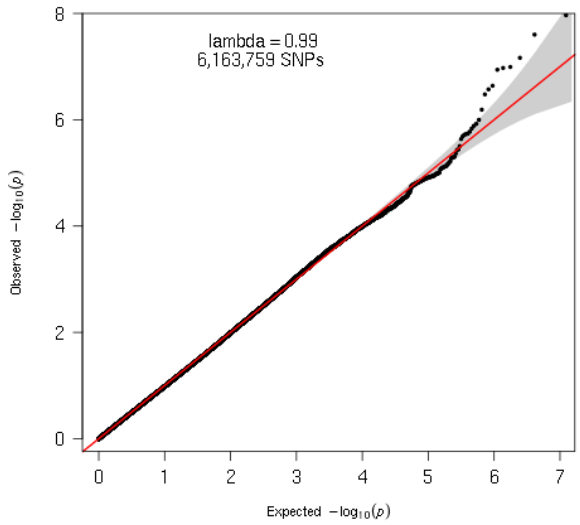
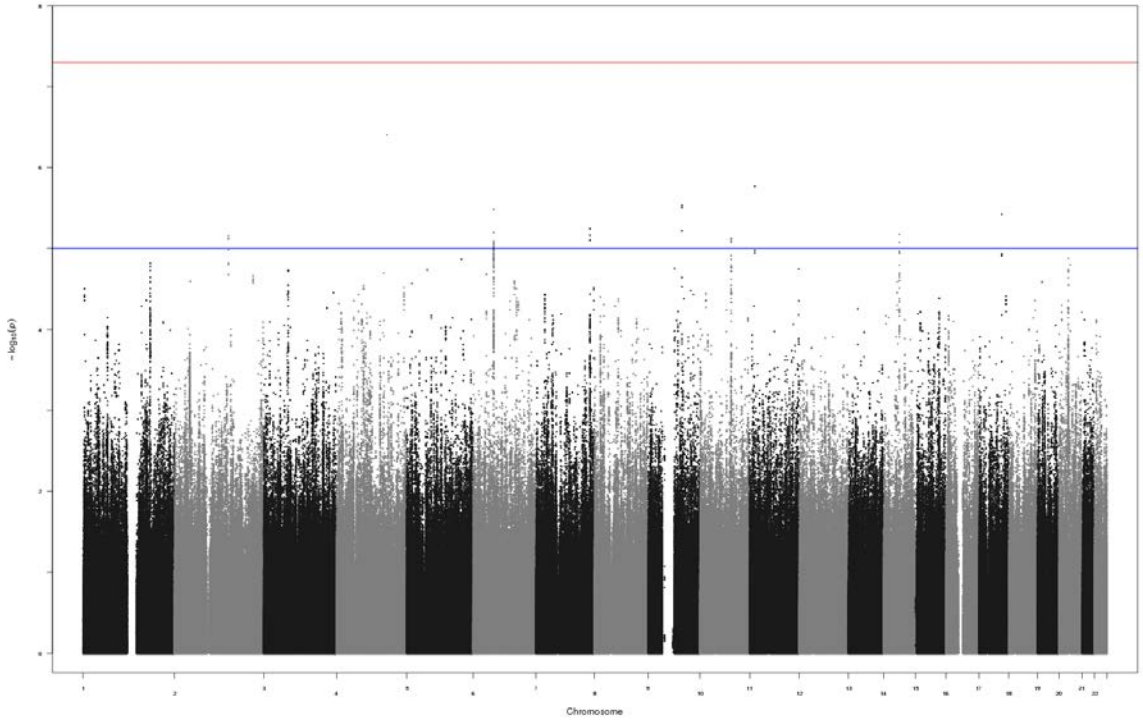


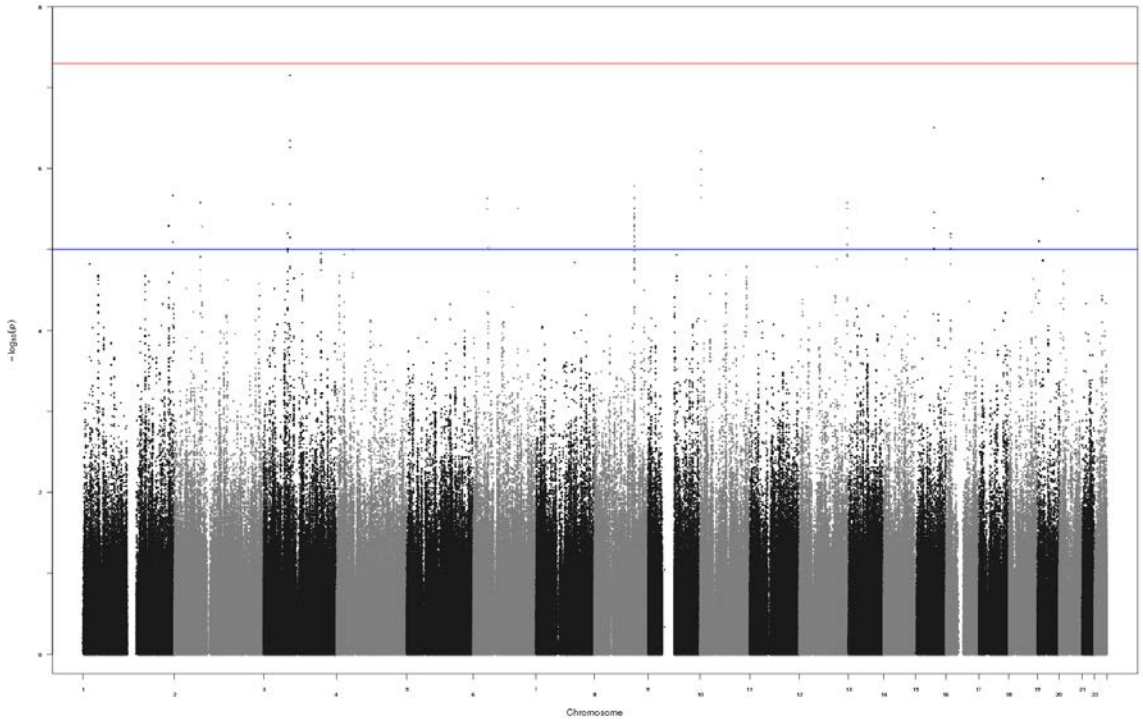
Figure S2a-i. Manhattan plots of the individual samples in the discovery meta-analysis (lamdba corrected)

The y-axes show the strength of association($-\log_{10}(P)$) and the x-axes the chromosomal position. The blue line indicates suggestive significance level (P-values < 10^{-5}) and the red line indicates genome-wide significance level (P-values < $5x10^{-8}$).

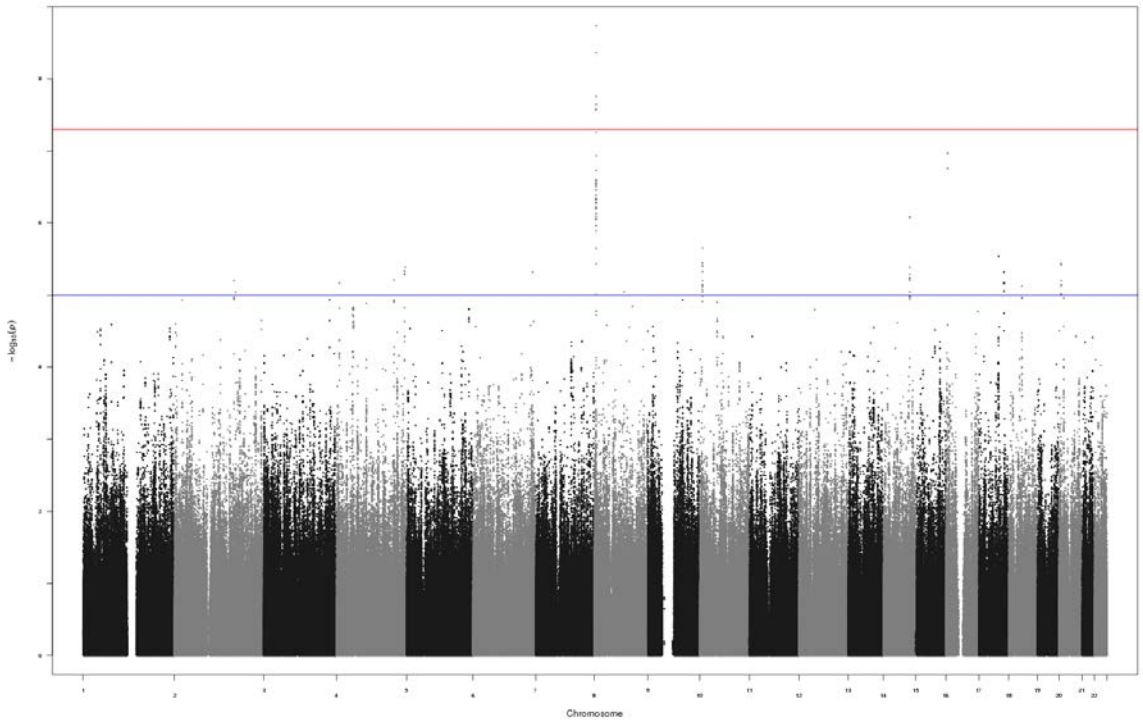
a. Avon Longitudinal Study of Parents and Children (ALSPAC)



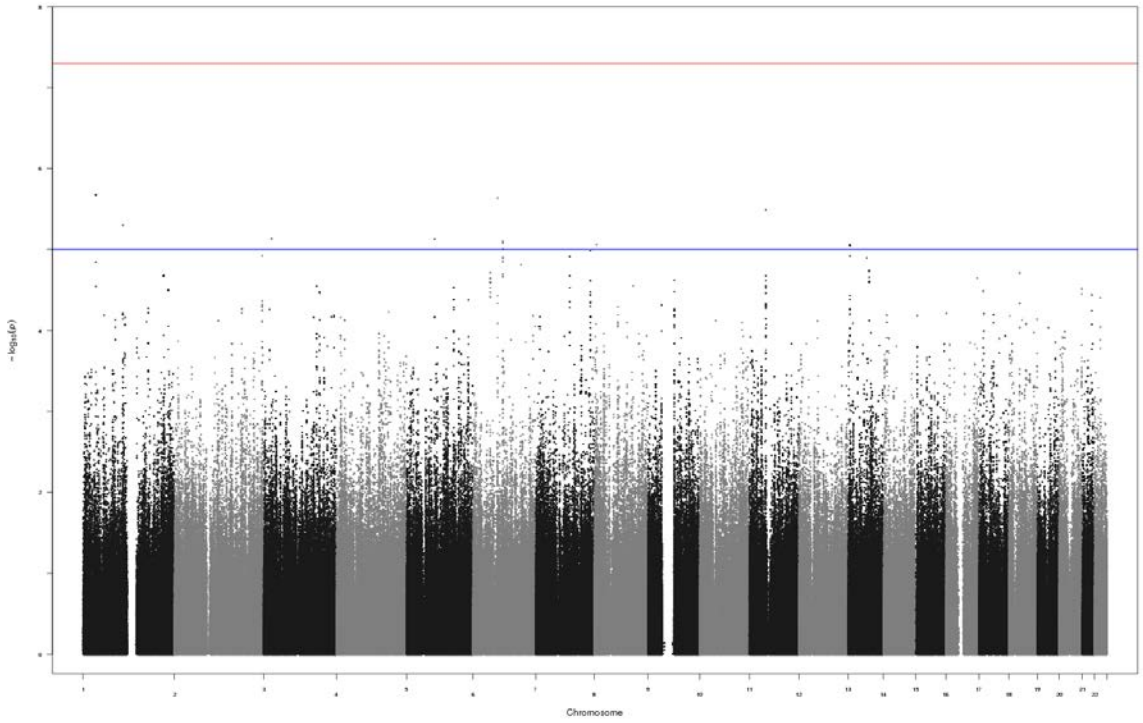
b. Brisbane Longitudinal Twin Study (BLTS)



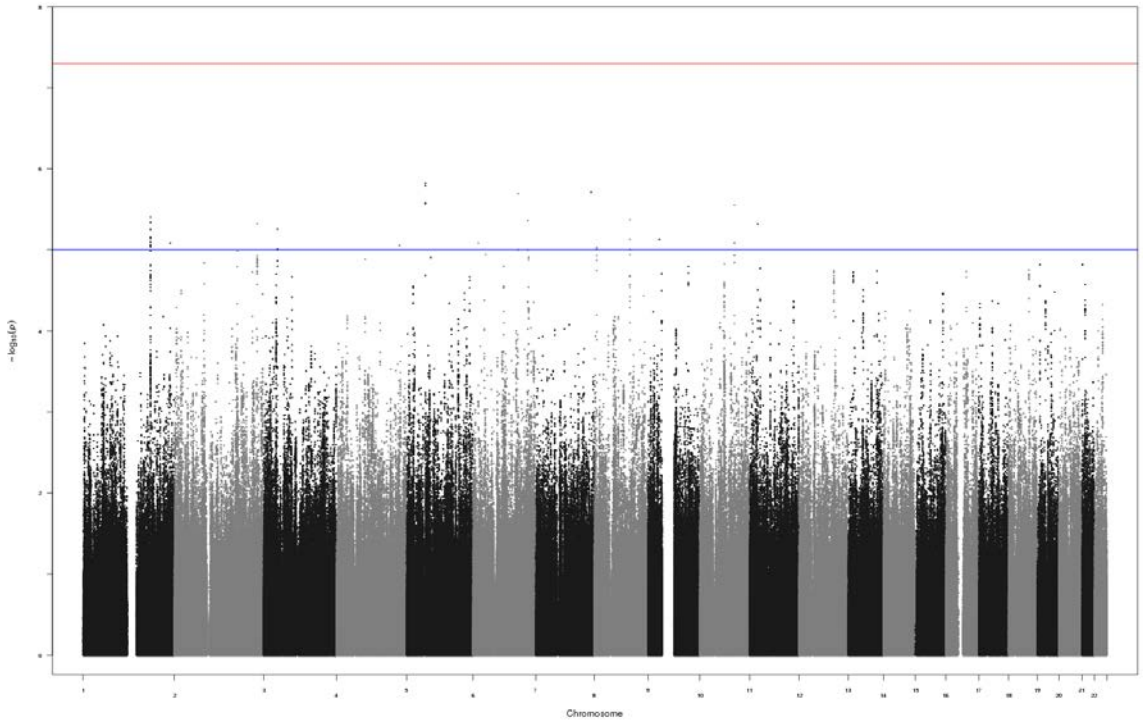
c. Finnish Twin Cohort (FinnTwin12 & FinnTwin16)



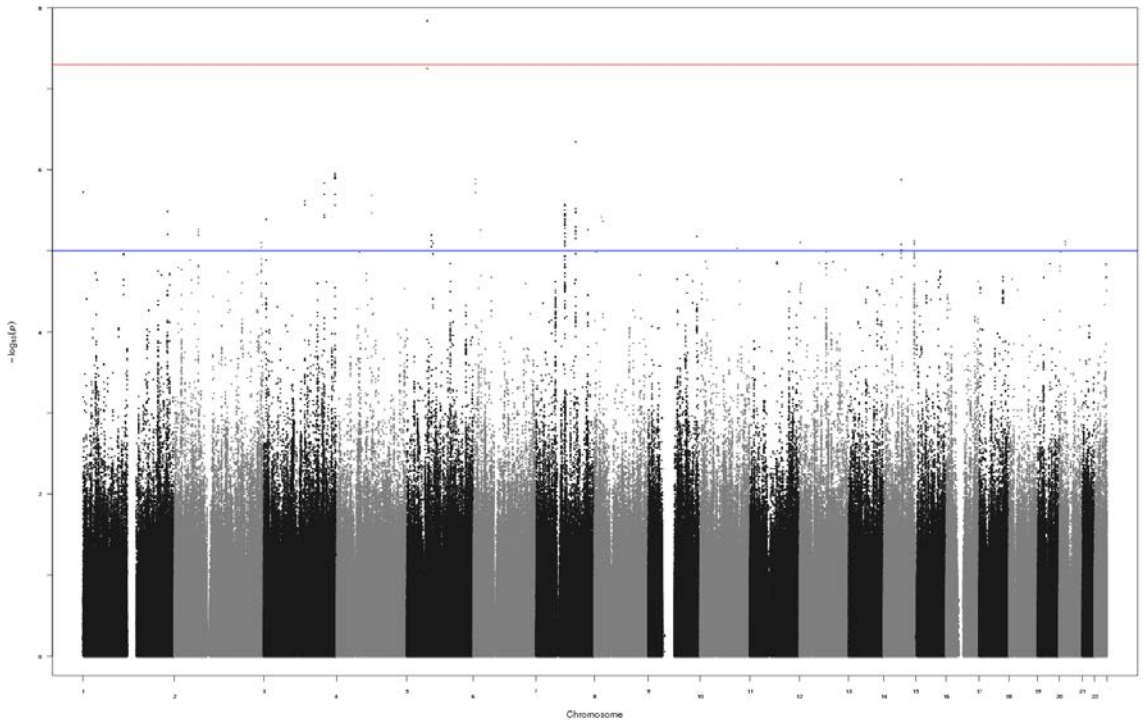
d. Hospital Universitari Vall d'Hebron – Barcelona (HUVH-Barcelona)



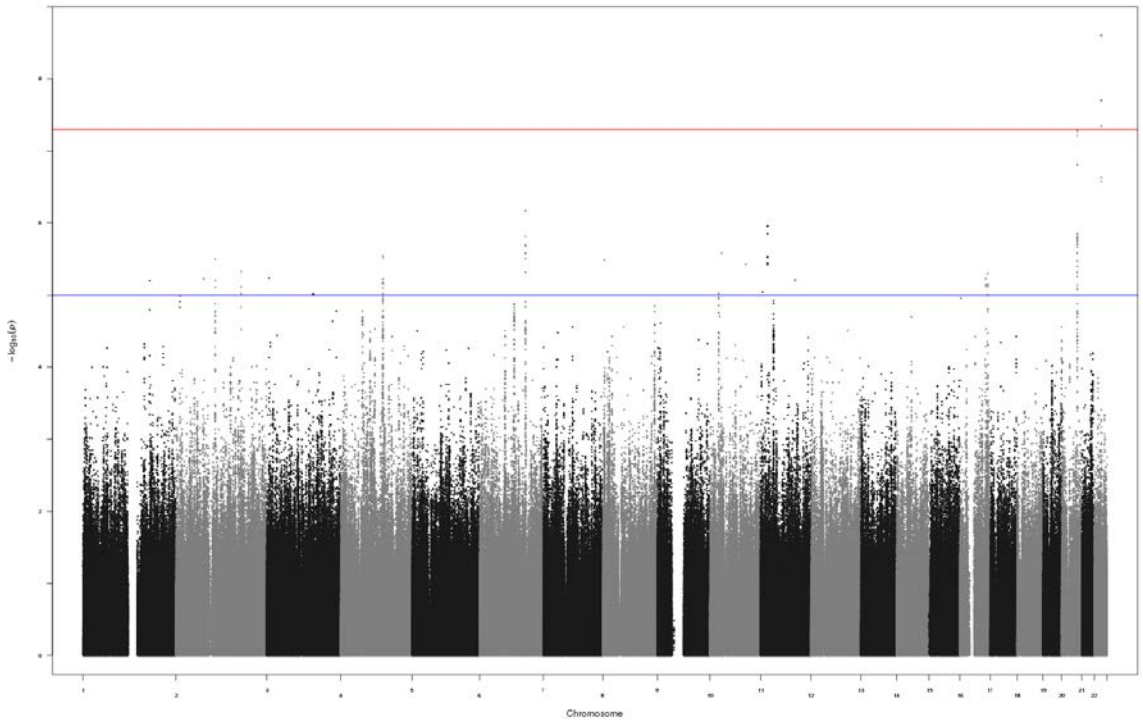
e. Netherlands Twin Register (NTR)



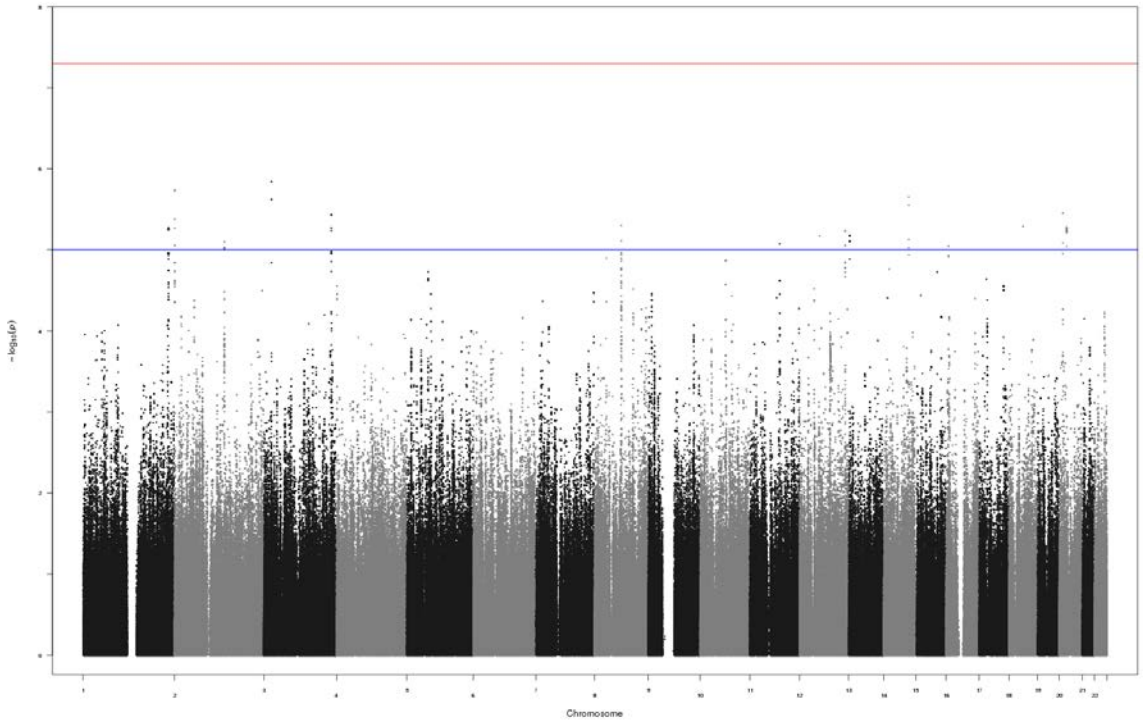
f. Queensland Institute of Medical Research Berghofer adults (QIMR Berghofer adults)



g. Tracking Adolescents'Individual Lives Survey (TRAILS)



h. Utrecht



i. Genetics of Substance Dependence - European American (Yale-Penn)

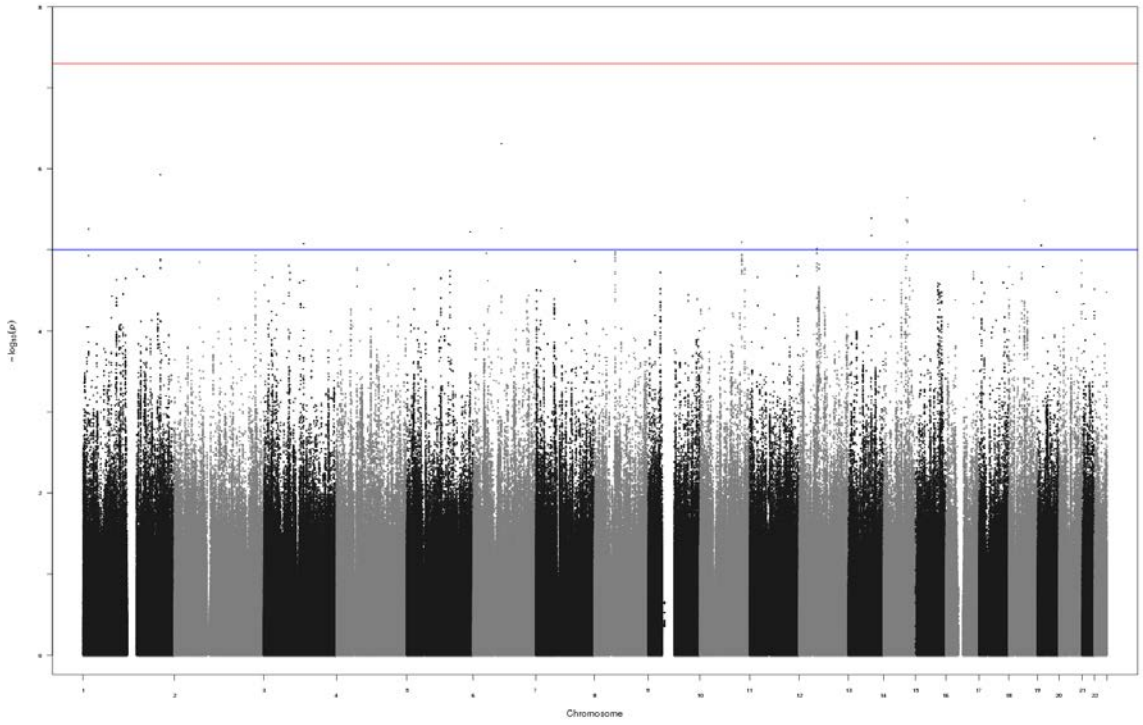
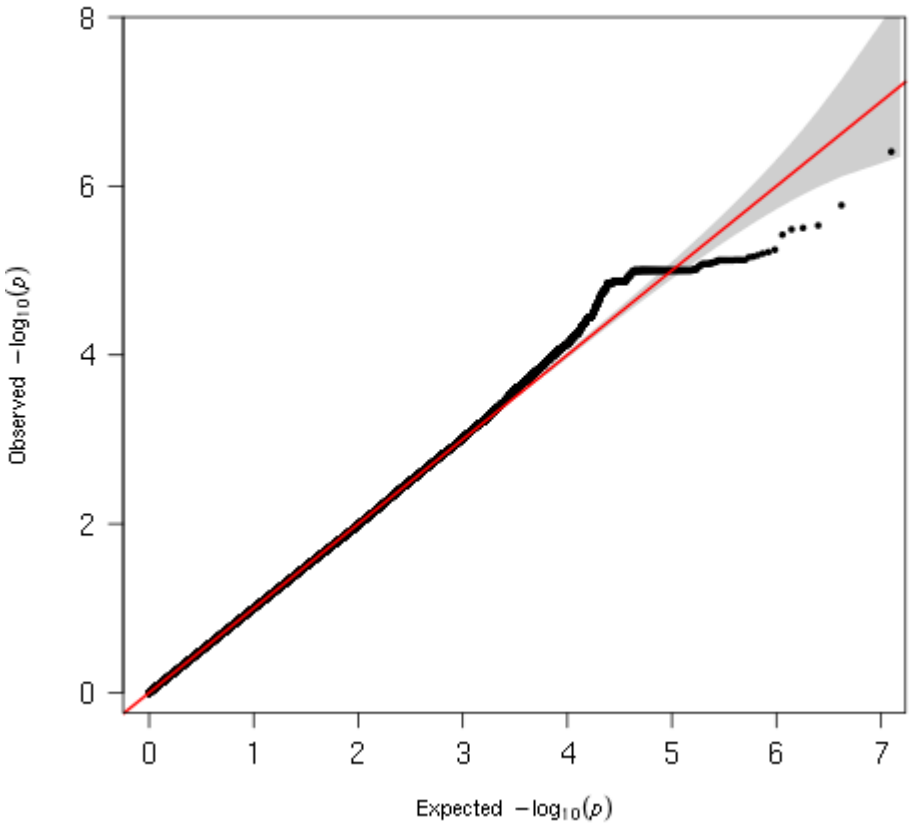


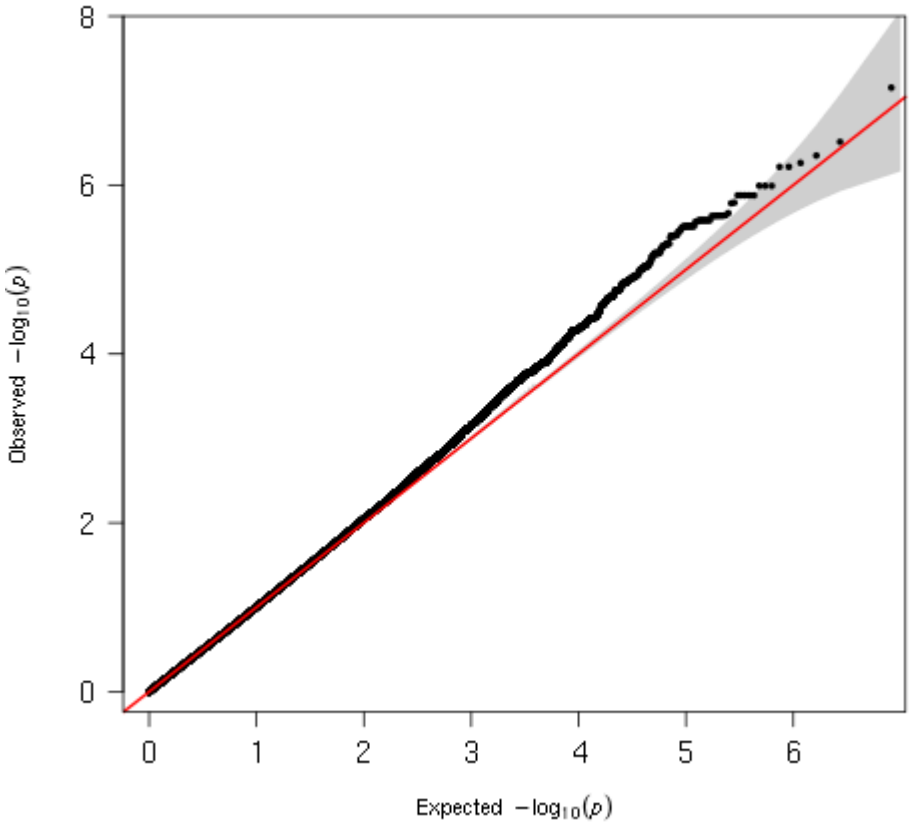
Figure S3a-i. Quantile-quantile plots of the individual samples in the discovery meta-analysis (lambda corrected).

