Accepted Article

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

Peristera Paschou(1)*, Electra Stylianopoulou(1) Iordanis Karagiannidis(1), Renata Rizzo(2), Zsanett Tarnok(3), Tomasz Wolanczyk(4), Johannes Hebebrand(5), Markus M. Nöthen(6), Gerd Lehmkuhl (7), Luca Farkas(3), Peter Nagy(3), Urszula Szymanska(4), Dimitris Lykidis(1), Christos Androutsos (8), Vaia Tsironi (8), Anastasia Koumoula (8), Csaba Barta(9), Sotiris Klidonas(1), Petros Ypsilantis(10), Constantinos Simopoulos(10), TSGeneSEE(11), George Skavdis(1) and Maria Grigoriou(1)*

***Corresponding authors**

1- Department of Molecular Biology and Genetics, Democritus University of Thrace, Greece
2- Materno-Infantile and Radiological Science Department, University of Catania, Italy
3- Vadaskert Clinic for Child and Adolescent Psychiatry, Hungary
4- Department of Child Psychiatry, Medical University of Warsaw, Poland
5- Department of Child and Adolescent Psychiatry, University of Duisburg-Essen, Essen, Germany
6- Institute for Human Genetics, University of Bonn, Germany
7- Department of Child and Adolescent Psychiatry of the University of Cologne, Germany
8- Child and Adolescent Psychiatry Clinic, Sismanoglio General Hospital of Attica, Greece
9- Department of Medical Chemistry, Molecular Biology and Pathobiochemistry, Semmelweis University, Hungary
10- School of Medicine, Democritus University of Thrace, Alexandroupolis, Greece
11- The Tourette Syndrome Genetics-Southern and Eastern Europe Initiative

Running Title: *LHX6* and *LHX8* as candidate genes for TS

Date of submission of revised version: 29/1/2012

Number of words in the Abstract: 246

Number of words in the Introduction: 506

Number of words in Discussion: 1137

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/j.1601-183X.2012.00778.x

© 2012 The Authors
Genes, Brain and Behavior © 2012 Blackwell Publishing Ltd and International Behavioural and Neural Genetics Society

Accepted Article

Abstract

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-SNPs-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: Tourette Syndrome, *LHX6, LHX8,* neuroanatomical expression, genetic association

Accepted Article

Introduction

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 (e.g. see Grados, 2010; O'Rourke et al., 2009; State, 2011 for a review) 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, 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 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;Fragkouli et al., 2009;Zhao et al., 2003).

Accepted Article

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 Demokritos 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 to 22 °C, on a 12:12h light:dark cycle, and were given commercial pelleted diet (4RF25, Mucedola, Milan, Italy) and water ad libitum. Animal breeding and experiments were done in accordance with the European Communities Council Directive of 24 November 1986 (86/609/EEC). 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 x 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), according to manufacturer's instructions, using Digoxigenin-11-UTP (Roche). 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 cDNA) 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

Accepted Article

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 supplementary table 1. 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 the supplementary material. Tagging SNPs (tSNPs) for both *LHX6* and *LHX8* were selected using the HapMap CEPH European population as reference. A total of 10 tagging SNPs (tSNPs) were selected for the *LHX6* region (supplementary table 2) 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 (supplementary table 2) (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 was performed, as implemented in Haploview (Barrett et al., 2005). Both single SNPs as well as three-SNP haplotypes were analyzed (tests with one degree of freedom, as implemented by Haploview). In order to correct for multiple comparisons, 1,000 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

Accepted Article

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 et al., 2008). PCA scores were calculated as we have previously described (Paschou et al., 2007;Paschou et al., 2008), and compared to 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, revealed 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* mRNAs first appeared at E11.5 in the newly formed subventricular zone (SVZ) of the medial ganglionic eminences (MGE), the region of the ventral telencephalon that gives rise to the PV+, the somatostatin (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 1G). At E13.5 the spatial distribution of *Lhx6* and *Lhx8* mRNAs remained the same (Fig 1B, 1H). 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

Accepted Article

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 in Fig. 1, A, B to D, E and G, H to J, K). In the adult mouse striatum expression of *Lhx6* and *Lhx8* was observed in distinct subsets of cells (compare in Fig. 1 serial sections C and I). *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 revealed that the pattern of expression of both genes is highly conserved (compare in Fig. 1, C with F, and I with L). 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.

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

At both *LHX6* and *LHX8*, we run the 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, supplementary table 3, 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 (χ^2=3.24).

At *LHX6*, the situation was a bit more complex (table 1 and supplementary table 3). When analyzing the complete TSGeneSEE dataset (127 trios), allele A of SNP rs3808901 was found to be over-transmitted to children with TS (χ^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 supplementary table 4), was also found to be significantly over-transmitted to TS patients (χ^2=11.28, P value=8X10^{-4}) and this time the test remained statistically significant even after performing 1,000 permutations of the data (P value = 0.006). However, 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 (χ^2=3.53, P value = 0.06).

