Published in final edited form as:

Am J Med Genet B Neuropsychiatr Genet. 2012 January ; 159B(1): 53–60. doi:10.1002/ajmg.b.32001.

Homeobox Genes in Obsessive-Compulsive Disorder

G Nestadt¹, Y Wang¹, MA Grados¹, MA Riddle¹, BD Greenberg⁴, JA Knowles⁸, AJ Fyer³, JT McCracken⁷, SL Rauch⁵, DL Murphy⁶, SA Rasmussen⁴, B Cullen¹, J Piacentini⁷, D Geller⁵, D Pauls⁵, OJ Bienvenu¹, Y Chen², KY Liang², FS Goes¹, B Maher¹³, AE Pulver¹, Y Y Shugart^{11,12}, D Valle⁹, JF Samuels¹, and YC Chang¹⁰

¹Department of Psychiatry and Behavioral Sciences, School of Medicine, Johns Hopkins University

²Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University

³College of Physicians and Surgeons at Columbia University

⁴Department of Psychiatry and Human Behavior, Brown Medical School, Butler Hospital

⁵Departments of Psychiatry and Psychiatric and Neurodevelopmental Genetics Unit, Massachusetts General Hospital and Harvard Medical School

⁶Laboratory of Clinical Science, NIMH, NIH, Bethesda

⁷Department of Psychiatry and Biobehavioral Sciences, School of Medicine, University of California, Los Angeles

⁸Department of Psychiatry, Keck Medical School, University of Southern California

⁹Department of Pediatrics, School of Medicine, Johns Hopkins University

¹⁰Department of Medicine, University of Maryland School of Medicine

¹¹Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD

¹²Genomic Research Branch, Division of Neuroscience and Basic Behavioral Science, National Institute of Mental Health, National Institutes of Health, Bethesda, MD

¹³Department of Mental Health, Bloomberg School of Public Health, Johns Hopkins University

Abstract

Background—Despite evidence that obsessive-compulsive disorder (OCD) is a familial neuropsychiatric condition, progress aimed at identifying genetic determinants of the disorder has been slow. The OCD Collaborative Genetic Study (OCGS) has identified several OCD susceptibility loci through linkage analysis.

Methods—In this study we investigate two regions on chromosomes 15q and 1q by first refining the linkage region using additional STRP markers. We then performed association analysis on SNP genotyped (markers placed every 2–4kb) in the linkage regions in the OCGS sample of 376 rigorously phenotyped affected families.

Results—Three SNPs are most strongly associated with OCD: rs11854486 ($p=0.00005$ [0.046 after adjustment for multiple tests]; Genetic Relative Risk (GRR)=11.1 homozygous and 1.6

Address reprint requests and correspondence to: Gerald Nestadt, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, 600 N. Wolfe Street, Meyer 109, Baltimore, MD 21287, Phone: 410-614-4942. Fax: 410-614-8137, gnestadt@jhmi.edu.

Financial disclosures & Conflict of Interest:

The authors report no disclosures.

heterozygous) and rs4625687 ($p=0.00007$ (after adjustment = 0.06); $GRR=2.4$) on 15q; and rs4387163 ($p=0.0002$ (after adjustment = 0.08); $GRR=1.97$) on 1q. The first SNP is adjacent to *NANOGP8*, the second SNP is in *MEIS2*, and the third is 150kb between *PBX1* and *LMX1A*.

Conclusions—All the genes implicated by association signals are homeobox genes and are intimately involved in neurodevelopment. *PBX1* and *MEIS2* exert their effects by the formation of a heterodimeric complex which is involved in development of the striatum, a brain region involved in the pathophysiology of OCD. *NANOGP8* is a retrogene of *NANOG*, a homeobox transcription factor known to be involved in regulation of neuronal development. These findings need replication; but support the hypothesis that genes involved in striatal development are implicated in the pathogenesis of OCD.

Keywords

OCD; *PBX1*; *MEIS2*; *LMX1A*; *NANOG*; genetic association

Introduction

Obsessive-compulsive disorder (OCD) is a psychiatric condition first described over a century ago (Janet, 1903). The pathognomonic features of the disorder are persistent, intrusive, senseless thoughts and impulses (obsessions) and repetitive, intentional behaviors (compulsions). Patients with the disorder recognize that their thoughts and behaviors are excessive and unreasonable, and they struggle to resist them. The lifetime prevalence of OCD is estimated to be 1-3%, based on population-based surveys conducted in many communities nationally and internationally, (Karno and Golding 1991; Weissman *et al* 1994). Conventional behavioral and medication treatments for this chronic, disabling disorder, while helpful, generally bring only partial relief. No cure exists. While neurocircuitry including the striatum and its key functional connections has been consistently implicated in OCD, application of that knowledge to treatment remains very limited (Greenberg *et al* 2010). The discovery of genes is crucial for elucidation of pathogenic mechanisms and for developing new rational treatments.

Several family studies of OCD indicate a genetic component for its origin (Black *et al* 1992; do Rosario-Campos *et al* 2005; Lenane *et al* 1990; Pauls *et al* 1995; Nestadt *et al* 2000). The Johns Hopkins OCD family study (Nestadt *et al* 2000) reported that the prevalence of OCD in the first-degree relatives of probands was 11.7%, compared to 2.7% in the relatives of controls; also, the prevalence of OCD in the siblings of early onset probands was 17.9% ($\lambda_{sib}=7.8$). There are several reports of monozygotic twins concordant for obsessive-compulsive symptoms (Woodruff and Pitts, 1964; Inyoue, 1965; Carey and Gottesman II, 1981).