The plots compare observed p-values with p-values expected from a normal distribution.

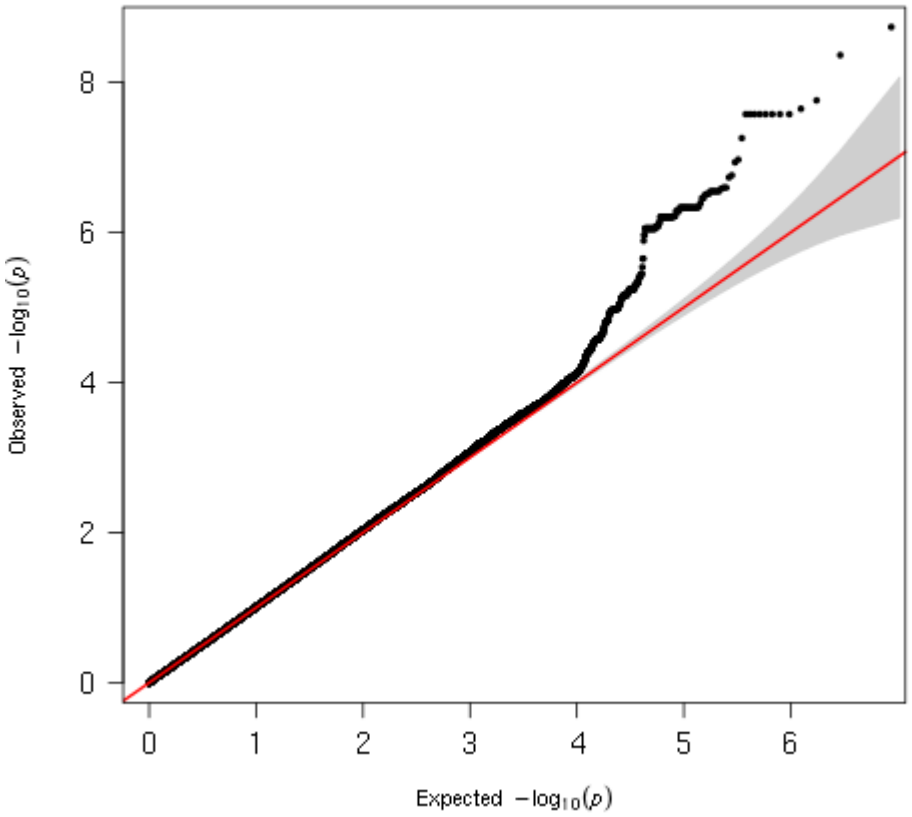
a. Avon Longitudinal Study of Parents and Children (ALSPAC) lambda=.96



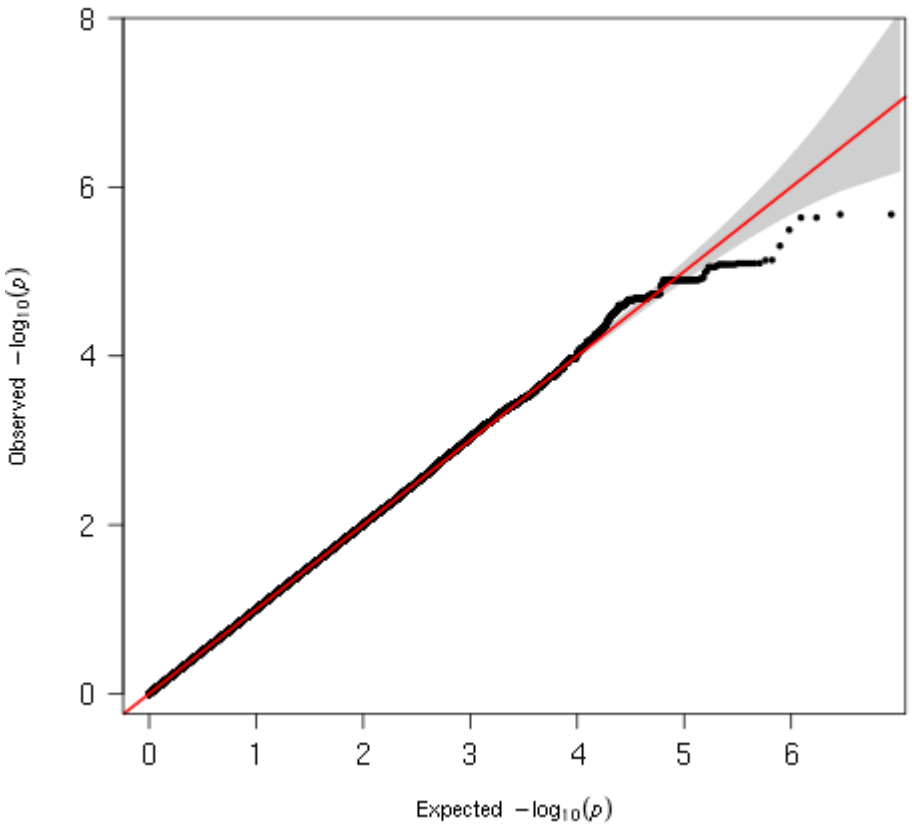
b. Brisbane Longitudinal Twin Study (BLTS) lambda=1.08



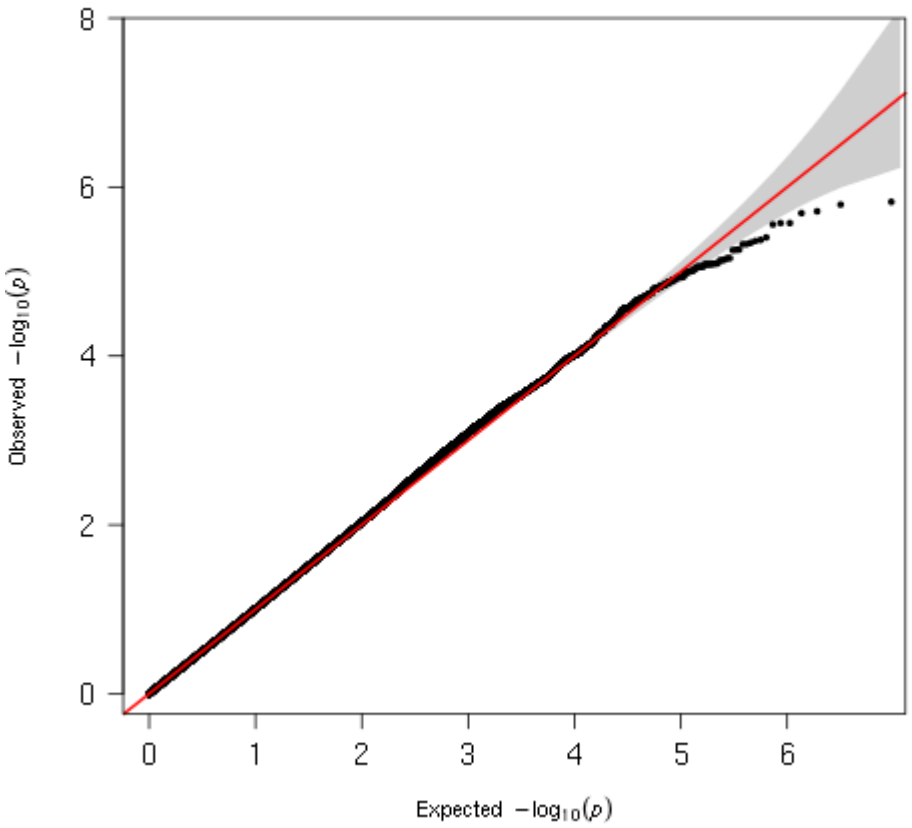
c. Finnish Twin Cohort (FinnTwin12 & FinnTwin16) lambda=1.17



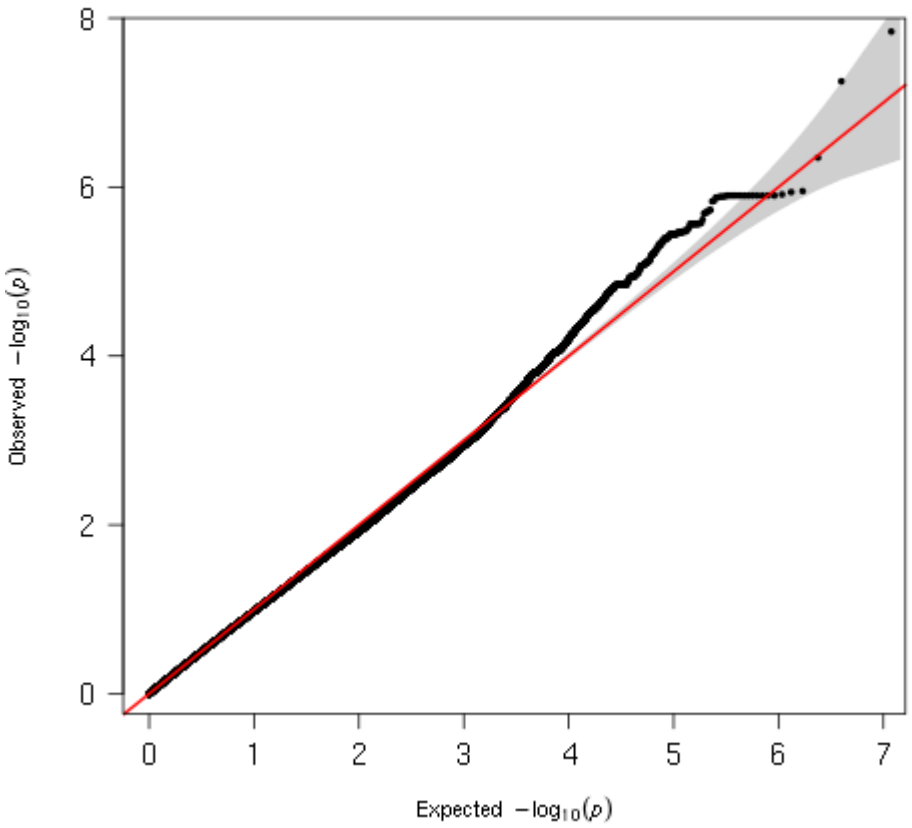
d. Hospital Universitari Vall d'Hebron – Barcelona (HUVH-Barcelona) lambda=1.11



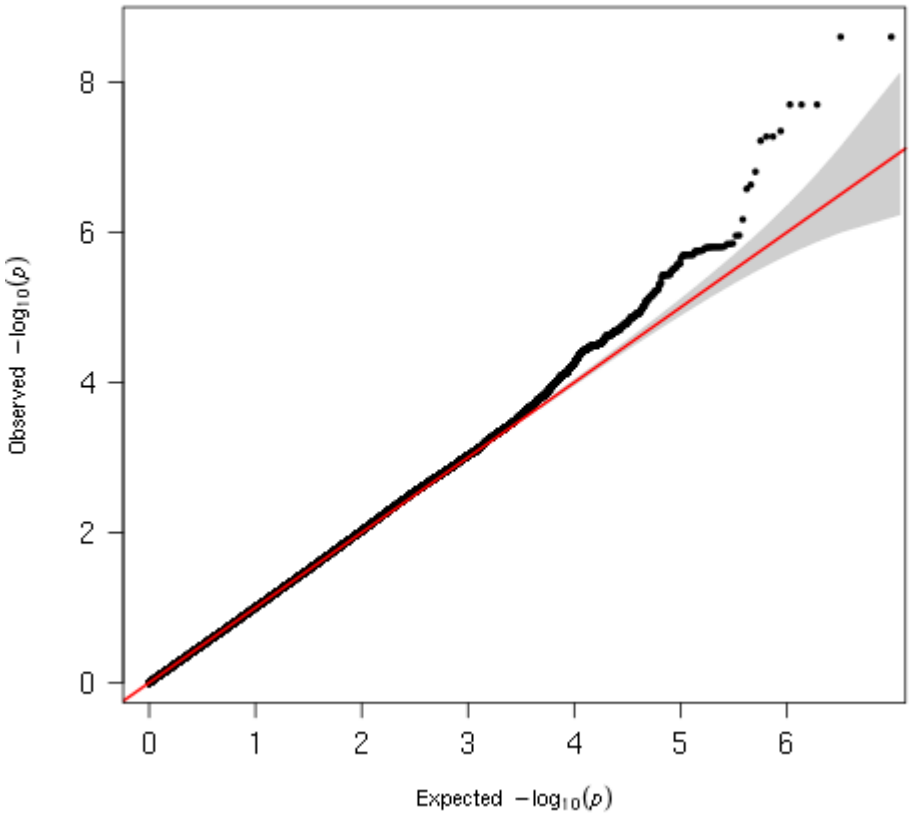
e. Netherlands Twin Register (NTR) lambda=1.08



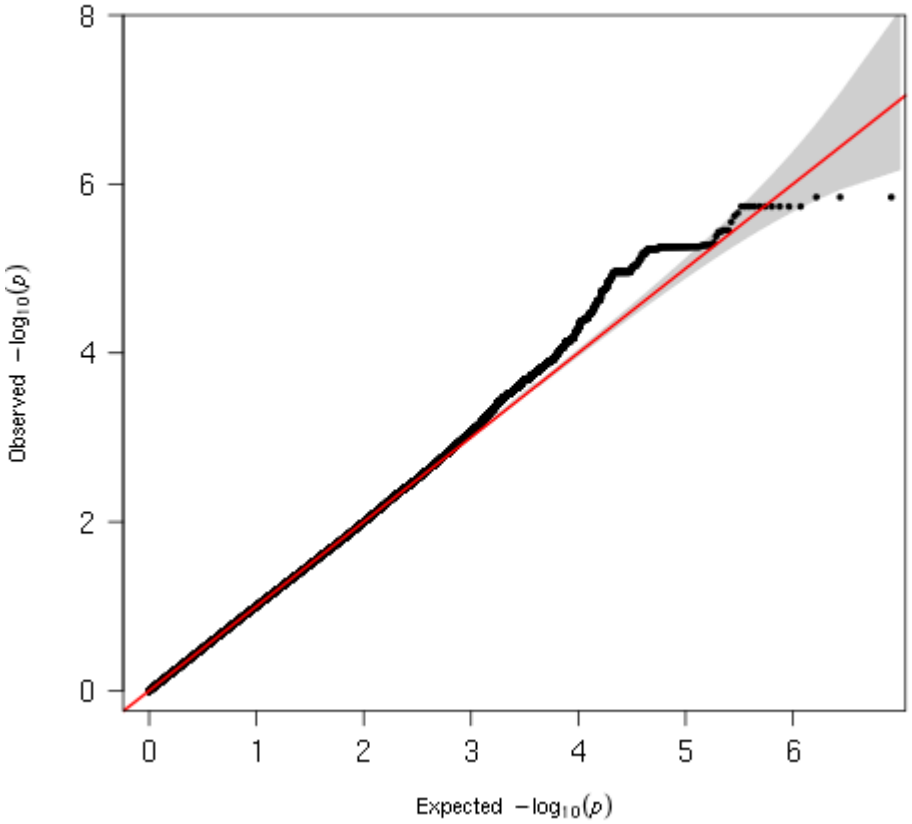
f. Queensland Institute of Medical Research Berghofer adults (QIMR Berghofer adults) lambda=0.8



g. Tracking Adolescents'Individual Lives Survey (TRAILS) lambda=.87



h. Utrecht lambda=.99



i. Genetics of Substance Dependence - European American (Yale Penn) lambda=1.02

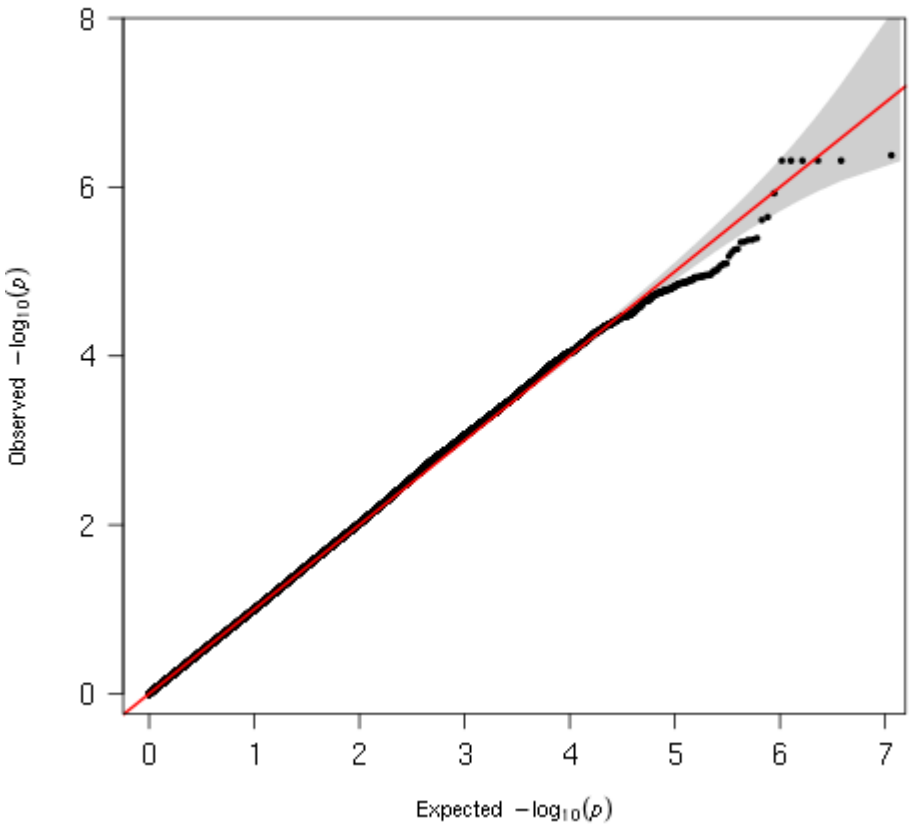
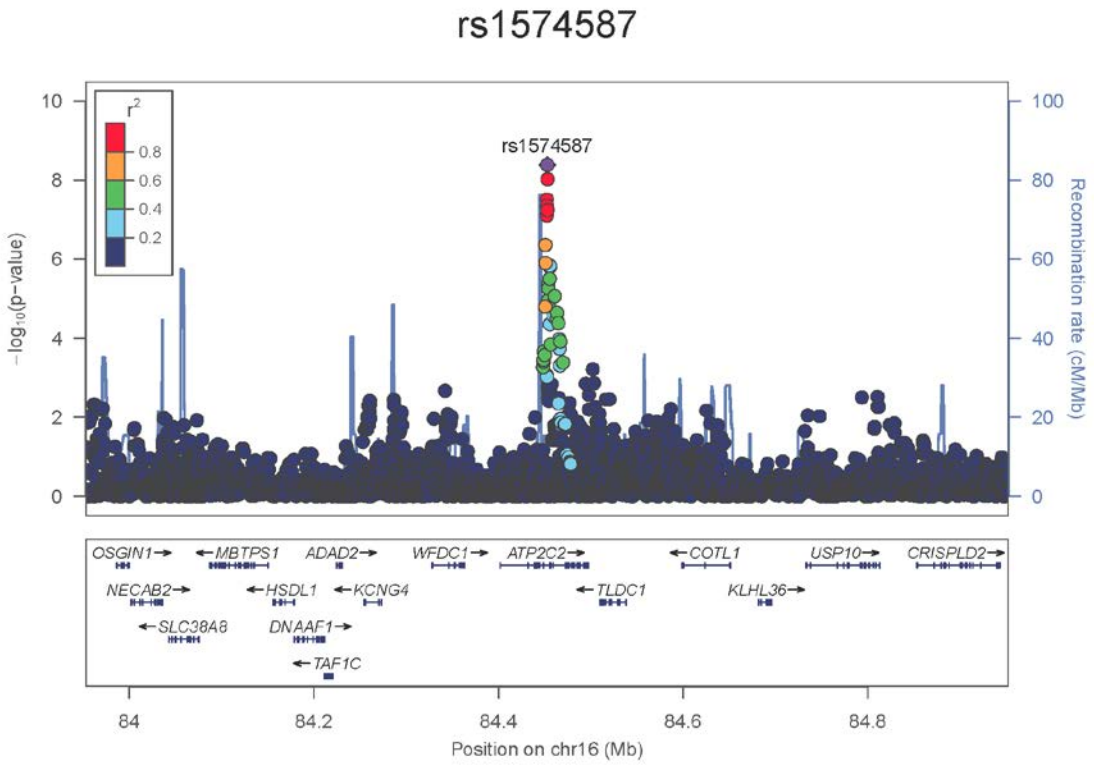


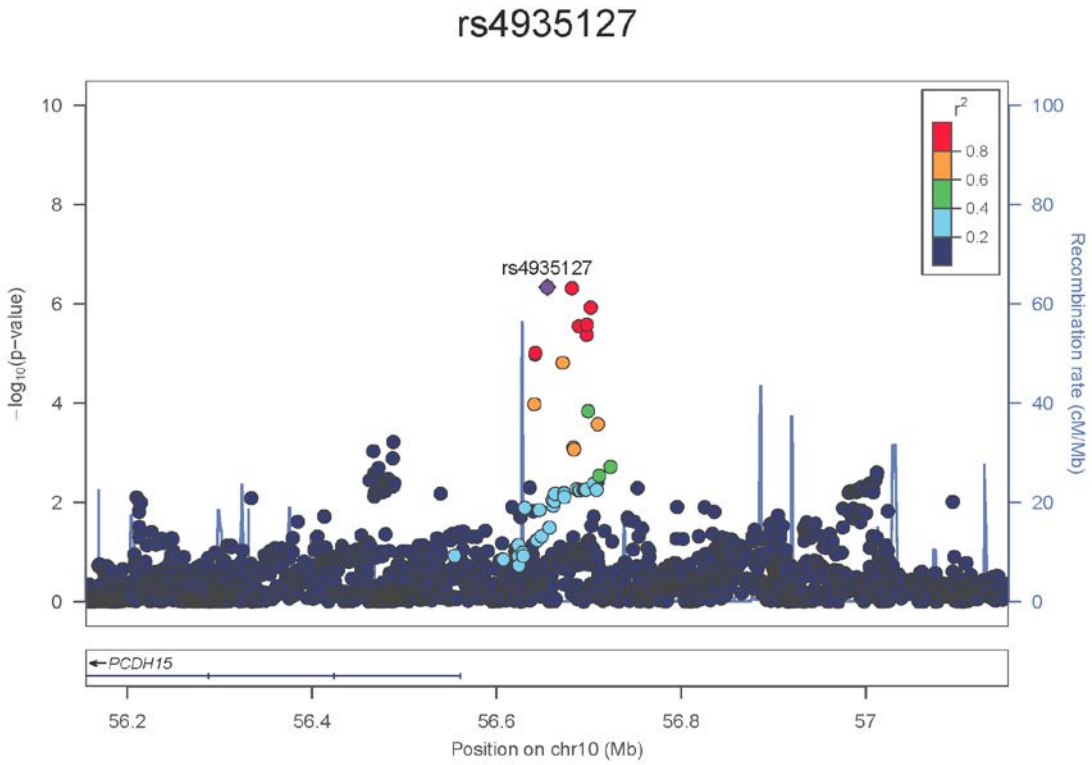
Figure S4a-l. Regional association plots showing signal around top SNPs

Regional plots showing the top independent SNPs (displayed as a violet dot at the top of the figure) in the discovery meta-analysis. The strength of the signal is displayed on the *y*-axis (–$\log_{10}$ p-value), and the chromosomal position is shown on the *x*-axis. The blue lines represent the recombination rate. The LD (r^2) of the SNPs within the region with the top signal is colour coded.

