

Synaptic processes and immune-related pathways implicated in Tourette Syndrome

Fotis Tsetsos, PhD¹, Dongmei Yu, MS^{2,3}, Jae Hoon Sul, PhD^{4,5}, Alden Y. Huang, PhD^{4,5,6}, Cornelia Illmann, PhD², Lisa Osiecki, BA², Sabrina M. Darrow, PhD⁷, Matthew E. Hirschtritt, MD, MPH⁷, Erica Greenberg, MD⁸, Kirsten R. Muller-Vahl, MD⁹, Manfred Stuhmann, MD¹⁰, Yves Dion, MD¹¹, Guy A. Rouleau, MD¹², Harald Aschauer, MD^{13,14}, Mara Stamenkovic, MD¹³, Monika Schlögelhofer, MA¹⁴, Paul Sandor, MD¹⁵, Cathy L. Barr, PhD¹⁶, Marco A. Grados, MD, MPH¹⁷, Harvey S. Singer, MD¹⁷, Markus M. Nöthen, MD¹⁸, Johannes Hebebrand, MD¹⁹, Anke Hinney, PhD¹⁹, Robert A. King, MD²⁰, Thomas V. Fernandez, MD²⁰, Csaba Barta, PhD²¹, Zsanett Tarnok, PhD²², Peter Nagy, MD²², Christel Depienne, PhD^{23,24}, Yulia Worbe, MD, PhD^{24,25,26,27}, Andreas Hartmann, MD^{24,25,26}, Cathy L. Budman, MD²⁸, Renata Rizzo, MD, PhD²⁹, Gholson J. Lyon, MD, PhD³⁰, William M. McMahon, MD³¹, James R. Batterson, MD³², Danielle C. Cath, MD³³, Irene A. Malaty, MD³⁴, Michael S. Okun, MD³⁴, Cheston Berlin, MD³⁵, Douglas W. Woods, PhD³⁶, Paul C. Lee, MD, MPH³⁷, Joseph Jankovic, MD³⁸, Mary M. Robertson, MD, PhD³⁹, Donald L. Gilbert, MD, MS⁴⁰, Lawrence W. Brown, MD⁴¹, Barbara J. Coffey, MD⁴², Andrea Dietrich, PhD⁴³, Pieter J. Hoekstra, MD, PhD⁴³, Samuel Kuperman, MD⁴⁴, Samuel H. Zinner, MD⁴⁵, Michael Wagner, PhD⁴⁶, James A Knowles, MD, PhD⁴⁷, A. Jeremy Willsey, PhD⁴⁸, Jay A. Tischfield, PhD⁴⁹, Gary A. Heiman, PhD⁴⁹, Nancy J. Cox, PhD⁵⁰, Nelson B. Freimer, MD^{4,5}, Benjamin M. Neale, PhD^{2,3,51}, Lea K. Davis, PhD⁵⁰, Giovanni Coppola, MD^{4,5}, Carol A. Mathews, MD⁵², Jeremiah M. Scharf, MD, PhD^{2,3,53}, Peristera Paschou, PhD⁵⁴, and on behalf of the Tourette Association of America International Consortium for Genetics, the Gilles de la Tourette GWAS Replication Initiative, the Tourette International Collaborative Genetics Study, and the Psychiatric Genomics Consortium Tourette Syndrome Working Group¹

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

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

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

⁴Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

⁵Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles, CA, USA

- ⁶*Bioinformatics Interdepartmental Program, University of California, Los Angeles, CA, USA*
- ⁷*Department of Psychiatry, UCSF Weill Institute for Neurosciences, University of California, San Francisco, CA, USA*
- ⁸*Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA*
- ⁹*Clinic of Psychiatry, Social Psychiatry, and Psychotherapy, Hannover Medical School, Hannover, Germany*
- ¹⁰*Institute of Human Genetics, Hannover Medical School, Hannover, Germany*
- ¹¹*McGill University Health Center, University of Montreal, McGill University Health Centre, Montreal, Canada*
- ¹²*Montreal Neurological Institute, Department of Neurology and Neurosurgery, McGill University, Montreal, Canada*
- ¹³*Department of Psychiatry and Psychotherapy, Medical University Vienna, Vienna, Austria*
- ¹⁴*Biopsychosocial Corporation, Vienna, Austria*
- ¹⁵*University Health Network, Youthdale Treatment Centres, and University of Toronto, Toronto, Canada*
- ¹⁶*Krembil Research Institute, University Health Network, Hospital for Sick Children, and University of Toronto, Toronto, Canada*
- ¹⁷*Johns Hopkins University School of Medicine and the Kennedy Krieger Institute, Baltimore, MD, USA*
- ¹⁸*Institute of Human Genetics, University Hospital Bonn, University of Bonn Medical School, Bonn, Germany*
- ¹⁹*Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, University Hospital Essen, University of Duisburg-Essen, Essen, Germany*
- ²⁰*Yale Child Study Center and the Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA*
- ²¹*Institute of Medical Chemistry, Molecular Biology, and Pathobiochemistry, Semmelweis University, Budapest, Hungary*
- ²²*Vadaskert Child and Adolescent Psychiatric Hospital, Budapest, Hungary*
- ²³*Institute of Human Genetics, University Hospital Essen, University Duisburg-Essen, Essen, Germany*
- ²⁴*Sorbonne Universités, UPMC Université Paris 06, UMR S 1127, CNRS UMR 7225, ICM, Paris, France*
- ²⁵*French Reference Centre for Gilles de la Tourette Syndrome, Groupe Hospitalier Pitié-Salpêtrière, Paris, France*
- ²⁶*Assistance Publique-Hôpitaux de Paris, Department of Neurology, Groupe Hospitalier Pitié-Salpêtrière, Paris, France*
- ²⁷*Assistance Publique Hôpitaux de Paris, Hopital Saint Antoine, Paris France*
- ²⁸*Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, USA*
- ²⁹*Child Neuropsychiatry, Department of Clinical and Experimental Medicine, University of Catania, Catania, Italy*
- ³⁰*Jervis Clinic, NYS Institute for Basic Research in Developmental Disabilities (IBR), Staten Island, NY, USA*
- ³¹*Department of Psychiatry, University of Utah, Salt Lake City, UT, USA*
- ³²*Children's Mercy Hospital, Kansas City, MO, USA*
- ³³*Department of Psychiatry, University Medical Center Groningen and Rijksuniversiteit Groningen, and Drenthe Mental Health Center, Groningen, the Netherlands*
- ³⁴*Department of Neurology, Norman Fixel Institute for Neurological Diseases, University of Florida Health, Gainesville, FL, USA*

³⁵ *Pennsylvania State University College of Medicine, Hershey, PA, USA*

³⁶ *Marquette University and University of Wisconsin-Milwaukee, Milwaukee, WI, USA*

³⁷ *Tripler Army Medical Center and University of Hawaii John A. Burns School of Medicine, Honolulu, HI, USA*

³⁸ *Parkinson's Disease Center and Movement Disorders Clinic, Department of Neurology, Baylor College of Medicine, Houston, TX, USA*

³⁹ *Division of Psychiatry, Department of Neuropsychiatry, University College London, UK*

⁴⁰ *Division of Pediatric Neurology, Cincinnati Children's Hospital Medical Center; Department of Pediatrics, University of Cincinnati*

⁴¹ *Children's Hospital of Philadelphia, PA, USA*

⁴² *Department of Psychiatry and Behavioral Sciences, University of Miami Miller School of Medicine, Miami, FL, USA*

⁴³ *Department of Child and Adolescent Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands*

⁴⁴ *University of Iowa Carver College of Medicine, Iowa City, IA, USA*

⁴⁵ *Department of Pediatrics, University of Washington, Seattle, WA, USA*

⁴⁶ *Department of Psychiatry and Psychotherapy, University Hospital Bonn, Bonn, Germany*

