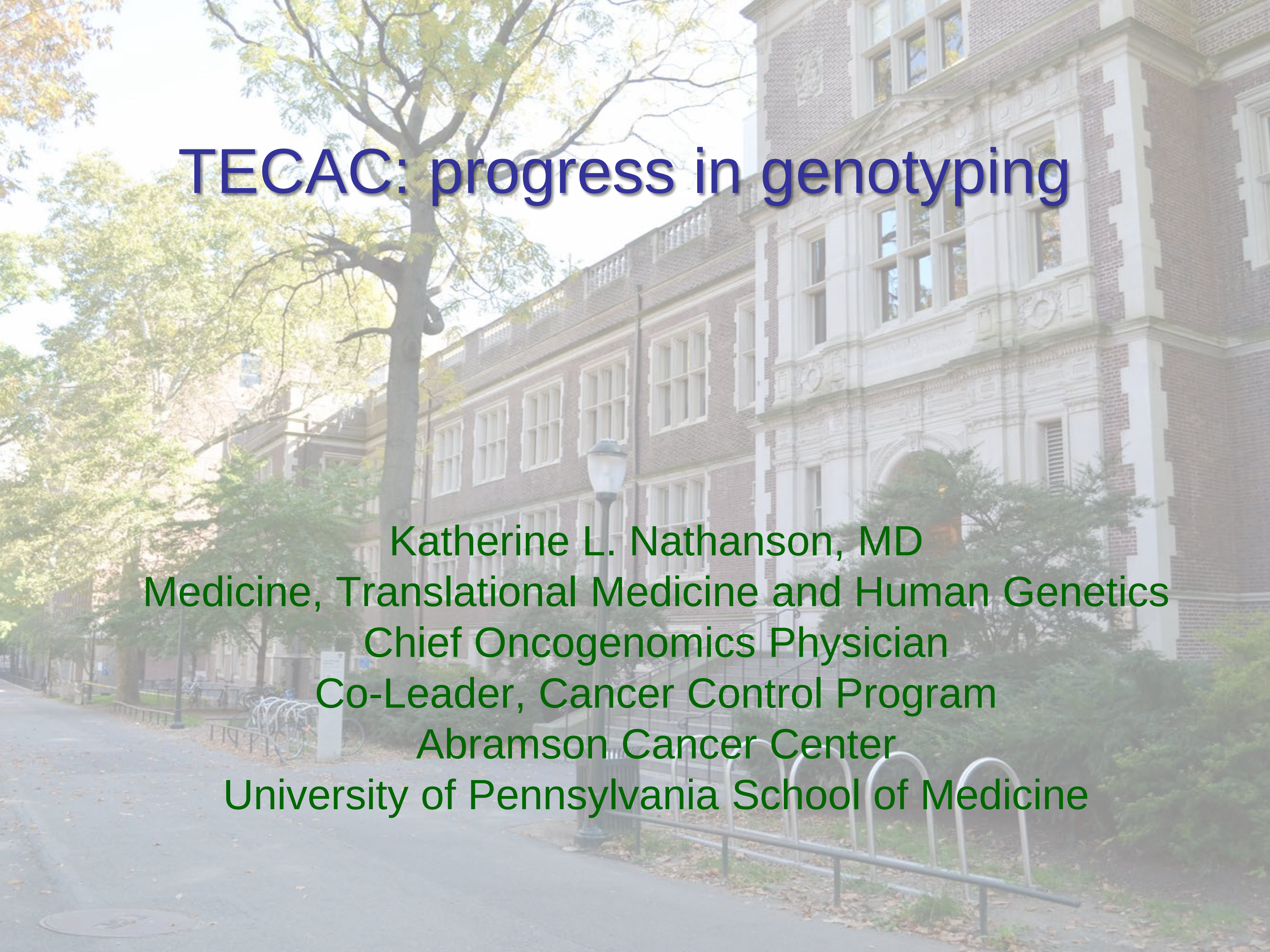


TECAC: progress in genotyping

Katherine L. Nathanson, MD
Medicine, Translational Medicine and Human Genetics
Chief Oncogenomics Physician
Co-Leader, Cancer Control Program
Abramson Cancer Center
University of Pennsylvania School of Medicine

