SUPPLEMENTARY MATERIALS

Suppl. Table S3: Annotated list of SNPs with $p<10^{-3}$ from the secondary meta-analysis	**NA**
Suppl. Table S4: Annotated list of SNPs with $p<10^{-3}$ from each subpopulation analysis	**NA**
Suppl. Table S5: Top LD-independent signals in the secondary meta-analysis of all TS samples	**35**
Suppl. Table S6: List of previously reported TS candidate genes	**23**
Suppl. Table S7: Evaluation of previously reported TS candidate genes	**38**

Abbreviations: AJ, Ashkenazi-Jewish European-derived isolate samples collected from the US and Israel; ANT, Latin American population isolate samples from Antioquia, Colombia; CVCR, Latin American population isolate samples from the Central Valley of Costa Rica; EU, European-ancestry, non-isolate samples collected from the US, Canada and Europe; FC, French-Canadian European-derived isolate samples collected from Quebec, Canada.

Supplementary Files. In addition to the figures and tables provided in this document, three additional supplementary tables (Supplementary Tables S2-S4) are provided in Excel format which contain an annotated list of all SNPs with $p<1 \times 10^{-3}$ from:

Supplementary Table S2: primary meta-analysis of European-ancestry samples (EU, FC, AJ)

Supplementary Table S3: secondary meta-analysis including all TS case-control samples (EU, FC, AJ and the CVCR/ANT Latin American population isolates)

Supplementary Table S4: results from each of the individual subpopulation analyses (EU, FC, AJ, CVCR/ANT)

SUBJECTS

Case subjects were recruited and assessed as described in the main text. Cases were recruited primarily from TS specialty clinics, with supplemental recruitment at each site through local advertising. In addition, population isolate cases were each required to have a specified number of grandparents (≥ 3/4 for AJ; 4/4 for FC) or great-grandparents (≥5/8 for CVCR; ≥6/8 for ANT) with self-reported ancestry from their specific population. Unselected, healthy FC controls were recruited from the general population of Quebec, Canada and were required to have 4/4 grandparents with FC ancestry. Unselected, healthy ANT controls were recruited from the general population of Medellin, Colombia either for population genetics studies or as controls for this and other disease studies in the ANT population. ANT Controls were required to have ≥6 great-grandparents from Antioquia (similar criterion as for ANT TS cases). Recruitment and assessment of unselected controls from the SAGE (Study of Addiction: Genetics and Environment), Illumina iControl and Dutch control samples were performed as described elsewhere[1-4] and online (www.Illumina.com, Illumina, San Diego, CA, USA).

GENOTYPING

As described in the main text, 1654 cases and 975 controls were genotyped on the Illumina Human610-Quadv1_B SNP array (Illumina, San Diego, CA, USA) at the Broad Institute of Harvard and MIT Center for Genotyping and Analysis (CGA) (Cambridge, MA, USA) in two batches (Sept-Nov 2008 and Dec 2008-Feb 2009). Genotypes were called from intensity data using Illumina BeadStudio (version 3.1.3.0) and GenTrain (version 3.2.32). Initial calls were made using a standard cluster file, followed by re-clustering using a revised custom cluster file for genotype clusters in which the automated calling algorithm did not adequately separate genotypes into appropriate clusters. SNPs were considered as technical failures if any of the following criteria were met: SNP call rate <97%, number of replicate sample genotype discordances >2, GenTrain score values <0.6 and BeadStudio cluster separation values <0.4. All genotypes were converted to missing for technically failed SNPs. All scans were subsequently re-called using the final, project-specific cluster definitions.

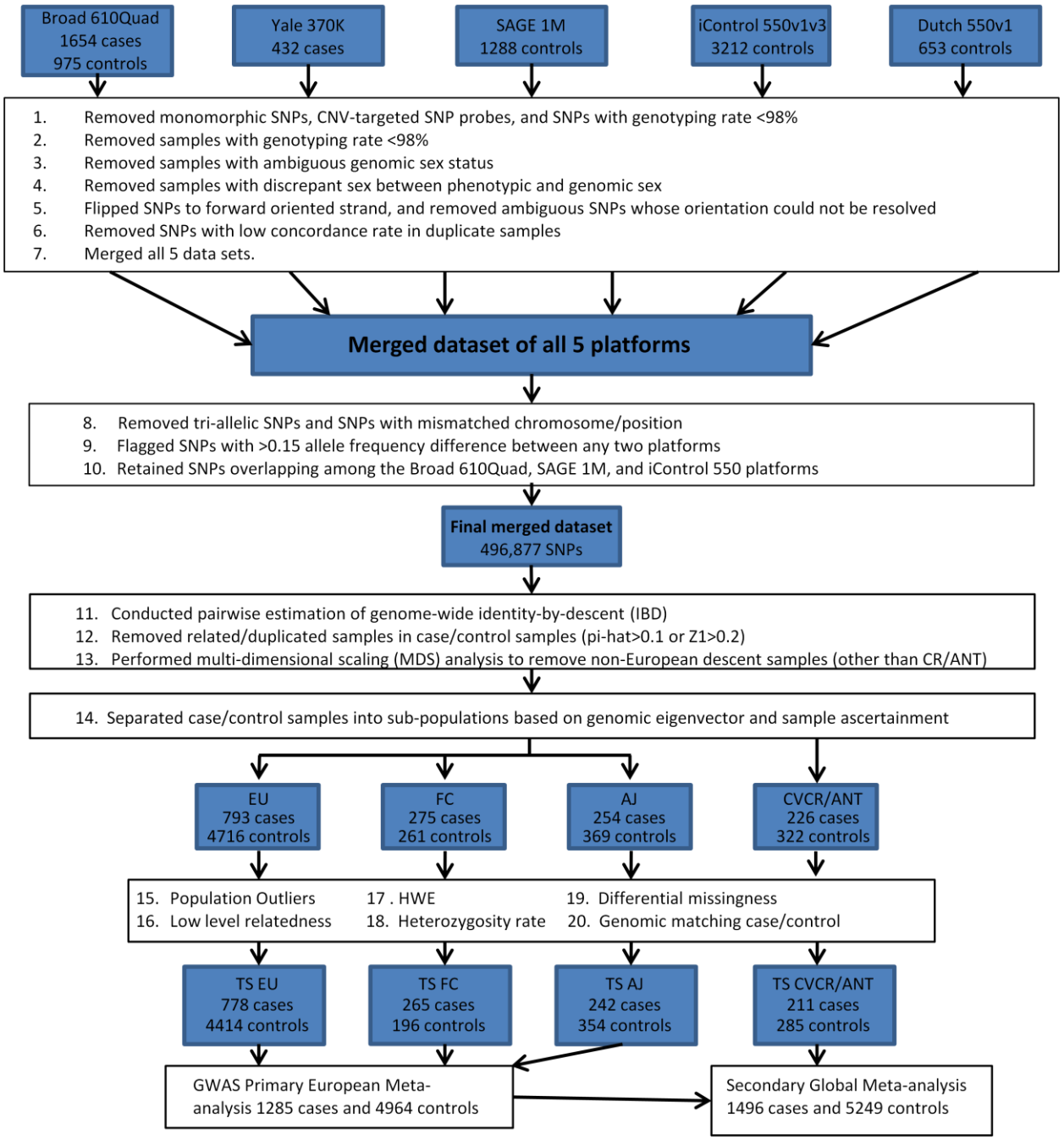
An additional set of 432 cases were genotyped on the Illumina HumanCNV370-Duo_v1 at the Yale Center for Genome Analysis (New Haven, CT, USA), including 88 duplicate EU samples overlapping with those genotyped on the Illumina 610Quad platform at the Broad Institute to allow for cross-platform checks of concordance.

Additional control genotypes were derived from three datasets previously genotyped on two nested Illumina SNP microarray platforms: **1) Illumina HumanHap550v1 & HumanHap550v3:** 3212 controls from the Illumina Genotype Control Database, Illumina, San Diego, CA; **2) Illumina HumanHap550v1:** 653 Dutch cohort controls genotyped at the Southern California Genotyping Center, UCLA, Los Angeles, CA, USA, including 48 duplicates also genotyped on the Illumina 610Quad at the Broad Institute for cross-platform QC. **3) Illumina Hap1Mv1_C:** 1288 European ancestry controls from the Study of Addiction: Genetics and Environment (SAGE) cohort genotyped at the Center for Inherited Disease Research (CIDR), Johns Hopkins University, Baltimore, MD, USA and accessed through dbGaP (download date: 08/25/2009). Of note, in order to increase the power of detecting poorly performing SNPs in this dataset, the initial platform-specific SNP QC steps **(Supplementary Figure S1)** utilized the entire European ancestry SAGE cohort of 2781 individuals, including 1181 SAGE cases consented for general health research. All SAGE samples other than the 1288 European ancestry SAGE controls were removed prior to multi-dimensional scaling analyses.

QUALITY CONTROL PROCEDURES

QC Overview

A schematic of the ordered QC pipeline is provided in **Supplementary Figure S1**.



Supplementary Figure S1. Quality control pipeline for the TS GWAS project.

Initial QC steps were performed in parallel within each of the five datasets, including two concordance checks of duplicate samples genotyped on two different platforms (610Quad vs 550v1 and 610Quad vs 370K) to confirm the robustness of Illumina genotyping across different platforms and sites as well as to remove SNPs

with discrepant calls across platforms. Platform-specific QC results at each filtering step are provided in **Supplementary Table S1.**

As noted in the main text, two SNP QC thresholds were generally used for each step: a more stringent threshold at which SNPs were removed from the analysis, and a second liberalized threshold for which SNPs were flagged in an annotation file and re-examined later for potential QC-related bias.

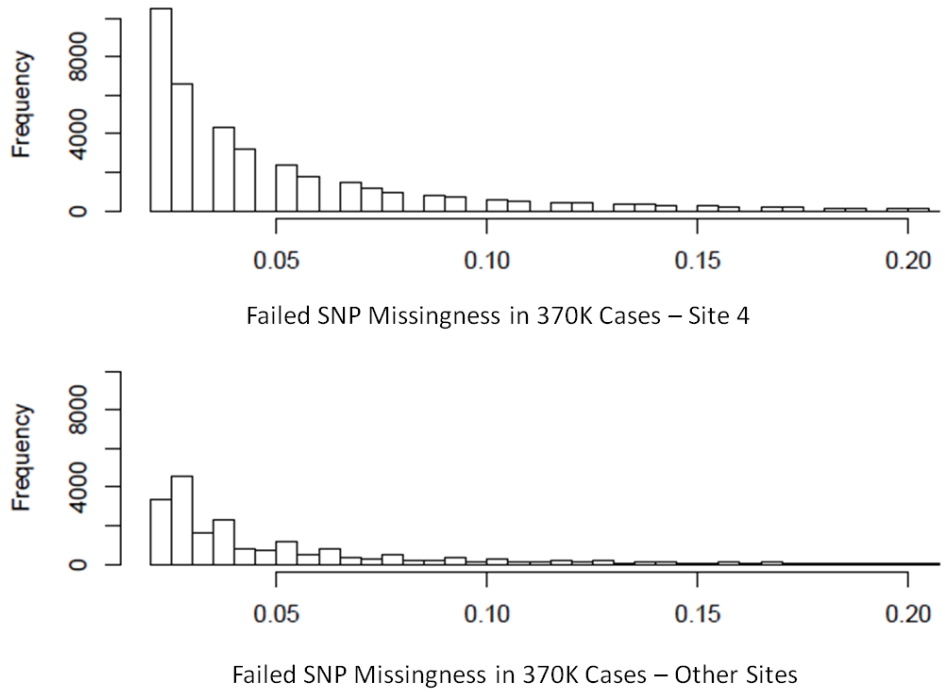
	TS Cases 610Quad	TS Controls 610Quad	TS Cases 370K	SAGE Controls 1M	Illumina iControls 550v1/v3	Dutch Controls 550v1
Total samples entering QC	1654	975	432	1288	3212	653
Low call rate	98	98	157‡	8	0	5
Sex discrepancy/ ambiguous genetic sex	22	19	8	1	3	1
Duplicates/Relatives	46	28	94*	0	21	49*
Abnormal heterozygosity (F_{het} outlier)	4	21	0	0	3	1
Related to multiple samples	1	2	0	1	2	0
Population stratification outliers	111	10	21	28	232	19
Unmatched controls/cases	24	70	4	40	214	3
Final samples for analysis	1348	727	148	1210	2737	575

Supplementary Table S1. Summary of sample-level QC for each genotyping platform. Cells indicate the number of samples removed at each QC step. ‡151 samples from one site were removed for increased rates of SNP missingness relative to other sites (Supplementary Figure S2). *Includes known duplicate pairs intended for cross-platform concordance checks. A total of 502 cases and 831 controls were excluded during the QC process.

Call rate

Iterative removal of poorly performing samples and SNPs was conducted to achieve a sample call rate of 98% and SNP call rate threshold of 98%. After setting an initial sample call rate threshold of 97%, SNPs with genotyping rate <97% were removed. Subsequently, remaining samples with call rates <98% were removed, followed by a final SNP genotyping rate threshold of 98%. In addition, when comparing genotyping call rates across the case samples genotyped on the Illumina 370K, a subset of samples, all originating from a single contributing site, were found to have consistently increased rates of SNP missingness even after restricting them to high call rate samples (>98% call rate) **(Figure S2).** Thus, the 151 samples from this site were

removed entirely from the analysis. Mitochondrial SNPs, Y chromosome SNPs, monomorphic SNPs and CNV-targeted SNP probes were also removed.



Supplementary Figure S2. Increased rate of SNP missingness in samples from one site (top) relative to other sites genotyped on the same platform (bottom). Axes indicates the number of SNPs with missing rates >2% in high call-rate samples (sample call rate >98%).

Genomic sex

Genetic sex status was determined using the F statistic to estimate homozygosity on the X chromosome. A genomic male was defined as F>0.75, while a genomic female was defined as F<0.25. Any samples with F statistic between 0.25 and 0.75 were defined as having ambiguous genomic sex and were removed from the study. In addition, samples with discrepant sex between reported phenotypic and genomic sex were also removed.

Resolution of strand-ambiguous SNPs

Within each platform, SNPs were oriented to genome-forward orientation using 3.96 million SNPs from all HapMap2 individuals as a reference dataset (release 23, NCBI build 36 coordinates). Strand-ambiguous

SNPs with minor allele frequency (MAF) < 0.3 and an allele frequency difference <0.2 relative to the reference data were retained. However, strand-ambiguous SNPs either with MAF > 0.3 or with an allele frequency difference of >0.2 between the observed, platform-specific MAF and the HapMap CEU reference data were excluded. As an additional screen for ambiguous SNPs oriented to the incorrect strand, each strand-ambiguous SNP was evaluated for inverted linkage disequilibrium (LD) patterns in relation to adjacent SNPs with LD correlations in the HapMap2 CEU sample serving as a reference (--flip-scan procedure in PLINK). Finally, any strand-ambiguous SNPs not present in the HapMap reference dataset were removed.

