Epigenome-Wide Association Study of Tic Disorders

Nuno R. Zilhão,1,2,a Shanmukha S. Padmanabhuni,3,+,a Luca Pagliaroli,4 Csaba Barta,4 BIOS Consortium,∗ Dirk J. A. Smit,2 Danielle Cath,2,5 Michel G. Nivard,1 Bart M. L. Baselmans,1 Jenny van Dongen,1 Peristera Paschou,3, b and Dorret I. Boomsma1,b

1Department of Biological Psychology, VU University Amsterdam, Neuroscience Campus Amsterdam, Amsterdam, the Netherlands
2Department of Clinical and Health Psychology, Utrecht University, Utrecht, the Netherlands
3Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupolis, Greece
4Institute of Medical Chemistry, Molecular Biology and Pathobiochemistry, Semmelweis University, Budapest, Hungary
5Altrecht Academic Anxiety Disorders Center, Utrecht, the Netherlands

Tic disorders are moderately heritable common psychiatric disorders that can be highly troubling, both in childhood and in adulthood. In this study, we report results obtained in the first epigenome-wide association study (EWAS) of tic disorders. The subjects are participants in surveys at the Netherlands Twin Register (NTR) and the NTR biobank project. Tic disorders were measured with a self-report version of the Yale Global Tic Severity Scale Abbreviated version (YGTSS-ABBR), included in the 8th wave NTR data collection (2008). DNA methylation data consisted of 411,169 autosomal methylation sites assessed by the Illumina Infinium HumanMethylation450 BeadChip Kit (HM450k array). Phenotype and DNA methylation data were available in 1,678 subjects (mean age \(= 41.5\)). No probes reached genome-wide significance \((p < 1.2 \times 10^{-7})\). The strongest associated probe was cg15583738, located in an intergenic region on chromosome 8 \((p = 1.98 \times 10^{-6})\). Several of the top ranking probes \((p < 1 \times 10^{-4})\) were in or nearby genes previously associated with neurological disorders (e.g., GABBR1, BLM, and ADAM10), warranting their further investigation in relation to tic disorders. The top significantly enriched gene ontology (GO) terms among higher ranking methylation sites included anatomical structure morphogenesis (GO:0009653, \(p = 4.6 \times 10^{-13}\)) developmental process (GO:0032502, \(p = 2.96 \times 10^{-12}\)), and cellular developmental process (GO:0048869, \(p = 1.96 \times 10^{-12}\)). Overall, these results provide a first insight into the epigenetic mechanisms of tic disorders. This first study assesses the role of DNA methylation in tic disorders, and it lays the foundations for future work aiming to unravel the biological mechanisms underlying the architecture of this disorder.

Keywords: Tourette syndrome, tics, epigenetics, DNA methylation, EWAS

Tic disorders form a broad spectrum encompassing four different clinical entities: Tourette syndrome (TS), chronic (motor or vocal) tic disorder, transient tic disorder, and tic disorder not otherwise specified. Tics are characterized by sudden, rapid, motor movements, or vocalizations, performed in a ritualized, recurrent, and stereotypical fashion (DSM-IV-TR; American Psychiatric Association, 2000).

Multiple lines of evidence suggest that both genetic and interacting environmental factors are causes underlying the etiology of these phenotypes (Paschou, 2013). Twin and family studies have shown that the prevalence of TS or chronic tic disorders among first-degree relatives of affected individuals varies between 15% and 53% (Towbin, 2010). Furthermore, heritability estimates on tic disorders or TS range between 0.28 and 0.56 (Anckarsäter et al., 2011; Bolton et al., 2007; de Haan et al., 2015; Lichtenstein et al., 2010; Mathews & Grados, 2011; Ooki, 2005). However, most genetic and clinical studies so far have been hampered by small sample sizes, ambiguity in phenotype definitions, and
difficulties in disentangling genetic and familial environmental effects. Thus, molecular genetic studies have thus far not yet yielded robust findings (Pauls et al., 2014).

