Joint Bayesian inference of risk variants and tissue-specific epigenomic enrichments across multiple complex human diseases

Yue Li[1,2,*] and Manolis Kellis[1,2,*]

[1]Computer Science and Artificial Intelligence Lab, Massachusetts Institute of Technology, 32 Vassar St, Cambridge, Massachusetts 02139, USA
[2]The Broad Institute of Harvard and MIT, 415 Main Street, Cambridge, Massachusetts 02142, USA
[*]Correspondence to liyue@mit.edu or manoli@mit.edu

Contents

List of Figures

List of Tables

1 Supplementary Figures

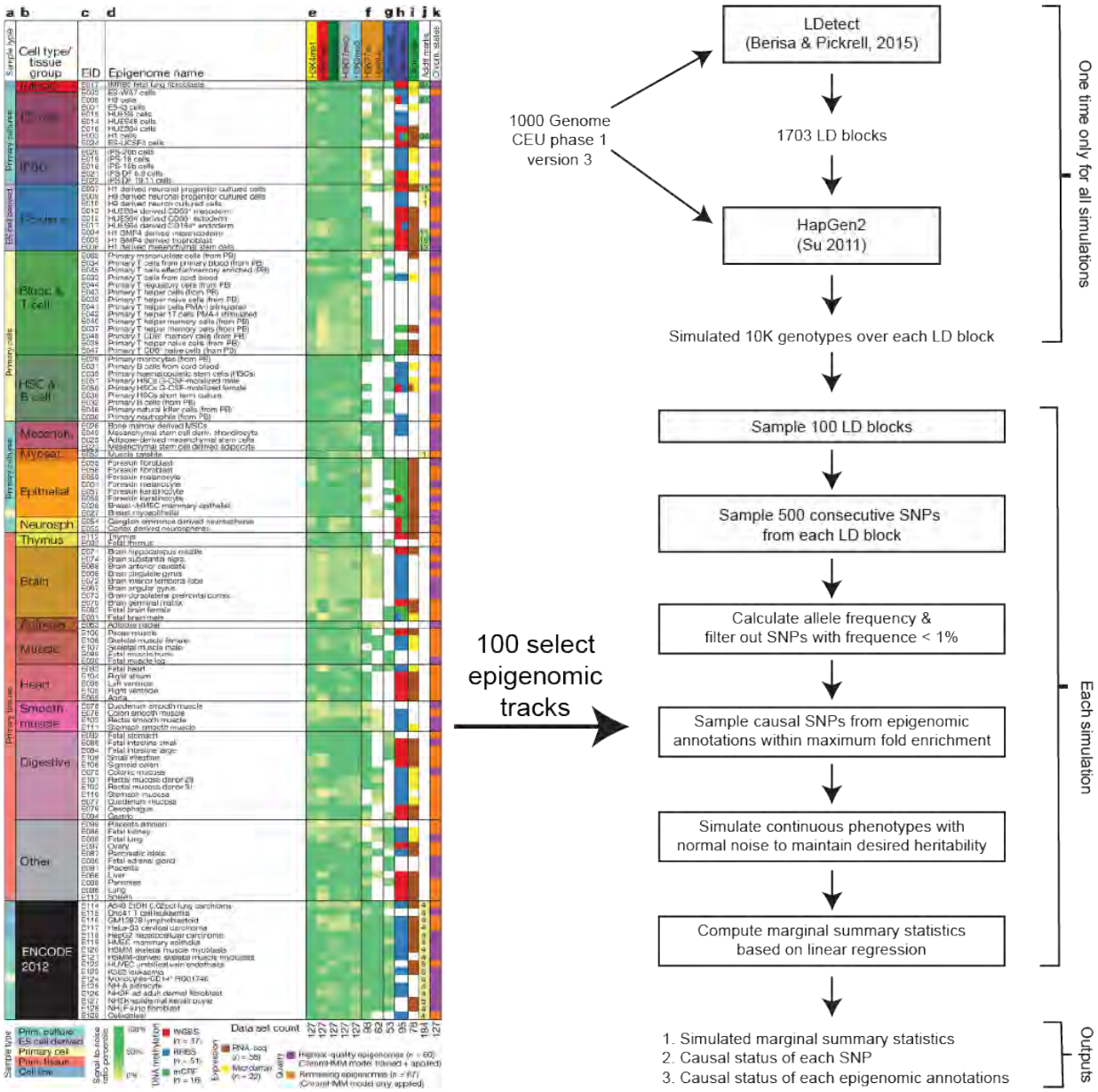


Fig. S1: Simulation pipeline to generate genotypes. HapGen2 [1] was used to sample haplotypes from 1000 Genome CEU data. For computational efficiency, we applied HapGen2 separately to each LD blocks approximated by LDetect [2]. We simulated 10,000 genotypes for each LD block. At each simulation run, we sampled 100 LD blocks each containing at maximum 500 SNPs to simulate risk loci for the 10,000 individuals. The causal variants were randomly sampled from the epigenomic annotations with predefined fold-enrichment until reaching the requried total number of causal variants. The continuous phenotypes were then simulated as a linear combination of the normalized genotypes weighted by the effect sizes of the causal variants. Finally, least-square linear regression on the 10,000 continuous phenotypes was performed for each variant to generate summary statistics in terms of both p-values and t-statistics that were input to the fine-mapping methods in comparison.

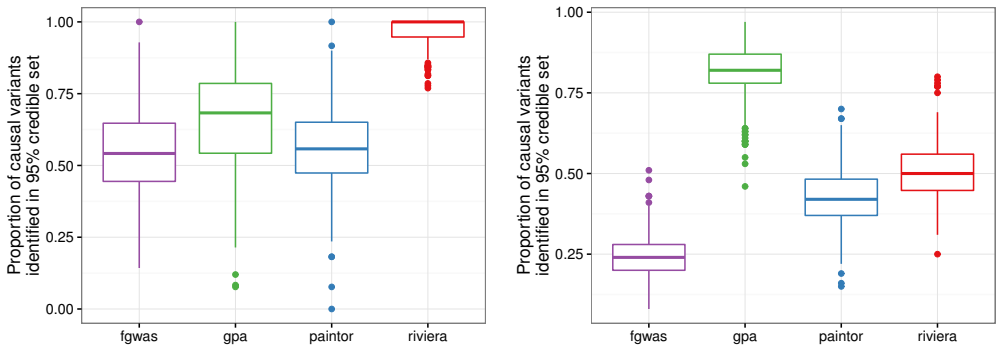


Fig. S2: Credible set calibration on simulated datasets with one variant per locus (left) and more than one causal variants per locus (right).

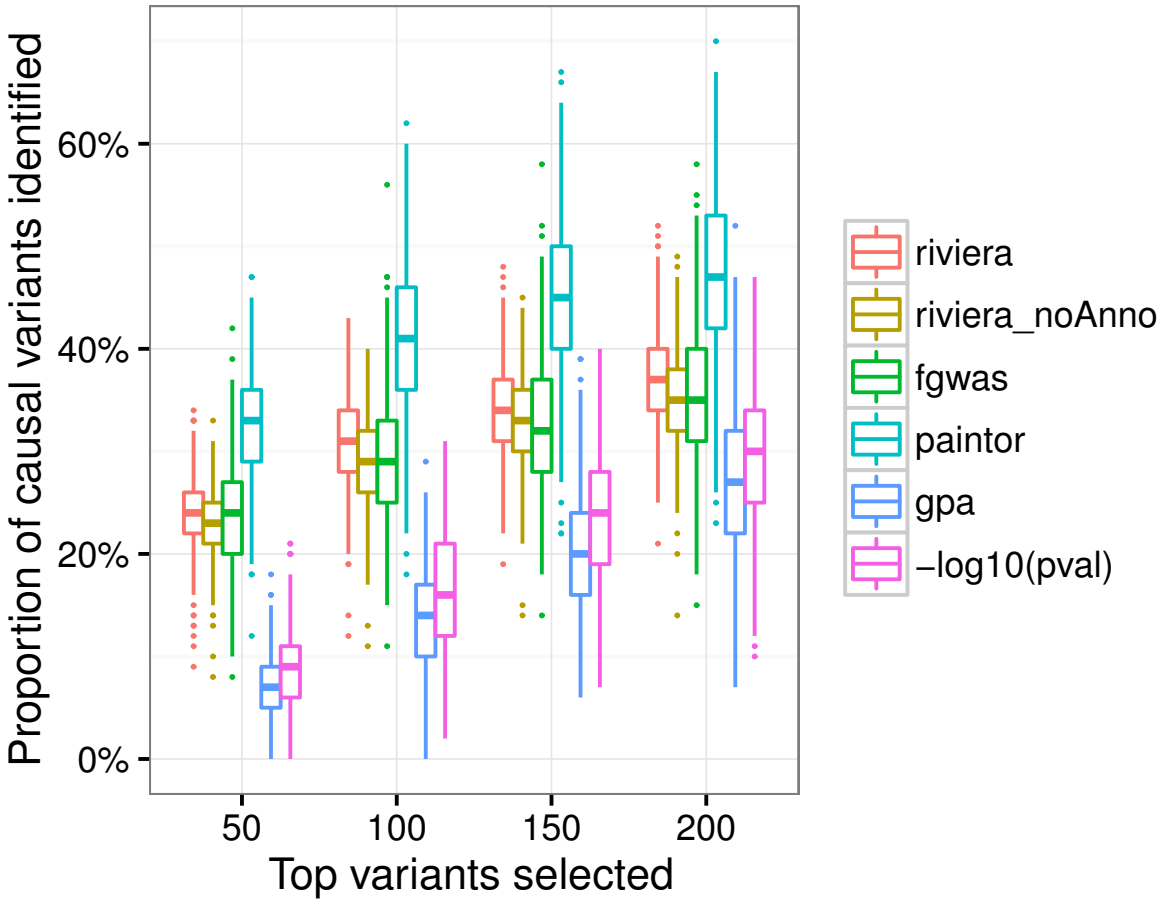


Fig. S3: Model performance on loci harboring multiple causal variants

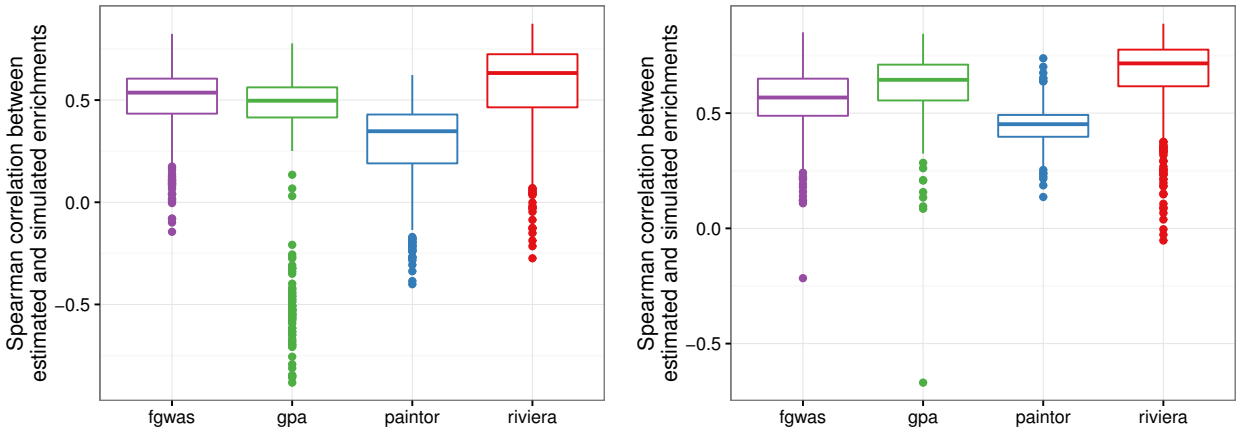


Fig. S4: Spearman correlation between estimated and simulated fold-enrichment. Left: single-causal variant per locus; Right: one or more causal variant per locus

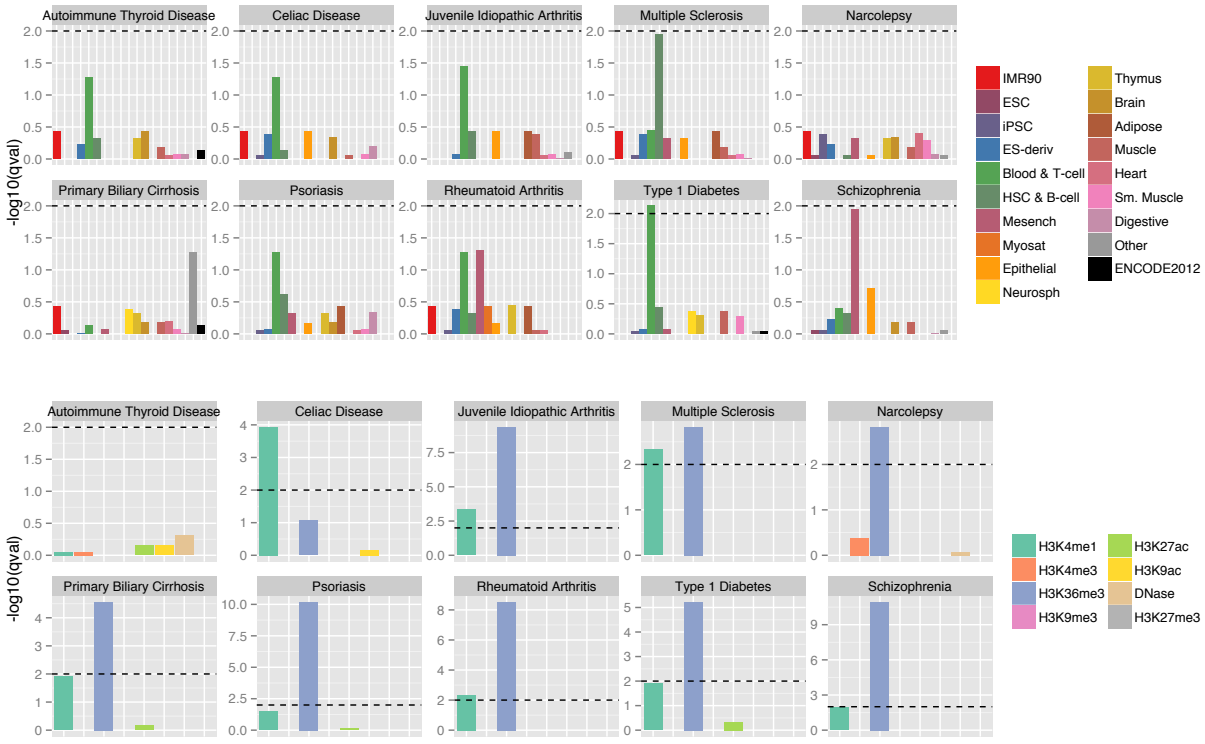
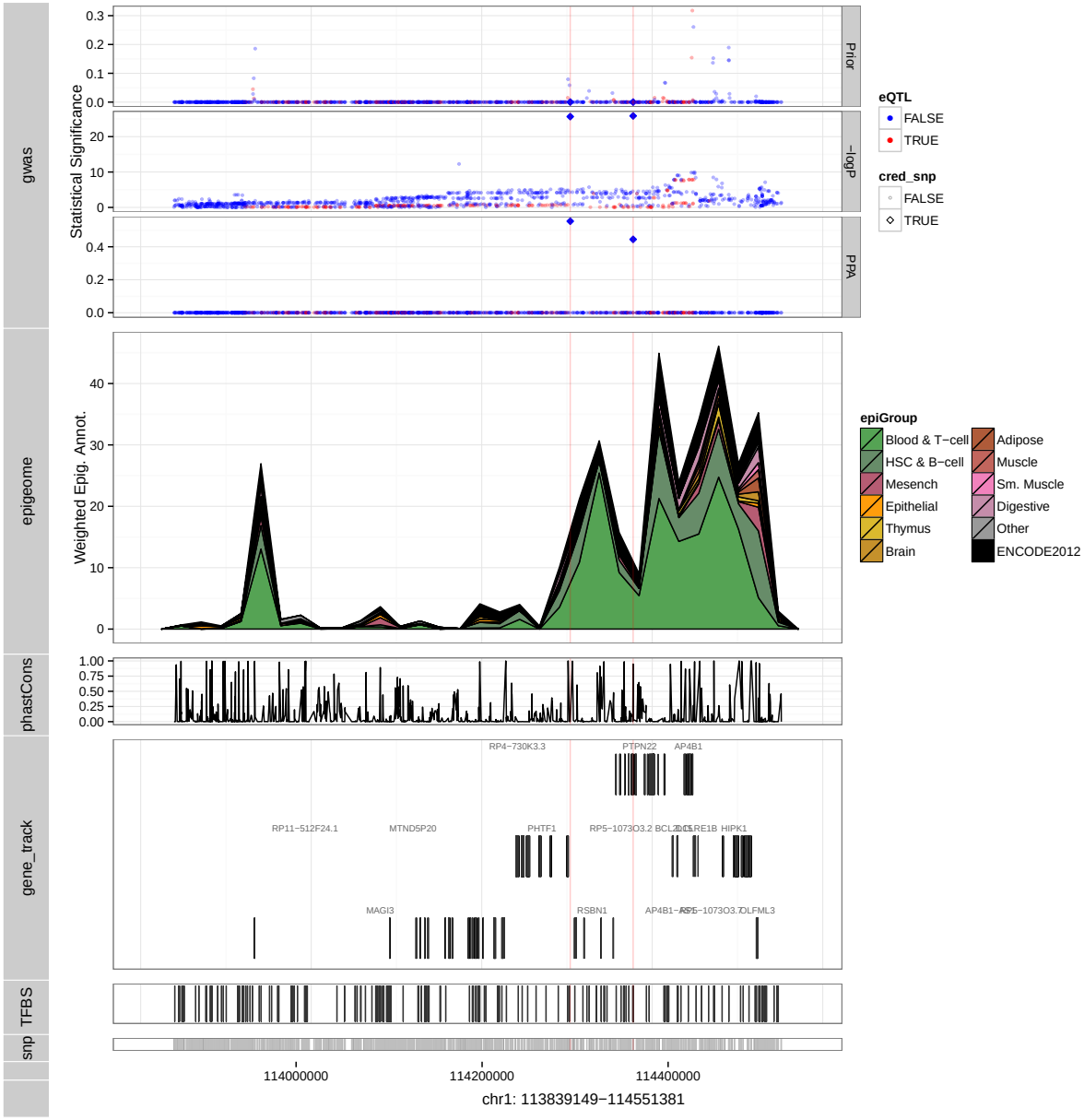
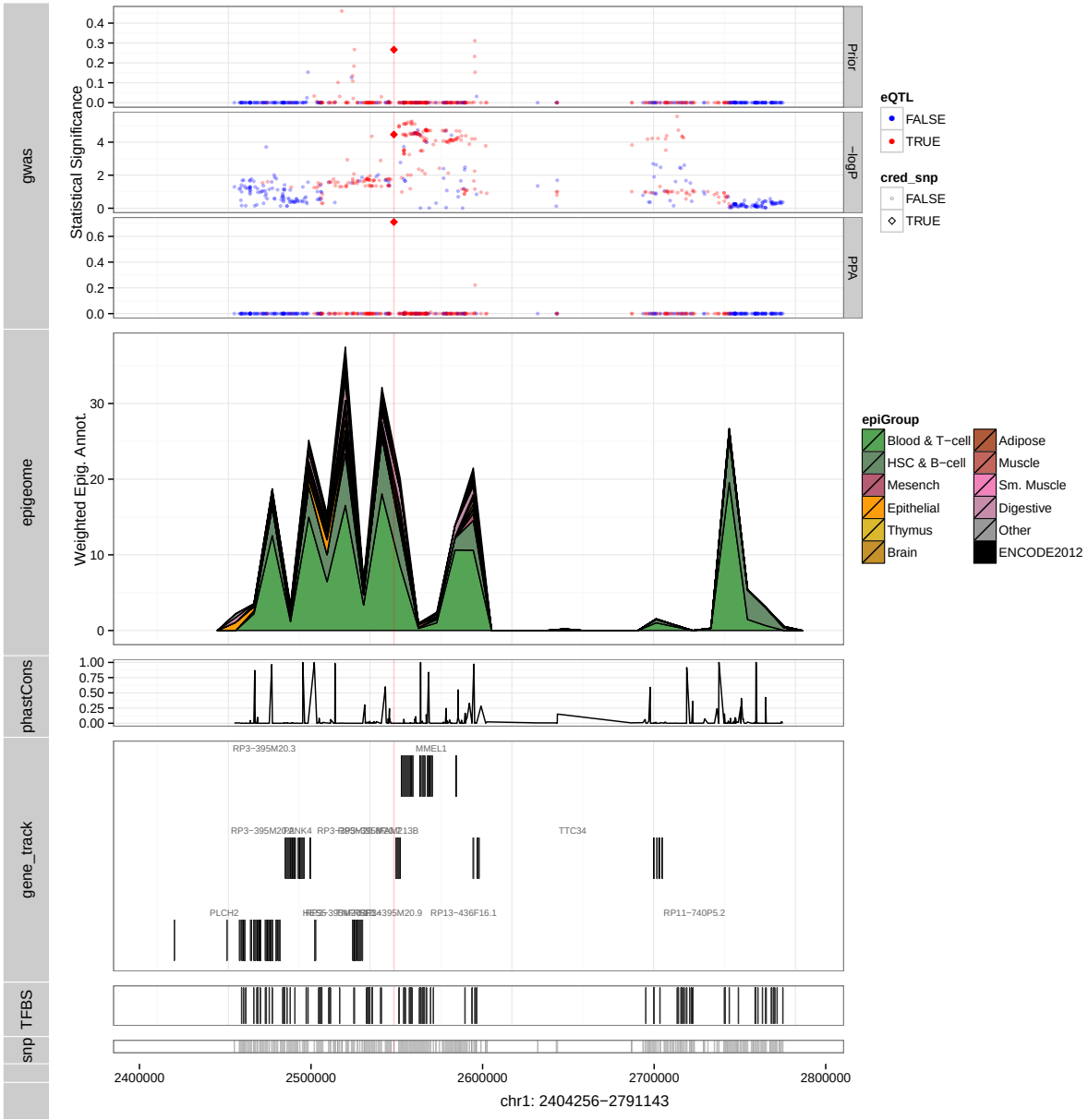


Fig. S5: Enrichments on permuted data. To further ascertain the enrichment results observed in Fig. 3, we permuted the data by randomly shuffling the SNPs and annotations and repeating the same enrichment tests on 19 tissue groups (upper panels) and 8 epigenomic marks (lower panels). Please refer to Fig. 3 for more details.

