Mapping pathogenic regulatory regions and genes

Chris Cotsapas

Yale/Broad

Mapping pathogenic regulatory regions

Chris Cotsapas

Yale/Broad

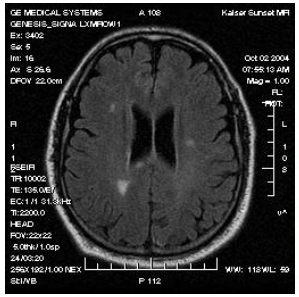
Mapping pathogenic regulatory regions

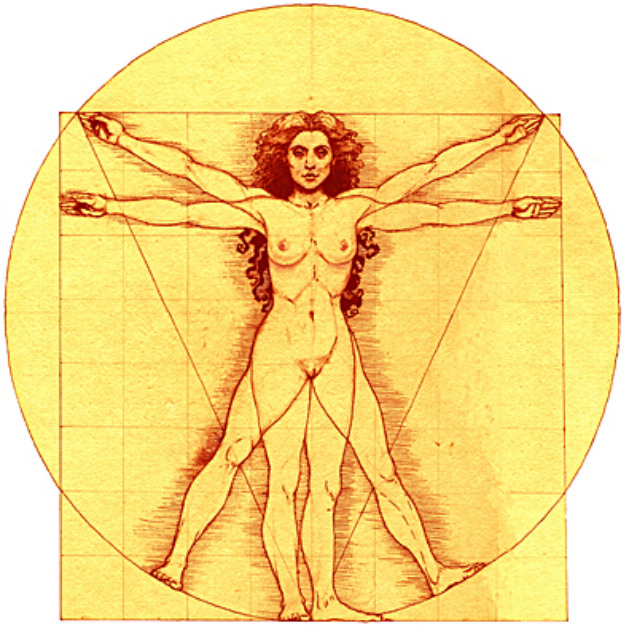
Chris Cotsapas

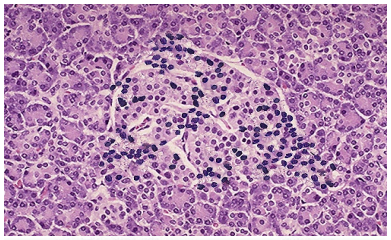
Yale/Broad

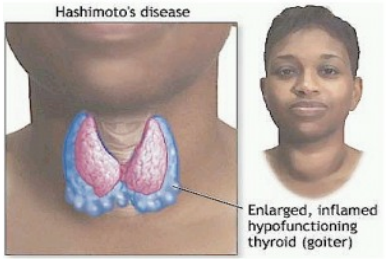
cotsapaslab.info/positions

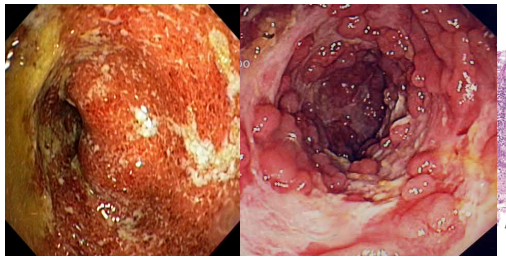
http://biorxiv.org/content/early/2016/05/19/054361