The study of epigenetic mechanisms on a genome-wide scale in humans represents the bridge between disease susceptibility and gene expression variation. It is known that epigenetic mechanisms regulate gene expression, and that these epigenetic mechanisms change during life, up- and down-regulating different genes in response to external environmental conditions. DNA methylation at CpG dinucleotides is the most studied epigenetic mechanism in humans. It regulates gene expression by targeting, for example, promoters and enhancers, and its pattern may change as a consequence of external and internal stimuli (Hackett et al., 2013). DNA methylation is also involved in other biological processes such as genomic imprinting, where CpG sites are methylated based on their parental origin, which are different between the paternal and maternal branch (Liu et al., 2010), and chromosome X inactivation, where one copy of the female X chromosome is inactivated (Heard et al., 1997). Genomic imprinting has also been suggested to be involved in TS in a study evaluating parental gender-influenced differences in childhood TS phenotype. Greater motor tic complexity was associated with maternal transmission, whereas higher frequency in vocal tics was associated with paternal transmission (Lichter et al., 1995).

EWAS have thus far been used to reveal altered methylation patterns in several complex disorders such as obesity, diabetes, schizophrenia, and autism (Dempster et al., 2011; Moore et al., 2014; Rakyan et al., 2011; Wockner et al., 2014). This approach is becoming increasingly accessible and it is likely that DNA methylation is also involved in other neurological disorders. Currently, no EWAS of tic disorders have been performed. These studies may clarify underlying mechanisms that differentially regulate genes in individuals manifesting tics.

Here, we report on the first EWAS of tic disorders performed in a population-based sample from the NTR.

**Material and Methods**

**Subjects’ Demographics**

Subjects in this study are participants in the NTR biobank Project (Willemsen et al., 2010, 2013). Since 1991, the NTR has been collecting information on a broad range of phenotypes in twins and family members (Willemsen et al., 2013). In total, 3,264 peripheral blood samples from 3,221 NTR participants have been assessed for genome-wide methylation data. After quality control (QC) on the methylation data, the total final selection comprised 3,089 samples, for a total of 3,057 individuals (32 subjects had methylation data for two time points). For a complete description of the entire methylation dataset from the NTR, please see Van Dongen et al. (2015).

In the present study, we analyzed tic data from the 2008 wave of collection in a subset of individuals in whom genome-wide methylation data were available. A total of 1,678 individuals (twins, siblings, and parents) from 1,057 families were included in the analysis. Table 1 gives an overview of the subjects entered in the analysis and demographics. Zygosity was assessed by DNA polymorphisms as described by van Beijsterveldt et al. (2013). The study was approved by the Central Ethics Committee on Research involving human subjects of the VU University Amsterdam.

**Phenotype**

Tics were measured using an abbreviated 12-item self-report version of the YGTSS, the latter being a well-validated interview with a high internal consistency (Cronbach’s alpha > 0.90; Leckman et al., 1989). The YGTSS-abbreviated (YGTSS-ABBR) contains the 12 most frequently occurring tics, assessing their occurrence: never (0), < 1 year ago (1), between 15 years ago (2) or as a child (3). Three additional questions are asked on age at onset, duration, and severity, to enable establishing a probable diagnosis according to DSM-IV-TR criteria (American Psychiatric Association, 2000). Table 2 shows the YGTSS-ABBR questionnaire used for measuring tics. A diagnosis of probable chronic tic disorder was established if the person had: (1) one or more chronic motor or one or more vocal tics, that (2) occurred before age 21, and (3) had been present for > 1 year. Probable TS diagnosis was established when two or more motor and one or more vocal tics were reported that occurred before age 21 and had lasted for > 1 year, and probable transient tic disorder was established when motor and/or vocal tics had occurred before age 21 for less than one year. From these categories, we derived a dichotomous variable on absence or presence of a probable tic disorder diagnosis — chronic tic disorder, transient tic disorder or TS, as referenced by the Tourette Syndrome Classification Study Group (1993). An extensive genetic analysis on the heritability of tic disorders has been performed (Zilhão et al., 2015). Since smoking is known to have an effect on DNA methylation (Lee & Pausova, 2013), we controlled for smoking status in

---

**Table 1**

**Number of Participants Included in the Analysis**

<table>
<thead>
<tr>
<th></th>
<th>MZ twins</th>
<th>DZ/DOS twins</th>
<th>Relatives</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male (case/control)</td>
<td>Female (case/control)</td>
<td>Male (case/control)</td>
</tr>
<tr>
<td>N</td>
<td>241 (30/211)</td>
<td>680 (74/606)</td>
<td>213 (31/182)</td>
</tr>
<tr>
<td>Age (SE)</td>
<td>42.48 (14.00)</td>
<td>40.57 (12.08)</td>
<td>37.61 (10.11)</td>
</tr>
</tbody>
</table>

Note: MZ = monozygotic twins, DZ = dizygotic twins, DOS = dizygotic opposite-sex twins.
the epigenome-wide association analysis. Smoking status was assessed at the moment of blood draw by interview.

