PNAS

Supplementary Information for
Genome Wide Association Studies of the Kynurenine Pathway

Chinh Bkrong Nguyen[1]*, Arve Ulvik[2], Tore Ivar Aarsland[3], Adrian McCann[2], Øyvind Helgeland[1], Grethe Sopp Tell[4], Rolv Lie[4], Jan Haavik[3,5], Ottar Nygård[6,7], Per Magne Ueland[2], Pål Rasmus Njølstad[1 8], Stefan Johansson[1,9]

1 Center for Diabetes Research, Department of Clinical Science, University of Bergen, Norway
2 Bevital AS, University of Bergen, Norway
3 Department of Biomedicine, University of Bergen, Norway
4 Department of Global Public Health and Primary Care, University of Bergen, Norway
5 Division of Psychiatry, Haukeland University Hospital, Bergen, Norway
6 Department of Heart Disease, Haukeland University Hospital, Bergen, Norway
7 Mohn Nutrition Research Laboratory, Department of Clinical Science, University of Bergen, 5020 Bergen, Norway
8 Children and Youth Clinic, Haukeland University Hospital, Bergen, Norway
9 Department for Medical Genetics, Haukeland University Hospital, Bergen, Norway

* Corresponding author: Chinh Bkrong Nguyen
Email: chinh.nguyen@uib.no

This PDF file includes:

Supplementary text (Quality Control/Methods)
Supplementary Figures: *Fig. S1-4*
SI References

Other supplementary materials (.zip) for this manuscript include the following:

Tables S1 to S7 (.xlsx): Characteristics of the study participants, GWAS summary statistics of top hits and Molecular/Trait annotation for all GWAS signals.

Supplementary Information Text

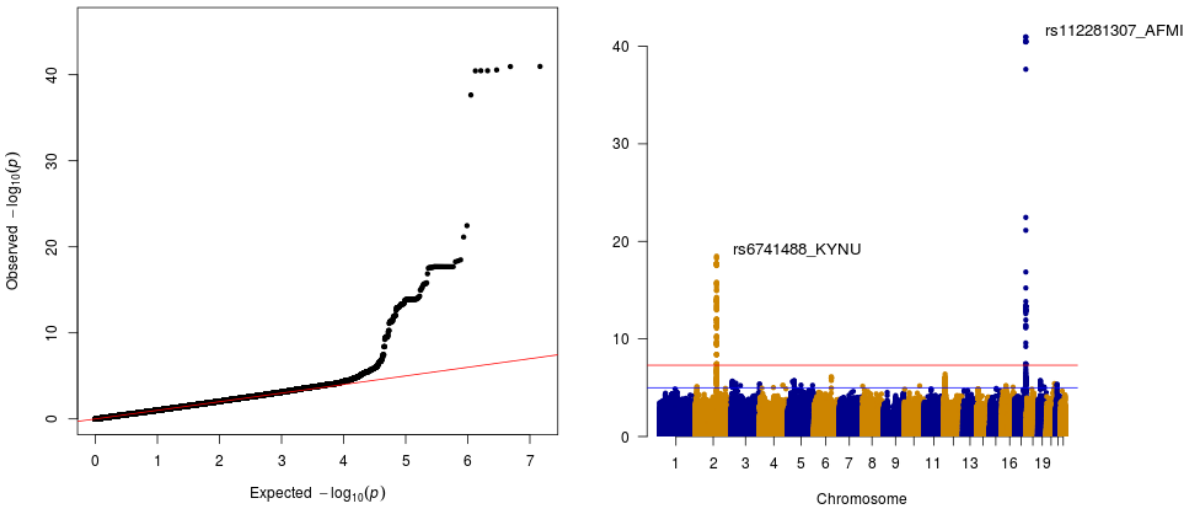
Methods

Quality Control

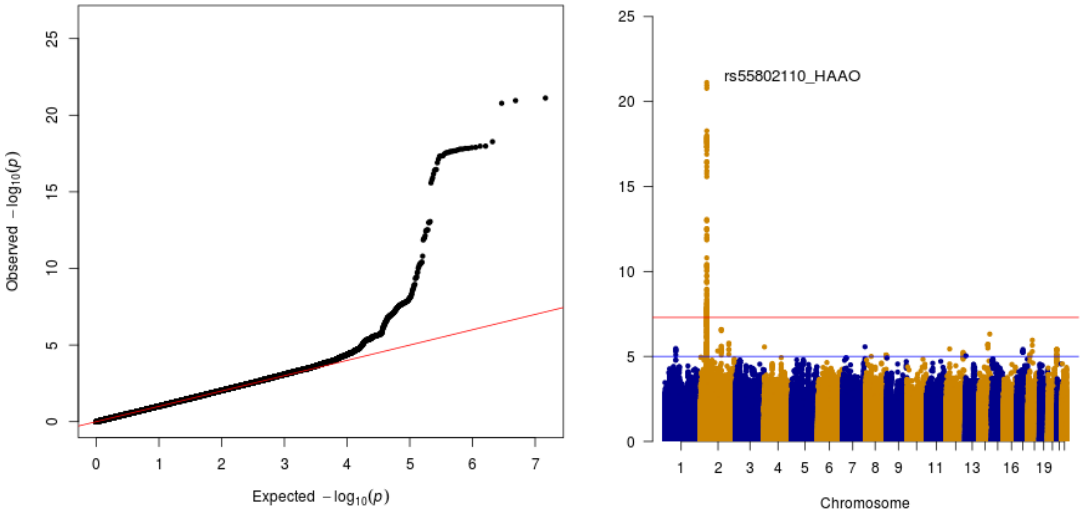
The genotype quality control (QC) procedure was set up based on “per-individual’s cleaning prior to per-marker’s cleaning“ as described by Anderson et al(1). We performed QC using PLINK 1.9 with the following criteria: a) sample call rate >=98%, b) marker call rate >=98%, c) heterozygosity rate<=4SD, d) minor allele frequency (MAF)>=1%, e) deviation from Hardy-Weinberg equilibrium (HWE) <1E-6, f) sex check (X chromosome homozygosity [Female <0,2 and Male >0,8]), g) homogenous population and non-relatedness.

The first round of QC was aimed at removing poorly genotyped samples and identifying a set of unrelated samples of European ancestry. Temporarily, strict thresholds were used to clean markers to facilitate sample cleaning. These compose of removal of markers lower than 10% minor allele frequency (MAF), 98% call rate, 1E-4 *P-val* HWE deviation and higher than 0.1 for LD squared correlation (r2). Next, non-autosomal and ambiguous markers filter and in-depth pairwise pruning (a 200bp window) returned 31373 SNPs which were subsequently used for sample cleaning. Samples with low call rate (<95%) and the first sample in each pair with Pihat >0.1 for identity by descent (IBD) were removed. In addition to Pihat, we also investigated potential presence of non-European or admixture samples by comparing our data with the European ancestry residents in Utah, USA (CEU) and other populations of the Hapmap v3 project using FlashPCA (2). K nearest neighborhood learning was used to assign the remaining CEU samples to their closest cluster. We used the igraph R package to build networks of related individuals of which edge weights were measured based on pairwise kinship coefficients (computed by relationship inference KING software(3)) and nodes are related HUSK individuals. Pairwise-IBD sharing indicating siblings, parents-offspring, 2nd and 3rd degree relationships of the complex networks (>5 nodes) were validated using IBD segment inference computation and visualization in KING (3). These procedures resulted in a homogeneous ancestral set of 4974 individuals (core samples).

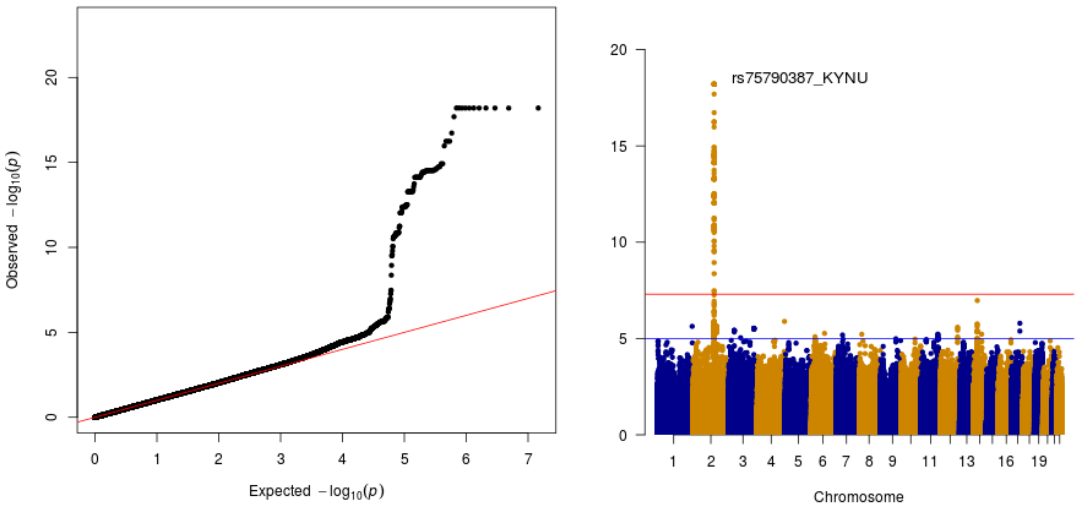
In the second round, we used the core sample set to reassess all genotyped markers of the original array. 492845 markers were cleaned similarly as above using cut-offs for genotyping rate (90%, 95% and 98% respectively) and sample call rate (95%, 97%, 98%); markers with HWE deviation < 10E-7 and 10E-6. Samples with heterozygosity rate outside 4SD were removed. Sex markers in strong LD (>0.5) were assessed by in-depth pairwise pruning of each 2000 markers in a 20000bp window. A threshold for X chromosome homozygosity (Female <0,2 and Male >0,8) was used to remove abnormal X chromosome samples. HWE metric for the X chromosome of only female individuals was assessed. In addition, we removed failed HWE markers belonging to the pseudo-autosomal (PAR region) as well as the heterozygous male X markers and markers in extra physical locations of the PAR and X-translocation regions. After the second round, 4931 individuals and 491901 markers remained.



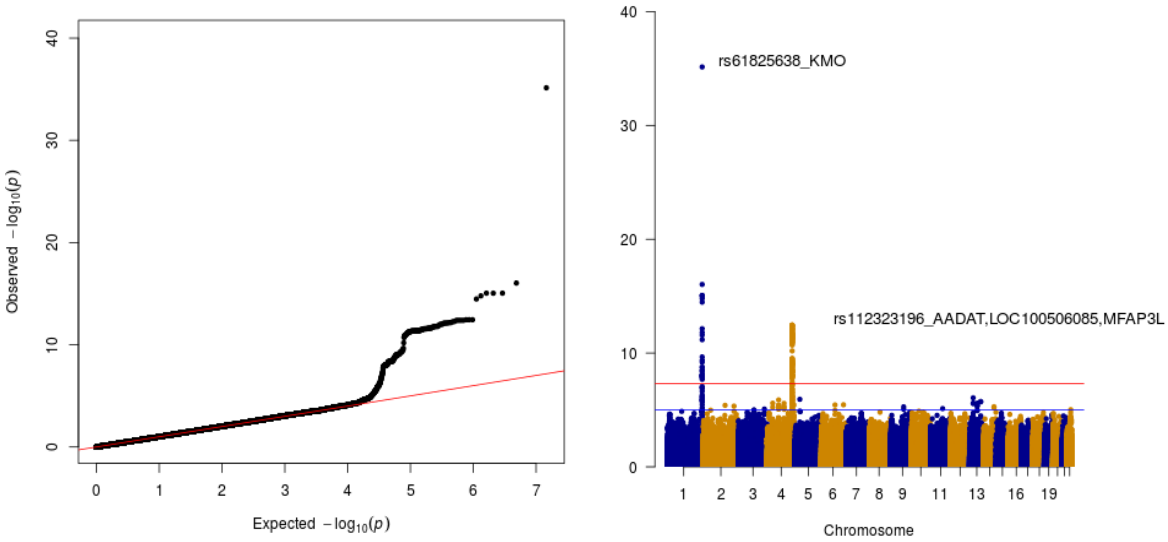
Quantile-Quantile (QQ) plot and Manhattan plot of Anthranilic acid (AA) GWAS



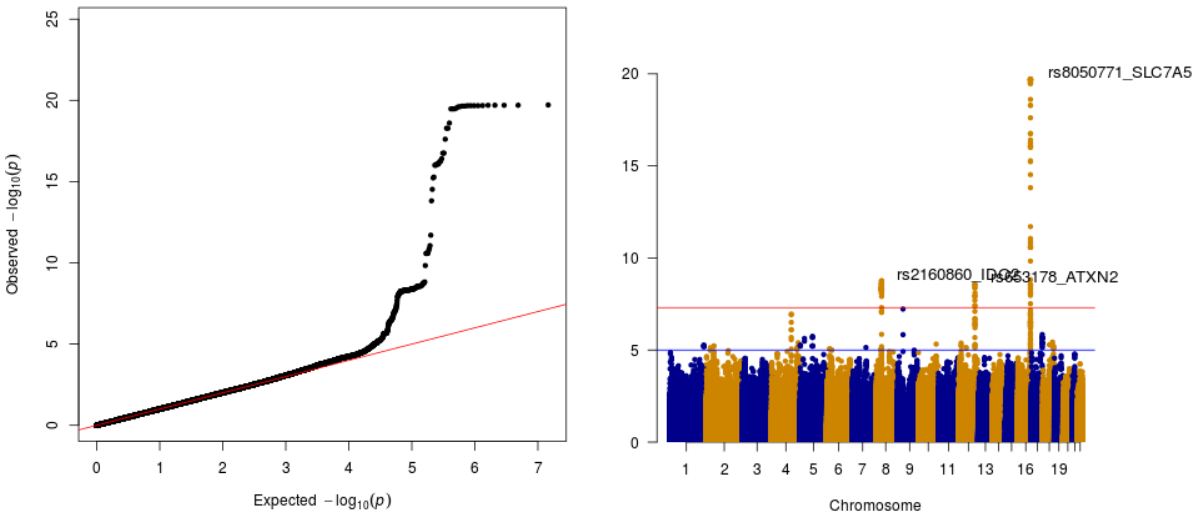
QQ plot and Manhattan plot of Hydroxyanthranilic acid (HAA) GWAS



