Genes, Brain and Behavior
Official publication of the International Behavioural and Neural Genetics Society

Genes, Brain and Behavior (2012) **11:** 444–451 doi: 10.1111/j.1601-183X.2012.00778.x

Evaluation of the LIM homeobox genes *LHX6* and *LHX8* as candidates for Tourette syndrome

P. Paschou*,†, E. Stylianopoulou†, I. Karagiannidis†, R. Rizzo‡, Z. Tarnok§, T. Wolanczyk¶, J. Hebebrand, M. M. Nöthen††, G. Lehmkuhl‡‡, L. Farkas§, P. Nagy§, U. Szymanska¶, D. Lykidis†, C. Androutsos§§, V. Tsironi§§, A. Koumoula§§, C. Barta¶¶, S. Klidonas†, P. Ypsilantis***, C. Simopoulos***, TSGeneSEE Consortium†††, G. Skavdis† and M. Grigoriou*,†**

*† Department of Molecular Biology and Genetics, Democritus University of Thrace, Alexandroupoli, Greece ‡ Materno-Infantile and Radiological Science Department, University of Catania, Catania, Italy § Vadaskert Clinic for Child and Adolescent Psychiatry, Budapest, Hungary ¶ Department of Child Psychiatry, Medical University of Warsaw, Warsaw, Poland ** Department of Child and Adolescent Psychiatry, University of Duisburg-Essen, Essen, Germany †† Institute for Human Genetics, University of Bonn, Bonn, Germany ‡‡ Department of Child and Adolescent Psychiatry of the University of Cologne, Cologne, Germany §§ Child and Adolescent Psychiatry Clinic, Sismanoglio General Hospital of Attica, Athens, Greece ¶¶ Department of Medical Chemistry, Molecular Biology and Pathobiochemistry, Semmelweis University, Budapest, Hungary *** School of Medicine, Democritus University of Thrace, Alexandroupolis, Greece ††† See Appendix S2 for list of the Tourette Syndrome Genetics-Southern and Eastern Europe Initiative Consortium members*

**Corresponding author: P. Paschou, Department of Molecular Biology and Genetics, Democritus University of Thrace, Panepistimioupoli, Dragana, Greece. E-mail: ppaschou@mbg.duth.gr*
M. Grigoriou, Department of Molecular Biology and Genetics, Democritus University of Thrace, Panepistimioupoli, Dragana, Greece. E-mail: mgrigor@mbg.duth.gr

The etiology and pathophysiology of Tourette Syndrome (TS) remain poorly understood. Multiple lines of evidence suggest that a complex genetic background and the cortico-striato-thalamo-cortical circuit are involved. The role of *Lhx6* and *Lhx8* in the development of the striatal interneurons, prompted us to investigate them as novel candidate genes for TS. We performed a comparative study of the expression of *Lhx6* and *Lhx8* and investigated genetic association with TS using two samples of trios (TSGeneSEE and German sample - 222 families). We show that *Lhx6* and *Lhx8* expression in the forebrain is evolutionarily conserved, underlining their possible importance in TS-related pathophysiological pathways. Our tagging-single nucleotide polymorphism (tSNP)-based association analysis was negative for association with *LHX8*. However, we found positive association with *LHX6* in the TSGeneSEE sample (corrected *P*-value = 0.006 for three-site haplotype around SNP rs3808901) but no association in the sample of German families. Interestingly, the SNP allele that was identified to be significantly associated in the TSGeneSEE dataset, showed an opposite trend of transmission in the German dataset. Our analysis of the correlation of the *LHX6* region with individual ancestry within Europe, revealed the fact that this particular SNP demonstrates a high degree of population differentiation and is correlated with the North to South axis of European genetic variation. Our results indicate that further study of the *LHX6* gene in relation to the TS phenotype is warranted and suggest the intriguing hypothesis that different genetic factors may contribute to the etiology of TS in different populations, even within Europe.

Keywords: Genetic association, *LHX6*, *LHX8*, neuroanatomical expression, Tourette syndrome

Received 19 October 2011, revised 29 January 2012, 20 February 2012, and 28 February 2012, accepted for publication 3 March 2012

Tourette syndrome (TS) is a childhood-onset neuropsychiatric disorder characterized by multiple motor and vocal tics and high comorbidity rates with obsessive compulsive disorder (OCD) and attention deficit and hyperactivity disorder (ADHD) (Swain *et al*. 2007). The fact that symptoms appear to peak during adolescence and often remit as individuals progress into adulthood, suggests the intriguing hypothesis that neurodevelopmental pathways are involved (Bloch and Leckman 2009). It is currently thought that environmental and genetic factors interact in order to lead to the onset of symptoms. However, the exact role and the contribution of each of these factors have not been yet elucidated. Although multiple genes and chromosomal regions have been implicated in TS etiology (see Grados 2010; O'Rourke *et al*. 2009; State 2011) the difficulty in replicating the original positive results, underlines the complexity and heterogeneity of the etiological background of the disorder.

Several lines of evidence indicate that the basal ganglia are affected in TS individuals. Imaging studies have shown that in children and adults with TS, the volume of caudate is smaller (Peterson *et al*. 2003) suggesting that there is a decrease in the number of the cells in the striatum. In accordance with these observations, recent studies of postmortem tissues from TS-affected individuals have shown a significant selective decrease in the number of the striatal cholinergic interneurons (Chat+) as well as of the

striatal interneurons expressing parvalbumin (PV+) (Kalanithi *et al*. 2005; Kataoka *et al*. 2010) raising the possibility that genes involved in the development and function of these neuronal subpopulations are implicated in the etiology of TS.

