Methylation of novel markers of fragile X alleles is inversely correlated with FMRP expression and FMR1 activation ratio

David Eugeny Godler1,*, Flora Tassone7, Danuta Zuzanna Loesch1,9, Annette Kimball Taylor10, Freya Gehling1, Randi Jenssen Hagerman8, Trent Burgess2, Devika Ganesamoorthy2, Debbie Hennerich10, Lavinia Gordon3, Andrew Evans4,5, K.H. Choo1,6 and Howard Robert Slater2,6

1Chromosome and Chromatin Research Laboratory, The Murdoch Childrens Research Institute, Royal Children's Hospital, 2VCGS Cytogenetics Laboratory, The Murdoch Childrens Research Institute, Royal Children's Hospital, 3Department of Pediatrics, The Bioinformatics Unit, The Murdoch Childrens Research Institute, Royal Children's Hospital, 4Department of Neurology, Royal Melbourne Hospital, 5Department of Medicine and 6Department of Paediatrics, University of Melbourne, Melbourne, Victoria 3052, Australia, 7Department of Biochemistry and Molecular Medicine, School of Medicine and 8The MIND Institute and Department of Pediatrics, University of California at Davis Medical Center, Sacramento, CA 95817, USA, 9School of Psychological Science, La Trobe University, Melbourne, Victoria 3084, Australia and 10Kimball Genetics, Inc., Denver, CO 80202, USA

Received October 25, 2009; Revised and Accepted January 25, 2010

The fragile X syndrome (FXS) is caused by silencing of the fragile X mental retardation gene (FMR1) and the absence of its product, fragile X mental retardation protein (FMRP), resulting from CpG island methylation associated with large CGG repeat expansions (more than 200) termed full mutation (FM). We have identified a number of novel epigenetic markers for FXS using matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS), naming the most informative fragile X-related epigenetic element 1 (FREE1) and 2 (FREE2). Methylation of both regions was correlated with that of the FMR1 CpG island detected using Southern blot (FREE1 \( R = 0.97; P < 0.00001, n = 23 \) and FREE2 \( R = 0.93; P < 0.00001, n = 23 \)) and negatively correlated with lymphocyte expression of FMRP (FREE1 \( R = -0.62; P = 0.01, n = 15 \) and FREE2 \( R = -0.55; P = 0.03, n = 15 \)) in blood of partially methylated ‘high functioning’ FM males. In blood of FM carrier females, methylation of both markers was inversely correlated with the FMR1 activation ratio (FREE1 \( R = -0.93; P < 0.0001, n = 12 \) and FREE2 \( R = -0.95; P < 0.0001, n = 9 \)). In a sample set of 49 controls, 18 grey zone (GZ 40–54 repeats), 22 premutation (PM 55–170 repeats) and 22 (affected) FXS subjects, the FREE1 methylation pattern was consistent between blood and chorionic villi as a marker of methylated FM alleles and could be used to differentiate FXS males and females from controls, as well as from carriers of GZ/PM alleles, but not between GZ and PM alleles and controls. Considering its high-throughput and specificity for pathogenic FM alleles, low cost and minimal DNA requirements, FREE MALDI-TOF MS offers a unique tool in FXS diagnostics and newborn population screening.

INTRODUCTION

The fragile X mental retardation-1 (FMR1) gene encodes fragile X mental retardation protein (FMRP), essential for normal neuronal functioning (1–3). The CGG expansion within the 5'-untranslated region of the gene has been implicated in a number of pathologies including the fragile X syndrome (FXS) (4), fragile X-associated tremor/ataxia syndrome (FXTAS) (5–8) and fragile X-associated premature ovarian insufficiency (POI) (9,10). FXS, the focus of this study, is associated with a fragile site at the Xq27.3 locus,
with approximate frequencies of one in 4000 males and one in 8000 females (4). Apart from cognitive and behavioural deficits, FXS individuals usually manifest physical malformations (11). FXS is caused by FMR1 promoter methylation within the CpG island, which occurs as a result of CGG triplet repeat expansion beyond 200 in number, termed full mutation (FM). Of the male FM carriers, 10% have the CpG island that is partially methylated (11,12), and in ~13% of the FM carriers, the intelligence quotient (IQ) is greater than 70 and are thus termed ‘high functioning’ (12). Of the high function- ing males, ~5% have completely unmethylated FM alleles and produce levels close to that of normal FMRP, with FXS phenotype much less severe than that of methylated FM subjects (12). The more common intermediate and small expansion alleles are termed premutations (PM, 55–199 repeats) (13) and grey zone (GZ, 41–54 repeats), respectively (14,15). The PM alleles have been reported to have an unmethylated CpG island within the FMR1 promoter (16,17), whereas the methylation status of the CpG island on the GZ alleles has not yet been assessed. Up to 15% of individu- als who have an FM are also mosaics for the intermediate expansion alleles and have milder FXS phenotype than the non-mosaic FM carriers (17). As for PM alleles, increased levels of FMR1 mRNA have been reported in the GZ individual, proportional to the size of CGG expansion (18,19). The elevated FMR1 mRNA level has been associated with neurodegeneration (5,20) as well as with increased risk for POI (9,10).

