Published in final edited form as:

Am J Med Genet A. 2010 October ; 152A(10): 2512–2520. doi:10.1002/ajmg.a.33626.

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

Stephen C. Collins¹, Steven M. Bray¹, Joshua A. Suhl¹, David J. Cutler¹, Bradford Coffee¹, Michael E. Zwick¹, and Stephen T. Warren^{1,2,*}

¹Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA

²Departments of Biochemistry and Pediatrics, Emory University School of Medicine, Atlanta, GA 30322, USA

Abstract

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 catalogue of known *FMR1* sequence variants and suggest that *FMR1* sequence variants may represent an important cause of developmental delay.

Keywords

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 of the *FMR1* gene (MIM 309550) [Ashley et al., 1993; Verkerk et al., 1991]. 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

*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, Telephone: 404-727-5624, Fax: 404-727-3949, swarren@emory.edu.

gene product, FMRP [Chiurazzi et al., 1998; Sutcliffe et al., 1992]. *FMRI* 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 *FMRI* was identified in 1991 [De Boulle et al., 1993; Lugenbeel et al., 1995]. Therefore, *FMRI* 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 *FMRI* sequence variation in a large number of patients, leaving the true significance of pathogenic sequence variants in *FMRI* unknown [Castellvi-Bel et al., 1999; Chiurazzi et al., 1994; Collins et al., 2010; Gronskov et al., 1998; Reyniers et al., 1996; Shinahara et al., 2004; Vincent et al., 1996; Wang et al., 1997].

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 Gyllenstein, 2009; Koboldt et al., 2009; Out et al., 2009].

To detect potentially pathogenic *FMRI* sequence variants, we employed pooled-template MPS to assess the promoter, all 17 exons, and a substantial portion of the intronic sequence of *FMRI* in 963 developmentally delayed males referred for *FMRI* 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 of FMRP. Furthermore, we report three novel promoter variants, all of which reduce the *in vitro* expression of *FMRI*, and several novel sequence variants in conserved noncoding regions of *FMRI*, including two predicted splice site mutations. Together, these novel variants suggest that there may be clinical utility in diagnostic *FMRI* sequencing for developmentally delayed males.

MATERIALS AND METHODS

Clinical Population

While loss of function mutations in *FMRI* 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 *FMRI* in patients who had tested negative for *FMRI* repeat expansion at the Emory Genetics Laboratory over a five year span. Because the current standard of care is for all children presenting with developmental delay to be tested for *FMRI* 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 *FMRI*. 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 *FMRI* 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, 1392 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 supporting information Figure 1, four long range PCR (LR-PCR) amplifications were designed to target *FMR1* (supporting information Figure 1 can be found in the online version of this article). The LR-PCR primer pairs were as follows: *FMR1A*-F: 5'-CAGACTGCGCTACTTTGAACC-3' and *FMR1A*-R: 5'-CTACATACCAACAAACGCACTACTGCTACAT-3'; *FMR1B*-F: 5'-AATTTCCAGTATACTTGTCTATTTTTCGAGATG-3' and *FMR1B*-R: 5'-TTTTGGGAGATAGCTACCTACAGGGTATCTGATT-3'; *FMR1C*-F: 5'-GTTGAACATTAAATTGCAGTTCAGAATACATAG-3' and *FMR1C*-R: 5'-GAGACATATCCAATCCACTTGCCGTTATAGT-3'; *FMR1D*-F: 5'-AATAATCTGATACGTTTAAAGGTTGCTATTGA-3' and *FMR1D*-R: 5'-TTAATATGGTTTAGTGGCACCCCTATGTAATAAA-3'. Each LR-PCR-A reaction contained 50 ng of genomic DNA, 100 ng of each primer, 5 µl of dNTPs (Takara Bio Inc., Otsu, Shiga, Japan), 12.5 µl of 2x GC Buffer II (Takara), and 0.5 µl of Ex Taq (Takara), in a total of 25 µl. The following PCR conditions were used for LR-PCR-A: initialization at 95°C for 4 minutes; 37 cycles of denaturation at 95°C for 30 seconds and annealing/elongation at 60°C for 4 minutes; and a final elongation step of 72°C for 9 minutes. Each LR-PCR-B, -C, and -D reaction contained 50 ng of genomic DNA, 100 ng of each primer, 4 µl of dNTPs (Takara Bio Inc., Otsu, Shiga, Japan), 2.5 µl of Ex Taq Buffer (Takara), and 0.4 µl of Ex Taq (Takara), in a total of 25 µl. The following PCR conditions were used for LR-PCR-B: initialization at 94°C for 4 minutes; 30 cycles of denaturation at 94°C for 20 seconds and annealing/elongation at 64°C for 8 minutes; and a final elongation step of 68°C for 13 minutes. The same conditions were used for LR-PCR-C, but 35 cycles of denaturation and annealing/elongation were used instead of 30. The same conditions used for LR-PCR-C were used for LR-PCR-D, but the annealing/elongation at 64°C was continued for 9 minutes instead of 8 minutes. The expected sizes of the LR-PCR amplicons were visually confirmed with gel electrophoresis on the E-Gel system (Invitrogen, Carlsbad, CA). Failed LR-PCRs were repeated one time before the sample was excluded from the study. Since clinical testing for the full mutation would have revealed large deletions of *FMR1*, the most common cause of repeated failure of amplification was DNA degradation. After LR-PCR, 963 samples remained for sequencing.