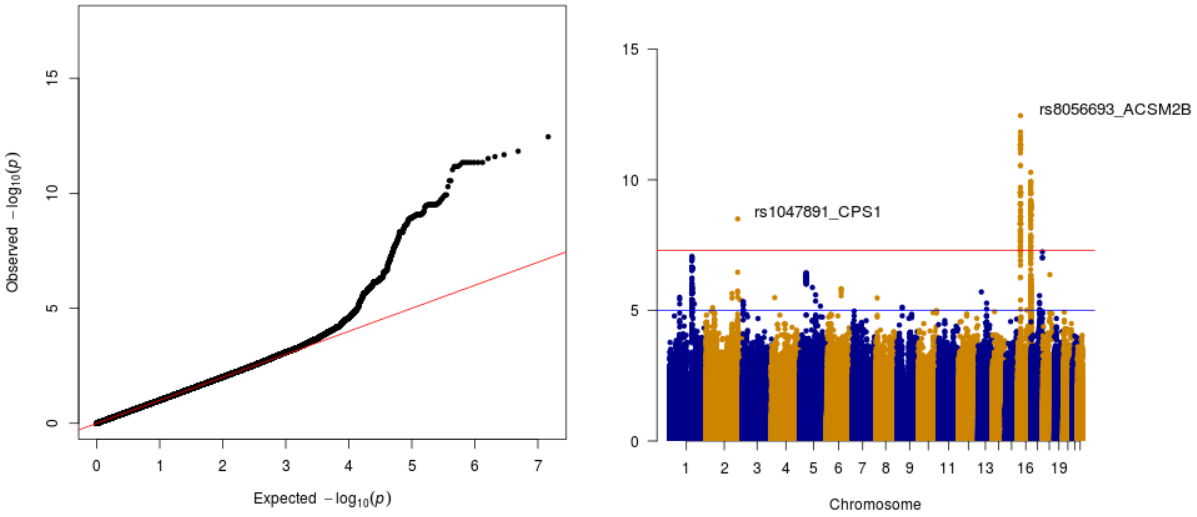
QQ plot and Manhattan plot of Hydroxykynurenine (HK) GWAS



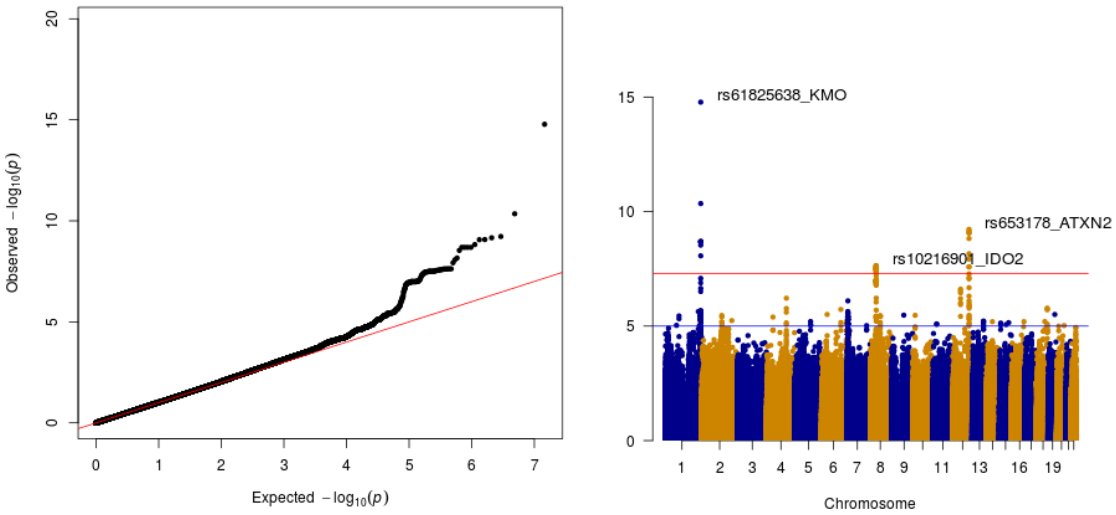
QQ plot and Manhattan plot of Kynurenic acid (KA) GWAS



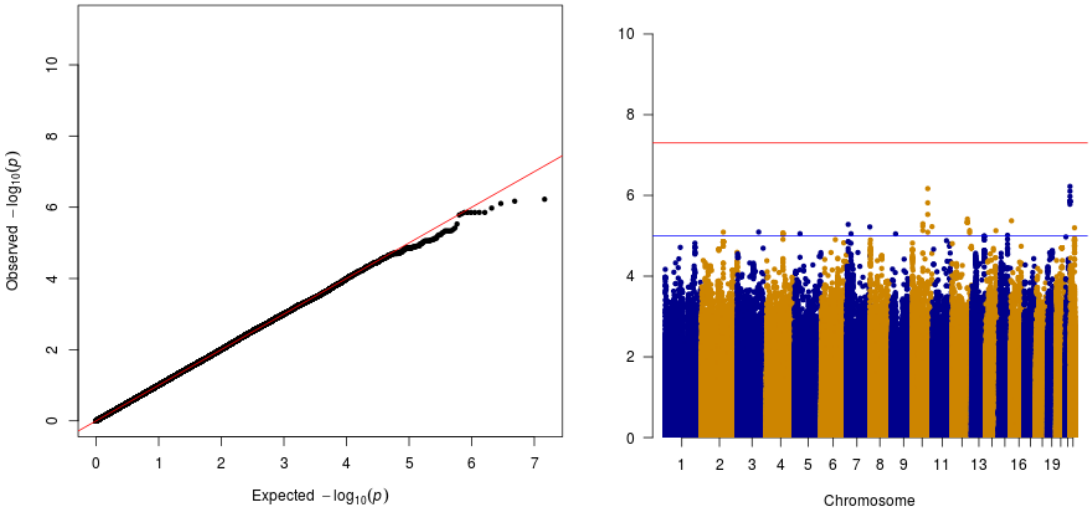
QQ plot and Manhattan plot of Kynurenine (Kyn) GWAS



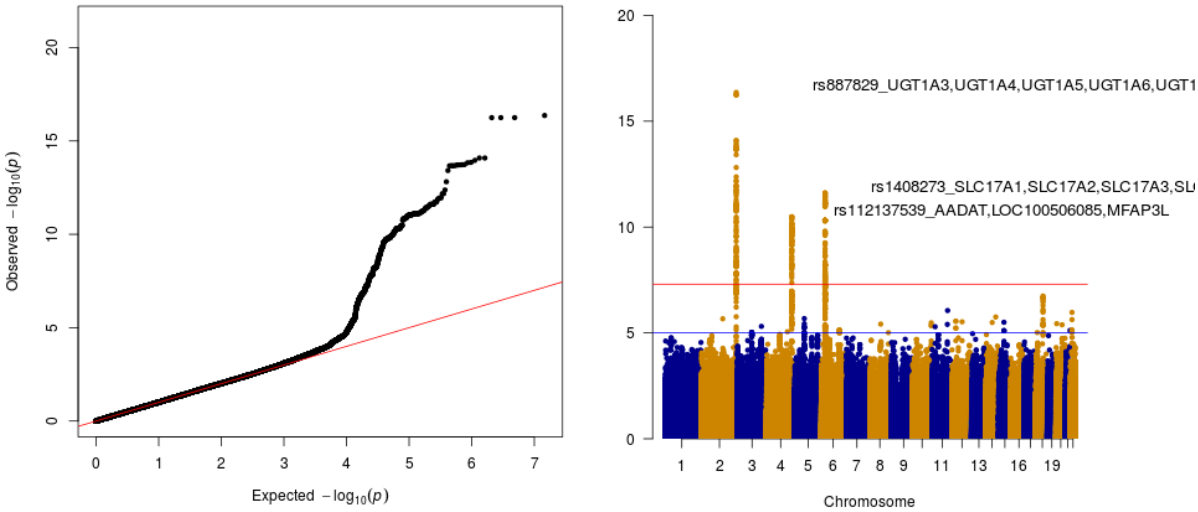
QQ plot and Manhattan plot of Picolinic acid (Pic) GWAS



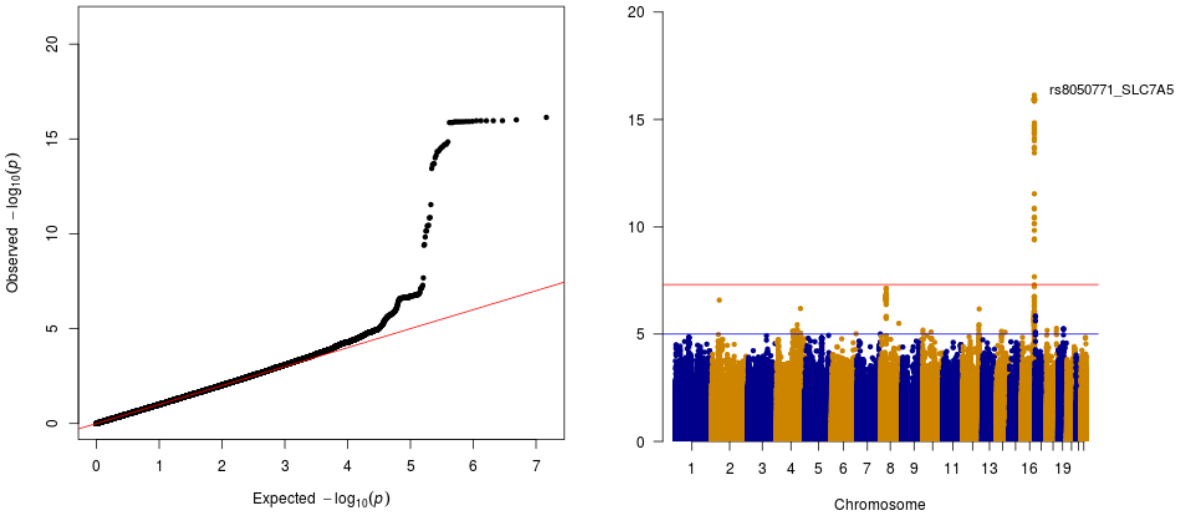
QQ plot and Manhattan plot of Qinolinic acid (QA) GWAS



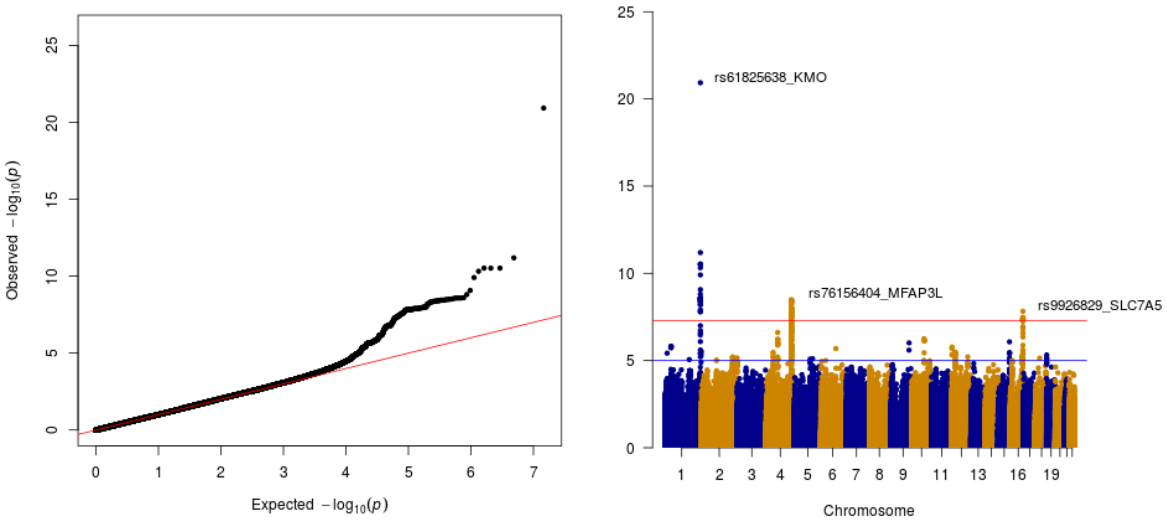
QQ plot and Manhattan plot of Tryptophan (Trp) GWAS



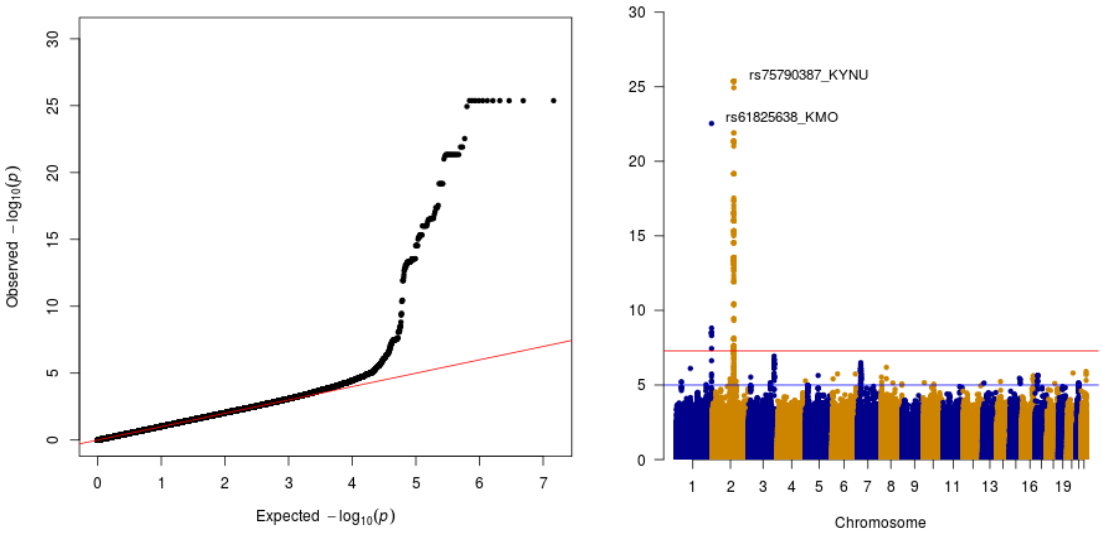
QQ plot and Manhattan plot of Xanthurenic acid (XA) GWAS



