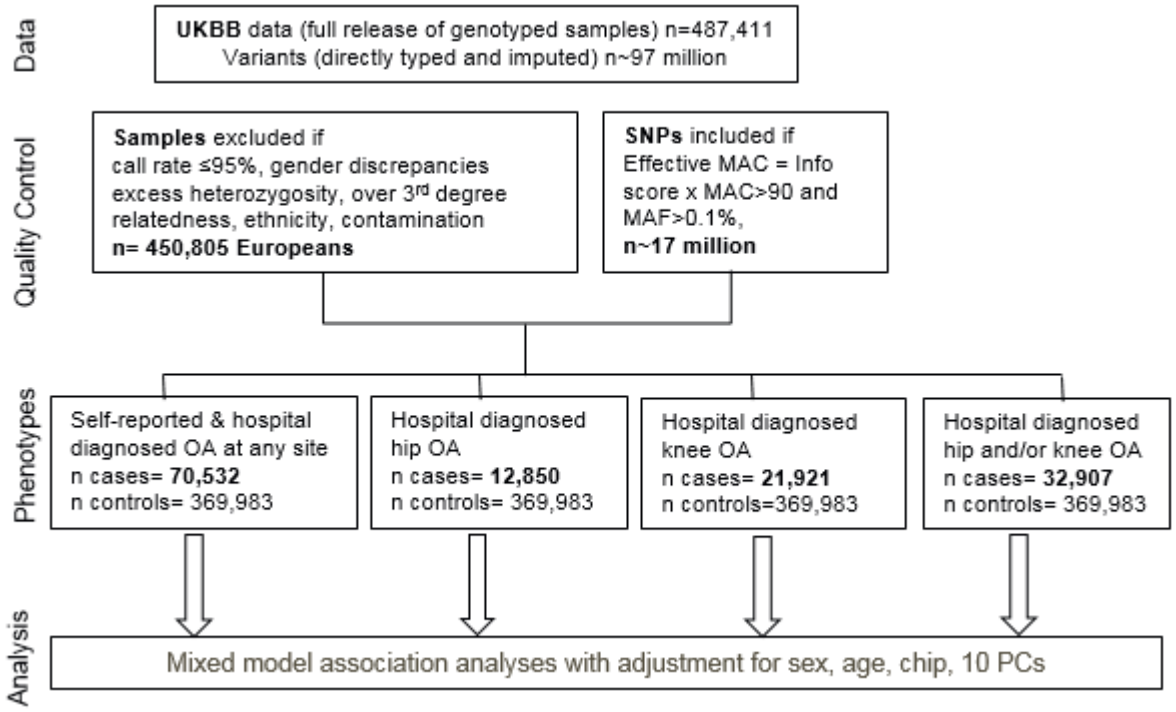
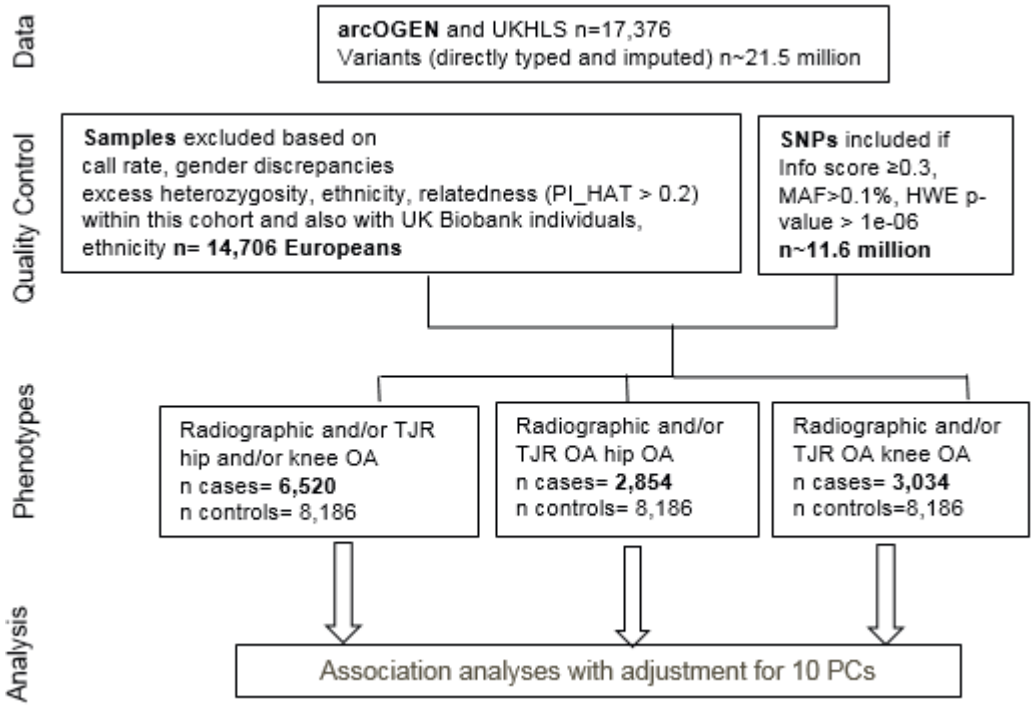
Supplementary Figure 1: Flowchart depicting the two GWAS and their meta-analysis. In UKBB (A), we performed sample and variant-based QC to produce a set of approximately 450K related Europeans and 17M variants. We derived four osteoarthritis definitions from self-report and hospital records. We analysed the data adjusting for relatedness. In arcOGEN (B), we imputed the data into HRC panel and performed sample and variant-based QC to produce a set of approximately 14K unrelated Europeans and 11M variants. We derived three osteoarthritis definitions from total joint replacement records or radiographic evidence of disease. Subsequently, we performed a genome-wide fixed-effects meta-analysis (C) across the two cohorts, identifying 64 unique independent signals across the four osteoarthritis definitions.

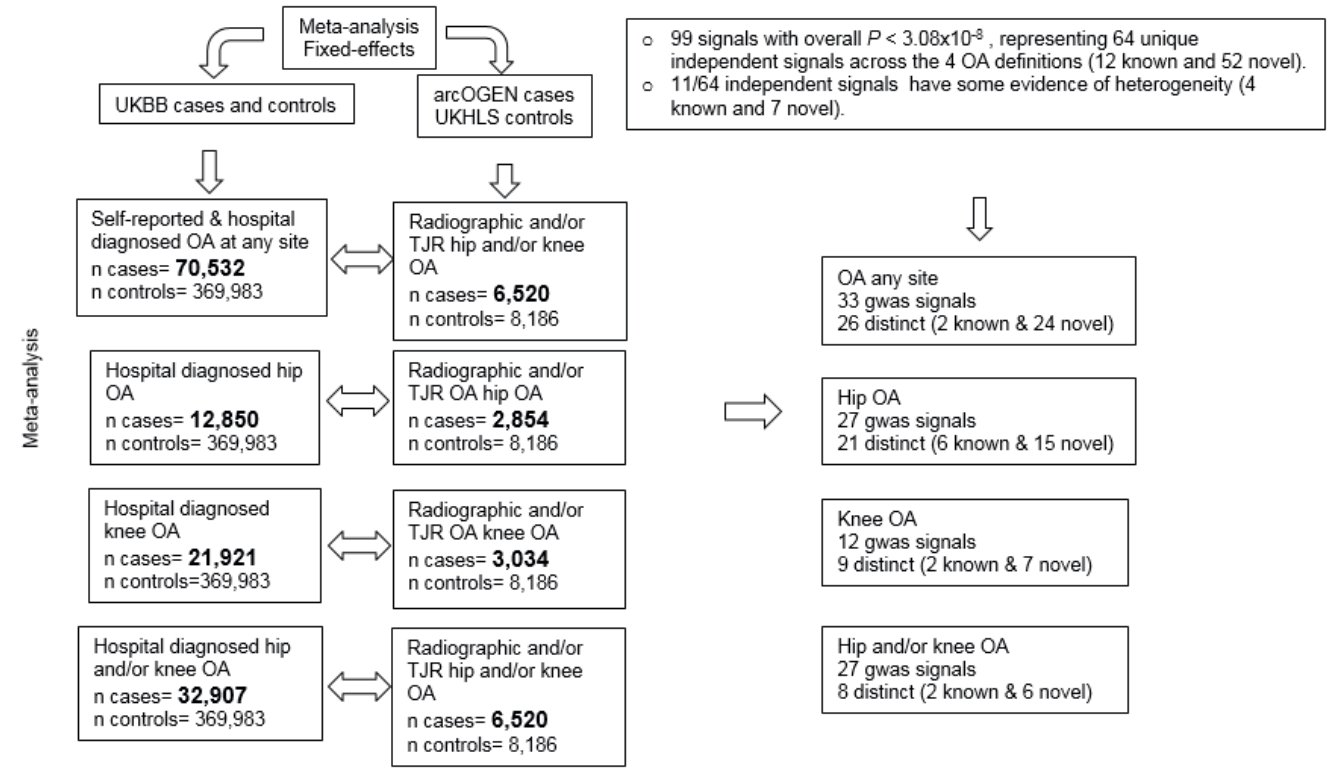
A



B

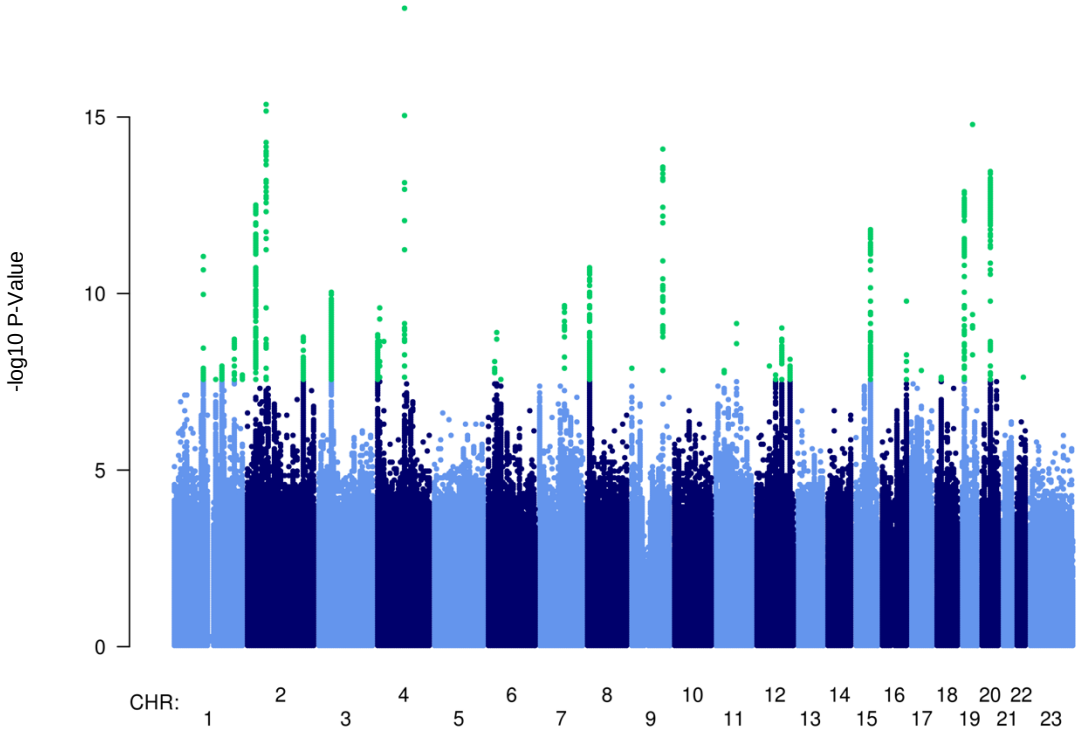

Data
Quality Control
Phenotypes
Analysis
arcOGEN and UKHLS n=17,376
Variants (directly typed and imputed) n~21.5 million
Samples excluded based on
call rate, gender discrepancies
excess heterozygosity, ethnicity, relatedness (PI_HAT > 0.2)
within this cohort and also with UK Biobank individuals,
ethnicity n= 14,706 Europeans
SNPs included if
Info score ≥0.3,
MAF>0.1%, HWE p-value > 1e-06
n~11.6 million
Radiographic and/or TJR hip and/or knee OA
n cases= 6,520
n controls= 8,186
Radiographic and/or TJR OA hip OA
n cases= 2,854
n controls= 8,186
Radiographic and/or TJR OA knee OA
n cases= 3,034
n controls=8,186
Association analyses with adjustment for 10 PCs

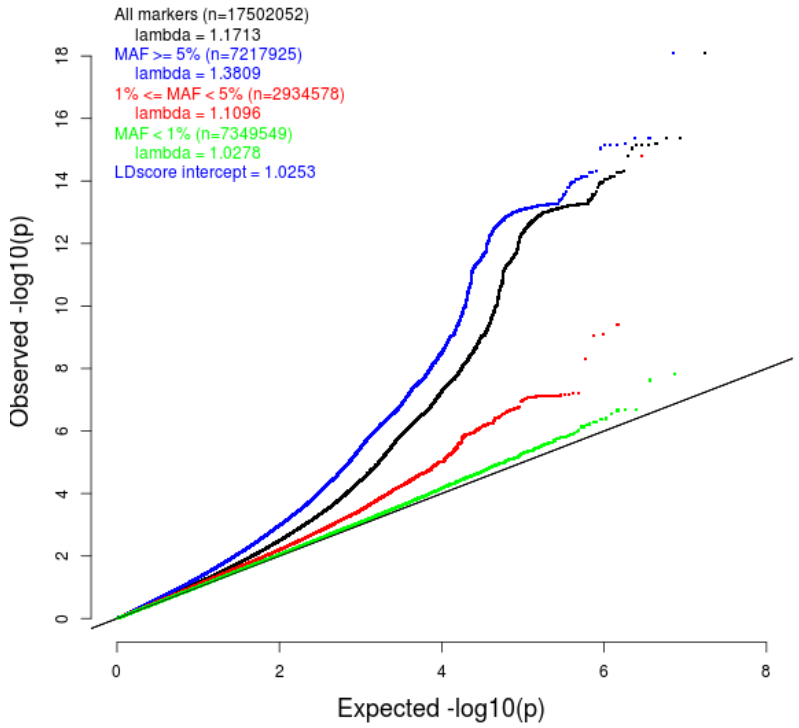
C


Meta-analysis
Fixed-effects
UKBB cases and controls
arcOGEN cases
UKHLS controls
99 signals with overall P < 3.08x10-8 , representing 64 unique independent signals across the 4 OA definitions (12 known and 52 novel).
11/64 independent signals have some evidence of heterogeneity (4 known and 7 novel).
Meta-analysis
Self-reported & hospital diagnosed OA at any site
n cases= 70,532
n controls= 369,983
Radiographic and/or TJR hip and/or knee OA
n cases= 6,520
n controls= 8,186
Hospital diagnosed hip OA
n cases= 12,850
n controls= 369,983
Radiographic and/or TJR OA hip OA
n cases= 2,854
n controls= 8,186
Hospital diagnosed knee OA
n cases= 21,921
n controls=369,983
Radiographic and/or TJR OA knee OA
n cases= 3,034
n controls= 8,186
Hospital diagnosed hip and/or knee OA
n cases= 32,907
n controls= 369,983
Radiographic and/or TJR hip and/or knee OA
n cases= 6,520
n controls= 8,186
OA any site
33 gwas signals
26 distinct (2 known & 24 novel)
Hip OA
27 gwas signals
21 distinct (6 known & 15 novel)
Knee OA
12 gwas signals
9 distinct (2 known & 7 novel)
Hip and/or knee OA
27 gwas signals
8 distinct (2 known & 6 novel)

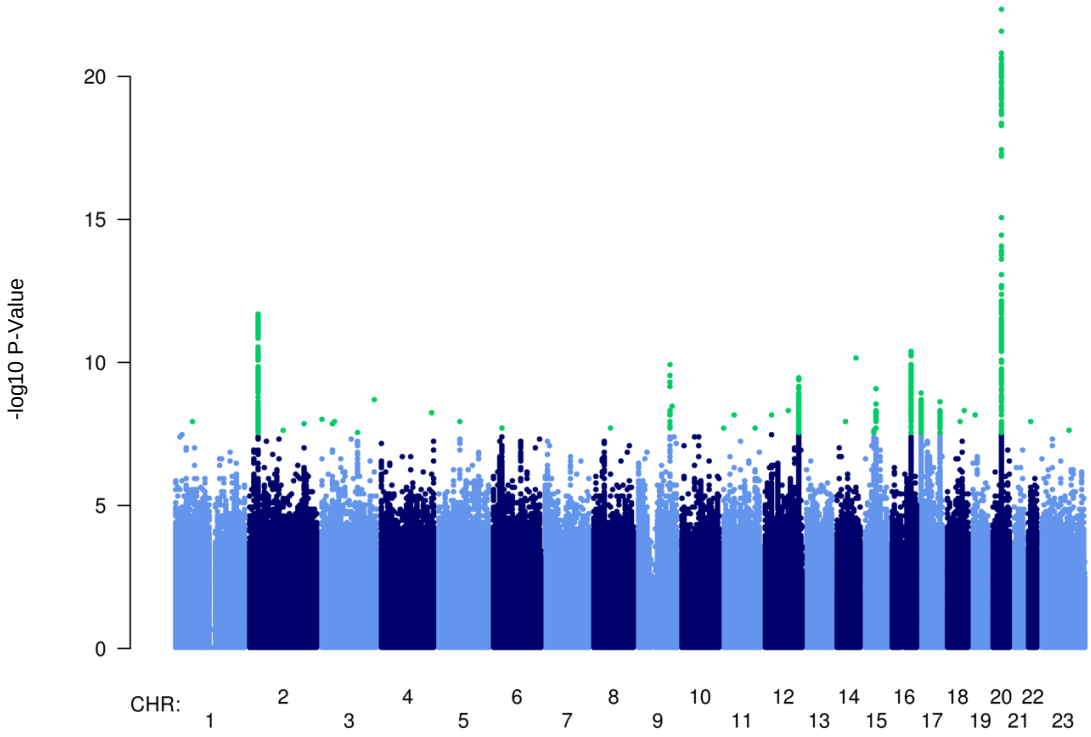
Supplementary Figure 2: Summary plots of summary statistics in UK Biobank and arcOGEN meta-analysis. Manhattan and quantile-quantile plots for osteoarthritis (A), knee osteoarthritis (B), hip osteoarthritis (C), knee and/or hip osteoarthritis (D) in UK Biobank and arcOGEN meta-analysis.

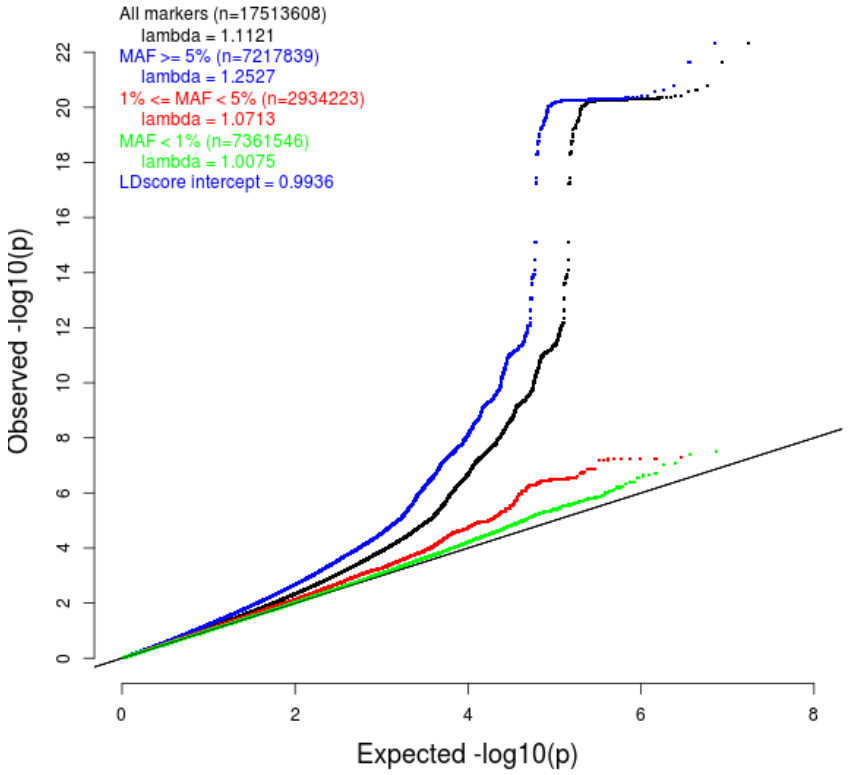
A



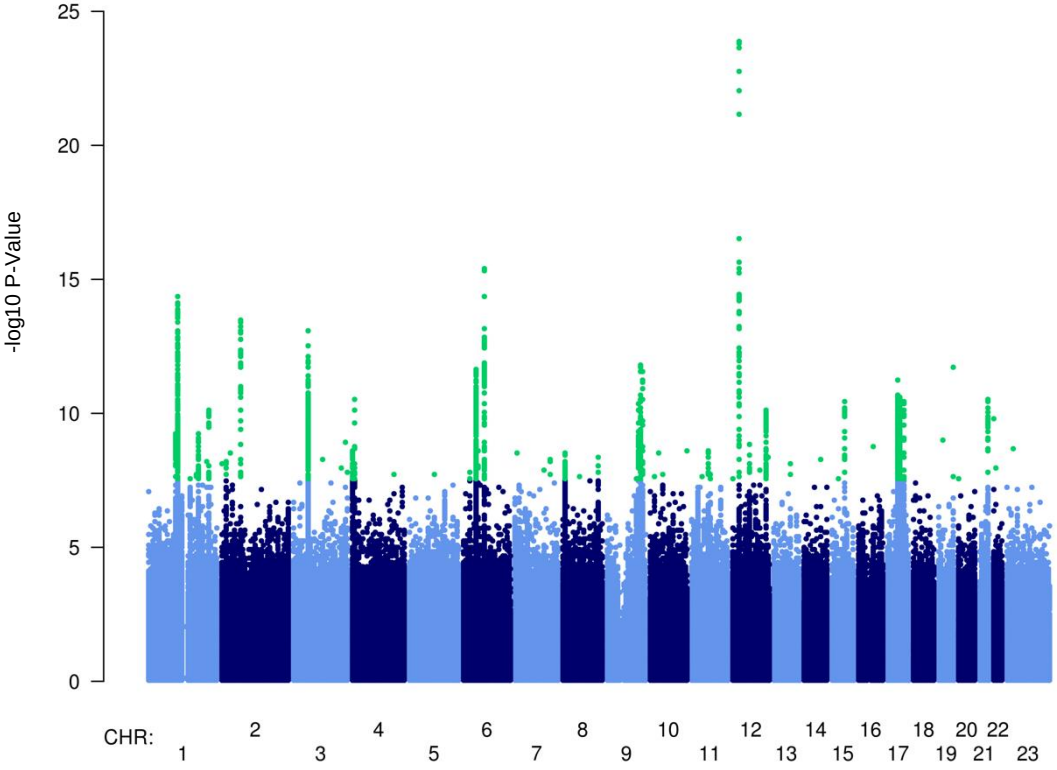


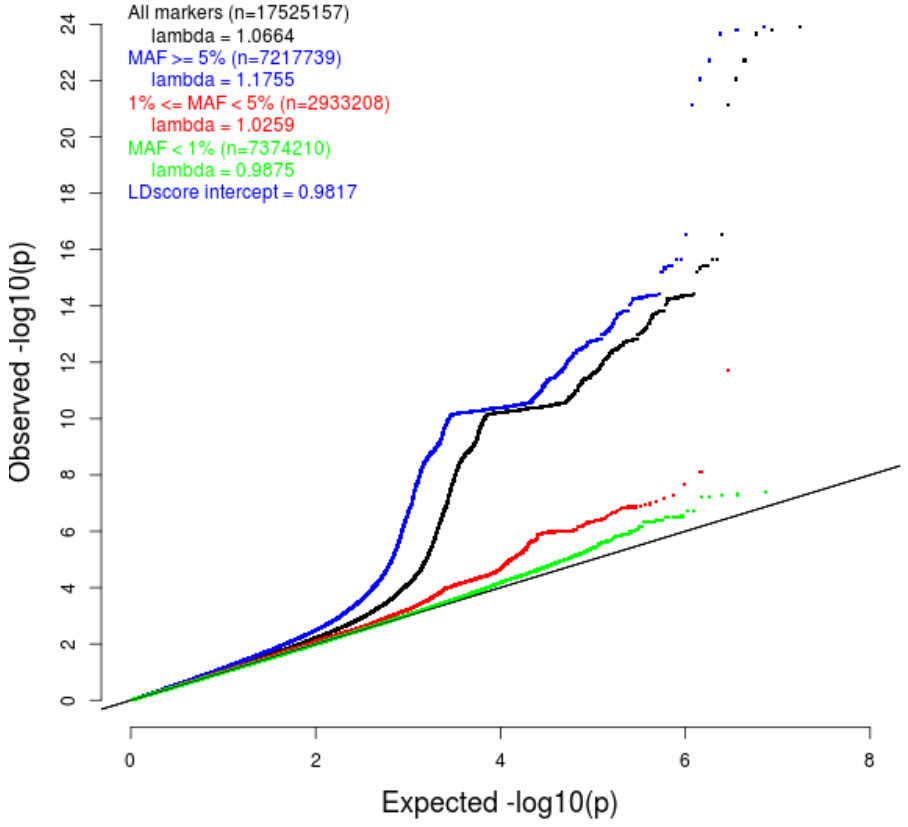
B


-log10 P-Value
20
15
10
5
0
CHR:
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

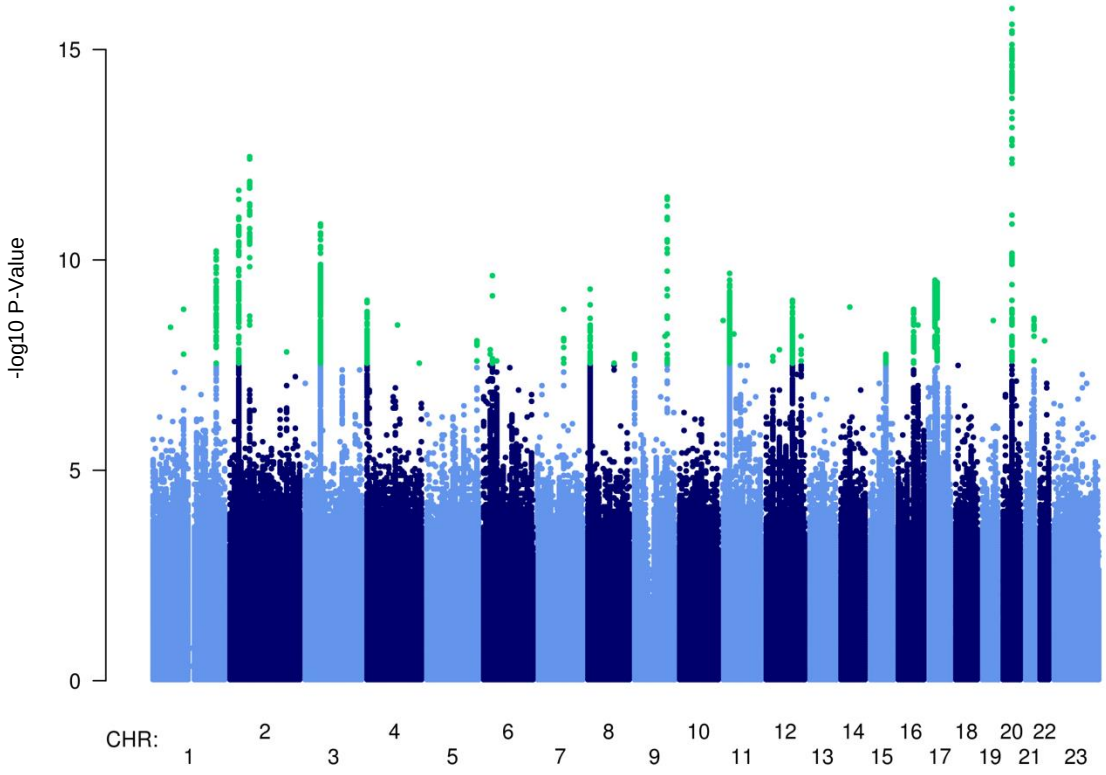

All markers (n=17513608)
lambda = 1.1121
MAF >= 5% (n=7217839)
lambda = 1.2527
1% <= MAF < 5% (n=2934223)
lambda = 1.0713
MAF < 1% (n=7361546)
lambda = 1.0075
LDscore intercept = 0.9936
Observed -log10(p)
Expected -log10(p)
0
2
4
6
8
10
12
14
16
18
20
22

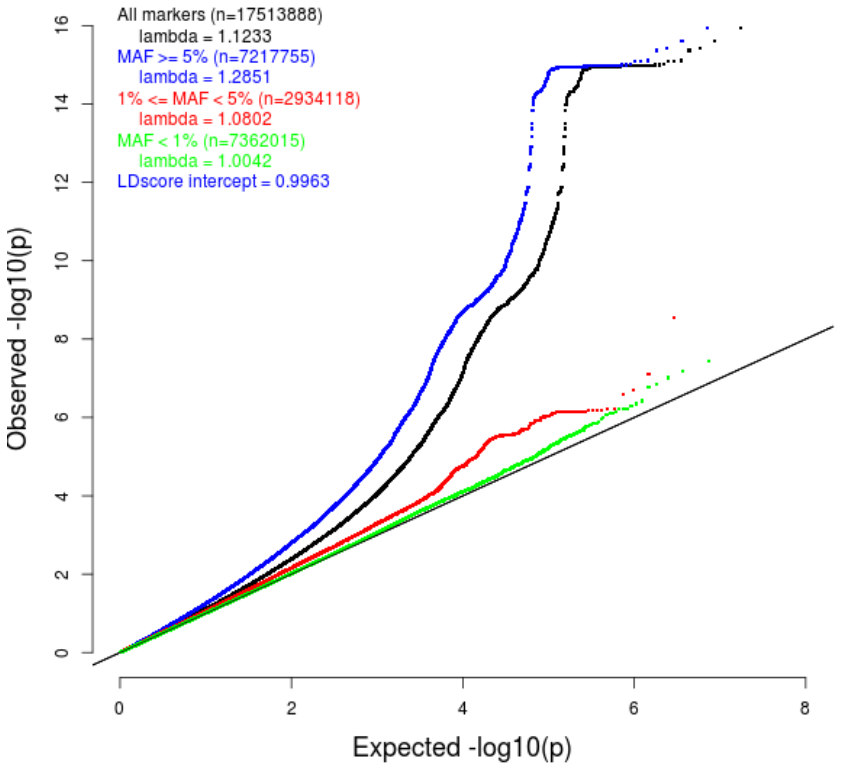
C


-log10 P-Value
25
20
15
10
5
0
CHR:
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

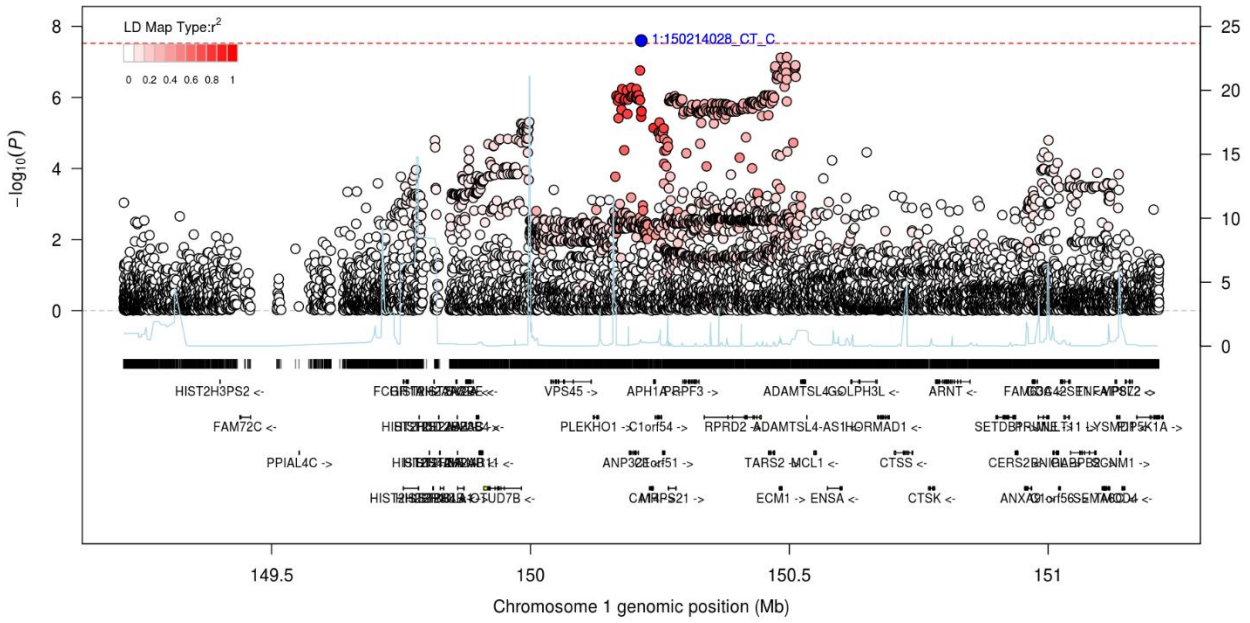

All markers (n=17525157)
lambda = 1.0664
MAF >= 5% (n=7217739)
lambda = 1.1755
1% <= MAF < 5% (n=2933208)
lambda = 1.0259
MAF < 1% (n=7374210)
lambda = 0.9875
LDscore intercept = 0.9817
Observed -log10(p)
Expected -log10(p)
0
2
4
6
8
10
12
14
16
18
20
22
24

