Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Stephen Charles Collins Date

Assessment of the clinical impact of sequence variants in the *FMR1* gene

By

Stephen Charles Collins
Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Neuroscience

_____
Stephen T. Warren, Ph.D.
Advisor

_____
Joseph F. Cubells, M.D., Ph.D.
Committee Member

_____
Andrew Escayg, Ph.D.
Committee Member

_____
David J. Cutler, Ph.D.
Committee Member

_____
Peng Jin, Ph.D.
Committee Member

_____
Michael E. Zwick, Ph.D.
Committee Member

Accepted:

_____
Lisa A. Tedesco, Ph.D.
Dean of the Graduate School

_____
Date

Assessment of the clinical impact of sequence variants in the *FMR1* gene

By

Stephen Charles Collins
B.S., The Pennsylvania State University, 2002

Advisor: Stephen T. Warren, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the Graduate School of Emory University in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Neuroscience

2010

Abstract

Assessment of the clinical impact of sequence variants in the *FMR1* gene

By Stephen Charles Collins

Fragile X syndrome, the most common inherited form of developmental delay, is typically caused by trinucleotide repeat expansion in *FMR1*. However, little is known about the clinical significance of sequence variants in *FMR1*. Only three pathogenic sequence variants in *FMR1* have previously been described: one missense mutation, one splice site mutation, and one frameshift nonsense mutation. While these variants were detected in patients with classic fragile X-like features, it is possible that other mutations may result in a more subtle developmental phenotype. Therefore, to assess the clinical impact of *FMR1* sequence variants, we performed DNA sequencing on large populations of patients with either a fragile X-like phenotype or general developmental delay. Through the use of novel high-throughput technologies, namely array-based sequencing and massively parallel sequencing, we were able to sequence *FMR1* in significantly more patients than was previously feasible. Notably, we detected no pathogenic *FMR1* sequence variants in 51 fragile X-like patients. However, in 963 patients with developmental delay, we detected one missense variant and three promoter variants, all of which show evidence of a possible functional effect. If a functional effect is verified, these variants would represent the second *FMR1* missense mutation and the first three *FMR1* promoter mutations to cause developmental delay. These data suggest that sequence variants in *FMR1* may indeed be a significant contributor to the heterogeneous etiology of developmental delay.

Assessment of the clinical impact of sequence variants in the *FMR1* gene

By

Stephen Charles Collins
B.S., The Pennsylvania State University, 2002

Advisor: Stephen T. Warren, Ph.D.

A dissertation submitted to the Faculty of the Graduate School of Emory University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Graduate Division of Biological and Biomedical Science
Neuroscience

2010

Table of Contents

Figures

Tables

Chapter 1. Introduction

1.1 Intellectual Disability: Diagnosis, Prevalence, and Impact

Intellectual disability (ID) is the name currently preferred (Schalock et al. 2007) for the condition classified as mental retardation in the Diagnostic and Statistical Manual of Mental Disorders, fourth edition, text revision. DSM-IV-TR specifies that this diagnosis is to be given to individuals with significantly reduced intellectual function, as determined by an IQ less than 70, concomitant with impaired adaptive functioning in the realms of communication, self-care, home living, and work, among others. Furthermore, DSM-IV-TR stipulates that the onset of the reduced intellectual function must be before age 18 for a diagnosis of intellectual disability, thus distinguishing ID from dementias with an adult onset (American Psychiatric Association 2000).

Prevalence estimates of intellectual disability vary widely, from as low as 0.2% to as high as 8.5% (Roeleveld et al. 1997). This variability results from differences and difficulties in the diagnosis of ID (Leonard and Wen 2002). Like many disorders with a clinical definition, rather than a biochemical or molecular one, ID has been diagnosed differently at different points in time and continues to be diagnosed differently across different societies (Durkin 2002). For example, in the United States, the IQ cutoff for ID diagnosis was lowered over the span of 40 years from 84 (Heber 1961) to its current cutoff of 70 (American Psychiatric Association 2000; Luckasson and Reeve 2001). Also, because ID is inherently a disorder of functioning *within a society*, sociocultural differences in language (Reschly and Jipson 1976) and behavior (Zigler 1987) can impact

diagnosis and prevalence rates. Considering these challenges to an accurate estimate of population prevalence, the most informative measure of the frequency of ID, perhaps, is one that is functionally derived. For this reason, the Centers for Disease Control and Prevention recognize the prevalence of ID to be roughly 1%, based upon data from special education programs in US public schools (Yeargin-Allsopp et al. 1992; Centers for Disease Control and Prevention 1996).

The high frequency of intellectual disability and the lack of functional independence inherent to the condition underlie its broad and costly impact. In fact, the CDC estimates that the lifetime cost associated with ID for all patients born in a given year is $51.2 billion. Approximately 80% of this is the indirect cost of productivity loss due to impaired occupational functioning, while the remainder is the expense of direct medical costs and support services (Centers for Disease Control and Prevention 2004). The burden of ID is often most greatly borne by the family of the patient. In addition to shouldering much of the financial cost, parents and caregivers of the intellectually disabled have more subjective feelings of stress, fatigue, and social hindrance (Chou et al. 2008) and are more susceptible to depression (Olsson and Hwang 2001). Furthermore, siblings of people with ID receive less time and attention from their parents and often face social obstacles as a result of their affected sibling (Mulroy et al. 2008)

1.2 Etiologies of Intellectual Disability

Intellectual disability arises from a variety of causes, including congenital, prenatal, perinatal, toxic, infectious, psychosocial, chromosomal, and hereditary factors (Curry et al. 1997). The most common etiologies are chromosomal abnormalities, such as

aneuploidies and segmental aneusomies. Chief among these is trisomy 21, which causes 9.2% of all ID. Other noteworthy causes include the DiGeorge and Williams-Beuren microdeletion syndromes, which represent 2.4% and 1.3% of ID, respectively (Rauch et al. 2006). It is hypothesized that haploinsufficiency or genetic overdosage of a specific gene or genes leads to the ID seen in many aneuploidies and segmental aneusomies (Rachidi and Lopes 2008). However, while genomic critical regions have been defined in Down, DiGeorge, and Williams-Beuren syndromes, no single genes have yet been identified as the sole cause of ID in these chromosomal abnormalities (Rahmani et al. 1989; Ferrero et al. 2009; Meechan et al. 2009).

The second most frequent causes of intellectual disability are single gene mutations (Rauch et al. 2006). In fact, over 100 genes have been associated with ID (Basel-Vanagaite 2007; Gecz et al. 2009). Mutations in some of these genes lead to a collection of recurrent, recognizable signs and symptoms which define a clinical syndrome. The most common of these is fragile X syndrome, which is caused by trinucleotide repeat expansion in the *FMR1* gene (Verkerk et al. 1991). Other significant monogenic ID syndromes include Smith-Lemli-Opitz syndrome, Rett syndrome, and Coffin-Lowry syndrome, caused, respectively, by mutations in *DHCR7*, *MECP2*, and *RPS6KA3* (Hanauer and Young 2002; Porter 2008; Matijevic et al. 2009). More common than syndromic ID, however, is nonsyndromic ID, in which there is no specific clinical phenotype beyond the intellectual disability. While this nonsyndromic preponderance may be an artifact of incomplete phenotyping and small sample sizes of discrete disorders, the observed clinical homogeneity has made genetic linkage studies quite challenging (Ropers 2008).

Although much progress has been made in elucidating the varied causes of ID, 50% to 80% of intellectual disability remains of unknown origin (Rauch et al. 2006). The lack of a diagnosis can be quite troubling, particularly to the parents of someone affected with ID. Feelings of guilt or culpability and a lack of acceptance of the condition's permanence and significance are commonly reported by the parents of children with undiagnosed ID. Furthermore, parents of children without a diagnosis have no way to determine recurrence risk for reproductive decision-making, have less access to support groups, and have less guidance regarding management and/or treatment of their child (Rosenthal et al. 2001). Compared to the parents of healthy children and parents of children diagnosed with Down syndrome, parents of children with undiagnosed ID report greater emotional strain and are more likely to wish their child had not been born and to regret the difficulties the child brought to their lives (Lenhard et al. 2005). These effects of having no diagnosis underscore the need for improved etiologic understanding and diagnostic measures for intellectual disability.

1.3 Fragile X Syndrome

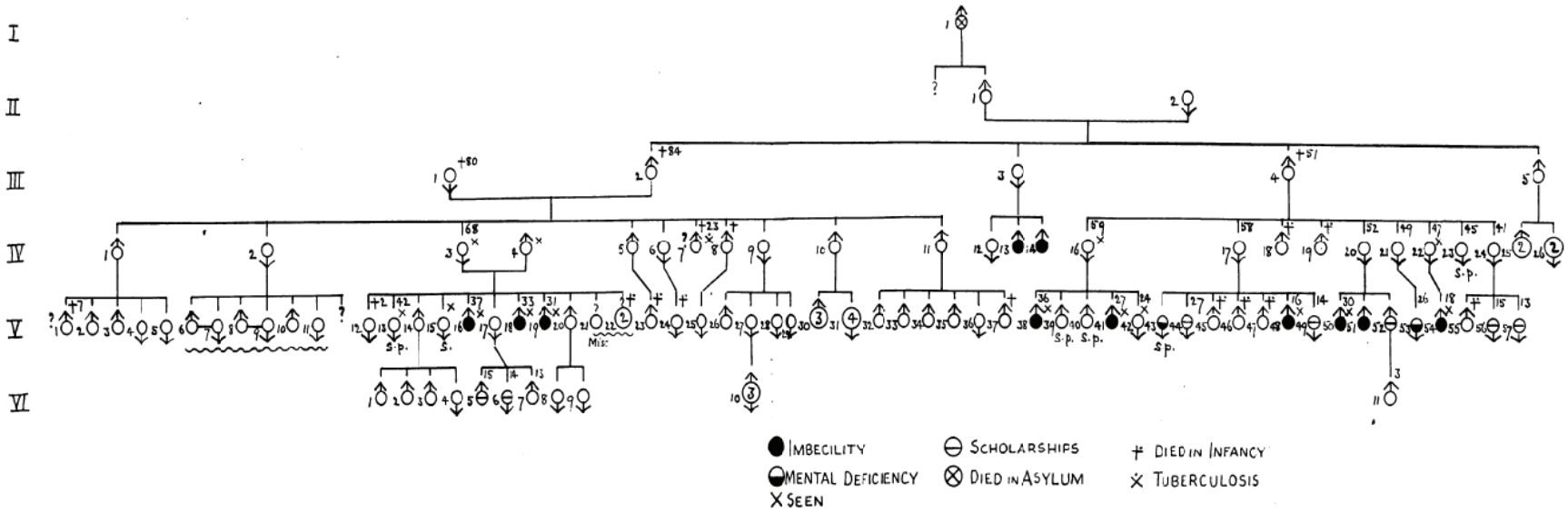
Fragile X syndrome (FXS) is the most common inherited form of intellectual disability. Independent studies conducted at sites throughout Europe and India have shown that 2-3% of all ID patients have FXS (Mila et al. 1997; Patsalis et al. 1999; Hecimovic et al. 2002; Pandey et al. 2002; Major et al. 2003; Biancalana et al. 2004; Rauch et al. 2006). Across the population, FXS affects 1 in 5161 males (Coffee et al. 2009) and, while measures are cruder, an estimated 1 in 8000 females (Crawford et al. 2001).

1.3.1 History of Fragile X Syndrome

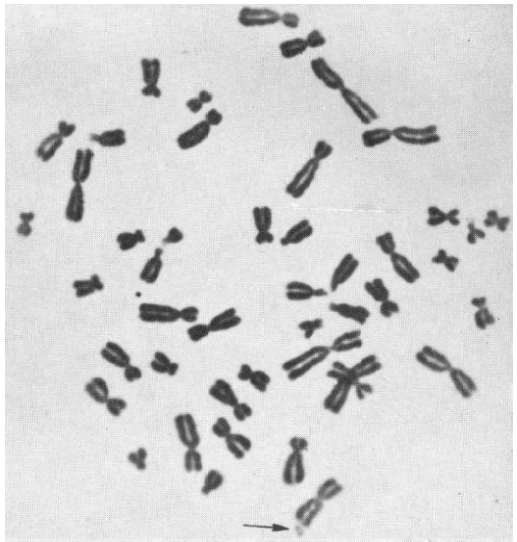
In 1943, J. Purdon Martin and Julia Bell published their seminal report on a family with eleven intellectually disabled males. While no females in the pedigree demonstrated "imbecility" (the antiquated term for severe intellectual disability), approximately 50% of the male offspring of unaffected females were affected (Figure 1.1A). Thus, the authors presented this pedigree as the first evidence of X-linked intellectual disability. Two features of the pedigree that puzzled the authors, however, were the lack of ID in the males of ancestral generations and the fact that two females exhibited some measure of mental deficiency, as both of these phenomena conflict with the hypothesis of recessive sex-linked Mendelian inheritance. With the exception of their ID, the affected males did not share any gross impairments or common physical features to distinguish their disorder from that of others affected with ID (Martin 1943).

H. A. Lubs encountered a family in 1969 that also demonstrated what appeared to be an X-linked recessive intellectual disability. Subsequent karyotyping of the affected males revealed a novel variant X chromosome with an unusual secondary constriction near the end of the long arm of the chromosome (Figure 1.1B). This "marker X" was found in all four affected males and two unaffected females in the pedigree, leading the author to conclude that the chromosomal abnormality associated with the recessive allele that led to intellectual disability (Lubs 1969). Indeed, cytogenetic evaluation of affected males from the family initially described by Martin and Bell revealed that they too featured a "fragile X" chromosome (Richards et al. 1981).

A


I
II
III
IV
V
VI
IMBECILITY
MENTAL DEFICIENCY
SEEN
SCHOLARSHIPS
DIED IN ASYLUM
DIED IN INFANCY
TUBERCULOSIS

B



C

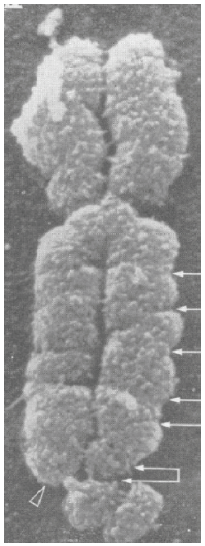


Figure 1.1. History of Fragile X Syndrome.

A. The original pedigree by Martin and Bell of the first family described in the literature with what we now know as fragile X syndrome. The transmission pattern is largely consistent with X-linked recessive inheritance of the intellectual disability, with approximately 50% of the unaffected females transmitting the ID to 50% of their male offspring. (Reprinted with permission from Martin, J.P., Bell, J. 1943. A pedigree of mental defect showing sex-linkage. *J Neurol Psychiatry* **6**: 154-157.)

B. The original metaphase spread by Lubs demonstrating a secondary constriction on the long arm of chromosome X in an intellectually disabled male. The appearance of this constriction would later inspire the name of the disorder, fragile X syndrome. (Reprinted with permission from Lubs, H.A. 1969. A marker X chromosome. *Am J Hum Genet* **21**(3): 231-244. Copyright The American Society of Human Genetics 1969.)

C. Scanning electron microscopic image of a G banded fragile X chromosome. SEM was used by Harrison and colleagues to determine the precise location of the fragile site to be Xq27.3. (Reprinted with permission from Harrison, C.J., Jack, E.M., Allen, T.D., and Harris, R. 1983. The fragile X: a scanning electron microscope study. *J Med Genet* **20**(4): 280-285.)

To identify the cause of the "fragile X syndrome," scientists turned their attention to the fragile site, in the hopes that the involved gene would reside either at it or nearby. Using scanning electron microscopy on G banded metaphase chromosomes (Figure 1.1C), Harrison and colleagues localized the fragile site to Xq27.3 (Harrison et al. 1983). Molecular mapping of Xq27.3 revealed a CpG island methylated in affected males (Heitz et al. 1991) and an unstable sequence (Yu et al. 1991) comprised of variable numbers of CGG trinucleotide repeats. In situ hybridization studies showed that these CGG repeats map precisely to the fragile site on chromosome X (Kremer et al. 1991). Finally, in 1991, a gene containing both the methylated CpG island and CGG repeat tract was identified. Designated *FMR1* (fragile X mental retardation-1), the gene was shown to be expressed in the brain, consistent with a role in fragile X etiology (Verkerk et al. 1991). Further supporting this role is the lack of *FMR1* expression in fragile X patients (Pieretti et al. 1991), while the gene is expressed widely in healthy individuals (Hinds et al. 1993).

Despite the refined understanding of the genetic cause of fragile X syndrome, the two puzzling observations about the original pedigree of Martin and Bell still remained to be resolved. The first of these, the presence of impaired females, was found to be a common phenomenon in fragile X families, with 30% of carrier females showing some measure of intellectual disability. Thus, rather than a recessive disorder, fragile X is considered to be an X-linked dominant disorder with reduced penetrance, likely due to patterns of X inactivation (Sherman et al. 1985). The second puzzling feature of no affected males in the Martin and Bell pedigree's ancestral generations was also seen in many other fragile X pedigrees. The perplexing inheritance pattern of increased risk of fragile X in successive generations, despite the low reproductive fitness of affected

males, became known as the "Sherman paradox" (Sherman et al. 1985). The explanation for this inheritance pattern, it would turn out, lay in the nature of the fragile X mutation. It was noted that affected males inherited an expanded allele of *FMR1* as compared to their mothers and maternal grandfathers, both of whom had the same size allele. The carrier mothers and their normal transmitting fathers, when compared to normal individuals, also had expanded alleles, termed premutations (Oberle et al. 1991). As was surmised, the mutation and premutation expansions originated in the variable length CGG trinucleotide repeat tract within *FMR1*, which ranges from a normal repeat count of 6-54 to a premutation repeat count of 55-200. The premutation allele can expand in oogenesis to a full mutation allele, which has more than 200 CGG repeats and leads to hypermethylation of the nearby CpG island, thereby explaining both the nature of the typical fragile X mutation and the Sherman paradox (Fu et al. 1991).

1.3.2 Fragile X Syndrome Phenotype

Intellectual disability is the hallmark feature of fragile X syndrome. The level of impairment ranges from mild to severe, with IQ scores of affected males typically between 25 and 70. Arithmetic, spatial memory, abstract thought, and complex reasoning are particular areas of weakness (Kemper et al. 1988), while verbal skills, including vocabulary and reading, are less significantly impaired (Hagerman et al. 1985). This is particularly true of receptive verbal skills, which are less impaired than expressive verbal skills (Roberts et al. 2001). Additionally, fragile X males tend to perform better at fact learning and school-related skills than at processing novel information (Kemper et al. 1988). Furthermore, longitudinal studies suggest that the IQ of affected males declines

over time, beyond their initial levels of impairment. The pattern of this decline is not consistent with a degenerative process, but rather a slowing of the intellectual growth of the fragile X male when compared to his unaffected peers and relatives (Skinner et al. 2005; Hall et al. 2008).

There are three subtle physical findings which eluded Martin and Bell in their 1943 pedigree but are now known to be common in fragile X syndrome. The first of these is the typical facial appearance of fragile X patients. Although not grossly dysmorphic, certain common features are shared: a long, narrow face, prominent jaw, large forehead, and large everted ears (Turner et al. 1980). A subset of these facial features can be seen in over 90% of affected males, but less than half of affected females (Rousseau et al. 1994). Among the specific features, large ears are the most closely associated with fragile X as compared to other causes of intellectual disability (Guruju et al. 2009).

Macroorchidism, the second noteworthy physical feature, is present in nearly all postpubertal fragile X males (Rousseau et al. 1994; Guruju et al. 2009). After multiple families were described with both X-linked ID and macroorchidism (Turner et al. 1975; Cantu et al. 1976; Ruvalcaba et al. 1977; Bowen et al. 1978), fragile X families were re-evaluated for the presence of enlarged testes, and macroorchidism was indeed found to be associated with fragile X syndrome (Turner et al. 1980). While a mouse model of fragile X showed increased Sertoli cell proliferation to be the cause of murine macroorchidism (Slegtenhorst-Eegdeman et al. 1998), no human studies have corroborated this finding; thus, the cause of macroorchidism in fragile X remains unclear

A third class of physical features associated with fragile X is an assortment of connective tissue abnormalities. Common findings include joints that are hyperextensible

and able to be voluntarily displaced, a high arched palate, pes planus, pectus excavatum, and velvety skin, particularly on the dorsum of the hand (Hagerman et al. 1984; Opitz et al. 1984). It is important to note, however, that joint hyperextensibility and soft skin are the only two of these features to show significant association with fragile X when compared to other causes of ID (Lachiewicz et al. 2000). When taken together, however, these signs indicate an underlying connective tissue dysplasia in fragile X. In fact, the previously mentioned fragile X facial dysmorphology, especially the large, everted ears, and the common finding of macroorchidism have also been attributed to connective tissue dysplasia, although this remains largely unsubstantiated
(Hagerman et al. 1984; Opitz et al. 1984).

Beyond the physical manifestations associated with fragile X syndrome, there are several noteworthy behavioral abnormalities that are commonly found in affected males. Hyperactivity and short attention span are two frequent findings (Guruju et al. 2009), such that over 50% of school-aged boys with fragile X meet DSM-IV criteria for attention-deficit/hyperactivity disorder (Sullivan et al. 2006). Additionally, autistic-like behaviors, including social impairments and stereotyped behaviors, are noted in more than 90% of males with fragile X (Hernandez et al. 2009). Common social impairments include eye gaze avoidance, social anxiety, and shyness. Frequently-seen stereotyped behaviors include perseveration of speech, hand-flapping, hand-biting and other self-injurious behaviors, and tactile defensiveness (Bailey et al. 1998; Hernandez et al. 2009). The presence of sufficient autistic-like behaviors leads to a dual diagnosis with autism in 25%-33% of fragile X males (Bailey et al. 1998; Rogers et al. 2001; Kaufmann et al. 2004; Hatton et al. 2006). While autism is commonly diagnosed among fragile X males,

fragile X is also commonly diagnosed among autistic males. In fact, with 7-8% prevalence of fragile X among autistic males, mutations in *FMR1* are the most frequent single-gene cause of autism (Muhle et al. 2004).

Several medical issues are also known to affect patients with fragile X syndrome. As is true of the intellectually disabled in general, fragile X patients have a higher rate of epilepsy than the general population. Affecting 22% of fragile X patients (Sabaratnam et al. 2001), these seizures appear to be both generalized (Wisniewski et al. 1991) and focal, and, while more commonly seen in childhood, can occur throughout adulthood in some fragile X patients (Sabaratnam et al. 2001). Another medical issue impacting many fragile X patients is recurrent otitis media. Over 60% of fragile X patients have more than 5 occurrences of otitis media during childhood. Many of these children are treated with the surgical insertion of polyethylene tympanostomy tubes (Hagerman et al. 1987). Two other medical issues that often affect fragile X patients are mitral valve prolapse and aortic root dilatation. While the clinical impact of these cardiovascular findings is unclear, they provide further evidence of connective tissue abnormalities in fragile X syndrome (Loehr et al. 1986).

Compared to the well-defined clinical picture of the fragile X male, the fragile X female often presents with a more subtle phenotype, thanks to the presence of a wild-type *FMR1* allele. Unlike in males, where intellectual disability is the hallmark feature, only 50% of fragile X females have an IQ less than 70. Another 21% have borderline IQ scores between 70 and 85 (de Vries et al. 1996). Even among fragile X females of normal intellect, though, learning disabilities are often encountered, especially involving mathematics and auditory linguistic processing (Wolff et al. 1988; Murphy and

Mazzocco 2008). Like the cognitive profile, the facial features typically seen in fragile X males are far more subtle in fragile X females, as are the autistic-like behaviors (Rousseau et al. 1994; Hatton et al. 2006). The social impairments seen in males, however, are not spared or dampened in females. In fact, the extreme shyness and social anxiety that affect many females with fragile X are profound enough to merit a diagnosis of avoidant personality disorder (Freund et al. 1993).

1.3.3 Diagnosing Fragile X Syndrome

The quest for a fragile X diagnosis is often a frustrating and slow-moving process. Fragile X is unlike many other causes of intellectual disability in that there are no obvious signs of disability at birth. Parents usually believe they have a healthy, "normal" child; it is not until developmental milestones begin to be delayed around the twelfth month of life that parents start to have concerns. These concerns are often met with a "wait and see" response from the primary care provider to whom the parents first turn for answers, resulting in delays to diagnosis and subsequent interventions. Further delaying diagnosis is the fact that many of the physical features associated with fragile X are subtle or completely missing in infants and young children. In all, it takes an average of more than 18 months after parents first have concerns about their child for fragile X to be correctly diagnosed (Bailey et al. 2003).

The diagnosis of fragile X syndrome begins with the clinical evaluation of the patient for the signs and symptoms comprising the fragile X phenotype. Several diagnostic checklists have been developed to assist clinicians as they attempt to identify patients who likely have fragile X (Butler et al. 1991a; Butler et al. 1991b; Hagerman et

al. 1991; Lachiewicz et al. 2000; Maes et al. 2000; Guruju et al. 2009). While each different checklist, upon presentation, was intended to include only diagnostically informative findings, subsequent studies have called the diagnostic value of many of the checklist items into question. For instance, transverse palmar and plantar creases, while sometimes seen in fragile X patients, have not consistently shown a specific association with fragile X as compared to other causes of ID (Lachiewicz et al. 2000; Guruju et al. 2009). The most predictive fragile X features are macroorchidism, large ears, hyperextensible joints, hyperactivity, poor eye contact, and a family history of intellectual disability (Lachiewicz et al. 2000; Guruju et al. 2009). Despite such checklists for the clinical prediction of which ID patients have fragile X, the diagnosis is ultimately made at the molecular level, where a definitive answer can be found. As a result, diagnostic checklists are not often used in practice; rather, the current standard of care dictates that all patients who present with ID of unknown etiology should be tested for the typical fragile X mutation (Garber et al. 2008).

According to the American College of Medical Genetics, over 99% of cases of fragile X are caused by CGG repeat expansion mutations in *FMR1* and the resultant hypermethylation of the *FMR1* promoter (Maddalena et al. 2001). As mentioned previously, a full mutation has more than 200 CGG repeats. Thus, molecular evidence of an *FMR1* allele with more than 200 CGG repeats establishes the fragile X syndrome diagnosis. This molecular evidence is provided by two complementary laboratory tests, PCR and Southern blotting. Using primers that flank the CGG repeat, PCR can generate amplicons of different sizes, as determined by the length of the CGG repeat tract. The sizes of the amplicons are determined with high resolution by gel electrophoresis.

However, PCR amplification of CGG repeats is challenging and prone to failure, particularly when the number of repeats approaches the premutation (55-200) or full mutation (>200) range. Thus, for accurate assessment of premutations and full mutations, Southern blotting is the preferred technique (Maddalena et al. 2001).

In Southern blotting, fragments of the patient's genomic DNA produced by digestion with various restriction enzymes are size-separated on an agarose gel. After blotting the gel to transfer the DNA fragments to a membrane, labeled DNA probes are hybridized to the DNA on the membrane to reveal the sizes of the DNA fragments, thus providing a measure of the number of CGG repeats within *FMR1*. This approach can be further enhanced with the use of a methylation-sensitive restriction enzyme, which will not be able to digest methylated DNA, allowing a hypermethylated promoter to stand out. Furthermore, because the inactive X chromosome in females is methylated, one can determine which *FMR1* allele is active and which is inactive; this inactivation pattern has implications for the severity of the female fragile X phenotype (Maddalena et al. 2001).

Because the standard of care is to test every patient who presents with ID of unknown etiology for the *FMR1* repeat expansion, 97 – 98% of tested patients receive a negative test result (Rauch et al. 2006). This does not rule out the possibility that the patient has a different type of mutation in *FMR1*. There are examples in the literature of point mutations that change the coding sequence (De Boulle et al. 1993) and alter splicing of *FMR1* (Lugenbeel et al. 1995), as well as deletions within or entirely encompassing *FMR1* (Coffee et al. 2008). Although the ACMG considers these less common mutations to represent fewer than 1% of the cases of fragile X (Maddalena et al. 2001), it is likely that this figure is an underestimate, resulting from incomplete

investigation into their frequency and contribution to disease. While some gross deletions may be made apparent in the process of testing for CGG repeat expansion, smaller deletions that are not in the vicinity of the CGG repeat tract will only be detected by other approaches, such as exon sequencing and sufficiently high resolution array comparative genomic hybridization (aCGH). Gene sequencing is also essential for the detection of point mutations and small indels in *FMR1*. However, with no data to support the value of such alternative methods, neither *FMR1* sequencing nor high resolution aCGH have become standard diagnostic tools and are not widely available. Further investigation into the frequency of point mutations and deletions in *FMR1* and their association with disease will be necessary to determine the clinical value of such approaches.

1.3.4 Managing Fragile X Syndrome

Despite increasing knowledge of the molecular underpinnings of fragile X, current treatment of the disorder remains symptomatic. ADHD and ADHD-like symptoms are typically well controlled by stimulants, including methylphenidate and Adderall, and α2-agonists, such as clonidine (Hagerman et al. 1988; Berry-Kravis and Potanos 2004). Selective serotonin reuptake inhibitors (SSRIs) are commonly used to counter perseverative behaviors, compulsions, and social anxiety. Antipsychotics, such as risperidone and aripiprazole, are well tolerated and can be useful for the reduction of aggression, irritability, and other aberrant social behaviors. Fragile X patients with seizures are treated with various anticonvulsants, including valproic acid, carbamazepine, and gabapentin (Berry-Kravis and Potanos 2004).

While current therapies are limited to the reduction of symptoms, several drugs are under development for the treatment of the underlying defect in fragile X syndrome, excessive mGluR5 signaling. FMRP, the protein encoded by *FMR1*, represses the postsynaptic translation of mRNAs essential for the process known as mGluR-LTD, or long-term depression triggered by metabotropic glutamate receptors. In mGluR-LTD, the activation of mGluR5 leads to increased protein translation in the synapse. An ensuing cascade of events results in the internalization of AMPA and NMDA receptors and an overall reduction in strength or elimination of the synaptic connection. In the functional absence of FMRP, as is seen in fragile X syndrome, there is a lack of repression of the mGluR5-dependent synaptic protein translation, leading to increased LTD, increased synapse loss, and decreased intellectual function (Bear et al. 2004). Hypothesizing that they could reduce this excessive LTD, positive modulators of AMPA-Rs and several selective antagonists of mGluR5 have entered clinical trials (Berry-Kravis et al. 2006; Berry-Kravis et al. 2009; Hagerman et al. 2009). While preliminary results on the ampakine CX516 did not reveal any significant cognitive or behavioral improvements from this class of drug (Berry-Kravis et al. 2006), work in fragile X animal models has shown strong benefits from mGluR5 blockade, giving much optimism for the success of this novel therapeutic strategy (McBride et al. 2005; Yan et al. 2005; Tucker et al. 2006).

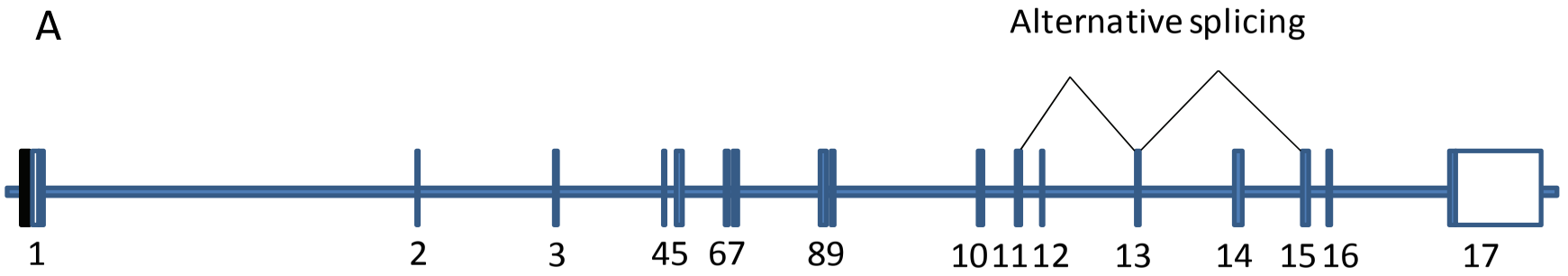
In addition to currently employed and investigational pharmacologic therapies, fragile X syndrome is also managed through educational and behavioral interventions. Individualized educational plans and speech therapy are useful methods to maximize the learning opportunities and communication skills of the fragile X patient. Specific behavioral interventions suggested to benefit affected individuals include those aimed at

improving eye contact, reducing stress, and improving sleep patterns. Additionally, interventions designed for autistic patients, such as social skills group therapy, may be useful for fragile X patients with autistic features (Hagerman et al. 2009).

While management of fragile X syndrome begins with efforts to improve the health and quality of life of the affected individual, the ramifications of a fragile X diagnosis extend throughout the family. The parents of a fragile X patient should be counseled on the risk of recurrence so they can make informed reproductive decisions (Bailey et al. 2003). Other relatives, especially those exhibiting developmental delay or social/behavioral disorders and those at risk of transmission of the mutation, should also be encouraged to seek testing (McConkie-Rosell et al. 2007). Furthermore, it is known that individuals who carry a premutation allele of *FMR1* are at increased risk for a tremor/ataxia disorder known as FXTAS (Jacquemont et al. 2003). Females carrying the premutation allele are also at increased risk for primary ovarian insufficiency (POI); 21% experience menopause before age 40 (Sherman 2000). Through cascade testing and genetic counseling, premutation carriers can be better prepared for these mid- to late-life changes.