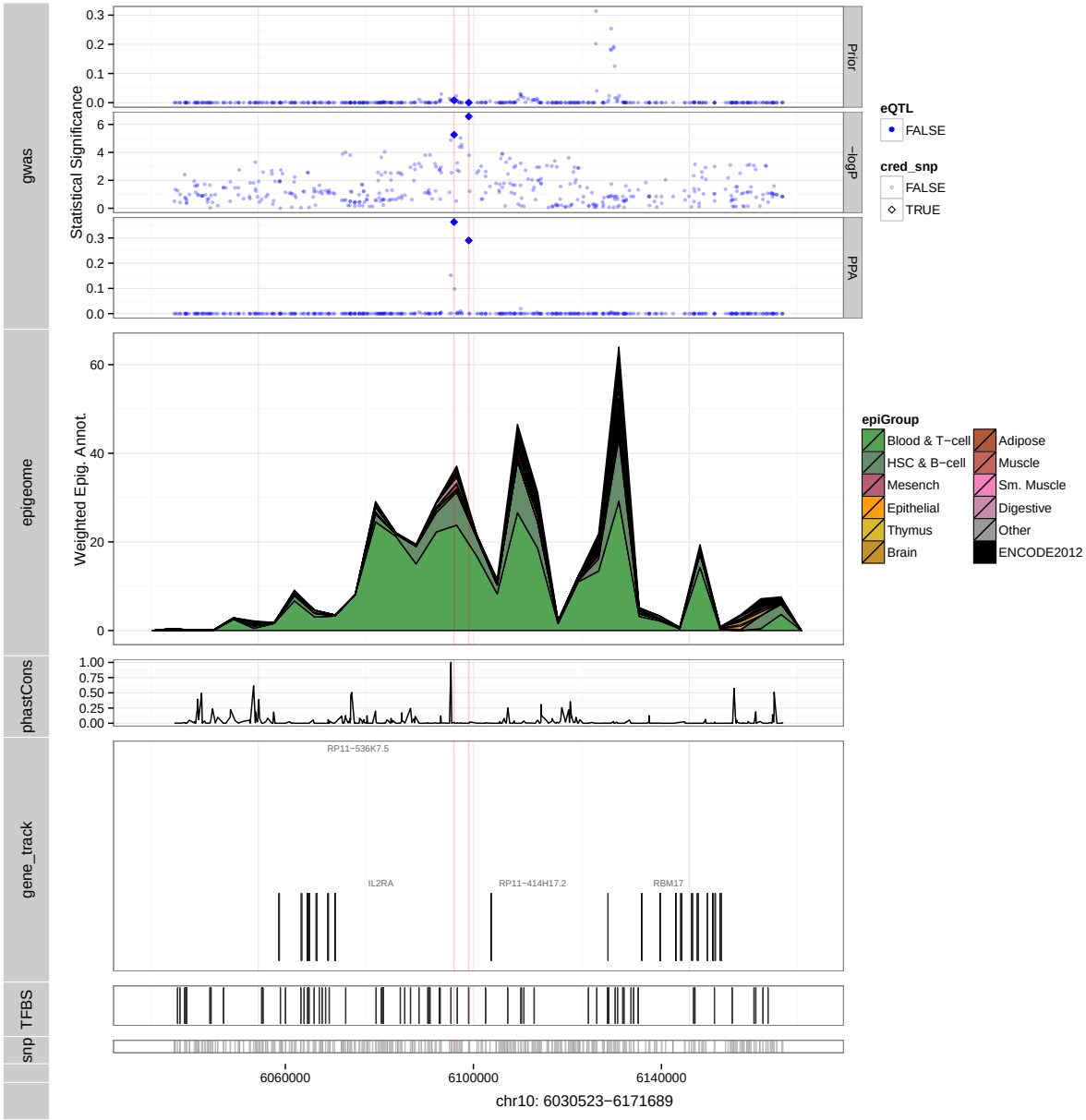
Autoimmune Thyroid Disease



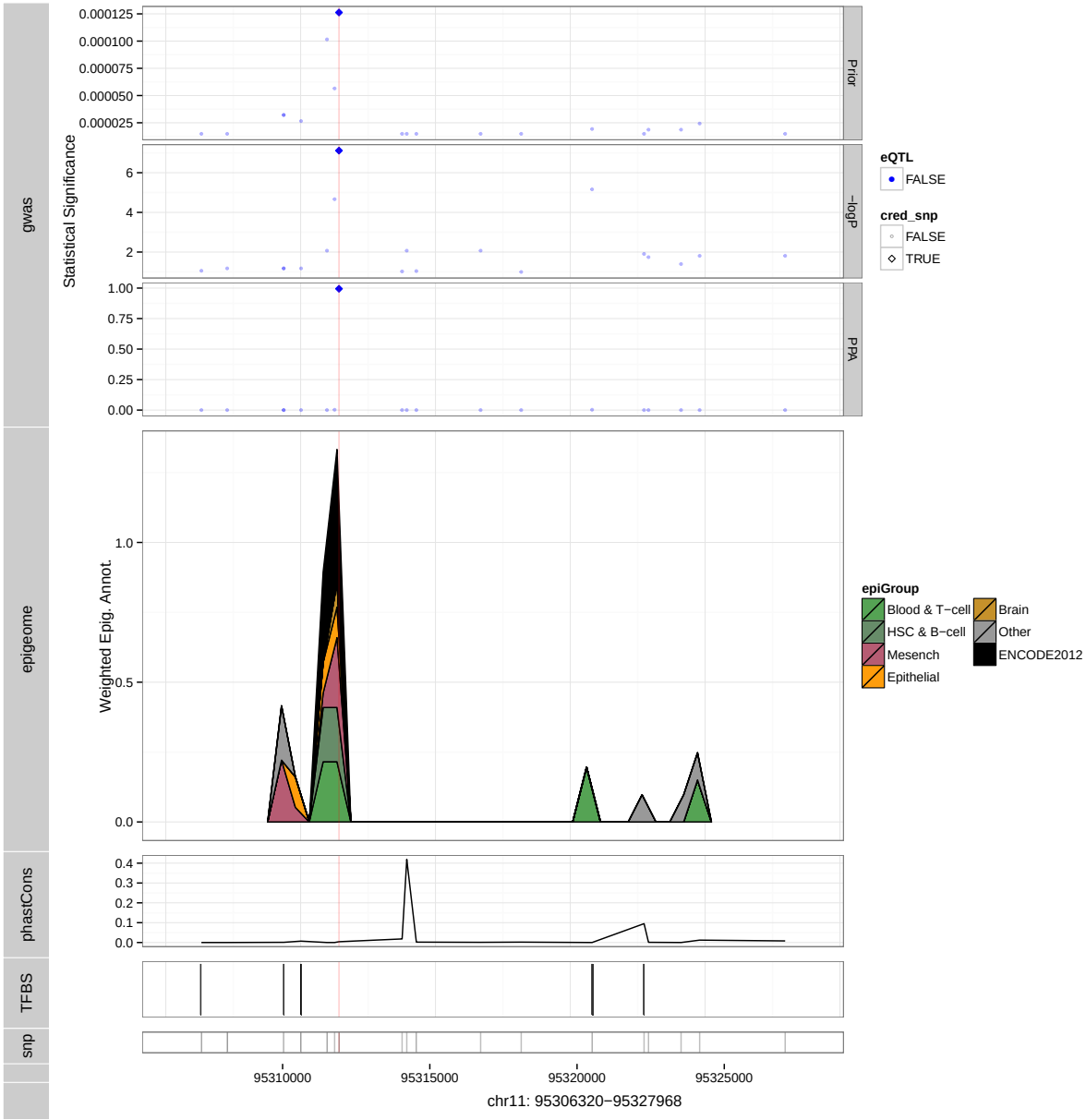
Autoimmune Thyroid Disease



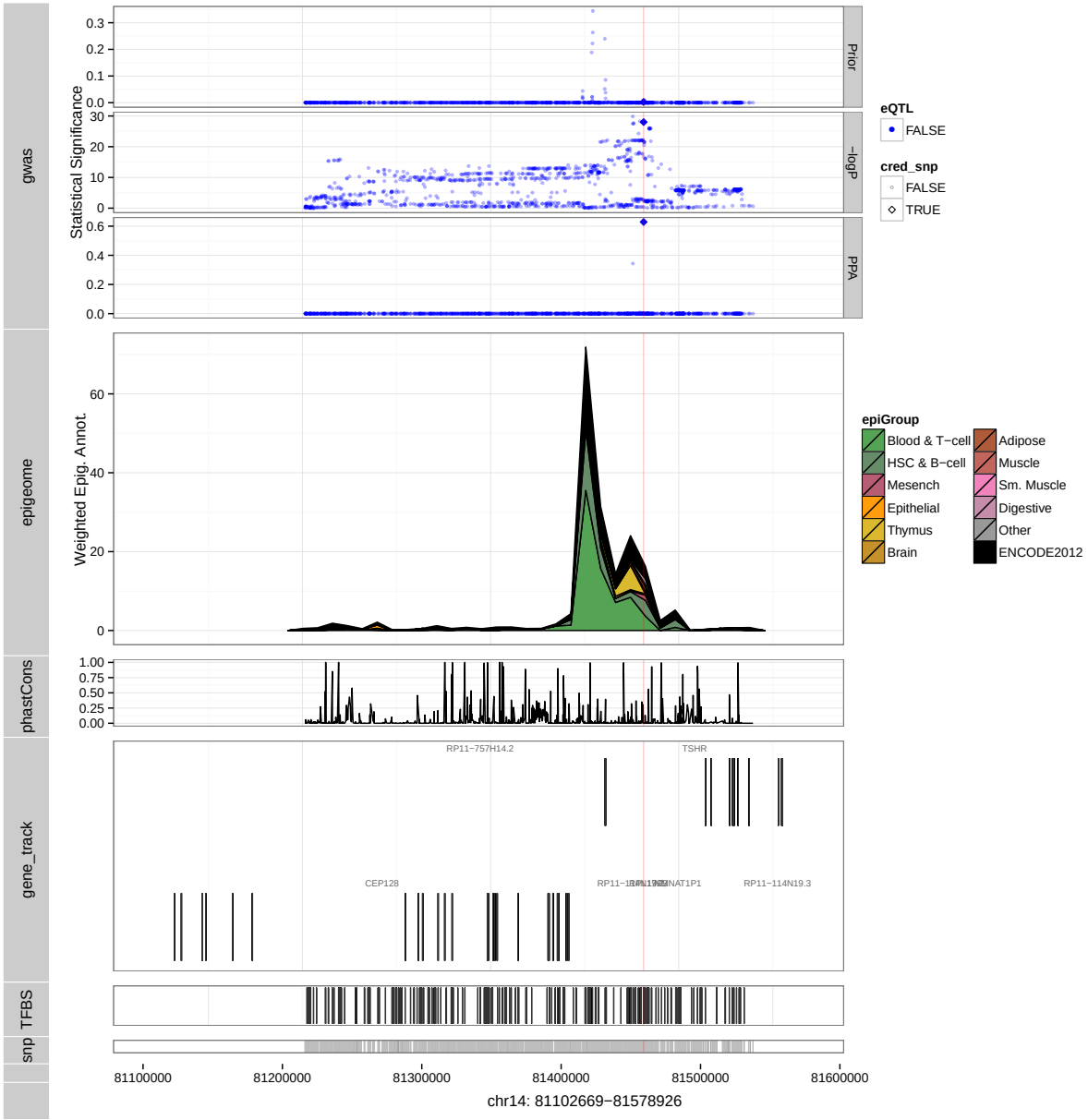
Autoimmune Thyroid Disease



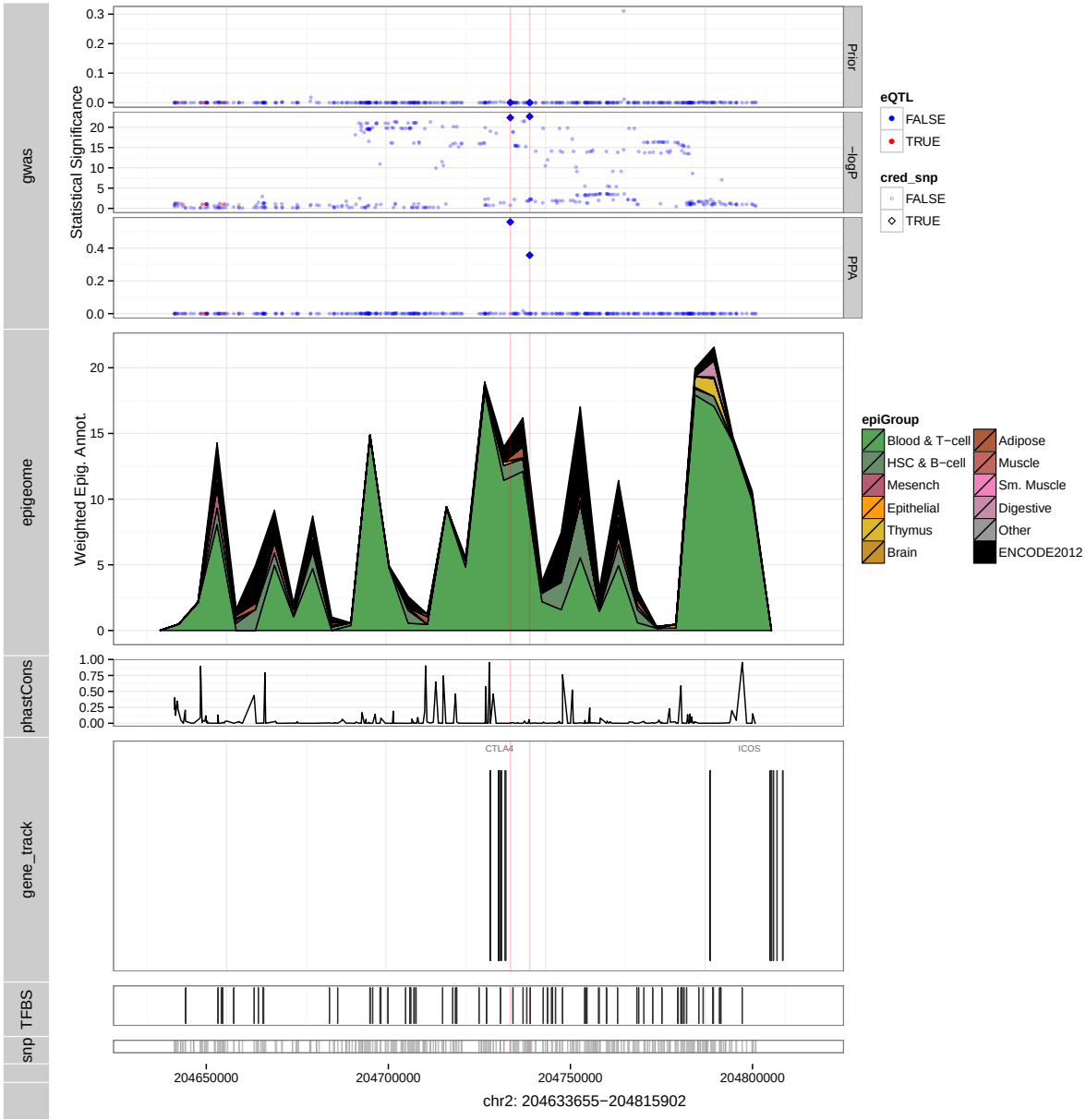
Autoimmune Thyroid Disease



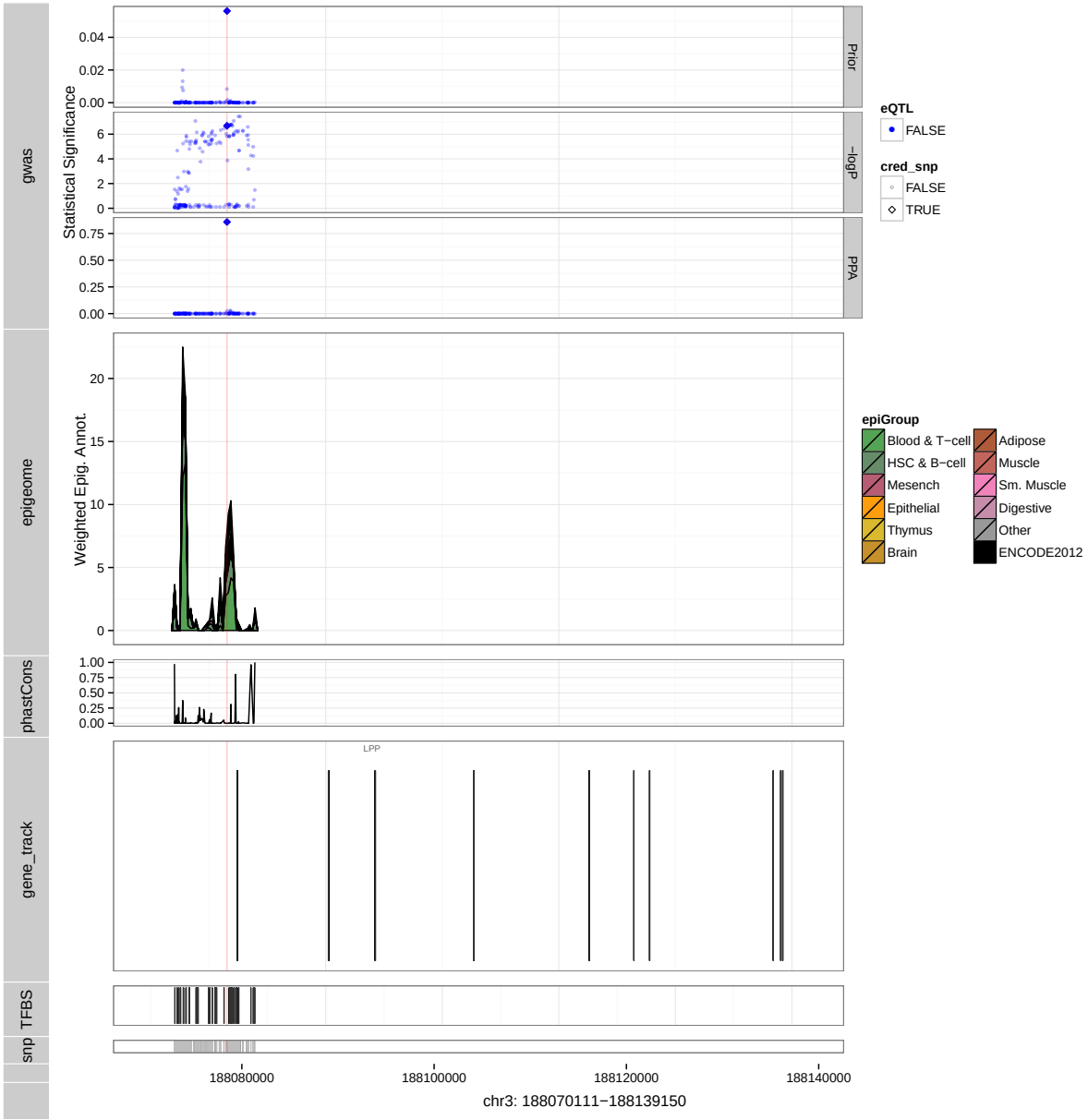
Autoimmune Thyroid Disease



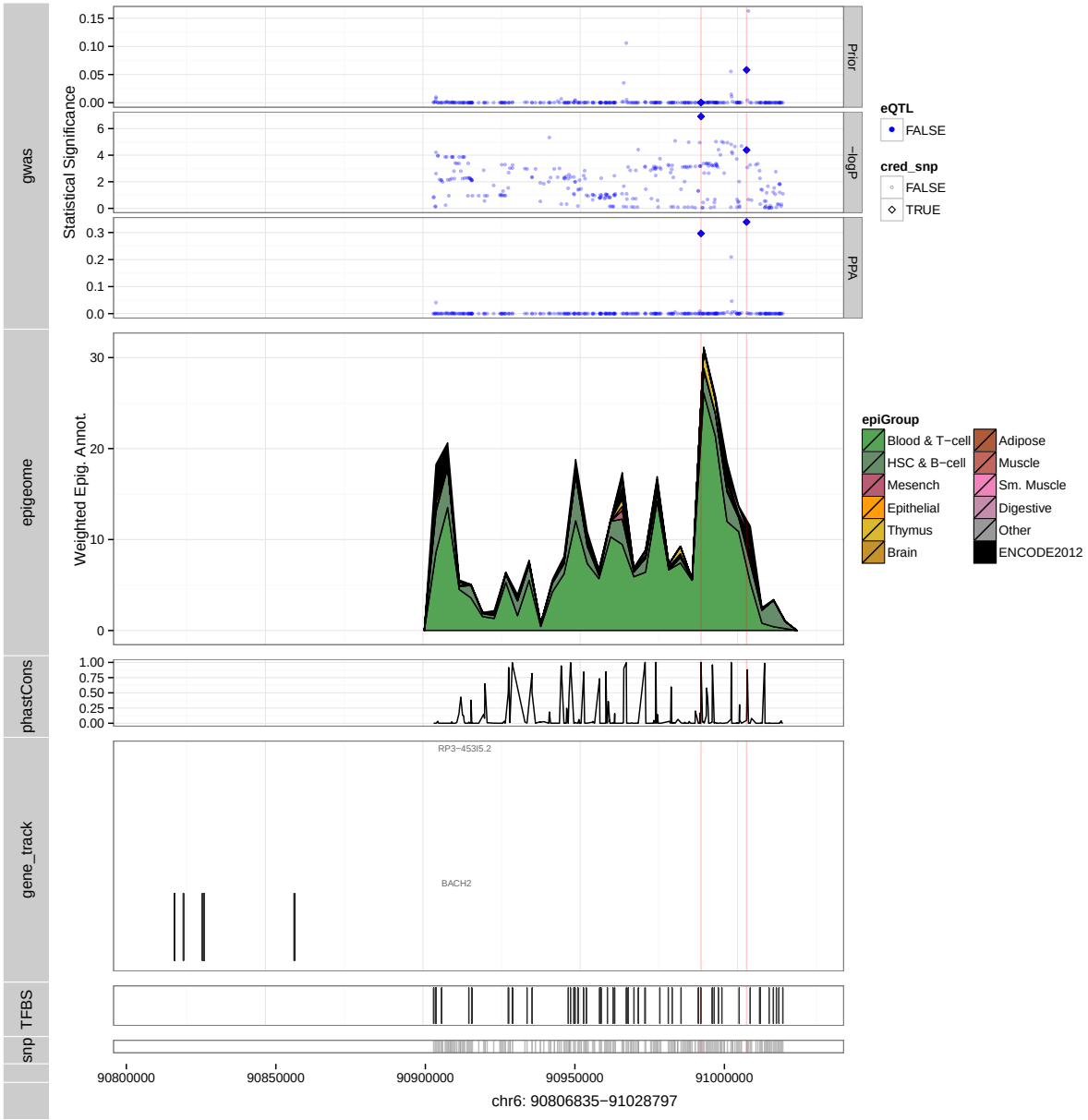
Autoimmune Thyroid Disease



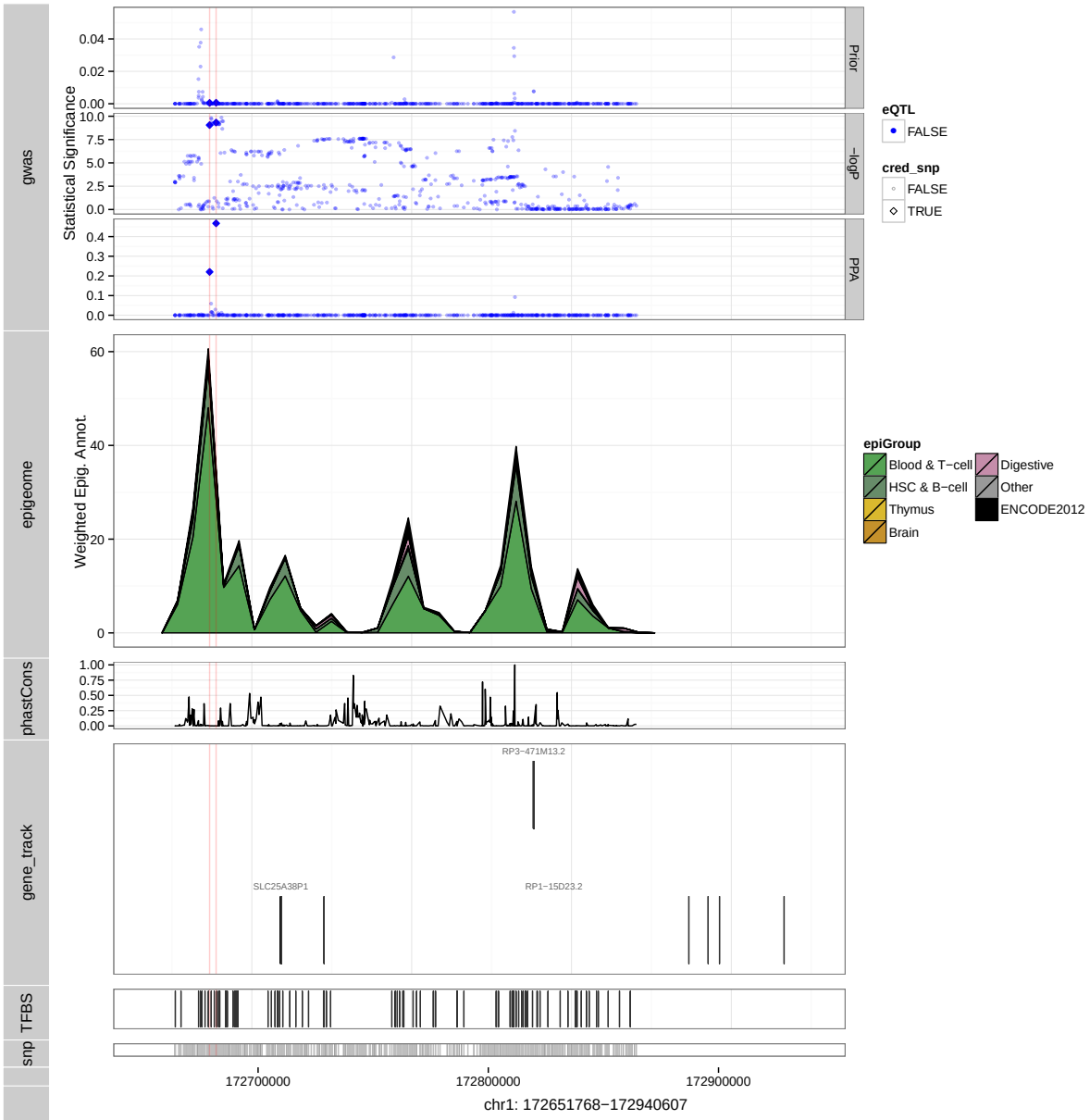
Autoimmune Thyroid Disease



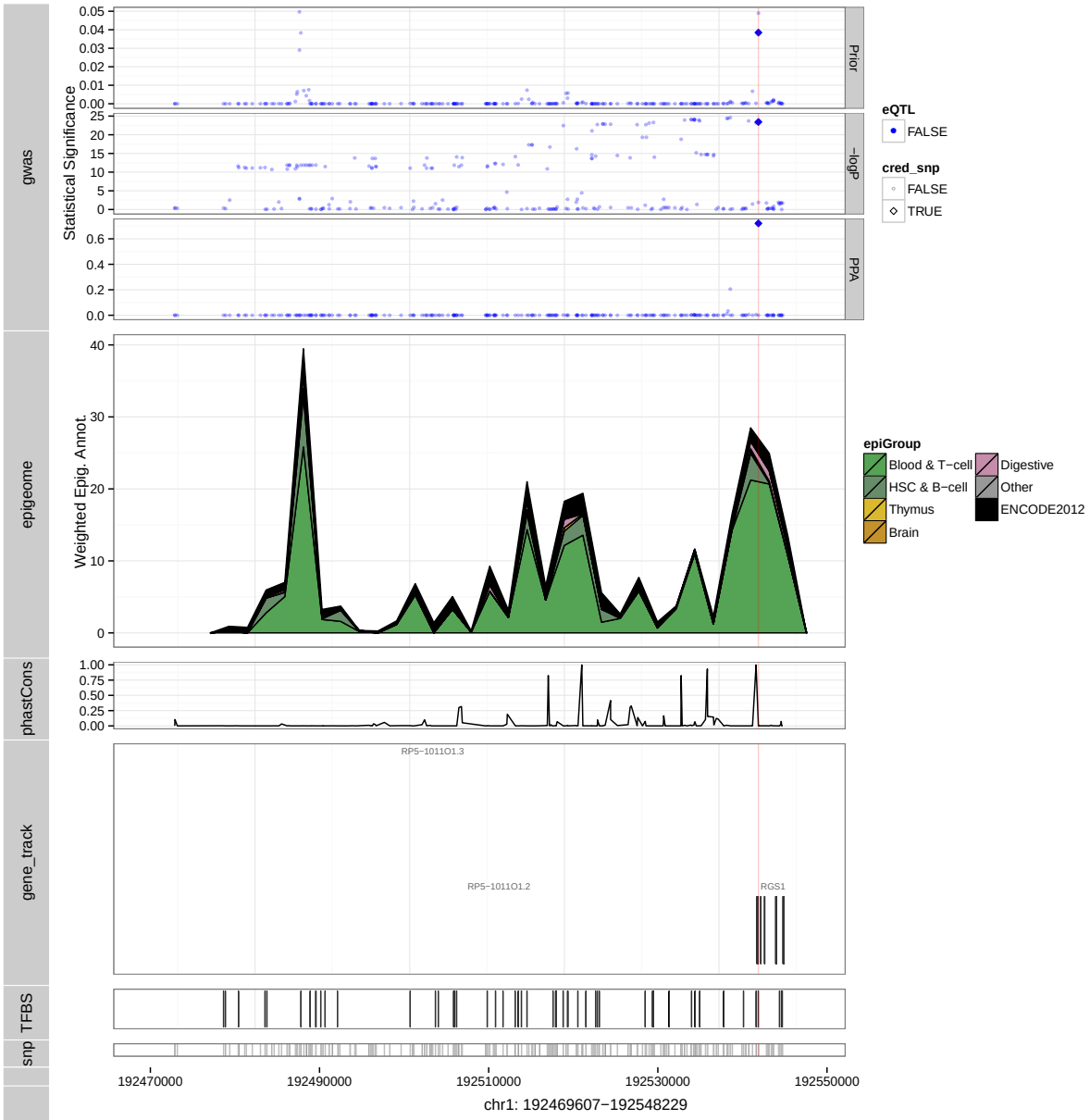
Autoimmune Thyroid Disease



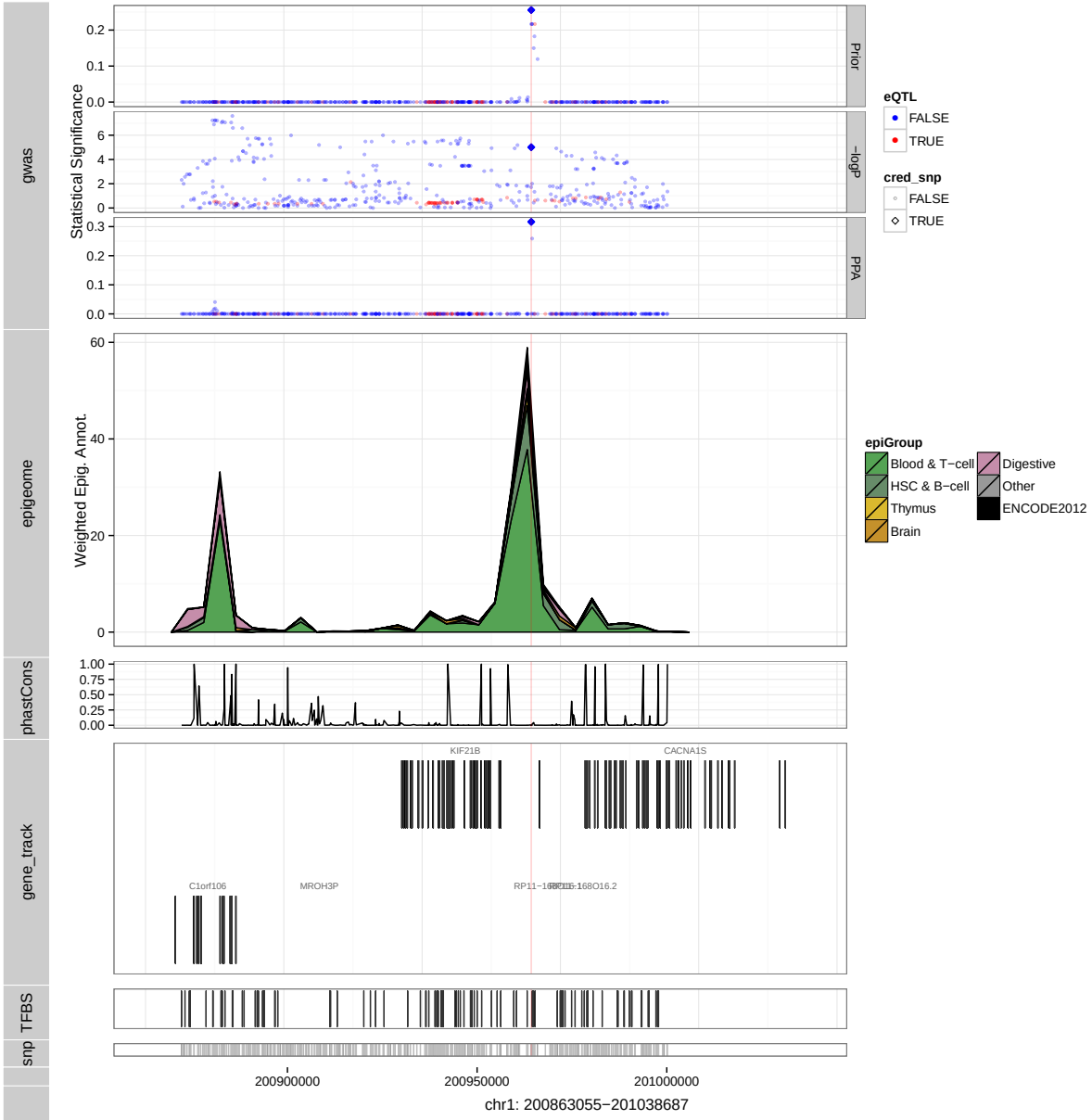
Celiac Disease



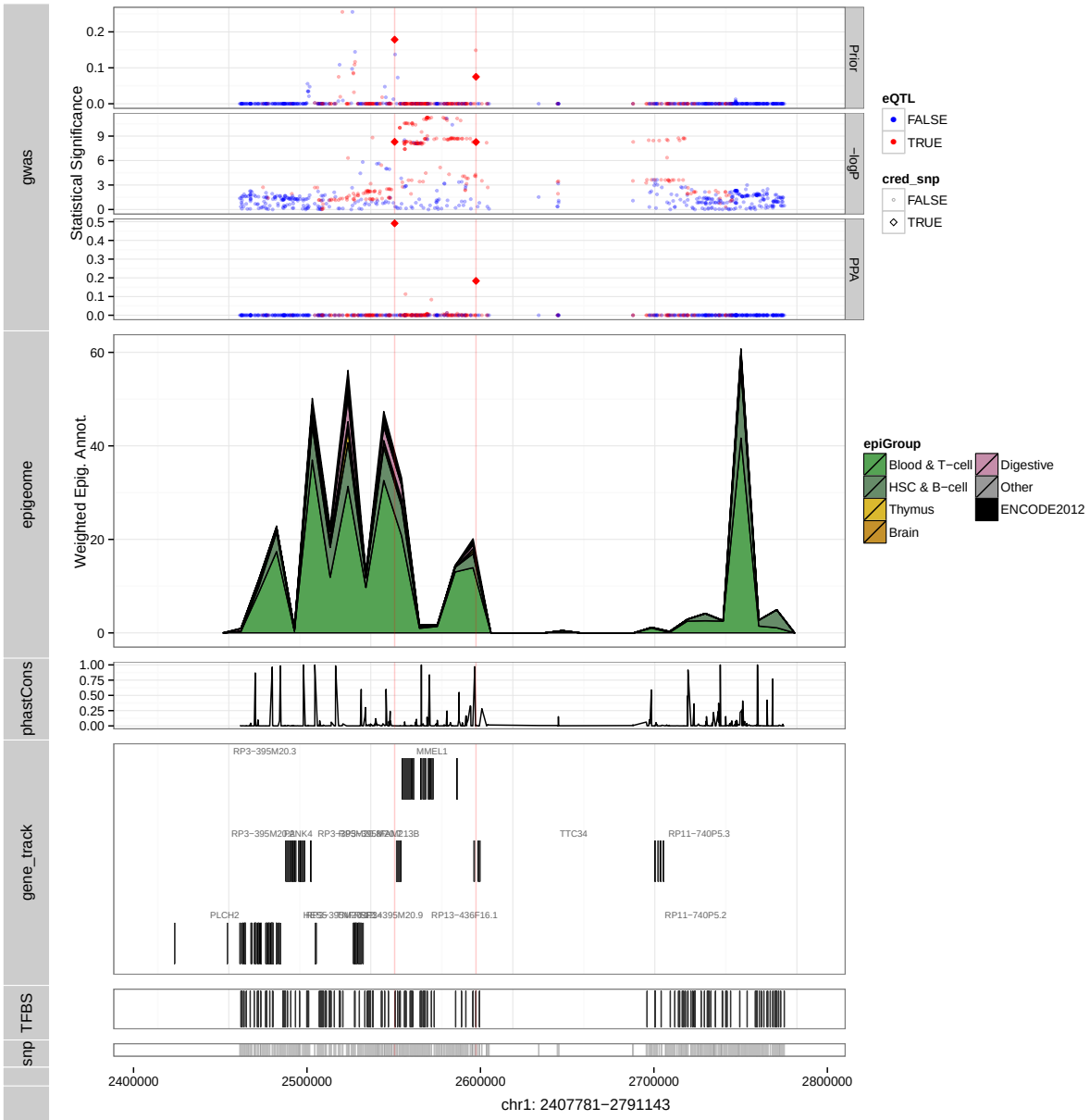
Celiac Disease



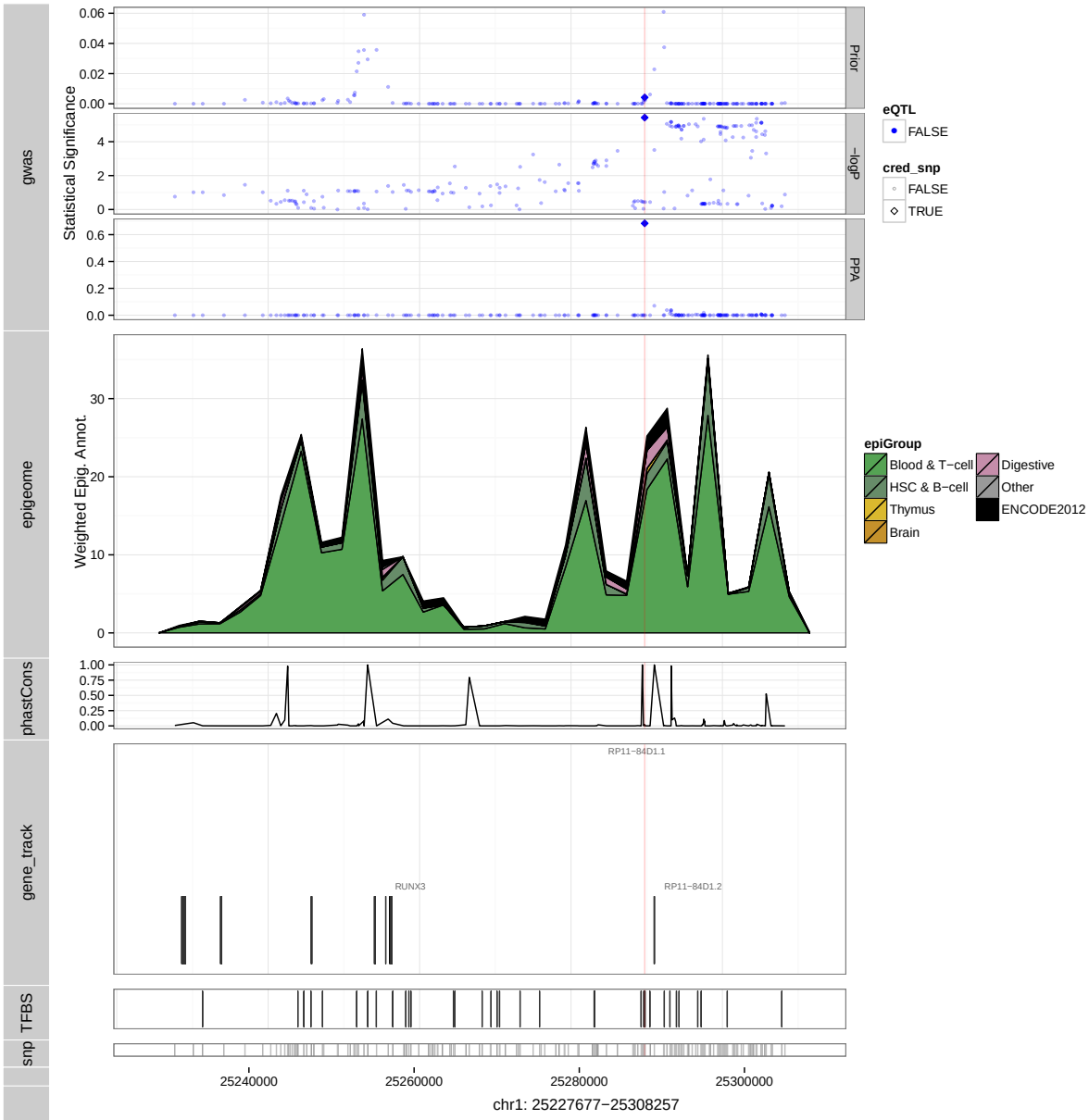
Celiac Disease



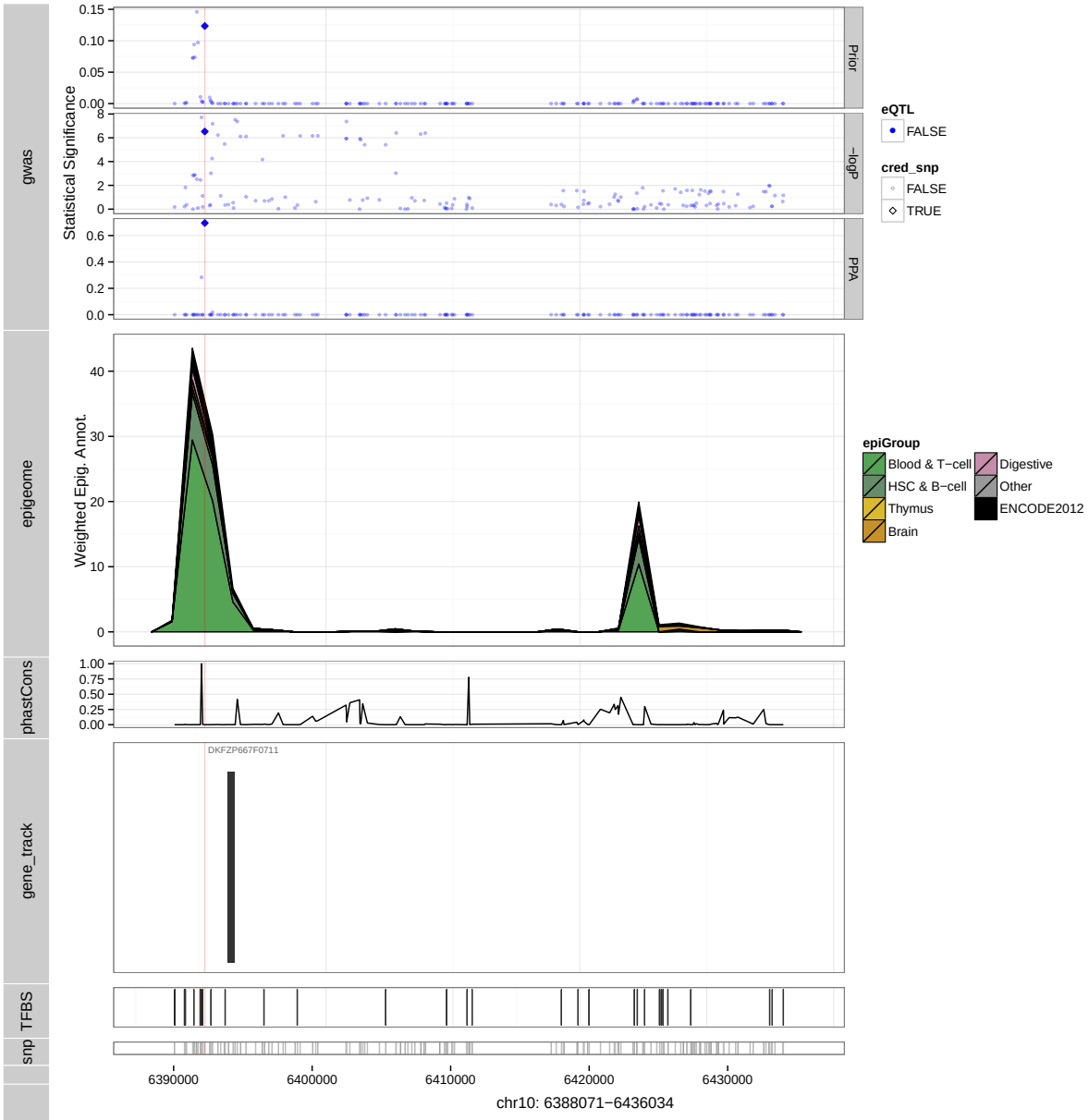
Celiac Disease



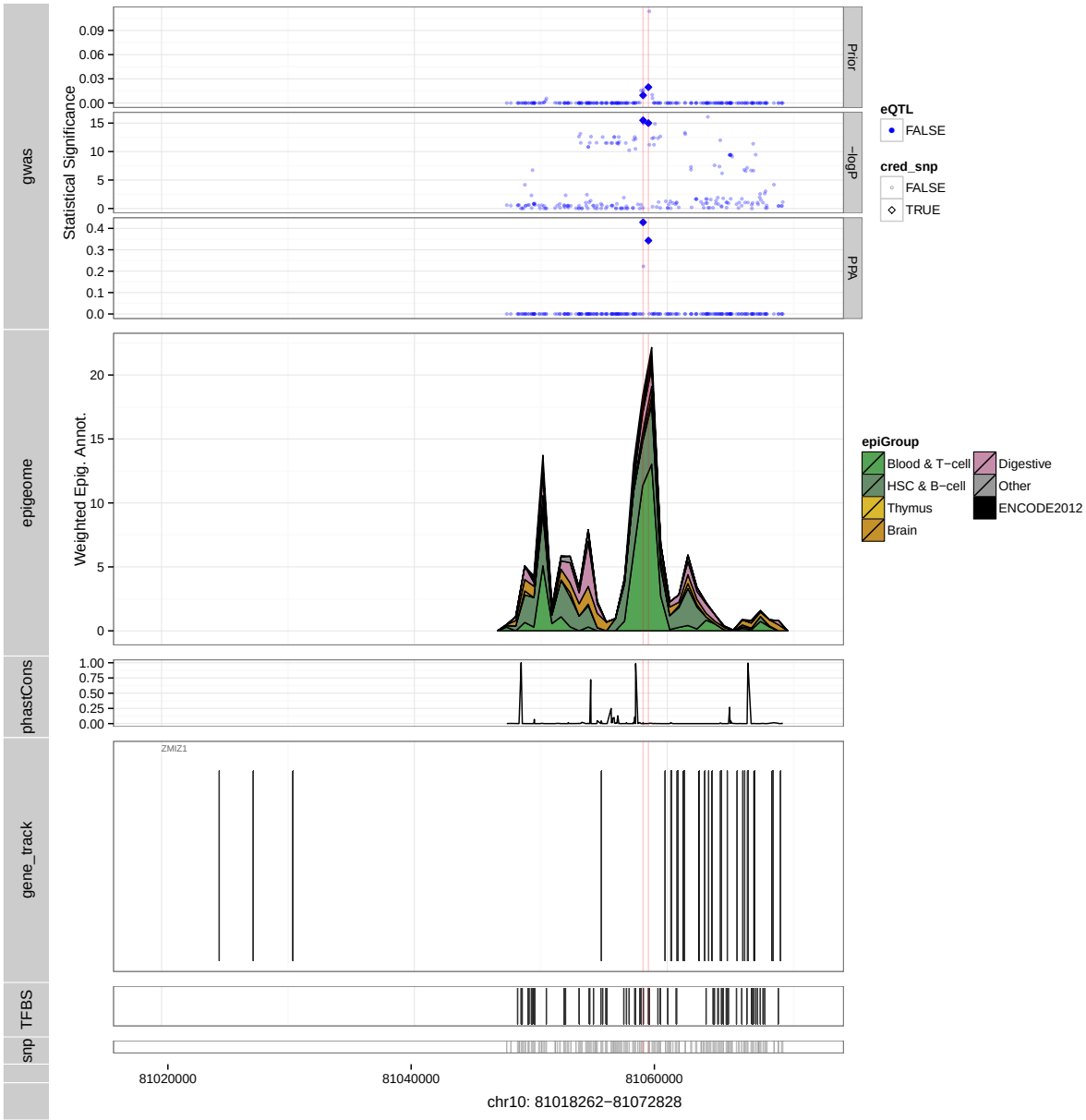
Celiac Disease



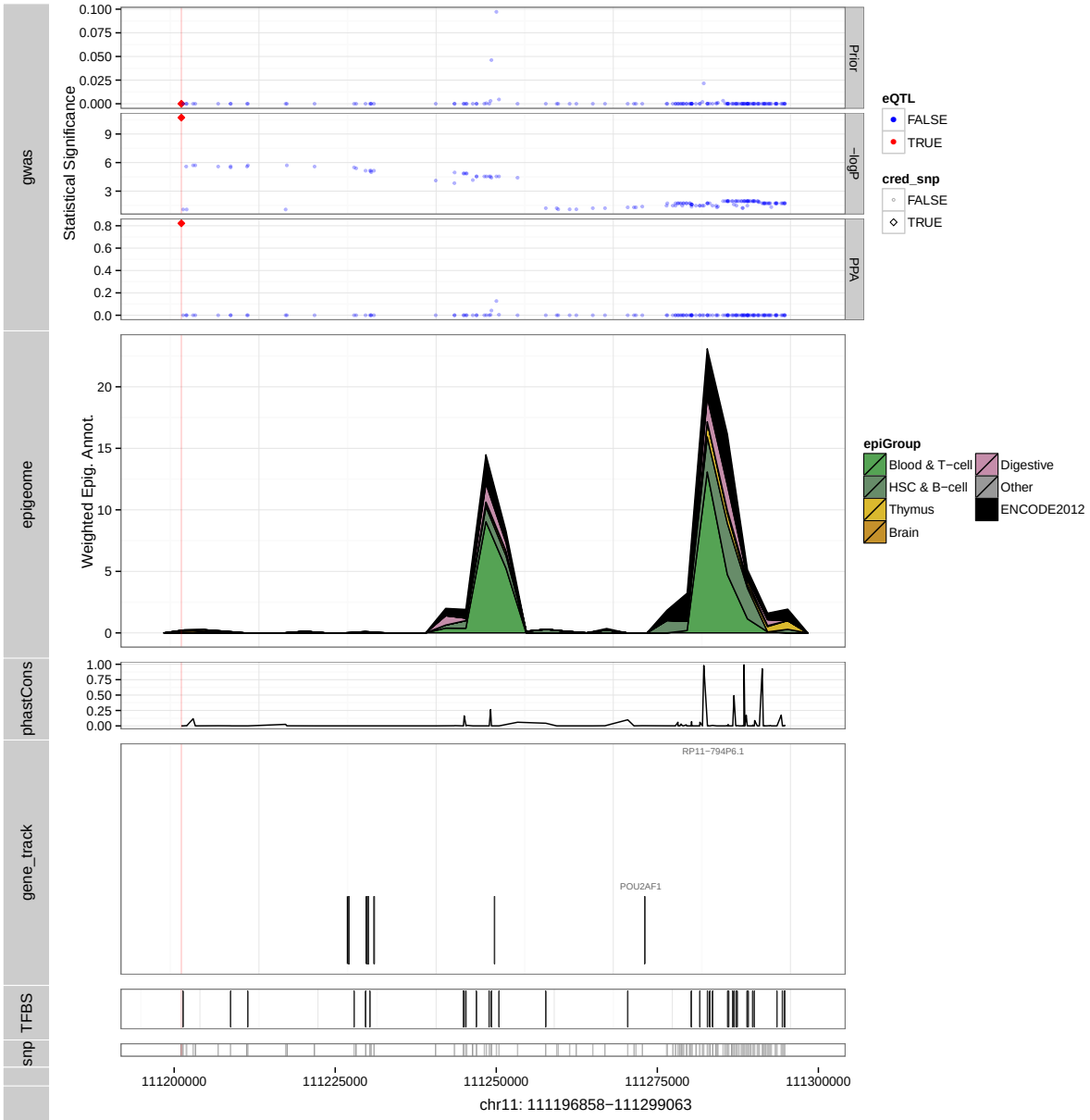
Celiac Disease



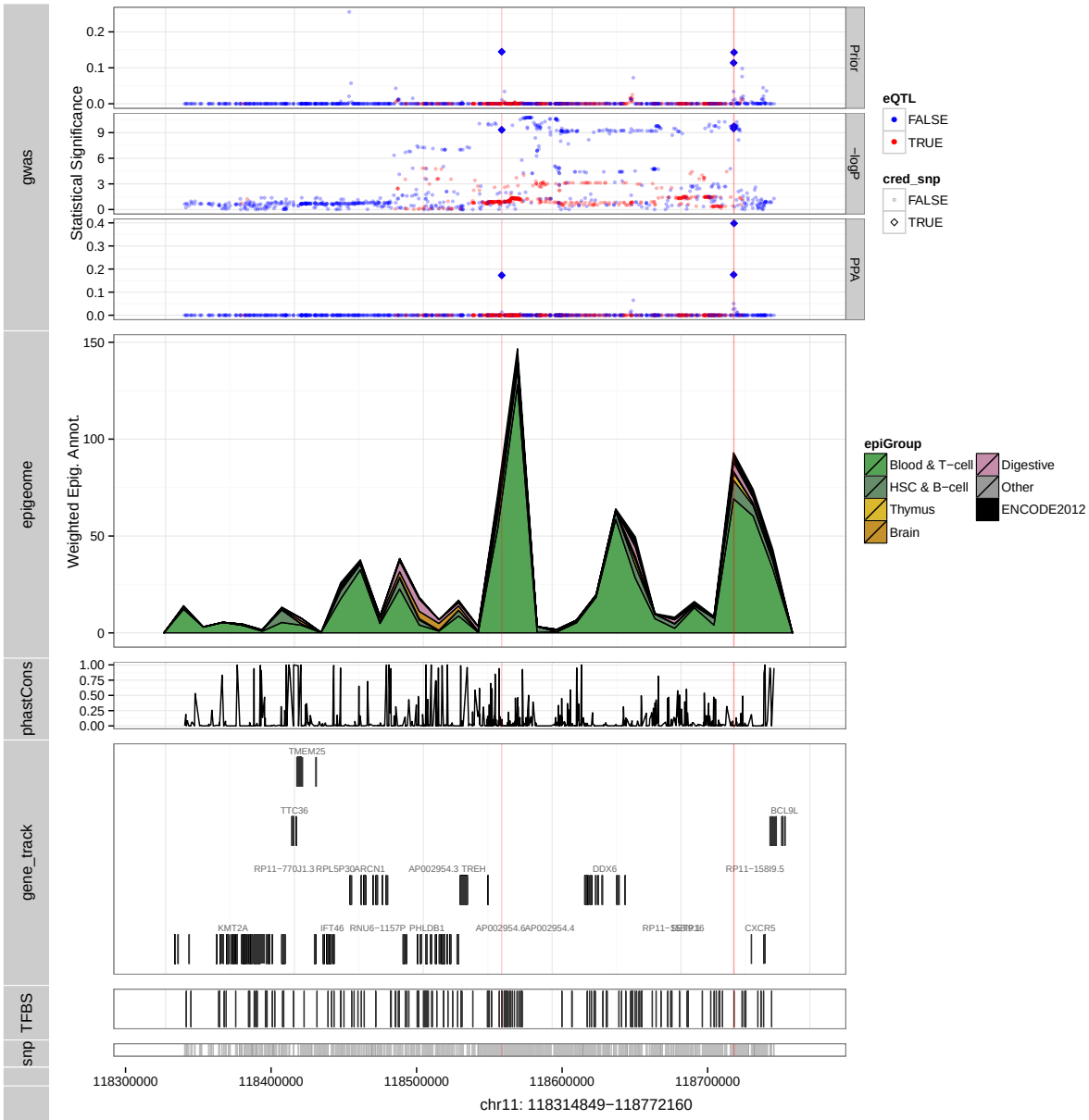
Celiac Disease



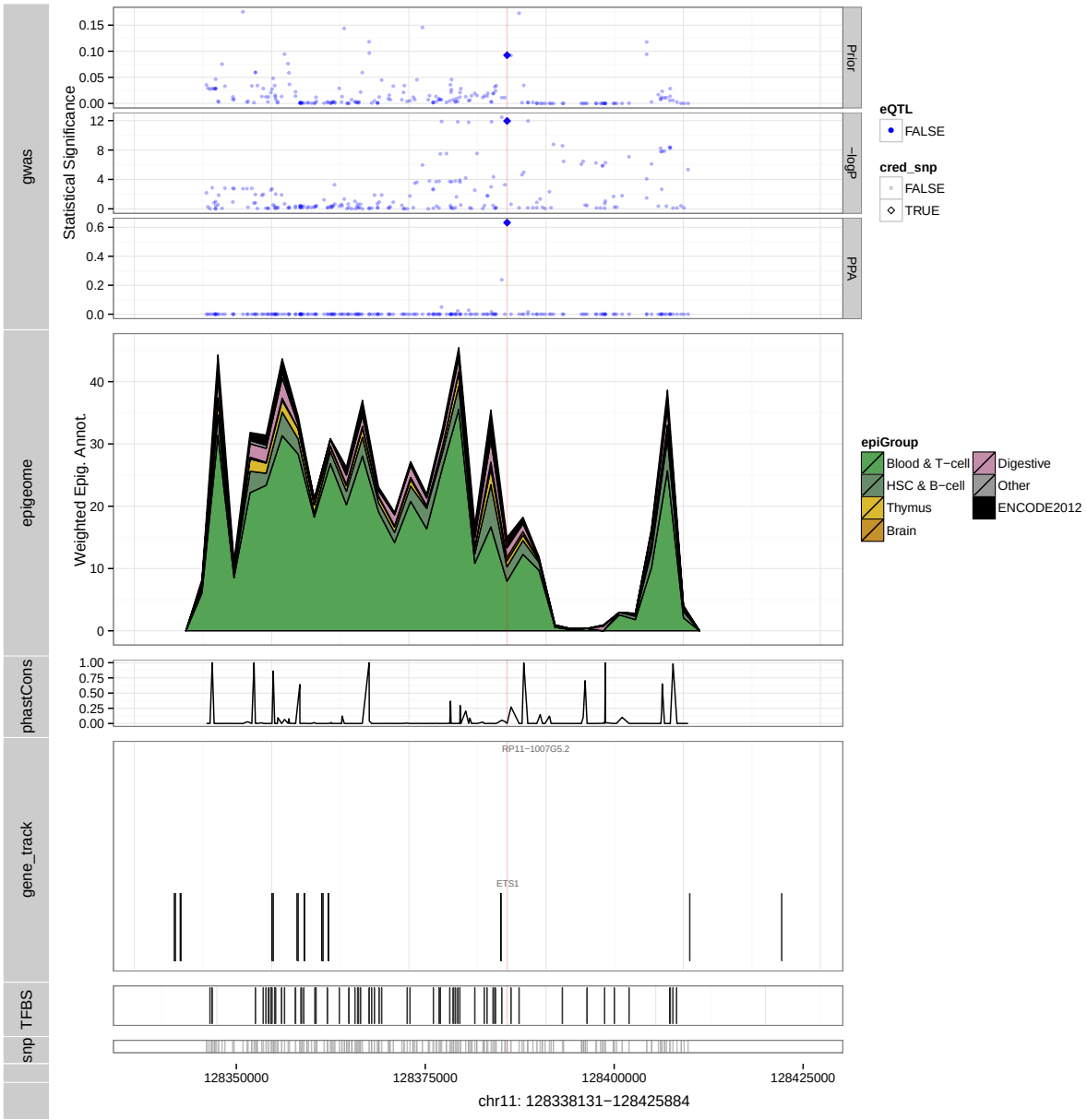
Celiac Disease



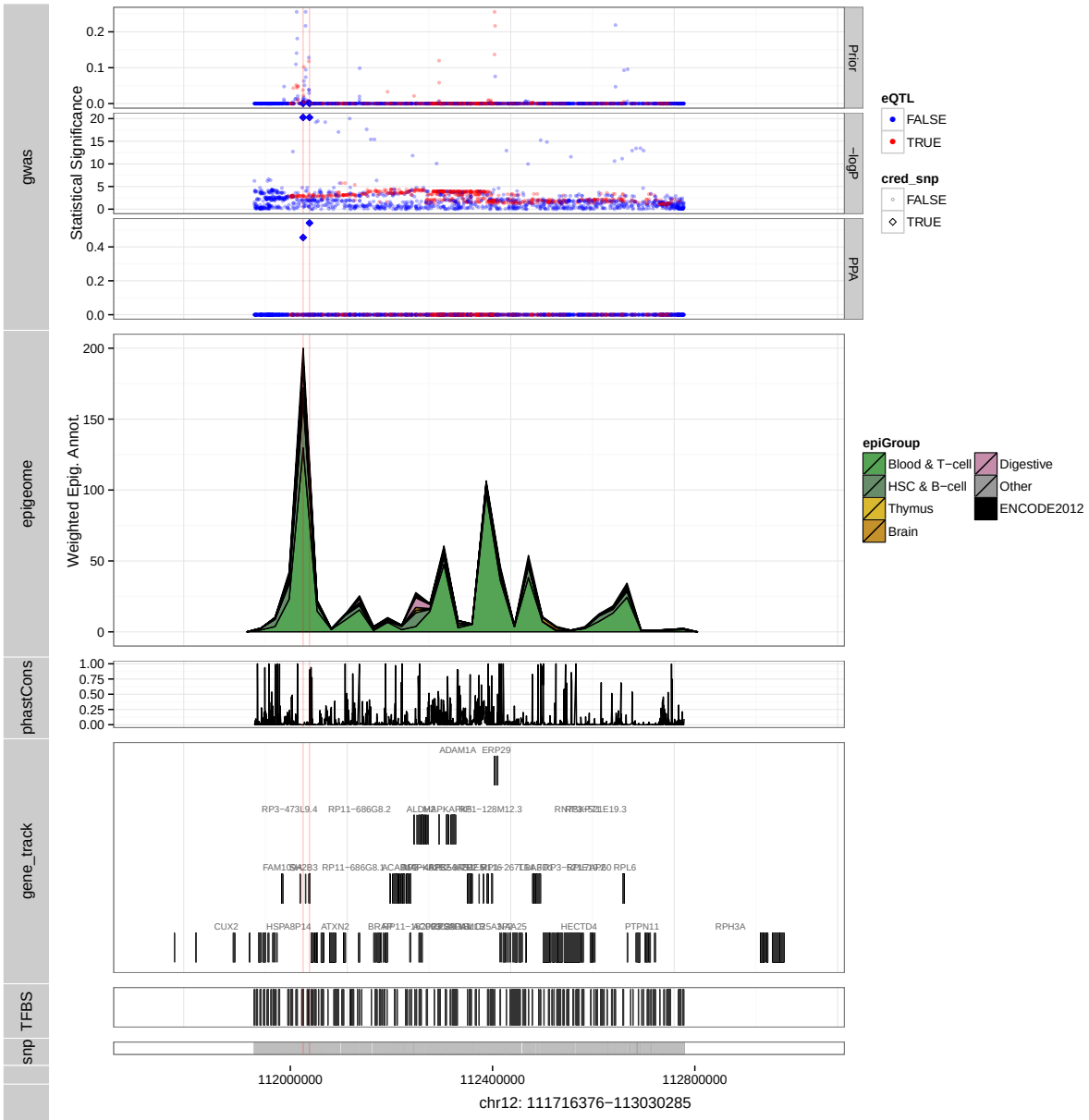
Celiac Disease



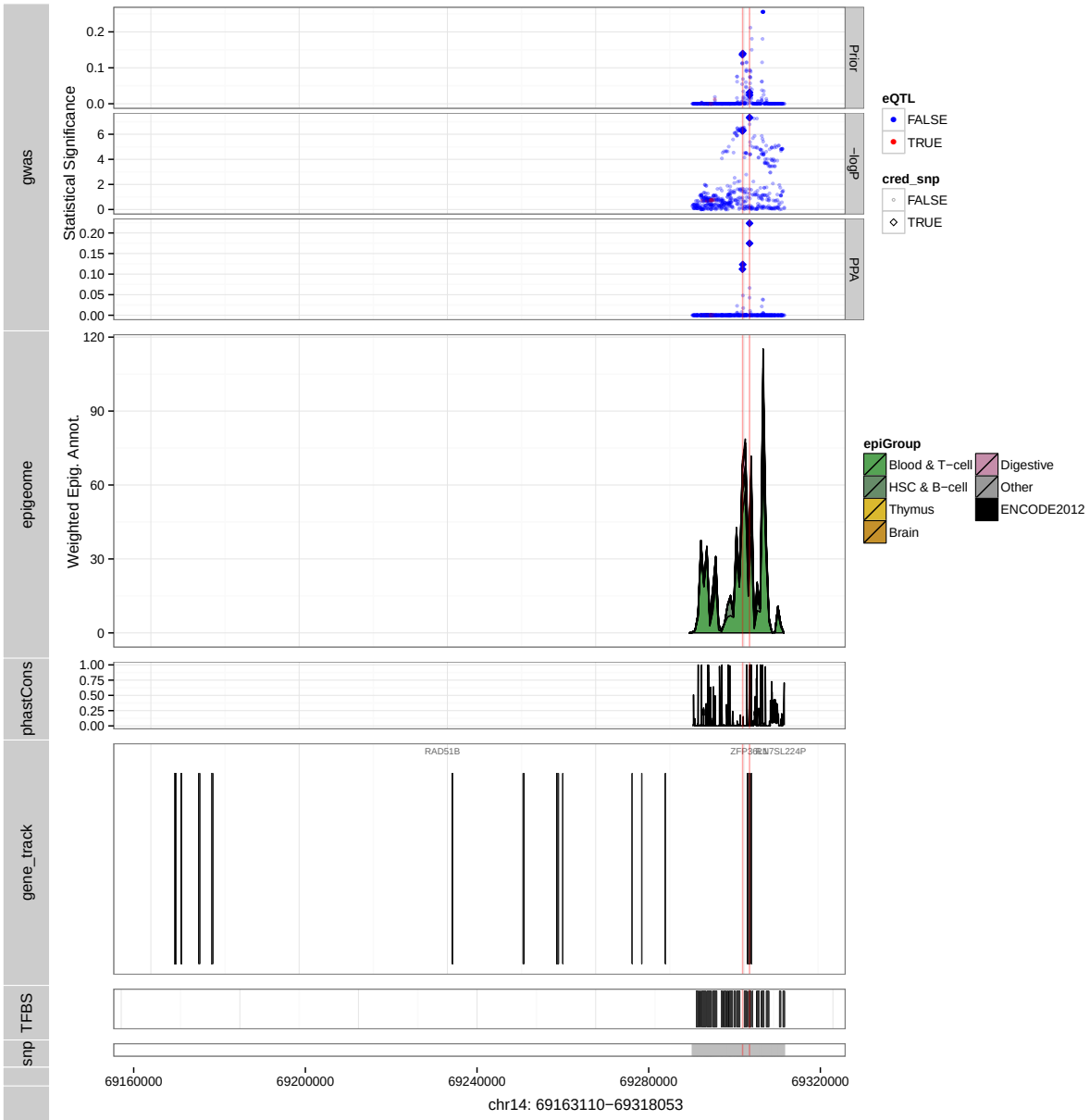
Celiac Disease



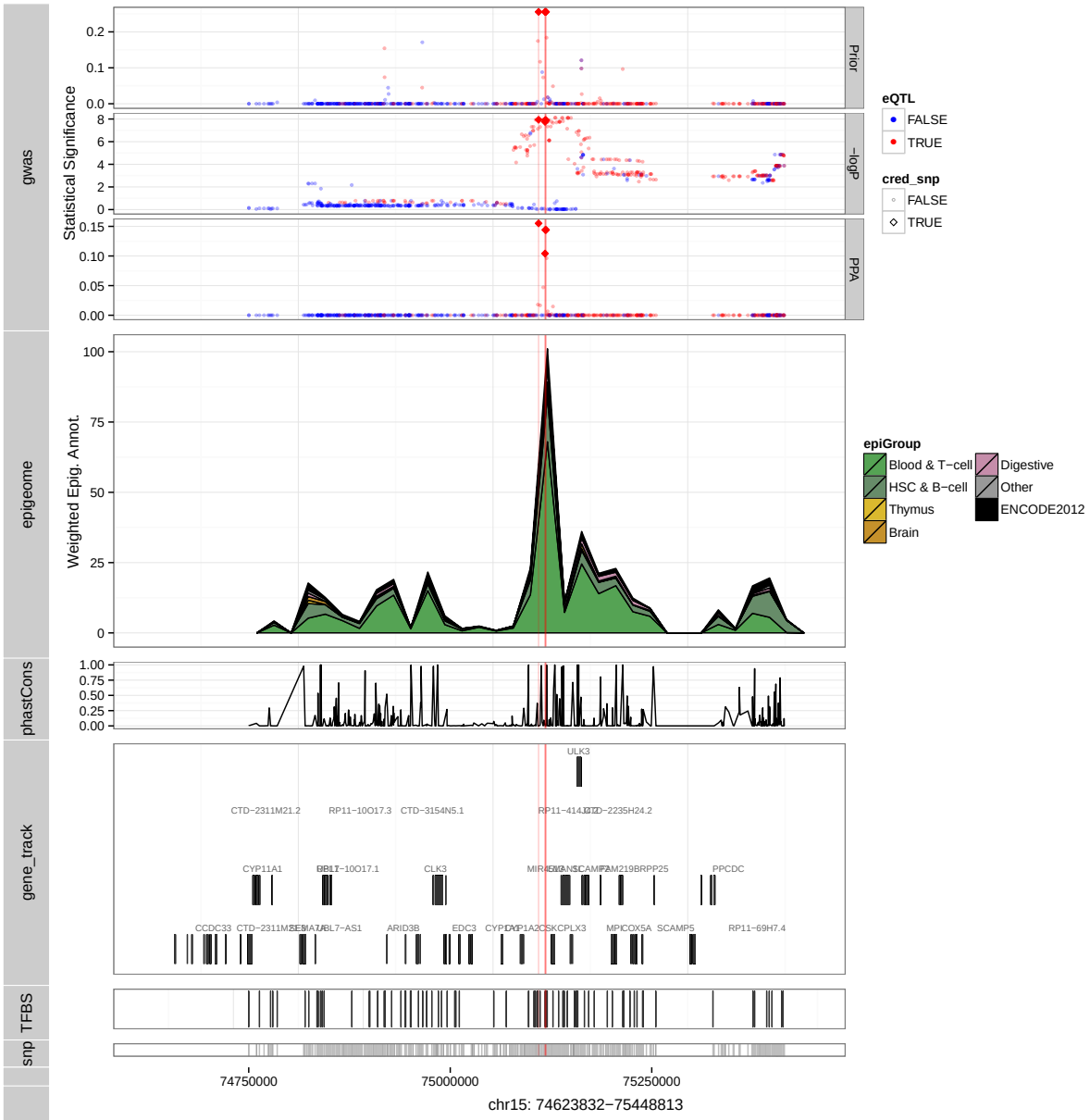
Celiac Disease



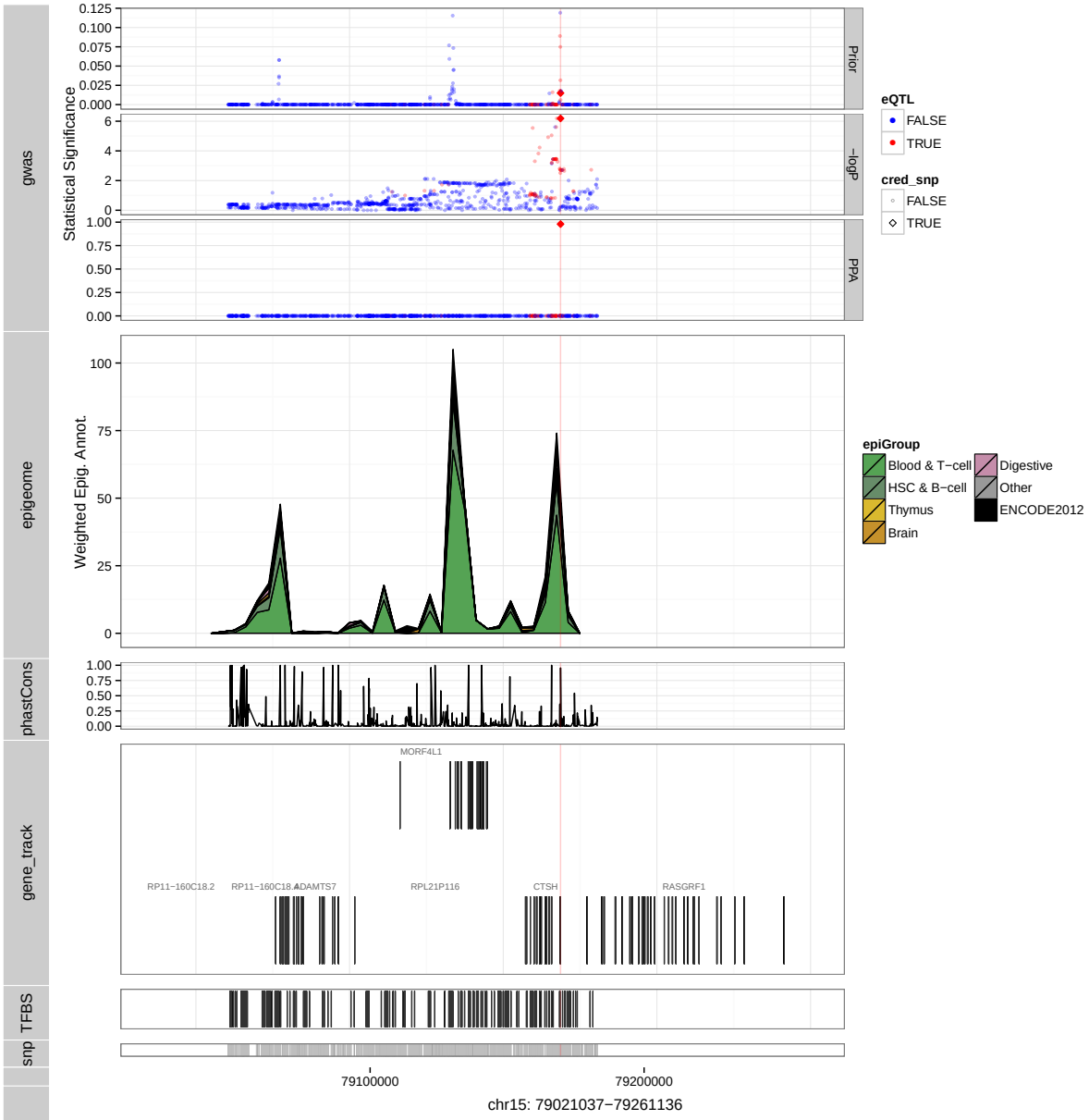
Celiac Disease



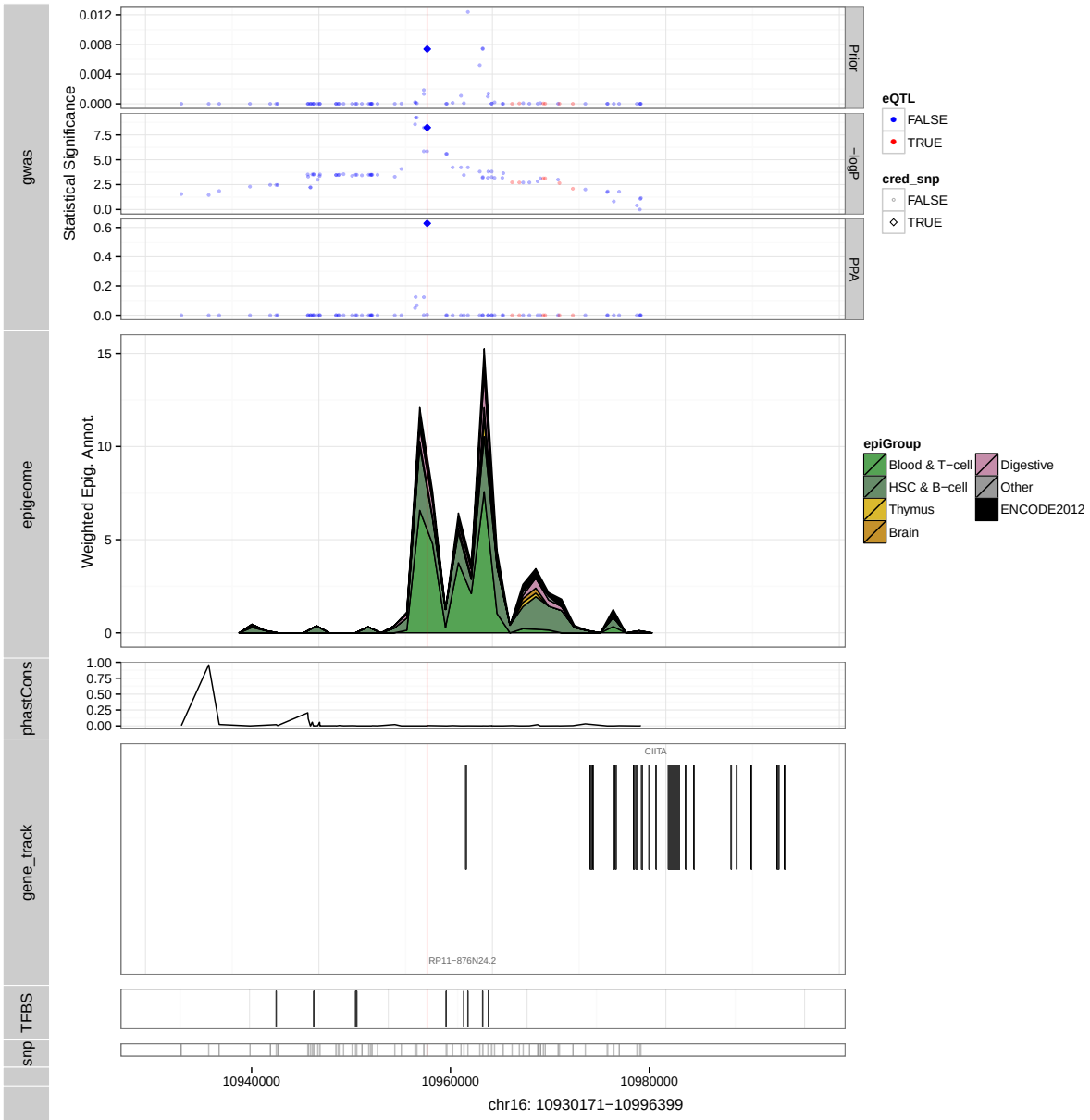
Celiac Disease



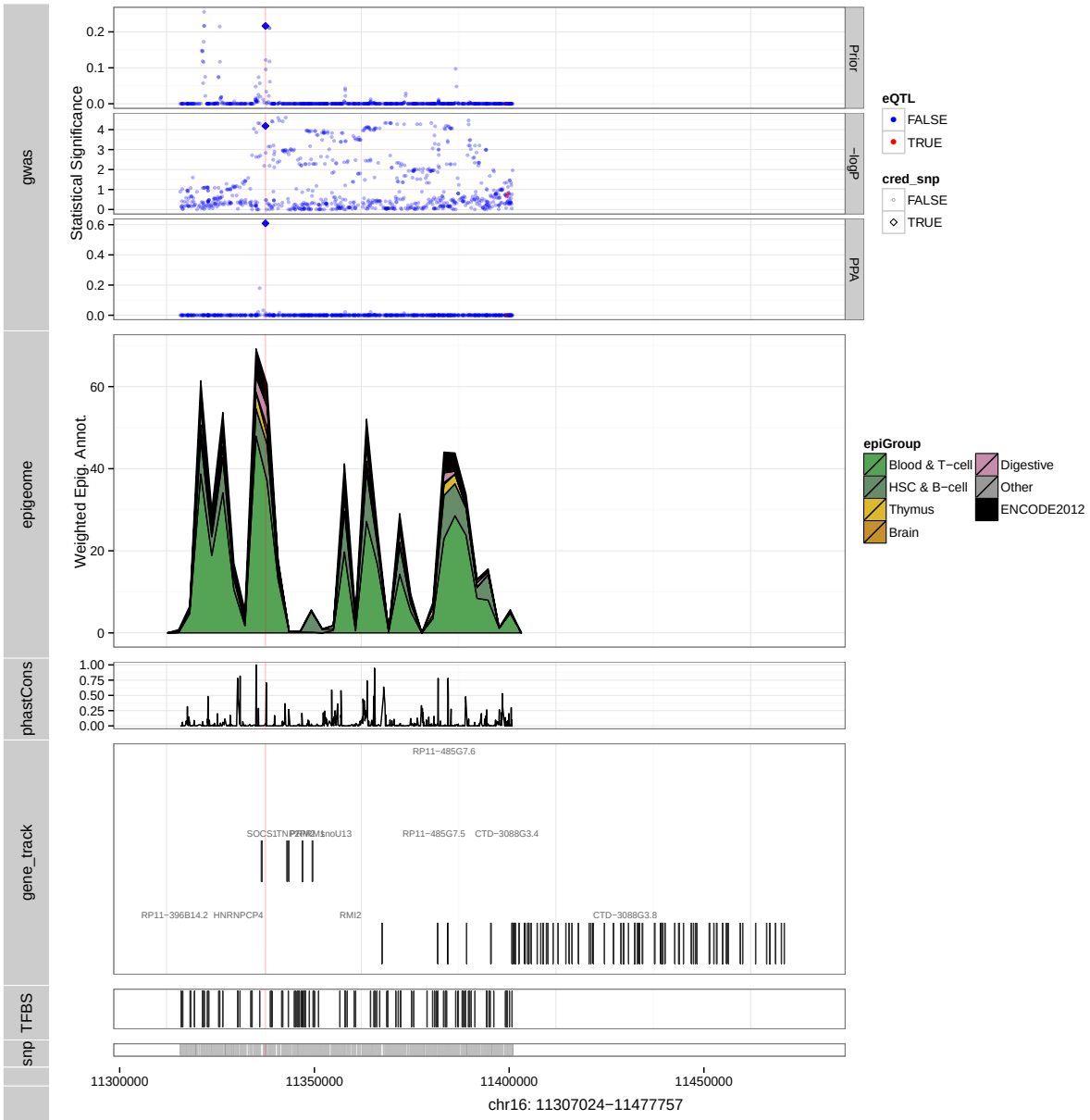
Celiac Disease



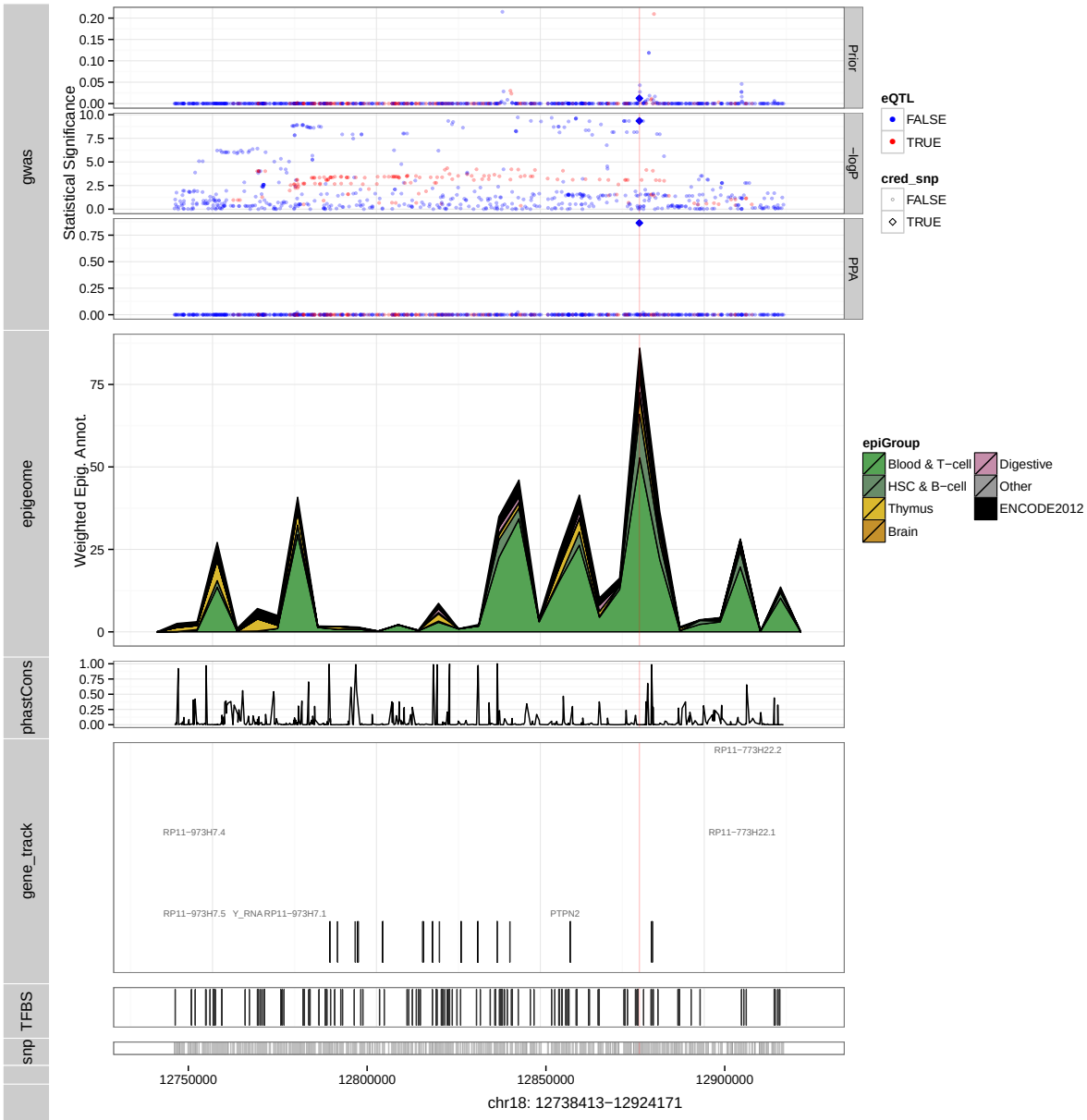
Celiac Disease



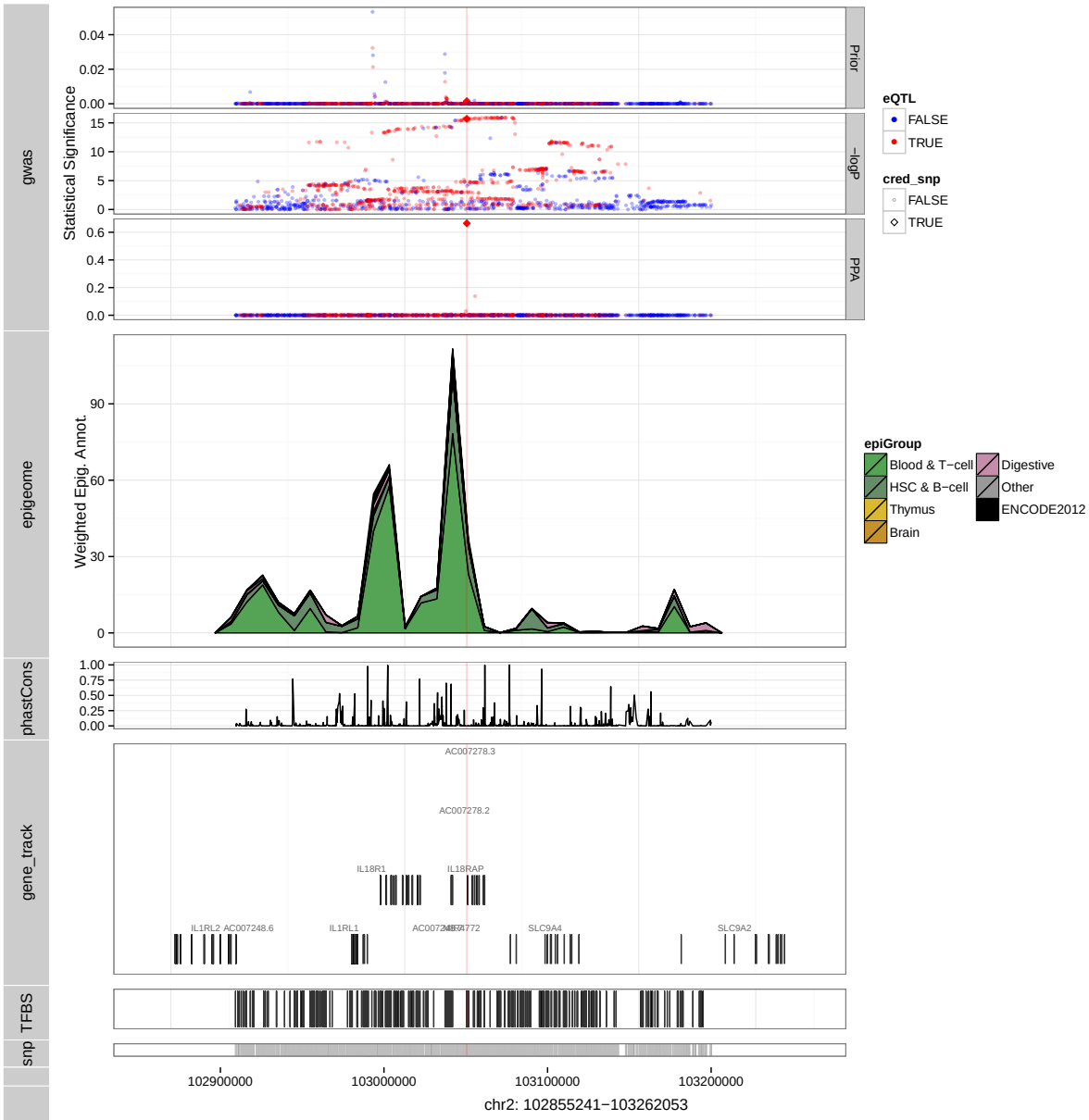
Celiac Disease



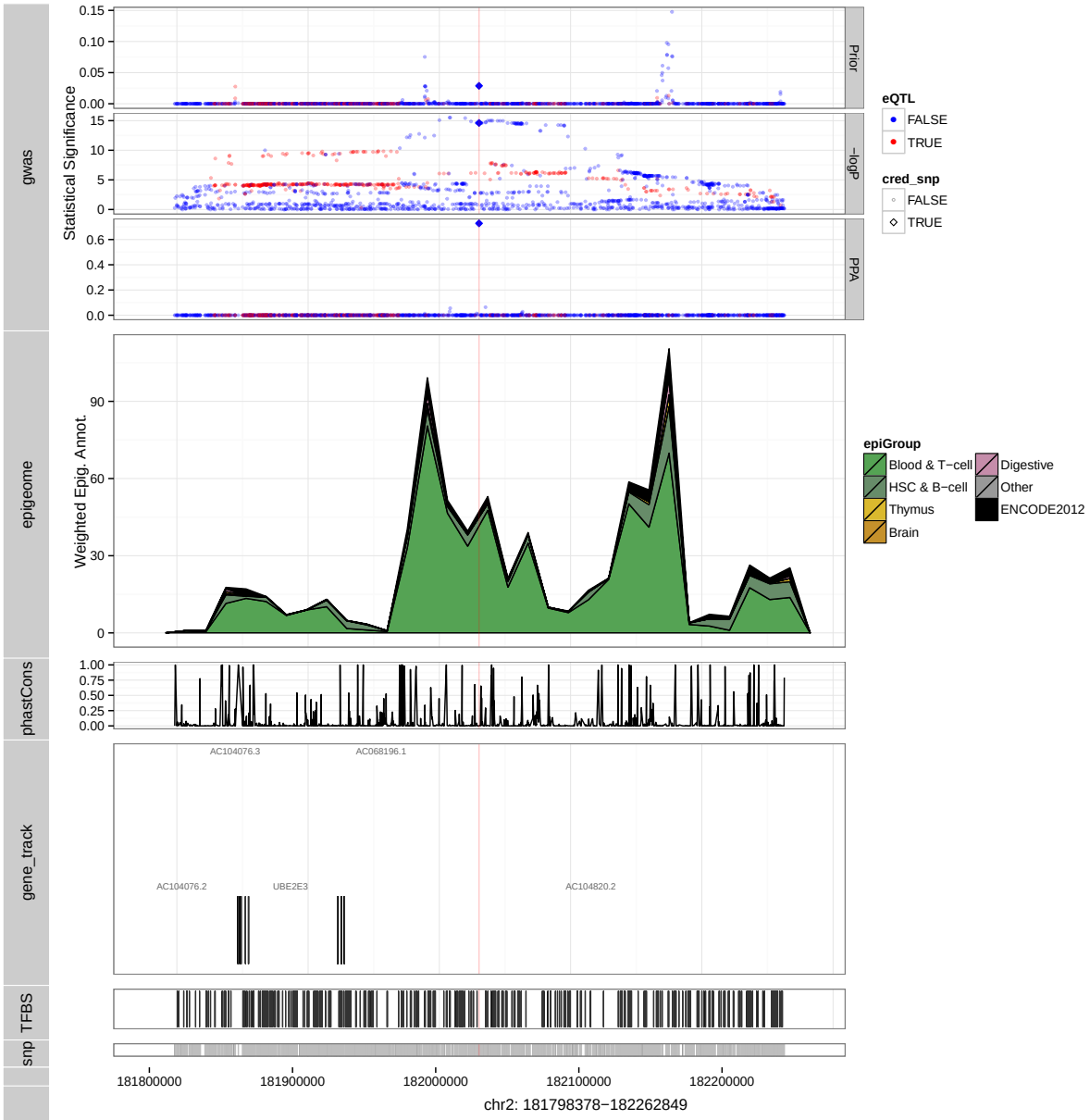
Celiac Disease



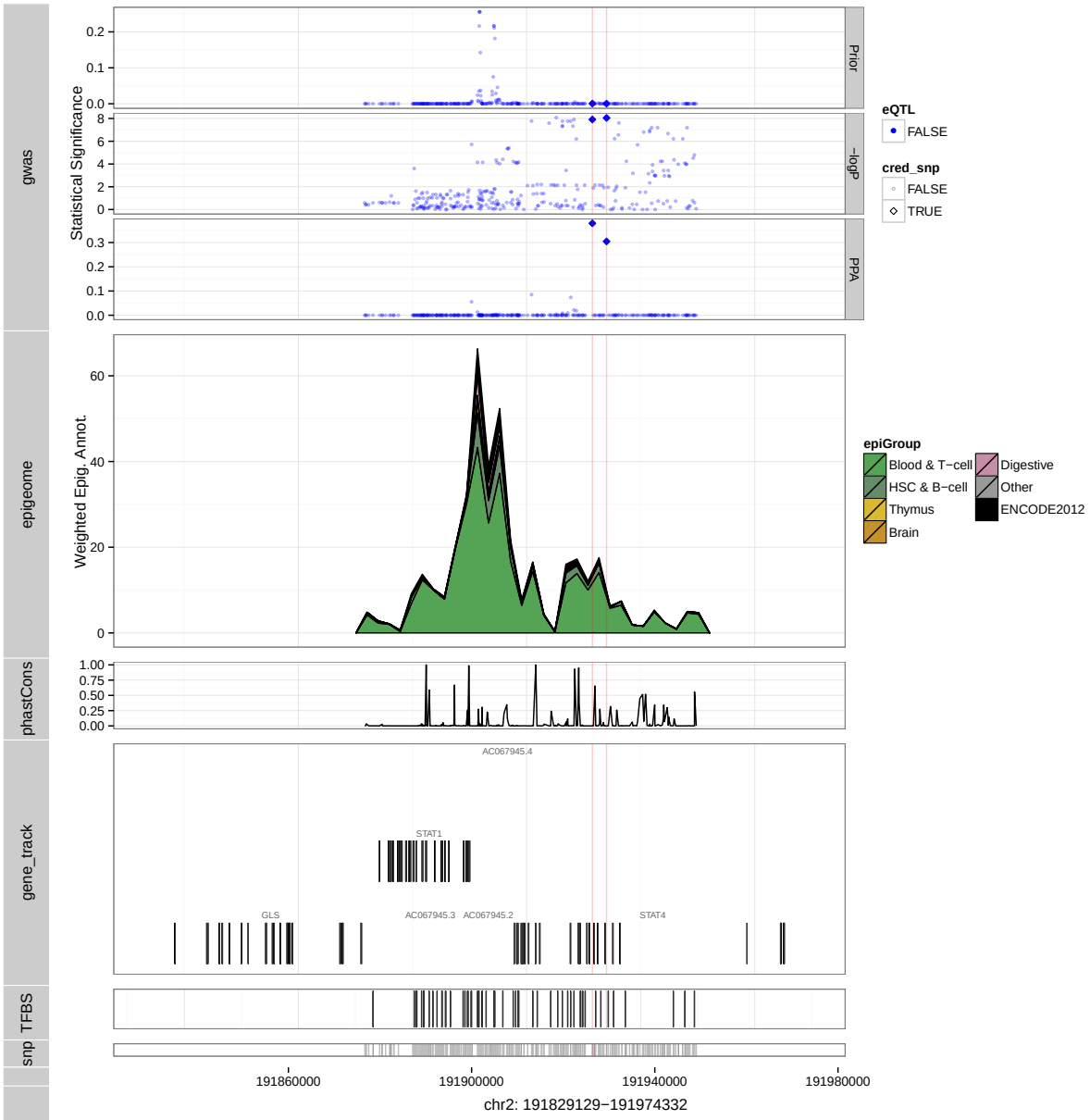
Celiac Disease



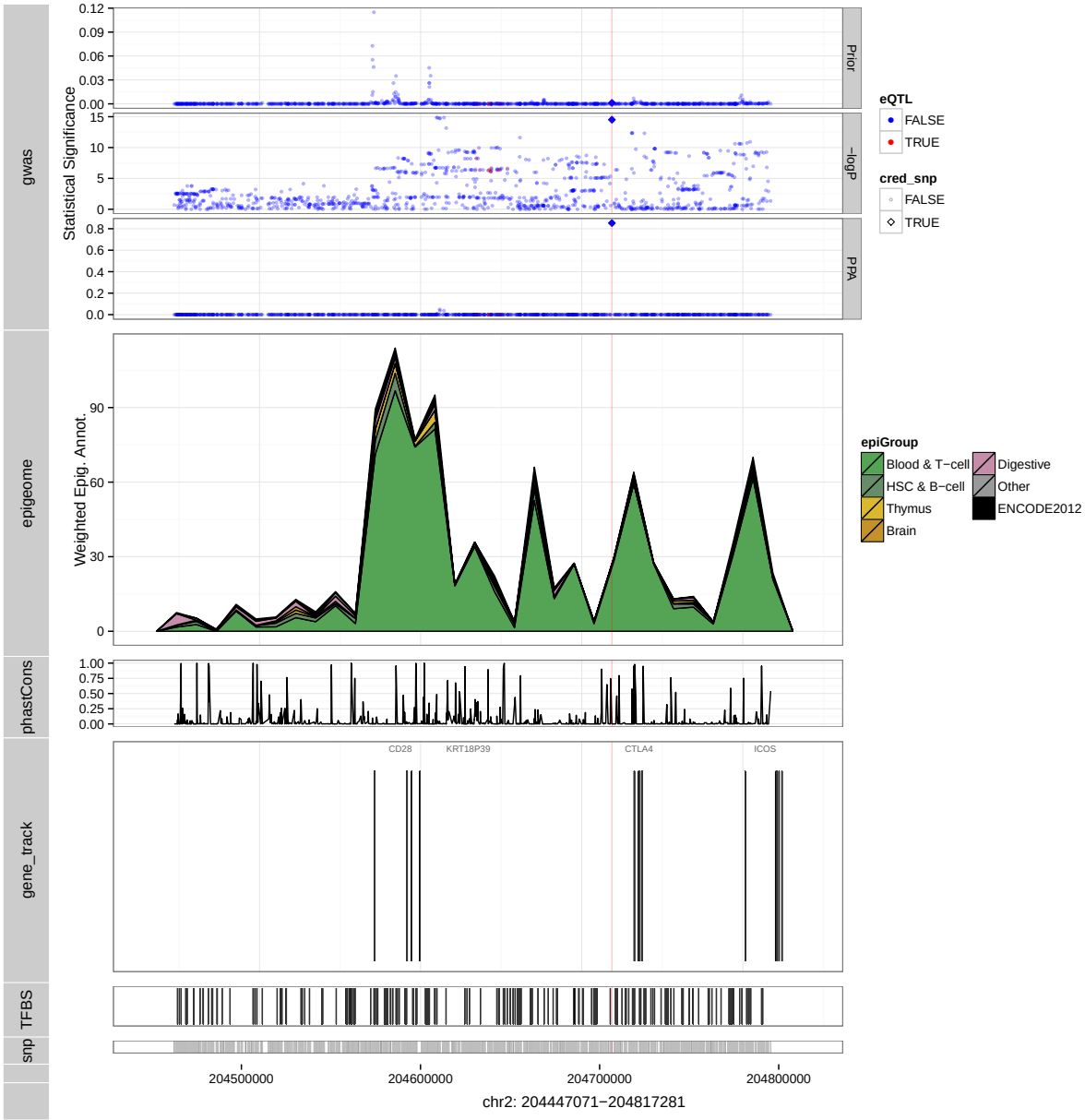
Celiac Disease



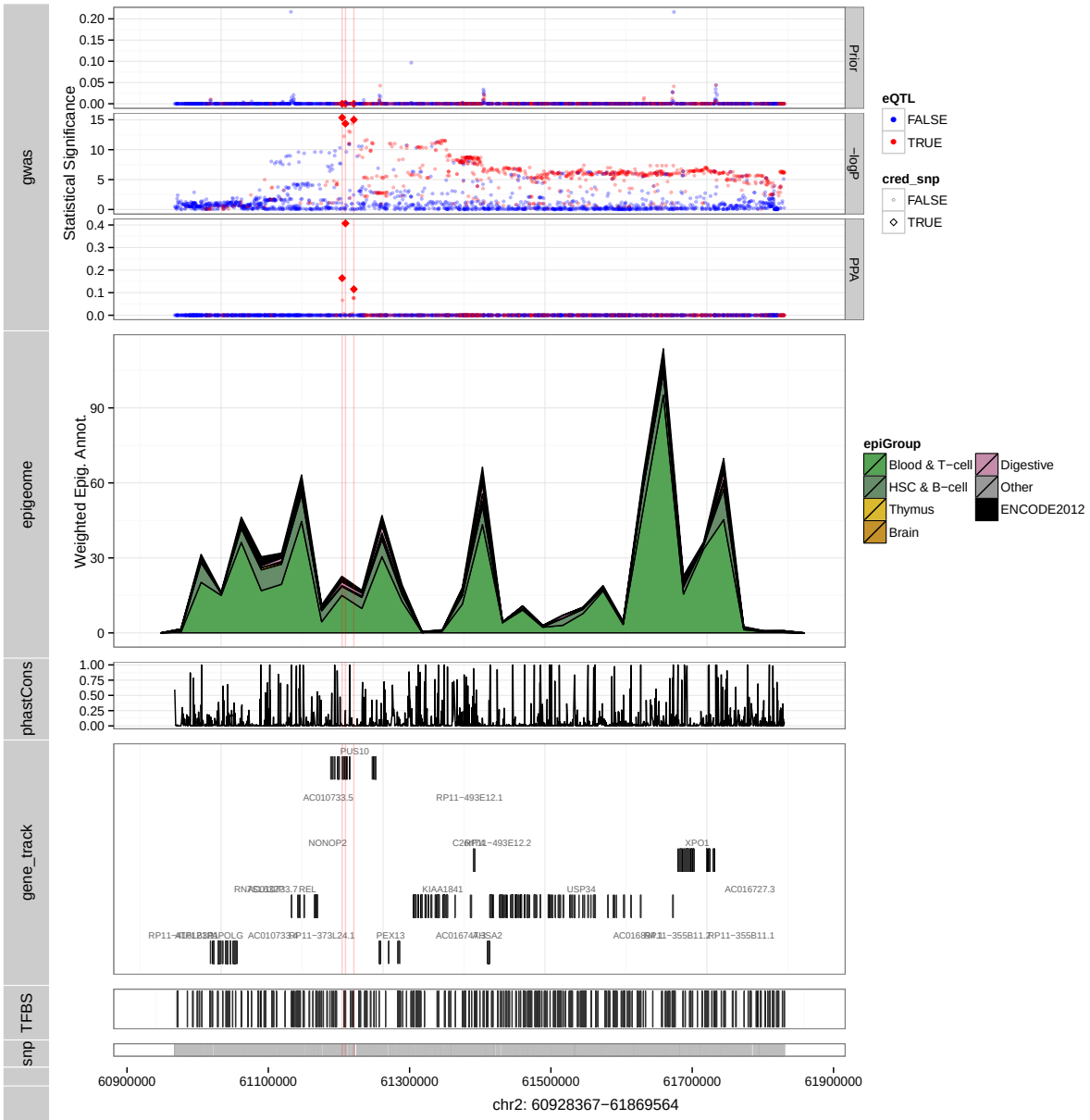
Celiac Disease



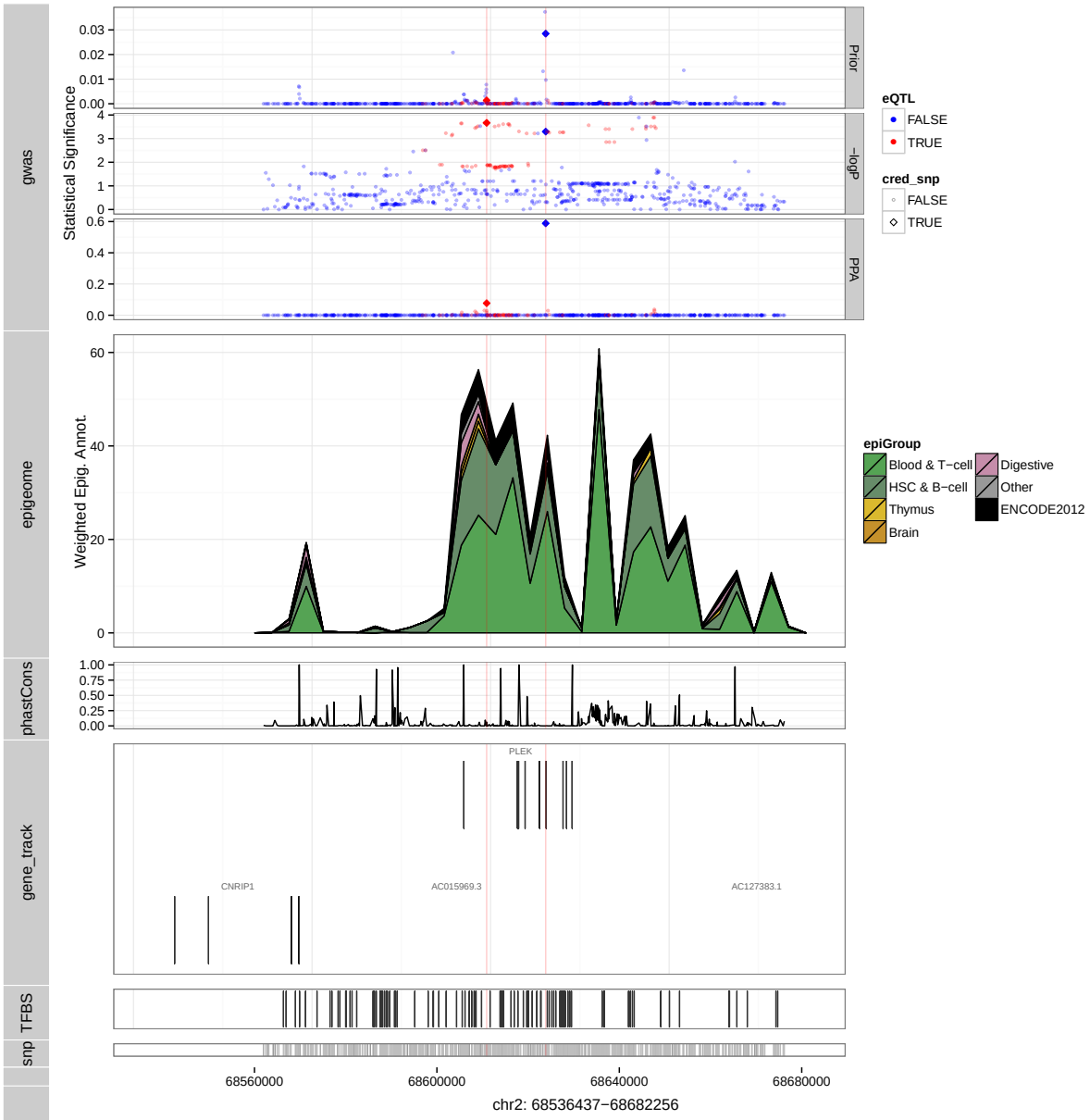
Celiac Disease



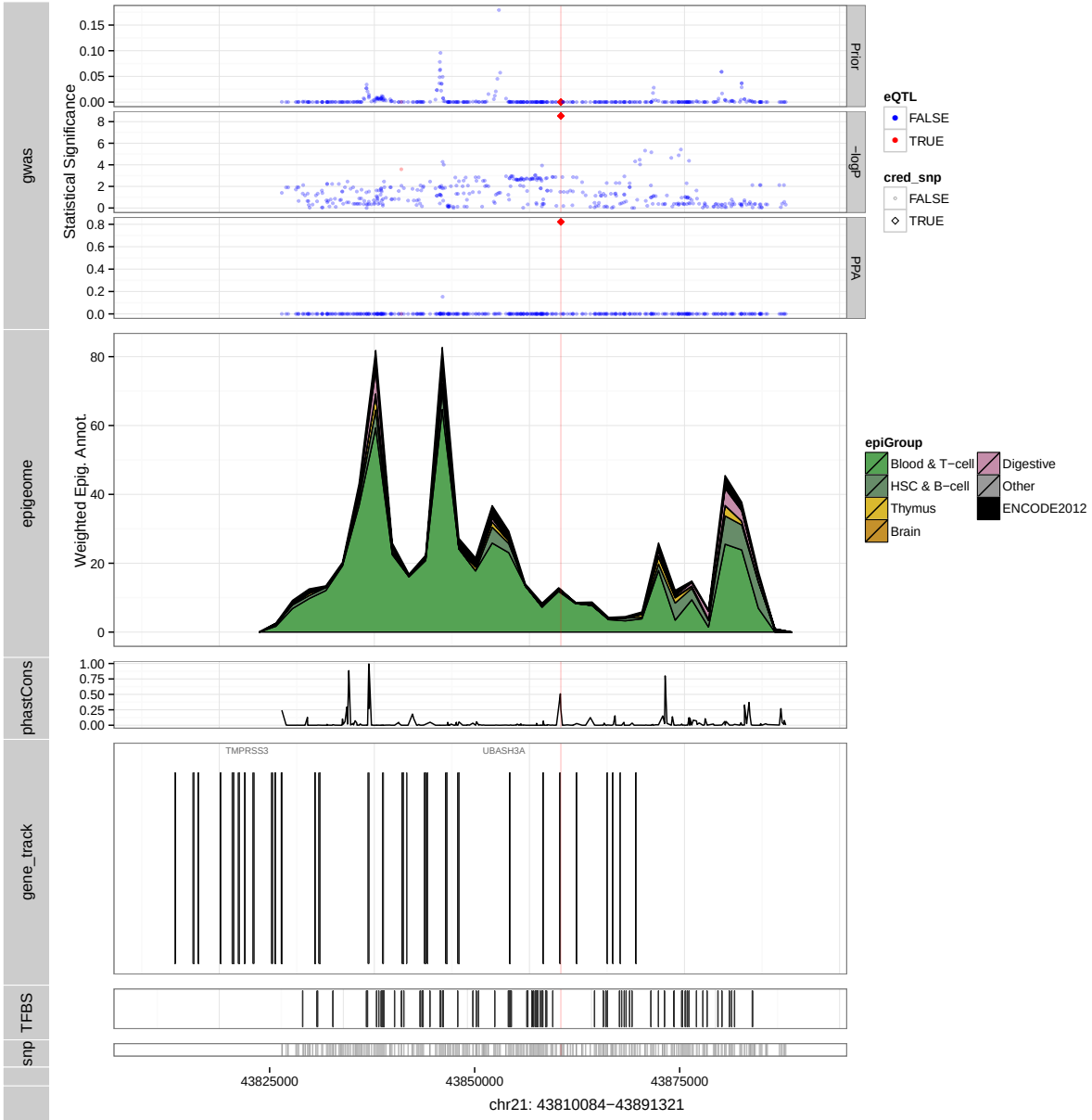
Celiac Disease



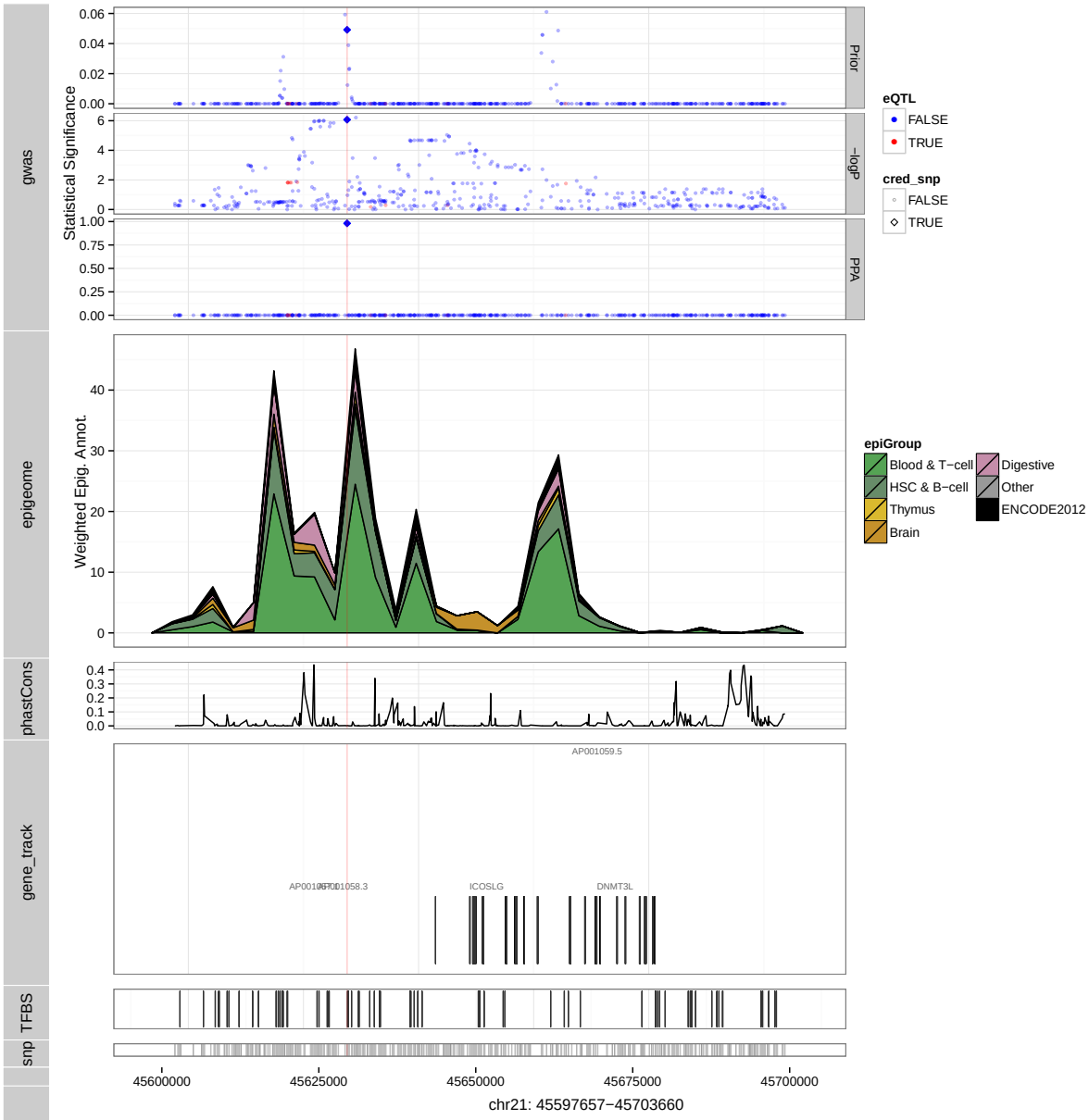
Celiac Disease



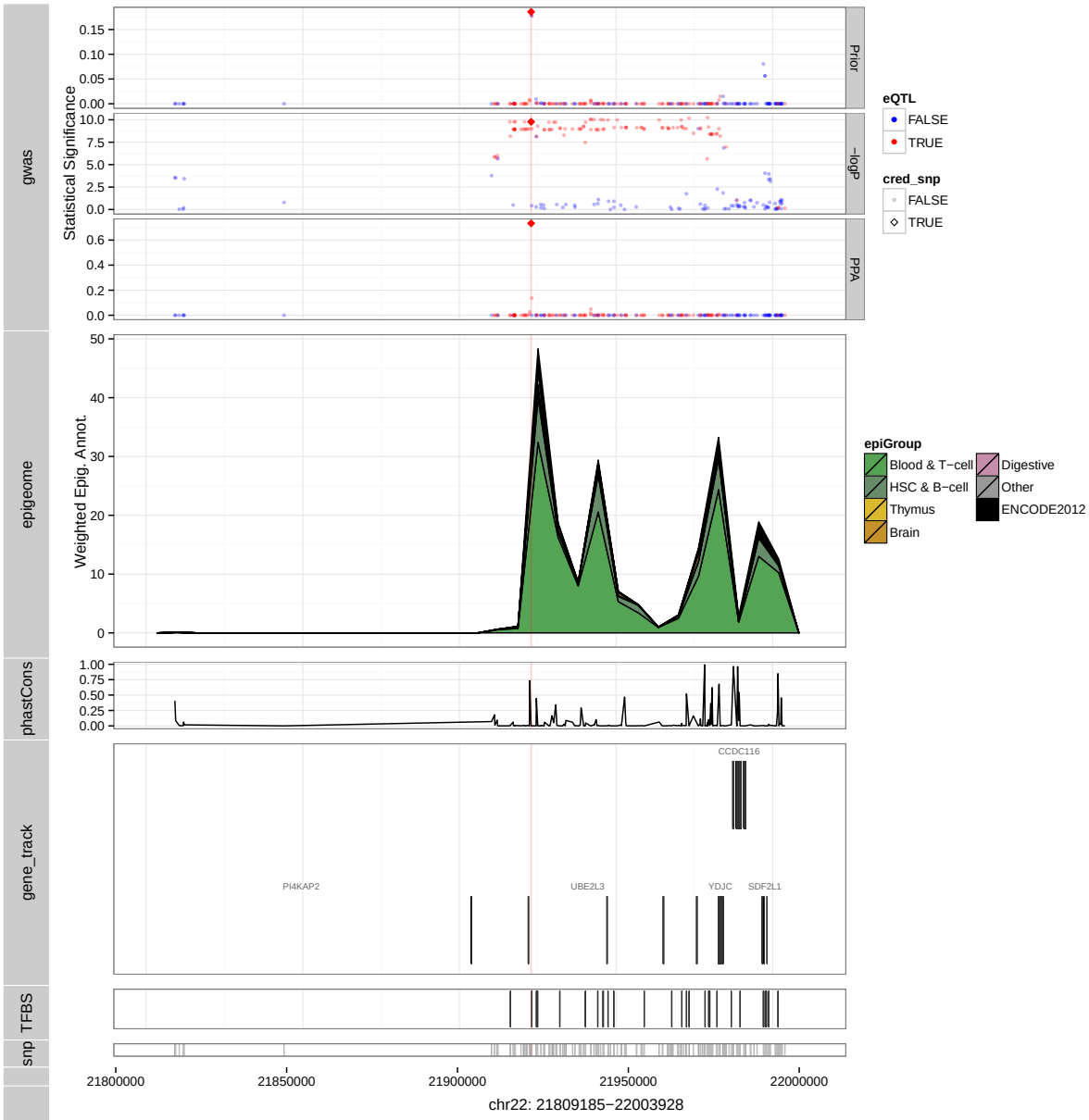
Celiac Disease



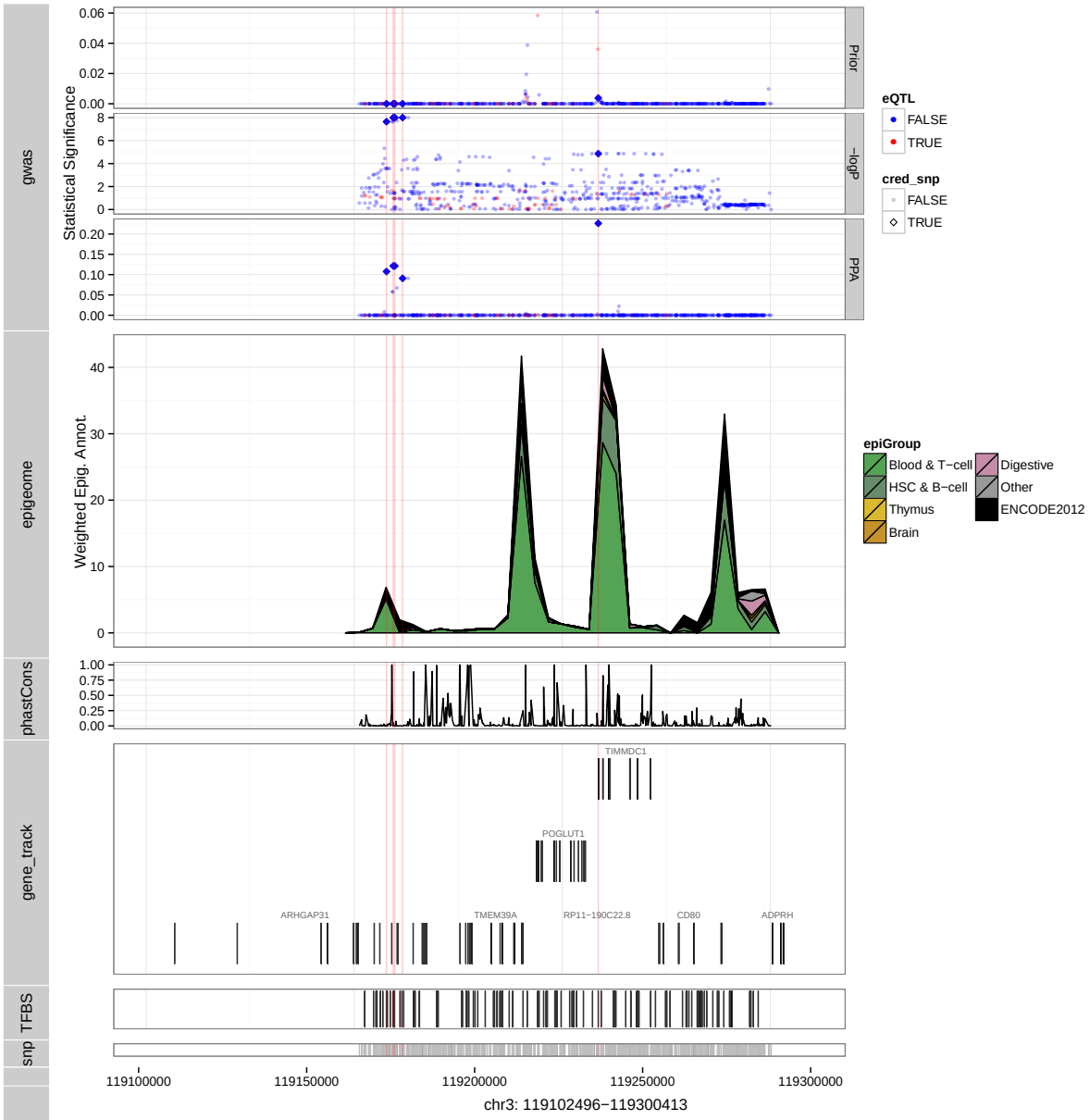
Celiac Disease



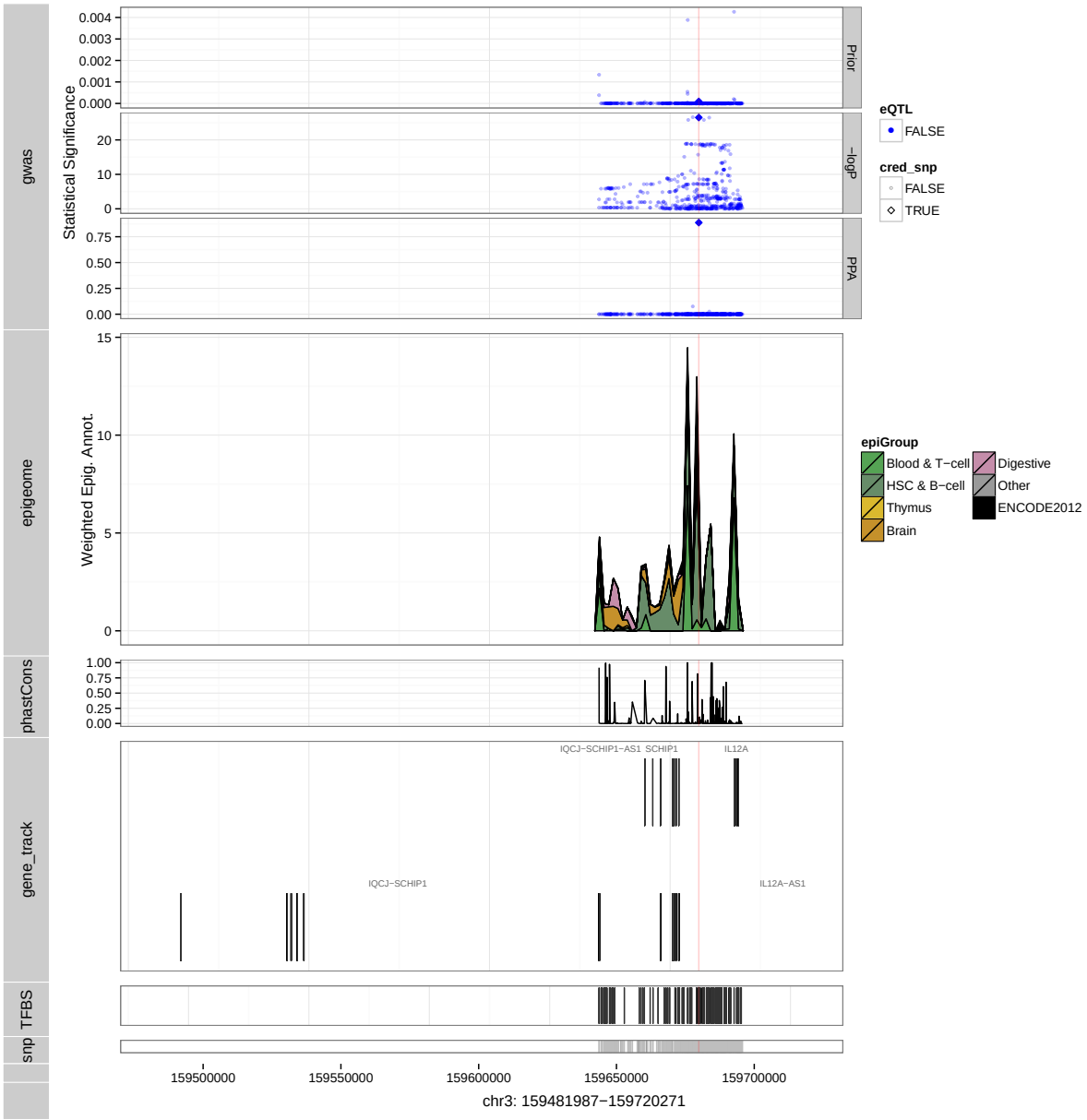
Celiac Disease



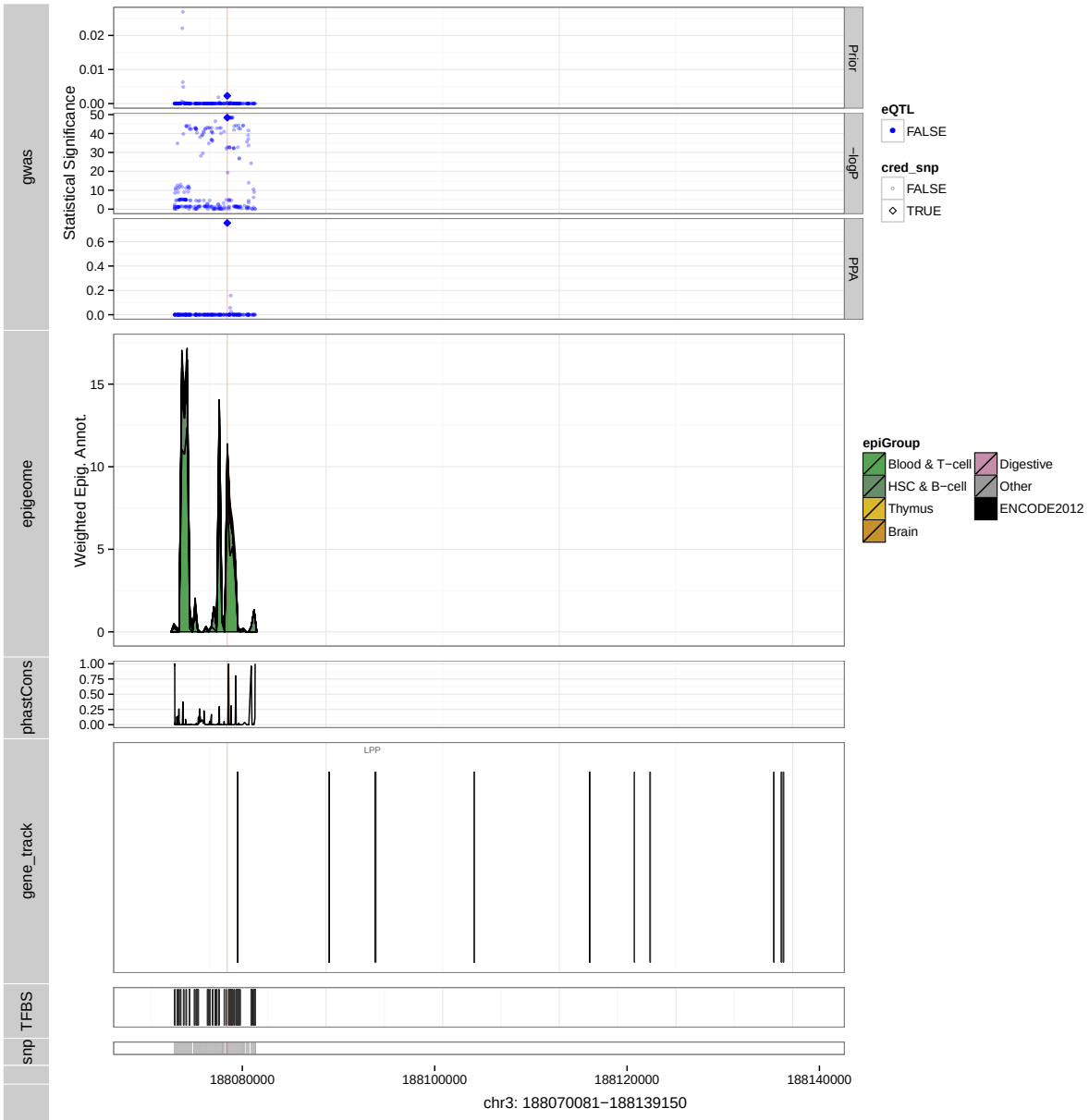
Celiac Disease



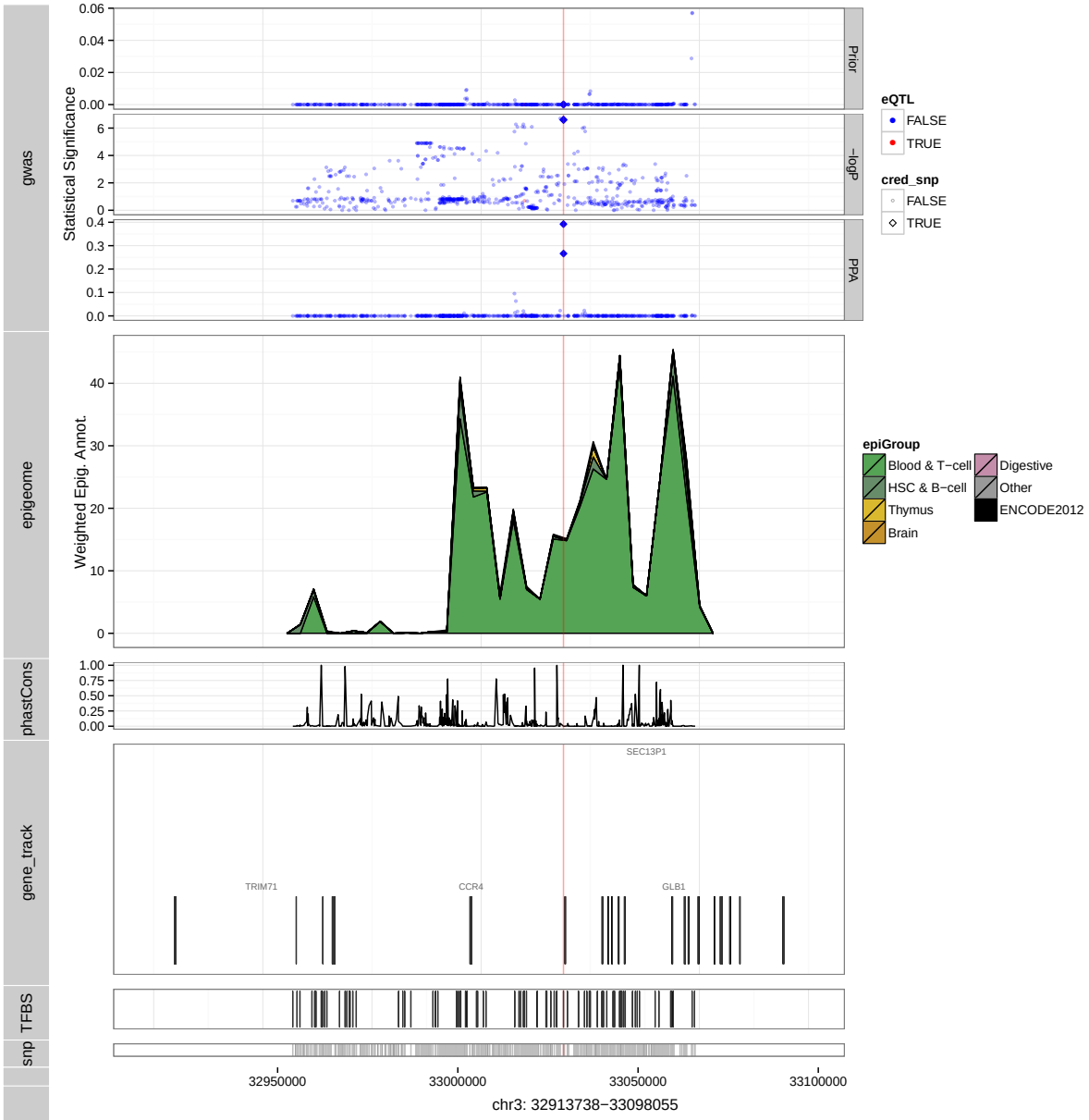
Celiac Disease



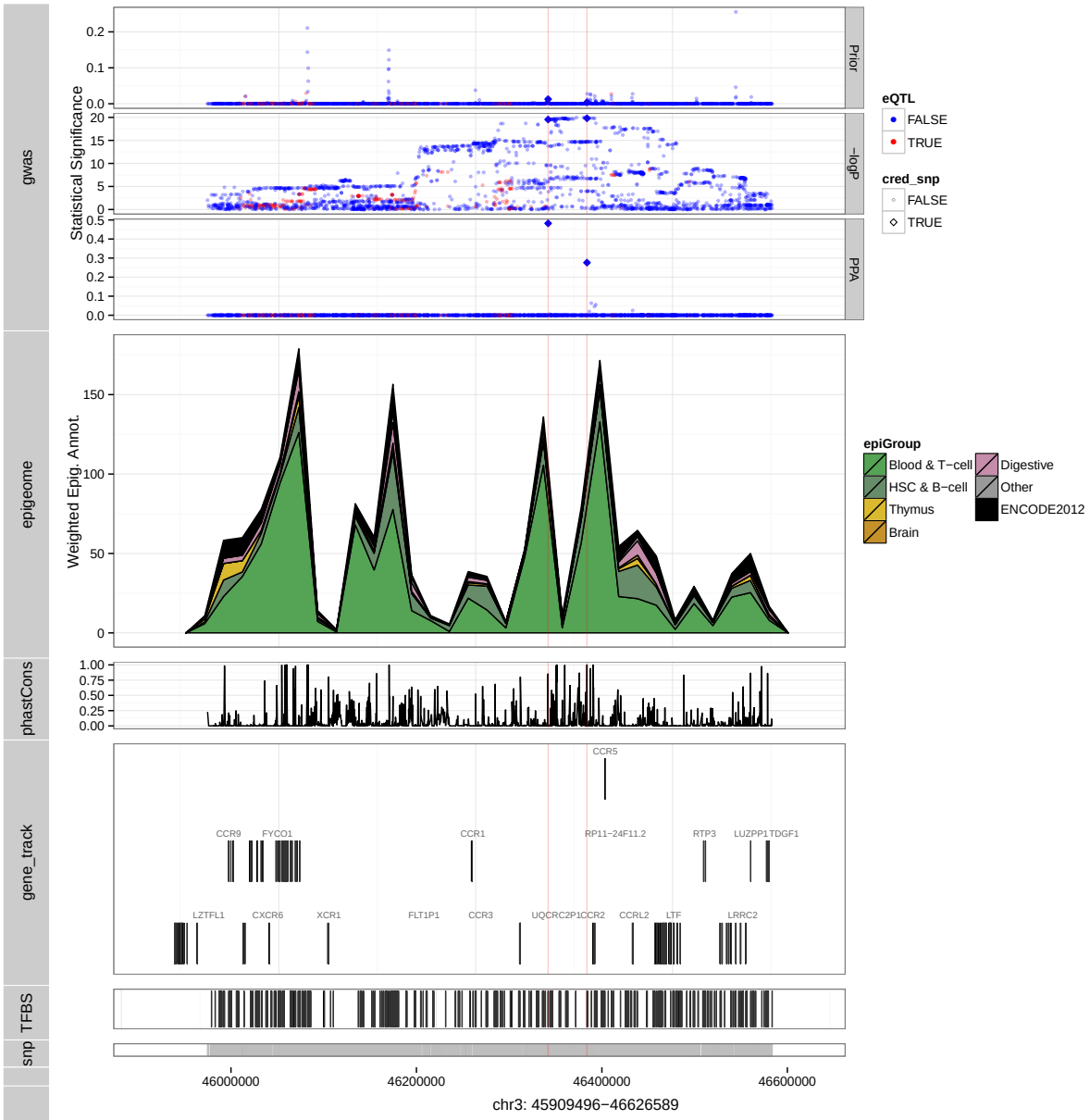
Celiac Disease



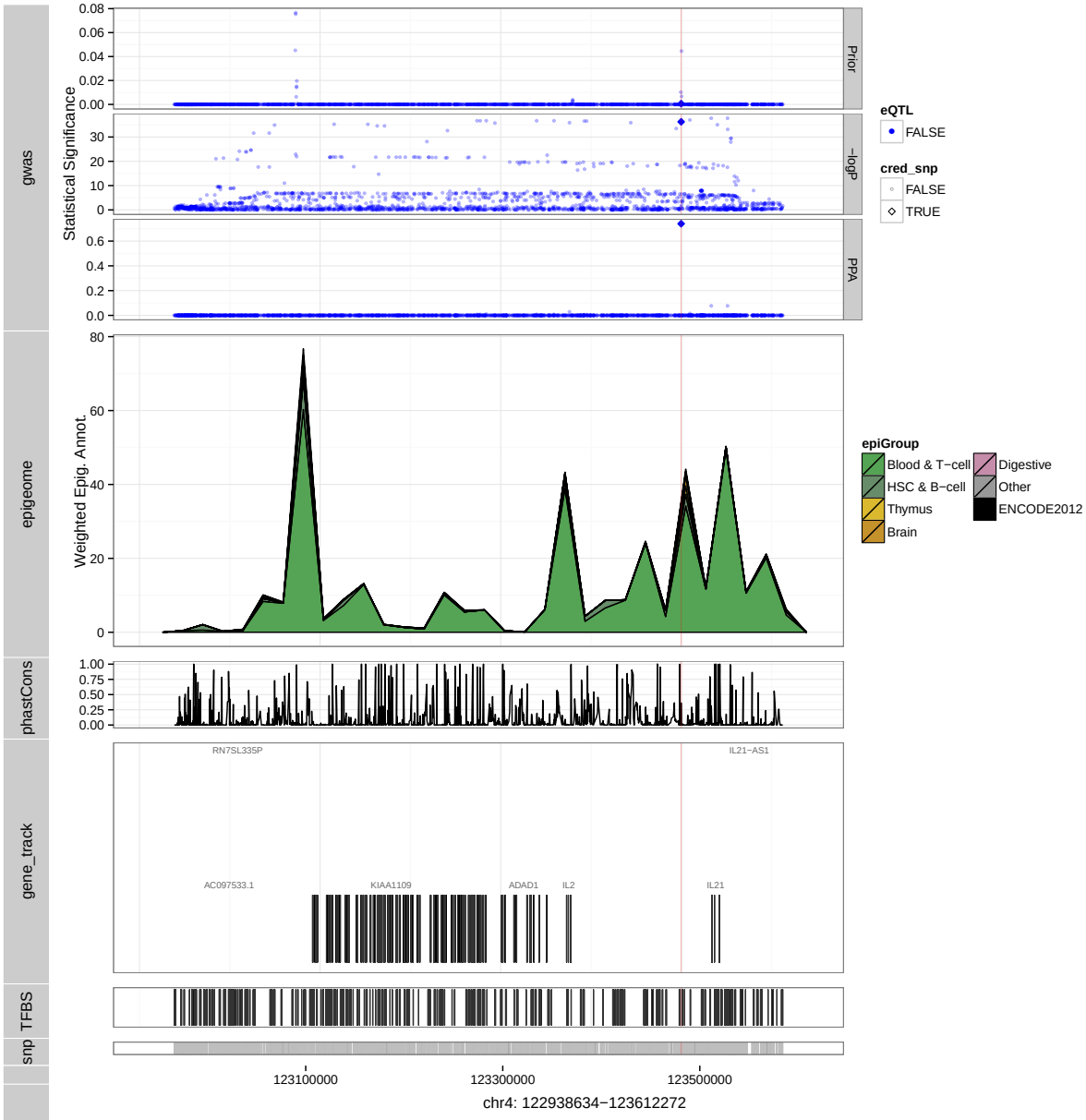
Celiac Disease



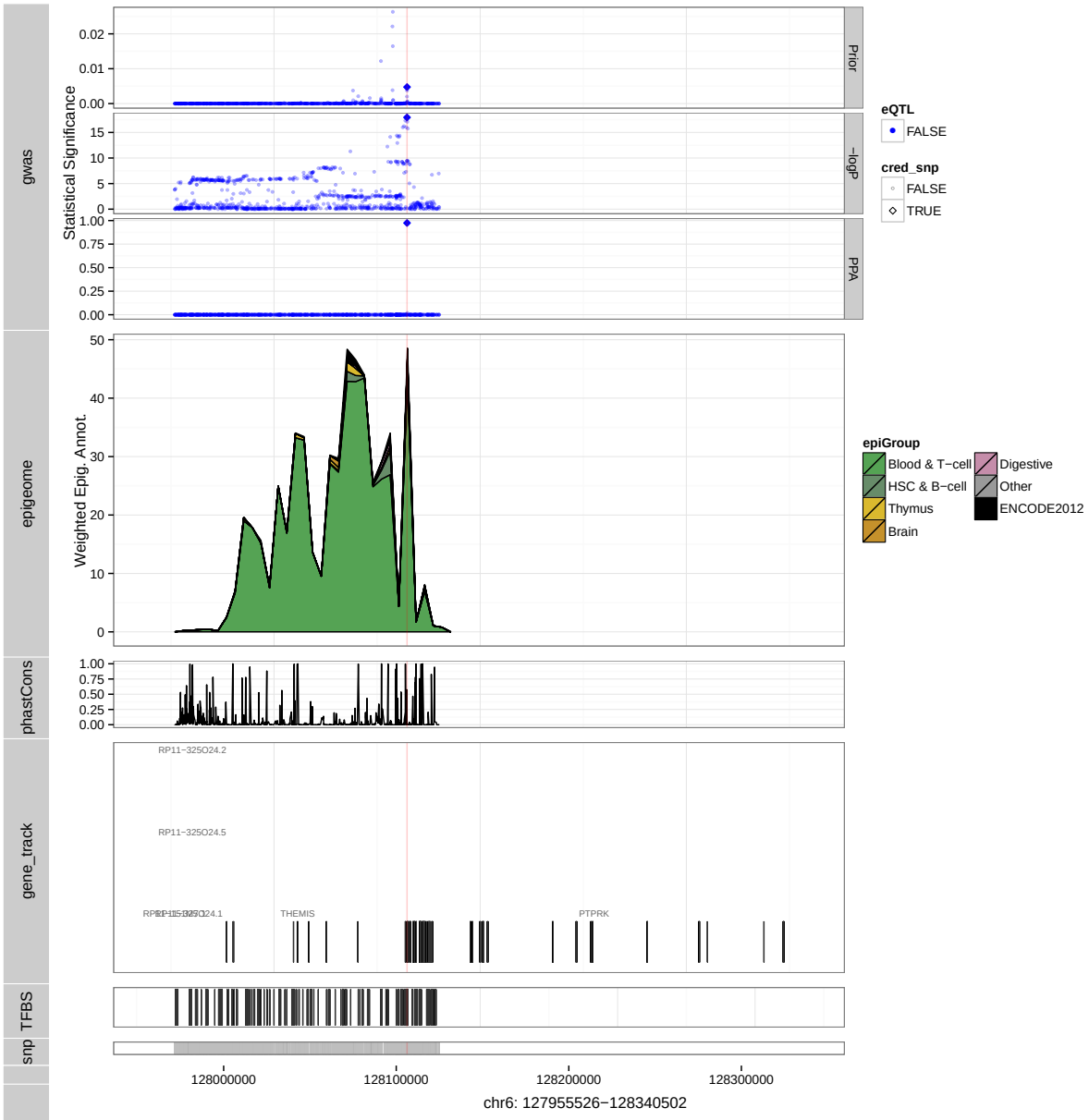
Celiac Disease



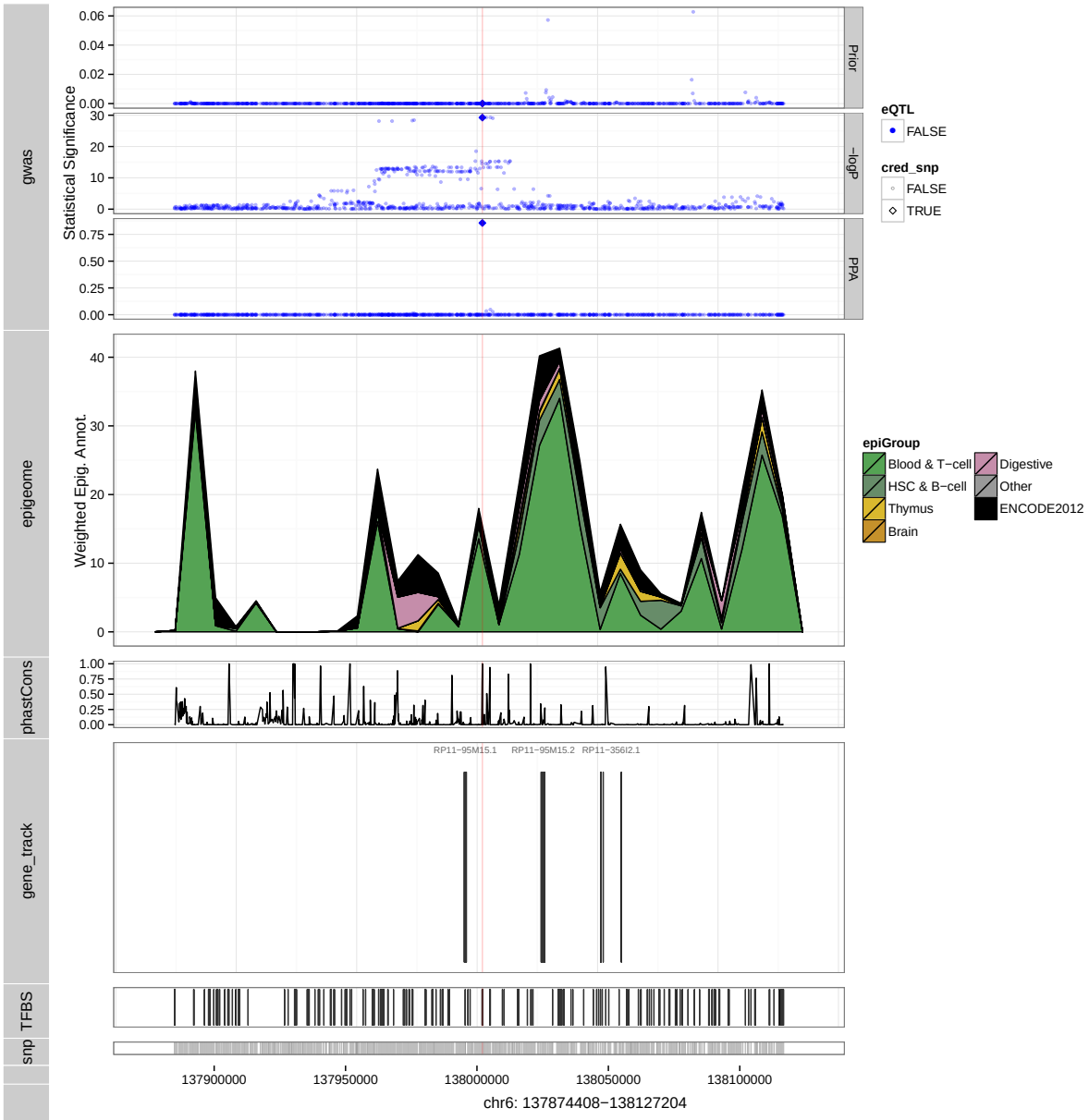
Celiac Disease



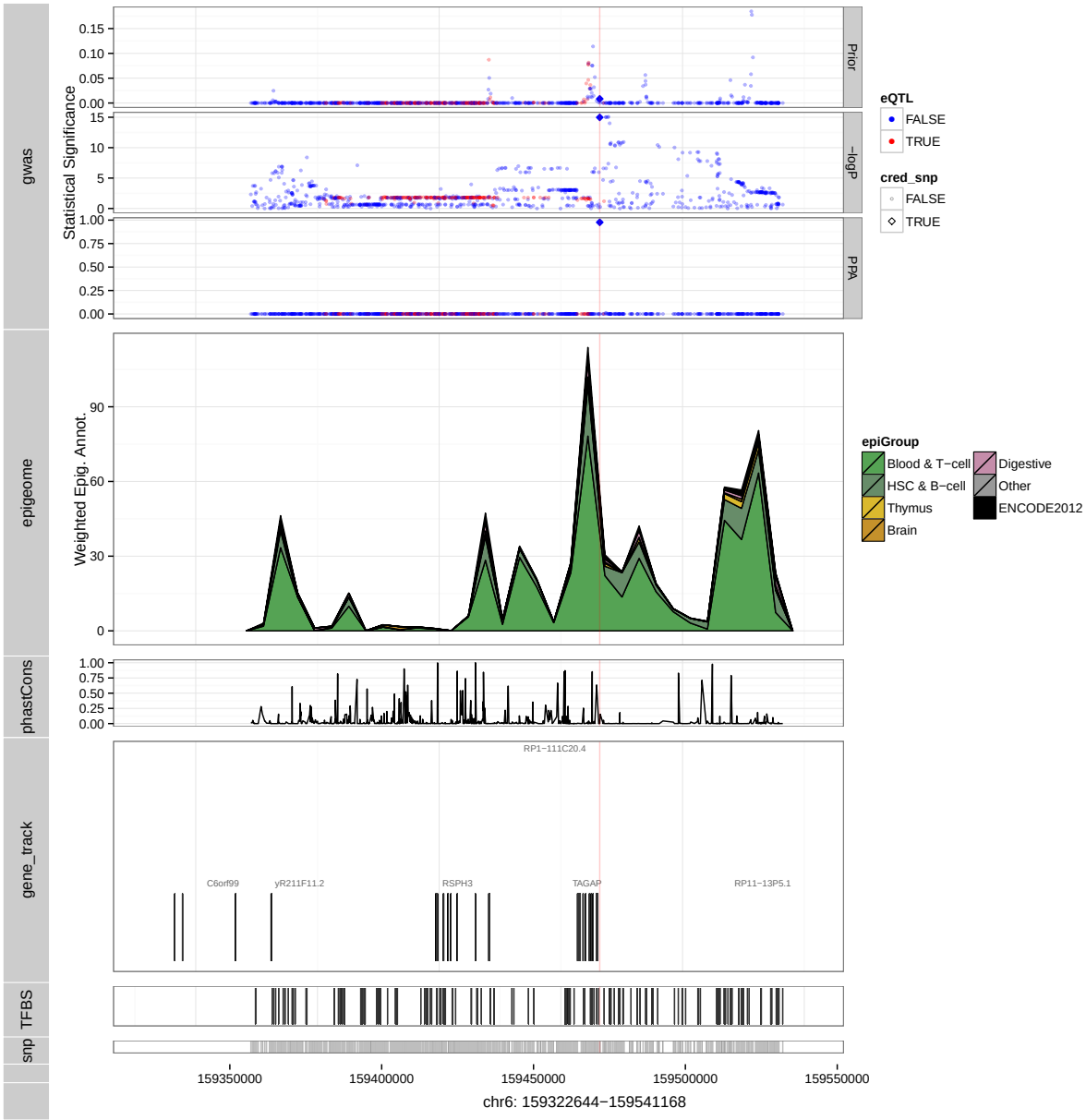
Celiac Disease



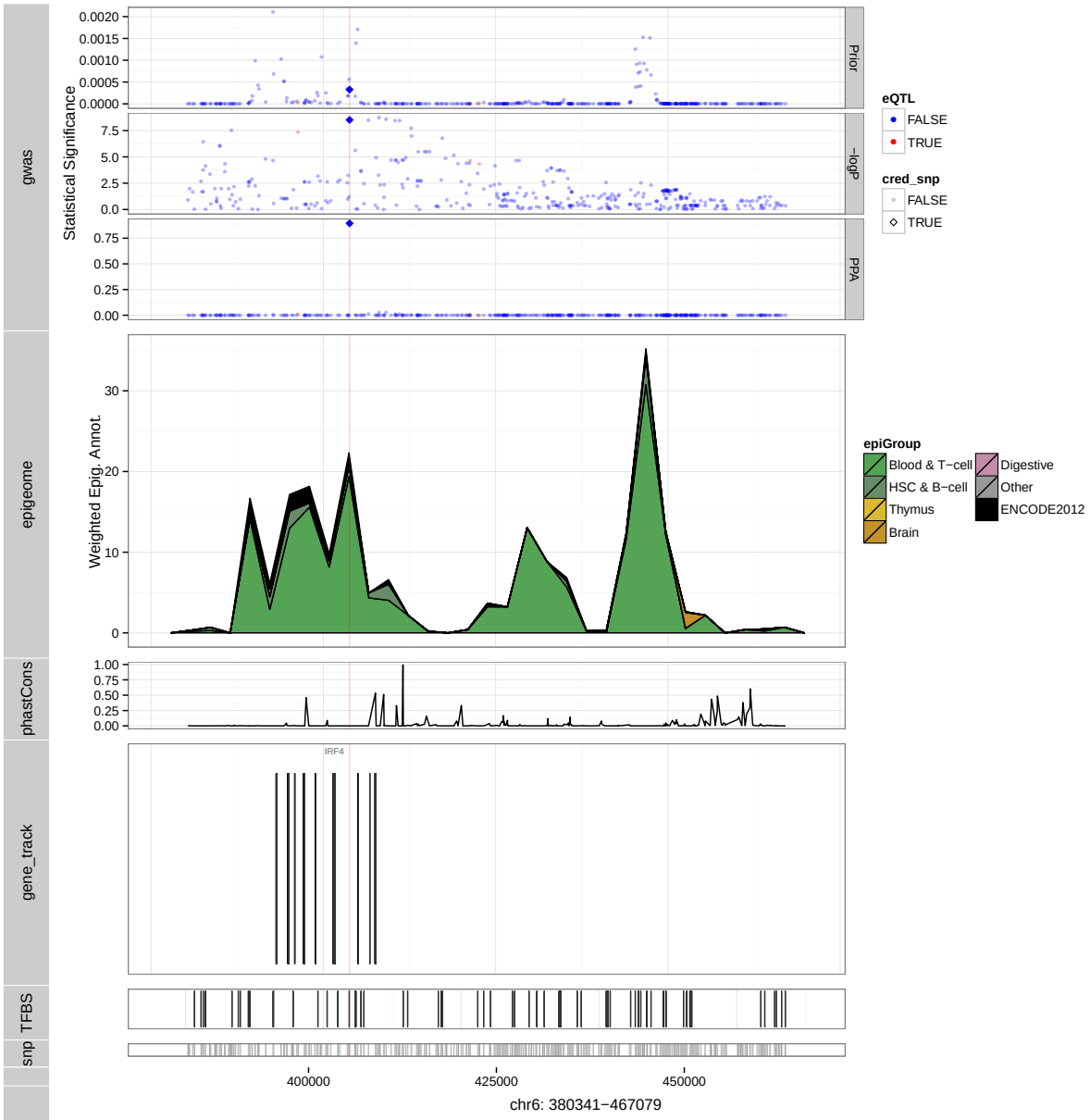
Celiac Disease



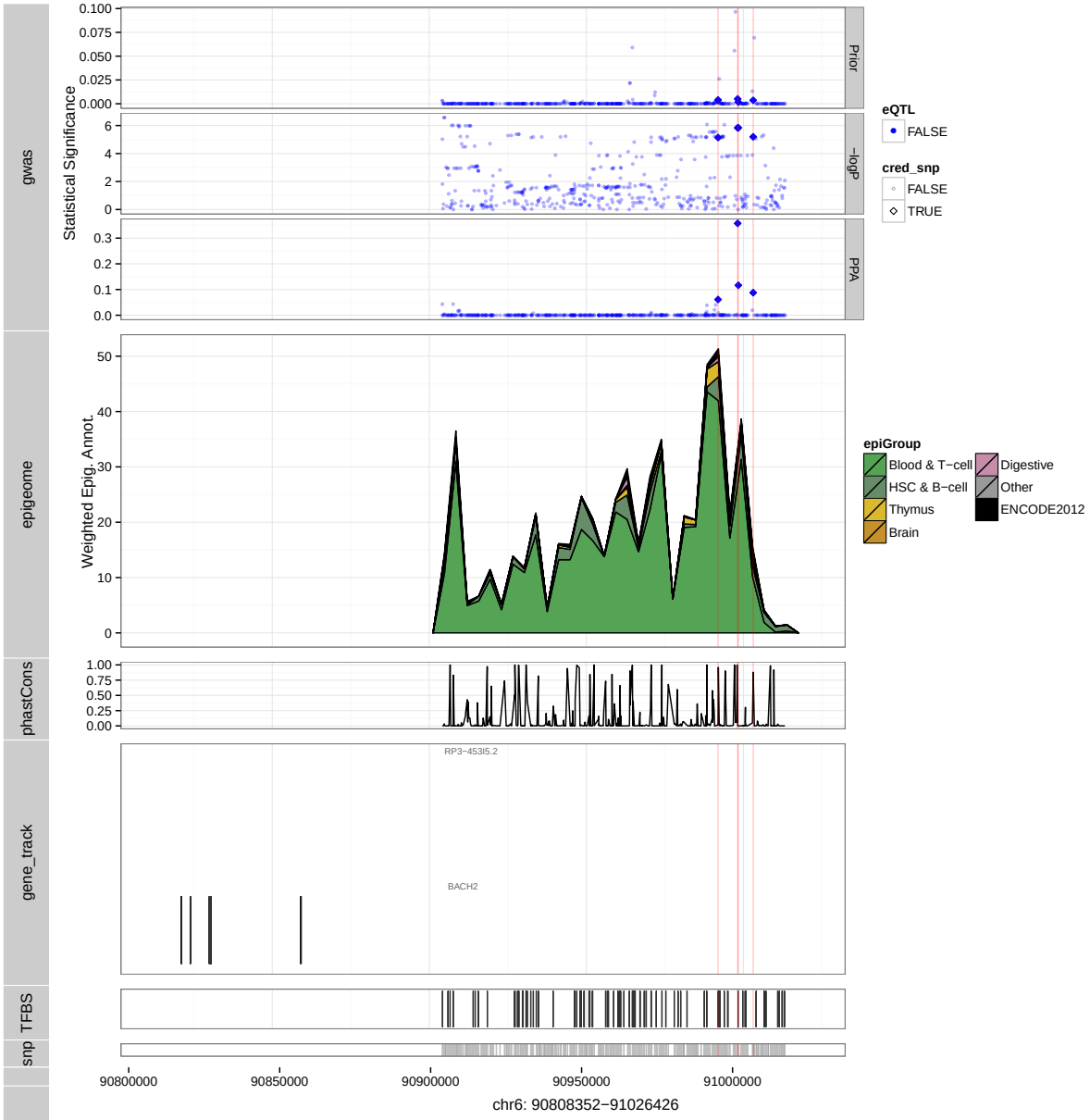
Celiac Disease



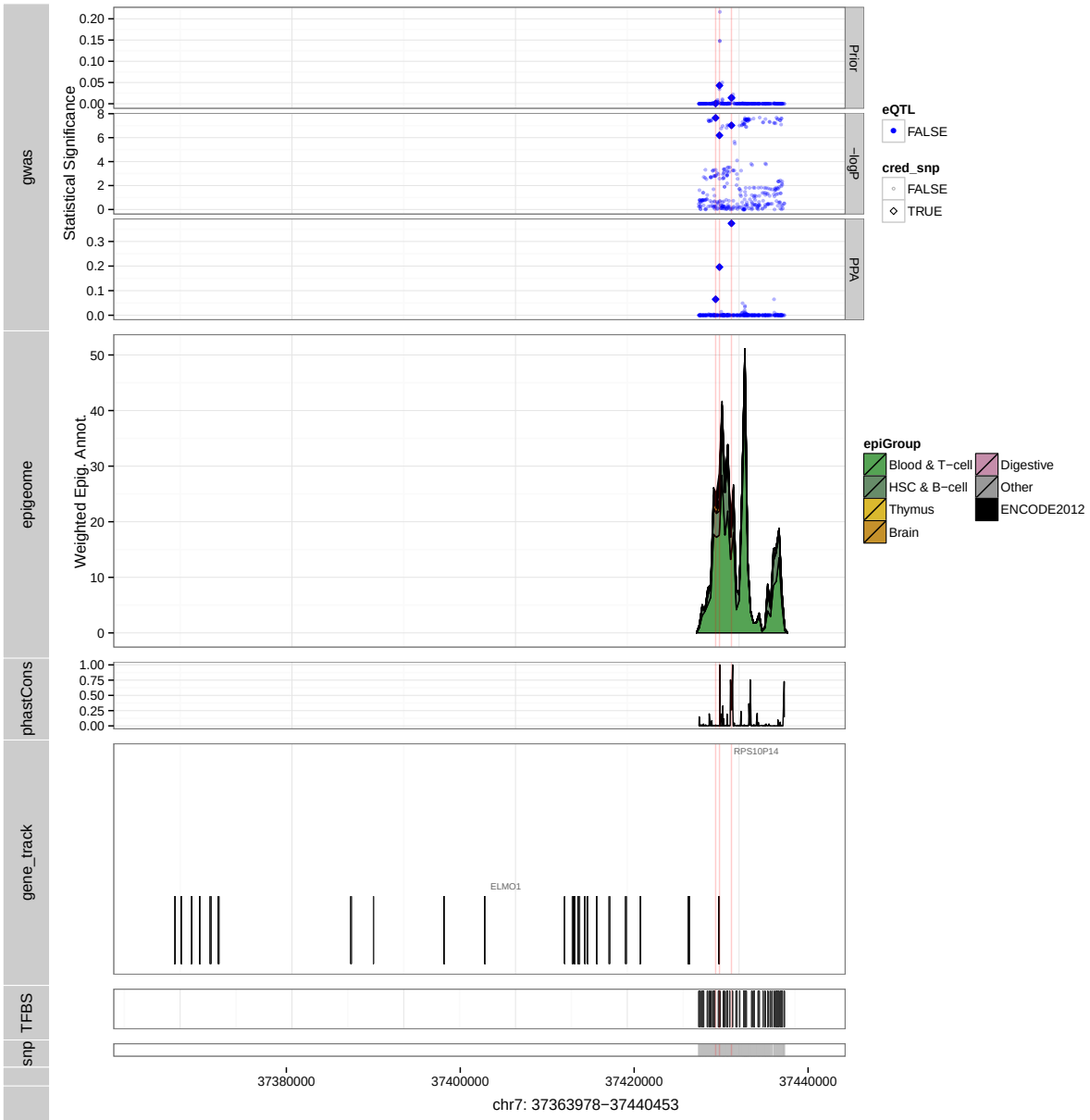
Celiac Disease



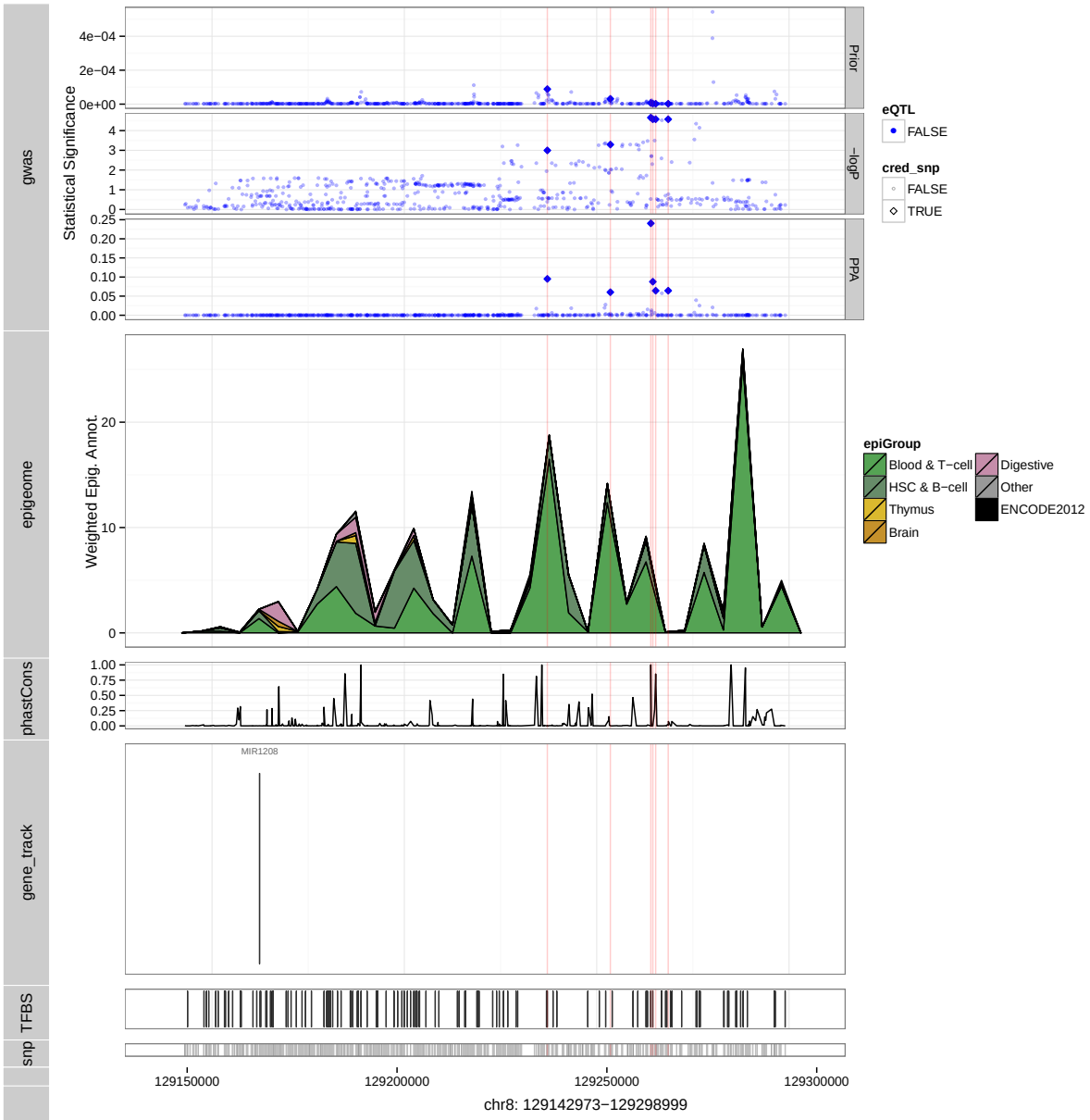
Celiac Disease



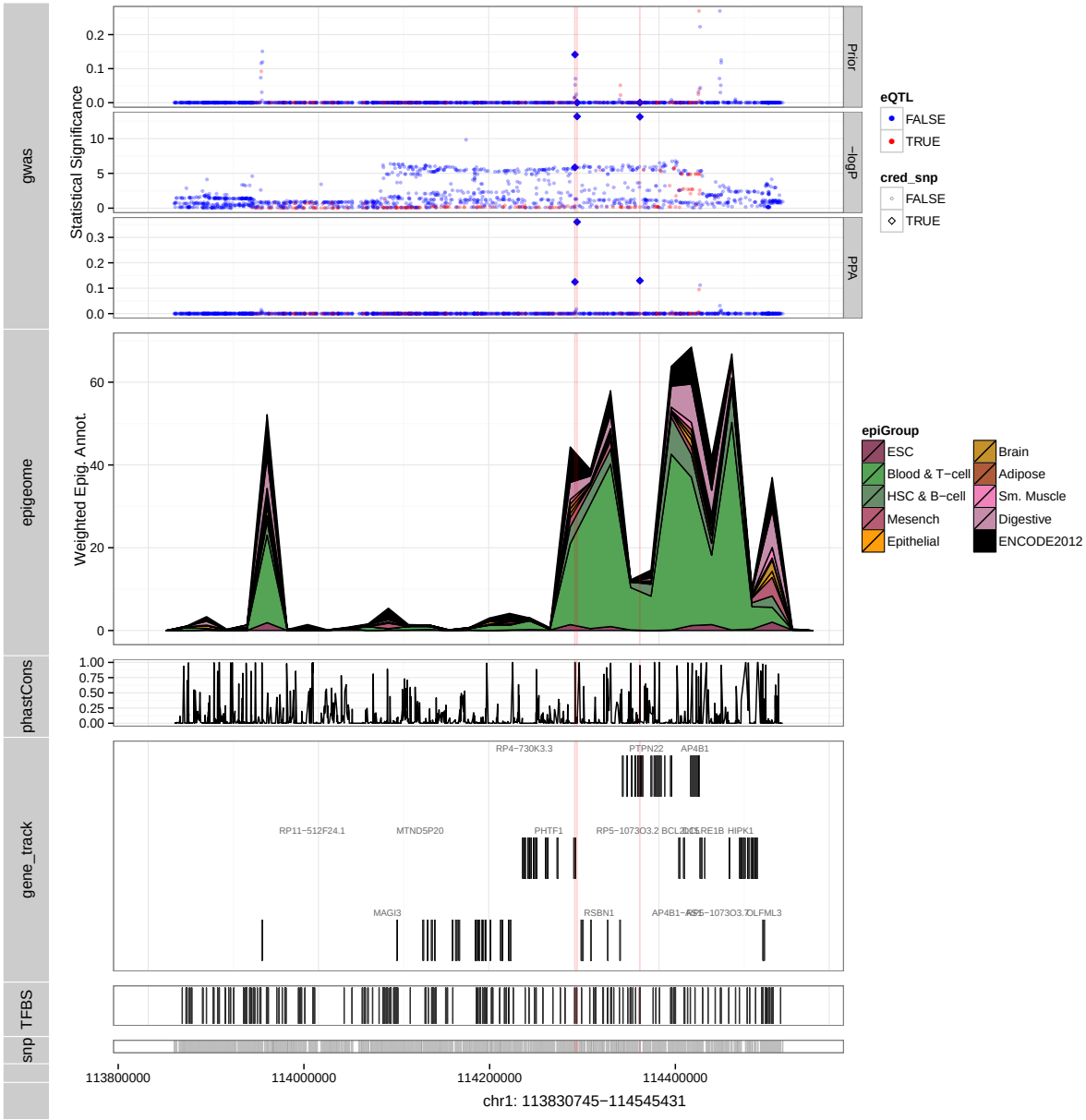
Celiac Disease



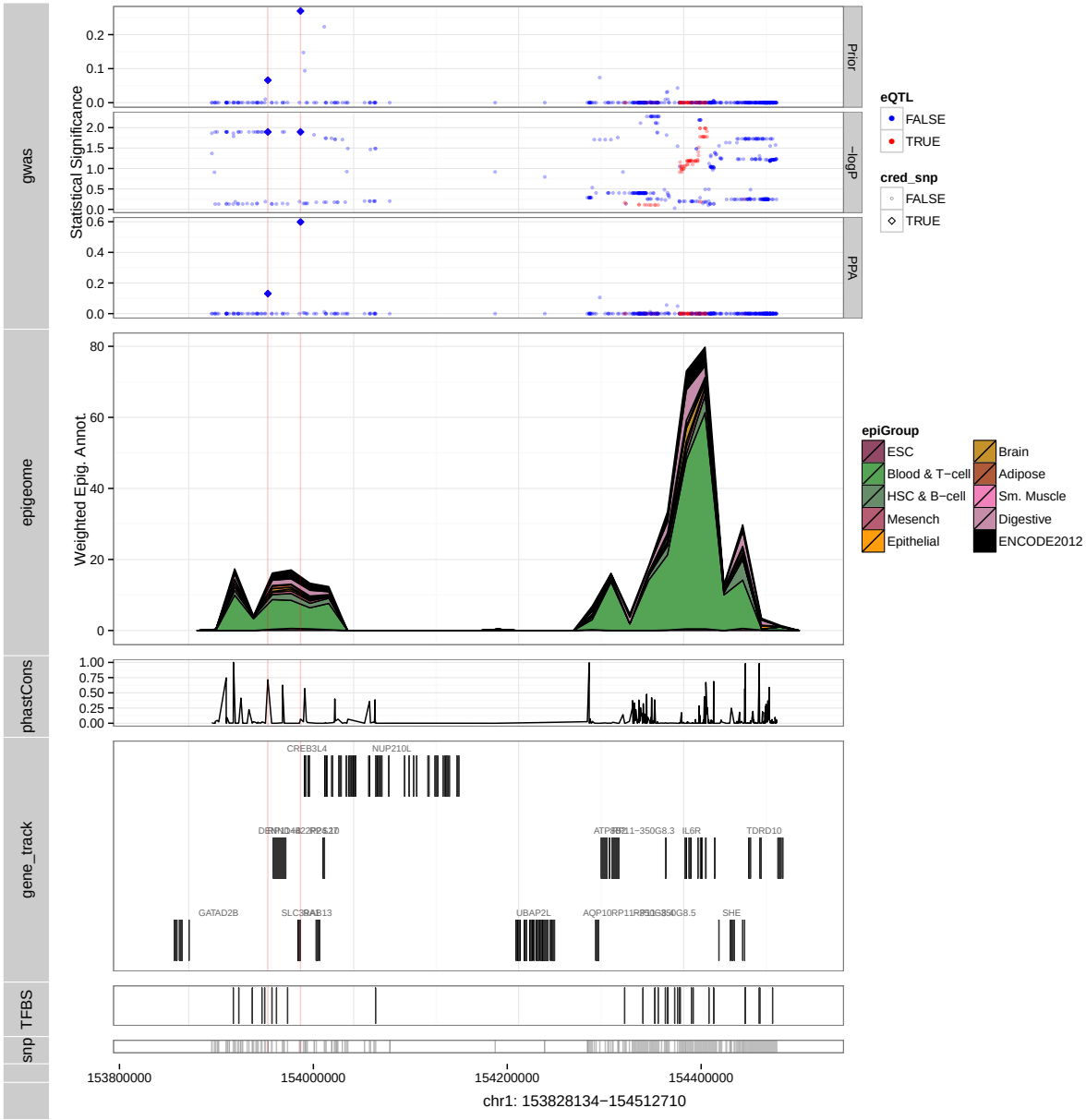
Celiac Disease



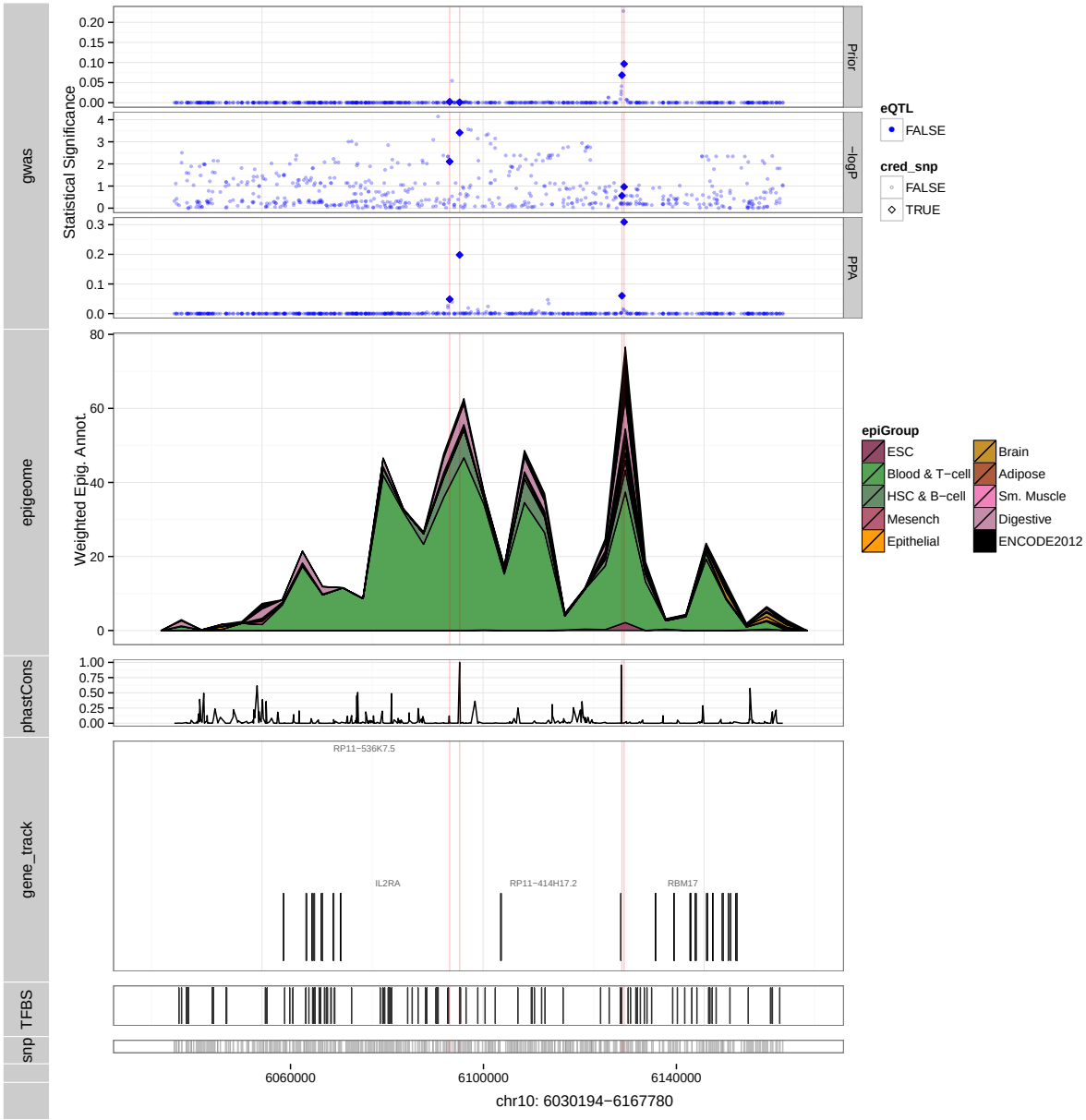
Juvenile Idiopathic Arthritis



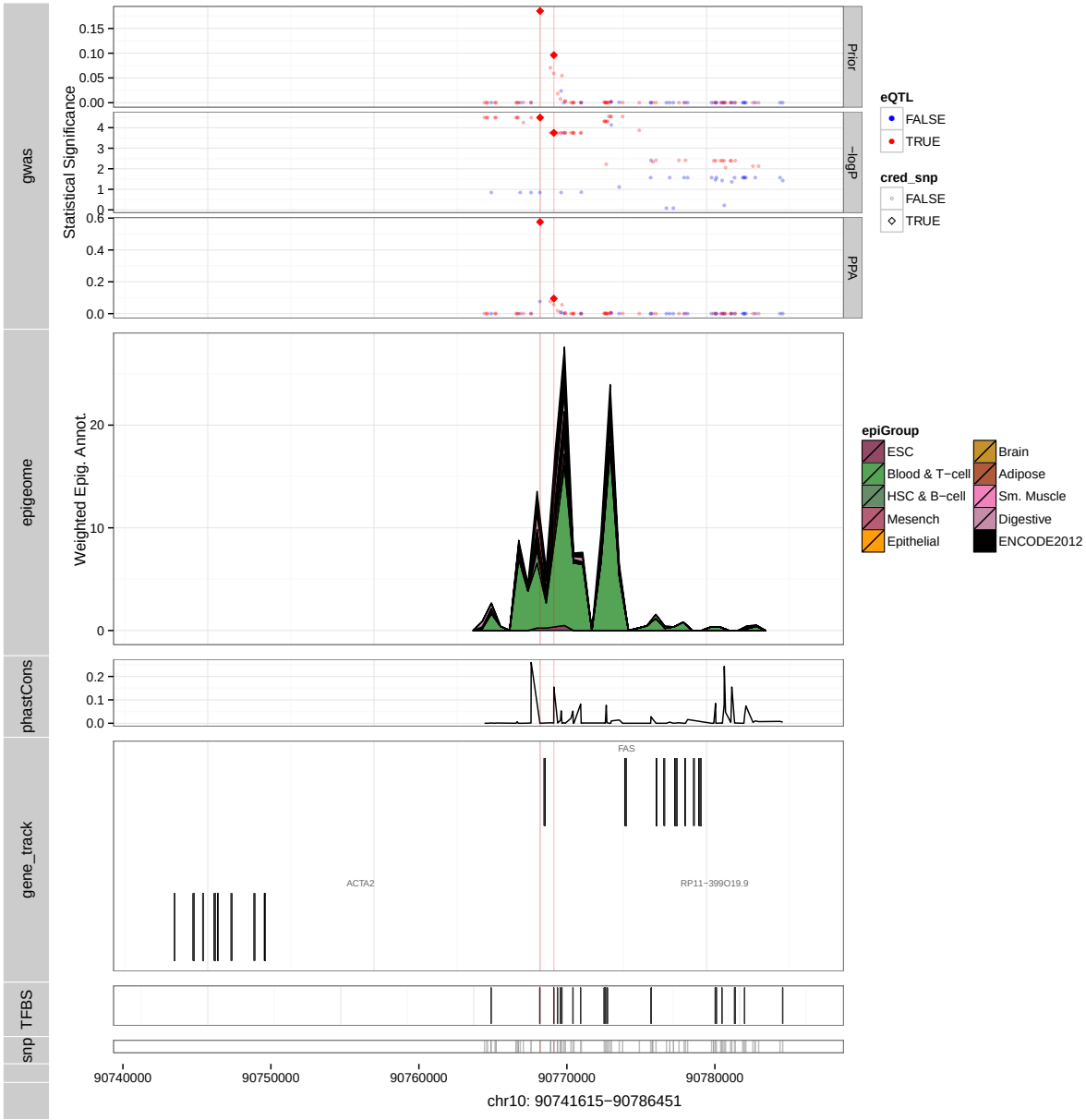
Juvenile Idiopathic Arthritis



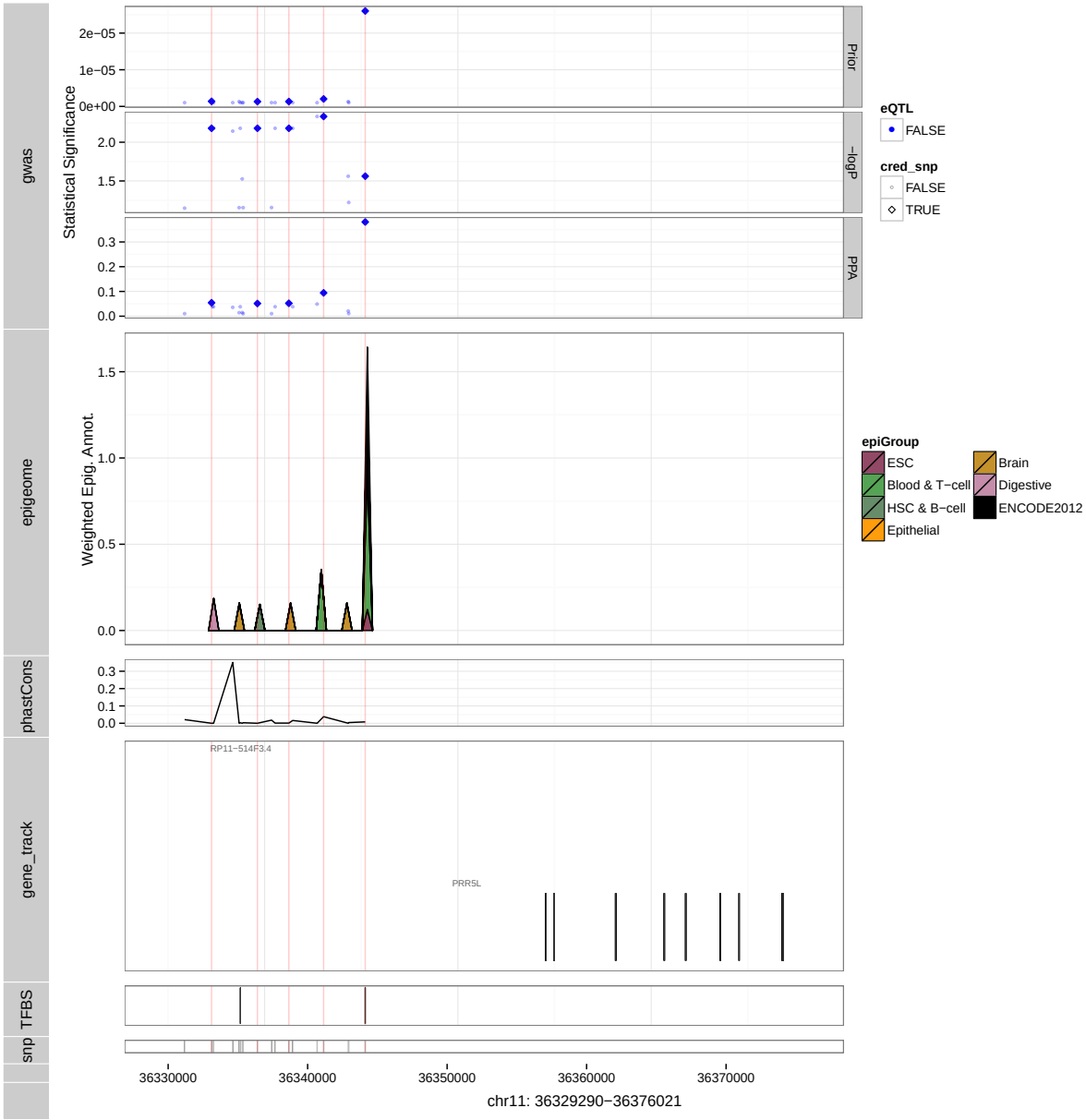
Juvenile Idiopathic Arthritis



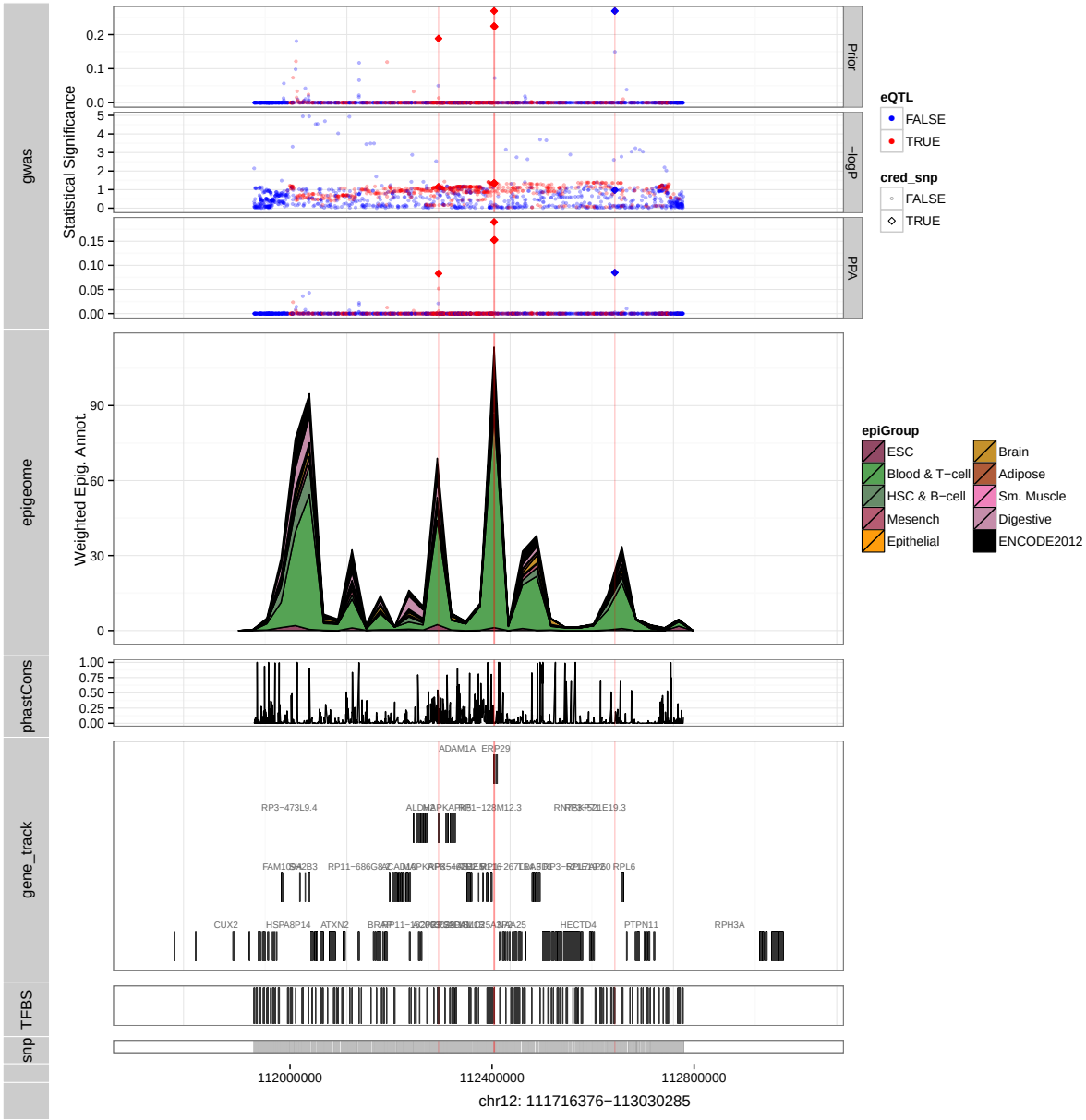
Juvenile Idiopathic Arthritis



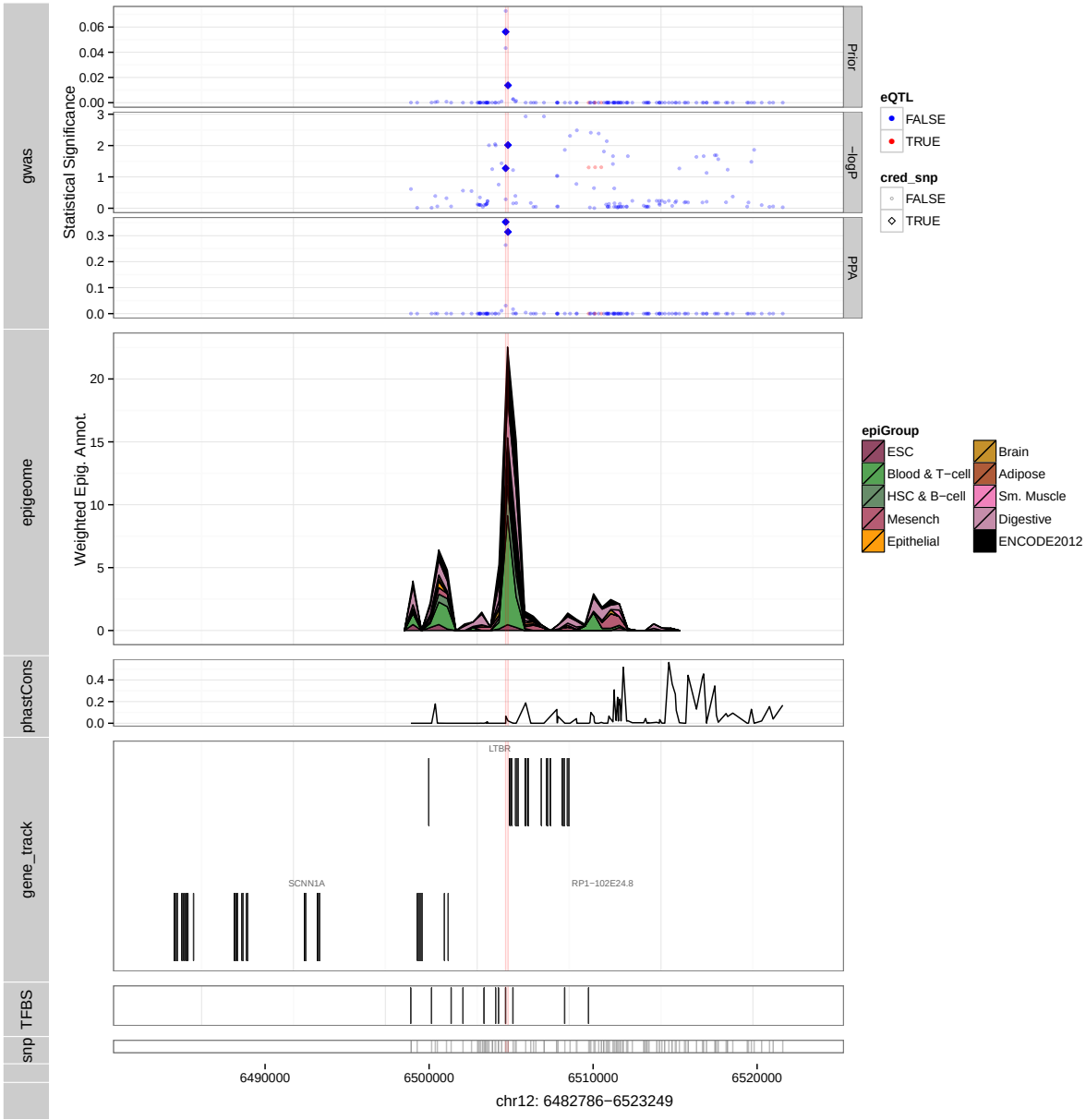
Juvenile Idiopathic Arthritis



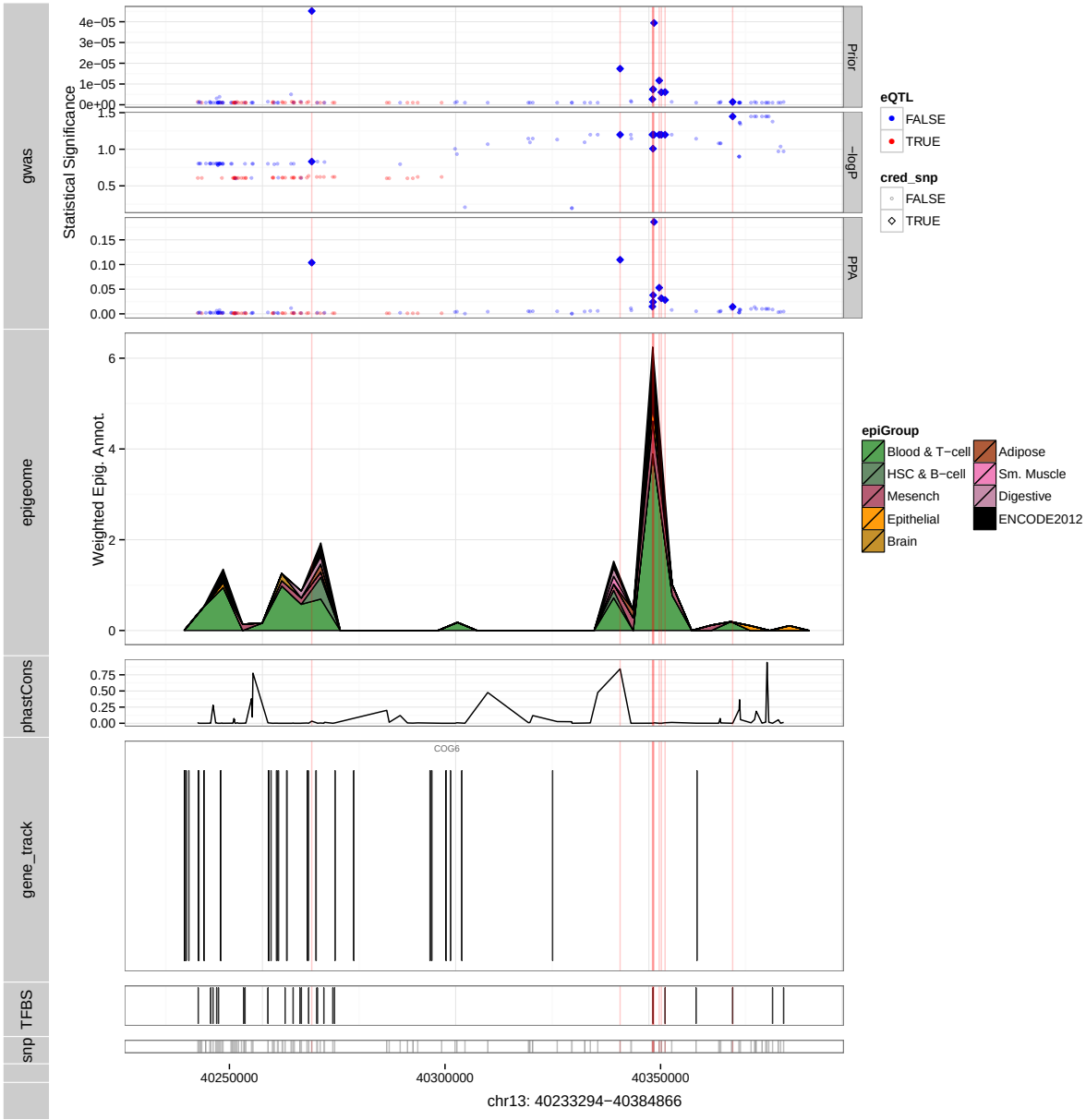
Juvenile Idiopathic Arthritis



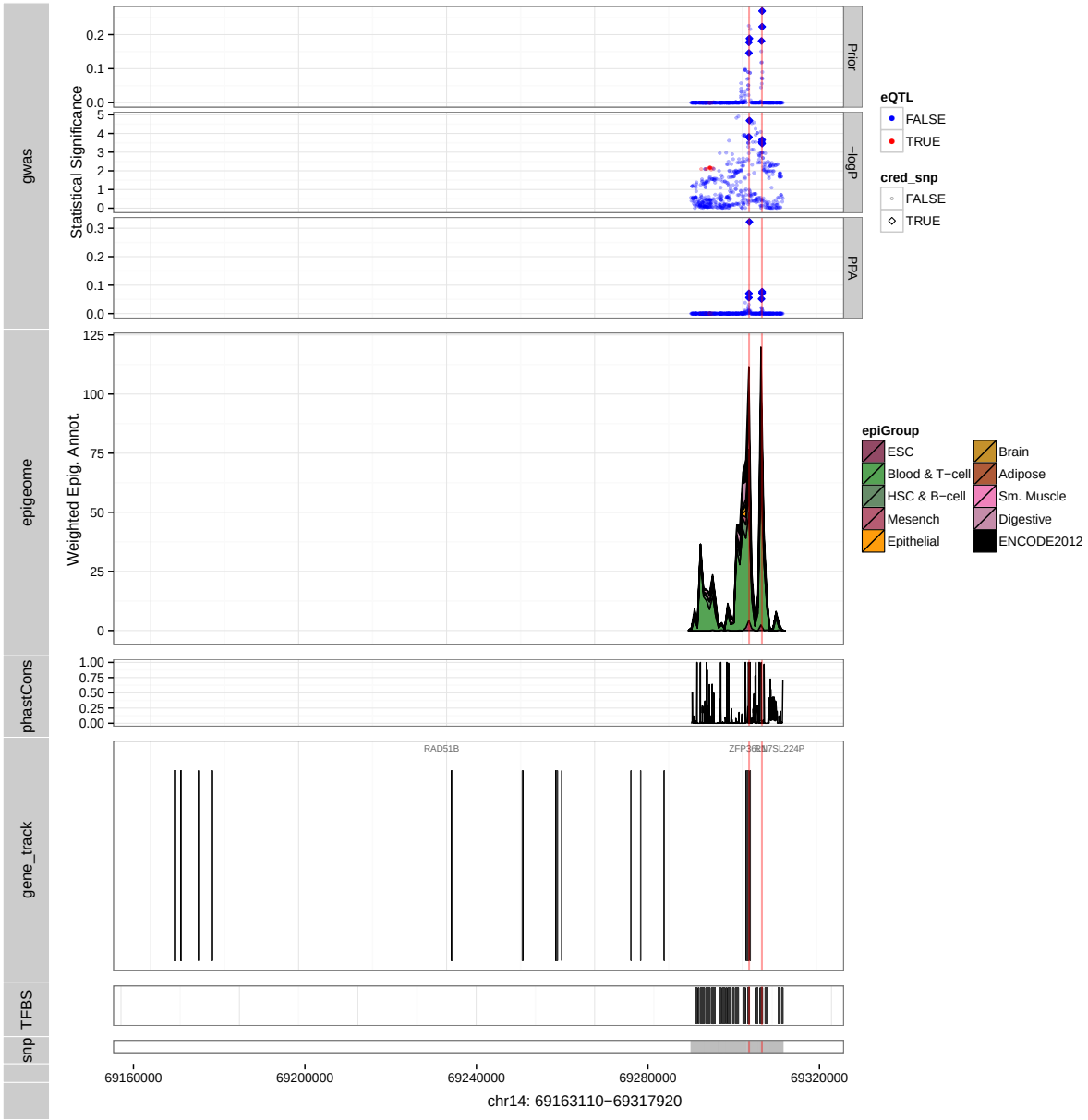
Juvenile Idiopathic Arthritis



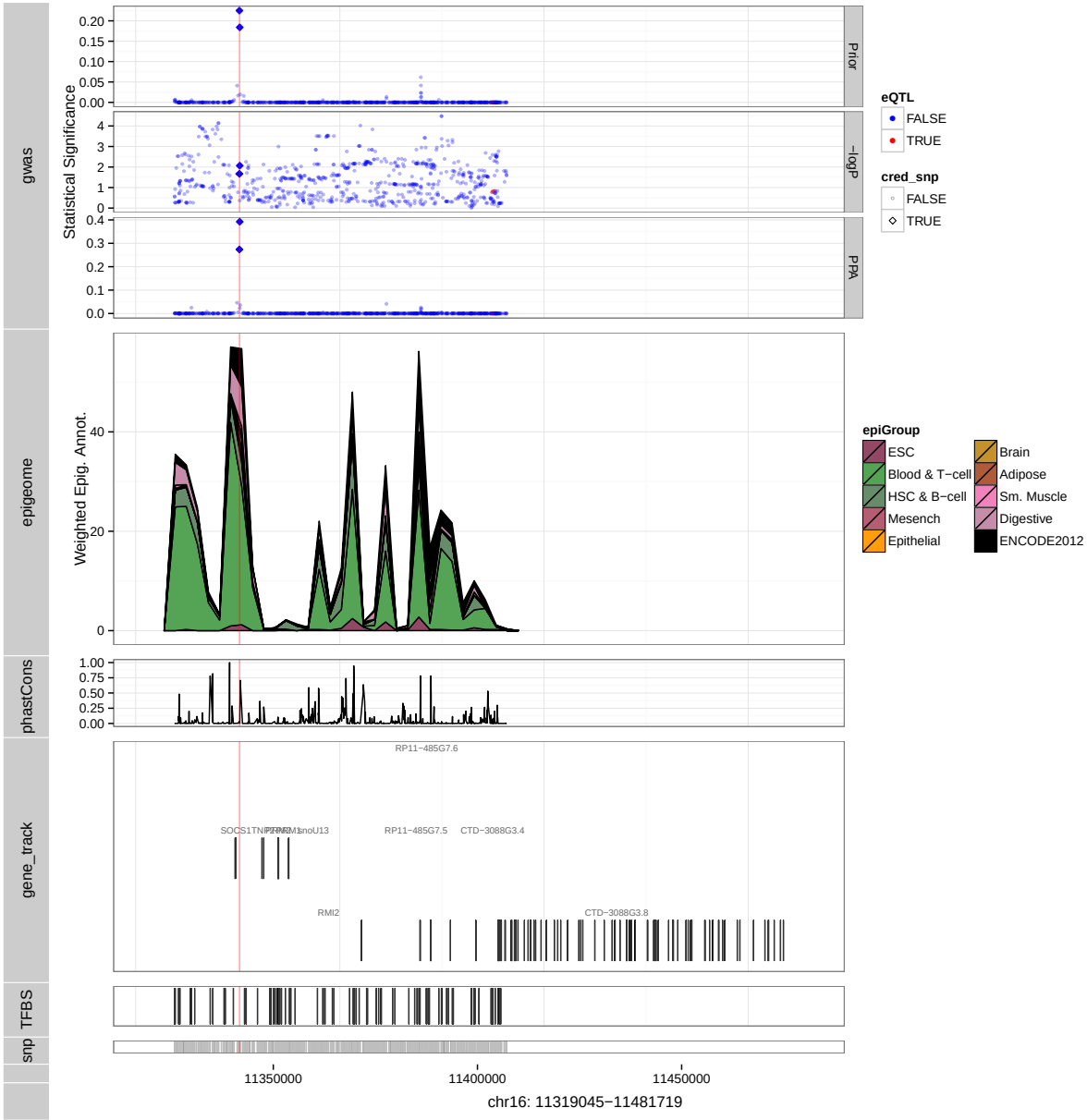
Juvenile Idiopathic Arthritis



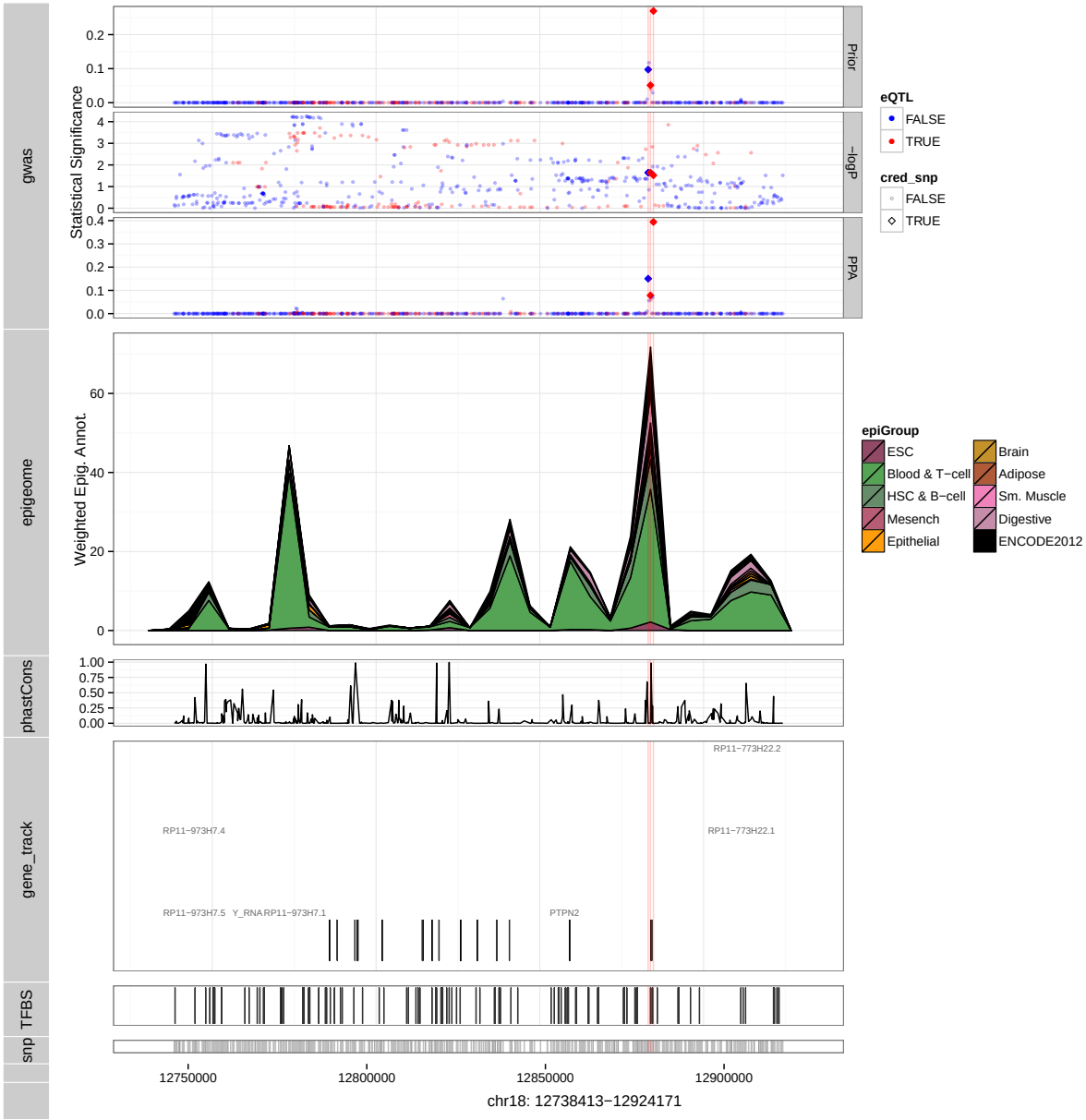
Juvenile Idiopathic Arthritis



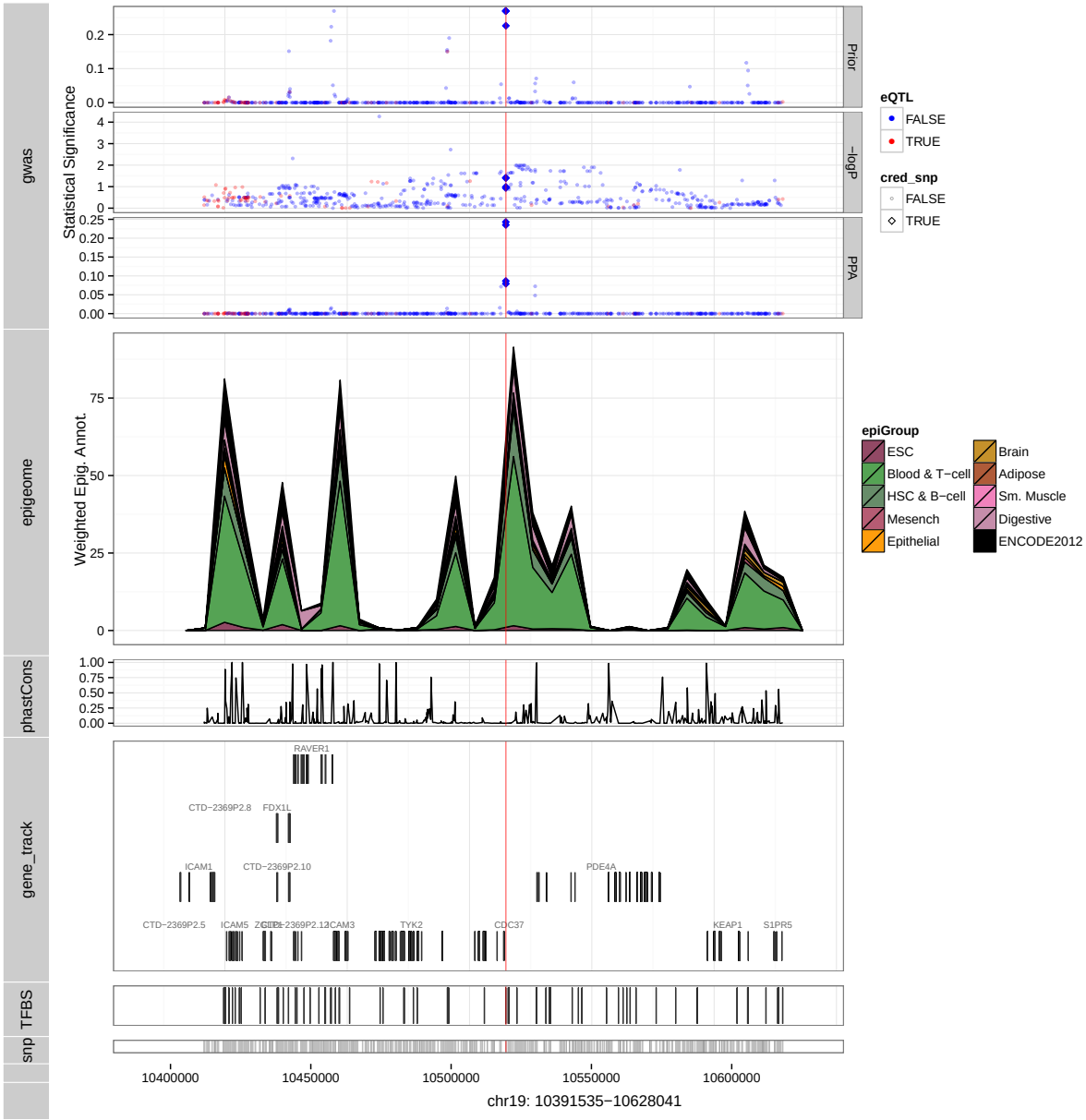
Juvenile Idiopathic Arthritis



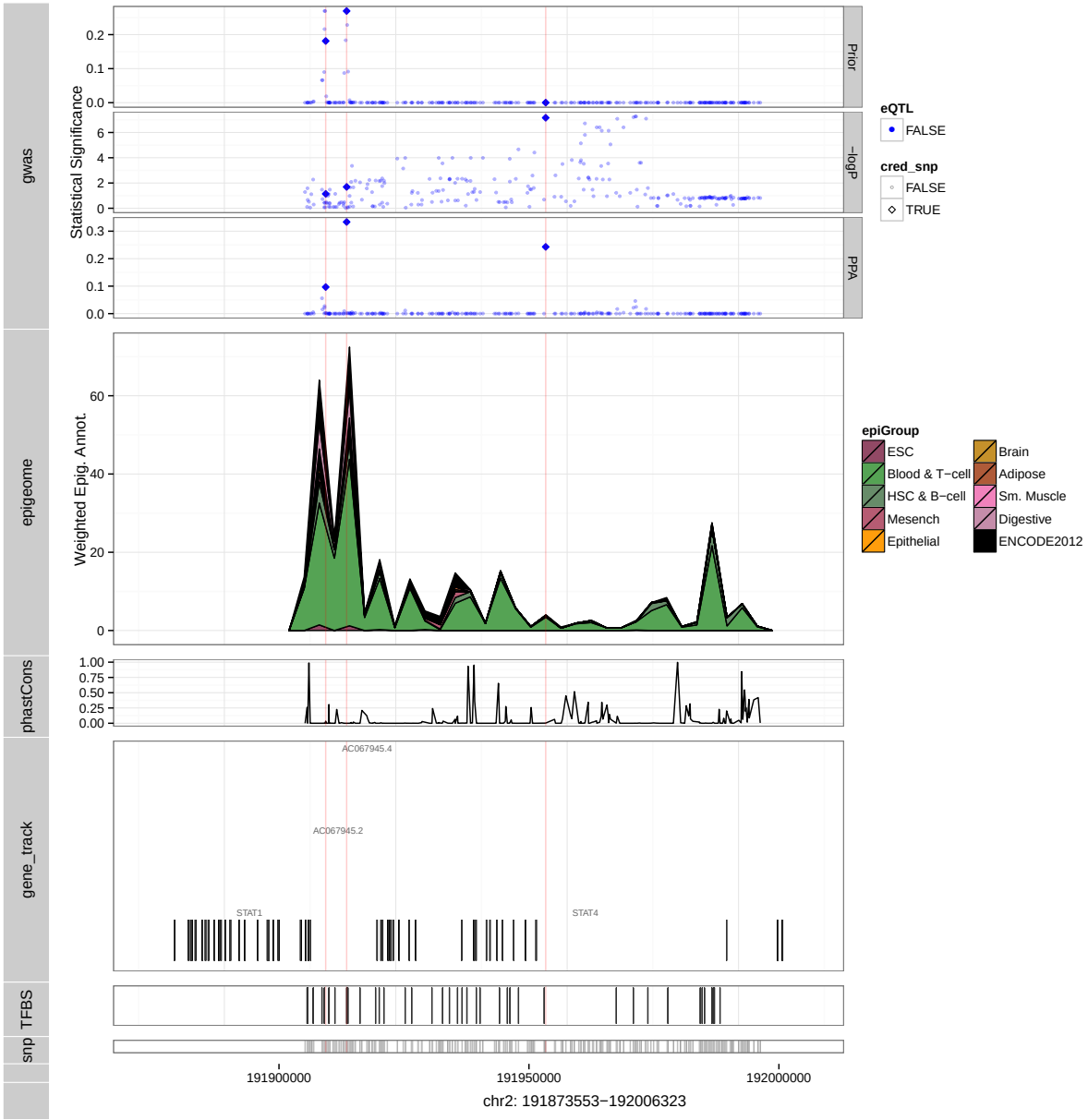
Juvenile Idiopathic Arthritis



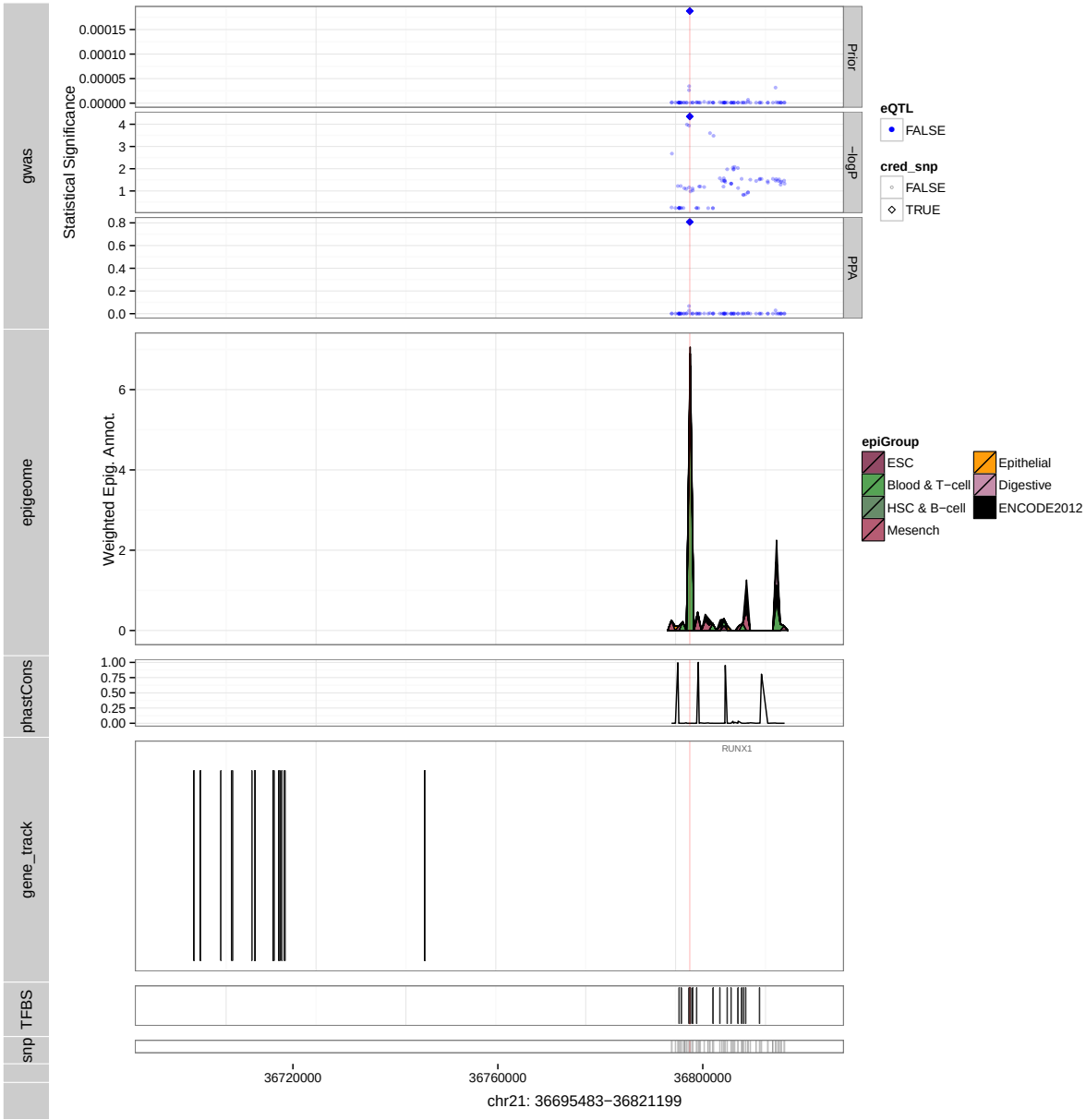
Juvenile Idiopathic Arthritis



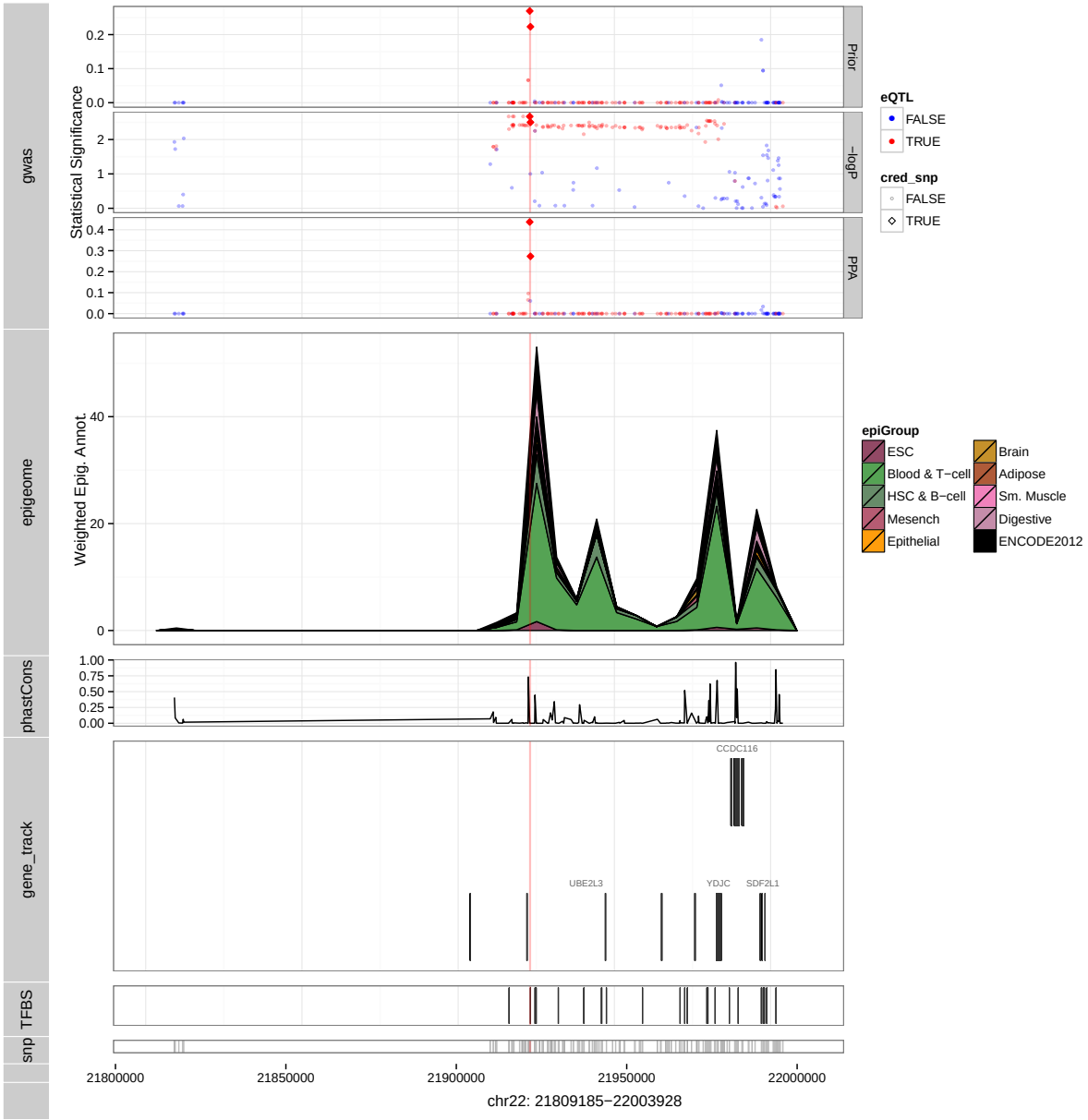