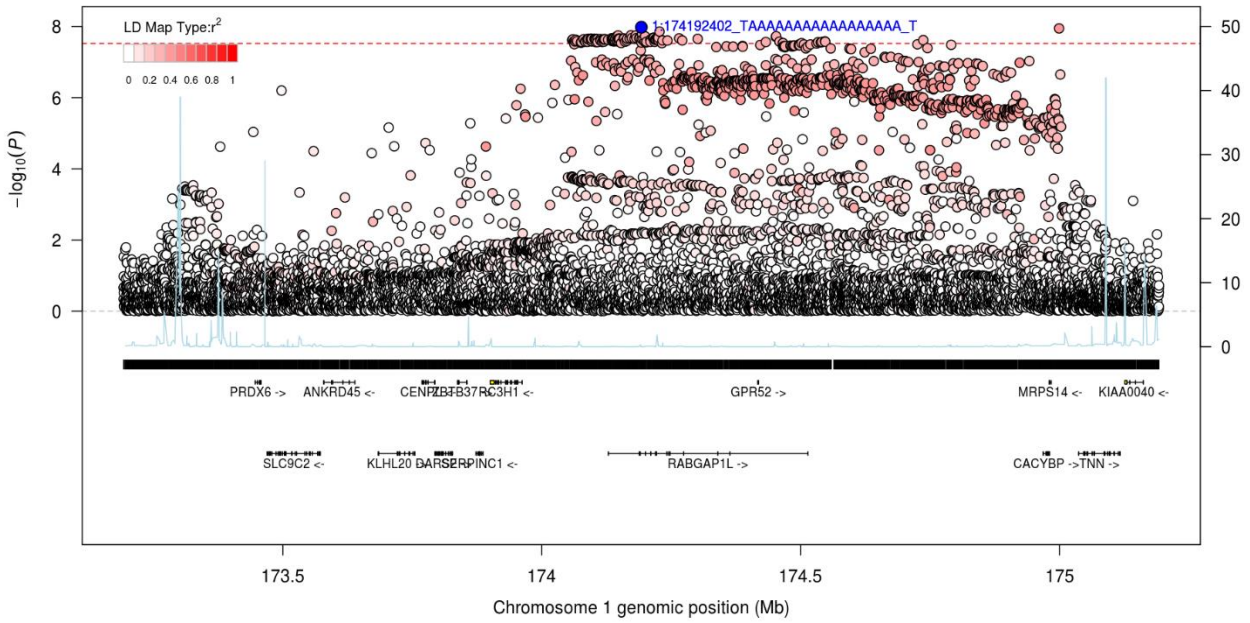
D

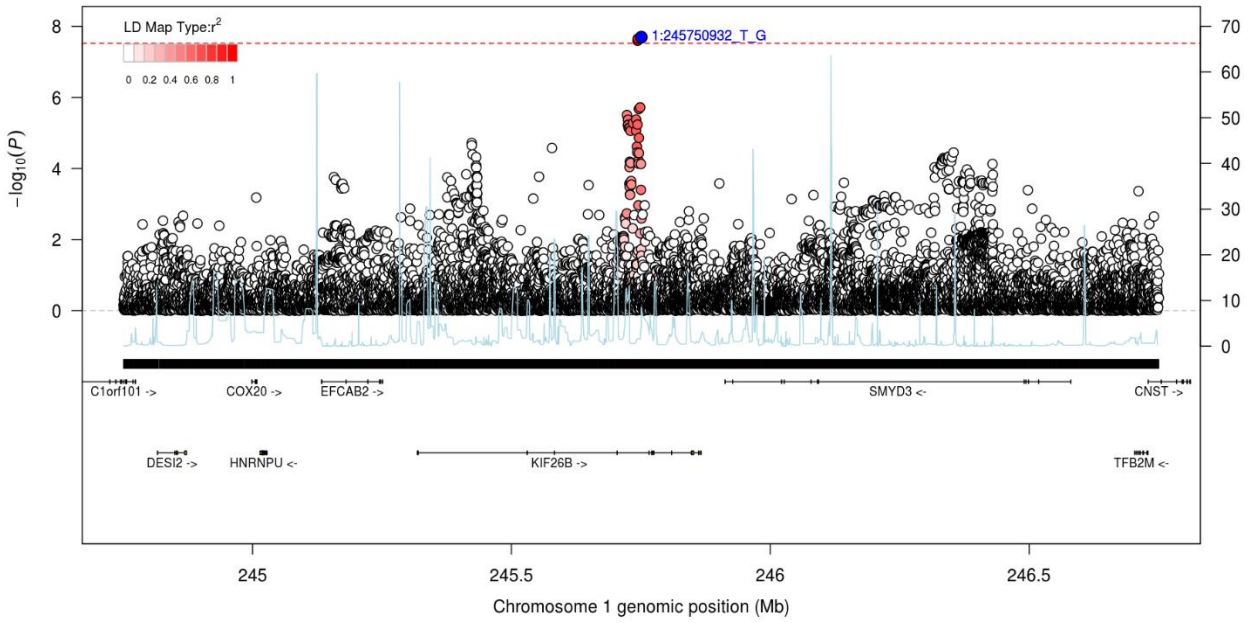

-log10 P-Value
15
10
5
0
CHR:
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

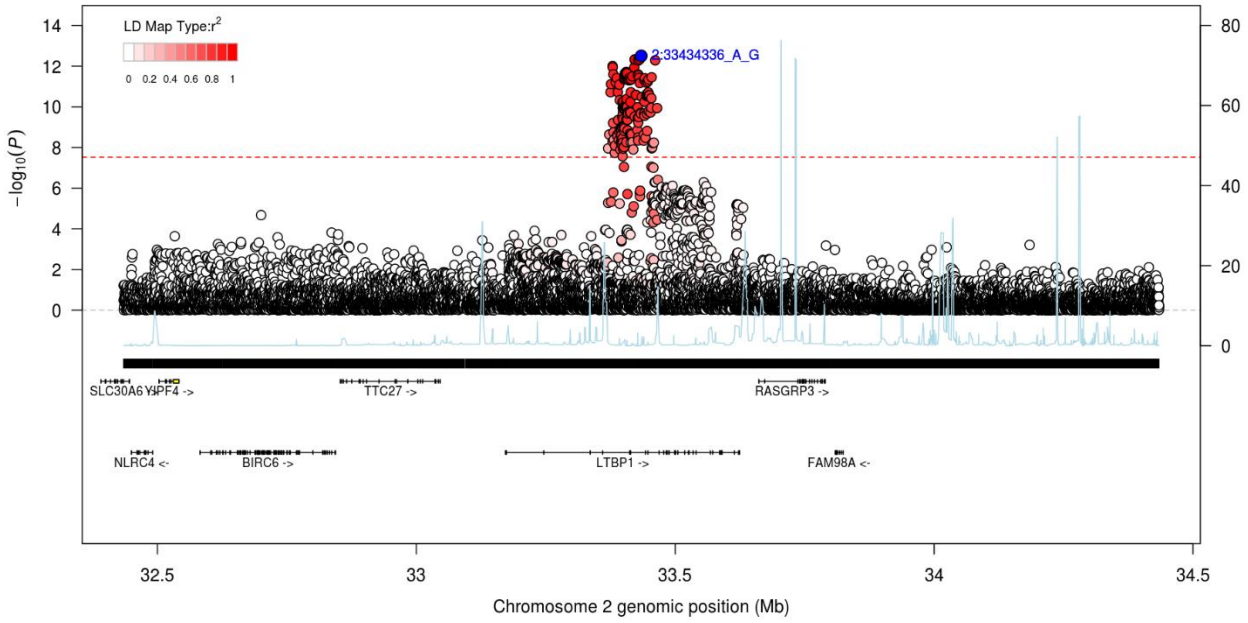

All markers (n=17513888)
lambda = 1.1233
MAF >= 5% (n=7217755)
lambda = 1.2851
1% <= MAF < 5% (n=2934118)
lambda = 1.0802
MAF < 1% (n=7362015)
lambda = 1.0042
LDscore intercept = 0.9963
Observed -log10(p)
Expected -log10(p)
0
2
4
6
8
10
12
14
16

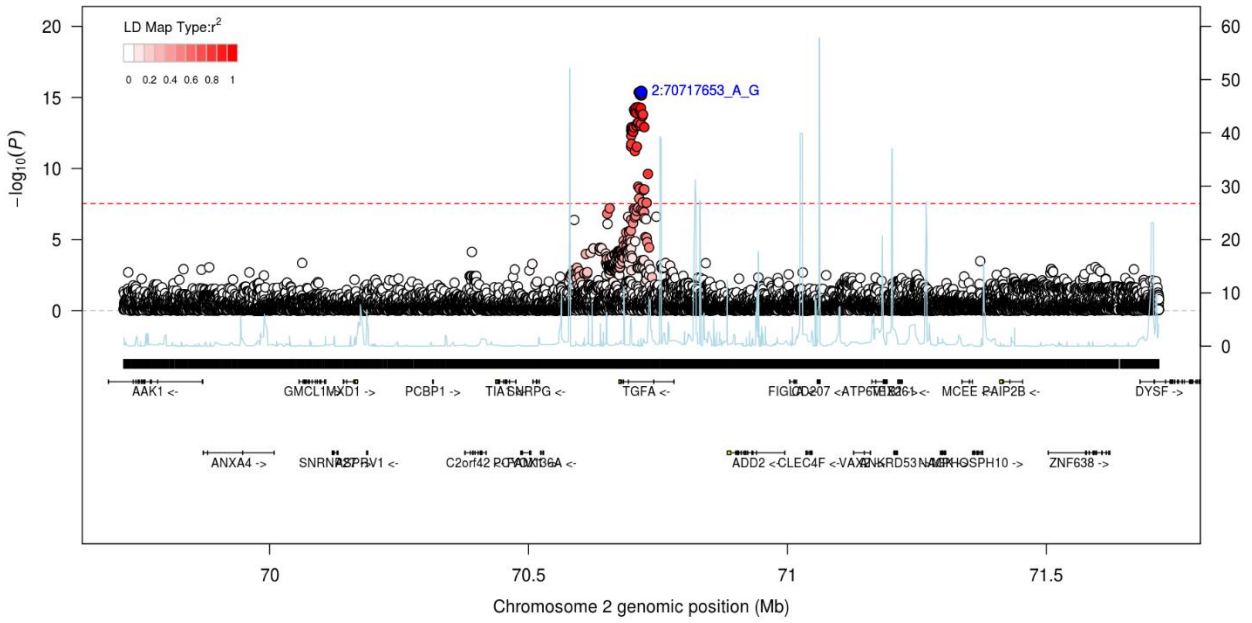
Supplementary Figure 3: Locus zoom plots for 26 osteoarthritis loci that reached genome-wide significance in the meta-analysis (as reported in Table 1 and Supplementary Table 3), displaying 1Mb each side of the top variant, where the blue circle represents the index variant on the meta-analysis across UK Biobank and arcOGEN. LD is calculated from UK Biobank.

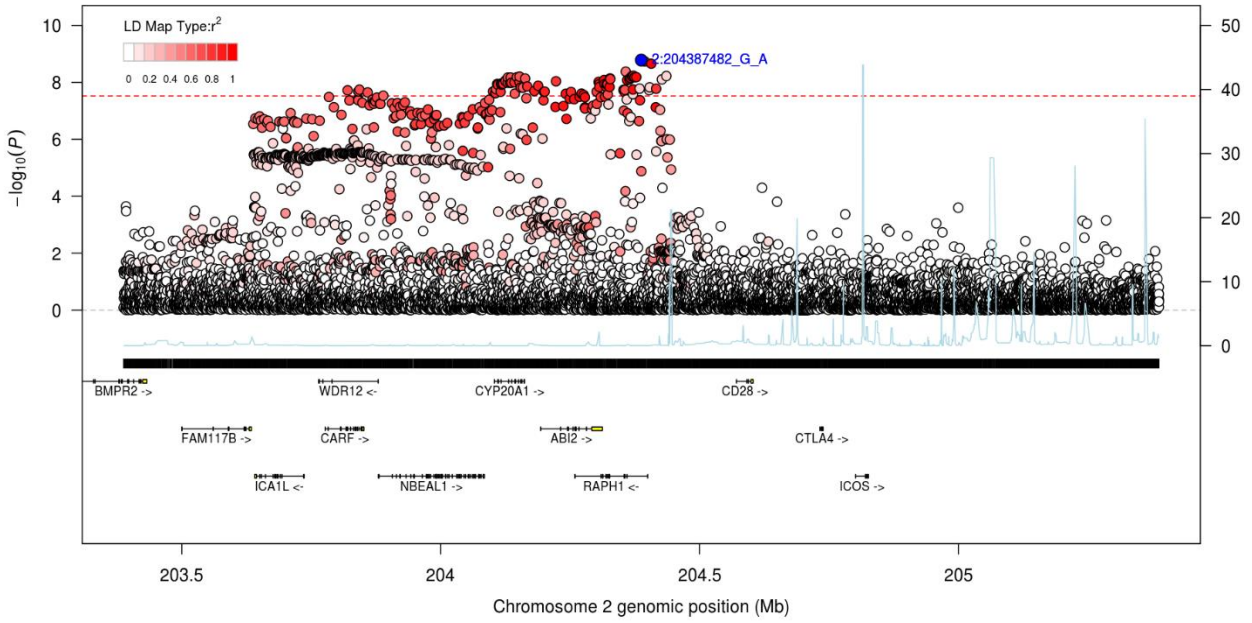


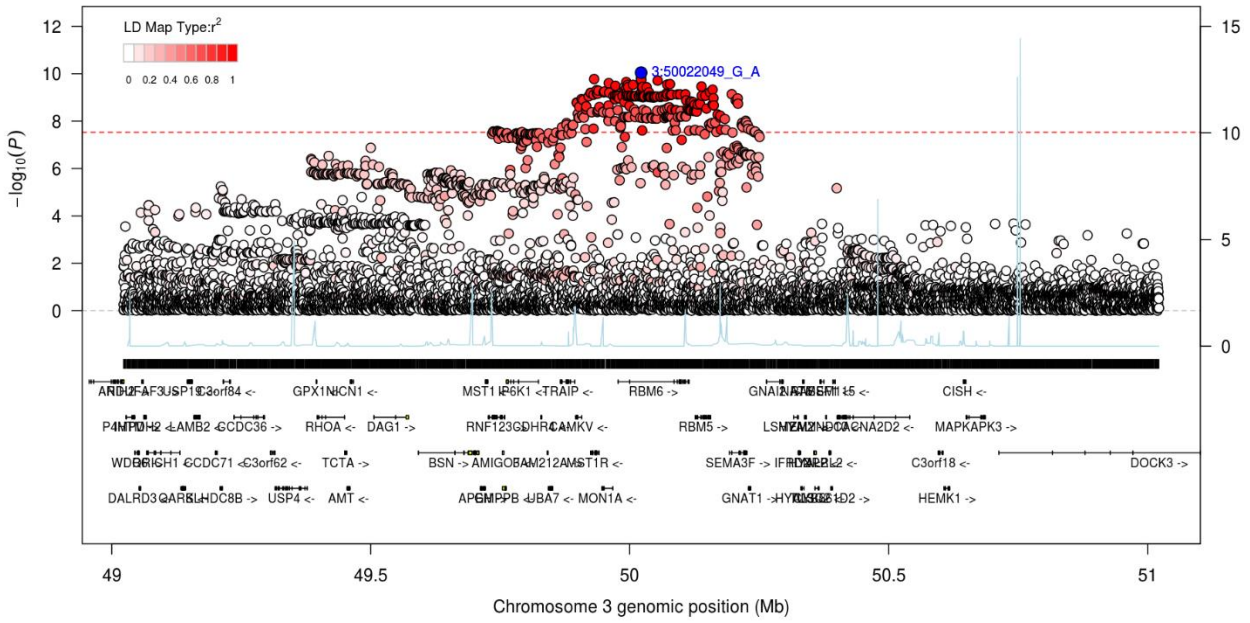


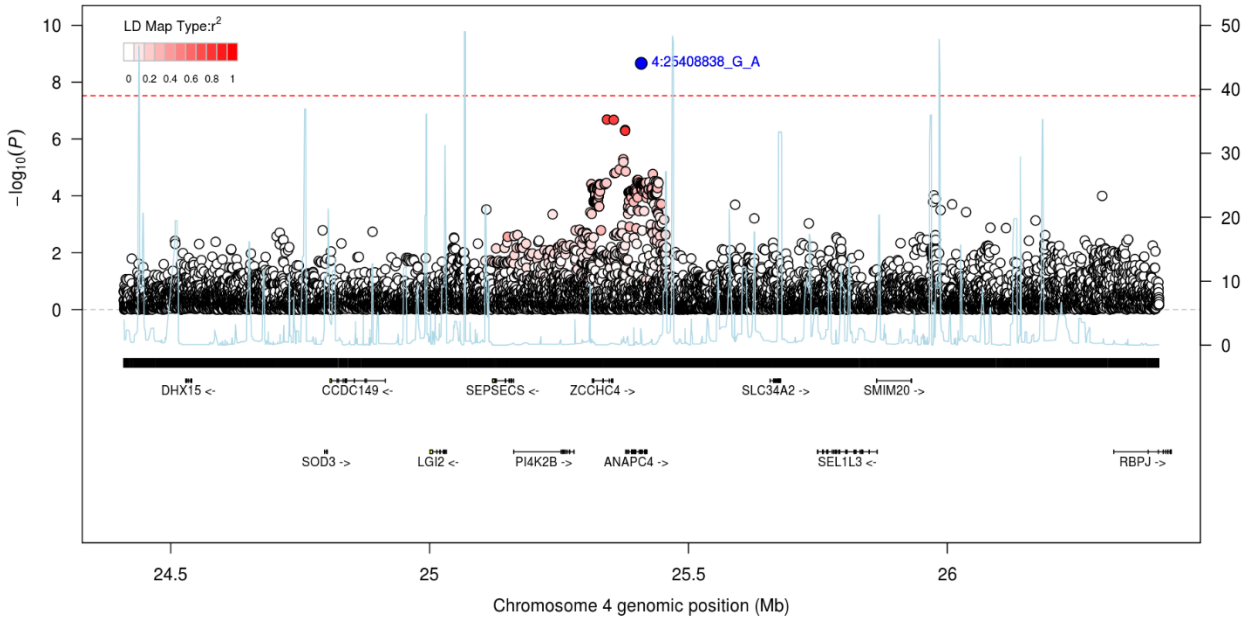

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
1:245750932_T_G
−log10(P)
C1orf101 ->
COX20 ->
EFCAB2 ->
SMYD3 <-
CNST ->
DESI2 ->
HNRNPU <-
KIF26B ->
TFB2M <-
245
245.5
246
246.5
Chromosome 1 genomic position (Mb)

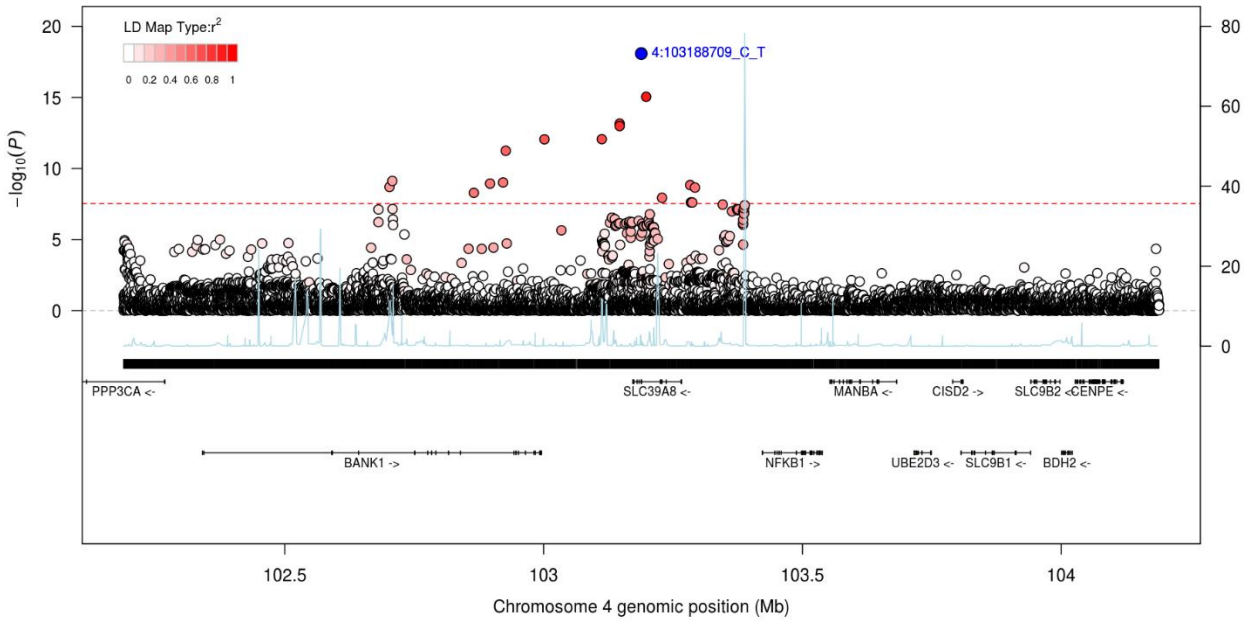

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
2:33434336_A_G
−log10(P)
SLC30A6
YPF4 ->
TTC27 ->
RASGRP3 ->
NLRC4 <-
BIRC6 ->
LTBP1 ->
FAM98A <-
32.5
33
33.5
34
34.5
Chromosome 2 genomic position (Mb)

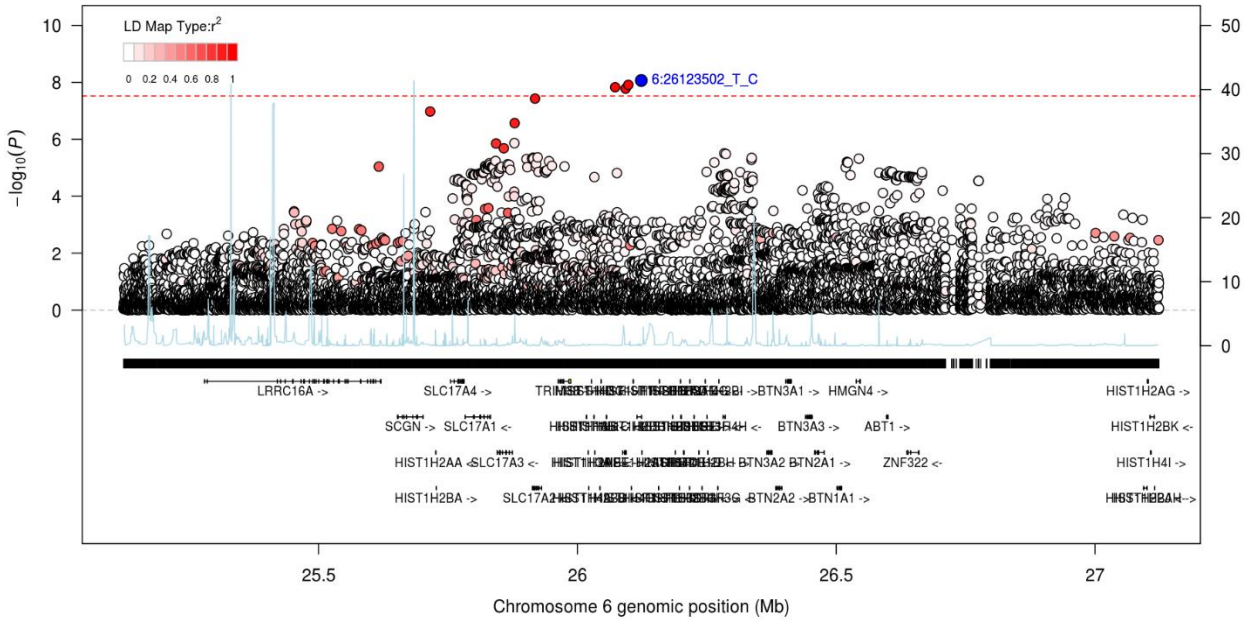

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
2:70717653_A_G
−log10(P)
AAK1 <-
GMCL1
MXD1 ->
PCBP1 ->
TGFA <-
DYSF ->
ANXA4 ->
ADD2 <-
CLEC4F <-
ZNF638 ->
70
70.5
71
71.5
Chromosome 2 genomic position (Mb)

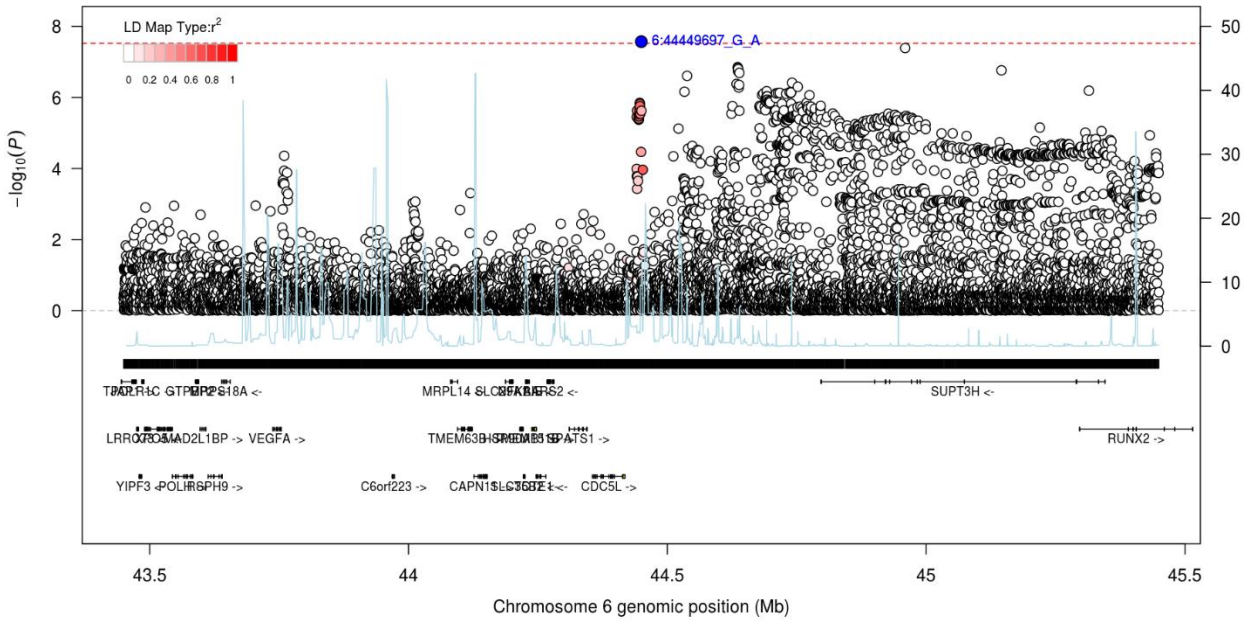

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
2:204387482_G_A
−log10(P)
BMPR2 ->
WDR12 <-
CYP20A1 ->
CD28 ->
FAM117B ->
CARF ->
ABI2 ->
CTLA4 ->
ICA1L <-
NBEAL1 ->
RAPH1 <-
ICOS ->
203.5
204
204.5
205
Chromosome 2 genomic position (Mb)

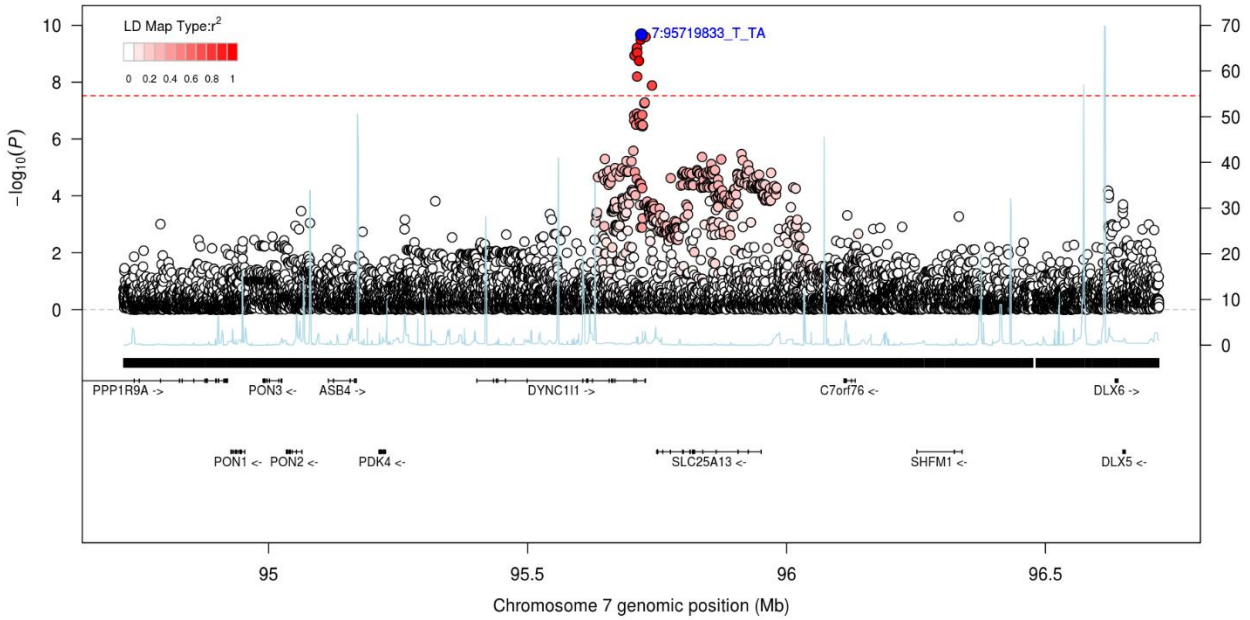

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
3:50022049_G_A
−log10(P)
12
10
8
6
4
2
0
15
10
5
0
RBM6 ->
CISH <-
RHOA <-
DAG1 ->
RBM5 ->
MAPKAPK3 ->
TCTA ->
BSN ->
SEMA3F ->
C3orf18 <-
DOCK3 ->
USP4 <-
AMT <-
MON1A <-
HEMK1 ->
49
49.5
50
50.5
51
Chromosome 3 genomic position (Mb)

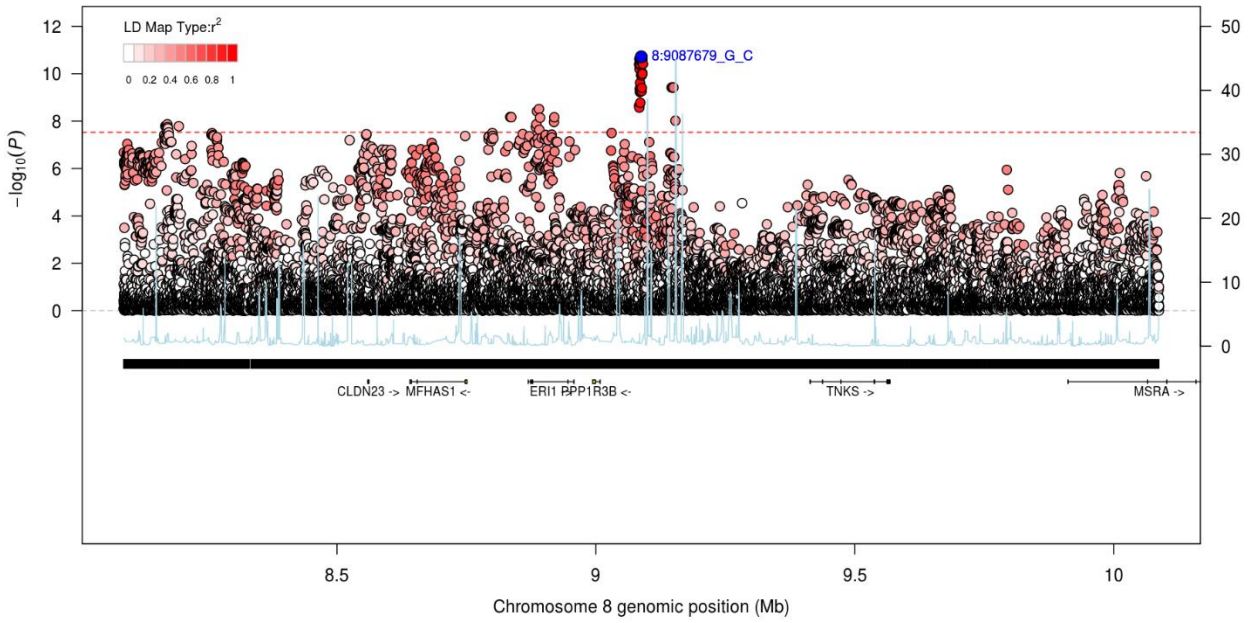

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
4:25408838_G_A
−log10(P)
10
8
6
4
2
0
50
40
30
20
10
0
DHX15 <-
CCDC149 <-
SEPSECS <-
ZCCHC4 ->
SLC34A2 ->
SMIM20 ->
SOD3 ->
LGI2 <-
PI4K2B ->
ANAPC4 ->
SEL1L3 <-
RBPJ ->
24.5
25
25.5
26
Chromosome 4 genomic position (Mb)

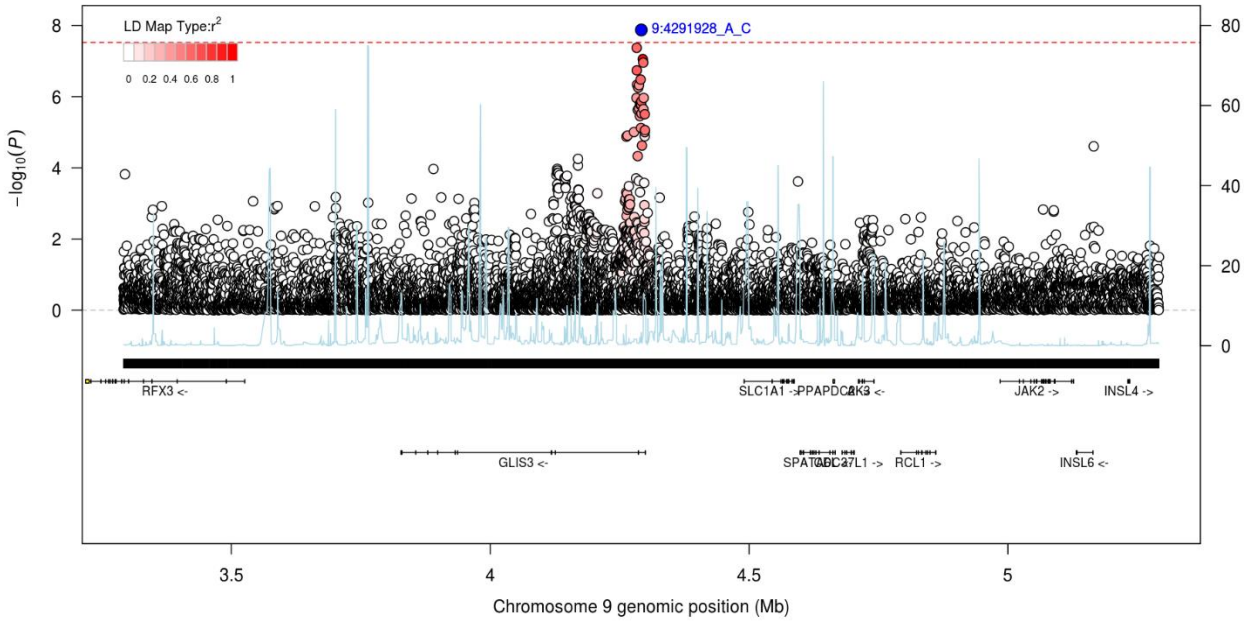

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
4:103188709_C_T
−log10(P)
PPP3CA <-
SLC39A8 <-
MANBA <-
CISD2 ->
SLC9B2 <-
CENPE <-
BANK1 ->
NFKB1 ->
UBE2D3 <-
SLC9B1 <-
BDH2 <-
102.5
103
103.5
104
Chromosome 4 genomic position (Mb)

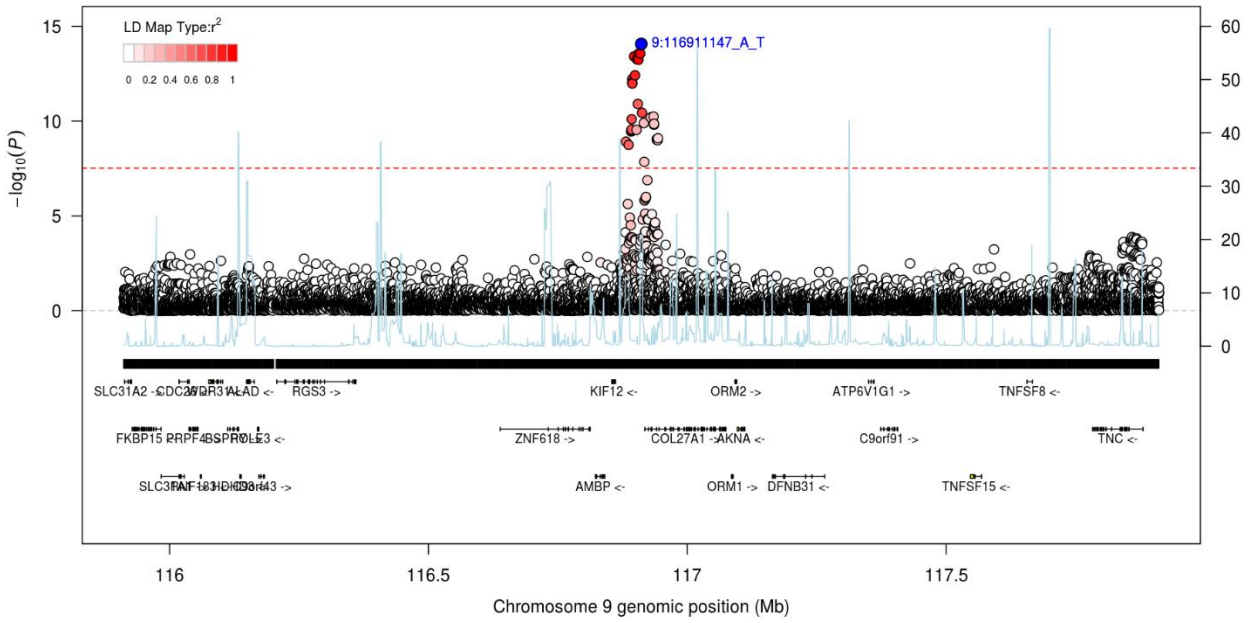

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
6:26123502_T_C
−log10(P)
LRRC16A ->
SLC17A4 ->
BTN3A1 ->
HMGN4 ->
HIST1H2AG ->
SCGN ->
SLC17A1 <-
BTN3A3 ->
ABT1 ->
HIST1H2BK <-
HIST1H2AA <-
SLC17A3 <-
BTN3A2
BTN2A1 ->
ZNF322 <-
HIST1H4I ->
HIST1H2BA ->
BTN2A2 ->
BTN1A1 ->
25.5
26
26.5
27
Chromosome 6 genomic position (Mb)