LR-PCR Amplicon Pooling—The concentrations of LR-PCR amplicons were measured by fluorometric quantification with PicoGreen dsDNA reagent (Molecular Probes, Eugene, OR). Equimolar 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 Boulle et al., 1993]. This

I304N 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, Carlsbad, CA). 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 μ g.

Target Library Construction—Amplicon pools were fragmented by sonication (Sonicator S-4000, Misonix, Farmingdale, NY) with the following parameters: 8 pulses of 30 seconds each, with 2 minutes rest between pulses, at a power output of 20%. The fragmentation range, as visualized by gel electrophoresis, was from 100 – 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 *FMRI* region of the human genome reference (NT_011681.15, g.3435700 to 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 rs68020458. 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 greater than or equal to 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 $p < 1 \times 10^{-4}$.

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 (Supporting information Table I, available 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 SacII, 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 phastCons and the basewise conservation across placental mammals was assessed by phyloP; these values were obtained from the UCSC Genome Browser [Rhead et al., 2009; Siepel et al., 2005]. 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 [Betel et al., 2008; John et al., 2004].

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 TOPO TA Cloning kit (Invitrogen, Carlsbad, CA). 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 6-well cell culture dishes. Transfections were carried out in Opti-MEM (Invitrogen, Carlsbad, CA), using 10 µl of Lipofectamine 2000 (Invitrogen, Carlsbad, CA) 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 μ l 1x Passive Lysis Buffer (Promega, Madison, WI) by rocking at room temperature for 15 minutes. Lysates were cleared of cell debris by centrifugation at 14000 rpm for 5 minutes at 4°C.

Luciferase Assays—Protein concentrations of the lysates were measured by the Bradford assay. The Dual-Luciferase Reporter Assay System (Promega, Madison, WI) was used to measure luciferase activity. From each lysate, 5 μ g of protein was added in 20 μ l total volume to a luminometer tube. To each luminometer tube, 100 μ l of LAR II was added. A manual-load luminometer was used to measure the luminescence over a 10 second period, following a 2 second premeasurement delay. The luminometer measurement was repeated after the addition of 100 μ l 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 catalogued 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 1401 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 and colleagues 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 GGC; 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 catalogued in dbSNP, nor were they detected in the first nine publicly available personal genomes [Ahn et al., 2009; Bentley et al., 2008; Levy et al., 2007; Siva, 2008; Wang et al., 2008; Wheeler et al., 2008]. Among these variants, 63.1% were only detected in one individual, while 36.9% were detected in multiple individuals. 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 supporting information Table II, available in the online version of this article.

Although pooled-template MPS is a highly sensitive method [Druley et al., 2009; Ingman and Gyllenstein, 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 $0.235 \times 130 / (1 - 0.235) = 39.9$.] However, it would be expected that we missed no more than one true promoter variant [$0.235 \times 3 / 0.765 = 0.92$] and no more than one true missense variant [$0.235 \times 1 / 0.765 = 0.31$], 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 (Figure 1A), was identified in a sample from a patient of European descent, but was not detected in 1385 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 nuclear localization signal (NLS) of FMRP [Bardoni et al., 1997; Eberhart et al., 1996] and is highly conserved through *Drosophila* (Figure 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* (Figures 2A, B), each occurring in only one of the 963 sequenced developmentally delayed males. The c.-332G>C variant is located within overlapping binding sites for the Sp1 and AP-2 α transcription factors [Lim et al., 2005; Smith et al., 2004]. The c.-293T>C variant is located near transcription start site II, within both an Initiator-like (Inr-like) and a TATA-like sequence [Beilina et al., 2004; Hwu et al., 1993]. The third variant, c.-254A>G, is located within an Inr-like sequence near the primary transcription start site [Beilina et al., 2004; Kumari and Usdin, 2001]. All three of these variant bases are conserved in mammals (Figure 2C).