⁴⁷ *SUNY Downstate Medical Center Brooklyn, NY, USA*

⁴⁸ *Institute for Neurodegenerative Diseases, UCSF Weill Institute for Neurosciences, University of California San Francisco, San Francisco, CA, USA*

⁴⁹ *Department of Genetics and the Human Genetics Institute of New Jersey, Rutgers, the State University of New Jersey, Piscataway, NJ, USA*

⁵⁰ *Division of Genetic Medicine, Vanderbilt Genetics Institute, Vanderbilt University Medical Center, Nashville, TN, USA*

⁵¹ *Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA*

⁵² *Department of Psychiatry, Genetics Institute, University of Florida, Gainesville, FL, USA*

⁵³ *Department of Neurology, Brigham and Women's Hospital, and the Department of Neurology, Massachusetts General Hospital, Boston, MA, USA*

⁵⁴ *Department of Biological Sciences, Purdue University, West Lafayette, IN, USA*

1 **Abstract**

2 Tourette Syndrome (TS) is a neuropsychiatric disorder of complex genetic architecture involving
3 multiple interacting genes. Here, we sought to elucidate the pathways that underlie the neurobiology
4 of the disorder through genome-wide analysis. We analyzed genome-wide genotypic data of 3581
5 individuals with Tourette Syndrome (TS) and 7682 ancestry-matched controls and investigated as-
6 sociations of TS with sets of genes that are expressed in particular cell types and operate in specific
7 neuronal and glial functions. We employed a self-contained, set-based association method (SBA) as
8 well as a competitive gene set method (MAGMA) using individual-level genotype data to perform a
9 comprehensive investigation of the biological background of TS. Our SBA analysis identified three
10 significant gene sets after Bonferroni correction, implicating Ligand-gated Ion Channel Signaling,
11 Lymphocytic, and Cell Adhesion and Trans-synaptic Signaling processes. MAGMA analysis further
12 supported the involvement of the Cell Adhesion and Trans-synaptic Signaling gene set. The Lympho-
13 cytic gene set was driven by variants in *FLT3*, raising an intriguing hypothesis for the involvement of
14 a neuroinflammatory element in TS pathogenesis. The indications of involvement of Ligand-gated
15 Ion Channel Signaling reinforce the role of GABA in TS, while the association of Cell Adhesion and
16 Trans-synaptic Signaling gene set provides additional support for the role of adhesion molecules in
17 neuropsychiatric disorders.

18 Introduction

19 Tourette Syndrome (TS) is a chronic neurodevelopmental disorder characterized by several motor
20 tics and at least one vocal tic that persist more than a year [1]. Its prevalence is between 0.6-1%
21 in school-aged children [2, 3]. Although TS is highly polygenic in nature, it is also highly heritable
22 [4]. The population-based heritability is estimated at 0.7 [5, 6], with SNP-based heritability ranging
23 from 21% [7] to 58% [4] of the total. The genetic risk for TS that is derived from common variants is
24 spread throughout the genome [4]. The two genome-wide association studies (GWAS) conducted to
25 date [7, 8] suggest that TS genetic variants may be associated, in aggregate, with tissues within the
26 cortico-striatal and cortico-cerebellar circuits, and in particular, the dorsolateral prefrontal cortex.
27 The GWAS results also demonstrated significant ability to predict tic severity using TS polygenic risk
28 scores [7, 9]. A genome-wide CNV study identified rare structural variation contributing to TS on the
29 *NRXN1* and *CNTN6* genes [10]. De novo mutation analysis studies in trios have highlighted two high
30 confidence genes, *CELSR* and *WWC1*, and four probable genes, *OPAI*, *NIPBL*, *FNI*, and *FBN2* to be
31 associated with TS [11, 12].

32 Investigating clusters of genes, rather than relying on single marker tests is an approach that
33 can significantly boost power in a genomewide setting [13]. This approach has already proven
34 useful in early genome-wide studies of TS. The first published TS GWAS, which included 1285 cases
35 and 4964 ancestry-matched controls did not identify any genome-wide significant loci. However,
36 by partitioning functional- and cell type-specific genes into gene sets, an involvement of genes
37 implicated in astrocyte carbohydrate metabolism was observed, with a particular enrichment in
38 astrocyte-neuron metabolic coupling [14]. Here, we investigated further the pathways that underlie
39 the neurobiology of TS, performing gene set analysis on a much larger sample of cases with TS and
40 controls from the second wave TS GWAS. We employed both a competitive gene set analysis as
41 implemented through MAGMA, as well as a self-contained analysis through a set-based association
42 method (SBA).

43 Methods and Materials

44 Samples and quality control

45 The sample collection has been extensively described previously [7, 8]. IRB approvals and consent
46 forms were in place for all data collected and analyzed as part of this project. For the purposes of our
47 analysis, we combined 1285 cases with TS and 4964 ancestry-matched controls from the first wave

48 TS GWAS, with 2918 TS cases and 3856 ancestry-matched controls from the second wave TS GWAS.
49 Standard GWAS quality control procedures were employed [15, 16]. The data were partitioned first by
50 genotyping platform and then by ancestry. The sample call rate threshold was set to 0.98, and the
51 inbreeding coefficient threshold to 0.2. A marker call rate threshold was defined at 0.98, case-control
52 differential missingness threshold at 0.02, and Hardy-Weinberg Equilibrium (HWE) threshold to 10^{-6}
53 for controls and 10^{-10} for cases. Before merging the partitioned datasets, we performed pairwise
54 tests of association and missingness between the case-only and control-only subgroups to address
55 potential batch effect issues. All SNPs with p-values $\leq 10^{-06}$ in any of these pairwise quality control
56 analyses were removed. After merging all datasets, Principal Component Analysis was utilized to
57 remove samples that deviated more than 6 standard deviations and to ensure the homogeneity of our
58 samples in the ancestry space of the first 10 principal components, through the use of the EIGENSOFT
59 suite [17]. Identity-by-descent analysis with a threshold of 0.1875 was used to remove related samples,
60 and thus to avoid confounding by cryptic relatedness. After quality control, the final merged dataset
61 consisted of 3581 cases with TS and 7682 ancestry-matched controls on a total of 236,248 SNPs,
62 annotated using dbSNP version 137 and the hg19 genomic coordinates.

63 We assessed the genomic variation in our data through PCA analysis to identify potential popula-
64 tion structure. The variation in our data was reduced to a triangular shape in the two-dimensional
65 space of the first two principal components. One tip was occupied by Ashkenazi Jewish samples,
66 the second by the Southern European samples and the other by the North Europeans. Depicting
67 geography, the Southern to Northern axis was populated by European-ancestry samples. The first five
68 principal components were added to the association model as covariates, in order to avoid population
69 structure influencing our results.

70 **Gene sets**

71 We collected neural-related gene sets from multiple studies on pathway analyses in neuropsychiatric
72 disorders [14, 18–22]. These studies relied on an evolving list of functionally-partitioned gene sets, fo-
73 cusing mainly on neural gene sets, including synaptic, glial sets, and neural cell-associated processes.
74 We added a lymphocytic gene set also described in these studies [21], in order to also investigate
75 potential neuro-immune interactions.

76 In total, we obtained 51 gene sets, which we transcribed into NCBI entrez IDs and subsequently
77 filtered by removing gene sets that contained fewer than 10 genes. 45 gene sets fit our criteria and
78 were used to conduct the analyses.

79 We examined two primary categories of pathway analysis methods, the competitive [23] and
80 the self-contained test [24]. The competitive test compares the association signal yielded by the
81 tested gene set to the association signals that do not reside in it [23, 25]. In this type of test, the null
82 hypothesis is that the tested gene set attains the same level of association with disease as equivalent
83 random gene sets. In contrast, the self-contained test investigates associations of each tested gene
84 set with the trait, and not with other gene sets, meaning that the null hypothesis in this case is that
85 the genes in the gene set are not associated with the trait [24, 25]. Therefore, for a competitive test,
86 there should be data for the whole breadth of the genome, but this test cannot provide information
87 regarding how strongly the gene set is associated with the trait [26]. We employ both methods for a
88 comprehensive investigation into the neurobiological background of TS.

