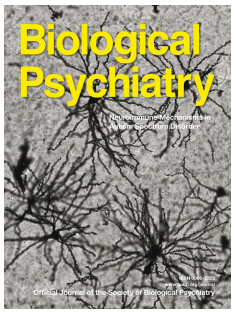
Journal Pre-proof



Genome-wide Association Study points to novel locus for Gilles de la Tourette Syndrome

Fotis Tsetsos, Apostolia Topaloudi, Pritesh Jain, Zhiyu Yang, Dongmei Yu, Petros Kolovos, Zeynep Tumer, Renata Rizzo, Andreas Hartmann, Christel Depienne, Yulia Worbe, Kirsten R. Müller-Vahl, Danielle C. Cath, Dorret I. Boomsma, Tomasz Wolanczyk, Cezary Zekanowski, Csaba Barta, Zsofia Nemoda, Zsanett Tarnok, Shanmukha S. Padmanabhuni, Joseph D. Buxbaum, Dorothy Grice, Jeffrey Glennon, Hreinn Stefansson, Bastian Hengerer, Evangelia Yannaki, John A. Stamatoyannopoulos, Noa Benaroya-Milshtein, Francesco Cardona, Tammy Hedderly, Isobel Heyman, Chaim Huyser, Pablo Mir, Astrid Morer, Norbert Mueller, Alexander Munchau, Kerstin J. Plessen, Cesare Porcelli, Veit Roessner, Susanne Walitza, Anette Schrag, Davide Martino, The PGC TS Working Group, The TSAICG, The TSGeneSEE initiative, The EMTICS collaborative group, The TS-EUROTRAIN network, The TIC Genetics collaborative group, Jay A. Tischfield, Gary A. Heiman, A. Jeremy Willsey, Andrea Dietrich, Lea K. Davis, James J. Crowley, Carol A. Mathews, Jeremiah M. Scharf, Marianthi Georgitsi, Pieter J. Hoekstra, Peristera Paschou

PII: S0006-3223(23)00051-3

DOI: https://doi.org/10.1016/j.biopsych.2023.01.023

Reference: BPS 15107

To appear in: *Biological Psychiatry*

Received Date: 11 May 2022

Revised Date: 23 November 2022

Accepted Date: 24 January 2023

Please cite this article as: Tsetsos F., Topaloudi A., Jain P., Yang Z., Yu D., Kolovos P., Tumer Z., Rizzo R., Hartmann A., Depienne C., Worbe Y., Müller-Vahl K.R., Cath D.C., Boomsma D.I., Wolanczyk T., Zekanowski C., Barta C., Nemoda Z., Tarnok Z., Padmanabhuni S.S., Buxbaum J.D., Grice D., Glennon J., Stefansson H., Hengerer B., Yannaki E., Stamatoyannopoulos J.A., Benaroya-Milshtein N., Cardona F., Hedderly T., Heyman I., Huyser C., Mir P., Morer A., Mueller N., Munchau A., Plessen K.J, Porcelli C., Roessner V., Walitza S., Schrag A., Martino D., The PGC TS Working Group, The TSAICG, The TSAICG, Barr C.L., Batterson J.R., Berlin C., Budman C.L., Coppola G., Cox N.J., Darrow S., Dion Y., Freimer N.B., Grados M.A., Greenberg E., Hirschtritt M.E., Huang A.Y., Illmann C., King R.A., Kurlan R.,

Leckman J.F., Lyon G.J., Malaty I.A., McMahon W.M., Neale B.M., Okun M.S., Osiecki L., Robertson M.M., Rouleau G.A., Sandor P., Singer H.S., Smit J.H., Sul J.H., The TSGeneSEE initiative, The EMTICS collaborative group, The TS-EUROTRAIN network, & The TIC Genetics collaborative group, Genome-wide Association Study points to novel locus for Gilles de la Tourette Syndrome, *Biological Psychiatry* (2023), doi: https://doi.org/10.1016/j.biopsych.2023.01.023.

Genome-wide Association Study points to novel locus for Gilles de la Tourette Syndrome

Running title: Tourette Syndrome GWAS analysis

Keywords: Tourette Syndrome, meta-analysis, GWAS, NR2F1

Fotis Tsetsos[1**], Apostolia Topaloudi[2***], Pritesh Jain[2], Zhiyu Yang[2], Dongmei Yu[3,4], Petros Kolovos[1], Zeynep Tumer[5,6], Renata Rizzo[7], Andreas Hartmann[8], Christel Depienne[9], Yulia Worbe[10,11], Kirsten R. Müller-Vahl[12], Danielle C. Cath[13], Dorret I. Boomsma[14,15], Tomasz Wolanczyk[16], Cezary Zekanowski[17], Csaba Barta[18], Zsofia Nemoda[18], Zsanett Tarnok[19], Shanmukha S. Padmanabhuni[2], Joseph D. Buxbaum[20,21,22,23,24,25], Dorothy Grice[20,24,26,25], Jeffrey Glennon[27], Hreinn Stefansson[28], Bastian Hengerer[29], Evangelia Yannaki[30,31], John A. Stamatoyannopoulos[32,33,34], Noa Benaroya-Milshtein[35], Francesco Cardona[36], Tammy Hedderly[37], Isobel Heyman[38], Chaim Huyser[39,40], Pablo Mir[41,42], Astrid Morer[43,44,45], Norbert Mueller[46], Alexander Munchau[47], Kerstin J Plessen[48,49], Cesare Porcelli[50], Veit Roessner[51], Susanne Walitza[52], Anette Schrag[53], Davide Martino[54], The PGC TS Working Group, The TSAICG, The TSGeneSEE initiative, The EMTICS collaborative group, The TS-EUROTRAIN network, The TIC Genetics collaborative group, Jay A. Tischfield[55], Gary A. Heiman[55], A. Jeremy Willsey[56,57], Andrea Dietrich[58], Lea K. Davis[59,60], James J. Crowley[61,62,63], Carol A. Mathews[64], Jeremiah M. Scharf[3,65,66], Marianthi Georgitsi[1,67], Pieter J. Hoekstra*[58], Peristera Paschou*[2]

[1]Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis, Greece

[2]Department of Biological Sciences, Purdue University, West Lafayette, IN, USA

[3]Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA

[4]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[5]Department of Clinical Genetics, Kennedy Center, Copenhagen University Hospital, Rigshospitalet, Denmark

[6]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen

[7]Child and Adolescent Neurology and Psychiatry, Department of Clinical and Experimental Medicine, University of Catania, Catania, Italy

[8]Department of Neurology, Hôpital de la Pitié-Salpêtrière, Paris, France

[9]Institute for Human Genetics, University Hospital Essen, Essen, Germany

[10]Assistance Publique Hôpitaux de Paris, Hopital Saint Antoine, Paris France

[11]French Reference Centre for Gilles de la Tourette Syndrome, Groupe Hospitalier Pitié-Salpêtrière, Paris, France

[12]Department of Psychiatry, Social psychiatry and Psychotherapy, Hannover Medical School, Hannover, Germany

[13]Department of Clinical and health Psychology, Utrecht University, Utrecht, Netherlands

[14]Institute for Anatomy and Cell Biology, Ulm University, Ulm, Germany

[15]EMGO+ Institute for Health and Care Research, VU University Medical Centre, Amsterdam, Netherlands

[16]Department of Child Psychiatry, Medical University of Warsaw, Warsaw, Poland

[17]Laboratory of Neurogenetics, Department of Neurodegenerative Disorders, Mossakowski Medical Research Institute, Polish Academy of Sciences, Warsaw, Poland

[18]Department of Molecular Biology, Institute of Biochemistry and Molecular Biology, Semmelweis University, Budapest, Hungary

[19]Vadaskert Clinic for Child and Adolescent Psychiatry, Hungary

[20]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, USA

[21]Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, USA

[22]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, USA

[23]Department of Neuroscience, Icahn School of Medicine at Mount Sinai, USA

[24]The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, USA

[25]Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, USA

[26]Division of Tics, OCD, and Related Disorders, Icahn School of Medicine at Mount Sinai, USA

[27]Department of Cognitive Neuroscience, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Netherlands