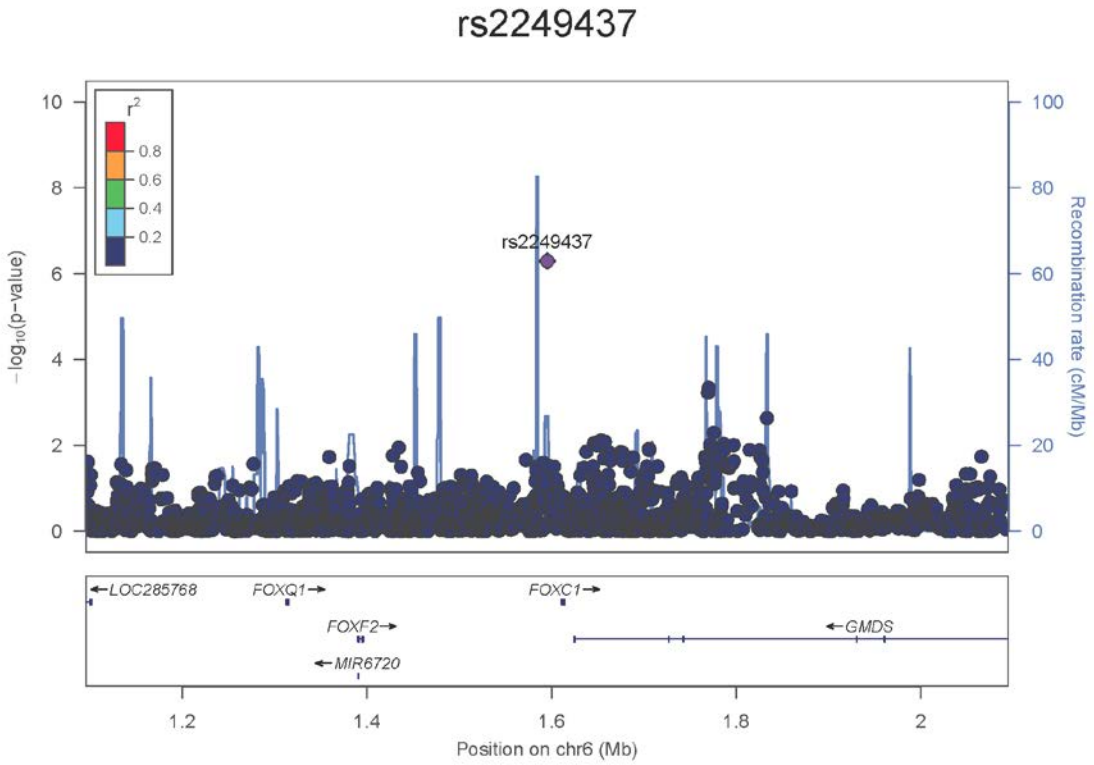
a.



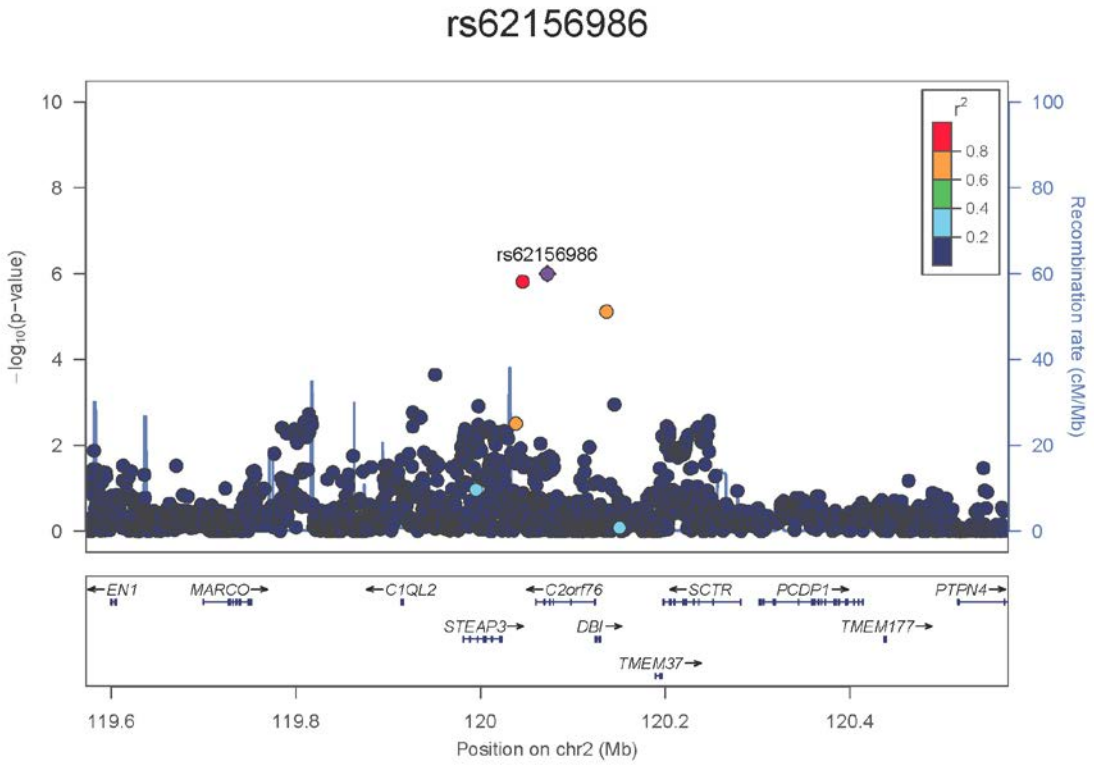
b.


rs4935127
rs4935127
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
PCDH15
Position on chr10 (Mb)
56.2
56.4
56.6
56.8
57

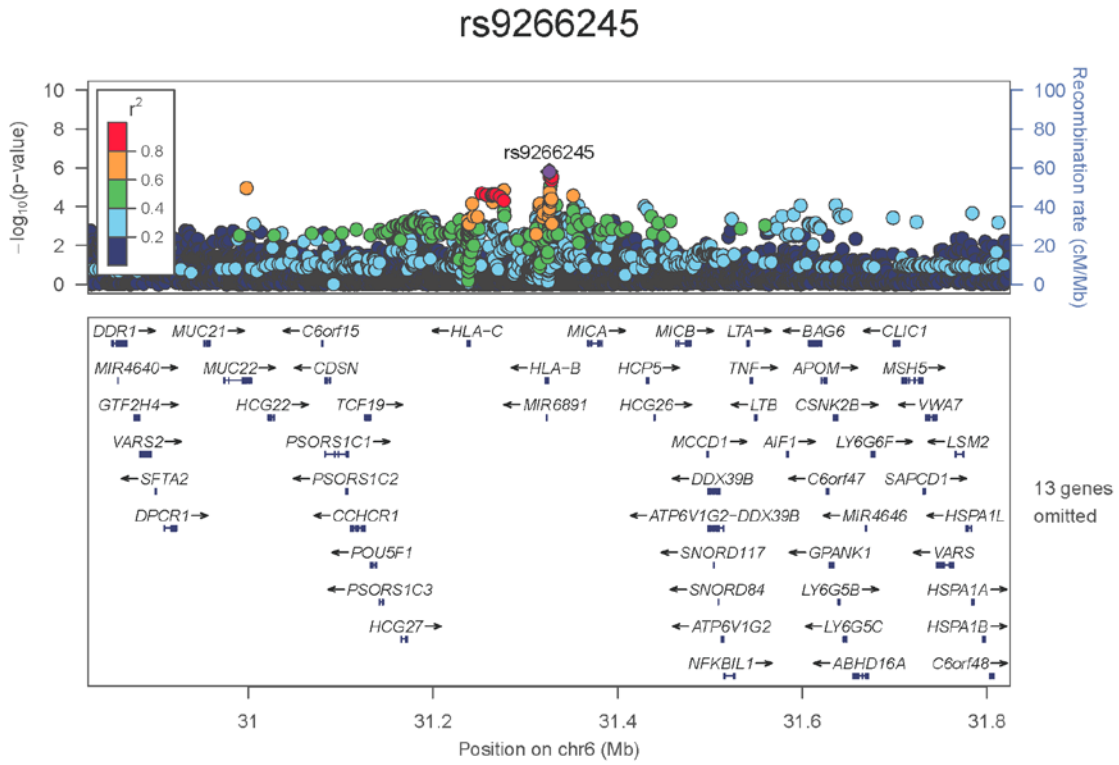
c.


rs2249437
r²
0.8
0.6
0.4
0.2
rs2249437
−log10(p-value)
Recombination rate (cM/Mb)
LOC285768
FOXQ1
FOXC1
FOXF2
GMDS
MIR6720
Position on chr6 (Mb)

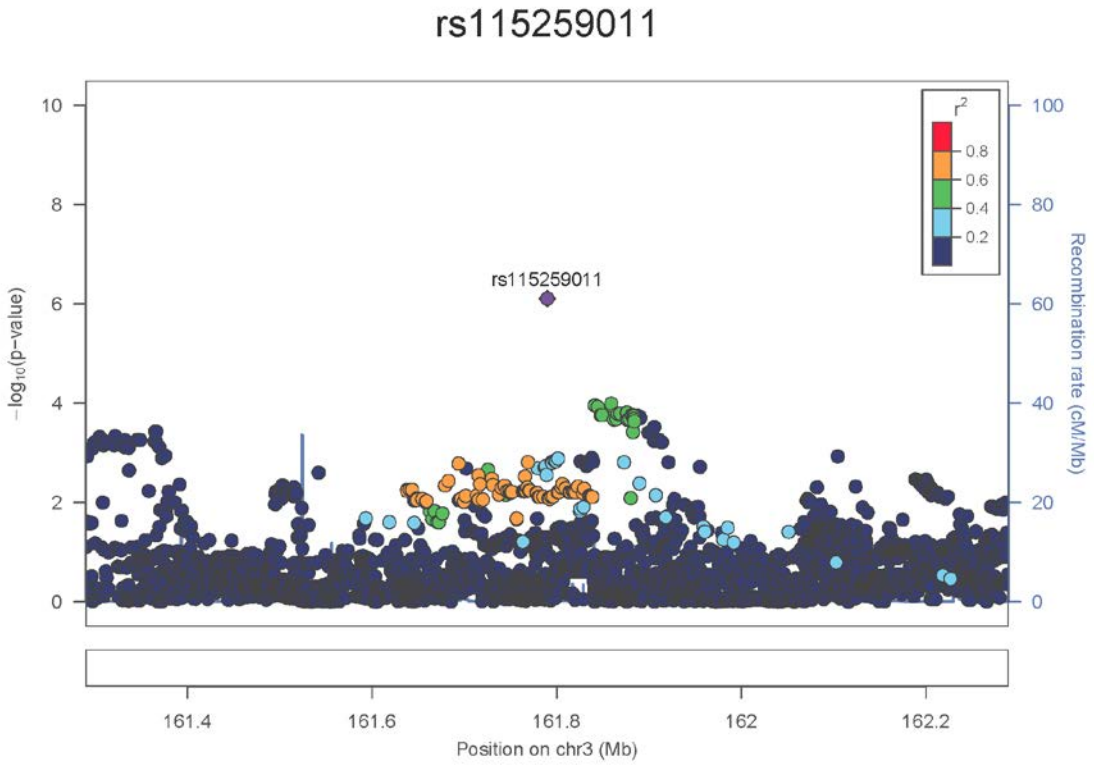
d.


rs62156986
rs62156986
r^2
0.8
0.6
0.4
0.2
$-\log_{10}$(p-value)
Recombination rate (cM/Mb)
10
8
6
4
2
0
100
80
60
40
20
0
EN1
MARCO
C1QL2
C2orf76
SCTR
PCDP1
PTPN4
STEAP3
DBI
TMEM177
TMEM37
119.6
119.8
120
120.2
120.4
Position on chr2 (Mb)

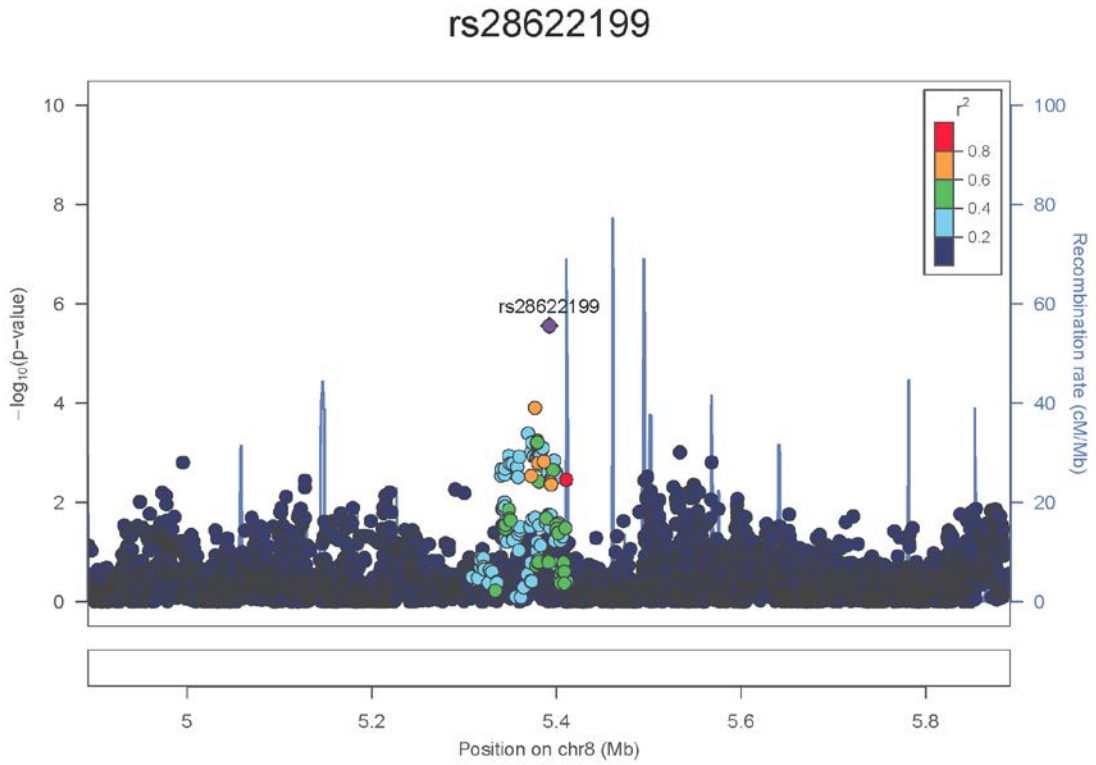
e.


rs9266245
rs9266245
r²
0.8
0.6
0.4
0.2
−log10(p-value)
Recombination rate (cM/Mb)
DDR1 MUC21 C6orf15 HLA-C MICA MICB LTA BAG6 CLIC1
MIR4640 MUC22 CDSN HLA-B HCP5 TNF APOM MSH5
GTF2H4 HCG22 TCF19 MIR6891 HCG26 LTB CSNK2B VWA7
VARS2 PSORS1C1 MCCD1 AIF1 LY6G6F LSM2
SFTA2 PSORS1C2 DDX39B C6orf47 SAPCD1
DPCR1 CCHCR1 ATP6V1G2-DDX39B MIR4646 HSPA1L
POU5F1 SNORD117 GPANK1 VARS
PSORS1C3 SNORD84 LY6G5B HSPA1A
HCG27 ATP6V1G2 LY6G5C HSPA1B
NFKBIL1 ABHD16A C6orf48
13 genes omitted
Position on chr6 (Mb)

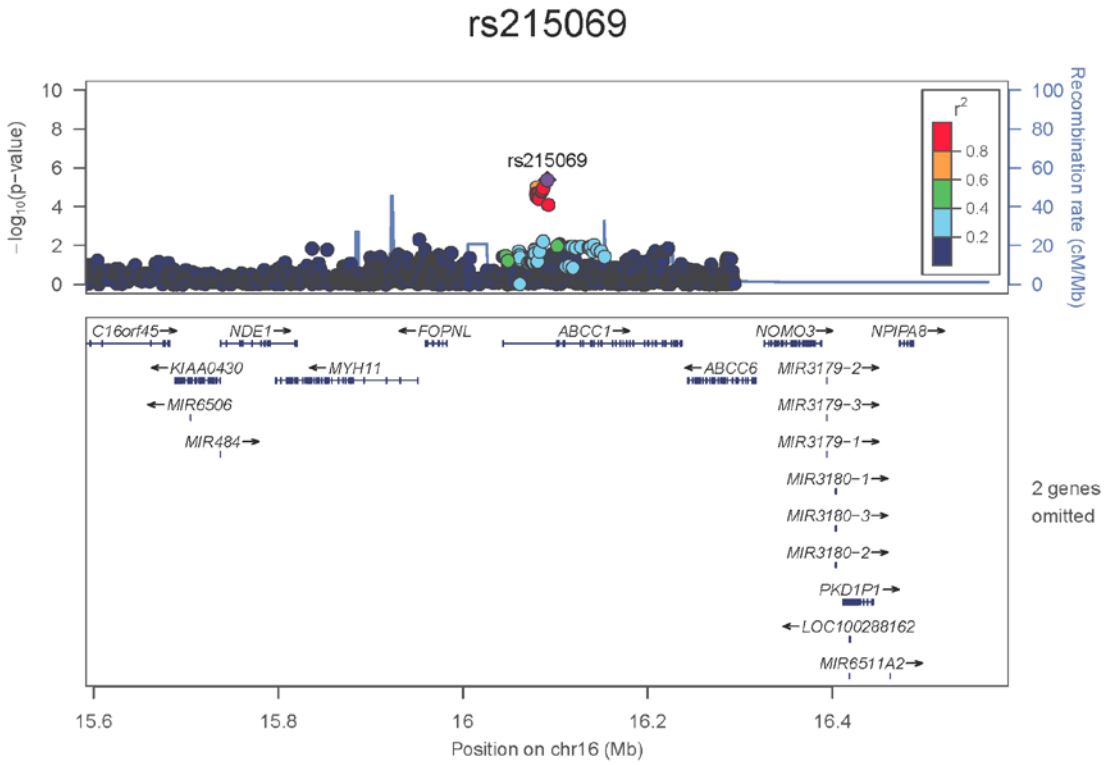
f.


rs115259011
rs115259011
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
10
8
6
4
2
0
100
80
60
40
20
0
161.4
161.6
161.8
162
162.2
Position on chr3 (Mb)

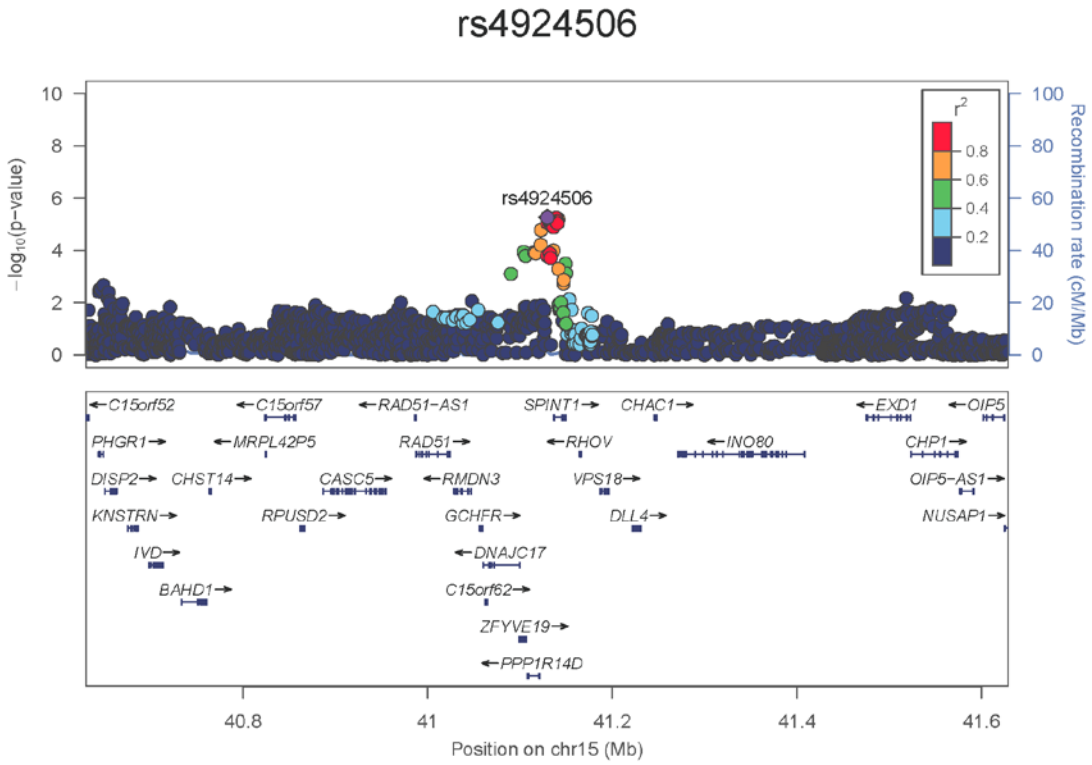
g.


rs28622199
r²
0.8
0.6
0.4
0.2
rs28622199
−log10(p−value)
Recombination rate (cM/Mb)
Position on chr8 (Mb)
10
8
6
4
2
0
100
80
60
40
20
0
5
5.2
5.4
5.6
5.8

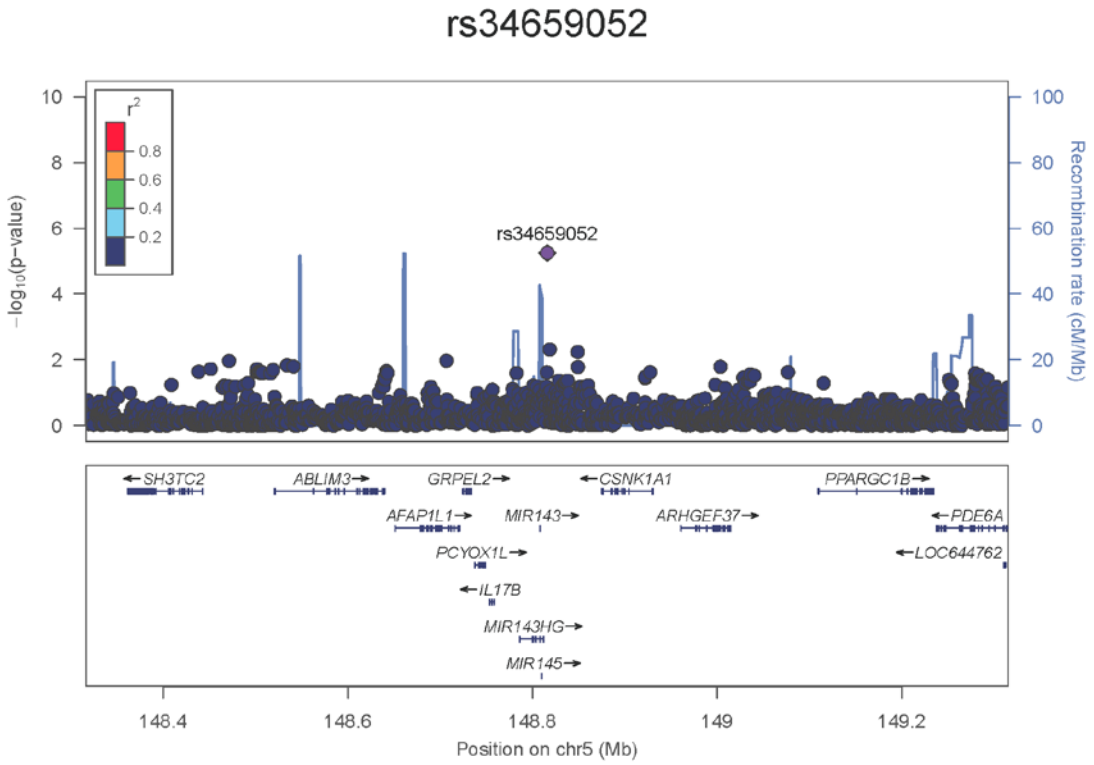
h.


rs215069
rs215069
−log10(p-value)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
C16orf45
NDE1
FOPNL
ABCC1
NOMO3
NPIPA8
KIAA0430
MYH11
ABCC6
MIR3179-2
MIR6506
MIR3179-3
MIR484
MIR3179-1
MIR3180-1
MIR3180-3
MIR3180-2
PKD1P1
LOC100288162
MIR6511A2
2 genes omitted
15.6
15.8
16
16.2
16.4
Position on chr16 (Mb)

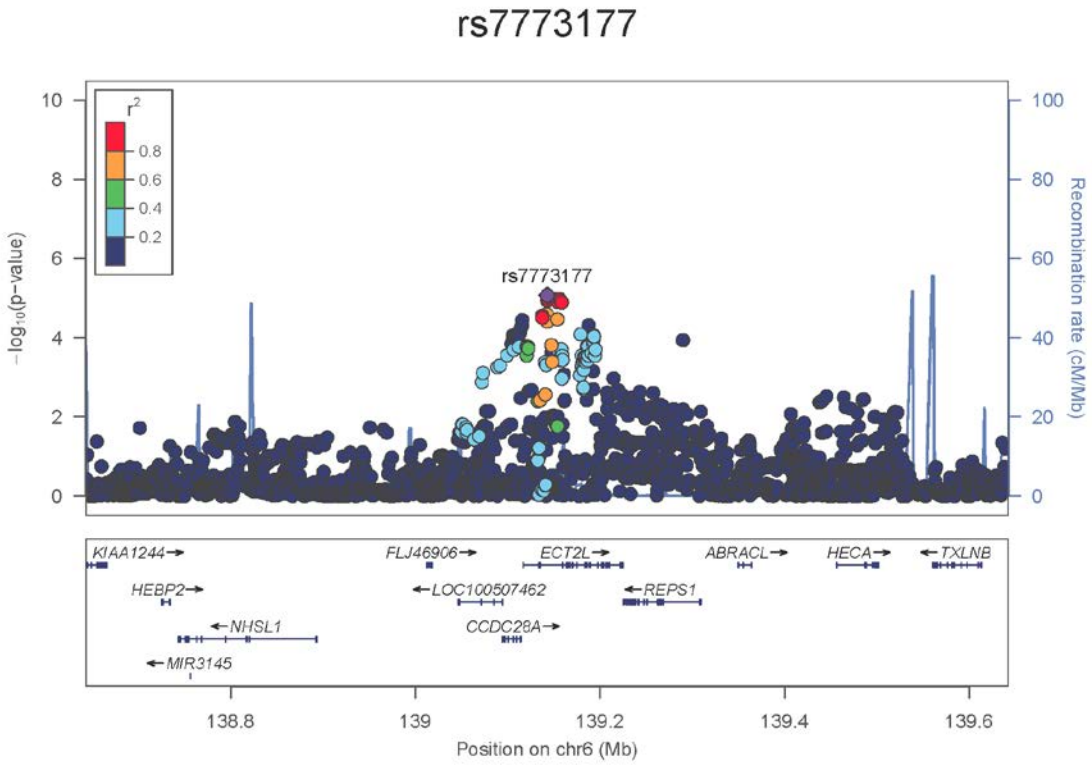
i.


rs4924506
rs4924506
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
C15orf52
C15orf57
RAD51−AS1
SPINT1
CHAC1
EXD1
OIP5
PHGR1
MRPL42P5
RAD51
RHOV
INO80
CHP1
DISP2
CHST14
CASC5
RMDN3
VPS18
OIP5−AS1
KNSTRN
RPUSD2
GCHFR
DLL4
NUSAP1
IVD
DNAJC17
BAHD1
C15orf62
ZFYVE19
PPP1R14D
Position on chr15 (Mb)

j.


rs34659052
r²
0.8
0.6
0.4
0.2
rs34659052
−log10(p-value)
Recombination rate (cM/Mb)
SH3TC2
ABLIM3
GRPEL2
CSNK1A1
PPARGC1B
AFAP1L1
MIR143
ARHGEF37
PDE6A
PCYOX1L
LOC644762
IL17B
MIR143HG
MIR145
148.4
148.6
148.8
149
149.2
Position on chr5 (Mb)

k.


rs7773177
r²
0.8
0.6
0.4
0.2
rs7773177
−log10(p-value)
Recombination rate (cM/Mb)
KIAA1244
FLJ46906
ECT2L
ABRACL
HECA
TXLNB
HEBP2
LOC100507462
REPS1
NHSL1
CCDC28A
MIR3145
138.8
139
139.2
139.4
139.6
Position on chr6 (Mb)

l.