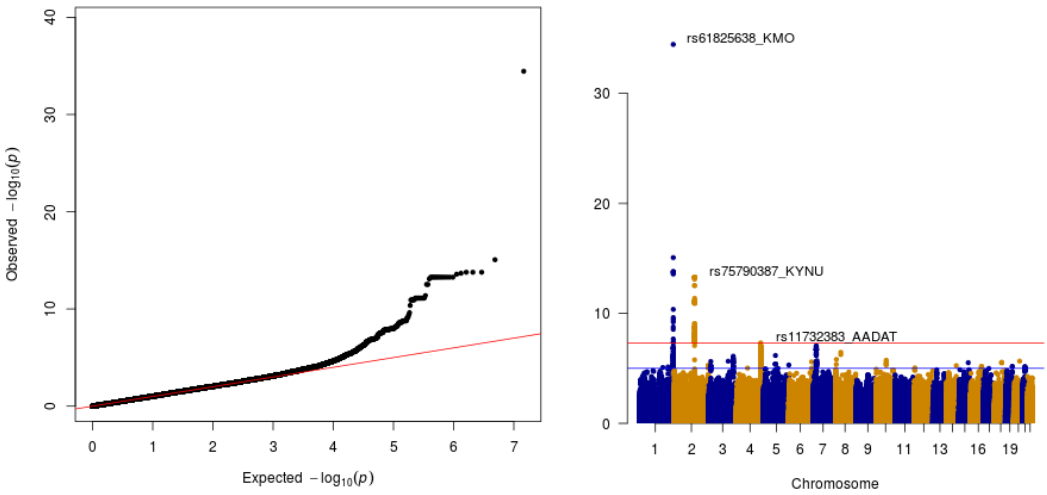
QQ plot and Manhattan plot of Kyn/Trp GWAS



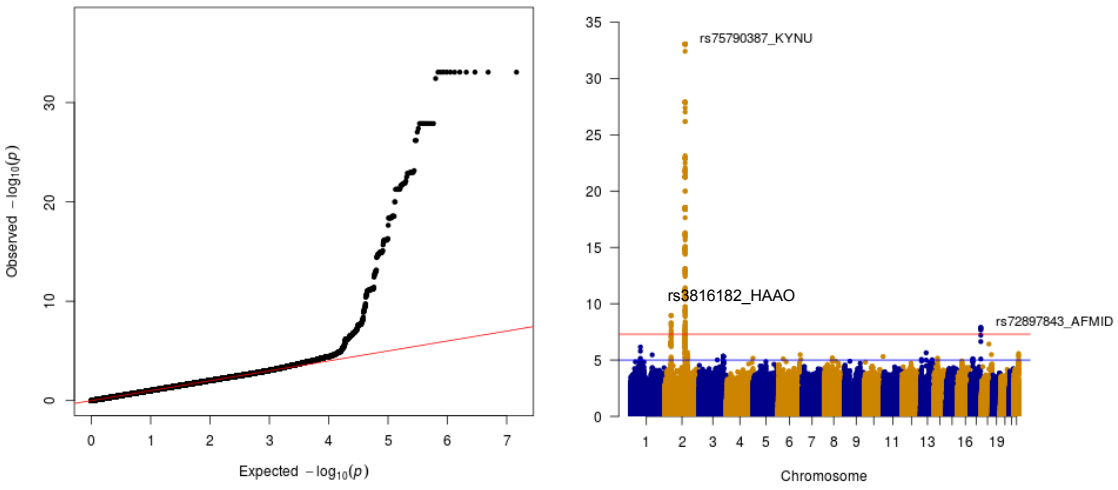
QQ plot and Manhattan plot of Kyn/KA GWAS



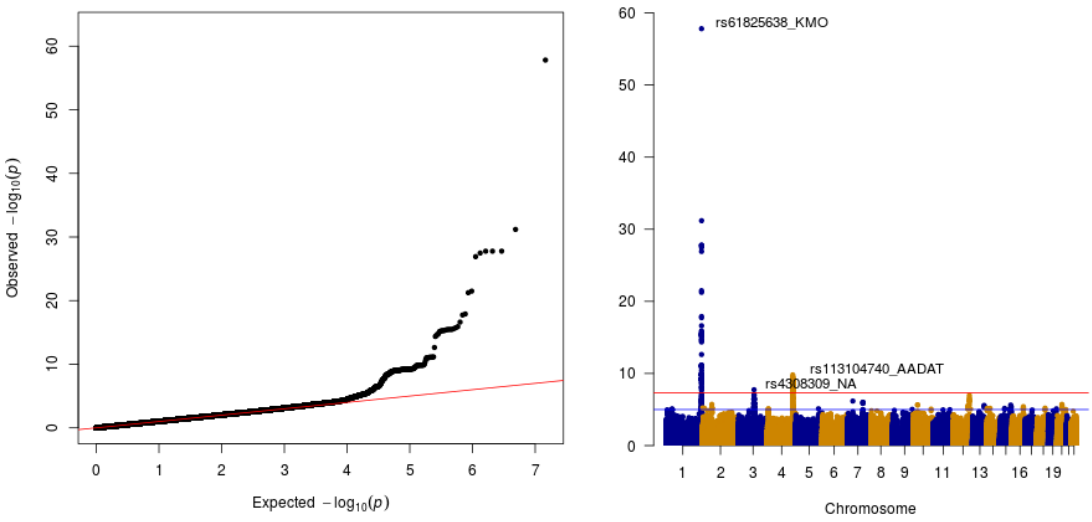
QQ plot and Manhattan plot of HKr GWAS



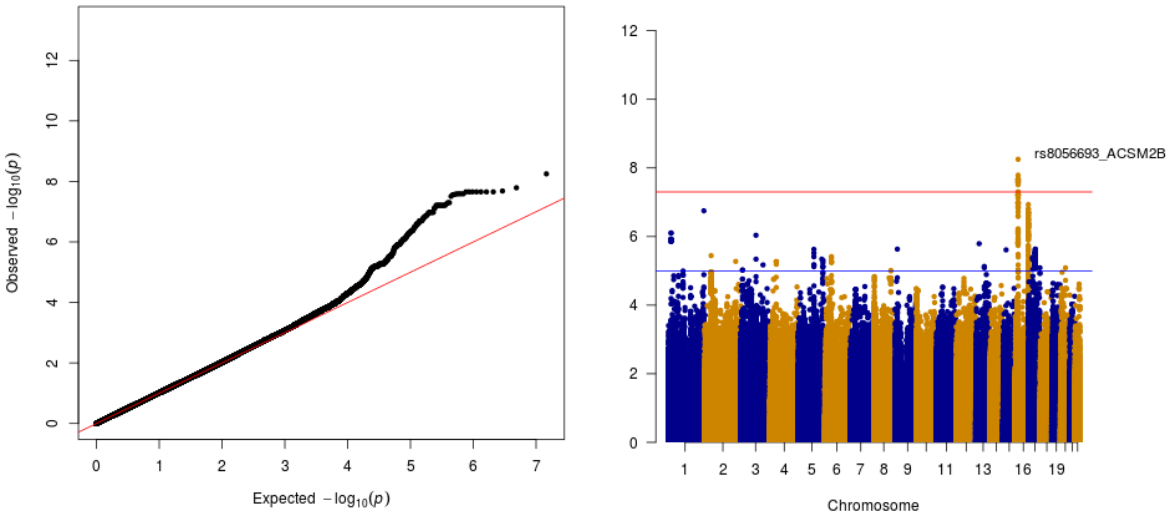
QQ plot and Manhattan plot of HKr_a GWAS



QQ plot and Manhattan plot of HKr_b GWAS

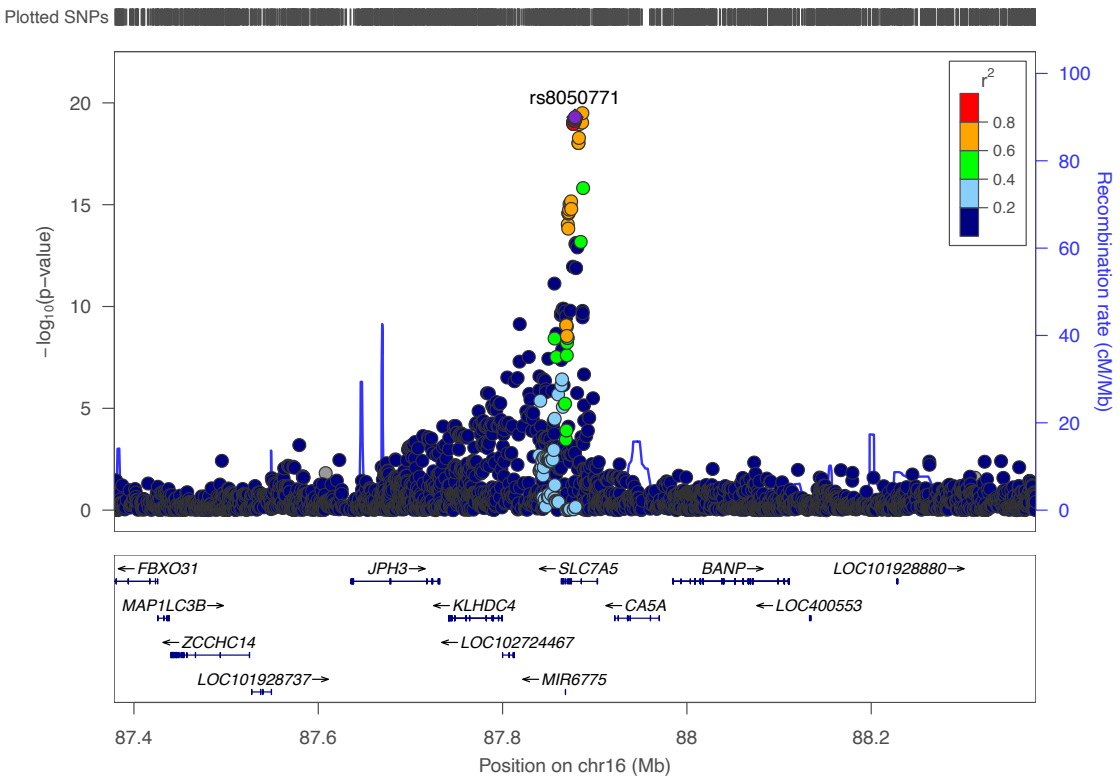


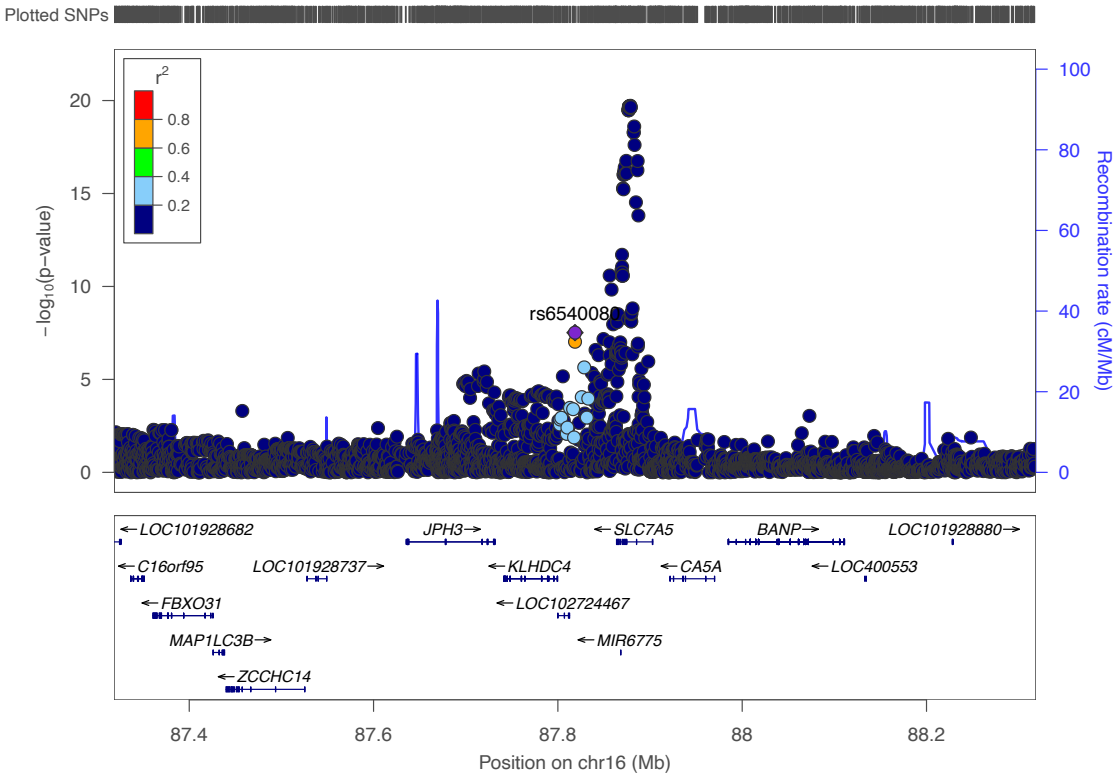
QQ plot and Manhattan plot of KA/QA GWAS