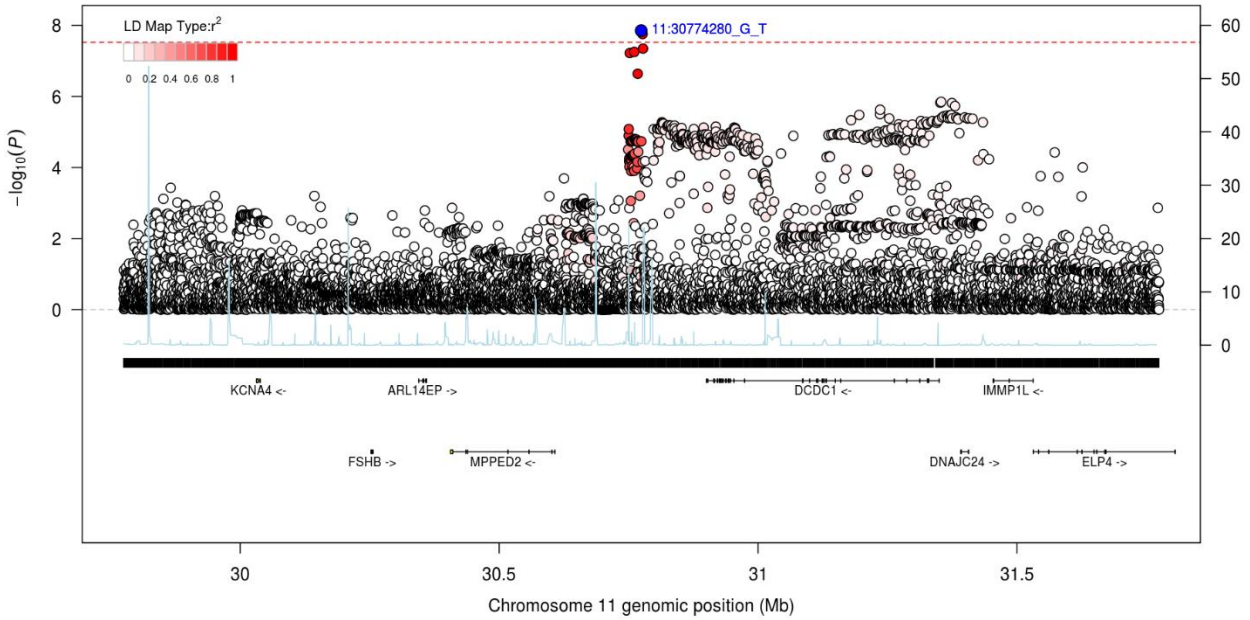

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
6:44449697_G_A
−log10(P)
SUPT3H <-
RUNX2 ->
VEGFA ->
C6orf223 ->
CDC5L ->
43.5
44
44.5
45
45.5
Chromosome 6 genomic position (Mb)

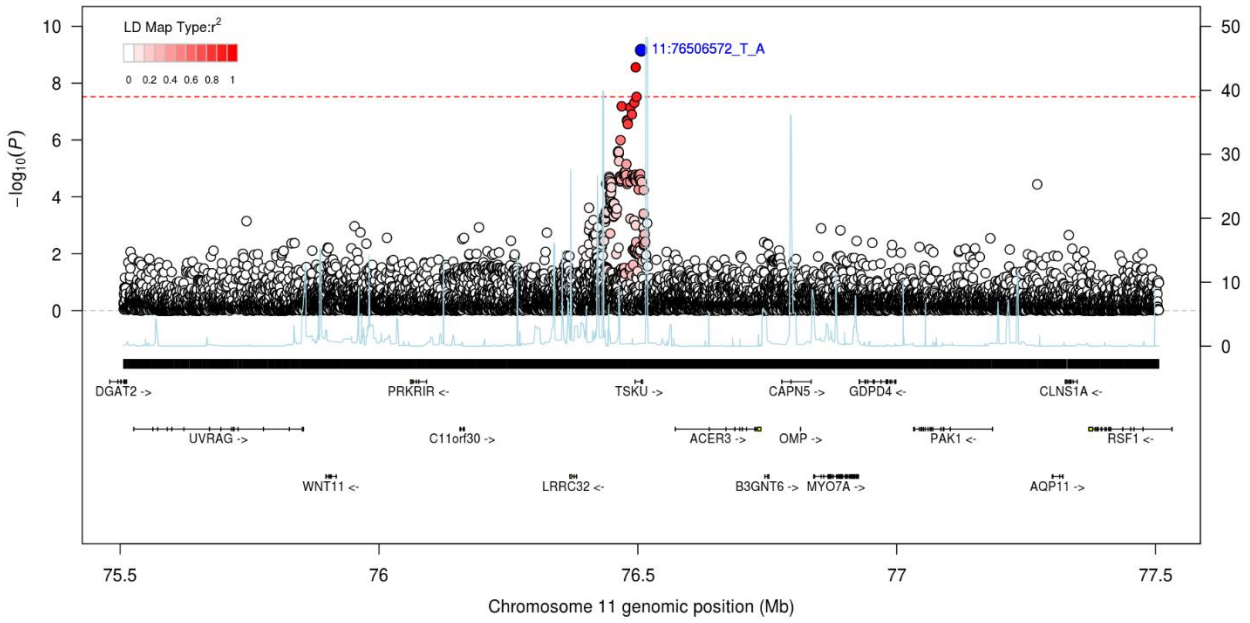

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
7:95719833_T_TA
−log10(P)
PPP1R9A ->
PON3 <-
ASB4 ->
DYNC1I1 ->
C7orf76 <-
DLX6 ->
PON1 <-
PON2 <-
PDK4 <-
SLC25A13 <-
SHFM1 <-
DLX5 <-
95
95.5
96
96.5
Chromosome 7 genomic position (Mb)

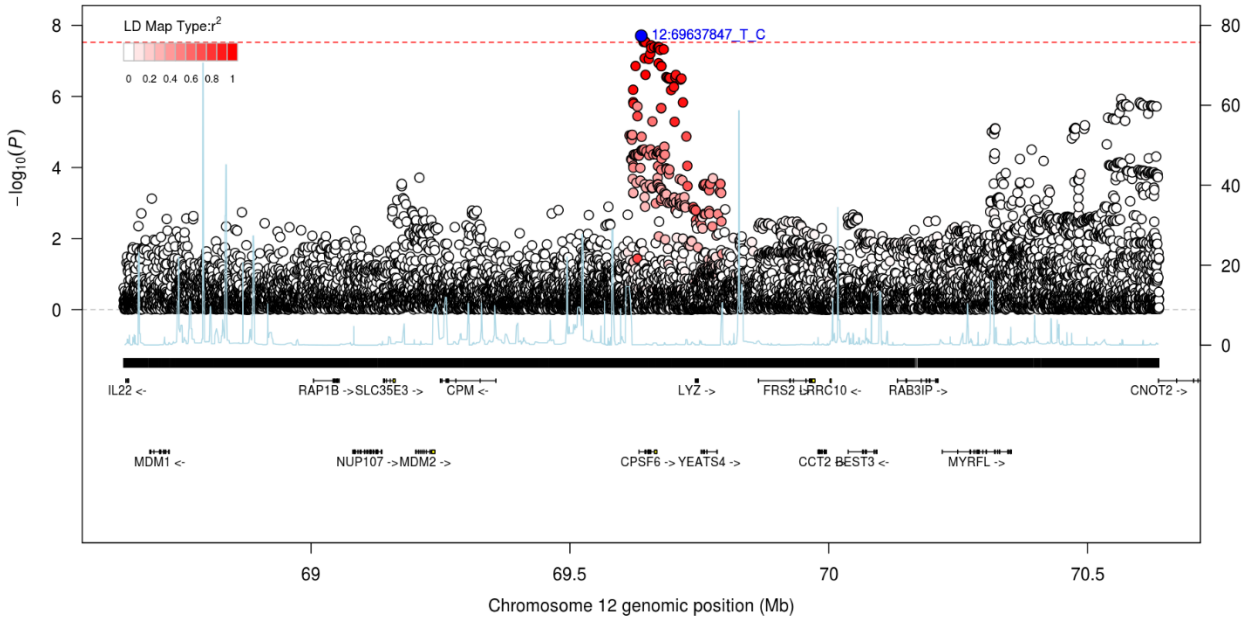

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
8:9087679_G_C
−log10(P)
CLDN23 ->
MFHAS1 <-
ERI1
PPP1R3B <-
TNKS ->
MSRA ->
Chromosome 8 genomic position (Mb)

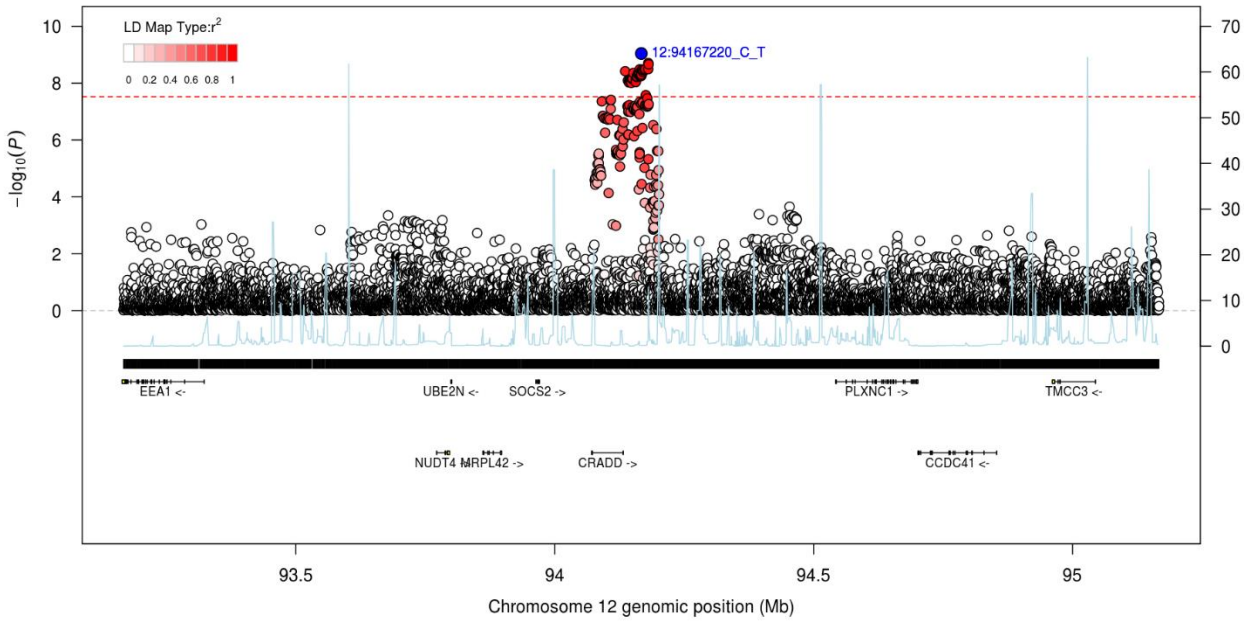

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
9:4291928_A_C
−log10(P)
RFX3 <-
SLC1A1 ->
JAK2 ->
INSL4 ->
GLIS3 <-
RCL1 ->
INSL6 <-
Chromosome 9 genomic position (Mb)

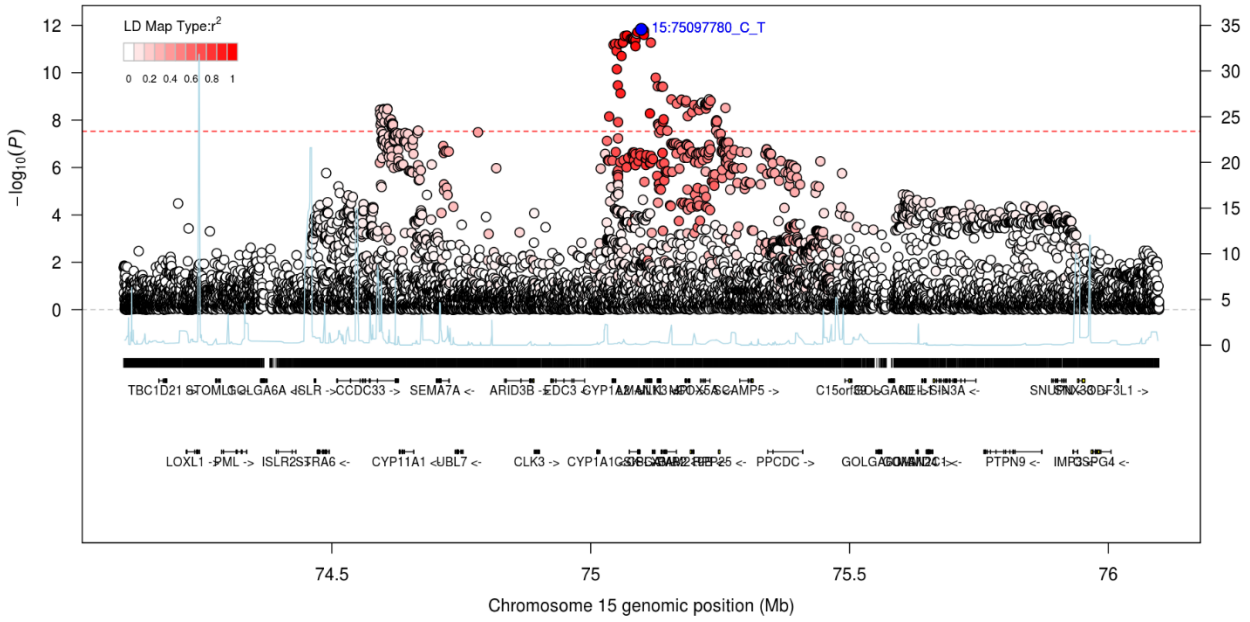

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
9:116911147_A_T
−log10(P)
SLC31A2 ->
RGS3 ->
KIF12 <-
ORM2 ->
ATP6V1G1 ->
TNFSF8 <-
FKBP15
ZNF618 ->
COL27A1 ->
AKNA <-
C9orf91 ->
TNC <-
AMBP <-
ORM1 ->
DFNB31 <-
TNFSF15 <-
116
116.5
117
117.5
Chromosome 9 genomic position (Mb)

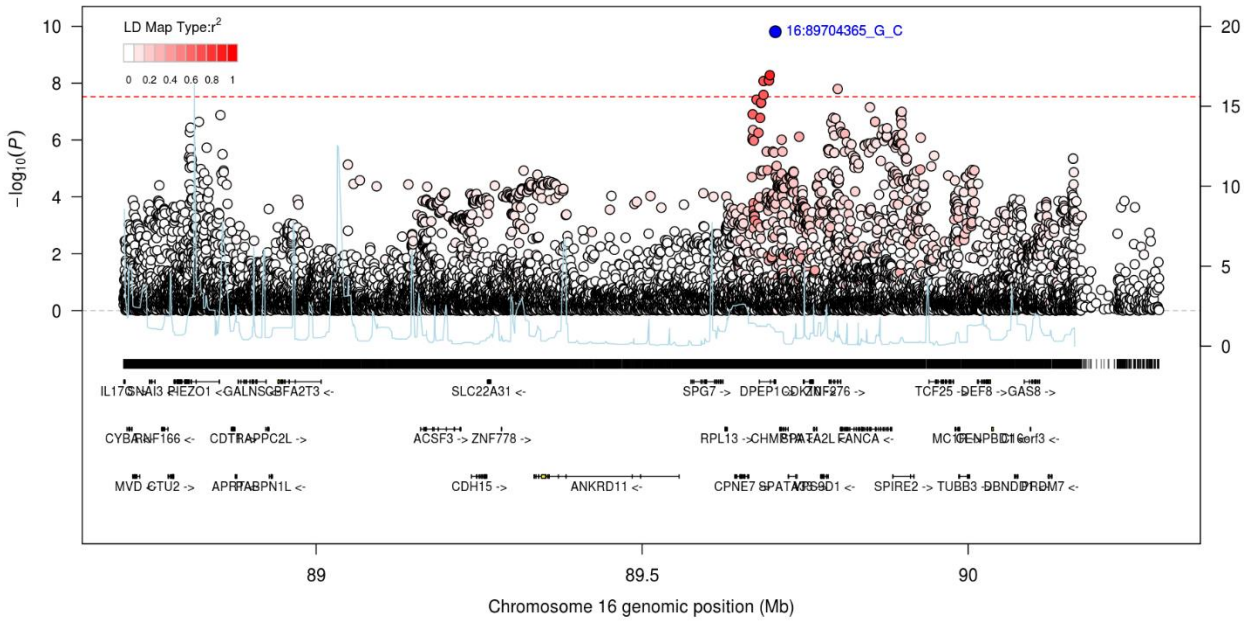

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
11:30774280_G_T
−log10(P)
KCNA4 <-
ARL14EP ->
DCDC1 <-
IMMP1L <-
FSHB ->
MPPED2 <-
DNAJC24 ->
ELP4 ->
30
30.5
31
31.5
Chromosome 11 genomic position (Mb)

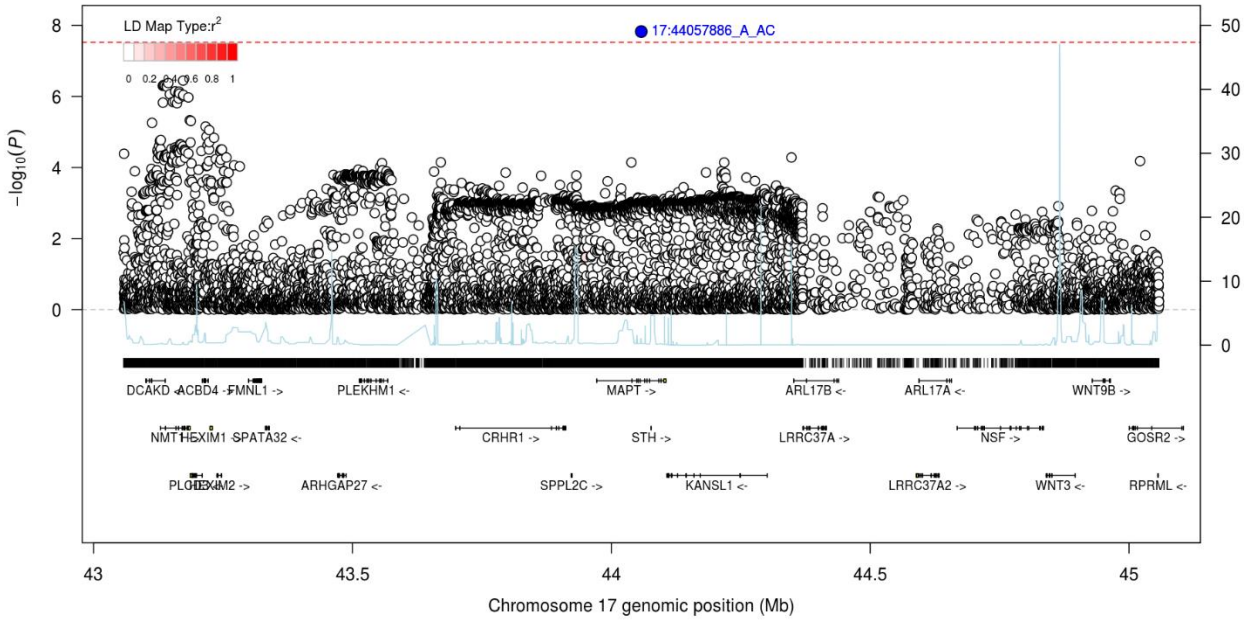

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
11:76506572_T_A
−log10(P)
DGAT2 ->
PRKRIR <-
TSKU ->
CAPN5 ->
GDPD4 <-
CLNS1A <-
UVRAG ->
C11orf30 ->
ACER3 ->
OMP ->
PAK1 <-
RSF1 <-
WNT11 <-
LRRC32 <-
B3GNT6 ->
MYO7A ->
AQP11 ->
75.5
76
76.5
77
77.5
Chromosome 11 genomic position (Mb)

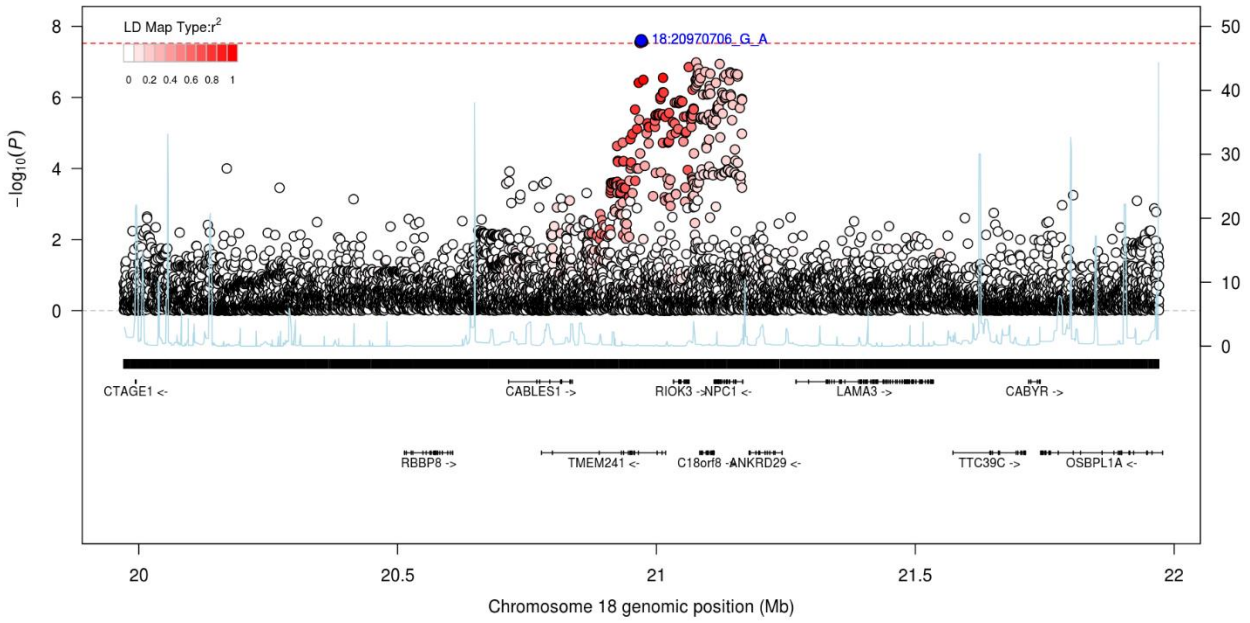

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
12:69637847_T_C
−log10(P)
IL22 <-
RAP1B ->
SLC35E3 ->
CPM <-
LYZ ->
FRS2
LRRC10 <-
RAB3IP ->
CNOT2 ->
MDM1 <-
NUP107 ->
MDM2 ->
CPSF6 ->
YEATS4 ->
CCT2
BEST3 <-
MYRFL ->
69
69.5
70
70.5
Chromosome 12 genomic position (Mb)

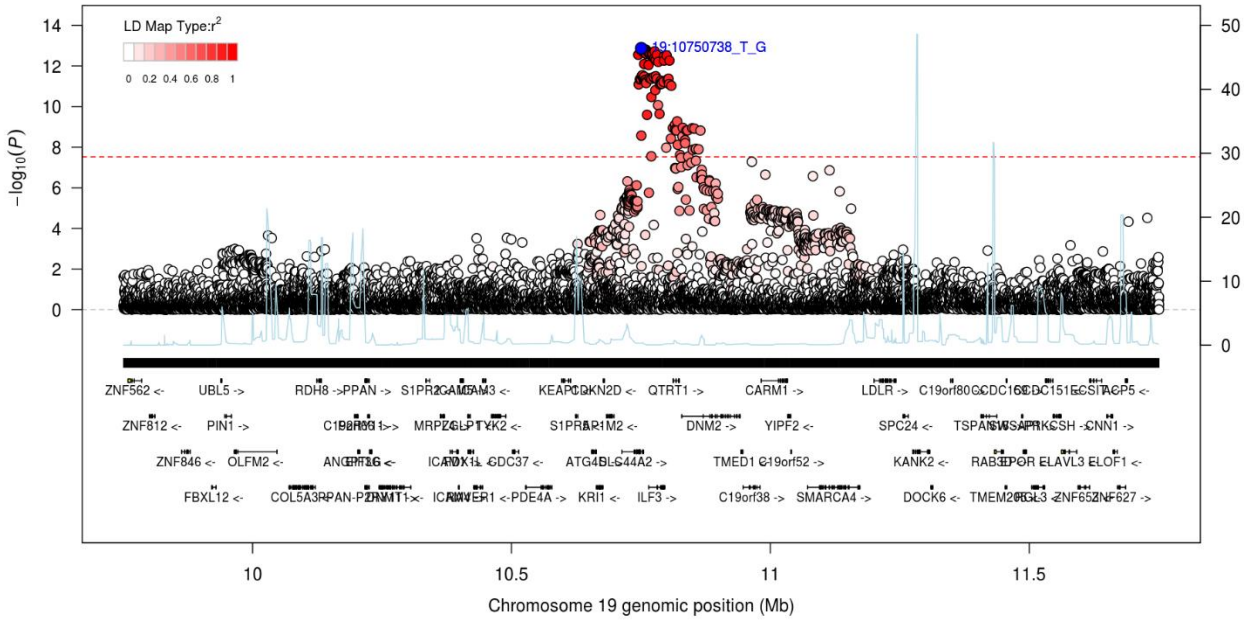

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
12:94167220_C_T
−log10(P)
EEA1 <-
UBE2N <-
SOCS2 ->
PLXNC1 ->
TMCC3 <-
NUDT4
MRPL42 ->
CRADD ->
CCDC41 <-
93.5
94
94.5
95
Chromosome 12 genomic position (Mb)

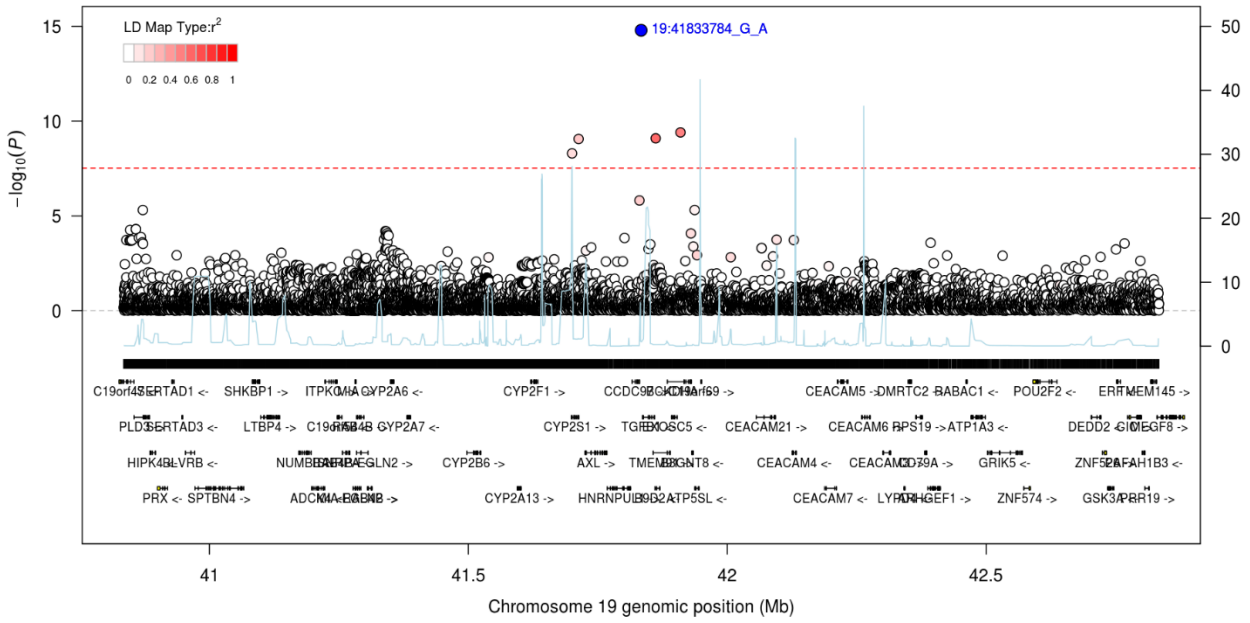

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
15:75097780_C_T
−log10(P)
CCDC33 ->
SEMA7A <-
ARID3B
CAMP5 ->
LOXL1 -
PML ->
ISLR2
CYP11A1 <-
UBL7 <-
CLK3 ->
PPCDC ->
PTPN9 <-
74.5
75
75.5
76
Chromosome 15 genomic position (Mb)

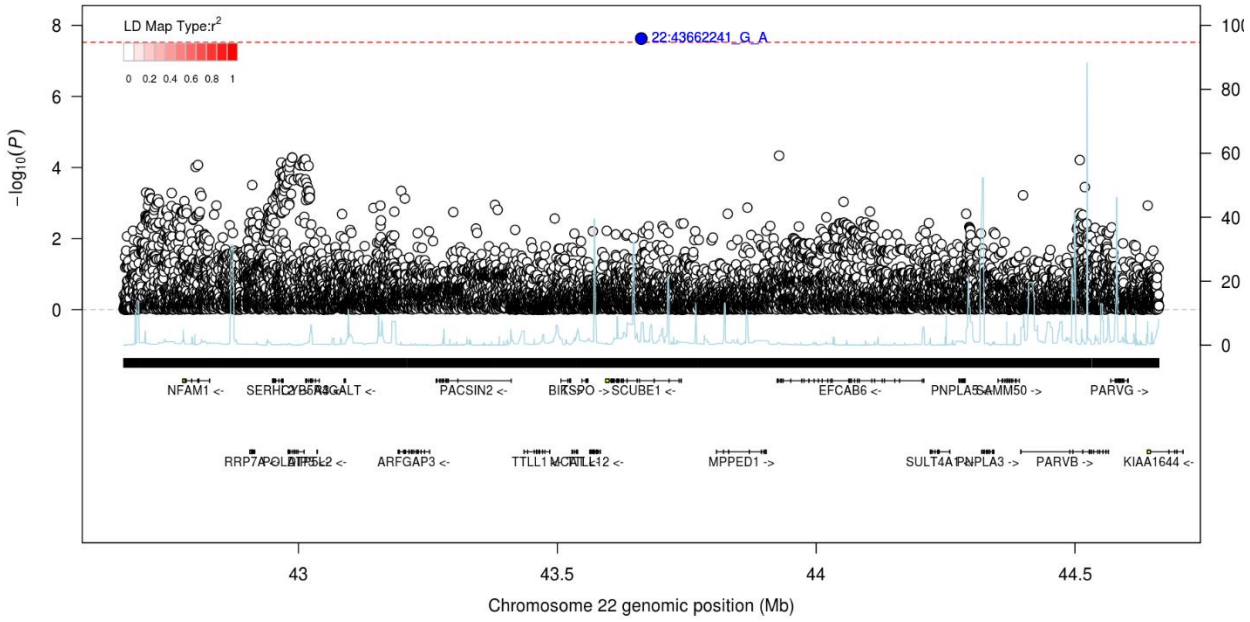

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
16:89704365_G_C
−log10(P)
10
8
6
4
2
0
20
15
10
5
0
89
89.5
90
Chromosome 16 genomic position (Mb)


LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
17:44057886_A_AC
−log10(P)
8
6
4
2
0
50
40
30
20
10
0
43
43.5
44
44.5
45
Chromosome 17 genomic position (Mb)

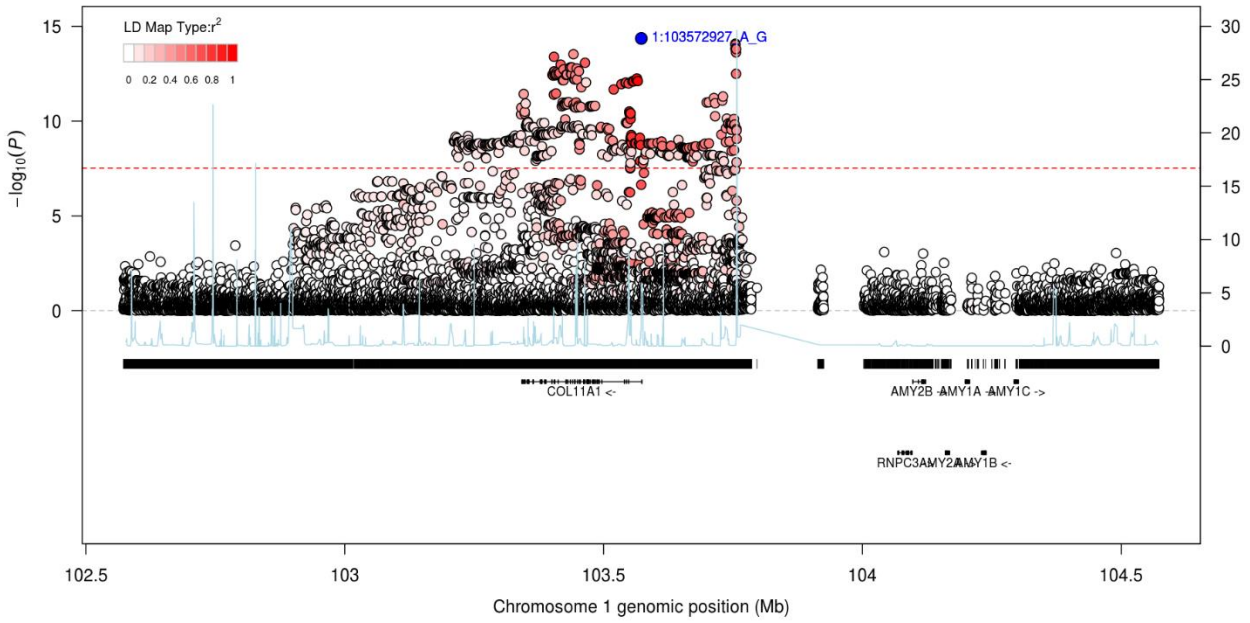

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
18:20970706_G_A
−log10(P)
CTAGE1 <-
CABLES1 ->
RIOK3 ->
NPC1 <-
LAMA3 ->
CABYR ->
RBBP8 ->
TMEM241 <-
C18orf8 ->
ANKRD29 <-
TTC39C ->
OSBPL1A <-
Chromosome 18 genomic position (Mb)