89 **MAGMA on raw genotypes**

90 We ran MAGMA [23] on the individual-level genotype data using the aforementioned filtered gene set
91 lists. MAGMA performs a three step analytic process. First it annotates the SNPs by assigning them to
92 genes, based on their chromosomal location. Then it performs a gene prioritization step, which is
93 used to perform the final gene set analysis step. We used a genomic window size of $\pm 10\text{kb}$ and the
94 top 5 principal components as covariates to capture population structure. SNP to gene assignments
95 were based on the NCBI 37.3 human gene reference build. The number of permutations required for
96 the analysis was determined by MAGMA, using an adaptive permutation procedure leading to 11,263
97 permutations. MAGMA employs a family-wise error correction calculating a significance threshold of
98 0.00100496.

99 **Set-Based Association (SBA) test**

100 We conducted SBA tests on the raw individual genotype data, as described in PLINK [24, 27] and
101 adapted in a later publication [28]. This test relies on the assignment of individual SNPs to a gene,
102 based on their position, and thus to a pathway, according to the NCBI 37.3 human gene reference
103 build. After single marker association analysis, the top LD-independent SNPs from each set are
104 retained and selected in order of decreasing statistical significance, and the mean of their association
105 p-values is calculated. We permuted the case/control status, repeating the previous association
106 and calculation steps described above, leading to the empirical p-value for each set. The absolute
107 minimum number of permutations required for crossing the significance level is dictated by the
108 number of gene sets tested. Testing for 45 gene sets requires at least 1000 permutations to produce

109 significant findings. PLINK's max(t) test recommends at least 64,000 permutations. We opted to
110 increase the number of permutations to one million, the maximum that was computationally feasible,
111 to maximize our confidence in the outcomes, given our large sample size.

112 We used logistic regression as the association model on the genotypes and the first five principal
113 components as covariates on the genotype data to conduct the SBA test with the collected neural
114 gene sets. Another repetition of this step was performed with a simple association test, to test for this
115 method's robustness to population structure. We proceeded to run the analysis on all samples, using
116 all gene sets at a 10kb genomic window size, the first five principal components as covariates, and
117 one million permutations. Since the permutations were performed on the phenotypic status of the
118 samples, and only served as a method of association of the trait with the gene sets, we also corrected
119 the results by defining the significance threshold through Bonferroni correction at 1.1×10^{-3} (0.05/45).

120 **Results**

121 For the gene set association analysis, we ran PLINK's self-contained set-based association method and
122 MAGMA's competitive association method, using the same 45 gene sets on the processed genotyped
123 data of 3581 cases and 7682 ancestry-matched controls on a total of 236,248 SNPs. By performing
124 both methods of analysis we aimed to obtain a global assessment of the gene sets' relationship with
125 TS.

126 MAGMA analysis identified one significant gene set (Figure 1), Cell Adhesion and Trans-synaptic
127 Signaling (CATS), which achieved a nominal p-value of 6.2×10^{-5} (permuted p-value of 0.0032). While
128 the CATS gene set is comprised of 83 genes, MAGMA's annotation step prioritized 72 of its genes
129 for the gene set analysis. It involves 3290 variants that were reduced to 1627 independent variants
130 in our data. Results were mainly driven by associations in the CDH26, CADM2 and OPCML genes
131 as indicated by MAGMA gene-based analysis (Table 1). In the gene-based tests, CDH26 attained a
132 p-value of 8.9526×10^{-6} , CADM2 a p-value of 4.6253×10^{-4} , and OPCML a p-value of 7.9851×10^{-4} ,
133 neither crossing the genome-wide significance threshold for gene tests (2.574×10^{-6} calculated on
134 19,427 genes contained in the NCBI37.3 version of RefGene).

135 We next ran SBA, which conducts an initial single-marker association step before performing
136 permutations to calculate empirical p-values for the gene sets. This association step is performed on
137 the total number of variants that are associated with the genes involved in the gene sets, leading to a
138 subset of 25,630 variants in our data, which are then filtered based on their LD. Analysis identified three
139 gene sets as significant (Table 2), the Ligand-gated Ion Channel Signaling (LICS) (P: 2.67×10^{-4}), the

140 Lymphocytic ($P: 3.5 \times 10^{-4}$), and the Cell Adhesion and Trans-synaptic Signaling (CATS) ($P: 1.07 \times 10^{-3}$).
141 Detailed results for all the tested gene sets are shown in Figure 2.

142 The LICS gene set was the top scoring gene set, including 38 genes and involving 683 variants,
143 66 of which were associated with TS. The gene set's signal was primarily driven by variants residing
144 in the genes of the γ -aminobutyric acid receptors *GABRG1* and *GABBR2*, the *HCN1* channel gene
145 and the glutamate receptor gene *GRIK4*. This signal was driven primarily by an association with SNP
146 rs9790873, which is an eQTL for *HCN1* in tibial nerve, according to GTEx [29]. *GABBR2* is represented
147 by two top SNPs, that are LD-independent, and removing either of those SNPs from the gene set did
148 not cause the gene set to drop under the significance threshold.

149 The Lymphocytic gene set was the next top scoring gene set, including 143 genes that translated
150 to 799 variants in our data, with 50 of these variants associated with TS. Its signal was driven by a
151 missense variant inside the *FLT3* gene and an intergenic variant between *NCR1* and *NLRP7*, followed
152 by *IL12A*, *HDAC9*, *CD180*. The rs1933437 SNP is the top variant for *FLT3*, and is a possibly damaging
153 missense variant [30], located in the sixth exon of the *FLT3* gene leading to a p.Thr227Met mutation.
154 It is a very common variant and the sixth exon appears to be less expressed than downstream exons.
155 Given the tissues in which this eQTL affects *FLT3*'s expression, we tested the Lymphocytic gene set
156 by removing *FLT3* from it, to identify whether the lymphocytic gene set association was biased by
157 the presence of *FLT3*. After removing *FLT3*, the Lymphocytic gene set association statistic decreased
158 slightly ($P: 0.00012$), driven mainly by *NCR1/NLRP7*.

159 The third significant gene set, CATS, consisted of 83 genes, including multiple large genes. CATS
160 was identified by both SBA and MAGMA in our analyses, and both gene set approaches identified
161 *CDH26* as the gene with the lowest p-value. Both SBA and MAGMA also identified *NCAM2*, *NTM* and
162 *ROBO2* as strongly associated with TS, with *NTM* represented by two LD-independent SNPs. CATS's
163 top SNP, rs1002762, resides in the *CDH26* gene on chromosome 20, and is the top associated SNP in
164 our data ($P: 2.031 \times 10^{-6}$) with an odds-ratio of 1.178.

165 Notable results from the SBA also include the Astrocyte small GTPase mediated signaling (ASGMS)
166 and the Astrocyte-neuron metabolic coupling (ANMC) gene sets, with a p-values slightly under the
167 significance thresholds. These gene sets attained a p-value of 0.00137 and 0.001504 respectively.

168 Discussion

169 Seeking to elucidate the neurobiology of TS, we present here the largest study to date aiming to
170 interrogate the involvement of sets of genes that are related to neuronal and glial function in TS. We

171 analyzed data from our recently performed TS GWAS and conducted two distinct types of testing,
172 a competitive, regression-based test (MAGMA) and a self-contained, p-value combining test (SBA).
173 Self-contained tests investigate for associations with a phenotype, while competitive tests compare a
174 specific gene set against randomly generated gene sets. We employed both methods to perform a
175 comprehensive investigation of the biological background of TS.

176 MAGMA's regression-based algorithm has been reported to account for gene size biases, as can
177 be also discerned by the variable sizes of the top genes. MAGMA's top prioritized gene, *CDH26*, is
178 represented by 4 SNPs in our data, *CADM2* by 42, while *OPCML* is represented by 210 SNPs, as it
179 covers an extensive genomic region. We addressed such issues in SBA by setting a low r^2 threshold
180 and conditioning on any LD-independent SNPs that resided on the same gene.