Accepted Article

This association did not remain statistically significant when 1,000 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 supplementary figure 1, the analysis of LD between studied SNPs in each of the studied populations, revealed comparable patterns. The rare allele frequencies for each of the SNPs analyzed in the studied populations are presented in supplementary table 5.

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;Paschou et al., 2008;Paschou et al., 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 demonstrate 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.

Accepted Article

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, 2003; Frey et al., 2006;Kalanithi et al., 2005;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;Zhao et al., 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+ subpopulations 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 (Courtemanche et al., 2003;Berke et al., 2004). 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;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 behaviour in obsessive compulsive disorders (Aliane et al., 2010;Bonsi et al., 2011;Silver et al., 2001;Cubo et al., 2008;Shprecher and Kurlan, 2009).

Accepted Article

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 $P=8x10^{-4}$, 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 due to 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) due to collection at different sites and different time-points. For instance, as shown in supplementary table 1, the German sample contains a much larger

Accepted Article

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 (Novembre et al. 2008, Lao et al. 2008, Drineas et al. 2010). 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 revealed 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, revealed the fact that this particular SNP demonstrates 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, due to 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.

Acknowledgements

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

Correspondence to

Peristera Paschou

Dept. of Molecular Biology and Genetics, Democritus University of Thrace

Panepistimioupoli, Dragana, Greece

ppaschou@mbg.duth.gr

Accepted Article

Maria Grigoriou

Dept. of Molecular Biology and Genetics, Democritus University of Thrace

Panepistimioupoli, Dragana, Greece

mgrigor@mbg.duth.gr

Accepted Article

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. *Developmental Biology*, **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. *Journal of Psychosomatic Research*, **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. *Frontiers in Neuroanatomy*, **5**, 1- 9 .

Cath DC, Hedderly T, Ludolph AG, Stern JS, Murphy T, Hartmann A, Czernecki V, Robertson MM, 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(4):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. *Journal of Neuroscience*, **23**, 11741-11752.

Cragg, S.J. (2006). Meaningful silences: how dopamine listens to the ACh pause. *Trends in Neurosciences*, **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. *Clinical Therapeutics*, **30**, 182–189.

Accepted Article

De Boer, P., & Abercrombie, E.D. (1996). Physiological release of striatal acetylcholine in vivo: modulation by D1 and D2 dopamine receptor subtypes. *Journal of Pharmacology and Experimental Therapeutics*, **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. *Journal of Neuroscience*, **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. *European Journal of Neuroscience*, **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. *Journal of Child Neurology*, 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. *Journal of the American Academy of Child and Adolescent 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 SA, Wettke-Schäfer R, Deget F, Camps A, Lisch S, Hebebrand K, von Gontard A, Lehmkuhl G, Poustka F, Schmidt M, Baur MP, 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(5):519-530.

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

Accepted Article

distribution in basal ganglia of individuals with Tourette syndrome. *Proceedings of the National Academy of Sciences of the USA*, **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. *Journal of Comparative Neurology*, **518**, 277–291.

Kreitzer, A.C. (2009). Physiology and pharmacology of striatal neurons. *Annual Review of Neuroscience*, **32**, 127-147.

Laird NM, Lange C (2006). Family-based designs in the age of large-scale gene-association studies. Nat Rev Genet 7(5):385-394.

Lao, O., Lu, T.T., Nothnagel, M., Junge, O., Freitag-Wolf, S., Caliebe, A., Balascakova, M., Bertranpetit, J., Bindoff, L.A., Comas, D., Holmlund, G., Kouvatsi, A., Macek, M., Mollet, I., Parson, W., Palo, J., Ploski, R., Sajantila, A., Tagliabracci, A., Gether, U., Werge, T., Rivadeneira, F., Hofman, A., Uitterlinden, A.G., Gieger, C., Wichmann, H.E., Rüther, A., Schreiber, S., Becker, C., Nürnberg, P., Nelson, M.R., Krawczak, M., & Kayser, M. (2008). Correlation between genetic and geographic structure in Europe. *Current Biology*, **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. *Journal of Neuroscience*, **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. *Journal of Neuroscience*, **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. *Neuroscience Letters*, **204**, 113-116.

Nelson, M.R., Bryc, K., King, K.S., Indap, A., Boyko, A.R., Novembre, J., Briley, L.P., Maruyama, Y., Waterworth, D.M., Waeber, G., Vollenweider, P., Oksenberg, J.R., Hauser, S.L., Stirnadel, H.A., Kooner, J.S., Chambers, J.C., Jones, B., Mooser, V., Bustamante, C.D., Roses, A.D., Burns, D.K., Ehm, M.G., & Lai, E.H. (2008). The Population Reference

Accepted Article

Sample, POPRES: a resource for population, disease, and pharmacological genetics research. *American Journal of Human Genetics*, **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. *Journal of Psychosomatic Research*, **67**, 533-545.

Pakhotin, P., & Bracci, E. (2007). Cholinergic interneurons control the excitatory input to the striatum. *Journal of Neuroscience*, **27**, 391–400.

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 Genetics*, **4**,e1000114.