Lhx6 and *Lhx7* genes (also called *Lhx8* or *Lhx8/L3*–we will use *Lhx8*, as the human homolog is known by this name only) encode for two closely related LIM homeodomain transcription factors (Grigoriou *et al*. 1998; Matsumoto *et al*. 1996). During mouse brain development these genes are expressed in the medial ganglionic eminences (MGE), the structures from which striatal interneurons originate (Bachy and Retaux 2006; Flames *et al*. 2007; Grigoriou *et al*. 1998; Lavdas *et al*. 1999). *Lhx6* and *Lhx8* play a critical role in the specification of forebrain interneurons: *Lhx6* is required for the specification of the PV+ and somatostatin (SST+) expressing interneurons of the cerebral cortex and the striatum (Liodis *et al*. 2007; Zhao *et al*. 2008), while *Lhx8* for the specification of two other subpopulations, namely the cholinergic interneurons of the striatum and the cholinergic projection neurons of the basal forebrain (Fragkouli *et al*. 2005, 2009; Zhao *et al*. 2003).

The results from the analysis of postmortem basal ganglia from TS-affected individuals, along with the role of *Lhx6* and *Lhx8* in the specification of specific subtypes of striatal interneurons implicated in TS, prompted us to investigate their involvement in the pathophysiological pathways of TS. We first performed a comparative study of the expression of *Lhx6* and *Lhx8* in order to establish that their expression during mammalian forebrain development is conserved. We then proceeded to investigate the genetic variation across *Lhx6* and *Lhx8* genes in relation to the TS phenotype in two independent family samples of European descent.

Materials and methods

Comparative analysis of Lhx6 and Lhx8 expression in the embryonic and adult brain of mouse and rat

Mice (C57BL/6J) and rats (Sprague Dawley) were obtained from the breeding facilities of the Institute of Molecular Biology and Biotechnology (Crete, Greece) and the Democritus National Center for Scientific Research (Athens, Greece) respectively. Animals used for the study were obtained from in-house breeding colonies, were housed in polycarbonate cages, at 20–22°C, on a 12:12 h light:dark cycle, and were given commercial pelleted diet (4RF25, Mucedola, Milan, Italy) and water ad libitum. Animal breeding and experiments were carried out in accordance with the European Communities Council Directive of 24 November 1986 (86/609/EEC); see Appendix S1, Supporting information. Time mated pregnant female mice (C57BL/6J) or rats (Sprague Dawley) were sacrificed at various stages (vaginal plug was considered as day 0.5) and embryos were dissected free of maternal tissues in 1× phosphate buffered saline (PBS). Tissue preparation, fixation and sectioning was performed as previously described (Grigoriou *et al*. 1998). Antisense RNA probes were synthesized by *in vitro* transcription with T3 RNA polymerase (Takara, Japan), according to manufacturer's instructions, using Digoxigenin-11-UTP (Roche, Switzerland). For *Lhx6*, a 410 bp fragment (nt 699-1110 of the mouse cDNA) was used as template. For the *Lhx8* an 163 bp fragment (nt 1022-1185 of the mouse complementary DNA) was used as template. For *in situ hybridization* on rat sections the mouse probes were used. *In situ* hybridization experiments on cryostat sections were performed as previously described. (Grigoriou *et al*. 1998).

Genetic association study samples

Two independently collected samples of TS trios (one individual with TS and two parents) were analyzed (222 families). Details of age of onset per sample, sex distribution as well as comorbid OCD and ADHD are shown in Table S1, Supporting information. The TSGeneSEE sample (127 trios – 83.7% male, 16.3% female) included European-descent families of Polish (27 trios), Italian (43 trios), and Hungarian origin (57 trios) (TSGeneSEE: the Tourette Syndrome Genetics – Southern and Eastern Europe Initiative). Assessment was performed by on-site clinicians using the tools provided by the Tourette Syndrome Association International Consortium for Genetics (TSAICG 2006). TS was ascertained according to DSM-IV-TR criteria for Italy and Hungary and DSM-IV for Poland. The second sample consisted of families of German descent (95 trios – 80% male, 20% female) that were collected independently of the TSGeneSEE study, using DSM-III-R criteria (Hebebrand *et al*. 1997; Schoenian *et al*. 2003). Differences between DSM-III-R, DSM-IV, and DSM-IV-TR are minimal and therefore, although studied samples were collected at different times, we expect very little (if any) heterogeneity among patients; the upper age of onset is 18 in DSM-IV (and DSM-IV-TR) and 21 in DSM-III-R, and the 'marked distress' criterion, possibly pointing to more severe cases, only appears in DSM-IV (Cath *et al*. 2011) (applied only for the 27 Polish families). For both samples, collection was approved by local Ethics Boards and informed consent was taken from all participating individuals or their parents.

Statistical analysis

Our methods are reported in detail in Supporting information. Tagging single nucleotide polyphormisms (tSNPs) for both *LHX6* and *LHX8* were selected using the HapMap CEPH European population as reference. A total of 10 tSNPs were selected for the *LHX6* region (Table S2, Supporting information) capturing variation at an additional 22 SNPs with a mean r^2 of 0.942. At the much smaller *LHX8* gene, three SNPs were selected as tagging capturing variation at an additional 13 SNPs (Table S2, Supporting information) (mean $r^2 = 0.988$). In order to test for association of the studied SNPs with the TS phenotype, the transmission test for linkage disequilibrium (TDT) was performed, as implemented in Haploview (Barrett *et al*. 2005). Both single SNPs as well as three-SNP haplotypes were analyzed (tests with 1 degree of freedom, as implemented by Haploview). In order to correct for multiple comparisons, 1000 permutations were performed and the adjusted *P*-values are also reported here.

In order to test the possible association of variation at *LHX6* and *LHX8* with ancestry, data from the POPRES (population reference sample) (Nelson *et al*. 2008) was extracted from both regions. For each SNP, we computed its correlation with the top two principal components of the dataset (PCA scores), which have been shown to capture the most significant axes of genetic variation within Europe and particularly the North to South and West to East distribution of variation (Lao *et al*. 2008; Novembre e*t al.* 2008). PCA scores were calculated as previously described (Paschou *et al*. 2007, 2008), and compared with the distribution of PCA scores for all 447 212 available SNPs.

