

## Supplementary Data

**Supplementary Table 1. Numbers of cases, controls, and trios by recruitment site.**

Continental Group	Site	Cleft Lip with or without Cleft Palate		
		Controls	Cases	Trios
<b>European</b>	Denmark	0	18	28
	Hungary	253	23	82
	Spain	0	2	31
	Turkey	171	57	115
	United States	411	70	150
	<b>TOTAL</b>	<b>835</b>	<b>170</b>	<b>406</b>
<b>Central/South American</b>	Argentina	30	68	43
	Colombia	277	276	405
	Guatemala	208	47	55
	Puerto Rico	106	33	51
	United States	5	25	47
	<b>TOTAL</b>	<b>626</b>	<b>449</b>	<b>601</b>
<b>Asian</b>	China	27	32	125
	India	38	19	32
	Philippines	96	32	127
	<b>TOTAL</b>	<b>161</b>	<b>83</b>	<b>284</b>
<b>African</b>	Ethiopia	0	80	3
	Nigeria	68	34	16
	United States	5	7	7
	Hungary	1	0	0
	<b>TOTAL</b>	<b>74</b>	<b>121</b>	<b>26</b>
<b>No subgroup (included in multiethnic analysis)</b>		<b>4</b>		<b>2</b>
<b>GRAND TOTAL</b>		<b>1700</b>	<b>823</b>	<b>1319</b>

**Supplementary Table 2.** Summary of the SNP filters recommended by the GCC. The number and percentage of SNPs removed is indicated for each filter individually. The number and percentage of SNPs retained is shown for the filters cumulatively applied in sequential order down the rows of the table.

Filter	SNPs filtered		Cumulative SNPs Retained	
	number	percent	number	percent
none (all SNP probes)			557,677	100.00%
CIDR technical filters	9,625	1.73%	548,052	98.27%
missing call rate greater than 2%	5,891	1.06%	542,161	97.22%
2 or more discordant call in 264 study duplicates	42	0.01%	542,119	97.22%
20 or more Mendelian errors in 5,288 parent-offspring trios or dyads	2,410	0.43%	539,709	96.78%
HWE p-value < 0.0001 in participants of genetically confirmed European ancestry	194	0.03%	539,515	96.74%
sex differences in allele frequency of 0.2 or greater for autosomes or XY pseudoautosomal region	42	0.01%	539,473	96.74%
sex differences in heterozygosity of 0.3 or greater for autosomes or XY pseudoautosomal region	0	0.00%	539,473	96.74%
positional duplicates	7,278	1.31%	532,195	95.43%
MAF = 0	76,746	13.76%	455,449	81.67%
MAF < 0.01	161,816	29.02%	293,633	52.65%

**Supplementary Table 3. Population-based case-control cohorts used for replication**

	<b>CL</b>	<b>CLP</b>	<b>Controls</b>	<b>TOTAL</b>
<b>United States</b>				
<i>Iowa Case-Control Study</i>	25	45	146	216
<i>Utah Child and Family Health Study</i>	42	59	220	321
<b>Norway</b>				
Norwegian Mother and Child Cohort	17	57	354	428
Norway Facial Clefts Study	117	178	455	750
<b>Denmark</b> ( <i>National Birth Cohort</i> )	31	36	510	577
<b>TOTAL</b>	<b>232</b>	<b>375</b>	<b>1685</b>	<b>2292</b>