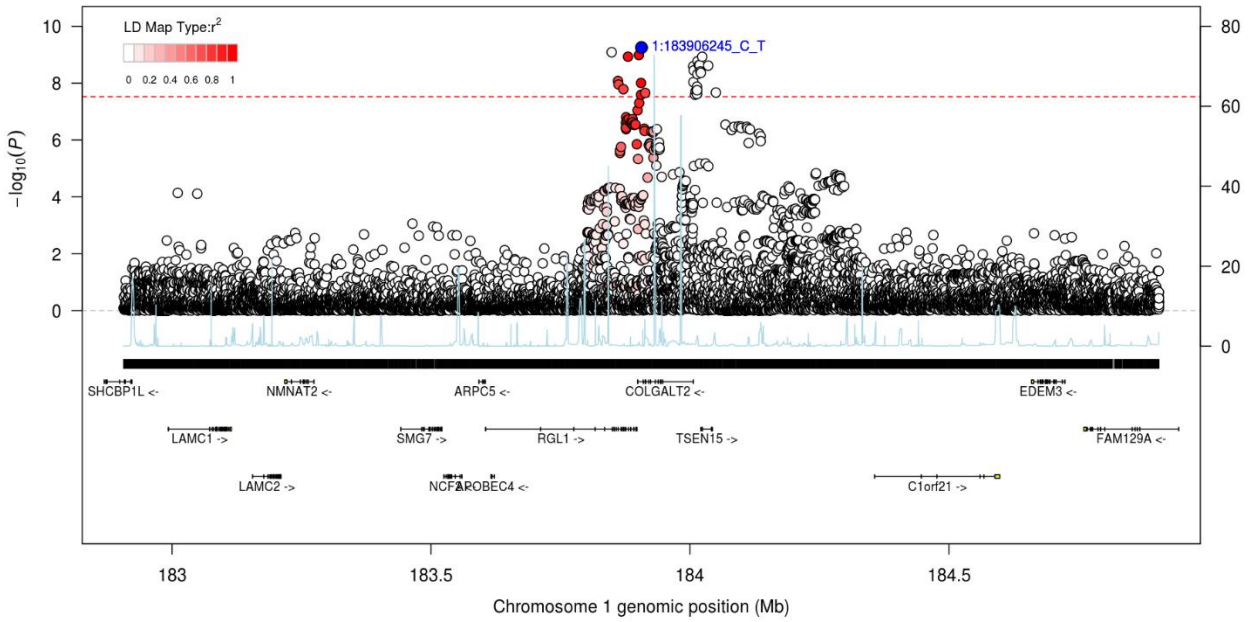

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
19:10750738_T_G
−log10(P)
ZNF562 <-
UBL5 ->
RDH8 ->
PPAN ->
KEAP1 ->
QTRT1 ->
CARM1 ->
LDLR ->
ZNF812 <-
PIN1 ->
DNM2 ->
YIPF2 ->
SPC24 <-
CNN1 ->
ZNF846 <-
OLFM2 <-
TMED1 ->
KANK2 <-
FBXL12 <-
PDE4A ->
KRI1 <-
ILF3 ->
SMARCA4 ->
DOCK6 <-
Chromosome 19 genomic position (Mb)

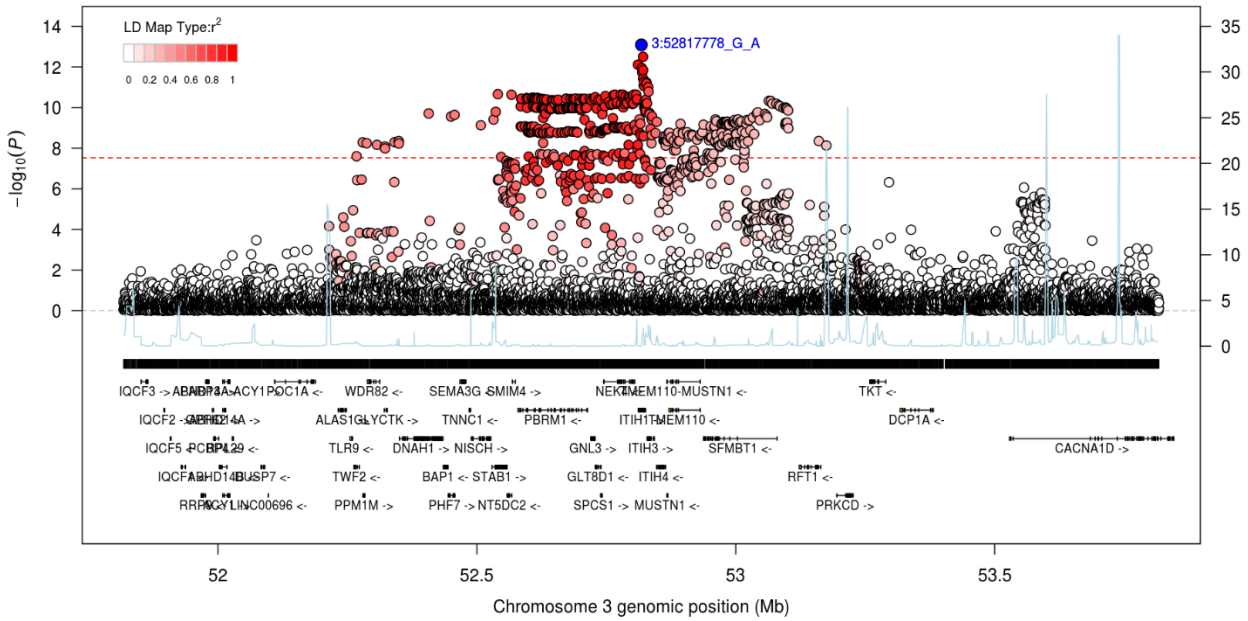

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
19:41833784_G_A
−log10(P)
Chromosome 19 genomic position (Mb)

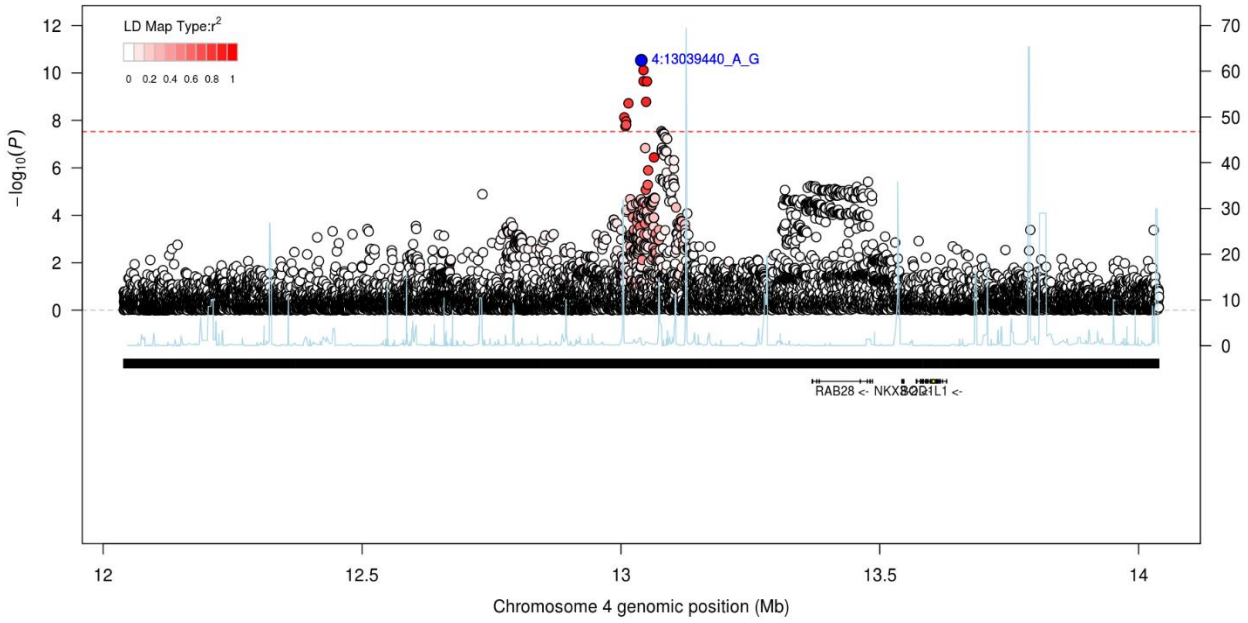

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
22:43662241_G_A
−log10(P)
Chromosome 22 genomic position (Mb)

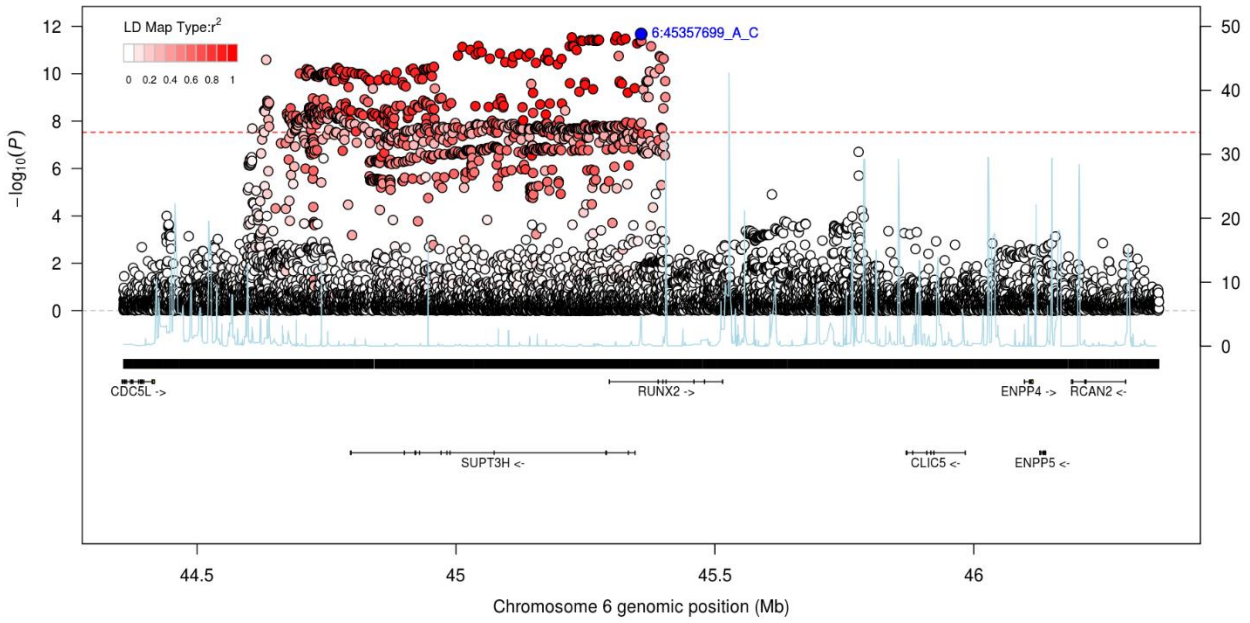
Supplementary Figure 4: Locus zoom plots for 21 hip osteoarthritis loci that reached genome-wide significance in the meta-analysis (as reported in Table 1 and Supplementary Table 3), displaying 1Mb each side of the top variant, where the blue circle represents the index variant on the meta-analysis across UK Biobank and arcOGEN. LD is calculated from UK Biobank.

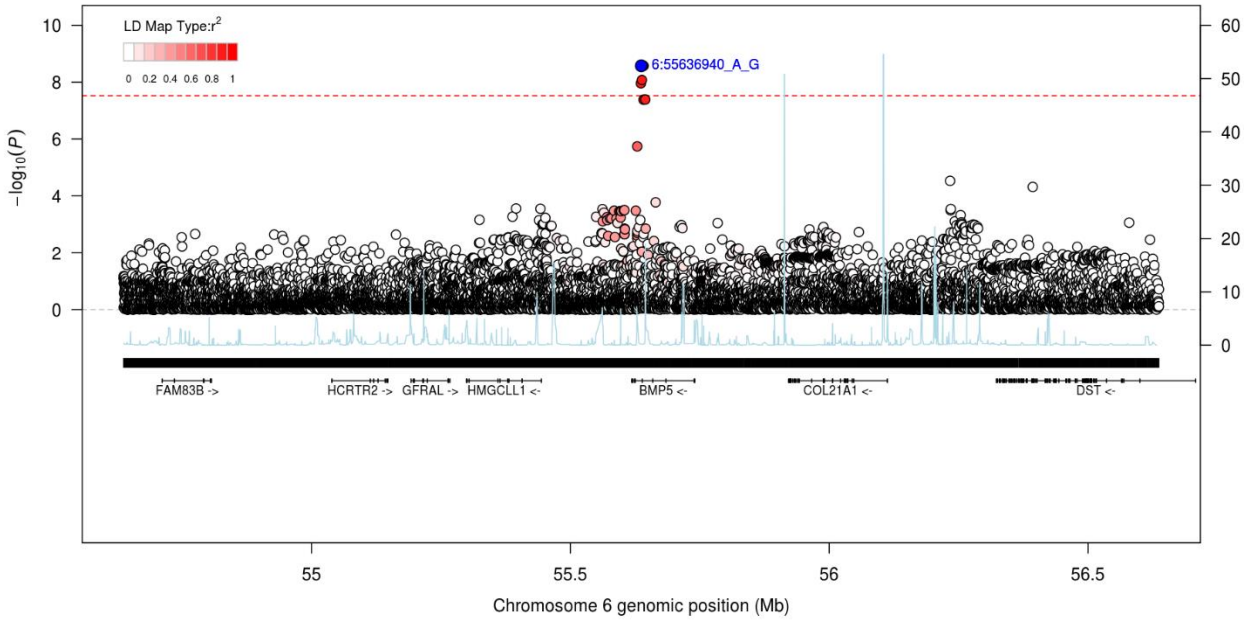


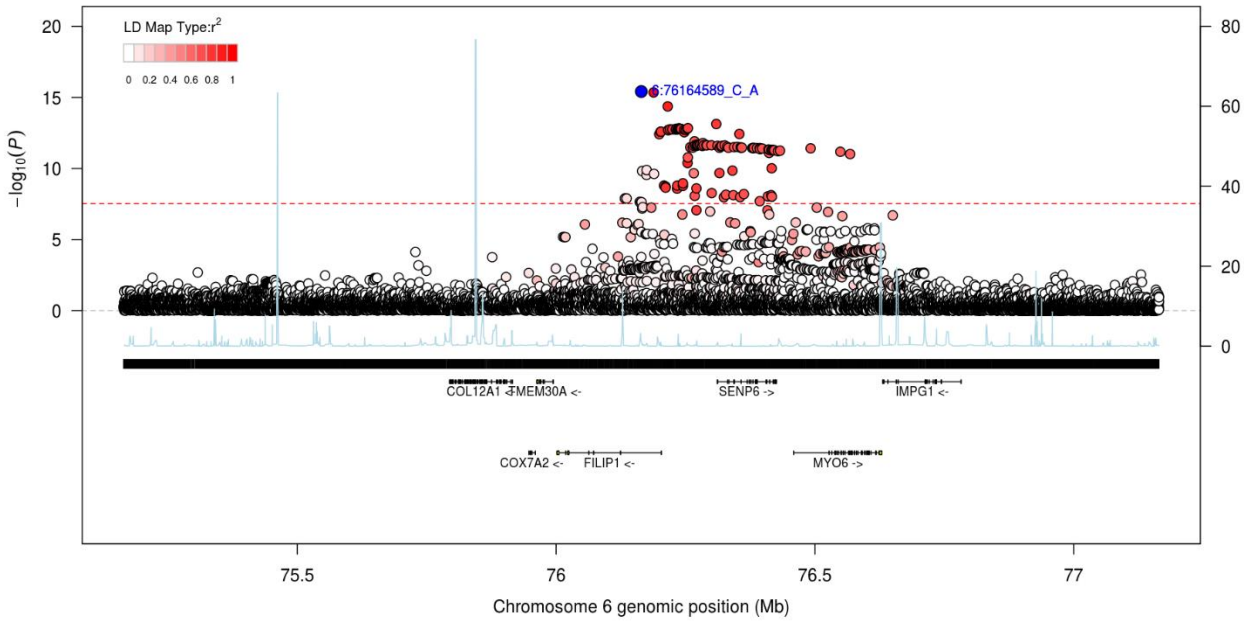


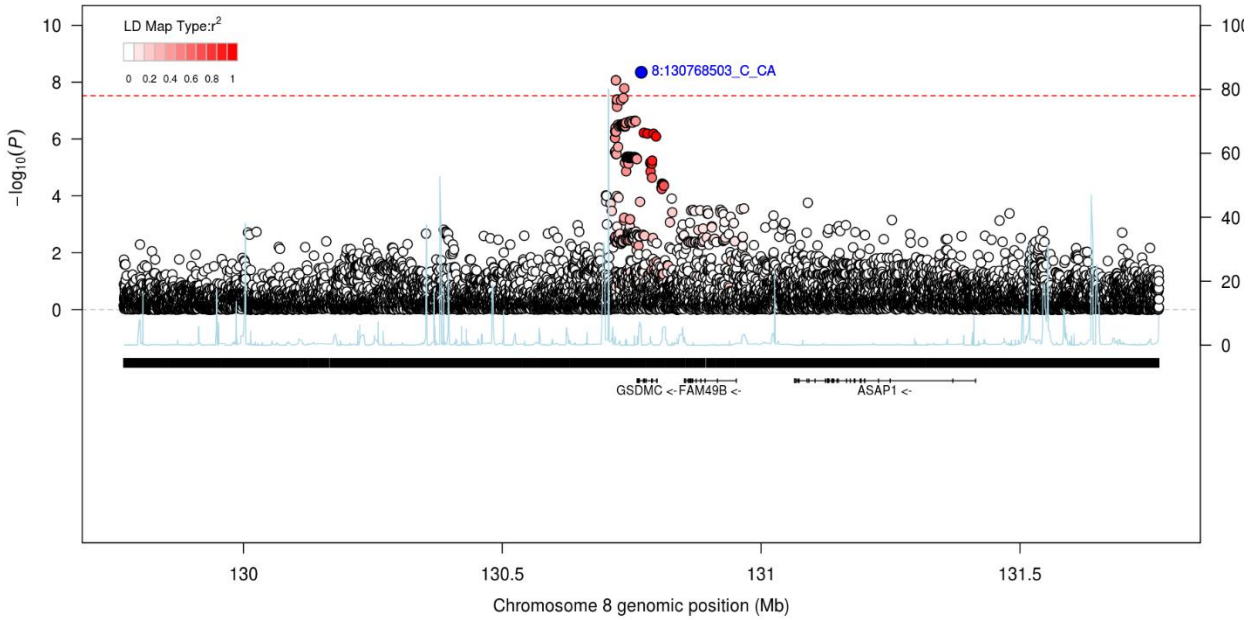

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
3:52817778_G_A
−log10(P)
Chromosome 3 genomic position (Mb)

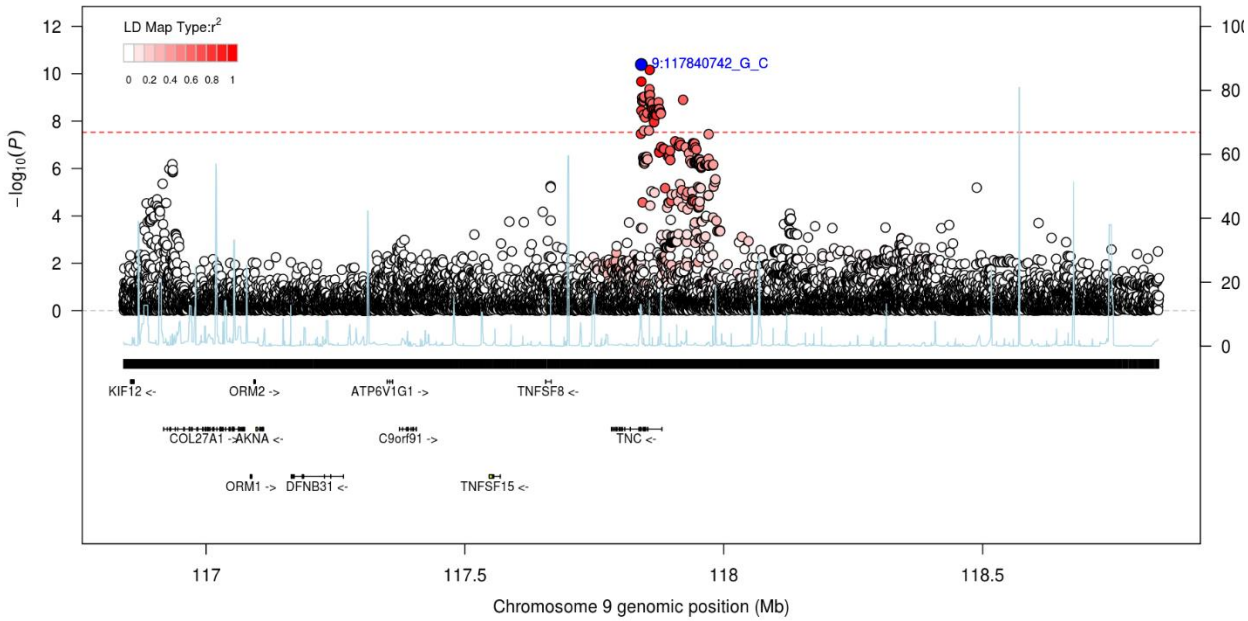

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
4:13039440_A_G
−log10(P)
Chromosome 4 genomic position (Mb)

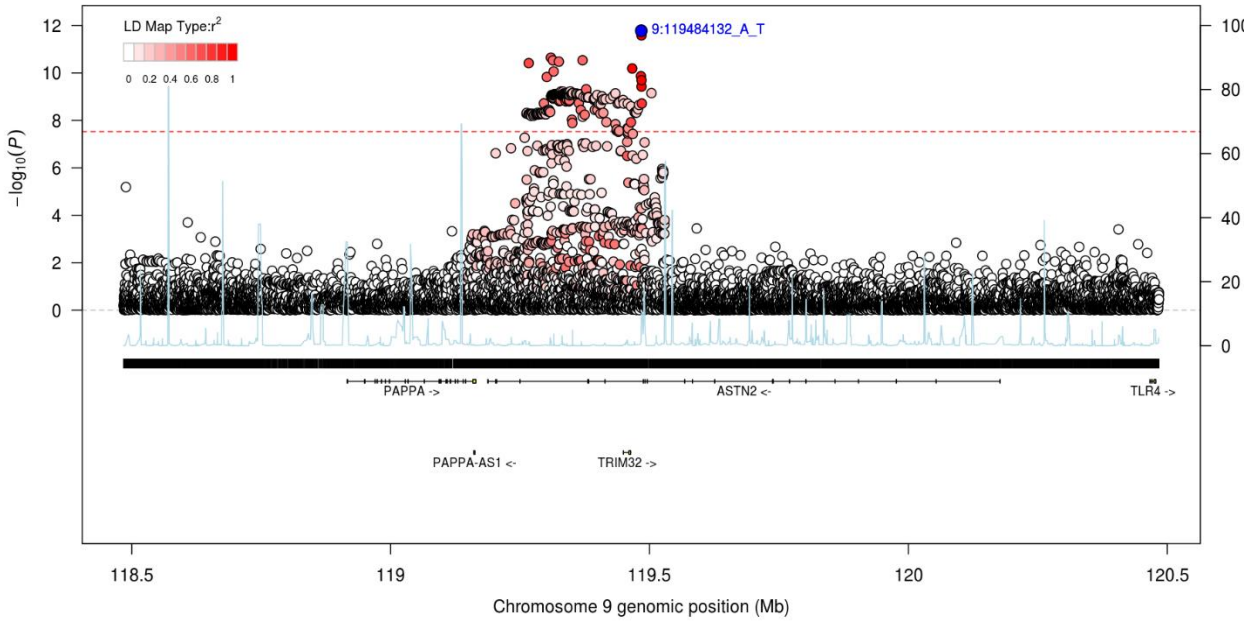

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
6:45357699_A_C
−log10(P)
CDC5L ->
RUNX2 ->
ENPP4 ->
RCAN2 <-
SUPT3H <-
CLIC5 <-
ENPP5 <-
44.5
45
45.5
46
Chromosome 6 genomic position (Mb)

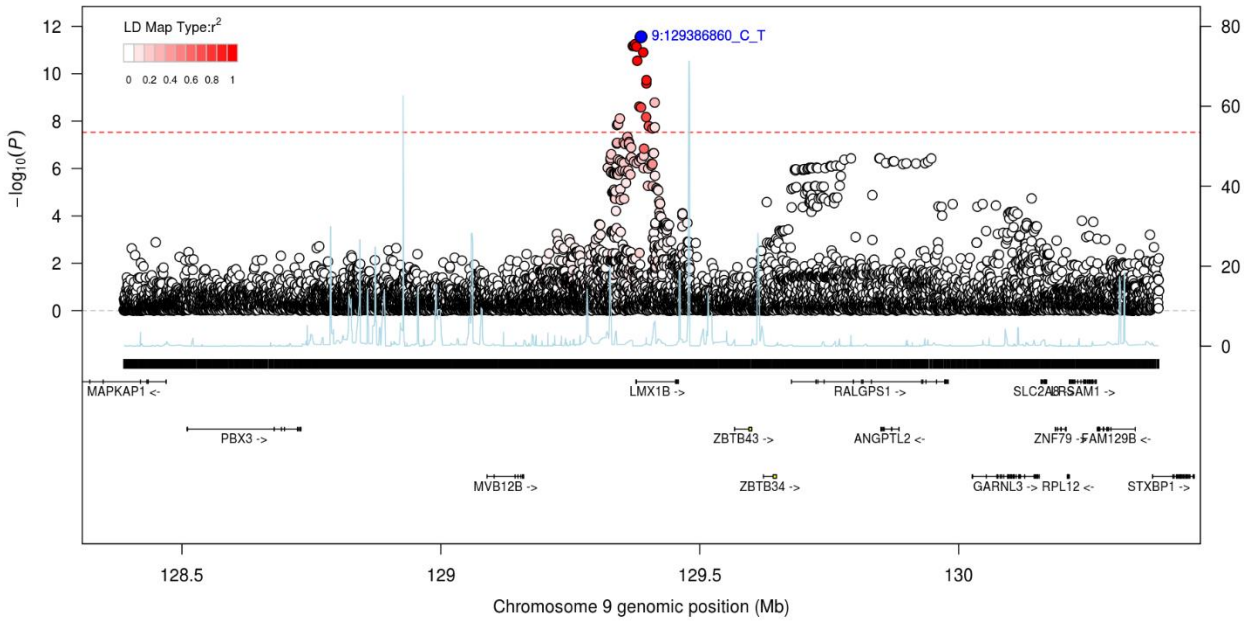

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
6:55636940_A_G
−log10(P)
FAM83B ->
HCRTR2 ->
GFRAL ->
HMGCLL1 <-
BMP5 <-
COL21A1 <-
DST <-
55
55.5
56
56.5
Chromosome 6 genomic position (Mb)

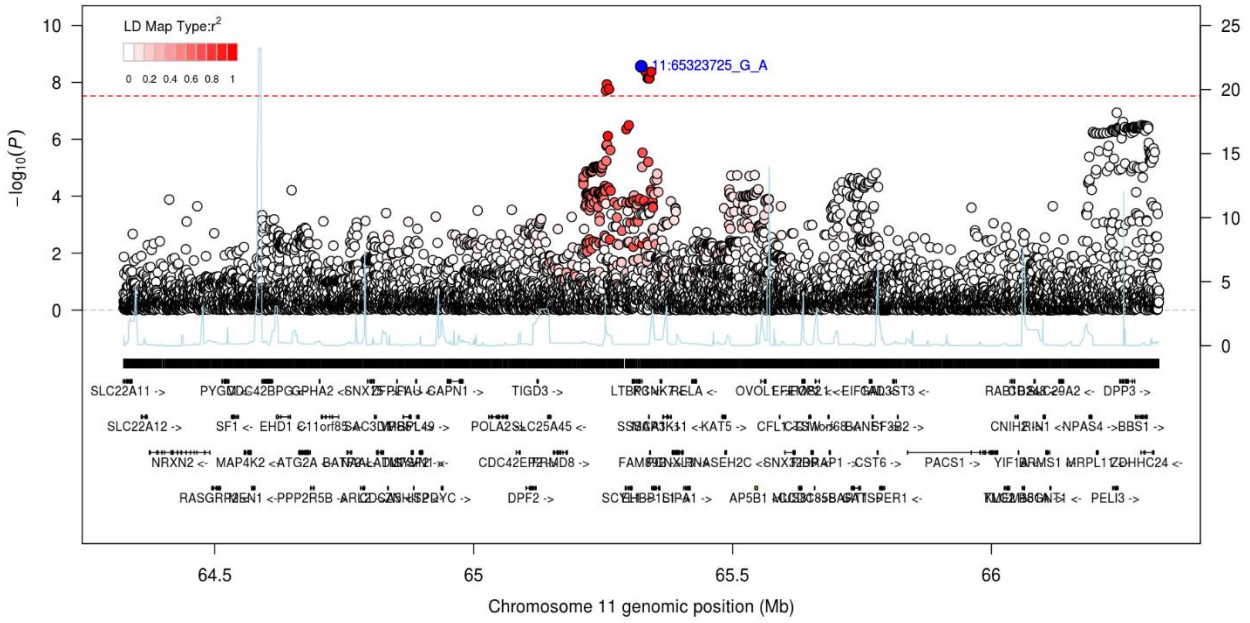

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
6:76164589_C_A
−log10(P)
COL12A1
MEM30A
SENP6
IMPG1
COX7A2
FILIP1
MYO6
75.5
76
76.5
77
Chromosome 6 genomic position (Mb)

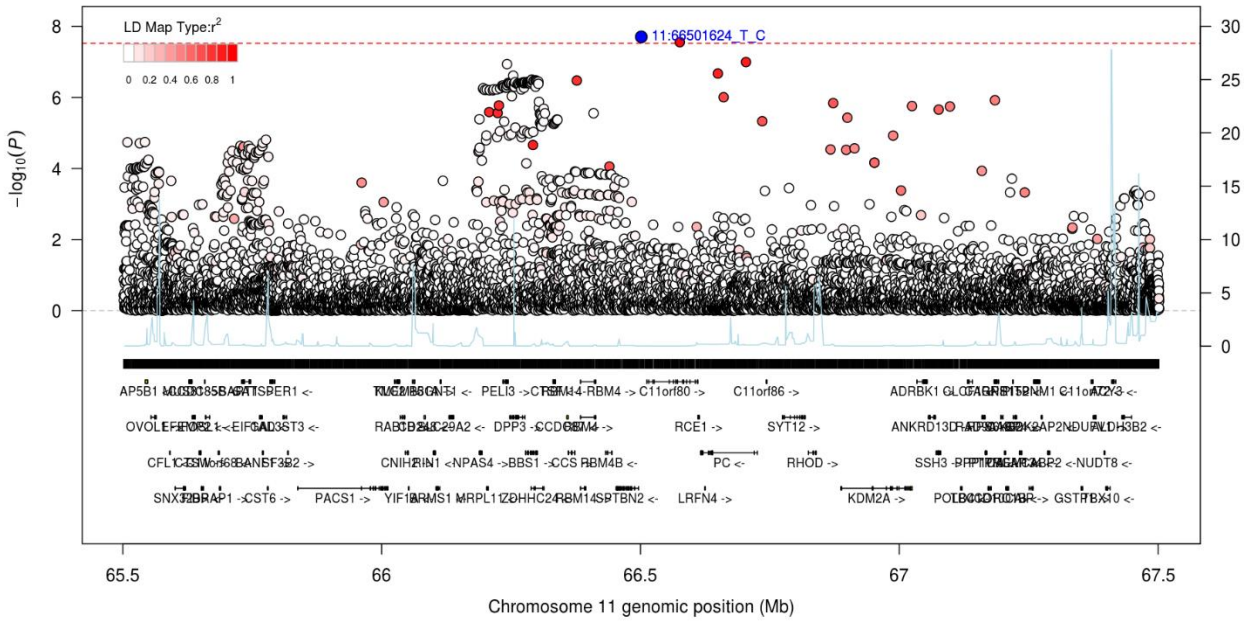

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
8:130768503_C_CA
−log10(P)
GSDMC
FAM49B
ASAP1
130
130.5
131
131.5
Chromosome 8 genomic position (Mb)

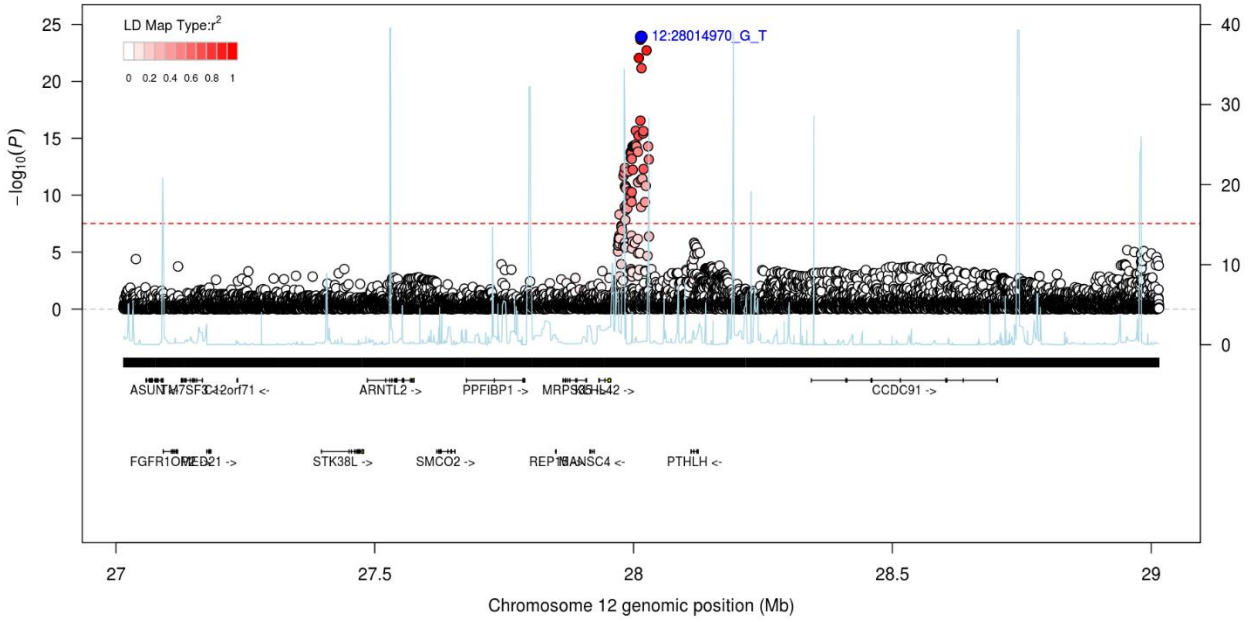

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
9:117840742_G_C
−log10(P)
KIF12 <-
ORM2 ->
ATP6V1G1 ->
TNFSF8 <-
COL27A1 ->
AKNA <-
C9orf91 ->
TNC <-
ORM1 ->
DFNB31 <-
TNFSF15 <-
117
117.5
118
118.5
Chromosome 9 genomic position (Mb)