Meta-analysis (discovery) progress overview

- Methods for meta-analysis
- Preliminary data from meta-analysis
- Next steps

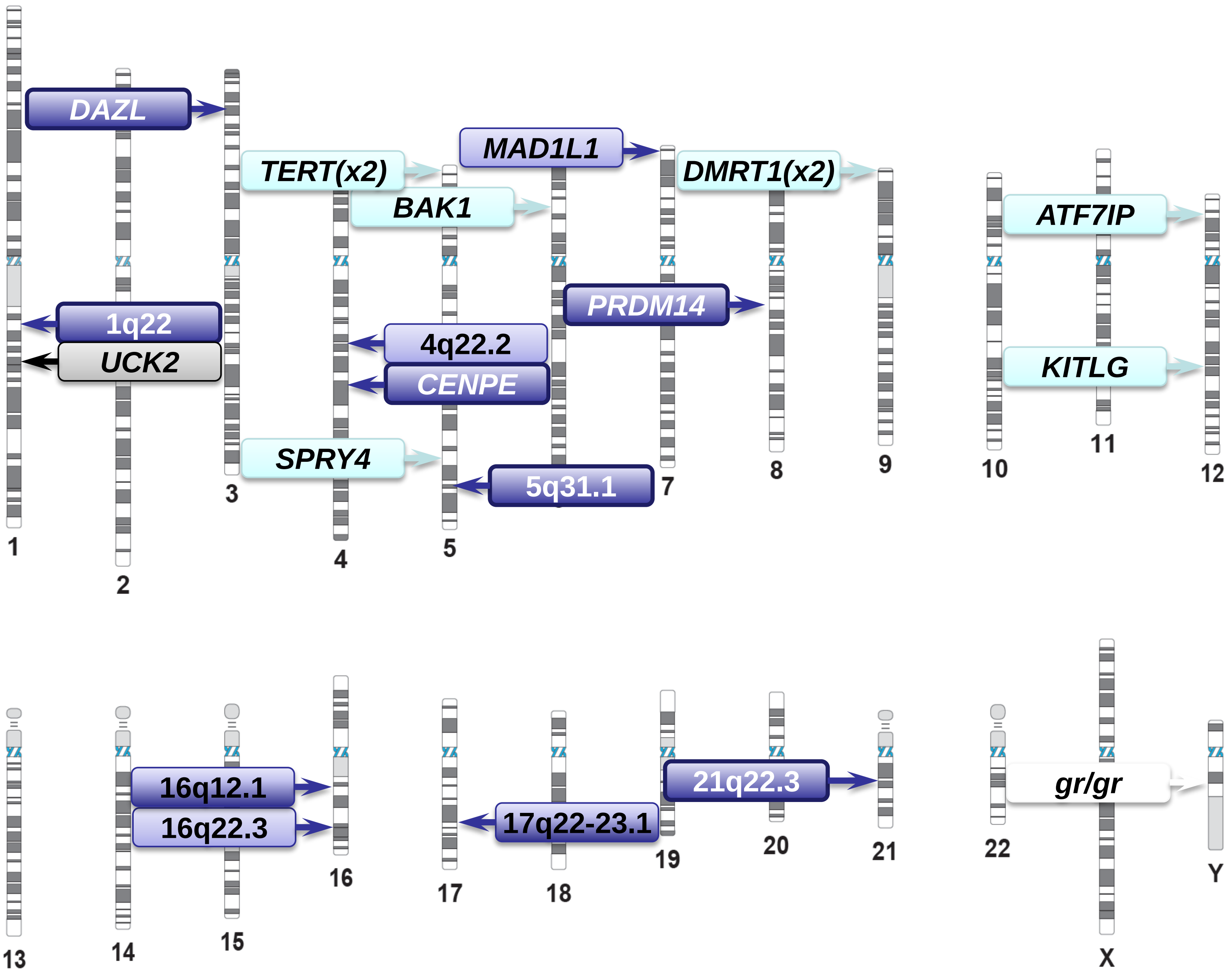
Data contributed

- NCI data (STEED, familial studies)
  - Individual case and control raw and imputed genotype data used for meta-analysis
- Penn data
  - Raw and imputed genotype data for cases deposited
  - Control data will be deposited
  - Aggregate data used for main meta-analysis
- UK data
  - Raw and imputed genotype data for cases deposited
  - WTCCC control data in process
  - Aggregate data used for main meta-analysis

Methods

- Analysis plan
  - Imputation of each GWAS scan using 1K Genomes V3 as reference set
  - Poorly imputed SNPs (INFO <0.3) or MAF < 0.01 removed
- Fixed effect model meta-analysis done to combine all three GWAS scans
- 2043 cases/6920 controls
  - NCI 582/1056, UK 986/4946, Penn 480/918
- Analysis done by Zhaoming Wang at NCI
  - Steve Chanock's group

