The American Journal of Human Genetics, Volume *101*

Supplemental Data

A Dementia-Associated Risk Variant near *TMEM106B*

Alters Chromatin Architecture and Gene Expression

Michael D. Gallagher, Marijan Posavi, Peng Huang, Travis L. Unger, Yosef Berlyand, Analise L. Gruenewald, Alessandra Chesi, Elisabetta Manduchi, Andrew D. Wells, Struan F.A. Grant, Gerd A. Blobel, Christopher D. Brown, and Alice S. Chen-Plotkin

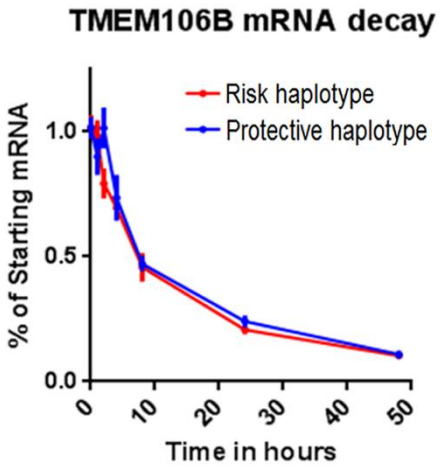
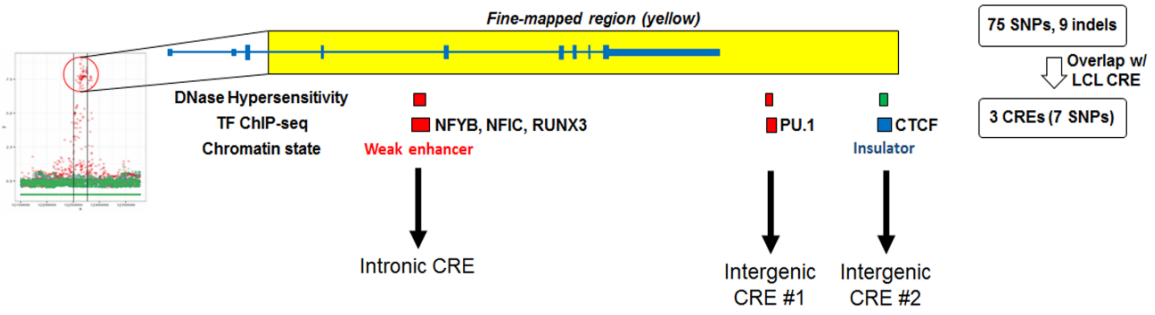


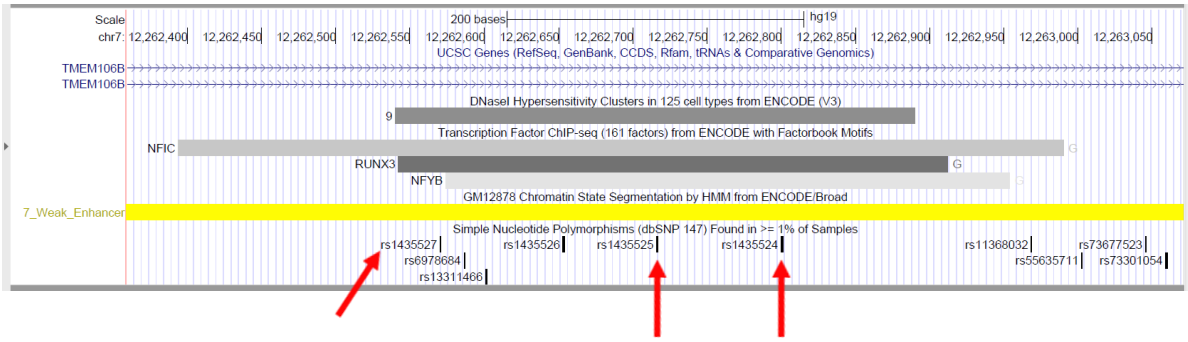
Figure S1. *TMEM106B* genotype does not affect *TMEM106B* mRNA stability.
Lymphoblastoid cell lines homozygous for the risk (n=3) and protective (n=3) *TMEM106B* haplotypes were treated with Actinomycin D to inhibit transcription. RNA was isolated at 0h, 1h, 2h, 4h, 8h, and 24h after treatment, and *TMEM106B* was quantified by RT-qPCR. Data were analysed with a two-way ANOVA.