Juvenile Idiopathic Arthritis



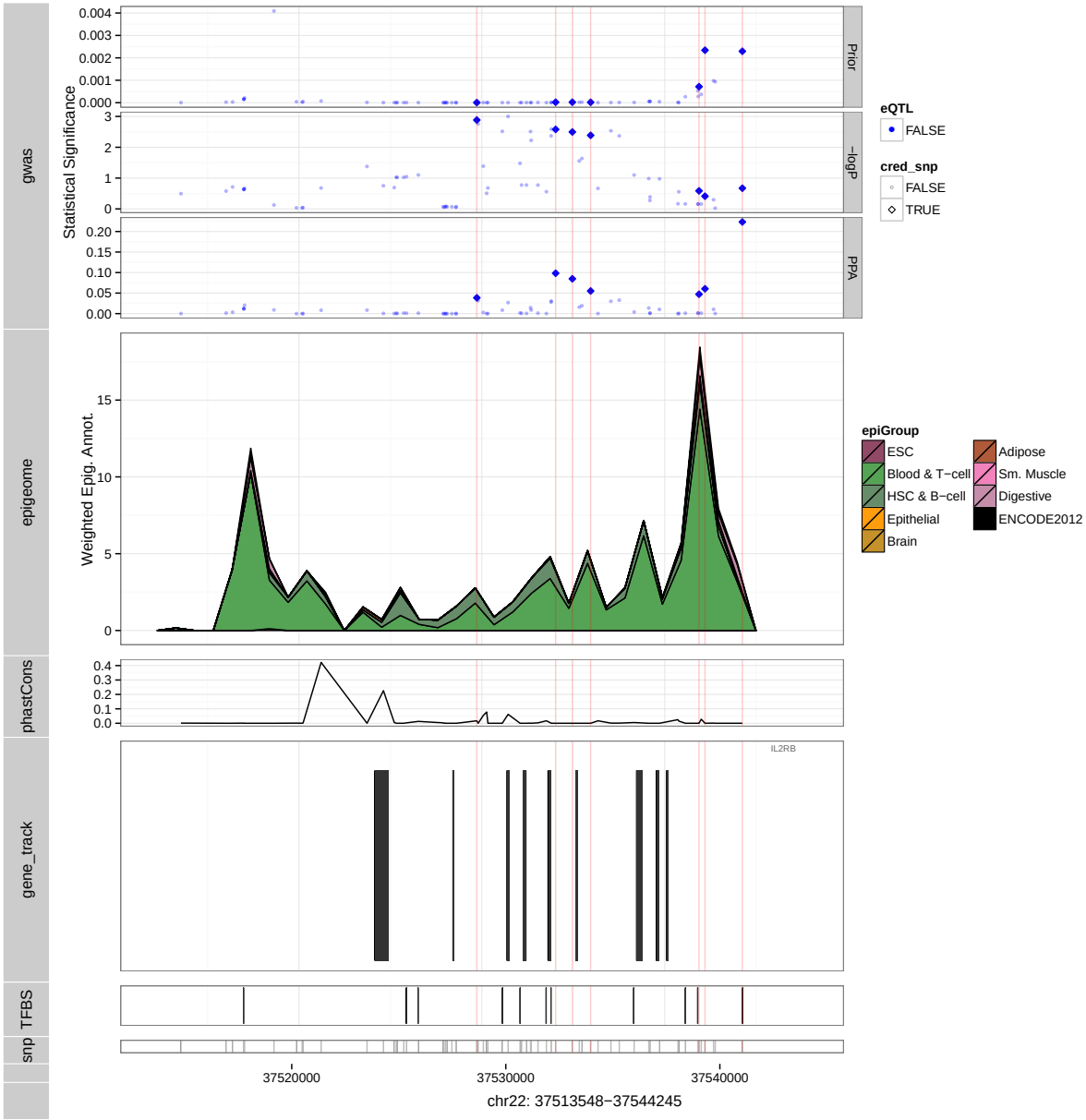
Juvenile Idiopathic Arthritis



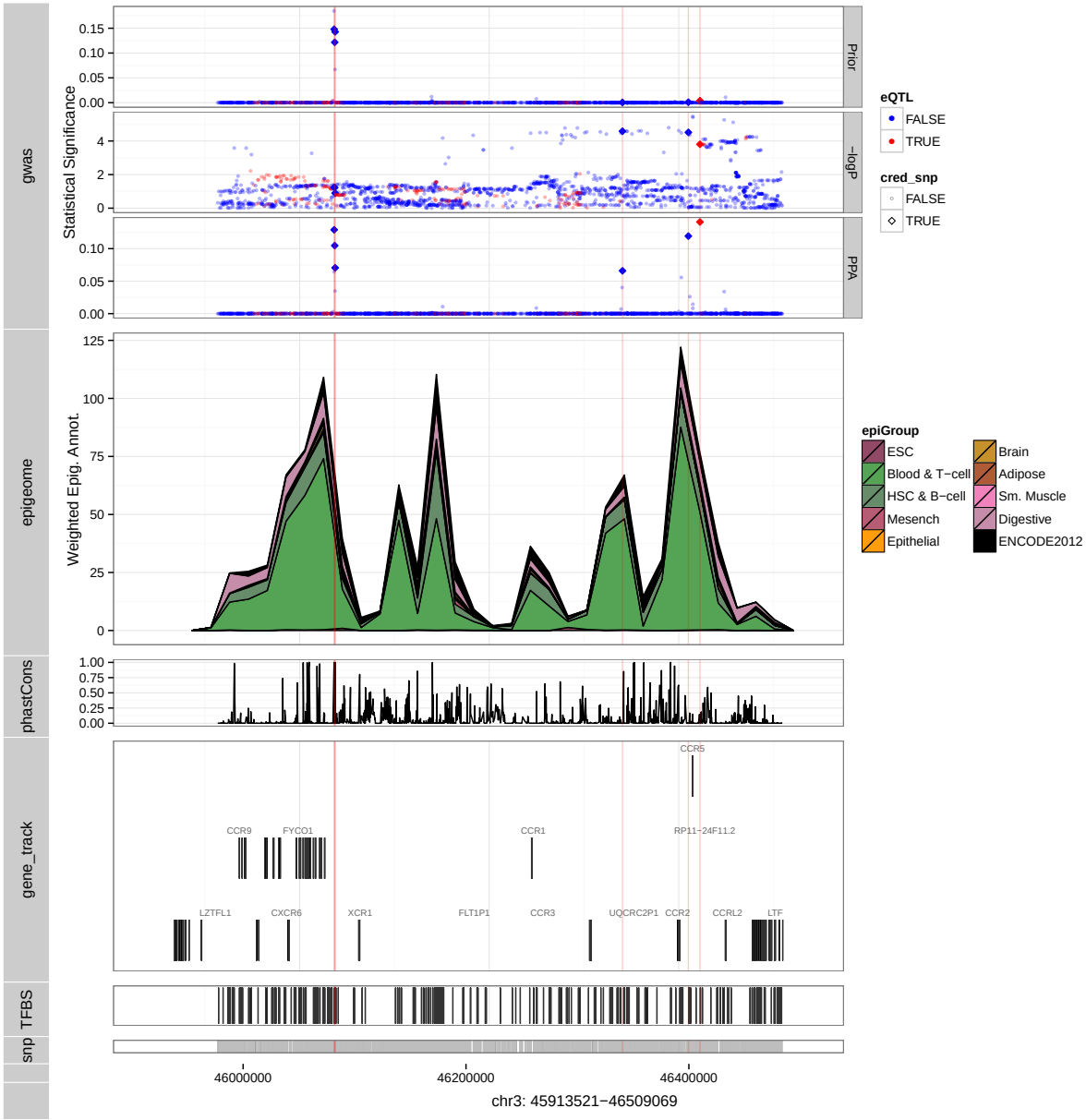
Juvenile Idiopathic Arthritis



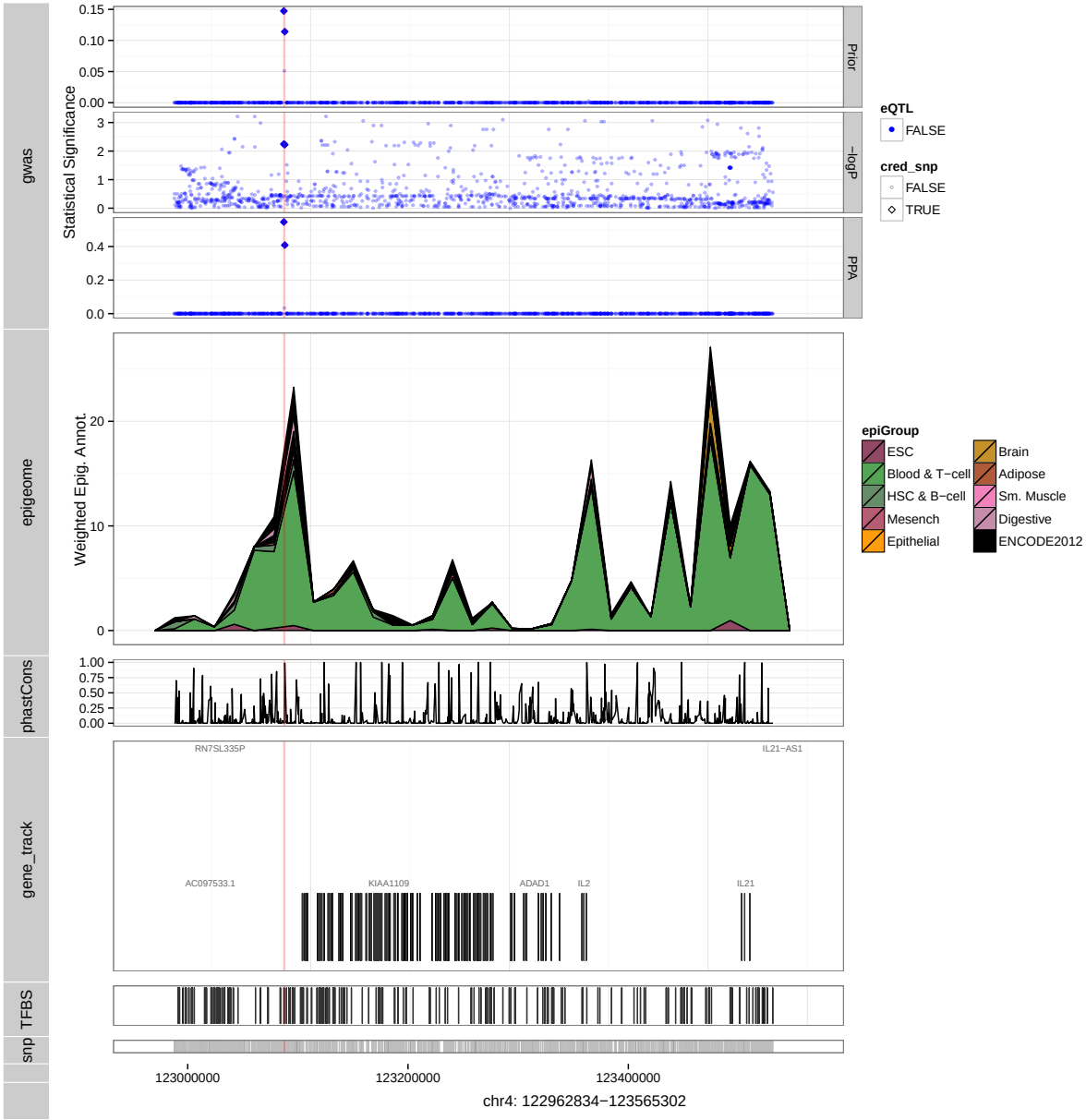
Juvenile Idiopathic Arthritis



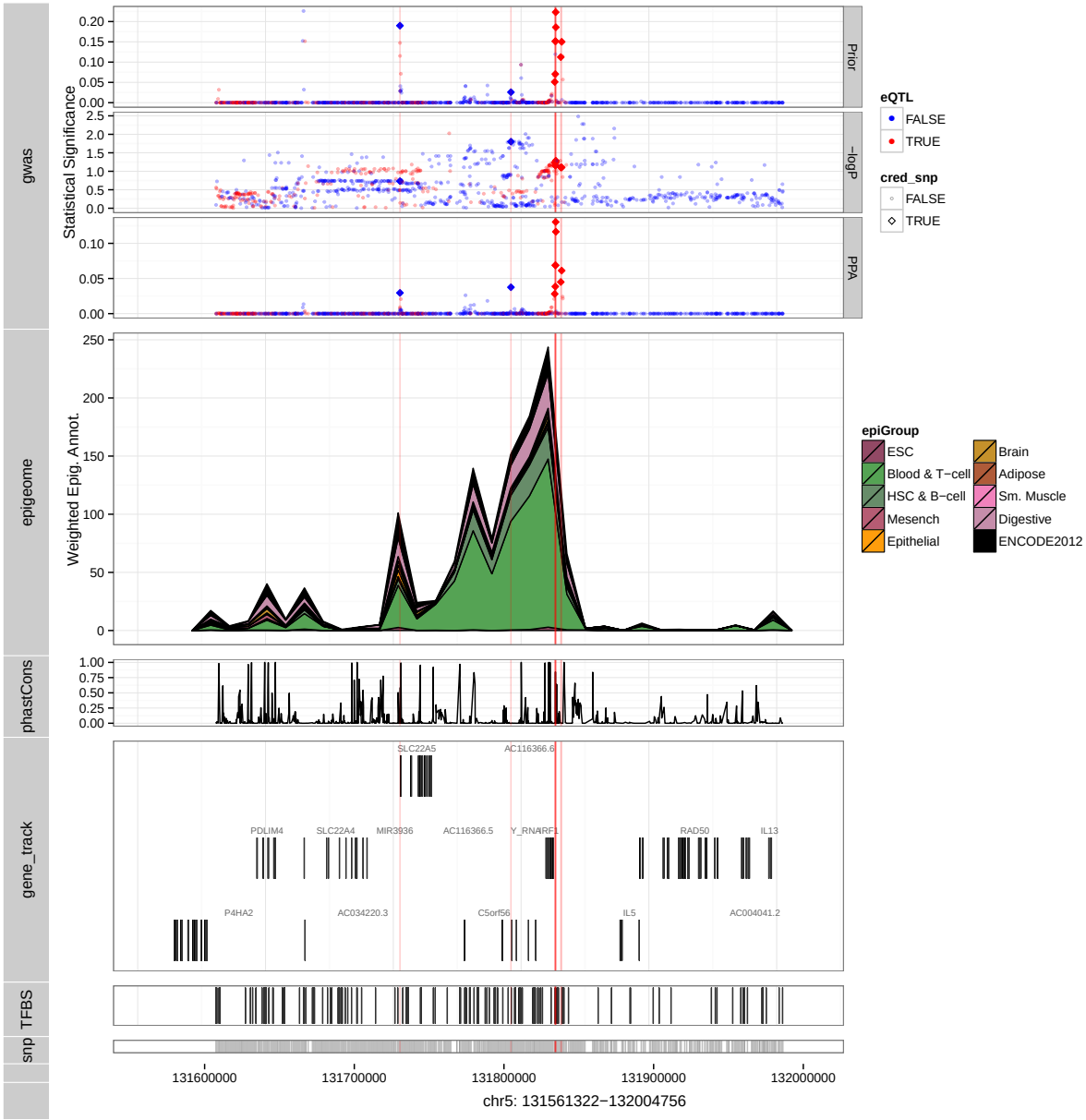
Juvenile Idiopathic Arthritis



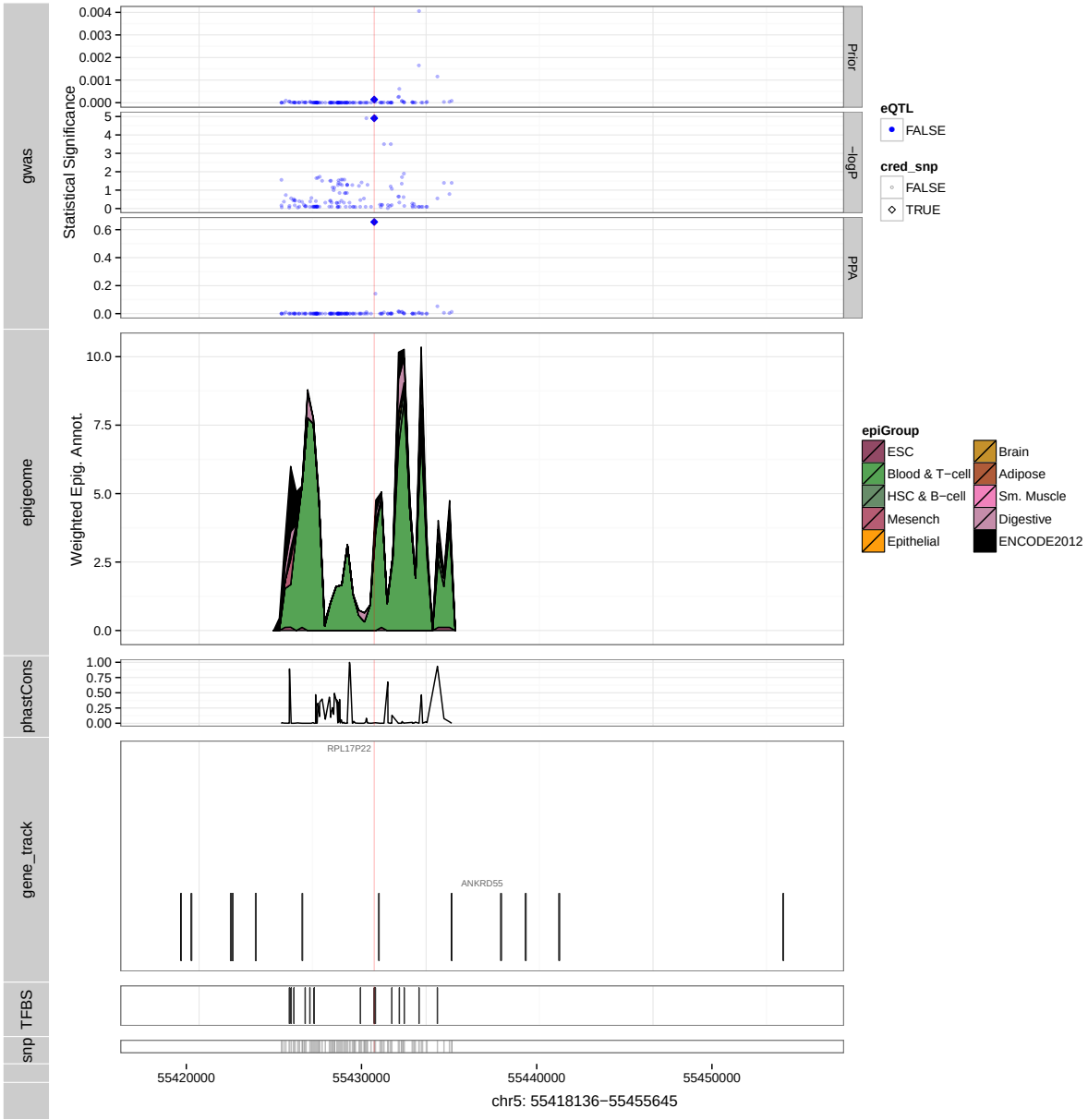
Juvenile Idiopathic Arthritis



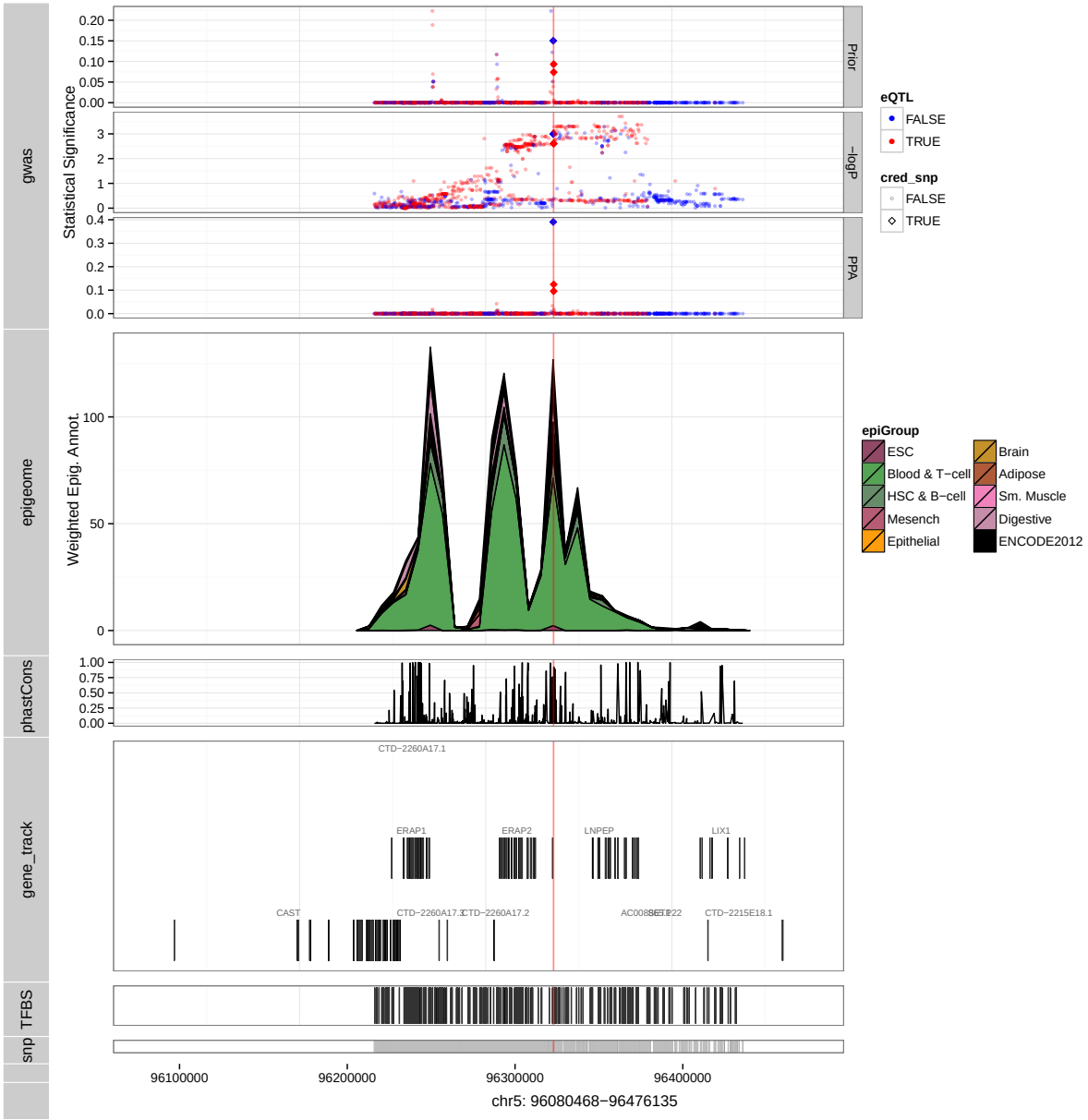
Juvenile Idiopathic Arthritis



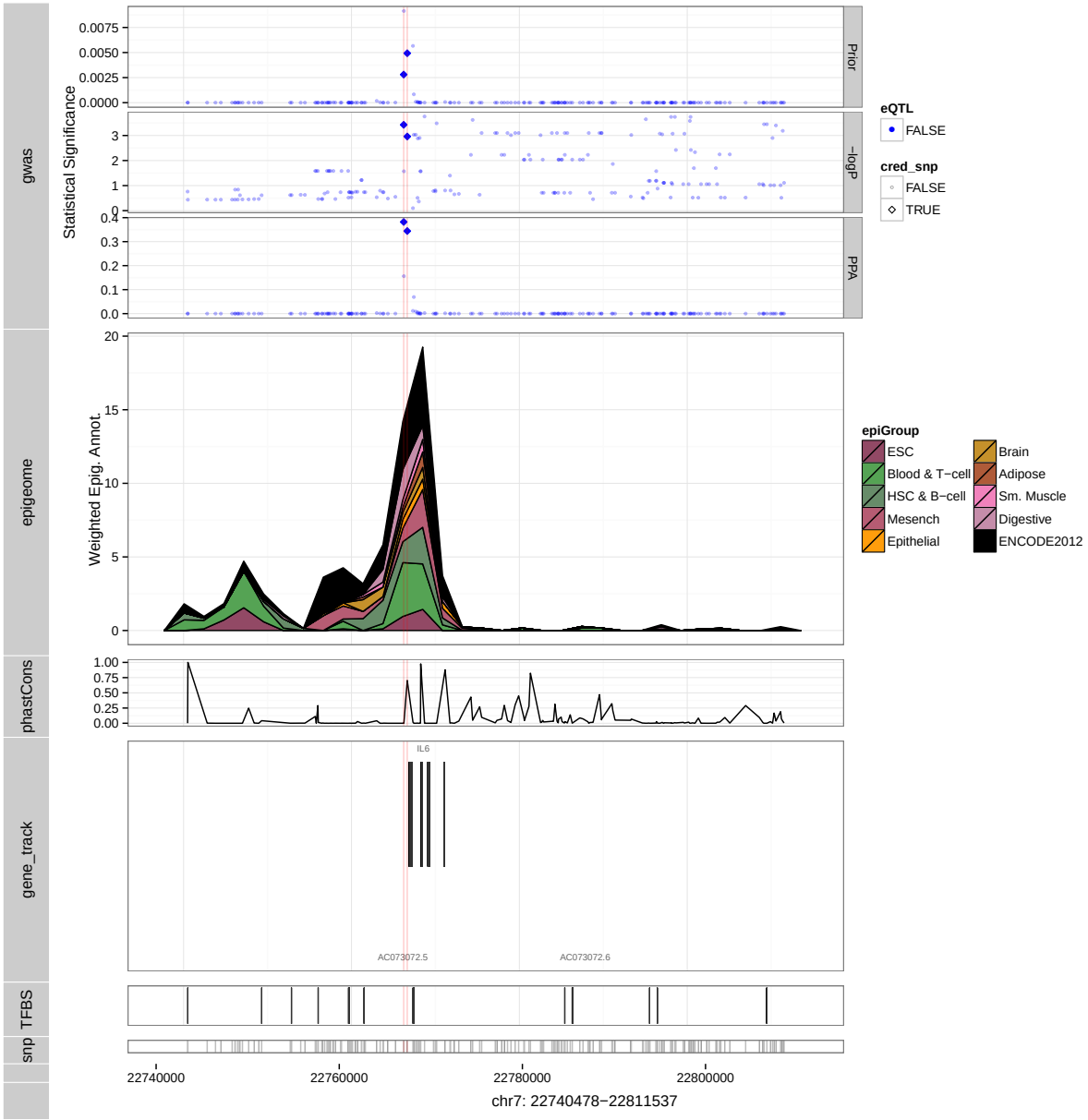
Juvenile Idiopathic Arthritis



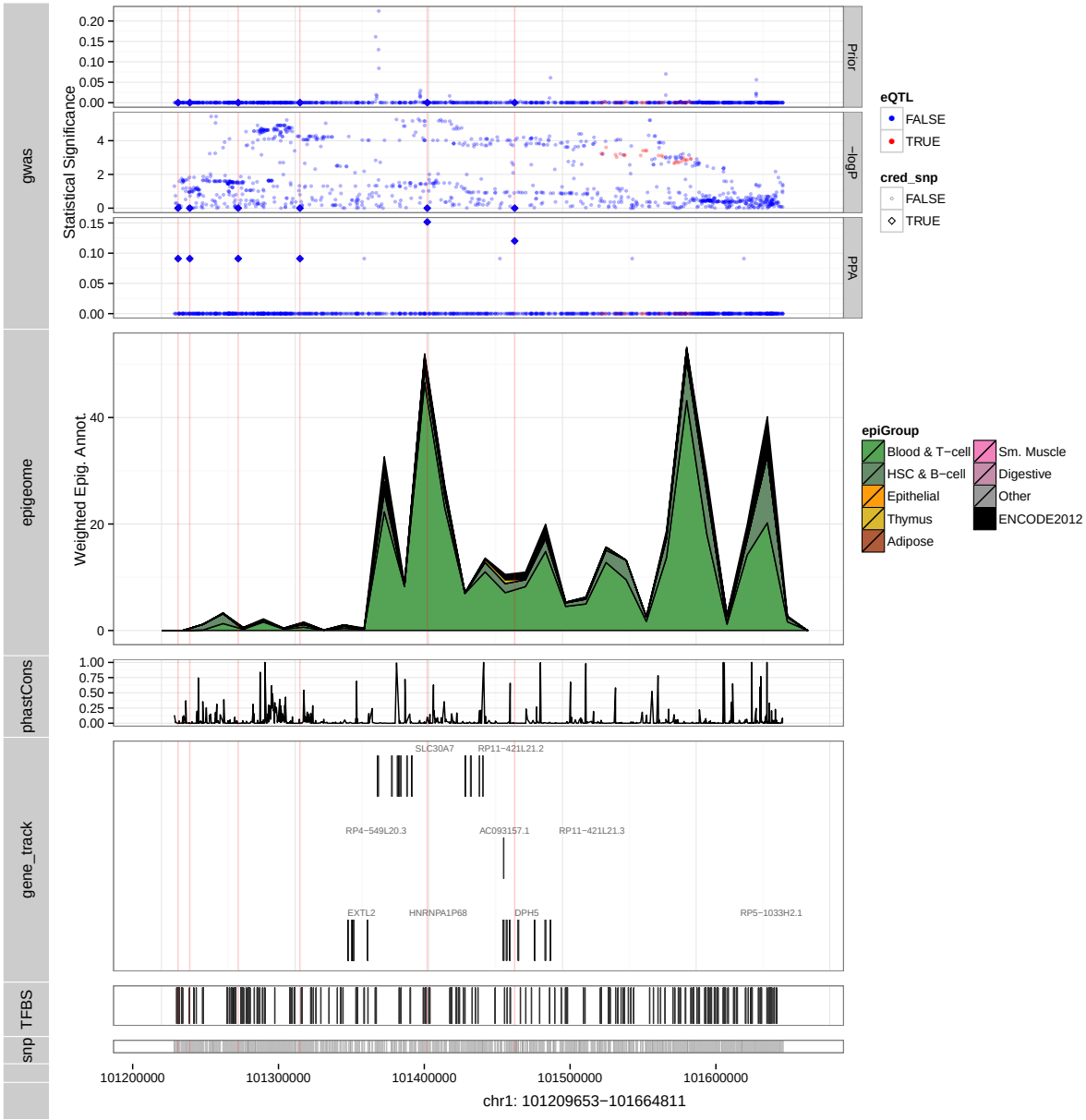
Juvenile Idiopathic Arthritis



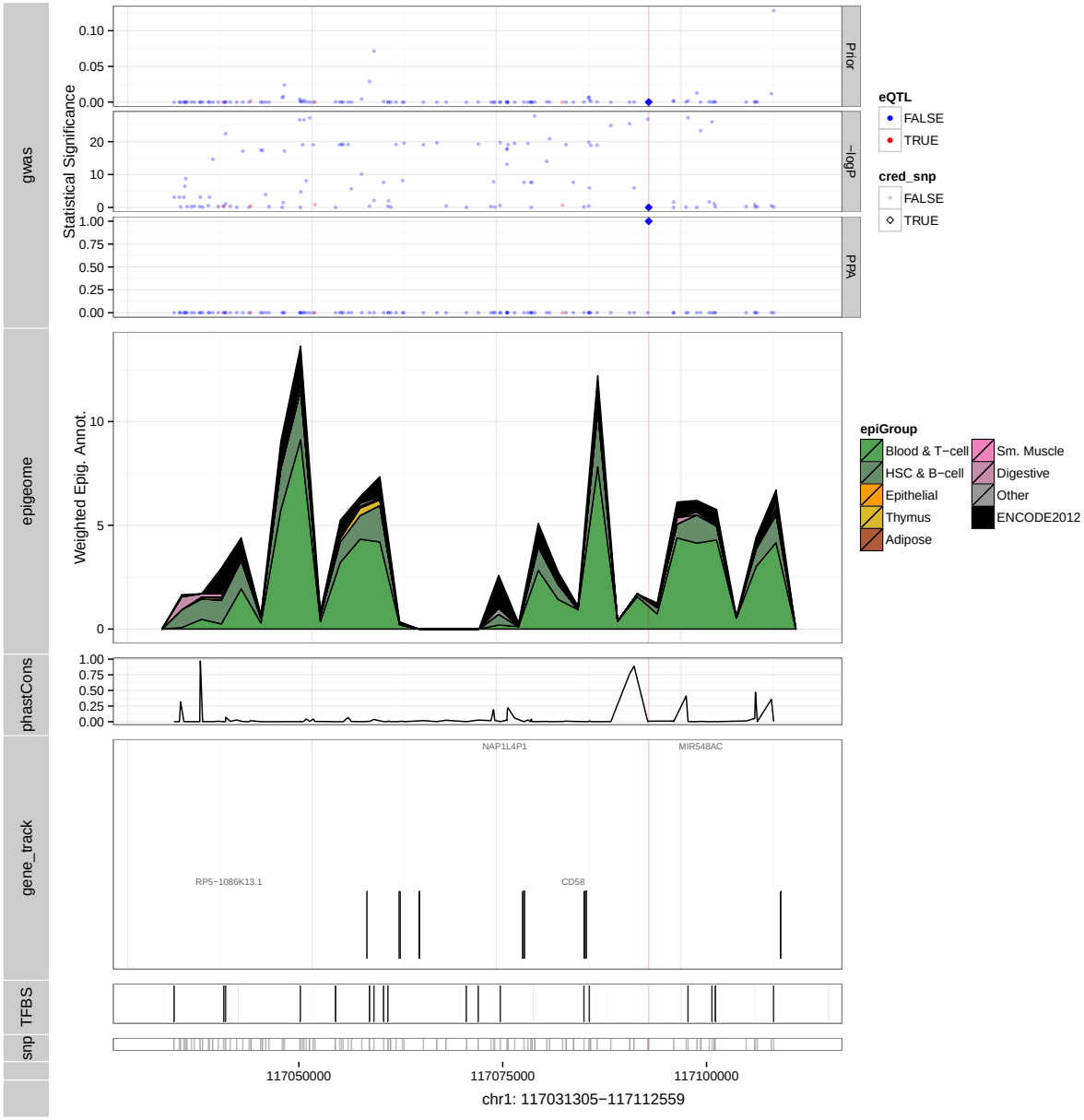
Juvenile Idiopathic Arthritis



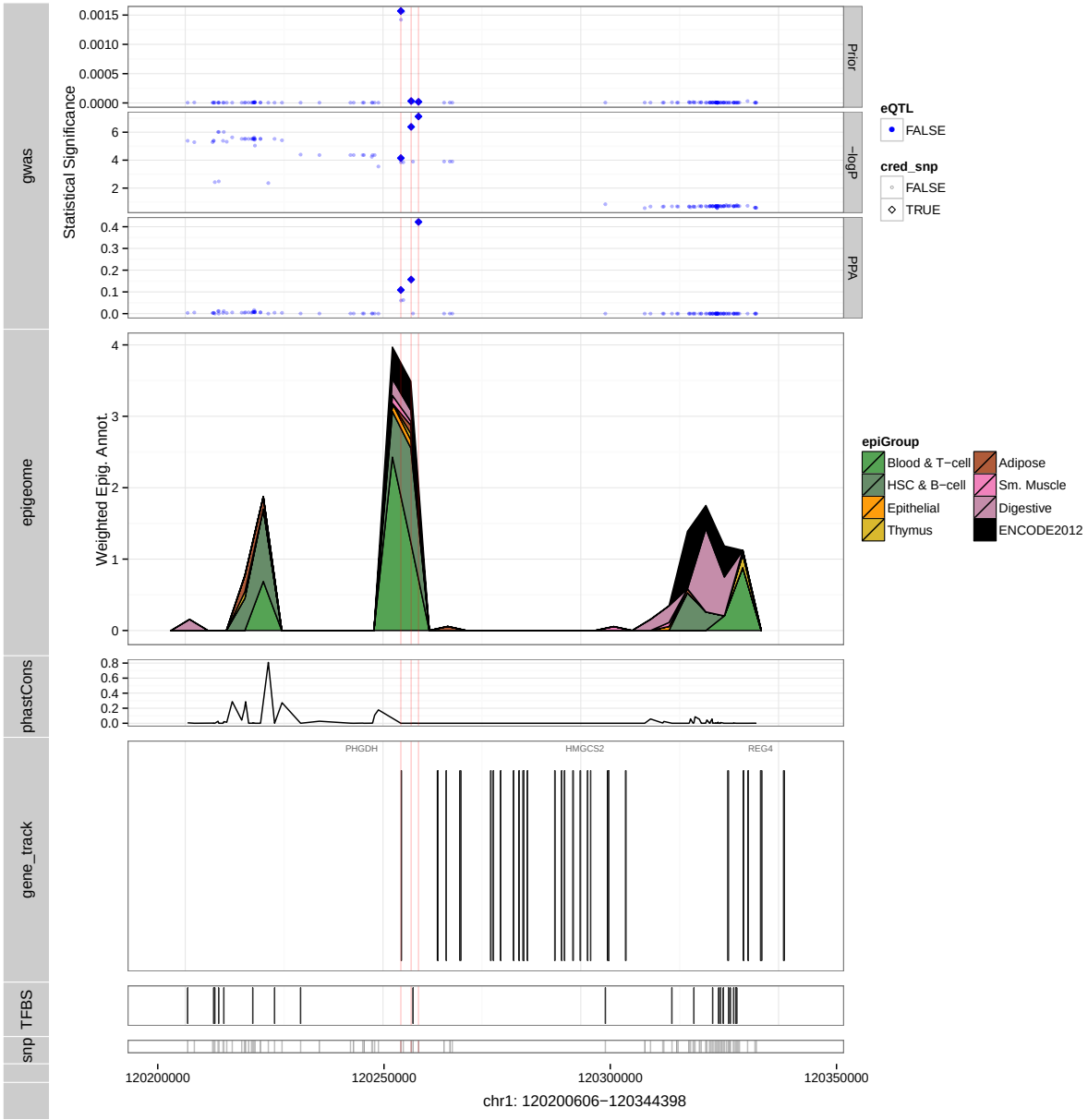
Multiple Sclerosis



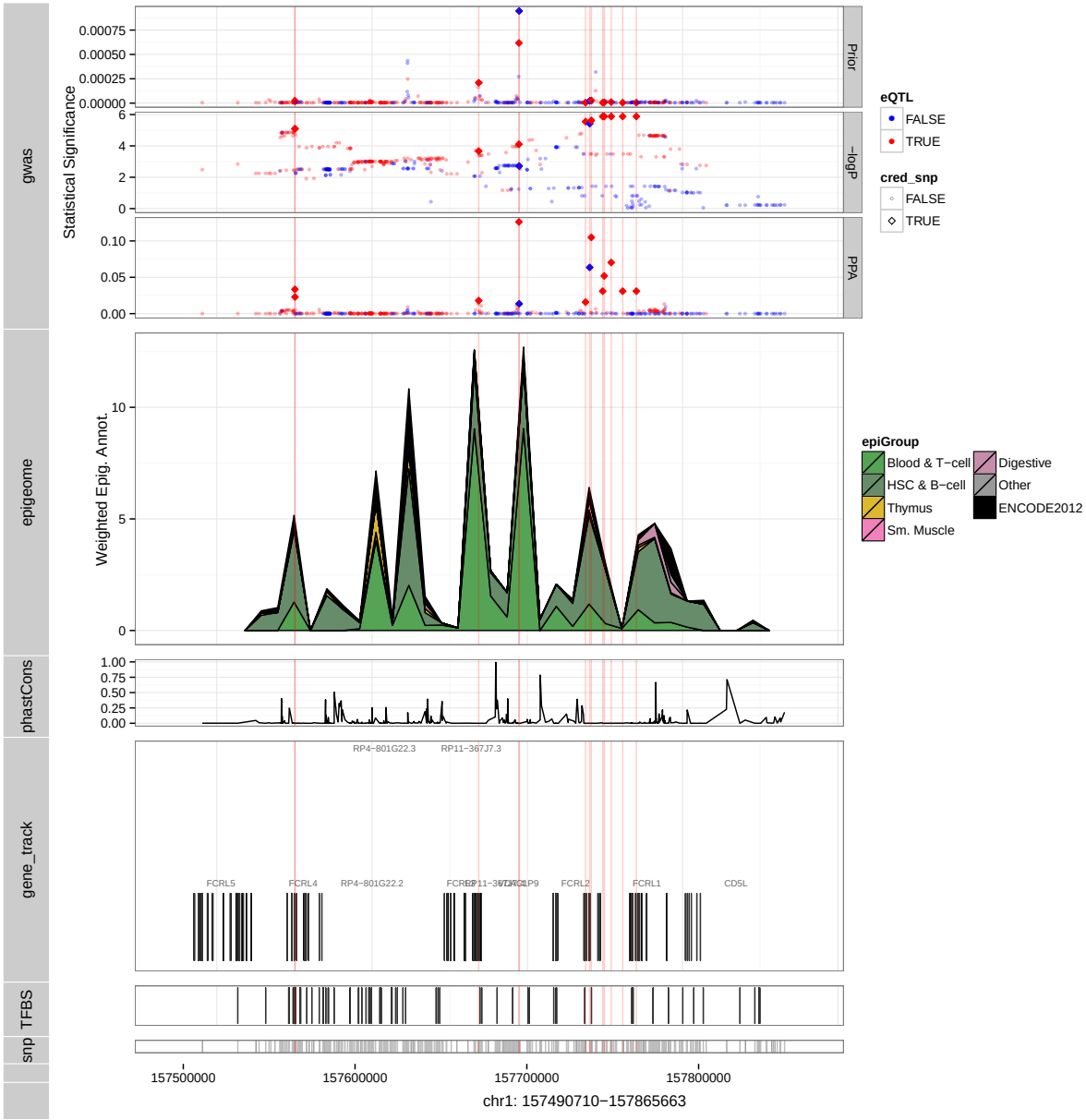
Multiple Sclerosis



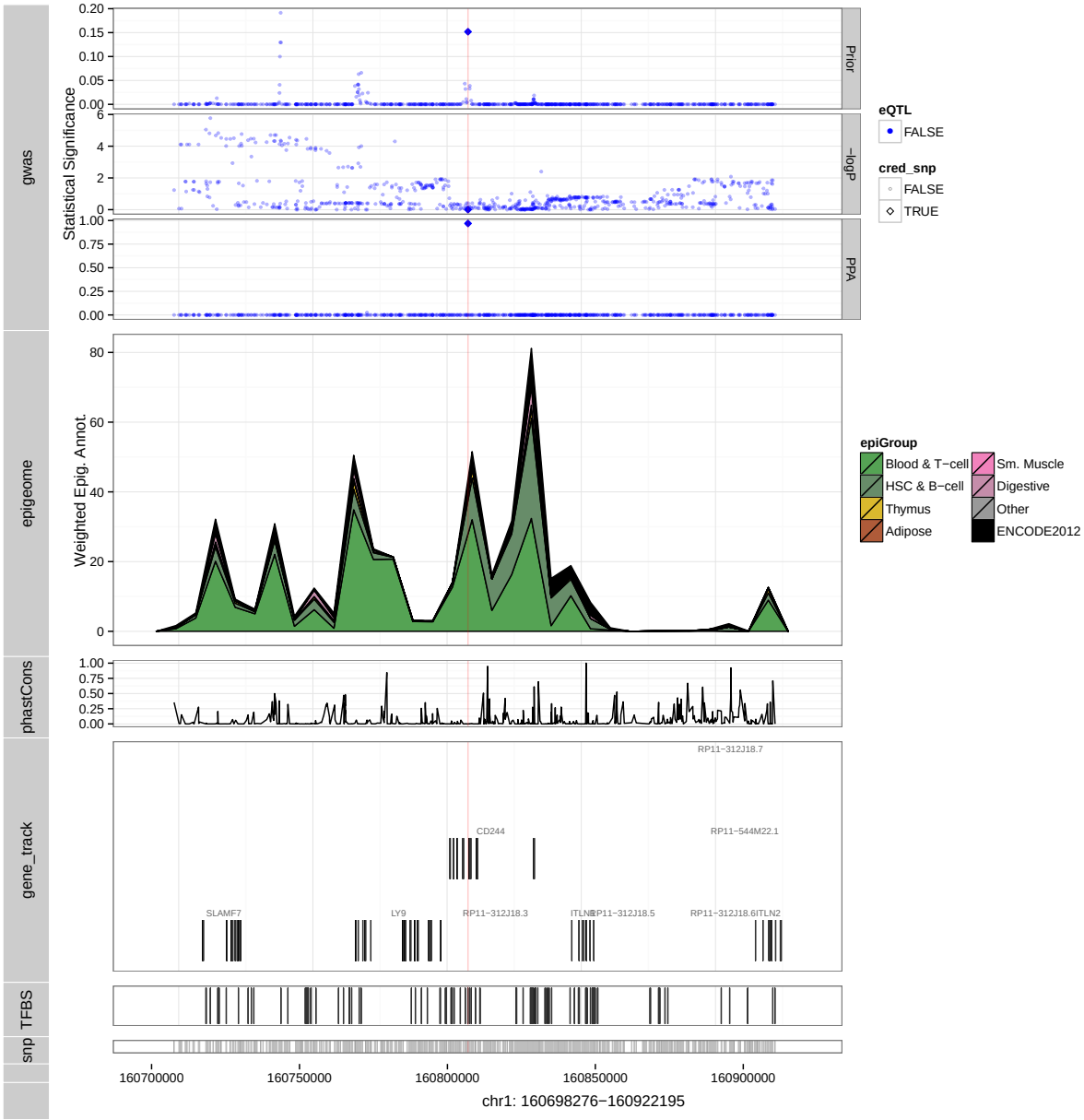
Multiple Sclerosis



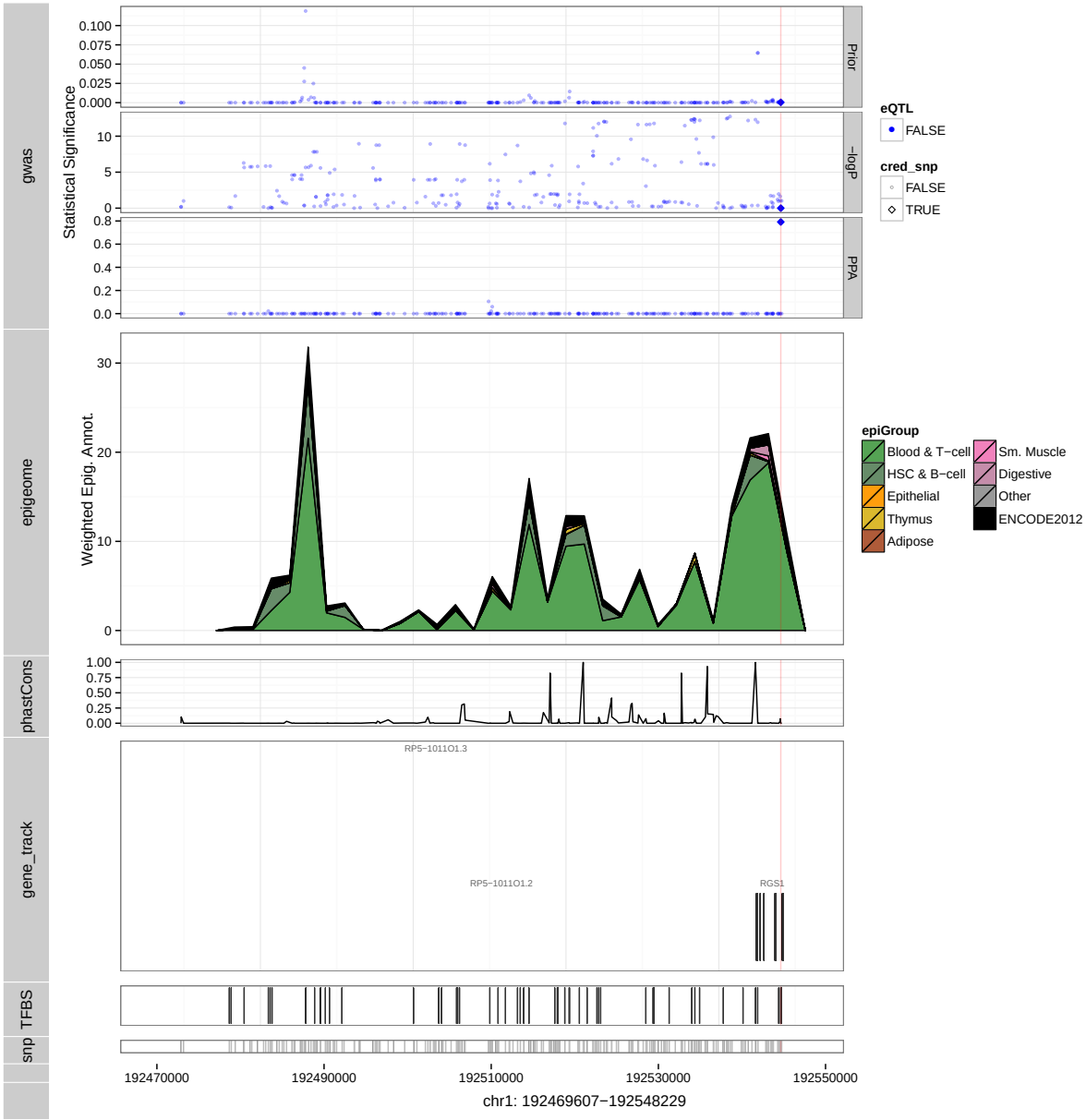
Multiple Sclerosis



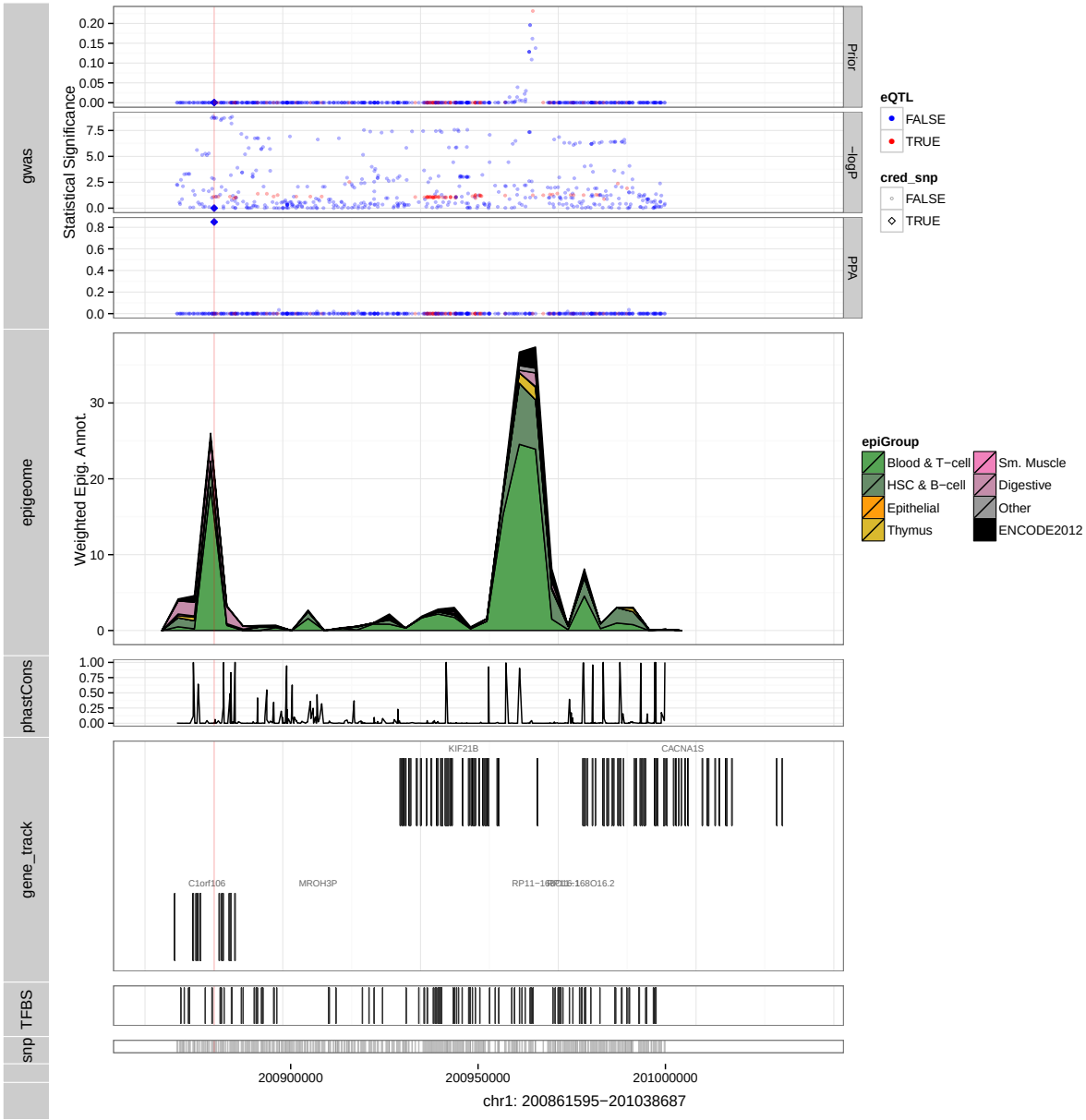
Multiple Sclerosis



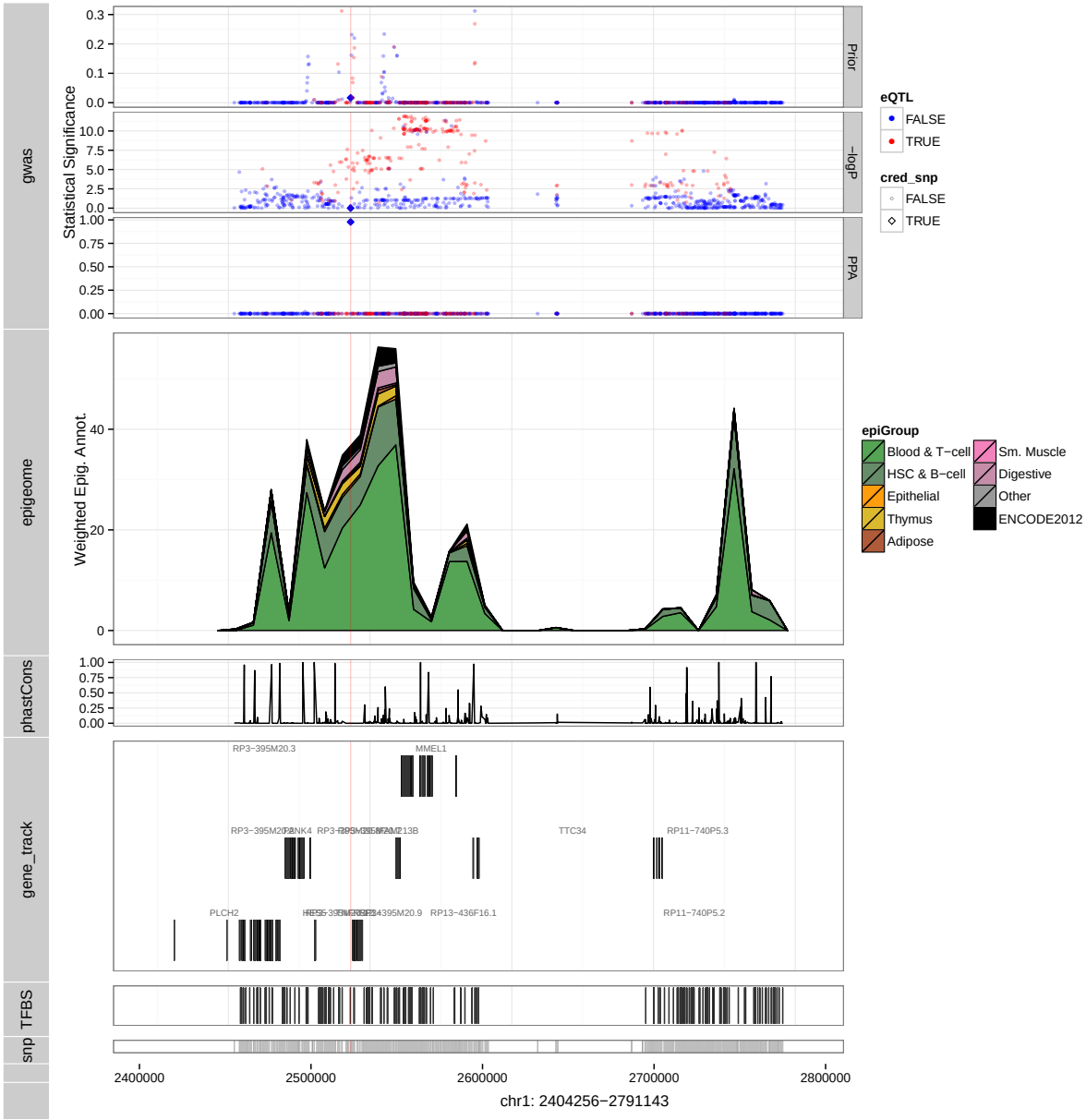
Multiple Sclerosis



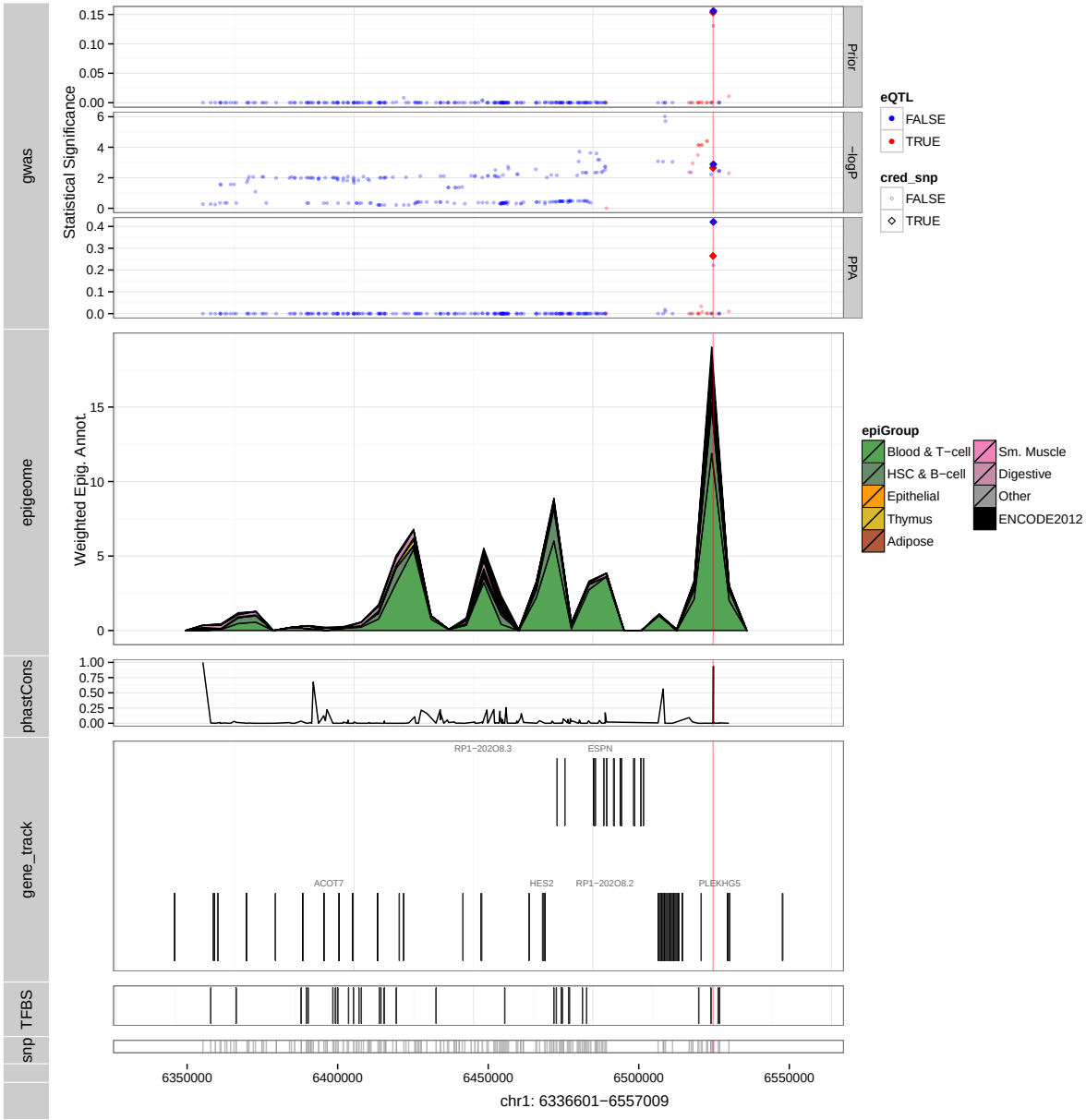
Multiple Sclerosis



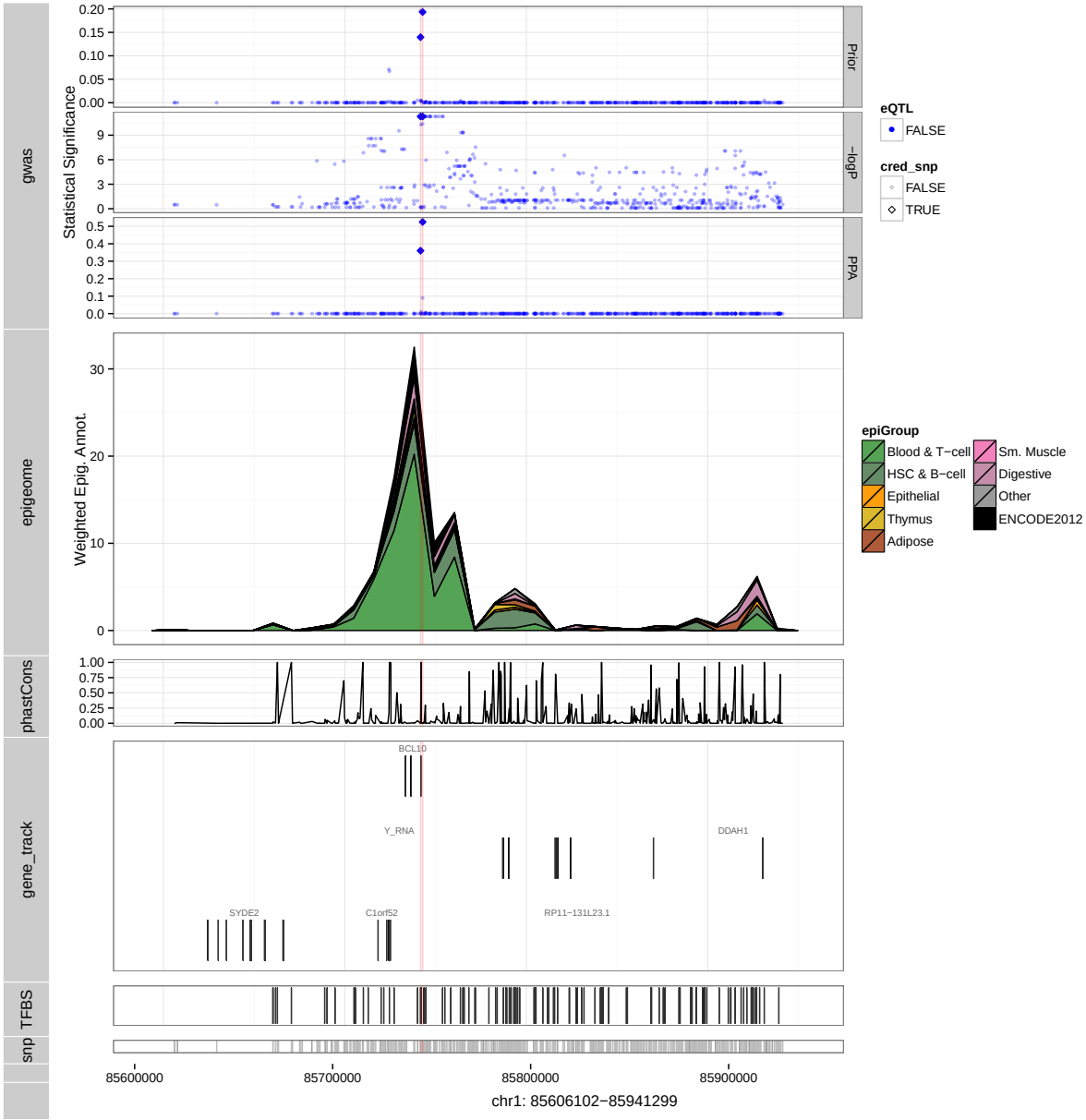
Multiple Sclerosis



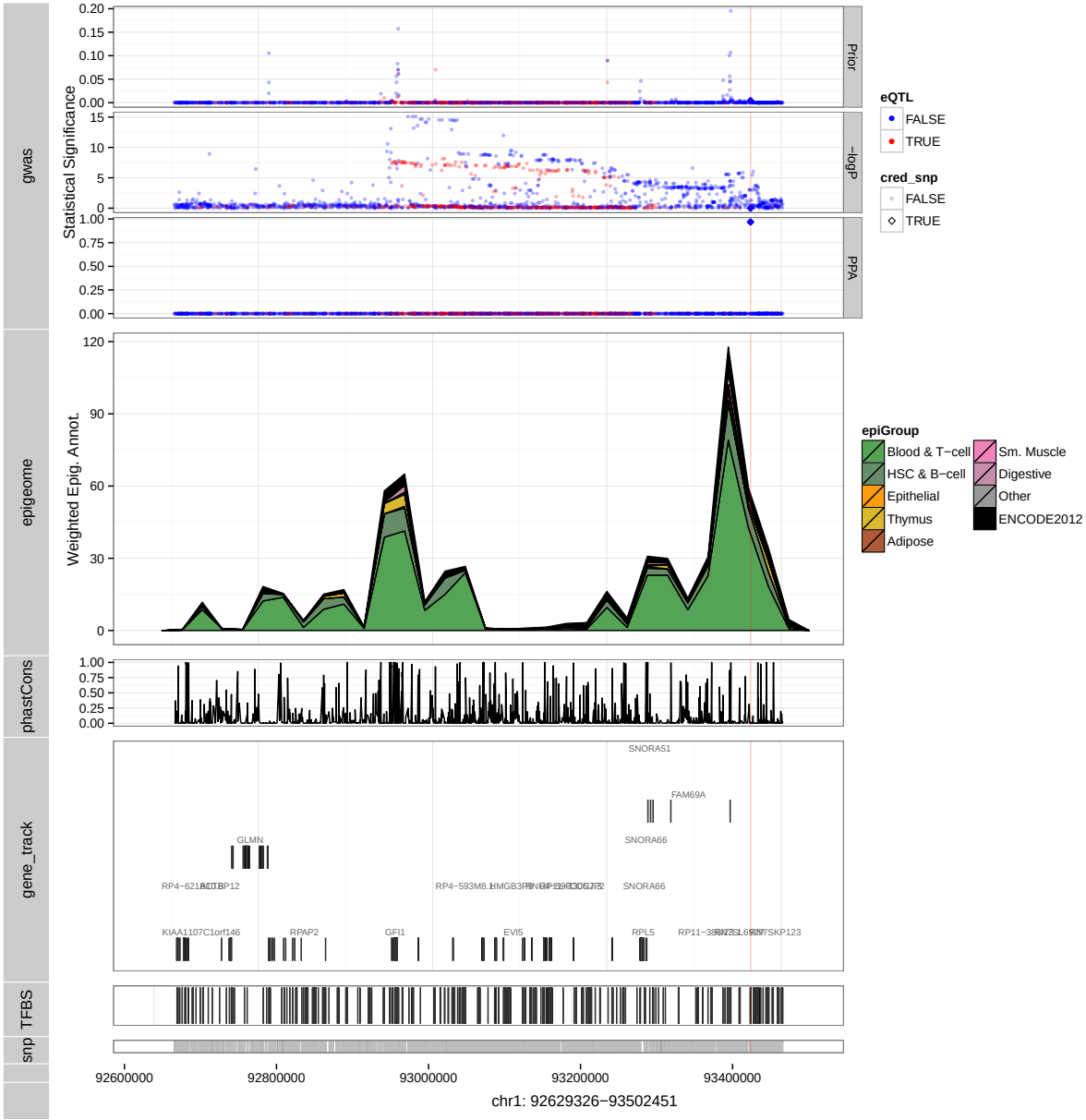
Multiple Sclerosis



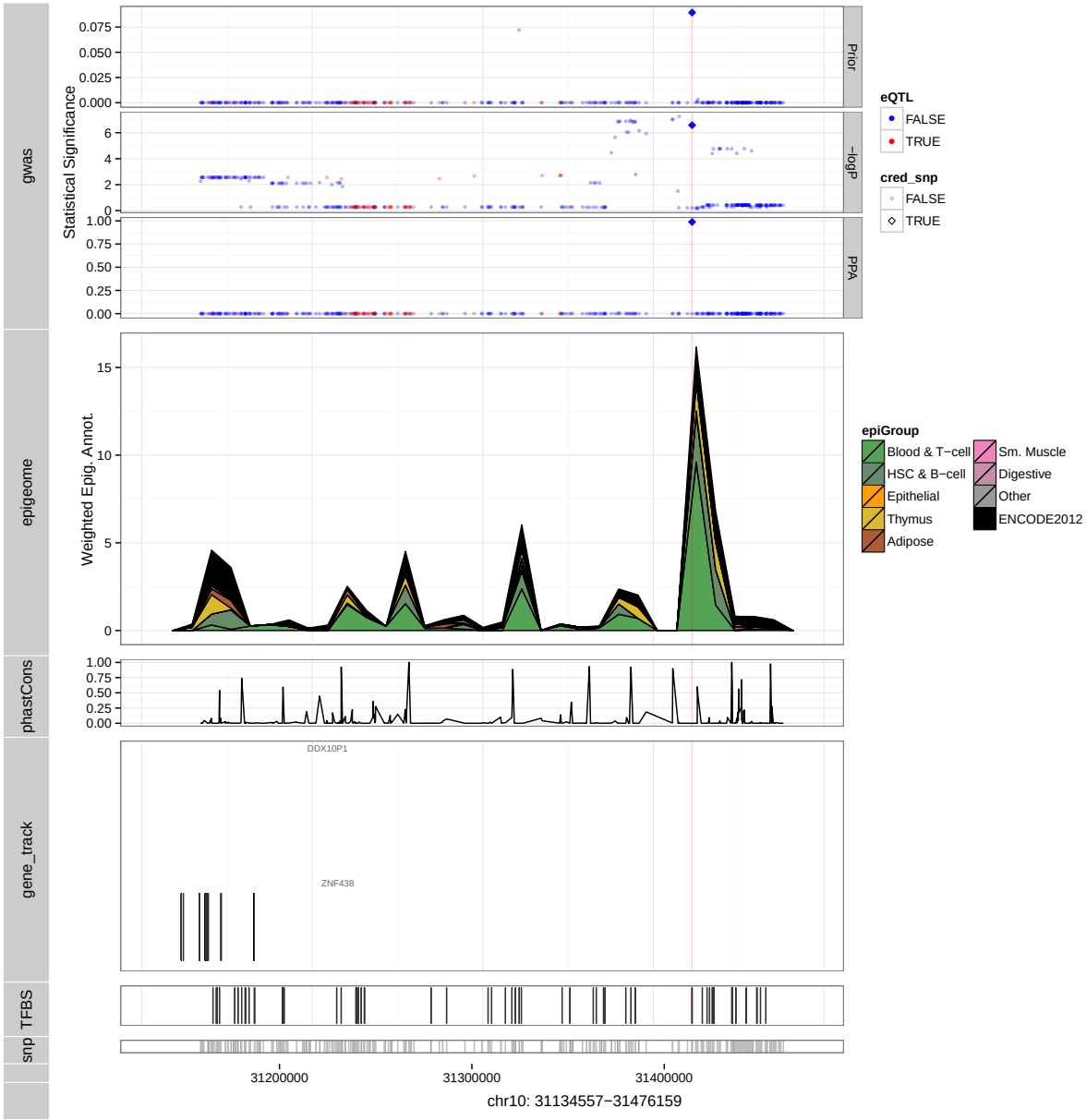
Multiple Sclerosis



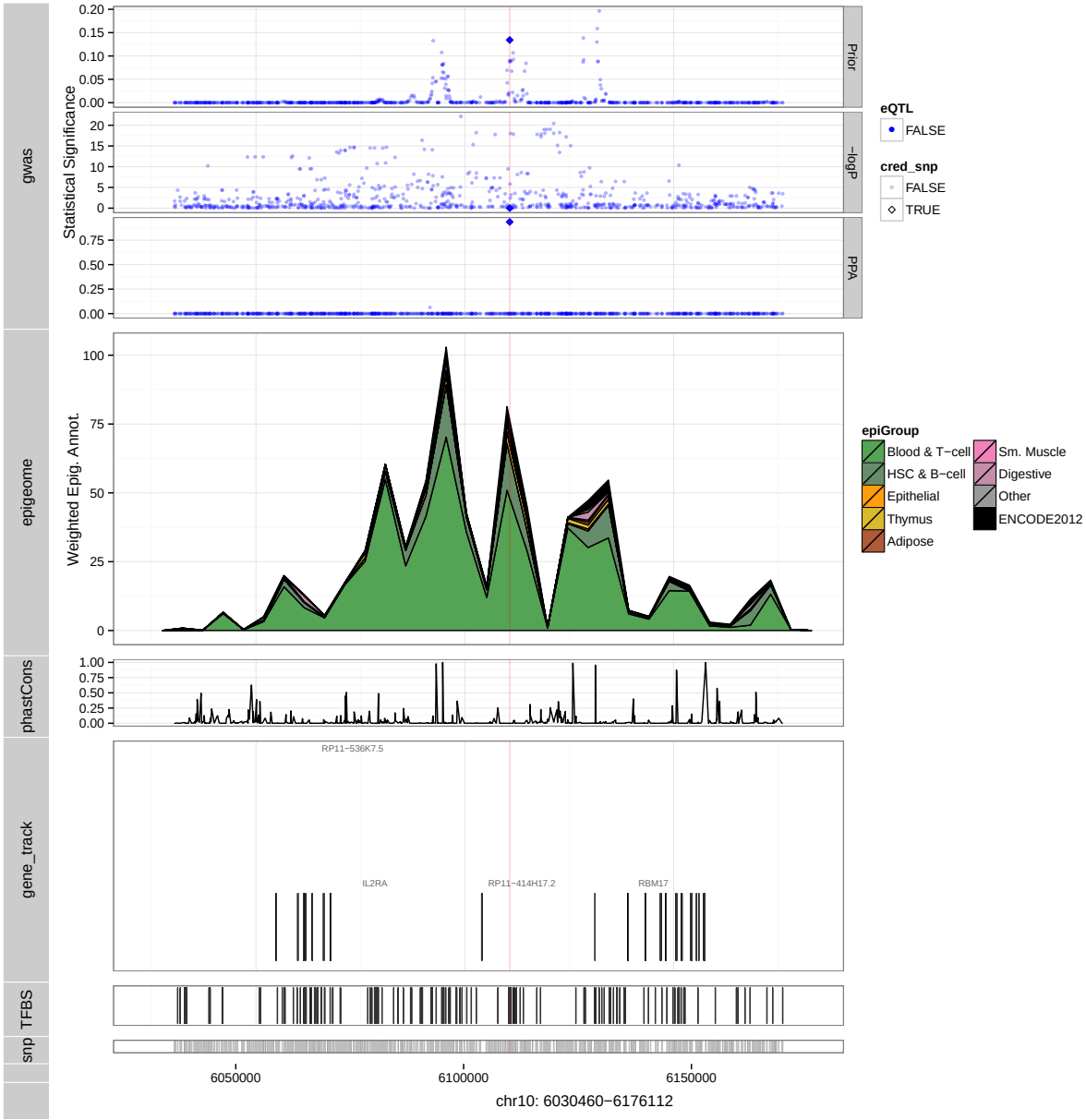
Multiple Sclerosis



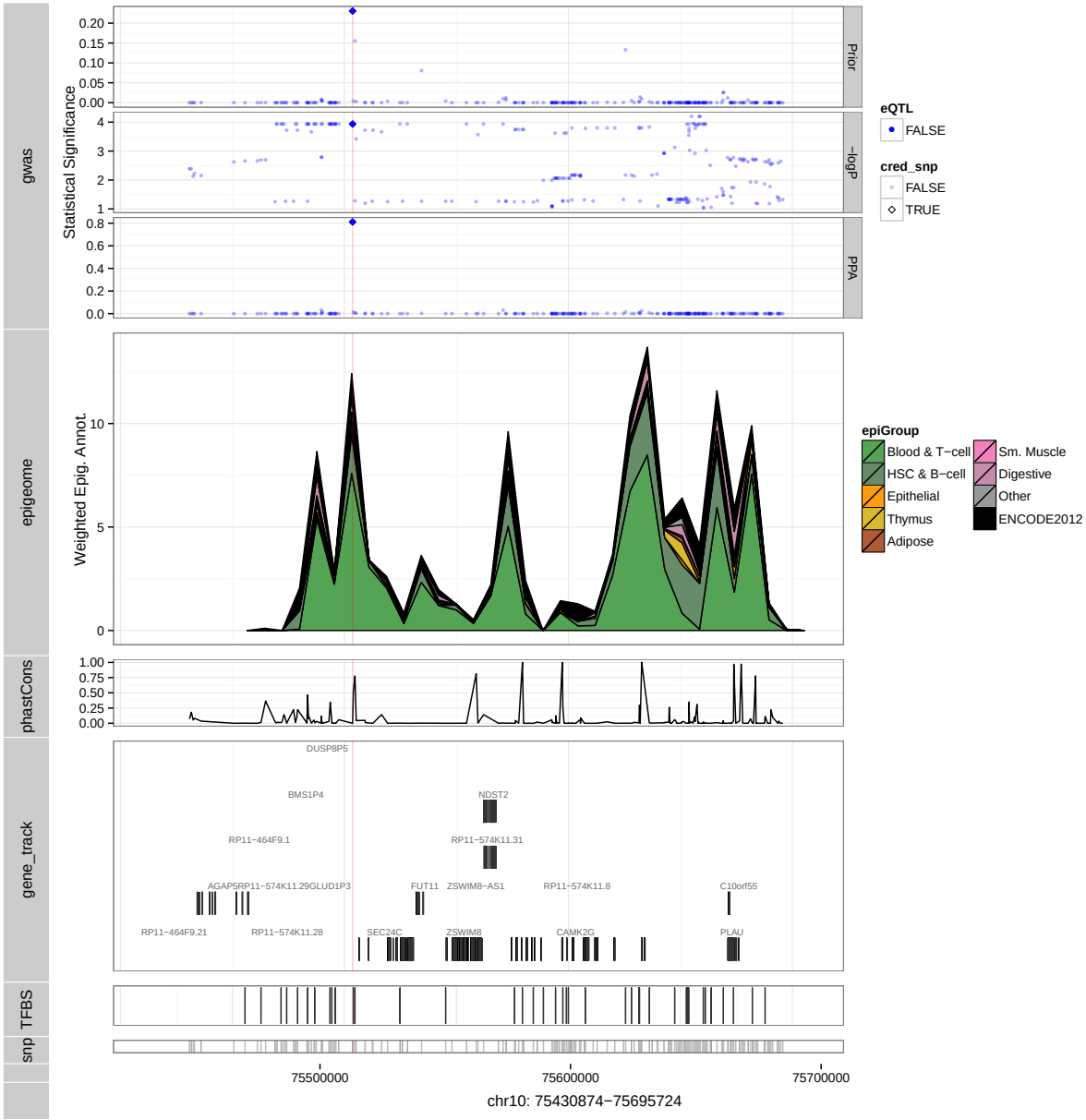
Multiple Sclerosis



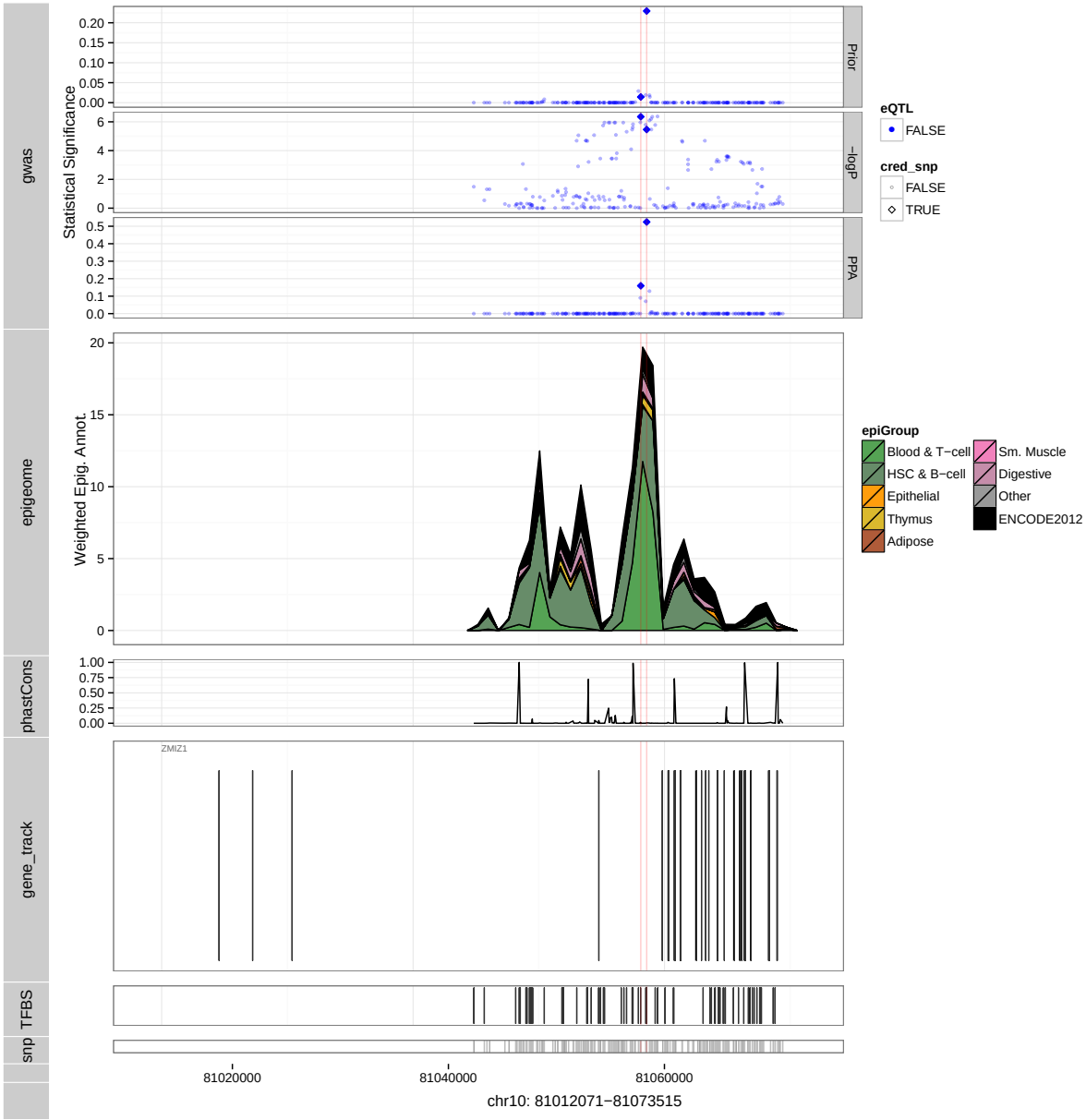
Multiple Sclerosis



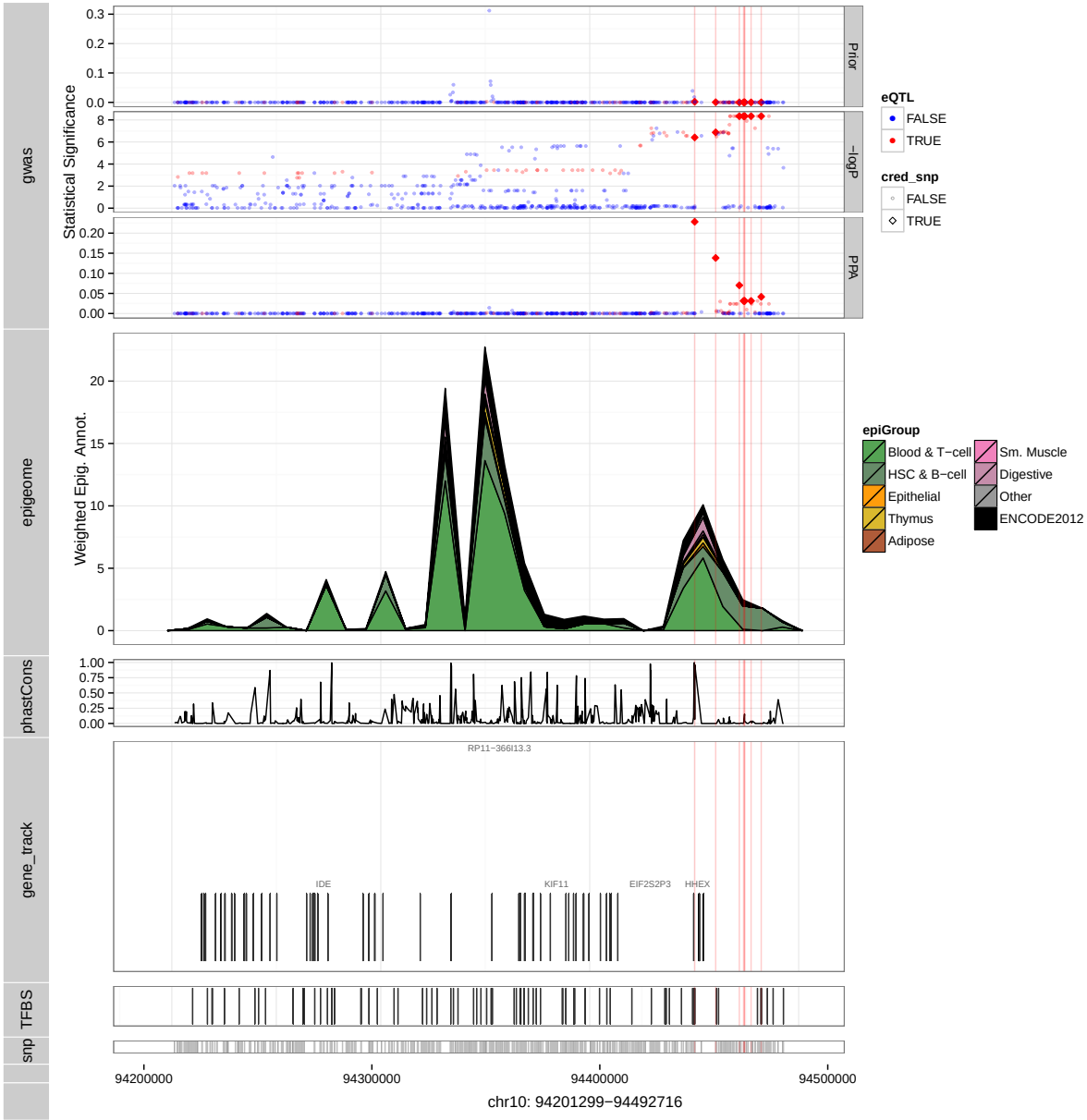
Multiple Sclerosis



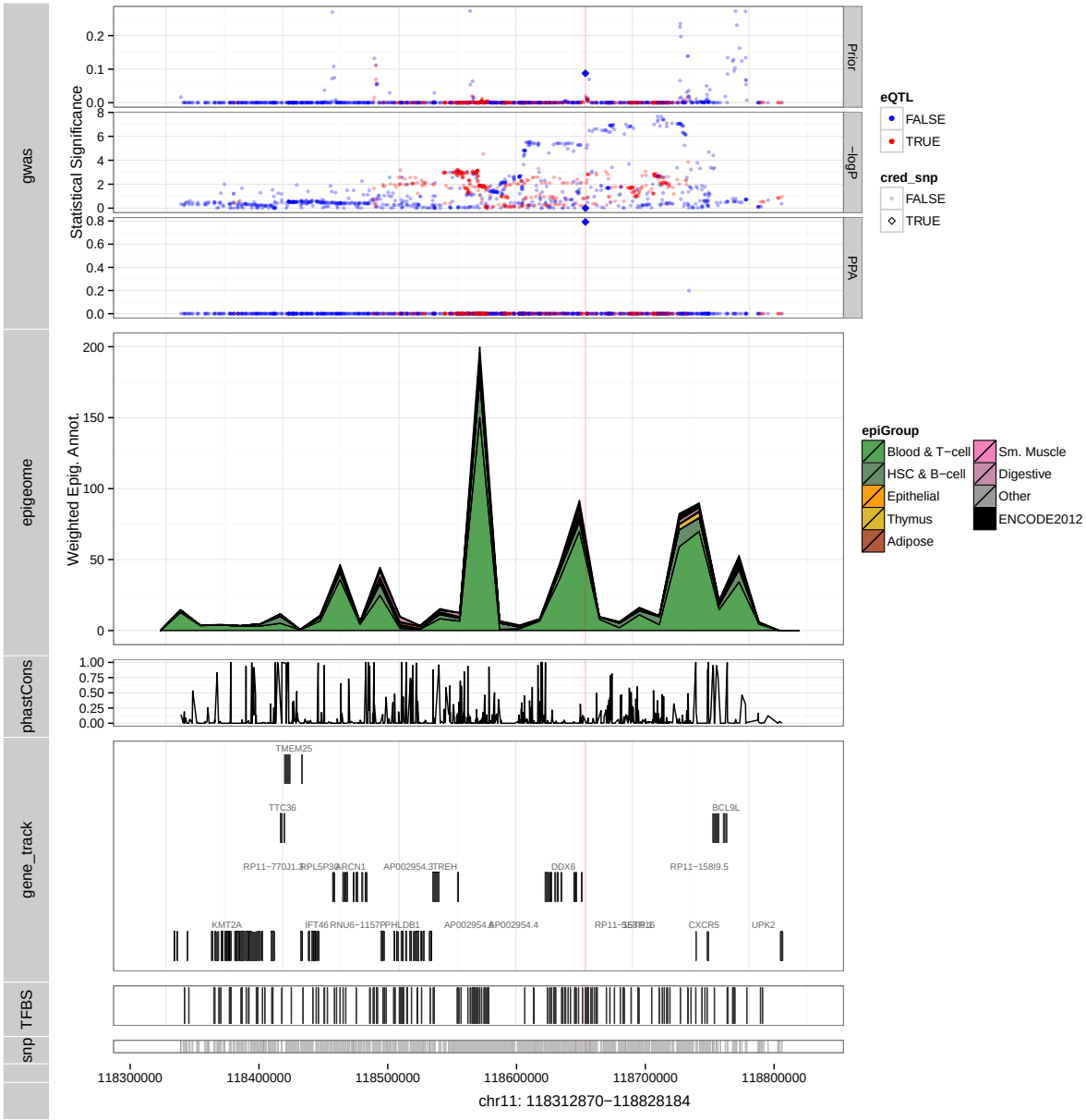
Multiple Sclerosis



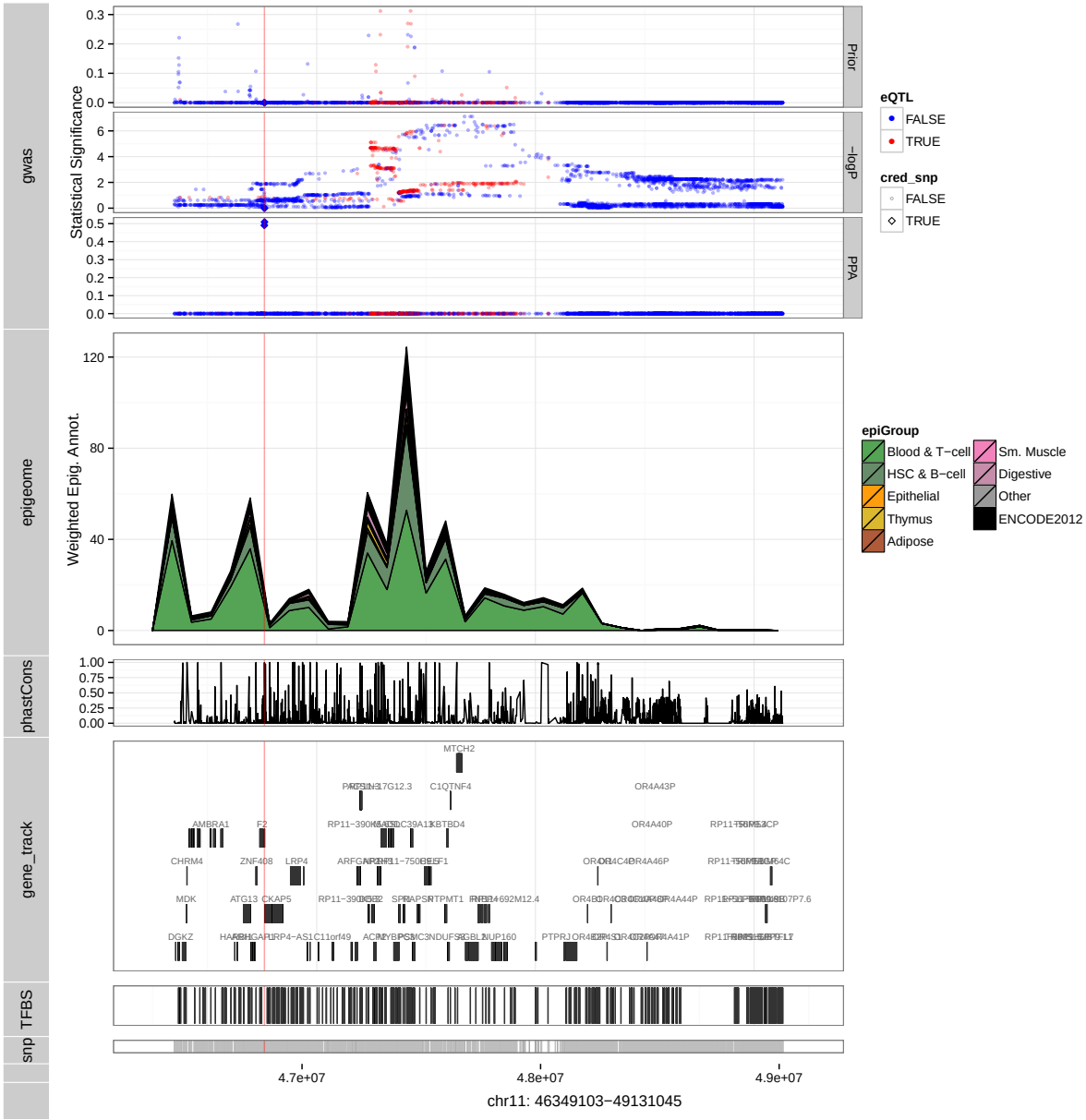
Multiple Sclerosis



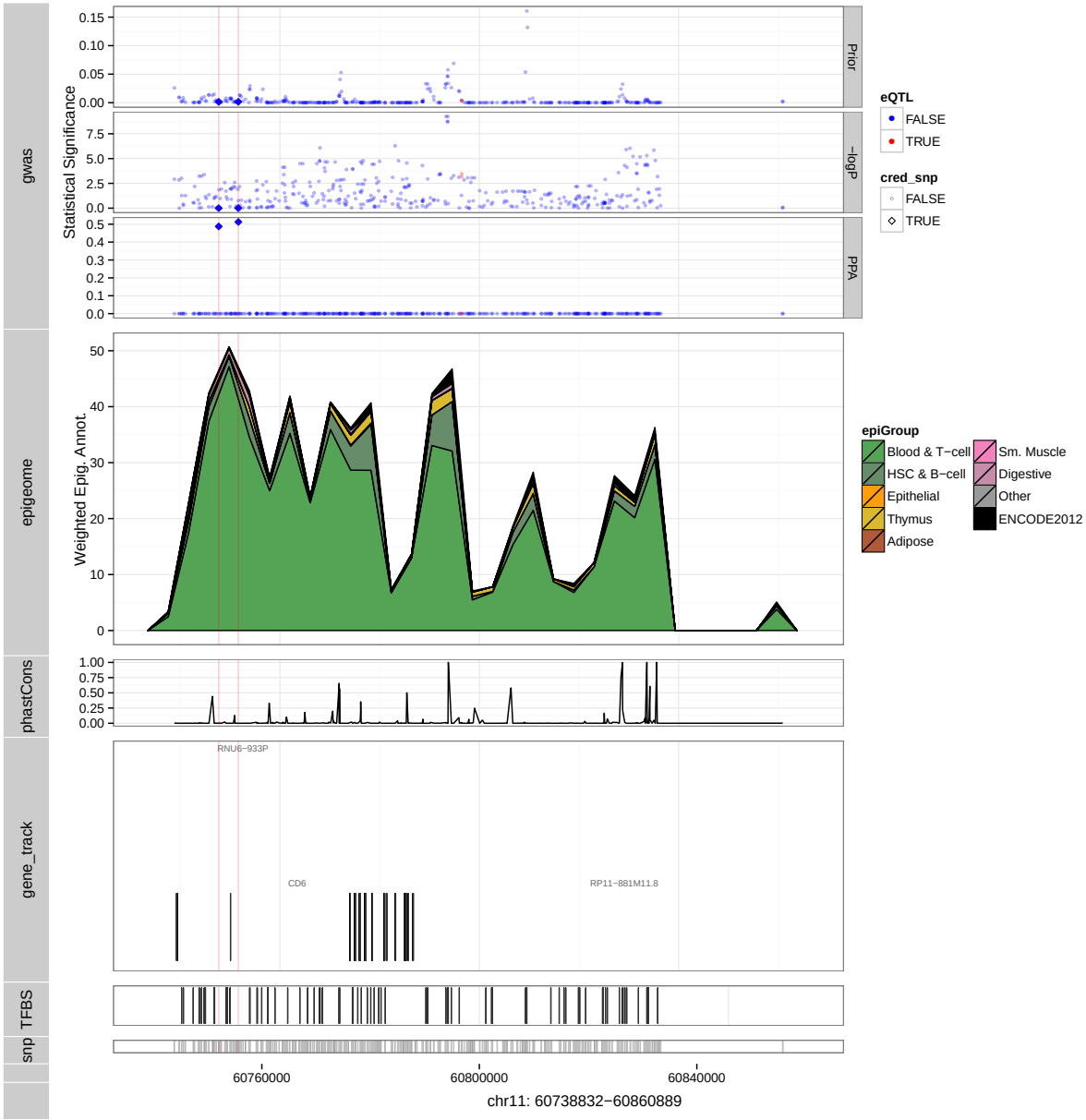
Multiple Sclerosis



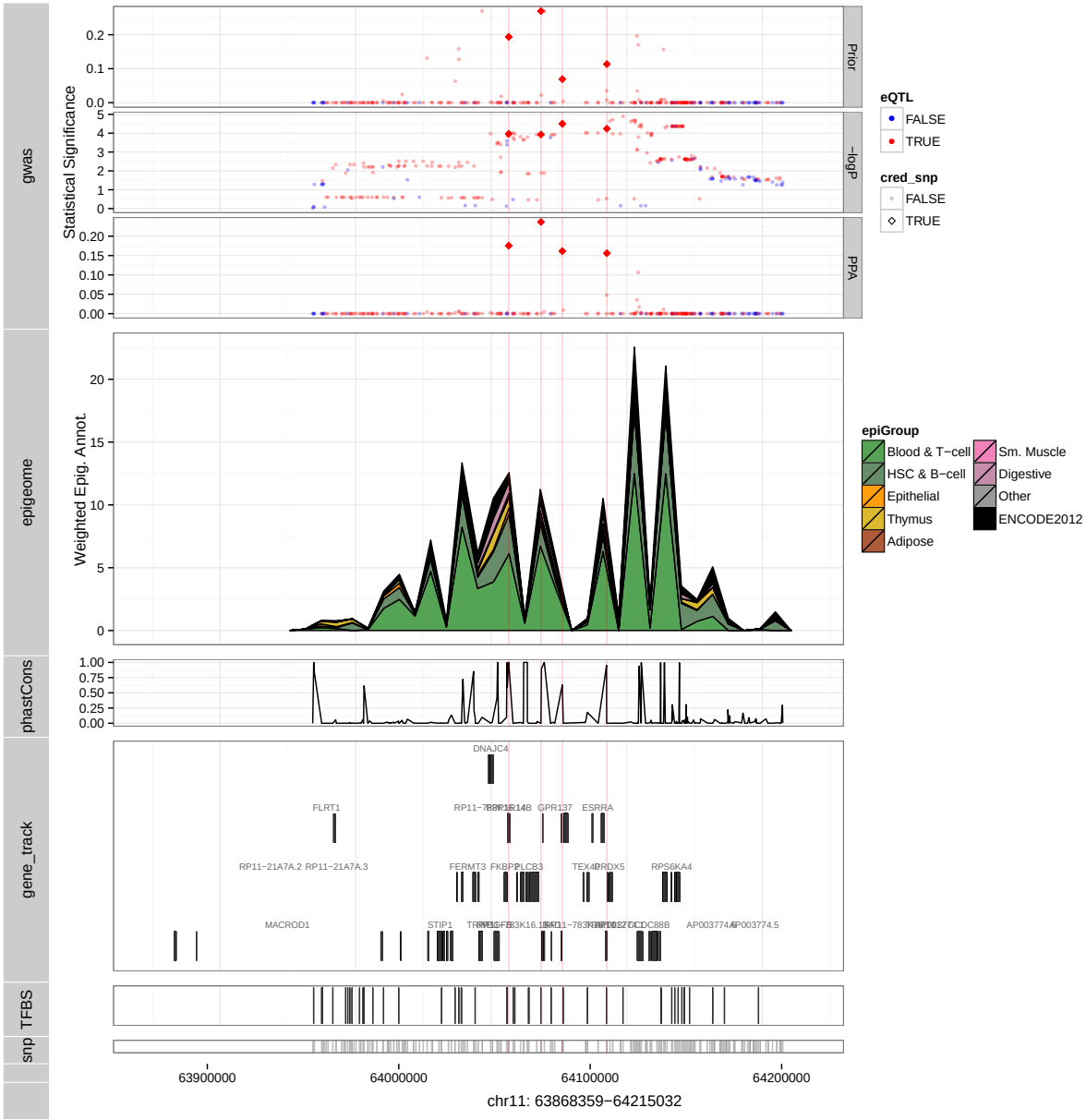
Multiple Sclerosis



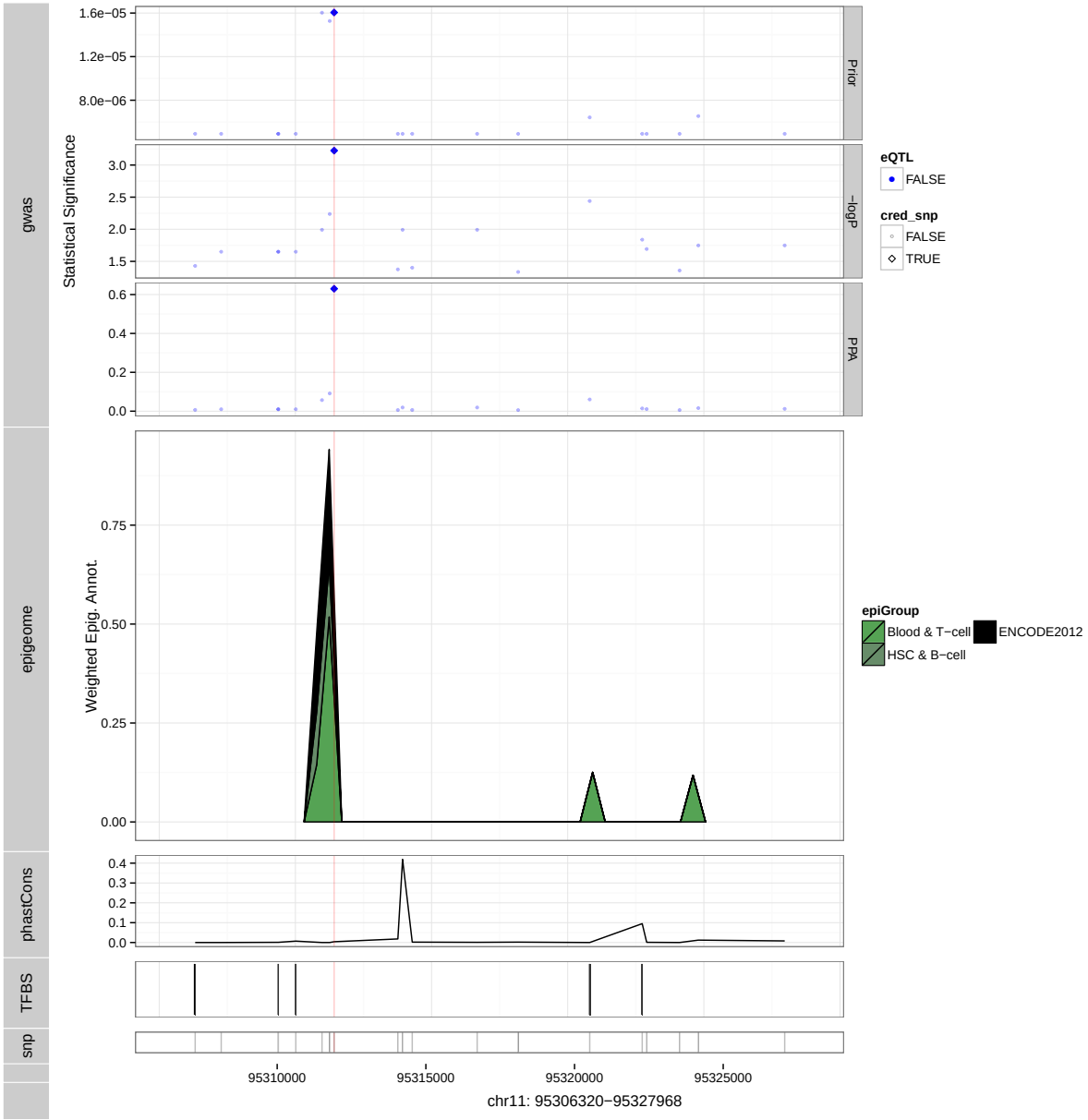
Multiple Sclerosis



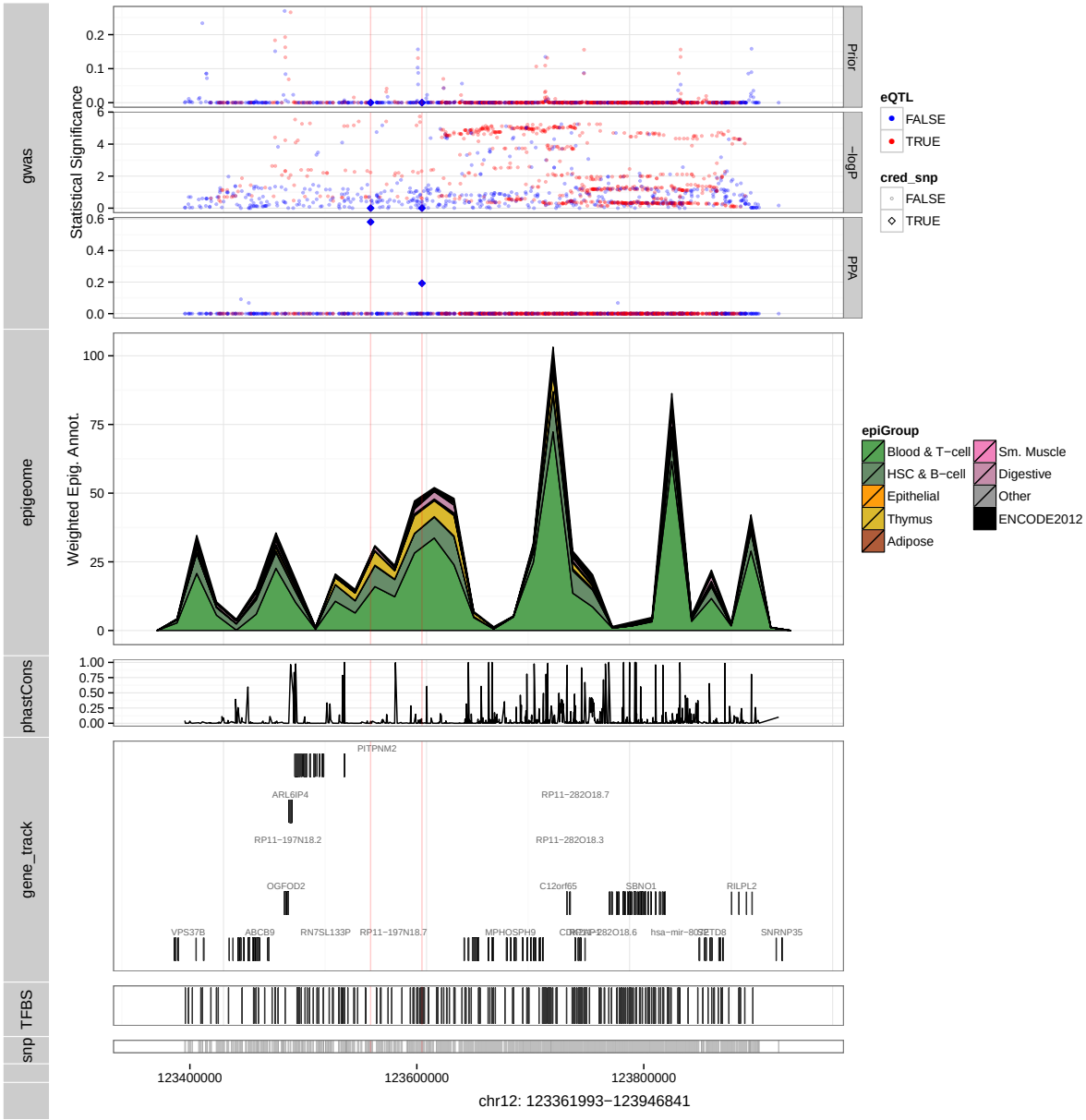
Multiple Sclerosis



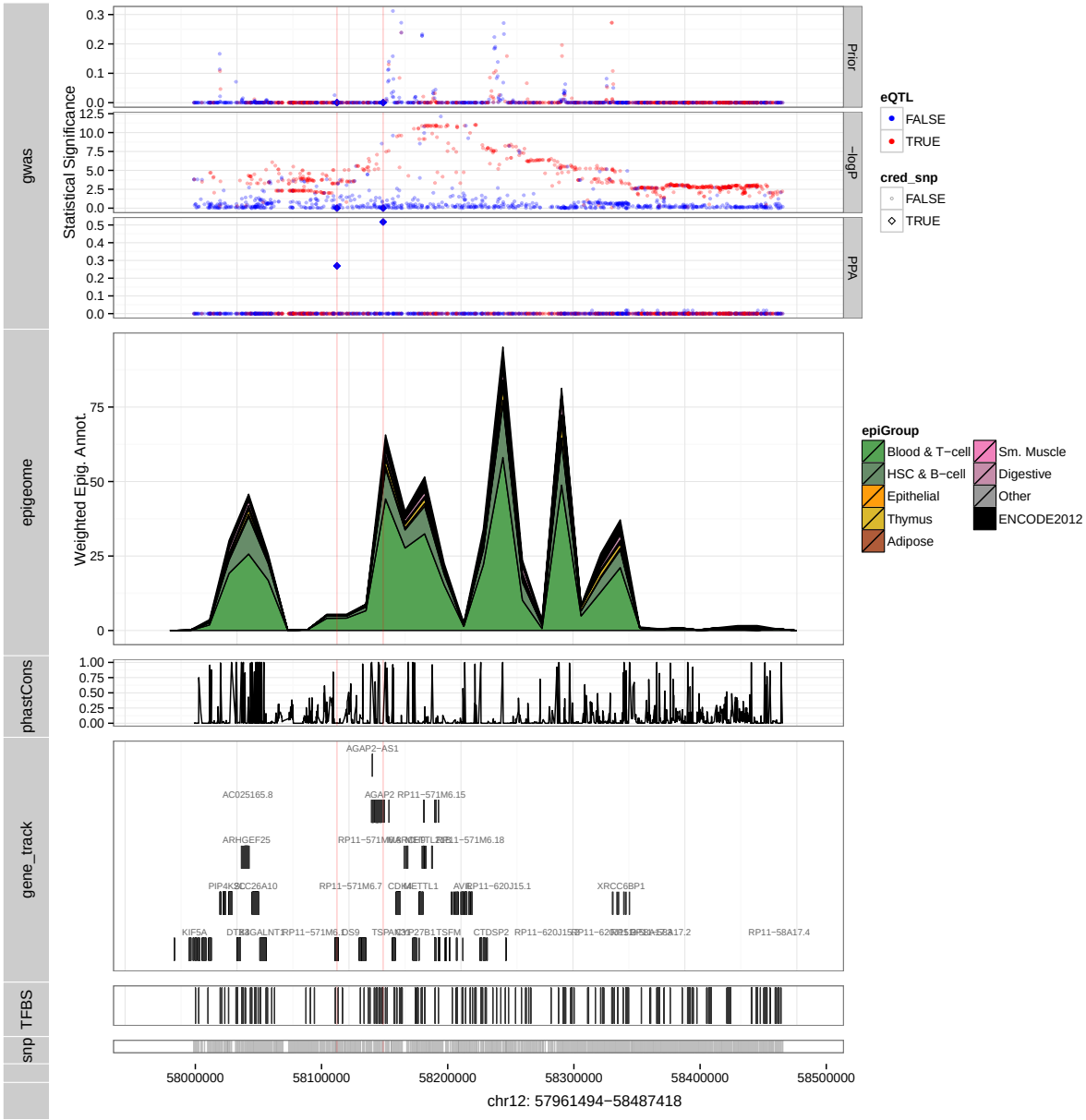
Multiple Sclerosis



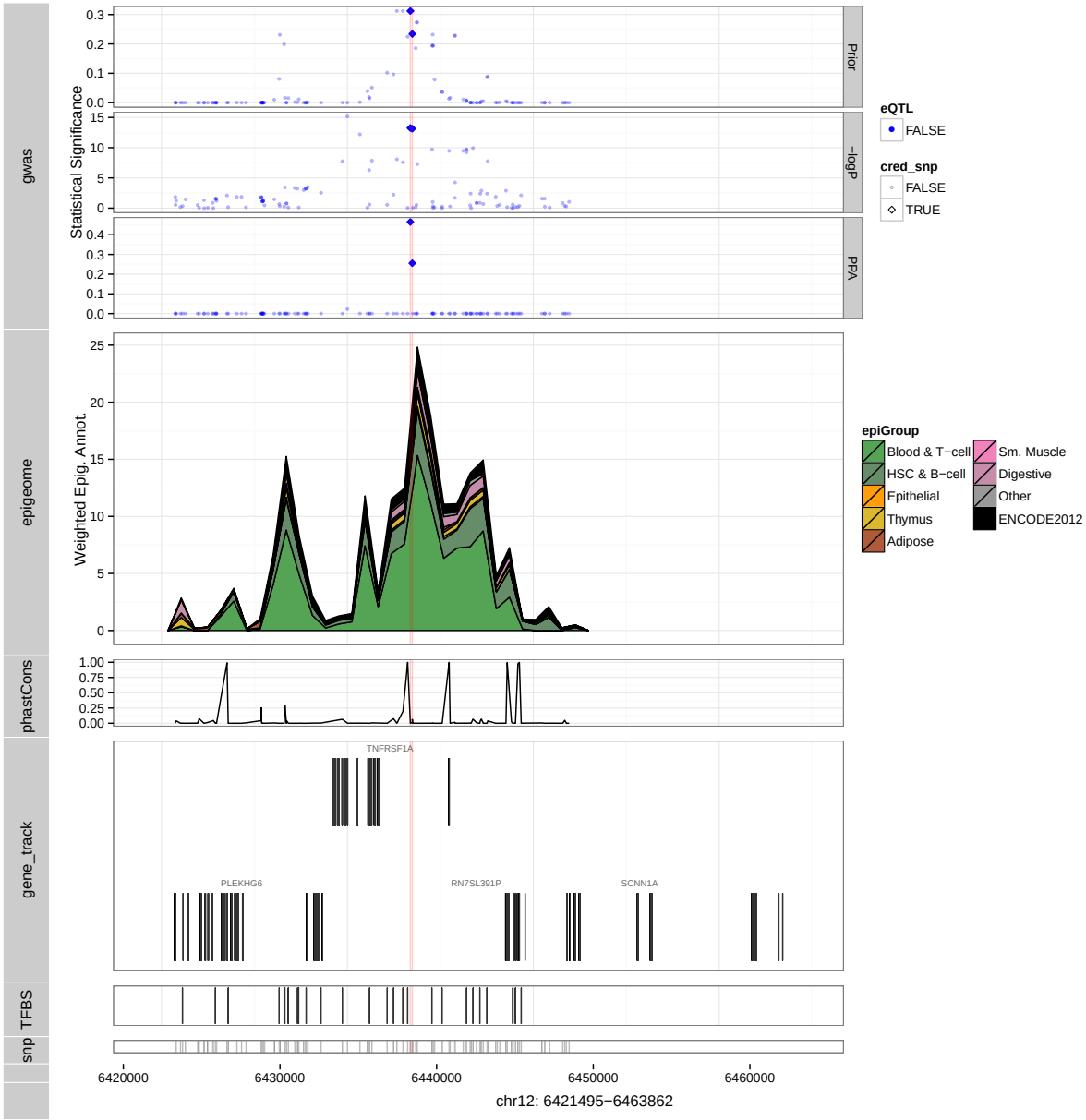
Multiple Sclerosis



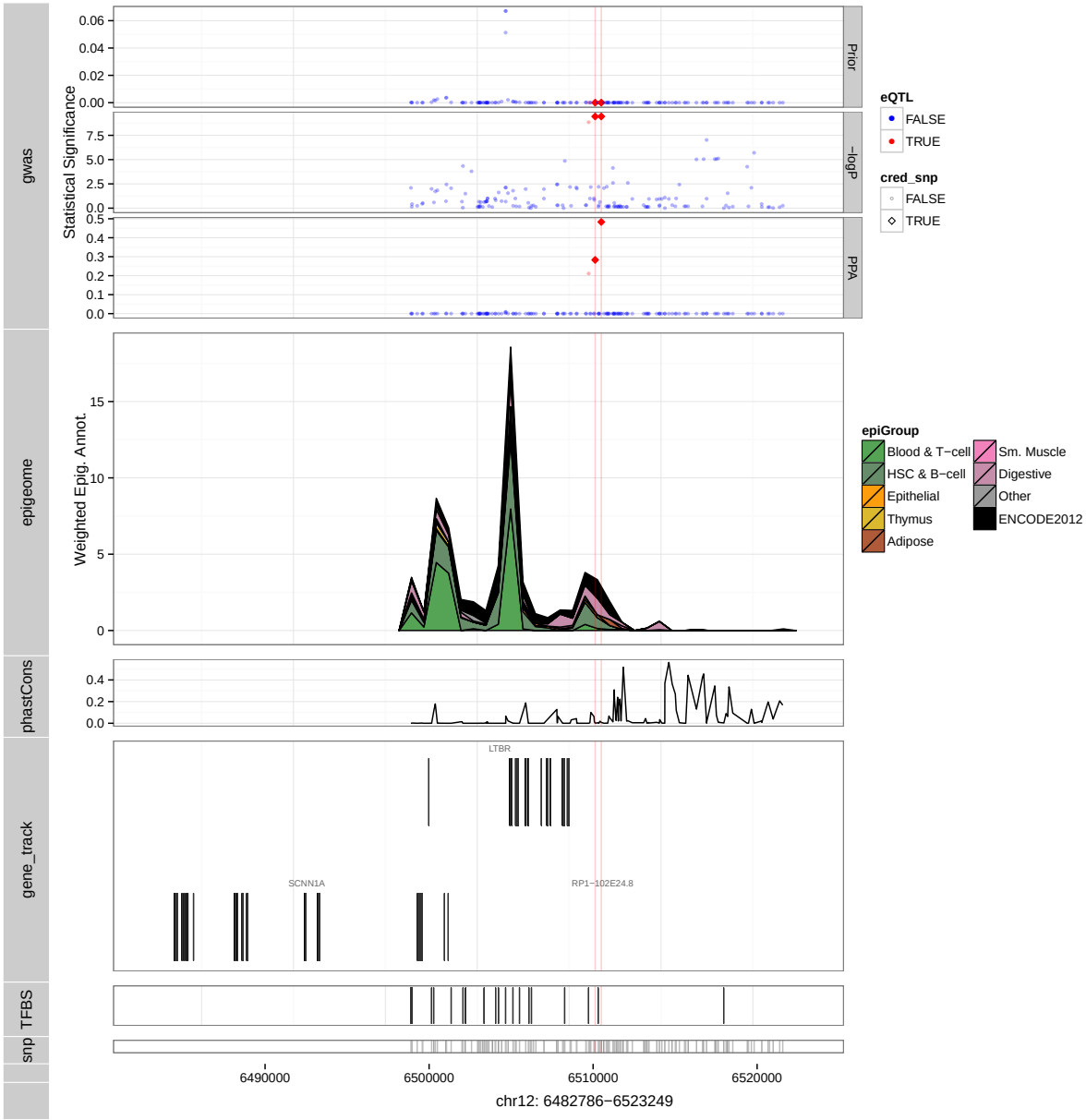
Multiple Sclerosis



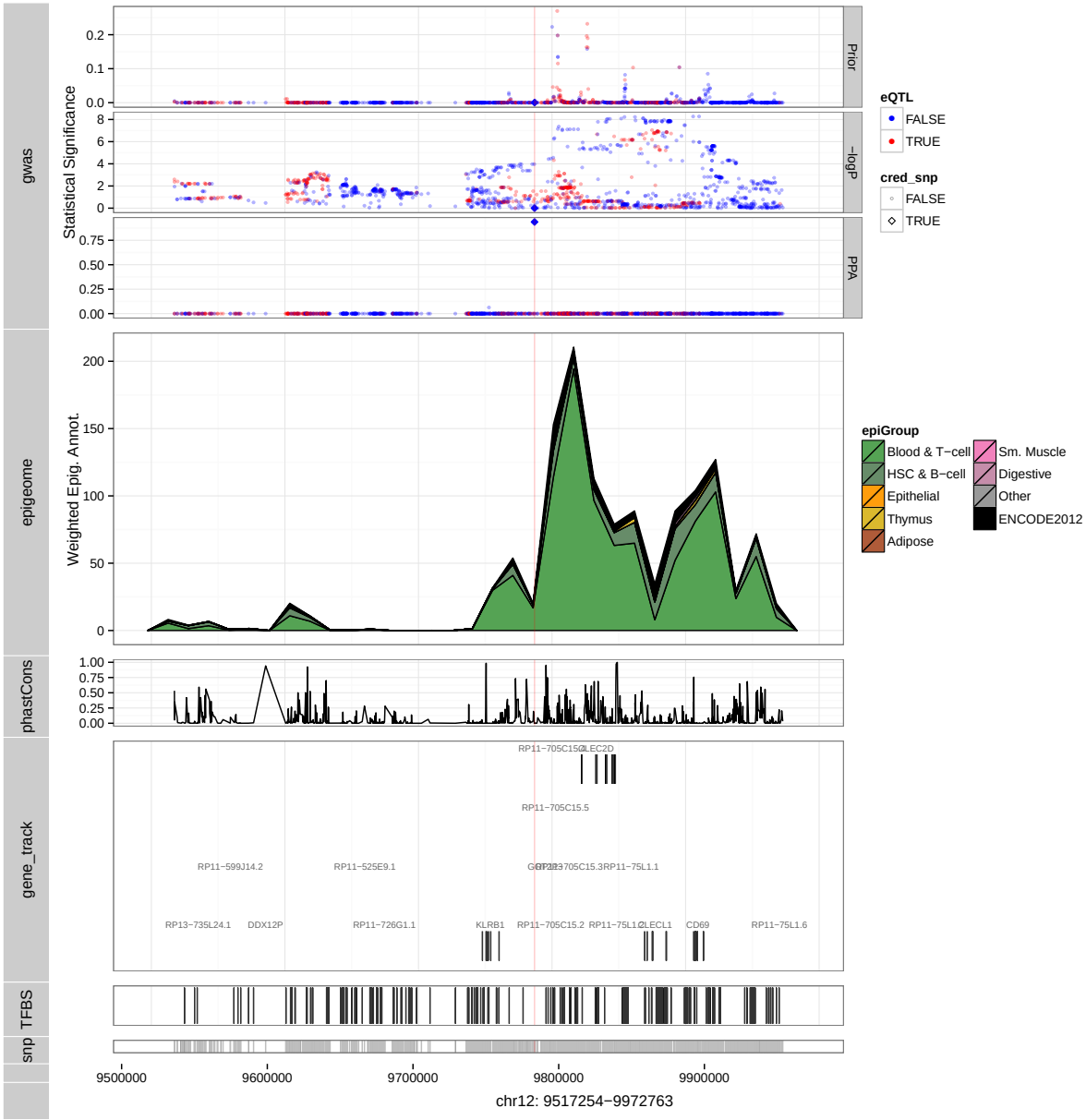
Multiple Sclerosis



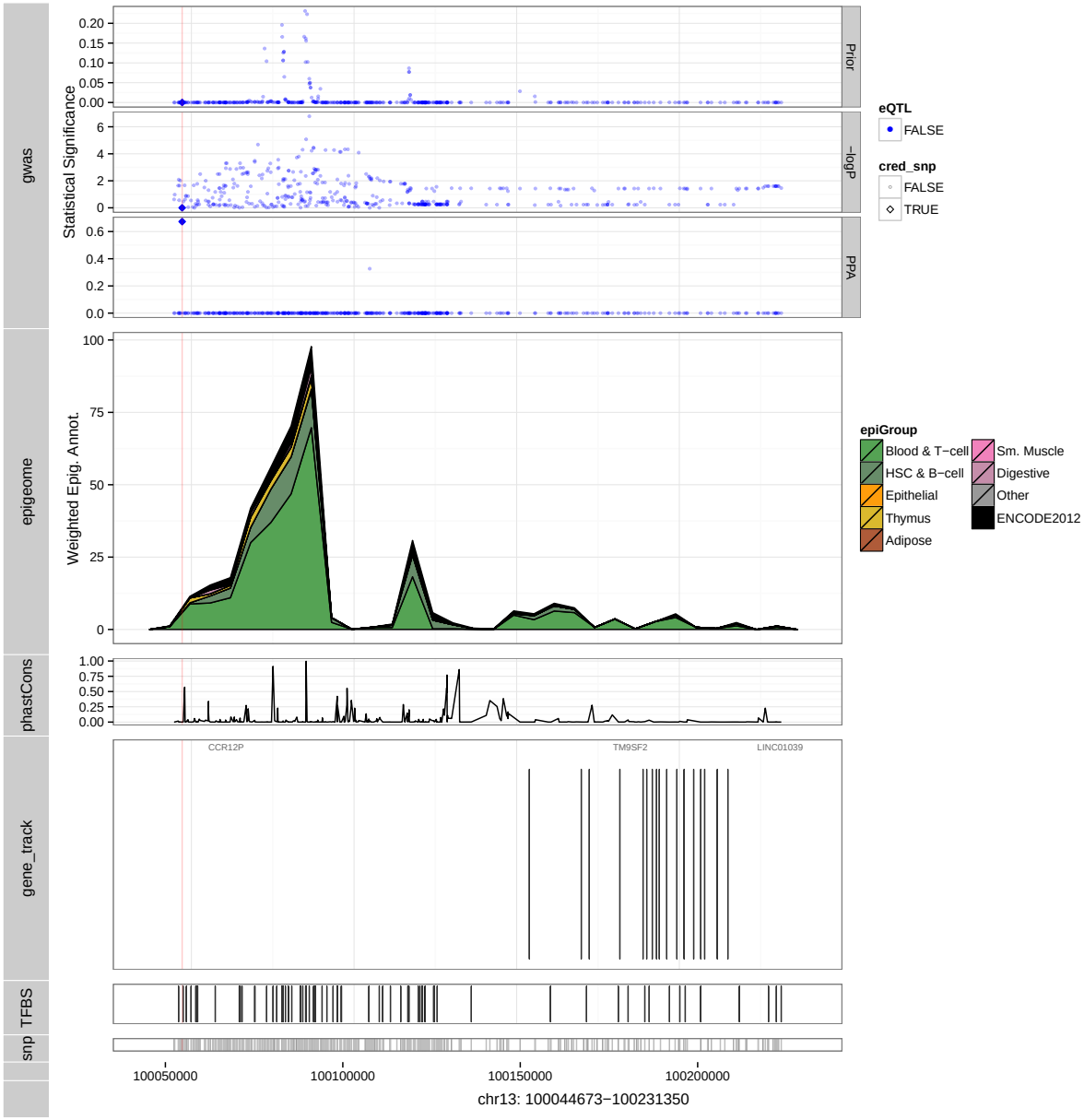
Multiple Sclerosis



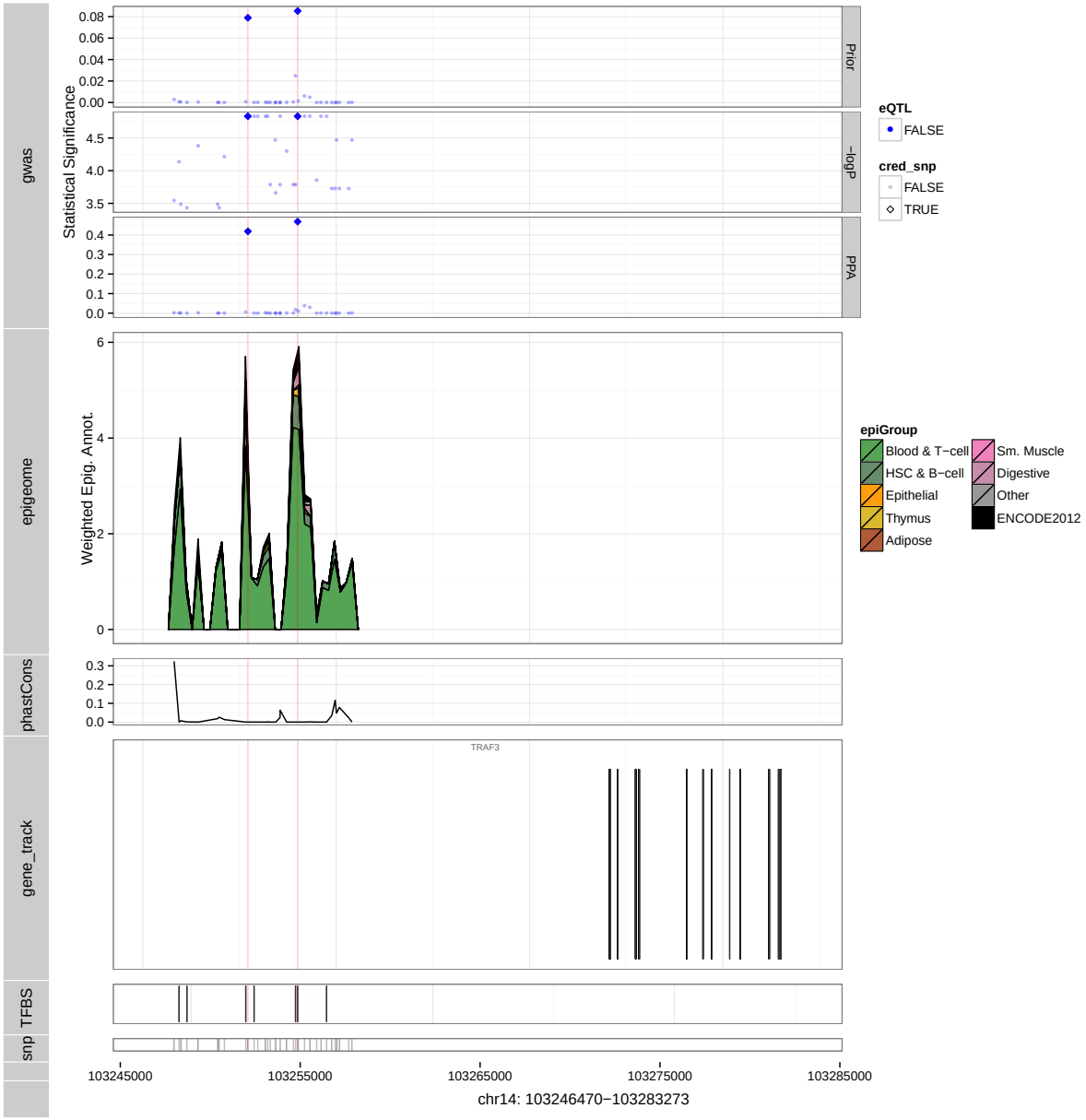
Multiple Sclerosis



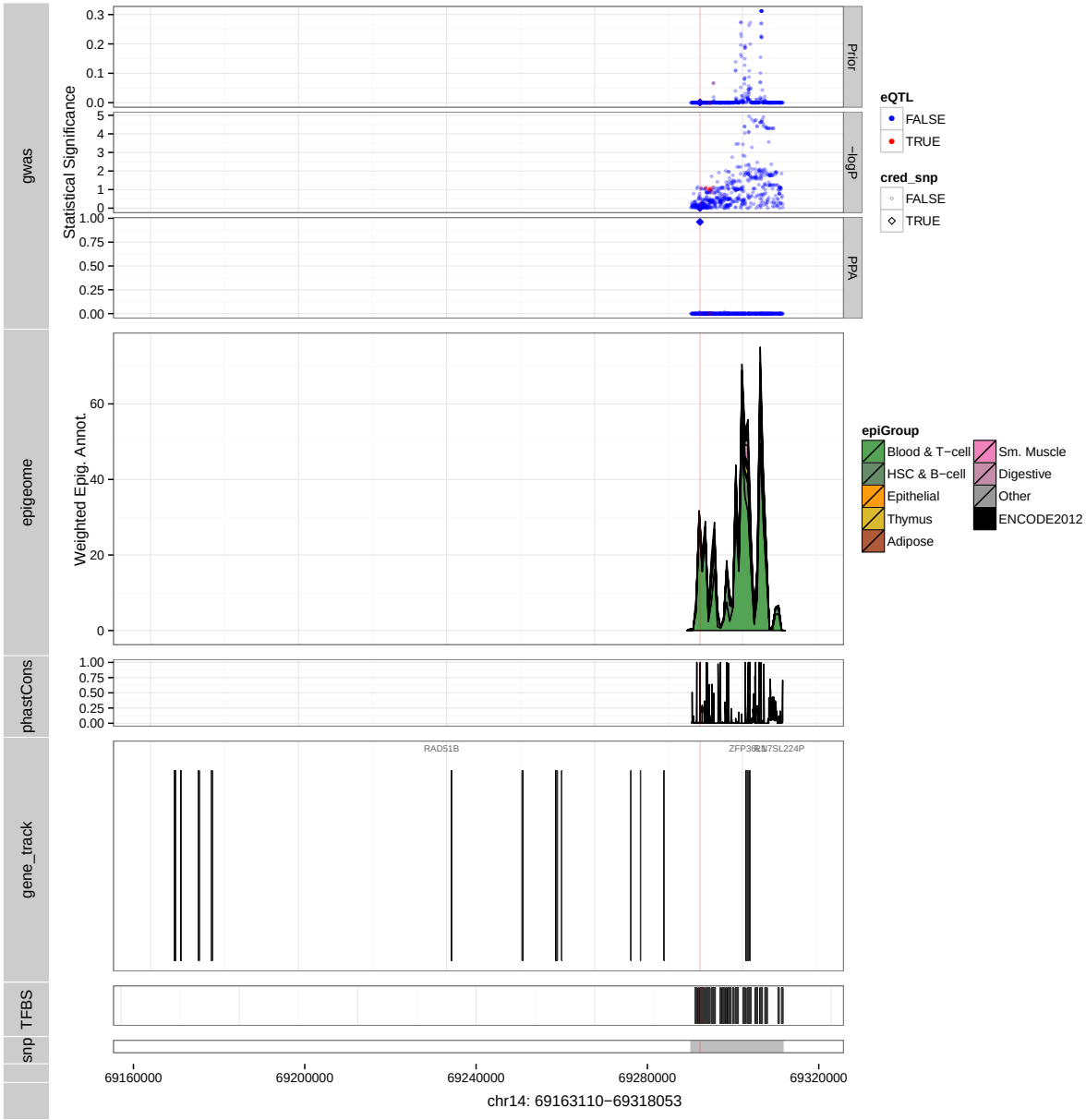
Multiple Sclerosis



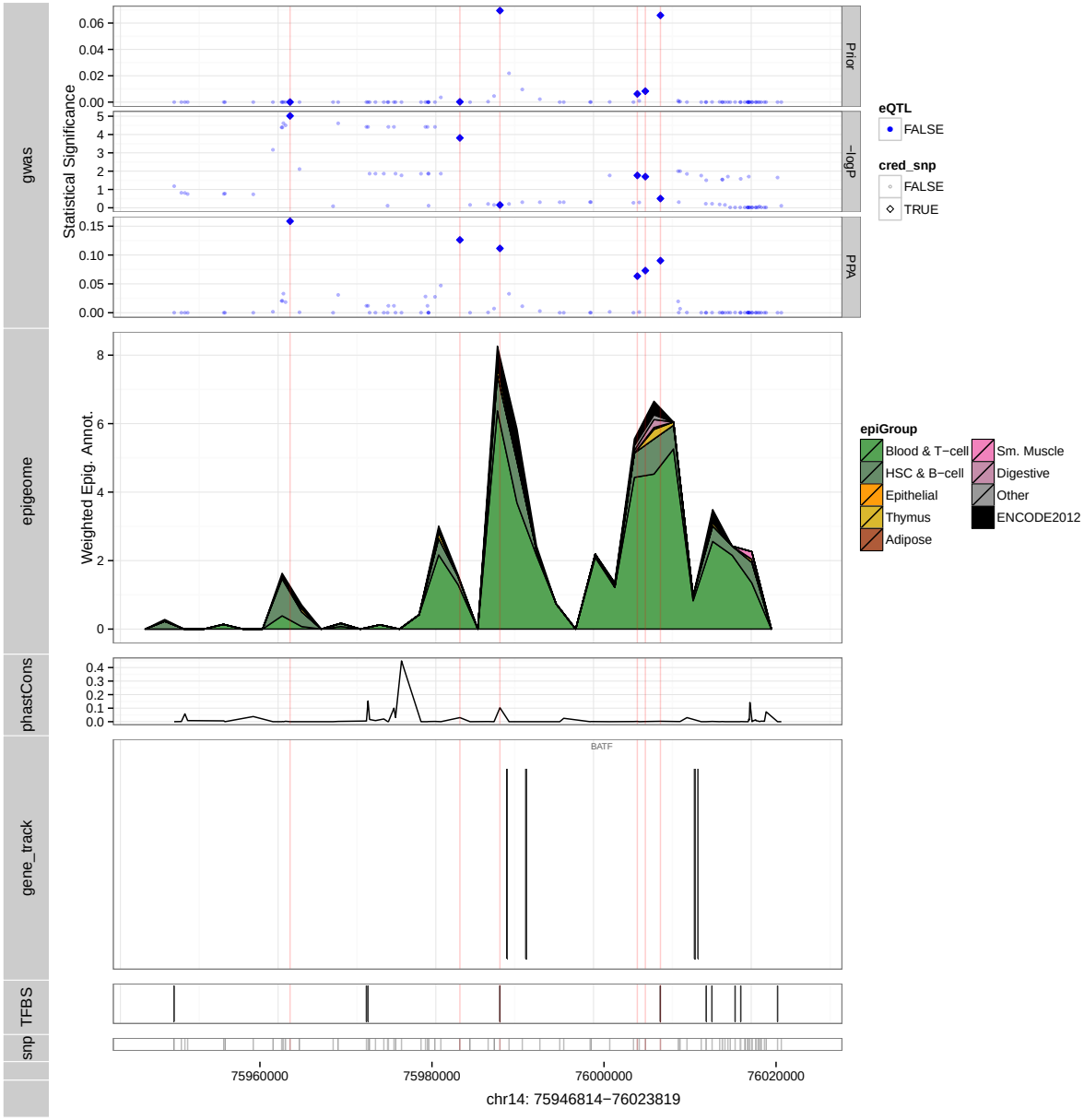
Multiple Sclerosis



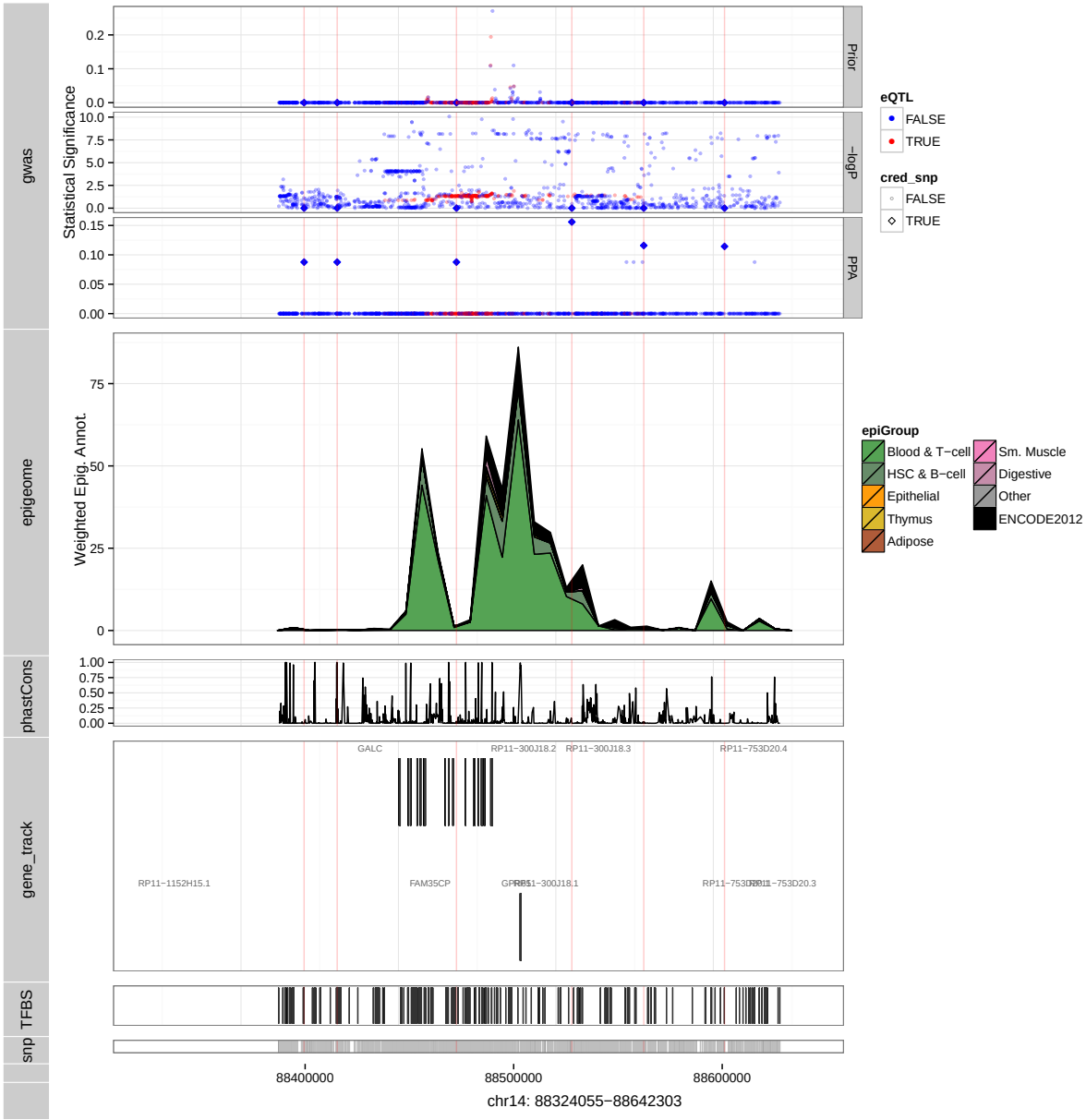
Multiple Sclerosis



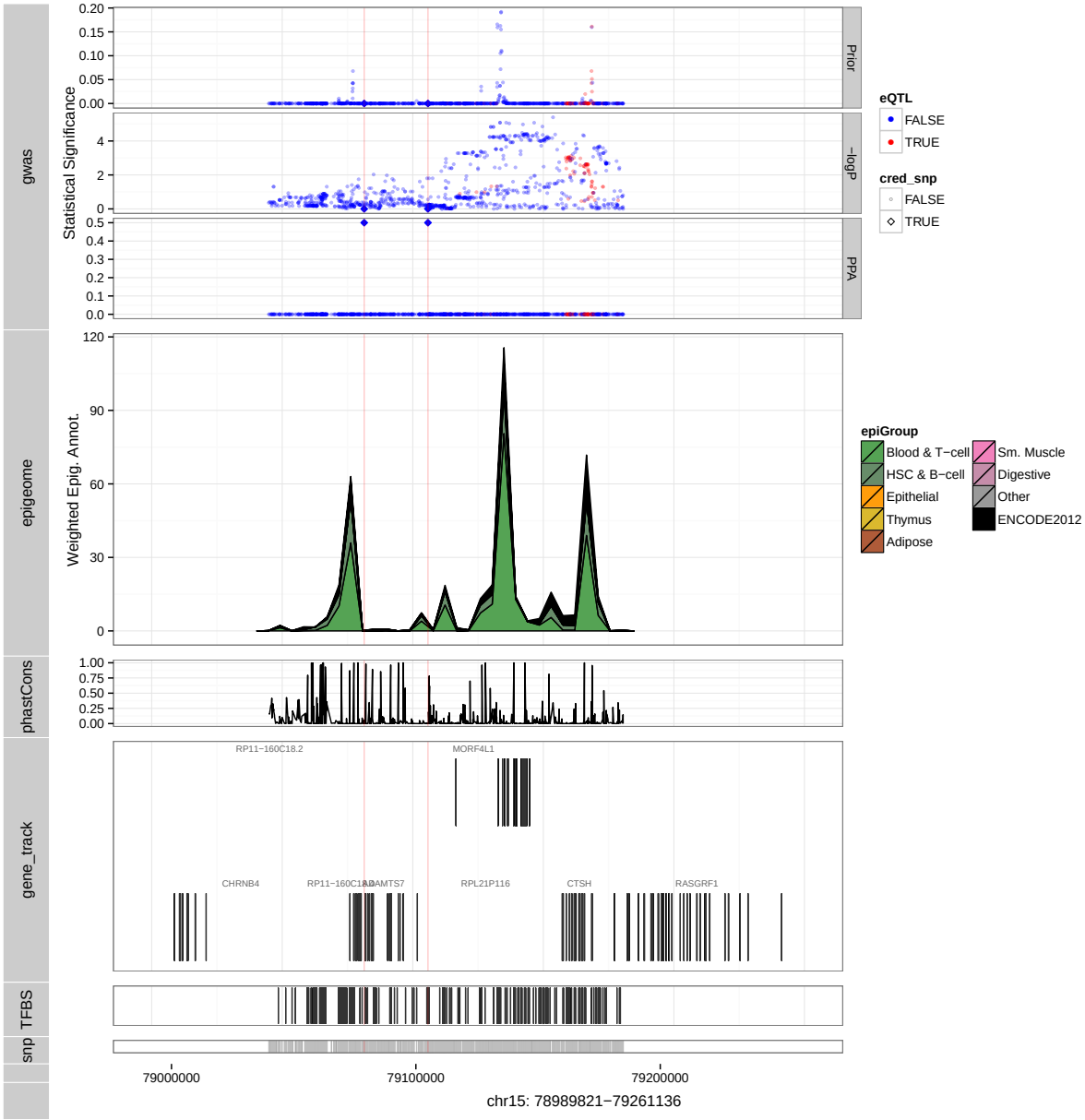
Multiple Sclerosis



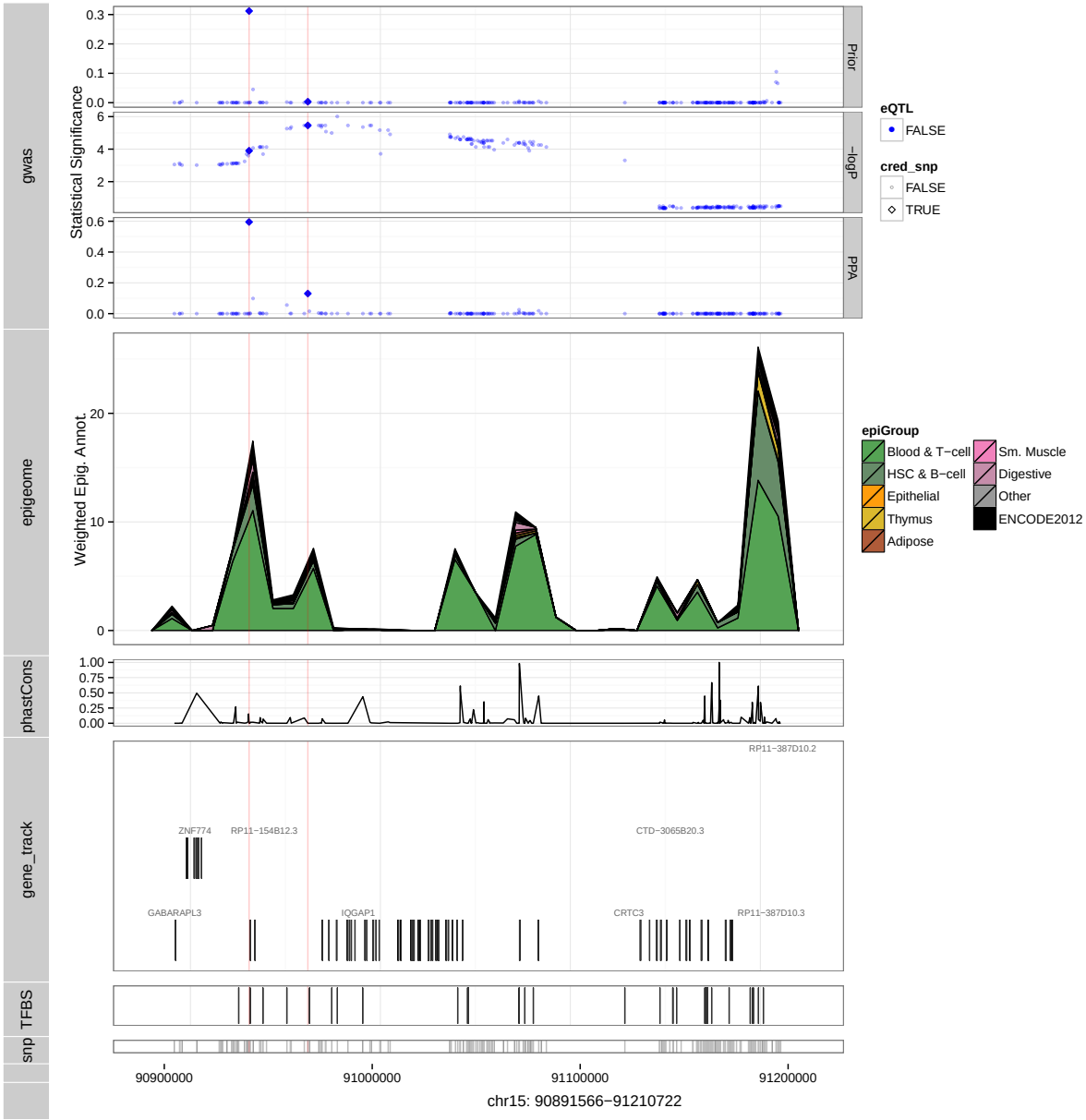
Multiple Sclerosis



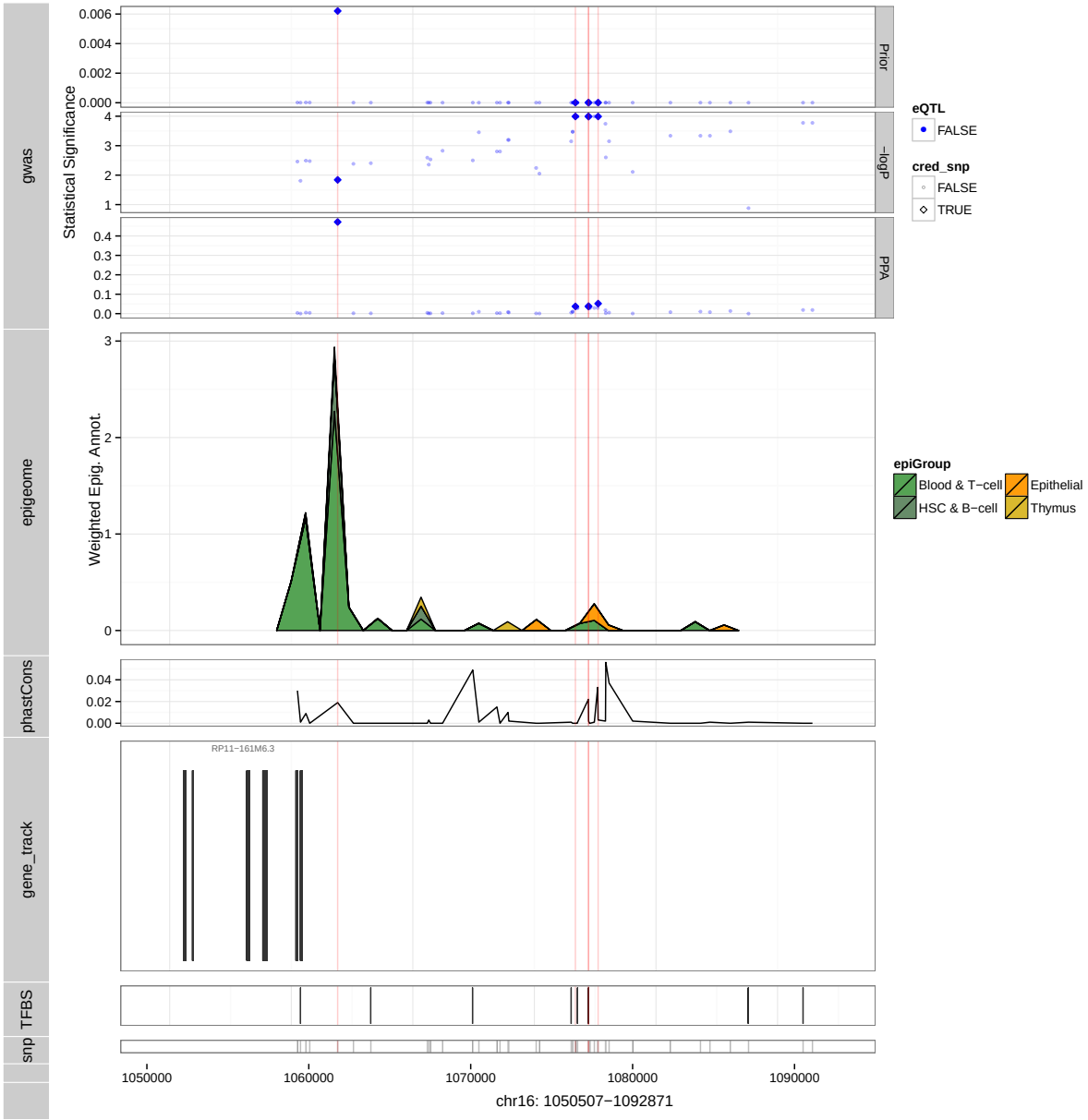
Multiple Sclerosis



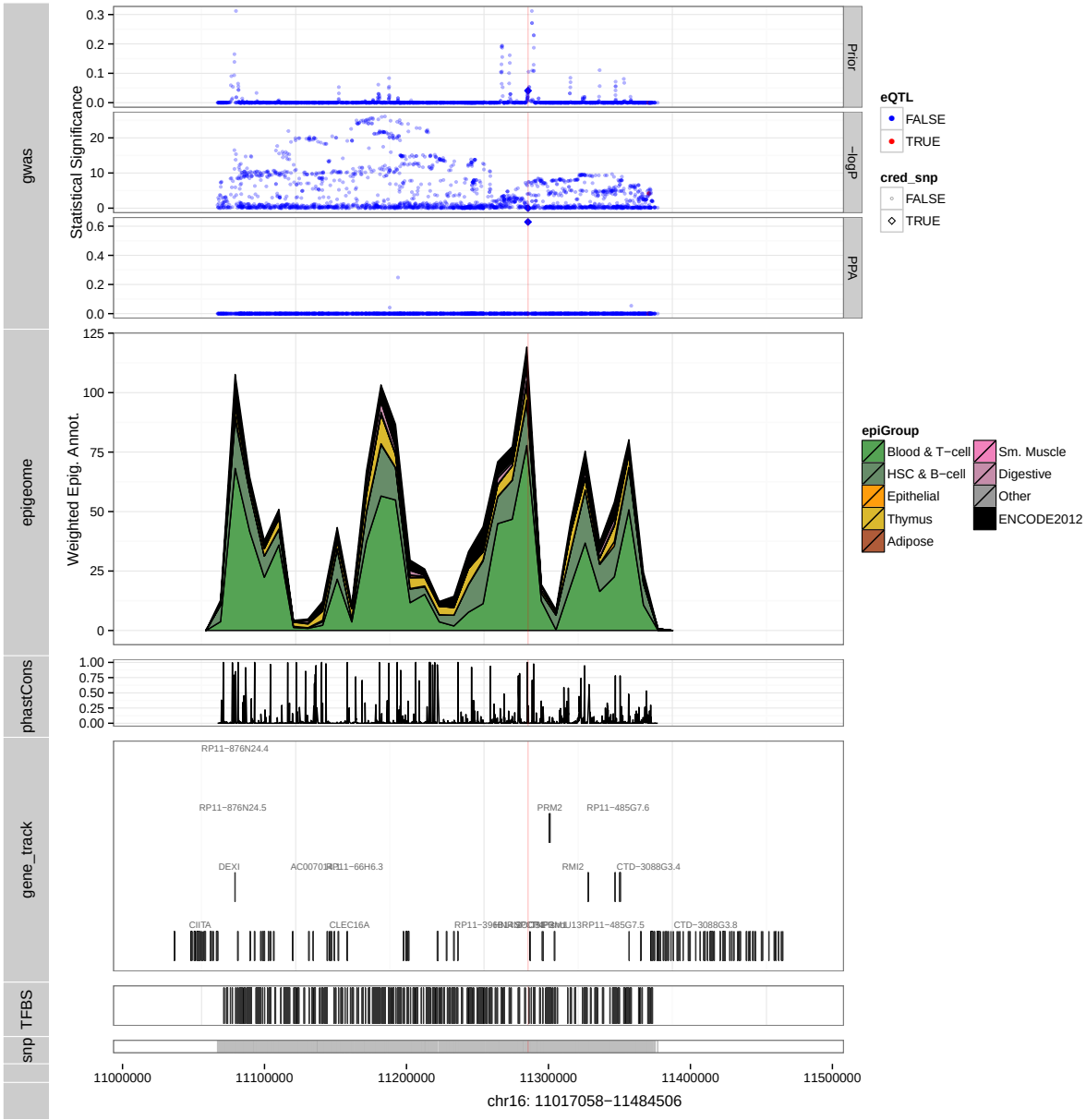
Multiple Sclerosis



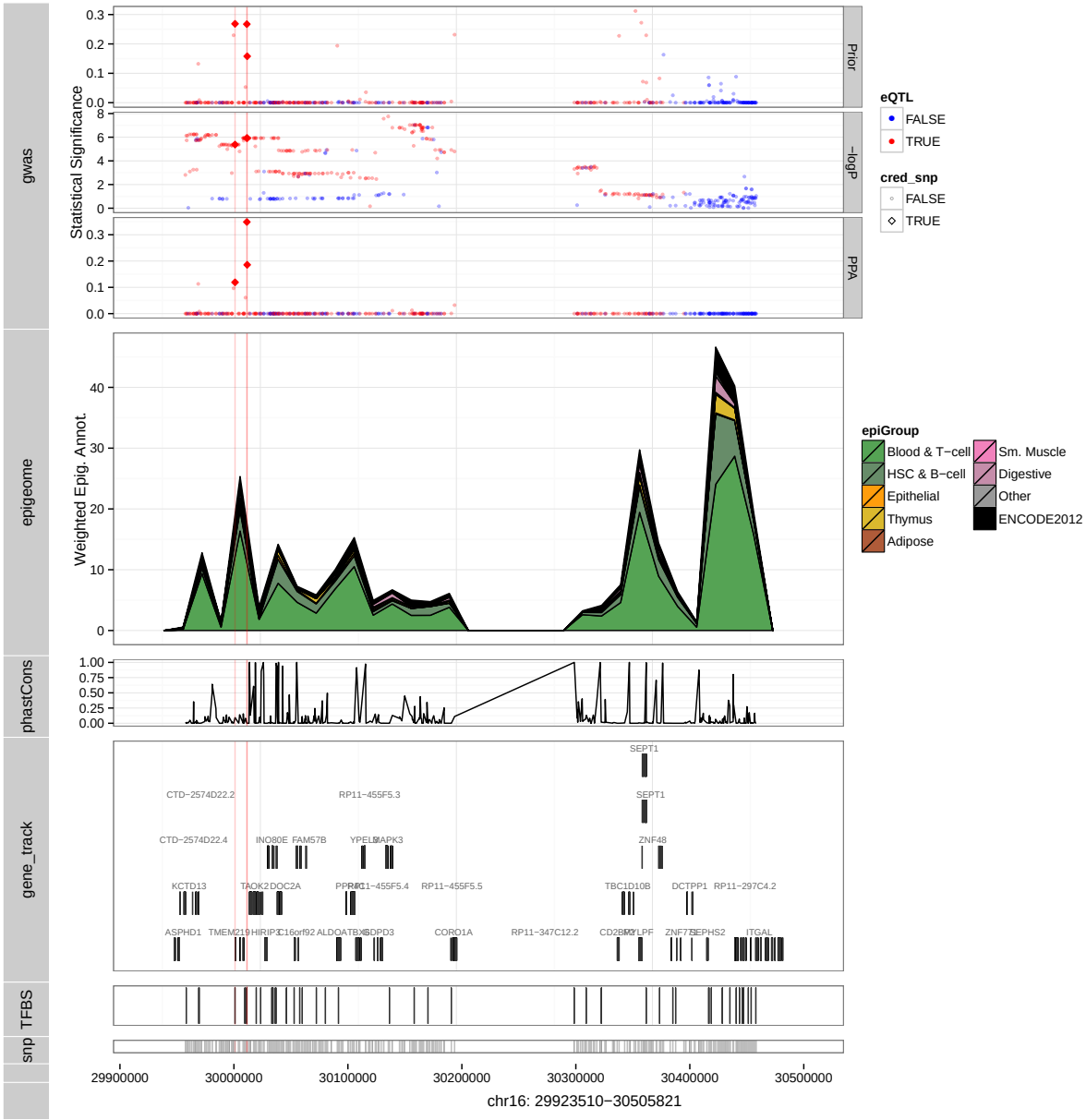
Multiple Sclerosis



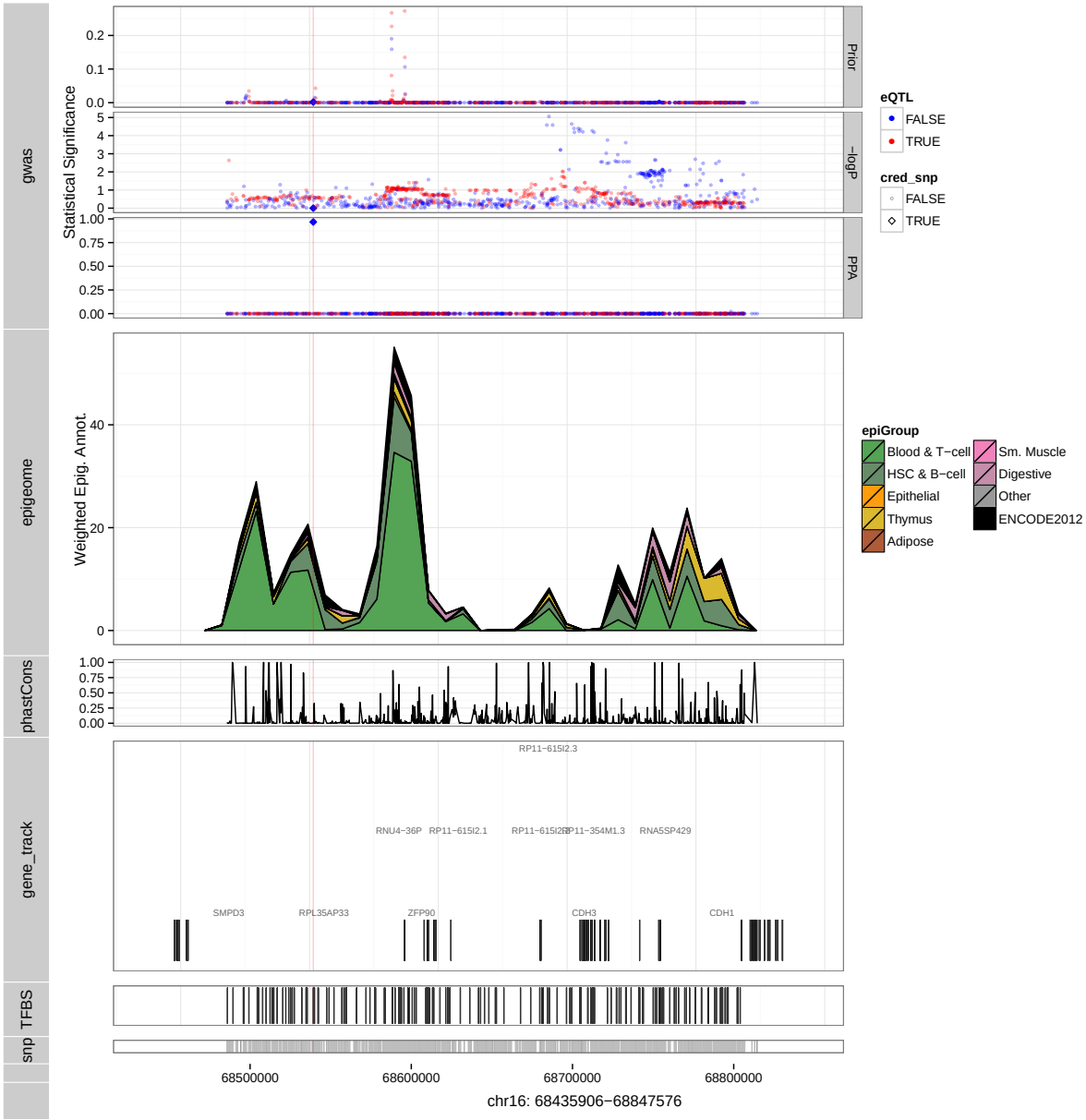
Multiple Sclerosis



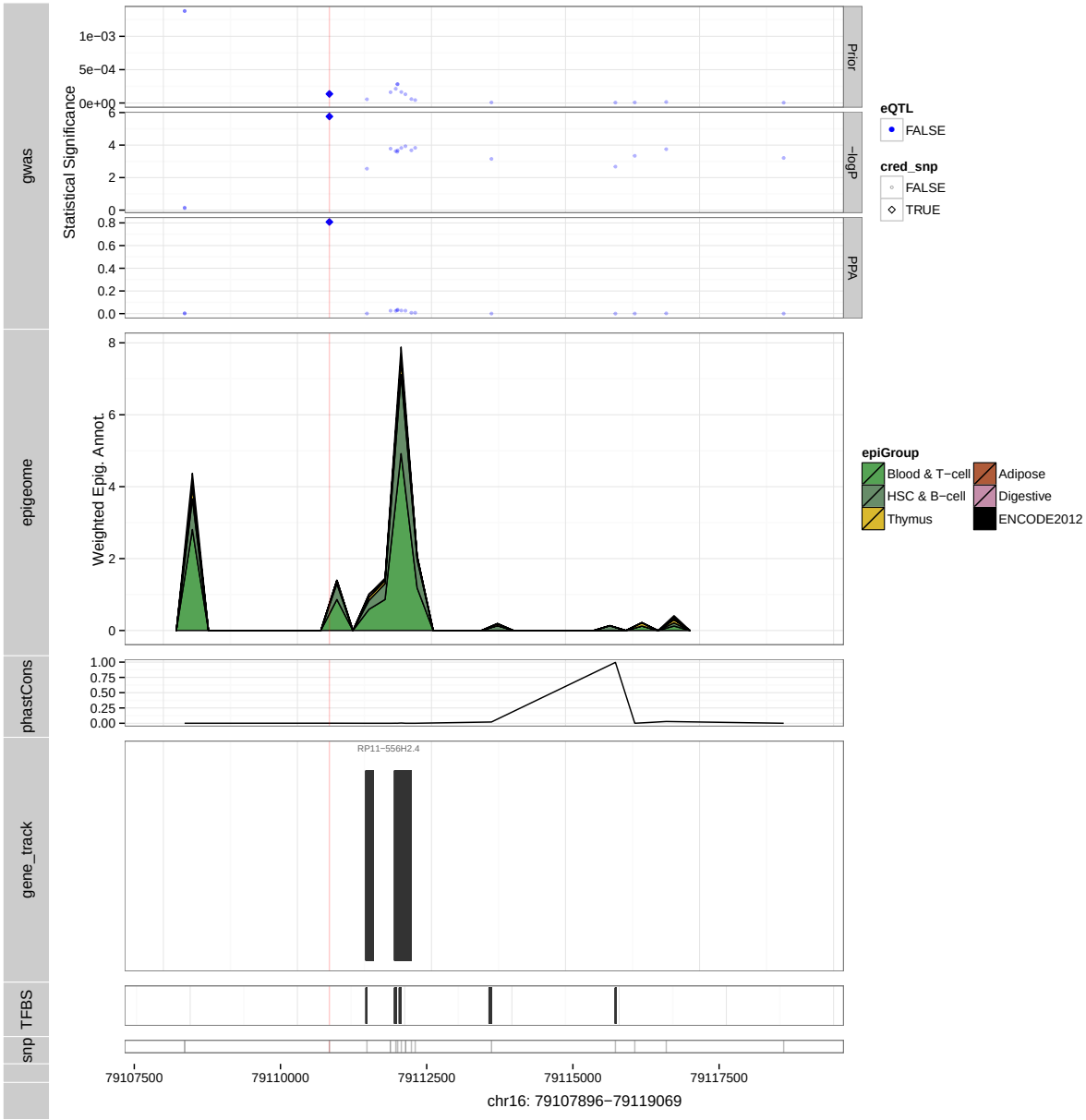
Multiple Sclerosis



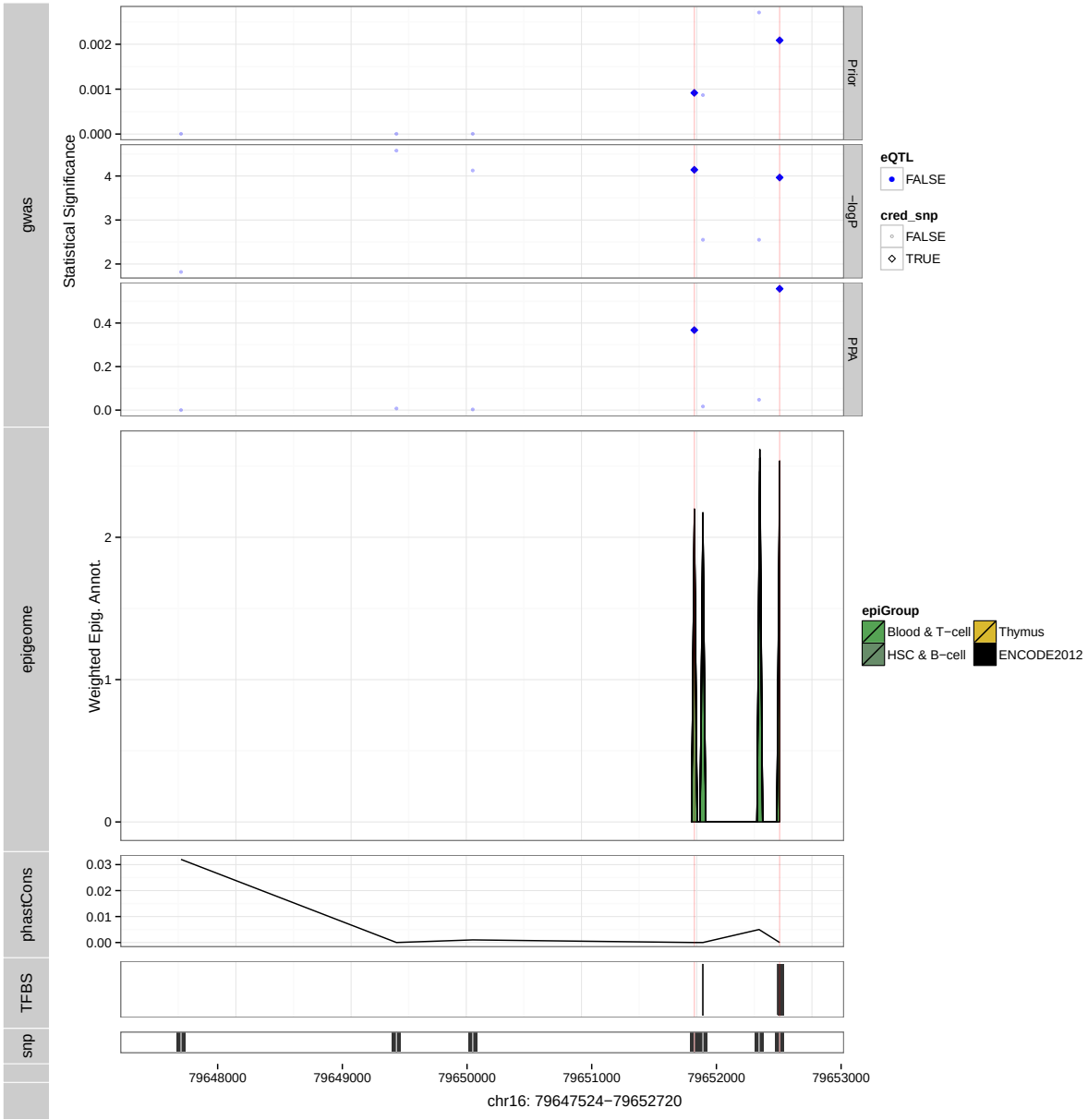
Multiple Sclerosis



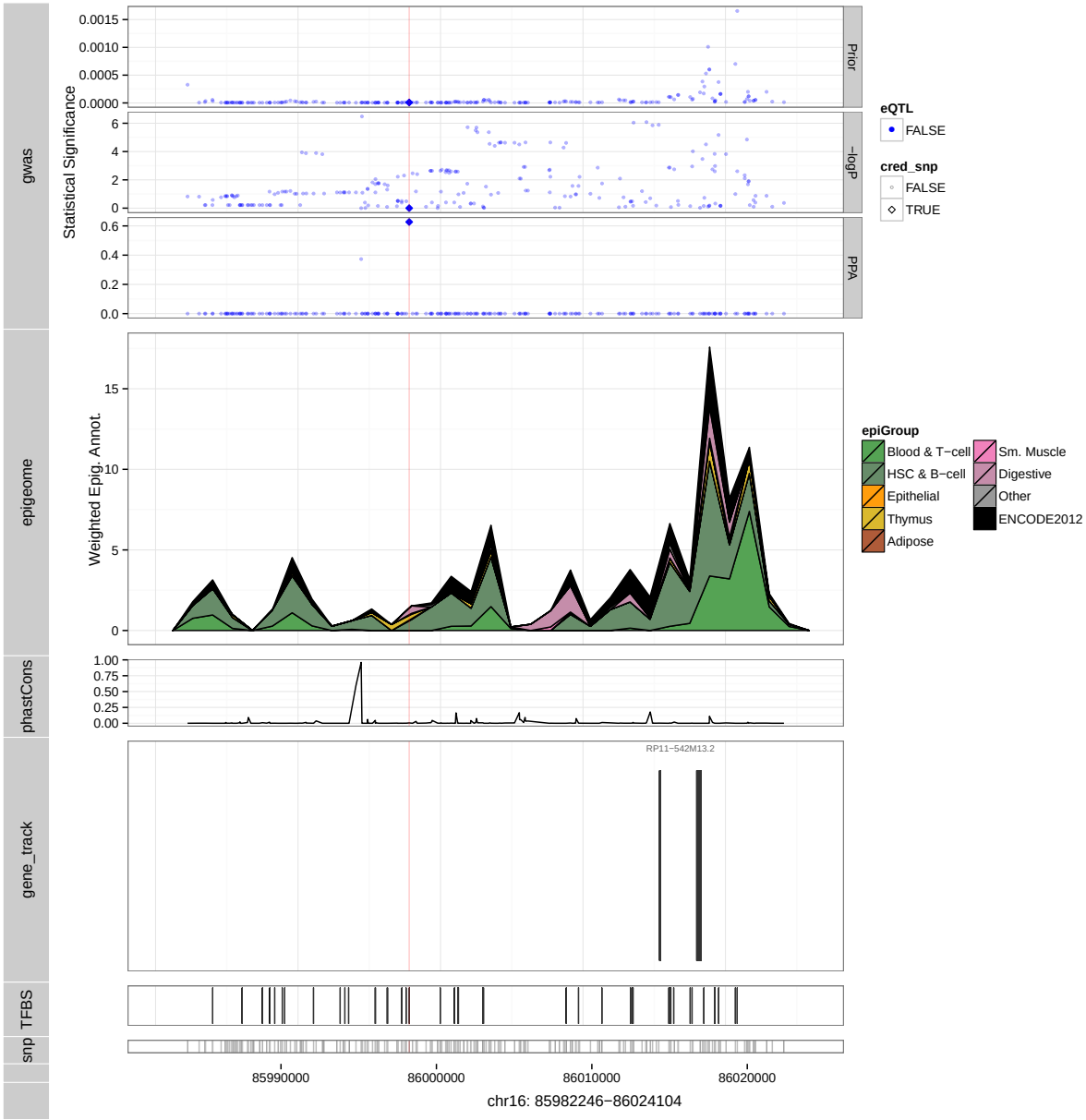
Multiple Sclerosis



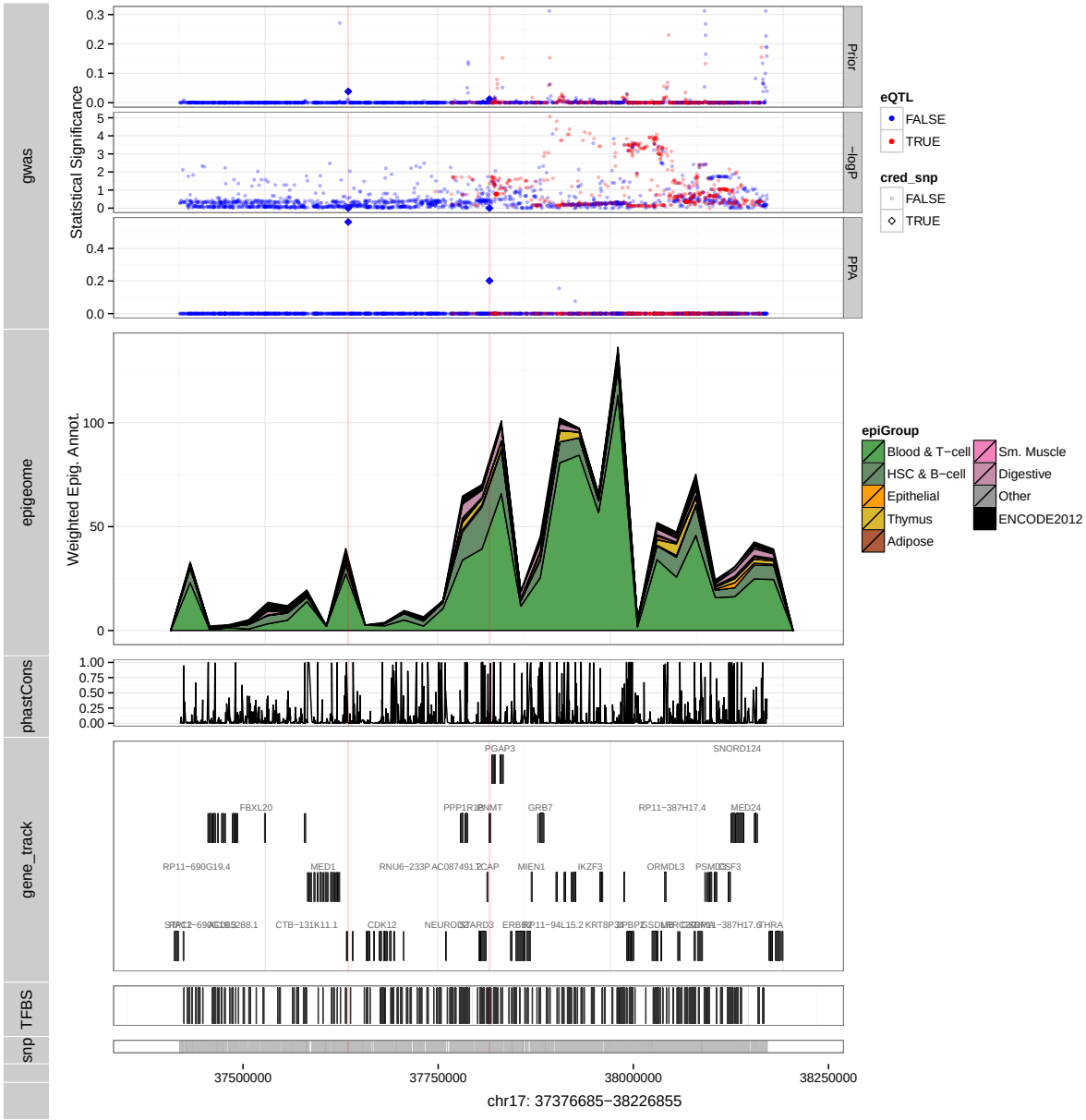
Multiple Sclerosis



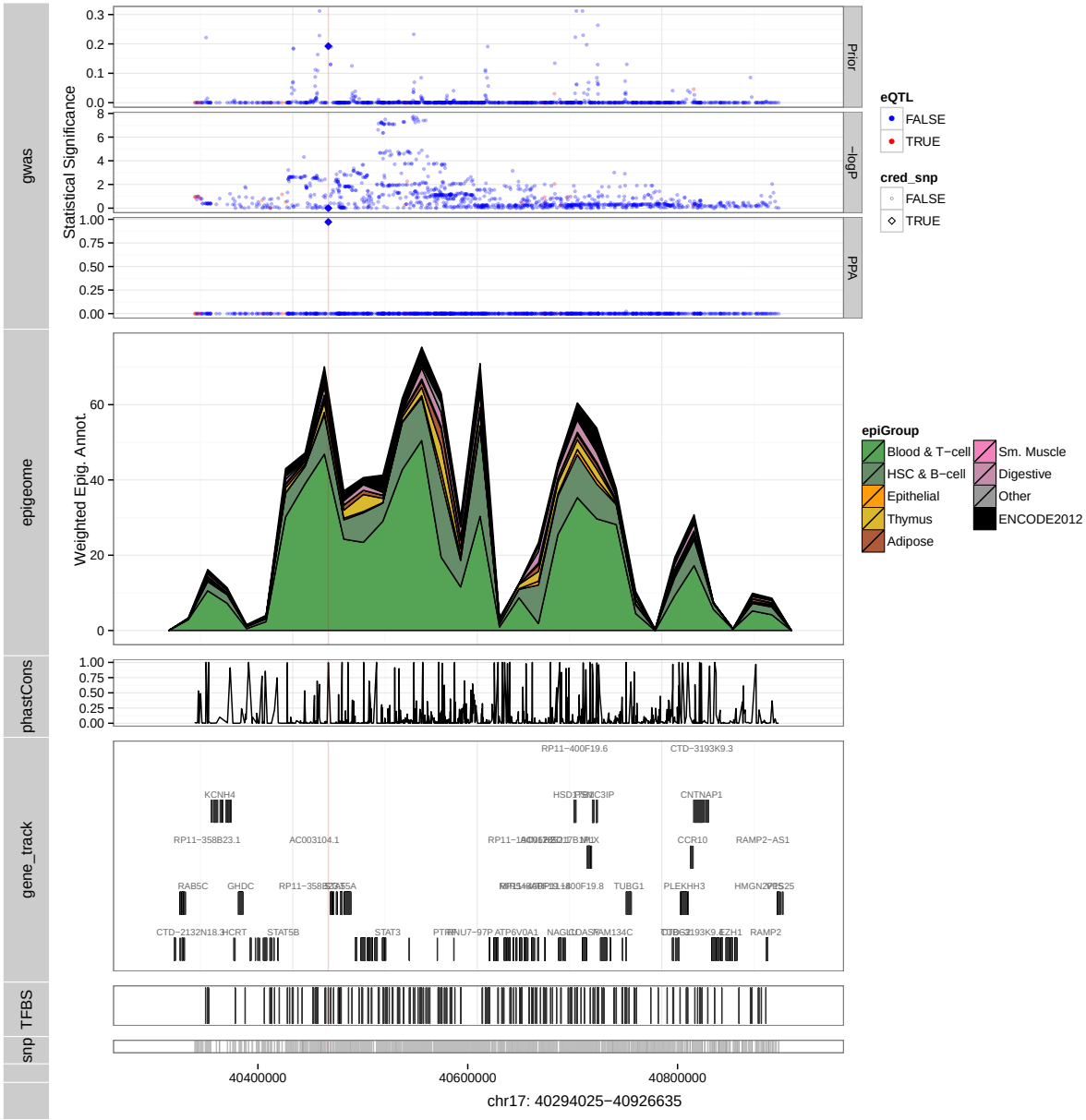
Multiple Sclerosis



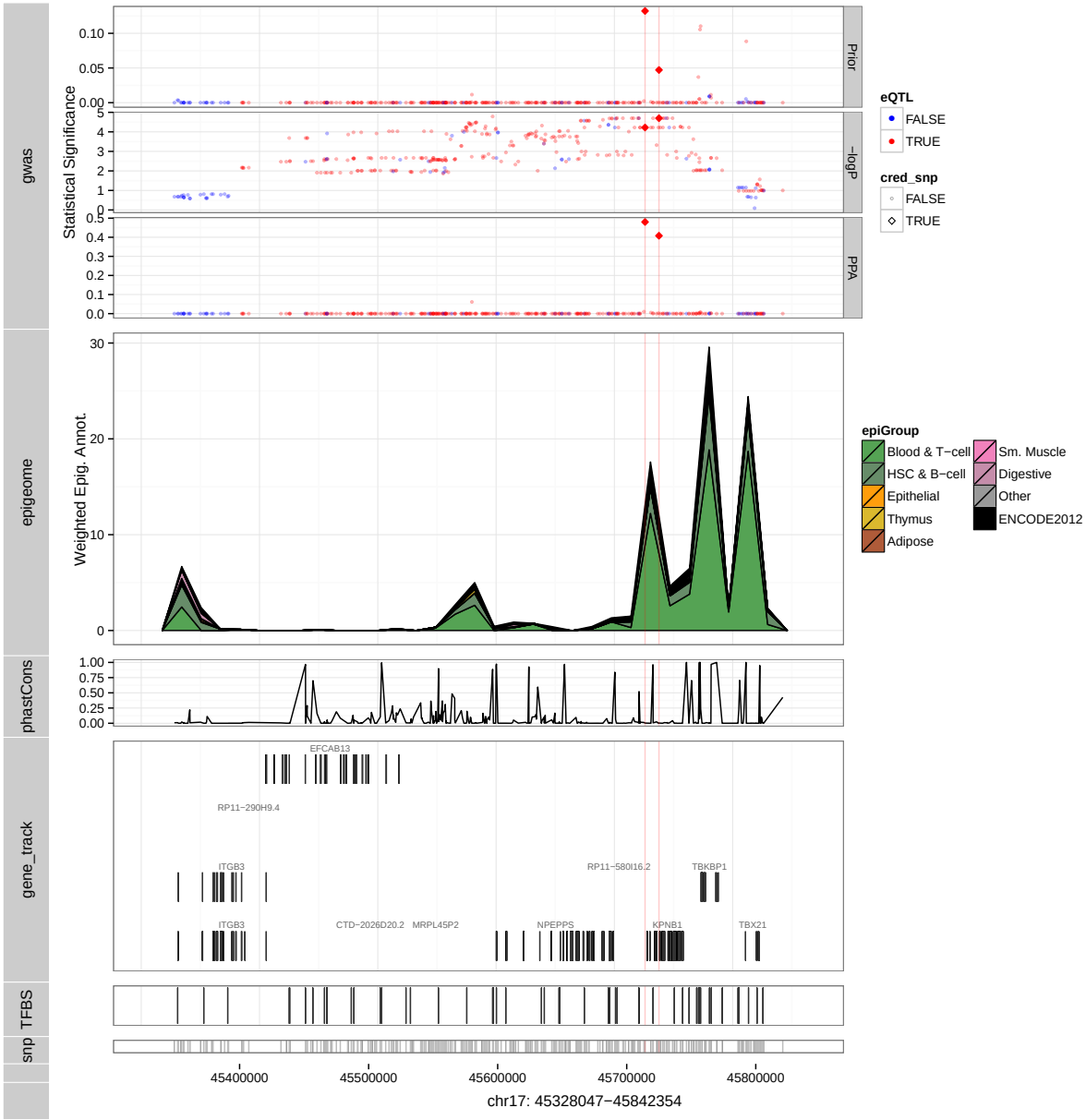
Multiple Sclerosis



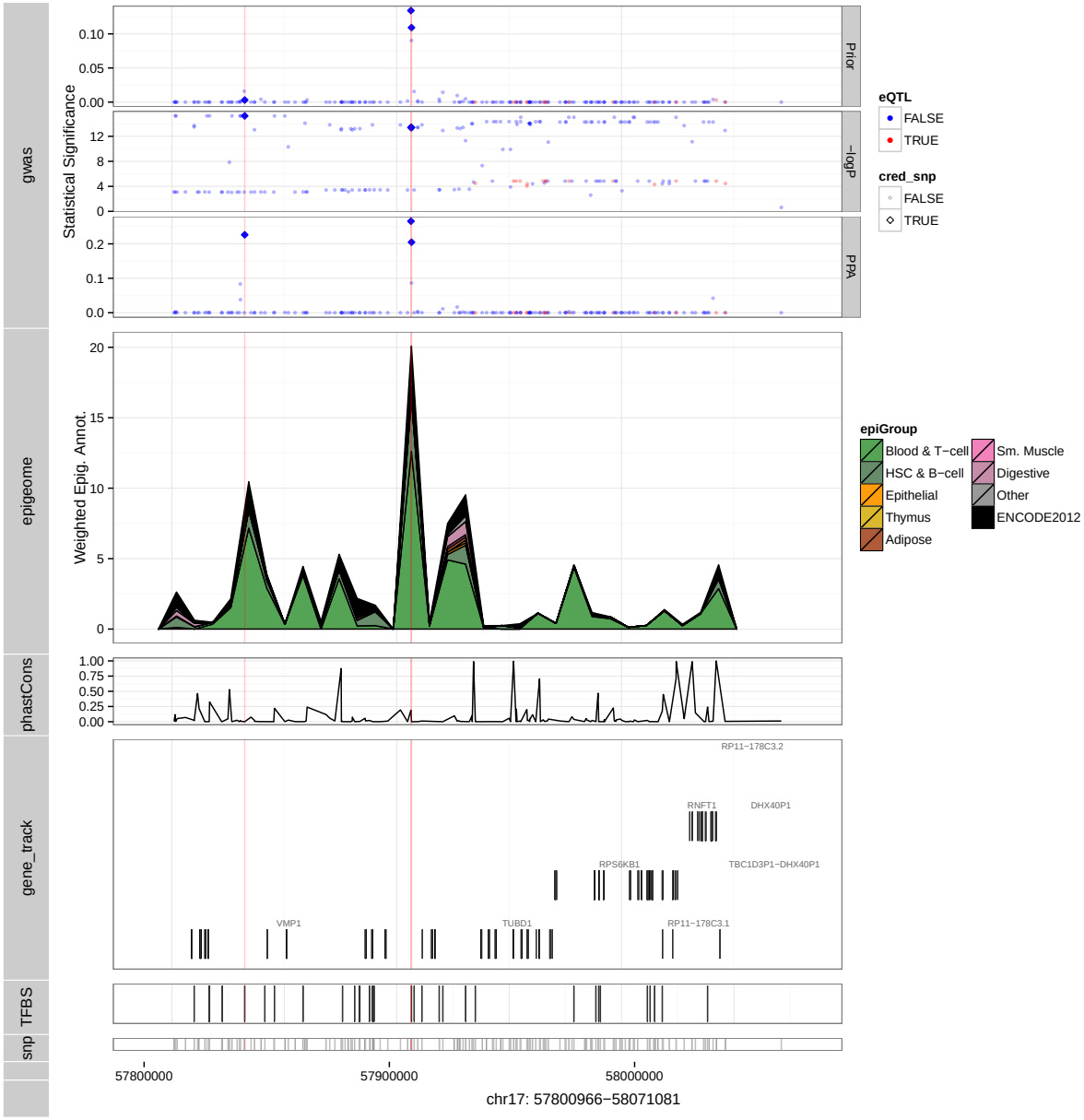
Multiple Sclerosis



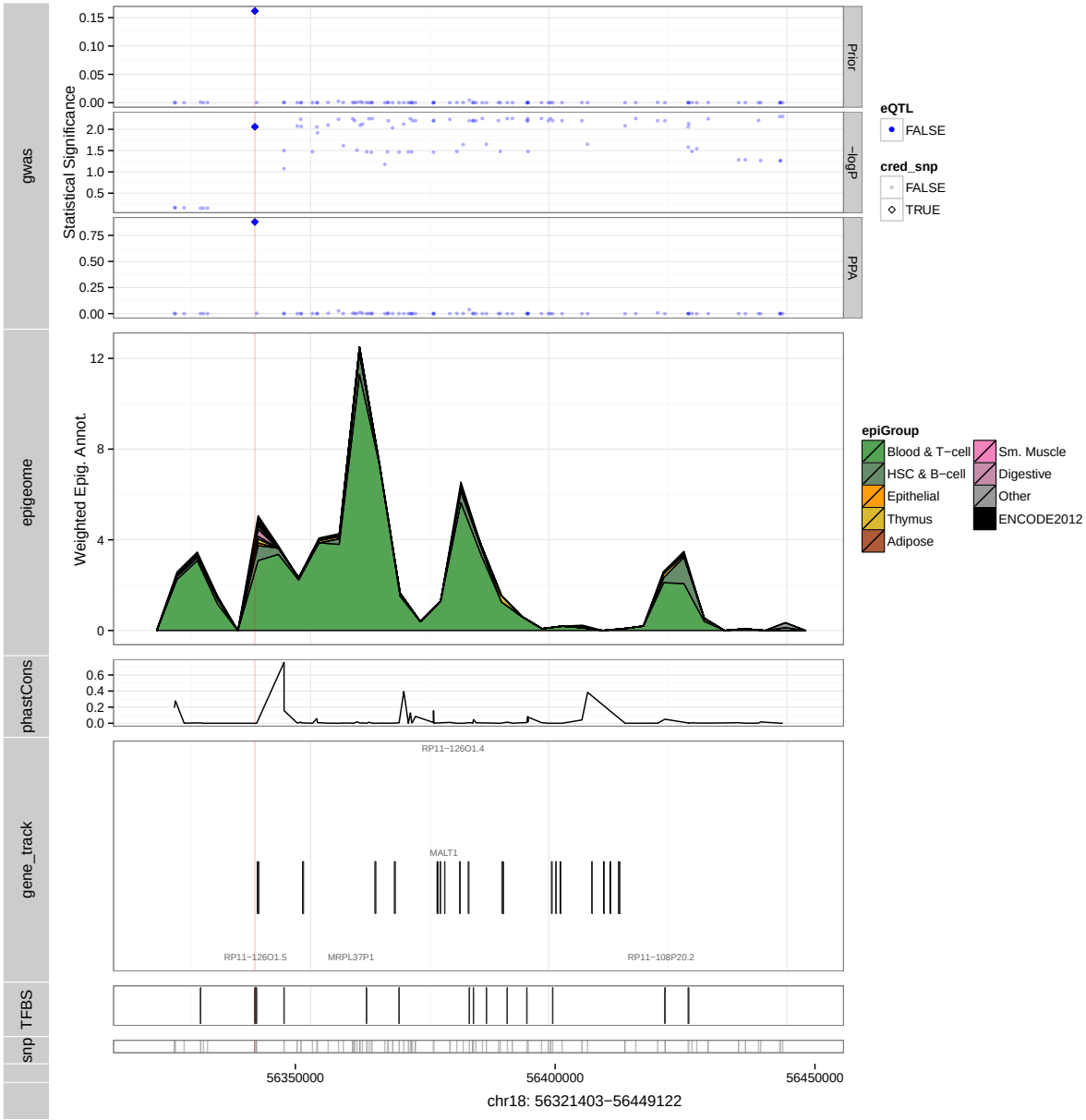
Multiple Sclerosis



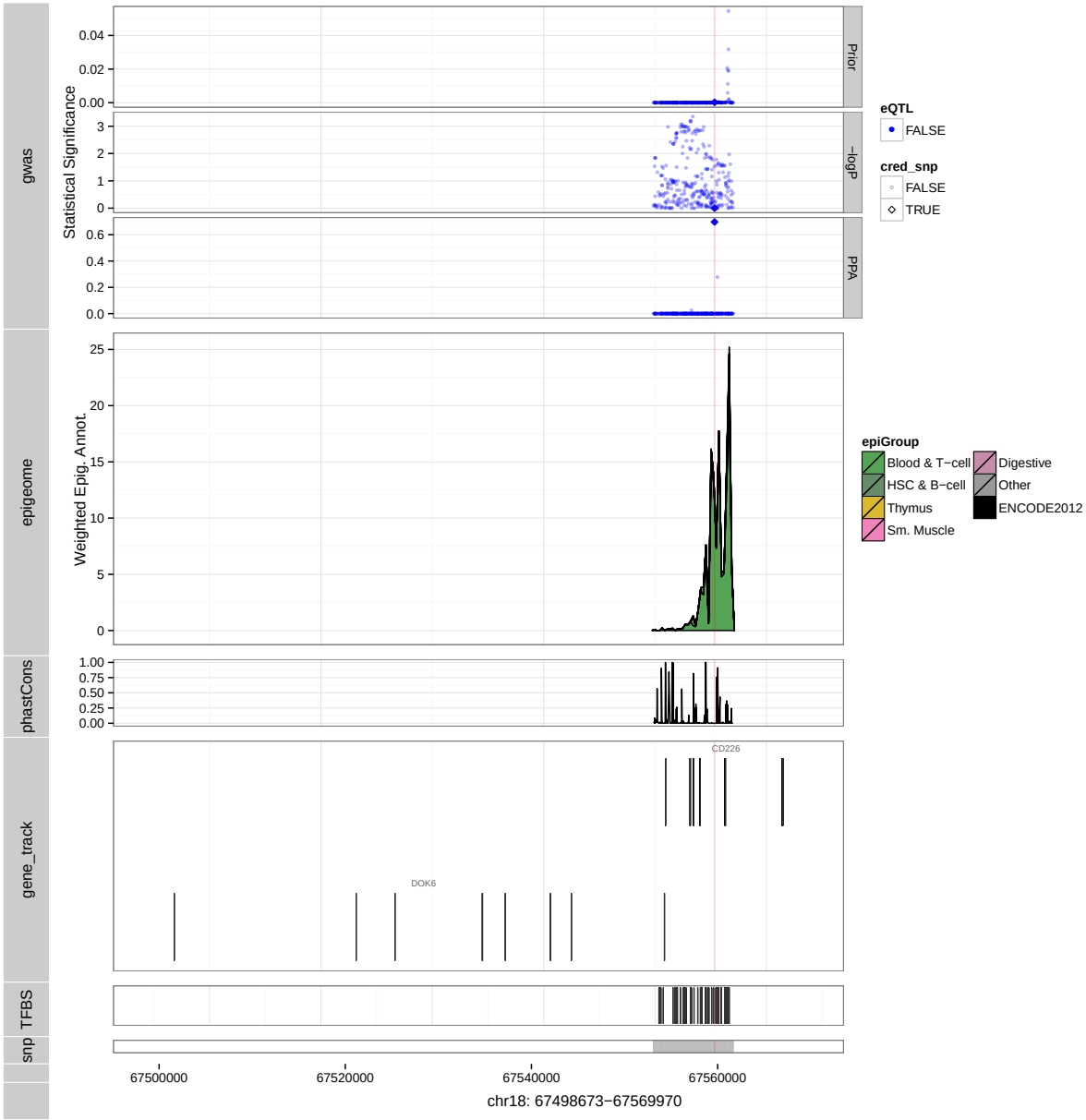
Multiple Sclerosis



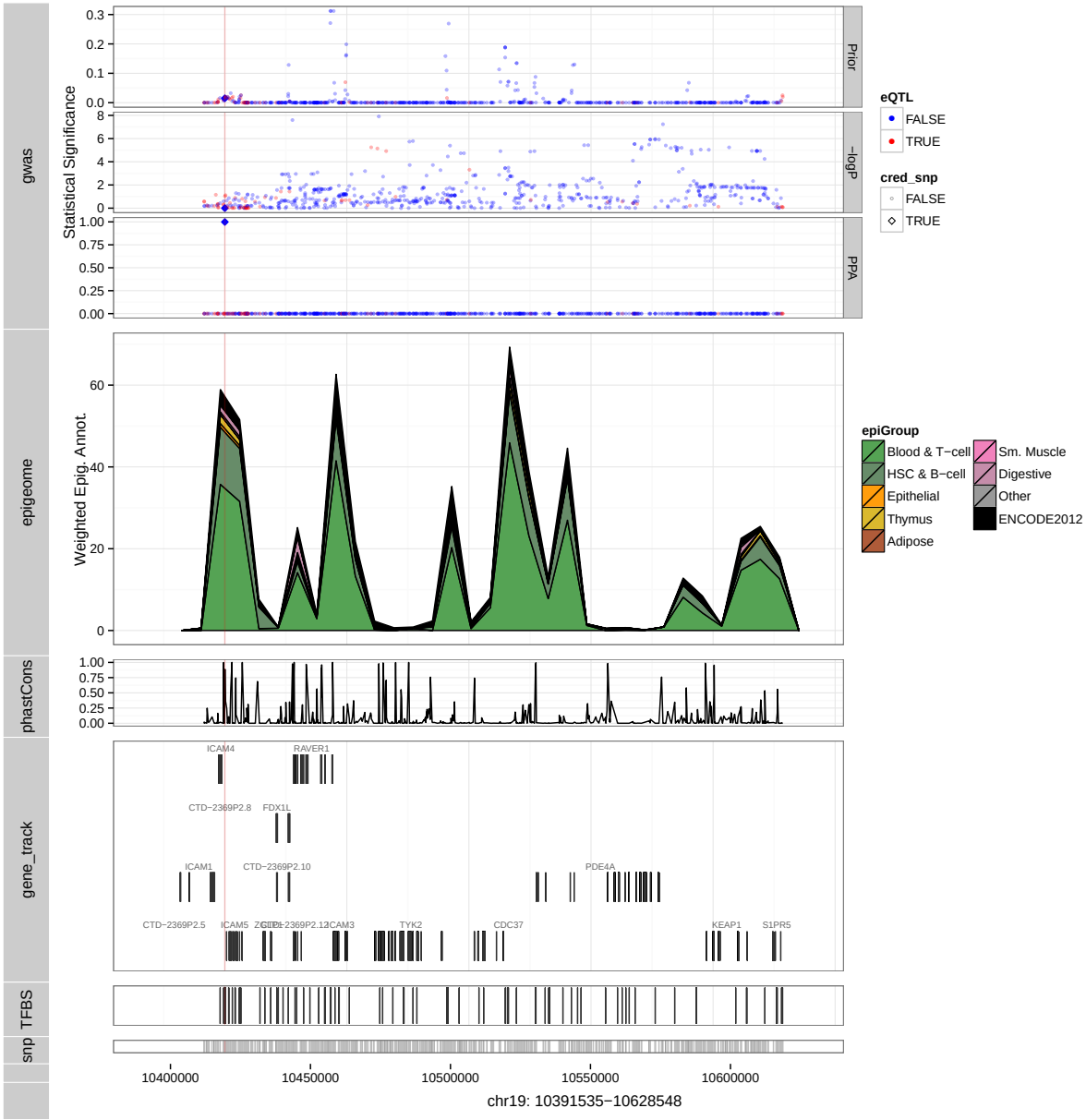
Multiple Sclerosis



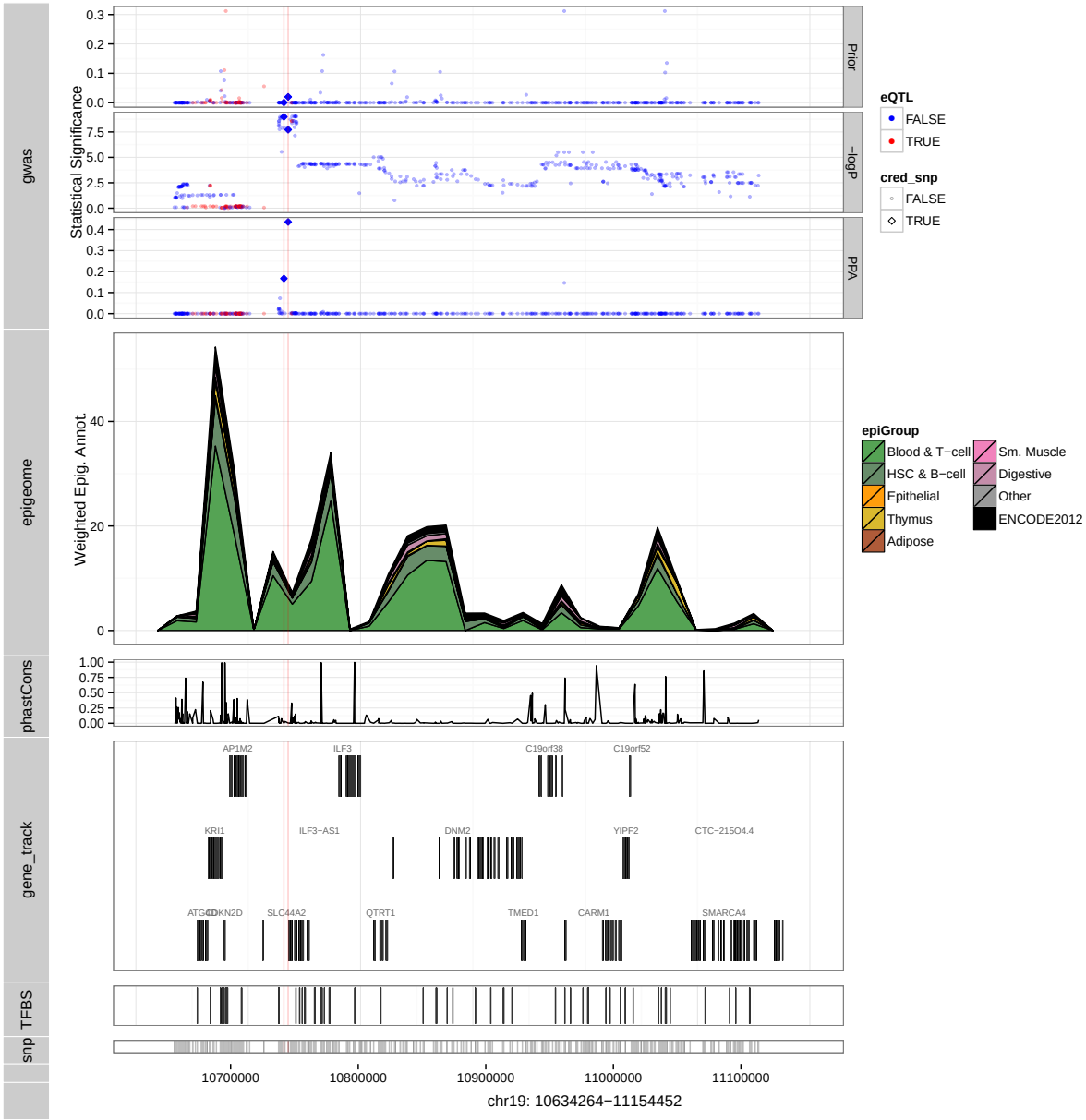
Multiple Sclerosis



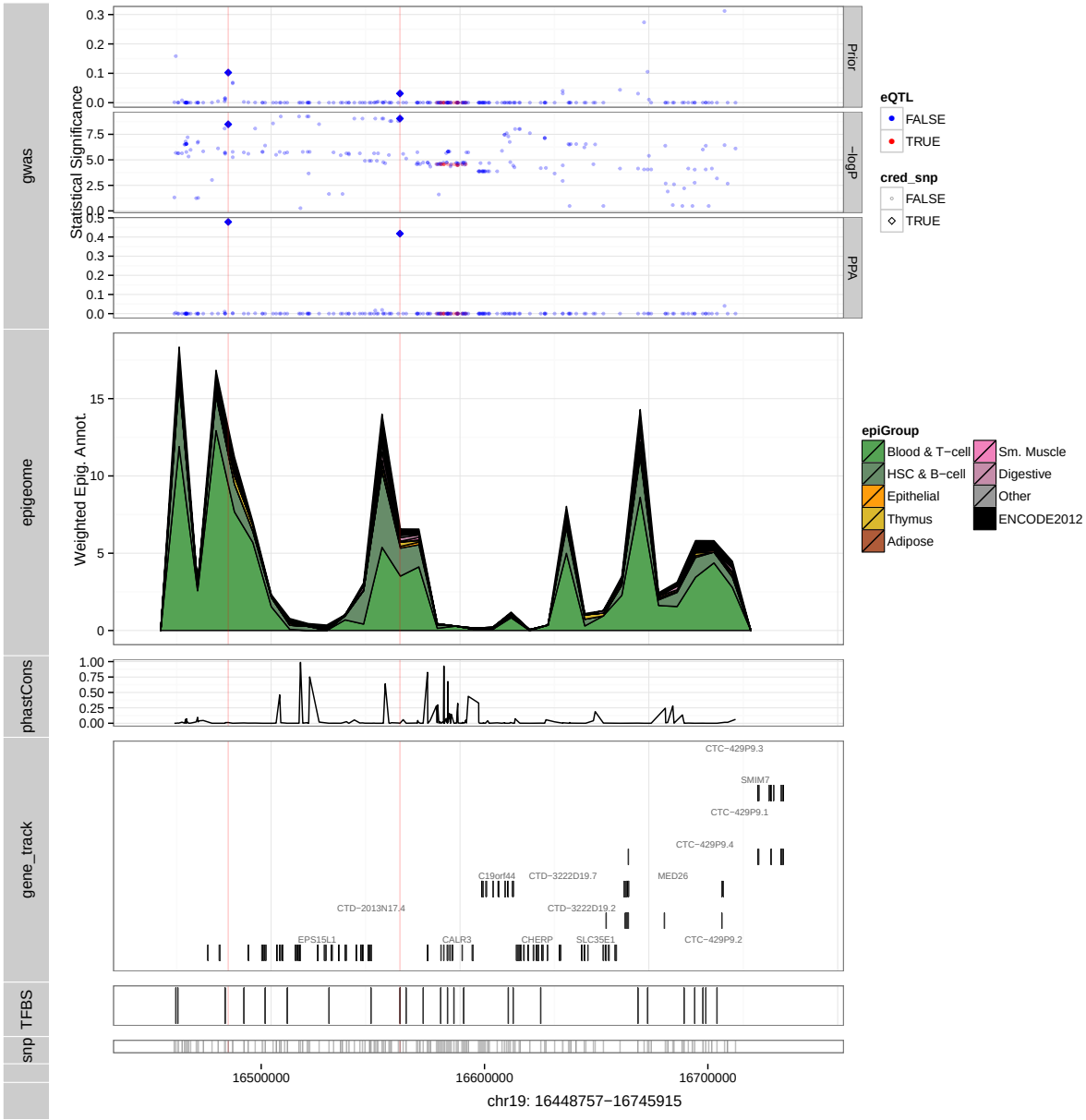
Multiple Sclerosis



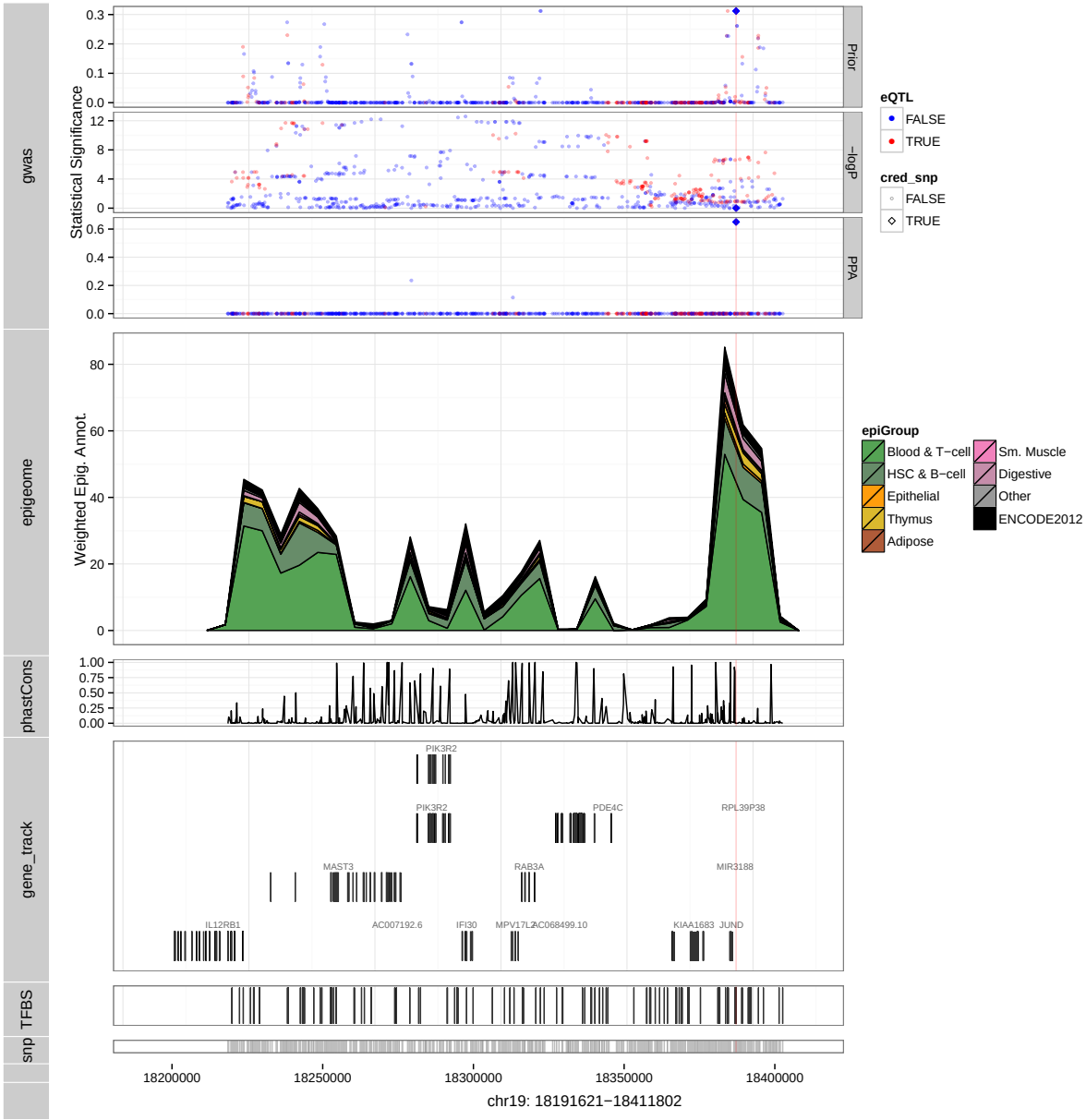
Multiple Sclerosis



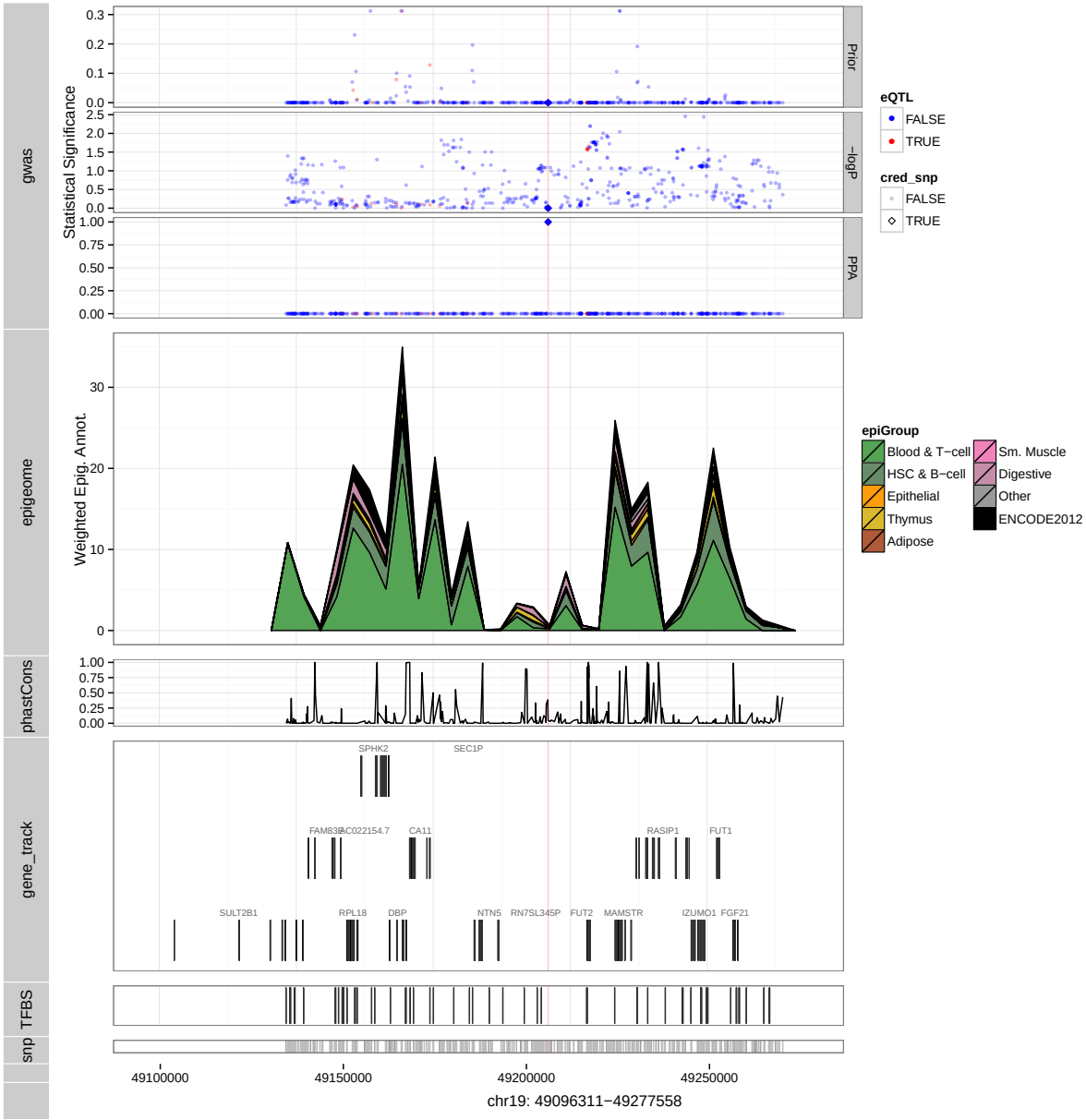
Multiple Sclerosis



