

## ELEVATED COMMON VARIANT GENETIC RISK FOR TOURETTE SYNDROME IN A DENSELY AFFECTED PEDIGREE

Matthew Halvorsen<sup>1</sup>, Jin Szatkiewicz<sup>1</sup>, Poorva Mudgal<sup>1</sup>, Dongmei Yu<sup>2,3</sup>, Psychiatric Genomics Consortium TS/OCD Working Group, Ashley E. Nordsletten<sup>4,5</sup>, David Mataix-Cols<sup>4</sup>, Carol A. Mathews<sup>6</sup>, Jeremiah M. Scharf<sup>2,3,7</sup>, Manuel Mattheisen<sup>4, 8-10</sup>, Mary M. Robertson<sup>11</sup>, Andrew McQuillin<sup>12a</sup>, James J. Crowley<sup>1,4,13a</sup>

<sup>1</sup> Department of Genetics, University of North Carolina at Chapel Hill, NC, USA

<sup>2</sup> Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA

<sup>3</sup> Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

<sup>4</sup> Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

<sup>5</sup> Department of Psychiatry, University of Michigan, Ann Arbor, MI, USA

<sup>6</sup> Department of Psychiatry and Genetics Institute, University of Florida, Gainesville, FL, USA

<sup>7</sup> Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

<sup>8</sup> Department of Biomedicine, Aarhus University, Aarhus, Denmark

<sup>9</sup> Center for Psychiatry Research, Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden

<sup>10</sup> Department of Psychiatry, Psychosomatics and Psychotherapy, University of Würzburg, Germany

<sup>11</sup> University College London, London, United Kingdom

<sup>12</sup> Molecular Psychiatry Laboratory, Division of Psychiatry, University College London, UK

<sup>13</sup> Department of Psychiatry, University of North Carolina at Chapel Hill, NC, USA

<sup>a</sup> Co-senior authors

Correspond with Dr. James Crowley ([crowley@unc.edu](mailto:crowley@unc.edu)) at the University of North

Carolina at Chapel Hill

### DOCUMENT INFORMATION

Submitting to:	<i>Molecular Psychiatry</i>
Running head:	Tourette pedigree study
Document:	Tourette_manuscript.docx
Date:	10/21/2021 2:41 PM
Abstract word count:	252 words
Body word count:	5,240 words
Additional:	3 figures, 2 supplemental files
Key words:	Tourette's syndrome, tic, genetic, gene, genomic, pedigree

**ABSTRACT**

Tourette syndrome (TS) is a highly heritable neuropsychiatric disorder with complex patterns of genetic inheritance. Recent genetic findings in TS have highlighted both numerous common variants with small effects and a few rare variants with moderate or large effects. Here we searched for genetic causes of TS in a large, densely-affected British pedigree using a systematic genomic approach. This pedigree spans six generations and includes 122 members, 85 of whom were individually interviewed, and 53 of whom were diagnosed as ‘cases’ (consisting of 28 with definite or probable TS, 20 with chronic multiple tics (CMT), and 5 with obsessive-compulsive behavior (OCB)). A total of 66 DNA samples were available (25 TS, 15 CMT, 4 OCB cases and 22 unaffecteds) and all were genotyped using a dense SNP array to identify shared segments, copy number variants, and to calculate genetic risk scores. Eight cases were also whole genome sequenced to test whether any rare variants were shared identical by descent. While we did not identify any notable CNVs, SNVs, indels or repeat expansions of near-Mendelian effect, the most distinctive feature of this family proved to be an unusually high load of common risk alleles for TS. We found that affecteds within this family carried a higher load of TS common variant risk similar to that of separate unrelated TS cases. Thus far, the strongest evidence we have in genetic data for contribution to risk seems to come from multiple common risk variants rather than one or a few variants of extremely strong effect.

## INTRODUCTION

While the etiology of Tourette syndrome (TS)<sup>1</sup> remains unknown, in the 1884 description of the condition that now bears his name, George Gilles de la Tourette noted a role for heredity. Today, we recognize TS as a highly heritable (~60-80%)<sup>2-4</sup> neuropsychiatric disorder characterized by recurrent, non-rhythmic tics with a childhood onset and male bias.<sup>1</sup> First-degree relatives of TS cases have a 15-25 fold increased risk of TS<sup>5, 6</sup> compared with the general population, representing one of the highest recurrence risks for common neuropsychiatric disorders.

Gene discovery for TS is in its infancy, yet as with other psychiatric disorders,<sup>7</sup> the available results suggest that much of overall genetic risk derives from a combination of many common loci of small effect and fewer rare variants of large effect. Regarding common variation, published TS GWAS<sup>8, 9</sup> to date reveal substantial SNP heritability (~30%) and genetic overlap with OCD,<sup>10</sup> though sample sizes to date are relatively small (4,819 cases and 9,488 controls; larger studies are underway). Regarding rare variation, there is now evidence for several genes with likely gene-disrupting mutations<sup>11-16</sup> or copy number variants.<sup>17-20</sup> The relative importance of common versus rare variation in TS, particularly in individual families, has yet to be determined. To begin to address this question, we evaluated a notable TS pedigree using modern genomic methods.

Large pedigrees with an unusually high rate of symptom presentation are appealing targets for psychiatric genetic study. There is already strong evidence that rare coding variants contribute to psychiatric illness risk across multiplex families, for example in autism.<sup>21</sup> More recently, there have been multiple studies focused on schizophrenia and bipolar disorder that have found that families with multiple affecteds display higher common variant genetic risk.<sup>22-24</sup>

The pedigree studied here was first described by Robertson and Gourdie in 1990.<sup>25</sup> They reported a multiply-affected six-generation kindred with one founding couple and affected family members in five generations. A total of 122 members were identified, 85 were individually interviewed, and 53 were diagnosed as ‘cases’. ‘Cases’ consisted of 28 with definite or probable TS, 20 with chronic multiple motor or vocal tics (CMT), and 5 with obsessive-compulsive behavior (OCB). Only eight of the 53 cases had previously consulted a doctor for their symptoms, with one receiving a diagnosis of TS. Many of those diagnosed later in adulthood presented with mild symptoms. For a full clinical and psychopathological description of the pedigree, see Robertson and Gourdie (1990).<sup>25</sup>