1.4 *FMR1*

FMR1 is a 39 kb gene located at Xq27.3. The gene contains 17 exons, which are all of average size (Figure 1.2A). While most of the introns are also of average size (mean of 2.2 kb), the first intron is exceptionally large (9.9 kb), comprising a substantial portion of the length of the gene. The 5'UTR of *FMR1* is also unusually long (264 bp, on average) and contains the well-described $(CGG)_n$ element, a trinucleotide repeat of


A
Alternative splicing
1
2
3
45
67
89
10
11
12
13
14
15
16
17

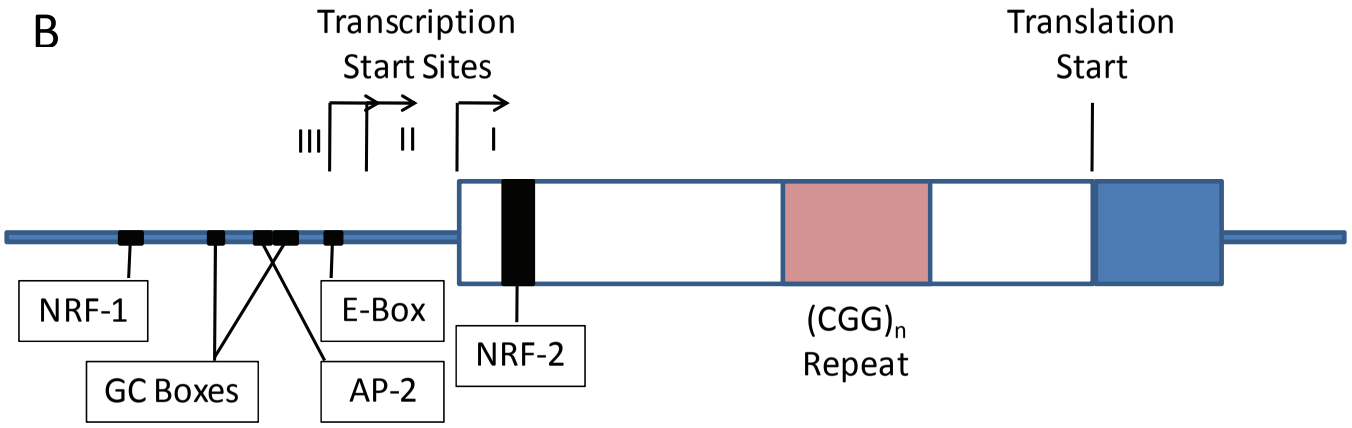

B
Transcription
Start Sites
III
II
I
Translation
Start
NRF-1
GC Boxes
E-Box
AP-2
NRF-2
(CGG)n
Repeat

Figure 1.2. Diagram of *FMR1*.

A. Diagram of *FMR1*. The horizontal axis is formed by intronic sequence, and the numbered vertical spokes represent the 17 exons of *FMR1*. Coding exonic sequence is shown in blue, while noncoding exonic sequence is shown in white. The black region upstream of exon 1 is the minimal promoter of *FMR1*. As depicted, alternative splicing can result in the omission of exons 12 and/or 14.

B. Diagram of the promoter and exon 1 of *FMR1*. There are three known transcription start sites (I-III) in *FMR1*, all of which coincide with an Inr-like element. Transcription factors that bind to the *FMR1* promoter include NRF-1, Sp1 (at the GC Boxes), AP-2, CREB (at the E-box), and NRF-2. The long 5'UTR of *FMR1* contains the well-described CGG trinucleotide repeat tract.

variable length (Eichler et al. 1993). Two paralogs of *FMR1*, known as *FXR1* and *FXR2*, have been described (Siomi et al. 1995; Zhang et al. 1995), as have conserved orthologs in species from vertebrates to Drosophila (Ashley et al. 1993b; Price et al. 1996; Wan et al. 2000).

FMR1 expression begins early in development and continues throughout the lifespan (Hinds et al. 1993). Through the use of multiple transcription start sites and alternative splicing, several distinct mRNA transcripts are encoded by *FMR1*. The most common of these is 4.4 kb in length and contains all 17 exons of the gene. The known alternative spliceforms (Figure 1.2A) selectively include or omit exons 12 and 14 and utilize one of three splice acceptor sites within exon 15 and one of two splice acceptor sites within exon 17 (Ashley et al. 1993b; Eichler et al. 1993). All of the distinct *FMR1* transcripts are broadly expressed in a variety of tissues (Verkerk et al. 1993), notably the hippocampus, cerebellum, cerebral cortex, and testis (Devys et al. 1993; Hinds et al. 1993).

1.4.1 *FMR1* Promoter

Through the use of reporter constructs and comparative genomics, the minimal promoter of *FMR1* (Figures 1.2B, 1.3) has been identified as the region extending from approximately 400 bp to 200 bp upstream of the translation start site (Hwu et al. 1993; Hergersberg et al. 1995; Kumari and Usdin 2001). Within this region, footprint analysis of DNA extracted from peripheral tissues revealed four transcription factor binding sites (Drouin et al. 1997; Schwemmle et al. 1997). The most upstream of these is a confirmed α-PAL/NRF-1 binding site. Reporter constructs containing either a deletion or

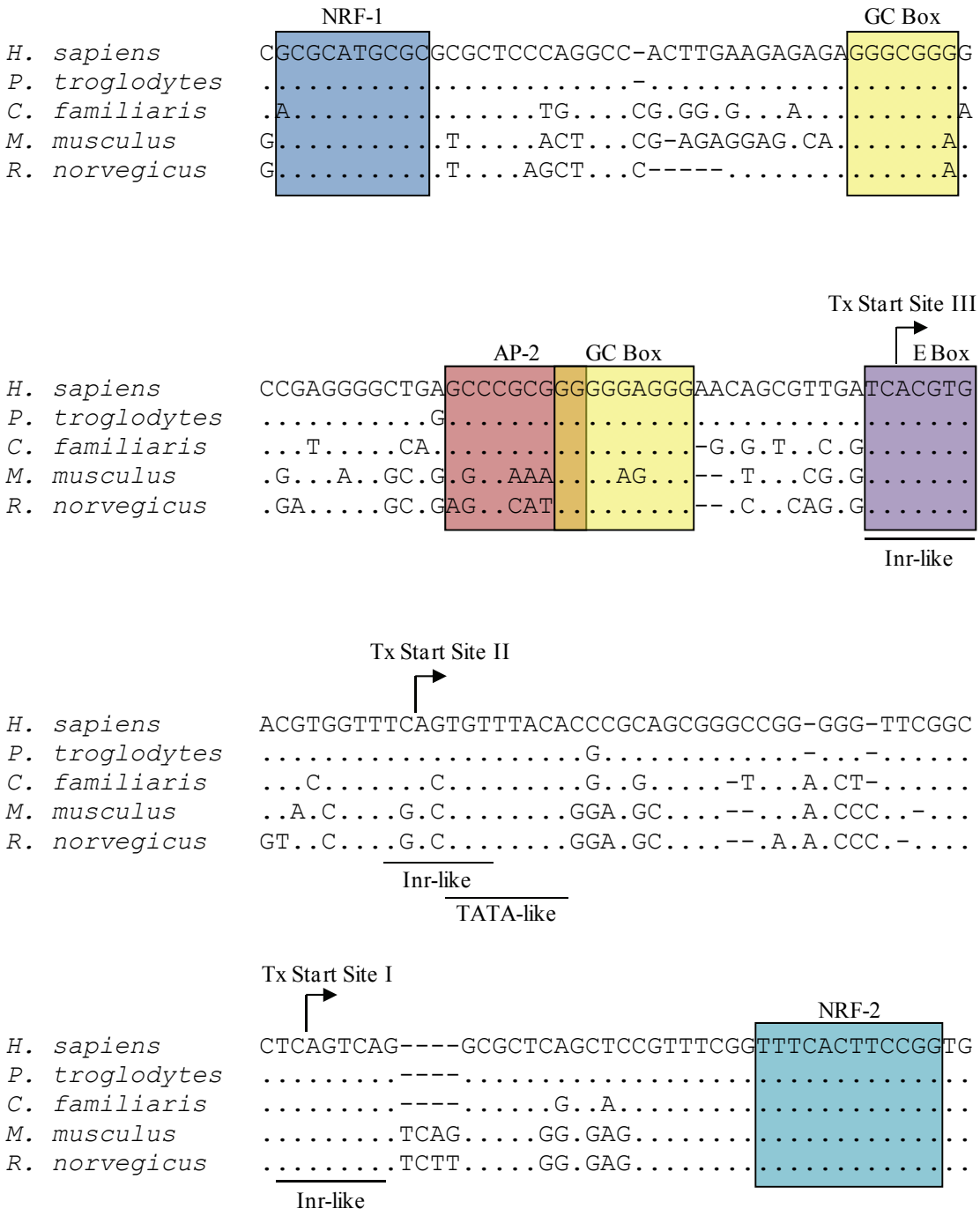

NRF-1
GC Box
H. sapiens CGCGCATGCGCGCGCTCCCAGGCC-ACTTGAAGAGAGAGGGCGGGG
P. troglodytes .....-.....
C. familiaris .A.....TG....CG.GG.G...A.....A
M. musculus G.....T.....ACT...CG-AGAGGAG.CA.....A.
R. norvegicus G.....T....AGCT...C-----.....A.
Tx Start Site III
AP-2
GC Box
E Box
H. sapiens CCGAGGGGCTGAGCCCGCGGGGGGAGGGAACAGCGTTGATCACGTG
P. troglodytes .....G.....
C. familiaris ...T.....CA.....-G.G.T..C.G.....
M. musculus .G...A..GC.G.G..AAA....AG...--.T...CG.G.....
R. norvegicus .GA.....GC.GAG..CAT.....--.C..CAG.G.....
Inr-like
Tx Start Site II
H. sapiens ACGTGGTTTCAGTGTTTACACCCGCAGCGGGCCGG-GGG-TTCGGC
P. troglodytes .....G.....-...-.....
C. familiaris ...C.....C.....G..G.....-T...A.CT-.....
M. musculus ..A.C....G.C.....GGA.GC....--...A.CCC..-...
R. norvegicus GT..C....G.C.....GGA.GC....--.A.A.CCC.-....
Inr-like
TATA-like
Tx Start Site I
NRF-2
H. sapiens CTCAGTCAG----GCGCTCAGCTCCGTTTCGGTTTCACTTCCGGTG
P. troglodytes .....----.....
C. familiaris .....----.....G..A.....
M. musculus .....TCAG.....GG.GAG.....
R. norvegicus .....TCTT.....GG.GAG.....
Inr-like

Figure 1.3. Conservation of the *FMR1* promoter.

The minimal promoter of *FMR1* extends from approximately 400 bp to 200 bp upstream of the translation start site. It contains several functional elements that are conserved through mammals; depicted here is the alignment of the promoter sequences from human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), dog (*Canis familiaris*), mouse (*Mus musculus*), and rat (*Rattus norvegicus*). Highlighted boxes represent the transcription factor binding sites for NRF-1, AP-2, and NRF-2; the two GC boxes, which bind Sp1 and Sp3; and the E-box, which binds CREB *in vivo*. The conserved Inr-like and TATA-like elements are underlined. The three transcription start sites are denoted with arrows.

methylation of this site exhibit reduced transcription, confirming its essential role in *FMR1* transcriptional activation (Kumari and Usdin 2001). The interior two transcription factor binding sites detected by footprinting are GC boxes, which have been shown to bind the strong transcriptional activators Sp1 and Sp3 (Smith et al. 2004; Kumari et al. 2005). The most downstream transcription factor binding site detected by footprinting is an E-box. While many transcription factors are capable of binding to the E-box consensus sequence, studies of human lymphoblasts suggest that CREB is the transcription factor that binds *in vivo* to the *FMR1* E-box (Smith et al. 2006). Remaining unresolved are the mechanisms of action for USF1, USF2, and Max, three other E-box-binding transcription factors that have been shown to inhibit *FMR1* transcription (Smith et al. 2004; Smith et al. 2006). While USF1 and USF2 were shown to bind to the *FMR1* E-box *in vitro* (Kumari and Usdin 2001), *in vivo* assays have not been able to replicate this finding (Smith et al. 2004). Therefore, it remains to be seen whether USF1, USF2, and Max bind the *FMR1* E-box in some context-specific fashion or if they simply act via a different binding site.

In addition to the four sites identified by footprint analysis, several putative transcription factor binding sites have been identified by the presence of their consensus sequences in the *FMR1* promoter. Among these are four potential AP-2 biding sites (Lim et al. 2005). Initially, these sites were ignored, due to their reduced conservation and the lack of an AP-2 footprint at any of the four sites in adult peripheral tissues. However, chromatin-IP studies have revealed an interaction between AP-2α and the *FMR1* promoter in HeLa cells. Additionally, decreased expression of *Fmr1* was noted in *AP-2α* knockout mice. Interestingly, this phenomenon was dependent on the age of the mouse,

as *Fmr1* levels were normal in adult conditional *AP-2α* knockouts, suggesting that AP-2α acts as a transcriptional activator of *FMR1* specifically during early development (Lim et al. 2005). Another transcription factor binding site that was not revealed by DNA footprinting was the conserved NRF-2 binding site. Located downstream of the typical transcription start site of *FMR1*, this binding site mediates the transactivating effect of NRF-2 (Smith et al. 2006). Together, these data suggest that other unconfirmed putative transcription factor binding sites in the *FMR1* promoter, such as Zeste (Drouin et al. 1997), may yet prove to be functional, possibly in a specific cell type or stage of development.

FMR1 transcription has been shown to begin at three discrete start sites (Figures 1.2B, 1.3). All three of these sites show substantial sequence similarity to the Initiator (Inr) consensus sequence (Kumari and Usdin 2001; Beilina et al. 2004). Inr elements direct RNAP II transcription initiation, either synergistically with a TATA box or in promoters that lack a TATA box (Smale and Baltimore 1989). While there appears to be no TATA box in *FMR1*, there is a TATA-like element located 26 bp upstream of start site I, the typical transcription start site (Hwu et al. 1993). Studies of reporter constructs containing deletions of the TATA-like element and the most downstream Inr element suggest that they may not be essential for transcription (Kumari and Usdin 2001). However, the 3-D structure of the *FMR1* promoter must be maintained for proper function (Kumari et al. 2005), so such deletion studies may not accurately reflect the *in vivo* behavior of the *FMR1* promoter.

1.4.2 *FMR1*: CGG Repeat

Located within the 5'UTR of *FMR1* (Figure 1.2B), the CGG repeat tract is an unstable genomic region that is prone to meiotic and mitotic expansion and contraction (Kremer et al. 1991). The likelihood of expansion is known to be influenced by three factors. First, larger repeat tracts are more susceptible to these changes in size. Thus, it is quite common for a female with a premutation allele of *FMR1* (i.e. with 55-200 CGG repeats) to have a child with a full mutation allele (i.e. more than 200 CGG repeats), as was noted in the "Sherman paradox" (Sherman et al. 1985; Fu et al. 1991). Furthermore, expansion of premutations to full mutations only occurs via maternal transmission, as spermatogenesis cannot maintain full length CGG repeat tracts (Malter et al. 1997). Finally, a third factor that influences the propensity for expansion is the presence of interspersed AGG repeats within the CGG repeat tract; interspersed AGG repeats confer stability and reduce the likelihood of expansion (Kunst and Warren 1994; Eichler et al. 1996).

While it is known that CGG repeat expansion is most common in females who have premutation-sized alleles with few interspersed AGGs, the exact mechanism of the expansion has been difficult to assess. A mouse model containing a human *FMR1* repeat tract with 98 CGG's was established to study the molecular mechanisms of repeat expansion. However, despite the large repeat tract that was introduced, only subtle expansions occurred in the mouse, as compared to the large expansions which are often seen in the human transmission of a full mutation by a premutation female (Bontekoe et al. 2001). Subsequently, larger expansions were elicited by crossing premutation mice to heterozygous *ATR* knockout mice which show impaired DNA repair at stalled replication forks (Entezam and Usdin 2008). While this suggests that DNA repair pathways are

responsible for CGG repeat expansion, this conclusion remains controversial. An alternative hypothesis attributes CGG repeat expansion to slippage during lagging strand synthesis. While also largely unsubstantiated, this hypothesis is supported by the similar 3' orientation of repeat tract expansion in human *FMR1* (Eichler et al. 1996) and in *in vitro* models of CGG repeat instability (Hirst and White 1998).

Although the cause of repeat tract expansion has been elusive, its effects on gene transcription have been well described. In fact, the size of the CGG repeat tract impacts the transcription of *FMR1* in two important and opposite ways. In individuals with intermediate-sized (i.e. 40-54 CGG repeats) or premutation alleles, the amount of *FMR1* mRNA transcribed is directly proportional to the number of CGG repeats, such that individuals with premutations have three- to five-times as much *FMR1* mRNA as people with normal alleles (Kenneson et al. 2001; Loesch et al. 2007). This excess of *FMR1* mRNA has been proposed to cause the premutation phenotypes of FXTAS and POI (Tassone et al. 2000; Jin et al. 2003). Along with the increase in transcription, there is a shift in the predominant transcription start site, from I to II and III, in individuals with larger repeat tracts; both the mechanism and consequence of this shift is unclear (Beilina et al. 2004). In contrast to the increase in transcription seen in premutations, full mutations cause *FMR1* to become transcriptionally silent. This silencing results from methylation of CpGs in the *FMR1* promoter (Sutcliffe et al. 1992; Chiurazzi et al. 1998) and the concomitant deacetylation and altered methylation pattern of histone H3 (Coffee et al. 1999; Coffee et al. 2002). Although the mechanism by which full mutation CGG expansions lead to these epigenetic changes is unclear, the existence of rare individuals in fragile X pedigrees who have full mutations without transcriptional silencing or a fragile

X phenotype suggests that as-yet-undescribed *trans* effects are involved (Pietrobono et al. 2005).

In addition to its effects on *FMR1* transcription, the CGG repeat tract also affects the translation of *FMR1*. In reporter expression systems, FMRP levels (normalized to the amount of transcript) increase as the repeat tract increases from 0 to 30 CGGs, and then decrease as the repeat tract increases from 30 to 60 CGGs and throughout the premutation range (Chen et al. 2003). In fact, for premutation carriers, FMRP levels are inversely proportional to the number of CGG repeats (Kenneson et al. 2001). This appears to result from impaired translation initiation of *FMR1* mRNA with premutation-sized CGG repeat tracts (Primerano et al. 2002). The decreased FMRP levels in premutation carriers have led to the as-yet-untested hypothesis that the increased transcription seen in premutations is a biological attempt to balance the decreased translational efficiency. However, it is important to note that this proposed balance mechanism is not sufficient to maintain normal levels of FMRP (Kenneson et al. 2001). Thus, it is unlikely that attempts to correct fragile X syndrome by reactivating the transcription of *FMR1* will provide adequate levels of FMRP for normal function.

While the CGG repeat tract is known to cause fragile X through expansion mutation and subsequent transcriptional silencing, the instability of this region also allows for repeat contraction, which results in intragenic *FMR1* deletions (de Graaff et al. 1995). The deletion breakpoints often lie outside of the CGG repeat tract. In fact, fragile X-like patients have been seen with deletions extending upstream that remove the promoter (Meijer et al. 1994; Hirst et al. 1995) and deletions extending downstream that remove exon 1 (Quan et al. 1995). The effect of such deletions is similar to that of CGG

repeat expansion: a lack of *FMR1* transcription and FMRP translation. Often these deletions are found in patients mosaic for the CGG repeat expansion (de Graaff et al. 1995), suggesting that a common mechanism of instability is indeed at work. Larger deletions encompassing the majority (Hirst et al. 1995) or entirety (Gedeon et al. 1992) of *FMR1* have also been noted; however, these are not believed to typically result from the instability of the CGG repeat in FMR1 (Coffee et al. 2008).

1.4.3 Conventional Mutations in *FMR1*

Shortly after the identification of *FMR1* as the gene located at the chromosome Xq27.3 fragile site, a patient with a severe fragile X-like phenotype was identified with a point mutation in the coding sequence of *FMR1* (De Boulle et al. 1993). This mutation, commonly referred to as I304N for the missense change it introduces in FMRP, was a monumental discovery for three important reasons. First, it confirmed *FMR1* as the cause of fragile X syndrome. Although *FMR1* was at the right genomic location and is not expressed in fragile X, one could not be certain that the lack of functional FMRP truly caused the phenotype until the effects of a specific disruption of function were observed. Second, the I304N mutation became a valuable tool in the molecular and physiological characterization of FMRP. As will be discussed in the following section, innumerable studies would later compare FMRP with and without the I304N mutation in an attempt to better understand the normal function of the protein. Third, the discovery of a conventional point mutation in *FMR1* suggested that mutations other than CGG repeat expansion could cause fragile X. While previous work had focused upon the cytogenetic

fragile site and the novel trinucleotide repeat expansion mutation, a more thorough search for *FMR1* mutations may be necessary in the clinical diagnosis of fragile X syndrome.

Soon after this initial discovery, two more patients were found to have intragenic mutations in *FMR1*. One was a three year-old male presenting with developmental delay, poor eye contact, hyperactivity, and facial features reminiscent of fragile X. Mutation detection electrophoresis and subsequent DNA sequencing revealed a single nucleotide deletion within exon 5, c.373delA, which results in a frameshift and premature translational stop. This *de novo* mutation resulted in no detectable FMRP expression. The other intragenic mutation discovered was a two basepair substitution in the splice acceptor site at the boundary of intron 1 and exon 2. This change, c.[52-1G>T;52G>A], results in the skipping of exon 2, causing a frameshift and premature translational stop and the loss of FMRP expression. The mutations were inherited by the fragile X-like proband from his mildly intellectually disabled mother, but no additional family members were found to share these mutations (Lugenbeel et al. 1995).

Excited by this initial success, many groups sought to identify further intragenic mutations in *FMR1* that cause a fragile X-like phenotype (Table 1.1). The group which had identified the first missense mutation, I304N, sequenced the *FMR1* cDNA from a patient presenting with a history of X-linked intellectual disability and a clinical picture highly reminiscent of the I304N patient, but found no mutations in the coding sequence (Reyniers et al. 1996). Another group examined five fragile X-like males by the same method, but also found no *FMR1* mutations (Chiurazzi et al. 1994). With conventional mutations in *FMR1* clearly not the only cause of a fragile X-like phenotype in patients negative for repeat expansion, subsequent attempts to identify such mutations examined

Table 1.1. Screening for conventional mutations in *FMR1* in various patient populations

N	Patient Diagnosis	Mutation Detection Method	Putative Conventional Mutations Identified	Reference
5	FXS-like	cDNA sequencing	None	Chiurazzi et al 1994
1	FXS-like	cDNA sequencing	None	Reyniers et al 1996
27	FXS-like	SSCP	c.1637G>A (R546H)	Wang et al 1997
31	FXS-like	SSCP	None	Castellvi-Bel et al 1999
6	FXS-like	SSCP	None	Gronskov et al 1998
27	Autism	SSCP	None	Vincent et al 1996
90	Autism and/or nonspecific ID	SSCP	c.879A>C (eliminated splice donor; V296SfsX8)	Shinahara et al 2004
112	General ID	SSCP	None	Gronskov et al 1998
208	XLID	Sanger sequencing	None	Tarpey et al 2009

Notes: The c.1637G>A and c.879A>C variants identified in affected individuals have not been established as causative of the patient phenotype. The number of individuals with autism, nonspecific ID, or both is not specified in Shinahara et al 2004.

larger numbers of individuals. A mutational screening method popular in the mid- to late-1990s was single-strand confirmation polymorphism analysis (SSCP). Although less sensitive than sequencing, its cost-effectiveness and technical simplicity made SSCP a screening method of choice (Dean 1995). Using this approach, two more groups attempted to identify conventional *FMR1* mutations in fragile X-like patients (Wang et al. 1997; Castellvi-Bel et al. 1999). In their screening of a combined 58 fragile X-like patients, only one possible causal mutation was unveiled, c.1637G>A (Wang et al. 1997). This variant is at a highly conserved nucleotide in exon 15 of *FMR1*, and results in the missense change R546H at a highly conserved residue in a known domain of FMRP, the RGG box. However, distracted by what later was shown to be the erroneous finding (Vincent and Gurling 1998) of a splice variant which would have introduced a premature stop codon, the authors did not follow up on this coding variant; hence, its clinical and functional significance cannot be determined.

Noting that two of the three confirmed conventional mutations found in *FMR1* in fragile X-like patients resulted in a complete elimination of FMRP expression, some groups hypothesized that more subtle changes to FMRP, such as would be encoded by most point mutations, would result in more subtle phenotypes than fragile X. Thus, groups began to use SSCP to screen for *FMR1* coding mutations in patients with autism (Vincent et al. 1996; Shinahara et al. 2004) and non-syndromic intellectual disability (Gronskov et al. 1998; Shinahara et al. 2004). While these studies did not reveal any missense or nonsense changes in *FMR1* in these patient populations (Vincent et al. 1996; Gronskov et al. 1998; Shinahara et al. 2004), one female with severe intellectual disability and autistic features was found to be heterozygous for a synonymous variant in

exon 9, c.879A>C. This variant alters the conserved splice donor site, resulting in the inclusion of 47 bp of intron 9 in the mRNA transcript and the creation of a premature stop codon (Shinahara et al. 2004). Unfortunately, the significance of this variant is difficult to interpret for three reasons. First, while the presence of the aberrant mRNA transcript was demonstrated in the patient's peripheral blood cells, no assessment was made of the patient's FMRP levels. Ultimately, this is where a functional effect would manifest. Second, the authors did not adequately demonstrate that this variant is only found in affected individuals. No family studies were described, including the simple assessment of if the variant is *de novo* or inherited, and only 50 controls were shown to be negative for this variant. Third, the phenotype of females heterozygous for mutations in *FMR1* can be highly variable, due to the presence of an intact *FMR1* allele and the variability in X inactivation. However, it is uncommon for fragile X females to be as severely affected as this patient, suggesting that other factors may be causing or influencing the reported phenotype.

All of these attempts to identify causal mutations in the coding sequence of *FMR1* were limited in several notable ways. First, the studies proved to be quite low-powered, as among individuals lacking a CGG repeat expansion, conventional *FMR1* mutations are apparently less frequent than 1/30 in the fragile X-like phenotype, 1/80 in autism, and 1/120 in general intellectual disability (for details, see Table 1.1). Second, the sensitivity of the mutation detection technique SSCP is less than direct sequencing and likely inadequate for this purpose (Dean 1995). Third, the inclusion of females in a screen for conventional mutations in a gene on the X chromosome can make the interpretation of findings more challenging, as it was for the c.879A>C variant (Shinahara et al. 2004). An

improvement upon each of these shortcomings was made in a recent large-scale resequencing of patients with demonstrated X-linked intellectual disability, or XLID (Tarpey et al. 2009). In this study, the X chromosome exome of 208 males with XLID was sequenced with conventional Sanger sequencing. Although the intention was to identify novel genes involved in XLID, this study also interrogated known XLID genes such as *FMR1*. However, in these 208 males, no missense or nonsense changes were uncovered in *FMR1* (Tarpey et al. 2009). Thus, in 70 males with a fragile X-like phenotype, 396 males with general intellectual disability, 103 males with autism, and 14 females with either autism or intellectual disability, only 3 known and 2 putative causal conventional mutations in *FMR1* have been identified (Table 1.2). Larger patient populations will be vital for the accurate estimation of the disease burden of conventional *FMR1* mutations.

1.5 FMRP

FMRP, the gene product of *FMR1*, is predominantly a 632 amino acid, 71 kDa protein, although several isoforms exist as a result of alternative splicing of *FMR1* (Devys et al. 1993; Verheij et al. 1993; Sittler et al. 1996). Like the *FMR1* mRNA transcript, FMRP is expressed broadly in various tissue types. Notable tissues expressing FMRP include the testis and brain, particularly in neurons of the hippocampus, cerebellum, and cortex. On the other hand, tissues of mesodermal origin, such as heart and skeletal muscle, show little to no FMRP expression (Devys et al. 1993). Within cells, FMRP is predominantly localized to the cytoplasm (Verheij et al. 1993); however, there is *in vitro* evidence that isoforms translated from alternative spliceforms lacking exon 14

Table 1.2. Conventional mutations in *FMR1* with verified or suggested causative role in intellectual disability

Variant class	cDNA change	Protein change	Segregation within family	Controls negative for variant	Functional Evidence	Patient phenotype	References
Missense/ Point mutation	c.911T>A	p.I304N	-*de novo* in proband -No ID in family	130	Interferes with FMRP-ribosome association	-IQ < 20 -Impressive macroorchidism -Acromegalic facies -Large ears	De Boulle et al 1993; Feng et al 1997a
Frameshift/ Single base deletion	c.373delA	p.T125LfsX66	-*de novo* in proband -No mention of ID in family	0	No detectable FMRP	-Developmental delay -Poor eye contact -Hyperactivity -Fragile X facies -No macroorchidism	Lugenbeel et al 1995
Altered splice acceptor/ Two basepair mutation	c.[52-1G>T; 52G>A]	p.A18LfsX4	-Inherited from mildly ID mother -Mutation and ID otherwise absent from family	0	No detectable FMRP	-Developmental delay -Poor eye contact -Autism -Fragile X facies -Connective tissue abnormalities -Macroorchidism	Lugenbeel et al 1995
Missense/ Point mutation	c.1637G>A	p.R546H	-Variant not present in unaffected sister	1	None	-"Typical fragile X phenotype"	Wang et al 1997
Altered splice donor/ Point mutation (**heterozygous)	c.879A>C	p.V296SfsX8	-No mention of inherited vs. *de novo* -No family studies	50	Aberrantly spliced transcript, but no evidence of change in FMRP levels	-Severe intellectual disability -Autistic features	Shinahara et al 2004

The variants above the double line have been confirmed as causal mutations, while the variants below the double line are suggested to be causal mutations.

are localized to the nucleus (Sittler et al. 1996). These two patterns of intracellular localization likely derive from the functional nuclear localization signal (NLS) and nuclear export signal (NES) of FMRP, two of the protein's known domains (Eberhart et al. 1996). Other known domains (Figure 1.4) include the N-terminal domain (NDF), which is involved in protein-protein interactions (Adinolfi et al. 2003), two K homology domains (KH1 and KH2), which mediate polyribosome interaction and RNA binding, and the RGG box, another RNA-binding domain (Ashley et al. 1993a; Siomi et al. 1993).

Much of our understanding of the physiological role of FMRP has come from situations in which it is absent: that is, through direct examination of fragile X patients and studies of animal models. As discussed previously, the neurocognitive phenotype of fragile X patients suggests that FMRP plays an important role in learning and memory. Consistent with this, autopsy examination of the brains of several fragile X patients has revealed an abundance of immature dendritic spines, which appear long, thin, and tortuous (Rudelli et al. 1985; Hinton et al. 1991). This neuroanatomical phenotype had previously been seen in other forms of intellectual disability, including Down syndrome, and is believed to reflect impaired synaptic plasticity (Purpura 1974; Marin-Padilla 1976). Much like fragile X patients, the mouse model of fragile X exhibits dendritic spine dysgenesis (Comery et al. 1997; Irwin et al. 2002). This mouse model, created by homologous recombination-mediated knockout of murine *Fmr1*, recapitulates much of the fragile X phenotype, including deficiencies in learning, macroorchidism, and abnormal social behavior (Bakker CE et al. 1994; Slegtenhorst-Eegdeman et al. 1998; Dobkin et al. 2000; Spencer et al. 2005). The similarity in both the neuronal morphology and the cognitive and behavioral features between fragile X patients and the *Fmr1*

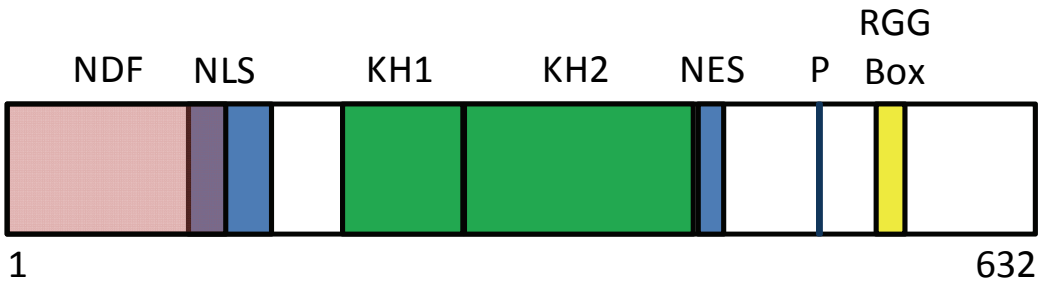


Figure 1.4. Diagram of FMRP.

The known domains of FMRP are identified. The N-terminal Domain of FMRP (NDF) mediates homodimerization and other protein-protein interactions. The nuclear localization signal (NLS) and nuclear export signal (NES) mediate the subcellular trafficking of FMRP. The K homology domains (KH1 and KH2) mediate polyribosome interaction and RNA binding. The RGG Box also mediates RNA binding. The primary site of FMRP phosphorylation, serine 500, is denoted with *P*.

knockout mouse suggest that this mouse model is a useful tool for the study of FMRP function and the effects of its absence.