Paschou, P., Lewis, J., Javed, A., & Drineas, P. (2010). Ancestry informative markers for fine-scale individual assignment to worldwide populations. *Journal of Medical Genetics*, **47**, 835-847.

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 Genetics* **3**, e160.

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. *Archives of General 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. *American Journal of Medical Genetics Part B*: *Neuropsychiatric Genetics*, **119B**, 60-64.

Shprecher, D., & Kurlan, R. (2009). The management of tics. *Movement disorders*, **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. *Journal of the American Academy of Child and Adolescent Psychiatry*, **40**, 1103-1110.

Accepted Article

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. *Nature Neuroscience*, **11**, 637– 639.

State, M.W. (2011). The genetics of Tourette disorder. *Current Opinion in Genetics and Development* in press.

Swain, J.E., Scahill, L., Lombroso, P.J., King, R.A., & Leckman, J.F. (2007). Tourette syndrome and tic disorders: a decade of progress. *Journal of the American Academy of Child and Adolescent Psychiatry*, **46**, 947-968.

Tepper, J.M., Tecuapetla, F., Koós, T., & Ibáñez-Sandoval, O. (2010). Heterogeneity and diversity of striatal GABAergic interneurons. *Frontiers in Neuroanatomy*, **4**, 1-18.

Tepper, J.M., Abercrombie, E.D., & Bolam, J.P. (2007). Basal ganglia macrocircuits. *Progress in Brain Research*, **160**, 3-7.

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(2):265-272.

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. *Journal of Comparative Neurology*, **510**, 79-99.

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. *Proceedings of the National Academy of Sciences of the USA*, **100**, 9005-9010.

Accepted Article

Figure Legends

Fig. 1. The expression pattern of *Lhx6* and *Lhx8* is highly conserved in the developing and adult striatum. A, B, D, E : Comparative analysis of the expression of *Lhx6* on saggital sections of embryonic mouse (A, B) and rat (D, 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, K: Comparative analysis of the expression of *Lhx8* on serial saggital sections of embryonic mouse (G, H) and rat (J, 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, L: Comparative analysis of the expression of *Lhx6* (C, F) and Lhx8 (I, L) on serial saggital sections of mouse (C, I) and rat (F, 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.