This pedigree has previously been examined by linkage and candidate gene analysis. While an initial complex segregation analysis by Curtis et al. (1992)<sup>26</sup> suggested the possibility that GTS could have been inherited as a single major gene with autosomal dominant inheritance, subsequent work by Brett et al. (1994, 1995a, b, 1996, 1997)<sup>27-31</sup> was inconsistent with this hypothesis as areas of interest on chromosome 3, chromosome 8, and several candidate genes were ruled out. A genome-wide linkage scan was performed by Curtis et al (2004)<sup>32</sup>, pointing to areas of modest interest (at least one result significant at 0.005 using two-point analysis) on chromosomes 5, 10 and 13. A subsequent factor analytic study of phenotypes from this pedigree by Robertson and Cavanna (2007)<sup>33</sup> suggested that the genetics of TS in this pedigree may be more complex than previously appreciated (i.e. more than a few segregating risk loci).

We reanalyzed this TS pedigree using modern genomic methods. All 66 available DNA samples (25 TS, 15 CMT, 4 OCB cases and 22 unaffecteds) were genotyped using a dense SNP array to identify shared segments, copy number variants, and to calculate genetic risk scores. Eight cases were also whole genome sequenced to test whether any rare variants were shared

identical by descent. These eight samples were selected from the most recent generation with DNA available and were spread across the pedigree in order to maximize the genetic distance between samples.

## **METHODS AND MATERIALS**

### Samples

All individuals provided written informed consent. The structure of the pedigree studied here is shown in *Figure 1*. We had access to a total of 66 DNA samples (25 TS, 15 CMT, 4 OCB cases and 22 unaffecteds). For a full clinical and psychopathological description of the pedigree, see Robertson and Gourdie (1990).<sup>25</sup> DNA was extracted from peripheral blood using standard methods.

### SNP array

All 66 samples were genotyped on the Illumina Multi-Ethnic Genotyping Array (MEGA) with ~1.7 million markers. Genotyping was performed at the SNP&SEQ Technology Platform at Uppsala University, Sweden and the results were analyzed using the software GenomeStudio 2011.1 from Illumina Inc. The raw dataset provided contains data from a total of 1,779,819 markers across a total of 67 samples (1 of which was a known duplicate). For downstream shared segment analysis, as quality control, we verified sex and relatedness and included only variants that were missing calls in  $\leq 2\%$  of samples, carried a Mendelian error rate  $< 0.02$  and had a MAF within the cohort of at least 0.01 ( $N = 736,144$  SNPs). All 66 distinct samples had a genotype missingness  $< 0.02$  and qualified for inclusion within the analysis.

### Shared segment analysis

We used the MEGA array genotype data to identify segments of the genome identical by descent (IBD). This was performed across all 66 samples included in the MEGA array data.

Beagle (version r1399, settings: impute=false, ibd=true) was used to collect pairwise IBD segments and DASH (v1.1.0, settings: default) was used to cluster pairwise IBD segment calls. We converted these IBD segments into PLINK bed/bim/bam files, and then used GCTA v1.92.3 to transform these data into a genetic relationship matrix (GRM) for pedigree members using the subset of segments found at  $MAF \geq 0.01$ . We used GCTA v1.92.3 again (`--mlma -maf 0.01`) to determine if any of these segments are associated with affected status within the pedigree to a multiple test corrected level of significance (ie. case/control with 44 cases and 22 controls), using the computed GRM to control for pairwise relatedness within the family. Cases are reported as having CMT, TS or OCB.

#### CNV analysis (arrays)

We used PennCNV v1.0.4 to call CNVs from genotype array data. The PennCNV script 'detect\_cnv.pl' was used to call copy number variants from sample-level B Allele Frequency (BAF) and Log R Ratio (LRR) data from all 1,779,819 markers that were on the array. We used 'filter\_cnv.pl' to subset on CNVs supported by at least 10 SNPs and have a confidence score of at least 10. We then used 'clean\_cnv.pl' (specifically, the function 'combineseg') to merge separate CNV calls on single samples that are likely part of one large single CNV, in order to form a final CNV callset. From this callset we identified a subset of 5 samples that had an LRR standard deviation or BAF drift  $> 95^{\text{th}}$  percentile within the pedigree and excluded these from further analysis, leaving 61 samples (42 affecteds, 19 unaffecteds) for analysis. We note that of these 5 samples, 4 were  $>95^{\text{th}}$  percentile for both LRR standard deviation and BAF drift, offering strong support for their exclusion from analysis.

Using the CNV callset, we performed standard CNV association testing, and also looked for specific CNV calls associated with a psychiatric phenotype from prior studies. We performed

CNV case/control association testing using PLINK v1.07, using 100,000 permutations and a seed value of 892523. We understood that this analysis was particularly underpowered, and thus took note of any CNVs with an association p-value  $< 0.05$ . We also examined in detail any CNV that carried a reciprocal overlap greater than 0.5 with an entry from a list of 54 previously identified neurodevelopmental CNVs.<sup>34</sup>

### Whole genome sequencing (WGS)

A total of 8 cases were sequenced on the Illumina HiSeq X to a mean coverage of 36X with paired-end 150 bp reads. Sequencing was performed at the SNP&SEQ Technology Platform at Uppsala University, Sweden. We used the BAM and genomic VCF files produced by the Uppsala University bioinformatics core, aligned on GRCh37 and produced following GATK best practices, for downstream analyses. We performed joint calling using the genomic VCFs followed by variant quality score recalibration, once again according to Broad GATK best practices workflows. As quality control, we tested for concordance between array-based variant genotypes and sequence-based genotypes, which would be consistent with sample concordance. Disagreement between the genotypes would suggest that sample switching had occurred in WGS data.