The c.-332G>C, c.-293T>C, and c.-254A>G variants were not detected in 1308, 1266, and 1304 control males of European descent, respectively. However, the ethnicity of the patients in whom these variants were detected is unknown. Thus, it remains possible that the suggested association of these promoter variants with developmental delay results from population substructure.

Effects on Promoter Activity—To assess the functional significance of the three promoter variants, we introduced each variant individually into the previously described pFMR1-luc plasmid, which uses the human *FMR1* promoter to drive expression of firefly luciferase [Smith et al., 2004]. Each resulting plasmid was co-transfected with the control plasmid pRL-TK into HeLa cells. Firefly luciferase activity, measured 48 hours post-transfection, was normalized to *Renilla* luciferase activity and expressed relative to the wild-type *FMR1* promoter activity. As seen in Figure 3, the c.-332G>C variant reduced *FMR1* promoter activity to 5.9% of wild-type levels, the c.-293T>C variant reduced *FMR1*

promoter activity to 29.2% of wild-type levels, and the c.-254A>G variant reduced *FMRI* promoter activity to 36.2% of wild-type levels. All of these reductions from wild-type promoter activity were statistically significant ($p < 1 \times 10^{-7}$).

Noncoding Variants in *FMRI*

To determine if any of the 127 novel noncoding variants in *FMRI* 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 [Betel et al., 2008; John et al., 2004] 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 *FMRI*. 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 *FMRI* 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 *FMRI* 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 *FMRI*. However, the functional significance of p.R138Q remains unclear. The arginine-138 residue is largely conserved through *Drosophila*, but the pufferfish species *Takifugu rubripes* and *Tetraodon nigroviridis* 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 [Bardoni et al., 1997; Eberhart et al., 1996; Fridell et al., 1996; Kim et al., 2009; Sittler et al., 1996]. Moreover, arginine-138 also lies within the N-terminus of FMRP, which has been implicated in RNA binding, domain dimerization and protein-protein interactions [Adinolfi et al., 2003]. Therefore, more thorough analyses of FMRP function will need to be performed to determine if the p.R138Q missense change has a deleterious effect.

Another important finding in this clinical population was the identification of three novel variants in the *FMR1* promoter. To our knowledge, these are the first three variants to be identified in defined elements of the *FMR1* promoter. Furthermore, all three independently reduced the activity of the *FMR1* promoter in a luciferase reporter assay. While the ultimate measure of the effects of *FMR1* promoter variants is the associated level of FMRP expression in patient tissues, this reporter assay data suggests that the three identified variants are likely to have a functional effect.

While the functional effects of the novel *FMR1* variants identified in this study remain unclear, it is likely that several do indeed associate with developmental delay, such as the missense variant, the three promoter variants and the two splice site mutations. If these variants are all verified to be pathogenic, the frequency of *FMR1* sequence variants causing developmental delay in our study population would be 0.8%, when accounting for our false negative rate. In comparison, the frequency of *FMR1* repeat expansions, the most common inherited cause of developmental delay, is 2 – 3% in the same population [Biancalana et al., 2004; Hecimovic et al., 2002; Major et al., 2003; Pandey et al., 2002; Patsalis et al., 1999; Rauch et al., 2006]. Therefore, if the novel *FMR1* 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 Gyllenstein, 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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 Dement, 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