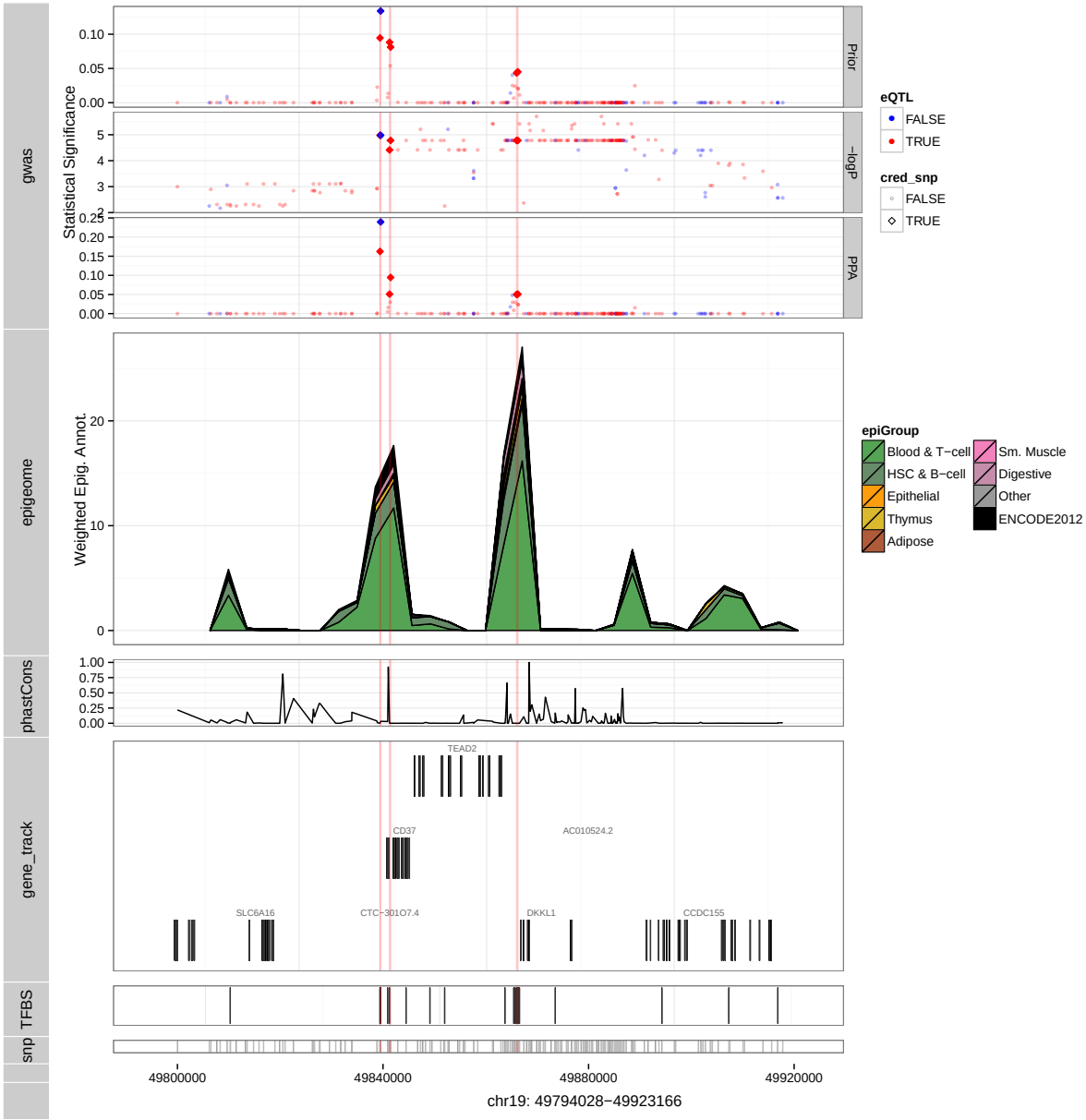
Multiple Sclerosis



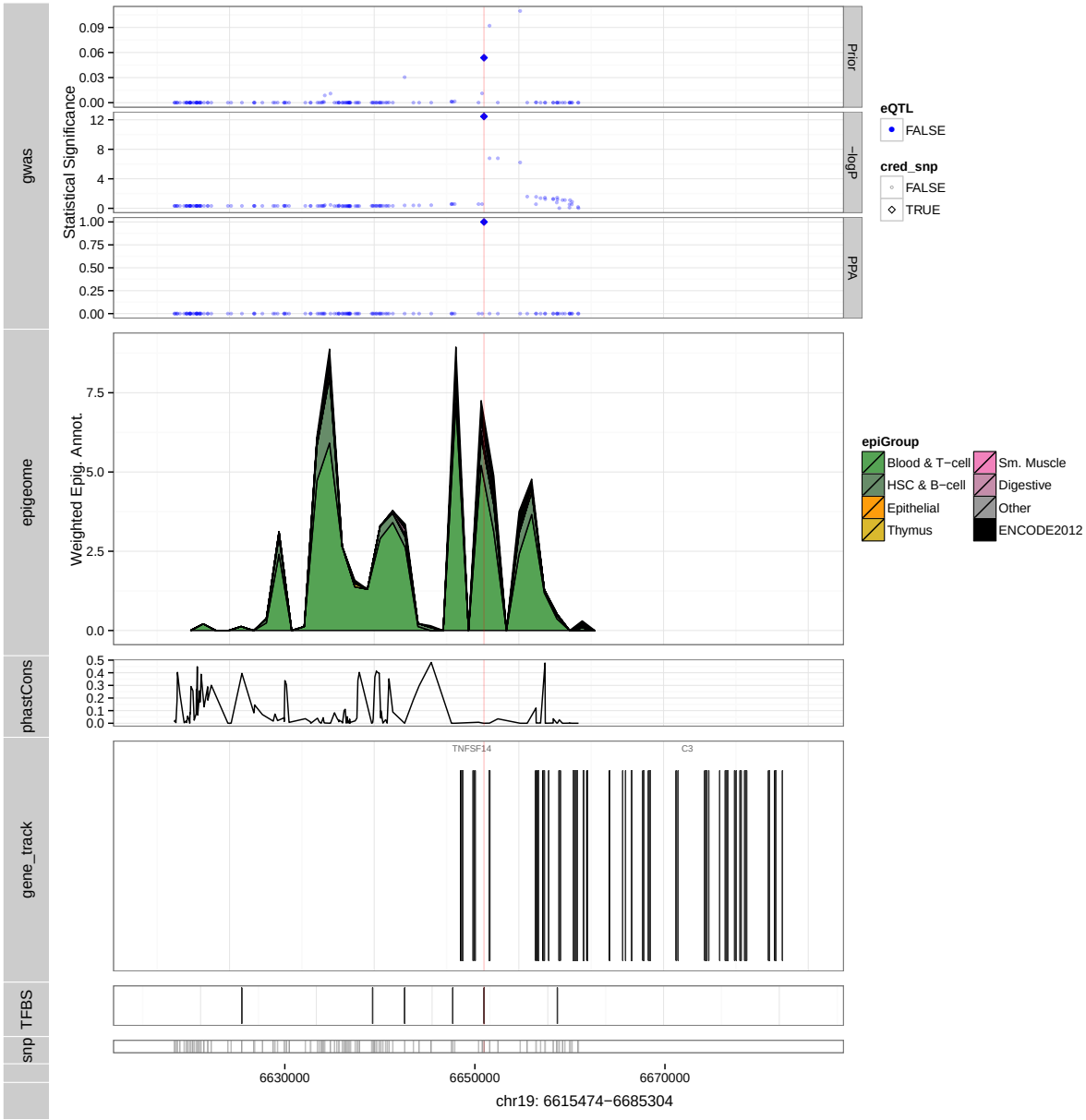
Multiple Sclerosis



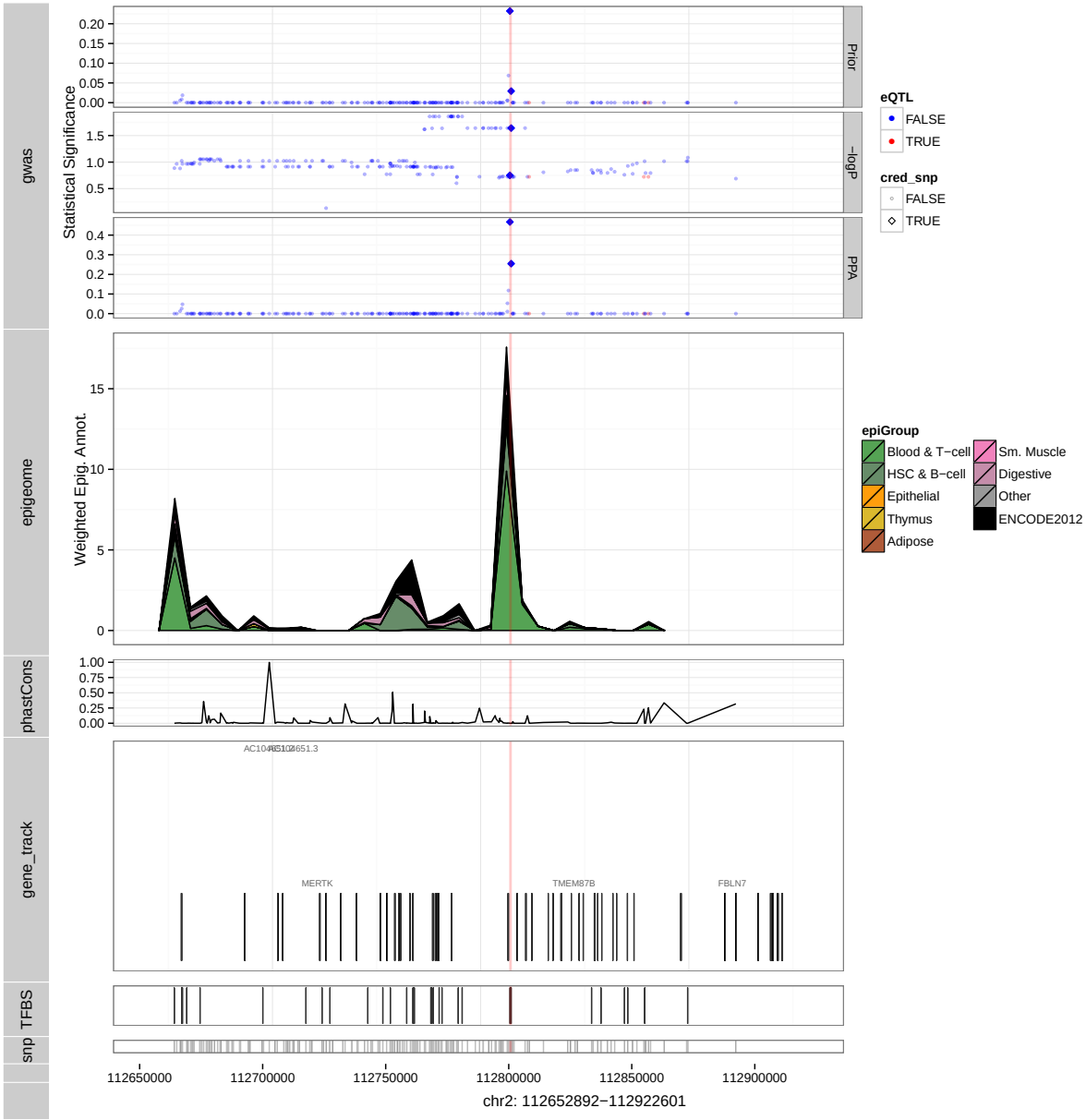
Multiple Sclerosis



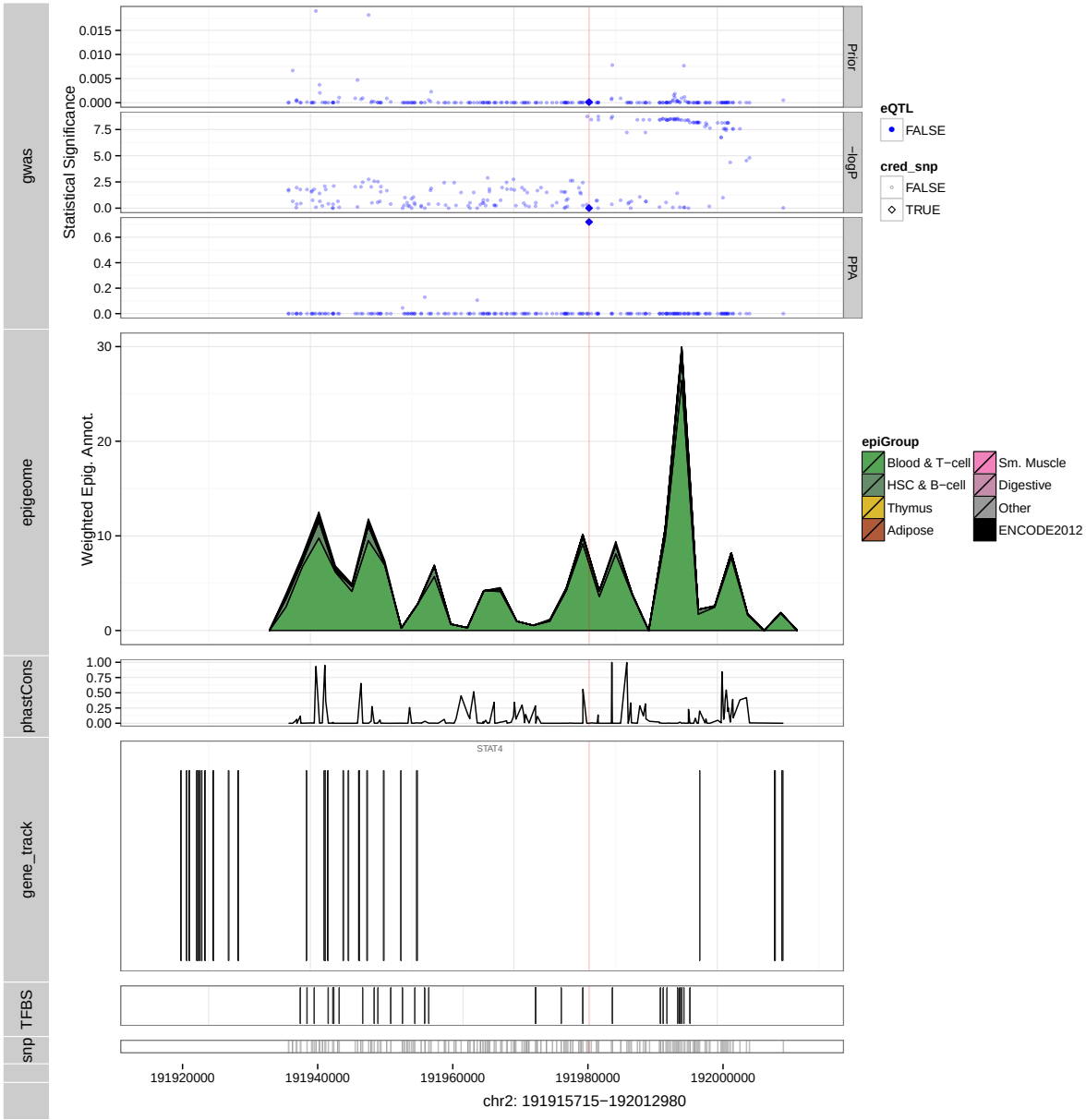
Multiple Sclerosis



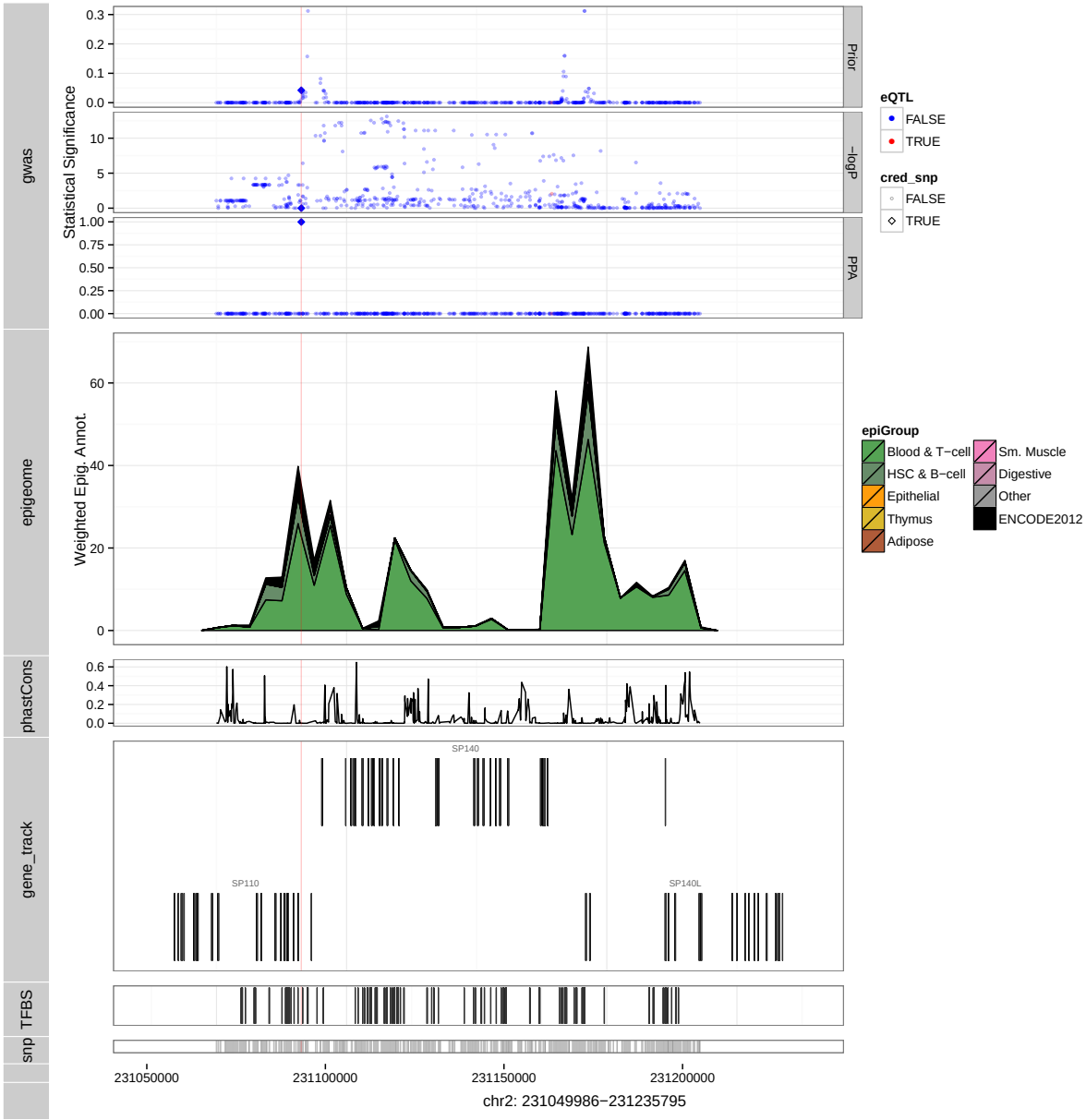
Multiple Sclerosis



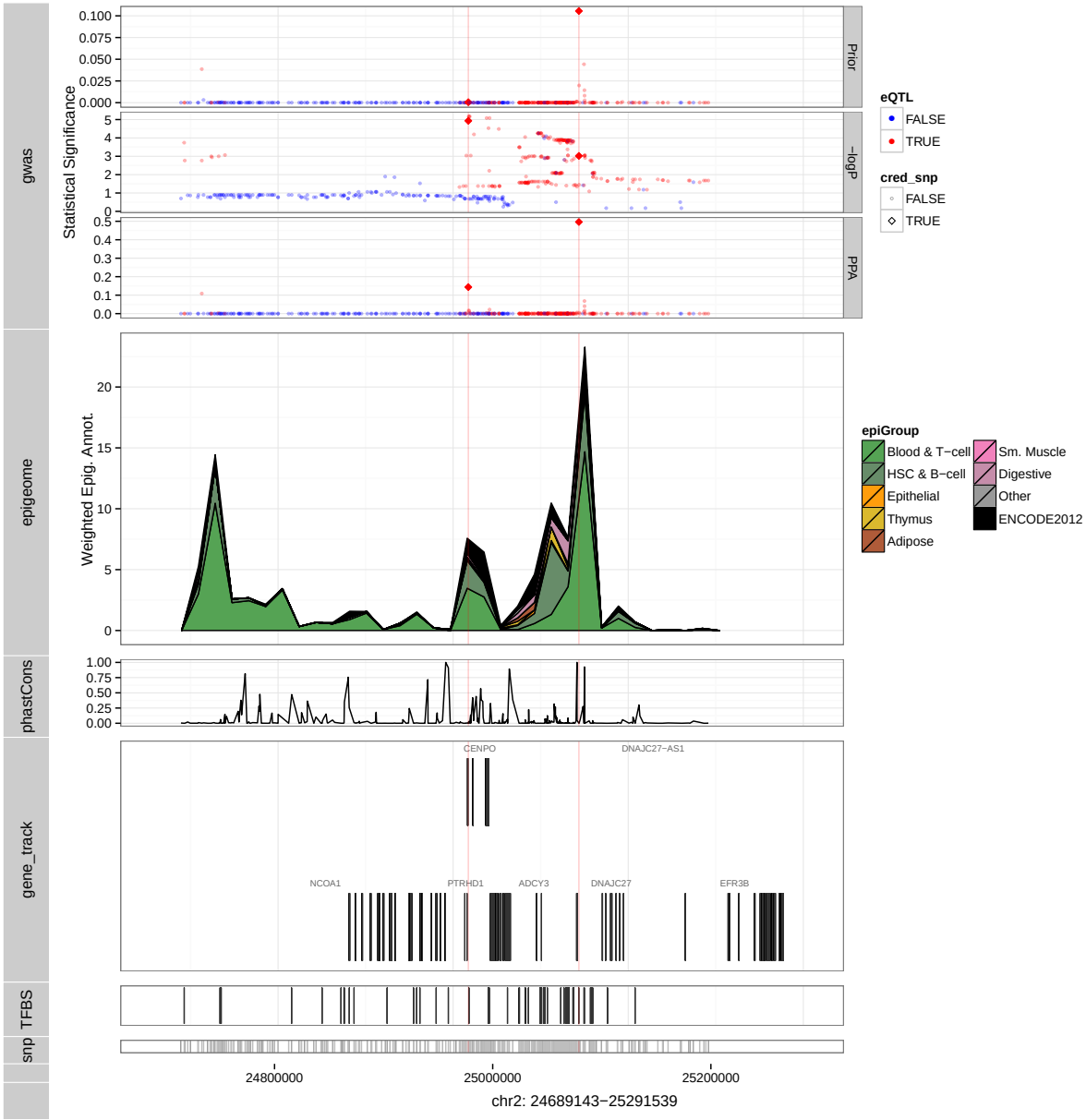
Multiple Sclerosis



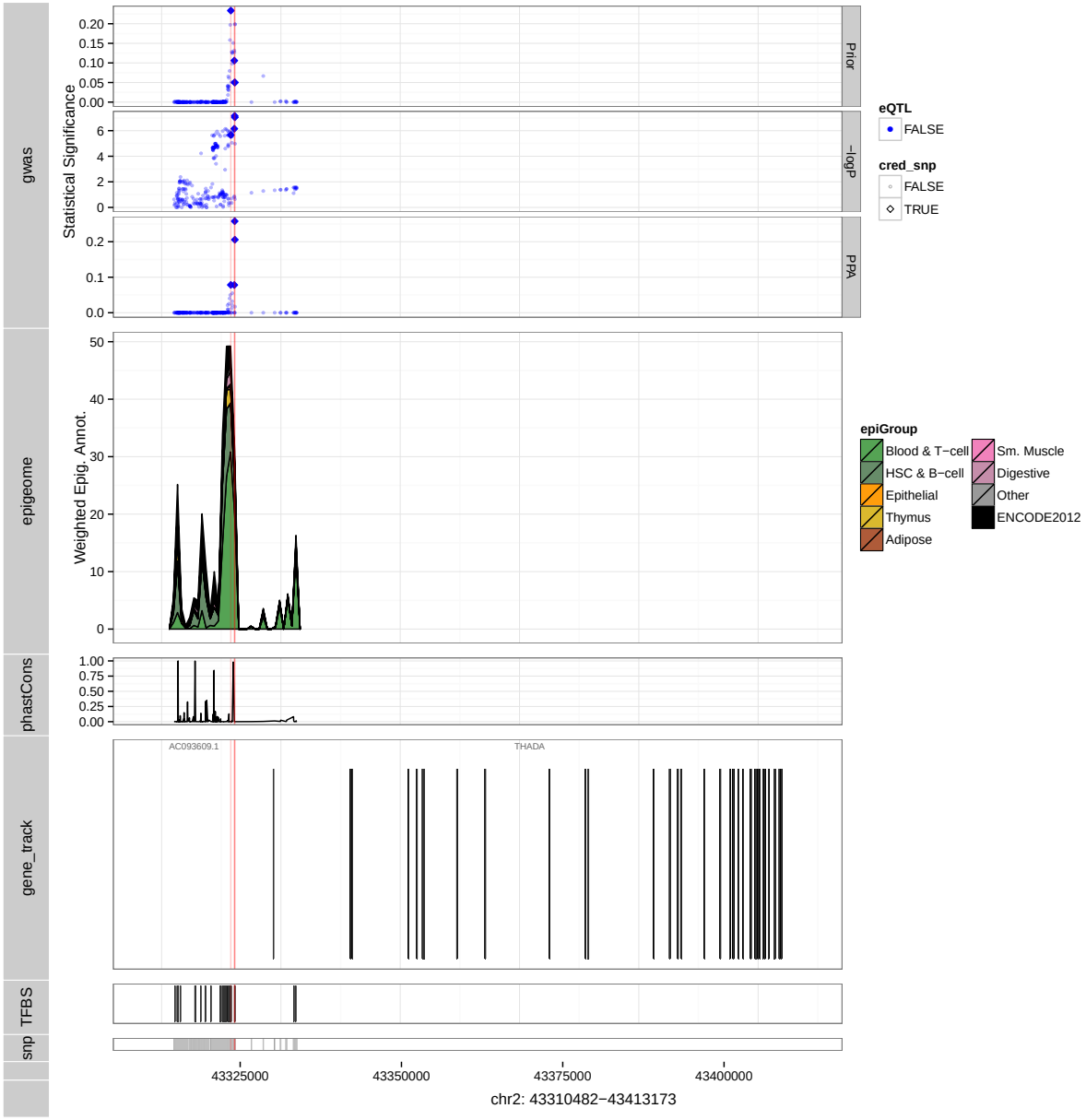
Multiple Sclerosis



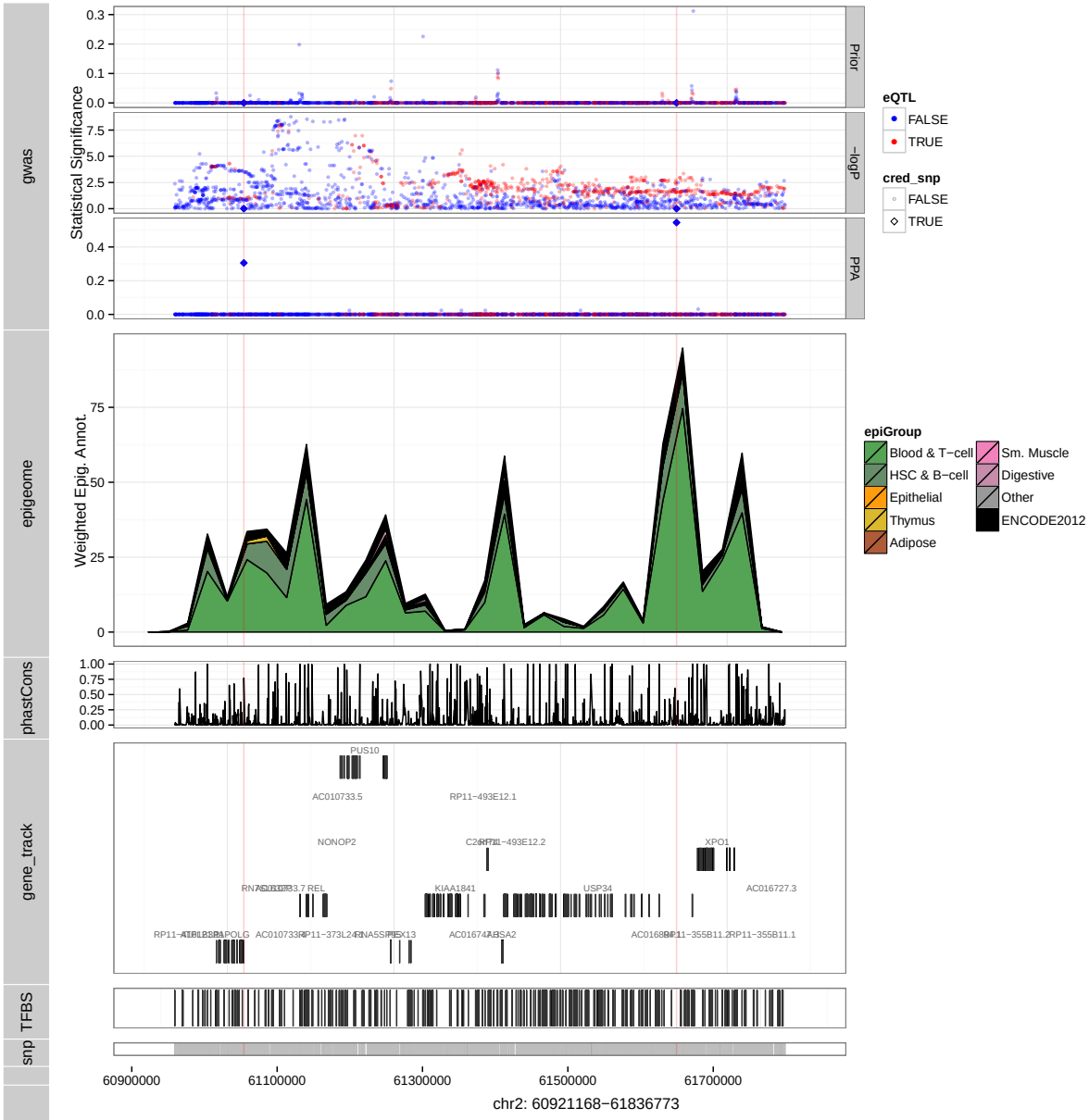
Multiple Sclerosis



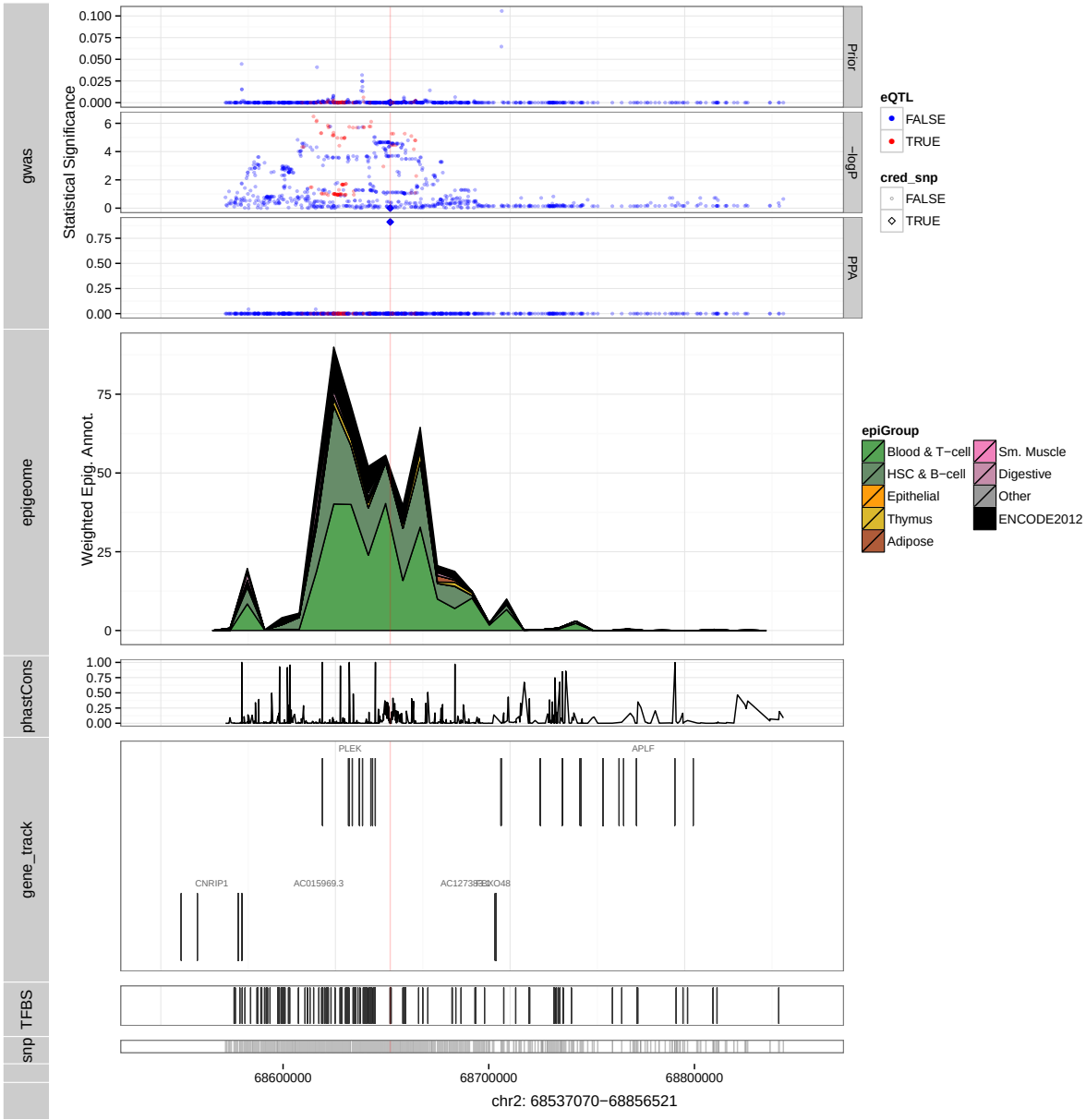
Multiple Sclerosis



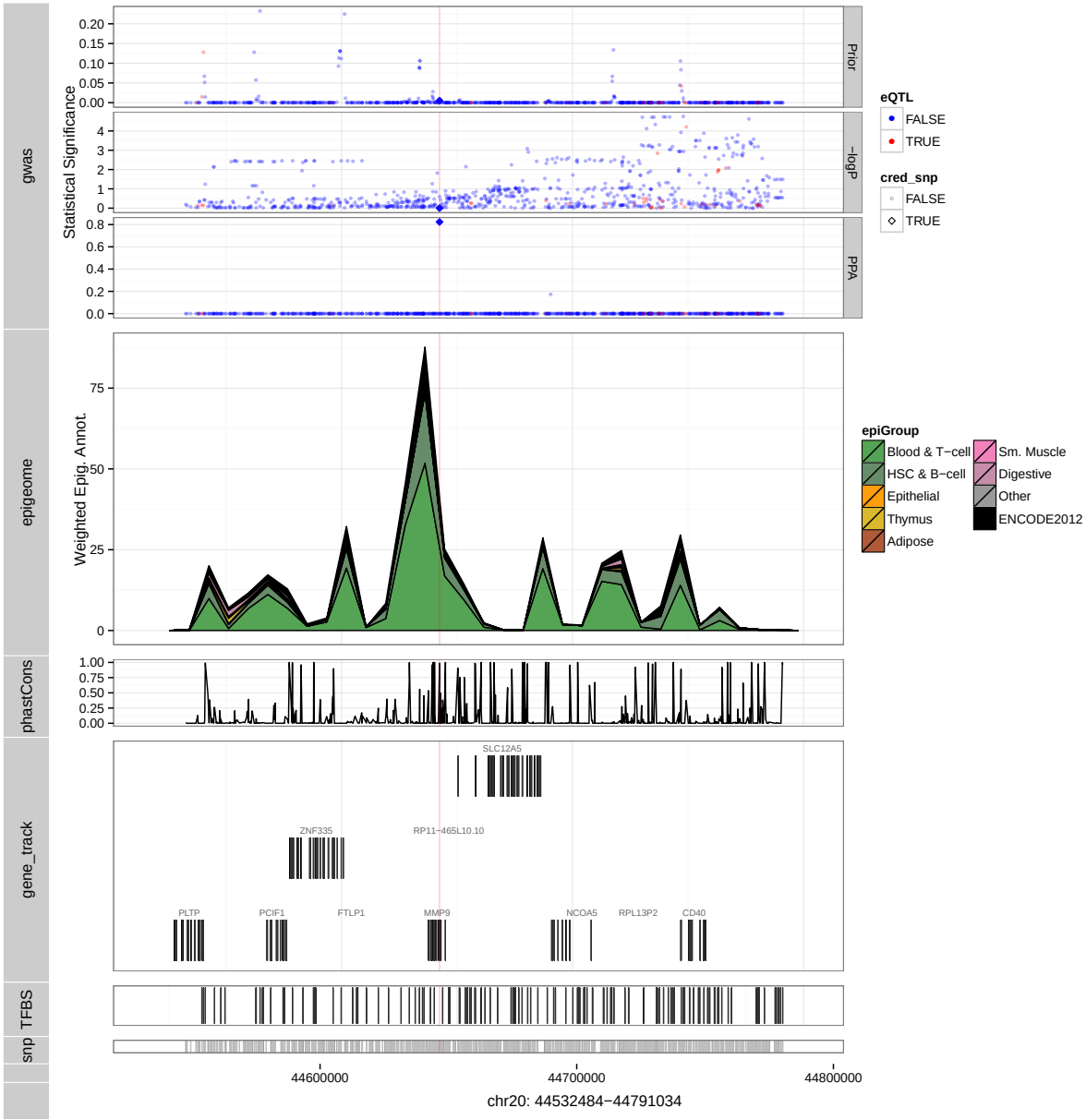
Multiple Sclerosis



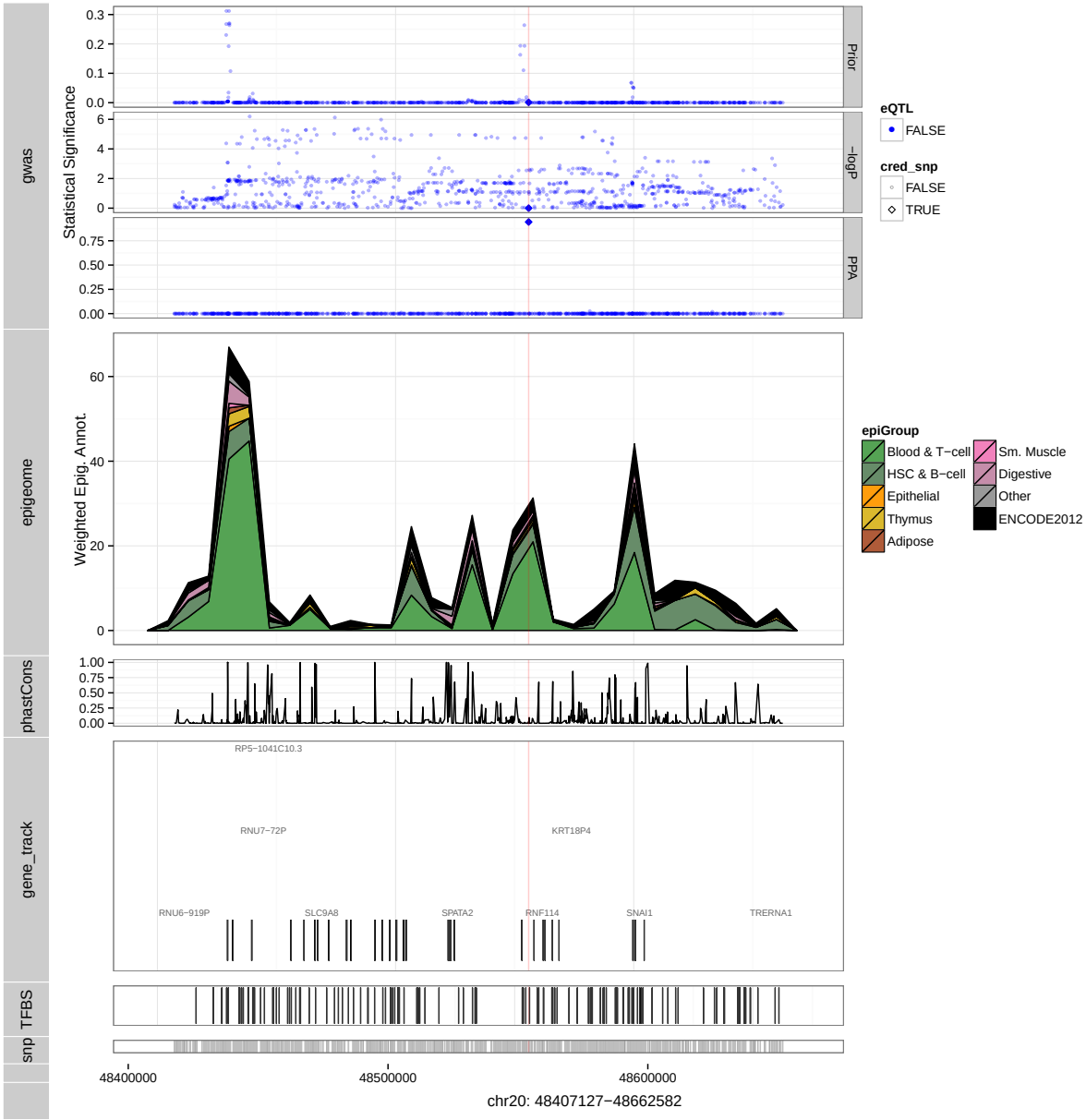
Multiple Sclerosis



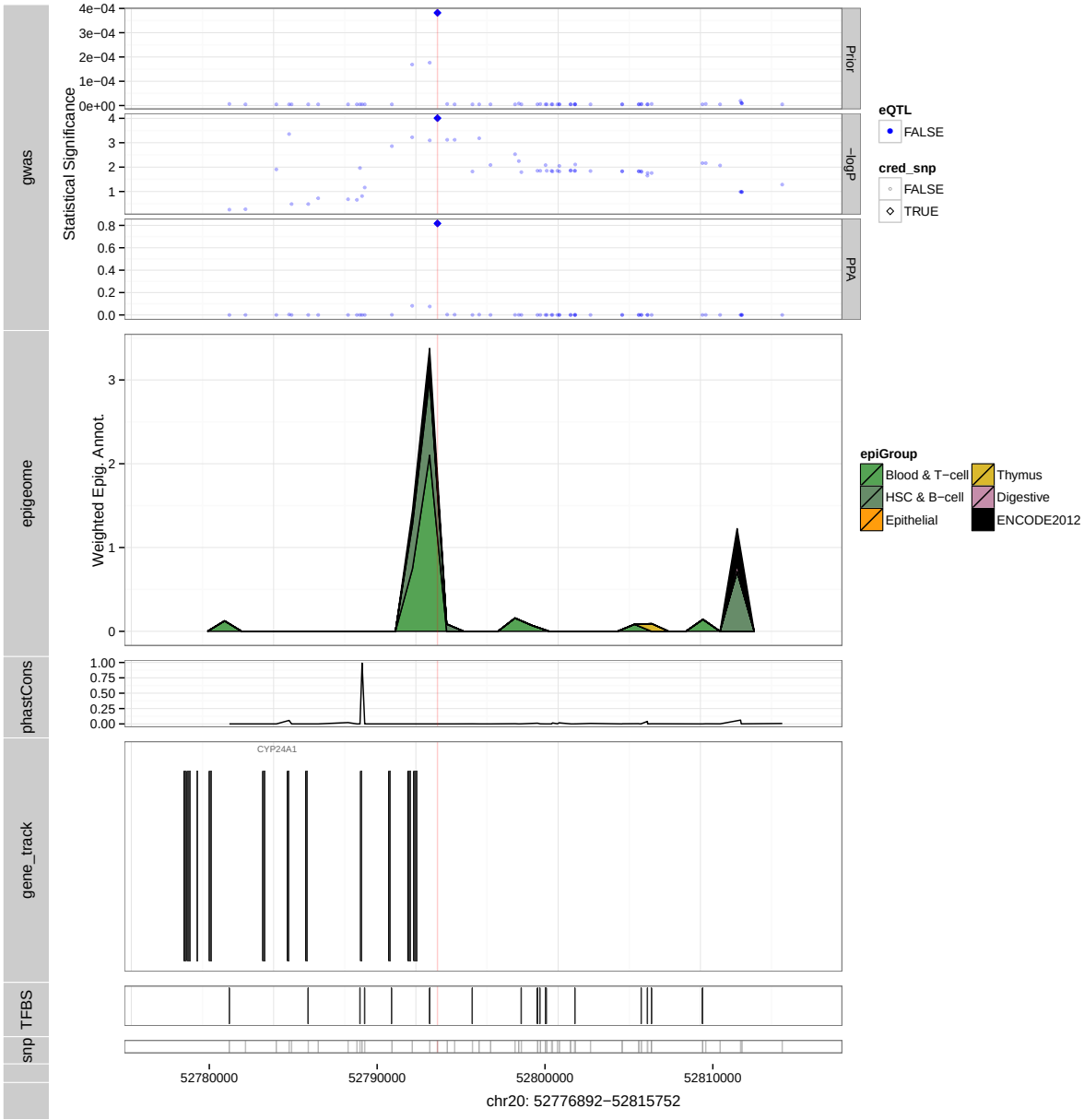
Multiple Sclerosis



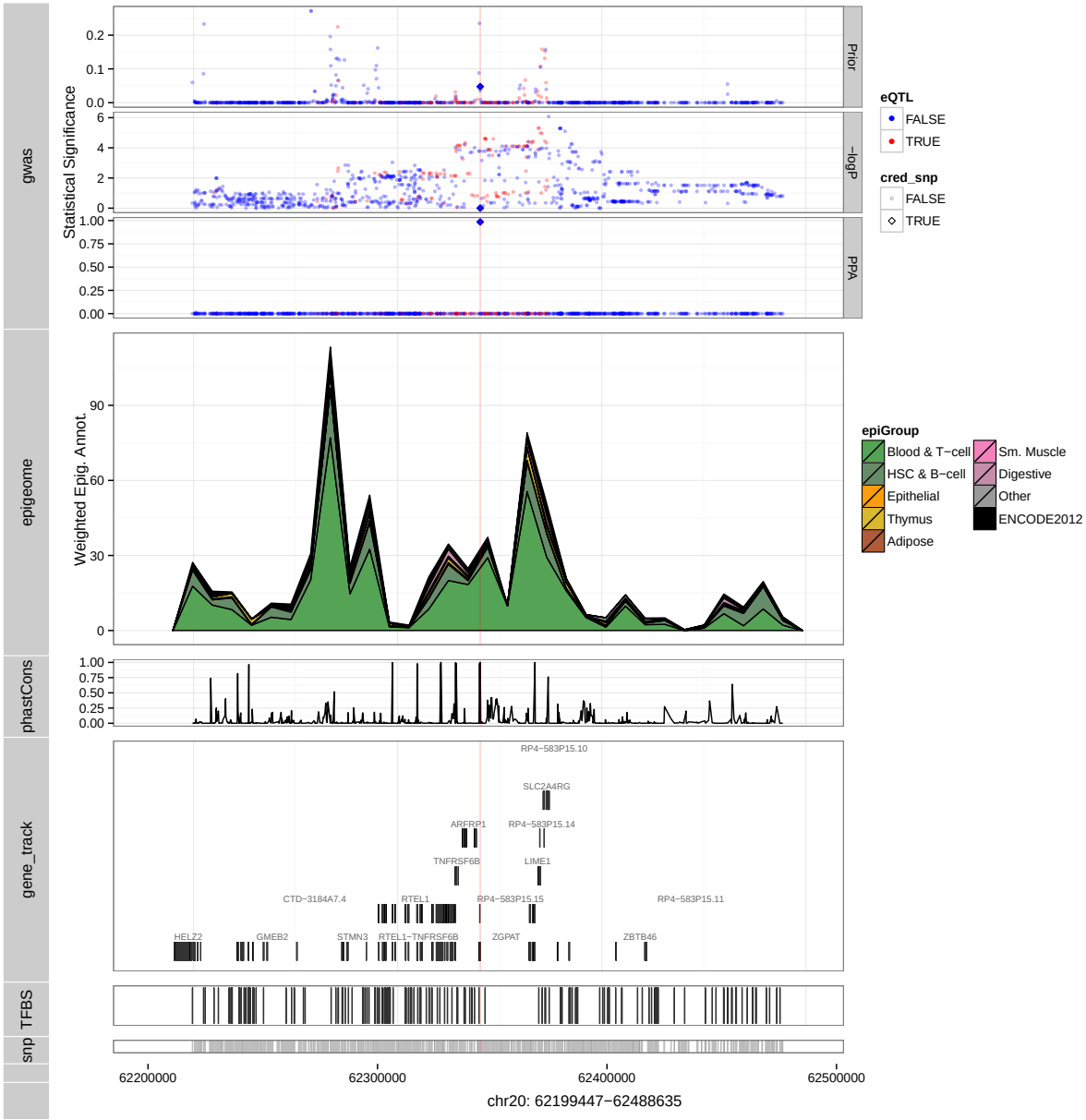
Multiple Sclerosis



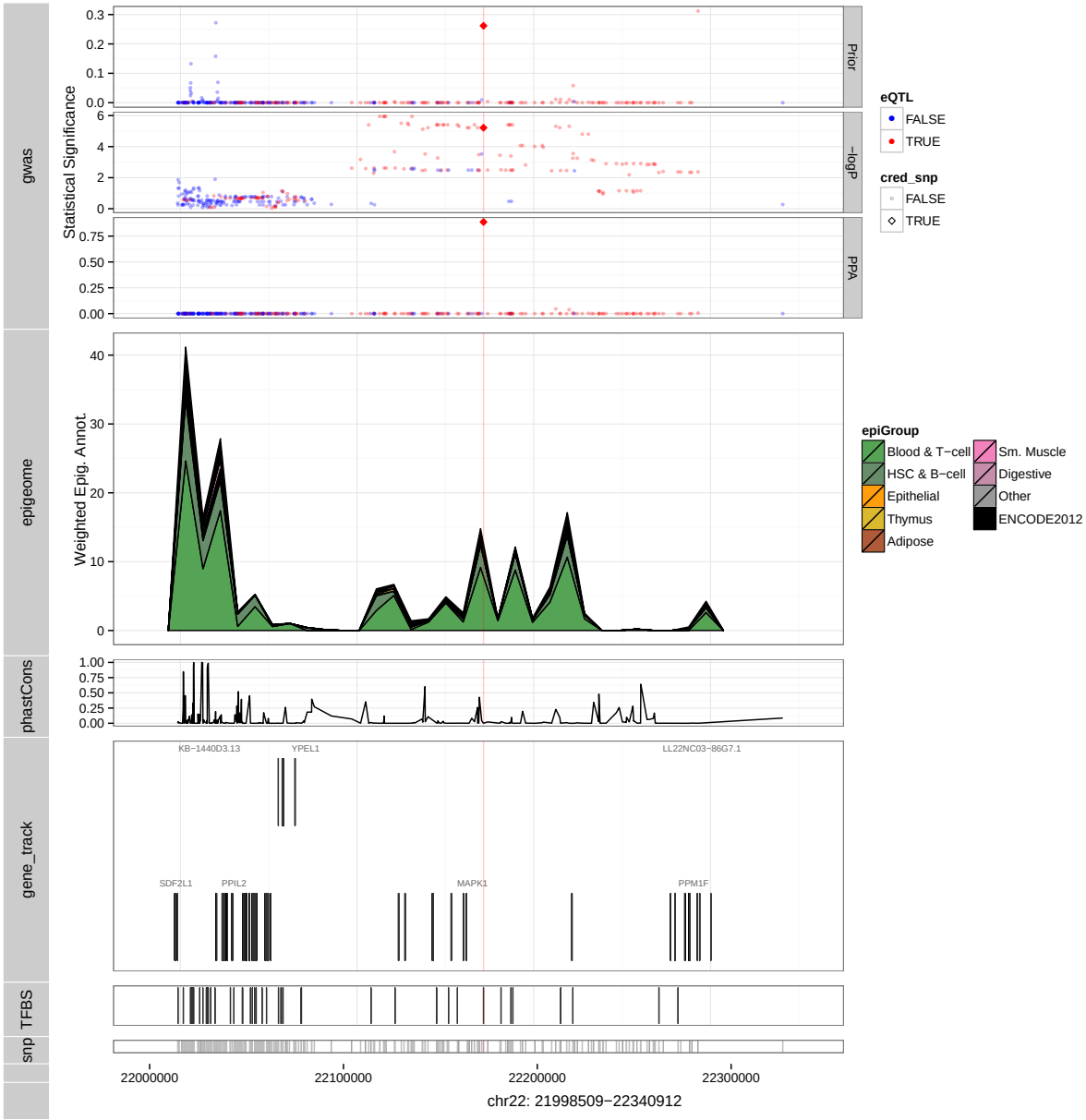
Multiple Sclerosis



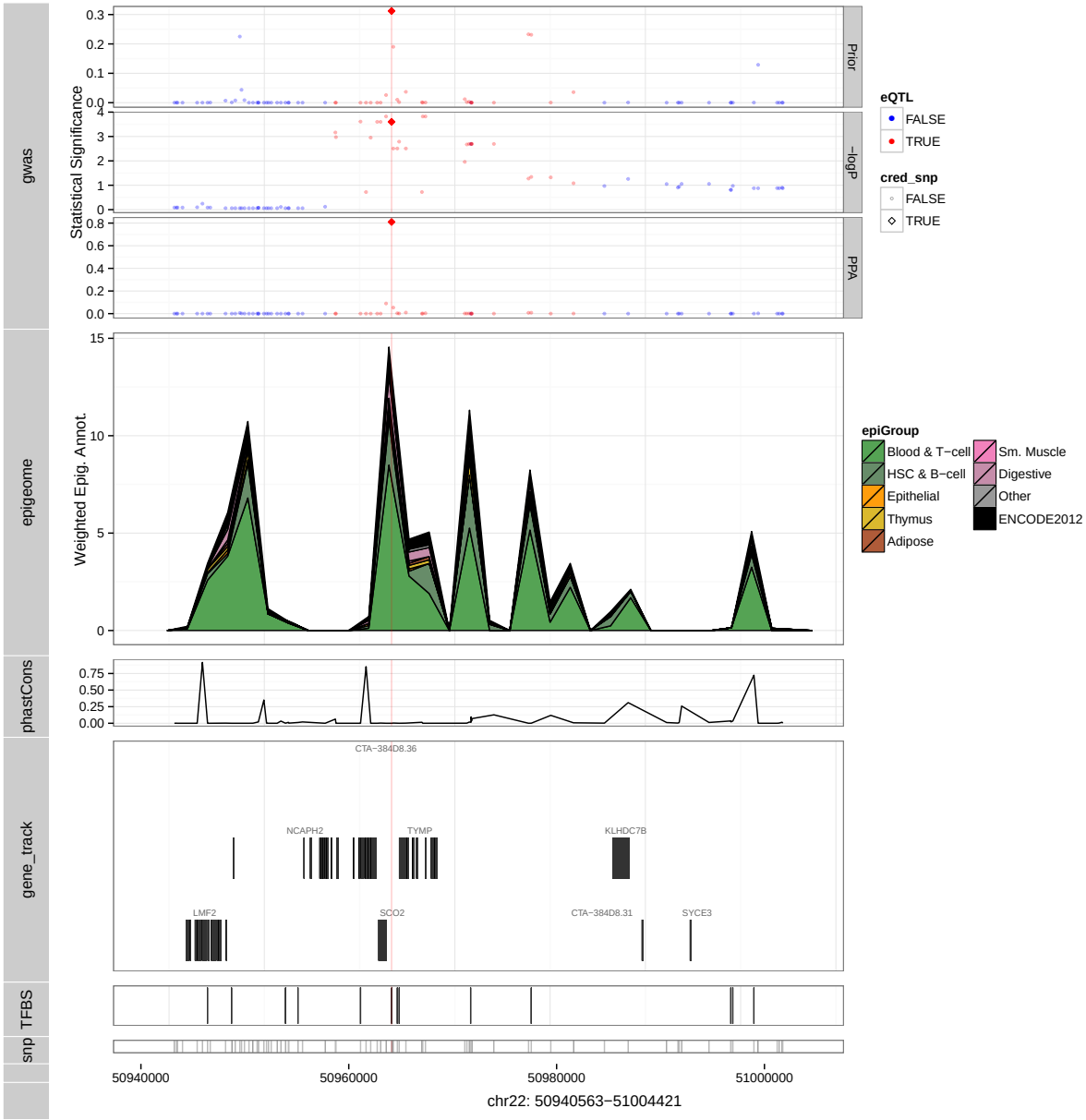
Multiple Sclerosis



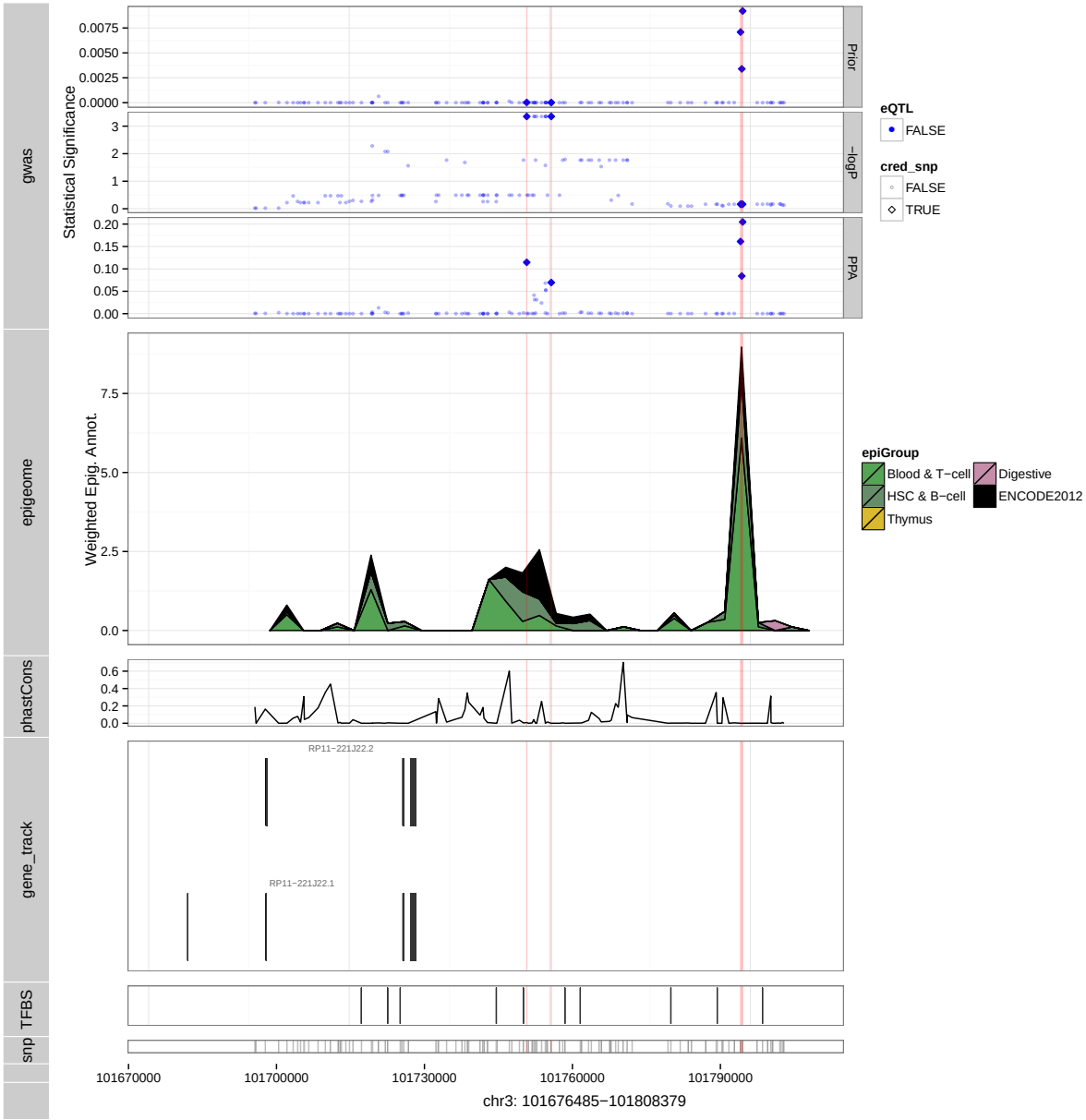
Multiple Sclerosis



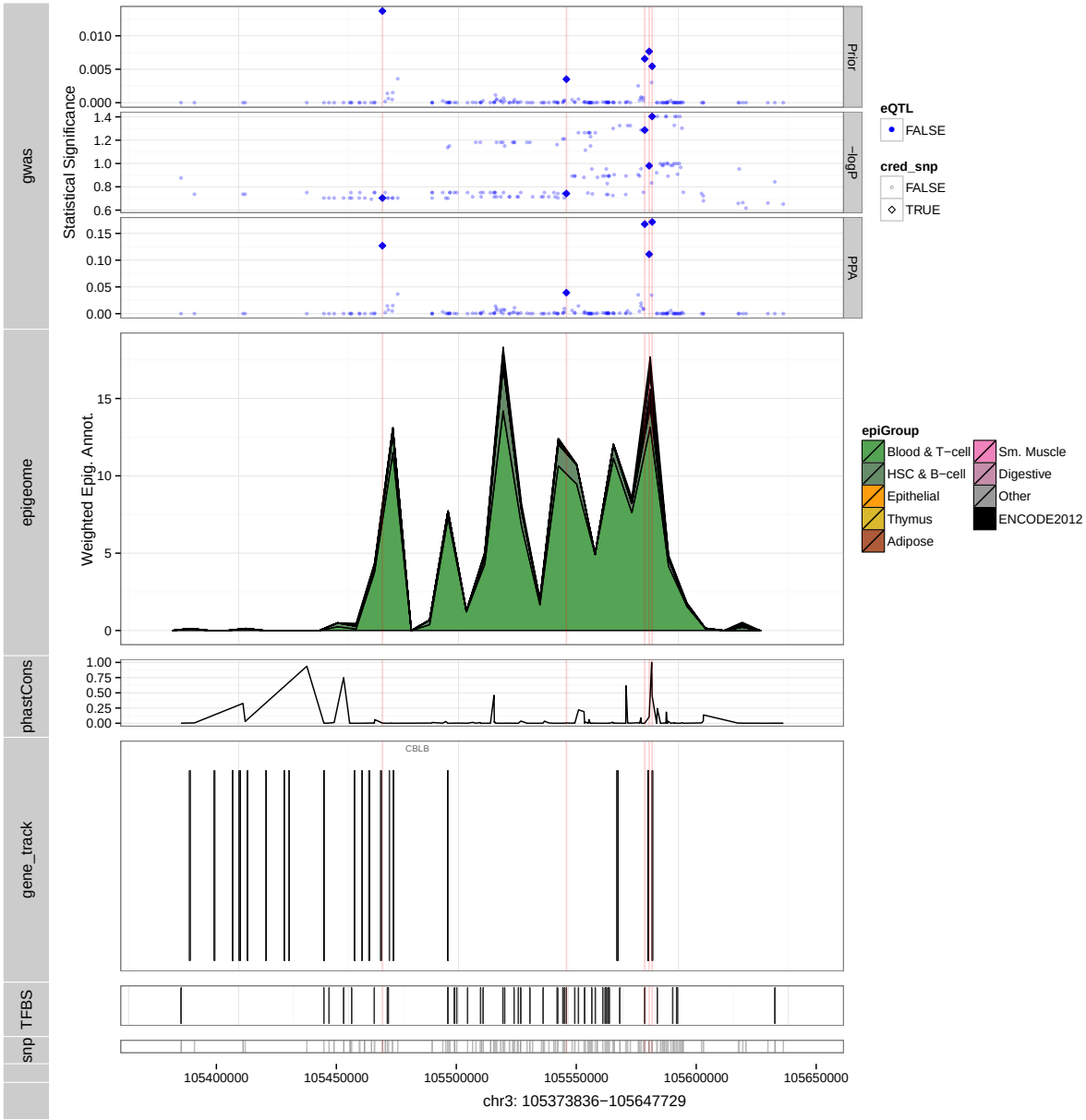
Multiple Sclerosis



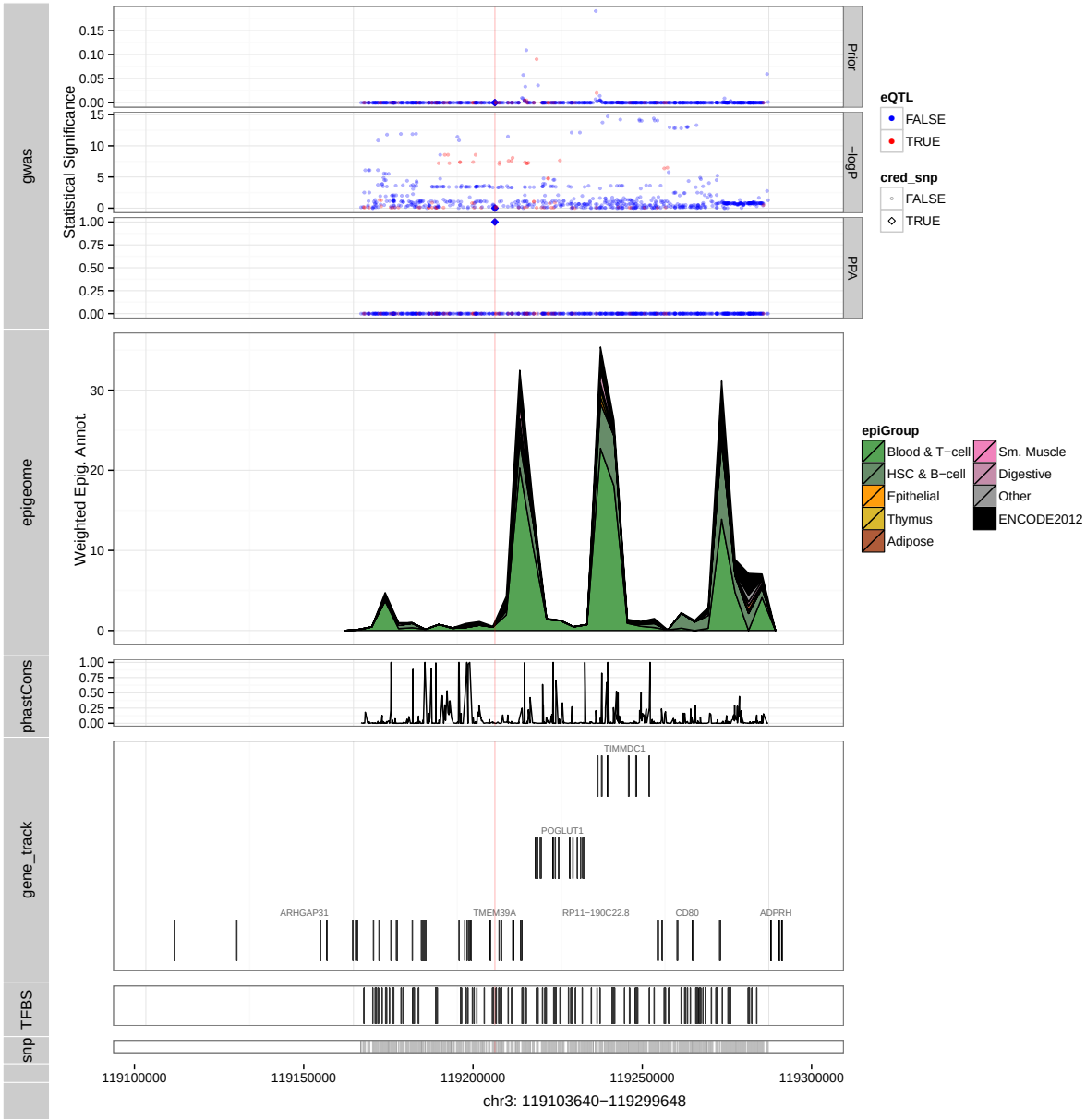
Multiple Sclerosis



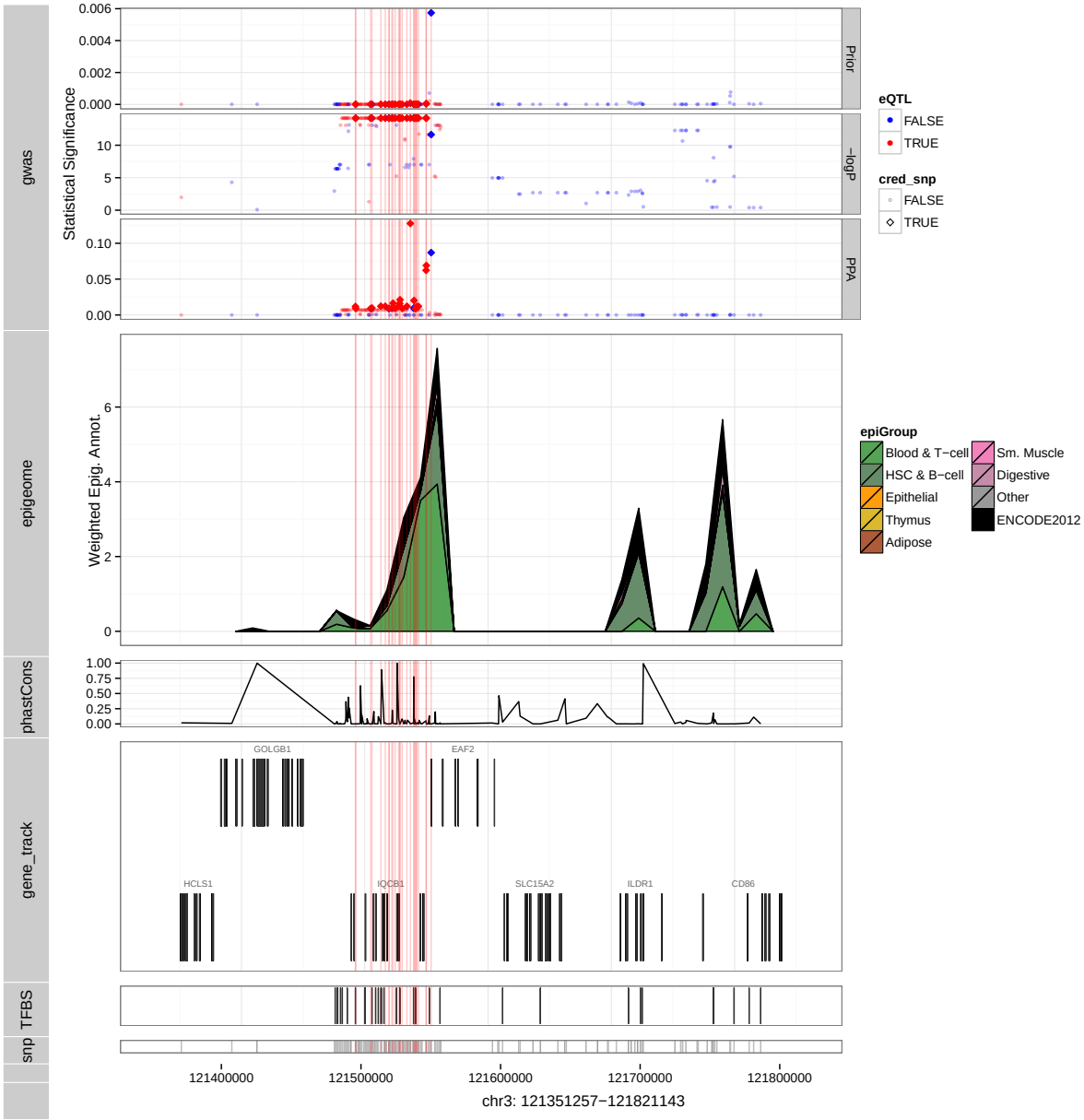
Multiple Sclerosis



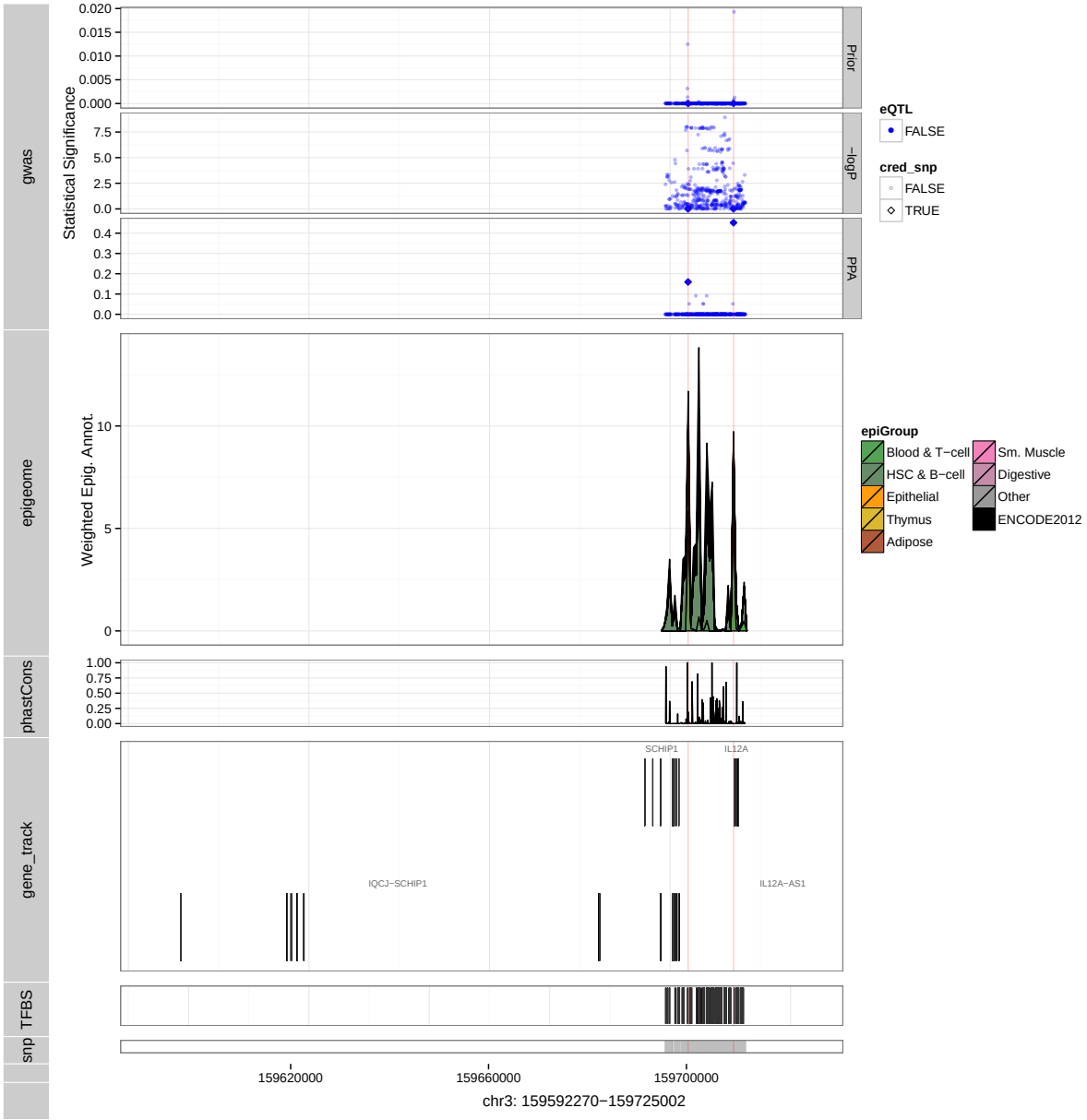
Multiple Sclerosis



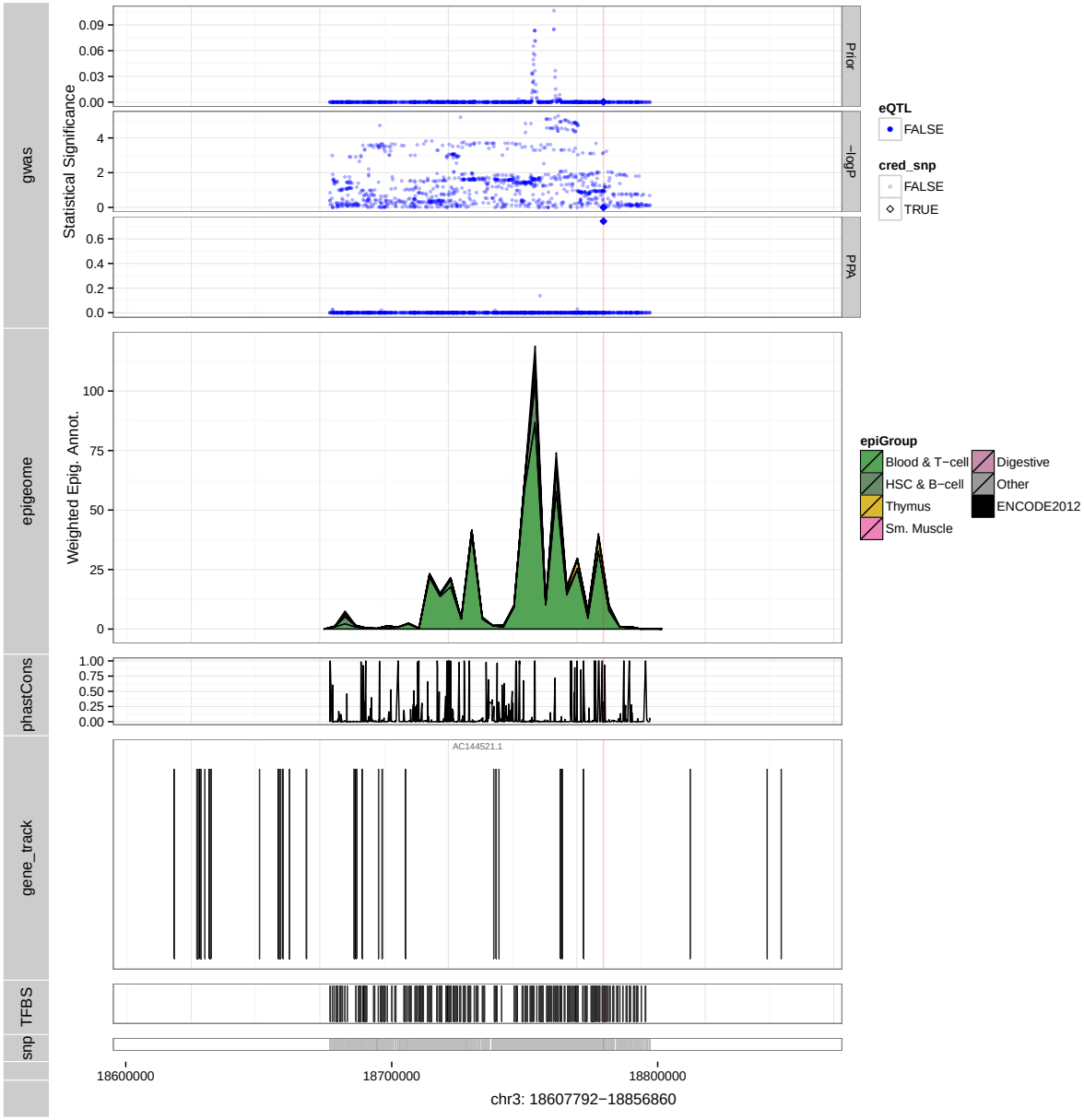
Multiple Sclerosis



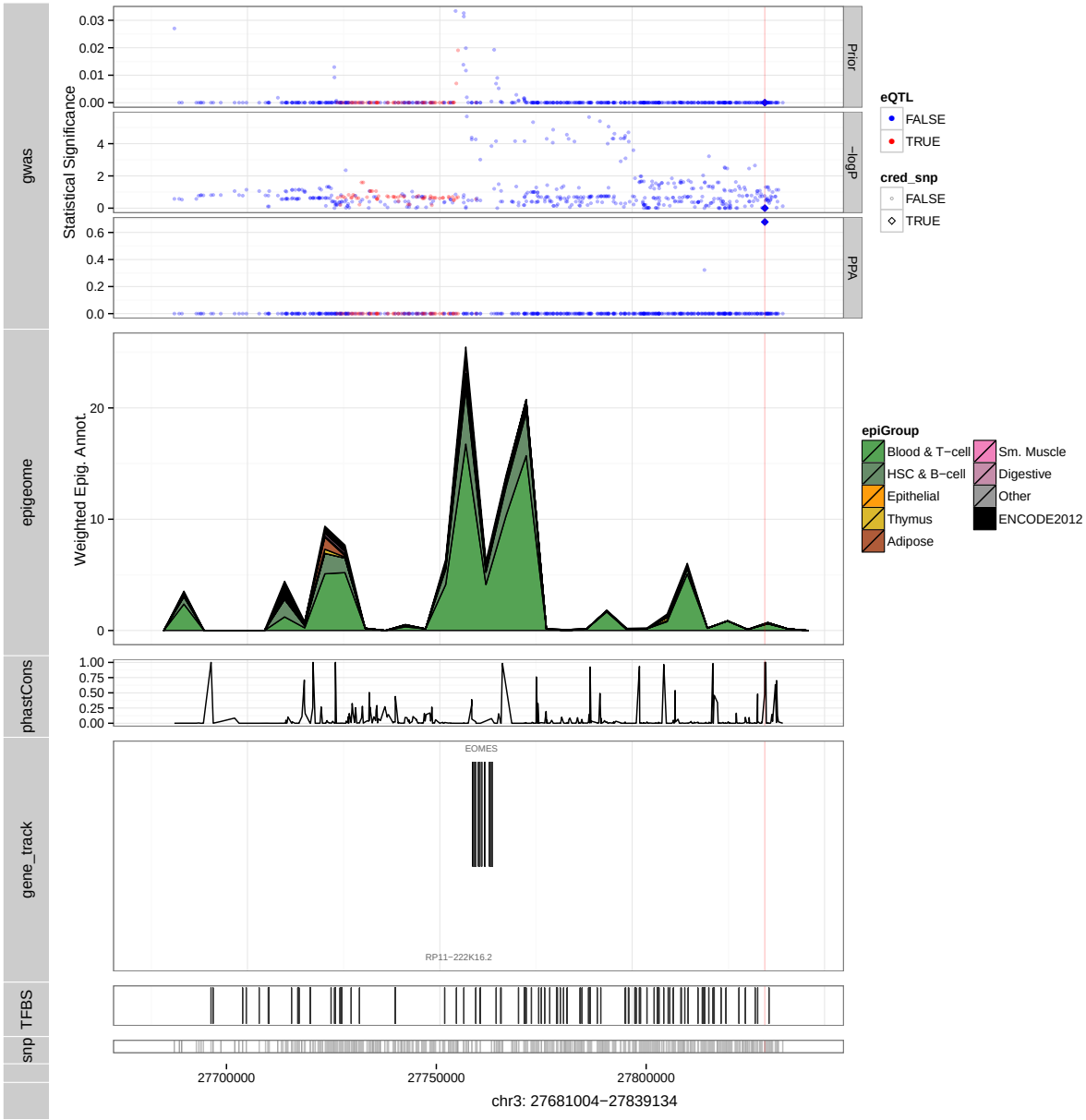
Multiple Sclerosis



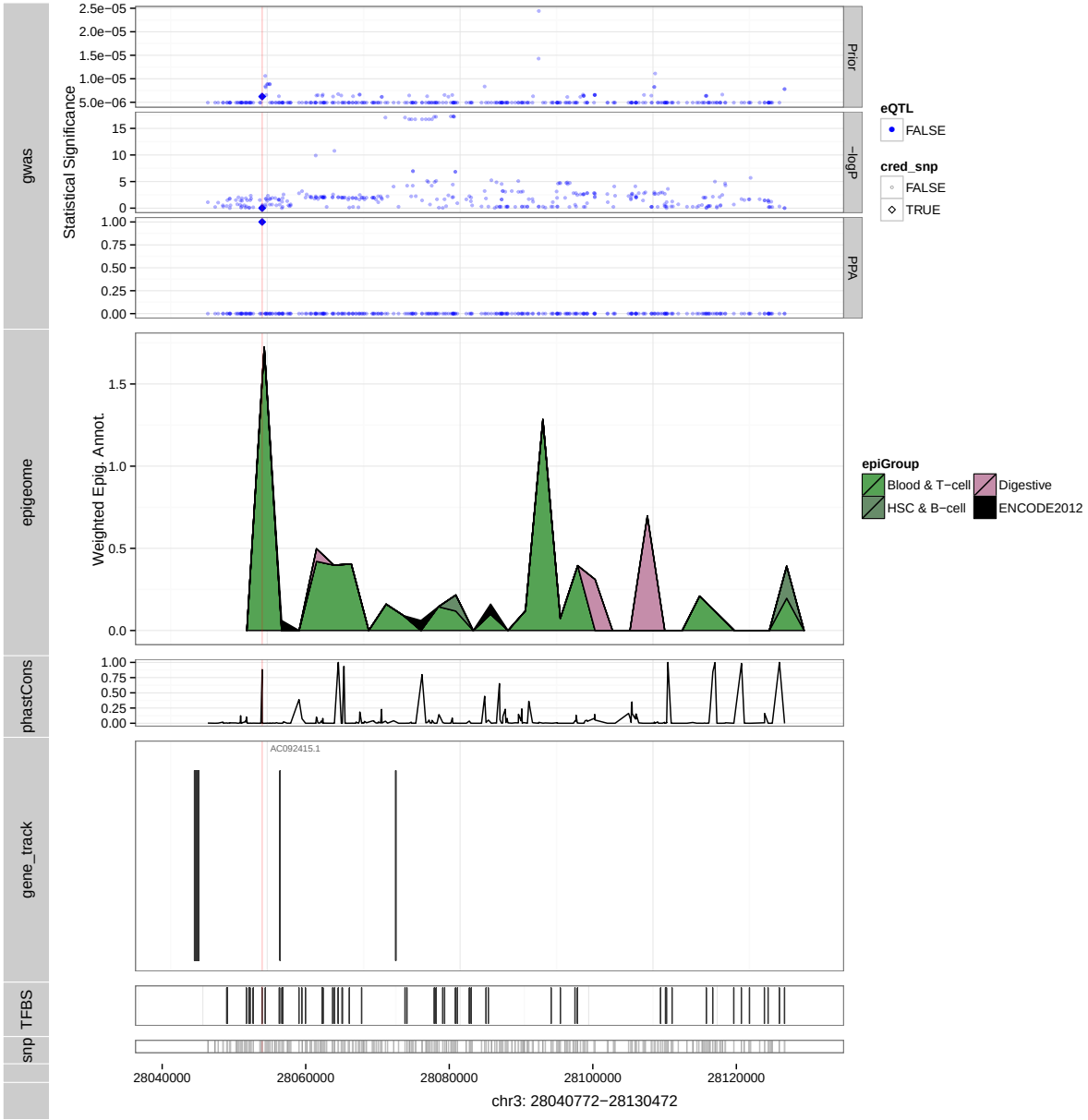
Multiple Sclerosis



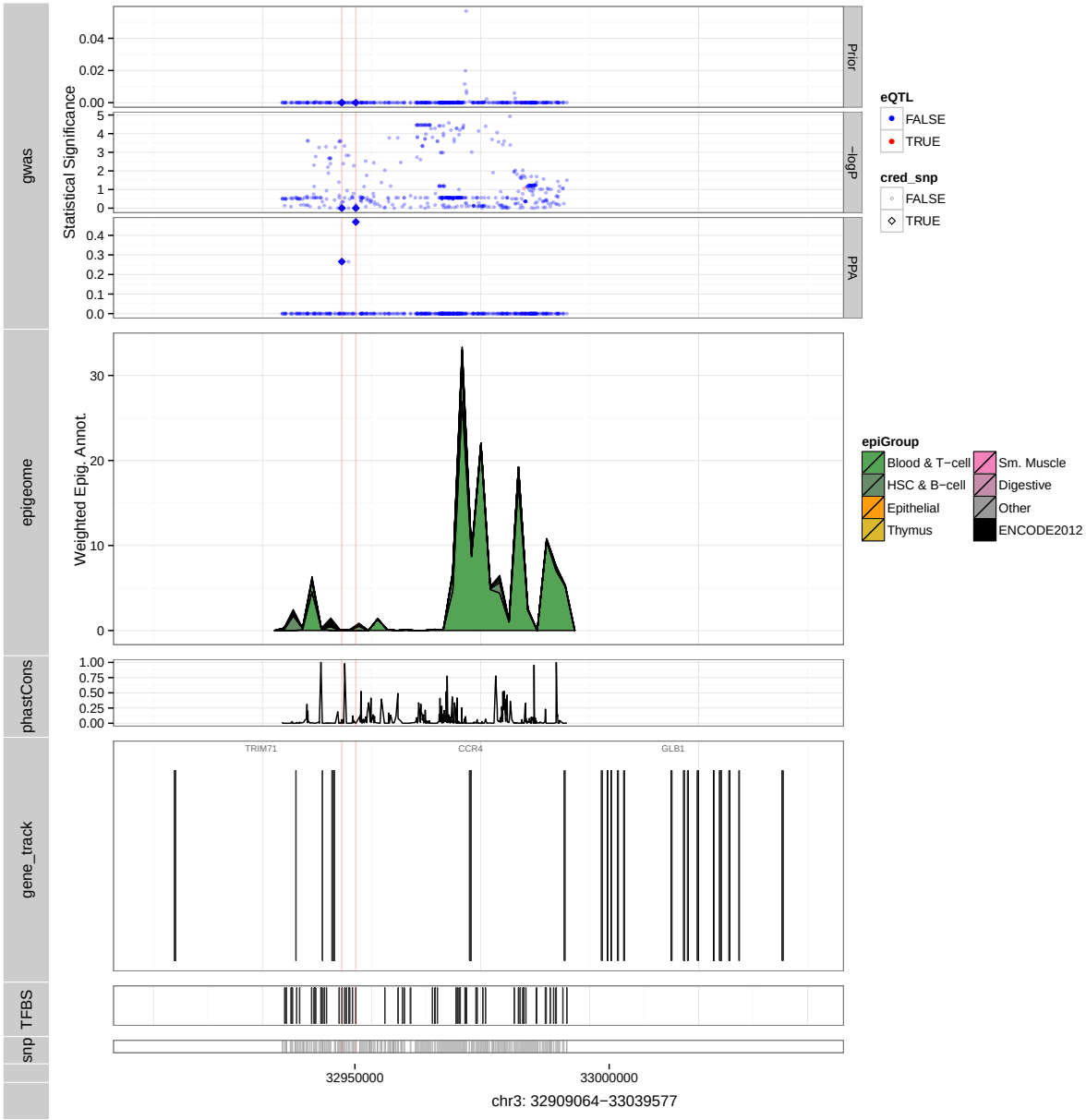
Multiple Sclerosis



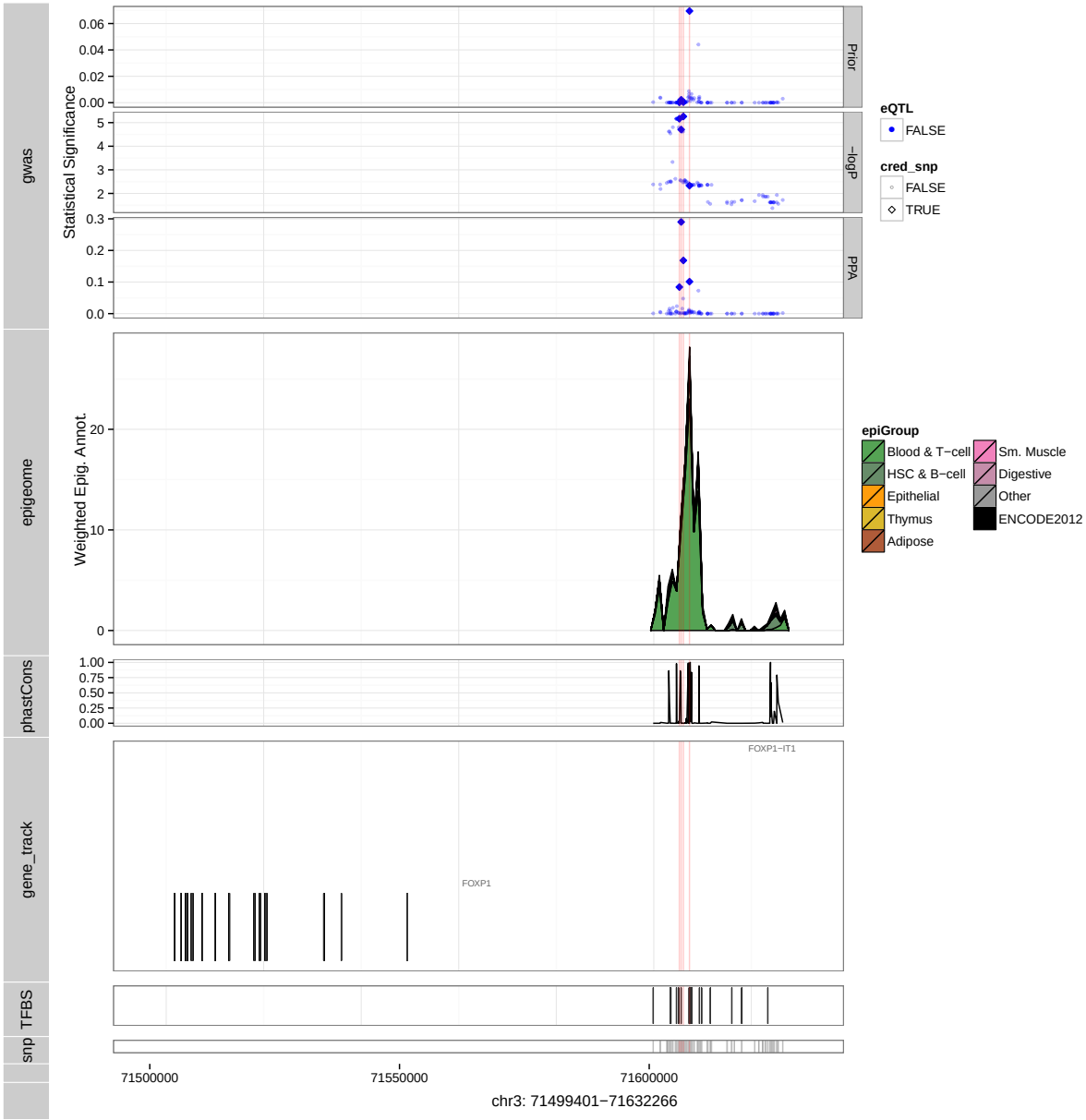
Multiple Sclerosis



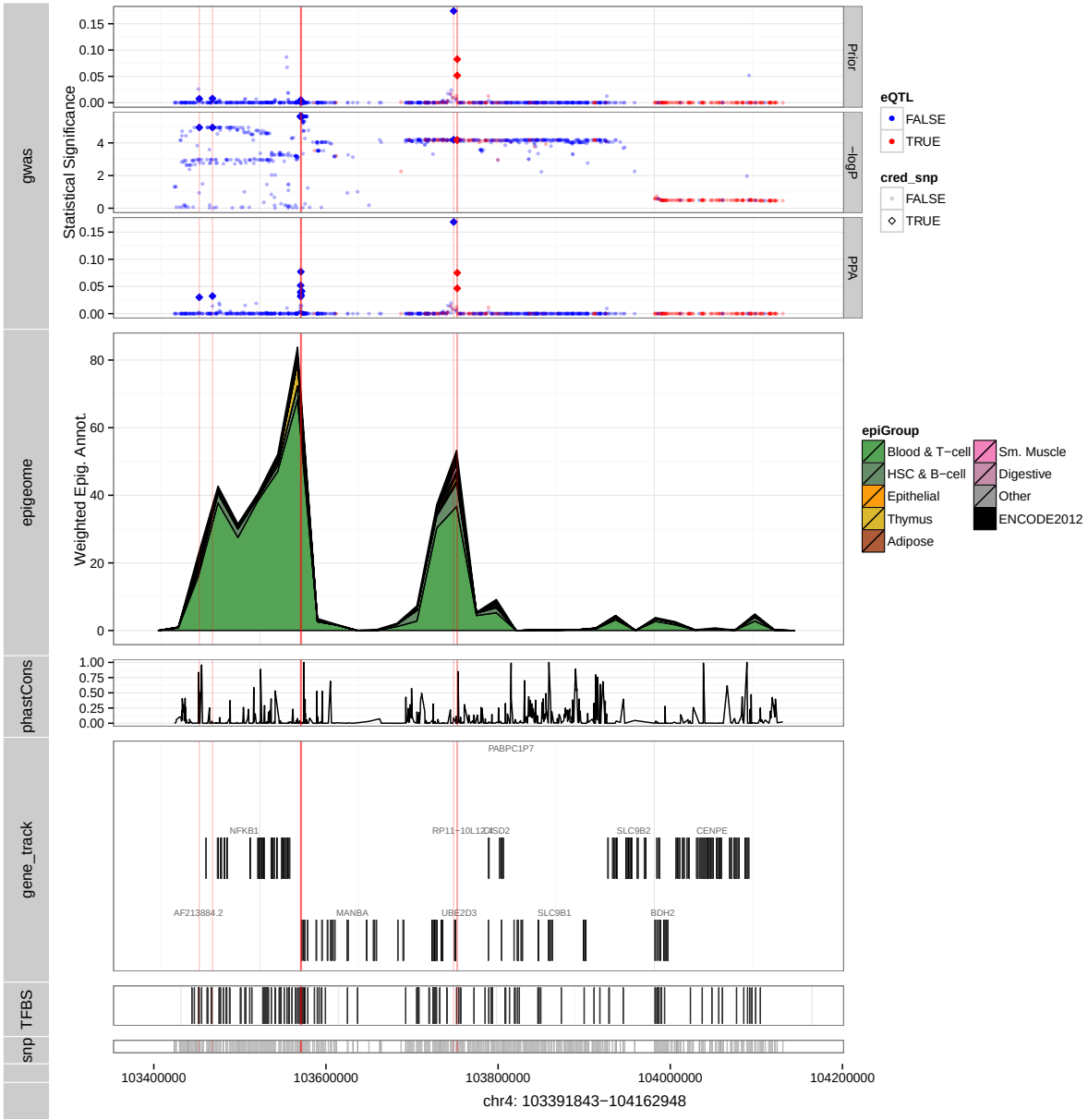
Multiple Sclerosis



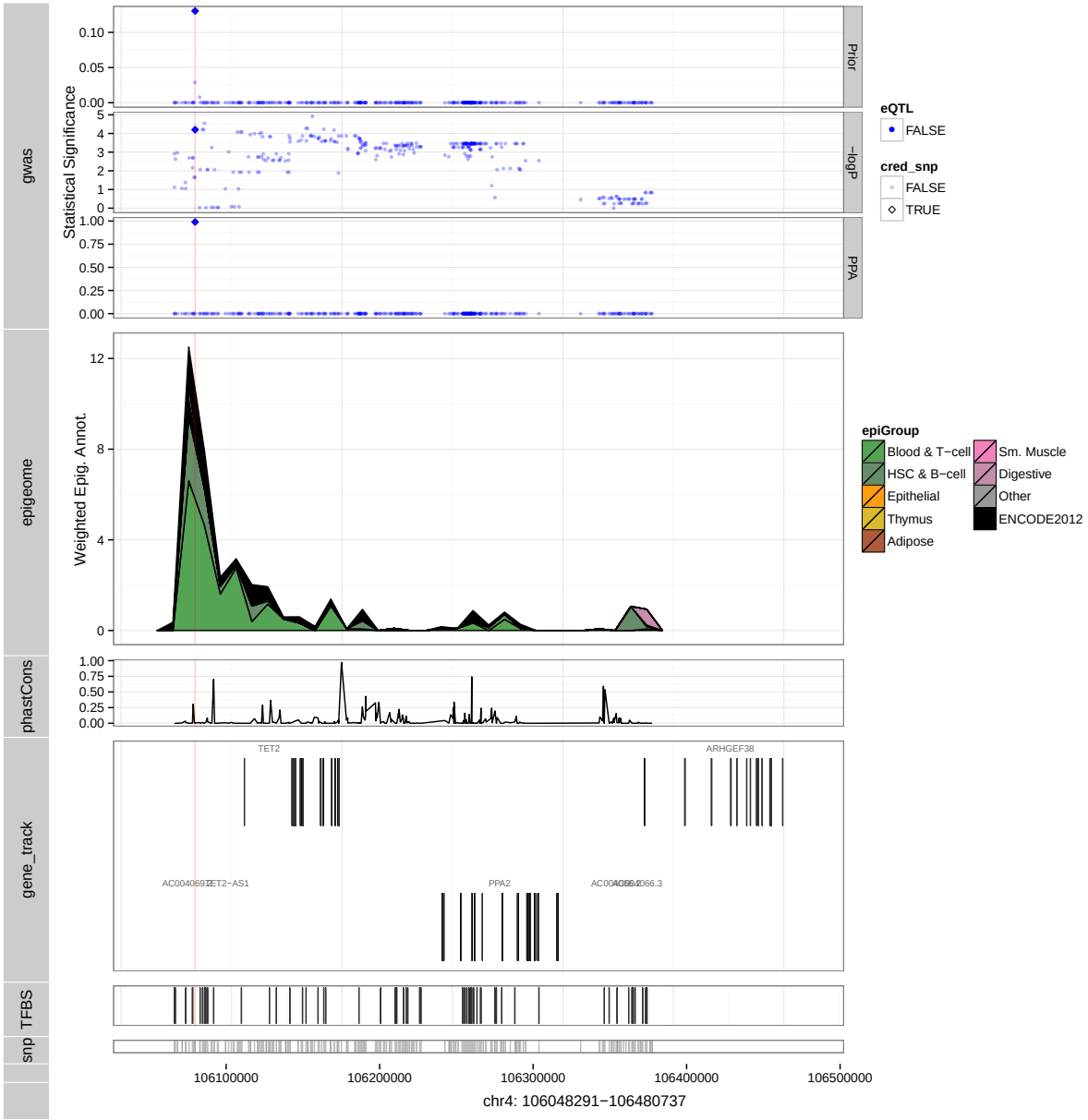
Multiple Sclerosis



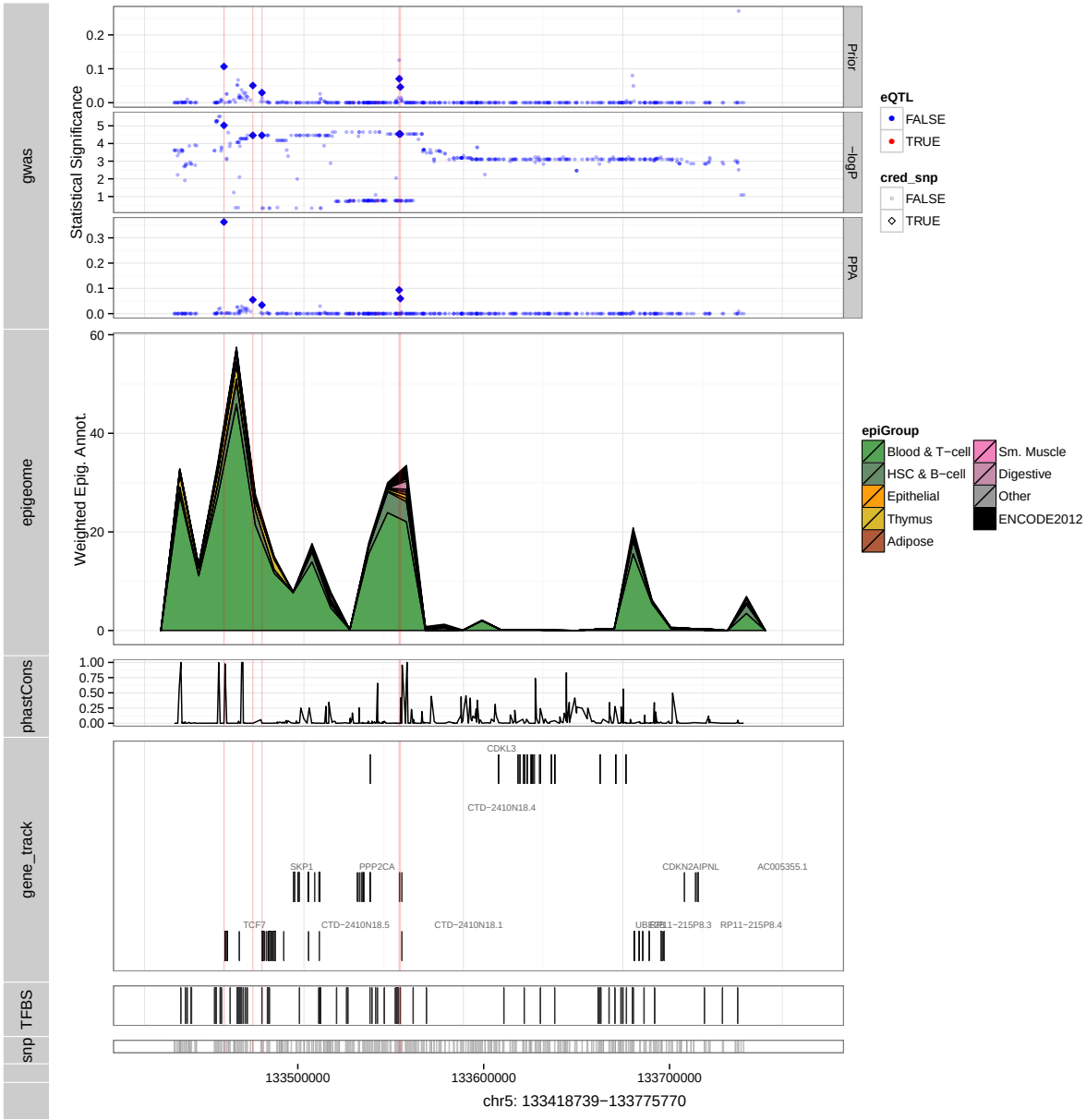
Multiple Sclerosis



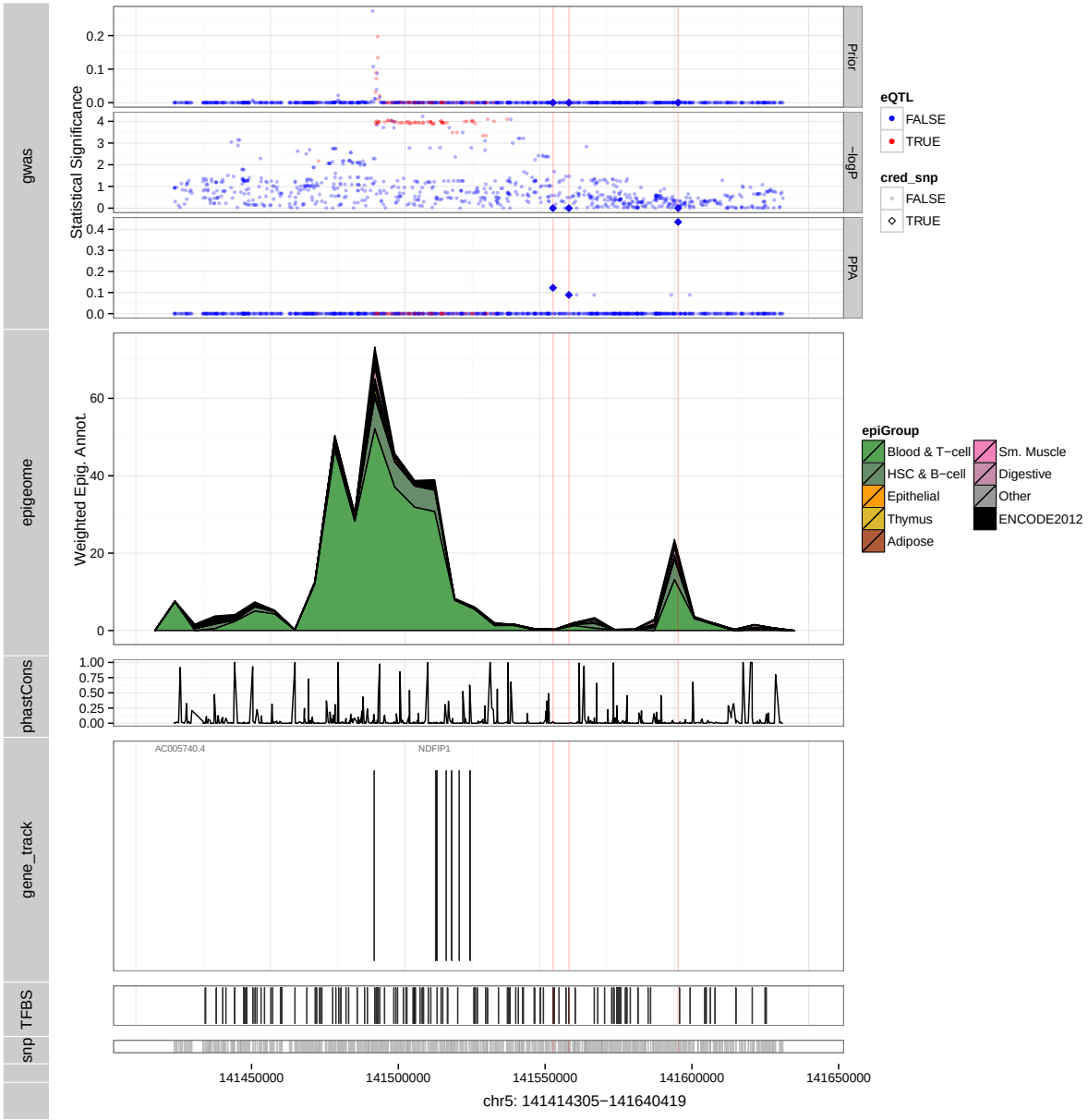
Multiple Sclerosis



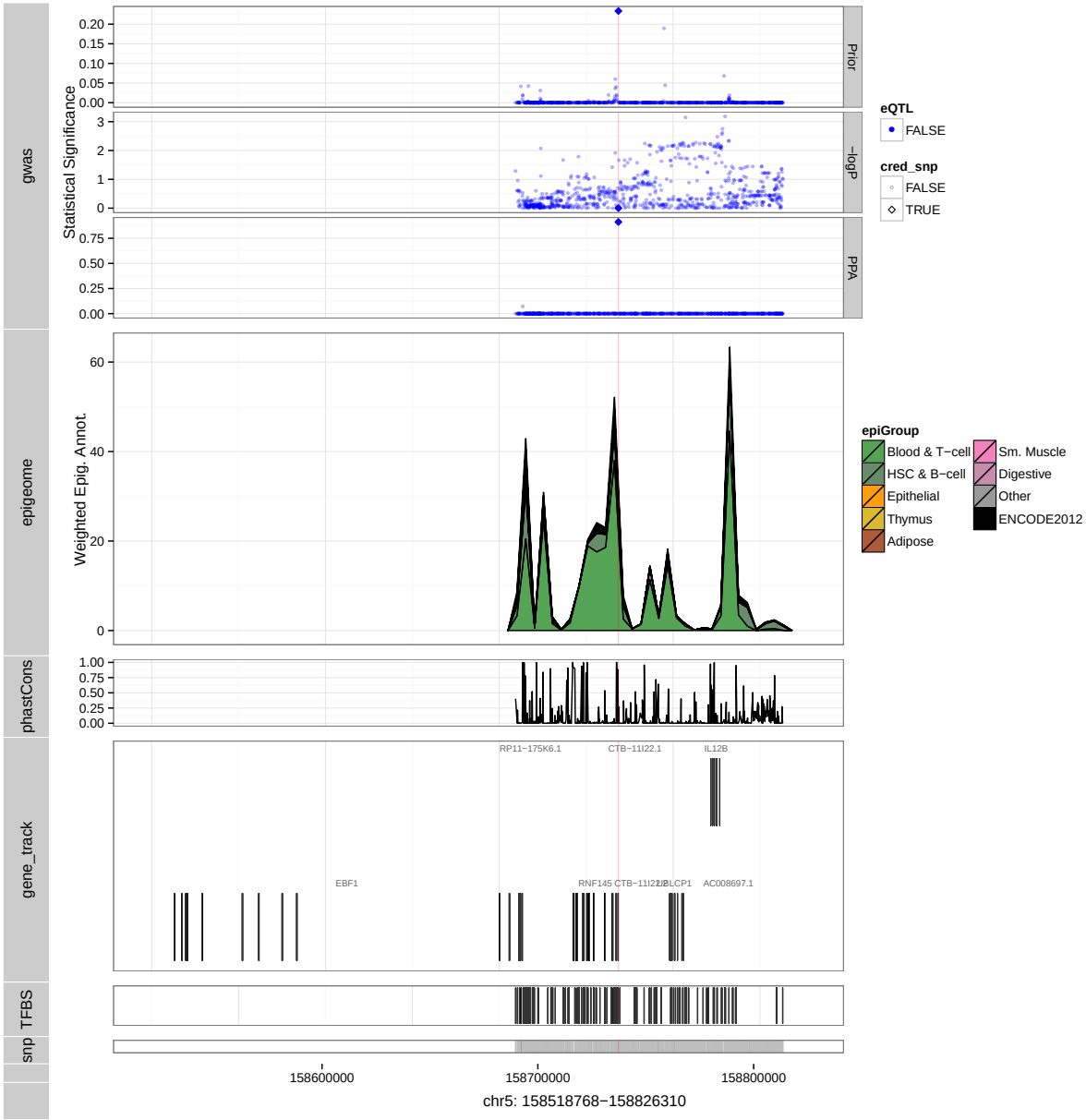
Multiple Sclerosis



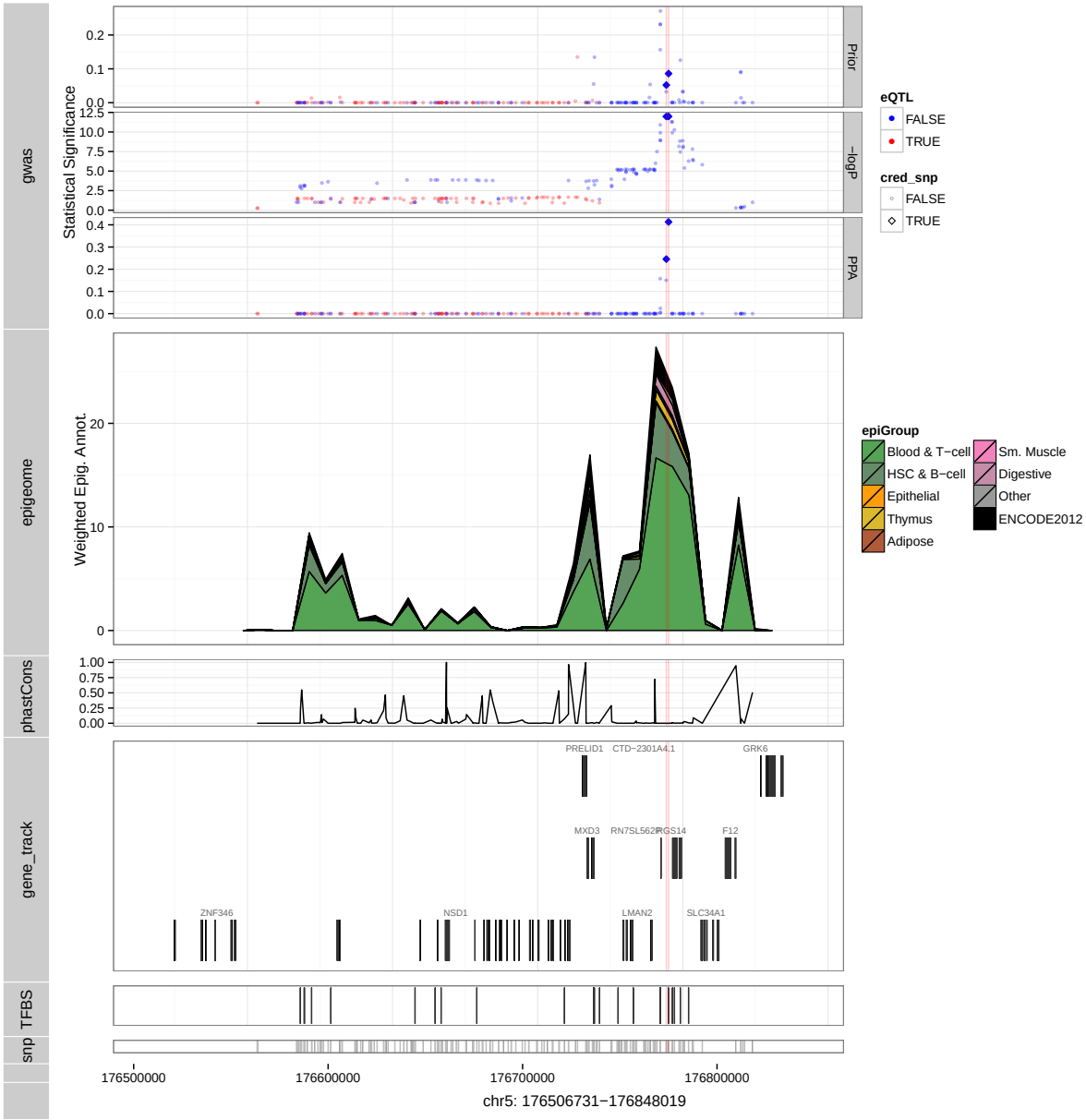
Multiple Sclerosis



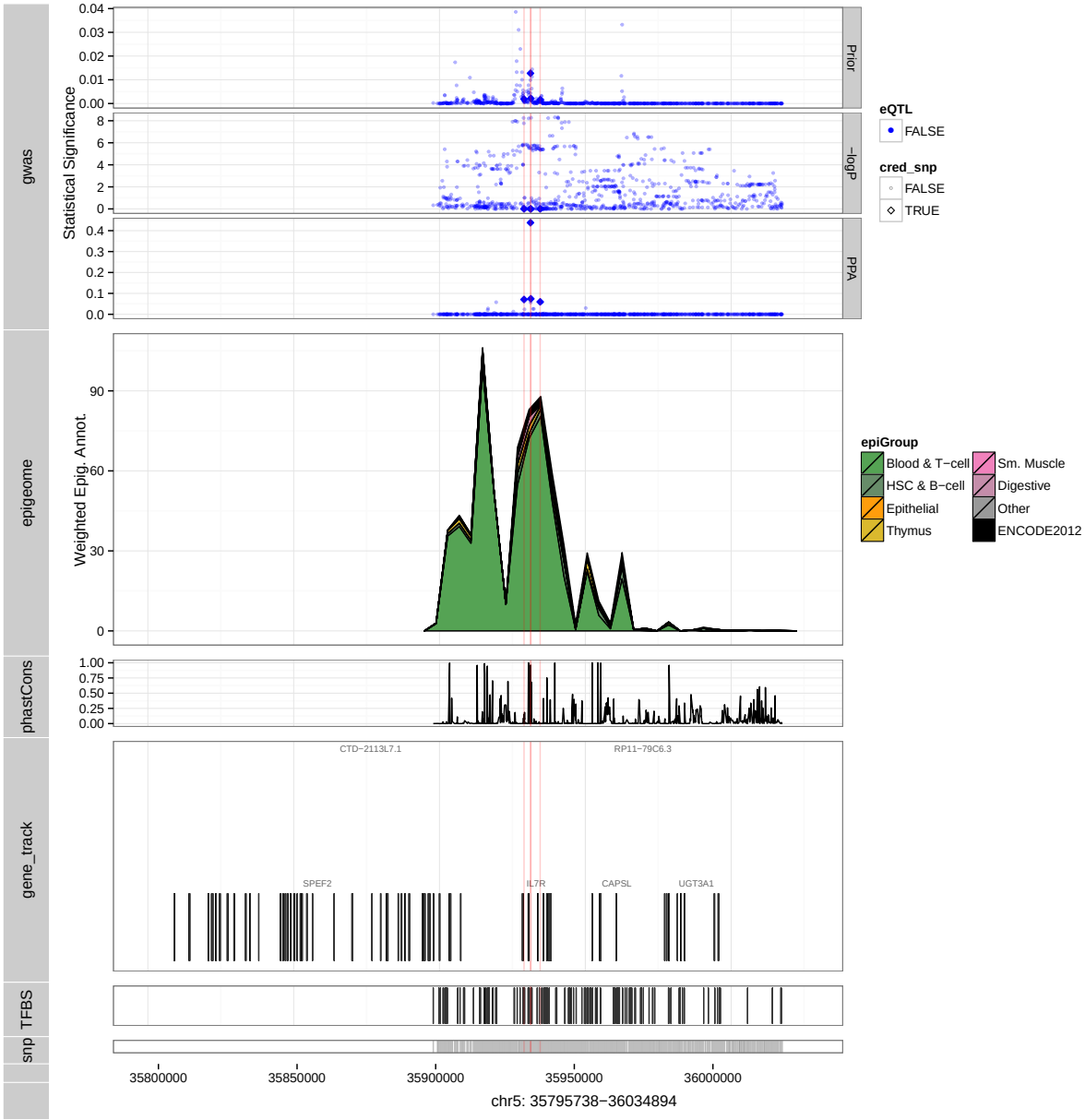
Multiple Sclerosis



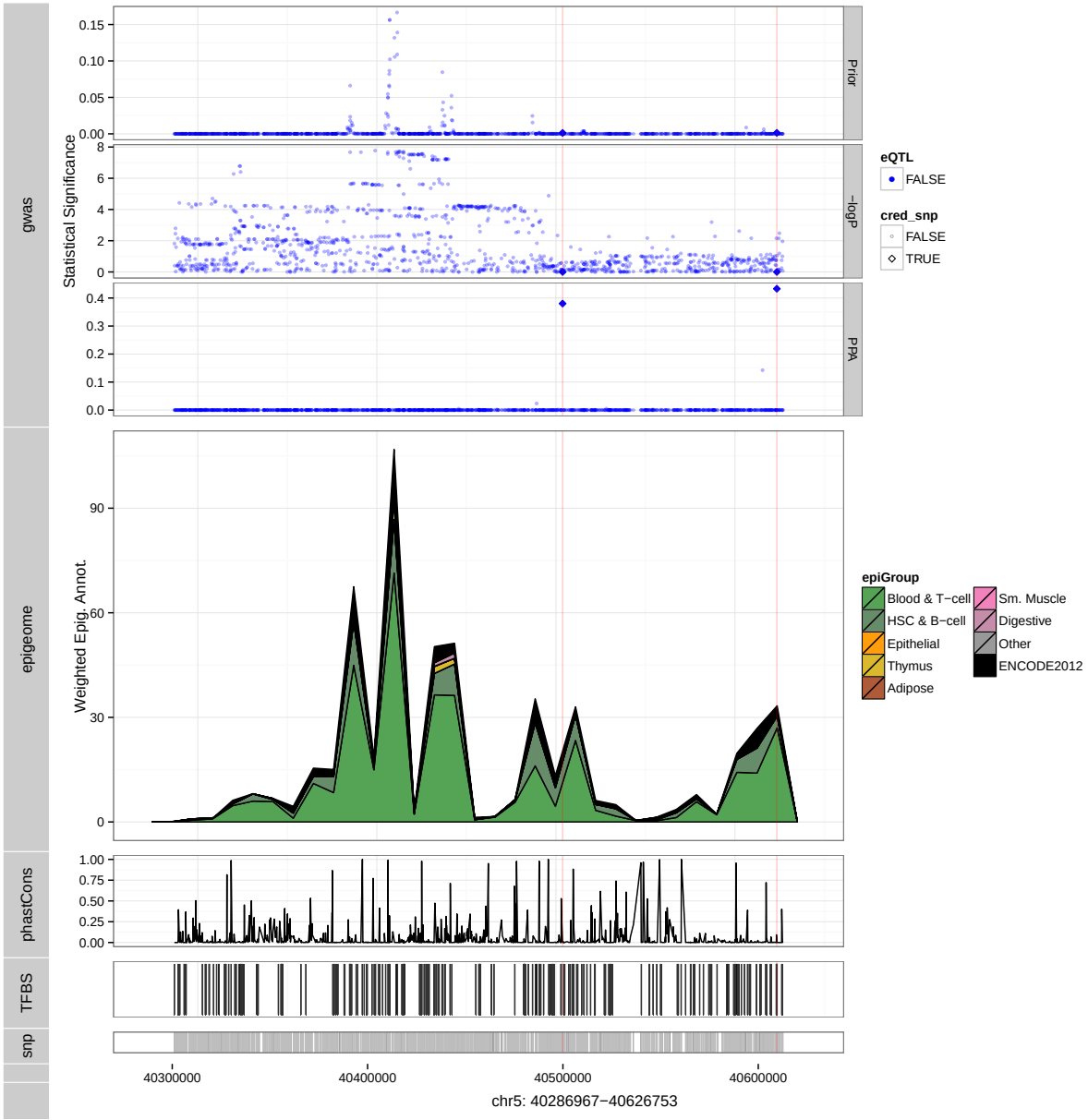
Multiple Sclerosis



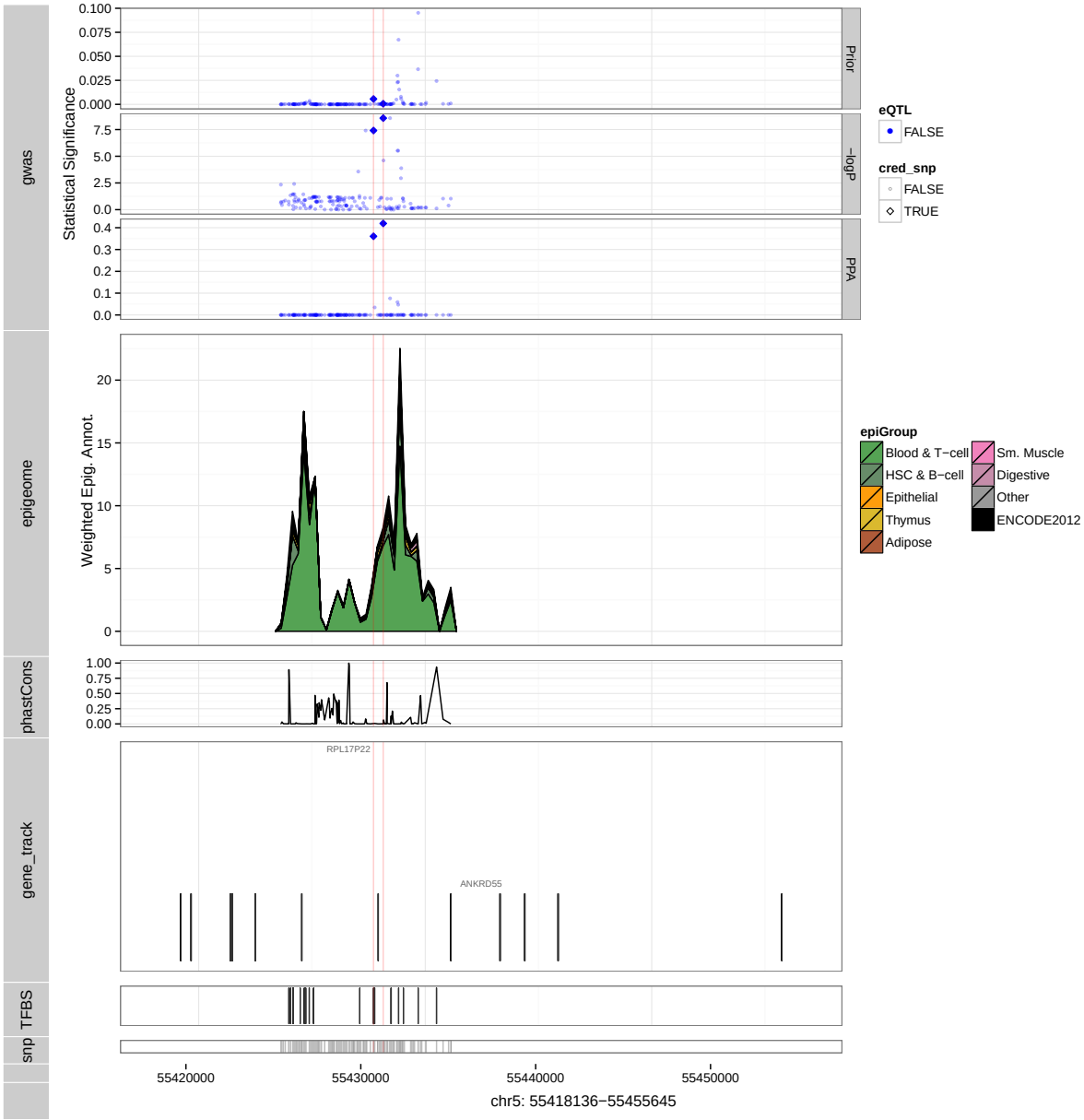
Multiple Sclerosis



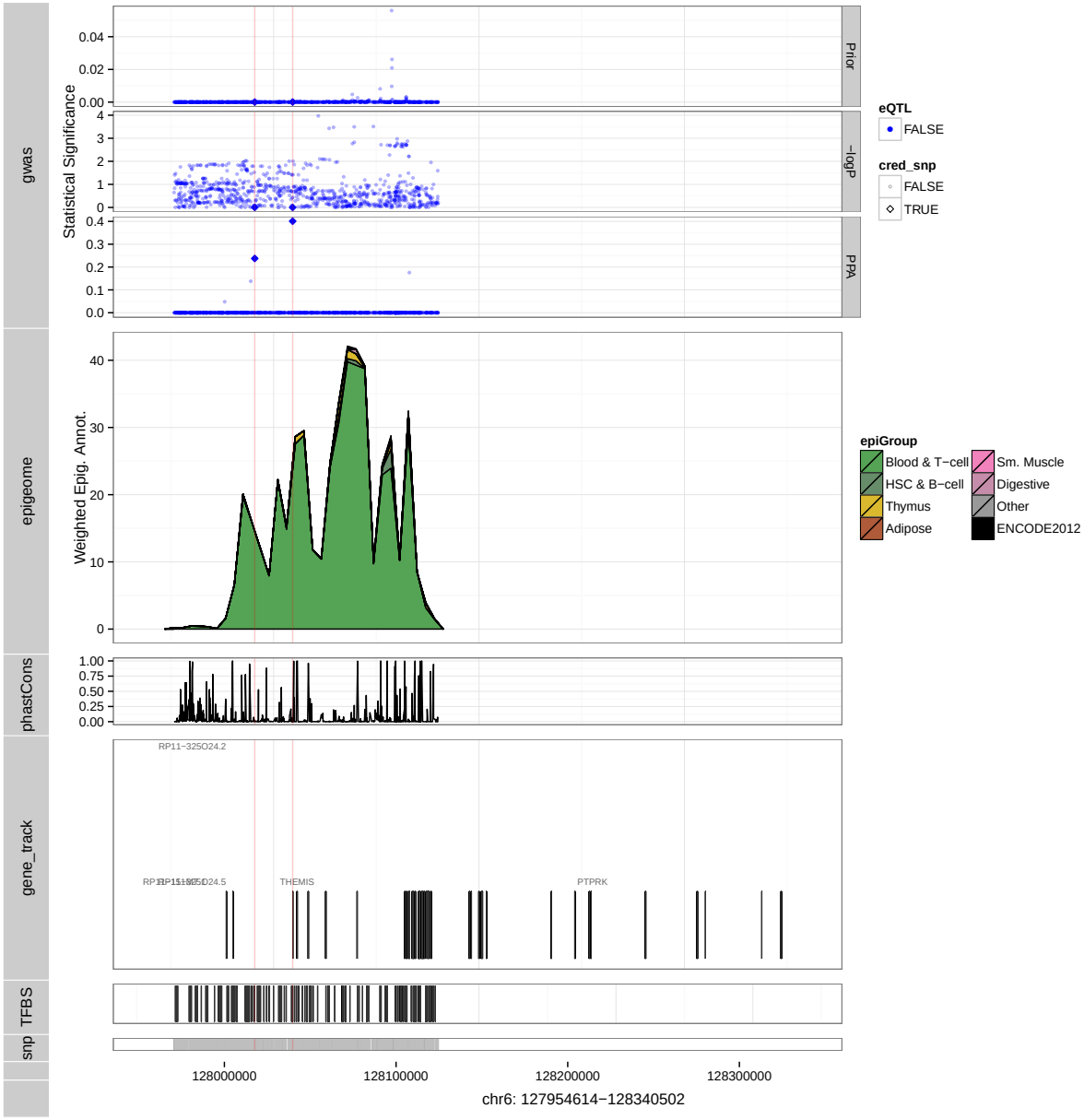
Multiple Sclerosis



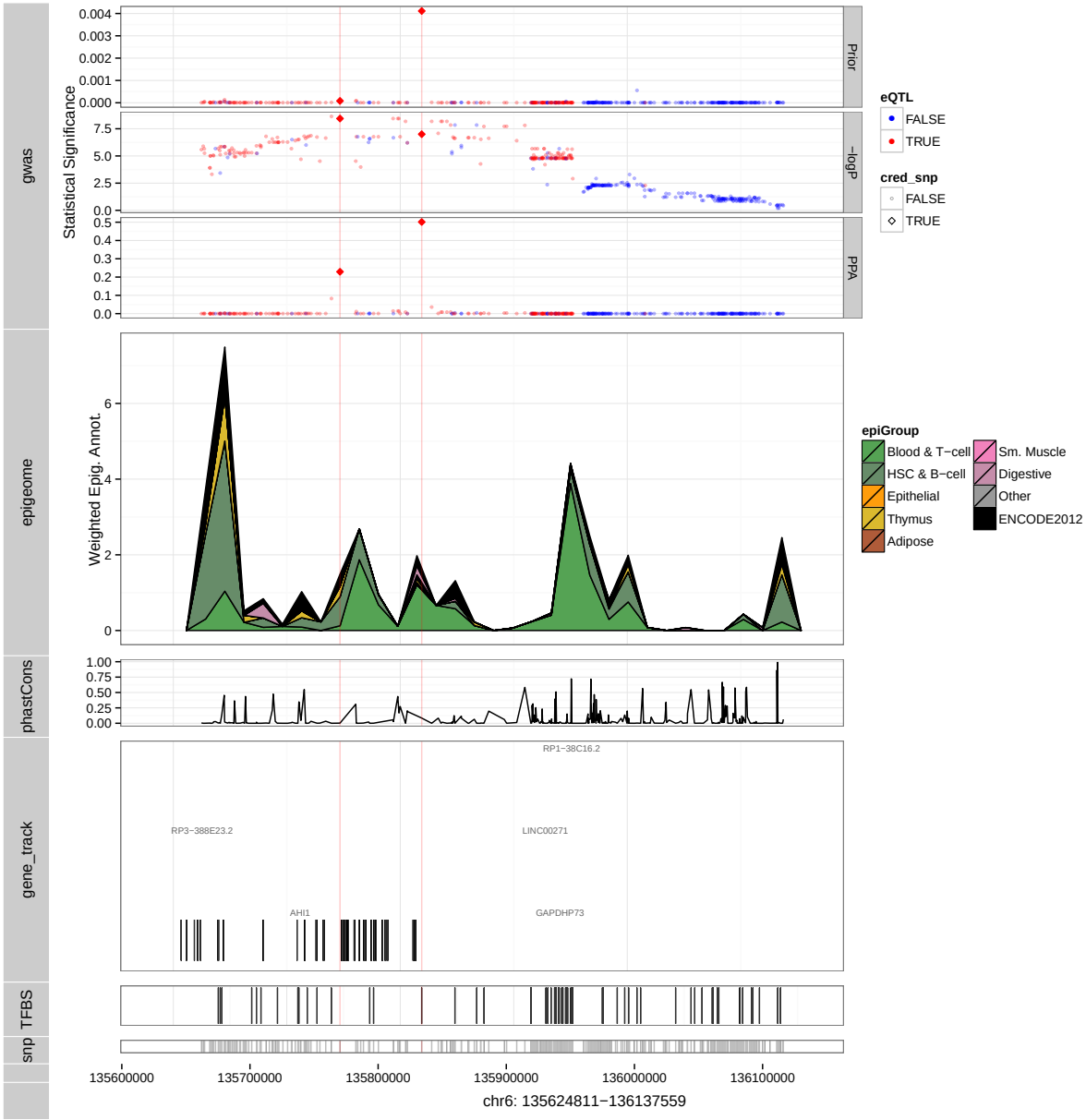
Multiple Sclerosis



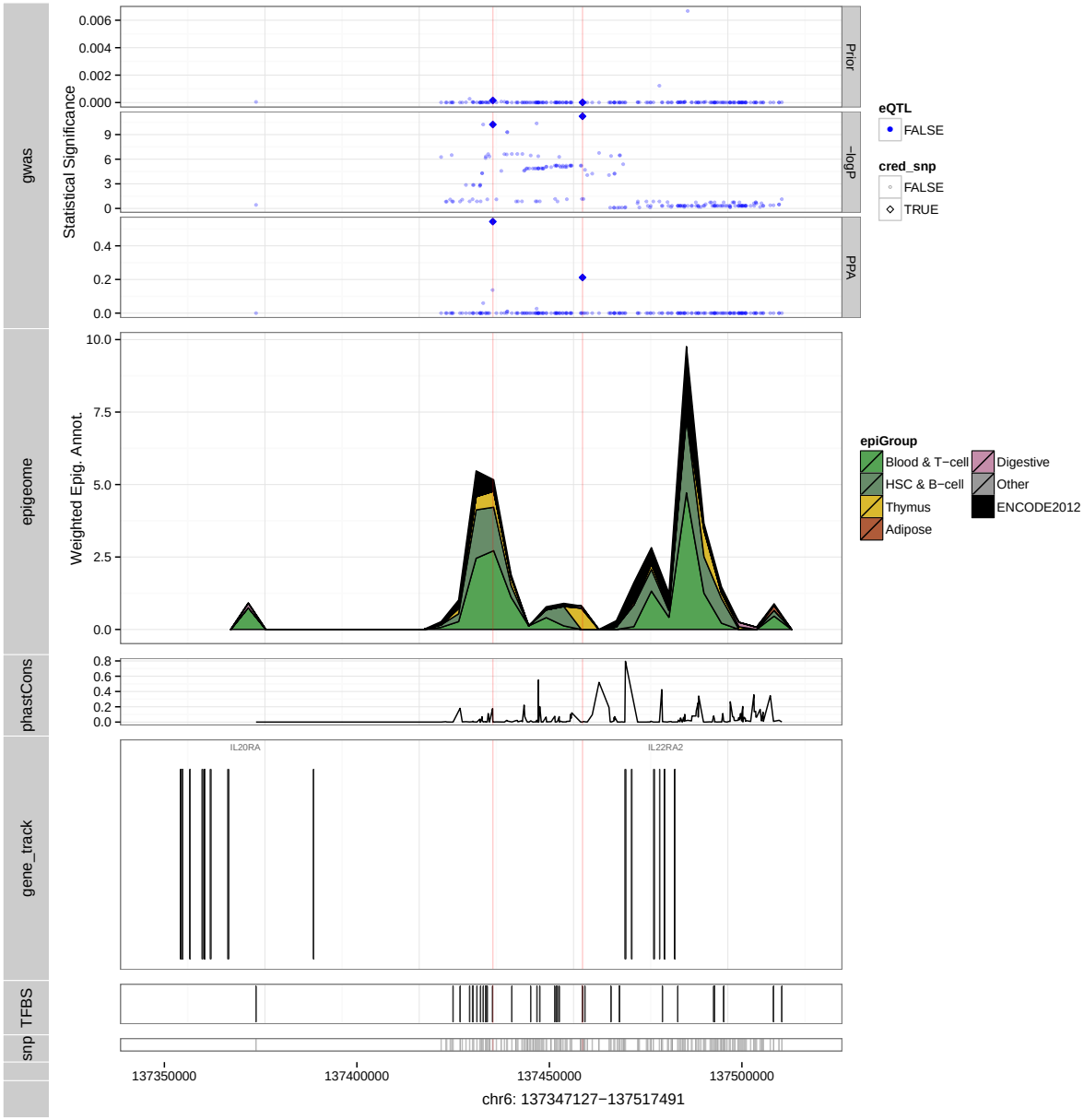
Multiple Sclerosis



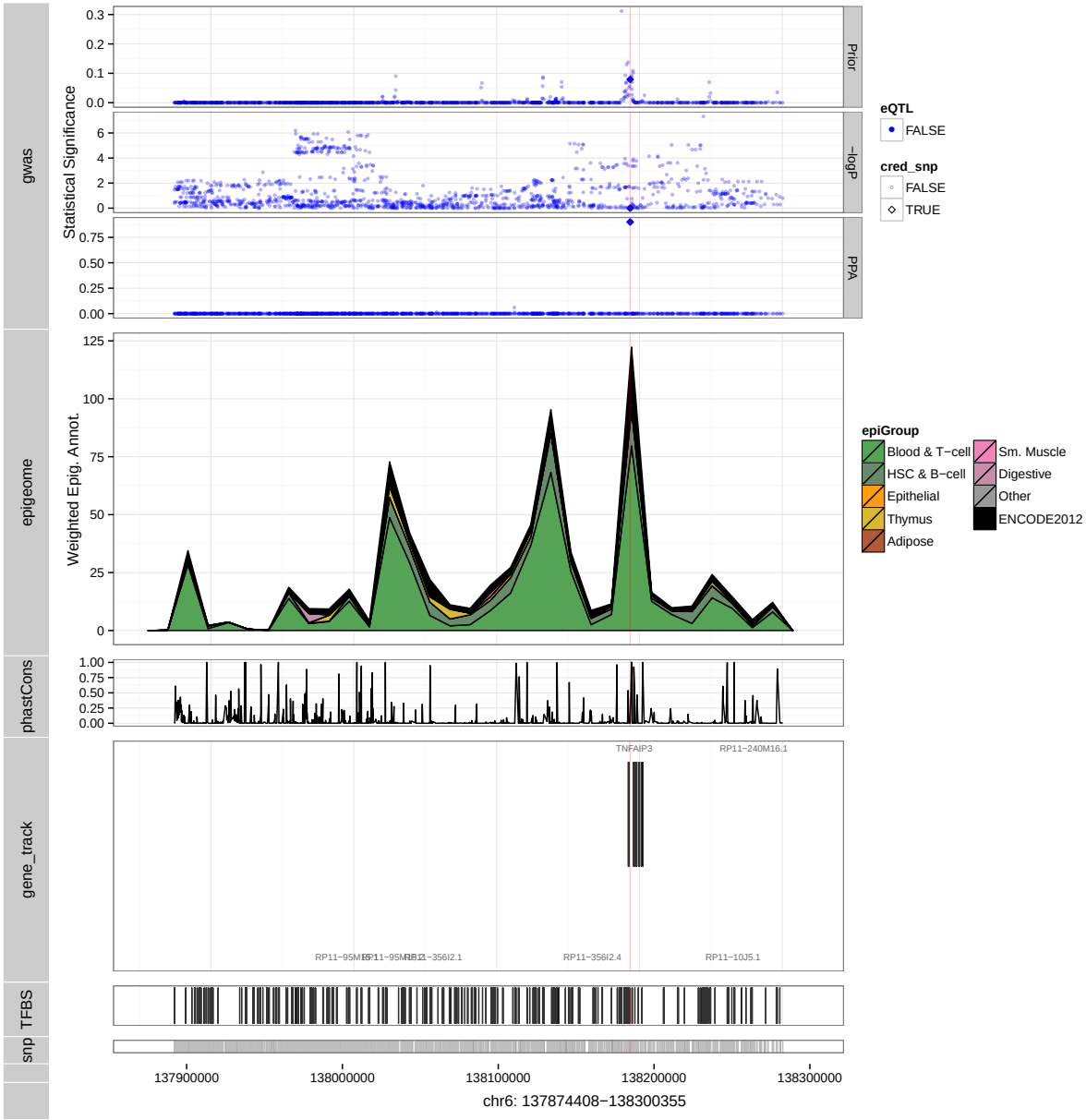
Multiple Sclerosis



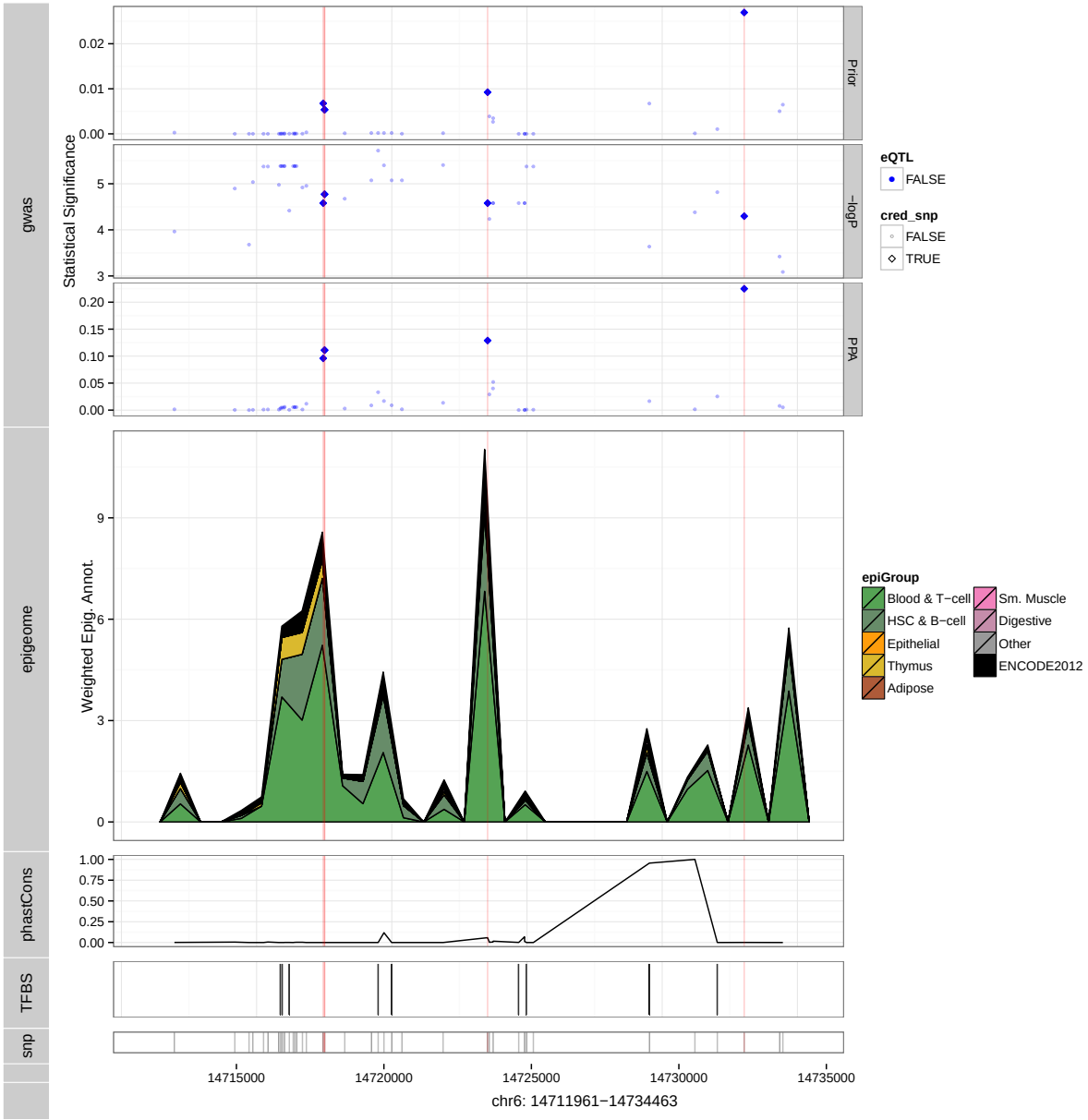
Multiple Sclerosis



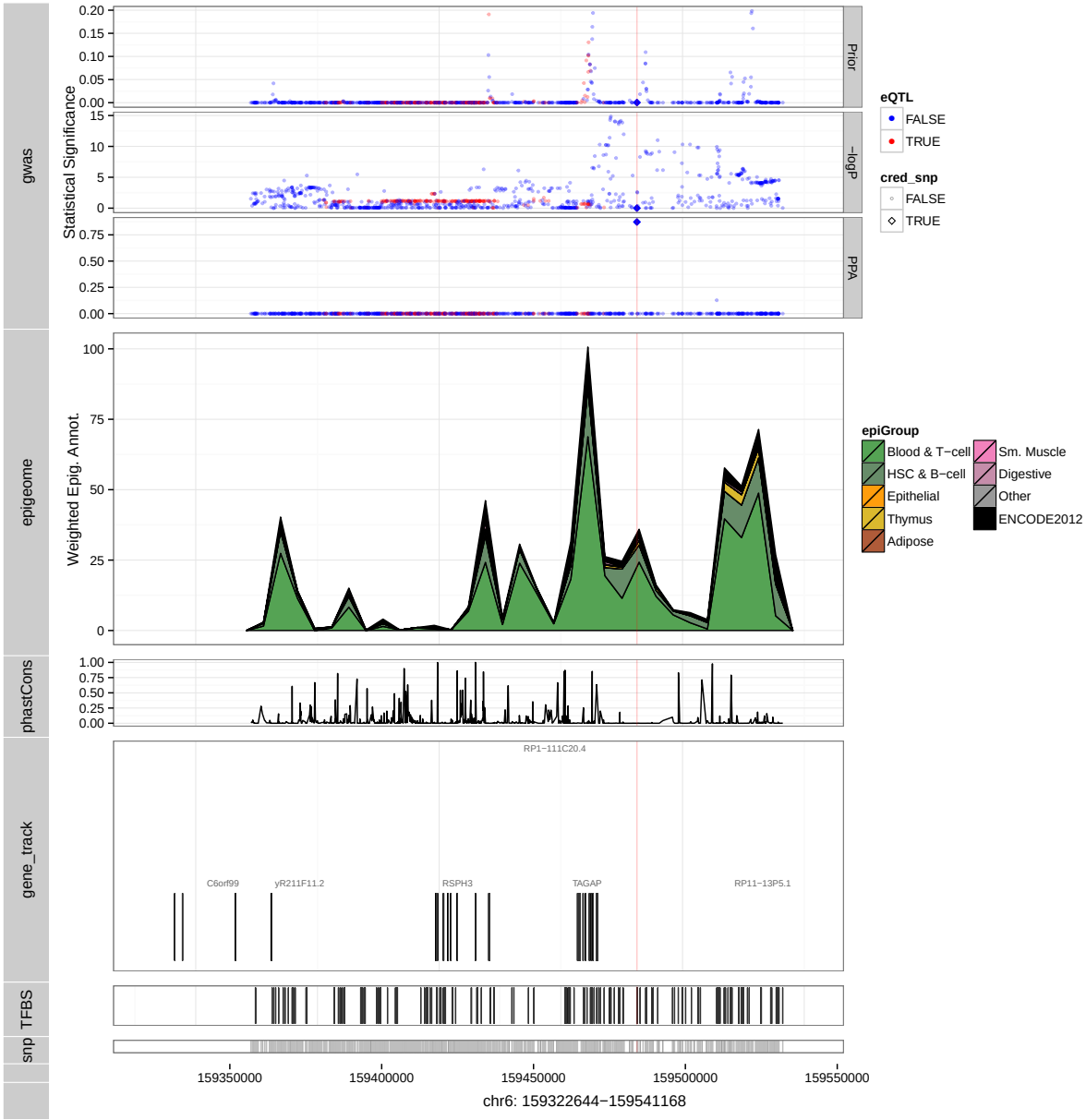
Multiple Sclerosis



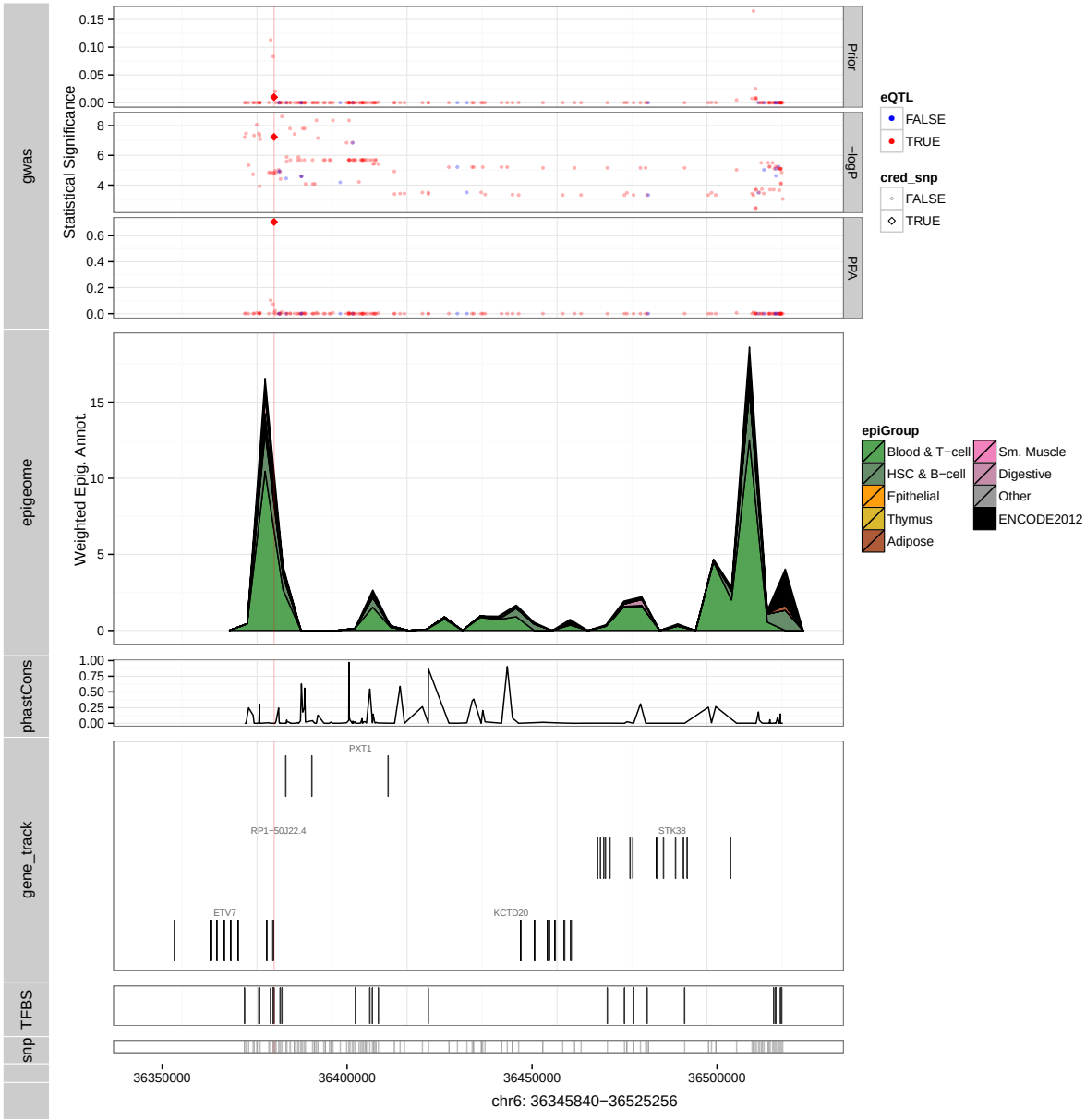
Multiple Sclerosis



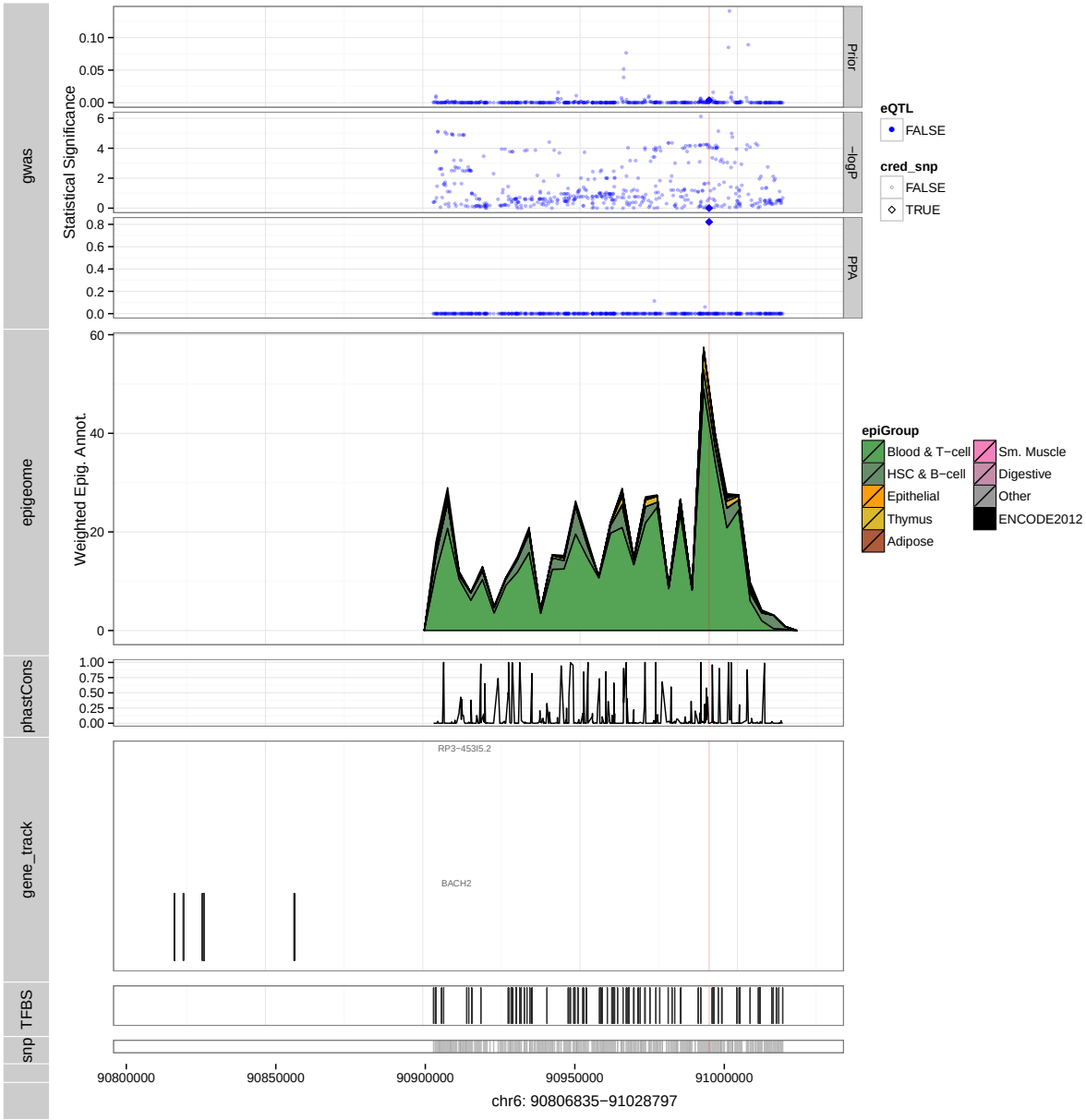
Multiple Sclerosis



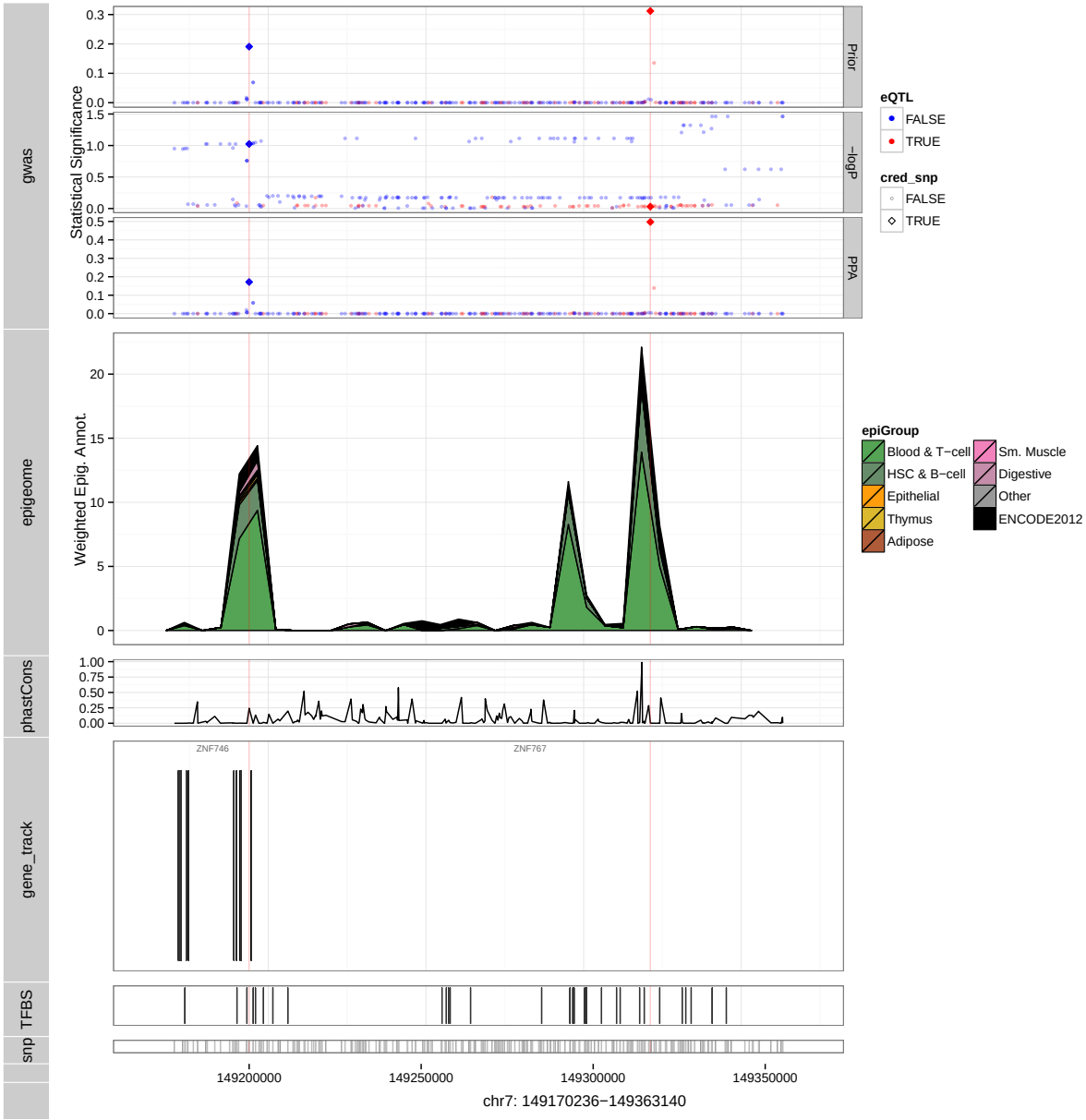
Multiple Sclerosis



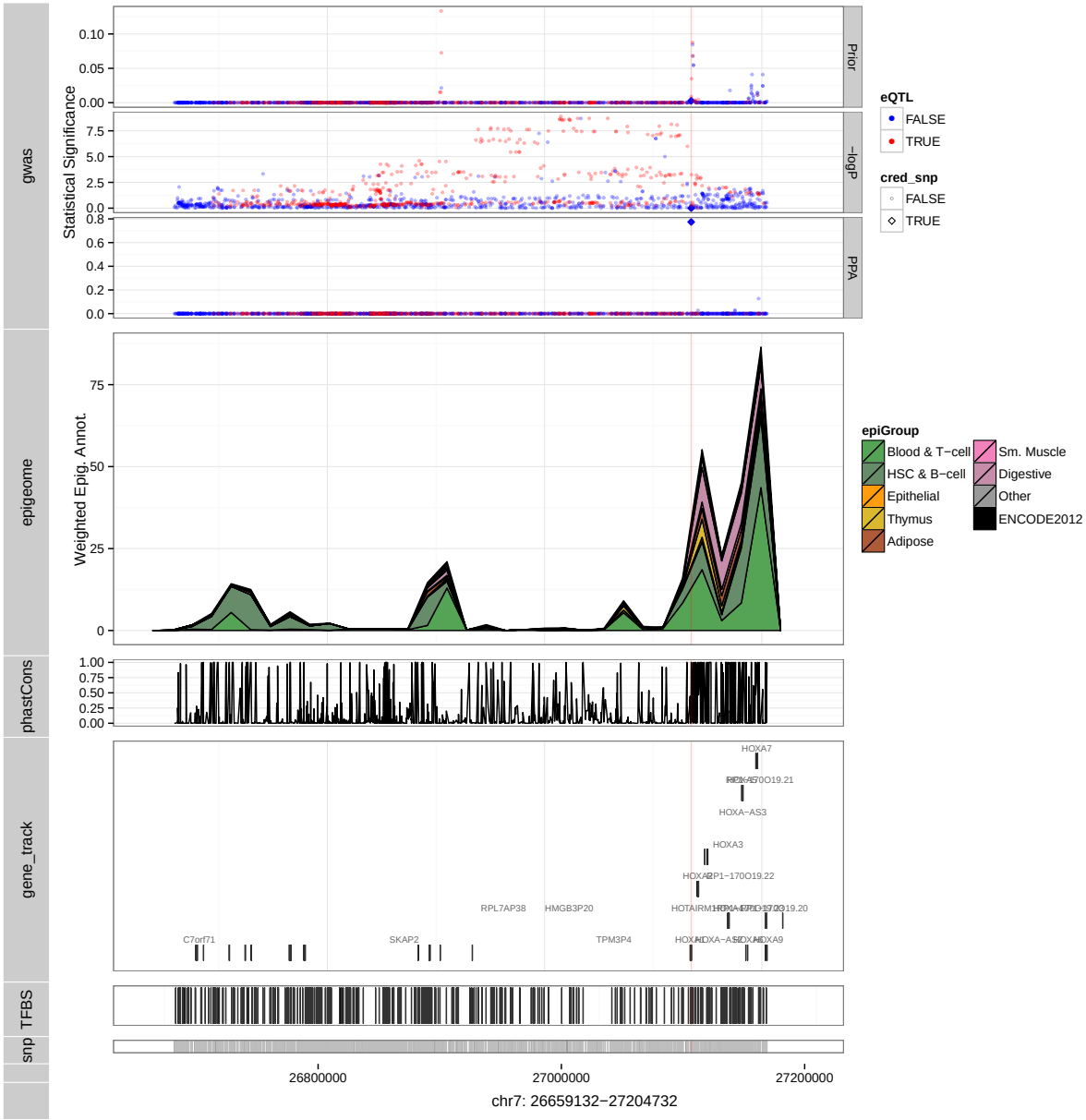
Multiple Sclerosis



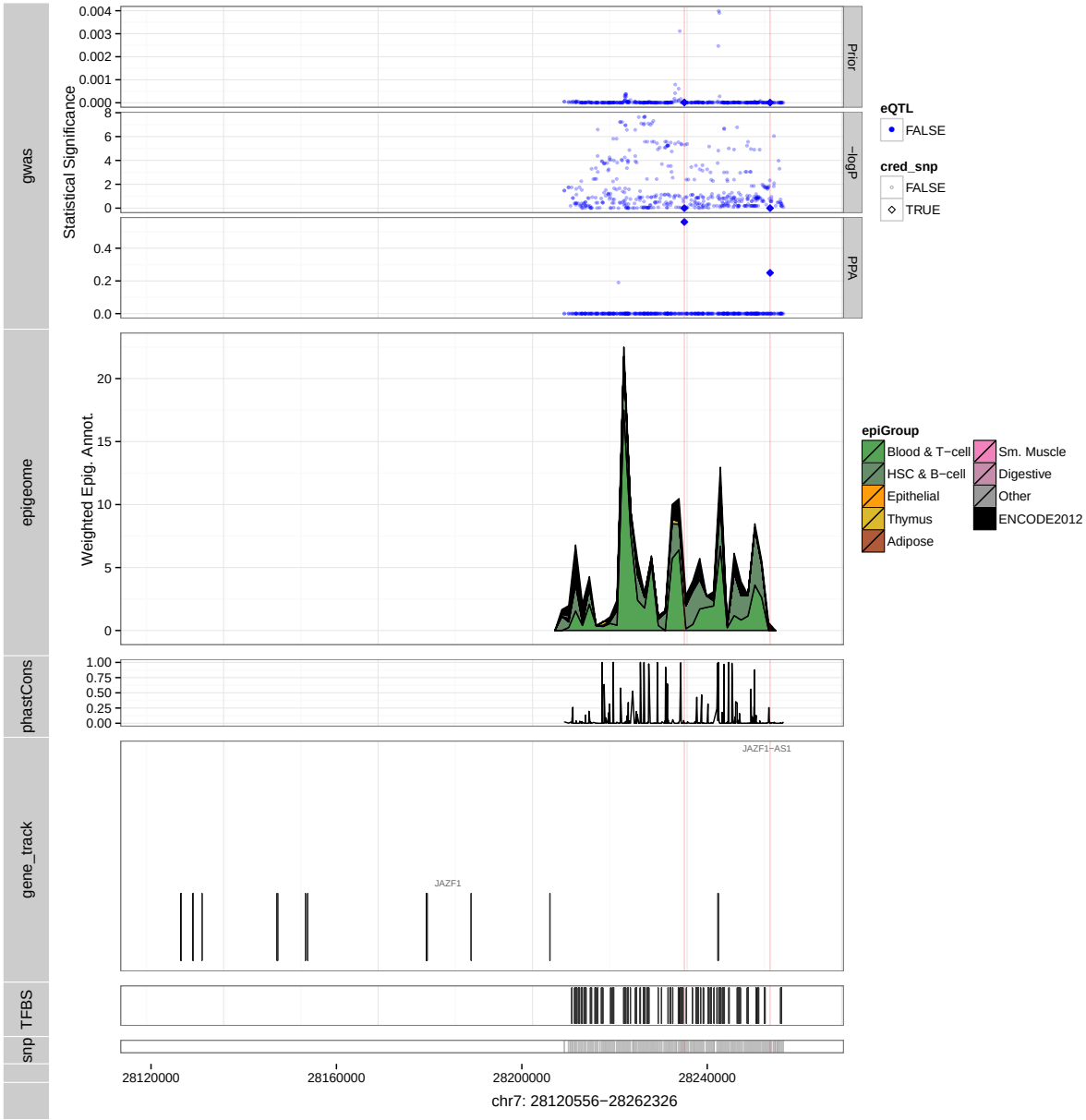
Multiple Sclerosis



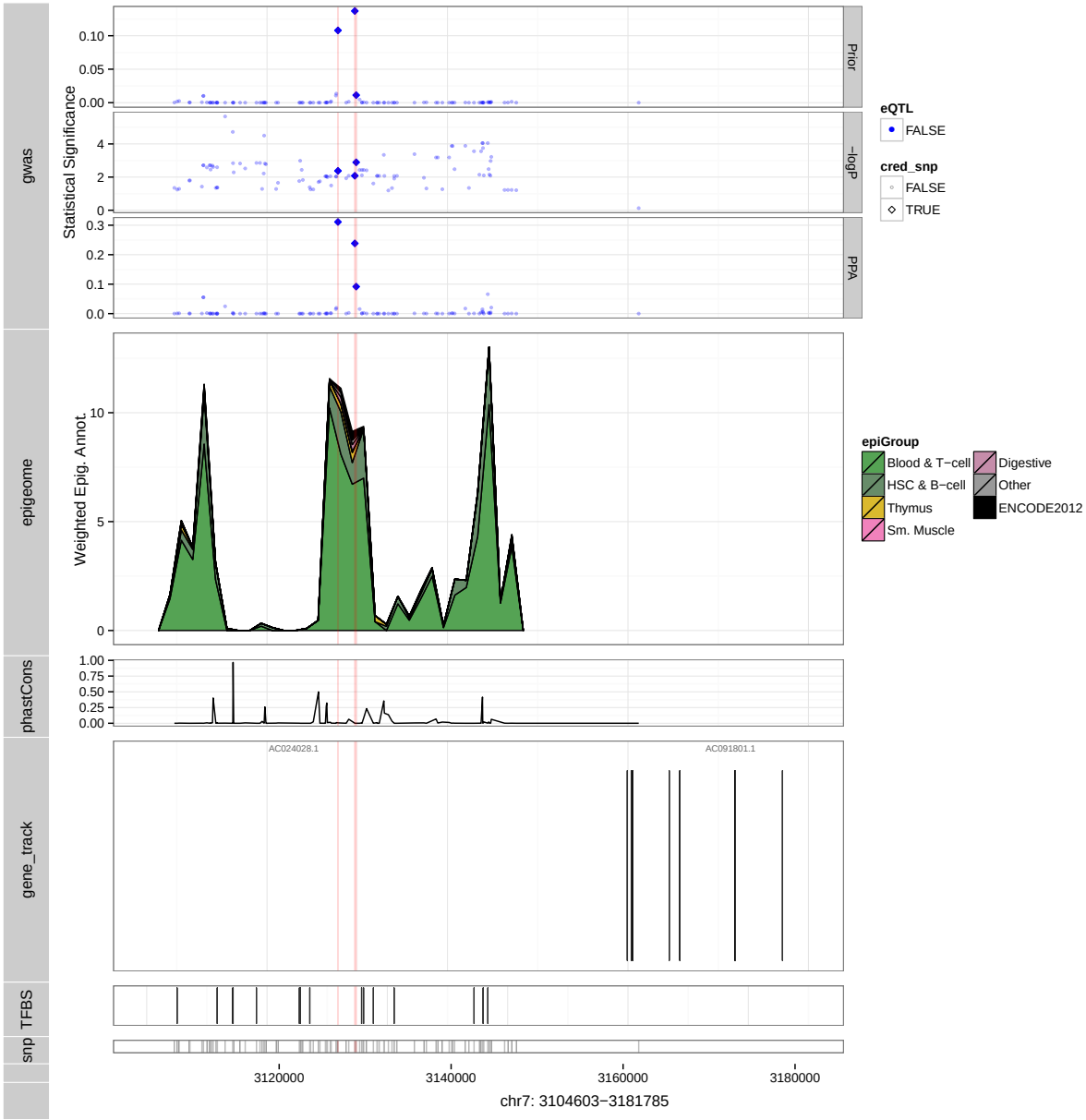
Multiple Sclerosis



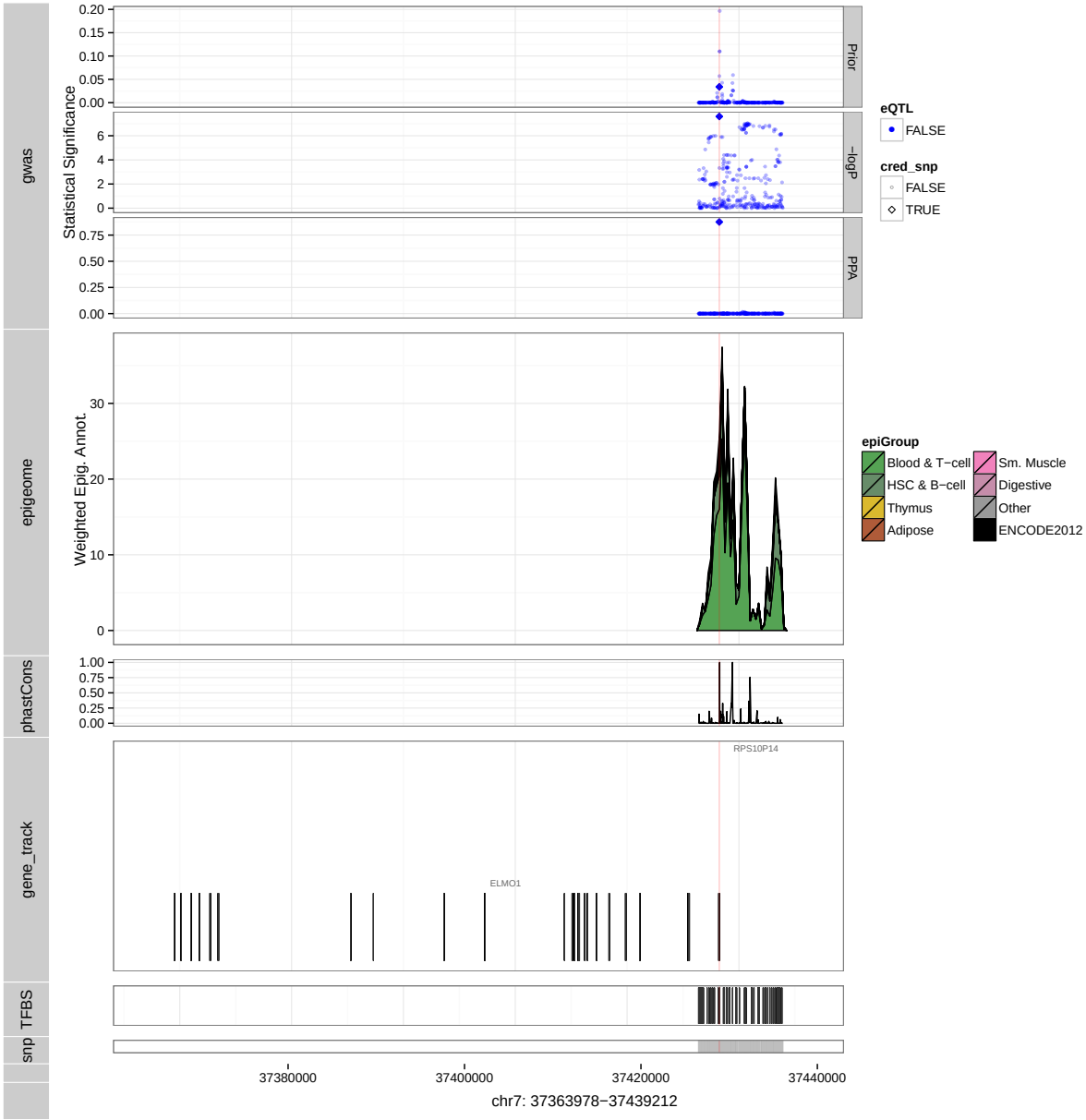
Multiple Sclerosis



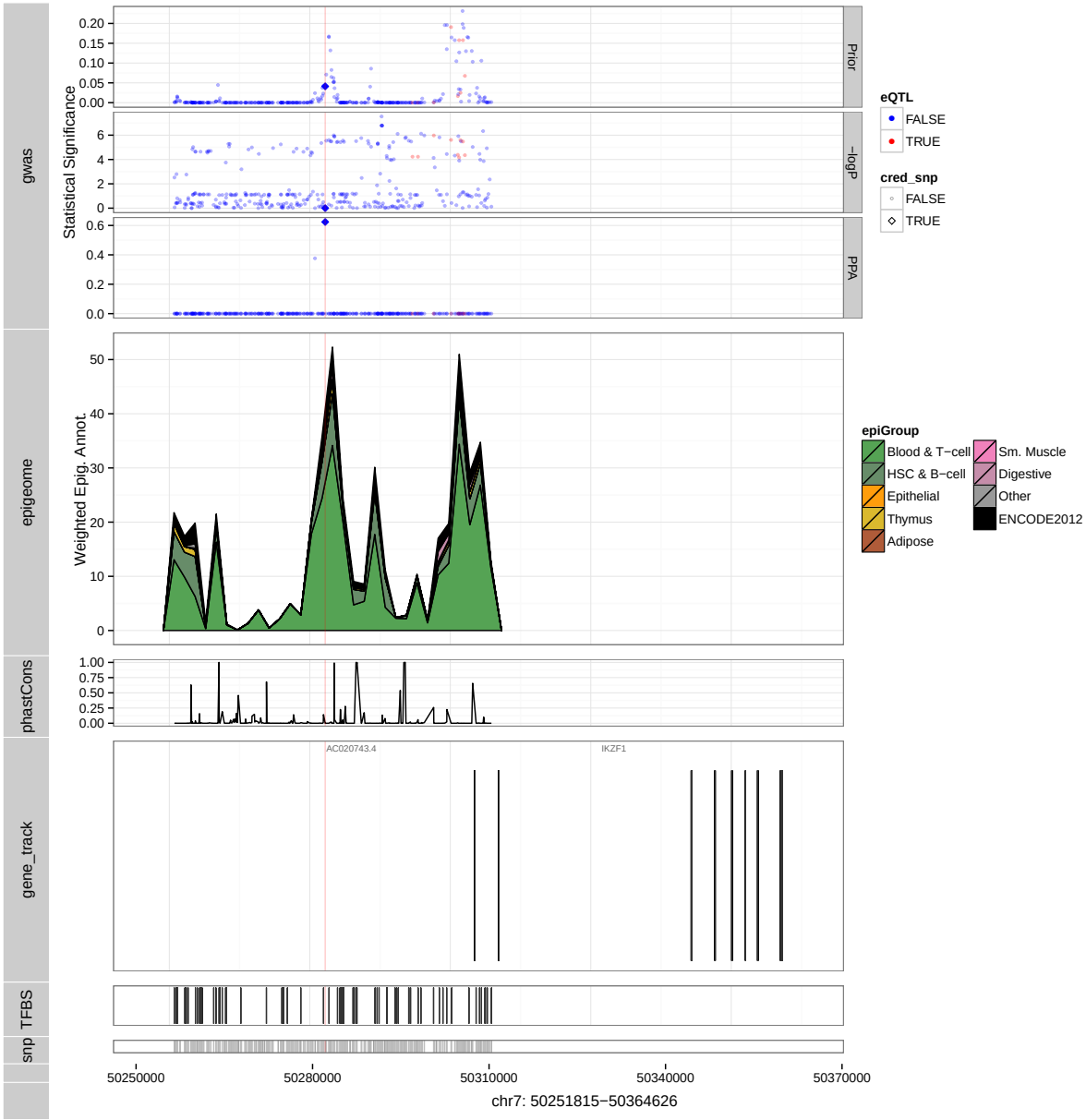
Multiple Sclerosis



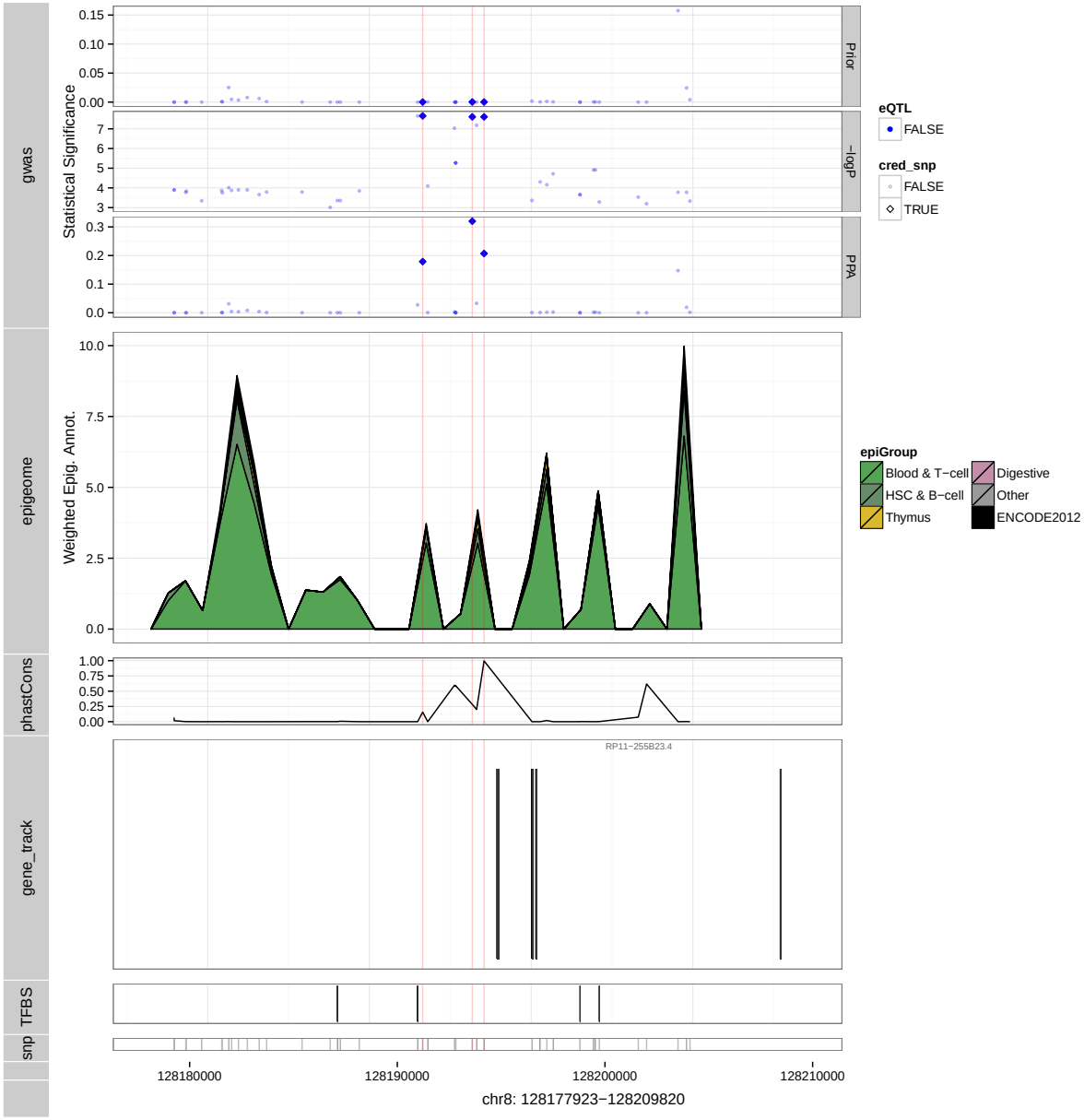
Multiple Sclerosis



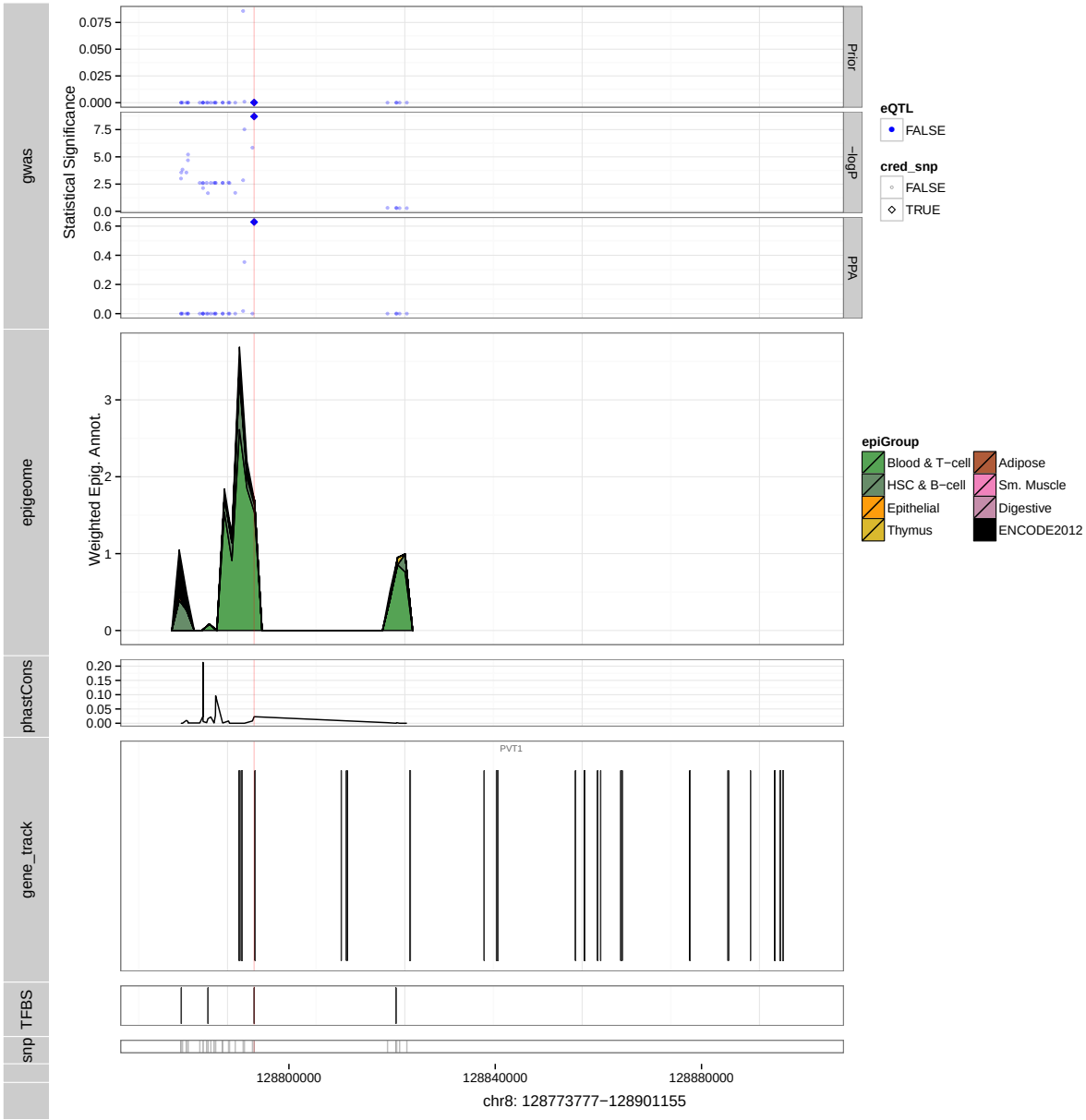
Multiple Sclerosis



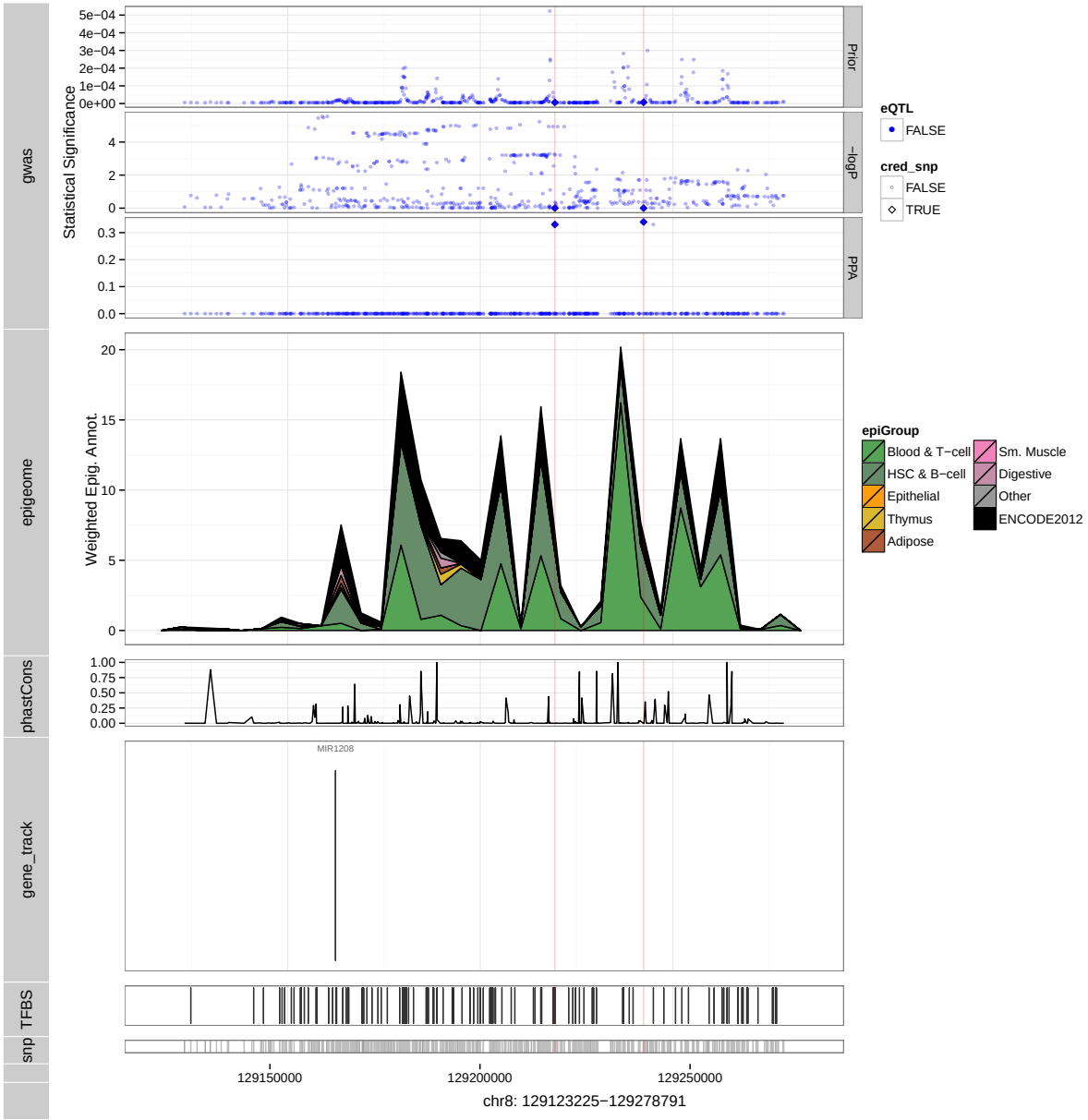
Multiple Sclerosis



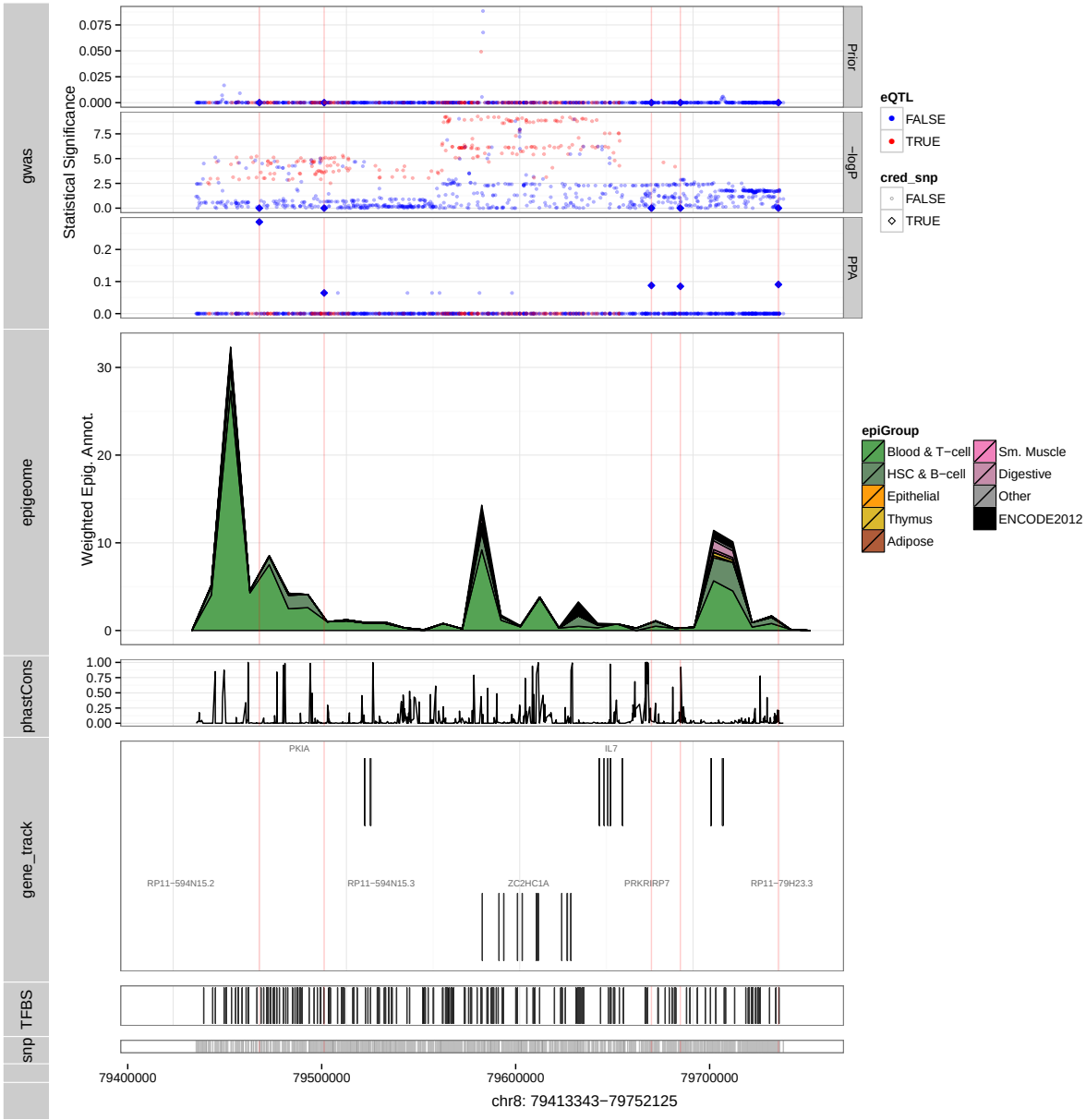
Multiple Sclerosis



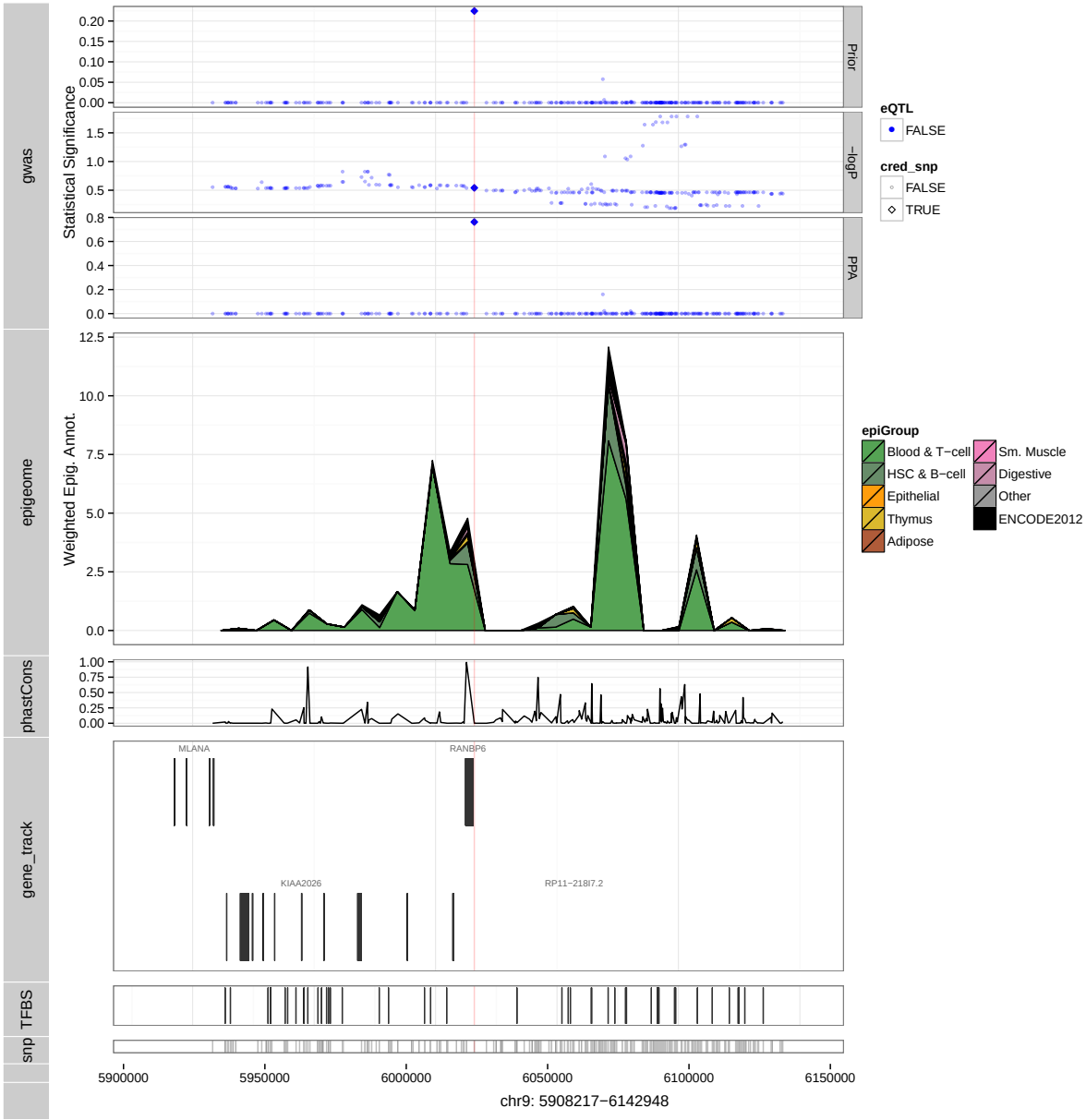
Multiple Sclerosis



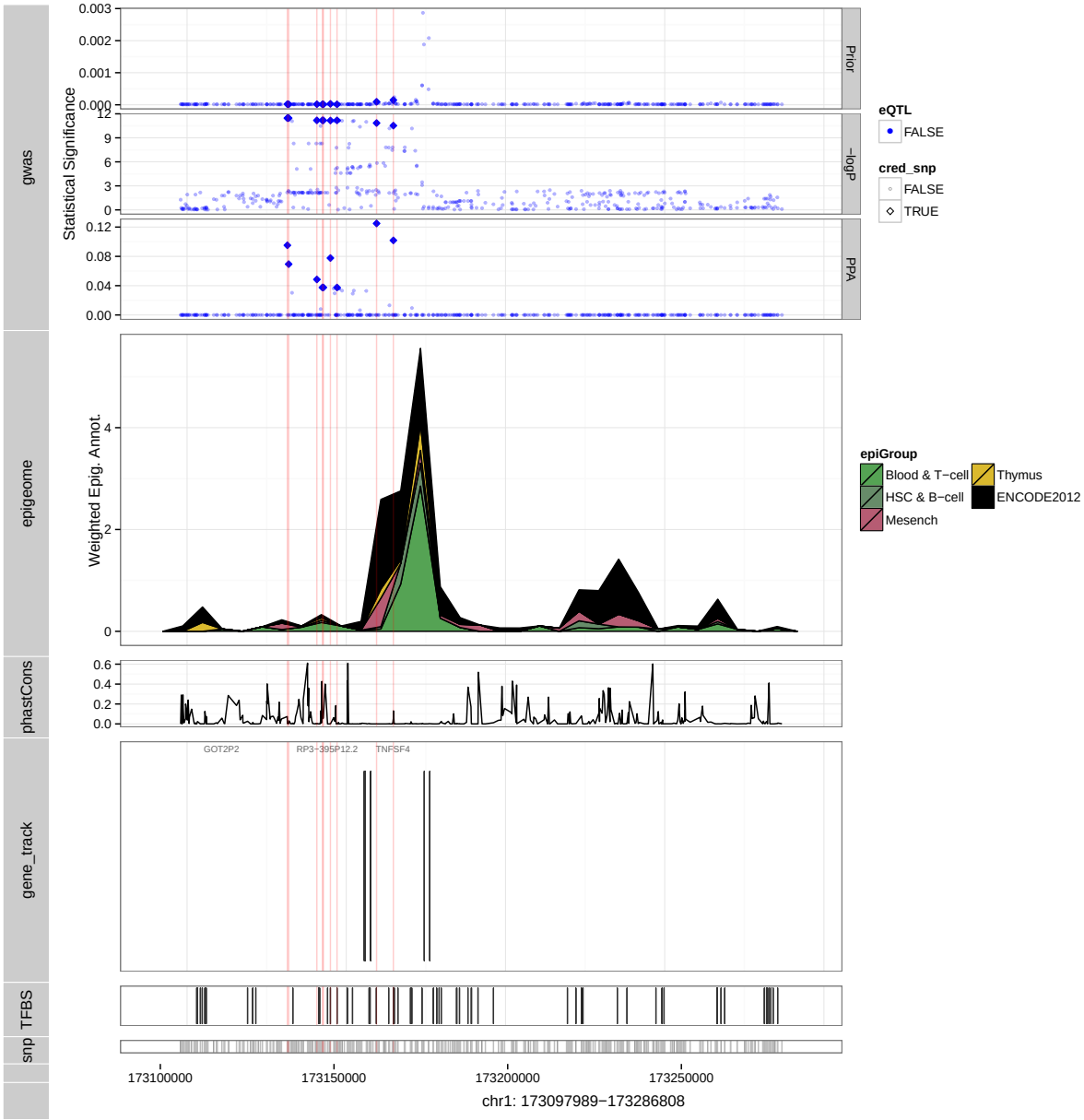
Multiple Sclerosis



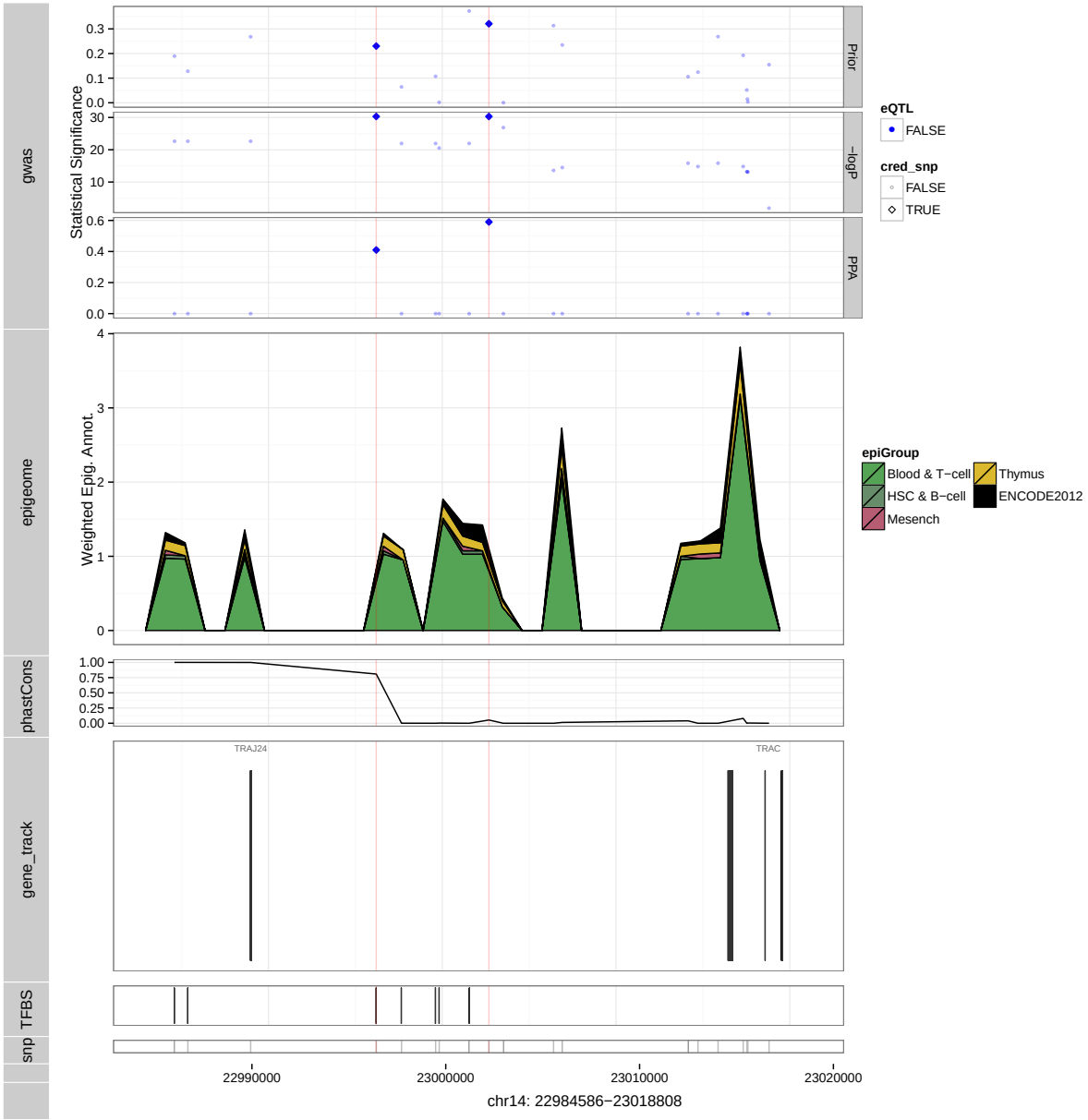
Multiple Sclerosis



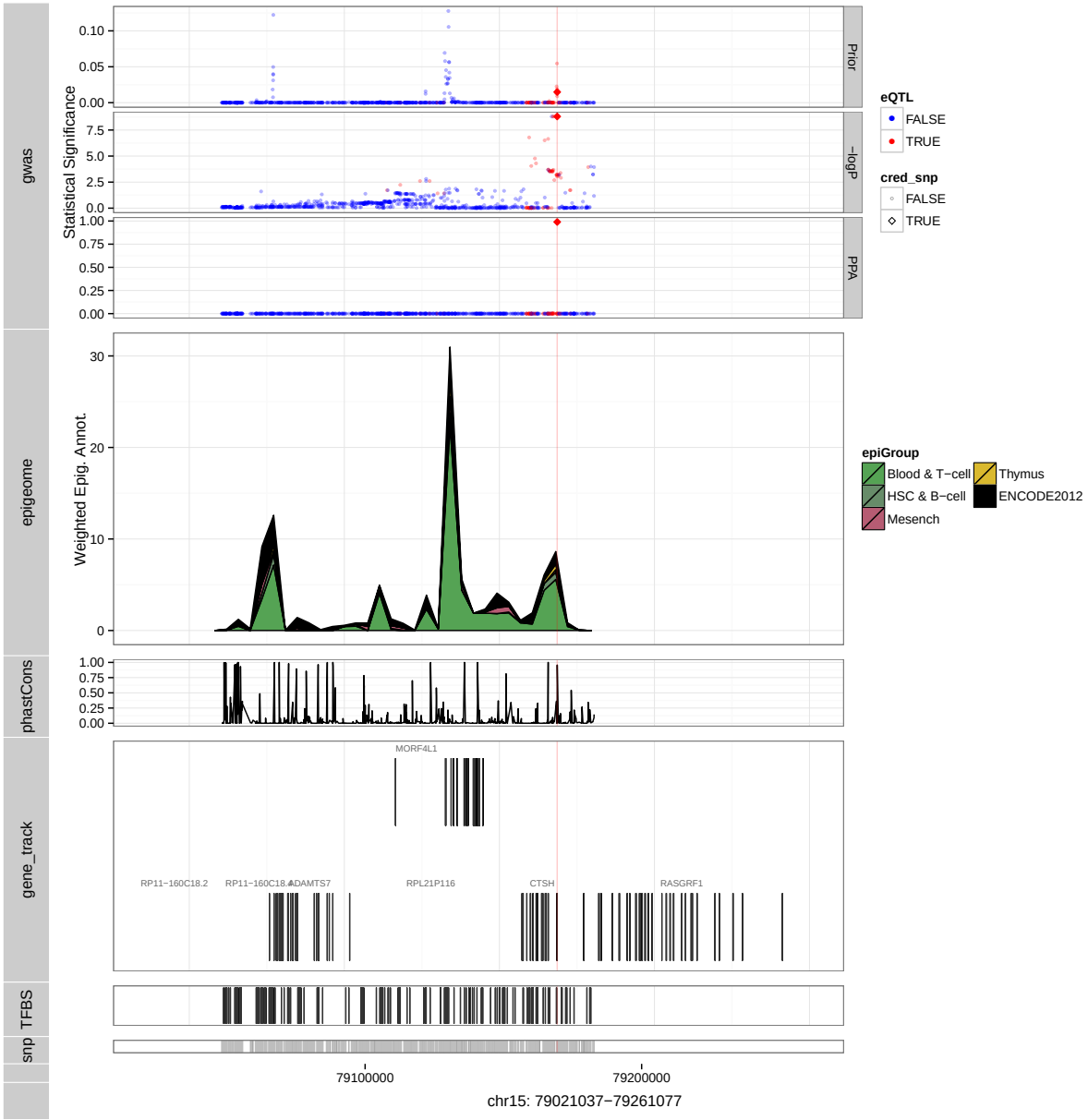
Narcolepsy


gwas
epigeome
phastCons
gene_track
snp
TFBS
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
GOT2P2
RP3−395P12.2
TNFSF4
173100000
173150000
173200000
173250000
chr1: 173097989−173286808
eQTL
FALSE
cred_snp
FALSE
TRUE
epiGroup
Blood & T−cell
HSC & B−cell
Mesench
Thymus
ENCODE2012