Currently established TGCT susceptibility loci



Risk associated with known loci

New meta-analysis top SNPs

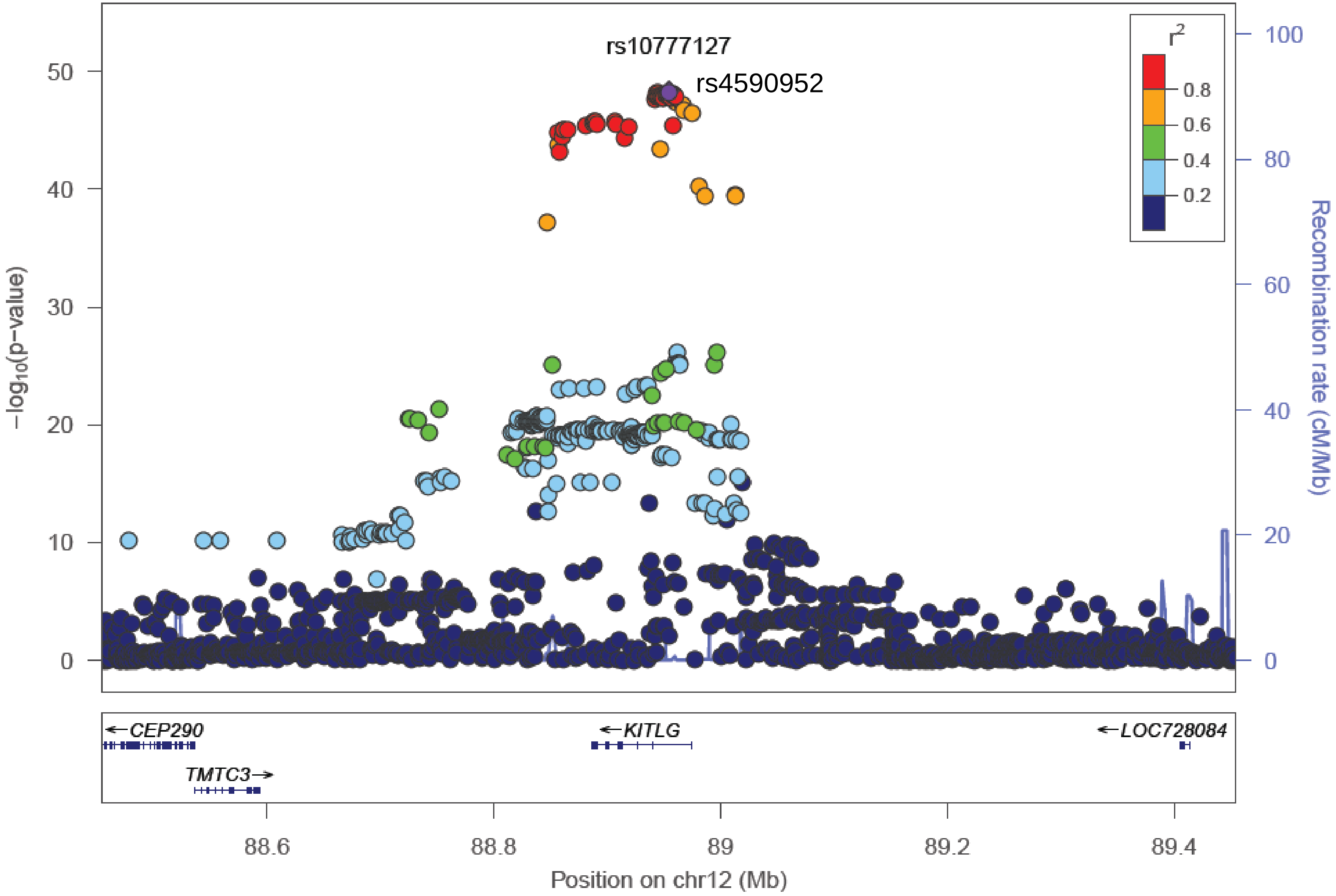
	Cytoband	SNP	Location	OR	CI	P value	Gene
1	1q22	rs909269	156192028	1.24	1.13-1.56	7.14E-08	*SLC25A44, PMF1*
2	1q24.1	rs10800166	165895525	1.28	1.18-1.39	2.95E-09	*UCK2*
3	3p24.3	rs11713721	16616628	0.79	0.73-0.85	1.12E-10	*DAZL*
4	4q22.3	rs6840913	95215892	0.80	0.74-0.87	3.16E-08	*SMARCAD1, HPGDS*
5	4q24	rs2711897	104023567	0.77	0.72-0.83	1.98E-11	*BDH2, UNQ6308, CENPE*
6	5p15.33	rs35953391	1312329	**1.65**	1.51-1.80	7.17E-29	*CLPTM1L*
7	5q31.1	rs3805663	134394099	1.25	1.13-1.38	*	*PTIX1*
8	5q31.3	rs6580240	141684389	**1.52**	1.42-1.64	5.14E-30	*SPRY4*
9	6p21.31	rs210138	33542538	**1.56**	1.43-1.71	7.40E-23	*BAK1*
10	7p22.3	rs13226911	1885144	1.36	1.25-1.48	5.75E-13	*MAD1L1*
11	8q13.3	rs6992772	70976756	0.82	0.76-0.88	2.69E-07	*PRDM14*
12	9p24.3	rs7863616	861220	**1.6**	1.49-1.73	4.62E-34	*DMRT1*
13	12p13.1	rs11055966	rs11055966	0.81	0.75-0.88	9.52E-08	*ATF7IP*
14	12q21.32	rs10777127	88954295	**2.0**	1.82-2.19	4.68E-49	*KITLG*
15	16q12.1	rs8046148	48700445	1.30	1.18-1.42	8.31E-09	*HEATR3*
16	16q23.1	rs7188581	73227959	0.81	0.75-0.87	1.53E-08	*RFWD3*
17	17q22	rs80030779	56765478	0.77	0.71-0.83	4.09E-12	*TEX14, RAD51C*
18	21q22.3	rs2839186	46514465	1.26	1.15-1.38	*	*MCM3AP*