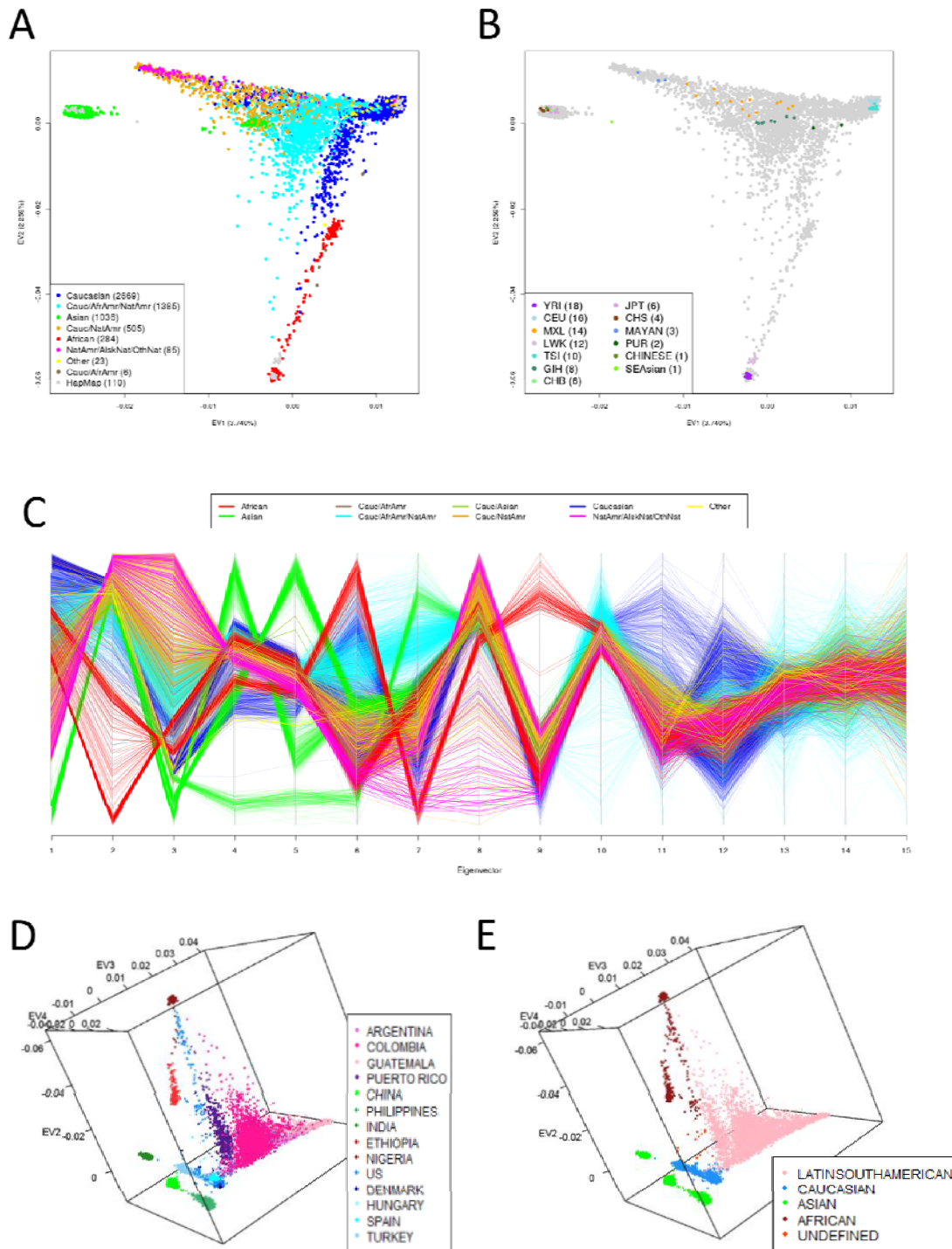


**Supplementary Table 8. Association results for previous reported CL/P GWAS loci**

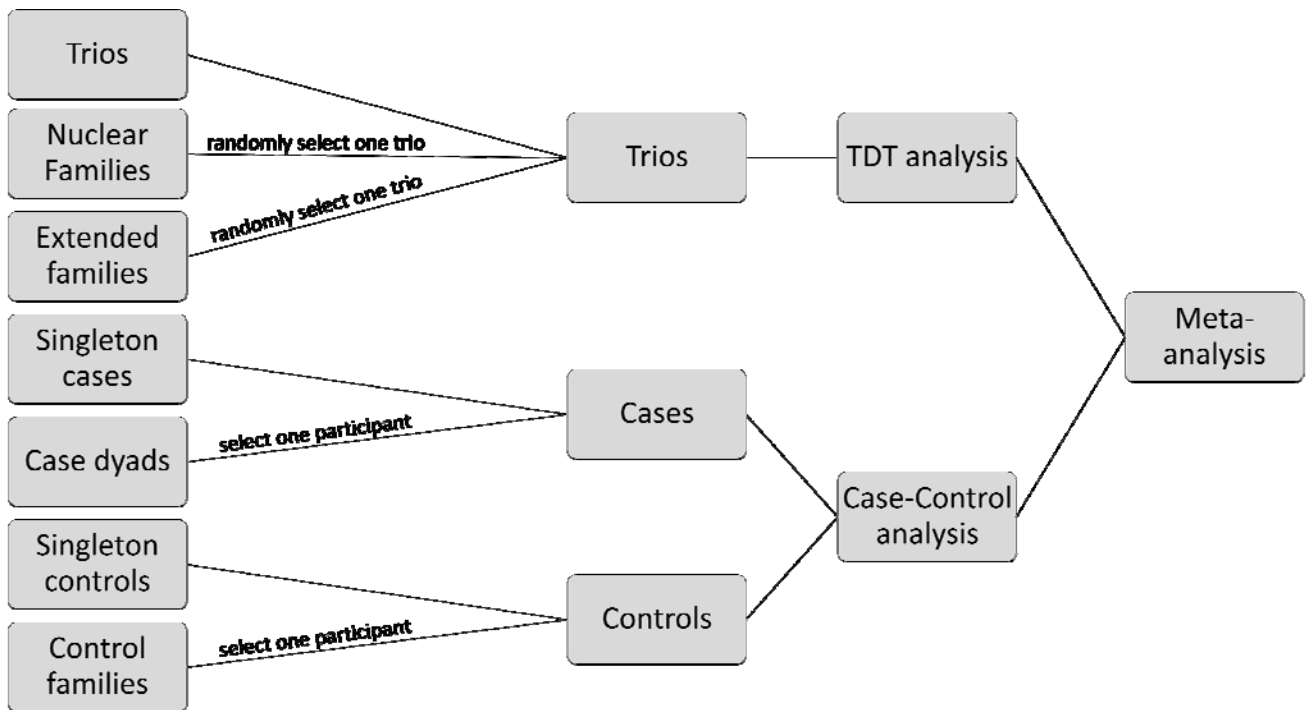
Locus	Gene	SNP	Population	P-value previous SNP	Best SNP in locus*	Best p-value in locus*
1p36	PAX7	rs742071	All Populations	6.80E-09	rs4920524	3.72E-09
			Europeans	9.16E-07	rs9439714	1.91E-07
			Asians	3.95E-01	rs12402156	9.27E-04
			Latin/South Americans	9.27E-04	rs6658147	1.25E-04
1p22	ARHGAP29	rs560426	All Populations	5.37E-05	rs4147882	5.59E-11
			Europeans	4.98E-01	rs12088309	3.63E-05
			Asians	1.58E-02	rs3789431	1.53E-05
			Latin/South Americans	3.39E-04	rs4147882	3.99E-07
1q32	IRF6	rs861020	All Populations	1.75E-08	rs11119345	2.52E-17
			Europeans	5.96E-02	rs56057644	2.92E-02
			Asians	2.35E-06	rs1044516	1.01E-06
			Latin/South Americans	6.94E-04	rs11119345	3.40E-13
1q32	IRF6	rs642961	All Populations	4.02E-09	-	-
			Europeans	4.66E-02	-	-
			Asians	2.16E-06	-	-
			Latin/South Americans	2.50E-04	-	-
2p21	THADA	rs7590268	All Populations	9.23E-03	rs10196106	2.98E-04
			Europeans	3.95E-01	rs67905560	3.22E-03
			Asians	6.24E-01	rs72613898	2.92E-02
			Latin/South Americans	2.98E-03	rs7604544	6.46E-04
3p11	EPHA3	rs7632427	All Populations	1.66E-02	3:89544366:C:T	3.90E-04
			Europeans	4.94E-02	rs9865804	3.67E-03
			Asians	3.94E-01	rs199821294	1.01E-02
			Latin/South Americans	1.58E-01	3:89544366:C:T	5.30E-04
8q21	DCAF4L2	rs12543318	All Populations	2.11E-07	same	same
			Europeans	1.41E-05	same	same
			Asians	9.74E-01	rs4961067	1.57E-02
			Latin/South Americans	2.84E-05	rs16875013	7.09E-06
8q24	--	rs987525	All Populations	6.58E-14	rs55658222	5.63E-30
			Europeans	1.47E-13	rs72728734	7.33E-15
			Asians	8.00E-01	rs112704402	3.74E-03
			Latin/South Americans	3.13E-06	rs55658222	3.98E-14
10q25	VAX1	rs7078160	All Populations	3.68E-06	rs10886040	7.19E-07
			Europeans	5.73E-03	rs201128098	9.04E-04
			Asians	7.37E-05	rs7072663	1.21E-05
			Latin/South Americans	6.31E-02	rs113339720	2.81E-02
13q31	SPRY2	rs8001641	All Populations	1.06E-05	rs11841646	7.44E-07
			Europeans	4.10E-02	rs11345095	1.66E-03
			Asians	6.00E-03	rs1854110	5.10E-04
			Latin/South Americans	1.27E-03	rs17072219	1.50E-04
16p13	CRBBP	rs8049367	All Populations	3.24E-02	rs2526687	1.35E-04
			Europeans	2.58E-01	rs137967419	1.84E-05
			Asians	3.70E-01	rs146255910	1.76E-03
			Latin/South Americans	2.44E-01	rs11076796	7.11E-04