Together, the unmethylated/partially methylated FM carri- ers and the mosaic individuals comprise up to 25% of all FM carriers (12,17). Furthermore, 22–50% of all FM carrier females have IQ scores less than 70, with moderate-to-severe mental retardation (21,22), and these low IQ levels have been related to X-inactivation as determined by Southern blot analysis (21). Such a high proportion of male and female FM carriers with variable phenotype, unrelated to the size of expansion in the FM range at least in females (21), stresses the importance of methylation analysis in FXS testing. Current FXS diagnostic tests involve a combination of poly- merase chain reaction (PCR)-based techniques and Southern blot analysis (23,24). Several PCR-based approaches have been developed recently to reliably amplify PM and FM alleles up to a certain size (23). Use of CGG-based tests, including Southern blot, results in the detection of highly prevalent GZ and PM alleles. This may inadvertently turn a screen for FXS into a predictive assay for a late onset disorder, raising a number of challenging issues associated with the ethics and high counselling costs related to follow-up of GZ and PM carriers.

Here, we present a novel approach that may address some of these issues by using a high-throughput method developed by Sequenom EpiTYPER system (Sequenom Inc., San Diego, CA, USA) (25,26), which is specific for FXS-affected individu- als, but does not differentiate between controls, GZ and PM carriers. The DNA methylation analysis uses bisulphite-PCR amplification, in vitro transcription and base-specific cleavage of amplified products. We successfully applied this high-throughput technology to quantitatively assess methylation differences between normal and fragile X samples of the FMR1 promoter and the adjacent regions spanning 97 CpG sites and have identified novel markers of FXS located 5’ and 3’ of the previously described CpG island (27). We named the most informative regions the fragile X-related epigenetic element 1 (FREE1) and 2 (FREE2). We tested the methylation status of these sites in male and female samples that included control subjects (carrying FMR1 alleles within the normal range—HC alleles), high functioning FM carriers including mosiacs, GZ and PM allele carriers (42–170 repeats) and FM allele carriers clinically manifesting FXS. We demonstrated that the quantitative methylation results of both markers allow accurate determination of FMRP expression and FMR1 activation ratio and can be used to differentiate FXS males and females from their control counterparts as well as from carriers of GZ and high PM alleles.

RESULTS

Characterization of novel epigenetic elements

MALDI-TOF MS application for the FMR1 locus methylation mapping. The technology that we have employed for methyl- ation mapping uses base-specific cleavage followed by matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS), in which the size ratio of the cleaved products provides quantitative methylation esti- mates for CpG sites within a target region. The shift in mass for non-methylated (NM) from methylated (M) fragments for a single CpG site is ~16 Da due to the presence of an adenosine residue in the place of a guanosine (Fig. 1A). The EpiTYPER software calculates methylation for each fragment based on this difference in mass, in which the output methyl- ation ratios are the intensities of methylated signal/ unmethylated signal) (see CpG2, Fig. 1B). If the fragment sizes overlap for different CpGs, their methyl- ation output ratio is calculated on the basis of the sum of intensities for methylated/(methylated + unmethylated signal) (see CpGs 3–5 and 11, Fig. 1B). If the mean methylation value is an average: (i) of less than two sites, we presented the site denominations separated by a forward slash (e.g. CpGs 5 and 6 presented as CpGs 5/6) and (ii) for more than two sites, we presented the site denominations separated by a hyphen (e.g. CpGs 23,24,25,26 presented as CpGs 23–26).