Risk associated with known loci

New meta-analysis top SNPs

Cytoband	SNP	Location	OR	CI	P value	Gene
12q21.32	rs10777127	88954295	2	1.82-2.19	4.68E-49	*KITLG*
9p24.3	rs7863616	861220	1.6	1.49-1.73	4.62E-34	*DMRT1*
5q31.3	rs6580240	141684389	1.5	1.42-1.64	5.14E-30	*SPRY4*
5p15.33	rs35953391	1312329	1.7	1.51-1.80	7.17E-29	*CLPTM1L*
6p21.31	rs210138	33542538	1.6	1.43-1.71	7.40E-23	*BAK1*
7p22.3	rs13226911	1885144	1.4	1.25-1.48	5.75E-13	*MAD1L1*
17q22	rs80030779	56765478	0.8	0.71-0.83	4.09E-12	*TEX14, RAD51C*
4q24	rs2711897	104023567	0.8	0.72-0.83	1.98E-11	*BDH2, UNQ6308, CENPE*
3p24.3	rs11713721	16616628	0.8	0.73-0.85	1.12E-10	*DAZL*
1q24.1	rs10800166	165895525	1.3	1.18-1.39	2.95E-09	*UCK2*
16q12.1	rs8046148	48700445	1.3	1.18-1.42	8.31E-09	*HEATR3*
16q23.1	rs7188581	73227959	0.8	0.75-0.87	1.53E-08	*RFWD3*
4q22.3	rs6840913	95215892	0.8	0.74-0.87	3.16E-08	*SMARCAD1, HPGDS*
1q22	rs909269	156192028	1.2	1.13-1.56	7.14E-08	*SLC25A44, PMF1*
12p13.1	rs11055966	rs11055966	0.8	0.75-0.88	9.52E-08	*ATF7IP*
8q13.3	rs6992772	70976756	0.8	0.76-0.88	2.69E-07	*PRDM14*
5q31.1	rs3805663	134394099	1.3	1.13-1.38	*	*PTIX1*
21q22.3	rs2839186	46514465	1.3	1.15-1.38	*	*MCM3AP*