The OCD Collaborative Genetics Study (OCGS) conducted by our group, was the first large-scale genome-wide linkage scan searching for genetic loci that predispose individuals to OCD (Shugart *et al* 2006); this study included 219 families. The strongest signal was observed at chromosome 3q27/28, with a LOD score of 2.67 (P -value=0.0002). Regions on chromosomes 1, 6, 7, and 15 also showed evidence of linkage, with P -values <0.01 (Shugart *et al* 2006).

The aim of the current study is to follow up the chromosomal regions identified in our linkage study using genetic association methods. This report describes the most convincing results in regions on chromosomes 15q and 1q, and centers on four genes (*Meis* homeobox 2 (*MEIS2*), pre-B-cell leukemia homeobox 1 (*PBX1*), LIM homeobox transcription factor 1, alpha (*LMX1A*), and *Nanog* homeobox pseudogene 8 (*NANOGP8*) that not only have the

most significant findings, but suggest a novel and testable hypothesis about the origin of OCD.

Materials and Methods

Sample

The OCGS, which commenced in 2001, is an ongoing, NIMH-funded collaboration among investigators at six sites in the US (Brown University, Columbia University, Johns Hopkins University (JHU), Massachusetts General Hospital, University of California at Los Angeles, and the National Institute of Mental Health). The methods are described in detail elsewhere (Samuels *et al* 2006) and summarized below.

The OCGS ascertained families with OCD-affected sibling pairs, and extended these when possible through affected first- and second-degree relatives; in addition, two sites also collected other pedigrees with multiple-affected relatives when these were available. Subjects were recruited into the study from outpatient and inpatient clinics, referrals from clinicians in the community, web sites, media advertisements, self-help groups, and annual conventions of the Obsessive Compulsive Disorder Foundation. To be considered affected, a subject had to meet Diagnostic and Statistical Manual (DSM)-IV OCD diagnostic criteria at any time in his/her life (APA, 2000). Probands were included if, in addition to meeting DSM-IV criteria, their first onset of obsessions and/or compulsions occurred before 18 years of age. Probands with schizophrenia, severe mental retardation, Tourette disorder, or secondary OCD (OCD occurring exclusively in the context of depression) were excluded. Subjects had to be at least 7 years old to participate in the study. Written, informed consent (or assent, for children) to study procedures was obtained before the clinical interview. The protocol was approved by the institutional review board at each site.

Diagnostic assessments were conducted by psychiatrists or PhD-level psychologists experienced with clinical evaluations using the OCGS assessment package, modified and developed for the study, as a semi-structured format for the evaluation of psychopathology. The OCD section was adapted from the Schedule for Affective Disorders and Schizophrenia-Lifetime Anxiety Version (SADS-LA-R; Mannuzza *et al* 1986) and included detailed screening questions; the Yale Brown Obsessive Compulsive Scale (YBOCS; Goodman *et al* 1989) and symptom checklist, and additional questions on onset, course, and treatment of symptoms. A similar section was developed for assessing tics, Tourette disorder, and other tic disorders. The Structured Clinical Interview for DSM-IV (SCID; Spitzer *et al* 1992) was used for assessing other major Axis I diagnoses and a semi-structured assessment protocol was used for additional diagnoses of interest. In adults, items from the Structured Instrument for the Diagnosis of DSM-IV Personality Disorders (SIDP; Pföhl *et al* 1989) were used to assess the presence of personality disorders. The Family Informant Schedule and Criteria (Mannuzza *et al* 1985) was used to obtain additional information about each subject from a knowledgeable informant. For subjects who had received psychiatric treatment, consent was obtained to review relevant medical records and to contact treatment providers, if such information was deemed useful for making diagnoses. Examiners completed a narrative formulation for each case.

The JHU OCD diagnostic assignment checklist was used to collate all the clinical information from a variety of sources (the semi-structured direct interview, case formulation, informant interview, and medical records). The checklist presents logical algorithms with specified rules, allowing assignment of definite, probable, absent, or unknown for each disorder. All psychiatric diagnoses were made according to strict DSM-IV criteria (APA, 2000). At each site, each case was reviewed independently by two expert diagnosticians who reviewed all case materials. Final diagnoses were assigned by

diagnosticians at Johns Hopkins University. If all criteria required for having the disorder were met, then a 'definite' diagnosis was given. If any required criterion was clearly not met, then the diagnosis was considered 'absent'. If it appeared likely that the subject had the diagnosis, but the diagnosticians could not be certain of a given criterion (required for definite diagnosis), then the diagnosis was made at the 'probable' level. If the diagnosticians could not be sure of the presence or absence of a given diagnosis, then that diagnosis was recorded as 'unknown'.

Blood samples were collected from all affected probands, their parents, and their affected relatives. If available, blood samples from unaffected relatives were also collected to help determine phase used for computing identity by descent (IBD) sharing probabilities.

Molecular genotyping

Different approaches were taken and different genotyping laboratories used for the two chromosomal regions. Specifically, the region on chromosome 15q was studied in a single fine-mapping phase, whereas, the region on chromosomes 1q was studied in three phases, such that the final phase represented the result of sequential testing. Furthermore, the sample size increased between the initial linkage study and the association studies.

Fine mapping of the regions on chromosome 15q and 1q, using 14 short tandem repeat polymorphic markers (STRP) markers across 12.8Mb region of 15q, and 21 STRP markers spaced every 1.2Mb across a 25.8Mb region on 1q, was conducted at deCODE genetics service laboratory in Iceland. Samples were analyzed by capillary electrophoresis utilizing ABI DNA analyzers. Alleles were called by proprietary deCODE Allele Caller (DAC) software. The genotype success rate was 96.1%. A total of 945 subjects in 214 families were genotyped. Compared to the genome-wide linkage scan, five families were excluded from the linkage fine mapping due to insufficient amount of DNA for genotyping.