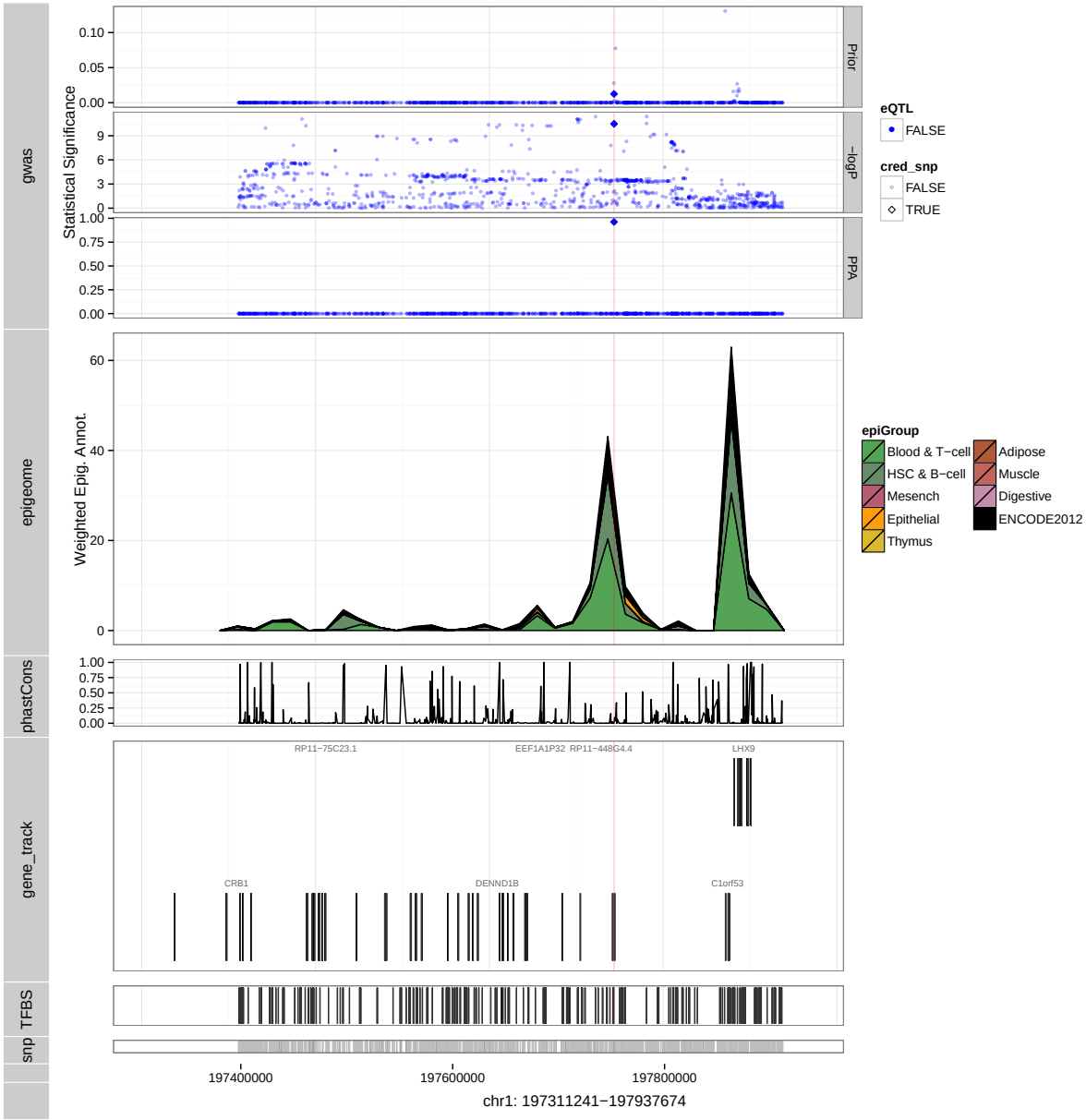
Narcolepsy


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
TRAJ24
TRAC
22990000
23000000
23010000
23020000
chr14: 22984586−23018808
eQTL
FALSE
cred_snp
FALSE
TRUE
epiGroup
Blood & T−cell
HSC & B−cell
Mesench
Thymus
ENCODE2012

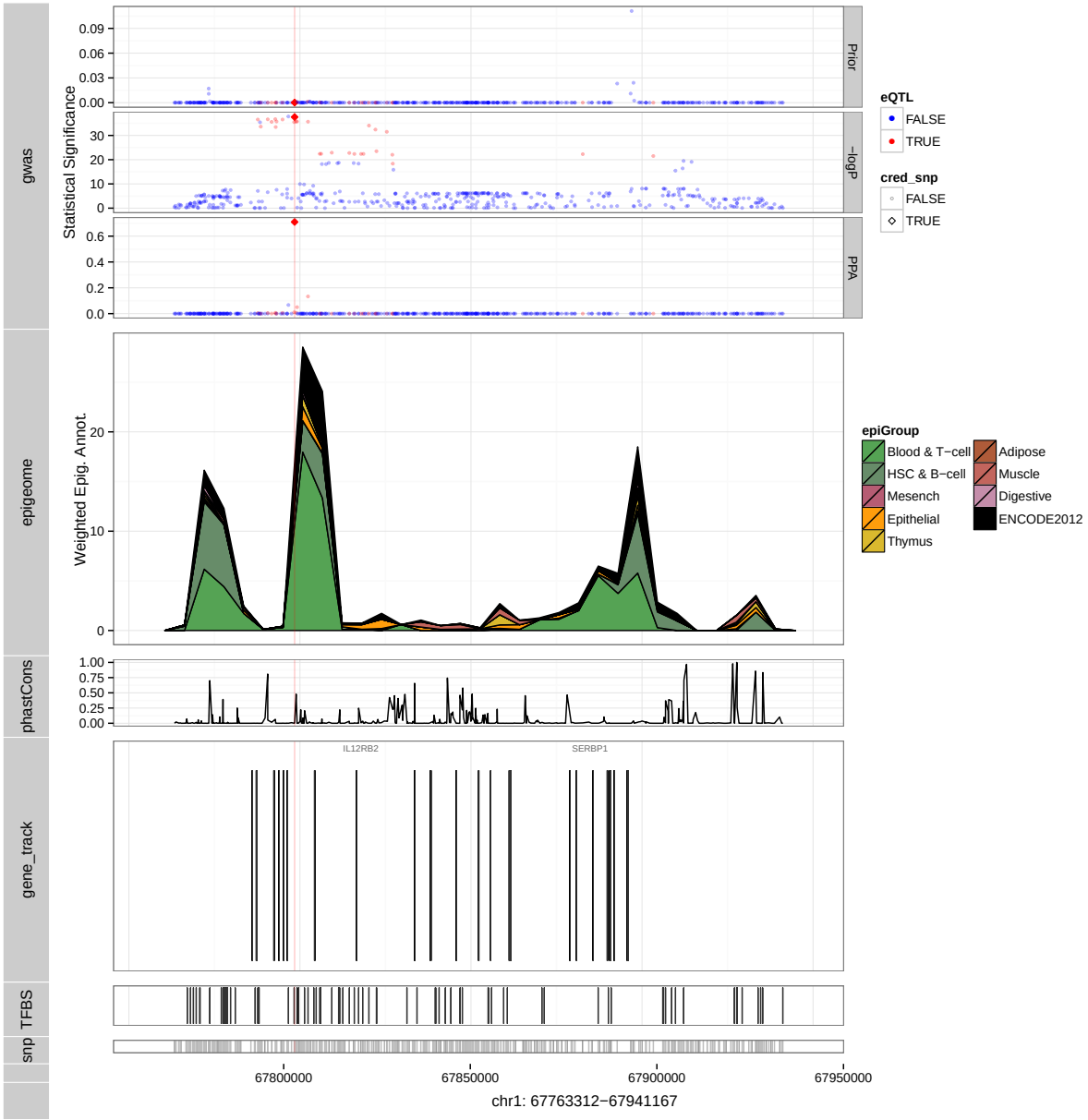
Narcolepsy


gwas
epigeome
phastCons
gene_track
snp
TFBS
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
epiGroup
Blood & T−cell
HSC & B−cell
Mesench
Thymus
ENCODE2012
MORF4L1
RP11−160C18.2
RP11−160C18.4
ADAMTS7
RPL21P116
CTSH
RASGRF1
79100000
79200000
chr15: 79021037−79261077

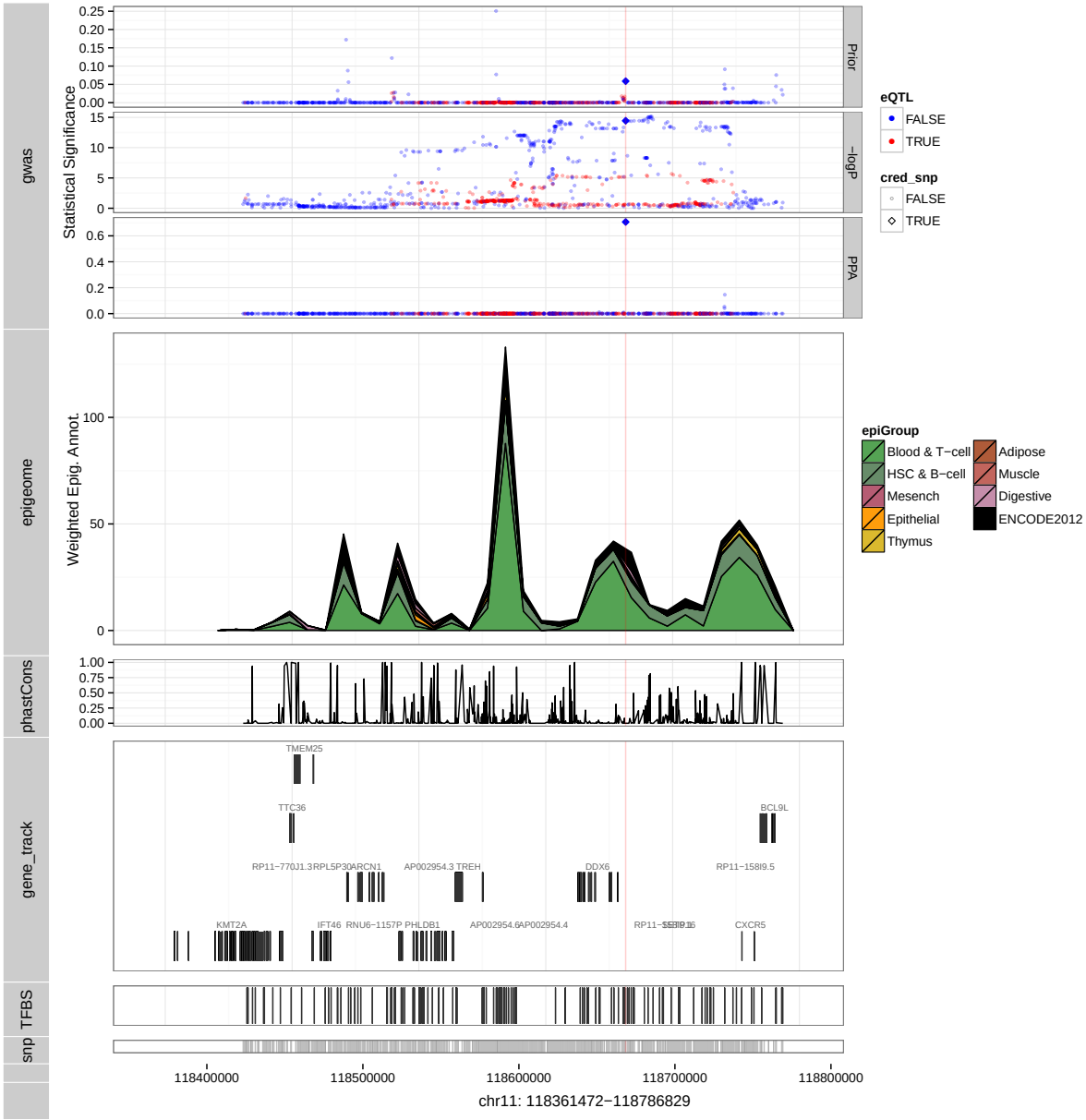
Primary Biliary Cirrhosis



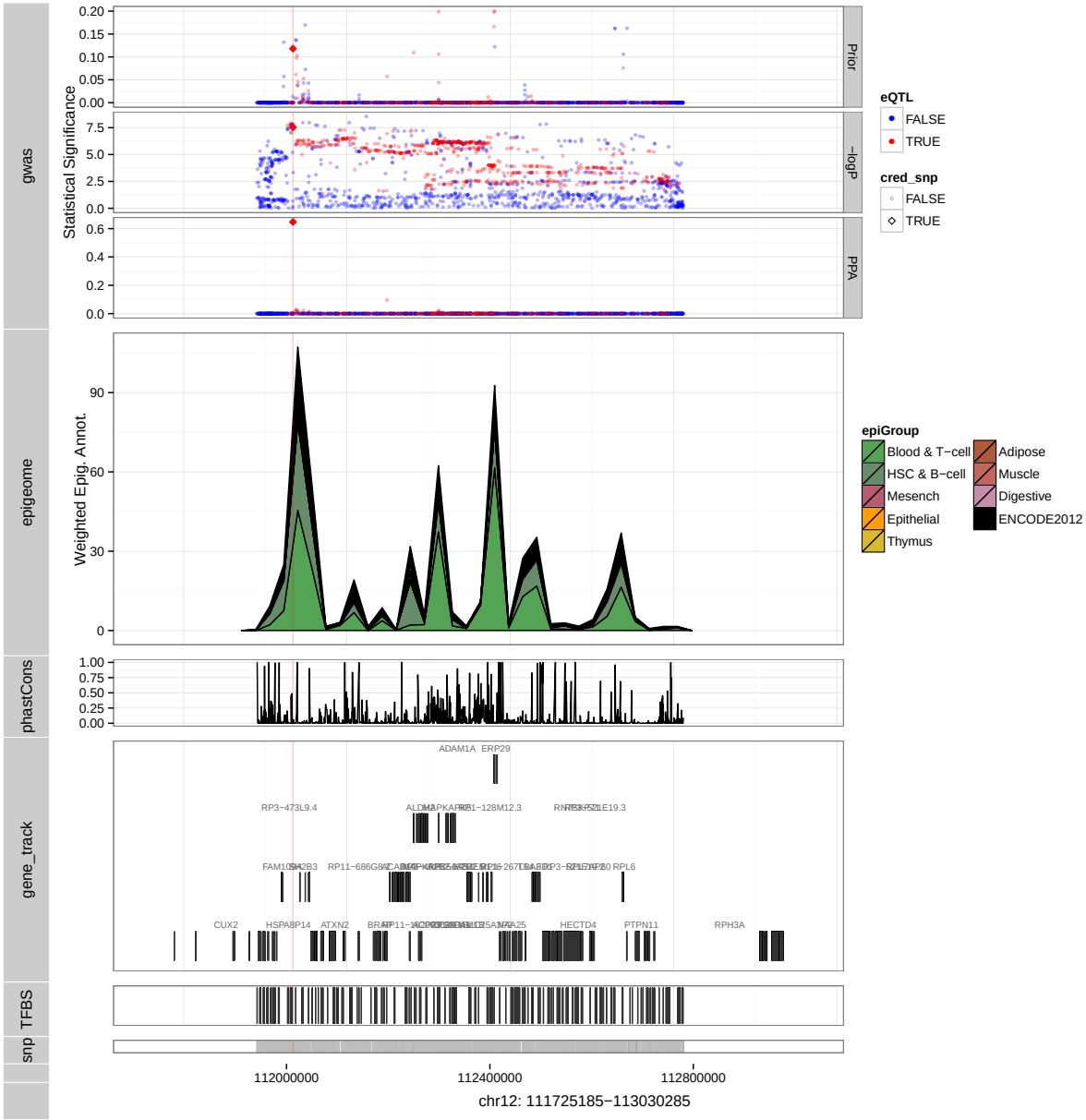
Primary Biliary Cirrhosis



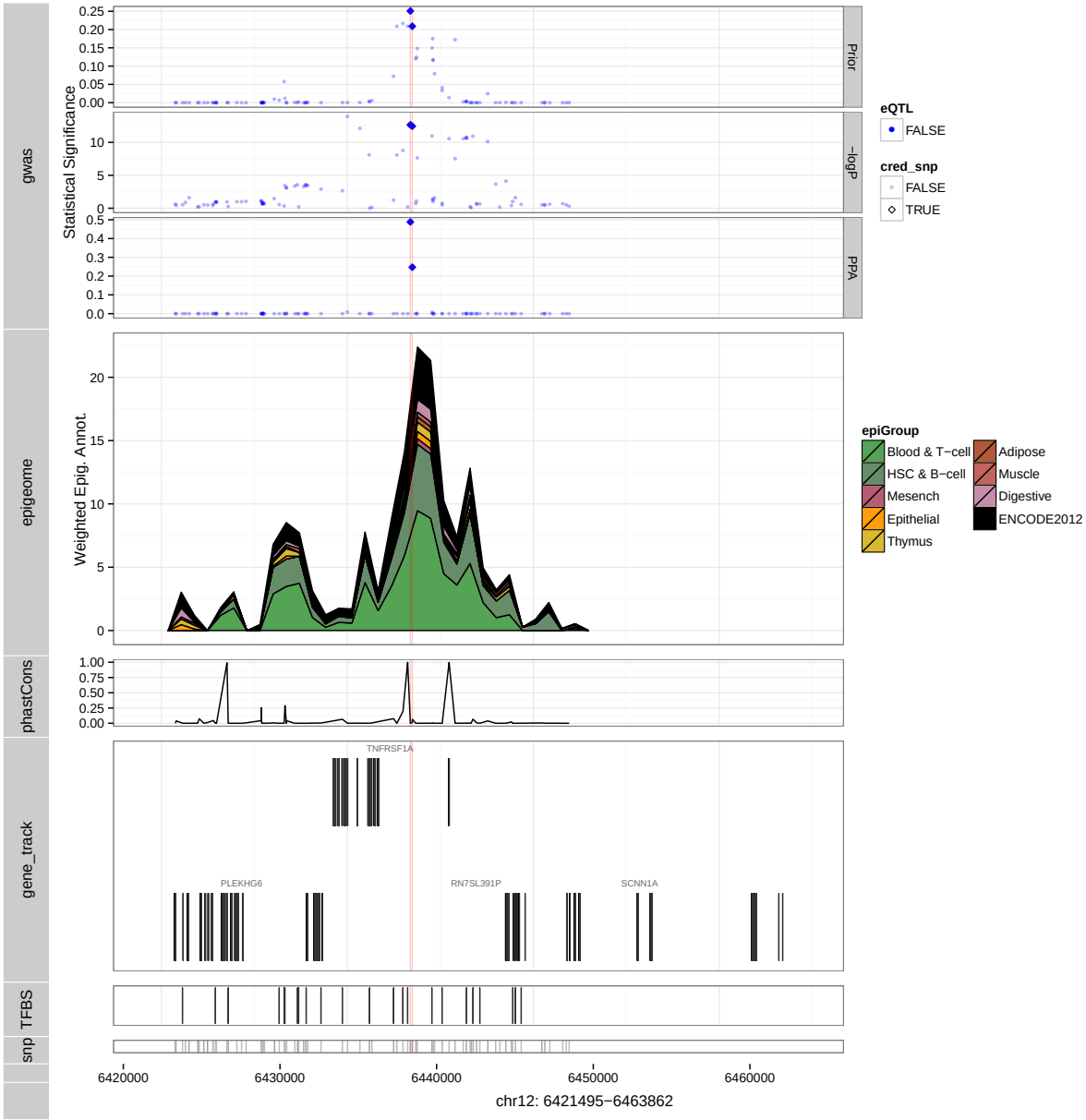
Primary Biliary Cirrhosis



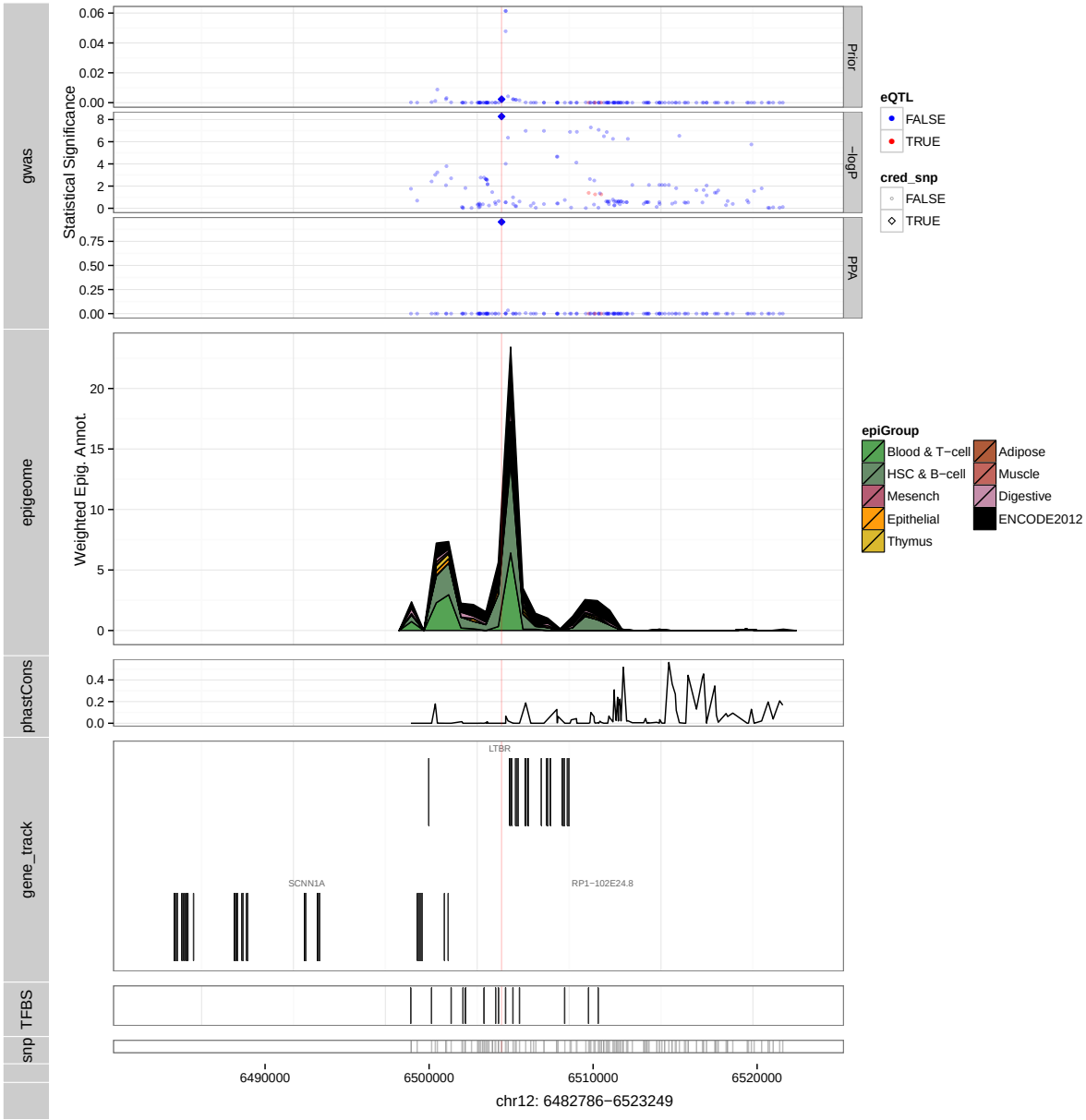
Primary Biliary Cirrhosis



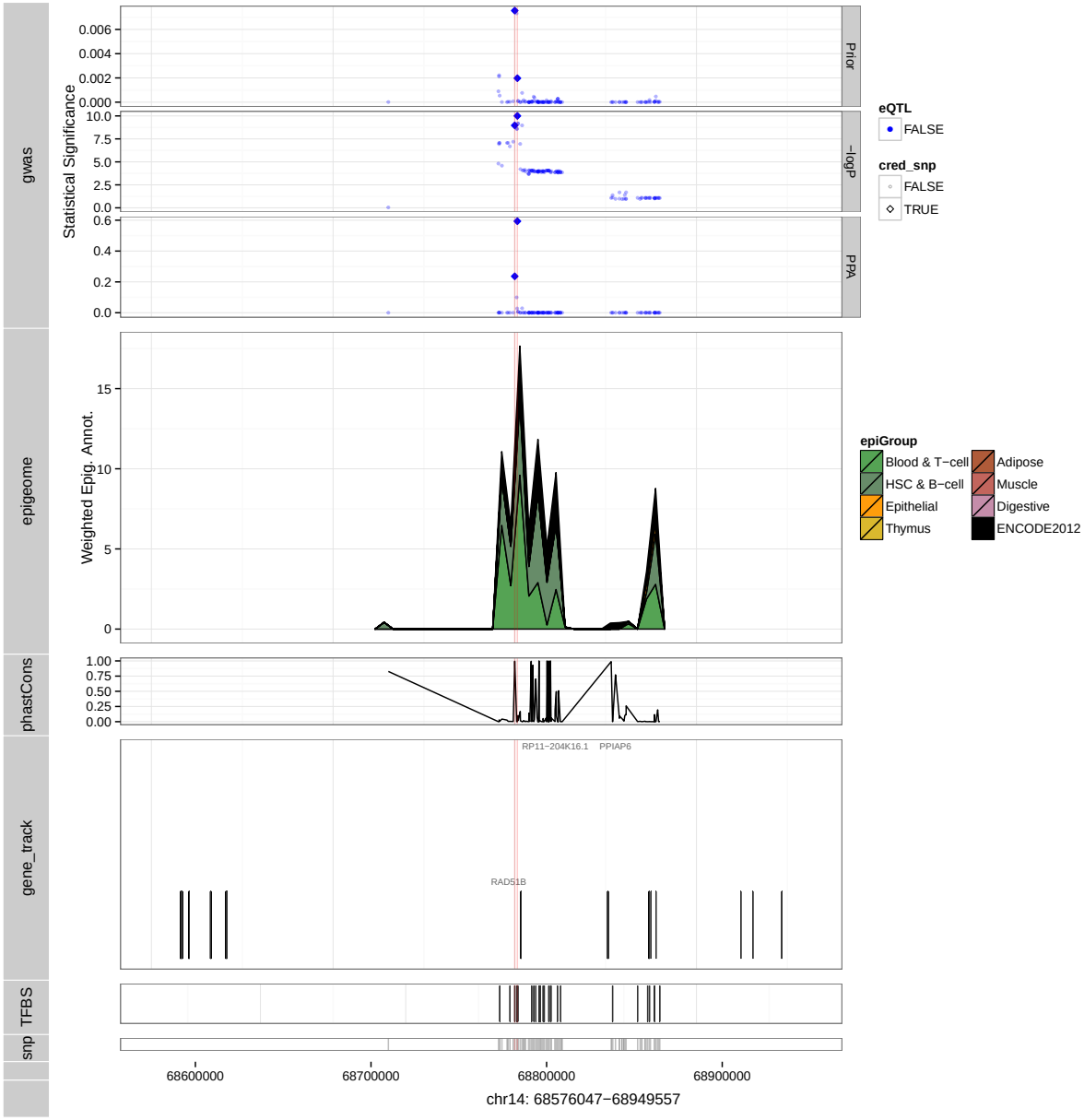
Primary Biliary Cirrhosis



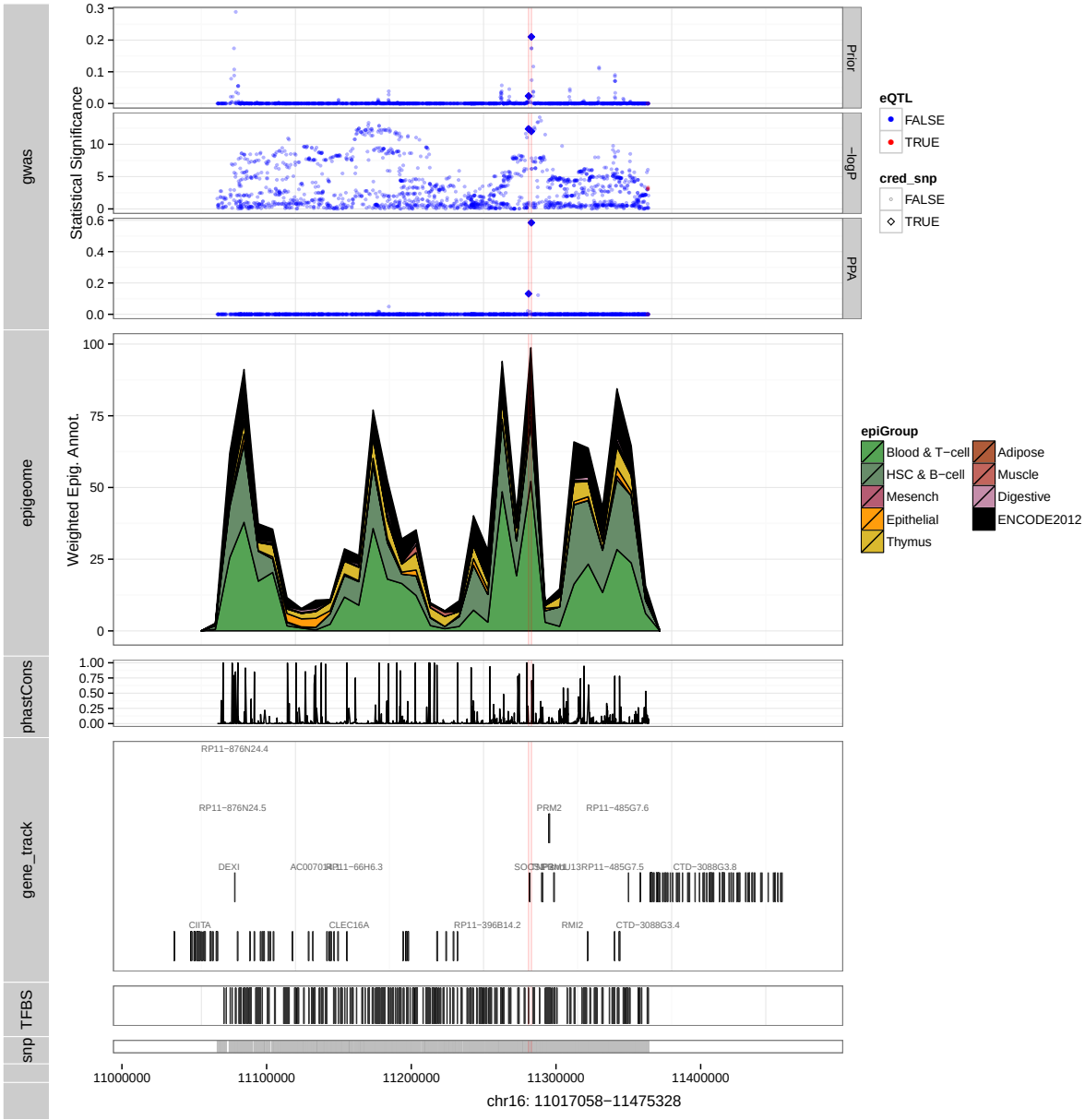
Primary Biliary Cirrhosis



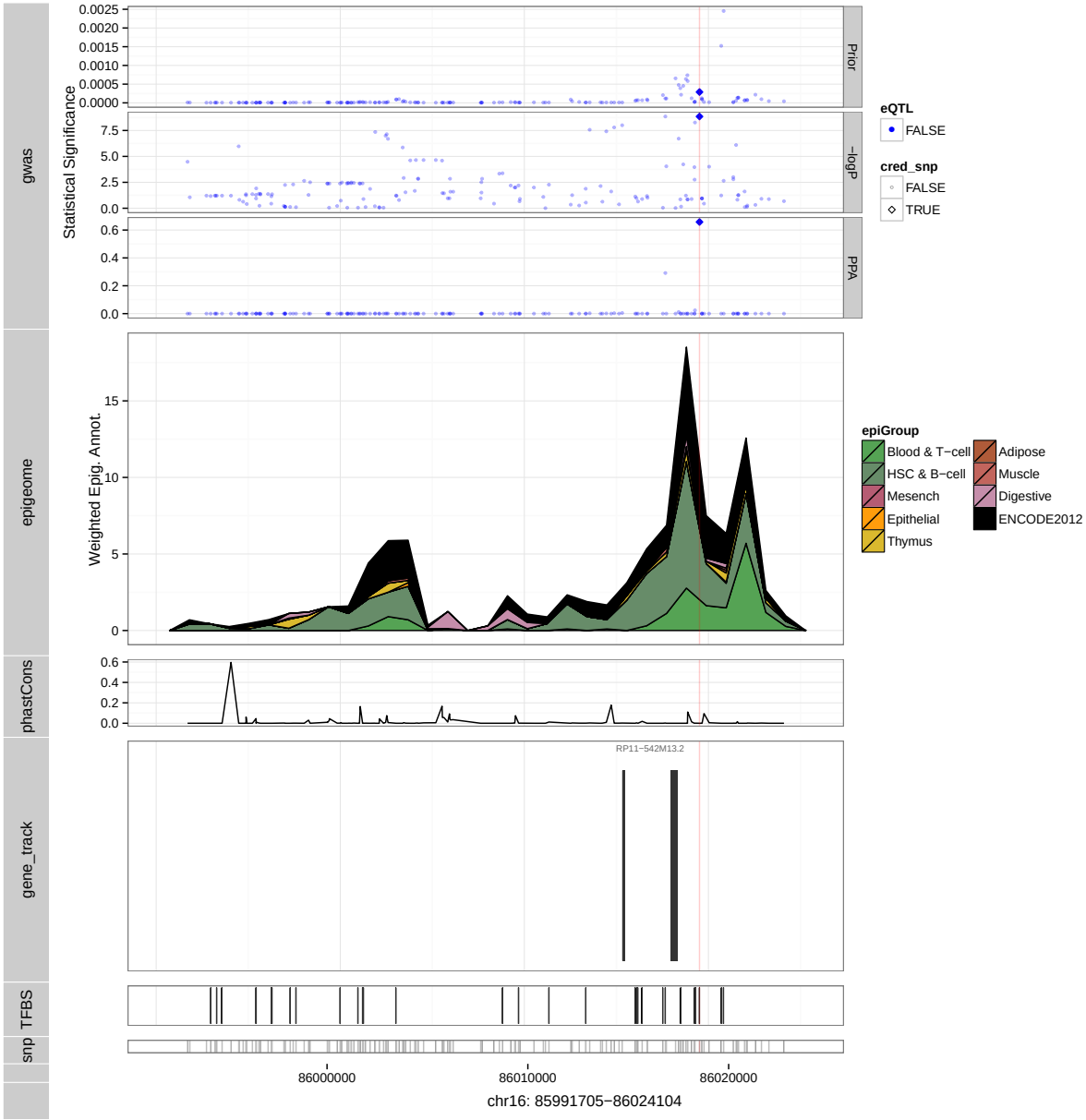
Primary Biliary Cirrhosis



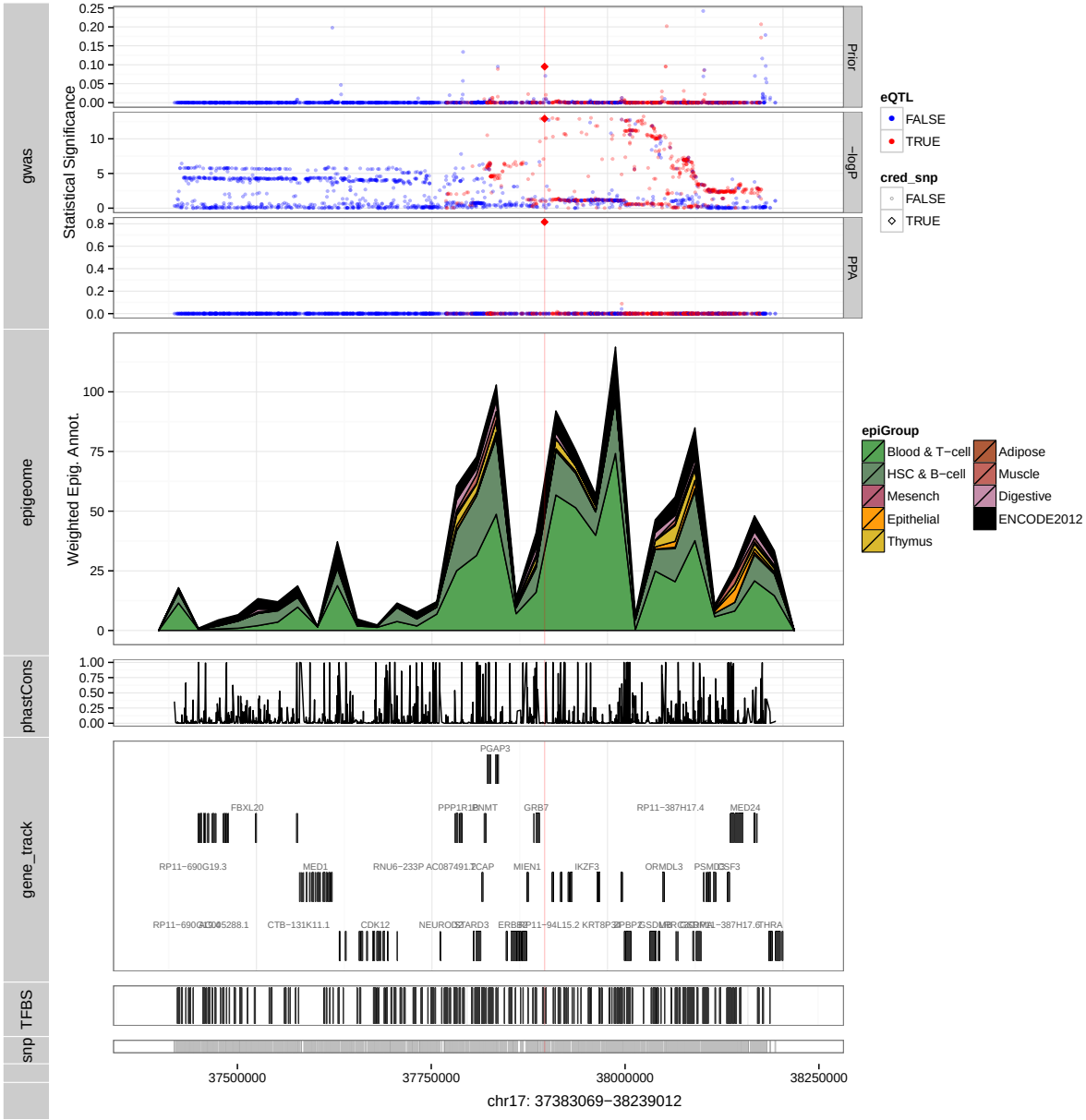
Primary Biliary Cirrhosis



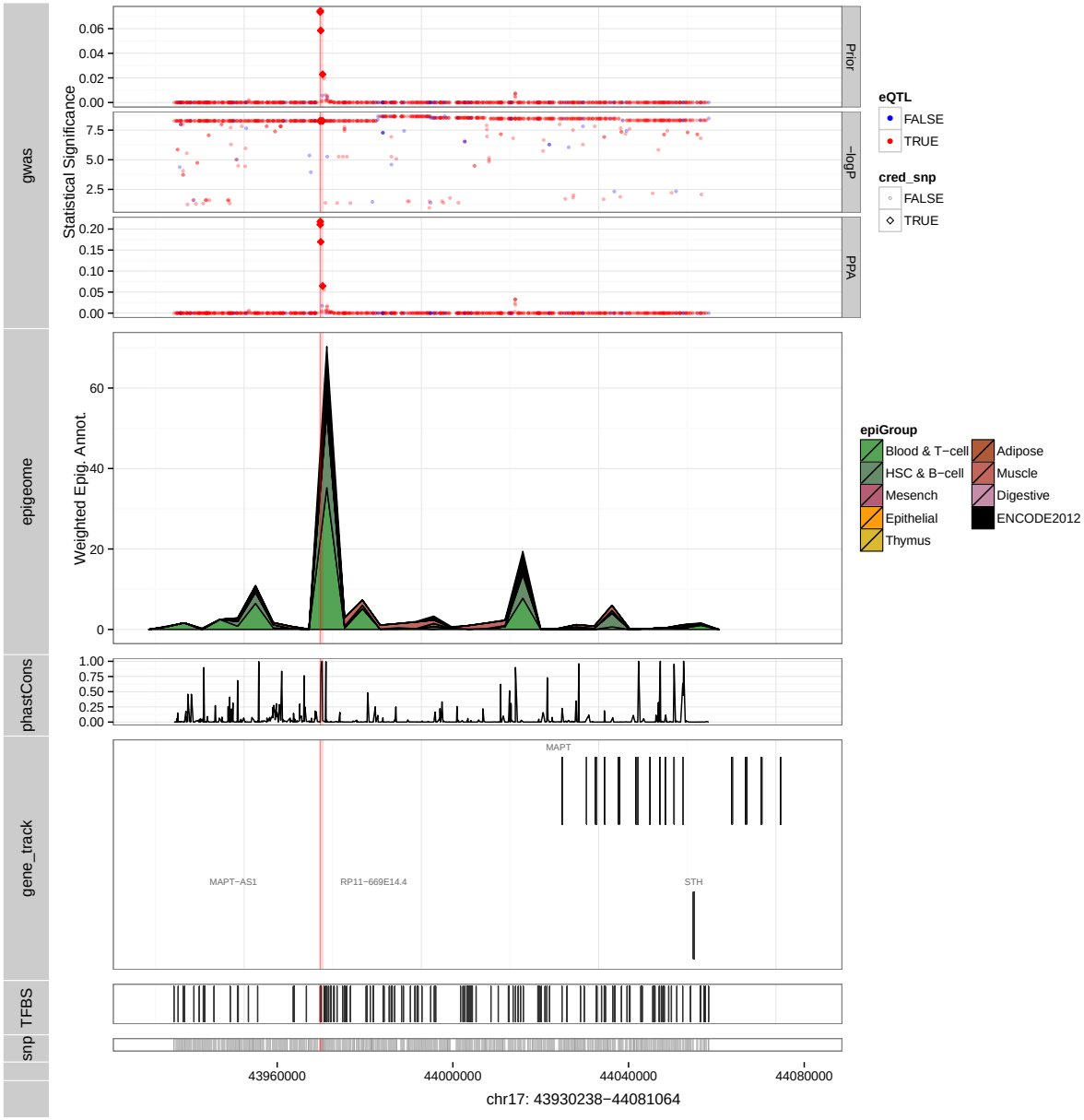
Primary Biliary Cirrhosis



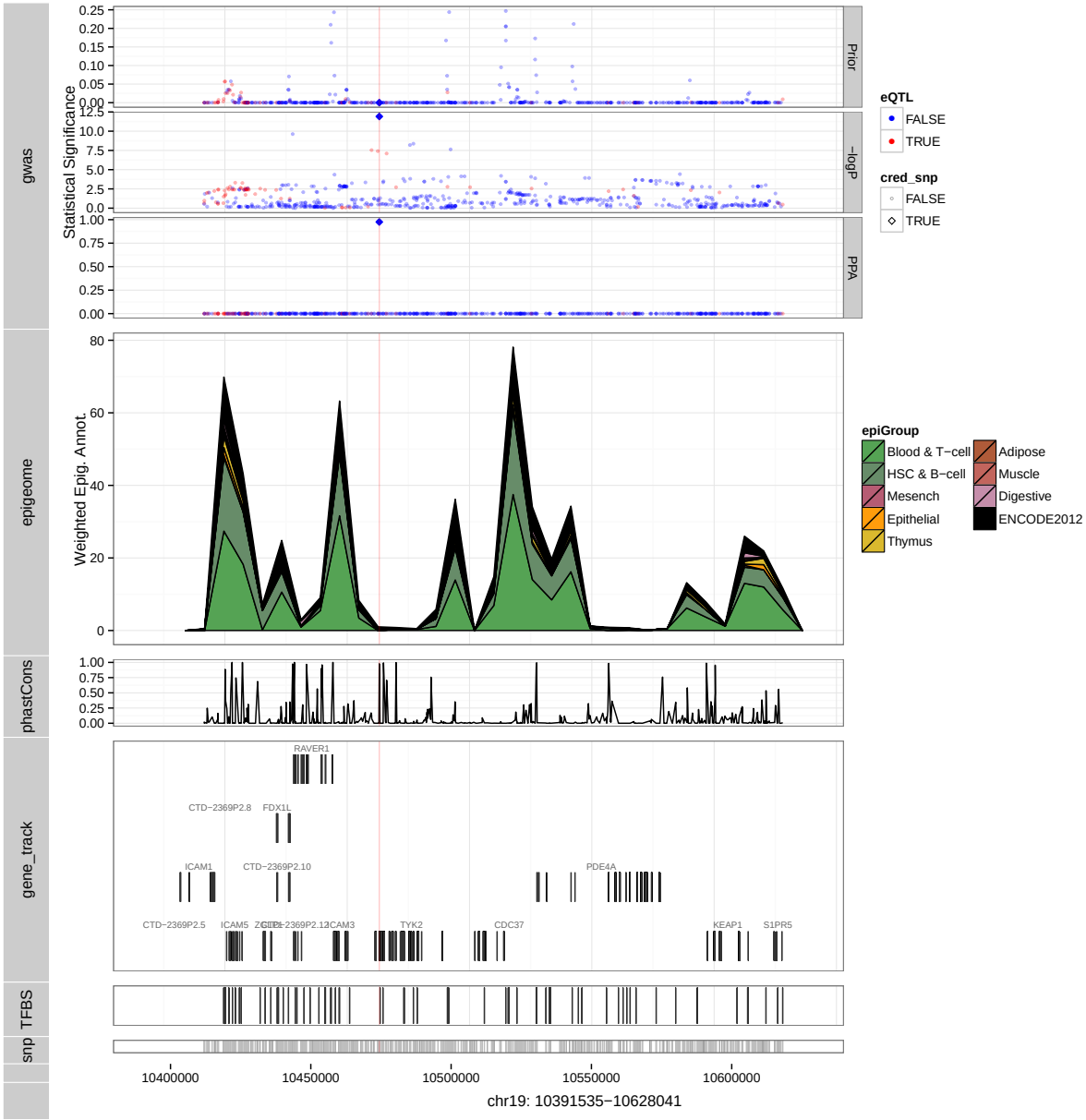
Primary Biliary Cirrhosis



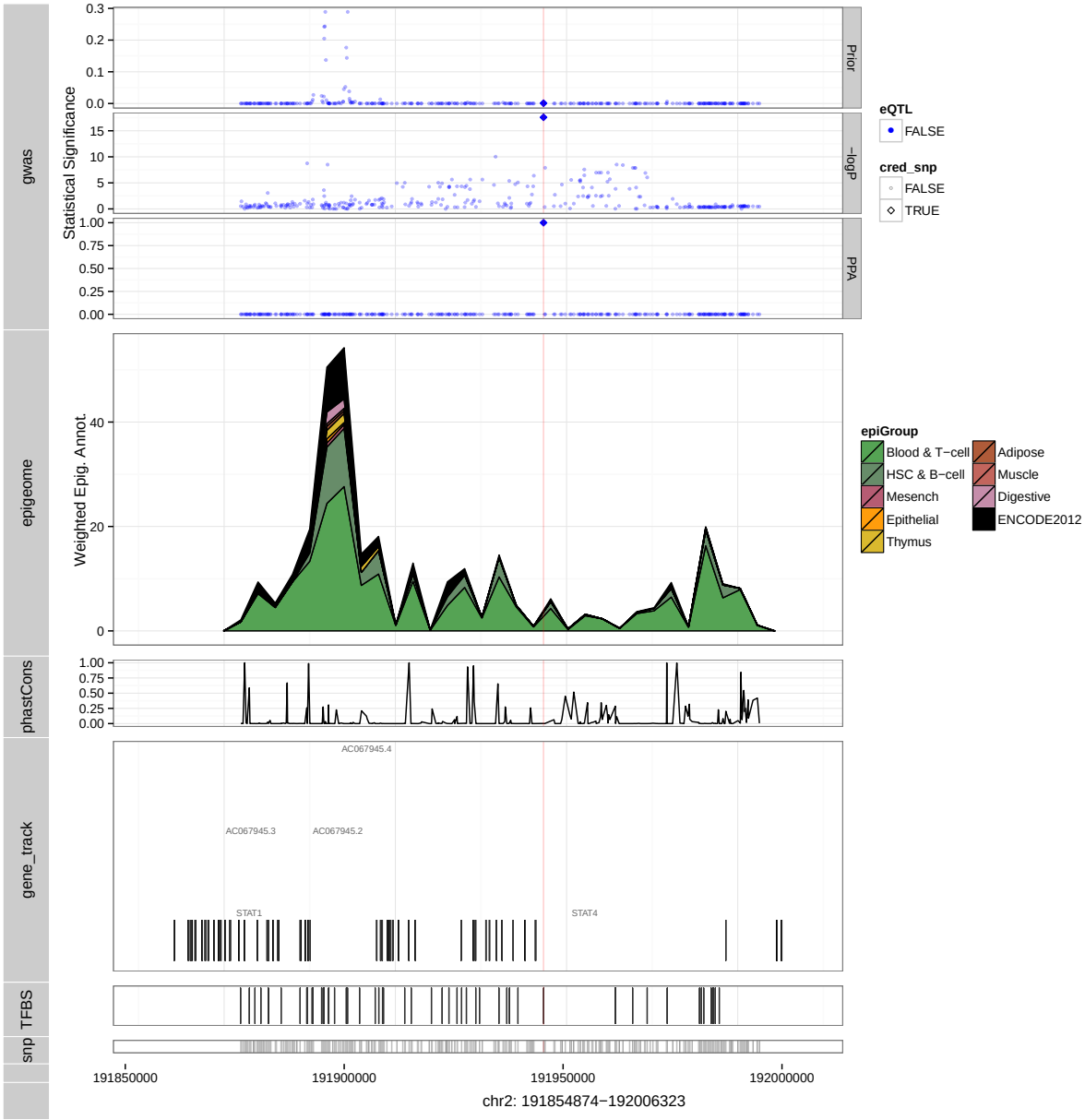
Primary Biliary Cirrhosis



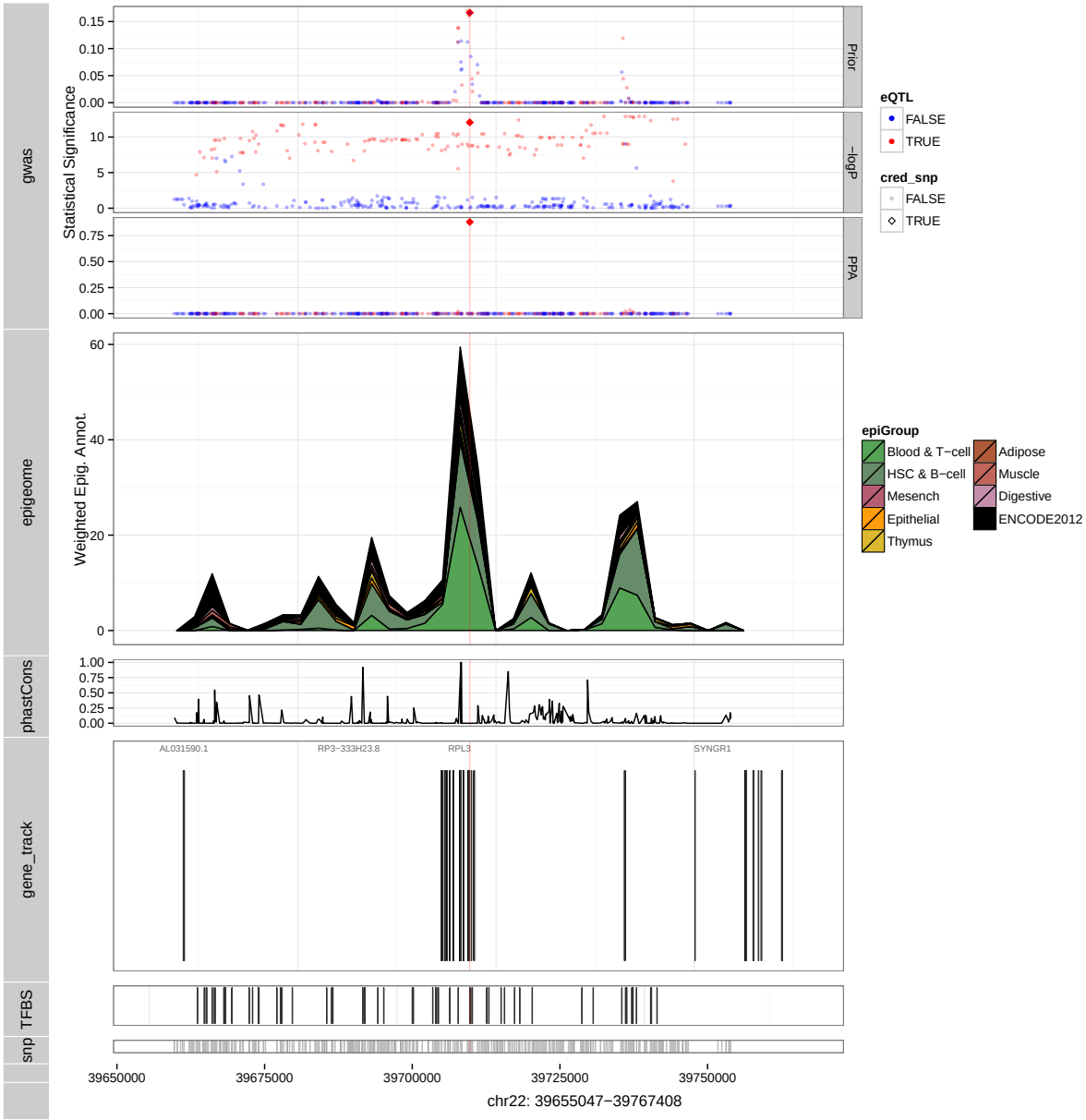
Primary Biliary Cirrhosis



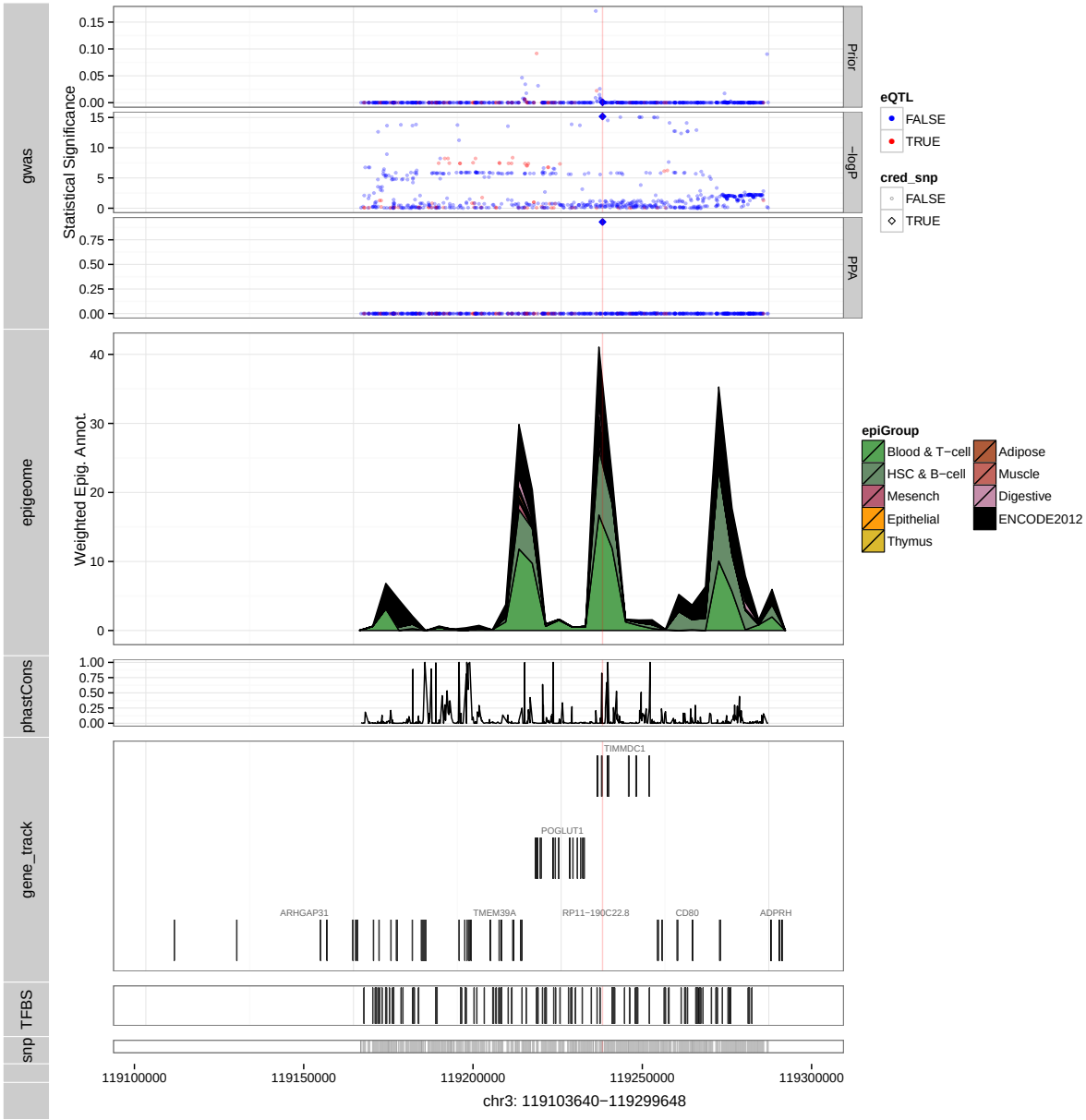
Primary Biliary Cirrhosis



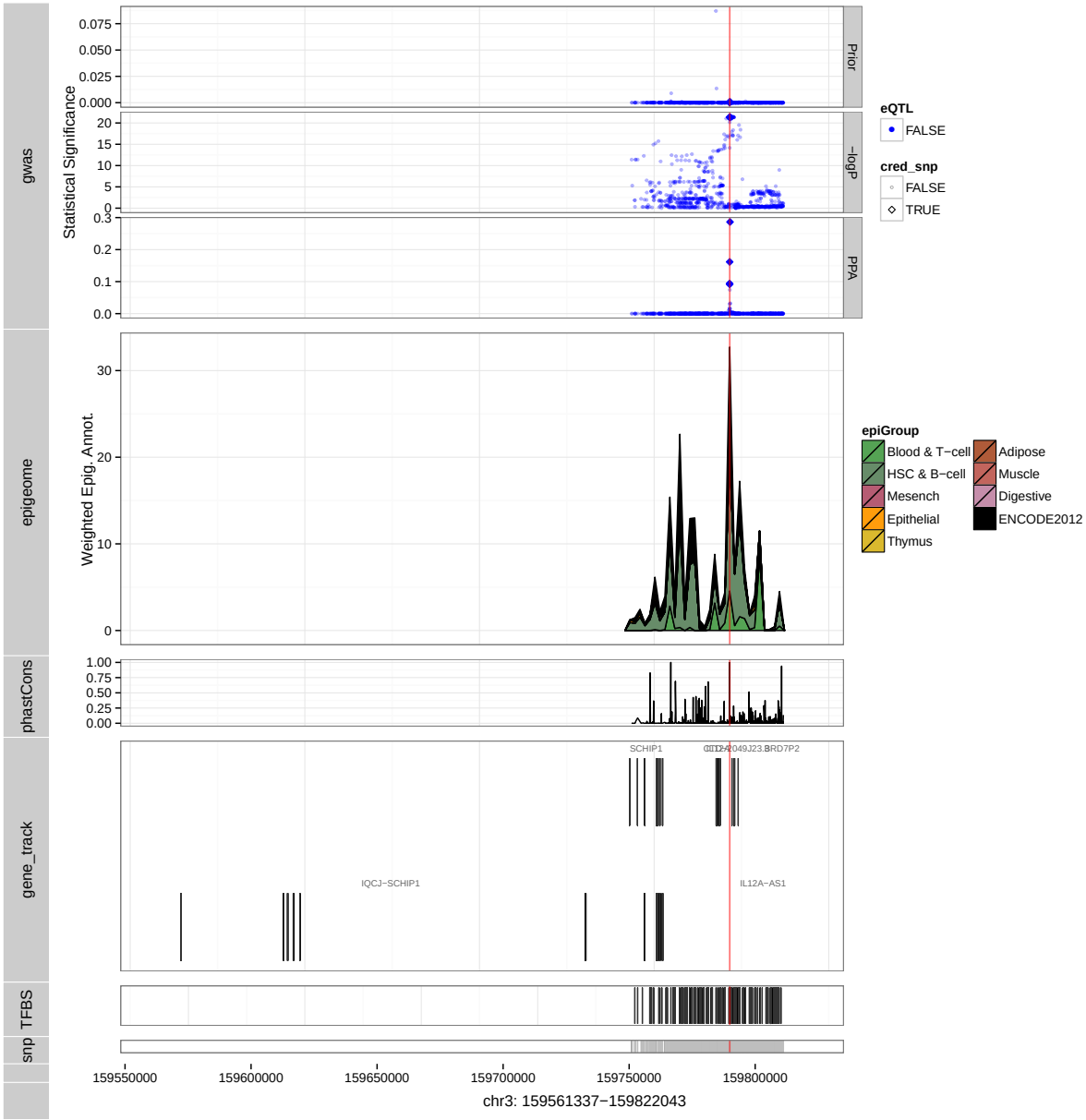
Primary Biliary Cirrhosis



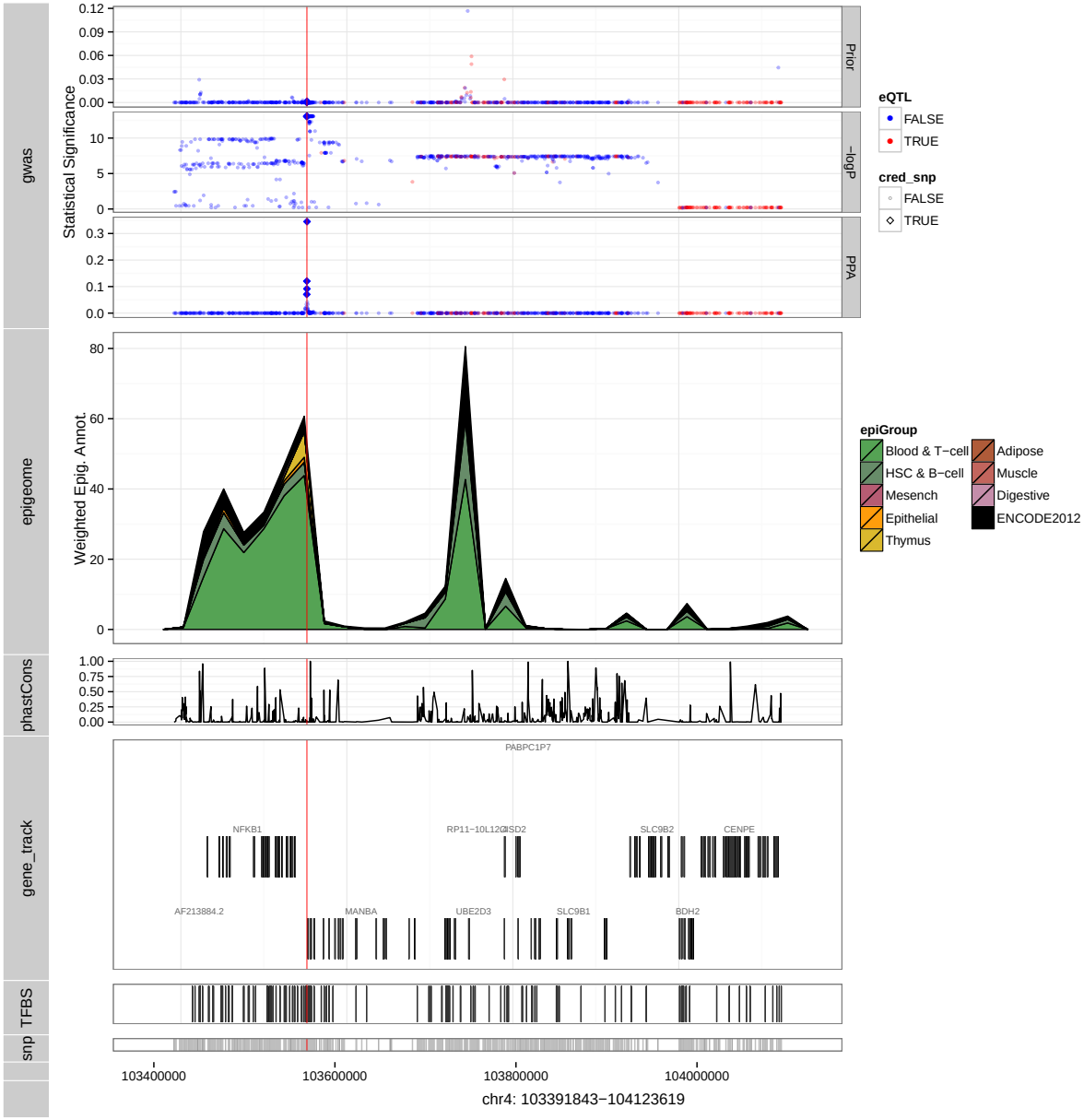
Primary Biliary Cirrhosis



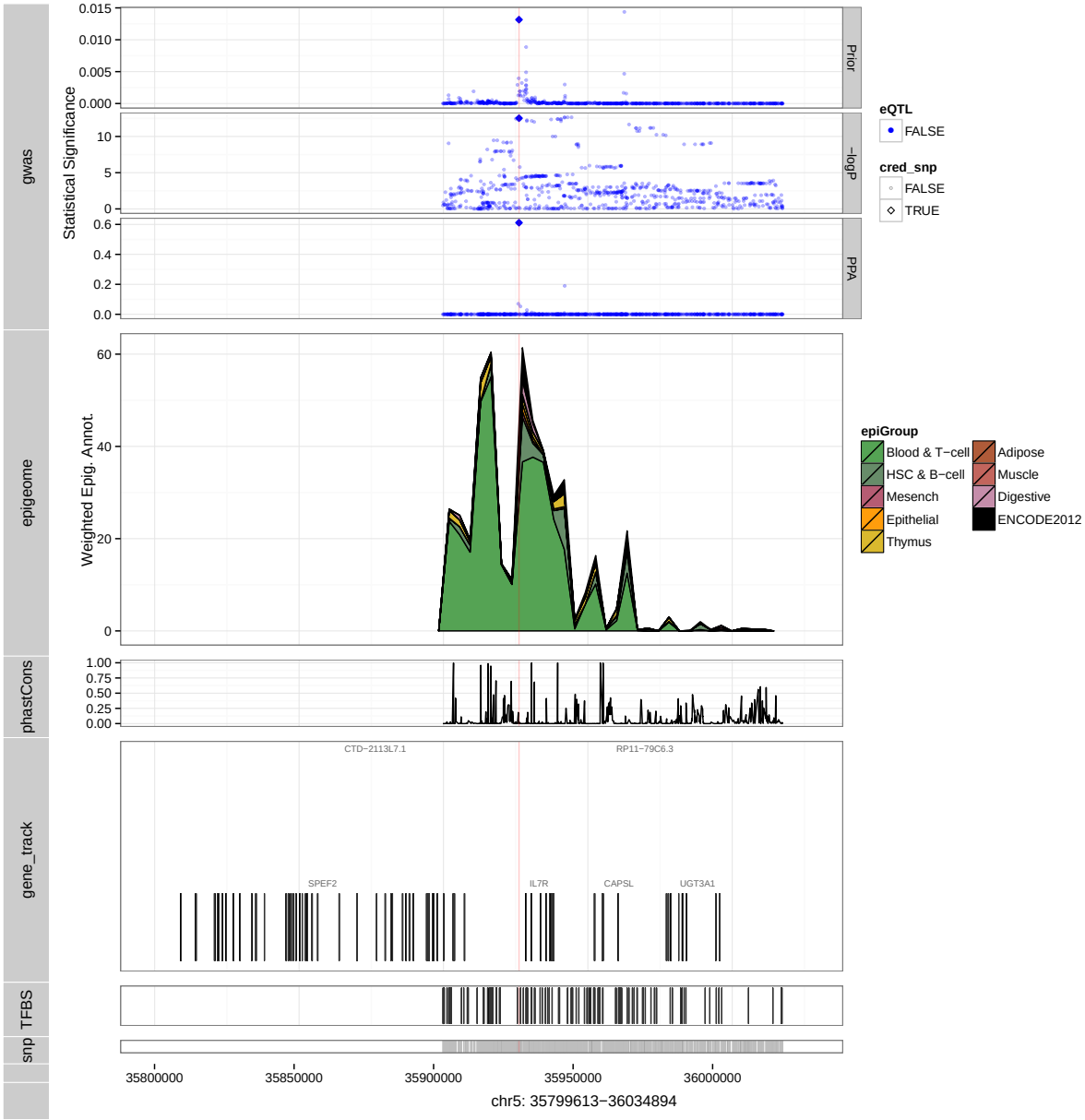
Primary Biliary Cirrhosis



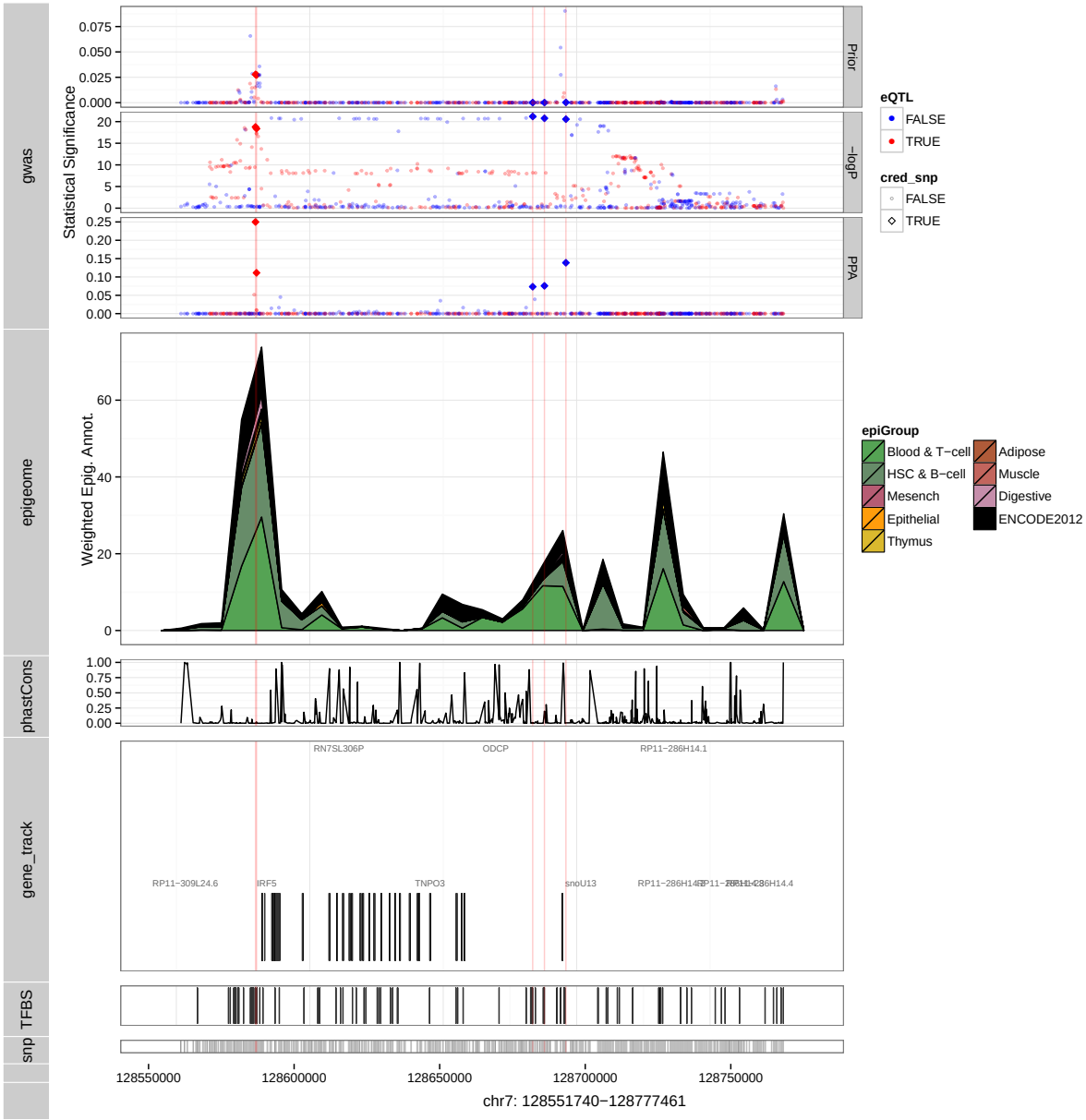
Primary Biliary Cirrhosis



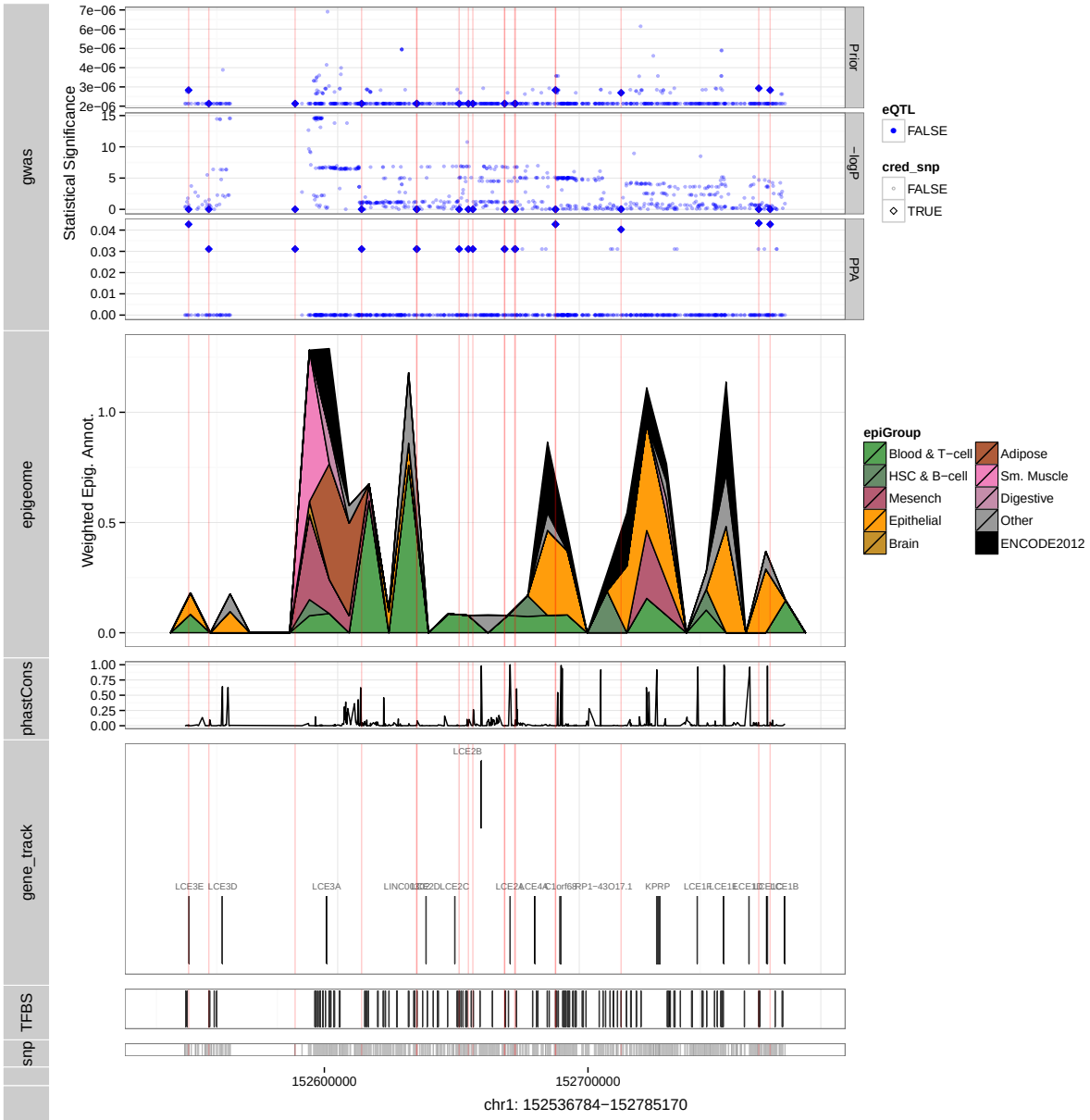
Primary Biliary Cirrhosis



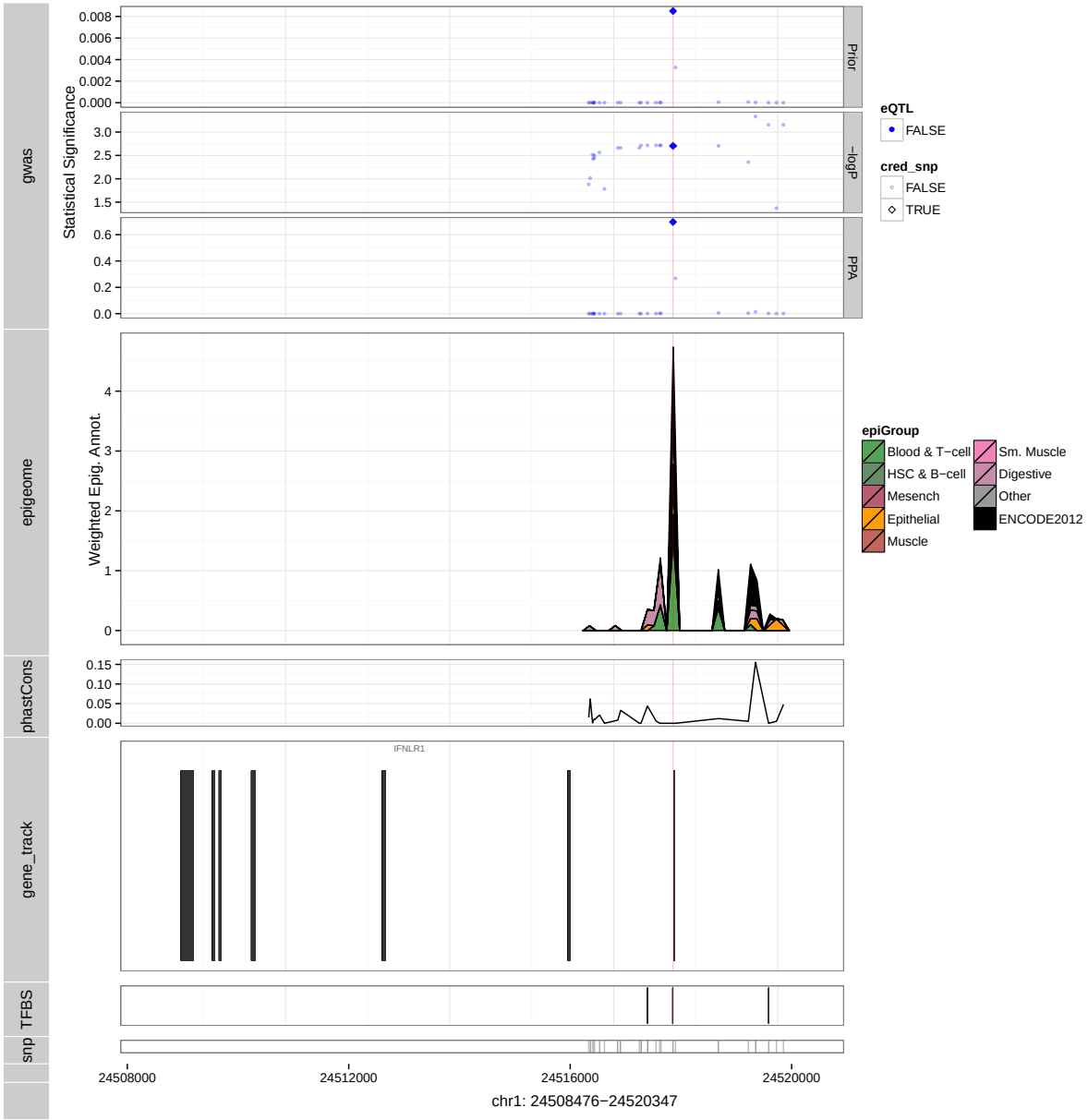
Primary Biliary Cirrhosis



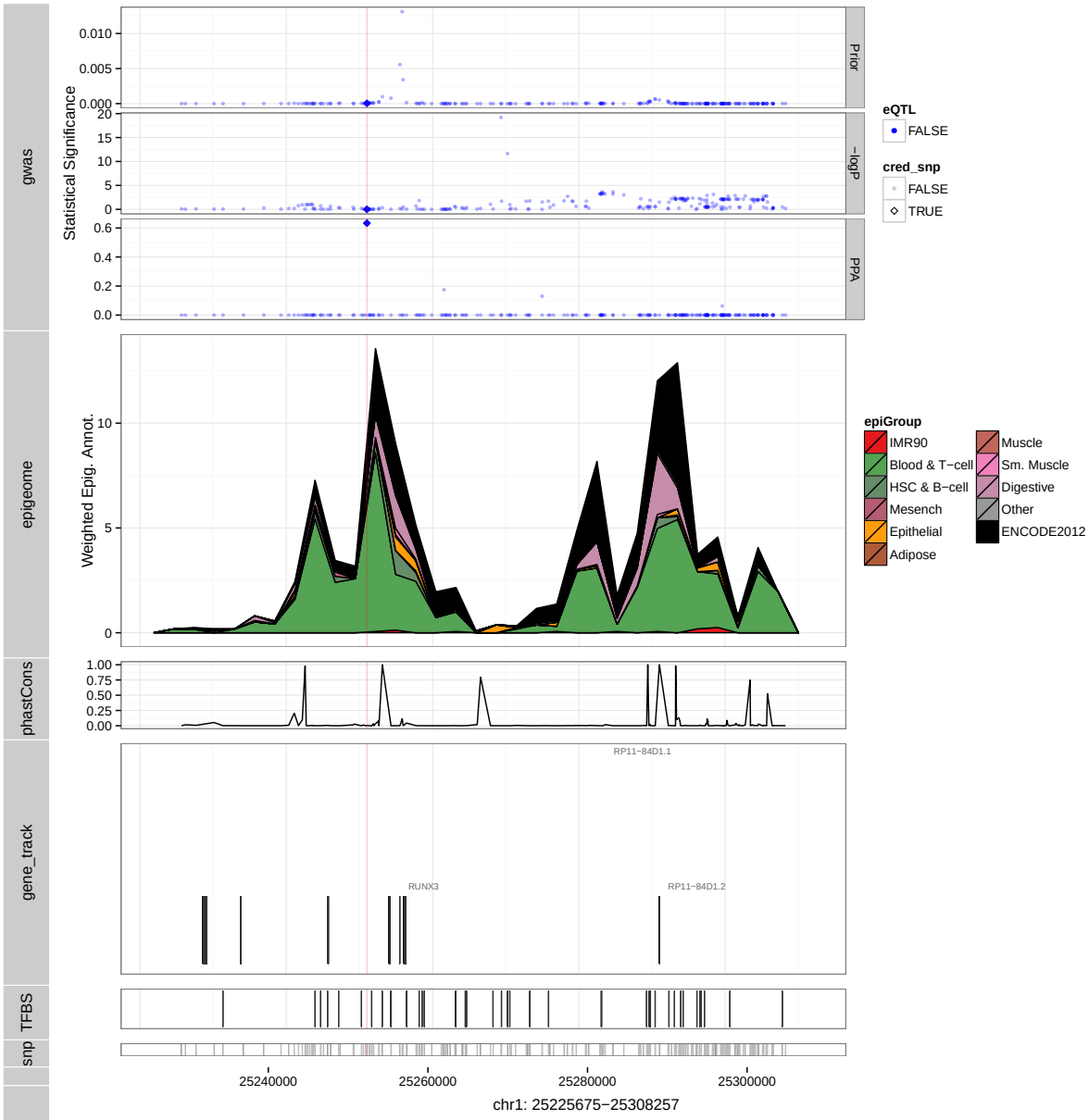
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
eQTL
FALSE
cred_snp
FALSE
TRUE
Weighted Epig. Annot.
epiGroup
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Sm. Muscle
Digestive
Other
ENCODE2012
LCE2B
LCE3E
LCE3D
LCE3A
LCE2D
LCE2C
KPRP
152600000
152700000
chr1: 152536784−152785170

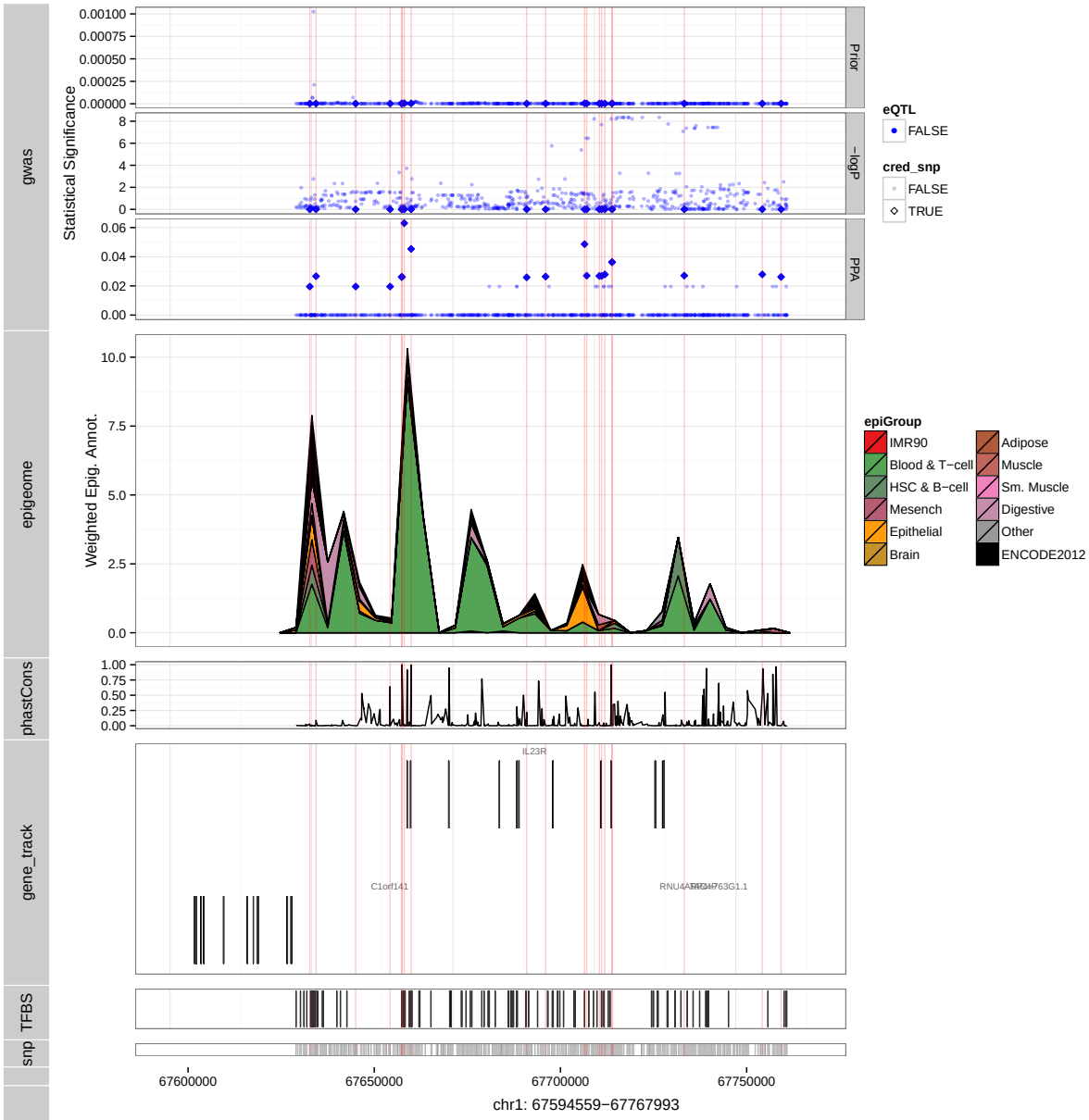
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
IFNLR1
eQTL
FALSE
cred_snp
FALSE
TRUE
epiGroup
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
24508000
24512000
24516000
24520000
chr1: 24508476−24520347

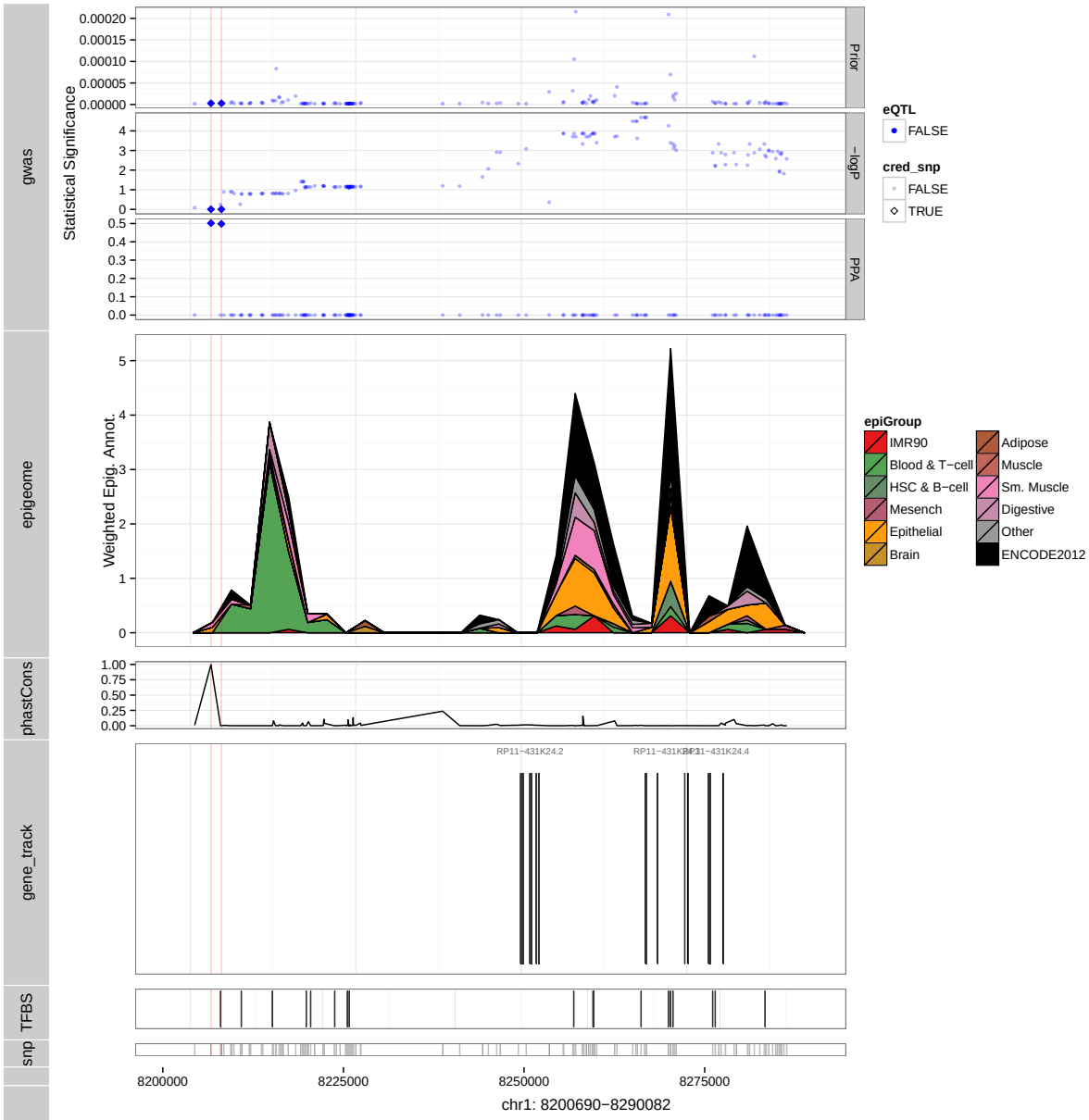
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
eQTL
FALSE
cred_snp
FALSE
TRUE
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
RP11−84D1.1
RUNX3
RP11−84D1.2
25240000
25260000
25280000
25300000
chr1: 25225675−25308257

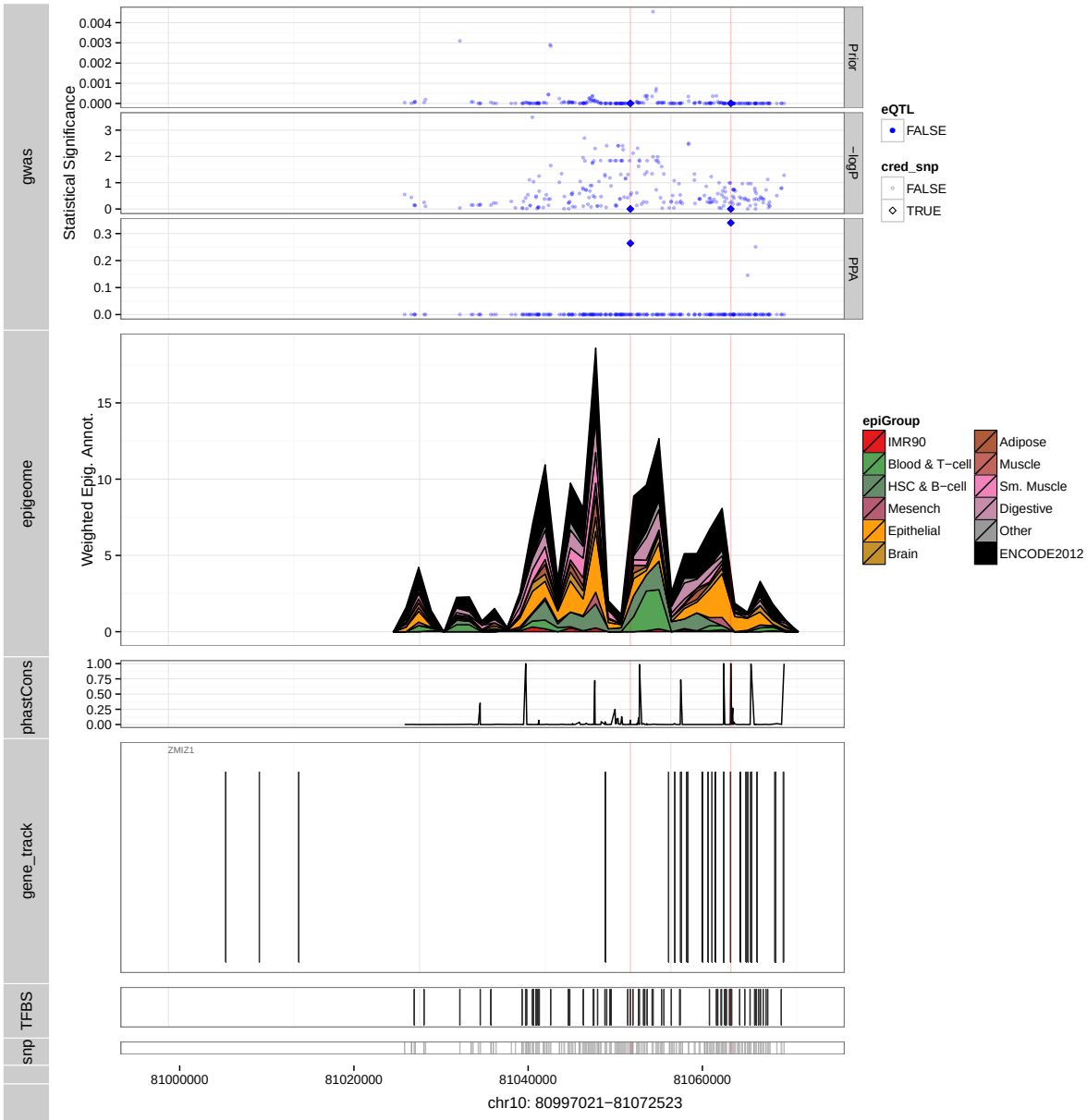
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
eQTL
FALSE
cred_snp
FALSE
TRUE
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
IL23R
C1orf141
67600000
67650000
67700000
67750000
chr1: 67594559−67767993

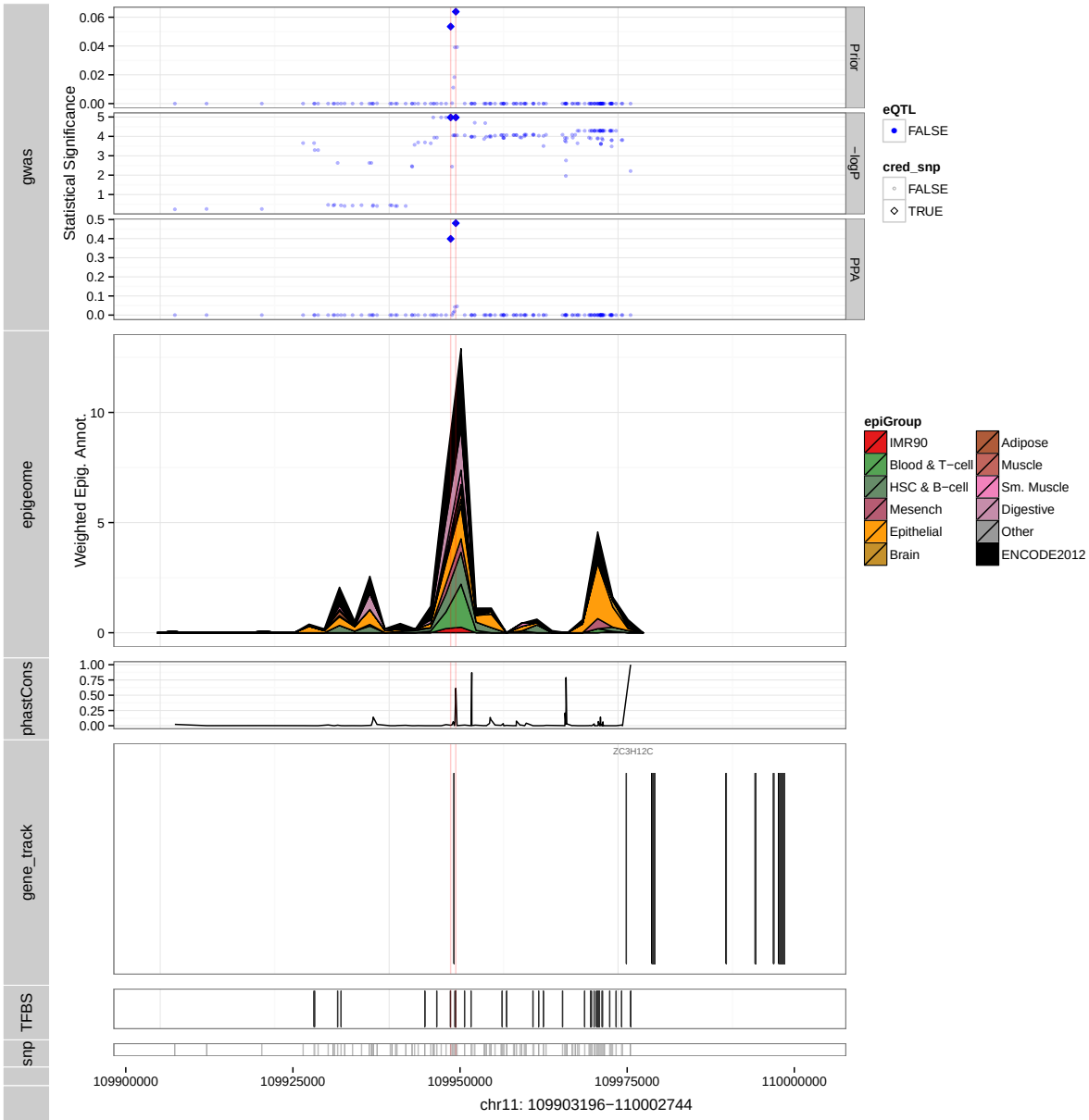
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
eQTL
FALSE
cred_snp
FALSE
TRUE
Weighted Epig. Annot.
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
RP11−431K24.2
RP11−431K24.3
RP11−431K24.4
8200000
8225000
8250000
8275000
chr1: 8200690−8290082

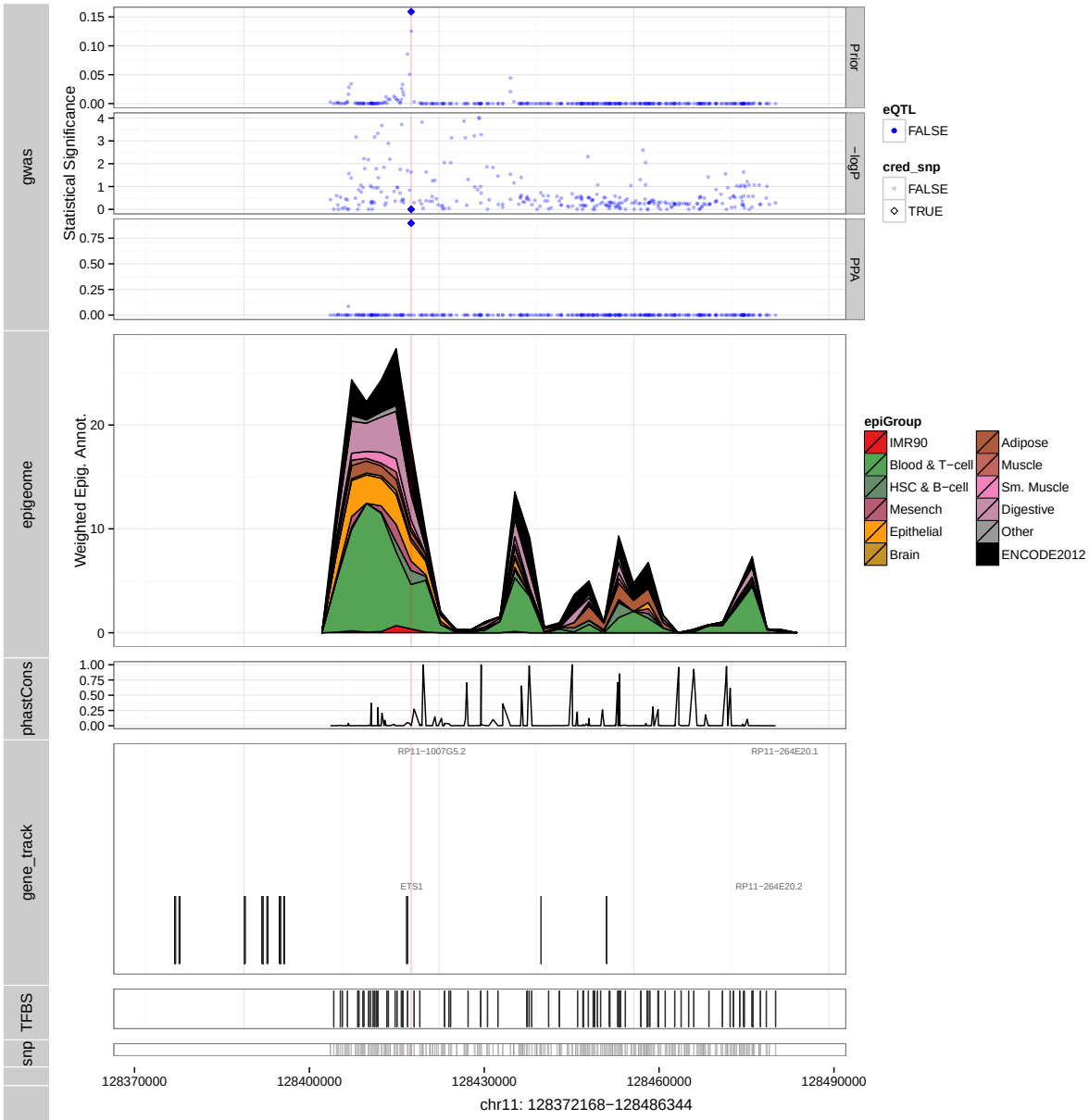
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
eQTL
FALSE
cred_snp
FALSE
TRUE
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
ZMIZ1
81000000
81020000
81040000
81060000
chr10: 80997021−81072523

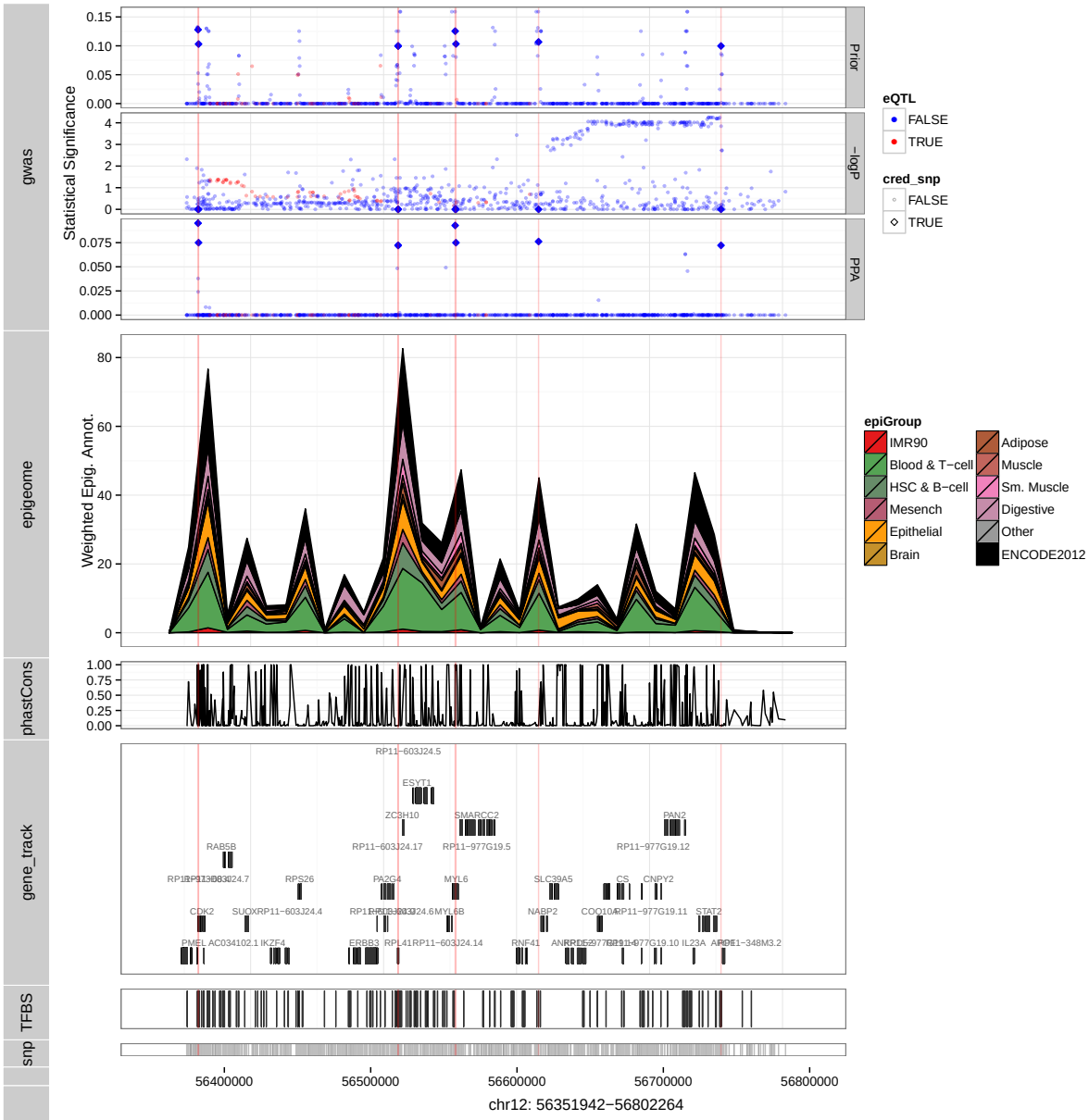
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
ZC3H12C
eQTL
FALSE
cred_snp
FALSE
TRUE
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
109900000
109925000
109950000
109975000
110000000
chr11: 109903196−110002744

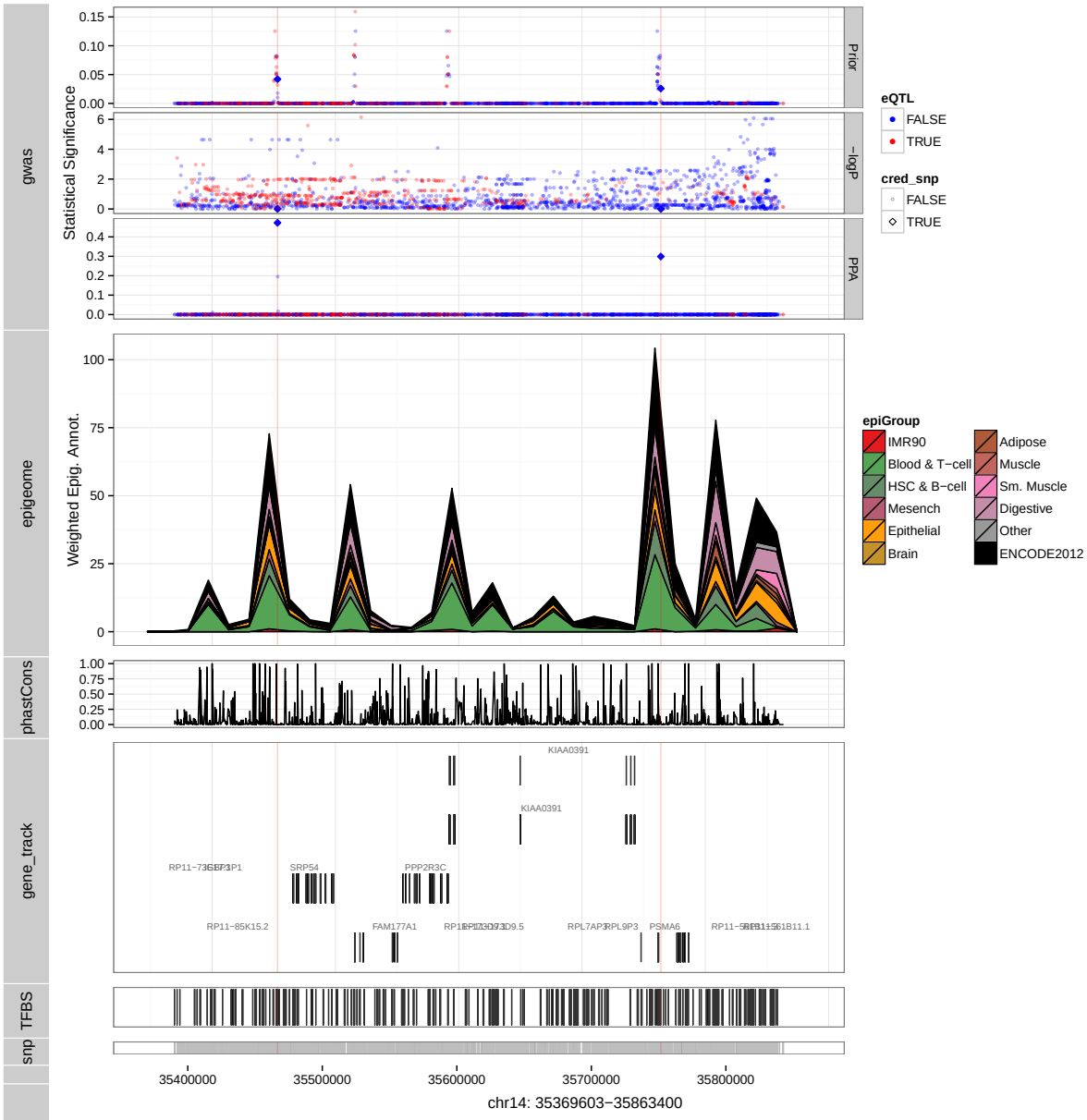
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
0.15
0.10
0.05
0.00
4
3
2
1
0
0.75
0.50
0.25
0.00
Weighted Epig. Annot.
20
10
0
1.00
0.75
0.50
0.25
0.00
RP11−1007G5.2
RP11−264E20.1
ETS1
RP11−264E20.2
128370000
128400000
128430000
128460000
128490000
chr11: 128372168−128486344
eQTL
FALSE
cred_snp
FALSE
TRUE
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012

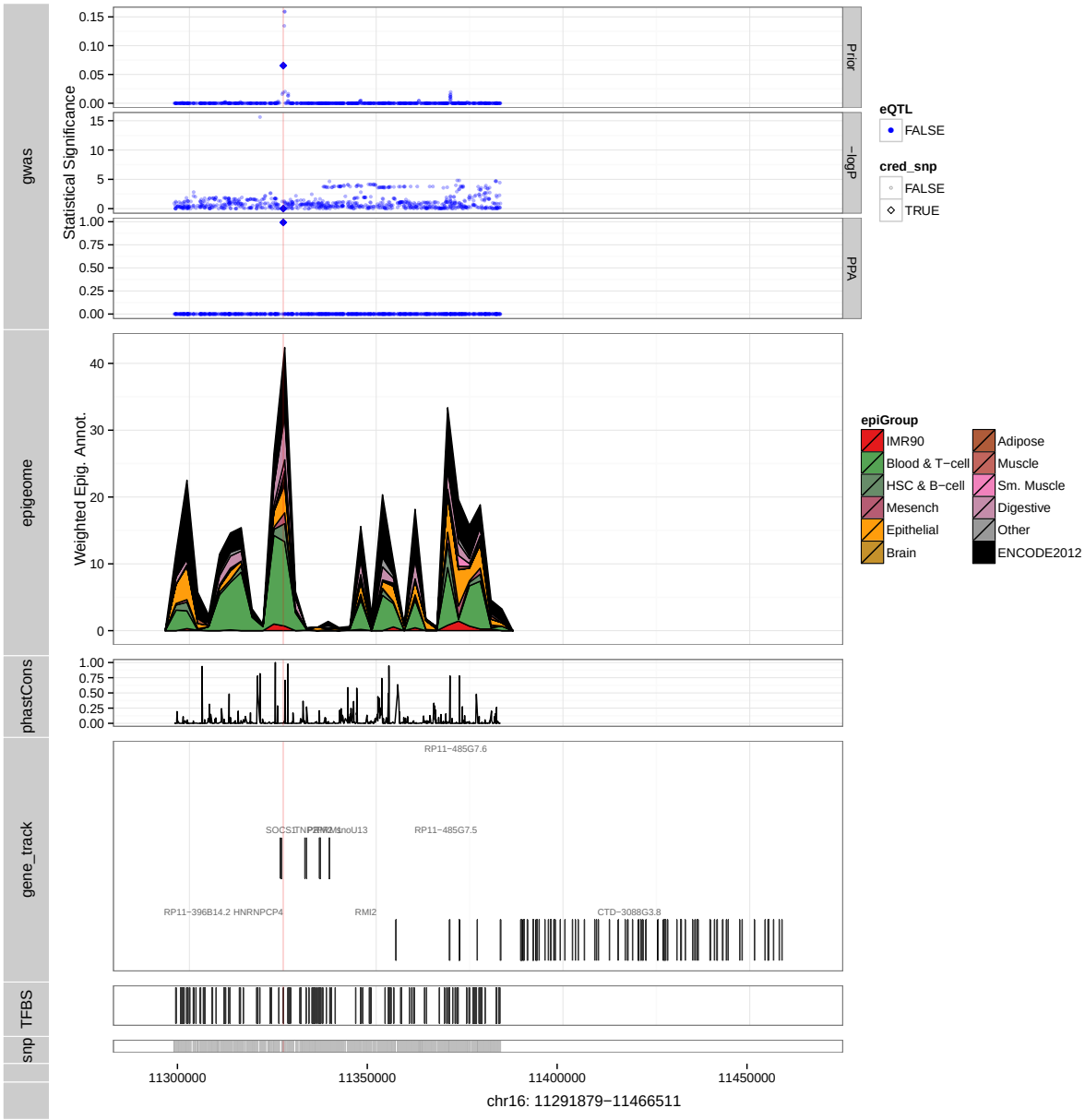
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
RP11−603J24.5
ESYT1
ZC3H10
SMARCC2
PAN2
RAB5B
RP11−603J24.17
RP11−977G19.5
RP11−977G19.12
RPS26
PA2G4
MYL6
SLC39A5
CS
CNPY2
CDK2
SUOX
RP11−603J24.4
NABP2
STAT2
RP11−977G19.11
PMEL
AC034102.1
IKZF4
ERBB3
RPL41
RP11−603J24.14
RNF41
RP11−977G19.10
IL23A
RP11−348M3.2
56400000
56500000
56600000
56700000
56800000
chr12: 56351942−56802264

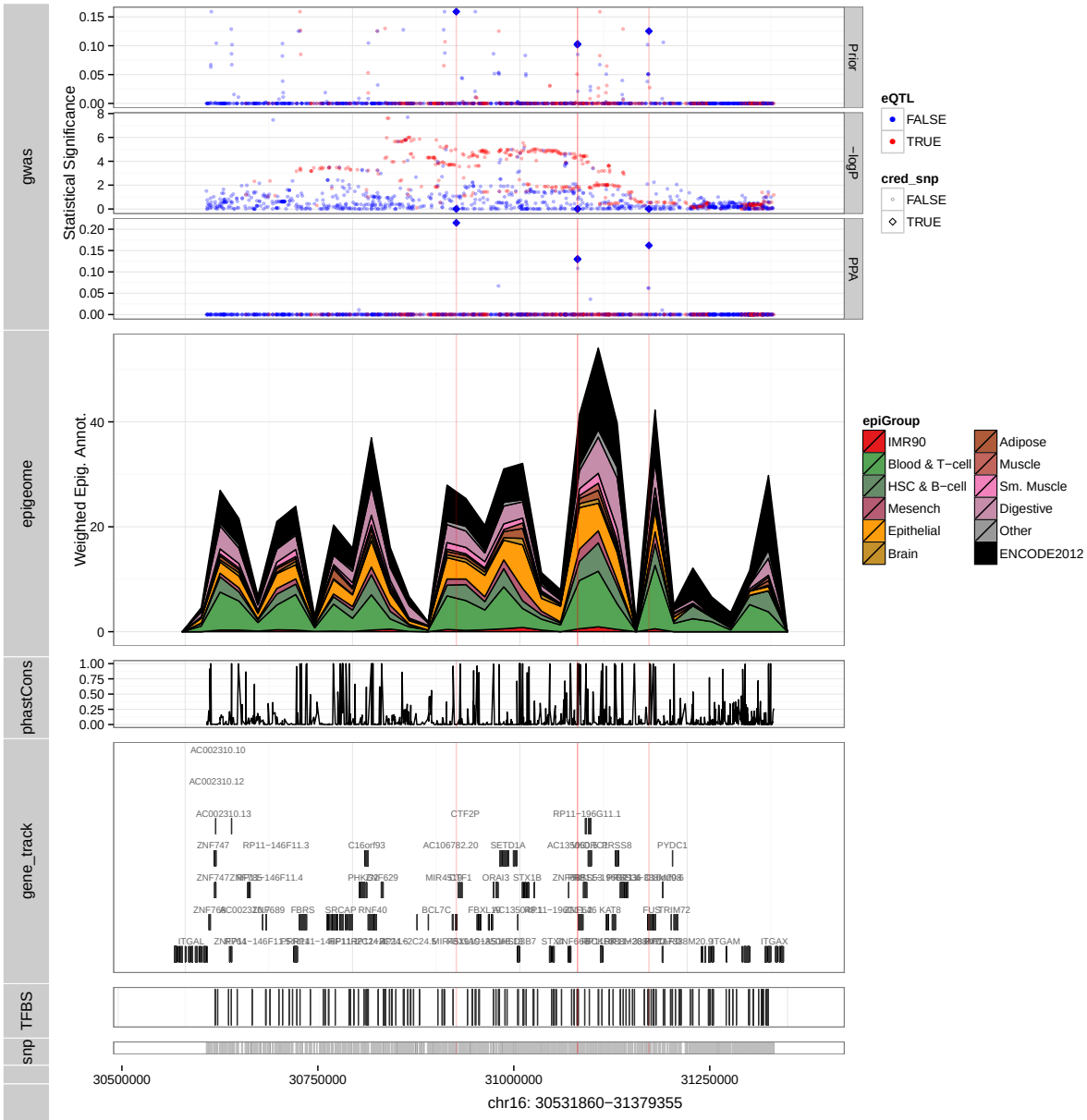
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
KIAA0391
SRP54
PPP2R3C
FAM177A1
PSMA6
35400000
35500000
35600000
35700000
35800000
chr14: 35369603−35863400

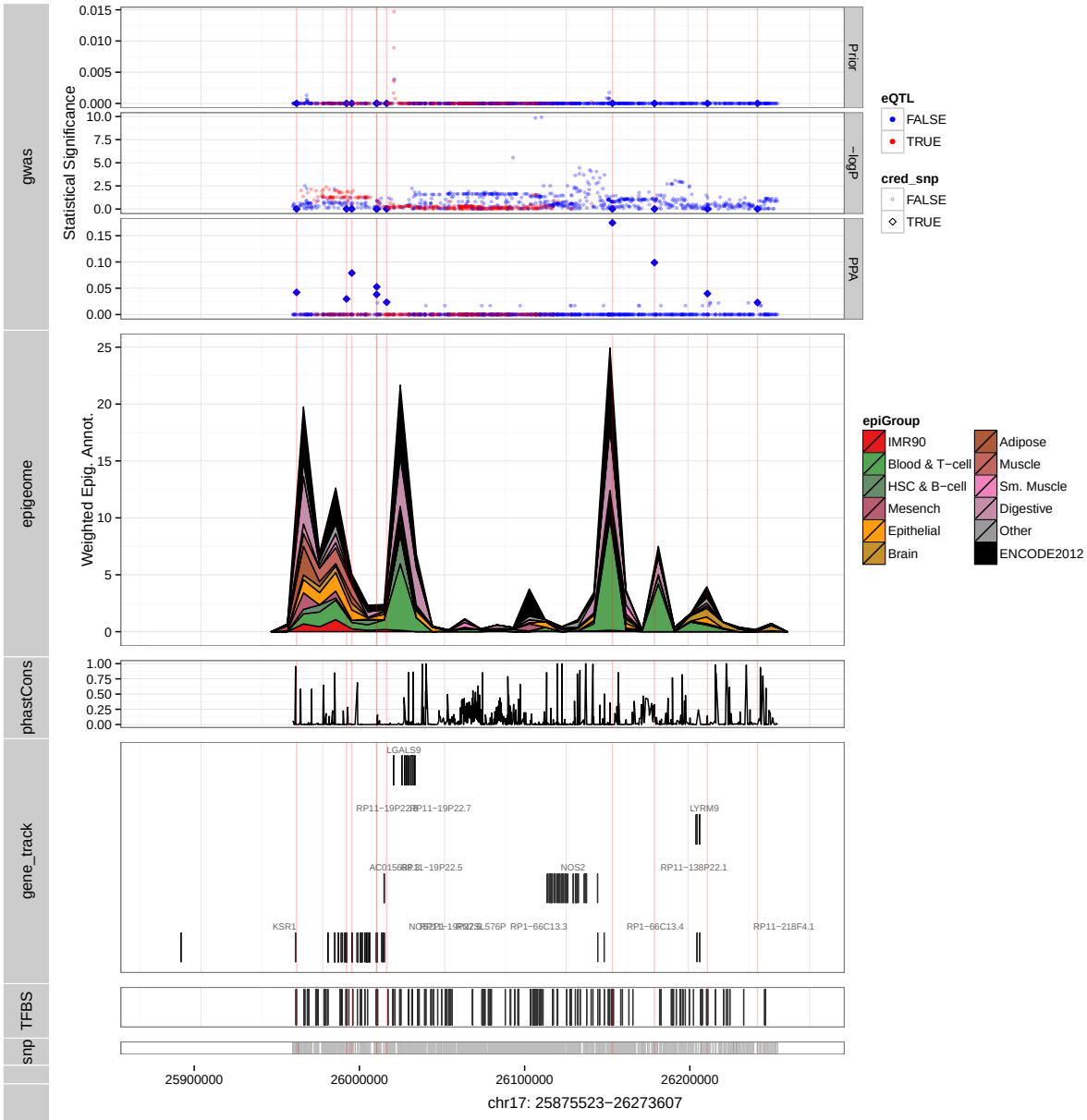
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
eQTL
FALSE
cred_snp
FALSE
TRUE
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
RP11−485G7.6
RP11−485G7.5
RP11−396B14.2
HNRNPCP4
RMI2
CTD−3088G3.8
11300000
11350000
11400000
11450000
chr16: 11291879−11466511

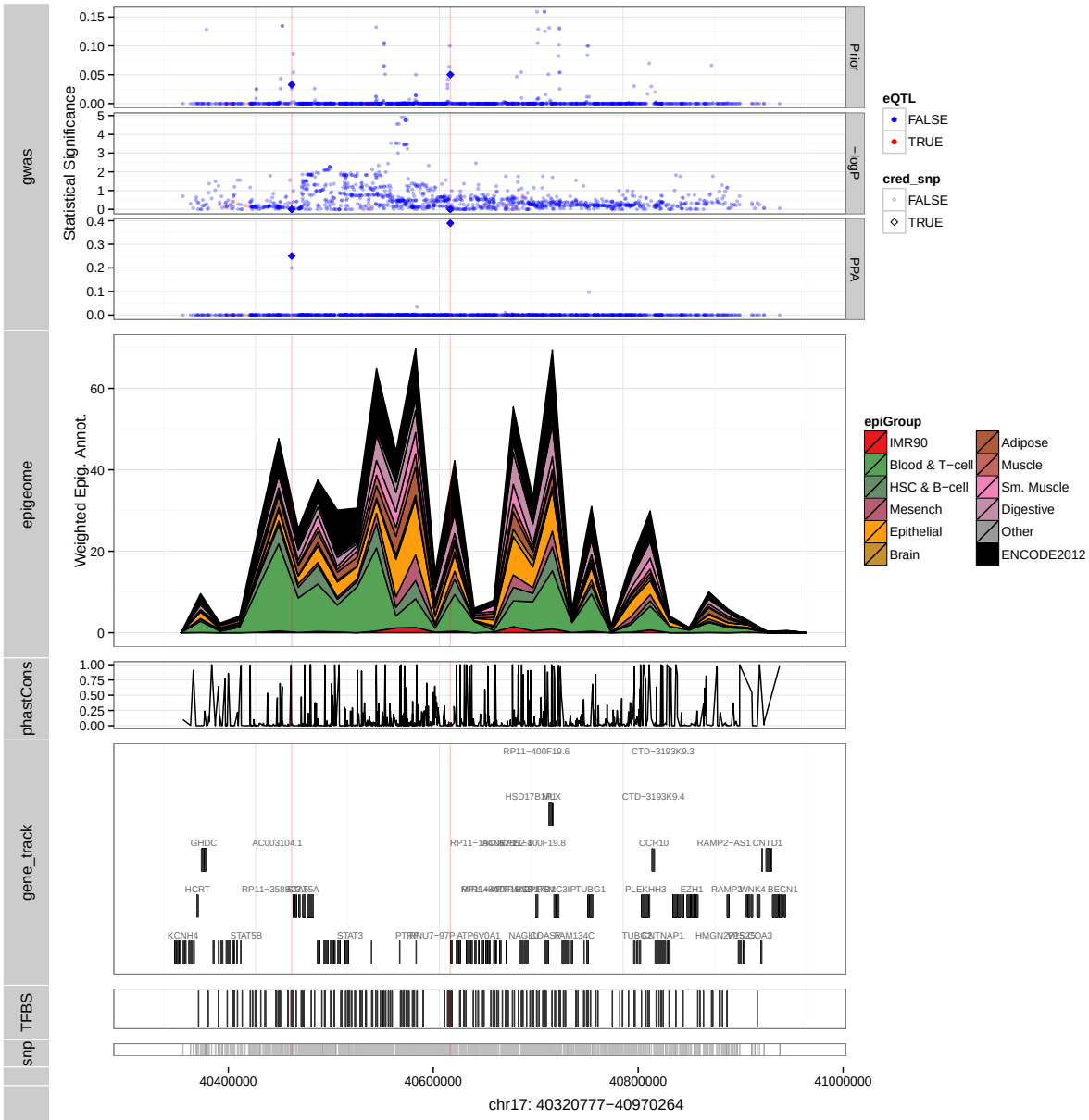
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
Weighted Epig. Annot.
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
AC002310.10
AC002310.12
AC002310.13
CTF2P
RP11−196G11.1
ZNF747
RP11−146F11.3
C16orf93
AC106782.20
SETD1A
PYDC1
RP11−146F11.4
PHKG2
ZNF629
ORAI3
STX1B
FBRS
SRCAP
RNF40
BCL7C
KAT8
FUS
TRIM72
ITGAL
ITGAX
30500000
30750000
31000000
31250000
chr16: 30531860−31379355

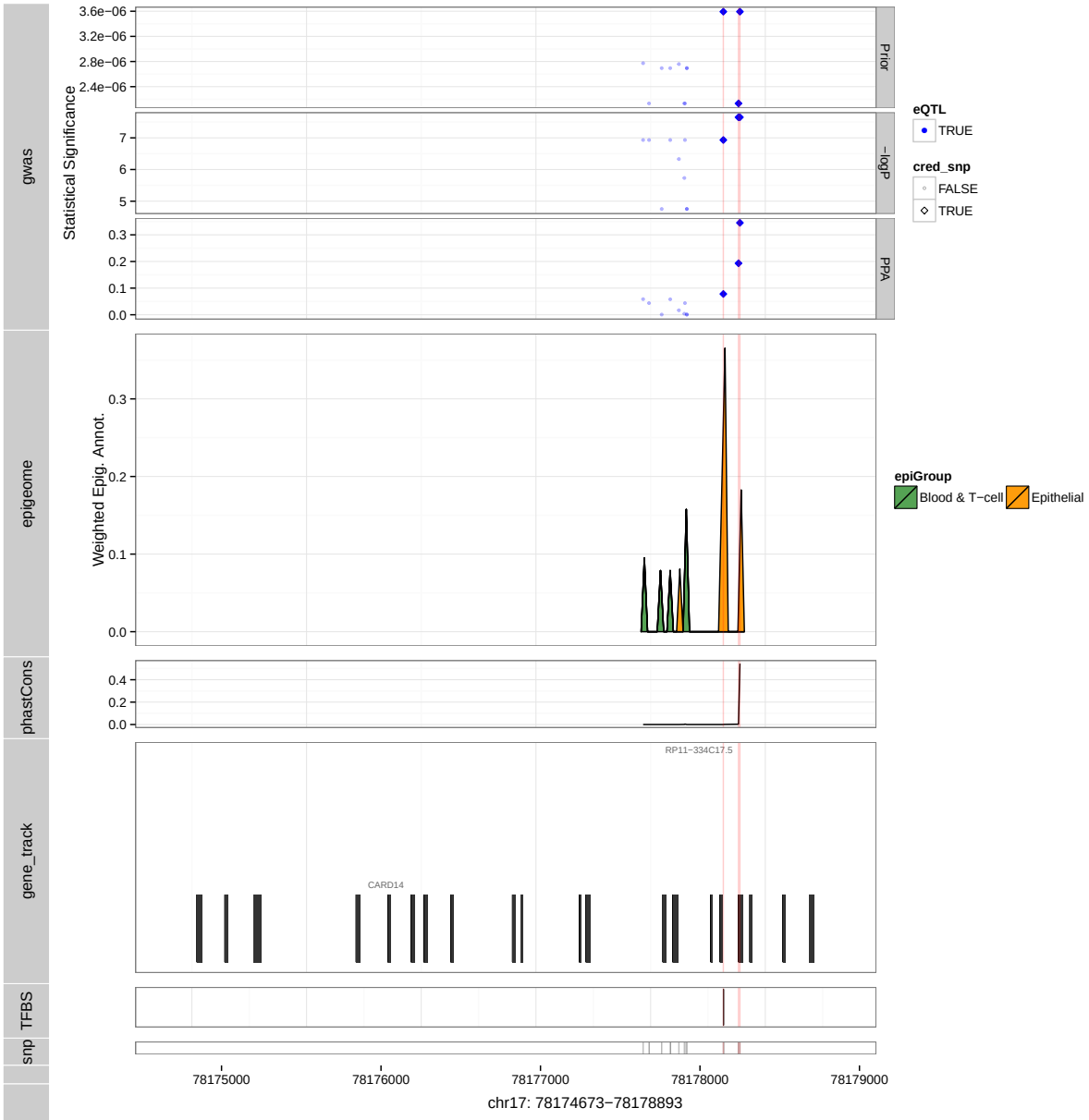
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
LGALS9
LYRM9
NOS2
KSR1
RP11−138P22.1
RP1−66C13.3
RP1−66C13.4
RP11−218F4.1
25900000
26000000
26100000
26200000
chr17: 25875523−26273607

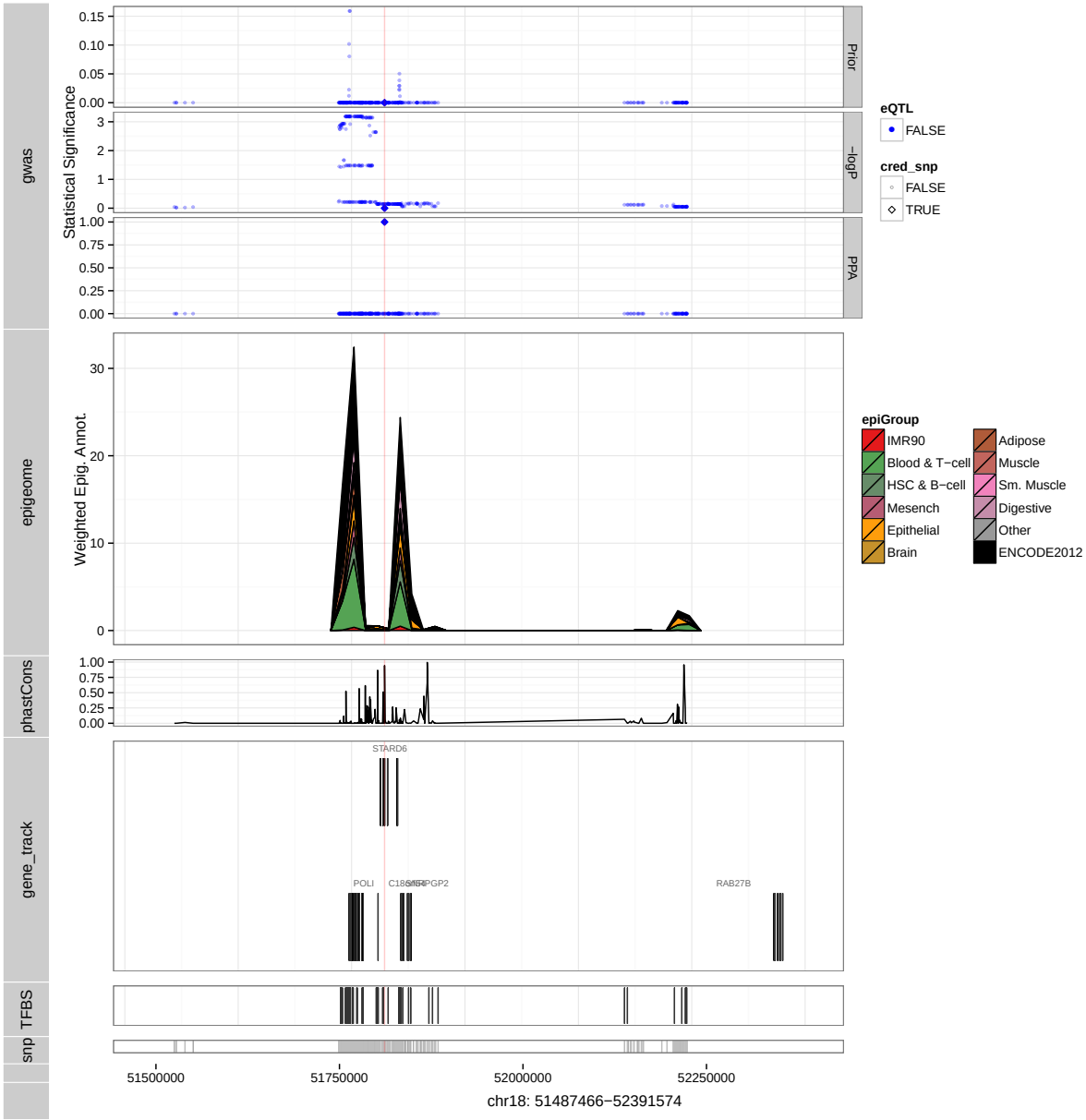
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
Weighted Epig. Annot.
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
RP11−400F19.6
CTD−3193K9.3
CTD−3193K9.4
GHDC
AC003104.1
CCR10
RAMP2−AS1
CNTD1
HCRT
PLEKHH3
EZH1
BECN1
KCNH4
STAT5B
STAT3
40400000
40600000
40800000
41000000
chr17: 40320777−40970264

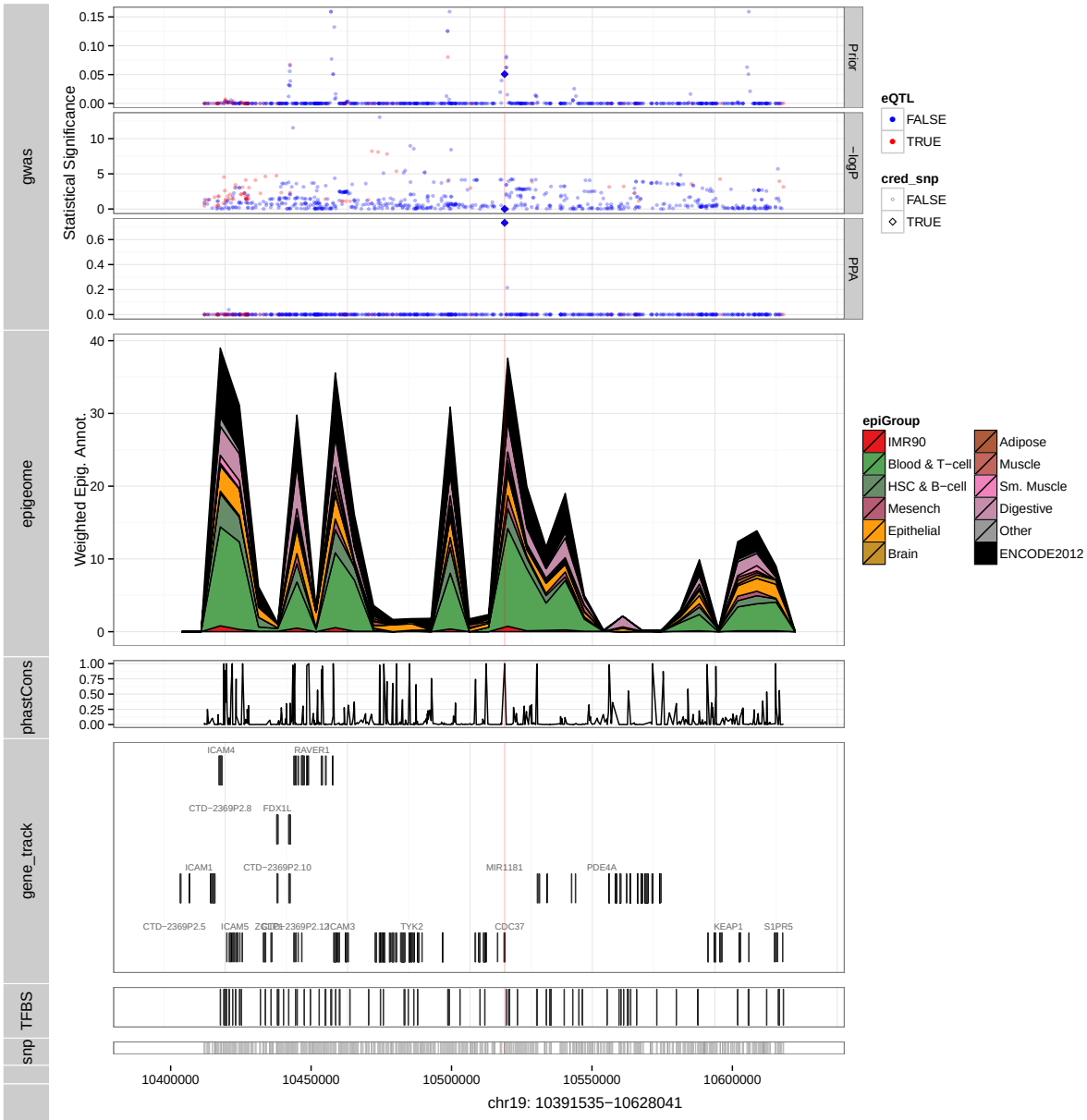
Psoriasis


gwas
epigeome
phastCons
gene_track
snp_TFBS
Statistical Significance
3.6e−06
3.2e−06
2.8e−06
2.4e−06
Prior
7
6
5
−logP
0.3
0.2
0.1
0.0
PPA
eQTL
TRUE
cred_snp
FALSE
TRUE
Weighted Epig. Annot.
0.3
0.2
0.1
0.0
epiGroup
Blood & T−cell
Epithelial
0.4
0.2
0.0
RP11−334C17.5
CARD14
78175000
78176000
78177000
78178000
78179000
chr17: 78174673−78178893

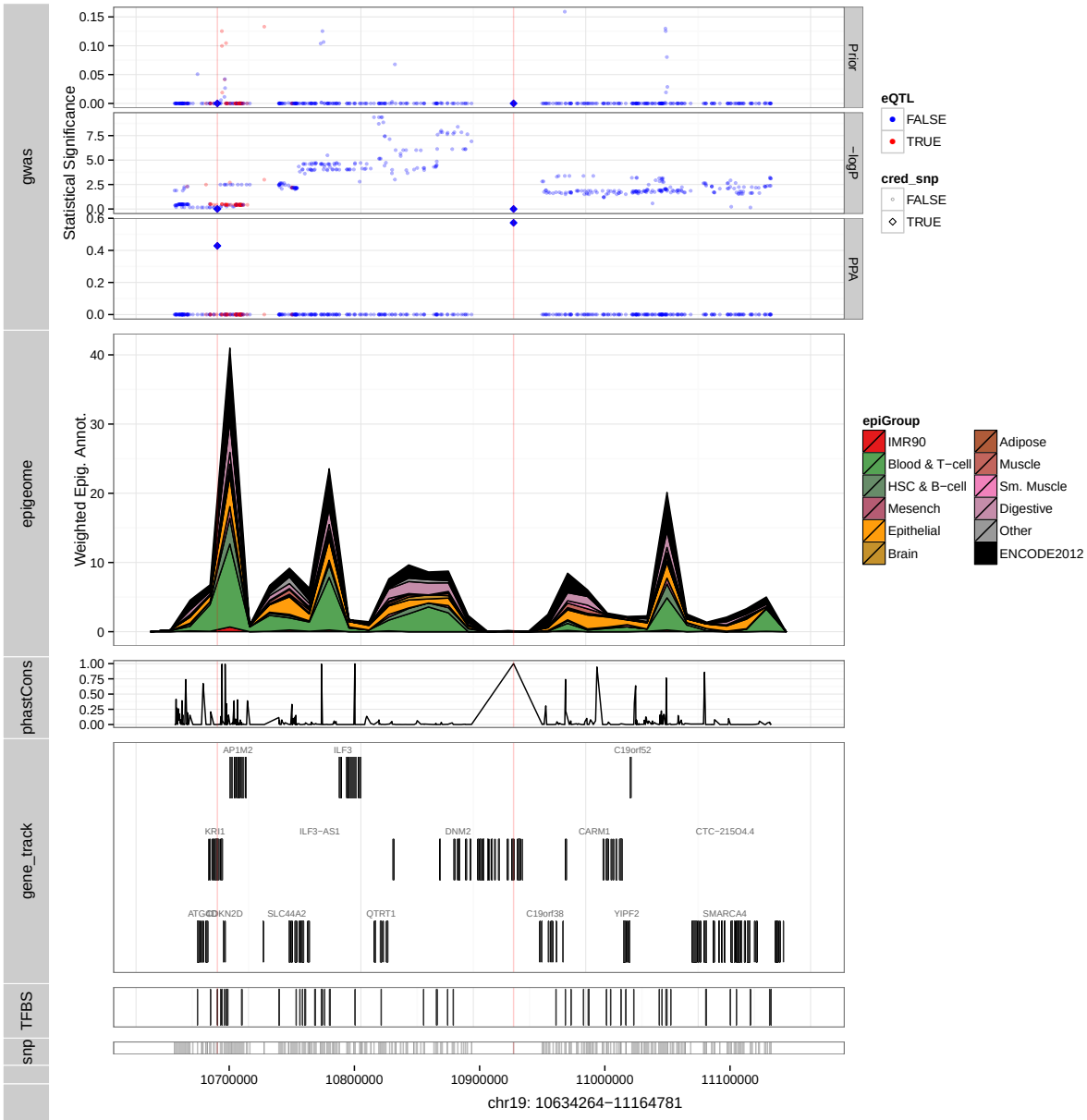
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
eQTL
FALSE
cred_snp
FALSE
TRUE
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
STARD6
POLI
RAB27B
51500000
51750000
52000000
52250000
chr18: 51487466−52391574

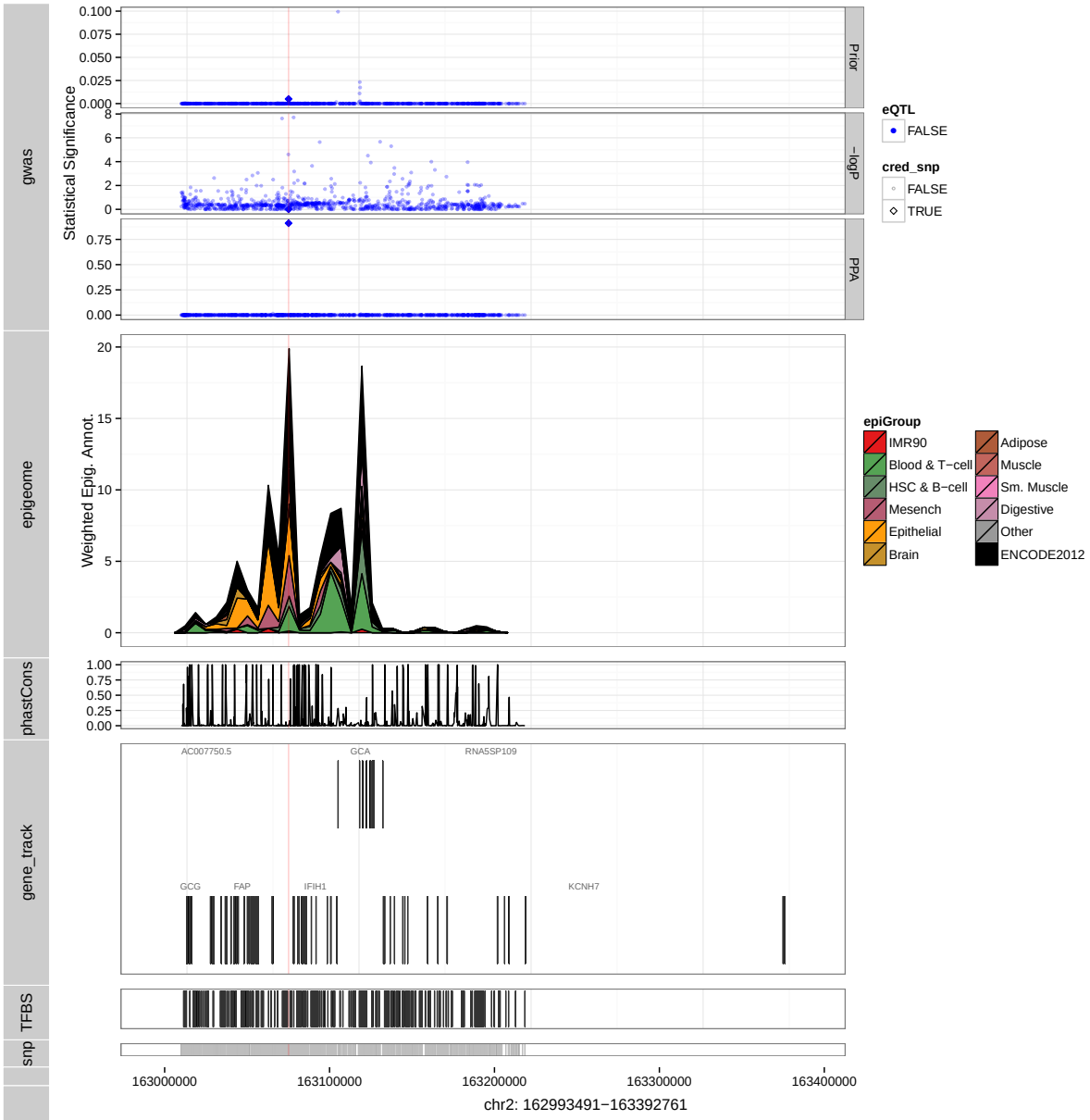
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
ICAM4
RAVER1
CTD−2369P2.8
FDX1L
ICAM1
CTD−2369P2.10
MIR1181
PDE4A
CTD−2369P2.5
ICAM5
TYK2
CDC37
KEAP1
S1PR5
10400000
10450000
10500000
10550000
10600000
chr19: 10391535−10628041

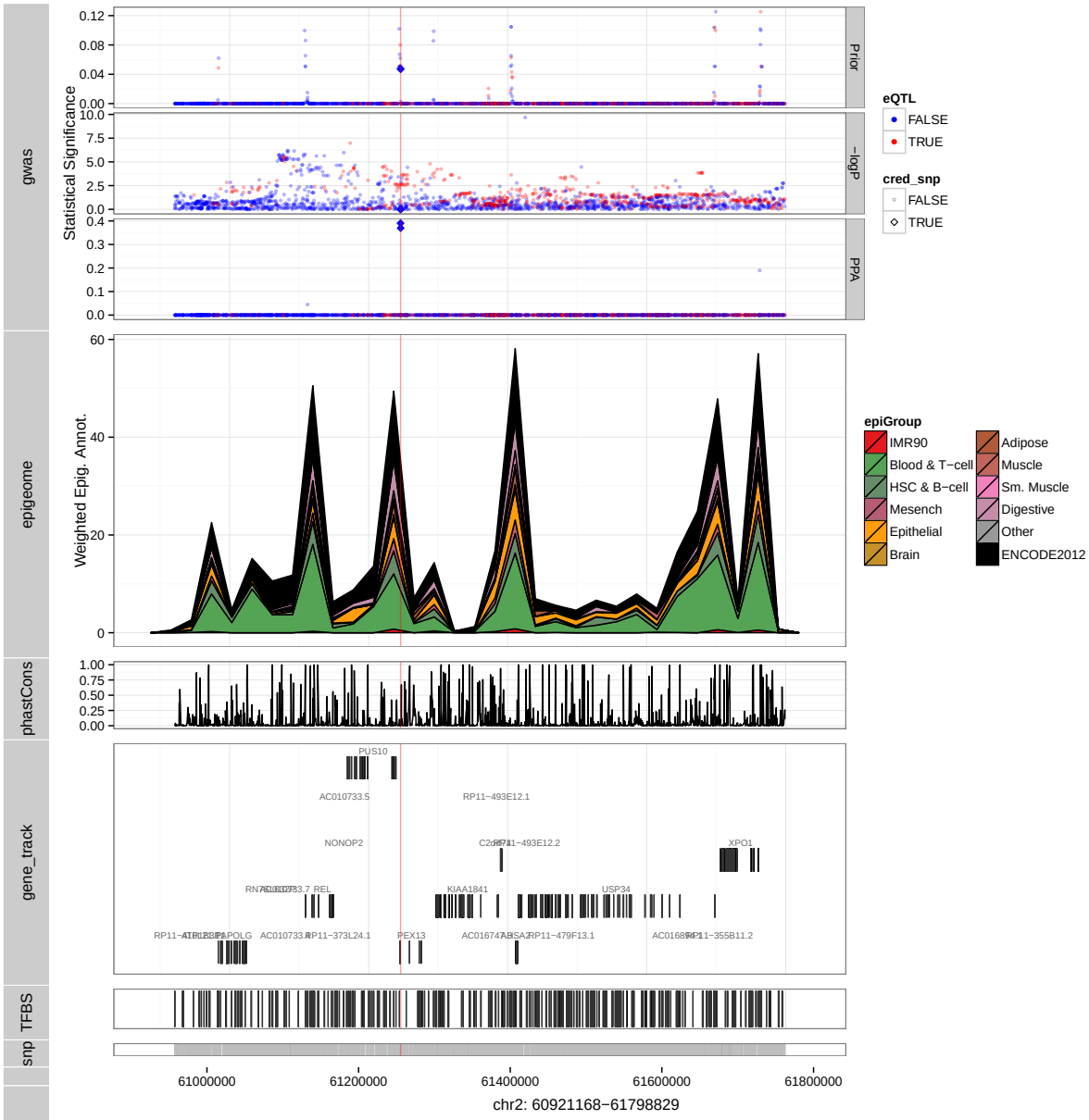
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
Weighted Epig. Annot.
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
AP1M2
ILF3
C19orf52
KRI1
ILF3−AS1
DNM2
CARM1
CTC−215O4.4
CDKN2D
SLC44A2
QTRT1
C19orf38
YIPF2
SMARCA4
10700000
10800000
10900000
11000000
11100000
chr19: 10634264−11164781

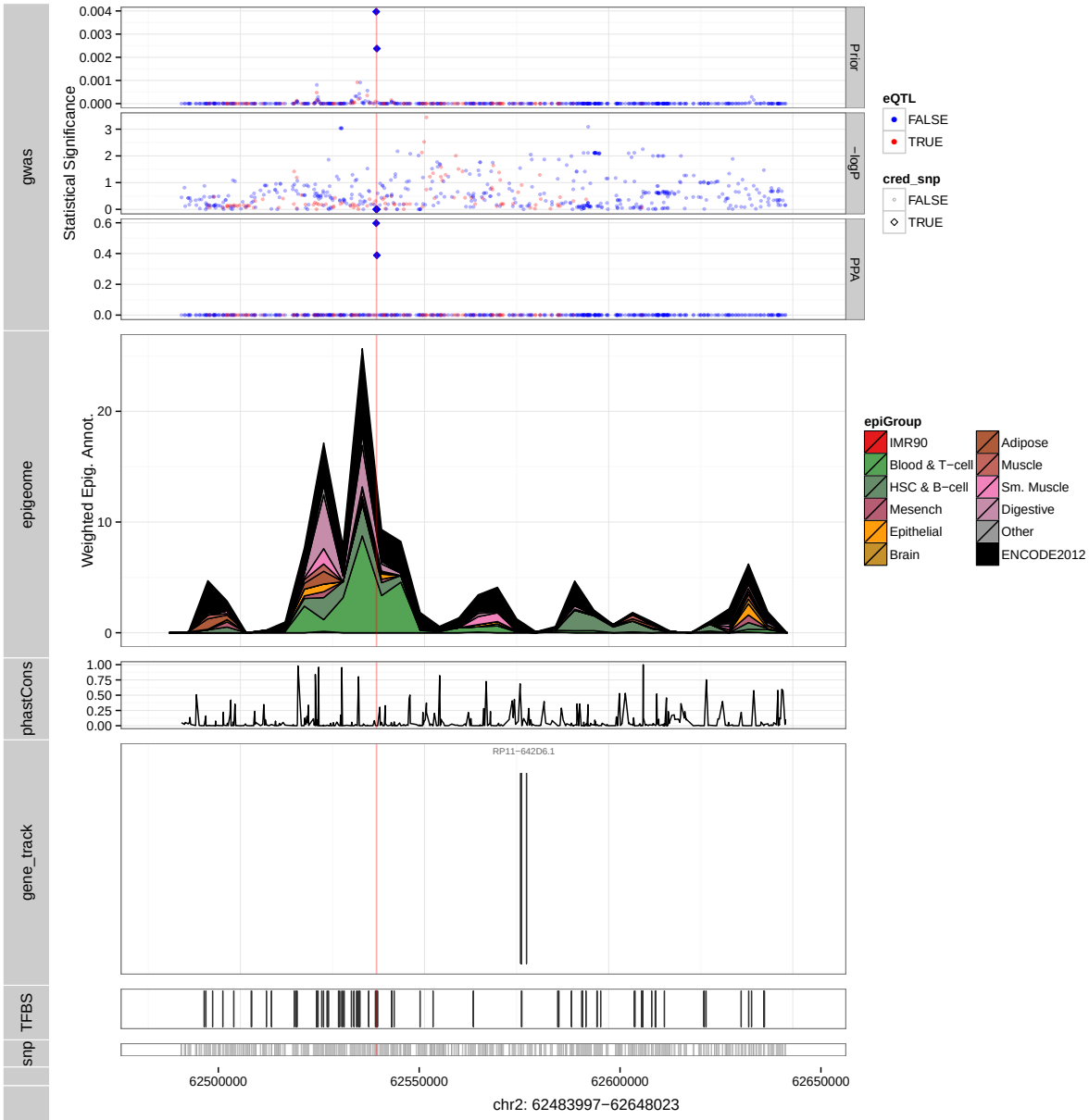
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
eQTL
FALSE
cred_snp
FALSE
TRUE
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
AC007750.5
GCA
RNASSP109
GCG
FAP
IFIH1
KCNH7
163000000
163100000
163200000
163300000
163400000
chr2: 162993491−163392761

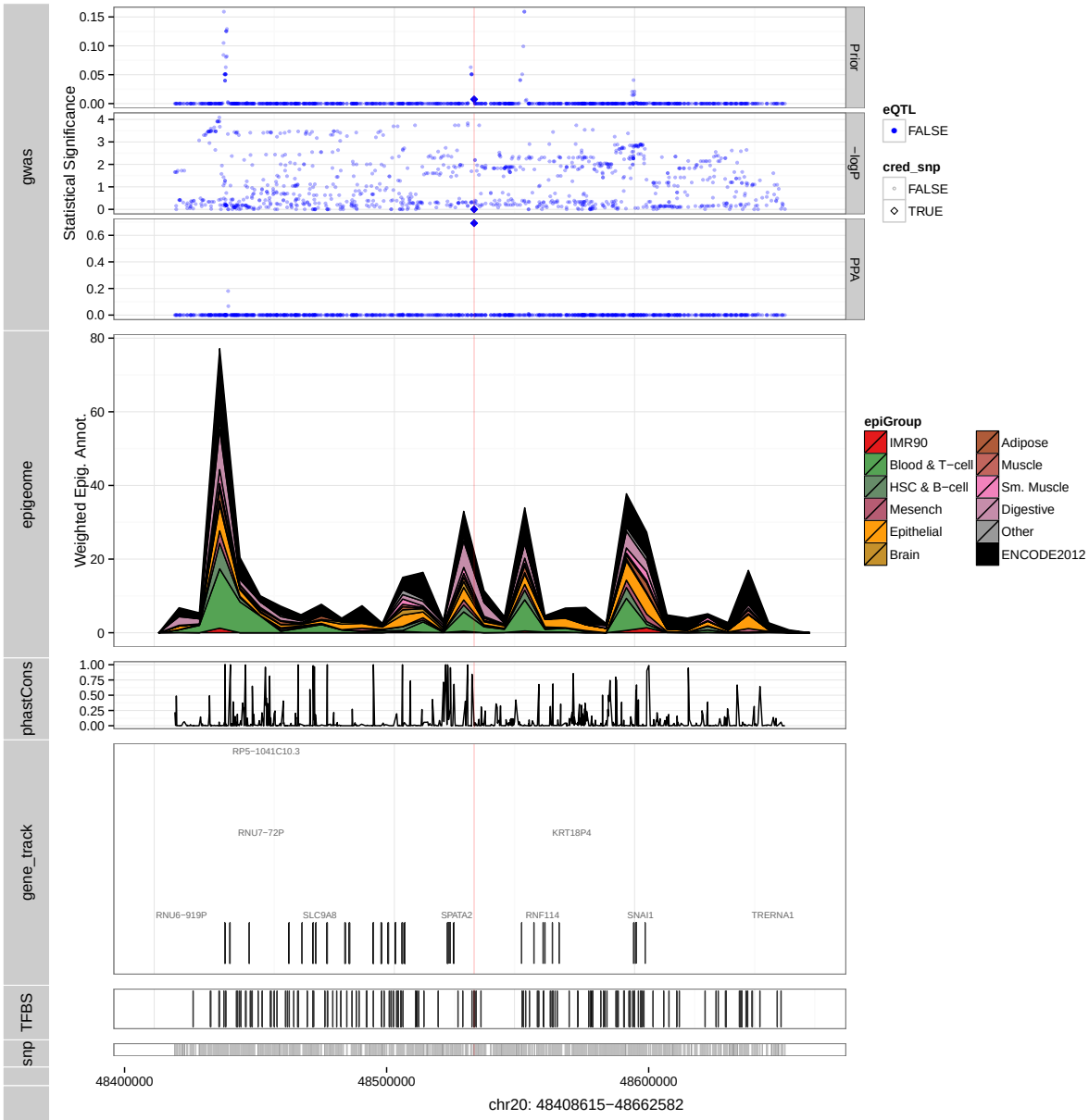
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
Weighted Epig. Annot.
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
PUS10
AC010733.5
RP11−493E12.1
NONOP2
XPO1
REL
KIAA1841
USP34
PEX13
RP11−373L24.1
RP11−479F13.1
61000000
61200000
61400000
61600000
61800000
chr2: 60921168−61798829

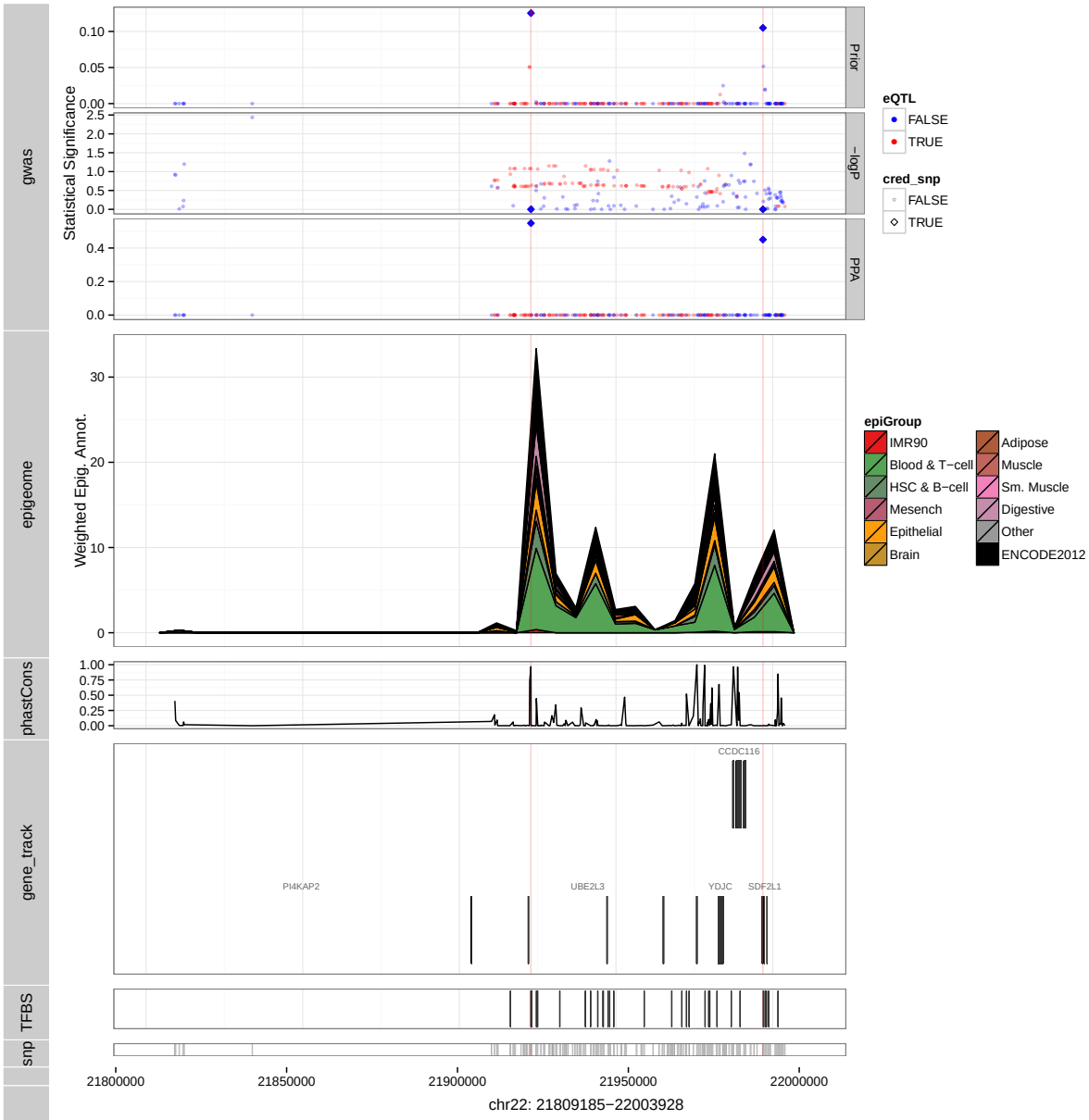
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
Weighted Epig. Annot.
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
RP11−642D6.1
62500000
62550000
62600000
62650000
chr2: 62483997−62648023

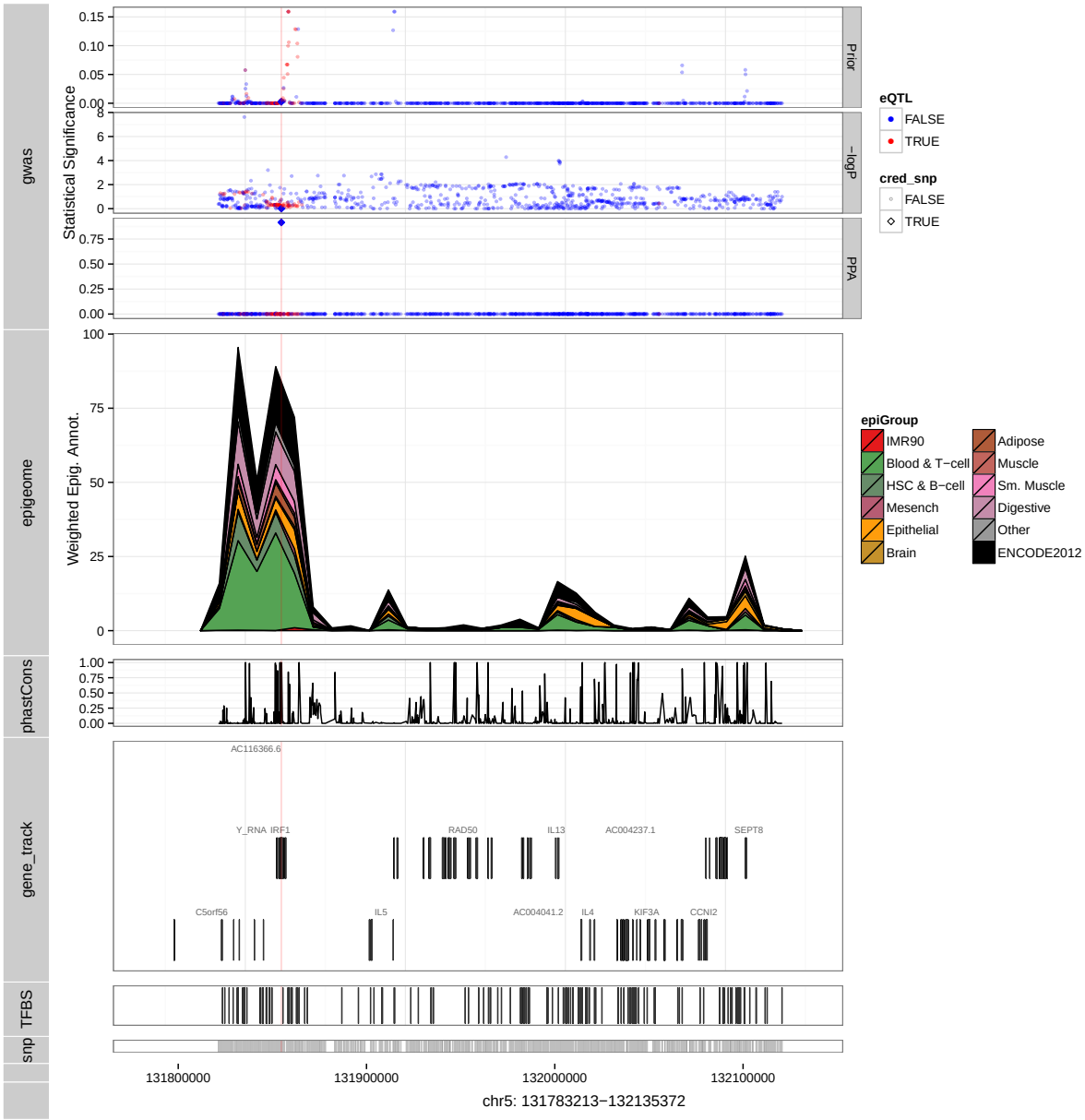
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
eQTL
FALSE
cred_snp
FALSE
TRUE
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
RP5−1041C10.3
RNU7−72P
KRT18P4
RNU6−919P
SLC9A8
SPATA2
RNF114
SNAI1
TRERNA1
48400000
48500000
48600000
chr20: 48408615−48662582

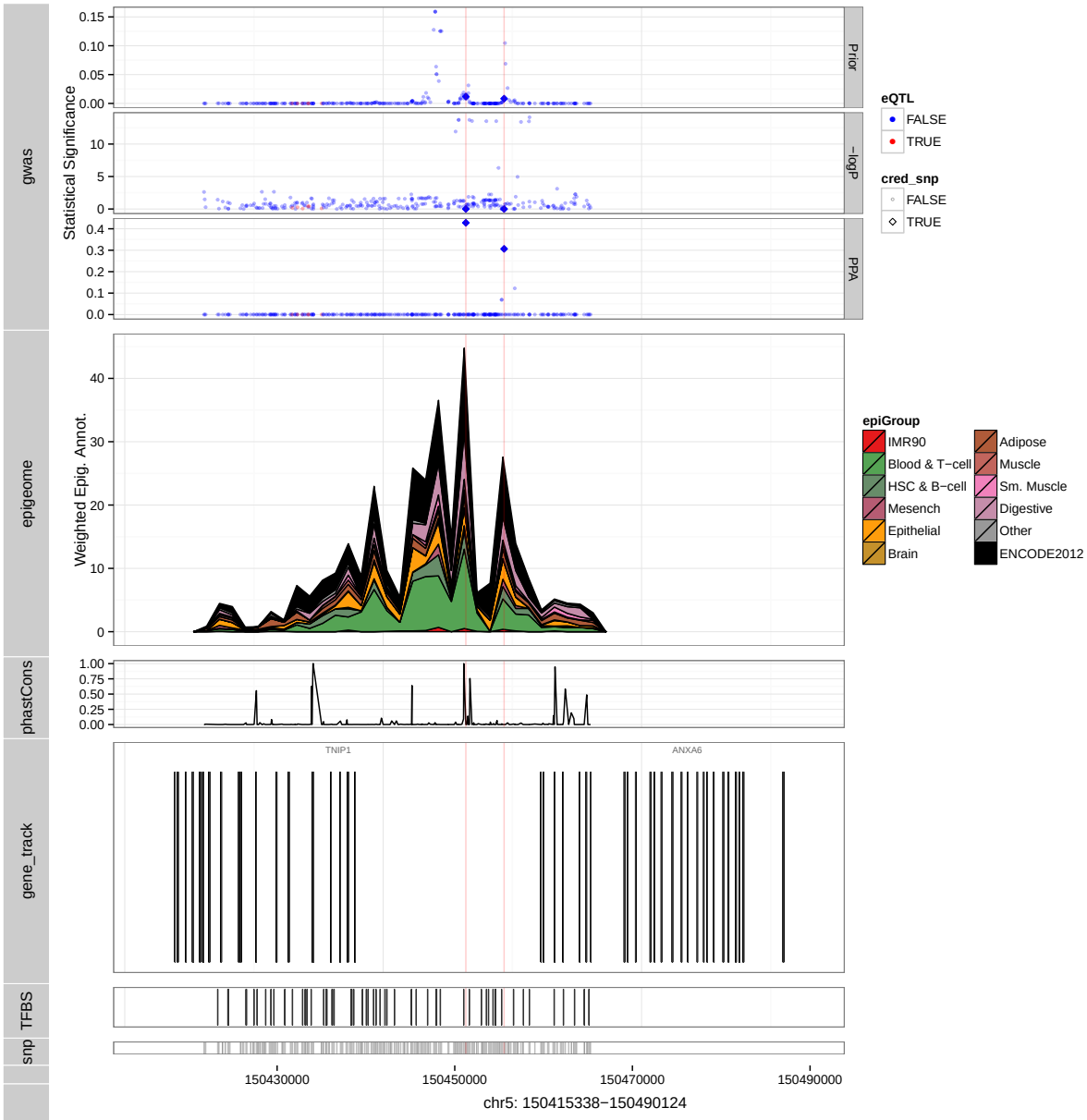
Psoriasis


gwas
epigeome
phastCons
gene_track
snp_TFBS
Statistical Significance
Prior
−logP
PPA
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
Weighted Epig. Annot.
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
CCDC116
PI4KAP2
UBE2L3
YDJC
SDF2L1
21800000
21850000
21900000
21950000
22000000
chr22: 21809185−22003928

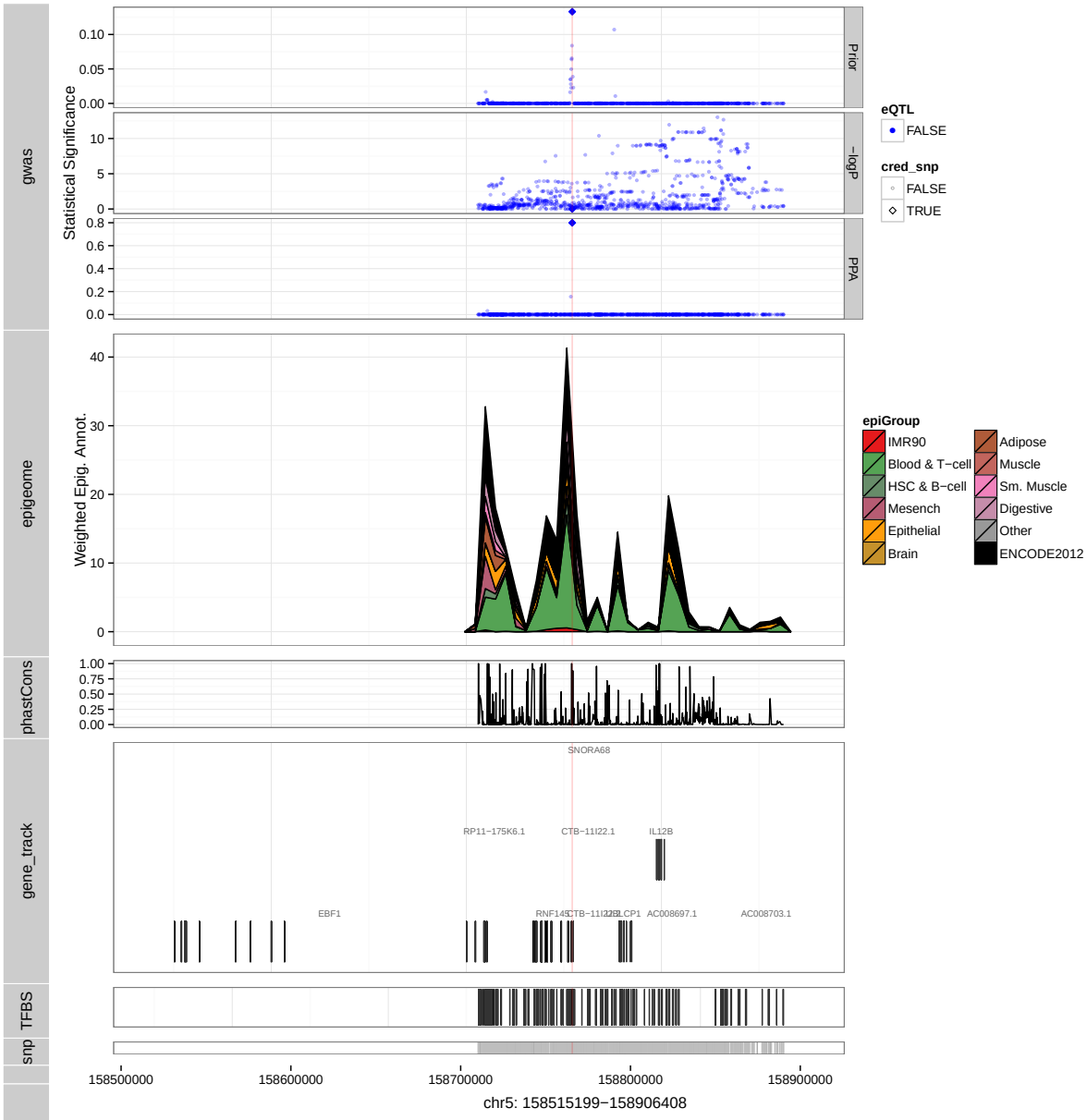
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
AC116366.6
Y_RNA
IRF1
RAD50
IL13
AC004237.1
SEPT8
C5orf56
IL5
AC004041.2
IL4
KIF3A
CCNI2
131800000
131900000
132000000
132100000
chr5: 131783213−132135372

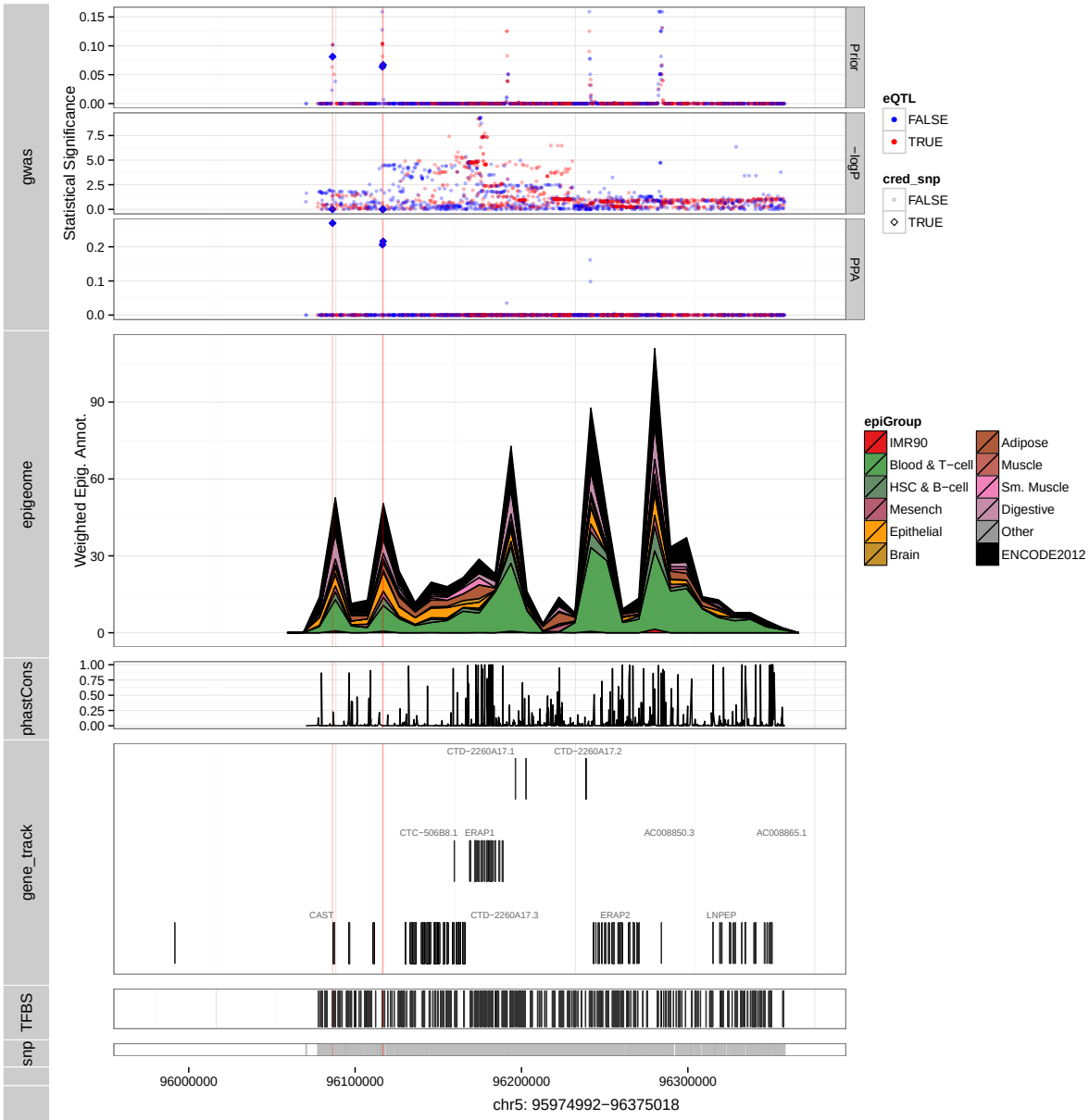
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
TNIP1
ANXA6
150430000
150450000
150470000
150490000
chr5: 150415338−150490124

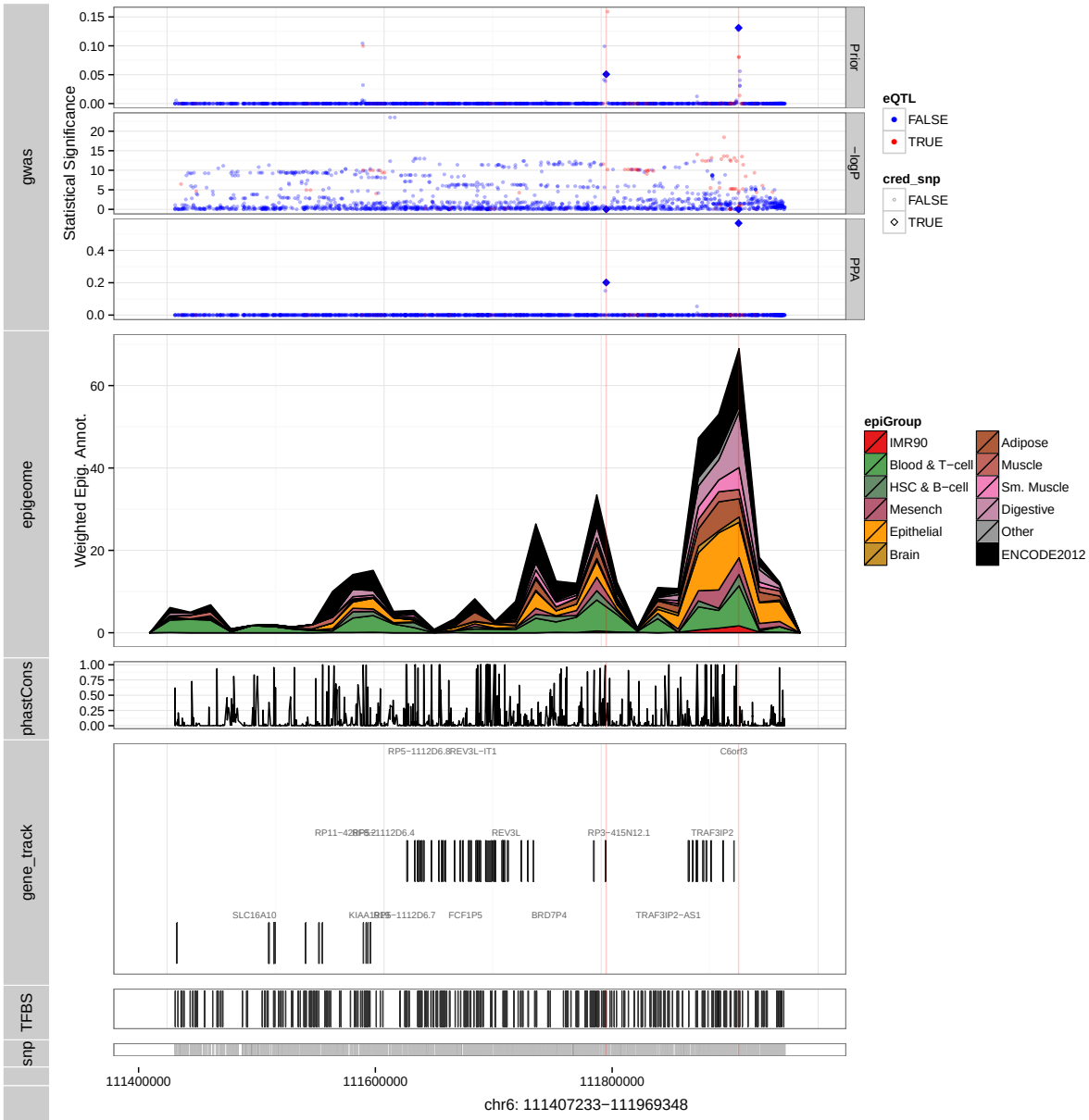
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
eQTL
FALSE
cred_snp
FALSE
TRUE
Weighted Epig. Annot.
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
SNORA68
RP11−175K6.1
CTB−11I22.1
IL12B
EBF1
AC008697.1
AC008703.1
158500000
158600000
158700000
158800000
158900000
chr5: 158515199−158906408

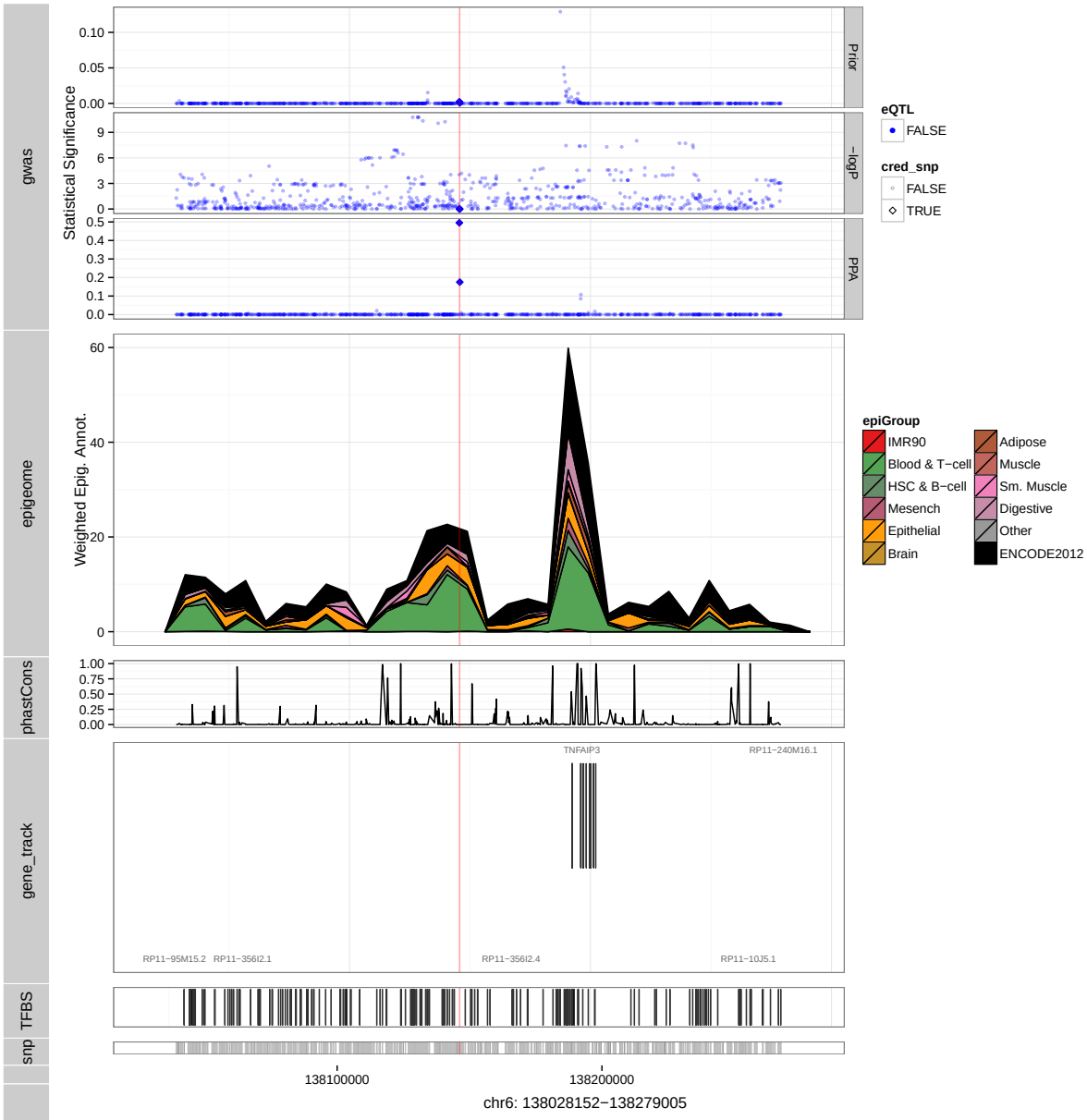
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
CTD−2260A17.1
CTD−2260A17.2
CTC−506B8.1
ERAP1
AC008850.3
AC008865.1
CAST
CTD−2260A17.3
ERAP2
LNPEP
96000000
96100000
96200000
96300000
chr5: 95974992−96375018

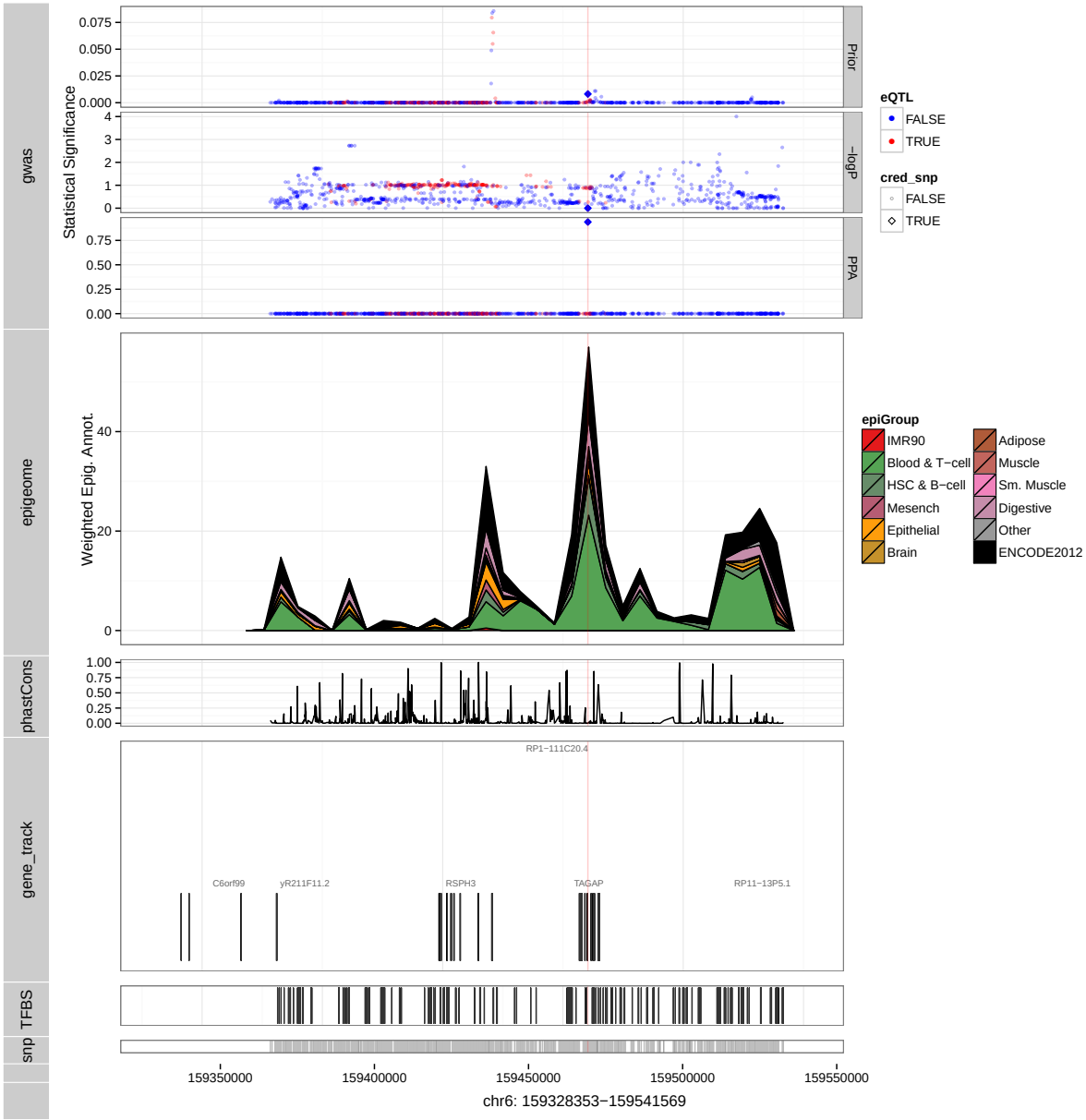
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
Weighted Epig. Annot.
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
RP5−1112D6.8
REV3L−IT1
C6orf3
RP11−42B5.2
RP5−1112D6.4
REV3L
RP3−415N12.1
TRAF3IP2
SLC16A10
RP5−1112D6.7
FCF1P5
BRD7P4
TRAF3IP2−AS1
111400000
111600000
111800000
chr6: 111407233−111969348

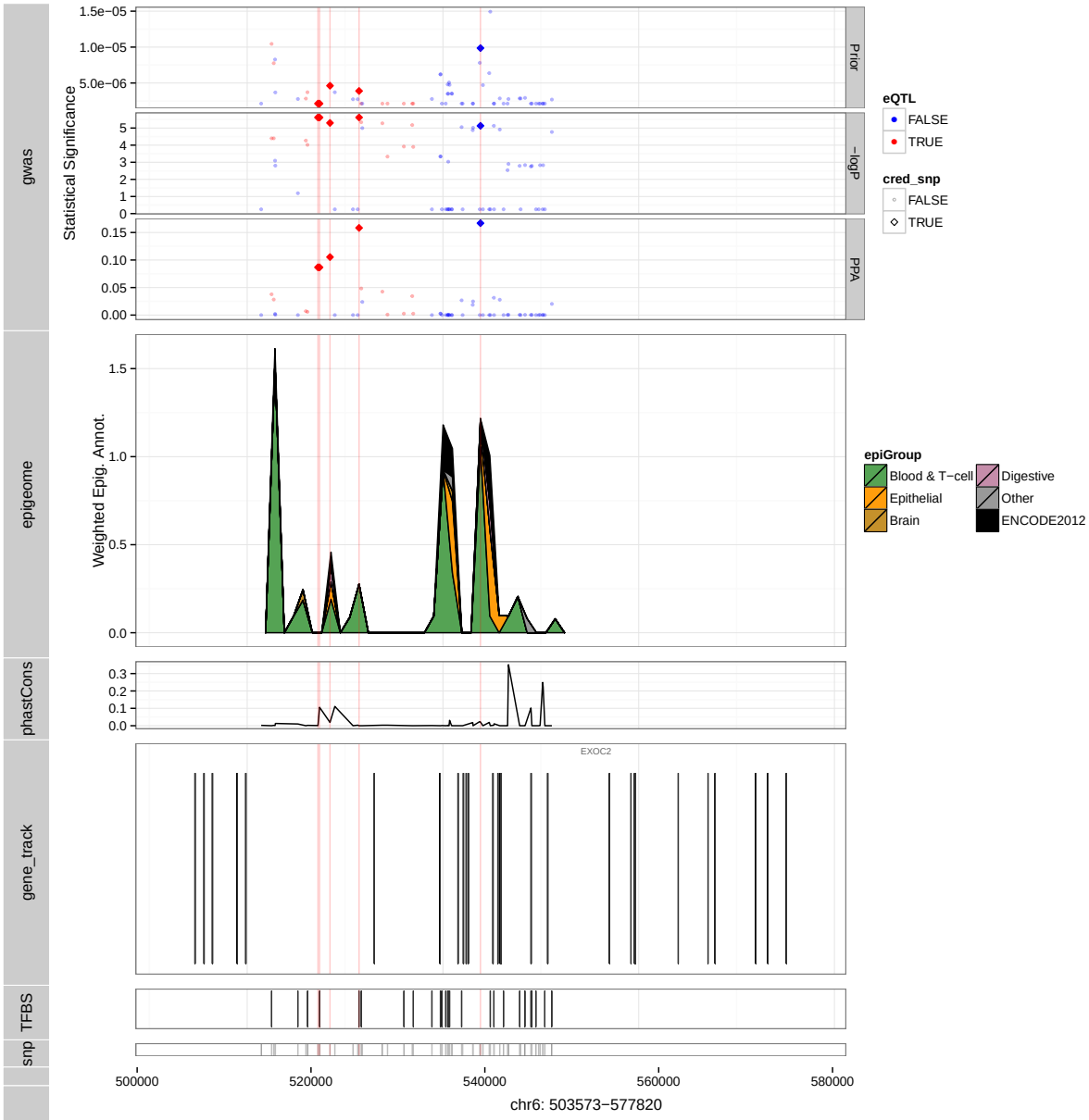
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
eQTL
FALSE
cred_snp
FALSE
TRUE
Weighted Epig. Annot.
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
TNFAIP3
RP11−240M16.1
RP11−95M15.2
RP11−356I2.1
RP11−356I2.4
RP11−10J5.1
138100000
138200000
chr6: 138028152−138279005

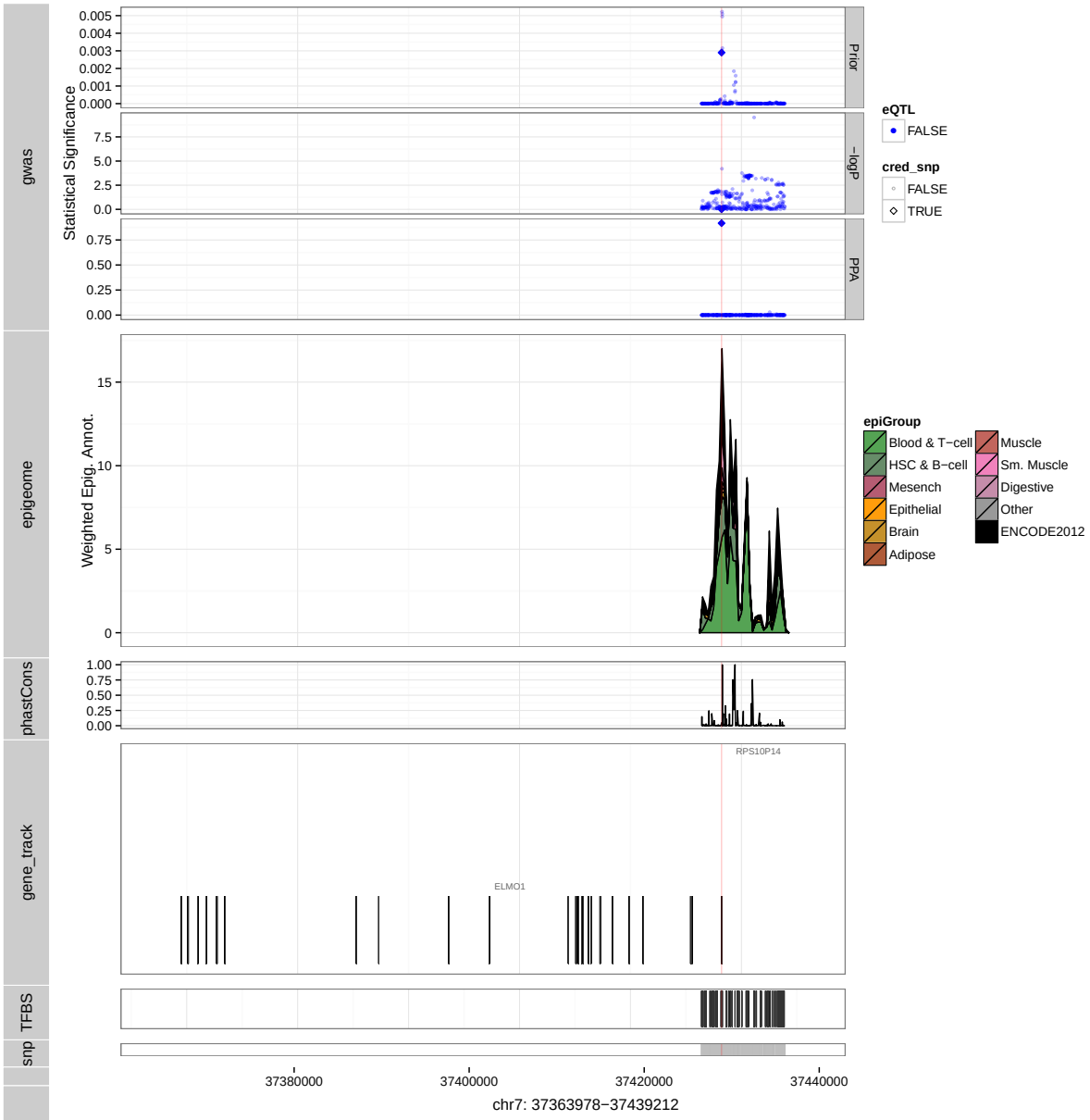
Psoriasis


gwas
epigeome
phastCons
gene_track
snp
TFBS
Statistical Significance
Prior
−logP
PPA
0.075
0.050
0.025
0.000
4
3
2
1
0
0.75
0.50
0.25
0.00
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
Weighted Epig. Annot.
40
20
0
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
1.00
0.75
0.50
0.25
0.00
RP1−111C20.4
C6orf99
yR211F11.2
RSPH3
TAGAP
RP11−13P5.1
159350000
159400000
159450000
159500000
159550000
chr6: 159328353−159541569

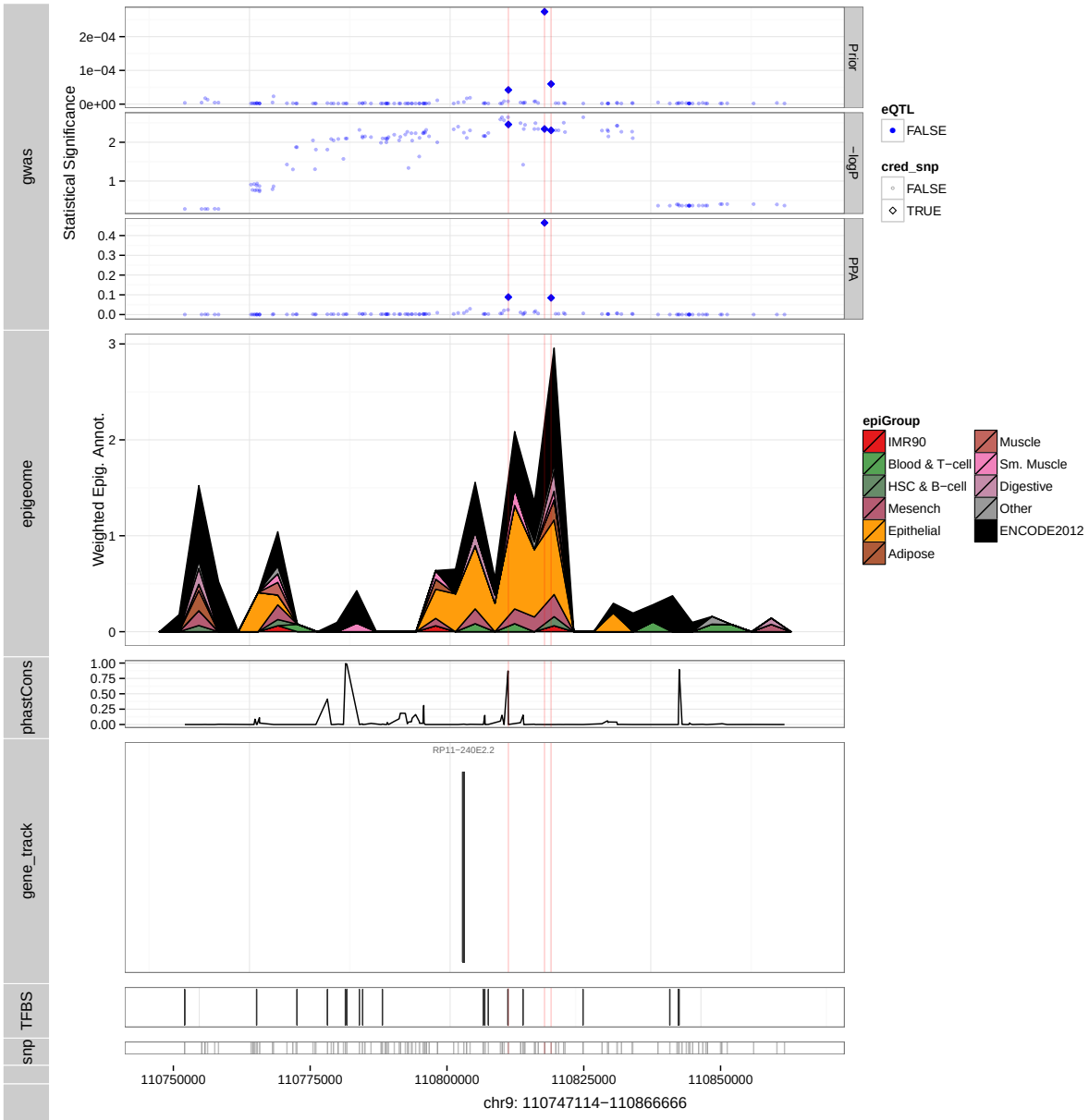
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
EXOC2
500000
520000
540000
560000
580000
chr6: 503573−577820
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
epiGroup
Blood & T−cell
Epithelial
Brain
Digestive
Other
ENCODE2012

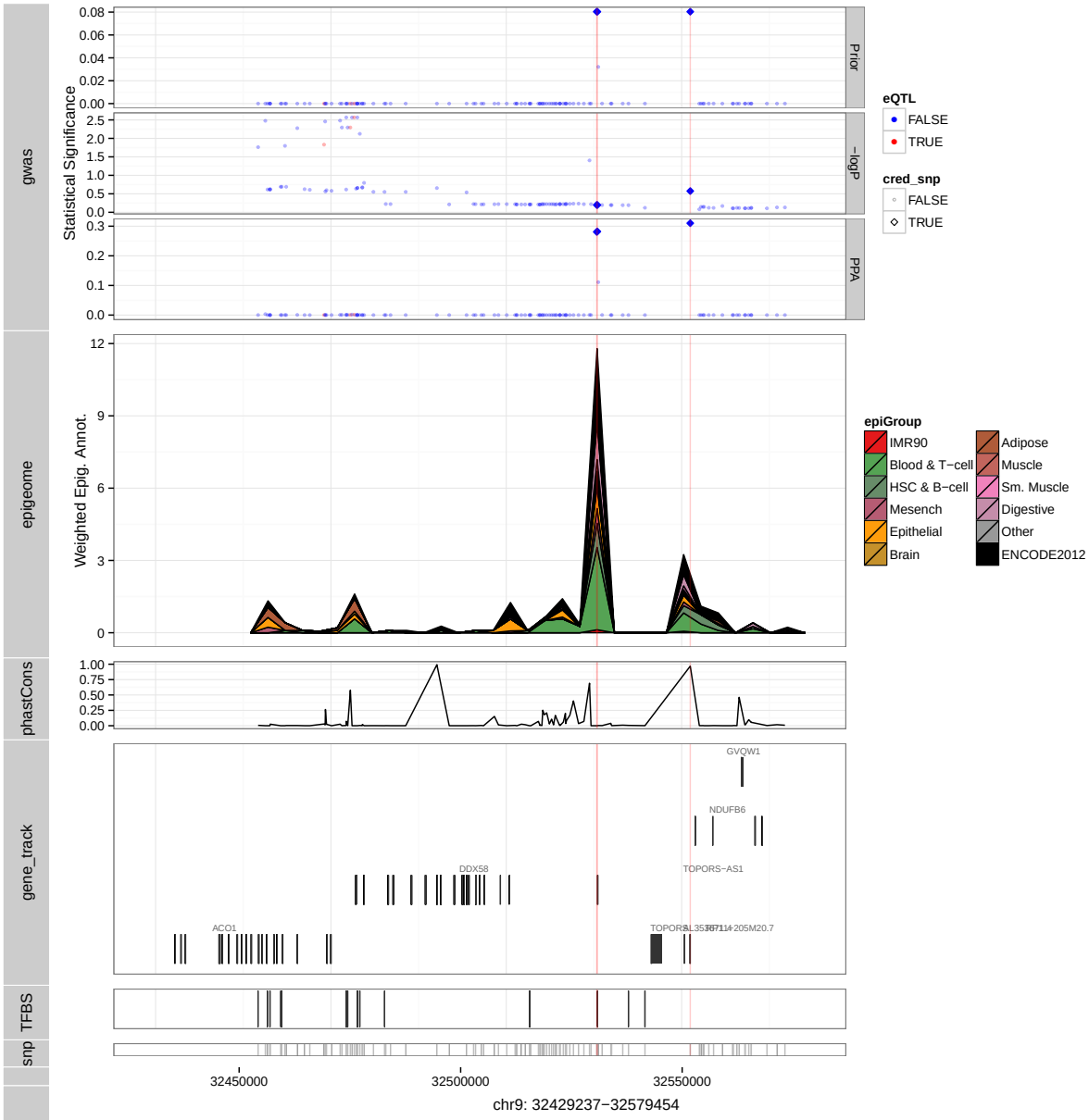
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
RPS10P14
ELMO1
37380000
37400000
37420000
37440000
chr7: 37363978−37439212
eQTL
FALSE
cred_snp
FALSE
TRUE
epiGroup
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012

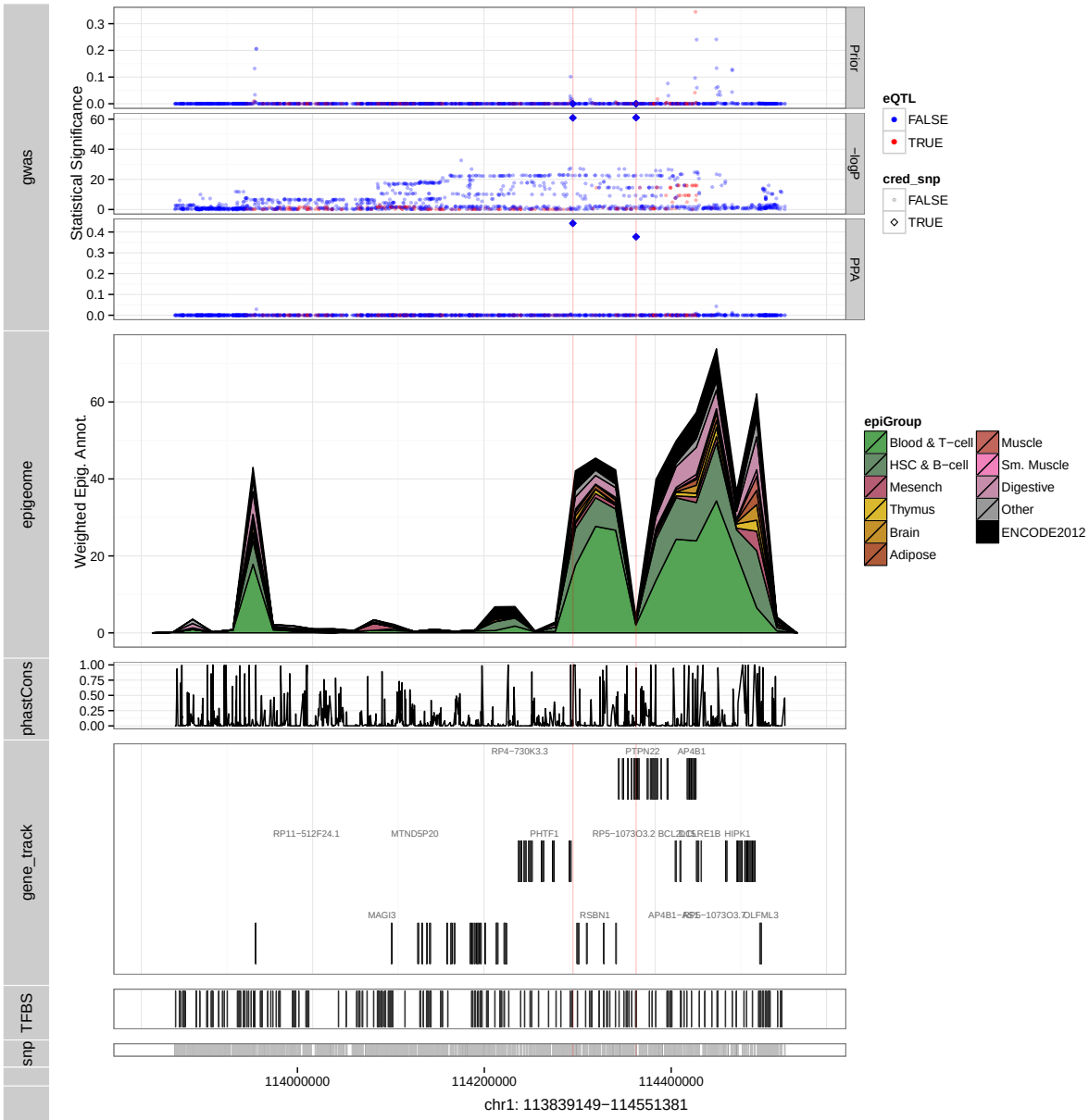
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
2e−04
1e−04
0e+00
2
1
0.4
0.3
0.2
0.1
0.0
eQTL
FALSE
cred_snp
FALSE
TRUE
Weighted Epig. Annot.
3
2
1
0
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
1.00
0.75
0.50
0.25
0.00
RP11−240E2.2
110750000
110775000
110800000
110825000
110850000
chr9: 110747114−110866666

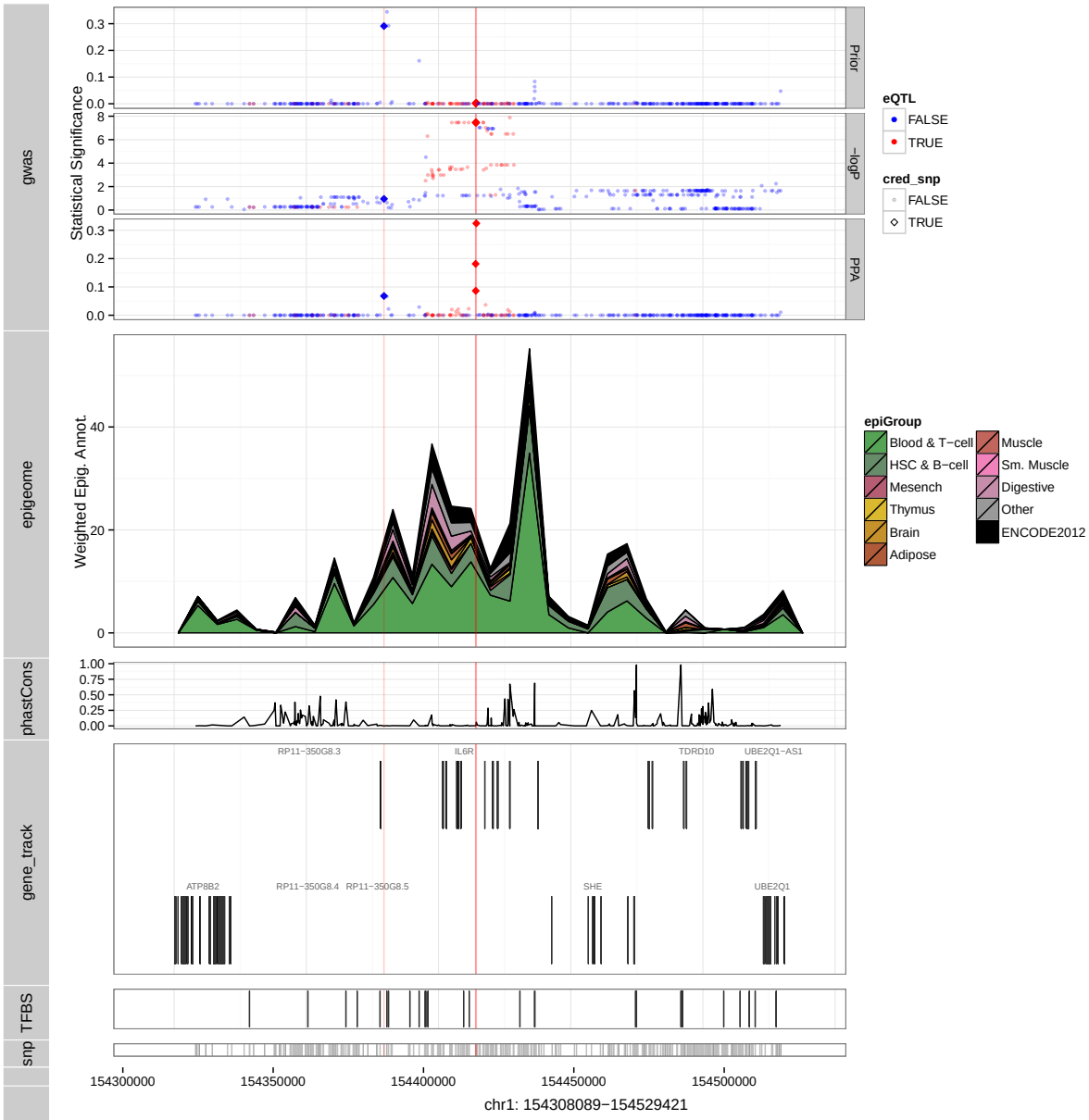
Psoriasis


gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
eQTL
FALSE
TRUE
cred_snp
FALSE
TRUE
Weighted Epig. Annot.
epiGroup
IMR90
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Brain
Adipose
Muscle
Sm. Muscle
Digestive
Other
ENCODE2012
GVQW1
NDUFB6
DDX58
TOPORS−AS1
ACO1
32450000
32500000
32550000
chr9: 32429237−32579454