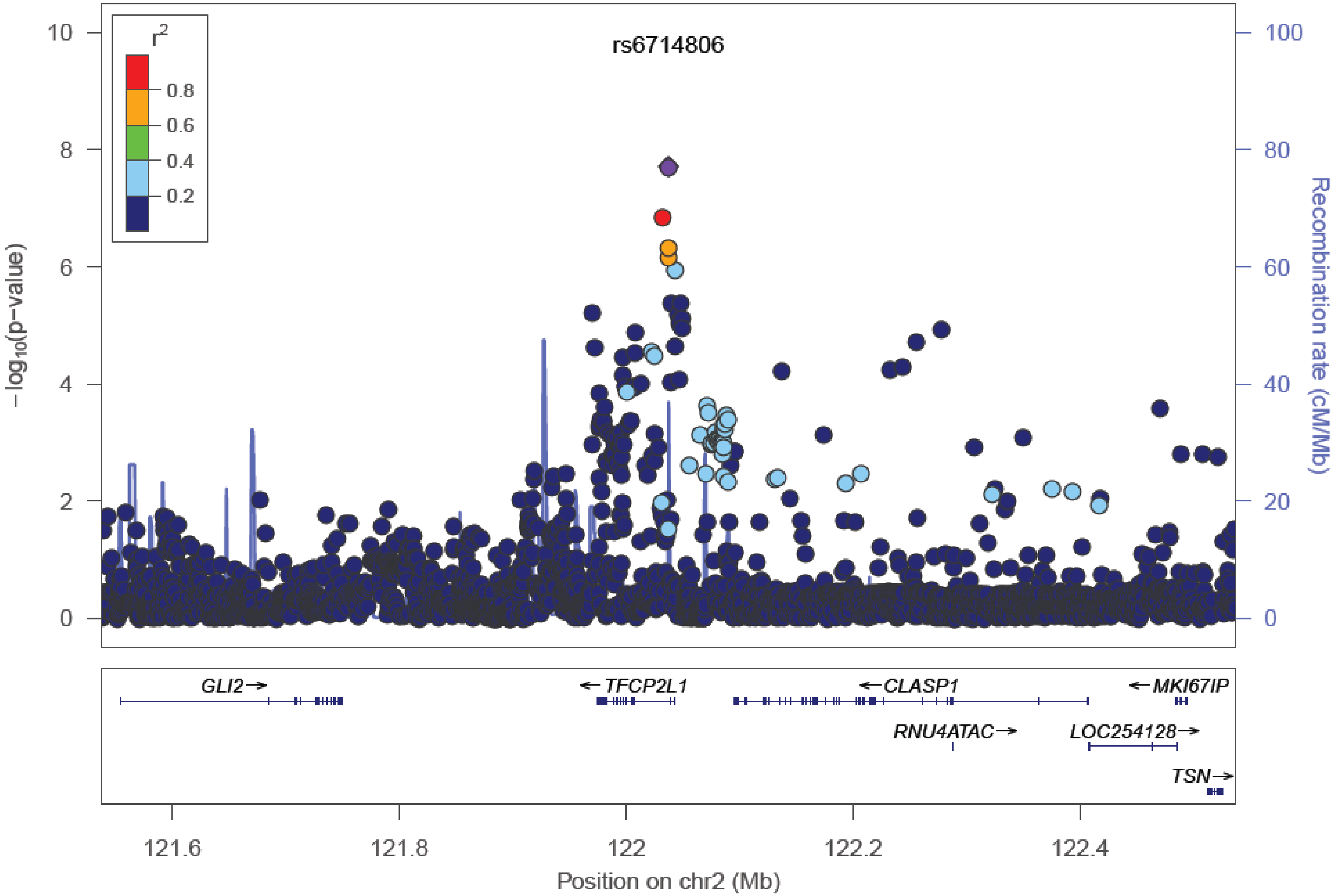
KITLG

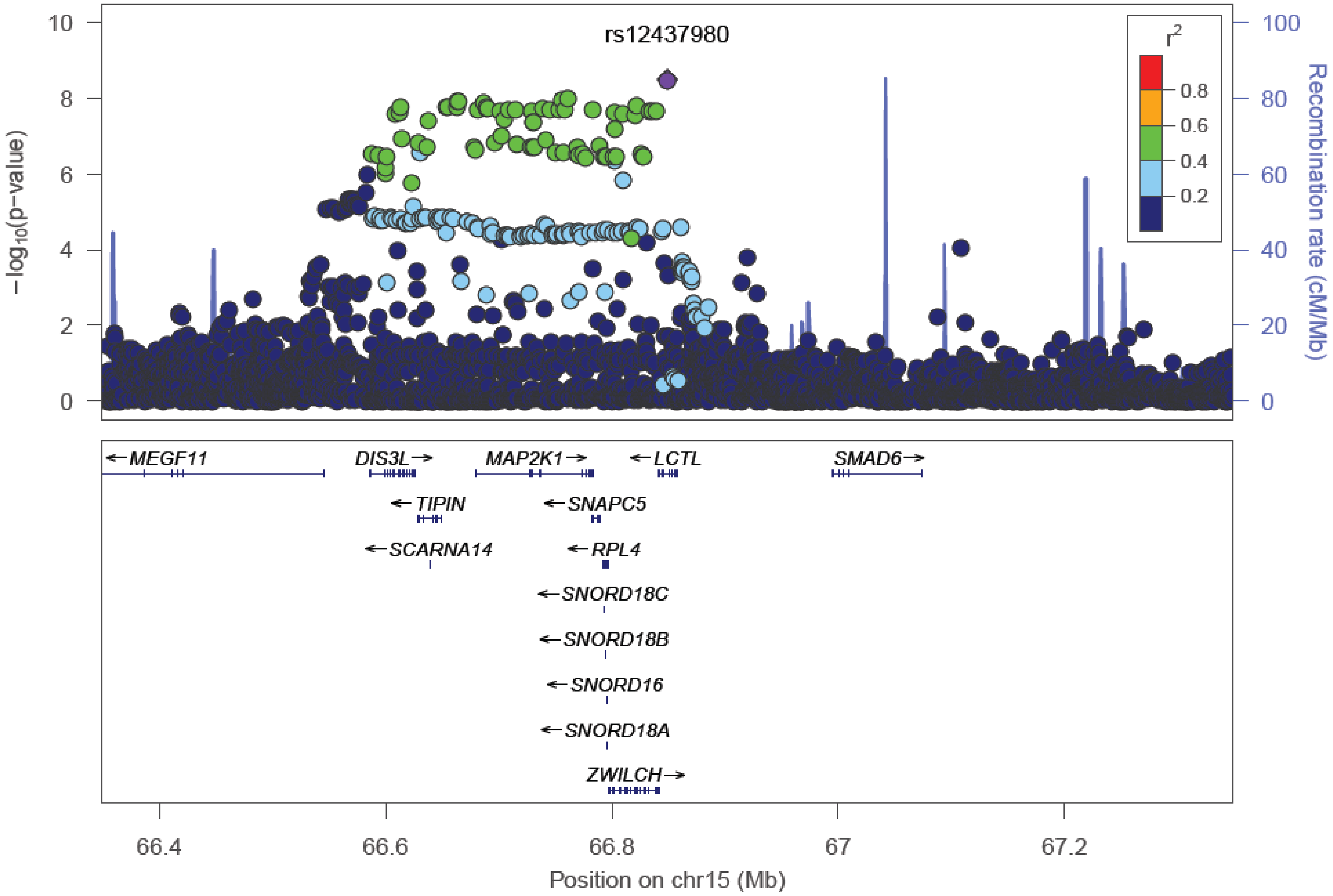

rs10777127
rs4590952
r^2
0.8
0.6
0.4
0.2
−log10(p−value)
50
40
30
20
10
0
100
80
60
40
20
0
Recombination rate (cM/Mb)
←CEP290
TMTC3→
←KITLG
←LOC728084
88.6
88.8
89
89.2
89.4
Position on chr12 (Mb)