While the fragile X patient phenotype and the *Fmr1* knockout mouse have illuminated the general role of FMRP in normal synaptic functioning, biochemical studies have been most effective at establishing the specific molecular mechanisms by which FMRP acts. The first biochemical property of FMRP to be noted was its ability to bind RNA (Ashley et al. 1993a; Siomi et al. 1993). Rather than serving as a general RNA binding protein, FMRP binds specific mRNA ligands representing approximately 4% of the brain mRNA (Ashley et al. 1993a). Structurally, the RNA species bound by FMRP often feature stem-loop G-quartets. Such mRNA species bind to the RGG box of FMRP *in vitro* (Darnell et al. 2001; Schaeffer et al. 2001). Several mRNA transcripts containing G-quartets have been verified as *in vivo* targets of FMRP, including *MAP1B*, *SEMA3F*, *SAPAP3*, *SAPAP4* (Brown et al. 2001), *PSD-95* (Todd et al. 2003; Zalfa et al. 2007), *APP* (Westmark and Malter 2007), and *FMR1* itself (Ashley et al. 1993a; Schaeffer et al. 2001). Another mRNA structural motif that is capable of binding to FMRP is the loop-loop pseudoknot termed the "kissing complex". While this motif is bound by the KH2 domain *in vitro*, this interaction has not been observed *in vivo* (Darnell et al. 2005). It has thus been postulated that the "kissing complex" motif may actually represent the *in vivo* interaction between two RNA molecules, such as an mRNA and a noncoding RNA (Bassell and Warren 2008).

Another essential biochemical property of FMRP is its ability to bind to ribosomes (Khandjian et al. 1996; Siomi et al. 1996; Tamanini et al. 1996). Although capable of binding to monosomes and ribosomal subunits, FMRP preferentially binds to

actively translating polyribosomes (Eberhart et al. 1996; Corbin et al. 1997). Studies of the I304N missense mutation, which alters a conserved residue in the KH2 domain of FMRP in a patient with a severe fragile X phenotype (De Boulle et al. 1993), have implicated the KH domains in this interaction (Feng et al. 1997a). Through its association with polyribosomes, FMRP regulates the translation of its bound mRNA ligands. *In vitro*, FMRP exclusively inhibits mRNA translation (Laggerbauer et al. 2001; Li et al. 2001). However, the *in vivo* regulatory effects of FMRP are more complex. In fragile X patient cells, where there is no FMRP expressed, some mRNA ligands have increased polyribosome loading, likely due to decreased inhibition of translation, but the association with polyribosomes is decreased for other ligands (Brown et al. 2001). Thus, fragile X syndrome is best viewed as a state of translational dysregulation.

As mentioned previously, the phenotype of fragile X patients and the mouse model suggest that the primary defect caused by the lack of FMRP is synaptic dysfunction. Thus, it is not surprising that the key site of translational dysregulation in fragile X is at the synapse. Several mRNA ligands of FMRP, including *MAP1B* and *PSD-95*, have been shown to colocalize with FMRP in dendrites (Antar et al. 2005; Muddashetty et al. 2007; Zalfa et al. 2007). Furthermore, excess translation of the FMRP mRNA ligands *MAP1B*, *α-CaMKII*, and *ARC* has been demonstrated in synaptoneurosomes from *Fmr1* knockout mice, a model system for fragile X synapses (Zalfa et al. 2003). However, potentially more deleterious than the elevated basal protein levels seen in fragile X synapses is the lack of regulation of mGluR-induced translation. In the process of long-term depression triggered by metabotropic glutamate receptors (mGluR-LTD), group I mGluR activation results in increased dendritic protein synthesis,

the internalization of AMPA receptors, and the weakening of synaptic connections. FMRP normally regulates mGluR-LTD by inhibiting the dendritic protein synthesis, but in fragile X, the process can continue in unabated fashion (Huber et al. 2002; Bear et al. 2004; Nakamoto et al. 2007). The consequent increase in LTD is believed to be associated with the deficiencies in learning and abnormal dendritic spine morphology seen in fragile X, as genetic and pharmacologic reductions in group I mGluR activity are able to correct these features in the mouse model (Dolen et al. 2007; de Vrij et al. 2008).

One important detail of this "mGluR theory" of fragile X syndrome is how FMRP normally responds to mGluR activation. It is known that the phosphorylation status of FMRP affects its association with polyribosomes. When FMRP is phosphorylated at the key residue serine-500, it associates with stalled polyribosomes; dephosphorylated FMRP, on the other hand, associates with actively translating polyribosomes (Ceman et al. 2003). In response to group I mGluR activation, FMRP is rapidly dephosphorylated by the phosphatase PP2A, enabling association with translating ribosomes and resulting in the increased translation of FMRP ligands (Narayanan et al. 2007). However, several minutes after group I mGluR activation, the mTOR signaling cascade is stimulated, which culminates in the S6K1-mediated phosphorylation of FMRP and subsequent repression of translation (Narayanan et al. 2008). These two opposing responses are believed to mediate the lack of mGluR activation-induced translation and the general increase in translation seen in fragile X synapses.

Another important detail of the "mGluR theory" of fragile X is which aberrantly translated proteins are responsible for the altered LTD. One protein that has received much recent attention is ARC. Encoded by a known FMRP mRNA ligand (Zalfa et al.

2003), ARC is a cytoskeletal protein that mediates the endocytic internalization of AMPA receptors (Chowdhury et al. 2006). In response to group I mGluR activation, ARC is translated in the synapse (Park et al. 2008; Waung et al. 2008). In the absence of FMRP, however, this rapid induction of ARC is impaired (Park et al. 2008). *Arc* knockout mice demonstrate reduced mGluR-LTD (Park et al. 2008), and blockade of ARC translation reduces the mGluR-mediated internalization of AMPA receptors (Waung et al. 2008). These data suggest that ARC may play a central role in the aberrant mGluR-LTD seen in fragile X syndrome. Other proteins whose dysregulation may contribute to altered mGluR-LTD include MAP1B, PSD-95, and APP, all of which have also been shown to affect AMPA receptor internalization (Hsieh et al. 2006; Davidkova and Carroll 2007; Xu et al. 2008).

Beyond its synaptic role as a regulator of local translation, FMRP also plays a role in RNA trafficking. The first clue of this was the description of functional nuclear localization and export signals in FMRP (Eberhart et al. 1996; Fridell et al. 1996; Sittler et al. 1996; Bardoni et al. 1997). These signals appear to mediate nucleocytoplasmic shuttling, as evinced by immunogold electron microscopy images of FMRP located within nuclear pores (Feng et al. 1997b). Structural similarities between FMRP and the RNA-binding protein hnRNP A1 led to the hypothesis that FMRP, like hnRNP A1, enters the nucleus to associate with RNA and other proteins and is exported to the cytoplasm as a ribonucleoprotein (mRNP) particle (Eberhart et al. 1996). Indeed, recent studies have shown that FMRP binds RNA in the nucleus, and the subsequent export of the FMRP-mRNA complex is aided by the RNA export protein Tap/NXF1 (Kim et al. 2009). The mRNP particles containing FMRP have been shown through co-immunoprecipitation to

include an assortment of RNA-binding proteins, such as FXR1, FXR2, nucleolin (Ceman et al. 1999), YB1/p50 (Ceman et al. 2000), Purα, and mStaufen (Ohashi et al. 2002). These mRNP particles have been shown to associate with microtubules as they are trafficked to dendrites in a translationally repressed state (Wang et al. 2008a). Activation of group I mGluRs stimulates the trafficking of the FMRP-containing mRNPs into dendrites (Antar et al. 2004) in an FMRP-dependent manner (Dictenberg et al. 2008). Thus, FMRP facilitates both the nuclear export and the intracellular trafficking of mRNA.

Several additional novel roles of FMRP have emerged over the past few years. First among these was the description of FMRP's interaction with the microRNA pathway (Caudy et al. 2002; Ishizuka et al. 2002; Jin et al. 2004). Although evidence suggests that miRNA may assist FMRP in the translational regulation of its specific targets (Jin et al. 2004), more research is needed to describe the mechanism of their interaction. A second novel role recently ascribed to FMRP is in axonal development. FMRP has been detected in the growth cones of developing axons; colocalized with it are the mRNA ligand *MAP1B* (Antar et al. 2006) and the miRNA-associated RISC (Hengst et al. 2006). These findings correlate with recently noted defects in axonal arborization in the *Fmr1* knockout mouse (Bureau et al. 2008), suggesting that FMRP plays an as-yet-uncharacterized role in axonal development. Additionally, FMRP has a third novel role in the modulation of neurotransmitter receptors other than AMPA receptors, namely GABA and dopamine receptors. The delta subunit of the $GABA_A$ receptor is encoded by an mRNA ligand of FMRP (Miyashiro et al. 2003) and exhibits reduced protein levels in *Fmr1* knockout mouse brains (D'Hulst et al. 2006), suggesting that its translation is regulated by FMRP. This results in abnormal GABAergic transmission in the *Fmr1*

knockout mouse (Centonze et al. 2008). Thus, inhibitory neurotransmission may play an understudied role in the fragile X phenotype. Dopaminergic neurotransmission is also impaired in the fragile X mouse. By unknown mechanisms, the subcellular localization of the kinase GRK2 is altered in the absence of FMRP. This seems to be the cause of D1 receptor hyperphosphorylation and the downstream effect of reduced surface expression of AMPA receptors. The administration of D1 agonists to the *Fmr1* knockout mice reduced hyperactive behaviors, leading to the hypothesis that the behavioral features of fragile X syndrome may result from aberrant dopamine signaling (Wang et al. 2008b).

1.6 *ASFMR1*

Several antisense transcripts that overlap *FMR1* have recently been identified in humans, non-human primates, and mice (Ladd et al. 2007; Khalil et al. 2008). Collectively attributed to the novel gene *ASFMR1* (Ladd et al. 2007; Gecz et al. 2009), these polyadenylated transcripts are expressed from two different promoters (Figure 1.5). The more 3' *ASFMR1* promoter is located in the vicinity of the *FMR1* promoter, while the more 5' *ASFMR1* promoter is located within intron 2 of *FMR1* (Ladd et al. 2007). Although it is unclear if the transcripts derived from these two distant promoters, separated by more than 10 kb of sequence, should truly be considered products of the same gene (Khalil et al. 2008), they appear to share a similar expression pattern with high levels of transcript in the brain and kidney and low levels in the heart and skeletal muscle (Ladd et al. 2007). Also uncertain is whether the products of *ASFMR1* are solely noncoding RNAs or whether a protein is translated from the transcripts. A putative open reading frame has been identified in mRNA transcribed from the more 5' promoter

Figure 1.5. Diagram of the overlapping locus of *ASFMR1* and *FMR1*.

The horizontal axes are formed by intronic sequence, and the numbered vertical spokes represent exons. Coding exonic sequence is shown in blue, while noncoding exonic sequence is shown in white. The black regions are the promoters of *ASFMR1* and *FMR1*. The 5' promoter of *ASFMR1* is located within intron 2 of *FMR1*. Its transcription product uses the same splice donor and acceptor sites as *FMR1* for the splicing out of intron 1. The coding exonic sequence of this *ASFMR1* transcript overlaps with the 5'UTR of *FMR1*, and thus includes the *FMR1* CGG repeat. The more 3' promoter of *ASFMR1* overlaps with the *FMR1* promoter; as depicted, it yields only noncoding RNA transcripts.

(Figure 1.5). The resultant 100 amino acid protein contains a polyproline stretch, encoded by the antisense CCG repeat, the size of which would be determined by the number of CGG repeats in *FMR1*. This protein is highly conserved in mammals, with 79.8% identity between human and mouse (Figure 1.6). The *in vivo* expression and possible functional role of this protein have not yet been examined (Ladd et al. 2007).

Much of the excitement about the *ASFMR1* gene comes from the finding that CGG repeat expansion impacts *ASFMR1* transcription in a similar fashion to its effect on *FMR1* transcription. Like *FMR1*, *ASFMR1* expression increases with increasing numbers of CCG/CGG repeats. Repeat expansion also causes a shift to the more 5' promoter in both genes. In fact, the more 3' promoter of *ASFMR1*, located near the *FMR1* promoter, is not used at all when associated with a premutation-sized repeat; rather, the more 5' promoter, located within intron 2 of *FMR1*, acts alone with increased activity. Furthermore, when CGG repeat expansion exceeds full mutation size, *ASFMR1* is transcriptionally silent, similar to *FMR1* (Ladd et al. 2007). These findings of differential expression have led to speculation that *ASFMR1* may play a role in fragile X syndrome or the fragile X premutation phenotype (Ladd et al. 2007; Gecz et al. 2009). Further studies will be necessary to determine if *ASFMR1* is indeed involved in the etiology of these disorders, and, if so, whether it is through an RNA- or protein-mediated mechanism.

1.7 Proposed Research

In 1991, mutations in the *FMR1* gene were identified to be the cause of fragile X syndrome (Pieretti et al. 1991; Verkerk et al. 1991). The novel mutational mechanism of trinucleotide repeat expansion was shown to be responsible for both the classic

```
H. sapiens      1    MNILYKCLVESAIGAPHFHHQLLHLLFSPASAGSPPPRGG   40
M. mulatta      1    MNILYKCLVESAIGAPHFHHQLLHLLFKKTRAGSPPPRGG   40
C. familiaris   1    MNILYKCLVESAIGAPHFHHQLLHLLESPARAGSPPPRGG   40
B. taurus       1    MNIFYKCLVESAIGAPHFHHQILHLLVRPARFGSPPSRGG   40
M. musculus     1    MDIFYKCLVESAIGAPHFHHQLLHLLVRP-RAGSPPPGGG   39
R. norvegicus   1    MDIFYKCLVESAIGAPHFHHQLLHLLVLP-RAGSPPPRGG   39
                     *:*:*****************:****      ****. **

H. sapiens      41   LRALEAQPPPPP--------PPPPPPPPPPPPPPPRCRTP   72
M. mulatta      41   LRALAAAAAAAAAAAAAARRPPPPPPPPPPPPPPRCRTP   80
C. familiaris   41   LRAQEAQ----------------PPPPPPPPPPPPRSRTP   64
B. taurus       41   LQALEAQ-----------------------PPPPRYRTP   56
M. musculus     40   LQALEAQ-------------------PPSPPPPPPRYRTP   60
R. norvegicus   40   LQALEAQ----------------------PSPPPPRYRTP   57
                     *:*  *                        ***** ***

H. sapiens      73   PGSGASVTAAARARR---RPAARSEAALHRK           100
M. mulatta      81   PGSGASVTAAARARR---RPAARSEAALHRK           108
C. familiaris   65   PGSGAAVTAATRARR---RLTARSEAAPHRK            92
B. taurus       57   PGSGAAVTAAACARL---RLAARSEAAPHRK            84
M. musculus     61   PGSGAAVTATARARRPFASLPALSGAAPHRK             91
R. norvegicus   58   PGSGAAVTDTARARRPLASLPALSGAAPHRK            88
                     *****:** :: **      .* * ** ***
```

Figure 1.6. Mammalian conservation of the putative *ASFMR1* ORF protein.

Comparison of the putative *ASFMR1* ORF protein across mammalian species. The amino acid sequences were aligned with Clustal. Asterisks denote identical amino acids. Colons denote conservative substitutions. Periods denote similar amino acid residues. With the exception of the variable length polyproline domain, the protein encoded by *ASFMR1* is largely conserved across mammals, particularly in the N-terminal region.

cytogenetic finding of a fragile site at chromosome Xq27.3 (Kremer et al. 1991) and the transcriptional silencing (Sutcliffe et al. 1992) that commonly leads to loss of *FMR1* function and fragile X syndrome. However, other types of mutations were also shown to cause a fragile X-like phenotype, including large deletions encompassing the gene (Gedeon et al. 1992) and conventional mutations changing the gene sequence. Interestingly, in the eighteen years since the gene was discovered, only three conventional mutations in *FMR1* have clearly been shown to cause fragile X: one missense mutation (De Boulle et al. 1993), one frameshift/nonsense mutation, and one splice junction mutation (Lugenbeel et al. 1995). Considering that a complete loss of gene function is compatible with life, one would expect that most mutations in *FMR1* would also be compatible with life. On the other hand, with only one missense change in *FMR1* known to be polymorphic in the population, one would not expect that such changes in *FMR1* would simply be tolerated without phenotypic effect. Thus, we hypothesize that conventional mutations in *FMR1* represent a heretofore understudied cause of disease.

Several groups have attempted to identify conventional mutations in *FMR1* in various patient populations, including patients with a fragile X-like phenotype (Chiurazzi et al. 1994; Reyniers et al. 1996; Wang et al. 1997; Gronskov et al. 1998; Castellvi-Bel et al. 1999), patients with autism (Vincent et al. 1996; Shinahara et al. 2004), patients with uncharacterized intellectual disability (Gronskov et al. 1998; Shinahara et al. 2004), and patients with X-linked intellectual disability (Tarpey et al. 2009). These attempts, however, were underpowered and generally used less sensitive techniques for mutation detection (Table 1.1). Therefore, the goal of this dissertation is to use novel sequencing

technologies on large samples of patients with either a fragile X-like phenotype or uncharacterized developmental delay in order to more thoroughly assess the frequency of conventional mutations in *FMR1*.

Aim 1: To assess the frequency of pathogenic *FMR1* point mutations in fragile X-like patients

Some clinical laboratories have begun to offer *FMR1* sequencing as a second-line diagnostic measure for patients presenting with developmental delay. Such clinical *FMR1* sequencing is generally ordered for patients who "look like fragile X." Our first aim was to assess the yield of this diagnostic approach. Therefore, we used resequencing arrays to sequence the promoter, exons, and splice junctions of *FMR1* in 51 unrelated fragile X-like males.

Aim 2: To develop a pooled-template method for the high-throughput resequencing of a candidate gene in a large patient population

The two greatest drawbacks with the previous attempts to identify conventional *FMR1* mutations in various patient populations were the number of patients assessed and the sensitivity of the detection techniques. With the advent of next-generation sequencing, both of these deficiencies can be readily improved upon. While Sanger sequencing costs approximately $0.50 per kilobase, Illumina Genome Analyzer (IGA) sequencing costs a mere $0.002 per kilobase, making large-scale sequencing projects more efficient and affordable (Shendure and Ji 2008). Because a single sequencing run with IGA provides an estimated 64 Megabases of sequence data, one can use this approach to obtain either great breadth or great depth of coverage. As our interest was to sequence a single candidate gene, *FMR1*, we had little use for breadth of coverage.

Rather, we sought to utilize the depth provided by this sequencing method to obtain sequence data from pools of templates from several patients. Through the work described herein, we have determined the utility and limitations of a pooled-template approach for targeted next-generation sequencing.

Aim 3: To assess the frequency of pathogenic *FMR1* sequence variants in a developmentally delayed clinical population

Although the current clinical application of *FMR1* sequencing is largely limited to fragile X-like patients, it is unclear if point mutations in *FMR1* would manifest in this phenotype. Thus, to investigate a less specific phenotype, we used a pooled-template method to sequence *FMR1* in 963 developmentally delayed males who had been referred to our clinical laboratory for fragile X testing. In the process, we were also able to create a more complete catalogue of normal variants in *FMR1*.

Aim 4: To determine the role of the *ASFMR1*-encoded protein in fragile X syndrome

ASFMR1, an antisense gene overlapping the *FMR1* locus, has been proposed to play a role in the etiology of fragile X syndrome, possibly due to the absence of the *ASFMR1*-encoded protein. To determine if the ASFMR1 protein is essential for normal cognitive function, we looked for sequence variants in both developmentally delayed and control males that would alter the function of the ASFMR1 protein. This application of the dataset generated in Aim 3 extended our findings to a second candidate gene for intellectual disability.

Chapter 2. Point mutations in *FMR1* are not a major cause of fragile X syndrome

2.1 Introduction

Fragile X syndrome (FXS), an X-linked dominant disorder with reduced penetrance, is the most frequent inherited form of intellectual disability. In 1991, the common causal mutation in FXS was identified to be a large CGG trinucleotide expansion in the 5'-untranslated region of the gene *FMR1* (Verkerk et al. 1991). Shortly thereafter, several groups identified *FMR1* deletions in FXS patients, suggesting that multiple mutational mechanisms could give rise to the disorder (Gedeon et al. 1992; Wohrle et al. 1992; Tarleton et al. 1993). The subsequent identification of an *FMR1* missense mutation in a severely affected FXS patient validated the role of *FMR1* in FXS and suggested that yet another class of *FMR1* mutation may be a significant cause of disease (De Boulle et al. 1993). However, while both trinucleotide repeat expansion (Garber et al. 2008) and *FMR1* deletions (Coffee et al. 2008) have proven to be common causes of FXS, no additional missense mutations have been identified.

Several groups have previously attempted to identify additional *FMR1* missense mutations in patients presenting with an FXS-like phenotype (Chiurazzi et al. 1994; Reyniers et al. 1996; Wang et al. 1997; Gronskov et al. 1998; Castellvi-Bel et al. 1999). However, as summarized in Table 2.1, these previous studies were not designed to comprehensively evaluate the frequency of *FMR1* missense mutations in FXS, but rather

Table 2.1. Previous screens for missense mutations in *FMR1* in FXS-like patients

***FMR1* Coverage**	**Mutation Detection Method**	**N**	**Reference**
Complete coding sequence	cDNA sequencing	3	Chiurazzi et al 1994
Complete coding sequence	cDNA sequencing	1	Reyniers et al 1996
Exons 1-10, 15	SSCP	27	Wang et al 1997
Exons 2-17	SSCP	31	Castellvi-Bel et al 1999
Complete coding sequence	SSCP	6	Gronskov et al 1998

served as mutational screens. Three of the studies surveyed fewer than ten FXS-like patients (Chiurazzi et al. 1994; Reyniers et al. 1996; Gronskov et al. 1998), while the other two studies used less proven detection methods to survey only a portion of the *FMR1* coding sequence (Wang et al. 1997; Castellvi-Bel et al. 1999). Furthermore, *FMR1* sequencing is rarely performed in the clinical setting, so the lack of case reports detailing individuals with coding changes in *FMR1* is rather uninformative regarding the frequency of such mutations.

In this study, we used array-based resequencing to search for point mutations in *FMR1* in a population of 51 unrelated FXS-like males. Despite achieving a high level of sequence coverage and accuracy, we did not identify any missense variants in *FMR1*, nor did we identify any novel noncoding variants likely to have a functional effect. Our method did, however, identify a pathogenic *FMR1* deletion in a patient with FXS.

2.2 Subjects and Methods

2.2.1 Subjects and Samples

In the clinical setting, *FMR1* sequencing is most frequently pursued for patients presenting with typical fragile X features but lacking *FMR1* repeat expansion. Therefore, we chose to examine the frequency of *FMR1* mutations in this patient population. We recruited 51 unrelated intellectually disabled males who exhibited at least two of the FXS-like features listed in Table 2.2. Forty-seven of the patients were of European descent and four were of African descent. A focused clinical history and either a blood or saliva specimen were obtained from each patient. DNA was extracted from the obtained

Table 2.2. Phenotypic characteristics of FXS-like patients

Characteristic	Examples
FXS-like facial features	Elongated face, everted ears, macrocephaly
Macroorchidism	
Connective tissue abnormalities	Hyperextensible finger joints, velvety skin, or recurrent ear infections
Shyness or poor eye contact	
Attention deficit/ hyperactivity	
Language delay	
Repetitive behaviors	Hand flapping, hand biting
Evidence of X-linked inheritance	Similarly affected male sibling, affected second-degree male relative through maternal lineage

Patients enrolled as FXS-like exhibited at least two of these characteristics

specimens using standard methods. All patients and/or legal guardians gave informed consent to participate in this study, which was approved by the Emory University Institutional Review Board.

2.2.2 *FMR1* LR-PCR

Four long range PCR (LR-PCR) amplifications were designed to target *FMR1* (Figure 2.1). The LR-PCR primer pairs were as follows: *FMR1*A-F: 5'-CAGACTGCGCTACTTTGAACC-3' and *FMR1*A-R: 5'-CTACATACCAACAAACGCACTACTGCTACAT-3'; *FMR1*B-F: 5'-AATTTCCAGTATACTTGTCTATTTTTCGAGATG-3' and *FMR1*B-R: 5'-TTTTGGGAGATAGCTACCTACAGGGTATCTGATT-3'; *FMR1*C-F: 5'-GTTGAACATTAAATTGCAGTTCAGAATACATAG-3' and *FMR1*C-R: 5'-GAGACATATCCAATCCACTTGCCGTTATAGT-3'; *FMR1*D-F: 5'-AATAATCTGATACGTTTAAAAGGTTGCTATTGA-3' and *FMR1*D-R: 5'-TTAATATGGTTTAGTGGCACCCTATGTAATAAA-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), 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-

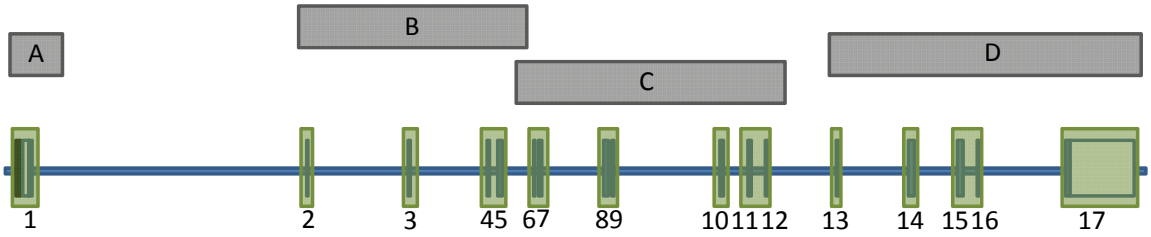


Figure 2.1. Diagram of *FMR1*.

The horizontal axis is formed by intronic sequence, and the numbered vertical spokes represent the 17 exons of *FMR1*. Coding exonic sequence is shown in blue, while noncoding exonic sequence is shown in white. The black region upstream of exon 1 is the minimal promoter of *FMR1*. The grey bars represent the four LR-PCR amplicons used for sequencing. The green boxes represent the *FMR1* regions sequenced with the custom resequencing array.

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 confirmed by gel electrophoresis.

2.2.3 Sequencing-by-hybridization

FMR1 sequencing was performed on Custom Resequencing Arrays (Affymetrix, Santa Clara, CA), designed to provide coverage of all 17 *FMR1* exons and the *FMR1* promoter, plus at least 200 bp of flanking intronic sequence (Figure 2.1). Patient sample amplicons were processed for sequencing-by-hybridization according to the Affymetrix CustomSeq Resequencing Array protocol, Version 2.1, with the following exceptions. The four LR-PCR amplicons per patient were pooled in equimolar fashion to a total of 4 µg and digested with 0.2 units of DNase I (Promega, Madison, WI) at 37°C for 3 minutes, yielding digestion products between 100 – 600 bp. Labeling, hybridization, and array processing were performed as per the protocol.

2.2.4 Variant Detection and Confirmation

Base-calling was performed with the ABACUS statistical method (Cutler et al. 2001) using the POPGEN genotyping software (Okou et al. 2007). Putative variants were confirmed by traditional Sanger sequencing of fresh LR-PCR amplicons. Both POPGEN

data and DNA chromatograms were inspected manually with the SeqScape software (Applied Biosystems, Foster City, CA).

2.2.5 Western Blotting

Immunoblotting was performed using standard methods. Briefly, patient and control lymphoblastoid cells were lysed with a standard Triton X-100-based lysis buffer. The lysate protein concentrations were measured with the Bradford assay. Proteins were denatured by heating at 95°C for 3 minutes and separated by polyacrylamide gel electrophoresis and transferred to a nitrocellulose membrane. To assess protein loading and transfer, the membrane was reversibly stained with Ponceau S. The membrane was blocked for one hour in blocking buffer (10 g dry milk, 200 µl Tween-20, and 100 ml PBS), probed with primary antibody (anti-FMRP 1a or anti-eIF4e) overnight, and probed for one hour with horseradish-peroxidase conjugated anti-mouse secondary antibodies. Proteins were detected by chemiluminescence (ECL, GE Healthcare, Piscataway, NJ).

2.3 Results

2.3.1 Sequence Accuracy

Across the 51 FXS-like patients sequenced by array hybridization, 99.6% of bases were called with high reliability, as determined by a quality score of 30 or greater. The high level of sequence accuracy achieved is further demonstrated by the identification of known polymorphisms. As seen in Table 2.3, we detected all seven SNPs catalogued in dbSNP (build 130) for which the population frequency has been measured in HapMap

Table 2.3. Detection of known polymorphisms in *FMR1* by array resequencing

SNP	FXS-like patient frequency	Weighted HapMap frequency	p-value
rs25726	0.176	0.073	0.23
rs25731	0.078	0.062	1
rs25707	0.137	0.072	0.53
rs29281	0.039	0.007	0.50
rs25714	0.078	0.084	1
rs29285	0.039	0.007	0.50
rs25704	0.353	0.280	0.52

P-values reflect the result of Fisher's exact tests.

samples. For the sake of comparison, we weighted the HapMap frequency data by the racial distribution of our patient population. None of the SNPs were found to be at a statistically different frequency in the FXS-like patients from in the HapMap controls, thus suggesting that the *FMR1* resequencing arrays reliably detect sequence variants.

2.3.2 Novel *FMR1* Sequence Variants

Notably, no novel variants were detected in the *FMR1* coding sequence in the population of 51 FXS-like males. However, two novel intronic variants, c.52-47A>G and c.105-179G>T, were identified in *FMR1* (Table 2.4). As an assessment of possible functional relevance, we examined the mammalian conservation of these nucleotide positions and their genomic regions using phyloP and phastCons scores, respectively (Siepel et al. 2005). Because both variants are located in poorly conserved genomic regions (phastCons of 0.01), it is likely that they represent novel rare polymorphisms.

2.3.3 Array-based Deletion Detection

In addition to detecting point mutations, resequencing arrays allow the detection of deletions. In one FXS-like patient, we identified a 355 bp deletion extending from 220 bp upstream of the CGG repeat through the second codon of the *FMR1* coding sequence. After confirming this deletion with Sanger sequencing, we assessed its effects on FMRP translation. As shown in Figure 2.2, immunoblot analysis of patient lymphoblastoid cell line lysates revealed an absence of FMRP expression.

2.4 Discussion

Table 2.4. Novel *FMR1* sequence variants identified in FXS-like males

Location	cDNA Variant	PhastCons	PhyloP	Patient Frequency
Intron 1	c.52-47A>G	0.01	1.27	1/51
Intron 2	c.105-179G>T	0.01	1.06	1/51

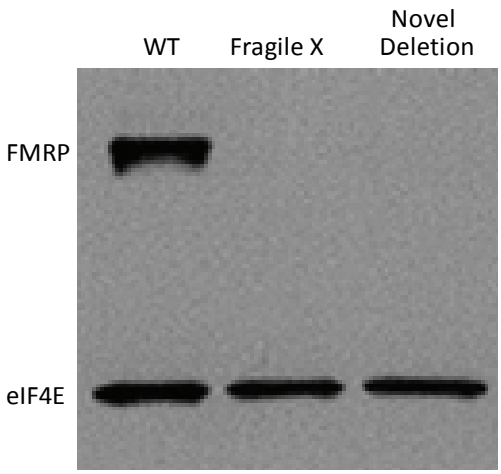


Figure 2.2. FMRP absent in patient harboring a 5'UTR deletion in *FMR1*.

Western blot of lymphoblastoid cell lysate from a healthy control, a fragile X patient, and a patient harboring a novel deletion in the 5'UTR of *FMR1*. The protein eIF4e was assessed as a loading control.

Through the use of resequencing arrays, we have sequenced the promoter, exons, and splice junctions of *FMR1* in 51 unrelated patients with several classic features of fragile X syndrome. We identified two novel intronic variants which likely have no functional effect, but notably detected no missense or promoter mutations. As the largest sequencing analysis of FXS-like patients performed to date, these data suggest that *FMR1* sequence variants are not a common cause of the FXS phenotype.

At the present time, two missense changes in *FMR1* have been identified, the polymorphic p.A145S (rs29281) and the p.I304N mutation previously detected in a severely affected FXS-like patient (De Boulle et al. 1993). It would be surprising if these are the only missense changes that occur in *FMR1*. In comparison, over 100 distinct point mutations in the nearby gene *MECP2* have been shown to cause Rett syndrome, despite the fact that the gene is smaller and more recently identified than *FMR1* (Chahrour and Zoghbi 2007). Furthermore, because a functional absence of the *FMR1* gene product is compatible with life, albeit associated with the symptoms of FXS, we would hypothesize that missense changes in *FMR1*, which in many cases would be less damaging than a loss-of-function, would not lead to embryonic lethality.

If *FMR1* missense mutations are truly more frequent than the current data suggest, it is likely that the phenotypic consequence of such mutations is distinct from FXS. One possibility is that missense mutations lead to a more subtle developmental phenotype, such as autism or nonspecific intellectual disability. While this possibility has fruitlessly been investigated by several groups, the known etiological heterogeneity of these phenotypes would require a much greater sample size than has been evaluated (Vincent et

al. 1996; Gronskov et al. 1998; Shinahara et al. 2004). Another possibility is that an *FMR1* missense mutation would alter the function of only one domain of FMRP, thereby causing a specific FXS-like symptom, such as connective tissue defects or attention-deficit/ hyperactivity disorder, in the absence of an overall FXS-like phenotype. This possibility has not, to our knowledge, been previously investigated.