To distinguish how well the methylation output ratio for multiple fragments of a similar size represented methylation of separate CpG sites for some amplicons, we performed both C and T cleave reactions (that produced fragments of different sizes) prior to fragment analysis. We represented the results from the C cleave reactions by underlining the relevant CpG sites. The sites that were not underlined represent the T cleave reactions. Silent peaks (S)—fragments of unknown origin—were not taken into consideration if their size did not overlap with the fragments of interest (Fig. 1B). Methyl- ation output of CpG sites that had silent peaks that overlapped with the fragments of interest was included in the analysis (see CpG21), and this was indicated by an S placed next to the CpG site number (Fig. 2; FREE1 CpG 9S and CpG 10S).

Methylation mapping of the FMR1 promoter and the adjacent regions using MALDI-TOF MS. We have defined the
boundaries for the FMR1 promoter (Fig. 2) based on Kumari and Usdin (27,28). The promoter begins 355 bp 5' of the (CGG)n region and ends 60 bp 3' of the (CGG)n region. The promoter encompasses a CpG island, located 5' of the CGG expansion (27,28), and includes a large number of regulatory motifs (Table 1). We have mapped methylation patterns of the promoter, as well as the 5' and 3' adjacent regions, initially in DNA from four controls and three cell lines from FXS patients, where FMR1 and FMRP expressions were absent in the FM cell lines but not in the control samples (Supplementary Material, Fig. S1). Five target regions were chosen for MALDI-TOF methylation analysis of the Xq27.3 locus (amplicons 1–5). Sequences for these regions are presented in Figure 2.

Regions identified as biologically significant showed consistent differences in methylation between controls and FXS samples. The main FMR1 transcription start site and the alternative Inr-like start site I were found between CpGs 11 and 12, Inr-like start site II between CpGs 7 and 8 and Inr-like start site 3 between CpGs 5 and 7 of amplicon 2 (Fig. 2). One of the two main FMR1 regulatory transcription factor binding sites previously reported to be functionally influenced by methylation—USF1/2 binding site footprint I (28,29)—was found at CpG unit 6 of amplicon 2. As expected, our analysis of HC and FXS DNA demonstrated that these CpG units (5–12) surrounding the four FMR1 transcription start sites and the USF1/2 binding site footprint I were fully methylated in FXS, but not in HC DNA (Fig. 3). There was also a weak methylation signal detected for CpGs 8–11 of amplicon 2 in HC cell lines.

From the analysis of the proximal amplicons, we have identified two novel regions adjacent to the classical FMR1 CpG
island. Amplicons 1 (FREE1) and 5 (FREE2) were consistently hypermethylated in FXS samples, but unmethylated or partially methylated in HC cell lines. These regions have not been previously shown to have any biological significance in FXS or any other FMR1/CGG expansion-related disorders. There are also no previously described functional regulatory motifs for these regions. Thus, we initially examined the sequences for putative transcription factor binding sites with TFSEARCH (30) (http://www.cbrc.jp/research/db/TFSEARCH.html), using vertebrate matrices and a standard threshold of 0.85. Subsequently, a number of shared prominent putative transcription factor binding sites were identified at both locations (Table 1).

For the FREE1 region (detected by amplicon 1), CpG 1 is located 9 bp upstream of the NruI site and corresponds to the CpG 1 site of the previously described CpG island (27).

We have found that except for this CpG 1, all FREE1 CpG sites reflected methylation of the classical FMR1 island described (27) in FXS and controls. CpG sites 2–6 of amplicon 1 were consistently fully methylated for FXS DNA and unmethylated for control DNA (Fig. 3A and B). For FXS DNA, CpGs 8–10 were also fully methylated, but for controls, they were partially methylated (Fig. 3A and B). Furthermore, between CpGs 1–6 and CpGs 8–10 regions, we have localized a (CAAAC)n repetitive element that had more than 11 putative binding sites with 87% homology to the SRY binding motif.