[28]deCODE Genetics/Amgen, Iceland

[29]Boehringer Ingelheim Pharma GmbH & Co. KG, CNS Research, Germany

[30]Hematology Department- Hematopoietic Cell Transplantation Unit, Gene and Cell Therapy Center, George Papanikolaou Hospital, Greece

[31]Department of Medicine, University of Washington, WA, USA

[32]Altius Institute for Biomedical Sciences, WA, USA

[33]Department of Genome Sciences, University of Washington, WA, USA

[34]Department of Medicine, Division of Oncology, University of Washington, WA, USA

[35]Child and Adolescent Psychiatry Department, Schneider Children's Medical Centre of Israel, Petah-Tikva. Affiliated to Sackler Faculty of Medicine, Tel Aviv University, Israel

[36]Department of Human Neurosciences, University La Sapienza of Rome, Rome, Italy

[37]Evelina London Children's Hospital GSTT, Kings Health Partners AHSC, London, UK

[38]Psychological Medicine, Great Ormond Street Hospital NHS Foundation Trust, Great Ormond Street, London, UK

[39]Levvel, Academic Center for Child and Adolescent Psychiatry, Amsterdam, The Netherlands

[40]Amsterdam UMC, Department of Child and Adolescent Psychiatry, Amsterdam, The Netherlands

[41]Unidad de Trastornos del Movimiento. Instituto de Biomedicina de Sevilla (IBiS). Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla. Seville, Spain

[42]Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Madrid, Spain

[43]Department of Child and Adolescent Psychiatry and Psychology, Institute of Neurosciences, Hospital Clinic Universitari, Barcelona, Spain

[44]Institut d'Investigacions Biomediques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

[45]Centro de Investigacion en Red de Salud Mental (CIBERSAM), Instituto Carlos III, Spain

[46]Department of Psychiatry and Psychotherapy, University Hospital, LMU Munich, Munich, Germany

[47]Institute of Systems Motor Science, University of Lübeck, Lübeck, Germany

[48]Child and Adolescent Mental Health Centre, Mental Health Services, Capital Region of Denmark and University of Copenhagen, Copenhagen, Denmark

[49]Division of Child and Adolescent Psychiatry, Department of Psychiatry, Lausanne University Hospital, Lausanne, Switzerland

[50]ASL BA, Maternal and Childood Department; Adolescence and Childhood Neuropsychiatry Unit; Bari, Italy

[51]Department of Child and Adolescent Psychiatry, Medical Faculty Carl Gustav Carus, TU Dresden, Dresden, Germany

[52]Department of Child and Adolescent Psychiatry and Psychotherapy, University of Zurich, Zurich, Switzerland

[53]Department of Clinical Neuroscience, UCL Institute of Neurology, University College London, London, UK

[54]Department of Clinical Neurosciences, Cumming School of Medicine & Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada

[55]Department of Genetics and the Human Genetics Institute of New Jersey, Rutgers, the State University of New Jersey, Piscataway, NJ, USA

[56]Department of Psychiatry and Behavioral Sciences, UCSF Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, USA.

[57]Quantitative Biosciences Institute, University of California, San Francisco, San Francisco, CA, USA.

[58]University of Groningen, University Medical Centre Groningen, Department of Child and Adolescent Psychiatry, Groningen, the Netherlands

[59]Division of Genetic Medicine, Department of Medicine Vanderbilt University Medical Center Nashville, Nashville, TN, USA

[60]Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA

[61]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[62]Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

[63]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[64]Department of Psychiatry and Genetics Institute, University of Florida College of Medicine, USA

[65]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

[66]Department of Neurology, Brigham and Women's Hospital, and the Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

[67]1st Laboratory of Medical Biology-Genetics, School of Medicine, Aristotle University of Thessaloniki, Thessaloniki, Greece

*co-corresponding authors:

Dr. Peristera Paschou (ppaschou@purdue.edu), Dr Pieter Hoekstra (p.hoekstra@accare.nl)

**equal contribution first authors

Abstract

Background: Tourette Syndrome (TS) is a childhood-onset neurodevelopmental disorder of complex genetic architecture, characterized by multiple motor tics and at least one vocal tic persisting for more than one year.

Methods: We performed a genome-wide meta-analysis integrating a novel TS cohort with previously published data, resulting in a sample size of 6,133 TS individuals and 13,565 ancestry-matched controls.

Results: We identified a genome-wide significant locus on chromosome 5q15. Integration of eQTL, Hi-C and GWAS data implicated the NR2F1 gene and associated lncRNAs within the 5q15 locus. Heritability partitioning identified statistically significant enrichment in brain tissue histone marks, while polygenic risk scoring on brain volume data identified statistically significant associations with right and left thalamus volumes and right putamen volume.

Conclusions: Our work presents novel insights in the neurobiology of TS opening up new directions for future studies.

Introduction

Tourette Syndrome (TS) is a childhood-onset neurodevelopmental disorder characterized by multiple motor tics and at least one vocal tic persisting for more than one year (1). The prevalence of TS is estimated in the range of 0.6-1% in school-aged children (2,3). It is a highly heritable disorder (4) with a population-based heritability estimated at 0.7 (5,6) and SNP-based heritability estimates ranging from 0.21 (7) to 0.58 (4) of the total heritability. TS exhibits high polygenicity and its genetic background is influenced by both common and rare variants of small effect spread throughout the genome (4,8,9). Two previously conducted genome-wide association studies (GWAS) (7,10), have indicated enrichment of TS genetic susceptibility variants in tissues within the cortico-striatal and cortico-cerebellar circuits, and in particular, the dorsolateral prefrontal cortex (7,10). Furthermore, gene set analyses of GWAS data implicated ligand-gated ion channel signaling, lymphocytic, and cell adhesion and trans-synaptic signaling processes as potential biological underpinnings in the pathogenesis of TS (11). Polygenic risk scores derived from the second TS GWAS can predict tic presence and severity at a statistically significant level (7,12). Additionally, the hypothalamus-pituitary-adrenal (HPA) axis was implicated in recent cross-disorder GWAS analysis for TS, attention deficit hyperactivity disorder (ADHD) and autism spectrum disorders (ASD) (13).

Here, we present a genome-wide meta-analysis for TS integrating novel and previously published data resulting in a combined sample size of 6,133 TS individuals and 13,565 ancestry-matched controls. We identify a novel genome-wide significant locus in the novel (TS-EUROTRAIN) GWAS and the TS GWAS meta-analysis. Our results provide further insight into the genetic basis of TS.

Methods and Materials

Datasets

The TS-EUROTRAIN dataset brings together three major TS cohorts, including 632 participants from the European Multicenter Tics in Children Study (EMTICS) (14), 763 participants from the TS-EUROTRAIN study (15), 238 participants from the TSGeneSEE study (16), and 52 participants from Sweden. These studies included participants from multiple European sites who were diagnosed using DSM-IV-TR or DSM-5 criteria for TS, consistent with previously published TS studies. In total, we collected samples from 1,685 individuals with TS (Supplementary Table 1). Additionally, 4,454 population-matched control individuals were recruited. Ancestry-matched controls were also used from the following public datasets following appropriate approvals: British WTCCC2 1958 Birth Cohort samples (Study accession code: EGAS00000000028), German control samples obtained from the POPGEN biobank (17), and French controls from the Three City Study (18), leading to a total of 8,558 general population controls (Supplementary Table 1). Written informed consent was obtained from all participants, as approved by the ethics committees of all participating institutions.

Genotyping, merging and imputation

Samples from the TS-EUROTRAIN dataset were genotyped on the Illumina HumanOmniExpress BeadChip at Decode genetics. The control samples obtained from collaborators and public repositories were genotyped on multiple Illumina arrays and full details are provided in the respective references and Supplementary Table 1. We applied standard GWAS quality control procedures to our data before and after the imputation, as described in previous GWAS performed by the Psychiatric Genomics Consortium (PGC) (7). Quality control procedures included the removal of samples that fit any of the following criteria: call rate < 0.98, absolute value of inbreeding coefficient < 0.2, genomic sex

discrepancy with reported sex, and formation of pairs with relatedness > 0.1875. We applied variant quality control, excluding markers with call rate < 0.98, differential missingness between cases and controls < 0.02, Hardy–Weinberg equilibrium P-value < 10^{-6} in controls and < 10^{-10} in cases.