Finally, the current study reiterates the known importance of *FMR1* deletions in FXS. The deletion we identified extends from 220 bp upstream of the CGG repeat through the second codon of the *FMR1* coding sequence, and results in the absence of FMRP expression in patient tissues. While it's exact breakpoints are unique, this deletion belongs to a well-characterized class of deletions that result from the instability of the CGG trinucleotide repeat (de Graaff et al. 1995; Coffee et al. 2008). This study is consistent with the notion that *FMR1* deletions play a larger role in FXS than *FMR1* missense changes.

Chapter 3. Development of Pooled-template Massively Parallel Sequencing for the Identification of Novel Sequence Variants in a Disease Gene

3.1 Introduction

Massively parallel sequencing (MPS) technologies vastly improve 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 these new technologies 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 is generally inefficient and thus is rarely performed. 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.

Although pooled-template MPS has previously been shown to provide adequate sensitivity for rare variant detection (Ingman and Gyllensten 2009; Koboldt et al. 2009), its usefulness and efficiency for novel variant detection have not been fully demonstrated.

In the first presentation of this approach, Ingman and Gyllensten presumed that erroneous base calls by MPS would be caused by PCR errors and thus would rarely recur at the same nucleotide position in multiple sequence reads. Therefore, the authors proposed that adequate depth of coverage for each pooled template would be sufficient to prevent false positives (Ingman and Gyllensten 2009). However, it has become clear that each MPS technology has systematic biases in sequencing which result in recurring errors (Dohm et al. 2008; Harismendy et al. 2009). These recurring errors would likely create more false positives in pooled-template MPS, where a low-frequency recurring error would resemble a low-frequency sequence variant. For pooled-template MPS to be a practical, efficient approach for novel variant detection, it is important to minimize the occurrence of false positives and the resultant need for excess secondary confirmation.

In this study, we used pooled-template MPS to identify sequence variants in *FMR1*, the gene implicated in fragile X syndrome, in a population of 963 developmentally delayed males. By determining the basal error rate of MPS on the Illumina GA platform, we were able to develop a novel variant detection algorithm which uses base-specific sequencing error rates to enhance the reliability of variant calls. Through this approach, we demonstrate that pooled-template MPS provides an adequate positive predictive value for efficient use in novel variant detection.

3.2 Material and Methods

3.2.1 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 tested negative for FXS 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, CLIA-certified methods. The patient's race was noted when available. For the 1392 samples obtained, racial identification was only available for 370 (26.6%). Among these, 251 (67.8%) were Caucasian, 111 (30.0%) were African-American, and 8 (2.2%) were Asian-American. For the subset of 963 patients that were sequenced, racial identification was only available for 241 (25.0%). Among these, 164 (68.1%) were Caucasian, 74 (30.7%) were African-American, and 3 (1.2%) were Asian-American. The Emory University Institutional Review Board approved this use of deidentified clinical samples.

3.2.2 LR-PCR Targeting of *FMR1*

Four long range PCR (LR-PCR) amplifications were designed to target *FMR1*. The LR-PCR primer pairs were as follows: *FMR1*A-F: 5'-CAGACTGCGCTACTTTGAACC-3' and *FMR1*A-R: 5'-CTACATACCAACAAACGCACTACTGCTACAT-3'; *FMR1*B-F: 5'-AATTTCCAGTATACTTGTCTATTTTTCGAGATG-3' and *FMR1*B-R: 5'-TTTTGGGAGATAGCTACCTACAGGGTATCTGATT-3'; *FMR1*C-F: 5'-GTTGAACATTAAATTGCAGTTCAGAATACATAG-3' and *FMR1*C-R: 5'-GAGACATATCCAATCCACTTGCCGTTATAGT-3'; *FMR1*D-F: 5'-AATAATCTGATACGTTTAAAAGGTTGCTATTGA-3' and *FMR1*D-R: 5'-

TTAATATGGTTTAGTGGCACCCTATGTAATAAA-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), 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. After LR-PCR, 963 samples remained for sequencing.

3.2.3 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). 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 were 172.1 ng of LR-PCR-C from a patient with a known missense mutation in *FMR1* (De Boulle et al. 1993). This 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 μg.

3.2.4 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).

3.2.5 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 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.

3.2.6 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." The measured "minor allele frequency" at each nucleotide position was used as the base error rate for the pooled sequencing runs. For each pool, the likelihood that the minor allele frequency at a given base position differed significantly from the base error rate was determined by using a binomial distribution model. The threshold for variant calling was empirically set at $p < 1 \times 10^{-4}$.

3.2.7 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.

3.3 Results

3.3.1 Illumina GA Performance

In all, we sequenced 51 amplicon pools with the Illumina GA. The average depth of coverage was 130.5 ± 51.1 (mean ± 1 S.D.) per patient, across all bases sequenced. As seen in Table 3.1, there is a large standard deviation for the depths of LR-PCR amplicon A; this variation derives from several pools in which the depth for LR-PCR-A was exceptionally great. Furthermore, there were significant differences in average depth between the four LR-PCR amplicons (ANOVA: $p = 3.6 \times 10^{-5}$). Despite these differences, however, every amplicon provided a minimum average of 30-fold coverage for each pooled individual. Such depth has proven adequate for diploid sequencing (Bentley et al. 2008), and thus was inferred to be acceptable for our haploid X chromosome sequencing in males.

3.3.2 GA Sequencing of *FMR1* in a Single Individual

Table 3.1. Sequencing depth across pooled amplicons

LR-PCR Amplicon	Mean depth (±1 S.D.)	Mean depth per patient (±1 S.D.)
A	3423 ± 2407	171 ± 127
B	2263 ± 922	113 ± 49
C	2273 ± 863	114 ± 45
D	3000 ± 860	150 ± 45

In order to assess the base-by-base error rate of MPS for the *FMR1* gene, we performed Illumina GA sequencing on LR-PCR amplicons derived from a single individual. Each base was sequenced an average of 1367 times. The percentage of incorrect sequence reads at each base position was subsequently used as the error rate for that base. Although 95.2% of the bases sequenced had less than 0.5% incorrect reads, 1088 bases had between 0.5% and 1% incorrect reads, 104 bases had between 1% and 2% incorrect reads, and 62 bases had more than 2% incorrect reads, as seen in Figure 3.1.

3.3.3 Variant Detection

To assess the sensitivity of our method for pooled-template MPS, 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 reflects the frequency of successful detection of other singleton variants in a pool. At our detection threshold, p.I304N was detected in 39/51 pools, giving a sensitivity of 76.5%.

Another important measure of the performance of pooled-template MPS is the positive predictive value (PPV). PPV is defined as the number of true variants divided by the total number of putative variants identified. However, many of the putative variants identified by pooled-template MPS were recurring (i.e. called in multiple pools). Sanger sequencing confirmation revealed that recurring variant calls were either confirmed in every pool or determined to be a false positive in every pool. Therefore, to reduce the influence of common variants and recurring errors, we defined a functional PPV, in

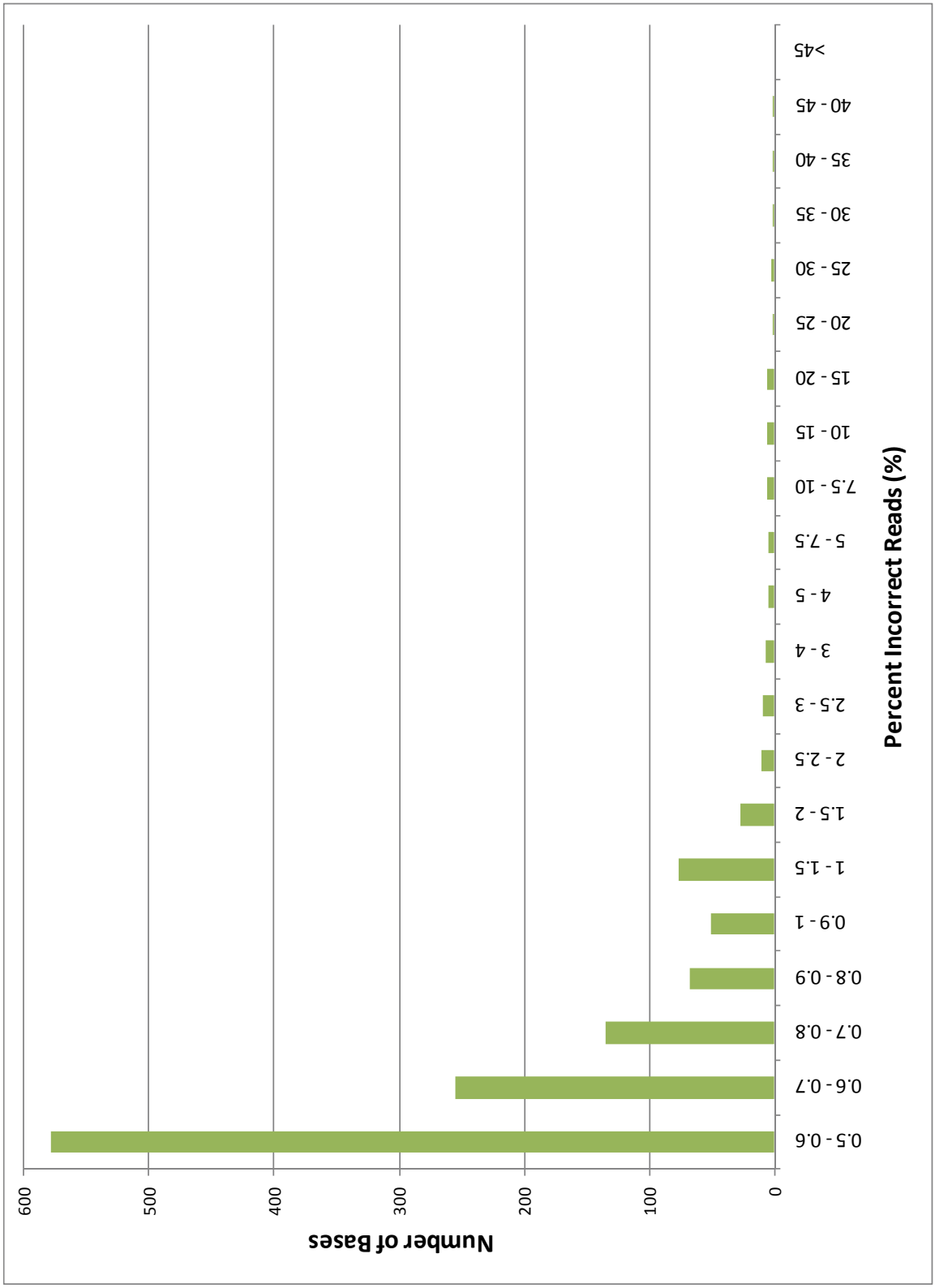

Number of Bases
600
500
400
300
200
100
0
Percent Incorrect Reads (%)
0.5 - 0.6
0.6 - 0.7
0.7 - 0.8
0.8 - 0.9
0.9 - 1
1 - 1.5
1.5 - 2
2 - 2.5
2.5 - 3
3 - 4
4 - 5
5 - 7.5
7.5 - 10
10 - 15
15 - 20
20 - 25
25 - 30
30 - 35
35 - 40
40 - 45
>45

Figure 3.1. Base-by-base error rate for pooled-template MPS.

The number of bases showing varying percentages of incorrect reads by Illumina GA sequencing, as assessed from a single-sample sequencing experiment. The existence of error-prone bases reinforces the notion of systematic sequencing errors in MPS and underscores the importance of accounting for these errors when detecting novel variants.

which the number of distinct true positives was divided by the number of distinct variant calls. In our study population, this was determined to be 51.5%.

To assess the nature of the false positives identified by our sequencing approach, we first segregated the uniquely-occurring false positives from the recurring errors. As seen in Figure 3.2, there were slightly more recurring errors than uniquely-occurring false positives. Because the single-individual sequencing run was intended to account for bases with high error rates, we investigated whether any of the false positives occurred at noted error-prone bases. Among unique false positives, 13.3% occurred at bases identified as having a sequencing error rate greater than 1%, while 25.3% occurred at bases identified as having a sequencing error rate greater than 0.5%. In contrast, 35.8% of recurring false positives occurred at bases identified as having a sequencing error rate greater than 1%, while 55.8% occurred at bases identified as having a sequencing error rate greater than 0.5%.

3.4 Discussion

In the current study, we have demonstrated that pooled-template MPS is an appropriate method for the detection of novel variants. Similar to previous publications (Ingman and Gyllensten 2009; Koboldt et al. 2009), we have demonstrated that rare variants can be detected by MPS as singletons in a pool. Extending these previous studies, we have shown that the PPV obtained from pooled-template MPS is adequately high for its use as a screening method. Essential to its usefulness is the fact that over 50% of called variants are true sequence changes, thereby reducing the number of confirmatory studies performed on false positive putative variants.

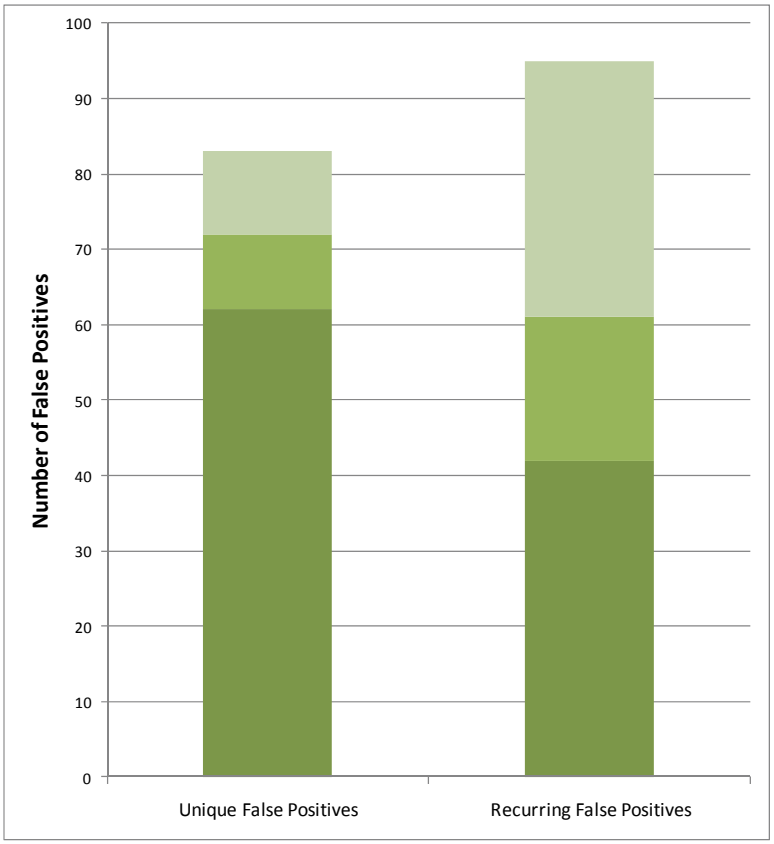


Figure 3.2. Error-prone bases are common sites of false positives.

The number of unique and recurring false positives detected in *FMR1* across 51 pools. Shown in dark green are the false positives occurring at bases with an error rate less than 0.5%. Shown in intermediate green are the false positives occurring at bases with an error rate between 0.5% and 1%. Shown in light green are the false positives occurring at bases with an error rate greater than 1%.

One important aspect of our method that improves the PPV of pooled-template MPS is the initial assessment of the base-by-base error rate. Because of the systematic biases in the Illumina GA sequencing technology (Dohm et al. 2008; Harismendy et al. 2009) and recurring errors in the mapping of short sequence reads (Li et al. 2008), assuming a constant error rate for all bases is inappropriate. We found that nearly 5% of the bases sequenced had an error rate of 0.5% or greater, and over 150 bases had error rates above 1%. By identifying putative variants as those base positions where the minor allele frequency in a pool is significantly different than this more refined estimate of erroneous reads, we likely reduced the number of false positives substantially.

Despite our adjustment for base-by-base error rate, a large proportion of the false positives that we detected still occurred at error-prone bases. This suggests that improvements can be made to our analysis approach. One possible improvement would be to perform multiple single-individual sequencing runs and average the error-rate across experiments. This would undoubtedly create a more accurate estimate of the base-by-base error rate. Another possibility would be to discard known error-prone bases from subsequent confirmatory analyses. This, however, may result in the omission of valuable data or necessitate the use of a secondary method to obtain the missing sequence information, thereby reducing the usefulness of the pooled-template MPS approach.

Although we achieved an acceptable sensitivity of 76.5%, our data suggest that accurate equimolar pooling is essential for the performance of pooled-template MPS. While some applications of this method may be more tolerant of false negatives, missed variants would be highly concerning in studies of rare variants or in clinical applications. More accurate concentration measurements and the avoidance of small-volume pipetting

would likely improve the sensitivity of pooled-template MPS. Moreover, it is possible that improvements to the PPV could also reduce the number of false negatives by allowing a lower threshold to be set for variant detection.

One technological development in MPS that emerged while we were performing the current study was sample indexing (Craig et al. 2008). In this process, sequencing libraries are created with unique adapter sequences. Several libraries are then pooled together for a given sequence run. The unique adapters allow the source library to be identified for each sequence read, thus enabling ready identification of the individuals within a pool. Therefore, sample indexing eliminates inaccurate pooling as a source of false negatives, as a sample that is underrepresented in a pool will be noted in the sequence output and can be addressed appropriately.

While sample indexing is undoubtedly an important development for MPS, our method of pooled-template MPS offers two distinct advantages over indexed MPS, as currently available. First, the pool size of indexed MPS is limited by the number of unique adapter sequences, which currently is twelve. On the other hand, the pool size of pooled-template MPS is only limited by the constraints of sequence depth and sequencing error rate. In the current study, we pooled 19 patient samples and one control sample, and achieved a two-fold greater depth than necessary. Second, the creation of individual sequencing libraries for each sample carries a substantial burden of time and expense. For smaller projects, this may not be an issue. However, in the current study, our method reduced the number of sequencing libraries by twenty-fold, making it much less costly to sequence nearly a thousand samples by MPS.

Finally, while this manuscript was in preparation, Out and colleagues published a manuscript detailing a similar approach to pooled-template MPS (Out et al. 2009). Notably, they used the SHORE software (Ossowski et al. 2008) instead of MAQ for sequence alignment and used a Poisson-distribution model rather than a binomial-distribution model for identifying variants. Additionally, they chose to use a constant error rate across all bases instead of using a base-by-base assessment of sequencing error. Despite this difference, Out and colleagues achieved a higher PPV than we achieved in the current study (88.2% vs. 51.5%). However, Out and colleagues also demonstrated a lower sensitivity for singleton detection (60.0% vs. 76.5%). This suggests that the higher PPV achieved by Out and colleagues came at the cost of singleton sensitivity. The lower sensitivity in the Out et al paper may result from the larger pool size they employed (88 vs. 20) to interrogate a smaller genomic region (5.7 kb vs. 25.9 kb). Thus, it remains to be seen what pool size is optimal for reliable singleton detection with an acceptably high PPV by pooled-template MPS.

In summary, we have developed a pooled-template MPS method for novel rare variant detection in a candidate disease gene. While contingent upon accurate equimolar pooling, our method achieves a relatively high sensitivity and an acceptably high PPV. It is likely that continued increases in the sequencing capacity of MPS will make this approach even more valuable in the future for the efficient use of MPS technology.

Chapter 4. Novel missense and promoter variants in *FMR1* are associated with developmental delay

4.1 Introduction

Fragile X syndrome (FXS) 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 macroorcchidism and an elongated face with large everted ears. However, due to the subtlety and reduced penetrance of the more distinguishing characteristics, the identification of a causal mutation is necessary for the diagnosis of FXS (Garber et al. 2008).

FXS is most frequently caused by the expansion of the CGG trinucleotide repeat located within the 5'-untranslated region of the *FMR1* gene (Verkerk et al. 1991; Ashley et al. 1993b). Repeat expansion 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 have also been shown to be a common cause of FXS (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 have been reported in the literature (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. However, methodological constraints have previously prevented a thorough assessment of *FMR1* sequence variation in a large number of patients, leaving the true frequency of pathogenic sequence variants in *FMR1* unknown (Chiurazzi et al. 1994; Reyniers et al. 1996; Vincent et al. 1996; Wang et al. 1997; Gronskov et al. 1998; Castellvi-Bel et al. 1999; Shinahara et al. 2004).

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 determine if *FMR1* sequence variants are a significant cause of morbidity, we employed pooled-template MPS to sequence the promoter, all 17 exons, and a substantial portion of the intronic sequence of *FMR1* in 963 developmentally delayed males. 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 expression of *FMR1*. Together, these four novel variants suggest that there is significant clinical utility in diagnostic *FMR1* sequencing for developmentally delayed males.

4.2 Material and Methods

4.2.1 Clinical Population

While novel 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 all patients who had tested negative for *FMR1* 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 *FMR1* repeat expansion, patients in this clinical population do not necessarily exhibit the classic FXS phenotype, but rather represent the more general diagnosis of developmental delay. 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 relate to transmission risk or a premutation-like tremor/ataxia phenotype than developmental delay. For the 963 patients sequenced, racial identification was only available for 241 (25.0%). Among these, 164 (68.1%) were Caucasian, 74 (30.7%) were African-American, and 3 (1.2%) were Asian-American.

4.2.2 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 tested negative for FXS from April 2002 to August 2007. In total, 1392 aliquots were obtained. The patient's race was noted when available. The genomic DNA samples had previously been extracted from whole blood by standard, CLIA-certified methods. The Emory University Institutional Review Board approved this use of deidentified clinical samples.

4.2.3 Massively-Parallel Sequencing

Pooled-template MPS was performed as previously described in Chapter 3.2.

4.2.4 Control Genotyping

The control samples used for the genotyping of novel variants were obtained from the NIMH Human Genetics Initiative. All controls were Caucasian adult males 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 (Table 4.1) 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. 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

Table 4.1. *FMR1* Sequenom genotyping assay primers

Variant	Amplification Primer 1	Amplification Primer 2	Extension Primer
c.-293T>C	ACGTTGGATGCGGGGCCGAGGGGCTGAG	ACGTTGGATGGTGAAACCGAAACGGAGCTG	aCACGTGACGTGGTTTCAG
c.18G>T	ACGTTGGATGAGGGCTGAAGAGAAGATGGA	ACGTTGGATGAGTACCTTGTAGAAAGCGCC	GATGGAGGAGCTGGTGGT
c.52-47A>G	ACGTTGGATGTGCCTGTGTAAGAAAAAACG	ACGTTGGATGAAAACGTTTGGTATCACTG	gggtgACTTGTGAGCTTAAAGATAGT
c.105-8A>C	ACGTTGGATGATCTGCCTATCAGGCTGCCA	ACGTTGGATGTCAGTTAAACATGAAAAGC	AGGCTGCCAGCTGTAAA
c.199-27A>T	ACGTTGGATGCCTCGATATCTGAAAATCTG	ACGTTGGATGGCTCTGGAATACACCTACAC	GATTTCAAAATTATGTTAATCATGAA
c.375T>C	ACGTTGGATGCGTCTAAGATCTGTTAATCCC	ACGTTGGATGATCACTTCCAAGTTCTCAGC	CCTGCCACAAAAGATAC
c.413G>A	ACGTTGGATGCGTCTAAGATCTGTTAATCCC	ACGTTGGATGATCACTTCCAAGTTCTCAGC	ATGTGCCAGAAGACTTAC
c.433G>T	ACGTTGGATGTTAAATTTCTAGGTGTGCC	ACGTTGGATGGTTACAGAAAAGGCACCAAC	CTAGGTGTGCCAAAGAG
c.630+438A>C	ACGTTGGATGAACTCCTGTCTTCAGATTCC	ACGTTGGATGGATAAAATGCACCTTAAGAG	CTTCAGATTCCCGCCAG
c.630+569A>G	ACGTTGGATGACTATCAAGTACGTCCATCC	ACGTTGGATGGAATATAGTAGAGCTGGCAC	cgCAAGTACGTCCATCCATTACTAT
c.631-840G>A	ACGTTGGATGGTGTTCAGTTCTAGCATGCC	ACGTTGGATGGTTCCTATAAGAAATTCCC	gtgcAAACCAAACGTATTAGTGG
c.651G>A	ACGTTGGATGGCTAGACCCATCAGATCTTC	ACGTTGGATGTTCTCTTCAGAGTTCAAGGC	CGATAAACTGTTCATGAAATCT
c.880+885A>G	ACGTTGGATGCATATCAAAACCAGGTGCAG	ACGTTGGATGAGGCAGTATGTCCCAATTTT	ctTGCAGTTTTATCAAGAAAGCTAC
c.990+4T>C	ACGTTGGATGGGTGAGGATTGAGGCTGAAA	ACGTTGGATGTATGTGCCACAAAATATTC	ATGTTCCACAAGAAGAGGTA
c.1126-78_1126-77 dupAAAAGTCCTGCAGT	ACGTTGGATGCTAATCTGTTTAGAAATGGG	ACGTTGGATGATACACAGAAGCCTGAACGC	GTTTAAAAGTCCTGCAGT
c.1276-325C>T	ACGTTGGATGACTGCTAAGGCACAGATATG	ACGTTGGATGGGTTAGCTGGTTATACCTTG	ccGGCACAGATATGCTATTTTG
c.1471+274T>A	ACGTTGGATGGATGATTTTCTTAAGGCCTC	ACGTTGGATGTGAGCCACAGTACACAAATG	CTTAAGGCCTCTCCTGAT
c.1471+420T>C	ACGTTGGATGCTCTTGTGGCTCATAAGGTG	ACGTTGGATGCTTCGCCTACTTCTTGACTG	tAGGTGATTTTGAGTGACA
c.1471+624C>T	ACGTTGGATGCTTGCTAATGATGGTATAAGG	ACGTTGGATGGGCTTTTCCATCTTAAAGAG	TGGTATAAGGTATAATCCATTTCA
c.1471+625G>C	ACGTTGGATGGGCTTTTCCATCTTAAAGAG	ACGTTGGATGCTTGCTAATGATGGTATAAGG	TCCTTTTTGAAATGCAAATATG
c.1472-521C>G	ACGTTGGATGTCAGGGAAGGGTCTCTTGTG	ACGTTGGATGTTGTGGACCAAACATCAGGC	GGTCATGGATTGGATCTA
c.*23T>C	ACGTTGGATGCCACTCGTGAATGGAGTACC	ACGTTGGATGGAATAAGAATTACGGAAATGG	ACTGCATAATTCTGAAGTTATA
c.*32C>G	ACGTTGGATGGGCCTAACAAAGTTTTCTAA	ACGTTGGATGCCCTAAACTGCATAATTCTG	TGGAATAAGAATTACGGAAATG
c.*60G>C	ACGTTGGATGGCCATCTTGCCTACTATTTG	ACGTTGGATGAATGGAGTACCCTAAACTGC	cttcGTCTTTGGCCTAACAAAGTTTT
c.*541T>C	ACGTTGGATGTCCTACCTGCCAAGATATAA	ACGTTGGATGGTGTGTGATGTGACATTTG	gaggaCAAGATATAAAAAAGGCACAG
c.*746T>C	ACGTTGGATGTATCAAACTTCAGTGCAAC	ACGTTGGATGCATAAACATCAGGTTAGGCAG	CAAACTTCAGTGCAACAAAACAA
c.*1482C>A	ACGTTGGATGAAATGGCAACAAACTGCAC	ACGTTGGATGGAAATCAACATTAATGTTTG	ccACAAACTGCACATGATTTCA
c.*1867G>A	ACGTTGGATGTAAGCTAGGAAAAGAAATC	ACGTTGGATGTCCTATGACGTGAAATTTCC	aTAGGAAAAGAAATCTATAGAAAGT
c.*2035C>T	ACGTTGGATGAGCAAGTTAGCGCCTTGCTG	ACGTTGGATGGTACAAAACCACAGTGTACC	cTAGCGCCTTGCTGAATACA

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.

4.2.5 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 (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 determined with the program miRanda (John et al. 2004; Betel et al. 2008).

4.2.6 Clinical Assessment

The patient in whom the novel missense variant p.R138Q was detected was contacted for clinical, genetic, and functional evaluations. All clinical assessments were carried out by a board-certified developmental pediatrician (J. Visootsak) and a clinical geneticist/dysmorphologist (M. P. Adam). Genotyping of family members was performed with targeted Sanger sequencing. All genotyped family members gave informed consent

to participate in this study. All aspects of the patient recontact and evaluation were approved by the Emory University Institutional Review Board.

4.2.7 Luciferase Assays

4.2.7.1 Plasmid Construction

The pFMR1-luc plasmid was a kind gift from Dr. Daniel Reines (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).

4.2.7.2 Cell Culture and Transfections

HeLa cells were cultured at 37°C with 5% CO_2 in DMEM with 10% fetal bovine serum. Twenty-four hours before transfection, $1x10^6$ 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 µl 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.

4.2.7.3 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 µg of protein was added in 20 µl total volume to a luminometer tube. To each luminometer tube, 100 µl 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 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.

4.3 Results

4.3.1 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. As shown in Table 4.2, we identified 59 known polymorphisms in *FMR1* in this population. Among these were two variants, c.-418_-417insGGC and c.18G>T, which had previously been identified in intellectually disabled patients with normal FMRP expression (Gronskov et al. 1998; Mila et al. 2000). The other 57 known variants are included in dbSNP (build 130). Providing evidence of the sensitivity of pooled-template MPS, the only *FMR1* SNPs catalogued in dbSNP that we did not detect were those with low or unknown population frequencies, which were likely not present in our patient population.

Additionally, we detected 130 variants in *FMR1* which, to our knowledge, have not previously been reported. None are catalogued 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. 2008c; 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 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 shown in Table 4.3.

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

Table 4.2. *FMR1* sequence variants detected in 963 developmentally delayed males

	Novel variants	Known polymorphisms
5'US	7	1
Promoter	3	0
5'UTR	1	0
Exon: Nonsynonymous	1	1
Exon: Synonymous	2	3
Intron	103	51
3'UTR	12	3
3'DS	1	0
Total	130	59

US = upstream sequence; UTR = untranslated region; DS = downstream sequence

Table 4.3. *FMR1* variants detected in 963 developmentally delayed males

Variant	rs ID #	Location	PhastCons	PhyloP	Patient frequency	Control frequency (Caucasian)
c.-602G>A		5' US	0	0.65	7/963 (0.7%)	-
c.-546G>T		5' US	0	-0.86	1/963 (0.1%)	-
c.-539C>A		5' US	0	0.50	1/963 (0.1%)	-
c.-534G>T		5' US	0	-0.41	1/963 (0.1%)	-
c.-494_-493dupCT		5' US	0.01	1.15	1/963 (0.1%)	-
c.-439C>A		5' US	0	0.85	1/963 (0.1%)	-
c.-436G>A		5' US	0	-0.61	1/963 (0.1%)	-
c.-418_-417insGGC		5' US	0	-1.21	8/963 (0.8%)	-
c.-332G>C		Promoter	0.72	0.65	1/963 (0.1%)	0/1308 (0%)
c.-293T>C		Promoter	0.94	1.95	1/963 (0.1%)	0/1266 (0%)
c.-254A>G		Promoter	1	2.10	1/963 (0.1%)	0/1304 (0%)
c.-67G>C		5'UTR	0.30	0.45	4/963 (0.4%)	1/1197 (0.1%)
c.18G>T		Exon 1	1	2.41	13/963 (1.4%)	19/1401 (1.4%)
c.51+94G>A		Intron 1	0	-1.70	6/963 (0.6%)	-
c.51+176C>G		Intron 1	0.10	0.64	1/963 (0.1%)	-
c.51+258C>G		Intron 1	0	-0.53	1/963 (0.1%)	-
c.51+353C>A	rs28900	Intron 1	0.01	-0.48	54/963 (5.6%)	-
c.51+371A>G		Intron 1	0.01	0.58	1/963 (0.1%)	-
c.51+491G>C		Intron 1	0	-0.59	23/963 (2.4%)	-
c.51+637T>G		Intron 1	0.02	-0.96	1/963 (0.1%)	-
c.51+880A>T		Intron 1	0.30	0.74	1/963 (0.1%)	-
c.52-112A>G	rs25726	Intron 1	0	0.31	120/963 (12.5%)	-
c.52-58_52-55delCTAA		Intron 1	0.01	0.51	3/963 (0.3%)	-
c.52-47A>G		Intron 1	0.01	1.27	7/963 (0.7%)	15/1382 (1.1%)
c.104+42A>G		Intron 2	0.01	0.41	14/963 (1.4%)	-
c.104+291T>C	rs25727	Intron 2	0.04	1.17	121/963 (12.6%)	-

Table 4.3. *FMR1* variants detected in 963 developmentally delayed males (cont.)