### CNV analysis (WGS)

To call structural variants (deletions, duplications, inversions and translocations) from WGS data, we used Delly2 v0.7.7. For the calling of each distinct SV type we followed the suggested workflow for germline SV calling by the developers: 1) single sample calling using developer-provided exclusion regions for hg19, and using the 1kg\_human\_ref.fa file; 2) merging of sample level genotype calls into a unified site list with constraints  $\text{min\_sv\_size} = 500$ ,  $\text{max\_sv\_size} = 1000000$ ,  $\text{max\_breakpoint\_offset} = 500$ ,  $\text{min\_reciprocal\_overlap} = 0.5$ ; 3) re-

genotyping each single sample using the unified site list; 4) merging of re-genotyped single-sample call BCFs to get one merged call BCF; and 5) application of the germline SV filter to the merged call BCF. We screened for CNVs (deletions, duplications, translocations and inversions) that overlapped exons from coding genes in Ensembl 37.75, and for deletions and duplications, did not overlap by more than 50% with CNVs found in > 1% of the general population. We screened for CNVs that were absent from 9 internal non-pedigree samples and were found in at least 1 of the 8 index cases. We also screened for CNVs that had reciprocal overlap with the previously described 54 neurodevelopmental CNV regions<sup>34</sup> and were found in at least one sample.

#### Repeat expansion analysis (WGS)

We screened the 8 WGS samples for shared previously reported pathogenic repeat expansions using TREDPARSE v0.7.8.<sup>35</sup> A total of 30 tri-nucleotide repeat expansion loci were screened for the presence of an expansion size with prior evidence of pathogenicity. Any positive results were listed, in particular any that were shared between more than one sample. In addition, we used ExpansionHunter v2.5.3, and screened for instances of expansions at the sample level where the majority of the reads spanned the entirety of the repeat, as opposed to containing the repeat region as a subset of the read.

#### SNV/indel analysis (WGS)

For alignment and variant calling, GATK best practices were followed. Realignment, base quality score recalibration, and realignment at indel sites were performed at SciLifeLab. From genome-level VCF files we performed joint genotyping on 8 TS pedigree members alongside 9 additional samples in-batch, forming a single callset. We then performed variant quality score recalibration following GATK best practices, requiring that all variants fall within a

VQSR tranche of >99.9% for SNVs and >95% for indels. After this we performed transcript-level variant annotation on the VCF using Variant Effect Predictor v90. Finally, additional annotation of the SNV/indel VCF was carried out using vcfanno v0.2.6, and included global and subpopulation allele frequency across exome and genome samples in gnomAD v2.0.2, as well as noncoding constraint percentiles derived from human genome data.<sup>36</sup> We prioritized loss of function coding variation based on the probability of loss-of-function intolerance (pLI) from the nonpsychiatric subset of ExAC v0.3.1,<sup>37</sup> and missense coding variation based on the MPC score (<https://doi.org/10.1101/148353>).

We screened for coding and noncoding variants with potential for contribution of risk to TS based on sharing between sequenced index cases to a degree that defied expectation. Three analyses of rare variation were constructed: 1) a genome-level screen for variants with evidence of excess sharing among index cases at a level of significance that survives a multiple test threshold, and 2) tests for an excess of rare nonsynonymous variant sharing relative to synonymous variants of similar MAF, and 3) tests for an excess of rare constrained (context-dependent tolerance score or CDTS<1%) noncoding variant sharing relative to unconstrained (CDTS>=1%) noncoding variant sharing. P-values from variant-level association testing were derived from the R package RVS.<sup>38</sup> We merged genomewide tests for oversharing with data from WGS deletions, duplications, translocations and inversions described previously. We also catalogued all coding SNVs and indels that were rare in all populations in gnomAD (gnomAD v2.0.2 popmax MAF < 0.0001), found in at least one sequenced case and carried strong evidence for being subject to purifying selection, either based on being loss of function within a gene that is highly loss of function intolerant (pLI > 0.995) or based on being a missense variant with an MPC score >= 2.

### Genetic risk scores (GRS)

In order to generate discovery and target TS datasets for GRS analyses, we split the individual-level data from Yu et al<sup>9</sup> into two datasets. The first comprised 2108 cases and 5726 controls and was used for the discovery GWAS. The second comprised 2191 cases and 2848 controls and was used as the target sample. To process the pedigree data the same way as the unaffected and TS sample, we first redid variant and sample-level QC using the ricopili function 'preimp\_dir\_12' on default settings resulting in a dataset of 762,276 variants across all 66 pedigree samples. We then used the ricopili 'impute\_dirsub\_57' function on the dataset to impute common variation using 1000 genomes phase 3 as a reference panel, just as was done with external samples. Next, we subsetted on samples that could be classified as European-ancestry according to PEDDY v0.4.2, resulting in a target sample set of 2832 unaffected controls, 2171 TS-affected cases, 22 unaffected pedigree members and 43 affected pedigree members. We also utilized published GWAS summary statistics for OCD<sup>39</sup> and ADHD<sup>40</sup>, consisting of case/control cohort sizes of 2688/7037 and 19,099/43,194 respectively.

Genetic risk score calculation was done using PRSice v2.2.6. Before proceeding with GRS calculation, we first cleaned our target genotype data using the following PRSice-recommended filter settings in PLINK : MAF  $\geq$  0.05; individual-level missingness  $<$  0.1; variant-level missingness  $<$  0.1; Hardy-Weinberg equilibrium  $P >$   $1 \times 10^{-6}$ ; exclude regions in MHC region (hg19 chr6:26,000,000-33,000,000). We ran all PRSice GRS calculations using the following settings: --base SUMSTATS\_FILE --seed 8653426 --target TARGET\_GENOTYPES - -keep SAMPLES\_KEEP\_FIDIID --binary-target T --fastscore --score avg --no-regress --stat OR. For a given GWAS summary statistic dataset, we derived the optimal p-value for partitioning unrelated TS cases and unrelated controls using PRSice with option '- -perm 10000' for a p-

value derived from 10,000 permutations of case/control status. These p-value thresholds were 0.1642, 0.1054 and 0.2964 for TS, OCD and ADHD respectively. We then split the set of lead SNPs into two groups: 1) SNPs with  $P_T \geq 0.5$ , and 2) SNPs with  $P_T \leq$  derived threshold for GRS. The first group was meant to act as a negative control that might capture instances where differences between external samples and pedigree samples are due to some bias that is not biological in nature, while the second group represents the set of SNPs where most real common risk variation might reside. For each group we calculated GRS as above, with the option `--extract QUALIFYING_SNP_FILE` added.