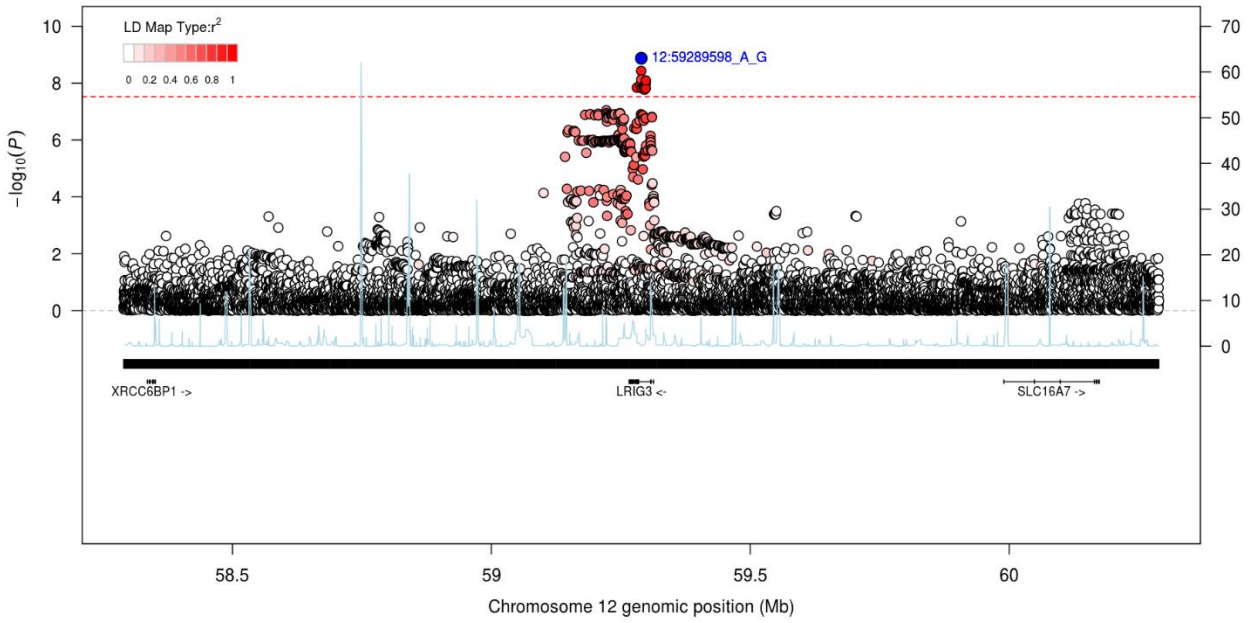

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
9:119484132_A_T
−log10(P)
PAPPA ->
ASTN2 <-
TLR4 ->
PAPPA-AS1 <-
TRIM32 ->
118.5
119
119.5
120
120.5
Chromosome 9 genomic position (Mb)

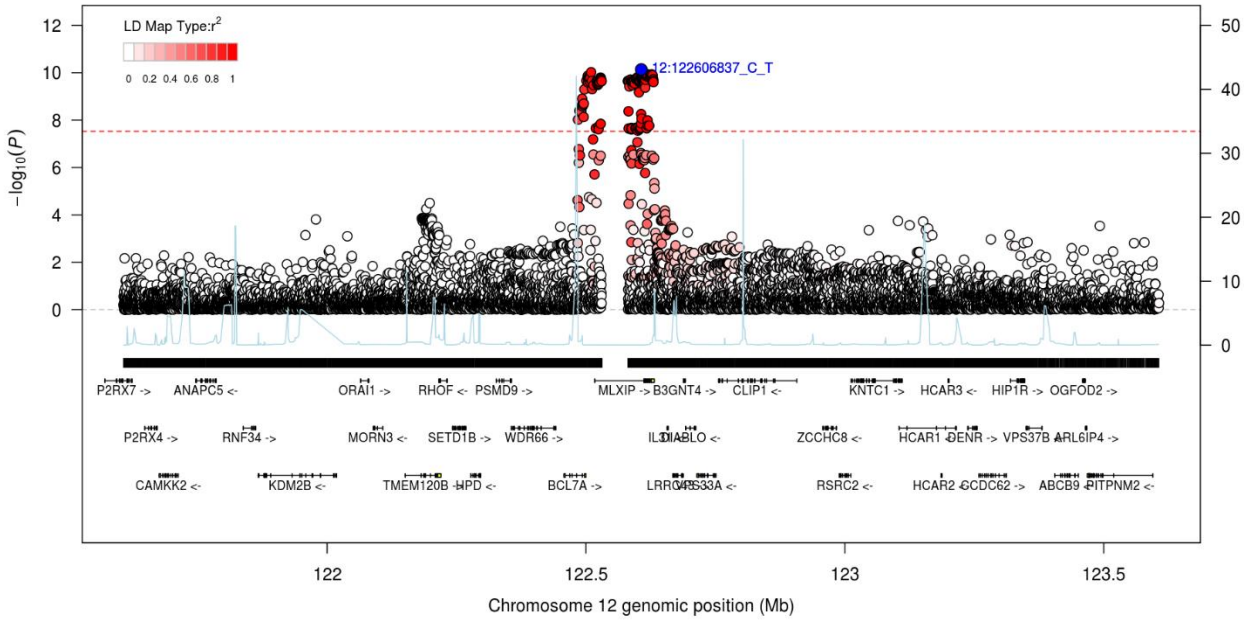

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
9:129386860_C_T
−log10(P)
MAPKAP1 <-
LMX1B ->
RALGPS1 ->
PBX3 ->
ZBTB43 ->
ANGPTL2 <-
ZNF79 ->
FAM129B <-
MVB12B ->
ZBTB34 ->
GARNL3 ->
RPL12 <-
STXBP1 ->
128.5
129
129.5
130
Chromosome 9 genomic position (Mb)

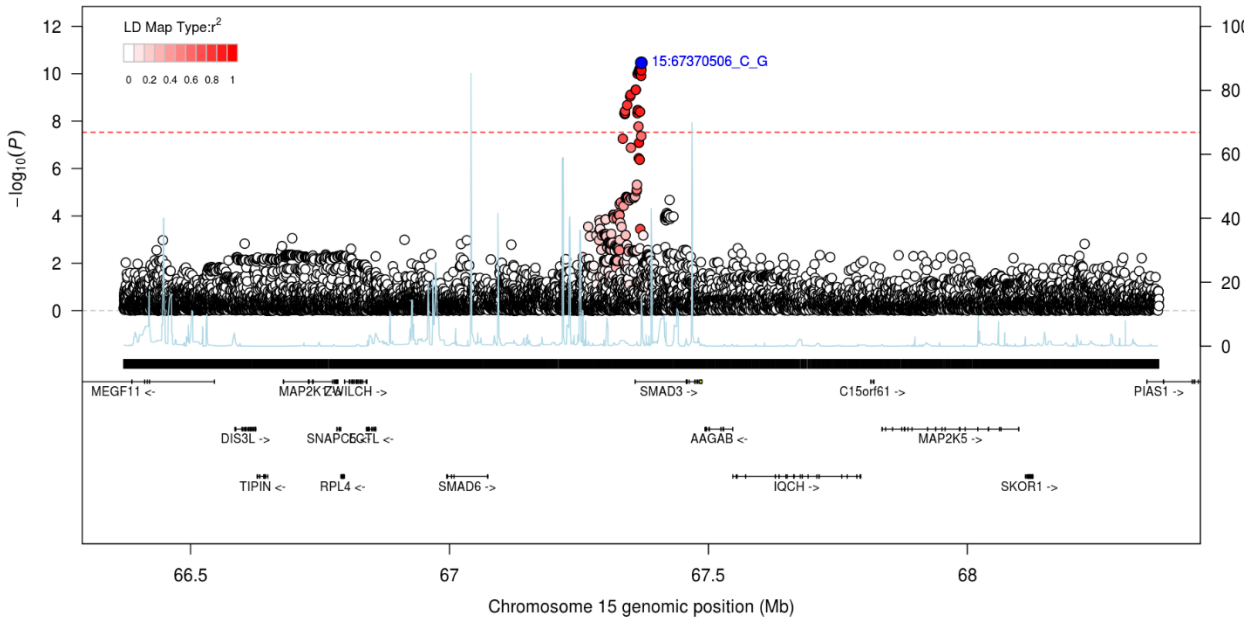

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
11:65323725_G_A
−log10(P)
SLC22A11 ->
TIGD3 ->
DPP3 ->
SLC22A12 ->
SF1 <-
EHD1 <-
POLA2
SLC25A45 <-
KAT5 ->
CFL1
CNIH2
NPAS4 ->
BBS1 ->
NRXN2 <-
MAP4K2 <-
ATG2A
CDC42EP1
CST6 ->
PACS1 ->
MRPL17
DHHC24 <-
DPF2 ->
AP5B1
PELI3 ->
64.5
65
65.5
66
Chromosome 11 genomic position (Mb)

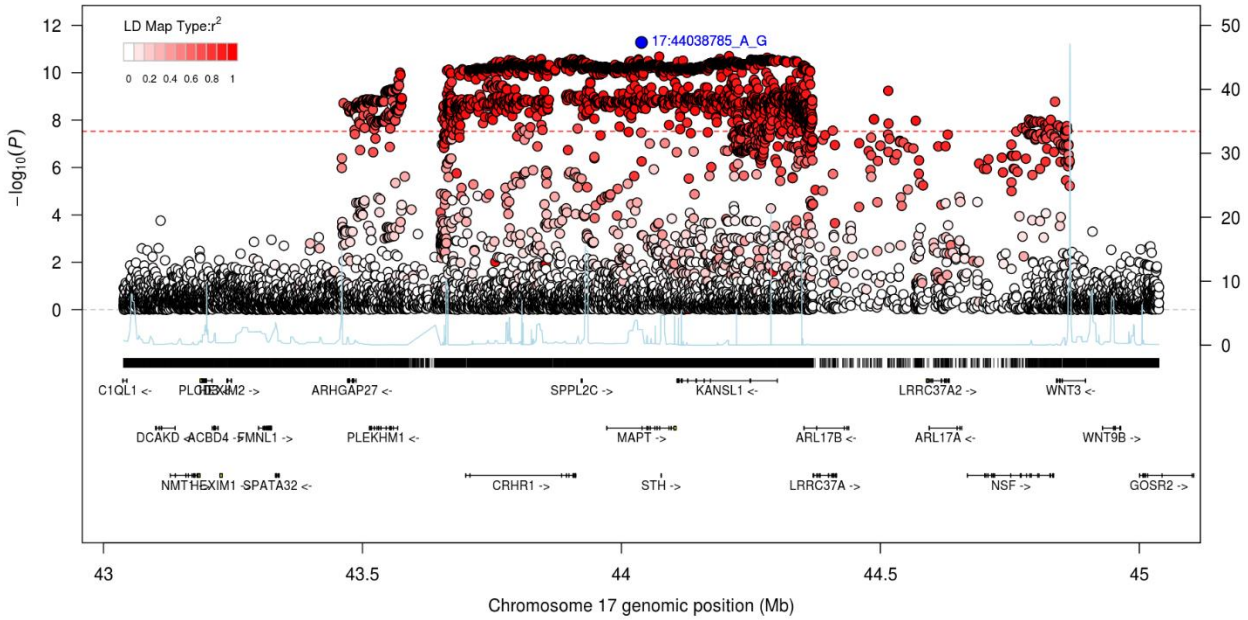

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
11:66501624_T_C
−log10(P)
Chromosome 11 genomic position (Mb)

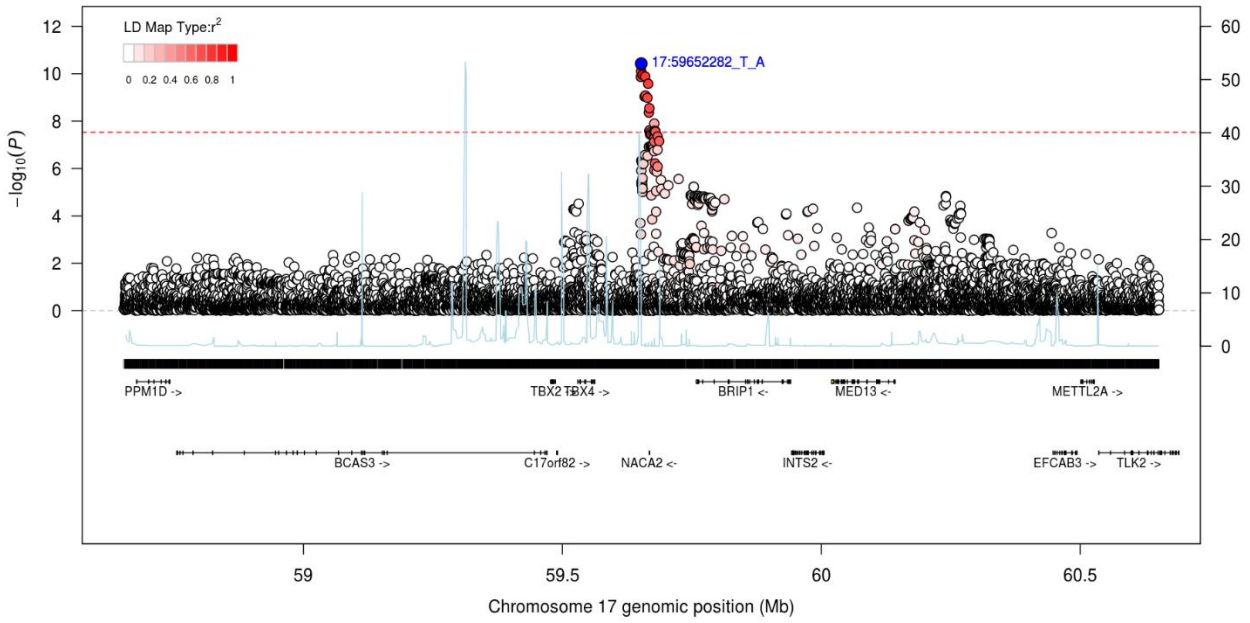

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
12:28014970_G_T
−log10(P)
Chromosome 12 genomic position (Mb)

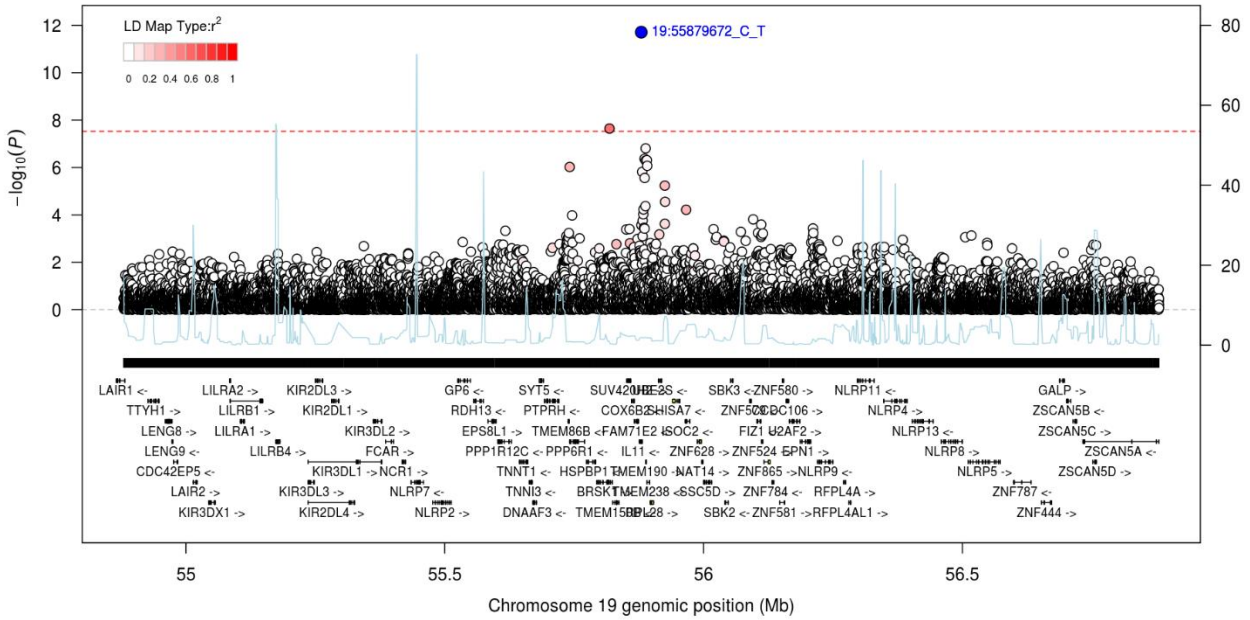

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
12:59289598_A_G
−log10(P)
XRCC6BP1 ->
LRIG3 <-
SLC16A7 ->
58.5
59
59.5
60
Chromosome 12 genomic position (Mb)

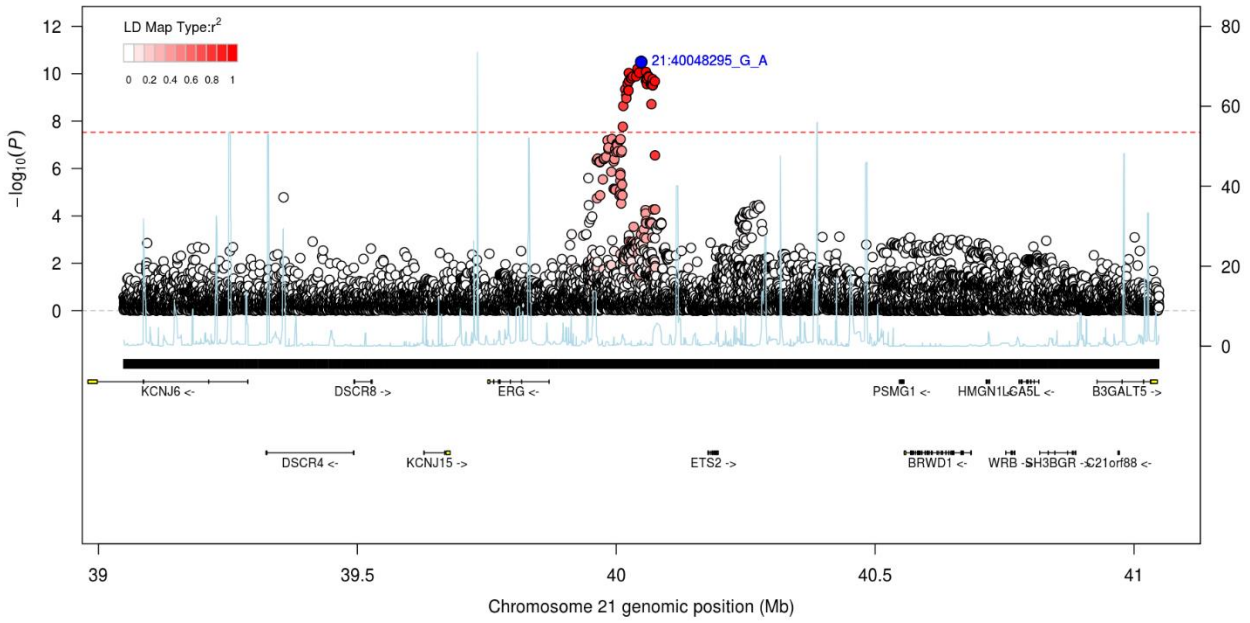

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
12:122606837_C_T
−log10(P)
P2RX7 ->
ANAPC5 <-
ORAI1 ->
RHOF <-
PSMD9 ->
MLXIP ->
B3GNT4 ->
CLIP1 <-
KNTC1 ->
HCAR3 <-
HIP1R ->
OGFOD2 ->
P2RX4 ->
RNF34 ->
MORN3 <-
SETD1B ->
WDR66 ->
ZCCHC8 <-
HCAR1 <-
DENR ->
VPS37B
ARL6IP4 ->
CAMKK2 <-
KDM2B <-
TMEM120B
HPD <-
BCL7A ->
RSRC2 <-
HCAR2 <-
CCDC62 ->
ABCB9 <-
PITPNM2 <-
122
122.5
123
123.5
Chromosome 12 genomic position (Mb)


LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
15:67370506_C_G
−log10(P)
MEGF11 <-
MAP2K1
ILCH ->
SMAD3 ->
C15orf61 ->
PIAS1 ->
DIS3L ->
SNAPC5
GTL <-
AAGAB <-
MAP2K5 ->
TIPIN <-
RPL4 <-
SMAD6 ->
IQCH ->
SKOR1 ->
66.5
67
67.5
68
Chromosome 15 genomic position (Mb)

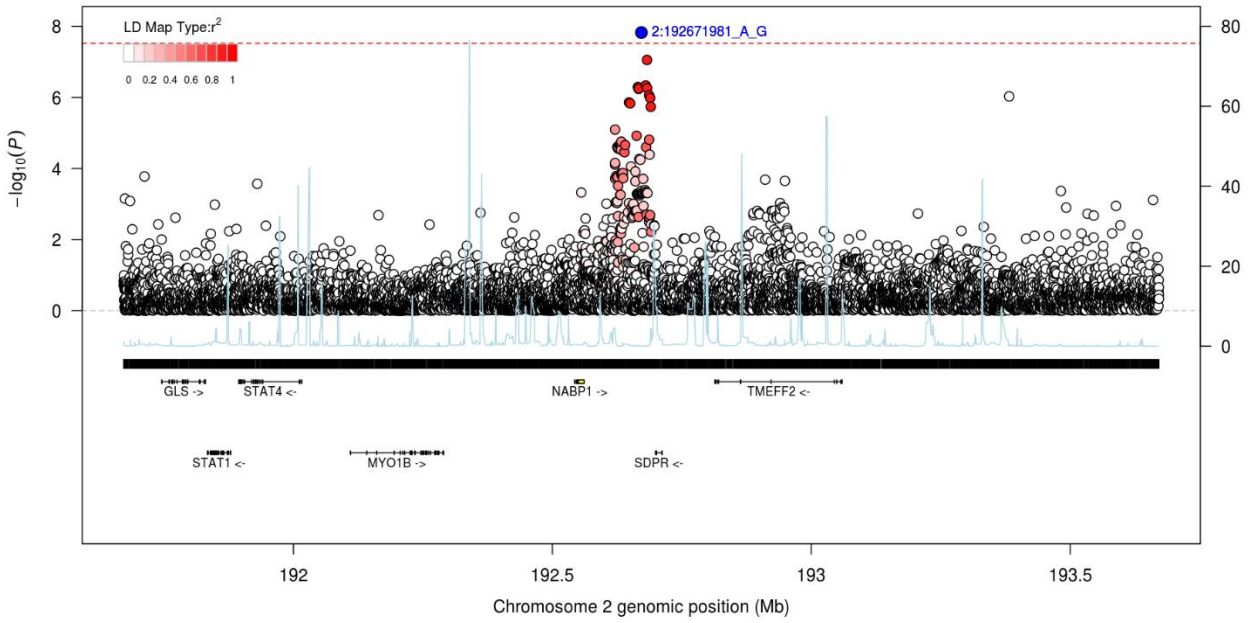

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
17:44038785_A_G
−log10(P)
C1QL1 <-
ARHGAP27 <-
SPPL2C ->
KANSL1 <-
LRRC37A2 ->
WNT3 <-
DCAKD
ACBD4
FMNL1 ->
PLEKHM1 <-
MAPT ->
ARL17B <-
ARL17A <-
WNT9B ->
NMT1
EXIM1
SPATA32 <-
CRHR1 ->
STH ->
LRRC37A ->
NSF ->
GOSR2 ->
43
43.5
44
44.5
45
Chromosome 17 genomic position (Mb)

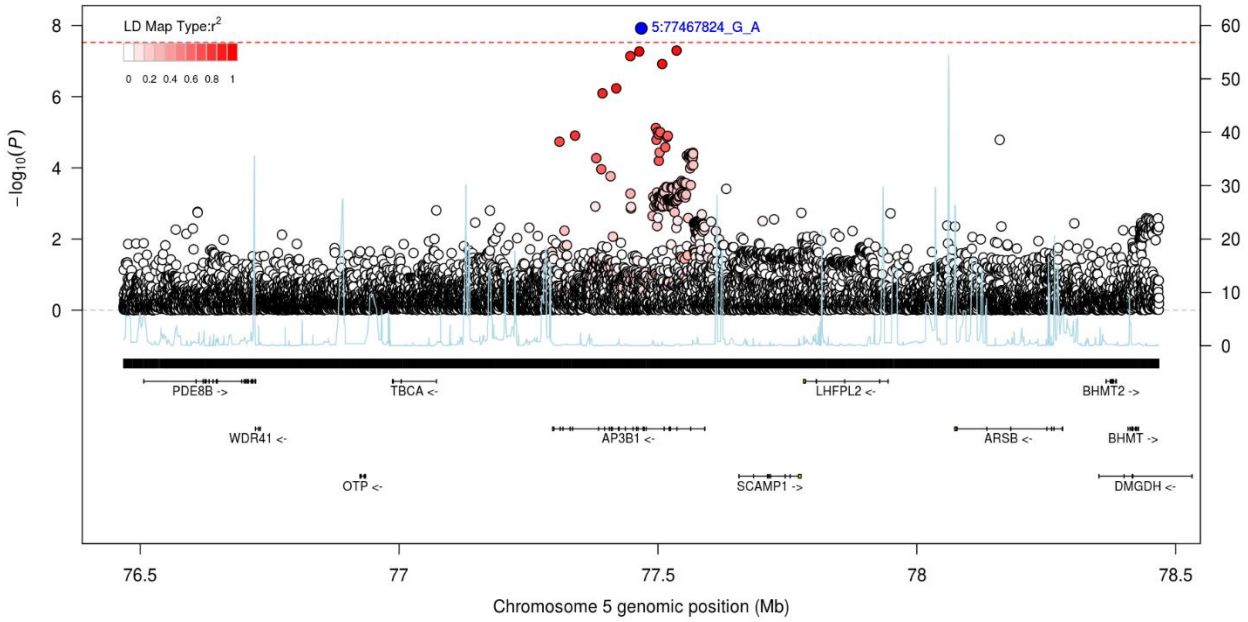

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
17:59652282_T_A
−log10(P)
PPM1D ->
TBX2
TBX4 ->
BRIP1 <-
MED13 <-
METTL2A ->
BCAS3 ->
C17orf82 ->
NACA2 <-
INTS2 <-
EFCAB3 ->
TLK2 ->
Chromosome 17 genomic position (Mb)

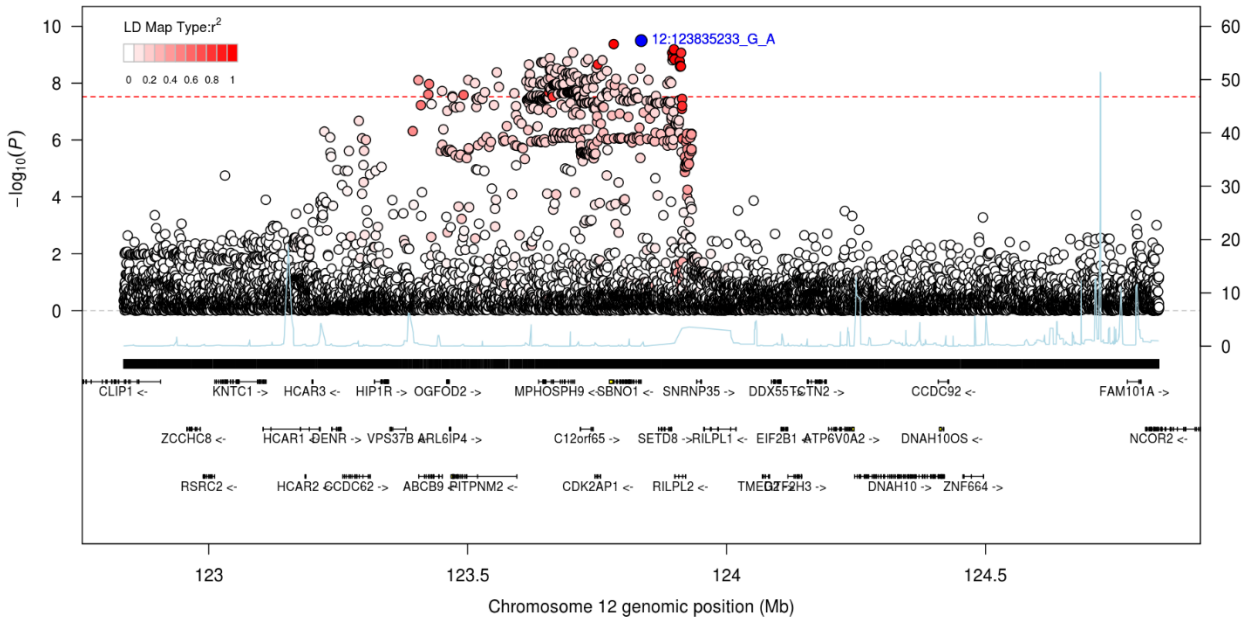

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
19:55879672_C_T
−log10(P)
Chromosome 19 genomic position (Mb)

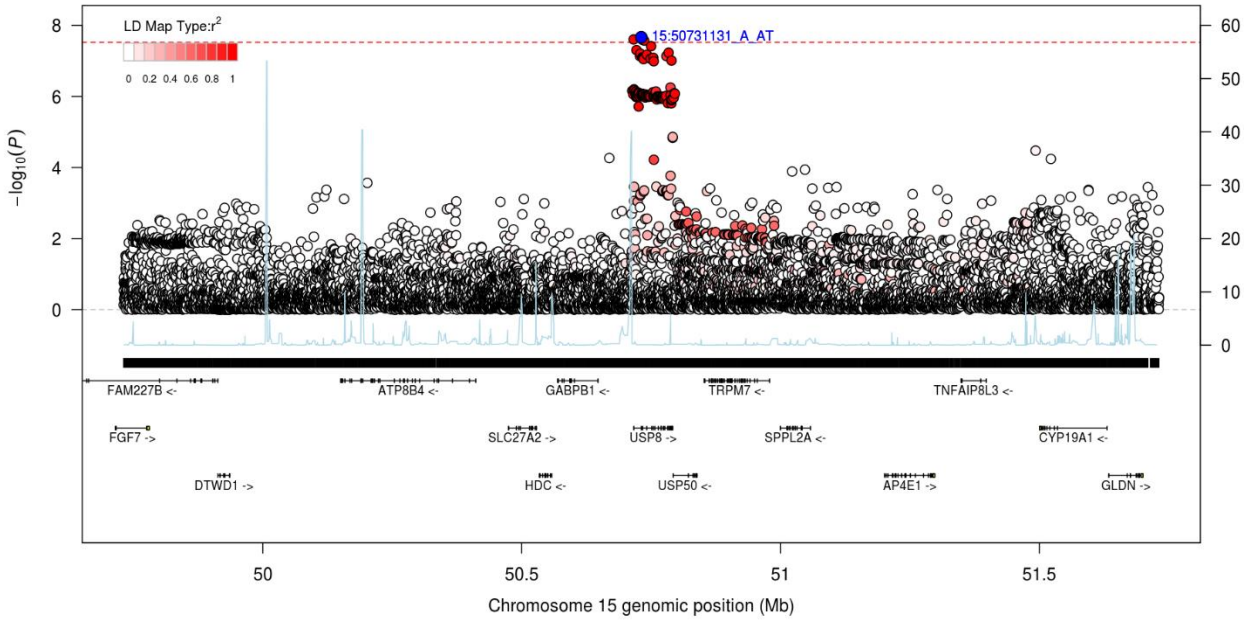

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
21:40048295_G_A
−log10(P)
12
10
8
6
4
2
0
80
60
40
20
0
KCNJ6 <-
DSCR8 ->
ERG <-
PSMG1 <-
HMGN1L
CA5L <-
B3GALT5 ->
DSCR4 <-
KCNJ15 ->
ETS2 ->
BRWD1 <-
WRB
SH3BGR ->
C21orf88 <-
39
39.5
40
40.5
41
Chromosome 21 genomic position (Mb)

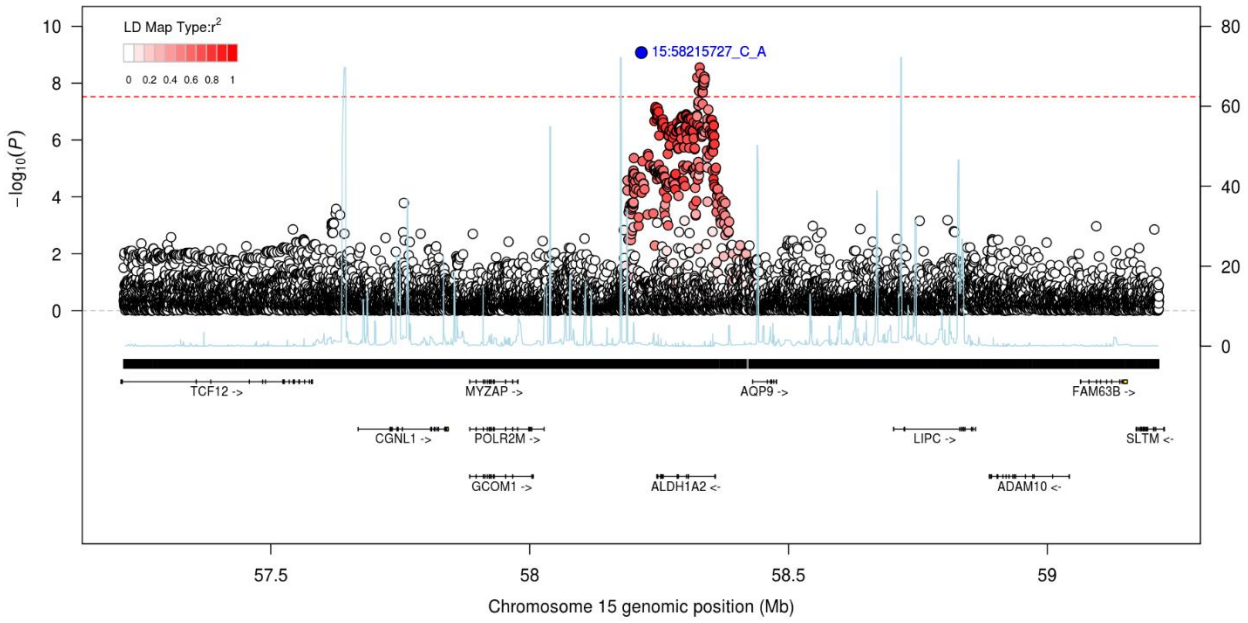
Supplementary Figure 5: Locus zoom plots for 9 knee osteoarthritis loci that reached genome-wide significance in the meta-analysis (as reported in Table 1 and Supplementary Table 3), displaying 1Mb each side of the top variant, where the blue circle represents the index variant on the meta-analysis across UK Biobank and arcOGEN. LD is calculated from UK Biobank.