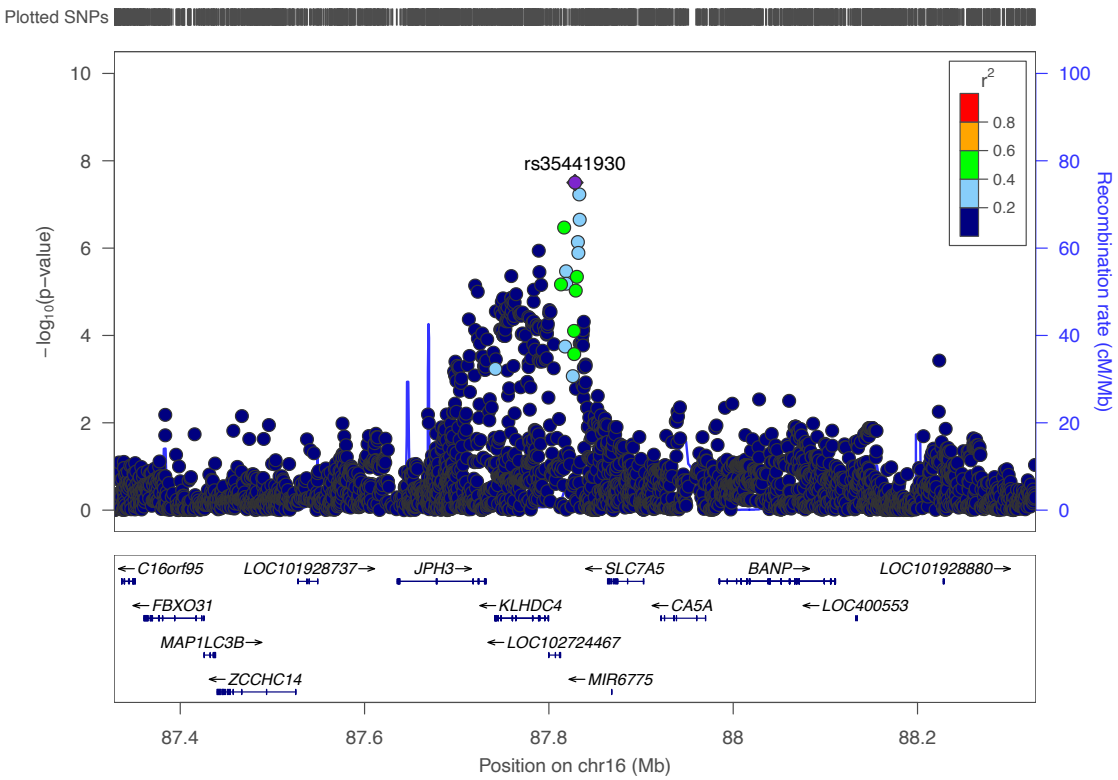


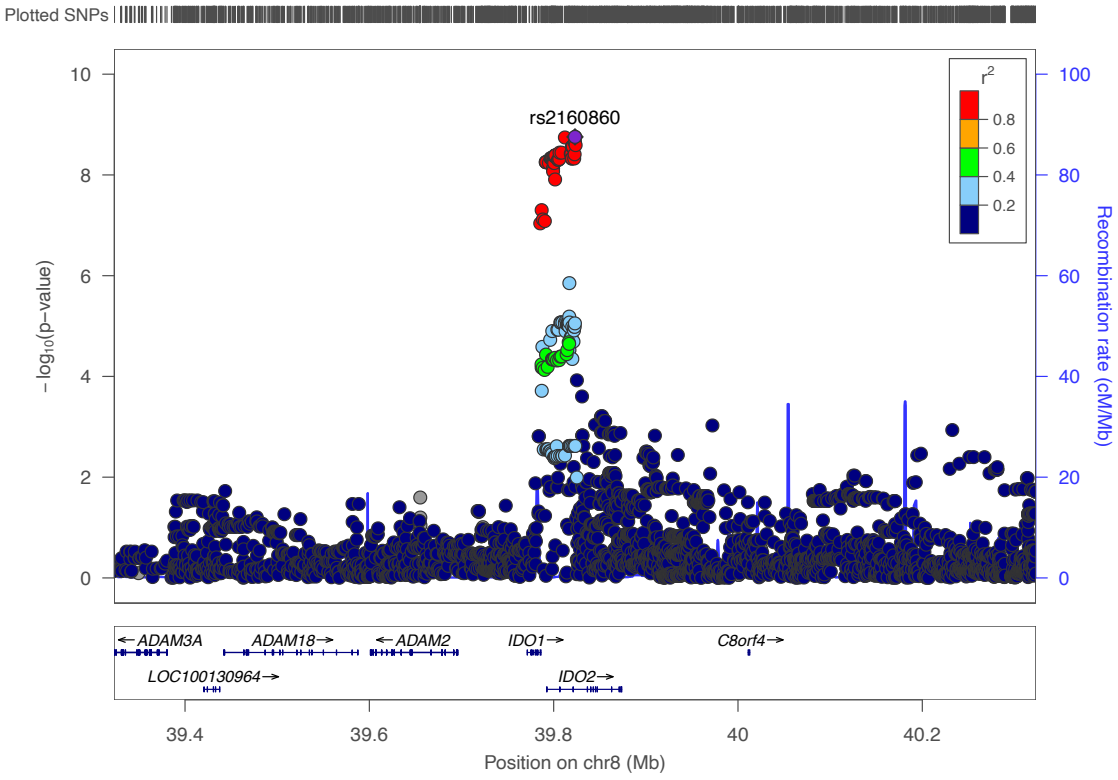
QQ plot and Manhattan plot of Pic/QA GWAS

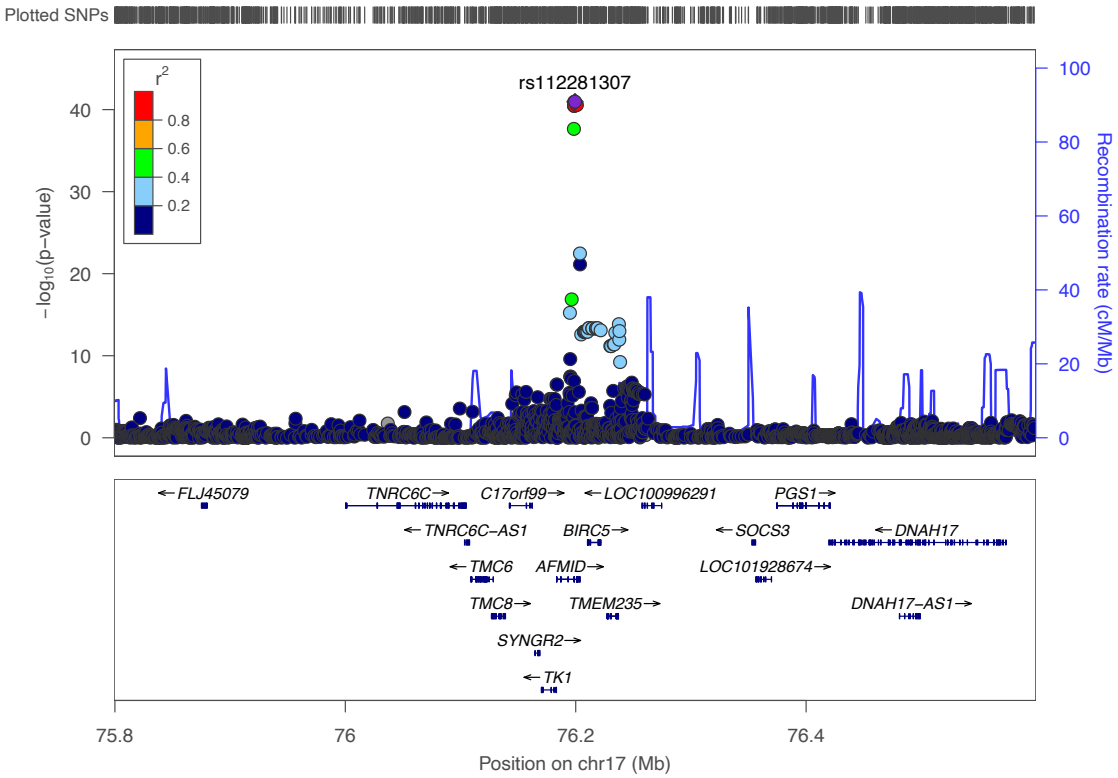
Figure S1: *Quantile-Quantile and Manhattan plots of GWAS results for 9 metabolites and 7 ratios.*

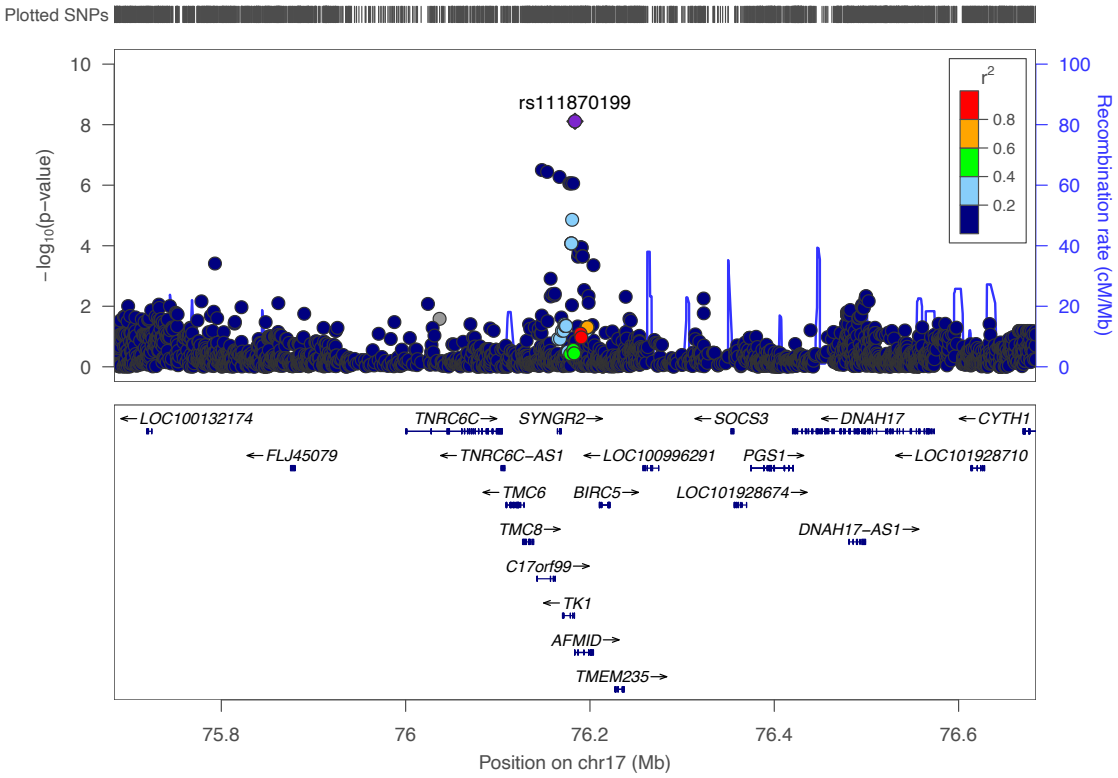

Plotted SNPs
rs8050771
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
FBXO31
JPH3
SLC7A5
BANP
LOC101928880
MAP1LC3B
KLHDC4
CA5A
LOC400553
ZCCHC14
LOC102724467
LOC101928737
MIR6775
87.4
87.6
87.8
88
88.2
Position on chr16 (Mb)

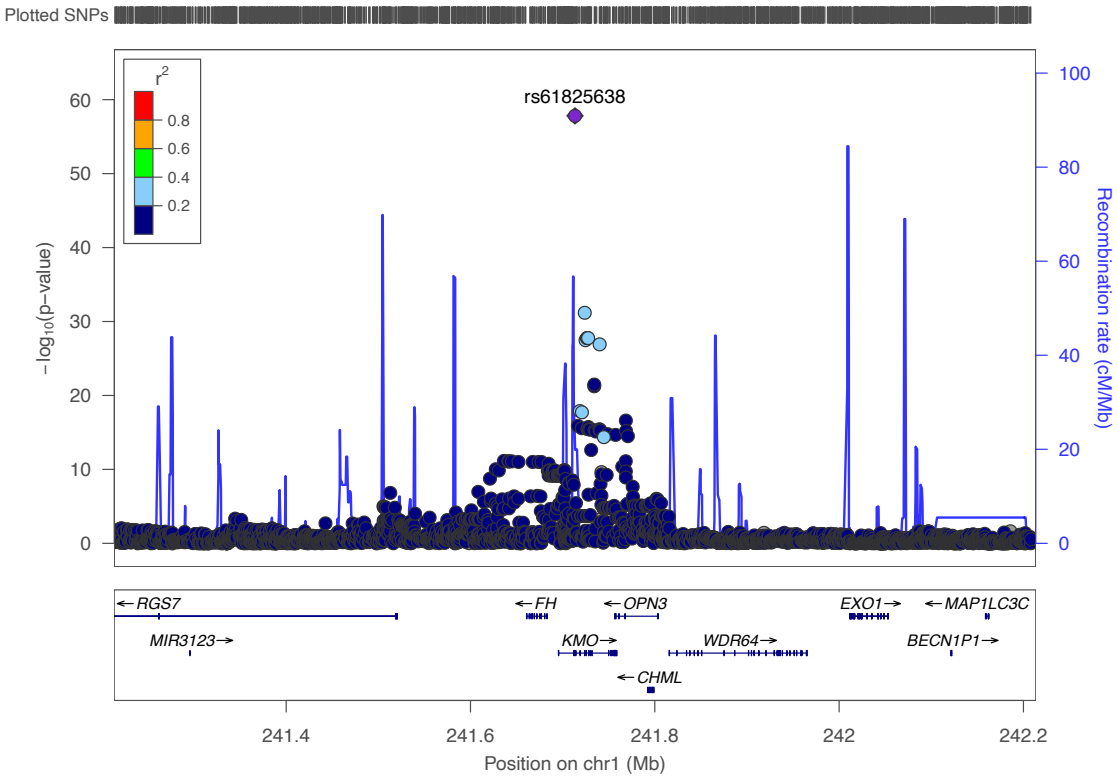

Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs6540080
−log10(p−value)
Recombination rate (cM/Mb)
LOC101928682
JPH3
SLC7A5
BANP
LOC101928880
C16orf95
LOC101928737
KLHDC4
CA5A
LOC400553
FBXO31
LOC102724467
MAP1LC3B
MIR6775
ZCCHC14
87.4
87.6
87.8
88
88.2
Position on chr16 (Mb)

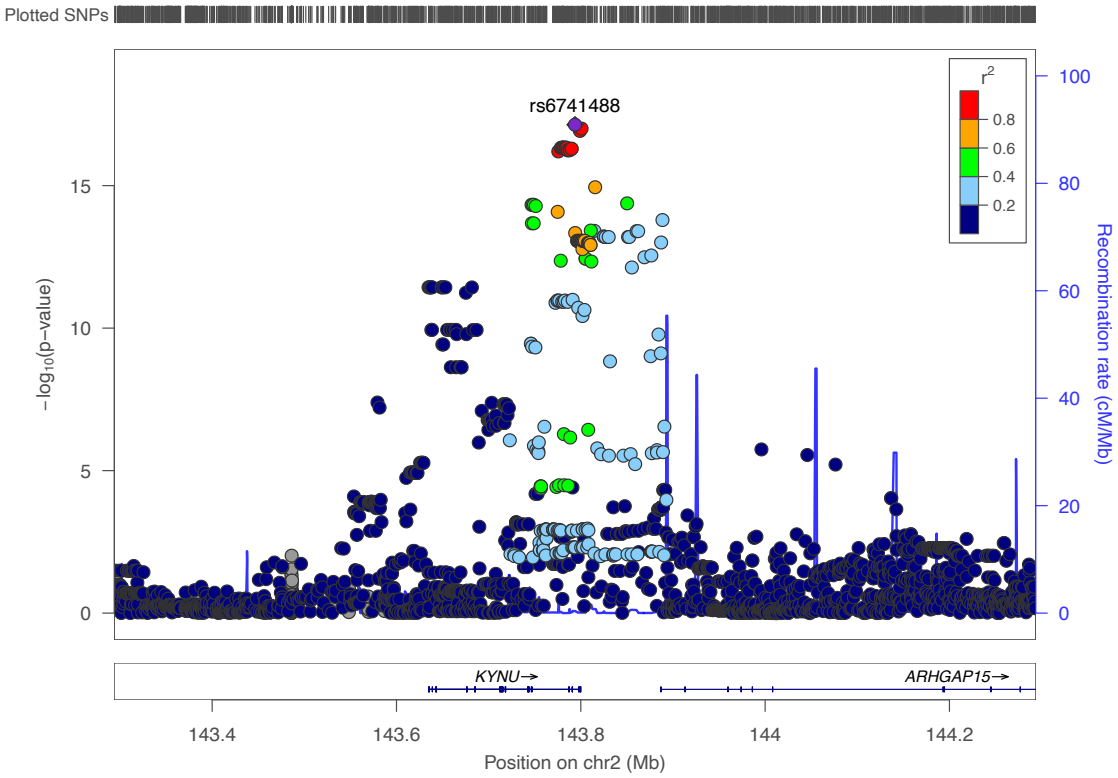

Plotted SNPs
rs35441930
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
C16orf95
LOC101928737
JPH3
SLC7A5
BANP
LOC101928880
FBXO31
KLHDC4
CA5A
LOC400553
MAP1LC3B
LOC102724467
ZCCHC14
MIR6775
87.4
87.6
87.8
88
88.2
Position on chr16 (Mb)