All normalization and analysis of computed GRS values were done using R v3.4.3. For a given set of GRS values relevant to a phenotype, we Z-transformed the values specifically for the subset of samples that were in the pedigree or in the set of unrelated TS cases or population controls. We conducted multiple-group comparisons with integrated multiple correction using the Tukey honest significant difference test, implemented in R via the function ‘TukeyHSD’. Plotting was done using ggplot2 (Wickham H (2016). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org/>), with code for raincloud plots taken from <https://micahallen.org/2018/03/15/introducing-raincloud-plots/> (<https://wellcomeopenresearch.org/articles/4-63/v1>).

For comparisons of GRS that effectively controlled for pairwise relatedness, we used an approach described previously<sup>22</sup> that relies on the R package GMMAT.<sup>41</sup> We formed a kinship matrix across all pedigree samples and all external control samples based on autosomal SNPs from the previously described set of genotypes that have a MAF of at least 0.05 in the full cohort. In R, we then used the GMMAT function `glmm.wald()` to form a logistic mixed model where for a given binary comparison, membership of one group versus the other was the

outcome, and the only listed fixed effect was gender. The kinship relationship matrix for included samples was provided as a random effect in the model. As described in<sup>22</sup> we fitted this model by maximum likelihood using Nelder–Mead optimization. To make the fitting tractable, we selected a set of 260 random samples from the control cohort set (based on retaining a 1:4 pedigree/control sample ratio) for use in our analysis versus the 65 assessable samples in the pedigree dataset.

## **RESULTS**

### Quality Control for array and WGS data

The genotype array data proved to be high quality. Genotype concordance between duplicates in the data was 99.99%. We used PLINK to assess sequence-derived sex and pairwise relatedness in array data. We found that all 66 genotyped family member sexes were concordant with those reported in the provided family pedigree. We next performed pairwise relatedness with the plink ‘--genome’ function on LD-pruned variants with MAF > 10% (see Methods and Materials), and verified that all relationships reported in the pedigree up to the second degree are supported by the array data.

Like the array data, WGS data also was of high quality. On average, 98.9% of reads per sample were successfully aligned to the reference and mean coverage was 36.1X (range: 31.4X-39.0X). We used PLINK to perform relatedness checks between WGS data and reportedly matching array data, using genotypes from 50,108 easily-intersected variants between the data. All WGS/array sample pairs reported as coming from the same individual showed near 100% relatedness, consistent with the DNA from these pairs coming from a single individual.

### Shared segments in array data

Using high density array data, we identified 57,533 total distinct shared segments (ie. haplotypes) within the pedigree. We tested for the presence of a shared segment found at a higher dosage in affected family members (CMT, OCB, TS) at a significance level that survives multiple test correction ( $0.05 / 57,533 = 8.7 \times 10^{-7}$ ). To accomplish this, we utilized a mixed model approach that takes familial pairwise relationships as covariates in comparing haplotype dosage, as described elsewhere.<sup>23</sup>

We were unable to identify any shared haplotype that passed the preset significance threshold (Figure 2A, Supplemental Table 1). The region that came closest was a segment which maps to chr7:34282415 ( $P = 1.7 \times 10^{-4}$ ). The genes nearby (*BMPER* and *NPSRI*) have no reported relevance to neuropsychiatric phenotypes. These null results are consistent with the absence of a single fully penetrant variant that might explain the high rate of TS within the pedigree.

#### CNV screen in array data

We called CNVs from array intensity data using PennCNV v1.0.4, and tested for evidence of overrepresentation of single CNVs in affecteds, as well as the presence of CNVs with known neuropsychiatric consequence.<sup>34</sup> We focused the analysis on a subset of 6513 QC-passing calls found within 61 QC-passing samples (42 affecteds, 19 unaffecteds). We first used PLINK v1.07 to determine if there is a CNV where the degree of elevated frequency in affecteds versus unaffecteds passes the multiple test corrected threshold. Here, since we were able to detect a total of 12,037 distinct markers for CNVs found at least once across the pedigree, we set the threshold as  $0.05 / 12,037 = 4.2 \times 10^{-6}$ . We were again unable to find a CNV that passed the preset threshold (Figure 2B, Supplemental Table 2). The most significant results were duplications overlapping the genes *SLC2A3* and *SLC2A14*, neither of which carry any known

relevance to neuropsychiatric phenotypes. We next screened the CNV callset for overlap with a set of 53 previously described neurodevelopmental CNVs. We detected 2 overlapping calls mapping to a single parent-child transmission of a 450kb 15q11.2 deletion. Given that this is the only observance of this CNV in the cohort, and that both parent and child are reported as unaffected, it seems unlikely that this CNV is relevant to the phenotype.

#### Oversharing of SNVs, indels and SVs in WGS data

For the 8 affected cases that we whole genome sequenced, we used the RVS package in R to compute, given the pedigree, the probability of the sharing pattern observed by chance. All CNVs that overlapped a coding base were included, while only SNVs and indels that were either nonsynonymous or in a region with CDTS < 5% were considered. We weighted this probability for SNVs and indels by their frequency in gnomAD v2.0.2, and corrected for the total number of variants assessed in this way (5349 total), searched for a variant where the degree of oversharing passed the multiple test corrected threshold ( $0.05 / 5349 = 9.3 \times 10^{-6}$ ). We were unable to identify a single variant where the probability of the sharing occurred by chance passed the significance threshold. The most significant instance of oversharing was a noncoding SNV within an intronic region of the protein-coding gene *WDR37* (4 of 8 pedigree samples,  $p=2.8 \times 10^{-5}$ ). The most significant oversharing of a CNV is that of a duplication overlapping the gene *CGNLI* (4 of 8 pedigree samples,  $p=1 \times 10^{-4}$ ).

We did identify 3 variants that were found in 5 of 8 samples. One was a noncoding SNV in the 5' untranslated region of the gene *BBS10*, another was a noncoding SNV in the 3' untranslated region of the gene *SLC2A4RG*, and the final variant was a missense SNV in *RTELI* gene. None of the overlapping protein-coding genes have known psychiatric relevance. Only *RTELI* showed particularly high gene expression in brain in GTEx.