Two thousand, two hundred-and-fifty-five single nucleotide polymorphisms (SNPs) were genotyped in a 10.7 Mb region with average spacing of 4.8kb in the 15q linkage region at deCODE, in Iceland. Fine mapping of the refined linkage region on 1q, using 860 SNP markers spaced every 6.3kb across a 5.4 Mb region (1q23-24), was conducted at the Illumina service laboratory. Public databases (dbSNP Build 126) were used to select the (SNPs) genotyped by Illumina with their BeadArray system with the GoldenGate assay (Steemers and Gunderson, 2005). We gave preference to validated SNPs with minor allele frequency >0.05 and high anticipated Illumina assay success rate (>80%). The genotype success rate was 99.46%.

We included the families described above and an additional 162 families (631 individuals) recruited and evaluated after the initial study in both the 15q and 1q association studies; i.e. a total of 376 families, including 1,576 individuals. Among these new families, 73 had at least one OCD-affected sibling pair or other relative pair; 89 were trio families. 999 participants in this study were diagnosed with probable or definite OCD. 622 (62.3%) of these participants were female and 208 (20.8%) were younger than eighteen years; the average age of the cases was 35 years and their means score on the Y-BOCS (lifetime worst episode) was 24.6(SD 7.2).

An additional 160 SNP markers were genotyped, at the deCODE service laboratory in Iceland using the same methods, in regions identified based on the significance of the signal in either the prior fine mapping phase, or that were in a gene or a conserved region according to the University of California Santa Cruz (UCSC) database.

Analysis

Multipoint allele-sharing methods were used to assess evidence for linkage by testing the null hypothesis given $\theta = 0$, where θ is the extent deviated from expected IBD sharing (Kong and Cox, 1997). Using Merlin version 0.10.2 (Abecasis *et al* 2002), we computed the Kong and Cox logarithm of odds (LOD)_{all} (KAC_{all}) statistic.

We used the Family-Based Association Test (FBAT; Horvath *et al* 2001) to test each SNP separately for association with OCD under the additive model. Then, we used the Haplotype-Based Association Test (HBAT; Horvath *et al* 2001) to test haplotypes of SNPs for association with OCD. We chose to use predefined blocks to specify the haplotypes for testing. Finally, we used the *gtrr* function of the GenAssoc STATA (Clayton, 2007) to obtain the estimated odds ratio as well as calculate global likelihood ratio test P-values. Empirical P-values for each significant association signal were also calculated with FBAT using 100,000 permutations.

Two-point P-values from the FBAT were corrected for multiple-marker testing in the presence of linkage disequilibrium (LD). The P-values Adjusted for Correlated Tests (PACT) approach was employed to adjust for multiple testing (Conneely and Boehnke, 2007).

Estimates of the interaction between identified markers were accomplished by performing conditional logistic regression. The transmission disequilibrium test (TDT) test is the McNemar's chi-squared test, and can be embedded in the more general conditional logistic regression framework. With two or more markers, the "case" is the transmission of a parental haplotype instead of an allele, and the "control" is non-transmission of a parental haplotype. Therefore, with L markers considered simultaneously, each case-parent trio contributes two entries in a $2^L \times 2^L$ matrix. The test of association tests the asymmetry of such a matrix. The test can be performed in conditional logistic regression with L covariates of exposure and possible interaction terms.

Results

Chromosome 15

The linkage signal at 15q increased only marginally after genotyping additional STRP markers in the original 214 families; five families were not genotyped. The KAC_{all} increased from 1.32 ($p = 0.007$) in the original genome wide linkage scan to 1.43 ($p = 0.005$) at marker DG15S14.

SNPs were genotyped every 4.3kb across the 1-LOD linkage region. As shown in Figure 1, SNP rs4625687, in the gene *MEIS2*, was associated with OCD ($p = 0.000079$; empirical $p = 0.00012$; adjusted for multiple tests, $p = 0.06$). The GRR at this SNP was 2.4 (95% C.I.: 1.5, 3.8). The SNP rs11854468, located approximately 64 kb downstream from a *NANOG* homeobox transcription factor (OMIM 607937) pseudogene, *NANOGP8*, was significantly associated with OCD ($p = 0.00005$; adjusted for multiple tests $p = 0.05$; GRR = 11.1, 95% C.I.: 100.0, 1.4 (GG) and GRR = 1.6, 95% C.I.: 2.3, 1.1 (GA). Several SNPs in or near *formin1* (*FMN1*) were significantly associated with OCD; the strongest association was at rs345804 ($p = 0.00049$; GRR = 1.6, 95% C.I.: 1.2, 2.0) in an expressed sequence tag (AI040235) about 5.7kb upstream of *FMN1*. One associated SNP rs2306277 in *exon 1* of *FMN1* is reported in Polyphen as a missense mutation with "probable protein damage." No interesting haplotype was identified in the 15q region.

Chromosome 1q

The linkage signal in this region increased after genotyping additional STRP markers in the original sample of families (214 families). The KAC_{all} increased from 1.42 ($p = 0.005$) in the original genome wide linkage scan to 2.06 ($p = 0.001$) at marker D1S2762 (figure 2).