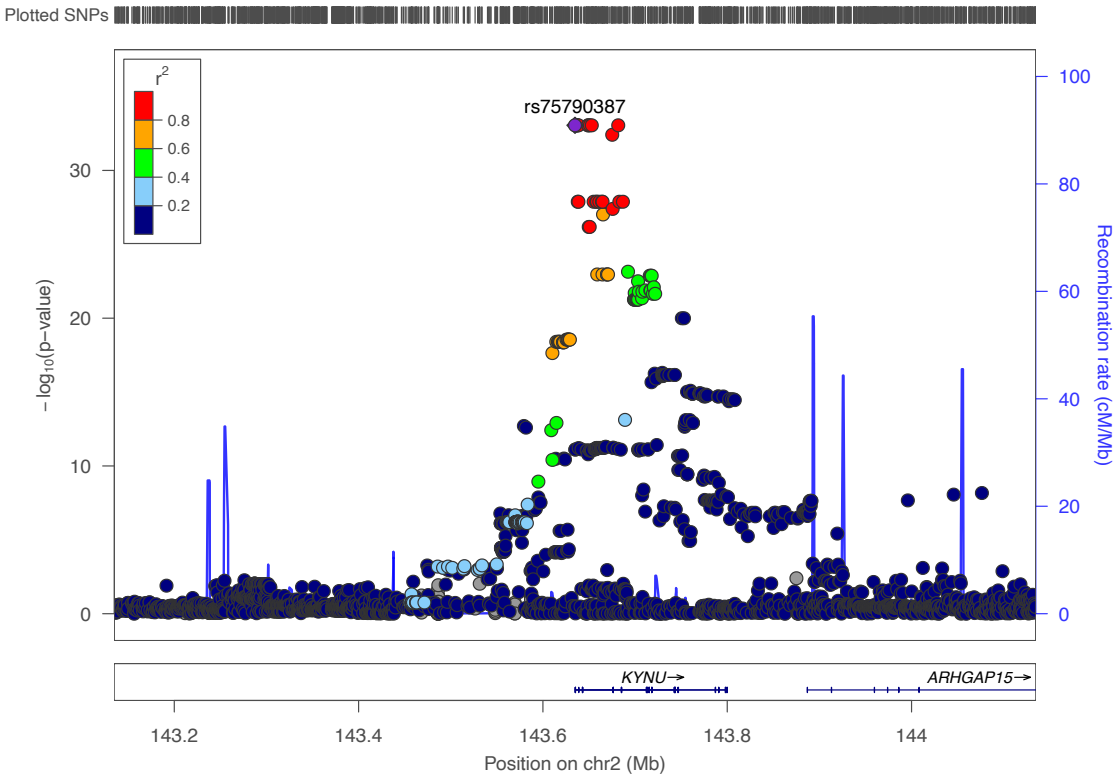

Plotted SNPs
rs2160860
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
ADAM3A
ADAM18
ADAM2
IDO1
C8orf4
LOC100130964
IDO2
39.4
39.6
39.8
40
40.2
Position on chr8 (Mb)

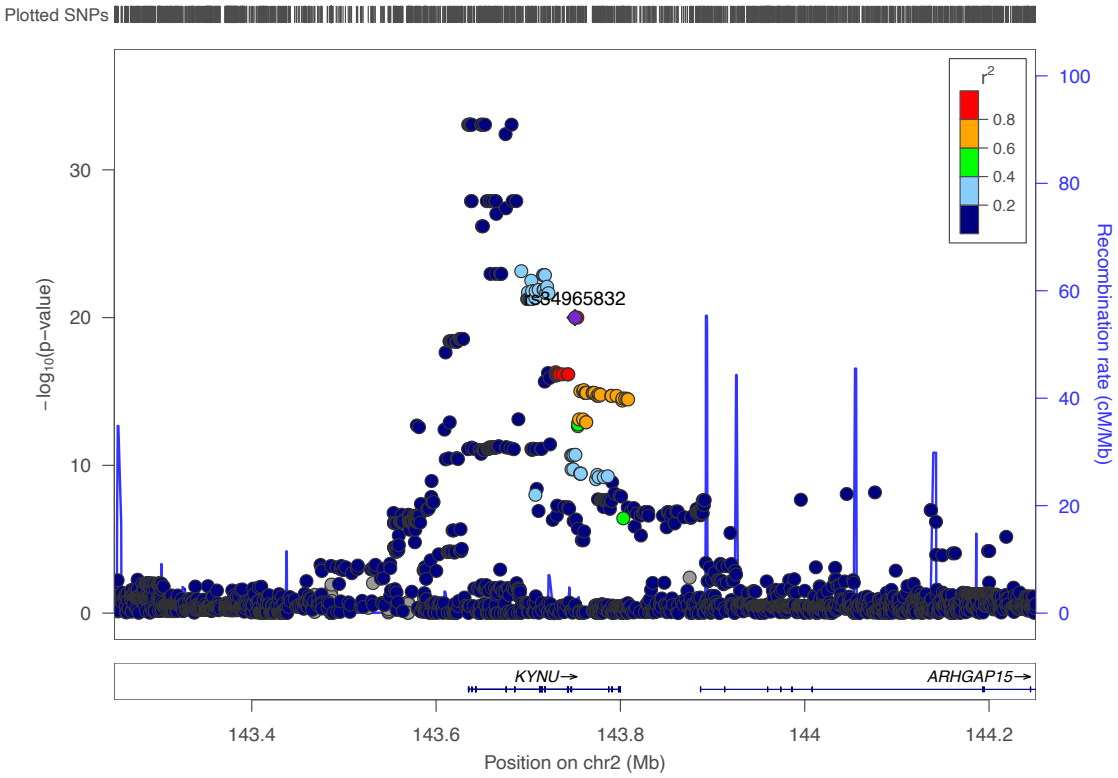

Plotted SNPs
rs112281307
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
FLJ45079
TNRC6C
C17orf99
LOC100996291
PGS1
TNRC6C−AS1
BIRC5
SOCS3
DNAH17
TMC6
AFMID
LOC101928674
TMC8
TMEM235
DNAH17−AS1
SYNGR2
TK1
75.8
76
76.2
76.4
Position on chr17 (Mb)

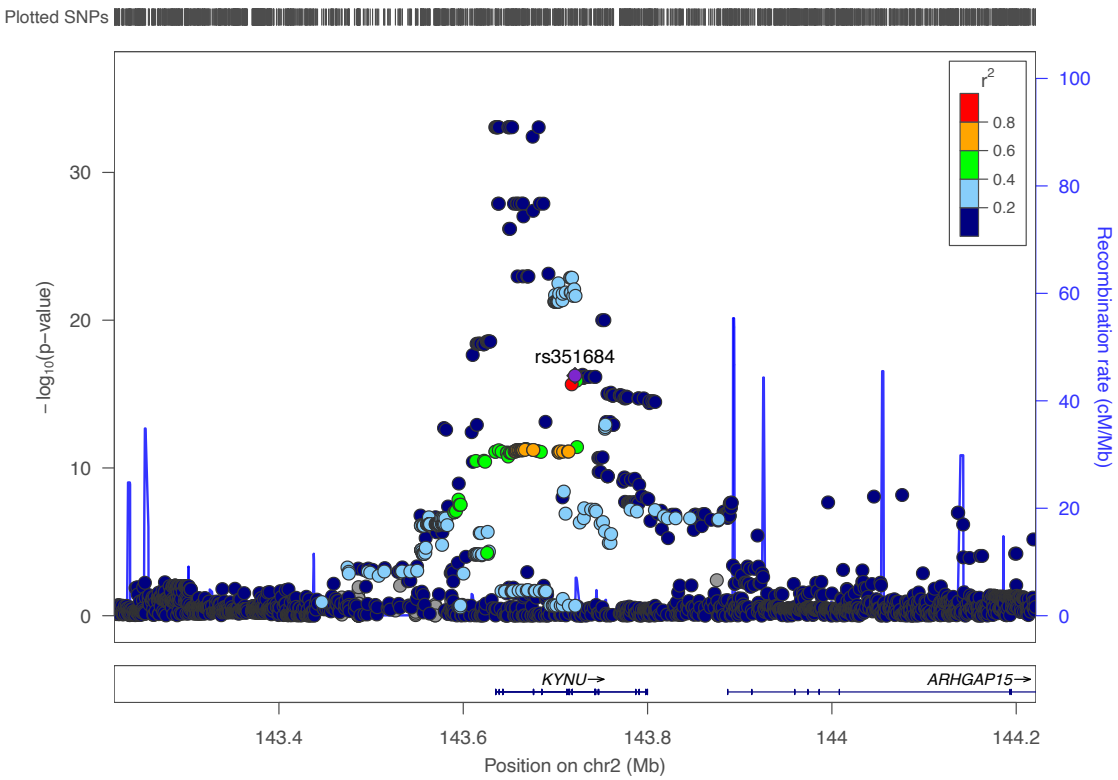

Plotted SNPs
rs111870199
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
LOC100132174
TNRC6C
SYNGR2
SOCS3
DNAH17
CYTH1
FLJ45079
TNRC6C−AS1
LOC100996291
PGS1
LOC101928710
TMC6
BIRC5
LOC101928674
TMC8
DNAH17−AS1
C17orf99
TK1
AFMID
TMEM235
75.8
76
76.2
76.4
76.6
Position on chr17 (Mb)

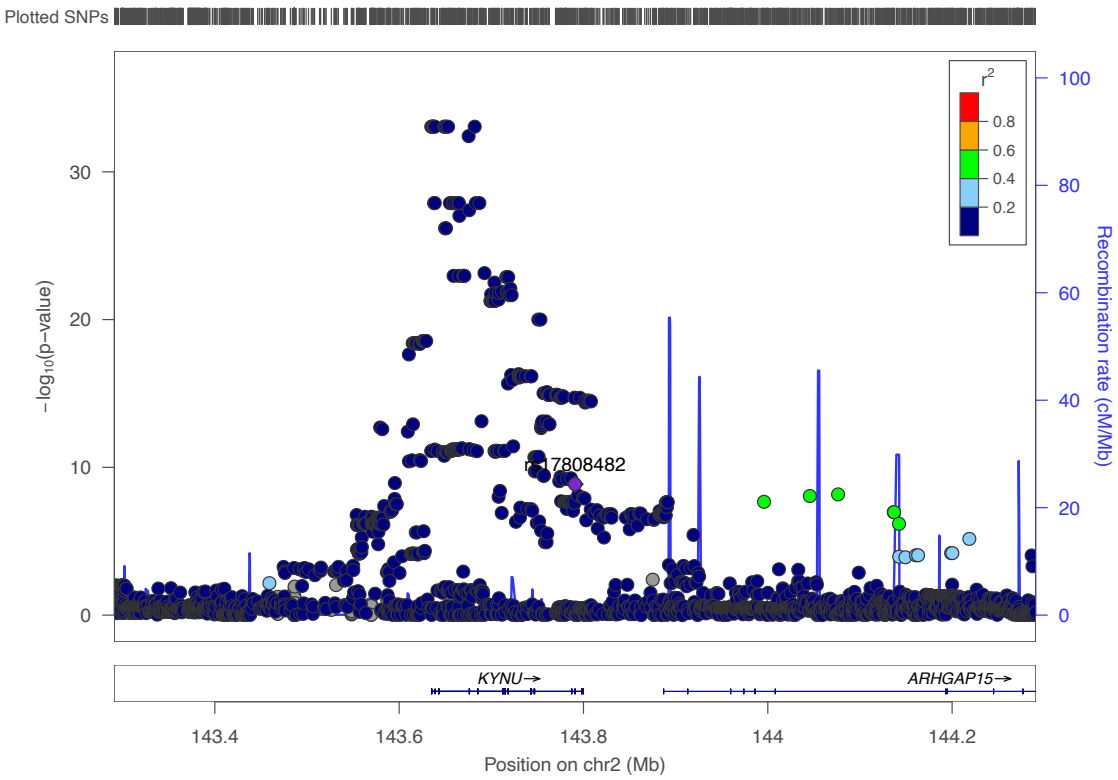

Plotted SNPs
rs61825638
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
RGS7
FH
OPN3
EXO1
MAP1LC3C
MIR3123
KMO
WDR64
BECN1P1
CHML
241.4
241.6
241.8
242
242.2
Position on chr1 (Mb)