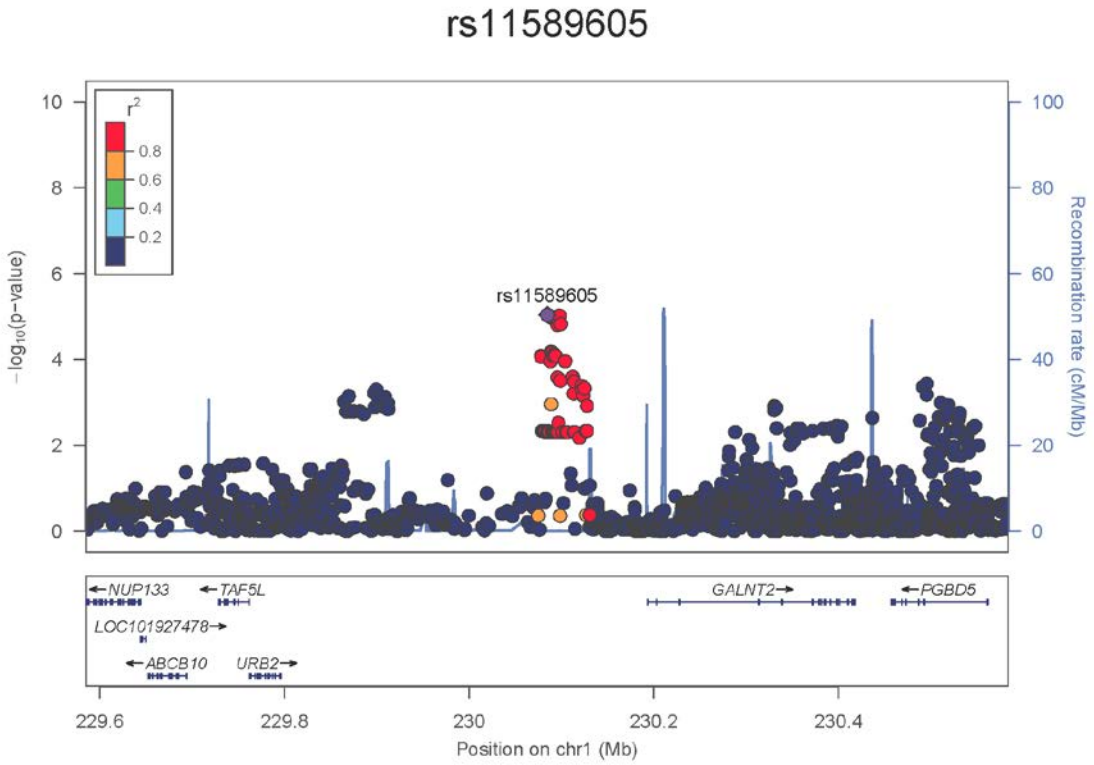
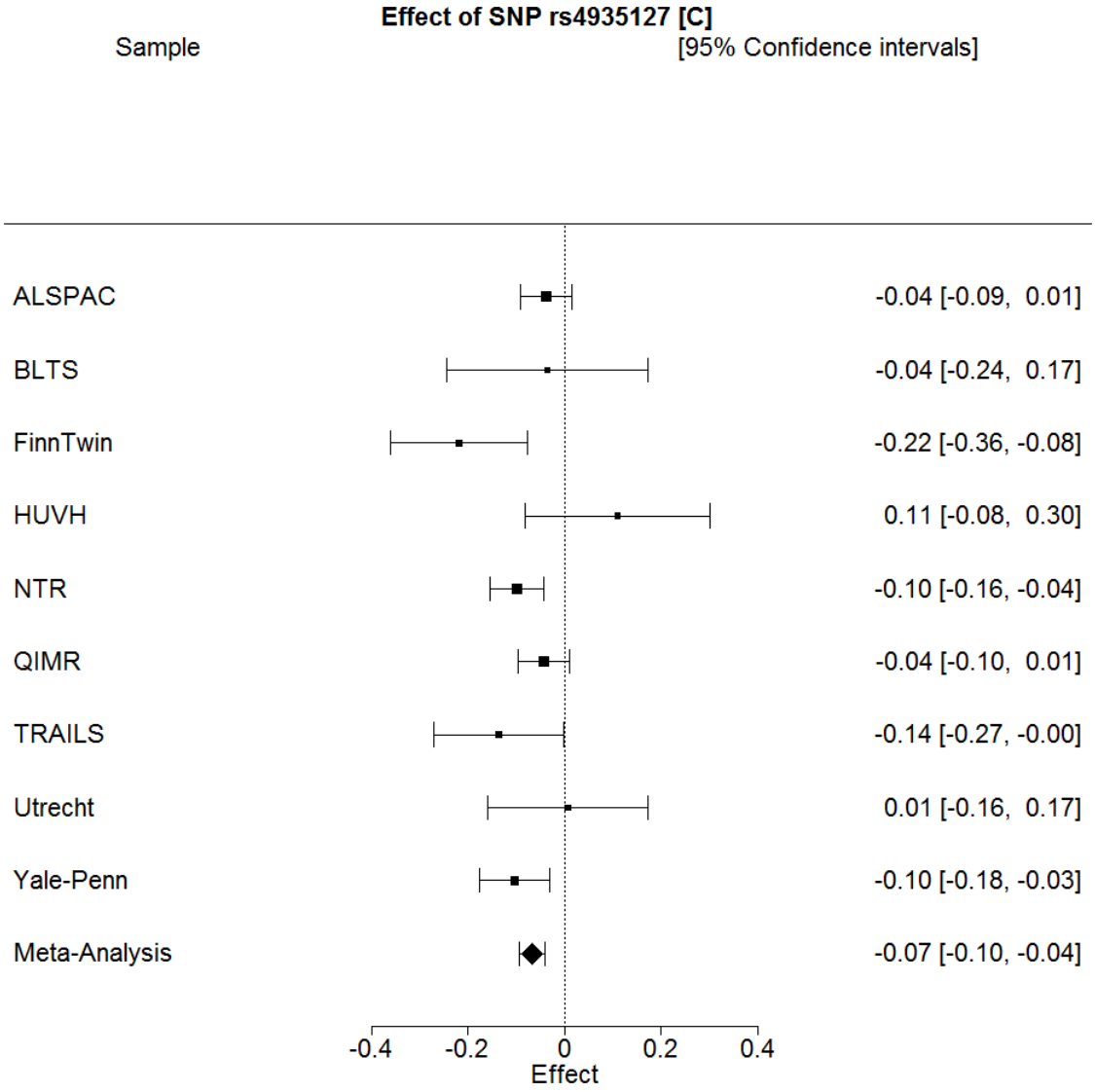

rs11589605
r²
0.8
0.6
0.4
0.2
rs11589605
−log10(p-value)
Recombination rate (cM/Mb)
NUP133
TAF5L
GALNT2
PGBD5
LOC101927478
ABCB10
URB2
229.6
229.8
230
230.2
230.4
Position on chr1 (Mb)

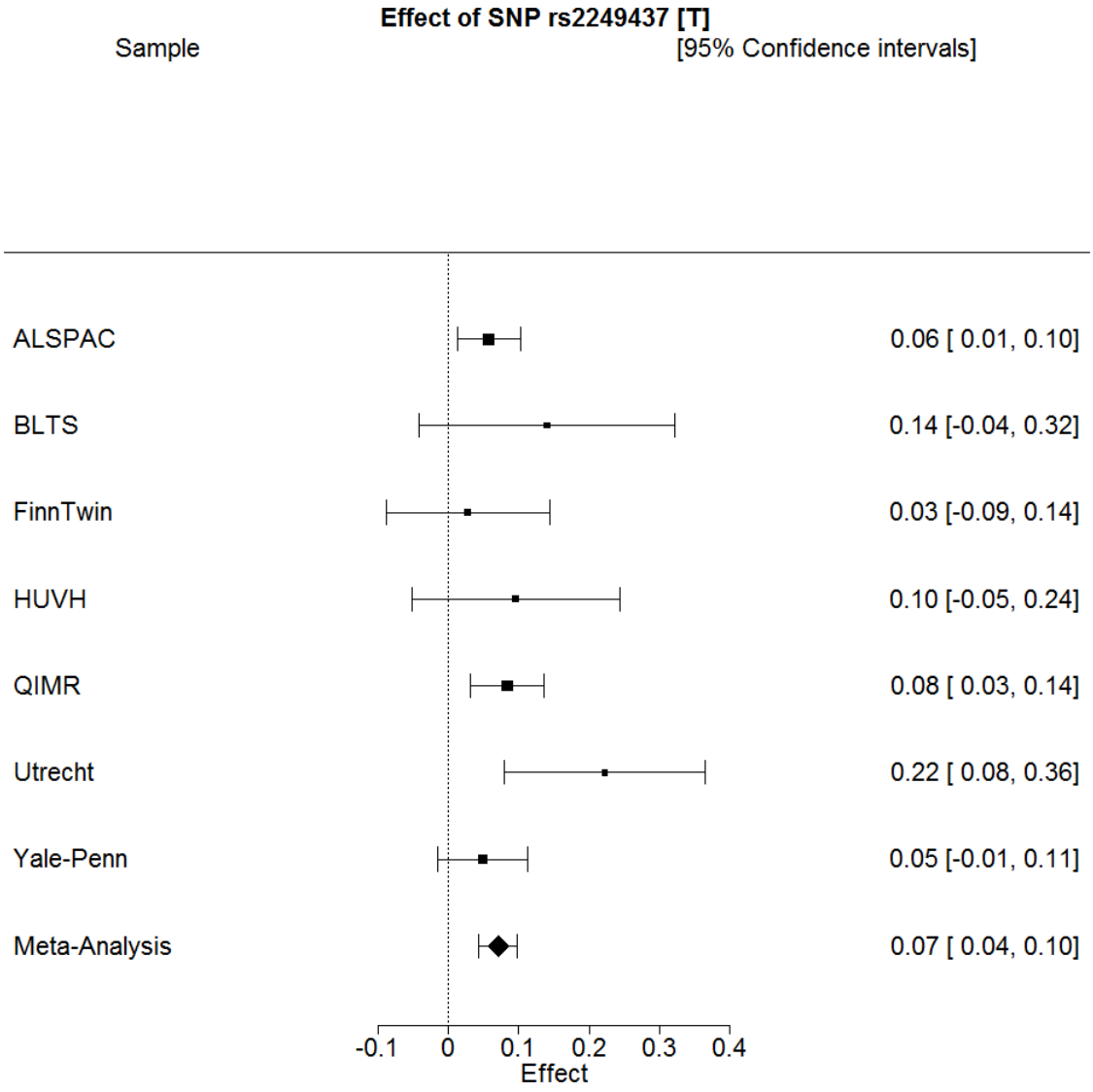
Figure S5 a-k. Forest plots top SNPs

These plots compare the effect-sizes and 95% confidence intervals for the top SNPs in the individual discovery cohorts as well as the meta-analysis. See main manuscript for forest plot of the top SNP rs1574587.

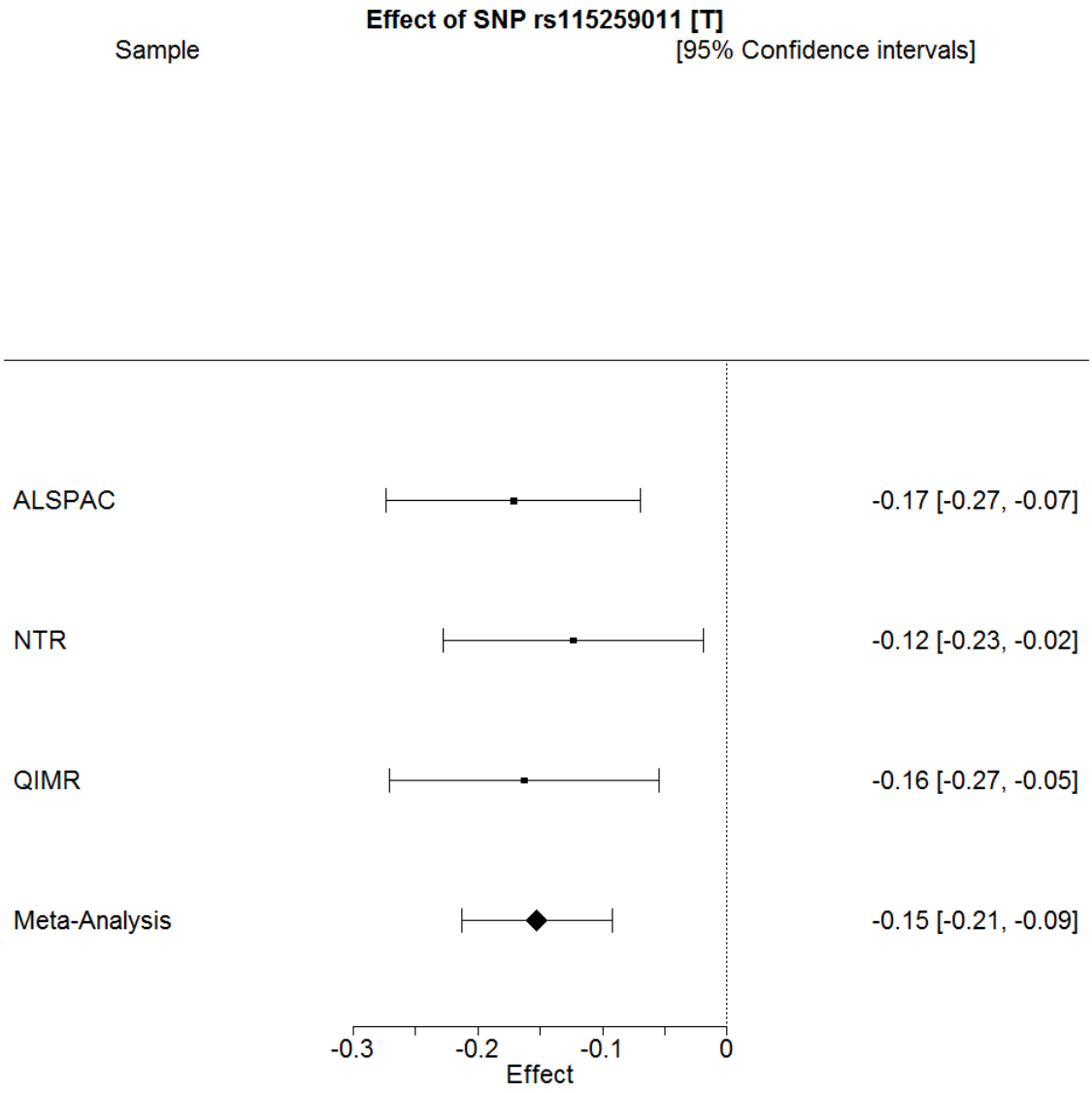
a.



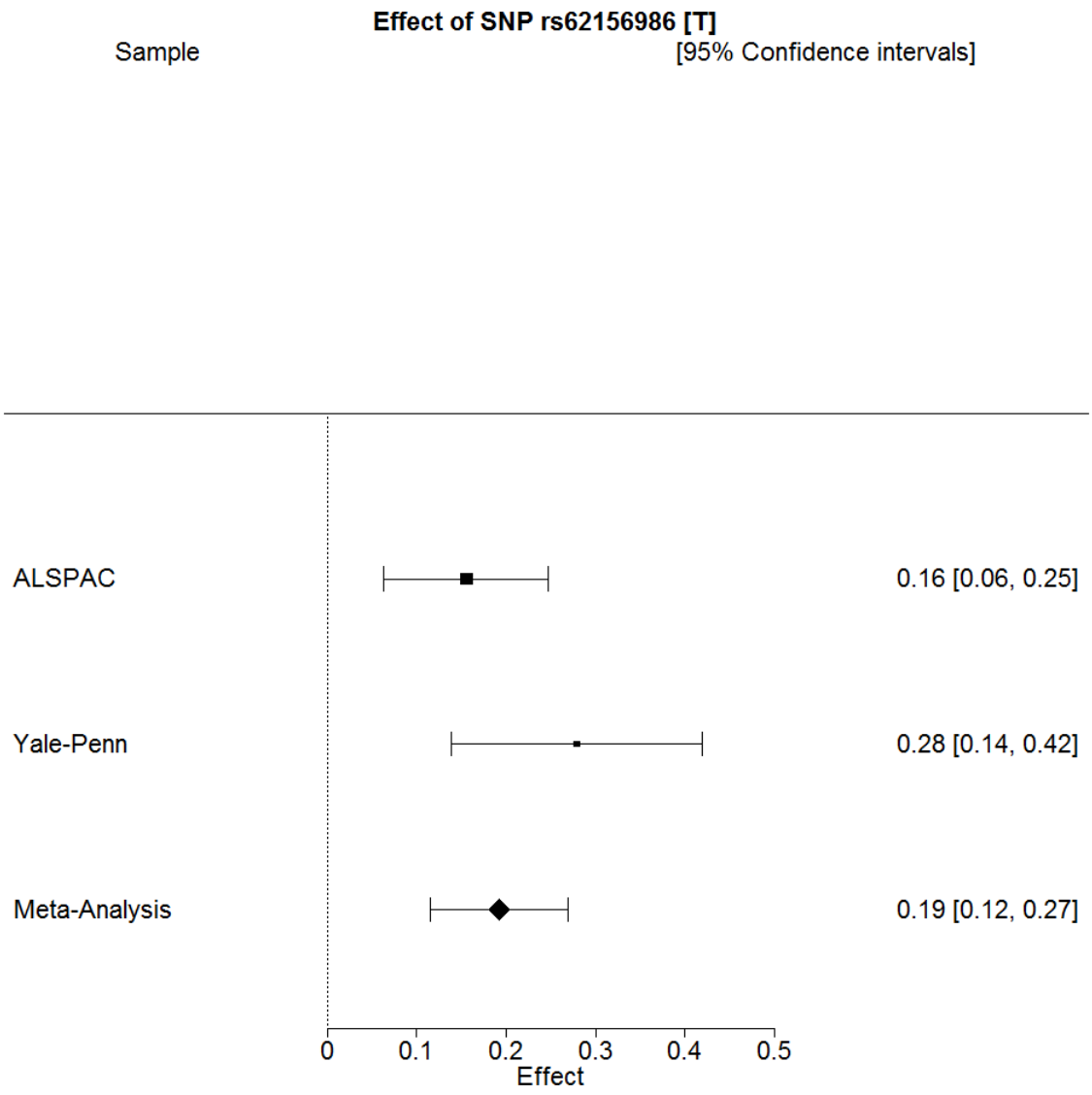
b.


Effect of SNP rs2249437 [T]
Sample
[95% Confidence intervals]
ALSPAC
0.06 [0.01, 0.10]
BLTS
0.14 [-0.04, 0.32]
FinnTwin
0.03 [-0.09, 0.14]
HUVH
0.10 [-0.05, 0.24]
QIMR
0.08 [0.03, 0.14]
Utrecht
0.22 [0.08, 0.36]
Yale-Penn
0.05 [-0.01, 0.11]
Meta-Analysis
0.07 [0.04, 0.10]
-0.1
0
0.1
0.2
0.3
0.4
Effect

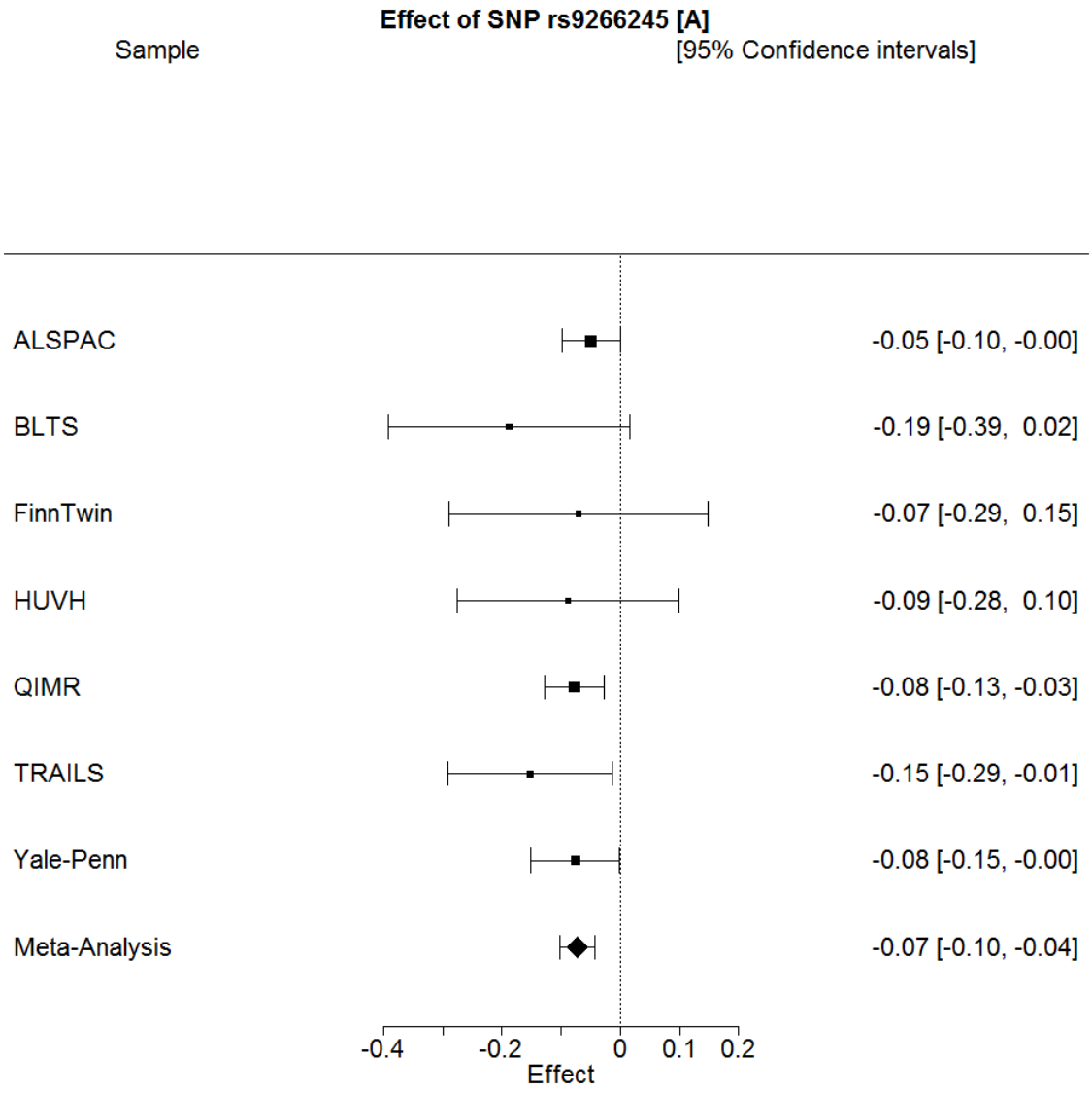
c.