Concordance checks

To verify the robustness of Illumina genotyping across different platforms and sites, rates of cross-platform concordance were assessed between two sets of duplicate samples genotyped on different platforms and centers. After the first two QC steps, 82 duplicate case pairs remained that were genotyped both on the 610Quad at the Broad Institute and the 370K at Yale. Of these 82 pairs, there were 4253 discordant SNP calls out of 25,218,718 SNPs genotyped on both platforms, resulting in a concordance rate of 99.98%. In addition, there were 41 duplicate control pairs genotyped both on the Illumina 610Quad at the Broad Institute and the Illumina 550v1 at UCLA. Of these 41 duplicate pairs, there were 2337 discordant SNP calls out of 20,575,349 SNPs genotyped on both platforms, resulting in a cross-platform concordance rate of 99.99%. Across both data sets, only 15 SNPs were discordant in >2 samples and were removed from the analysis. The remaining discordant SNPs were flagged in the annotation file, none of which had $p<1.0x10^{-3}$ in any analysis. In addition, 4106 SNPs were removed based on reported poor concordance across different Illumina platforms in other studies[5] (N. Cox, personal communication). Lastly, GWAS data from HapMap samples in the Illumina database (www.icom, Illumina.com, Illumina, San Diego, CA, USA) genotyped previously on at least 2 of the 5 platforms used in this study were used to examine SNP concordance rates between all possible cross-platform duplicate pairs (total of 1068 pairwise comparisons). Any SNP with >1 cross-platform discordant call within each HapMap population was flagged. Flagged SNPs with subsequent GWAS association p-values $<1.0 \times 10^{-3}$ are annotated in **Table S2-S4** (only 2 SNPs (ranked 429th and 490th) were flagged for cross-platform

discordances in the primary European meta-analysis; no SNPs were flagged for cross-platform discordances in the secondary global ancestry meta-analysis).

Batch effect testing

Since the Illumina 610Quad samples were genotyped in two batches (Sept-Nov 2008 and Dec 2008-Feb 2009), a logistic regression analysis was performed using batch indicator as phenotype with case-control status and MDS dimensions as covariates in the regression model. The resulting Q-Q plot showed no evidence of overdispersion to suggest a batch effect (λ=0.956). Three SNPs with $p<10^{-5}$ in the batch effect regression analysis were flagged and none of these appeared in the top 551 SNPs in the final primary association meta-analysis.

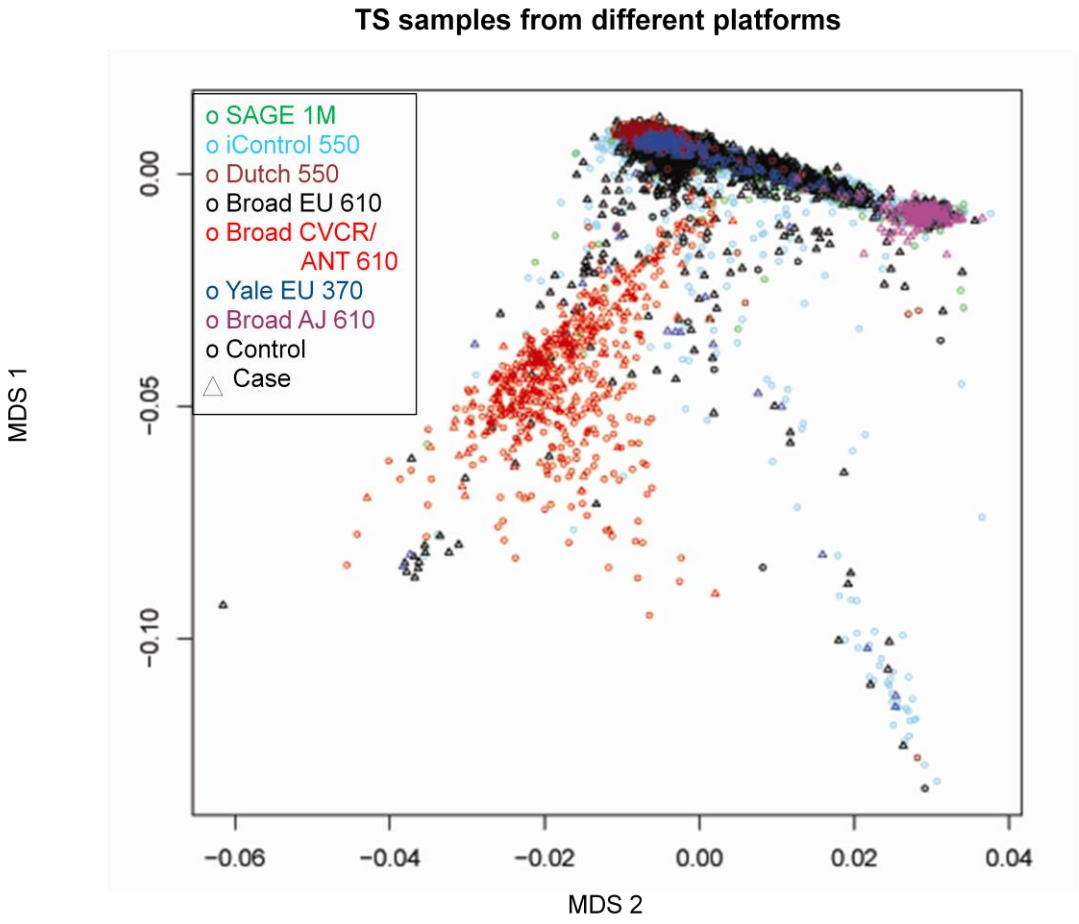
Platform merging and initial cross-platform comparisons

At this stage in the QC process, all samples were merged into a single dataset using PLINK. Following the merge, 23 SNPs were either mismatched or tri-allelic and were removed. SNP allele frequencies were compared among each platform and any SNP with an absolute allele frequency difference >0.15 between two platforms were flagged. SNPs were also flagged in all samples if they generated >1% Mendelian errors in a sample of 400 OCD trios genotyped in parallel with the TS cases on the Illumina 610Quad at the Broad Institute (see accompanying manuscript by Stewart et al.) Lastly, any SNP not in common between the cleaned Illumina 1M, 610Quad and 550K iControl samples were removed, leaving 496 877 cleaned SNPs for subsequent analyses.

Removal of duplicates, related samples and individuals of non-European descent

For all 7651 case-control samples remaining in the common dataset, pairwise estimation of genome-wide identity-by-descent (IBD) was conducted with an LD-pruned set of 51 516 SNPs using PLINK. One individual from each sample pair with either a pi-hat>0.1 or Z1>0.2, representing unexpected duplicates or relatives, was

removed from subsequent analyses. All remaining case-control samples were subjected to a multi-dimensional scaling (MDS) analysis to identify individuals of non-European ancestry (**Figure S3**).

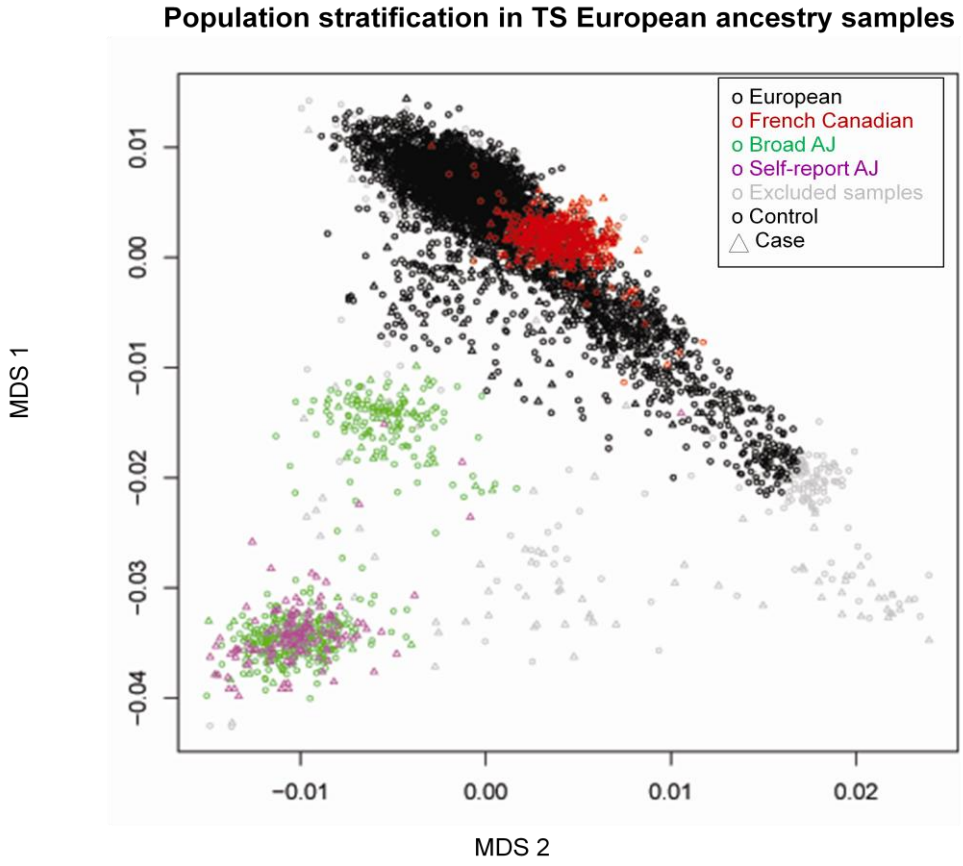


Supplementary Figure S3. Multi-dimensional scaling (MDS) plot of all TS GWAS case-control samples.

CVCR and ANT cases and controls (red) were set aside for further subpopulation-specific QC and analyses (see below). The majority of non-CVCR/ANT samples clustered along a diagonal with samples of Dutch origin at the top left (brown) and self-reported Ashkenazi Jewish (AJ) samples at the bottom right (purple), consistent with the expected distribution of European ancestry samples along a Northern to Southern European cline (**Figure S3**). However, 69 cases and 138 controls fell far outside this general European ancestry cluster and thus were removed from analysis due to the presence of non-European genetic ancestry (**Table S1**).

Separation of case-control samples into genetically homogeneous subpopulations of European-ancestry derived samples: EU, FC, and AJ

After removing all individuals with non-European genetic ancestry, a second European ancestry MDS analysis was performed to stratify remaining samples into more homogeneous subpopulations and to re-assign individuals whose self-reported ancestry did not reflect observed genetic ancestry (**Figure S4**).



Supplementary Figure S4. MDS plot TS GWAS case-control samples of European ancestry.
As expected, most case-control subjects clustered together in a homogeneous cloud along the expected Northern-Southern European cline (black). These individuals were separated out as a non-isolate European ancestry stratum (EU) for subpopulation-specific QC and analysis.

AJ Subpopulation

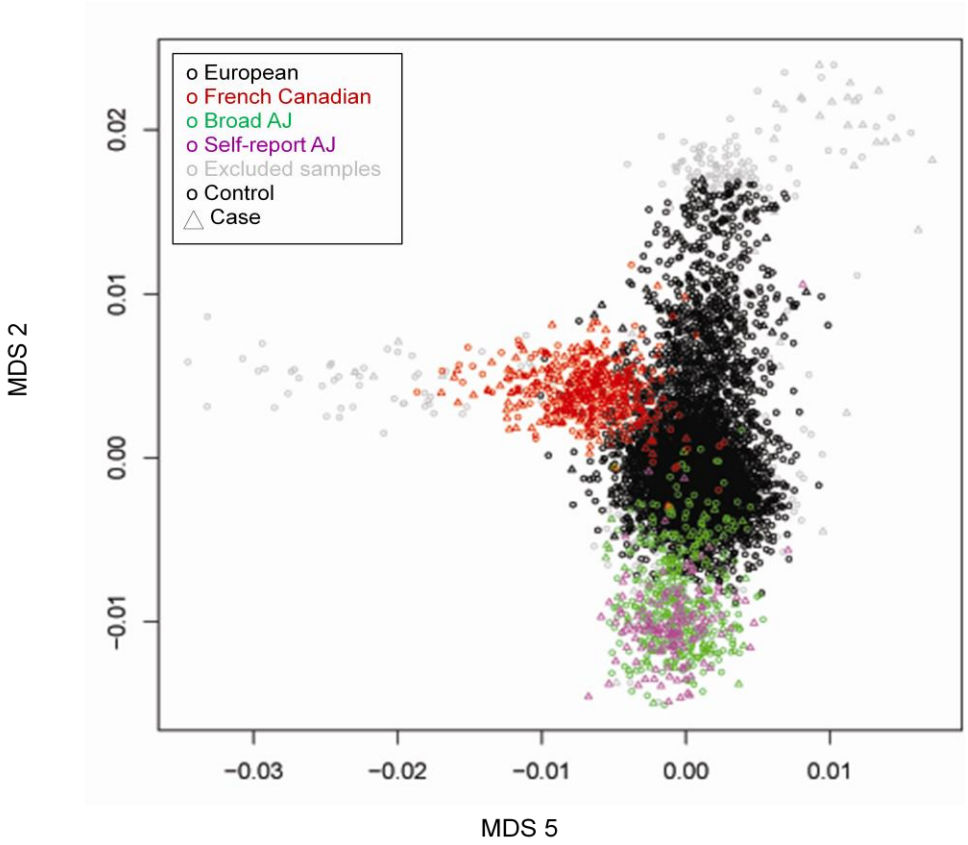
Two major clusters of individuals were identified that were distinct from the main EU sample in the European ancestry MDS analysis (**Figure S4**). The majority of individuals with self-reported AJ ancestry (purple)

segregated in one of these clusters along with an additional 101 cases and 369 controls from the general EU case-control sample (green). These "general EU" individuals did not differ significantly from self-reported AJ samples on any of the first 10 MDS dimensions and thus were re-assigned as "genetic AJ" individuals and analyzed with the AJ subpopulation. Of note, the middle green cluster contained many samples with self-reported half-AJ/half-EU ancestry. Due to the small number of samples, this "half-AJ cluster" was combined with the main AJ cluster and analyzed together as a single "broad AJ" stratum.

Two additional clusters within the European ancestry MDS analysis were noted with MDS1 scores ~-0.03 and MDS2 scores centering around 0.005 and 0.02, respectively (**Figure S4**, bottom middle and bottom right, grey). These clusters contained samples from the US and Israel who reported either Sephardic Jewish ancestry (bottom right, grey) or half AJ-half Sephardic ancestry (bottom middle, grey). However, since these clusters primarily contained TS cases without closely matched controls, they were removed from the association analyses.