Variant	rs ID #	Location	PhastCons	PhyloP	Patient frequency	Control frequency (Caucasian)
c.104+353T>C	rs29295	Intron 2	0	-2.02	20/963 (2.1%)	-
c.104+453T>C		Intron 2	0	-0.45	1/963 (0.1%)	-
c.104+462A>C		Intron 2	0	0.76	3/963 (0.3%)	-
c.104+745T>C		Intron 2	0	-1.79	1/963 (0.1%)	-
c.104+1067C>G		Intron 2	0	-0.44	1/963 (0.1%)	-
c.104+1195G>A		Intron 2	0	-0.93	3/963 (0.3%)	-
c.104+1312A>G		Intron 2	0	-0.57	3/963 (0.3%)	-
c.104+1516_104+1517delTT	rs10567806	Intron 2	0	-1.61	27/963 (2.8%)	-
c.105-1715T>C		Intron 2	0	-1.45	1/963 (0.1%)	-
c.105-1648G>A	rs6626955	Intron 2	0	-0.58	2/963 (0.2%)	-
c.105-1595A>G	rs29296	Intron 2	0	-2.66	58/963 (6.0%)	-
c.105-1418T>A	rs29297	Intron 2	0.02	-0.75	71/963 (7.4%)	-
c.105-339A>G		Intron 2	0.52	1.36	3/963 (0.3%)	-
c.105-333G>A		Intron 2	0	-0.70	23/963 (2.4%)	-
c.105-260G>A		Intron 2	0.35	0.47	1/963 (0.1%)	-
c.105-179G>T		Intron 2	0.01	1.06	11/963 (1.1%)	-
c.105-133G>T		Intron 2	0.02	-0.95	21/963 (2.2%)	-
c.105-8A>C		Intron 2	1	2.48	6/963 (0.6%)	0/1262 (0%)
c.198+69T>A	rs25731	Intron 3	0	-0.63	78/963 (8.1%)	-
c.198+170C>T		Intron 3	0	-2.37	3/963 (0.3%)	-
c.198+243G>C		Intron 3	0	0.36	3/963 (0.3%)	-
c.198+278C>T	rs73606778	Intron 3	0	-0.75	22/963 (2.3%)	-
c.198+367A>G	rs29299	Intron 3	0	-0.34	48/963 (5.0%)	-
c.198+400A>G		Intron 3	0.36	0.53	6/963 (0.6%)	-
c.198+423T>A	rs29300	Intron 3	0	-0.86	18/963 (1.9%)	-
c.198+476A>T		Intron 3	0	1.11	2/963 (0.2%)	-

Table 4.3. *FMR1* variants detected in 963 developmentally delayed males (cont.)

Variant	rs ID #	Location	PhastCons	PhyloP	Patient frequency	Control frequency (Caucasian)
c.198+540A>G	rs25730	Intron 3	0.03	0.30	2/963 (0.2%)	-
c.198+576T>C		Intron 3	0	-0.63	4/963 (0.4%)	-
c.198+729A>G		Intron 3	0	0.53	10/963 (1.0%)	-
c.198+731T>A		Intron 3	0	0.41	4/963 (0.4%)	-
c.199-886C>T	rs29277	Intron 3	0	-1.44	69/963 (7.2%)	-
c.199-829C>T		Intron 3	0	-0.63	4/963 (0.4%)	-
c.199-664C>A	rs25706	Intron 3	0.03	0.94	1/963 (0.1%)	-
c.199-661C>G	rs29278	Intron 3	0	-1.74	1/963 (0.1%)	-
c.199-520A>G		Intron 3	0	-0.47	96/963 (10.0%)	-
c.199-349T>A		Intron 3	0	0.57	5/963 (0.5%)	-
c.199-256T>A		Intron 3	0	-0.47	1/963 (0.1%)	-
c.199-117_199-116delAT	rs36079315	Intron 3	0	-0.36	2/963 (0.2%)	-
c.199-27A>T		Intron 3	1	0.74	1/963 (0.1%)	-
c.271-57A>G		Intron 4	0	0.34	1/963 (0.1%)	-
c.271-19A>G	rs29279	Intron 4	0.93	2.54	23/963 (2.4%)	-
c.375T>C		Exon 5	1	0.56	4/963 (0.4%)	1/1357 (0.1%)
c.413G>C (p.R138Q)		Exon 5	0.99	2.94	1/963 (0.1%)	0/1385 (0%)
c.414G>A	rs25707	Exon 5	0.98	-0.70	70/963 (7.3%)	-
c.419+73T>G		Intron 5	0.10	0.71	7/963 (0.7%)	-
c.419+204G>A	rs25708	Intron 5	0	-0.41	53/963 (5.5%)	-
c.419+353C>T		Intron 5	0	-1.53	1/963 (0.1%)	-
c.419+389A>T	rs29280	Intron 5	0	-0.46	78/963 (8.1%)	-
c.419+482G>A		Intron 5	0	-1.04	2/963 (0.2%)	-
c.433G>T	rs29281	Exon 6	0.98	0.64	14/963 (1.5%)	0/1383 (0%)
c.630+78G>A		Intron 7	0	0.33	1/963 (0.1%)	-
c.630+326C>A		Intron 7	1	1.14	8/963 (0.8%)	-

Table 4.3. FMR1 variants detected in 963 developmentally delayed males (cont.)

Variant	rs ID #	Location	PhastCons	PhyloP	Patient frequency	Control frequency (Caucasian)
c.630+438A>C		Intron 7	1	2.55	1/963 (0.1%)	0/1263 (0%)
c.630+523T>C	rs25699	Intron 7	1	0.83	519/963 (53.9%)	-
c.630+569A>G		Intron 7	1	1.32	1/963 (0.1%)	3/1270 (0.2%)
c.630+821C>G		Intron 7	0.03	0.87	5/963 (0.5%)	-
c.630+841T>C	rs29282	Intron 7	0.01	-1.19	74/963 (7.7%)	-
c.630+849C>T		Intron 7	0.11	2.35	1/963 (0.1%)	-
c.630+852T>C		Intron 7	0.12	0.53	1/963 (0.1%)	-
c.630+875T>C		Intron 7	0.01	1.12	1/963 (0.1%)	-
c.630+1045T>G	rs25710	Intron 7	0.40	2.15	25/963 (2.6%)	-
c.631-984A>G		Intron 7	0.13	1.02	1/963 (0.1%)	-
c.631-952A>T	rs7885132	Intron 7	0	0.68	39/963 (4.0%)	-
c.631-909A>G	rs5951855	Intron 7	0.05	-2.96	7/963 (0.7%)	-
c.631-840G>A		Intron 7	1	1.76	1/963 (0.1%)	0/1239 (0%)
c.631-700T>A		Intron 7	0	0.33	1/963 (0.1%)	-
c.631-418A>G		Intron 7	0	-0.64	1/963 (0.1%)	-
c.631-240A>G	rs29283	Intron 7	1	1.28	15/963 (1.6%)	-
c.631-124C>A		Intron 7	0	-0.50	1/963 (0.1%)	-
c.651G>A		Exon 8	0.99	-0.44	1/963 (0.1%)	1/1392 (0.1%)
c.801+31C>T		Intron 8	0	-1.56	21/963 (2.2%)	-
c.880+577A>T		Intron 9	0.10	0.94	2/963 (0.2%)	-
c.880+885A>G		Intron 9	0	-0.73	1/963 (0.1%)	0/1084 (0%)
c.880+1173G>A		Intron 9	0.40	0.68	2/963 (0.2%)	-
c.880+1392G>A		Intron 9	0.03	1.58	1/963 (0.1%)	-
c.880+1442_880+1443insT	rs11425567	Intron 9	0.01	0.82	53/963 (5.5%)	-
c.880+1446_880+1447insG		Intron 9	0.02	1.11	1/963 (0.1%)	-
c.881-1813C>T		Intron 9	0	0.44	2/963 (0.2%)	-

Table 4.3. FMR1 variants detected in 963 developmentally delayed males (cont.)

Variant	rs ID #	Location	PhastCons	PhyloP	Patient frequency	Control frequency (Caucasian)
c.881-1756A>T		Intron 9	0	-0.65	2/963 (0.2%)	-
c.881-1731C>A		Intron 9	0	-0.46	1/963 (0.1%)	-
c.881-1686C>G		Intron 9	0.01	0.59	1/963 (0.1%)	-
c.881-1455C>T		Intron 9	0	-0.55	1/963 (0.1%)	-
c.881-1312C>T		Intron 9	0	-0.76	1/963 (0.1%)	-
c.881-1146A>G		Intron 9	0	-0.98	1/963 (0.1%)	-
c.881-790A>G		Intron 9	0	-1.67	1/963 (0.1%)	-
c.881-731T>C	rs25715	Intron 9	0.13	-0.39	453/963 (47.1%)	-
c.881-700G>A		Intron 9	0.17	0.94	1/963 (0.1%)	-
c.881-607G>A		Intron 9	0.03	0.82	1/963 (0.1%)	-
c.881-319G>A		Intron 9	0	-1.12	1/963 (0.1%)	-
c.881-261A>G		Intron 9	0	-0.97	1/963 (0.1%)	-
c.881-116A>G		Intron 9	0	-1.22	1/963 (0.1%)	-
c.881-94C>T	rs29284	Intron 9	0	-0.53	28/963 (2.9%)	-
c.990+4T>C		Intron 10	1	2.46	1/963 (0.1%)	0/1248 (0%)
c.990+14C>T	rs25714	Intron 10	0	0.47	88/963 (9.2%)	-
c.990+35A>G		Intron 10	0.07	0.82	1/963 (0.1%)	-
c.990+48G>A		Intron 10	0	-0.38	1/963 (0.1%)	-
c.990+141T>C	rs25713	Intron 10	0	-1.62	70/963 (7.3%)	-
c.991-311G>A		Intron 10	0	-0.41	1/963 (0.1%)	-
c.991-54A>G	rs41311668	Intron 10	0	0.66	2/963 (0.2%)	-
c.1125+73G>A	rs25700	Intron 11	0	-1.57	161/963 (16.7%)	-
c.1125+176A>G		Intron 11	0.14	-0.71	1/963 (0.1%)	-
c.1126-132T>G	rs29285	Intron 11	0.88	2.23	16/963 (1.7%)	-
c.1126-78_1126-77dupAAAAGTCCTGCAGT		Intron 11	1	1.71	1/963 (0.1%)	0/443 (0%)
c.1188+94G>C	rs29286	Intron 12	0.99	1.66	48/963 (5.0%)	-

Table 4.3. FMR1 variants detected in 963 developmentally delayed males (cont.)

Variant	rs ID #	Location	PhastCons	PhyloP	Patient frequency	Control frequency (Caucasian)
c.1275+103A>G	rs25701	Intron 13	0	0.68	117/963 (12.2%)	-
c.1275+207T>C		Intron 13	0	0.99	1/963 (0.1%)	-
c.1275+264G>T		Intron 13	0	0.43	12/963 (1.2%)	-
c.1275+337_1275+338delGT		Intron 13	0.08	-0.37	1/963 (0.1%)	-
c.1275+407C>A		Intron 13	0	-0.97	1/963 (0.1%)	-
c.1275+505G>A	rs25702	Intron 13	0	-0.83	99/963 (10.3%)	-
c.1275+793T>G		Intron 13	0.01	2.30	1/963 (0.1%)	-
c.1275+1020_1275+1021dupT		Intron 13	0.27	1.38	1/963 (0.1%)	-
c.1275+1027G>A	rs29288	Intron 13	0.55	1.10	51/963 (5.3%)	-
c.1275+1186G>A	rs25718	Intron 13	0.76	0.69	7/963 (0.7%)	-
c.1275+1209delC	rs1805425	Intron 13	0.99	2.58	1/963 (0.1%)	-
c.1276-1018G>A	rs29289	Intron 13	0.54	0.61	40/963 (4.2%)	-
c.1276-1011C>T		Intron 13	0	-0.91	1/963 (0.1%)	-
c.1276-940C>T		Intron 13	0	0.32	1/963 (0.1%)	-
c.1276-910A>G		Intron 13	0	1.01	1/963 (0.1%)	-
c.1276-526C>T		Intron 13	0	0.32	1/963 (0.1%)	-
c.1276-518A>G		Intron 13	0.38	1.01	1/963 (0.1%)	-
c.1276-516G>A	rs73606788	Intron 13	0.51	0.42	26/963 (2.7%)	-
c.1276-389T>C	rs29290	Intron 13	1	2.33	19/963 (2.0%)	-
c.1276-325C>T		Intron 13	0.60	0.71	1/963 (0.1%)	0/1396 (0%)
c.1276-93A>T	rs25719	Intron 13	1	1.13	102/963 (10.6%)	-
c.1471+274T>A		Intron 14	0.99	1.17	1/963 (0.1%)	0/1241 (0%)
c.1471+420T>C		Intron 14	1	1.13	1/963 (0.1%)	0/1213 (0%)
c.1471+624C>T		Intron 14	1	1.33	1/963 (0.1%)	0/1251 (0%)
c.1471+625G>C		Intron 14	0.99	1.33	1/963 (0.1%)	0/1127 (0%)
c.1472-581A>T		Intron 14	0	-0.52	2/963 (0.2%)	-

Table 4.3. FMR1 variants detected in 963 developmentally delayed males (cont.)

Variant	rs ID #	Location	PhastCons	PhyloP	Patient frequency	Control frequency (Caucasian)
c.1472-521C>G		Intron 14	0	0.40	1/963 (0.1%)	0/1254 (0%)
c.1472-469G>A		Intron 14	0.36	1.13	2/963 (0.2%)	-
c.1472-356A>G		Intron 14	0	0.77	1/963 (0.1%)	-
c.1472-350C>G	rs29291	Intron 14	0	-0.89	17/963 (1.8%)	-
c.1472-104_1472-103insTGGG	rs71836009	Intron 14	0.98	0.71	24/963 (2.5%)	-
c.1654+48T>C	rs7054612	Intron 15	0	0.65	21/963 (2.2%)	-
c.1654+225C>A		Intron 15	0	-0.47	2/963 (0.2%)	-
c.1655-91G>T		Intron 15	0	-0.36	1/963 (0.1%)	-
c.1737+8C>A	rs73606790	Intron 16	0	-0.78	28/963 (2.9%)	-
c.1737+185G>A		Intron 16	0	-0.47	1/963 (0.1%)	1/1401 (0.1%)
c.1737+600C>G	rs12012926	Intron 16	0	-0.79	16/963 (1.7%)	-
c.1737+1107T>C		Intron 16	0.08	1.04	2/963 (0.2%)	-
c.1738-900C>A	rs25723	Intron 16	0.01	1.96	126/963 (1.3%)	-
c.1738-556C>T	rs25720	Intron 16	0	-0.44	6/963 (0.6%)	-
c.1738-276A>G		Intron 16	0.76	0.45	1/963 (0.1%)	-
c.1738-235C>G		Intron 16	0.16	-0.50	1/963 (0.1%)	-
c.1857C>T	rs45540244	Exon 17	0.99	2.79	2/963 (0.2%)	-
c.*23T>C		3'UTR	0.99	2.40	1/963 (0.1%)	0/900 (0%)
c.*32C>G		3'UTR	1	2.64	2/963 (0.2%)	1/1195 (0.1%)
c.*60G>C		3'UTR	1	3.26	1/963 (0.1%)	1/1408 (0.1%)
c.*261C>T		3'UTR	0.17	0.36	9/963 (0.9%)	-
c.*541T>C		3'UTR	1	2.22	1/963 (0.1%)	1/1238 (0.1%)
c.*746T>C		3'UTR	1	2.22	6/963 (0.6%)	0/1260 (0%)
c.*952_953dupT		3'UTR	0.98	0.75	1/963 (0.1%)	-
c.*1062A>T	rs11798323	3'UTR	1	2.45	20/963 (2.1%)	-
c.*1169A>G		3'UTR	0.34	0.61	11/963 (1.2%)	-

Table 4.3. FMR1 variants detected in 963 developmentally delayed males (cont.)

Variant	rs ID #	Location	PhastCons	PhyloP	Patient frequency	Control frequency (Caucasian)
c.*1174C>T	rs25704	3'UTR	0	-0.87	297/963 (30.8%)	-
c.*1185delT		3'UTR	0	-0.36	3/963 (0.3%)	-
c.*1482C>A		3'UTR	0.99	1.25	1/963 (0.1%)	-
c.*1513C>T	rs25722	3'UTR	1	2.47	1/963 (0.1%)	-
c.*1867G>A		3'UTR	1	1.54	12/963 (1.2%)	0/951 (0%)
c.*2035C>T		3'UTR	1	2.47	3/963 (0.3%)	0/1270 (0%)
c.*2382A>C		3'DS	0.05	1.12	1/963 (0.3%)	-

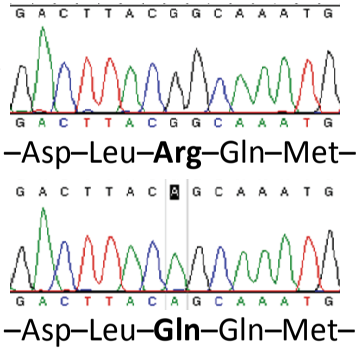
amplicon was pooled in the same quantity and fashion as all of the patient amplicons, the frequency of successful detection of p.I304N reflects the frequency of successful detection of other singleton variants in a pool. At our detection threshold, p.I304N was not detected in 12/51 pools. Based upon this false negative rate of 23.5%, it is possible that we missed up to 40 true variants in the population of 963 developmentally delayed males. However, it would be expected that we missed no more than one true promoter variant and no more than one true missense variant, thus suggesting that we detected the majority of *FMR1* variants that are likely to have a functional effect in this population.

4.3.2 Characterization of the Novel Missense Variant p.R138Q

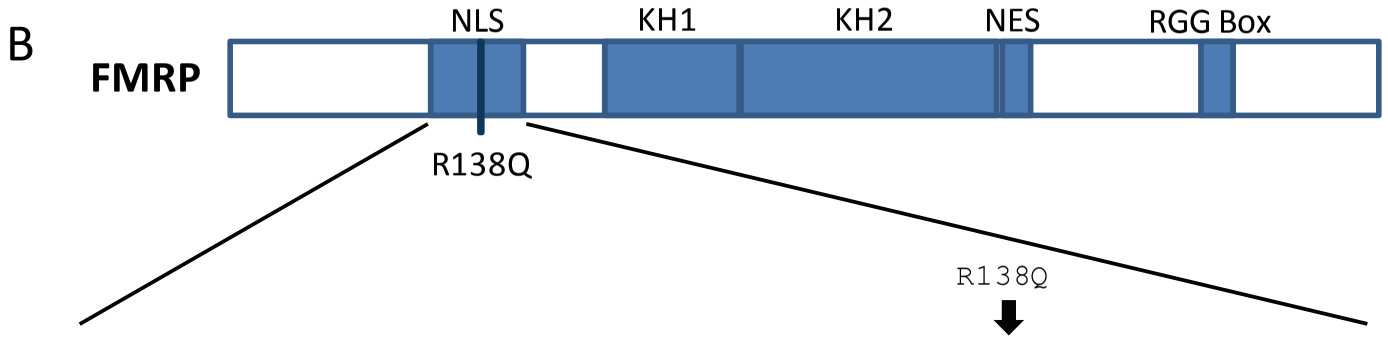
4.3.2.1 Identification and Preliminary Functional Evaluation

In the 963 developmentally delayed males sequenced, we detected only one novel missense variant in *FMR1*. This variant, c.413G>C (Figure 4.1A), was not detected in 1385 control Caucasian males. The c.413G>C variant encodes an arginine-to-glutamine substitution at codon 138. Arginine-138 has been identified as a key basic residue in the nuclear localization signal (NLS) of FMRP (Eberhart et al. 1996; Bardoni et al. 1997) and is conserved through *Drosophila* (Figure 4.1B). The R-to-Q missense substitution is predicted to be pathological by PMut (NN output: 0.84) (Ferrer-Costa et al. 2005) and functional by PANTHER PSEC (subPSEC: -4.3) (Thomas and Kejariwal 2004), but tolerated by SIFT (score: 0.22) (Ng and Henikoff 2003) and benign by PolyPhen (PSIC Score Δ: 0.11) (Ramensky et al. 2002).

A



B



R138Q

```
Homo sapiens       111 RLRSVNPNKP ATKDTFHKIK LDVPEDLRQM CAKEAAHKDF KK 152
Macaca mulatta     111 RLRSVNPNKP ATKDTFHKIK LDVPEDLRQM CAKESAHKDF KK 152
Canis familiaris   111 RLRSVNPNKP ATKDTFHKIK LDVPEDLRQM CAKESAHKDF KK 152
Rattus norvegicus  111 RLRSVNPNKP ATKDTFHKIK LEVPEDLRQM CAKESAHKDF KK 152
Mus musculus       111 RLRSVNPNKP ATKDTFHKIK LEVPEDLRQM CAKESAHKDF KK 152
Gallus gallus      111 RLRSVNPNKP ATKDTFHKIK LEVPEDLRQM CAKESAHKDF KK 152
Xenopus laevis     111 RLRSVNPNKP ATKSSFHKVK LDVPEDLRQM CAKDSAHKDF KK 152
Danio rerio        111 RLRPVNPNKA ATKNTFLKTR LDVPEDLRQM CAKDSAHKDF KK 152
D. melanogaster    111 RLRAKNSNPP ITAKTFYQFT LPVPEELREE AQKDGIHKEF QR 152
                       ***. *.* .  * .:* :   * ***:**:  . *:. **:* ::
```

Figure 4.1. Functional implications of the p.R138Q missense change.

A. DNA chromatograms comparing the wild-type and c.413G>A 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.

4.3.2.2 Clinical Evaluation

The patient in whom the p.R138Q variant was detected is an 8 year old Caucasian male with a history of developmental delay. He was born full-term with a birth weight of 8 pounds 12 ounces. He was delivered by spontaneous vaginal delivery and stayed in the hospital for 5 days due to feeding and respiratory problems.

At 1 month of age, the patient had a non-accidental trauma due to shaking and developed seizures after the event. He was subsequently diagnosed with shaken baby syndrome. He began to experience seizures associated with fever at approximately 18 months of age. This consisted of a period of blank staring, falling, and moving his arms. The second seizure, unassociated with fever, occurred at 24 months of age. He continues to have complex partial seizures every 3 – 4 months with approximately 12 seizures per day every 2 – 3 days. In addition to these seizures, the patient also had a history of frequent ear infections that required placement of tympanostomy tubes at 9 months of age.

In terms of the patient's developmental milestones, he rolled over at 6 months, sat independently at 8 months, and walked alone at 1½ years. He began to say his first word at 12 months of age. He began to combine words at 3 years and spoke in complete sentences at 7 years. His speech is intelligible to strangers. He has poor handwriting, and is unable to write his name. He is able to use utensils to feed himself.

Behaviorally, the patient is very social and outgoing. He is not shy or timid in new situations, nor does he have any difficulties with changes or transitions. He does not have stereotypic behaviors, such as chewing on his nails, scratching his body, hand

flapping, or hand biting. He is not hyperactive or impulsive nor does he have temper tantrums or physical aggressiveness.

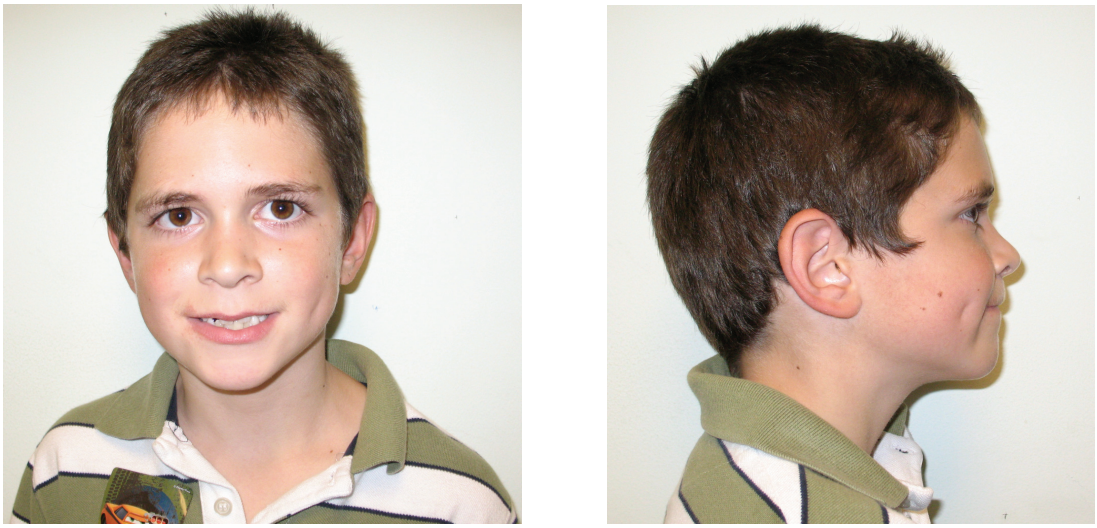
Assessments of cognitive function and adaptive behaviors were performed using the Kaufman Brief Intelligence Test, Second Edition, and the Vineland Adaptive Behavior Scales, Second Edition (VABS). At the age of 8 years and 2 months, the patient's verbal IQ is 72 (90% CI = 66-83), non-verbal IQ is 77 (90% CI = 68-89), and IQ composite is 71 (90% CI = 65-79). Results from the VABS revealed scores of 70 for communication, 76 for daily living skills, 104 for socialization, 78 for motor skills, yielding an adaptive behavior composite of 81.

Physical examination reveals a nondysmorphic male (Figure 4.2A) with several noteworthy features. His head circumference is 54 cm, the 95th percentile for his age. His ears appear normal in size, location, and morphology. He exhibits no cardiac murmurs, rubs, or gallops. There is no evidence of macroorchidism. The patient's hands and feet show normal flexion creases. The patient has partial 2/3 toe syndactyly and striking hirsutism of the forearms bilaterally.

4.3.2.3 Patient Pedigree

The proband inherited the p.R138Q variant from his mother, who is heterozygous for the variant allele (Figure 4.2B). The proband's mother does not display overt intellectual disability and successfully graduated from high school, but stated that she struggled academically and is currently unemployed. Furthermore, the proband's mother exhibits a high level of social anxiety, a common phenotypic finding in females heterozygous for an *FMR1* repeat expansion (Freund et al. 1993). The proband's father,

A



B

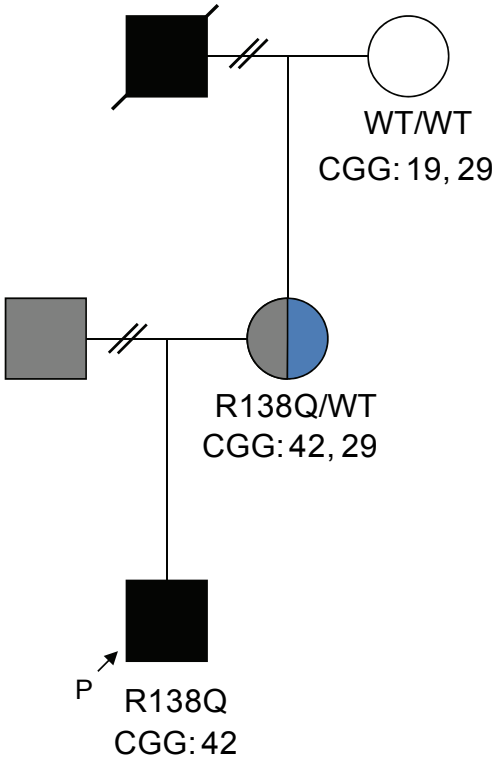


Figure 4.2. Clinical implications and inheritance of the p.R138Q missense change.

A. Photographs of the patient in whom the p.R138Q missense variant was detected.

B. Pedigree of the family in which the p.R138Q missense variant was detected. Black shading denotes intellectual disability, grey shading denotes learning difficulties, and blue shading denotes social anxiety. Genotype data is in reference to the p.R138Q variant. Also noted is the number of CGG repeats in *FMR1*, where available.

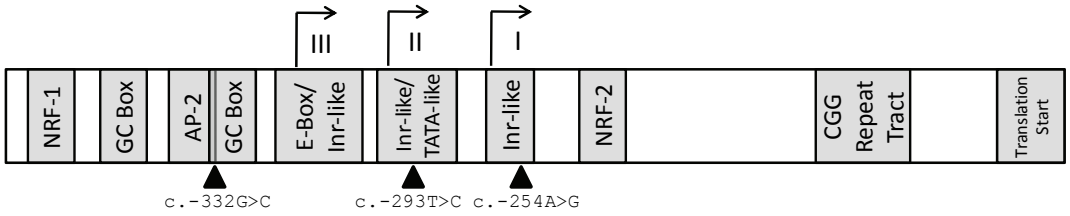
with whom the proband is no longer in contact, is stated to have had learning difficulties and failed to graduate high school. The proband's maternal grandmother is not stated to exhibit difficulties with learning or socialization, and was found to be negative for the p.R138Q variant. The proband's maternal grandfather, who is deceased, is stated to have been unable to read or write. We performed *FMR1* repeat testing on the proband, the proband's mother, and the proband's maternal grandmother and determined that the p.R138Q variant allele is, in this family, associated with a repeat tract 42 CGGs in length. Because the proband's maternal grandmother lacks an *FMR1* allele with 42 CGG repeats, this allele was inherited by the proband's mother from the proband's maternal grandfather. Thus, the p.R138Q variant may have been inherited from the proband's cognitively impaired maternal grandfather or it may have arisen *de novo* in the proband's mother.

4.3.3 Novel Variants in the *FMR1* Promoter

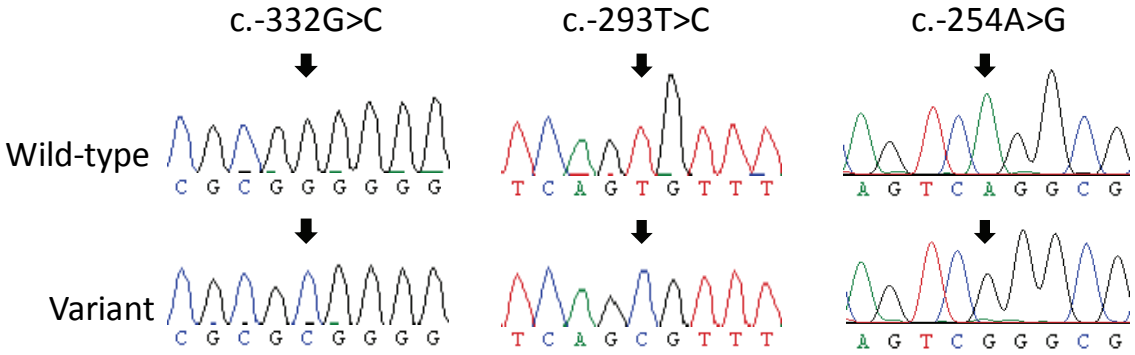
4.3.3.1 Identification and Preliminary Functional Evaluation

We detected three novel sequence variants in the minimal promoter of *FMR1* (Figure 4.3), 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-2α transcription factors (Smith et al. 2004; Lim et al. 2005). The c.-293T>C variant is located near transcription start site II, within both an Initiator-like (Inr-like) and a TATA-like sequence (Hwu et al. 1993; Beilina et al. 2004). The third variant, c.-254A>G, is located within an Inr-like sequence near the primary transcription start site (Kumari and

A



B



C

	c.-332G>C	c.-293T>C	c.-254A>G
Homo sapiens	GCCCGCGGGGGGAGGG	TCAGTGTTTACA	TCAGTCAG
Macaca mulatta	GCCCGCTGGGGGAGGG	TCAGTGTTTACA	TCAGTCAG
Canis familiaris	GCCCGCGGGGGGAGGG	TCACTGTTTACA	TCAGTCAG
Rattus norvegicus	AGCCCATGGGGGAGGG	TGACTGTTTACA	TCAGTCAG
Mus musculus	GGCCAAAGGGGAGGGG	TGACTGTTTACA	TCAGTCAG
	** **** ***	* * ********	********
Promoter element:	**AP-2 binding/GC Box**	**Inr-like/TATA-like**	**Inr-like**

Figure 4.3. Novel variants in the *FMR1* promoter.

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.-332G>C, c.-293T>C, and c.-254A>G.

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.

Usdin 2001; Beilina et al. 2004). All three of these variant bases are conserved in mammals (Figure 4.3C).

The c.-332G>C, c.-293T>C, and c.-254A>G variants were not detected in 1308, 1266, and 1304 Caucasian male controls, respectively. However, the race of the patients in whom these variants were detected is unknown. Thus, it remains possible that these promoter variants are highly rare in Caucasians but polymorphic in other populations.

4.3.3.2 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 4.4, 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 *FMR1* promoter activity to 36.2% of wild-type levels. All of these reductions from wild-type promoter activity were statistically significant ($p < 1x10^{-7}$).