**Infinium HumanMethylation450 BeadChip Data**

DNA methylation was assessed with the Infinium HumanMethylation450 BeadChipKit (Illumina, Inc.; Bibikova et al., 2011). Genomic DNA (500ng) from whole blood was bisulfite treated using the ZymoResearch EZ DNA Methylation kit (Zymo Research Corp, Irvine, CA, USA) following the standard protocol for Illumina 450K micro-arrays, by the department of Molecular Epidemiology from the Leiden University Medical Center (LUMC), The Netherlands. Subsequent steps (i.e., sample hybridization, staining, scanning) were performed by the Erasmus Medical Center micro-array facility, Rotterdam, the Netherlands. QC and processing of the blood methylation dataset has been described in detail previously (Van Dongen et al., 2015). In short, a number of sample-level and probe-level quality checks were performed. Sample-level QC was performed using MethylAid (van Iterson et al., 2014). Probes were set to missing in a sample if they had an intensity value of exactly zero, or a detection p-value > .01, or a bead count < 3. After these steps, probes that failed based on the above criteria in >5% of the samples were excluded from all samples (only probes with a success rate ≥95% were retained). Probes were also excluded from all samples if they mapped to multiple locations in the genome (Chen et al., 2013), and/or had a SNP within the CpG site (at the C or G position) irrespective of minor allele frequency in the Dutch population (Genome of the Netherlands Consortium, 2014). Only autosomal methylation sites were analyzed in the EWAS. The methylation data were normalized with Functional Normalization (Fortin et al., 2014) and normalized intensity values were converted into beta (β)-values. The β-value represents the methylation level at a site, ranging from 0 to 1 and is calculated as

$$\beta = \frac{M}{M + U + 100},$$

where $M$ = methylated signal, $U$ = unmethylated signal, and 100 represents a correction term to control the β-value of probes with very low overall signal intensity. After QC, from an initial total of 485,577 methylation sites, the final total selection of methylation sites was 411,169.

**Statistical Analysis**

EWAS was performed using linear regression under an additive model correcting for principal components (PCs) and covariates. PCs were calculated from the methylation data after QC and normalization. The PCA plot calculated can be seen in Supplementary Figure S2. Supplementary Figure S3 provides the correlation plot between the first 20 PCs and the covariates (monocyte count, eosinophils count, neutrophils count, array row number, smoking status, age, and sex) in our dataset. Generalized estimation equation (GEE) models were used to test whether tics were associated with DNA methylation. In the final model, DNA methylation level was used as outcome with the following predictors: tics, top five PCs, smoking status, eosinophil percentage, and monocyte percentage. Additional models were tested to evaluate the inflation factor with different covariates (Supplementary material). Sex, neutrophils count,
and age were not included due to their correlation with the top five PCs (Supplementary Figure S3). Basophil percentage was not included because it showed little variation between subjects, with a large number of subjects having 0% of basophils. Smoking status was coded as '0 = non-smoker', '1 = former-smoker', '2 = current-smoker'. GEE uses cluster robust standard errors with family ID as cluster variable, and thus standard errors are robust for the presence of related individuals in the sample. A permuted p-value was calculated by random sampling of the phenotype (n = 10,000), defining the proportion of permutations meeting or exceeding our p-value estimate based on the actual data. CpGs with p-values < \(1.2 \times 10^{-7}\) (Bonferroni correction of 0.05/411169 — autosomal sites) were considered statistically significant.

Enrichment of Gene Ontology Terms
Enrichment of GO terms among methylation sites having a stronger association with tics was done by ranking all methylation sites based on the EWAS p-value and the resulting ranked gene list was supplied to the online software tool GOrilla (Eden et al., 2009). The GO tool GOrilla takes into account the rank of the gene whereby p-value cut-off is not required in creating the gene list. A false discovery rate (FDR) q-value <0.05 was considered for a GO term to be statistically significant.

