Identification of Novel FMR1 Variants by Massively Parallel Sequencing in Developmentally Delayed Males


Department of Human Genetics, Emory University School of Medicine, Atlanta, Georgia
Department of Biochemistry and Pediatrics, Emory University School of Medicine, Atlanta, Georgia

Received 7 May 2010; Accepted 1 July 2010

Fragile X syndrome (FXS), the most common inherited form of developmental delay, is typically caused by CGG-repeat expansion in FMR1. However, little attention has been paid to sequence variants in FMR1. Through the use of pooled-template massively parallel sequencing, we identified 130 novel FMR1 sequence variants in a population of 963 developmentally delayed males without CGG-repeat expansion mutations. Among these, we identified a novel missense change, p.R138Q, which alters a conserved residue in the nuclear localization signal of FMRP. We have also identified three promoter mutations in this population, all of which significantly reduce in vitro levels of FMR1 transcription. Additionally, we identified 10 noncoding variants of possible functional significance in the introns and 3′-untranslated region of FMR1, including two predicted splice site mutations. These findings greatly expand the catalog of known FMR1 sequence variants and suggest that FMR1 sequence variants may represent an important cause of developmental delay.

Key words: FMR1; massively parallel sequencing; developmental delay

INTRODUCTION

Fragile X syndrome (FXS, MIM 300624) is the most common inherited cause of intellectual disability. In addition to a variable degree of intellectual impairment, FXS patients often exhibit autism-like behaviors, such as gaze avoidance, hand-flapping, and tactile defensiveness. Other classic features of FXS include macroorchidism and an elongated face with large everted ears. However, due to the subtlety and variable expressivity of the more distinguishing characteristics, the identification of a causal mutation is necessary for the diagnosis of FXS [Garber et al., 2008].

The most common causal mutation leading to FXS is the expansion of the CGG trinucleotide repeat located within the 5′-untranslated region (UTR) of the FMR1 gene (MIM 309550) [Verkerk et al., 1991; Ashley et al., 1993]. This expansion, referred to as the full mutation, represents expansion beyond 200 repeats. The full mutation leads to hypermethylation of the FMR1 promoter, thereby preventing expression of FMR1 and its gene product, FMRP [Sutcliffe et al., 1992; Chiurazzi et al., 1998]. FMR1 deletions are the second most common known cause of FXS, although far less common than the repeat expansion mutation [Coffee et al., 2008]. While it seems plausible that sequence variants affecting the expression or function of FMRP could represent a third important cause of FXS, only three such mutations, a missense mutation (I304N) and two small deletion nonsense mutations, have been reported since FMR1 was identified in 1991 [De Boulle et al., 1993; Lugenbeel et al., 1995]. Therefore, FMR1 sequencing is rarely performed in the clinical setting, due to the expectation of a low diagnostic yield. Also, methodological constraints have previously prevented a thorough assessment of FMR1 sequence variation in a large number of patients, leaving the true significance of pathogenic sequence variants in FMR1 unknown [Chiurazzi et al., 1994; Additional supporting information may be found in the online version of this article.

Grant sponsor: National Institutes of Health Grants; Grant numbers: AG029749, HD020521, HD024064; Grant sponsor: FRAXA Foundation Fellowship.

*Correspondence to:
Stephen T. Warren, Department of Human Genetics, Emory University School of Medicine, Whitehead Biomedical Research Building, 615 Michael Street, Suite 301, Atlanta, GA 30322. E-mail: swarren@emory.edu
Published online 26 August 2010 in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/ajmg.a.33626

How to Cite this Article:
Reyniels et al., 1996; Vincent et al., 1996; Wang et al., 1997; Gronskov et al., 1998; Castelli-Bel et al., 1999; Shinahara et al., 2004; Collins et al., 2010].