Accepted Article

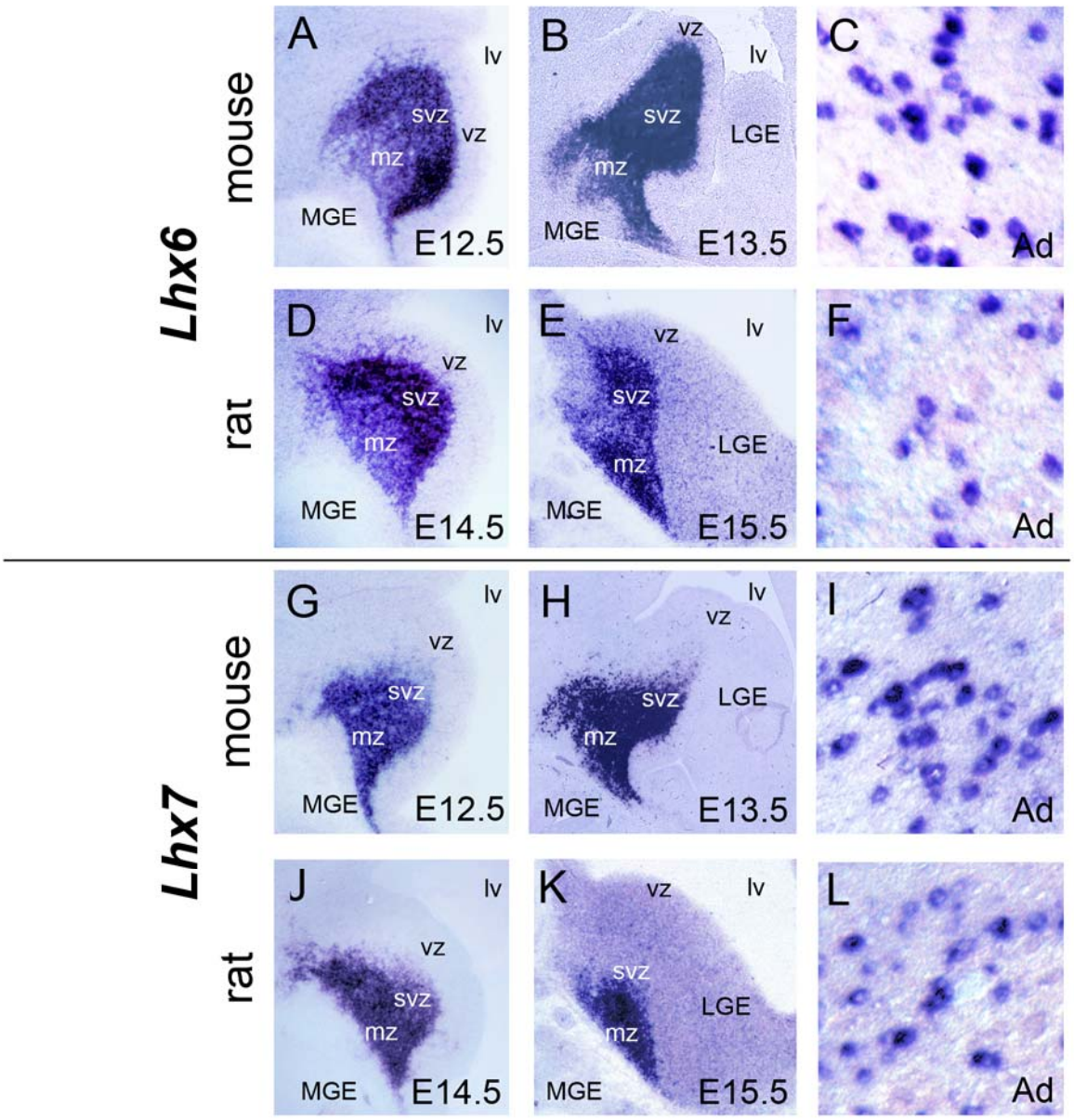


Table 1. Transmission test for linkage disequilibrium in four different populations (total of 222 trios) for single markers tested at LHX6 and LHX8 genes. The TSGeneSEE sample corresponds to the Hungarian, Italian, and Polish samples analyzed jointly.
(T:U=Transmitted:Untransmitted, P=P value of TDT, Ov.=Overtransmitted allele)

	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**[1]	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

[1] Permutation P value=0.32 (1000 permutations performed)

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. The TSGeneSEE sample corresponds to the Hungarian, Italian, and Polish samples analyzed jointly. (T:U=Transmitted:Untransmitted, P=P value of TDT).

	Hungarian (57 trios)			Italian (43 trios)			Polish (27 trios)			German (95 trios)			TSGeneSEE (127 trios)			TSGeneSEE + German (222 trios)		
Hapl.	**freq**	**T:U**	**P**	**freq**	**T:U**	**P**	**freq**	**T:U**	**P**	**freq**	**T:U**	**P**	**freq**	**T:U**	**P**	**freq**	**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**[1]	**0.22**	**20.8 : 8.7**	**0.02**[2]	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**[3]	**0.18**	**77.5 : 55.9**	**0.06**[4]
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						

[1] Permutation P value=0.20 (1000 permutations performed)
[2] Permutation P value=0.09 (1000 permutations performed)
[3] Permutation P value=0.006 (1000 permutations performed)
[4] Permutation P value=0.27 (1000 permutations performed)

Accepted Article

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

LHX6 chr9:123999000-124036000

rs#	position	percentage
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

rs#	position	percentage
1526512	75385685	42.67
12138223	75387614	7.21

LCT chr2:136540000-136600000;

rs#	position	percentage
6707117	136571975	78.28
2734871	136586343	0.83

Supplementary material

Comparative analysis of *Lhx6* and *Lhx8* expression in the developing and adult brain of mouse and rat.

Tissue preparation, fixation and sectioning

All experiments were done in accordance with the European Communities Council Directive of 24 November 1986 (86/609/EEC).

a. Embryos: Time mated pregnant female mice (C57BL/6J) and rats (Sprague Dawley) were euthanized at various stages of pregnancy (the day of vaginal plug detection was considered as E 0.5) and embryos were dissected free of maternal tissues in phosphate-buffered saline (PBS, pH 7.4). Embryos were then fixed in 4% w/v paraformaldehyde (PFA) for 24 h at 4 ^{o}C, washed with PBS, cryoprotected in 30% w/v sucrose in 0.1 M PBS, embedded in Tissue Freezing Medium (Leica Microsystems), sectioned at 12 μm using a cryostat (Leica 1900UV) and transferred to superfrost plus (ROTH) slides. The sections were air dried for at least 30 min and stored at -80^{o}C until later use.

b. Adult brains: Animals were anesthetized and perfused transcardially with cold saline (4 °C) followed by 4 % w/v PFA in PBS. Brains were removed, postfixed in the same fixative overnight at 4 °C, washed thoroughly with PBS, cryoprotected in 30% w/v sucrose in 0.1 M PBS, embedded in Tissue Freezing Medium (Leica Microsystems), sectioned at 12 μm using a cryostat (Leica 1900UV) and transferred to superfrost plus (ROTH) slides. The sections were air dried for at least 30 min and stored at -80^{o}C until later use .

Riboprobes

Antisense RNA probes were synthesized by *in vitro* transcription with T3 RNA polymerase (Takara), according to manufacturer's instructions, using Digoxigenin-11-UTP (Roche). 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 cDNA) was used as template. For *in situ hybridization* experiments on sections of rat embryonic and adult brain the mouse probes were used.