181 The gene sets used in our study come from a line of function-based analyses, aiming to investigate
182 neurobiological mechanisms in neuropsychiatric disorders. A previous pathway analysis using
183 individual level genotype data of the first wave TS GWAS identified genes involved in astrocytic-
184 neuron metabolic coupling, implicating astrocytes in TS pathogenesis [14]. In this study, we took
185 advantage of the increased sample size of the second wave TS GWAS and the mechanics of the
186 two distinct methods to identify gene sets associated with TS that provide a novel insight into the
187 pathogenesis of TS, and substantiate the role of neural processes in this neuropsychiatric disorder.

188 The ANMC gene set that contains genes involved in carbohydrate metabolism in astrocytes was
189 the single identified gene set in the previous pathway analysis study on TS [14], raising a hypothesis on
190 a potential mechanism that involves altered metabolism of glycogen and glutamate/ γ -aminobutyric
191 acid in the astrocytes. In our study the ANMC gene set scored slightly under the significance threshold.

192 Here, analyzing a much larger sample size we identified 3 sets of genes as significantly associated
193 to the TS phenotype. Among them the LICS gene set, which involves genes implicated in ion channel
194 signaling through γ -aminobutyric acid and glutamate. Several genes in the LICS gene set have been
195 previously implicated in neuropsychiatric phenotypes. *HCN1*, a hyperpolarization-activated cation
196 channel involved in native pacemaker currents in neurons and the heart, has been significantly
197 associated with schizophrenia and autism [31–33]. *GABRG1*, an integral membrane protein that
198 inhibits neurotransmission by binding to the benzodiazepine receptor, has yielded mild associations
199 with general cognitive ability [34] and epilepsy [35], while *GABBR2*, a g-protein coupled receptor
200 that regulates neurotransmitter release, with schizophrenia [36] and post-traumatic stress disorder
201 [37] in multiple studies. The GABA-ergic pathway has been previously implicated in TS, and recent
202 advances showcased the possibility that a GABA-ergic transmission deficit can contribute towards
203 TS symptoms[38]. *GRIK4*, encoding a glutamate-gated ionic channel, has shown associations with

204 mathematical ability and educational attainment [39] and weaker associations with attention-
205 deficit hyperactivity disorder [40]. The γ -aminobutyric acid receptors and the HCN channel, are
206 features of inhibitory interneurons [41] and also identified in the brain transcriptome of individuals
207 with TS [42], adding to the evidence that the phenotype of TS could be influenced by an inhibitory
208 circuit dysfunction, as has previously been proposed [43].

209 Individuals with TS are reported to present elevated markers of immune activation [42, 44].
210 Additionally, a number of studies have implicated neuroimmune responses with the pathogenesis of
211 TS [45–47]. We investigated neuro-immune interactions by interrogating association to a gene set
212 designed by Goudriaan et al [21] to study enrichment in lymphocytic genes. Indeed, our analysis
213 yielded a statistically significant signal. The *FLT3* association coincides with the results of the second
214 wave TS GWAS, in which *FLT3* was the only genome-wide significant hit [7]. *FLT3* and its ligand,
215 *FLT3LG*, have a known role in cellular proliferation in leukemia, and have been found to be expressed
216 in astrocytic tumors [48]. The rs1933437 variant in *FLT3* is an eQTL in the brain cortex and the
217 cerebellum [29], and has also been implicated in the age at the onset of menarche [49]. Variants in
218 *FLT3* have attained genome-wide significance in a series of studies focusing on blood attributes
219 in populations of varying ancestry, and our current insights into its role are mostly based on these
220 associations with blood cell counts, serum protein levels, hypothyroidism and autoimmune disorders
221 [49–52].

222 *FLT3* could play a role in neuroinflammation as supported by its intriguing association with
223 peripheral neuropathic pain. The inhibition of *FLT3* is reported to alleviate peripheral neuropathic
224 pain (PNP) [53], a chronic neuro-immune condition that arises from aberrations in the dorsal root
225 ganglia. Cytokines and their receptors have been at the epicenter of the neuro-immune interactions,
226 with microglia contributing significantly to chronic phenotypes of such states [54]. *FLT3* is a critical
227 component for neuro-immune interactions, especially in the case of the development and sustenance
228 of the PNP phenotype. Interestingly, pain follows sex-specific routes, with glia having a prominent
229 role for pain propagation in males, while females involve adaptive immune cells instead [55]. These,
230 paired with previous evidence of glial involvement in TS [14], raise an interesting hypothesis for
231 TS symptom sustenance, since *FLT3* has been shown to be critical for the chronicity of neuronal
232 dysregulations [53].

233 *FLT3* is expressed in the cerebellum and whole blood, while *FLT3*'s top variant, rs1933437, is
234 an eQTL for *FLT3* on GTEx [29] in various brain tissues, such as the cortex, the cerebellum, the
235 hypothalamus, the frontal cortex (BA9) and non-brain tissues, such as the skin, the pancreas and
236 adipose tissues. In order to test the robustness of the lymphocytic association in our findings, we

237 repeated the analysis after removing *FLT3* from the lymphocytic gene set. The p-value of the gene set
238 decreased, but still remained significant, due to the association in the *NCR1/NLRP7* locus. Besides
239 *FLT3*, the other genes included in this gene set are also quite intriguing to consider as potential
240 candidates that could underlie the pathophysiology of TS. In the same vein with *FLT3*, common
241 variants in *NCR1* have also been significantly associated with blood protein levels [56]. *HDAC9*
242 has been significantly associated with androgenetic alopecia [49, 57], hair color [49], and ischemic
243 stroke [58]. These seem to follow previous knowledge, given that genes involved in ischemic stroke
244 have been identified as a common component between TS and ADHD [59], and that TS, similar to
245 other neuropsychiatric disorders, demonstrates a distinct preference for males. *CD180* has shown
246 associations with general cognitive ability [34].

247 The CATS gene set involves many cadherins, with the top signals being in *CDH26*. *CDH26* is
248 a cadherin that regulates leukocyte migration, adhesion and activation, especially in the case of
249 allergic inflammation [60]. Cell adhesion molecules have been consistently implicated in phenotypes
250 pertaining to brain function, with the latest addition of the high confidence TS gene *CELSR3*, a
251 flamingo cadherin, that was identified in a large scale de novo variant study for TS [12].

252 Most of the genes contained in the identified gene sets in this study are involved in cognitive
253 performance, mathematical ability and educational attainment [39]. *OPCML*, *CADM2*, and *ROBO2*
254 have been implicated in neuromuscular and activity phenotypes, such as grip strength [61], physical
255 activity [62], and body mass index [49]. *ROBO2* has been associated with depression [63], expressive
256 vocabulary in infancy [64], while *CADM2* is associated to a multitude of phenotypes, including anxiety
257 [63], risk taking behavior, smoking [65]. *NTM* displays similar patterns of pleiotropy, associated with
258 smoking [49], myopia [57], hair color [66], anxiety [63], asperger's syndrome [67], bipolar disorder with
259 schizophrenia [68], and eating disorders [69]. *NCAM2* and *NTM*, similarly to the lymphocytic genes,
260 have been significantly associated with blood protein levels [70] and leukocyte count [49] respectively.
261 Many of these phenotypes are known TS comorbidities, presenting themselves commonly or less
262 commonly in TS cases, and other are related to functions that get impaired in TS symptomatology.

263 The CATS gene set was identified in both methods indicating the involvement of cell adhesion
264 molecules in trans-synaptic signaling. Using genotypes with both methods as a means of identifying
265 pathways instead of summary statistics, gave our study the edge of sample-specific linkage dise-
266 quilibrium rather than relying on an abstract linkage disequilibrium pattern reference. Our current
267 understanding for the regional structures of the genome and the cis effects of genomic organization
268 will aid the refinement of these associations as well as help shape our understanding of the pleiotropic
269 mechanisms in the identified loci potentially responsible for disease pathogenesis.