### Sharing of damaging SNVs and indels in WGS data

While there is no compelling evidence in the data for a single penetrant variant that is found among all 8 index cases that underwent WGS, we also tested the hypothesis that there may be damaging rare variants in the pedigree that altogether are overshared to a greater degree than expected by chance. For a given selection of genomic bases that fit a broad classification of annotation (coding, noncoding) we used a corresponding set of variant level annotations that could be considered neutral within these spaces (synonymous variation and variants in regions with CDTS > 1%, respectively). We compared the ratio of sharing/nonsharing of variants that are rare (gnomAD popmax MAF < 0.0001) with a deleterious classification to that of neutral variation, since these variants best represent patterns of sharing specific to this pedigree.

We noted some oversharing of loss of function coding variation relative to presumably neutral synonymous variation (Odds Ratio=3.39,  $p=0.009$ , Supplemental Figure S1). There was no significant evidence for oversharing of noncoding variation in constrained versus unconstrained regions, or for missense variation relative to synonymous variation. Oversharing of loss of function variants would be consistent with a non-zero amount of these variants contributing to total genetic risk for TS for at least a fraction of cases. Despite this, there is no single variant with deleterious annotation that is found in all 8 index cases that has potential to explain the heritability within this pedigree. Ultimately, these data do not offer compelling evidence that there are discernable shared rare variants that contribute to TS risk within this particular pedigree.

### Damaging coding SNVs and indels present in the pedigree

We identified a total of 27 variants that could be classified as damaging coding variants (Table S3). Of these, 23 could be classified as damaging based on having missense annotation

with an MPC score  $\geq 2$ , while 4 could be classified as damaging based on having loss of function annotation and falling within a gene that has a pLI score in the nonpsychiatric subset of ExAC v0.3.1  $> 0.995$ . Of these, only 2 were shared, a damaging missense variant in *ALKBH6* (carried by 3 of 8 index cases) and a damaging missense variant in *PDE4A* (found in 2 of 8 index cases).

#### Genetic risk scores of unrelated TS cases and healthy controls

As a final analysis of the genetic variation within the pedigree, we tested the hypothesis that the family carries a burden of genetic risk on the more common end of the allele frequency spectrum, detectable through computation of common variant genetic risk scores (GRS). We calculated sample-level GRS values for TS, OCD and ADHD across a European ancestry cohort consisting of 65 genotyped pedigree members (43 affected, 22 unaffected), 2171 unrelated TS cases and 2832 healthy controls. See Methods and Materials for details.

Before proceeding, we checked for evidence of any sort of systemic stratification of pedigree and unrelated case/control data, and were unable to find any. We selected a subset of 13 unrelated married-in members of the pedigree and constructed a principal components analysis (PCA) of genetic variation with MAF  $> 0.05$  in these individuals alongside unrelated TS cases and unaffected controls. Since the 13 pedigree members are unrelated we do not expect that the PCA will be impacted by any patterns of relatedness, and should only be influenced by ethnicity. We found no evidence for stratification of these individuals and the rest of the unrelated case/control cohort across 20 principal components (Supplemental Figure S2). We also conducted PCA on the unrelated case/control cohort with a representative in-family member (sample 14, close in relationship to pedigree founders) and do not see any separation of the in-family member from other unrelated samples across PCs (Supplemental Figure S3).

We constructed a set of GRS comparisons meant to 1) test the null hypothesis of an absence of genetic signal in a scenario where GRS calculations only include flat summary statistics ( $P \geq 0.5$ ), and 2) test the null hypothesis that there is an absence of group differences with regards to GRS calculated from meaningful summary statistics (P-value selected using the permutation procedure in PRSice v2.2.6, set as 0.1642, 0.2964 and 0.1054 for TS, OCD and ADHD respectively). We partition the samples into four different groups, 1) population controls, 2) unrelated TS cases, 3) unaffected pedigree samples and 4) affected pedigree samples. If sample GRS values are comparable and there are no unknown covariates biasing GRS values, then the values computed using null summary statistics should be comparable between cohorts. Indeed we find this to be the case for all 3 comparisons (Figure 3, Supplemental Figures S4 and S5).

When using GRS values computed from clumps with  $P < 0.5$ , we found substantial differences between groups. We find that as expected, TS cases have a higher GRS than population controls (difference = 0.30, Tukey's HSD adjusted  $P < 1 \times 10^{-8}$ ). We failed to find statistically significant evidence that unaffected family members have a higher load than controls (difference = 0.42, adjusted  $P = 0.19$ ). We do however observe a higher level of GRS in affected family members than controls (difference = 0.79, adjusted  $P = 1.3 \times 10^{-6}$ ). We fail to observe a statistically significant difference in GRS between affected and unaffected family members, potentially due to low sample size (difference = 0.37, adjusted  $P = 0.49$ ).

Similar analyses of the load of common risk variation for OCD and ADHD did not yield similar results. We did not find significant evidence of elevated OCD GRS in affected pedigree samples relative to controls (difference = -0.08, adjusted  $P = 0.96$ , Supplemental Figure S4). We found that pedigree affected samples had unusually low ADHD GRS values relative to controls

(difference = -1.01,  $P = 2 \times 10^{-10}$ , Supplemental Figure S5). This is surprising given the genetic correlation between ADHD and TS, but could simply reflect the genetic profile of this family primarily being centered on TS risk. Interestingly, unaffected pedigree samples do not display a significant difference in ADHD GRS from controls (difference = -0.1,  $P = 0.97$ ).

We performed several follow-up analyses to determine if the elevation in GRS among pedigree affecteds could be explained via upstream batch effects or biasing that is technical in nature, and were unable to identify any. We repeated the GRS procedure with variants that had  $MAF > 0.20$  in unaffected samples from the TS GWAS sumstats to determine if the GRS elevation was limited to rare variation, and found that there was still a clear elevated GRS in affected pedigree members relative to controls (difference = 0.61,  $P = 4.3 \times 10^{-5}$ ). To determine if the results are consistent when a different means of computing GRS is utilized, we used PRS-CS<sup>42</sup> to compute SNP-level weights followed by PLINK to compute GRS based on these weights. The results of a comparison between pedigree affecteds and unrelated controls with regards to TS GRS was consistent with the PRSice result described earlier (difference = 0.67,  $P = 5.1 \times 10^{-5}$ ).