The quality control steps were applied on each dataset separately. Imputation was performed on the Sanger Imputation Server using a reference panel of 64,940 European ancestry haplotypes (v1.1) from the Haplotype Reference Consortium (HRC) (19,20). We performed batch effect tests in samples of same status (case/control) between different sources, as they are shown in Supplementary Table 1, excluding markers that achieved a p-value < 10^{-5}. X chromosome data were excluded from the final analysis. To avoid ancestry bias, we matched the ancestry of the controls to TS individuals, at a three to one ratio, using the first five principal components as basis for a final dataset of 1,438 individuals with TS and 4,356 controls on 2,949,675 markers.

Genome-Wide Association Study

We conducted a GWAS using an additive logistic regression model on the best guess genotypes produced by imputation. We incorporated the principal components identified by Tracy-Widom statistics, as calculated by EIGENSOFT(21,22), as well as sex and imputation batch. We excluded SNPs when their MAF was <1% and minor allele count was <10, in either cases or controls. We set the level of genome-wide significance at $P = 5\times10^{-8}$. We estimated confounding bias in our results by performing LD Score Regression with *ldsc* (23) and using the attenuation ratio, as well as the p-value of the intercept to evaluate our results. Results were plotted using *matplotlib* in Python and the package *region-plot* (24).

Meta-analysis

We conducted a meta-analysis with the results of the second TS GWAS study conducted by the TS Working Group of the PGC (TSGWAS2) (7). Sample overlap was verified through genotypic identity-by-descent analysis, so the TS-EUROTRAIN GWAS was re-analyzed after excluding the overlapping samples (124 cases and 279 controls), leading to a non-overlapping sample size of 1,314 cases and 4,077 ancestry-matched controls. The summary statistics were used as input to METASOFT (25). METASOFT implements an array of methods for meta-analysis, especially in the case of heterogeneity in the results. In our study, we employed Han and Eskin's random effects model (RE2), which separates hypothesis testing from effect-size estimation, and is demonstrated to increase statistical power under heterogeneity compared to the conventional random effects model (26,27). We also employed METASOFT to produce m-values, that is, estimates of the posterior probability that an effect exists, with small values indicating absence of effect, large values indicating presence of effect, and intermediate values indicating ambiguity (25).

Heritability and heritability partitioning

SNP-based heritability was estimated through LD Score Regression (LDSC) (23). We further investigated heritability partitioning into functional categories using stratified (LDSC) (28). TS heritability was partitioned into 53 functional categories as well as into 220 cell-type-specific and 10 cell-type-group-specific annotations produced on the data derived from the Roadmap Epigenomics Project (29). The significance threshold for the heritability enrichments was defined at a Benjamini-Hochberg false discovery rate (FDR) < 0.05.

Genetic correlations

Bivariate LD score regression (23) was conducted to identify genetic correlations between the TS-EUROTRAIN GWAS, the TS EUROTRAIN/GWAS2 meta-analysis, and TSGWAS2 (7). We then examined each of these studies' cross-disorder correlations with multiple psychiatric and neurological disorders for which data were accessible and are listed in Supplementary Table 2. To avoid confounding due to sample size, we selected summary statistics from studies with more than 5,000 samples and all studies, except anxiety (h^2_{SNP} z-score= 2.5), have a heritability z-score >4. For the correlation analysis we used the European LD scores and merged alleles based on the HapMap3 reference panel for each trait, excluding markers residing in the Major Histocompatibility Complex region on chromosome 6. Significance threshold was defined by Benjamini-Hochberg FDR as $p < 0.05$.

Polygenic Risk Scoring

We used PRSice-2 (30) for our Polygenic Risk Score (PRS) analysis. We performed a unilateral PRS analysis between the TS-EUROTRAIN cohort and the TSGWAS2 (7) cohort, using the TSGWAS2 summary statistics as discovery and the TS-EUROTRAIN GWAS, after excluding the overlapping samples, as the target dataset. The TSGWAS2 summary statistics were clumped on the LD information of the TS-EUROTRAIN GWAS, using a window of 250kb and an r^2 threshold of 0.1. PRSice performed the scoring on subsets of the dataset based on nine thresholds of p-value leniency (5×10^{-8}, 10^{-4}, 10^{-3}, 0.01, 0.05, 0.1, 0.2, 0.5, 1). The resulting PRS was tested for association with TS, using logistic regression with the previously mentioned ancestry components, sex, and imputation batch as covariates. The model fit for best p-value threshold was run using 10,000 permutations. Liability scale was calculated on the variance explained by the PRS (R^2), using a TS population prevalence of 1%.

Biological annotation of results

We used FUMA (31) to perform gene-based and gene-set analyses on the results from the TS-EUROTRAIN GWAS and the subsequent TS-EUROTRAIN/GWAS2 meta-analysis. The genetic variants were assigned to protein-coding genes based on their GRCh37 build genomic position, using a ±20kb window size. After quality control, 18,089 genes contained at least one variant and as such were used for the gene-based analysis, thus setting the Bonferroni threshold at $p < 2.764\times10^{-6}$. The gene-based association results were subsequently used for gene-set analysis under a competitive model. Tissue Expression Analysis was conducted on the GTEx v8 expression data (32,33). We investigated chromatin contact points through Capture Hi-C data available from the 3D Genome Browser (34), using promoter-centered long-range chromatin interaction data derived from human dorsolateral prefrontal cortex tissues(35).

We performed a set-based association analysis using PLINK (36,37) on the gene-sets that were previously identified as significantly associated with TS (11,38). We used logistic regression as the association model on the genotypes and principal components that were identified by Tracy-Widom statistics in the GWAS. Another repetition of this step was performed with the χ^2 association test, to test for this method's robustness to population structure. We proceeded to run the analysis on all samples, using a 10kb genomic window size and a million permutations. Since the permutations were performed on the phenotypic status of the samples, and only served as a method of association of the trait with the gene sets, we also corrected the results by defining the significance threshold through Bonferroni correction.

Investigating genetic relationships with Subcortical Brain Volumes

Using the TS-EUROTRAIN/GWAS2 meta-analysis summary statistics as base, we computed TS PRS (PRS_{TS}) for individuals in the UK Biobank (UKBB) (39) using PRScs (40) and subsequently evaluated the association between risk scores and subjects' 14 subcortical

volumes (FIRST) that were available in the UKBB (category 1102). Since we had access to individual-level UKBB data, it was possible to calculate the association of PRS_{TS} and different brain volumes, instead of just evaluating their genetic correlation with LDSC. After quality controls (see Supplementary methods), 29,798 samples with brain MRI phenotypes available were included in the analysis. To calculate the PRS for these individuals we applied the continuous shrinkage method implemented in PRScs to obtain the updated effect sizes for all SNPs in TS-EUROTRAIN/GWAS2 meta-analysis summary statistics using European 1000 Genomes data as the reference for LD structure. We then used these updated effect sizes to calculate the PRS using the score function in PLINK (37). Regressions between PRS_{TS} and brain volume measurements were performed using the PHESANT tool (41) while using age, sex, genotyping batch and the first ten UKBB PCs, as covariates. Significance threshold was defined by Benjamini-Hochberg FDR as $p < 0.05$. Additionally, to leverage results from larger datasets for which individual-level data were not accessible, we used LDSC to test the genetic correlation of the TS-EUROTRAIN/GWAS2 meta-analysis and seven subcortical brain volumes GWAS summary statistics from the ENIGMA consortium (42).

Results

Mega-analysis of TS-EUROTRAIN GWAS

GWAS analysis was performed as a mega-analysis, on the combined genetic data of all TS-EUROTRAIN samples (1,438 TS cases and 4,356 controls), using a logistic regression model on the best-guess genotypes (genotype probability > 0.9) with INFO score > 0.9 and MAF> 0.01. As covariates, we included the ancestry components 1,2,4, and 5 to account for population stratification as identified by ANOVA statistics (Supplementary Table 3), sex, and imputation batch.