270 In conclusion, our analysis provides further support for the role of *FLT3* in TS, strengthens the
271 possibility for the involvement of the GABA-ergic biological pathway in TS pathogenesis, and provides
272 parallel insights into possible mechanisms underlying tic persistence, and a possible correlation with
273 glial-derived neuro-immune phenotypes.

274 **Conflict of interest**

275 IM has participated in research funded by the Parkinson Foundation, Tourette Association, Dystonia
276 Coalition, AbbVie, Biogen, Boston Scientific, Eli Lilly, Impax, Neuroderm, Prilenia, Revance, Teva
277 but has no owner interest in any pharmaceutical company. She has received travel compensation or
278 honoraria from the Tourette Association of America, Parkinson Foundation, International Association
279 of Parkinsonism and Related Disorders, Medscape, and Cleveland Clinic, and royalties for writing a
280 book with Robert rose publishers.

281 KMV has received financial or material research support from the EU (FP7-HEALTH-2011 No.
282 278367, FP7-PEOPLE-2012-ITN No. 316978), the German Research Foundation (DFG: GZ MU 1527/3-
283 1), the German Ministry of Education and Research (BMBF: 01KG1421), the National Institute of Men-
284 tal Health (NIMH), the Tourette Gesellschaft Deutschland e.V., the Else-Kroner-Fresenius-Stiftung,
285 and GW, Almirall, Abide Therapeutics, and Therapix Biosciences and has received consultant's hono-
286 raria from Abide Therapeutics, Tilray, Resalo Vertrieb GmbH, and Wayland Group, speaker's fees from
287 Tilray and Cogitando GmbH, and royalties from Medizinisch Wissenschaftliche Verlagsgesellschaft
288 Berlin, Elsevier, and Kohlhammer; and is a consultant for Nuvelution TS Pharma Inc., Zynerva Phar-
289 maceuticals, Resalo Vertrieb GmbH, CannaXan GmbH, Therapix Biosciences, Syqe, Nomovo Pharma,
290 and Columbia Care.

291 BMN is a member of the scientific advisory board at Deep Genomics and consultant for Camp4
292 Therapeutics, Takeda Pharmaceutical and Biogen.

293 MMN has received fees for memberships in Scientific Advisory Boards from the Lundbeck Foun-
294 dation and the Robert-Bosch-Stiftung, and for membership in the Medical-Scientific Editorial Office
295 of the Deutsches Ärzteblatt. MMN was reimbursed travel expenses for a conference participation by
296 Shire Deutschland GmbH. MMN receives salary payments from Life & Brain GmbH and holds shares
297 in Life & Brain GmbH. All this concerned activities outside the submitted work.

298 MSO serves as a consultant for the Parkinson's Foundation, and has received research grants from
299 NIH, Parkinson's Foundation, the Michael J. Fox Foundation, the Parkinson Alliance, Smallwood
300 Foundation, the Bachmann-Strauss Foundation, the Tourette Syndrome Association, and the UF

301 Foundation. MSO's DBS research is supported by: NIH R01 NR014852 and R01NS096008. MSO is PI of
302 the NIH R25NS108939 Training Grant. MSO has received royalties for publications with Demos, Man-
303 son, Amazon, Smashwords, Books4Patients, Perseus, Robert Rose, Oxford and Cambridge (movement
304 disorders books). MSO is an associate editor for New England Journal of Medicine Journal Watch Neu-
305 rology. MSO has participated in CME and educational activities on movement disorders sponsored
306 by the Academy for Healthcare Learning, PeerView, Prime, QuantiaMD, WebMD/Medscape, Medicus,
307 MedNet, Einstein, MedNet, Henry Stewart, American Academy of Neurology, Movement Disorders
308 Society and by Vanderbilt University. The institution and not MSO receives grants from Medtronic,
309 Abbvie, Boston Scientific, Abbott and Allergan and the PI has no financial interest in these grants.
310 MSO has participated as a site PI and/or co-I for several NIH, foundation, and industry sponsored
311 trials over the years but has not received honoraria. Research projects at the University of Florida
312 receive device and drug donations.

313 DW receives royalties for books on Tourette Syndrome with Guilford Press, Oxford University
314 Press, and Springer Press.

315 The rest of the authors declare that they have no conflict of interest.

316 **Acknowledgements**

317 This research is co-financed by Greece and the European Union (European Social Fund- ESF) through
318 the Operational Programme «Human Resources Development, Education and Lifelong Learning» in
319 the context of the project “Reinforcement of Postdoctoral Researchers - 2nd Cycle” (MIS-5033021),
320 implemented by the State Scholarships Foundation (IKY).

321 LKD was supported by grants from the National Institutes of Health including U54MD010722-04,
322 R01NS102371, R01MH113362, U01HG009086, R01MH118223, DP2HD98859, R01DC16977, R01NS105-
323 746, R56MH120736, R21HG010652, and RM1HG009034.

324 The members of the Tourette Association of America International Consortium for Genetics are
325 Cathy L. Barr, James R. Batterson, Cheston Berlin, Cathy L. Budman, Danielle C. Cath, Giovanni
326 Coppola, Nancy J. Cox, Sabrina Darrow, Lea K. Davis, Yves Dion, Nelson B. Freimer, Marco A. Grados,
327 Erica Greenberg, Matthew E. Hirschtritt, Alden Y. Huang, Cornelia Illmann, Robert A. King, Roger
328 Kurlan, James F. Leckman, Gholson J. Lyon, Irene A. Malaty, Carol A. Mathews, William M. McMahon,
329 Benjamin M. Neale, Michael S. Okun, Lisa Osiecki, Danielle Posthuma, Mary M. Robertson, Guy A.
330 Rouleau, Paul Sandor, Jeremiah M. Scharf, Harvey S. Singer, Jan Smit, Jae Hoon Sul, and Dongmei Yu.

331 The members of the Gilles de la Tourette Syndrome Genome-Wide Association Study Replication

332 Initiative are Harald Aschauer, Csaba Barta, Cathy L. Budman, Danielle C. Cath, Christel Depienne,
333 Andreas Hartmann, Johannes Hebebrand, Anastasios Konstantinidis, Carol A. Mathews, Kirsten R.
334 Muller-Vahl, Peter Nagy, Markus M. Nöthen, Peristera Paschou, Renata Rizzo, Guy A. Rouleau, Paul
335 Sandor, Jeremiah M. Scharf, Monika Schlögelhofer, Mara Stamenkovic, Manfred Stuhmann, Fotis
336 Tsetsos, Zsanett Tarnok, Tomasz Wolanczyk, and Yulia Worbe.

337 The members of the Tourette International Collaborative Genetics consortium are Lawrence W.
338 Brown, Keun-Ah Cheon, Barbara J. Coffey, Andrea Dietrich, Thomas V. Fernandez, Blanca Garcia-
339 Delgar, Donald L. Gilbert, Dorothy E. Grice, Julie Hagström, Tammy Hedderly, Gary A. Heiman, Isobel
340 Heyman, Pieter J. Hoekstra, Chaim Huyser, Young Key Kim, Young-Shin Kim, Robert A. King, Yun-Joo
341 Koh, Sodahm Kook, Samuel Kuperman, Bennett L. Leventhal, Marcos Madruga-Garrido, Pablo Mir,
342 Astrid Morer, Alexander Münchau, Kerstin J. Plessen, Veit Roessner, Eun-Young Shin, Dong-Ho Song,
343 Jungeun Song, Jay A. Tischfield, A. Jeremy Willsey, and Samuel H. Zinner.