For the FREE2 region covered by amplicon 5 and located 3' of the CGG expansion, we have found that all of the CpG sites reflected methylation of the classical CpG island in FXS and controls. Specifically, we observed complete methylation of CpGs 1, 2, 10–12 and partial-to-full methylation of CpG units 3–9 in all FM cell lines from FXS patients; however, in the control samples, all of the CpG units within the amplicon were largely unmethylated. It should be noted that the ATG start site for FMRP translation is located in close proximity to CpG 1 of FREE2. CpGs 3–12 of FREE2
Table 1. Prominent regulatory motif locations and methylation-sensitive restriction sites within the classical FMR1 CpG island (amplicon 2), FREE1 (amplicon 1) and FREE2 (amplicon 5) regions

<table>
<thead>
<tr>
<th>Transcription factor sites/potential regulatory motifs</th>
<th>Sequence</th>
<th>Amplicon</th>
<th>CpG unit location</th>
</tr>
</thead>
<tbody>
<tr>
<td>GATA1/GATA2-SRY-Ik2-c-Ets</td>
<td>ACTGGGATAACCGG ATGCATTTG ATTTCCCACGCCACTG</td>
<td>1</td>
<td>CpG 3 and CpG2</td>
</tr>
<tr>
<td>GATA2</td>
<td>ATCCCAAGA</td>
<td>1</td>
<td>Between CpG 4 and CpG 5/6</td>
</tr>
<tr>
<td>Putative SRY binding sites (CAAAAC)n</td>
<td>CGGGACACAAA CCAACAAAAA CCAACAAAAA CCAACACC</td>
<td>1</td>
<td>Between CpG 5/6 and CpG 8</td>
</tr>
<tr>
<td>Footprint IV-α-PAL/NRF1 binding site</td>
<td>CGCGCATGCGC</td>
<td>Between 1 and 2</td>
<td>N/A</td>
</tr>
<tr>
<td>NruI</td>
<td>TCGCGA</td>
<td>Between 1 and 2</td>
<td>N/A—detected using Southern blot analysis</td>
</tr>
<tr>
<td>Eagl</td>
<td>CCGCCG</td>
<td>Between 1 and 2</td>
<td>N/A</td>
</tr>
<tr>
<td>Footprint III—Sp1 binding site</td>
<td>GAGGGCG</td>
<td>2</td>
<td>CpG 1</td>
</tr>
<tr>
<td>Footprint II—AP2 H4tf1/Sp1 binding site</td>
<td>CGCGGGGGGA</td>
<td>2</td>
<td>CpG 3/4</td>
</tr>
<tr>
<td>Footprint I—USF1/2 binding site</td>
<td>TCACG</td>
<td>2</td>
<td>CpG 6</td>
</tr>
<tr>
<td>Zeste site</td>
<td>CGCTCA</td>
<td>2</td>
<td>CpG 12</td>
</tr>
<tr>
<td>GATA1/GATA2</td>
<td>GAAGATGGAG 5 5</td>
<td>5</td>
<td>CpG 3</td>
</tr>
<tr>
<td>GATA1/GATA2/GATA3</td>
<td>GCCCCATCTTCG</td>
<td>5</td>
<td>CpG 12</td>
</tr>
</tbody>
</table>

are located within intron 1 of the FMR1 gene and encompass seven putative GATA binding sites (Figs 1 and 2 and Table 1). In contrast to FREE1 and FREE2, other regions examined, covered by amplicons 3 and 4, did not reflect the methylation pattern of the classical CpG island in FXS patients and controls (Fig. 3). These regions thus served as appropriate controls for FREE1 and FREE2 methylation mapping using MALDI-TOF MS analysis in FXS and control samples, while assisting in re-defining the 5′ border of the methylated locus in DNA from FXS patients. In addition, portions of these upstream regions (particularly the amplicon 3 CpGs 6 and 7 and amplicon 4 CpGs 7–10) were largely hypomethylated in FXS samples, but hypermethylated in half of the controls. This may indicate that in FXS, the methylation shifts from these upstream regions to FREE1, and this may have implications for transcription from the FMR1 locus.

Sequence analysis of FREE1, FREE2 and the classical FMR1 CpG island. We performed sequence analysis to determine whether the FREE1 and FREE2 regions can be classified as extensions of the previously described CpG island (27). To date, the bioinformatics criteria most widely used to classify CpG islands are either DNA regions >200 bp long with a GC content of >50% and an observed/expected CpG ratio of >0.6 (31) or DNA regions >500 bp long with a GC content of >55% and an observed/expected CpG ratio of >0.65 (32). By comparing these parameters (Supplementary Material, Table S1), we have found that the FREE1 and FREE2 regions had >50% GC content, length >200 bp but <500 bp. This indicated that neither FREE1 nor FREE2 could be considered as a CpG island using the criteria of Takai and Jones (32). However, when the FREE1 sequence was combined with the classical CpG island, the total length exceeded 500 bp, and thus in this context, the FREE1 region could be classified as an extension of the previously described FMR1 CpG island (27) and does fall within the CpG island criteria of both Takai and Jones (32) and Gardiner-Garden and Frommer (31). In contrast, the FREE2 region must be considered as a separate epigenetic entity, as it is separated from the classical CpG island by the CGG expansion, which may be up to several kilobases long in FM subjects. Its observed/expected CpG site content is also markedly lower than the cut-offs used for both Takai and Jones (32) and Gardiner-Garden and Frommer (31) criteria, 0.47 (Supplementary Material, Table S1).