4.3.4 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 Caucasian male controls for

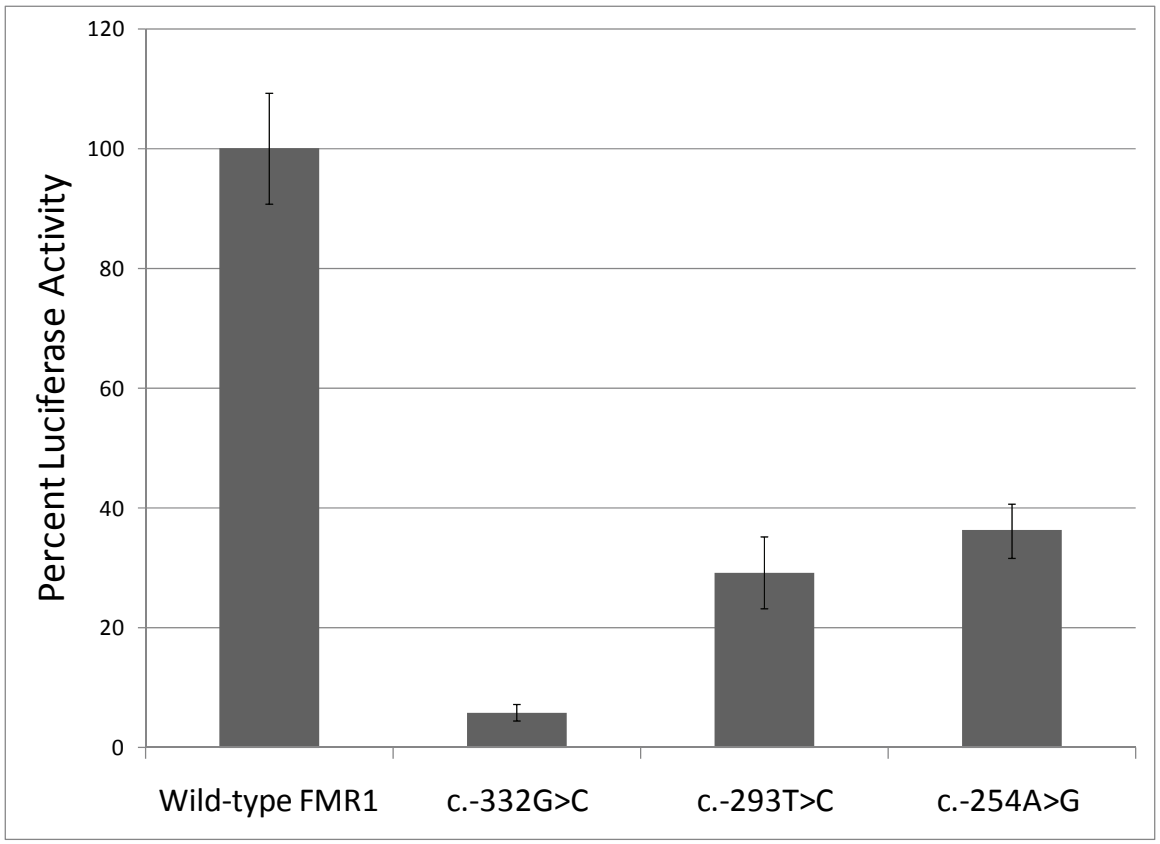


Figure 4.4.Luciferase assays of the variant *FMR1* promoters.

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 ± 1 S.D.

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 a phastCons score > 0.8 and phyloP score > 1.5 were defined to be highly conserved. Similarly, we genotyped control Caucasian males 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 4.4, 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. However, because the race of most of the patients in whom these variants were detected is unknown, it is possible that these variants are very rare in Caucasians but polymorphic in other populations.

4.4 Discussion

Through the use of pooled-template MPS, we have identified 130 novel sequence variants in *FMR1* in a population of 963 developmentally delayed males. 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.

Table 4.4. Novel *FMR1* variants only found in developmentally delayed males

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

The variants c.880+885A>G and c.1472-521C>G are predicted to introduce a splice donor site;

c.*746T>C is predicted to reduce miRNA binding

We identified four *FMR1* sequence variants that show evidence of association with developmental delay. Among these are the third missense variant and the first three promoter variants to be detected in *FMR1*. If these novel variants are all verified to be pathogenic, the frequency of *FMR1* sequence variants causing developmental delay in our study population would be between 0.4 – 0.5%, 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 (Patsalis et al. 1999; Hecimovic et al. 2002; Pandey et al. 2002; Major et al. 2003; Biancalana et al. 2004; Rauch et al. 2006). Therefore, *FMR1* sequence variants should be viewed as a significant contributor to the heterogeneous diagnosis of developmental delay.

It remains unclear in which patients pathogenic *FMR1* sequence variants are most likely to be found. Currently, *FMR1* sequencing is most often ordered for patients who present with a phenotype strongly resembling classical FXS. However, this may not be the phenotypic expression of many *FMR1* sequence variants. For instance, the patient harboring the p.R138Q variant did not exhibit macroorchidism, autism-like behaviors, or the facial features commonly seen in FXS. Furthermore, the study presented in Chapter 2 and several previous studies of FXS-like patients have failed to find pathogenic sequence variants in *FMR1* (Chiurazzi et al. 1994; Wang et al. 1997; Castellvi-Bel et al. 1999), suggesting that the phenotype caused by such variants may be more subtle or distinct from classical FXS. As more patients are identified with pathogenic *FMR1* sequence variants, careful phenotyping will be necessary to more accurately define the typical clinical presentation.

While diagnostic *FMR1* sequencing may be relatively low-yield, positive findings would empower developmentally delayed patients and their parents in several important ways. It is likely that therapies used to manage FXS, including several targeted pharmaceuticals currently in development (Chang et al. 2008; Hagerman et al. 2009), would also be beneficial for developmentally delayed patients with pathogenic *FMR1* sequence variants. Furthermore, the identification of an underlying cause for developmental delay has been shown to benefit the patient's parents by assuaging feelings of guilt and regret (Lenhard et al. 2005), guiding family planning decisions, and calming anxieties about their child's prognosis (Rosenthal et al. 2001).

In addition to the clinical benefit of identifying patients with pathogenic *FMR1* sequence variants, our data provide several avenues for molecular inquiries into the structure and function of FMRP. The first missense mutation identified in *FMR1*, p.I304N (De Boulle et al. 1993), has guided many of the studies of FMRP functions, including ribosomal interaction (Feng et al. 1997a), RNA binding (Darnell et al. 2005), and intracellular transport (Schrier et al. 2004). Two of the less well-studied properties of FMRP, its nuclear localization signal (NLS) and protein-protein interactions, may be disrupted by the novel p.R138Q substitution. Arginine-138 is one of the basic residues proposed to be essential for the function of FMRP's nonclassical NLS (Eberhart et al. 1996; Bardoni et al. 1997). If p.R138Q disrupts NLS function, it will allow a direct assessment of the functional importance of nucleocytoplasmic shuttling by FMRP (Feng et al. 1997b). Furthermore, the N-terminal domain of FMRP mediates the protein-protein interactions of FMRP with FXR1, FXR2, and NUFIP (Bardoni et al. 1999; Adinolfi et al.

2003). It is possible that p.R138Q could disrupt these interactions, allowing an assessment of their heretofore uncharacterized functional role.

The novel noncoding variants identified in *FMR1* will also be functionally informative. Previous studies of the *FMR1* promoter have largely relied on deletion constructs and *in vitro* assays for the characterization of various promoter elements (Kumari and Usdin 2001; Kumari et al. 2005). The three functional variants we identified in the *FMR1* promoter will provide a more accurate way to assess the *in vivo* function of the promoter elements that they disrupt. Additionally, the 3'UTR of *FMR1* shows a high level of sequence conservation, suggesting that it plays an important functional role. The characterization of this function will be aided by the identification of variants associated with developmental delay, as well as by the identification of novel variants found to be polymorphic in the population.

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 suggested 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 truly 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, and the first three sequence variants to be identified in the *FMR1* promoter, all of which reduce transcriptional activity. Taken together, these results suggest that pathogenic sequence variants in *FMR1* represent a significant cause of developmental delay.

Chapter 5. Polymorphic missense and frameshift variants in the putative *ASFMR1*-encoded protein refute its proposed role in fragile X syndrome

5.1 Introduction

Fragile X syndrome (FXS) is the most common inherited form of intellectual disability. As the first described trinucleotide repeat disorder, FXS is typically caused by an expansion of the CGG repeat tract in the 5'UTR of *FMR1*, which results in the gene's transcriptional silencing (Garber et al. 2008). The identification of patients who harbor deletions (Coffee et al. 2008) and missense mutations (De Boulle et al. 1993) in *FMR1* has verified the gene's vital role in the etiology of FXS. However, many FXS-like patients have been described who have no detectable *FMR1* mutations (Castellvi-Bel et al. 1999) or change in *FMR1* expression (Chiurazzi et al. 1994; Reyniers et al. 1996), suggesting that other genes may also be involved in the development of FXS.

Recently, Ladd and colleagues identified a novel gene overlapping the *FMR1* locus in the antisense direction (Ladd et al. 2007). Dubbed *ASFMR1*, this gene produces multiple polyadenylated transcripts that are exported from the nucleus into the cytoplasm. The expression pattern of *ASFMR1* mirrors that of *FMR1*, with high expression in the

brain and minimal expression in skeletal muscle and heart. Intriguingly, *ASFMR1* is affected by the shared trinucleotide repeat in the same fashion as *FMR1*, with increasing gene expression over the premutation size range (55-200 repeats) and a lack of expression in the context of a full mutation (more than 200 repeats). Thus, *FMR1* is not the only gene whose expression is directly inhibited by trinucleotide repeat expansion in FXS. This finding led the authors to hypothesize that *ASFMR1* may play a role in the pathogenesis of FXS, either through a noncoding RNA mechanism or through the involvement of a 100 amino acid protein putatively encoded by *ASFMR1* (Ladd et al. 2007). In the current study, we investigate the proposed role of the ASFMR1 protein in FXS.

5.2 Methods and Results

We first examined the domain structure and conservation of the ASFMR1 protein to determine if it is likely to play an essential functional role. The *ASFMR1* transcript containing the identified open reading frame (ORF) overlaps with exons 1 and 2 of *FMR1* and includes the trinucleotide repeat tract (Figure 5.1). Putatively encoded by this transcript is a 100 amino acid protein containing a polyproline domain, which derives from translation across the antisense CCG repeat. Polyproline domains are a common motif in transcription factors (Gerber et al. 1994), suggesting a possible role for the ASFMR1 protein in transcriptional activation. The length of the polyproline domain, and thus of the protein itself, is determined by the size of the CCG repeat tract. No other known domains have been described in the ASFMR1 protein. Alignment of the amino acid sequences of the ASFMR1 protein from six species reveals that the protein is largely

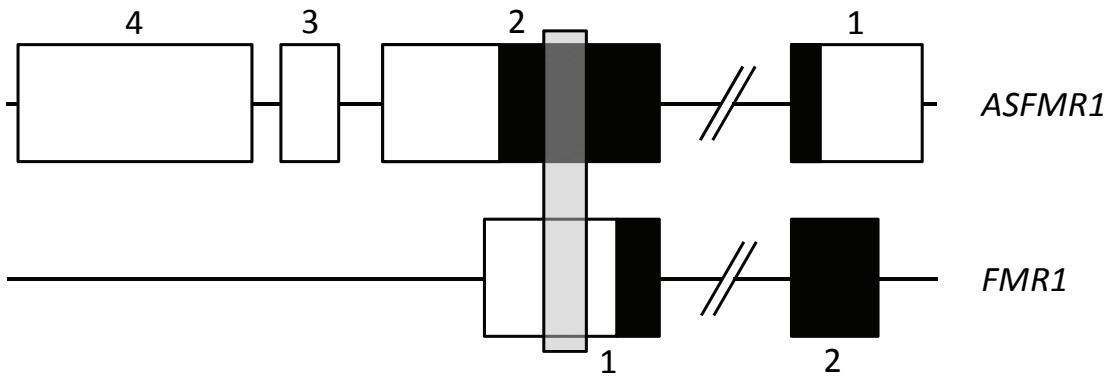


Figure 5.1. Diagram of *ASFMR1* and *FMR1*.

The overlapping *ASFMR1* and *FMR1* genes. Exons are depicted as black (coding) and white (noncoding) numbered boxes. The CGG/CCG trinucleotide repeat shared by *ASFMR1* and *FMR1* is highlighted with a grey box.

conserved in mammals, with the exception of the size of the polyproline domain (Figure 5.2).

If the loss of ASFMR1 protein expression is truly an additional cause of FXS, one would expect that intragenic mutations in the *ASFMR1* coding sequence could cause FXS independently of *FMR1* involvement. To assess this possibility, we sequenced the *ASFMR1* ORF in 963 males referred to our clinical laboratory for fragile X testing who demonstrated no evidence of trinucleotide repeat expansion or hypermethylation of the *FMR1* promoter. Deidentified aliquots of genomic DNA were obtained from every male patient who tested negative for fragile X in our laboratory from February 2003 through August 2007. Because the standard of care in the United States is for clinicians to order fragile X testing for every child presenting with developmental delay, the patient population assessed represents a broader phenotypic spectrum than classic FXS, enhancing the likelihood that causal mutations with less specific effects would be detected. *ASFMR1* was targeted with two long PCR amplifications of patient genomic DNA. Pools of long PCR products from 19 patients and one control sample were sequenced with the Genome Analyzer (Illumina, San Diego, California, USA). The single-end 26-bp sequence reads were aligned to the *FMR1* region of the human genome reference (NT_011681.15, g.3435700 to g.3475545) with the assembly software MAQ (Li et al. 2008). On average, 130-fold sequence coverage was obtained for each patient; however, no sequence data was derived for the antisense CCG repeat tract, due to the inability of short sequence reads covering stretches of microsatellites to map uniquely to a genomic position. Variant bases were detected using an in-house algorithm, and these putative sequence variants were confirmed with Sanger sequencing.

```
H. sapiens      1    MNILYKCLVESAIGAPHFHHQLLHLLFSPASAGSPPPRGG   40
M. mulatta      1    MNILYKCLVESAIGAPHFHHQLLHLLFKKTRAGSPPPRGG   40
C. familiaris   1    MNILYKCLVESAIGAPHFHHQLLHLLESPARAGSPPPRGG   40
B. taurus       1    MNIFYKCLVESAIGAPHFHHQILHLLVRPARFGSPPSRGG   40
M. musculus     1    MDIFYKCLVESAIGAPHFHHQLLHLLVRP-RAGSPPPGGG   39
R. norvegicus   1    MDIFYKCLVESAIGAPHFHHQLLHLLVLP-RAGSPPPRGG   39
                     *:*:******************:****      ****. **

H. sapiens      41   LRALEAQPPPPP--------PPPPPPPPPPPPPPRCRTP   72
M. mulatta      41   LRALAAAAAAAAAAAAAARRPPPPPPPPPPPPPPRCRTP   80
C. familiaris   41   LRAQEAQ----------------PPPPPPPPPPPRSRTP   64
B. taurus       41   LQALEAQ------------------------PPPPRYRTP   56
M. musculus     40   LQALEAQ--------------------PPSPPPPPRYRTP   60
R. norvegicus   40   LQALEAQ----------------------PSPPPPRYRTP   57
                     *:*  *                       ****** ***

H. sapiens      73   PGSGASVTAAARARR---RPAARSEAALHRK           100
M. mulatta      81   PGSGASVTAAARARR---RPAARSEAALHRK           108
C. familiaris   65   PGSGAAVTAATRARR---RLTARSEAAPHRK            92
B. taurus       57   PGSGAAVTAAACARL---RLAARSEAAPHRK            84
M. musculus     61   PGSGAAVTATARARRPFASLPALSGAAPHRK            91
R. norvegicus   58   PGSGAAVTDTARARRPLASLPALSGAAPHRK            88
                     *****:** :: **      .* * ** ***
```

Figure 5.2. Mammalian conservation of the putative *ASFMR1* ORF protein.

Comparison of the putative *ASFMR1* ORF protein across mammalian species. The amino acid sequences were aligned with Clustal. Asterisks denote identical amino acids. Colons denote conservative substitutions. Periods denote similar amino acid residues. With the exception of the variable length polyproline domain, demarcated with the black box, the protein encoded by *ASFMR1* is largely conserved across mammals, particularly in the N-terminal region.

We identified and confirmed two sequence variants in the *ASFMR1* ORF in our developmentally delayed patient population, both of which encode a nonsynonymous change in the ASFMR1 protein. Neither variant appears in dbSNP (build 130) or in any of the first nine publicly available personal genomes. The novel variant c.139C>G encodes the conservative missense change p.Q47E at the amino acid residue immediately preceding the ASFMR1 polyproline domain. Detected in 0.4% of the sequenced patients (Table 5.1), this variant has not, to our knowledge, been previously reported. Another variant, c.55C>A, encodes the conservative missense change p.H19N at a residue conserved in mammals. We detected c.55C>A in 1.4% of the developmentally delayed males sequenced (Table 5.1). The same variant, which manifests as a synonymous variant in codon 6 of *FMR1*, had previously been identified by Gronskov and colleagues in 1.7% of intellectually disabled males. Despite the fact that the authors did not detect the variant in 83 genotyped controls, they dismissed it as a potential cause of intellectual disability when the associated *FMR1* mRNA splicing and FMRP levels were shown to be unaffected by this sequence change (Gronskov et al. 1998). However, since the effects of the variant could be mediated through the p.H19N change in the ASFMR1 protein rather than through *FMR1*, the association between c.55C>A and intellectual disability should be more thoroughly evaluated.

Although both p.H19N and p.Q47E are conservative amino acid changes that may not be expected to disrupt the function of the ASFMR1 protein, the fact that they have only been seen in developmentally delayed males necessitates further investigation of a possible pathogenic role. To determine whether these two variants associate with developmental delay, we genotyped a large number of healthy adult Caucasian males

Table 5.1. *ASFMR1* sequence variants detected in developmentally delayed males

Variant	Missense Change	Patient Frequency	Control Frequency
c.55C>A	p.H19N	13/963 (1.4%)	19/1401 (1.4%)
c.139C>G	p.Q47E	4/963 (0.4%)	1/1197 (0.1%)

from the NIMH Human Genetics Initiative for both c.55C>A and c.139C>G. Although our developmentally delayed patient population was racially heterogeneous, both variants were detected in Caucasians, making this an appropriate control population. The c.139C>G substitution obliterates a BseYI recognition site, so restriction digestion was used to screen for this variant in controls. A single-base primer extension assay on the iPlex Gold platform (Sequenom, San Diego, California, USA) was used to genotype for c.55C>A. Positive controls were genotyped to validate the assays, and all detected minor alleles were confirmed with Sanger sequencing. As seen in Table 5.1, both c.55C>A and c.139C>G were detected in healthy controls, refuting a causal role of these *ASFMR1* missense changes in FXS.

While the sequencing and genotyping data together suggest that intragenic mutations in the *ASFMR1* ORF do not play a significant role in developmental delay, we still sought to determine if the lack of ASFMR1 protein caused by trinucleotide repeat expansion plays any role in the FXS pathogenesis. Although no polymorphisms within the *ASFMR1* ORF are included in dbSNP, this genomic region is routinely interrogated in the course of clinical fragile X testing. Thus, we performed a detailed literature search for *ASFMR1* sequence variants, focusing especially on reports of aberrant fragile X test findings associated with normal levels of FMRP. This literature search revealed six previously reported variants in the *ASFMR1* ORF, three of which encode missense substitutions and three of which encode truncations of the protein (Figure 5.3). The variants encoding missense changes were identified as “pseudodeletions” in the course of fragile X testing by Southern blot. All three are within the trinucleotide repeat tract, and therefore encode substitutions within the polyproline domain of the ASFMR1 protein.

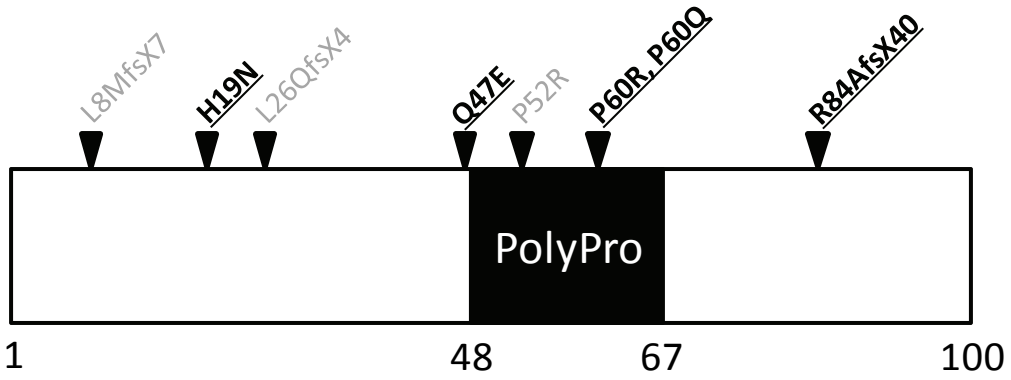


Figure 5.3. Missense and truncating variants found in the ASFMR1 protein.

The missense and truncating variants in grey have only been detected in developmentally delayed males. The underlined missense and truncating variants have been detected in healthy controls. While this depiction features a polyproline stretch 20 amino acids in length, based on the reference human genome (hg18), the length of this domain is dependent on the number of trinucleotide repeats in *ASFMR1*.

While p.P52R was detected in a developmentally delayed male with reduced FMRP expression (Tarleton et al. 2002), p.P60R and p.P60Q were detected in unaffected individuals (Cecconi et al. 2008). The first reported truncating variant, detected via direct sequencing, changes the splice donor site of exon 1, resulting in a premature stop codon. However, this variant also changes the splice acceptor site of exon 2 in *FMR1* and results in a lack of FMRP expression (Lugenbeel et al. 1995). The other two truncating variants in *ASFMR1* were detected via aberrant fragile X test results by PCR. In one case, testing of a developmentally delayed male with normal FMRP expression revealed a 5 bp deletion (Hegde et al. 2001) which encodes the p.L26QfsX4 frameshift. In a second report, however, a 49 bp duplication, initially detected in multiple patients referred to rule out FXS, was shown to be polymorphic in the Finnish population with a minor allele frequency of 0.2% (Mononen et al. 2007). This polymorphic duplication encodes the frameshift mutation p.R84AfsX40, which would be predicted to severely alter the structure and function of the ASFMR1 protein. The detection of a frameshift-inducing duplication in healthy controls suggests that the ASFMR1 protein is not essential for normal cognitive development.

5.3 Discussion

In summary, we sequenced 963 developmentally delayed males for intragenic mutations in the *ASFMR1* ORF, which is predicted to encode a 100 amino acid protein containing a polyproline domain. We identified two missense variants, p.H19N and p.Q47E, which had not previously been reported in healthy controls. Despite the high level of amino acid sequence conservation in the ASFMR1 protein among mammals,

control genotyping studies demonstrated that these variants do not associate with developmental delay. The absence of any causal *ASFMR1* mutations in such a large clinical sample discounts the gene as a significant single-locus cause of developmental delay.

Our subsequent search of the literature for rare variants in the *ASFMR1* ORF revealed three missense variants and three truncating variants whose effects on *ASFMR1* had not been previously characterized. One particular variant, which causes a frameshift in *ASFMR1*, has been shown to be polymorphic in a Finnish population (Mononen et al. 2007). The compatibility of a frameshift mutation with normal cognitive development suggests that the ASFMR1 protein does not play a role in the etiology of FXS.

Although these data show that the loss of ASFMR1 protein expression in FXS does not contribute to developmental delay, it remains possible that increased *ASFMR1* expression from fragile X premutation alleles may be involved in the premutation phenotypes FXTAS and primary ovarian insufficiency (Hagerman and Hagerman 2004). Furthermore, while our data refute the proposed role of the ASFMR1 protein in FXS, they do not rule out the possibility of the *ASFMR1* transcript contributing to FXS pathogenesis. Thus, further investigations into the role of *ASFMR1* in the etiology of FXS should focus upon a possible RNA-mediated mechanism.

Chapter 6. Concluding Remarks

6.1 Summary

Intellectual disability (ID) is a highly heterogeneous disorder, but the underlying cause is unknown in the majority of cases (Rauch et al. 2006). The most common inherited cause of ID is fragile X syndrome, a disorder resulting from mutations in the *FMR1* gene (Garber et al. 2008). Although several types of *FMR1* mutations have been identified, the standard of care is to test only for the most common of these, the expansion of a CGG trinucleotide repeat in the 5'UTR of the gene. Testing for repeat expansion can also, as a side benefit, reveal *FMR1* deletions that encompass the promoter and exon 1. Rarely assessed, however, is the possibility of sequence variants in *FMR1*.

In truth, there is no direct evidence to suggest that *FMR1* sequence variants are a significant cause of disease. However, previous attempts to detect point mutations in *FMR1* have been limited in size, scope, and sensitivity (see Table 1.1 for details). Furthermore, the prevailing school of thought has been self-fulfilling in the realm of clinical diagnostics, where the lack of known *FMR1* point mutations begets infrequent *FMR1* sequencing, which begets a lack of known *FMR1* point mutations. The possible clinical implications of this unanswered question, combined with the opportunity for increased functional understanding of *FMR1*, necessitated a more thorough assessment of the frequency of pathogenic sequence variants in *FMR1*.

Our initial hypothesis was that point mutations in *FMR1* would result in a fragile X-like phenotype. This hypothesis echoes the clinical use of diagnostic *FMR1*

sequencing, where patients who "look like fragile X" are the most likely to be tested for pathogenic *FMR1* sequence variants. To test this hypothesis, we collected DNA from 51 unrelated males who exhibit several fragile X-like features. We used resequencing arrays to sequence the promoter, coding sequence, and splice sites of *FMR1* in these patients, but found no functional point mutations. The lack of pathogenic sequence variants in this patient cohort suggests that *FMR1* point mutations are not a common cause of fragile X. Thus, the current clinical practice of pursuing *FMR1* sequencing in patients with a strong fragile X-like phenotype should be reevaluated. While this study does not rule out the possibility of *FMR1* point mutations causing a fragile X-like phenotype, it suggests that such sequencing is likely to be low in yield.

In light of our finding that *FMR1* point mutations are not a common cause of fragile X, we speculated that sequence variants may instead cause a more subtle, nonspecific phenotype. After all, missense changes are typically less damaging than a complete loss of gene expression, as occurs in the context of trinucleotide repeat expansion. Therefore, we collected aliquots of DNA from 963 males who tested negative for fragile X by repeat expansion testing at our affiliated clinical laboratory. Because every patient who presents with developmental delay is tested for the *FMR1* repeat expansion, this population has a less specific phenotype than fragile X.

With nearly a thousand samples to sequence, we decided to first develop a new method for efficient high-throughput sequencing. At that point in time, massively-parallel sequencing (MPS) platforms were entering the market. We designed a pooled-template method to allow the Gigabase-scale sequencing capacity of MPS to be efficiently used on a kilobase-scale genomic region, the *FMR1* gene. Through our design, we were able to

sequence 19 individuals at a time, vastly reducing the time and reagent cost of such a large sequencing project. Our method achieved the relatively high sensitivity of 76.5% and an acceptably high positive predictive value of 51.5%. Although other groups were concurrently developing similar methods (Druley et al. 2009; Ingman and Gyllensten 2009; Koboldt et al. 2009; Out et al. 2009), and ultimately published their techniques before we did, the performance metrics we achieved surpass those of many other groups.

By applying our pooled-template MPS method to the developmentally delayed male patient samples, we detected 130 novel sequence variants in *FMR1*. This catalogue of variants nearly doubles the number of known sequence variants in *FMR1*. While most of these are likely novel polymorphisms, several variants stood out for their functional implications. The solitary missense change we detected, p.R138Q, alters a conserved amino acid residue within the nuclear localization signal of FMRP. Three novel variants in the minimal *FMR1* promoter reduce the transcription of *FMR1*, as shown through luciferase assays. Also identified were several noncoding sequence variants of possible functional relevance, as determined by their sequence conservation, effects on splicing, and impact on miRNA binding.

Finally, in the course of sequencing *FMR1* in developmentally delayed patients, we were able to also examine the overlapping gene *ASFMR1* for variants in its coding sequence. Because *ASFMR1* is transcriptionally silent in the context of trinucleotide repeat expansion, much like *FMR1*, several groups have hypothesized that it may play a role in the etiology of fragile X (Ladd et al. 2007; Gecz et al. 2009). One mechanism by which it may act is through a putatively encoded 100 amino acid protein (Ladd et al. 2007). To investigate whether the functional absence of this protein plays a role in fragile

X, we looked for sequence variants in the *ASFMR1* open reading frame that would likely ablate the function of the ASFMR1 protein. While we identified no likely pathogenic missense changes in *ASFMR1* in developmentally delayed males, we did identify a polymorphism in the literature (Mononen et al. 2007) which causes a frameshift in *ASFMR1*. This frameshift would be expected to severely alter the ASFMR1 protein. Because this variant has been seen in several healthy controls (Mononen et al. 2007), we conclude that the ASFMR1 protein is not necessary for normal cognition and well-being, and therefore does not play a role in the etiology of fragile X.

6.2 Future Directions

In our work and in the work of others (Druley et al. 2009; Ingman and Gyllensten 2009; Koboldt et al. 2009; Out et al. 2009), the pooled-template MPS method has shown much promise as an effective strategy for novel and/or rare variant detection in a small genomic region across a large population. Several improvements would enhance its value for future use. First, the technical aspects of equimolar pooling should be improved. While the singleton sensitivity we achieved was higher than what others achieved (Out et al. 2009), our sensitivity would be unacceptably low for some applications, such as studies of the frequency of rare variants or the clinical detection of mutations. This, we believe, largely derives from inaccuracies in the creation of equimolar pools. An increased understanding of, and improved method of accounting for, the systematic biases in different MPS platforms would be another valuable development for pooled-template MPS. To our knowledge, we took greater steps than any other group to account for the systematic errors seen in MPS. Despite this, more than half of the false positives

detected by our method were seen recurrently across pools, and nearly half of our false positives occurred at known error-prone bases. By eliminating predictable false positives, pooled-template MPS would become a more specific test.

The findings of the pooled-template MPS of *FMR1* in developmentally delayed males also create many avenues for future study. First among these is the missense variant p.R138Q. Because this variant alters a residue believed to be functionally important in the nuclear localization signal of FMRP (Eberhart et al. 1996), a colleague (S. M. Bray) performed a preliminary assessment of the intracellular localization of FMRP in patient lymphoblasts. Surprisingly, no striking changes were noted in the nuclear and cytoplasmic proportions of FMRP. Furthermore, another colleague (M. Nakamoto) used a viral vector to transfect primary mouse hippocampal neurons in culture with *FMR1* containing the p.R138Q variant. The resulting dendritic AMPA receptor internalization was indistinguishable from neurons transfected with wild-type *FMR1*, indicating that the p.R138Q variant does not alter the synaptic function of FMRP.

Although it is possible that the p.R138Q variant is simply not functional, several experiments should be considered before ruling out a functional effect. In addition to its proposed role in the nuclear localization signal of FMRP, arginine-138 is also located at the edge of the N-terminal domain of FMRP (Figure 1.4), which is known to mediate many of the protein-protein interactions of FMRP (Adinolfi et al. 2003). Thus, a relatively simple first experiment would be to assess the protein-protein interactions of the variant FMRP by immunoprecipitation. A similar approach could also be employed to assess the mRNA transcripts bound by the variant FMRP. Because FMRP is proposed to enter the nucleus to bind RNA (Feng et al. 1997b; Kim et al. 2009), it is possible that a

subtle alteration to the nuclear localization signal would result in an expanded, reduced, or simply different subset of RNA species bound. The most informative evidence of functionality, however, may come from an animal model, in which a more complete range of functional and phenotypic assessments can be made.

While the p.R138Q variant is lacking functional evidence of a pathogenic role, the three promoter variants are currently lacking a clinical correlate for their observed reduction of *FMR1* transcription. We hypothesize that the reduced, but not ablated, promoter activity likely causes a mild intellectual disability or a mild fragile X-like phenotype, similar to what has been seen in patients with reduced *FMR1* expression due to mosaic deletions (Han et al. 2006; Coffee et al. 2008). It will be important to characterize the patients' phenotypes to test this hypothesis and to guide the future clinical use of *FMR1* promoter sequencing. Furthermore, patient re-contact may create the opportunity for family studies, which could support or refute an etiologic role for the variants.

The promoter variants also create the opportunity for functional assessment of the *FMR1* promoter. Previous studies of the *FMR1* promoter have largely relied on deletion constructs to determine the functional roles of various promoter elements (Kumari and Usdin 2001). However, deletions can disrupt the three-dimensional structure of the promoter, thus creating an artificial context that may not accurately represent the *in vivo* function (Kumari et al. 2005). The novel variants we identified will allow functional studies to be carried out in a more representative context. For instance, footprinting studies may reveal that a lack of Sp1 binding to the variant-containing GC box explains the associated reduction in transcriptional activity. Additionally, with promoter variants

identified in two of the three Initiator-like elements, functional studies of these changes may illuminate the relative importance of the three described transcriptional start sites in *FMR1*.

The novel noncoding variants identified in developmentally delayed males are a third class of variant worthy of functional investigation. While their possible functional roles are not as obvious as changes to the coding sequence or promoter, several variants were noted to alter conserved bases, suggesting that they may have a functional impact. Several of these variants were not identified in a large population of Caucasian controls. However, some of the variants were originally identified in patients of unknown ancestry or in African-American patients, making our control population less informative. For these variants, control genotyping should be pursued in African-Americans or in the Human Genome Diversity Panel to determine if they simply represent polymorphisms that, while rare in Caucasians, are at an appreciable minor allele frequency in a certain race. Following this analysis, any variants showing evidence of association with developmental delay should then be assessed for a functional role in splicing or transcriptional regulation.