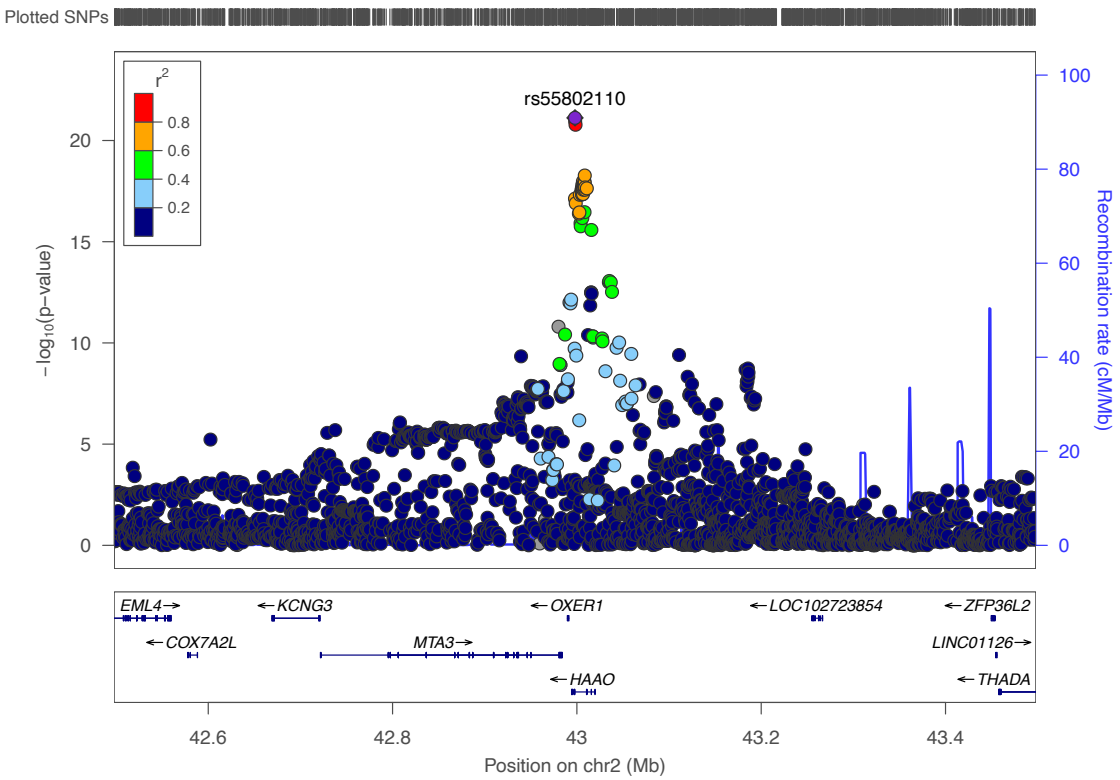

Plotted SNPs
rs6741488
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
KYNU
ARHGAP15
143.4
143.6
143.8
144
144.2
Position on chr2 (Mb)

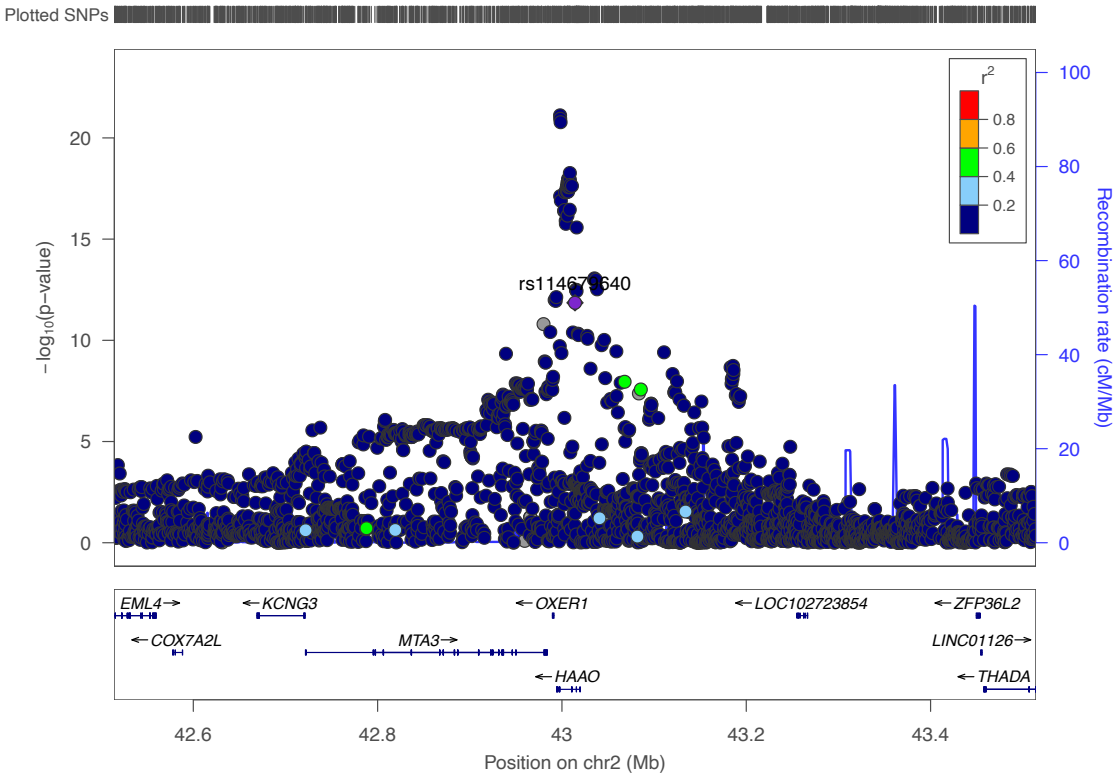

Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs75790387
30
20
10
0
100
80
60
40
20
0
− log10(p−value)
Recombination rate (cM/Mb)
KYNU→
ARHGAP15→
143.2
143.4
143.6
143.8
144
Position on chr2 (Mb)

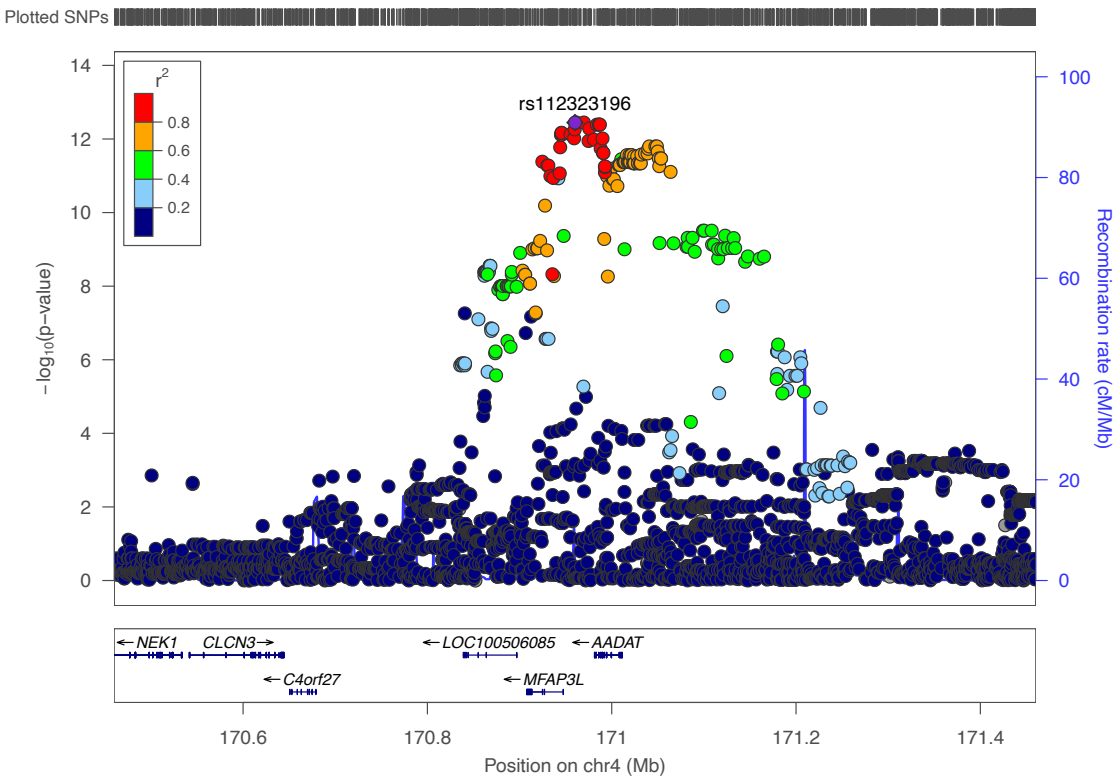

Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs34965832
30
20
10
0
100
80
60
40
20
0
− log10(p−value)
Recombination rate (cM/Mb)
KYNU→
ARHGAP15→
143.4
143.6
143.8
144
144.2
Position on chr2 (Mb)

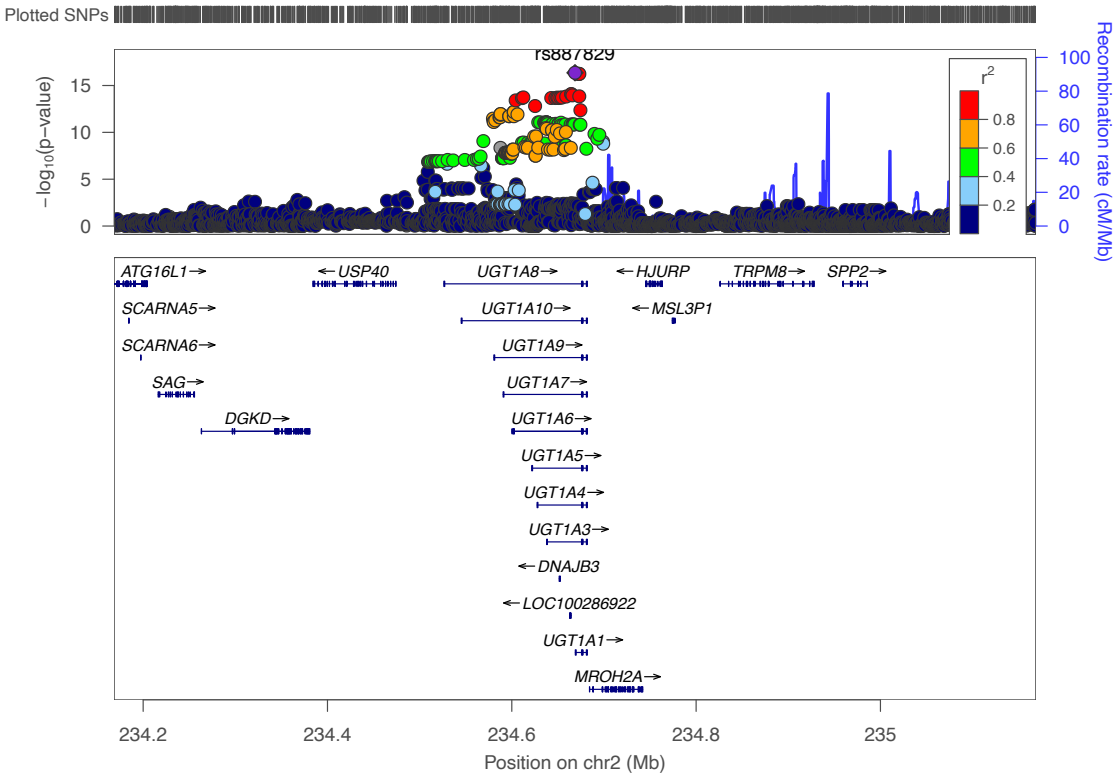

Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs351684
− log10(p−value)
Recombination rate (cM/Mb)
KYNU→
ARHGAP15→
143.4
143.6
143.8
144
144.2
Position on chr2 (Mb)

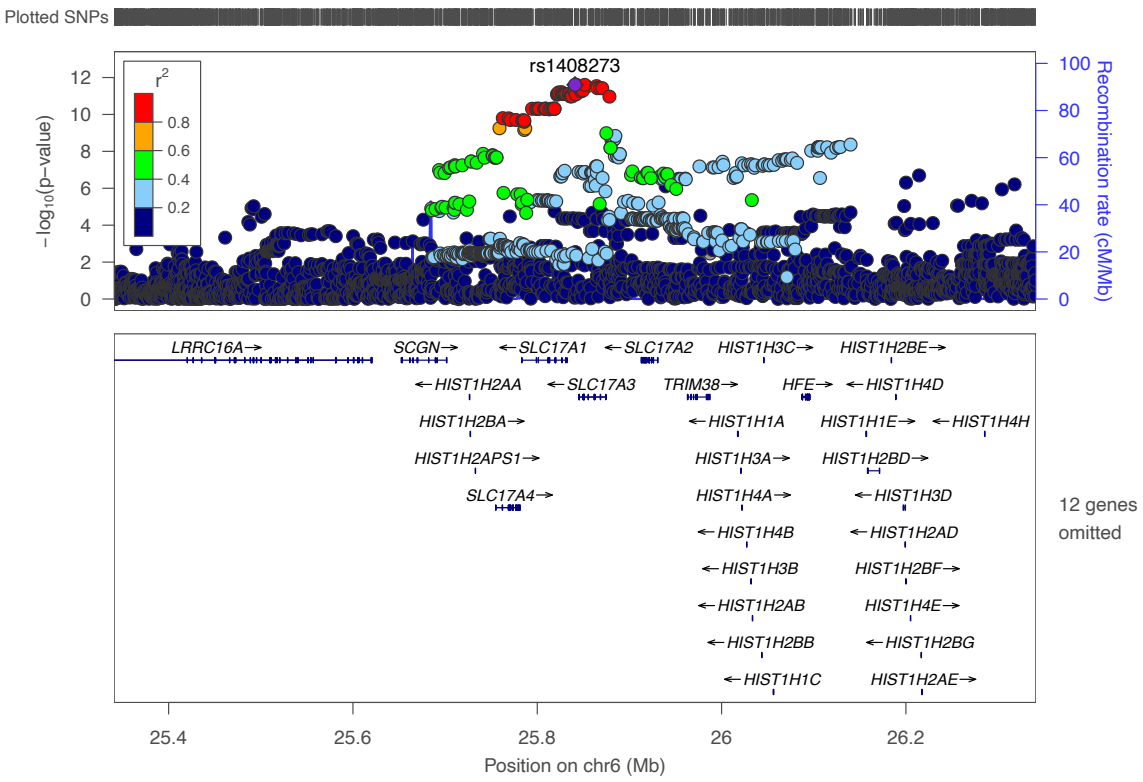

Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs17808482
− log10(p−value)
Recombination rate (cM/Mb)
KYNU→
ARHGAP15→
143.4
143.6
143.8
144
144.2
Position on chr2 (Mb)

