Identification of **EFHC2** as a quantitative trait locus for fear recognition in Turner syndrome

Lauren A. Weiss¹,³, Shaun Purcell¹,³, Skye Waggoner³, Kate Lawrence⁴, David Spektor⁴, Mark J. Daly³,², Pamela Sklar¹,³,*, and David Skuse⁴,†

¹Psychiatric and Neurodevelopmental Genetics Unit, ²Department of Medicine, Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA, ³Medical and Population Genetics Group, Broad Institute of MIT and Harvard, USA and ⁴Behavioural and Brain Sciences Unit, Institute of Child Health, University College London, UK

Received July 28, 2006; Revised October 5, 2006; Accepted November 21, 2006

One-third of women with Turner syndrome (45,X) have autism-like social and communication difficulties, despite normal verbal IQ. Deletion mapping of the X-chromosome implicated 5 Mb of Xp11.3–4 as critical for recognition of facial fear, a quantitative measure of social cognition. Variability in fear recognition accuracy in Turner syndrome suggested the existence of a quantitative trait locus (QTL) revealed by X-monosomy. We aimed to identify the gene(s) influencing fear recognition by dense mapping of the 5 Mb region. Initial regression-based association mapping of fear recognition in 93 women with Turner syndrome across the critical region was performed, using genotype data at 242 single nucleotide polymorphisms (SNPs). We identified three regions of interest, in which 52 additional SNPs were genotyped. The third region then contained four SNPs associated with fear recognition (0.0030 > P > 0.00046). We obtained an independent sample of 77 Turner syndrome females that we genotyped for 77 SNPs in the initial regions of interest. Region three showed association in the same direction, maximal at SNPs rs7055196 and rs7887763 (P = 0.022 each). Four SNPs in strong linkage disequilibrium (LD), including this pair, span 40 kb within a novel transcript, **EF-hand domain containing 2 (EFHC2)**. In the combined Turner syndrome samples, the most strongly associated SNP (P = 0.00007) has frequency of 8.8% and an estimated effect size accounting for over 13% of the variance in fear recognition. **EFHC2** shows genealogy and extended LD consistent with directional selection. This novel QTL may influence social cognition in the general population and in autism.

**INTRODUCTION**

Facial expressions provide critical information about emotional states. Without conscious effort, most of us respond appropriately to these social cues because we accurately interpret the facial configuration—however, some read faces more adeptly than others. Innate social cognition, the ability to recognize, manipulate and behave with respect to socially relevant information (1), may be reduced in schizophrenia (2), vascular dementia (3), conduct disorder (4) and psychopathy (5), and are by definition impaired in autism (6). Social cognition influences the ability to function productively in society, and people with autism can be profoundly disabled by social deficits despite otherwise average or above-average IQ (7).

The development of social cognition depends both on an innate biological framework and a social and cultural context. Inborn components of social cognition are not well-understood, and few genetic studies have investigated this phenotype. Most commonly, social cognition has been assessed by a series of pictures developed to investigate the ability to recognize emotions as conveyed by facial expression (8). The ability to recognize (negative, in particular) facial emotions is a key measure of social-cognitive skills (9), in which the amygdala plays a central role (10). Specifically, efficient functional connectivity between amygdala and regions of temporal cortex specialized for face processing is necessary for normal social perception (6). Recognition of fear in faces, the measure we use in this study, activates the amygdala.
and fusiform cortex, and has been shown to be a very sensitive measure of social cognitive competence in autism and other conditions (11, 12). Increased amygdala size has been reported in autism (13), and is correlated with impairment in fear recognition (14).