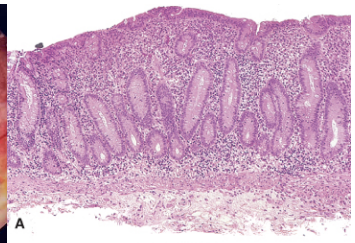




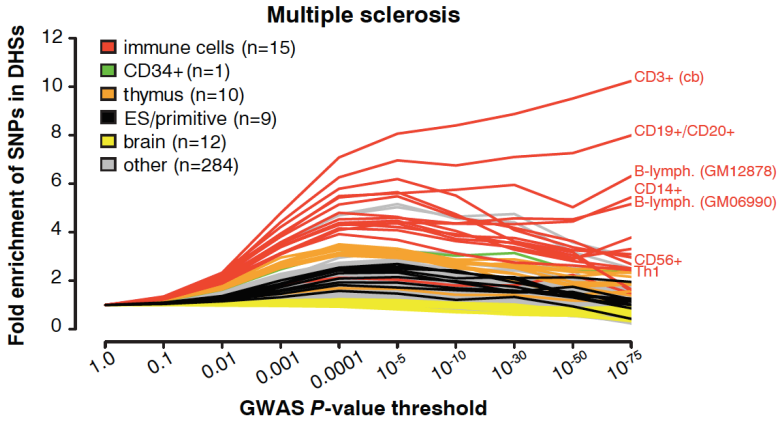




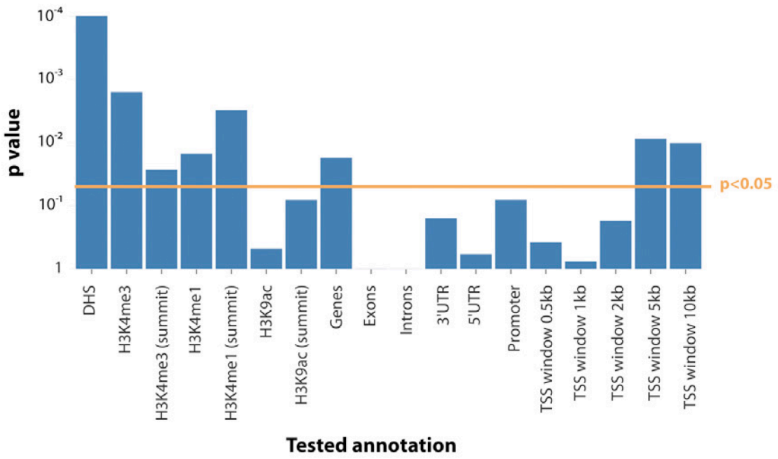




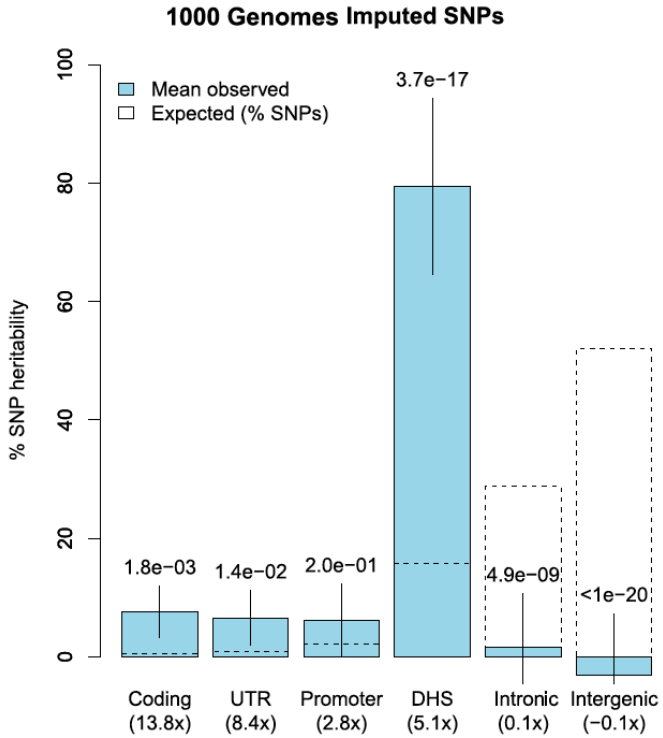
Common risk variants localize to DHS



Maurano et. al, Science 2012

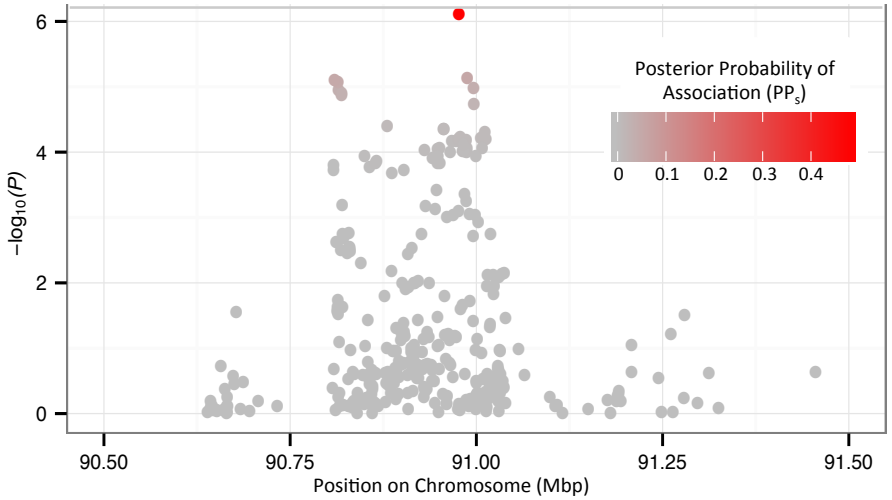


Trynka et al, AJHG 2015