Massively parallel sequencing (MPS) vastly improves upon the cost-effectiveness and throughput of traditional Sanger sequencing, enabling facile detection of sequence variation at a scale that was previously impractical [Shendure and Ji, 2008]. One of the many applications that have emerged for MPS is targeted resequencing to detect novel mutations in particular genomic regions, such as a collection of candidate genes [Dahl et al., 2007] or the entire exome [Ng et al., 2009]. Due to the Gigabase-scale capacity of MPS, targeting of a single candidate gene in a single patient is generally inefficient. However, through the use of a pooled-template design, a single gene can be sequenced in multiple individuals simultaneously to screen for the presence of rare or novel sequence variants, thus allowing for efficient, cost-effective large-scale targeted resequencing [Druley et al., 2009; Ingman and Gyllensten, 2009; Koboldt et al., 2009; Out et al., 2009].

To detect potentially pathogenic FMR1 sequence variants, we employed pooled-template MPS to assess the promoter, all 17 exons, and a substantial portion of the intronic sequence of FMR1 in 963 developmentally delayed males referred for FMR1 repeat testing but found not to have the full mutation. We identified one patient with the novel missense change p.R138Q, which alters a conserved residue within the nuclear localization signal (NLS) of FMRP. Furthermore, we report three novel promoter variants, all of which reduce the in vitro expression of FMR1, and several novel sequence variants in conserved noncoding regions of FMR1, including two predicted splice site mutations. Together, these novel variants suggest that there may be clinical utility in diagnostic FMR1 sequencing for developmentally delayed males.

MATERIALS AND METHODS

Clinical Population
While loss of function mutations in FMR1 may cause a phenotype resembling FXS, we would anticipate that mutations that simply reduce FMRP function or expression would result in a more subtle phenotype. Therefore, we decided to sequence FMR1 in patients who had tested negative for FMR1 repeat expansion at the Emory Genetics Laboratory over a 5-year span. Because the current standard of care is for all children presenting with developmental delay to be tested for FMR1 repeat expansion, the patients in this clinical population do not necessarily exhibit the classic FXS phenotype, but rather represent the more general diagnosis of developmental delay. Indeed, only 2–3% of such referred samples test positive for the full mutation of FMR1. For ease of interpretation, we elected to only sequence males, in which a variant would be hemizygous and more likely to be penetrant. Patients older than age 18 at the time of testing were excluded, as the clinical indication for their FMR1 repeat test was more likely to have been triggered by a question concerning transmission risk or the appearance of later onset premutation-like tremor/ataxia phenotype, not an early onset developmental delay. Racial identification was available for only 241 of the 963 patients sequenced (25.0%). Among these, 164 (68.1%) were of European descent, 74 (30.7%) were of African descent, and 3 (1.2%) were of Asian descent.

Genomic DNA Samples
We obtained deidentified aliquots of genomic DNA from the Emory Genetics Laboratory, Department of Human Genetics, Emory University School of Medicine, for every male under age 18 who was referred for and tested negative for the FMR1 full mutation from April 2002 to August 2007. In total, 1,392 aliquots were obtained. The genomic DNA samples had previously been extracted from whole blood by standard methods in a CLIA-certified environment. The Emory University Institutional Review Board approved this use of deidentified clinical samples.

Massively Parallel Sequencing

Targeting FMR1. As seen in Figure S1, four long-range PCR (LR-PCR) amplifications were designed to target FMR1 (supporting information Fig. S1 may be found in the online version of this article). The LR-PCR primer pairs were as follows: FMR1A-R: 5’-CACGTGGGTTGTGACC-3’ and FMR1B-F: 5’-CTACAATCAACAAACACGCACTACTGCTACAT-3’; FMR1B-R: 5’-AATTTCACGATATCTGTGG-3’; FMR1C-F: 5’-GTGAGAAATTTAATGAG-3’ and FMR1C-R: 5’-GAGACATATGCCCCCATGTTATAGT-3’; FMR1D-F: 5’-CCTATGCTGTTATCTATGTTAAAG GGTGCTATAG-3’ and FMR1D-R: 5’-TATAATGCTTATGTTAG GGCACCCATGTAATA AA-3’.

Genotyping.
Massively Parallel Sequencing

LR-PCR amplicon pooling.