SNPs were genotyped, on average every 6.3kb, across the 5.4 Mb region of the 1-LOD interval of the linkage peak. As shown in Table 1 and figure 3, five SNPs between PBX1 and LMX1A were associated with OCD. rs4387163 gave the most significant signal ($p=0.0002$; empirical $p=0.0002$; adjusted for multiple tests $p=0.08$) with a genetic relative risk (GRR) of 1.97 (95% C.I.:1.33, 2.92) for the homozygote (G/G) and 1.58 (95% C.I.:1.19, 2.10) for the heterozygote (G/A). Haploview (confidence interval method) identified three contiguous haplotype blocks in the region of this signal, as described in Figure 3. The first two blocks contained SNPs with signals of $p<0.01$. The most common haplotype in Block 1 (frequency = 0.55) was associated with OCD ($p = 0.0012$); the most common haplotype in Block 2 associated with OCD had a frequency = 0.55 ($p = 0.0004$); and an extended haplotype across both blocks had a frequency = 0.52 and was also associated with OCD ($P = 0.0004$). A fourth haplotype based on “pair-wise tagging” with a threshold of $r^2>0.5$, computed in Haploview (most common haplotype frequency = 0.526), provided similar results ($p=0.0002$).

Statistical interaction between significant SNP markers

Evidence for statistical interaction was investigated using conditional logistic regression. The best markers on 1q are rs4387163 and rs986362 and those on 15q are rs4625687, rs4244559, and rs1464284. No evidence for interaction was detected in tests including each of the markers on one chromosome with those on the other chromosome. It should be noted that the power to detect evidence of interaction is limited; a much larger sample size than this one would be required to determine the presence of statistical interaction.

Clinical characterization

We explored the clinical data for association between clinical characteristics and the associated allele of SNP rs4625687 in *MEIS2*. Three clinical characteristics were significantly associated with this allele; the presence of any tic disorder ($p < 0.05$); body dysmorphic disorder ($p<0.05$); and the symmetry/ordering symptom factor, derived by factor analysis (Pinto *et al* 2008), from the overall range of OCD symptoms ($p=0.02$).

Discussion

This study represents a classical approach for the identification of genes underlying common and complex disorders, such as the disabling neuropsychiatric condition, OCD. The approach entailed a genome-wide linkage scan in a large sample of rigorously assessed families, each with multiple members diagnosed with DSM-IV OCD (Shugart *et al* 2006). Chromosomal regions with the strongest linkage signals were then investigated further; first with additional linkage fine-mapping using STRP markers, and then with staged, association fine-mapping using SNP markers.

The original linkage study suggested that six regions were most promising for further investigation. After additional linkage and association fine-mapping, two regions on chromosomes 15q and 1q, and, more specifically, two genes in each region, were associated with OCD. These genes are *MEIS2* and *NANOGP8* (15q) and *PBX1* and *LMX1A* (1q).

The strongest signals on chromosome 15q are at rs11854468 ($p=0.000054$), located approximately 60 kb from *NANOGP8*, and at rs4625687 ($p=0.00007$), located in the intron

of *MEIS2*. There is another strong signal at rs11858145 ($p=0.0007$) located in a coding region of *FMN1*. The best signal on chromosome 1q is at rs4387163 ($p=0.0002$) which is located at 1q23.3. The GRR at this locus is 1.9. This SNP is located 150kb equidistant from both *PBX1* and *LMX1A*. SNP rs11854468 on 15q remained significant after adjustment for multiple testing ($p<0.05$). Although statistically this is the most robust signal, rs4625687 in the *MEIS2* gene (p after adjustment = 0.065) remains an interesting signal, because of its location and the signals at surrounding SNPs.

An outstanding feature of these results is that *PBX1*, *MEIS2*, *LMX1A*, and the *NANOG*-related *NANOGP8* gene are all transcription factors, or closely related to transcription factors, important in neurodevelopment. While we recognize that there are many transcription factors involved in neurodevelopment, and even more that could plausibly be implicated in some hypothetical relationship to the origin of OCD, we assert that the probability is exceedingly small that genes with these functional properties would be identified. Specifically, of the thousands of genes across the genome there are approximately 233 Homeobox genes identified; therefore identifying these four genes by chance is highly improbable.

MEIS2, *PBX1*, *LMX1A*, and *NANOG* code for homeobox proteins: the former two are members of the three amino acid loop extension (TALE) superclass, whereas *LMX1A* is a member of the LIM superclass. All are cofactors for transcription regulation and have important roles in brain development. *MEIS2* and *PBX1* are intimately involved as cofactors with homeobox (*HOX*) and myogenic differentiation (*MyoD*) genes (Berkas *et al* 2004). It has been shown that *MEIS2* and *PBX1* form a heterodimeric complex which acts together with *HOX* genes to activate transcription (Sagerström, 2004) and *MEIS2* has been found to be the preferred partner of *PBX1* (Chang, *et al* 1997). *NANOG* also binds to *PBX1* which regulate its function (Chan *et al.*, 2009). *NANOGP8* is a retrogene (Zhang, *et al.*, 2006) which belongs to a family of 11 *NANOG* pseudogenes. The *NANOGP8* gene in 15q14 is unique among *NANOG* pseudogenes in that it has an intact open reading frame and is actively transcribed (Jeter *et al* 2009). It is the only *NANOG*-related gene that is uniquely expressed in humans compared to chimpanzees (Fairbanks *et al* 2006). *FMN1* is a regulator of another important transcription factor in neurodevelopment, glioma-associated oncogene family zinc finger 3 (*GLI3*; Zeller *et al* 1999; Zuniga & Zeller, 1999).

Given the relationship between *MEIS2* and *PBX1* and the *HOX* class of transcription factors, and in particular the interaction of *PBX1* and *HOXB8* (Knoepfler *et al* 2001; Neuteboom *et al* 1995; Sanchez *et al* 1997), it is fascinating that Greer and Capecchi (2002) reported that *HOXB8* knockout mice exhibit compulsive hair pulling behavior. This group recently reported that this behavior is attributable to bone marrow-derived microglia and hypothesize an immunologic basis to the behavior (Chen *et al* 2010). It should be noted that both *MEIS2* and *PBX1* are also intimately involved in hematopoiesis (Ficara *et al* 2008; Wilson *et al* 2009).