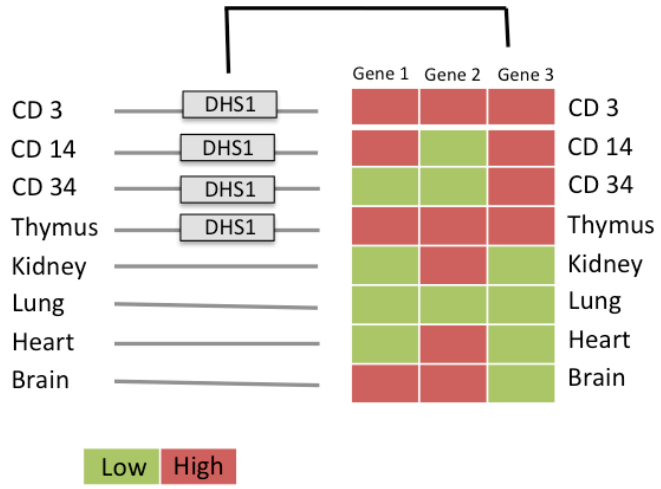


Gusev et. al, AJHG 2014

Regulatory fine-mapping model



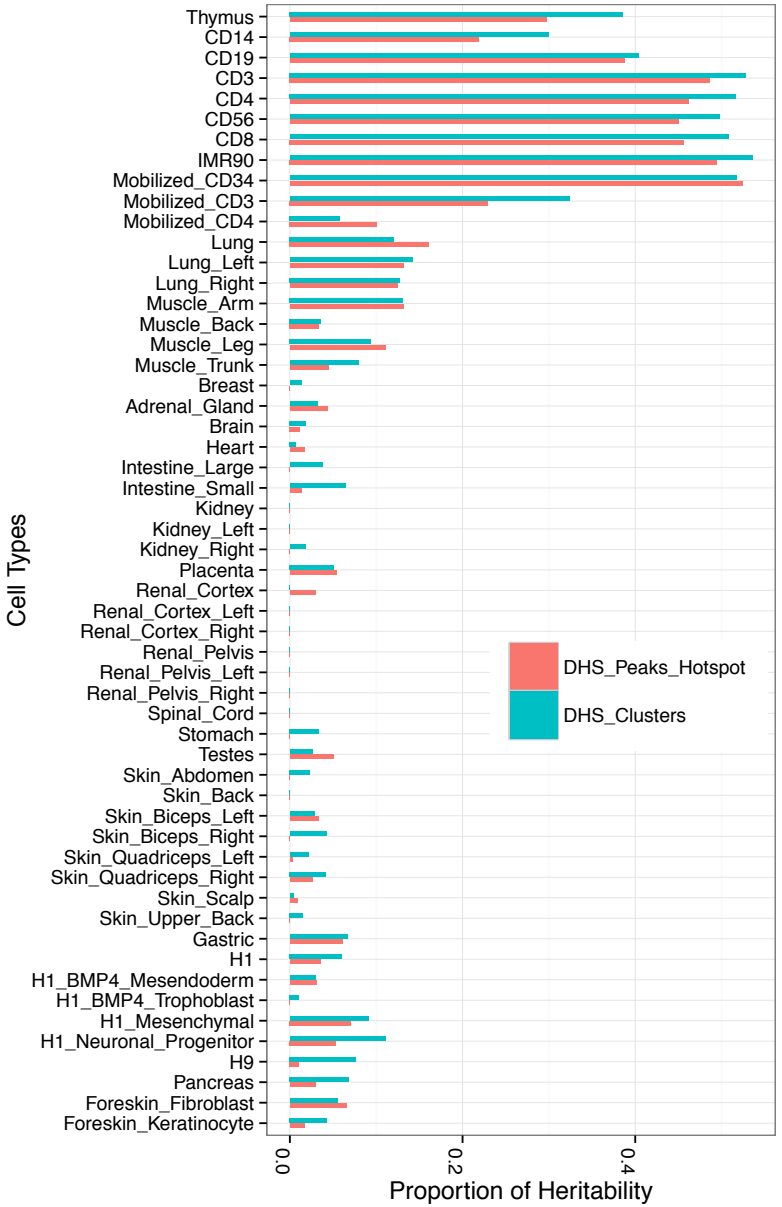
Summary *P* va
association da



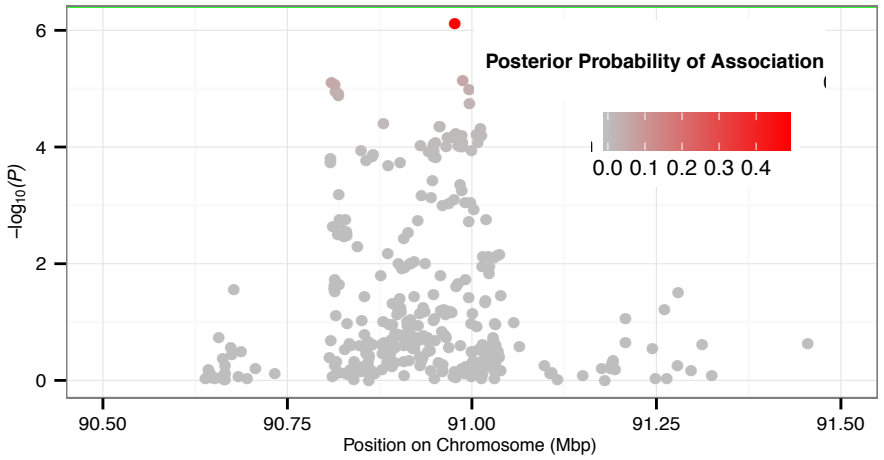
Aligning DHSs Over Cell Types

1,079,138/1,994,675 clusters (~54%) pass QC

8% of genome (cf. 14% all peaks)