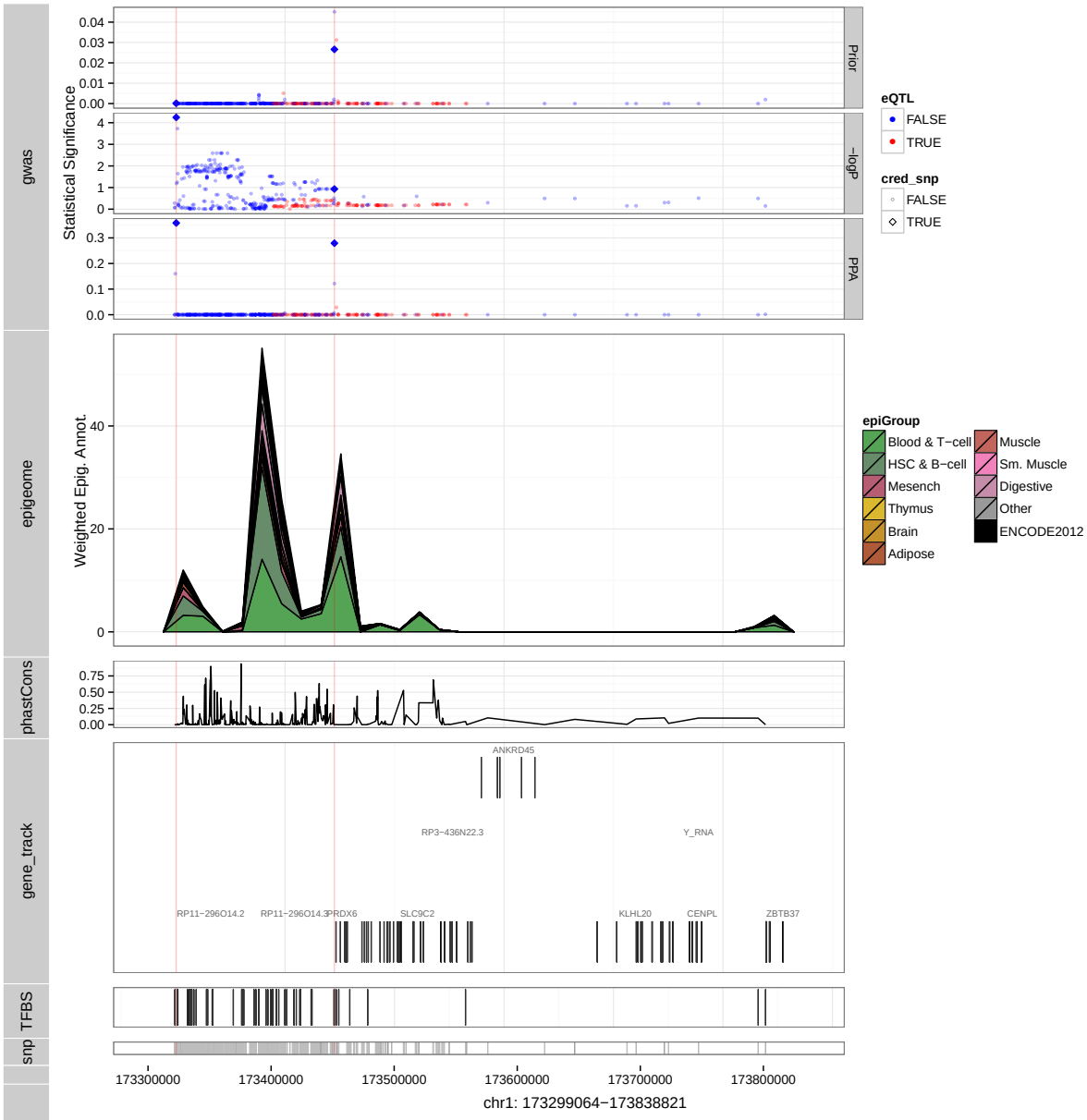
Rheumatoid Arthritis



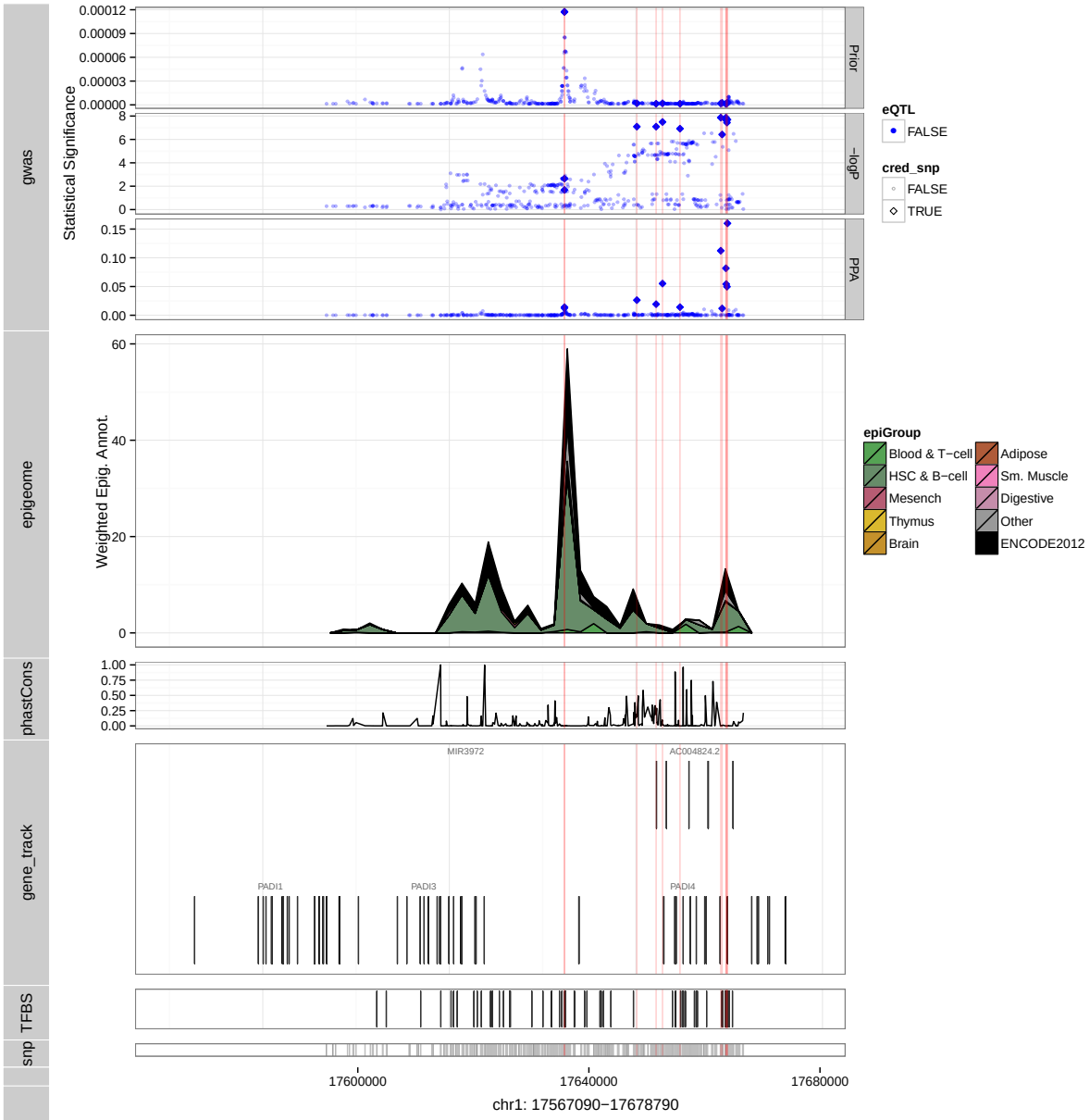
Rheumatoid Arthritis



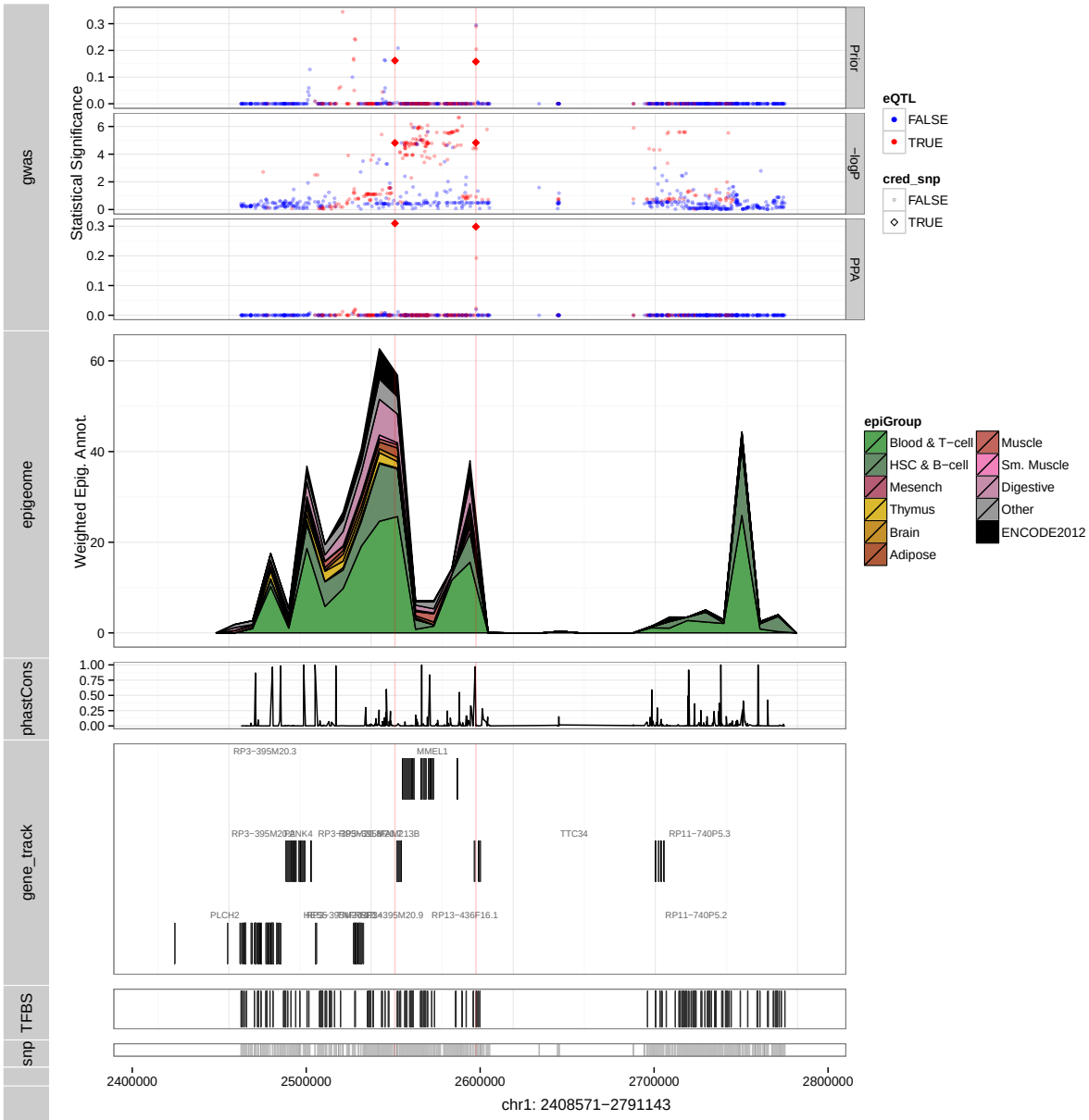
Rheumatoid Arthritis



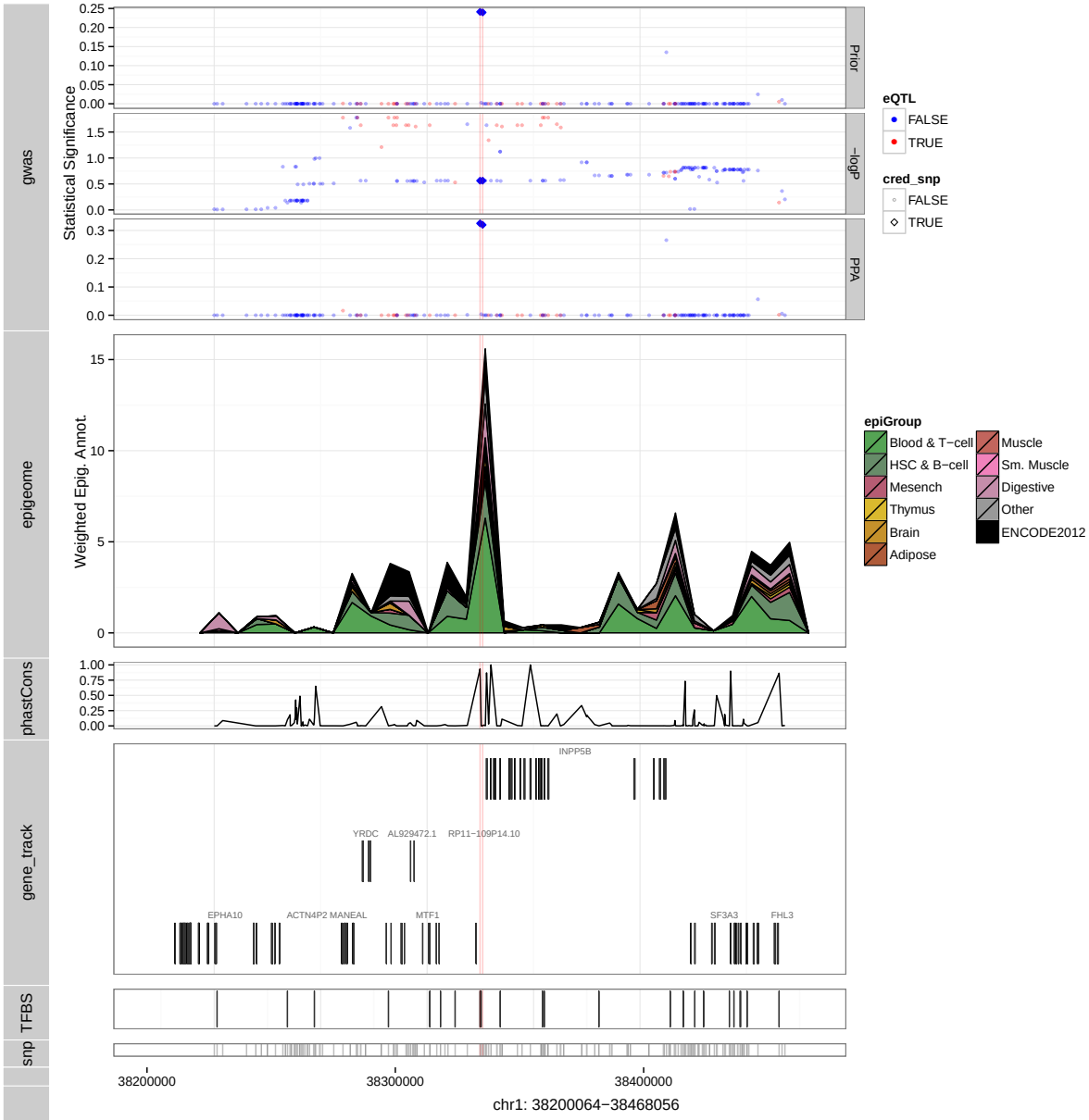
Rheumatoid Arthritis



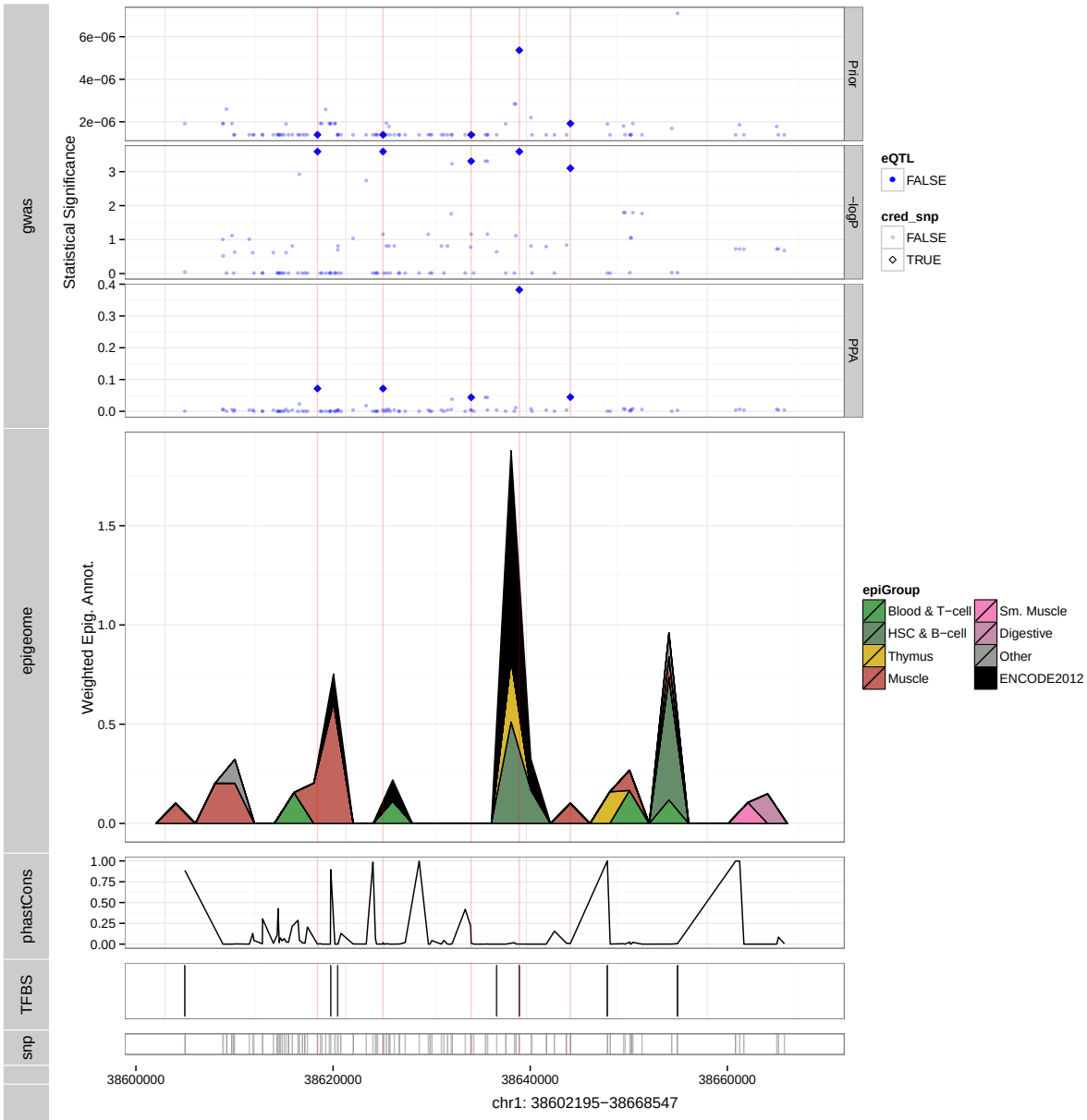
Rheumatoid Arthritis



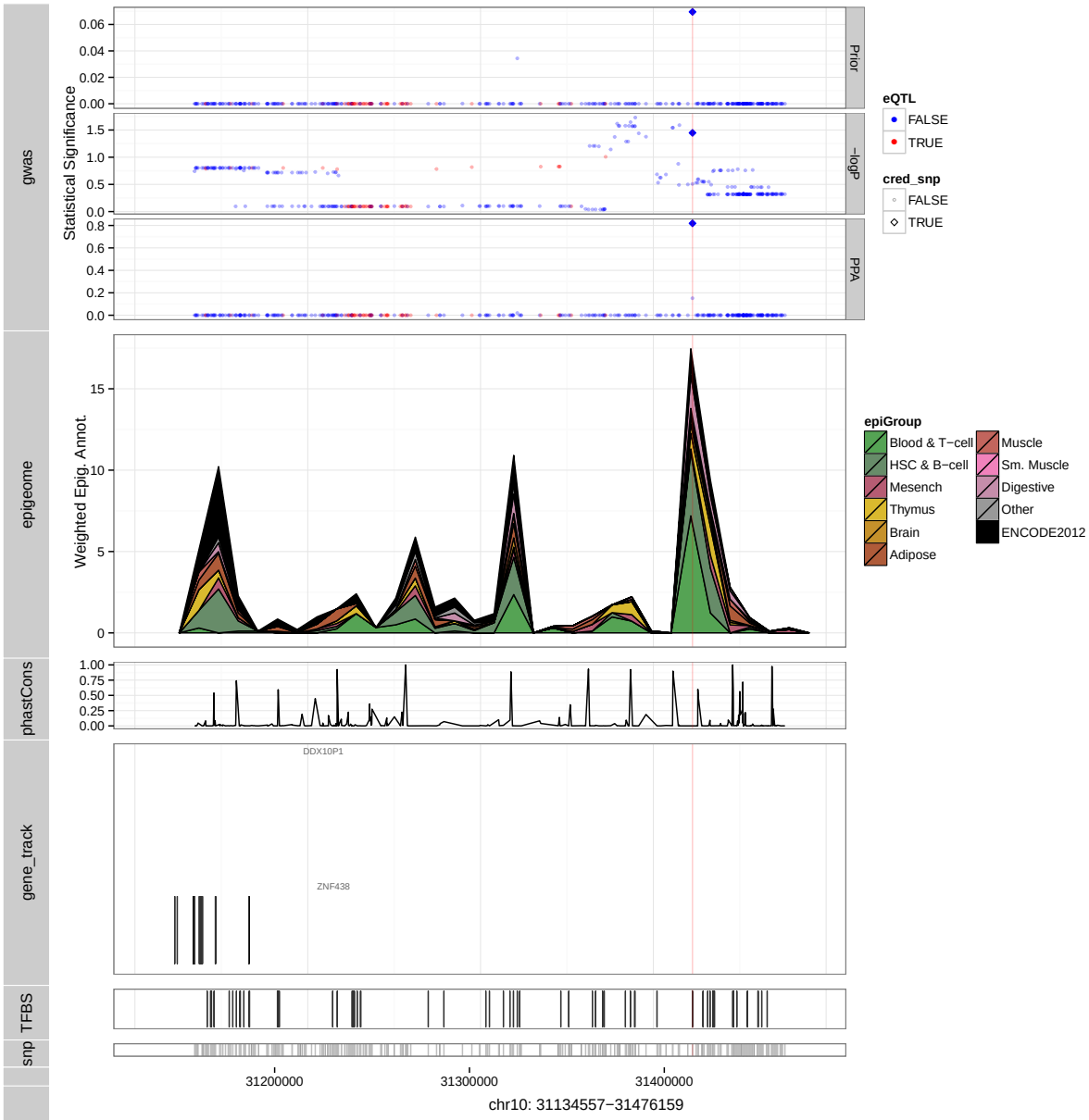
Rheumatoid Arthritis



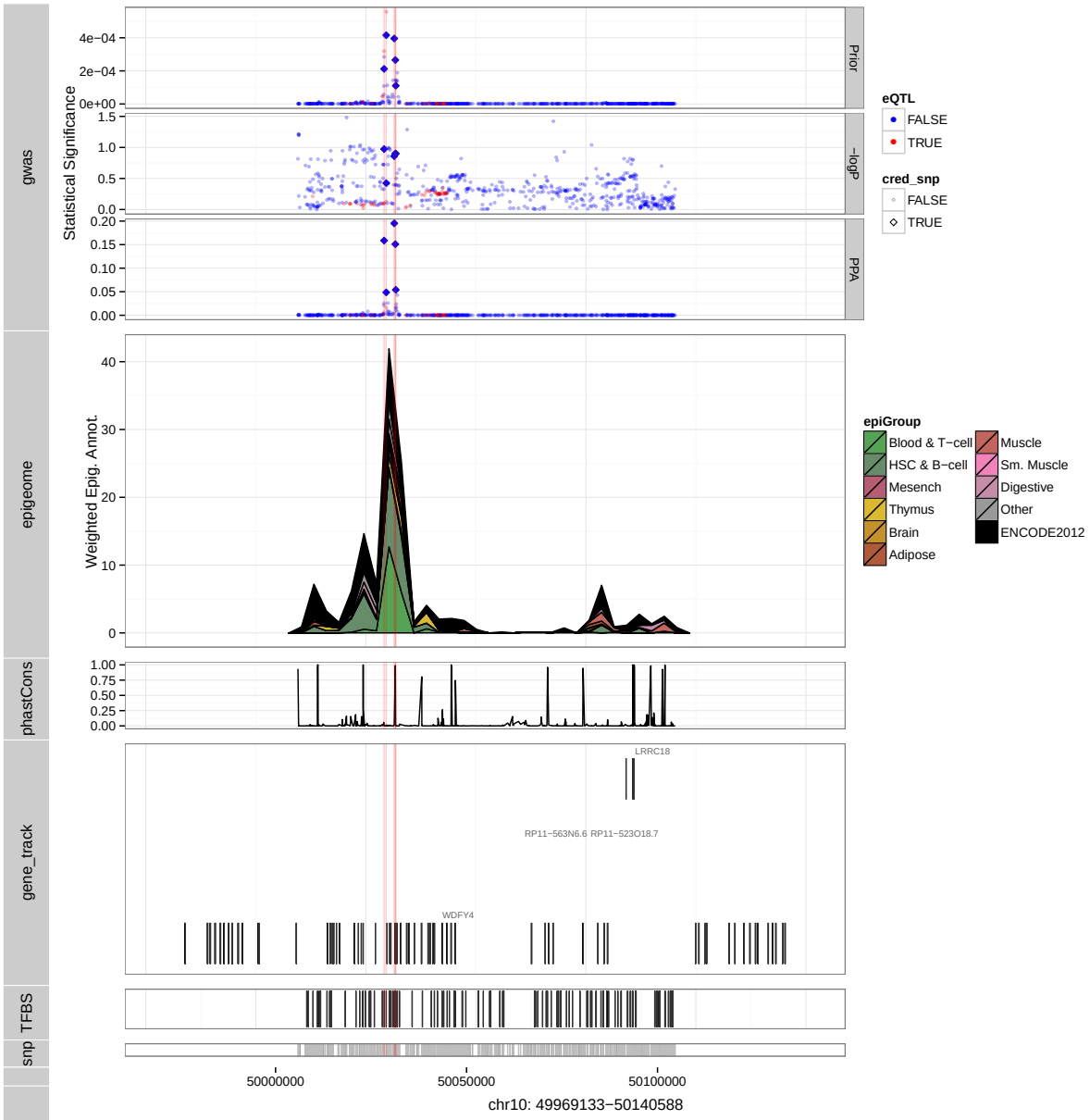
Rheumatoid Arthritis



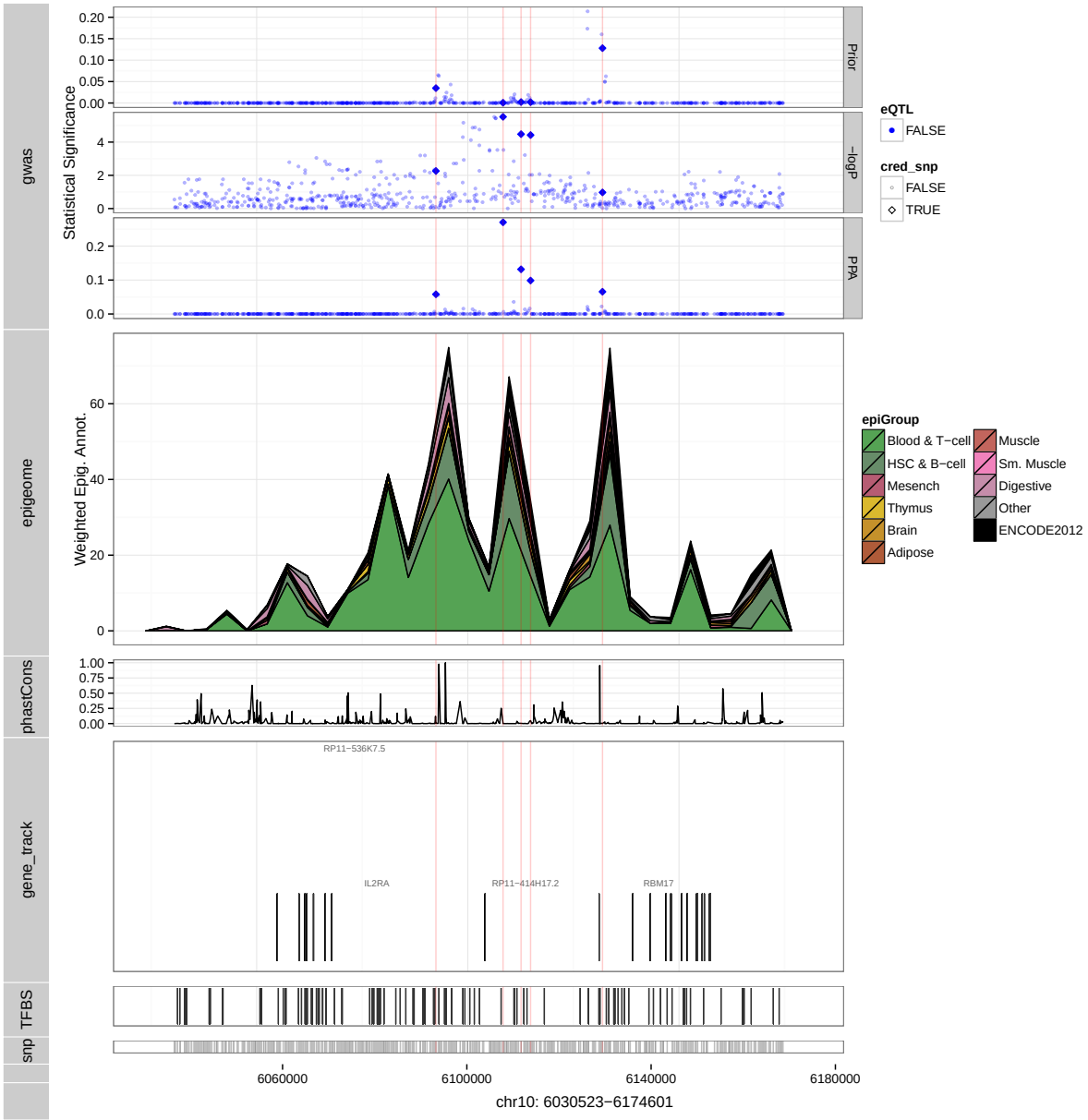
Rheumatoid Arthritis



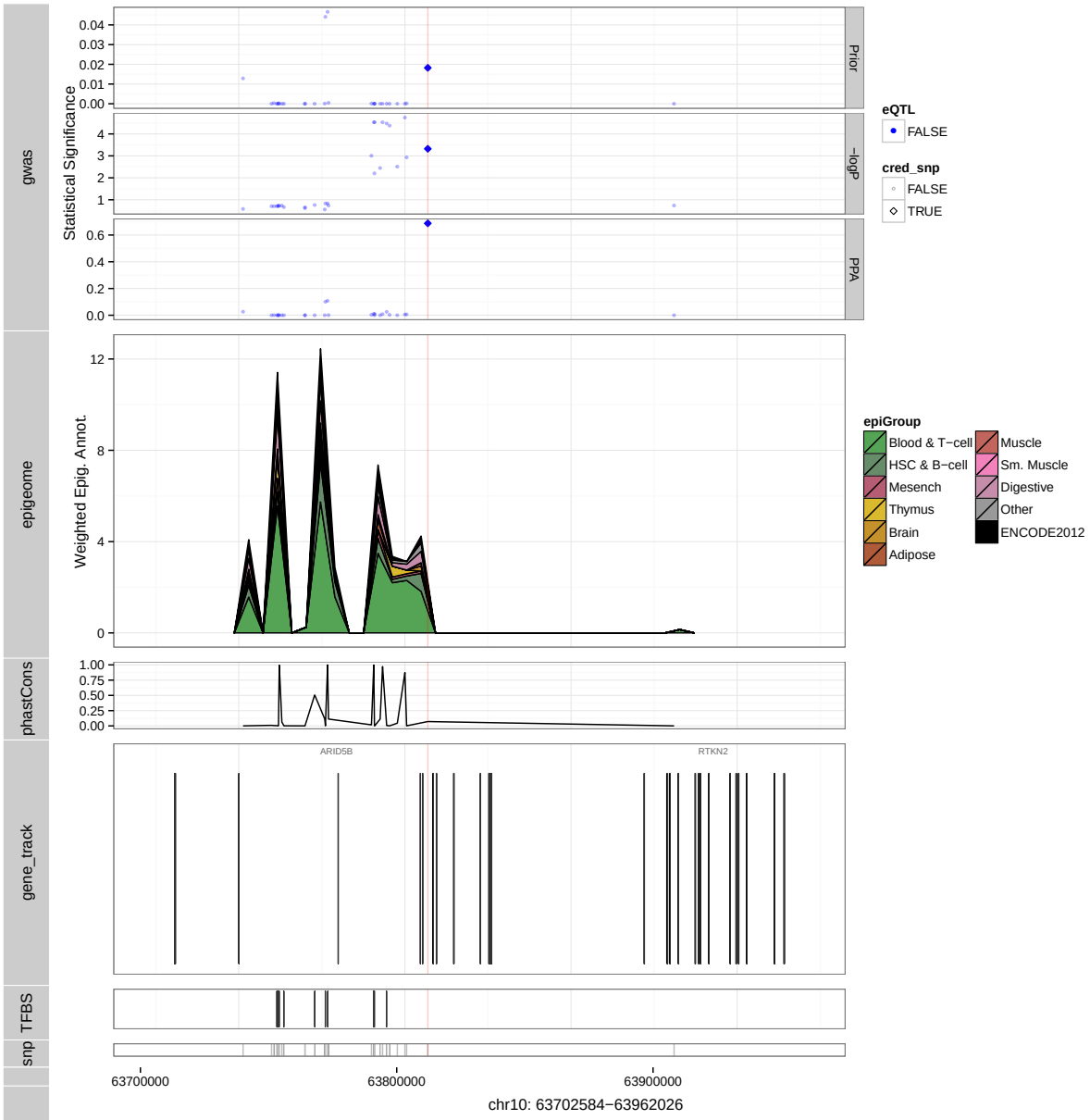
Rheumatoid Arthritis



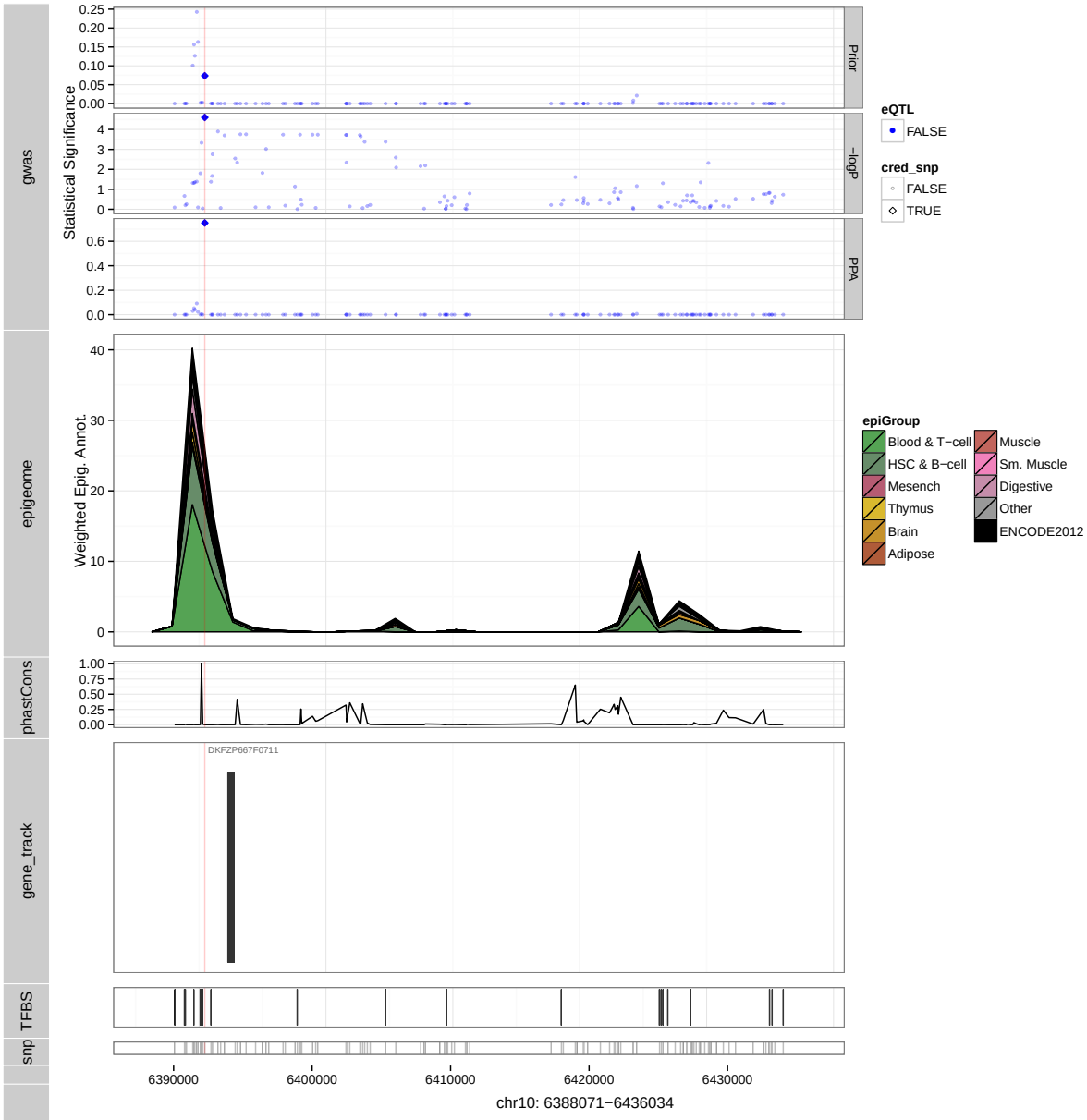
Rheumatoid Arthritis



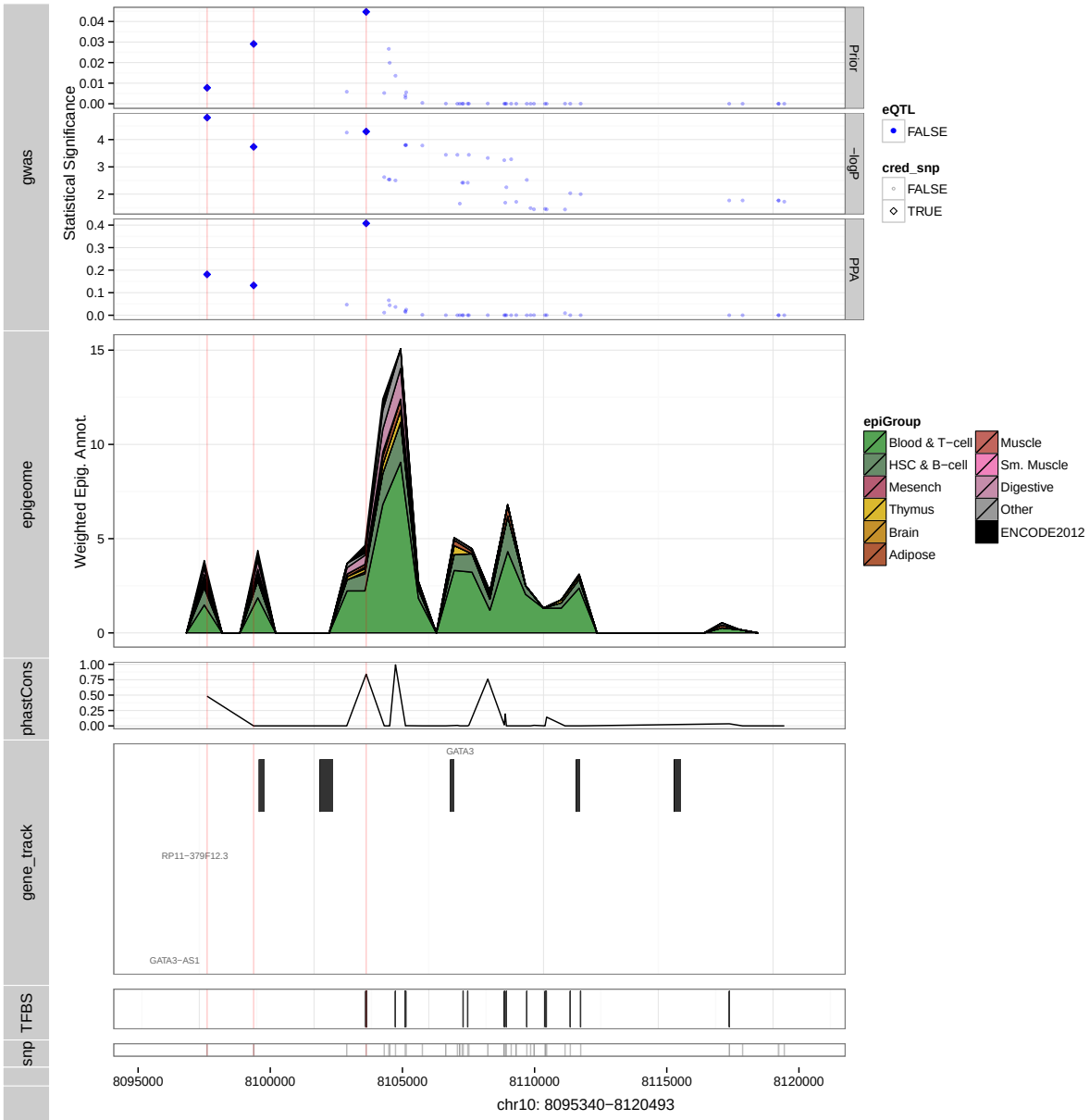
Rheumatoid Arthritis



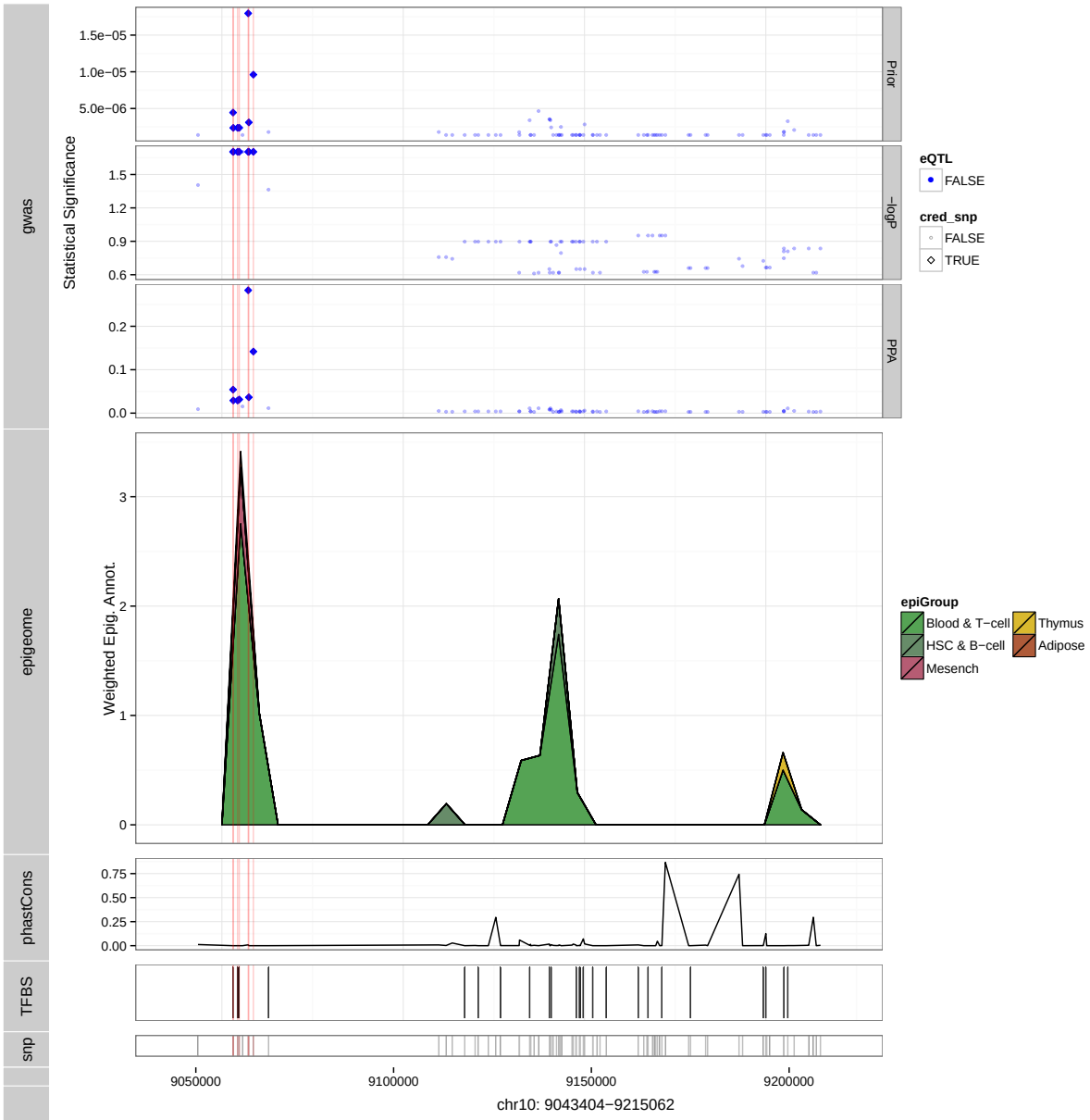
Rheumatoid Arthritis



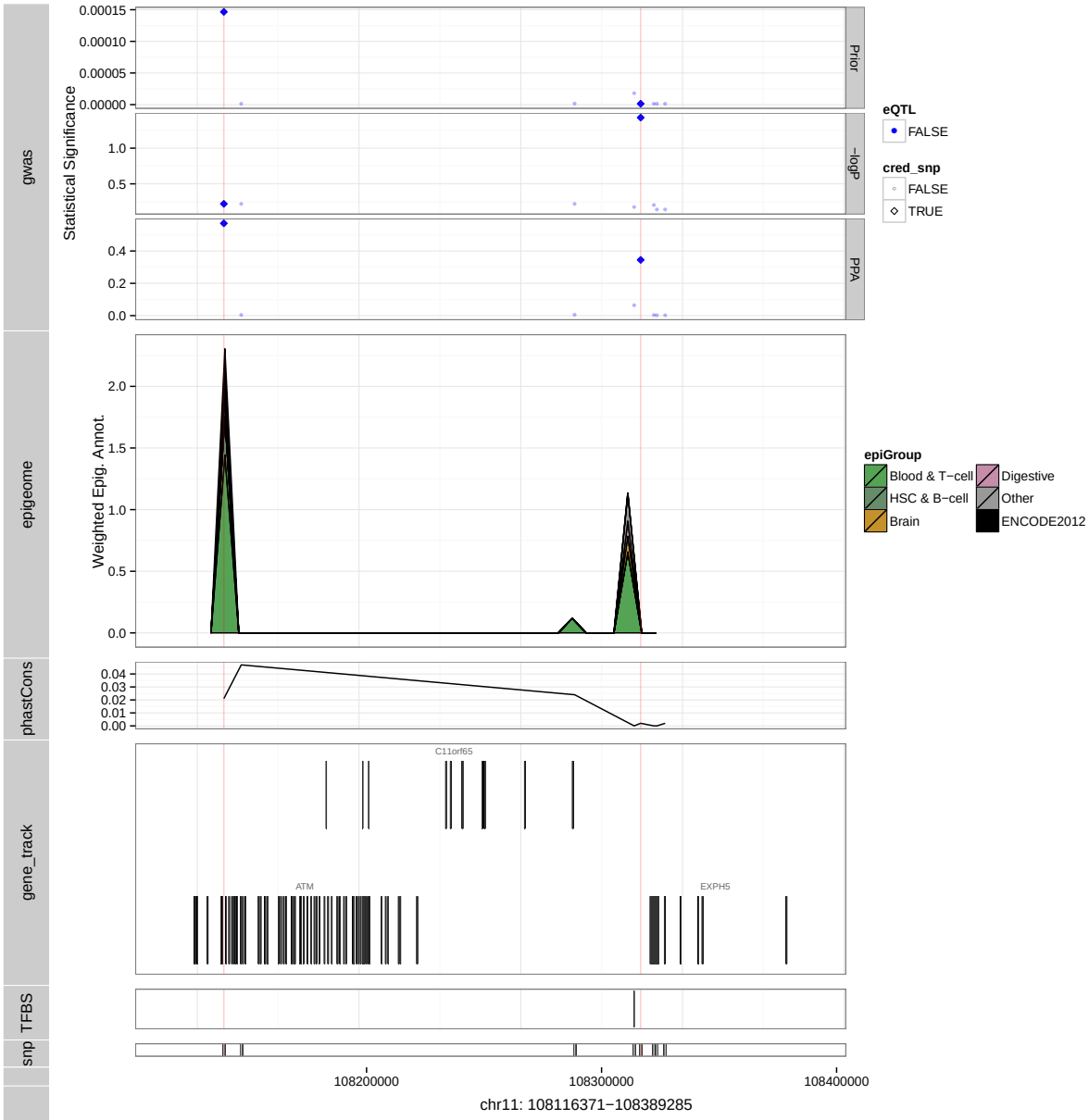
Rheumatoid Arthritis



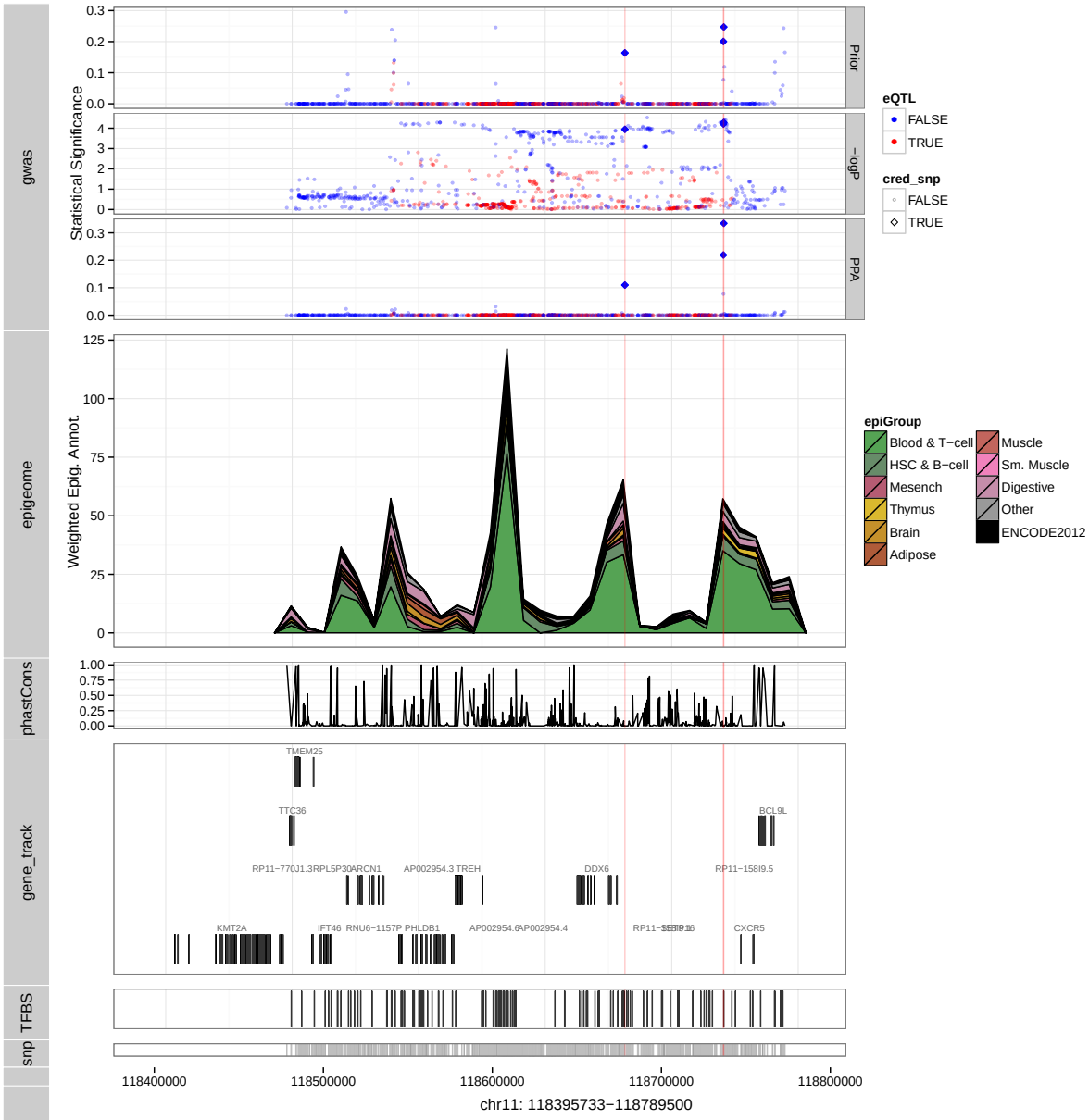
Rheumatoid Arthritis



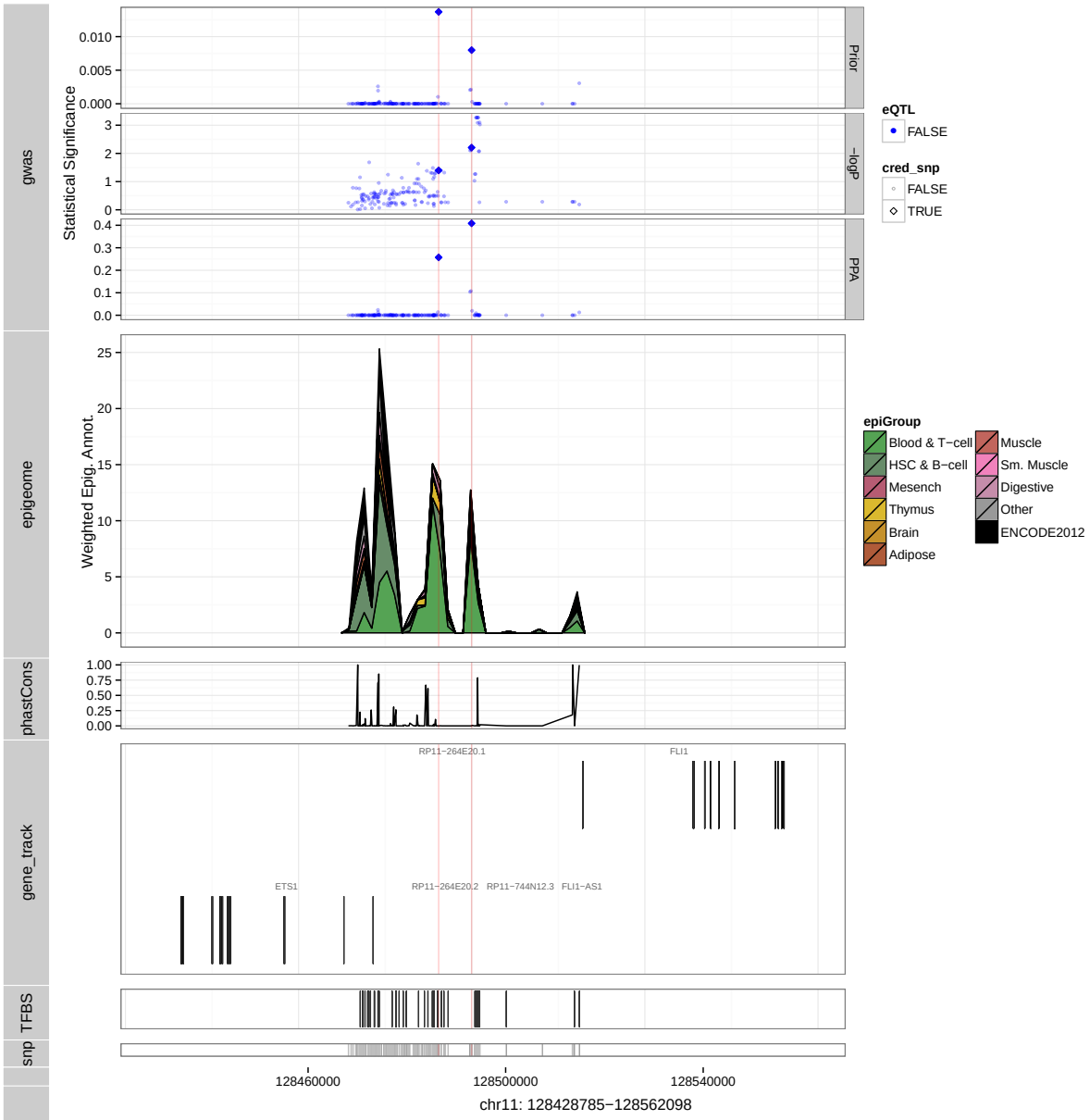
Rheumatoid Arthritis



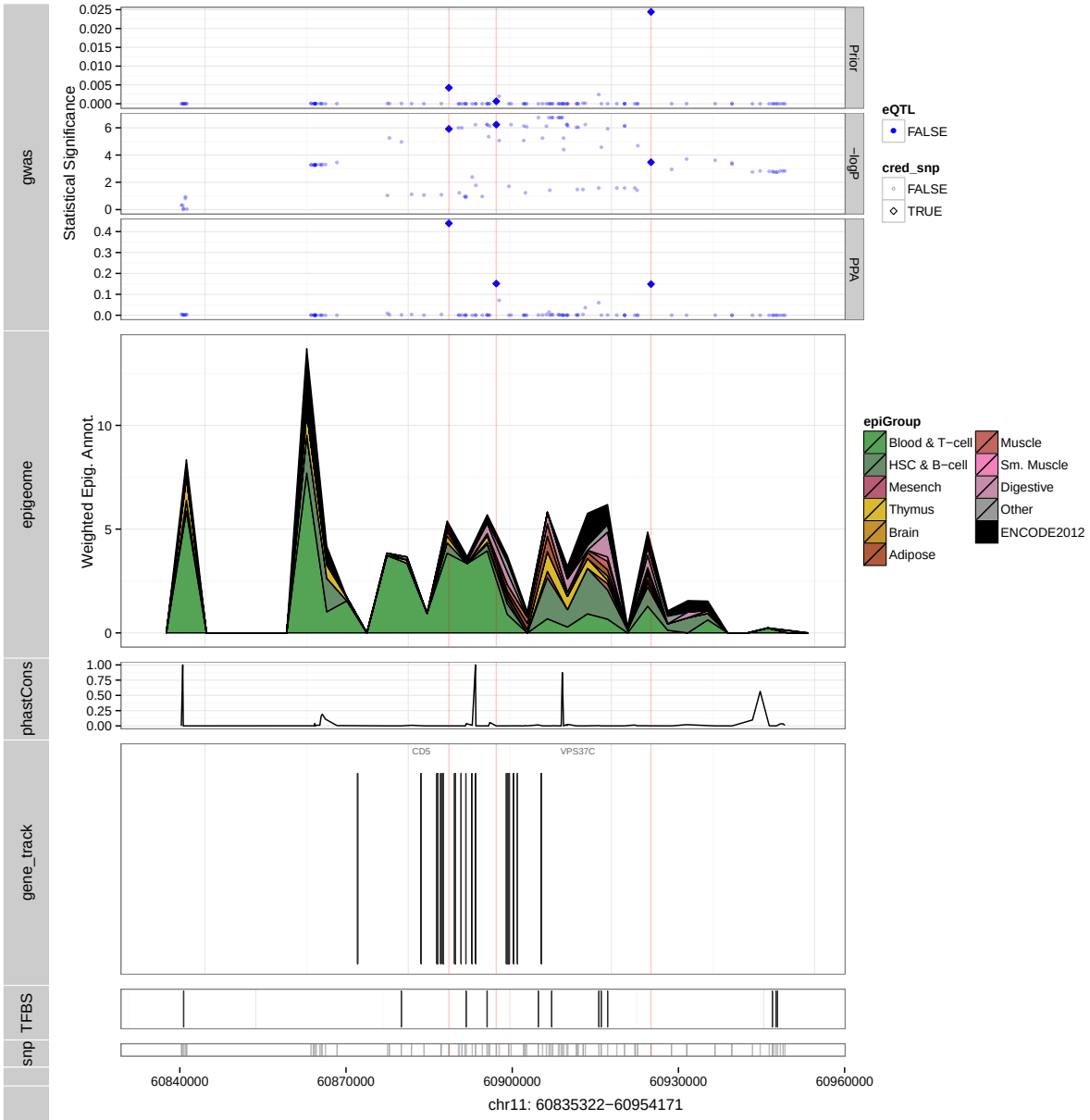
Rheumatoid Arthritis



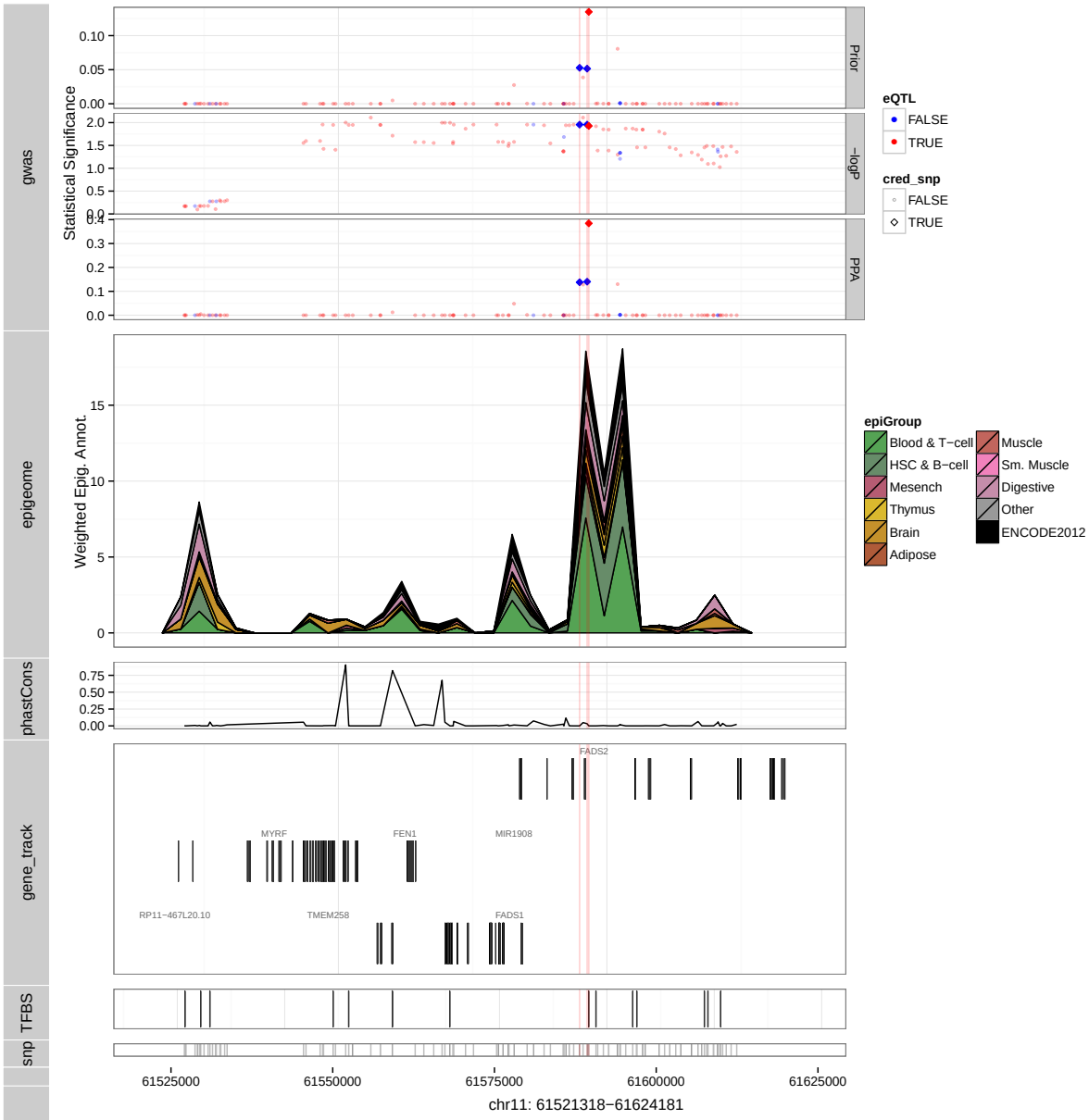
Rheumatoid Arthritis



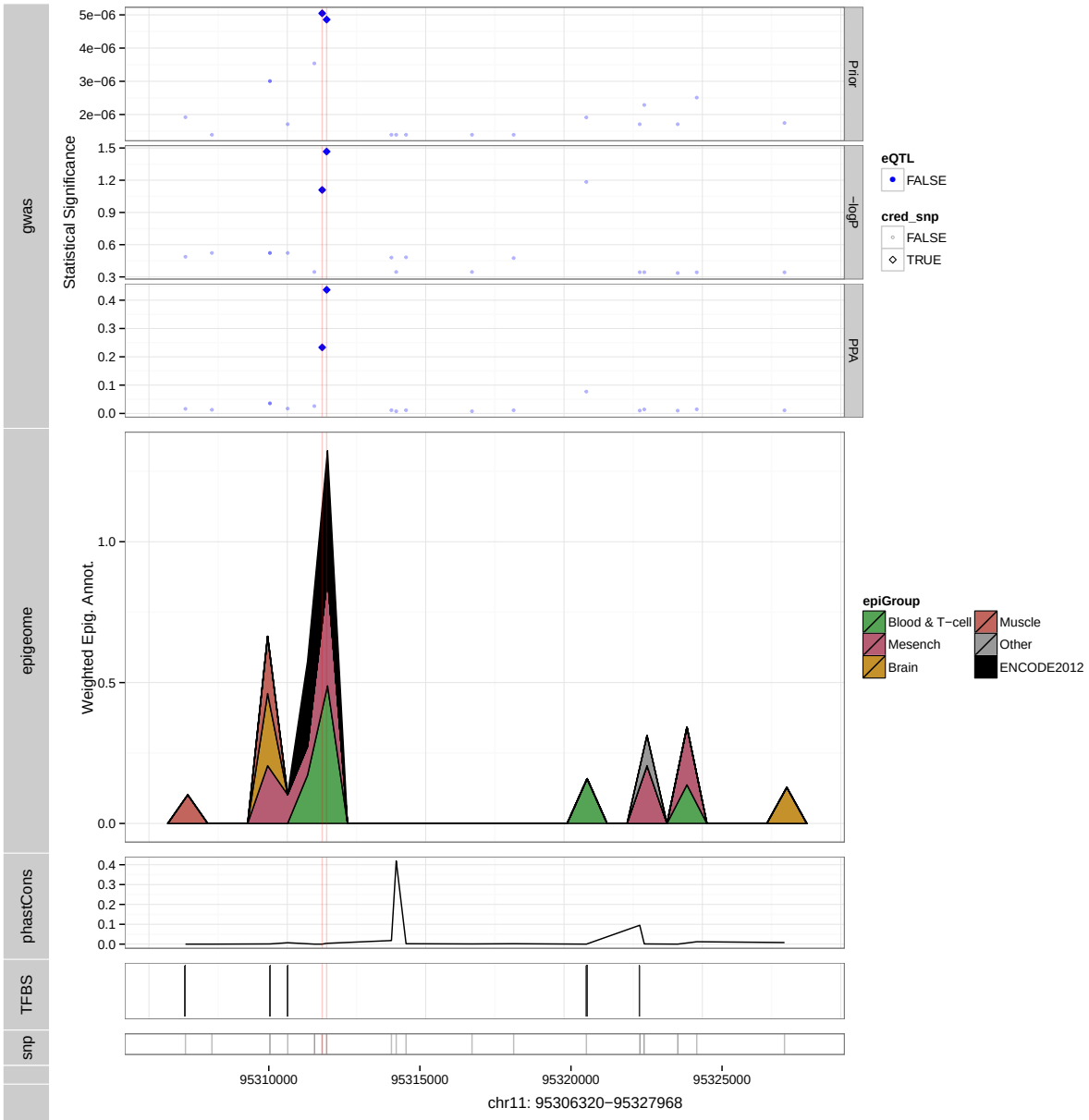
Rheumatoid Arthritis



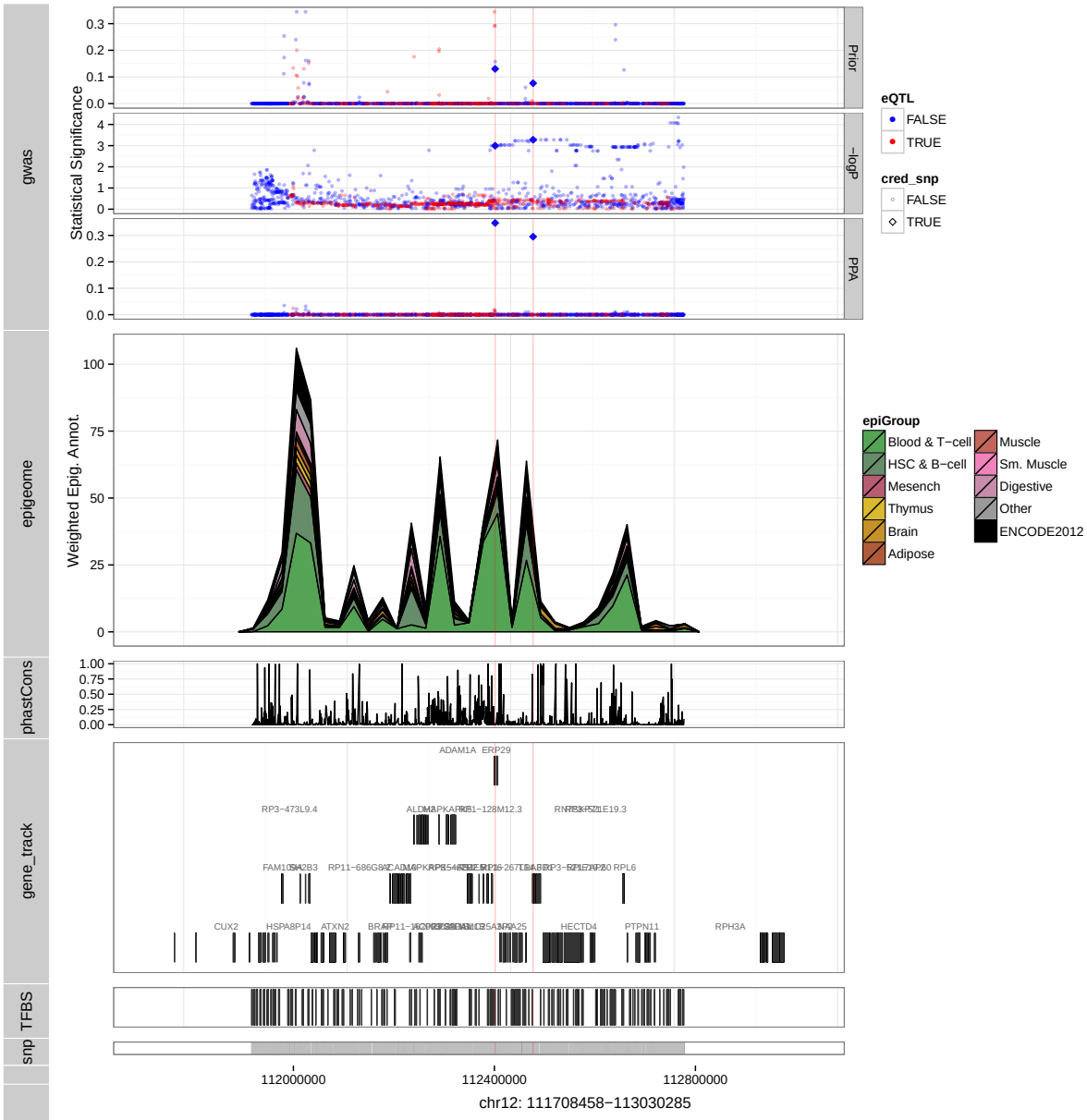
Rheumatoid Arthritis



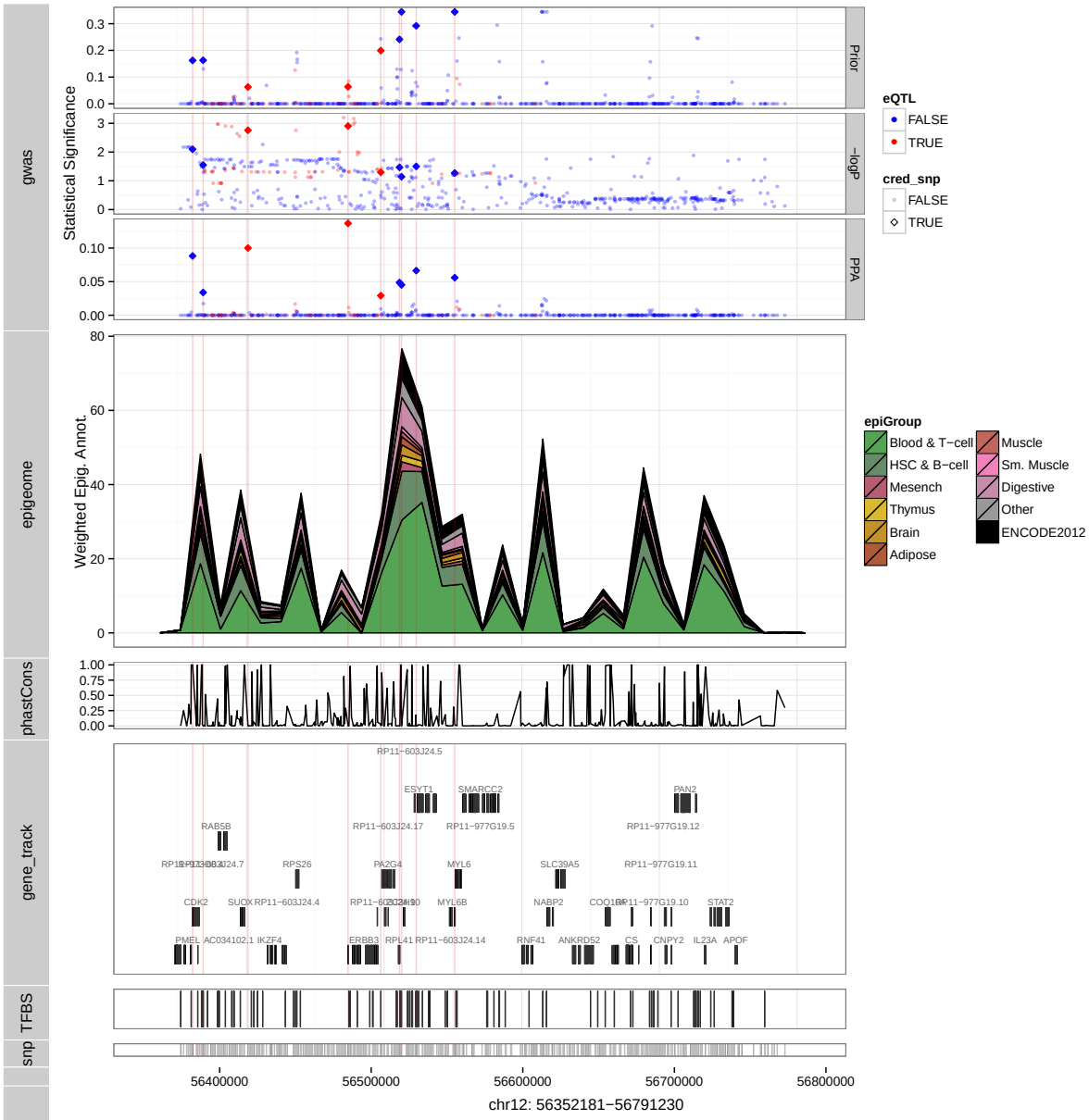
Rheumatoid Arthritis



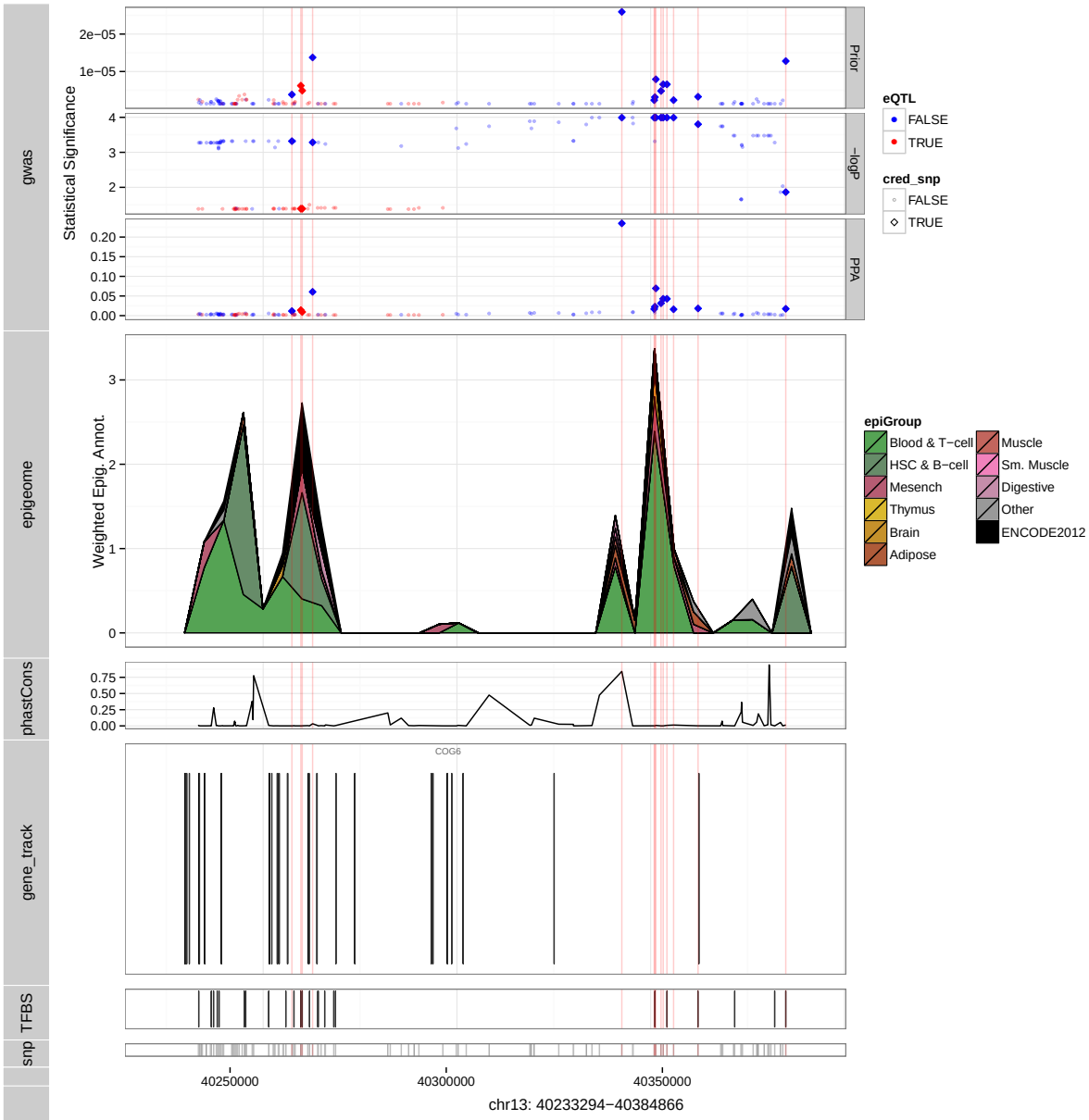
Rheumatoid Arthritis



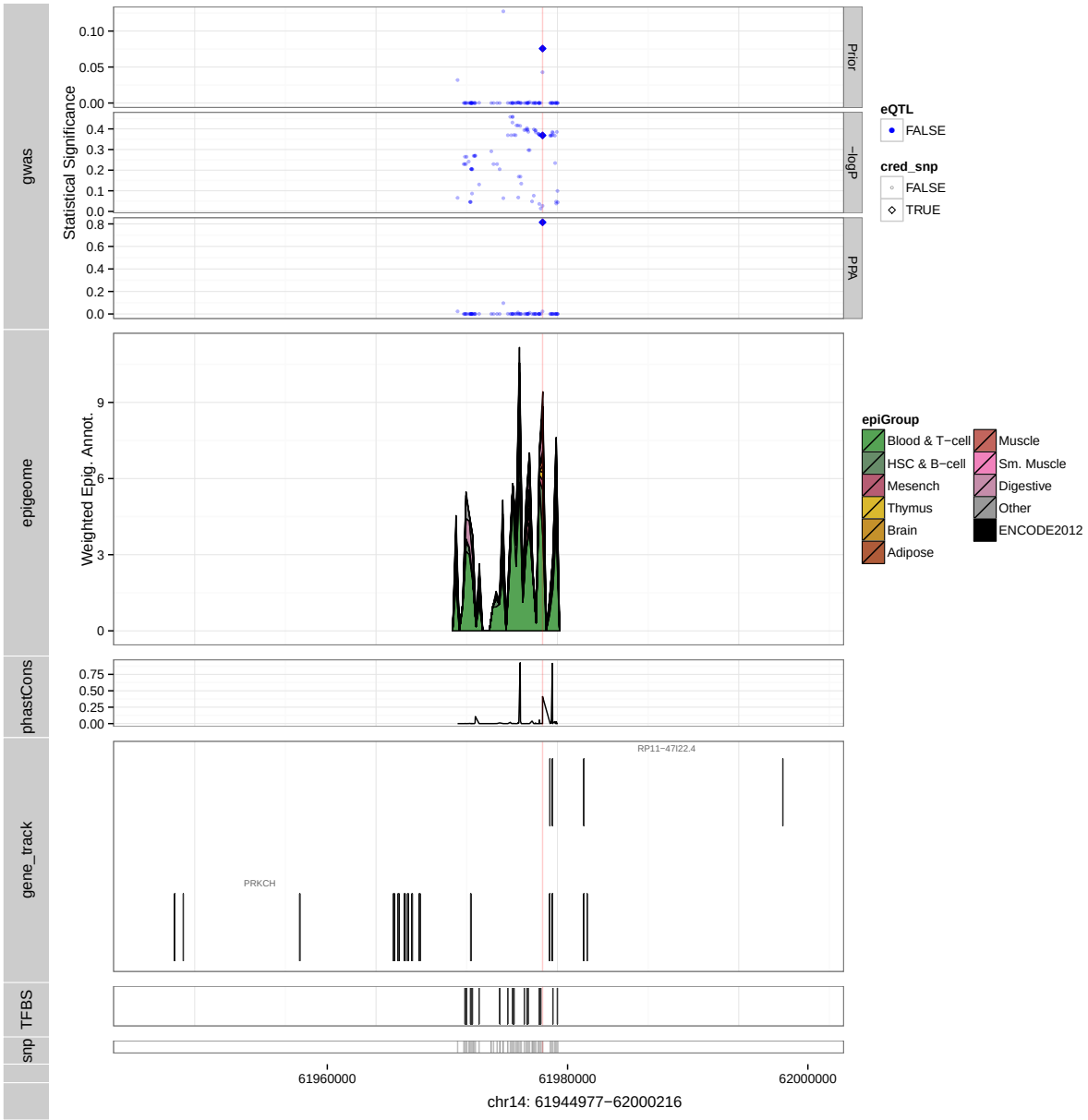
Rheumatoid Arthritis



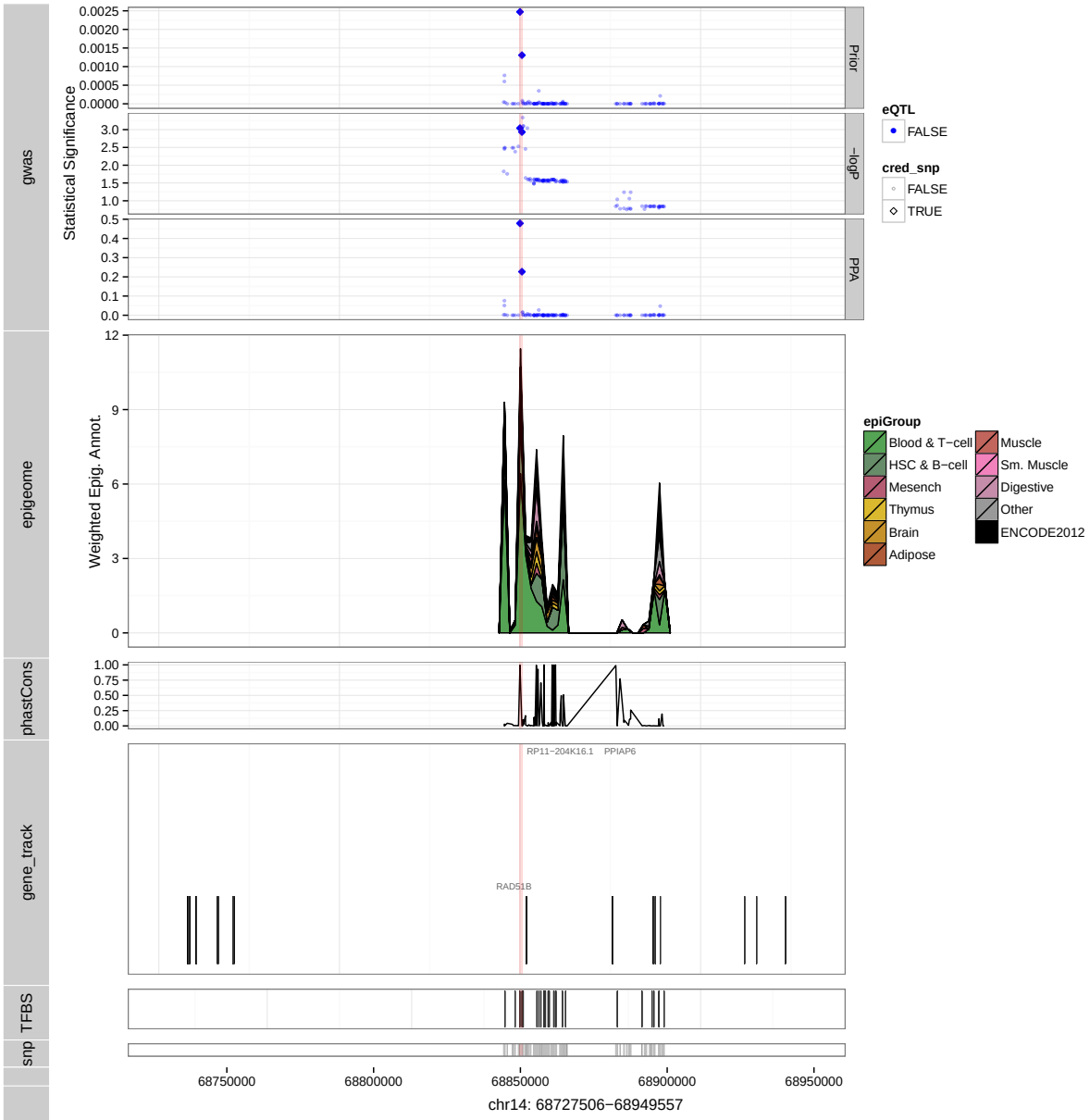
Rheumatoid Arthritis



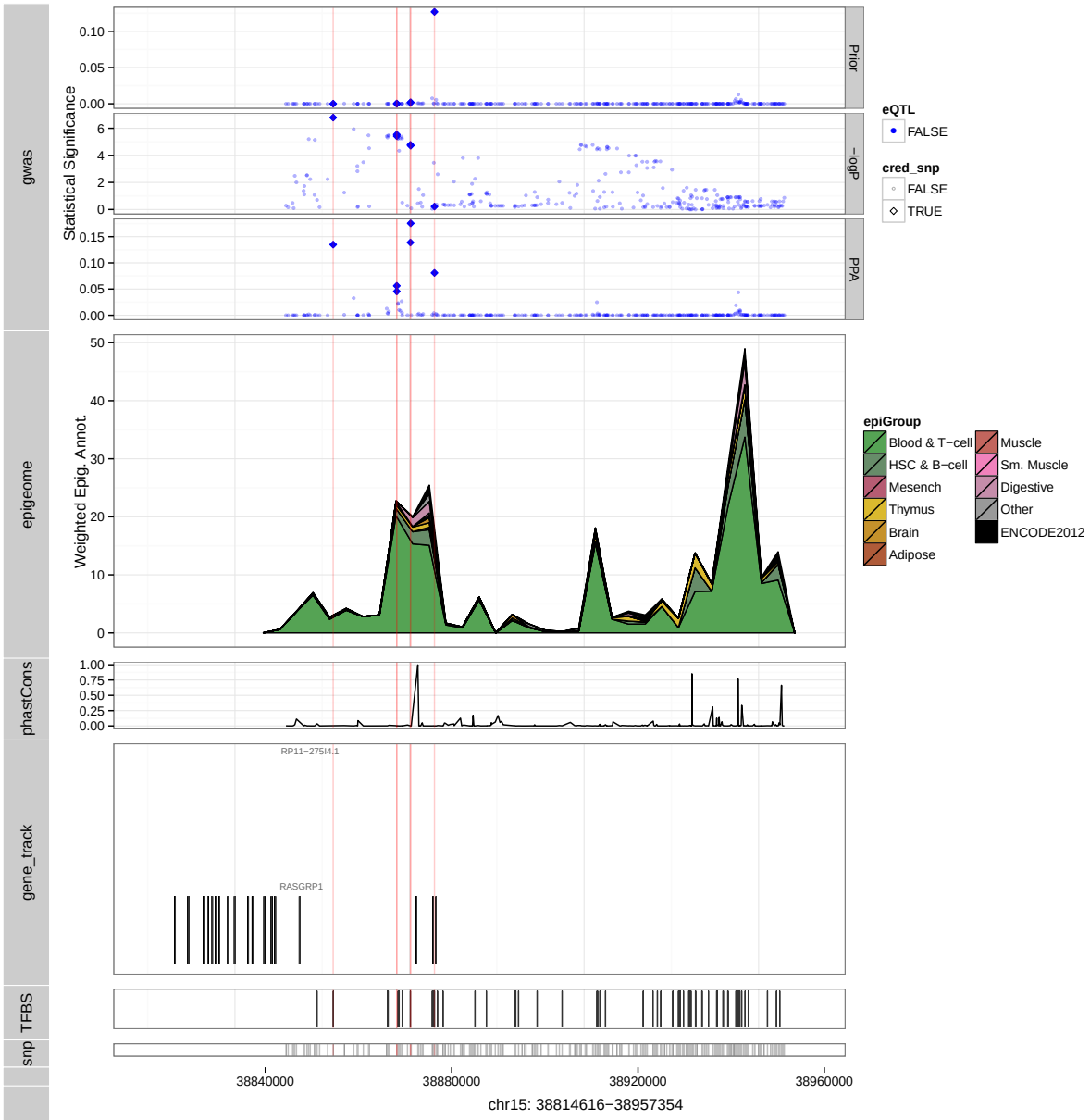
Rheumatoid Arthritis



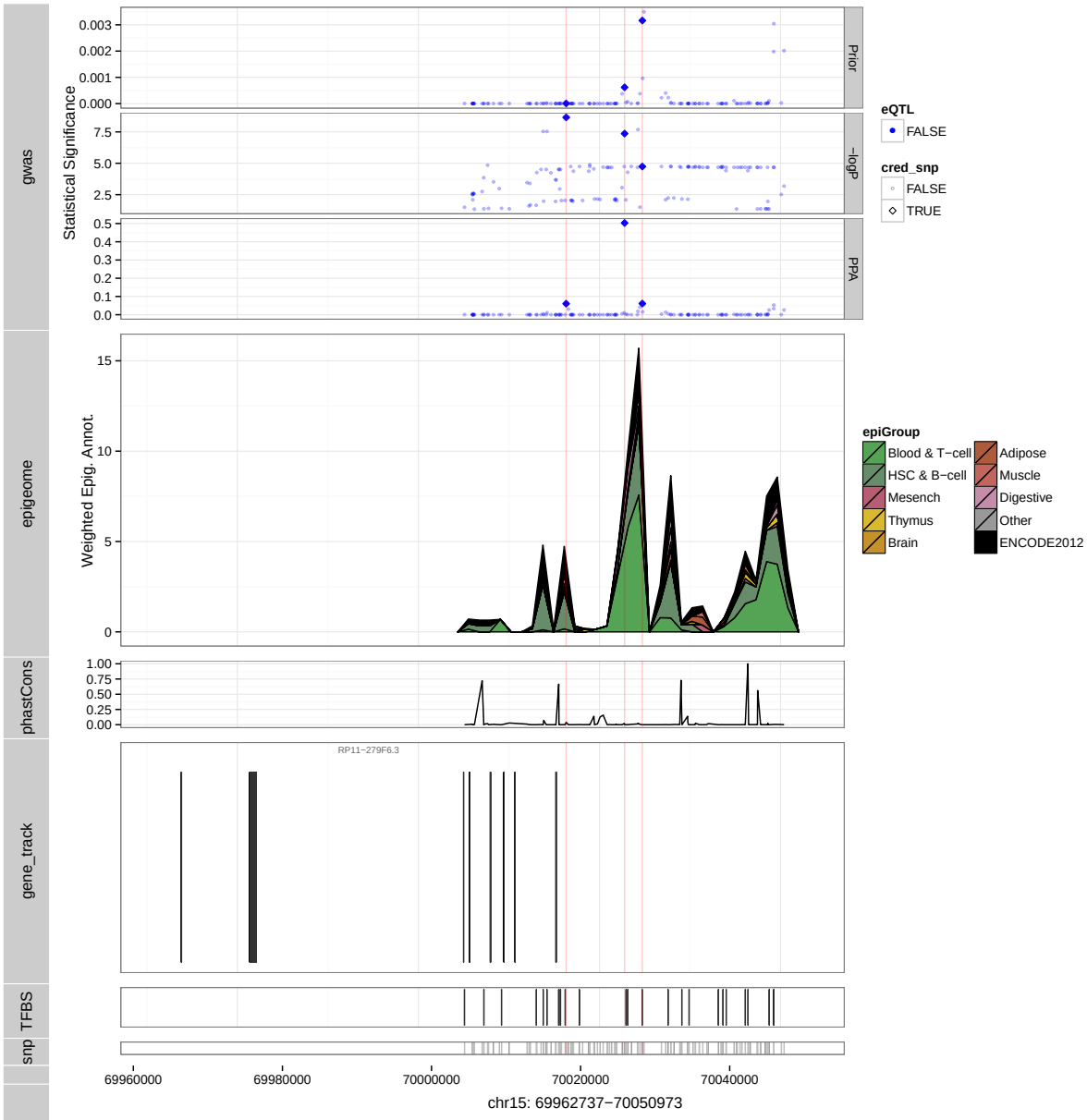
Rheumatoid Arthritis



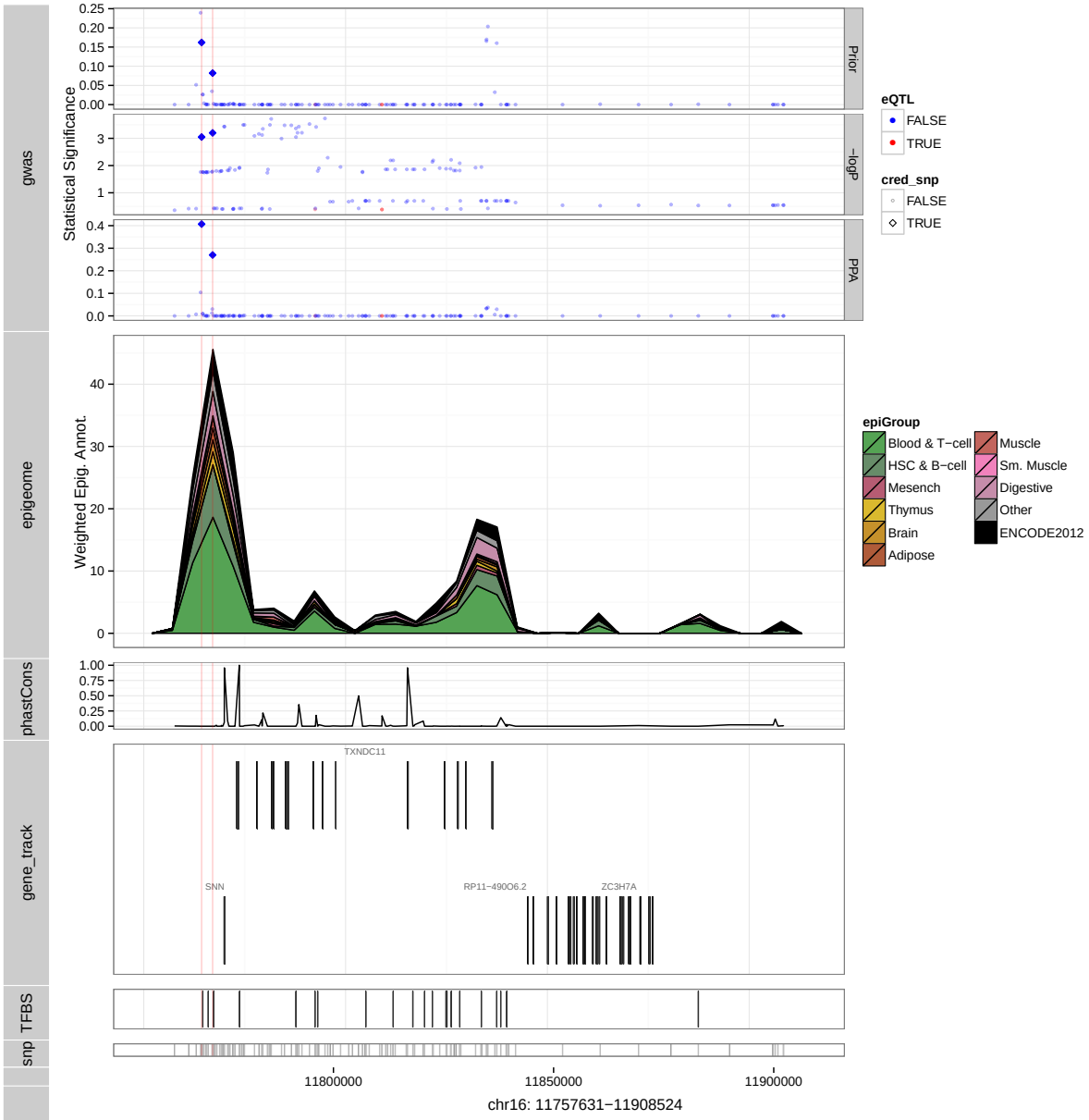
Rheumatoid Arthritis



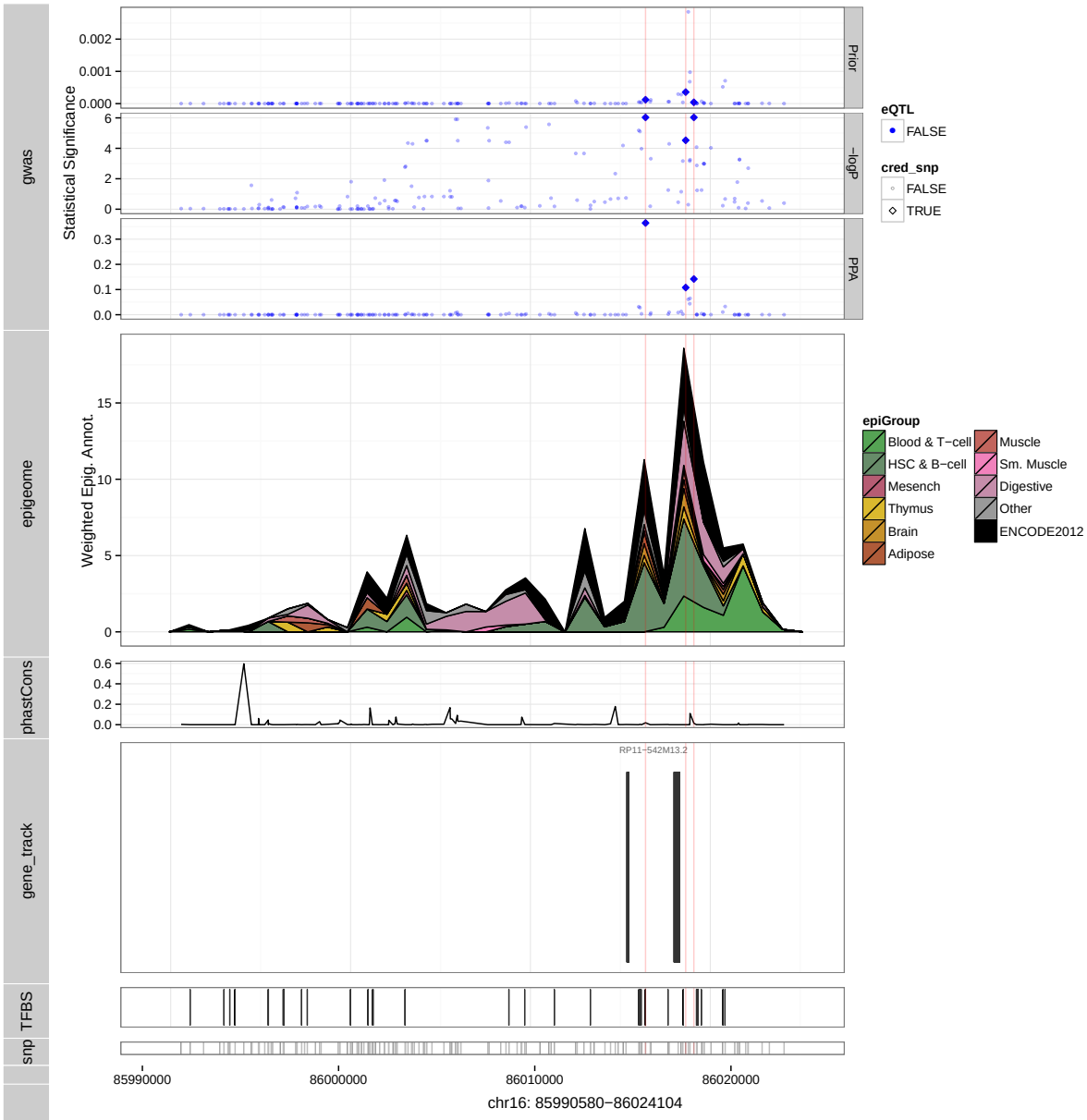
Rheumatoid Arthritis



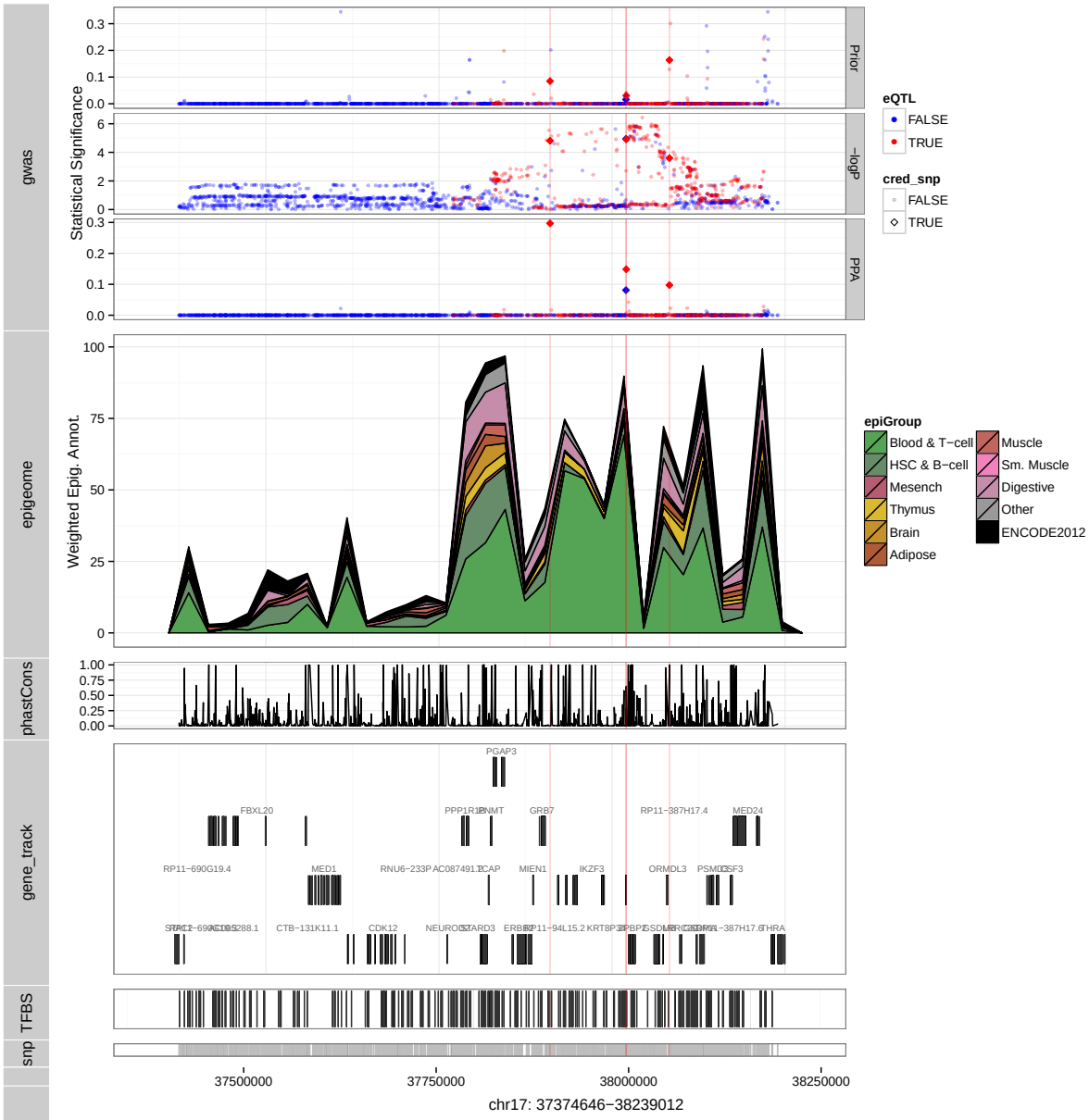
Rheumatoid Arthritis



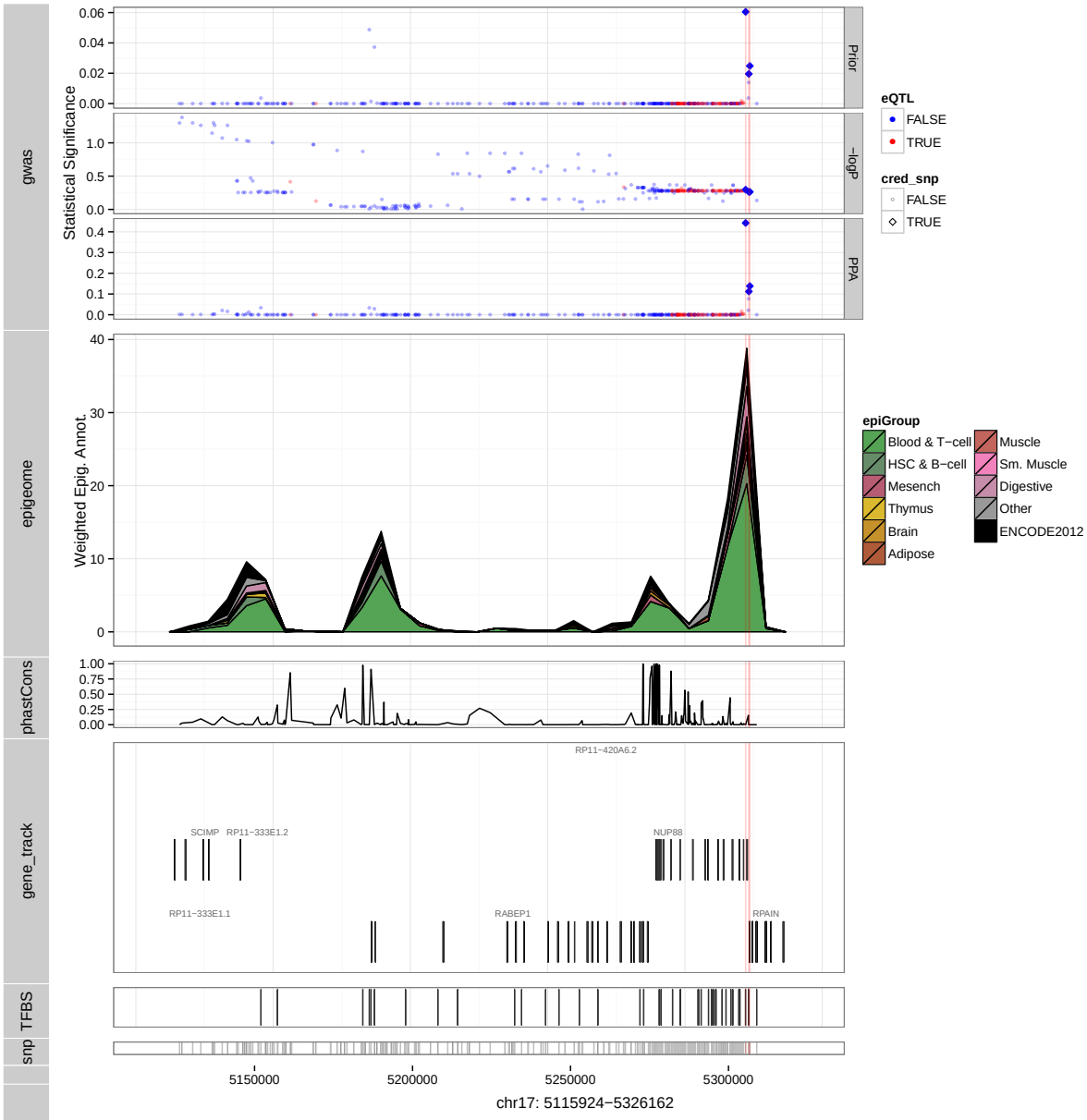
Rheumatoid Arthritis



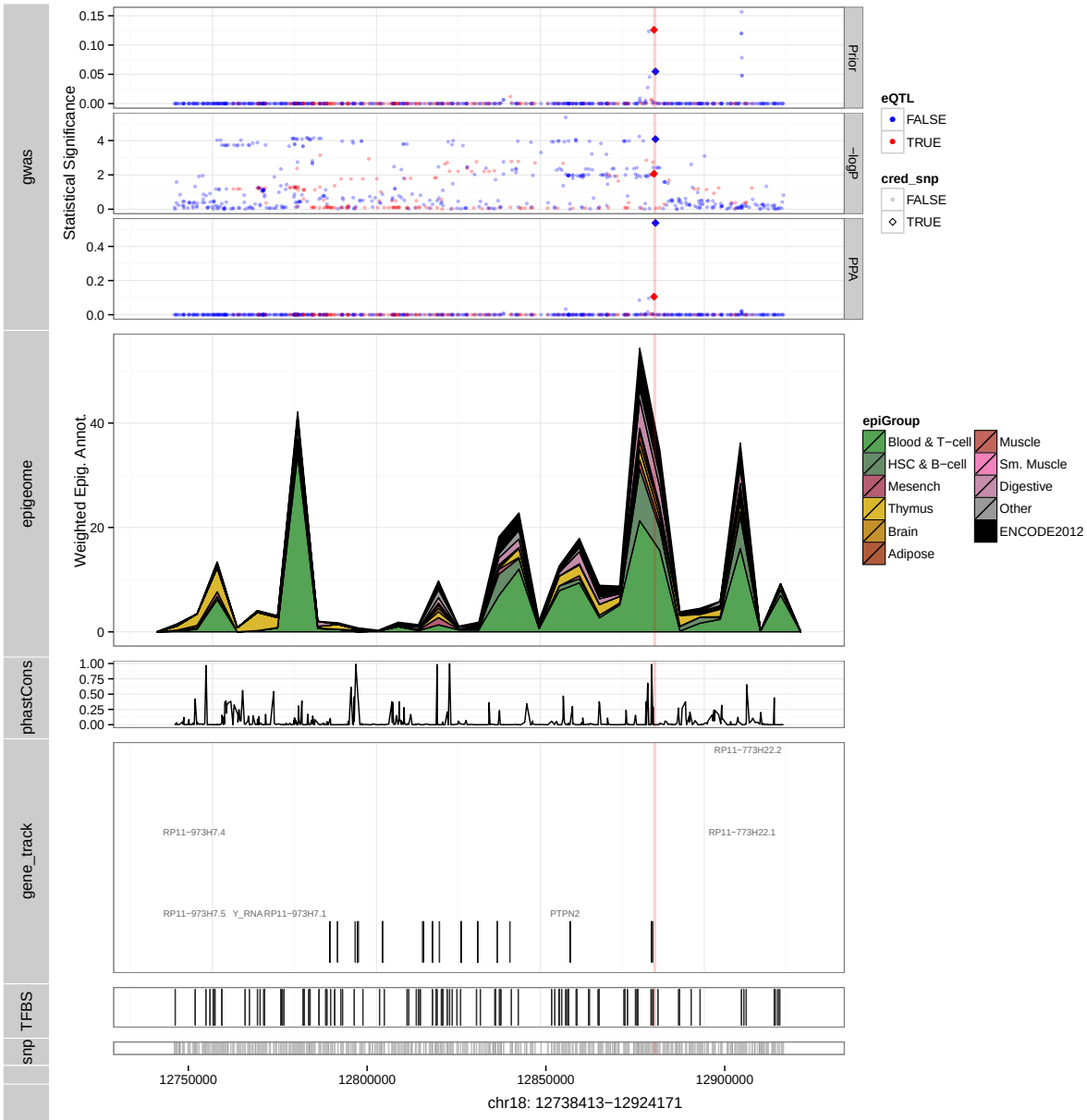
Rheumatoid Arthritis



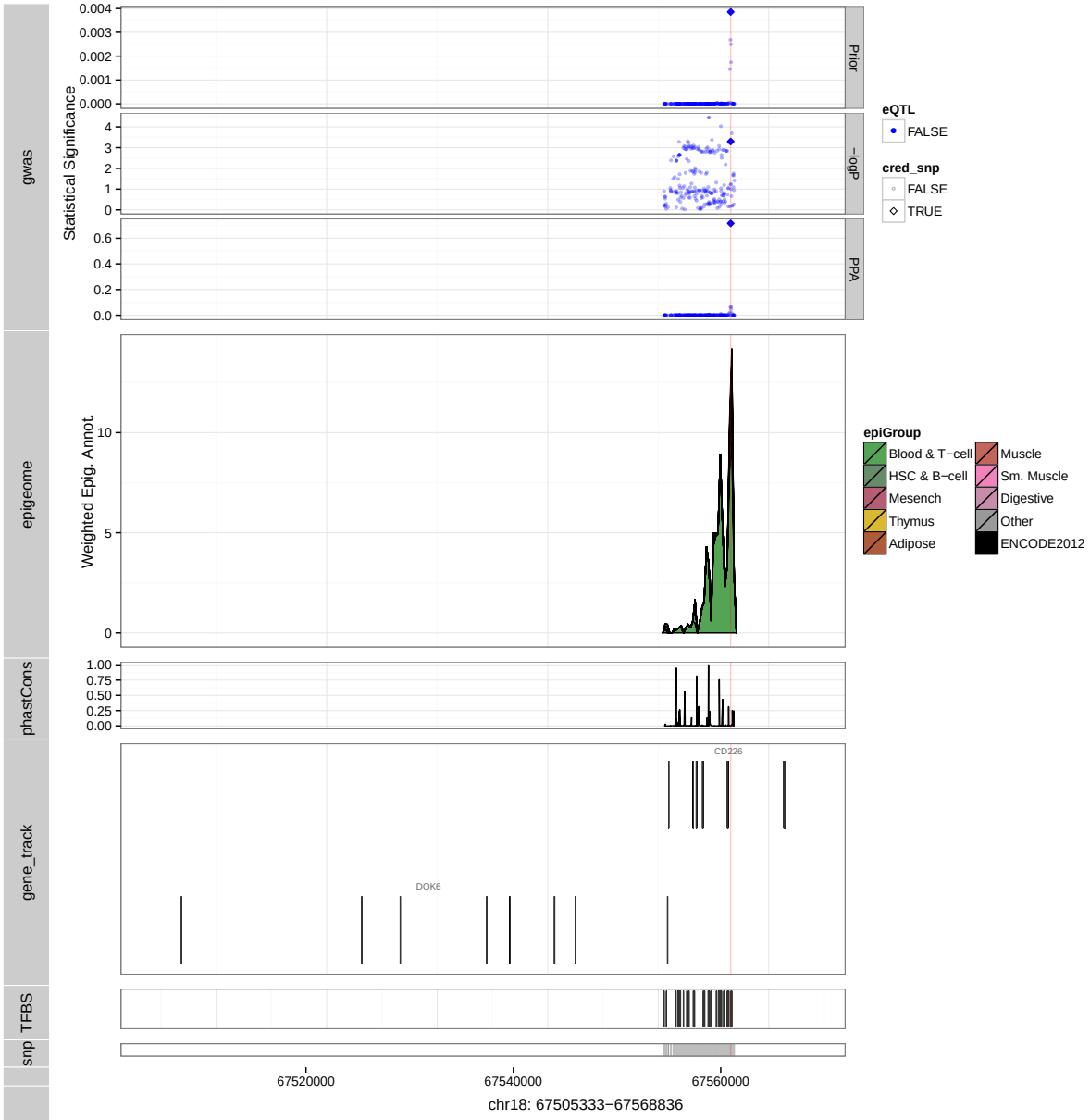
Rheumatoid Arthritis



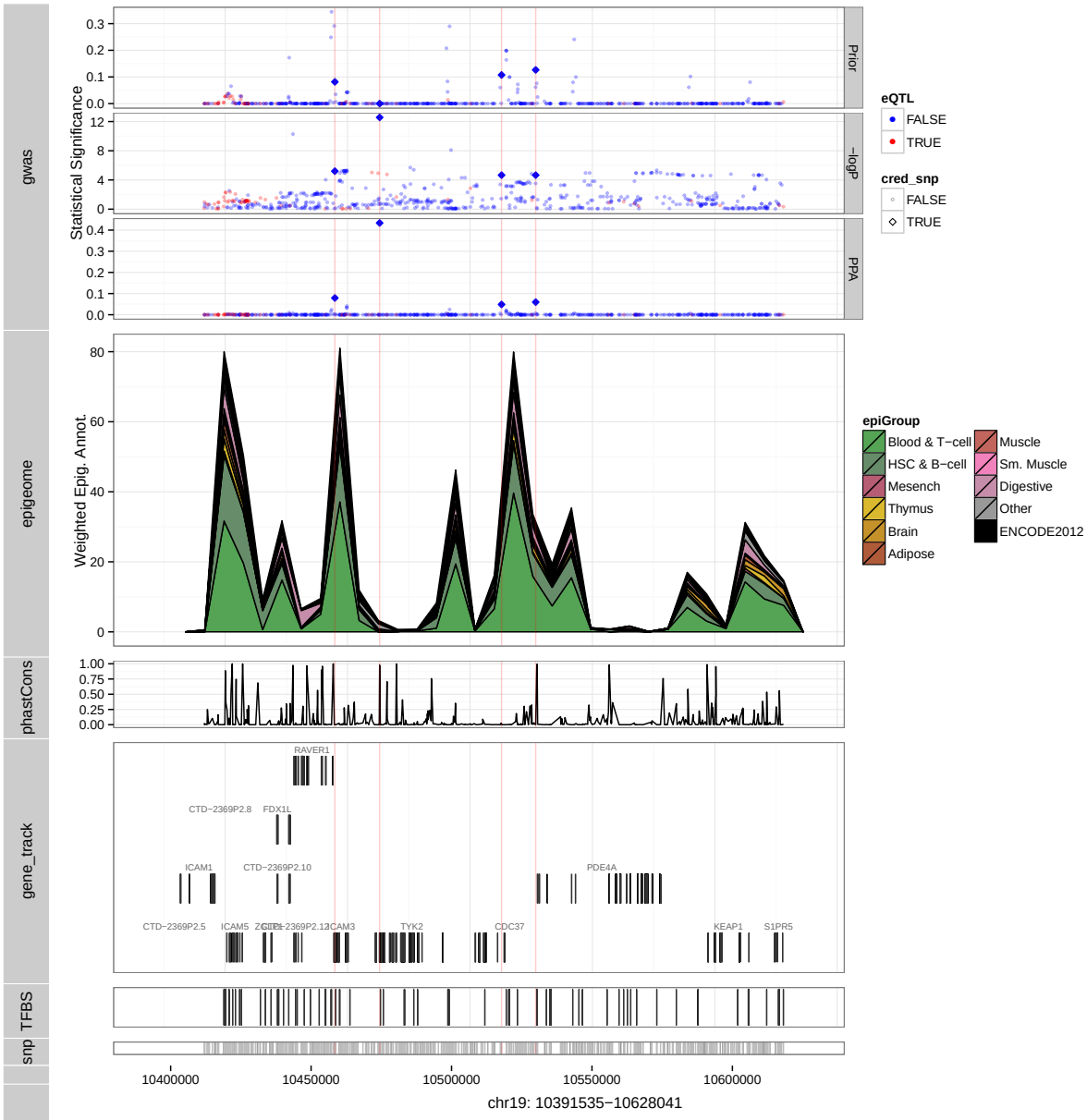
Rheumatoid Arthritis



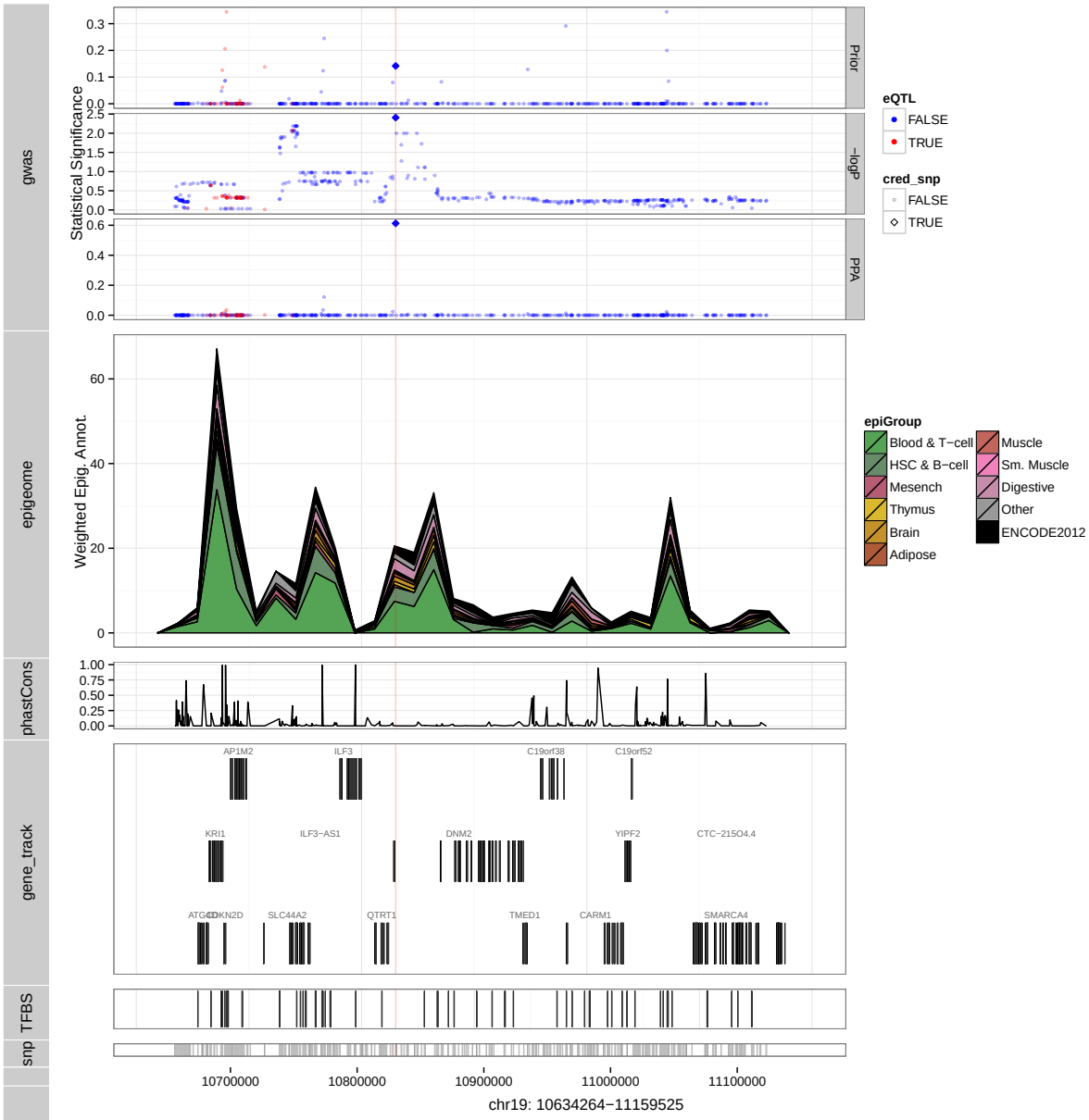
Rheumatoid Arthritis



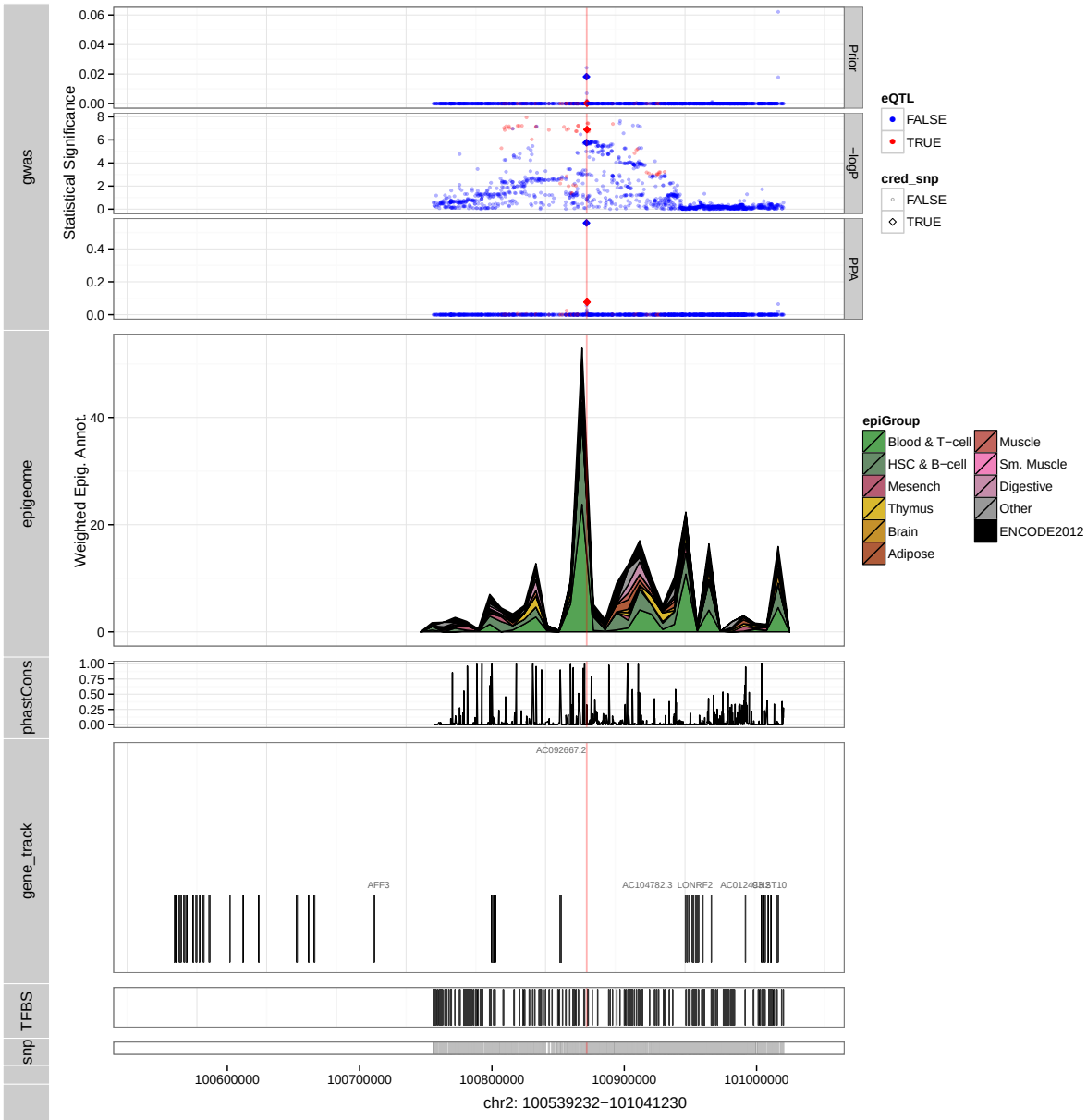
Rheumatoid Arthritis



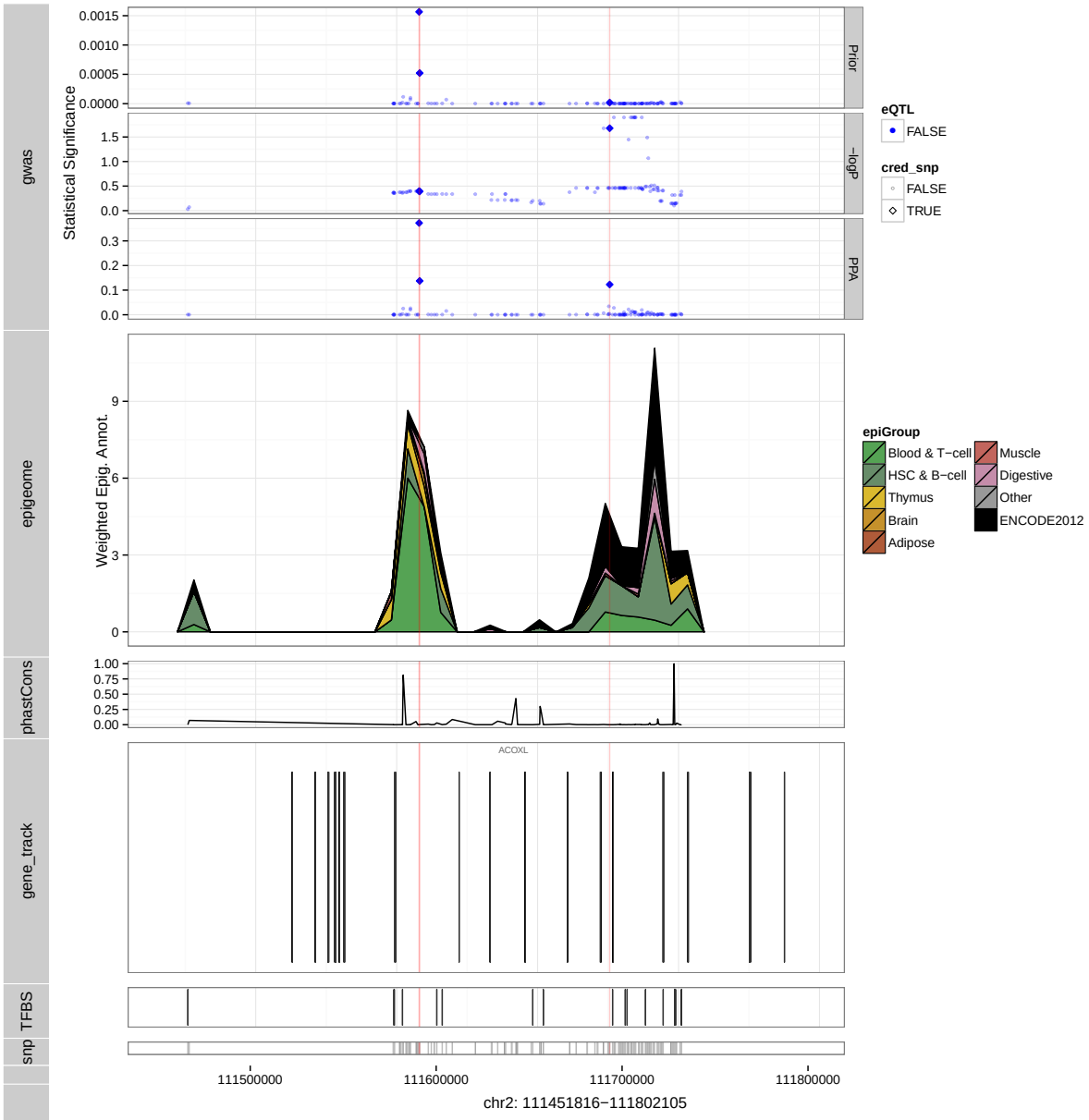
Rheumatoid Arthritis



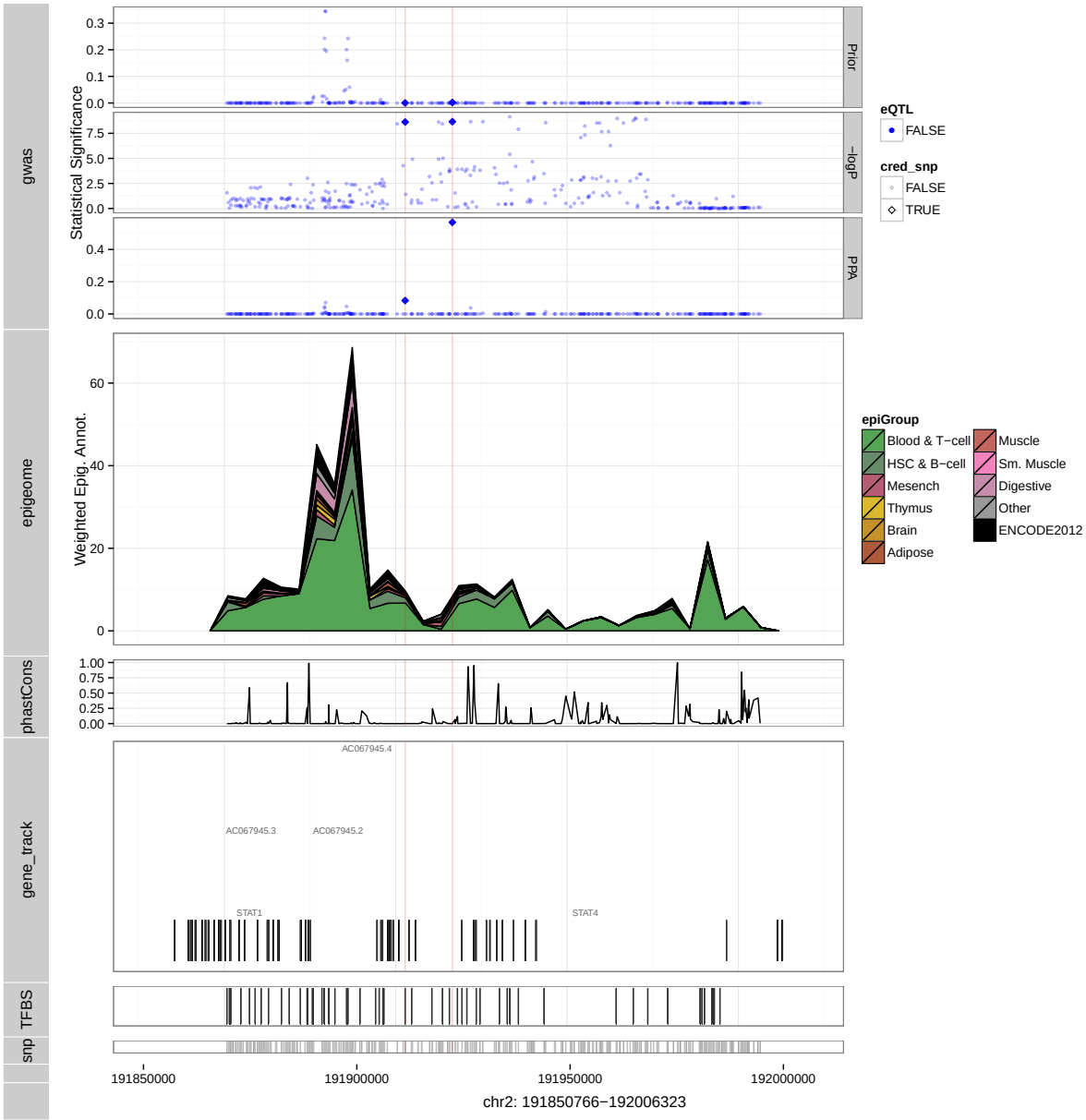
Rheumatoid Arthritis



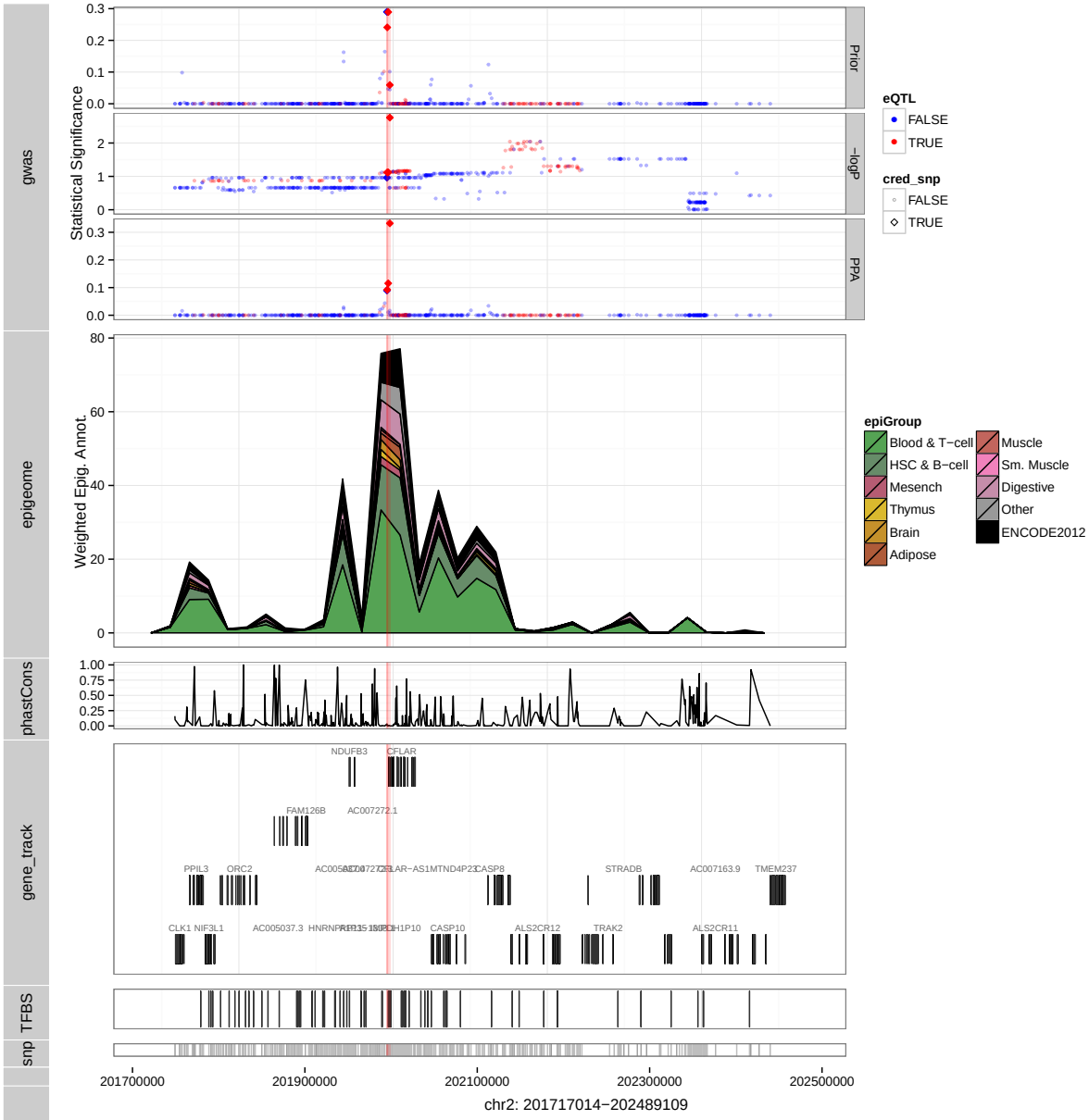
Rheumatoid Arthritis



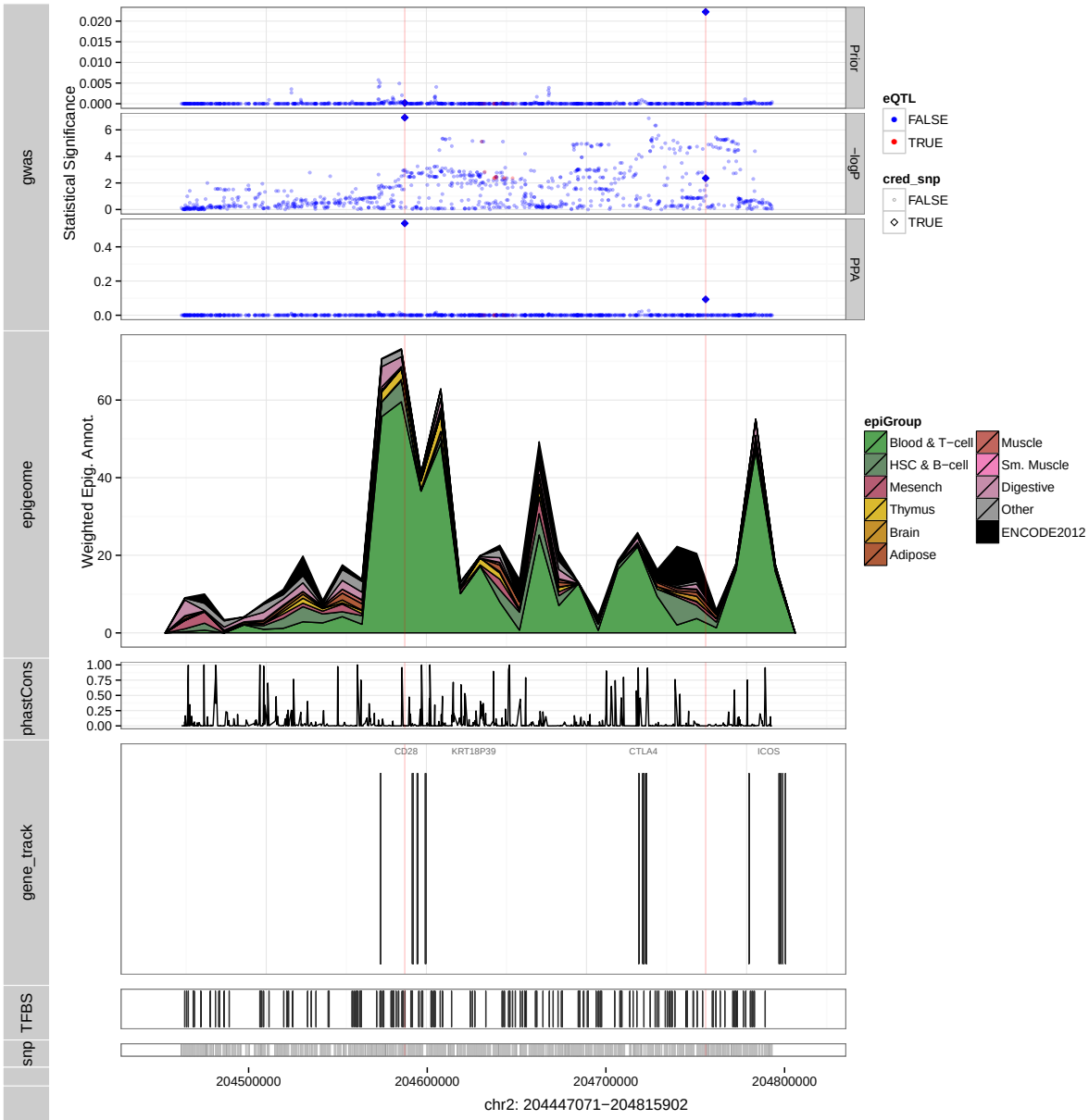
Rheumatoid Arthritis



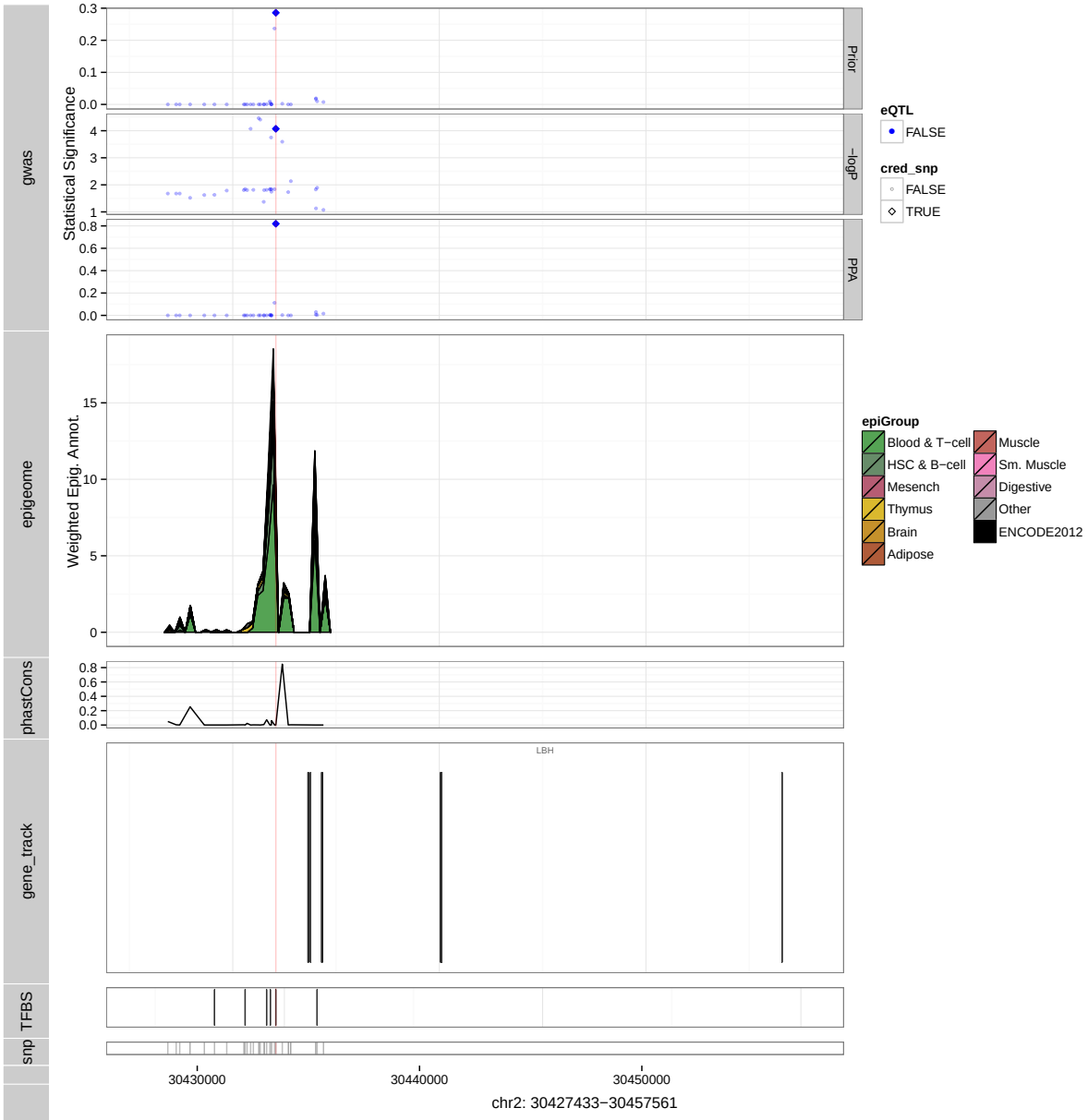
Rheumatoid Arthritis



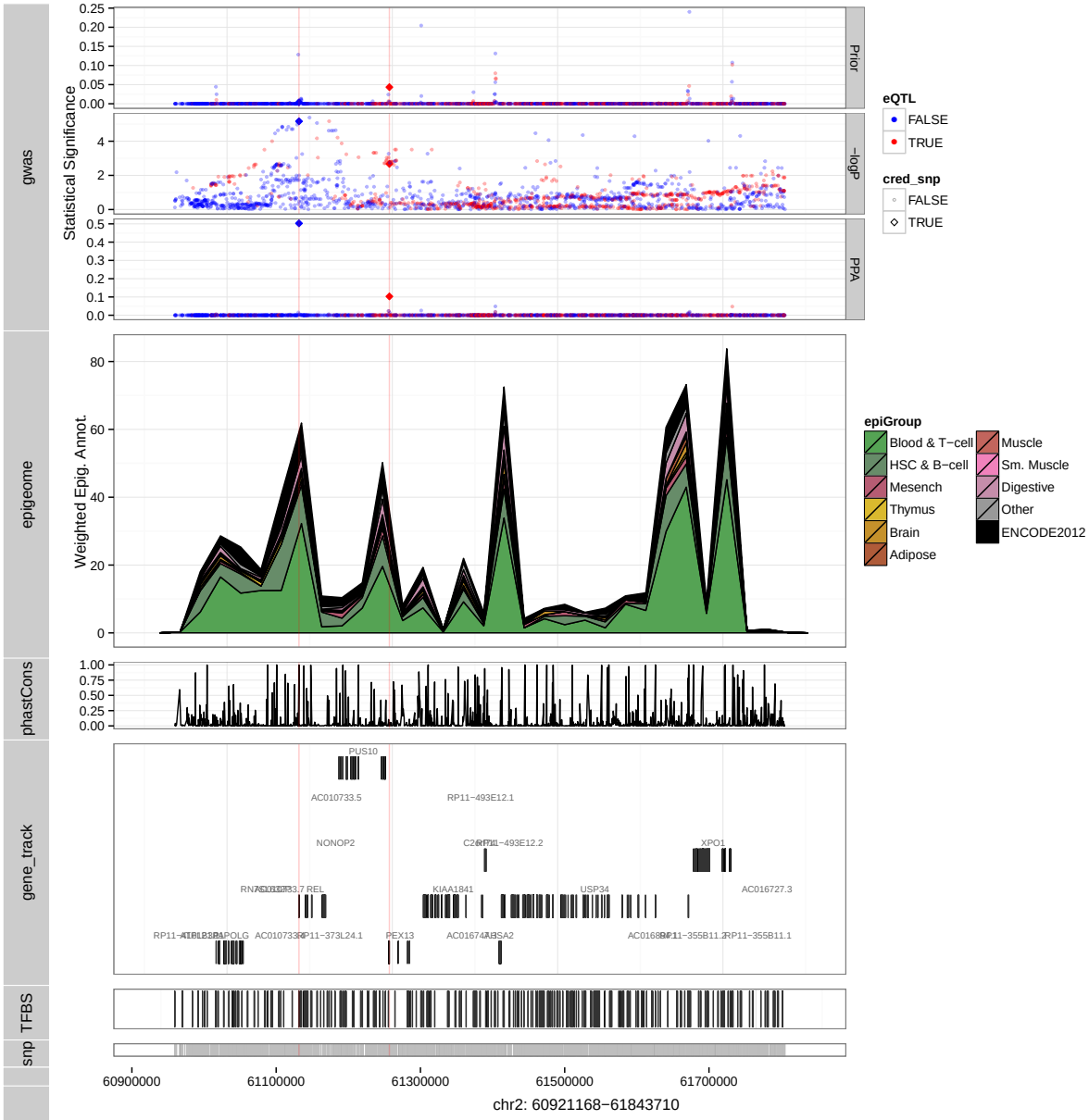
Rheumatoid Arthritis



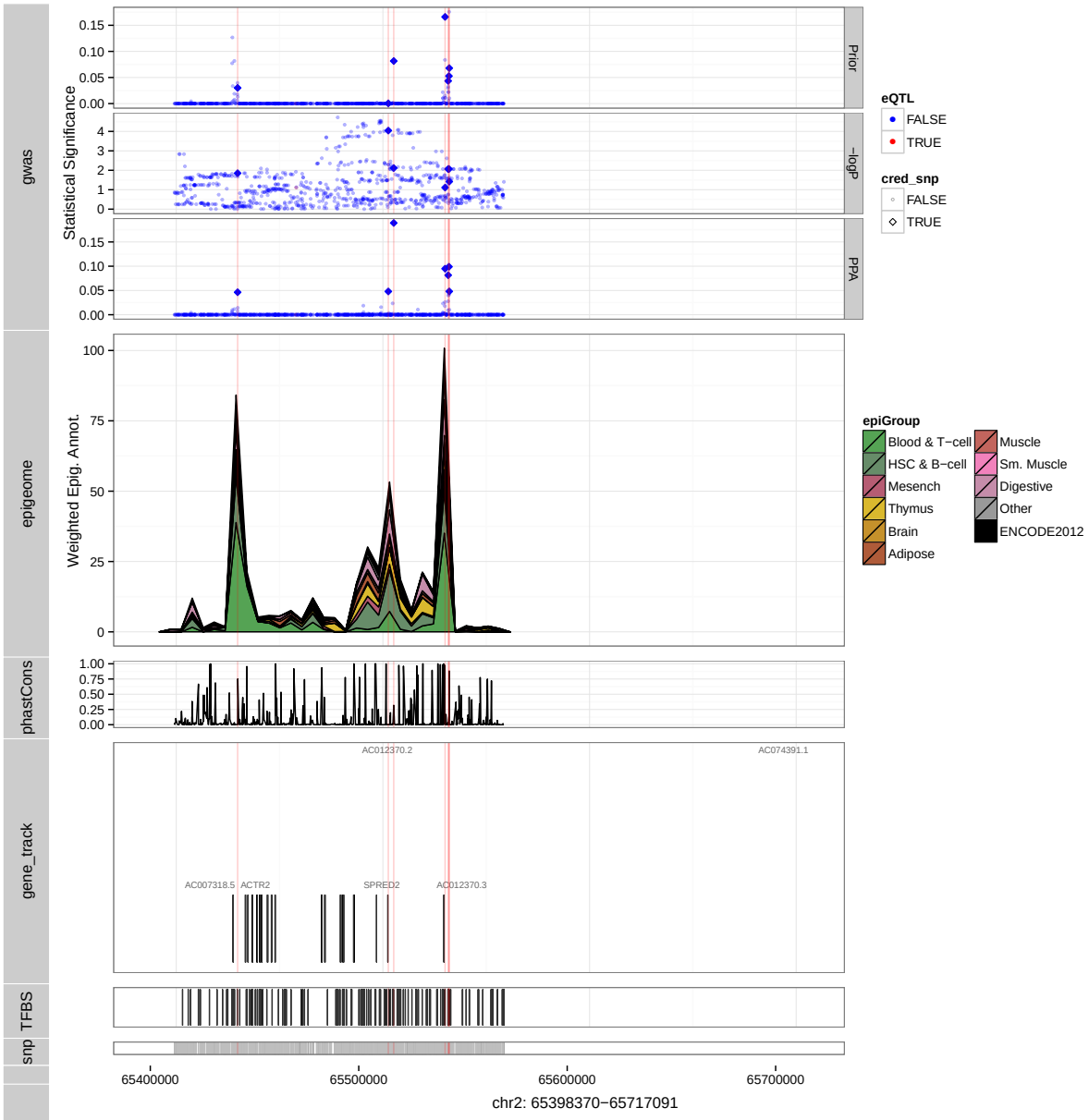
Rheumatoid Arthritis



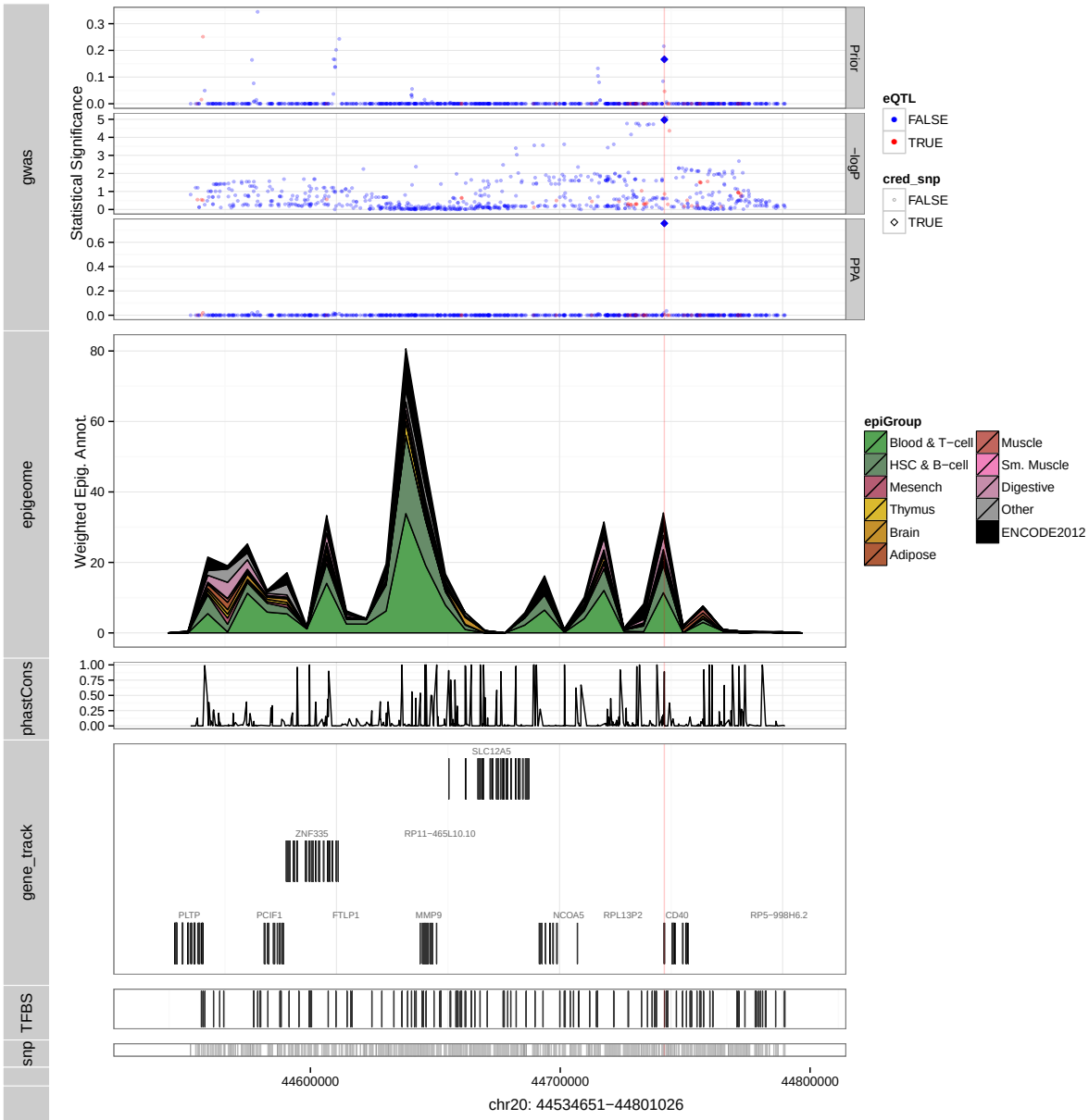
Rheumatoid Arthritis



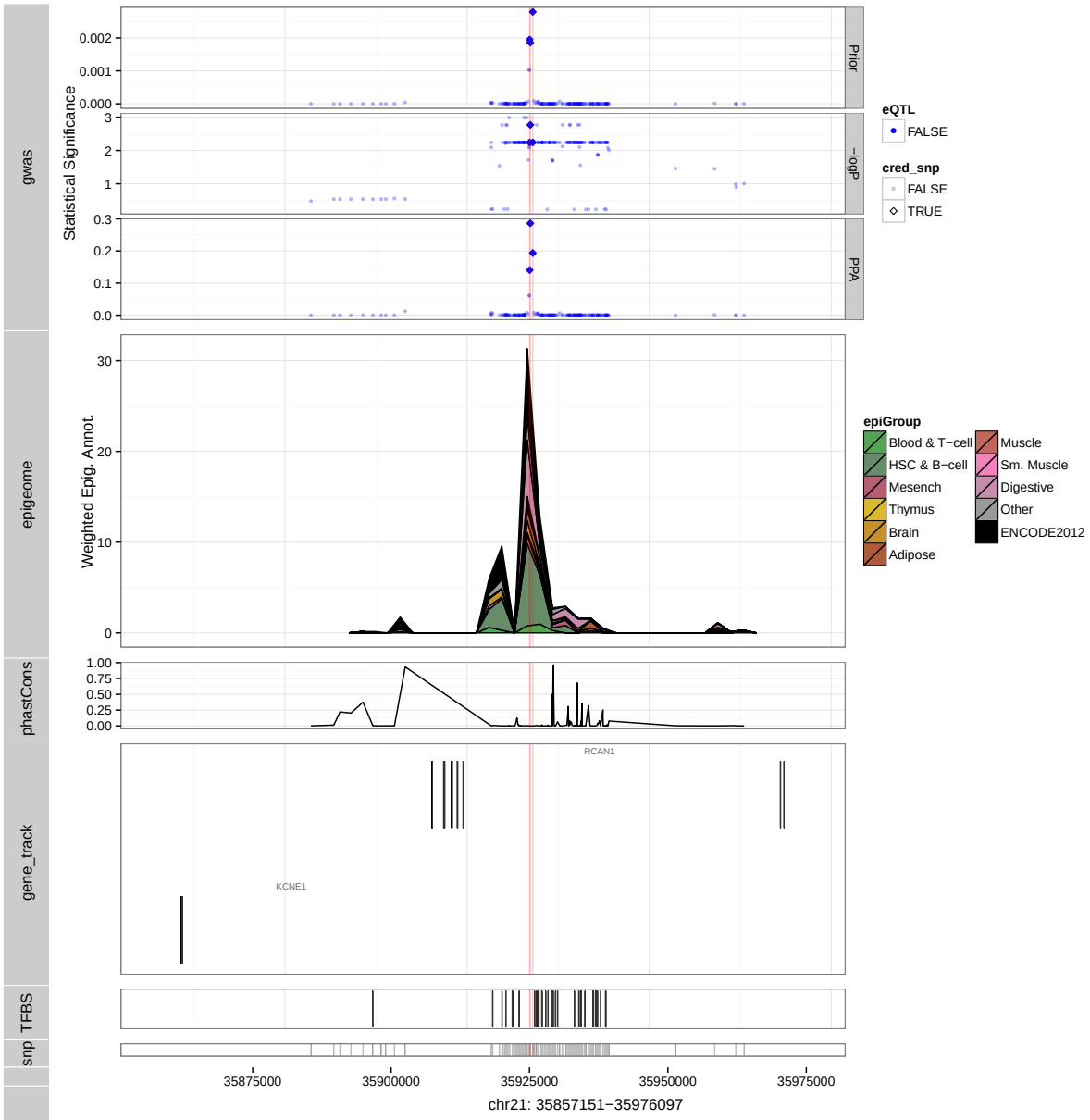
Rheumatoid Arthritis



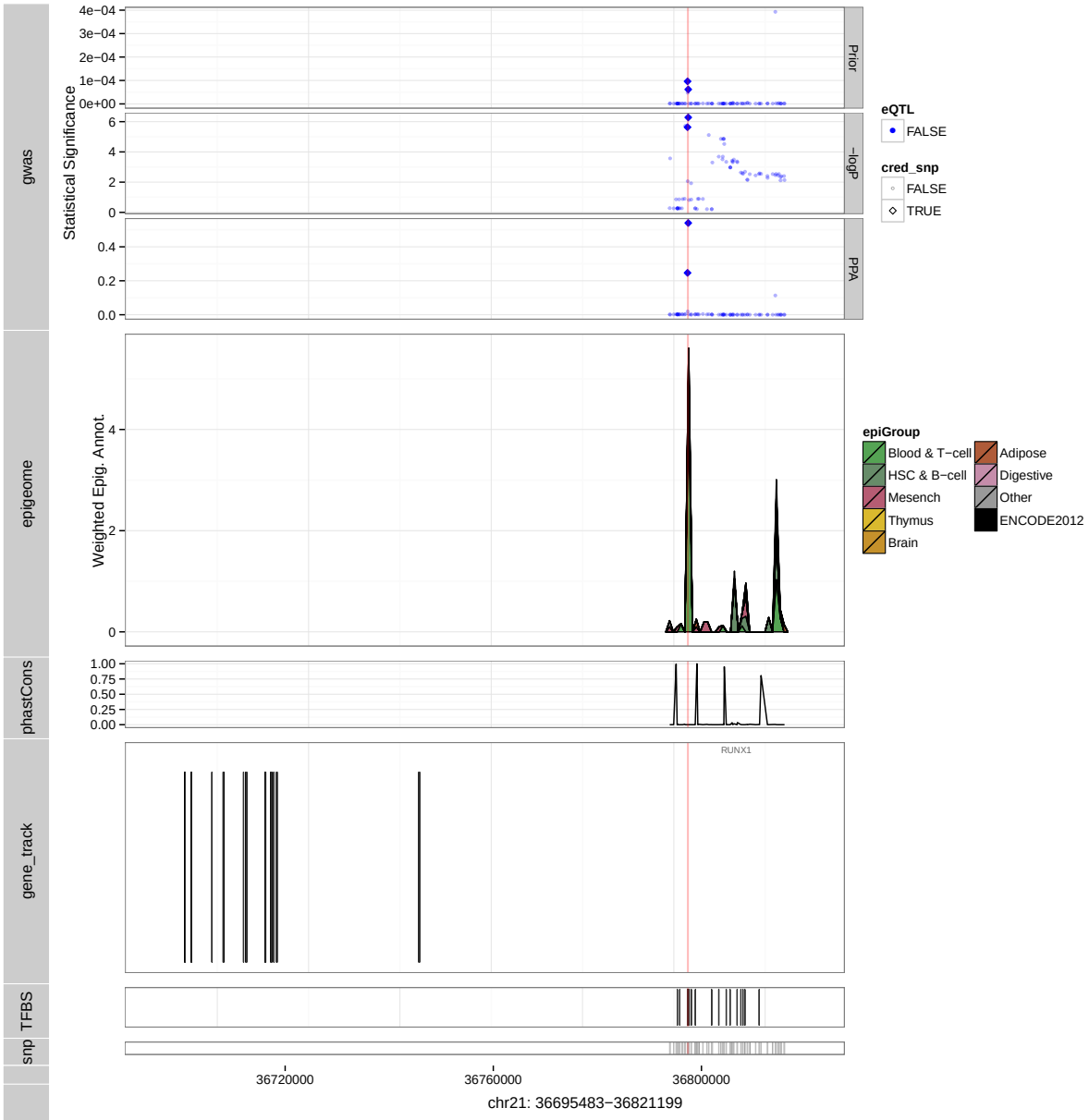
Rheumatoid Arthritis



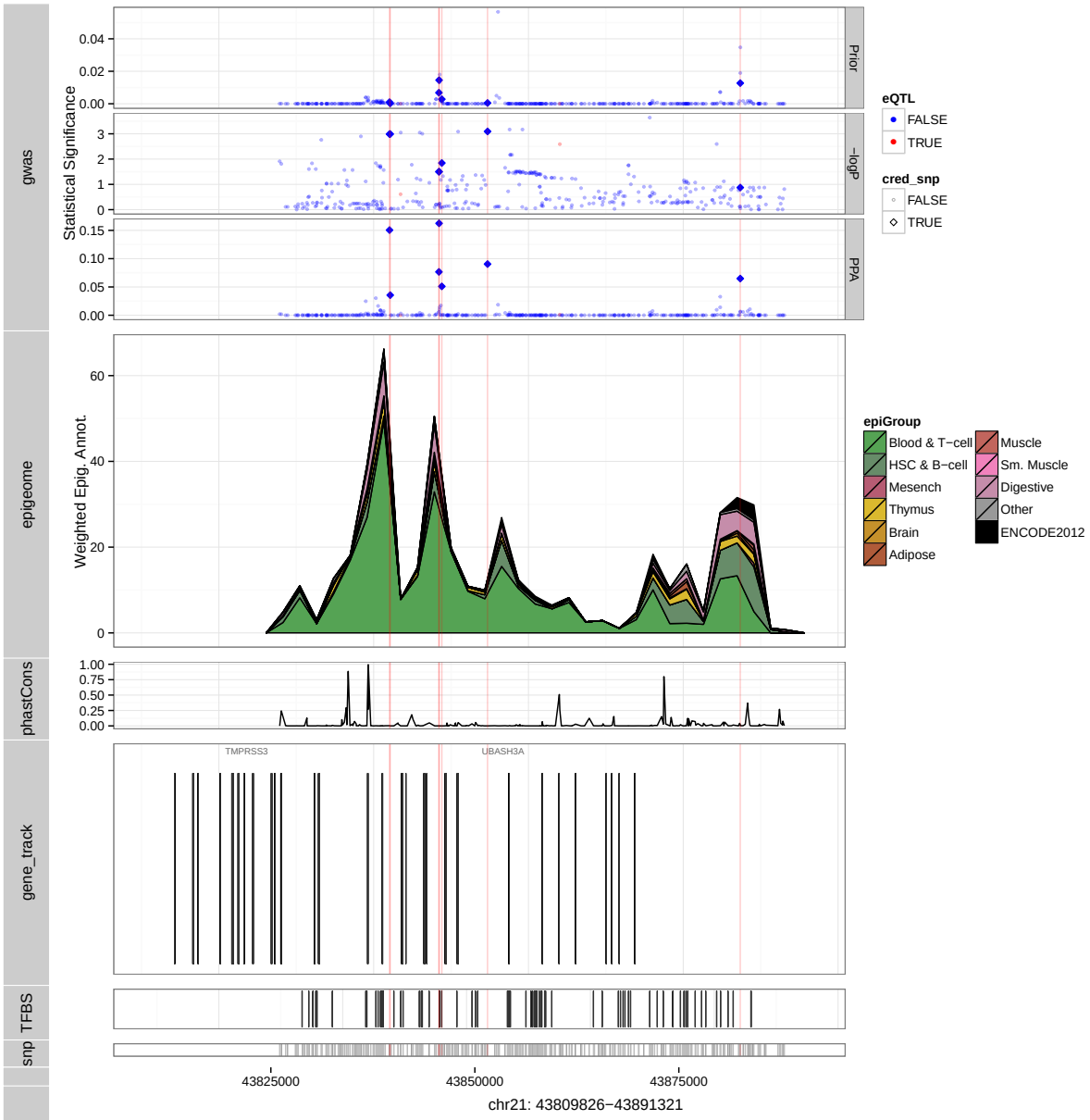
Rheumatoid Arthritis



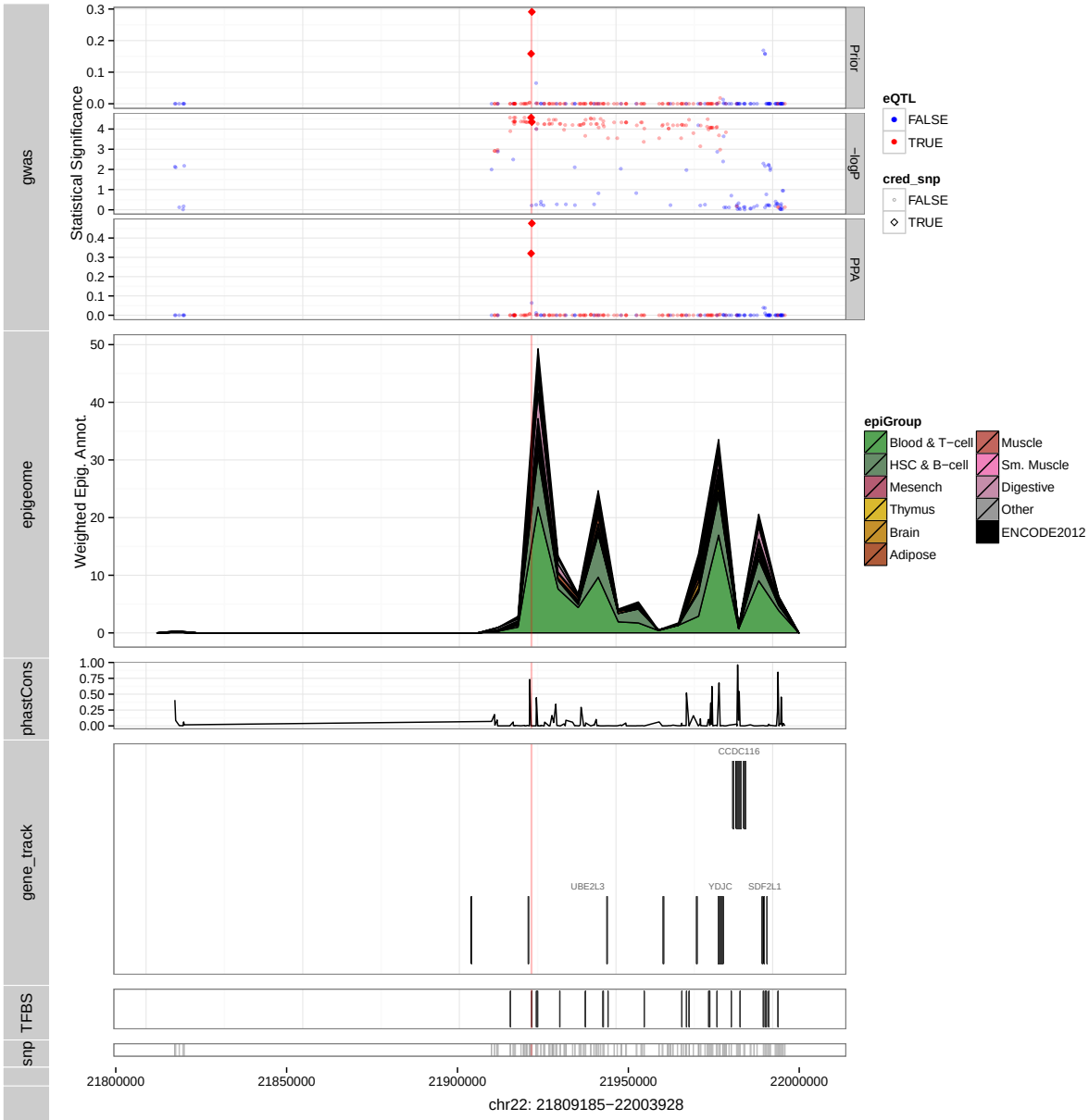
Rheumatoid Arthritis



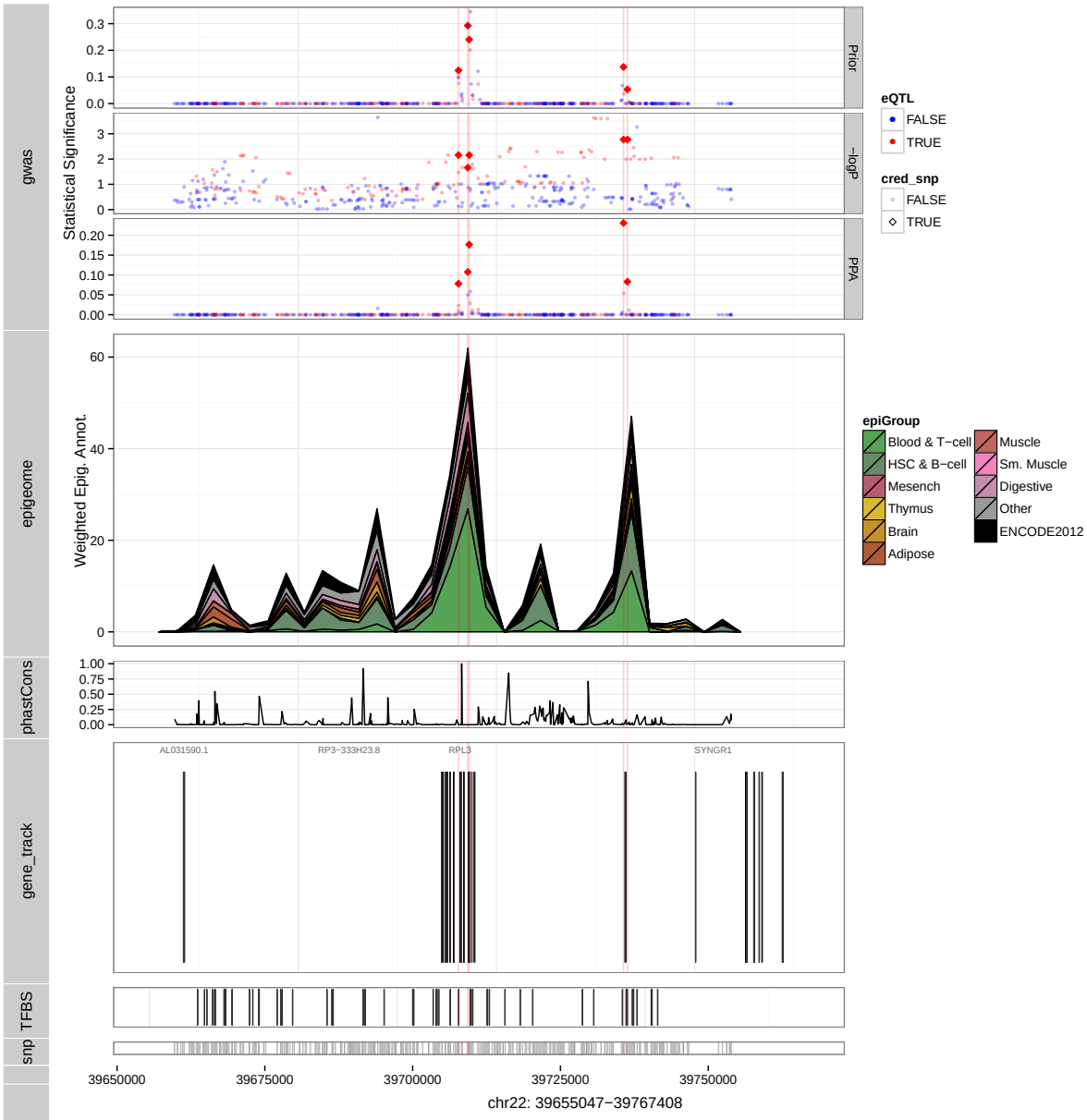
Rheumatoid Arthritis



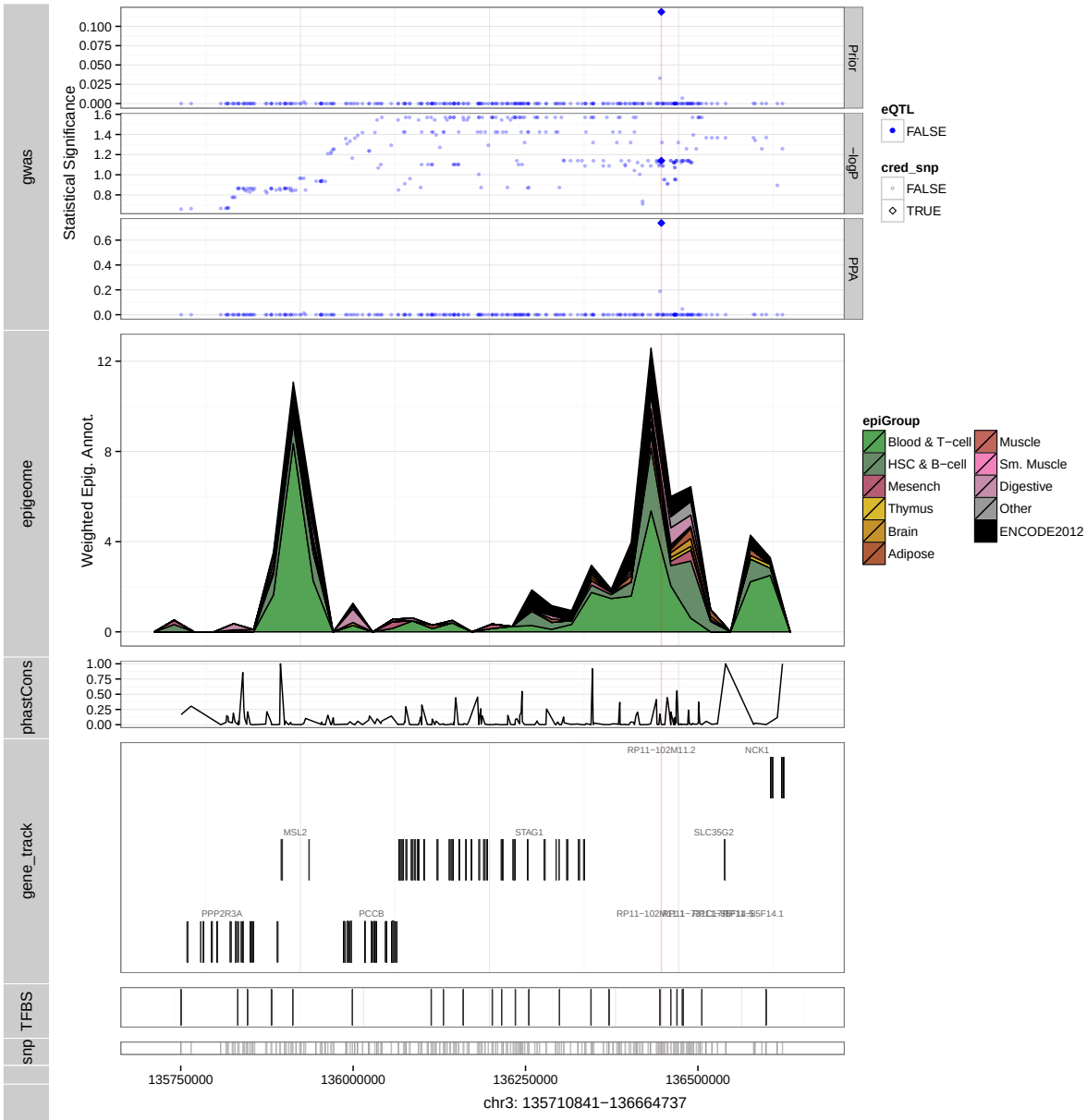
Rheumatoid Arthritis



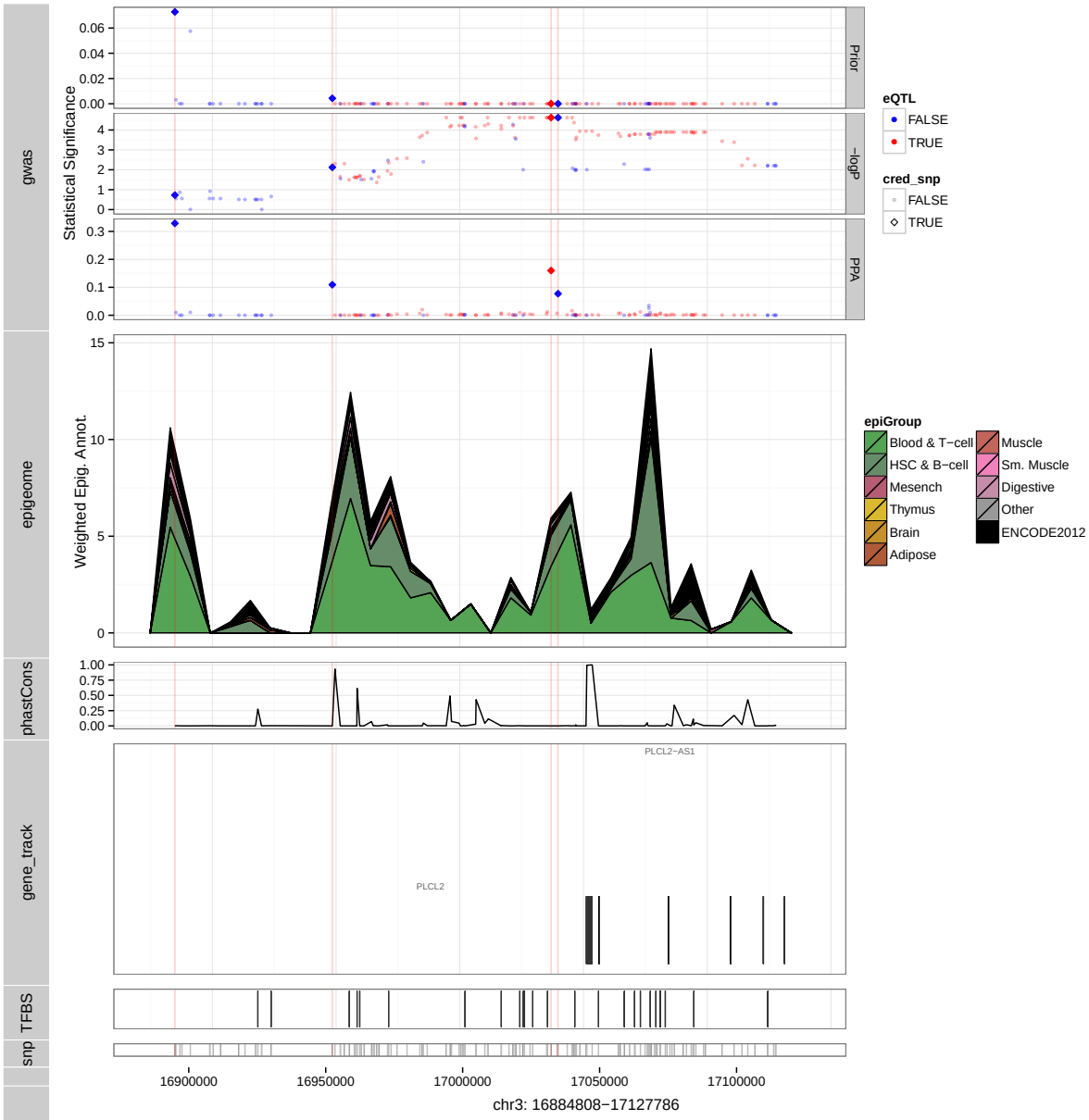
Rheumatoid Arthritis



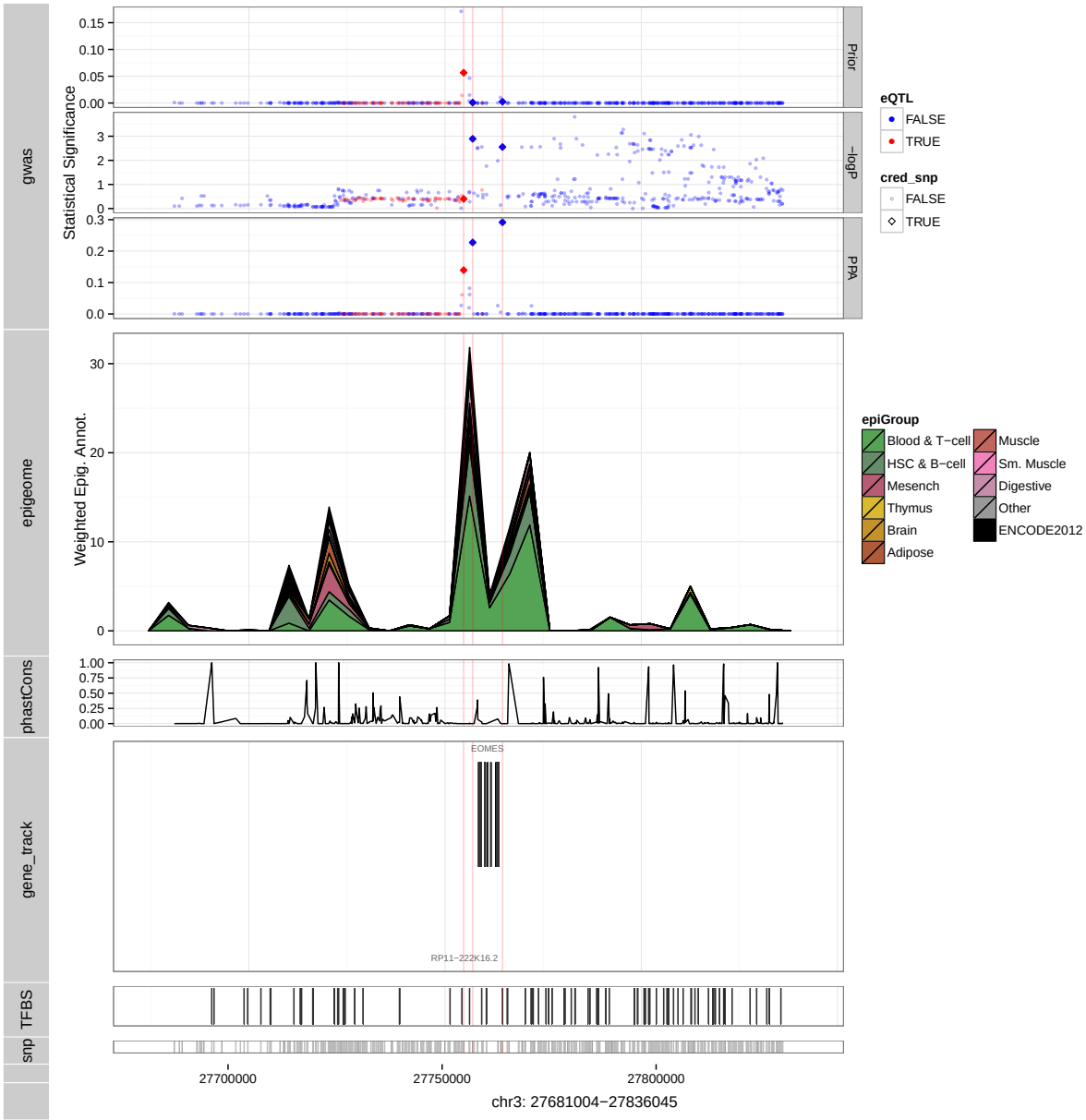
Rheumatoid Arthritis



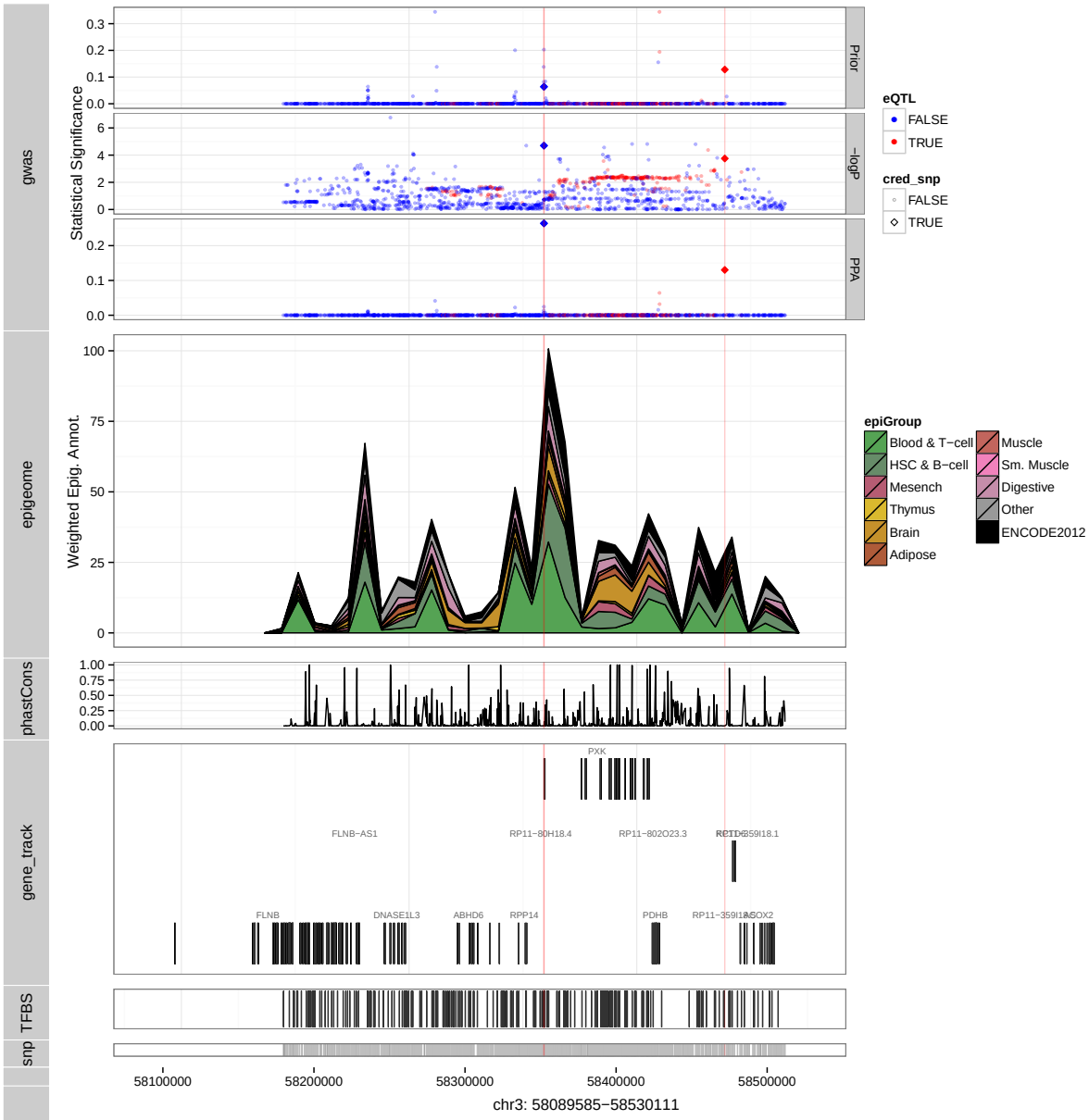
Rheumatoid Arthritis



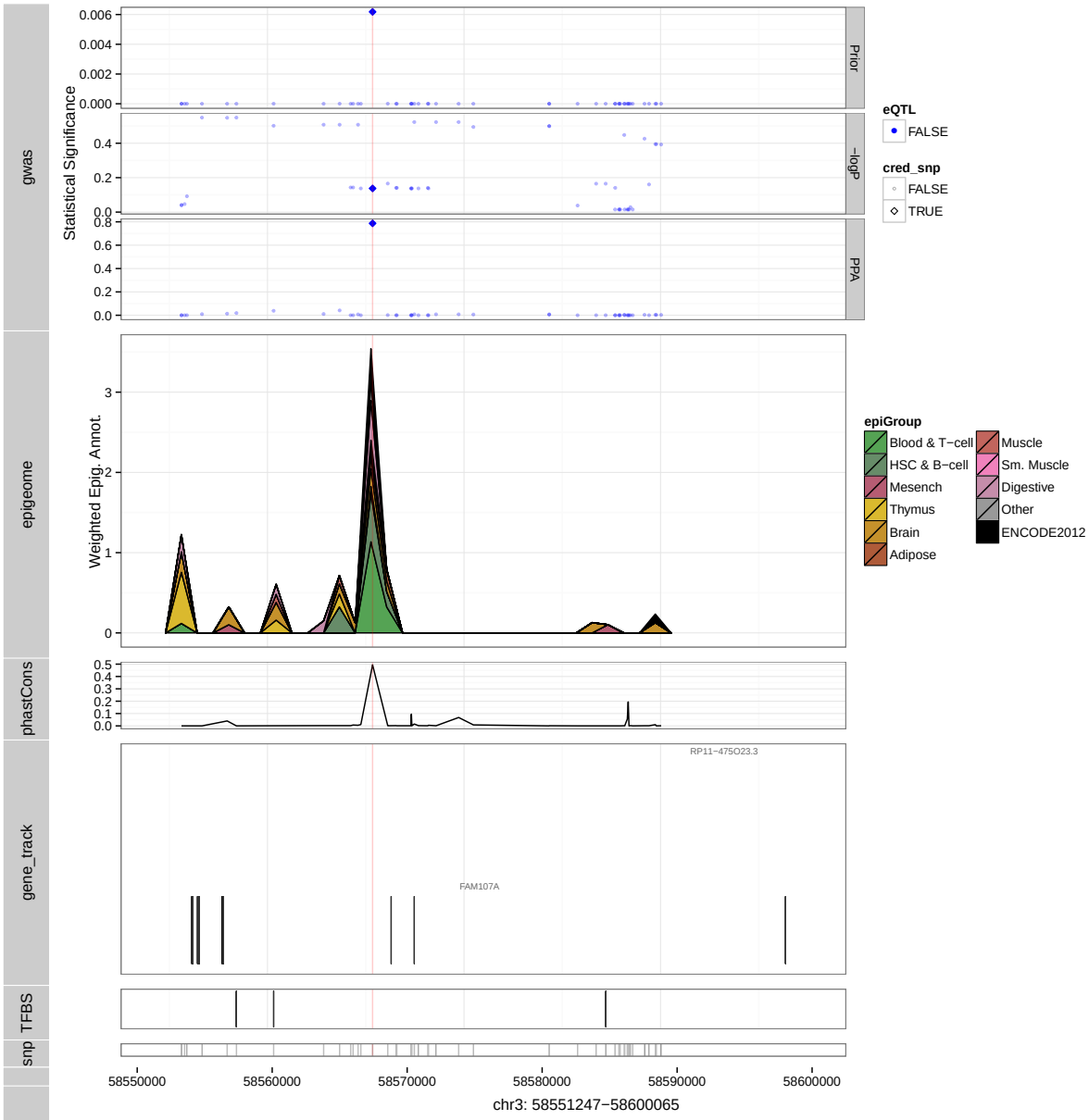
Rheumatoid Arthritis



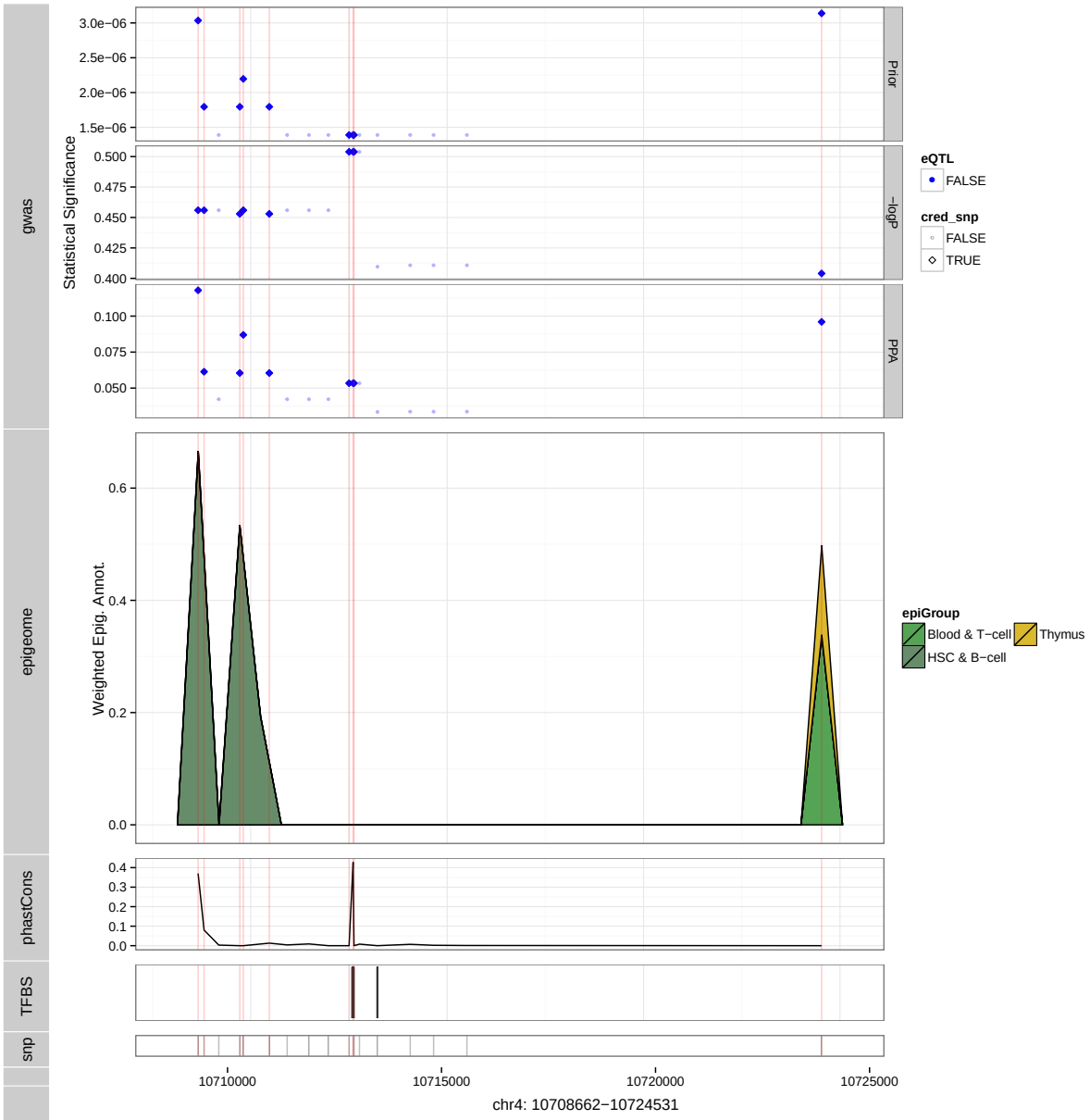
Rheumatoid Arthritis



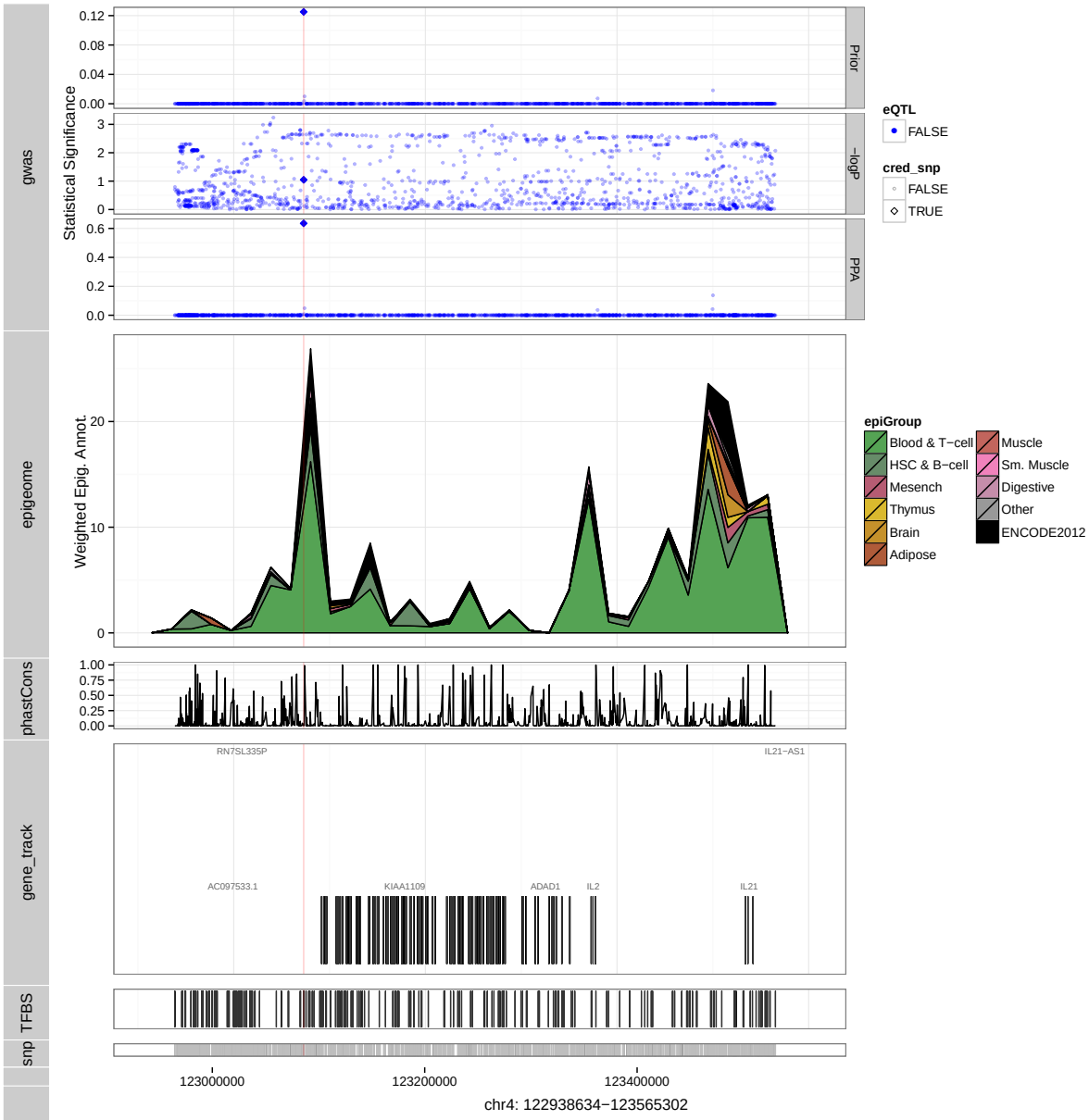
Rheumatoid Arthritis



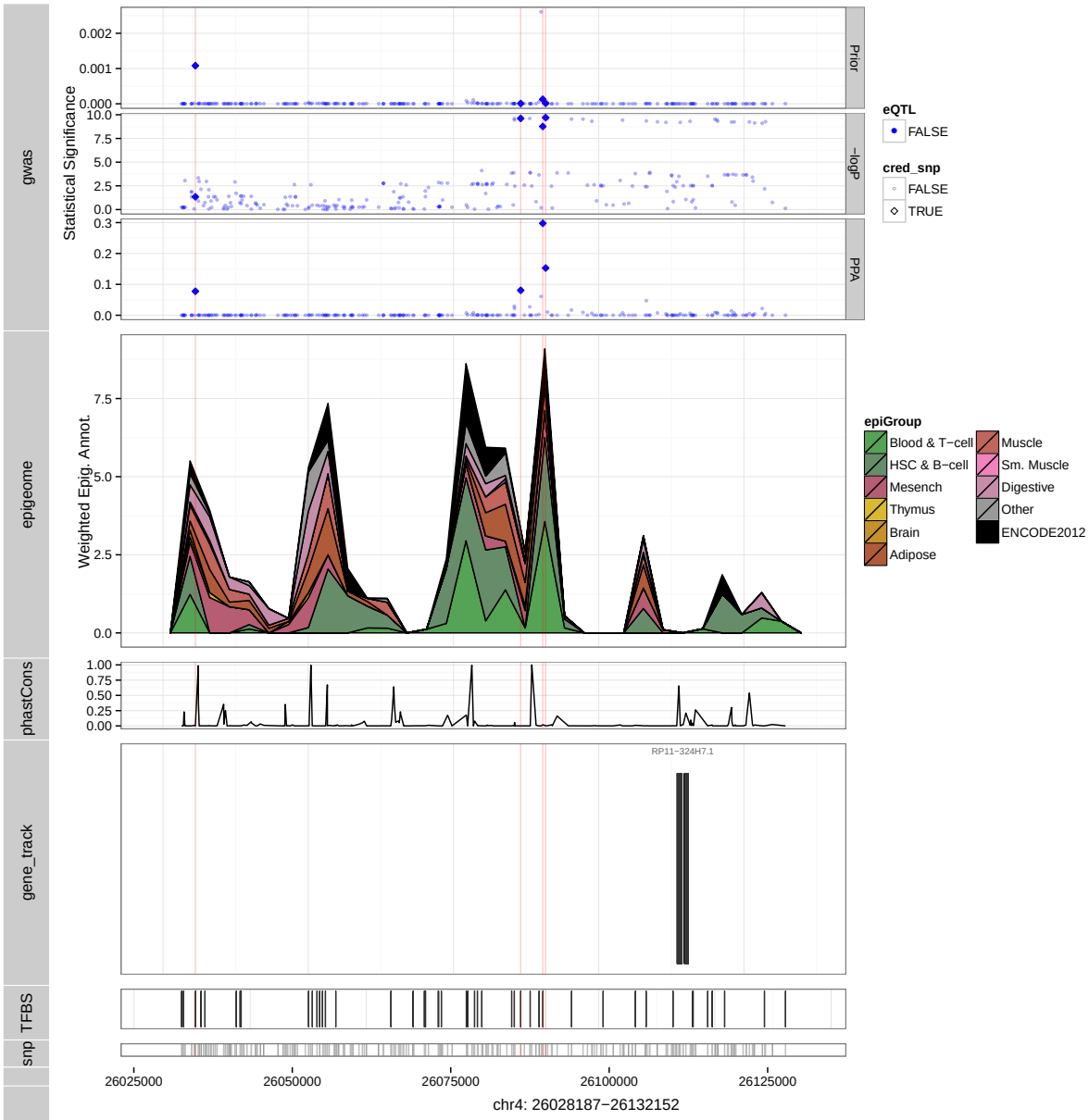
Rheumatoid Arthritis



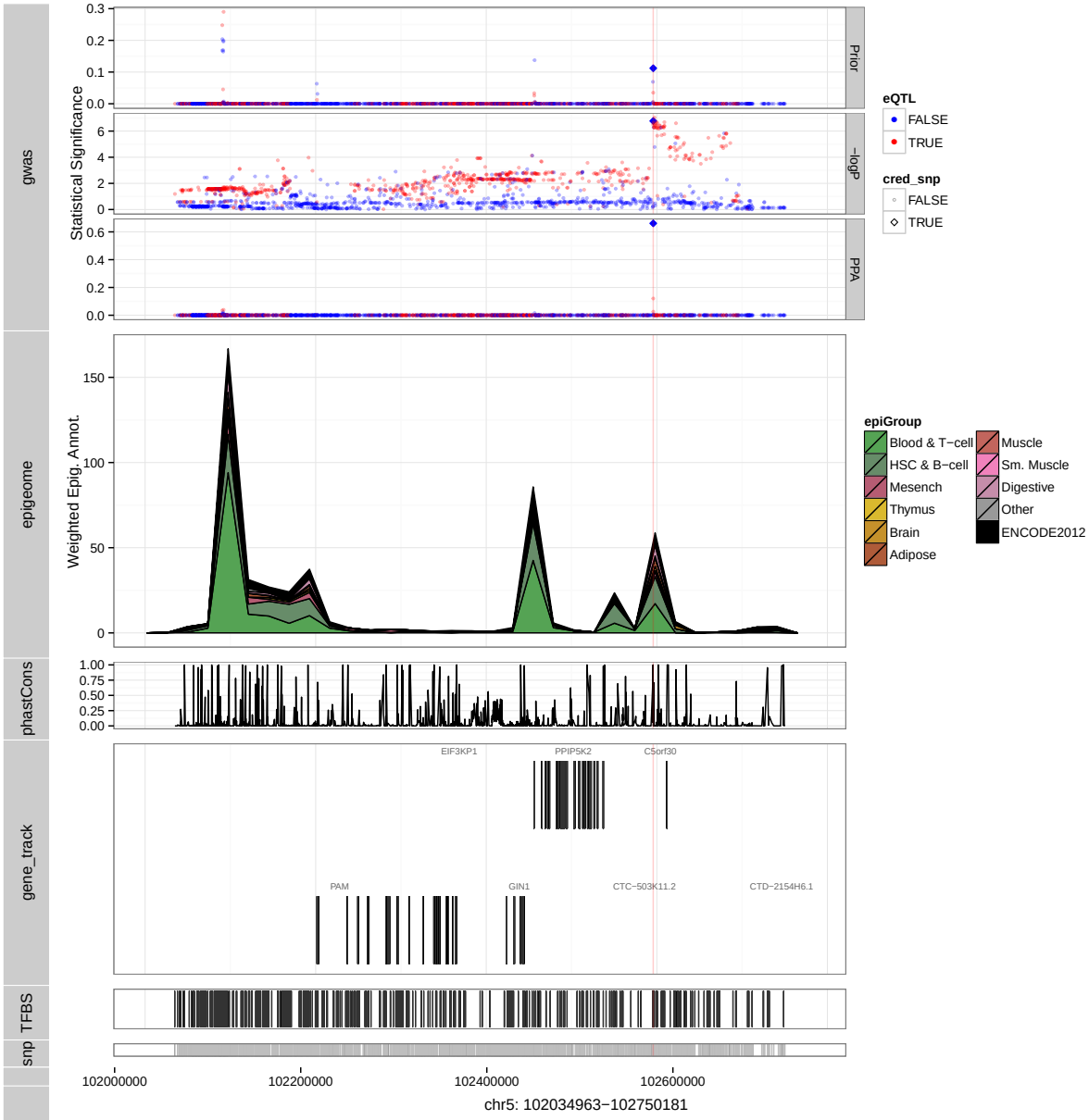
Rheumatoid Arthritis



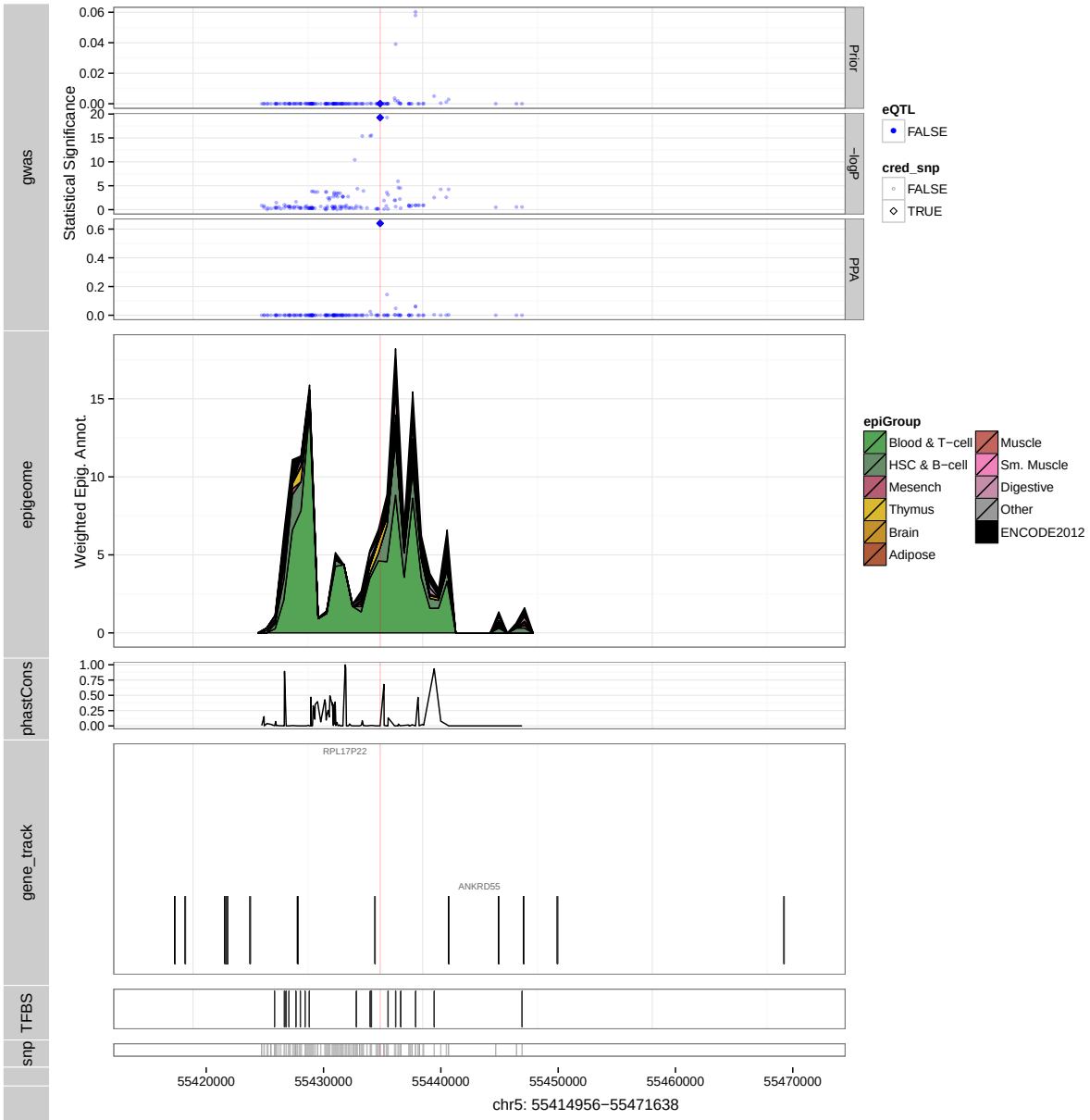
Rheumatoid Arthritis



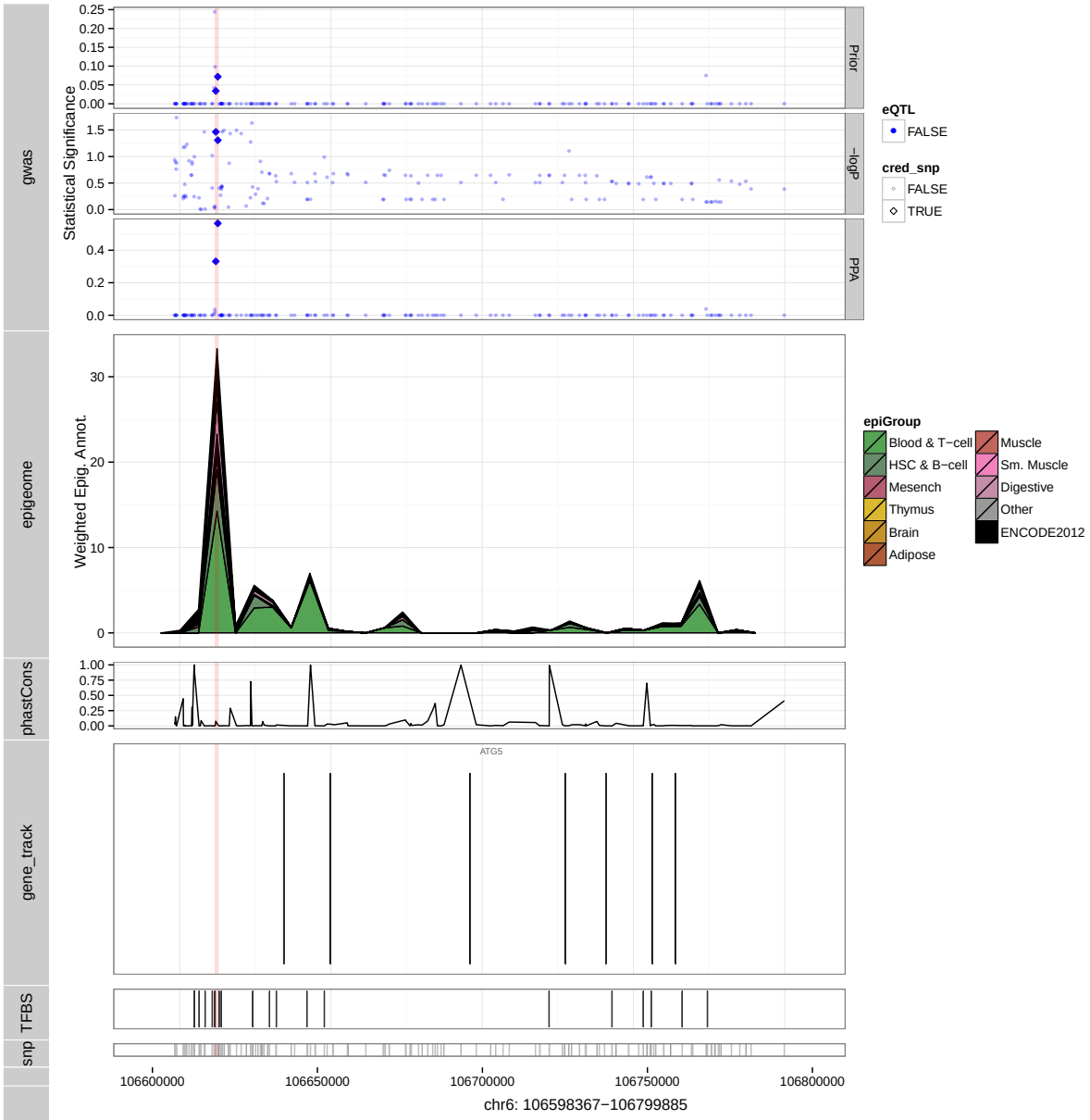
Rheumatoid Arthritis



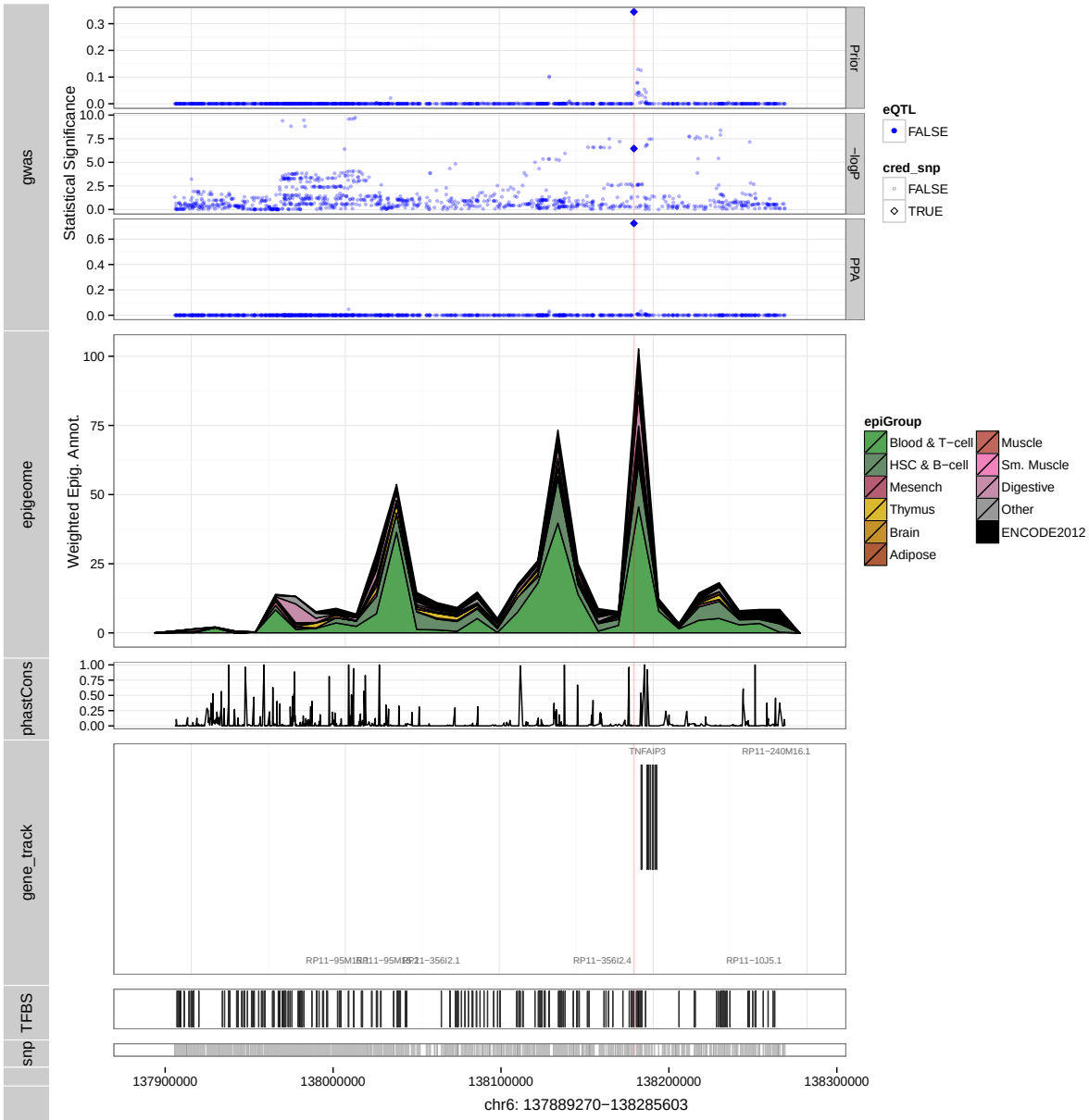
Rheumatoid Arthritis



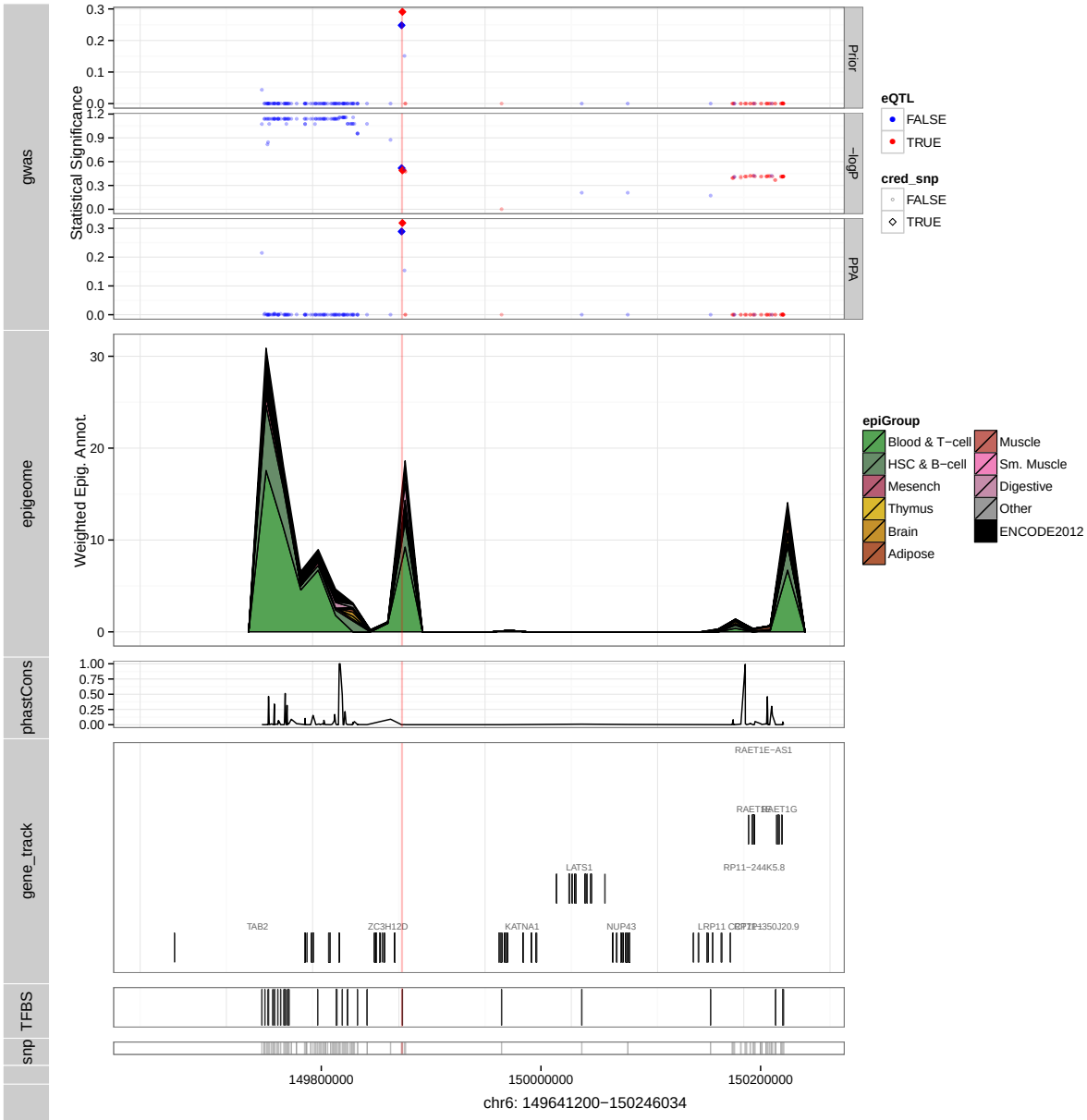
Rheumatoid Arthritis



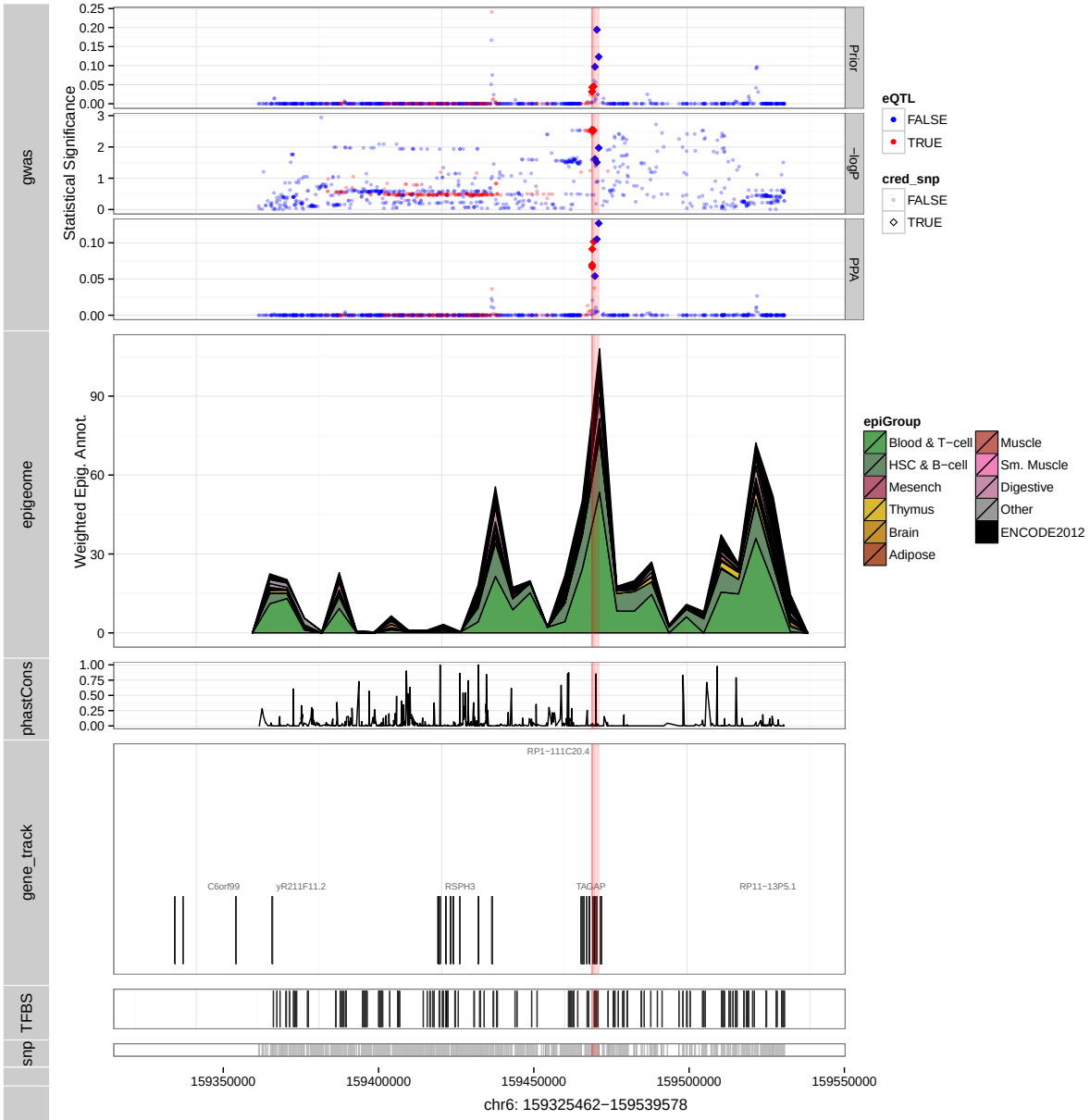
Rheumatoid Arthritis



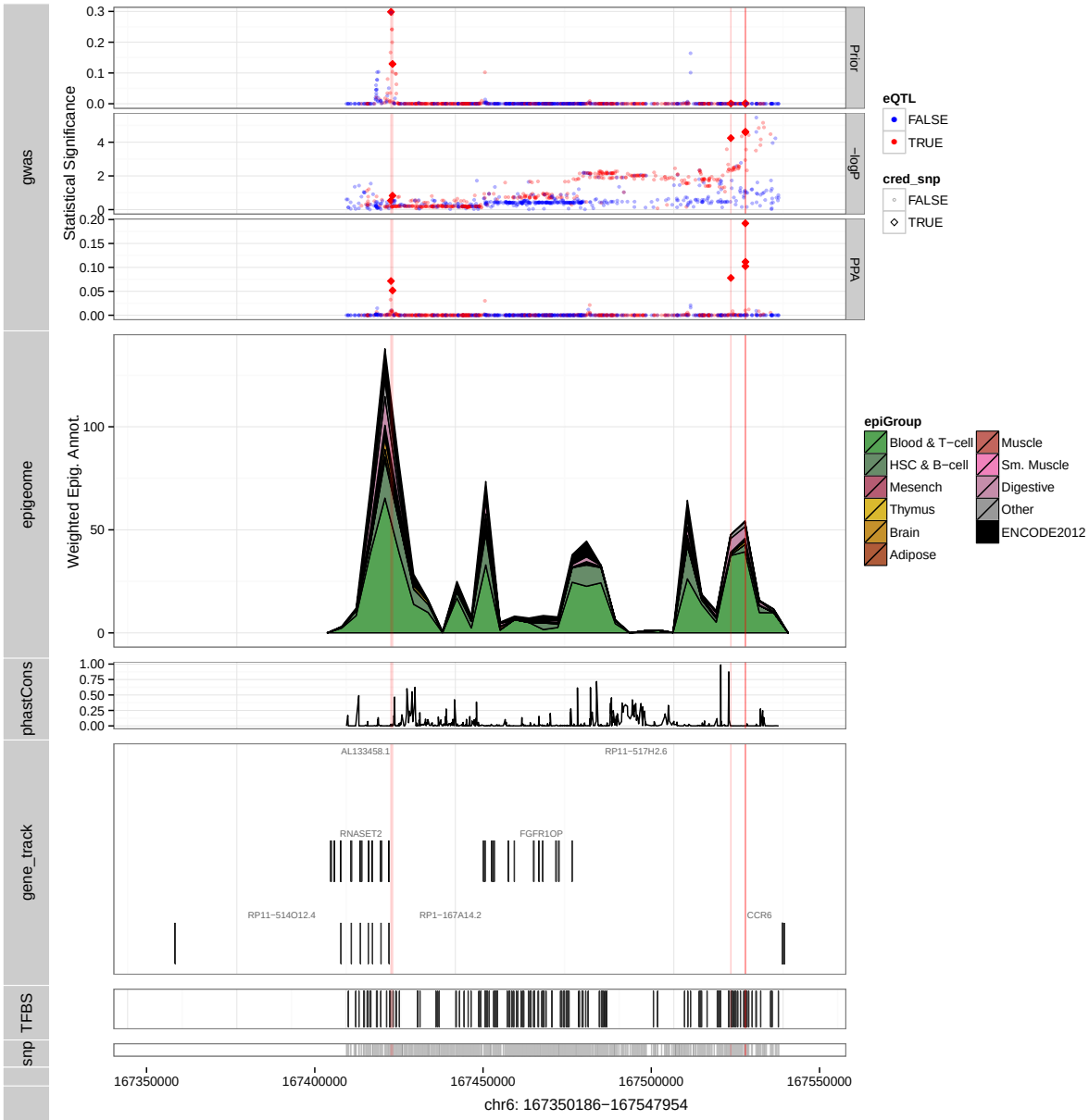
Rheumatoid Arthritis



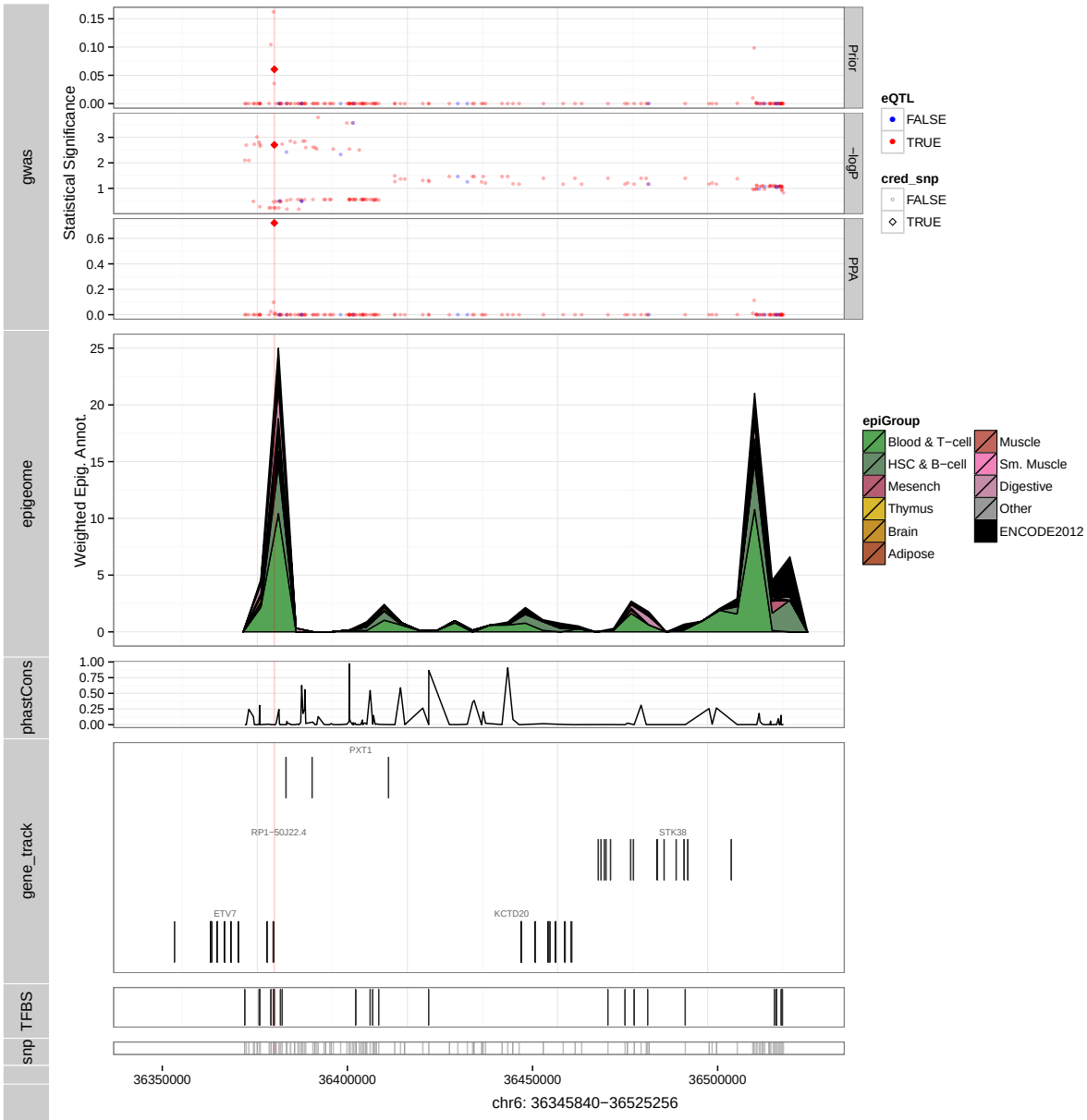
Rheumatoid Arthritis



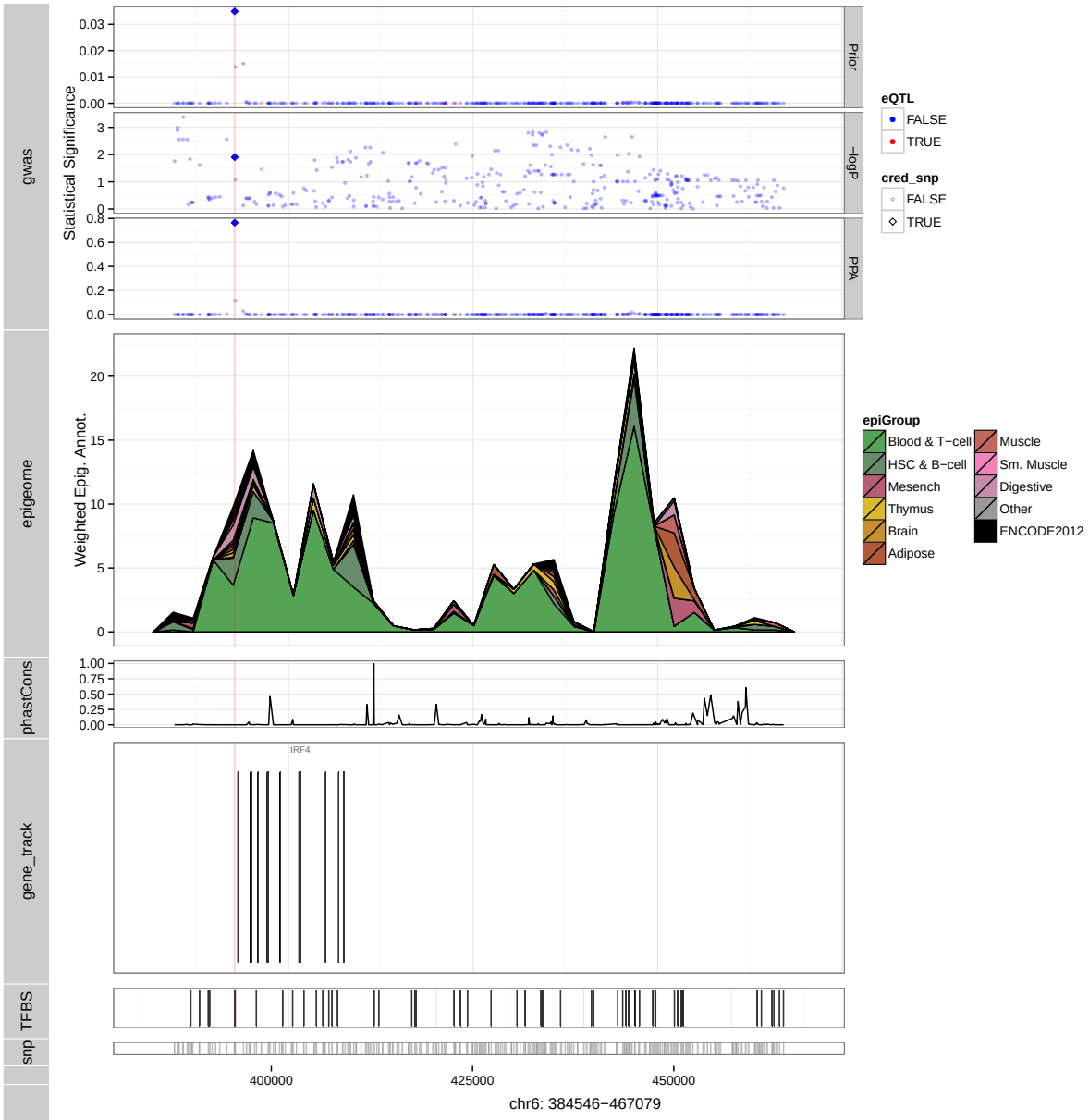
Rheumatoid Arthritis



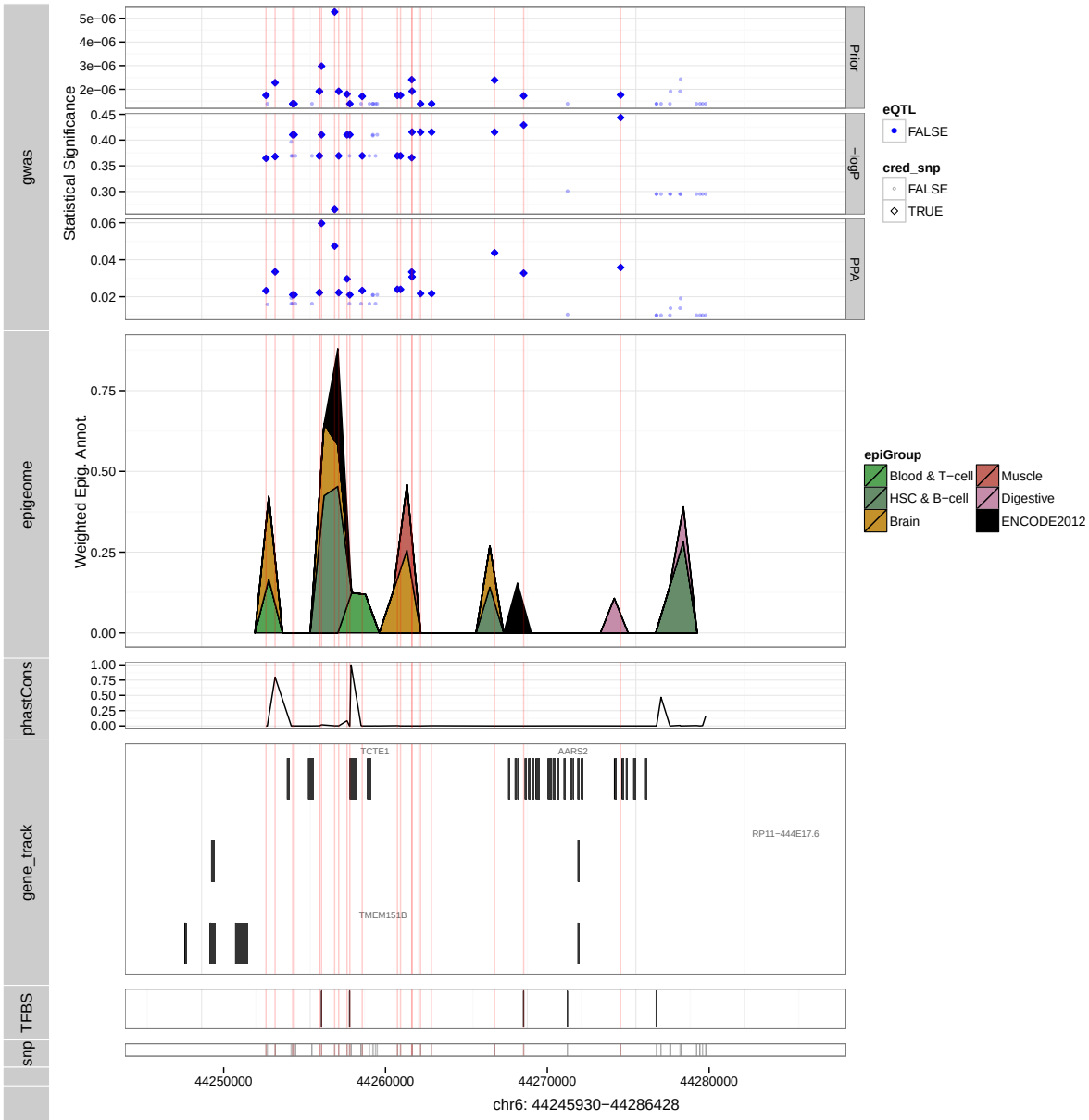
Rheumatoid Arthritis



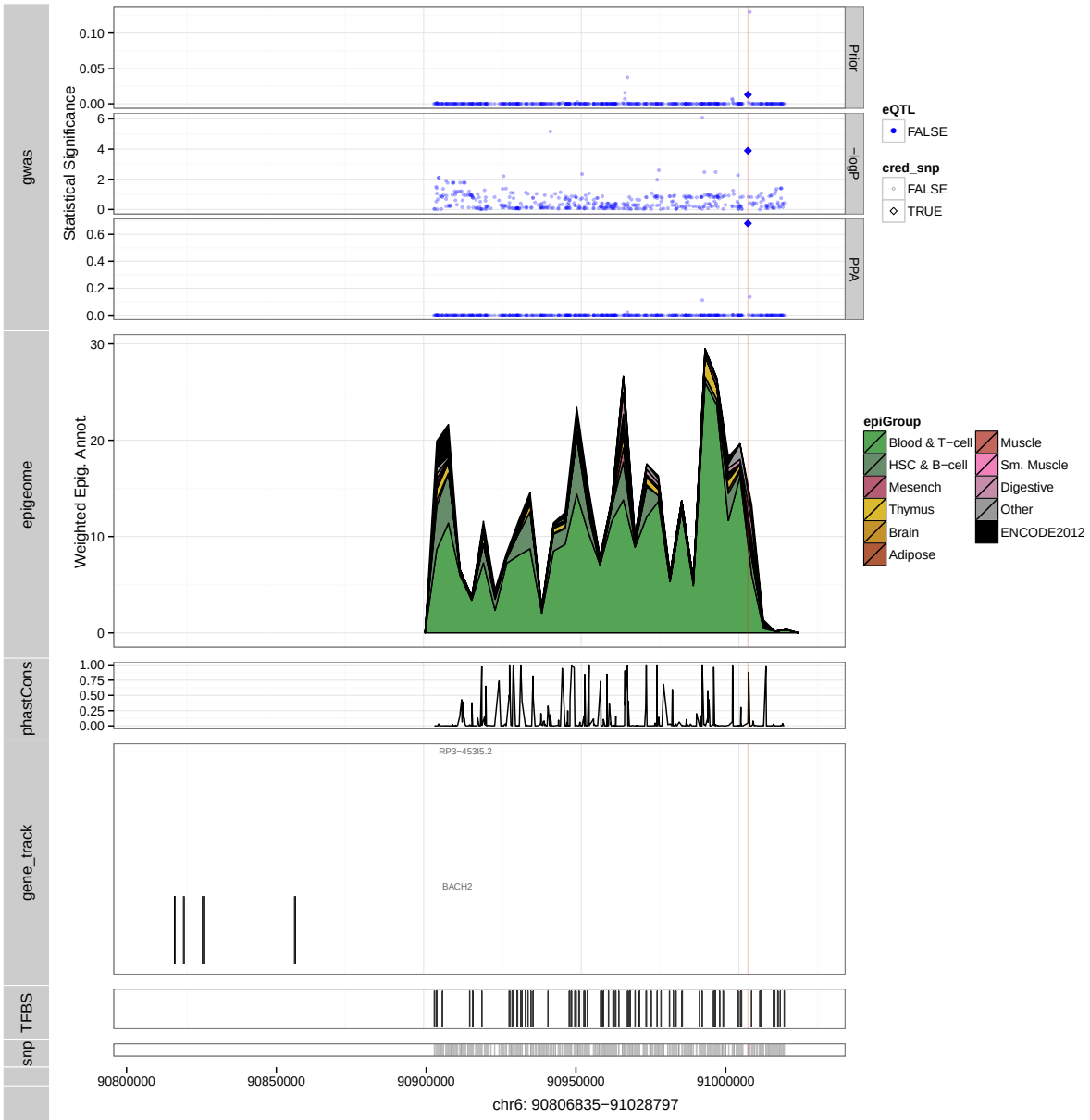
Rheumatoid Arthritis



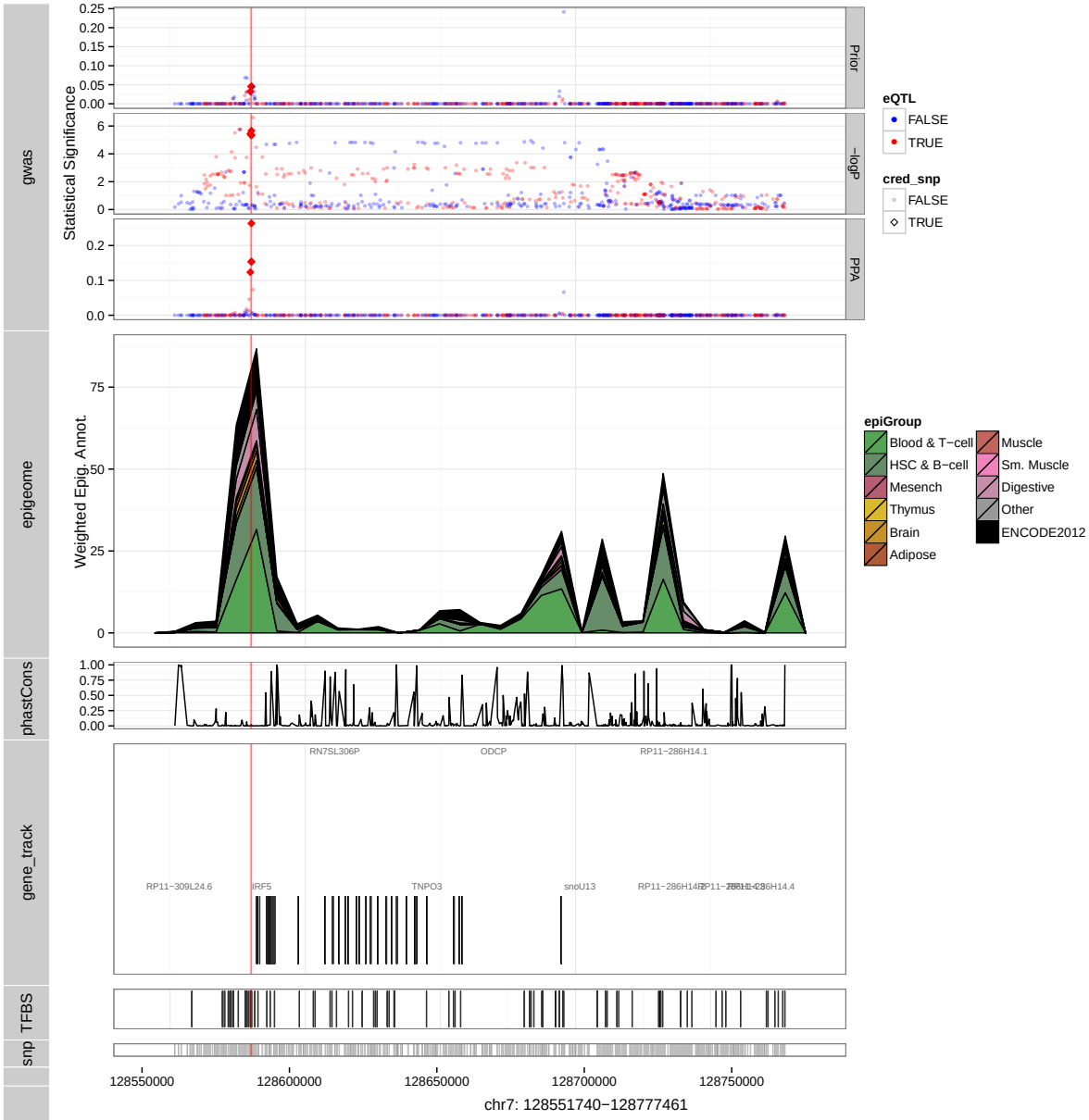
Rheumatoid Arthritis



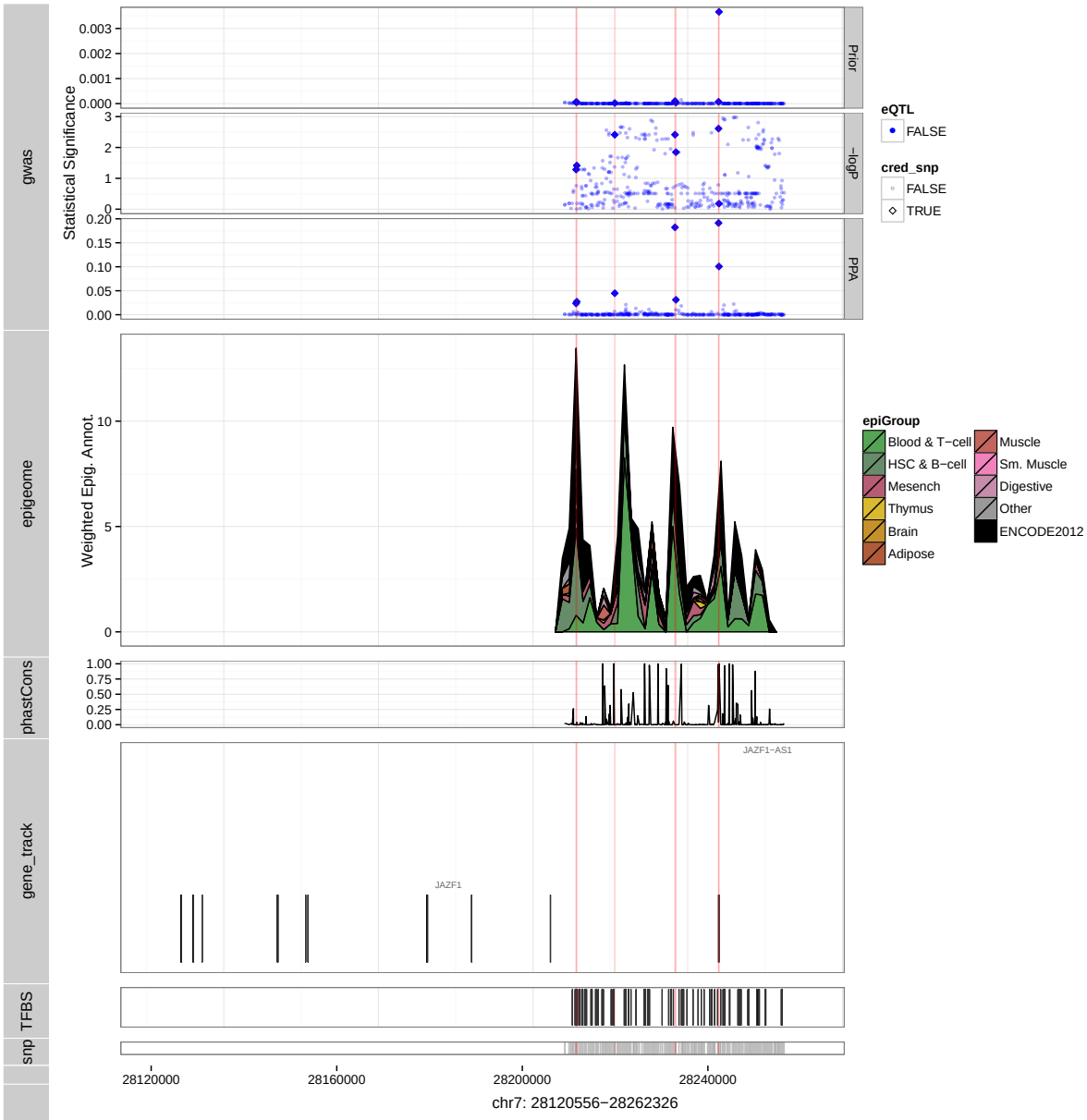
Rheumatoid Arthritis



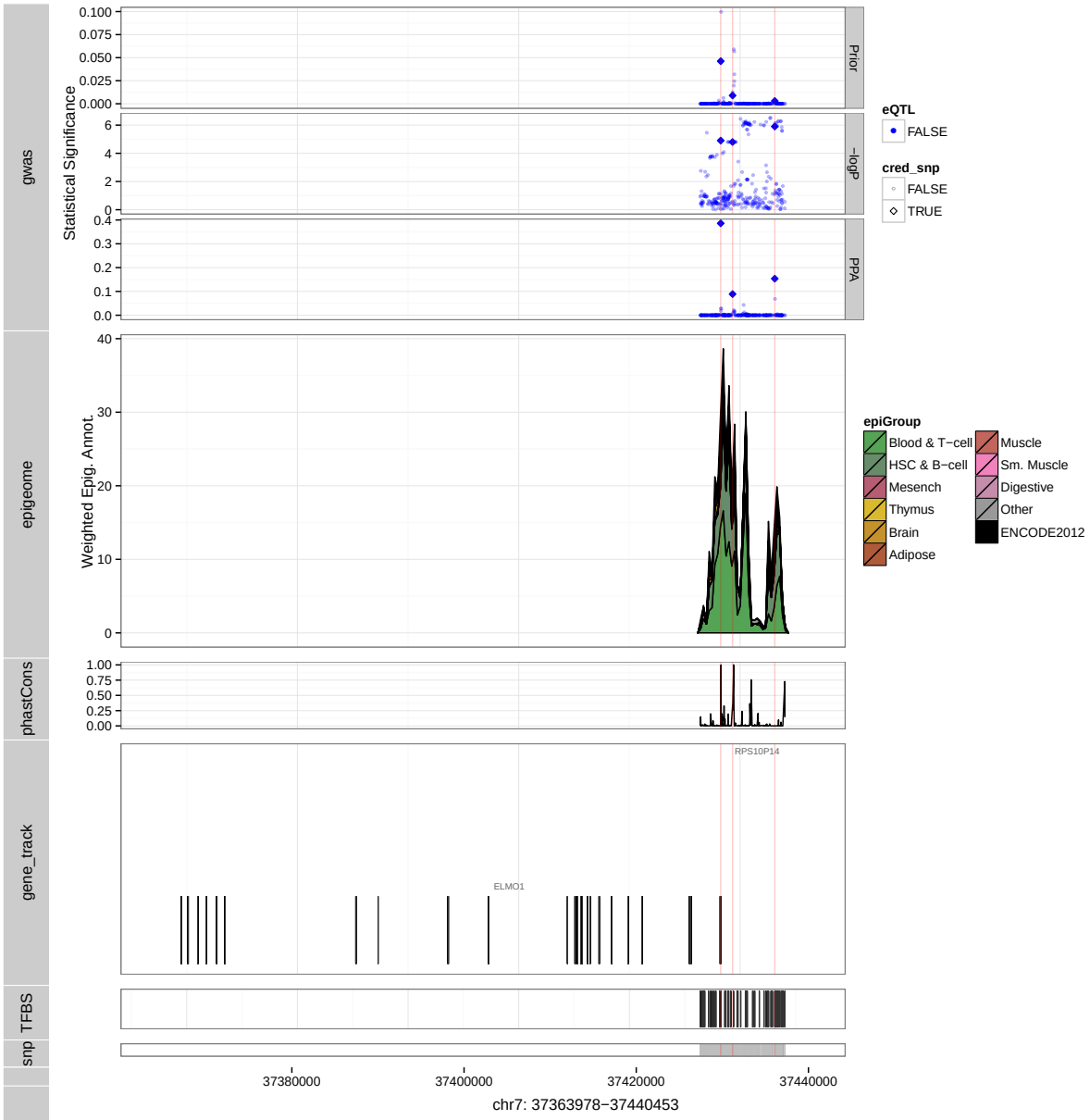
Rheumatoid Arthritis



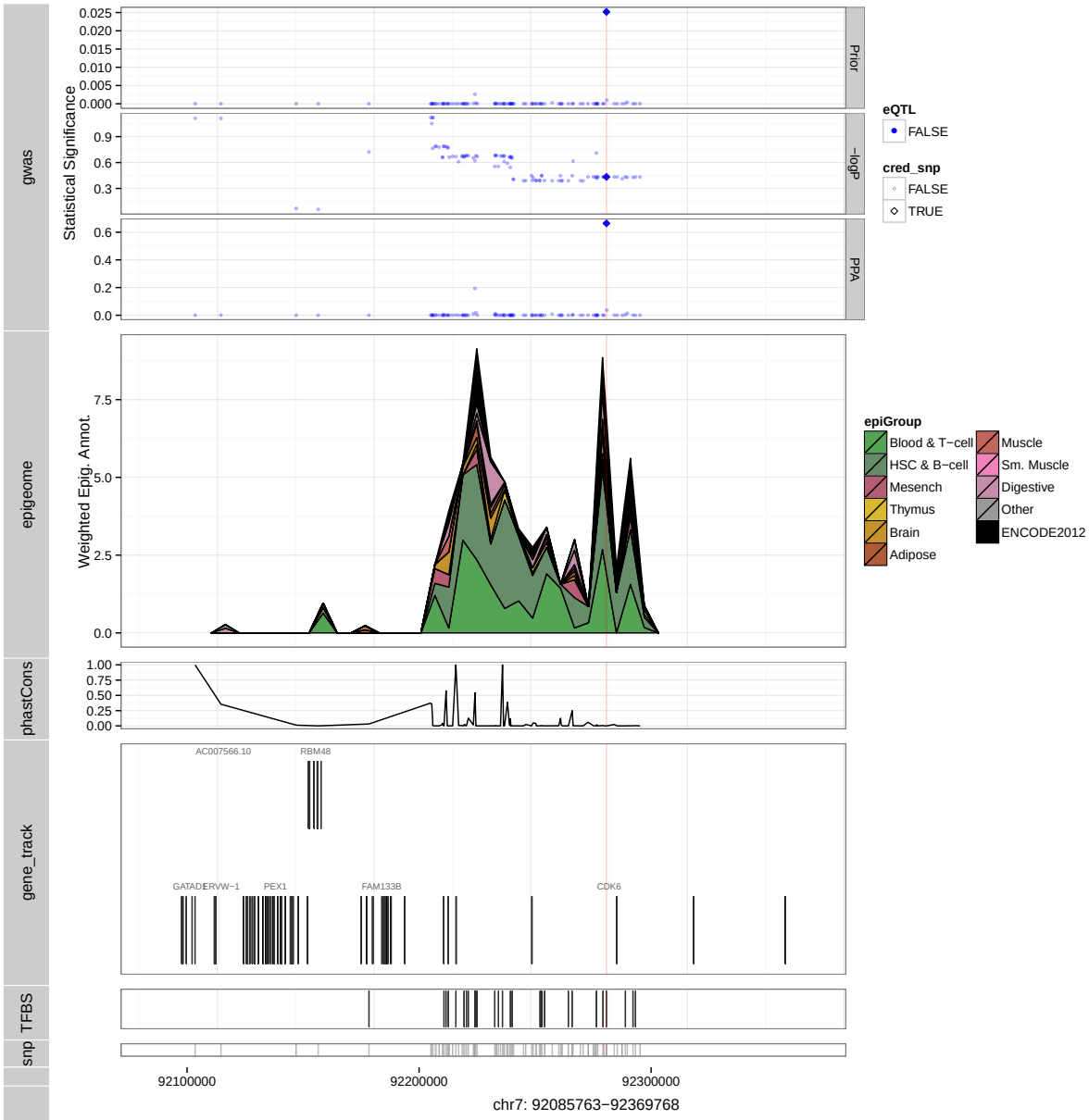
Rheumatoid Arthritis



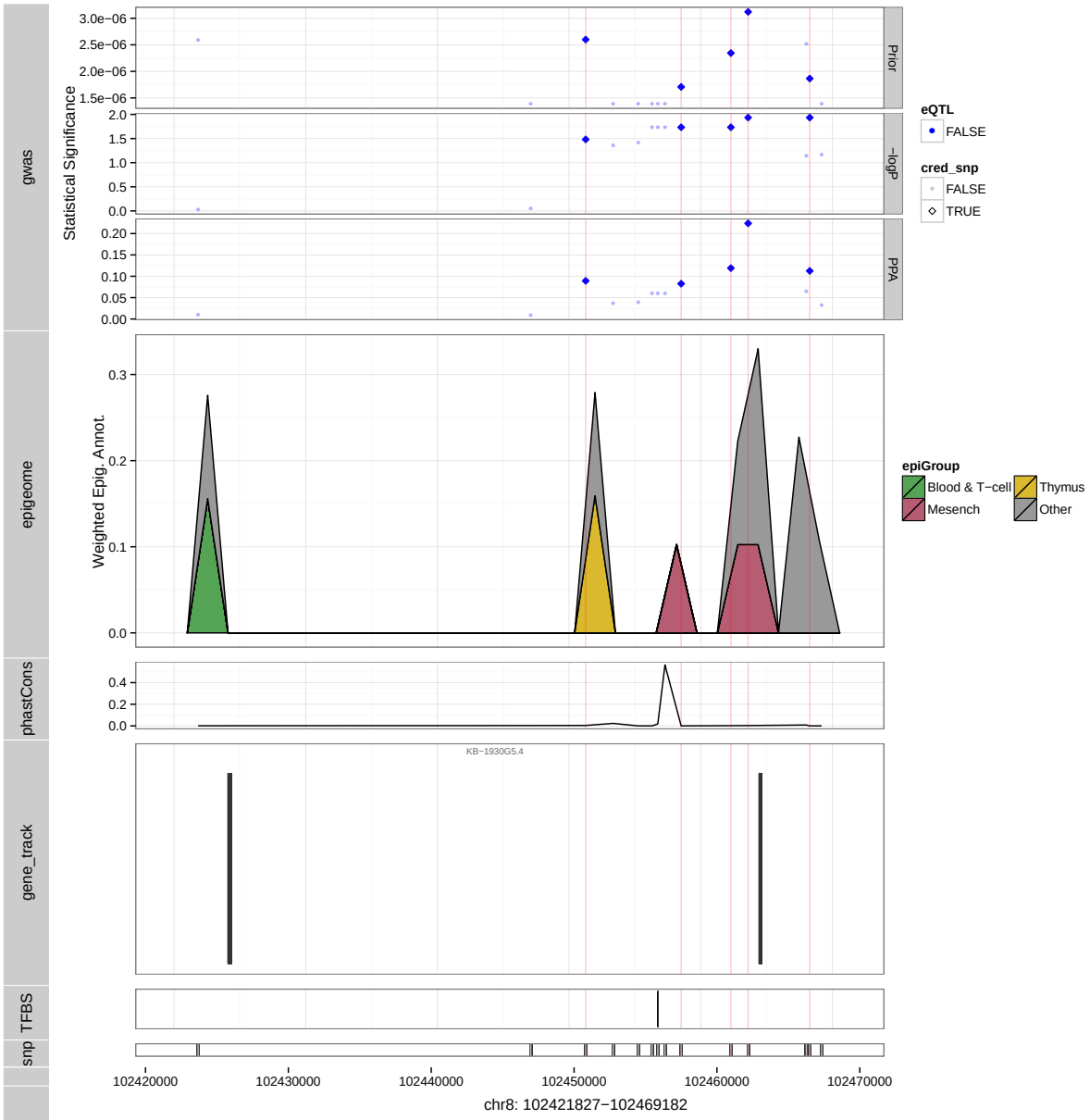
Rheumatoid Arthritis



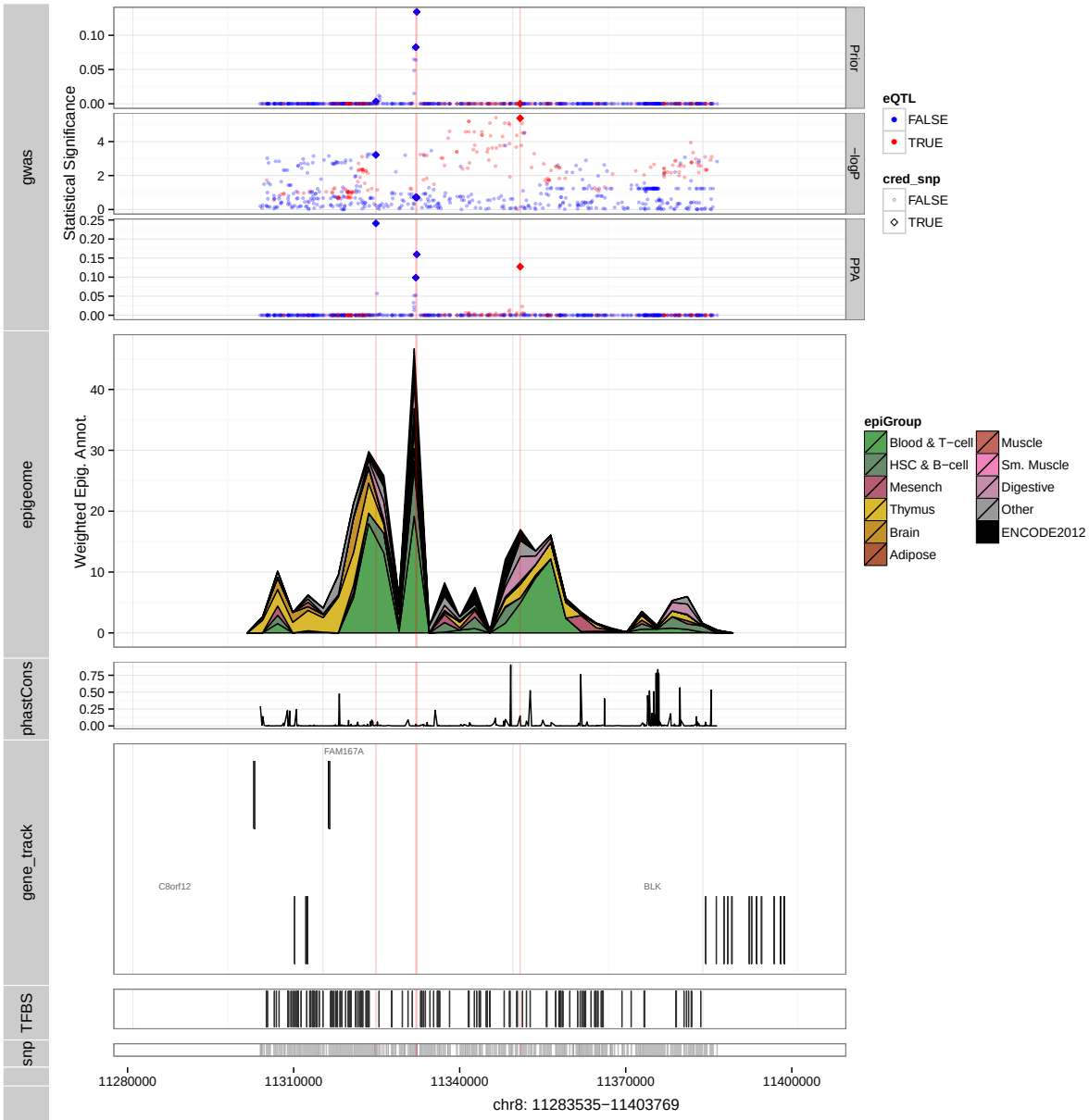
Rheumatoid Arthritis



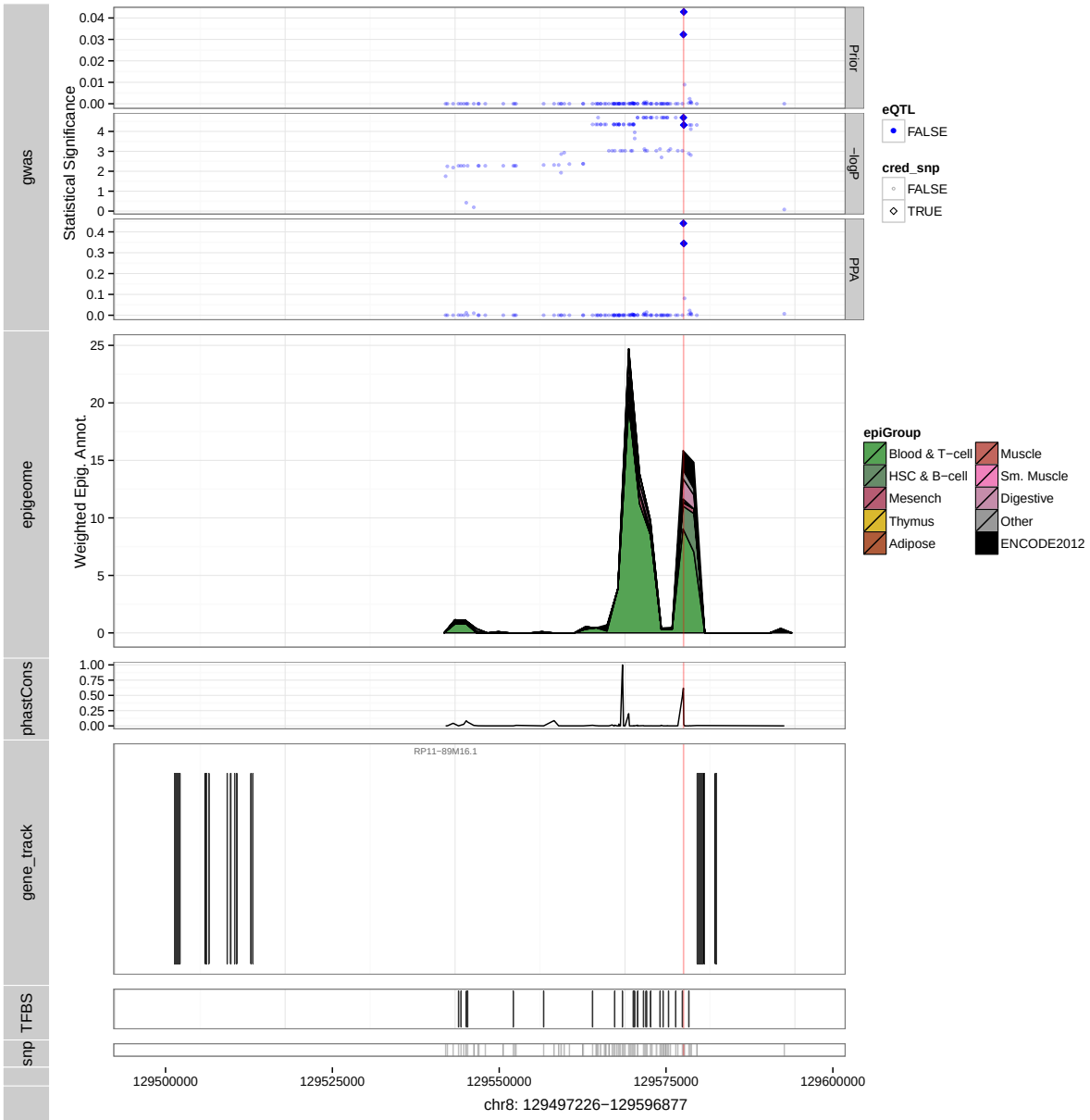
Rheumatoid Arthritis



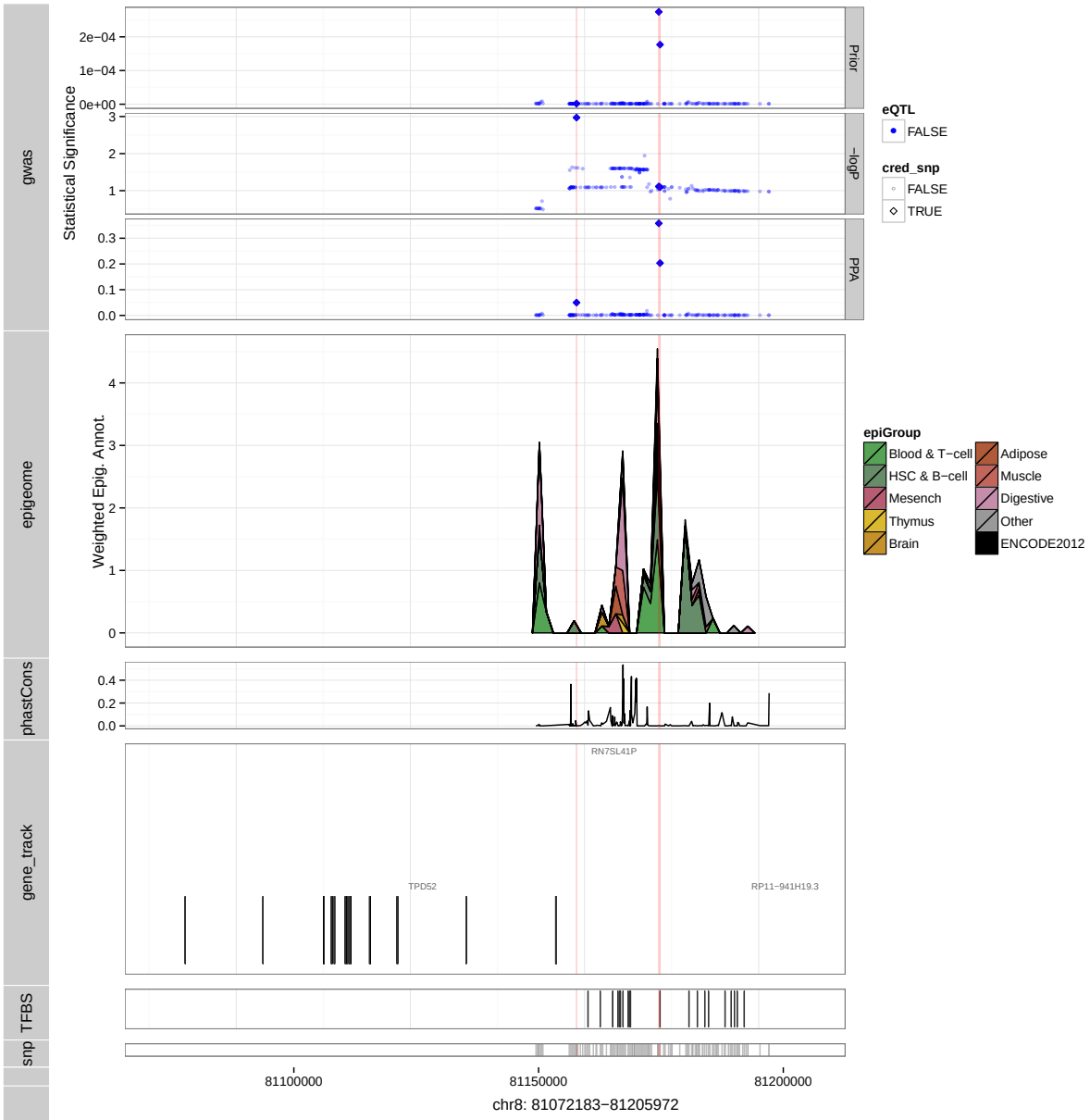
Rheumatoid Arthritis



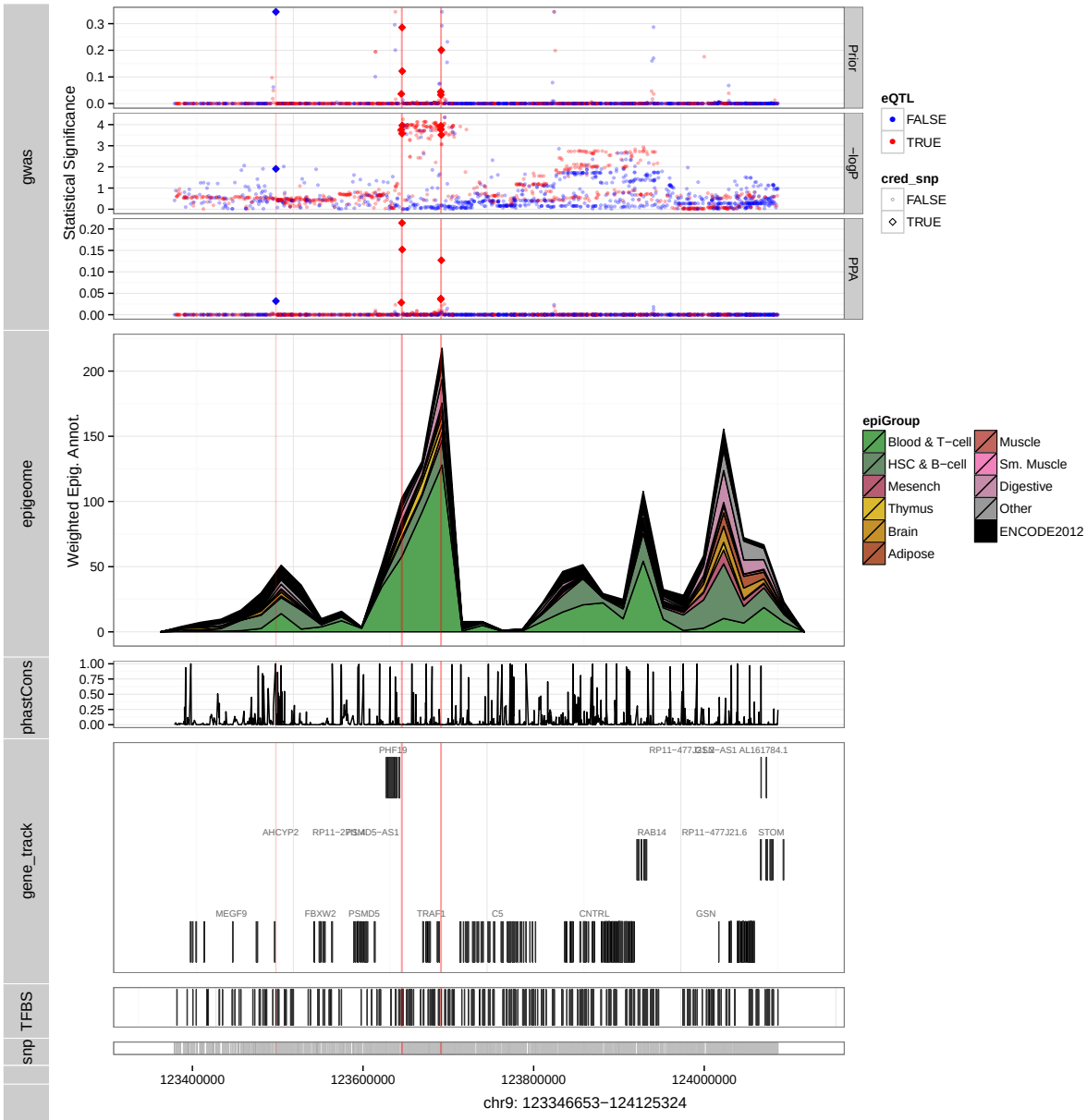
Rheumatoid Arthritis



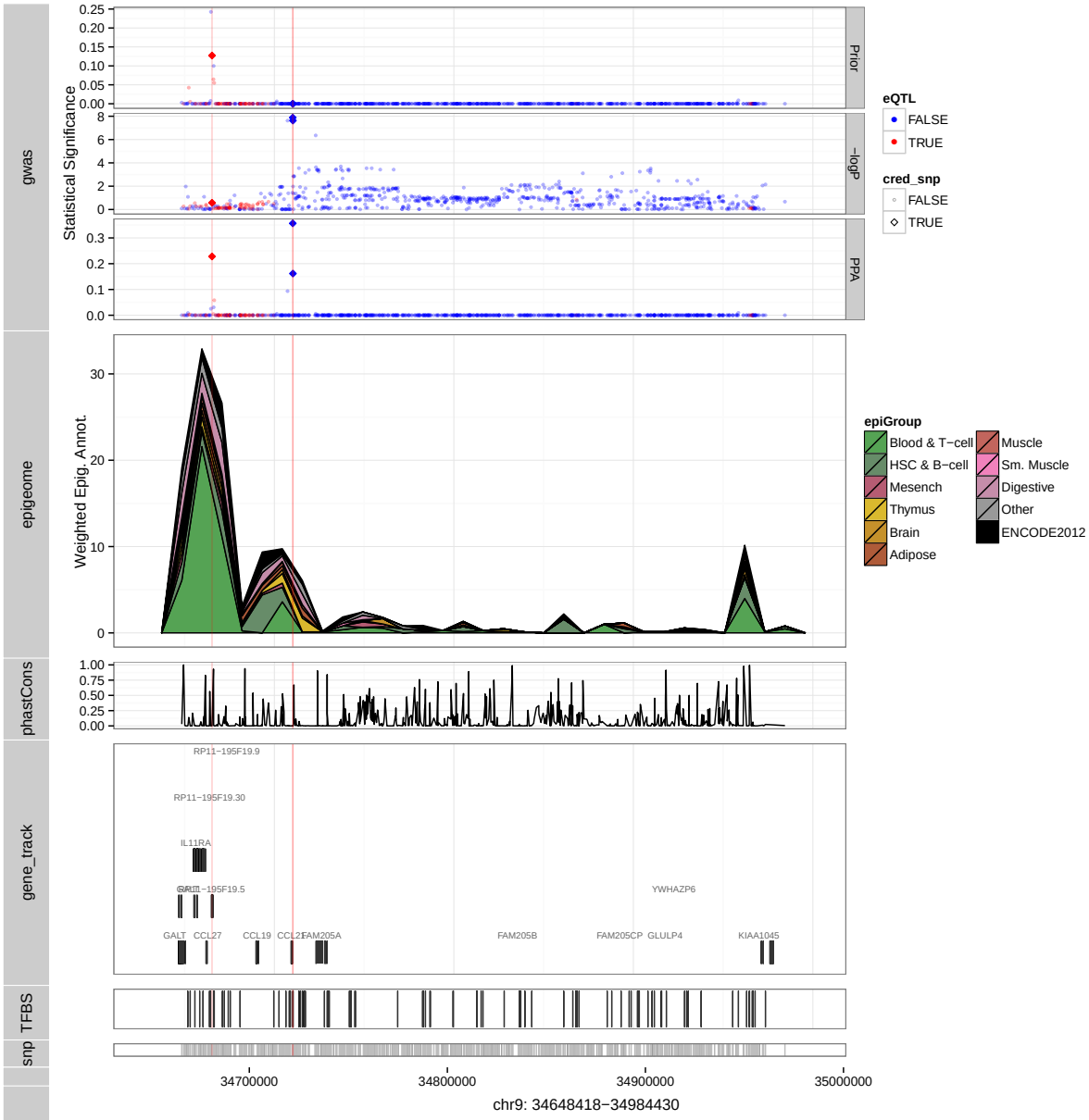
Rheumatoid Arthritis



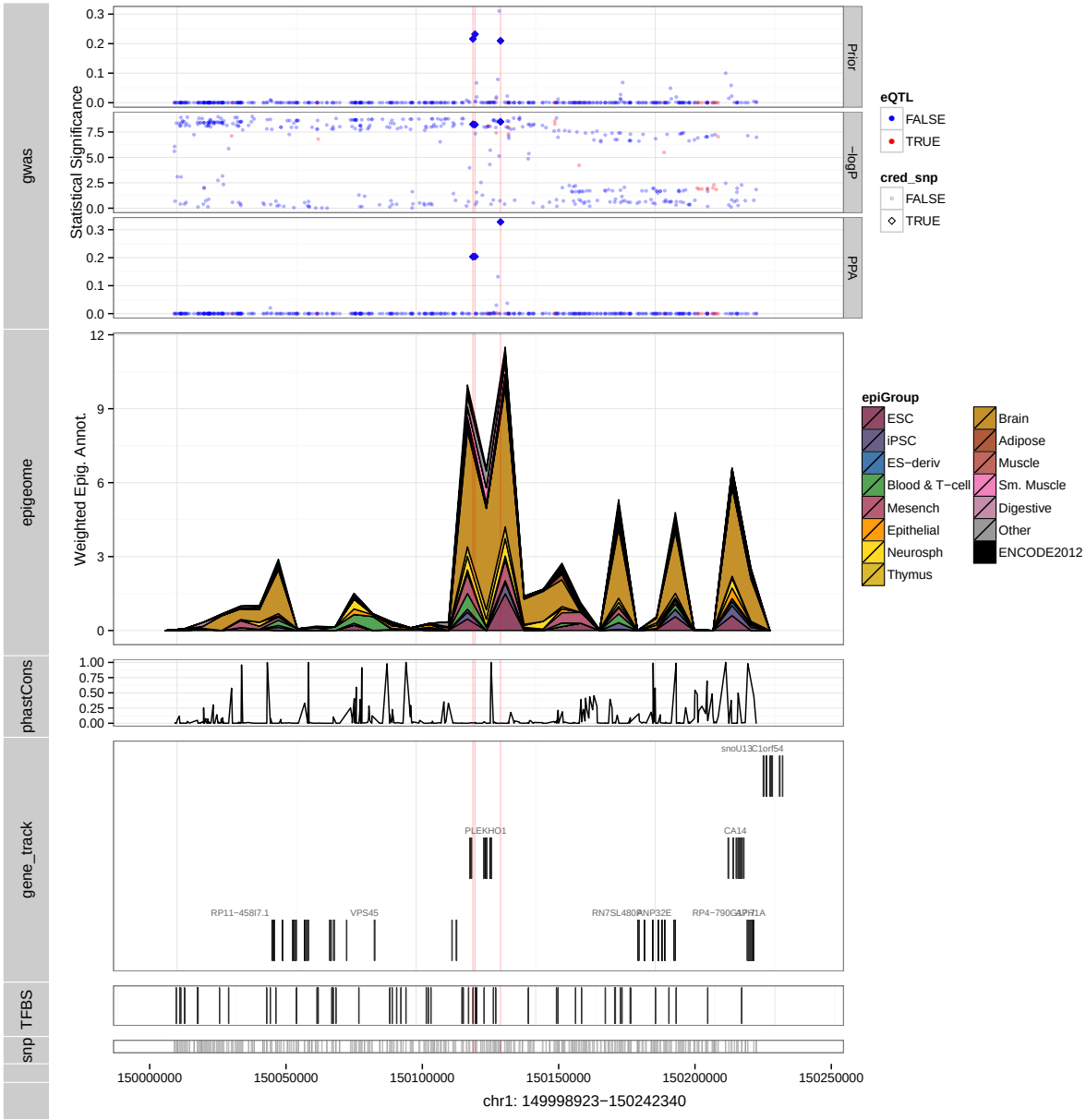
Rheumatoid Arthritis



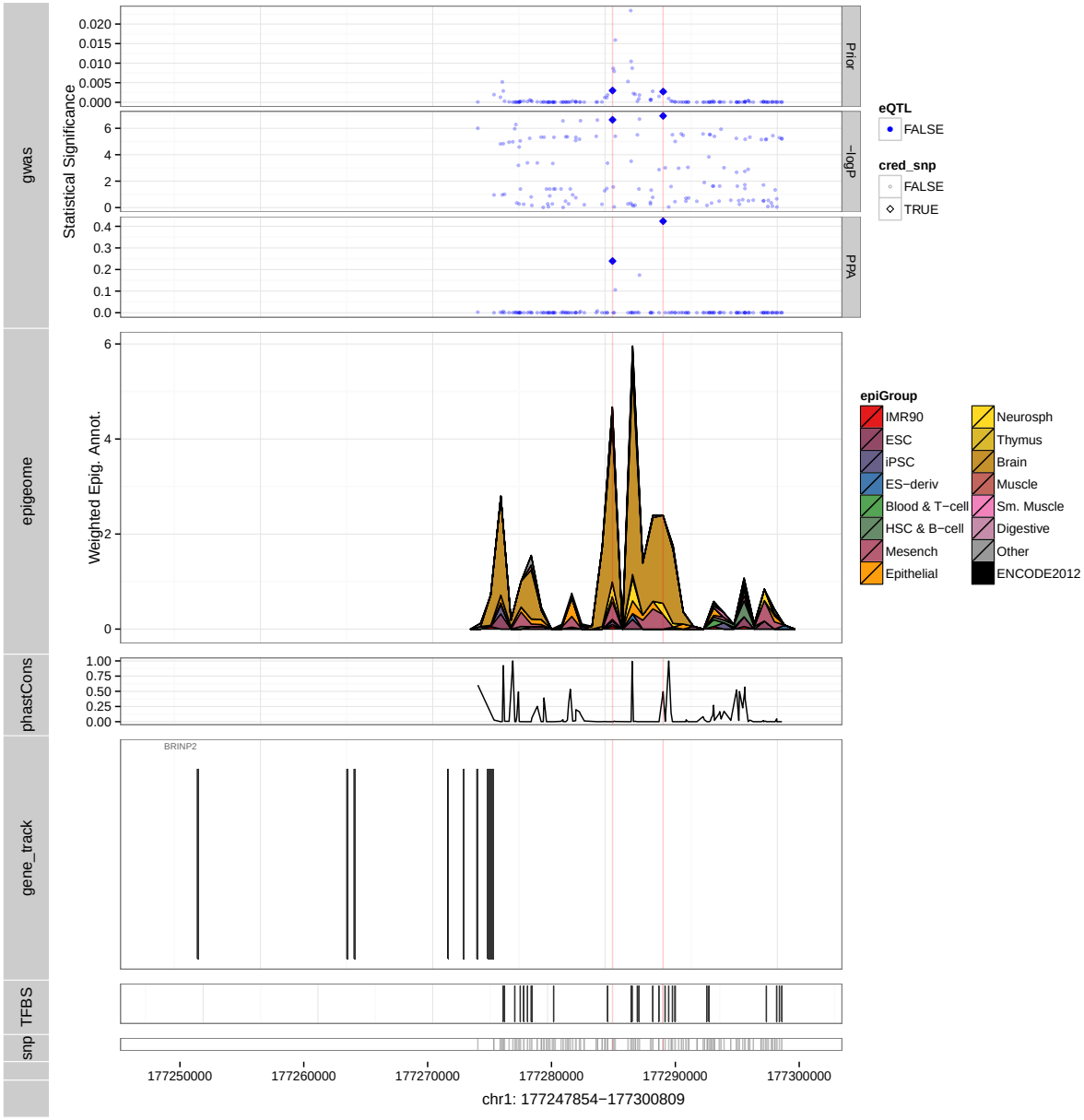
Rheumatoid Arthritis



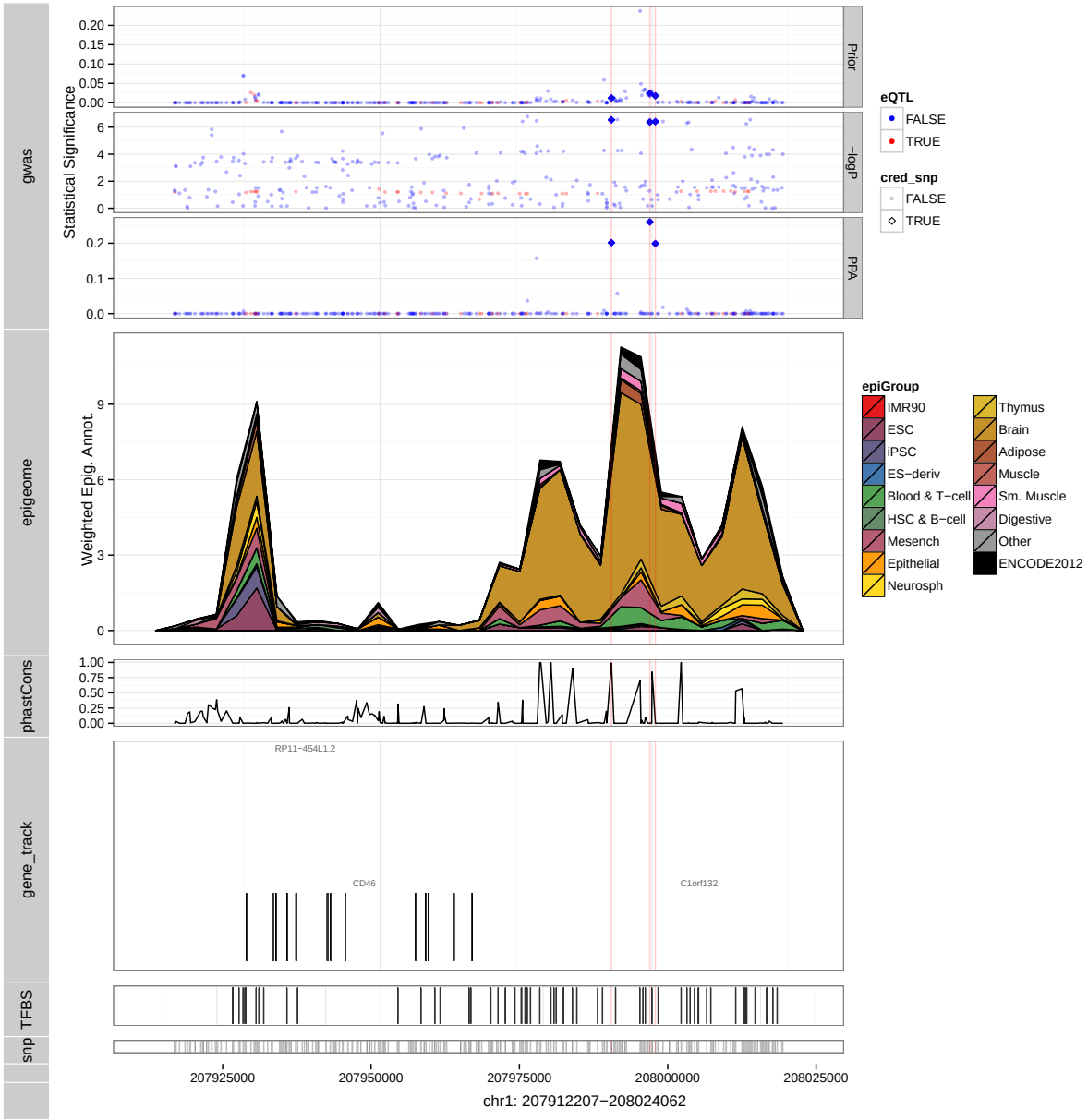
Schizophrenia



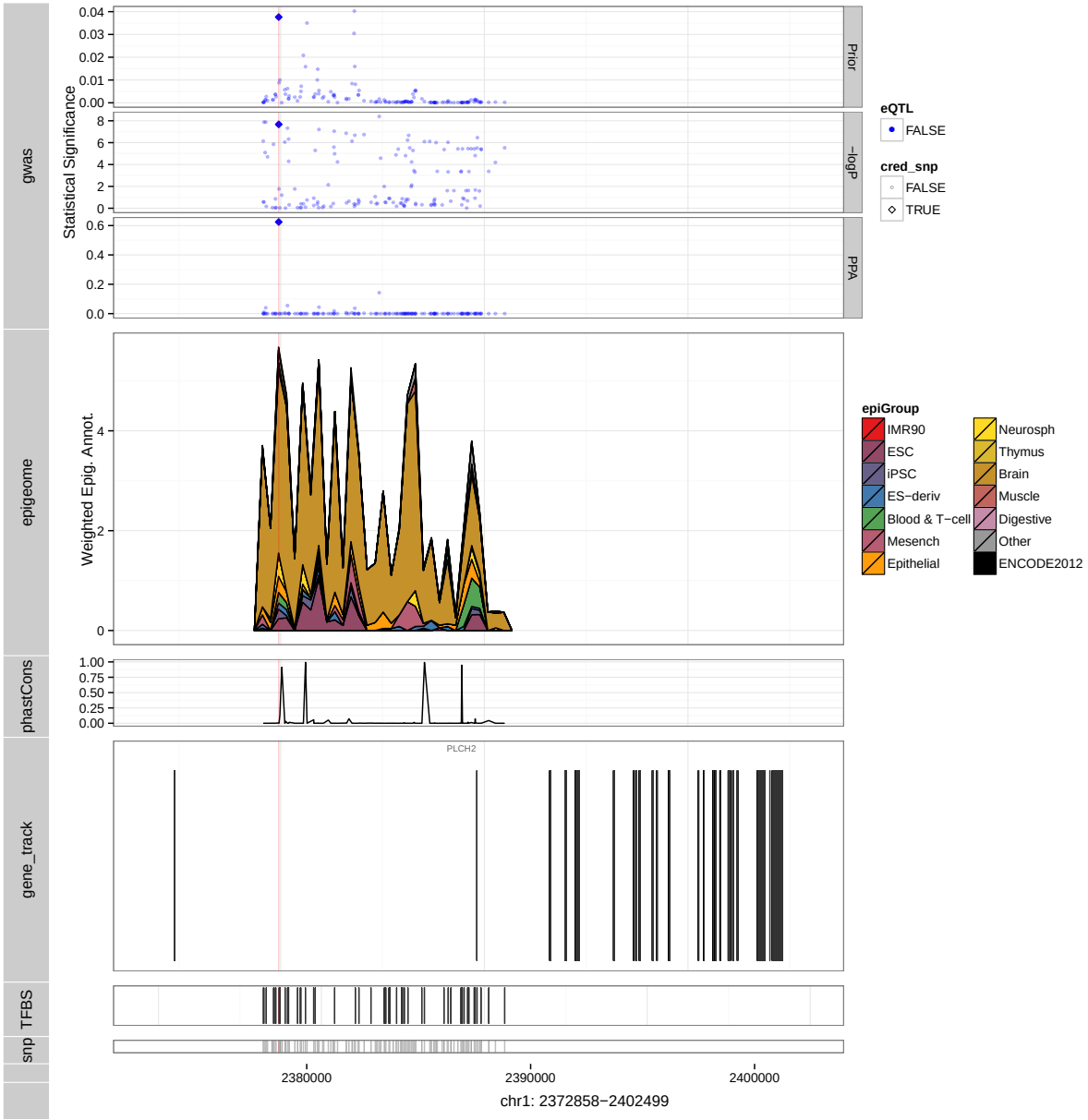
Schizophrenia



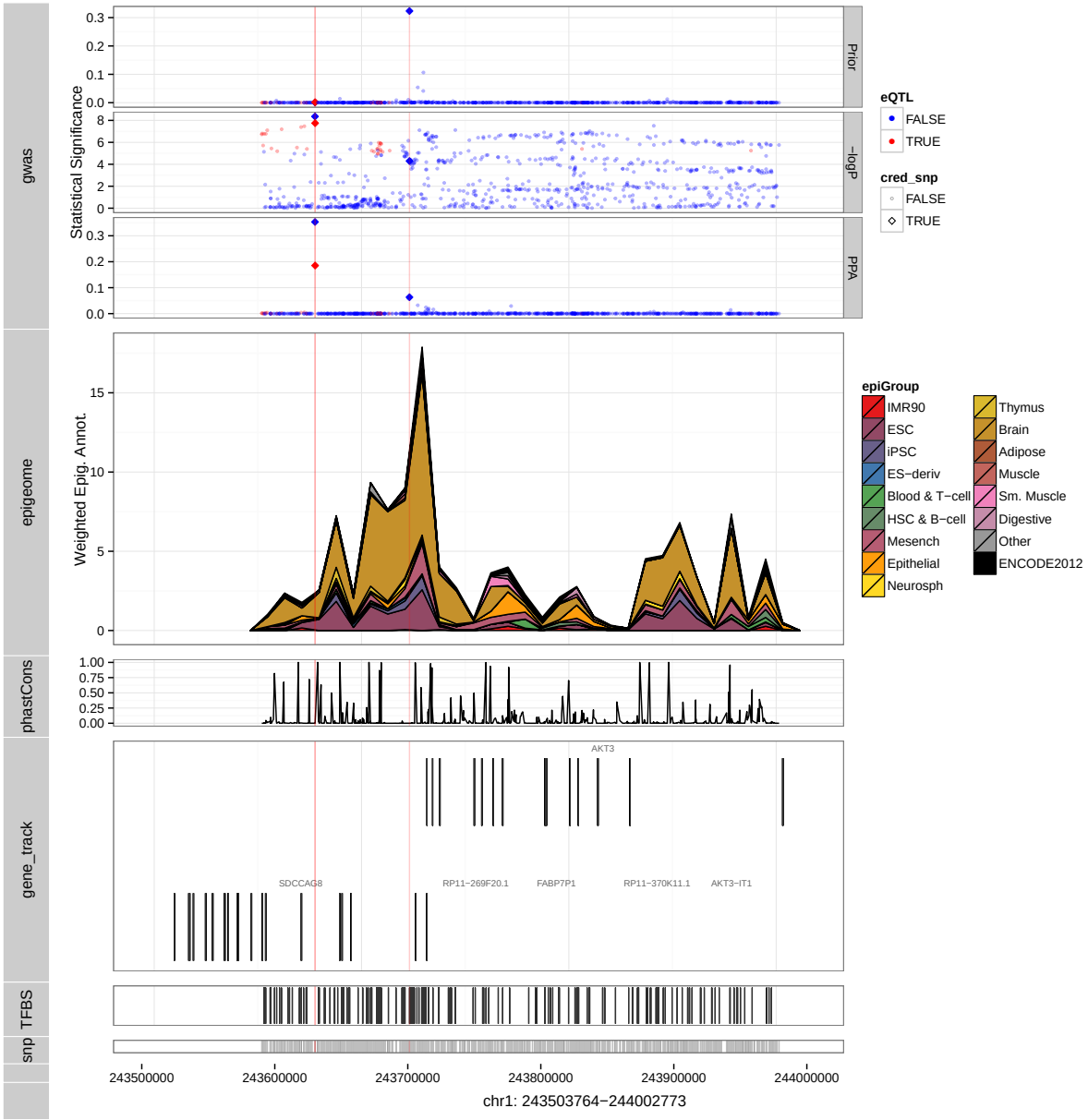
Schizophrenia



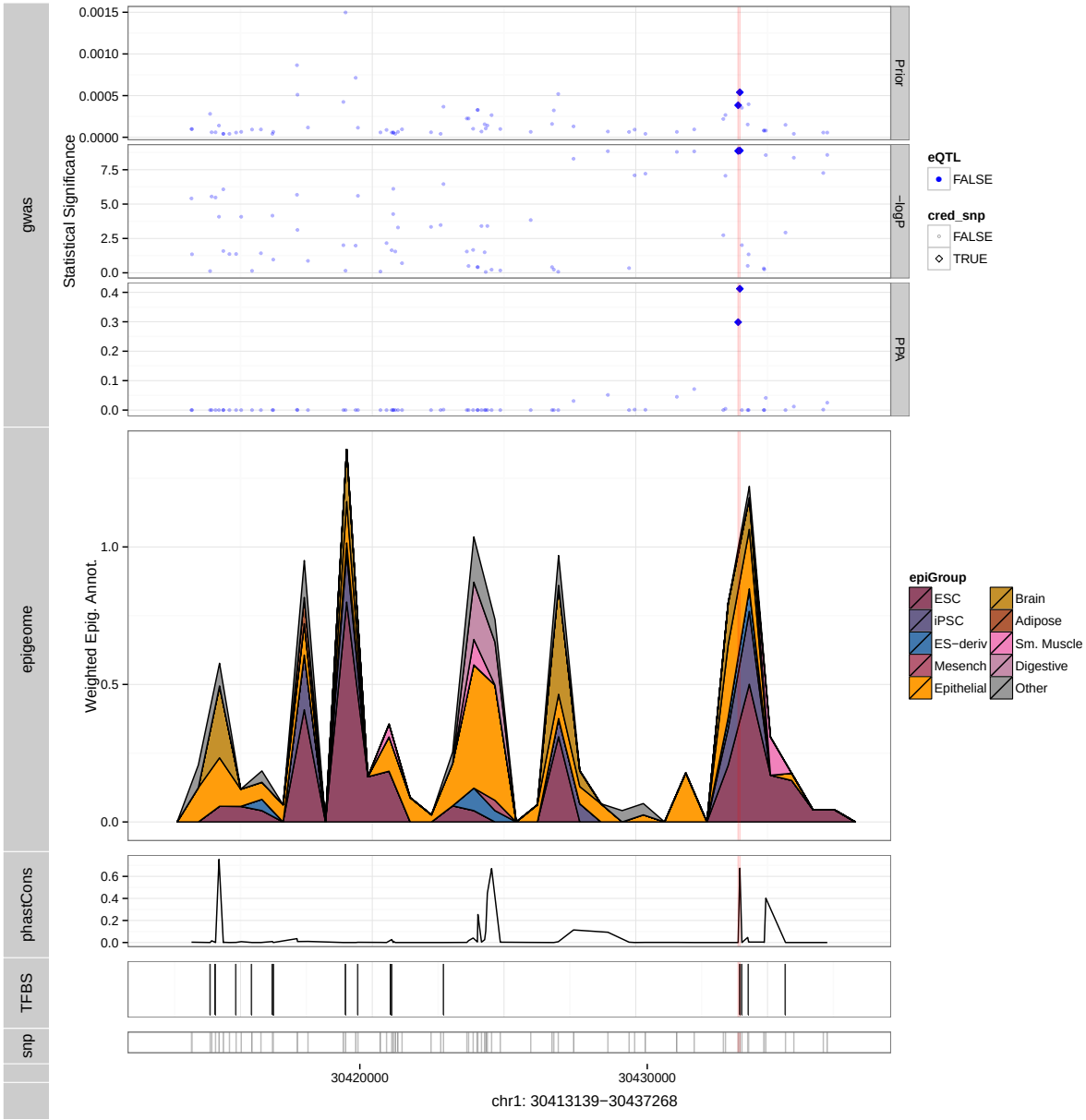
Schizophrenia



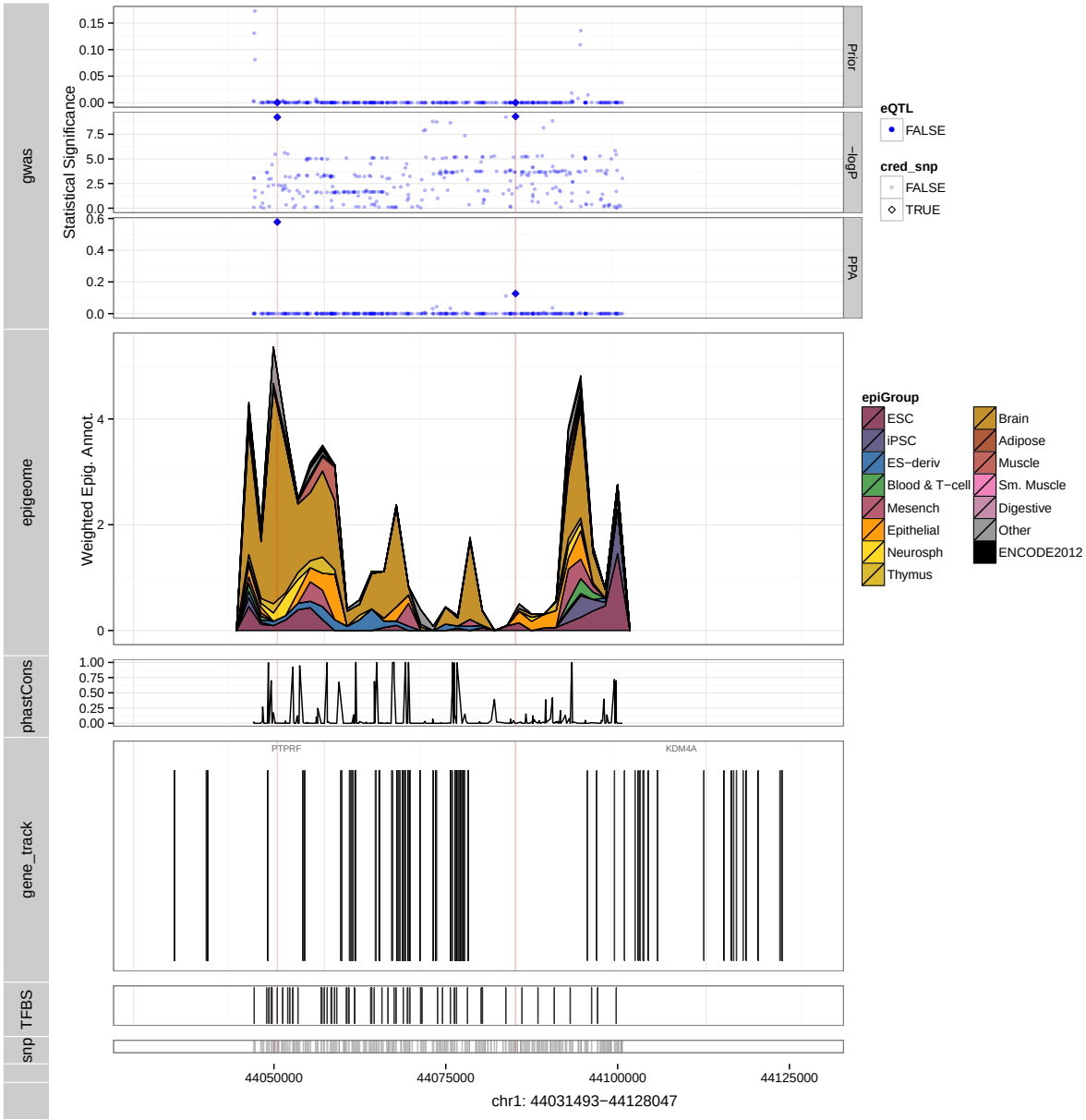
Schizophrenia



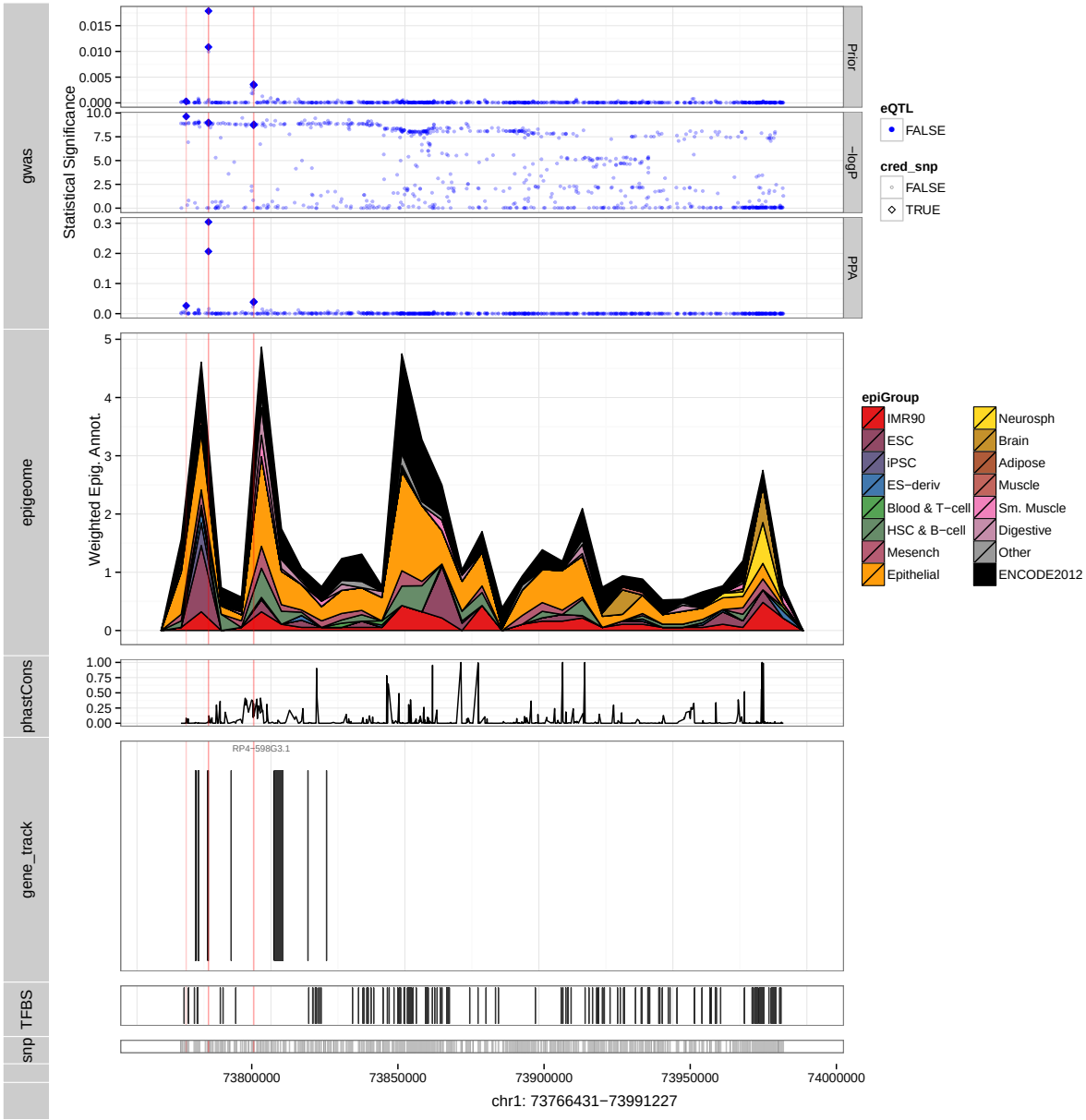
Schizophrenia



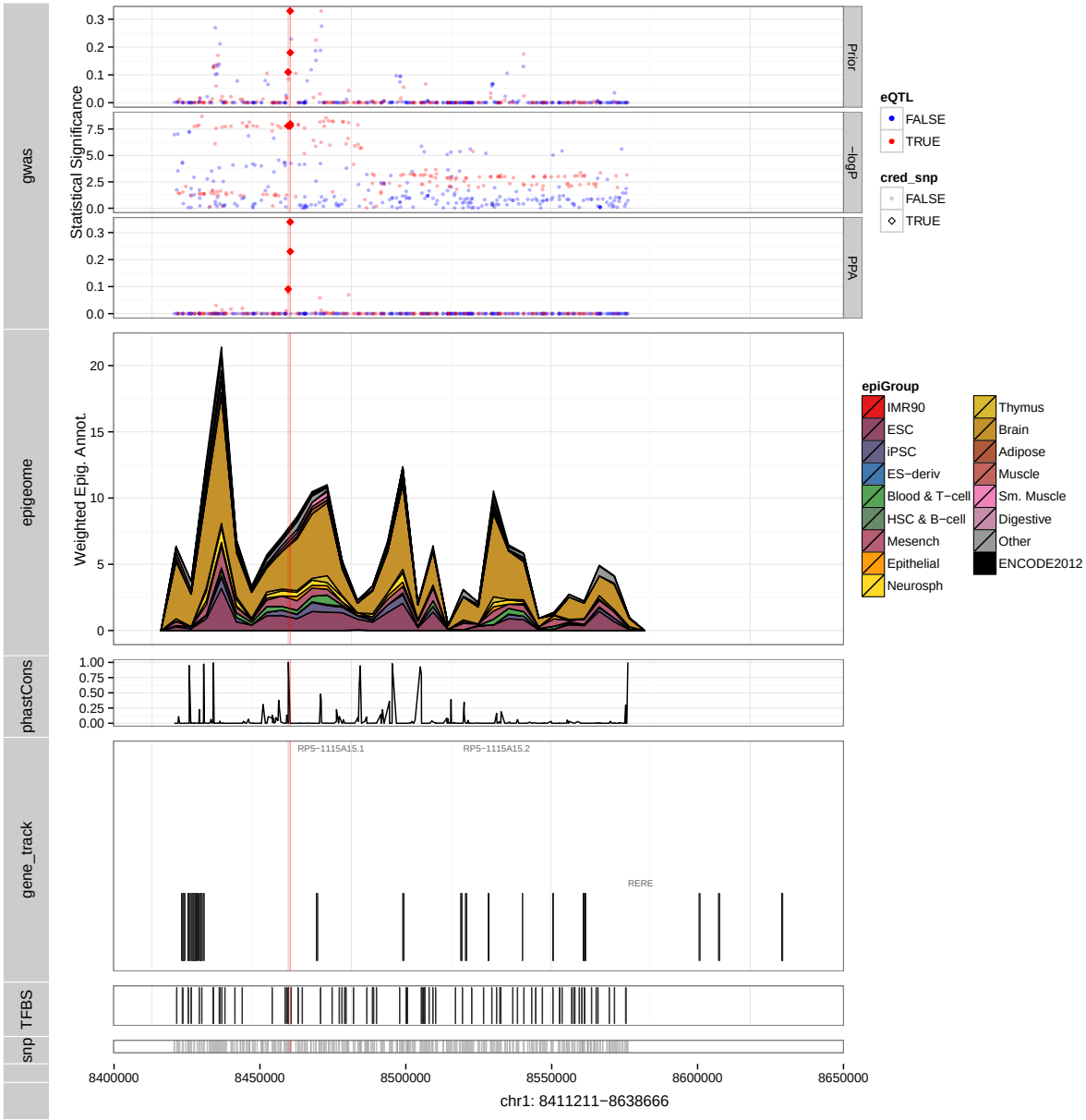
Schizophrenia



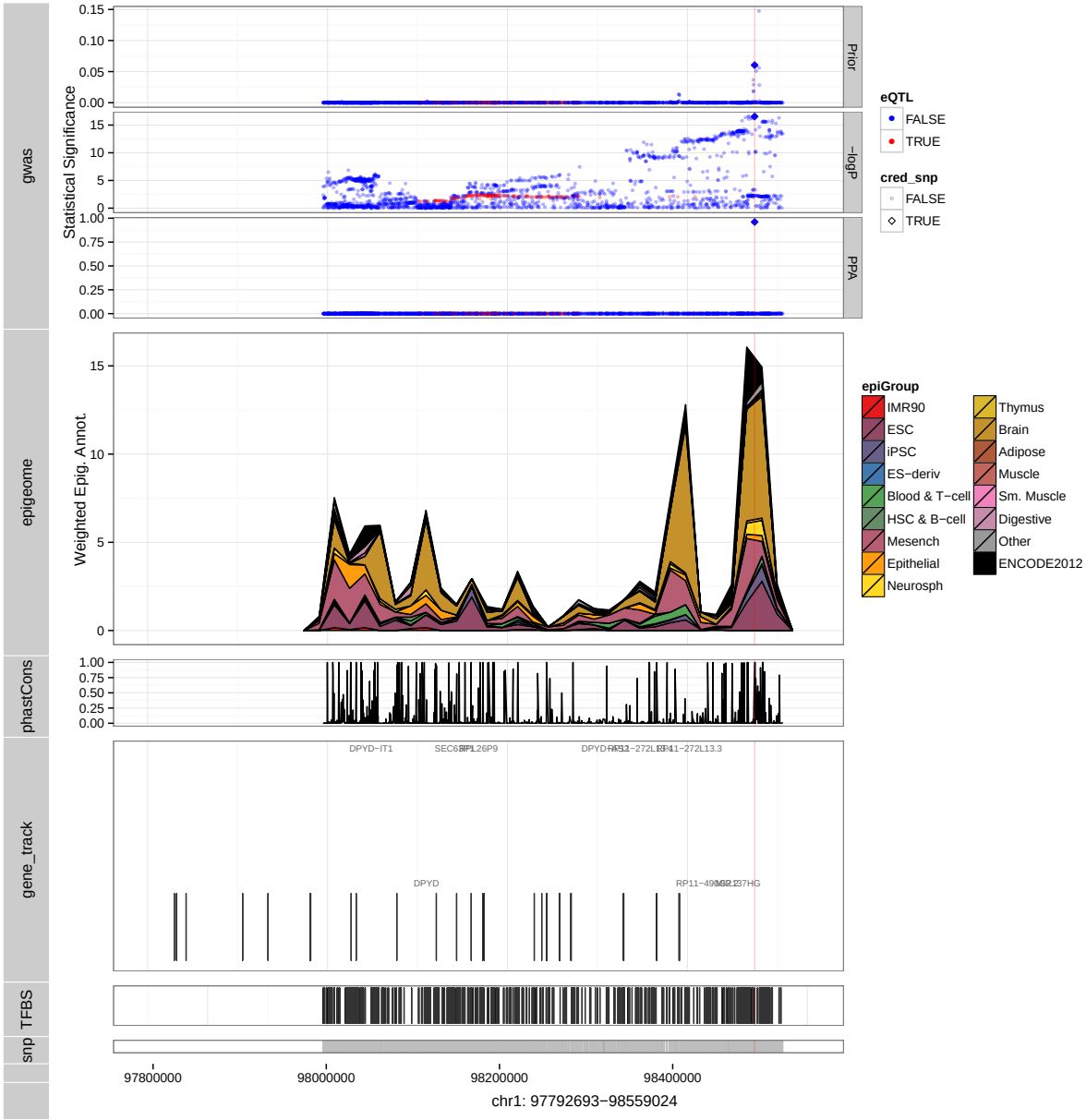
Schizophrenia



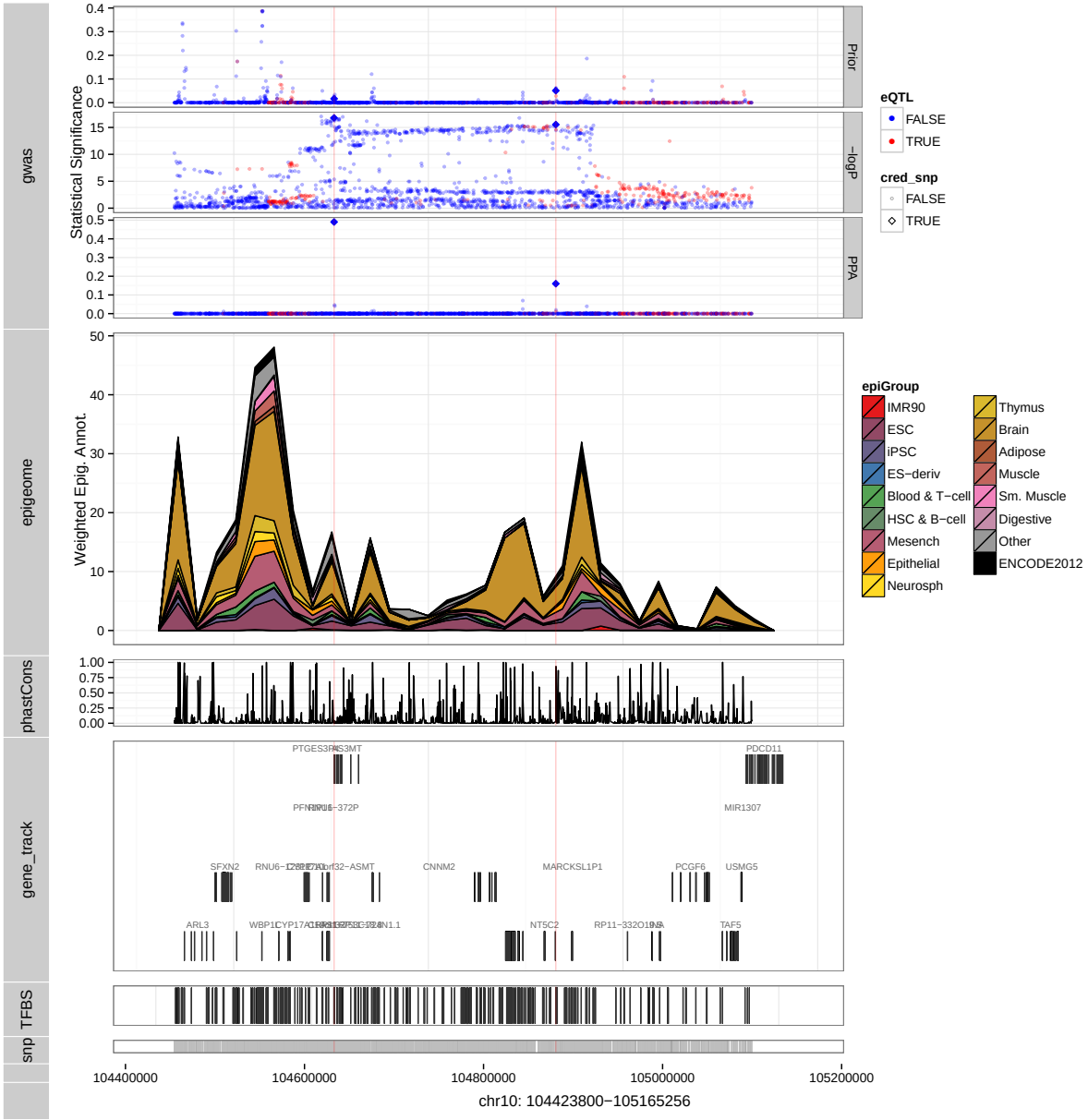
Schizophrenia



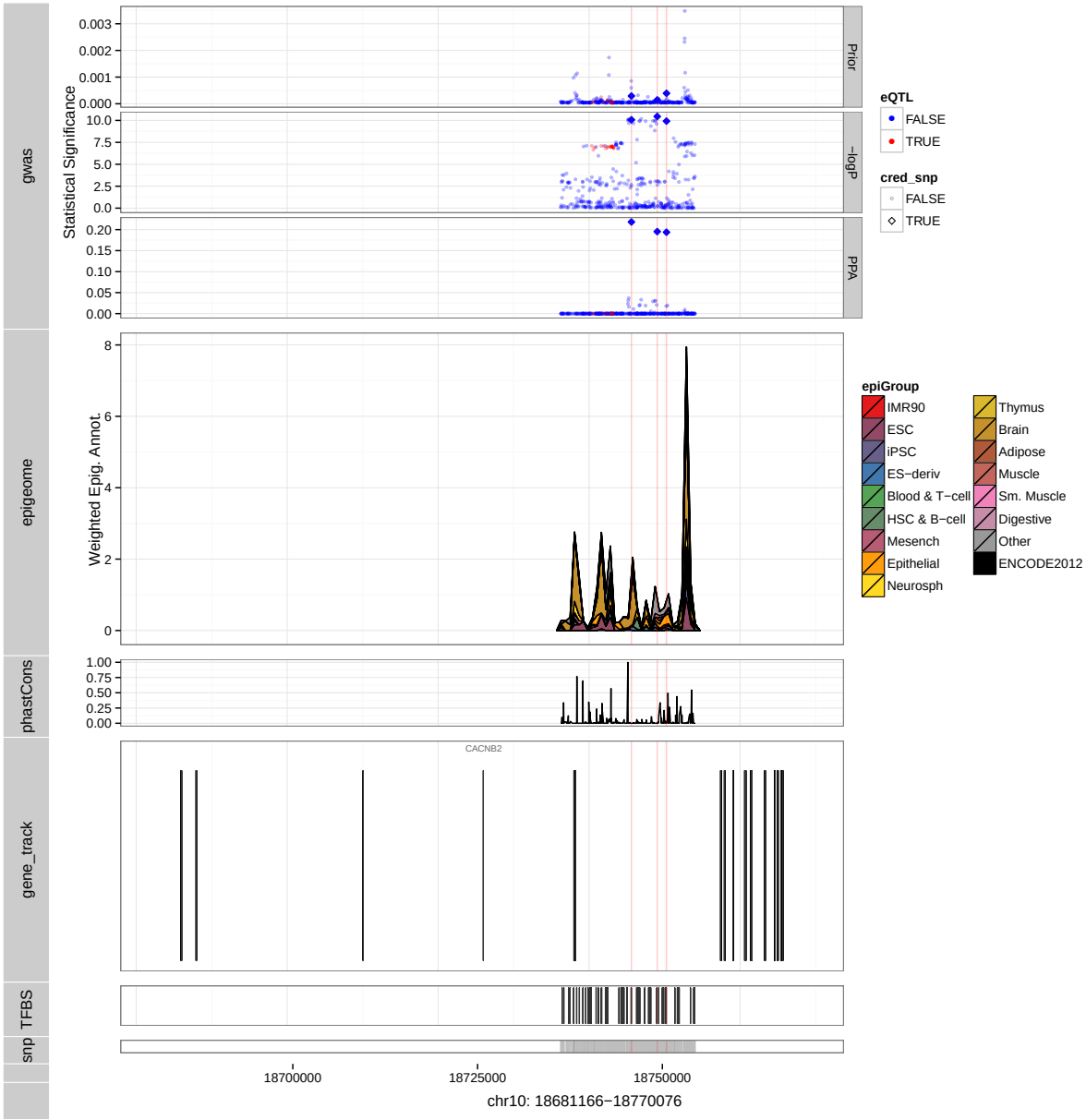
Schizophrenia



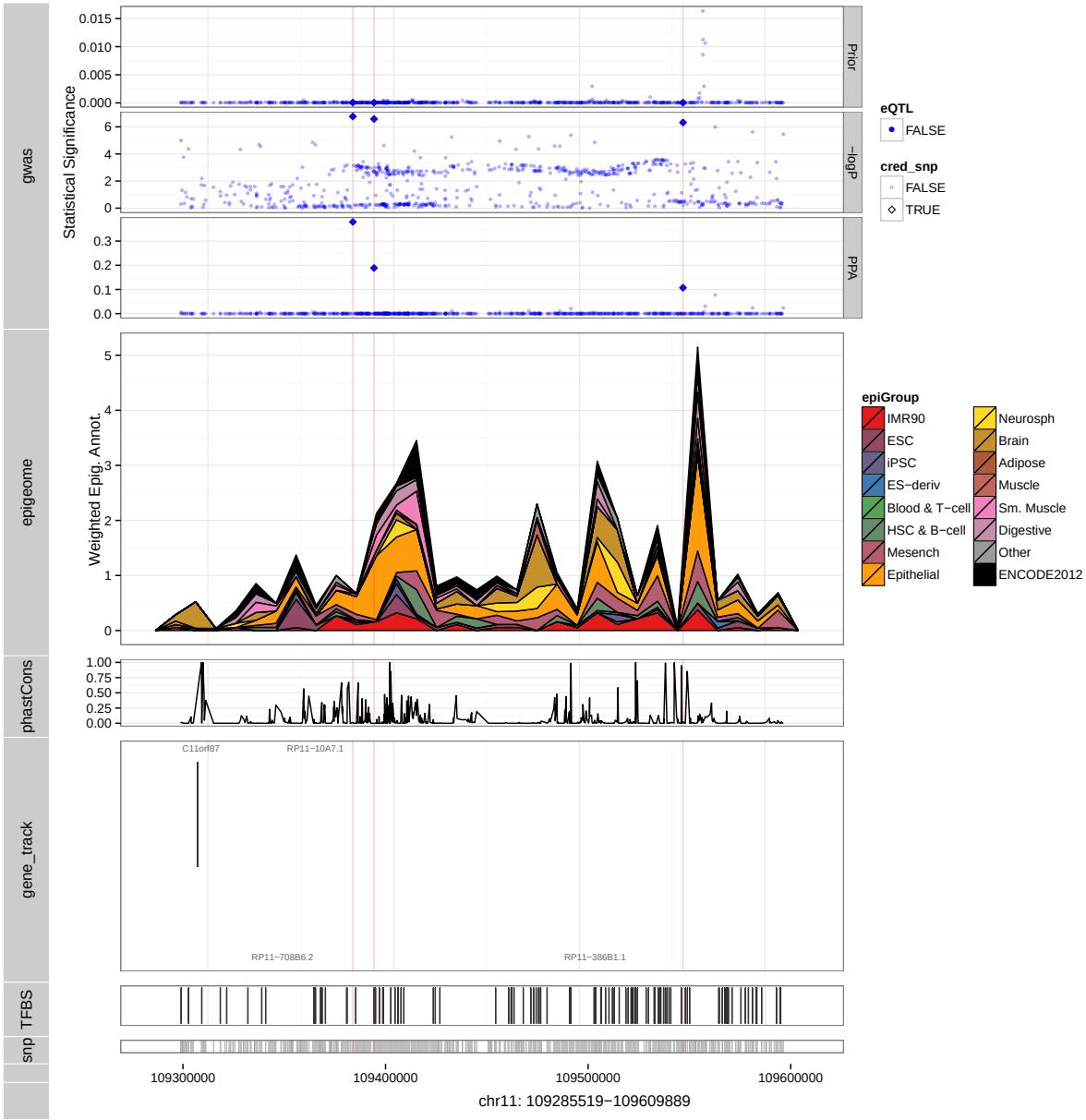
Schizophrenia



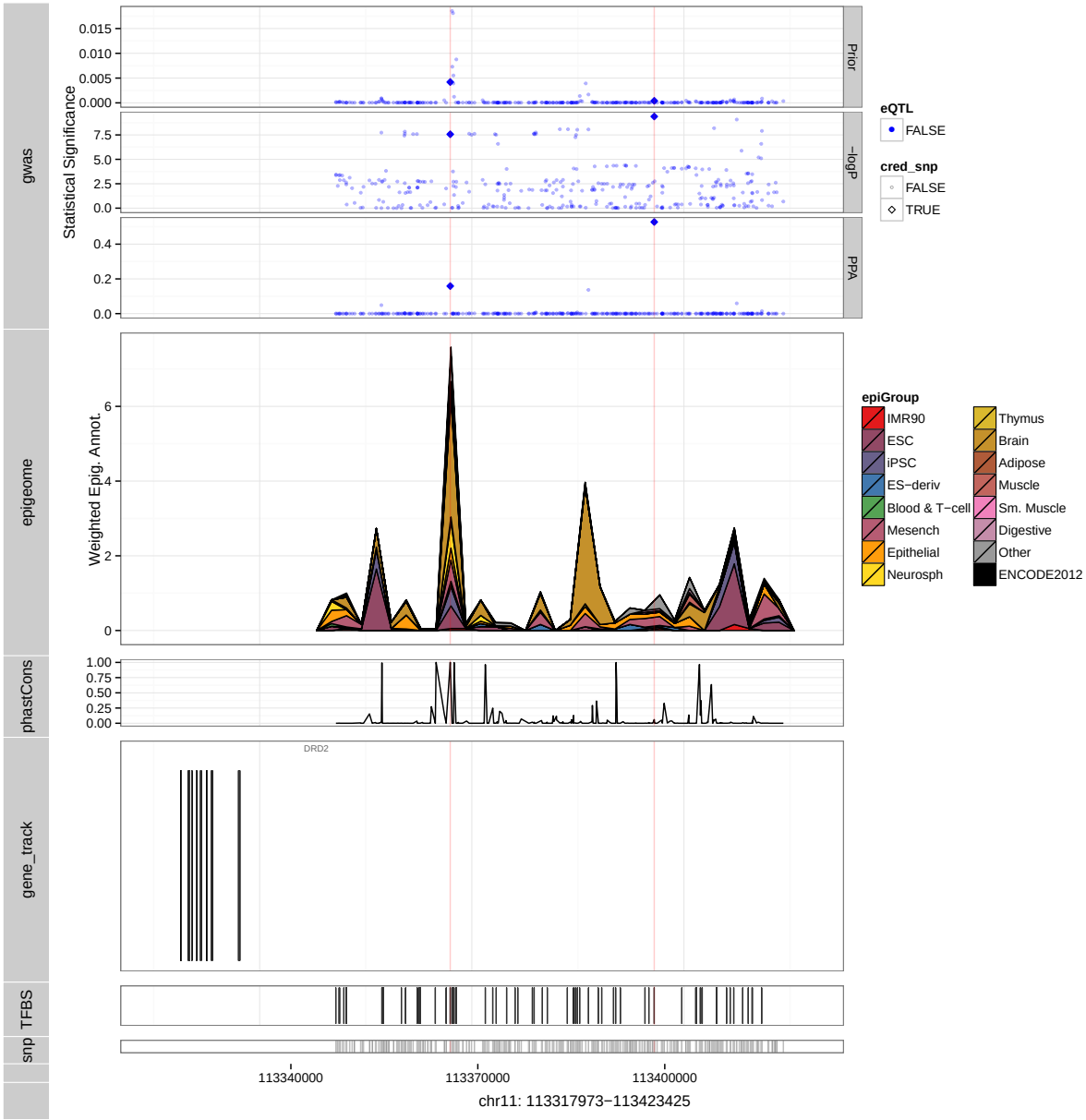
Schizophrenia



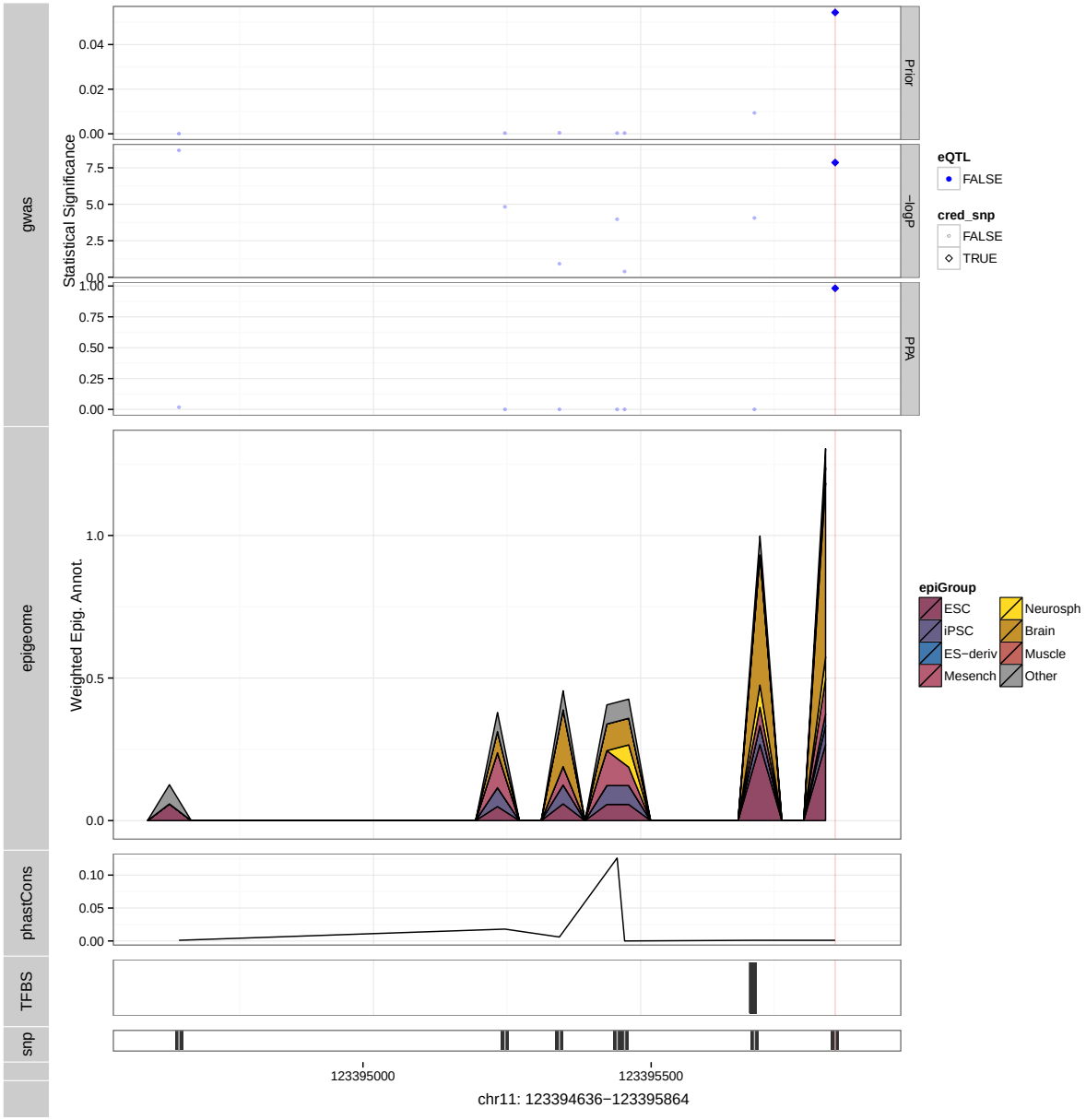
Schizophrenia



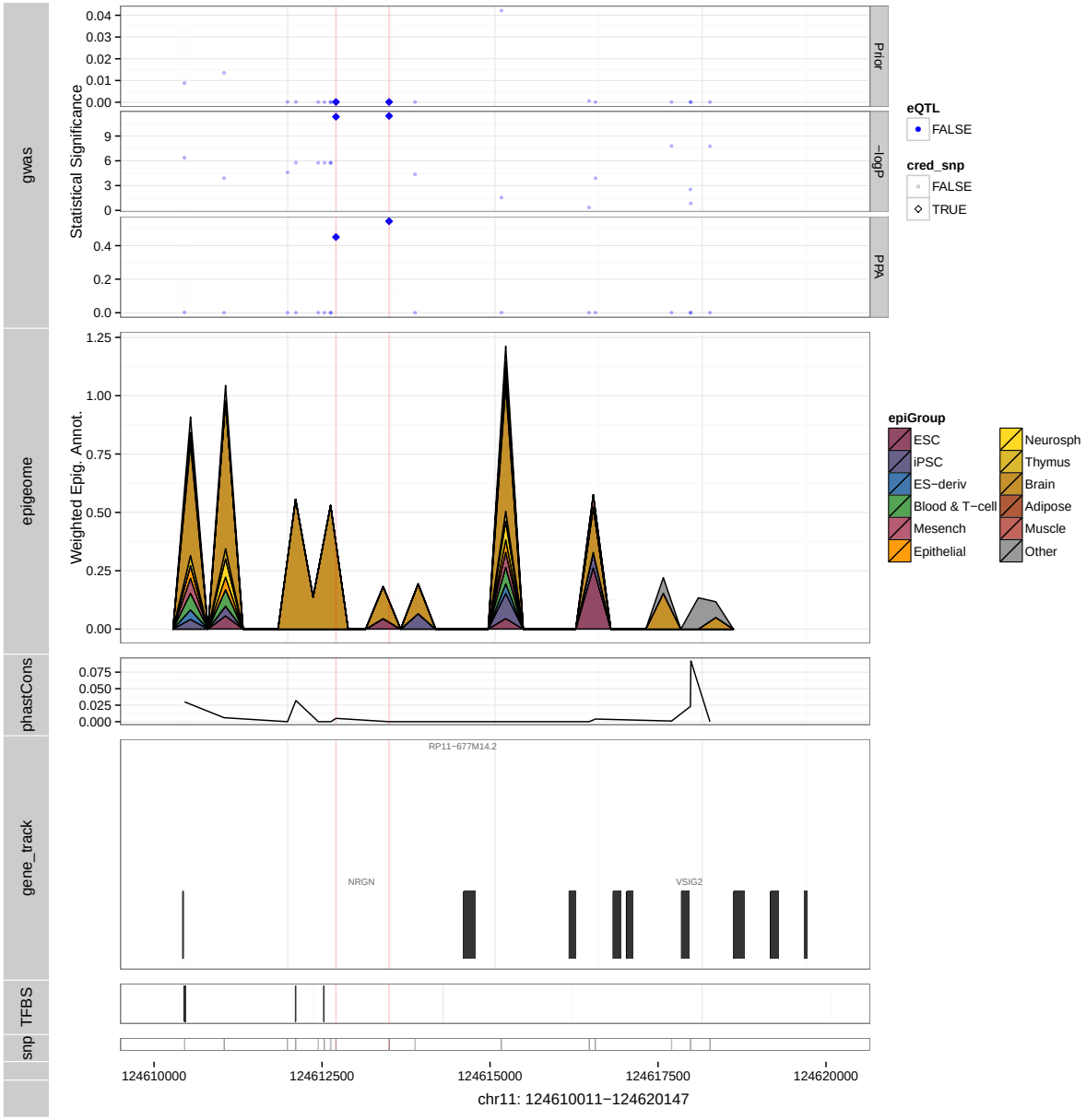
Schizophrenia



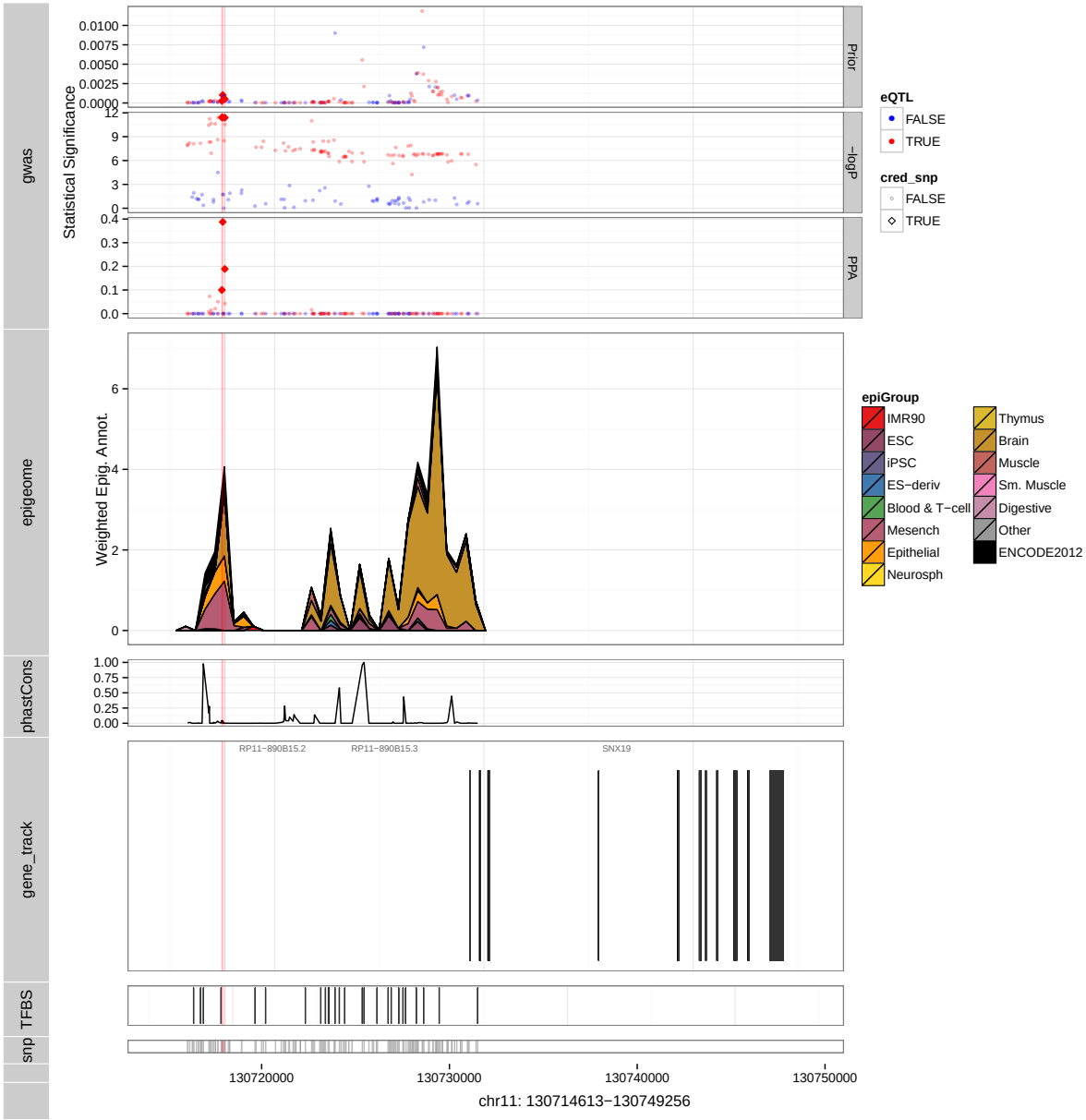
Schizophrenia



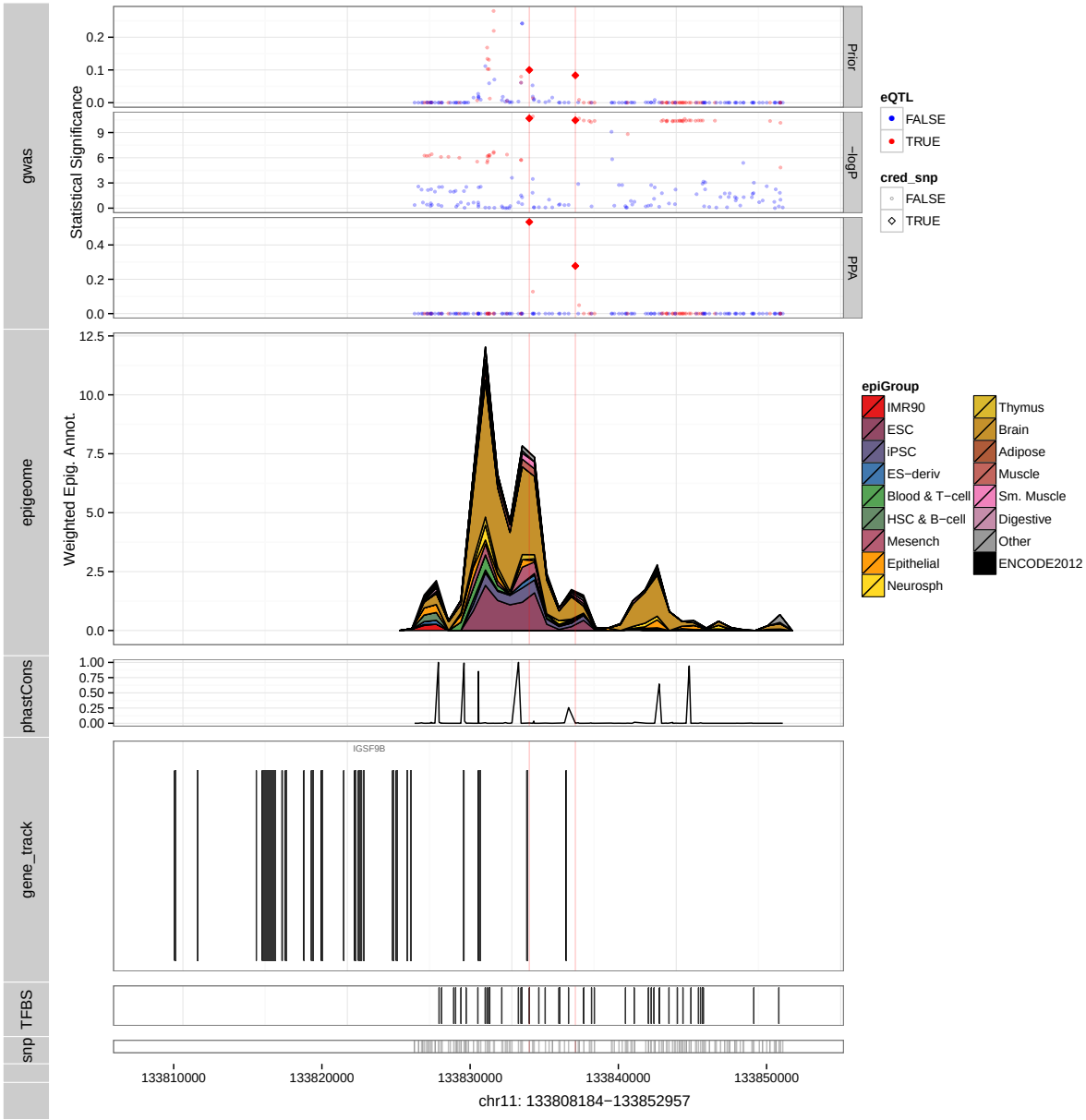
Schizophrenia



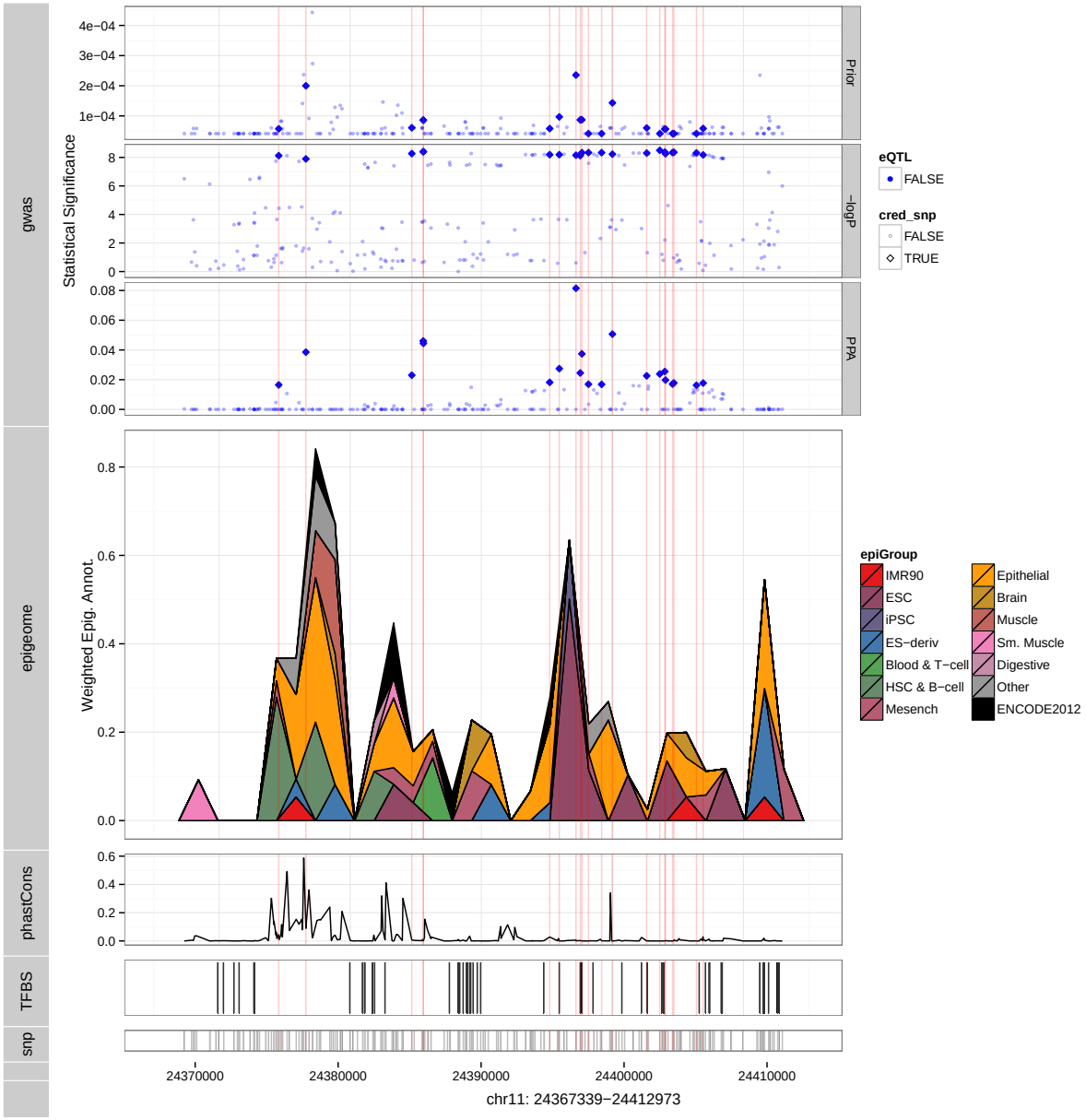
Schizophrenia



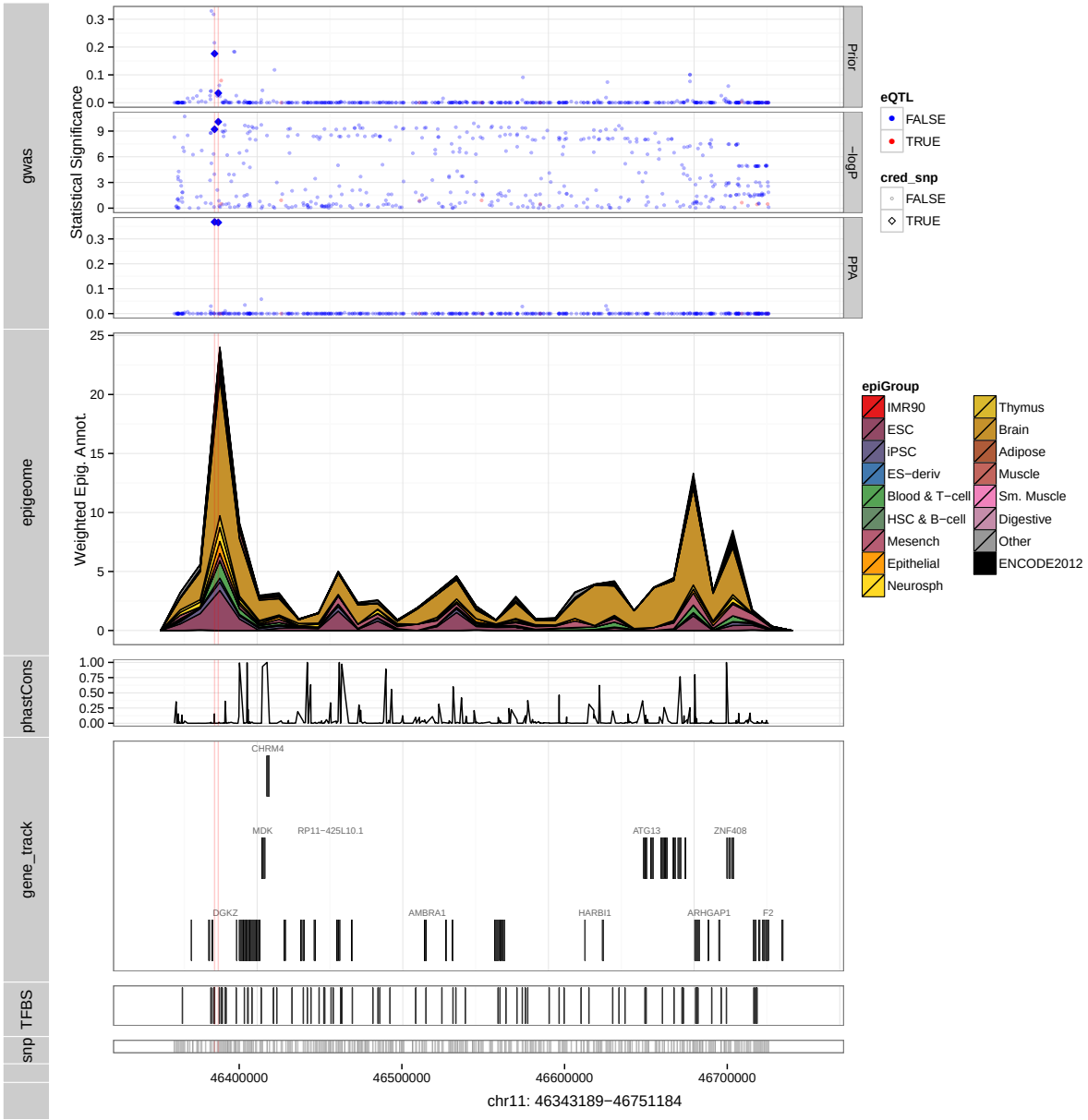
Schizophrenia



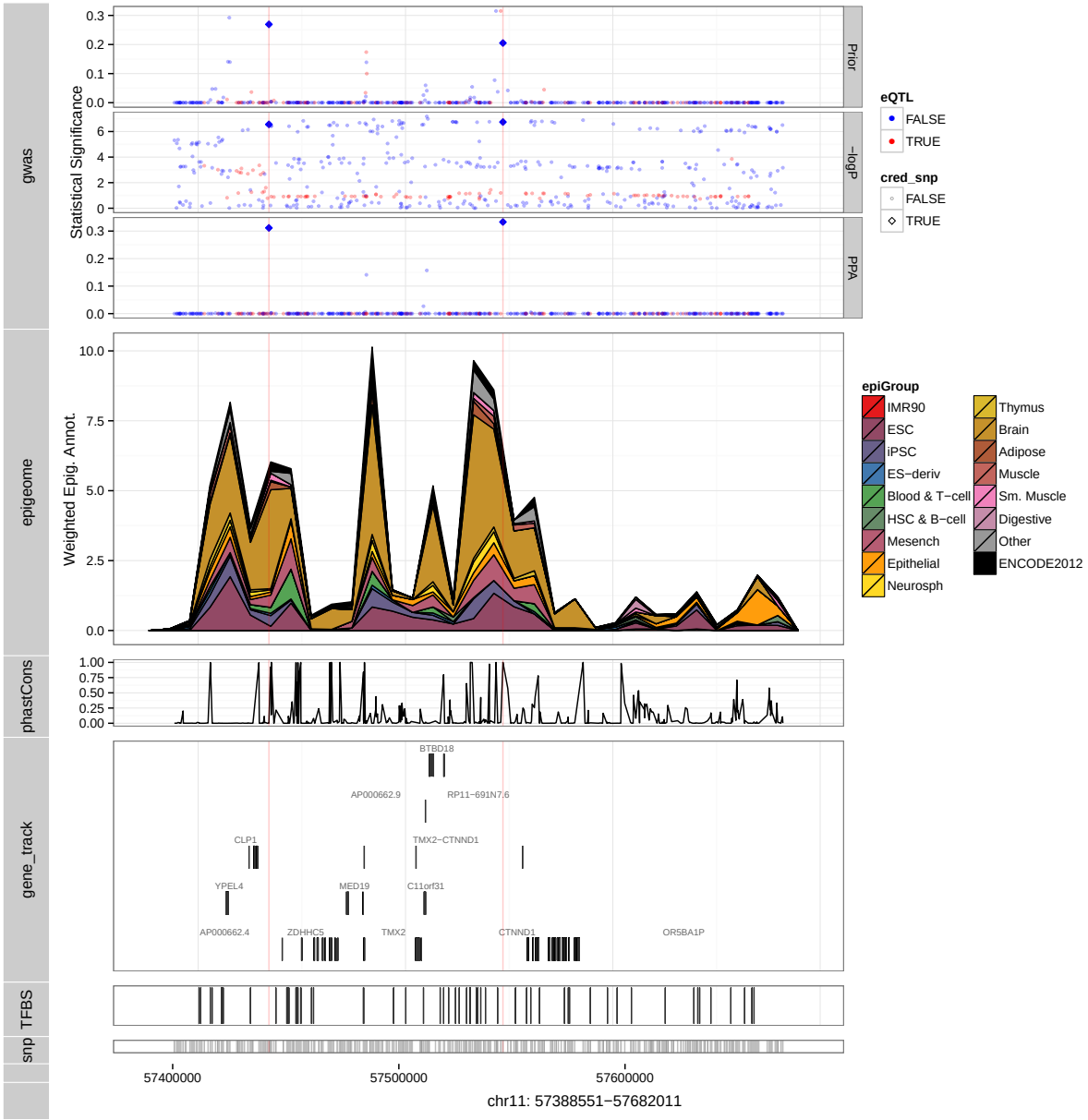
Schizophrenia



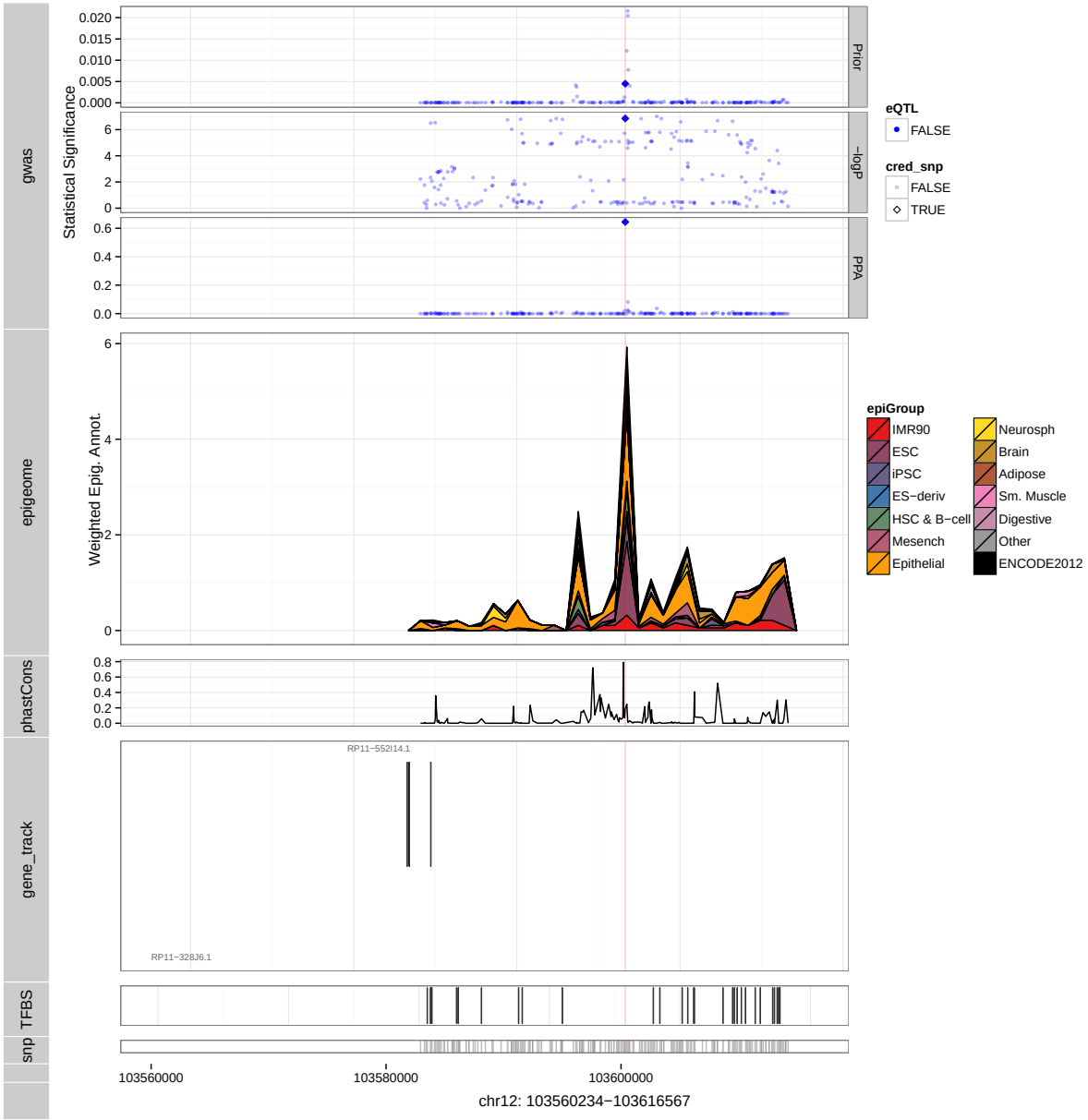
Schizophrenia



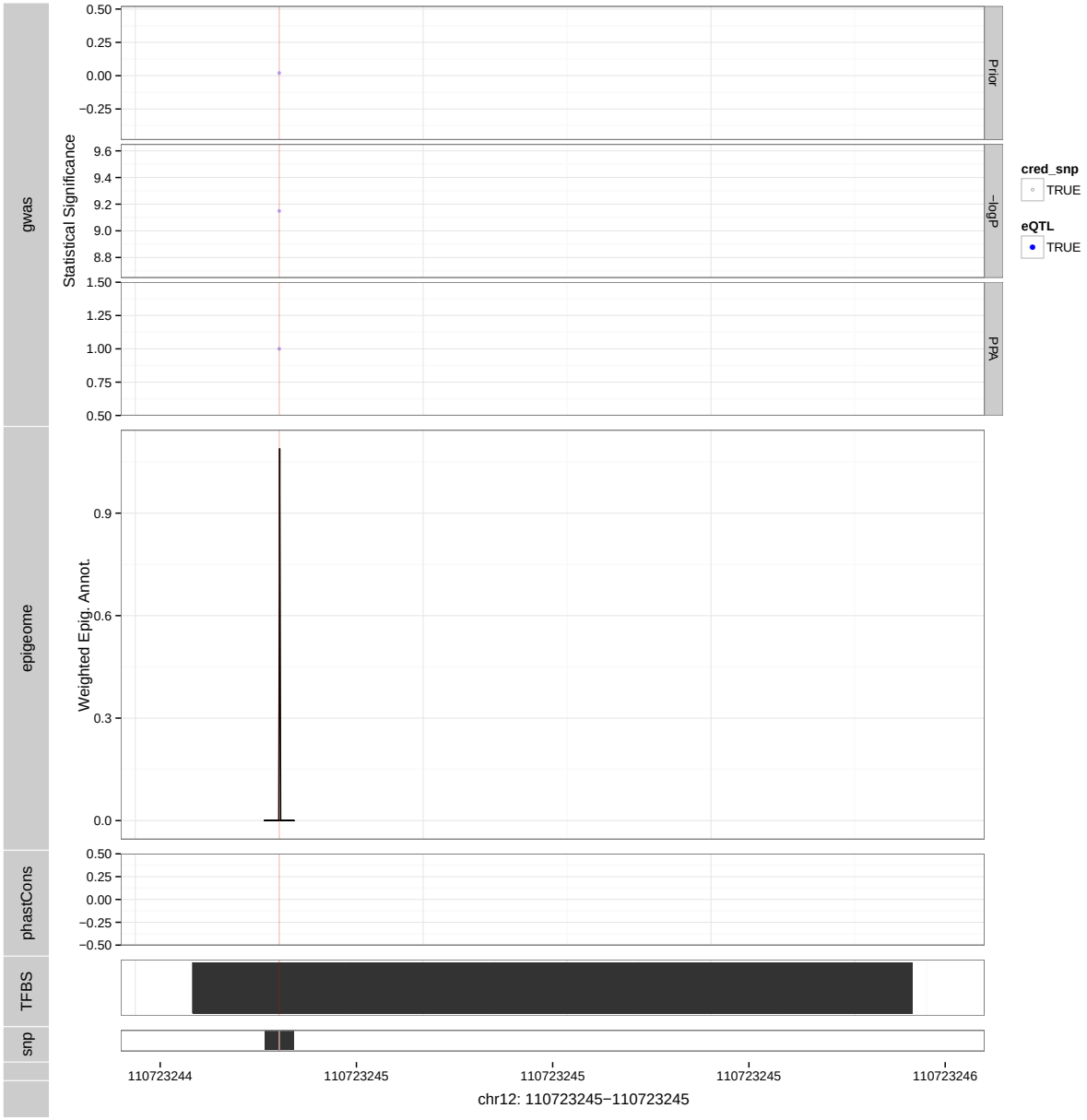
Schizophrenia



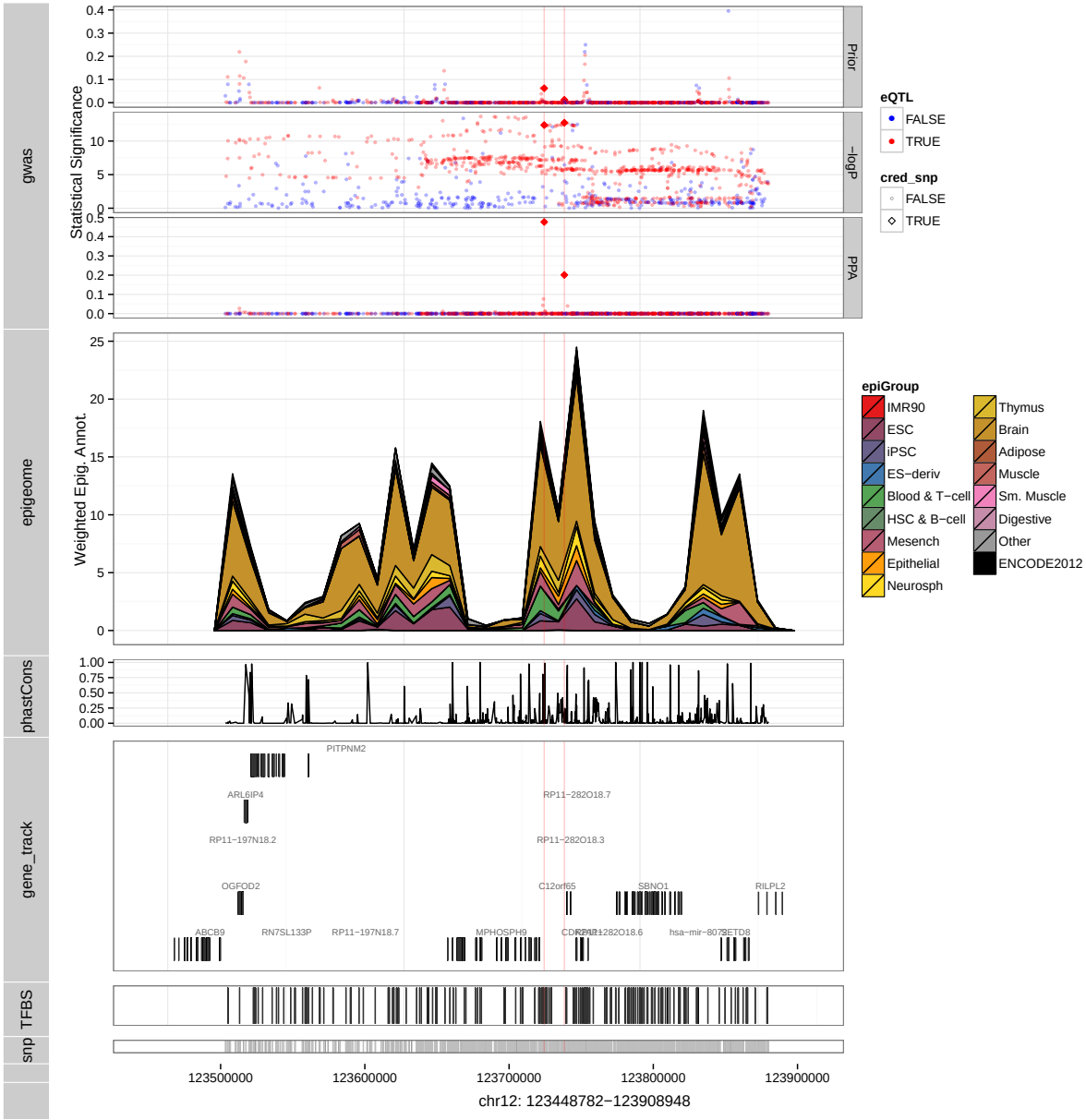
Schizophrenia



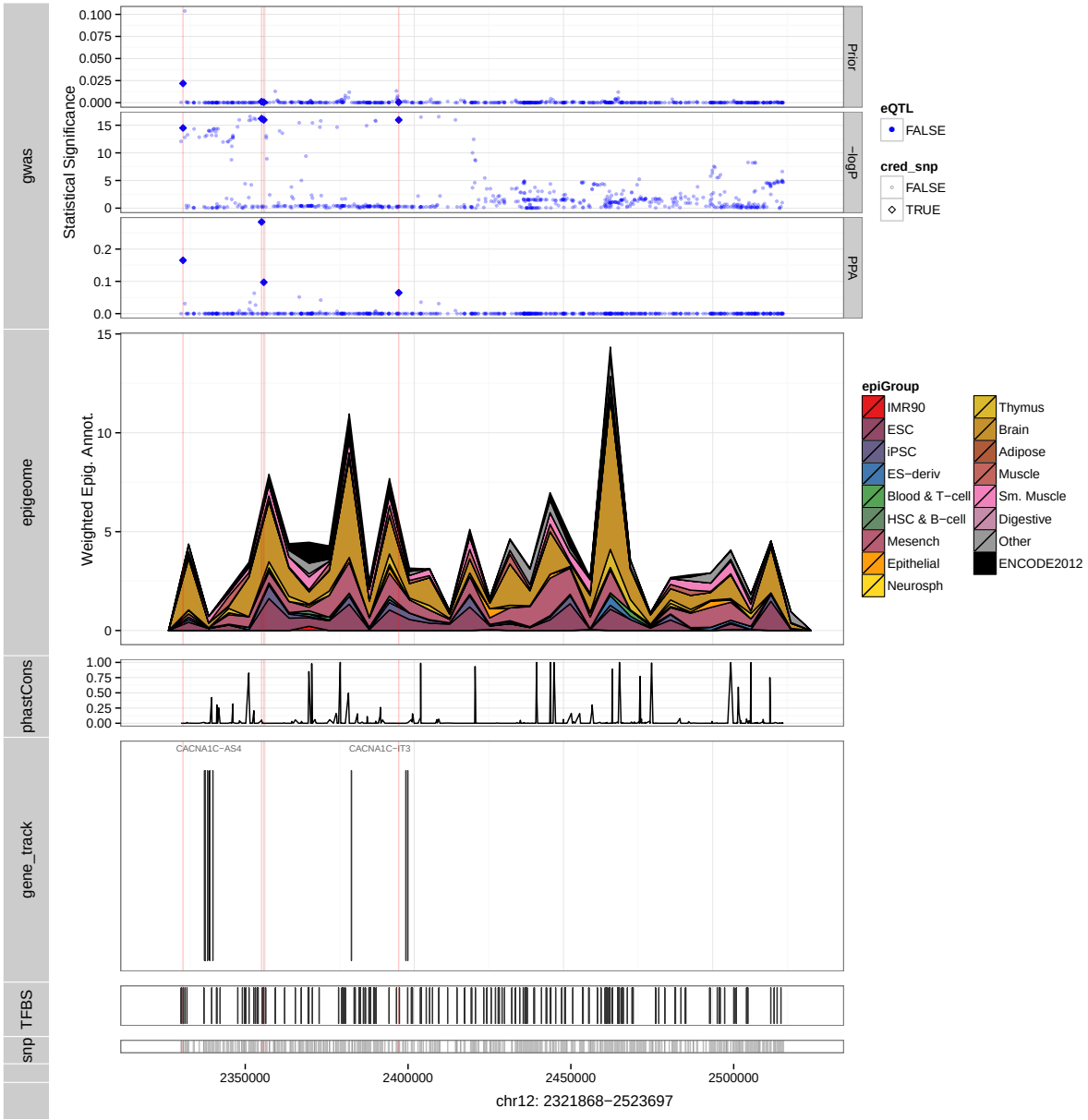
Schizophrenia



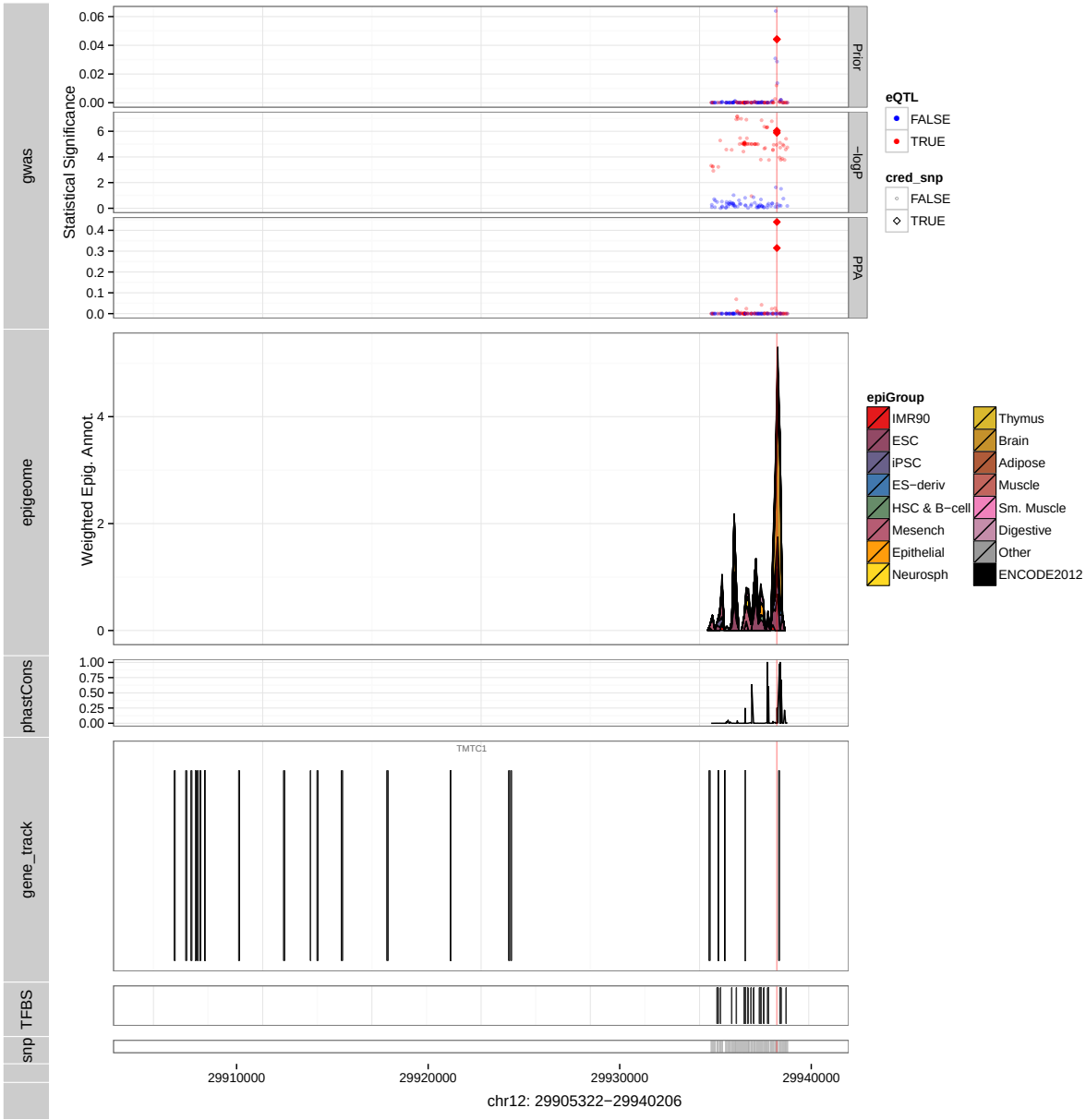
Schizophrenia



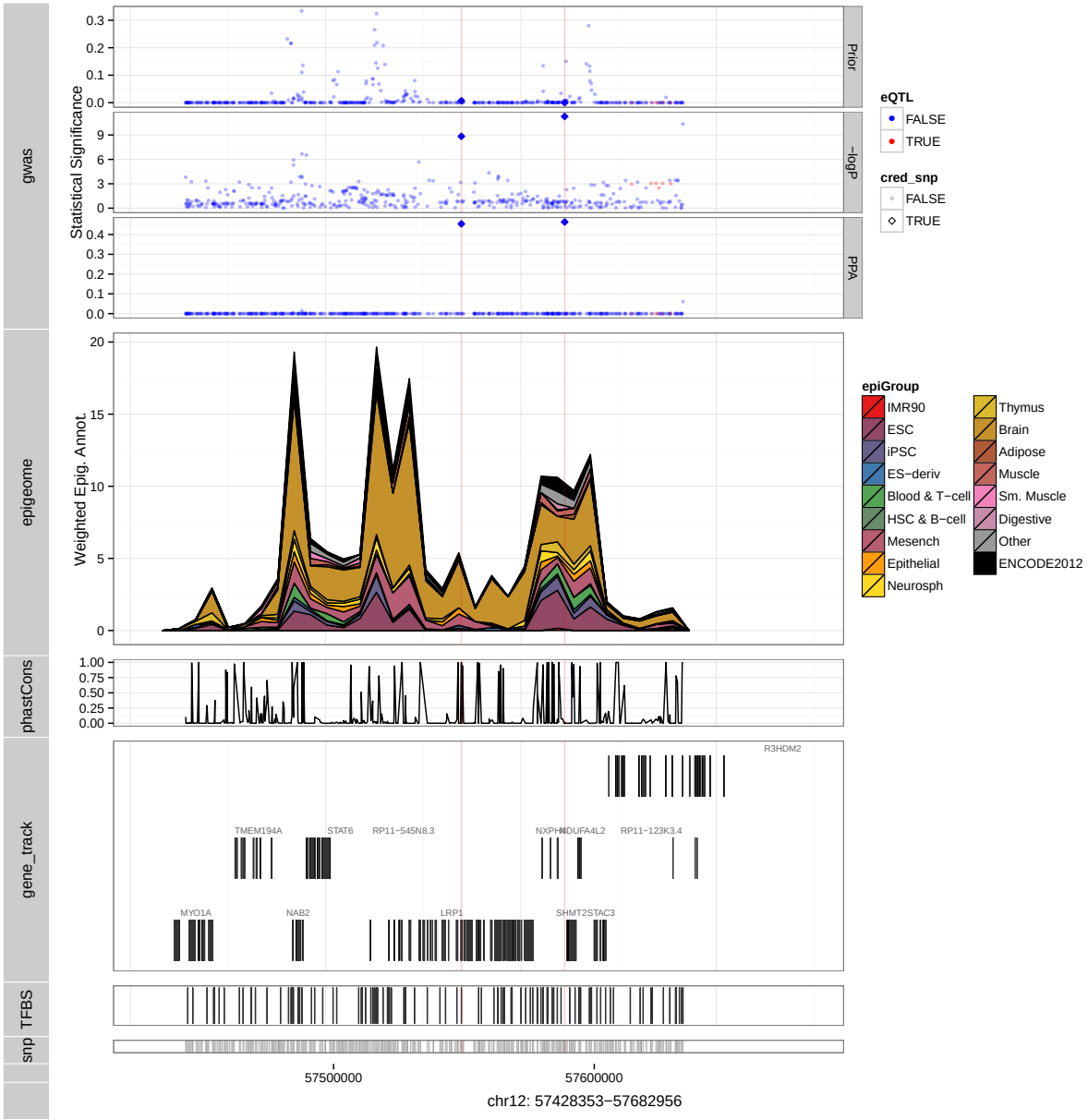
Schizophrenia



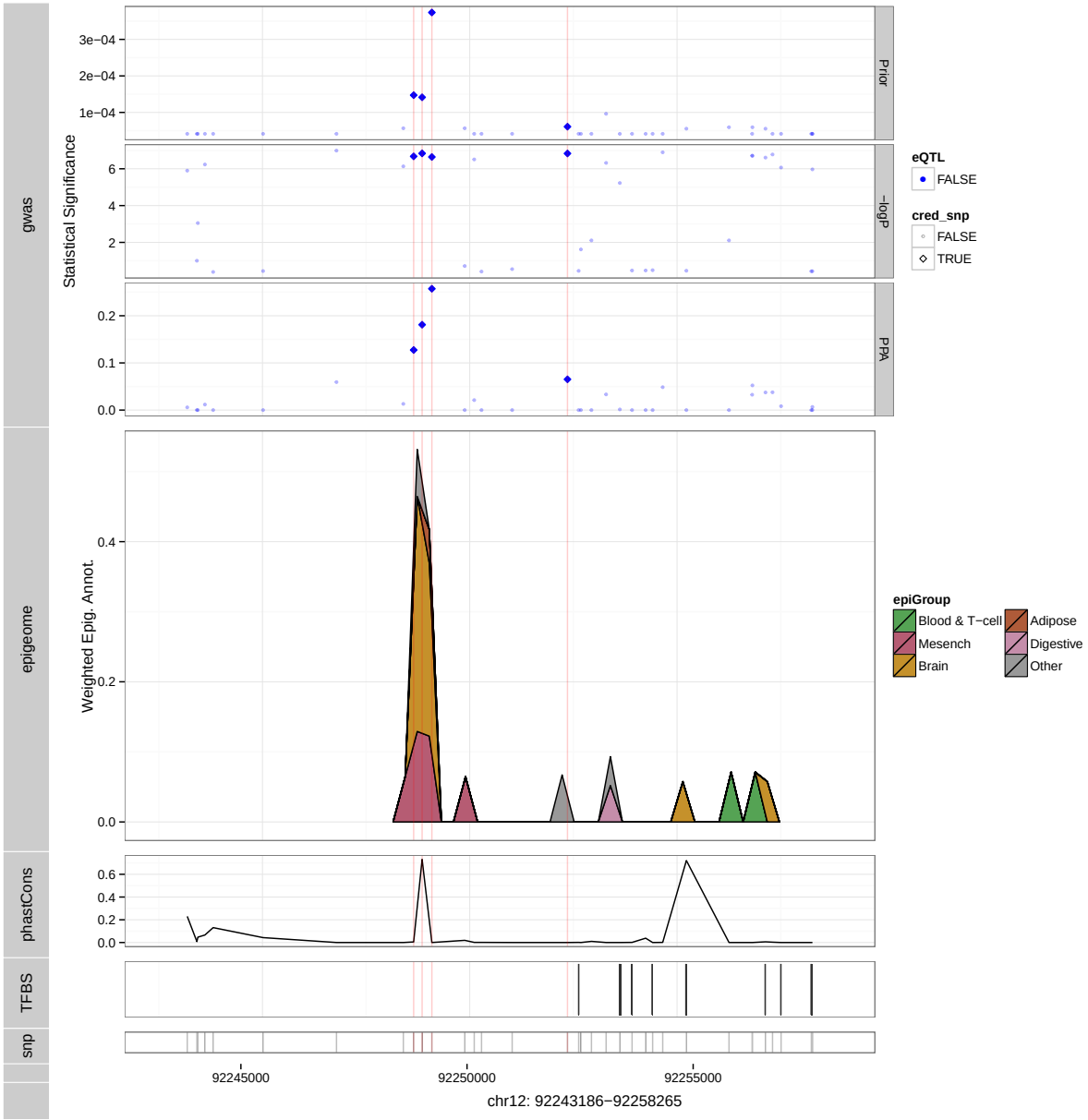
Schizophrenia



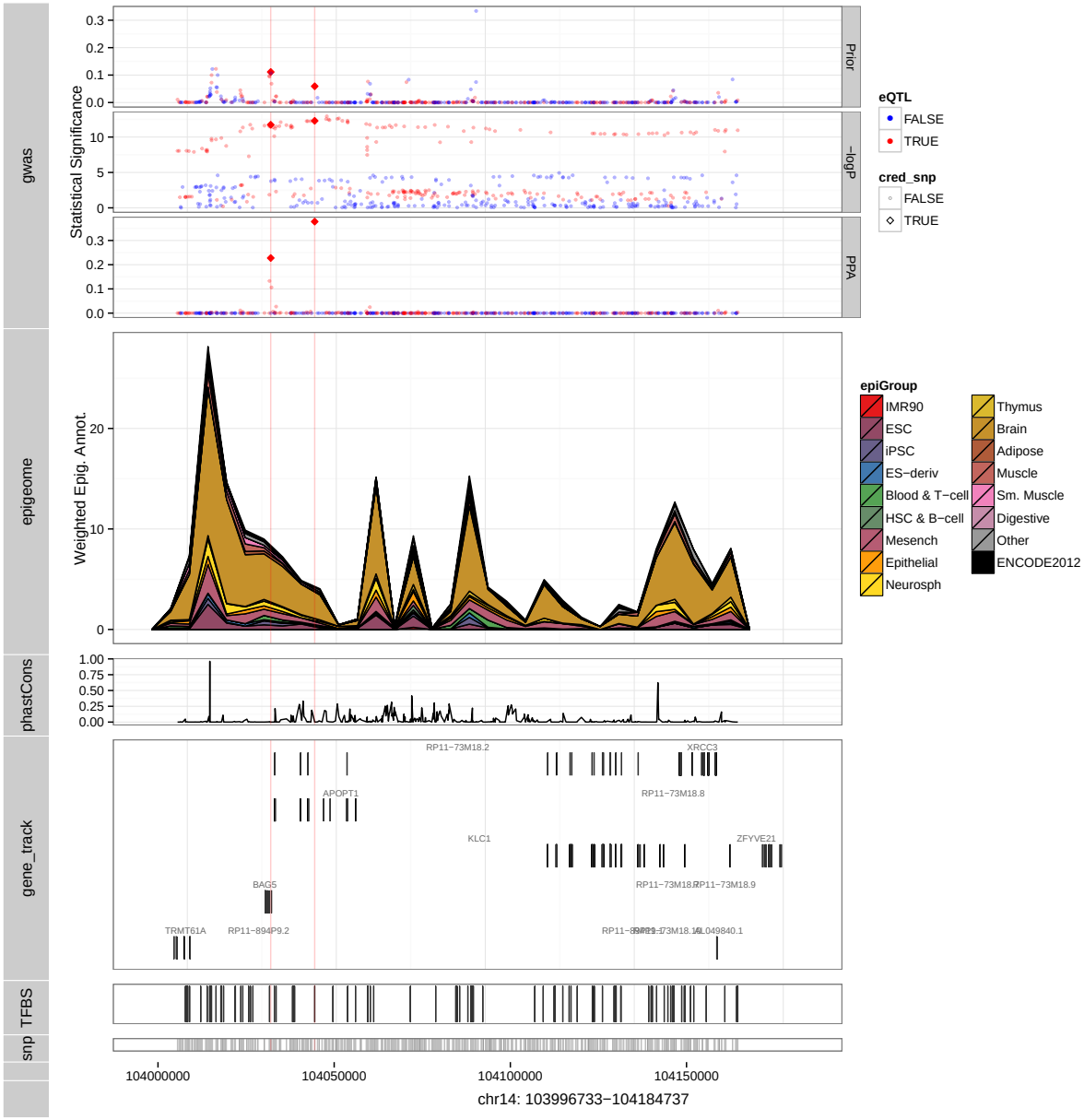
Schizophrenia

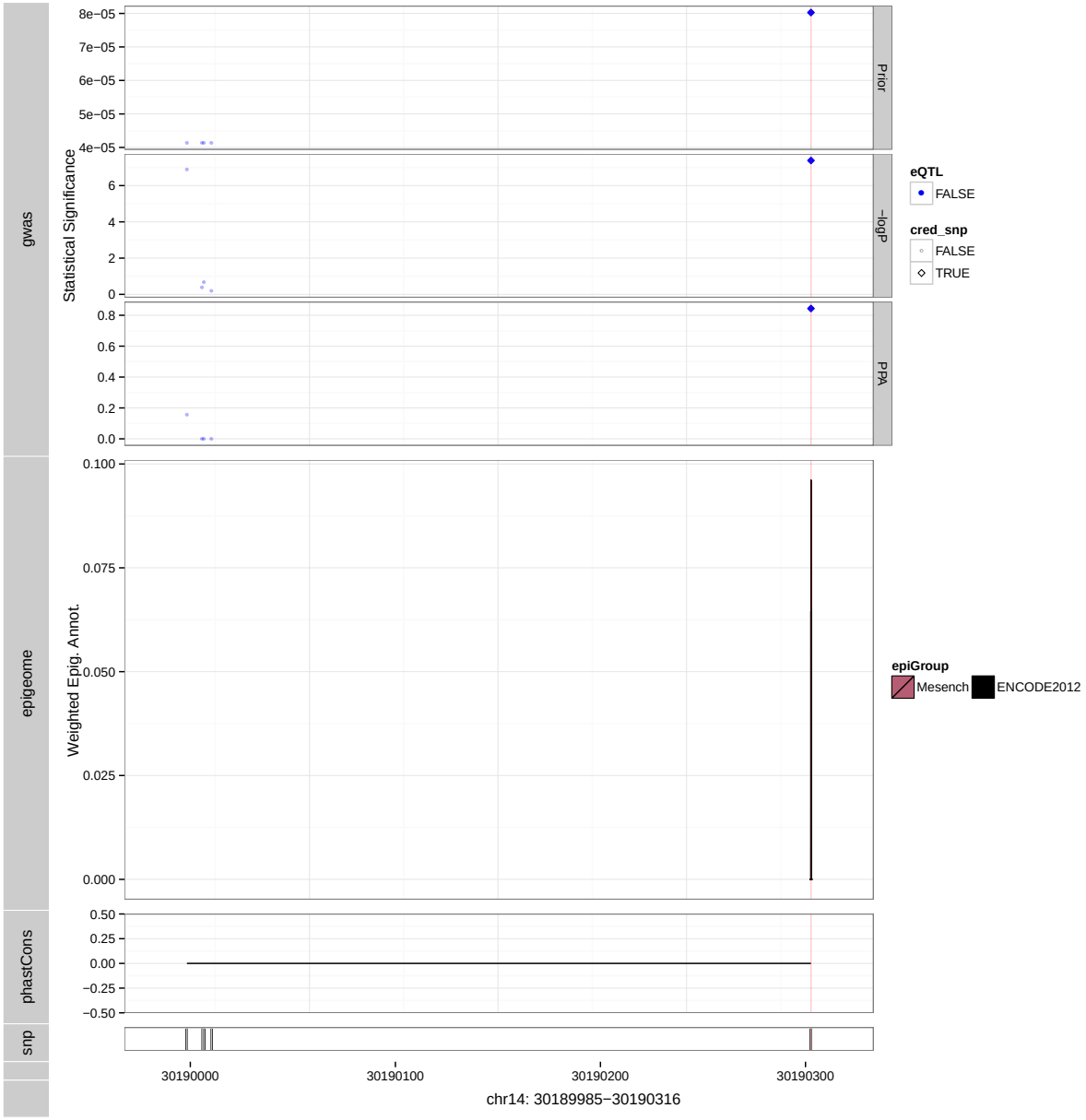


Schizophrenia

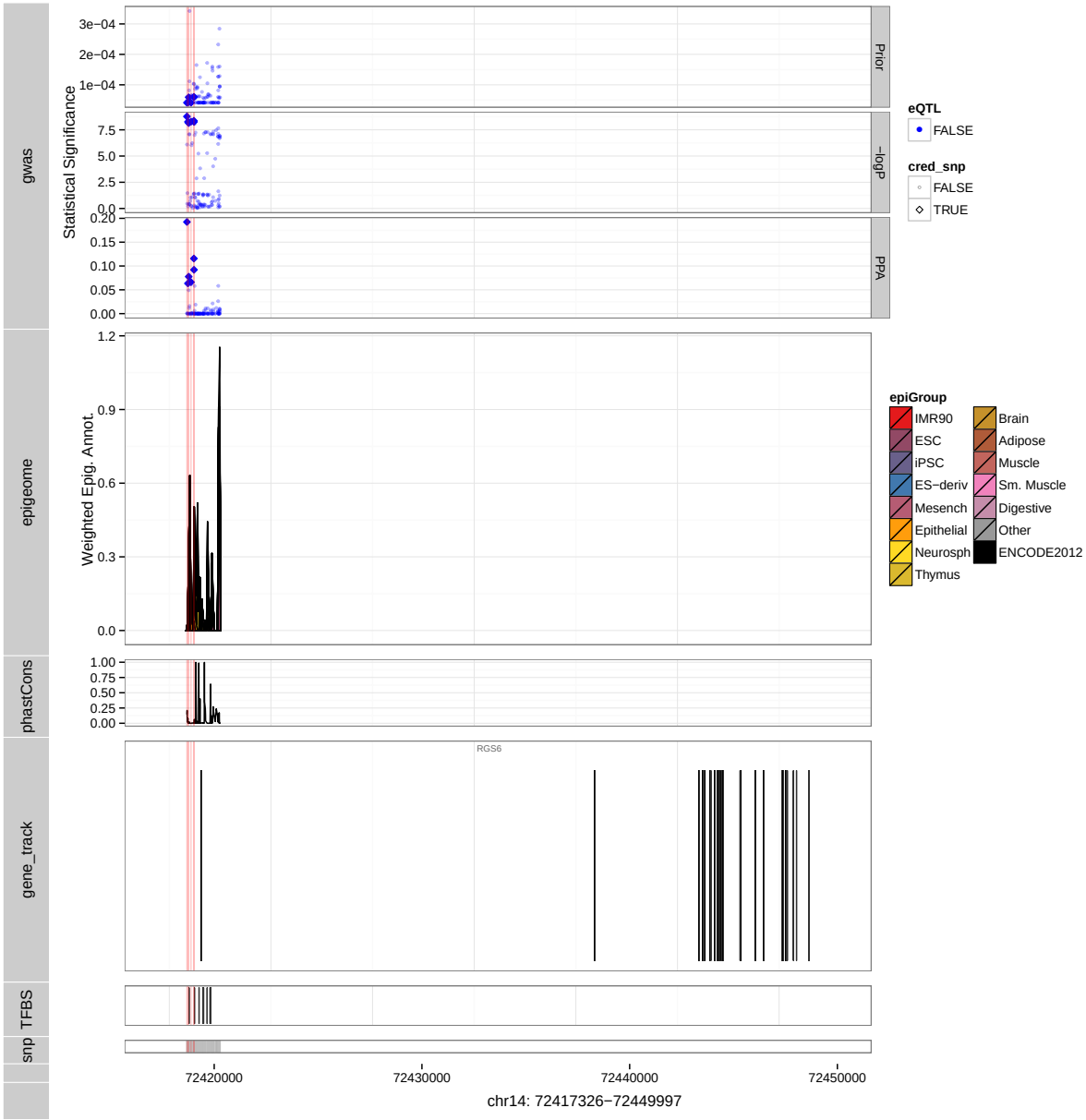


Schizophrenia

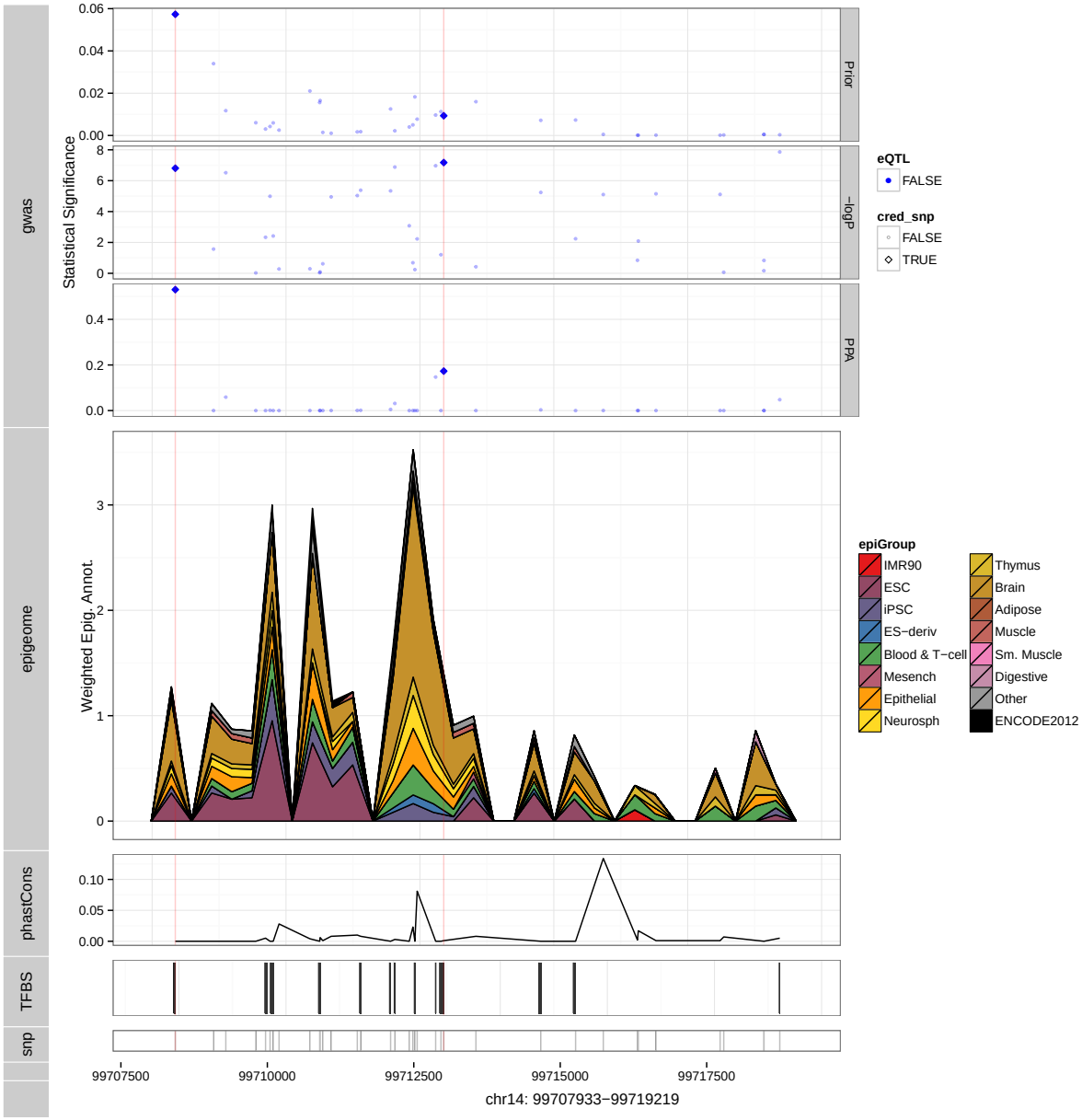



gwas
epigeome
phastCons
snp
Statistical Significance
Prior
−logP
PPA
8e−05
7e−05
6e−05
5e−05
4e−05
6
4
2
0
0.8
0.6
0.4
0.2
0.0
eQTL
FALSE
cred_snp
FALSE
TRUE
Weighted Epig. Annot.
0.100
0.075
0.050
0.025
0.000
epiGroup
Mesench
ENCODE2012
0.50
0.25
0.00
−0.25
−0.50
30190000
30190100
30190200
30190300
chr14: 30189985−30190316