Targeting FMR1. As seen in Figure S1, four long-range PCR (LR-PCR) amplifications were designed to target FMR1 (supporting information Fig. S1 may be found in the online version of this article). The LR-PCR primer pairs were as follows: FMR1A-R: 5’-CACGTGGGTTGTGACC-3’ and FMR1B-F: 5’-CTACAATCAACAAACACGCACTACTGCTACAT-3’; FMR1B-R: 5’-AATTTCACGATATCTGTGG-3’; FMR1C-F: 5’-GTGAGAAATTTAATGAG-3’ and FMR1C-R: 5’-GAGACATATGCCCCCATGTTATAGT-3’; FMR1D-F: 5’-CCTATGCTGTTATCTATGTTAAAG GGTGCTATAG-3’ and FMR1D-R: 5’-TATAATGCTTATGTTAG GGCACCCATGTAATA AA-3’.

LR-PCR amplicon pooling.

The concentrations of LR-PCR amplicons were measured by fluorometric quantification with PicoGreen dsDNA reagent (Molecular Probes, Eugene, OR). Equi-
molar pools were created by first combining across 19 patients within a given LR-PCR (i.e., A, B, C, or D as shown in Fig. S1). The following amplicon amounts were used per patient: 33.4 ng of LR-PCR-A; 145.8 ng of LR-PCR-B; 172.1 ng of LR-PCR-C; and 200 ng of LR-PCR-D. Included in each LR-PCR-C pool was 172.1 ng of LR-PCR-C from a patient with the single known missense mutation in FMR1 [De Boule et al., 1993]. This 1304N positive control was intended to serve as sentinel for singleton detection in each pool and as a direct measure of the false negative rate of our approach. Each amplicon pool was purified to remove excess primers with the PureLink PCR purification kit (Invitrogen). Purified amplicon pools were quantified with the Nanodrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE). For each set of 19 patients, purified amplicon pools A, B, C, and D were combined in equimolar fashion to a total of 4 µg.

Target library construction. Amplicon pools were fragmented by sonication (Sonicator S-4000, Misonix, Farmingdale, NY) with the following parameters: eight pulses of 30 sec each, with 2 min rest between pulses, at a power output of 20%. The fragmentation range, as visualized by gel electrophoresis, was from 100 to 400 bp. End-repair, adaptor ligation, and PCR amplification were performed as described previously [Okou et al., 2009].

Genome analyzer sequencing and analysis. From each processed amplicon pool, a total of 6 pmol of the library were added to one lane of the Genome Analyzer (GA) flowcell (Illumina, San Diego, CA). Single-end sequencing for 26 cycles was performed on the GA using the Illumina protocols for cluster generation and sequencing-by-synthesis. GA image analysis and base-calling were performed, respectively, with the Firecrest and Bustard software packages from the Illumina GA Pipeline. Mapping was performed with the MAQ software package [Li et al., 2008], using the following mapping parameters: a maximum mismatch (-n) of 3, a mutation rate (-m) of 0.01, and a maximum sum of mismatching base qualities (-e) of 140. The short sequence reads were mapped to the FMR1 region of the human genome reference (NT_011681.15, g.3435700–g.3475545). The reference sequence was modified at the following SNPs, where the minor allele is incorrectly included in the reference sequence: rs1270092, rs1270091, rs4824232, rs4824233, rs5904650, rs11342854, rs61419778, rs5904816, and rs668020458. The reference sequence was further altered by masking repeat elements with RepeatMasker.