Population data have shown that males do not score quite as well as females in their recognition of fearful facial expressions (15). Approximately 8.8% of otherwise normal males are less than 50% accurate (>2 standard deviations below the mean), whereas only 4% of females show this level of disability (9). However, nearly 20% of females with Turner syndrome score <50% (16). The deficit is associated with impaired amygdala-fusiform connectivity, similar in nature to that described in autistic individuals (6). X-monosomy may ‘unmask’ social disability; for example, mild features of autism, such as failure to understand nonverbal social cues, poor maintenance of eye contact and limited ‘Theory of Mind’ skills affect up to 30% of Turner syndrome females (16, 17). Like autistic adults, Turner syndrome females are significantly impaired at recognizing fear (P < 0.001) and anger (P < 0.006) in comparison with matched controls (46,XX) using the Ekman Pictures of Facial Affect (14). Females with Turner syndrome (45,X) have a 5% incidence of strictly defined autism (over 10 times higher than the general population) and substantially increased (≥200 times) risk of autism-related disorders (18), which are up to 10 times as common in males as females with normal-range intelligence (19). There are also structural brain similarities: both conditions have been associated with abnormally large amygdala (13, 14). Males (also X-monosomic) are at higher risk for neurodevelopmental disorders associated with poor social cognitive abilities, including psychopathy, early onset schizophrenia and especially disorders on the autistic spectrum (19).

Taken together, these observations suggest a potential role for X-linked genes in susceptibility to autism and associated deficits in social cognition; the possession of two intact X-chromosomes may be protective for normal females (46,XX) (20); males (46,XY) may be at increased risk for social deficits, and Turner syndrome females (45,X) may be highly susceptible to social disability. The variability in social phenotype among Turner syndrome females with similar IQ suggests a quantitative trait locus (QTL) on the X-chromosome with a specific influence on social cognition.

In order to narrow down the critical region for social cognitive ability on the X-chromosome, we recently performed deletion mapping in 13 women with partial deletions of the short arm of the X-chromosome (46,XXp−) (14). Among the partially deleted sample, there was no correlation between the cognitive phenotype and ovarian hormonal deficiency, as evidenced by ovarian dysgenesis, which is almost invariable in X-monosomy and may independently influence cognitive skills (21). Women with a deletion including a nearly 5 Mb region of Xp11.3–Xp11.4 (markers DXS1368 to DXS8083) in one copy had a fear recognition and/or amygdala size phenotype similar to X-monosomy, while women with deletions excluding this region were indistinguishable from (46,XX) women (14). Therefore, we have previously defined a 5 Mb region of the X-chromosome in which a QTL for social cognition is likely to exist. The objective of our study was to identify this putative QTL through dense mapping of the 5 Mb region.

RESULTS

Identification of an association with fear recognition

First stage low-coverage genotyping and analysis was performed in 93 Turner syndrome women to identify regions of potential association with fear recognition within the 5 Mb critical region that could be followed up in a second stage of denser coverage. Mass spectrometry was used to initially genotype 242 single nucleotide polymorphisms (SNPs) that cover the region at ~20 kb average spacing and 4 kb spacing in six genes of interest (MAOA, MAOB, NDP, UTX, DDX3 and USP9X) that have been reported to escape X-inactivation and be expressed in the brain (Supplementary Material, Table S1) (22–26). These six genes were considered particularly good candidates because a transcript influencing fear recognition (and potentially amygdala size), which are sexually dimorphic, would be likely to be present in the brain and have different expression dosage in males and females. Quantitative association analysis identified three areas of putative association (P < 0.01) (Fig. 1).

For the second stage analysis, we increased the SNP density in three areas of putative association (P < 0.01) by genotyping 52 additional SNPs in these regions of interest (Supplementary Material, Table S1). After including the additional data, only the third region contained (four) SNPs associated with decreased fear recognition (0.0030 > P > 0.00046) (Fig. 1). Empirically correcting for all 294 SNPs in a single-SNP analysis gives a P-value of 0.060. We considered this result very promising, given the small sample size, and wanted to validate this association in a second sample.

Second sample confirms EFHC2 as a QTL for fear recognition

Follow-up was performed in an additional 77 adult Turner syndrome females, by genotyping 77 SNPs in the three regions of interest identified in the first stage analysis (Supplementary Material, Table S1). Only region three showed an association between variation at two SNPs and fear recognition scores (rs7055196, rs7887763; P = 0.022 each). These two SNPs were in perfect linkage disequilibrium (LD) in our sample (r² = 1), and therefore rs7055196 was used in subsequent analyses. The associated SNPs span 40 kb within a novel transcript, EF-hand domain containing 2 (EFHC2), previously identified as FLJ22843. In the combined Turner syndrome sample, the most strongly associated SNP (rs7055196, P = 0.00007) has a frequency of 8.8% and an estimated effect size that accounts for over 13% of the variance in fear recognition. When the combined Turner syndrome sample was dichotomized into normal (greater than ~80% accuracy; N = 96) versus low (less than ~70% accuracy; N = 56) fear recognition scores, the frequency of the risk allele at rs7055196 was 1% in normal scorers and 20% in low scorers, with no one carrying the risk allele found above the population mean (risk allele carrier Z ≤ −0.36). Therefore, variation in EFHC2 is associated with social cognition, as measured by recognition of fear in faces, in two Turner syndrome samples.