In situ hybridization on sections

Cryostat sections were fixed in 4% w/v PFA in PBS for 10 min, washed three times in PBS and incubated in acetylating solution (1.3% v/v triethanolamine, 0.03 N HCl and 0.25% acetic anydrite) for 10 min. Sections were then washed in PBS, incubated in 1% v/v Triton-X-100 in PBS for 30 min and washed three times in PBS. Prehybridization was performed for 4-6 h in buffer H [50% v/v formamide, 5X SSC (0.75 M NaCl, 0.075 M sodium citrate), 5X Denhardt's (0.1% bovine serum albumin, 0.1% and 0.1% Polyvinylpyrrolidone), 250 μg/ml yeast RNA and 500 μg/ml salmon sperm DNA]. Hybridization was performed in humidified conditions for 16 h at 65 ^{o}C in buffer H with DIG-labeled probe added (400ng/ml). Following hybridization sections were sequentially washed in 5X SSC (5 min, 65^{o}C), 0.2X SSC (1 h, 65^{o}C), 0.2X SSC (5 min, room temperature). Then, they were incubated in AB buffer (100 mM Tris pH 7.5, 150 mM NaCl) for 5 min, and in blocking solution (10% v/v FCS in AB) for 1-2 h at room temperature. Antibody reaction was performed by incubating the sections for 16 h at 4^{o}C in a 1:5000 dilution of anti-DIG alkaline phosphatase-coupled Fab fragment (Roche) in 1% v/v FCS in AB solution. All subsequent steps were performed at room temperature. Sections were washed thoroughly in solution AB and equilibrated in alkaline phosphatase buffer (100 mM Tris–HCl pH: 9.5, 100 mM NaCl, 50 mM $MgCl_2$) for 5 min. Alkaline phosphatase activity was detected with 45 mg/ml 4-nitrobluetetrazolium chloride (NBT, Roche) and 35 mg/ml 5-bromo-4-chloro-3-indolyl-phosphate (BCIP, Roche) in alkaline phosphatase buffer. The reaction was stopped with PBS and the sections were mounted in Glycergel (Dako).

LHX6 and *LHX8* Genetic association study samples and methods

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 supplementary table 1. 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). DNA was extracted from whole blood using the Qiagen Puregene kit. For both samples, collection was approved by local Ethics Boards and informed consent was taken from all participating individuals or their parents.

In order to study variation in the *LHX6* and *LHX8* genes, tagging SNPs (tSNPs) were selected using the HapMap CEPH European population as reference and running the algorithms implemented in Tagger (De Bakker et al., 2005), with the r^2 threshold for tSNP selection set to 0.8 and allowing multi-marker tagging. The studied region extended before the start and after the end of each gene. A total of 10 SNPs were selected for the *LHX6* region (supplementary table 1) 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 (supplementary table 1) (mean r^2=0.988). Samples were genotyped using the fluorescently based KASPAR assay, as developed by KBiosciences.

Genotyping results for all SNPs conformed to Hardy-Weinberg equilibrium proportions, and Mendelian inheritance within families. The average genotyping success was greater than 99%. The linkage disequilibrium (LD) structure of each gene in the different studied populations was analyzed using Haploview (Barrett et al., 2005). In order to test for association of the studied SNPs with the TS phenotype, the Transmission Test for Linkage Disequilibrium was performed, as implemented in Haploview (Barrett et al., 2005). Both single SNPs as well as three-SNP haplotypes were analyzed using the same software. In order to correct for multiple comparisons, 1,000 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. The subset of the POPRES dataset that we analyzed comprises 1,200 individuals from 11 European populations and has been described in detail previously (Novembre et al., 2008). 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 (Lao et al., 2008, Novembre et al., 2008). PCA scores were computed as we have previously described (Paschou et al., 2007; Paschou et al., 2008), and compared to the distribution of PCA scores for all available SNPs (447,212 SNPs).

Supplementary methods - References

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

Cath DC, Hedderly T, Ludolph AG, Stern JS, Murphy T, Hartmann A, Czernecki V, Robertson MM, 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(4):155-171.

De Bakker, P.I.W., Yelensky, R., Peter, I., Gabriel, S.B., Daly, M.J., & Altshuler, D. (2005). Efficiency and power in genetic association studies. *Nature Genetics*, **37**, 1217-1223

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.

Hebebrand J, Klug B, Fimmers R, Seuchter SA, Wettke-Schäfer R, Deget F, Camps A, Lisch S, Hebebrand K, von Gontard A, Lehmkuhl G, Poustka F, Schmidt M, Baur MP, 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(5):519-530.

Lao, O., Lu, T.T., Nothnagel, M., Junge, O., Freitag-Wolf, S., Caliebe, A., Balascakova, M., Bertranpetit, J., Bindoff, L.A., Comas, D., Holmlund, G., Kouvatsi, A., Macek, M., Mollet, I., Parson, W., Palo, J., Ploski, R., Sajantila, A., Tagliabracci, A., Gether, U., Werge, T., Rivadeneira, F., Hofman, A., Uitterlinden, A.G., Gieger, C., Wichmann, H.E., Rüther, A., Schreiber, S., Becker, C., Nürnberg, P., Nelson, M.R., Krawczak, M., & Kayser, M. (2008). Correlation between genetic and geographic structure in Europe. *Current Biology*, **18**, 1241-1248.