6.3 A final word

Ultimately, the greatest impact of the research presented in this dissertation will hopefully be upon the families affected by intellectual disability of currently unknown origin. Although all families affected by intellectual disability face struggles, stresses, and financial burdens (Olsson and Hwang 2001; Centers for Disease Control and Prevention 2004; Chou et al. 2008; Mulroy et al. 2008), the strains are even greater when

families have no diagnosis, prognosis, or treatment options for their child (Rosenthal et al. 2001; Lenhard et al. 2005). By identifying four *FMR1* sequence variants that show evidence of association with developmental delay, we have demonstrated the value of *FMR1* sequencing as a diagnostic test for intellectually disabled patients. If clinically implemented, diagnostic *FMR1* sequencing would offer answers, and hope, to some number of families affected by ID. Fragile X support groups and novel, targeted therapeutics currently in development may be as helpful for this group of patients as for those with classic fragile X. There certainly will be challenges in the broad clinical implementation of *FMR1* sequencing, such as the difficulty of interpreting the functionality of novel variants, even those in the coding sequence. Additionally, the low frequency of functional *FMR1* variants in our undifferentiated population of developmentally delayed males suggests that a more specific phenotype may need to be described for such testing to have a high yield. However, with the increasing capacity and decreasing cost of next-generation sequencing, one can imagine a not-so-distant future where a collection of known ID genes are sequenced simultaneously for the efficient and thorough detection of causal mutations, thereby decreasing the burden of intellectual disability of unknown origin.

Chapter 7. References

Adinolfi, S., Ramos, A., Martin, S.R., Dal Piaz, F., Pucci, P., Bardoni, B., Mandel, J.L., and Pastore, A. 2003. The N-terminus of the fragile X mental retardation protein contains a novel domain involved in dimerization and RNA binding. *Biochemistry* **42**(35): 10437-10444.

Ahn, S.M., Kim, T.H., Lee, S., Kim, D., Ghang, H., Kim, D.S., Kim, B.C., Kim, S.Y., Kim, W.Y., Kim, C. et al. 2009. The first Korean genome sequence and analysis: full genome sequencing for a socio-ethnic group. *Genome Res* **19**(9): 1622-1629.

American Psychiatric Association. 2000. *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*. American Psychiatric Association, Washington, DC.

Antar, L.N., Afroz, R., Dictenberg, J.B., Carroll, R.C., and Bassell, G.J. 2004. Metabotropic glutamate receptor activation regulates fragile x mental retardation protein and FMR1 mRNA localization differentially in dendrites and at synapses. *J Neurosci* **24**(11): 2648-2655.

Antar, L.N., Dictenberg, J.B., Plociniak, M., Afroz, R., and Bassell, G.J. 2005. Localization of FMRP-associated mRNA granules and requirement of microtubules for activity-dependent trafficking in hippocampal neurons. *Genes Brain Behav* **4**(6): 350-359.

Antar, L.N., Li, C., Zhang, H., Carroll, R.C., and Bassell, G.J. 2006. Local functions for FMRP in axon growth cone motility and activity-dependent regulation of filopodia and spine synapses. *Mol Cell Neurosci* **32**(1-2): 37-48.

Ashley, C.T., Jr., Wilkinson, K.D., Reines, D., and Warren, S.T. 1993a. FMR1 protein: conserved RNP family domains and selective RNA binding. *Science* **262**(5133): 563-566.

Ashley, C.T., Sutcliffe, J.S., Kunst, C.B., Leiner, H.A., Eichler, E.E., Nelson, D.L., and Warren, S.T. 1993b. Human and murine FMR-1: alternative splicing and translational initiation downstream of the CGG-repeat. *Nat Genet* **4**(3): 244-251.

Bailey, D.B., Jr., Mesibov, G.B., Hatton, D.D., Clark, R.D., Roberts, J.E., and Mayhew, L. 1998. Autistic behavior in young boys with fragile X syndrome. *J Autism Dev Disord* **28**(6): 499-508.

Bailey, D.B., Jr., Skinner, D., and Sparkman, K.L. 2003. Discovering fragile X syndrome: family experiences and perceptions. *Pediatrics* **111**(2): 407-416.

Bakker CE, V.C., Willemsen R, van der Helm R, Oerlemans F, Vermey, M, B.A., Hoogeveen AT, Oostra BA, Reyniers E, De Boulle D,, and Nagels G, M.J.-J., De Deyn PP, Darby JK, Willems PJ. 1994. Fmr1 knockout mice: a model to study fragile X mental retardation. The Dutch-Belgian Fragile X Consortium. *Cell* **78**(1): 23-33.

Bardoni, B., Schenck, A., and Mandel, J.L. 1999. A novel RNA-binding nuclear protein that interacts with the fragile X mental retardation (FMR1) protein. *Hum Mol Genet* **8**(13): 2557-2566.

Bardoni, B., Sittler, A., Shen, Y., and Mandel, J.L. 1997. Analysis of domains affecting intracellular localization of the FMRP protein. *Neurobiol Dis* **4**(5): 329-336.

Basel-Vanagaite, L. 2007. Genetics of autosomal recessive non-syndromic mental retardation: recent advances. *Clin Genet* **72**(3): 167-174.

Bassell, G.J. and Warren, S.T. 2008. Fragile X syndrome: loss of local mRNA regulation alters synaptic development and function. *Neuron* **60**(2): 201-214.

Bear, M.F., Huber, K.M., and Warren, S.T. 2004. The mGluR theory of fragile X mental retardation. *Trends Neurosci* **27**(7): 370-377.

Beilina, A., Tassone, F., Schwartz, P.H., Sahota, P., and Hagerman, P.J. 2004. Redistribution of transcription start sites within the FMR1 promoter region with expansion of the downstream CGG-repeat element. *Hum Mol Genet* **13**(5): 543-549.

Bentley, D.R. Balasubramanian, S. Swerdlow, H.P. Smith, G.P. Milton, J. Brown, C.G. Hall, K.P. Evers, D.J. Barnes, C.L. Bignell, H.R. et al. 2008. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**(7218): 53-59.

Berry-Kravis, E., Hessl, D., Coffey, S., Hervey, C., Schneider, A., Yuhas, J., Hutchison, J., Snape, M., Tranfaglia, M., Nguyen, D.V. et al. 2009. A pilot open label, single dose trial of fenobam in adults with fragile X syndrome. *J Med Genet* **46**(4): 266-271.

Berry-Kravis, E., Krause, S.E., Block, S.S., Guter, S., Wuu, J., Leurgans, S., Decle, P., Potanos, K., Cook, E., Salt, J. et al. 2006. Effect of CX516, an AMPA-modulating compound, on cognition and behavior in fragile X syndrome: a controlled trial. *J Child Adolesc Psychopharmacol* **16**(5): 525-540.

Berry-Kravis, E. and Potanos, K. 2004. Psychopharmacology in fragile X syndrome--present and future. *Ment Retard Dev Disabil Res Rev* **10**(1): 42-48.

Betel, D., Wilson, M., Gabow, A., Marks, D.S., and Sander, C. 2008. The microRNA.org resource: targets and expression. *Nucleic Acids Res* **36**(Database issue): D149-153.

Biancalana, V., Beldjord, C., Taillandier, A., Szpiro-Tapia, S., Cusin, V., Gerson, F., Philippe, C., and Mandel, J.L. 2004. 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* **129A**(3): 218-224.

Bontekoe, C.J., Bakker, C.E., Nieuwenhuizen, I.M., van der Linde, H., Lans, H., de Lange, D., Hirst, M.C., and Oostra, B.A. 2001. Instability of a (CGG)98 repeat in the Fmr1 promoter. *Hum Mol Genet* **10**(16): 1693-1699.

Bowen, P., Biederman, B., and Swallow, K.A. 1978. The X-linked syndrome of macroorchidism and mental retardation: further observations. *Am J Med Genet* **2**(4): 409-414.

Brown, V., Jin, P., Ceman, S., Darnell, J.C., O'Donnell, W.T., Tenenbaum, S.A., Jin, X., Feng, Y., Wilkinson, K.D., Keene, J.D. et al. 2001. Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* **107**(4): 477-487.

Bureau, I., Shepherd, G.M., and Svoboda, K. 2008. Circuit and plasticity defects in the developing somatosensory cortex of FMR1 knock-out mice. *J Neurosci* **28**(20): 5178-5188.

Butler, M.G., Allen, G.A., Haynes, J.L., Singh, D.N., Watson, M.S., and Breg, W.R. 1991a. Anthropometric comparison of mentally retarded males with and without the fragile X syndrome. *Am J Med Genet* **38**(2-3): 260-268.

Butler, M.G., Mangrum, T., Gupta, R., and Singh, D.N. 1991b. A 15-item checklist for screening mentally retarded males for the fragile X syndrome. *Clin Genet* **39**(5): 347-354.

Cantu, J.M., Scaglia, H.E., Medina, M., Gonzalez-Diddi, M., Morato, T., Moreno, M.E., and Perez-Palacios, G. 1976. Inherited congenital normofunctional testicular hyperplasia and mental deficiency. *Hum Genet* **33**(1): 23-33.

Castellvi-Bel, S., Sanchez, A., Badenas, C., Mallolas, J., Barcelo, A., Jimenez, D., Villa, M., Estivill, X., and Mila, M. 1999. Single-strand conformation polymorphism analysis in the FMR1 gene. *Am J Med Genet* **84**(3): 262-265.

Caudy, A.A., Myers, M., Hannon, G.J., and Hammond, S.M. 2002. Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev* **16**(19): 2491-2496.

Cecconi, M., Forzano, F., Rinaldi, R., Cappellacci, S., Grammatico, P., Faravelli, F., Dagna Bricarelli, F., Di Maria, E., and Grasso, M. 2008. A single nucleotide variant in the FMR1 CGG repeat results in a "Pseudodeletion" and is not associated with the fragile X syndrome phenotype. *J Mol Diagn* **10**(3): 272-275.

Ceman, S., Brown, V., and Warren, S.T. 1999. Isolation of an FMRP-associated messenger ribonucleoprotein particle and identification of nucleolin and the fragile X-related proteins as components of the complex. *Mol Cell Biol* **19**(12): 7925-7932.

Ceman, S., Nelson, R., and Warren, S.T. 2000. Identification of mouse YB1/p50 as a component of the FMRP-associated mRNP particle. *Biochem Biophys Res Commun* **279**(3): 904-908.

Ceman, S., O'Donnell, W.T., Reed, M., Patton, S., Pohl, J., and Warren, S.T. 2003. Phosphorylation influences the translation state of FMRP-associated polyribosomes. *Hum Mol Genet* **12**(24): 3295-3305.

Centers for Disease Control and Prevention. 1996. State-specific rates of mental retardation--United States, 1993. *MMWR Morb Mortal Wkly Rep* **45**(3): 61-65.

-. 2004. Economic costs associated with mental retardation, cerebral palsy, hearing loss, and vision impairment--United States, 2003. *MMWR Morb Mortal Wkly Rep* **53**(3): 57-59.

Centonze, D., Rossi, S., Mercaldo, V., Napoli, I., Ciotti, M.T., De Chiara, V., Musella, A., Prosperetti, C., Calabresi, P., Bernardi, G. et al. 2008. Abnormal striatal GABA transmission in the mouse model for the fragile X syndrome. *Biol Psychiatry* **63**(10): 963-973.

Chahrour, M. and Zoghbi, H.Y. 2007. The story of Rett syndrome: from clinic to neurobiology. *Neuron* **56**(3): 422-437.

Chang, S., Bray, S.M., Li, Z., Zarnescu, D.C., He, C., Jin, P., and Warren, S.T. 2008. Identification of small molecules rescuing fragile X syndrome phenotypes in Drosophila. *Nat Chem Biol* **4**(4): 256-263.

Chen, L.S., Tassone, F., Sahota, P., and Hagerman, P.J. 2003. The (CGG)n repeat element within the 5' untranslated region of the FMR1 message provides both

positive and negative cis effects on in vivo translation of a downstream reporter. *Hum Mol Genet* **12**(23): 3067-3074.

Chiurazzi, P., de Graaff, E., Ng, J., Verkerk, A.J., Wolfson, S., Fisch, G.S., Kozak, L., Neri, G., and Oostra, B.A. 1994. No apparent involvement of the FMR1 gene in five patients with phenotypic manifestations of the fragile X syndrome. *Am J Med Genet* **51**(4): 309-314.

Chiurazzi, P., Pomponi, M.G., Willemsen, R., Oostra, B.A., and Neri, G. 1998. In vitro reactivation of the FMR1 gene involved in fragile X syndrome. *Hum Mol Genet* **7**(1): 109-113.

Chou, Y.C., Tzou, P.Y., Pu, C.Y., Kroger, T., and Lee, W.P. 2008. Respite care as a community care service: factors associated with the effects on family carers of adults with intellectual disability in Taiwan. *J Intellect Dev Disabil* **33**(1): 12-21.

Chowdhury, S., Shepherd, J.D., Okuno, H., Lyford, G., Petralia, R.S., Plath, N., Kuhl, D., Huganir, R.L., and Worley, P.F. 2006. Arc/Arg3.1 interacts with the endocytic machinery to regulate AMPA receptor trafficking. *Neuron* **52**(3): 445-459.

Coffee, B., Ikeda, M., Budimirovic, D.B., Hjelm, L.N., Kaufmann, W.E., and Warren, S.T. 2008. 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* **146A**(10): 1358-1367.

Coffee, B., Keith, K., Albizua, I., Malone, T., Mowrey, J., Sherman, S.L., and Warren, S.T. 2009. Incidence of fragile X syndrome by newborn screening for methylated FMR1 DNA. *Am J Hum Genet* **85**(4): 503-514.

Coffee, B., Zhang, F., Ceman, S., Warren, S.T., and Reines, D. 2002. Histone modifications depict an aberrantly heterochromatinized FMR1 gene in fragile x syndrome. *Am J Hum Genet* **71**(4): 923-932.

Coffee, B., Zhang, F., Warren, S.T., and Reines, D. 1999. Acetylated histones are associated with FMR1 in normal but not fragile X-syndrome cells. *Nat Genet* **22**(1): 98-101.

Comery, T.A., Harris, J.B., Willems, P.J., Oostra, B.A., Irwin, S.A., Weiler, I.J., and Greenough, W.T. 1997. Abnormal dendritic spines in fragile X knockout mice: maturation and pruning deficits. *Proc Natl Acad Sci U S A* **94**(10): 5401-5404.

Corbin, F., Bouillon, M., Fortin, A., Morin, S., Rousseau, F., and Khandjian, E.W. 1997. The fragile X mental retardation protein is associated with poly(A)+ mRNA in actively translating polyribosomes. *Hum Mol Genet* **6**(9): 1465-1472.

Craig, D.W., Pearson, J.V., Szelinger, S., Sekar, A., Redman, M., Corneveaux, J.J., Pawlowski, T.L., Laub, T., Nunn, G., Stephan, D.A. et al. 2008. Identification of genetic variants using bar-coded multiplexed sequencing. *Nat Methods* **5**(10): 887-893.

Crawford, D.C., Acuna, J.M., and Sherman, S.L. 2001. FMR1 and the fragile X syndrome: human genome epidemiology review. *Genet Med* **3**(5): 359-371.

Curry, C.J., Stevenson, R.E., Aughton, D., Byrne, J., Carey, J.C., Cassidy, S., Cunniff, C., Graham, J.M., Jr., Jones, M.C., Kaback, M.M. et al. 1997. Evaluation of mental retardation: recommendations of a Consensus Conference: American College of Medical Genetics. *Am J Med Genet* **72**(4): 468-477.

Cutler, D.J., Zwick, M.E., Carrasquillo, M.M., Yohn, C.T., Tobin, K.P., Kashuk, C., Mathews, D.J., Shah, N.A., Eichler, E.E., Warrington, J.A. et al. 2001. High-

throughput variation detection and genotyping using microarrays. *Genome Res* **11**(11): 1913-1925.

D'Hulst, C., De Geest, N., Reeve, S.P., Van Dam, D., De Deyn, P.P., Hassan, B.A., and Kooy, R.F. 2006. Decreased expression of the GABAA receptor in fragile X syndrome. *Brain Res* **1121**(1): 238-245.

Dahl, F., Stenberg, J., Fredriksson, S., Welch, K., Zhang, M., Nilsson, M., Bicknell, D., Bodmer, W.F., Davis, R.W., and Ji, H. 2007. Multigene amplification and massively parallel sequencing for cancer mutation discovery. *Proc Natl Acad Sci U S A* **104**(22): 9387-9392.

Darnell, J.C., Fraser, C.E., Mostovetsky, O., Stefani, G., Jones, T.A., Eddy, S.R., and Darnell, R.B. 2005. Kissing complex RNAs mediate interaction between the Fragile-X mental retardation protein KH2 domain and brain polyribosomes. *Genes Dev* **19**(8): 903-918.

Darnell, J.C., Jensen, K.B., Jin, P., Brown, V., Warren, S.T., and Darnell, R.B. 2001. Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell* **107**(4): 489-499.

Davidkova, G. and Carroll, R.C. 2007. Characterization of the role of microtubule-associated protein 1B in metabotropic glutamate receptor-mediated endocytosis of AMPA receptors in hippocampus. *J Neurosci* **27**(48): 13273-13278.

De Boulle, K., Verkerk, A.J., Reyniers, E., Vits, L., Hendrickx, J., Van Roy, B., Van den Bos, F., de Graaff, E., Oostra, B.A., and Willems, P.J. 1993. A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nat Genet* **3**(1): 31-35.

de Graaff, E., Rouillard, P., Willems, P.J., Smits, A.P., Rousseau, F., and Oostra, B.A. 1995. Hotspot for deletions in the CGG repeat region of FMR1 in fragile X patients. *Hum Mol Genet* **4**(1): 45-49.

de Vries, B.B., Wiegers, A.M., Smits, A.P., Mohkamsing, S., Duivenvoorden, H.J., Fryns, J.P., Curfs, L.M., Halley, D.J., Oostra, B.A., van den Ouweland, A.M. et al. 1996. Mental status of females with an FMR1 gene full mutation. *Am J Hum Genet* **58**(5): 1025-1032.

de Vrij, F.M., Levenga, J., van der Linde, H.C., Koekkoek, S.K., De Zeeuw, C.I., Nelson, D.L., Oostra, B.A., and Willemsen, R. 2008. Rescue of behavioral phenotype and neuronal protrusion morphology in Fmr1 KO mice. *Neurobiol Dis* **31**(1): 127-132.

Dean, M. 1995. Resolving DNA mutations. *Nat Genet* **9**(2): 103-104.

Devys, D., Lutz, Y., Rouyer, N., Bellocq, J.P., and Mandel, J.L. 1993. The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. *Nat Genet* **4**(4): 335-340.

Dictenberg, J.B., Swanger, S.A., Antar, L.N., Singer, R.H., and Bassell, G.J. 2008. A direct role for FMRP in activity-dependent dendritic mRNA transport links filopodial-spine morphogenesis to fragile X syndrome. *Dev Cell* **14**(6): 926-939.

Dobkin, C., Rabe, A., Dumas, R., El Idrissi, A., Haubenstock, H., and Brown, W.T. 2000. Fmr1 knockout mouse has a distinctive strain-specific learning impairment. *Neuroscience* **100**(2): 423-429.

Dohm, J.C., Lottaz, C., Borodina, T., and Himmelbauer, H. 2008. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res* **36**(16): e105.

Dolen, G., Osterweil, E., Rao, B.S., Smith, G.B., Auerbach, B.D., Chattarji, S., and Bear, M.F. 2007. Correction of fragile X syndrome in mice. *Neuron* **56**(6): 955-962.

Drouin, R., Angers, M., Dallaire, N., Rose, T.M., Khandjian, E.W., and Rousseau, F. 1997. Structural and functional characterization of the human FMR1 promoter reveals similarities with the hnRNP-A2 promoter region. *Hum Mol Genet* **6**(12): 2051-2060.

Druley, T.E., Vallania, F.L., Wegner, D.J., Varley, K.E., Knowles, O.L., Bonds, J.A., Robison, S.W., Doniger, S.W., Hamvas, A., Cole, F.S. et al. 2009. Quantification of rare allelic variants from pooled genomic DNA. *Nat Methods* **6**(4): 263-265.

Durkin, M. 2002. The epidemiology of developmental disabilities in low-income countries. *Ment Retard Dev Disabil Res Rev* **8**(3): 206-211.

Eberhart, D.E., Malter, H.E., Feng, Y., and Warren, S.T. 1996. The fragile X mental retardation protein is a ribonucleoprotein containing both nuclear localization and nuclear export signals. *Hum Mol Genet* **5**(8): 1083-1091.

Eichler, E.E., Macpherson, J.N., Murray, A., Jacobs, P.A., Chakravarti, A., and Nelson, D.L. 1996. Haplotype and interspersion analysis of the FMR1 CGG repeat identifies two different mutational pathways for the origin of the fragile X syndrome. *Hum Mol Genet* **5**(3): 319-330.

Eichler, E.E., Richards, S., Gibbs, R.A., and Nelson, D.L. 1993. Fine structure of the human FMR1 gene. *Hum Mol Genet* **2**(8): 1147-1153.

Entezam, A. and Usdin, K. 2008. ATR protects the genome against CGG.CCG-repeat expansion in Fragile X premutation mice. *Nucleic Acids Res* **36**(3): 1050-1056.

Feng, Y., Absher, D., Eberhart, D.E., Brown, V., Malter, H.E., and Warren, S.T. 1997a. FMRP associates with polyribosomes as an mRNP, and the I304N mutation of severe fragile X syndrome abolishes this association. *Mol Cell* **1**(1): 109-118.

Feng, Y., Gutekunst, C.A., Eberhart, D.E., Yi, H., Warren, S.T., and Hersch, S.M. 1997b. Fragile X mental retardation protein: nucleocytoplasmic shuttling and association with somatodendritic ribosomes. *J Neurosci* **17**(5): 1539-1547.

Ferrer-Costa, C., Gelpi, J.L., Zamakola, L., Parraga, I., de la Cruz, X., and Orozco, M. 2005. PMUT: a web-based tool for the annotation of pathological mutations on proteins. *Bioinformatics* **21**(14): 3176-3178.

Ferrero, G.B., Howald, C., Micale, L., Biamino, E., Augello, B., Fusco, C., Turturo, M.G., Forzano, S., Reymond, A., and Merla, G. 2009. An atypical 7q11.23 deletion in a normal IQ Williams-Beuren syndrome patient. *Eur J Hum Genet*.

Freund, L.S., Reiss, A.L., and Abrams, M.T. 1993. Psychiatric disorders associated with fragile X in the young female. *Pediatrics* **91**(2): 321-329.

Fridell, R.A., Benson, R.E., Hua, J., Bogerd, H.P., and Cullen, B.R. 1996. A nuclear role for the Fragile X mental retardation protein. *EMBO J* **15**(19): 5408-5414.

Fu, Y.H., Kuhl, D.P., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J., Holden, J.J., Fenwick, R.G., Jr., Warren, S.T. et al. 1991. Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* **67**(6): 1047-1058.

Garber, K.B., Visootsak, J., and Warren, S.T. 2008. Fragile X syndrome. *Eur J Hum Genet* **16**(6): 666-672.

Gecz, J., Shoubridge, C., and Corbett, M. 2009. The genetic landscape of intellectual disability arising from chromosome X. *Trends Genet* **25**(7): 308-316.

Gedeon, A.K., Baker, E., Robinson, H., Partington, M.W., Gross, B., Manca, A., Korn, B., Poustka, A., Yu, S., Sutherland, G.R. et al. 1992. Fragile X syndrome without CCG amplification has an FMR1 deletion. *Nat Genet* **1**(5): 341-344.

Gerber, H.P., Seipel, K., Georgiev, O., Hofferer, M., Hug, M., Rusconi, S., and Schaffner, W. 1994. Transcriptional activation modulated by homopolymeric glutamine and proline stretches. *Science* **263**(5148): 808-811.

Gronskov, K., Hallberg, A., and Brondum-Nielsen, K. 1998. Mutational analysis of the FMR1 gene in 118 mentally retarded males suspected of fragile X syndrome: absence of prevalent mutations. *Hum Genet* **102**(4): 440-445.

Guruju, M.R., Lavanya, K., Thelma, B.K., Sujatha, M., OmSai, V.R., Nagarathna, V., Amarjyothi, P., Jyothi, A., and Anandaraj, M.P. 2009. Assessment of a clinical checklist in the diagnosis of fragile X syndrome in India. *J Clin Neurosci* **16**(10): 1305-1310.

Hagerman, P.J. and Hagerman, R.J. 2004. The fragile-X premutation: a maturing perspective. *Am J Hum Genet* **74**(5): 805-816.

Hagerman, R., Kemper, M., and Hudson, M. 1985. Learning disabilities and attentional problems in boys with the fragile X syndrome. *Am J Dis Child* **139**(7): 674-678.

Hagerman, R.J., Altshul-Stark, D., and McBogg, P. 1987. Recurrent otitis media in the fragile X syndrome. *Am J Dis Child* **141**(2): 184-187.

Hagerman, R.J., Amiri, K., and Cronister, A. 1991. Fragile X checklist. *Am J Med Genet* **38**(2-3): 283-287.

Hagerman, R.J., Berry-Kravis, E., Kaufmann, W.E., Ono, M.Y., Tartaglia, N., Lachiewicz, A., Kronk, R., Delahunty, C., Hessl, D., Visootsak, J. et al. 2009. Advances in the treatment of fragile X syndrome. *Pediatrics* **123**(1): 378-390.

Hagerman, R.J., Murphy, M.A., and Wittenberger, M.D. 1988. A controlled trial of stimulant medication in children with the fragile X syndrome. *Am J Med Genet* **30**(1-2): 377-392.

Hagerman, R.J., Van Housen, K., Smith, A.C., and McGavran, L. 1984. Consideration of connective tissue dysfunction in the fragile X syndrome. *Am J Med Genet* **17**(1): 111-121.

Hall, S.S., Burns, D.D., Lightbody, A.A., and Reiss, A.L. 2008. Longitudinal changes in intellectual development in children with Fragile X syndrome. *J Abnorm Child Psychol* **36**(6): 927-939.

Han, X.D., Powell, B.R., Phalin, J.L., and Chehab, F.F. 2006. Mosaicism for a full mutation, premutation, and deletion of the CGG repeats results in 22% FMRP and elevated FMR1 mRNA levels in a high-functioning fragile X male. *Am J Med Genet A* **140**(13): 1463-1471.

Hanauer, A. and Young, I.D. 2002. Coffin-Lowry syndrome: clinical and molecular features. *J Med Genet* **39**(10): 705-713.

Harismendy, O., Ng, P.C., Strausberg, R.L., Wang, X., Stockwell, T.B., Beeson, K.Y., Schork, N.J., Murray, S.S., Topol, E.J., Levy, S. et al. 2009. Evaluation of next

generation sequencing platforms for population targeted sequencing studies. *Genome Biol* **10**(3): R32.

Harrison, C.J., Jack, E.M., Allen, T.D., and Harris, R. 1983. The fragile X: a scanning electron microscope study. *J Med Genet* **20**(4): 280-285.

Hatton, D.D., Sideris, J., Skinner, M., Mankowski, J., Bailey, D.B., Jr., Roberts, J., and Mirrett, P. 2006. Autistic behavior in children with fragile X syndrome: prevalence, stability, and the impact of FMRP. *Am J Med Genet A* **140A**(17): 1804-1813.

Heber, R. 1961. Modifications in the manual on terminology and classification in mental retardation. *Am J Ment Defic* **65**: 499-500.

Hecimovic, S., Tarnik, I.P., Baric, I., Cakarun, Z., and Pavelic, K. 2002. Screening for fragile X syndrome: results from a school for mentally retarded children. *Acta Paediatr* **91**(5): 535-539.

Hegde, M.R., Chong, B., Fawkner, M., Lambiris, N., Peters, H., Kenneson, A., Warren, S.T., Love, D.R., and McGaughran, J. 2001. Microdeletion in the FMR-1 gene: an apparent null allele using routine clinical PCR amplification. *J Med Genet* **38**(9): 624-629.

Heitz, D., Rousseau, F., Devys, D., Saccone, S., Abderrahim, H., Le Paslier, D., Cohen, D., Vincent, A., Toniolo, D., Della Valle, G. et al. 1991. Isolation of sequences that span the fragile X and identification of a fragile X-related CpG island. *Science* **251**(4998): 1236-1239.

Hengst, U., Cox, L.J., Macosko, E.Z., and Jaffrey, S.R. 2006. Functional and selective RNA interference in developing axons and growth cones. *J Neurosci* **26**(21): 5727-5732.

Hergersberg, M., Matsuo, K., Gassmann, M., Schaffner, W., Luscher, B., Rulicke, T., and Aguzzi, A. 1995. Tissue-specific expression of a FMR1/beta-galactosidase fusion gene in transgenic mice. *Hum Mol Genet* **4**(3): 359-366.

Hernandez, R.N., Feinberg, R.L., Vaurio, R., Passanante, N.M., Thompson, R.E., and Kaufmann, W.E. 2009. Autism spectrum disorder in fragile X syndrome: a longitudinal evaluation. *Am J Med Genet A* **149A**(6): 1125-1137.

Hinds, H.L., Ashley, C.T., Sutcliffe, J.S., Nelson, D.L., Warren, S.T., Housman, D.E., and Schalling, M. 1993. Tissue specific expression of FMR-1 provides evidence for a functional role in fragile X syndrome. *Nat Genet* **3**(1): 36-43.

Hinton, V.J., Brown, W.T., Wisniewski, K., and Rudelli, R.D. 1991. Analysis of neocortex in three males with the fragile X syndrome. *Am J Med Genet* **41**(3): 289-294.

Hirst, M., Grewal, P., Flannery, A., Slatter, R., Maher, E., Barton, D., Fryns, J.P., and Davies, K. 1995. Two new cases of FMR1 deletion associated with mental impairment. *Am J Hum Genet* **56**(1): 67-74.

Hirst, M.C. and White, P.J. 1998. Cloned human FMR1 trinucleotide repeats exhibit a length- and orientation-dependent instability suggestive of in vivo lagging strand secondary structure. *Nucleic Acids Res* **26**(10): 2353-2358.

Hsieh, H., Boehm, J., Sato, C., Iwatsubo, T., Tomita, T., Sisodia, S., and Malinow, R. 2006. AMPAR removal underlies Abeta-induced synaptic depression and dendritic spine loss. *Neuron* **52**(5): 831-843.

Huber, K.M., Gallagher, S.M., Warren, S.T., and Bear, M.F. 2002. Altered synaptic plasticity in a mouse model of fragile X mental retardation. *Proc Natl Acad Sci U S A* **99**(11): 7746-7750.

Hwu, W.L., Lee, Y.M., Lee, S.C., and Wang, T.R. 1993. In vitro DNA methylation inhibits FMR-1 promoter. *Biochem Biophys Res Commun* **193**(1): 324-329.

Ingman, M. and Gyllensten, U. 2009. SNP frequency estimation using massively parallel sequencing of pooled DNA. *Eur J Hum Genet* **17**(3): 383-386.

Irwin, S.A., Idupulapati, M., Gilbert, M.E., Harris, J.B., Chakravarti, A.B., Rogers, E.J., Crisostomo, R.A., Larsen, B.P., Mehta, A., Alcantara, C.J. et al. 2002. Dendritic spine and dendritic field characteristics of layer V pyramidal neurons in the visual cortex of fragile-X knockout mice. *Am J Med Genet* **111**(2): 140-146.

Ishizuka, A., Siomi, M.C., and Siomi, H. 2002. A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes Dev* **16**(19): 2497-2508.

Jacquemont, S., Hagerman, R.J., Leehey, M., Grigsby, J., Zhang, L., Brunberg, J.A., Greco, C., Des Portes, V., Jardini, T., Levine, R. et al. 2003. Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. *Am J Hum Genet* **72**(4): 869-878.

Jin, P., Zarnescu, D.C., Ceman, S., Nakamoto, M., Mowrey, J., Jongens, T.A., Nelson, D.L., Moses, K., and Warren, S.T. 2004. Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. *Nat Neurosci* **7**(2): 113-117.

Jin, P., Zarnescu, D.C., Zhang, F., Pearson, C.E., Lucchesi, J.C., Moses, K., and Warren, S.T. 2003. RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in Drosophila. *Neuron* **39**(5): 739-747.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., and Marks, D.S. 2004. Human MicroRNA targets. *PLoS Biol* **2**(11): e363.

Kaufmann, W.E., Cortell, R., Kau, A.S., Bukelis, I., Tierney, E., Gray, R.M., Cox, C., Capone, G.T., and Stanard, P. 2004. Autism spectrum disorder in fragile X syndrome: communication, social interaction, and specific behaviors. *Am J Med Genet A* **129A**(3): 225-234.

Kemper, M.B., Hagerman, R.J., and Altshul-Stark, D. 1988. Cognitive profiles of boys with the fragile X syndrome. *Am J Med Genet* **30**(1-2): 191-200.

Kenneson, A., Zhang, F., Hagedorn, C.H., and Warren, S.T. 2001. Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Hum Mol Genet* **10**(14): 1449-1454.

Khalil, A.M., Faghihi, M.A., Modarresi, F., Brothers, S.P., and Wahlestedt, C. 2008. A novel RNA transcript with antiapoptotic function is silenced in fragile X syndrome. *PLoS One* **3**(1): e1486.