References

1. Robertson MM, Cavanna AE, Eapen V. Gilles de la Tourette syndrome and disruptive behavior disorders: prevalence, associations, and explanation of the relationships. *The Journal of neuropsychiatry and clinical neurosciences* **27**, 33–41 (2015).
2. Scharf JM, Miller LL, Gauvin CA, Alabiso J, Mathews CA, Ben-Shlomo Y. Population prevalence of Tourette syndrome: A systematic review and meta-analysis. *Movement Disorders* **30**, 221–228 (2015).
3. Robertson MM, Eapen V, Cavanna AE. The international prevalence, epidemiology, and clinical phenomenology of Tourette syndrome: A cross-cultural perspective. *Journal of Psychosomatic Research* **67**, 475–483 (2009).
4. Davis LK *et al.* Partitioning the Heritability of Tourette Syndrome and Obsessive Compulsive Disorder Reveals Differences in Genetic Architecture. *PLoS Genetics*. doi:10.1371/journal.pgen.1003864 (2013).
5. Robertson MM *et al.* Gilles de la Tourette syndrome. *Nature Reviews Disease Primers* **3**, 16097 (2017).
6. Mataix-Cols D *et al.* Familial Risks of Tourette Syndrome and Chronic Tic Disorders: A Population-Based Cohort Study. *JAMA psychiatry* **72**, 787–793 (2015).
7. Yu* D *et al.* Interrogating the genetic determinants of Tourette syndrome and other tic disorders through genome-wide association studies. *American Journal of Psychiatry* (2018).
8. Scharf JM *et al.* Genome-wide association study of Tourette's syndrome. *Molecular psychiatry* **18**, 721–8 (2013).
9. Abdulkadir M *et al.* 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* **85**. Neurodevelopmental Alterations and Childhood Behavior, 298–304 (2019).
10. Huang A *et al.* Rare Copy Number Variants in NRXN1 and CNTN6 Increase Risk for Tourette Syndrome. *Neuron* **94**. doi:10.1016/j.neuron.2017.06.010 (2017).
11. Willsey AJ *et al.* De Novo Coding Variants Are Strongly Associated with Tourette Disorder. *Neuron* **94**, 486–499.e9 (2017).

12. Wang S *et al.* De Novo Sequence and Copy Number Variants Are Strongly Associated with Tourette Disorder and Implicate Cell Polarity in Pathogenesis. *Cell Reports* **24**, 3441–3454.e12 (2018).
13. Ballard DH, Cho J, Zhao H. Comparisons of multi-marker association methods to detect association between a candidate region and disease. *Genetic Epidemiology* **34**, 201–212 (2010).
14. De Leeuw C *et al.* Involvement of astrocyte metabolic coupling in Tourette syndrome pathogenesis. *European Journal of Human Genetics* **23**, 1519–1522 (2015).
15. Turner S *et al.* Quality Control Procedures for Genome-Wide Association Studies. *Current Protocols in Human Genetics*, 1.19.1–1.19.18 (2011).
16. Lam M *et al.* RICOPIILI: Rapid Imputation for COnsortias PIpeLIine. *Bioinformatics*. btz633. doi:10.1093/bioinformatics/btz633 (2019).
17. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* **38**, 904–909 (2006).
18. Ruano D *et al.* Functional Gene Group Analysis Reveals a Role of Synaptic Heterotrimeric G Proteins in Cognitive Ability. *American Journal of Human Genetics* **86**, 113–125 (2010).
19. Lips ES *et al.* Functional gene group analysis identifies synaptic gene groups as risk factor for schizophrenia. *Molecular Psychiatry* **17**, 996–1006 (2012).
20. Duncan LE *et al.* Pathway analyses implicate glial cells in schizophrenia. *PLoS ONE* **9**. doi:10.1371/journal.pone.0089441 (2014).
21. Goudriaan A *et al.* Specific glial functions contribute to Schizophrenia susceptibility. *Schizophrenia Bulletin* **40**, 925–935 (2014).
22. Jansen A *et al.* Gene-set analysis shows association between FMRP targets and autism spectrum disorder. *European Journal of Human Genetics* **25**, 863–868 (2017).
23. De Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Computational Biology* **11**. doi:10.1371/journal.pcbi.1004219 (2015).
24. Purcell S *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).
25. Goeman JJ, Buhlmann P. Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics* **23**, 980–987 (2007).

26. Mooney MA, Wilmot B. Gene set analysis: A step-by-step guide. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **168**, 517–527 (2015).
27. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
28. Skafidas E, Testa R, Zantomio D, Chana G, Everall IP, Pantelis C. Predicting the diagnosis of autism spectrum disorder using gene pathway analysis. *Molecular Psychiatry* **19**, 504–510 (2014).
29. Carithers LJ *et al.* A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreservation and Biobanking* **13**, 311–319 (2015).
30. Karczewski KJ *et al.* Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*. doi:10.1101/531210 (2019).
31. Pardiñas AF *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nature genetics* **50**, 381–389 (2018).
32. Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Molecular Autism* **8**, 21 (2017).
33. Okbay A *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).
34. Davies G *et al.* Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications* **9**, 2098 (2018).
35. International League Against Epilepsy Consortium on Complex Epilepsies. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. *The Lancet Neurology* **13**, 893–903 (2014).
36. Ikeda M *et al.* Genome-Wide Association Study Detected Novel Susceptibility Genes for Schizophrenia and Shared Trans-Populations/Diseases Genetic Effect. *Schizophrenia bulletin*. doi:10.1093/schbul/sby140 (2018).
37. Xie P, Kranzler HR, Yang C, Zhao H, Farrer LA, Gelernter J. Genome-wide Association Study Identifies New Susceptibility Loci for Posttraumatic Stress Disorder. *Biological Psychiatry* **74**, 656–663 (2013).
38. Puts NAJ *et al.* Reduced GABAergic inhibition and abnormal sensory symptoms in children with Tourette syndrome. *Journal of Neurophysiology* **114**, 808–817 (2015).

39. Lee JJ *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics* **50**, 1112–1121 (2018).
40. Yang L *et al.* Polygenic transmission and complex neuro developmental network for attention deficit hyperactivity disorder: genome-wide association study of both common and rare variants. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **162B**, 419–430 (2013).
41. Kelsom C, Lu W. Development and specification of GABAergic cortical interneurons. *Cell & bioscience* **3**, 19 (2013).
42. Lenington JB *et al.* Transcriptome Analysis of the Human Striatum in Tourette Syndrome. *Biological Psychiatry* **79**, 372–382 (2016).
43. Rapanelli M, Frick LR, Pittenger C. The Role of Interneurons in Autism and Tourette Syndrome. *Trends in Neurosciences* **40**, 397–407 (2017).
44. Krause DL, Müller N. The Relationship between Tourette's Syndrome and Infections. *The open neurology journal* **6**, 124–8 (2012).
45. Ercan-Sencicek AG *et al.* L-Histidine Decarboxylase and Tourette's Syndrome. *New England Journal of Medicine* **362**, 1901–1908 (2010).
46. Castellan Baldan L *et al.* Histidine Decarboxylase Deficiency Causes Tourette Syndrome: Parallel Findings in Humans and Mice. *Neuron* **81**, 77–90 (2014).
47. Alexander J *et al.* Targeted Re-Sequencing Approach of Candidate Genes Implicates Rare Potentially Functional Variants in Tourette Syndrome Etiology. *Frontiers in Neuroscience* **10**, 428 (2016).
48. Eßbach C *et al.* Abundance of Flt3 and its ligand in astrocytic tumors. *OncoTargets and therapy* **6**, 555–61 (2013).
49. Kichaev G *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *American journal of human genetics* **104**, 65–75 (2019).
50. Astle WJ *et al.* The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. *Cell* **167**, 1415–1429.e19 (2016).
51. Jain D *et al.* Genome-wide association of white blood cell counts in Hispanic/Latino Americans: the Hispanic Community Health Study/Study of Latinos. *Human molecular genetics* **26**, 1193–1204 (2017).