A



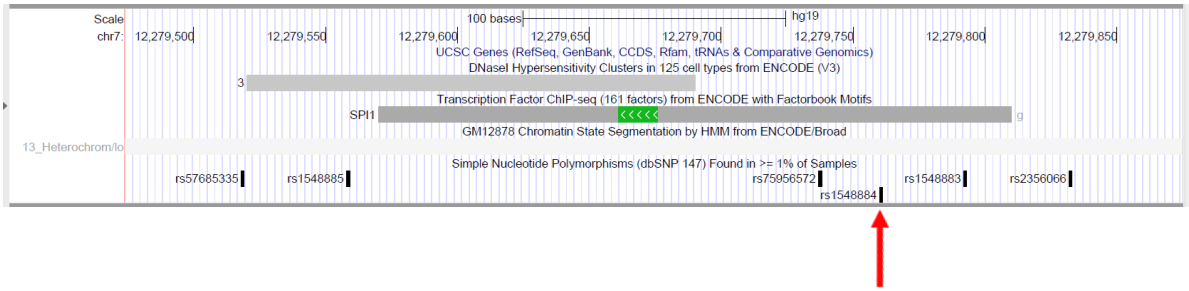
B

Intronic CRE



C

Intergenic CRE #1



D

Intergenic CRE #2

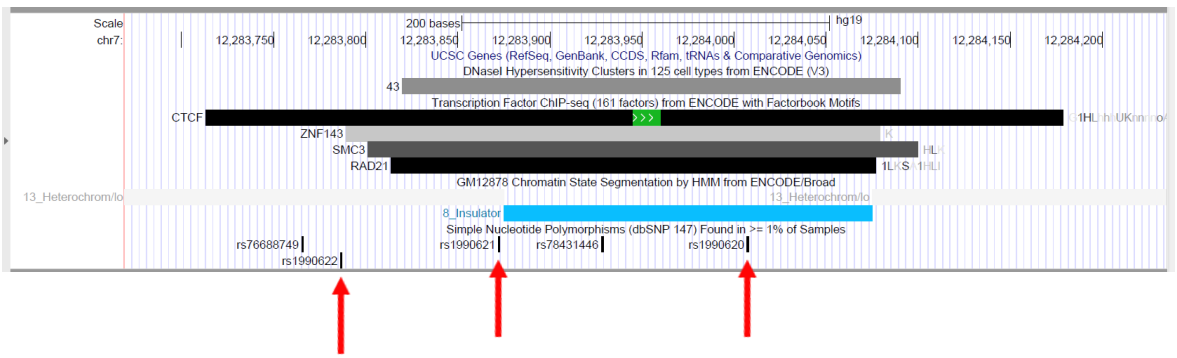


Figure S2. Seven candidate causal variants in three candidate *cis*-regulatory regions (CREs).

(A) Epigenomic prioritization of the 84 fine-mapped candidate causal variants identifies seven candidate causal variants in three candidate CREs.

For panels B-D, UCSC Genome Browser snapshots displaying ENCODE DNase hypersensitivity (DHS), transcription factor (TF) ChIP-seq, and the GM12878 LCL chromatin state segmentation tracks are shown. DHS was performed on 125 cell lines, and the number directly to the left of the peak indicates the number of cell types that displays DHS at that region. TF ChIP-seq was performed for 161 TFs on varying numbers of cell lines, and the letters directly to the right of a peak indicate the cell type abbreviations for which the peak was detected ("g" or "G" indicates LCLs).

(B) Three variants are located in an intronic CRE that displays DHS, binding of the TFs NFIC, RUNX3, and NFYB, and a "weak enhancer" chromatin state in LCLs.

(C) One variant overlaps an intergenic region of SPI1 (PU.1) binding in LCLs (SPI1 motif indicated in green, with arrows indicating directionality of motif). DHS is not observed in LCLs, but is observed in three other white blood cell types. This region has an assigned heterochromatic/low signal chromatin state in LCLs, which is characterized by the lack of histone marks and lack of transcription (Ernst et al., 2011).

(D) Three variants overlap a ubiquitous intergenic CTCF binding region (CTCF motif indicated in green, with arrows indicating directionality of motif). DHS and binding of the TF ZNF143 and the cohesin subunits SMC3 and RAD21 are seen in non-LCL and non-brain-relevant cell types. This region has an assigned "insulator" chromatin state in LCLs, characterized by the lack of histone marks in the presence of CTCF binding (Ernst et al., 2011, Nature), and is flanked by regions of heterochromatin/low signal chromatin states.