17p13	NTN1	rs9788972	All Populations	1.42E-07	rs11273201	7.84E-12
			Europeans	6.26E-03	rs7406226	2.16E-07
			Asians	2.66E-02	rs1880646	1.29E-04
			Latin/South Americans	1.10E-05	rs7406226	1.46E-08
17p13	NTN1	rs4791774	All Populations	1.21E-04	-	-
			Europeans	1.66E-02	-	-
			Asians	7.80E-02	-	-
			Latin/South Americans	5.50E-03	-	-
17q22	NOG	rs227731	All Populations	9.77E-07	-	-
			Europeans	6.06E-02	-	-
			Asians	3.89E-01	-	-
			Latin/South Americans	8.56E-07	-	-
17q22	NOG	rs227727	All Populations	9.28E-07	same	same
			Europeans	5.20E-02	rs78717626	8.53E-03
			Asians	3.63E-01	rs62074819	9.76E-03
			Latin/South Americans	5.73E-07	same	same
20q12	MAFB	rs13041247	All Populations	4.23E-04	rs6016399	8.61E-07
			Europeans	2.31E-01	rs11698990	1.19E-02
			Asians	1.04E-01	rs6102085	1.05E-05
			Latin/South Americans	9.07E-04	rs4810299	3.25E-05

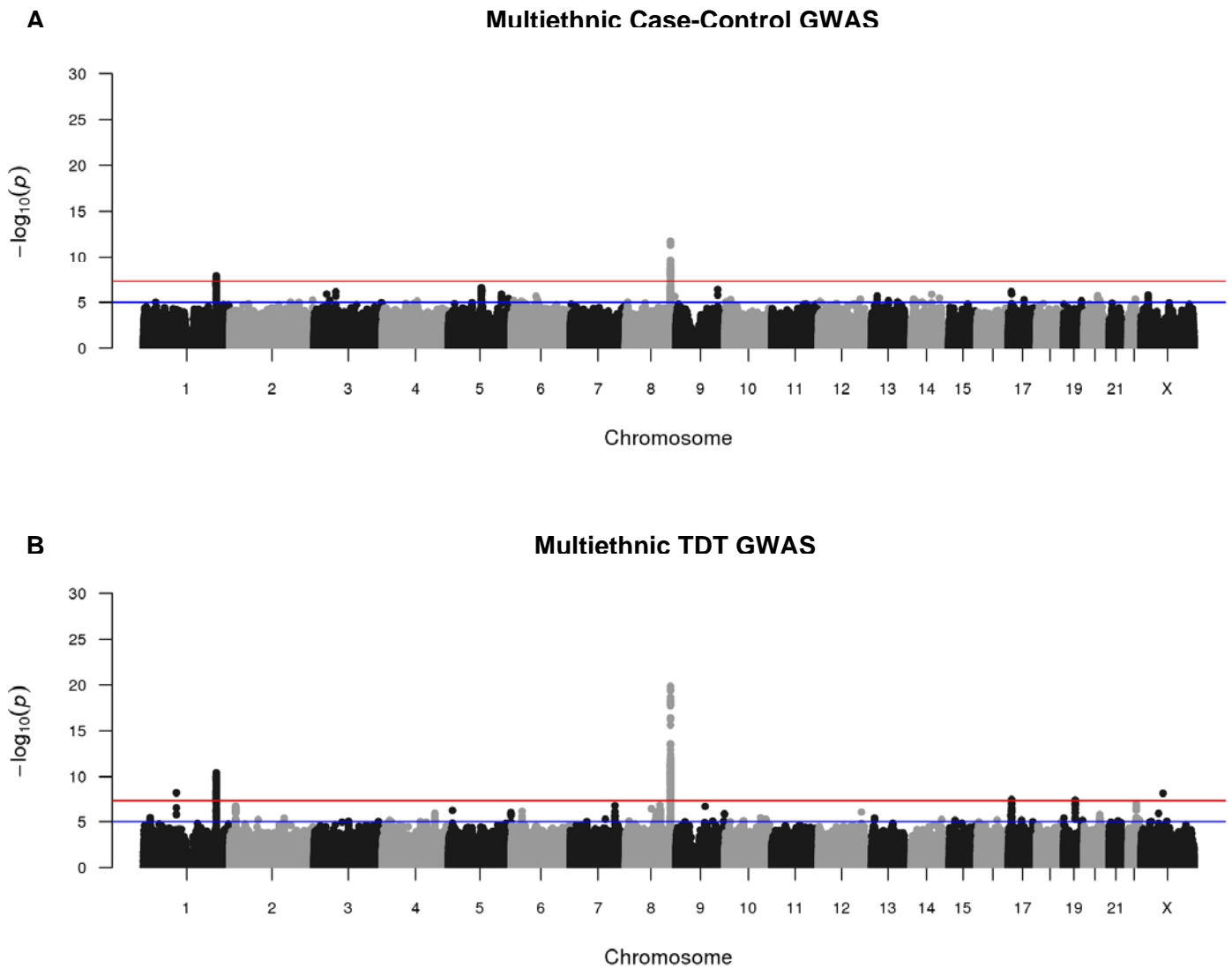
\* locus defined by +/- 100kb region around previously reported SNP