- Adinolfi S, Ramos A, Martin SR, Dal Piaz F, Pucci P, Bardoni B, Mandel JL, Pastore A. The N-terminus of the fragile X mental retardation protein contains a novel domain involved in dimerization and RNA binding. *Biochemistry*. 2003; 42:10437–10444. [PubMed: 12950170]
- Ahn SM, Kim TH, Lee S, Kim D, Ghang H, Kim DS, Kim BC, Kim SY, Kim WY, Kim C, Park D, Lee YS, Kim S, Reja R, Jho S, Kim CG, Cha JY, Kim KH, Lee B, Bhak J, Kim SJ. The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res*. 2009; 19:1622–1629. [PubMed: 19470904]
- Ashley CT, Sutcliffe JS, Kunst CB, Leiner HA, Eichler EE, Nelson DL, Warren ST. Human and murine FMR-1: alternative splicing and translational initiation downstream of the CGG-repeat. *Nat Genet*. 1993; 4:244–251. [PubMed: 8358432]
- Bardoni B, Schenck A, Mandel JL. A novel RNA-binding nuclear protein that interacts with the fragile X mental retardation (FMR1) protein. *Hum Mol Genet*. 1999; 8:2557–2566. [PubMed: 10556305]
- Bardoni B, Sittler A, Shen Y, Mandel JL. Analysis of domains affecting intracellular localization of the FMRP protein. *Neurobiol Dis*. 1997; 4:329–336. [PubMed: 9440121]
- Beilina A, Tassone F, Schwartz PH, Sahota P, Hagerman PJ. Redistribution of transcription start sites within the FMR1 promoter region with expansion of the downstream CGG-repeat element. *Hum Mol Genet*. 2004; 13:543–549. [PubMed: 14722156]
- Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, Hall KP, Evers DJ, Barnes CL, Bignell HR, Boutell JM, Bryant J, Carter RJ, Keira Cheetham R, Cox AJ, Ellis DJ, Flatbush MR, Gormley NA, Humphray SJ, Irving LJ, Karbelashvili MS, Kirk SM, Li H, Liu X, Maisinger KS, Murray LJ, Obradovic B, Ost T, Parkinson ML, Pratt MR, Rasolonjatovo IM, Reed MT, Rigatti R, Rodighiero C, Ross MT, Sabot A, Sankar SV, Scally A, Schroth GP, Smith ME, Smith VP, Spiridou A, Torrance PE, Tzonev SS, Vermaas EH, Walter K, Wu X, Zhang L, Alam MD, Anastasi C, Aniebo IC, Bailey DM, Bancarz IR, Banerjee S, Barbour SG, Baybayan PA, Benoit VA, Benson KF, Bevis C, Black PJ, Boodhun A, Brennan JS, Bridgham JA, Brown RC, Brown AA, Buermann DH, Bundu AA, Burrows JC, Carter NP, Castillo N, Chiara ECM, Chang S, Neil Cooley R, Crake NR, Dada OO, Diakoumakos KD, Dominguez-Fernandez B, Earnshaw DJ, Egbujor UC, Elmore DW, Etchin SS, Ewan MR, Fedurco M, Fraser LJ, Fuentes Fajardo KV, Scott Furey W, George D, Gietzen KJ, Goddard CP, Golda GS, Granieri PA, Green DE, Gustafson DL, Hansen NF, Harnish K, Haudenschild CD, Heyer NI, Hims MM, Ho JT, Horgan AM, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008; 456:53–59. [PubMed: 18987734]
- Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. *Nucleic Acids Res*. 2008; 36:D149–D153. [PubMed: 18158296]
- Biancalana V, Beldjord C, Taillandier A, Szpiro-Tapia S, Cusin V, Gerson F, Philippe C, Mandel JL. Five years of molecular diagnosis of Fragile X syndrome (1997–2001): a collaborative study reporting 95% of the activity in France. *Am J Med Genet A*. 2004; 129A:218–224. [PubMed: 15326620]
- Castellvi-Bel S, Sanchez A, Badenas C, Mallolas J, Barcelo A, Jimenez D, Villa M, Estivill X, Mila M. Single-strand conformation polymorphism analysis in the FMR1 gene. *Am J Med Genet*. 1999; 84:262–265. [PubMed: 10331603]
- Chiurazzi P, de Graaff E, Ng J, Verkerk AJ, Wolfson S, Fisch GS, Kozak L, Neri G, Oostra BA. No apparent involvement of the FMR1 gene in five patients with phenotypic manifestations of the fragile X syndrome. *Am J Med Genet*. 1994; 51:309–314. [PubMed: 7942992]
- Chiurazzi P, Pomponi MG, Willemsen R, Oostra BA, Neri G. In vitro reactivation of the FMR1 gene involved in fragile X syndrome. *Hum Mol Genet*. 1998; 7:109–113. [PubMed: 9384610]
- Coffee B, Ikeda M, Budimirovic DB, Hjelm LN, Kaufmann WE, Warren ST. Mosaic FMR1 deletion causes fragile X syndrome and can lead to molecular misdiagnosis: a case report and review of the literature. *Am J Med Genet A*. 2008; 146A:1358–1367. [PubMed: 18412117]
- Collins SC, Coffee B, Benke PJ, Berry-Kravis E, Gilbert F, Oostra B, Halley D, Zwick ME, Cutler DJ, Warren ST. Array-based FMR1 sequencing and deletion analysis in patients with a fragile X syndrome-like phenotype. *PLoS One*. 2010; 5:e9476. [PubMed: 20221430]