52. Kanai M *et al.* Genetic analysis of quantitative traits in the Japanese population links cell types to complex human diseases. *Nature genetics* **50**, 390–400 (2018).
53. Rivat C *et al.* Inhibition of neuronal FLT3 receptor tyrosine kinase alleviates peripheral neuropathic pain in mice. *Nature Communications* **9**, 1042 (2018).
54. Marchand F, Perretti M, McMahon SB. Role of the Immune system in chronic pain. *Nature Reviews Neuroscience* **6**, 521–532 (2005).
55. Sorge RE *et al.* Different immune cells mediate mechanical pain hypersensitivity in male and female mice. *Nature Neuroscience* **18**, 1081–1083 (2015).
56. Sun BB *et al.* Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).
57. Pickrell JK, Berisa T, Liu JZ, Séguirel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nature genetics* **48**, 709–17 (2016).
58. Malik R *et al.* Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nature genetics* **50**, 524–537 (2018).
59. Tsetsos F *et al.* Meta-analysis of tourette syndrome and attention deficit hyperactivity disorder provides support for a shared genetic basis. *Frontiers in Neuroscience* **10**. doi:10.3389/fnins.2016.00340 (2016).
60. Caldwell JM *et al.* Cadherin 26 is an alpha integrin-binding epithelial receptor regulated during allergic inflammation. *Mucosal Immunology* **10**, 1190–1201 (2017).
61. Tikkanen E *et al.* Biological Insights Into Muscular Strength: Genetic Findings in the UK Biobank. *Scientific reports* **8**, 6451 (2018).
62. Klimentidis YC *et al.* Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including CADM2 and APOE. *International Journal of Obesity* **42**, 1161–1176 (2018).
63. Nagel M, Watanabe K, Stringer S, Posthuma D, van der Sluis S. Item-level analyses reveal genetic heterogeneity in neuroticism. *Nature Communications* **9**, 905 (2018).
64. St Pourcain B *et al.* Common variation near ROBO2 is associated with expressive vocabulary in infancy. *Nature communications* **5**, 4831 (2014).
65. Clifton EAD *et al.* Genome-wide association study for risk taking propensity indicates shared pathways with body mass index. *Communications biology* **1**, 36 (2018).

66. Morgan MD *et al.* Genome-wide study of hair colour in UK Biobank explains most of the SNP heritability. *Nature communications* **9**, 5271 (2018).
67. Salyakina D *et al.* Variants in several genomic regions associated with asperger disorder. *Autism research : official journal of the International Society for Autism Research* **3**, 303–10 (2010).
68. Wang KS, Liu XF, Aragam N. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. *Schizophrenia research* **124**, 192–9 (2010).
69. Cornelis MC *et al.* A genome-wide investigation of food addiction. *Obesity (Silver Spring, Md.)* **24**, 1336–41 (2016).
70. Emilsson V *et al.* Co-regulatory networks of human serum proteins link genetics to disease. *Science (New York, N.Y.)* **361**, 769–773 (2018).

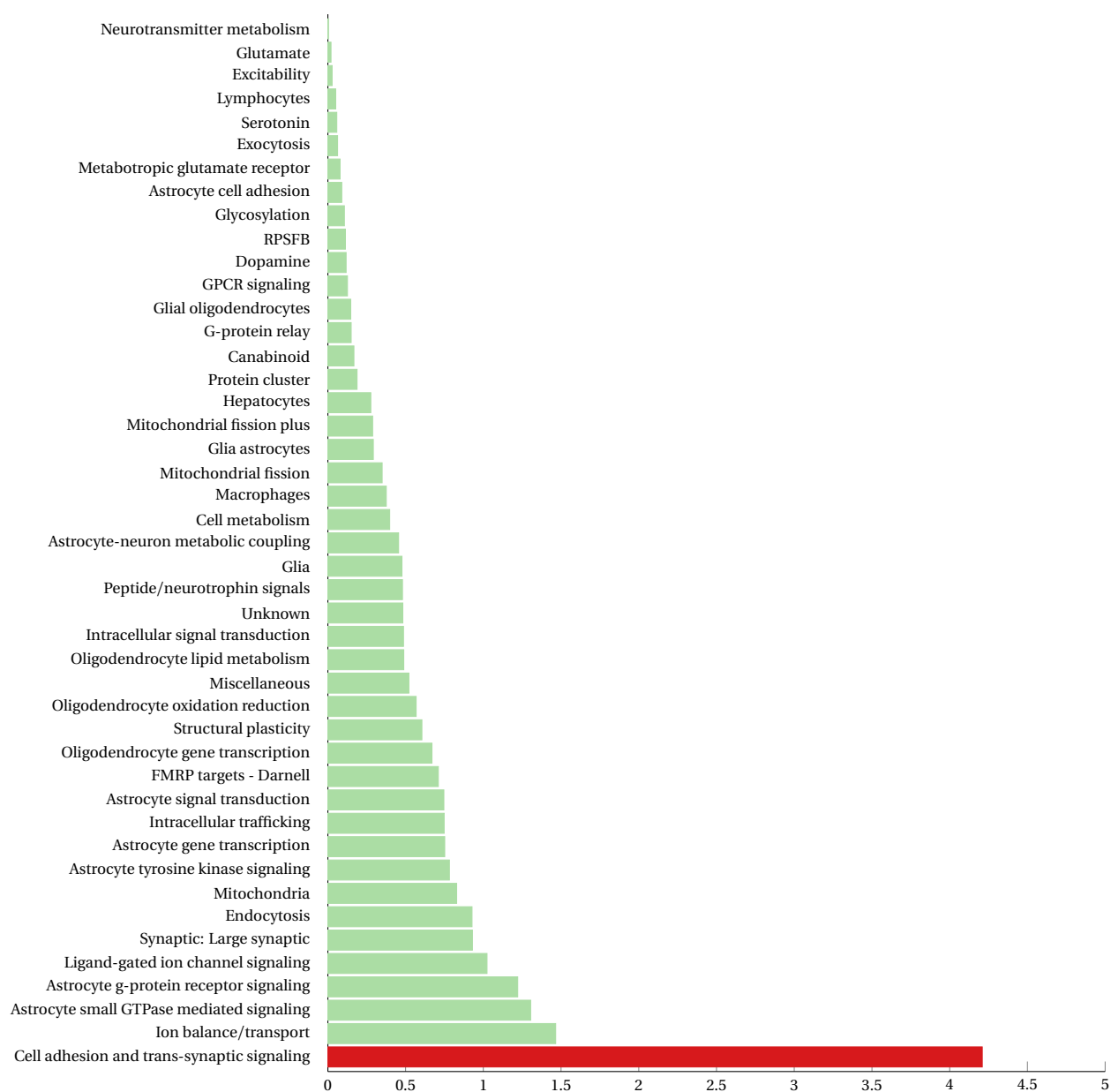


Figure 1: Results of gene set analysis as implemented by MAGMA. The gene set that crossed the significance threshold is depicted in red.