Variant detection. Only base-calls with a GA quality score greater than 45, corresponding to an error rate of 0.003%, were considered for variant detection. The frequency of each base was calculated at each nucleotide position in the pileup of mapped short reads. To account for context-specific errors in GA sequencing [Dohm et al., 2008; Harismendy et al., 2009], we measured the error rate at each nucleotide position by performing a similar GA sequencing run on a single male individual, in which every base position should have a theoretical 100% “major allele frequency.” Call this experiment the “null experiment.” The observed “minor allele frequency” at each nucleotide position in this “null experiment” was used as the expected error rate for each of the pooled sequencing runs. Since each pool contained 20 individual alleles, true variants would be expected to be observed in 5% or more of the reads. To call any given base, we compared two models. In the first model, the number of minor allele reads was assumed to be binomially distributed, with minor allele frequency equal to the error rate estimated in the “null experiment.” This model corresponds to the “null” model of no variants present. In the second model, we assume the minor allele count is also binomially distributed, but with frequency ≥5%. We compare the two models with a likelihood ratio test, and call a variant if the variant model fits better than the null model with P < 1 × 10−4.

Variant confirmation. Standard Sanger sequencing was used to confirm all variant calls. Fresh LR-PCRs were generated from each individual sample in a pool positive for a given putative variant. The amplicons were purified and Sanger sequenced using primers targeted to confirm the variant. Each chromatogram was visually inspected for the presence of the putative variant.

Control Genotyping

The control samples used for the genotyping of novel variants were obtained from the NIMH Human Genetics Initiative. All controls were adult males of European descent who had been screened to rule out psychiatric disorders. Genotyping was performed by the iPlex Gold method (Sequenom, San Diego, CA) as per the manufacturer’s instructions, using primers (see supporting information Table S1 which may be found in the online version of this article) designed with the Sequenom Assay Design 3.1 software. The single-base primer extension method failed for three variants, c.−332G>C, c.−254A>G, and c.−67G>C, described below. These three variants obliterate restriction sites for SacI, EcoNI, and BseYI, respectively. Thus, restriction digestion was used to genotype for these three variants. For both iPlex and restriction digestion genotyping, a positive control was included in every plate to confirm the sensitivity of the assay. After genotyping, a fresh PCR was produced for all control samples in which a minor allele was detected, and traditional Sanger sequencing was used to confirm the presence of the minor allele.

In Silico Analysis

Assessments of the cross-species conservation of the FMRP amino acid sequence and the FMR1 promoter sequence were performed with the ClustalW2 sequence alignment program. Predictions of the effects of amino acid substitution were performed with the programs PMut [Ferrer-Costa et al., 2005], PANTHER PSEC [Thomas and Kejariwal, 2004], SIFT [Ng and Henikoff, 2003], and PolyPhen [Ramensky et al., 2002]. For each variant position, the regional conservation across placental mammals was assessed by phyloP; these values were obtained from the UCSC Genome Browser [Siepel et al., 2005; Rhead et al., 2009]. The program NNsplice was used to predict splice sites that may be created or obliterated by novel sequence variants [Reese et al., 1997]. Variants predicted to alter miRNA binding to the FMR1 3’UTR were identified with the program miRanda [John et al., 2004; Betel et al., 2008].

Luciferase Assays

Plasmid construction. The pFMR1-luc plasmid has been previously reported [Smith et al., 2004]. A multistep process was used
to introduce the novel promoter variants into pFMR1-luc. First, the LR-PCR-A amplicon from each of the patients identified with novel promoter variants was cloned using the TOPOTA Cloning kit (Invitrogen). The TA clones were cut with NarI and NruI and ligated into pBluescript, and then cut with KpnI and HindIII and ligated into the pFMR1-luc plasmid. Sanger sequencing was used to confirm that the three variant plasmids contained the novel promoter variants and that all four plasmids contained an equivalent number of CGG repeats, which was determined to be 8. The pGL3-Basic and pRL-TK plasmids were purchased from Promega (Madison, WI).

**Cell culture and transfections.** HeLa cells were cultured at 37°C with 5% CO₂ in DMEM with 10% fetal bovine serum. Twenty-four hours before transfection, 1 × 10⁶ cells were plated in 2 ml of media in each well of six-well cell culture dishes. Transfections were carried out in Opti-MEM (Invitrogen), using 10 μl of Lipofectamine 2000 (Invitrogen) and 1 μg of total plasmid in a 10:1 ratio (firefly plasmid:control Renilla plasmid). Each plasmid was transfected into six separate wells. Four hours after transfection, the media containing the transfection reagent and plasmids was replaced with DMEM with 10% fetal bovine serum. Forty-eight hours after transfection, cells were harvested with 500 μl 1× Passive Lysis Buffer (Promega) by rocking at room temperature for 15 min. Lysates were cleared of cell debris by centrifugation at 14,000 rpm for 5 min at 4°C.