Finally, it was important to compile a comparison of GRS values between pedigree samples and unaffected controls in a manner that controls for pairwise relatedness between all samples, since in this comparison, essentially all cases are from the same family. To do this we followed a protocol described in separate work<sup>22</sup> that relies on the R package GMMAT<sup>41</sup> for forming and analyzing a logistic mixed model. In the model we defined as the outcome membership within the group being tested, defined sex as a fixed effect, and defined the genetic relationship matrix between all included samples as a random effect. As with before we found that samples with higher GRS are more likely to be cases within the pedigree than they are to be

controls outside of the pedigree, even taking pairwise relatedness into account (OR=2.03, 95% CI=1.15-3.58, P=0.01). This difference was significant after adjusting for a false discovery rate of 0.1, given the number of comparisons conducted in total (Supplemental Figure 6).

## DISCUSSION

Our comprehensive analyses did not conclusively identify any rare variants of strong effect in this British pedigree. However, we cannot exclude the presence of risk variants with more complex inheritance patterns, variants with more cryptic functional effects, or variants missed due to coverage or individuals sequenced. Nonetheless, our results suggest that the presence of many common TS risk alleles may partly explain why there is a high incidence of tic disorders in this pedigree.

Prior studies of large pedigrees with other psychiatric disorders have found similar results. Collins et al<sup>43</sup> analyzed a large Spanish pedigree densely affected by bipolar disorder and found no rare variants of strong effect, but an unusually high load of common risk alleles for bipolar disorder. Szatkiewicz et al<sup>23</sup> genomically characterized a large pedigree from Northern Sweden with a high density of psychotic disorders and likewise found no rare variants of strong effect, but rather an elevated common variant risk profile. Diniz et al<sup>44</sup> studied a large, multigenerational pedigree from a rural village in Brazil with a high prevalence of mood disorders and noted a substantial role for common genetic variation, given the strength of linkage signals observed.

One possible explanation for the accumulation of common risk alleles in a family is non-random mating, where partners with similar phenotypes are more likely to mate. In support of this idea, Nordsletten et al<sup>45</sup> found pervasive patterns of nonrandom mating within and across 11 major psychiatric disorders, including OCD (tic disorders were not reported due to insufficient

power for analysis). Furthermore, in the Brazilian pedigree mentioned above, de Jong et al<sup>24</sup> found that polygenic risk contributions from married-in individuals, caused by non-random mating on phenotype, appeared to increase risk for psychiatric disorders. Finally, Kurlan et al<sup>46</sup> observed bilineal (from maternal and paternal sides) transmission in TS families.

In conclusion, we systematically assessed this large pedigree dense with TS, CMT and OCB for genetic variants of strong effect. This comprehensive analysis did not conclusively identify any CNVs, SNVs, indels or repeat expansions of near-Mendelian effect. However, the common variant risk profiles of cases in this pedigree suggest that the etiology of TS, CMT and OCB in this pedigree is more related to the cumulative burden of TS genetic risk rather than one or a few variants of extremely strong effect.

#### **PGC TS/OCD WORKING GROUP**

Members of the PGC TS/OCD Working Group are:

Dongmei Yu, M.S., Jae Hoon Sul, Ph.D., Fotis Tsetsos, Ph.D., Muhammad S. Nawaz, Ph.D., Alden Y. Huang, Ph.D., Ivette Zelaya, Ph.D., Cornelia Illmann, Ph.D., Lisa Osiecki, B.A., Sabrina M. Darrow, Ph.D., Matthew E. Hirschtritt, M.D., M.P.H., Erica Greenberg, M.D., Kirsten R. Muller-Vahl, M.D., Manfred Stuhmann, M.D., Yves Dion, M.D., Guy Rouleau, M.D., Harald Aschauer, M.D., Mara Stamenkovic, M.D., Monika Schlögelhofer, M.A., Paul Sandor, M.D., Cathy L. Barr, Ph.D., Marco Grados, M.D., M.P.H., Harvey S. Singer, M.D., Markus M. Nöthen, M.D., Johannes Hebebrand, M.D., Anke Hinney, Ph.D., Robert A. King, M.D., Thomas V. Fernandez, M.D., Ph.D., Csaba Barta, Ph.D., Zsanett Tarnok, Ph.D., Peter Nagy, M.D., Christel Depienne, Ph.D., Yulia Worbe, M.D., Ph.D., Andreas Hartmann, M.D., Cathy L. Budman, M.D., Renata Rizzo, M.D., Ph.D., Gholson J. Lyon, M.D., Ph.D., William M.

McMahon, M.D., James R. Batterson, M.D., Danielle C. Cath, M.D., Irene A. Malaty, M.D., Michael S. Okun, M.D., Cheston Berlin, M.D., Douglas W. Woods, Ph.D., Paul C. Lee, M.D., M.P.H., Joseph Jankovic, M.D., Mary M. Robertson, M.D., Ph.D., Donald L. Gilbert, M.D., Lawrence W. Brown, M.D., Barbara J. Coffey, M.D., Andrea Dietrich, Ph.D., Pieter J. Hoekstra, M.D., Ph.D., Samuel Kuperman, M.D., Samuel H. Zinner, M.D., Pétur Luðvigsson, M.D., Evald Sæmundsen, M.D., Ólafur Thorarensen, M.D., Gil Atzmon, Ph.D., Nir Barzilai, M.D., Michael Wagner, Ph.D., Rainald Moessner, M.D., Roel Ophoff, Ph.D., Carlos N. Pato, M.D., Ph.D., Michele T. Pato, M.D., James A. Knowles, M.D., Ph.D., Joshua L. Roffman, M.D., Jordan W. Smoller, M.D., Sc.D., Randy L. Buckner, Ph.D., A. Jeremy Willsey, Ph.D., Jay A. Tischfield, Ph.D., Gary A. Heiman, Ph.D., Hreinn Stefansson, Ph.D., Kári Stefansson, M.D., Danielle Posthuma, Ph.D., Nancy J. Cox, Ph.D., David L. Pauls, Ph.D., Nelson B. Freimer, M.D., Benjamin M. Neale, Ph.D., Lea K. Davis, Ph.D., Peristera Paschou, Ph.D., Giovanni Coppola, M.D., Carol A. Mathews, M.D., Jeremiah M. Scharf, M.D., Ph.D.