The TS-EUROTRAIN GWAS identified three highly-correlated (r^2>0.8) genome-wide significant SNPs, located near the *NR2F1 Antisense RNA 1* long non-coding RNA (*NR2F1-AS1* lncRNA) locus (Supplementary Figures 1 and 2a). The strongest signal was found for rs2453763 (chr5:92376460:T/A, *OR* = 0.7512, $P = 2.62\times10^{-8}$, MAF=0.3581), a variant 350kb upstream of *NR2F1-AS1*, and associated with decreased risk for TS. The imputation statistics for this SNP indicate high imputation quality (MAF= 0.3581, INFO= 0.99). The proximal SNPs were rs2009416 (chr5:92415111:T/C, *OR* = 0.7532, $P = 3.31\times10^{-8}$, MAF=0.3562) and rs1496337 (chr5:92411293:T/C, *OR*= 0.7534, $P = 3.33\times10^{-8}$, MAF=0.3563). Conditional analysis using the lead SNP as covariate showed no secondary signals in the region. The top ($P < 10^{-5}$) loci detected in the novel GWAS are reported in Table 1. LD Score Regression analysis of the summary statistics did not provide evidence for genomic inflation (λ_{GC} = 1.07, intercept=1.0061, intercept p-value=0.28, attenuation ratio=0.0887), while using the full GWAS SNPs the λ_{GC} is 1.05 (Supplementary Figure 1b).

TS-EUROTRAIN Meta-analysis with TSGWAS2

The TS-EUROTRAIN GWAS was then meta-analyzed with summary statistics results from the previous largest meta-analysis of TS to date (TSGWAS2) (7) using Han and Eskin's random effects model (25,26). Since there was a small but known sample overlap between the two studies, the TS-EUROTRAIN GWAS was re-analyzed after excluding the overlapping samples (124 cases and 279 controls) resulting in a dataset of 1,314 cases and 4,077 ancestry-matched controls, with the results being very similar to the full dataset TS-EUROTRAIN GWAS (Supplementary Figure 3). Only variants overlapping in both studies were included, leading to a total of 6,133 cases, 13,565 controls and 1,955,677 variants in the final meta-analysis.

The TS-EUROTRAIN/GWAS2 meta-analysis (Figure 1, Supplementary Table 4) pointed again to the three genome-wide significant SNPs of the TS-EUROTRAIN GWAS, which remained genomewide significant in the meta-analysis with rs2453763 again being the most strongly associated (chr5:92376460:T/A, random effects $P = 4.05\times10^{-8}$). SNP, rs10209244 (chr2:161561898:A/G, random effects $P = 6.16\times10^{-8}$, $MAF = 0.01$) that resides 200kb downstream of *RBMS1* (Supplementary Figure 2b) was the next hit and did not manage to achieve genomewide significance. The top loci detected from the meta-analysis are reported in Table 2. LD Score Regression analysis of the summary statistics did not provide evidence for genomic inflation ($\lambda_{GC} = 1.16$, intercept=1.016, intercept p-value=0.11, attenuation ratio=0.0869).

Genetic relationship between the TS-EUROTRAIN GWAS and TSGWAS2

The SNP with the strongest signal in the previously published TSGWAS2 study was absent from the TS-EUROTRAIN dataset due to the differences in reference panels used (1000 Genomes for TSGWAS2 and HRC for the novel study) and stringent batch effect quality control performed on the novel dataset. However, we observed no genomewide significant heterogeneity (Cochran's Q-test p-value$< 5\times10^{-8}$) in the meta-analysis.

To explore the relationship between the TS-EUROTRAIN GWAS and TSGWAS2, we used LDSC (23) to compute their genetic correlation, after excluding the overlapping samples. LDSC identified a strong genetic correlation between the two studies (rg = 0.95, $p = 6\times10^{-8}$), and provided evidence of consistency across them (Figure 2). Investigation of the gene sets found previously associated with TS (11,38) also successfully replicated the associations for the lymphocytic, the ligand-gated ion channel signaling, the cell adhesion and trans-synaptic signaling, as well as the astrocyte-neuron metabolic coupling gene sets (Supplementary Table 5).

PRS analysis displayed consistency between the two studies. PRS were computed using the summary statistics of TSGWAS2 as a training dataset and the TS-EUROTRAIN raw genotypes as discovery in PRSice (43). The best fit p-value threshold was determined at p = 0.1182 (model fit $p = 1.26\times10^{-28}$) (Figure 3a-b). Maximum variance explained at the best fit model was estimated by Nagelkerke's R^2 at 3.3%.

Cross-disorder analysis

Pairwise genetic correlations were computed between ten psychiatric and six neurological traits using LDSC (23) and results are shown in Figure 2. Benjamini-Hochberg FDR correction with an α=0.05 was used to correct for multiple testing. After correction, the TS-EUROTRAIN/GWAS2 meta-analysis was significantly correlated with OCD (*rg* = 0.39, $p_{FDR} = 3.4\times10^{-3}$. We also observed genetic correlations with p<0.05 for the meta-analysis with ADHD, SCZ, and Migraine which were however not significant after FDR correction.

Heritability estimation and partitioning

We used LDSC (23) to estimate the SNP-based heritability (h^2_{SNP}) using the summary statistics of the novel GWAS and the meta-analysis. The summary statistics were merged with the HapMap3 marker panel provided by the authors. For the TS-EUROTRAIN GWAS, analysis yielded an h^2_{SNP} estimate of 0.4385 (*SE*: 0.1167) on the observed scale, and 0.2736 (*SE*: 0.0728) assuming a TS prevalence of 0.01 on the liability scale, while the LD score regression analysis intercept was computed at 1.0157 (*SE*: 0.013) (p-value 0.028) and the ratio of stratification to polygenicity was estimated at 0.0863 (*SE*: 0.0711). For the meta-analysis the h^2_{SNP} estimate was 0.3504 (*SE* : 0.0439) and 0.2184 (*SE*: 0.0269) on the liability scale.

We proceeded to partition the heritability of the meta-analysis GWAS by functional genomic categories using stratified LD Score Regression (28) on the full baseline model and a model based on the Roadmap epigenomics data, as provided by the authors (28). The full baseline

model included 24 main overlapping functional categories and identified statistically significant enrichment in two categories, after Benjamini-Hochberg FDR correction at an α=0.05. The H3K4me1 sites category (enrichment value 1.61, $P = 9.5\times10^{-4}$) was the top significant signal in the analysis, with the conserved elements category (enrichment value 2.05, $P = 3.8\times10^{-3}$) being the second significant signal. The Roadmap model includes epigenomic mapping data from 395 tissues (29) and when applied to our data for heritability partitioning, yielded 13 statistically significant modifications after Benjamini-Hochberg FDR correction at an α=0.05. These 13 signals marked the enrichment of the histone marks H3K27ac, H3K4me1, and H3K9ac on five brain tissues. H3K27ac was identified on the angular gyrus, the cingulate gyrus, the dorsolateral prefrontal cortex, and the inferior temporal lobe; H3K4me1 on the angular gyrus, the cingulate gyrus, the dorsolateral prefrontal cortex, the inferior temporal lobe, and the substantia nigra; H3K9ac on the angular gyrus, the anterior caudate, the dorsolateral prefrontal cortex, and the inferior temporal lobe (Supplementary Tables 6 and 7).

Biological annotation and enrichment analysis

Functional mapping, annotation, and gene set enrichment using the FUMA pipeline did not produce significant results. The identified top signals from the TS-EUROTRAIN GWAS and the TS-EUROTRAIN/GWAS2 meta-analysis reside in large intergenic regions exceeding the distance limits set by the software, and were thus excluded from the annotation step of the pipeline. The top signal of the TS-EUROTRAIN/GWAS2 meta-analysis gene-based analysis was *RANGAP1* ($P = 3.36\times10^{-6}$) on chromosome 22; it did not meet the genome-wide significance threshold ($P = 2.8\times10^{-6}$) (Supplementary Table 8). MAGMA tissue expression analysis using FUMA did not produce any statistically significant results for the TS-EUROTRAIN GWAS or the meta-analysis (Supplementary Figures 4a-b and 5a-b). MAGMA tissue expression analysis of the meta-analysis, using the 53 tissue sample set from GTEx, indicated stronger putative enrichment in various brain tissues, with the top signals in the

cerebellar hemisphere, cerebellum, and frontal cortex (area BA9). Using the 30 tissue sample set from GTEx, stronger evidence of potential enrichment could be identified in the brain, followed by the pituitary and the ovary (Supplementary Figure 5a-b).