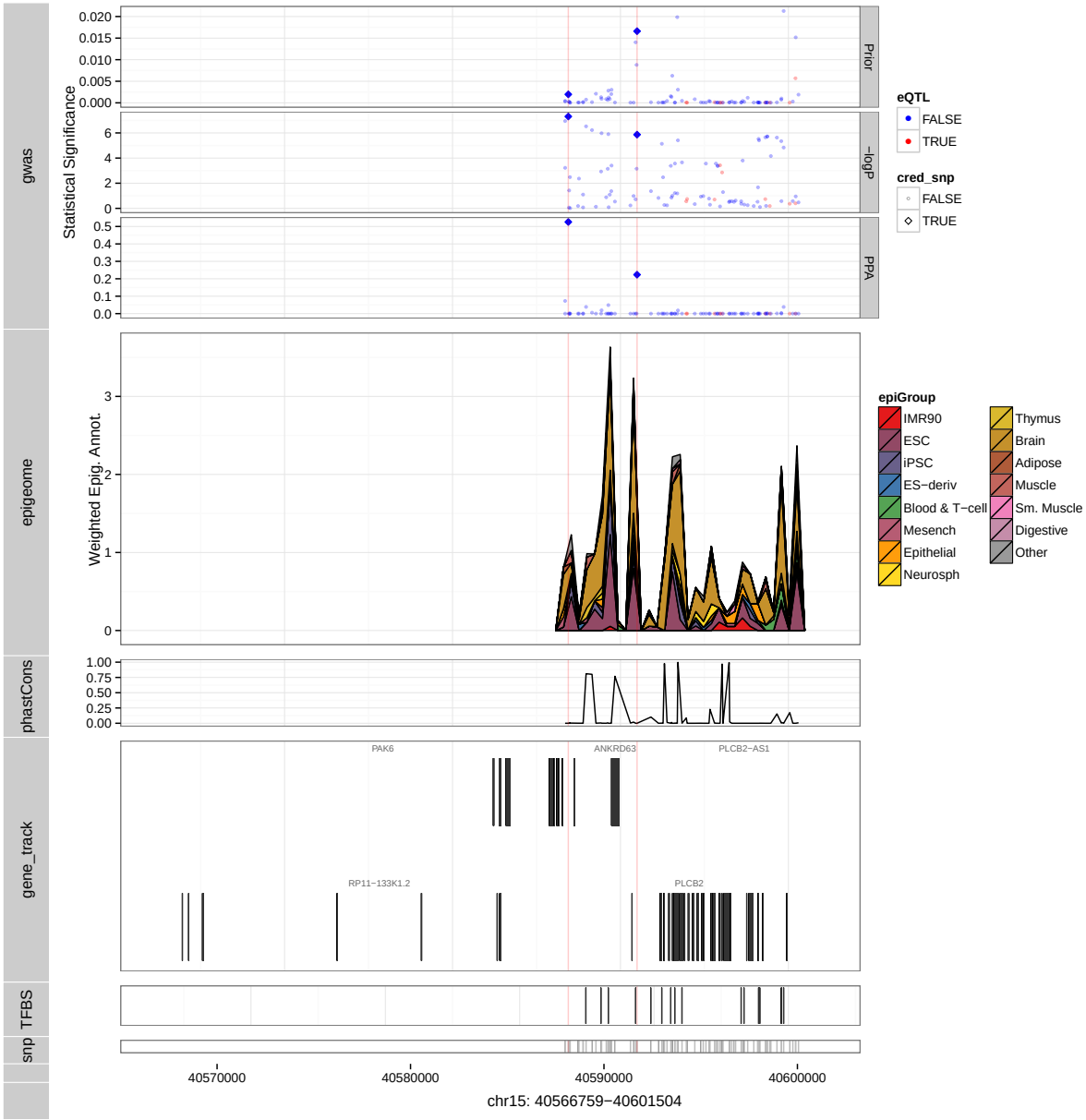
Schizophrenia



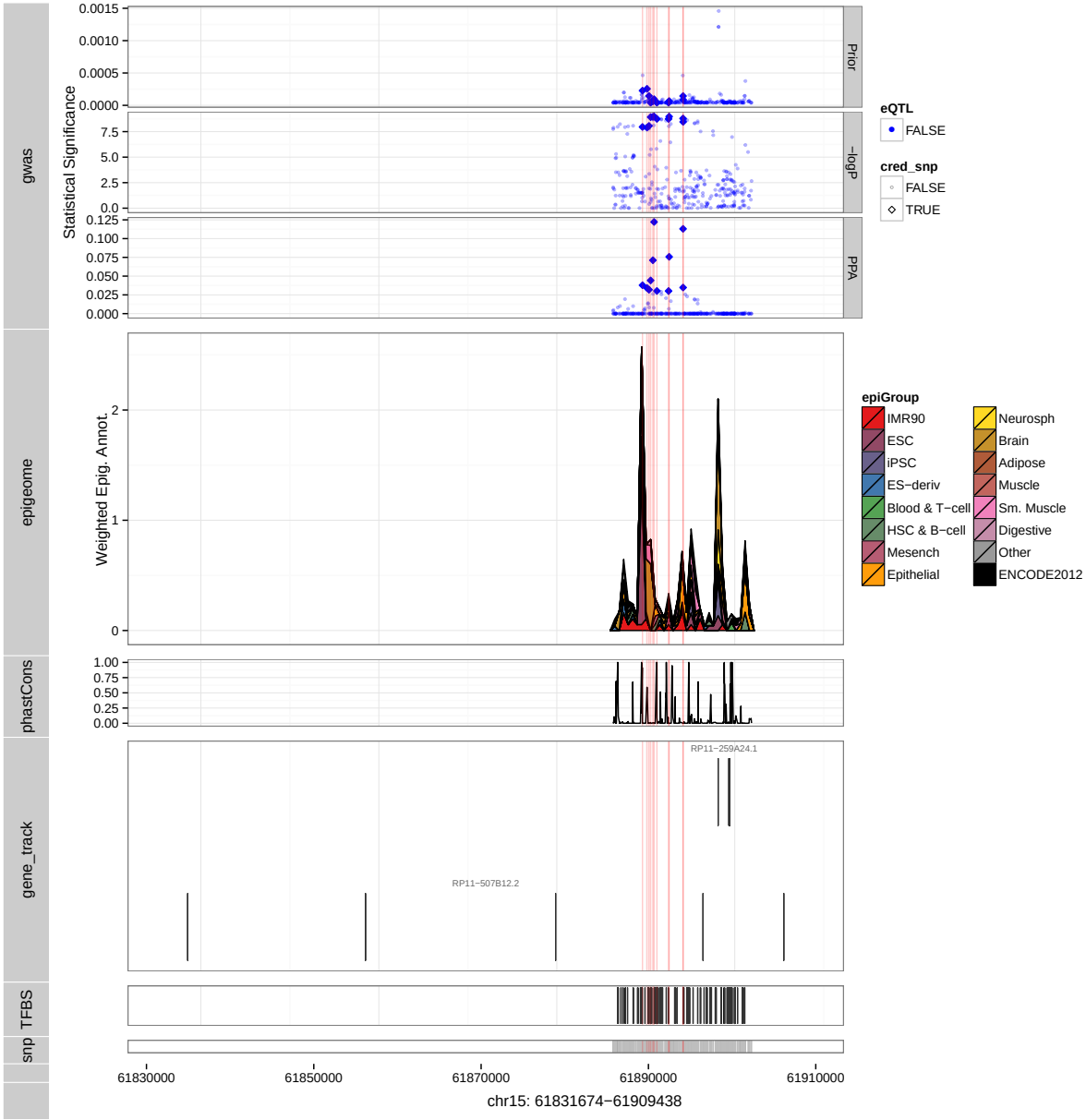
Schizophrenia



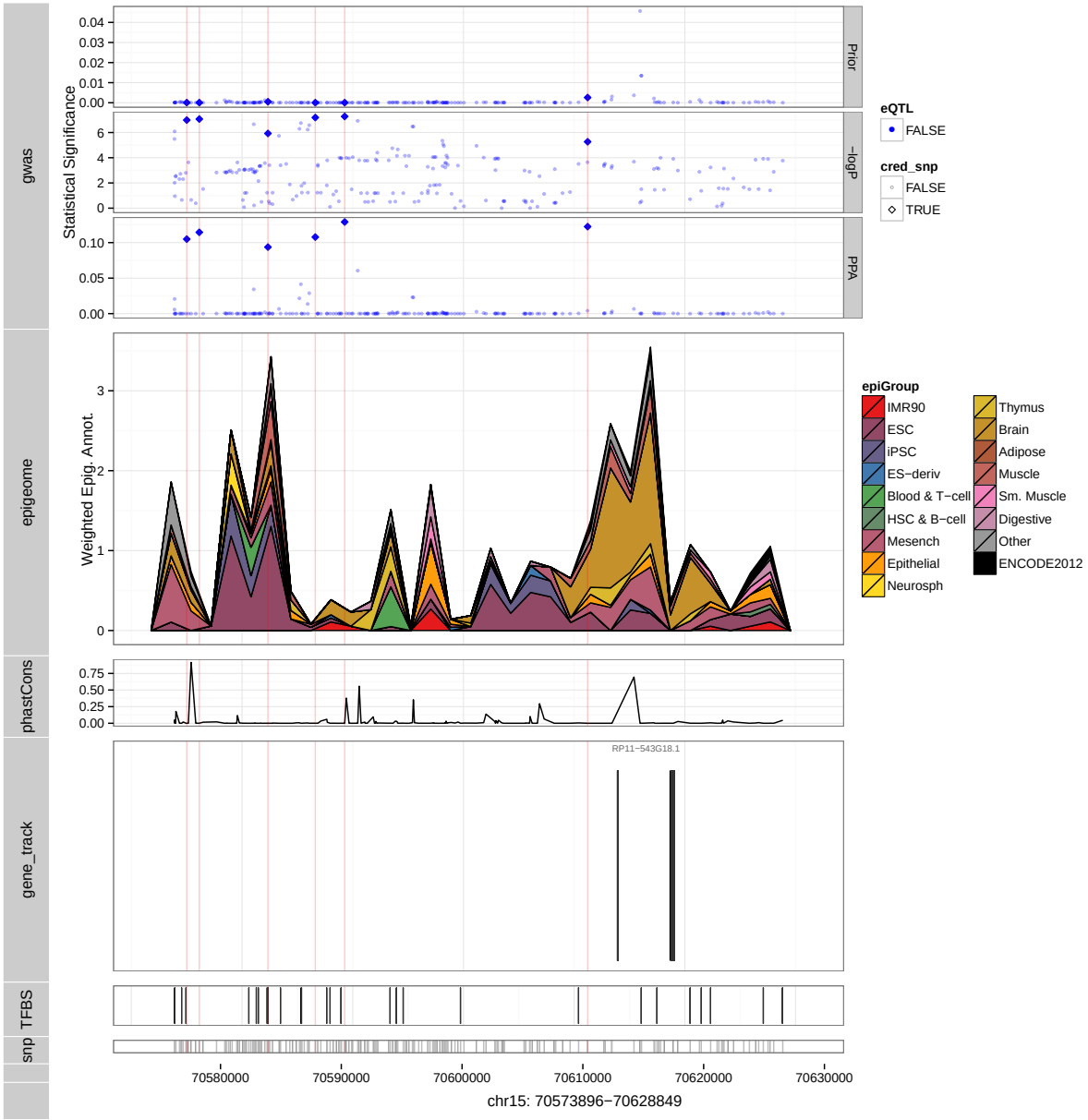
Schizophrenia



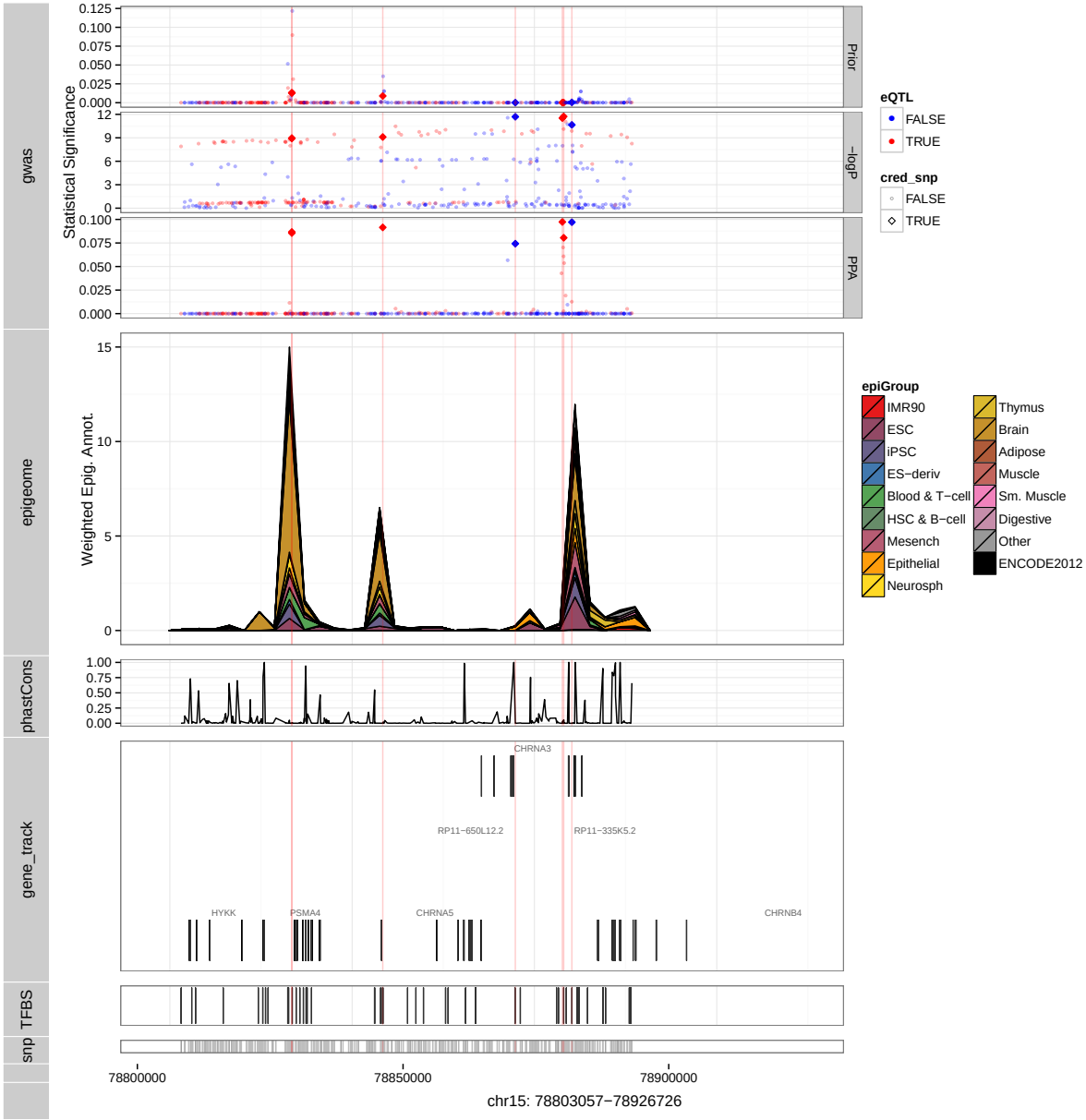
Schizophrenia



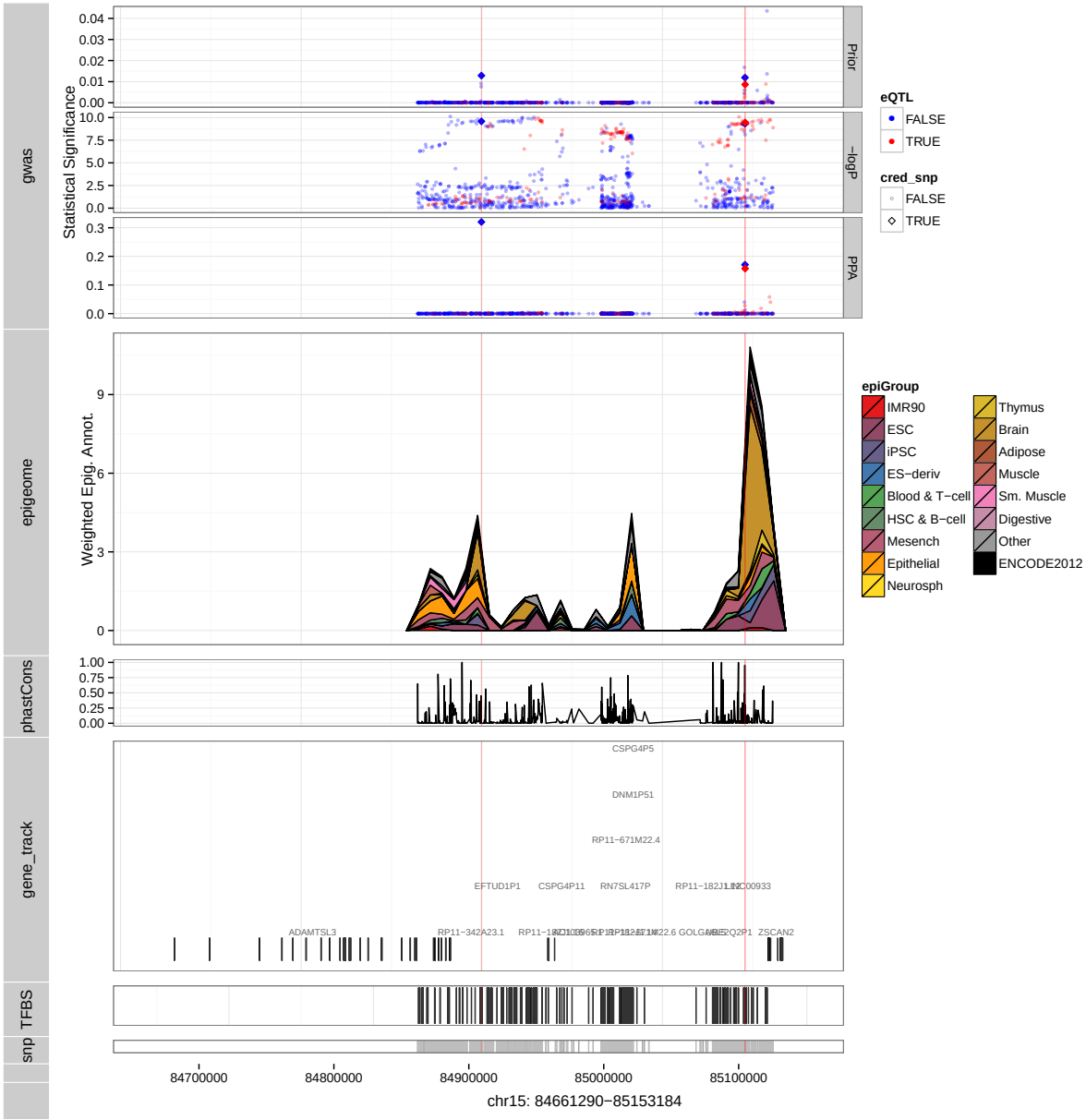
Schizophrenia



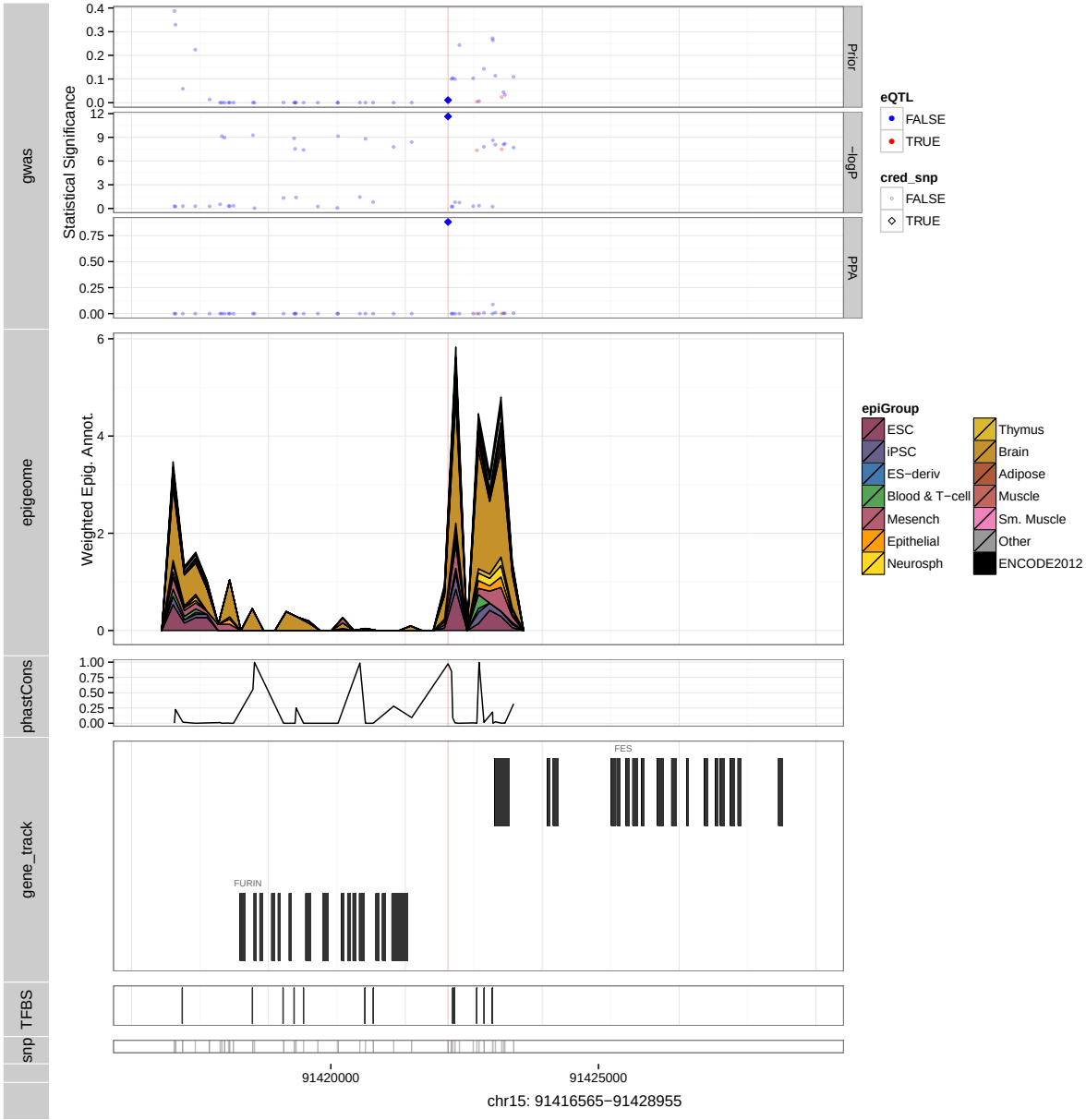
Schizophrenia



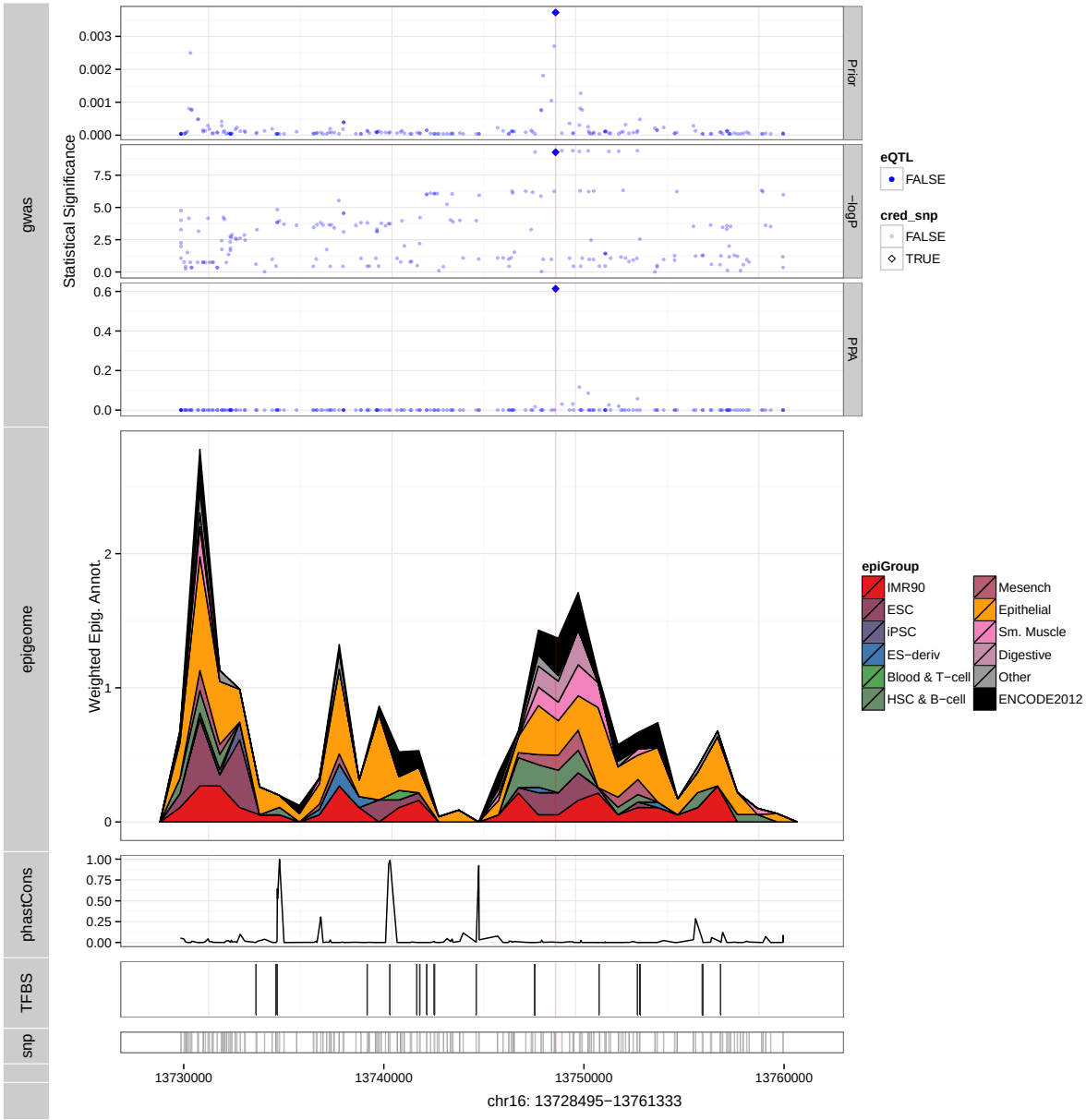
Schizophrenia



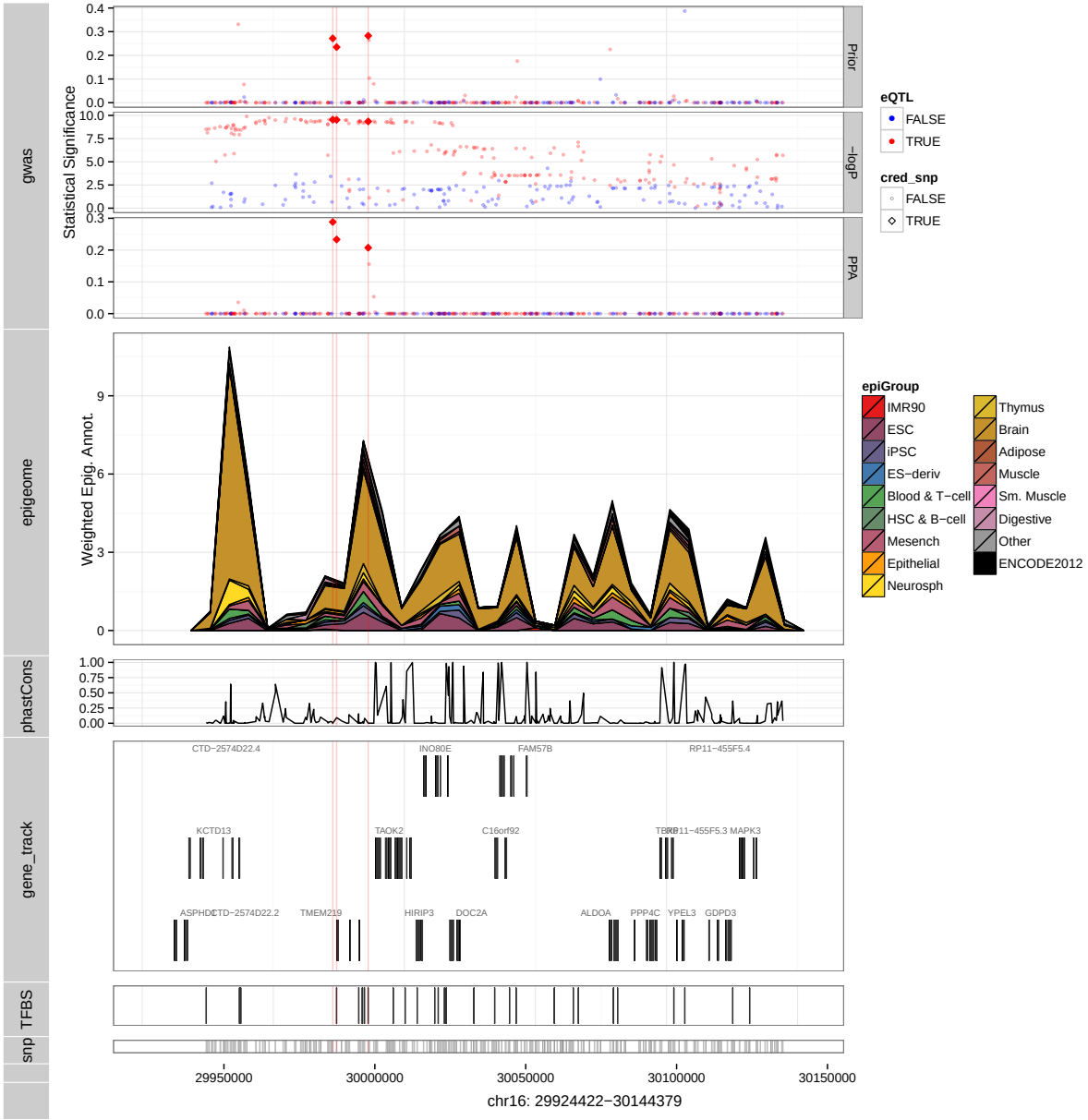
Schizophrenia



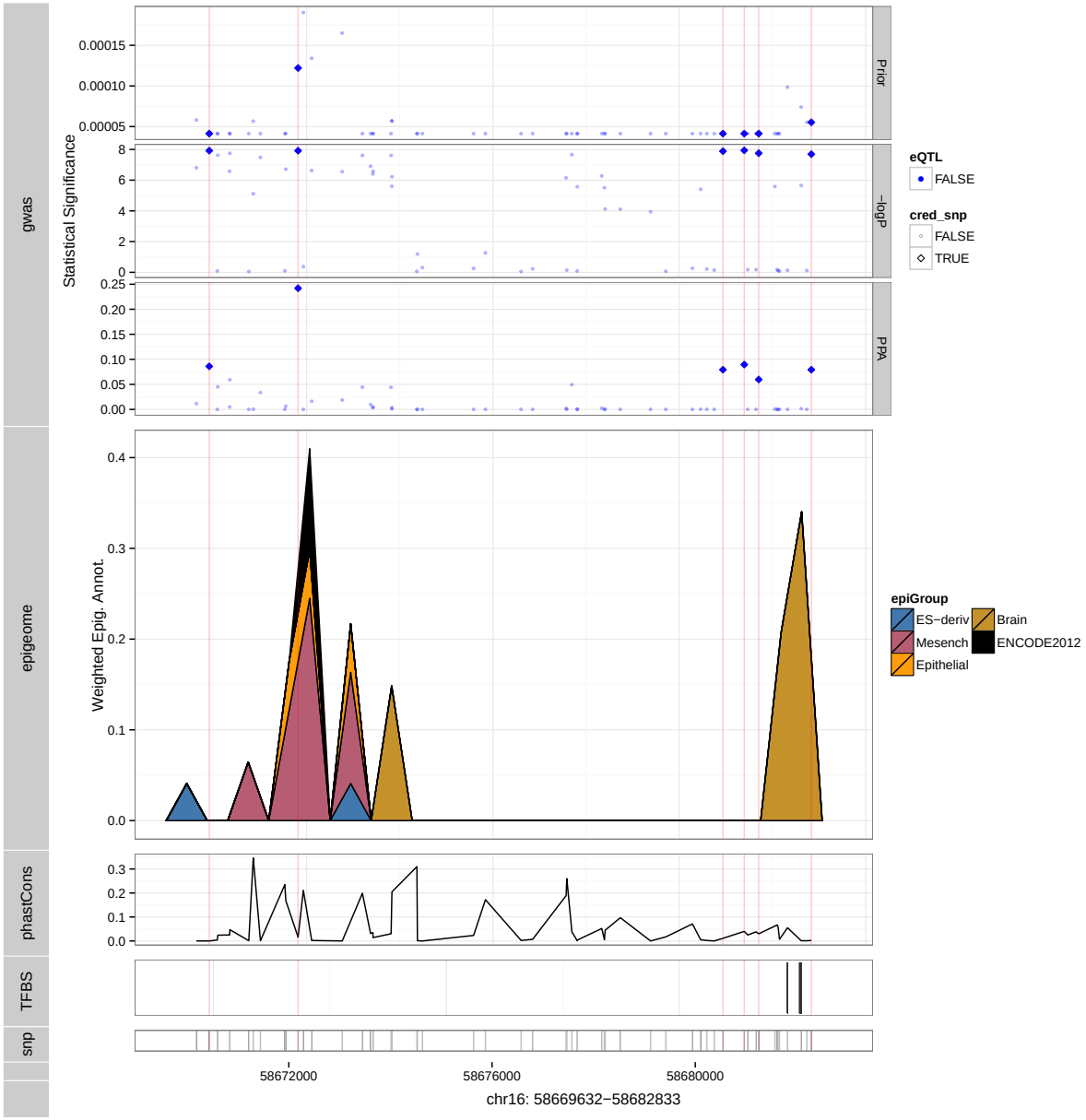
Schizophrenia



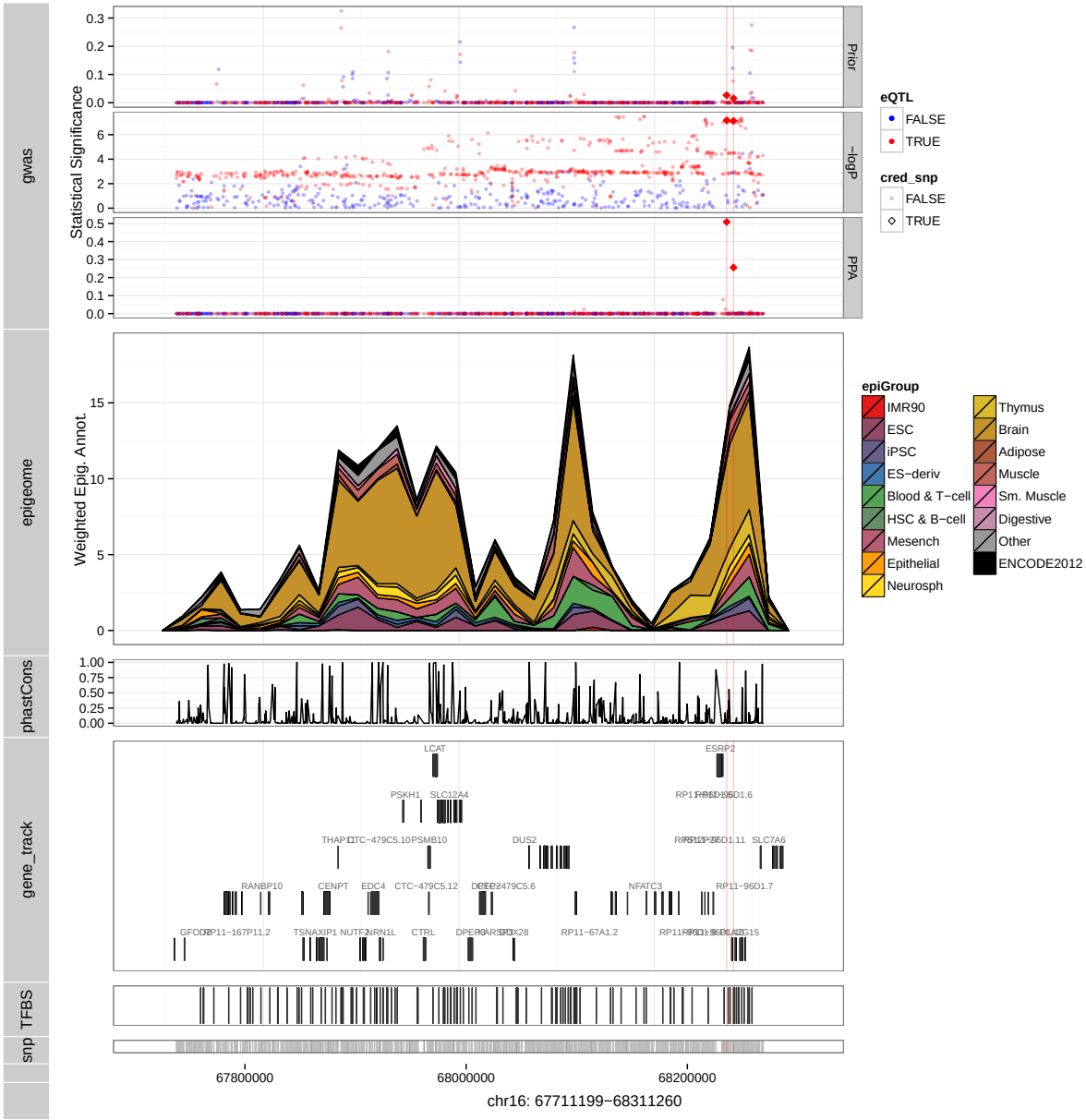
Schizophrenia



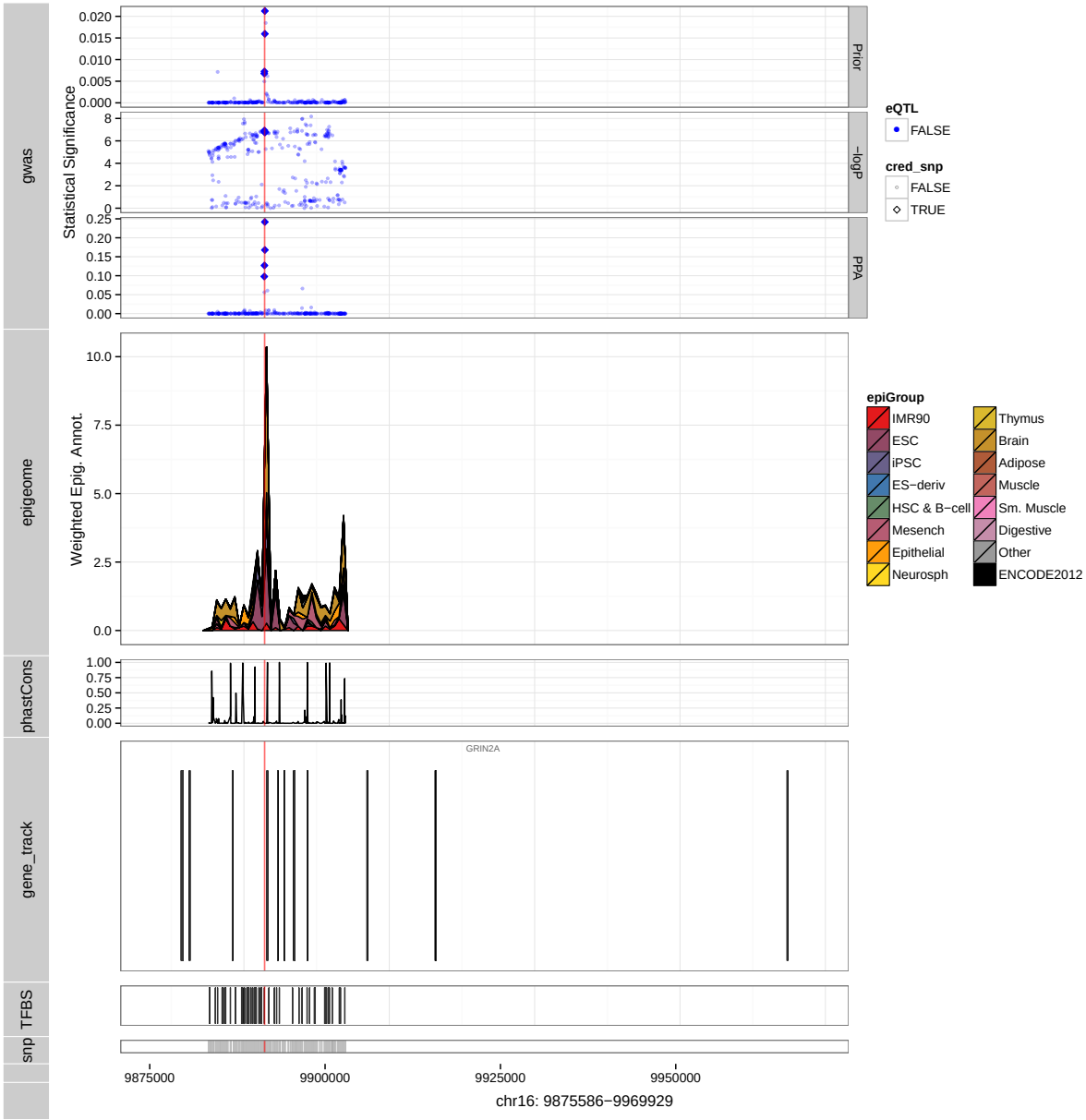
Schizophrenia



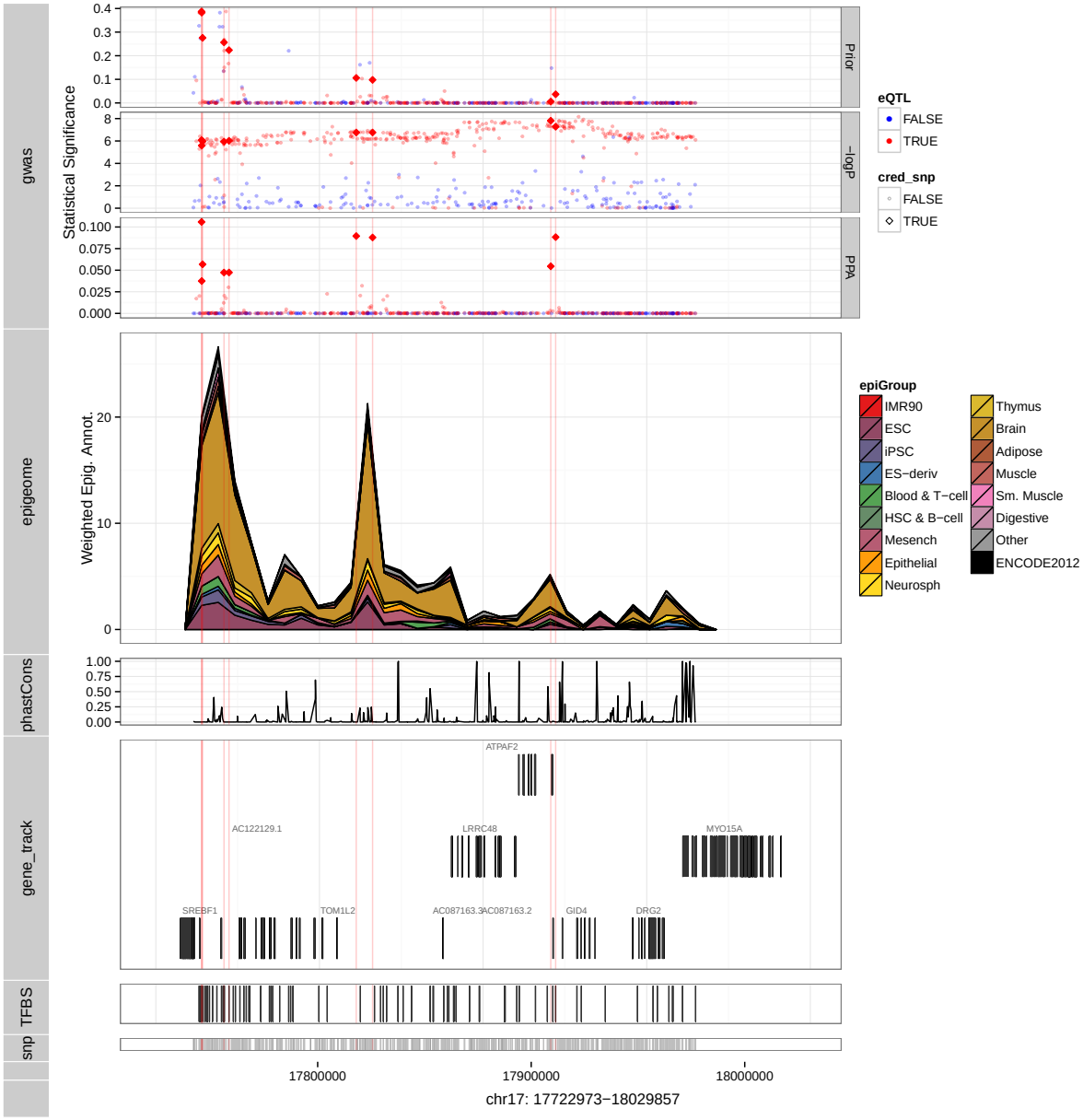
Schizophrenia



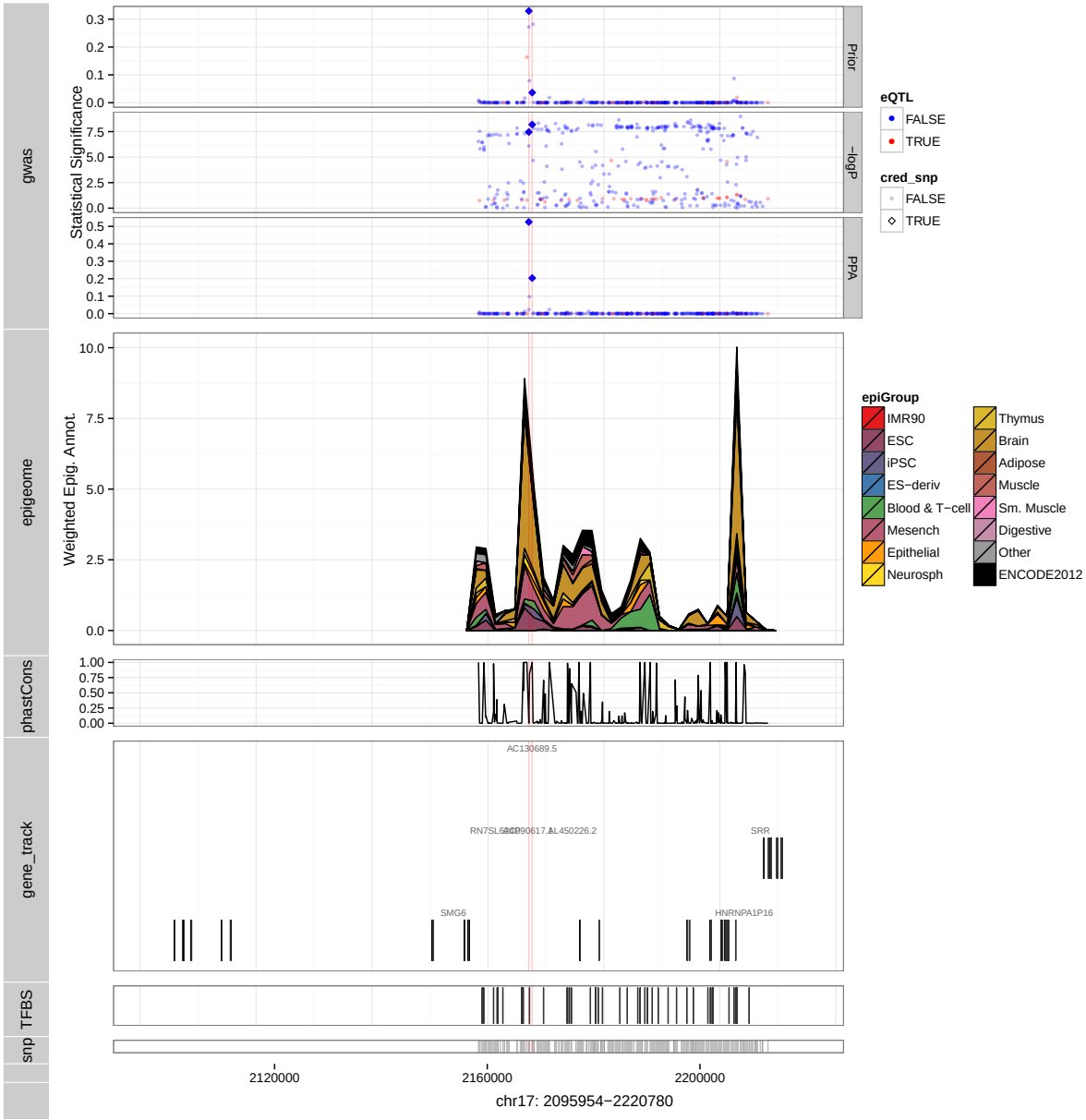
Schizophrenia



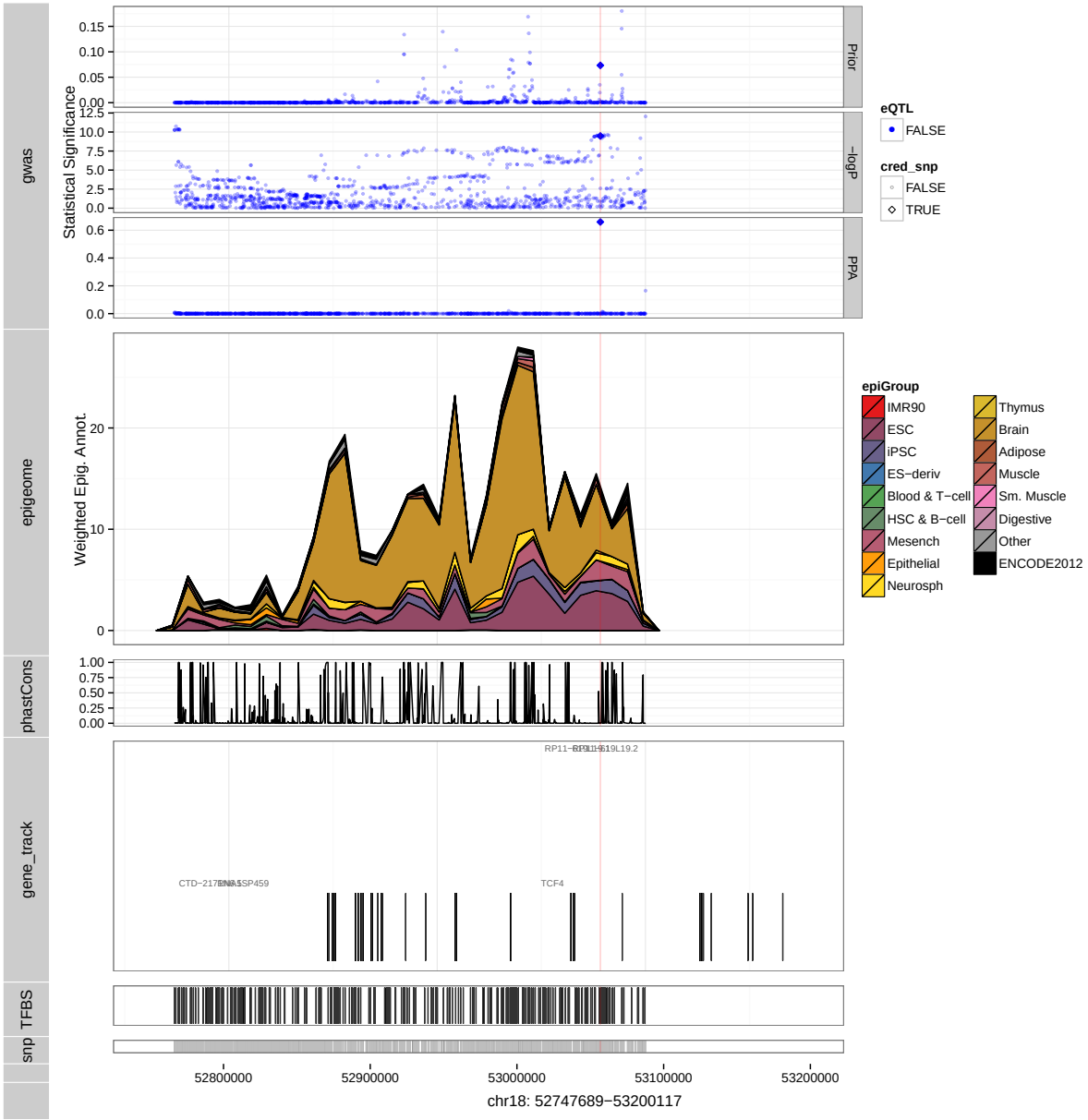
Schizophrenia



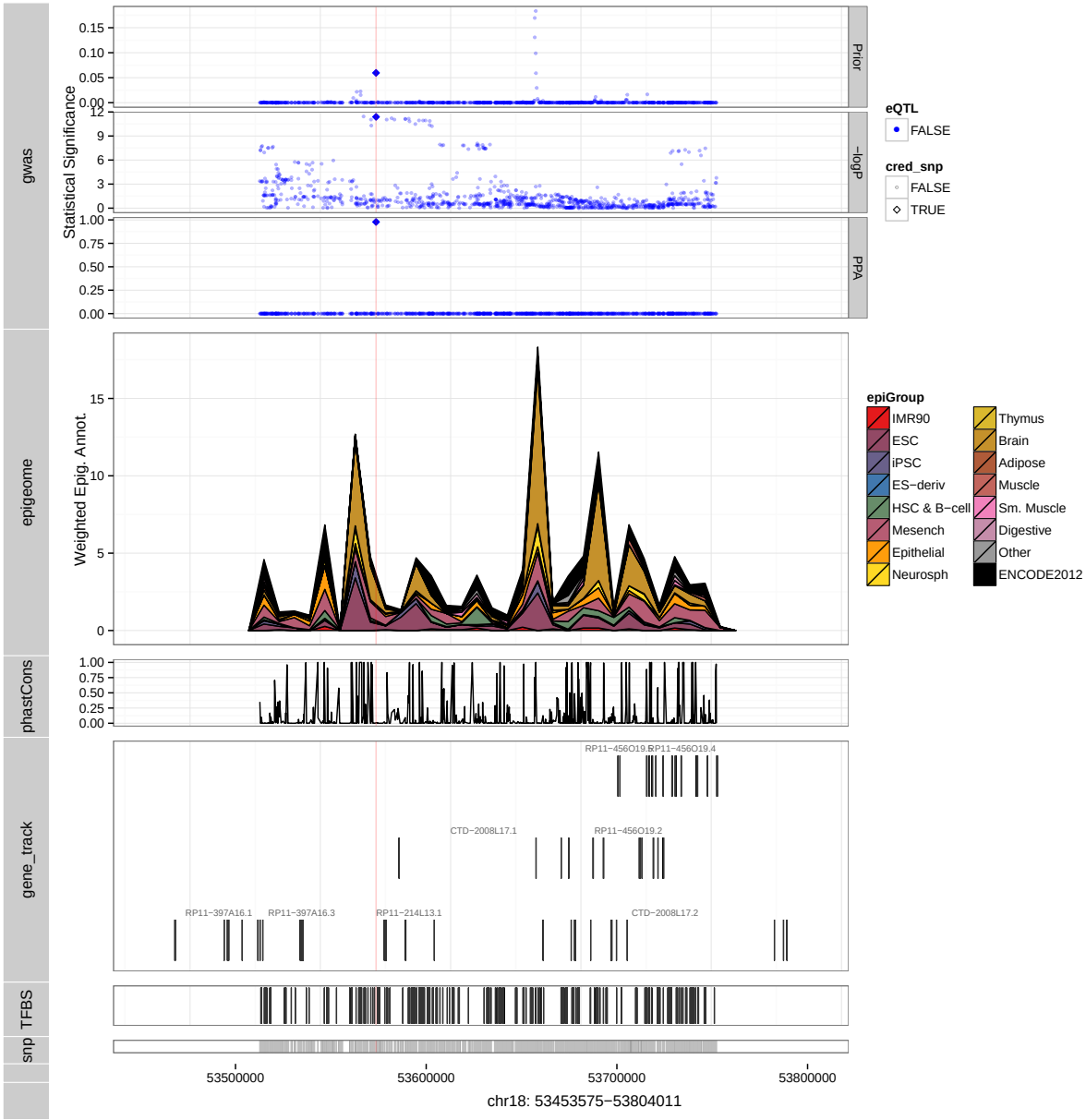
Schizophrenia



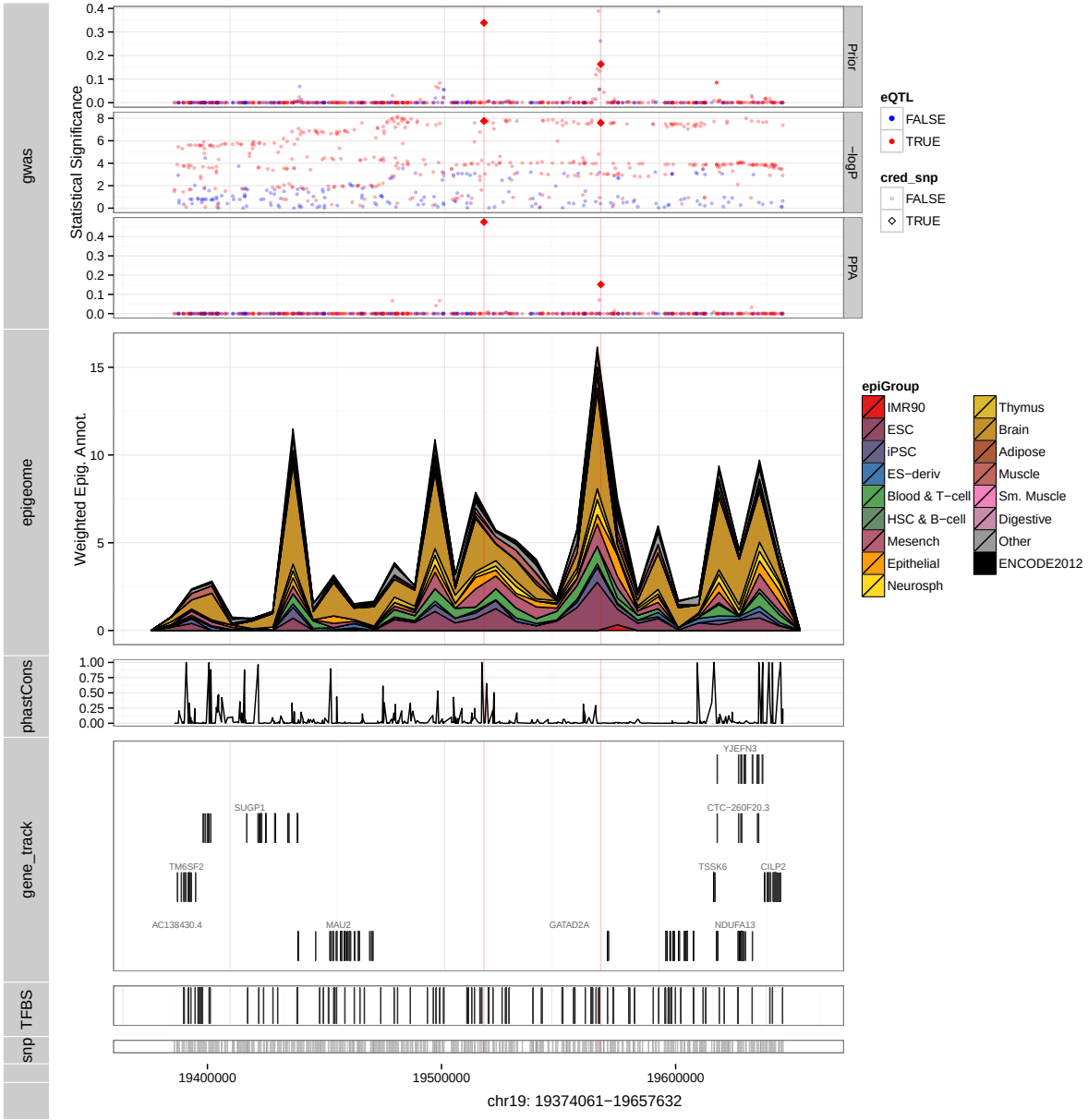
Schizophrenia



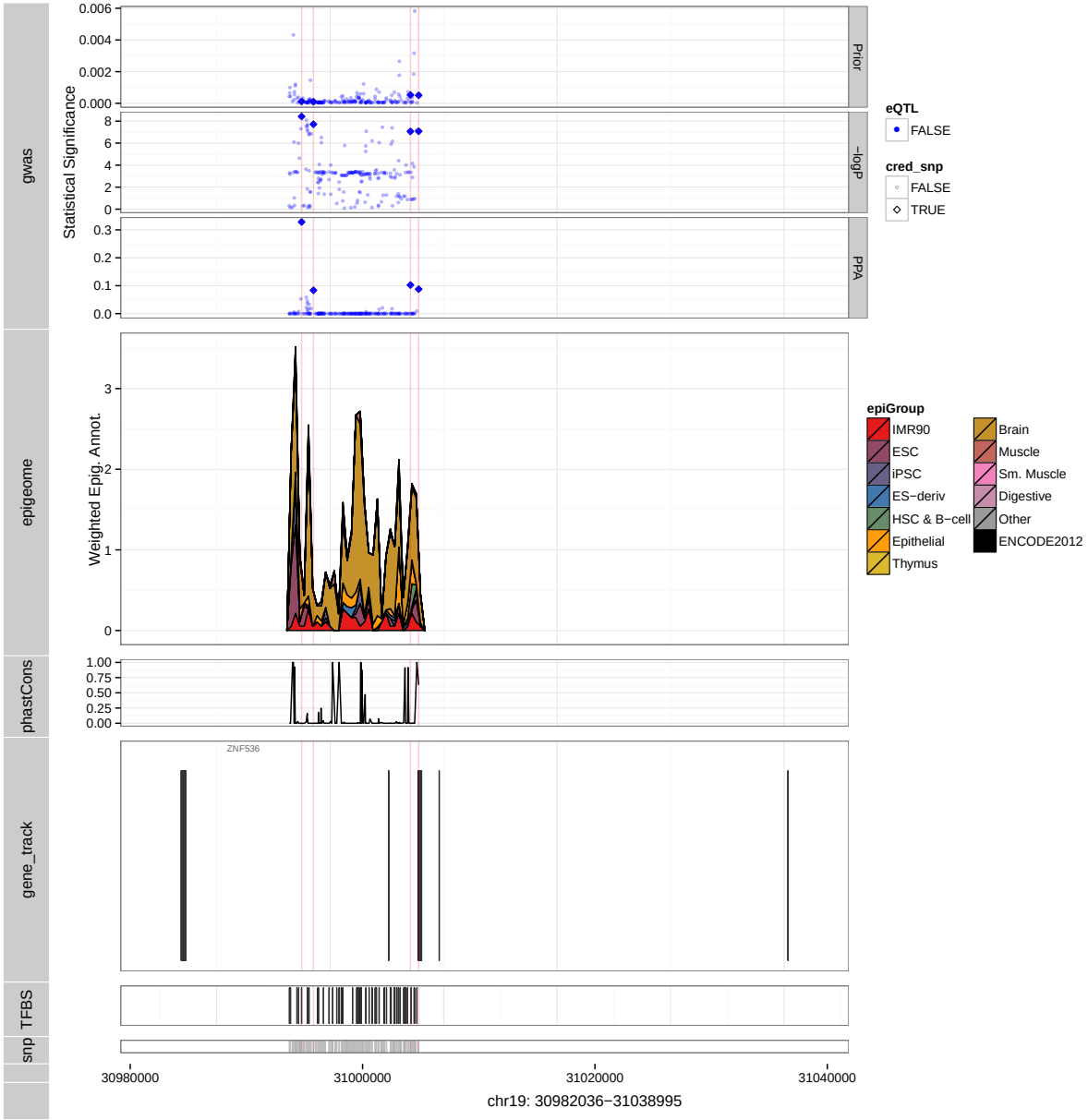
Schizophrenia



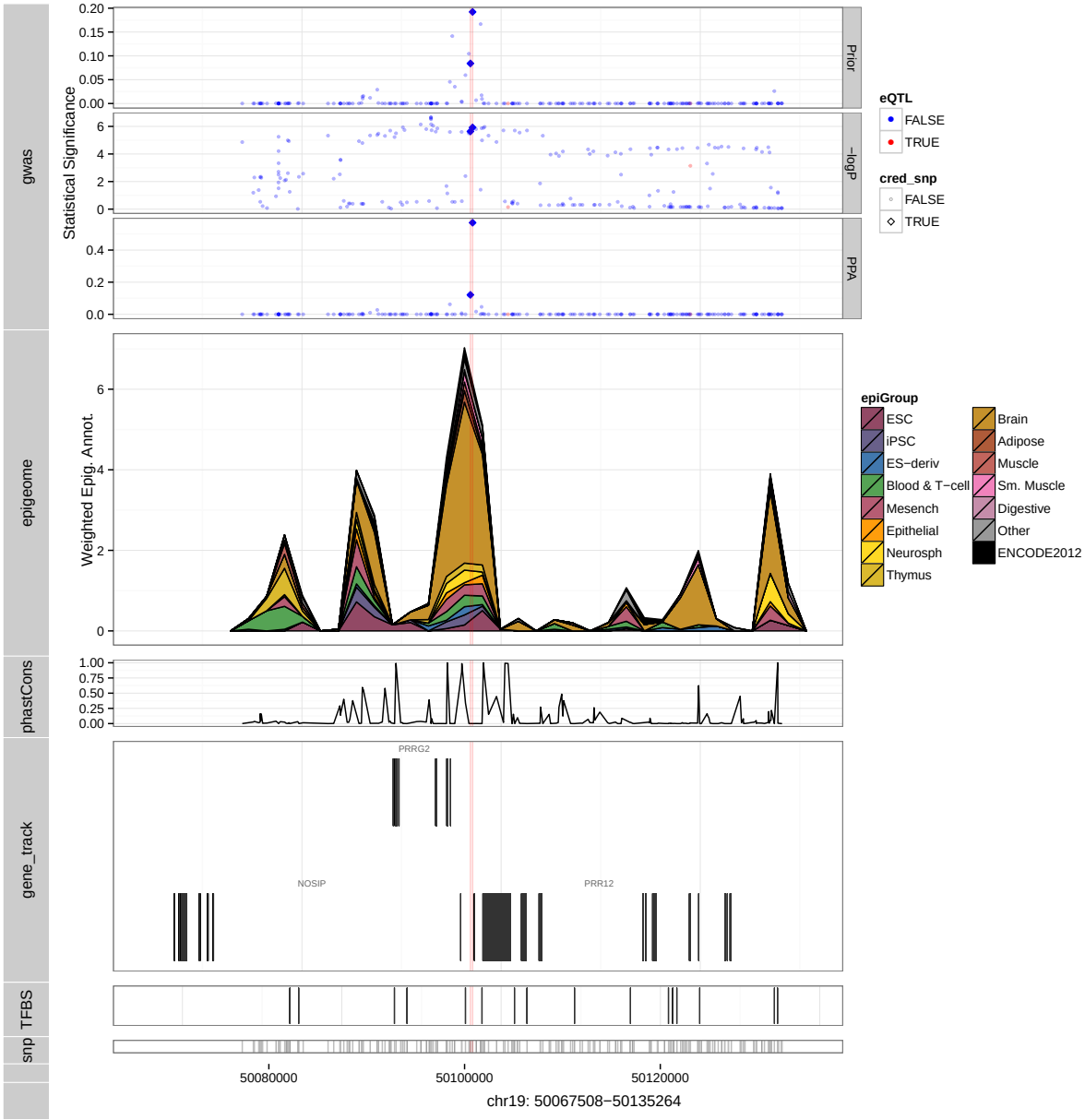
Schizophrenia



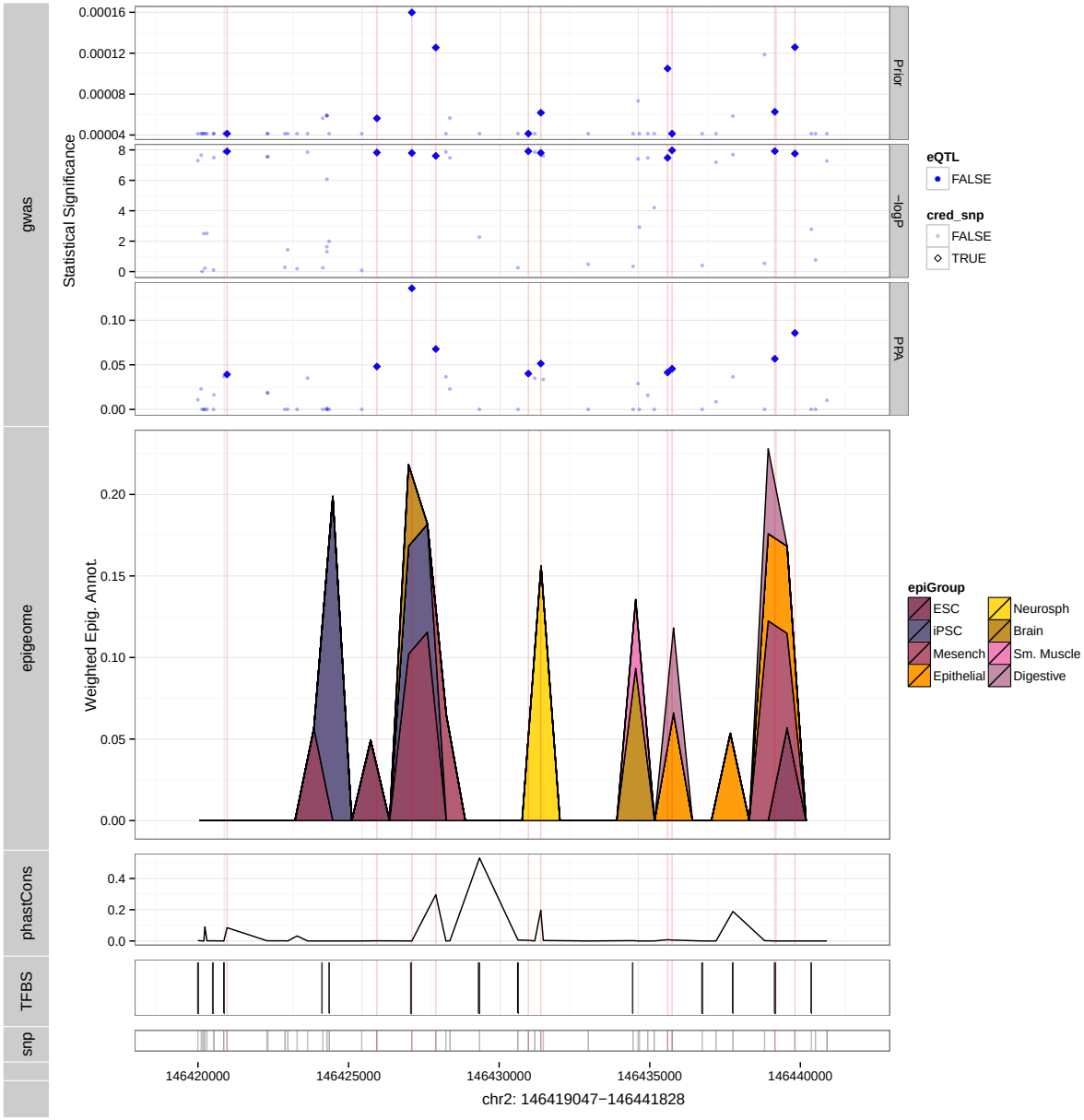
Schizophrenia



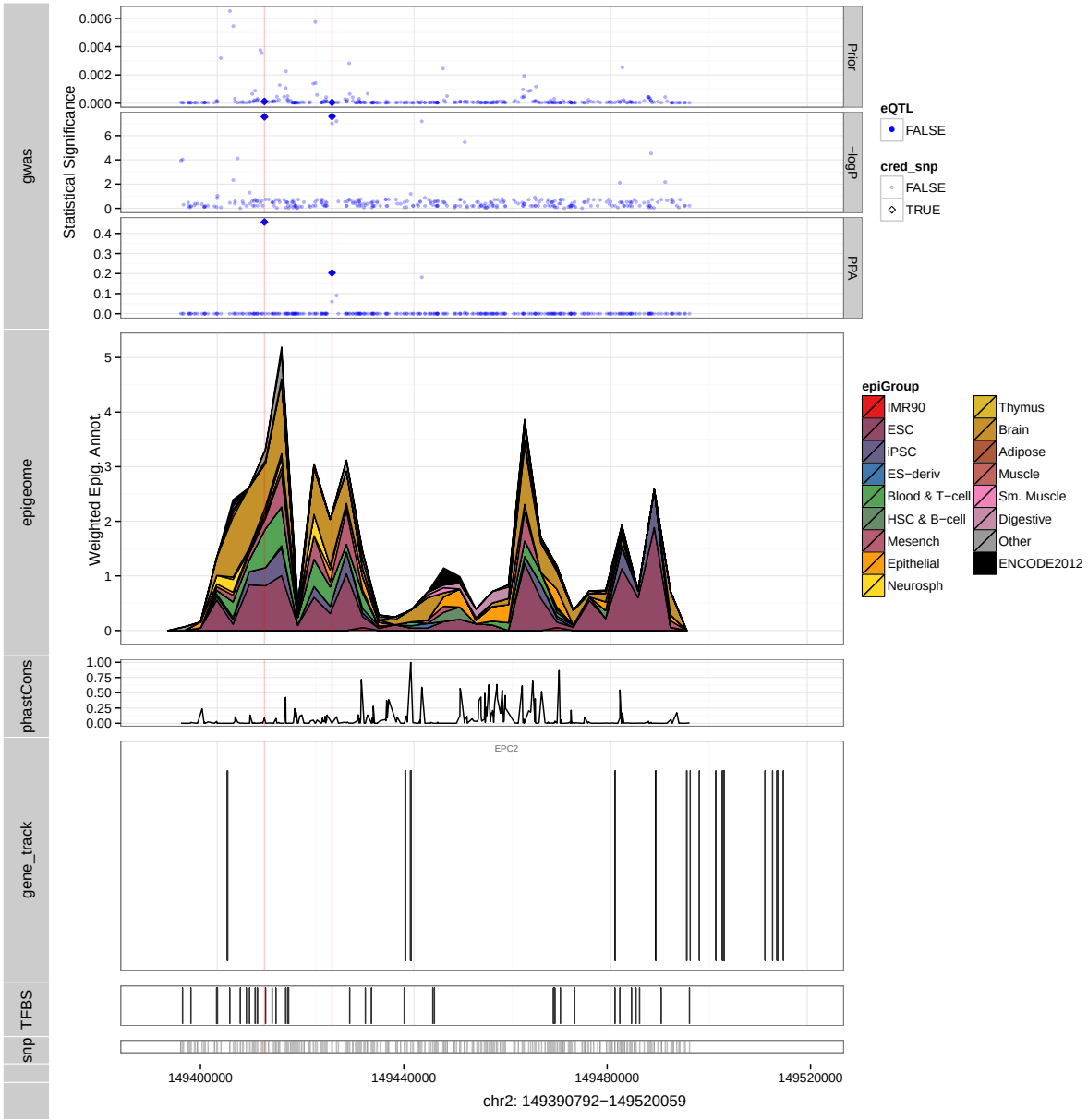
Schizophrenia



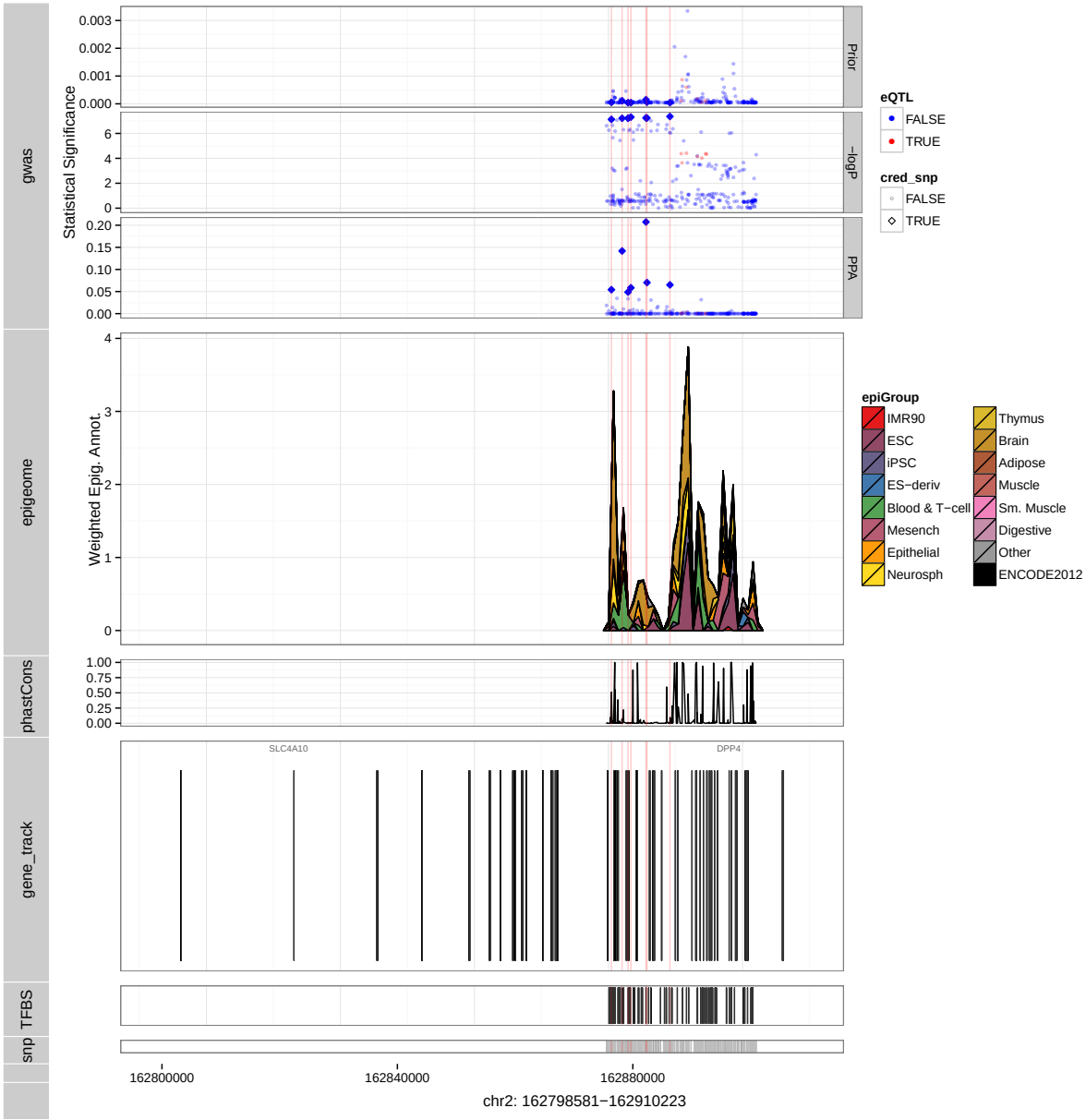
Schizophrenia



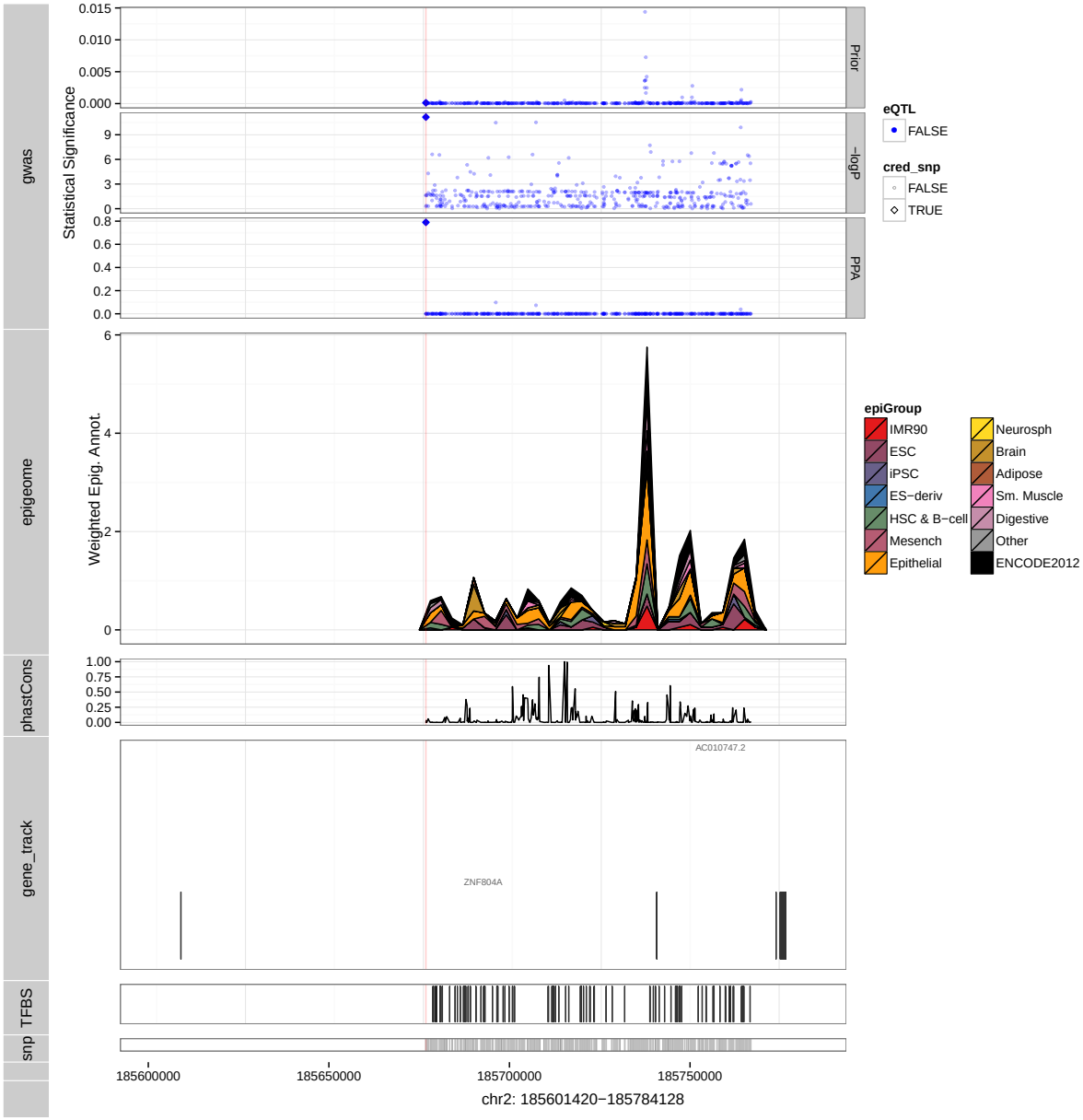
Schizophrenia



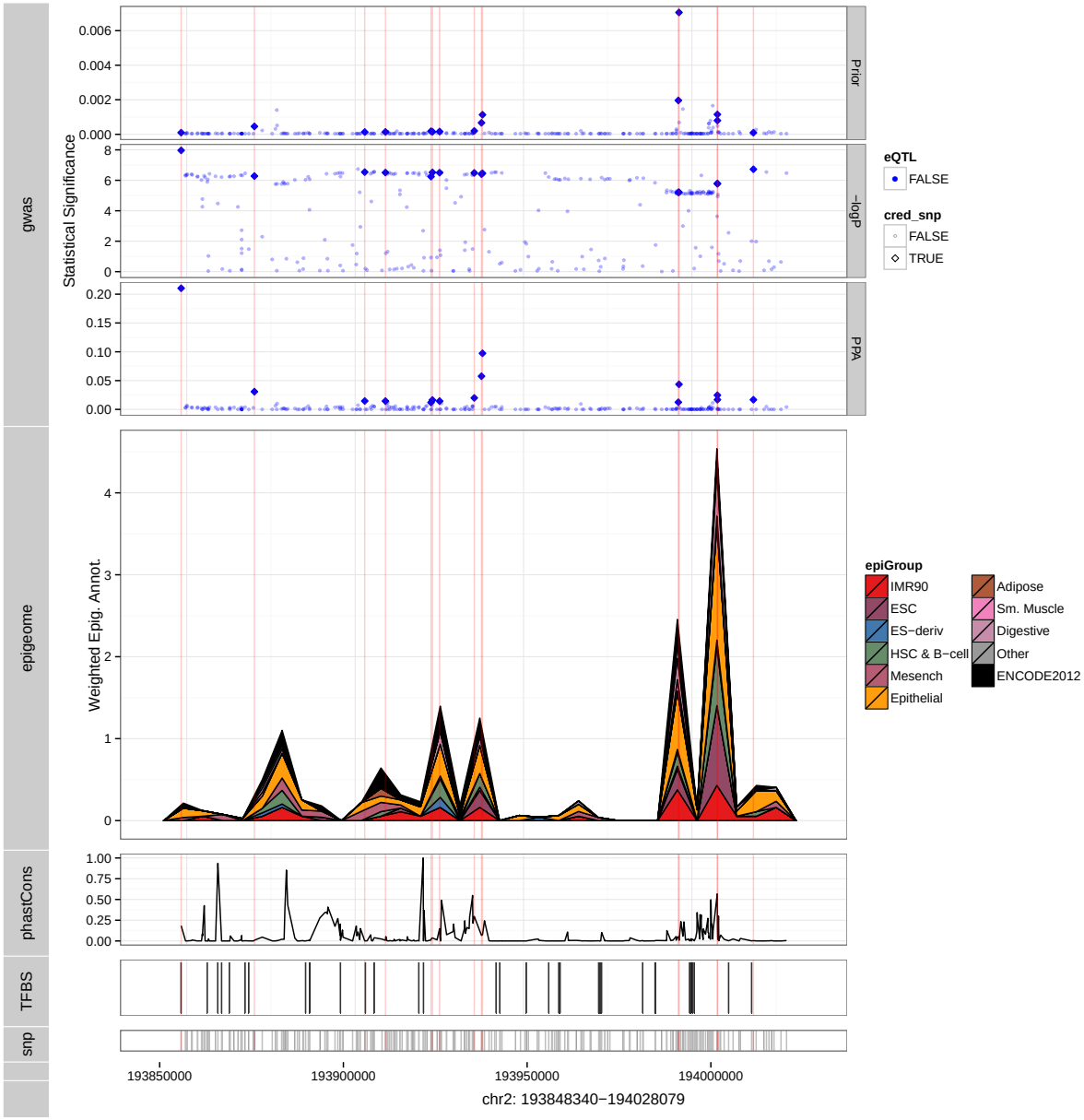
Schizophrenia



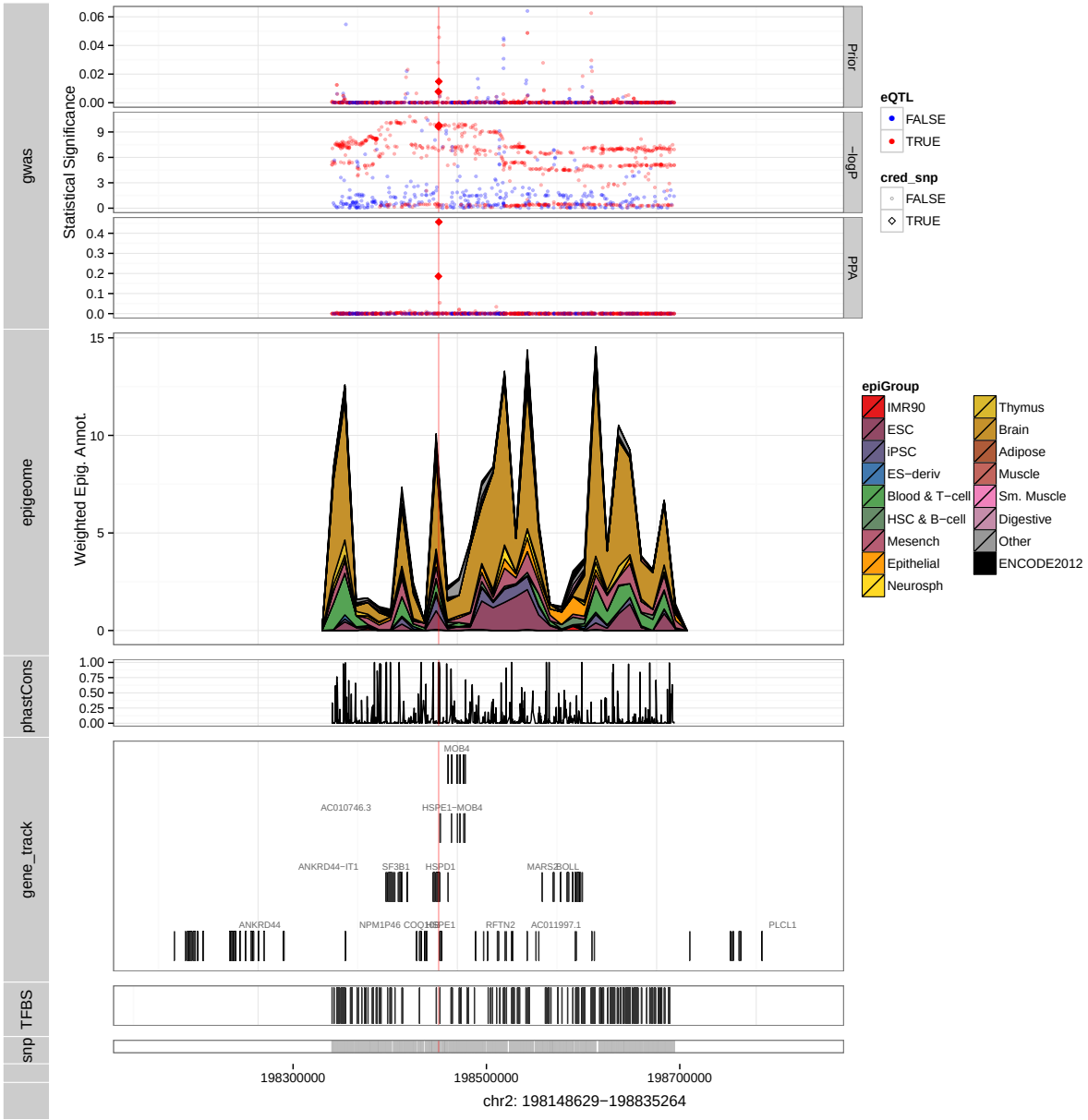
Schizophrenia



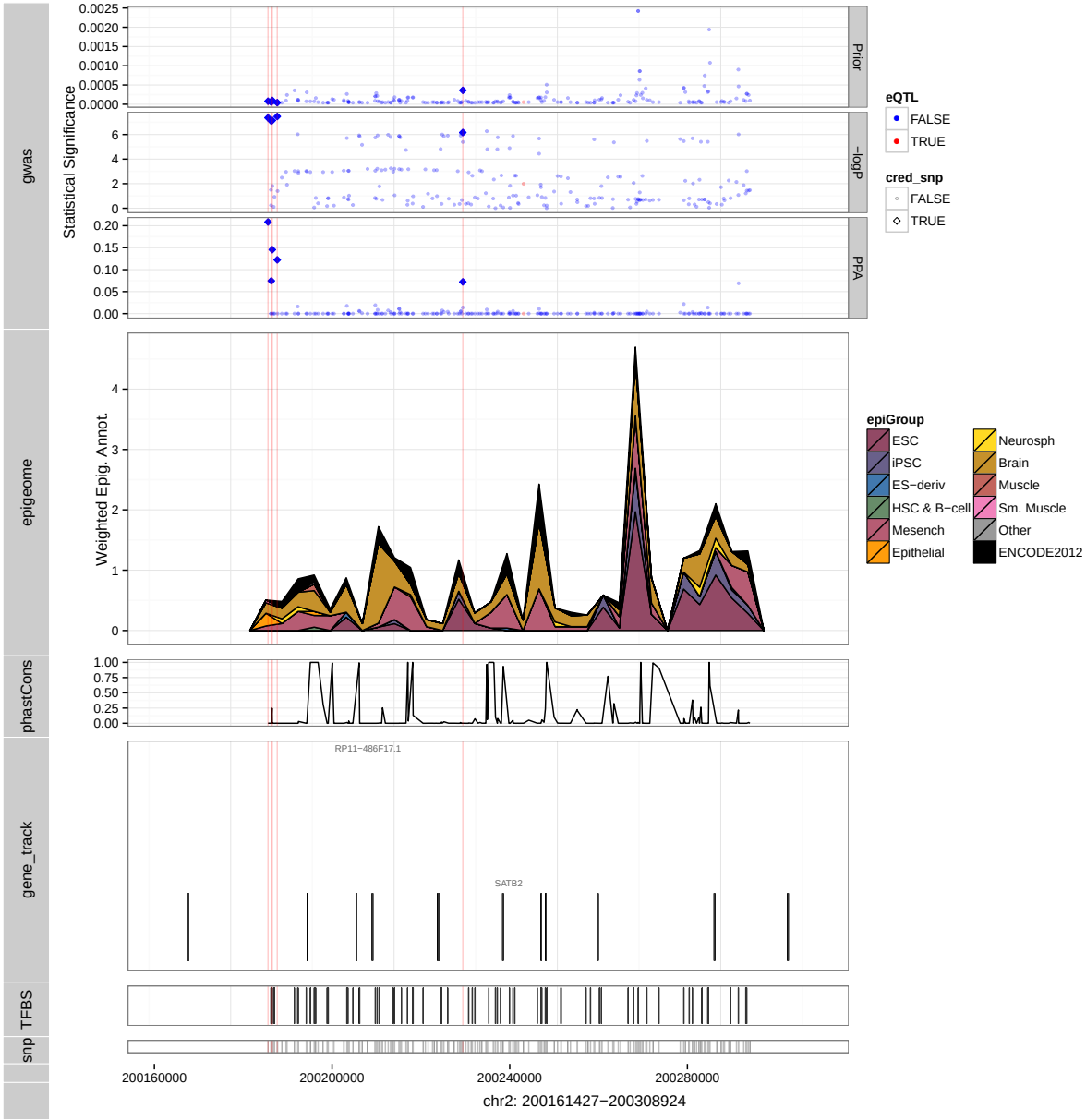
Schizophrenia



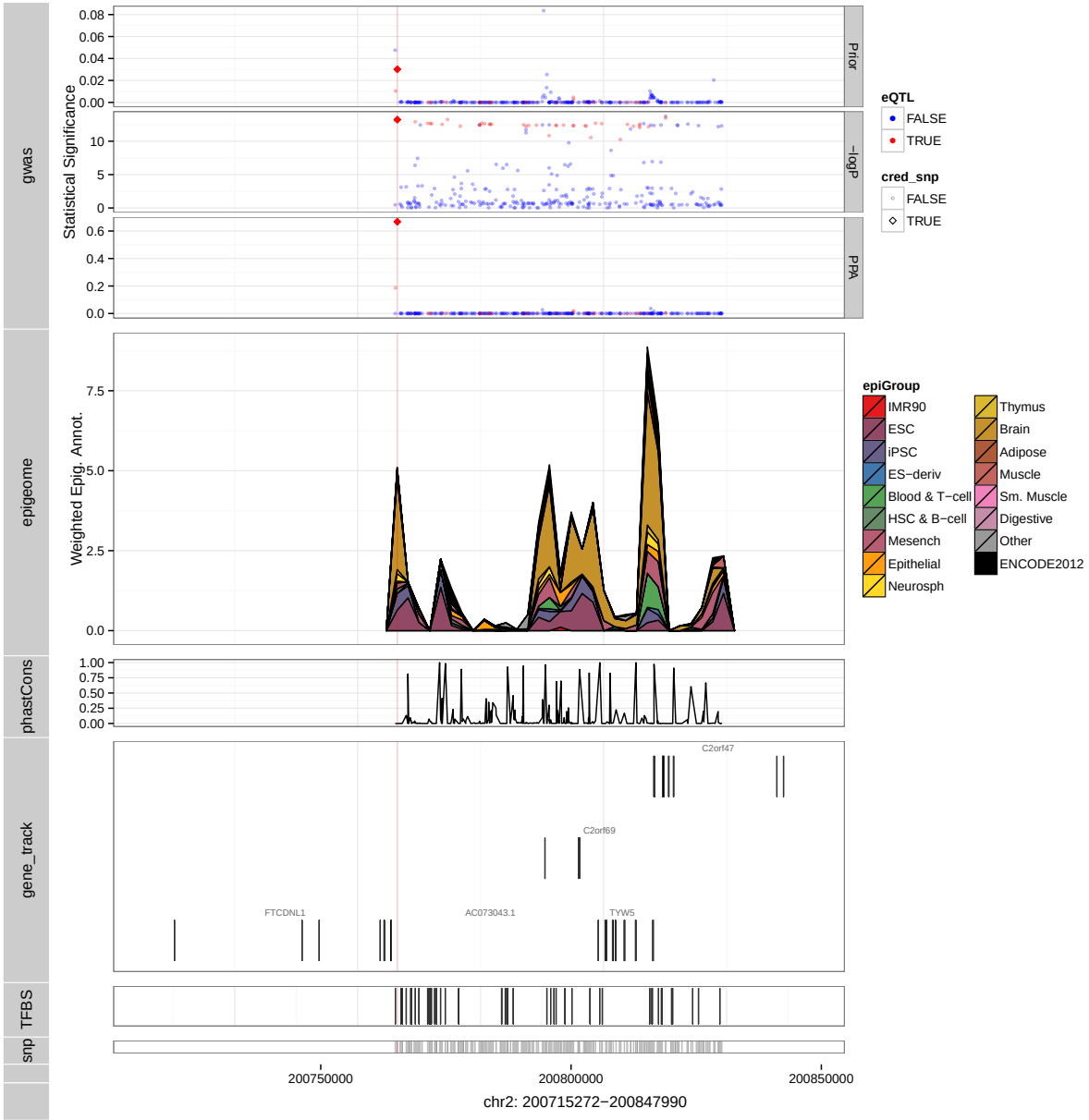
Schizophrenia



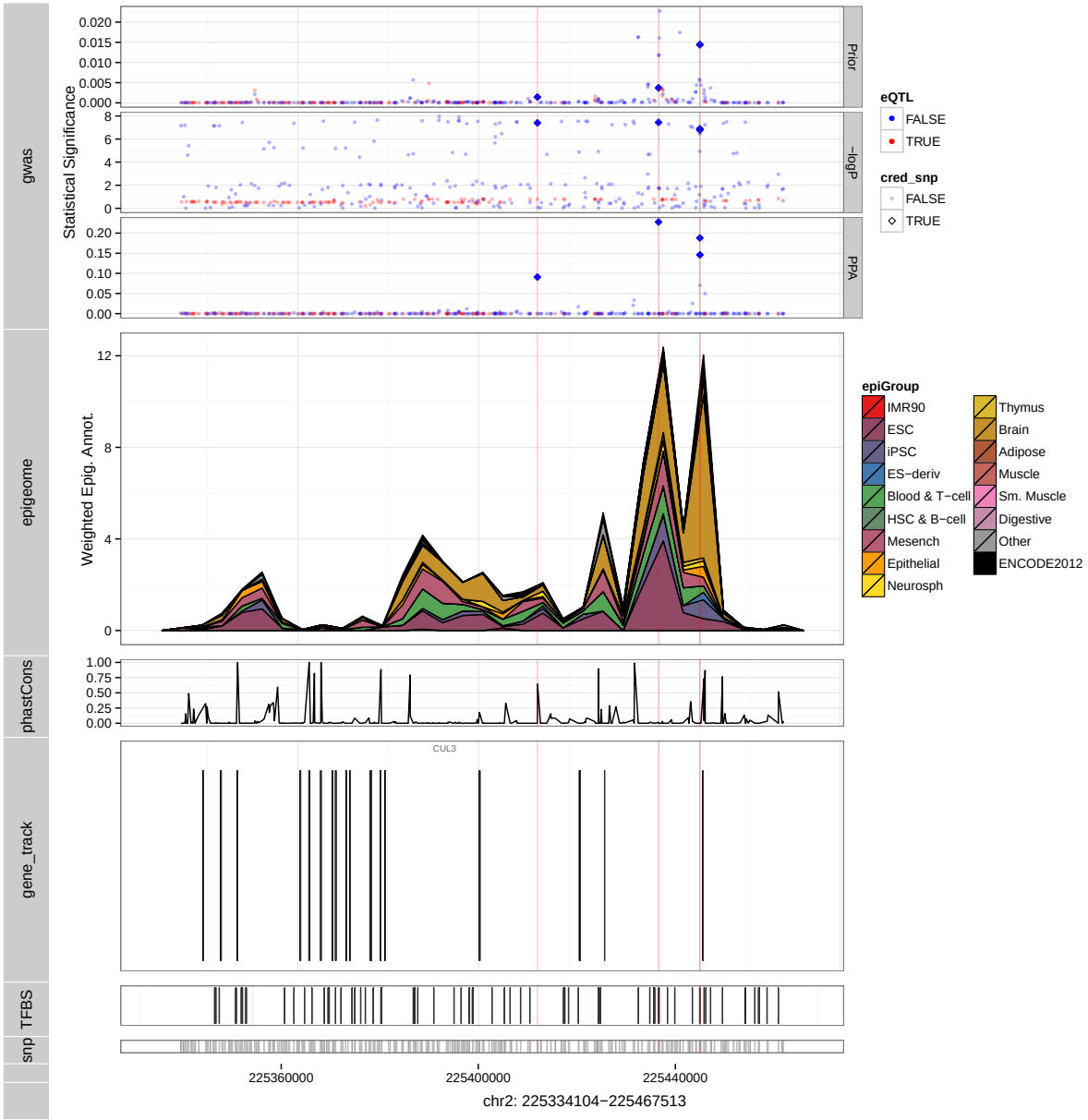
Schizophrenia



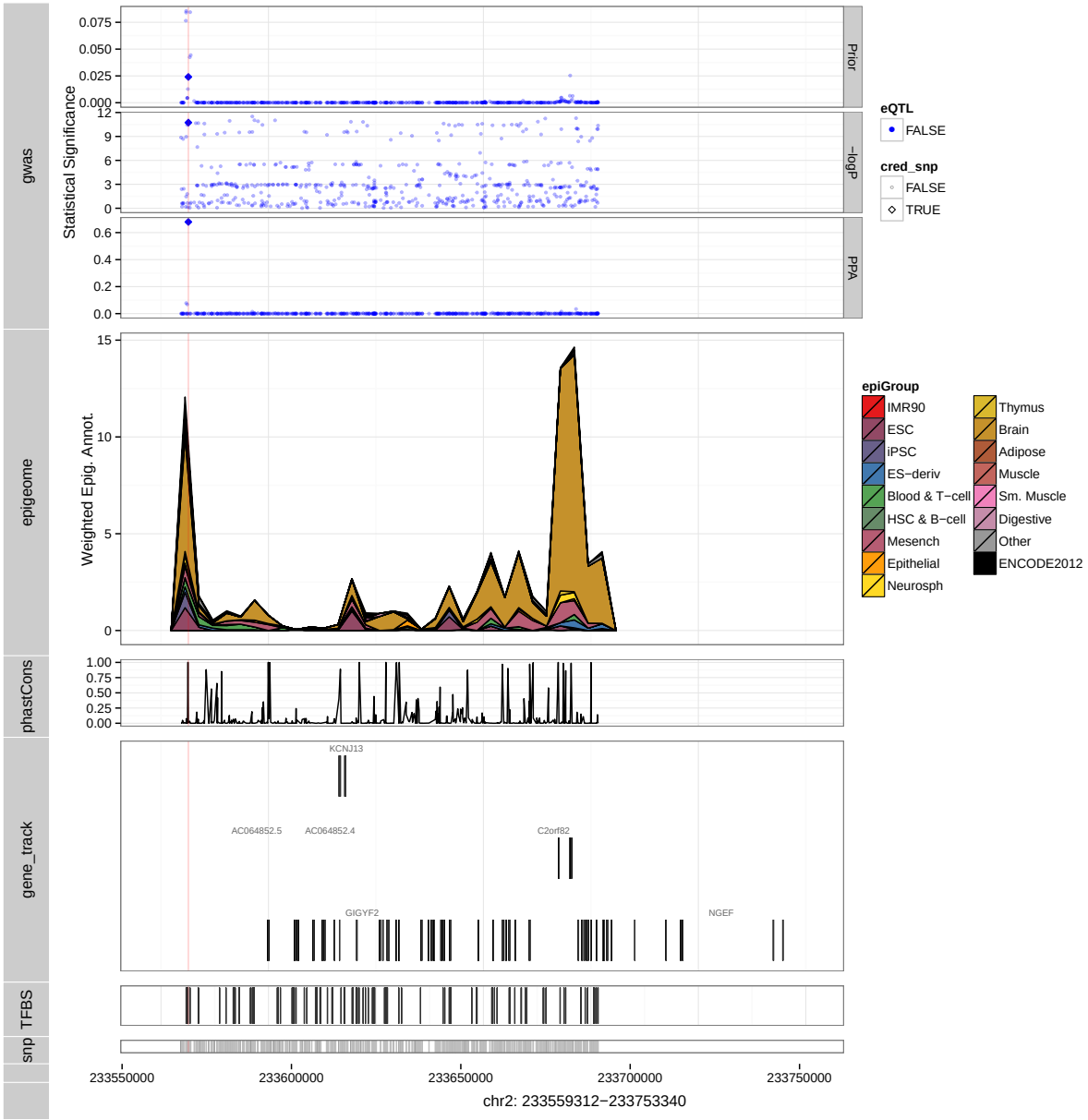
Schizophrenia



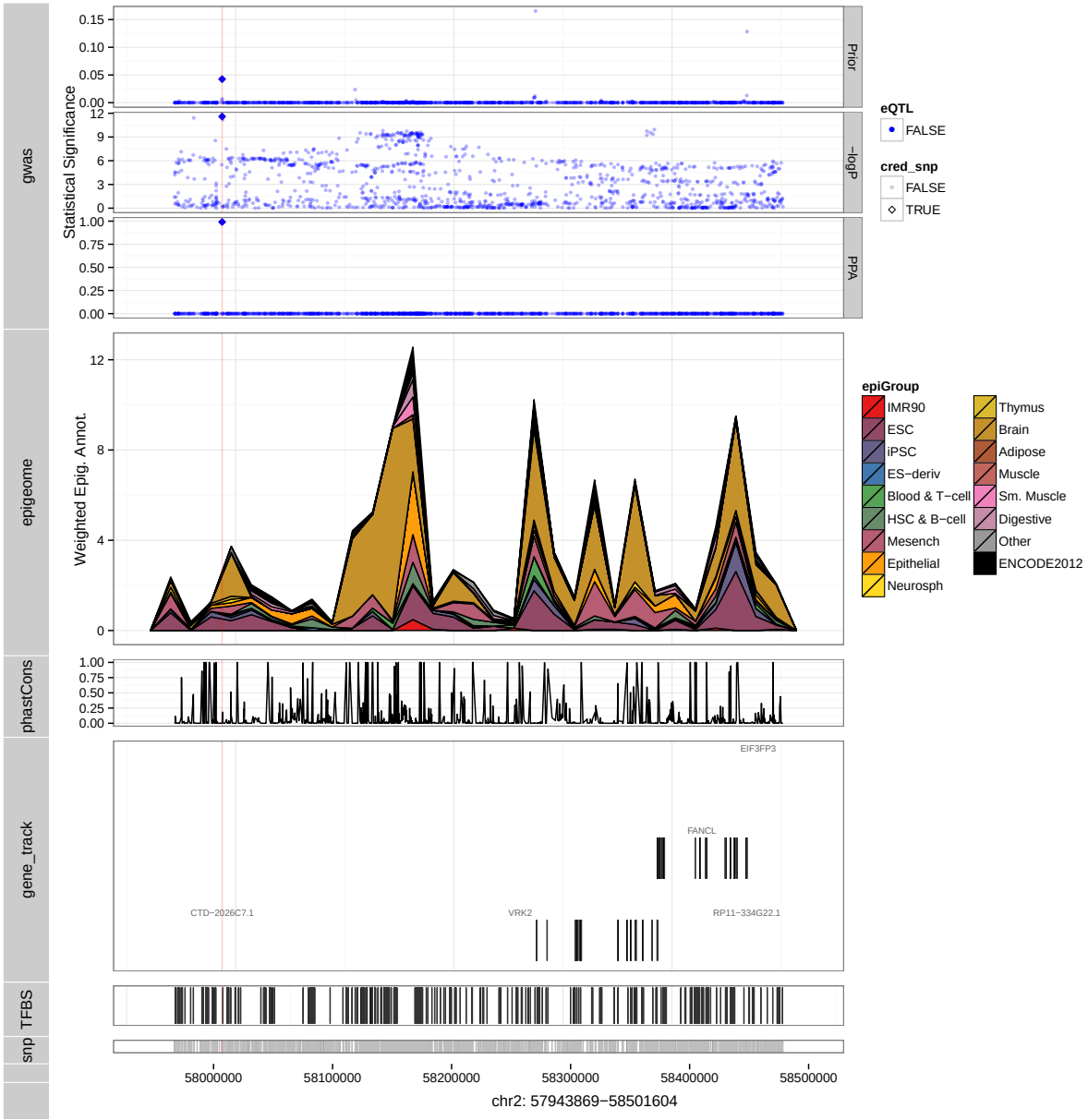
Schizophrenia



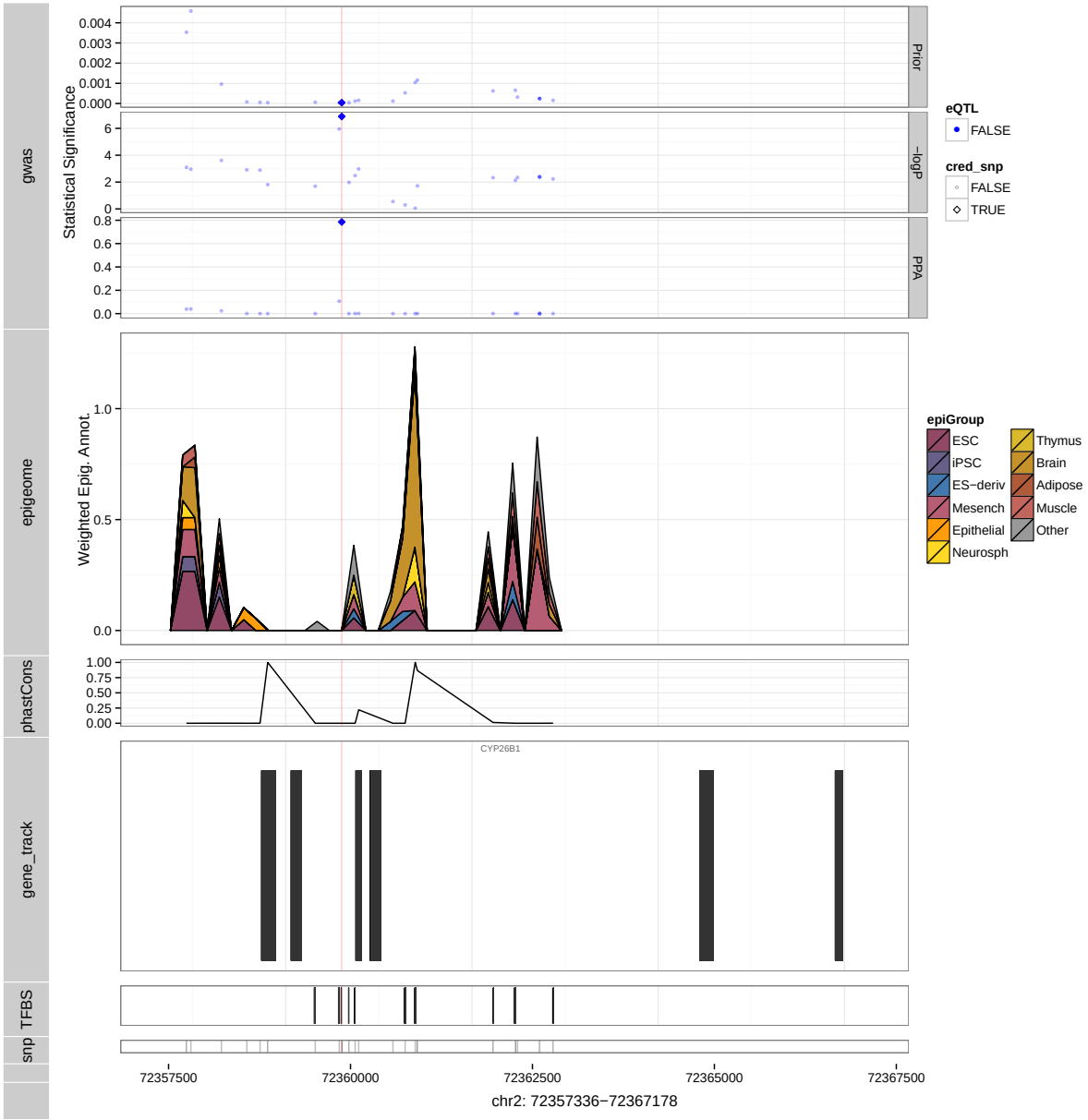
Schizophrenia



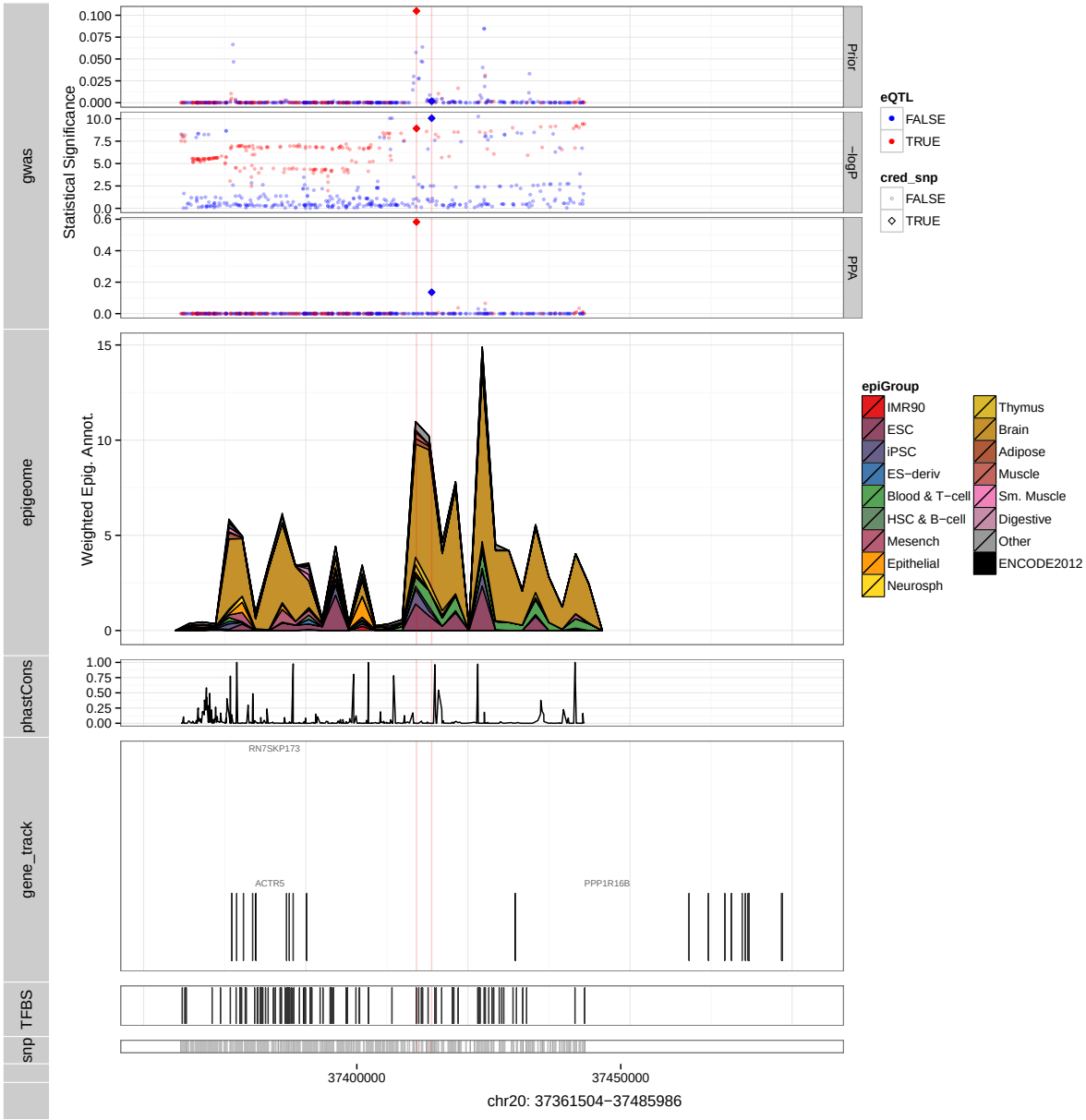
Schizophrenia



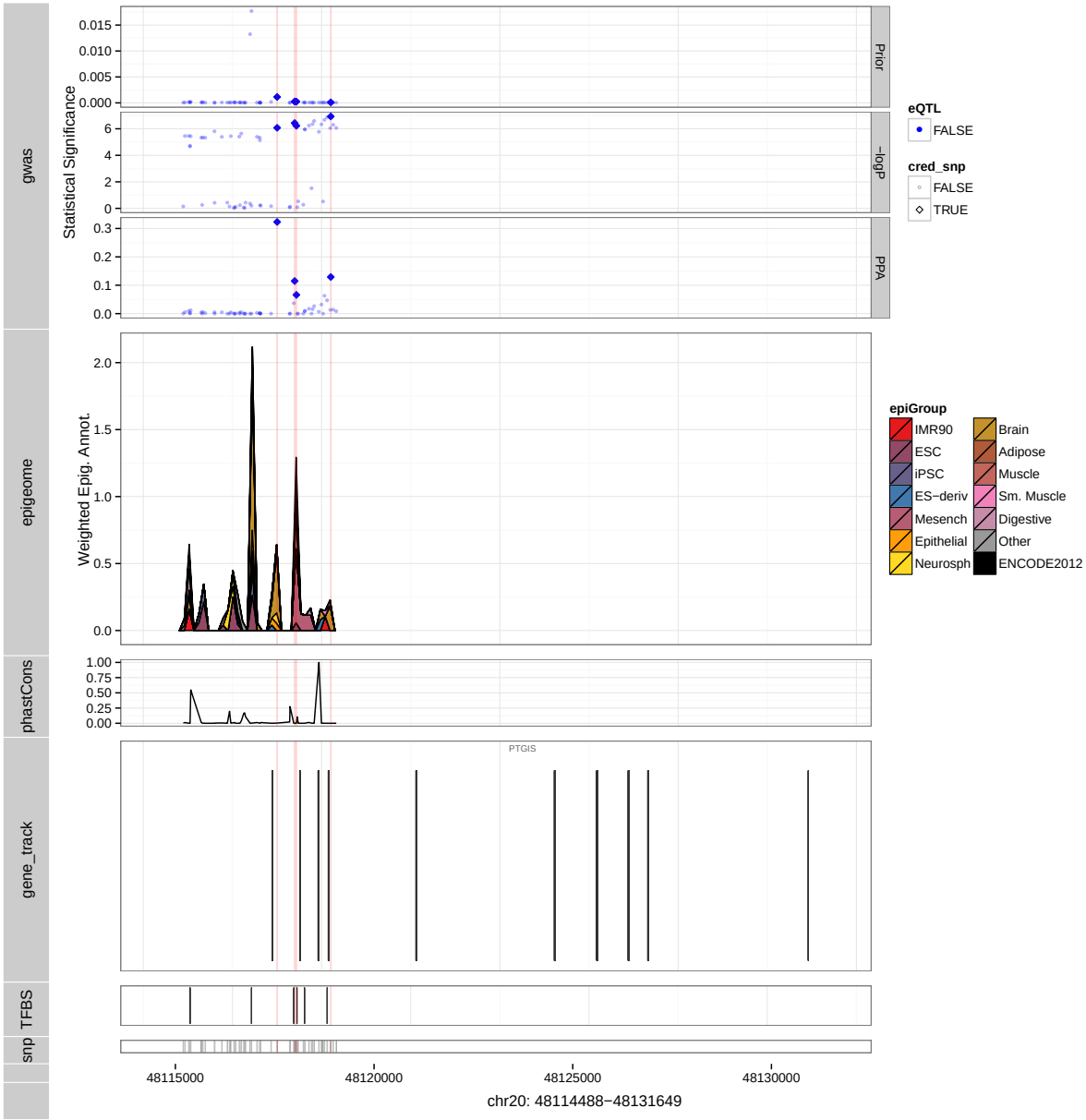
Schizophrenia



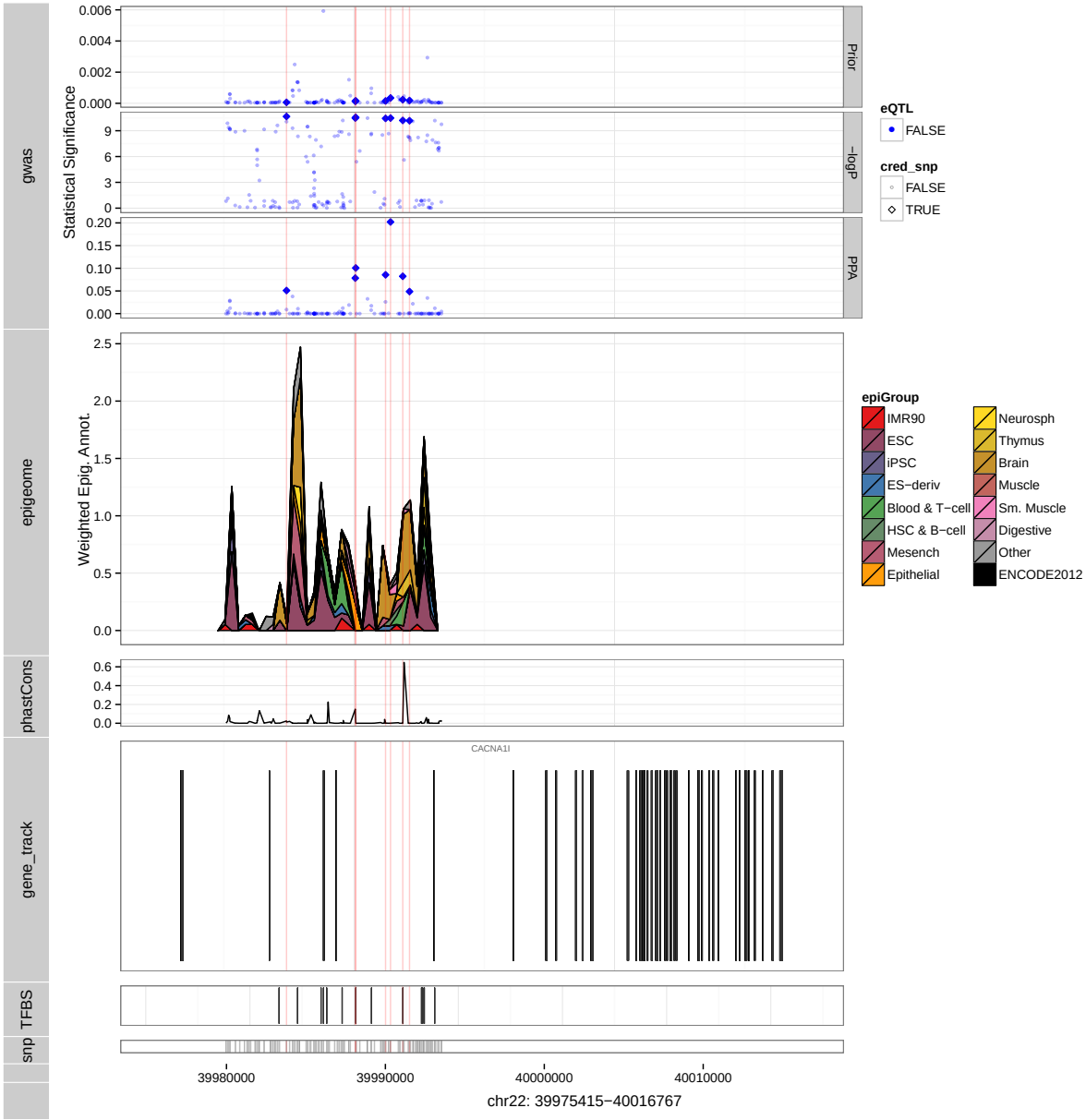
Schizophrenia



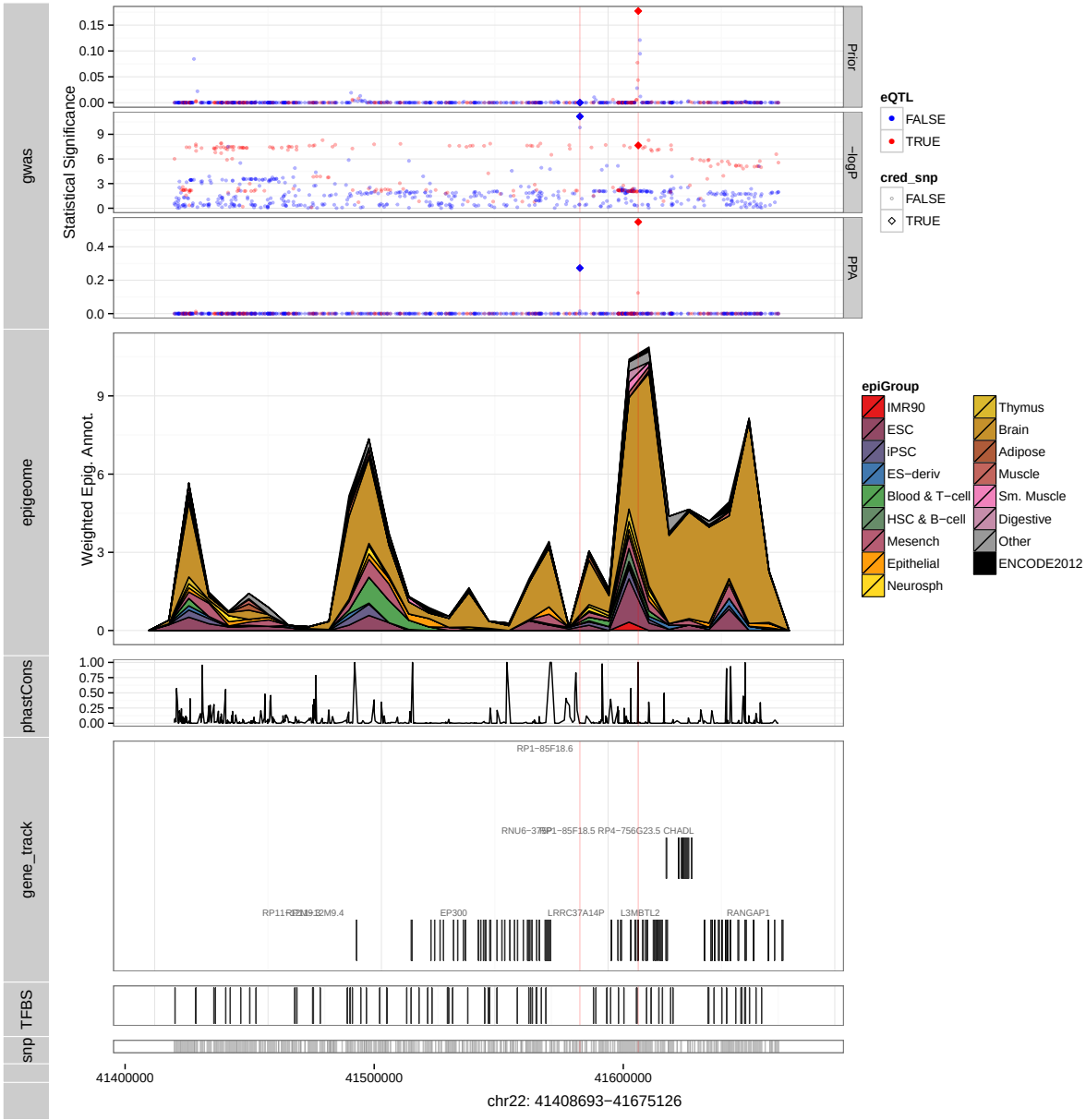
Schizophrenia



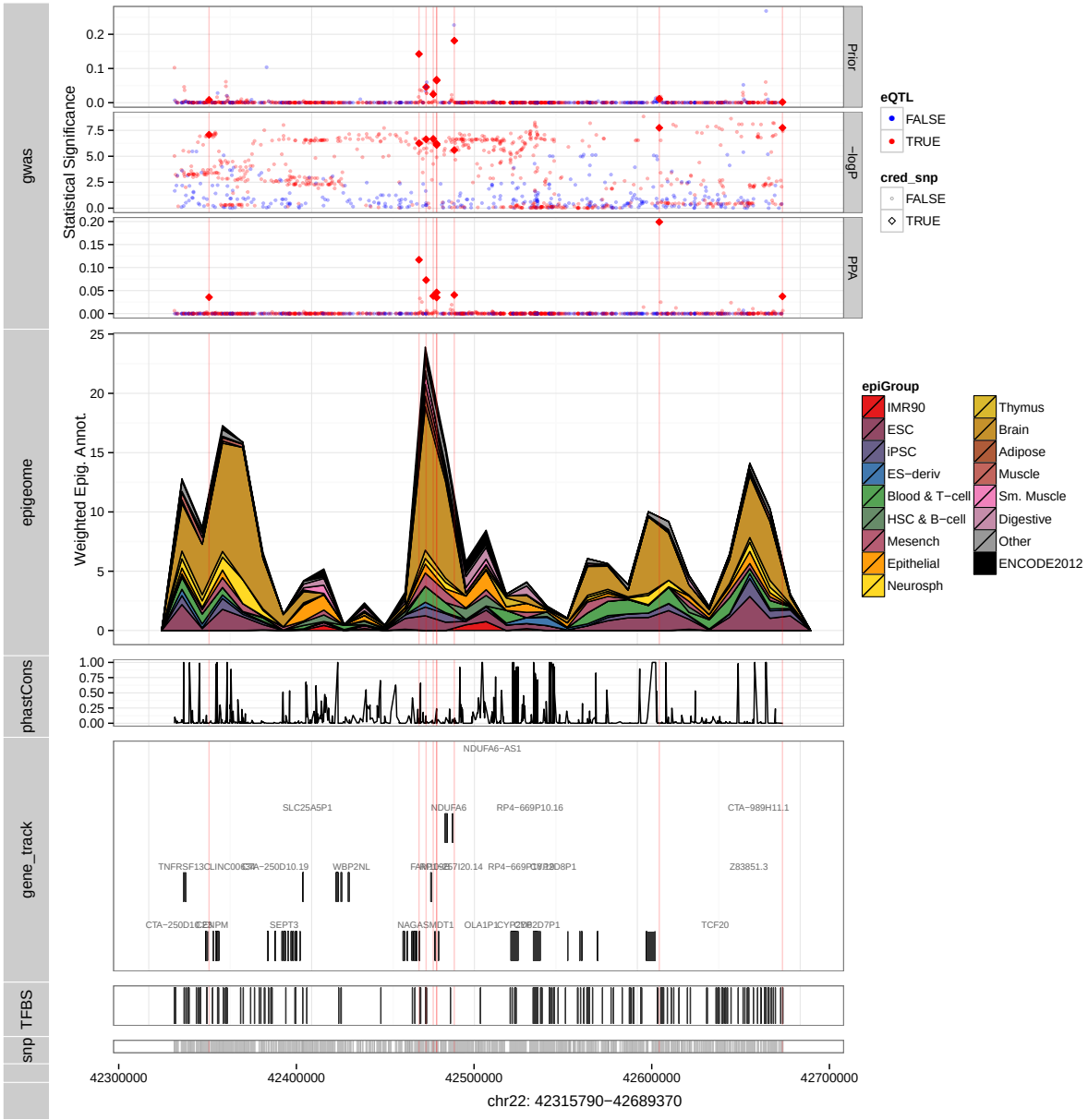
Schizophrenia



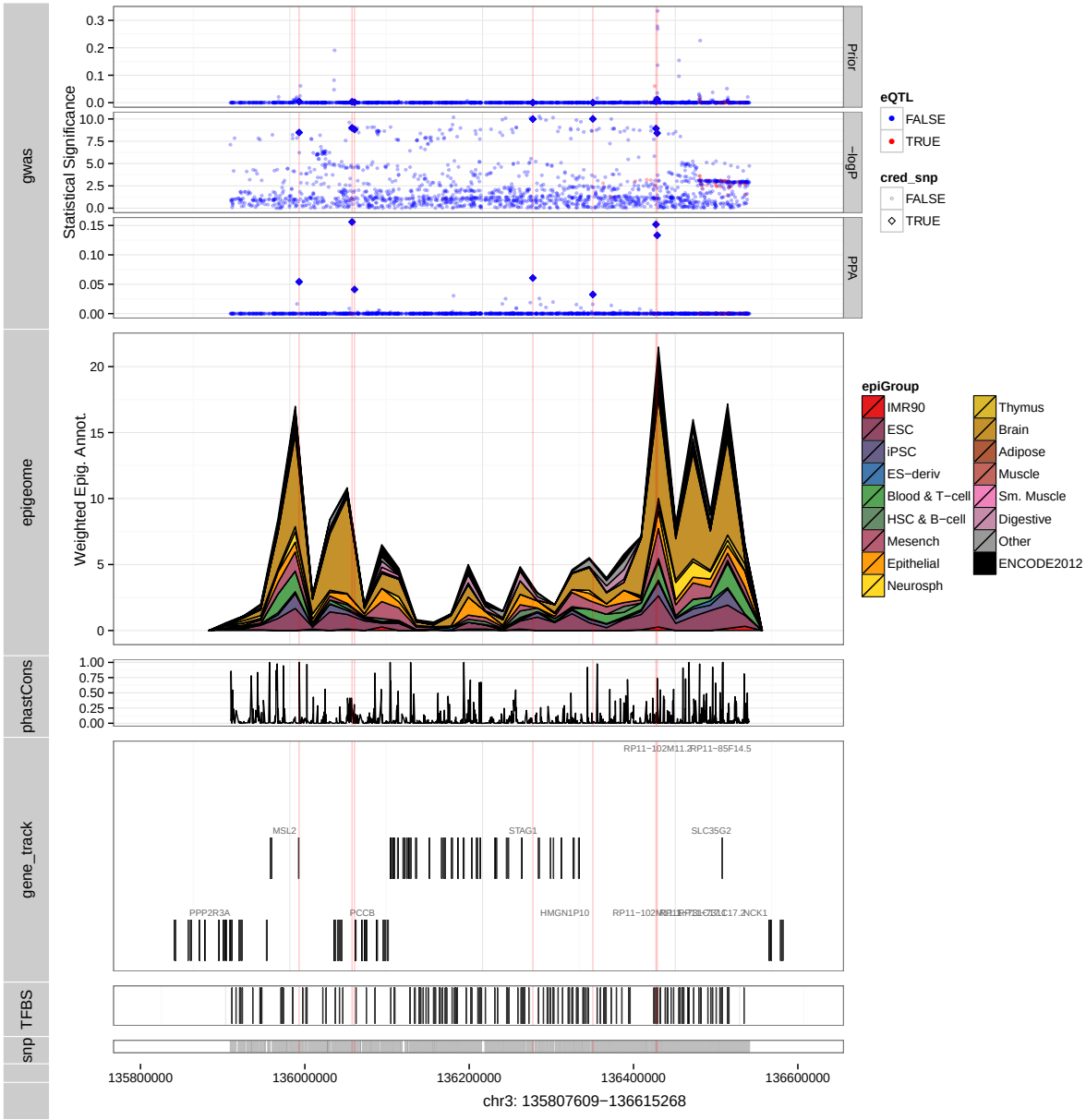
Schizophrenia



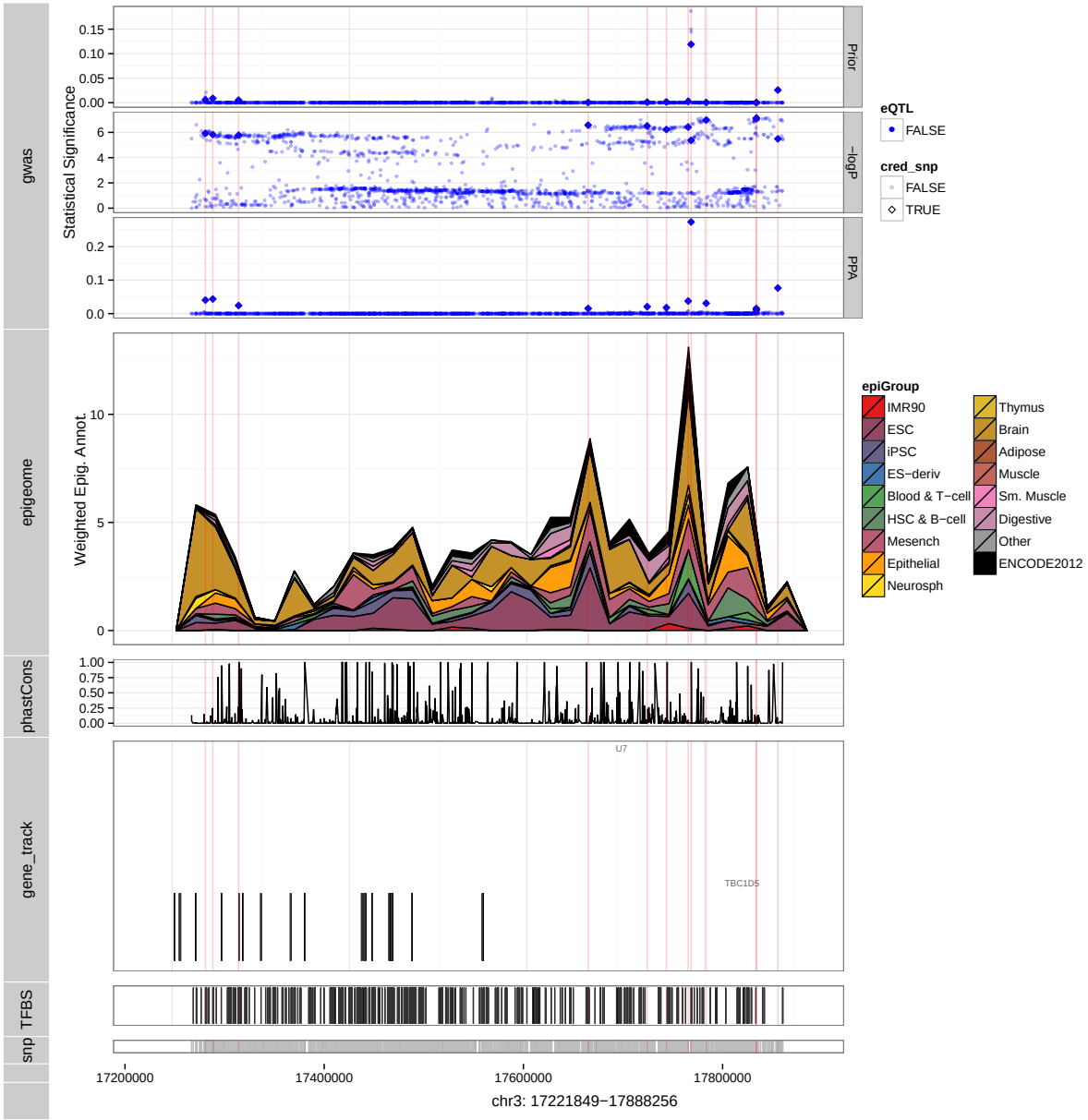
Schizophrenia



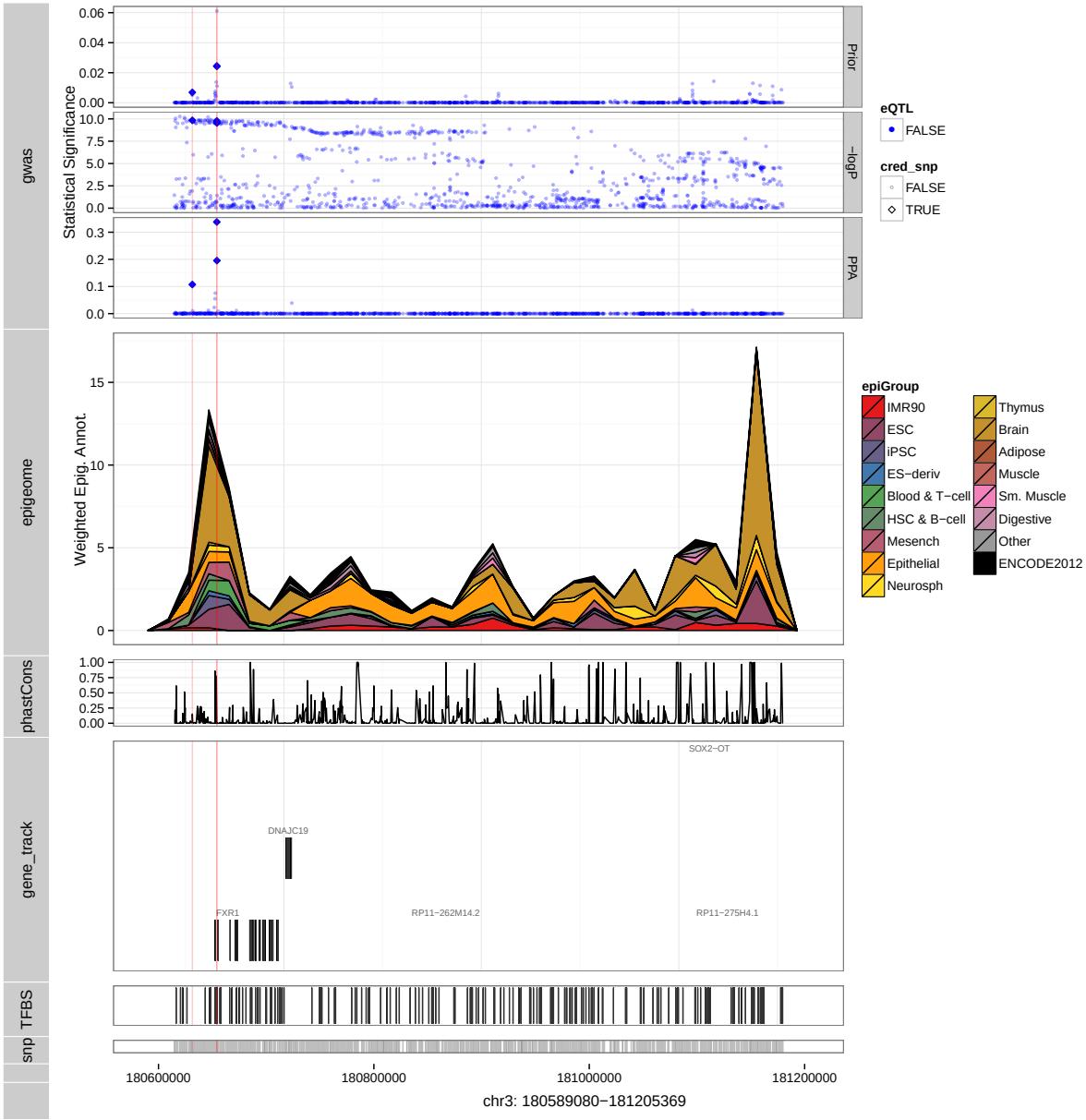
Schizophrenia



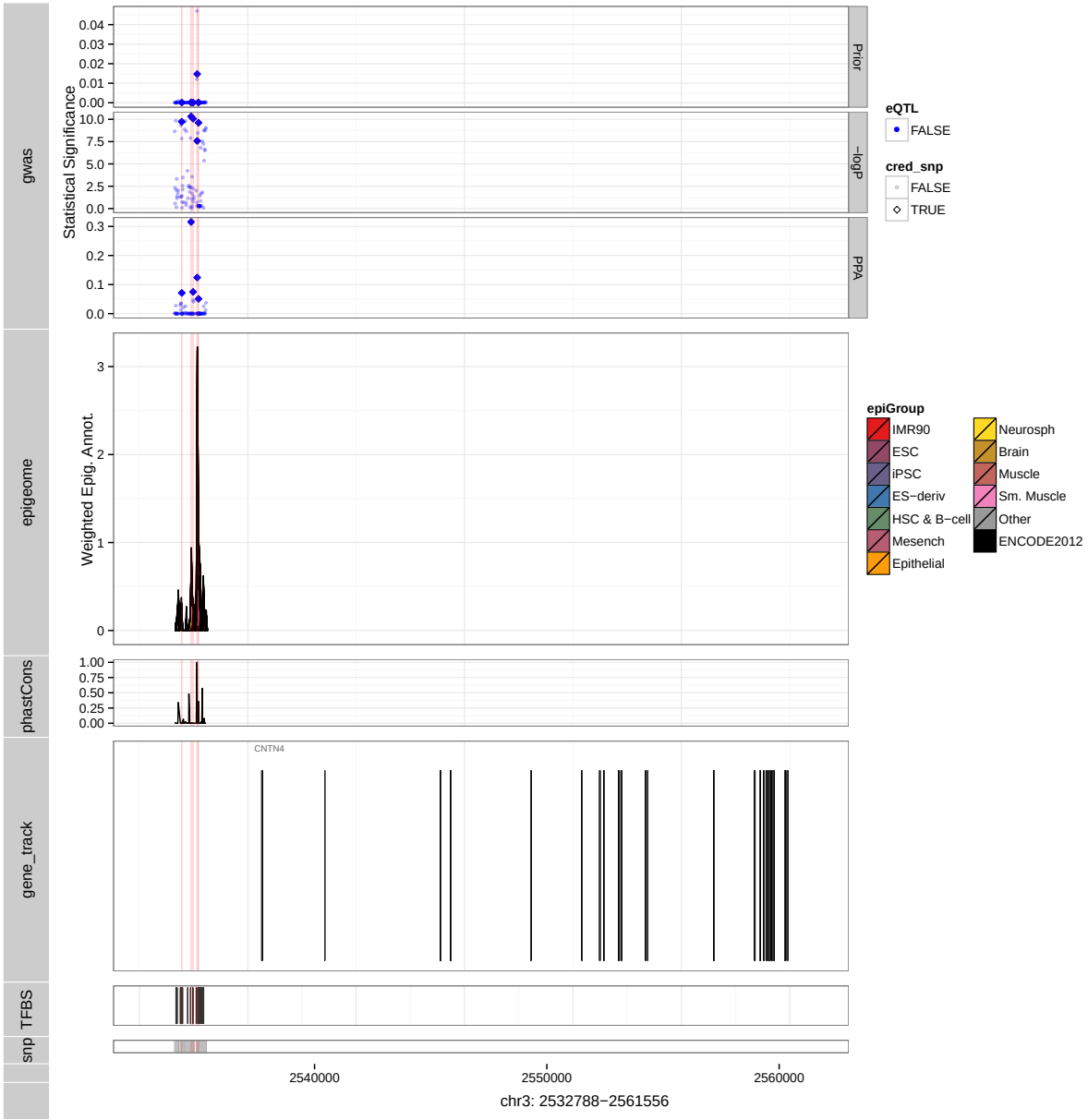
Schizophrenia



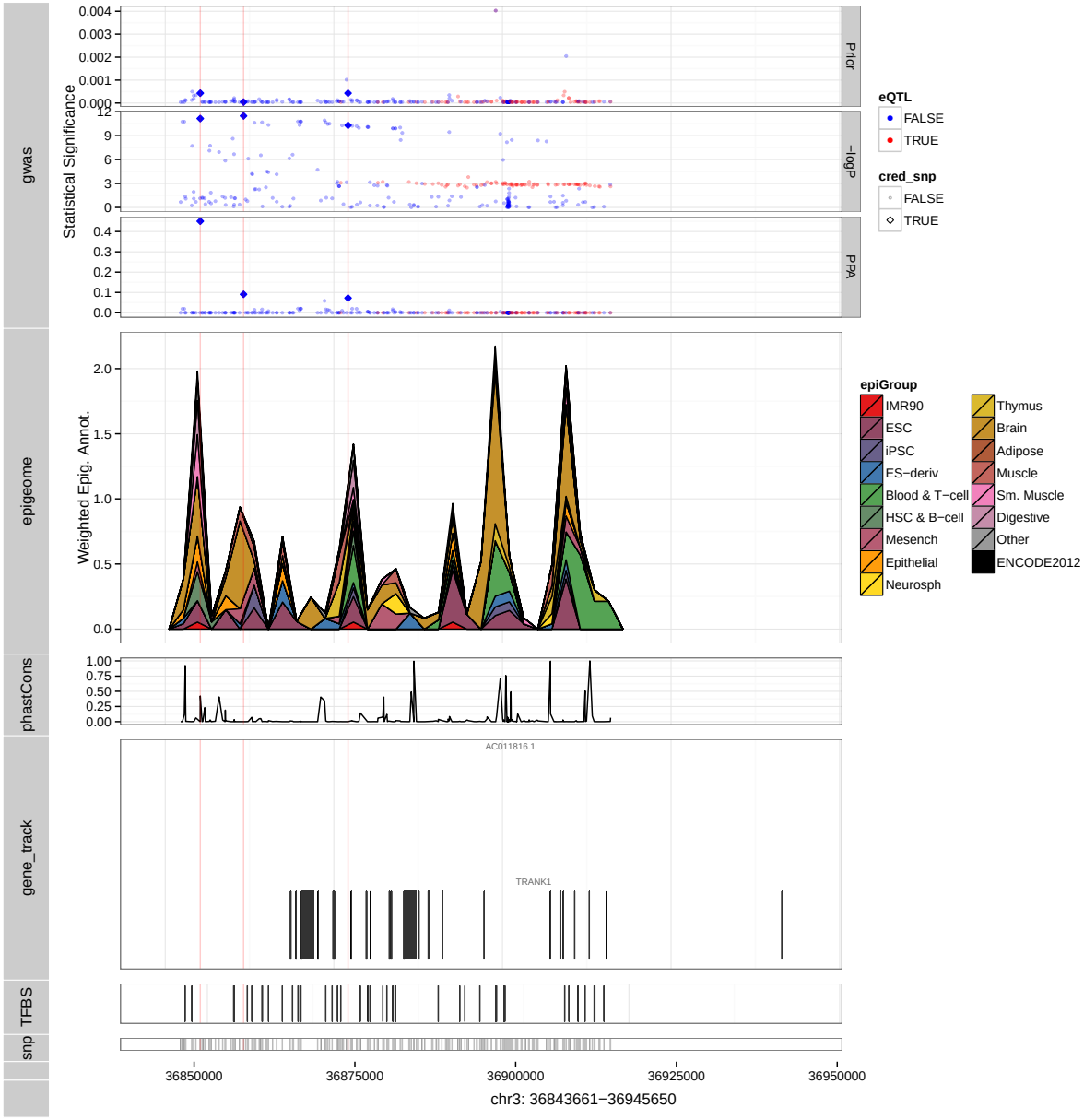
Schizophrenia



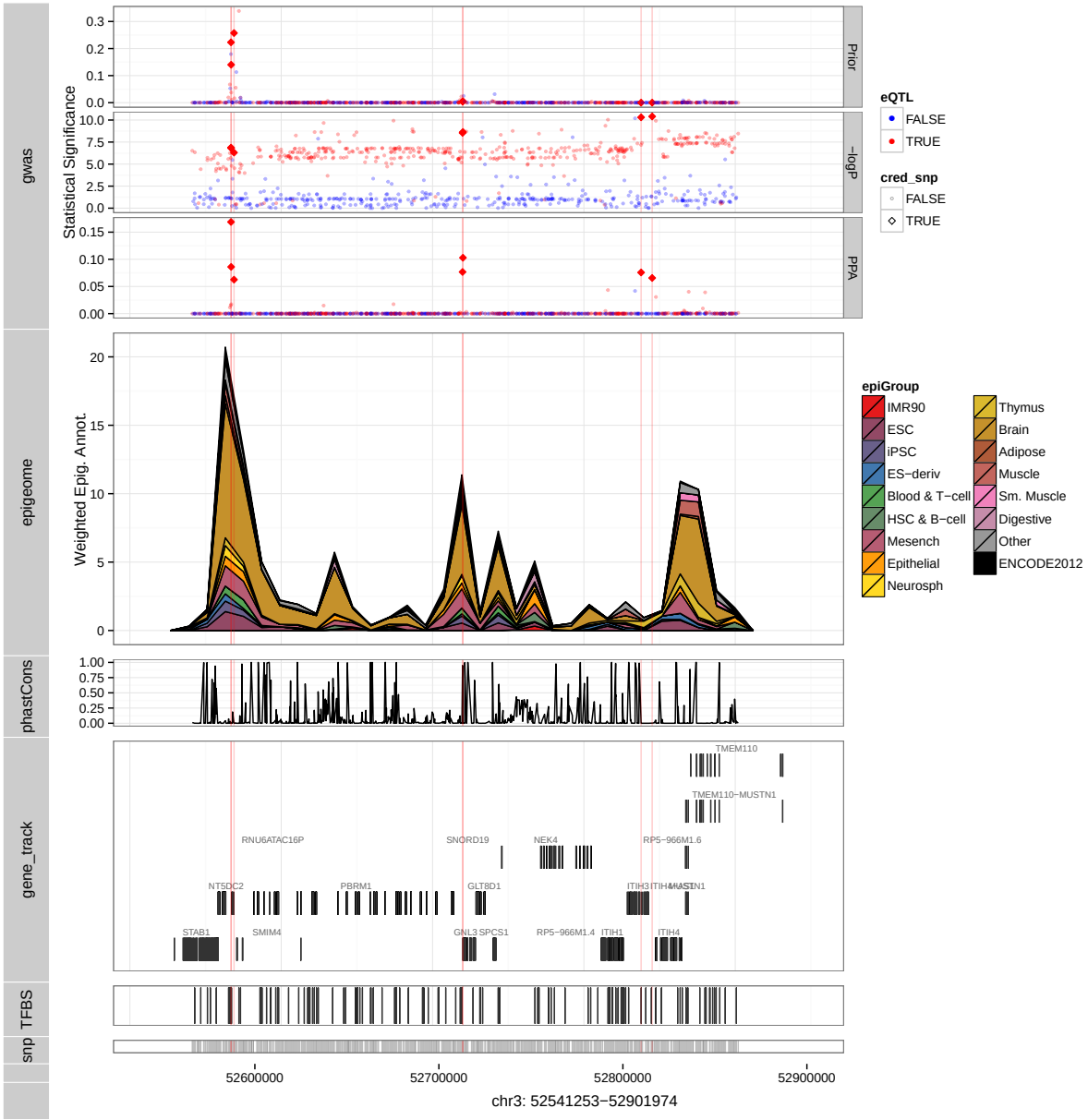
Schizophrenia



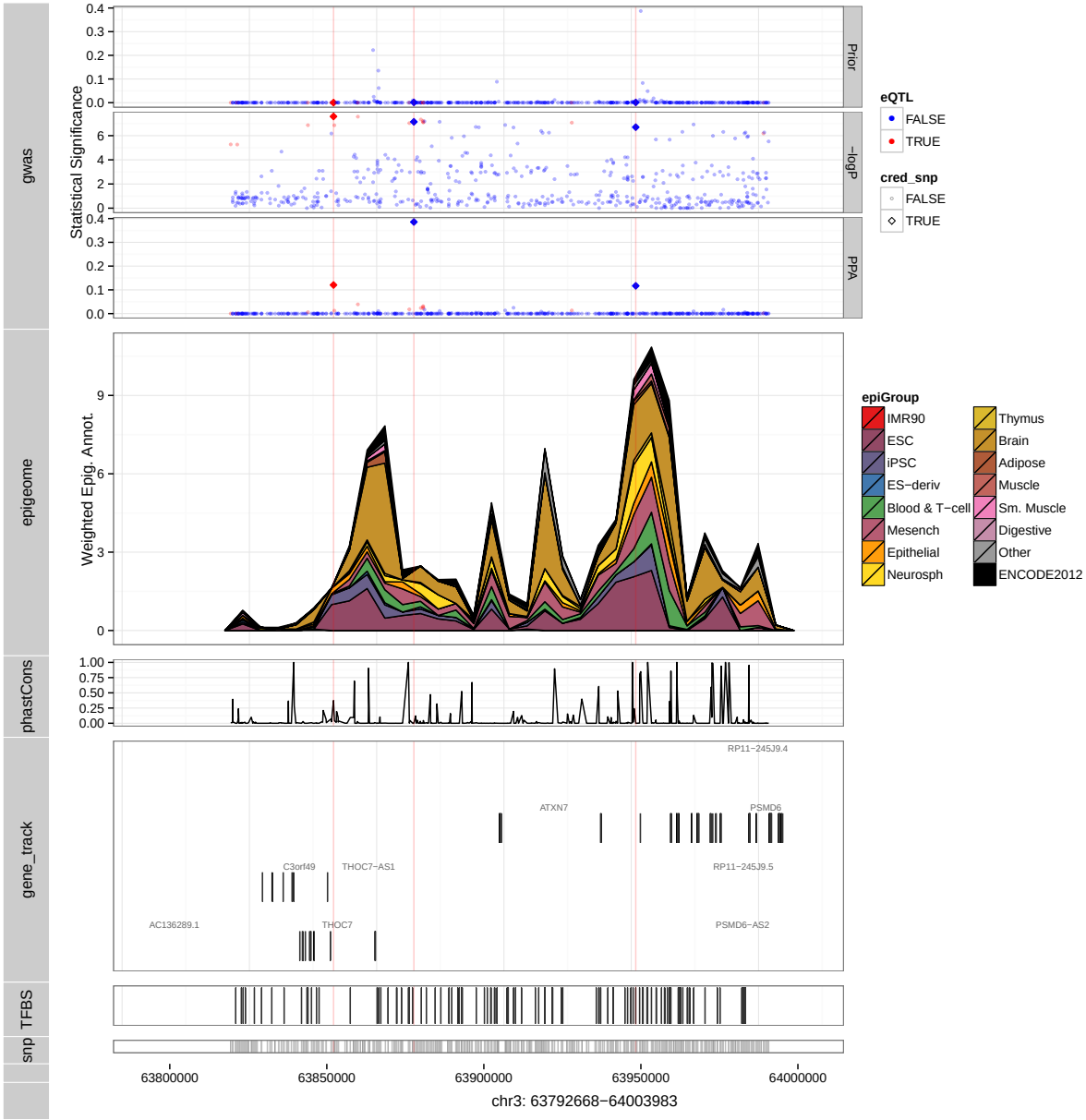
Schizophrenia



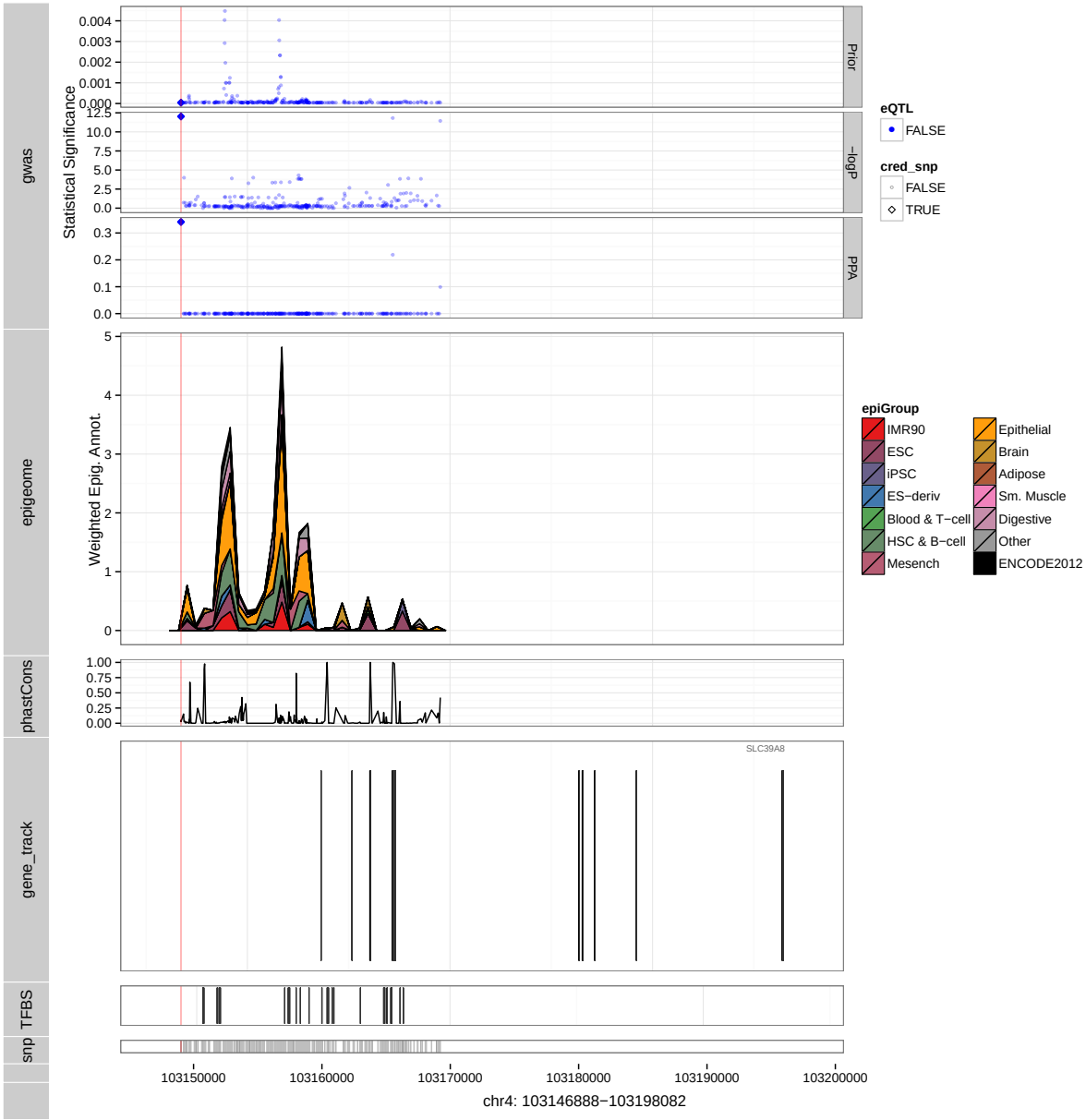
Schizophrenia



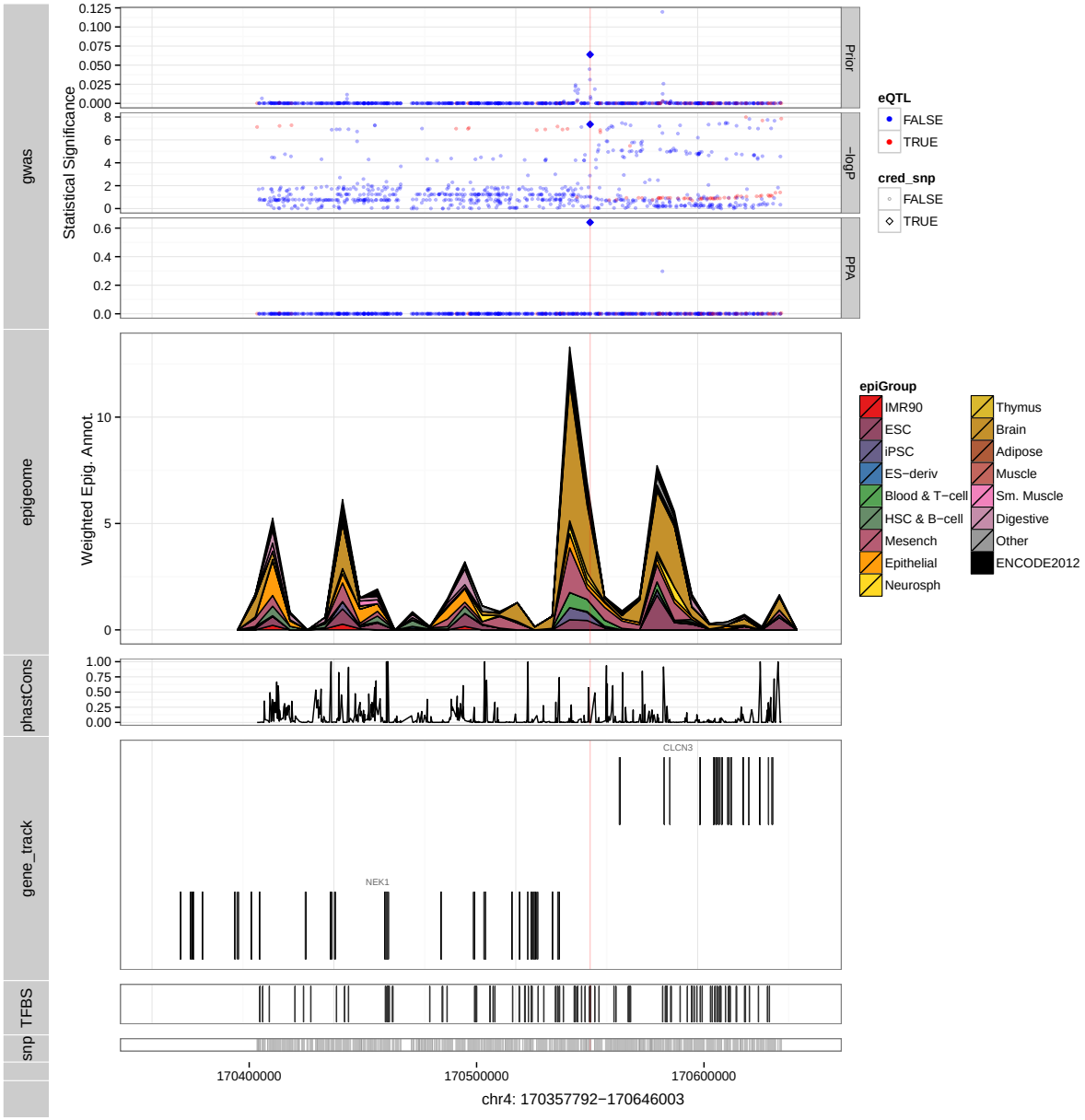
Schizophrenia



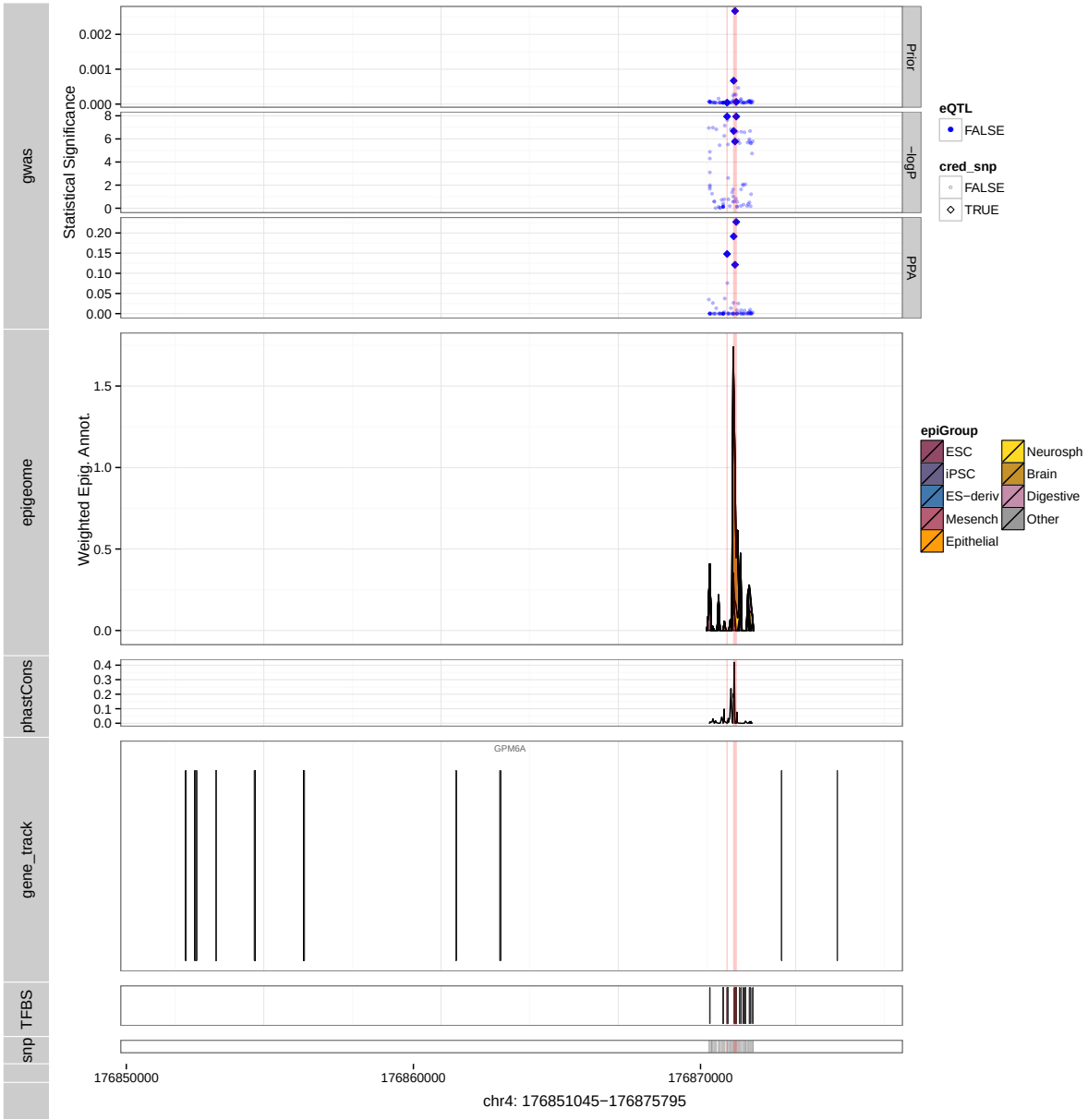
Schizophrenia



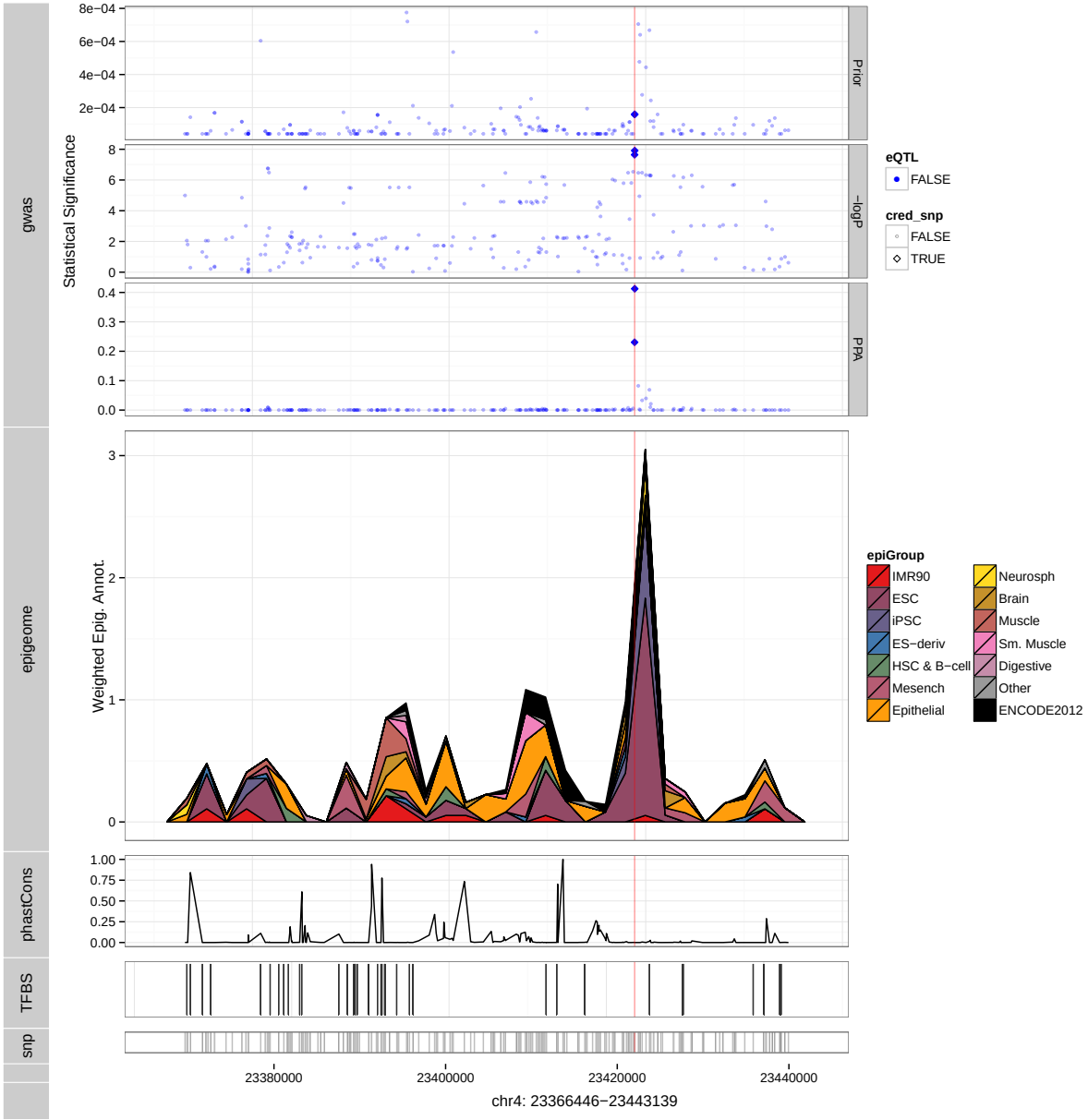
Schizophrenia



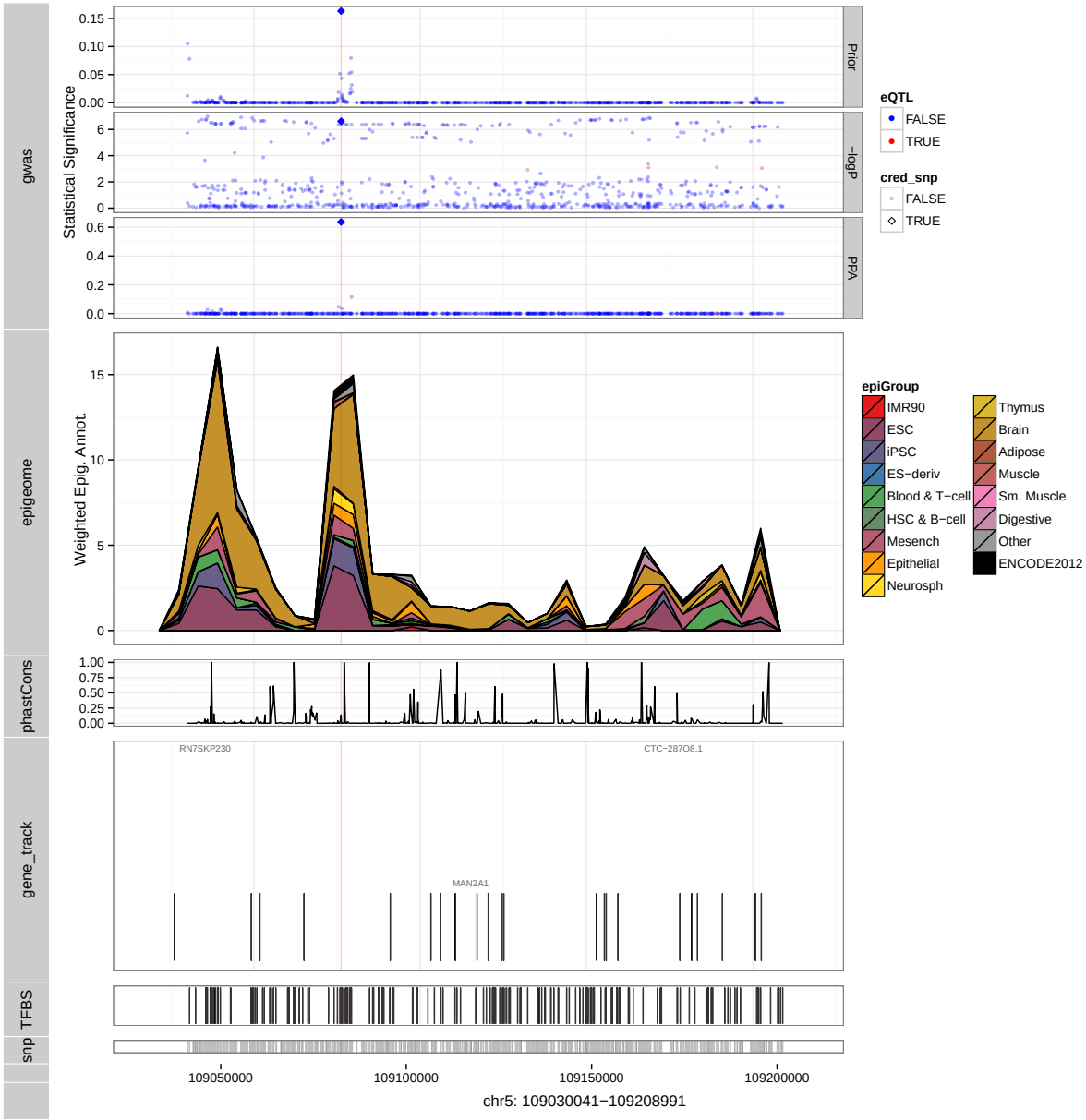
Schizophrenia



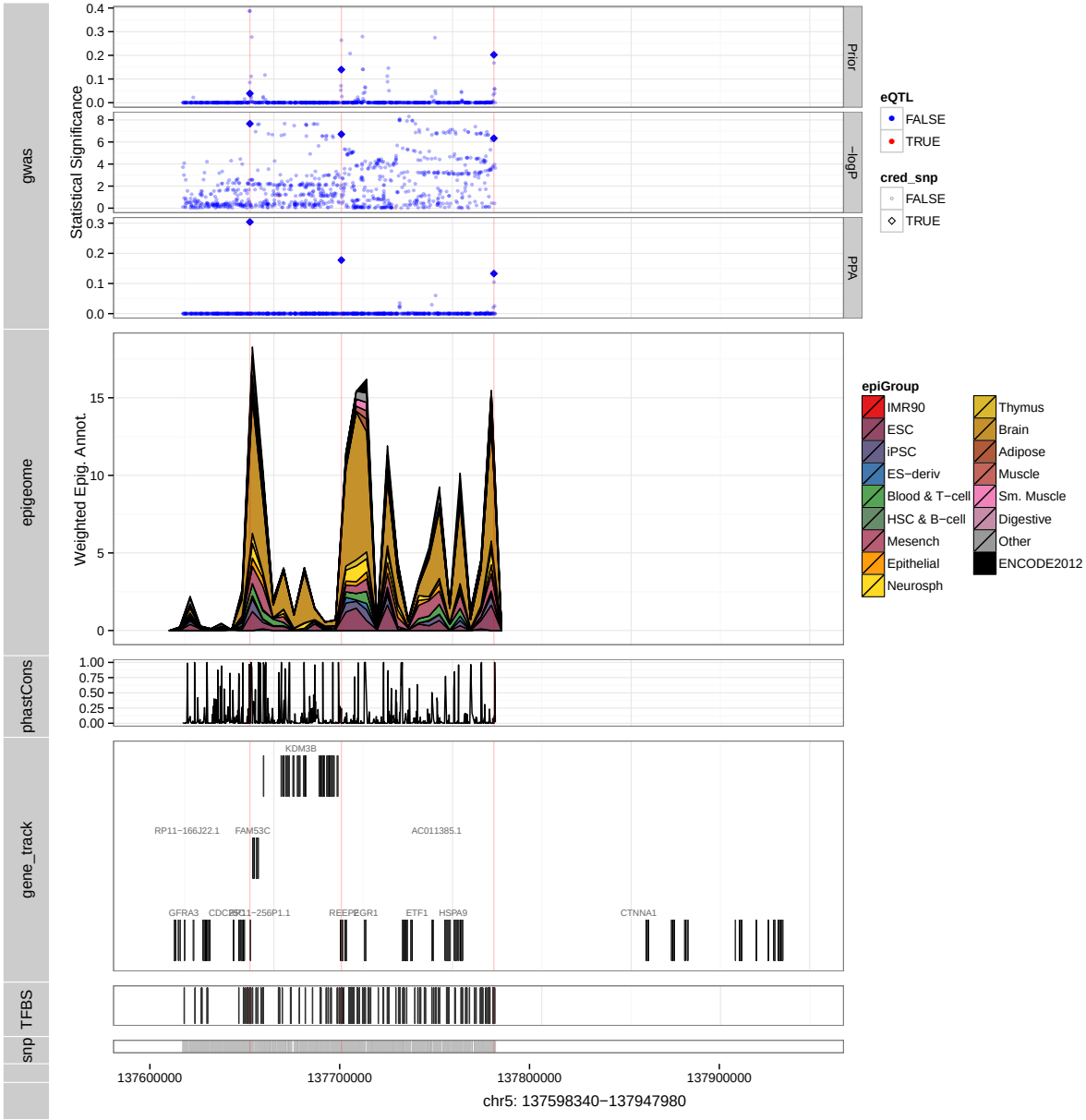
Schizophrenia



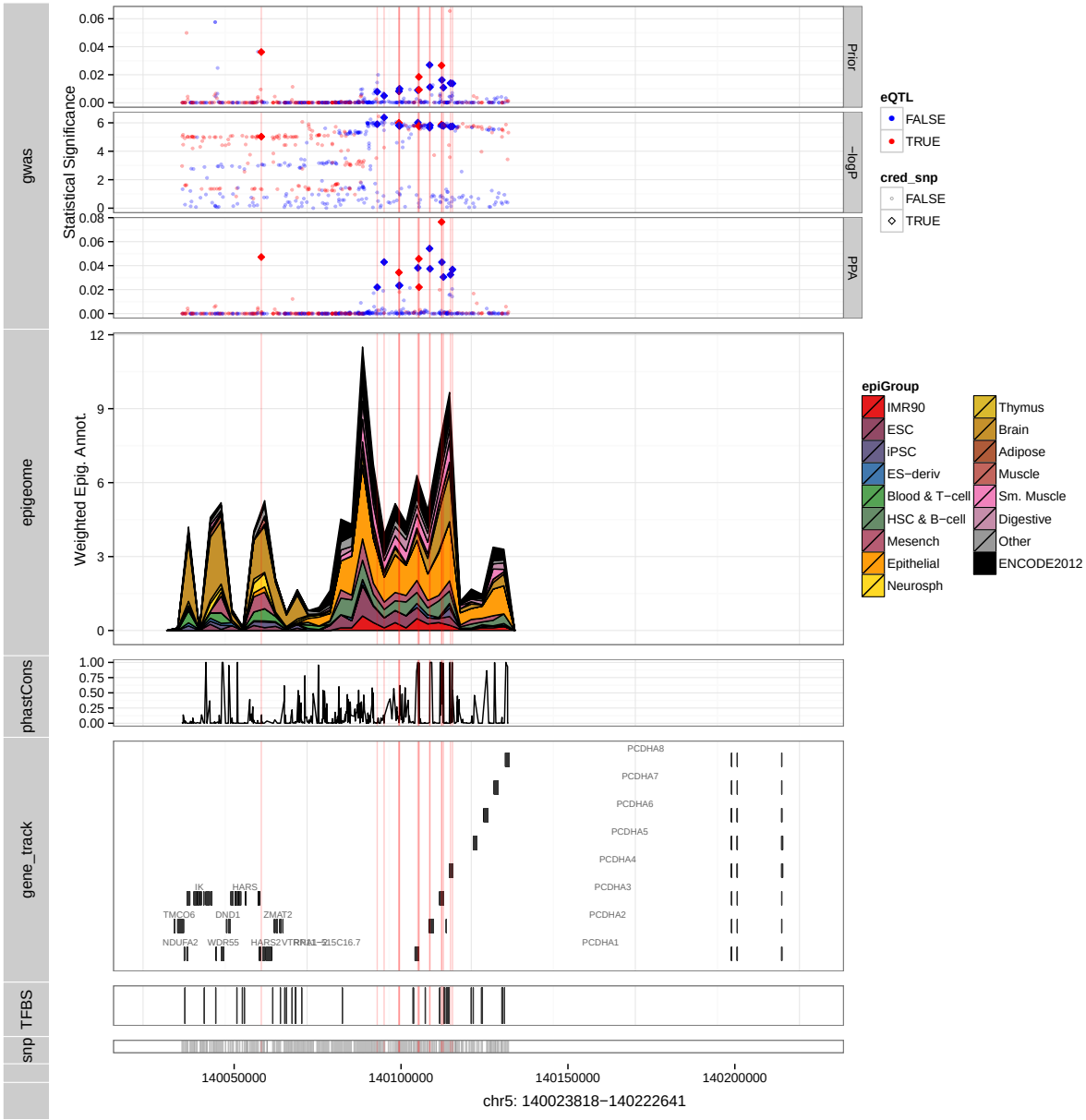
Schizophrenia



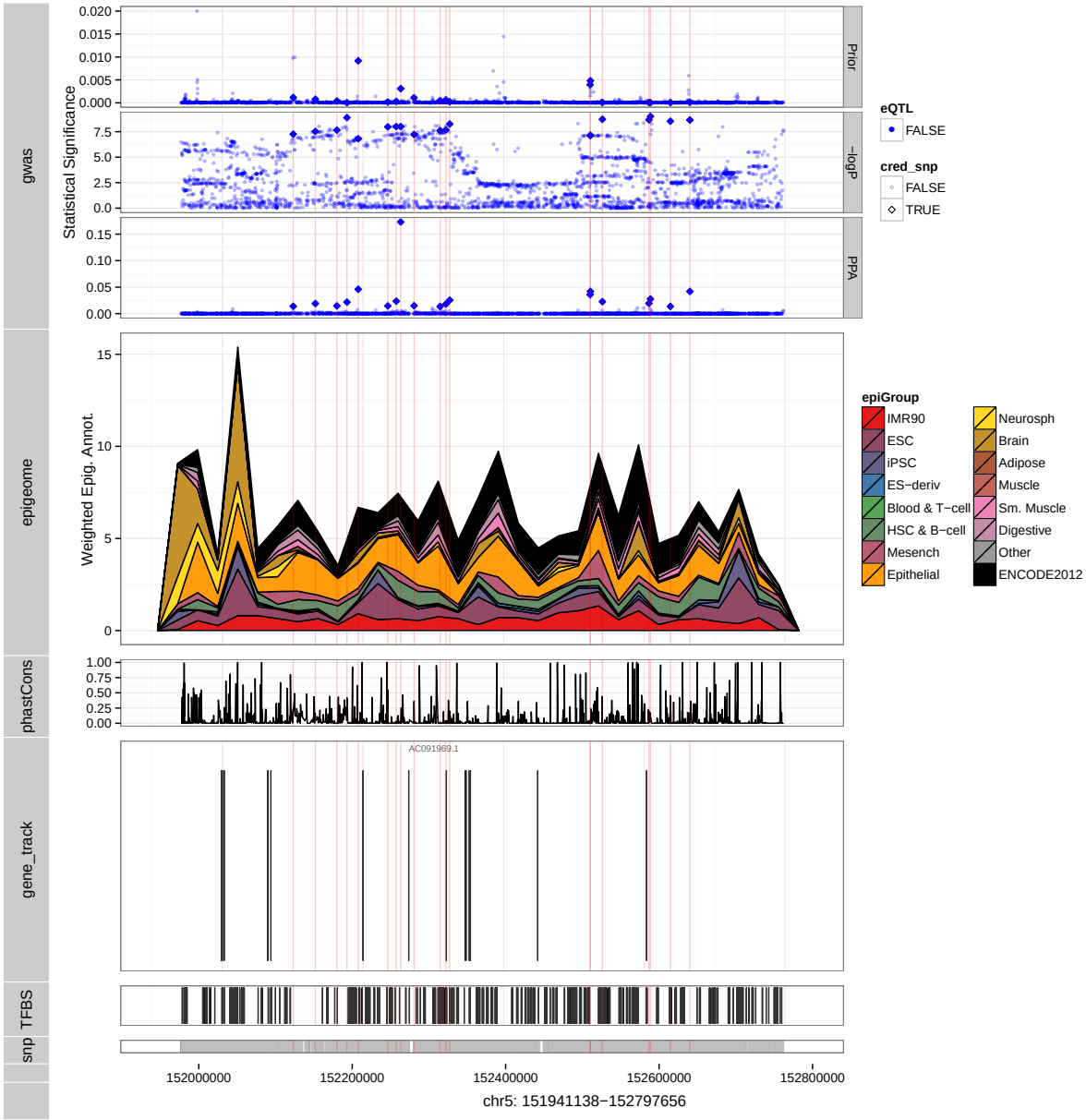
Schizophrenia



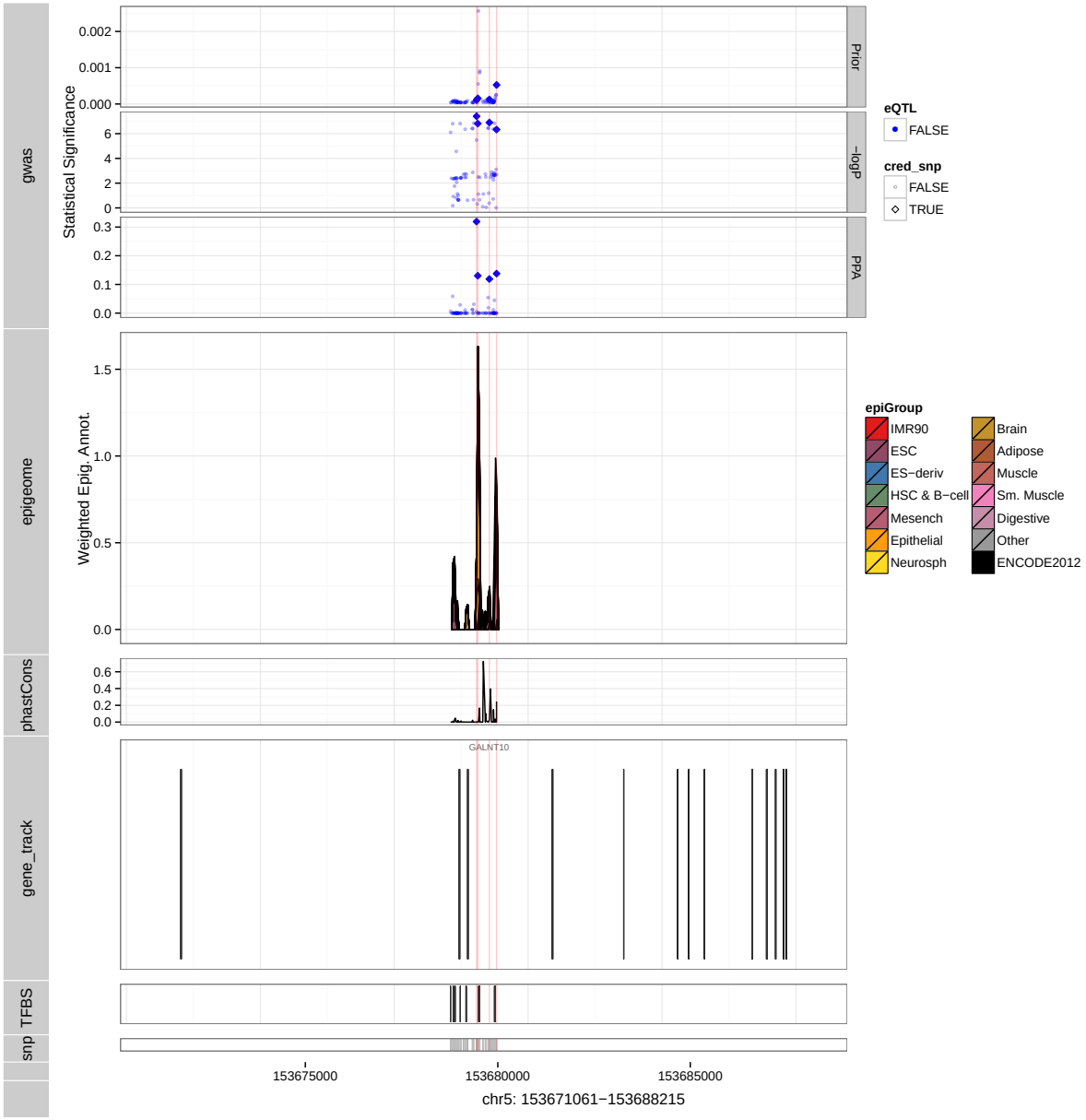
Schizophrenia



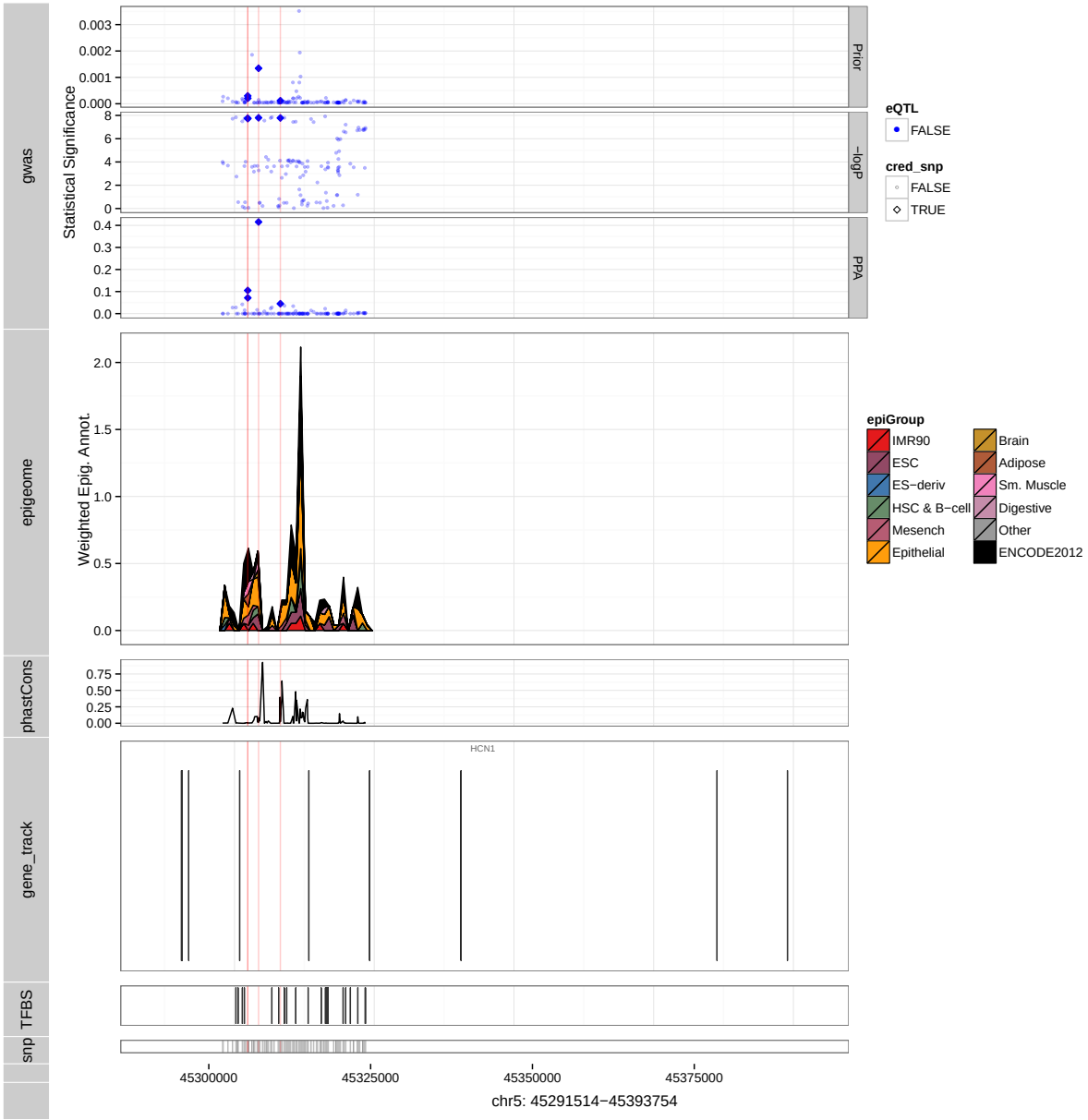
Schizophrenia



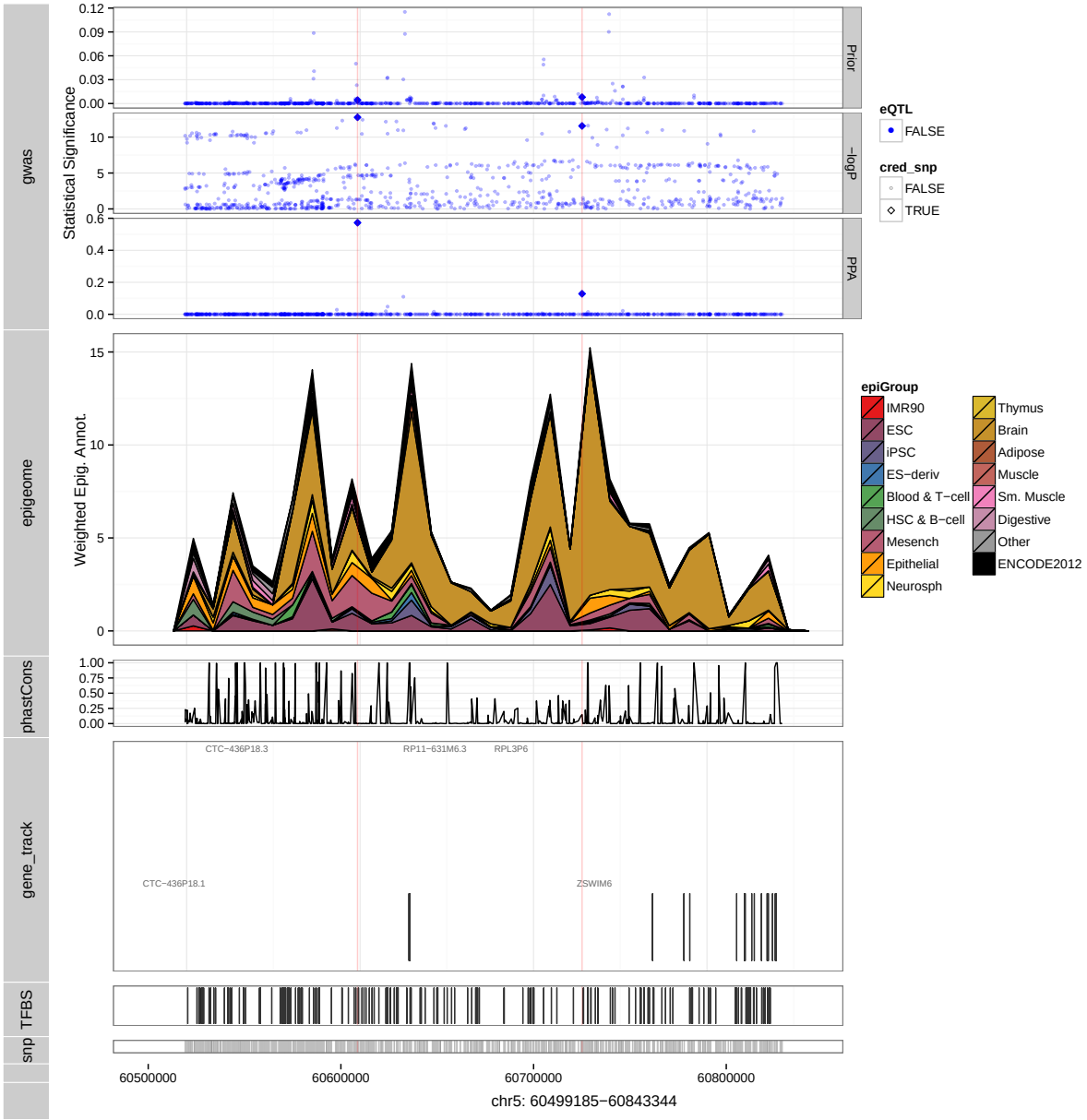
Schizophrenia



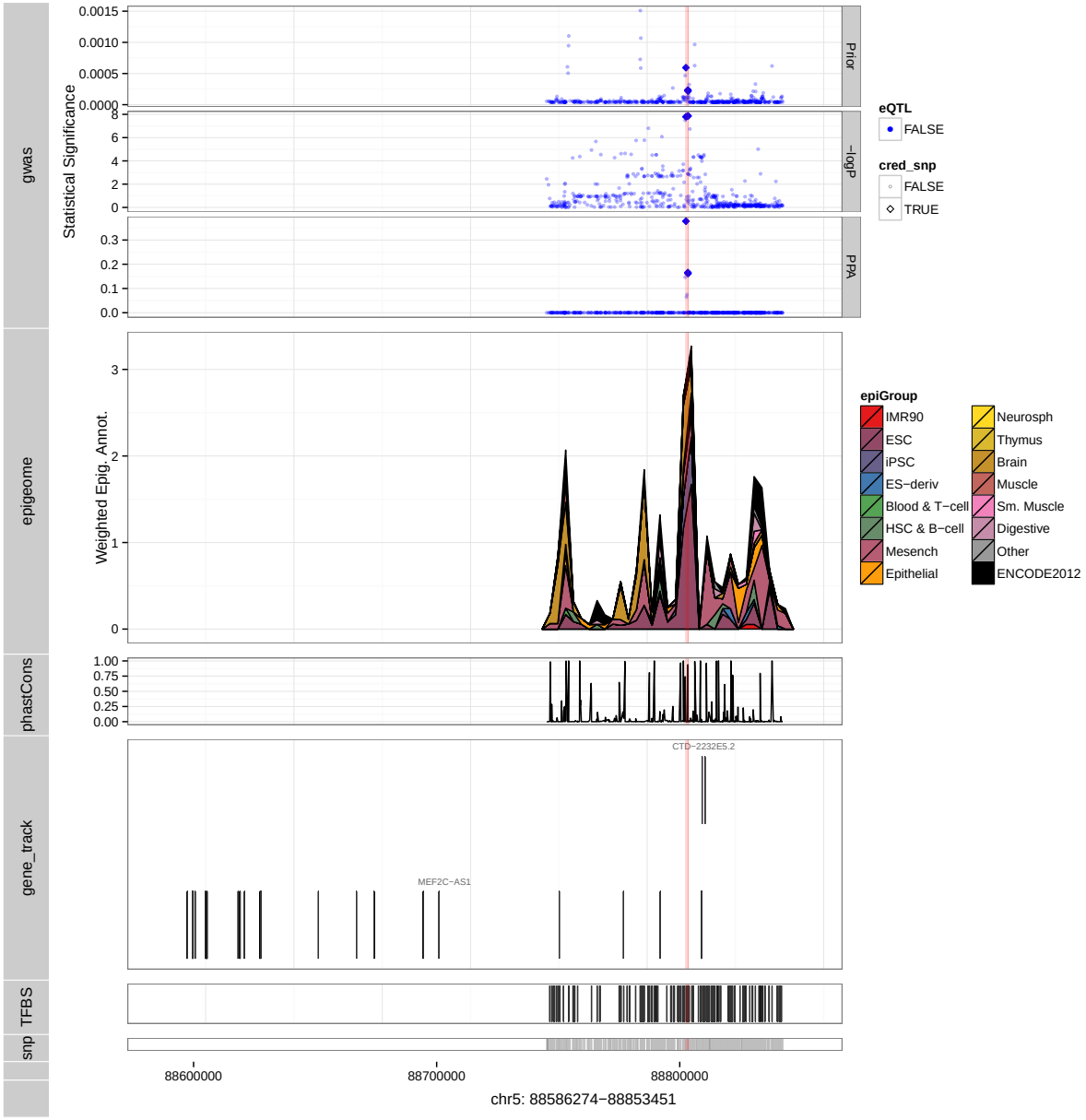
Schizophrenia



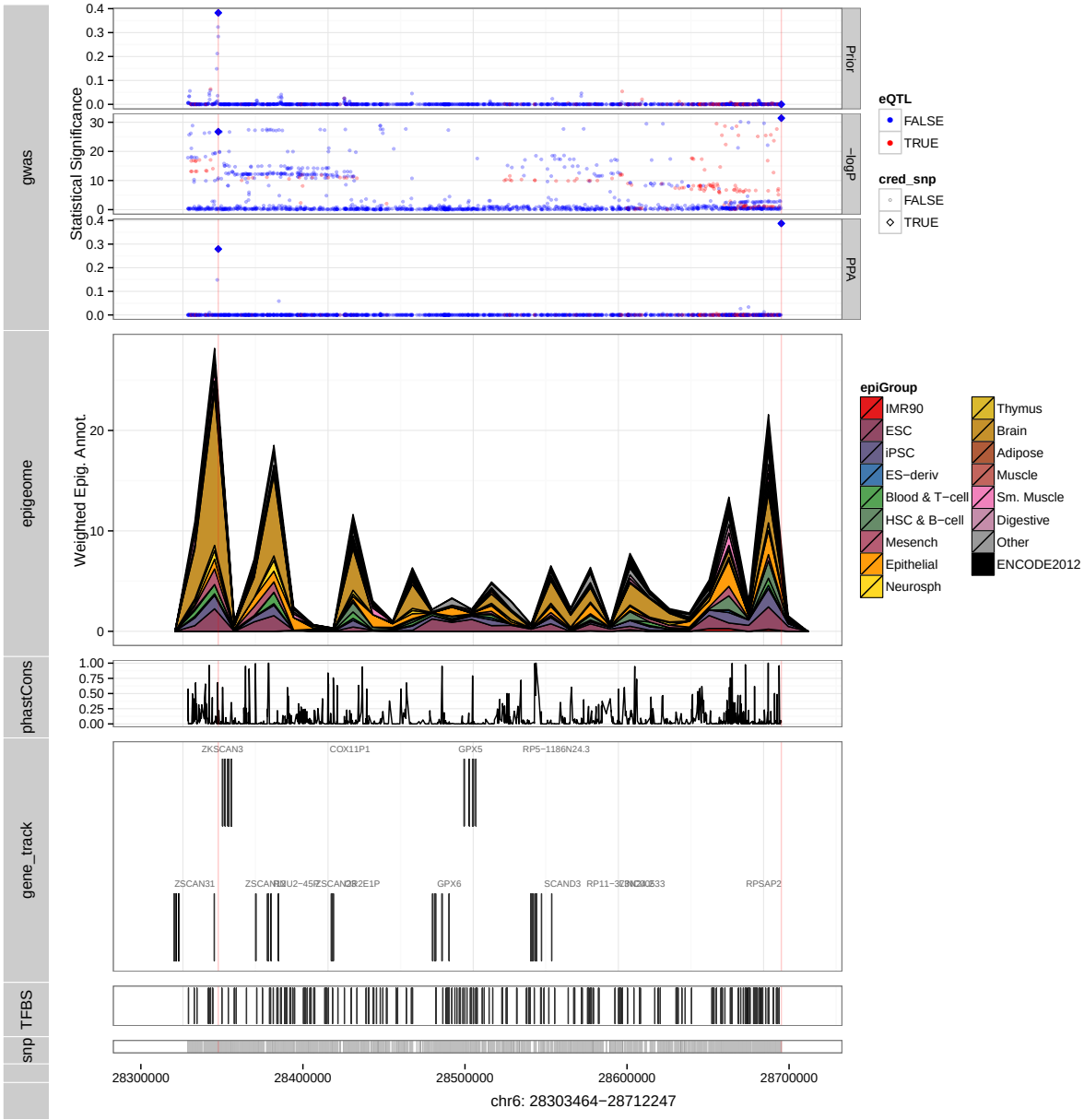
Schizophrenia



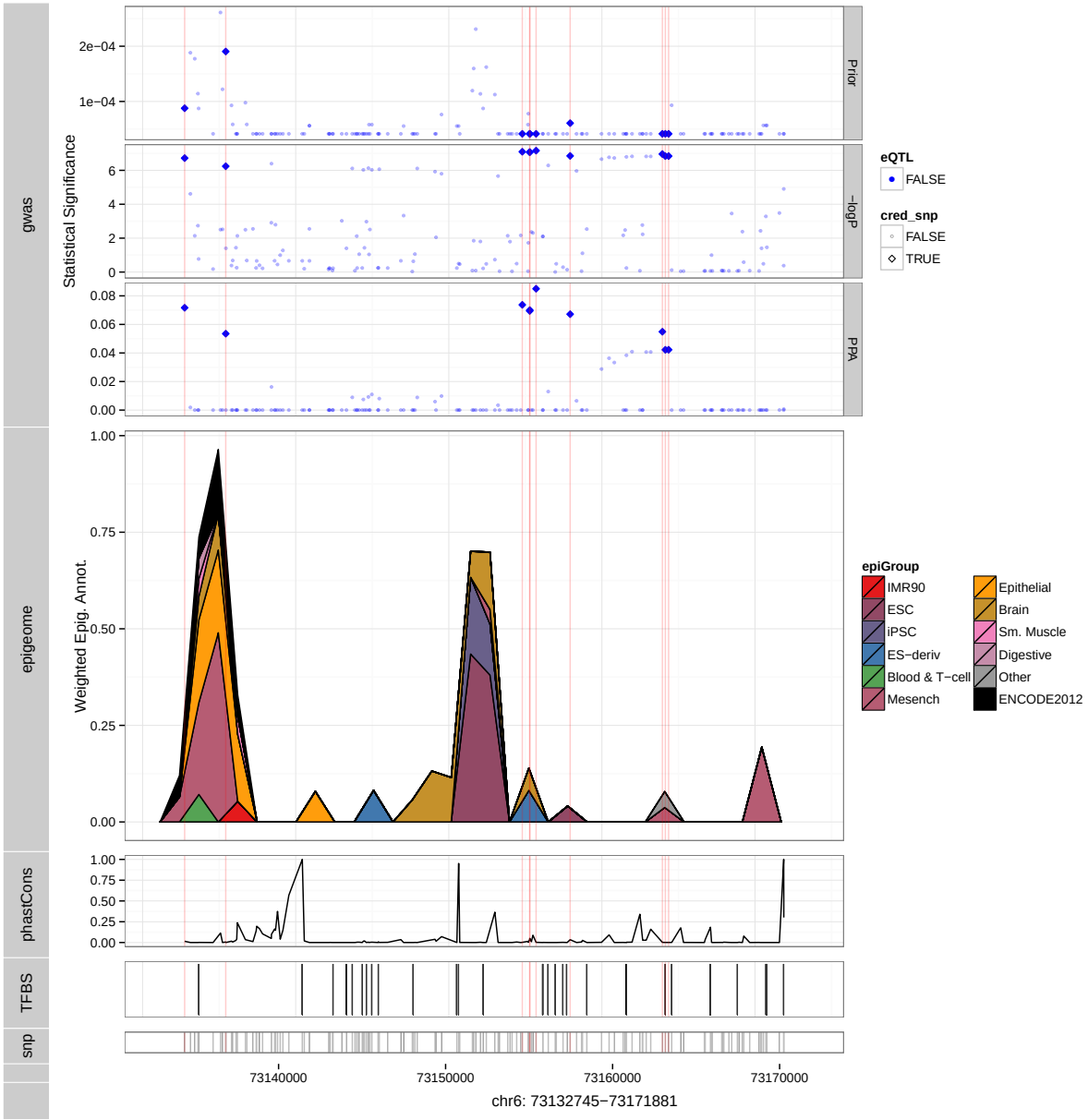
Schizophrenia



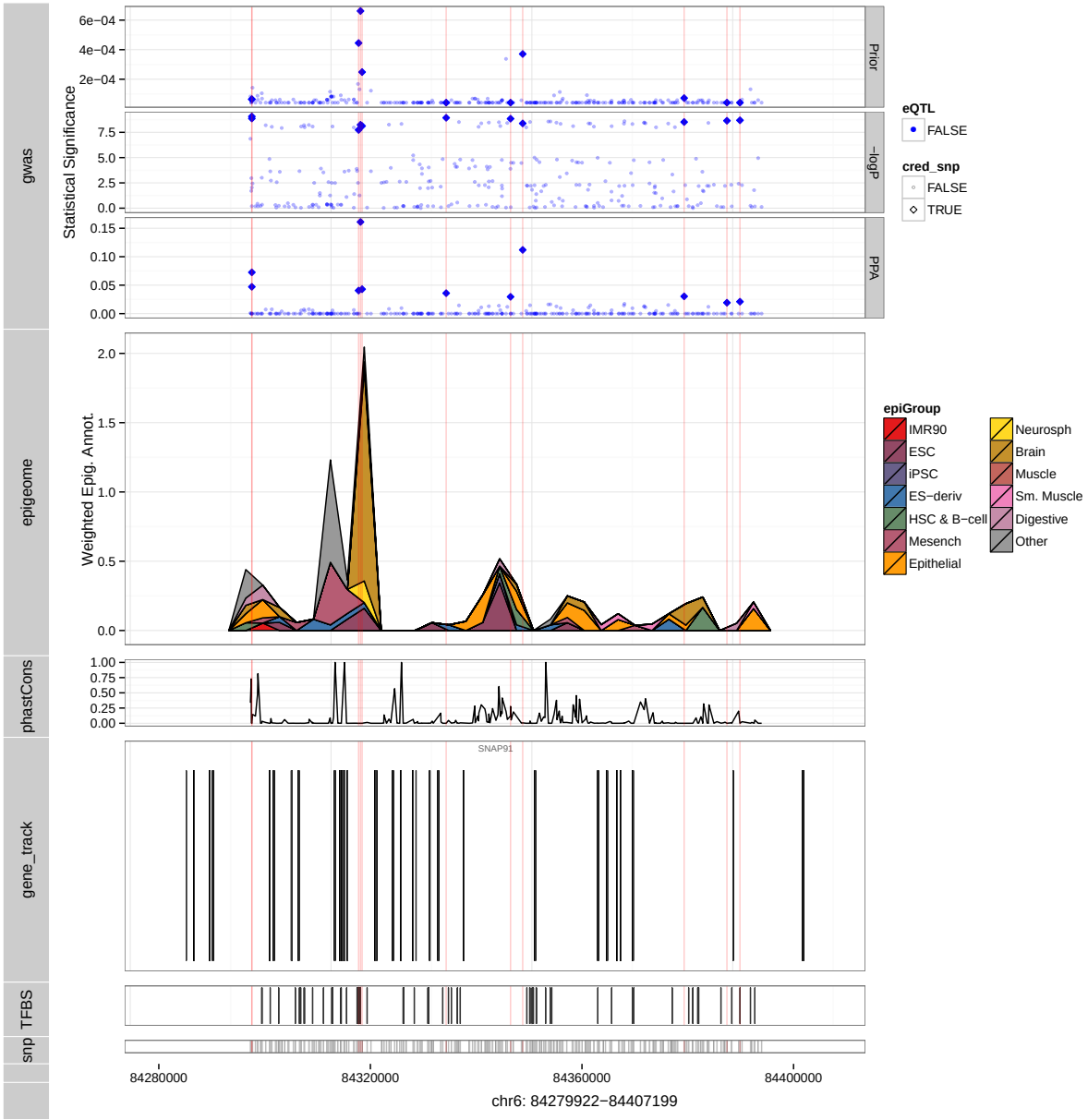
Schizophrenia



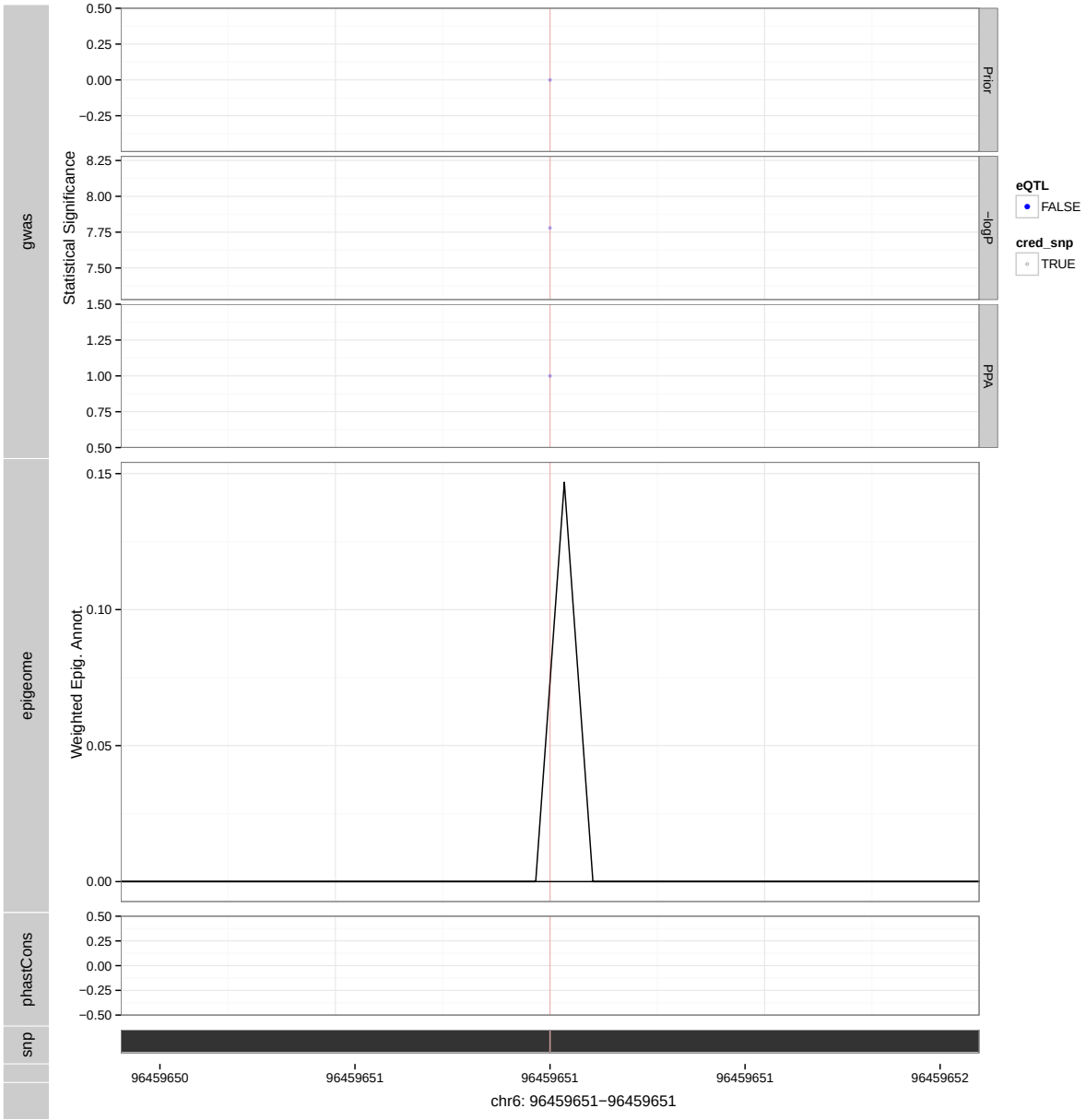
Schizophrenia



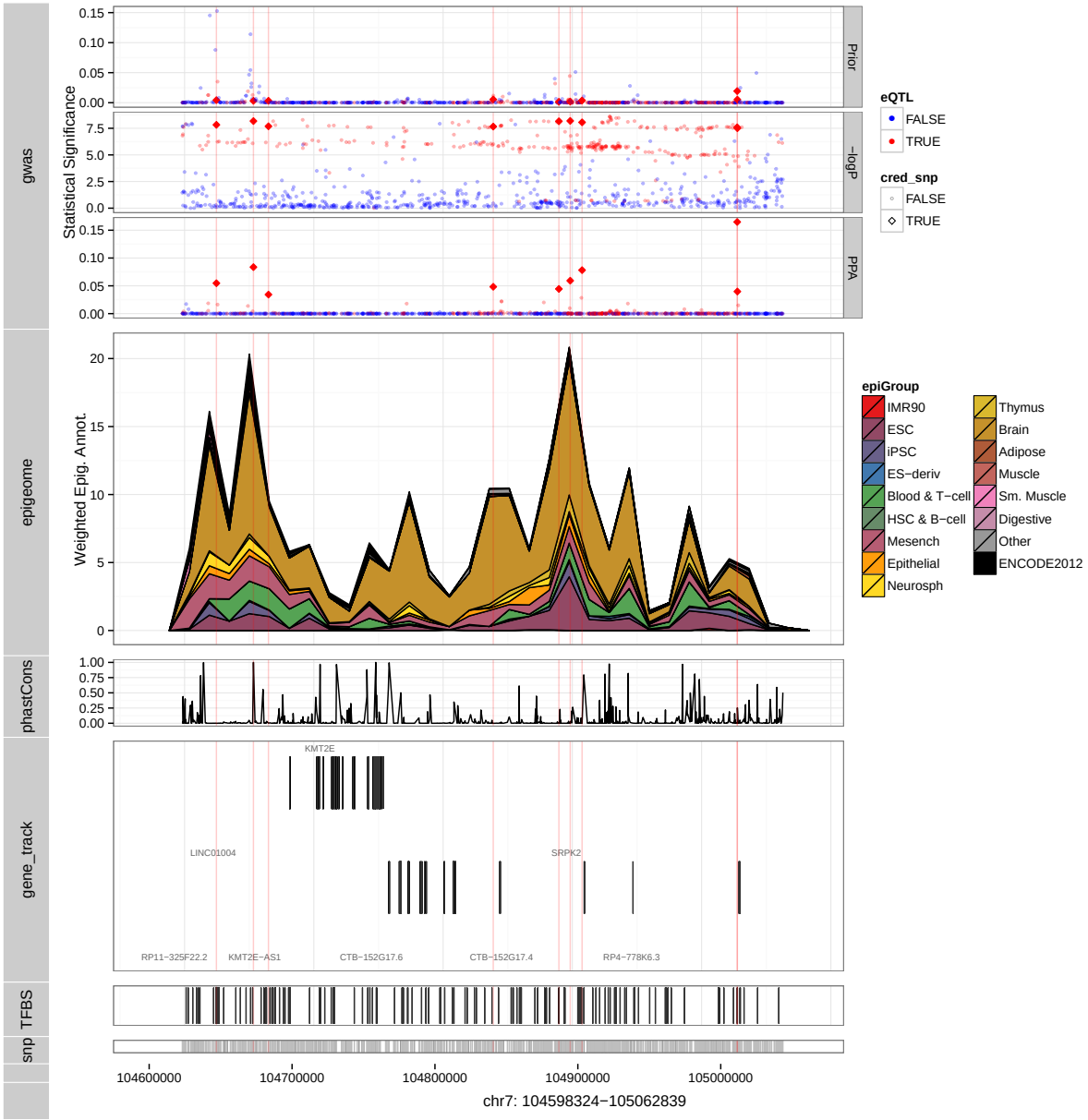
Schizophrenia



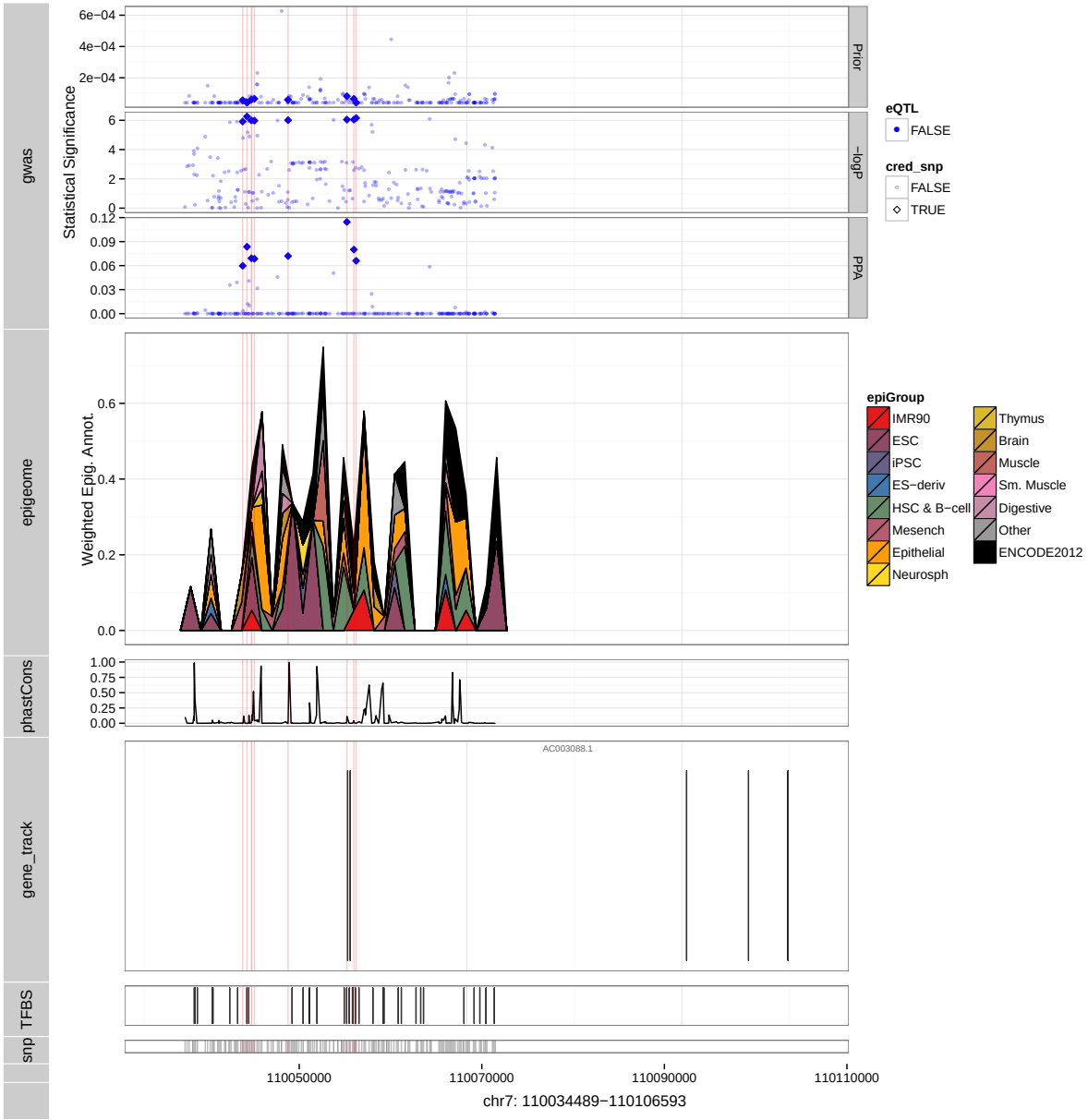
Schizophrenia



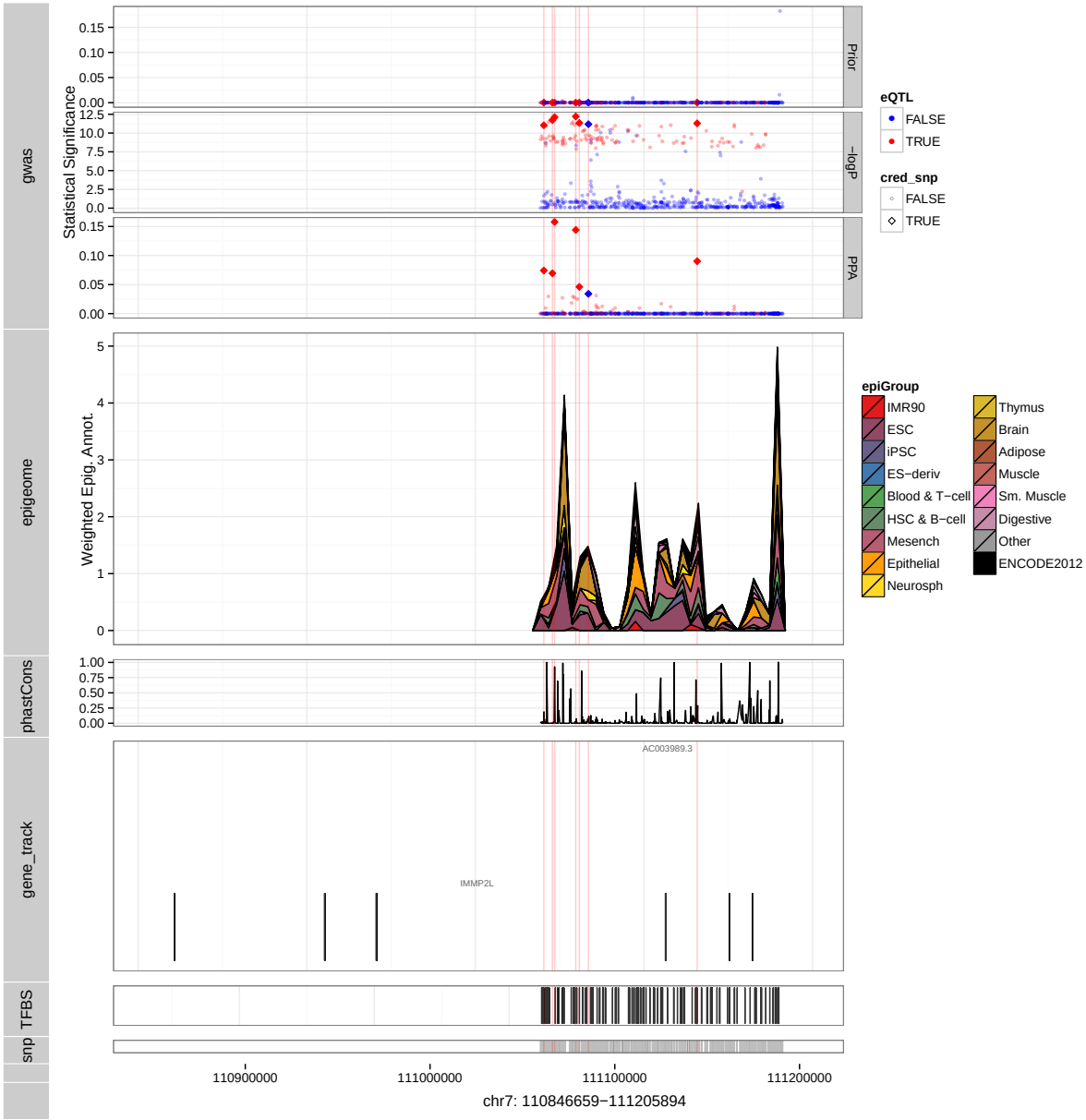
Schizophrenia



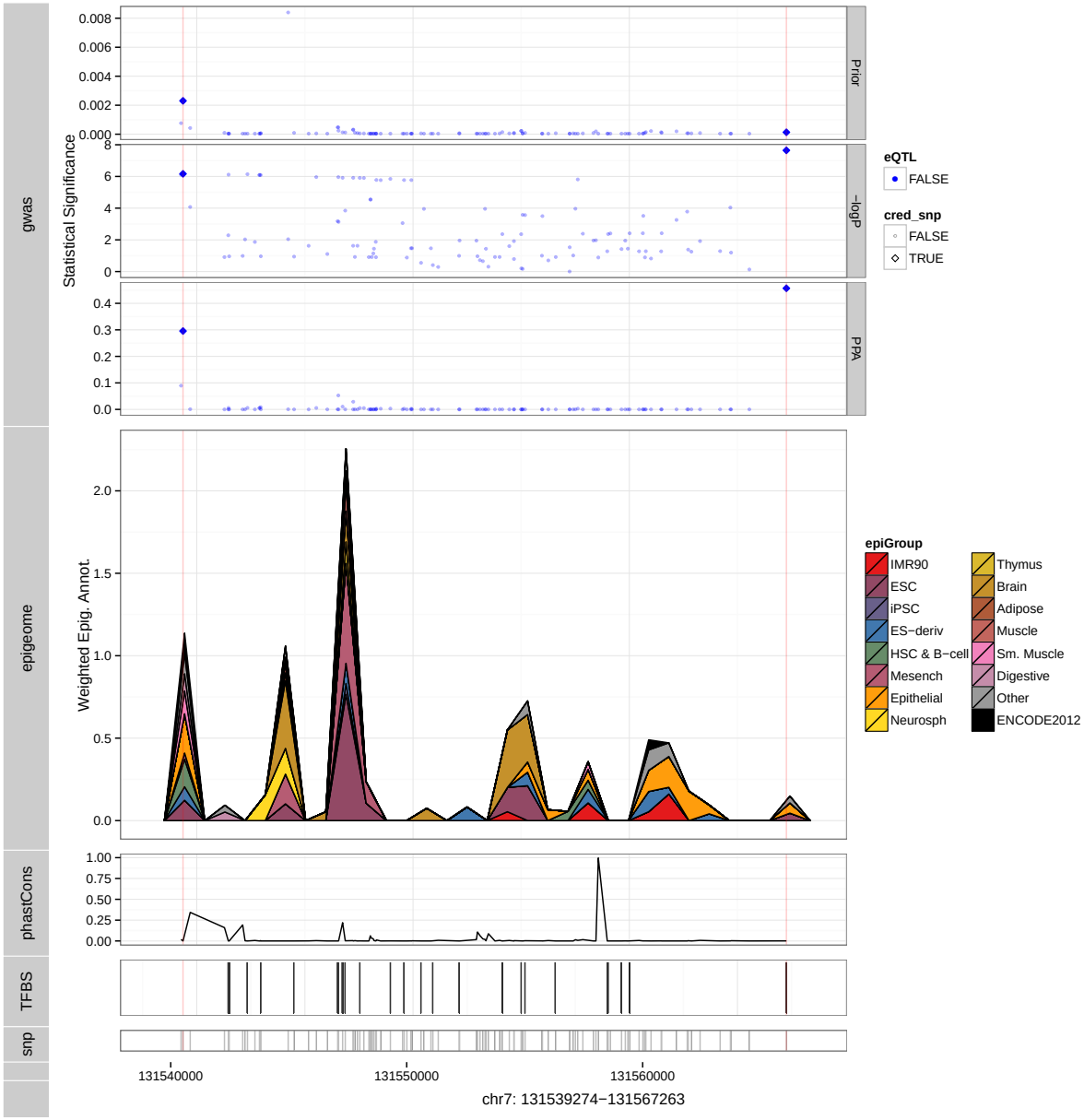
Schizophrenia



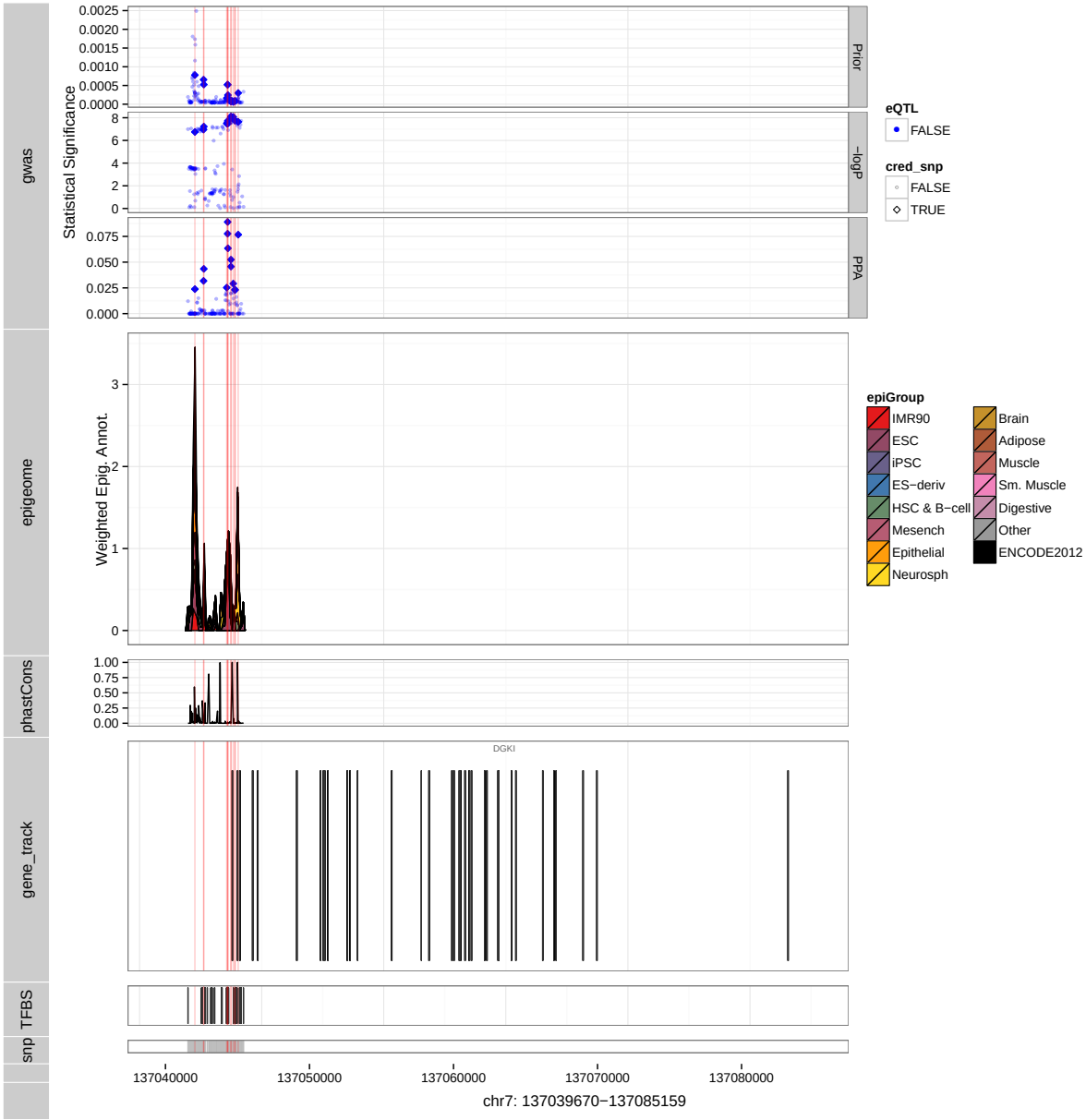
Schizophrenia



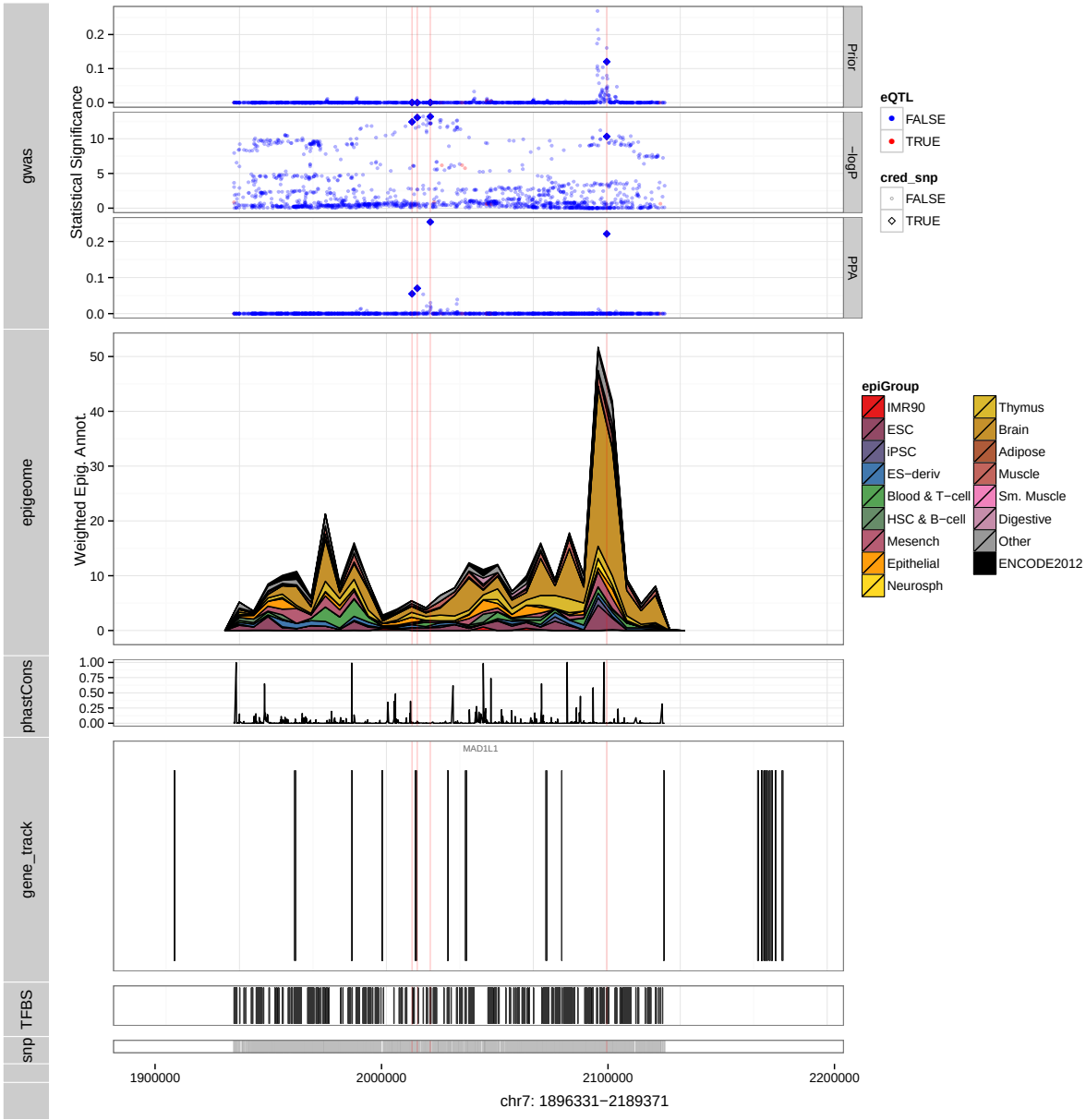
Schizophrenia



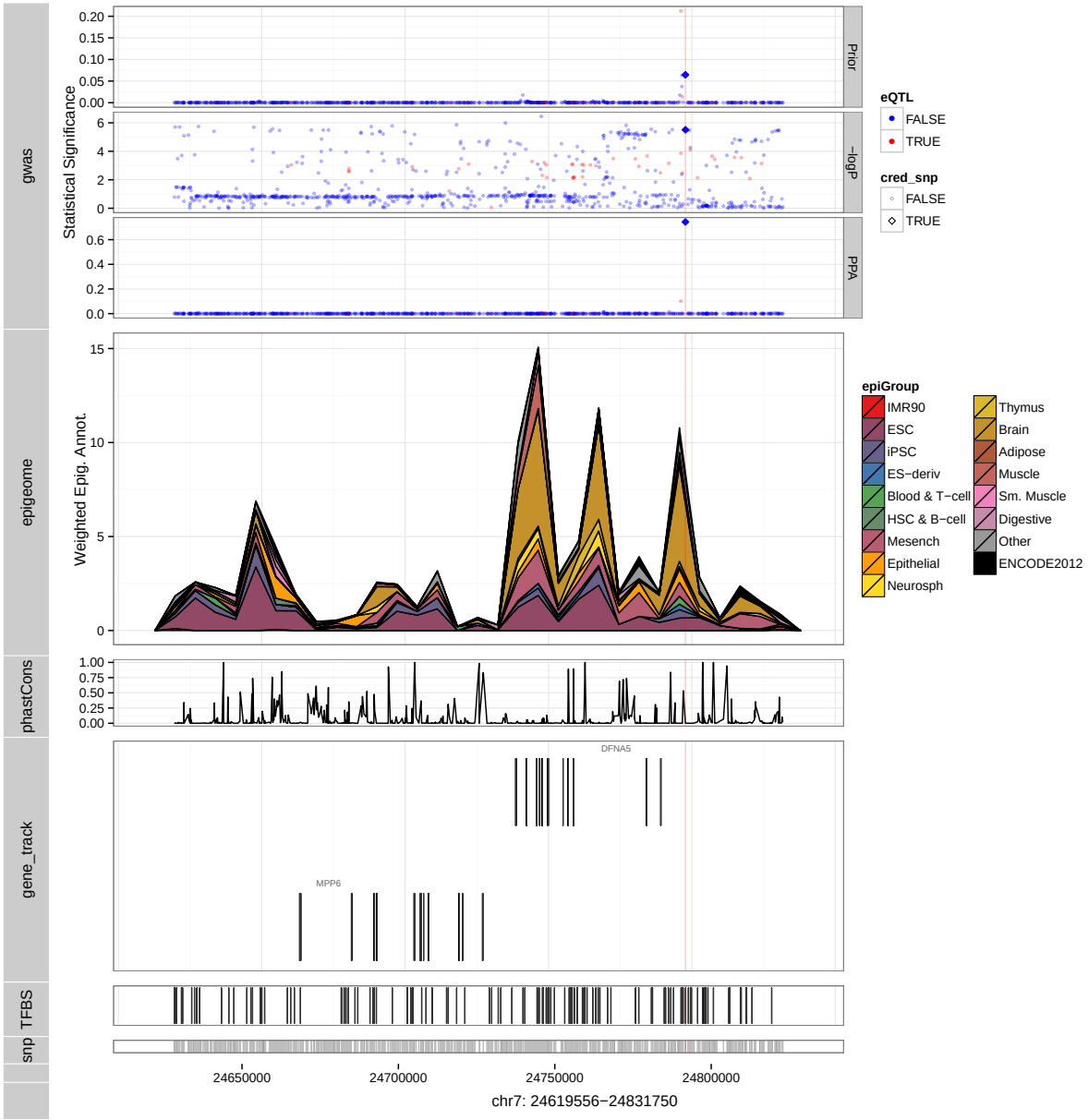
Schizophrenia



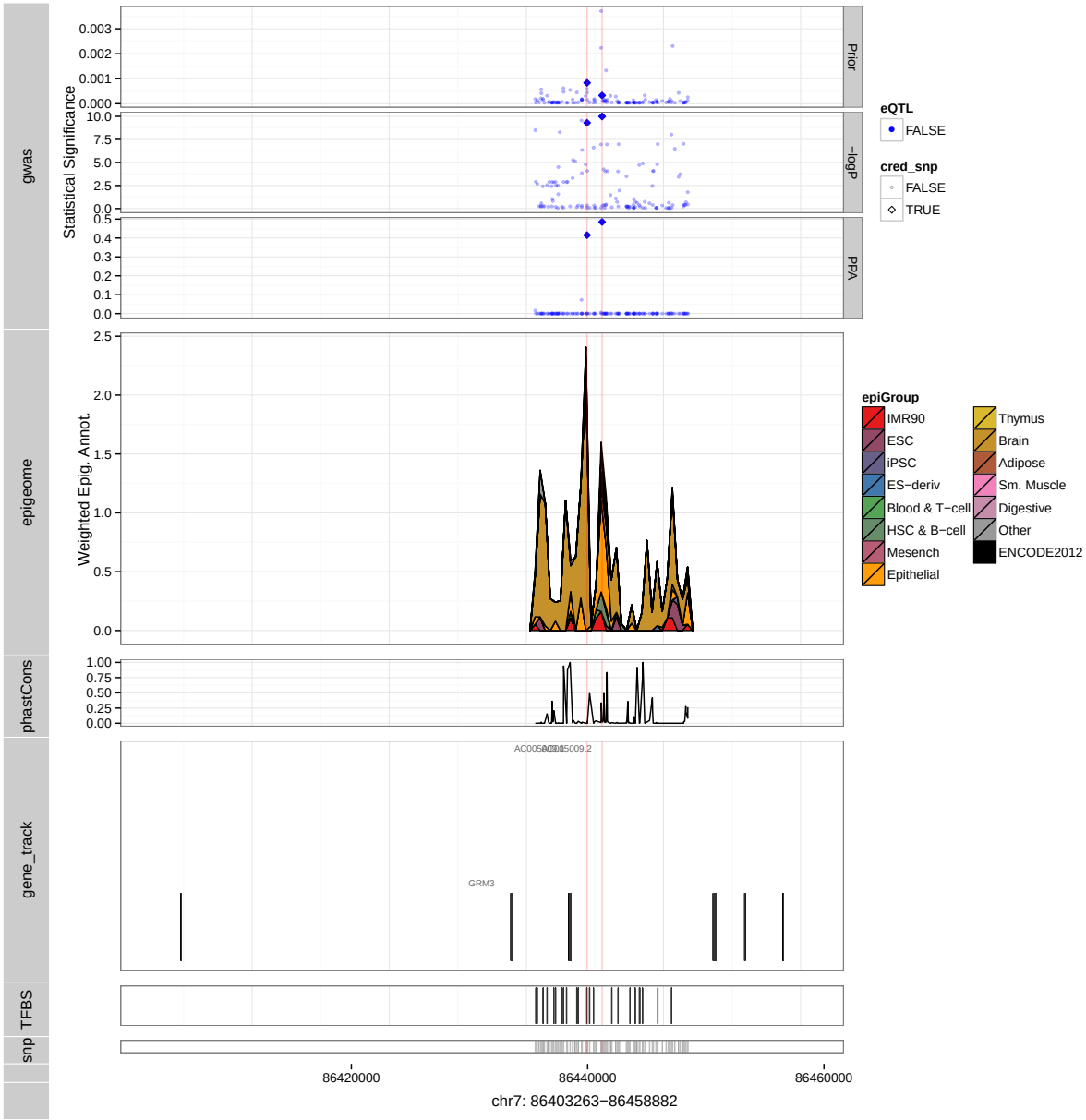
Schizophrenia



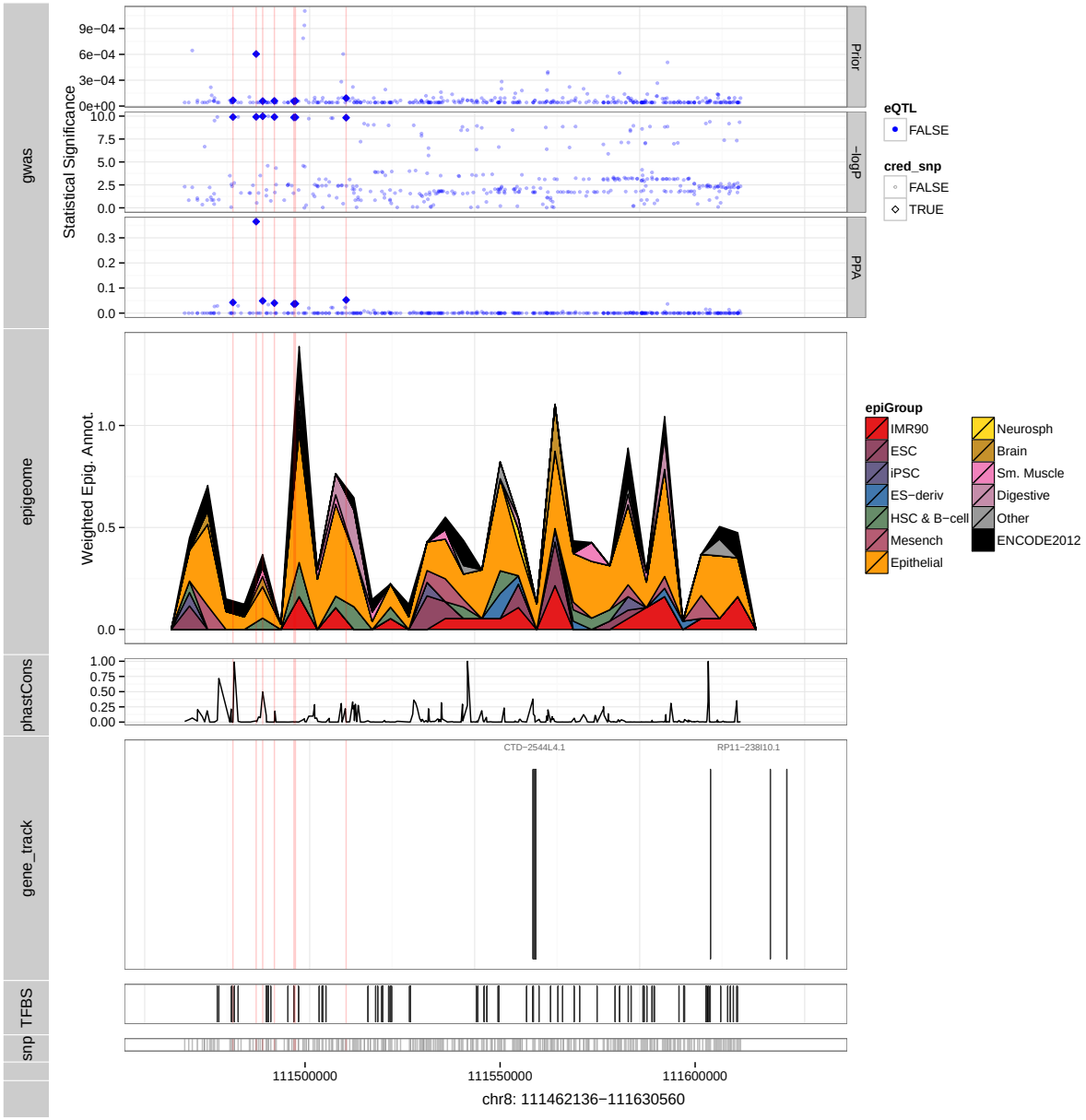
Schizophrenia

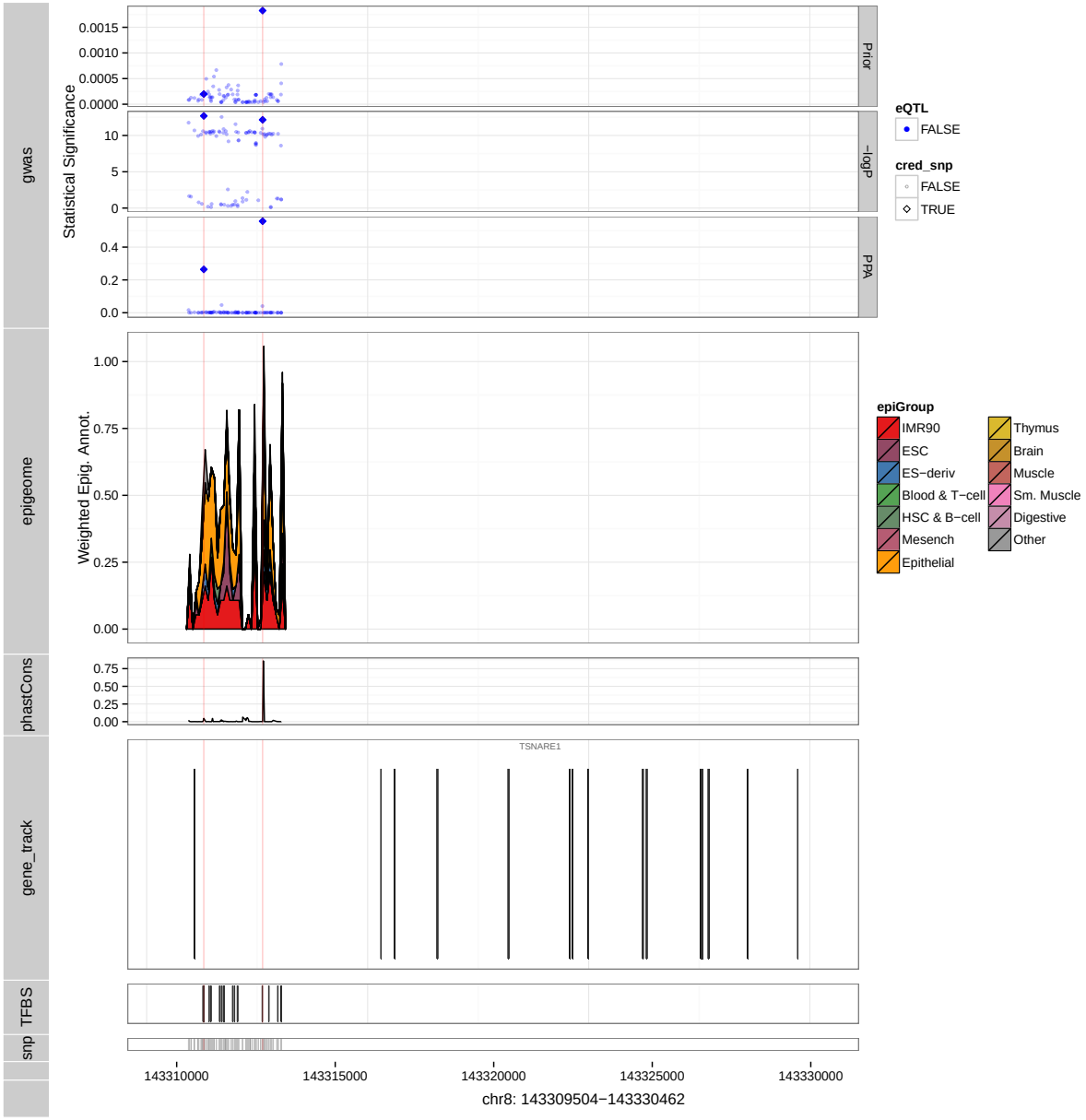


Schizophrenia

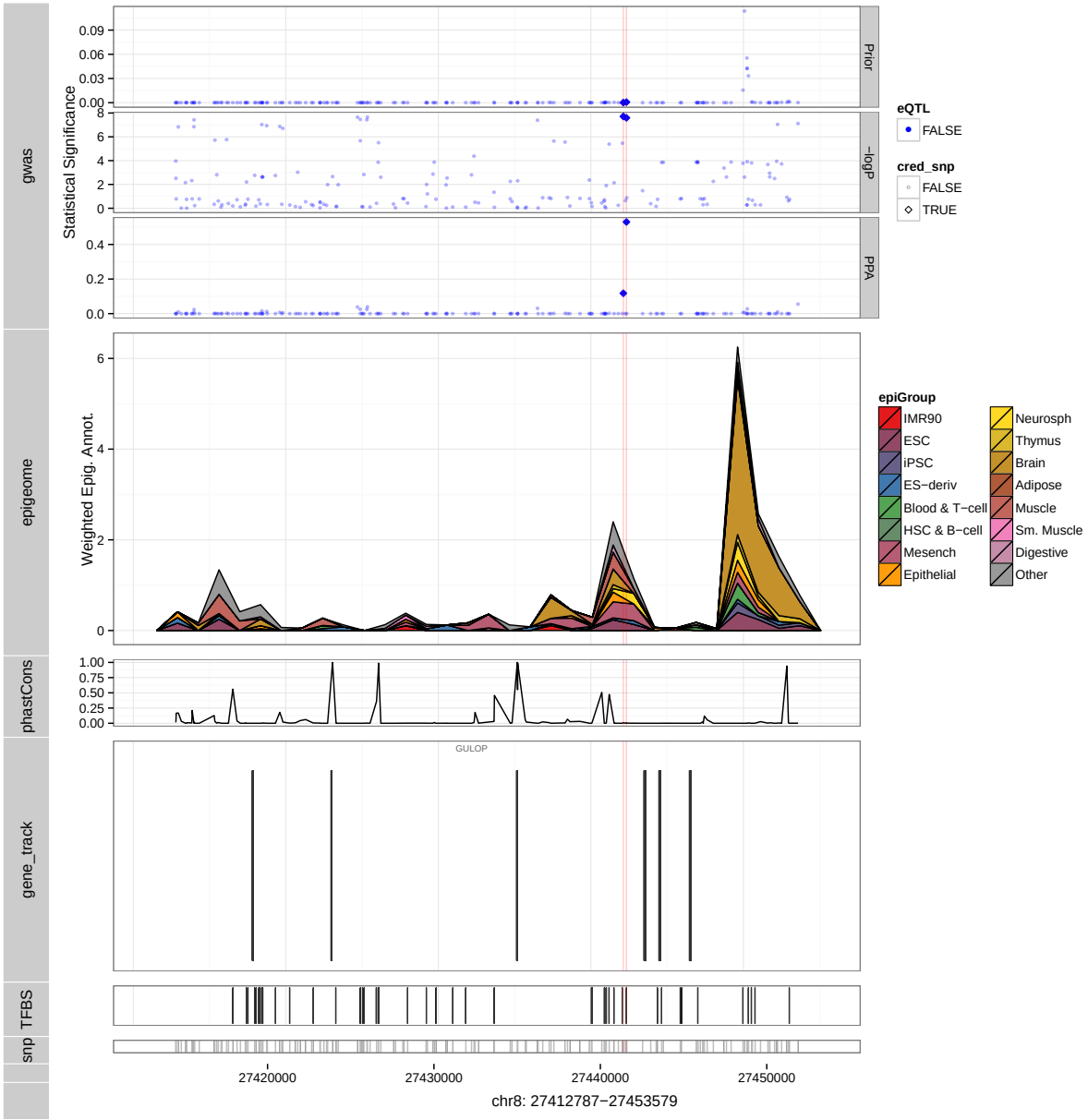


Schizophrenia

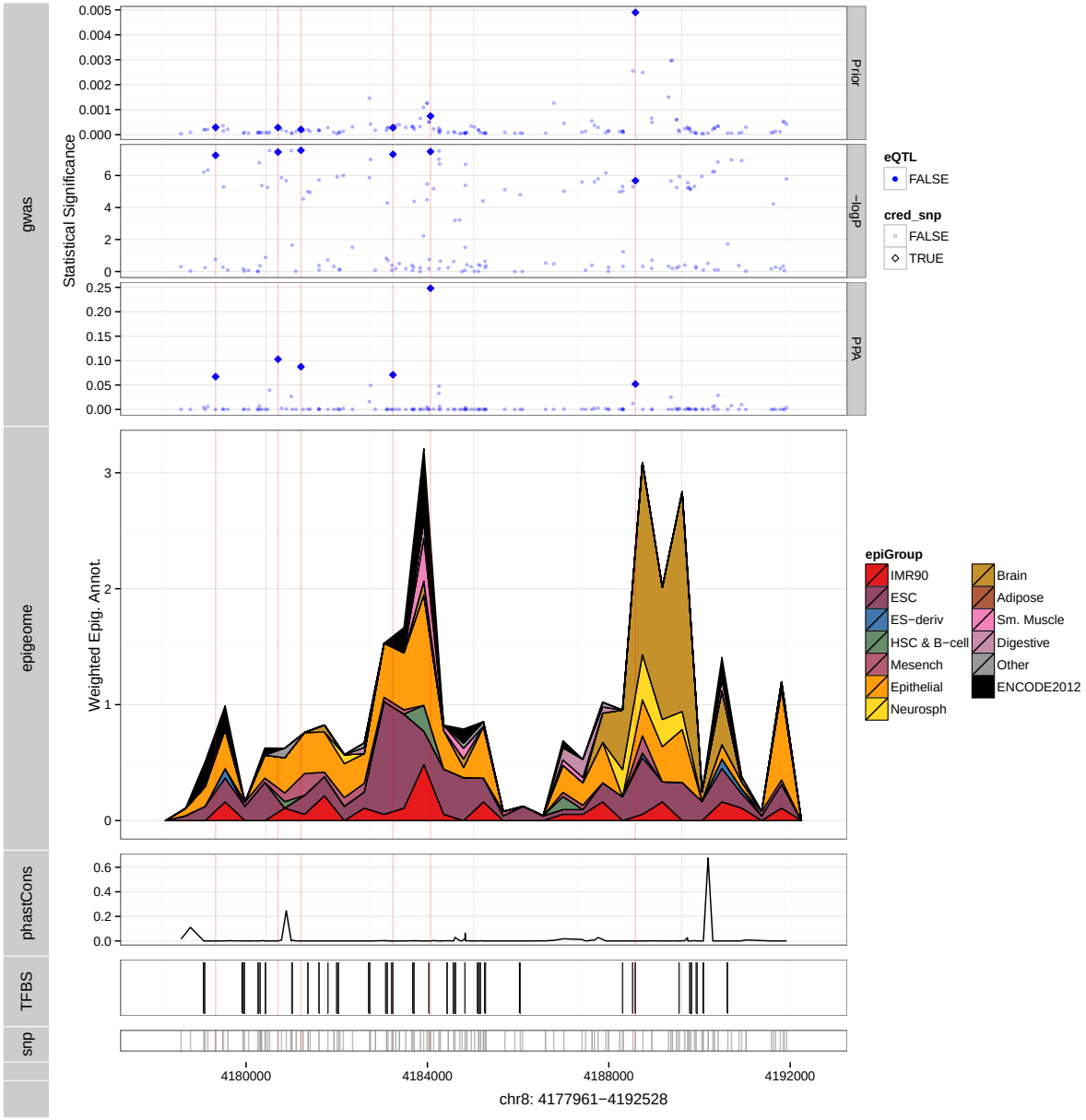



gwas
epigeome
phastCons
gene_track
TFBS
snp
Statistical Significance
Prior
−logP
PPA
Weighted Epig. Annot.
TSNARE1
eQTL
FALSE
cred_snp
FALSE
TRUE
epiGroup
IMR90
ESC
ES−deriv
Blood & T−cell
HSC & B−cell
Mesench
Epithelial
Thymus
Brain
Muscle
Sm. Muscle
Digestive
Other
143310000
143315000
143320000
143325000
143330000
chr8: 143309504−143330462