## ACKNOWLEDGEMENTS

Genotyping and sequencing was performed by the SNP&SEQ Technology Platform in Uppsala, Sweden. The facility is part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. Dr. Crowley was supported by NIH grants R01MH105500 and R01MH110427 and the Foundation of Hope. We finally wish to thank once again, the pedigree members, albeit years ago, for their extraordinary help and cooperation with the interviews, examination and venepunctures.

## FINANCIAL DISCLOSURES

Prof Mataix-Cols reports receiving personal fees for editorial tasks (Elsevier) and contributing articles (UpToDate, Inc), both outside the submitted work. Dr. Mathews has received fees from W.W. Norton and Co. for a book on hoarding disorder, also outside the submitted work. The remaining authors report no biomedical financial interests or potential conflicts of interest.

## FIGURE LEGENDS

**Figure 1.** The pedigree for the family being studied. The plus signs indicate individuals where we were able to obtain DNA and conduct array-based genotyping. Individuals being pointed to with an arrow were subjected to whole genome sequencing. We highlight inbreeding loops which represent potential means through which recessive genotypes or polygenic risk for Tourette syndrome could accumulate.

**Figure 2.** Manhattan plots for single locus and variant tests in affected pedigree members versus unaffected controls. We were unable to identify a single array IBD segment (top), array CNV (middle) or WGS SNV/indel/CNV (bottom) which was found in cases relative to controls at a degree of significance that survives multiple testing. This is not to say that there is no contribution of risk from rare variants within this particular pedigree, but rather that we lack evidence for a particular single rare variant contributing to risk in the pedigree.

**Figure 3.** Comparisons of common variant genetic risk scores for Tourette syndrome in unrelated controls (n=2832), unrelated TS cases (n=2171), unaffected family members (n=22) and affected family members (n=43). There is no difference between groups with regards to a 'null' TS GRS, computed using  $P_T > 0.5$  (top). We defined  $P_T \leq 0.1642$  using PRSice, derived as the threshold that maximized the difference between unrelated controls and TS cases, and detected a significant elevation of TS GRS in affected family members (bottom).

## REFERENCES

1. Leckman JF. Tourette's syndrome. *Lancet* 2002; **360**(9345): 1577-1586.
2. Price RA, Kidd KK, Cohen DJ, Pauls DL, Leckman JF. A twin study of Tourette syndrome. *Arch Gen Psychiatry* 1985; **42**(8): 815-820.

3. Deng H, Gao K, Jankovic J. The genetics of Tourette syndrome. *Nat Rev Neurol* 2012; **8**(4): 203-213.
4. O'Rourke JA, Scharf JM, Yu D, Pauls DL. The genetics of Tourette syndrome: a review. *J Psychosom Res* 2009; **67**(6): 533-545.
5. Mataix-Cols D, Isomura K, Perez-Vigil A, Chang Z, Ruck C, Larsson KJ *et al*. Familial Risks of Tourette Syndrome and Chronic Tic Disorders. A Population-Based Cohort Study. *JAMA Psychiatry* 2015; **72**(8): 787-793.
6. Browne HA, Hansen SN, Buxbaum JD, Gair SL, Nissen JB, Nikolajsen KH *et al*. Familial clustering of tic disorders and obsessive-compulsive disorder. *JAMA Psychiatry* 2015; **72**(4): 359-366.
7. Sullivan PF, Daly MJ, O'Donovan M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. *Nat Rev Genet* 2012; **13**(8): 537-551.
8. Scharf JM, Yu D, Mathews CA, Neale BM, Stewart SE, Fagerness JA *et al*. Genome-wide association study of Tourette's syndrome. *Mol Psychiatry* 2013; **18**(6): 721-728.
9. Yu D, Sul JH, Tsetsos F, Nawaz MS, Huang AY, Zelaya I *et al*. Interrogating the Genetic Determinants of Tourette's Syndrome and Other Tic Disorders Through Genome-Wide Association Studies. *Am J Psychiatry* 2019; **176**(3): 217-227.
10. Davis LK, Yu D, Keenan CL, Gamazon ER, Konkashbaev AI, Derks EM *et al*. Partitioning the heritability of Tourette syndrome and obsessive compulsive disorder reveals differences in genetic architecture. *PLoS Genet* 2013; **9**(10): e1003864.
11. 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.
12. Castellan Baldan L, Williams KA, Gallezot JD, Pogorelov V, Rapanelli M, Crowley M *et al*. Histidine decarboxylase deficiency causes tourette syndrome: parallel findings in humans and mice. *Neuron* 2014; **81**(1): 77-90.

13. 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.
14. O'Roak BJ, Morgan TM, Fishman DO, Saus E, Alonso P, Gratacos M *et al.* Additional support for the association of SLITRK1 var321 and Tourette syndrome. *Mol Psychiatry* 2010; **15**(5): 447-450.
15. 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.
16. Willsey AJ, Fernandez TV, Yu D, King RA, Dietrich A, Xing J *et al.* De Novo Coding Variants Are Strongly Associated with Tourette Disorder. *Neuron* 2017; **94**(3): 486-499 e489.
17. Sundaram SK, Huq AM, Wilson BJ, Chugani HT. Tourette syndrome is associated with recurrent exonic copy number variants. *Neurology* 2010; **74**(20): 1583-1590.
18. 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.
19. Robertson MM, Shelley BP, Dalwai S, Brewer C, Critchley HD. A patient with both Gilles de la Tourette's syndrome and chromosome 22q11 deletion syndrome: clue to the genetics of Gilles de la Tourette's syndrome? *J Psychosom Res* 2006; **61**(3): 365-368.
20. Huang AY, Yu D, Davis LK, Sul JH, Tsetsos F, Ramensky V *et al.* Rare Copy Number Variants in NRXN1 and CNTN6 Increase Risk for Tourette Syndrome. *Neuron* 2017; **94**(6): 1101-1111 e1107.
21. Ruzzo EK, Perez-Cano L, Jung JY, Wang LK, Kashef-Haghighi D, Hartl C *et al.* Inherited and De Novo Genetic Risk for Autism Impacts Shared Networks. *Cell* 2019; **178**(4): 850-866 e826.
22. Andlauer TFM, Guzman-Parra J, Streit F, Strohmaier J, Gonzalez MJ, Gil Flores S *et al.* Bipolar multiplex families have an increased burden of common risk variants for psychiatric disorders. *Mol Psychiatry* 2019.