**Luciferase assays.** Protein concentrations of the lysates were measured by the Bradford assay. The Dual-Luciferase Reporter Assay System (Promega) was used to measure luciferase activity. From each lysate, 5 μg of protein was added in 20 μl total volume to a lumimeter tube. To each lumimeter tube, 100 μl of LAR II was added. A manual-load lumimeter was used to measure the luminescence over a 10-sec period, following a 2-sec premeasurement delay. The lumimeter measurement was repeated after the addition of 100 μl of Stop & Glo reagent. For each lysate, the firefly luciferase values were divided by the Renilla luciferase values. The results of six independent transfections were averaged and the standard deviation was calculated for each plasmid.

## RESULTS

### Sequence Variants in FMR1

Through the use of pooled-template MPS, we sequenced FMR1 in 963 developmentally delayed males, each to an average sequence depth of 130-fold coverage per sample. As shown in Table I, we identified 59 known polymorphisms in FMR1 in this population, 57 of which were previously cataloged in dbSNP (build 130). This provides evidence of the sensitivity of pooled-template MPS. The only FMR1 SNPs included in dbSNP that we did not detect were those with low or unknown population frequencies; these were likely not present in our patient population.

Two other variants, c.18G>T and c.–418_.–417insGGC, although not previously identified in controls, were previously identified in intellectually disabled patients [Gronskov et al., 1998; Mila et al., 2000]. We observed the c.18G>T variant in 13 samples (1.4%) but also in 19 of 1,401 controls (1.4%), indicating this is a benign polymorphism of FMR1. Intriguingly, the previously identified c.–418_.–417insGGC variant was detected in the current study in 8/963 patients (0.83%), but in 0/700 Caucasian control males (P < 0.05, Fisher’s exact test). Consistent with this finding, Mila et al. identified the variant in 1/25 fragile X-like males, but in none of 250 control males. It is not obvious how this insertion might influence FMRP expression as the site of insertion is not well conserved and since the site of insertion is itself GCC; thus this variant is a 3-bp duplication. Moreover, Mila et al. saw no difference in steady-state FMRP levels by Western analysis nor any differences in gel shift or gene reporter analysis. Since we do not know the ethnicities in our patient samples, it is possible that this significant association results from population substructure.

Additionally, we detected 130 variants in FMR1, which, to our knowledge, have not previously been reported. None were previously cataloged in dbSNP, nor were they detected in the first nine publicly available personal genomes [Levy et al., 2007; Bentley et al., 2008; Siva, 2008; Wang et al., 2008; Wheeler et al., 2008; Ahn et al., 2009]. Among these variants, 63.1% were only detected in one individual, while 36.9% were detected in multiple individuals. The