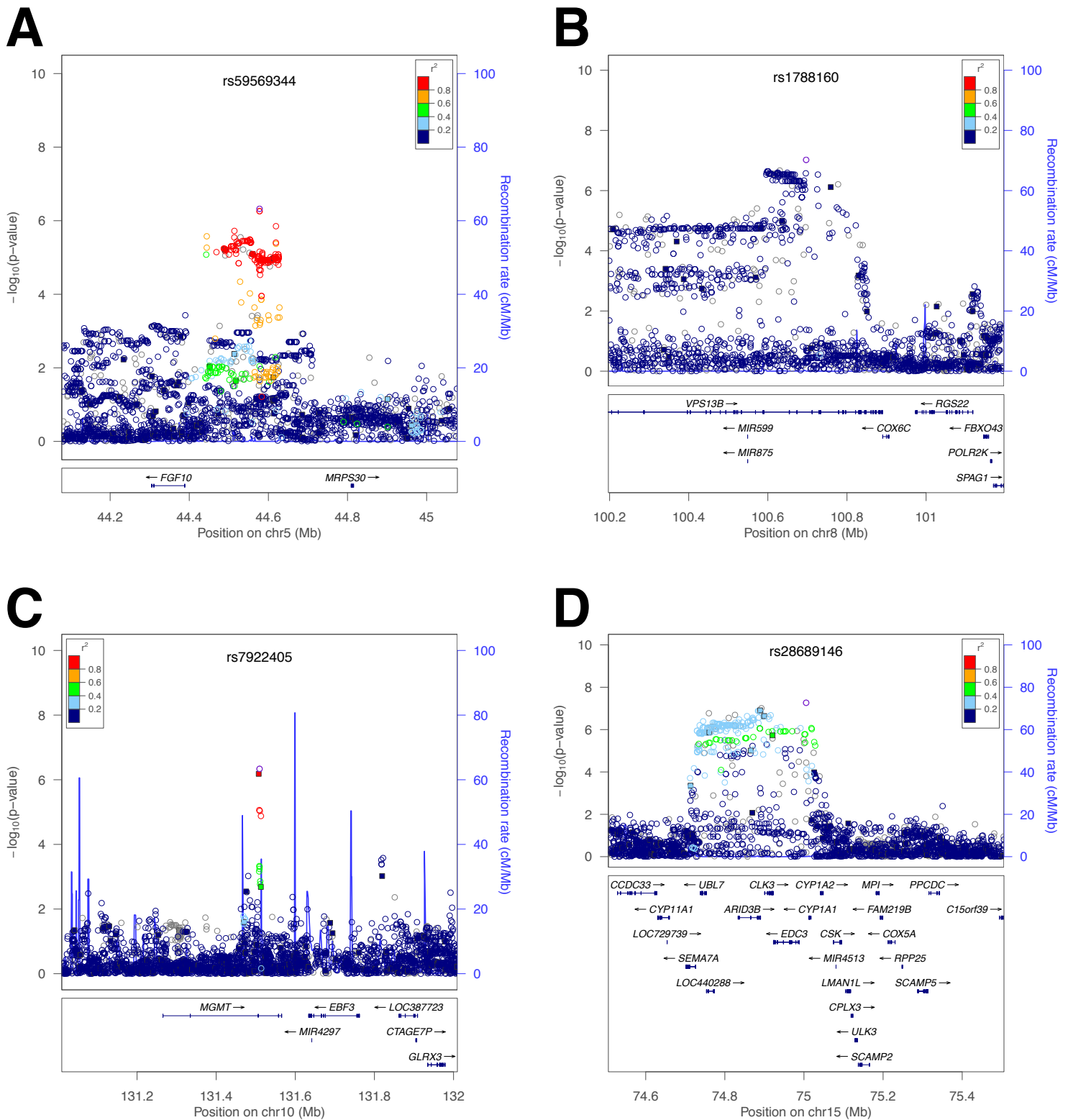
**Supplementary Figure 1. Population structure in the sample.** The first and second PCA eigenvectors, colored by (A) participant self-reported ancestry, and (B) HapMap control recruitment site, show the customary triangular pattern of ancestry with vertices comprising African, Asian, and European continental groups. (C) Parallel coordinate graph depicting scaled PCA eigenvectors 1-15 colored by self-reported ancestry shows many dimensions of population structure are represented in the sample. PCA eigenvectors 2-4 colored by (D) recruitment site, and (E) four genetic analysis groups, shows that the sample can be partitioned into ancestry groups, and that substantial population structure exists within each of these groups.