EFHC2 is a novel transcript spanning 195 kb with a predicted mRNA of 2.5 kb and 15 exons encoding a protein of...
Figure 1. Quantitative trait analysis of fear recognition in Turner syndrome. (A) Linkage disequilibrium plot of 5 Mb region generated by Haploview (http://www.broad.mit.edu/mpg/haploview/), showing $D'$ statistics (red = high, white = low). (B) The three regions identified by locus-wide SNP analysis. (C) Results of whap analysis. The blue line represents a 4-SNP sliding window haplotype analysis for the original sample, the red line shows significance after adding the second sample. (D) Block structure of region three indicates three haplotype blocks; the most significant associations are in the second block, 3.II, which spans 40 kb within EFHC2. The genomic layout of EFHC2 is shown (not to scale), with the approximate positions of genotyped SNPs indicated by arrows. (E) The four SNPs in the associated haplotype are shown, with a plot of the quantitative trait distribution by allele at SNP rs7055196.
749 amino acids. Orthologs are reported in diverse species including *Drosophila* and *Danio*. This transcript is widely expressed in peripheral tissues and the central nervous system. *EFHC2* is predicted to contain two recognizable types of domain: three DM10 domains of unknown function and one EF-hand domain. EF-hand domains include a 12-residue loop with two 12-residue alpha-helical domains on either side, and are generally found in calcium-binding proteins (27). The four risk SNPs include one nonsynonymous change (S430Y, rs2208592) at the residue before the start of the third predicted DM10 domain. The coding variant Tyr430 did not account for all of the evidence for association, as analysis of the other risk variation conditional on this allele showed modest association (*P* = 0.047).

**Analysis of parent-of-origin at EFHC2**

Because of the previous work suggesting that Turner syndrome individuals whose X-chromosome is maternally inherited show decreased social skills compared to those with paternally inherited X-chromosomes (28), we wanted to investigate whether the risk allele is underrepresented on paternal copies of *EFHC2*. Overall in our sample, nearly 77% of the chromosomes of known parental origin were maternally inherited, which is consistent with other estimates (29). There was no mean difference in fear recognition between individuals with maternally inherited (mean *Z* = −0.74) and paternally inherited (mean *Z* = −0.74) X-chromosomes in this sample. Eleven individuals with an X-chromosome of known parental origin have the risk allele at rs7055196. Of these, only one is of paternal origin, compared with the 2.6 expected based on the overall rate of 23% paternal. The individual with a paternally inherited X-chromosome had a fear score near to the average of the risk group (*Z* = −1.77). Therefore, although we see a slight excess of maternal inheritance, this could be observed by chance in such a small sample.

**Signatures of evolutionary selection at EFHC2**

Both coding and noncoding regions of the *EFHC2* locus are highly evolutionarily conserved (data not shown), suggesting the possibility of a unique evolutionary history for this region. We investigated this hypothesis using dense genotype data in multiple populations made available by the International HapMap project (www.hapmap.org) and an analysis tool that looks for extended haplotype homozygosity (EHH) as a signature of natural selection (30). One would expect common alleles in the population to be old and show low levels of LD, therefore common alleles with high LD extending long distances are suggestive of positive selection. Among the populations included in the International HapMap, the European–American (CEPH) population is likely to be most similar to the Turner syndrome samples from Great Britain. The haplotype block [as determined by Gabriel et al. (31)] containing the two risk SNPs genotyped in this sample had a very unusual genealogy, described below. In order to examine the mutational history, we modified the defined block slightly (shortened by four SNPs) to eliminate recombinant haplotypes, which did not change the overall pattern. The haplotype containing the risk alleles (frequency 7%) is most similar to the ancestral haplotype and 12 mutational steps from the common haplotype (frequency 83%) in the population (Fig. 2). We next investigated haplotypes that were outliers in their frequency bin. *EFHC2* contained 56 haplotypes in nine blocks, of which four were above the 95th percentile of all haplotypes on the X-chromosome for relative EHH (REHH). A haplotype at the 3’ end of *EFHC2* showed a significant relative REHH score (*P* = 0.0063). Markers in this haplotype block showed LD with the risk variation (0.6 ≤ *D’* ≤ 1.0). This 3’ haplotype was in the top 1/2 percentile of the X-chromosome, a chromosome known to be subject to strong selective forces (32,33).