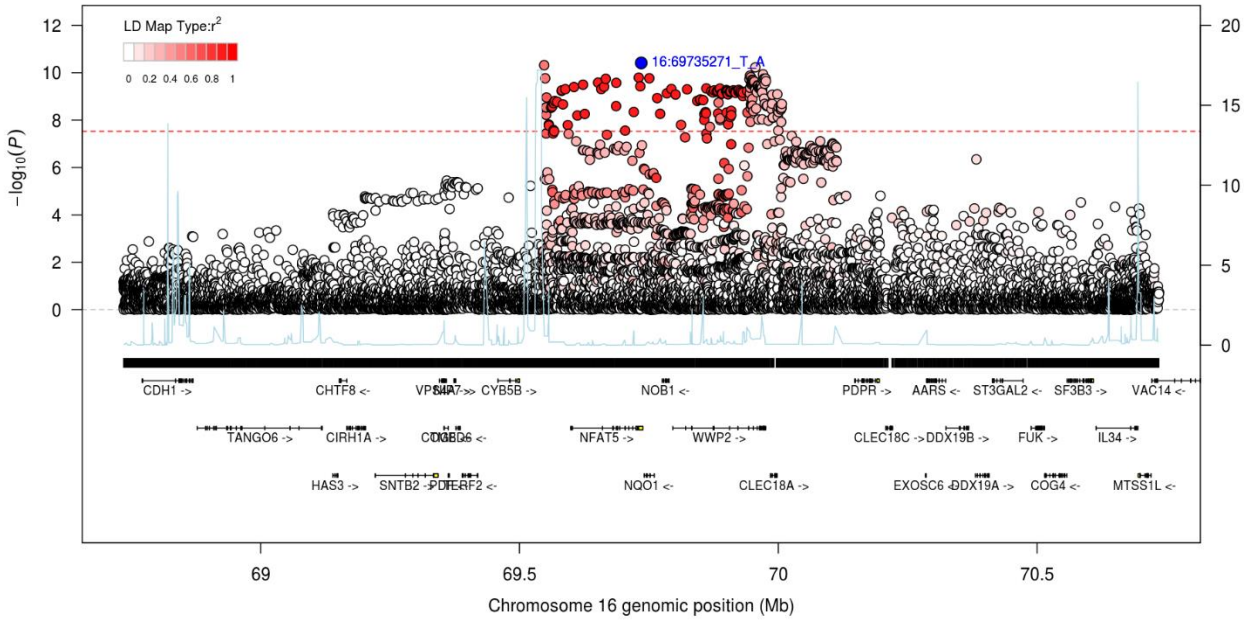


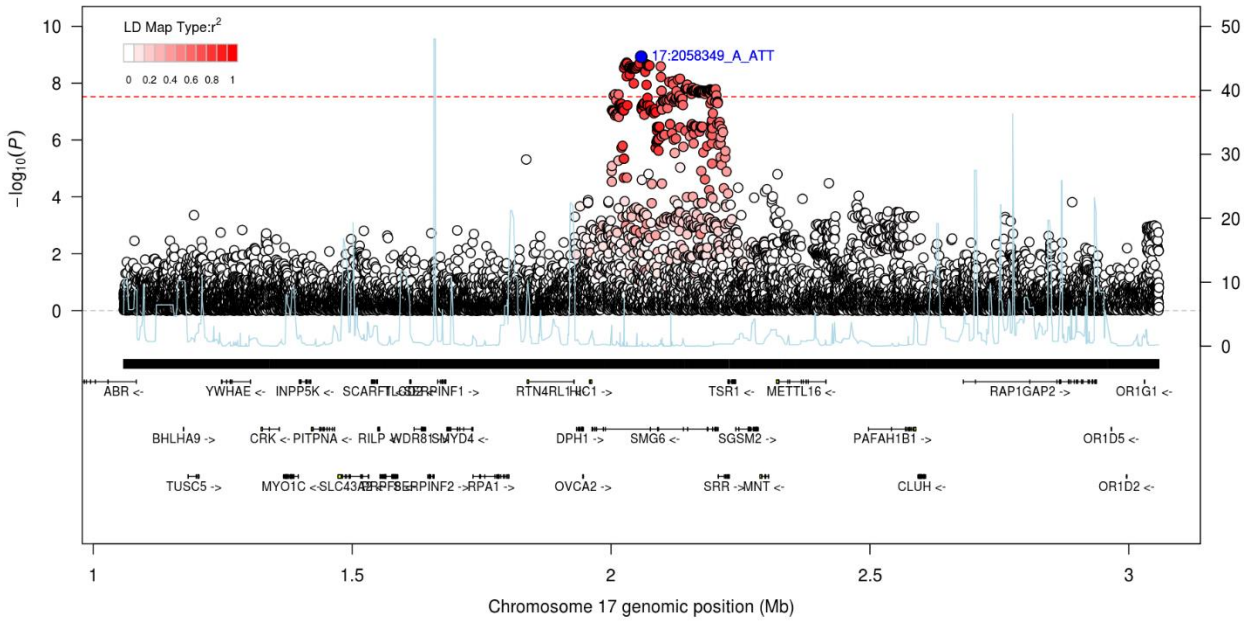


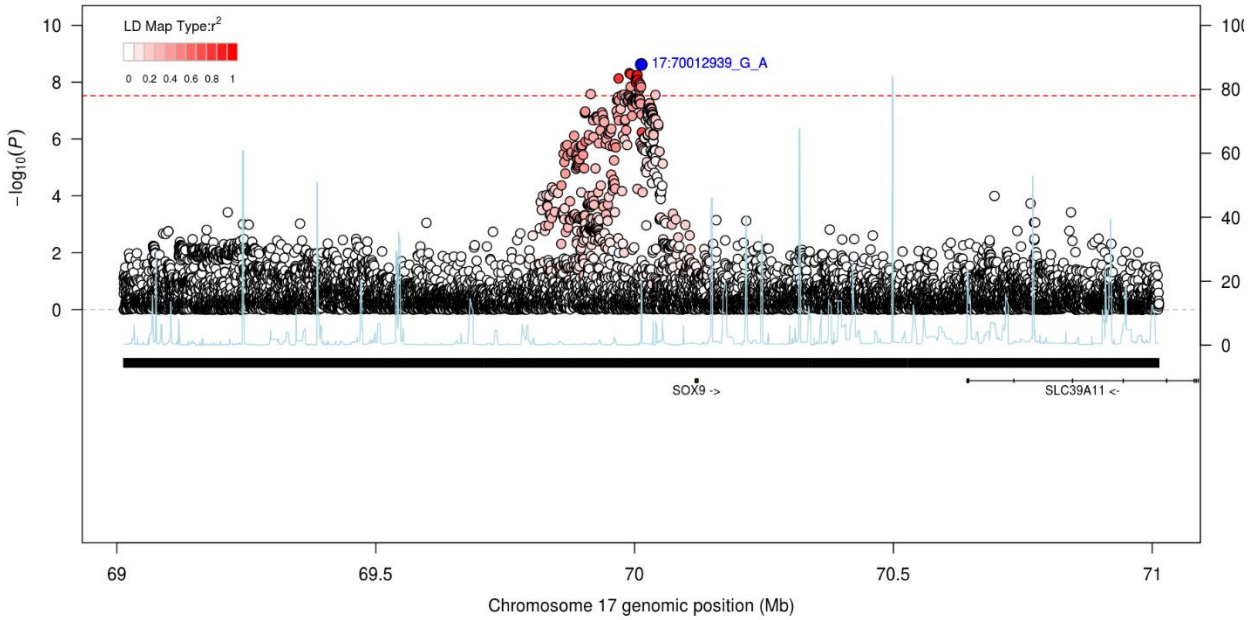

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
12:123835233_G_A
−log10(P)
CLIP1 <-
KNTC1 ->
HCAR3 <-
HIP1R ->
OGFOD2 ->
MPHOSPH9 <-
SBNO1 <-
SNRNP35 ->
DDX55 <-
CTN2 ->
CCDC92 <-
FAM101A ->
ZCCHC8 <-
HCAR1 <-
DENR ->
VPS37B <-
ARL6IP4 ->
C12orf65 ->
SETD8 ->
RILPL1 <-
EIF2B1 <-
ATP6V0A2 ->
DNAH10OS <-
NCOR2 <-
RSRC2 <-
HCAR2 <-
CCDC62 ->
ABCB9 <-
PITPNM2 <-
CDK2AP1 <-
RILPL2 <-
TMEM72 <-
ZFP2H3 ->
DNAH10 ->
ZNF664 ->
123
123.5
124
124.5
Chromosome 12 genomic position (Mb)

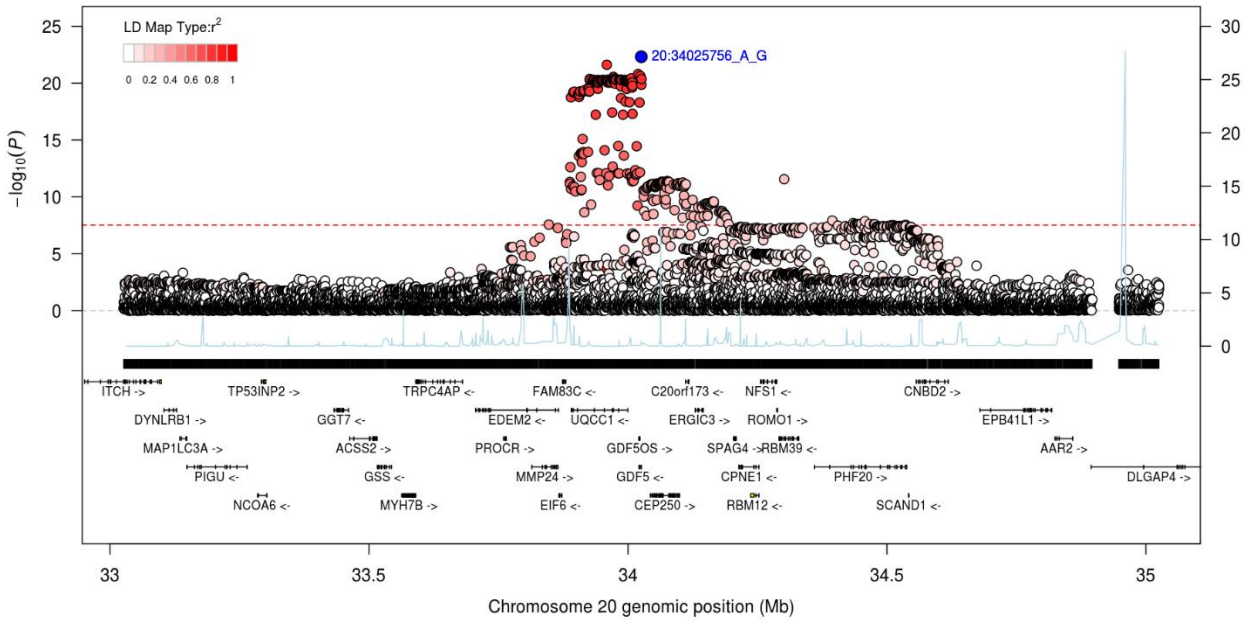

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
15:50731131_A_AT
−log10(P)
FAM227B <-
ATP8B4 <-
GABPB1 <-
TRPM7 <-
TNFAIP8L3 <-
FGF7 ->
SLC27A2 ->
USP8 ->
SPPL2A <-
CYP19A1 <-
DTWD1 ->
HDC <-
USP50 <-
AP4E1 ->
GLDN ->
50
50.5
51
51.5
Chromosome 15 genomic position (Mb)


LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
15:58215727_C_A
−log10(P)
TCF12 ->
MYZAP ->
AQP9 ->
FAM63B ->
CGNL1 ->
POLR2M ->
LIPC ->
SLTM <-
GCOM1 ->
ALDH1A2 <-
ADAM10 <-
57.5
58
58.5
59
Chromosome 15 genomic position (Mb)

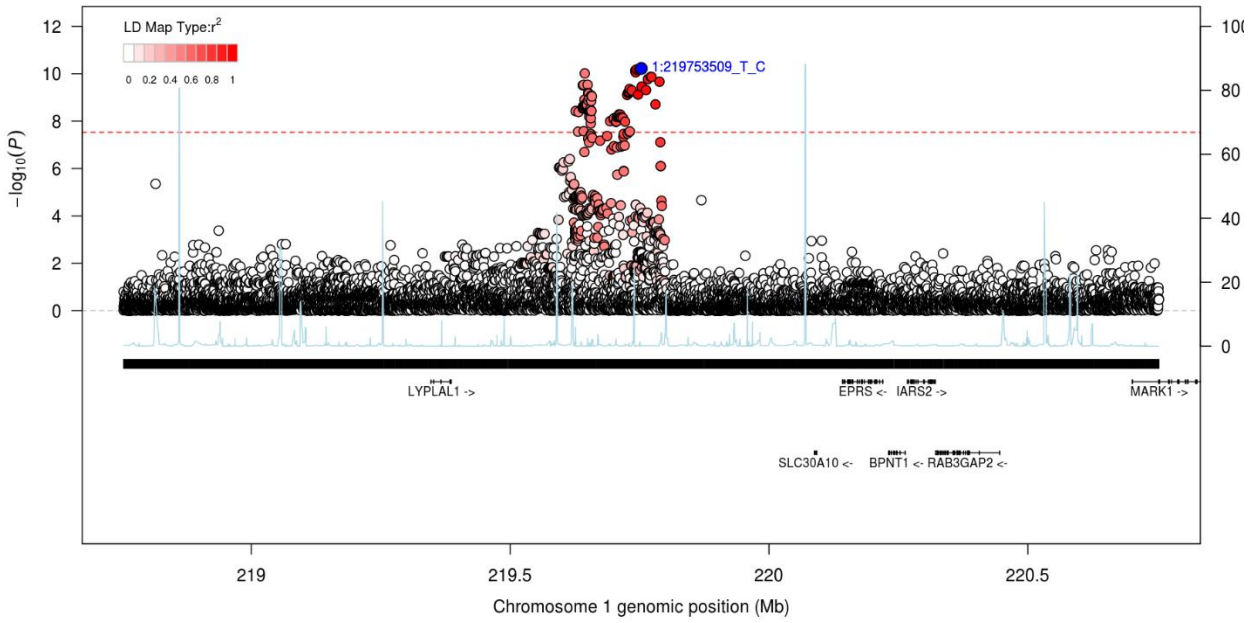

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
16:69735271_T_A
−log10(P)
CDH1 ->
CHTF8 <-
CYB5B ->
NOB1 <-
PDPR ->
AARS <-
ST3GAL2 <-
SF3B3 ->
VAC14 <-
TANGO6 ->
CIRH1A ->
NFAT5 ->
WWP2 ->
CLEC18C ->
DDX19B ->
FUK ->
IL34 ->
HAS3 ->
NQO1 <-
CLEC18A ->
EXOSC6 ->
DDX19A ->
COG4 <-
MTSS1L <-
69
69.5
70
70.5
Chromosome 16 genomic position (Mb)

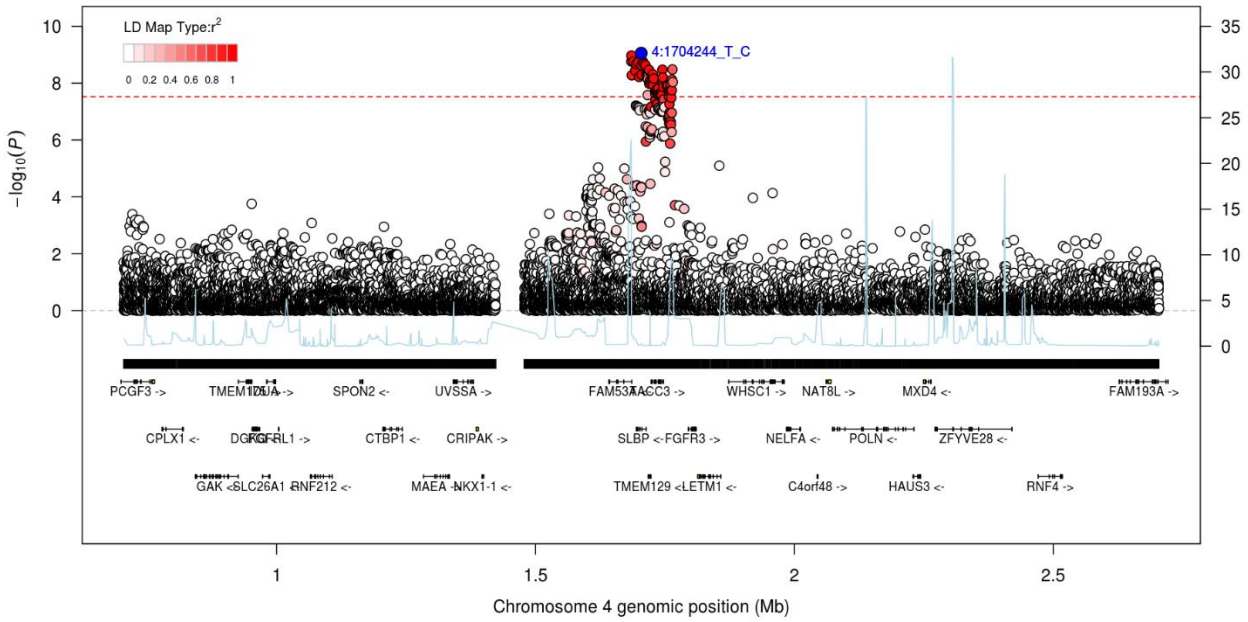

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
17:2058349_A_ATT
−log10(P)
ABR <-
YWHAE <-
INPP5K <-
RTN4RL1
TSR1 <-
METTL16 <-
RAP1GAP2 ->
OR1G1 <-
BHLHA9 ->
CRK <-
PITPNA <-
RILP
DPH1 ->
SMG6 <-
SGSM2 ->
PAFAH1B1 ->
OR1D5 <-
TUSC5 ->
MYO1C <-
OVCA2 ->
SRR ->
MNT <-
CLUH <-
OR1D2 <-
Chromosome 17 genomic position (Mb)

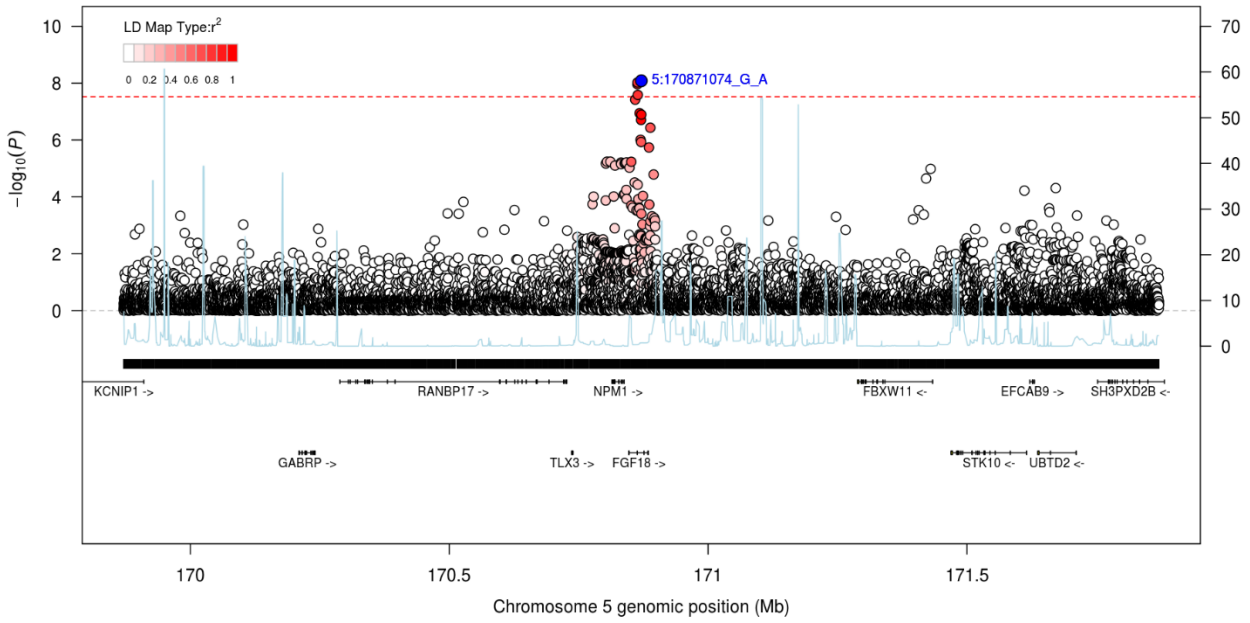

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
17:70012939_G_A
−log10(P)
SOX9 ->
SLC39A11 <-
Chromosome 17 genomic position (Mb)

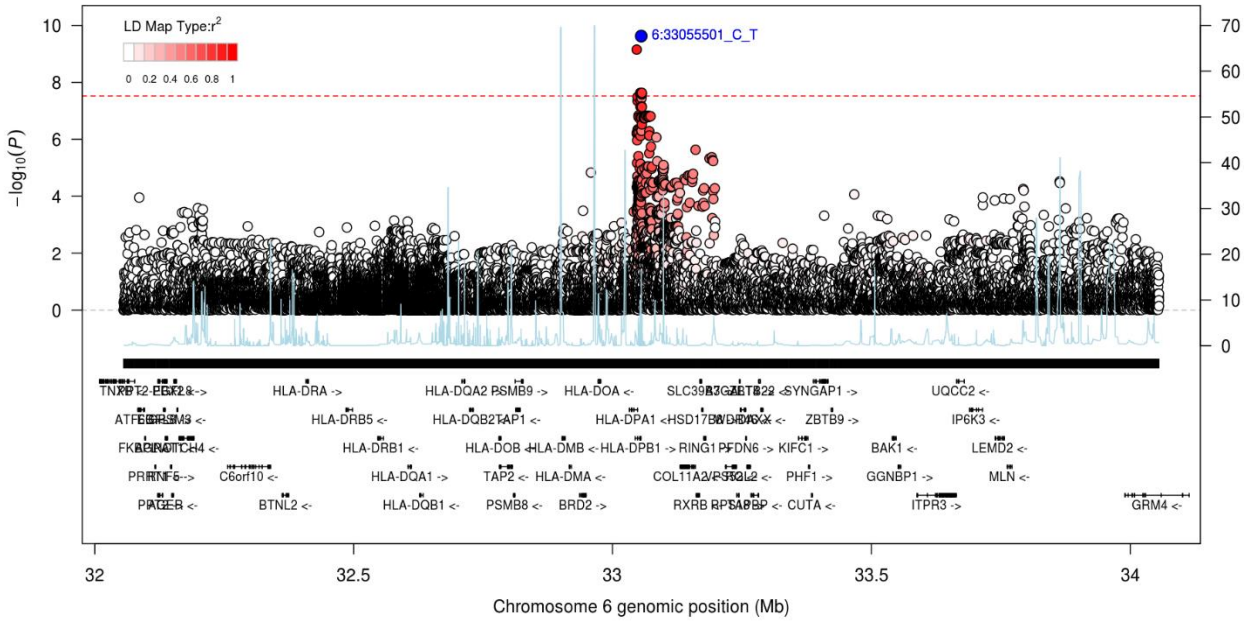

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
20:34025756_A_G
−log10(P)
25
20
15
10
5
0
30
25
20
15
10
5
0
ITCH ->
TP53INP2 ->
TRPC4AP <-
FAM83C <-
C20orf173 <-
NFS1 <-
CNBD2 ->
DYNLRB1 ->
GGT7 <-
EDEM2 <-
UQCC1 <-
ERGIC3 ->
ROMO1 ->
EPB41L1 ->
MAP1LC3A ->
ACSS2 ->
PROCR ->
GDF5OS ->
SPAG4 ->
RBM39 <-
AAR2 ->
PIGU <-
GSS <-
MMP24 ->
GDF5 <-
CPNE1 <-
PHF20 ->
DLGAP4 ->
NCOA6 <-
MYH7B ->
EIF6 <-
CEP250 ->
RBM12 <-
SCAND1 <-
33
33.5
34
34.5
35
Chromosome 20 genomic position (Mb)

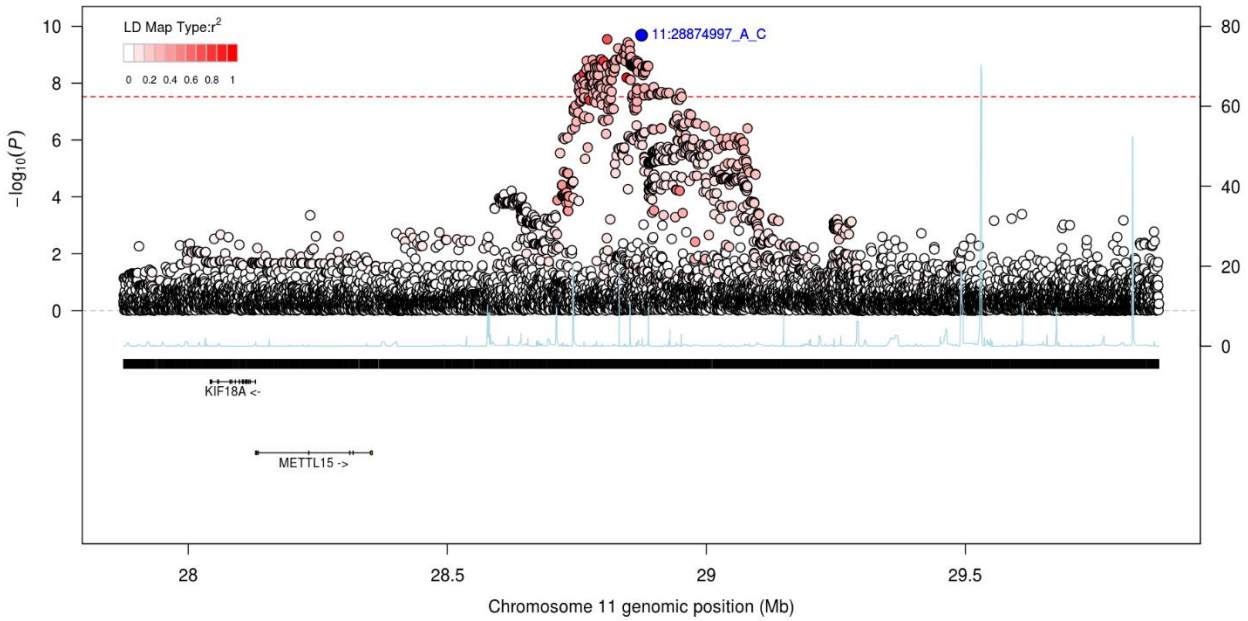
Supplementary Figure 6: Locus zoom plots for 8 knee and/or hip osteoarthritis loci that reached genome-wide significance in the meta-analysis (as reported in Table 1 and Supplementary Table 3), displaying 1Mb each side of the top variant, where the blue circle represents the index variant on the meta-analysis across UK Biobank and arcOGEN. LD is calculated from UK Biobank.

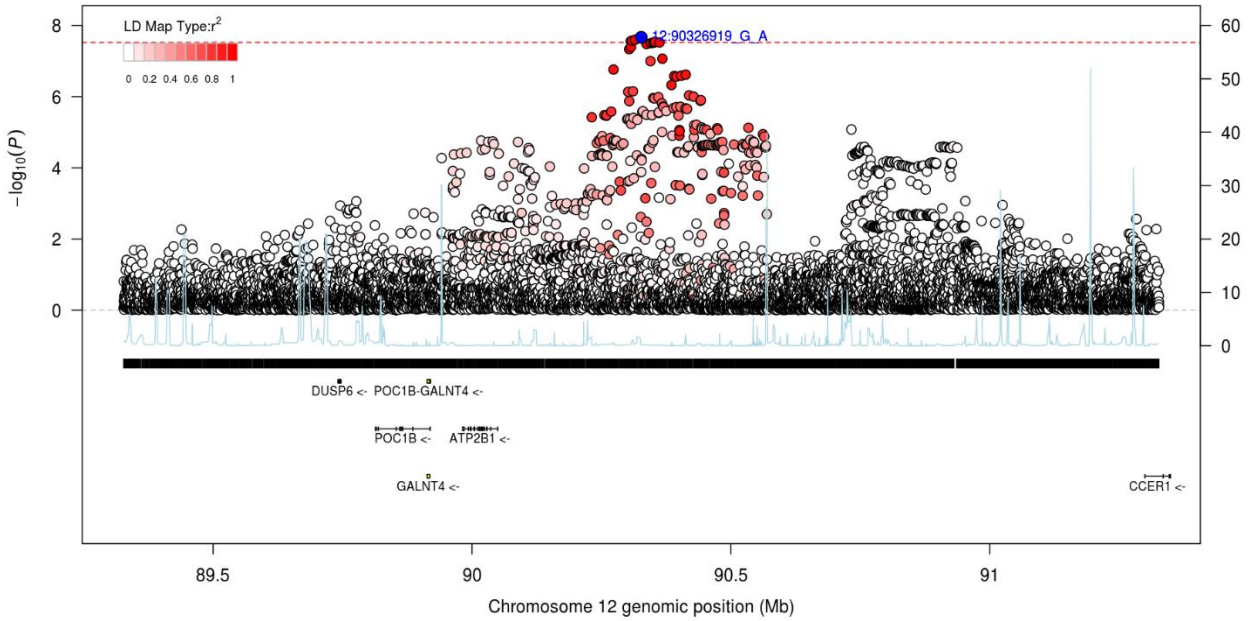


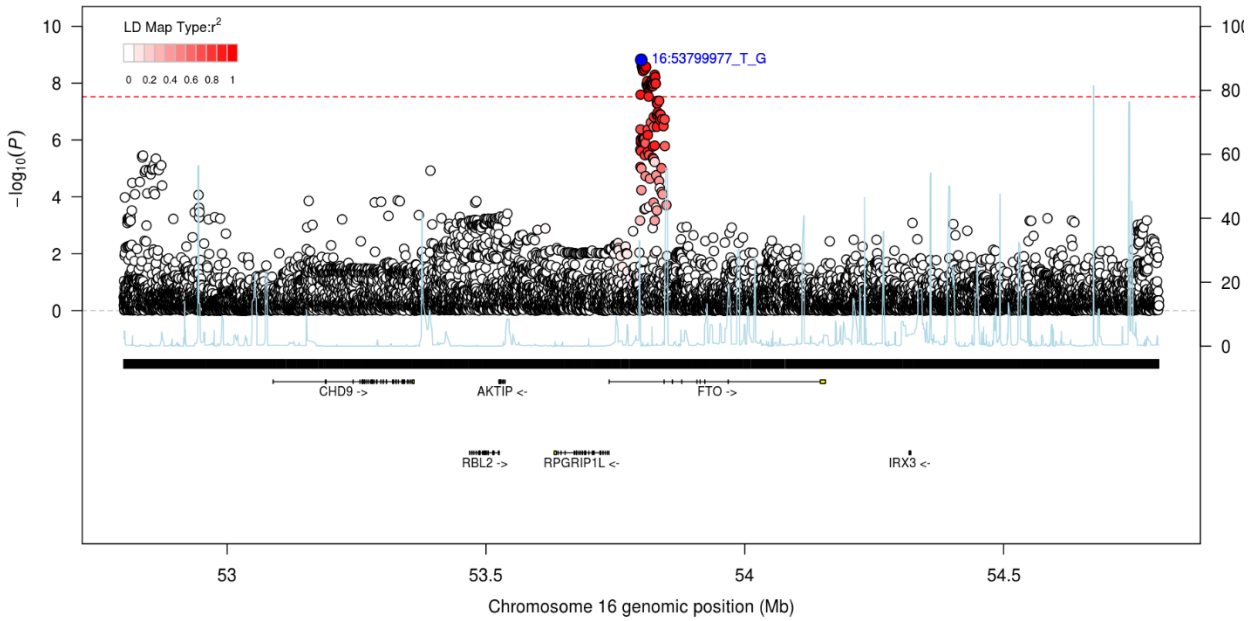


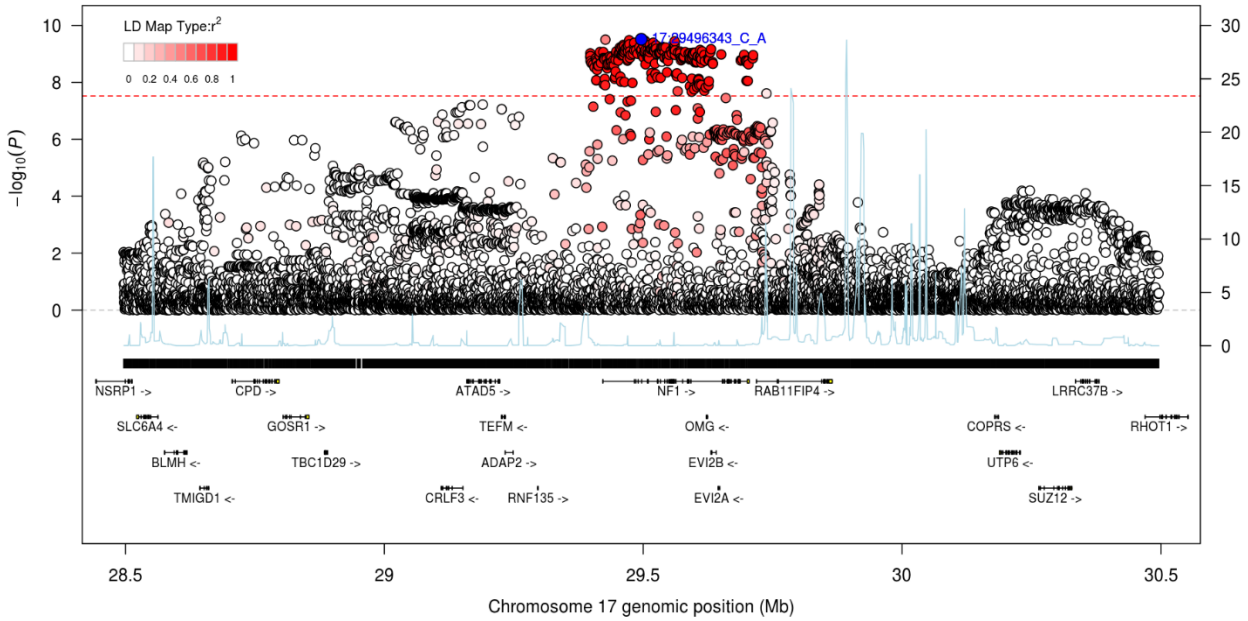

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
5:170871074_G_A
−log10(P)
KCNIP1 ->
RANBP17 ->
NPM1 ->
FBXW11 <-
EFCAB9 ->
SH3PXD2B <-
GABRP ->
TLX3 ->
FGF18 ->
STK10 <-
UBTD2 <-
170
170.5
171
171.5
Chromosome 5 genomic position (Mb)


LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
6:33055501_C_T
−log10(P)
HLA-DRA ->
HLA-DQA2
PSMB9 ->
HLA-DOA <-
SYNGAP1 ->
UQCC2 <-
HLA-DRB5 <-
HLA-DQB2
TAP1 <-
HLA-DPA1 <-
ZBTB9 ->
IP6K3 <-
HLA-DRB1 <-
HLA-DOB
HLA-DMB <-
HLA-DPB1 ->
RING1
PFDN6 ->
KIFC1 ->
BAK1 <-
LEMD2 <-
C6orf10 <-
HLA-DQA1 ->
TAP2 <-
HLA-DMA <-
PHF1 ->
GGNBP1 ->
MLN <-
BTNL2 <-
HLA-DQB1 <-
PSMB8 <-
BRD2 ->
CUTA <-
ITPR3 ->
GRM4 <-
32
32.5
33
33.5
34
Chromosome 6 genomic position (Mb)

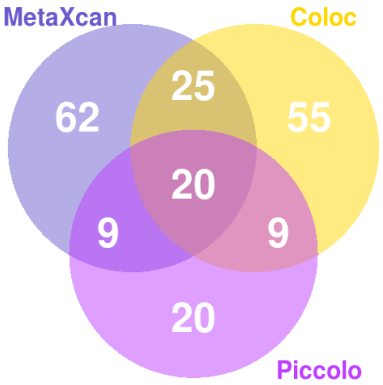

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
11:28874997_A_C
−log10(P)
KIF18A <-
METTL15 ->
28
28.5
29
29.5
Chromosome 11 genomic position (Mb)


LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
12:90326919_G_A
−log10(P)
DUSP6 <-
POC1B-GALNT4 <-
POC1B <-
ATP2B1 <-
GALNT4 <-
CCER1 <-
89.5
90
90.5
91
Chromosome 12 genomic position (Mb)


LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
16:53799977_T_G
−log10(P)
CHD9 ->
AKTIP <-
FTO ->
RBL2 ->
RPGRIP1L <-
IRX3 <-
53
53.5
54
54.5
Chromosome 16 genomic position (Mb)

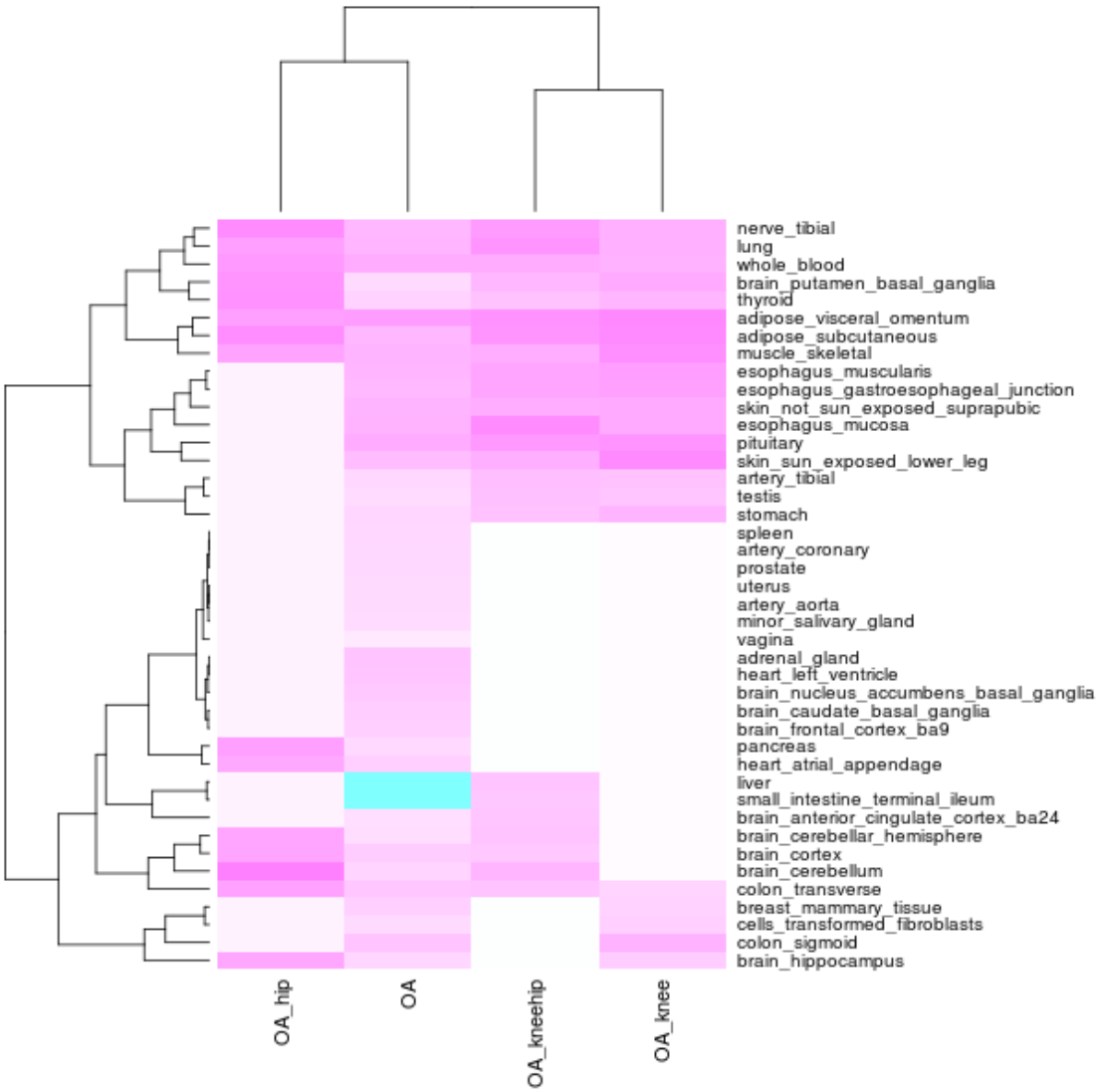

LD Map Type:r2
0 0.2 0.4 0.6 0.8 1
17:29496343_C_A
−log10(P)
NSRP1 ->
CPD ->
ATAD5 ->
NF1 ->
RAB11FIP4 ->
LRRC37B ->
SLC6A4 <-
GOSR1 ->
TEFM <-
OMG <-
COPRS <-
RHOT1 ->
BLMH <-
TBC1D29 ->
ADAP2 ->
EVI2B <-
UTP6 <-
TMIGD1 <-
CRLF3 <-
RNF135 ->
EVI2A <-
SUZ12 ->
28.5
29
29.5
30
30.5
Chromosome 17 genomic position (Mb)

Supplementary Figure 7: Overlap of genes implicated by different colocalisation analyses. Gene overlap from associations between the osteoarthritis traits and expression levels in at least one human tissue using MetaXcan, Coloc and Picolo.