MEIS2 and *PBX1* have been shown to be expressed in the developing rat striatum (Cecconi *et al* 1997; Toresson *et al* 2009), as has *NANOG* (Molero *et al* 2009). Takahashi *et al* (2008) suggest that *MEIS2* and *PBX1* might play a “pivotal role in establishing the neurons of the islands of Calleja, within the ventral striatum, as well as striosomal neurons in the striatum” (Takahashi *et al* 2008). These transcription factors have also been reported to be associated with the *ephA8* regulatory sequence in mesencephalon development (Sungbo *et al* 2007). *MEIS2* (with *ISLET1*) has also been shown to activate transcription through *I56ii* (a conserved regulatory element of *DLX* genes) during mid-gestation. Additionally, both *PBX1* and *MEIS2* have been shown to have a functional link to *SOX3*, which is involved in specifying neural fate (Mojsin and Stevanovic, 2009). Determining the pathways through

which these and other transcription factors interact in the development of the relevant brain regions is in progress (Long *et al* 2009a; Long *et al* 2009b), and will ultimately enhance our understanding of brain development and possibly the origin of neuropsychiatric conditions.

Either of these two genes (i.e. *MEIS2* and *PBX1*) alone could be considered interesting candidate genes involved in the pathogenesis of OCD (Takahashi *et al* 2008). However, identifying signals in the region of both genes, on different chromosomes, makes these findings more compelling; both are involved in brain development in a brain region, the striatum, that is the primary focus of interest in OCD; and they exert their developmental effect in tandem, forming a heterodimeric complex for this purpose. The genetic effects of *FMN1* and *NANOGP8* also plausibly impinge on neurodevelopment, albeit in different ways. While *FMN1* directly supports cell structure via actin/microtubules and regulates a neurodevelopment gene, *GLI3*, *NANOGP8* is related to *NANOG* which maintains stem cell pluripotency, and therefore conceivably affects neurogenesis generally, and early stage serotonin neurons specifically (Bethea *et al* 2009).

An obvious step to investigate a possible inter-relationship between these two genes in the origin of OCD is to determine whether there is a statistical interaction between them in their association with OCD. We did not find evidence for statistical interaction in this sample using the method described above; however, definitive identification of interaction would require a far larger sample of families than were available in this study.

LMX1A is adjacent to *PBX1* on 1q and both are equidistant to the association signal in that region. *LMX1A* acts in a pathway that involves the development of dopamine neurons in the midbrain. It has also been reported to be involved in postnatal brain development and the maintenance of neuronal function (Andersson *et al* 2006). This gene is therefore also an interesting candidate in OCD. Dopaminergic function has been hypothesized to be involved in the etiology (and treatment) of OCD (Graybiel and Rauch, 2000).

FMN1 is a member of the formin family of proteins that remodel the actin and microtubule cytoskeletons (Goode and Eck, 2007); however, it also directly regulates *GLI3*, a key gene in development, including cortical morphology (Friedrichs *et al* 2008; Theil, 2005). Thus, a plausible explanation for the involvement of *FMN1* in OCD may be related to the involvement of microtubule and actin cytoskeleton in neural migration (Kerjan and Gleeson, 2007) or in the gene network that regulates brain morphology (Yu *et al.*, 2009; Quinn *et al.*, 2009).

The likelihood of genetic heterogeneity is a fundamental concern in conducting studies such as this one. We investigated whether any clinical characteristic of the cases were associated with SNP rs4625687 in *MEIS2*. Only this SNP was investigated, out of several potential SNPs, as an exploratory exercise. The presence of body dysmorphic disorder, tics, and symmetry/ordering OCD symptoms were significantly more common. It should be noted that probands with Tourette Disorder were excluded. Nevertheless, these three clinical features frequently co-occur in OCD cases who exhibit tics; this and suggests that a 'tic-form' of OCD may be a relevant sub-phenotype underlying these findings.

Patients with OCD typically score extremely high on 'Neuroticism', a well-characterized personality trait in the psychological literature (Samuels *et al* 2000). It is therefore interesting that a WGA study of 'neuroticism' reported a single nucleotide polymorphism (SNP) in *PBX1* as one of the ten best signals (7.18×10^{-6}), and this SNP was most significant for the "anxiety" facet of neuroticism (van den Oord *et al* 2008). The hypothesis that 'neuroticism' and OCD share a common etiology would be borne out if the relationship between *PBX1* and both 'neuroticism' and OCD were replicated.

Limitations

The original linkage findings, and the subsequent fine mapping linkage results, do not meet statistical significance, as defined by Lander and Kruglyak (1995). Nevertheless, after fine-mapping the LOD score was 2.06 ($p=0.001$) on chromosome 1 and 1.43 ($p=0.0035$) on chromosome 15. There could be several reasons that these signals did not meet that threshold, including power and genetic heterogeneity. Furthermore, because the linkage signal does not reach the conventional level of significance, it could be argued that the significance of the association signal ought to be adjusted for across the entire genome. Nevertheless, this is the largest OCD sample studied and therefore these are the most likely known genomic regions to harbor OCD susceptibility genes. Therefore, it seems appropriate to both investigate these chromosomal regions and to test for significance within the linkage region. This limitation emphasizes the need for replication of these findings.

The association signals do not specifically identify which of the genes may be involved in the genesis of OCD. For instance the strongest signal on chromosome 1 is between *LMX1A* and *PBX1*. We therefore discuss the potential relationship of both genes to the development of OCD.