Regarding the top SNP in our GWAS, the GTEx portal (32) reports SNP rs2453763 to be significantly associated as a splicing quantitative trait locus (sQTL) for *CTD-2091N23.1* in the tibial nerve, and as an expression quantitative trait locus (eQTL) for *NR2F1* and *NR2F1-AS* in the esophagus smooth muscles and for *CTD-2091N23.1* in cultured fibroblasts (Table 3). Capture Hi-C (34) showed strong evidence for the SNP being related to the regulation of *NR2F1* (Supplementary Figure 6a).

Genetic correlation to Subcortical Brain volumes

The genetic correlation analysis for our TS-EUROTRAIN/GWAS2 meta-analysis and ENIGMA GWAS summary statistics for seven subcortical brain volumes (42) did not reveal any significant correlation (Supplementary Table 9). However, with individual-level data available within the UKBB, we were also able to test for association of TS PRS_{TS} to specific brain phenotypes. Our results highlighted the previously described relationship (44) between genetic risk for TS and putamen volume. We observed that increase in the genetic risk of TS was associated with decrease in right putamen (beta: -0.0175, adj.p: 0.0069) and left pallidum (beta: -0.0137, adj.p: 0.043) volumes. Significant associations were also observed between PRS_{TS} and bilateral thalamic volume indicating that increase in PRS_{TS} was associated with decrease in right thalamus volume (beta: -0.0138, adj.p: 0.035) and left thalamus volume (beta: -0.0132, adj.p: 0.037) (Supplementary Table 10).

Discussion

We present results from a novel TS GWAS and integrate with a previous study to report the largest GWAS meta-analysis on TS at this time (6,133 TS individuals of European ancestry and 13,565 matched controls). We find one novel independent genome-wide significant locus associated with TS on chromosome 5q15 upstream of a gene cluster that harbors *NR2F1-AS*, *NR2F1*, and *lnc-NR2F1*. The top associated SNP is located within *CTD-2091N23.1*, a gene encoding a long non-coding RNA that has yet to be functionally characterized. Our study provides novel insights into the genetic cause and pathophysiology of TS. However, replication of results in independent samples and even larger studies are needed in order to ultimately elucidate the background of this complex disorder and lead towards interventions that will be informed by genetic discoveries.

The 5q15 region that is highlighted by our study has previously been implicated in neurodevelopmental phenotypes (45–48). The 5q14-5q15 regions have been reported to contain fragile sites that are associated with genomic and epigenomic instability as well as linked to schizophrenia and autism (49). The exact genes are yet to be identified, with recent evidence suggesting a role for *NR2F1*-related genes, and more intriguingly, the *lnc-NR2F1* gene. *lnc-NR2F1* is a long non-coding RNA locus discovered to be recurrently mutated in individuals with autism spectrum disorders and intellectual disability, with translocations in this locus reported to show patterns of Mendelian inheritance (50). A functional study of *lnc-NR2F1* identified its role in neuronal maturation *in vitro* through expression regulation of a network of genes that have been linked to autism (50). Functional studies of the *NR2F1* gene also have indicated its critical role for neurodevelopment through investigations into human and mouse haploinsufficiency (51), insertion of point mutations in mouse models that lead to excitatory/inhibitory neuronal imbalance (52), and the study of knock-out mouse models (53). Notably, in the absence of *NR2F1*, an imbalance between oligodendrocytes and astrocytes develops, leading to postnatal hypomyelination and astrogliosis (51).

NR2F1 is a highly conserved orphan nuclear receptor which is a regulator of transcription. It belongs to the steroid/thyroid hormone nuclear receptor superfamily, involved in a wide range of roles, including cell differentiation, cancer progression, and central and peripheral neurogenesis (54). A multitude of pathogenic variants have been identified in NR2F1, leading to Bosch-Boonstra-Schaaf optic atrophy syndrome, and autosomal dominant neurodevelopmental disorder (55). *NR2F1* is also known by its historical name, *COUP-TF1*; it is a target of the androgen receptor (AR), along with *SOX9* and *OCT4* (56). NR2F1 interacts with SOX9 (56,57)) and represses a host of targets in multiple tissues, including *CYP17A1*, Oxytocin gene *OXT*, and *OCT4* (58). Especially in the case of *CYP17A1*, encoding for a key enzyme of steroid biosynthesis, NR2F1 and SF-1 exert opposing effects, as repressor and activator, respectively (59).

A limitation of our study is that we did not replicate our findings in independent samples due to lack of data availability. Future studies will help evaluate the magnitude of the contribution from the locus implicated by our study. Nevertheless, we sought to validate our results through means of heritability correlation patterns and polygenic risk scoring. Heritability analysis in the TS-EUROTRAIN dataset indicated that a large proportion of TS SNP-based heritability can be attributed to common variants (h^2_{SNP} =0.4385), in concordance with the estimate ranges in previous investigations (4,7). The lower SNP heritability (h^2_{SNP} =0.3504) we observed in meta-analysis, we suspect could be due to the heterogeneity with additional samples and different populations included in the meta-analysis. Polygenic prediction in the TS-EUROTRAIN cohort using the TSGWAS2 results as discovery achieved significant predictive levels, on par with the inter-cohort predictive Nagelkerke's R^2 levels in the previous TS GWAS (7) and substantially increased by more than an order of magnitude compared to tic prediction in a general population cohort (12).

Like most GWAS, this study does not yet have direct clinical implications. However, this work can motivate future studies that may have significant clinical impact, like for instance patient stratification according to genetic risk to inform clinical practice or drug discovery based on genetic loci that are identified through genetic association. Our recently described phenomewide association analysis of TS genetic risk and a large number of clinical traits available in the UKBB represents one further step towards clinically related discoveries (60). Our study is an important stepping stone towards understanding the genetic background of TS and confirms the value of collaborative efforts towards expanding sample size and datasets available for analysis (such as the TS-EUROTRAIN, EMTICS, TSGeneSEE, and PGC initiatives). However, we still only capture a small fraction of the risk for TS attributable to common variants. Larger studies, bringing together even larger datasets are necessary and warranted and indeed are currently underway, with participation from authors of this work. Increased statistical power will further enable the identification of more leads towards the elucidation of the underlying biology of TS and potential future interventions.

Conflicts of interest

The authors report no biomedical financial interests or potential conflicts of interest.

Funding

This study was supported by the EMTICS (FP7-HEALTH, Grant agreement ID: 278367), TSEUROTRAIN (FP7-PEOPLE, Grant agreement ID: 316978), NINDS (R01NS105746), NIMH (R01MH126213) grants to PP. AJW was funded by R01NS105746, the Tourette Association of America, and the Weill Institute for Neurosciences. AM was funded by the Deutsche Forschungsgemeinschaft (DFG; FOR 2698). AS received support from the NIHR UCL/H Biomedical Research Centre. BH is an employee of Boehringer Ingelheim Pharma.