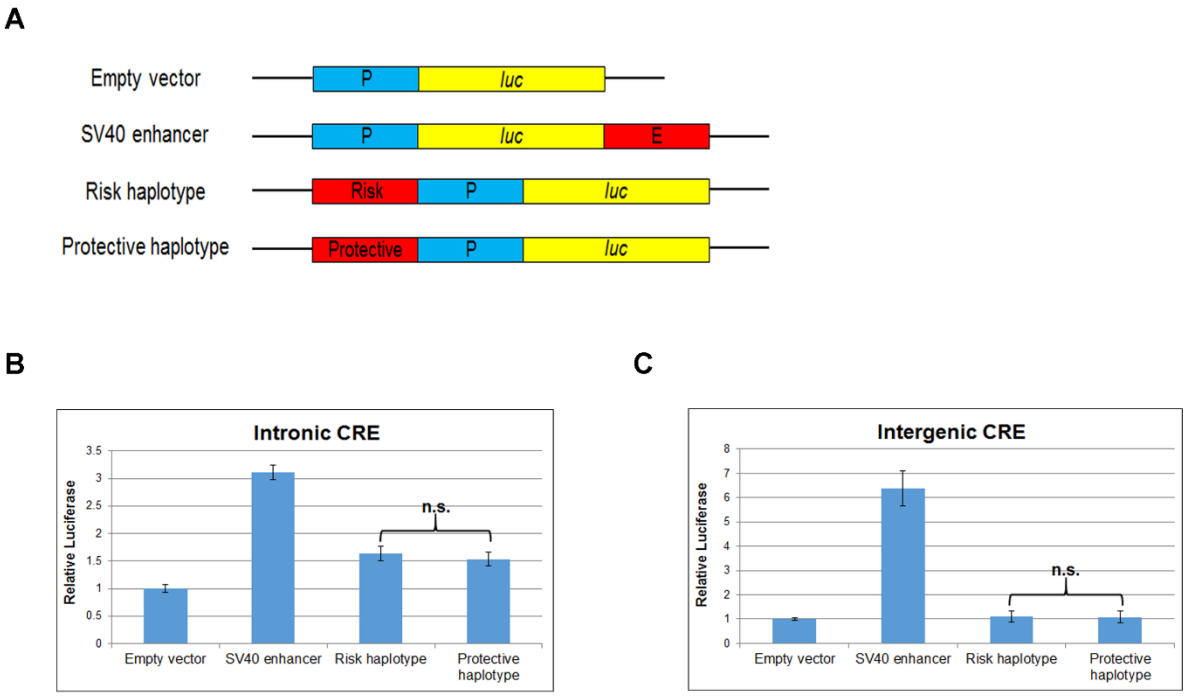


Figure S3. Intronic CRE and intergenic CRE #1 do not harbour functional variants.
(A) Schematic of luciferase reporter constructs used to test candidate CREs for enhancer activity. Empty vector construct contains no enhancer (negative control); SV40 enhancer construct serves as a positive control; test constructs contain putative CREs with either the risk or protective SNP alleles cloned upstream of the SV40 promoter. P=SV40 promoter; *luc*=luciferase; E=SV40 enhancer.
(B,C) The intronic CRE displays weak ~1.5-fold enhancer activity with both SNP haplotypes (P<0.001), whereas neither SNP haplotype of the intergenic CRE displays any activity. n.s.=non-significant.

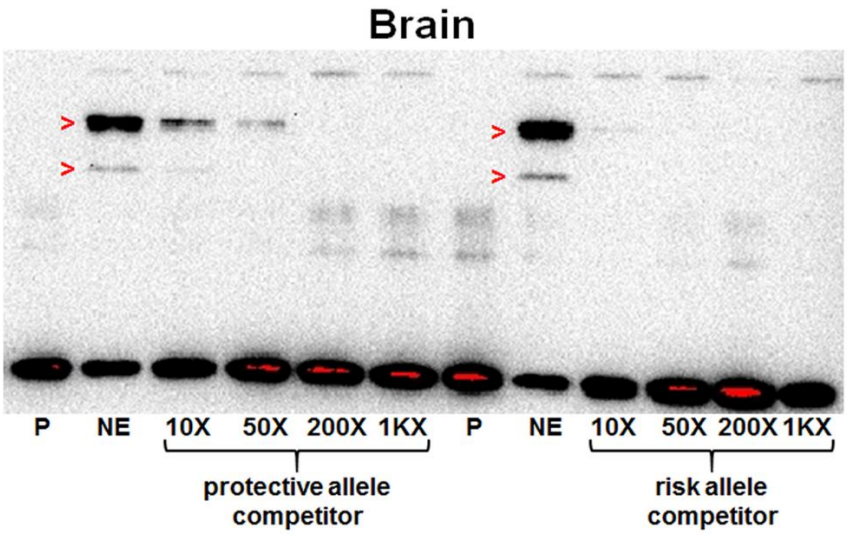


Figure S4. The risk allele of rs1990620 preferentially recruits a nuclear factor in brain nuclear extract.

A 5' biotinylated probe containing the rs1990620 protective allele was incubated with human brain nuclear extract, resulting in a similar shift (red arrowheads) to that seen with the rs1990620 risk allele (**Figure 5D**). The shifted complex was better competed with excess amounts of unlabelled oligo containing the risk allele of rs1990620 than that containing the protective allele.

P=probe, NE=probe+nuclear extract, 1KX=1,000X.