FC Subpopulation

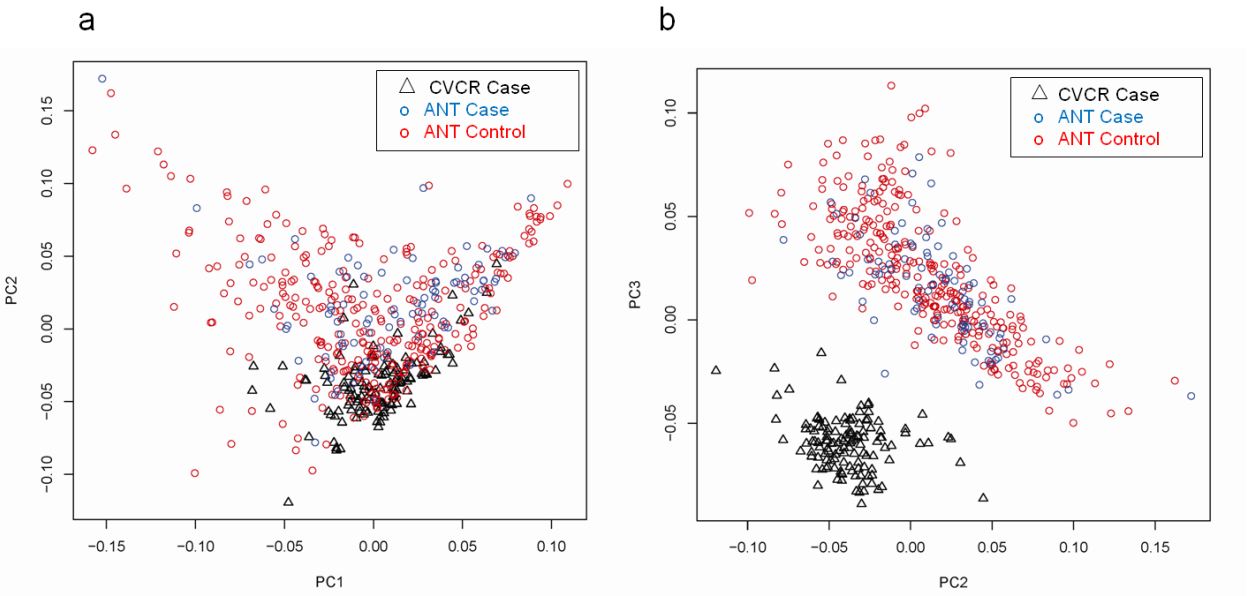
Although the French Canadian (FC) cases and controls (red) also fell within the general EU cluster, further MDS analyses identified additional dimensions that distinguished FC cases and controls from the EU samples, and thus they were analyzed separately as an FC-specific stratum (**Figure S5,** red).



Supplementary Figure S5. MDS plot of European ancestry TS GWAS samples. Plots of additional MDS dimensions (here the 2nd and 5th dimensions) demonstrated a separation of the French Canadian (FC) case-control sample (red) from the other European ancestry (EU) samples (black).

Latin-American ancestry samples: CVCR and ANT

Previous analyses have indicated that the CVCR and ANT population isolates are closely related Mestizo populations with common European, African and Native American founders.[6, 7] Thus, ANT controls were used for a pooled Latin American population isolate analysis of both ANT and CVCR TS cases. Whole-genome principal component analyses confirmed the close relationship between CVCR and ANT individuals, but also identified a distinct component of genetic ancestry between the two populations (**Figure S6**).



Supplementary Figure S6. PCA plots of the Latin American TS GWAS population isolates. a. Plots of the first two principal components (PCs) in EIGENSTRAT demonstrate the expected overlap between TS cases from the Central Valley of Costa Rica (CVCR, black) and TS cases (blue) and controls (red) from Antioquia, Colombia (ANT). **b.** Plot of the 2nd and 3rd principal components indicate residual stratification between CVCR and ANT.

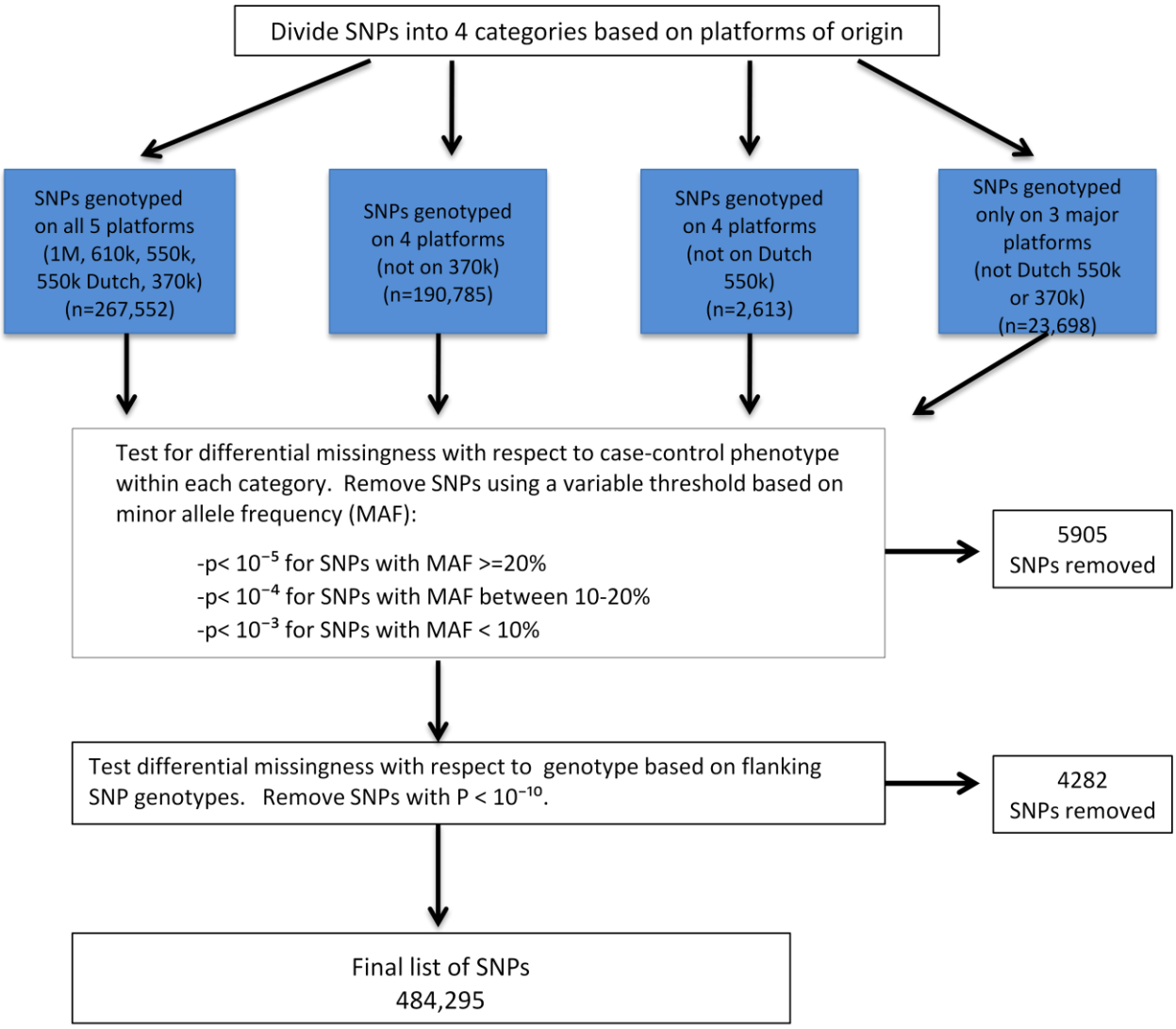
The effects of these subtle ancestry differences were able to be attenuated by incorporating significant MDS axes as covariates in the association analyses (λ_{GC}=1.04, see below). However, due to this residual population stratification, we opted to use a meta-analysis of the three European-derived subpopulations (EU, AJ, FC) for the primary analysis and reserve the CVCR/ANT samples for a secondary meta-analysis of all TS samples.

Subpopulation-specific QC

After separating the final samples into four subpopulation-specific strata (EU, AJ, FC, CVCR/ANT), an additional set of QC analyses were undertaken within each subpopulation to optimize case-control matching and to remove remaining poorly performing samples and SNPs (**Figure S1**). First, samples were removed that demonstrated low-level relatedness (Z1>0.1 with a large number (≥20) other samples in the subpopulation). Second, samples within each subpopulation were subjected to a cluster analysis (--cluster in PLINK), and any sample whose pairwise identity-by-state distance from the closest samples was > 5 standard deviations

compared to the rest of samples was removed. Average heterozygosity was calculated, and any sample with F_{het} > ±0.05 was also removed from the final analysis. Following these sample QC steps, SNPs were tested for the presence of Hardy-Weinberg disequilibrium in controls from each subpopulation. Any SNPs with HWE $p<10^{-10}$ were removed; those with HWE $p<10^{-5}$ were flagged. For the CVCR/ANT populations, Hardy Weinberg equilibrium testing was performed both in the ANT controls alone as well as compared across the CVCR and ANT TS cases; SNPs with HWE $p<10^{-10}$ across the two Latin American populations were removed, while those with HWE $p<10^{-5}$ were flagged. Allele frequencies (AF) were also examined between CVCR and ANT cases, and any SNPs with an AF difference >0.15 were flagged.

Given the use of five different datasets across four nested Illumina platforms, we performed an additional QC step within each subpopulation to identify SNPs with differential missingness between cases and controls (i.e., differential missingness by phenotype) across 4 cross-platform combinations (**Figure S7**). For each of these comparisons, SNPs were removed for cross-platform differential missingness using increasing levels of stringency with decreasing minor allele frequency thresholds. For SNPs with MAF≥0.2, SNPs were excluded with Chi-square test $p<10^{-5}$ (χ^2 test, 1df). For SNPs with MAF<0.2, but ≥0.1, SNPs were excluded with $p<10^{-4}$. Lastly, for SNPs with MAF<0.1, SNPs were excluded with $p<10^{-3}$. In addition, a haplotype-based test for non-random missingness by genotype was performed (--test-mishap in PLINK). This test utilized haplotypes generated by two SNPs flanking the reference SNP as a means of predicting the reference SNP's true genotype when that SNP is missing from the dataset. The haplotypes are then tested as a proxy for the presence of differential (i.e., non-random) missingness between the reference SNP's true genotype (e.g., whether an A/G SNP has higher rates of missingness when the true underlying genotype is AA rather than AG or GG) (http://pngu.mgh.harvard.edu/~purcell/plink/summary.shtml#testmiss). SNPs were excluded if surrounding flanking SNP haplotypes were associated with missingness at the reference SNP site at $p< 10^{-10}$. Any SNP that failed either differential missingness test in one subpopulation was subsequently removed from all populations prior to association analysis.



Supplementary Figure S7. Schematic of differential SNP missingness tests for cross-platform comparisons. Each of the cross-platform missingness tests above were performed separately in each of the four ancestral subpopulations (EU, AJ, FC, CVCR/ANT). However, any SNP that failed one of the missingness tests in any individual subpopulation was subsequently removed from all subpopulations prior to association analysis. A total of 5905 SNPs were removed from all samples based on differential missingness with respect to phenotype (i.e. between cases and controls). An additional 4282 SNPs were removed from all subpopulations due to differential missingness with respect to flanking SNP genotypes.

In addition, for EU cases of known Dutch ancestry genotyped on the 610Quad platform, all SNPs absent from the Dutch Hap550v1 control dataset were removed to reduce any differential missingness specific to these ancestry-matched samples.

Two further rounds of MDS analyses were conducted within each ancestry-specific subpopulation. The first set of subpopulation-specific MDS analyses was used to remove any remaining samples with poor case-control matching. A final MDS analysis was then performed to identify MDS dimensions which could explain any residual population stratification. MDS dimensions were retained for subsequent association analysis if: 1) they were associated with the TS phenotype at $p<0.01$; or 2) for dimensions association with the TS phenotype at $0.01<p<0.05$, a scree plot was generated to determine the corresponding genomic control (λ) values associated with inclusion of each MDS dimension. All dimensions demonstrating a notable drop in genomic control values relative to prior MDS dimensions were retained. These MDS dimensions were included as covariates in the logistic regression model used for tests of association.

X chromosome QC

QC steps for X chromosome SNPs followed the same pipeline as for autosomal SNPs (**Figure S1**) with a few modifications. In the first QC step, a SNP call rate threshold of 98% was used as calculated based on female samples only. Similarly, for resolution of strand-ambiguous SNPs (**Figure S1**, step 5 and pg. 8 above), allele frequencies were estimated based on female samples only. Third, prior to merging samples from each platform, samples with a call rate <95% on the X chromosome were removed from analysis. After dataset merging, 1915 SNPs were removed for having heterozygous genotypes in males. Finally, in the subpopulation-specific QC (**Figure S1**, step 17), a more conservative cutoff for SNPs in Hardy-Weinberg disequilibrium was used (HWE $p<0.001$ in female controls).

Of note, no pre-defined pseudo-autosomal SNPs were genotyped on the 610Quad and thus were not available for analysis. Similarly, since only 129 Y chromosome SNPs passed QC with a call rate>98% in males, all Y chromosome SNPs were removed from the analysis.