CJ was funded by Lundbeck Fonden, grant number R100-2011-9332. CB was supported by funding from the Merit-prize fellowship of Semmelweis University, the Bolyai Janos research fellowship of the Hungarian Academy of Sciences BO/00987/16/5, the UNKP-18-4 of the new National Excellence Program of the Ministry of Human Capacities and the Baron Munchausen Program of the Institute of Medical Chemistry, Molecular Biology and Pathobiochemistry, Semmelweis University. DCC was funded by the TSAA, by the Stichting VC-GGZ, and by TS-EUROTRAIN. DM has received research support from Ipsen Corporate and funding grants from Dystonia Medical Research Foundation Canada, Parkinson Canada, The Owerko Foundation, and the Michael P Smith Family. LKD was supported by grants from the National Institutes of Health including R01NS102371 and R01NS105746. PM has received grants from the Spanish Ministry of Science and Innovation [RTC2019-007150-1], the Instituto de Salud Carlos III-Fondo Europeo de Desarrollo Regional (ISCIII-FEDER), PI16/01575, PI19/01576], the Consejeria de Economia, Innovacion, Ciencia y Empleo de la Junta de Andalucia [CVI-02526, CTS-7685], the Consejeria de Salud y Bienestar Social de la Junta de Andalucia, PE-0210-2018]. PJ and CZ were funded by the National Science Center, Poland: UMO-2016/23/B/NZ2/03030. ZT funded by Lundbeck Fonden, grant number R100-2011-9332. TIC Genetics was supported by grants from NIH (MH115958, MH115960, MH115962, MH115961, MH115993, MH115963, MH115959) and the New Jersey Center for Tourette Syndrome and Associated Disorders (NJCTS).

Acknowledgements

The TSAICG includes Cathy L. Barr, James R. Batterson, Cheston Berlin, Cathy L. Budman, Giovanni Coppola, Nancy J. Cox, Sabrina Darrow, Yves Dion, Nelson B. Freimer, Marco A. Grados, Erica Greenberg, Matthew E. Hirschtritt, Alden Y. Huang, Cornelia Illmann, Robert A. King, Roger Kurlan, James F. Leckman, Gholson J. Lyon, Irene A. Malaty, William M.

McMahon, Benjamin M. Neale, Michael S. Okun, Lisa Osiecki, Mary M. Robertson, Guy A. Rouleau, Paul Sandor, Harvey S. Singer, Jan H. Smit, Jae Hoon Sul

The TSGeneSEE initiative includes Christos Androutsos, Entela Basha, Luca Farkas, Jakub Fichna, Piotr Janik, Mira Kapisyzi, Iordanis Karagiannidis, Anastasia Koumoula, Peter Nagy, Joanna Puchala, Natalia Szejko, Urszula Szymanska, Vaia Tsironi

The EMTICS collaborative group includes Alan Apter, Juliane Ball, Benjamin Bodmer, Emese Bognar, Judith Buse, Marta Correa Vela, Carolin Fremer, Blanca Garcia-Delgar, Mariangela Gulisano, Annelieke Hagen, Julie Hagstrøm, Marcos Madruga-Garrido, Peter Nagy, Alessandra Pellico, Daphna Ruhrman, Jaana Schnell, Paola Rosaria Silvestri, Liselotte Skov, Tamar Steinberg, Friederike Tagwerker Gloor, Victoria L. Turner, Elif Weidinger

The TS-EUROTRAIN network includes John Alexander, Tamas Aranyi, Wim R. Buisman, Jan K. Buitelaar, Nicole Driessen, Petros Drineas, Siyan Fan, Natalie J. Forde, Sarah Gerasch, Odile A. van den Heuvel, Cathrine Jespersgaard, Ahmad S. Kanaan, Harald E. Möller, Muhammad S. Nawaz, Ester Nespoli, Luca Pagliaroli, Geert Poelmans, Petra J. W. Pouwels, Francesca Rizzo, Dick J. Veltman, Ysbrand D. van der Werf, Joanna Widomska, Nuno R. Zilhäo The TIC Genetics authors collaborative group includes Lawrence W. Brown, Keun-Ah Cheon, Barbara J. Coffey, Thomas V. Fernandez, Blanca Garcia-Delgar, Donald L Gilbert, Julie Hagstrøm, Hyun Ju Hong, Laura Ibanez-Gomez, Eun-Joo Kim, Young Key Kim, Young-Shin Kim, Robert A. King, Yun-Joo Koh, Sodahm Kook, Samuel Kuperman, Bennett L. Leventhal, Marcos Madruga-Garrido, Athanasios Maras, Tara L. Murphy, Eun-Young Shin, Dong-Ho Song, Jungeun Song, Matthew W State, Frank Visscher, Sheng Wang, Samuel H. Zinner

Data availability

Summary statistics data are available upon request to the corresponding authors.

References

1. Robertson MM, Cavanna AE, Eapen V (2015): Gilles de la Tourette syndrome and disruptive behavior disorders: prevalence, associations, and explanation of the relationships. *J Neuropsychiatry Clin Neurosci* 27: 33–41.

2. Scharf JM, Miller LL, Gauvin CA, Alabiso J, Mathews CA, Ben-Shlomo Y (2015): Population prevalence of Tourette syndrome: A systematic review and meta-analysis. *Movement Disorders*, vol. 30. pp 221–228.

3. Robertson MM, Eapen V, Cavanna AE (2009): The international prevalence, epidemiology, and clinical phenomenology of Tourette syndrome: a cross-cultural perspective. *J Psychosom Res* 67: 475–483.

4. Davis LK, Yu D, Keenan CL, Gamazon ER, Konkashbaev AI, Derks EM, *et al.* (2013): Partitioning the heritability of Tourette syndrome and obsessive compulsive disorder reveals differences in genetic architecture. *PLoS Genet* 9: e1003864.

5. Robertson MM, Eapen V, Singer HS, Martino D, Scharf JM, Paschou P, *et al.* (2017): Gilles de la Tourette syndrome. *Nat Rev Dis Primers* 3: 16097.

6. Mataix-Cols D, Isomura K, Pérez-Vigil A, Chang Z, Rück C, Larsson KJ, *et al.* (2015): Familial Risks of Tourette Syndrome and Chronic Tic Disorders. A Population-Based Cohort Study. *JAMA Psychiatry* 72: 787–793.

7. Yu D, Sul JH, Tsetsos F, Nawaz MS, Huang AY, Zelaya I, *et al.* (2019): Interrogating the Genetic Determinants of Tourette's Syndrome and Other Tic Disorders Through Genome-Wide Association Studies. *Am J Psychiatry* 176: 217–227.

8. Huang AY, Yu D, Davis LK, Sul JH, Tsetsos F, Ramensky V, *et al.* (2017): Rare Copy Number Variants in NRXN1 and CNTN6 Increase Risk for Tourette Syndrome. *Neuron* 94: 1101–1111.e7.

9. Wang S, Mandell JD, Kumar Y, Sun N, Morris MT, Arbelaez J, *et al.* (2018): De Novo Sequence and Copy Number Variants Are Strongly Associated with Tourette Disorder and Implicate Cell Polarity in Pathogenesis. *Cell Rep* 25: 3544.

10. Scharf JM, Yu D, Mathews CA, Neale BM, Stewart SE, Fagerness JA, *et al.* (2013): Genome-wide association study of Tourette's syndrome. *Mol Psychiatry* 18: 721–728.

11. Tsetsos F, Yu D, Sul JH, Huang AY, Illmann C, Osiecki L, *et al.* (2021): Synaptic processes and immune-related pathways implicated in Tourette syndrome. *Transl Psychiatry* 11: 56.

12. Abdulkadir M, Mathews CA, Scharf JM, Yu D, Tischfield JA, Heiman GA, *et al.* (2019): Polygenic Risk Scores Derived From a Tourette Syndrome Genome-wide Association Study Predict Presence of Tics in the Avon Longitudinal Study of Parents and Children Cohort. *Biological Psychiatry*, vol. 85. pp 298–304.

13. Yang Z, Wu H, Lee PH, Tsetsos F, Davis LK, Yu D, *et al.* (2021): Investigating Shared Genetic Basis Across Tourette Syndrome and Comorbid Neurodevelopmental Disorders Along the Impulsivity-Compulsivity Spectrum. *Biol Psychiatry* 90: 317–327.

14. Schrag A, EMTICS Collaborative Group, Martino D, Apter A, Ball J, Bartolini E, *et al.* (2019): European Multicentre Tics in Children Studies (EMTICS): protocol for two cohort studies to assess risk factors for tic onset and exacerbation in children and adolescents. *European Child & Adolescent Psychiatry*, vol. 28. pp 91–109.

15. Forde NJ, Kanaan AS, Widomska J, Padmanabhuni SS, Nespoli E, Alexander J, *et al.* (2016): TS-EUROTRAIN: A European-Wide Investigation and Training Network on the Etiology and Pathophysiology of Gilles de la Tourette Syndrome. *Front Neurosci* 10: 384.