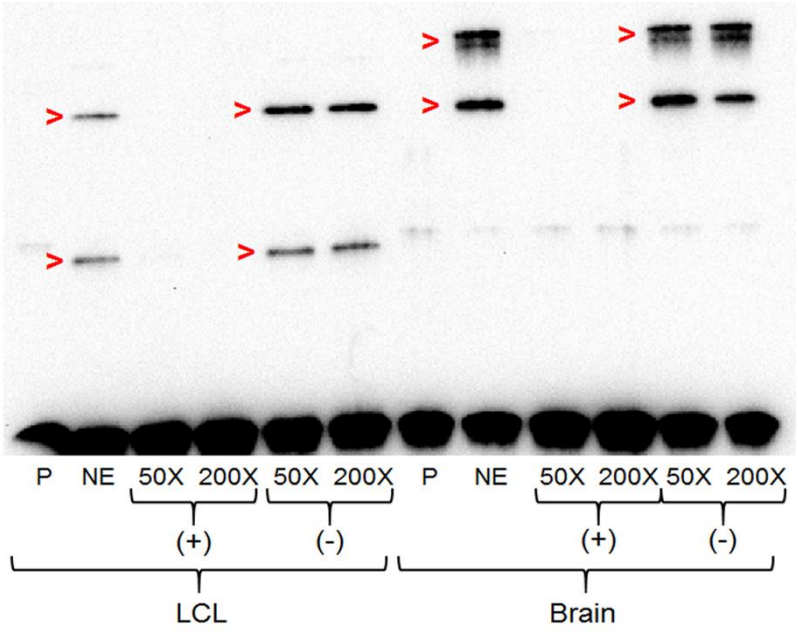


Figure S5. The rs1990620 probe is specifically competed by unlabelled oligonucleotides containing the probe sequence.
The probe containing the risk allele of rs1990620 was incubated with nuclear extract from both LCLs (left) and brain (right). Shifts are indicated by the red arrowheads. Both 50X and 200X excess unlabelled competitor oligo containing the probe sequence (indicated with a (+), representing positive control) efficiently outcompeted the probe for its bound nuclear factor(s), whereas the same concentrations of a negative control (-) competitor did not outcompete the probe.

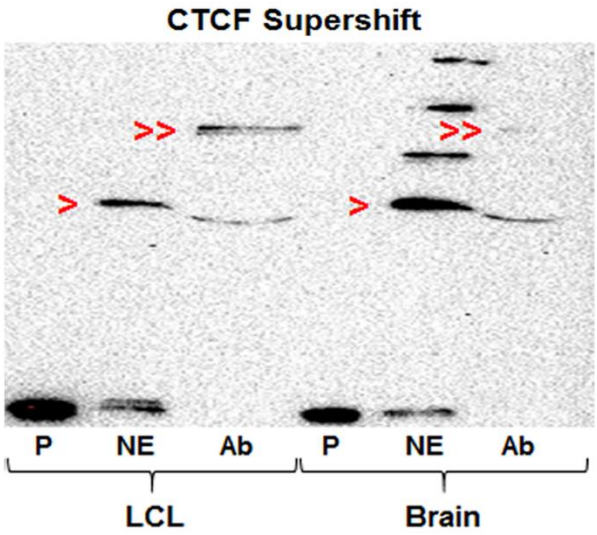


Figure S6. CTCF binds a 3' biotinylated rs1990620 probe.
A 3' biotinylated rs1990620 risk allele probe is bound by one or more nuclear factors in LCL and brain nuclear extract (red arrowheads), and this probe/protein complex can be supershifted by addition of an anti-CTCF antibody (Ab, double red arrowheads).