**DISCUSSION**

Variation in *EFHC2* is associated with social cognitive competence, as measured by recognition of fear in faces, in two Turner syndrome samples. The most strongly associated (noncoding) SNP accounts for 13% of the variance in fear recognition in our combined (45,X) sample. We do not find evidence of parent-of-origin bias contributing to variation in social cognition in this sample. *EFHC2* is a novel transcript expressed in the brain and other regions predicted to contain a calcium-binding domain as well as conserved domains of unknown function. The known coding change (Ser430Tyr) present on the risk haplotype did not appear to account for all of the evidence for association at this locus. However, there were only two individuals with Ser430 (nonrisk) and risk alleles at the other three SNPs (which were statistically inseparable due to strong LD), so although our data is suggestive that noncoding variation or yet untyped variation is contributing to the association at *EFHC2*, we do not consider it definitive. A third sample that replicated the association between *EFHC2* and fear recognition could help narrow the associated region and clarify any parent-of-origin bias.

The haplotype including risk variation in this sample is found at <10% frequency in the population. Although not conclusive, the signature at this locus is consistent with the common allele being influenced by directional selection and the susceptibility allele being ancestral in the population. Recently, ancestral-susceptibility models have been explored and found to be consistent with examples of common human disease (34).

As a candidate to be involved in the development of neural circuits that are critical for normal social cognitive competence, a calcium-binding protein could have diverse neuronal functions. Perhaps the best-studied calcium-sensing protein is calmodulin, which regulates many downstream proteins critical to neuronal signaling. Preliminary studies have reported associations between calcium-sensing proteins and schizophrenia (35) and Alzheimer’s disease (36). In addition, *EFHC1*, which has 37% amino acid identity with *EFHC2*, has been recently shown to specifically increase R-type calcium currents and facilitate apoptosis, potentially involved in developmental ‘pruning’ of neurons (37) or neurotransmission (38). Mutations in *EFHC1* cause juvenile myoclonic epilepsy (37). Notably, not only have defects in developmental pruning been hypothesized to account for observations of increased brain size in autism (13,39), but at least 30% of individuals with
Figure 2. Long-range haplotype analysis in EFHC2. (A) EFHC2 is represented (top) with vertical bars representing HapMap (release #16) SNPs genotyped in the CEPH sample. The haplotype block under examination (circled) was shortened at the 5' end to exclude recombination. The haplotype tree is shown with the size of each numbered box representing the frequency of the corresponding haplotype. Each common human (CEPH) haplotype (frequency > 1%) is shown under the chimpanzee haplotype, with the fear recognition risk variation denoted with arrows and rs numbers below the SNPs. The chimpanzee sequence at rs2208592 is based on a BLAST comparison, but falls within a gap in the aligned chimpanzee sequence. (B) EFHC2 is represented (top) with vertical bars representing HapMap (release #16) SNPs genotyped in the CEPH sample. The haplotype block under examination (circled) is at the 3' end of the gene. Below, REHH is plotted by frequency for each haplotype across the X-chromosome, with the 95th and 99th percentiles of the distribution represented by thin and thick lines, respectively.
In Supplementary Material, Figure S1.
The distribution of scores in the combined sample is shown not differ in fear recognition distribution (data not shown). EFHC2 is a candidate to play a role in increased susceptibility to autism among males, and should be examined for evidence of imprinting and X-inactivation. Further investigation of EFHC2 and the protein it encodes could provide insight into the biological and evolutionary underpinnings of normal social development and deficits therein.