- Dahl F, Stenberg J, Fredriksson S, Welch K, Zhang M, Nilsson M, Bicknell D, Bodmer WF, Davis RW, Ji H. Multigene amplification and massively parallel sequencing for cancer mutation discovery. *Proc Natl Acad Sci U S A*. 2007; 104:9387–9392. [PubMed: 17517648]
- Darnell JC, Fraser CE, Mostovetsky O, Stefani G, Jones TA, Eddy SR, Darnell RB. Kissing complex RNAs mediate interaction between the Fragile-X mental retardation protein KH2 domain and brain polyribosomes. *Genes Dev*. 2005; 19:903–918. [PubMed: 15805463]
- De Boule K, Verkerk AJ, Reyniers E, Vits L, Hendrickx J, Van Roy B, Van den Bos F, de Graaff E, Oostra BA, Willems PJ. A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nat Genet*. 1993; 3:31–35. [PubMed: 8490650]
- Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res*. 2008; 36:e105. [PubMed: 18660515]
- Druley TE, Vallania FL, Wegner DJ, Varley KE, Knowles OL, Bonds JA, Robison SW, Doniger SW, Hamvas A, Cole FS, Fay JC, Mitra RD. Quantification of rare allelic variants from pooled genomic DNA. *Nat Methods*. 2009; 6:263–265. [PubMed: 19252504]
- Eberhart DE, Malter HE, Feng Y, Warren ST. The fragile X mental retardation protein is a ribonucleoprotein containing both nuclear localization and nuclear export signals. *Hum Mol Genet*. 1996; 5:1083–1091. [PubMed: 8842725]
- Feng Y, Absher D, Eberhart DE, Brown V, Malter HE, Warren ST. FMRP associates with polyribosomes as an mRNP, and the I304N mutation of severe fragile X syndrome abolishes this association. *Mol Cell*. 1997a; 1:109–118. [PubMed: 9659908]
- Feng Y, Gutekunst CA, Eberhart DE, Yi H, Warren ST, Hersch SM. Fragile X mental retardation protein: nucleocytoplasmic shuttling and association with somatodendritic ribosomes. *J Neurosci*. 1997b; 17:1539–1547. [PubMed: 9030614]
- Ferrer-Costa C, Gelpi JL, Zamakola L, Parraga I, de la Cruz X, Orozco M. PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics*. 2005; 21:3176–3178. [PubMed: 15879453]
- Fridell RA, Benson RE, Hua J, Bogerd HP, Cullen BR. A nuclear role for the Fragile X mental retardation protein. *EMBO J*. 1996; 15:5408–5414. [PubMed: 8895584]
- Garber KB, Visootsak J, Warren ST. Fragile X syndrome. *Eur J Hum Genet*. 2008; 16:666–672. [PubMed: 18398441]
- Gronskov K, Hallberg A, Brondum-Nielsen K. Mutational analysis of the FMR1 gene in 118 mentally retarded males suspected of fragile X syndrome: absence of prevalent mutations. *Hum Genet*. 1998; 102:440–445. [PubMed: 9600241]
- Harismendy O, Ng PC, Strausberg RL, Wang X, Stockwell TB, Beeson KY, Schork NJ, Murray SS, Topol EJ, Levy S, Frazer KA. Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol*. 2009; 10:R32. [PubMed: 19327155]
- Hecimovic S, Tarnik IP, Baric I, Cakarun Z, Pavelic K. Screening for fragile X syndrome: results from a school for mentally retarded children. *Acta Paediatr*. 2002; 91:535–539. [PubMed: 12113322]
- Hwu WL, Lee YM, Lee SC, Wang TR. In vitro DNA methylation inhibits FMR-1 promoter. *Biochem Biophys Res Commun*. 1993; 193:324–329. [PubMed: 8503923]
- Ingman M, Gyllensten U. SNP frequency estimation using massively parallel sequencing of pooled DNA. *Eur J Hum Genet*. 2009; 17:383–386. [PubMed: 18854868]
- John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. Human MicroRNA targets. *PLoS Biol*. 2004; 2:e363. [PubMed: 15502875]
- Kim M, Bellini M, Ceman S. Fragile X mental retardation protein FMRP binds mRNAs in the nucleus. *Mol Cell Biol*. 2009; 29:214–228. [PubMed: 18936162]
- Koboldt DC, Chen K, Wylie T, Larson DE, McLellan MD, Mardis ER, Weinstock GM, Wilson RK, Ding L. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics*. 2009; 25:2283–2285. [PubMed: 19542151]
- Kumari D, Gabrielian A, Wheeler D, Usdin K. The roles of Sp1, Sp3, USF1/USF2 and NRF-1 in the regulation and three-dimensional structure of the Fragile X mental retardation gene promoter. *Biochem J*. 2005; 386:297–303. [PubMed: 15479157]