### Table I. FMR1 Sequence Variants Detected in 963 Developmentally Delayed Males

<table>
<thead>
<tr>
<th>Novel variants</th>
<th>Unique</th>
<th>Recurrent</th>
</tr>
</thead>
<tbody>
<tr>
<td>S’US</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Promoter</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>5’UTR</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Exon: nonsynonymous</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Exon: synonymous</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Intron</td>
<td>65</td>
<td>38</td>
</tr>
<tr>
<td>3’UTR</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>3’DS</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Known polymorphisms</th>
<th>Unique</th>
<th>Recurrent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Unique variants were detected in only one sample pool, while recurrent variants were detected in more than one sample pool. US, upstream sequence; UTR, untranslated region; DS, downstream sequence.
The majority of the novel variants were found in the introns of FMR1. However, novel sequence variants were detected in all regions of the gene. A full summary of the sequence variants detected is provided in Table SII (supporting information Table SII may be found in the online version of this article). Although pooled-template MPS is a highly sensitive method [Druley et al., 2009; Ingman and Gyllensten, 2009; Out et al., 2009], subtle imbalances in pool construction can result in the underrepresentation of a given template in the sequence output, which leads to false negatives. To assess our false negative rate, we included a LR-PCR amplicon containing the rare mutation p.I304N in every pool. Because the positive control amplicon was pooled in the same quantity and fashion as all of the patient amplicons, the frequency of successful detection of p.I304N likely reflects the frequency of successful detection of other singleton variants in a pool. At our detection threshold, p.I304N was not detected in 12/51 pools. Based upon this false negative rate of 23.5%, it is possible that we missed up to 40 previously unknown true variants in the population of 963 developmentally delayed males. [Note that we successfully detected 130 previously unknown variants, and assuming we missed 23.5% of the variants that were present, then we missed 0.235 × 30/0.765 = 0.92] and no more than one true missense variant (0.235 × 1/0.765 = 0.31), thus suggesting that we detected the majority of FMR1 variants that are most likely to have a functional effect in this population.

Identification of the Novel Missense Variant p.R138Q

In the 963 developmentally delayed males sequenced, we detected only one novel missense variant in FMR1. This variant, c.413G>A (Fig. 1A), was identified in a sample from a patient of European descent, but was not detected in 1,385 control males of European descent. The c.413G>A variant encodes an arginine-to-glutamine substitution at codon 138. Arginine-138 is believed to be one of the basic residues comprising the NLS of FMRP [Eberhart et al., 1996; Bardoni et al., 1997] and is highly conserved through Drosophila (Fig. 1B). The sole exceptions to this conservation are the pufferfish species Takifugu rubripes and Tetraodon nigroviridis, in which glutamine is used at this position. The R-to-Q missense substitution is predicted to be pathological by PMut (NN output: 0.84) [Ferrer-Costa et al., 2005] and PANTHER PSEC (subPSEC: −4.3) [Thomas and Kejariwal, 2004], but tolerated by SIFT (score: 0.22) [Ng and Henikoff, 2003] and PolyPhen (PSIC Score Δ: 0.11) [Ramensky et al., 2002].

Novel Variants in the FMR1 Promoter

Identification and preliminary functional evaluation. We detected three novel sequence variants in the minimal promoter of FMR1. Roman numerals I–III represent the three transcription start sites of FMR1. The GC boxes bind the transcription factor Sp1. (FIG. 2) A: DNA chromatograms of the three novel promoter variants. C: Mammalian conservation of the overlapping AP-2 binding site and GC box, the overlapping Inr-like and TATA-like sequences at transcription start site II, and the Inr-like sequence at transcription start site I.
Noncoding Variants in FMR1

To determine if any of the 127 novel noncoding variants in FMR1 are associated with developmental delay, we genotyped large numbers of control males of European descent for all variants occurring at highly conserved bases. We assessed the sequence conservation of the genomic region and nucleotide position for each variant by phastCons and phyloP scores, respectively [Siepel et al., 2005]. Variant positions with phastCons score >0.8 and phyloP score >1.5 were defined to be highly conserved. Similarly, we genotyped controls for variants predicted to alter splicing or miRNA binding. The splice prediction program NNSplice [Reese et al., 1997] was used to identify any variants that obliterate known splice sites or introduce novel splice sites. Two variants, c.880 + 885A>G and c.1472 − 521C>G, were predicted to introduce novel splice donor sites with high likelihood (>0.85). The miRNA target prediction software miRanda [John et al., 2004; Betel et al., 2008] suggested that the novel variant c.*746T>C may reduce the binding of miR-548p, miR-891a, and miR-454 to the 3’UTR of FMR1. As seen in Table II, six novel intronic variants and four novel 3’UTR variants of possible functional impact were not identified in a large sample of control Caucasian males. Furthermore, one intronic variant and two 3’UTR variants exhibited a significantly different minor allele frequency in developmentally delayed subjects as compared to controls (Table II). However, because the ethnicity of most of the patients in whom these variants were detected is unknown, it is possible that these associations result from population substructure.