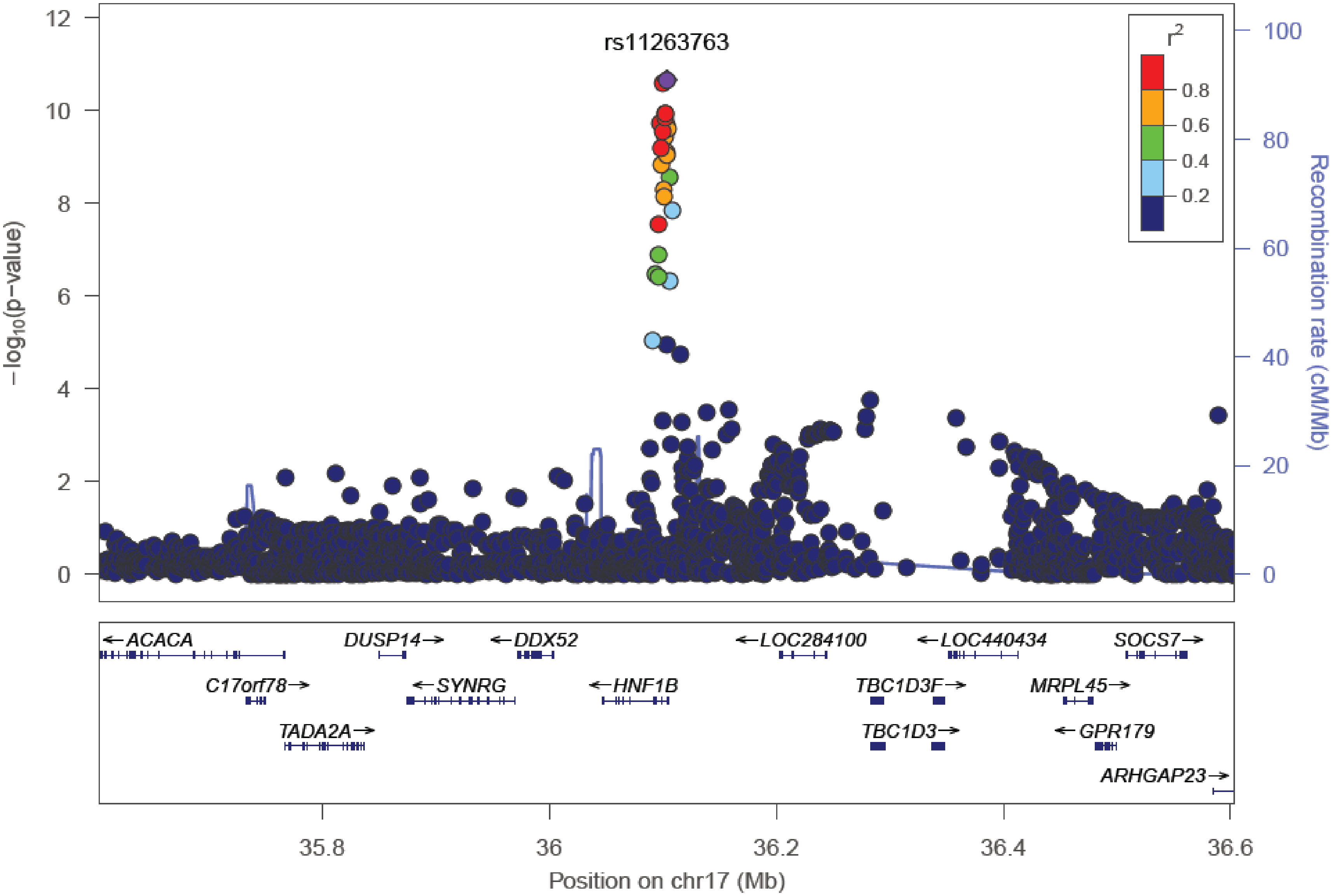
Meta-analysis results

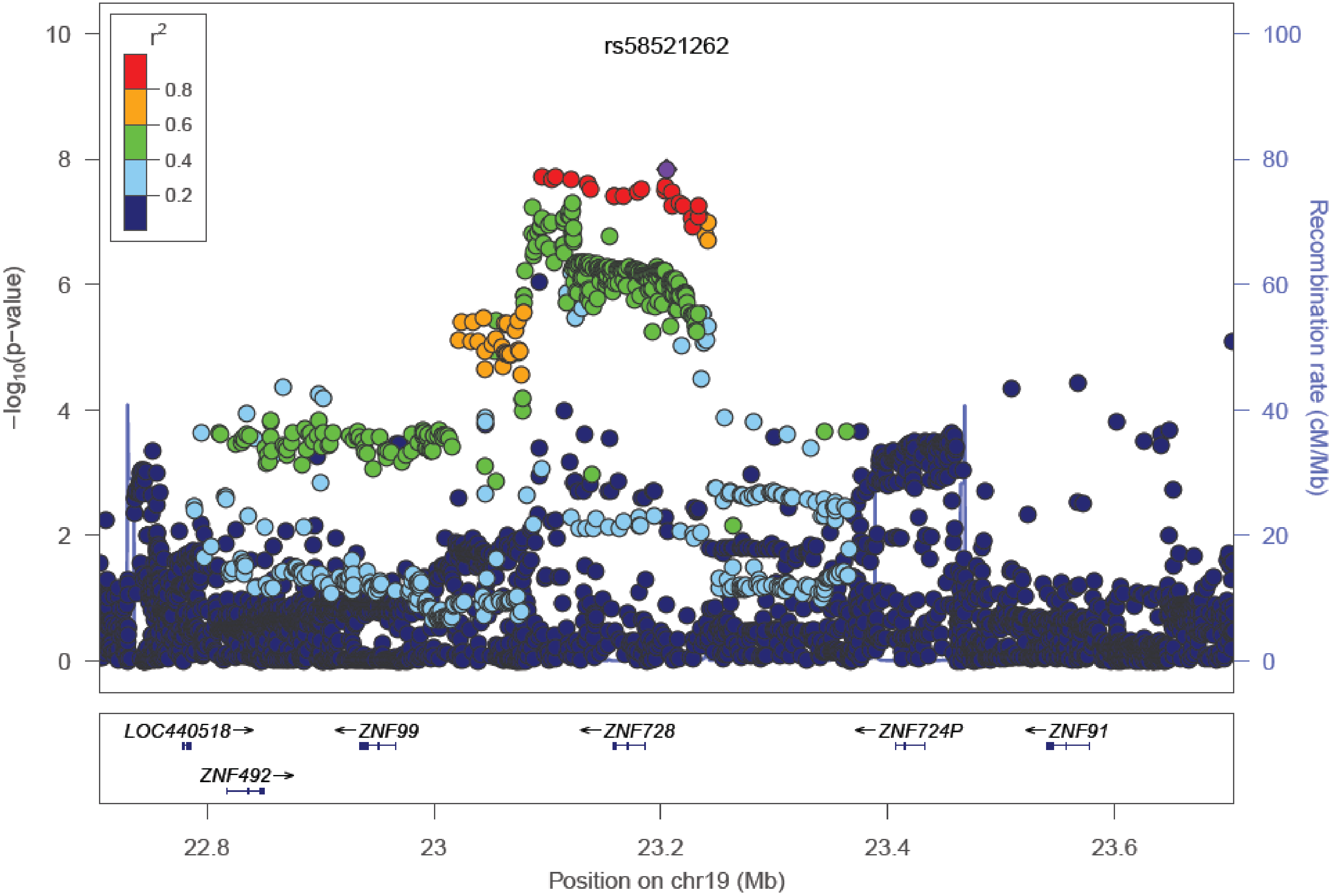
- Known loci
  - 13 of 18 known loci at genome wide significance level
  - Three additional loci at close to genome wide significance level
  - Two loci not identified at near genome wide significance level
- Four novel loci
  - At genome wide significance level

	Cytoband	SNP	Location	OR	CI	P value	Gene
1	2q14.2	rs6714806	122037433	1.33	1.21-1.48	1.93E-08	*TFCP2L1*
2	15q22.31	rs12437980	66848938	1.26	1.17-1.36	3.17E-09	*ZWILCH, LCTL, UNQ3022*
3	17q12	rs11263763	36103565	1.29	1.20-1.39	2.18E-11	*HNF1B*
4	19p12	rs58521262	23205184	0.74	0.66-0.82	1.44E-08	*ZNF728-ZNF730*


rs6714806
r^2
0.8
0.6
0.4
0.2