Nelson, M.R., Bryc, K., King, K.S., Indap, A., Boyko, A.R., Novembre, J., Briley, L.P., Maruyama, Y., Waterworth, D.M., Waeber, G., Vollenweider, P., Oksenberg, J.R., Hauser, S.L., Stirnadel, H.A., Kooner, J.S., Chambers, J.C., Jones, B., Mooser, V., Bustamante, C.D., Roses, A.D., Burns, D.K., Ehm, M.G., & Lai, E.H. (2008). The Population Reference Sample, POPRES: a resource for population, disease, and pharmacological genetics research. *American Journal of Human Genetics*, **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.

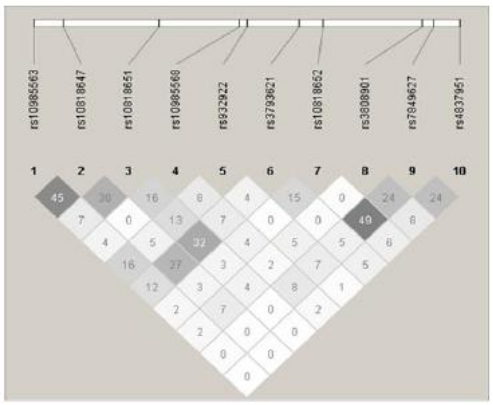
Paschou, P., Drineas, P., Lewis, J., Nievergelt, C.M., Nickerson, D.A., Smith, J.D., Ridker, P.M., Chasman, D.I., Krauss, R.M., & Ziv, E. (2008). Tracing sub-structure in the European American population with PCA-informative markers. *PLoS Genetics*, **4**,e1000114.

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 Genetics* **3**, e160.

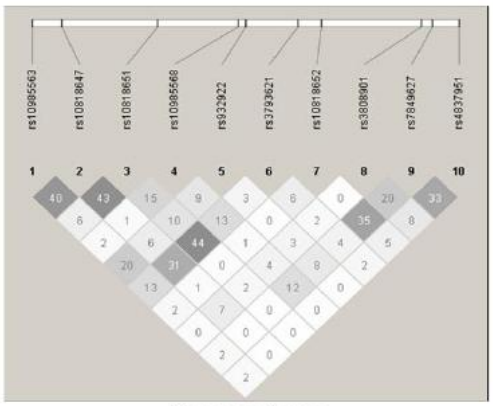
Schoenian, S., Konig, I., Oertel, W., Remschmidt, H., Ziegler, A., Hebebrand, J., & Bandmann, O. (2003). HLA-DRB genotyping in Gilles de la Tourette patients and their parents. *American Journal of Medical Genetics Part B*: *Neuropsychiatric Genetics*, **119B**, 60-64.

Supplementary Figure 1. Pairwise LD tests (r^2) in our German, Hungarian, Italian, and Polish samples, for all studied SNPs in the LHX6 and the LHX8 genes.

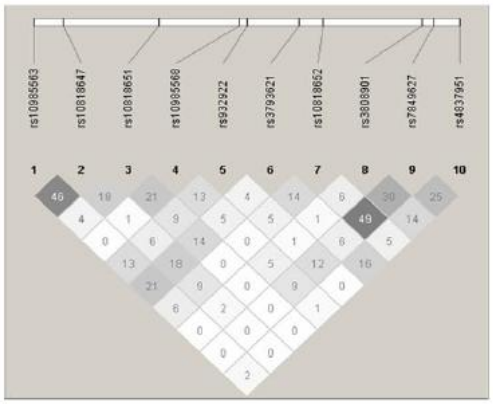
LHX6



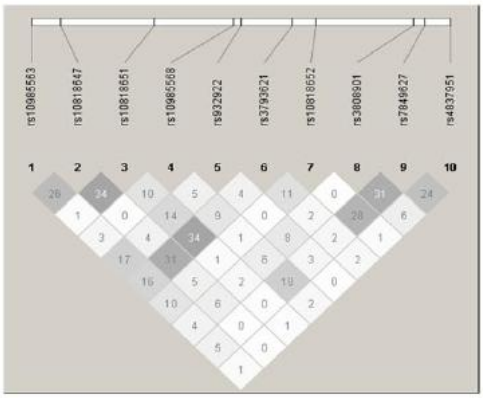
Germans



Hungarians



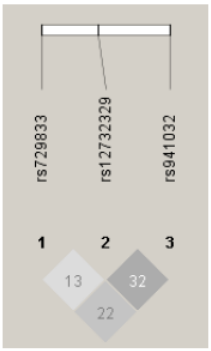
Polish



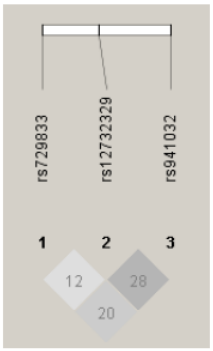
Italians

Fig

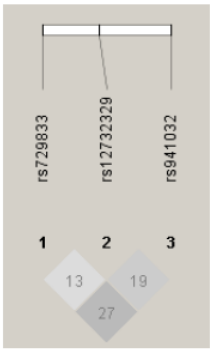
LHX8



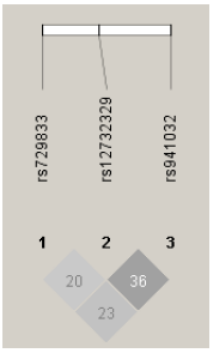
Germans



Hungarians



Polish



Italians

Supplementary Table 1. Phenotypic details of TSGeneSEE (Hungarian, Italian, and Polish), and German sample. Patient numbers are shown in parentheses.