Effect of SNP rs115259011 [T]
Sample
[95% Confidence intervals]
ALSPAC
-0.17 [-0.27, -0.07]
NTR
-0.12 [-0.23, -0.02]
QIMR
-0.16 [-0.27, -0.05]
Meta-Analysis
-0.15 [-0.21, -0.09]
-0.3
-0.2
-0.1
0
Effect

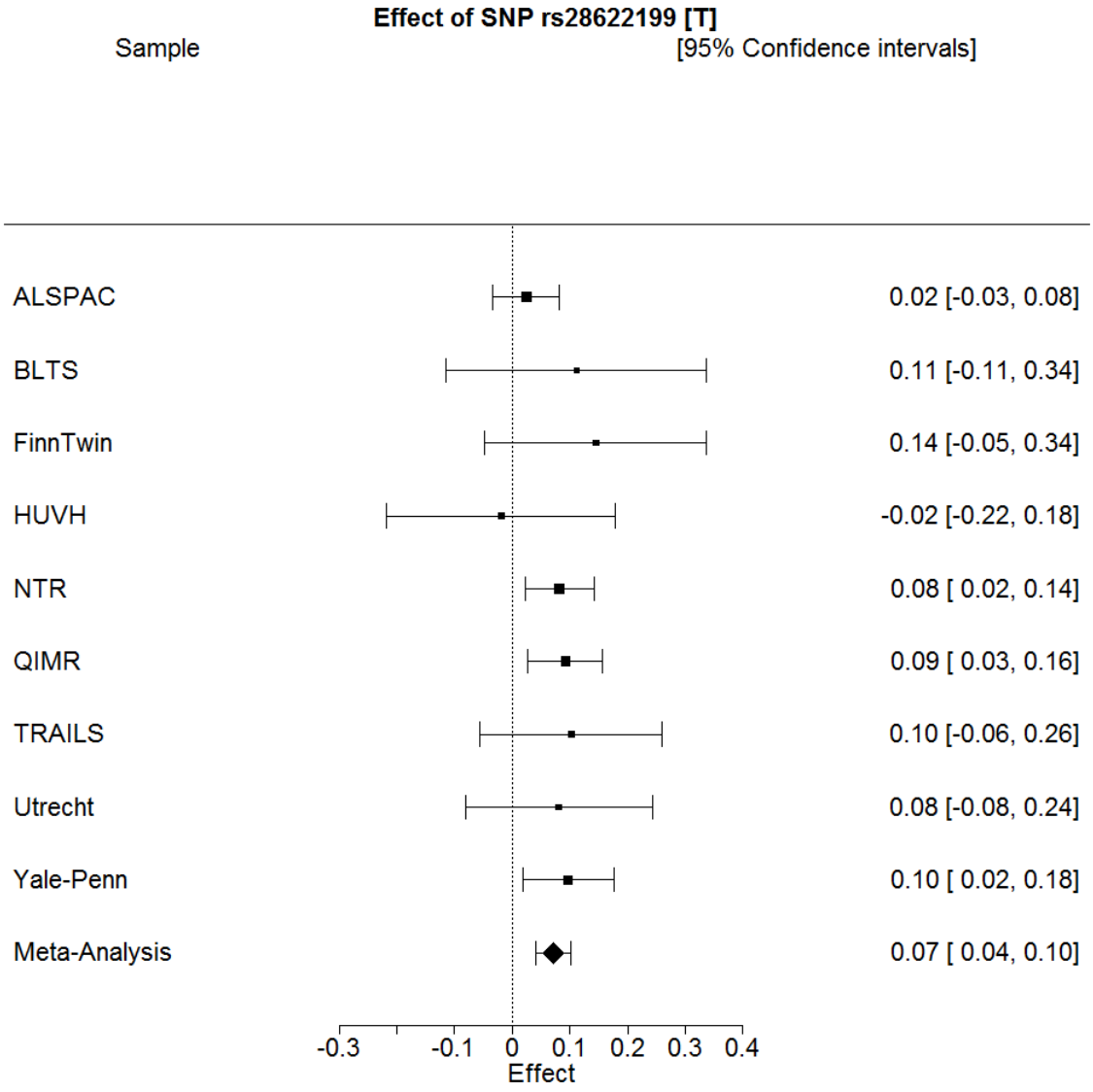
d.


Effect of SNP rs62156986 [T]
Sample
[95% Confidence intervals]
ALSPAC
0.16 [0.06, 0.25]
Yale-Penn
0.28 [0.14, 0.42]
Meta-Analysis
0.19 [0.12, 0.27]
0
0.1
0.2
0.3
0.4
0.5
Effect

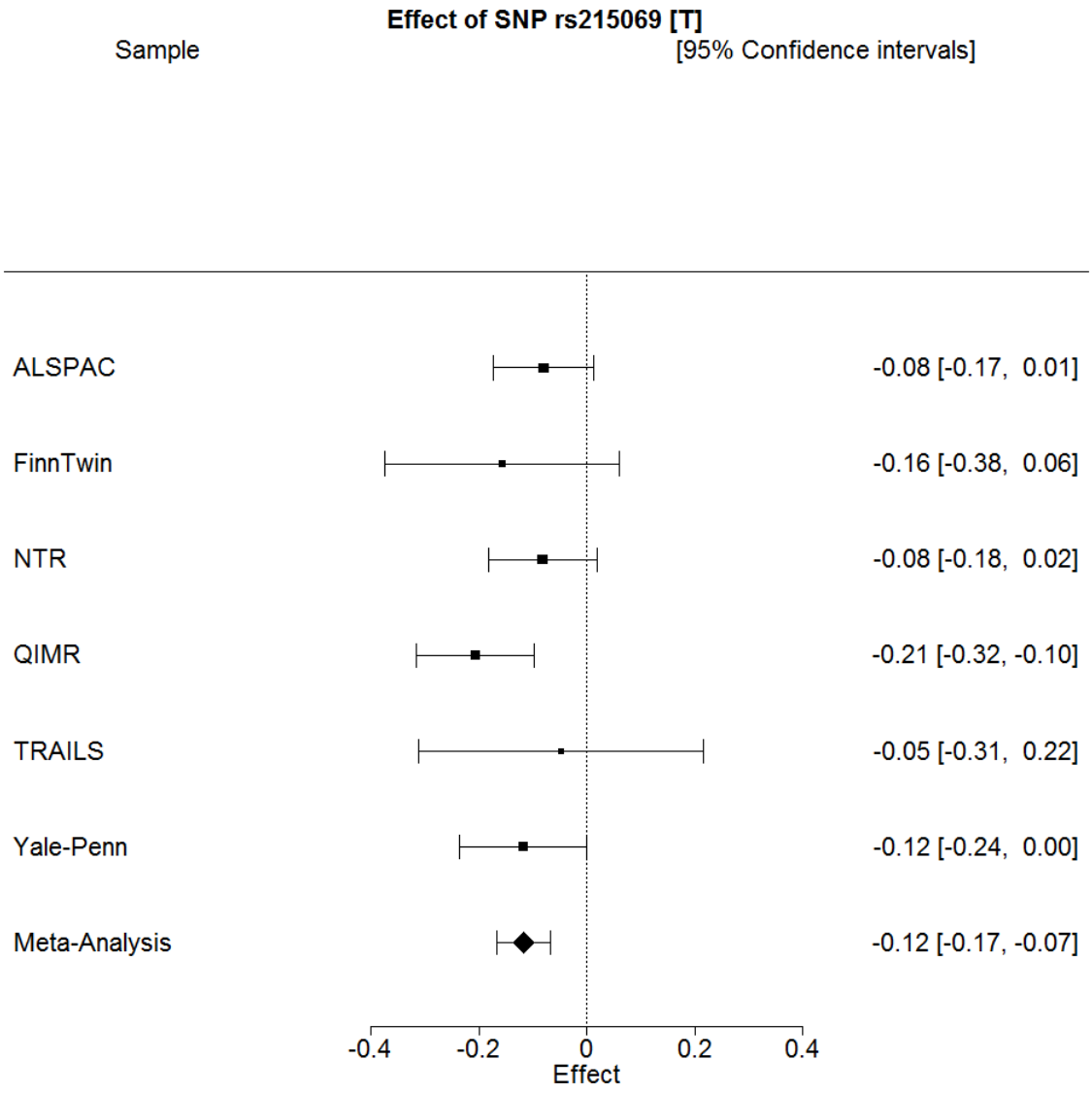
e.


Effect of SNP rs9266245 [A]
Sample
[95% Confidence intervals]
ALSPAC
-0.05 [-0.10, -0.00]
BLTS
-0.19 [-0.39, 0.02]
FinnTwin
-0.07 [-0.29, 0.15]
HUVH
-0.09 [-0.28, 0.10]
QIMR
-0.08 [-0.13, -0.03]
TRAILS
-0.15 [-0.29, -0.01]
Yale-Penn
-0.08 [-0.15, -0.00]
Meta-Analysis
-0.07 [-0.10, -0.04]
-0.4
-0.2
0
0.1
0.2
Effect

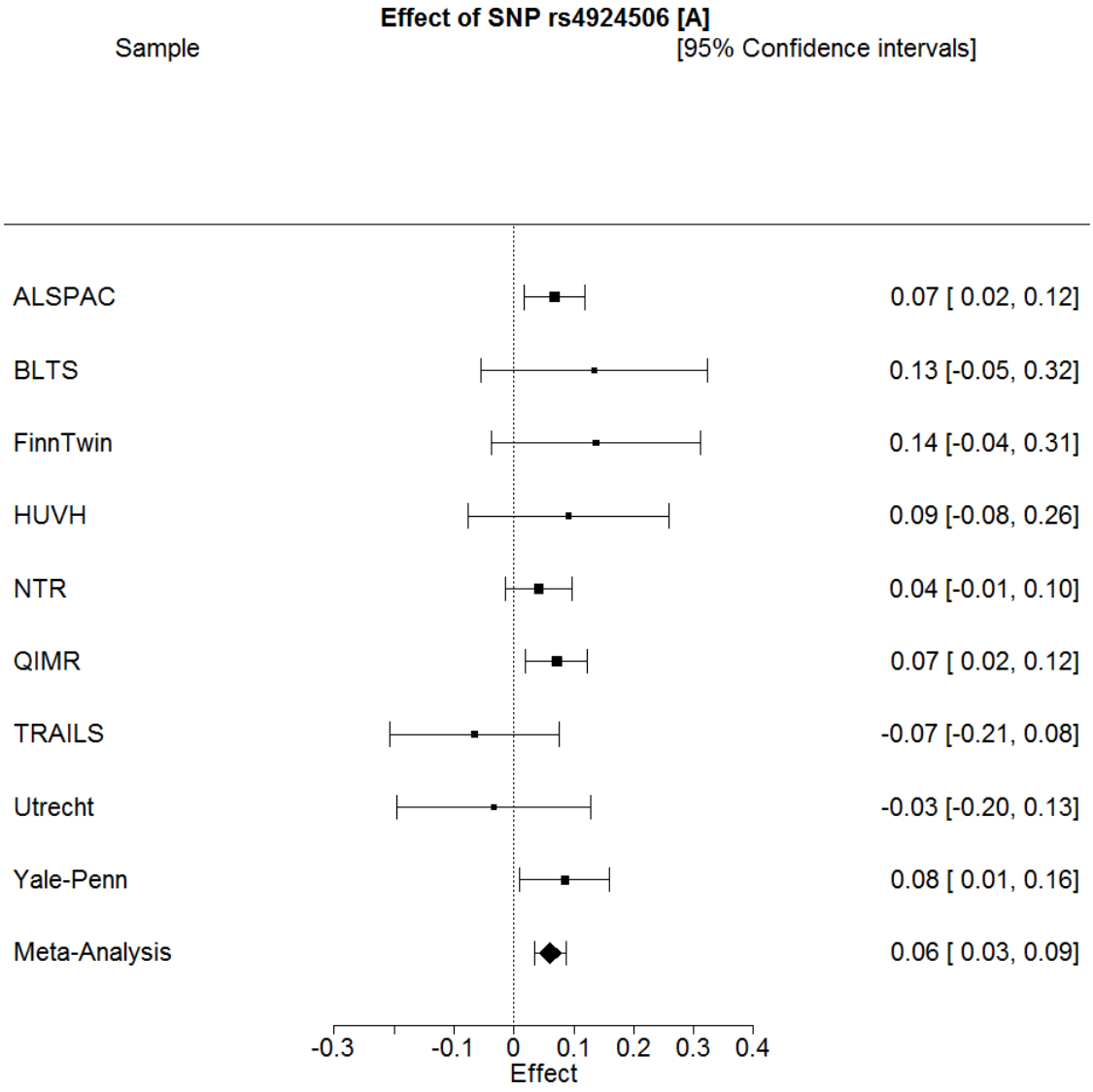
f.


Effect of SNP rs28622199 [T]
Sample
[95% Confidence intervals]
ALSPAC
0.02 [-0.03, 0.08]
BLTS
0.11 [-0.11, 0.34]
FinnTwin
0.14 [-0.05, 0.34]
HUVH
-0.02 [-0.22, 0.18]
NTR
0.08 [0.02, 0.14]
QIMR
0.09 [0.03, 0.16]
TRAILS
0.10 [-0.06, 0.26]
Utrecht
0.08 [-0.08, 0.24]
Yale-Penn
0.10 [0.02, 0.18]
Meta-Analysis
0.07 [0.04, 0.10]
-0.3
-0.1
0
0.1
0.2
0.3
0.4
Effect

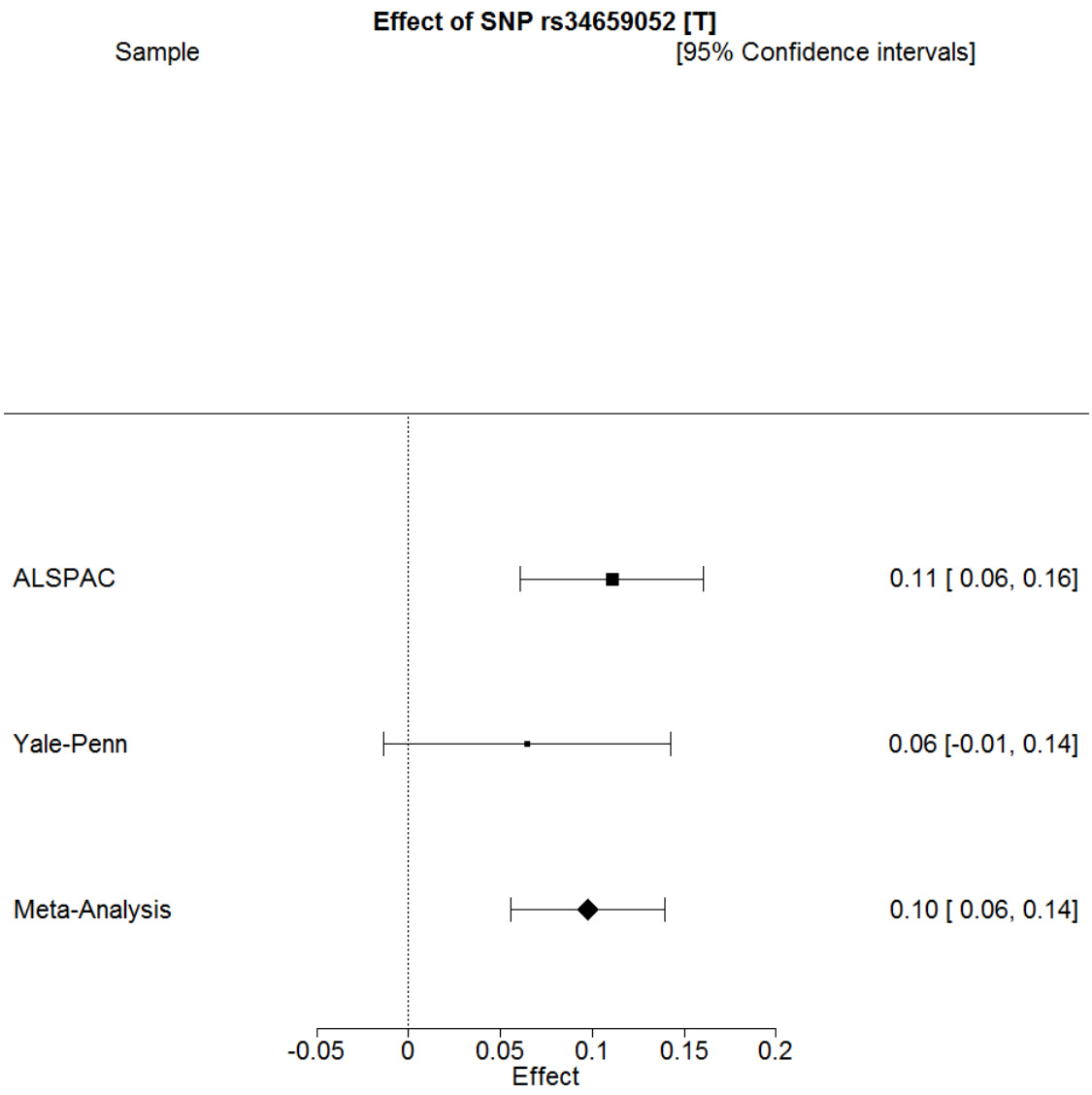
g.


Effect of SNP rs215069 [T]
Sample
[95% Confidence intervals]
ALSPAC
-0.08 [-0.17, 0.01]
FinnTwin
-0.16 [-0.38, 0.06]
NTR
-0.08 [-0.18, 0.02]
QIMR
-0.21 [-0.32, -0.10]
TRAILS
-0.05 [-0.31, 0.22]
Yale-Penn
-0.12 [-0.24, 0.00]
Meta-Analysis
-0.12 [-0.17, -0.07]
-0.4
-0.2
0
0.2
0.4
Effect

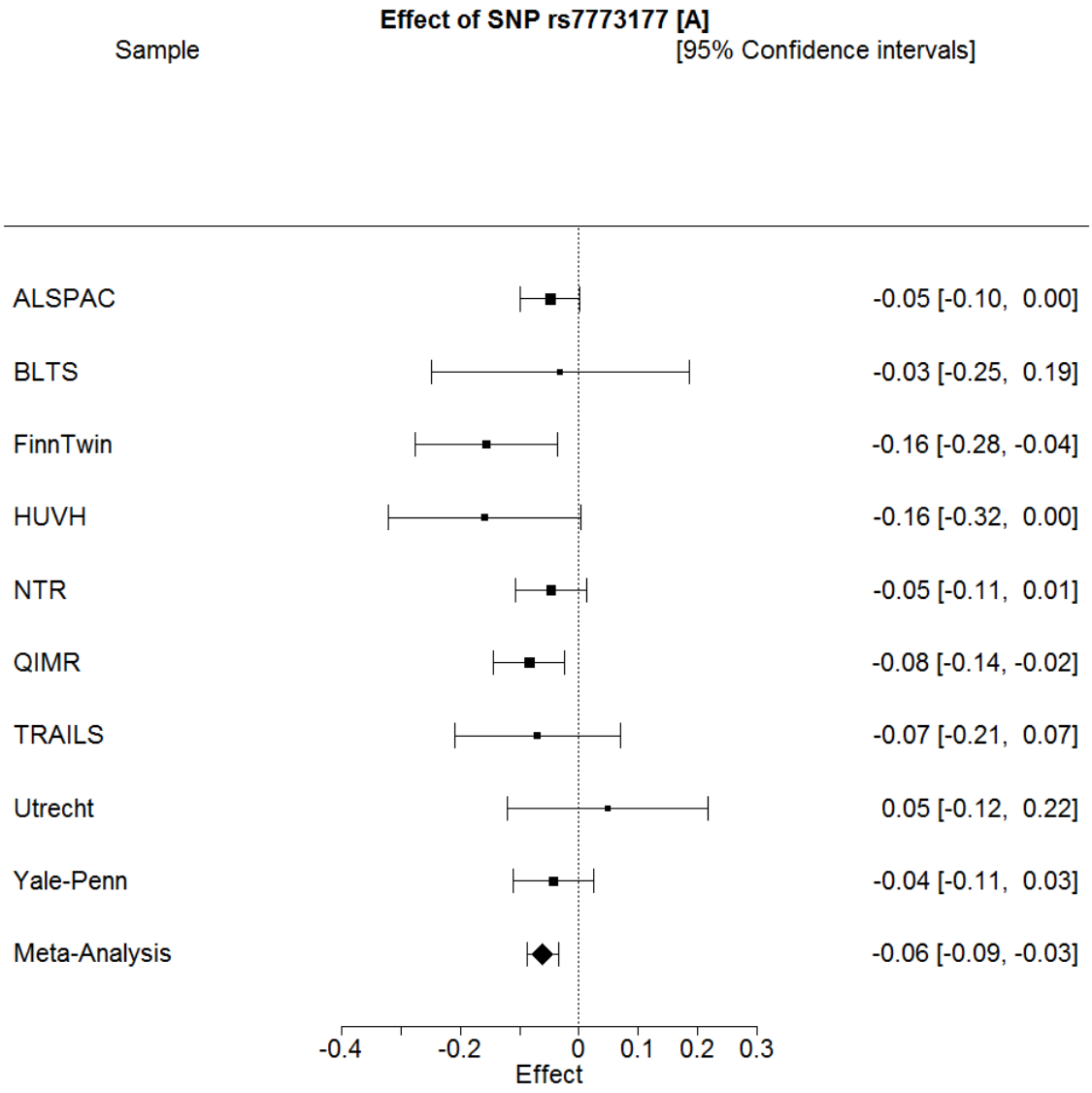
h.


Effect of SNP rs4924506 [A]
Sample
[95% Confidence intervals]
ALSPAC
0.07 [0.02, 0.12]
BLTS
0.13 [-0.05, 0.32]
FinnTwin
0.14 [-0.04, 0.31]
HUVH
0.09 [-0.08, 0.26]
NTR
0.04 [-0.01, 0.10]
QIMR
0.07 [0.02, 0.12]
TRAILS
-0.07 [-0.21, 0.08]
Utrecht
-0.03 [-0.20, 0.13]
Yale-Penn
0.08 [0.01, 0.16]
Meta-Analysis
0.06 [0.03, 0.09]
-0.3
-0.1
0
0.1
0.2
0.3
0.4
Effect

i.


Effect of SNP rs34659052 [T]
Sample
[95% Confidence intervals]
ALSPAC
0.11 [0.06, 0.16]
Yale-Penn
0.06 [-0.01, 0.14]
Meta-Analysis
0.10 [0.06, 0.14]
-0.05
0
0.05
0.1
0.15
0.2
Effect

j.


Effect of SNP rs7773177 [A]
Sample
[95% Confidence intervals]
ALSPAC
-0.05 [-0.10, 0.00]
BLTS
-0.03 [-0.25, 0.19]
FinnTwin
-0.16 [-0.28, -0.04]
HUVH
-0.16 [-0.32, 0.00]
NTR
-0.05 [-0.11, 0.01]
QIMR
-0.08 [-0.14, -0.02]
TRAILS
-0.07 [-0.21, 0.07]
Utrecht
0.05 [-0.12, 0.22]
Yale-Penn
-0.04 [-0.11, 0.03]
Meta-Analysis
-0.06 [-0.09, -0.03]
-0.4
-0.2
0
0.1
0.2
0.3
Effect

k.

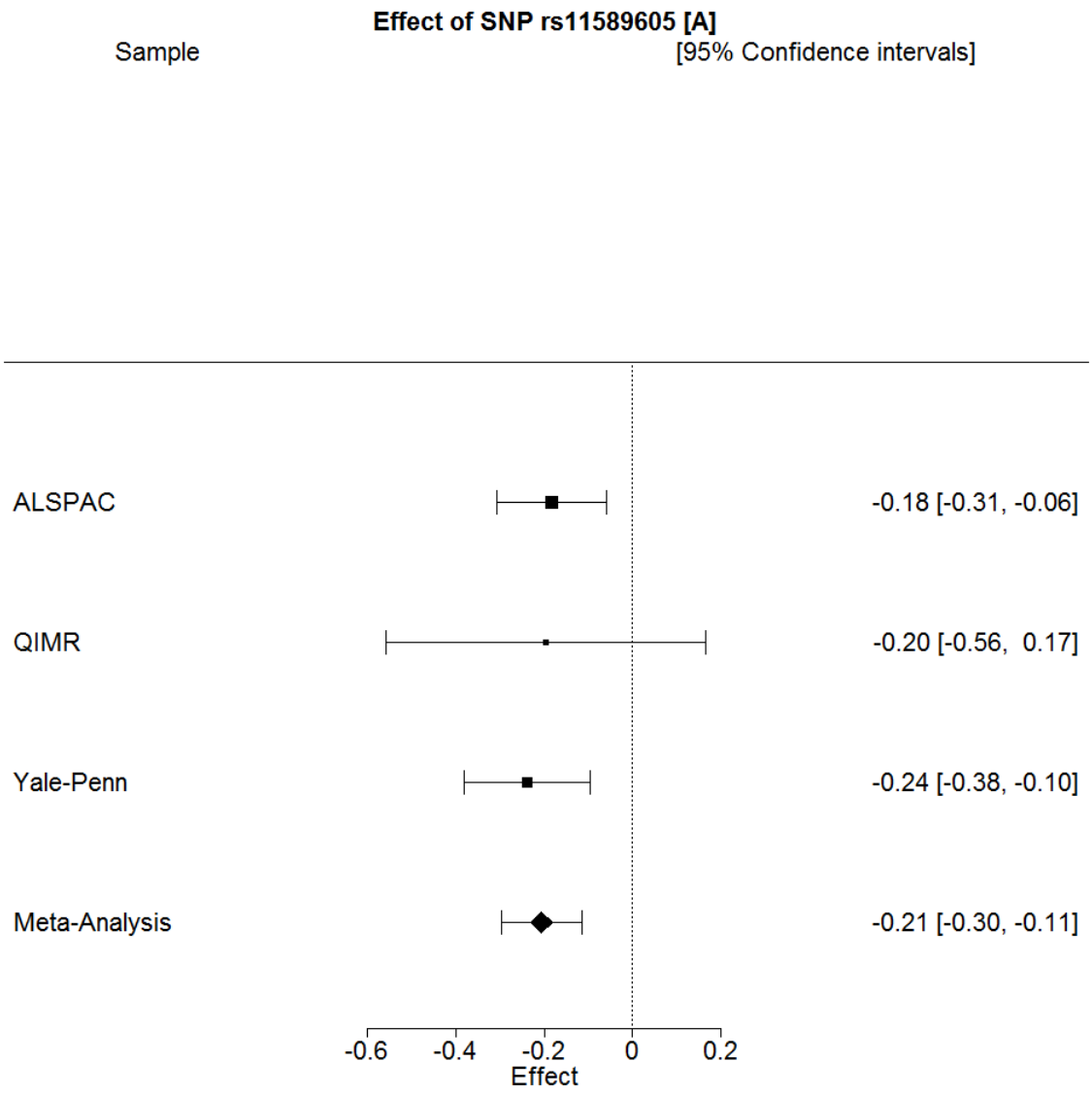
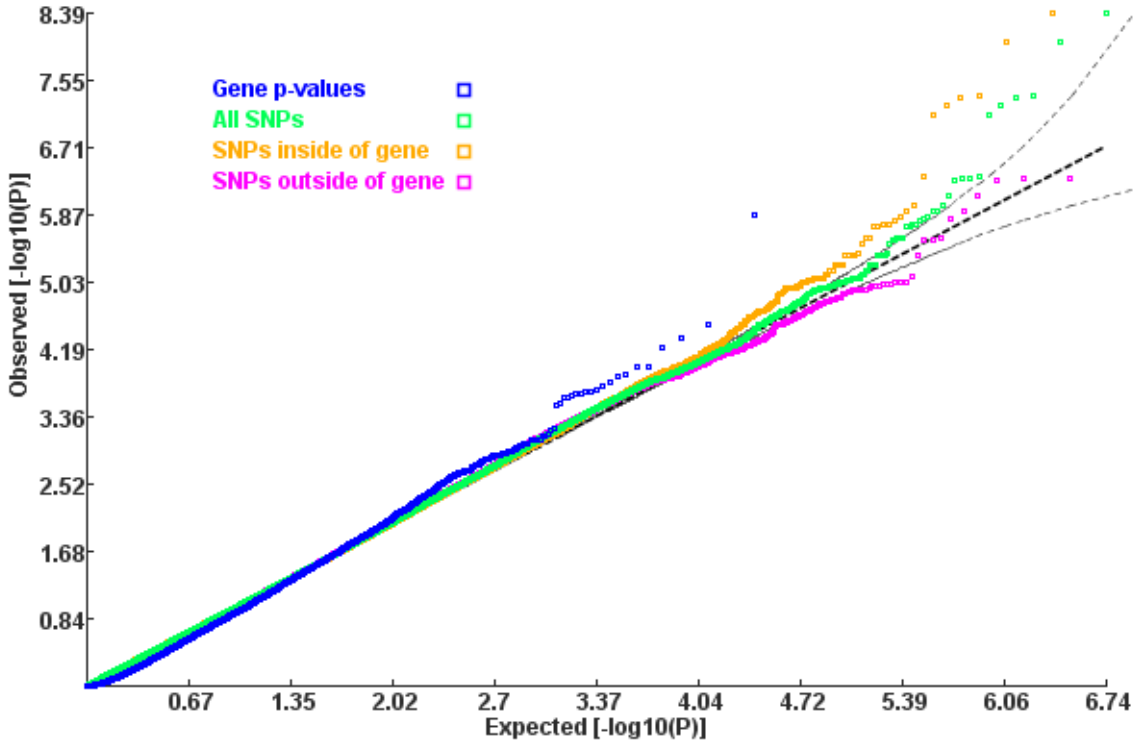