Results

Lhx6 and Lhx8 expression in the developing and adult basal ganglia is conserved in mammalian species

Lhx6 and Lhx8 proteins are highly conserved across vertebrates. In six mammalian species (*Homo sapiens, Pan troglodytes, Macaca mulatta, Bos taurus, Mouse musculus and Rattus norvegicus*), any pairwise comparison of the Lhx6 amino acid sequences (NP_055183, XP_001135172, XP_001089041, NP_001179777, NP_032526 and NP_001101307 respectively) or the Lhx8 amino acid sequences (NP_001001933, XP_524738, XP_001097664, XP_589896.4, NP_034843 and NP_001012219 respectively) using NCBI's

BLAST, showed at least 97% identity between them. These results indicate that Lhx6 and Lhx8 are subject to strong evolutionary pressure and reflect the conservation of their mode of action in the developing and adult nervous system.

We then performed comparative analysis of the expression of *Lhx6* and *Lhx8* by *in situ* hybridization on serial saggital sections of embryonic and adult brain of mouse and rat focusing our analysis in the striatum.

In the developing mouse brain low levels of *Lhx6* and *Lhx8* messenger RNAs first appeared at E11.5 in the newly formed subventricular zone (SVZ) of the MGE, the region of the ventral telencephalon that gives rise to the PV+, the SST+ and the cholinergic striatal interneurons. One day later, at E12.5, high levels of *Lhx6* and *Lhx8* expression were detected within the SVZ and the mantle zone (MZ) of the MGE with an overlapping but clearly distinct pattern: *Lhx6* expression domain spanned the entire SVZ and the dorsal part of the MZ, while *Lhx8* domain covered the ventral part of the SVZ and the MZ (compare Fig. 1a with Fig. 1g). At E13.5 the spatial distribution of *Lhx6* and *Lhx8* mRNAs remained the same (Fig. 1b,h). Notably, the levels of expression of *Lhx6* and *Lhx8* peak around this stage in which the generation of the majority of the striatal interneurons occurs. At late embryonic stages (E15.5–E18.5) lower levels of *Lhx6* and *Lhx8* mRNAs were detected in the striatum (not shown). A very similar pattern of expression was observed for both genes in the rat embryos at equivalent embryonic stages (E14.5 and E15.5 – compare Fig. 1a,b with Fig. 1d, e and Fig. 1g,h with Fig. 1j,k). In the adult mouse striatum expression of *Lhx6* and *Lhx8* was observed in distinct subsets of cells (compare Fig. 1c with Fig. 1i). *Lhx6* expressing cells (Fig. 1c) represent the PV+ and SST+ subpopulations of striatal interneurons while *Lhx8* expressing cells (Fig. 1i) the cholinergic internneurons. Analysis of the expression of *Lhx6* and *Lhx8* in the adult rat striatum showed that the pattern of expression of both genes is highly conserved (compare Fig. 1c with Fig. 1f and Fig. 1i with Fig. 1l). In summary, these data show that *Lhx6* and *Lhx8* expression in subpopulations of striatal interneurons that have been implicated in TS is conserved across mammalian species, suggesting that mutations in these genes may contribute to the development of TS.

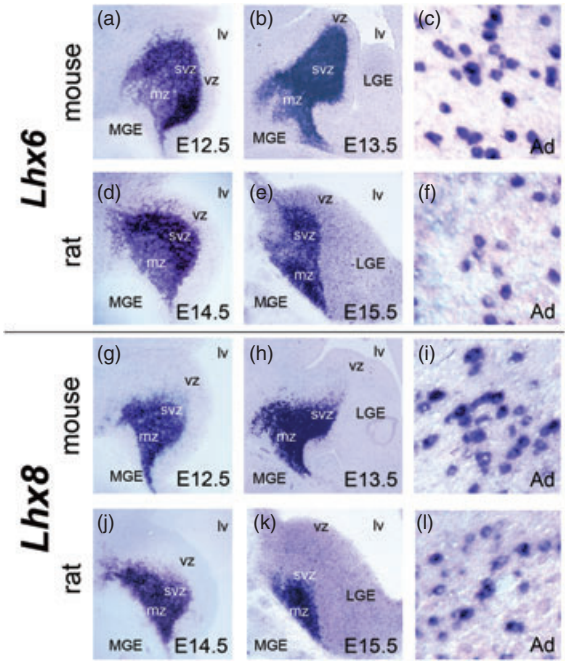

Figure 1: The expression pattern of *Lhx6* and *Lhx8* is highly conserved in the developing and adult striatum. a, b, d and e: Comparative analysis of the expression of *Lhx6* on saggital sections of embryonic mouse (a and b) and rat (d and e) at equivalent embryonic stages (compare a with d, and b with e). In both species *Lhx6* is expressed in the subventricular and mantle zone of the developing MGE which gives rise to the striatal interneurons. g, h, j and k: Comparative analysis of the expression of *Lhx8* on serial saggital sections of embryonic mouse (g and h) and rat (j and k) at equivalent embryonic stages (compare g with j, and h with k). In both species *Lhx8* is expressed in the subventricular and mantle zone of the developing MGE in a pattern that is overlapping with the expression pattern of *Lhx6* but distinct. *Lhx6* expression domain spanned the entire subventricular zone and the dorsal part of the mantle zone, while *Lhx8* domain covered the ventral part of the subventricular zone and the entire mantle zone (compare a with g, d with j, b with h and e with k). c, f, i and l: Comparative analysis of the expression of *Lhx6* (c and f) and *Lhx8* (i and l) on serial saggital sections of mouse (c and i) and rat (f and l) adult striatum. The expression of *Lhx6* and *Lhx8* is confined in different cellular populations (compare c with i, and f with l). Ad, adult; MGE, medial ganglionic eminence; LGE, lateral ganglionic eminence; vz, ventricular zone; svz, subventricular zone; mz, mantle zone; lv, lateral ventricle.

Investigating variation at LHX6 and LHX8 as candidate susceptibility regions for TS

At both *LHX6* and *LHX8*, we run the transmission/disequilibrium test (TDT) for families within each individual population as well as for the total TSGeneSEE dataset and the total TSGeneSEE and German datasets. At *LHX8*, results were negative for the single association tests as well as for the test of different haplotypes comprised of alleles at all three SNPs, both for the complete dataset and the subsets of the data that were studied (Table 1, Table S3, Supporting information, haplotype data not shown). At SNP rs729833, we observed a slight but non-significant over-transmission of one allele in the Italian population, yielding an uncorrected *P*-value of 0.07 for the TDT ($\chi^2 = 3.24$).