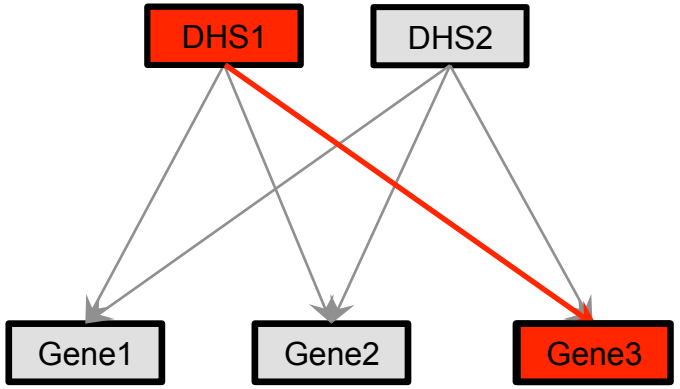
Posterior probabilities of association



SNPs

$$PPS_i$$

DHSs

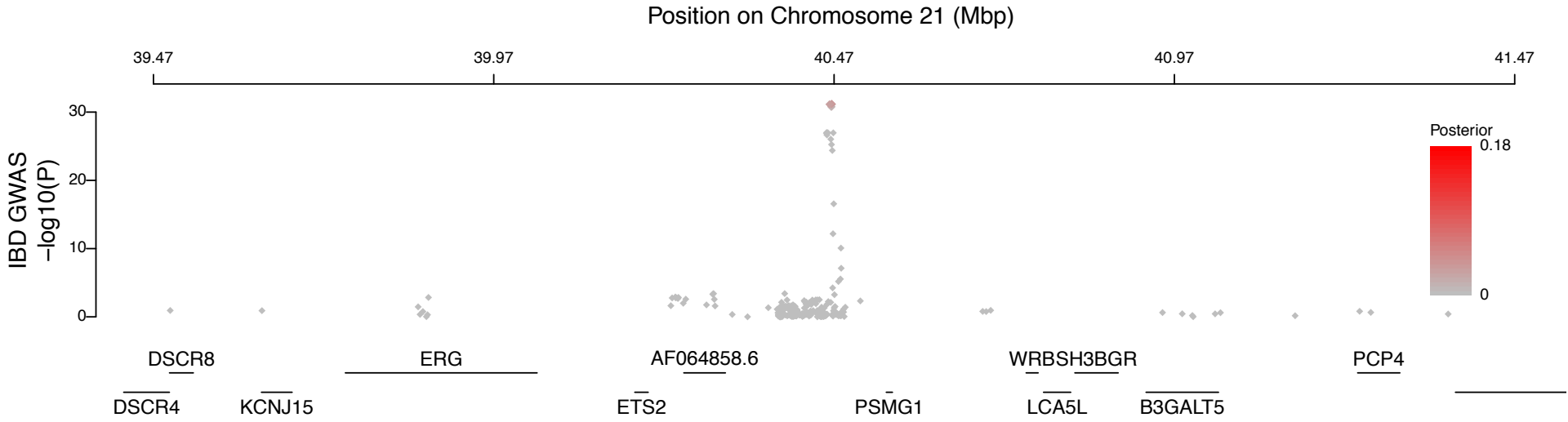


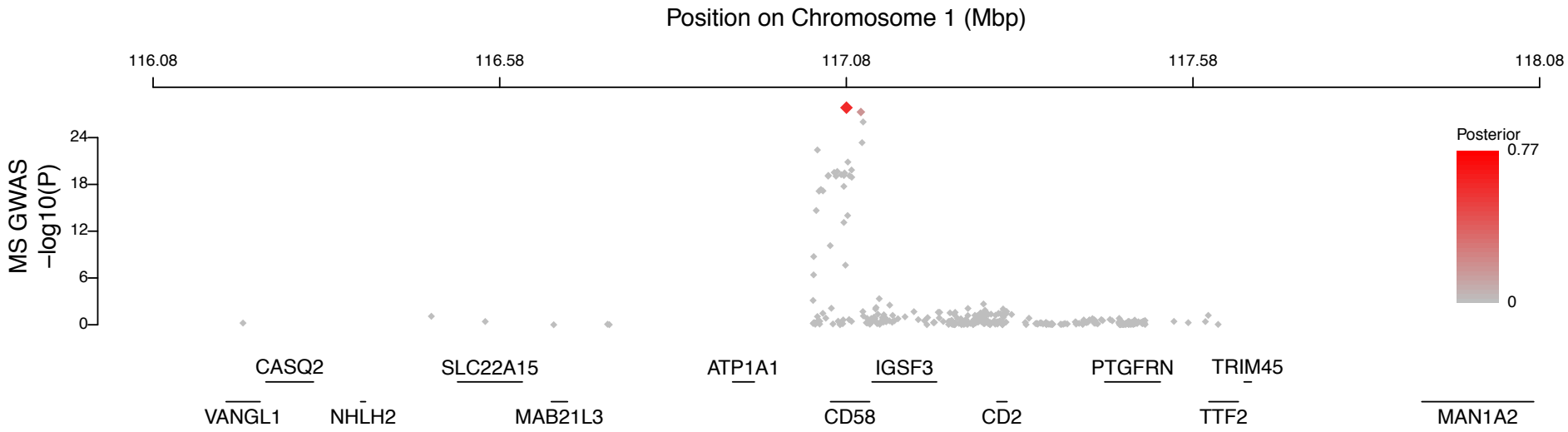