**Supplementary Figure 2. Flowchart showing the partitioning of the total sample into analysis sets.** The transmission disequilibrium test (TDT) analysis set comprised case-parent trios either recruited as trios or randomly selected from nuclear or extended families. The case-control analysis set comprised unrelated cases either recruited as singletons or chosen from case dyads and unrelated controls either recruited as singletons or chosen from control families. TDT and case-control results were combined via weighted odds ratio meta-analysis to generate study-wide association results.

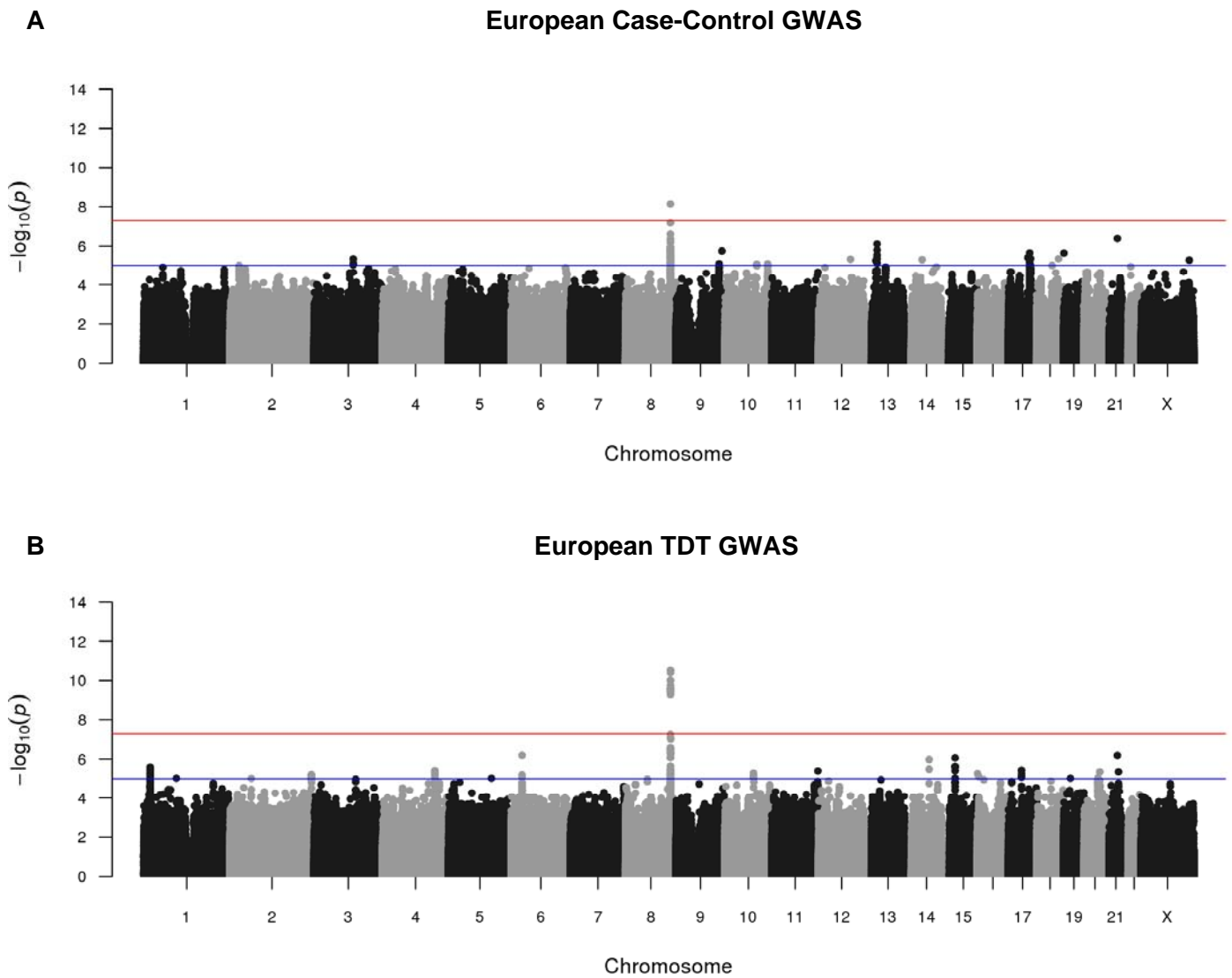


**Supplementary Figure 3. Manhattan plots for case-control and TDT analyses in multiethnic sample.** (A) Results for logistic regression analysis using 18 PCs of ancestry. Genomic inflation factor,  $\lambda$ , was 1.043. (B) Results for TDT analysis. Genomic inflation factor ( $\lambda$ ) was 1.068.