At *LHX6*, the situation was a bit more complex (Table 1 and Table S3, Supporting information). When analyzing the complete TSGeneSEE dataset (127 trios), allele A of SNP rs3808901 was found to be overtransmitted to children with TS ($\chi^2 = 5.14$, *P*-value = 0.02), although this slight over-transmission did not withstand correction for multiple testing. Nevertheless, a three-SNP haplotype around this SNP (Table 2 and Table S4, Supporting information), was also found to be significantly over-transmitted to TS patients ($\chi^2 = 11.28$, *P*-value $= 8 \times 10^{-4}$) and this time the test remained statistically significant even after performing 1000 permutations of the data (*P*-value = 0.006). However,

Table 1: Transmission test for linkage disequilibrium in four different populations (total of 222 trios) for single markers tested at *LHX6* and *LHX8* genes

	Hungarian (57 trios)			Italian (43 trios)			Polish (27 trios)			German (95 trios)			TSGeneSEE (127 trios)			TSGenesee + German (222 trios)		
rs#	Ov.	T:U	*P*-value	Ov.	T:U	*P*-value	Ov.	T:U	*P*-value	Ov.	T:U	*P*-value	Ov.	T:U	*P*-value	Ov.	T:U	*P*-value
LHX6																		
rs10985563	C	31:28	0.6961	T	23:17	0.3428	T	16:15	0.8575	C	47:45	0.8348	T	67:63	0.7257	T	112:110	0.8932
rs10818647	–	21:21	1.0	T	15:8	0.1444	C	15:10	0.3173	C	39:35	0.6419	T	46:44	0.833	C	83:81	0.8759
rs10818651	G	33:31	0.8026	A	25:18	0.2858	G	15:11	0.4328	G	53:39	0.1444	A	67:66	0.9309	G	119:106	0.3861
rs10985568	T	22:17	0.4233	C	14:12	0.6949	T	15:8	0.1444	C	30:24	0.4142	T	49:39	0.2864	T	73:69	0.7371
rs932922	–	22:22	1.0	C	14:13	0.8474	G	11:8	0.4913	C	31:30	0.8981	G	46:44	0.833	G	76:75	0.9351
rs3793621	A	31:27	0.5994	G	24:13	0.0705	A	15:13	0.7055	A	39:33	0.4795	G	64:59	0.6521	A	98:97	0.9429
rs10818652	A	29:26	0.6858	G	27:17	0.1317	G	19:13	0.2888	A	49:43	0.5316	G	72:59	0.256	G	115:108	0.6392
rs3808901	A	30:19	0.1161	A	20:14	0.3035	A	13:7	0.1797	G	42:35	0.425	**A**	**63:40**	**0.0234***	A	98:82	0.233
rs7849627	G	29:28	0.8946	A	22:16	0.3304	A	20:12	0.1573	G	47:46	0.9174	A	70:57	0.2487	A	116:104	0.4185
rs4837951	A	30:28	0.7928	A	20:17	0.6219	G	22:12	0.0863	G	53:48	0.6188	G	67:62	0.6598	G	120:110	0.5097
LHX8																		
rs729833	T	20:14	0.3035	C	17:08	0.0719	T	11:10	0.8273	C	35:33	0.8084	C	41:39	0.8231	C	76:72	0.7423
rs12732329	G	27:21	0.3865	G	23:19	0.5371	G	10:06	0.3173	G	33:32	0.9013	G	60:46	0.1739	G	93:78	0.2513
rs941032	C	32:28	0.6056	–	18:18	1.0	T	14:09	0.2971	T	48:41	0.4581	T	60:59	0.927	T	108:100	0.5791

The TSGeneSEE sample corresponds to the Hungarian, Italian, and Polish samples analyzed jointly.
Boldface indicates the significant findings.
Ov., over-transmitted allele; *P*, *P*-value of TDT; T:U, transmitted:untransmitted.
*Permutation *P*-value = 0.32 (1000 permutations performed).

analyzing the TSGeneSEE and German datasets jointly, yielded no statistically significant transmission disequilibrium of alleles to the affected children. Nevertheless, although no over-transmission was seen when tests for single SNPs were performed, the same three-SNP haplotype was over-transmitted ($\chi^2 = 3.53$, *P*-value = 0.06). This association did not remain statistically significant when 1000 permutation tests were performed.

Upon closer examination, we noted that, whereas the A allele of SNP rs3808901, which is found at the core of this haplotype, is over-transmitted in Hungarians, Polish and Italian patients, in the German population, it is the opposite allele (G) that is most often transmitted to individuals with TS, suggesting the possibility that different alleles may be implicated in increasing risk for the development of TS symptomatology in different populations.

Examining variation at LHX6 and LHX8 across different European populations

As shown in Fig. S1, Supporting information, the analysis of LD between studied SNPs in each of the studied populations, showed comparable patterns. The rare allele frequencies for each of the SNPs analyzed in the studied populations are presented in Table S5, Supporting information.

In order to investigate the possible differentiation of variation at *LHX6* and *LHX8* across European populations, we calculated the correlation of SNPs across these genes extracted from the POPRES dataset, with the top two principal components of the dataset (Table 3), that have been shown to correlate with ancestry across Europe (Drineas *et al*. 2010; Lao *et al*. 2008; Novembre *et al*. 2008). As we have analyzed in detail in earlier work, a high PCA score is expected for SNPs that show high association with population ancestry, and are thus highly differentiating among studied populations (Paschou *et al*. 2007, 2008, 2010). Remarkably, for *LHX6* the PCA score at SNP rs3808901 lies in the top 6% of all studied SNPs. As already described above, this SNP was shown to be significantly associated with TS in a subset of the families studied (TSGeneSEE sample) but not in the German population. For the sake of comparison we note that the most differentiated SNP in the lactase gene (*LCT*), which resides in one of the most prominent chromosomal regions that show high population differentiation within Europe and is considered to be the target of natural selection (Novembre *et al*. 2008), yields a PCA score that resides in the top 0.83%. Consequently, The SNP of interest in this study, rs3808901, with a PCA-score in the top 6% is also highly responsible for population differentiation among European populations.