Effect of SNP rs11589605 [A]
Sample
[95% Confidence intervals]
ALSPAC
-0.18 [-0.31, -0.06]
QIMR
-0.20 [-0.56, 0.17]
Yale-Penn
-0.24 [-0.38, -0.10]
Meta-Analysis
-0.21 [-0.30, -0.11]
-0.6
-0.4
-0.2
0
0.2
Effect

Figure S6. The quantile-quantile (QQ) plot of the gene-based test



Supplementary Table S1. Characteristics of individual samples

Study Name	Characteristics of sample used for lifetime cannabis use GWA a		
	N	% ever used cannabis	N ever users
ALSPAC	6,147	38.4	2360
BLTS	721	59.5	428
FinnTwin12 & FinnTwin16	1,029	27.5	283
HUVH-Barcelona	581	30.3	176
NTR	5,148	16.6	852
QIMR	6,758	51.3	3469
TRAILS population cohort	1,249	61.7	771
Utrecht	958	59.0	565
Yale Penn EA	2,362	92.6	2187
CADD	1,060	78	827
NTR2-RADAR (combined)	2,082	27.9	581
NTR2	1740	22.2	386
RADAR	342	57.0	195
Saguenay Youth Studie (SYS)	593	47.7	283

analyses N never users	Mean age at initation (sd) (in users)	Range age at initiation (in users)
3787	14.8 (1.6)	7-19
293	18.8 (2.8)	10-32
746	18.0 (2.5)	13-25
405	16.0 (3.0)	9-32
4296	18.99 (5.1)	12-72
3289	19.9 (5.8)	8-23
478	16.3 (2.0)	13-22
393	15.5 (2.1)	11-23
175	17.0 (9.4)	5-76
233	14.3 (3.2)	6-26
1501	17.3 (3.5)	10-47
1354	18.0 (4.0)	11-47
147	15.9 (1.7)	10-20
310	16.2 (1.7)	12-18

% female	N females	mean age (sd)	age range	Birth cohorts
51.91	3191	17.3 (1.73)	12-20	1991-1992
57.1	411	26.2 (3.3)	18-33	1979-1993
51.7	532	22.8 (1.3)	20-29	1974-1987
31.3	182	28.7 (12.5)	9-66	1944-1992
				reference category: 1944-1963
				dummy 1: 1964-1983
				dummy 2: 1984-1992
62.3	3205	46.9 (17.5)	16-99	1915 - 1998
				reference category: up to 1950
				dummy 1: 1951-1970
				dummy 2: 1971-1998
53.8	3638	45.2 (10.9)	18-85	1917-1986
				reference category: up to 1950
				dummy 1: 1951-1970
				dummy 2: 1971-1986
53.88	673	20.0 (1.6)	18-24	1989-1991
51.3	491	17.4 (3.2)	18-36	1970-1991
41.2	977	38.2 (10.6)	16-76	reference category: 1921-1940
				1941-1960
				1961-1980
40.18	426	24.06 (4.2)	13-36	1976-1994
60.6	1262	32.4 (14.5)	12-79	1928-1995
				reference category: up to 1960
				dummy 1: 1961-1980
				dummy 2: 1981 and later
63.7	1109	35.0 (14.6)	12-79	
44.7	153	19.5 (0.8)	13-22	
55.3	328	49.4 (5.1)	37-65	1947-1977

Genome wide association analysis	
software	covariates
Plink	Sex
	4 Principal components
Plink	Sex
	4 principal components
Plink	Sex
	4 principal components
plink	Sex
	Birth Cohort (2 dummy variables)
	4 Principal components
Plink	Sex
	Birth Cohort (2 dummy variables)
	4 principal components
	Note: The sample included an on extre
	covariates were used to deal with the c
Plink	Sex
	Birth Cohort (2 dummy variables)
	4 principal components
Plink	Sex
	4 principal components
Plink	Sex
	4 principal components
	Genotype array
Plink	Sex
	4 Principal components
	Birth Cohort (2 dummy variables)
Plink	Sex
	4 principal components
Plink	Sex
	Birth Cohort (2 dummy variables)
	3 principal components
	Corrections were made to include fami
Plink	Birth cohort (dummy)
	4 principal components
	Sex

Reference (if available)
http://www.ncbi.nlm.nih.gov/pubmed/22507743
http://www.ncbi.nlm.nih.gov/pubmed/23187020
http://onlinelibrary.wiley.com/doi/10.1080/00049530410001734865/abstract
http://www.ncbi.nlm.nih.gov/pubmed/18455885
http://www.ncbi.nlm.nih.gov/pubmed/17254406
http://www.ncbi.nlm.nih.gov/pubmed/23298696
PubMed PMID: 25284319
PubMed PMID: 24269040
http://www.ncbi.nlm.nih.gov/pubmed/21489006
http://www.ncbi.nlm.nih.gov/pubmed/19793625
http://www.ncbi.nlm.nih.gov/pubmed/23186620
http://www.ncbi.nlm.nih.gov/pubmed/23298648
different input genotype chips (basically a stratified analysis). Also corrections were made to include family m
http://www.ncbi.nlm.nih.gov/pubmed/22823124
http://www.ncbi.nlm.nih.gov/pubmed/15724882
http://www.ncbi.nlm.nih.gov/pubmed/11825135
http://www.ncbi.nlm.nih.gov/pubmed/21529783
https://www.ncbi.nlm.nih.gov/pubmed/25431468
ly members using the --family option in Plink.
http://www.ncbi.nlm.nih.gov/pubmed/23348558
http://www.ncbi.nlm.nih.gov/pubmed/25466800
http://www.ncbi.nlm.nih.gov/pubmed/23982435
http://www.ncbi.nlm.nih.gov/pubmed/22181711

embers using the 'cluster' option. 1 of a MZ pair was excluded prior to the analyses.

Supplementary Table S2. Genotyping and imputation

Study Name
ALSPAC
BLTS
FinnTwin12 & FinnTwin16
HUVH-Barcelona (HUVH)
NTR
QIMR
TRAILS population cohort
Utrecht sample
Yale Penn EA
CADD

NTR-RADAR
Saguenay Youth Study (SYS)

information per sample

Sample QC
call rate
<95%
≤98%
<95%
<95%
>95%
At least all with ≤95%; generally stricter (eg. ≤98%)
< 95%
< 95%
≤98%
<95%

<90%, most are around 0.98

<95%

Other exclusion criteria
Excessive or minimal heterozygosity Cryptic relatedness as measured by IBD Sample duplication Non-European Ancestry Missing phenotype information Sex discrepancy with genetic data from X-linked markers Autosomal chromosome abberations
Sex discrepancy with genetic data from X-linked markers Unexpected relatives
Sex discrepancy with genetic data from X-linked markers Duplicates Heterozygosity Unexpected relatives Missing phenotype information
gender discrepancy with genetic data from X-linked markers duplicates and first/second degree relatives Log R ratio (LRR) < 0,30
sex discrepancy with genetic data from X-linked markers duplicates and first/second degree relatives ancestry outliers heterozygosity autosomal chromosome abberations missing phenotype information
Ancestry outliers [outside 6sd from European mean in PCA analysis - 1st two PCs] missing phenotype information Wrong gender
heterozygosity >4SD duplicates sex mismatch non-caucasian missing phenotype information
heterozygosity >3SD duplicates/relatedness sex mismatch non-caucasian missing phenotype information genotyping platform
sex discrepancy with genetic data from X-linked markers
ID labeled incorrectly Removed older samples (18) because it was too small to be used as a reference cohort. Removed those initiated at age < 10. Removed those who only donated DNA sample.

- Heterozygosity F<-0.1 & F>0.1
- IBD and sex mismatch
- CQC<0.4
- Ethnic outliers
duplicates
sex discrepancy with genetic data
ancestry outliers (non-European)
heterozygosity
missing phenotype information
duplicates

Genotyping QC
Platform
Illumina HumanHap550
Illumina 610K-quad
Illumina 670K
Illumina HumanOmni1-Quad
Affymetrix 6.0, Perlegen-Affymetrix 5.0, Illumina 660, Illumina 1M, Illumina 370K
Illumina 317K Illumina HumanCNV370-Quadv3 Illumina HumanCNV370-CNV370v1 Illumina Human610-Quad
Illumina Cyto SNP12 v2
Illumina HumanOmniExpress (733,202 SNPs, 576 individuals) Illumina Human610-Quad Beadchip (620,901 SNPs, 768 individuals)
Illumina HumanOmni1-Quad v1.0
Affymetrix 6.0

Affymetrix 6.0
Illumina Human610W-Quad Beadchip (Quad), Illumina HumanOmniExpress BeadChip (Omni)

Genotype calling	exclusion criteria: MAF	Call rate	P HWE
GenomeStudio	<1%	<95%	<5x10e-7
GenomeStudio	<1%	≤98%	<10e-6
GenomeStudio	<1%	<95%	<1x10-6
Beadstudio	<1%	<99%	1.00E-06
Birdseed + Affymetrix Genotyper	<1%	<0.90, per	<10e-4
GenomeStudio genotyping module	<1%	≤95%	<1e-6
Illumina GenomeStudio	<1%	<95%	<0.0001
Illumina GenomeStudio	0.01 0.01	0.95 0.95	1.00E-06 1.00E-06
GenomeStudio software V2011.1 and genotyping module V1.8.4	<1%	≤98%	<10e-4
GeneChip Targeted Genotyping Analysis Software	<1%	>99%	1.00E-06

Birdseed 2.6, Affymetrix APT 3.3	0	<0.95	<10e-6
GenomeStudio	<1%	<0.95 per c	<10e-6

	SNPs passing QC
other	
	500527
	529269
	6728589
	780165
Mendelian < 2%	6379086
mean GenCall score <0.7	273158
[for imputation purposes]	
chr X SNPs >1% heterozygous in men	255254
	277957
overlap after merging the three datasets	
	889,659
Removed mendel errors as identified by PI	541,445

> 1% inconsistent calls in controls	733,971
>5sd Mendel errors (1%)	
	542,345 (Quad); 644,272 (Omni); 313,653 (ove

	Imputation
Genomic control lambdaGC	Software
lambdaGC=0.96	minimac
lambdaGC=1.085	Minimac
lambdaGC=1.173	IMPUTE 2.2.2
lambdaGC=1.116	BEAGLE
lambdaGC=1.087	Mach / minimach
lambdaGC=0.802	MACH (phasing); minimac (imputation)
lambdaGC=0.871	Impute v2
lambdaGC=0.996	MACH (phasing); minimac (imputation)
lambdaGC=1.028	Impute2
lambdaGC=1.118	miniMACH

lambdaGC=1.038	Shapeit (Phasing) Impute2 (Imputation)
lambdaGC=1.221	SHAPEIT/IMPUTE2

Reference Panel
1000 genome, release March 2012
1000 Genomes project uses NCBI Build 37 ('GIANT' marker subset - no SNPs with <2 copies of minor allele) of
1000 Genomes 2012 March release
1000 genome, release March 2012
March 2012 1000G release, GIANT ALL panel
1000 Genome, 20101123 v3 [GIANT subset, ie. no markers with <2 copies of minor allele in reference panel]]
1000 genomes, release march 2012
1000 genome, release March 2012
1000 genome, release March 2012
1000 genome, phase 1, release March 2012

1000G Phase 1, june 2014
March 2012 1000G release (EUR reference panel)

Reference (if available)

20101123 v3

http://www.ncbi.nlm.nih.gov/pubmed/23298648

http://www.ncbi.nlm.nih.gov/pubmed/18263649

http://www.ncbi.nlm.nih.gov/pubmed/24166409

https://imputationserver.sph.umich.edu/index.html

Supplemental Table S3: Top 100 SNPs in the discovery meta-analysis.

SNP	CHR	BP	Allele1	Allele2	Freq1	FreqSE
rs1574587	16	84453056	t	c	0.1415	0.0111
rs12922606	16	84453352	a	g	0.8585	0.0113
rs11644628	16	84452597	t	c	0.1431	0.0113
rs11644673	16	84452771	a	g	0.8626	0.0109
rs11644663	16	84452541	a	g	0.1465	0.0102
rs12922477	16	84453332	a	c	0.8598	0.0118
rs79927873	16	84452497	a	c	0.1392	0.0099
rs1008994	16	84450857	c	g	0.1454	0.0118
rs4935127	10	56654986	c	g	0.7741	0.0166
rs1733786	10	56681617	a	g	0.7742	0.0196
rs2249437	6	1595216	t	c	0.4595	0.0184
rs115259011	3	161789904	t	g	0.9572	0.0043
rs62156986	2	120072326	t	g	0.9349	0.0018
rs1349893	10	56701951	t	c	0.7658	0.0191
rs11643072	16	84451155	a	g	0.1475	0.0109
rs3943846	16	84455781	a	g	0.8146	0.0146
rs62159383	2	120045513	t	c	0.9347	0.0003
rs2163036	16	84455766	t	c	0.173	0.0135
rs9266245	6	31325702	a	g	0.2655	0.0313
rs9266262	6	31325932	a	g	0.7251	0.0279
rs9266244	6	31325692	a	g	0.7345	0.0312
rs141294240	6	31325822	a	g	0.7296	0.0319
rs1733762	10	56697898	a	g	0.2167	0.0193
rs28622199	8	5392103	t	c	0.8012	0.009
rs1670812	10	56689178	t	c	0.2164	0.0195
rs2523578	6	31328542	a	g	0.7333	0.0258
rs8045313	16	84455540	t	g	0.8158	0.0142
rs215069	16	16091237	t	c	0.0687	0.0058
rs1733763	10	56697536	a	c	0.7839	0.0196
rs55966520	16	84454043	a	g	0.1636	0.0119
rs2523582	6	31328092	a	g	0.2632	0.032
rs59006942	16	84454029	a	g	0.1632	0.011
rs71386833	16	84454170	a	g	0.1636	0.0118
rs4924506	15	41129467	a	c	0.7318	0.017
rs34659052	5	148816223	t	c	0.7351	0.0061
rs689589	15	41139250	t	g	0.258	0.0154
rs647930	15	41141459	a	g	0.7218	0.0208
rs2412569	15	41140159	a	g	0.254	0.0139
rs114529675	2	120136433	t	c	0.0614	0.0029
rs2395475	6	31326920	a	g	0.6563	0.0221
rs2917953	15	41131916	t	c	0.2558	0.0141
rs690660	15	41139165	t	c	0.254	0.014
rs7773177	6	139143088	a	g	0.7383	0.018
rs2326270	16	84461051	a	c	0.0994	0.01
rs11639292	15	41129528	a	g	0.2601	0.0153
rs668750	15	41135827	a	t	0.7415	0.0157

rs689618	15	41133008	t	c	0.2558	0.0142
rs11589605	1	230084670	a	t	0.9599	0.0072
rs2412570	15	41140168	t	c	0.2555	0.0141
rs12193938	6	139142855	c	g	0.7383	0.0179
rs16850641	1	230097368	a	g	0.9596	0.0075
rs7528099	1	230097883	c	g	0.9596	0.0075
rs60064513	1	230098302	t	g	0.0404	0.0075
rs12413522	10	56642064	t	c	0.7838	0.0149
rs73113155	1	230087856	a	g	0.0401	0.0072
rs452277	16	16079948	t	c	0.9385	0.0071
rs435683	16	16079952	t	c	0.9385	0.0071
rs6570291	6	139143274	a	g	0.7383	0.0181
rs10872495	6	139155677	t	g	0.2614	0.0178
rs11155012	6	139151784	a	g	0.2614	0.0177
rs11004618	10	56641728	t	c	0.7836	0.0148
rs11155013	6	139151838	t	c	0.2866	0.0197
rs6683692	1	230089628	a	g	0.0401	0.0072
rs116395818	1	230091683	t	c	0.0401	0.0072
rs2523595	6	31326618	a	g	0.6583	0.0229
rs690572	15	41135406	t	c	0.2557	0.0144
rs1865255	6	40407132	a	g	0.1029	0.0157
rs56169108	16	84453872	a	g	0.8376	0.012
rs116815096	6	30997692	c	g	0.7014	0.0232
rs119774	16	16086833	t	c	0.0695	0.0051
rs2523594	6	31326629	t	c	0.6587	0.0229
rs622538	15	41133744	c	g	0.738	0.0167
rs7459013	7	10887478	a	c	0.5912	0.0169
rs12193993	6	139142909	t	c	0.2611	0.0176
rs12207788	6	139157314	t	g	0.7399	0.0175
rs114645619	1	230092054	c	g	0.0401	0.0072
rs6541302	1	230093400	c	g	0.9599	0.0071
rs689736	15	41136030	t	g	0.258	0.0154
rs11800137	1	230096121	t	c	0.96	0.0071
rs16841453	1	230095259	t	g	0.9599	0.0071
rs57801654	1	230095573	a	g	0.0401	0.0071
rs11122468	1	230096338	a	t	0.0401	0.0071
rs7037098	9	83219719	c	g	0.5246	0.0214
rs4714345	6	40406829	c	g	0.8974	0.0159
rs10453983	10	130604719	t	g	0.8304	0.0109
rs12205070	6	139154247	a	g	0.2599	0.0176
rs680493	15	41136116	a	g	0.2609	0.0163
rs670321	1	108648369	a	g	0.4473	0.0223
rs10872496	6	139158676	t	g	0.2615	0.0176
rs1371884	1	108644530	a	g	0.444	0.0215
rs2326269	16	84454552	t	c	0.1679	0.0121
rs215072	16	16085551	a	g	0.0707	0.0047
rs115283957	6	31276435	c	g	0.7147	0.025
rs476769	1	108651825	t	g	0.5532	0.0222

rs1968096	7	10885886	t	c	0.4073	0.0157
rs16850649	1	230099183	a	g	0.0402	0.0073
rs1454487	10	56671487	a	g	0.274	0.0124
rs614372	1	108656156	t	c	0.4466	0.0216
rs60601651	1	230095614	t	c	0.9596	0.0075
rs614456	1	108656210	t	c	0.4465	0.0216

MinFreq	MaxFreq	Effect	StdErr	P	Direction	HetISq
0.1054	0.1853	0.098	0.0167	4.09E-09	+?+++++-+	32.6
0.8132	0.8948	-0.0952	0.0166	9.39E-09	-?-----+-	20.7
0.1145	0.1898	0.094	0.017	3.07E-08	+?+++++-+	32
0.8196	0.8919	-0.0955	0.0174	4.37E-08	-?-----+-	16.9
0.1385	0.1903	0.0938	0.0172	4.69E-08	+??++++-+	42.4
0.814	0.8948	-0.0935	0.0172	5.68E-08	-?-----+-	22.5
0.1308	0.1818	0.0943	0.0176	7.79E-08	+??++++-+	53.4
0.102	0.1843	0.0845	0.0167	4.34E-07	+?+++++-+	12.5
0.7081	0.8168	-0.0684	0.0136	4.65E-07	---+---+-	41.7
0.6892	0.8241	-0.0685	0.0136	4.83E-07	---+---+-	26.1
0.3977	0.4759	0.0707	0.0141	5.06E-07	++++?+?++	9.1
0.9526	0.9632	-0.1528	0.0309	7.77E-07	-???--???	0
0.9322	0.9361	0.1925	0.0393	1.00E-06	+???????+	51.8
0.6856	0.8155	-0.0656	0.0135	1.18E-06	---+---+-	12.1
0.1014	0.1812	0.0808	0.0167	1.23E-06	+?+++++-+	21.7
0.7844	0.8793	-0.0762	0.0158	1.45E-06	-?-------	0
0.9342	0.9349	0.1889	0.0393	1.53E-06	+???????+	69.5
0.1159	0.2108	0.0785	0.0163	1.56E-06	+?+++++++	0
0.1537	0.2962	-0.0728	0.0152	1.57E-06	----?--?-	0
0.6912	0.7835	0.0735	0.0154	1.74E-06	+++??++?+	0
0.7038	0.8455	0.0722	0.0152	1.87E-06	++++?++?+	0
0.7015	0.8466	0.0709	0.0151	2.57E-06	++++?++?+	0
0.1664	0.3051	0.0666	0.0142	2.62E-06	+++-+++-+	12.5
0.7836	0.8162	0.0712	0.0152	2.75E-06	+++-+++++	0
0.1664	0.3052	0.0664	0.0142	2.80E-06	+++-+++-+	14.2
0.7068	0.7868	0.0718	0.0154	2.91E-06	+++??++?+	0
0.7819	0.8796	-0.074	0.0159	3.10E-06	-?-------	0
0.0639	0.085	-0.1167	0.0253	4.14E-06	-?-?---?-	0
0.695	0.8336	-0.0653	0.0142	4.21E-06	---+---+-	8.2
0.1168	0.194	0.0766	0.0167	4.23E-06	+?+++++++	0
0.1581	0.2934	-0.0694	0.0151	4.27E-06	----?--?-	0
0.1327	0.1979	0.0768	0.0167	4.37E-06	+?+++++++	0
0.1168	0.1937	0.0757	0.0167	5.48E-06	+?+++++++	0
0.7082	0.7827	0.0608	0.0134	5.50E-06	++++++--+	0
0.7312	0.7446	0.0974	0.0214	5.59E-06	+???????+	0
0.2137	0.2793	-0.0608	0.0134	5.72E-06	------++-	0
0.6899	0.7754	0.0601	0.0133	6.51E-06	++++++--+	0
0.2135	0.2709	-0.0603	0.0135	7.52E-06	------++-	0
0.0595	0.0658	-0.1771	0.0396	7.66E-06	-???????-	1.1
0.6112	0.7328	0.0643	0.0144	8.25E-06	++-+?++?+	17.5
0.2136	0.274	-0.0599	0.0134	8.26E-06	------++-	0.8
0.2134	0.2708	-0.0599	0.0134	8.28E-06	------++-	0
0.6823	0.7564	-0.0613	0.0138	8.50E-06	-------+-	0
0.0907	0.1184	0.0896	0.0201	8.54E-06	+??++++-+	0
0.214	0.2817	-0.0595	0.0134	8.76E-06	------++-	0
0.7196	0.7864	0.0595	0.0134	8.89E-06	++++++--+	0

0.2136	0.274	-0.0596	0.0134	9.12E-06	------++-	0
0.9513	0.9663	-0.2056	0.0463	9.13E-06	-????-??-	0
0.2136	0.2737	-0.0596	0.0134	9.28E-06	------++-	0
0.6823	0.7564	-0.0611	0.0138	9.29E-06	-------+-	0
0.9506	0.9663	-0.2047	0.0462	9.50E-06	-????-??-	0
0.9506	0.9663	-0.2047	0.0462	9.52E-06	-????-??-	0
0.0337	0.0494	0.2047	0.0463	9.62E-06	+????+??+	0
0.7157	0.8267	-0.0609	0.0138	9.73E-06	---+---+-	28.1
0.0337	0.0487	0.2049	0.0463	9.73E-06	+????+??+	0
0.9183	0.9465	0.1266	0.0286	9.77E-06	+?+?++??+	0
0.9183	0.9465	0.1266	0.0286	9.81E-06	+?+?++??+	0
0.6823	0.7564	-0.0608	0.0138	1.00E-05	-------+-	0
0.2436	0.3175	0.0613	0.0139	1.02E-05	+++++++-+	0
0.2434	0.3174	0.0613	0.0139	1.03E-05	+++++++-+	0
0.7157	0.8267	-0.0607	0.0138	1.03E-05	---+---+-	27.5
0.2649	0.3613	0.0594	0.0135	1.03E-05	+++++++-+	0
0.0337	0.0487	0.2042	0.0463	1.05E-05	+????+??+	0
0.0337	0.0486	0.2041	0.0463	1.06E-05	+????+??+	0
0.6105	0.7318	0.0635	0.0144	1.07E-05	++-+?++?+	28.5
0.2136	0.2746	-0.0591	0.0134	1.08E-05	------++-	0
0.0748	0.13	0.0848	0.0193	1.08E-05	+-+++++++	2.1
0.8024	0.8831	-0.0731	0.0166	1.09E-05	-?-------	0
0.6728	0.7469	0.064	0.0146	1.09E-05	+++??++?+	0
0.0657	0.086	-0.1114	0.0253	1.10E-05	-?-?---?-	0
0.6117	0.731	0.0633	0.0144	1.12E-05	++-+?++?+	29
0.7155	0.7831	0.0589	0.0134	1.12E-05	++++++--+	0
0.5449	0.6093	0.0532	0.0121	1.12E-05	+++++++-+	0
0.2436	0.3177	0.0605	0.0138	1.14E-05	+++++++-+	0
0.6821	0.7567	-0.0612	0.0139	1.14E-05	-------+-	0
0.0337	0.0486	0.2033	0.0463	1.15E-05	+????+??+	0
0.9514	0.9663	-0.2032	0.0463	1.16E-05	-????-??-	0
0.2135	0.2792	-0.0588	0.0134	1.16E-05	------++-	0
0.9514	0.9663	-0.2033	0.0464	1.17E-05	-????-??-	0
0.9514	0.9663	-0.2033	0.0464	1.17E-05	-????-??-	0
0.0337	0.0486	0.2033	0.0464	1.18E-05	+????+??+	0
0.0337	0.0486	0.2032	0.0464	1.18E-05	+????+??+	0
0.4726	0.5618	-0.0516	0.0118	1.23E-05	-+-------	0
0.8703	0.9252	-0.0842	0.0193	1.24E-05	-+-------	1.9
0.8029	0.8544	-0.0711	0.0163	1.24E-05	------+++	24.8
0.243	0.3174	0.0608	0.0139	1.25E-05	+++++++-+	0
0.2169	0.2829	-0.0585	0.0134	1.26E-05	------++-	0.1
0.3431	0.4855	-0.0526	0.0121	1.26E-05	---------	0
0.2438	0.3189	0.0609	0.014	1.28E-05	+++++++-+	0
0.3415	0.465	-0.0527	0.0121	1.32E-05	---------	0
0.1174	0.1921	0.0717	0.0165	1.34E-05	+?+++++-+	0
0.0661	0.0859	-0.109	0.0251	1.39E-05	-?-?---?-	0
0.675	0.7681	0.0678	0.0156	1.40E-05	+????++?+	0
0.5144	0.6569	0.0523	0.0121	1.42E-05	+++++++++	0