- Kumari D, Usdin K. Interaction of the transcription factors USF1, USF2, and alpha -Pal/Nrf-1 with the FMR1 promoter. Implications for Fragile X mental retardation syndrome. *J Biol Chem*. 2001; 276:4357–4364. [PubMed: 11058604]
- Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, Axelrod N, Huang J, Kirkness EF, Denisov G, Lin Y, MacDonald JR, Pang AW, Shago M, Stockwell TB, Tsiamouri A, Bafna V, Bansal V, Kravitz SA, Busam DA, Beeson KY, McIntosh TC, Remington KA, Abril JF, Gill J, Borman J, Rogers YH, Frazier ME, Scherer SW, Strausberg RL, Venter JC. The diploid genome sequence of an individual human. *PLoS Biol*. 2007; 5:e254. [PubMed: 17803354]
- Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res*. 2008; 18:1851–1858. [PubMed: 18714091]
- Lim JH, Booker AB, Luo T, Williams T, Furuta Y, Lagutin O, Oliver G, Sargent TD, Fallon JR. AP-2alpha selectively regulates fragile X mental retardation-1 gene transcription during embryonic development. *Hum Mol Genet*. 2005; 14:2027–2034. [PubMed: 15930016]
- Lugenbeel KA, Peier AM, Carson NL, Chudley AE, Nelson DL. Intragenic loss of function mutations demonstrate the primary role of FMR1 in fragile X syndrome. *Nat Genet*. 1995; 10:483–485. [PubMed: 7670500]
- Major T, Culjkovic B, Stojkovic O, Gucsekic M, Lakic A, Romac S. Prevalence of the fragile X syndrome in Yugoslav patients with non-specific mental retardation. *J Neurogenet*. 2003; 17:223–230. [PubMed: 14668200]
- Mila M, Castellvi-Bel S, Sanchez A, Barcelo A, Badenas C, Mallolas J, Estivill X. Rare variants in the promoter of the fragile X syndrome gene (FMR1). *Mol Cell Probes*. 2000; 14:115–119. [PubMed: 10799273]
- Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003; 31:3812–3814. [PubMed: 12824425]
- Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, Bamshad M, Nickerson DA, Shendure J. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*. 2009; 461:272–276. [PubMed: 19684571]
- Okou DT, Locke AE, Steinberg KM, Hagen K, Athri P, Shetty AC, Patel V, Zwick ME. Combining microarray-based genomic selection (MGS) with the Illumina Genome Analyzer platform to sequence diploid target regions. *Ann Hum Genet*. 2009; 73:502–513. [PubMed: 19573206]
- Out AA, van Minderhout IJ, Goeman JJ, Ariyurek Y, Ossowski S, Schneeberger K, Weigel D, van Galen M, Taschner PE, Tops CM, Breuning MH, van Ommen GJ, den Dunnen JT, Devilee P, Hes FJ. Deep sequencing to reveal new variants in pooled DNA samples. *Hum Mutat*. 2009
- Pandey UB, Phadke S, Mittal B. Molecular screening of FRAXA and FRAXE in Indian patients with unexplained mental retardation. *Genet Test*. 2002; 6:335–339. [PubMed: 12537661]
- Patsalis PC, Sismani C, Hettinger JA, Boumba I, Georgiou I, Stylianidou G, Anastasiadou V, Koukoulli R, Pagoulatos G, Syrrou M. Molecular screening of fragile X (FRAXA) and FRAXE mental retardation syndromes in the Hellenic population of Greece and Cyprus: incidence, genetic variation, and stability. *Am J Med Genet*. 1999; 84:184–190. [PubMed: 10331587]
- Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res*. 2002; 30:3894–3900. [PubMed: 12202775]
- Rauch A, Hoyer J, Guth S, Zweier C, Kraus C, Becker C, Zenker M, Huffmeier U, Thiel C, Ruschendorf F, Nurnberg P, Reis A, Trautmann U. Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *Am J Med Genet A*. 2006; 140:2063–2074. [PubMed: 16917849]
- Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *J Comput Biol*. 1997; 4:311–323. [PubMed: 9278062]
- Reyniers E, Wolff G, Tariverdian G, De Boulle K, Storm K, Kooy RF, Willems PJ. Severe mental retardation and macroorchidism without mutation in the FMR1 gene. *Am J Med Genet*. 1996; 64:408–412. [PubMed: 8844093]
- Rhead B, Karolchik D, Kuhn RM, Hinrichs AS, Zweig AS, Fujita PA, Diekhans M, Smith KE, Rosenbloom KR, Raney BJ, Pohl A, Pheasant M, Meyer LR, Learned K, Hsu F, Hillman-Jackson