Clinical applications of FREE1 and FREE2 methylation analyses

Methylation status of FREE1 and FREE2 in high functioning males and FM carrier females. To determine which region is the most informative in biological settings, we examined the methylation of FREE1 and FREE2 in blood samples from 21 high functioning males including mosaic individuals and compared the results with those from Southern blot methylation of the classical CpG island and with FMRP expression. The relationship between FMRP-positive cell counts in blood smears and Southern blot methylation analyses, FMR1 mRNA levels and standardized IQ for some of these patients was previously described by Tassone et al. (33). Here we used complete data set for 15 individuals for the analysis of the relationship between FMRP levels and methylation. For correlation between FREE1 and FREE2 methylation and Southern blot analysis, we have included two additional blood samples from FXS individuals with low IQ (approximately 47) and no FMRP expression, making the total number of samples analysed to 23.

We found that mean methylation across FREE1 (CpG units 2–10) was most closely correlated with the number of FMRP-positive lymphocytes of all regions examined (Fig. 4A) (mean methylation across FREE1 $R = -0.62$, $P = 0.01$; FREE2 $R = -0.55$, $P = 0.03$ and Southern blot methylation $R = -0.59$, $P = 0.01$). Southern blot methylation was positively correlated with both FREE1 ($R = 0.97$;
It is also of interest that in three FM carriers with high standardized IQ (65–76) and mild-to-moderate deficits in FMRP, the methylation of both FREE1 and FREE2 was not related to elevated FMR1 mRNA levels (Fig. 4A). This indicates that mean FREE1 and FREE2 methylation analyses do not reflect RNA toxicity associated with the expanded alleles, while being specific for deficits in FMRP and thus FXS-related phenotype. We also found that, in rare cases of PM alleles that have marginally (10–30%) methylated classical CpG island, both FREE1 and FREE2 were methylated at a low level.

To determine which region is the most informative in female samples, we performed the FREE1 and FREE2 analyses in 12 FM allele carriers with variable FMR1 activation ratios and 11 controls. We found that the FMR1 activation ratio determined by Southern blot was inversely correlated with the methylation status of both FREE1 ($R = -0.93; P < 0.00001$) and FREE2 ($R = -0.95; P < 0.00001$) (Fig. 4B). As an internal control, we measured $N_{ru/l}$ methylation using Southern blot in nine control females that showed $0.39 \pm 0.045$ (mean $\pm$ SD) X-inactivation. This was consistent with the FREE1 analysis ($0.43 \pm 0.04$), whereas the FREE2 analysis showed much lower X-inactivation values ($0.25 \pm 0.025$) in female controls (Fig. 4B).

FREE1 methylation analysis in male and female clinical samples with the full range of CGG expansions. We performed methylation analysis of the FREE1 region (as it had the largest number of informative sites, Supplementary Material, Table S3) for the larger sample set composed of DNA from 49 controls, 18 GZ, 22 PM carriers and 22 (clinically affected) FXS subjects in males and females. The FREE1 region methylation was examined in duplicate bisulphite reactions per sample, each amplified with a single PCR reaction. Methylation output for each sample was expressed as a mean of two technical replicates, provided that the duplicate measurements were within 35% of their mean. Based on
We found that in male blood and chorionic villus sampling (CVS), FREE1 CpG units 2–10 were consistently hypermethylated in FXS subjects (~70–100%) (Fig. 5A and C),
but unmethylated in controls, GZ and PM allele carriers (Fig. 5A). Hypomethylation in control CVS was mirrored in a parallel amniocyte sample (Fig. 5C). In FM alleles of females with clinical FXS, FREE1 CpG units 2–9 were also consistently hypermethylated (methylation output ratios of approximately 0.8–1), whereas in control females with normal size alleles, the methylation output ratio approached 0.5 in blood (Fig. 5B). In female blood, there were also no significant differences in FREE1 CpG units 2–10 methylation between GZ and PM groups (55 to approximately 130 repeats) with uncharacterized phenotype (Fig. 5B). However, there was a significant increase in methylation of CpG units 1, 2, 3, 5/6 and 9 in females with GZ/PM allele carriers compared with controls likely to be due to skewed X-inactivation (Fig. 5B).