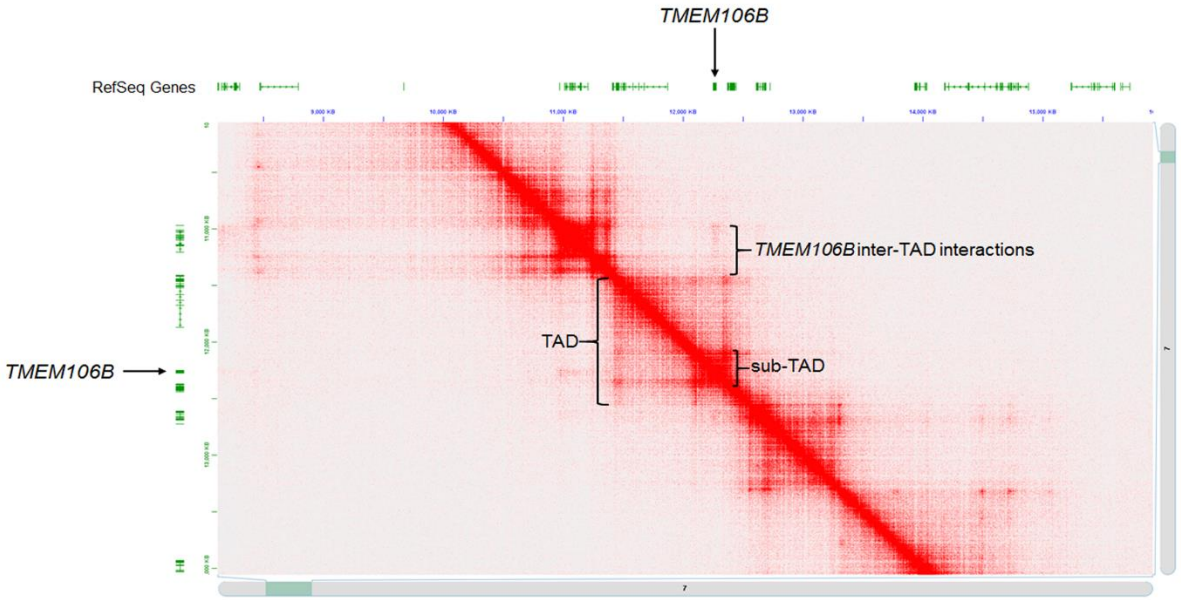
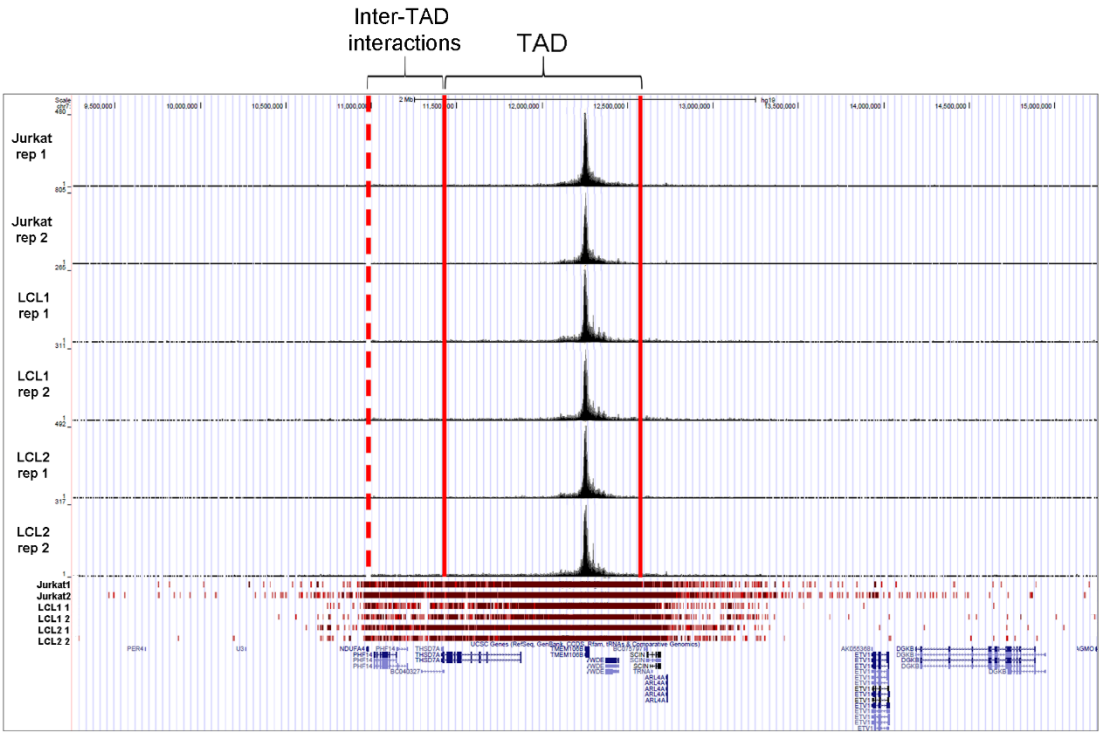


Figure S7. Lymphoblastoid cell line *in situ* Hi-C data reveals the chromatin architecture at the *TMEM106B* locus.

Hi-C heatmap for chromosome 7p21 displays a ~250kb topologically associating domain (sub-TAD) within a larger ~1Mb TAD, containing *TMEM106B* and no other genes. *TMEM106B* also interacts with genomic regions extending several hundred kilobases upstream of the TAD (inter-TAD interactions).