In conclusion, the genes with the strongest signals in this systematic genome-wide linkage and association study are all Homeobox genes, with transcription factor activity. They are important in development, and possibly to ongoing neurogenesis in brain regions strongly implicated in the pathophysiology of OCD. Thus, these findings provide important and novel evidence that OCD is a neurodevelopmental condition. The hypothesis that emerges is that developmental influences on the neuronal architecture in the striatum results in a vulnerability to OCD. An alternate hypothesis is that individuals with OCD have a deficit in neurogenesis that prevents/ limits appropriate learning mechanisms, resulting in the pathognomonic symptoms of OCD.

Acknowledgments

The authors thank the many families participating in the study; David Housman, MD, Kathleen Merikangas, PhD, Alec Wilson, PhD, and Dani Fallin PhD, for consultation; and Dalin Li PhD for analytic support; and clinicians and coordinators at each site. Yin Yao Shugart worked on this project as approved outside activities. The views expressed in this manuscript do not necessarily represent the views of the NIMH, NIH, HHS, or the United States Government.

This work was funded by the National Institute of Health grants R01-MH-50214, and NIH/NCRR/OPD-GCRC RR00052.

References

- Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin - Rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet.* 2002; 30:97–101. [PubMed: 11731797]
- Andersson E, Tryggvason U, Deng U, Friling S, Alekseenko Z, Robert B, et al. Identification of intrinsic determinants of midbrain dopamine neurons. *Cell.* 2006; 124:393–405. [PubMed: 16439212]
- APA. Diagnostic and Statistical Manual of Mental Disorders. APA; Washington, DC: 2000.
- Berkes CA, Bergstrom DA, Penn BH, Seaver KJ, Knoepfler PS, Tapscott SJ. Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. *Mol. Cell.* 2004; 14:465–477. [PubMed: 15149596]
- Bethea CL, Reddy AP, Pedersen D, Tokuyama Y. Expression profile of differentiating serotonin neurons derived from rhesus embryonic stem cells and comparison to adult serotonin neurons. *Gene Expr Patterns.* 2009; 9(2):94–108. [PubMed: 18996226]

- Black DW, Noyes R Jr, Goldstein RB, Blum N. A family study of obsessive-compulsive disorder. *Arch. Gen. Psychiatry.* 1992; 49:362–368. [PubMed: 1586271]
- Carey, G.; Gottesman, II. Twin and family studies of anxiety, phobic, and obsessive disorders *Anxiety: New research and changing concepts.* In: Klein, DF.; Rabkin, JG., editors. *Anxiety: New research and changing concepts.* Raven Press; New York: 1981. p. 117-136.
- Cecconi F, Proetzel G, Alvarez-Bolado G, Jay D, Gruss P. Expression of Meis2, a Knotted-related murine homeobox gene, indicates a role in the differentiation of the forebrain and the somitic mesoderm. *Dev Dyn.* 1997; 210(2):184–90. [PubMed: 9337138]
- Chan KK, Zhang J, Chia NY, Chan YS, Sim HS, Tan KS, Oh SK, Ng H, Choo AB. KLF4 and PBX1 Directly Regulate NANOG Expression in Human Embryonic Stem Cells. *Stem Cells.* 2009; 27:2114–2125. [PubMed: 19522013]
- Chang CP, Jacobs Y, Nakamura T, Jenkins NA, Copeland NG, Cleary ML. Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pbx proteins *Mol. Cell. Biol.* 1997; 17:5679–5687.
- Chen SK, Tvrdik P, Peden E, Cho S, Wu S, Spangrude G, Capecchi MR. Hematopoietic origin of pathological grooming in Hoxb8 mutant mice. *Cell.* 2010; 141(5):775–85. [PubMed: 20510925]
- Clayton D. David Clayton's Web site. 2007 <http://www.gene.cimr.cam.ac.uk/clayton/software/stata/>.
- Conneely KN, Boehnke M. So Many Correlated Tests, So Little Time! Rapid Adjustment of P Values for Multiple Correlated Tests. *Am J Hum Genet.* 2007; 81(6):1158–1168. [PubMed: 17966093]
- do Rosario-Campos MC, Leckman JF, Curi M, Quatrano S, Katsovitch L, Miguel EC, Pauls DL. A family study of early-onset obsessive-compulsive disorder. *Am J Med Genet B Neuropsychiatr Genet.* 2005; 136B(1):92–7. [PubMed: 15892140]
- Fairbanks DJ, Peter J, Maughan PJ. Evolution of the NANOG pseudogene family in the human and chimpanzee genomes *BMC Evolutionary Biology.* 2006; 6:12.
- Ficara F, Murphy MJ, Lin M, Cleary ML. Pbx1 regulates self-renewal of long-term hematopoietic stem cells by maintaining their quiescence. *Cell Stem Cell.* 2008; 2(5):484–96. [PubMed: 18462698]
- Friedrichs M, Larralde O, Skutella T, Theil T. Lamination of the cerebral cortex is disturbed in Gli3 mutant mice. *Dev Biol.* 2008; 318(1):203–14. [PubMed: 18448089]
- Goode BL, Eck JM. Mechanism and Function of Formins in the Control of Actin Assembly *Annu. Rev. Biochem.* 2007; 76:593–627.
- Goodman WK, Price LH, Rasmussen SA, Mazure C, Fleischmann RL, Hill CL, et al. The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. *Arch Gen Psychiatry.* 1989; 46:1006–1011. [PubMed: 2684084]
- Graybiel AM, Rauch SL. Toward a neurobiology of obsessive-compulsive disorder. *Neuron.* 2000; 28(2):343–7. [PubMed: 11144344]
- Greenberg BD, Rauch SL, Haber SN. Invasive circuitry-based neurotherapeutics: stereotactic ablation and deep brain stimulation for OCD. *Neuropsychopharmacology.* 2010; 35(1):317–36. [PubMed: 19759530]
- Greer JM, Capecchi MR. Hoxb8 Is Required for Normal Grooming Behavior in Mice. *Neuron.* 2002; 33(1):23–34. [PubMed: 11779477]
- Horvath S, Xu X, Laird NM. The family based association test method: Strategies for studying general genotype - Phenotype associations. *Eur J Hum Genet.* 2001; 9:301–306. [PubMed: 11313775]
- Inyoue E. Similar and dissimilar manifestations of obsessive-compulsive neurosis in monozygotic twins. *Am J Psych.* 1965; 121:1171–1175.
- Janet, P. *Les Obsessions et Psychasthenie.* Bailliere; Paris: 1903.
- Jeter CR, Badeaux M, Choy G, Chandra D, Patrawala L, Liu C, et al. Functional evidence that the self-renewal gene NANOG regulates human tumor development. *Stem Cells.* 2009; 27(5):993–1005. [PubMed: 19415763]
- Karno, M.; Golding, JM.; Robins, LN.; Regier, DA. *Psychiatric Disorders in America.* Free Press; New York: 1991. Obsessive compulsive disorder; p. 204-219.
- Kerjan G, Gleeson JG. Genetic mechanisms underlying abnormal neuronal migration in classical lissencephaly. *Trends Genet.* 2007; 23(12):623–30. [PubMed: 17997185]