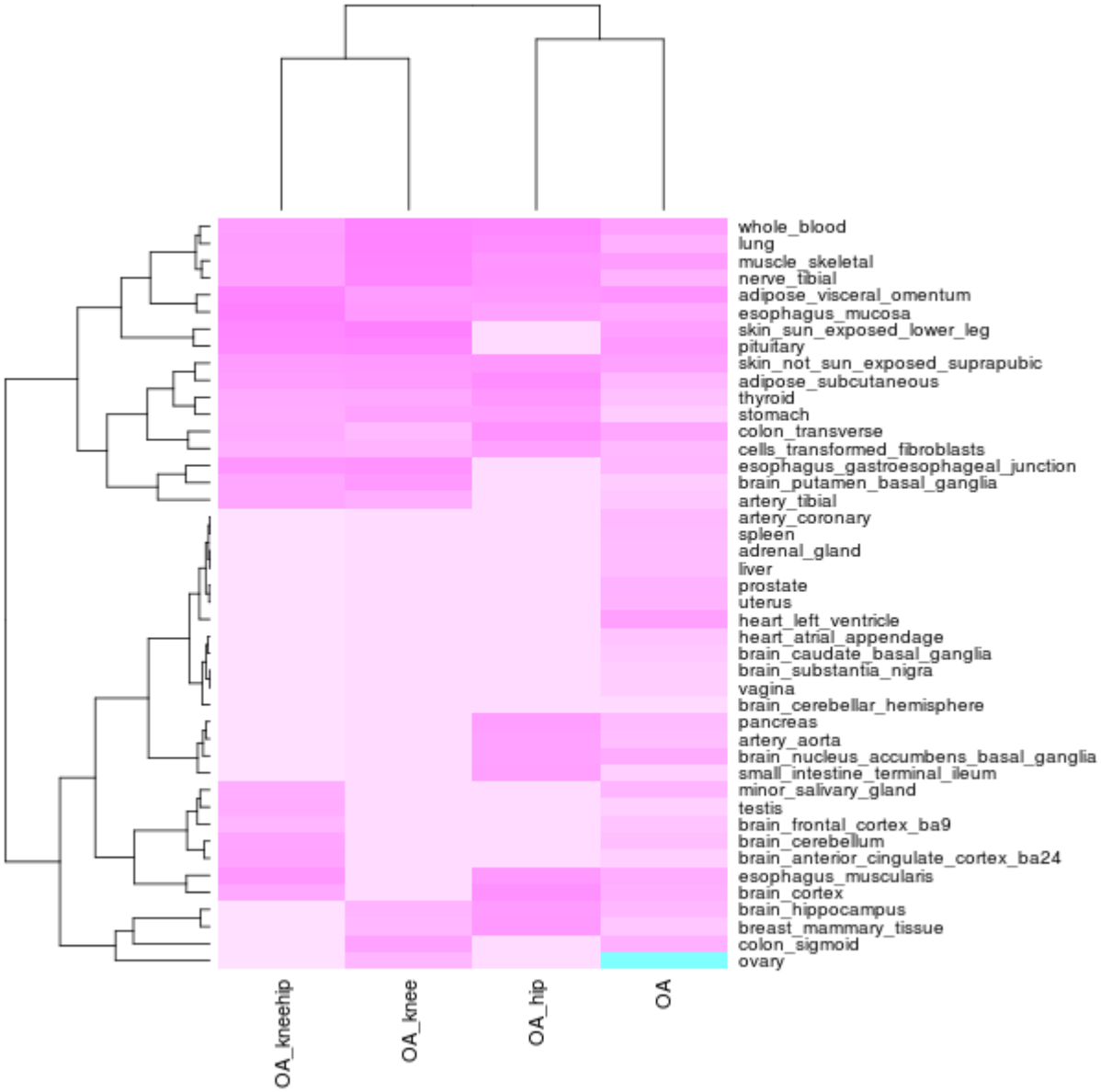


Supplementary Figure 8: Average enrichment of genes by tissue. We display the enrichment for the osteoarthritis phenotypes measured as the mean of the square of the Z-scores (effect size/standard error) of the association between the genetic component of gene expression and phenotype, as calculated by MetaXcan. We average the Z^2 for only the Bonferroni corrected significant gene-tissue pairs (A), and we also further filter them to have probability of colocalisation of at least 80% (B). Pink and blue shades represent positive and negative effect sizes, respectively, while higher colour intensity indicates more significant association.

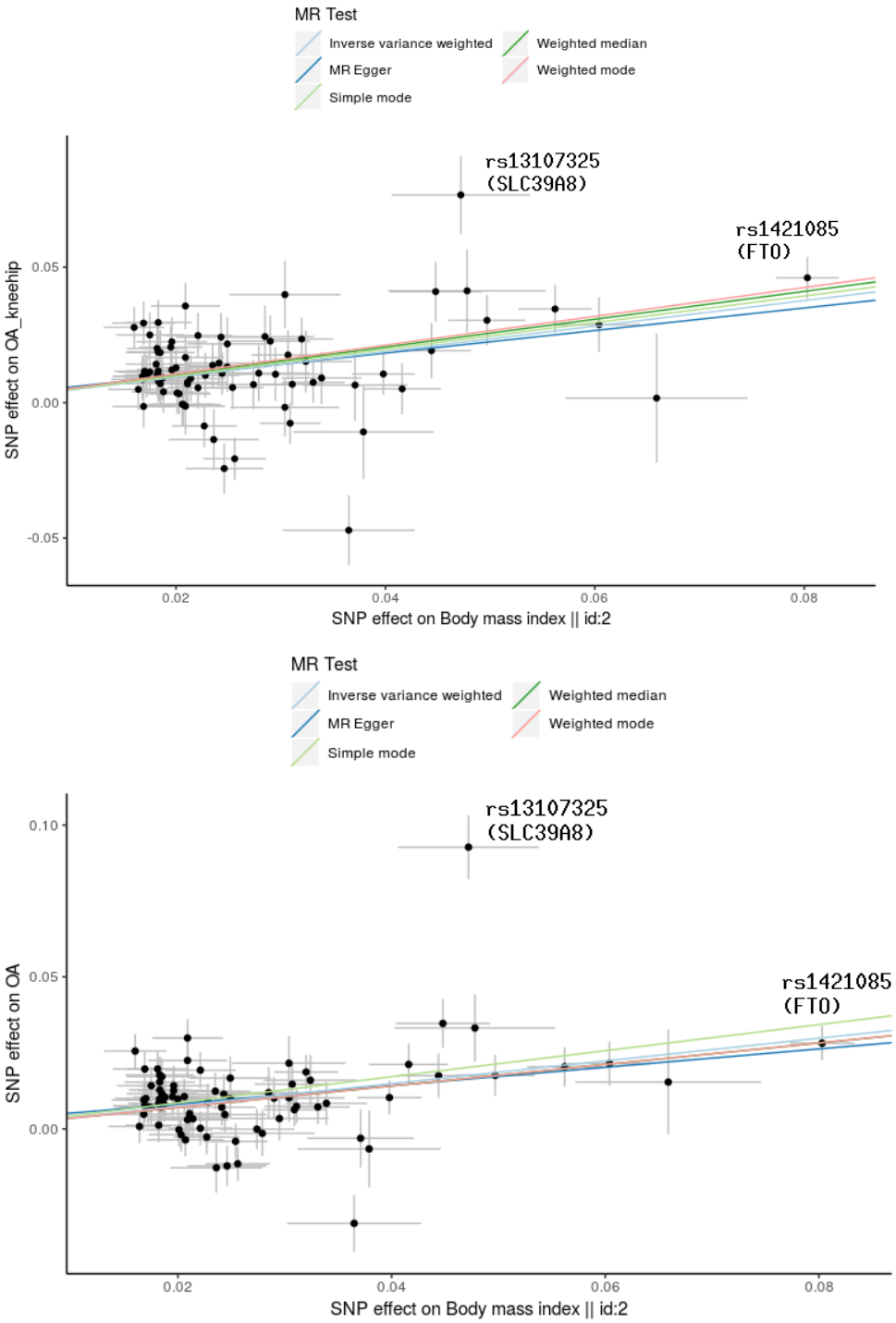
A



B


whole_blood
lung
muscle_skeletal
nerve_tibial
adipose_visceral_omentum
esophagus_mucosa
skin_sun_exposed_lower_leg
pituitary
skin_not_sun_exposed_suprapubic
adipose_subcutaneous
thyroid
stomach
colon_transverse
cells_transformed_fibroblasts
esophagus_gastroesophageal_junction
brain_putamen_basal_ganglia
artery_tibial
artery_coronary
spleen
adrenal_gland
liver
prostate
uterus
heart_left_ventricle
heart_atrial_appendage
brain_caudate_basal_ganglia
brain_substantia_nigra
vagina
brain_cerebellar_hemisphere
pancreas
artery_aorta
brain_nucleus_accumbens_basal_ganglia
small_intestine_terminal_ileum
minor_salivary_gland
testis
brain_frontal_cortex_ba9
brain_cerebellum
brain_anterior_cingulate_cortex_ba24
esophagus_muscularis
brain_cortex
brain_hippocampus
breast_mammary_tissue
colon_sigmoid
ovary
OA_kneehip
OA_knee
OA_hip
OA

Supplementary Figure 9: Effect sizes of each BMI instrument on BMI and osteoarthritis is shown. Two BMI-associated loci are genome-wide significantly associated with osteoarthritis (*SLC39A8* and *FTO*), and while the associations of *FTO* with osteoarthritis are of the magnitude expected given the causal effects of BMI on osteoarthritis, we find that *SLC39A8* missense SNV has a far larger effect on osteoarthritis than predicted based on its BMI-raising effects.



Supplementary Table 1: Genetic correlation across the 4 osteoarthritis trait definitions using genome-wide summary statistics of common variants from the UK Biobank cohort.

	Osteoarthritis	Knee osteoarthritis	Hip osteoarthritis	Hip and/or knee osteoarthritis
Osteoarthritis	1	0.914	0.678	0.947
Knee osteoarthritis		1	0.498	0.935
Hip osteoarthritis			1	0.772
Hip and/or knee osteoarthritis				1

Supplementary Table 3: **Individual cohort summary statistics for the independent variants with *P*<3x10^{-8} in an inverse-variance weighted fixed effects meta-analysis of UK Biobank and arcOGEN.** Variant positions are reported according to build 37 and their alleles are coded based on the positive strand.

RSID	Most Severe Consequence	EA	NEA	UK Biobank		arcOGEN		EAF						
				OR	PV	OR	PV	UK Biobank		arcOGEN		1000 Genomes		
								Cases	Controls	Cases	Controls	GBR	L95	U95
Newly Identified Loci														
rs4338381	sequence feature	A	G	1.1	4.70E-13	1.1	2.71E-03	0.653	0.631	0.654	0.632	0.62	0.55	0.69
1:150214028	downstream gene	C	CT	1.03	2.50E-08			0.368	0.36					
1:174192402	intron	TAAAAAAAAAAAAAA	T	1.03	1.10E-08			0.586	0.576					
rs11583641	3 prime UTR	C	T	1.09	3.10E-09	1.07	5.55E-02	0.74	0.725	0.735	0.722	0.68	0.61	0.74
rs10218792	intron	G	T	1.04	7.40E-08	1.04	5.13E-02	0.274	0.268	0.274	0.266	0.2	0.15	0.27
rs2061027	intron	A	G	1.04	9.70E-12	1.07	6.58E-03	0.514	0.504	0.518	0.501	0.55	0.47	0.62
rs12470967	intron	A	G	1.06	1.50E-08			0.442	0.429			0.55	0.48	0.63
rs62182810	intron	A	G	1.03	6.20E-09	1.04	5.44E-02	0.554	0.546	0.556	0.547	0.52	0.45	0.6
rs62262139	intron	A	G	1.04	9.10E-11			0.544	0.535			0.54	0.47	0.62
rs11732213	intron	T	C	1.06	2.30E-09	1.04	1.46E-01	0.814	0.805	0.816	0.809	0.73	0.66	0.79
rs1913707	intergenic region	A	G	1.07	9.00E-08	1.16	4.24E-06	0.627	0.611	0.642	0.608	0.59	0.52	0.66
rs34811474	missense	G	A	1.04	3.70E-09	1.03	1.93E-01	0.775	0.767	0.787	0.782	0.71	0.64	0.77
rs13107325	missense	T	C	1.1	4.10E-18	1.08	5.14E-02	0.08	0.074	0.087	0.081	0.1	0.06	0.15
rs35611929	intron	A	G	1.06	5.50E-08	1.06	7.93E-02	0.354	0.341	0.342	0.33	0.32	0.25	0.39
rs3884606	intron	G	A	1.04	2.10E-07	1.06	3.43E-03	0.498	0.487	0.5	0.484	0.47	0.39	0.54
rs115740542	upstream gene	C	T	1.06	2.00E-08	1.06	6.72E-02	0.078	0.073	0.074	0.07	0.05	0.02	0.09

rs9277552	downstream gene	C	T	1.06	9.30E-10	1.05	1.07E-01	0.796	0.786	0.825	0.817	0.7	0.62	0.76
rs12154055	intergenic region	G	A	1.03	6.90E-09	0.99	8.95E-01	0.619	0.612	0.611	0.613	0.66	0.59	0.73
rs80287694	intron	G	A	1.13	2.10E-09	1.05	2.64E-01	0.125	0.113	0.116	0.111	0.15	0.1	0.22
rs11409738	intron	TA	T	1.04	2.10E-10			0.375	0.366			0.39	0.32	0.47
rs330050	upstream gene	G	C	1.04	2.90E-10	1.06	9.51E-01	0.515	0.505	0.514	0.5	0.45	0.37	0.52
rs60890741	upstream gene	C	CA	1.11	4.50E-09			0.88	0.867			0.87	0.81	0.91
rs919642	intergenic region	T	A	1.05	1.10E-14	1.03	1.96E-01	0.27	0.26	0.277	0.272	0.27	0.21	0.35
rs1330349	upstream gene	C	G	1.08	2.40E-09	1.09	3.93E-03	0.593	0.574	0.599	0.577	0.6	0.53	0.68
rs62578127	intron	C	T	1.09	2.00E-11	1.07	2.15E-02	0.651	0.631	0.655	0.64	0.6	0.52	0.67
rs17659798	intron	A	C	1.05	2.50E-09	1.06	2.14E-02	0.725	0.714	0.727	0.715	0.69	0.61	0.75
rs11031191	intergenic region	T	G	1.03	3.30E-08	1.03	9.65E-02	0.354	0.347	0.355	0.348	0.36	0.29	0.44
rs10896015	upstream gene	G	A	1.08	2.30E-07	1.11	2.13E-03	0.745	0.73	0.749	0.728	0.69	0.62	0.76
rs34419890	upstream gene	T	C	1.13	2.70E-07	1.16	8.15E-03	0.933	0.925	0.94	0.931	0.93	0.88	0.96
rs1149620	upstream gene	T	A	1.04	2.50E-09	1.04	8.13E-02	0.58	0.571	0.576	0.567	0.58	0.51	0.65
rs79056043	intron	G	A	1.16	9.50E-07	1.28	6.44E-05	0.051	0.044	0.065	0.052	0.06	0.03	0.11
rs317630	intron	T	C	1.03	7.50E-08	1.04	1.79E-01	0.279	0.273	0.284	0.276	0.27	0.21	0.34
rs11105466	intron	A	G	1.04	8.30E-07	1.07	2.37E-03	0.415	0.415	0.429	0.412	0.41	0.34	0.48
rs2171126	intron	T	C	1.03	1.30E-08	1.06	6.78E-03	0.514	0.506	0.519	0.504	0.51	0.44	0.59
rs11059094	intron	T	C	1.07	9.80E-09	1.1	9.54E-04	0.494	0.477	0.497	0.472	0.45	0.37	0.52
rs56116847	upstream gene	A	G	1.07	4.40E-11	1	9.37E-01	0.369	0.354	0.364	0.363	0.35	0.28	0.43
rs35912128	upstream gene	AT	A	1.08	2.20E-08			0.172	0.161			0.25	0.19	0.32

rs35206230	downstream gene	T	C	1.04	8.10E-12	1.05	6.44E-02	0.68	0.67	0.673	0.662	0.68	0.61	0.74
rs6499244	3 prime UTR	A	T	1.06	6.40E-10	1.07	5.55E-03	0.427	0.441	0.577	0.559	0.52	0.45	0.6
rs1126464	missense	G	C	1.04	3.00E-11	0.99	8.88E-01	0.764	0.757	0.763	0.764	0.73	0.66	0.8
rs35087650	intron	ATT	A	1.07	1.20E-09			0.255	0.241			0.27	0.21	0.35
rs2953013	intron	C	A	1.05	1.90E-09	1.05	4.36E-02	0.305	0.214	0.307	0.296	0.27	0.21	0.34
rs62063281	intron	G	A	1.1	1.60E-10	1.1	9.13E-03	0.236	0.22	0.249	0.232	0.25	0.19	0.32
rs547116051	upstream gene	AC	A	1.83	1.50E-08			0.0006	0.0004			0		
rs7222178	intergenic region	A	T	1.1	2.60E-10	1.08	1.23E-02	0.213	0.197	0.205	0.193	0.18	0.13	0.25
rs8067763	intergenic region	G	A	1.06	2.50E-09	1.03	3.83E-01	0.418	0.404	0.413	0.406	0.44	0.37	0.51
rs10502437	intron	G	A	1.03	3.60E-08	1.02	3.39E-01	0.61	0.604	0.61	0.604	0.58	0.51	0.65
rs1560707	upstream gene	T	G	1.04	1.70E-13	1.02	1.91E-01	0.378	0.368	0.382	0.376	0.4	0.33	0.47
rs75621460	downstream gene	A	G	1.16	4.40E-15	1.12	1.24E-02	0.021	0.018	0.028	0.025	0.03	0.01	0.07
rs4252548	missense	T	C	1.37	5.30E-13	1.11	1.63E-01	0.028	0.021	0.026	0.023	0.01	0.002	0.04
rs2836618	intergenic region	A	G	1.08	1.90E-07	1.18	2.17E-06	0.275	0.261	0.285	0.252	0.19	0.14	0.26
rs528981060	upstream gene	A	G	1.68	2.40E-08			0.0012	0.0008					
Previously reported loci														
rs2820443	intergenic region	C	T	1.06	9.90E-10	1.06	1.20E-02	0.308	0.297	0.312	0.299	0.26	0.2	0.34
rs3771501	intron	A	G	1.05	6.00E-15	1.06	1.31E-02	0.484	0.473	0.489	0.475	0.48	0.4	0.55
rs3774355	upstream gene	A	G	1.09	3.00E-10	1.15	2.07E-05	0.378	0.358	0.389	0.357	0.39	0.32	0.47
rs2396502	intron	C	A	1.09	1.80E-10	1.1	2.23E-03	0.622	0.602	0.622	0.6	0.64	0.57	0.71
rs12209223	upstream gene	A	C	1.16	2.50E-12	1.23	7.79E-06	0.113	0.1	0.124	0.104	0.09	0.06	0.15

rs10974438	intron	A	C	1.03	1.10E-07	1.06	6.39E-03	0.657	0.649	0.658	0.646	0.63	0.55	0.7
rs34687269	intron	A	T	1.08	1.70E-10	1.1	1.67E-03	0.55	0.53	0.556	0.532	0.52	0.45	0.6
rs10492367	intergenic region	T	G	1.16	2.90E-19	1.21	4.38E-07	0.209	0.187	0.219	0.188	0.18	0.13	0.25
rs4775006	intergenic region	A	C	1.05	1.30E-07	1.12	2.58E-04	0.423	0.41	0.437	0.41	0.36	0.29	0.44
rs12901372	intron	C	G	1.07	2.90E-08	1.13	5.22E-05	0.549	0.532	0.565	0.535	0.47	0.4	0.55
rs9930333	intron	G	T	1.04	4.30E-07	1.1	2.57E-04	0.432	0.421	0.447	0.425	0.42	0.35	0.49
rs143384	5 prime UTR	A	G	1.1	3.60E-22	1.07	9.24E-03	0.619	0.596	0.603	0.587	0.64	0.57	0.71

Most Severe Consequence: Most severe consequence of index variant as predicted by SNPEff; EA: Effect allele; NEA: Non-effect allele; OR: Odds ratio for the index trait in Table 1; PV: *P* value (two-sided) for the index trait in Table 1; EAF: Effect allele frequency; GBR: 1000 Genomes Project GBR; L95: lower 95% confidence intervals of the EAF for 1000 Genomes Project GBR; U95: 95% confidence intervals of the EAF for 1000 Genomes Project GBR.

Supplementary Table 6: eQTL and pQTL studies used with Piccolo.

PMID	TISSUE	QTL type
24604203	Dendritic cells (derived from monocytes)	eQTL
22233810	Dendritic cells	eQTL
24604202	Monocytes	eQTL
28814792	Monocytes	eQTL
25214635	CD4+ T cells	eQTL
27768889	Monocyte derived Macrophage	eQTL
25874939	PBMCs	eQTL
26917434	Leucocytes	eQTL
27768888	Monocytes	eQTL
23878721	Whole blood	eQTL
26151758	Neutrophils	eQTL
24786080	CD4(+) T cells and monocytes	eQTL
24610777	CD4(+) T cells and T regs	eQTL
22446964	Monocytes and B-cells	eQTL
28248954	CD4+ and CD8+ T cells	eQTL
24013639	Whole blood	eQTL
27918533	Whole blood	eQTL
23209423	Lung	eQTL
21949713	Sputum from COPD patients (ECLIPSE)	pQTL
28564610	Dorsal Root Ganglion	eQTL
NA	Serum asthma	pQTL
18462017	Liver	eQTL
21637794	Liver	eQTL
28240269	Plasma	pQTL

29875488	Plasma	pQTL
28193859	Islet	eQTL
28575649	Kidney	eQTL

Supplementary Table 11: Genetic correlations between osteoarthritis phenotypes and 69 other traits and diseases. Estimates were made by LD score regression[1] as implemented in LDHub[2]. The table contains only traits with significant (*P*<0.05) correlation with at least one of the four definitions of osteoarthritis tested following multiple testing correction. Red font colour denotes non-significant correlation and negative values indicate negative correlations. AI: autoimmune; BV: brain volume; CM: cardiometabolic; C: cognitive; H: hormone; K: kidney; L: lipids; LF: lung function O: other

	Osteoarthritis	Knee and/or hip osteoarthritis	Knee osteoarthritis	Hip osteoarthritis	
Fathers age at death	-0.3167	-0.2569	-0.2989	-0.1314	Aging
Mothers age at death	-0.3325	-0.2782	-0.3202	-0.0973	
Parents age at death	-0.2959	-0.2323	-0.3467	0.0105	
Body fat	0.3578	0.3382	0.3506	0.2017	Anthropometric
Body mass index	0.4156	0.3923	0.3873	0.2464	
Childhood obesity	0.3257	0.3049	0.3008	0.2154	
Extreme bmi	0.3724	0.3849	0.3558	0.2662	
Hip circumference	0.2891	0.3043	0.2641	0.2424	
Obesity class 1	0.4061	0.3725	0.3822	0.2142	
Obesity class 2	0.4320	0.4002	0.3927	0.2508	
Obesity class 3	0.3968	0.3307	0.3433	0.1653	
Overweight	0.4210	0.4028	0.4015	0.2442	
Sitting height ratio	-0.1404	-0.1139	-0.1126	-0.0631	
Waist circumference	0.3626	0.3569	0.3271	0.2552	
Waist-to-hip ratio	0.3068	0.2815	0.2763	0.1573	
Asthma	0.2229	0.1275	0.1276	0.0535	AI
Rheumatoid Arthritis	0.1552	0.1110	0.1224	0.0475	
Femoral neck bone mineral density (PMID:22504420)	0.1425	0.1541	0.1473	0.0932	Bone
Femoral neck bone mineral density (PMID:26367794)	0.1147	0.1613	0.1505	0.1192	
Lumbar spine bone mineral density (PMID:22504420)	0.1939	0.2575	0.2114	0.2271	
Lumbar spine bone mineral density (PMID:26367794)	0.1197	0.2046	0.1717	0.1714	
Mean Caudate	0.1557	0.1810	0.1754	0.0937	BV
Lung cancer	0.2362	0.2014	0.2465	0.0635	Cancer
Lung cancer (all)	0.2034	0.1943	0.1985	0.1119	
Lung cancer (squamous cell) (PMID:24880342)	0.2913	0.2600	0.2853	0.1367	
Squamous cell lung cancer (PMID:27488534)	0.3359	0.2951	0.3412	0.1027	
Coronary artery disease	0.1775	0.1361	0.1469	0.0798	CM

Intelligence	-0.2572	-0.2245	-0.2350	-0.1361	C
Childhood IQ	-0.2411	-0.1919	-0.2368	-0.062	Education
College completion	-0.3600	-0.2346	-0.3027	-0.0527	
Years of schooling (proxy cognitive performance)	-0.3381	-0.2232	-0.2911	-0.0447	
Years of schooling 2013	-0.3513	-0.2295	-0.3019	-0.037	
Years of schooling 2016	-0.3797	-0.2755	-0.3225	-0.1144	
Fasting insulin main effect	0.217	0.1802	0.1953	0.0952	Glycemic
HbA1C	0.2155	0.1224	0.1623	0.0032	
HOMA-IR	0.2345	0.1951	0.1947	0.1011	
Type 2 Diabetes	0.1621	0.1173	0.1368	0.053	
Leptin_not_adjBMI	0.3051	0.2884	0.3466	0.1357	H
Serum cystatin c	-0.1491	-0.1227	-0.1231	-0.0692	K
HDL cholesterol	-0.1020	-0.0558	-0.0754	0.0059	L
Triglycerides	0.0967	0.0496	0.0955	-0.0617	
Forced expiratory volume in 1 second (FEV1)	-0.1511	-0.1176	-0.1024	-0.0844	LF
Forced expiratory volume in 1 second (FEV1)/Forced Vital capacity(FVC)	-0.1195	-0.1177	-0.1456	-0.1157	
Citrate	-0.2043	-0.1992	-0.2471	0.0021	Metabolites
Glycoprotein acetyls; mainly a1-acid glycoprotein	0.1963	0.1034	0.1510	-0.0085	
Mean diameter for HDL particles	-0.1592	-0.1271	-0.1218	-0.0243	
Tyrosine	0.1858	0.1573	0.0234	0.2528	
Urate	0.0865	0.0872	0.1101	0.0256	O
Anorexia Nervosa	-0.1127	-0.0995	-0.0885	-0.0871	Psychiatric
Attention deficit hyperactivity disorder	0.3614	0.325	0.3384	0.1257	
Attention deficit hyperactivity disorder (GC)	0.3276	0.2088	0.1746	0.1314	
Attention deficit hyperactivity disorder (No GC)	0.3280	0.2088	0.1740	0.1324	
Autism spectrum disorder	-0.1930	-0.2128	-0.1989	-0.171	
Depressive symptoms	0.3670	0.2192	0.2454	0.0771	
PGC cross-disorder analysis	-0.0496	-0.1013	-0.1345	-0.0222	
Schizophrenia	-0.0771	-0.0919	-0.1099	-0.0204	
Age at Menarche	-0.1179	-0.0865	-0.1023	-0.0391	Reproductive
Age at Menopause	-0.1063	-0.0663	-0.0908	-0.0113	
Age of first birth	-0.4432	-0.3073	-0.3557	-0.1259	
Number of children ever born	0.2390	0.1626	0.1766	0.0712	
Chronotype	0.0682	0.0724	0.1062	-0.0067	Sleeping
Insomnia (PMID:28604731)	0.3795	0.2383	0.2614	0.0953	
Insomnia (PMID:27992416)	0.3399	0.1921	0.2135	0.0624	
Sleep duration	-0.2465	-0.1624	-0.1869	-0.0422	
Excessive daytime sleepiness	0.1317	0.1281	0.0790	0.1530	
Age of smoking initiation	-0.2574	-0.1299	-0.0996	0.0254	Smoking behaviour
Cigarettes smoked per day	0.2474	0.2224	0.3306	-0.0163	
Ever vs never smoked	0.2614	0.2010	0.2211	0.0857	
Former vs Current smoker	-0.3613	-0.3435	-0.3015	-0.2839	

Supplementary Table 12: Genetic correlations between osteoarthritis phenotypes and 120 other traits and diseases from the UK Biobank resource. Cross-trait LD score regression[1] as implemented in LDHub[2] was performed to calculate the genetic correlation estimates. The table contains only traits with significant correlations (*P*<0.05) across all osteoarthritis definitions. Negative values indicate negative correlations. A: aging; B: bone; C: cognitive; D: Disabilities; H: Hereditary; M: Mouth /teeth problems; Occup: occupational; R: Reproductive.