Results
After QC, the final dataset consisted of 411,169 autosomal CpG sites, for 1,678 individuals (188 cases and 1,490 controls). The variation in the data captured by the first two PCs is shown in Figure 1. It can be seen that the main contributors to variation in DNA methylation are not associated with the phenotype. The first component represented sex and the second component correlated strongly with neutrophil count. Supplementary Figure S1 shows the
distribution of genome-wide methylation level for all individuals included in the analysis.

The quantile-quantile (QQ) plot based on the EWAS model is shown in Figure 2. The inflation factor (λ) is 1.03, indicating that the results are not stratified.

None of the CpG sites passed the Bonferroni p-value threshold ($p = 1.2 \times 10^{-7}$) for the association with Tics. Table 3 shows all CpG sites with a $p$-value $< 1 \times 10^{-4}$ ($N = 57$). Permutation tests for all of these 57 CpG sites resulted in a permutation $p$-value $< .05$, indicating that these CpG sites did not occur by chance. As an example, the distribution of permutation $p$-values for the top nine CpGs is shown in Supplementary Figure S7. The Manhattan plot can be seen in Figure 3 with the top 57 CpG sites highlighted in green. Figure 4 shows the methylation level in cases and controls for our top 15 CpGs.

Gene Ontology Enrichment Analysis
GO enrichment analysis identified a large number of GO terms that were significantly enriched (Supplementary Tables T1–T3). Top enriched GO Processes include developmental process (GO:0032502, $p = 3.98e^{-16}$, FDR $q$-value = $2.69e^{-12}$), cellular developmental process (GO:0048869, $p = 4.35e^{-16}$, FDR $q$-value = $1.96e^{-12}$), single-organism developmental process (GO:0044767, $p = 6.57e^{-23}$, FDR $q$-value = $2.22e^{-12}$), and other GO terms related to development. Brain processes, including regulation of neuron projection guidance (GO:0097485, $p = 8.33e^{-11}$, FDR $q$-value = $5.11e^{-08}$), axon guidance (GO:0007411, $p = 8.33e^{-11}$, FDR $q$-value = $4.89e^{-08}$), and neuron differentiation (GO:0030182, $p = 1.15e^{-10}$, FDR $q$-value = $6.24e^{-08}$) were also significantly enriched. The top most enriched GO component was cell junction (GO:0030054, $p = 6.73e^{-09}$, FDR $q$-value = $1.09e^{-05}$), followed by the neuron-specific GO components neuron part (GO:0097458, $p = 4.39 \times 10^{-07}$, FDR $q$-value = $1.78e^{-04}$), synapse part (GO:0044456, $p = 5.13 \times 10^{-06}$, FDR $q$-value = $1.39e^{-03}$), postsynaptic density (GO:0014069, $p = 1.67e^{-05}$, FDR $q$-value = $3.87e^{-03}$) and dendrite (GO:0030425, $p = 1.04e^{-04}$, FDR $q$-value = $1.05e^{-02}$). GO components related to histone modification, including MOZ/MORF histone
FIGURE 3
Manhattan plot showing the p-values of genome-wide CpG sites. The red line is the genome-wide threshold and the blue line is the suggestive threshold ($p < 1 \times 10^{-4}$). Top CpG sites are highlighted in green.

FIGURE 4
Average methylation levels of the top CpGs in cases versus controls.
acetyltransferase complex (GO:0070776, \( p = 4.21 \times 10^{-5} \), FDR \( q\)-value = 6.20e-03) and H3 histone acetyltransferase complex (GO:0070775, \( p = 4.21 \times 10^{-5} \), FDR \( q\)-value = 5.68e-03), were significantly enriched. The top enriched GO function was protein binding (GO:0005515, \( p = 1.10e^{-17} \), FDR \( q\)-value = 4.55e-14) followed by many functions involving DNA binding.