DISCUSSION

Through the use of pooled-template MPS, we have identified 130 novel sequence variants in FMR1 in a deidentified population of 963 developmentally delayed males. While many of these are likely novel polymorphisms, several show evidence of functional effects and association with developmental delay. These findings have important implications for the diagnosis of developmental delay, the structure and function of FMR1 and FMRP, and the utility of pooled-template MPS for novel variant detection in a disease gene.

One of the most noteworthy findings in the sequenced clinical population was the identification of the novel missense change p.R138Q. If proven to be pathogenic, this variant would be only the second missense mutation to be identified in FMR1. However, the functional significance of p.R138Q remains unclear. The arginine-138 residue is largely conserved through the pufferfish species Takifugu rubripes and Tetraodon nigroviridis but both use a glutamine at this position. Furthermore, algorithms for the prediction of the effects of amino acid substitution give conflicting results for this change, as PMut and PANTHER PSEC predict that it is pathological, while SIFT and PolyPhen predict that it is benign. Finally, our initial studies of the effects of the p.R138Q substitution on NLS function have not revealed any change in the intracellular localization of FMRP (data not shown). However, the function of the FMRP NLS has historically been difficult to assess in its endogenous context [Eberhart et al., 1996; Fridell et al., 1996; Sittler et al., 1996; Bardoni et al., 1997; Kim et al.,...
be considered for diagnostic testing. Therefore, if the novel sequence variants are indeed functional, such variants should be viewed as a significant contributor to the heterogeneous diagnosis of developmental delay, and FMR1 sequencing should be considered for diagnostic testing.

To our knowledge, this study represents the first application of pooled-template MPS for the identification of novel sequence variants in a clinical population. While several proof-of-principle papers had demonstrated that this approach was adequately sensitive and unbiased for the detection of rare novel variants [Druley et al., 2009; Ingman and Gyllensten, 2009; Out et al., 2009], the current study validates pooled-template MPS as a useful application of next-generation sequencing technologies for targeted studies of a single gene. With the continual market-driven increase in MPS capacity, pooled-template approaches will become even more important for the efficient use of MPS on single genes and other small genomic regions.

In summary, we have identified 130 novel sequence variants in FMR1 in a population of 963 developmentally delayed males. Among these variants are the novel missense change p.R138Q, which alters a conserved residue in the FMRP NLS, the first three sequence variants to be identified in the FMR1 promoter, all of which reduce transcriptional activity, and several noncoding variants of possible functional effect, including two splice site mutations. Taken together, these results provide avenues for structure–function studies of FMRP and FMR1, and suggest that pathogenic sequence variants in FMR1 may represent a significant cause of developmental delay.

ACKNOWLEDGMENTS

We are grateful to the patients and families for their invaluable contributions. We thank Ephrem Chin and the Emory Genetics Laboratory for assistance with sample collection; Viren Patel and Amol Shetty for assistance with GA sequence analysis; Daniel Reines for his gift of the pFMR1-luc plasmid; Kerry Ressler for the use of his Sequenom MassARRAY Compact Analyzer; Kimberly De-