A



B

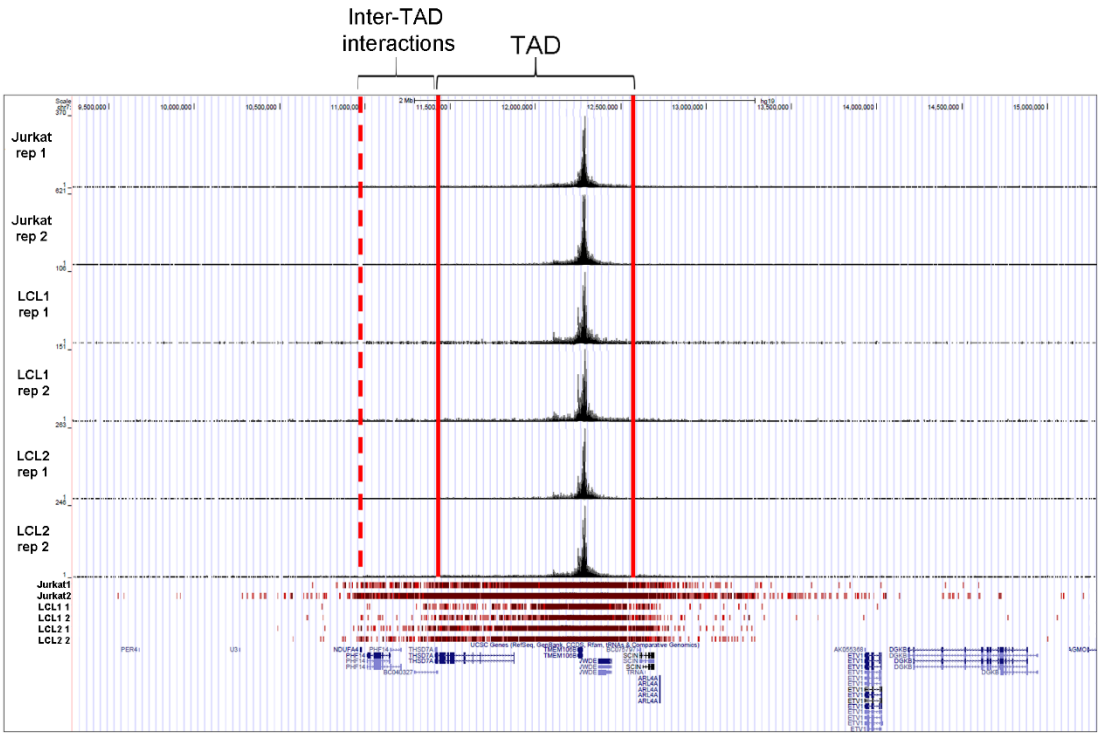


Figure S8. Capture-C recapitulates the predicted chromatin architecture at the *TMEM106B* locus.

(A) When analysing all long-range interactions mapping to chromosome 7 from the *TMEM106B* promoter viewpoint, statistically significant interactions are limited mostly to coordinates agreeing with the Hi-C TAD (solid red lines) and upstream inter-TAD interactions (dashed red line indicates endpoint of interactions observed in the Hi-C data).

(B) Similar results are seen from the CTCF site viewpoint.
The top of each panel shows raw read coverage for each sample and replicate, and the bottom of each panel shows statistically significant interactions (red bars indicate statistical significance, with darker shades of red indicating higher confidence interactions). Data are viewed in the UCSC Genome Browser with GRCh37/hg19 coordinates, with read counts on the y-axis of each track.