### Discussion

In this work, we present the first genome-wide epigenetic analysis on tics/tic disorders. Although no methylation site achieved significance at the genome-wide threshold, gene-ontology analysis of the top hits revealed enrichment in brain-specific and developmental processes. Thus, our findings provide interesting targets for further analysis.
Two of our top CpGs (cg08093277 and cg21879791, ranked 11 and 41) are located near the GABBR1 gene (39670 bp and 39201 bp, at 5', respectively), which represents a very interesting target for further analysis. GABBR1 encodes a subunit of the GABA receptor, the major inhibitory neurotransmitter in the central nervous system (CNS). GABA acts at brain inhibitory synapses where it binds transmembrane receptors of both presynaptic and postsynaptic neurons. Notably, GABBR1 has been previously associated with autism, schizophrenia, tremor, and obsessive-compulsive disorder (Fatemi et al., 2009; Hegyi, 2013; Luo et al., 2012; Richter et al., 2012).

The involvement of GABA in TS is well documented in the extant neurobiological literature. Tics have been associated with reduced basal ganglia volume and reduced cortical thickness in motor and sensorimotor areas controlling the facial, orolingular, and laryngeal muscles (Sowell et al., 2008). Studies of brain activity using PET or fMRI in TS subjects compared to controls have implicated dysfunctional striatal and thalamic, as well as cortical regions. The medium-sized spiny neurons in the basal ganglia have GABAergic inhibitory projections to the substantia nigra and Globus Pallidus (Ribak et al., 1979). It is hypothesized that the tonic activity of the striatum acts to inhibit unwanted motor patterns (Albin & Mink, 2006). Treatment strategies have aimed at increasing GABA levels (Awaad, 1999; Mink, 2001) to counteract decreased inhibitory output resulting in excessive activity in frontal cortical areas. Although not all studies showed significant results, one randomized double-blind study administering a GABA β-receptor agonist in TS resulted in reduced tic severity (Singer et al., 2001).

Importantly, 8 of the top 57 associated CpG sites (p < 1 x 10^-4) mapped to genes that have been previously associated with psychiatric or neurological disorders, some of them sharing neurodevelopmental aspects with tics (OCD, autism, bipolar disorder, schizophrenia, some forms of mental retardation), some disorders involving corticostriatal pathways similar to tics (as in the case of Parkinson's disease), and some in which the link with tics seems less clear (Alzheimer's disease) (Supplementary Table T4 summarizes the genes from our top list that have been previously associated with neurological disorders). CpG site cg06026425, located near CLINT1, has been linked to schizophrenia (Leon et al., 2011; Wang et al., 2010). Moreover, cg01321816, located near BLM, and cg0785856, located near ADAM10, involve genes that have been associated with Alzheimer's disease (Schröter et al., 2013; Vassar, 2013). Another example is cg26548492, which is located near LOC153328/SLC25A48, a gene that has been proposed as candidate for Parkinson's disease (Liu et al., 2011). Lastly, cg2051967, which is located near PLEKHG3, involves a gene that has been linked to mental retardation (Lehalle et al., 2014; Lybaek et al., 2008). Furthermore, some of the top hits lead to genes with a brain-specific function. This is the case for cg25086136, located near SNTG1, which is specifically expressed in the CNS (Hafner et al., 2010), and cg19830950, located near SEMA4G, which might play a role in cerebellar development, the cerebellum being a core structure involved in the precision of motor control (Maier et al., 2011).
Future studies might consider taking a trans-diagnostic neurodevelopmental approach, combining tic disorders and other neurodevelopmental disorders (including OCD, autism, ADHD, and childhood movement disorders), to disentangle common versus disorder-specific underlying methylation patterns. Finally, environmental factors assessed in a longitudinal study design should be incorporated in epigenetic studies, to investigate which environmental stressors/protectors at what age/stage of development influence DNA methylation.

This study provides a first step to unravel the role of epigenetic mechanisms in tic disorders. It should be noted that we analyzed an ‘inclusive’ tic phenotype definition that may obscure different underlying etiologies. Future studies of larger size or including clinical samples at the more extreme end of the tic phenotypic spectrum are required to improve statistical power. Future studies should also aim to examine different phenotypic tic dimensions (de Haan et al., 2015). Such studies hold the promise to shed light on the complex interaction between environmental and genetic factors leading to development and persistence of neuropsychiatric disorders.

Acknowledgments

We thank the twins and their family members who participated in the studies of the NTR. This study was supported by the BBRMI-NL-financed BIOS Consortium (NWO 184.021.007). This project was also financed by FP7-People-2012-ITN, project: TS-EUROTRAIN, grant number 316978; BBR Foundation (NARSAD) 21668; ZonMW (Ad-

Supplementary Material

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/thg.2015.72

References