MATERIALS AND METHODS

Subjects

The initial sample of 93 adult Turner syndrome (45,X) females and the second set of 77 Turner syndrome (45,X) females were selected from a national survey of Turner syndrome, from the records of the Wessex Regional Genetics Laboratory, and from a specialist service for women with Turner syndrome at the Middlesex Hospital in London. They were ascertained only for having Turner syndrome, without regard to fear recognition ability. The majority of cases of X-monosomy were determined at the Wessex Regional Genetics Laboratory by parental origin of the chromosome. The presence of two or more cells with the chromosome and/or the presence of a structurally abnormal chromosome. The presence of two or more cells with the same constitution in the 100 cells was considered to represent a cell line. Only non-mosaic (single cell-line) cases were utilized in our mapping study. Parental origin of the X-chromosome in Turner women was determined by PCR amplification of polymorphic microsatellite repeat markers on distal Xp, as previously reported (41).

Recently, a computerized test battery known as Schedules for the Assessment of Social Intelligence (SASI) has been established for the measurement of social-cognitive functions (16). SASI tasks include facial expression recognition, face recognition memory, gaze-monitoring skills and a ‘Theory of Mind’ task. Each of these domains may be impaired in autism (42). The Turner syndrome individuals were assessed by SASI to determine fear recognition ability (the only trait used in this study), but did not have amygdala volume measurements. The subjects have a mean IQ of 97.2 (SD = 16.6, range is 54–140) and average age 26.8 (SD = 10.9). The Z-scores for fear recognition are adjusted by age and sex, as described (16). For 13 of the 170 individuals fear Z-score was calculated based on 7 instead of 10 fear recognition tasks. The original and second sample did not differ in fear recognition distribution (data not shown). The distribution of scores in the combined sample is shown in Supplementary Material, Figure S1.

Genotyping and analysis

Mass spectrometry in the Mass ARRAY system (Sequenom, La Jolla, CA) was used to initially genotype 242 SNPs, and to genotype 52 additional SNPs in the regions of interest identified (Supplementary Material, Table S1). 77 SNPs in the three regions of interest identified in the first stage analysis were genotyped in the second sample (Supplementary Material, Table S1). Quantitative association analysis was performed using the whap program, which can perform single SNP as well as sliding window haplotype regression. We performed regression on fear recognition score and estimated significance at each SNP via permutation, which is robust to any non-normality of the data (http://pngu.mgh.harvard.edu/~purcell/whap/) (43). Empirical experiment-wide significance was determined in the initial sample. ‘Low’ fear scores (less than ~70% accuracy) were defined as below the 10th population centile, standardized by age and sex, while ‘normal’ scores were above the 10th centile (~80% or better accuracy).

We assessed evidence for evolutionary selection using dense genotype data in multiple populations made available by the International HapMap project (www.hapmap.org). We examined the phased HapMap data (release #16) from the European–American (CEPH) samples across the X-chromosome. The genotype data was broken into haplotype blocks (31), and the ancestral allele was determined by alignment with the chimpanzee genome sequence. The aligned chimpanzee sequence had a gap within EFHC2 across S430Y, so we determined the ancestral allele in this location by BLAST alignment with chimpanzee (98% match over 600 bp with accessions: AC159463, AC161010). We used an analysis tool that looks for extended EHH as a signature of natural selection (30), and compared REHH scores within EFHC2 to the rest of the X-chromosome. To compare across blocks, we matched by looking at a distance of 0.3 cm. In order to assess significance, haplotypes are binned by frequency (in units of 5%). For a given haplotype, the REHH score (log-transformed to achieve normality) can be compared to the mean in its bin, and a P-value assigned based on the number of standard deviations it is from the mean.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support of Eric Lander, who encouraged us to initiate this project and has provided thoughtful discussions of the data and manuscript. We also acknowledge Pardis Sabeti and Steve Schaffner, who provided assistance in data analysis and interpretation, Gerry Conway, who assisted in the identification of Turner syndrome women for both studies, Patricia Jacobs and the Salisbury District General Hospital, who did the karyotyping of most (45,X) subjects, and Leslie Gaffney who helped with figure preparation. Funding was provided by the National Alliance for Autism Research, the Nancy Lurie Marks Family Foundation, the Child Growth Foundation and the Wellcome Trust.

Conflict of Interest statement. None declared.