0.3904	0.4553	-0.0525	0.0121	1.44E-05	-------+-	0
0.0337	0.0488	0.2006	0.0463	1.49E-05	+????+??+	0
0.2461	0.3214	0.0552	0.0128	1.52E-05	+++-+++-+	48.6
0.3431	0.4648	-0.0521	0.0121	1.56E-05	---------	0
0.9505	0.9662	-0.2005	0.0464	1.57E-05	-????-??-	0
0.3431	0.4647	-0.0521	0.0121	1.57E-05	---------	0

HetChiSq	HetDf	HetPVal
10.383	7	0.1679
8.83	7	0.2651
10.295	7	0.1724
8.422	7	0.2968
10.417	6	0.1082
9.027	7	0.2508
12.874	6	0.04507
8.003	7	0.3323
13.714	8	0.08953
10.828	8	0.2116
6.603	6	0.3591
0.46	2	0.7944
2.073	1	0.1499
9.104	8	0.3336
8.939	7	0.257
1.066	7	0.9937
3.275	1	0.07035
0.517	7	0.9994
3.464	6	0.7487
3.095	5	0.6854
3.576	6	0.7338
3.245	6	0.7775
9.147	8	0.33
5.134	8	0.7431
9.322	8	0.3159
3.516	5	0.621
0.886	7	0.9965
4.111	5	0.5335
8.713	8	0.3671
0.935	7	0.9958
4.196	6	0.6502
1.139	7	0.9923
1.114	7	0.9928
6.921	8	0.5452
0.952	1	0.3293
7.432	8	0.4908
6.617	8	0.5785
7.693	8	0.464
1.011	1	0.3145
7.272	6	0.2964
8.063	8	0.4273
7.395	8	0.4946
6.778	8	0.5608
1.984	6	0.9211
7.471	8	0.4867
7.601	8	0.4733

7.9	8	0.4433
0.342	2	0.8428
7.515	8	0.4822
6.766	8	0.562
0.44	2	0.8026
0.439	2	0.8028
0.438	2	0.8035
11.124	8	0.1948
0.35	2	0.8395
1.007	4	0.9087
1.007	4	0.9087
6.265	8	0.6176
6.389	8	0.6038
6.477	8	0.594
11.037	8	0.1996
4.813	8	0.7774
0.366	2	0.8327
0.365	2	0.833
8.386	6	0.2112
7.75	8	0.4583
8.169	8	0.4172
1.07	7	0.9936
3.709	5	0.5921
3.863	5	0.5694
8.456	6	0.2065
7.484	8	0.4854
6.8	8	0.5584
6.809	8	0.5574
5.922	8	0.656
0.395	2	0.8207
0.395	2	0.8208
7.59	8	0.4745
0.395	2	0.8209
0.393	2	0.8218
0.392	2	0.8219
0.391	2	0.8222
6.36	8	0.6069
8.159	8	0.4181
10.641	8	0.2229
6.441	8	0.598
8.009	8	0.4326
5.045	8	0.7527
5.779	8	0.672
4.265	8	0.8324
1.372	7	0.9864
4.106	5	0.5343
1.273	3	0.7356
5.295	8	0.7257

6.404	8	0.602
0.5	2	0.7788
15.577	8	0.04885
5.082	8	0.7488
0.374	2	0.8293
5.059	8	0.7513

Table 4a: Association results and descriptive information for the top SNP rs1574587 based on th

Sample	**BETA**	**SE**	**P**	**N**	**EAF**	**INFO**	**λ_{GC}SE**	**λ_{GC}P**
ALSPAC	0.0858	0.0323	0.0079	6147	0.139	0.988	0.0317	0.0067
BLTS	NA	NA	NA	NA	NA	NA	NA	NA
FinnTwin	0.0535	0.0994	0.5898	1022	0.105	0.993	0.1076	0.6187
HUVH	0.167	0.1057	0.114	581	0.139	0.967	0.1116	0.1348
NTR	0.0697	0.0311	0.0248	5148	0.145	0.977	0.0324	0.0314
QIMR	0.1084	0.0404	0.0072	6758	0.134	0.987	0.0362	0.0027
TRAILS	0.263	0.0917	0.0041	1102	0.185	0.897	0.0857	0.0021
Utrecht	-0.127	0.1129	0.2605	958	0.156	0.976	0.1127	0.2597
Yale-Penn	0.1454	0.0435	0.0008	2362	0.139	0.975	0.0441	0.0009
CADD	NA	NA	NA	NA	NA	NA	NA	NA
NTR2/RADAR	NA	NA	NA	NA	NA	NA	NA	NA
SYS	-0.0904	0.1219	0.4581	533	0.133	0.981	0.1347	0.5018

Abbreviations: B – beta (effect size); SE – standard error of beta; N – sample size; EAF –effect a

Note: In the BLTS (dicovery) sample, and in the CADD and NTR2/RADAR (replication) samples the SNP rs 1574587 f

Note: rs1574587 is located at chromosome 16, base pair position 84453056, Allele 1=T, Allele 2 = C

e discovery samples and the replication samples.

ıllele frequency; INFO – imputation quality; λ_{GC} *SE – lambda corrected standard error;* λ_{GC} *P–*
ailed to meet the imputation quality threshold of >.8.

lambda corrected P-value;

Symbol	NominalP	CorrectedP	Chromosome	Start_Position
ATP2C2	0.00000133	0.034007967	16	84440193
SPINT1-AS1	0.0000312	0.371946674	15	41130613
SPINT1	0.0000455	0.371946674	15	41136642
C2orf76	0.0000618	0.371946674	2	120059792
HIRIP3	0.0001	0.371946674	16	30003641
ARG2	0.00011	0.371946674	14	68086578
MIR195	0.00013	0.371946674	17	6920933
MIR497	0.00014	0.371946674	17	6921229
PPP1R14D	0.00016	0.371946674	15	41107642
INO80E	0.00018	0.371946674	16	30007529
C17orf49	0.0002	0.371946674	17	6918055
CARMN	0.00021	0.371946674	5	148786407
MIR497HG	0.00021	0.371946674	17	6919136
RNASEK	0.00023	0.371946674	17	6915735
BCL6B	0.00023	0.371946674	17	6926368
RNASEK-C17orf49	0.00024	0.371946674	17	6915735
MIR6891	0.00025	0.371946674	6	31323000
HLA-B	0.00026	0.371946674	6	31321642
VTI1B	0.00031	0.410128365	14	68117866
LINC02287	0.00032	0.410128365	14	93372041
HS3ST1	0.00063	0.622432982	4	11399987
ECT2L	0.00065	0.622432982	6	139117247
CCDC168	0.00074	0.622432982	13	103381716
CCDC28A	0.00075	0.622432982	6	139094656
LOC105379511	0.00076	0.622432982	20	25733196
EPGN	0.00084	0.622432982	4	75174186
TAOK2	0.00084	0.622432982	16	29985187
MYMK	0.00085	0.622432982	9	136379707
LOC105372582	0.00086	0.622432982	20	25731843
ZFYVE19	0.00087	0.622432982	15	41099273
FAS-AS1	0.00093	0.622432982	10	90751180
CDH24	0.00094	0.622432982	14	23516269
ABCC1	0.00097	0.622432982	16	16043433
TMEM45A	0.00099	0.622432982	3	100211462
LOC158435	0.00101	0.622432982	9	98828120
MYBPH	0.00104	0.622432982	1	203136938
KRT32	0.00106	0.622432982	17	39615764
LRFN2	0.00114	0.622432982	6	40359372
HOXC12	0.00115	0.622432982	12	54348651
LY6G5B	0.00118	0.622432982	6	31638727
DOC2A	0.0012	0.622432982	16	30016834
SFTA2	0.00122	0.622432982	6	30899126
TMEM219	0.00124	0.622432982	16	29973350
RDH12	0.00125	0.622432982	14	68168602
ZFAT-AS1	0.00125	0.622432982	8	135610313
TIPIN	0.00129	0.622432982	15	66629007
MIR6832	0.0013	0.622432982	6	31601563
C5orf47	0.0013	0.622432982	5	173416161
RNASEH2B	0.00132	0.622432982	13	51483813

HOXC13	0.00134	0.622432982	12	54332575
MFN1	0.00135	0.622432982	3	179065479
PPIE	0.00136	0.622432982	1	40204516
RDH11	0.00136	0.622432982	14	68143518
TMEM212-AS1	0.0014	0.622432982	3	171594141
MIR3135B	0.00141	0.622432982	6	32717688
GPANK1	0.00148	0.622432982	6	31629005
HOXC13-AS	0.00148	0.622432982	12	54329111
MUC22	0.00148	0.622432982	6	30973728
GNB4	0.0015	0.622432982	3	179113875
GDPD3	0.00151	0.622432982	16	30116130
CSNK2B	0.00154	0.622432982	6	31633656
SPATA46	0.00155	0.622432982	1	162343514
PSMB7	0.00158	0.622432982	9	127115743
ZNF639	0.00164	0.622432982	3	179041550
LOC101928595	0.00165	0.622432982	16	30107750
TMPRSS11GP	0.00173	0.622432982	4	68857529
C1orf226	0.00173	0.622432982	1	162351519
TOMM6	0.00174	0.622432982	6	41755180
ARHGEF38	0.00175	0.622432982	4	106473776
LINC02451	0.00179	0.622432982	12	43040384
NUDCD1	0.00182	0.622432982	8	110253147
MT4	0.00189	0.622432982	16	56598960
SPATA5L1	0.0019	0.622432982	15	45694518
MIR876	0.00191	0.622432982	9	28863623
SMARCAL1	0.00202	0.622432982	2	217277472
LINC00207	0.00204	0.622432982	22	44965219
CCNG2	0.00204	0.622432982	4	78078356
SNAPC5	0.00204	0.622432982	15	66785805
TCL6	0.00207	0.622432982	14	96117514
TJP2	0.00209	0.622432982	9	71820077
DIS3L	0.0021	0.622432982	15	66586157
ATAD5	0.0021	0.622432982	17	29158987
CHI3L1	0.00213	0.622432982	1	203148058
ACIN1	0.00214	0.622432982	14	23527773
SH3GL3	0.00216	0.622432982	15	84159365
TTF1	0.00218	0.622432982	9	135250936
PPP2R5D	0.00219	0.622432982	6	42952236
GATM	0.0022	0.622432982	15	45653321
PUS3	0.00224	0.622432982	11	125763379
GET4	0.00225	0.622432982	7	916190
DNAJC17	0.00225	0.622432982	15	41060066
HLA-DQB2	0.00227	0.622432982	6	32723874
MIR6078	0.00229	0.622432982	10	4033351
MIR4512	0.0023	0.622432982	15	66789295
LOC101805491	0.00232	0.622432982	2	46656328
SNORD18C	0.00237	0.622432982	15	66793589
HLA-DQA2	0.00241	0.622432982	6	32709162
HMP19	0.00243	0.622432982	5	173472606
SNORA70C	0.00243	0.622432982	9	119943344

Group
protein-coding gene
non-coding RNA
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
non-coding RNA
non-coding RNA
protein-coding gene
protein-coding gene
protein-coding gene
non-coding RNA
non-coding RNA
protein-coding gene
protein-coding gene
other
non-coding RNA
protein-coding gene
protein-coding gene
non-coding RNA
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
unknown
protein-coding gene
protein-coding gene
protein-coding gene
unknown
protein-coding gene
non-coding RNA
protein-coding gene
protein-coding gene
protein-coding gene
unknown
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
non-coding RNA
protein-coding gene
non-coding RNA
protein-coding gene
protein-coding gene

protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
non-coding RNA
non-coding RNA
protein-coding gene
non-coding RNA
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
unknown
pseudogene
protein-coding gene
protein-coding gene
protein-coding gene
non-coding RNA
protein-coding gene
protein-coding gene
protein-coding gene
non-coding RNA
protein-coding gene
non-coding RNA
protein-coding gene
protein-coding gene
non-coding RNA
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
protein-coding gene
non-coding RNA
non-coding RNA
unknown
non-coding RNA
protein-coding gene
unknown
non-coding RNA

Supplemental Information 1. Information about sample collection

ALSPAC: Avon Longitudinal Study of Parents and Children — United Kingdom

The Avon Longitudinal Study of Parents and Children (ALSPAC) is a prospective cohort study which recruited 14,541 pregnant women residing in Avon, United Kingdom, with expected dates of delivery between 1 April 1991 and 31 December 1992. Of these initial pregnancies, there was a total of 14,676 fetuses, resulting in 14,062 live births and 13,988 children who were alive at 1 year of age. Full details of study recruitment and methodology have been published previously[1]. Detailed information on the mothers and their children has been collected from self-report questionnaires and attendance at clinics. Please note that the study website contains details of all the data that is available through a fully searchable data dictionary (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/). Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Data on cannabis use for this project was measured in the ALSPAC offspring between 17 and 20 years of age and limited to unrelated individuals only. The answer categories were recoded as uncensored (coded as 0, ever used cannabis) versus censored (coded as 1, never used cannabis during their lifetime) observations. Individuals were also asked about frequency of cannabis use if they answered yes to ever having used cannabis. Cannabis data were available for 6,230 individuals. Both genotype and lifetime cannabis use data were available for 6.147 individuals.

References:

1. Boyd et al. (2013), Cohort profile: The 'Children of the 90s' - the index offspring of the Avon Longitudinal Study of Parents and Children. *IJE, 42*, 111-127.

BLTS: Brisbane Longitudinal Twin Study - Australia

Beginning in 1992, the Brisbane Longitudinal Twin Study (BLTS) consists of 3,561 individuals: 1,422 twin pairs and 717 additional siblings first enrolled at age 12 years and now aged 30 years and older (Gillespie, 2013). The sample is: genetically informative (MZ and DZ twins, and often parents and siblings; genotyped for 610,000 common single nucleotide polymorphisms - SNPs); (b) large; (c) longitudinal with many participants have been assessed at 12, 14, 16 and 21 years of age; (d) well characterized for behavioral and brain-related outcomes; (e) rich in biological samples; and includes (f) a subgroup [n=969] who have undergone MRI scanning. As part of an ongoing US NIH/NIDA funded project beginning 2009, measures of lifetime cannabis use, abuse and dependence data are collected, along with diagnostic data for nicotine, alcohol, and other illicit substances, as well as pilot epidemiological data for ecstasy and methamphetamine use. The average age at interview is 25.65 yrs (SD=3.65, range=18-38yrs). Lifetime cannabis use was assessed by asking twins, "In your life, have you ever used cannabis (marijuana, pot, grass or hash)?". The answer categories were recoded as uncensored (coded as 0, if one ever used cannabis) versus censored (coded as 1, if one never used cannabis during the lifetime) observations. Age at initiation was assessed by the question "At what age did you first use cannabis?". The entire BLTS sample and 1,549 of their parents have GWAS data (Illumina 610k chip) (Medland et al., 2009) imputed on the GRCh37 assembly. The final sample included 721 individuals (314 males and 407 females) with both genotypic and age at cannabis iniation data.

References

Gillespie, N.A., A.K. Henders, T.A. Davenport, D.F. Hermens, M.J. Wright, N.G. Martin, and I.B. Hickie, The Brisbane Longitudinal Twin Study: Pathways to Cannabis Use, Abuse, and Dependence project-current status, preliminary results, and future directions. Twin Res Hum Genet, 2013. 16(1): p. 21-33. PMC3805122

Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G et al. Common variants in the trichohyalin gene are associated with straight hair in Europeans. Am J Hum Genet 2009; 85(5): 750-755.

CADD:

Data on cannabis use and DNAs were collected as part of two longitudinal centers based in Colorado and California, the Center on Antisocial Drug Dependence (CADD) and the Genetics of Antisocial Drug Dependence (GADD). Subjects from the centers included both clinical and community samples with up to three waves of data collection at approximately 5 year intervals. The original set of unrelated subjects from Derringer (2015) was augmented with additional family members. For these analyses, genotypes were available from 1806 individuals who were over-selected for adolescent behavioral disinhibition. 35.5% of the sample were females, and the mean age was 23.65 (4.36) years, with a range from 13-38 years. Lifetime cannabis use data were collected using a supplemental questionnaire appended to the CIDI-SAM, which asked participants to self report "Have you ever used_____ ?" (yes/no) for each of 14 substances (plus "other"), including cannabis. 81.9% of respondents reported ever using cannabis. If participants answered in the affirmative, they were asked to respond to questions about age at first use, age at regular use, typical pattern of use, and days used within the past six months; otherwise, participants were instructed to skip to questions about the next substance. If subjects had been assessed on more than one occasion, age of initiation was taken from the assessment when they first reported cannabis use. Genotyping was conducted using the Affymetrix 6 platform. Samples were excluded if their call rate was less than 95%, if they were one of 18 samples from the oldest cohort (as this was too few to be used as a reference cohort), of if they did not have phenotype data. After these exclusions, 541445 SNPs passed initial genotyping QC that removed SNPs with MAF <1%, or call rate <99%, or p(HWE) < 1.00E-06. Genotype calling used GeneChip Targeted Genotyping Analysis Software. Imputation to the 1000 genome, phase 1, release March 2012, and used miniMACh software via https://imputationserver.sph.umich.edu/index.html

Derringer J, Corley RP, Haberstick BC, Young SE, Demmitt BA, Howrigan DP, Kirkpatrick RM, Iacono WG, McGue M, Keller MC, Brown S, Tapert S, Hopfer CJ, Stallings MC, Crowley TJ, Rhee SH, Krauter K, Hewitt JK, McQueen MB.(2015). Genome-Wide Association Study of Behavioral Disinhibition in a Selected Adolescent Sample. Behav Genet. 2015 Jul;45(4):375-81. doi: 10.1007/s10519-015-9705-y. PubMed PMID: 25637581; PubMed Central PMCID: PMC4459903.

FinnTwin: Finnish Twin Cohort (FinnTwin12 & FinnTwin16) - Finland

In FinnTwin12 (FT12), data on lifetime cannabis use were collected as part of a longitudinal study targeting all Finnish twin pairs born in 1983-1987 (Kaprio, Pulkkinen, Rose 2002). Four waves of data collection have been completed (at ages 12, 14, 17.5, and in early adulthood age range 21-25) (Kaprio, 2013). In wave 4, using a SSAGA interview we asked whether participants ever experimented with cannabis (no, yes) and if yes, at what age they first experimented with cannabis. They were also asked how many times they have used cannabis. Cannabis abuse and dependence were also assessed. Lifetime cannabis use data were available for 1346 FT12 subjects (25.3% (N=341) ever users). Both genotype and cannabis data were available for 929 FT12 subjects (26.6% (N=247) ever users). The mean age of cannabis use initiation among ever users was 17.8 (SD 2.2).

In FinnTwin16 (FT16), data on lifetime cannabis use were collected as part of a longitudinal study targeting all Finnish twin pairs born in 1975-1979 (Kaprio, Pulkkinen, Rose, 2002). Five waves of data collection have been completed (at ages 16, 17, 18.5, mean age 24, and mean age 34). In wave 4 we conducted SSAGA interviews from a subsample. In the SSAGA interview we asked whether participants ever experimented with cannabis (no, yes) and if yes, at what age they first experimented with cannabis. They were also asked how many times they have used cannabis. Cannabis abuse and dependence were also assessed. Lifetime cannabis use data were available for 602 FT16 subjects (34.2% (N=206) ever users). Both genotype and cannabis data were available for 100 FT16 subjects (36.0% (N=36) ever users). The mean age of cannabis use initiation among ever users was 19.9 (SD 3.0).

FinnTwin12 and FinnTwin16 samples were analyzed together. The answer categories were recoded as uncensored (coded as 0, if one ever used cannabis) versus censored (coded as 1, if one never used cannabis during the lifetime) observations. Both genotype and cannabis use data were available for altogether 1029 subjects (27.5% (N=283) ever users). The mean age of cannabis use initiation among ever users was 18.0 (SD 2.5).

References

Kaprio J. The Finnish Twin Cohort Study: an update. Twin Res Hum Genet. 2013; 16(1):157-62.

Kaprio J, Pulkkinen L, Rose RJ. Genetic and environmental factors in health-related behaviors: studies on Finnish twins and twin families. Twin Res. 2002; 5(5):366-71.

HUVH: Hospital Universitari Vall d'Hebron – Barcelona - Spain

Data on lifetime cannabis use were collected as part of a GWAS study of persistent ADHD. Recruitment of participants was performed between 2004 and 2011 at the Department of Psychiatry of the Hospital Universitari Vall d'Hebron, Barcelona, Spain. Genotypes were available from 1039 unrelated Caucasian individuals. Age at initiation of lifetime cannabis use data was collected from 581 of them) and assessed by the Structured Clinical Interview for DSM-IV Axis I and Axis II Disorders (SCID-I and SCID-II). The average age at assessment was 28.7 years (SD = 12.5), 71% of participants were males. The study was approved by the ethics committee of the institution and informed consent was obtained from all subjects.

Reference

Sánchez-Mora C, Ramos-Quiroga JA, Bosch R, Corrales M, Garcia-Martínez I, Nogueira M, Pagerols M, Palomar G, Richarte V, Vidal R, Arias-Vasquez A, Bustamante M, Forns J, Gross-Lesch S, Guxens M, Hinney A, Hoogman M, Jacob C, Jacobsen KK, Kan CC, Kiemeney L, Kittel-Schneider S, Klein M, Onnink M, Rivero O, Zayats T, Buitelaar J, Faraone SV, Franke B, Haavik J, Johansson S, Lesch KP, Reif A, Sunyer J, Bayés M, Casas M, Cormand B, Ribasés M. Case-control genome-wide association study of persistent attention-deficit hyperactivity disorder identifies FBXO33 as a novel susceptibility gene for the disorder. Neuropsychopharmacology. 2015 Mar;40(4):915-26. doi: 10.1038/npp.2014.267.

Ramos-Quiroga JA, Sánchez-Mora C, Casas M, Garcia-Martínez I, Bosch R, Nogueira M, Corrales M, Palomar G, Vidal R, Coll-Tané M, Bayés M, Cormand B, Ribasés M. Genome-wide copy number variation analysis in adult attention-deficit and hyperactivity disorder. J Psychiatr Res. 2014 Feb;49:60-7.

NTR1 and NTR2: Netherlands Twin Register — The Netherlands

Data on lifetime cannabis use were collected as part of a longitudinal study on health, personality and lifestyle in adolescent and adult twins and their relatives (i.e., their non-twin siblings, parents, spouses and children). Ten waves of data collection have been completed (in 1991, 1993, 1995, 1997, 2000, 2002, 2004 and 2009-2010, 2011-2012, and 2013-3014). Questions on cannabis use were administered 5 times: in 1993 (wave 2), 1995 (wave 3), , 2000 (wave 5), 2009-2010 (wave 8), and 2013-2014 (wave 10). In wave 2, 3, and 5 the participants were asked at what age they had experimented with cannabis for the first time and at what age they had regularly used cannabis. The answer categories were: 1=never, 2=11 years or younger, 3=12 years, 4=13 years, 5=14 years, 6=15 years, 7=16 years, 8=17 years, 9=18 years or older in wave 2 and 3, and 1=11 years or younger, 2=12-13 years, 3=14-15 years, 4=16-17 years, 5=18 years or older and never in wave 5. In wave 8 and 10 participants were asked whether they had ever experimented with cannabis (no, yes) and if yes, at what age they started. They were also asked whether they had ever used cannabis on a regular basis (no, yes) and if yes, at what age. For the NTR1 sample, we limited the sample to data from wave 8, as in this wave age at initiation was collected based on an open question (see above) and wave 10 was not available yet during the discovery phase of our study. For NTR2 we used data from individuals that were no family members of the individuals in the NTR1 sample. For NTR2 we used the responses from wave 10, and when missing we also used the responses to waves 5, 3, and 2 (in this order).The answer categories were recoded to uncensored observations (coded as 0, if one ever used cannabis) versus censored observations (coded as 1, if one never used cannabis). Both genotype and age at initiation of cannabis use data were available for 5,148 subjects for NTR1 and for 1,740 individuals for NTR2.

QIMR: Queensland Institute of Medical Research Berghofer adults — Australia

Data from Australian adults were collected in twin family studies conducted at the QIMR Berghofer Medical Research Institute. Data on cannabis use were obtained from: 1) a series of studies conducted collaboratively by Nick Martin and Andrew Heath between 2001 and 2006 (Pergadia et al., 2009; Saccone et al., 2007; Distel et al., 2008), and 2) a study conducted between 1996 and 2000 of 6233 twin individuals from the young adult cohort (born between 1964 and 1971) (see Nelson et al. 2002; Knopik et al. 2004). In both studies individuals participated in semi-structured telephone interviews primarily focussed at psychiatric disorders. The interview was an adaptation of the SSAGA (Semi-Structured Assessment for the Genetics of Alcoholism). As part of this interview individuals were asked whether they had ever used cannabis and at what age they first used cannabis (with the question: "How old were you when you first used…?"). In case individuals participated in both studies, data from the last assessment were included.

The genotypic data are derived from multiple waves of genotyping. DNA samples were collected in accordance with standard protocols and submitted to different genotype centres using different Illumina SNP platforms (317 single, 370 single, 370 duo, 670 quad, 610 quad) (see Medland et al., 2009). Phenotypic and genotypic data collections were approved by the QIMR Human Research Ethics Committee and informed consent was obtained from all participants.

The final sample included 6,758 individuals with both genotype and phenotype data (NB. One twin per MZ twin pair was deleted).

References

Distel MA, Trull TJ, Derom CA, Thiery EW, Grimmer MA, Martin NG *et al.* Heritability of borderline personality disorder features is similar across three countries. *Psychol Med* 2008; 38(9): 1219-1229.

Knopik VS, Heath AC, Madden PAF, Bucholz KK, Slutske WS, Nelson EC, Statham D, Whitfield JB, Martin NG (2004). Genetic effects on alcohol dependence risk : re-evaluating the importance of psychiatric and other heritable risk factors. Psychological Medicine 34, 1519–1530.

Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G *et al.* Common variants in the trichohyalin gene are associated with straight hair in Europeans. *Am J Hum Genet* 2009; 85(5): 750-755.

Nelson EC, Heath AC, Madden PAF, Cooper ML, Dinwiddie SH, Bucholz KK, Glowinski A, McLaughlin T, Dunne MP, Statham DJ, Martin NG (2002). Association between self-reported childhood sexual abuse and adverse psychosocial outcomes: results from a twin study. Archives of General Psychiatry 59, 139–145.

Pergadia ML, Agrawal A, Loukola A, Montgomery GW, Broms U, Saccone SF. Genetic linkage findings for DSM-IV nicotine withdrawal in two populations. *Am J Med Genet B* 2009; 150B: 950-959.

Saccone SF, Pergadia ML, Loukola A, Broms U, Montgomery GW, Wang JC *et al.* Genetic linkage to chromosome 22q12 for a heavy-smoking quantitative trait in two independent samples. *Am J Hum Genet* 2007; 80(5): 856-866.

RADAR: Research on Adolescent Development and Relationships – The Netherlands

The RADAR study (Research on Adolescent Development and Relationships) is longitudinal research project in the Netherlands and focuses on the development of interpersonal relationships, personality, and psychopathology, in a sample of adolescents and their families that were followed from approximately ages 13 to 19. Currently, there are 7 waves available (collected between 2006 and 2013), but the study is still ongoing. The RADAR study has a focus on delinquency development, therefore adolescents at risk for externalizing behavior were oversampled, which was determined by a having a T-score > 60 on the externalizing scale of the Teacher's Report Form at age 12 (TRF; Achenbach, 1991; Verhulst, van der Ende, & Koot, 1997). In total, 497 adolescents were included in the study, of which 206 (41.45%) were at high risk for externalizing behavior (for more information on the sample see Creemers et al., 2015). The study was approved by the medical ethical committee of Utrecht University. Families received 100 Euros for each home visit. The data on lifetime cannabis use were collected every year as part of home assessments, during which research assistants visited the adolescents and their families (wave 1-7). Using self-report questionnaires, adolescents were asked to indicate how often they had used hash or weed in the past 12 months. Answer categories were 0 times, 1 time, 2 times, 3 times, 4 times, 5 times, 6 times, 7 times, 8 times, 9 times, 10 times, 11-19 times, 20-39 times, 40 times or more. Additionally, adolescents were asked how old they were when they had used weed or hash for the first time, but only in the last wave (wave 7). For the present analyses, age at onset of cannabis use was based on the response to wave 7, but if missing it was calculated based on all annual assessments.

In wave 5, 416 adolescents provided genotype data. For the current study, valid genotype and phenotype data were available for 342 adolescents. The mean age at wave 7 was 19.5 (SD = 0.8, range = 13–22).

References:

Achenbach, T. M. (1991). Manual for the youth self report and 1991 profiles. Burlington: Department of Psychiatry, University of Vermont.

Creemers, H. E., Buil, J. M., van Lier, P. A.C., Keijsers, L., Meeus, W., Koot, H. M., & Huizink, A. C. (2015). Early onset of cannabis use: Does personality modify the relation with changes in perceived parental involvement? *Drug and Alcohol Dependence, 146,* 61-67.

Verhulst, F. C., van der Ende, J., & Koot, H. M. (1997). Handleiding voor de Youth-Self-Report [Manual for the Youth Self-Report]. Rotterdam, The Netherlands: Afdeling jeugdpsychiatrie, Sophia Kinderziekenhuis/Academisch Ziekenhuis Rotterdam/Erasmus Universiteit Rotterdam.

Saguenay Youth Study

The Saguenay Youth Study (SYS) is a population-based study of adolescents and their middle-aged parents1. It is aimed at investigating the etiology, early stages and trans-generational trajectories of common cardio-metabolic and brain diseases. The SYS was designed as a two-generational cohort; it includes 1,029 adolescents and their 962 parents. The cohort was recruited via 12- to 18-year old adolescents attending high schools in the Saguenay Lac-Saint-Jean region of Quebec (Canada). Half of the adolescents were exposed prenatally to maternal cigarette smoking. The cohort is family-based (n=481 families), including only adolescents who have one or more siblings of similar age (i.e., 12 to 18 years) and both biological parents of the French-Canadian origin born in the region. The data collection occurred in two waves. Wave 1 (2003-2012) involved the recruitment and complete assessment of all 1,028 adolescents, as well as a partial ('soft') assessment of 962 parents. Wave 2 (2012-2015) involved the complete assessment of a subset of the parents (n=664). In Wave 2, parents answered a series of questions about their drug use; this questionnaire was based on the European School Survey Project on Alcohol and Other Drugs (http://www.espad.org/). The GWAS was based on answering (Yes/No) the following question: "Have you ever used marijuana (grass, pot) or hashish (hash, hash oil)?" with the following answer categories: 1=11 or less 2= 12 3=13 4=14 5=15 6=16 7=17 8=18 and more 9= prefer to not answer. Of the total sample (N=593), 310 individuals did not initiate cannabis use, 184 individuals initiated cannabis use at an age between 12 and 17 years and 99 individuals said they did so at 18 years of later. Given that the average age at initiation is around 15.4 (youth), 16.5 (young adults) and 18.8 (adults) (www.ccsa.ca/), the age at initiation for this last group is set on 18 years old in order to allow us to use the full sample and so, to maximize the power to replicate the results. We note that re-running the analysis by excluding this group did not change the results and the conclusions of the replication stage.

1. Paus T, Pausova Z, Abrahamowicz M, et al. Saguenay Youth Study: A multi-generational approach to studying virtual trajectories of the brain and cardio-metabolic health. Developmental cognitive neuroscience. Oct 23 2014.

TRAILS: TRacking Adolescents' Individual Lives Survey — The Netherlands.

Data on lifetime cannabis use were collected as part of a Dutch longitudinal study on the development of (mental) health in adolescence and young adulthood. Data on lifetime cannabis use were collected at the fourth wave, in 2008-2010, when the sample was 18-20 years old. The participants were asked if and how often they had used cannabis during (1) their lives, (2) the last year, and (3) the last month; and how old they were when they used cannabis for the first time. Both genotype and data on lifetime cannabis use and age at initiation were available for 1249 subjects.

Reference

Oldehinkel AJ, Rosmalen JGM, Buitelaar JK, Hoek HW, Ormel J, Raven D, Reijneveld SA, Veenstra R, Verhulst FC, Vollebergh WAM, Hartman CA. Cohort Profile update. The TRacking Adolescents' Individual Lives Survey (TRAILS). Int J Epidemiol 2015; 44(1):76-76n.

Utrecht: Utrecht Cannabis Cohort (CannabisQuest) –The Netherlands

Participants were recruited using a project website launched in 2006 targeted at Dutch young adults and adolescents from 18 to 25 years (www.cannabisquest.nl) (Schubart et al., 2010). Strategies to generate traffic on the project website included collaboration with over a hundred colleges, universities, and youth centres, as well as the use of online commercial advertisement products (i.e. banners and text links) (Schubart et al., 2010). The chance to win an Apple iPod™ or a Nintendo Wii™ was used as an incentive. Double entries were prevented by exclusion of subjects with an identical e-mail address, surname, and date of birth. Anonymous submission of data was not possible. The online assessment included verification questions to protect against random answers, and participants failing to correctly complete the verification questions were subsequently excluded. From the online data (N = 17,698), 1259 participants were included for subsequent genetic assessment in two waves. First, in order to increase power for gene × environment interactions (Boks et al., 2007), we prioritized a sample of 719 participants who belonged to the top or bottom quintile of total scores of psychotic experiences as measured by the Community Assessment of Psychic Experiences (CAPE) score (see below) that were either cannabis naïve (i.e. a lifetime cannabis exposure frequency less than 6 times) or were cannabis users (i.e. current expenditure for personal cannabis use exceeded 3€ weekly). Second, an unselected sample of 540 individuals was included. As ascertained with the validated Dutch version of either the Structured Clinical Interview (SCID) (First et al., 1997) or the MINI International Neuropsychiatric Interview (Sheehan et al., 1998), healthy controls had no history of any psychotic disorder. The possible concomitant use of recreational drugs was assessed with the substance abuse module of the Composite International Diagnostic Interview (Compton, 1993). Participants provided a urine sample to screen for the presence of recreational drugs in order to verify recent self-reported cannabis use. The study was approved by the Ethical Review Board of the University Medical Center Utrecht and all participants gave written informed consent. For a total of 1173 participants data on age at initiation of cannabis use and genotypes were available.

References

Schubart CD, van Gastel WA, Breetvelt EJ, Beetz SL, Ophoff RA, Sommer IE, Kahn RS, Boks MP. Cannabis use at a young age is associated with psychotic experiences. Psychol Med 2011;41:p 1301-1310.

Vinkers CH, Van Gastel WA, Schubart CD, Van Eijk KR, Luykx JJ, Van Winkel R, ; GROUP Investigators, Joëls M, Ophoff RA, Boks MP. The effect of childhood maltreatment and cannabis use on adult psychotic symptoms is modified by the COMT Val158Met polymorphism. Schizophr Res. 2013 Aug 15.

Yale Penn EA: Genetics of Substance Dependence - United States

Our sample included a total of 2,379 European American (EA) subjects from a cohort of small nuclear families and unrelated individuals originally collected to study the genetics of drug (opioid or cocaine) or alcohol dependence (Gelernter et al., 2014). Subjects gave written informed consent as approved by the institutional review board at each site, and certificates of confidentiality were obtained from NIDA and NIAAA. Yale/Penn subjects were administered the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA) (Pierucci-Lagha et al, 2005) to derive DSM-V diagnoses of lifetime cannabis dependence and other major psychiatric traits. Lifetime cannabis use was assessed by the following question: "Have you ever used marijuana to feel good or high, or to feel more active or alert." Age at first use, as well as a measure of frequency of during the period in the subject's life in which they used the drug most heavily, was also assessed. A total of 2.188 subjects for which both age at initiation of cannabis use and genotypes were available were used in this study.

References

Gelernter, J., Kranzler, H. R., Sherva, R., Almasy, L., Koesterer, R., Smith, A. H., ... & Farrer, L. A. (2014). Genome-wide association study of alcohol dependence: significant findings in African-and European-Americans including novel risk loci. *Molecular psychiatry*, *19*(1), 41-49.

Pierucci-Lagha, A., Gelernter, J., Feinn, R., Cubells, J. F., Pearson, D., Pollastri, A., ... & Kranzler, H. R. (2005). Diagnostic reliability of the Semi-structured Assessment for Drug Dependence and Alcoholism (SSADDA). *Drug and alcohol dependence*, *80*(3), 303-312.

Heritability study

Subjects: The sample for estimating the heritability for age at first cannabis use consisted of 8,055 twin pairs from three samples: the NTR sample comprising 3,798 twin pairs (Willemsen et al., 2013), the QIMR sample comprising 3,251 twin pairs (Heath et al., 2011; Knopik et al., 2004), and the BLTS sample comprising 1,006 twin pairs (Gillespie et al., 2013). The prevalence of lifetime cannabis use was 24%, 57%, and 51% in the NTR, QIMR, and BLTS samples respectively. Among ever or lifetime cannabis users, the percentage early initiators (before the age of 18) was 56%, 36% and 51% respectively.

Measures: To investigate the heritability of age at first cannabis use, this trait was considered to have an underlying, continuous liability. A threshold liability model was then used to divide age at first cannabis use into four ordinal categories: users who initiated cannabis before age 18, initiation between ages 18 to 20, initiation after age 20, and never users.

Genetic models: We fitted and compared three models to determine the relationship between risk of cannabis initiation *per se* and the age at first use. The single liability dimension model (SLD) postulates that the liability to cannabis initiation and the age at first cannabis use fall along the same dimension of risk or liability. In contrast, the independent liability dimension (ILD) model predicts that these liabilities are independent of one another. Finally, the combined model (CM) postulates the existence of separate dimensions, while allowing for the possibility that these dimensions are correlated. For more details, see Vink et al. (2005).

Model fitting: Age at first cannabis use in the first twin was cross-classified with age at first cannabis use in the second twin, resulting in 4x4 contingency tables for each of the five zygosity groups: monozygtic males, dizygotic males, monozygotic females, dizygotic females and dizygotic opposite sex twins. The three models were then fitted to the five contingency tables using maximum likelihood in the structural equation modelling package

MX (Neale et al., 2006), and the goodness-of-fit of the (nested) models was assessed using likelihood-ratio chi-square statistics. For the model that gave the best description of the data, the twin correlations in liability were expressed as a function of genetic and environmental parameters based on the classical twin design (Neale and Cardon, 1992). Sources of variation that were considered in modelling were additive genetic variation (A), shared environmental variation (C) and unique environmental variation that is not shared by twin pairs (E).

Effective sample size for SNP-based heritability estimation

Estimates of heritability from LD score regression and the So et al. method depend on the relationship between sample size, effect size, and corresponding test statistic. Using the Cox proportional hazards model and applying genomic control both affect that relationship in the current analysis. Therefore we approximate the effective sample size (i.e. the sample size with the intended statistical behavior for heritability analysis) of the current GWAS. For simplicity we motivate the derivation of the effective sample size calculation using LD score regression and apply the same calculation with the So et al. method.

First, we note that the statistical power of the Cox proportional-hazards regression model is directly proportional to the number of observed events rather than the number of individuals in study (Schoenfeld, 1983; Hsieh, 2000). As derived by Hsieh, the noncentrality parameter for the chi-square test of the regression coefficient β is:

$$D\sigma^2\beta^2$$

where D is the number of events and σ^2 is the variance of the regression covariate. By comparison, LD score regression assumes a chi square statistic with noncentrality parameter

$$N\beta^2$$

where the effect size beta is defined assuming a standardized genotype and standardized phenotype. Using a standardized genotype implies $\sigma^2 = 1$, leaving non-centrality parameter $D\beta^2$, confirming that the number of events D rather than the number of individuals N is the appropriate effective sample size in the Cox proportional hazards model.

Second, we note that the use of genomic control in each cohort will proportionately reduce the chi-square statistic. As derived by Bulik-Sullivan et al. (2015), the use of genomic control modifies the relationship between heritability and expected chi-square statistic to replace the sample size with

$$\frac{1}{N}\sum_{j}\sum_{k}\frac{N_j N_k}{\sqrt{\lambda_j \lambda_k}}$$

where λ_j is the genomic control factor and N_j is the sample size for each study *j* and *N* is to total sample size.

Putting this together with the adjustment for power in the Cox proportional hazard model gives a final effective sample size of:

$$N_{effective} = \frac{1}{D}\sum_{j}\sum_{k}\frac{D_j D_k}{\sqrt{\lambda_j \lambda_k}}$$

D is the number of uncensored observations (i.e. the number of individuals with observed cannabis initiation).

Note that neither of these adjustments account for the use of family data analzed allowing correlated standard errors, which will further reduce the effective sample size. As such, the effective sample size described above is likely to be too large, and thus our estimates of heritability will be deflated proportional to the amount of relatedness in the analyzed cohorts.

References:

Schoenfeld, D.A. (1983). Sample-size formula for the proportional-hazards regression model. *Biometrics*, 39, 499-503.

Hseih, F.Y. & Lavori, P.W. (2000). Sample-size calculations for the Cox proportional hazards regression model with nonbinary covariates. *Controlled Clinical Trials*, 21, 552-560.

Bulik-Sullivan, B.K. et al. (2015). LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, 47, 291-295.

Combined data of NTR, QIMR, BLTS

Table I. Descriptives of the three samples:

	QIMR	**BLTS**	**NTR**
N MZ twin pairs	1282	429	2027
N DZ twin pairs	1969	577	1771
Total N twin pairs (MZ + DZ)	3251	1006	3798
Percentage males	44%	43%	34%
% ever used cannabis	57%	51%	24%
% early initiators (<18 years) among ever users	36%	51%	56%

Table II. Goodness-of-Fit of the Single Liability Dimension (SLD), the Independent Liability Dimension (ILD) and the Combined Model (CM)

	df	χ^2	AIC
SLD	64	369.539	241.539
ILD	59	280.857	162.857
CM	**57**	**149.619**	**35.619**

df=degrees of freedom; AIC=χ^2 - 2df, this is a measure of the parsimony of the model, a lower AIC indicates a more parsimonious model.

Table III. Polychoric twin correlations for age at cannabis initiation under the Combined Model.

Combined model	R initiation	95% CI	R age at initiation	
MZM	.85	.77-.91	.78	.69-.86
DZM	.66	.51-.79	.61	.43-.74
MZF	.85	.78-.90	.78	.68-.85
DZF	.54	.37-.67	.69	.53-.80
DOS	.63	.52-.75	.50	.33-.63

Table IV. Model fitting results for a combined model with cannabis initiation and age at first use (best fitting model is given in boldface). Full = full model with qualitative (Rc dos free) and quantitative (ACE separately for males and females) sex difference. Rcdos fix = full model without qualitative sex differences for shared environment (shared environmental correlation in DOS twins is fixed at 1). ACE= full model without quantitative sex differenced (ACE m=f), AE = model without shared environmental factors, CE = model without additive genetic factors.

	Initiation	Age	Chi square	Df	Delta chi2	Delta df	versus	P	AIC
1	Full	Full	150.3980	72					6.398
2	Rcdos fix	Full	150.398	73	0.0001	1	1	.992	4.398
3	ACE	Full	152.976	76	2.577	3	2	.461	0.976
4	AE	Full	185.043	77	32.067	1	3	<.001	31.043
5	CE	Full	194.876	77	41.901	1	3	<.001	40.876
6	Full	Rcdos fix	152.442	73	2.044	1	1	.153	6.442
7	Full	ACE	157.457	76	5.015	3	6	.171	5.457
8	Full	AE	190.469	77	33.012	1	7	<.001	175.496
9	Full	CE	187.316	77	29.859	1	7	<.001	33.316
10	**ACE**	**ACE**	**159.361**	**80**	**8.964**	**8**	**1**		**-0.639**

Table V. A. Heritability estimates from the **full model** under the Combined Model.

	A2	C2	E2
Cannabis initiation M	.38 (.10-.65)	.48 (.23-.72)	.14 (.08-.22)
Cannabis initiation F	.54 (.31-.73)	.30 (.14-.52)	.16 (.10-.22)
Age at initiation M	.35 (.05-.70)	.43 (.11-.70)	.22 (.14-.31)
Age at initiation F	.20 (0 - .53)	.57 (.27 - .78)	.23 (.15-.33)

Estimate for shared environmental correlation in DOS twins for cannabis initiation: 1

Estimate for shared environmental correlation in DOS twins for age at initiation: .77

Table V. B. Heritability estimates from the **best fitting model** under the Combined Model.

	A2	C2	E2
Cannabis initiation	.48 (.30-.65)	.37 (.21-.52)	.15 (.11-.20)
Age at initiation	.38 (.19-.60)	.39 (.20-.56)	.22 (.16-.29)

Estimate for shared environmental correlation in DOS twins for cannabis initiation: fixed at 1

Estimate for shared environmental correlation in DOS twins for age at initiation: fixed at 1

GENOME-WIDE ASSOCIATION META-ANALYSIS OF AGE AT FIRST CANNABIS USE – REPLICATION ANALYSES

We performed a GWAS meta-analysis of age at first cannabis use in a discovery sample of 24,953 individuals from nine cohorts from Europe, Australia, and the United States. The top findings obtained in the single nucleotide polymorphisms (SNP) and gene-based analyses were tested for replication.

MATERIALS AND METHODS

The replication sample comprised of 3,735 individuals with a mean age ranging from 24 to 49.4 years (Table 1S5). Females represented 53.9% of the sample, and 45.3% of the observations were uncensored (see Supplementary Table S1 for more details on the samples).

Table 1S5: Descriptive information on the participating replication cohorts.

Cohort	N	% Females	%Uncensored Observations	Mean age (SD)	Mean age at first use (sd) (in users)	Number of SNPs
CADD	1060	40.18	78	24.06 (4.2)	14.2 (3.16)	8*
NTR2**	1740	63.7	22.2	35.0 (14.6)	18.0 (4.0)	8*
RADAR**	342	44.7	57.0	19.5 (0.8)	15.9 (1.7)	8*
SYS	593	55.3	47.7	49.4 (5.1)	16.2 (1.7)	8*

*CADD - the Center on Antisocial Drug Dependence; NTR – the Netherlands Twin Register sample 2; RADAR: Research on Adolescent Development and Relationships; N = sample size, % uncensored observations (i.e., individuals who have initiated cannabis use). Mean age: age when completing survey or interview. Mean age at first use: mean age at first cannabis use * In the replication samples only the top 8 independent SNPs were tested. ** The NTR2 and the RADAR samples were combined and the analysis was performed in this combined sample.*

Phenotyping

For details on phenotyping in the participating samples we refer to the main manuscript, and to Supplementary File S1 for information on the exact phrasing of the question used to collect the data by each replication cohort.

Genotyping

Details on genotyping in the GWAS participating cohorts are included in the main manuscript. For information on the extensive quality control (QC) performed by each participating cohort we refer to the Supplementary Table S2.

Imputation

All participating cohorts performed genotype imputation using the 1000 Genomes Phase 1 March 2012 release as reference (40) (see Supplementary Table S2 for further imputation details).

Quality checks prior to meta-analysis

The same quality checks performed in the discovery sample (see main manuscript for details) were applied in the replication cohorts.

Power analysis

We evaluated the power to detect a significant association in the replication sample using the R library "powerSurvEpi".

Statistical analysis of individual samples, meta-analysis and gene-based tests of association

We implemented the same procedures and options in the replication phase as in the discovery meta-analysis. Statistical analyses were performed on the Lisa Genetic Cluster Computer (http://www.geneticcluster.org).

Polygenic score analysis

Polygenic score analyses were carried out to determine if age at first cannabis use could be predicted in the replication samples. Results from the GWAS discovery meta-analysis were used to create polygenic scores in an independent sample from the Netherlands (the combined sample of NTR2-RADAR (see Table 1, and Supplementary Tables S1, S2, and Supplemental File S1). We used LDpred (55) to take into account LD among the SNPs when creating the polygenic scores. The polygenic scores were generated by calculating the mean causal effect size of each marker using the SNP effect sizes from the GWAS discovery meta-analysis and the LD structure from the European populations in the 1000 Genomes Phase I reference set.

Polygenic scores were computed based on the genome-wide meta-analysis results from which we selected only genotyped SNPs (i.e. non-imputed) present in at least 7 of the discovery cohorts. The final number of SNPs included was 376,819. Polygenic scores were calculated for several expected fractions of causal genetic markers to optimize prediction accuracy (0.1%, 1%, 10%, and 100%). The scores were transformed into z-values before analysis. We then tested if the computed polygenic scores predicted age at first cannabis use in the independent target cohort using a Cox proportional hazards regression in R (as applied in the main analyses). Age at first cannabis use (or age at the last survey for censored observations) was regressed on the polygenic scores. Sex, birth cohort, and three ancestry-

informative genetic PCs were included as covariates in the model. To account for family relatedness we used the 'cluster' option implemented in the survival R-package.

RESULTS

Power analysis

Table 2-S5 displays the results of our power analysis.

Beta	Hazard Ratio	MAF	alpha	Power
0.09	1.094174	0.14	0.005	0.07
-0.06	0.941765	0.77	0.005	0.04
0.07	1.072508	0.46	0.005	0.09
-0.07	0.932394	0.27	0.005	0.07
0.07	1.072508	0.8	0.005	0.05
-0.11	0.895834	0.07	0.005	0.04
0.06	1.061837	0.73	0.005	0.04
-0.06	0.941765	0.74	0.005	0.04

The power to replicate the top 8 SNPs was low, ranging from 0.04 to 0.10.

GWAS meta-analysis

In the independent replication samples, none of the 8 tested SNPs replicated (see Table 3-S5 for details). Note that the top SNP remained significant in the combined discovery and replication meta-analysis (P=8.7E-09).

Table 3-S5*. Results for the top 8 independent SNPs in the meta-analysis of the discovery samples (present in at least one replication sample), and results of the meta-analysis of combined discovery and replication samples. SNPs are displayed when not in linkage*

disequilibrium (R^2<0.1). For SNPs with R^2 >= 0.1 only the most significant SNP is shown in the top 8.

						Replication			***Combined$***	
SNP	**Chr**	**BP (hg19)**	**A1**	**A2**	**Freq A1**	**beta (s.e.)**	**P**	**Direction ***	**beta (s.e.)**	**P**
rs1574587	16	84453056	T	C	0.1415	-0.09 (0.135)	0.50	??-	0.09 (0.016)	8.7x10^{-9}
rs4935127	10	56654986	C	G	0.7741	0.07 (0.04)	0.07	?++	-0.05 (0.012)	2.9x10^{-5}
rs2249437	6	1595216	T	C	0.4595	-0.14 (0.093)	0.13	??-	0.06 (0.013)	2.1x10^{-6}
rs9266245	6	31325702	A	G	0.2655	-0.01 (0.042)	0.85	?-+	-0.06 (0.014)	4.7x10^{-6}
rs28622199	8	5392103	T	C	0.8012	-0.04 (0.039)	0.36	+--	0.05 (0.014)	5.4x10^{-5}
rs215069	16	16091237	T	C	0.0685	0.02 (0.069)	0.80	+-?	-0.10 (0.024)	2.3x10^{-5}
rs4924506	15	41129467	A	C	0.7318	-0.02 (0.039)	0.55	+-+	0.05 (0.012)	4.2x10^{-5}
rs7773177	6	139143088	A	G	0.7383	0.03 (0.038)	0.48	+++	-0.05 (0.013)	7.7x10^{-5}

** Direction per sample: allele A1 increases (+) or decreases (-) liability for cannabis use, or sample did not contribute to this SNP because it did not pass the post-imputation quality control (?). Order of samples in the replication meta-analysis: CADD, NTR2/RADAR, SYS. Sample information can be found in Table 1-S5.*

Chr = Chromosome; BP (hg19) = location in base pairs in human genome version 19, A1 = allele 1, A2 = allele 2, Freq A1 = Frequency of allele 1, s.e. = standard error, P = p-value.

$The combined sample contains the discovery samples and the CADD, NTR2/Radar and SYS replication samples.

Gene-based tests of association

The *ATP2C2* gene did not reach significance in the meta-analysis of the replication samples (P=0.47).

Polygenic score analysis

None of the polygenic scores for age at first cannabis use explained significant proportions of variance in age at first use in the independent target sample NTR2-RADAR; all polygenic scores based on different expected fractions of causal markers yielded p-values > 0.10.