**DISCUSSION**

Standard laboratory testing for FXS falls into the following categories. The first category applies to diagnostic testing of individuals with intellectual disability, in which a positive result leads to testing of other family members, in whom the risk of carrier status is high. The second category applies to pre-natal testing in known carrier pregnancies. The third category is the population screening of newborns. Earlier
treatment intervention, identification of probands pointing to high-risk relatives and provision of reproductive counselling are strong arguments in favour of newborn screening. One major impediment is the limited suitability of current test methods, which are a combination sizing of small and large CGG repeat expansions by PCR and Southern blot testing, respectively (23). The PCR-based assays only target CGG size and do not provide information on the state of the FMR1 promoter methylation (24). This shortcoming necessitates further testing using methylation-sensitive Southern blot. However, the use of Southern blot for any type of large-scale testing is primarily restricted by low throughput, cost and DNA quality and quantity limitations. Importantly, PCR testing detects GZ and PM carriers who are highly prevalent in the general population (for males one in approximately 30 for GZ and one in approximately 700 for PM carriers and for females one in approximately 15 for GZ and 1 in approximately 250 for PM carriers) (34,35). These small expansions do not cause FXS, but PM carriers have a high risk of transmitting them in an expanded form. Furthermore, both PM and GZ alleles have been associated with elevated FMR1 mRNA (18,19) and related to increased risks of developing FXPOI, and PM alleles have also been linked to FXTAS (5,9,10,15,20,34,36). Thus, a test that detects GZ and PM alleles may inadvertently turn a screen for FXS into a predictive assay for a late onset disorder raising unavoidable ethical issues. Furthermore, the counselling costs generated by finding these alleles are prohibitive and of questionable value. There has, therefore, been no ethically acceptable, simple, accurate and inexpensive test available for FXS population screening.

These problems might now be overcome by a novel MALDI-TOF MS test for methylated markers of FXS described in this study. The method is advantageous over most other MS-PCR-based assays (37,38) and enzyme-based methylation-sensitive multiplex ligation-dependent probe amplification (MS-MLPA) methods (39) developed to examine the methylation of the classical FMR1 CpG island, as MALDI-TOF MS can be used to rapidly examine large stretches of DNA for methylation. In contrast, most PCR or MS-MLPA and enzyme-based MS-MLPA methods are restricted to a few sites that are less biologically significant and/or more heavily affected by skewed X-inactivation (37–39). MALDI-TOF MS described in this study requires ~100-fold less DNA quantity than Southern blot and ~10-fold less DNA quantity than the MS-PCR established for FXS testing in males (38). It is less sensitive to DNA quality issues than Southern blot (as it was used to identify FM/FSX samples that failed when analysed via Southern blot; unpublished data), is far more rapid (<2 days versus 1–2 weeks for Southern) and is less expensive. Furthermore, MALDI-TOF MS has been used previously to assess methylation from both freshly prepared and archival dried blood spots (40).

Using this methodology, we identified two novel regions that were hypermethylated in FXS individuals, but unmethylated in carriers of smaller expansions. We have named the region 5′ of the promoter FREE1 and the region 3′ of the CGG expansion, largely located within intron 1 of the FMR1, FREE2. The two regions were of similar size and CG content, with the CpG unit density approximately one-third of that found in the 52 unit classical CpG island described previously (27). Based on the GC content, the observed/expected CpG ratio and the length of the fragment, we propose that the FREE1 region should be considered as an extension of the classical CpG island (27). In contrast, we could not classify FREE2 as a part of the island, as the regions are separated by the CGG expansion. When considered separately, the FREE2 observed/expected CpG ratio is significantly lower than that used for CpG island classification (32). However, due to its orthologous conservation (Supplementary Material, Table S2), it can be considered as a CG cluster (41). The presence of a number of conserved putative GATA binding sites within this region further suggests that FREE2 may be involved in transcriptional regulation (42–44).