- J, Harte RA, Giardine B, Dreszer TR, Clawson H, Barber GP, Haussler D, Kent WJ. The UCSC genome browser database: update 2010. *Nucleic Acids Res.* 2009
- Schrier M, Severijnen LA, Reis S, Rife M, van't Padje S, van Cappellen G, Oostra BA, Willemsen R. Transport kinetics of FMRP containing the I304N mutation of severe fragile X syndrome in neurites of living rat PC12 cells. *Exp Neurol.* 2004; 189:343–353. [PubMed: 15380484]
- Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol.* 2008; 26:1135–1145. [PubMed: 18846087]
- Shinahara K, Saijo T, Mori K, Kuroda Y. Single-strand conformation polymorphism analysis of the FMR1 gene in autistic and mentally retarded children in Japan. *J Med Invest.* 2004; 51:52–58. [PubMed: 15000256]
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, Haussler D. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 2005; 15:1034–1050. [PubMed: 16024819]
- Sittler A, Devys D, Weber C, Mandel JL. Alternative splicing of exon 14 determines nuclear or cytoplasmic localisation of fmr1 protein isoforms. *Hum Mol Genet.* 1996; 5:95–102. [PubMed: 8789445]
- Siva N. 1000 Genomes project. *Nat Biotechnol.* 2008; 26:256. [PubMed: 18327223]
- Smith KT, Coffee B, Reines D. Occupancy and synergistic activation of the FMR1 promoter by Nrf-1 and Sp1 in vivo. *Hum Mol Genet.* 2004; 13:1611–1621. [PubMed: 15175277]
- Sutcliffe JS, Nelson DL, Zhang F, Pieretti M, Caskey CT, Saxe D, Warren ST. DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum Mol Genet.* 1992; 1:397–400. [PubMed: 1301913]
- Thomas PD, Kejariwal A. Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc Natl Acad Sci U S A.* 2004; 101:15398–15403. [PubMed: 15492219]
- Verkerk AJ, Pieretti M, Sutcliffe JS, Fu YH, Kuhl DP, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang F, Eussen BE, Van Ommen GJ, Blonden LA, Riggins GJ, Chastain JL, Kunst CB, Galjaard H, Caskey CT, Nelson DL, Oostra BA, Warren ST. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell.* 1991; 65:905–914. [PubMed: 1710175]
- Vincent JB, Konecki DS, Munstermann E, Bolton P, Poustka A, Poustka F, Gurling HM. Point mutation analysis of the FMR-1 gene in autism. *Mol Psychiatry.* 1996; 1:227–231. [PubMed: 9118347]
- Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, Fan W, Zhang J, Li J, Guo Y, Feng B, Li H, Lu Y, Fang X, Liang H, Du Z, Li D, Zhao Y, Hu Y, Yang Z, Zheng H, Hellmann I, Inouye M, Pool J, Yi X, Zhao J, Duan J, Zhou Y, Qin J, Ma L, Li G, Zhang G, Yang B, Yu C, Liang F, Li W, Li S, Ni P, Ruan J, Li Q, Zhu H, Liu D, Lu Z, Li N, Guo G, Ye J, Fang L, Hao Q, Chen Q, Liang Y, Su Y, San A, Ping C, Yang S, Chen F, Li L, Zhou K, Ren Y, Yang L, Gao Y, Yang G, Li Z, Feng X, Kristiansen K, Wong GK, Nielsen R, Durbin R, Bolund L, Zhang X, Yang H. The diploid genome sequence of an Asian individual. *Nature.* 2008; 456:60–65. [PubMed: 18987735]
- Wang YC, Lin ML, Lin SJ, Li YC, Li SY. Novel point mutation within intron 10 of FMR-1 gene causing fragile X syndrome. *Hum Mutat.* 1997; 10:393–399. [PubMed: 9375856]
- Wheeler DA, Srinivasan M, Egholm M, Shen Y, Chen L, McGuire A, He W, Chen YJ, Makhijani V, Roth GT, Gomes X, Tartaro K, Niazi F, Turcotte CL, Irzyk GP, Lupski JR, Chinault C, Song XZ, Liu Y, Yuan Y, Nazareth L, Qin X, Muzny DM, Margulies M, Weinstock GM, Gibbs RA, Rothberg JM. The complete genome of an individual by massively parallel DNA sequencing. *Nature.* 2008; 452:872–876. [PubMed: 18421352]