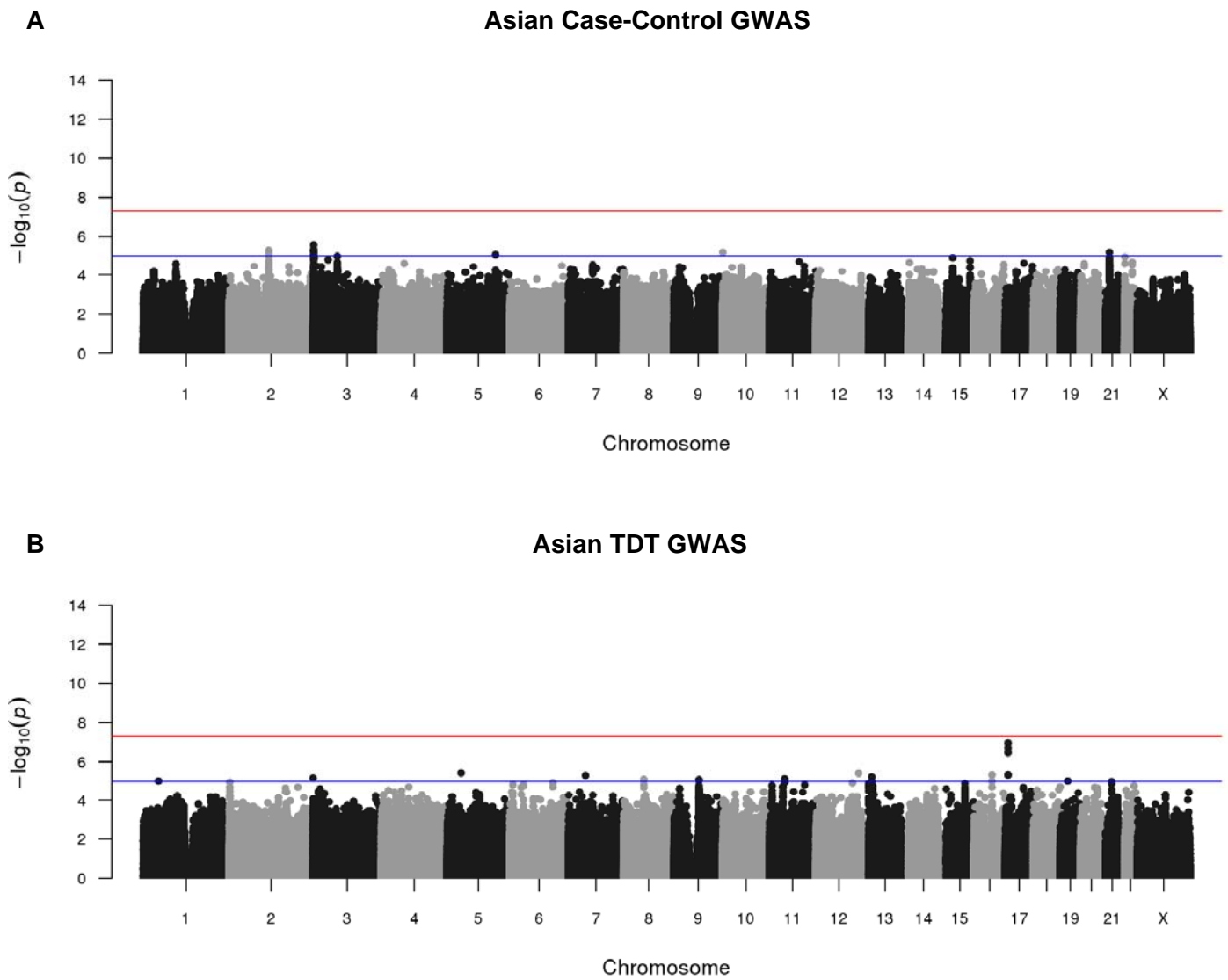


**Supplementary Figure 4. Novel loci approaching genome-wide significance in multiethnic meta-analysis.**

(A) 5p13 (B) 8q22 (C) 10q26 (D) 15q24. The lead SNP for each locus is colored purple. Genotyped SNPs are indicated by filled squares, imputed SNPs are open circles.