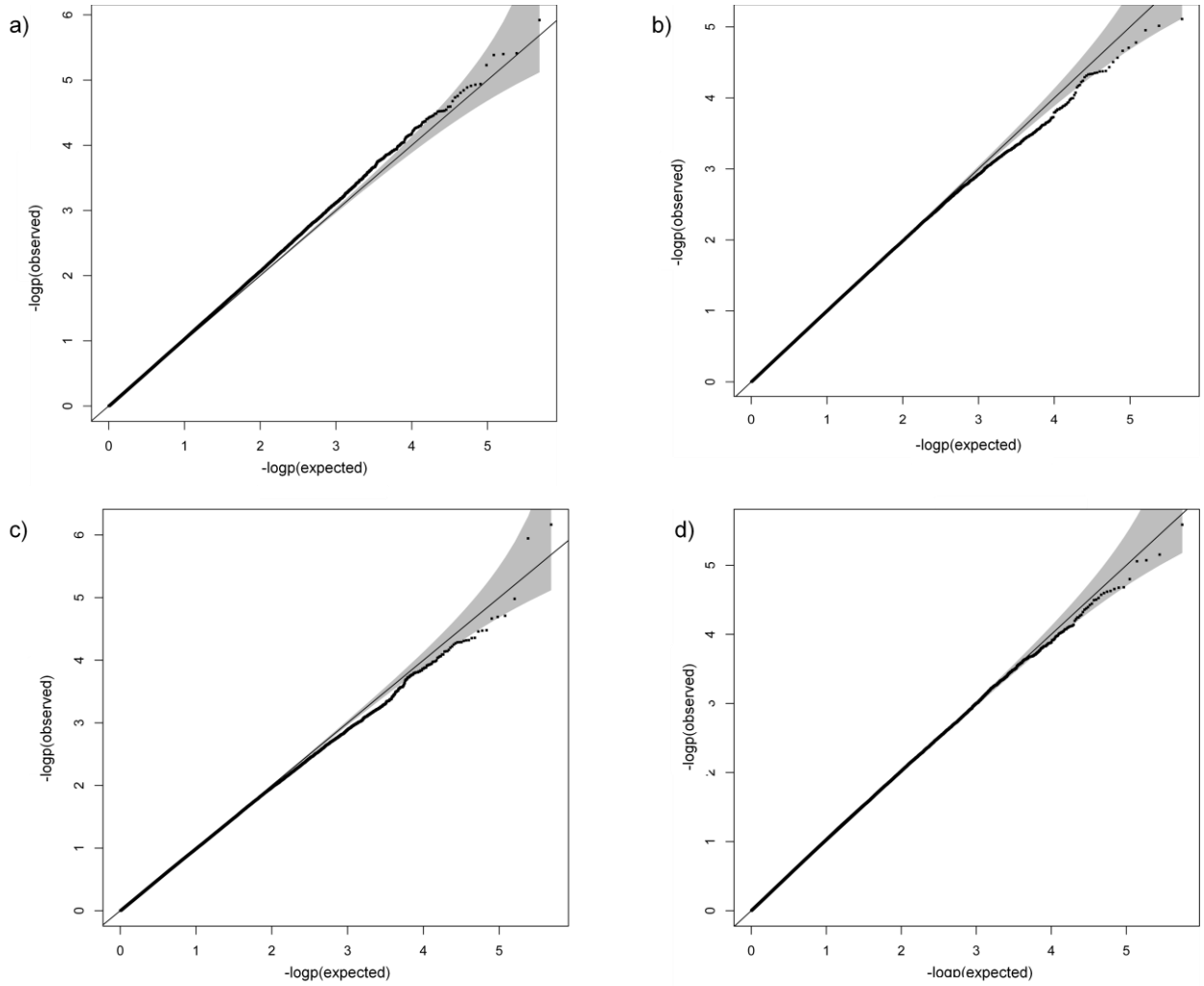
Post-hoc confirmation of QC analyses

As a final step to confirm the quality of the QC process, we examined the square of the GWAS test statistic for any correlation with residual call rate, Hardy-Weinberg p-value and minor allele frequency of the surviving SNPs, none of which were significant (data not shown). For the top SNP in both meta-analyses, rs7868992 on chromosome 9, the cross-platform concordance rates were re-examined both in sample duplicates genotyped in this study (see above) and in the HapMap samples from the Illumina control database that were genotyped on multiple Illumina platforms. In each of these comparisons the concordance rate for rs7868992 was 100%.

ANALYSES

Subpopulation-specific association analysis

Following QC, each of the four cleaned datasets (EU, AJ, FC, CVCR/ANT) were analyzed as separate subpopulations in PLINK using logistic regression under an additive model (Cochran-Armitage trend test) with subpopulation-specific MDS dimensions incorporated as covariates in each analysis (EU: 3 MDS dimensions, AJ: 1 MDS dimension; FC: 1 MDS dimension; CVCR/ANT: 6 MDS dimensions). As noted in the main text, X chromosome SNPs were first analyzed separately in males and females and subsequently combined by meta-analysis in METAL.[8] Quantile-quantile plots of each subpopulation-specific analysis revealed no evidence of residual population stratification or significant systematic technical artifacts (**Supplementary Figures S8a-d).**



Supplementary Figure S8. Quantile-quantile plots of observed vs. expected -log (p) values for the four subpopulation-specific GWAS analyses. The 95% confidence interval of expected values is indicated in grey. a) EU analysis, $\lambda = 1.011$; b) AJ analysis, $\lambda = 0.993$; c) FC analysis, $\lambda = 0.971$; d) CVCR/ANT analysis, $\lambda = 1.044$.

Meta-analysis

Meta-analysis was conducted using METAL, which combined the p-values using the number of cases in each subpopulation-specific stratum for weighting.[8] Two meta-analyses were conducted: a primary meta-analysis of the three European-derived populations (EU, FC, AJ) and a secondary meta-analysis of all four subpopulations (EU, FC, AJ, CVCR/ANT). Heterogeneity was assessed using Cochran's Q test and I^2 statistics.[9]

eQTL and mQTL enrichment tests

Expression quantitative trait loci (eQTL) data from lymphoblast cell lines (LCLs) were generated from 176 HapMap CEU and YRI cell lines as described previously.[10] Cerebellar cis- and trans-eQTLs were generated from 153 individuals of European ancestry obtained from the Stanley Medical Research Institute.[11] Similarly, cerebellar cis- and trans- methylation QTLs (mQTLs), which represent SNPs associated with variation in genome-wide patterns of methylation, were generated from the same 153 European ancestry individuals from the Stanley Medical Research Institute. Data on age, gender, brain pH, smoking and alcohol use, suicide status, and postmortem interval (PMI) were collected. SNP genotyping was performed on the Affymetrix GeneChip Mapping 5.0 K Array (Affymetrix, Santa Clara, CA, USA). The Affymetrix Human Gene 1.0 ST Array was used for gene expression profiling. Methylation profiling in the cerebellum samples was conducted using the Illumina HumanMethylation27 BeadChips array (Illumina, San Diego, CA, USA). Significance cutoffs were based on the estimated numbers of probes available for each analysis. For cis- eQTLs and mQTLs, defined as SNPs within 4 Mb of the probe site, the significance threshold was set at 0.001, based on the number of cis-probes tested for each SNP. For trans-QTLs, the significance threshold was corrected as 0.05/n, where n= 25 834 probes for the expression data and n= 8 597 for the methylation data. Frontal cortex cis-eQTL data were derived from frozen post-mortem brain tissue from 399 neurologically normal European ancestry subjects from the United States and Great Britain. Detailed methods are described previously based on data from the first 150 of these subjects.[12] Genotyping was performed on the Illumina HumanHap550 genotyping array (Illumina, San Diego, CA, USA). Expression analysis was performed using the Illumina HumanHT-12 v3 chip (Illumina, San Diego, CA, USA). SNP genotypes were imputed in MACH[13] using the March 2010 1000 Genomes CEU phased data. Expression probes were adjusted for known covariates as previously described.[12] Cis-eQTL analysis was performed using mach2qtl with cis-eQTLs defined as loci within 1 Mb of the probe.

Each of the SNPs with $p<1.0 \times 10^{-3}$ in the primary European-derived GWAS meta-analysis was subsequently annotated with eQTL and mQTL information, including the strength of the evidence for the impact of the

polymorphism on expression and methylation. To test for enrichment of eQTLs or mQTLs among these top SNP associations, one thousand randomized SNP sets were generated, each of the same size as the original list of the top associations, and each containing SNPs matching the minor allele frequency distribution of the top association SNP set and sampled without replacement from the set of typed SNPs on the Illumina HumanHap550 array. Minor allele frequency matching was conducted by classifying all Illumina 550K platform SNPs into discrete minor allele frequency bins at 5% intervals (0-5%, 5-10%, …,45-50%), followed by random selection of SNPs from the same allele-frequency bins as those in the top signals. All SNP sets (actual and simulated) were pruned for LD as previously described.[10] The number of eQTLs (or mQTLs) in each simulated set yields an empirical distribution and an enrichment p-value, calculated as the proportion of randomized sets in which the eQTL (or mQTL) count matches or exceeds the actual observed count in the list of top SNP associations.

Gene and coding SNP enrichment tests

Each polymorphic SNP in HapMap was assigned a function, following the dbSNP functional classification scheme, as previously described.[14] Briefly, a SNP was considered "genic" if it was located either within a coding region, intron or 2 kb of upstream or downstream flanking sequences. Coding SNPs were assigned a function depending on how each allele altered the translated amino acid sequence. If either allele is nonsynonymous, it was assigned a "missense," "nonsense," or "frameshift" annotation. To test for enrichment of genic SNPs and specifically for missense polymorphisms among the top GWAS SNPs with $p<0.001$ in the primary meta-analysis, a similar approach to that applied for eQTL enrichment was used (as described above).

Evaluation of previously reported TS candidate genes

A literature review was conducted to select TS candidate genes that were reported to have association p-values <0.05 in prior studies (**Table S6**).[15-28]

Gene Name	Gene	Type of Study	Ref #
5HT2A	Serotonin receptor 2A	Association	22
BTBD9	BTB/POZ domain-containing protein 9	Association	26
DBH	Dopamine β-hydroxylase	Association	18
DLGAP3 (SAPAP3)	Discs large-associated protein 3	Association	23
DRD1	Dopamine receptor D1	Association	20
DRD2	Dopamine Receptor D_2	Association	15
DRD3	Dopamine Receptor D_3	Association	16
DRD4	Dopamine Receptor D_4	Association	17
IL1RN	interleukin 1 receptor antagonist	Association	24
MOG	Myelin oligodendrocyte glycoprotein	Association	21
MAOA	Monoamine oxidase A	Association	27
SLC6A3	Dopamine-associated transporter	Association	19
AADAC	arylacetamide deacetylase	CNV analysis	32
CTNNA3	alpha 3 catenin	CNV analysis	32
FSCB	Fibrous sheath CABYR binding protein	CNV analysis	32
NRXN1	Neurexin 1	CNV analysis	32
NLGN4X	Neurolignin 4X	Chromosomal rearrangement	28
CNTNAP2	Contactin 2 associated protein 2	Chromosomal rearrangement	34
IMPP2L	Inner mitochondrial membrane protein 2L	Chromosomal rearrangement	31
SLITRK1	SLIT and NTRK-like family, member 1	Chromosomal rearrangement, sequencing and association	25, 29
DNAJC13	DnaJ (Hsp40) homolog, subfamily C, member 13	Exome sequencing	33
MRPL3	Mitochondrial ribosomal protein L3	Exome sequencing	33
OFCC1	Orofacial cleft 1 candidate 1	Exome sequencing	33
HDC	Histamine decarboxylase	Linkage and sequencing	30

Supplementary Table S6. List of previously reported TS candidate genes.

In addition, this list was supplemented by genes identified in TS chromosomal translocation breakpoint studies, copy-number variant analyses as well as recent linkage and sequencing studies.[25, 29-34] All 2135 SNPs in the primary European ancestry GWAS meta-analysis that were contained within 50 kb of each candidate gene based on hg19 coordinates were extracted in PLINK, and the SNP with the lowest p-value within each gene region was selected (**Supplementary Table S7**).

IMPUTATION

Pre-imputation QC

Prior to genotype imputation, additional quality control measures were conducted to ensure that the GWAS dataset was compatible with the reference data set (1000 Genomes June 2011 Data Release)[35] and that only the highest quality SNPs were included in the imputation so as not to adversely affect imputation accuracy. SNP positions were converted to NCBI Build 37 (GRCh37) using an Illumina-provided mapping file (Human1M-Duov3_B-H_MappingInformation.txt), and 211 SNPs were removed without updated mapping information. For each pair of Illumina genotyping platforms used in the study (1M, 610Quad,550v1, 550v3, 370K), SNP-level concordances were calculated using genotype data from HapMap2 samples previously genotyped on each of these platforms (www.icom, Illumina.com, Illumina, San Diego, CA, USA). For any pair of platforms, any SNP with a concordance <95% and >1 genotype mismatch in any one HapMap population, or <95% concordance in the combined populations were removed. In addition, SNPs with a call rate <80% in any combined HapMap population data set were removed. Additionally, SNP-level concordances with the 1000 Genomes June 2011 Data Release were calculated using Illumina-genotyped samples for the same HapMap population samples. The same criteria for the pairwise genotyping platform comparisons above were used to remove SNPs, with the exception that removed SNPs were also excluded from being re-imputed. Finally, SNPs with a HWE $p<10^{-5}$ in controls were removed (previously flagged in the main GWAS analyses) and excluded from being re-imputed.

Imputation

Imputation of genotypes was conducted using the IMPUTE2 software program (version 2.1.2)[36] and haplotypes from all 1,092 individuals in the 1000 Genomes June 2011 Data Release[35] as a reference dataset. Imputation was run separately for individuals of European and Latin-American ancestry, so that different values of the k_hap parameter could be used for each set. To facilitate parallel computation, the genome was divided into 1Mb chunks or chunks containing 10,000 reference panel SNPs, whichever was smaller. As no C/G or A/T SNPs were present in the study set, we allowed the software to automatically match the strand of the study

and reference set genotypes. Default program settings were used except for the following parameters: Ne=20000, iter=30, k=80, hap_specific_family, fix_strand_g, pgs_miss, k_hap=1038 (European ancestry samples), k_hap=1942 (Latin-American samples).

Post-imputation QC

Following imputation, the distribution of various quality score metrics of imputed SNPs were first examined within all European-derived ancestry (EU, AJ, FC) case-control samples together. SNPs were excluded for having an IMPUTE2 info score <0.5 or an IMPUTE2 certainty score <0.9. CVCR/ANT samples were examined separately and SNPs excluded at the same QC thresholds (IMPUTE2 info score <0.5 or certainty <0.9). Samples were then separated into genotyping platforms of origin, and imputed SNPs with a PLINK info<0.6 in any individual platform were excluded. Allele frequencies of imputed SNPs were also compared across pairs of platforms, where platform of origin was used as the phenotype for analysis (e.g., 610Quad cases vs 370K cases, or 610Quad controls vs. 550K controls). SNPs with cross-platform analysis p-values $<1x10^{-5}$ for SNPs with MAF>0.05 or $p<1x10^{-4}$ for SNPs with MAF≤0.05 were flagged. Of note, of the 1650 SNPs flagged at this step, only 8 SNPs had subsequent association p-values $<1x10^{-3}$ in either the European ancestry or global ancestry meta-analyses; none of these 8 SNPs had $p<1x10^{-4}$.

Third, subpopulation-specific QC of imputed SNPs was conducted, and SNPs were removed for MAF<0.01, PLINK info score <0.1, or a Hardy-Weinberg p-value $<10^{-5}$. Given the size of the EU sample relative to that of the other subpopulations, any imputed SNP that failed QC in the EU stratum was also excluded from the other three groups (AJ, FC, CVCR/ANT). Lastly, imputed SNPs were removed based on the presence of >4 Mendelian errors (>1%) in a parallel imputation of 400 OCD trios genotyped on the Illumina 610Quad simultaneously with the TS case-control sample. For the trio samples, genotype probabilities were converted to "best call" genotypes using GTOOL[37]; if no genotype probability was greater than 0.9, the "best call" genotype was set to missing. After all post-imputation QC, 6,955,061 imputed SNPs remained for dosage analysis.

Dosage Analysis

Allelic dosage analysis was performed in PLINK in each of the four subpopulations (EU, AJ, FC, CVCR/ANT) separately using logistic regression under an additive model incorporating the same MDS covariates as used for the analysis of genotyped data. Subpopulation-specific results were subsequently combined by case-weighted meta-analysis in METAL using a fixed-effects model.