Khandjian, E.W., Corbin, F., Woerly, S., and Rousseau, F. 1996. The fragile X mental retardation protein is associated with ribosomes. *Nat Genet* **12**(1): 91-93.

Kim, M., Bellini, M., and Ceman, S. 2009. Fragile X mental retardation protein FMRP binds mRNAs in the nucleus. *Mol Cell Biol* **29**(1): 214-228.

Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., Weinstock, G.M., Wilson, R.K., and Ding, L. 2009. VarScan: variant detection in

massively parallel sequencing of individual and pooled samples. *Bioinformatics* **25**(17): 2283-2285.

Kremer, E.J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S.T., Schlessinger, D., Sutherland, G.R., and Richards, R.I. 1991. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science* **252**(5013): 1711-1714.

Kumari, D., Gabrielian, A., Wheeler, D., and Usdin, K. 2005. 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* **386**(Pt 2): 297-303.

Kumari, D. and Usdin, K. 2001. 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* **276**(6): 4357-4364.

Kunst, C.B. and Warren, S.T. 1994. Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell* **77**(6): 853-861.

Lachiewicz, A.M., Dawson, D.V., and Spiridigliozzi, G.A. 2000. Physical characteristics of young boys with fragile X syndrome: reasons for difficulties in making a diagnosis in young males. *Am J Med Genet* **92**(4): 229-236.

Ladd, P.D., Smith, L.E., Rabaia, N.A., Moore, J.M., Georges, S.A., Hansen, R.S., Hagerman, R.J., Tassone, F., Tapscott, S.J., and Filippova, G.N. 2007. An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. *Hum Mol Genet* **16**(24): 3174-3187.

Laggerbauer, B., Ostareck, D., Keidel, E.M., Ostareck-Lederer, A., and Fischer, U. 2001. Evidence that fragile X mental retardation protein is a negative regulator of translation. *Hum Mol Genet* **10**(4): 329-338.

Lenhard, W., Breitenbach, E., Ebert, H., Schindelhauer-Deutscher, H.J., and Henn, W. 2005. Psychological benefit of diagnostic certainty for mothers of children with disabilities: lessons from Down syndrome. *Am J Med Genet A* **133A**(2): 170-175.

Leonard, H. and Wen, X. 2002. The epidemiology of mental retardation: challenges and opportunities in the new millennium. *Ment Retard Dev Disabil Res Rev* **8**(3): 117-134.

Levy, S., Sutton, G., Ng, P.C., Feuk, L., Halpern, A.L., Walenz, B.P., Axelrod, N., Huang, J., Kirkness, E.F., Denisov, G. et al. 2007. The diploid genome sequence of an individual human. *PLoS Biol* **5**(10): e254.

Li, H., Ruan, J., and Durbin, R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* **18**(11): 1851-1858.

Li, Z., Zhang, Y., Ku, L., Wilkinson, K.D., Warren, S.T., and Feng, Y. 2001. The fragile X mental retardation protein inhibits translation via interacting with mRNA. *Nucleic Acids Res* **29**(11): 2276-2283.

Lim, J.H., Booker, A.B., Luo, T., Williams, T., Furuta, Y., Lagutin, O., Oliver, G., Sargent, T.D., and Fallon, J.R. 2005. AP-2alpha selectively regulates fragile X mental retardation-1 gene transcription during embryonic development. *Hum Mol Genet* **14**(14): 2027-2034.

Loehr, J.P., Synhorst, D.P., Wolfe, R.R., and Hagerman, R.J. 1986. Aortic root dilatation and mitral valve prolapse in the fragile X syndrome. *Am J Med Genet* **23**(1-2): 189-194.

Loesch, D.Z., Bui, Q.M., Huggins, R.M., Mitchell, R.J., Hagerman, R.J., and Tassone, F. 2007. Transcript levels of the intermediate size or grey zone fragile X mental retardation 1 alleles are raised, and correlate with the number of CGG repeats. *J Med Genet* **44**(3): 200-204.

Lubs, H.A. 1969. A marker X chromosome. *Am J Hum Genet* **21**(3): 231-244.

Luckasson, R. and Reeve, A. 2001. Naming, defining, and classifying in mental retardation. *Ment Retard* **39**(1): 47-52.

Lugenbeel, K.A., Peier, A.M., Carson, N.L., Chudley, A.E., and Nelson, D.L. 1995. Intragenic loss of function mutations demonstrate the primary role of FMR1 in fragile X syndrome. *Nat Genet* **10**(4): 483-485.

Maddalena, A., Richards, C.S., McGinniss, M.J., Brothman, A., Desnick, R.J., Grier, R.E., Hirsch, B., Jacky, P., McDowell, G.A., Popovich, B. et al. 2001. Technical standards and guidelines for fragile X: the first of a series of disease-specific supplements to the Standards and Guidelines for Clinical Genetics Laboratories of the American College of Medical Genetics. Quality Assurance Subcommittee of the Laboratory Practice Committee. *Genet Med* **3**(3): 200-205.

Maes, B., Fryns, J.P., Ghesquiere, P., and Borghgraef, M. 2000. Phenotypic checklist to screen for fragile X syndrome in people with mental retardation. *Ment Retard* **38**(3): 207-215.

Major, T., Culjkovic, B., Stojkovic, O., Gucscekic, M., Lakic, A., and Romac, S. 2003. Prevalence of the fragile X syndrome in Yugoslav patients with non-specific mental retardation. *J Neurogenet* **17**(2-3): 223-230.

Malter, H.E., Iber, J.C., Willemsen, R., de Graaff, E., Tarleton, J.C., Leisti, J., Warren, S.T., and Oostra, B.A. 1997. Characterization of the full fragile X syndrome mutation in fetal gametes. *Nat Genet* **15**(2): 165-169.

Marin-Padilla, M. 1976. Pyramidal cell abnormalities in the motor cortex of a child with Down's syndrome. A Golgi study. *J Comp Neurol* **167**(1): 63-81.

Martin, J.P., Bell, J. 1943. A pedigree of mental defect showing sex-linkage. *J Neurol Psychiatry* **6**: 154-157.

Matijevic, T., Knezevic, J., Slavica, M., and Pavelic, J. 2009. Rett syndrome: from the gene to the disease. *Eur Neurol* **61**(1): 3-10.

McBride, S.M., Choi, C.H., Wang, Y., Liebelt, D., Braunstein, E., Ferreiro, D., Sehgal, A., Siwicki, K.K., Dockendorff, T.C., Nguyen, H.T. et al. 2005. Pharmacological rescue of synaptic plasticity, courtship behavior, and mushroom body defects in a Drosophila model of fragile X syndrome. *Neuron* **45**(5): 753-764.

McConkie-Rosell, A., Abrams, L., Finucane, B., Cronister, A., Gane, L.W., Coffey, S.M., Sherman, S., Nelson, L.M., Berry-Kravis, E., Hessl, D. et al. 2007. Recommendations from multi-disciplinary focus groups on cascade testing and genetic counseling for fragile X-associated disorders. *J Genet Couns* **16**(5): 593-606.

Meechan, D.W., Tucker, E.S., Maynard, T.M., and Lamantia, A.S. 2009. Diminished dosage of 22q11 genes disrupts neurogenesis and cortical development in a mouse model of 22q11 deletion/DiGeorge syndrome. *Proc Natl Acad Sci U S A* **106**(38): 16434-16445.

Meijer, H., de Graaff, E., Merckx, D.M., Jongbloed, R.J., de Die-Smulders, C.E., Engelen, J.J., Fryns, J.P., Curfs, P.M., and Oostra, B.A. 1994. A deletion of 1.6 kb

proximal to the CGG repeat of the FMR1 gene causes the clinical phenotype of the fragile X syndrome. *Hum Mol Genet* **3**(4): 615-620.

Mila, M., Castellvi-Bel, S., Sanchez, A., Barcelo, A., Badenas, C., Mallolas, J., and Estivill, X. 2000. Rare variants in the promoter of the fragile X syndrome gene (FMR1). *Mol Cell Probes* **14**(2): 115-119.

Mila, M., Sanchez, A., Badenas, C., Brun, C., Jimenez, D., Villa, M.P., Castellvi-Bel, S., and Estivill, X. 1997. Screening for FMR1 and FMR2 mutations in 222 individuals from Spanish special schools: identification of a case of FRAXE-associated mental retardation. *Hum Genet* **100**(5-6): 503-507.

Miyashiro, K.Y., Beckel-Mitchener, A., Purk, T.P., Becker, K.G., Barret, T., Liu, L., Carbonetto, S., Weiler, I.J., Greenough, W.T., and Eberwine, J. 2003. RNA cargoes associating with FMRP reveal deficits in cellular functioning in Fmr1 null mice. *Neuron* **37**(3): 417-431.

Mononen, T., von Koskull, H., Airaksinen, R.L., and Juvonen, V. 2007. A novel duplication in the FMR1 gene: implications for molecular analysis in fragile X syndrome and repeat instability. *Clin Genet* **72**(6): 528-531.

Muddashetty, R.S., Kelic, S., Gross, C., Xu, M., and Bassell, G.J. 2007. Dysregulated metabotropic glutamate receptor-dependent translation of AMPA receptor and postsynaptic density-95 mRNAs at synapses in a mouse model of fragile X syndrome. *J Neurosci* **27**(20): 5338-5348.

Muhle, R., Trentacoste, S.V., and Rapin, I. 2004. The genetics of autism. *Pediatrics* **113**(5): e472-486.

Mulroy, S., Robertson, L., Aiberti, K., Leonard, H., and Bower, C. 2008. The impact of having a sibling with an intellectual disability: parental perspectives in two disorders. *J Intellect Disabil Res* **52**(Pt 3): 216-229.

Murphy, M.M. and Mazzocco, M.M. 2008. Rote numeric skills may mask underlying mathematical disabilities in girls with fragile x syndrome. *Dev Neuropsychol* **33**(3): 345-364.

Nakamoto, M., Nalavadi, V., Epstein, M.P., Narayanan, U., Bassell, G.J., and Warren, S.T. 2007. Fragile X mental retardation protein deficiency leads to excessive mGluR5-dependent internalization of AMPA receptors. *Proc Natl Acad Sci U S A* **104**(39): 15537-15542.

Narayanan, U., Nalavadi, V., Nakamoto, M., Pallas, D.C., Ceman, S., Bassell, G.J., and Warren, S.T. 2007. FMRP phosphorylation reveals an immediate-early signaling pathway triggered by group I mGluR and mediated by PP2A. *J Neurosci* **27**(52): 14349-14357.

Narayanan, U., Nalavadi, V., Nakamoto, M., Thomas, G., Ceman, S., Bassell, G.J., and Warren, S.T. 2008. S6K1 phosphorylates and regulates fragile X mental retardation protein (FMRP) with the neuronal protein synthesis-dependent mammalian target of rapamycin (mTOR) signaling cascade. *J Biol Chem* **283**(27): 18478-18482.

Ng, P.C. and Henikoff, S. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**(13): 3812-3814.

Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E. et al. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* **461**(7261): 272-276.

Oberle, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M.F., and Mandel, J.L. 1991. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* **252**(5010): 1097-1102.

Ohashi, S., Koike, K., Omori, A., Ichinose, S., Ohara, S., Kobayashi, S., Sato, T.A., and Anzai, K. 2002. Identification of mRNA/protein (mRNP) complexes containing Puralpha, mStaufen, fragile X protein, and myosin Va and their association with rough endoplasmic reticulum equipped with a kinesin motor. *J Biol Chem* **277**(40): 37804-37810.

Okou, D.T., Locke, A.E., Steinberg, K.M., Hagen, K., Athri, P., Shetty, A.C., Patel, V., and Zwick, M.E. 2009. Combining microarray-based genomic selection (MGS) with the Illumina Genome Analyzer platform to sequence diploid target regions. *Ann Hum Genet* **73**(Pt 5): 502-513.

Okou, D.T., Steinberg, K.M., Middle, C., Cutler, D.J., Albert, T.J., and Zwick, M.E. 2007. Microarray-based genomic selection for high-throughput resequencing. *Nat Methods* **4**(11): 907-909.

Olsson, M.B. and Hwang, C.P. 2001. Depression in mothers and fathers of children with intellectual disability. *J Intellect Disabil Res* **45**(Pt 6): 535-543.

Opitz, J.M., Westphal, J.M., and Daniel, A. 1984. Discovery of a connective tissue dysplasia in the Martin-Bell syndrome. *Am J Med Genet* **17**(1): 101-109.

Ossowski, S., Schneeberger, K., Clark, R.M., Lanz, C., Warthmann, N., and Weigel, D. 2008. Sequencing of natural strains of Arabidopsis thaliana with short reads. *Genome Res* **18**(12): 2024-2033.

Out, A.A., van Minderhout, I.J., Goeman, J.J., Ariyurek, Y., Ossowski, S., Schneeberger, K., Weigel, D., van Galen, M., Taschner, P.E., Tops, C.M. et al. 2009. Deep sequencing to reveal new variants in pooled DNA samples. *Hum Mutat*.

Pandey, U.B., Phadke, S., and Mittal, B. 2002. Molecular screening of FRAXA and FRAXE in Indian patients with unexplained mental retardation. *Genet Test* **6**(4): 335-339.

Park, S., Park, J.M., Kim, S., Kim, J.A., Shepherd, J.D., Smith-Hicks, C.L., Chowdhury, S., Kaufmann, W., Kuhl, D., Ryazanov, A.G. et al. 2008. Elongation factor 2 and fragile X mental retardation protein control the dynamic translation of Arc/Arg3.1 essential for mGluR-LTD. *Neuron* **59**(1): 70-83.

Patsalis, P.C., Sismani, C., Hettinger, J.A., Boumba, I., Georgiou, I., Stylianidou, G., Anastasiadou, V., Koukoulli, R., Pagoulatos, G., and Syrrou, M. 1999. 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* **84**(3): 184-190.

Pieretti, M., Zhang, F.P., Fu, Y.H., Warren, S.T., Oostra, B.A., Caskey, C.T., and Nelson, D.L. 1991. Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* **66**(4): 817-822.

Pietrobono, R., Tabolacci, E., Zalfa, F., Zito, I., Terracciano, A., Moscato, U., Bagni, C., Oostra, B., Chiurazzi, P., and Neri, G. 2005. Molecular dissection of the events leading to inactivation of the FMR1 gene. *Hum Mol Genet* **14**(2): 267-277.

Porter, F.D. 2008. Smith-Lemli-Opitz syndrome: pathogenesis, diagnosis and management. *Eur J Hum Genet* **16**(5): 535-541.

Price, D.K., Zhang, F., Ashley, C.T., Jr., and Warren, S.T. 1996. The chicken FMR1 gene is highly conserved with a CCT 5'-untranslated repeat and encodes an RNA-binding protein. *Genomics* **31**(1): 3-12.

Primerano, B., Tassone, F., Hagerman, R.J., Hagerman, P., Amaldi, F., and Bagni, C. 2002. Reduced FMR1 mRNA translation efficiency in fragile X patients with premutations. *Rna* **8**(12): 1482-1488.

Purpura, D.P. 1974. Dendritic spine "dysgenesis" and mental retardation. *Science* **186**(4169): 1126-1128.

Quan, F., Grompe, M., Jakobs, P., and Popovich, B.W. 1995. Spontaneous deletion in the FMR1 gene in a patient with fragile X syndrome and cherubism. *Hum Mol Genet* **4**(9): 1681-1684.

Rachidi, M. and Lopes, C. 2008. Mental retardation and associated neurological dysfunctions in Down syndrome: a consequence of dysregulation in critical chromosome 21 genes and associated molecular pathways. *Eur J Paediatr Neurol* **12**(3): 168-182.

Rahmani, Z., Blouin, J.L., Creau-Goldberg, N., Watkins, P.C., Mattei, J.F., Poissonnier, M., Prieur, M., Chettouh, Z., Nicole, A., Aurias, A. et al. 1989. Critical role of the D21S55 region on chromosome 21 in the pathogenesis of Down syndrome. *Proc Natl Acad Sci U S A* **86**(15): 5958-5962.

Ramensky, V., Bork, P., and Sunyaev, S. 2002. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res* **30**(17): 3894-3900.

Rauch, A., Hoyer, J., Guth, S., Zweier, C., Kraus, C., Becker, C., Zenker, M., Huffmeier, U., Thiel, C., Ruschendorf, F. et al. 2006. Diagnostic yield of various genetic approaches in patients with unexplained developmental delay or mental retardation. *Am J Med Genet A* **140**(19): 2063-2074.

Reese, M.G., Eeckman, F.H., Kulp, D., and Haussler, D. 1997. Improved splice site detection in Genie. *J Comput Biol* **4**(3): 311-323.

Reschly, D.J. and Jipson, F.J. 1976. Ethnicity, geographic locale, age, sex, and urban-rural residence as variables in the prevalence of mild retardation. *Am J Ment Defic* **81**(2): 154-161.

Reyniers, E., Wolff, G., Tariverdian, G., De Boulle, K., Storm, K., Kooy, R.F., and Willems, P.J. 1996. Severe mental retardation and macroorchidism without mutation in the FMR1 gene. *Am J Med Genet* **64**(2): 408-412.

Rhead, B., Karolchik, D., Kuhn, R.M., Hinrichs, A.S., Zweig, A.S., Fujita, P.A., Diekhans, M., Smith, K.E., Rosenbloom, K.R., Raney, B.J. et al. 2009. The UCSC genome browser database: update 2010. *Nucleic Acids Res*.

Richards, B.W., Sylvester, P.E., and Brooker, C. 1981. Fragile X-linked mental retardation: the Martin-Bell syndrome. *J Ment Defic Res* **25 Pt 4**: 253-256.

Roberts, J.E., Mirrett, P., and Burchinal, M. 2001. Receptive and expressive communication development of young males with fragile X syndrome. *Am J Ment Retard* **106**(3): 216-230.

Roeleveld, N., Zielhuis, G.A., and Gabreels, F. 1997. The prevalence of mental retardation: a critical review of recent literature. *Dev Med Child Neurol* **39**(2): 125-132.

Rogers, S.J., Wehner, D.E., and Hagerman, R. 2001. The behavioral phenotype in fragile X: symptoms of autism in very young children with fragile X syndrome,

idiopathic autism, and other developmental disorders. *J Dev Behav Pediatr* **22**(6): 409-417.

Ropers, H.H. 2008. Genetics of intellectual disability. *Curr Opin Genet Dev* **18**(3): 241-250.

Rosenthal, E.T., Biesecker, L.G., and Biesecker, B.B. 2001. Parental attitudes toward a diagnosis in children with unidentified multiple congenital anomaly syndromes. *Am J Med Genet* **103**(2): 106-114.

Rousseau, F., Heitz, D., Tarleton, J., MacPherson, J., Malmgren, H., Dahl, N., Barnicoat, A., Mathew, C., Mornet, E., Tejada, I. et al. 1994. A multicenter study on genotype-phenotype correlations in the fragile X syndrome, using direct diagnosis with probe StB12.3: the first 2,253 cases. *Am J Hum Genet* **55**(2): 225-237.

Rudelli, R.D., Brown, W.T., Wisniewski, K., Jenkins, E.C., Laure-Kamionowska, M., Connell, F., and Wisniewski, H.M. 1985. Adult fragile X syndrome. Clinico-neuropathologic findings. *Acta Neuropathol* **67**(3-4): 289-295.

Ruvalcaba, R.H., Myhre, S.A., Roosen-Runge, E.C., and Beckwith, J.B. 1977. X-linked mental deficiency megalotestes syndrome. *JAMA* **238**(15): 1646-1650.

Sabaratnam, M., Vroegop, P.G., and Gangadharan, S.K. 2001. Epilepsy and EEG findings in 18 males with fragile X syndrome. *Seizure* **10**(1): 60-63.

Schaeffer, C., Bardoni, B., Mandel, J.L., Ehresmann, B., Ehresmann, C., and Moine, H. 2001. The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J* **20**(17): 4803-4813.

Schalock, R.L., Luckasson, R.A., Shogren, K.A., Borthwick-Duffy, S., Bradley, V., Buntinx, W.H., Coulter, D.L., Craig, E.M., Gomez, S.C., Lachapelle, Y. et al. 2007. The renaming of mental retardation: understanding the change to the term intellectual disability. *Intellect Dev Disabil* **45**(2): 116-124.

Schrier, M., Severijnen, L.A., Reis, S., Rife, M., van't Padje, S., van Cappellen, G., Oostra, B.A., and Willemsen, R. 2004. Transport kinetics of FMRP containing the I304N mutation of severe fragile X syndrome in neurites of living rat PC12 cells. *Exp Neurol* **189**(2): 343-353.

Schwemmle, S., de Graaff, E., Deissler, H., Glaser, D., Wohrle, D., Kennerknecht, I., Just, W., Oostra, B.A., Doerfler, W., Vogel, W. et al. 1997. Characterization of FMR1 promoter elements by in vivo-footprinting analysis. *Am J Hum Genet* **60**(6): 1354-1362.

Shendure, J. and Ji, H. 2008. Next-generation DNA sequencing. *Nat Biotechnol* **26**(10): 1135-1145.

Sherman, S.L. 2000. Premature ovarian failure in the fragile X syndrome. *Am J Med Genet* **97**(3): 189-194.

Sherman, S.L., Jacobs, P.A., Morton, N.E., Froster-Iskenius, U., Howard-Peebles, P.N., Nielsen, K.B., Partington, M.W., Sutherland, G.R., Turner, G., and Watson, M. 1985. Further segregation analysis of the fragile X syndrome with special reference to transmitting males. *Hum Genet* **69**(4): 289-299.

Shinahara, K., Saijo, T., Mori, K., and Kuroda, Y. 2004. Single-strand conformation polymorphism analysis of the FMR1 gene in autistic and mentally retarded children in Japan. *J Med Invest* **51**(1-2): 52-58.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S. et al. 2005. Evolutionarily

conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* **15**(8): 1034-1050.

Siomi, H., Siomi, M.C., Nussbaum, R.L., and Dreyfuss, G. 1993. The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. *Cell* **74**(2): 291-298.

Siomi, M.C., Siomi, H., Sauer, W.H., Srinivasan, S., Nussbaum, R.L., and Dreyfuss, G. 1995. FXR1, an autosomal homolog of the fragile X mental retardation gene. *EMBO J* **14**(11): 2401-2408.

Siomi, M.C., Zhang, Y., Siomi, H., and Dreyfuss, G. 1996. Specific sequences in the fragile X syndrome protein FMR1 and the FXR proteins mediate their binding to 60S ribosomal subunits and the interactions among them. *Mol Cell Biol* **16**(7): 3825-3832.

Sittler, A., Devys, D., Weber, C., and Mandel, J.L. 1996. Alternative splicing of exon 14 determines nuclear or cytoplasmic localisation of fmr1 protein isoforms. *Hum Mol Genet* **5**(1): 95-102.

Siva, N. 2008. 1000 Genomes project. *Nat Biotechnol* **26**(3): 256.

Skinner, M., Hooper, S., Hatton, D.D., Roberts, J., Mirrett, P., Schaaf, J., Sullivan, K., Wheeler, A., and Bailey, D.B., Jr. 2005. Mapping nonverbal IQ in young boys with fragile X syndrome. *Am J Med Genet A* **132A**(1): 25-32.

Slegtenhorst-Eegdeman, K.E., de Rooij, D.G., Verhoef-Post, M., van de Kant, H.J., Bakker, C.E., Oostra, B.A., Grootegoed, J.A., and Themmen, A.P. 1998. Macroorchidism in FMR1 knockout mice is caused by increased Sertoli cell proliferation during testicular development. *Endocrinology* **139**(1): 156-162.

Smale, S.T. and Baltimore, D. 1989. The "initiator" as a transcription control element. *Cell* **57**(1): 103-113.

Smith, K.T., Coffee, B., and Reines, D. 2004. Occupancy and synergistic activation of the FMR1 promoter by Nrf-1 and Sp1 in vivo. *Hum Mol Genet* **13**(15): 1611-1621.

Smith, K.T., Nicholls, R.D., and Reines, D. 2006. The gene encoding the fragile X RNA-binding protein is controlled by nuclear respiratory factor 2 and the CREB family of transcription factors. *Nucleic Acids Res* **34**(4): 1205-1215.

Spencer, C.M., Alekseyenko, O., Serysheva, E., Yuva-Paylor, L.A., and Paylor, R. 2005. Altered anxiety-related and social behaviors in the Fmr1 knockout mouse model of fragile X syndrome. *Genes Brain Behav* **4**(7): 420-430.

Sullivan, K., Hatton, D., Hammer, J., Sideris, J., Hooper, S., Ornstein, P., and Bailey, D., Jr. 2006. ADHD symptoms in children with FXS. *Am J Med Genet A* **140**(21): 2275-2288.

Sutcliffe, J.S., Nelson, D.L., Zhang, F., Pieretti, M., Caskey, C.T., Saxe, D., and Warren, S.T. 1992. DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum Mol Genet* **1**(6): 397-400.

Tamanini, F., Meijer, N., Verheij, C., Willems, P.J., Galjaard, H., Oostra, B.A., and Hoogeveen, A.T. 1996. FMRP is associated to the ribosomes via RNA. *Hum Mol Genet* **5**(6): 809-813.

Tarleton, J., Kenneson, A., Taylor, A.K., Crandall, K., Fletcher, R., Casey, R., Hart, P.S., Hatton, D., Fisch, G., and Warren, S.T. 2002. A single base alteration in the CGG

repeat region of FMR1: possible effects on gene expression and phenotype. *J Med Genet* **39**(3): 196-200.

Tarleton, J., Richie, R., Schwartz, C., Rao, K., Aylsworth, A.S., and Lachiewicz, A. 1993. An extensive de novo deletion removing FMR1 in a patient with mental retardation and the fragile X syndrome phenotype. *Hum Mol Genet* **2**(11): 1973-1974.

Tarpey, P.S., Smith, R., Pleasance, E., Whibley, A., Edkins, S., Hardy, C., O'Meara, S., Latimer, C., Dicks, E., Menzies, A. et al. 2009. A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat Genet* **41**(5): 535-543.

Tassone, F., Hagerman, R.J., Taylor, A.K., Gane, L.W., Godfrey, T.E., and Hagerman, P.J. 2000. Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome. *Am J Hum Genet* **66**(1): 6-15.

Thomas, P.D. and Kejariwal, A. 2004. Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. *Proc Natl Acad Sci U S A* **101**(43): 15398-15403.

Todd, P.K., Mack, K.J., and Malter, J.S. 2003. The fragile X mental retardation protein is required for type-I metabotropic glutamate receptor-dependent translation of PSD-95. *Proc Natl Acad Sci U S A* **100**(24): 14374-14378.

Tucker, B., Richards, R.I., and Lardelli, M. 2006. Contribution of mGluR and Fmr1 functional pathways to neurite morphogenesis, craniofacial development and fragile X syndrome. *Hum Mol Genet* **15**(23): 3446-3458.

Turner, G., Daniel, A., and Frost, M. 1980. X-linked mental retardation, macro-orchidism, and the Xq27 fragile site. *J Pediatr* **96**(5): 837-841.

Turner, G., Eastman, C., Casey, J., McLeay, A., Procopis, P., and Turner, B. 1975. X-linked mental retardation associated with macro-orchidism. *J Med Genet* **12**(4): 367-371.

Verheij, C., Bakker, C.E., de Graaff, E., Keulemans, J., Willemsen, R., Verkerk, A.J., Galjaard, H., Reuser, A.J., Hoogeveen, A.T., and Oostra, B.A. 1993. Characterization and localization of the FMR-1 gene product associated with fragile X syndrome. *Nature* **363**(6431): 722-724.

Verkerk, A.J., de Graaff, E., De Boulle, K., Eichler, E.E., Konecki, D.S., Reyniers, E., Manca, A., Poustka, A., Willems, P.J., Nelson, D.L. et al. 1993. Alternative splicing in the fragile X gene FMR1. *Hum Mol Genet* **2**(8): 1348.

Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F.P. et al. 1991. Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**(5): 905-914.

Vincent, J.B. and Gurling, H.M. 1998. Point mutation in intron 10 of FMR1 is unlikely to be a cause of fragile X syndrome. *Hum Mutat* **12**(6): 431-432.

Vincent, J.B., Konecki, D.S., Munstermann, E., Bolton, P., Poustka, A., Poustka, F., and Gurling, H.M. 1996. Point mutation analysis of the FMR-1 gene in autism. *Molecular psychiatry* **1**(3): 227-231.

Wan, L., Dockendorff, T.C., Jongens, T.A., and Dreyfuss, G. 2000. Characterization of dFMR1, a Drosophila melanogaster homolog of the fragile X mental retardation protein. *Mol Cell Biol* **20**(22): 8536-8547.

Wang, H., Dictenberg, J.B., Ku, L., Li, W., Bassell, G.J., and Feng, Y. 2008a. Dynamic association of the fragile X mental retardation protein as a messenger ribonucleoprotein between microtubules and polyribosomes. *Mol Biol Cell* **19**(1): 105-114.

Wang, H., Wu, L.J., Kim, S.S., Lee, F.J., Gong, B., Toyoda, H., Ren, M., Shang, Y.Z., Xu, H., Liu, F. et al. 2008b. FMRP acts as a key messenger for dopamine modulation in the forebrain. *Neuron* **59**(4): 634-647.

Wang, J., Wang, W., Li, R., Li, Y., Tian, G., Goodman, L., Fan, W., Zhang, J., Li, J., Guo, Y. et al. 2008c. The diploid genome sequence of an Asian individual. *Nature* **456**(7218): 60-65.

Wang, Y.C., Lin, M.L., Lin, S.J., Li, Y.C., and Li, S.Y. 1997. Novel point mutation within intron 10 of FMR-1 gene causing fragile X syndrome. *Hum Mutat* **10**(5): 393-399.

Waung, M.W., Pfeiffer, B.E., Nosyreva, E.D., Ronesi, J.A., and Huber, K.M. 2008. Rapid translation of Arc/Arg3.1 selectively mediates mGluR-dependent LTD through persistent increases in AMPAR endocytosis rate. *Neuron* **59**(1): 84-97.

Westmark, C.J. and Malter, J.S. 2007. FMRP mediates mGluR5-dependent translation of amyloid precursor protein. *PLoS Biol* **5**(3): e52.

Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T. et al. 2008. The complete genome of an individual by massively parallel DNA sequencing. *Nature* **452**(7189): 872-876.

Wisniewski, K.E., Segan, S.M., Miezejeski, C.M., Sersen, E.A., and Rudelli, R.D. 1991. The Fra(X) syndrome: neurological, electrophysiological, and neuropathological abnormalities. *Am J Med Genet* **38**(2-3): 476-480.

Wohrle, D., Kotzot, D., Hirst, M.C., Manca, A., Korn, B., Schmidt, A., Barbi, G., Rott, H.D., Poustka, A., Davies, K.E. et al. 1992. A microdeletion of less than 250 kb, including the proximal part of the FMR-I gene and the fragile-X site, in a male with the clinical phenotype of fragile-X syndrome. *Am J Hum Genet* **51**(2): 299-306.

Wolff, P.H., Gardner, J., Lappen, J., Paccia, J., and Meryash, D. 1988. Variable expression of the fragile X syndrome in heterozygous females of normal intelligence. *Am J Med Genet* **30**(1-2): 213-225.

Xu, W., Schluter, O.M., Steiner, P., Czervionke, B.L., Sabatini, B., and Malenka, R.C. 2008. Molecular dissociation of the role of PSD-95 in regulating synaptic strength and LTD. *Neuron* **57**(2): 248-262.

Yan, Q.J., Rammal, M., Tranfaglia, M., and Bauchwitz, R.P. 2005. Suppression of two major Fragile X Syndrome mouse model phenotypes by the mGluR5 antagonist MPEP. *Neuropharmacology* **49**(7): 1053-1066.

Yeargin-Allsopp, M., Murphy, C.C., Oakley, G.P., and Sikes, R.K. 1992. A multiple-source method for studying the prevalence of developmental disabilities in children: the Metropolitan Atlanta Developmental Disabilities Study. *Pediatrics* **89**(4 Pt 1): 624-630.

Yu, S., Pritchard, M., Kremer, E., Lynch, M., Nancarrow, J., Baker, E., Holman, K., Mulley, J., Warren, S., Schlessinger, D. et al. 1991. Fragile X genotype characterized by an unstable region of DNA. *Science* **252**(5009): 1179-1181.

Zalfa, F., Eleuteri, B., Dickson, K.S., Mercaldo, V., De Rubeis, S., di Penta, A., Tabolacci, E., Chiurazzi, P., Neri, G., Grant, S.G. et al. 2007. A new function for the fragile X mental retardation protein in regulation of PSD-95 mRNA stability. *Nat Neurosci* **10**(5): 578-587.

Zalfa, F., Giorgi, M., Primerano, B., Moro, A., Di Penta, A., Reis, S., Oostra, B., and Bagni, C. 2003. The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses. *Cell* **112**(3): 317-327.

Zhang, Y., O'Connor, J.P., Siomi, M.C., Srinivasan, S., Dutra, A., Nussbaum, R.L., and Dreyfuss, G. 1995. The fragile X mental retardation syndrome protein interacts with novel homologs FXR1 and FXR2. *EMBO J* **14**(21): 5358-5366.

Zigler, E. 1987. The definition and classification of mental retardation. *Ups J Med Sci Suppl* **44**: 9-18.