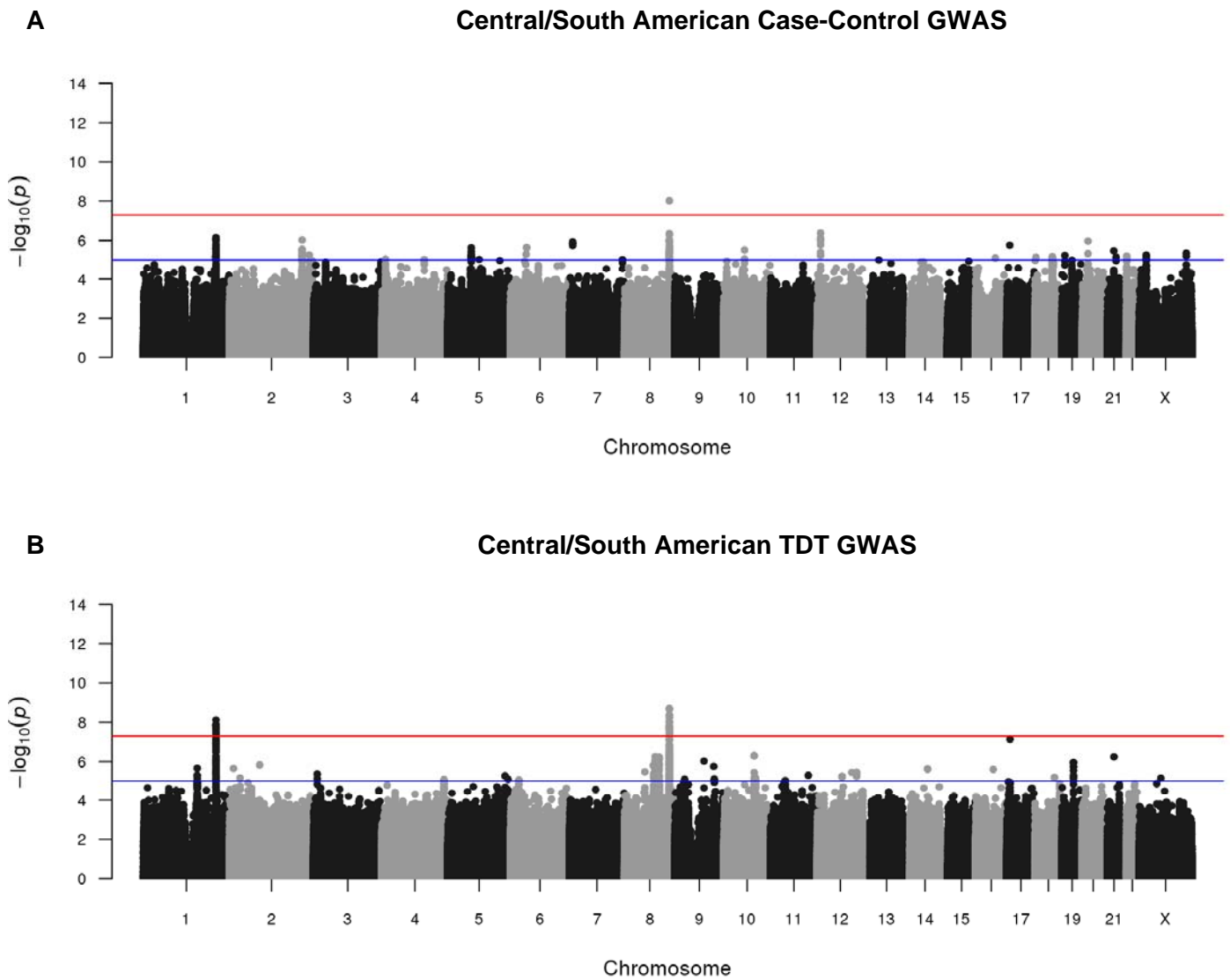


**Supplementary Figure 5. Manhattan plots for case-control and TDT analyses in the European subsample.** (A) Results for logistic regression analysis using 5 PCs of ancestry. Genomic inflation factor,  $\lambda$ , was 0.995. (B) Results for TDT analysis. Genomic inflation factor ( $\lambda$ ) was 1.041.

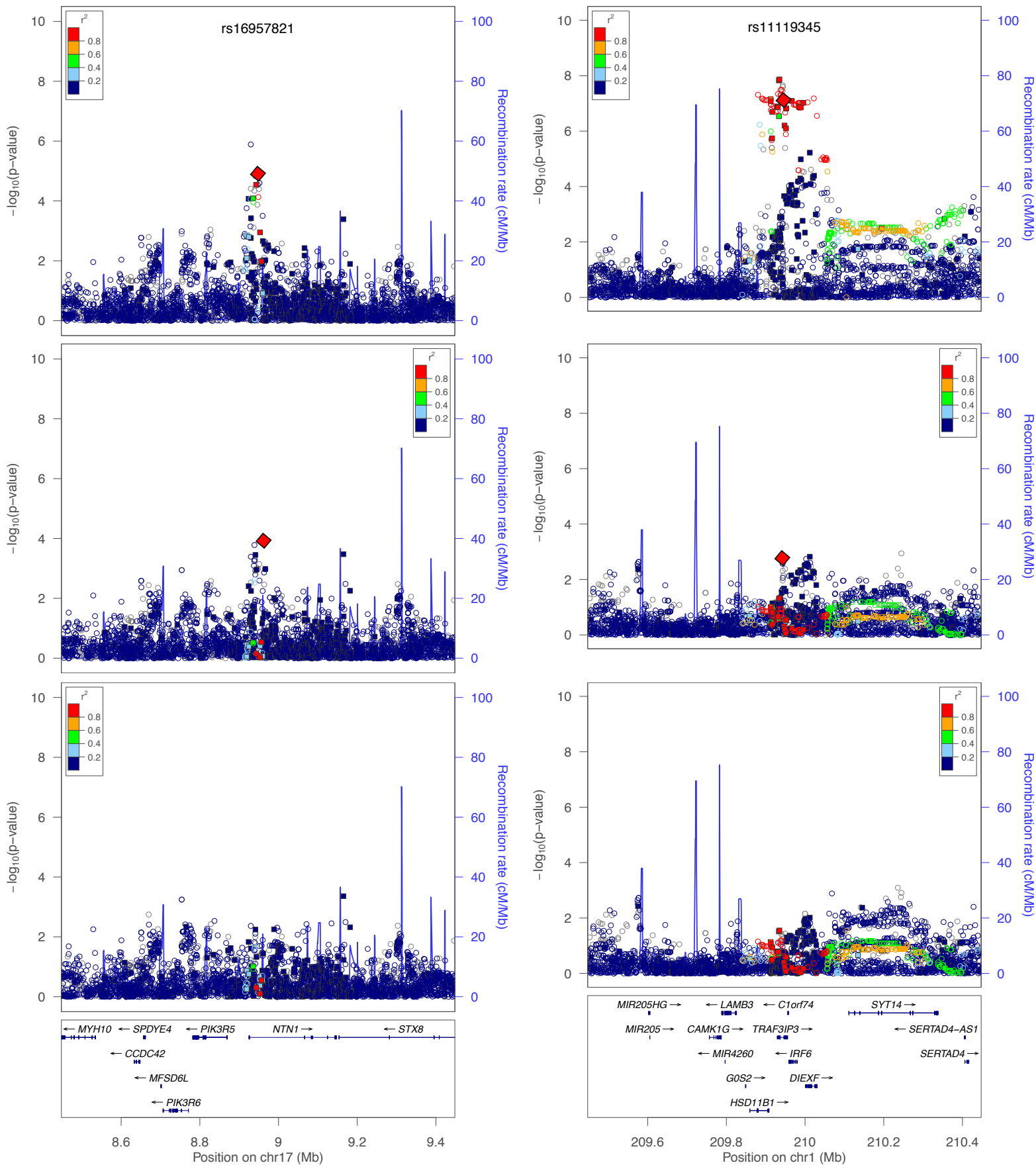


**Supplementary Figure 6. Manhattan plots for case-control and TDT analyses in the Asian subsample.** (A) Results for logistic regression analysis using 3 PCs of ancestry. Genomic inflation factor,  $\lambda$ , was 0.964. (B) Results for TDT analysis. Genomic inflation factor ( $\lambda$ ) was 1.039.