$$PPD_d = \sum_{s=1;\, s \subseteq i} PPS_s$$

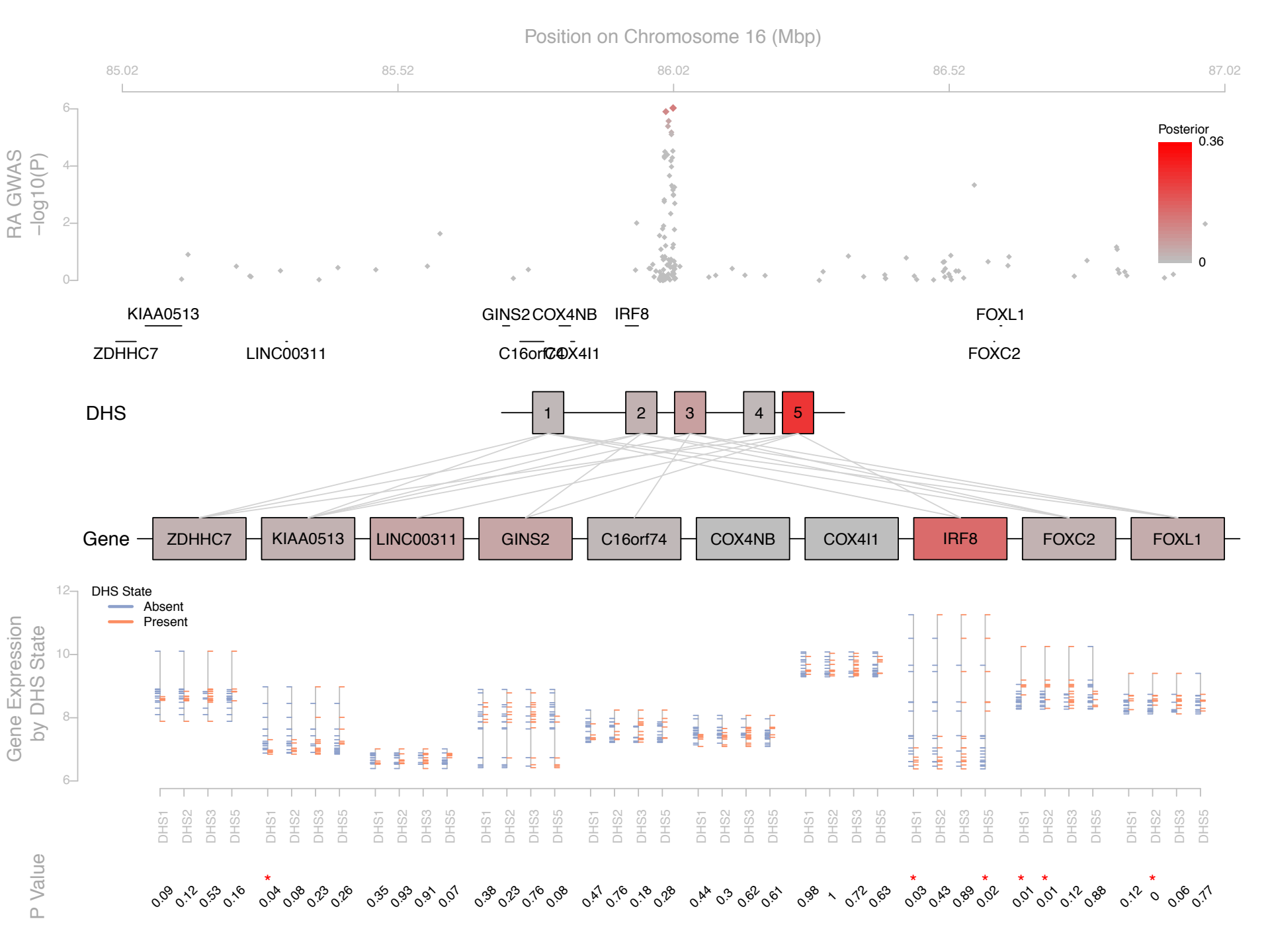
$$PPC_{d,g} = \frac{-log_{10}\, p_{d,g}}{\sum_G -log_{10} p_{d,G}}$$