Variation that is highly differentiated across European populations seems also to exist at *LHX8*, although we could not directly test the SNPs that we genotyped in our own sample, as they were not available in the POPRES dataset.

Discussion

Neuropathological and neurosurgical data as well as *in vivo* imaging studies, strongly implicate the basal ganglia and related cortical and thalamic structures in the pathobiology of TS (Albin *et al*. 2003; Frey *et al*. 2006; Kalanithi *et al*. 2005;

Table 2: Transmission test for linkage disequilibrium in four different populations (total of 222 trios) for a three SNP haplotype around SNP rs3808901 at the *LHX6* gene

	Hungarian (57 trios)			Italian (43 trios)			Polish (27 trios)			German (95 trios)			TSGeneSEE (127 trios)			TSGeneSEE + German (222 trios)		
Haplotype	Frequency	T:U	*P*	Frequency	T:U	*P*	Frequency	T:U	*P*	Frequency	T:U	*P*	Frequency	T:U	*P*	Frequency	T:U	*P*
GGA	**0.36**	**19:33**	**0.05**	0.28	17.2:18.0	0.90	0.26	12.9:11.9	0.84	0.33	47.0:42.7	0.65	0.32	49.1:62.9	0.19	0.32	96.1:105.5	0.51
AGG	0.27	26.9:23.9	0.67	0.27	15.8:19.7	0.52	0.43	11.0:18.8	0.15	0.36	44.9:42.9	0.83	0.30	53.7:62.2	0.43	0.32	98.6:105.2	0.65
GAA	0.18	20.9:9.9	0.05*	0.22	20.8:8.7	0.02†	0.19	11.1:5.0	0.12	0.16	24.8:32.3	0.32	0.19	52.8:23.5	8.0E-4‡	0.18	77.5:55.9	0.06§
AAA	0.12	15.1:13.1	0.71	0.13	6.2:12.3	0.15	0.05	2.9:3.0	0.95	0.08	14.2:13	0.82	0.11	24.2:28.5	0.55	0.10	38.5:41.5	0.73
GGG	0.07	6.1:8.1	0.60	0.09	6.2:7.6	0.70	0.07	4.0:4.3	0.92	0.07	11.1:12.1	0.84	0.08	16.3:20.1	0.53	0.08	27.4:32.1	0.54
AGA										0.01	1:1	0.99						

The TSGeneSEE sample corresponds to the Hungarian, Italian, and Polish samples analyzed jointly.
P = *P*-value of TDT; T:U = transmitted:untransmitted.
Boldface indicates the significant findings.
* Permutation *P*-value = 0.20 (1000 permutations performed).
† Permutation *P*-value = 0.09 (1000 permutations performed).
‡ Permutation *P*-value = 0.006 (1000 permutations performed).
§ Permutation *P*-value = 0.27 (1000 permutations performed).

Table 3: PCA scores for variation at *LHX6* and *LHX8* genes, studying 1200 SNPs from 11 European populations (POPRES dataset, Novembre *et al*. 2008)

rs#	Position	Percentage
LHX6 chr9:123999000-124036000		
885404	124000747	85.45
7030904	124000776	79.98
10818647*	124007597	24.95
10985564	124007731	8.63
7873828	124009654	84.95
10818651*	124014240	41.37
10985568*	124019847	61.98
10818652*	124025606	77.30
3808901*	124032492	5.94
3808899	124032645	40.28
7849627*	124033326	24.07
LHX8 chr1:75368000-75405000		
1526512	75385685	42.67
12138223	75387614	7.21
LCT chr2:136540000-136600000;		
6707117	136571975	78.28
2734871	136586343	0.83

The *LCT* gene results are shown for comparison. SNPs that were also genotyped in our sample are marked by an asterisk. Note that the percentage represents the number of SNPs (out of 447 212) with a higher PCA score with respect to the top two principal components in the POPRES dataset. Thus, if the percentage of SNP X is 7% that means that approximately 31 000 SNPs had a higher score (and thus were more correlated with the top two principal components) than SNP X.

Kataoka *et al*. 2010; Peterson *et al*. 2003; Sowell *et al*. 2008). Yet, the cellular and molecular mechanisms implicated in the pathophysiology of TS remain poorly understood. Lhx6 and Lhx8 transcription factors are expressed in the developing and adult mouse basal ganglia and are required for the specification of striatal interneurons: *Lhx6* is required for the specification of the PV+ and SST+ subpopulation while *Lhx8* for the specification of the cholinergic subpopulation (Bachy and Retaux 2006; Fragkouli *et al*. 2005, 2009; Liodis *et al*. 2007; Zhao *et al*. 2003, 2008). These two subpopulations of striatal interneurons are of paramount importance for the regulation of striatum activity (Kreitzer 2009; Tepper *et al*. 2007, 2010) as they modify the activity of the medium spiny projection neurons (MSNs), the principal cell type of the striatum which receives cortical input and targets the substantia nigra and the globus pallidus (Tepper *et al*. 2007, 2010). In the traditional neuronal circuits of the basal ganglia cortical input excites the GABAergic PV+ and SST+ subpopulation which acts to inhibit the MSNs. Thus, the absence of these populations would probably result in hyperactivity of the MSNs neurons. Notably, data from *in vivo* recording experiments suggest that the GABAergic PV+ are implicated in synchronizing striatal oscillations therefore, a deficit in the inhibitory network would probably lead to a tic-like behavior (Berke *et al*. 2004; Courtemanche *et al*. 2003). Interestingly, recent studies on postmortem forebrain tissues from TS-affected individuals have shown a significant selective decrease in the number of GABAergic PV+ interneurons of the basal ganglia (Kalanithi *et al*. 2005; Kataoka *et al*. 2010). Cholinergic striatal interneurons receive inputs from the cortex, the substantia nigra and the thalamus and modulate the activity of MSNs and GABAergic interneurons (Cragg 2006; de Boer & Abercrombie 1996; Pakhotin *et al*. 2007; Tepper *et al*. 2007, 2010). Analysis of postmortem tissues from TS-affected individuals has shown a significant decrease in the number of cholinergic striatal interneurons of the basal ganglia (Kataoka *et al*. 2010). In accordance, several studies have shown that neuroleptics and acetylcholinesterase inhibitors, which act by increasing the level and the duration of action of acetylcholine, are effective in treating motor and phonic tics in TS as well as stereotyped behavior in OCDs (Aliane *et al*. 2010; Bonsi *et al*. 2011; Cubo *et al*. 2008; Shprecher and Kurlan 2009; Silver *et al*. 2001).