SUPPLEMENTARY RESULTS

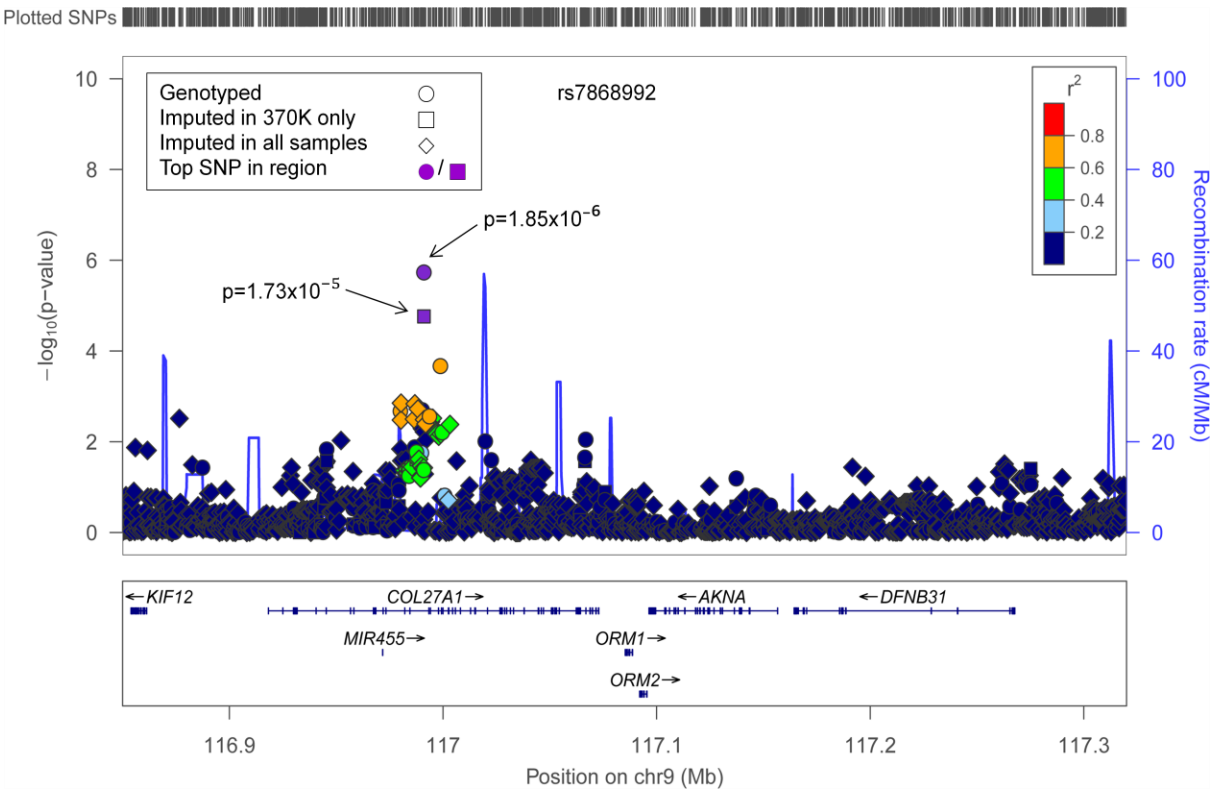
Population Isolate Results (FC, AJ, CVCR/ANT)

Analysis of population isolates has been proposed as an approach to reduce the genetic heterogeneity of traits with complex inheritance like TS.[38] Although this method has been extremely successful in identification of Mendelian disorders, it has proven to be more challenging for complex traits.[39] Given the small sample size of each TS population isolate, we opted not to consider results from each isolate population separately, but instead to perform a meta-analysis to identify common susceptibility alleles across all samples, i.e. those which arose prior to the separation of each isolate from the main outbred population. However, since these isolate-specific results may be instructive for future studies of TS in each subpopulation, the results of each isolate-specific analysis are provided in **Supplementary Table S4**.

Primary meta-analysis of European ancestry derived samples (EU, AJ, FC)

As mentioned in the main text, the primary European-ancestry meta-analysis produced 551 loci with association p-values $<1 \times 10^{-3}$ (**Table S2**). The complete list of these SNPs, with full annotation[14, 40, 41], including eQTL data from all three tissues (LCL, cerebellum, and frontal cortex) and cerebellar mQTL data are provided in **Table S2**. LocusZoom plots[42] and forest plots[43] of the top 5 loci in the primary analysis are shown in **Figures S9-S13.**

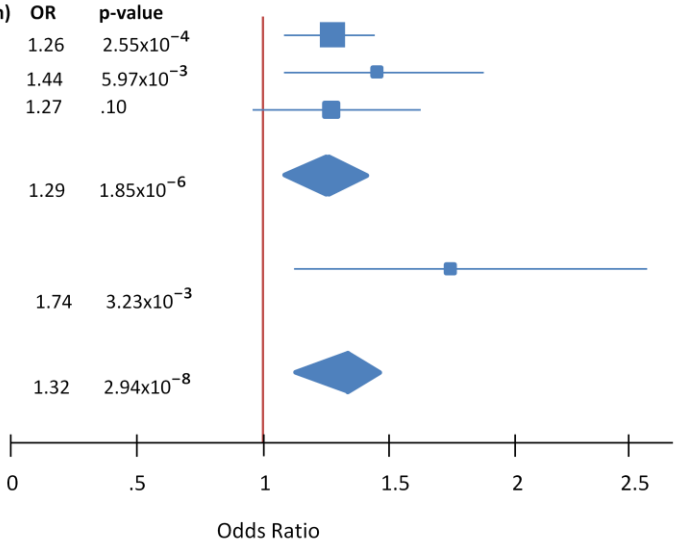
a)



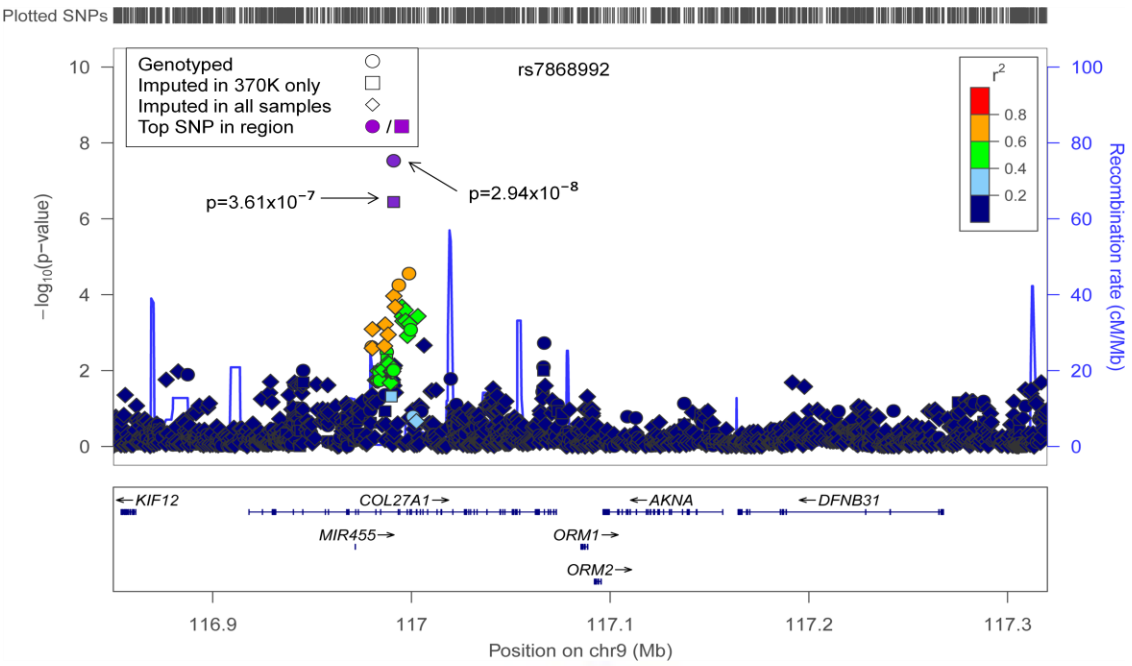
b)

rs7868992 **G/A**
Chr9:116030892

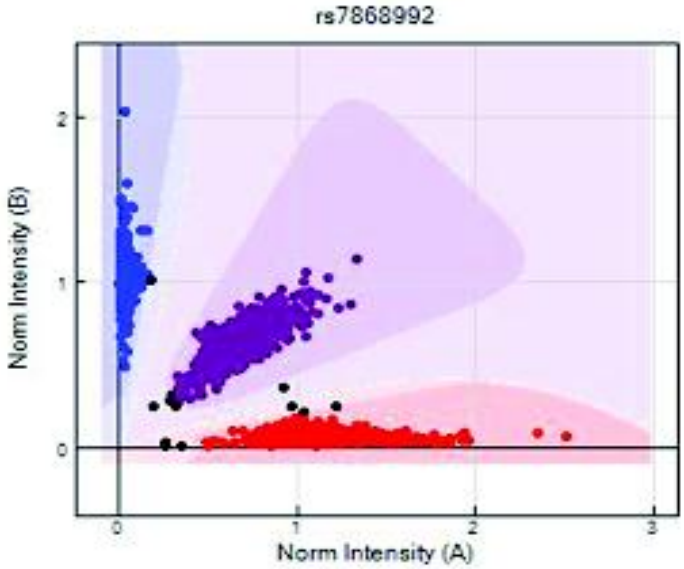
	Case MAF (n)	Control MAF (n)	OR	p-value
TS EU	0.327 (652)	0.281 (4401)	1.26	2.55×10^{-4}
TS AJ	0.377 (220)	0.304 (354)	1.44	5.97×10^{-3}
TS FC	0.326 (235)	0.277 (195)	1.27	.10
Primary European Ancestry Meta-Analysis	0.336 (1137)	0.283 (4950)	1.29	1.85×10^{-6}
Direction +++ **Het** p=.64; I^2= 0%				
TS CVCR/ANT	0.337 (211)	0.349 (285)	1.74	3.23×10^{-3}
Secondary Global Ancestry Meta-Analysis	0.343 (1348)	0.286 (5235)	1.32	2.94×10^{-8}
Direction ++++ **Het** p=.67; I^2= 0%				



c)

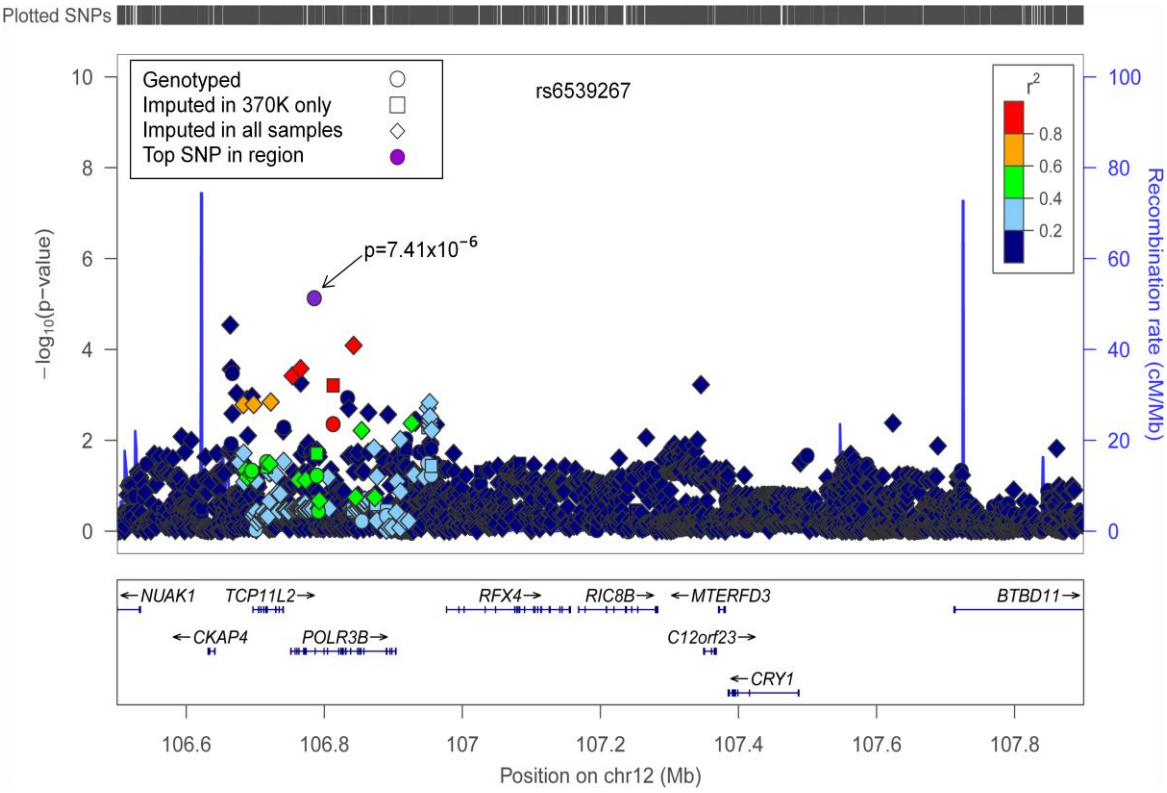


d)