	Hungarian (57)	Italian (43)	Polish (27)	German (95)
Male	81%	90%	80%	80%
Female	19%	10%	20%	20%
Average age at onset	5.46	5.3	6	6.3
Comorbid OCD	4.17%	10.21%	11.76%	33%
Comorbid ADHD	41.76%	68.96%	52.94%	Not available

Supplementary Table 2. 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.

tSNP/tSNP haplotype	SNPs captured
LHX6	
rs10818651	rs10760213,rs3793618,rs3793619,rs989798,rs10818651
rs10985563	rs10818649,rs10985564,rs10985563
rs10985568	rs12237553,rs10985568
rs7849627	rs7849627,rs3808899
rs10818647	rs10818647,rs3793617
rs3808901	rs3808901
rs932922	rs932922
rs3793621	rs3793621
rs10818652	rs10818652
rs4837951	rs4837951
rs3808901,rs7849627 : CA	rs3808900
rs932922,rs10818651 : GG	rs4837949
rs932922,rs10985568 : GT	rs1467737
LHX8	
rs12732329	rs1526511,rs12732329,rs12142460,rs12133614,rs6695423,rs7524020, rs1852348,rs12138223,rs12723906,rs1852349
rs941032	rs941032,rs957599
rs729833	rs729833

Supplementary table 3a. 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).
(T:U=Transmitted:Untransmitted, P=P value of TDT, Ov.=Overtransmitted allele)

	Hungarian (57 trios)				**Italian (43 trios)**				**Polish (27 trios)**				**German (95 trios)**			
rs#	**Ov.**	**T:U**	χ^2	**P**	**Ov.**	**T:U**	χ^2	**P**	**Ov.**	**T:U**	χ^2	**P**	**Ov.**	**T:U**	χ^2	**P**
LHX6																
rs10985563	C	31:28	0.15	0.70	T	23:17	0.90	0.34	T	16:15	0.03	0.86	C	47:45	0.04	0.83
rs10818647	-	21:21	0.00	1.00	T	15:8	2.13	0.14	C	15:10	1.00	0.32	C	39:35	0.22	0.64
rs10818651	G	33:31	0.06	0.80	A	25:18	1.14	0.29	G	15:11	0.62	0.43	G	53:39	2.13	0.14
rs10985568	T	22:17	0.64	0.42	C	14:12	0.15	0.69	T	15:8	2.13	0.14	C	30:24	0.67	0.41
rs932922	-	22:22	0.00	1.00	C	14:13	0.04	0.85	G	11:8	0.47	0.49	C	31:30	0.02	0.90
rs3793621	A	31:27	0.28	0.60	G	24:13	3.27	0.07	A	15:13	0.14	0.71	A	39:33	0.50	0.48
rs10818652	A	29:26	0.16	0.69	G	27:17	2.27	0.13	G	19:13	1.13	0.29	A	49:43	0.39	0.53
rs3808901	A	30:19	2.47	0.12	A	20:14	1.06	0.30	A	13:7	1.80	0.18	G	42:35	0.64	0.43
rs7849627	G	29:28	0.02	0.89	A	22:16	0.95	0.33	A	20:12	2.00	0.16	G	47:46	0.01	0.92
rs4837951	A	30:28	0.07	0.79	A	20:17	0.24	0.62	G	22:12	2.94	0.09	G	53:48	0.25	0.62
LHX8																
rs729833	T	20:14	1.06	0.30	C	17:08	3.24	0.07	T	11:10	0.05	0.83	C	35:33	0.06	0.81
rs12732329	G	27:21	0.75	0.39	G	23:19	0.38	0.54	G	10:06	1.00	0.32	G	33:32	0.02	0.90
rs941032	C	32:28	0.27	0.61	-	18:18	0.00	1.00	T	14:09	1.09	0.30	T	48:41	0.55	0.46

Supplementary table 3b. 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. (T:U=Transmitted:Untransmitted, P=P value of TDT, Ov.=Overtransmitted allele)