16. Karagiannidis I, Rizzo R, Tarnok Z, Wolanczyk T, Hebebrand J, Nöthen MM, *et al.* (2012): Replication of association between a SLITRK1 haplotype and Tourette Syndrome in a large sample of families. *Mol Psychiatry* 17: 665–668.

17. Nöthlings U, Krawczak M (2012): PopGen: Eine populationsbasierte Biobank mit Langzeitverfolgung der Kontrollkohorte (Leitthema). *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 55: 831–835.

18. 3C Study Group (2003): Vascular factors and risk of dementia: design of the Three-City Study and baseline characteristics of the study population. *Neuroepidemiology* 22: 316–325.

19. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, *et al.* (2016): Next-generation genotype imputation service and methods. *Nat Genet* 48: 1284–1287.

20. Consortium THR, the Haplotype Reference Consortium (2016): A reference panel of 64,976 haplotypes for genotype imputation. *Nature Genetics*, vol. 48. pp 1279–1283.

21. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006): Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38: 904–909.

22. Patterson N, Price AL, Reich D (2006): Population structure and eigenanalysis. *PLoS Genet* 2: e190.

23. Bulik-Sullivan BK, Loh P-R, Finucane H, Ripke S, Yang J, Patterson N, *et al.* (n.d.): LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies. https://doi.org/10.1101/002931

24. Perreault L-PL (2020): *Pgxcentre/region-Plot: Release 0.1.4*. https://doi.org/10.5281/zenodo.4095310

25. Han B, Eskin E (2012): Interpreting meta-analyses of genome-wide association studies. *PLoS Genet* 8: e1002555.

26. Han B, Eskin E (2011): Random-Effects Model Aimed at Discovering Associations in Meta-Analysis of Genome-wide Association Studies. *The American Journal of Human Genetics*, vol. 88. pp 586–598.

27. Urbut SM, Wang G, Carbonetto P, Stephens M (n.d.): Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. https://doi.org/10.1101/096552

28. Finucane HK, ReproGen Consortium, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, *et al.* (2015): Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, vol. 47. pp 1228–1235.

29. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, *et al.* (2015): Integrative analysis of 111 reference human epigenomes. *Nature* 518: 317–330.

30. Choi SW, O'Reilly PF (2019): PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* 8. https://doi.org/10.1093/gigascience/giz082

31. Watanabe K, Taskesen E, van Bochoven A, Posthuma D (2017): Functional mapping and

annotation of genetic associations with FUMA. *Nat Commun* 8: 1826.

32. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, *et al.* (2013): The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 45: 580–585.

33. GTEx Consortium (2020): The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* 369: 1318–1330.

34. Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, *et al.* (2018): The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol* 19: 151.

35. Jung I, Schmitt A, Diao Y, Lee AJ, Liu T, Yang D, *et al.* (2019): A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nature Genetics*, vol. 51. pp 1442–1449.

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

37. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ (2015): Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4: 7.

38. de Leeuw C, Goudriaan A, Smit AB, Yu D, Mathews CA, Scharf JM, *et al.* (2015): Involvement of astrocyte metabolic coupling in Tourette syndrome pathogenesis. *Eur J Hum Genet* 23: 1519–1522.

39. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, *et al.* (2018): The UK Biobank resource with deep phenotyping and genomic data. *Nature* 562: 203–209.

40. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW (2019): Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 10: 1776.

41. Millard LAC, Davies NM, Gaunt TR, Davey Smith G, Tilling K (2018): Software Application Profile: PHESANT: a tool for performing automated phenome scans in UK Biobank. *Int J Epidemiol* 47: 29–35.

42. Satizabal CL, Adams HHH, Hibar DP, White CC, Knol MJ, Stein JL, *et al.* (2019): Genetic architecture of subcortical brain structures in 38,851 individuals. *Nat Genet* 51: 1624–1636.

43. Euesden J, Lewis CM, O'Reilly PF (2015): PRSice: Polygenic Risk Score software. *Bioinformatics* 31: 1466–1468.

44. Mufford M, Cheung J, Jahanshad N, van der Merwe C, Ding L, Groenewold N, *et al.* (n.d.): Concordance of Genetic Variation that Increases Risk for Tourette Syndrome and that Influences its Underlying Neurocircuitry. https://doi.org/10.1101/366294

45. Goodart SA, Butler MG, Overhauser J (1996): Familial double pericentric inversion of chromosome 5 with some features of cri-du-chat syndrome. *Hum Genet* 97: 802–807.

46. Malan V, Martinovic J, Sanlaville D, Caillat S, Waill M-CP, Ganne M-LM, *et al.* (2006): Molecular characterisation of a prenatally diagnosed 5q15q21.3 deletion and review of the literature. *Prenat Diagn* 26: 231–238.

47. Brown KK, Alkuraya FS, Matos M, Robertson RL, Kimonis VE, Morton CC (2009): NR2F1 deletion in a patient with a de novo paracentric inversion, inv(5)(q15q33.2), and syndromic deafness. *Am J Med Genet A* 149A: 931–938.

48. Al-Kateb H, Shimony JS, Vineyard M, Manwaring L, Kulkarni S, Shinawi M (2013):

NR2F1 haploinsufficiency is associated with optic atrophy, dysmorphism and global developmental delay. *Am J Med Genet A* 161A: 377–381.

49. Smith CL, Bolton A, Nguyen G (2010): Genomic and epigenomic instability, fragile sites, schizophrenia and autism. *Curr Genomics* 11: 447–469.

50. Ang CE, Ma Q, Wapinski OL, Fan S, Flynn RA, Lee QY, *et al.* (2019): The novel lncRNA lnc-NR2F1 is pro-neurogenic and mutated in human neurodevelopmental disorders. *Elife* 8. https://doi.org/10.7554/eLife.41770

51. Bertacchi M, Gruart A, Kaimakis P, Allet C, Serra L, Giacobini P, *et al.* (2019): Mouse Nr2f1 haploinsufficiency unveils new pathological mechanisms of a human optic atrophy syndrome. *EMBO Mol Med* 11: e10291.

52. Zhang K, Yu F, Zhu J, Han S, Chen J, Wu X, *et al.* (2020): Imbalance of Excitatory/Inhibitory Neuron Differentiation in Neurodevelopmental Disorders with an NR2F1 Point Mutation. *Cell Rep* 31: 107521.

53. Del Pino I, Tocco C, Magrinelli E, Marcantoni A, Ferraguto C, Tomagra G, *et al.* (2020): COUP-TFI/Nr2f1 Orchestrates Intrinsic Neuronal Activity during Development of the Somatosensory Cortex. *Cereb Cortex* 30: 5667–5685.

54. Manikandan M, Abuelreich S, Elsafadi M, Alsalman H, Almalak H, Siyal A, *et al.* (2018): NR2F1 mediated down-regulation of osteoblast differentiation was rescued by bone morphogenetic protein-2 (BMP-2) in human MSC. *Differentiation* 104: 36–41.

55. Rech ME, McCarthy JM, Chen C, Edmond JC, Shah VS, Bosch DGM, *et al.* (2020): Phenotypic expansion ofBosch–Boonstra–Schaafoptic atrophy syndrome and further evidence for genotype–phenotype correlations. *American Journal of Medical Genetics Part A*, vol. 182. pp 1426–1437.

56. Perets R, Kaplan T, Stein I, Hidas G, Tayeb S, Avraham E, *et al.* (2012): Genome-wide analysis of androgen receptor targets reveals COUP-TF1 as a novel player in human prostate cancer. *PLoS One* 7: e46467.

57. Sosa MS, Parikh F, Maia AG, Estrada Y, Bosch A, Bragado P, *et al.* (2015): NR2F1 controls tumour cell dormancy via SOX9- and RARβ-driven quiescence programmes. *Nature Communications*, vol. 6. https://doi.org/10.1038/ncomms7170

58. Tang K, Tsai SY, Tsai M-J (2015): COUP-TFs and eye development. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, vol. 1849. pp 201–209.