Figure 1. The novel missense variant p.R138Q
(A) DNA chromatograms of the wild-type and variant alleles. (B) Diagram of FMRP, depicting the functional domains (NLS = nuclear localization signal; KH = K homology domain; NES = nuclear export signal) and R138Q missense substitution. The amino acid sequence alignment corresponds to the NLS. Functional basic residues are shown in red. Asterisks denote identical residues, colons denote conserved substitutions, and periods denote semi-conserved substitutions.

Figure 2. The novel promoter variants c.-332G>C, c.-293T>C, and c.-254A>G
(A) Diagram of the minimal promoter and 5'UTR of *FMR1*. Roman numerals I–III represent the three transcription start sites of *FMR1*. The GC boxes bind the transcription factor Sp1. (B) 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.

Figure 3. Functional effects of the novel promoter variants c.-332G>C, c.-293T>C, and c.-254A>G

Luciferase activity is depicted as the ratio of firefly luciferase to *Renilla* luciferase and normalized to the activity of the wild-type *FMR1* promoter. The results shown represent six independent transfections. Data are represented as mean $\pm$ 1 S.D.

Table I*FMRI* Sequence Variants Detected in 963 Developmentally Delayed Males

	Novel variants		Known Polymorphisms	
	Unique	Recurrent	Unique	Recurrent
5'US	6	1	0	1
Promoter	3	0	0	0
5'UTR	0	1	0	0
Exon: Nonsynonymous	1	0	0	1
Exon: Synonymous	1	1	0	3
Intron	65	38	3	48
3'UTR	5	7	1	2
3'DS	1	0	0	0

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

Table II

Novel *FMR1* Variants Only Found in Developmentally Delayed Males

Variant	Location	PhastCons	PhyloP	Patient frequency	Control frequency	P (Fisher's Exact Test)
Missense variant						
c.413G>A	Exon 5 (p.R138Q)	0.99	2.94	1/963 (0.1%)	0/1385 (<0.07%)	N.S.
Promoter variants						
c.-332G>C	GC Box/AP-2	0.72	0.65	1/963 (0.1%)	0/1308 (<0.08%)	N.S.
c.-293T>C	Inr-like/Tx. Start II/TATA-like	0.94	1.95	1/963 (0.1%)	0/1266 (<0.08%)	N.S.
c.-254A>G	Inr-like/Tx. Start I	1	2.10	1/963 (0.1%)	0/1304 (<0.08%)	N.S.
Other noncoding variants						
c.105-8A>C	Intron 2	1	2.48	6/963 (0.6%)	0/1262 (<0.08%)	0.007
c.630+438A>C	Intron 7	1	2.55	1/963 (0.1%)	0/1263 (<0.08%)	N.S.
c.631-840G>A	Intron 7	1	1.76	1/963 (0.1%)	0/1239 (<0.08%)	N.S.
c.880+885A>G	Intron 9 ^a	0	-0.73	1/963 (0.1%)	0/1084 (<0.09%)	N.S.
c.990+4T>C	Intron 10	1	2.46	1/963 (0.1%)	0/1248 (<0.08%)	N.S.
c.1472-521C>G	Intron 14 ^a	0	0.40	1/963 (0.1%)	0/1254 (<0.08%)	N.S.
c.*23T>C	3'UTR	0.99	2.40	1/963 (0.1%)	0/900 (<0.11%)	N.S.
c.*746T>C	3'UTR ^b	1	2.22	6/963 (0.6%)	0/1260 (<0.08%)	0.007
c.*1867G>A	3'UTR	1	1.54	12/963 (1.2%)	0/951 (<0.11%)	0.0005
c.*2035C>T	3'UTR	1	2.47	3/963 (0.3%)	0/1270 (<0.08%)	N.S.

^a denotes predicted splice site

^b denotes predicted miRNA binding site
 N.S. indicates p value is not significant.