Given the role of *Lhx6* and *Lhx8* in the development of specific subtypes of striatal interneurons and the results from the analysis of postmortem basal ganglia from TS-affected individuals we sought to investigate the involvement of these genes in the pathophysiological pathways of TS. To this end, we first studied the conservation of the sequence and the expression pattern of these two genes, in the striatum as this structure is implicated in the pathobiology of TS. *Lhx6* and *Lhx8* sequences are highly conserved across vertebrates a fact that reflects the conservation of their mode of action. Moreover, the expression patterns of *Lhx6* and *Lhx8* in the developing and adult striatum are highly conserved between mouse and rat. In both species these genes are expressed in the embryonic and adult striatum in specific subpopulations of differentiating and mature neuronal populations: *Lhx6* in the PV+ and SST+ striatal interneurons and *Lhx8* in the cholinergic striatal interneurons. These results suggest that the expression patterns of *Lhx6* and *Lhx8* are likely to be conserved in other mammalian species including humans. These data along with the high sequence conservation and the data from previous studies that have established the role of *Lhx6* and *Lhx8* in the aforementioned neuronal subpopulations as well as the implication of these interneurons in TS, suggest that *Lhx6* and *Lhx8* are probably involved in the pathophysiological pathways implicated in TS.

We then analyzed two independent family samples of European descent. Our analysis was negative for association of *LHX8*. However, we found positive association of the TS phenotype with *LHX6* in one of the samples studied (TSGeneSEE dataset), but no association in our sample of German families. Interestingly, the SNP allele that was identified to be significantly associated in the TSGeneSEE dataset, showed an opposite trend of transmission in the German dataset. Notably, a three-SNP haplotype around this SNP showed very significant association with TS in the first sample (uncorrected $P = 8 \times 10^{-4}$, corrected $P = 0.006$), while the result was diluted and did not withstand correction for multiple testing (uncorrected $P = 0.06$) when both datasets were analyzed jointly.

It should be noted that a limitation of our study is reduced power because of a relatively limited sample size, although, we should also note that trios-based designs are much more powerful than simple case-control studies (Laird and Lange

2006). In any case, our results should only be considered indicative and further studies of larger sample size are warranted in this region. An additional point to consider is the possible heterogeneity between the two samples (German and TSGeneSEE) because of collection at different sites and different time-points. For instance, as shown in Table S1, Supporting information, the German sample contains a much larger proportion of patients with OCD (33% vs. 8.71% on average for the TSGeneSEE sample). So it is possible to speculate that this heterogeneity, might also be the reason for the observed differences in genetic background.

Alternatively, our analysis indicates, that ancestry differences may also account for the observed contradictory findings at the *LHX6* locus. Within Europe, two major axes of genetic variation are observed, corresponding to the North–South and West–East axes of origin (Drineas *et al*. 2010; Lao *et al*. 2008; Novembre *et al*. 2008). Ancestry Informative SNPs lie at the extremes of this distribution and can even be used to predict ancestry down to a few hundred kilometers of self-reported origin (Drineas *et al*. 2010). Such differences may be subtle but can be showed through powerful techniques such as PCA (Novembre *et al*. 2008; Paschou *et al*. 2007). Although, the TDT is robust to population stratification, our analysis of the correlation of the *LHX6* region with individual ancestry within Europe, showed the fact that this particular SNP shows a high degree of population differentiation. The same would be true for unstudied variants that are in high LD with this SNP. This could be the result of stochastic factors and demographic history, but it could also reflect the pressure of natural selection. As noted earlier, because of the relatively small number of families in each individual sample, our results should be considered indicative and support from additional studies is needed. Nevertheless, it is intriguing to speculate that different genetic factors may contribute to the etiology of TS in different populations, even within Europe. Such a hypothesis would be in accordance with the observed difficulty in replicating findings among different studies.

References

Albin, R.L., Koeppe, R.A., Bohnen, N.I., Nichols, T.E., Meyer, P., Wernette, K., Minoshima, S., Kilbourn, M.R. & Frey, K.A. (2003) Increased ventral striatal monoaminergic innervation in Tourette syndrome. *Neurology* **12**, 310–315.

Aliane, V., Perez, S., Bohren, Y., Deniau, J.M. & Kemel, M.L. (2010) Key role of striatal cholinergic interneurons in processes leading to arrest of motor stereotypies. *Brain* **134**, 110–118.

Bachy, I. & Rétaux, S. (2006) GABAergic specification in the basal forebrain is controlled by the LIM-hd factor Lhx7. *Dev Biol* **291**, 218–226.

Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265.

Berke, J.D., Okatan, M., Skurski, J. & Eichenbaum, H.B. (2004) Oscillatory entrainment of striatal neurons in freely moving rats. *Neuron* **43**, 883–896.

Bloch, M.H. & Leckman, J.F. (2009) Clinical course of Tourette syndrome. *J Psychosom Res* **67**, 497–501.

Bonsi, P., Cuomo, D., Martella, G., Madeo, G., Schirinzi, T., Puglisi, F., Ponterio, G. & Pisani, A. (2011) Centrality of striatal cholinergic transmission in Basal Ganglia function. *Front Neuroanat* **5**, 1–9.