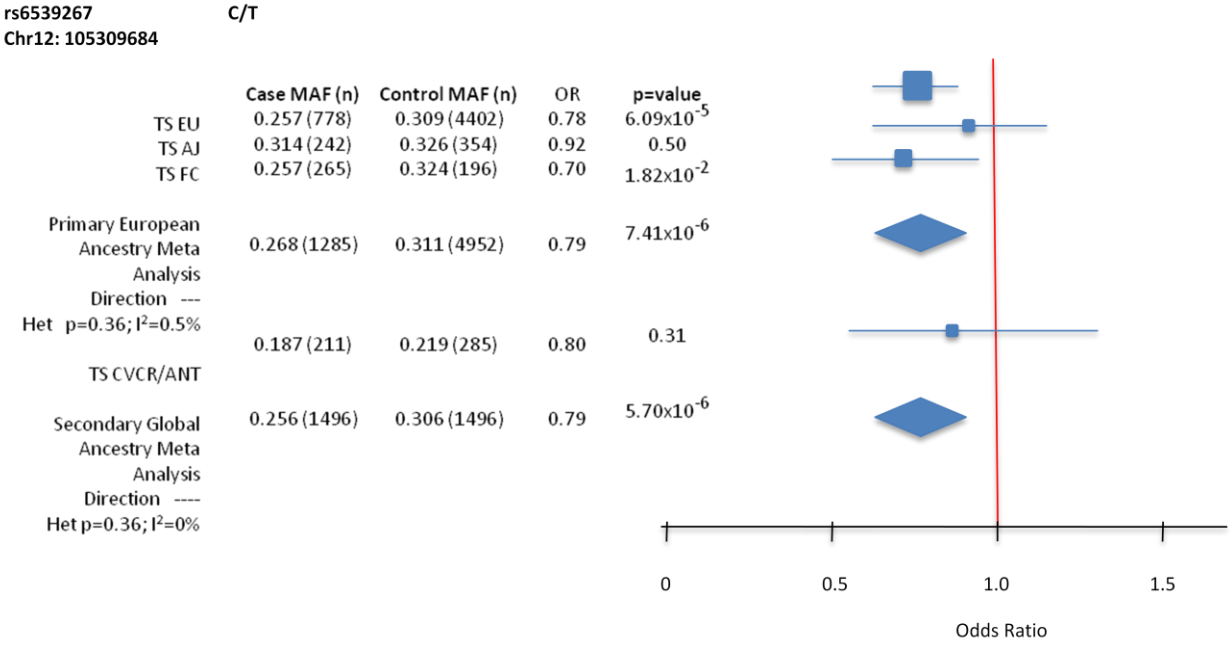


Supplementary Figure S9: Regional association and forest plots of rs7868992 on chromosome 9q32 containing the top signal in both the primary and secondary meta-analyses. a) LocusZoom regional association plot from the primary European ancestry meta-analysis including imputed data from the 1000 Genomes Project. Circles depict genotyped SNPs. For SNPs genotyped on the 4 major study platforms, but not present on the Illumina 370K, there are two symbols per SNP: a circle representing the signal from genotyped data only, and a square representing the signal incorporating imputed data from the 148 samples genotyped on the 370K; diamonds depict SNPs imputed in all samples. Red, orange, green and blue colors indicate the r^2 (derived from 1000 Genomes CEU data) between each plotted SNP and the top SNP in the region (i.e, rs7868992, in purple). Blue lines indicate the estimated recombination rate from HapMap release 22; **b)** Forest plot of rs7868992, adapted from Ripke et al[43]. Blue boxes indicate the odds ratio point estimate in each subpopulation-specific analysis; box size is proportional to the number of cases (i.e., the weighting factor in each meta-analysis). Blue lines denote the 95% CI. Diamonds indicate the 95% CI of the primary (top) and secondary (bottom) meta-analyses. MAF, minor allele frequency; n, number of samples in each analysis. Het p, p-value of Cochran's Q heterogeneity test; I^2 statistic, % variation attributed to heterogeneity; **c)** LocusZoom plot from the secondary meta-analysis of all samples; **d)** Normalized intensity plot of SNP rs7868992 genotype clusters from BeadStudio (Illumina, San Diego, CA, USA).

a)

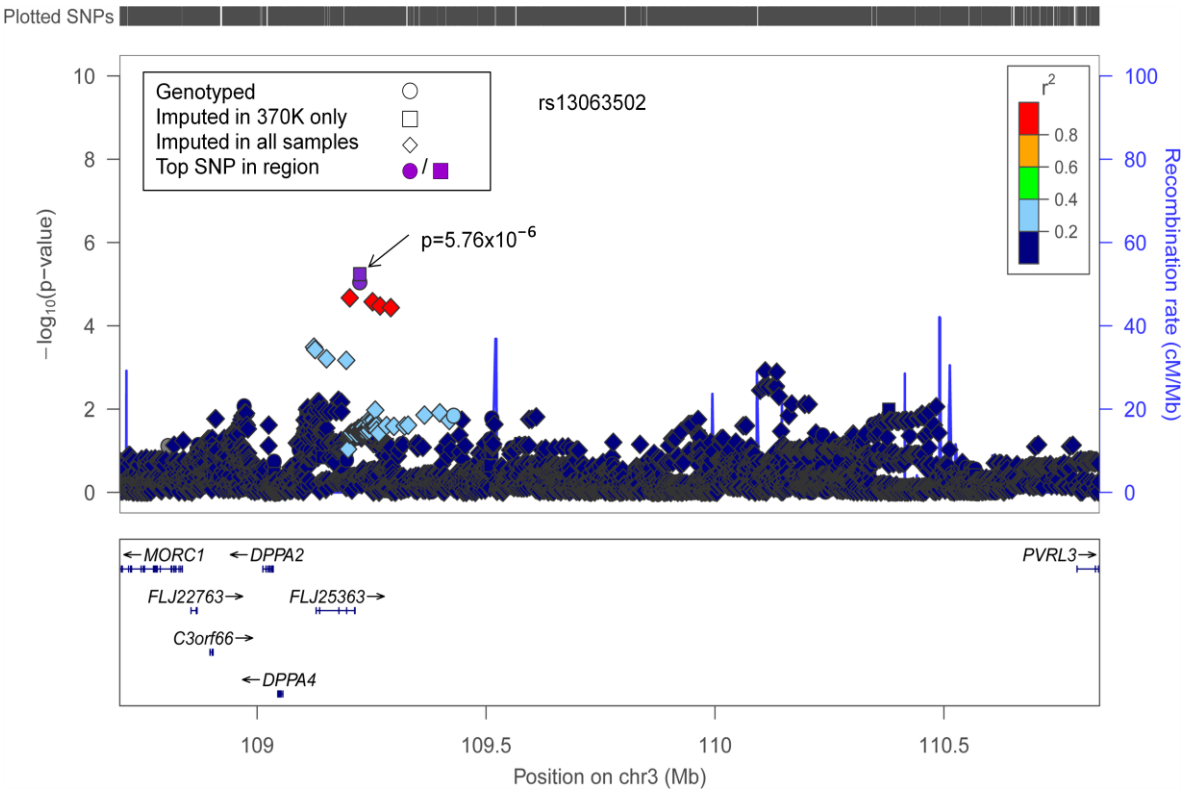


b)

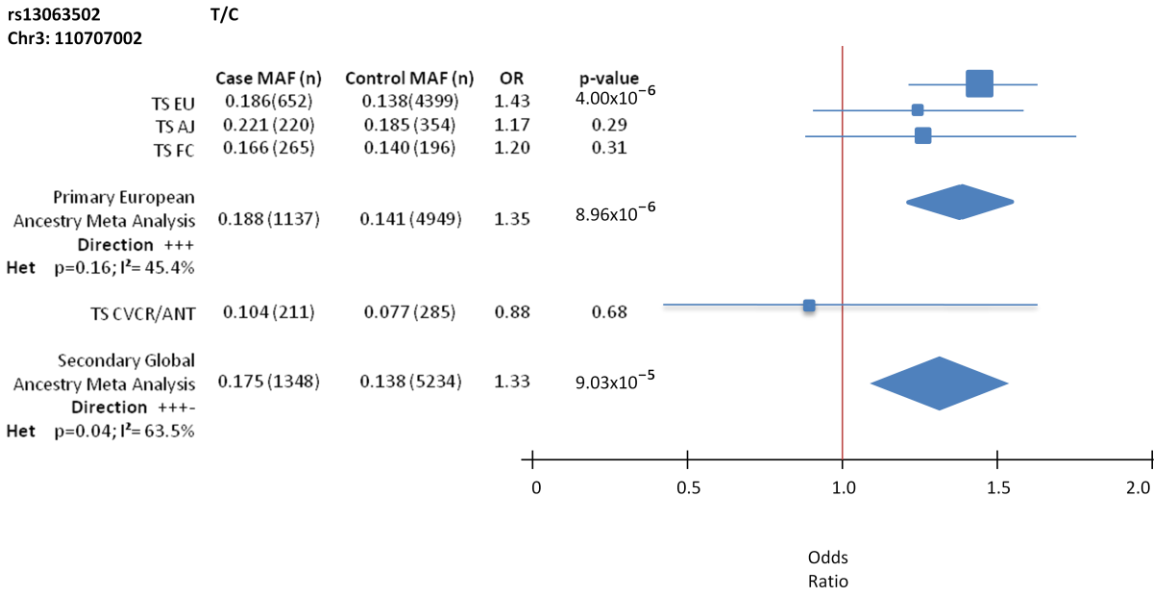


Supplementary Figure S10: Regional association and forest plots of rs6539267 on chromosome 12q23. a) LocusZoom regional association plot from the primary European ancestry meta-analysis including imputed data. Arrow indicates the top SNP in the region, rs6539267 (in purple). **b)** Forest plot and heterogeneity tests for rs6539267.

a)

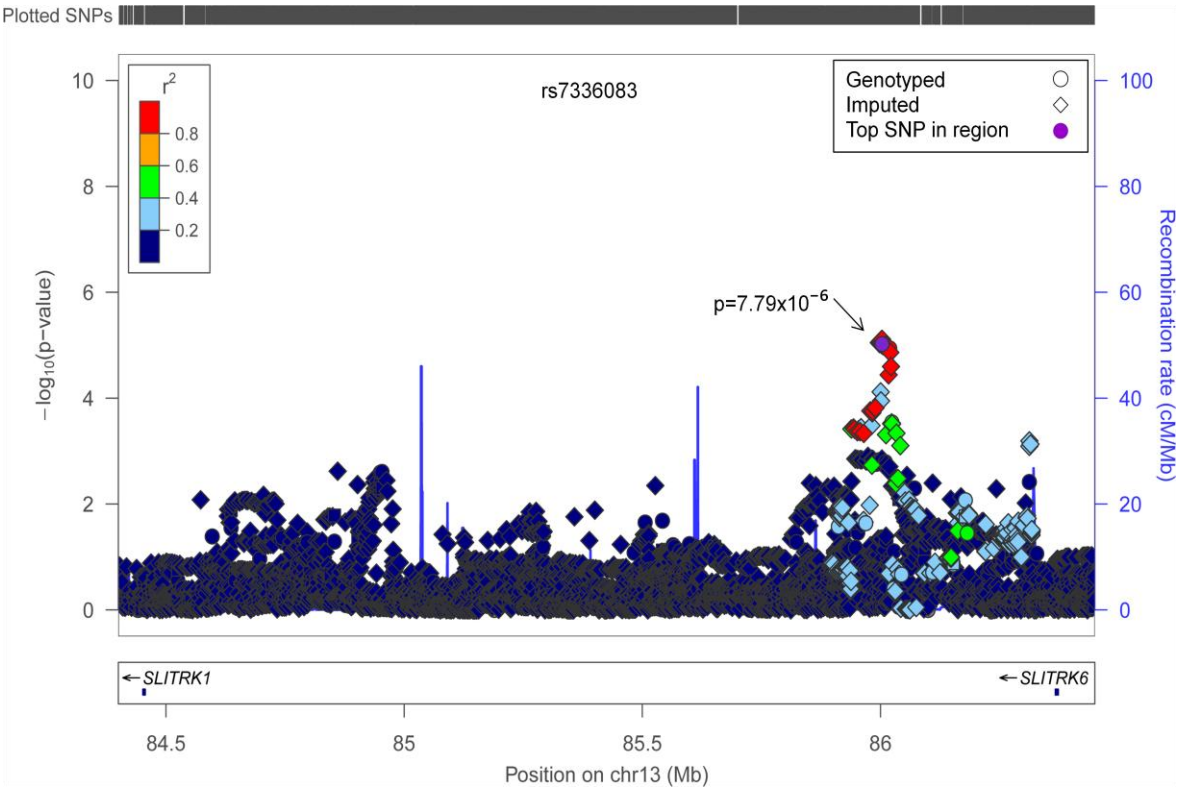


b)

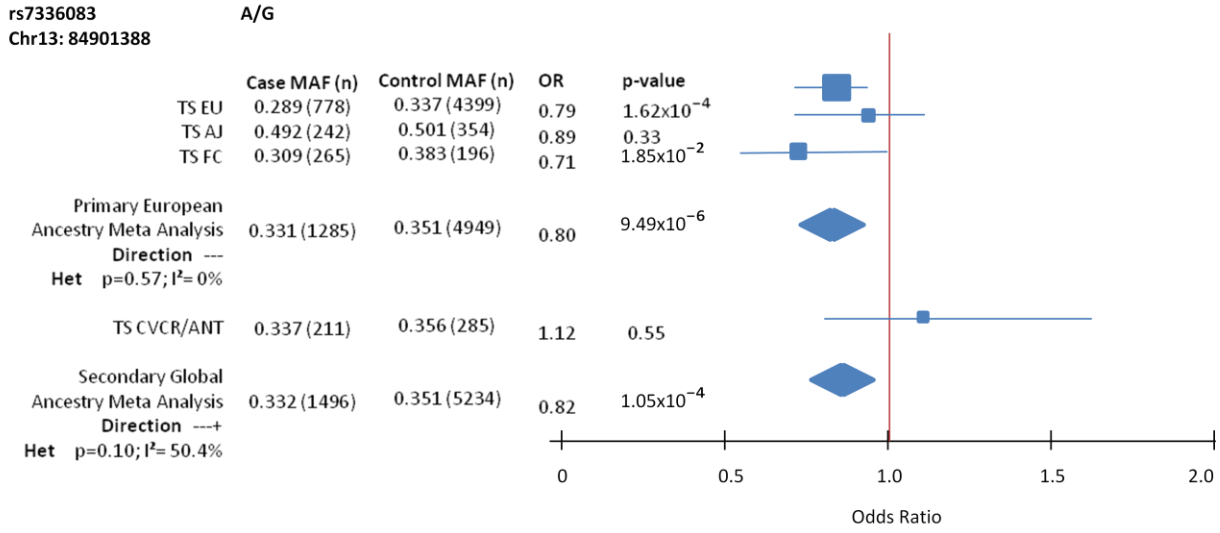


Supplementary Figure S11: Regional association and forest plots of rs13063502 on chromosome 3q13. a) LocusZoom regional association plot from the primary European ancestry meta-analysis including imputed data. Arrow indicates the top SNP in the region, rs13063502 (in purple); **b)** Forest plot and heterogeneity tests for rs13063502.

a)