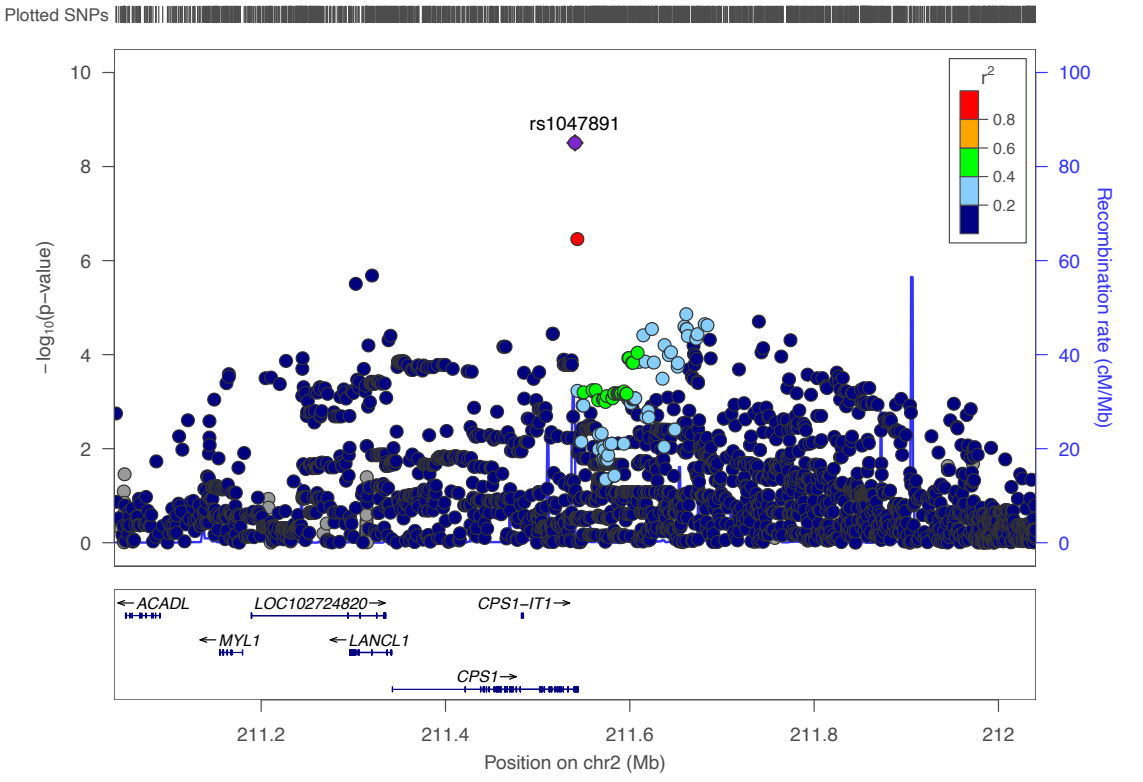

Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs55802110
−log10(p−value)
Recombination rate (cM/Mb)
EML4
KCNG3
OXER1
LOC102723854
ZFP36L2
COX7A2L
MTA3
LINC01126
HAAO
THADA
42.6
42.8
43
43.2
43.4
Position on chr2 (Mb)

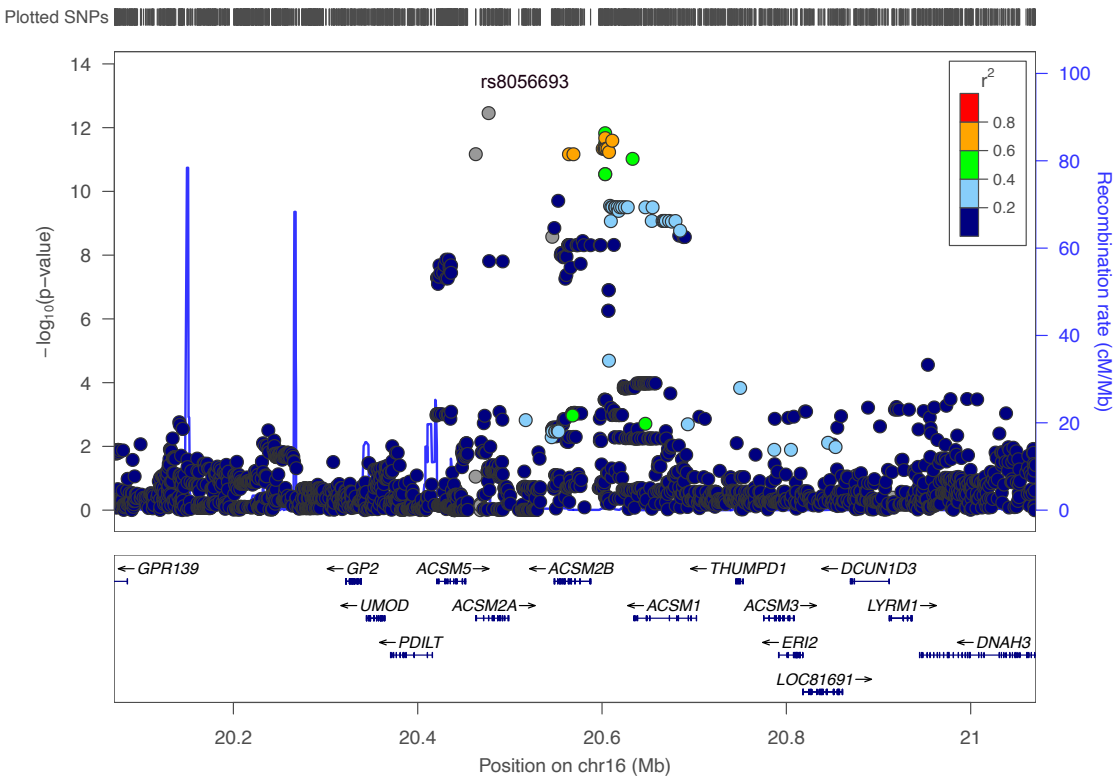

Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs114679640
−log10(p−value)
Recombination rate (cM/Mb)
EML4
KCNG3
OXER1
LOC102723854
ZFP36L2
COX7A2L
MTA3
LINC01126
HAAO
THADA
42.6
42.8
43
43.2
43.4
Position on chr2 (Mb)

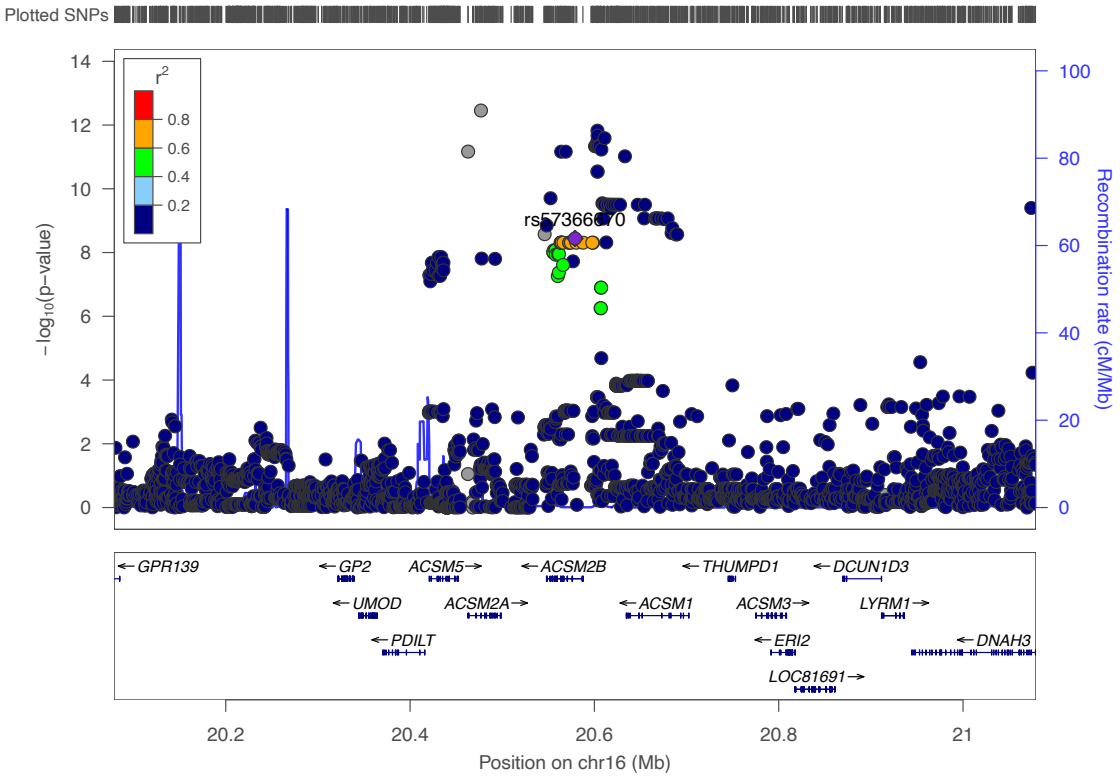

Plotted SNPs
rs112323196
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
NEK1
CLCN3
LOC100506085
AADAT
C4orf27
MFAP3L
170.6
170.8
171
171.2
171.4
Position on chr4 (Mb)

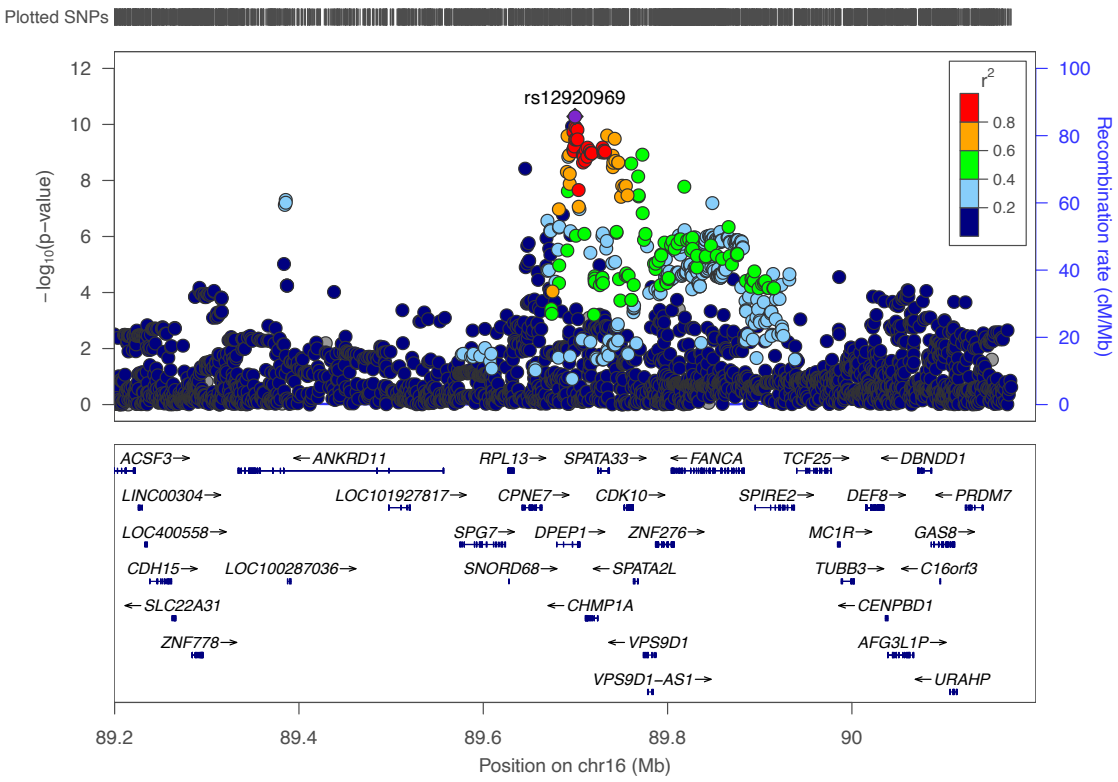

Plotted SNPs
rs887829
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
ATG16L1
USP40
UGT1A8
HJURP
TRPM8
SPP2
SCARNA5
UGT1A10
MSL3P1
SCARNA6
UGT1A9
SAG
UGT1A7
DGKD
UGT1A6
UGT1A5
UGT1A4
UGT1A3
DNAJB3
LOC100286922
UGT1A1
MROH2A
234.2
234.4
234.6
234.8
235
Position on chr2 (Mb)

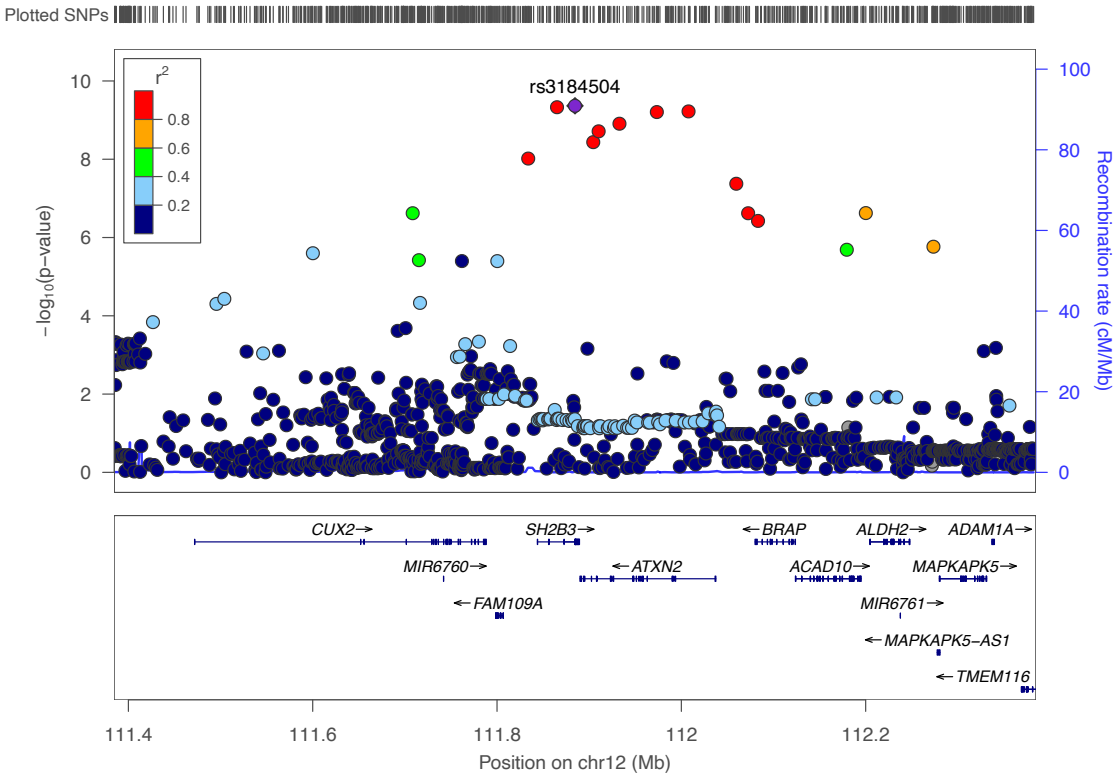

Plotted SNPs
rs1408273
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
LRRC16A→
SCGN→
←SLC17A1
←SLC17A2
HIST1H3C→
HIST1H2BE→
←HIST1H2AA
←SLC17A3
TRIM38→
HFE→
←HIST1H4D
HIST1H2BA→
←HIST1H1A
HIST1H1E→
←HIST1H4H
HIST1H2APS1→
HIST1H3A→
HIST1H2BD→
SLC17A4→
HIST1H4A→
←HIST1H3D
←HIST1H4B
←HIST1H2AD
←HIST1H3B
HIST1H2BF→
←HIST1H2AB
HIST1H4E→
←HIST1H2BB
←HIST1H2BG
←HIST1H1C
HIST1H2AE→
12 genes omitted
25.4
25.6
25.8
26
26.2
Position on chr6 (Mb)