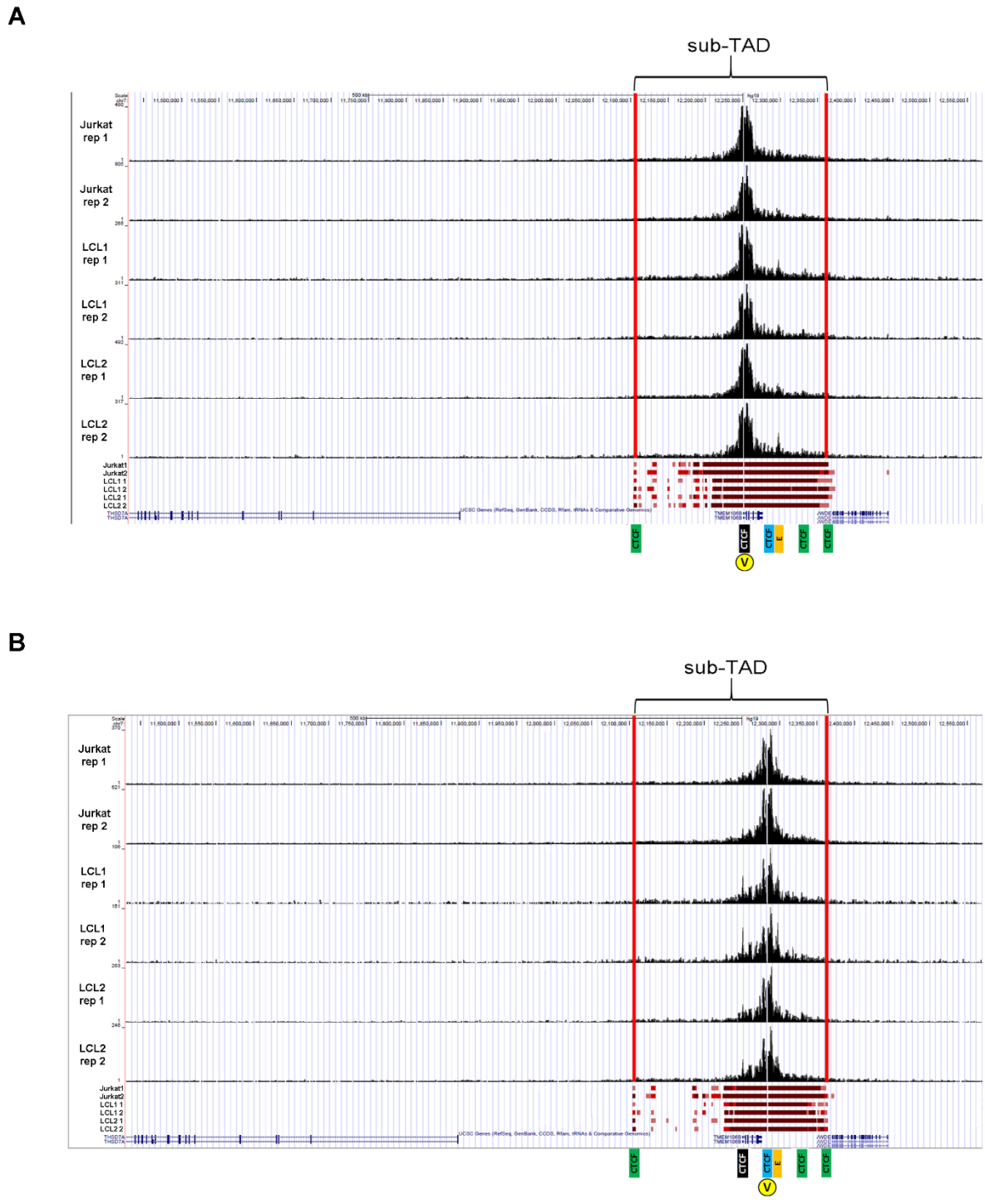


Figure S9. Capture-C identifies key interactions between *cis*-regulatory elements at the *TMEM106B* locus.

(A) When analyzing only the interactions mapping to the TAD from the *TMEM106B* promoter viewpoint, statistically significant interactions map entirely to the Hi-C sub-TAD coordinates (solid red lines).

(B) Similar results are seen from the CTCF site viewpoint.

Key interacting regions from **Figure 6A** are indicated below each browser snapshot. Yellow circles marked with a "V" indicate the Capture-C viewpoint. Data are viewed in the UCSC Genome Browser with GRCh37/hg19 coordinates, with read counts on the y-axis of each track.

Putative CRE	rsid	hg19	DNase HS		TF binding		Chromatin state		LD with rs1990622 (r^2)
			LCL	Brain	LCL	Brain	LCL	Brain	
intronic CRE	rs1435527	12262571	Yes	No	NFIC, RUNX3, NFYB	No	Weak enhancer	None	0.96
intronic CRE	rs1435525	12262717	Yes	No	NFIC, RUNX3, NFYB	No	Weak enhancer	None	0.96
intronic CRE	rs1435524	12262801	Yes	No	NFIC, RUNX3, NFYB	No	Weak enhancer	None	0.96
intergenic CRE	rs1548884	12279761	No	No	PU.1	No	Low signal	None	0.98
CTCF site	rs1990622	12283787	No	Yes	CTCF	CTCF	Insulator	None*	N/A
CTCF site	rs1990621	12283873	No	Yes	CTCF	CTCF	Insulator	None*	0.99
CTCF site	rs1990620	12284008	No	Yes	CTCF	CTCF	Insulator	None*	1

Table S1. Functional annotations of tested top eQTL SNPs variants.
The 7 SNPs overlapping either DNase hypersensitivity (HS), transcription factor (TF) binding, or an active chromatin state in lymphoblastoid cell lines (LCLs) are listed by location. DNase HS and TF binding in brain refers to experiments performed in brain-relevant ENCODE cell lines, whereas chromatin state in brain refers to Roadmap EpiGenome brain tissue data. CTCF binding at the region containing the GWAS sentinel SNP, rs1990622, and two other linked SNPs, is seen in brain-relevant cell types as well as LCLs. *Since the Roadmap EpiGenome chromatin state annotations are based only on histone marks, and not TF binding, the CTCF site has no annotation in Roadmap brain samples.

Experiment	Cell lines
CTCF ChIP-seq	BE2_C, Caco-2, GM06990, GM12864, GM12872, GM12873, GM12874, GM12891, GM12892, GM19238, H1-hESC, HBMEC, HCFA, HEK293, HepG2, HRPEpiC, HUVEC, HVMF, NHDF, RPTEC
DNase digital genomic footprinting	AoAf, HCFaa, NHDF-neo, HMVEC-dBl-Ad, H7-hESC, HMVEC-dLy-Neo

Table S3. Cell lines used for CTCF ChIP-seq and DNase digital genomic footprinting analyses.