	TSGeneSEE (127 trios)				**TSGenesee + German (222 trios)**			
rs#	**Ov.**	**T:U**	**χ^2**	**P**	**Ov.**	**T:U**	**χ^2**	**P**
LHX6								
rs10985563	T	67:63	0.12	0.73	T	112:110	0.02	0.89
rs10818647	T	46:44	0.04	0.83	C	83:81	0.02	0.88
rs10818651	A	67:66	0.01	0.93	G	119:106	0.75	0.39
rs10985568	T	49:39	1.14	0.29	T	73:69	0.11	0.74
rs932922	G	46:44	0.04	0.83	G	76:75	0.01	0.94
rs3793621	G	64:59	0.20	0.65	A	98:97	0.01	0.94
rs10818652	G	72:59	1.29	0.26	G	115:108	0.22	0.64
rs3808901	**A**	**63:40**	**5.14**	**0.02**[1]	A	98:82	1.42	0.23
rs7849627	A	70:57	1.33	0.25	A	116:104	0.66	0.42
rs4837951	G	67:62	0.19	0.66	G	120:110	0.44	0.51
LHX8								
rs729833	C	41:39	0.05	0.82	C	76:72	0.11	0.74
rs12732329	G	60:46	1.85	0.17	G	93:78	1.32	0.25
rs941032	T	60:59	0.01	0.93	T	108:100	0.31	0.58

[1] Permutation P value=0.32 (1000 permutations performed)

Supplementary table 4. 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. (T:U=Transmitted:Untransmitted. P=P value of TDT).

	Hungarian (57 trios)				**Italian (43 trios)**				**Polish (27 trios)**				**German (95 trios)**			
Hapl.	**freq**	**T:U**	χ^2	**P**	**freq**	**T:U**	χ^2	**P**	**freq**	**T:U**	χ^2	**P**	**freq**	**T:U**	χ^2	**P**
GGA	**0.36**	**19 : 33**	3.77	**0.05**	0.28	17.2 : 18.0	0.02	0.90	0.26	12.9 : 11.9	0.04	0.84	0.33	47.0 : 42.7	0,21	0.65
AGG	0.27	26.9 : 23.9	0.18	0.67	0.27	15.8 : 19.7	0.42	0.52	0.43	11.0 : 18.8	2.05	0.15	0.36	44.9 : 42.9	0,05	0.83
GAA	**0.18**	**20.9 : 9.9**	3.94	**0.05**[1]	**0.22**	**20.8 : 8.7**	**5.02**	**0.02**[2]	0.19	11.1 : 5.0	2.36	0.12	0.16	24.8 : 32.3	0,99	0.32
AAA	0.12	15.1 : 13.1	0.14	0.71	0.13	6.2 : 12.3	2.06	0.15	0.05	2.9 : 3.0	0.01	0.95	0.08	14.2 : 13	0,05	0.82
GGG	0.07	6.1 : 8.1	0.28	0.60	0.09	6.2 : 7.6	0.14	0.70	0.07	4.0 : 4.3	0.01	0.92	0.07	11.1 : 12.1	0,04	0.84
AGA					0.01	1 : 1	0.00	0.99								

	TSGeneSEE (127 trios)				**TSGeneSEE + German (222 trios)**			
Hapl.	**freq**	**T:U**	χ^2	**P**	**freq**	**T:U**	χ^2	**P**
GGA	0.32	49.1 : 62.9	1.68	0.19	0.32	96.1 : 105.5	0.44	0.51
AGG	0.30	53.7 : 62.2	0.63	0.43	0.32	98.6 : 105.2	0.21	0.65
GAA	**0.19**	**52.8 : 23.5**	**11.28**	**8.0E-4**[3]	**0.18**	**77.5 : 55.9**	**3.53**	**0.06**[4]
AAA	0.11	24.2 : 28.5	0.35	0.55	0.10	38.5 : 41.5	0.12	0.73
GGG	0.08	16.3 : 20.1	0.39	0.53	0.08	27.4 : 32.1	0.38	0.54

[1] Permutation P value=0.20 (1000 permutations performed)
[2] Permutation P value=0.09 (1000 permutations performed)
[3] Permutation P value=0.006 (1000 permutations performed)
[4] Permutation P value=0.27 (1000 permutations performed)

Supplementary table 5. Allele frequencies of studied SNPs at LHX6 and LHX8 in all studied populations. Position refers to Build 36 of the genome.

LHX6

rs#	Position	Alleles	Hungarians	German	Italian	Polish
rs10985563	124005523	T:C	0.464	0.447	0.462	0.468
rs10818647	124007597	C:T	0.254	0.264	0.184	0.29
rs10818651	124014240	G:A	0.35	0.388	0.346	0.325
rs10985568	124019847	T:C	0.23	0.209	0.168	0.315
rs932922	124020369	C:G	0.234	0.234	0.226	0.213
rs3793621	124023946	A:G	0.335	0.302	0.32	0.31
rs10818652	124025606	G:A	0.384	0.44	0.423	0.492
rs3808901	124032492	G:A	0.292	0.244	0.346	0.23
rs7849627	124033326	A:G	0.343	0.423	0.383	0.5
rs4837951	124035103	G:A	0.333	0.386	0.322	0.405

LHX8

rs#	Position	Alleles	Germans	Italians	Polish	Hungarian
rs729833	75373733	C:T	0.263	0.193	0.305	0.224
rs12732329	75383231	A:G	0.275	0.449	0.235	0.296
rs941032	75395204	C:T	0.467	0.306	0.385	0.448