<table>
<thead>
<tr>
<th>Variant</th>
<th>Location</th>
<th>PhastCons</th>
<th>PhyloP</th>
<th>Patient frequency</th>
<th>Control frequency</th>
<th>P (Fisher's exact test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Missense variant</td>
<td>c.413G&gt;A</td>
<td>Exon 5 [p.R138Q]</td>
<td>0.99</td>
<td>2.94</td>
<td>1/963 (0.1%)</td>
<td>0/1,385 (&lt;0.07%)</td>
</tr>
<tr>
<td>Promoter variants</td>
<td>c.—322G&gt;C</td>
<td>GC Box/AP-2</td>
<td>0.72</td>
<td>0.65</td>
<td>1/963 (0.1%)</td>
<td>0/1,308 (&lt;0.08%)</td>
</tr>
<tr>
<td></td>
<td>c.—293T&gt;C</td>
<td>Inr-like/Tx. Start II/TATA-like</td>
<td>0.94</td>
<td>1.95</td>
<td>1/963 (0.1%)</td>
<td>0/1,266 (&lt;0.08%)</td>
</tr>
<tr>
<td></td>
<td>c.—254A&gt;G</td>
<td>Inr-like/Tx. Start I</td>
<td>1</td>
<td>2.10</td>
<td>1/963 (0.1%)</td>
<td>0/1,304 (&lt;0.08%)</td>
</tr>
<tr>
<td>Other noncoding variants</td>
<td>c.105 — 8A&gt;C</td>
<td>Intron 2</td>
<td>1</td>
<td>2.48</td>
<td>6/963 (0.6%)</td>
<td>0/1,262 (&lt;0.08%)</td>
</tr>
<tr>
<td></td>
<td>c.630 — 438A&gt;C</td>
<td>Intron 7</td>
<td>1</td>
<td>2.55</td>
<td>1/963 (0.1%)</td>
<td>0/1,263 (&lt;0.08%)</td>
</tr>
<tr>
<td></td>
<td>c.631 — 840G&gt;A</td>
<td>Intron 7</td>
<td>1</td>
<td>1.76</td>
<td>1/963 (0.1%)</td>
<td>0/1,239 (&lt;0.08%)</td>
</tr>
<tr>
<td></td>
<td>c.880 — 885A&gt;G</td>
<td>Intron 9a</td>
<td>0</td>
<td>0.73</td>
<td>1/963 (0.1%)</td>
<td>0/1,084 (&lt;0.09%)</td>
</tr>
<tr>
<td></td>
<td>c.990 — 4T&gt;C</td>
<td>Intron 10</td>
<td>1</td>
<td>2.46</td>
<td>1/963 (0.1%)</td>
<td>0/1,248 (&lt;0.08%)</td>
</tr>
<tr>
<td></td>
<td>c.1472 — 521C&gt;G</td>
<td>Intron 14a</td>
<td>0</td>
<td>0.40</td>
<td>1/963 (0.1%)</td>
<td>0/1,254 (&lt;0.08%)</td>
</tr>
<tr>
<td></td>
<td>c.*23T&gt;C</td>
<td>3’UTR</td>
<td>0.99</td>
<td>2.40</td>
<td>1/963 (0.1%)</td>
<td>0/900 (&lt;0.11%)</td>
</tr>
<tr>
<td></td>
<td>c.*746T&gt;C</td>
<td>3’UTRb</td>
<td>1</td>
<td>2.22</td>
<td>6/963 (0.6%)</td>
<td>0/1,260 (&lt;0.08%)</td>
</tr>
<tr>
<td></td>
<td>c.*1867G&gt;A</td>
<td>3’UTR</td>
<td>1</td>
<td>1.54</td>
<td>12/963 (1.2%)</td>
<td>0/951 (&lt;0.11%)</td>
</tr>
<tr>
<td></td>
<td>c.*2035C&gt;T</td>
<td>3’UTR</td>
<td>1</td>
<td>2.47</td>
<td>3/963 (0.3%)</td>
<td>0/1,270 (&lt;0.08%)</td>
</tr>
</tbody>
</table>

N.S. indicates P value is not significant.
aPredicted splice site.
bPredicted miRNA binding site.
ment, Ann Dodd, Krayton Keith, Brian Lynch, Tamika Malone, Kristie Mercer, and Julie Mowrey for their assistance with genotyping; and the members of the Warren laboratory for insightful discussion. We acknowledge the assistance of the Emory Genomics Center and its support through the Georgia Research Alliance and the Atlanta Clinical & Translational Science Institute (UL1 RR025008). This research was supported by National Institutes of Health Grants AG029749 (to S.C.C.) and HD020521 and HD024064 (to S.T.W.) and a FRAXA Foundation Fellowship (to J.A.S.).

REFERENCES