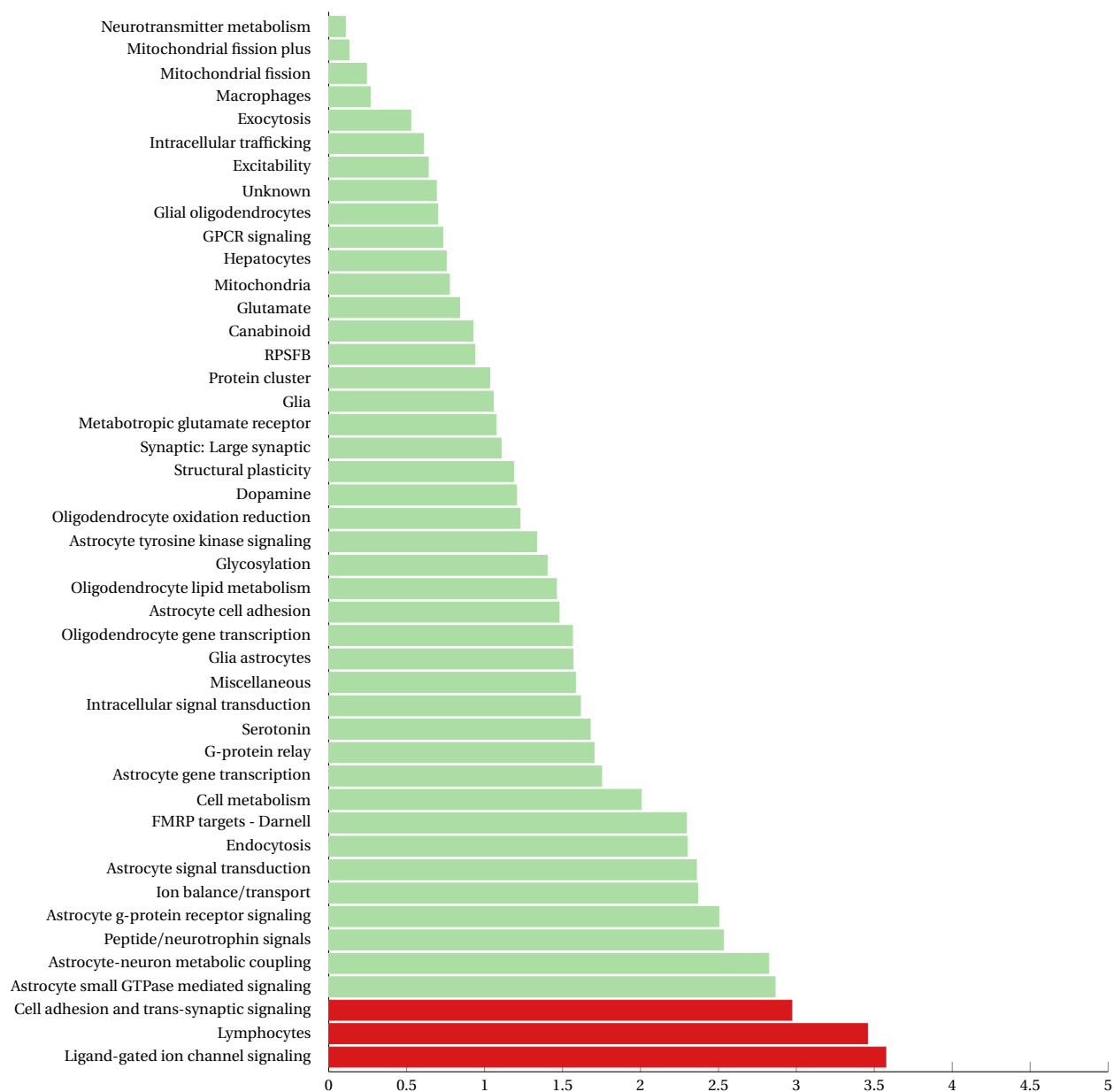


Figure 2: Results of gene set analysis as implemented by SBA. The gene sets that crossed the significance threshold are depicted in red.

Table 1: Statistically significant result of MAGMA gene set analysis. The Cell adhesion and trans-synaptic signaling gene set achieved statistical significance. Genes within this set that achieved nominal significance with gene-based test implemented by MAGMA are also listed here. Gene ID refers to Entrez ID, Param to the number of SNPs used for the snp-wise analysis.

Gene set				Genes	P-value	P_{corr}				
Gene ID	Chr	Start	End	SNPs	Param	N	Z-stat	P-value	Gene Name	
Cell Adhesion and Trans-synaptic Signaling				72	6.17e-05	0.00318				
60437	20	58528471	58593772	4	3	11263	4.2895	8.95e-06	cadherin 26 (CDH26)	
253559	3	85003133	86128579	42	18	11263	3.3124	0.00046	cell adhesion molecule 2 (CADM2)	
4978	11	132279875	133407403	210	106	11263	3.1564	0.00079	opioid binding protein/cell adhesion molecule like (OPCML)	
1007	5	26875709	27043689	14	7	11263	2.9627	0.0015	cadherin 9 (CDH9)	
4685	21	22365633	22918892	61	29	11263	2.7975	0.0025	neural cell adhesion molecule 2 (NCAM2)	
961	3	107756941	107814935	6	4	11263	2.6465	0.0040	CD47 molecule (CD47)	
1003	16	66395525	66443689	11	6	11263	2.0242	0.021	cadherin 5 (CDH5)	
199731	19	44121519	44148991	4	3	11263	1.984	0.023	CADM4 (cell adhesion molecule 4)	
708	17	5331099	5347471	1	1	11263	1.9269	0.026	C1QBP (complement C1q binding protein)	
2017	11	70239612	70287690	2	2	11263	1.8709	0.030	CTTN (cortactin)	
4045	3	115516210	116169385	56	29	11263	1.8095	0.035	limbic system-associated membrane protein (LSAMP)	
8502	2	159308476	159542941	19	9	11263	1.7503	0.040	plakophilin 4 (PKP4)	
5097	5	141227655	141263361	3	3	11263	1.6903	0.045	PCDH1 (protocadherin 1)	
26047	7	145808453	148123090	237	110	11263	1.6621	0.048	contactin associated protein-like 2 (CNTNAP2)	
4155	18	74685789	74849774	49	30	11263	1.6502	0.049	MBP (maltose-binding protein)	

Table 2: Statistically significant results of the SBA analysis. 3 pathways achieved significance. Association statistics for the top 5 SNPs driving the signal in each set are also shown. NSIG is the number of SNPs crossing the nominal significance threshold. EMP1 is the empirical p-value attained by the tested gene set. P is the p-value of the original single marker association, OR is the respective odds ratio. A1 is the minor allele and A2 the major allele. F_A and F_U are the frequencies of the minor allele in case and control samples respectively.

Gene set				SNPs	NSIG	EMP1				
Ligand-gated Ion Channel Signaling				683	66	0.000267				
Chr	SNP	BP	A1	F_A	F_U	A2	P	OR	Genes implicated	
4	rs1391174	46072596	T	0.4892	0.4586	C	1.764e-05	1.131	GABRG1(0)	
5	rs9790873	45291514	C	0.1535	0.1335	T	5.621e-05	1.177	HCN1(0)	
9	rs2259639	101317401	T	0.2751	0.2982	C	0.0003612	0.8928	GABBR2(0)	
9	rs1930415	101238974	T	0.2218	0.2424	C	0.0007006	0.8908	GABBR2(0)	
11	rs949054	120795888	C	0.2241	0.2053	T	0.001281	1.118	GRIK4(0)	

Gene set				SNPs	NSIG	EMP1				
Lymphocytes				799	50	0.00035				
Chr	SNP	BP	A1	F_A	F_U	A2	P	OR	Genes implicated	
19	rs16986092	55433696	T	0.1158	0.09473	C	1.093e-06	1.251	NCR1(+9.257kb) NLRP7(-1.18kb)	
13	rs1933437	28624294	G	0.4183	0.3871	A	8.482e-06	1.138	FLT3(0)	
3	rs2243123	159709651	C	0.2515	0.2759	T	0.0001167	0.8817	IL12A(0) IL12A-AS1(0)	
7	rs3801983	18683672	C	0.1928	0.2133	T	0.0003981	0.8808	HDAC9(0)	
5	rs2230525	66478626	C	0.08431	0.07127	T	0.0005641	1.2	CD180(0)	

Gene set				SNPs	NSIG	EMP1				
Cell Adhesion and Trans-synaptic Signaling				3290	292	0.00107				
Chr	SNP	BP	A1	F_A	F_U	A2	P	OR	Genes implicated	
20	rs1002762	58580885	G	0.2305	0.2028	A	2.031e-06	1.178	CDH26(0)	
21	rs2826825	22762779	G	0.376	0.3487	A	6.698e-05	1.126	NCAM2(0)	
11	rs7925725	131449365	C	0.3709	0.3979	A	0.0001099	0.8921	NTM(0)	
11	rs12224080	131816849	G	0.09841	0.08353	A	0.0002519	1.198	NTM(0)	
3	rs6773575	77060574	C	0.0964	0.1126	A	0.000256	0.8407	ROBO2(0)	