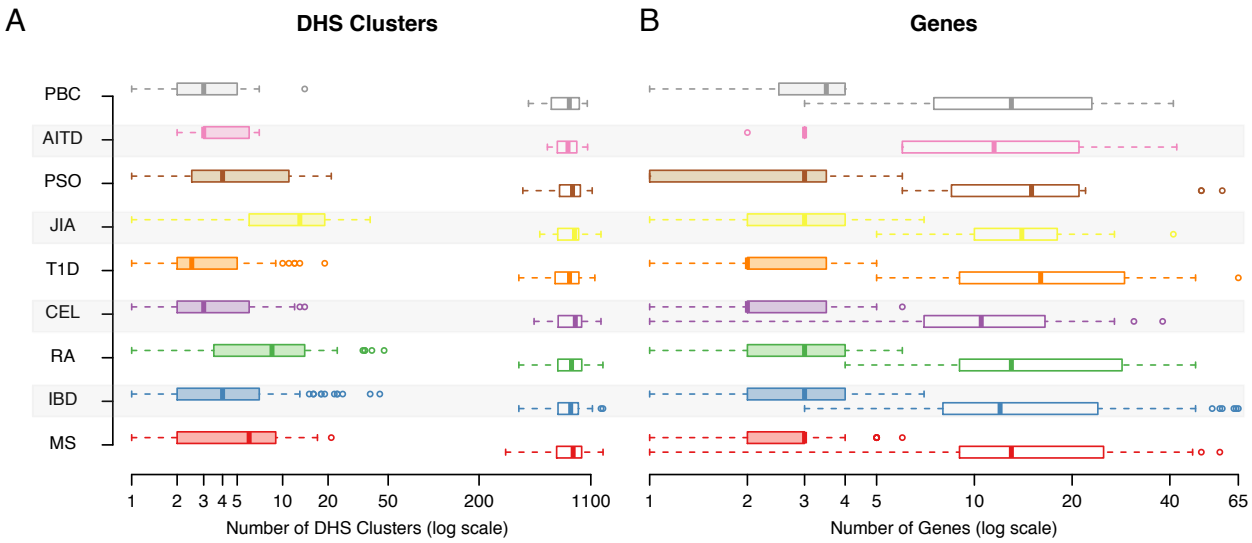
Genes

$$PPG_g = \sum_d PPD_d \times PPC_{d,g}$$

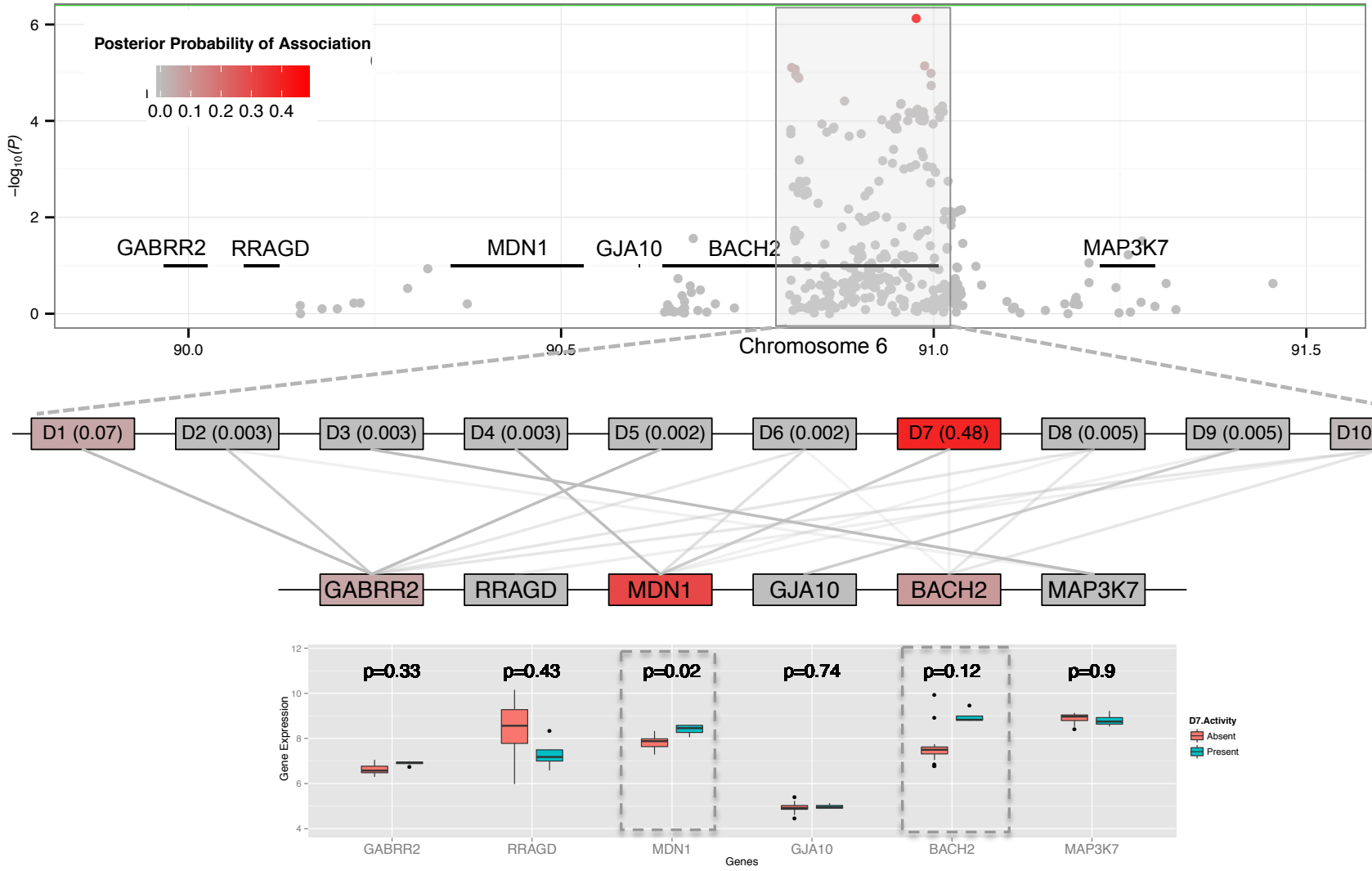

Position on Chromosome 21 (Mbp)
39.47
39.97
40.47
40.97
41.47
IBD GWAS
−log10(P)
30
20
10
0
Posterior
0.18
0
DSCR8
ERG
AF064858.6
WRBSH3BGR
PCP4
DSCR4
KCNJ15
ETS2
PSMG1
LCA5L
B3GALT5


Position on Chromosome 1 (Mbp)
116.08
116.58
117.08
117.58
118.08
MS GWAS
−log10(P)
24
18
12
6
0
Posterior
0.77
0
CASQ2
SLC22A15
ATP1A1
IGSF3
PTGFRN
TRIM45
VANGL1
NHLH2
MAB21L3
CD58
CD2
TTF2
MAN1A2