- Knoepfler PS, Sykes DB, Pasillas M, Kamps MP. HoxB8 requires its Pbx- interaction motif to block differentiation of primary myeloid progenitors and of most cell line models of myeloid differentiation. *Oncogene*. 2001; 20(39):5440–8. [PubMed: 11571641]
- Kong A, Cox NJ. Allele-sharing models: LOD scores and accurate linkage tests. *Am J Hum Genet*. 1997; 61:1179–1188. [PubMed: 9345087]
- Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet*. 1995; 11(3):241–7. [PubMed: 7581446]
- Lenane MC, Swedo SE, Leonard H, Pauls DL, Sceery W, Rapoport JL. Psychiatric disorders in first degree relatives of children and adolescents with obsessive compulsive disorder. *J. Am. Acad. Child Adolesc. Psychiatry*. 1990; 29:407–412. [PubMed: 2347838]
- Long JE, Cobos I, Potter GB, Rubenstein JL. Dlx1&2 and Mash1 transcription factors control MGE and CGE patterning and differentiation through parallel and overlapping pathways. *Cereb Cortex*. 2009; 19(Suppl 1):i96–106. [PubMed: 19386638]
- Long JE, Swan C, Liang WS, Cobos I, Potter GB, Rubenstein JL. Dlx1&2 and Mash1 transcription factors control striatal patterning and differentiation through parallel and overlapping pathways. *J Comp Neurol*. 2009; 512(4):556–72. [PubMed: 19030180]
- Mannuzza S, Fyer AJ, Klein DF, Endicott J. Schedule for Affective Disorders and Schizophrenia--Lifetime Version modified for the study of anxiety disorders (SADS-LA): rationale and conceptual development. *J Psychiatr Res*. 1986; 20:317–325. [PubMed: 3806426]
- Mannuzza, S.; Fyer, AJ.; Endicott, J.; Klein, DF. Family Informant Schedule and Criteria. New York State Psychiatric Institute: NY; Anxiety Family Study Unit: New York, NY: 1985.
- Mojsin M, Stevanovic M. PBX1 and MEIS1 up-regulate SOX3 gene expression by direct interaction with a consensus binding site within the basal promoter region. *Biochem J*. 2009; 425(1):107–16. [PubMed: 19799567]
- Molero AE, Gokhan S, Gonzalez S, Feig JL, Alexandre LC, Mehler MF. Impairment of developmental stem cell-mediated striatal neurogenesis and pluripotency genes in a knock-in model of Huntington's disease. *Proc Natl Acad Sci*. 2009; 106(51):21900–5. [PubMed: 19955426]
- Nestadt G, Samuels J, Riddle M, Bienvenu OJ III, Liang KY, LaBuda M, et al. A family study of obsessive-compulsive disorder. *Arch. Gen. Psychiatry*. 2000; 57:358–363. [PubMed: 10768697]
- Neuteboom ST, Peltenburg LT, van Dijk MA, Murre C. The hexapeptide LFPWMR in Hoxb-8 is required for cooperative DNA binding with Pbx1 and Pbx2 proteins. *Proc Natl Acad Sci U S A*. 1995; 92(20):9166–70. [PubMed: 7568094]
- Pauls DL, Alsobrook JP, Goodman W, Rasmussen S, Leckman JF. A family study of obsessive-compulsive disorder. *Am. J. Psychiatry*. 1995; 152:76–84. [PubMed: 7802125]
- Pfohl, B.; Blum, N.; Zimmerman, M.; Stangl, D. Structured interview for DSM-III-R personality SIDR-R. University of Iowa College of Medicine; Iowa City, IA: 1989.
- Pinto A, Greenberg BD, Grados MA, Bienvenu OJ 3rd, Samuels JF, Murphy DL, et al. Further development of YBOCS dimensions in the OCD Collaborative Genetics study: symptoms vs. categories. *Psychiatry Res*. 2008; 160(1):83–93.
- Quinn JC, Molinek M, Mason JO, Price DJ. Gli3 is required autonomously for dorsal telencephalic cells to adopt appropriate fates during embryonic forebrain development. *Dev Biol*. 2009; 327(1): 204–15. [PubMed: 19121302]
- Sagerström CG. Pbx marks the spot. *Dev Cell*. 2004; 6(6):737–8. [PubMed: 15177017]
- Samuels JF, Riddle MA, Greenberg BD, Fyer AJ, McCracken JT, Rauch SL, et al. The OCD collaborative genetics study: methods and sample description. *Am J Med Genet B Neuropsychiatr Genet*. 2006; 141B(3):201–7. 5. [PubMed: 16511842]
- Samuels J, Nestadt G, Bienvenu OJ, Costa PT Jr, Riddle MA, Liang KY, et al. Personality disorders and normal personality dimensions in obsessive-compulsive disorder. *Br J Psychiatry*. 2000; 77:457–62. [PubMed: 11060001]
- Sanchez M, Jennings PA, Murre C. Conformational changes induced in Hoxb- 8/Pbx-1 heterodimers in solution and upon interaction with specific DNA. *Mol Cell Biol*. 1997; 17(9):5369–76. [PubMed: 9271414]