All cell lines were confirmed to be heterozygous at rs1990620 by analysing reads containing rs1990620, as well as rs1990621 and rs1990622, if such reads existed. Risk and protective allele rs1990620 reads were enumerated and pooled across samples for each experiment. Deviation from an expected ratio of 50:50 was tested using a two-tailed binomial sign test. Cell lines are listed by their ENCODE identifiers, or spelled out if no identifiers exist.

SNP	Risk allele reads	Protective allele reads	Total reads	Enrichment of risk allele	*P*-value
rs1990622	40	26	66	53.8%	0.110
rs1990621	114	90	204	26.7%	0.107
rs1990620	207	167	374	24.0%	*0.0434*
TOTAL	361	283	644	27.6%	*0.00245*

Table S4. All three SNPs at the CTCF binding region show a pattern of increased CTCF binding on the risk haplotype.

The number of CTCF ChIP-seq reads from 20 heterozygous cell lines is shown for rs1990622, rs1990621 and rs1990620. All three SNPs are covered by more risk allele than protective allele reads, and this effect is statistically significant for rs1990620 alone and when totalling all three SNPs, by using a 2-tailed binomial sign test (italicized *P*-values). Since the ChIP-seq samples were sequenced with 36bp single end reads (Gerstein et al., 2015, Nature), which is less than the minimum distance between SNPs (86bp, see **Table S1**), reads can be summed across SNPs without double counting.

Probe	Sequence
Promoter probe 1	GAGTTGCTGTGTCGCCTCTAATGAGGCCCAGCCAGGGAACAC TCGGCTTCGGCCCAAGCC
Promoter probe 2	ACCTCTGGGTTGCCTCTAGGCCCTCACACCTTGAGCGCCAGGT GGCCCTCTTCCTTTTGC
Promoter probe 3	CTAGGCCCTCACACCTTGAGCGCCAGGTGGCCCTCTTCCTTTT GCTGTTGATGAATGTTC
Promoter probe 4	TTGAGCGCCAGGTGGCCCTCTTCCTTTTGCTGTTGATGAATGT TCTTGCCGTGGTGCCGG
CTCF site probe 1	CAATGGAGAACAGAGCTGCAGTACAACTATAGCATTTTGAGG GGCTGGGAGAACGGGATA
CTCF site probe 2	TACAGTACTACAATGGAGAACAGAGCTGCAGTACAACTATAG CATTTTGAGGGGCTGGGA
CTCF site probe 3	ATGAAGAGTCTAAGCAGAACTTTCCAGTCATTGTACTACAAT GGGTTGCAGGCGTGCCAG
CTCF site probe 4	TTTCCAGTCATTGTACTACAATGGGTTGCAGGCGTGCCAGCCC TCTGGTGGCCACATAAG
CTCF site probe 5	ATGGGTTGCAGGCGTGCCAGCCCTCTGGTGGCCACATAAGTC CTGGAGTGGCTGGTCATC
CTCF site probe 6	GCCCTCTGGTGGCCACATAAGTCCTGGAGTGGCTGGTCATCT CAGACTGGGAGCTTCCTC

Table S5. Probe sequences used for Capture-C.
Four and six 60bp biotinylated probes were used to capture interactions involving the *TMEM106B* promoter and rs1990620-containing CTCF site, respectively.

Sample	Test region	Risk allele reads	Protective allele reads	Total reads	Enrichment of risk allele	*P*-value
Jurkat	Chr7	28644	26917	55561	6.4%	*<1.0e-06*
LCL #1	Chr7	13189	12511	25700	5.4%	*2.5e-05*
LCL #2	Chr7	19309	17992	37301	7.3%	*<1.0e-06*
Jurkat	TAD	13730	12252	25982	12.1%	*<1.0e-06*
LCL #1	TAD	6098	5815	11913	4.9%	*9.88e-03*
LCL #2	TAD	9305	8606	17911	8.1%	*<1.0e-06*
Jurkat	sub-TAD	10184	9641	19825	5.6%	*1.2e-04*
LCL #1	sub-TAD	4652	4398	9050	5.8%	*7.8e-03*
LCL #2	sub-TAD	6812	6612	13424	3.0%	0.085

Table S6. Read count totals for long-range interactions containing the *TMEM106B* promoter SNP rs4721056.

Data show the raw read counts covering rs4721056, which was used a haplotype marker SNP in the Capture-C analyses. Data are organized by cell line and test region (i.e., which genomic region was analysed for long-range interactions). For all three cell lines and test regions, rs4721056 risk allele reads outnumbered protective allele reads, and these differences reached statistical significance in all but one sample (italicized *P*-values).