Cath, D.C., Hedderly, T., Ludolph, A.G., Stern, J.S., Murphy, T., Hartmann, A., Czernecki, V., Robertson, M.M., Martino, D., Munchau, A. & Rizzo, R., ESSTS Guidelines Group (2011) European clinical guidelines for Tourette syndrome and other tic disorders. Part I: assessment. *Eur Child Adolesc Psychiatry* **20**, 155–171.

Courtemanche, R., Fujii, N. & Graybiel, A.M. (2003) Synchronous, focally modulated beta-band oscillations characterize local field potential activity in the striatum of awake behaving monkeys. *J Neurosci* **23**, 11741–11752.

Cragg, S.J. (2006) Meaningful silences: how dopamine listens to the ACh pause. *Trends Neurosci* **29**, 125–131.

Cubo, E., Fernandez Jaen, A., Moreno, C., Anaya, B., Gonzalez, M. & Kompoliti, K. (2008) Donepezil use in children and adolescents with tics and attention-deficit/hyperactivity disorder: an 18-week, single-center, dose-escalating, prospective, open-label study. *Clin Ther* **30**, 182–189.

De Boer, P. & Abercrombie, E.D. (1996) Physiological release of striatal acetylcholine in vivo: modulation by D1 and D2 dopamine receptor subtypes. *J Pharmacol Exp Ther* **277**, 775–783.

Drineas, P., Lewis, J. & Paschou, P. (2010) Inferring Geographic Coordinates of Origin for Europeans using Small Panels of Ancestry Informative Markers. *PLoS ONE* **5**, e11892.

Flames, N., Pla, R., Gelman, D.M., Rubenstein, J.L., Puelles, L. & Marin, O. (2007) Delineation of multiple subpallial progenitor domains by the combinatorial expression of transcriptional codes. *J Neurosci* **27**, 9682–9695.

Fragkouli, A., Hearn, C., Errington, M., Cooke, S., Grigoriou, M., Bliss, T., Stylianopoulou, F. & Pachnis, V. (2005) Loss of forebrain cholinergic neurons and impairment in spatial learning and memory in LHX7-deficient mice. *Eur J Neurosci* **21**, 2923–2938.

Fragkouli, A., Verhey van Wijk, N., Lopes, R., Kessaris, N. & Pachnis, V. (2009) LIM homeodomain transcription factor-dependent specification of bipotential MGE progenitors into cholinergic and GABAergic striatal interneurons. *Development* **136**, 3841–3851.

Frey, K.A. & Albin, R.L. (2006) Neuroimaging of Tourette syndrome. *J Child Neurol* **21**, 672–677.

Grados, M.A. (2010) The genetics of obsessive-compulsive disorder and Tourette syndrome: an epidemiological and pathway-based approach for gene discovery. *J Am Acad Child Adolesc Psychiatry* **49**, 810–819.

Grigoriou, M., Tucker, A.S., Sharpe, P.T. & Pachnis, V. (1998) Expression and regulation of Lhx6 and Lhx7, a novel subfamily of LIM homeodomain encoding genes, suggests a role in mammalian head development. *Development* **125**, 2063–2074.

Hebebrand, J., Klug, B., Fimmers, R., Seuchter, S.A., Wettke-Schäfer, R., Deget, F., Camps, A., Lisch, S., Hebebrand, K., von Gontard, A., Lehmkuhl, G., Poustka, F., Schmidt, M., Baur, M.P. & Remschmidt, H. (1997) Rates for tic disorders and obsessive compulsive symptomatology in families of children and adolescents with Gilles de la Tourette syndrome. *J Psychiatr Res* **31** 519–530. 5.

Kalanithi, P.S., Zheng, W., Kataoka, Y., DiFiglia, M., Grantz, H., Saper, C.B., Schwartz, M.L., Leckman, J.F. & Vaccarino, F.M. (2005) Altered parvalbumin-positive neuron distribution in basal ganglia of individuals with Tourette syndrome. *Proc Natl Acad Sci U S A* **102**, 13307–13312.

Kataoka, Y., Kalanithi, S.A., Grantz, H., Schwartz, M.L., Saper, C., Leckman, J.F. & Vaccarino, F.M. (2010) Decreased Number of parvalbumin and cholinergic interneurons in the striatum of individuals with Tourette syndrome. *J Comp Neurol* **518**, 277–291.

Kreitzer, A.C. (2009) Physiology and pharmacology of striatal neurons. *Ann Rev Neurosci* **32**, 127–147.

Laird, N.M. & Lange, C. (2006) Family-based designs in the age of large-scale gene-association studies. *Nat Rev Genet* **7**, 385–394.

Lao, O., Lu, T.T., Nothnagel, M. *et al*. (2008) Correlation between genetic and geographic structure in Europe. *Curr Biol* **18**, 1241–1248.

Lavdas, A.A., Grigoriou, M., Pachnis, V. & Parnavelas, J.G. (1999) The medial ganglionic eminence gives rise to a population of