- Shugart YY, Samuels J, Willour VL, Grados MA, Greenberg BD, Knowles JA, et al. Genomewide linkage scan for obsessive-compulsive disorder: evidence for susceptibility loci on chromosomes 3q, 7p, 1q, 15q, and 6q. *Mol. Psychiatry*. 2006; 11:763–770. [PubMed: 16755275]
- Spitzer RL, Williams JBW, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. *Arch Gen Psychiatry*. 1992; 49:624–629. [PubMed: 1637252]
- Stemmers FJ, Gunderson KL, Illumina, Inc. *Pharmacogenomics*. 2005; 6(7):777–82. [PubMed: 16207153]
- Sungbo, Shim; Yujin, Kim; Jongdae, Shin; Jieun, Kim; Soochul, Park. Regulation of EphA8 Gene Expression by TALE Homeobox Transcription Factors during Development of the Mesencephalon. *Mol Cell Biol*. 2007; 27(5):1614–1630. [PubMed: 17178831]
- Takahashi K, Liu FC, Oishi T, Mori T, Higo N, Hayashi M, et al. Expression of FOXP2 in the developing monkey forebrain: comparison with the expression of the genes FOXP1, PBX3, and MEIS2. *J Comp Neurol*. 2008; 509(2):180–9. [PubMed: 18461604]
- Theil T. Gli3 is required for the specification and differentiation of preplate neurons. *Dev Biol*. 2005; 286(2):559–71. [PubMed: 16168404]
- Toresson H, Parmar M, Campbell K. Expression of Meis and Pbx genes and their protein products in the developing telencephalon: implications for regional differentiation. *Mech Dev*. 2009; 2000; 94(1-2):183–7. [PubMed: 10842069]
- van den Oord EJ, Kuo PH, Hartmann AM, Webb BT, Möller HJ, Hettema JM, et al. Genomewide association analysis followed by a replication study implicates a novel candidate gene for neuroticism. *Arch Gen Psychiatry*. 2008; 65(9):1062–71. [PubMed: 18762592]
- Weissman MM, Bland RC, Canino GJ, Greenwald S, Hwu HG, Lee CK, et al. The Cross National Collaborative Group. The cross national epidemiology of obsessive compulsive disorder. *J. Clin. Psychiatry*. 1994; 55(Suppl):5–10. [PubMed: 8077177]
- Wilson A, Laurenti E, Trumpp A. Balancing dormant and self-renewing hematopoietic stem cells. *Curr Opin Genet Dev*. 2009; 19(5):461–8. [PubMed: 19811902]
- Woodruff R, Pitts FM. Monozygotic twins with obsessional illness. *Am J Psych*. 1964; 120:1075–108.
- Yu T, Fotaki V, Mason JO, Price DJ. Analysis of early ventral telencephalic defects in mice lacking functional Gli3 protein. *J Comp Neurol*. 2009; 512(5):613–27. [PubMed: 19048639]
- Zeller R, Haramis AG, Zuniga A, McGuigan C, Dono R, Davidson G, Chabanis S, Gibson T. Formin defines a large family of morphoregulatory genes and functions in establishment of the polarising region. *Cell Tissue Res*. 1999; 296(1):85–93. [PubMed: 10199968]
- Zhang J, Wang X, Li M, Han J, Chen B, Wang B, Dai J. NANOGP8 is a retrogene expressed in cancers. *FEBS J*. 2006; 273(8):1723–30. [PubMed: 16623708]
- Zuniga A, Zeller R. Gli3 (Xt) and formin (ld) participate in the positioning of the polarising region and control of posterior limb-bud identity. *Development*. 1999; 126(1):13–21. [PubMed: 9834182]

Figure 1.

Figure 2.

Figure 3.

Association of specific SNPs to OCD (FBAT p value < 0.01) and Genetic Relative Risk (GRR) on chromosome 1q between *PBX1* and *LMX1A* (N=376 families).

Table 1

Marker	Allele	Allele frequency ¹	Position	S-E(S)	Var(S)	p	Genotype	GRR	95% C.I.
rs4387163	A/G	0.539	163,262,337	-39.15	114.752	0.0002 ²	A/G	1.58	1.19, 2.10
							G/G	1.97	1.33, 2.92
rs4657393	A/G	0.443	163,279,035	33.15	111.918	0.002	A/A	1.84	1.24, 2.74
							A/G	1.45	1.10, 1.91
rs7524769	A/G	0.442	163,280,159	34.15	112.918	0.001	A/A	1.85	1.24, 2.74
							A/G	1.44	1.09, 1.89
rs723821	A/C	0.141	163,283,336	22.733	53.928	0.002	A/C	1.57	1.15, 2.14
rs986362	A/G	0.529	163,283,623	-38.817	117.085	0.0003	A/G	1.58	1.19, 2.09
							G/G	1.97	1.33, 2.92

¹ Allele frequency based on first allele

² p=0.08; after adjustment for multiple tests; all others, p>0.1