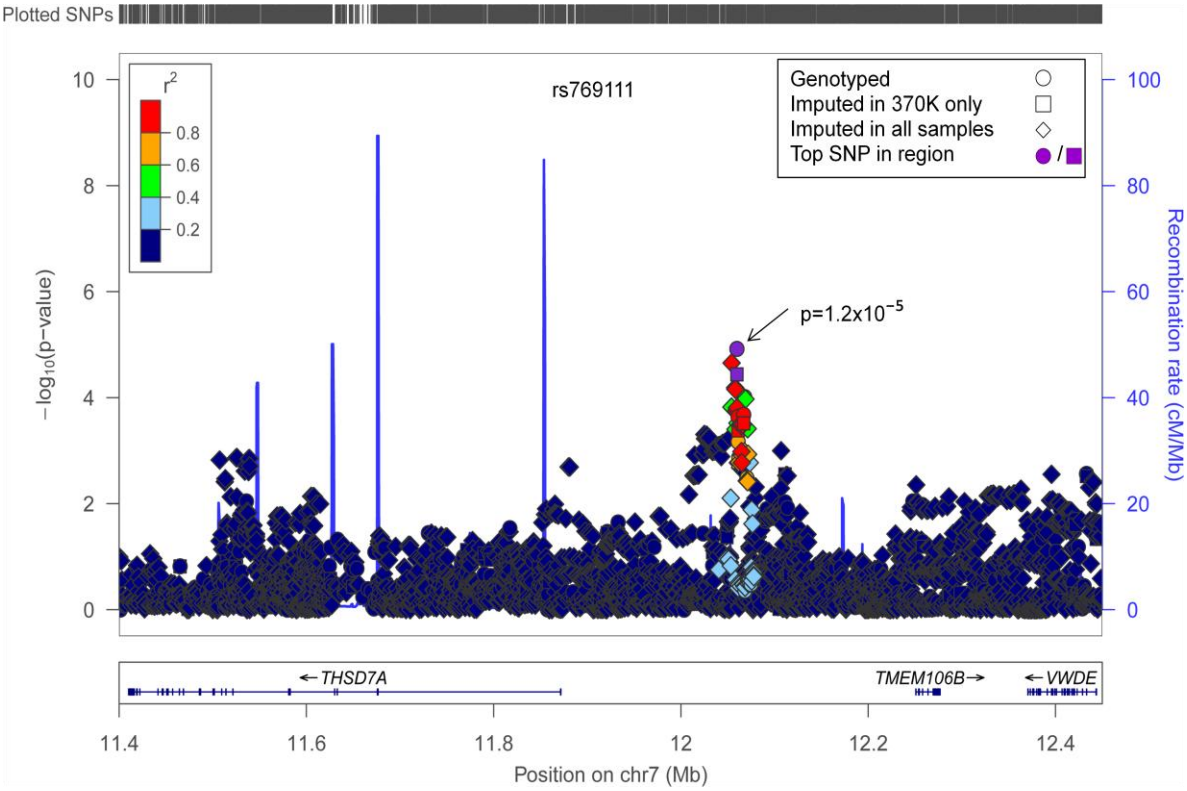


b)

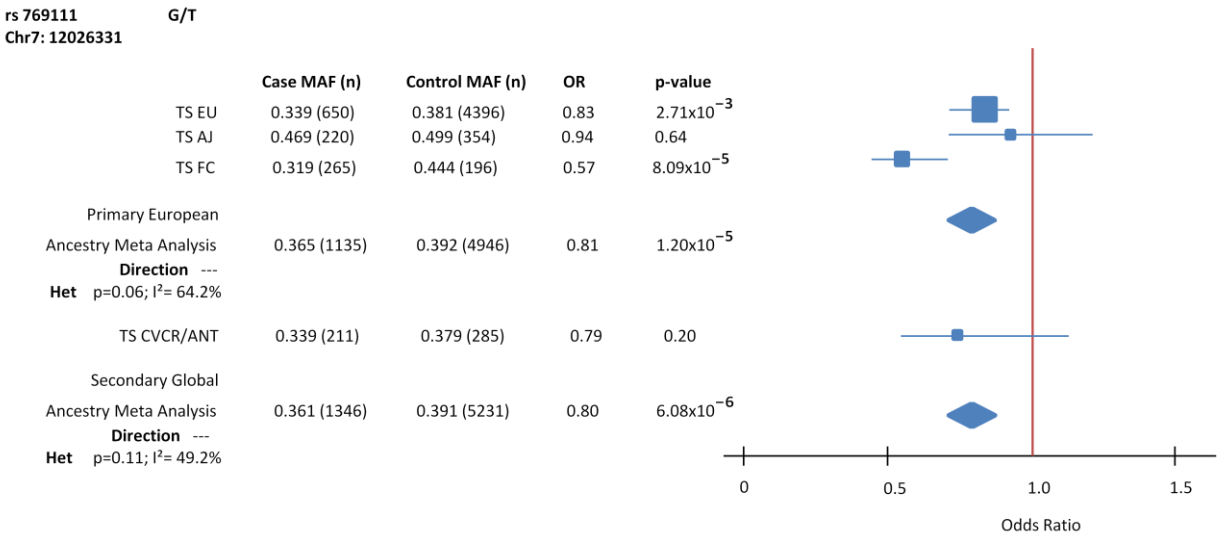


Supplementary Figure S12: Regional association and forest plots of rs7336083 on chromosome 13q31. a) LocusZoom regional association plot from the primary European ancestry meta-analysis including imputed data. Arrow indicates the top SNP in the region, rs7336083 (in purple); **b)** Forest plot and heterogeneity tests for rs7336083.

a)



b)



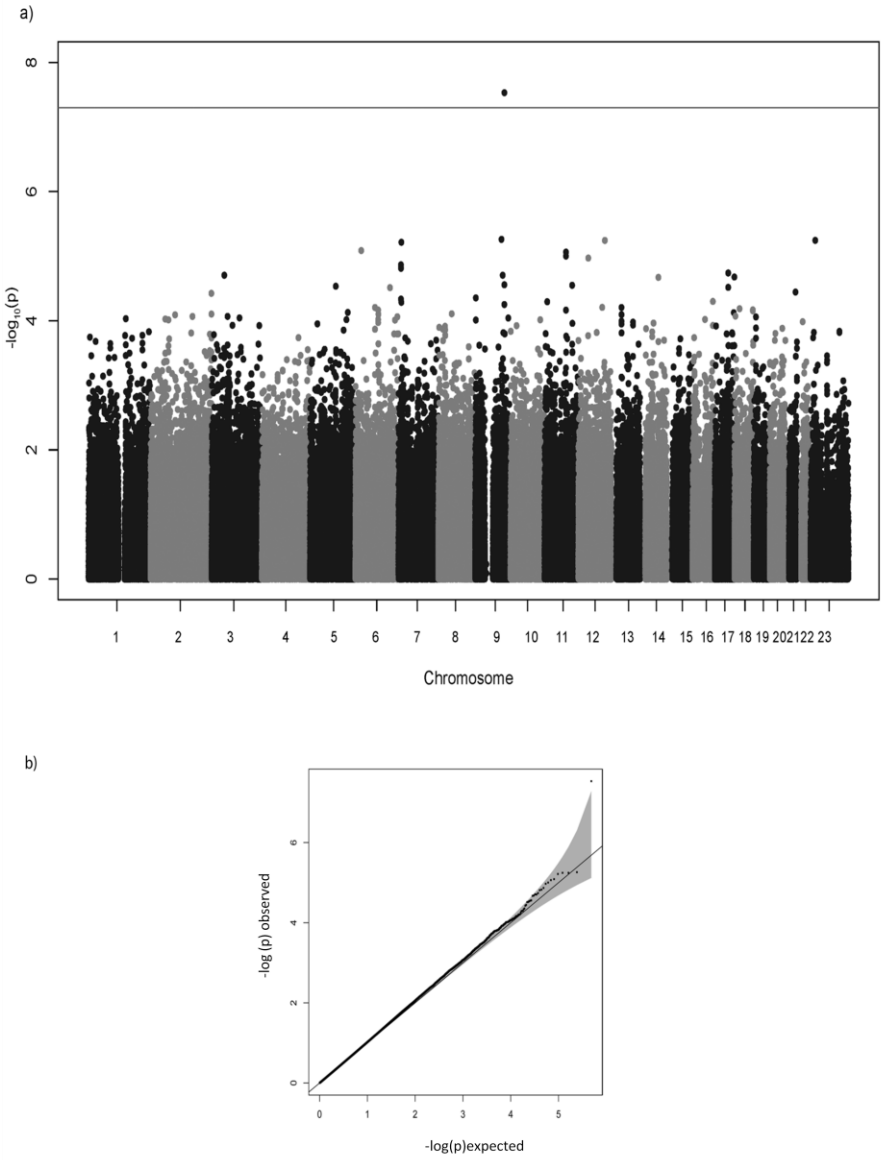
Supplementary Figure S13: Regional association and forest plots of rs769111 on chromosome 7p21. a) LocusZoom regional association plot from the primary European ancestry meta-analysis including imputed data. Arrow indicates the top SNP in the region, rs769111 (in purple); **b)** Forest plot and heterogeneity tests for rs769111.

X-Chromosome analysis

9647 X chromosome SNPs passed QC and were included in the association analysis. No X chromosome SNPs reached genome-wide significance; the strongest signal in the primary meta-analysis was found in rs5951698 ($p=7.7x10^{-5}$) within *PHEX*. In the male-specific analysis, there were no noteworthy signals in exonic variants or in LD with exonic variants (lowest p-value=0.01). All X-chromosome SNPs with $p<1x10^{-3}$ are provided in Tables **S2-S4**.

Secondary meta-analysis of all TS case-control samples (EU, AJ, FC, CVCR/ANT)

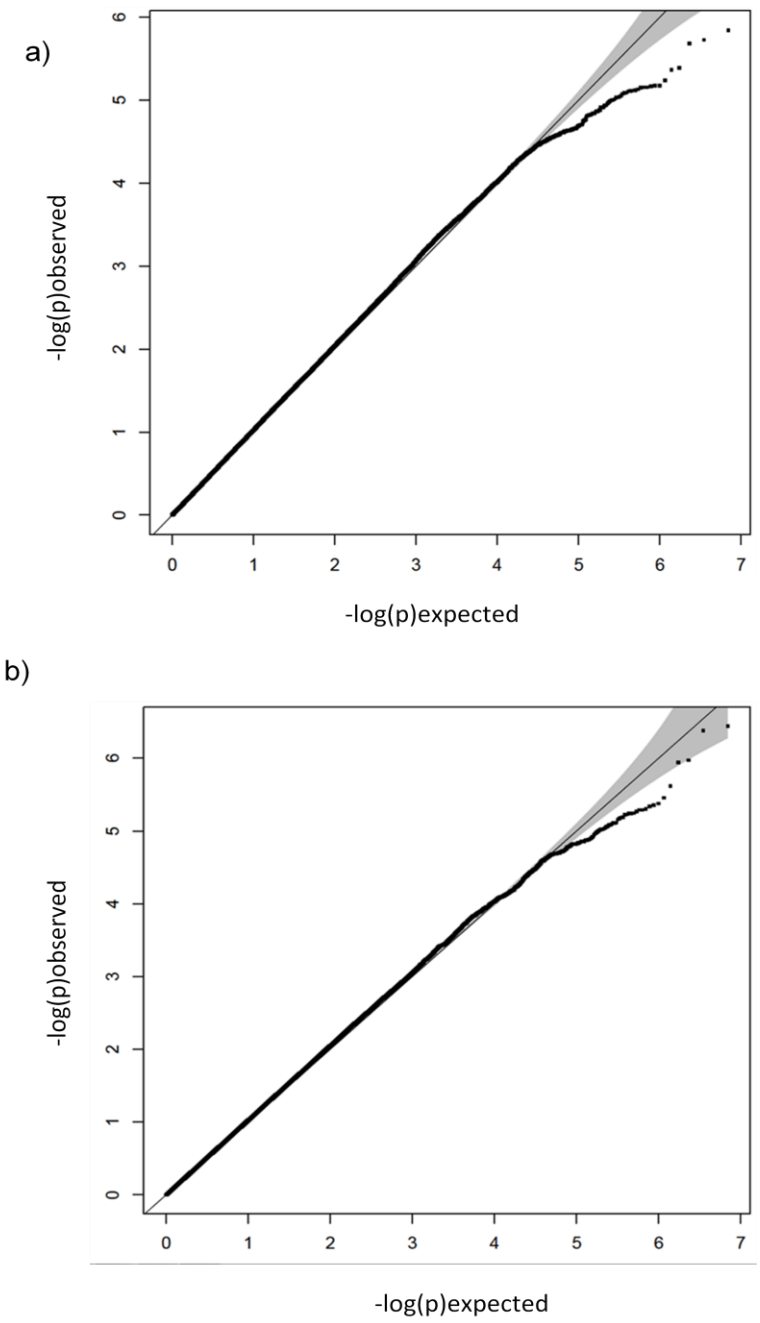
The secondary global meta-analysis of all four TS subpopulations consisting of 1496 TS cases and 5249 controls produced 560 loci with association p-values $<10^{-3}$ (**Figure S14**; complete annotated list provided in **Table S3**). The top signal in the primary European ancestry meta-analysis, rs7868992, was also the highest signal in the secondary meta-analysis of all four TS subpopulations (**Figure S9c**). Examination of the cluster plot for rs7868992 demonstrated no evidence of a genotype calling artifact as an explanation for the high test statistic (**Figure S9d**)**.** The top 5 loci in the secondary meta-analysis are provided in **Table S5**.



Supplementary Figure S14: Results of the secondary meta-analysis of all TS GWAS samples. a) Quantile-quantile plot of observed vs. expected –log(p) values from the secondary meta-analysis of 1496 TS cases and 5249 controls from the EU, FC, AJ and CVCR/ANT populations. The 95% confidence interval of expected values is indicated in grey. The genomic control λ value is 1.012. b) Manhattan plot of all genotyped SNPs from the secondary meta-analysis. Grey line indicates the genome-wide significant threshold of $p=5 \times 10^{-8}$.