early neurons in the developing cerebral cortex. *J Neurosci* **19**, 7881–7888.
Liodis, P., Denaxa, M., Grigoriou, M., Akufo-Addo, C., Yanagawa, Y. & Pachnis, V. (2007) Lhx6 activity is required for the normal migration and specification of cortical interneuron subtypes. *J Neurosci* **27**, 3078–3089.
Matsumoto, K., Tanaka, T., Furuyama, T., Kashihara, Y., Mori, T., Ishii, N., Kitanaka, J., Takemura, M., Tohyama, M. & Wanaka, A. (1996) L3, a novel murine LIM-homeodomain transcription factor expressed in the ventral telencephalon and the mesenchyme surrounding the oral cavity. *Neurosci Lett* **204**, 113–116.
Nelson, M.R., Bryc, K., King, K.S. *et al*. (2008) The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. *Am J Hum Genet* **83**, 347–358.
Novembre, J., Johnson, T., Bryc, K., Kutalik, Z., Boyko, A.R., Auton, A., Indap, A., King, K.S., Bergmann, S., Nelson, M.R., Stephens, M. & Bustamante, C.D. (2008) Genes mirror geography within Europe. *Nature* **456**, 98–101.
O'Rourke, J.A., Scharf, J.M., Yu, D. & Pauls, D.L. (2009) The genetics of Tourette syndrome: a review. *J Psychosom Res* **67**, 533–545.
Pakhotin, P. & Bracci E. (2007) Cholinergic interneurons control the excitatory input to the striatum. *J Neurosci* **27**, 391–400.
Paschou, P., Ziv, E., Burchard, E.G., Choudry, S., Rodriguez-Cintron, W., Mahoney, M.W. & Drineas, P. (2007) PCA-correlated SNPs for structure identification in worldwide human populations. *PLoS Genet* **3**, e160.
Paschou, P., Drineas, P., Lewis, J., Nievergelt, C.M., Nickerson, D.A., Smith, J.D., Ridker, P.M., Chasman, D.I., Krauss, R.M. & Ziv, E. (2008) Tracing sub-structure in the European American population with PCA-informative markers. *PLoS Genet* **4,** e1000114.
Paschou, P., Lewis, J., Javed, A. & Drineas, P. (2010) Ancestry informative markers for fine-scale individual assignment to worldwide populations. *J Med Genet* **47**, 835–847.
Peterson, B.S., Thomas, P., Kane, M.J., Scahill, L., Zhang, H., Bronen, R., King, R.A., Leckman, J.F. & Staib, L. (2003) Basal Ganglia volumes in patients with Gilles de la Tourette syndrome. *Arch Gen Psychiatry* **60**, 415–424.
Schoenian, S., Konig, I., Oertel, W., Remschmidt, H., Ziegler, A., Hebebrand, J. & Bandmann, O. (2003) HLA-DRB genotyping in Gilles de la Tourette patients and their parents. *Am J Med Genet B: Neuropsychiatr Genet* **119B**, 60–64.
Shprecher, D. & Kurlan, R. (2009) The management of tics. *Mov Disord* **24**, 15–24.
Silver, A.A., Shytle, R.D., Sheehan, K.H., Sheehan, D.V., Ramos, A. & Sanberg, P.R. (2001) Multicenter, double-blind, placebo-controlled study of mecamylamine monotherapy for Tourette's disorder. *J Am Acad Child Adolesc Psychiatry* **40**, 1103–1110.
Sowell, E.R., Kan, E., Yoshii, J., Thompson, P.M., Bansal, R., Xu, D., Toga, A.W. & Peterson, B.S. (2008) Thinning of sensorimotor cortices in children with Tourette syndrome. *Nat Neurosci* **11**, 637–639.
State, M.W. (2011) The genetics of Tourette disorder. *Curr Opin Genet Dev* **21**, 302–309.
Swain, J.E., Scahill, L., Lombroso, P.J., King, R.A. & Leckman, J.F. (2007) Tourette syndrome and tic disorders: a decade of progress. *J Am Acad Child Adolesc Psychiatry* **46**, 947–968.
Tepper, J.M., Abercrombie, E.D. & Bolam, J.P. (2007) Basal ganglia macrocircuits. *Prog Brain Res* **160**, 3–7.
Tepper, J.M., Tecuapetla, F., Koós, T. & Ibáñez-Sandoval, O. (2010) Heterogeneity and diversity of striatal GABAergic interneurons. *Front Neuroanat* **4**, 1–18.
Tourette Syndrome Association International Consortium for Genetics. (2007) Genome scan for Tourette disorder in affected-sibling-pair and multigenerational families. *Am J Hum Genet* **80**, 265–272.
Zhao, Y., Marin, O., Hermesz, E., Powell, A., Flames, N., Palkovits, M., Rubenstein, J.L. & Westphal, H. (2003) The LIM-homeobox gene Lhx8 is required for the development of many cholinergic neurons in the mouse forebrain. *Proc Natl Acad Sci U S A* **100**, 9005–9010.
Zhao, Y., Flandin, P., Long, J.E., Cuesta, M.D., Westphal, H. & Rubenstein, J.L. (2008) Distinct molecular pathways for development of telencephalic interneuron subtypes revealed through analysis of Lhx6 mutants. *J Comp Neurol* **510**, 79–99.

Acknowledgments

This study was funded by a Tourette Syndrome Association, Research Grant Award to PP, and supported by COST Action BM0905 (http://tourette-eu.org).

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1: Comparative analysis of *Lhx6* and *Lhx8* expression in the developing and adult brain of mouse and rat-*LHX6* and *LHX8* genetic association study samples and methods.

Appendix S2: Members of the TSGeneSEE consortium.

Figure S1: Pairwise LD tests (*r2*) in our German, Hungarian, Italian, and Polish samples, for all studied SNPs in the *LHX6* and the *LHX8* genes.

Table S1: Phenotypic details of TSGeneSEE (Hungarian, Italian, and Polish), and German sample. Patient numbers are shown in parentheses.

Table S2: Selection of tSNPs at *LHX6* and *LHX8* using the HapMap CEPH European population as reference (http://hapmap.ncbi.nlm.nih.gov/). The r^2 threshold for tSNP selection was set to 0.8. Variation at *LHX6* and *LHX8* was captured by the selected tSNPs with a mean $r^2$of 0.942 and 0.988, respectively.

Table S3a: Transmission test for linkage disequilibrium in four different populations (total of 222 trios) for single markers tested at *LHX6* and *LHX8* genes (test implemented and *P*-values determined by Haploview).

Table S3b: Transmission test for linkage disequilibrium in joint analysis of four different populations (total of 222 trios) for single markers tested at *LHX6* and *LHX8* genes (test implemented and *P*-values determined by Haploview). The TSGeneSEE sample corresponds to the Hungarian, Italian, and Polish samples analyzed jointly.

Table S4: Transmission test for linkage disequilibrium in four different populations (total of 222 trios) for a three SNP haplotype around SNP rs3808901 at the *LHX6* gene (test implemented and P values determined by Haploview). The TSGeneSEE sample corresponds to the Hungarian. Italian. and Polish samples analyzed jointly.

Table S5: Allele frequencies of studied SNPs at *LHX6* and *LHX8* in all studied populations. Position refers to Build 36 of the genome.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.