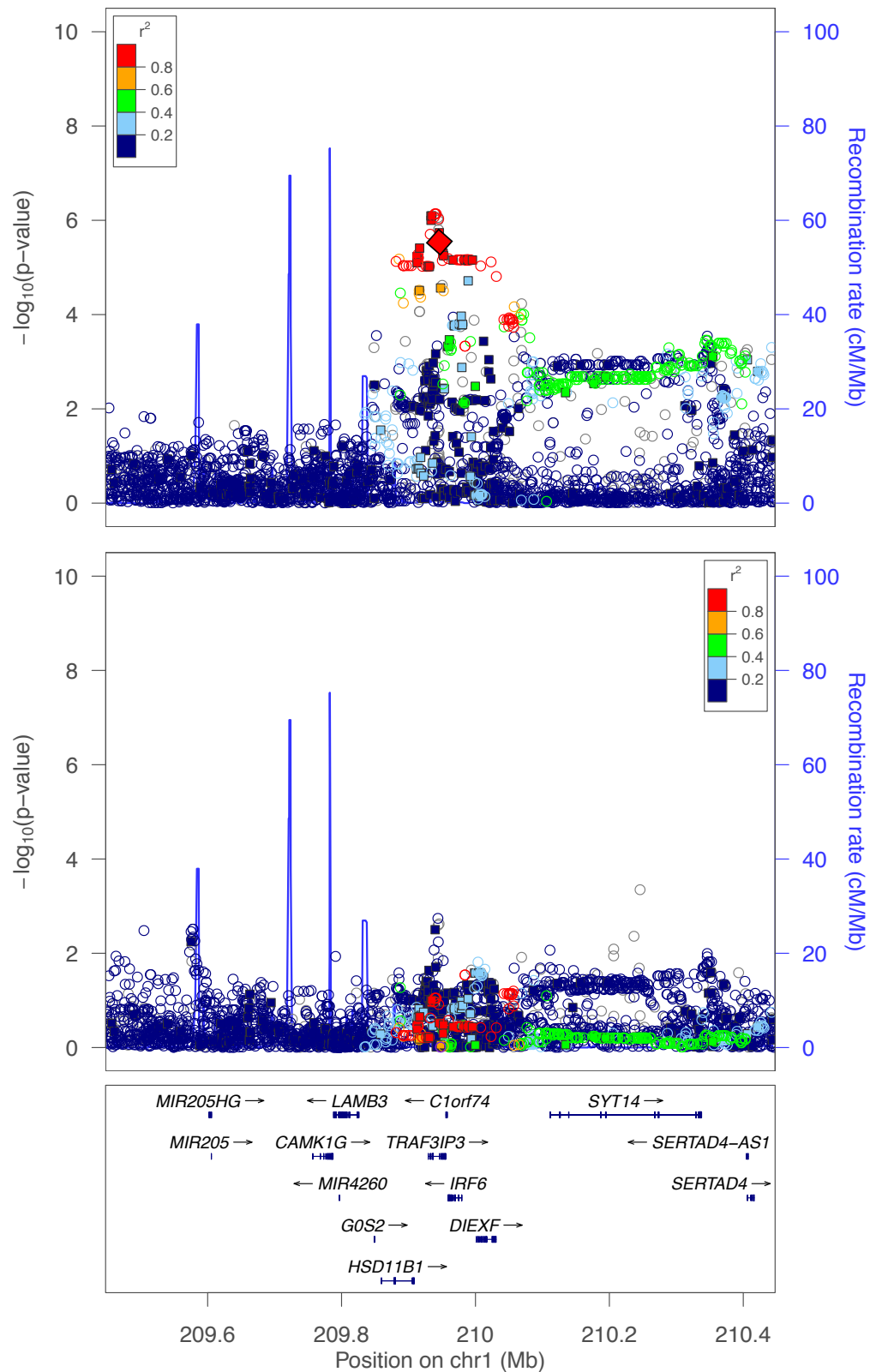




**Supplementary Figure 7. Manhattan plots for case-control and TDT analyses in the Central/South American subsample.** (A) Results for logistic regression analysis using 5 PCs of ancestry. Genomic inflation factor,  $\lambda$ , was 1.017. (B) Results for TDT analysis. Genomic inflation factor ( $\lambda$ ) was 1.049.



**Supplementary Figure 8. Conditional analysis at the 17p13 and 1q32 loci in multiethnic cases and controls.** (A) Stepwise conditional analysis in the 17p13 locus. (B) Stepwise conditional analysis in the 1q32 locus. The top panel shows the unconditioned results. The middle panel shows the results after conditioning on one of the lead SNPs (large red diamond). The bottom panel shows the results after conditioning on two SNPs.



**Supplementary Figure 9. Conditional analysis at the 1q32 locus in Central/South American cases and controls.** The top panel shows the unconditioned results, the SNP that was conditioned on is rs11119345 (large red diamond). The bottom panel shows the results after conditioning.