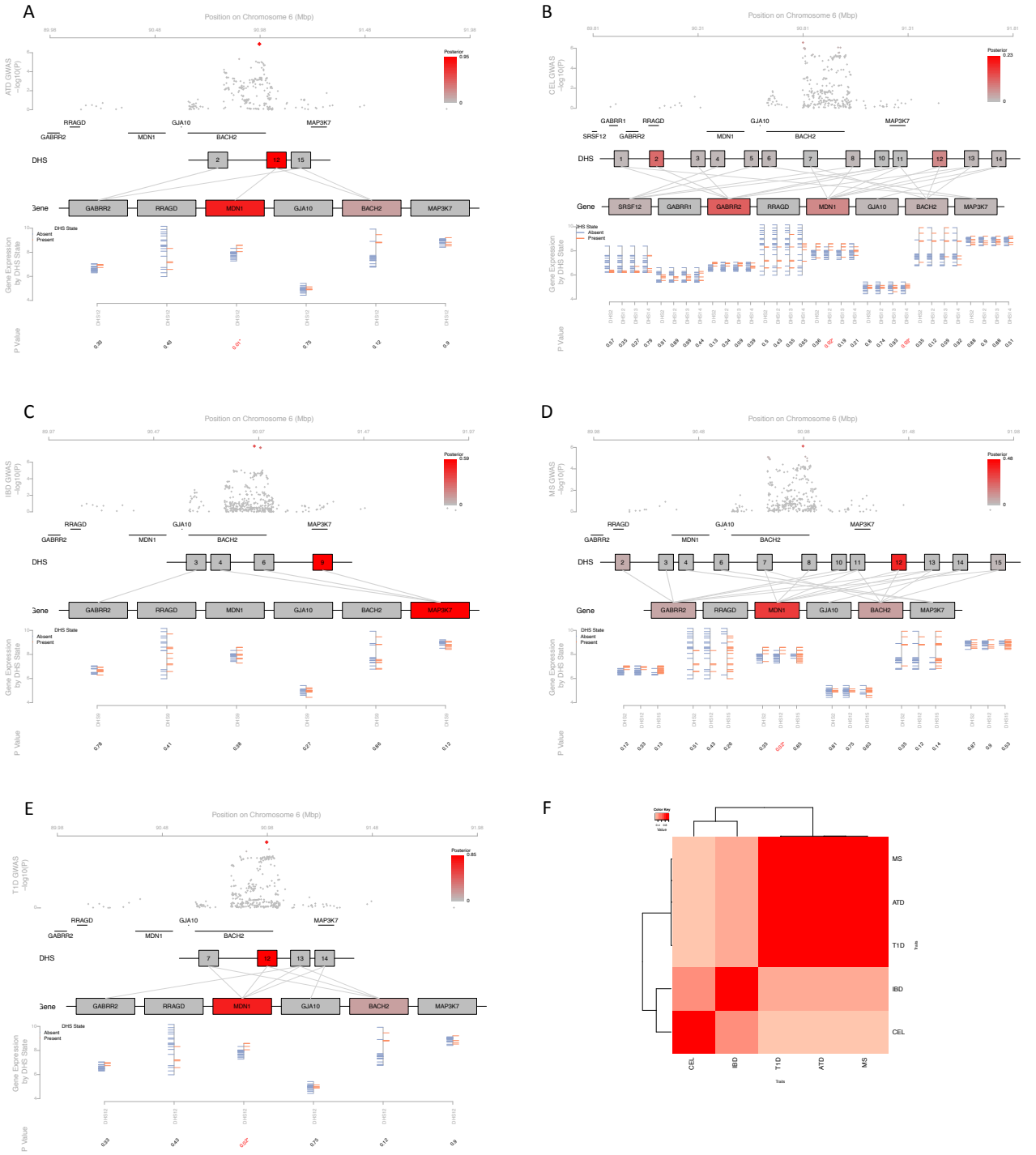

Position on Chromosome 16 (Mbp)
85.02
85.52
86.02
86.52
87.02
RA GWAS
−log10(P)
6
4
2
0
Posterior
0.36
0
KIAA0513
GINS2
COX4NB
IRF8
FOXL1
ZDHHC7
LINC00311
C16orf74
COX4I1
FOXC2
DHS
1
2
3
4
5
Gene
ZDHHC7
KIAA0513
LINC00311
GINS2
C16orf74
COX4NB
COX4I1
IRF8
FOXC2
FOXL1
Gene Expression
by DHS State
12
10
8
6
DHS State
Absent
Present
DHS1
DHS2
DHS3
DHS5
P Value
0.09 0.12 0.53 0.16 * 0.04 0.08 0.23 0.26 0.35 0.93 0.91 0.07 0.38 0.23 0.76 0.08 0.47 0.76 0.18 0.28 0.44 0.3 0.62 0.61 0.98 1 0.72 0.63 * 0.03 0.43 0.89 * 0.02 * 0.01 * 0.01 0.12 0.88 0.12 * 0 0.06 0.77


A
DHS Clusters
B
Genes
PBC
AITD
PSO
JIA
T1D
CEL
RA
IBD
MS
1
2
3
4
5
10
20
50
200
1100
Number of DHS Clusters (log scale)
1
2
3
4
5
10
20
40
65
Number of Genes (log scale)

BACH2 locus (chr6:91Mb) in MS



IBD (MAP3K7)


A
B
C
D
E
F
Position on Chromosome 6 (Mbp)
ATD GWAS −log10(P)
CEL GWAS −log10(P)
IBD GWAS −log10(P)
MS GWAS −log10(P)
T1D GWAS −log10(P)
Posterior
DHS
Gene
GABBR2
RRAGD
MDN1
GJA10
BACH2
MAP3K7
SRSF12
GABBR1
Gene Expression by DHS State
DHS State
Absent
Present
P Value
MS
ATD
T1D
IBD
CEL
Trait

	Concordance	Discordance	Jaccard Coefficient
Most Associated SNPs	6	45	0.12
CI SNPs (mean)	6.8	31.16	0.21
Prioritized CI SNPs (mean)	2.2	9.47	0.25
Prioritized DHS Clusters (mean)	2.47	8.51	0.27
Prioritized Genes	16	35	0.31

	Concordance	Discordance	Jaccard Coefficient
Most Associated SNPs	**6**	**45**	**0.12**
CI SNPs (mean)	6.8	31.16	0.21
Prioritized CI SNPs (mean)	2.2	9.47	0.25
Prioritized DHS Clusters (mean)	2.47	8.51	0.27
Prioritized Genes	**16**	**35**	**0.31**

Fisher's exact test **P = 0.001**