23. Szatkiewicz J, Crowley JJ, Adolfsson AN, Aberg KA, Alaerts M, Genovese G *et al.* The genomics of major psychiatric disorders in a large pedigree from Northern Sweden. *Translational psychiatry* 2019; **9**(1): 60.
24. de Jong S, Diniz MJA, Saloma A, Gadelha A, Santoro ML, Ota VK *et al.* Applying polygenic risk scoring for psychiatric disorders to a large family with bipolar disorder and major depressive disorder. *Commun Biol* 2018; **1**: 163.
25. Robertson MM, Gourdie A. Familial Tourette's syndrome in a large British pedigree. Associated psychopathology, severity, and potential for linkage analysis. *Br J Psychiatry* 1990; **156**: 515-521.
26. Curtis D, Robertson MM, Gurling HM. Autosomal dominant gene transmission in a large kindred with Gilles de la Tourette syndrome. *Br J Psychiatry* 1992; **160**: 845-849.
27. Brett PM, Le Bourdelles B, See CG, Whiting PJ, Attwood J, Woodward K *et al.* Genomic cloning and localization by FISH and linkage analysis of the human gene encoding the primary subunit NMDAR1 (GRIN1) of the NMDA receptor channel. *Ann Hum Genet* 1994; **58**(Pt 2): 95-100.
28. Brett PM, Curtis D, Robertson MM, Gurling HM. Exclusion of the 5-HT1A serotonin neuroreceptor and tryptophan oxygenase genes in a large British kindred multiply affected with Tourette's syndrome, chronic motor tics, and obsessive-compulsive behavior. *Am J Psychiatry* 1995; **152**(3): 437-440.
29. Brett PM, Curtis D, Robertson MM, Gurling HM. The genetic susceptibility to Gilles de la Tourette syndrome in a large multiple affected British kindred: linkage analysis excludes a role for the genes coding for dopamine D1, D2, D3, D4, D5 receptors, dopamine beta hydroxylase, tyrosinase, and tyrosine hydroxylase. *Biol Psychiatry* 1995; **37**(8): 533-540.
30. Brett PM, Curtis D, Robertson MM, Dahlitz M, Gurling HM. Linkage analysis and exclusion of regions of chromosomes 3 and 8 in Gilles de la Tourette syndrome following the identification of a balanced reciprocal translocation 46 XY, t(3:8)(p21.3 q24.1) in a case of Tourette syndrome. *Psychiatr Genet* 1996; **6**(3): 99-105.
31. Brett PM, Curtis D, Robertson MM, Gurling HM. Neuroreceptor subunit genes and the genetic susceptibility to Gilles de la Tourette syndrome. *Biol Psychiatry* 1997; **42**(10): 941-947.

32. Curtis D, Brett P, Dearlove AM, McQuillin A, Kalsi G, Robertson MM *et al.* Genome scan of Tourette syndrome in a single large pedigree shows some support for linkage to regions of chromosomes 5, 10 and 13. *Psychiatr Genet* 2004; **14**(2): 83-87.
33. Robertson MM, Cavanna AE. The Gilles de la Tourette syndrome: a principal component factor analytic study of a large pedigree. *Psychiatr Genet* 2007; **17**(3): 143-152.
34. Kendall KM, Rees E, Bracher-Smith M, Legge S, Riglin L, Zammit S *et al.* Association of Rare Copy Number Variants With Risk of Depression. *JAMA Psychiatry* 2019.
35. Tang H, Kirkness EF, Lippert C, Biggs WH, Fabani M, Guzman E *et al.* Profiling of Short-Tandem-Repeat Disease Alleles in 12,632 Human Whole Genomes. *Am J Hum Genet* 2017; **101**(5): 700-715.
36. di Iulio J, Bartha I, Wong EHM, Yu HC, Lavrenko V, Yang D *et al.* The human noncoding genome defined by genetic diversity. *Nat Genet* 2018; **50**(3): 333-337.
37. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T *et al.* Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; **536**(7616): 285-291.
38. Sherman T, Fu J, Scharpf RB, Bureau A, Ruczinski I. Detection of rare disease variants in extended pedigrees using RVS. *Bioinformatics* 2019; **35**(14): 2509-2511.
39. International Obsessive Compulsive Disorder Foundation Genetics Collaborative and OCD Collaborative Genetics Association Studies. Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol Psychiatry* 2018; **23**(5): 1181-1188.
40. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat Genet* 2019; **51**(1): 63-75.
41. Chen H, Wang C, Conomos MP, Stilp AM, Li Z, Sofer T *et al.* Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. *Am J Hum Genet* 2016; **98**(4): 653-666.
42. Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nature communications* 2019; **10**(1): 1776.

43. Collins AL, Kim Y, Szatkiewicz JP, Bloom RJ, Hilliard CE, Quackenbush CR *et al.* Identifying bipolar disorder susceptibility loci in a densely affected pedigree. *Mol Psychiatry* 2013; **18**(12): 1245-1246.
44. Diniz MJA, Rodrigues ACS, Gadelha A, Alsabban S, Guindalini C, Paya-Cano J *et al.* Whole genome linkage analysis in a large Brazilian multigenerational family reveals distinct linkage signals for Bipolar Disorder and Depression. *bioRxiv* 2017.
45. Nordsletten AE, Larsson H, Crowley JJ, Almqvist C, Lichtenstein P, Mataix-Cols D. Patterns of Nonrandom Mating Within and Across 11 Major Psychiatric Disorders. *JAMA Psychiatry* 2016; **73**(4): 354-361.
46. Kurlan R, Eapen V, Stern J, McDermott MP, Robertson MM. Bilineal transmission in Tourette's syndrome families. *Neurology* 1994; **44**(12): 2336-2342.