	Osteoarthritis	Knee and/or hip osteoarthritis	Knee osteoarthritis	Hip osteoarthritis	
Mothers age at death	-0.5121	-0.3996	-0.4200	-0.2384	A
Alcohol intake versus 10 years previously	0.3045	0.2217	0.2237	0.1386	Alcohol consumption
Average weekly champagne plus white wine intake	-0.2592	-0.1762	-0.1593	-0.1422	
Average weekly spirits intake	0.3743	0.3181	0.3208	0.1931	
Arm fat mass (left)	0.4504	0.4181	0.4408	0.2219	Anthropometric
Arm fat mass (right)	0.4526	0.4198	0.4436	0.2213	
Arm fat percentage (left)	0.4286	0.3819	0.4101	0.1925	
Arm fat percentage (right)	0.4285	0.3805	0.4104	0.1884	
Arm fat-free mass (left)	0.3139	0.3299	0.3259	0.2048	
Arm fat-free mass (right)	0.2976	0.3203	0.3130	0.2060	
Arm predicted mass (left)	0.3118	0.3258	0.3250	0.1985	
Arm predicted mass (right)	0.3002	0.3207	0.3135	0.2047	
Basal metabolic rate	0.3084	0.3295	0.3281	0.2011	
Body fat percentage	0.4038	0.3564	0.3781	0.1859	
Body mass index (BMI)	0.4932	0.4437	0.4735	0.2303	
Comparative body size at age 10	0.1908	0.2019	0.1680	0.1725	
Hip circumference	0.3906	0.3734	0.3826	0.2179	
Impedance of arm (left)	-0.3562	-0.3347	-0.3471	-0.1889	
Impedance of arm (right)	-0.3519	-0.3359	-0.3451	-0.1948	
Impedance of leg (left)	-0.2997	-0.3039	-0.3253	-0.1567	
Impedance of leg (right)	-0.2955	-0.3021	-0.3219	-0.1598	
Impedance of whole body	-0.3541	-0.3460	-0.3629	-0.1890	
Leg fat mass (left)	0.4662	0.4290	0.4462	0.2385	
Leg fat mass (right)	0.4656	0.4257	0.4429	0.2351	
Leg fat percentage (left)	0.4501	0.3907	0.4131	0.2062	
Leg fat percentage (right)	0.4470	0.3838	0.4069	0.1990	
Leg fat-free mass (left)	0.3118	0.3331	0.3358	0.1997	
Leg fat-free mass (right)	0.2932	0.3205	0.3214	0.1953	
Leg predicted mass (left)	0.3115	0.3337	0.3366	0.1990	

Leg predicted mass (right)	0.2926	0.3201	0.3210	0.1952	
Trunk fat mass	0.3904	0.3686	0.3819	0.2038	
Trunk fat percentage	0.3645	0.3274	0.3469	0.1710	
Trunk fat-free mass	0.2256	0.2607	0.2490	0.1710	
Trunk predicted mass	0.2251	0.2613	0.2499	0.1708	
Usual walking pace	-0.4197	-0.3475	-0.3695	-0.1802	
Waist circumference	0.4675	0.4208	0.4422	0.2257	
Weight	0.3951	0.3886	0.3971	0.2253	
Whole body fat mass	0.4249	0.3956	0.4128	0.2158	
Whole body fat-free mass	0.2683	0.2980	0.2929	0.1866	
Whole body water mass	0.2698	0.2986	0.2934	0.1873	
Heel bone mineral density (BMD) T-score_ automated	0.1214	0.1590	0.1243	0.1494	B
Fluid intelligence score	-0.2616	-0.2399	-0.2531	-0.1385	C
Vascular/heart problems diagnosed by doctor: High blood pressure	0.2335	0.1993	0.2112	0.1151	Cardiometabolic
Vascular/heart problems diagnosed by doctor: None of the above	-0.2540	-0.2141	-0.2245	-0.1249	
Non-cancer illness code_ self-reported: hypertension	0.2261	0.1945	0.2069	0.1084	
Mineral and other dietary supplements: Glucosamine	0.3414	0.3416	0.3062	0.2688	Diet
Mineral and other dietary supplements: None of the above	-0.1890	-0.1719	-0.1315	-0.1598	
Long-standing illness_ disability or infirmity	0.5812	0.4321	0.4822	0.1872	D
Diagnoses - main ICD10: M20 Acquired deformities of fingers and toes	0.4811	0.4543	0.4816	0.2428	Disorders
Diagnoses - main ICD10: M23 Internal derangement of knee	0.7779	0.7314	0.8463	0.2628	
Age completed full time education	-0.3601	-0.2721	-0.3230	-0.1061	Education
Qualifications: A levels/AS levels or equivalent	-0.3749	-0.3090	-0.3516	-0.1496	
Qualifications: College or University degree	-0.4027	-0.3173	-0.3538	-0.1486	
Qualifications: None of the above	0.3730	0.2998	0.3338	0.1410	
Qualifications: O levels/GCSEs or equivalent	-0.3313	-0.2916	-0.3165	-0.1498	
Qualifications: Other professional qualifications eg: nursing_ teaching	-0.2762	-0.1963	-0.2240	-0.1067	
Illnesses of siblings: Heart disease	0.3048	0.2590	0.2656	0.1567	H
Mouth/teeth dental problems: Dentures	0.3386	0.2811	0.3048	0.1527	M

Blood clot_ DVT_ bronchitis_ emphysema_ asthma_ rhinitis_ eczema_ allergy diagnosed by doctor: Blood clot in the leg (DVT)	0.3549	0.3130	0.2539	0.2769	Non-cancer illnesses
Diagnoses - main ICD10: G56 Mononeuropathies of upper limb	0.6068	0.4872	0.5455	0.2212	
Falls in the last year	0.4426	0.3589	0.3658	0.2008	
Non-cancer illness code_ self-reported: chronic obstructive airways disease/copd	0.4333	0.3455	0.2765	0.3291	
Non-cancer illness code_ self-reported: deep venous thrombosis (dvt)	0.3593	0.3266	0.2716	0.2805	
Diagnoses - main ICD10: J44 Other chronic obstructive pulmonary disease	0.5122	0.4077	0.3873	0.2973	
Number of self-reported non-cancer illnesses	0.4888	0.3475	0.3840	0.1423	
Non-cancer illness code_ self-reported: hiatus hernia	0.5070	0.3562	0.3384	0.2155	
Current employment status: Unable to work because of sickness or disability	0.5539	0.4366	0.4689	0.2270	Occup
Job involves shift work	0.3642	0.3065	0.3362	0.1631	
Diagnoses - main ICD10: T84 Complications of internal orthopaedic prosthetic devices_ implants and grafts	0.7541	0.7625	0.7416	0.5256	Operations
Had other major operations	0.5479	0.4541	0.4543	0.2562	
Number of operations_ self-reported	0.5480	0.4561	0.4919	0.1994	
Diagnoses - main ICD10: M16 Coxarthrosis [arthrosis of hip]	0.5742	0.7023	0.4091	0.8903	Osteoarthritis
Diagnoses - main ICD10: M17 Gonarthrosis [arthrosis of knee]	0.9190	0.8822	0.9628	0.4058	
Non-cancer illness code_ self-reported: osteoarthritis	0.9714	0.8368	0.8563	0.4633	
Health satisfaction	0.5927	0.4505	0.4793	0.2440	Overall health rating
Number of treatments/medications taken	0.5358	0.4134	0.4488	0.1940	
Overall health rating	0.5623	0.4233	0.4675	0.1989	
Taking other prescription medications	0.5390	0.3716	0.3873	0.1804	
Frequency of tiredness / lethargy in last 2 weeks	0.4468	0.2868	0.3443	0.0932	
Back pain for 3+ months	0.4754	0.3766	0.3723	0.2508	Pain
Chest pain or discomfort walking normally	0.4305	0.3877	0.3840	0.2404	
Diagnoses - main ICD10: M54 Dorsalgia	0.6127	0.5510	0.5534	0.3649	
Knee pain for 3+ months	0.6017	0.5468	0.5953	0.2366	
Leg pain on walking	0.8159	0.6896	0.7262	0.3685	
Medication for pain relief_ constipation_ heartburn: Laxatives (e.g. Dulcolax_ Senokot)	0.3990	0.3074	0.2975	0.2214	

Medication for pain relief_ constipation_ heartburn: None of the above	-0.5109	-0.4043	-0.4345	-0.1978	
Medication for pain relief_ constipation_ heartburn: Paracetamol	0.4576	0.3489	0.3815	0.1520	
Neck/shoulder pain for 3+ months	0.6995	0.4954	0.4739	0.329	
Pain type(s) experienced in last month: Back pain	0.5222	0.3813	0.4018	0.1929	
Pain type(s) experienced in last month: Hip pain	0.6440	0.5197	0.4342	0.4627	
Pain type(s) experienced in last month: Knee pain	0.7940	0.6979	0.7846	0.3007	
Pain type(s) experienced in last month: Neck or shoulder pain	0.5507	0.3920	0.4202	0.1855	
Pain type(s) experienced in last month: None of the above	-0.6271	-0.4974	-0.5440	-0.2350	
Wheeze or whistling in the chest in last year	0.4853	0.3539	0.3834	0.1701	
Medication for pain relief_ constipation_ heartburn: Omeprazole (e.g. Zanprol)	0.4935	0.3334	0.3657	0.1515	
Diagnoses - main ICD10: R07 Pain in throat and chest	0.4291	0.3028	0.3378	0.1523	
Pain type(s) experienced in last month: Pain all over the body	0.5292	0.3908	0.4098	0.1905	
Job involves heavy manual or physical work	0.3399	0.2701	0.3218	0.0964	Physical activity
Job involves mainly walking or standing	0.3141	0.252	0.2923	0.0977	
Shortness of breath walking on level ground	0.6177	0.4674	0.5047	0.2254	
Time spent driving	0.2415	0.2556	0.2376	0.1765	
Time spent watching television (TV)	0.3576	0.2917	0.3325	0.1103	
Transport type for commuting to job workplace: Car/motor vehicle	0.4611	0.4079	0.3762	0.2766	
Transport type for commuting to job workplace: Public transport	-0.3676	-0.3426	-0.3179	-0.2543	
Transport type for commuting to job workplace: Walk	-0.3926	-0.3798	-0.3483	-0.2808	
Types of physical activity in last 4 weeks: None of the above	0.4549	0.3301	0.3773	0.1213	
Types of physical activity in last 4 weeks: Walking for pleasure (not as a means of transport)	-0.3725	-0.2741	-0.3065	-0.1188	
Types of transport used (excluding work): Public transport	-0.3037	-0.2777	-0.263	-0.1918	
Types of transport used (excluding work): Walk	-0.3514	-0.297	-0.2841	-0.1961	
Illness_ injury_ bereavement_ stress in last 2 years: Financial difficulties	0.4292	0.3167	0.3582	0.1126	Psychiatric
Loneliness_ isolation	0.3249	0.2117	0.2191	0.1122	

Prospective memory result	0.2713	0.2330	0.2145	0.1647	
Risk taking	0.2249	0.1970	0.1904	0.1253	
Number of children fathered	0.2953	0.3065	0.3057	0.1756	R
Ever smoked	0.1839	0.126	0.1121	0.0865	Smoking behaviour
Light smokers_ at least 100 smokes in lifetime	0.2075	0.1967	0.1887	0.1358	
Pack years adult smoking as proportion of life span exposed to smoking PREVIEW ONLY	0.4288	0.3162	0.3696	0.1255	
Pack years of smoking PREVIEW ONLY	0.4423	0.3323	0.3836	0.1336	
Past tobacco smoking	-0.2327	-0.1755	-0.1617	-0.1170	
Smoking status: Previous	0.1878	0.1550	0.1307	0.1189	

Supplementary Table 17: Narrow sense osteoarthritis heritability explained by variants genome-wide (total h_{2L}) and the proportion of total h^2_L explained by genome-wide significant signals in the meta-analysis stage.

Trait	Total h^2_L (%)	Total h^2_L (%) explained by top signals	Proportion of total h^2_L (%) explained by top signals
Osteoarthritis	9.8 (8.57-9.91)	2.2	22.5
Hip osteoarthritis	16.38 (15.24-17.52)	8.5	51.9
Knee osteoarthritis	15.69 (14.89-16.49)	2.3	14.7
Knee and/or hip osteoarthritis	13.2 (12.56-13.84)	3.2	24.2

Supplementary Table 19: Enrichment analysis of monogenic bone diseases on osteoarthritis loci.

Bone disease, developmental [1]						
		Yes	No	OR	CI[2]	*P*[3]
OA Loci	Yes	6	574	8.87	2.87 - 23.42	$1.8x10^{4}$
	No	18	15281			
Early onset osteoarthritis [4]						
		Yes	No	OR	CI	*P*
OA Loci	Yes	3	577	8.83	1.53 - 35.52	0.008
	No	9	15290			

[1] Diseases annotated with the MeSH term "bone disease, developmental" (Online Methods). [2] 95% confidence interval of the enrichment odds ratio (OR). [3] Enrichment *P* value (two-sided) from a Fisher's exact test. [4] Genes underlying early onset osteoarthritis[3].

Supplementary Table 20: Genes in osteoarthritis loci causing monogenic bone diseases or early-onset osteoarthritis.

Gene	Disease type[1]	Disease	OMIM[2]	Orphanet[3]	Clinical symptoms
COL27A1	BDD	STEEL Syndrome	615155	ORPHA:438117	Characteristic facies, dislocated hips and radial heads, carpal coalition (fusion of carpal bones), short stature, scoliosis, and cervical spine anomalies
FGFR3	BDD	CATSHL syndrome	610474	ORPHA:85164	Camptodactyly (flexion deformity of the proximal interphalangeal joints), tall stature, and hearing loss syndrome
GDF5	BDD	Angel-shaped phalango-epiphyseal dysplasia	105835	ORPHA:63442	Angel-shaped middle phalanges, a typical metacarpophalangeal pattern profile (mainly affecting first metacarpals and middle phalanges of second, third and fifth digits, which all appear short), epiphyseal changes in the hips and, in some, abnormal dentition and delayed bone age
RNF135	BDD	Overgrowth-macrocephaly-facial dysmorphism syndrome	614192	ORPHA:137634	Tall stature, learning difficulties and facial dysmorphism

TBX4	BDD	Ischiocoxopodopatellar syndrome	147891	ORPHA:1509	Bone dysplasia affecting skeletal structures of the lower limb and the pelvis
TKT	BDD	SDDHD	617044	ORPHA:488618	Short stature, developmental delay, and congenital heart defects (SDDHD)
COL11A1	EO-OA	Stickler syndrome, type II	604841	ORPHA:90654	Clinically variable disorder characterized by ocular, auditory, skeletal, and orofacial abnormalities. Additional findings may multiple epiphyseal abnormalities and EO-OA
COL11A2	EO-OA	non-syndromic EO-OA	na	na	Age of onset <50 years, OA in multiple joint sites (hip, knee, hand, spine)
SMAD3	EO-OA	LOEYS-DIETZ SYNDROME 3 (aneurysm osteoarthritis syndrome)	613795	ORPHA:284984	Early-onset joint abnormalities, including OA, and arterial aneurysms.

[1] Indicates if the disease-causing gene was identified via the MeSH term “bone diseases, developmental (BDD) “or as a gene underlying “Early-onset Osteoarthritis” (EO-OA). [2] OMIM identification number from https://www.omim.org/. [3] Orphanet identification number from http://www.orpha.net/. “na” indicates not available.

SUPPLEMENTARY NOTE

STUDIES

UK Biobank case and control definitions: To define osteoarthritis cases, we used the self-reported status established during interview with a nurse (Field 20002; Initial assessment visit) and the Hospital Episode Statistics ICD10 primary and secondary codes. We conducted four osteoarthritis discovery GWAS: self-reported or hospital-diagnosed osteoarthritis at any site based on ICD10 hospital record codes M15-M19 (n=70,532); hospital-diagnosed hip osteoarthritis based on ICD10 hospital record M16 (n=12,850); hospital-diagnosed knee osteoarthritis based on ICD10 hospital record M17 codes (n=21,921); and hospital-diagnosed hip and/or knee osteoarthritis (M16 or M17; n=32,907). To minimise misclassification in the control datasets to the extent possible, in all analyses, from the controls, we excluded individuals with primary or secondary ICD10 codes M05 through M14, as well as arthrosis codes M15-M19. This excluded individuals with inflammatory polyarthropathies and resulted in a control sample size of 369,983 controls (Supplementary Figure 1).

Genotyping, imputation and association testing

UK Biobank: Briefly, 50,000 samples were genotyped using the UKBiLEVE array and the remaining samples were genotyped using the UK Biobank Axiom array (Affymetrix). After sample and SNP quality control (QC) of the directly-typed genotypes, followed by phasing and imputation, carried out centrally[4], there are approximately 97million variants in 487,411 individuals. Following additional QC checks, we excluded samples with call rate ≤95%. We checked samples for gender discrepancies, excess heterozygosity, over third degree relatedness, ethnicity and we removed possibly contaminated and withdrawn samples. Variants with minor allele frequency (MAF) of ≤0.1% or effective minor allele count (imputation quality score times minor allele count) ≤90 were discarded. In summary, we interrogated approximately 17 million variants from 450,805 related individuals of European descent (Supplementary Figure 1).

Association analysis: Principle components (PCs) were computed using fastPCA[5] and high quality directly typed markers from the unrelated set of Europeans. PCs were then projected to the related European sample using SNP weights. We tested for association using the non-infinitesimal mixed model association test implemented in BOLT-LMM v2.3[6] with adjustment for the first 10 PCs, sex, age at recruitment and genotyping chip. As BOLT-LMM implements a linear regression, effect size estimates and their standard errors for case-control outcomes have been calculated by dividing the effect size estimates and standard errors output by the software by $p \times (1-p)$, where p is the proportion of cases in the trait definition[7].

ArcOGEN and UKHLS: 7,410 arcOGEN cases were genotyped on the Illumina Human 610-Quad array, 670 arcOGEN cases were genotyped on the Illumina OmniExpress array, and 9,296 UKHLS controls were genotyped on the Illumina CoreExome array. Genotype QC has been previously described[8-10]. Prior to imputation all variants were mapped to GRCh37 and cases and controls were merged into a single dataset containing overlapping variants. We further excluded variants with MAF≤1, indels, and evidence for differential missingness between cases and controls (Fisher's exact test P<0.0001). We performed a case-control analysis and visually inspected the cluster plots for any variant with $P \leq 5 \times 10^{-8}$; variants with poor clusters were excluded. Final QC checks prior to imputation were performed using a HRC pre-imputation checking tool (http://www.well.ox.ac.uk/~wrayner/tools/#Checking). Imputation was performed on 17,376 individuals and 126,188 variants using the Michigan HRC[11,12] imputation server with Eagle2[13] for the prephasing. We excluded related individuals by performing pair-wise identity-by-descent (IBD) (using PI_HAT threshold > 0.2) in PLINK[14] using directly typed variants with MAF>1% and pruned for LD using r^2<0.2. In addition, we excluded any individuals that were related to individuals in the UK Biobank dataset using the imputed variants and the same IBD method as described above. We also excluded any variant that had an imputation information score (minimac R-square)<0.3, with a Hardy Weinberg $P \leq 1 \times 10^{-6}$ and MAF ≤ 0.1%The final dataset contained ~11.6 million variants in 6,520

cases and 8,186 controls. Association analysis was performed using SNPTEST[15] with the first ten principal components as covariates. We used the odds ratio in the meta-analysis. The arcOGEN dataset comprised three osteoarthritis phenotypes: hip osteoarthritis (2,854 cases and 8,186 controls), knee osteoarthritis (3,034 cases and 8,186 controls), and hip and/or knee osteoarthritis (6,520 cases and 8,186 controls).

TRANSCRIPTOME-WIDE ASSOCIATION

We observe 205 significant MetaXcan gene-trait associations in at least one tissue, which reduce to 69 when we filter out those with probability of colocalisation less than 80% (Supplementary Table 8). We observe 147 unique genes with at least one GTEx tissue significant association, while this number reduces to 52 genes if we filter the results by colocalisation probabilities (Supplementary Table 9). We calculated the average enrichment of genes by tissue for each of the osteoarthritis phenotypes as the mean of the square of the z-scores (effect size/standard error) of the association between the genetic component of gene expression and phenotype. We average the Z^2 for all the Bonferroni corrected significant gene-tissue pairs, and after further filtering them to have probability of colocalisation of at least 80% (Supplementary Figure 8). We observe enrichment across whole blood, skeletal muscle and brain tissues.

COLOCALISATION ANALYSIS

We observe evidence of colocalisation in at least one tissue for 50 out of our 64 loci using any of the 3 methods (MetaXcan, Coloc, Piccolo), 41 of which are at newly associated osteoarthritis signals (Supplementary Table 7). MetaXcan alone identified 119 genes, Coloc 113 and Piccolo 58, while the overlap of all 3 methods implicate 20 genes (*TGFA, ILF3, CSK, CYP1A1, ULK3, CHMP1A, TSKU, SUPT3H, GNL3, NT5DC2, LMX1B, SMAD3, MLXIP, COLGALT2, FAM89B, UQCC1, NFAT5, ALDH1A2, FAM53A, FGFR3*; Supplementary Figure 7). An interesting finding of the colocalisation analyses is for the known osteoarthritis locus around *GDF5*. There is a body of literature supporting *GDF5* as the susceptibility gene. In our analyses, we found more evidence of GWAS-eQTL association for nearby *UQCC1*. Piccolo reports significant GWAS-eQTL association in 2 tissues for *UQCC1* and none for *GDF5*, MetaXcan reports such associations in 15 tissues for *UQCC1* and 8 for *GDF5*, while Coloc reports associations in 17 tissues for *UQCC1* and 6 for *GDF5* (Supplementary Table 7). Specifically, higher predicted expression of *UQCC1* in fibroblast cells is associated with increased disease risk. Positive correlation was also predicted in skeletal muscle for *UQCC1*. No GWAS-eQTL association was predicted for *GDF5* in fibroblast cells or skeletal muscle. *UQCC1* has been previously associated in GWAS with developmental dysplasia of the hip, height, spine bone size and hip axis length. Further studies are needed to investigate the role of *UQCC1* at the same locus as *GDF5*.

FINE-MAPPING

One out of the 64 regions contains a secondary signal reaching genome-wide significance *(P*-value<3×10^{-8} conditioned on the top signal). Six regions fine-map to a single variant, 20 regions to less than ten variants, and 26 regions to less than twenty variants (Supplementary Table 18). If we restrict the 95% credible sets to variants with posterior probability of causality (PPC) over 0.03 or to variants with moderate or high consequence (irrespective of their PPC), 60 out of the 64 regions fine-map to up to 17 variants, with 33 of these 60 regions containing the causal variant(s) with probability of over 80%. In 6 regions, variants with probability of causality over 0.03 in the 95% credible sets have more severe consequences than the index variant of the region (Supplementary Table 18). There are 28 genes in the 95% credible intervals with at least one possibly causal variant with high or moderate consequence (Supplementary Table 18). When we sum the PPC of variants with moderate or high consequence in the 95% credible set, four of these 28 genes have over 89% probability of containing the causal variant (*IL11, SLC39A8, ANAPC4, DPEP1*) and another three over 1% (*TSEN15, HFE, SLBP*). Five of these 28 genes overlap with MetaXcan and Coloc results (*MST1R,*

RABGAP1L, TSEN15, UQCC1, CLEC18A), 3 genes overlap with Piccolo results (*RAPH1, ITIH1, UQCC1*), while *UQCC1* is identified by all 4 approaches.

TRANSCRIPTOMIC AND PROTEOMIC ANALYSIS

Proteomics: LC-MS analysis was performed on the Dionex Ultimate 3000 UHPLC system coupled with the Orbitrap Fusion Tribrid Mass Spectrometer. To account for protein loading, abundance values were normalised by the sum of all protein abundances in a given sample, then log2-transformed and quantile normalised. We restricted the analysis to 3917 proteins that were quantified in all samples. We tested proteins for differential abundance using limma[16] in R, based on a within-individual paired sample design. Significance was defined at 1% Benjamini-Hochberg FDR to correct for multiple testing. Of the 3732 proteins with unique mapping of gene name and Ensembl ID, we took forward 245 and 489 proteins with significantly different abundance between intact and degraded cartilage at 1% and 5% FDR, respectively (Supplementary Table 5).

RNA sequencing:

Multiplexed libraries were sequenced on the Illumina HiSeq 2000 (75bp paired-end read length). This yielded bam files for cohort 1 and cram files for cohorts 2 and 3. The cram files were converted to bam files using samtools 1.3.1[17] and then to fastq files using biobambam 0.0.191[18], after exclusion of reads that failed QC. We obtained transcript-level quantification using salmon 0.8.2[19] (with --gcBias and --seqBias flags to account for potential biases) and the GRCh38 cDNA assembly release 87 downloaded from Ensembl. We converted transcript-level to gene-level count estimates, with estimates for 39037 genes based on Ensembl gene IDs. After QC, we retained expression estimates for 15994 genes with counts per million of 1 or higher in at least 10 samples. Limma-voom[20] was used to remove heteroscedasticity from the estimated expression data. We tested genes for differential expression using limma in R (with lmFit and eBayes), based on a within-individual paired sample design. Significance was defined at 1% Benjamini-Hochberg FDR to correct for multiple testing. Of the 14408 genes with unique mapping of gene name and Ensembl ID, we took forward 1705 genes with significantly different abundance between intact and degraded cartilage at 1% FDR.

FUNCTIONAL GENOMICS

We examined quantitative proteomics and RNA sequencing data for all genes within 1Mb of the osteoarthritis-associated variants (Supplementary Table 5) to determine if they were differentially regulated between paired intact and degraded chondrocytes extracted from 38 osteoarthritis patients that had undergone total joint replacement surgery. At 1% FDR, 13 proteins show differential abundance (DA), and 122 genes show differential expression (DE) (Supplementary Table 5). Supplementary Table 5 additionally contains a systematic overview of all the analysis performed to identify the underlying causal gene/s on each region, including: eQTL colocalization analysis, variants with severe consequences on the credible sets (missense, etc), monogenic diseases and animal model data.

MENDELIAN RANDOMIZATION

In analyses of exposures in MR-Base to osteoarthritis outcomes, genetic predisposition to higher BMI, hip and waist circumference was associated with higher risk of osteoarthritis of both the knee and hip (Supplementary Table 13). We also saw evidence of protection against knee osteoarthritis arising from higher years of schooling (Supplementary Table 13) and likelihood of attaining a university degree, although we acknowledge the complex nature of these phenotypes, uncertainty over the mechanistic links between educational attainment and OA and the potential for confounding by other factors. For example, predisposition to higher educational attainment is also associated with lower BMI[21], and as such, we performed multivariable MR[22] to investigate the effect of educational attainment while adjusting for BMI. In these analyses, the association remained for education after adjusting for BMI (Supplementary Table 14). However, we acknowledge that the mechanisms and/or unknown confounders of this association are unclear. Predisposition to higher

LDL cholesterol (and various related lipid species) was protective against OA (Supplementary Table 13). In multivariable MR analyses of LDL-, HDL-cholesterol and triglycerides, LDL cholesterol associations remained, but we saw no associations with HDL-cholesterol or triglycerides (Supplementary Table 14). Among the exposures recently included in MR-Base[21] following large-scale GWAS analysis of UK Biobank, we saw similar associations with BMI, education and statin therapy as a proxy for cholesterolaemia (Supplementary Table 13). We also saw associations of genetic predisposition to knee pain (as instrumented in UK Biobank) with risk of osteoarthritis (Supplementary Table 13), although these associations may reflect reverse causality in that of the seven knee pain instruments, three were also genome-wide significantly associated with osteoarthritis. In analyses investigating the potential causal consequences of predisposition to osteoarthritis, all significant associations were observed in the outcome data derived from the same UK Biobank population in which the osteoarthritis instruments were derived (Supplementary Table 15). We saw potential causal relationships with pain phenotypes and medications, glucosamine supplementation, self-reported walking pace and hand grip strength, among other phenotypes. Self-report of both knee and hip pain in the last month were among the leading associations with predisposition to osteoarthritis (Supplementary Table 15). We next tested the association of predisposition to osteoarthritis with reported pain, but restricted to individuals in UK Biobank who had not reported, nor been hospital-diagnosed with osteoarthritis. In those analyses, we found that higher risk of osteoarthritis was associated with higher likelihood of reporting pain phenotypes (Supplementary Table 16) further suggesting that the apparent causal association of pain with OA is driven by reverse causality. Recent analyses have highlighted the potential for adjustment or stratification by heritable risk factors to induce collider bias[23], however, in this instance, the direction of bias is opposite to the effect that we observe, suggesting that the effects in the "undiagnosed- osteoarthritis" group are not the result of collider bias.

PATHWAY ANALYSIS

We conducted gene-set enrichment analyses in MAGMA v1.06[24], DEPICT[25] and PASCAL[26] to understand the potential functional effects of the identified variants and to identify biological pathways implicated in osteoarthritis by using widely-used gene annotation databases (Methods). MAGMA's pathway analysis identifies two biological processes associated with osteoarthritis: collagen formation (P=5.63x10^{-5} in self-reported or hospital diagnosed osteoarthritis; P=1.53x10^{-6} in hip osteoarthritis; P=4.74x10^{-7} in hip and/or knee osteoarthritis) and extracellular matrix (ECM) organization (Reactome database: P=1.98x10^{-5} in hip osteoarthritis; P=1.17x10^{-6} in hip and/or knee osteoarthritis; GO database: P=2.31x10^{-5} in hip and/or knee osteoarthritis) (Supplementary Table 10). Collagen formation (P=1.40x10^{-4} in osteoarthritis at any site; P=4.95x10^{-7} in hip osteoarthritis; P=1.37x10^{-5} in hip and/or knee osteoarthritis) and ECM organization (P=1.46x10^{-4} in osteoarthritis at any site; P=6.54x10^{-6} in hip osteoarthritis; P=1.10x10^{-6} in hip and/or knee osteoarthritis) are also identified as significant gene sets at FDR<0.05 in 3/4 osteoarthritis phenotypes by PASCAL analysis (Supplementary File 10). Protein binding is the most significant gene set in all phenotypes tested (P=2.81x10^{-8} in osteoarthritis at any site; P=3.56x10^{-7} in hip osteoarthritis; P=2.76x10^{-8} in knee osteoarthritis; P=7.47x10^{-9} in hip and/or knee osteoarthritis). Apart from the above 3, PASCAL reveals 7 more pathways of which collagen catabolic process is significant in hip (P=1.19x10^{-4}) and hip and/or knee osteoarthritis (P=9.95x10^{-5}). DEPICT pathway analysis resulted in 4, 22 and 35 significant gene sets at FDR<0.05 for osteoarthritis at any site, hip osteoarthritis and hip and/or knee osteoarthritis, respectively (Supplementary File 10). The 5 most significant gene sets are a) abnormal fibula morphology (P=8.05x10^{-7} in hip osteoarthritis), b) abnormal cartilage morphology (P=9.25x10^{-7} in hip and/or knee osteoarthritis), c) abnormal vertebral body morphology (P=2.09x10^{-6} in hip and/or knee osteoarthritis), d) regulation of growth (P=2.48x10^{-6} in osteoarthritis at any site) and regulation of cartilage development (P=2.56x10^{-6} in hip and/or knee osteoarthritis). The first three gene sets belong to Mammalian Phenotype Ontology (MP) database where the last two in Gene Ontology (GO).

Collagen formation and ECM organization were the two biological pathways identified by both MAGMA and PASCAL. ECM is a non-cellular component of all mammalian tissues, a network consisting of proteoglycans, laminins, fibronectin and collagen[27]. The relative abundance of ECM proteins and their interactions enables the ECM to carry out tissue-specific roles, including proliferation, adhesion, migration, cell differentiation and death[27,28]. Collagens are an essential component of the ECM and are involved in structural integrity[29] and processes such as tissue reconstruction[30]. Articular cartilage degeneration in osteoarthritis is associated with changes in type II collagen and further ECM components[31]. The two identified pathways are not independent of each other: the collagen formation gene-set is a subset of the ECM organization gene-set.

THE ARCOGEN CONSORTIUM

John Loughlin[1], Nigel Arden[2], Fraser Birrell[3,4], Andrew Carr[2,5], Panos Deloukas[6,7], Michael Doherty[8], Andrew W. McCaskie[9,10], William E. R. Ollier[11], Ashok Rai[12], Stuart H. Ralston[13], Tim D. Spector[14], Ana M. Valdes[8], Gillian A. Wallis[15], J. Mark Wilkinson[16], Eleftheria Zeggini[17,18].

Affiliations

1. Musculoskeletal Research Group, Institute of Cellular Medicine, Newcastle University, Newcastle-upon-Tyne, NE2 4HH, UK.
2. Botnar Research Centre, University of Oxford, Nuffield Orthopaedic Centre, Oxford, OX3 7LD UK.
3. Musculoskeletal Research Group, Institute of Cellular Medicine, Newcastle University, Newcastle upon-Tyne, NE2 4HH, UK.
4. Northumbria Healthcare NHS Foundation Trust, Wansbeck General Hospital, NE63 9JJ, UK.
5. NIHR Oxford Biomedical Research Centre John Radcliffe Hospital Oxford OX3 9DU UK.

6. William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University, London, EC1M 6BQ, UK.
7. Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University Jeddah, 21589, Saudi Arabia.
8. Academic Rheumatology, School of Medicine, University of Nottingham, UK, Nottingham, NG5 1PB, UK.
9. Division of Trauma and Orthopaedic Surgery, Department of Surgery, University of Cambridge Cambridge, CB2 0QQ, UK.
10. Musculoskeletal Research Group, Institute of Cellular Medicine, Newcastle University Newcastle-upon-Tyne, NE2 4HH, UK.
11. School of Healthcare Science, Manchester Metropolitan University, Manchester M1 5GD, UK.

12. Worcestershire Acute Hospitals NHS Trust, Worcester, UK.
13. Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, EH4 2XU, UK.
14. Department of Twin Research and Genetic Epidemiology, King's College London, London, SE1 7EH, UK.
15. Wellcome Centre for Cell Matrix Research, University of Manchester, Manchester M13 9PT UK.
16. Department of Oncology and Metabolism, University of Sheffield, Sheffield, S10 2JF, UK.
17. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, CB10 1SA, UK.
18. Institute of Translational Genomics, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, Germany

REFERENCES

1. Bulik-Sullivan, B.K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* **47**, 291-5 (2015).
2. Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272-279 (2017).
3. Aury-Landas, J., Marcelli, C., Leclercq, S., Boumediene, K. & Bauge, C. Genetic Determinism of Primary Early-Onset Osteoarthritis. *Trends Mol Med* **22**, 38-52 (2016).
4. Bycrof, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *Preprint at https://www.biorxiv.org/content/early/2017/07/20/166298* (2017).
5. Galinsky, K.J. *et al.* Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. *Am J Hum Genet* **98**, 456-472 (2016).
6. Loh, P.R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet* **47**, 284-90 (2015).
7. Cook, J.P., Mahajan, A. & Morris, A.P. Guidance for the utility of linear models in meta-analysis of genetic association studies of binary phenotypes. *Eur J Hum Genet* **25**, 240-245 (2017).
8. arcOGEN Consortium *et al.* Identification of new susceptibility loci for osteoarthritis (arcOGEN): a genome-wide association study. *Lancet* **380**, 815-23 (2012).
9. Evangelou, E. *et al.* A meta-analysis of genome-wide association studies identifies novel variants associated with osteoarthritis of the hip. *Ann Rheum Dis* **73**, 2130-6 (2014).
10. Prins, B.P. *et al.* Genome-wide analysis of health-related biomarkers in the UK Household Longitudinal Study reveals novel associations. *Sci Rep* **7**, 11008 (2017).
11. McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-83 (2016).
12. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat Genet* **48**, 1284-1287 (2016).
13. Loh, P.R. *et al.* Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet* **48**, 1443-1448 (2016).
14. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559-75 (2007).
15. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-13 (2007).
16. Ritchie, M.E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
17. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
18. Tischler, G. & Leonard, S. biobambam: tools for read pair collation based algorithms on BAM files. *Source Code for Biology and Medicine* **9**, 13-13 (2014).
19. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Meth* **14**, 417-419 (2017).
20. Law, C.W., Chen, Y., Shi, W. & Smyth, G.K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology* **15**, R29 (2014).
21. Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7:e34408**(2018).
22. Burgess, S. & Thompson, S.G. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* **181**, 251-60 (2015).
23. Day, F.R., Loh, P.R., Scott, R.A., Ong, K.K. & Perry, J.R. A Robust Example of Collider Bias in a Genetic Association Study. *Am J Hum Genet* **98**, 392-3 (2016).
24. de Leeuw, C.A., Mooij, J.M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* **11**, e1004219 (2015).

25. Pers, T.H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nat Commun* **6**, 5890 (2015).
26. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLoS Comput Biol* **12**, e1004714 (2016).
27. Mouw, J.K., Ou, G. & Weaver, V.M. Extracellular matrix assembly: a multiscale deconstruction. *Nat Rev Mol Cell Biol* **15**, 771-85 (2014).
28. Hynes, R.O. The extracellular matrix: not just pretty fibrils. *Science* **326**, 1216-9 (2009).
29. Gordon, M.K. & Hahn, R.A. Collagens. *Cell Tissue Res* **339**, 247-57 (2010).
30. Twardowski, T., Fertala, A., Orgel, J.P. & San Antonio, J.D. Type I collagen and collagen mimetics as angiogenesis promoting superpolymers. *Curr Pharm Des* **13**, 3608-21 (2007).
31. Tchetina, E.V. Developmental mechanisms in articular cartilage degradation in osteoarthritis. *Arthritis* **2011**, 683970 (2011).