−log$_{10}$(p−value)
10
8
6
4
2
0
Recombination rate (cM/Mb)
100
80
60
40
20
0
GLI2→
←TFCP2L1
←CLASP1
←MKI67IP
RNU4ATAC→
LOC254128→
TSN→
121.6
121.8
122
122.2
122.4
Position on chr2 (Mb)


rs12437980
r^2
0.8
0.6
0.4
0.2
−log10(p−value)
10
8
6
4
2
0
Recombination rate (cM/Mb)
100
80
60
40
20
0
MEGF11
DIS3L
TIPIN
SCARNA14
MAP2K1
SNAPC5
RPL4
SNORD18C
SNORD18B
SNORD16
SNORD18A
ZWILCH
LCTL
SMAD6
66.4
66.6
66.8
67
67.2
Position on chr15 (Mb)


rs11263763
r^2
0.8
0.6
0.4
0.2
12
10
8
6
4
2
0
−log₁₀(p−value)
100
80
60
40
20
0
Recombination rate (cM/Mb)
←ACACA
DUSP14→
←DDX52
←LOC284100
←LOC440434
SOCS7→
C17orf78→
←SYNRG
←HNF1B
TBC1D3F→
MRPL45→
TADA2A→
TBC1D3→
←GPR179
ARHGAP23→
35.8
36
36.2
36.4
36.6
Position on chr17 (Mb)