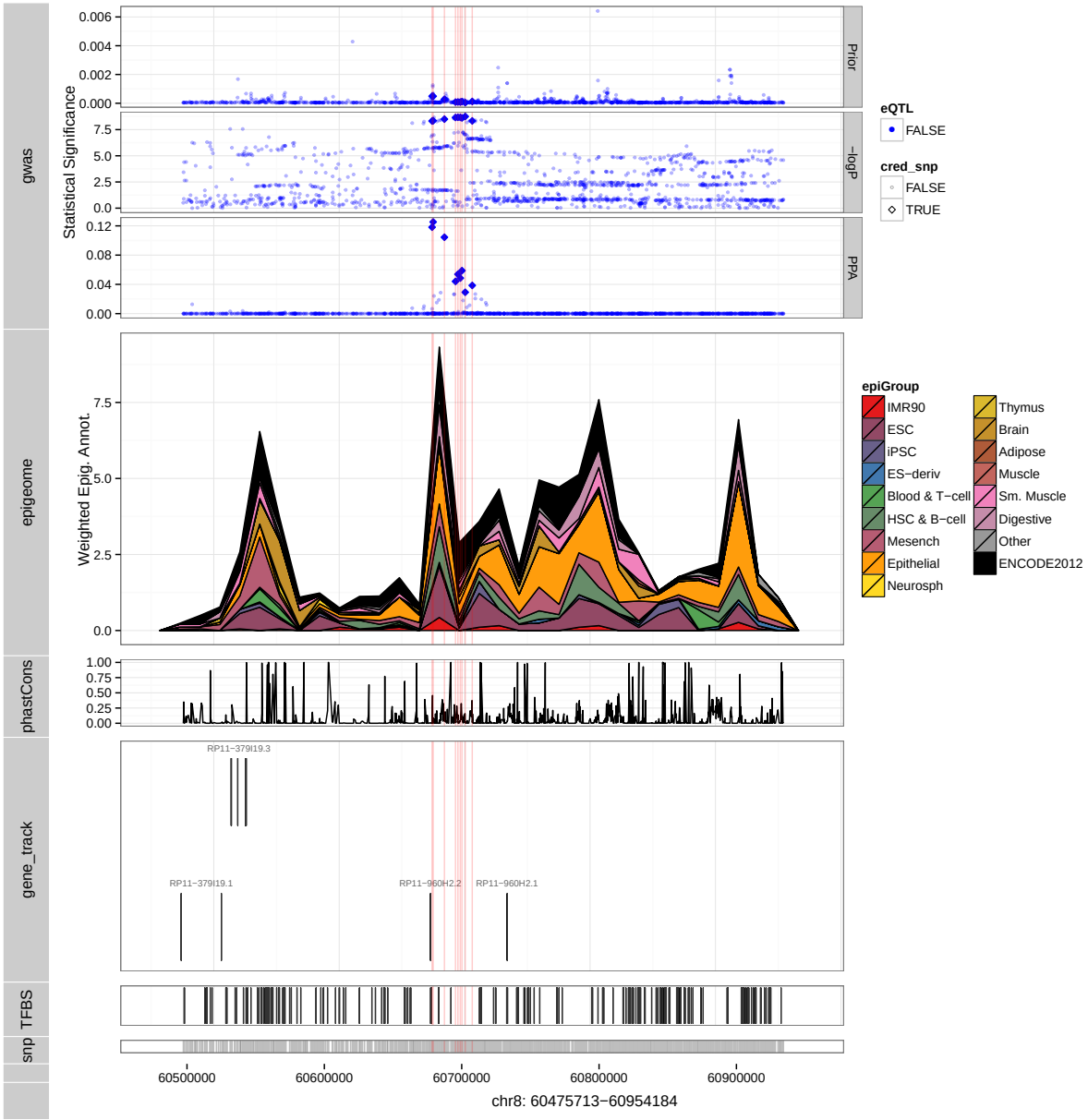
Schizophrenia



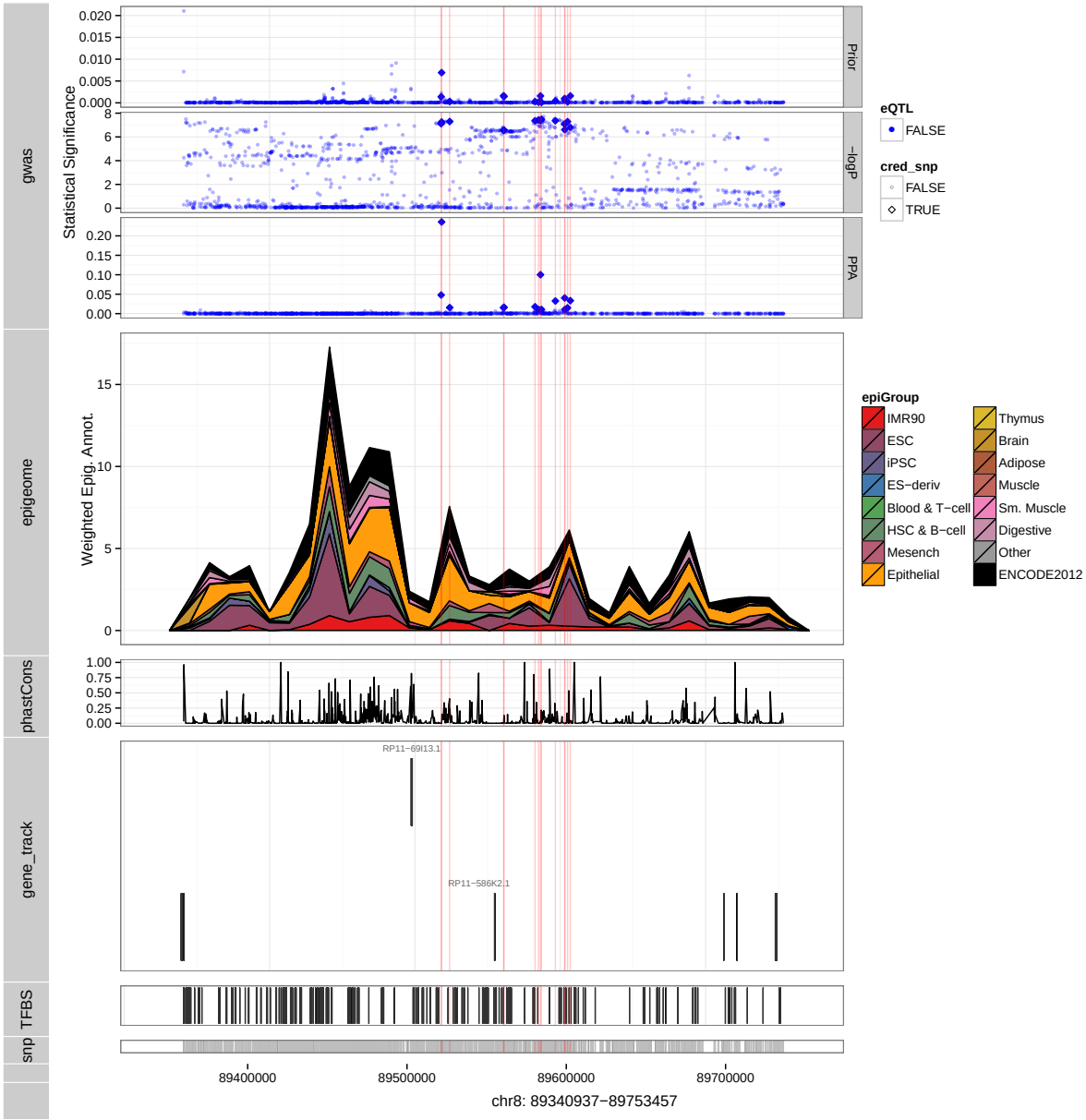
Schizophrenia



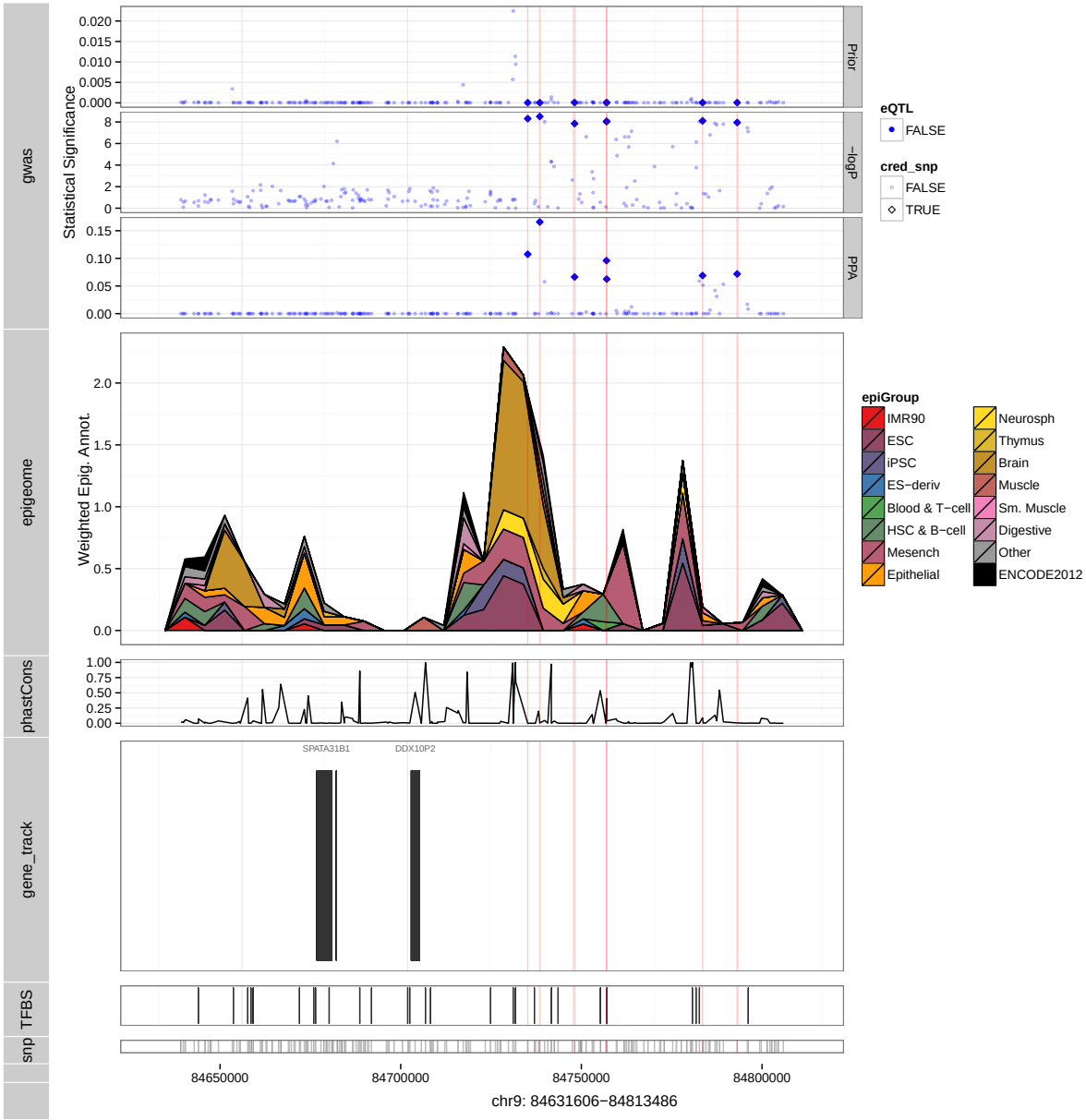
Schizophrenia



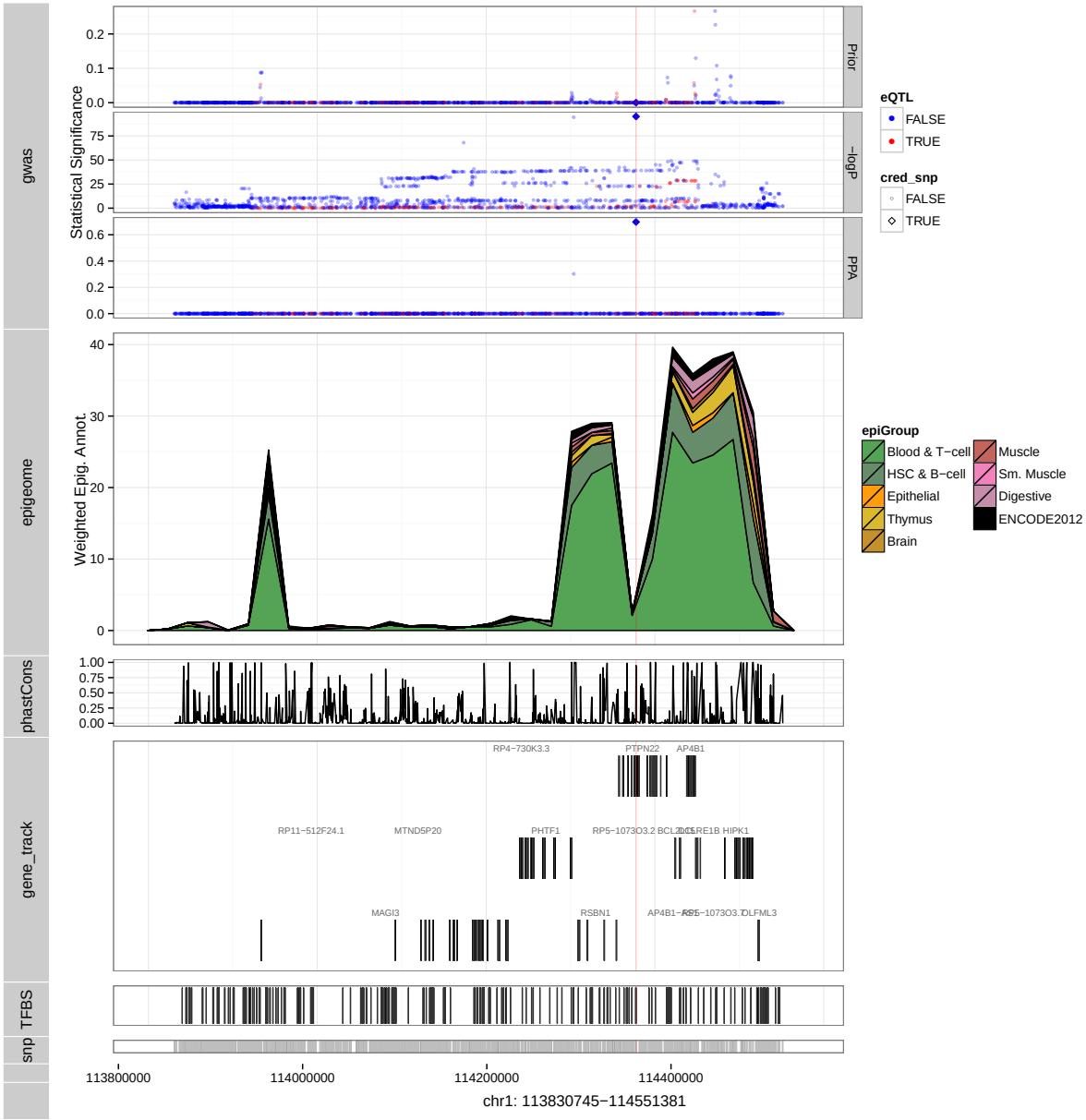
Schizophrenia



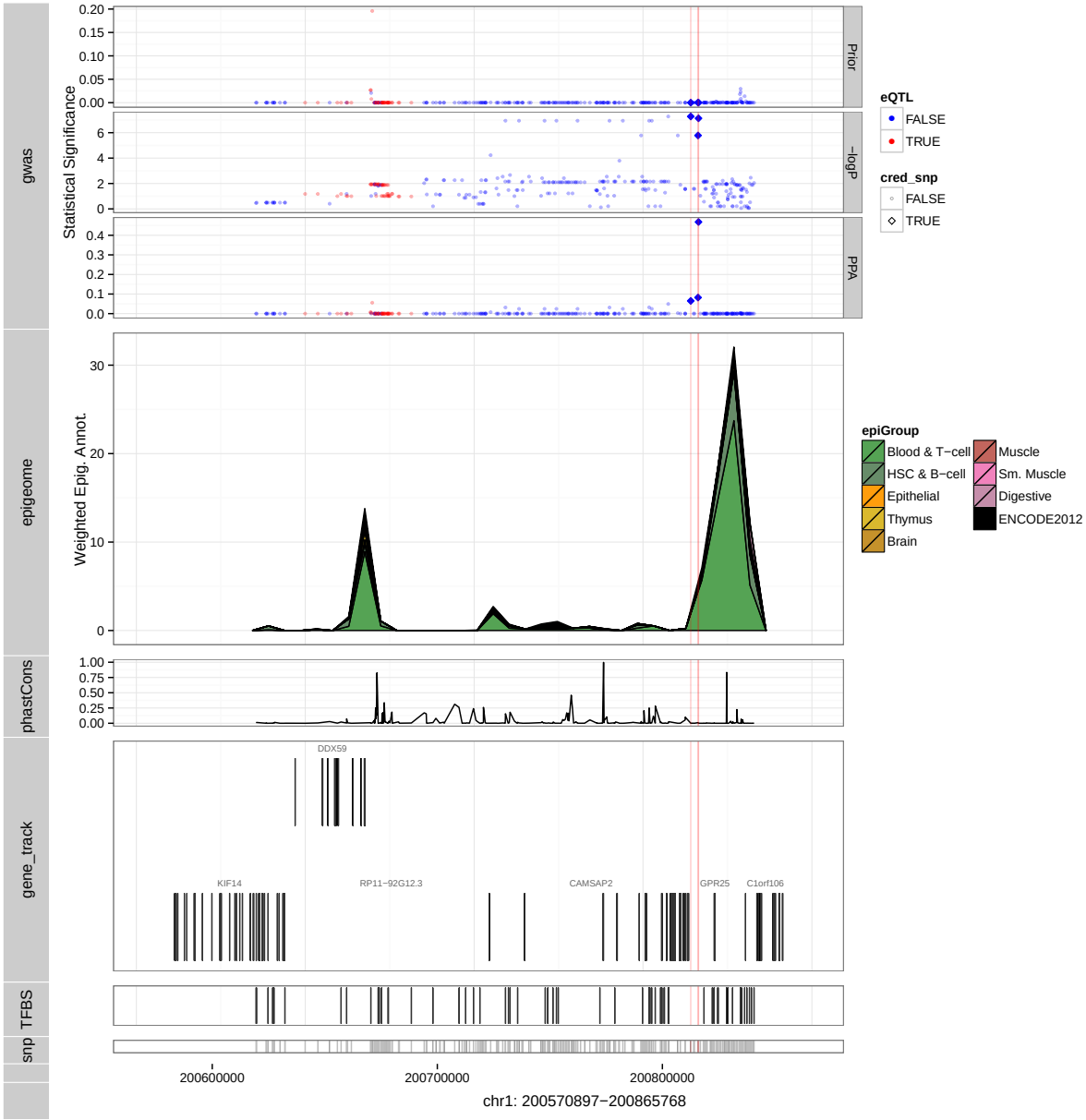
Schizophrenia



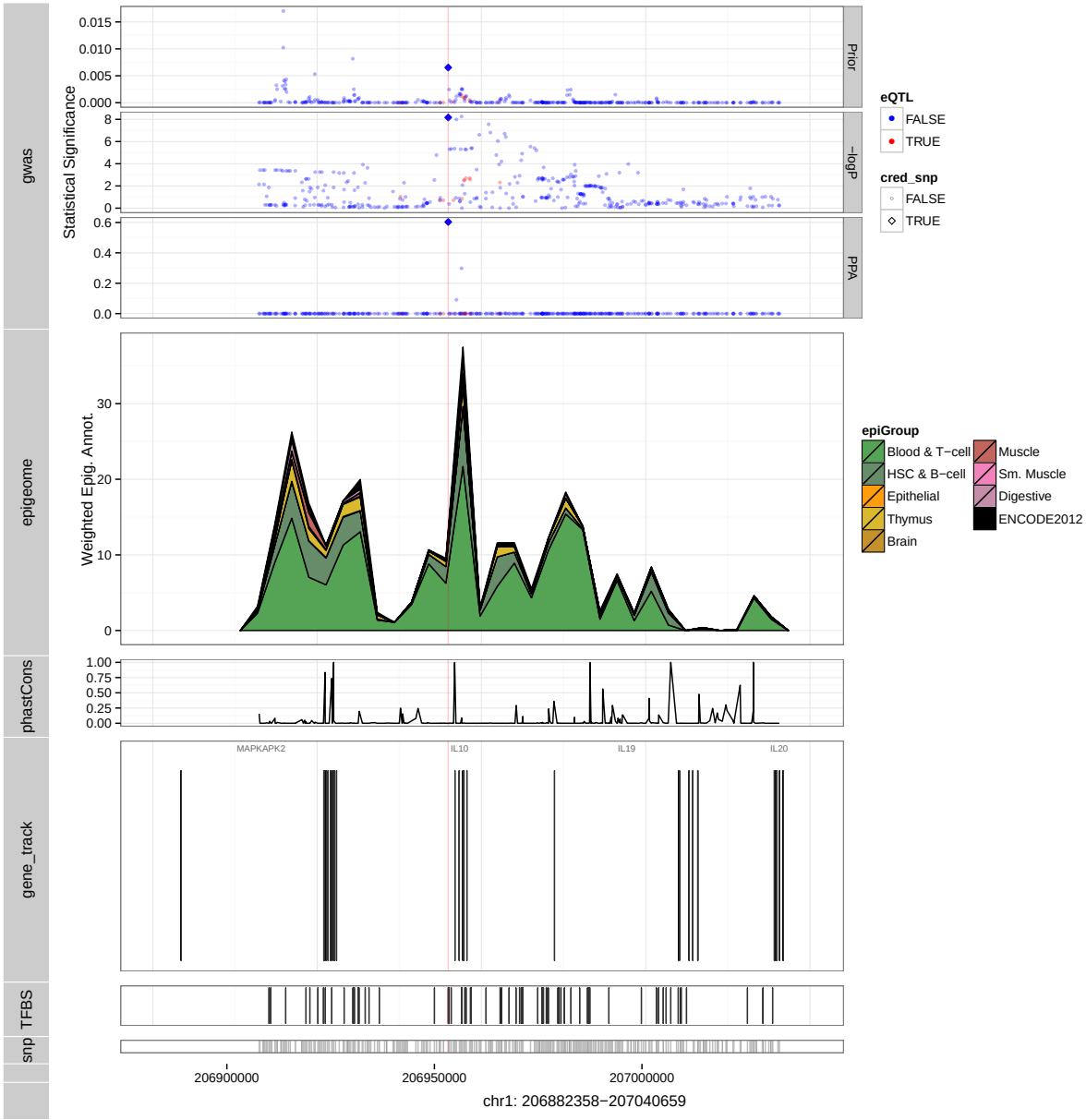
Type 1 Diabetes



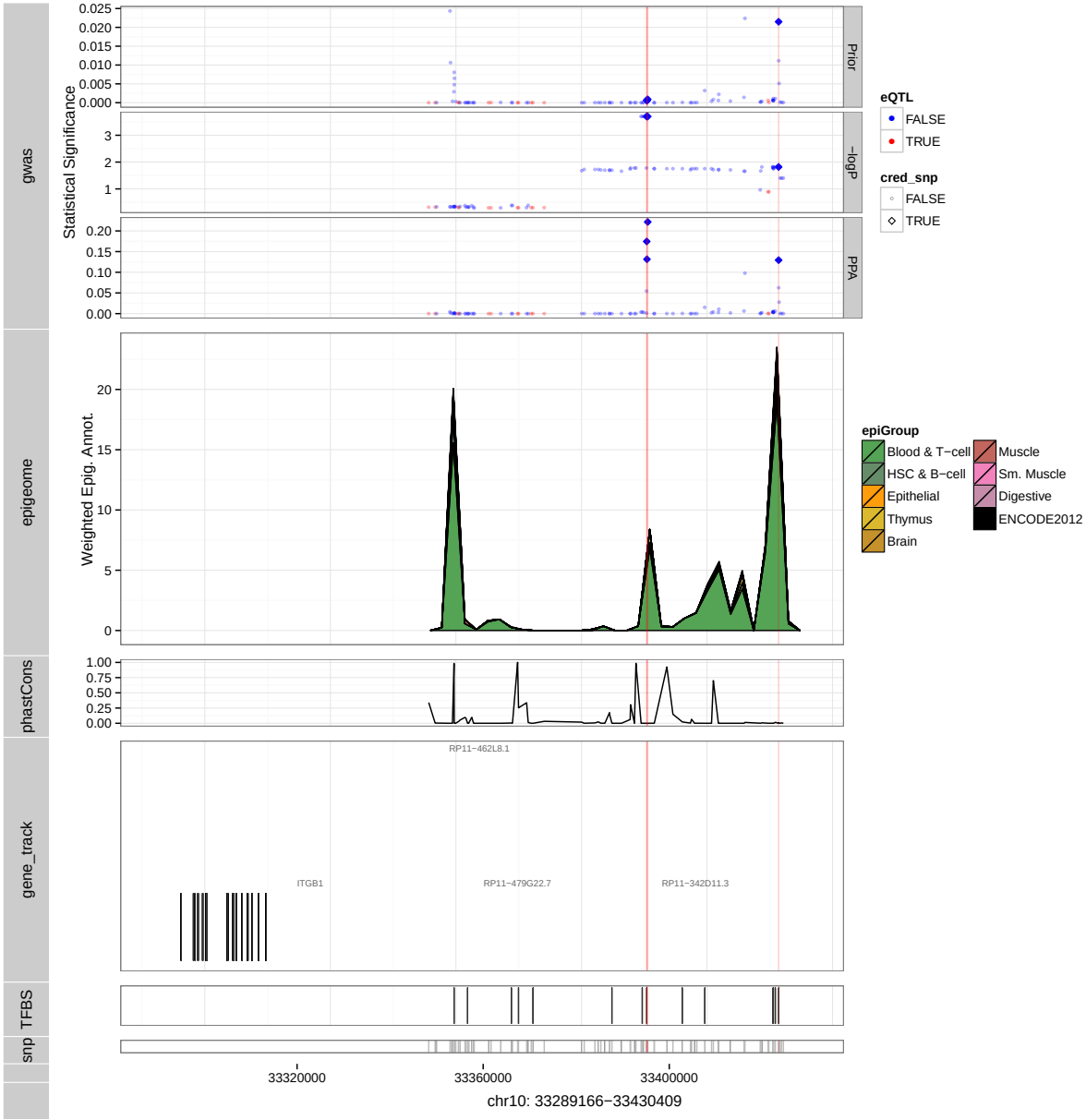
Type 1 Diabetes



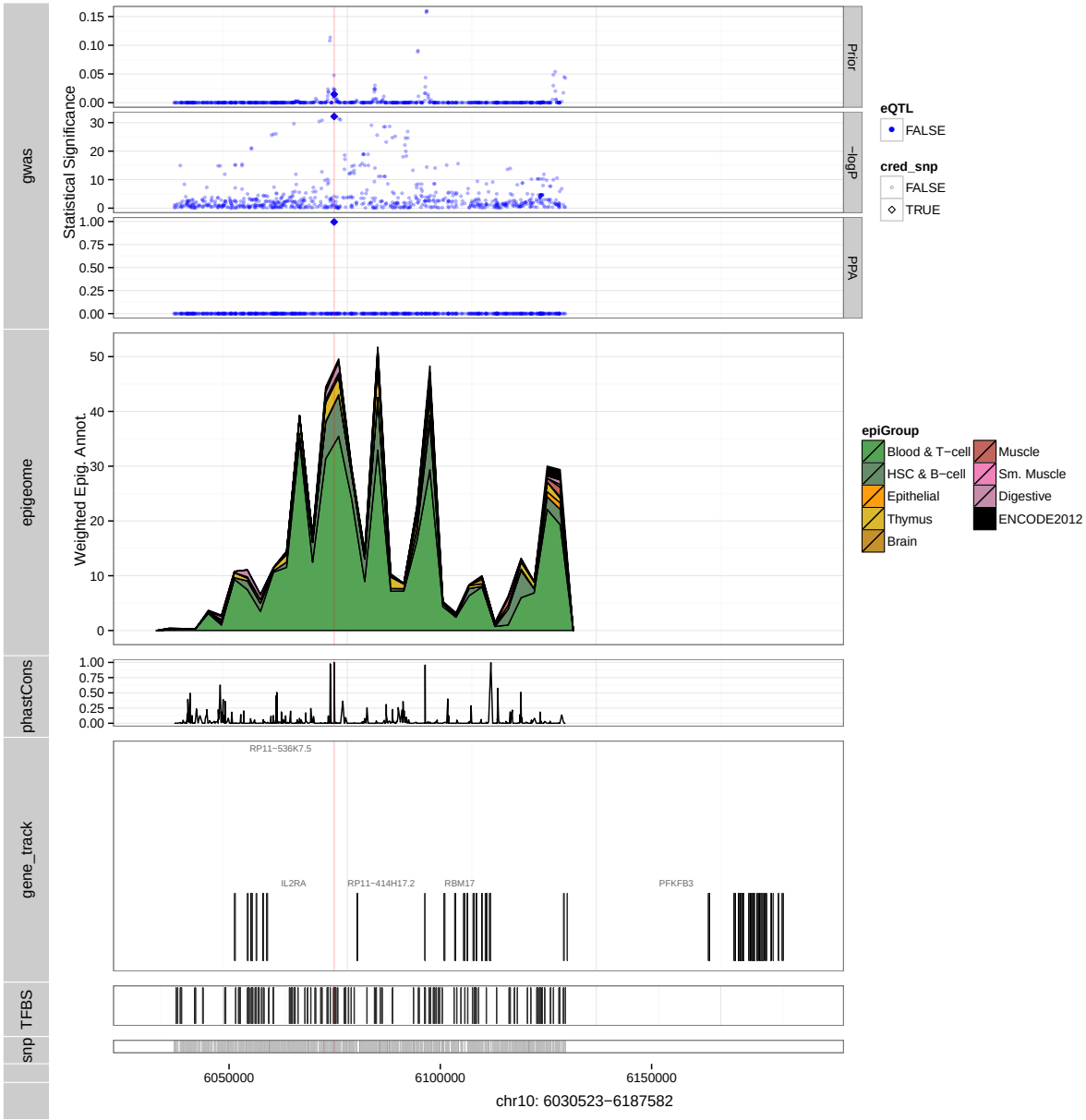
Type 1 Diabetes



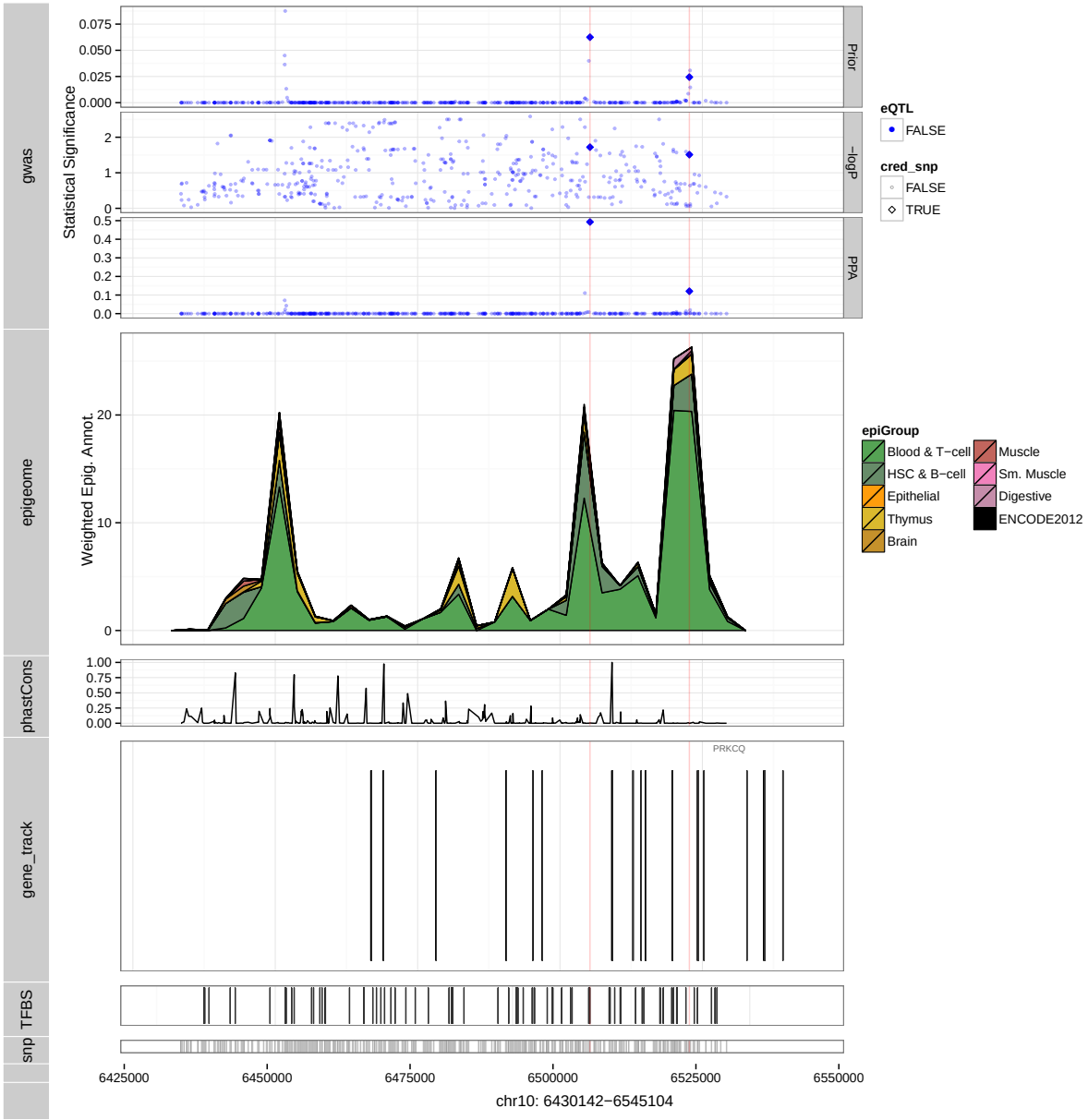
Type 1 Diabetes



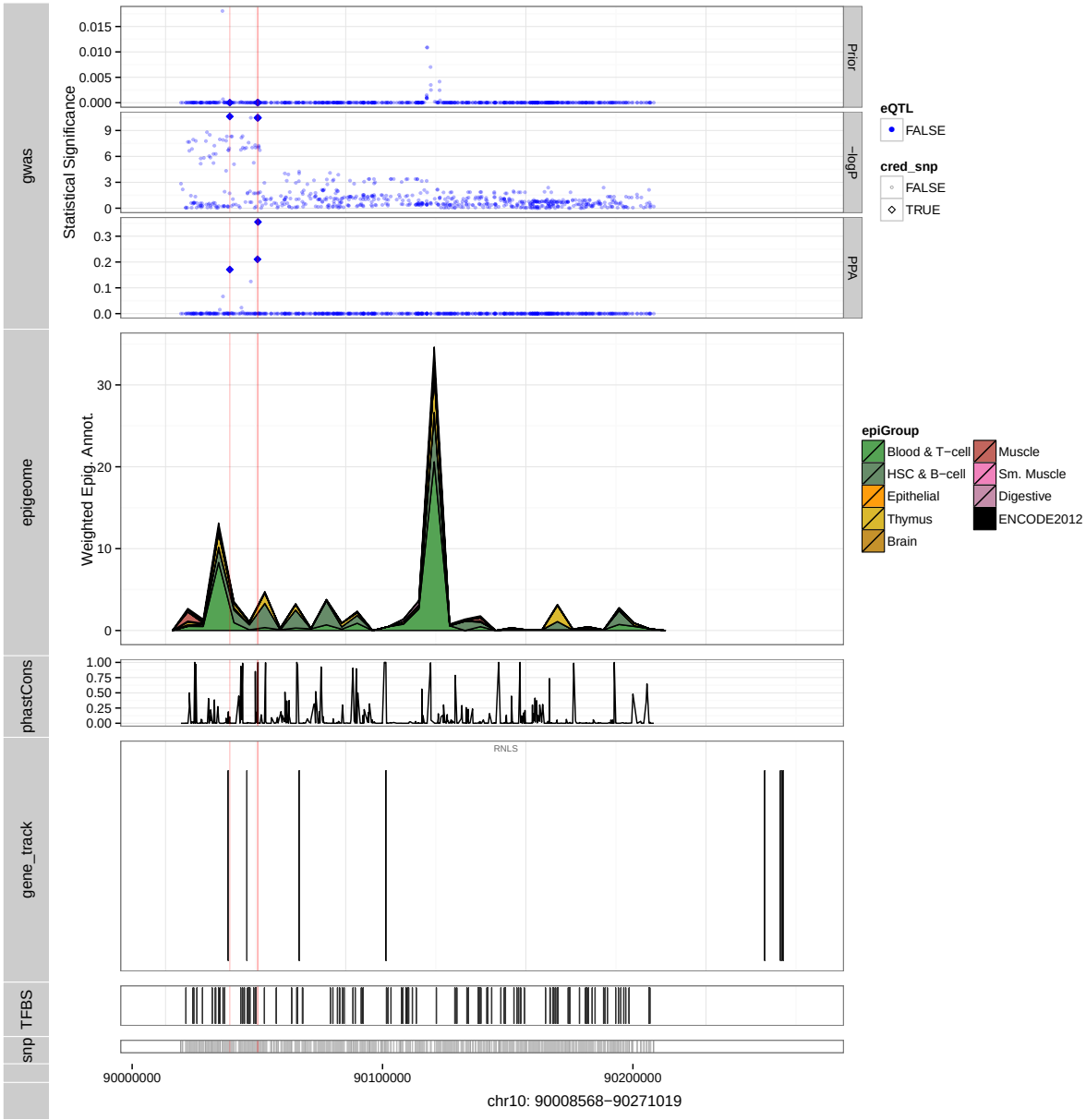
Type 1 Diabetes



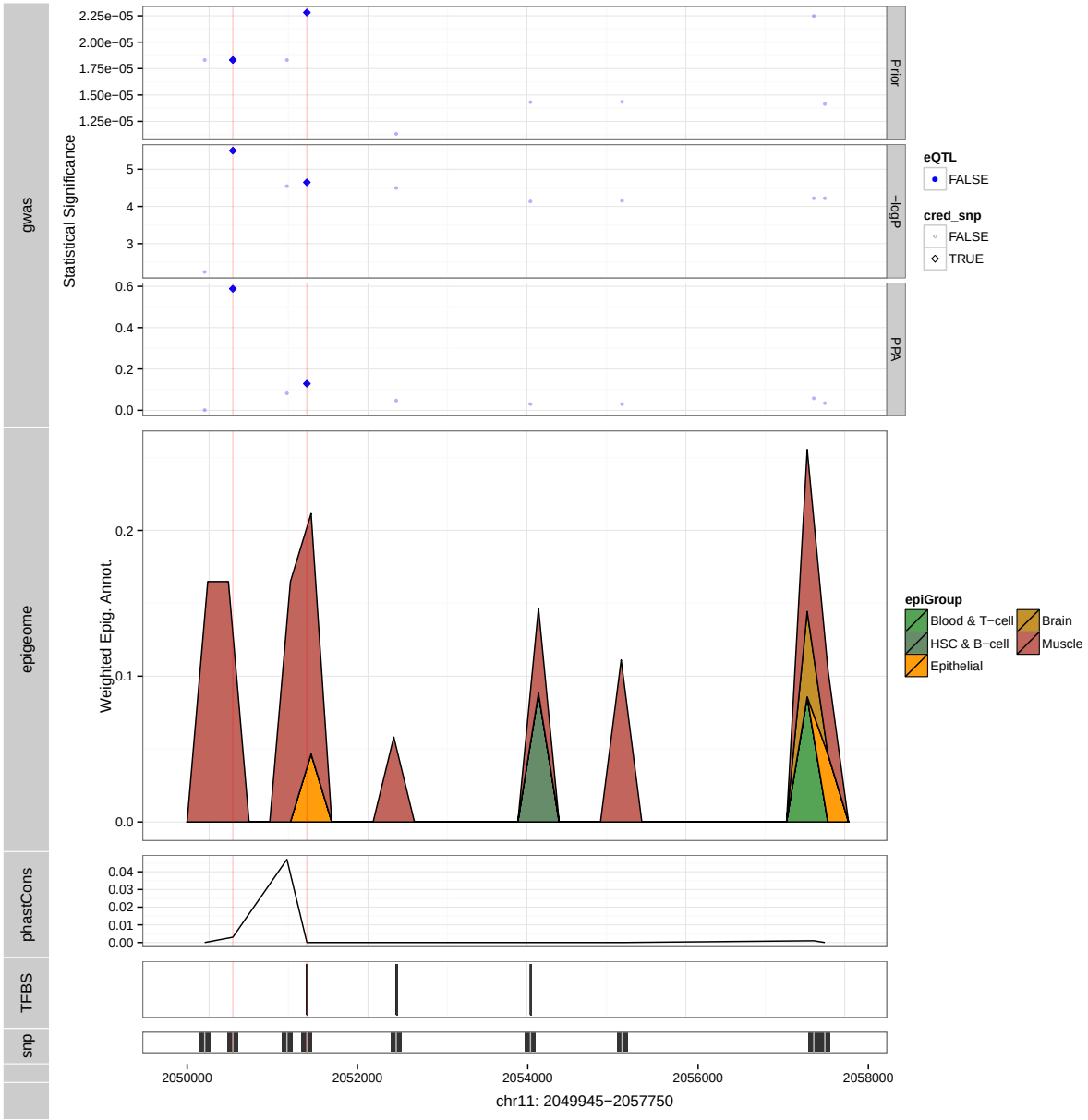
Type 1 Diabetes



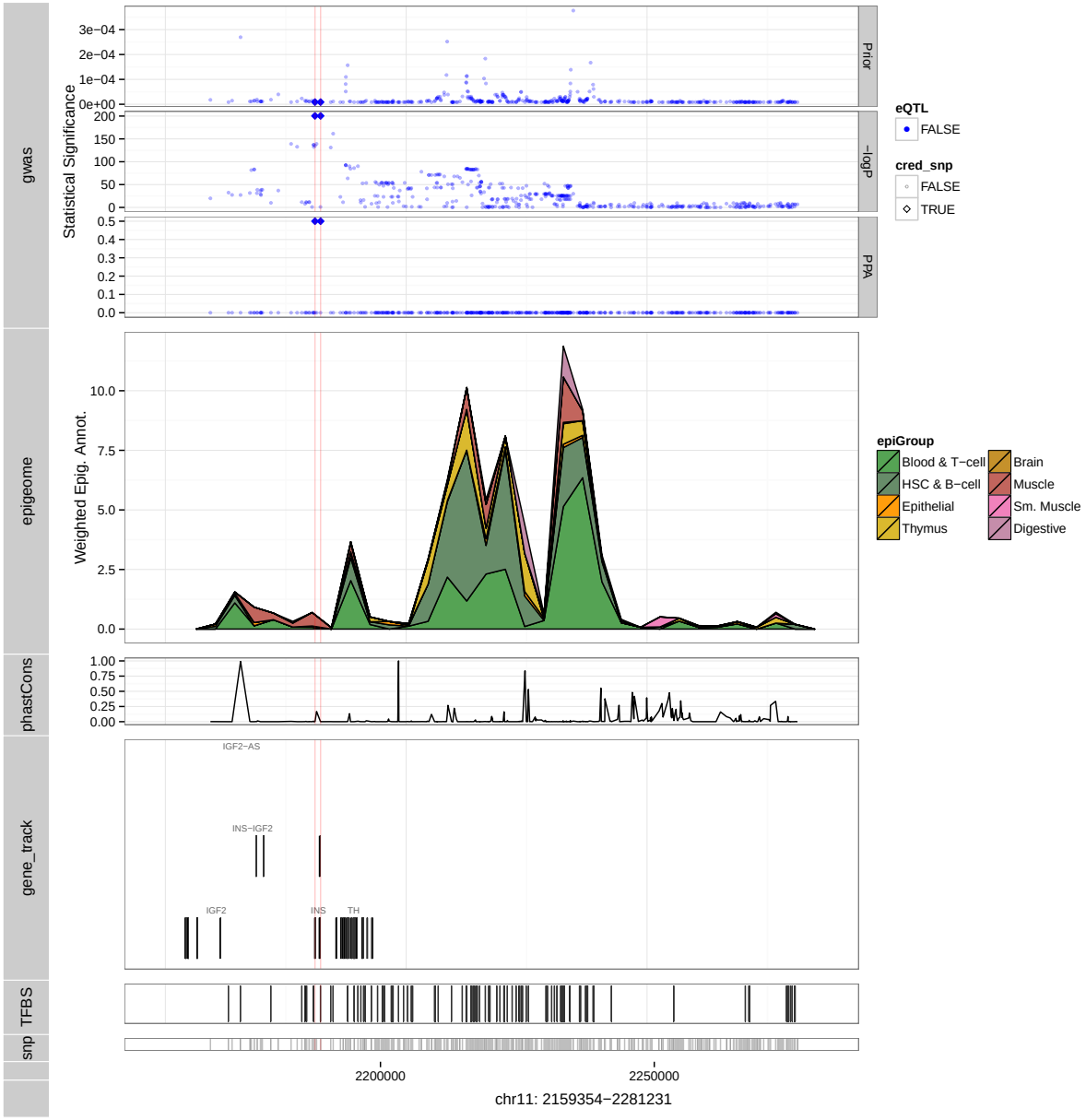
Type 1 Diabetes



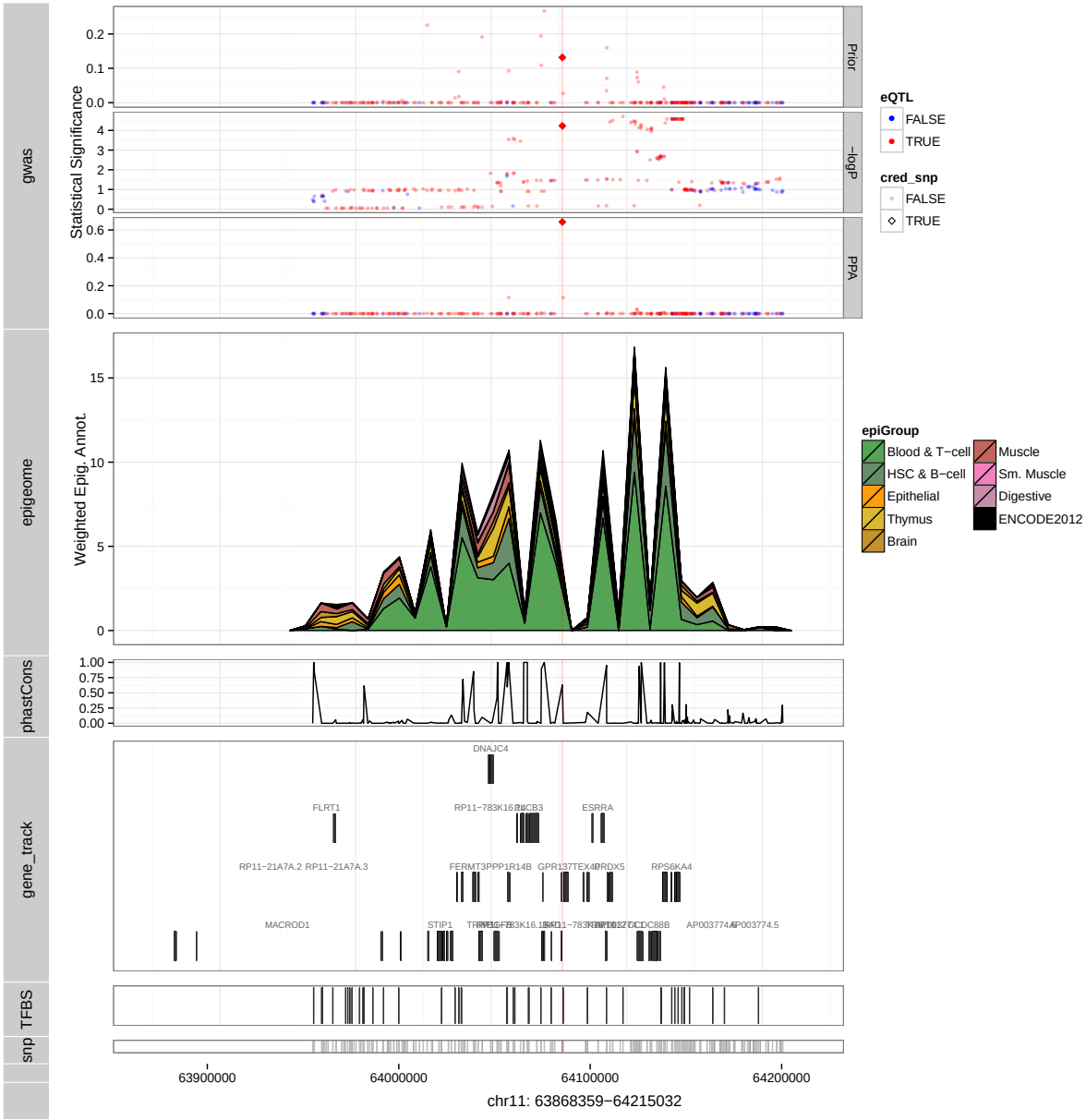
Type 1 Diabetes



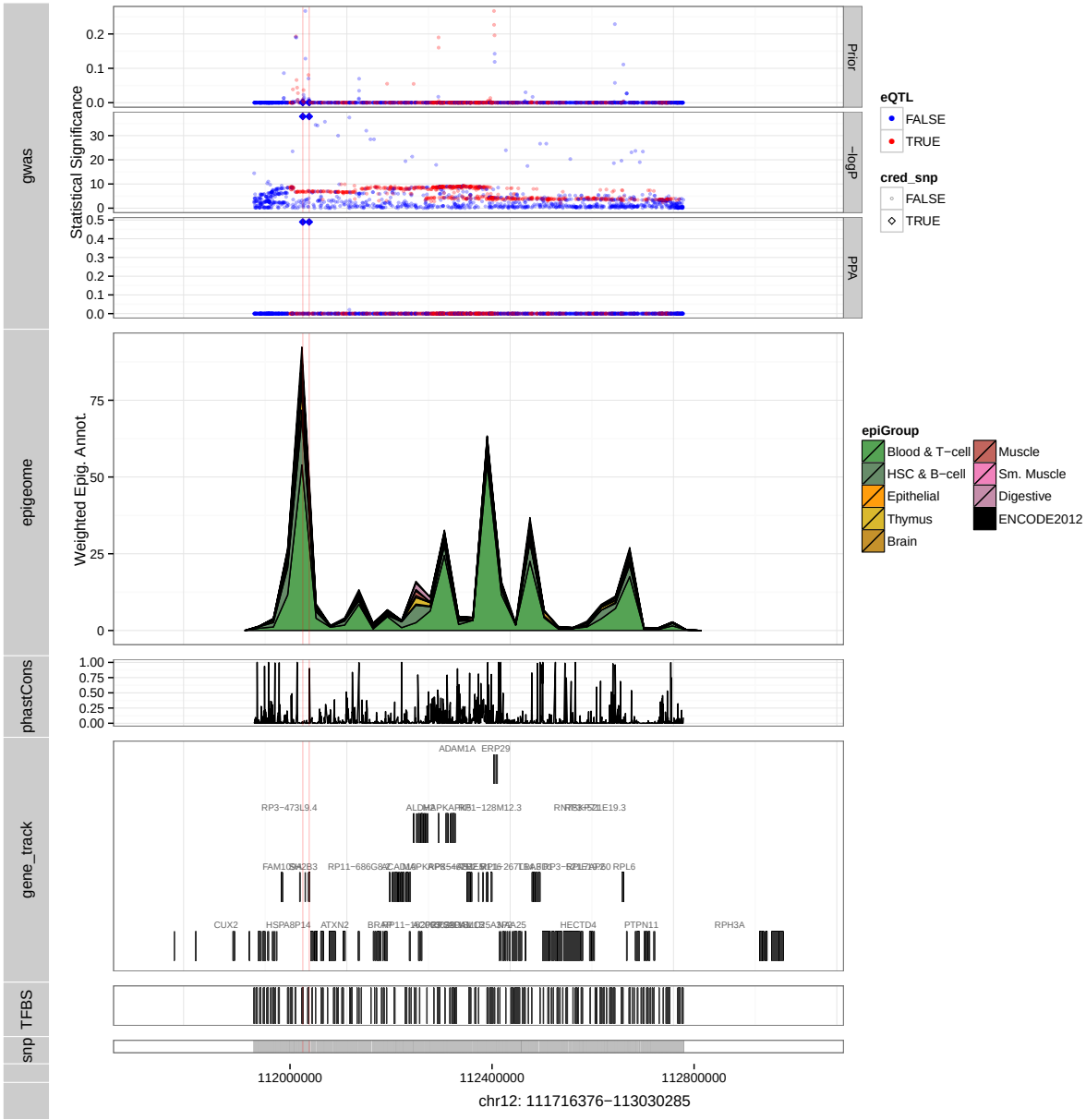
Type 1 Diabetes



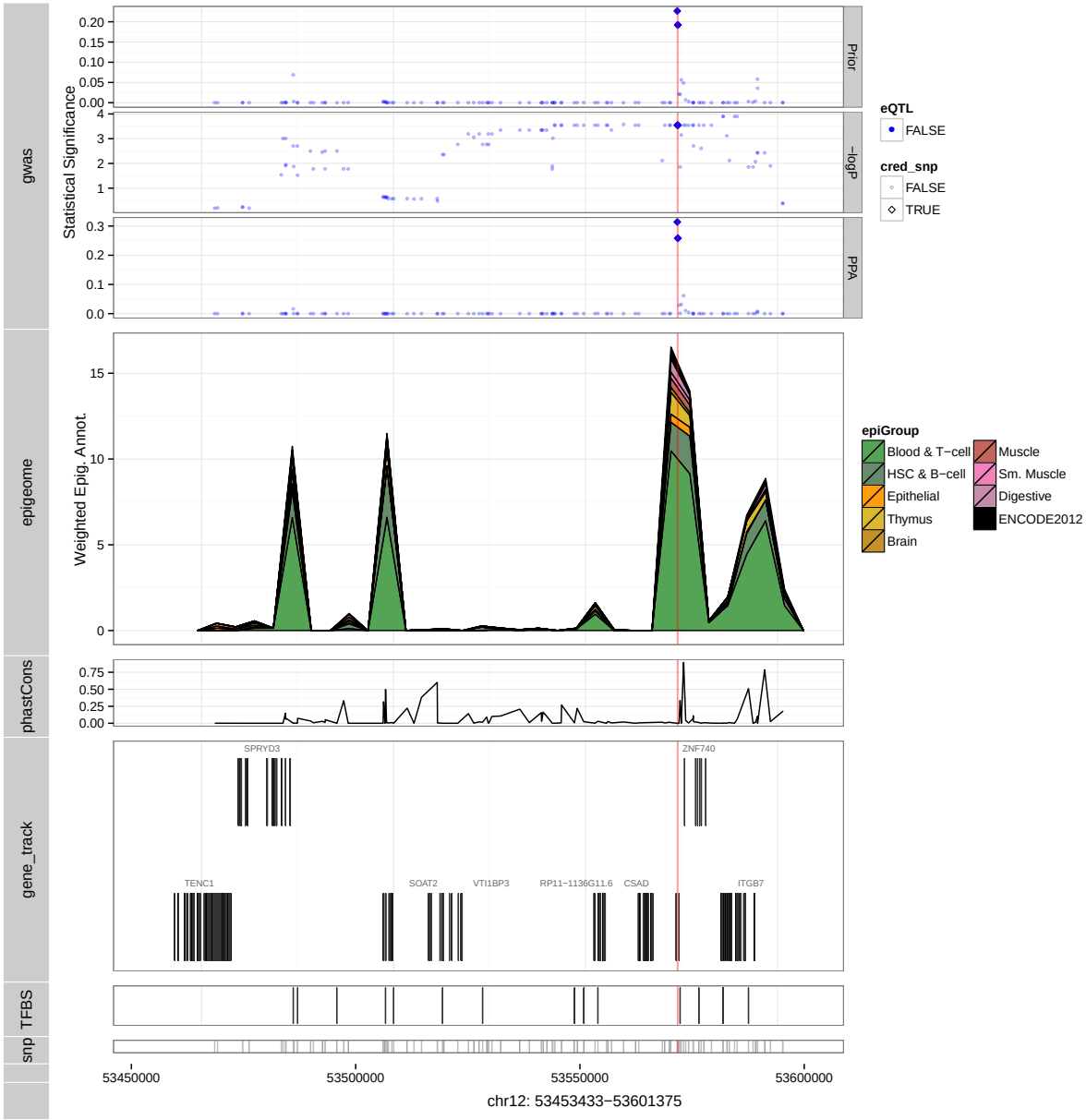
Type 1 Diabetes



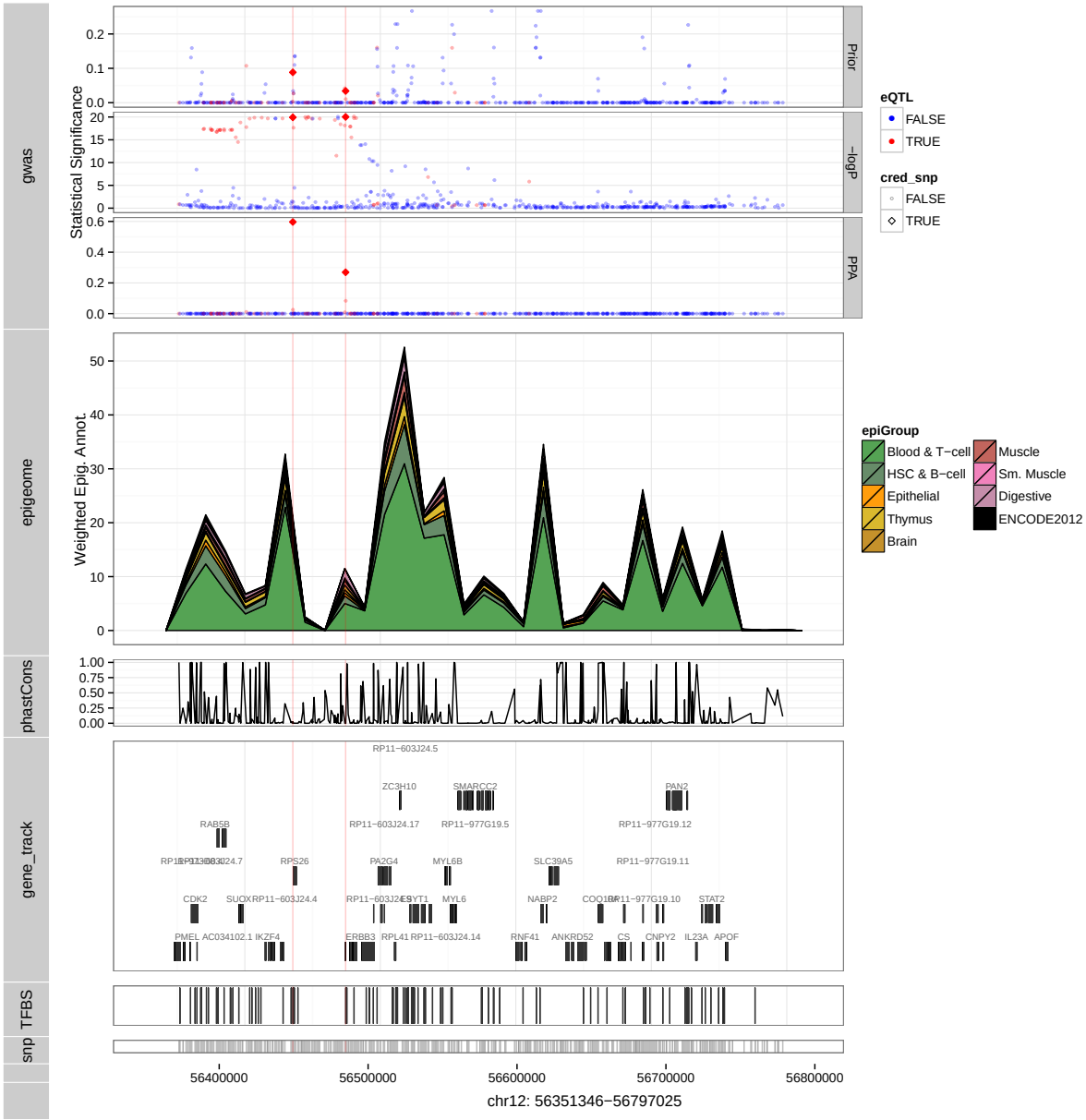
Type 1 Diabetes



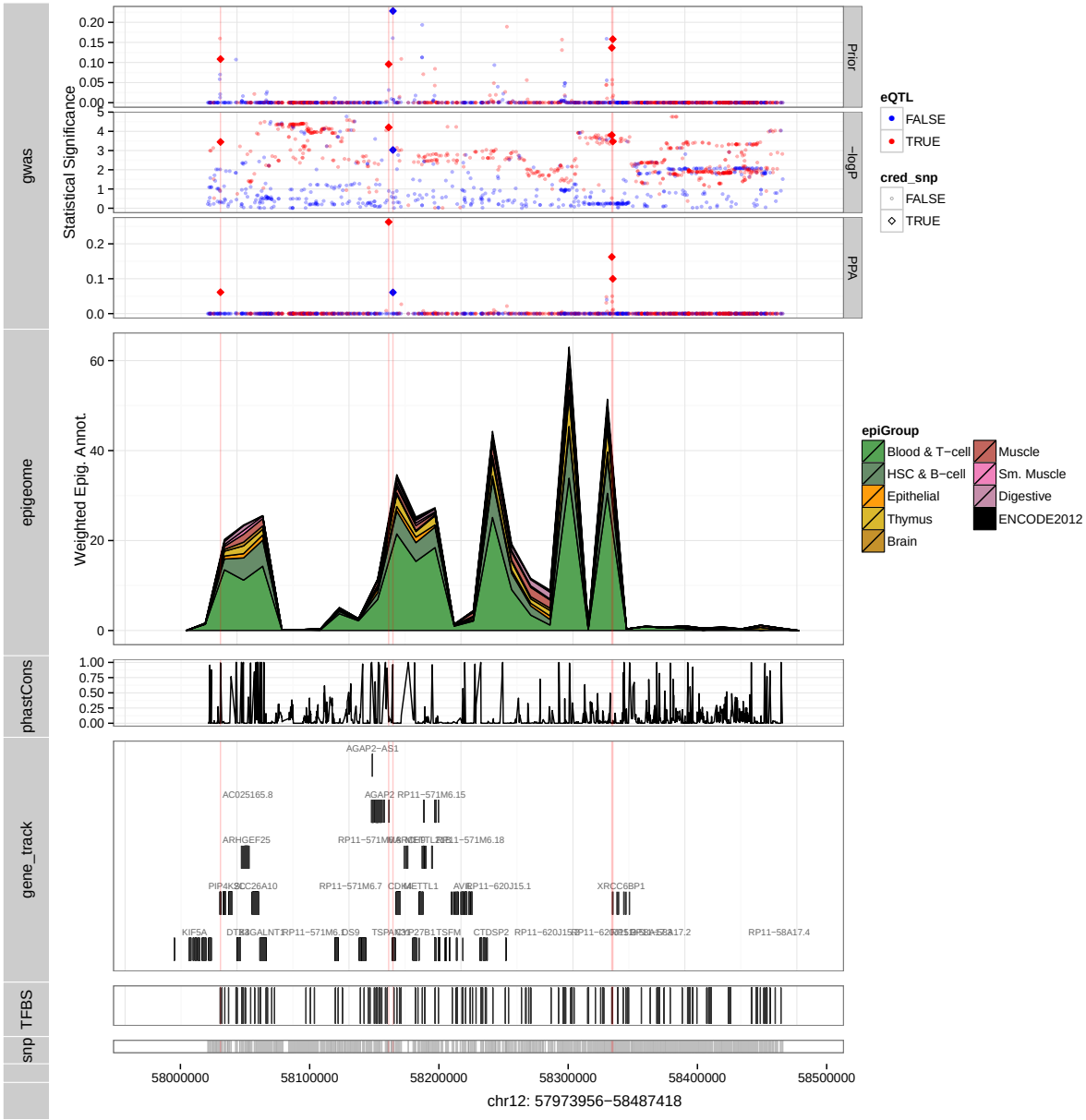
Type 1 Diabetes



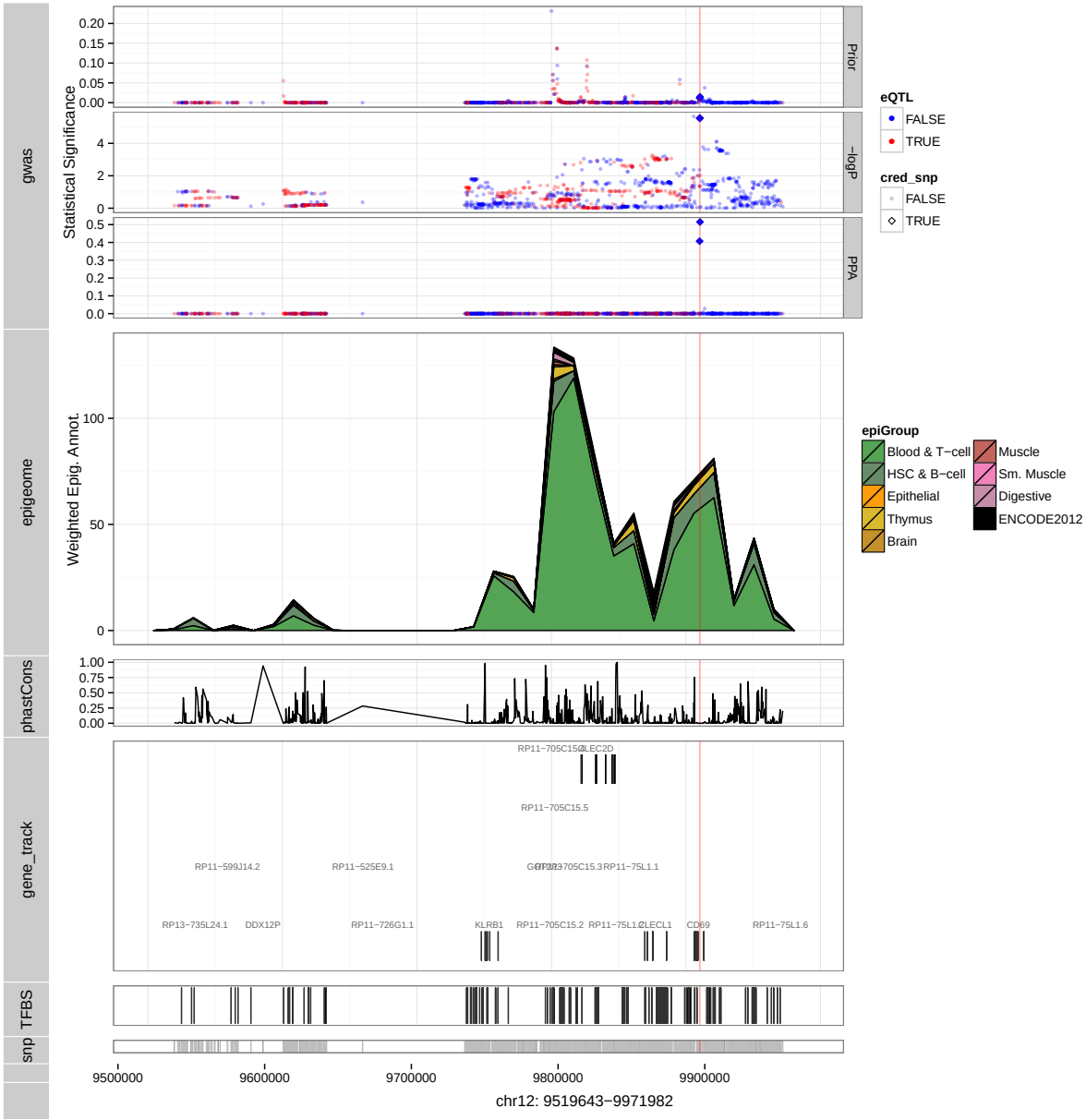
Type 1 Diabetes



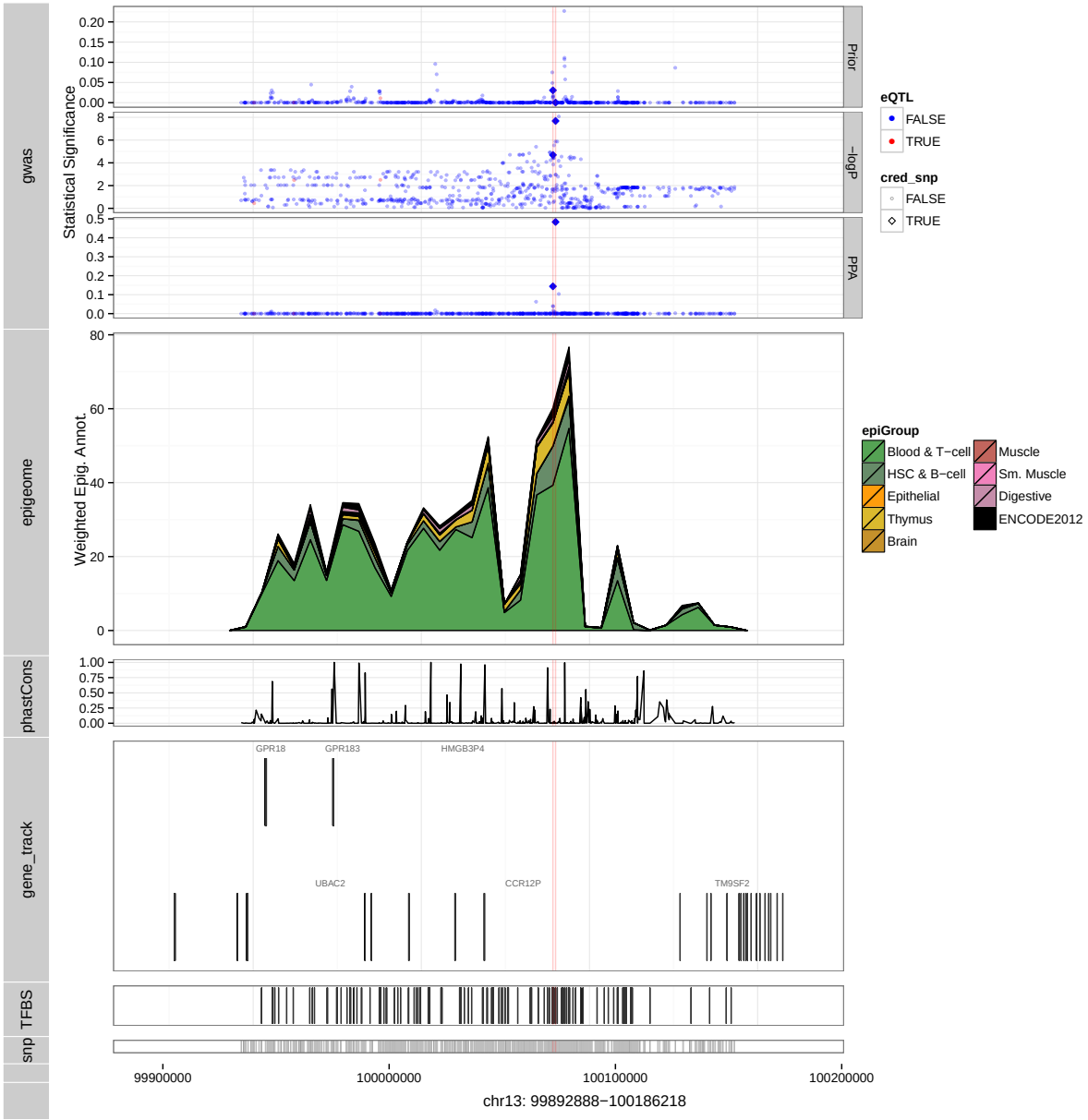
Type 1 Diabetes



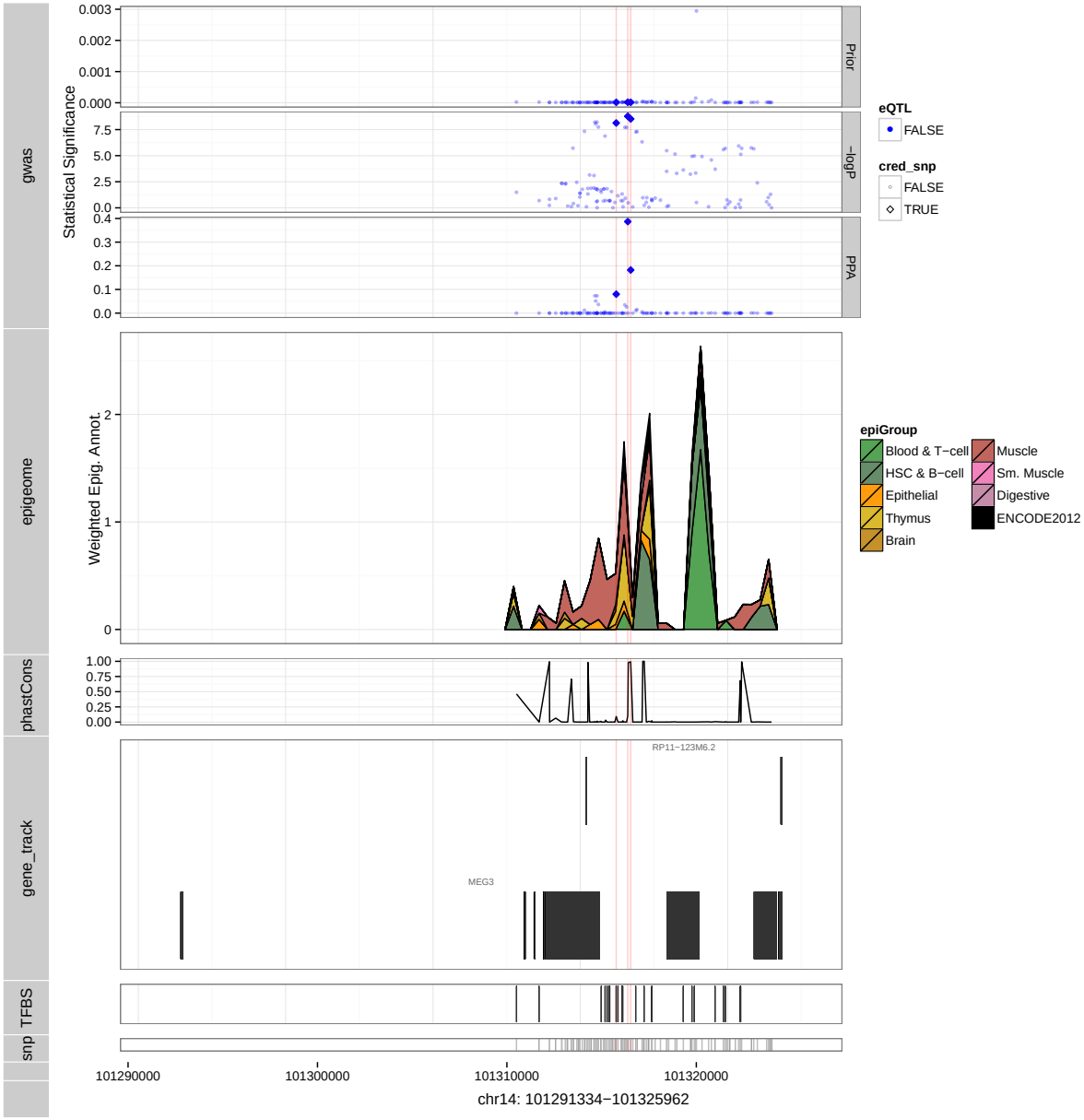
Type 1 Diabetes



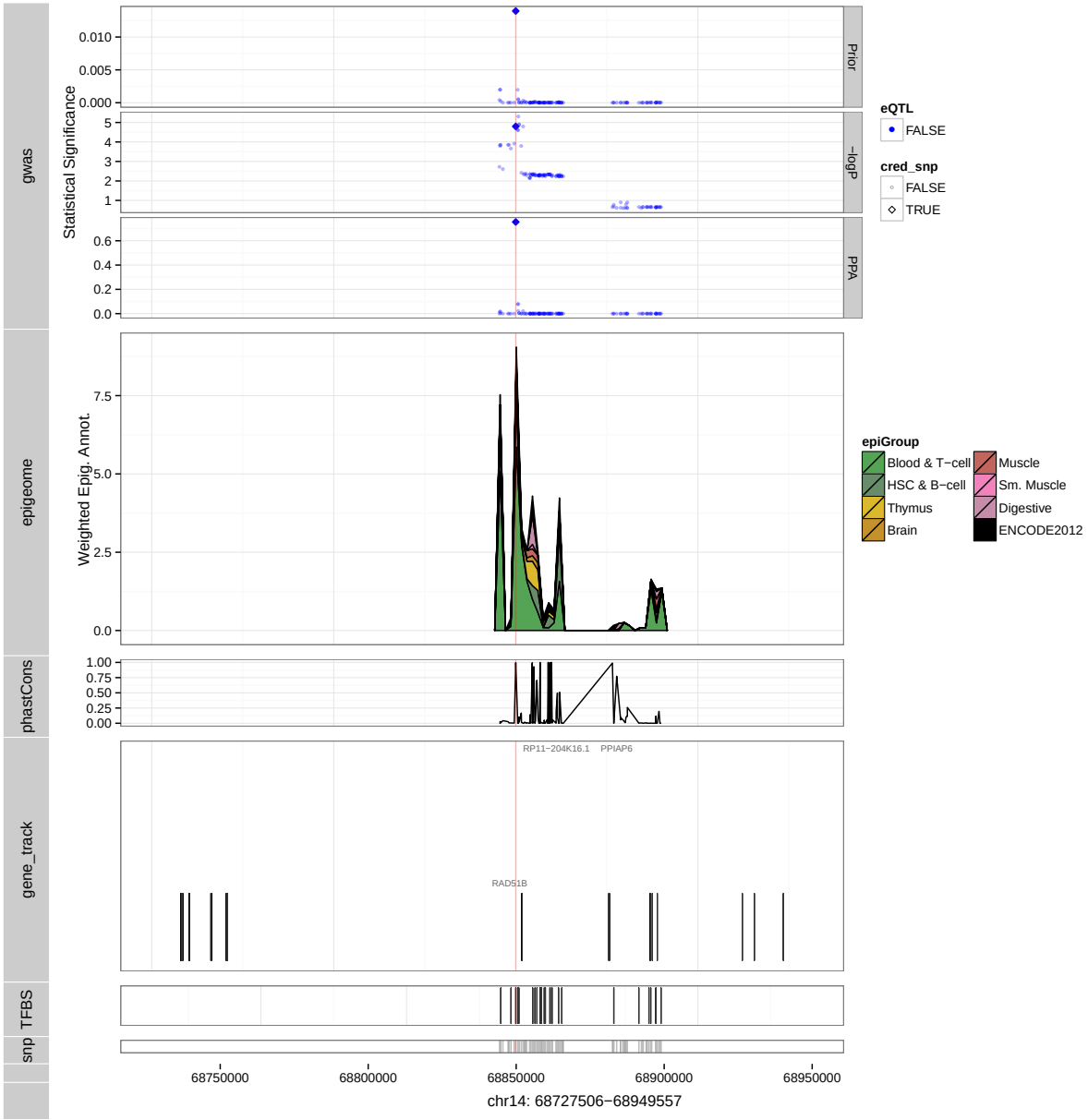
Type 1 Diabetes



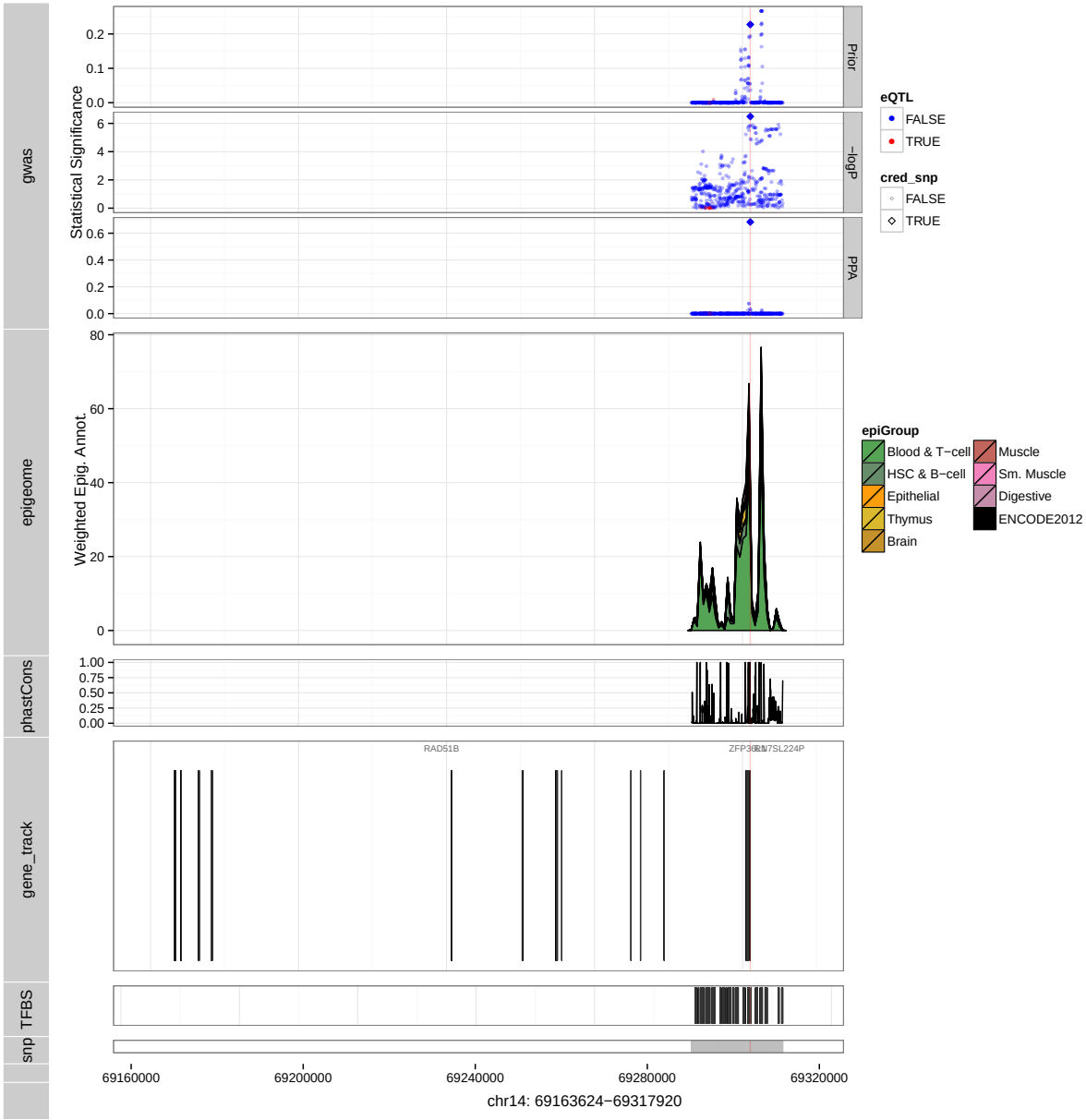
Type 1 Diabetes



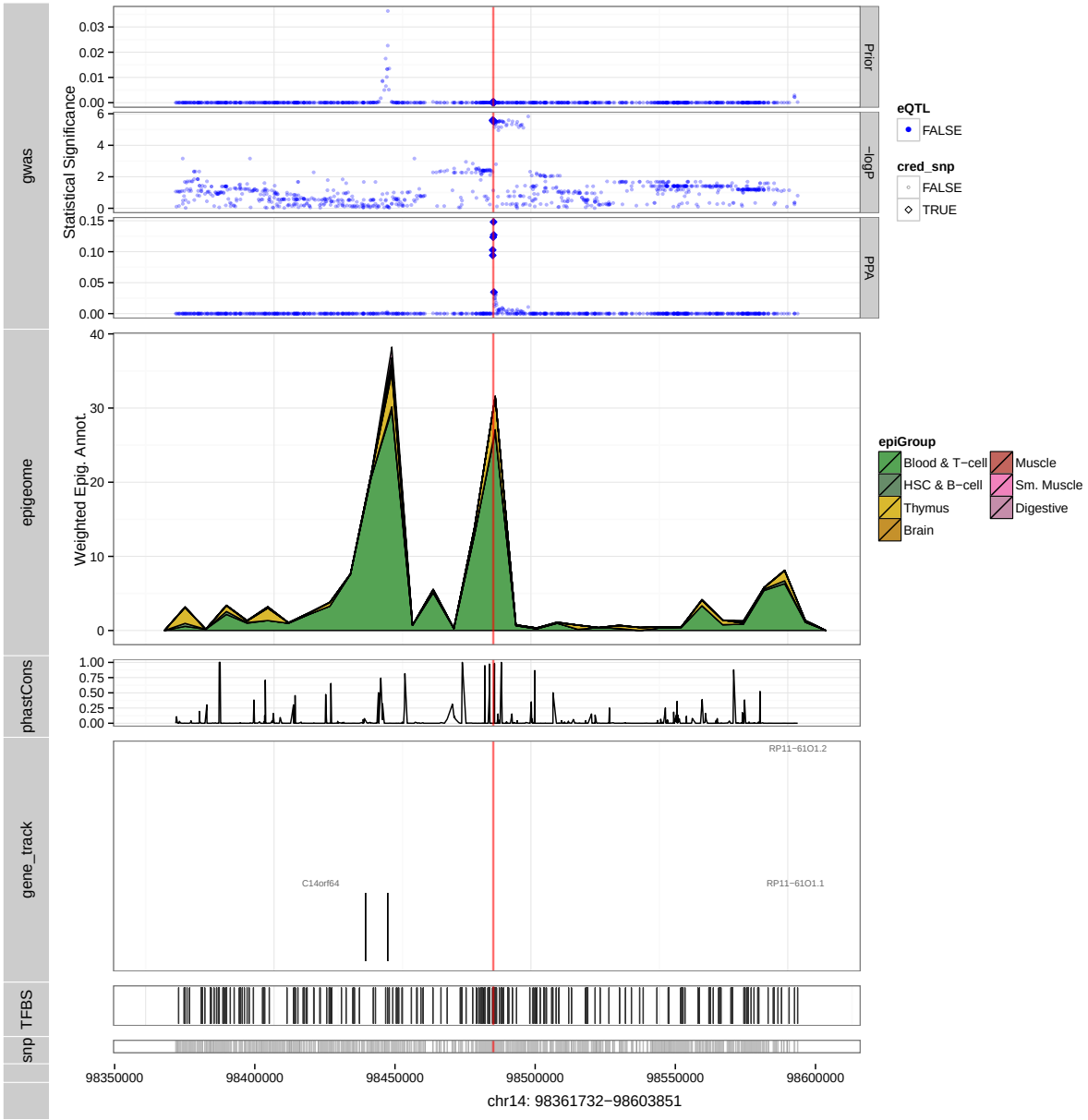
Type 1 Diabetes



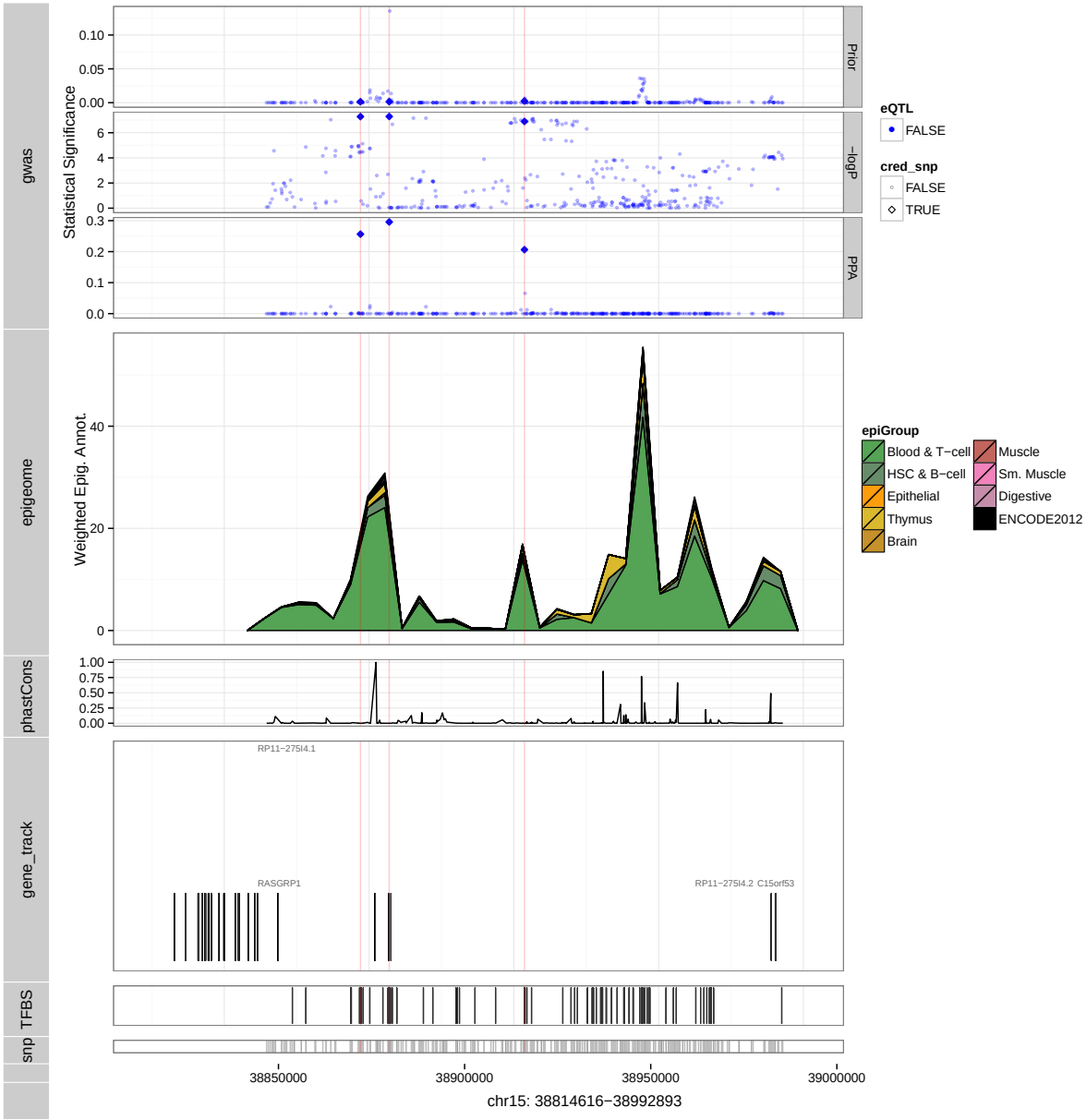
Type 1 Diabetes



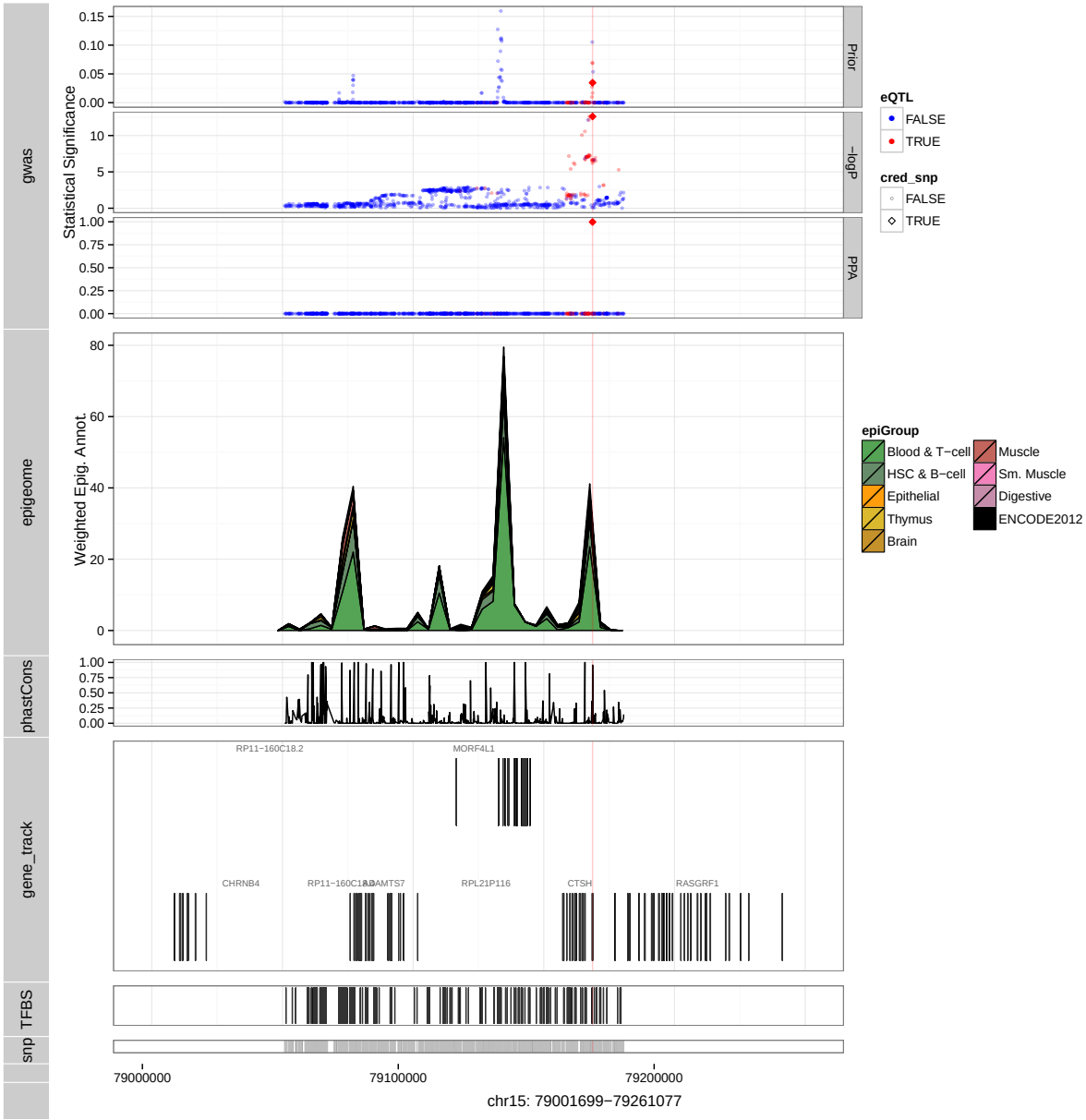
Type 1 Diabetes



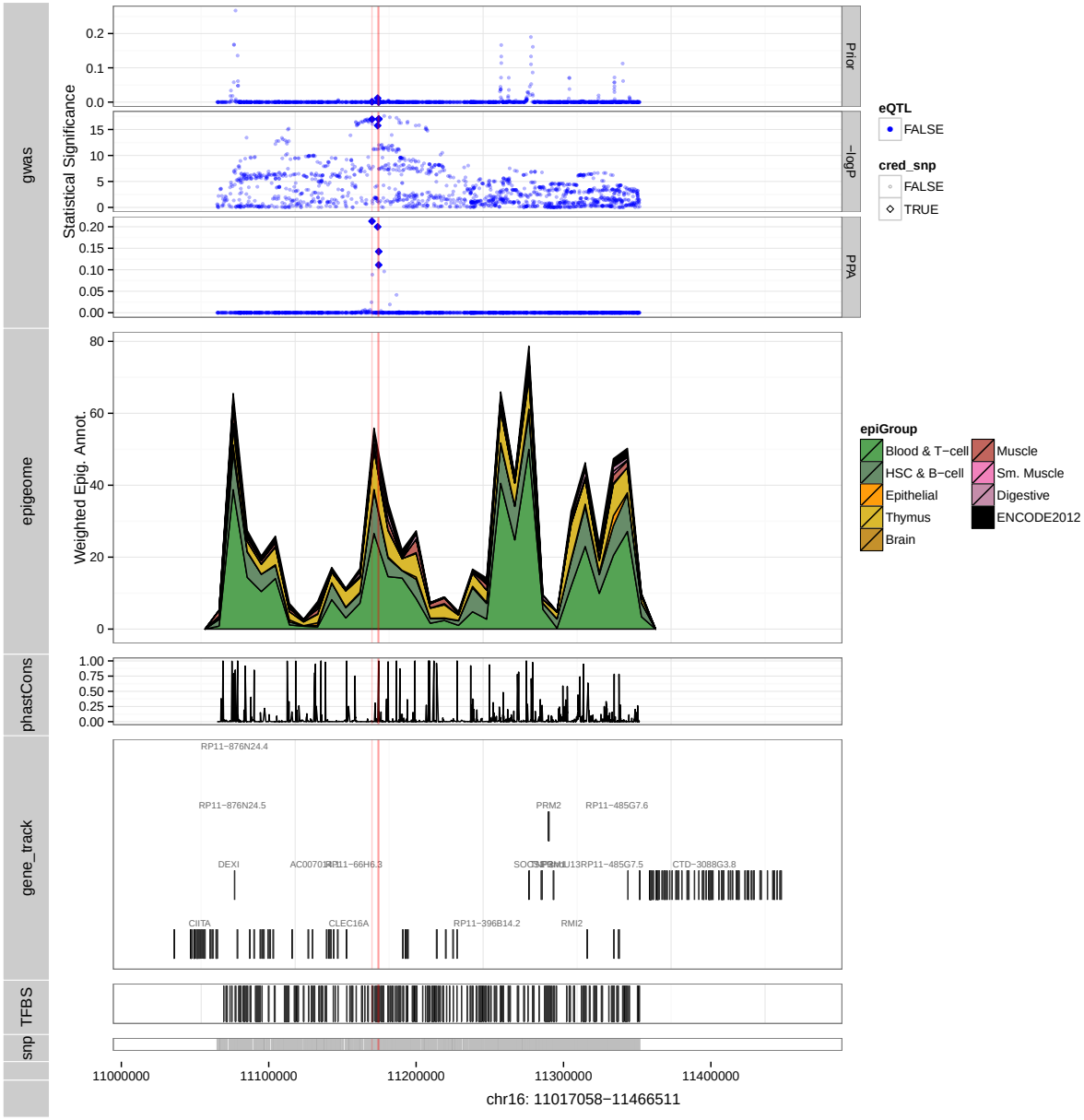
Type 1 Diabetes



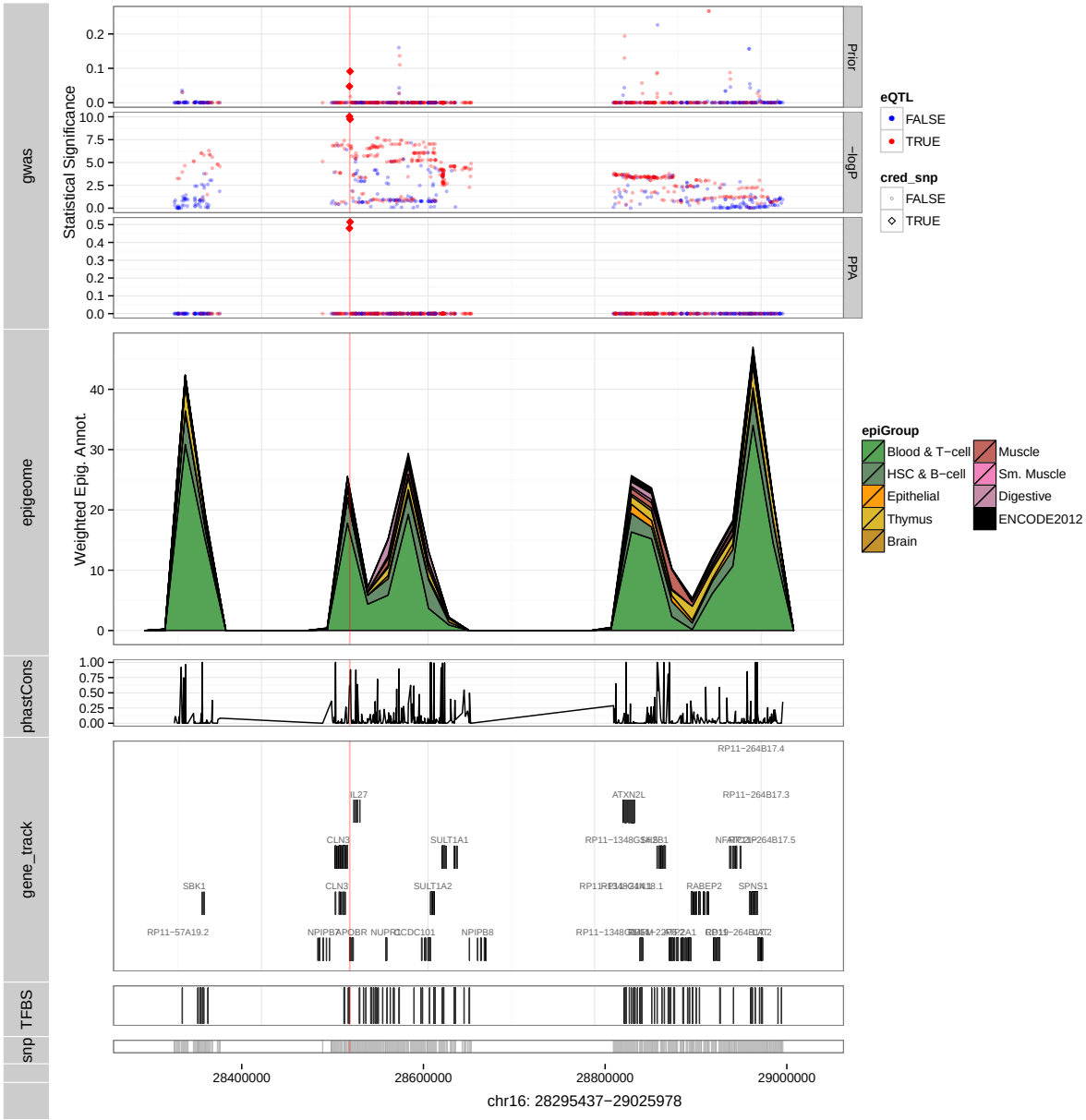
Type 1 Diabetes



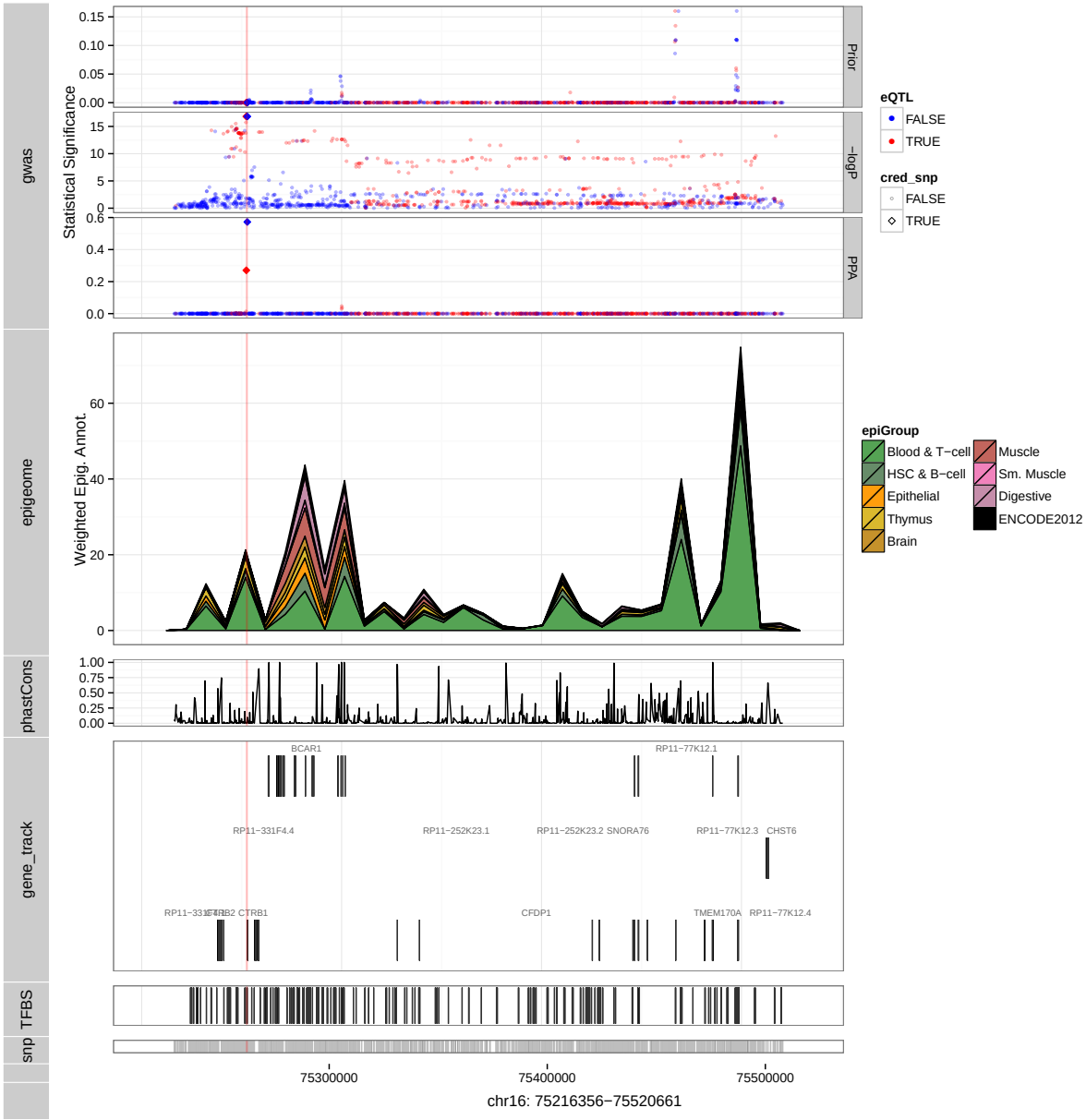
Type 1 Diabetes



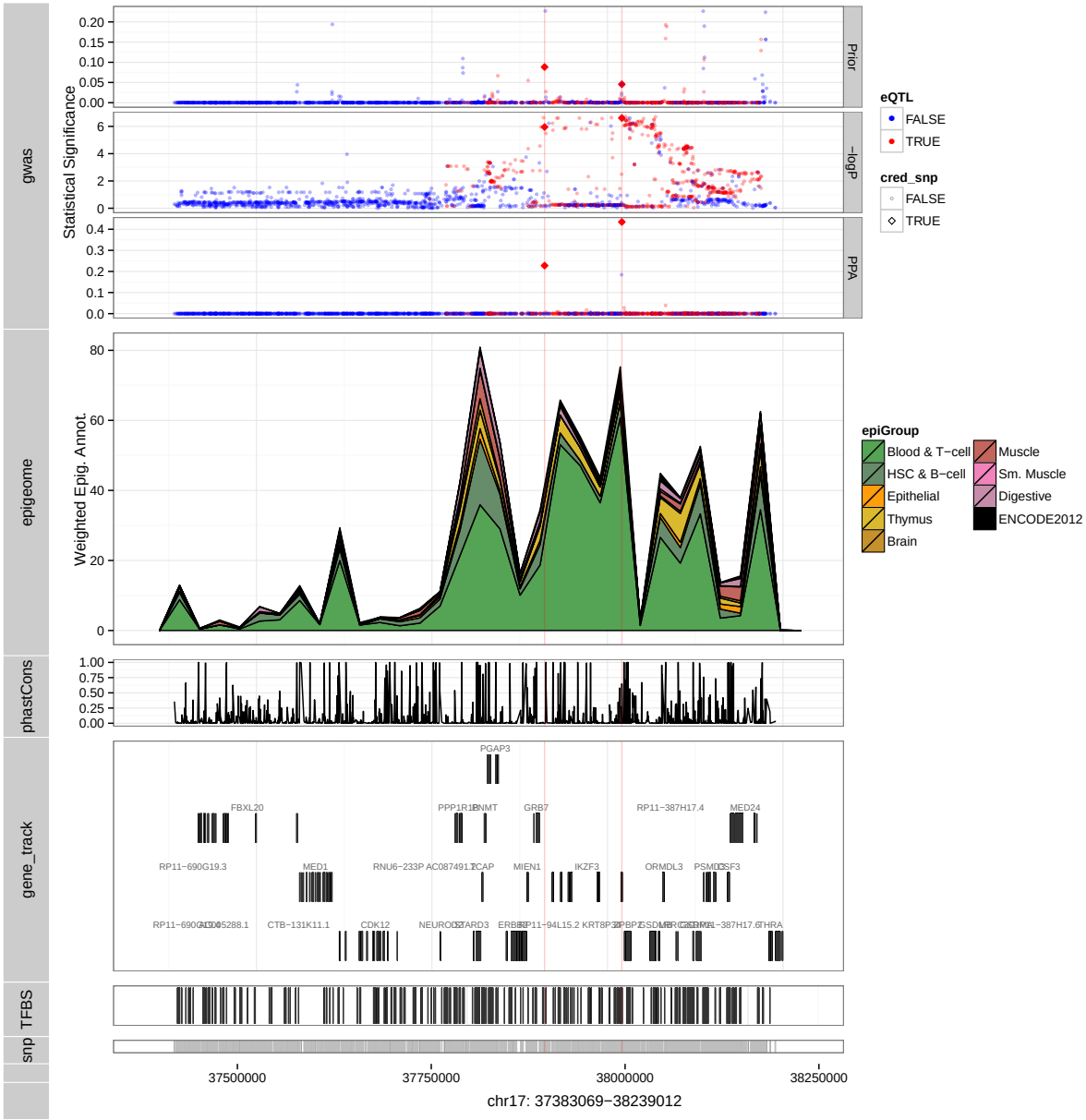
Type 1 Diabetes



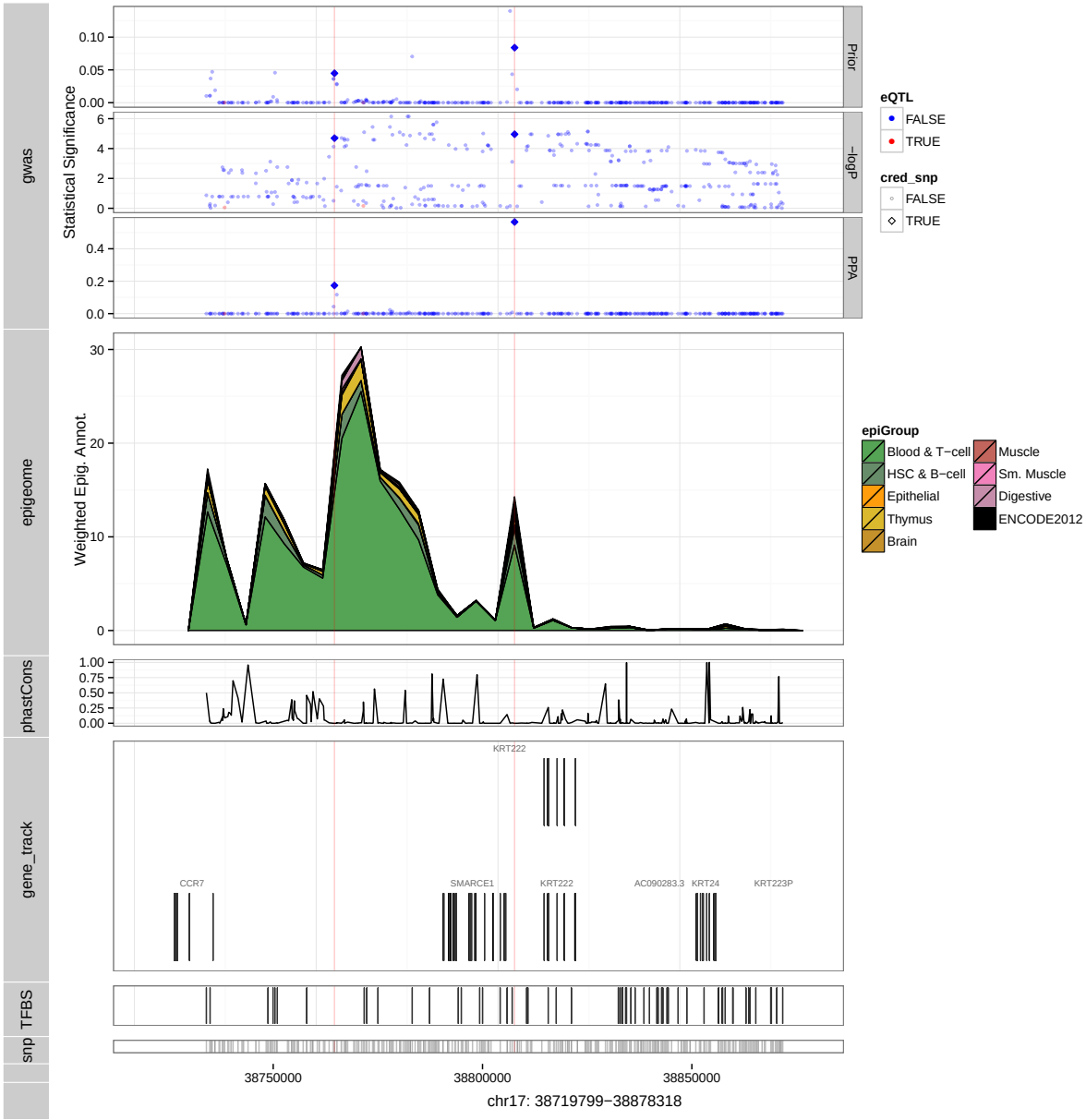
Type 1 Diabetes



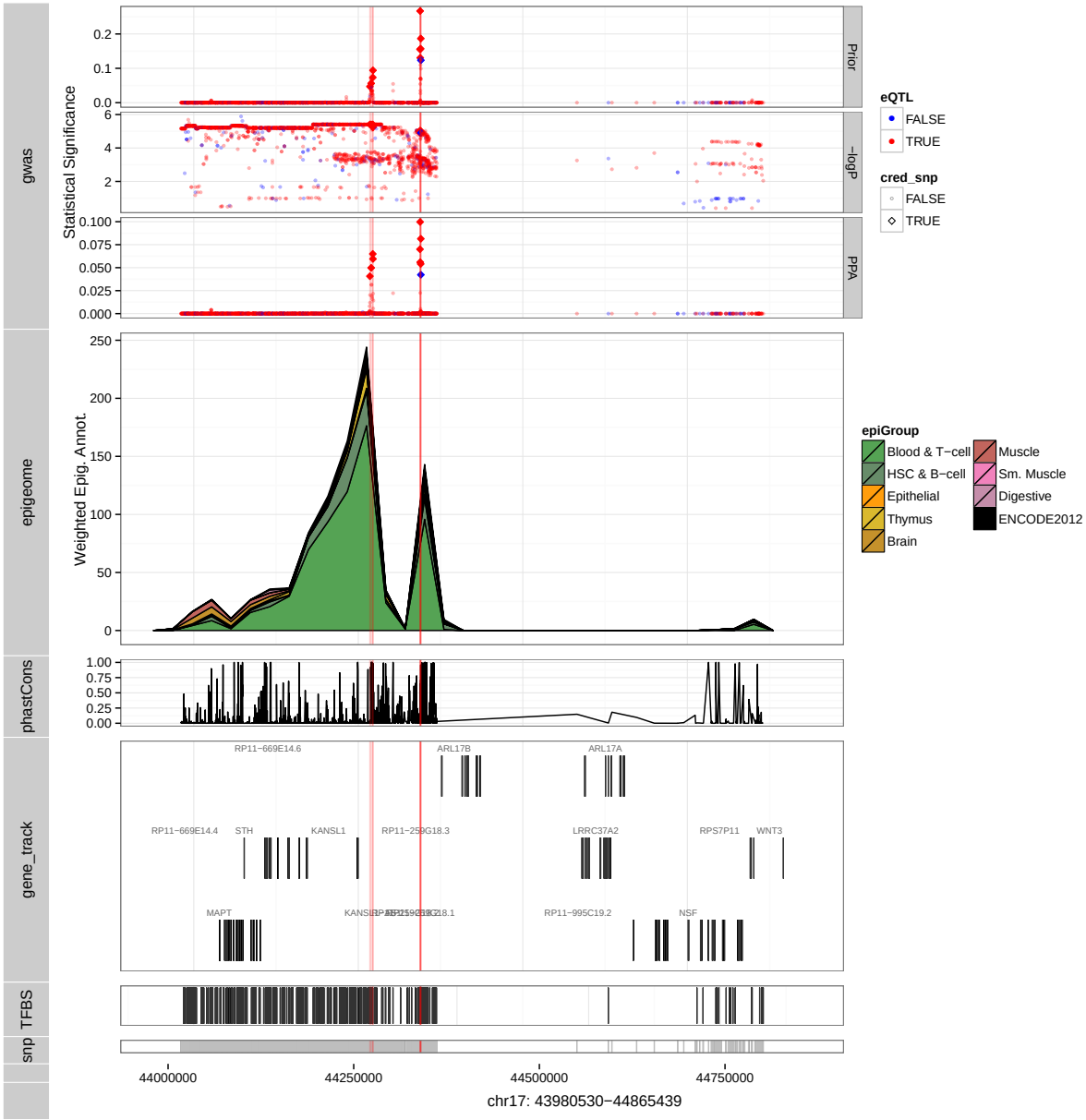
Type 1 Diabetes



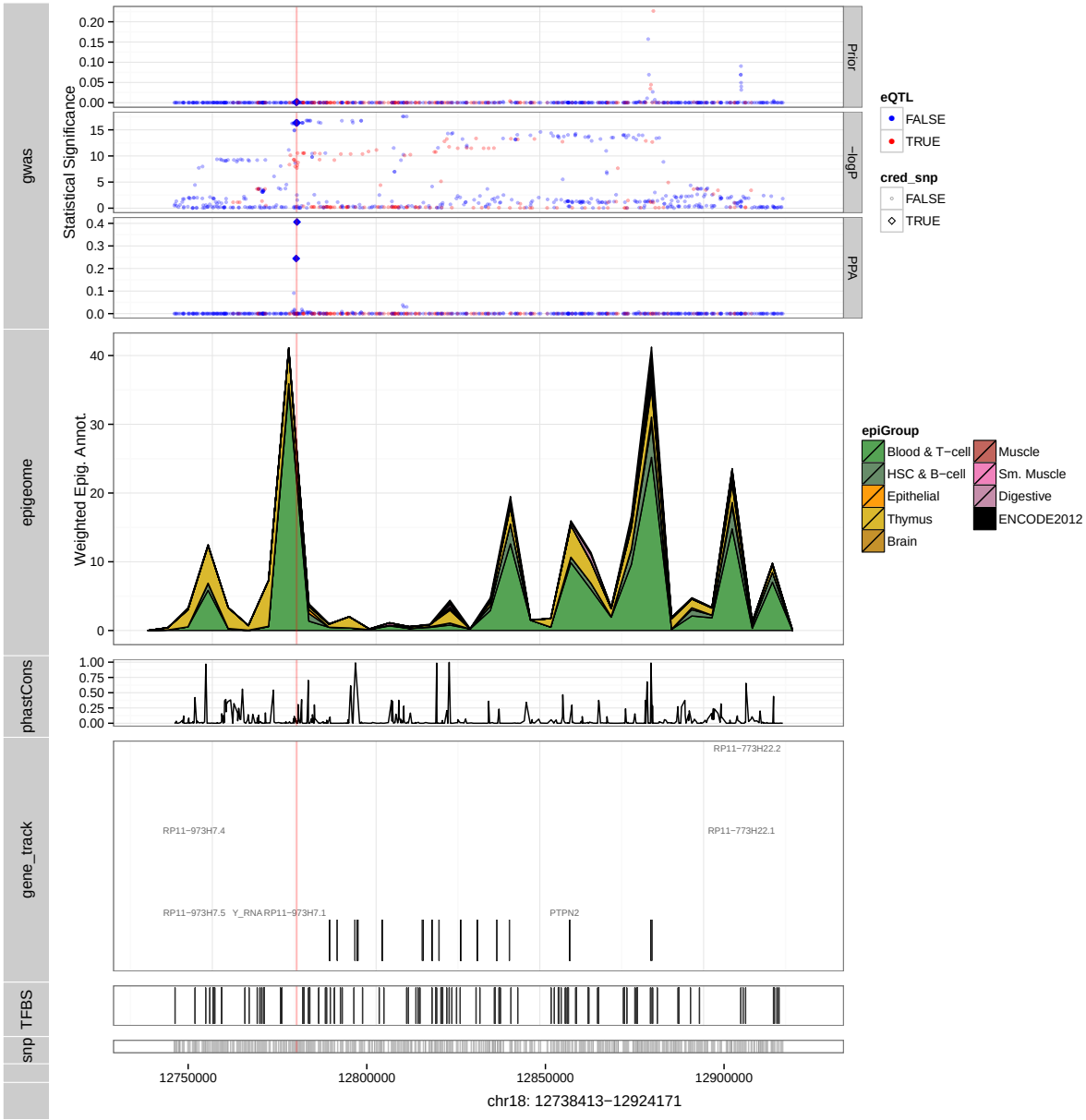
Type 1 Diabetes



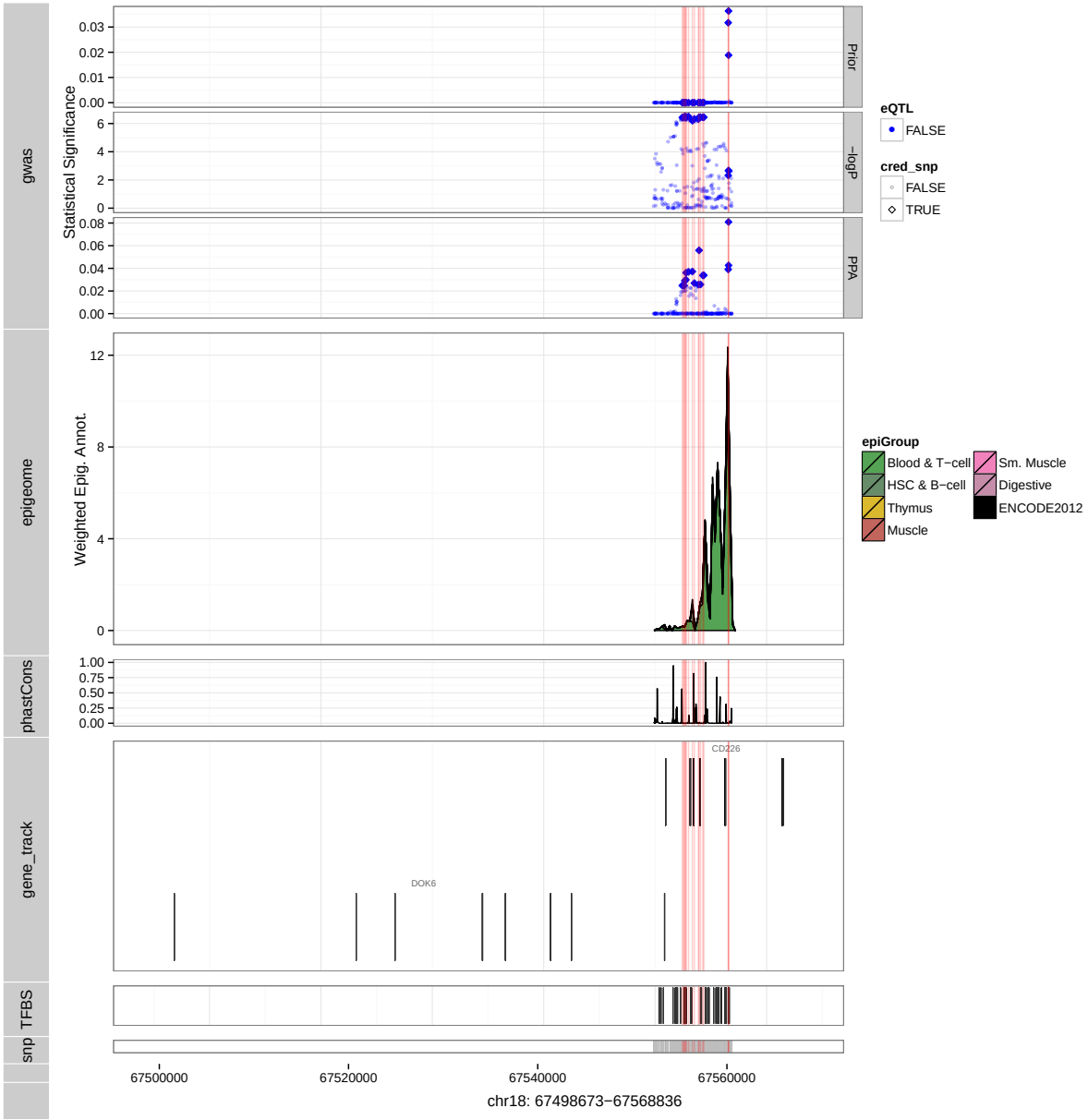
Type 1 Diabetes



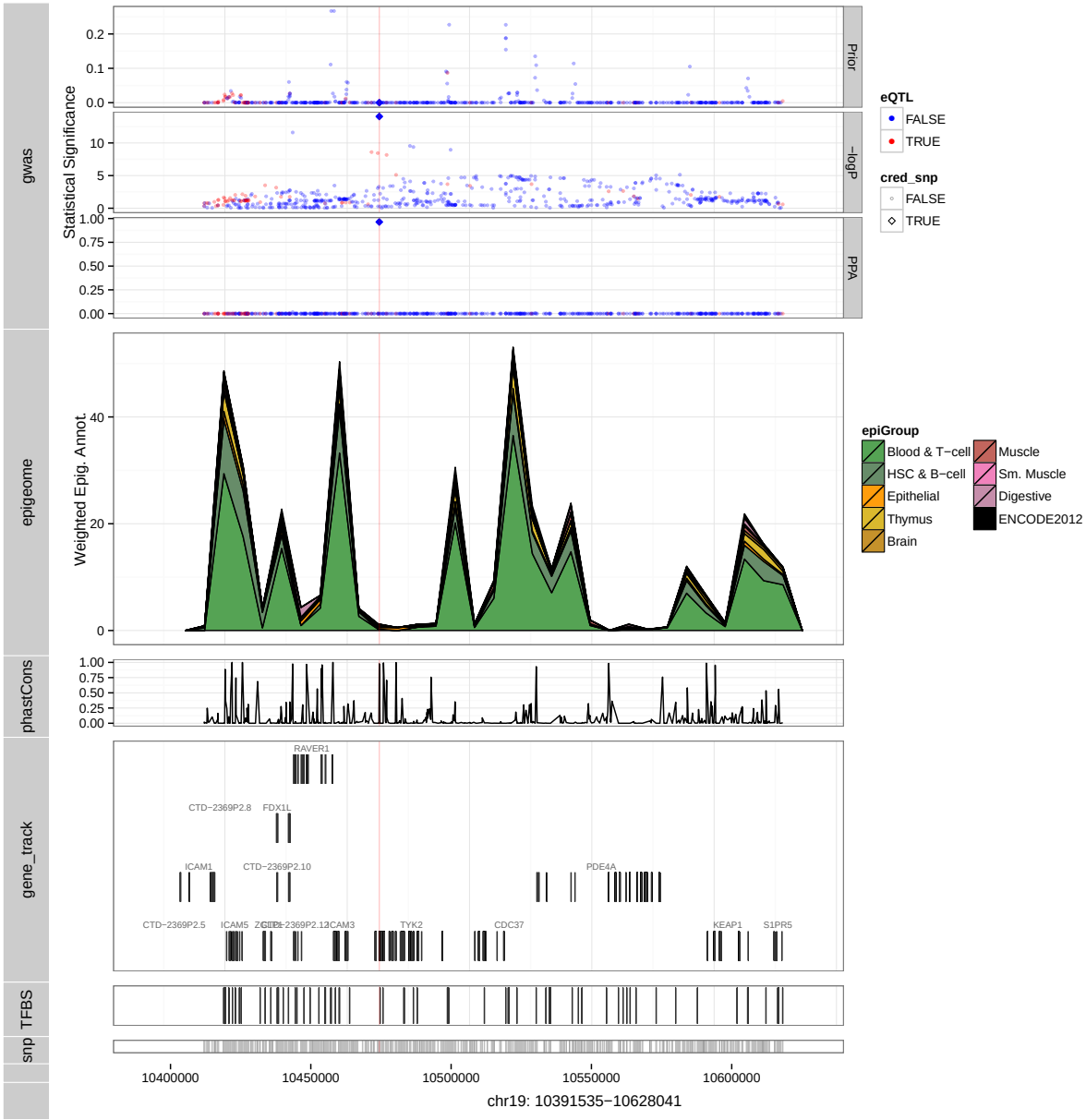
Type 1 Diabetes



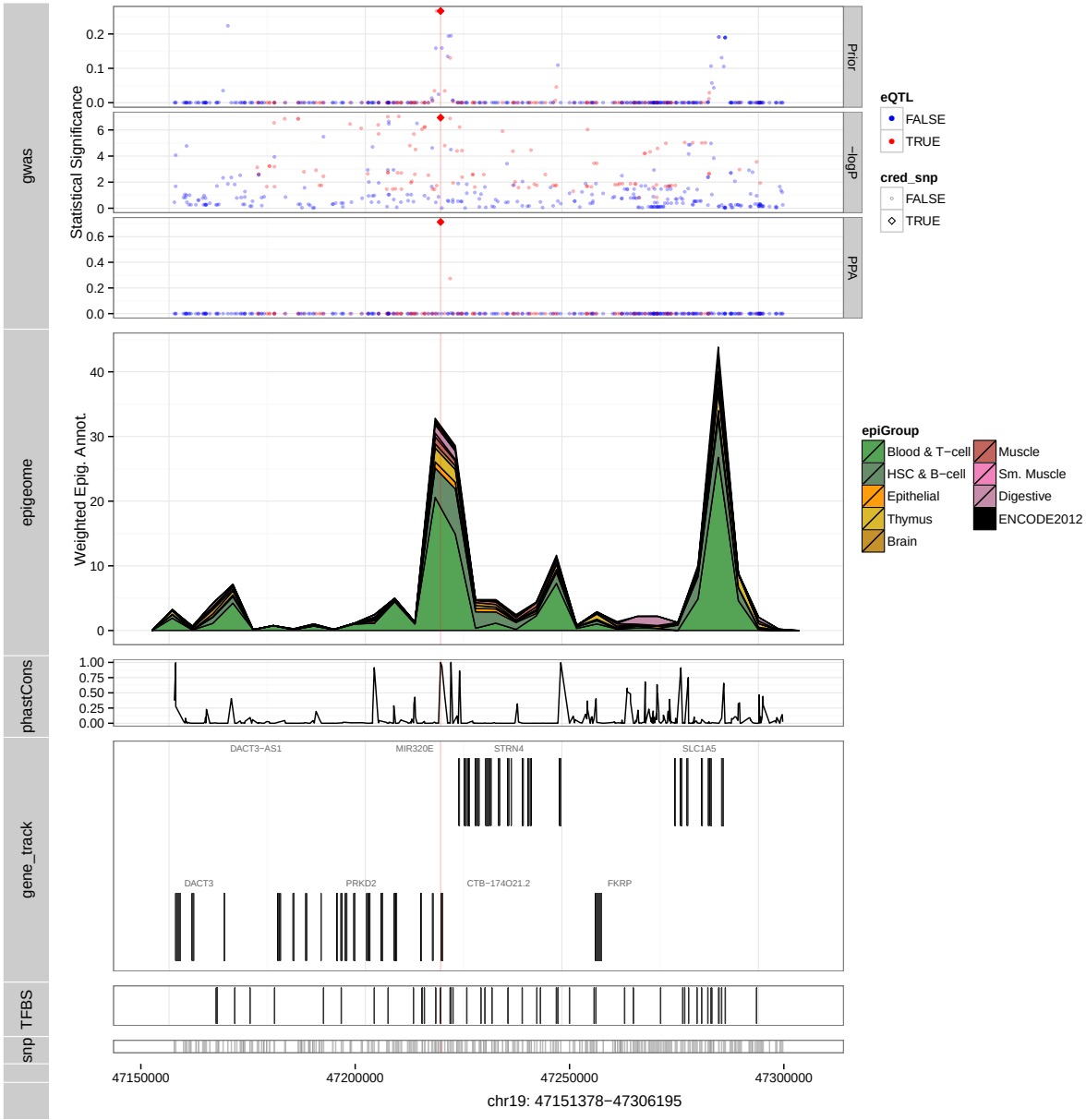
Type 1 Diabetes



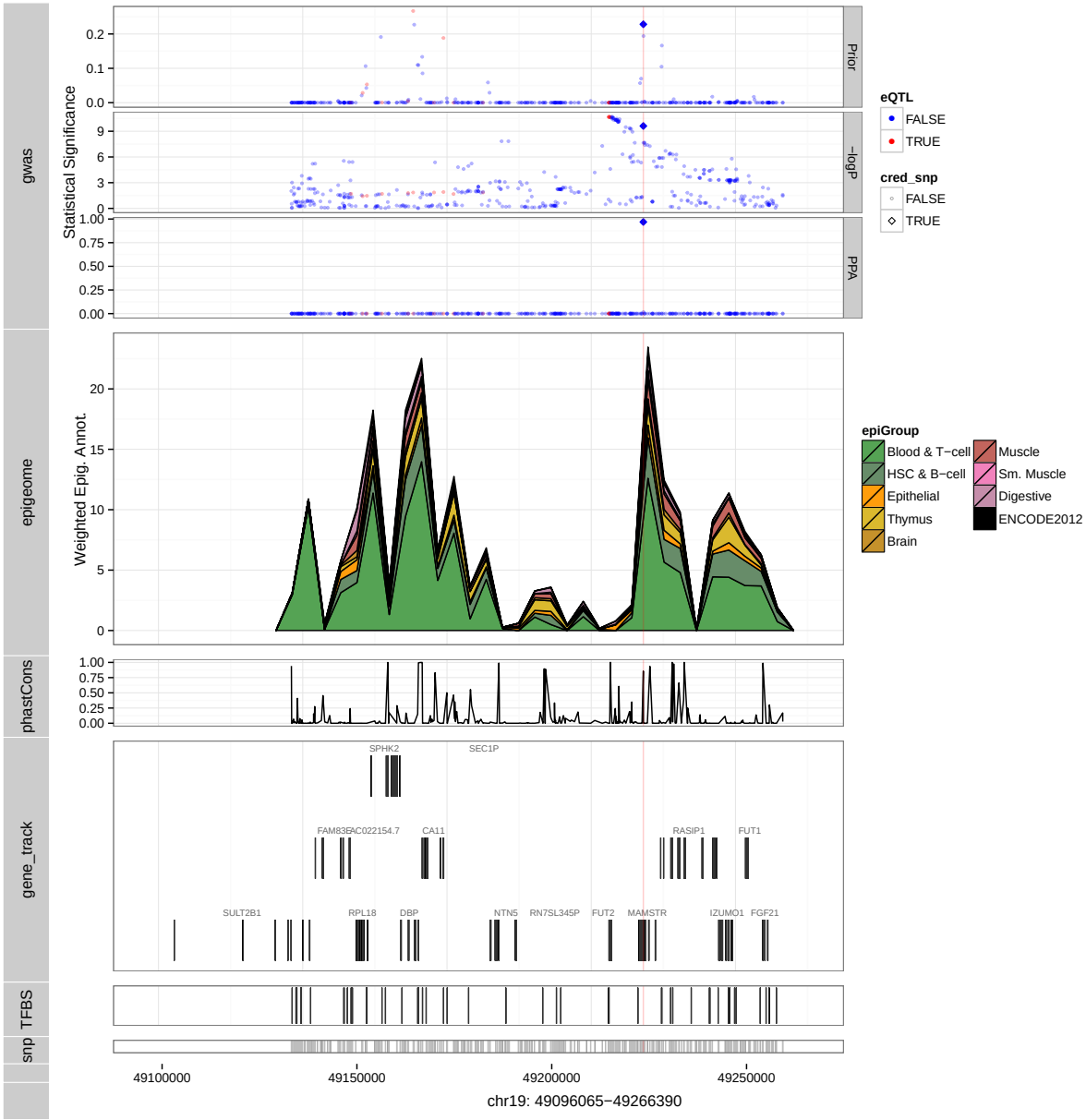
Type 1 Diabetes



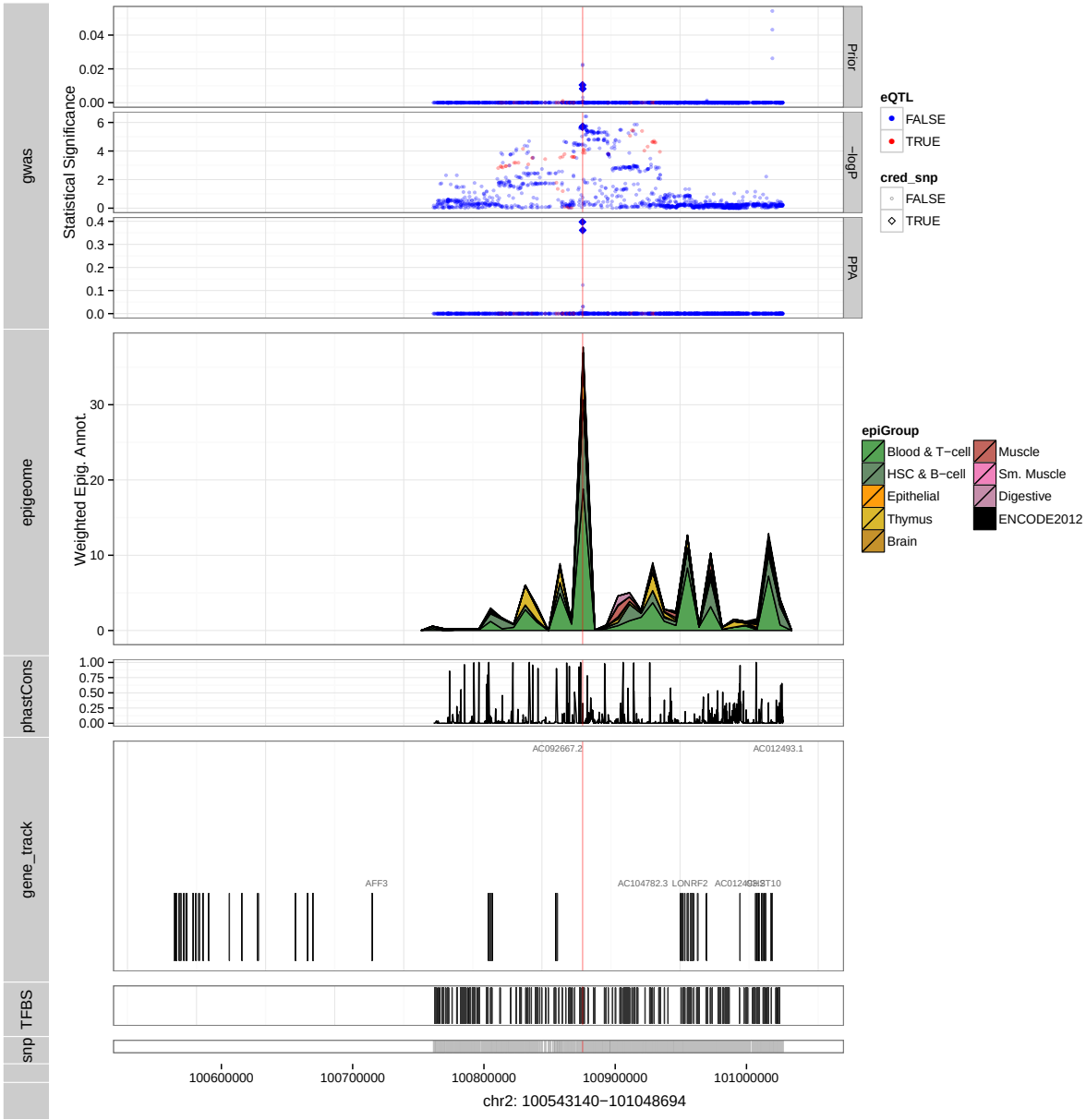
Type 1 Diabetes



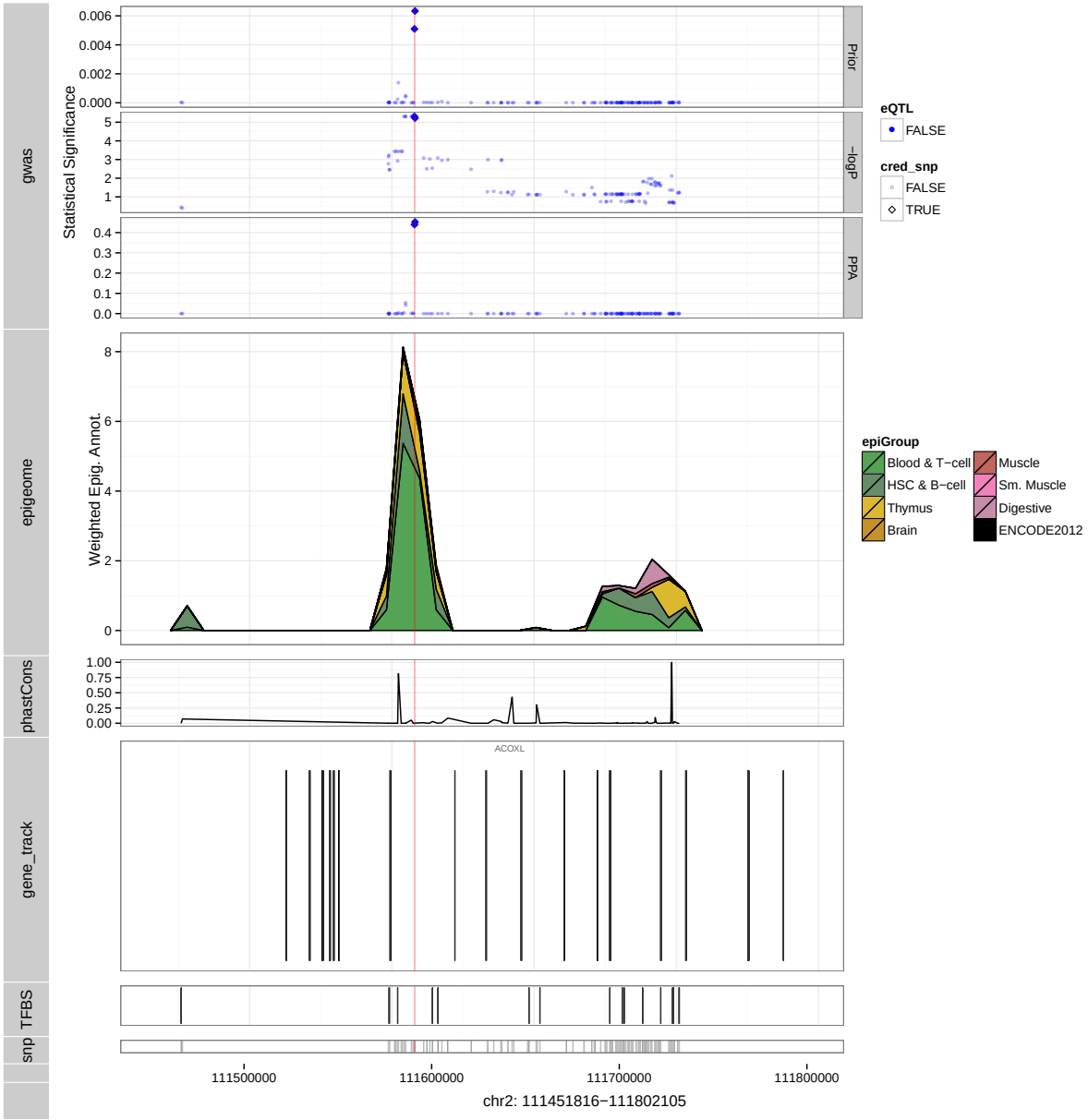
Type 1 Diabetes



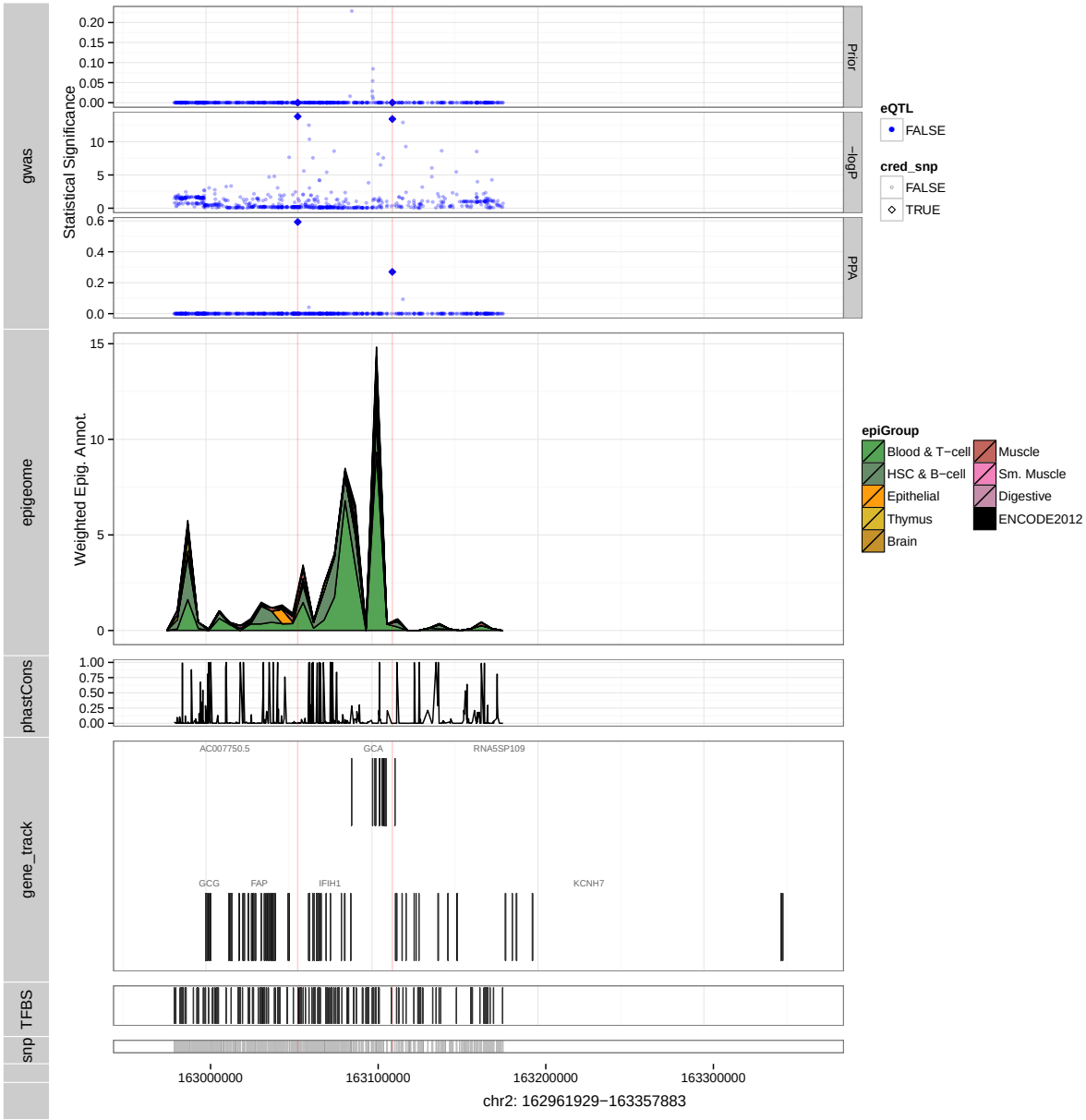
Type 1 Diabetes



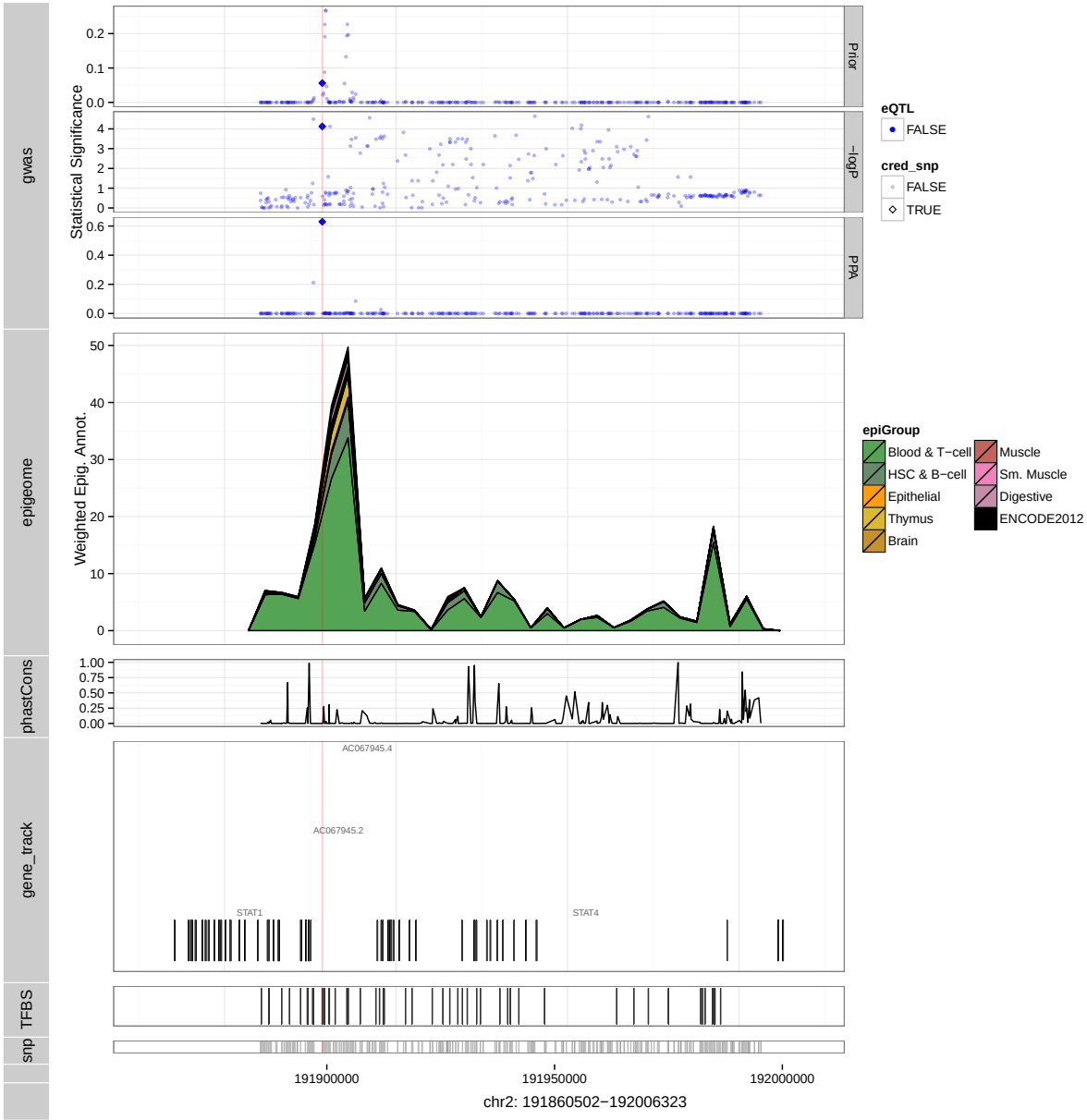
Type 1 Diabetes



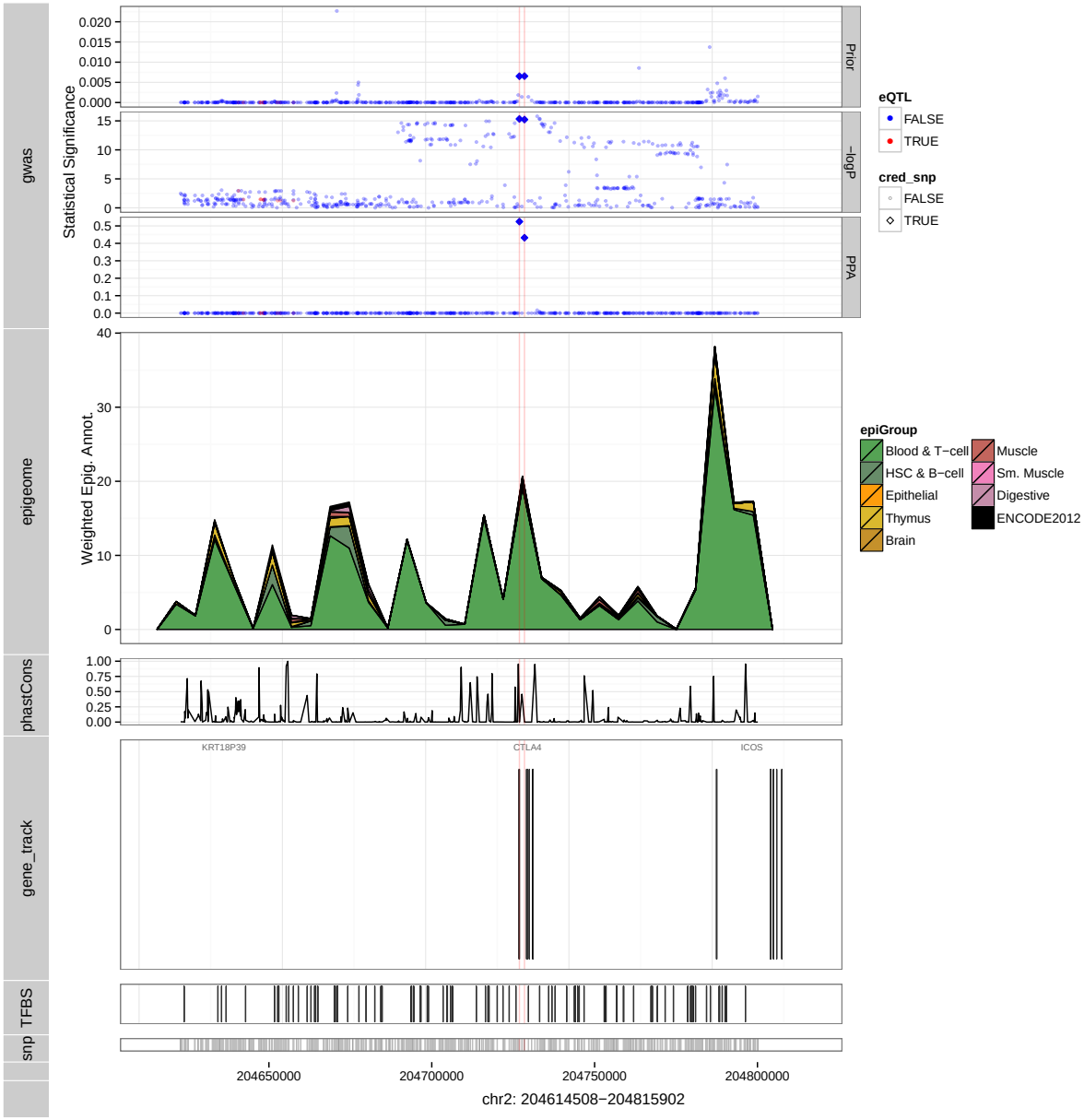
Type 1 Diabetes



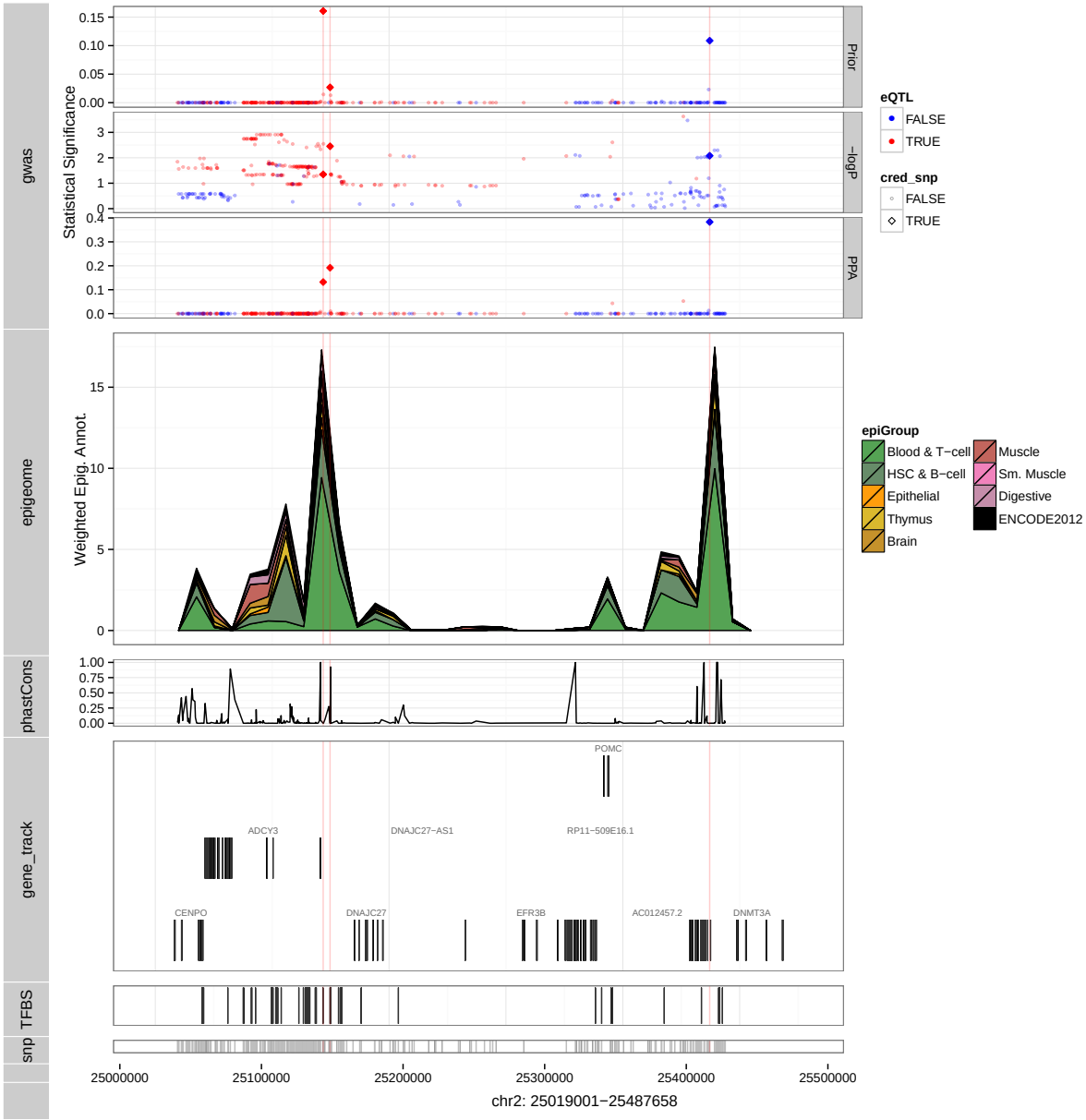
Type 1 Diabetes



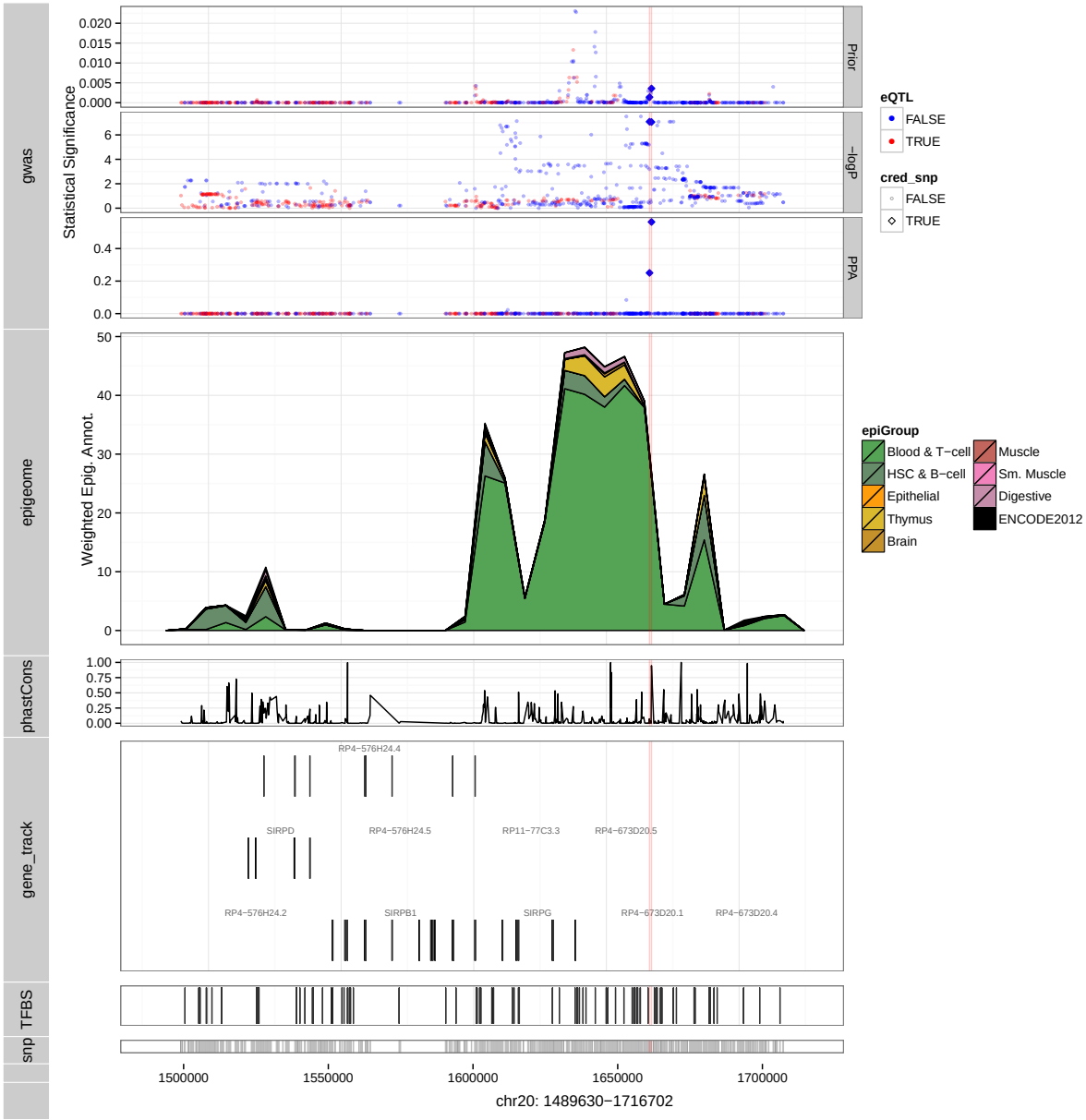
Type 1 Diabetes



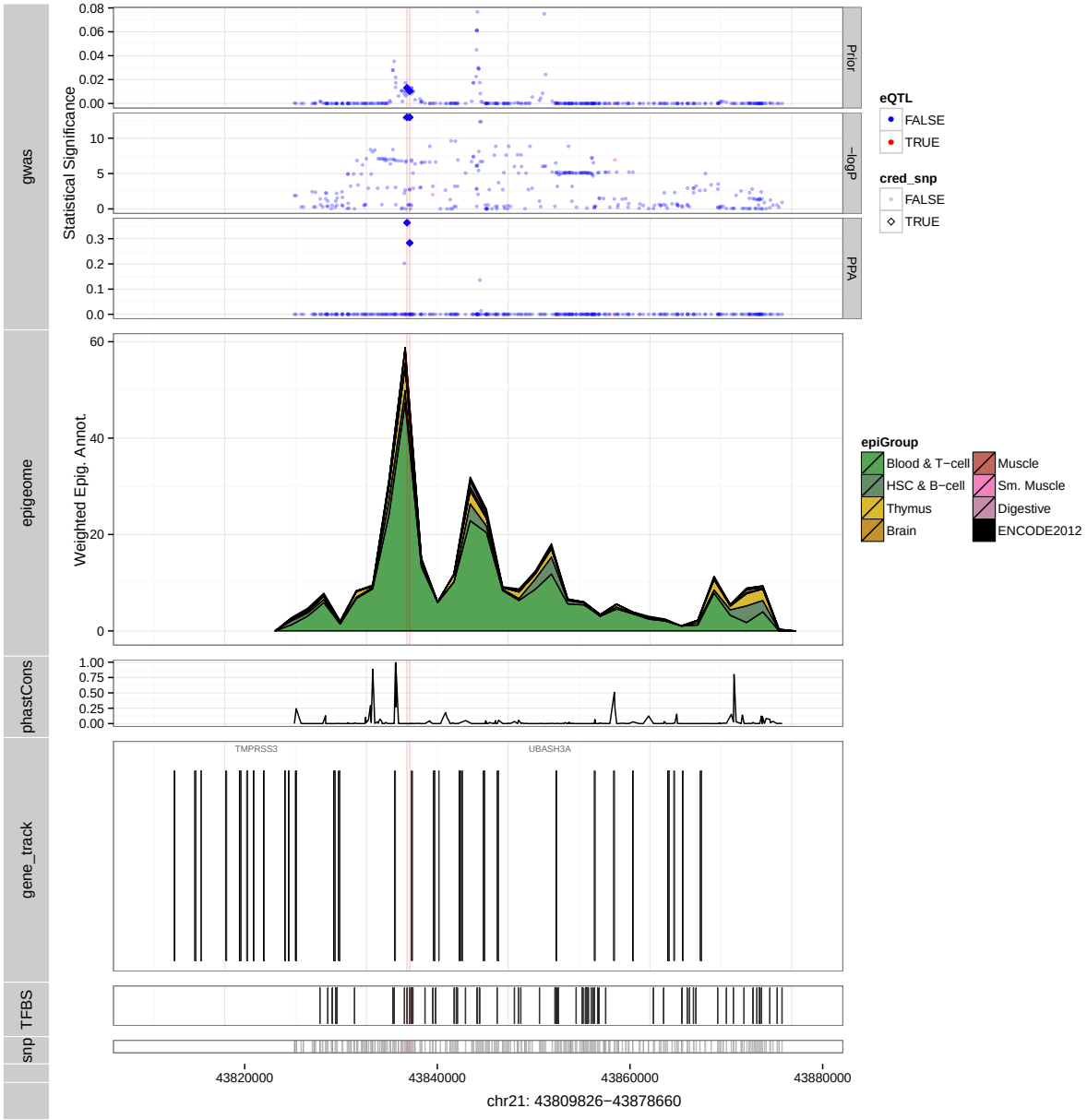
Type 1 Diabetes



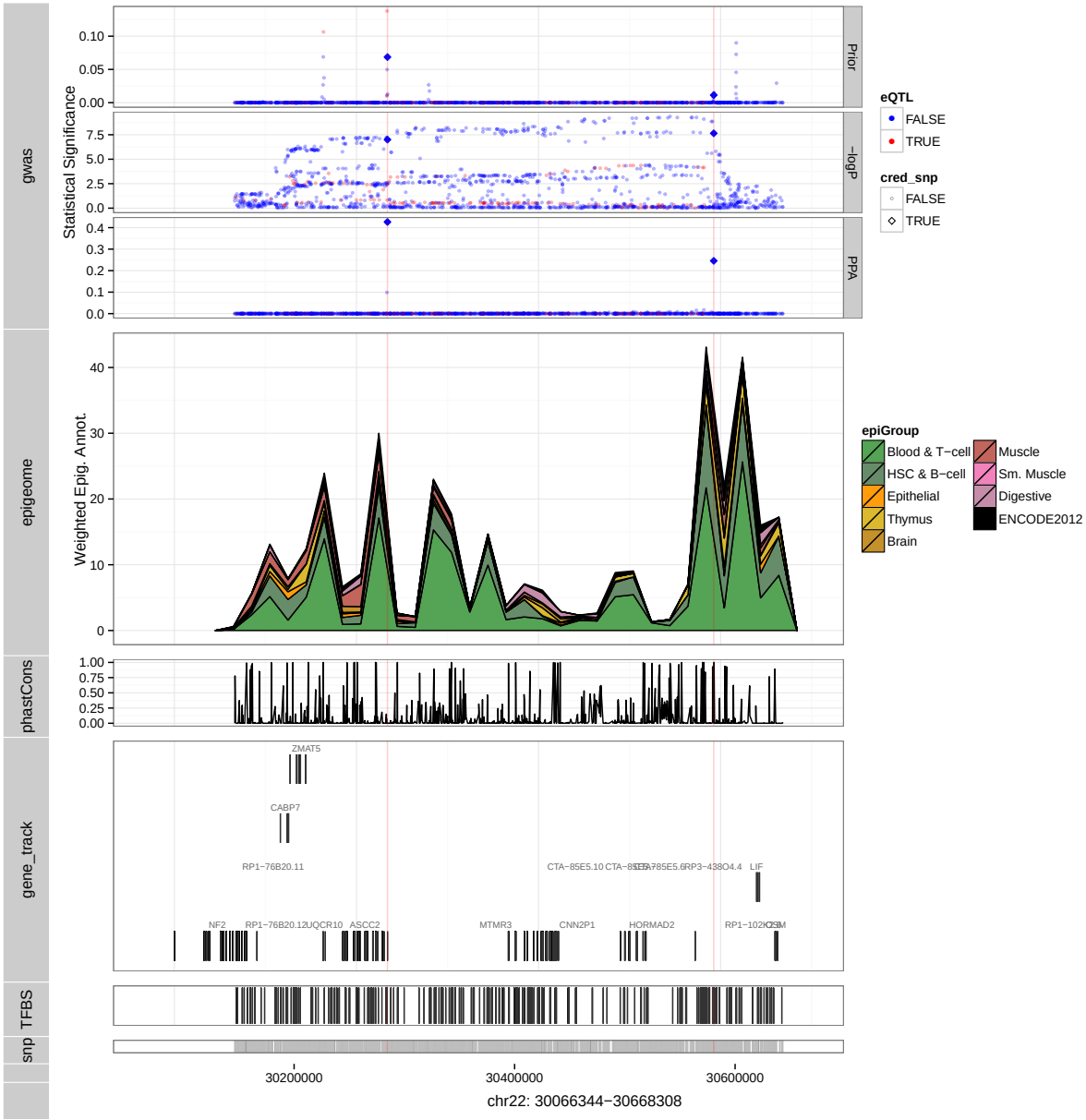
Type 1 Diabetes



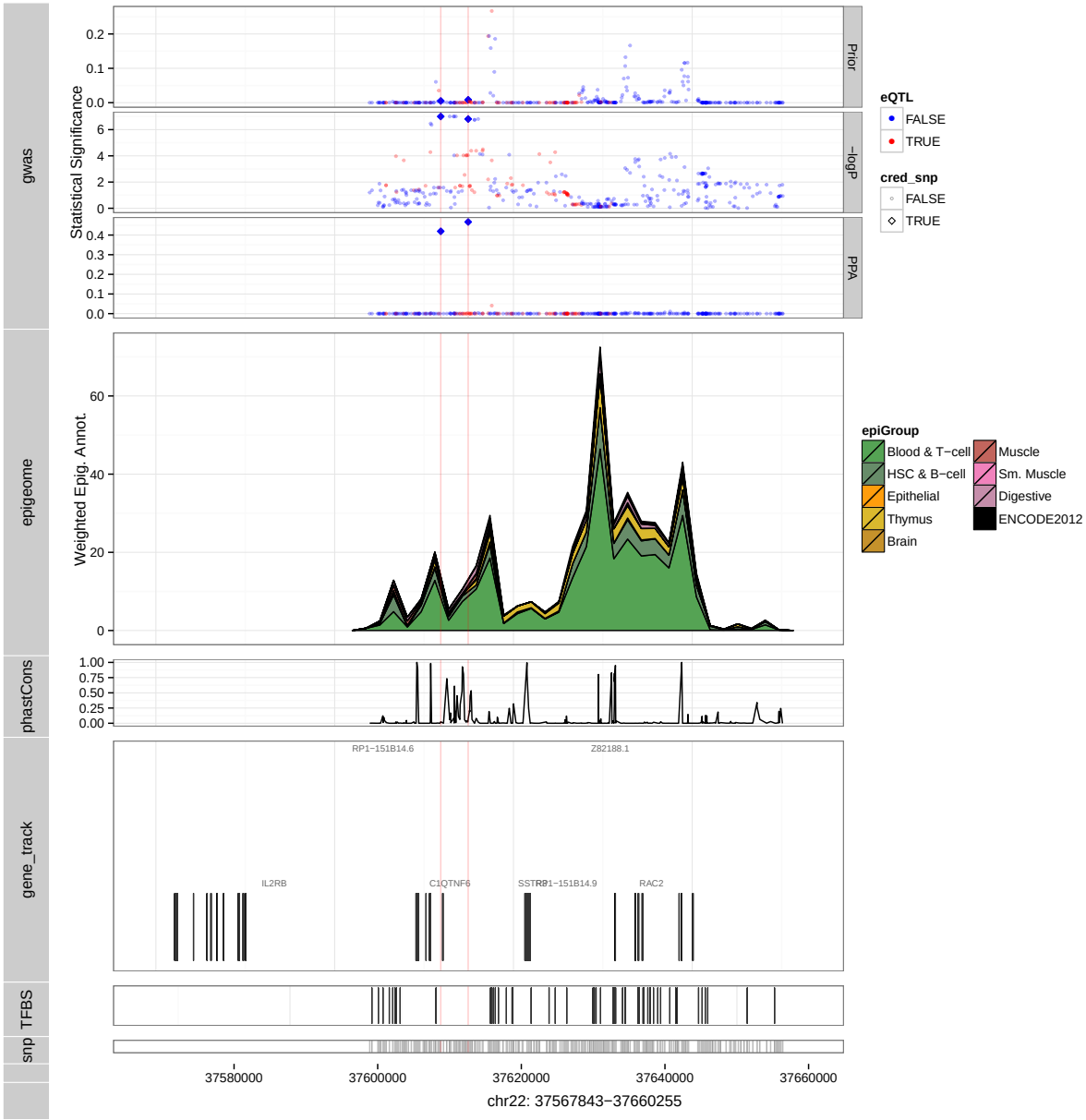
Type 1 Diabetes



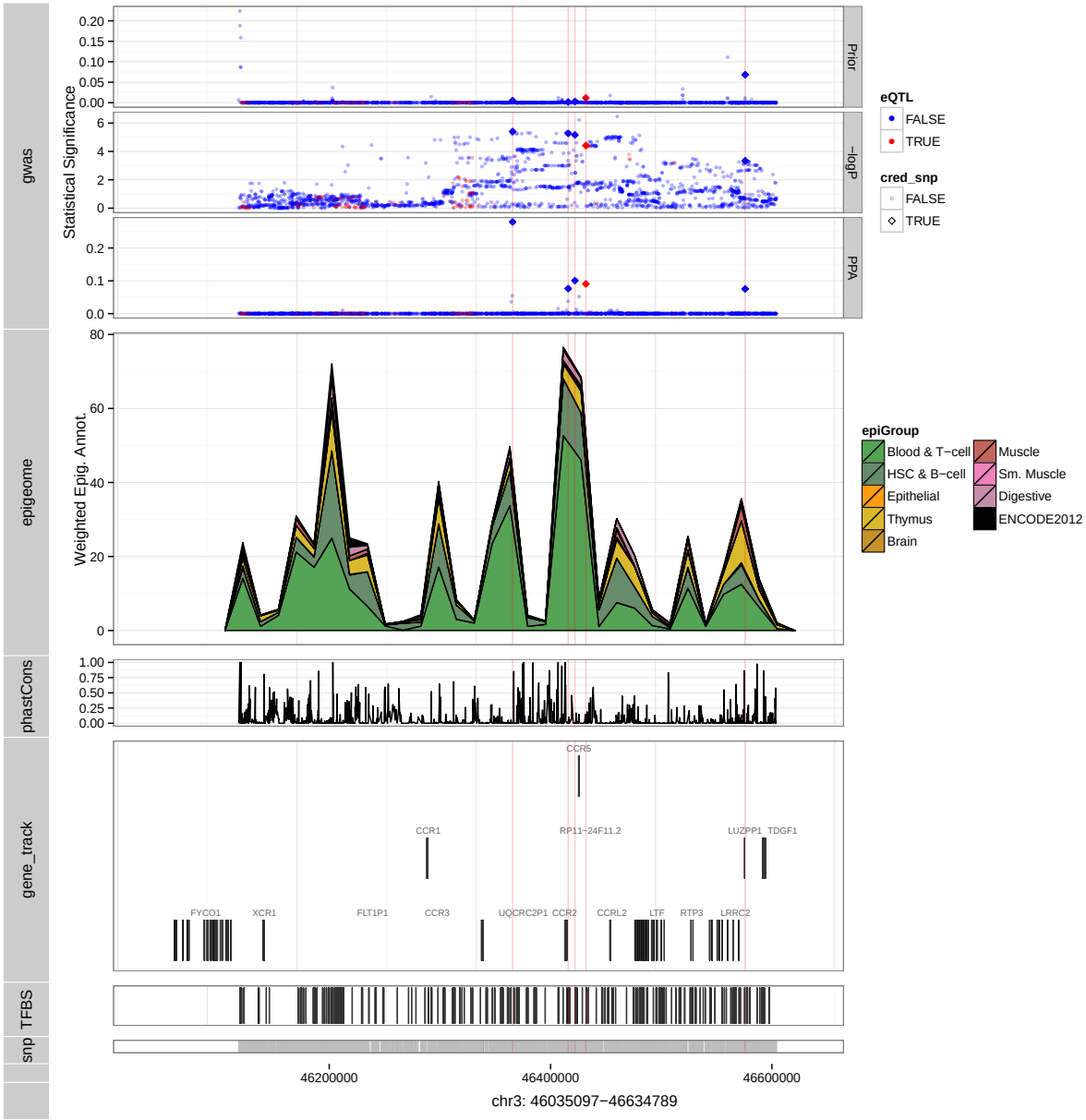
Type 1 Diabetes



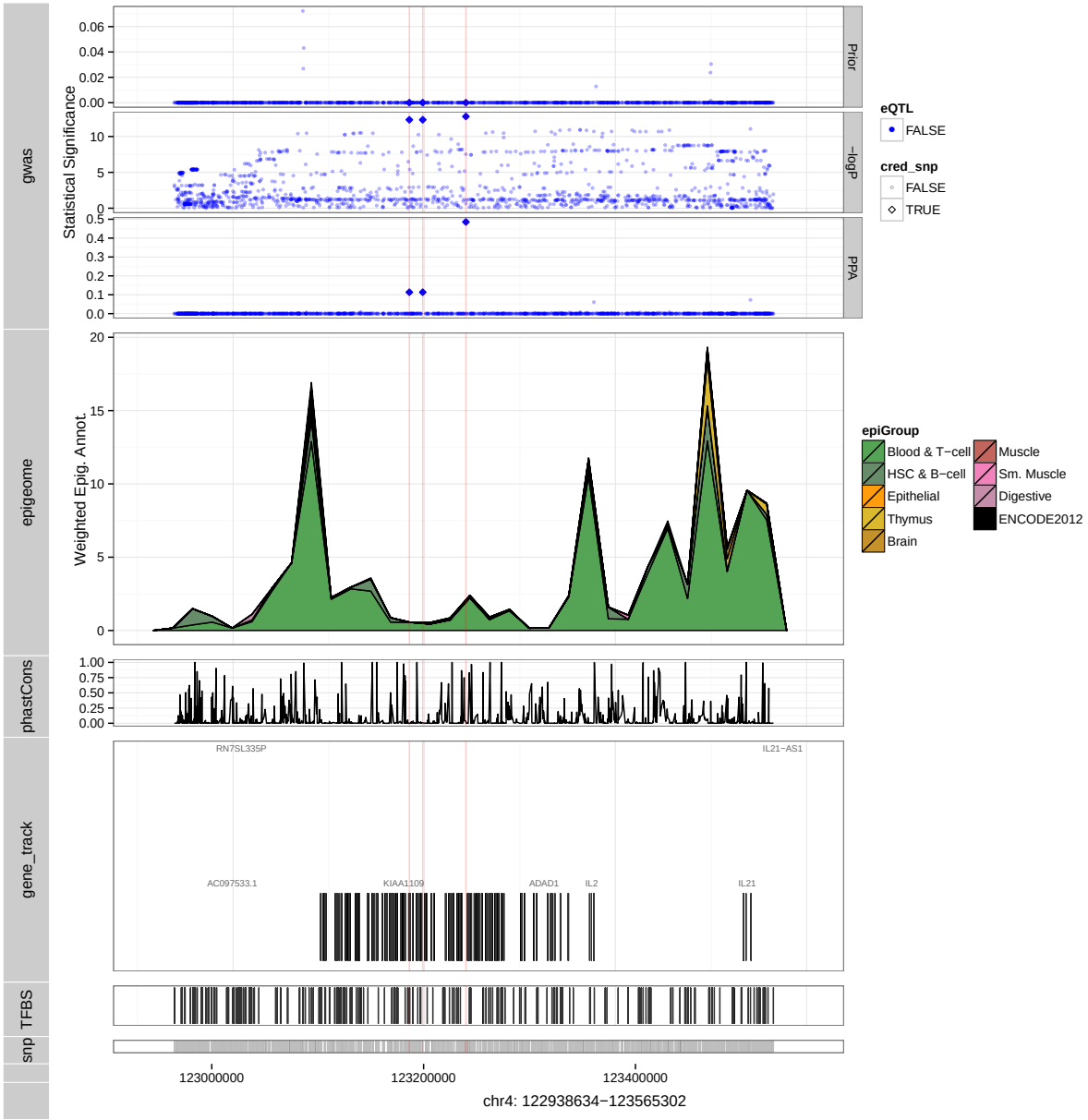
Type 1 Diabetes



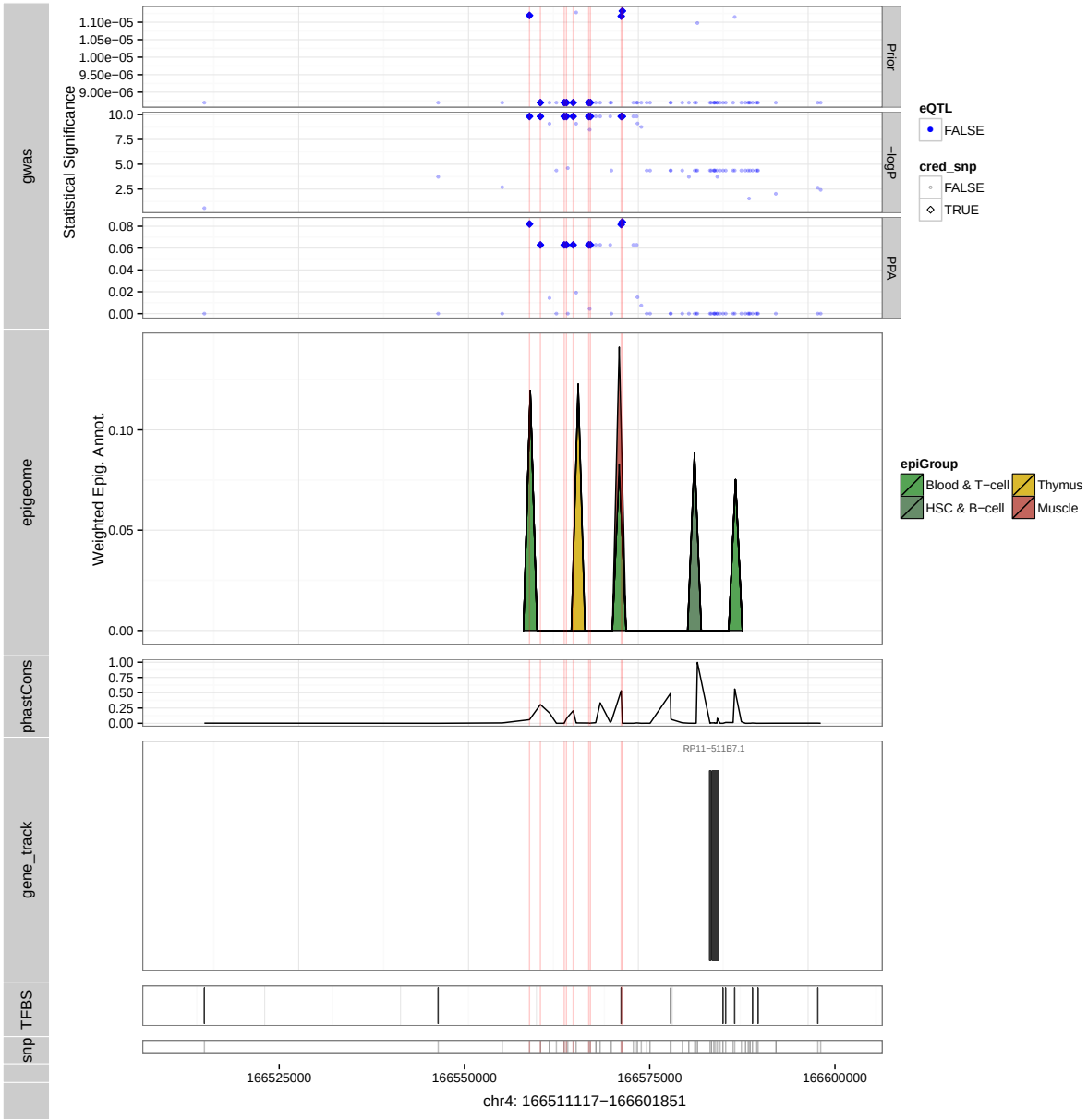
Type 1 Diabetes



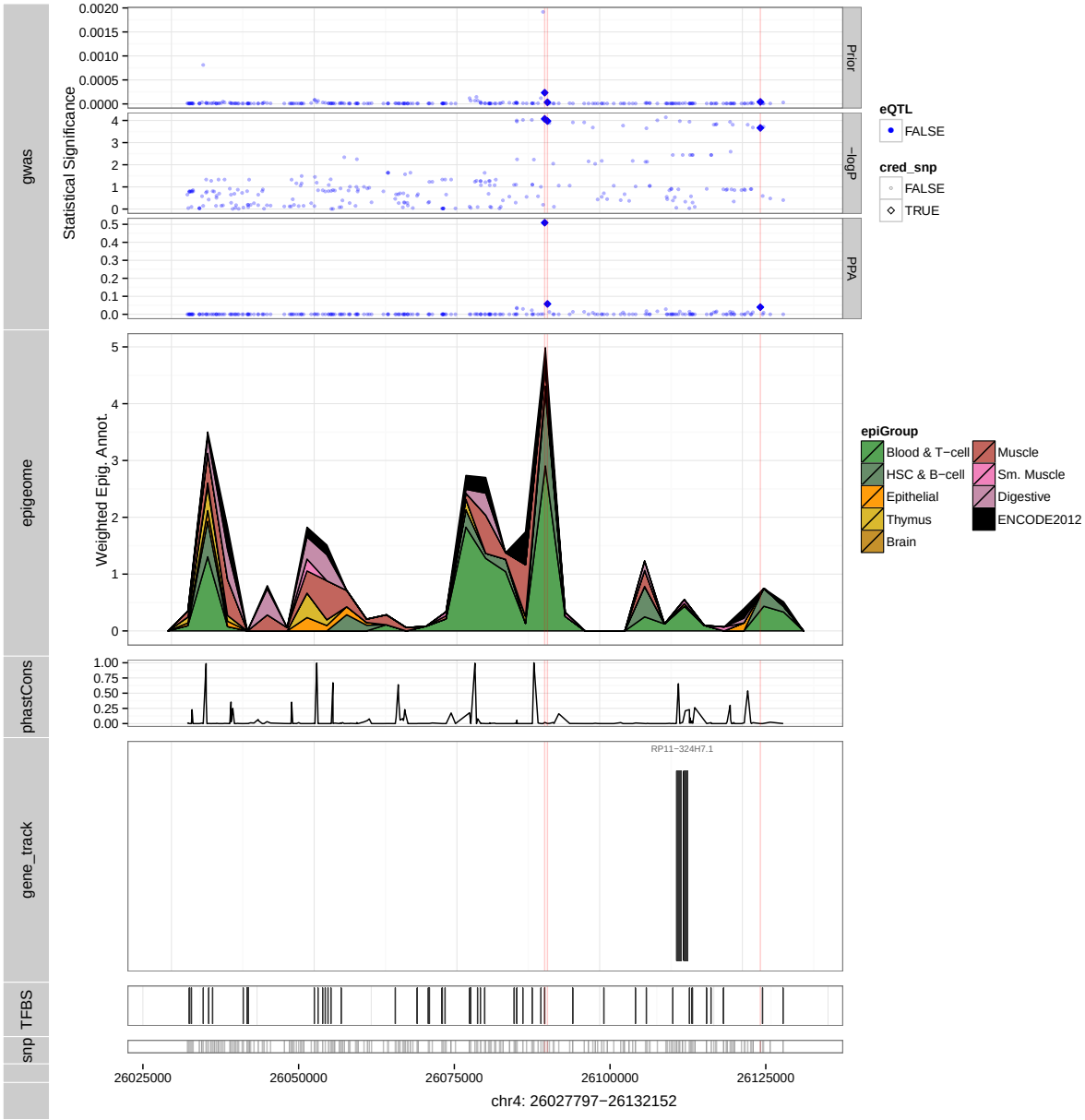
Type 1 Diabetes



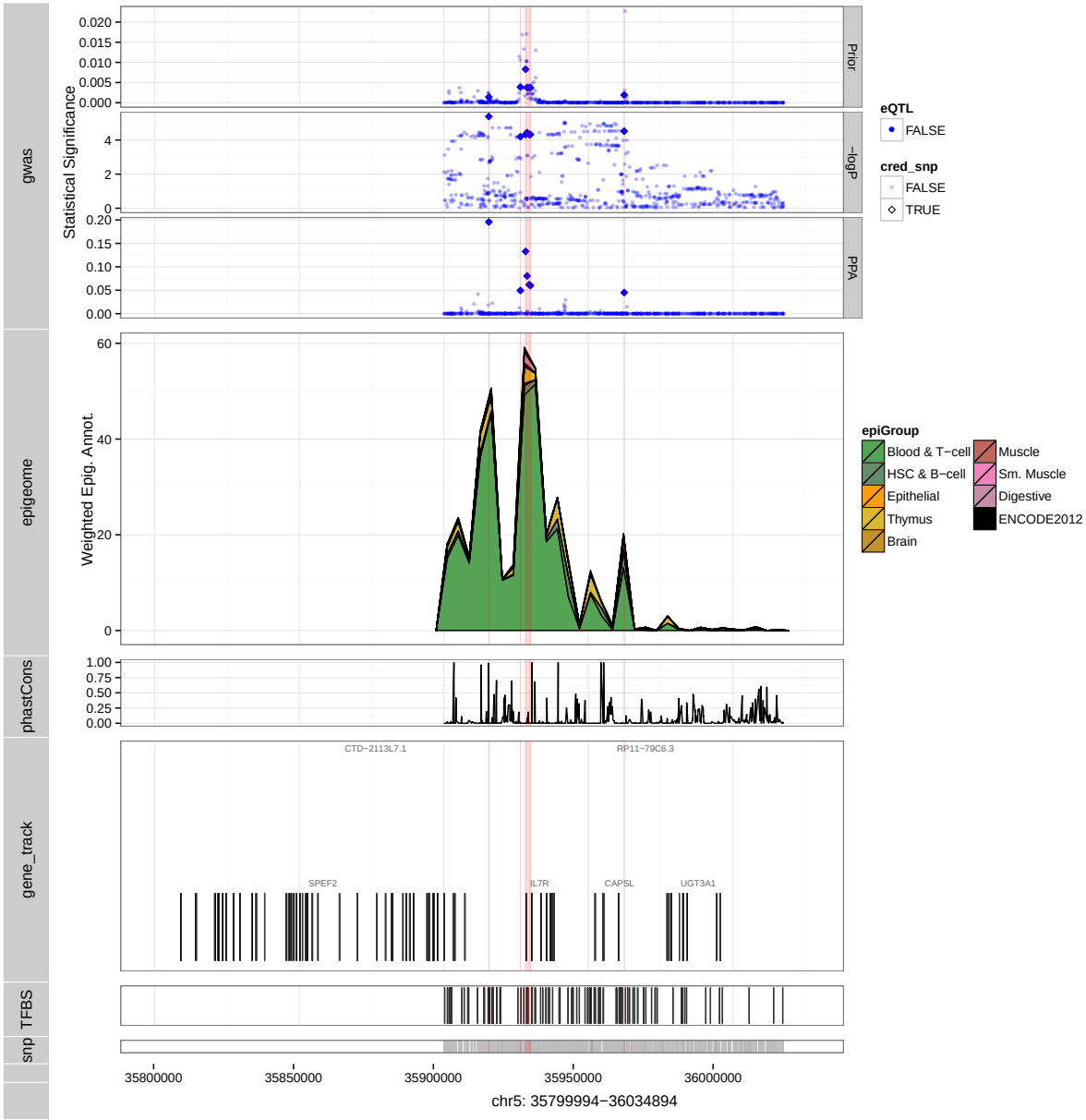
Type 1 Diabetes



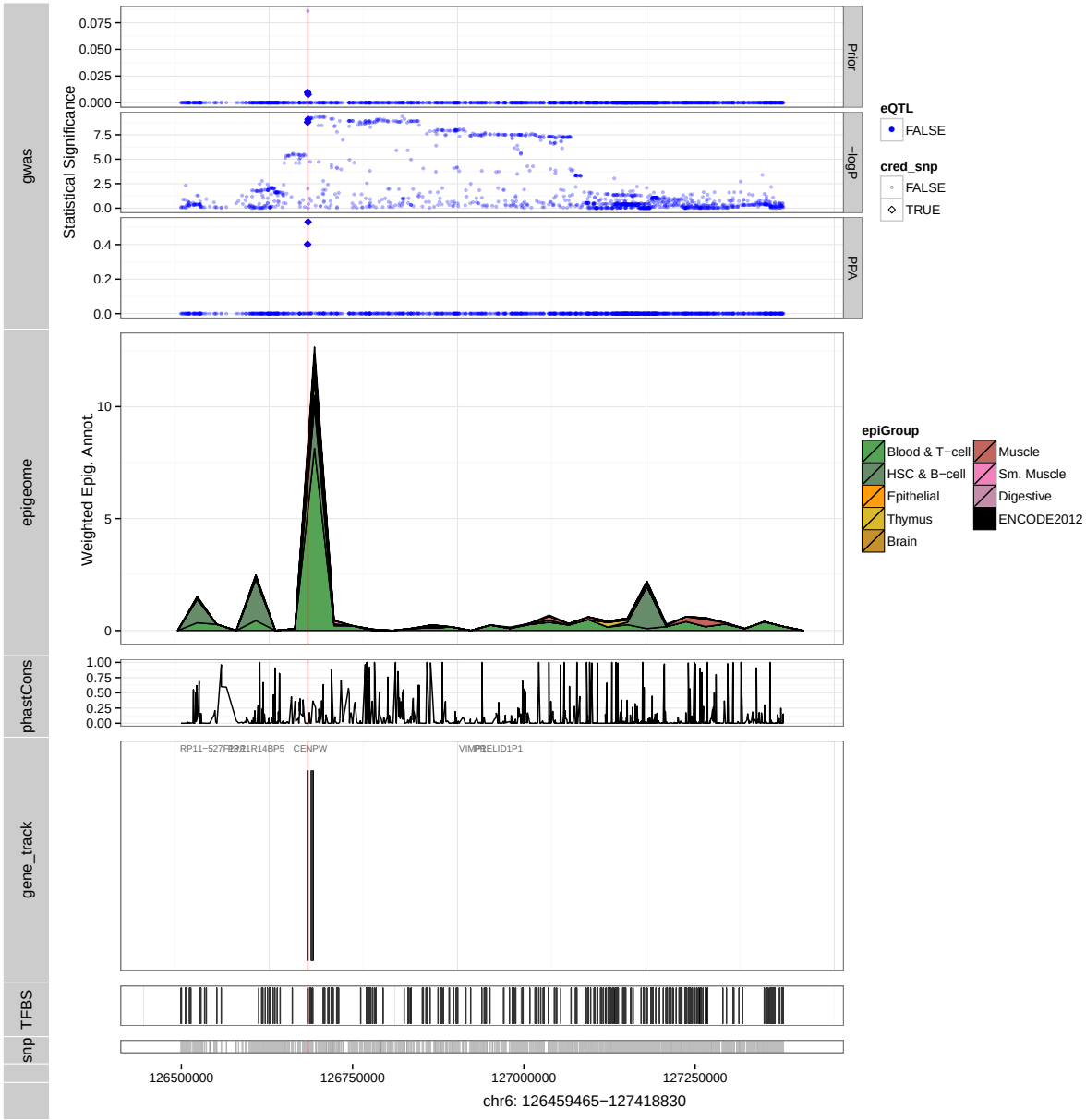
Type 1 Diabetes



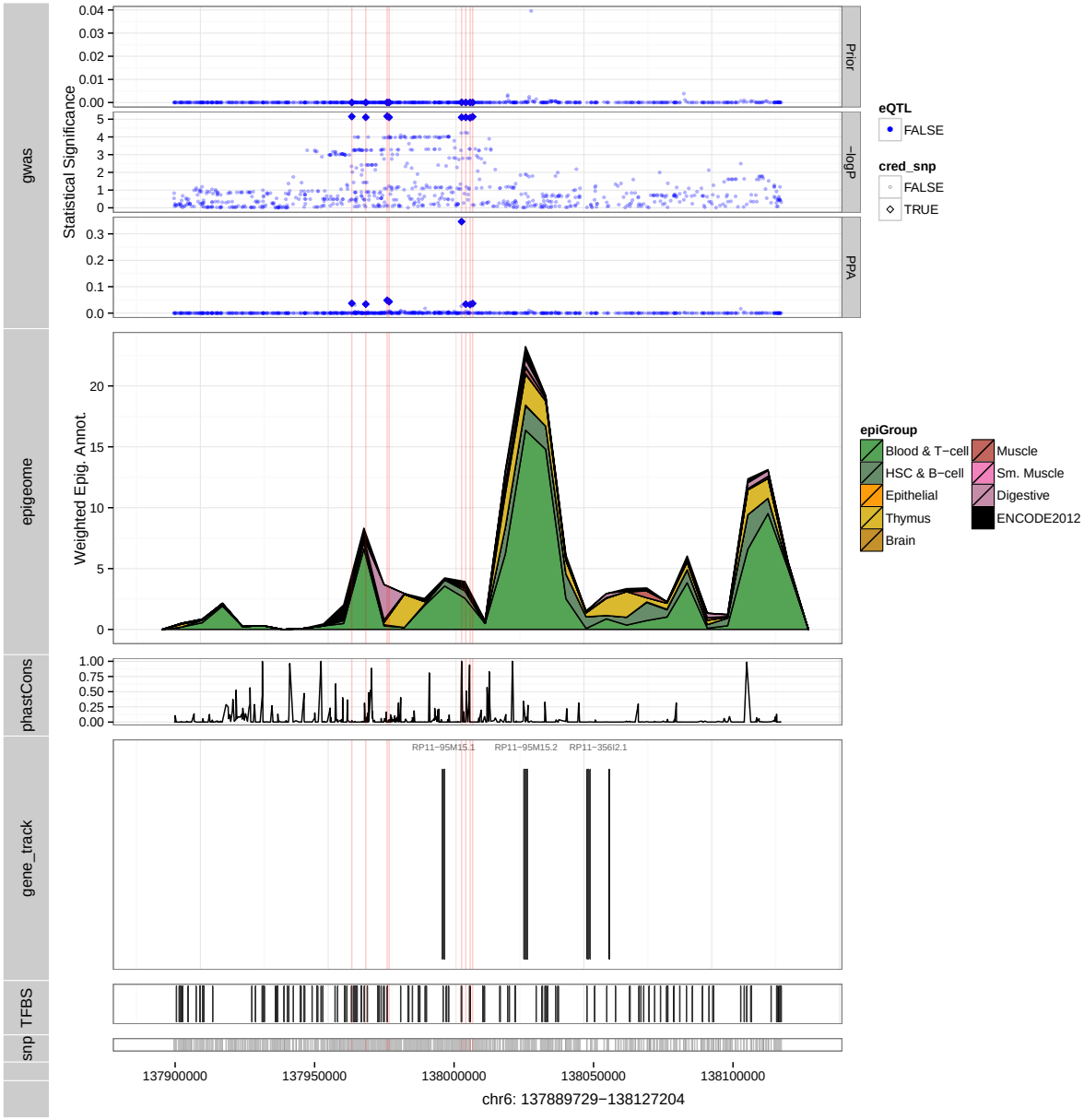
Type 1 Diabetes



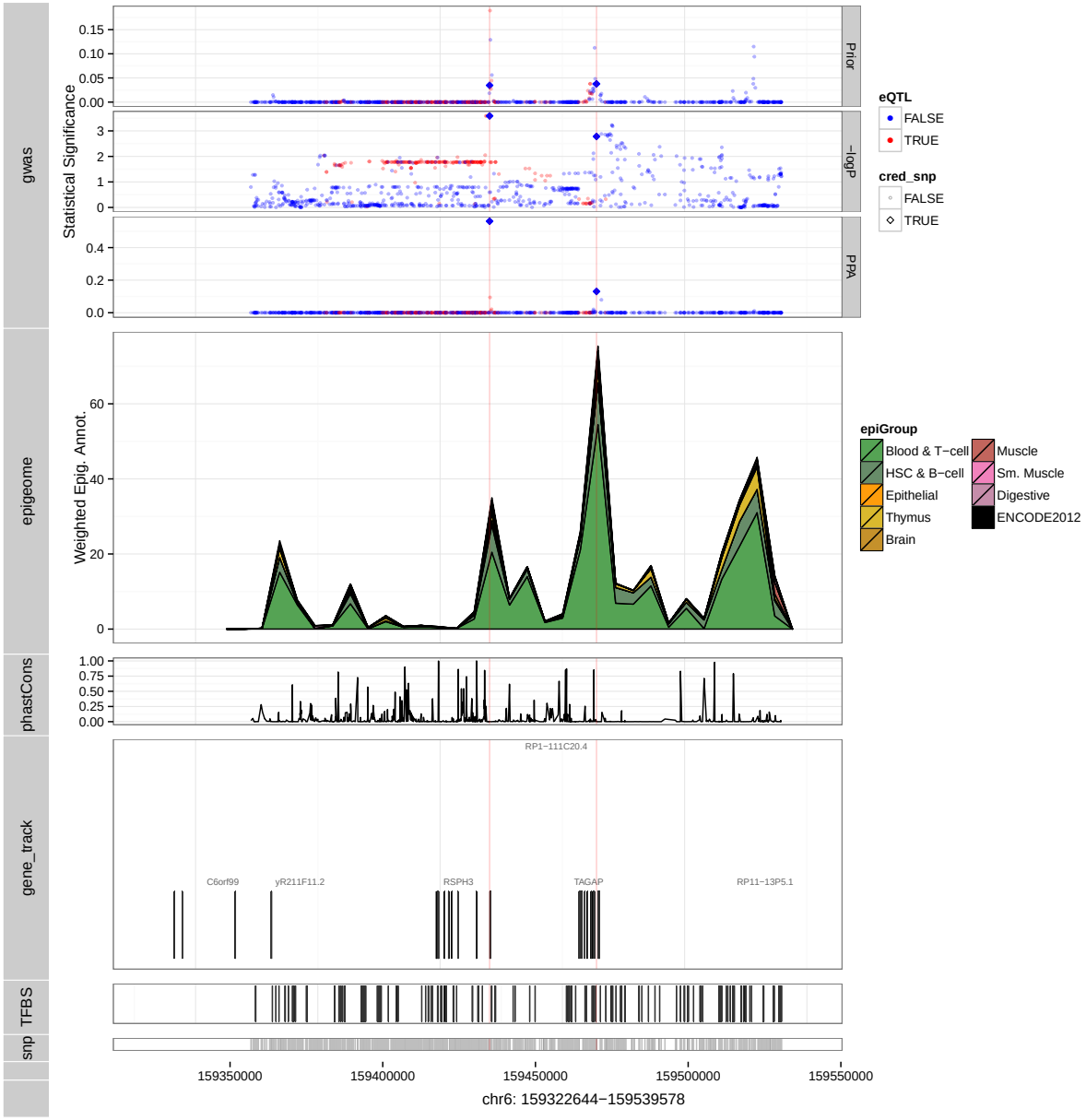
Type 1 Diabetes



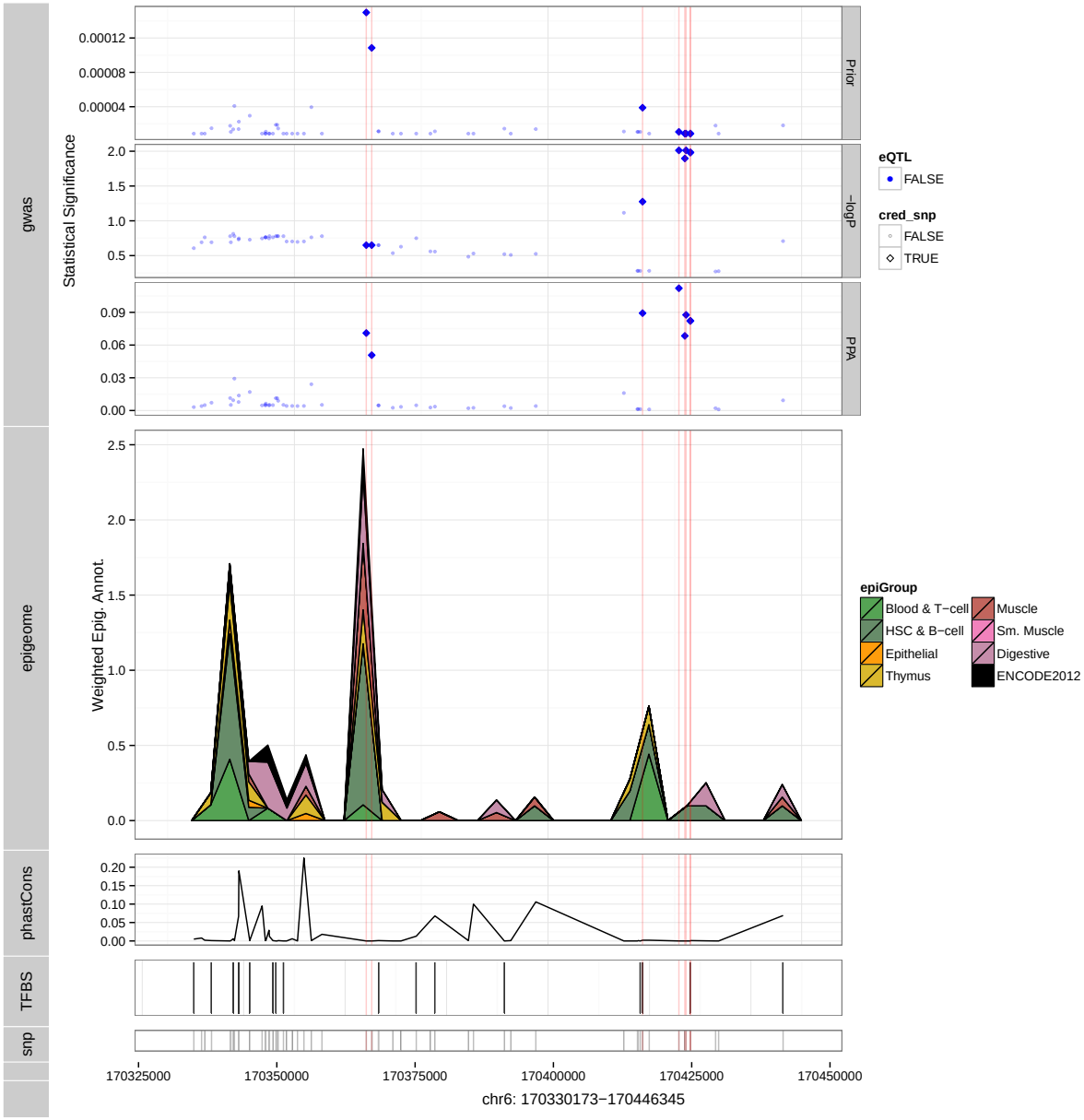
Type 1 Diabetes



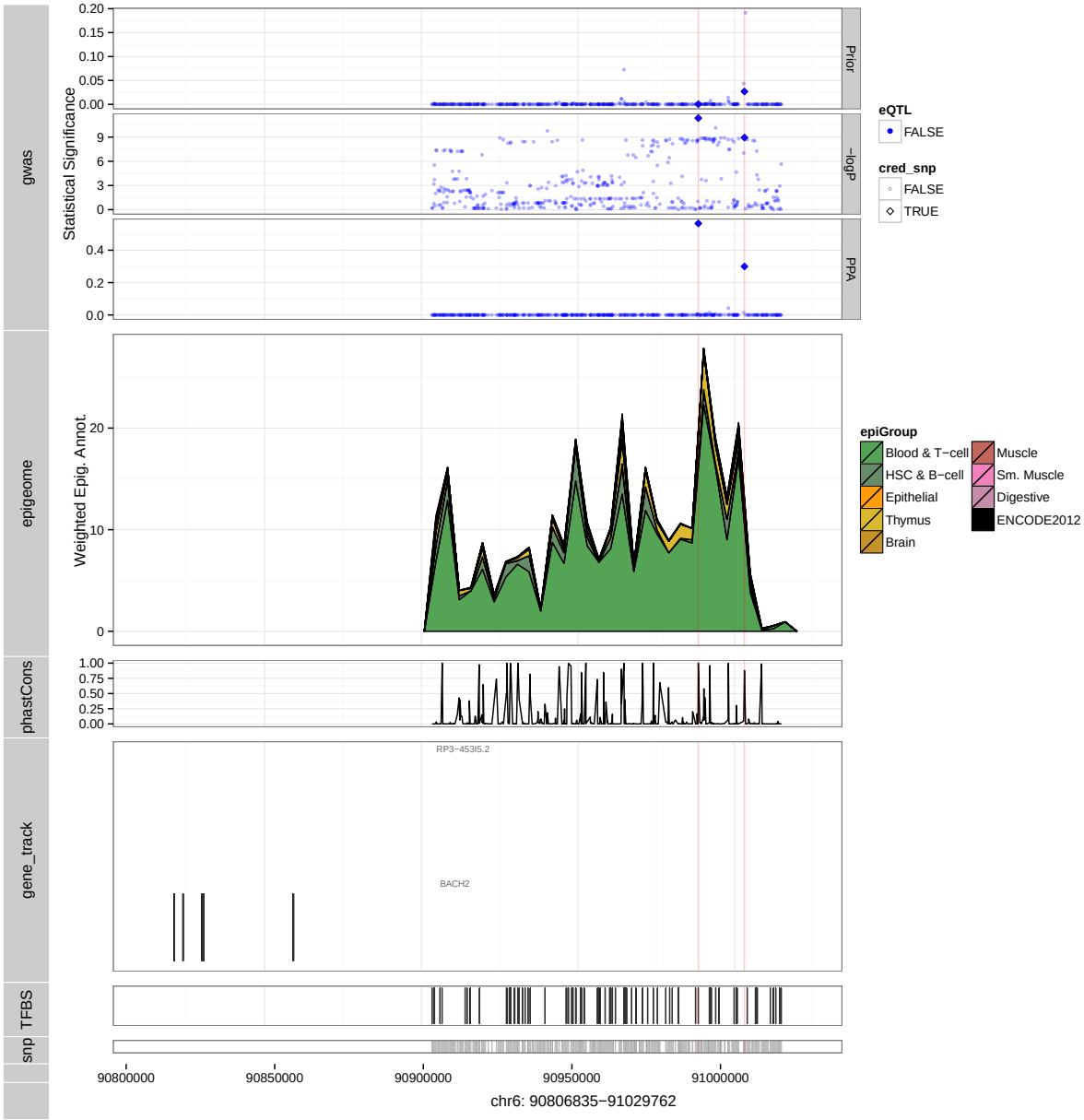
Type 1 Diabetes



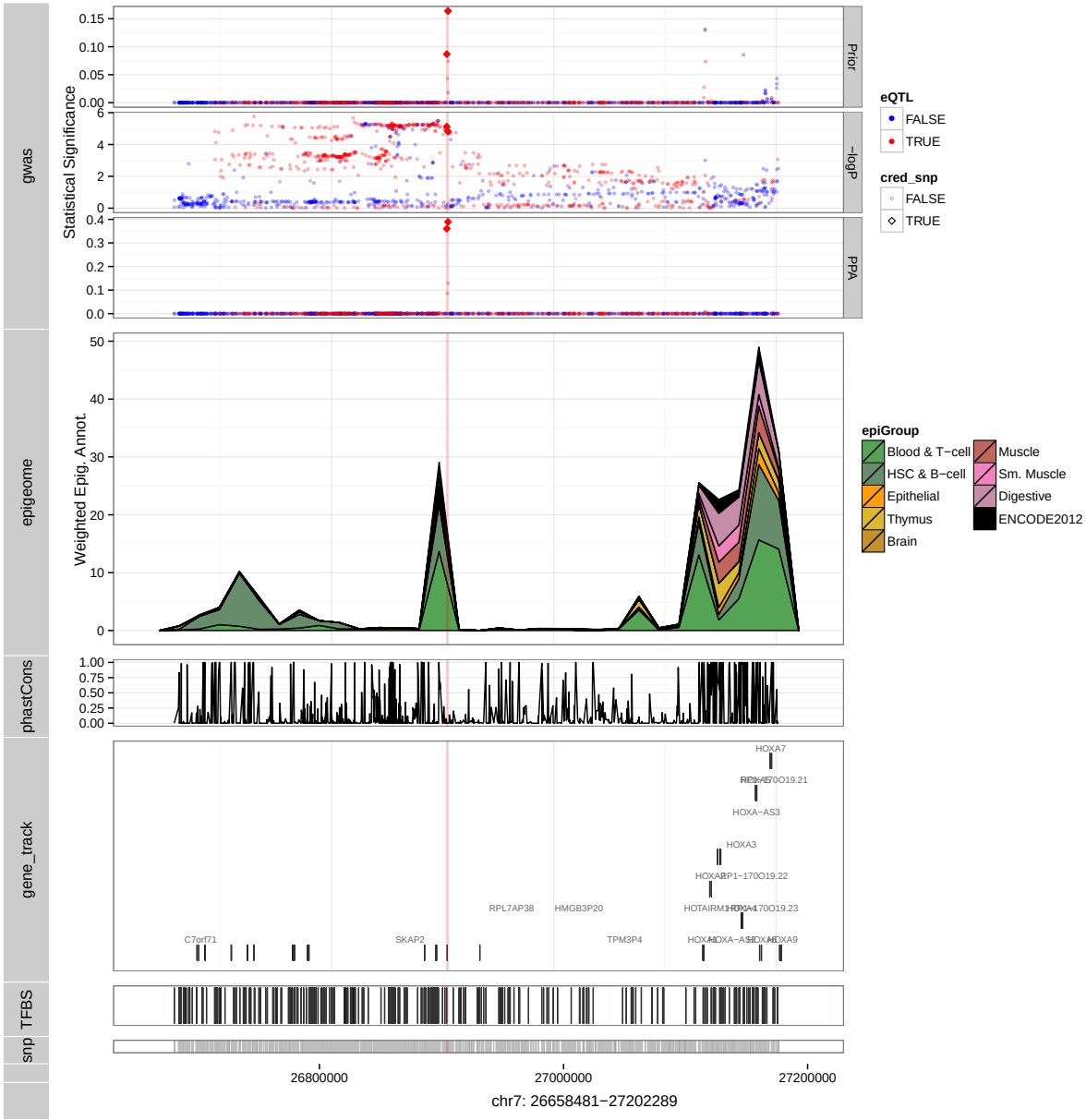
Type 1 Diabetes



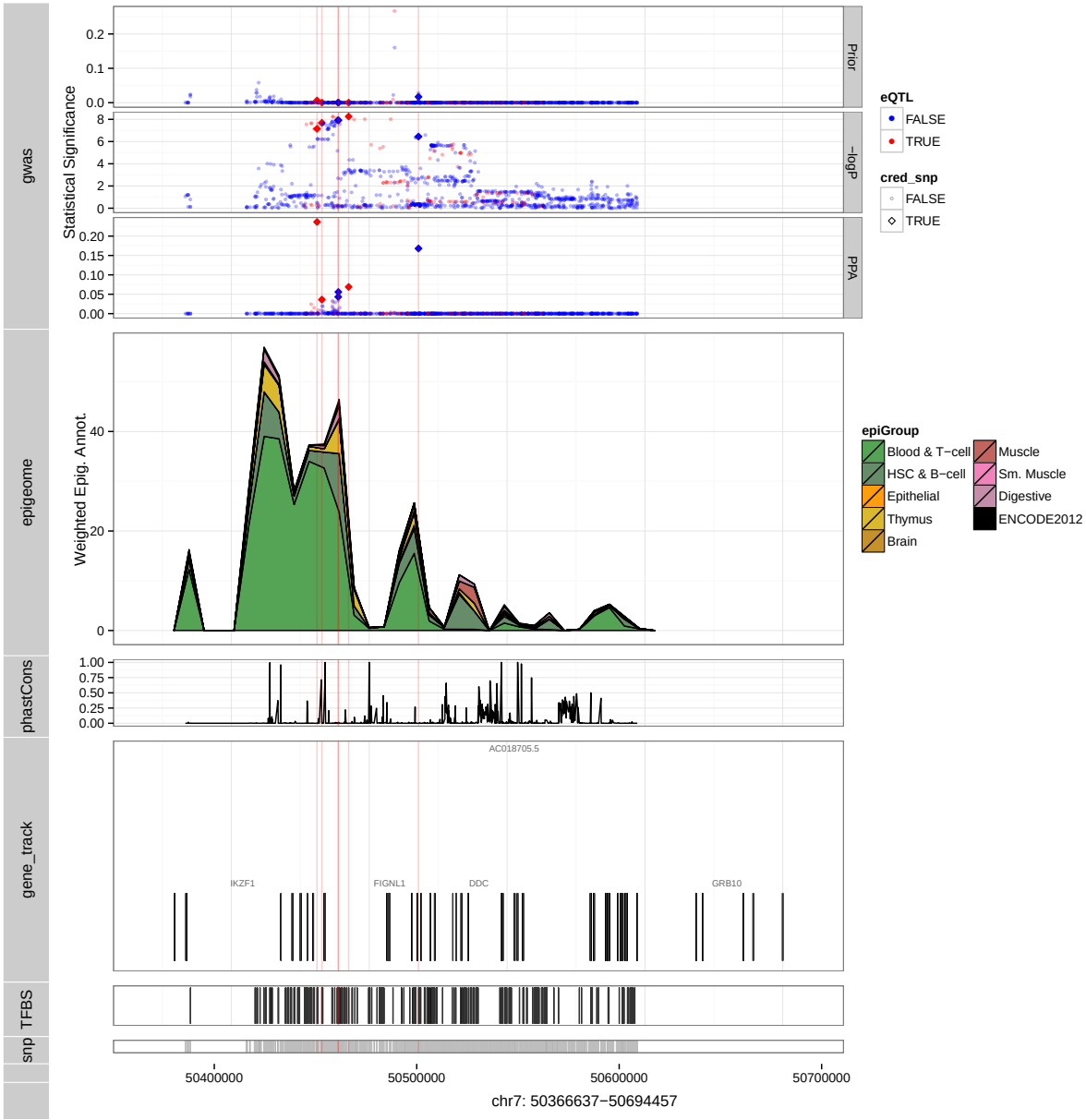
Type 1 Diabetes



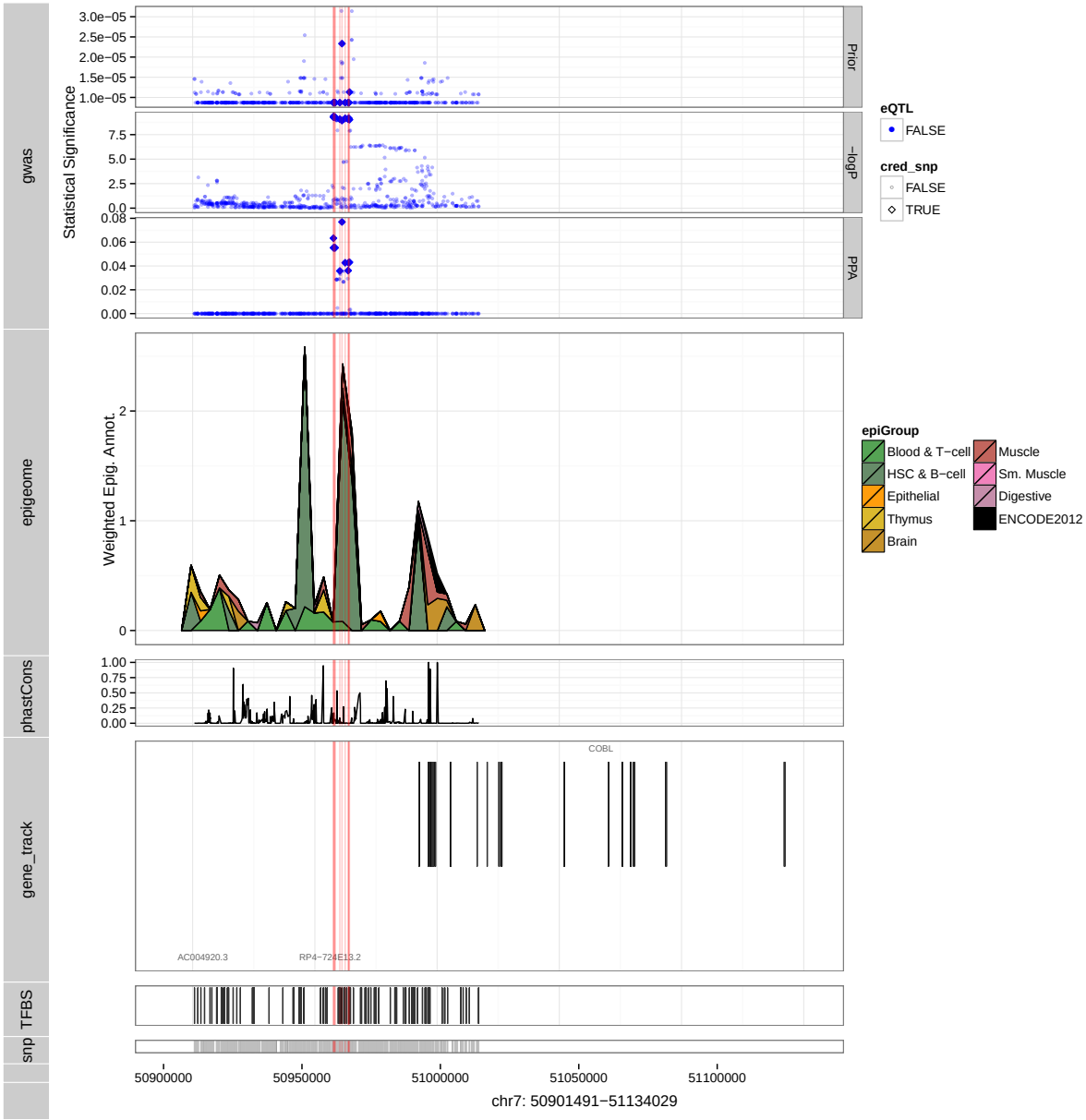
Type 1 Diabetes



Type 1 Diabetes



Type 1 Diabetes



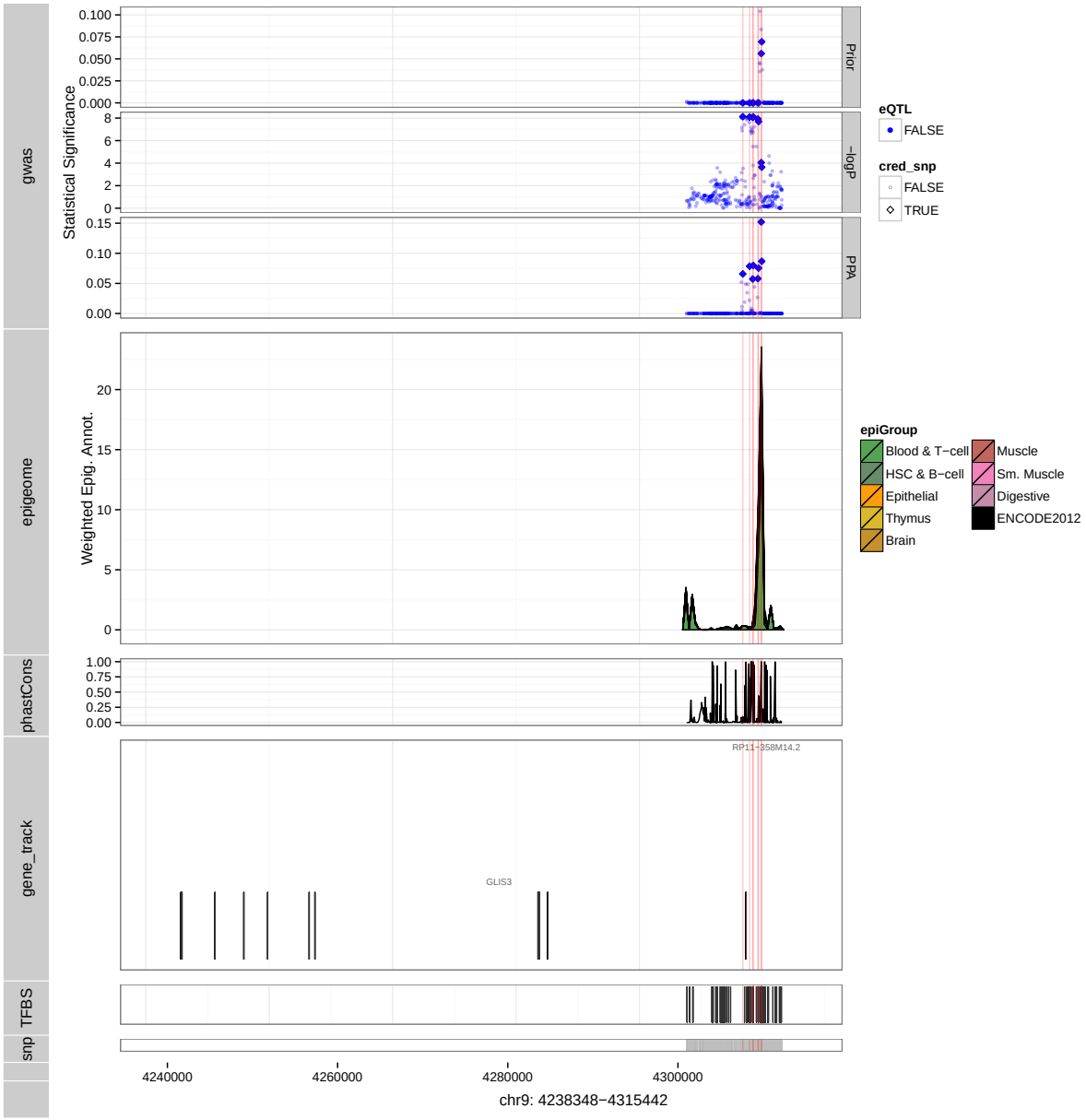


Fig. S5: Visualization of fine-mapping results on 469 risk loci across 10 traits. The upper panel display the model prior, the genetic signals of GWAS -log10 p-values, and PPA; the middle panel illustrates the cumulative density of weighted epigenomic profiles colored based on the epigenomic group; the bottom tracks shows the conservation, gene annotations (Gencode 19), transcription factor binding sites (TFBS), and SNP positions.

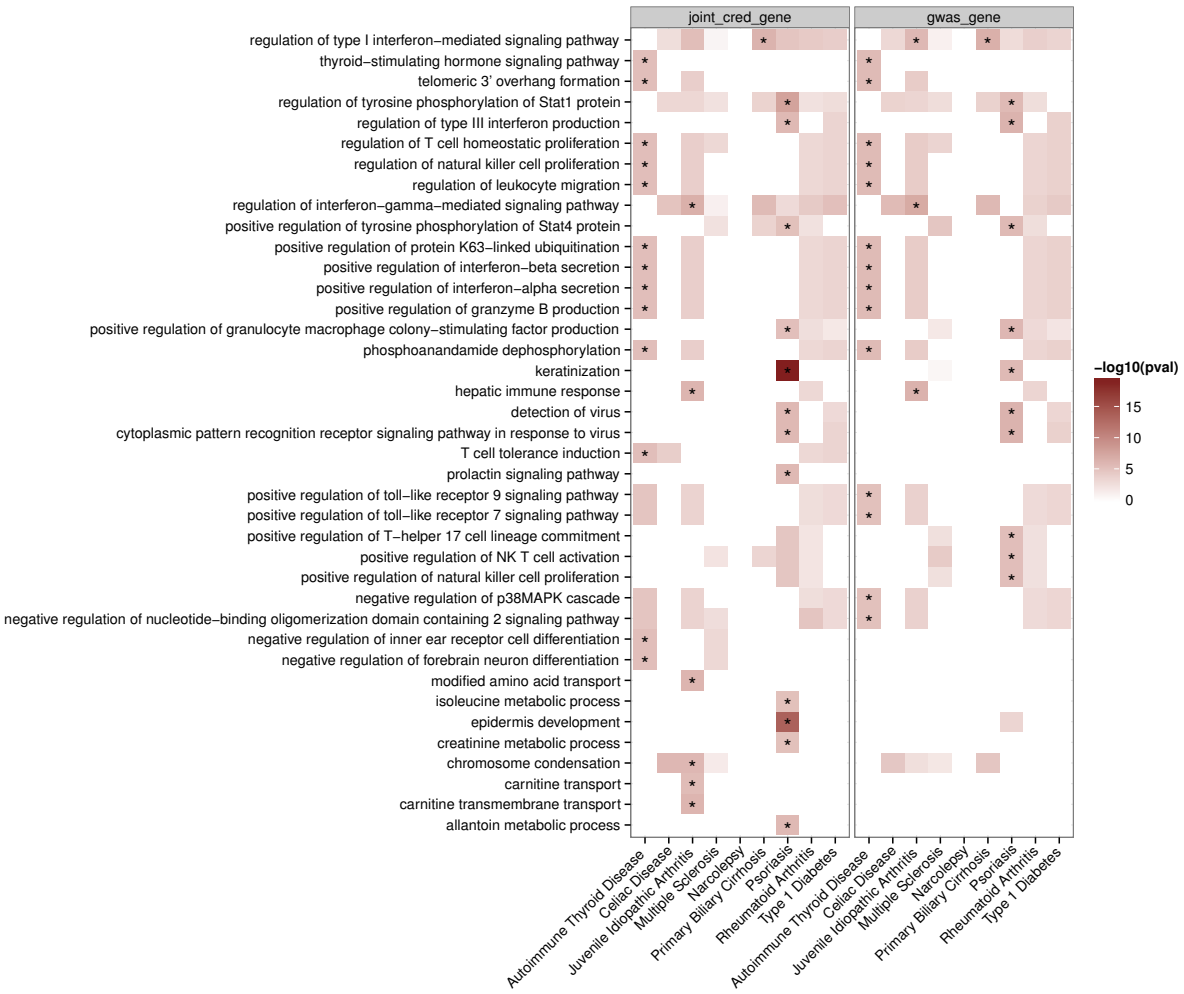


Fig. S6: Gene ontology enrichments across the 9 traits. Rows are the GO biological processes and columns are the 10 traits. Color intensities in each cell reflect the significance level in terms of -log10 p-value. Asterisks indicate q-values above significant cutoff after correcting for multiple testings (FDR < 0.005).

2 Supplementary Tables

Table S1: Full enrichment results for the 848 epigenomic annotations (Excel)

Table S2: Credible SNPs and related functional genomic information (Excel)

Table S3: Functional enrichments of credible SNPs (Excel)

Table S4: Gene ontology enrichment results using credible genes (Excel)

Table S5: Enrichments for eQTL of multi-trait model (Excel)

Table S6: Credible SNPs inferred from the multi-trait model on the 9 immune diseases and related functional genomic information (Excel)

Table S7: Gene ontology enrichment results using credible genes from cross-trait model (Excel)

3 Supplementary Text S1: HMC method details and RiVIERA algorithm outlines

3.1 Hamiltonian Monte Carlo (HMC) sampling for w_{kd}, w_{0d}, μ_d, ϕ_d

To sample epigenomic effects w_{kd}, causal mean μ_d, causal precision ϕ_d, we employ Hamiltonian Monte Carlo (also known as hybrid Monte Carlo) (HMC) [7,8]. HMC is one of the most powerful MCMC methods [7–9]. By extending the state space to include auxiliary momentum variables, and then using Hamiltonian dynamics to traverse long probability contours in the extended state space, HMC is able to move long distances in state space in a single update step and thus largely bypasses the slow exploration of the state space that occurs when Metropolis updates are done using a simple random-walk proposal distribution [9]. However, the essential elements for HMC are partial derivatives or gradients of the potential energy function $U = -\log p(\Theta|\mathcal{D})$ with respect to the variable of interests, which are not always tractable.

Sampling model parameters q by HMC can be intuitively visualized as a particle that slides over a frictionless surface [9]. When climbing on a rising slope, the particles momentum allows it to continue to explore the parameter space with decreasing kinetic energy and increasing potential energy (i.e., decreasing posterior), until the kinetic energy reaches zero, at which point it will slide back down with increasing kinetic energy and decreasing potential energy (i.e., increasing posterior). Thus, the momentum can be built up along the HMC trajectory to tolerate to some extent the deterioration of posterior probability upon hill climbing in order to seek for better parameter space. As a result, the HMC algorithm is able to explore much larger parameter space and much less prone to getting trapped within local suboptimal regions than simple Metropolis random walk.

In our case, solving HMC inference is feasible because $U = -\log p(\Theta|\mathcal{D})$ (i.e., joint posterior) is differentiable with respect to the model parameters $\mu_d, \phi_d, w_{kd}, w_{0d}$. With some simple linear algebra, we obtain the following gradients:

$$\frac{\partial U}{\partial \mu_d} = -\frac{\mu_0\phi_0 - 1}{\mu_d} + \frac{(1-\mu_0)\phi_0 - 1}{1-\mu_d} - \sum_{v=1}^{V} c_{vd} \frac{\partial \log f(a_{vd}; \mu_d, \phi_d)}{\partial \mu_d} \tag{1}$$

$$\frac{\partial U}{\partial \phi_d} = -\sum_{v=1}^{V} c_{vd} \left\{ [\mu(a^*_{vd} - \mu^*_d) + \log(1 - a_{vd}) - \psi((1-\mu_d)\phi_d)] + \psi(\phi_d) \right\} \tag{2}$$

$$\frac{\partial U}{\partial w_{kd}} = \sum_{d'=1}^{D} w_{kd'} \lambda_{dd'} + \sum_{v=1}^{V} (\pi_{vd} - c_{vd}) e_{vk} \tag{3}$$

$$\frac{\partial U}{\partial w_{0d}} = \sum_{v=1}^{V} (\pi_{vd} - c_{vd}) + \lambda_{0d}(w_{0d} - \log \frac{\pi_0}{1-\pi_0})) \tag{4}$$

where $\psi(.)$ is the digamma function and

$$\frac{\partial \log f(a_{vd}; \mu_d, \phi_d)}{\partial \mu_d} = \phi_d \left[\log \frac{a_{vd}}{1 - a_{vd}} - \{\psi(\mu_d\phi_d) - \psi((1-\mu_d)\phi_d)\} \right] \tag{5}$$

The gradient calculations were verified by comparing them individually to the changes in potential energy $\Delta U(q) = \frac{U(q+\xi)-U(q-\xi)}{2\times\epsilon}$, where $\xi = 10^{-4}$ and $\frac{\partial U}{\partial q} \approx \Delta U(q)$ within a very small numerical discrepancy (e.g., 10^{-8}).

Given the gradients, we apply the standard HMC procedure [8,9] known as the "leapfrog" method [8, 9]. In a nutshell, the Hamiltonian dynamics is defined as

$$H(q, p) = U(q) + K(p) \tag{6}$$

where the potential energy $U(q) = -\log p(\Theta|\mathcal{D})$ (i.e., negative log posterior defined as above) and q denotes the position variable in a fictitious physical system. In our case, the position variable $q \in \Theta = \{\mu_d, \phi_d, \mathbf{w}_d, w_{0d}\}$. $K(p)$ is the kinetic energy, where the auxiliary variable p is drawn independently from $\mathcal{N}(0, 1)$ to model the momentum for each position variable q.

As described in Algorithm 1, we simulate the Hamiltonian dynamics via the "leapfrog" method. The procedure starts with a half-step update of the momentum, followed by $L - 1$ full-step updates of both position and momentum variables, and ends with another half-step update of the momentum. During the leapfrog updates, we constrain $q \in \{\mu_d, \phi_d, \mathbf{w}_d\}$ to be positive by changing the sign of the momentum variable p and the position variable q once q becomes negative. This is equivalent to setting an infinite potential energy on the those values that violate the positive constraint, which will give the corresponding positions probability of zero. Intuitively, we can visualize it as reverting the trajectory of the particle as if it were to bounce off the wall given by the positive constraint [9].

At the end of this trajectory, we apply a Metropolis update to accept the proposed Hamiltonian state at the probability $\min(1, \exp[-\Delta H])$, where $\Delta H = H^* - H^{cur}$. Thus, we will always accept the proposed state if the Hamiltonian dynamics is lowered or unchanged in contrast to the current one. Because of the discrete-time procedure, we must use a small step size ϵ to compensate for the numerical imprecision introduced in simulating the Hamiltonian system. Automated procedure adapting the step size ϵ and step number L to the Metropolis acceptance rate is an active research field [9, 10]. A rule-of-thumb is to maintain the acceptance rate at around 65%. Here we set the step size ϵ to 0.001 and number of steps L to 100 based on some preliminary runs, where the acceptance rates are within the range of [50%, 90%] over 1000 MCMC samples.

3.2 RiVIERA inference algorithm

The pseudocode for RiVIERA is outlined in Algorithm 2.

Algorithm 1 HMC procedure to simultaneously sample $w_{kd}, w_{0d}, \mu_d, \phi_d$

Given ϵ, L, and q, where $q \in \{w_{kd}, w_{0d}, \mu_d, \phi_d\}$
Set $q^{cur} \leftarrow q^0 \leftarrow q$
Sample $p^{cur} \sim \mathcal{N}(0,1)$
Set $p \leftarrow p^{cur} - (\epsilon/2)\frac{\partial U}{\partial q}$
for $i = 1 \dots L$ **do**
    Set $q^i \leftarrow q^{i-1} + \epsilon p^{i-1}$
    if $i \neq L$ **then**
        Set $p^i \leftarrow p^{i-1} - \epsilon \frac{\partial U}{\partial q^i}$
    else
        Set $p^i \leftarrow p^{i-1} - (\epsilon/2)\frac{\partial U}{\partial q^i}$
    end if
    if $q \in \{(\mathbf{w}), \mu, \phi\}$ AND $q^i < 0$ **then**
        $q^i \leftarrow -q^i$, $p^i \leftarrow -p^i$
    end if
end for
Set $H^* \leftarrow U(q^L) + \frac{1}{2}\sum_m (p_m^L)^2$
Set $H^{cur} \leftarrow U(q^{cur}) + \frac{1}{2}\sum_m (p_m^{cur})^2$
if $Unif(0,1) < \exp(-\{H^* - H^{cur}\})$ **then**
    Set $q \leftarrow q^L$
else
    Set $q \leftarrow q^{cur}$
end if

Algorithm 2 RiVIERA algorithm

Input:

A: $V \times D$ GWAS summary statistics matrix as p-values for V variants across D diseases
E: $V \times K$ annotation matrix for V variants and K epigenomic annotations
L: $B \times 2$ matrix indicating start and end indices of the SNPs for each locus $b = 1, \dots, B$
Θ_0: π_0, μ_0, ϕ_0, ν_0, λ_0, α_0, β_0, T, ϵ, L (model settings or hyperparameters)

Output: :

$\hat{\pi}_v, \hat{c}_v, \hat{f}_k$: estimated expected values of priors, causal indicators, enrichments averaged across T' MCMC runs

Initialize model parameters $\mathbf{w} \sim \mathcal{N}(0,1), (\mu, \phi) \leftarrow (\mu_0, \phi_0), w_0 \leftarrow logit(\pi_0)$
for epoch = 1 To T **do**
    Gibbs sample $\Lambda_w | \mathbf{W} \sim Wishart((\Lambda_0^{-1} + \mathbf{S})^{-1}, \nu_0 + K)$
    Gibbs sample $\lambda_0 | w_0 \sim Gamma(\alpha_0 + 0.5, (\beta_0 + \frac{(w_{0d} - \mu_{w_0})^2}{2})^{-1})$
    Randomly sampled a locus b from $(1, \dots, B)$
    Compute empirical prior $\pi_{vd} \forall v \in \mathcal{V}_b; d = 1, \dots, D$
    Compute PPA $p(c_{vd} | a_{vd}, \pi_{vd}) \forall v \in \mathcal{V}_b; d = 1, \dots, D$
    HMC sample $w_{kd}, w_{0d}, \mu_d, \phi_d \forall k = 1, \dots, K; d = 1, \dots, D$
end for
Estimate $\hat{c}_{vd} = \frac{1}{T'}\sum_{t'=1}^{T'} c_{vd}^{(t')}$ and $\hat{f}_{kd} = \frac{1}{T'}\sum_{t=1}^{T'} \log \frac{1}{|\mathcal{C}_d|} \sum_{v \in \mathcal{C}_d} \frac{p(c_{vd} | \mathbf{w}_d^{(t)})}{p(c_{vd} | \mathbf{w}_d^{(t)}, w_{kd}^{(t)} = 0)}$

4 Supplementary Text S2: GWAS simulation

4.1 Simulating genotypes

As the first step of the simulation, we generated 10000 genotypes (**Fig. ??**). In particular, we use HapGen2 [1] to simulate haplotypes by conditioning on the 1000 Genome CEU population haplotypes (phase 1 version 3) so that the simulated data has the same LD patterns as the reference panel. However, HapGen2 is computationally demanding and can only simulate several thousands of SNPs for the chosen cohort size. To remedy that, we obtained the approximately independent LD blocks estimated from LDetect [2] and simulated 10,000 genotypes using HapGen2 together for each LD block separately. We only did this once for all of the simulation runs.

For each simulation run, we randomly sampled 100 LD blocks. For each sampled LD block, we then sampled 500 consecutive SNPs or used the whole block if it contains less or equal to 500 SNPs, which resulted in on average ∼38 kb per locus. Thus, the maximum number of SNPs generated per simulation run is 500,000 SNPs. The sampling step was done mainly for computational reason in order to run many simulation runs in parallel. We then filtered out SNPs with allele frequency lower than 1%.

4.2 Simulating epigenomic enrichments

We chose 20 representative cell types from the 19 primary groups defined by Epigenome Roadmap Consortium [3]. Whenever available, for each of the 20 chosen tissues or cell types, we obtained epigenomic profiles for DNase and six core histone marks namely H3K4me1, H3K4me3, H3K36me1, H3K27me3, H3K9me3, and H3K9ac (**Fig.** S1). In total, we obtained 100 epigenomic annotations in BigWig format. For each epigenomic annotation, we then thresholded their values to be 1 for p-values < 0.01 and 0 otherwise.

The fold-enrichment of annotation k is then:

$$f_k = \frac{m_k/m_c}{a_k/m} \tag{7}$$

where

- m_k: number of causal variants harbored in annotation k;
- m_c: the number of causal variants, which we set to 100, giving on average 1 causal variant per locus for the 100 simulated loci per simulated dataset;
- a_k: total number of variants from the 100 loci harbored in annotation k;
- m: total number of variants from the 100 loci;

The maximum fold-enrichment for annotation k is then:

$$f_{\max,k} = \begin{cases} m/a_k, & \text{if } a_k > m_c \\ m/m_c, & \text{if } a_k < m_c \end{cases} \tag{8}$$

We then sampled from uniform distribution the fold-enrichment for each annotation within the corresponding range defined by (8) and subsequently solved for m_k to sample causal variants from that annotation:

$$f_k \sim U(1, f_{max,k}) \tag{9}$$

$$m_k = \frac{f_k a_k m_c}{m} \tag{10}$$

Due to the previously observed enrichments [3,11], we started sampling causal variants from annotations corresponding to DNase, H3K4me1, and H3K4me3 and then the remaining annotations. The sampling continued until we reached or exceed the defined number of causal variants (i.e., 100). We then randomly chose 100 causal variants from the candidate causal variants. Because of the overlapping between annotations and the above rounding, we recomputed the fold-enrichment for each annotation using Eq (7).

4.3 Simulating GWAS summary statistics

We then simulated the effect sizes for causal SNPs as a function of minor allele frequency (p_i):

$$\beta_i = \frac{1}{\sqrt{2p_i(1-p_i)}} \tag{11}$$

We then simulated the quantitative phenotype or liability as:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon \tag{12}$$

$$\epsilon \sim \mathcal{N}(0, \sigma_e^2 I_n) \tag{13}$$

Here $\mathbf{X}$ is the standardized genotype of SNPs across individuals, and $\sigma_e^2 = \frac{\sigma_g^2 - h_g^2\sigma_g^2}{h_g^2}$ (fixed at 0.25), such that $h_g^2 = \frac{\sigma_g^2}{\sigma_e^2+\sigma_g^2}$ and $\sigma_g^2 = \mathrm{Var}(\mathrm{X}\beta) = \beta'\Sigma_g\beta$ is the phenotypic variance explained by the genetics [11], where Σ_g is the genotypic correlation matrix (i.e., linkage disequilibrium) estimated as the Pearson correlation from the 10,000 simulated individuals. Finally, we simulated p-value and z-score (as t-statistics) by regressing phenotype $\mathbf{y}$ on $\mathbf{X}$ for each SNP i: ($\mathbf{y} \sim \mathbf{x}_i$). We repeatedly generated simulated datasets using the approach described above and evaluated each fine-mapping method on each of the simulated dataset.

References

1. Su, Z., Marchini, J. & Donnelly, P. HAPGEN2: simulation of multiple disease SNPs. *Bioinformatics (Oxford, England)* **27**, 2304–2305 (2011).

2. Berisa, T. & Pickrell, J. K. Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics (Oxford, England)* (2015).

3. Consortium, R. E. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

4. Pickrell, J. K. Joint Analysis of Functional Genomic Dataand Genome-wide Association Studies of 18 Human Traits. *The American Journal of Human Genetics* **94**, 559–573 (2014).

5. Kichaev, G. *et al.* Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS genetics* **10**, e1004722 (2014).

6. Chung, D., Yang, C., Li, C., Gelernter, J. & Zhao, H. GPA: A Statistical Approach to Prioritizing GWAS Results by Integrating Pleiotropy and Annotation. *PLoS Genetics* **10**, e1004787 (2014).

7. Duane, S., Kennedy, A. D., Pendleton, B. J. & Roweth, D. Hybrid monte carlo. *Physics letters B* **195**, 216–222 (1987).

8. Neal, R. M. *Probabilistic inference using Markov chain Monte Carlo methods*. Ph.D. thesis, University of Toronto (1993).

9. Neal, R. M. Mcmc using hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo* **2** (2011).

10. Homan, M. D. & Gelman, A. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *The Journal of Machine Learning Research* **15**, 1593–1623 (2014).

11. Gusev, A. *et al.* Partitioning Heritability of Regulatory and Cell-Type-Specific Variants across 11 Common Diseases. *AJHG* **95**, 535–552 (2014).

12. Burton, P. R. *et al.* Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).