				Primary European Meta-analysis		CVCR/ANT		Secondary Global Meta-analysis		# of SNPs in LD	Annotation				
CHR	SNP	BP	A1/ A2	OR	p-value	OR	p-value	OR	p-value		Gene	Left Gene	Right Gene	eQTL	Cerebellar mQTL
9	rs7868992	116030892	G/A	1.29	1.85×10^{-6}	1.74	3.23×10^{-3}	1.32	2.94×10^{-8}	3	COL27A1 (intron)	KIF12	ORM1	-	SYTL4, AMBP, HSPC152, OAS2, PWP1, RALBP1
9	rs10990268	104542319	C/T	1.32	7.37×10^{-5}	1.61	0.02	1.34	5.49×10^{-6}	1	-	LOC 100127962	CYLC2	-	-
X	rs5951698	22134658	G/A	5.65	7.67×10^{-5}	3.25	4.86×10^{-3}	4.31	5.69×10^{-6}	0	PHEX (intron)	SMS	ZNF645	-	-
12	rs6539267	105309684	C/T	0.79	7.41×10^{-6}	0.80	0.31	0.79	5.70×10^{-6}	0	POLR3B (intron)	TCP11L2	FLJ45508	-	TMEM119
7	rs769111	12026331	G/T	0.81	1.20×10^{-5}	0.79	0.20	0.80	6.08×10^{-6}	2	-	THSD7A	TMEM106B	MEOX2 (cerebellum)	PLSCR1, PCDHB16

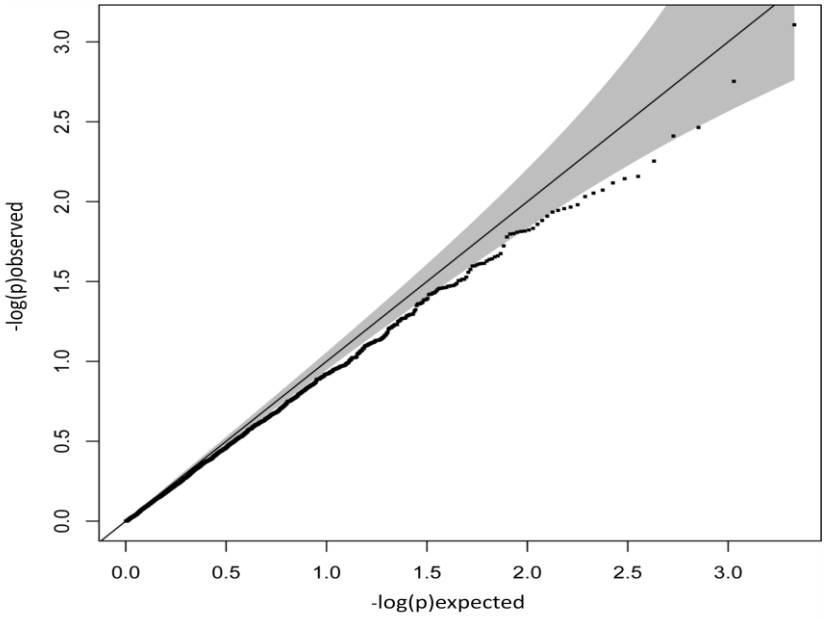
Supplementary Table S5: Top 5 LD-independent signals in the secondary global meta-analysis of all TS cases and controls. CHR, chromosome; BP, hg19 position; A1, reference allele; A2, alternative allele; OR, odds ratio; CVCR/ANT, Latin American TS population isolate samples from Central Valley Costa Rica and Antioquia,Colombia; [1]# SNPs in LD, number of additional SNPs in linkage disequilibrium (LD) with association p-values $<1 \times 10^{-3}$ in the secondary meta-analysis (LD defined as $r^2>0.5$). Complete annotation of all SNPs in the secondary meta-analysis with association p-values $<1 \times 10^{-3}$ are provided in Supplementary Table S3.



Supplementary Figure S15. Quantile-quantile plots of observed vs. expected -log (p) values for the dosage analysis of imputed data from the 1000 Genomes Project. The 95% confidence interval of expected values is indicated in grey. a) Primary meta-analysis of European ancestry samples (EU, AJ, FC), $\lambda = 1.031$; b) Secondary meta-analysis of all samples (EU, FC, AJ and CVCR/ANT), $\lambda = 1.033$.

Analysis previously reported TS candidate genes

24 genes previously reported as candidate TS susceptibility genes either in association, sequencing or copy-number variant studies were evaluated within the primary European-ancestry meta-analysis for evidence of association (**Figure S17; Table S7)**. A quantile-quantile plot of the 2135 SNPs lying within 50 kb of one of these 24 candidate genes demonstrated no deviation from the null distribution, indicating the absence of any strong GWAS signals among these candidate genes (**Figure S17**). While one gene, *CNTNAP2*, had an association p-value < $1x10^{-3}$ (rs10277969, $p=7.82x10^{-4}$), this gene locus is extremely large (>2.3 Mb) and contains 541 SNPs from the primary meta-analysis (Bonferroni corrected p-value for 266 LD-independent within-gene SNPs, p=0.21).



Supplementary Figure S16: Quantile-quantile plot of observed vs. expected -log (p) values in the primary GWAS meta-analysis for 2135 SNPs from prior TS candidate genes. The 95% confidence interval of expected values is indicated in grey.

Gene Name	# of SNPs	# of LD independent SNPs	Best SNP in Primary Meta-analysis	p in Primary European Meta-analysis	Distance from Gene (kb)
5HT2A (HTR2A)	60	32	rs7983914	0.04	-15.6
AADAC	26	19	rs9823831	0.08	38.4
BTBD9	113	61	rs7742915	0.01	0
CNTNAP2	542	266	rs10277969	7.82×10^{-4}	0
CTNNA3	416	196	rs17190635	8.85×10^{-3}	0
DBH	53	37	rs10993768	0.03	39.5
DLGAP3 (SAPAP3)	23	17	rs12076918	0.01	-14.5
DNAJC13	24	13	rs2270801	0.06	20.0
DRD1	35	22	rs265973	0.03	-7.0
DRD2	42	17	rs11214589	0.02	-35.3
DRD3	36	17	rs7631540	0.07	-17.0
DRD4	14	10	rs936469	0.03	-30.6
FSCB	17	8	rs12892077	7.63×10^{-3}	-48.7
HDC	26	15	rs7166052	0.02	-1.7
IL1RN	40	25	rs13030546	0.13	-35.5
IMMP2L	142	57	rs2190529	0.03	0
MAOA	8	6	rs3027449	0.11	43.2
MOG	29	14	rs29246	0.03	-36.7
MRPL3	23	13	rs1352108	5.58×10^{-3}	-44.3
NLGN4X	48	27	rs7055627	3.88×10^{-3}	0
NRXN1	291	133	rs9309197	9.30×10^{-3}	0
OFCC1	81	36	rs1206974	0.04	-49.3
SLC6A3	35	27	rs27056	0.02	28.4
SLITRK1	11	6	rs9593836	0.29	0

Supplementary Table S7. Evaluation of previously reported TS candidate genes. SNPs within 50kb of each candidate gene were selected for evaluation. The number of LD-independent SNPs within each locus was determined using an $r^2<0.5$ threshold between each SNP. The top SNP from the primary European ancestry meta-analysis is reported. Distance of this SNP from the specified candidate gene is listed based on hg19 coordinates.

SUPPLEMENTARY REFERENCES

1. Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF *et al.* Novel genes identified in a high-density genome wide association study for nicotine dependence. *Hum Mol Genet* 2007; **16**(1)**:** 24-35.

2. Bierut LJ, Saccone NL, Rice JP, Goate A, Foroud T, Edenberg H *et al.* Defining alcohol-related phenotypes in humans. The Collaborative Study on the Genetics of Alcoholism. *Alcohol Res Health* 2002; **26**(3)**:** 208-213.

3. Bierut LJ, Strickland JR, Thompson JR, Afful SE, Cottler LB. Drug use and dependence in cocaine dependent subjects, community-based individuals, and their siblings. *Drug Alcohol Depend* 2008; **95**(1-2)**:** 14-22.

4. Stefansson H, Ophoff RA, Steinberg S, Andreassen OA, Cichon S, Rujescu D *et al.* Common variants conferring risk of schizophrenia. *Nature* 2009; **460**(7256)**:** 744-747.

5. Sebastiani P, Solovieff N, Puca A, Hartley SW, Melista E, Andersen S *et al.* Genetic signatures of exceptional longevity in humans. *Science* 2010; **2010**.

6. Service S, DeYoung J, Karayiorgou M, Roos JL, Pretorious H, Bedoya G *et al.* Magnitude and distribution of linkage disequilibrium in population isolates and implications for genome-wide association studies. *Nat Genet* 2006; **38**(5)**:** 556-560.

7. Carvajal-Carmona LG, Ophoff R, Service S, Hartiala J, Molina J, Leon P *et al.* Genetic demography of Antioquia (Colombia) and the Central Valley of Costa Rica. *Hum Genet* 2003; **112**(5-6)**:** 534-541.

8. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26**(17)**:** 2190-2191.

9. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**(7414)**:** 557-560.

10. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* 2010; **6**(4)**:** e1000888.

11. Gamazon ER, Nicolae DL, Cox NJ. A study of CNVs as trait-associated polymorphisms and as expression quantitative trait loci. *PLoS Genet* 2011; **7**(2)**:** e1001292.

12. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL *et al.* Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet* 2010; **6**(5)**:** e1000952.

13. Li Q, Fallin MD, Louis TA, Lasseter VK, McGrath JA, Avramopoulos D *et al.* Detection of SNP-SNP interactions in trios of parents with schizophrenic children. *Genet Epidemiol* 2010; **34**(5)**:** 396-406.

14. Gamazon ER, Zhang W, Konkashbaev A, Duan S, Kistner EO, Nicolae DL *et al.* SCAN: SNP and copy number annotation. *Bioinformatics* 2010; **26**(2)**:** 259-262.

15. Comings DE, Comings BG, Muhleman D, Dietz G, Shahbahrami B, Tast D *et al.* The dopamine D2 receptor locus as a modifying gene in neuropsychiatric disorders. *JAMA* 1991; **266**(13)**:** 1793-1800.

16. Comings DE, Muhleman D, Dietz G, Dino M, LeGro R, Gade R. Association between Tourette's syndrome and homozygosity at the dopamine D3 receptor gene. *Lancet* 1993; **341**(8849)**:** 906.

17. Grice DE, Leckman JF, Pauls DL, Kurlan R, Kidd KK, Pakstis AJ *et al.* Linkage disequilibrium between an allele at the dopamine D4 receptor locus and Tourette syndrome, by the transmission-disequilibrium test. *Am J Hum Genet* 1996; **59**(3)**:** 644-652.

18. Comings DE, Wu S, Chiu C, Ring RH, Gade R, Ahn C *et al.* Polygenic inheritance of Tourette syndrome, stuttering, attention deficit hyperactivity, conduct, and oppositional defiant disorder: the additive and subtractive effect of the three dopaminergic genes--DRD2, D beta H, and DAT1. *American Journal of Medical Genetics* 1996; **67**(3)**:** 264-288.

19. Rowe DC, Stever C, Gard JM, Cleveland HH, Sanders ML, Abramowitz A *et al.* The relation of the dopamine transporter gene (DAT1) to symptoms of internalizing disorders in children. *Behav Genet* 1998; **28**(3)**:** 215-225.

20. Tarnok Z, Ronai Z, Gervai J, Kereszturi E, Gadoros J, Sasvari-Szekely M *et al.* Dopaminergic candidate genes in Tourette syndrome: association between tic severity and 3' UTR polymorphism of the dopamine transporter gene. *Am J Med Genet B Neuropsychiatr Genet* 2007; **144B**(7)**:** 900-905.

21. Huang Y, Li T, Wang Y, Ansar J, Lanting G, Liu X *et al.* Linkage disequilibrium analysis of polymorphisms in the gene for myelin oligodendrocyte glycoprotein in Tourette's syndrome patients from a Chinese sample. *Am J Med Genet B Neuropsychiatr Genet* 2004; **124**(1)**:** 76-80.

22. Huang Y, Liu X, Li T, Guo L, Sun X, Xiao X *et al.* [Cases-Control association study and transmission disequilibrium test of T102C polymorphism in 5HT2A and Tourette syndrome]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi* 2001; **18**(1)**:** 11-13.

23. Crane J, Fagerness J, Osiecki L, Gunnell B, Stewart SE, Pauls DL *et al.* Family-based genetic association study of DLGAP3 in Tourette Syndrome. *Am J Med Genet B Neuropsychiatr Genet* 2011; **156**(1)**:** 108-114.

24. Chou IC, Lin HC, Wang CH, Lin WD, Lee CC, Tsai CH *et al.* Polymorphisms of interleukin 1 gene IL1RN are associated with Tourette syndrome. *Pediatr Neurol* 2010; **42**(5)**:** 320-324.

25. Miranda DM, Wigg K, Kabia EM, Feng Y, Sandor P, Barr CL. Association of SLITRK1 to Gilles de la Tourette Syndrome. *Am J Med Genet B Neuropsychiatr Genet* 2009; **150B**(4)**:** 483-486.

26. Riviere JB, Xiong L, Levchenko A, St-Onge J, Gaspar C, Dion Y *et al.* Association of intronic variants of the BTBD9 gene with Tourette syndrome. *Arch Neurol* 2009; **66**(10)**:** 1267-1272.

27. Gade R, Muhleman D, Blake H, MacMurray J, Johnson P, Verde R *et al.* Correlation of length of VNTR alleles at the X-linked MAOA gene and phenotypic effect in Tourette syndrome and drug abuse. *Mol Psychiatry* 1998; **3**(1)**:** 50-60.

28. Lawson-Yuen A, Saldivar JS, Sommer S, Picker J. Familial deletion within NLGN4 associated with autism and Tourette syndrome. *Eur J Hum Genet* 2008; **16**(5)**:** 614-618.

29. Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM *et al.* Sequence variants in SLITRK1 are associated with Tourette's syndrome. *Science* 2005; **310**(5746)**:** 317-320.

30. Ercan-Sencicek AG, Stillman AA, Ghosh AK, Bilguvar K, O'Roak BJ, Mason CE *et al.* L-histidine decarboxylase and Tourette's syndrome. *N Engl J Med* 2010; **362**(20)**:** 1901-1908.

31. Petek E, Windpassinger C, Vincent JB, Cheung J, Boright AP, Scherer SW *et al.* Disruption of a novel gene (IMMP2L) by a breakpoint in 7q31 associated with Tourette syndrome. *American Journal of Human Genetics* 2001; **68**(4)**:** 848-858.

32. Sundaram SK, Huq AHM, Wilson BJ, Chugani HT. Tourette Syndrome is associated with recurrent exonic copy number variants. *Neurology* 2010**:** in press.

33. Sundaram SK, Huq AM, Sun Z, Yu W, Bennett L, Wilson BJ *et al.* Exome sequencing of a pedigree with tourette syndrome or chronic tic disorder. *Ann Neurol* 2011; **69**(5)**:** 901-904.

34. Verkerk AJ, Mathews CA, Joosse M, Eussen BH, Heutink P, Oostra BA. CNTNAP2 is disrupted in a family with Gilles de la Tourette syndrome and obsessive compulsive disorder. *Genomics* 2003; **82**(1)**:** 1-9.

35. 1000 Genomes Project. A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**(7319)**:** 1061-1073.

36. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; **5**(6)**:** e1000529.

37. GTOOL. http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html, 2011, Accessed Date Accessed 2011 Accessed.

38. Mathews CA, Reus VI, Bejarano J, Escamilla MA, Fournier E, Herrera LD *et al.* Genetic studies of neuropsychiatric disorders in Costa Rica: a model for the use of isolated populations. *Psychiatr Genet* 2004; **14**(1)**:** 13-23.

39. Lowe JK, Maller JB, Pe'er I, Neale BM, Salit J, Kenny EE *et al.* Genome-wide association studies in an isolated founder population from the Pacific Island of Kosrae. *PLoS Genet* 2009; **5**(2)**:** e1000365.

40. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**(3)**:** 559-575.

41. Saccone SF, Bolze R, Thomas P, Quan J, Mehta G, Deelman E *et al.* SPOT: a web-based tool for using biological databases to prioritize SNPs after a genome-wide association study. *Nucleic Acids Res* 2010; **38**(Web Server issue)**:** W201-209.

42. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010; **26**(18)**:** 2336-2337.

43. Ripke S, Sanders AR, Kendler KS, Levinson DF, Sklar P, Holmans PA *et al.* Genome-wide association study identifies five new schizophrenia loci. *Nat Genet* 2011; **43**(10)**:** 969-976.