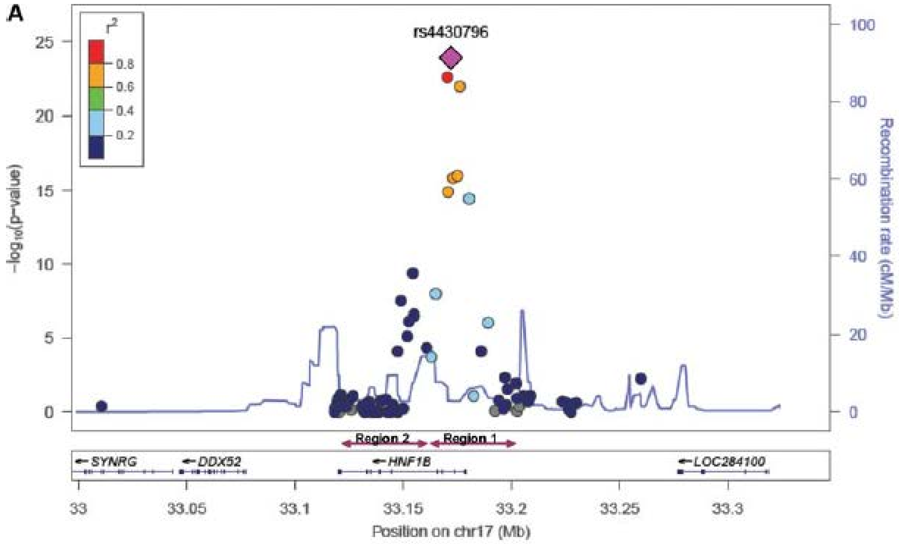

rs58521262
r²
0.8
0.6
0.4
0.2
10
8
6
4
2
0
−log₁₀(p−value)
100
80
60
40
20
0
Recombination rate (cM/Mb)
LOC440518→
ZNF492→
←ZNF99
←ZNF728
←ZNF724P
←ZNF91
22.8
23
23.2
23.4
23.6
Position on chr19 (Mb)

Implicated loci 1

- 2q14.2
  - *CLASP1* - regulation of microtubule dynamics at the kinetochore and throughout the spindle
  - *TFCP2L1* (LBP9, CRTR-1) – developmentally regulated transcription factor, maturation salivary glands and kidney, maintains pluripotency in ES cells
- 15q22.31
  - Five genes, seven non-coding RNAs
  - MAP2K1 (MEK1) – essential component of the MAPK signaling pathway
  - ZWILCH – kinetochore protein, part of complex (RZZ) essential for recruitment of MAD1/2 to the kinetochore
  - TIPIN - part of the replisome complex, a group of proteins that support DNA replication

Implicated loci 2

- 17q12
  - HNF1B – transcription factor
  - Mutations associated with familial multicystic dysplastic kidneys, renal failure, MODY5



  - Variation associated with increased risk of prostate cancer, type 2 diabetes, invasive serous and epithelial ovarian cancer, endometrial cancer
- 19p12
  - ZNF728, ZNF724P

Berndt et al., Hum Molec Genet 2011

GWAS significant loci

Cytoband	SNP	Location	OR	CI	P value	Gene
12q21.32	rs10777127	88954295	2	1.82-2.19	4.68E-49	*KITLG*
9p24.3	rs7863616	861220	1.6	1.49-1.73	4.62E-34	*DMRT1*
5q31.3	rs6580240	141684389	1.5	1.42-1.64	5.14E-30	*SPRY4*
5p15.33	rs35953391	1312329	1.7	1.51-1.80	7.17E-29	*CLPTM1L*
6p21.31	rs210138	33542538	1.6	1.43-1.71	7.40E-23	*BAK1*
7p22.3	rs13226911	1885144	1.4	1.25-1.48	5.75E-13	*MAD1L1*
17q22	rs80030779	56765478	0.8	0.71-0.83	4.09E-12	*TEX14, RAD51C*
4q24	rs2711897	104023567	0.8	0.72-0.83	1.98E-11	*BDH2, UNQ6308, CENPE*
17q12	**rs11263763**	**36103565**	**1.3**	**1.20-1.39**	**2.18E-11**	***HNF1B***
3p24.3	rs11713721	16616628	0.8	0.73-0.85	1.12E-10	*DAZL*
1q24.1	rs10800166	165895525	1.3	1.18-1.39	2.95E-09	*UCK2*
15q22.31	**rs12437980**	**66848938**	**1.3**	**1.17-1.36**	**3.17E-09**	***ZWILCH, LCTL, UNQ3022***
16q12.1	rs8046148	48700445	1.3	1.18-1.42	8.31E-09	*HEATR3*
19p12	**rs58521262**	**23205184**	**0.7**	**0.66-0.82**	**1.44E-08**	***ZNF728-ZNF730***
16q23.1	rs7188581	73227959	0.8	0.75-0.87	1.53E-08	*RFWD3*
2q14.2	**rs6714806**	**122037433**	**1.3**	**1.21-1.48**	**1.93E-08**	***TFCP2L1***
4q22.3	rs6840913	95215892	0.8	0.74-0.87	3.16E-08	*SMARCAD1, HPGDS*

Six additional novel loci just below genome wide significance

Stratified analyses

- Based on discussion of the meta-analysis working group
- Cryptorchidism
- Tumor Histology
  - Seminoma
  - NSGCT (with and without combined S/NSGCT)
- Family history
  - With and without bilateral disease

Analysis notes

- Stratified analyses
  - Either raw/imputed data for both cases and controls, along with phenotypic data has to go to NCI
  - Or stratified analyses are done locally and aggregate data sent to NCI
- Additional analyses (e.g. pathway, rare variant) will be reviewed by Peter

Timeline for completing meta-analysis

- Potential for inclusion of additional genome wide data sets
- Need to have all data (main and stratified) by XXX date
- Timeline based on grant budget
  - Need to order chips within this fiscal year to start to expend genotyping resources budgeted
  - Plan was for iSelect (2K novel, 500 old loci, 500 AIMs) but will review current possibilities (input welcome)