59. Shibata H, Ando T, Suzuki T, Kurihara I, Hayashi K, Hayashi M, *et al.* (1998): COUP-TFI Expression in Human Adrenocortical Adenomas: Possible Role in Steroidogenesis. *The Journal of Clinical Endocrinology & Metabolism*, vol. 83. pp 4520–4520.

60. Jain P, Miller-Fleming T, Topaloudi A, Yu D, Drineas P, Georgitsi M, *et al.* (2022, September 2): Polygenic risk score-based phenome-wide association study identifies novel associations for Tourette syndrome. *bioRxiv*. https://doi.org/10.1101/2022.09.01.22279340

Tables and figure legends

Table 1: **Top regions ($p<10^{-5}$) in the TS-EUROTRAIN GWAS.** The total sample size was 1,438 cases and 4,356 controls on 2,949,675 variants). One variant was identified as genome-wide significant ($p<5\times10^{-8}$). Chromosome and region (based on hg19) are shown for index SNPs (LD-r^2 >0.1), as well as the number of LD-associated markers in proximity (N). A1 refers to the associated allele. The odds ratio (OR) and standard error (SE) are shown for the association between A1 and TS. MAF indicates the allelic frequency of allele 1 in the dataset. The reported nearest genes were determined by genomic location (±500 kb). The analysis was restricted to variants with MAF ≥ 0.01 and information quality (INFO) score ≥ 0.9. Chromosome X was not analyzed, since it was absent from a significant portion of the acquired datasets.

Table 2: Top regions ($p<10^{-5}$) in the TS GWAS meta-analysis (TS-EUROTRAIN and TSGWAS2). The analysis performed on 6,133 cases, 13,565 controls on 1,955,677 variants using Han and Eskin's random effects model (24). One variant was identified as genome-wide significant ($p<5\times10^{-8}$) while the second top hit reached a ($p<10^{-7}$). Chromosome and region (based on hg19) are shown for index SNPs (LD-r^2>0.1), as well as number of LD-associated markers in proximity (N). The reported nearest genes were determined by genomic location (±500 kb). MAF indicates the allelic frequency of the minor allele in the dataset. MAF EUR (1KG) indicates the frequency of the minor allele in the meta-analysis in the 1000 Genomes Project European samples.

Table 3: Significant SNP-gene pairings identified through GTEx eQTL and sQTL data (37, 38). Four significant associations were identified for SNP rs2453763, while no significant associations were identified for rs10209244. rs2453763 is an eQTL for three genes on two tissues, and an sQTL for one gene on one tissue. Reported are the symbol of the associated gene, the respective associated tissue, and the normalized effect size (NES). a) GTEx eQTL associations for the top variant in NR2F1 (rs2453763). b) GTEx sQTL associations for the top variant in NR2F1 (rs2453763)

Figure 1. The Manhattan plot for the genome-wide association meta-analysis of Tourette Syndrome with the TS-EUROTRAIN and the TSGWAS2 datasets (6,133 TS cases and 13,565 controls of European descent on 1,955,677 variants) using Han and Eskin's random effects model (24). The -log10(p) values for the association tests (two-tailed) are shown on the y axis and the chromosomes are ordered on the x axis. One genetic locus on chromosome 5 surpassed the genome-wide significance threshold ($p<5\times10^{-8}$; indicated by the red line). Gray and black differentiate adjacent chromosomes.

Figure 2. Genetic correlations with Tourette Syndrome. The genetic correlations were estimated with bivariate LD score regression (21). We showcase the correlations between three TS studies (TS-EUROTRAIN, TS-EUROTRAIN/TSGWAS2 meta-analysis, and TSGWAS2) and 16 psychiatric and neurological traits (see supplementary Table 2 for full list of studies and abbreviations). The number in each square denotes the correlation rg. Two asterisks (**) denote the correlations that were identified as statistically significant after Benjamini-Hochberg FDR correction (a=0.05), while one asterisk (*) shows the correlations with nominal p<0.05.

Figure 3. Polygenic Risk Scoring analysis using the TSGWAS2 dataset (7) as discovery and the TS-EUROTRAIN dataset as target. Best fit p-value threshold was determined at p=0.1182 (model fit $p=1.26\times10^{-28}$). Maximum variance explained at the best fit model was estimated by Nagelkerke's R^2 at 3.3%. a) PRS distribution comparison between cases and controls for the best fit model. b) PRS

histogram for each p-value bin, including the best fit bin.

Tables

Table 1: Top regions ($p<10^{-5}$) in the TS-EUROTRAIN GWAS.

SNP ID	Chr	P-value	A1	OR	SE	N	MAF Cases	MAF Controls	Location	KB	Nearest Genes
rs2453763	5	2.623E-08	T	0.751	0.051	25	0.314	0.373	chr5:92322427..92559372	236.946	*NR2F1-AS1*
rs2197383	3	4.681E-07	A	0.595	0.103	165	0.050	0.077	chr3:79889497..80380401	490.905	*ROBO1*
rs3773057	3	1.388E-06	T	0.198	0.335	23	0.003	0.019	chr3:29563164..29627731	64.568	*RBMS3, RBMS3-AS3*
rs9382365	6	1.829E-06	G	0.677	0.081	77	0.087	0.112	chr6:54418351..54531232	112.882	*FAM83B, TINAG*
rs152061	5	2.085E-06	T	1.26	0.049	196	0.434	0.379	chr5:64778944..64989139	210.196	*ADAMTS6, CENPK, ERBB2IP, NLN, PPWD1, SGTB, TRAPPC13, TRIM23*
rs2278796	1	6.882E-06	T	1.26	0.052	9	0.330	0.279	chr1:204951209..204971553	20.345	*CNTN2, DSTYK, NFASC, RBBP5, TMCC2, TMEM81*
rs34940828	3	7.531E-06	C	2.10	0.16	64	0.027	0.015	chr3:123213895..123398466	184.572	*ADCY5, CCDC14, MYLK, MYLK-AS1, PTPLB, SEC22A*
rs2076218	1	9.746E-06	A	1.25	0.05	3	0.380	0.340	chr1:209745395..209768699	23.305	*C1orf74, CAMK1G, DIEXF, G0S2, HSD11B1, IRF6, LAMB3, MIR205, MIR205HG, MIR4260, TRAF3IP3*

Table 2: Top regions ($p<10^{-5}$) in the TS GWAS meta-analysis (TS-EUROTRAIN and TSGWAS2).

SNP ID	Chr	P-value	OR	MAF	MAF EUR (1KG)	N	Location	KB	Nearest Genes
rs2453763	5	4.054E-08	0.83	0.358	0.347	25	chr5:92322427..92559372	236.946	*NR2F1-AS1*
rs10209244	2	6.156E-08	2.32	0.012	0.001	29	chr2:161422880..161676570	253.691	*MIR4785, RBMS1*
rs13401916	2	2.441E-07	2.08	0.014	0.003	13	chr2:161945103..162055548	110.446	*LOC100996579, LOC101929512, PSMD14, TANK, TBR1*
rs139469	22	9.997E-07	0.89	0.345	0.327	33	chr22:41451185..41627527	176.343	*ACO2, CHADL, DNAJB7, EP300, EP300-AS1, L3MBTL2, MIR1281, MIR4766, MIR6889, PHF5A, RANGAP1, RBX1, SLC25A17, ST13, TEF, TOB2, XPNPEP3, ZC3H7B*

Table 3: Significant SNP-gene pairings identified through GTEx eQTL and sQTL data

a)

Gencode Id	Gene Symbol	SNP Id	P-Value	NES	Tissue
ENSG00000175745.11	*NR2F1*	rs2453763	5.7E-10	0.25	Esophagus - Muscularis
ENSG00000237187.8	*NR2F1-AS1*	rs2453763	1.9E-7	0.22	Esophagus - Muscularis
ENSG00000251361.1	*CTD-2091N23.1*	rs2453763	2.3E-4	-0.19	Cells - Cultured fibroblasts

b)

Gencode Id	Gene Symbol	SNP Id	Intron Id	P-Value	NES	Tissue
ENSG00000251361.1	*CTD-2091N23.1*	rs2453763	93051776:93075658:clu_40848	4.6E-09	0.39	Nerve - Tibial