Plotted SNPs
rs1047891
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
←ACADL
LOC102724820→
CPS1−IT1→
←MYL1
←LANCL1
CPS1→
211.2
211.4
211.6
211.8
212
Position on chr2 (Mb)


Plotted SNPs
rs8056693
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
GPR139
GP2
ACSM5
ACSM2B
THUMPD1
DCUN1D3
UMOD
ACSM2A
ACSM1
ACSM3
LYRM1
PDILT
ERI2
DNAH3
LOC81691
Position on chr16 (Mb)


Plotted SNPs
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
GPR139
GP2
ACSM5
ACSM2B
THUMPD1
DCUN1D3
UMOD
ACSM2A
ACSM1
ACSM3
LYRM1
PDILT
ERI2
DNAH3
LOC81691
Position on chr16 (Mb)


Plotted SNPs
rs12920969
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
ACSF3→ ←ANKRD11 RPL13→ SPATA33→ ←FANCA TCF25→ ←DBNDD1
LINC00304→ LOC101927817→ CPNE7→ CDK10→ SPIRE2→ DEF8→ ←PRDM7
LOC400558→ SPG7→ DPEP1→ ZNF276→ MC1R→ GAS8→
CDH15→ LOC100287036→ SNORD68→ ←SPATA2L TUBB3→ ←C16orf3
←SLC22A31 ←CHMP1A ←CENPBD1
ZNF778→ ←VPS9D1 AFG3L1P→
VPS9D1−AS1→ ←URAHP
89.2
89.4
89.6
89.8
90
Position on chr16 (Mb)


Plotted SNPs
rs3184504
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
CUX2→ SH2B3→ ←BRAP ALDH2→ ADAM1A→
MIR6760→ ←ATXN2 ACAD10→ MAPKAPK5→
←FAM109A MIR6761→
←MAPKAPK5−AS1
←TMEM116
111.4
111.6
111.8
112
112.2
Position on chr12 (Mb)

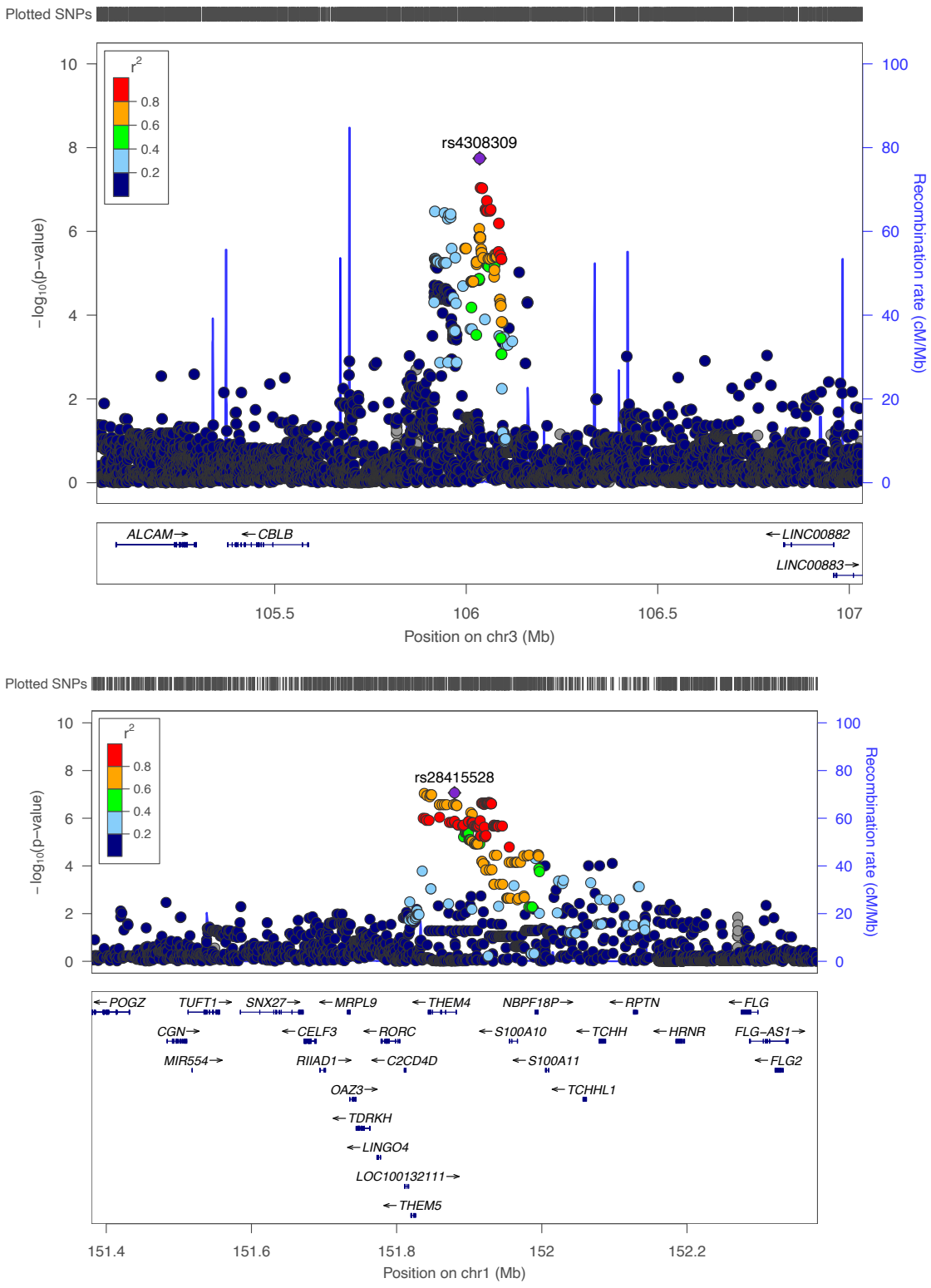


Figure S2: *Locus Zoom plots of all the 23 GWA signals (mQTL1-23) corresponding to the lead SNP-trait pairs in Table 1 and also THEM4 signal of Pic GWAS (near-GWA signal).*

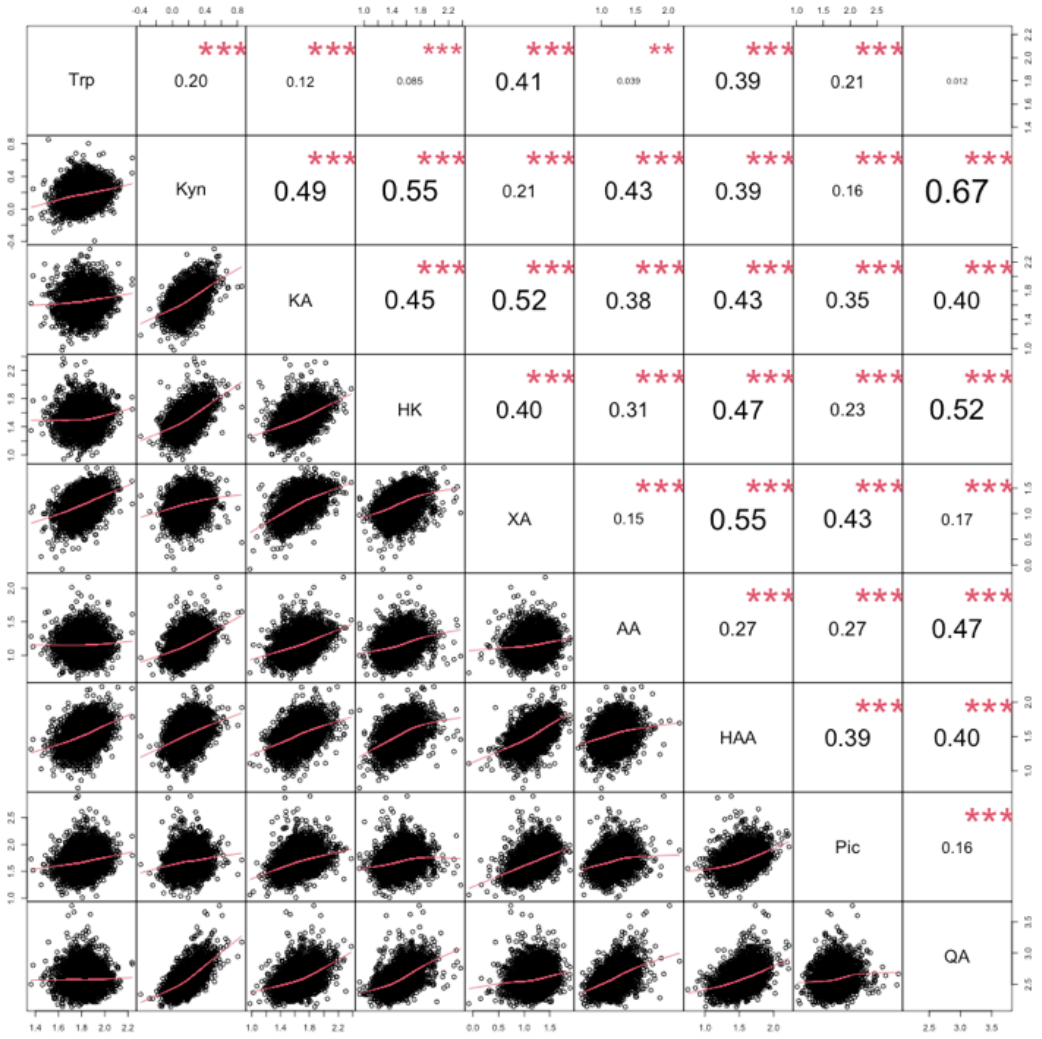


Fig. S3. *Pairwise Spearman correlation of the log transformed values of the nine studied compounds. Name of each metabolite is shown on the diagonal. The bottom of the diagonal shows the bivariate scatter plots with a fitted line in red. The top of the diagonal depicts the*

*correlation coefficients and the significance levels. Significance levels (P-values) shown as red stars (*** 0.001, ** 0.01, * 0.05, “ ” 1)*

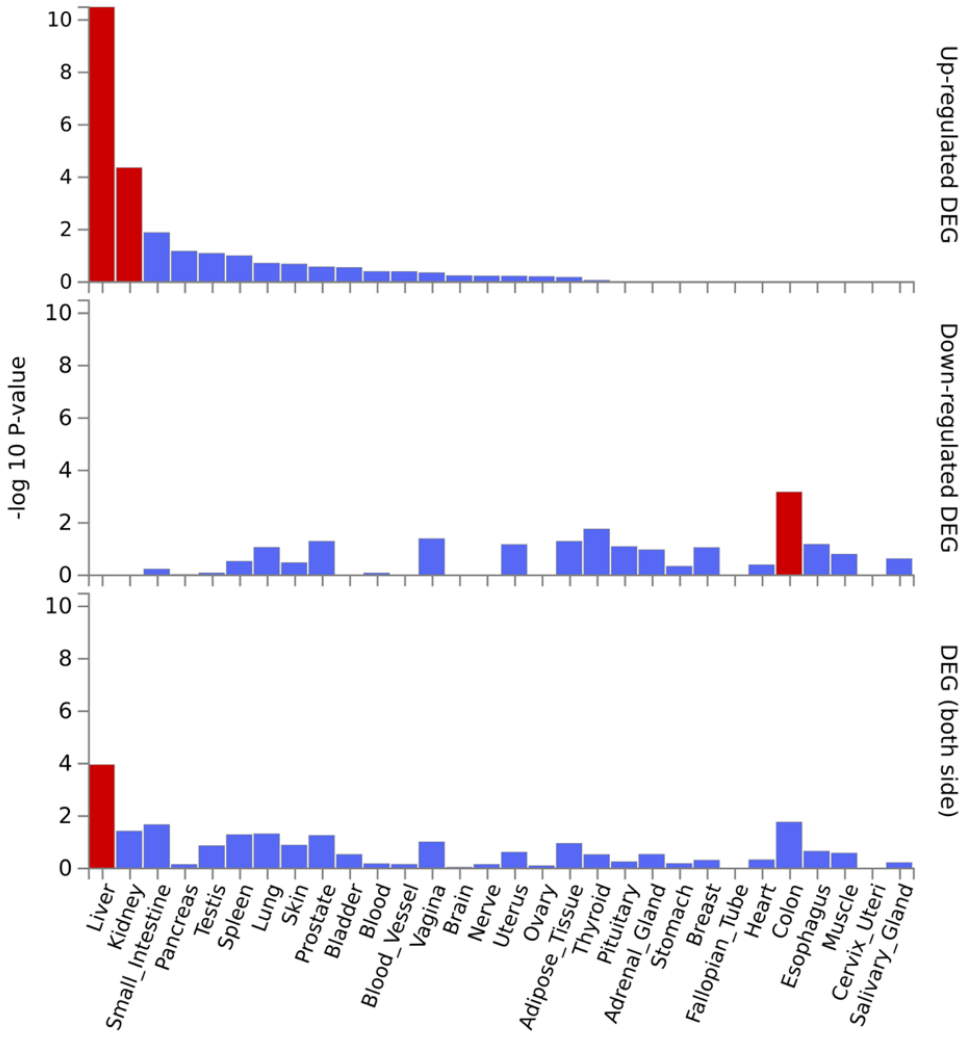


Fig. S4. *Upregulation and downregulation patterns of differentially expressed genes (DEG) in GTEX v7. These DEG are significantly upregulated in liver and kidney and down regulated in colon.*

Separated files:

Tables S1 to S7 (.xlsx) Characteristics of the study participants, GWAS summary statistics of top hits and Molecular/Trait annotation for all GWAS signals.

SI References

1. C. A. Anderson, *et al.*, Data quality control in genetic case-control association studies. *Nat. Protoc.* **5**, 1564–1573 (2010).
2. G. Abraham, Y. Qiu, M. Inouye, FlashPCA2: principal component analysis of Biobank-scale genotype datasets. *Bioinformatics* **33**, 2776–2778 (2017).
3. A. Manichaikul, *et al.*, Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).