In contrast to FREE2, the human FREE1 sequence is only homologous to that of the common chimp. This lack of conservation may explain why FREE1 was not previously considered functionally important in the regulation of FMR1 transcriptional activity (28) and was not included as part of the FMR1 CpG island (27). However, lack of sequence conservation does not exclude functional importance for long non-coding RNA species, particularly for those acting in cis (e.g. XIST) (45) and in trans (e.g. HOTAIR) (46). Since the FREE1 region falls at the 5′ end of FMR4—a long non-coding RNA (47)—it may also have functional significance despite the absence of interspecies homology. This is supported by our sequence analysis of FREE1, in which we have localized a number of putative transcription factor binding sites, suggesting the potential importance of this region in transcriptional regulation. Of particular interest is the (CAAAAC)n repetitive element that encompasses more than 11 putative SRY binding sites and the GATA1/2-SRY-Ik2-c-Ets transcription factor binding region in the 5′ portion of FREE1.

Since we found that the mean methylation of the FREE1 and FREE2 regions on either side of the previously described FMR1 promoter (27,28) was closely related, we propose that these markers are extensions of the same FMR1 regulatory region. This is further supported by our observations of FREE1 and FREE2 hypermethylation in the FXS cell lines with silenced FMR1 transcription and FMRP expression (Supplementary Material, Fig. S1). It is also of interest that one of these cell lines was taken from a female carrying FM with completely skewed X-inactivation, implying that in all cells of this patient the methylated FM was on the active X. This is consistent with the fact that this 37-year-old female was the most retarded carrier in her family (IQ<30; pedigree no. II.5), with the most severe emotional and physical abnormalities (48). It is fascinating that although this female had methylation of most CpG units within FREE1, FREE2 and the classical CpG island approaching 100%, the regions adjacent to FREE1 amelions 3 and 4 (that are largely hypermethylated in controls) were ~50% methylated in this female. Since a similar pattern was observed in the FXS cell lines from the affected males, this may suggest that in FXS, methylation shifts from the upstream regions (detected by amelions 3 and 4) to FREE1, and this may have implications for transcription from the FMR1 locus and the FXS phenotype.

The potential regulatory involvement of the FREE1 and FREE2 regions is further supported by our observations of a significant positive association between methylation status of
the two regions and that of the classical CpG island determined by Southern blot analysis in blood of high functioning males, including mosaic individuals. In addition, we have observed a significant inverse correlation between FREE1 and FREE2 and the number of FMRP-positive lymphocytes and the FMR1 activation ratio in blood of high functioning males and of the females with variable FMR1 activation ratio that carried FM alleles. It is of interest to note that mean FREE1 methylation was more closely related to FMRP expression in blood than methylation of the NrdI site examined using Southern blot, emphasizing its potential importance as a novel marker in FXS diagnostics.

In a larger sample set, the FREE1 methylation pattern was generally consistent between blood, amniocytes and chorionic villus as a marker of methylated FM alleles and X-inactivation and could be used to differentiate FXS males and females from controls, as well as from carriers of GZ and high PM alleles (expanded up to 170 repeats). However, the FREE1 methylation analysis could not differentiate between GZ and PM alleles and controls—a technical drawback that could also be a major advantage of the MALDI-TOF MS over the existing methodologies in a newborn screening programme.

In terms of potential pre-natal applications, it is important to note that in CVS of FM carriers, the methylation pattern is not necessarily consistent between all CpG sites within the classical CpG island, meaning that the analysis of a specific site using Southern blot may not necessarily reflect the methylation of a different restriction site and the CpG island as a whole. This is primarily based on reports of discordance in the methylation status of a specific site (BssHII, co-localized to footprint IV of the CpG island) in a proportion of samples between male foetal tissues and the corresponding CVS taken at 8 weeks of gestation (49,50), and lack of methylation at the FMR1 locus related to X-inactivation in CVS but not in foetal cells (17). In contrast, Devys et al. (17) have reported that hypermethylation of a different site (EagI located 10 bp upstream of footprint IV) in CVS of FM carrier males occurred as early as 10 weeks of gestation and was strikingly consistent between CVS, cord blood and different foetal tissues, as well as the clinical FXS phenotype after birth (17,51). The same may apply to different CpG sites within FREE1 analysed in chorionic villi taken between 10 and 20 weeks of gestation; however, because our analysis looks at methylation pattern across nine sites, instead of one (which is the case for Southern blot), FREE1 analysis may provide a better alternative for methylation analysis in pre-natal samples. However, before our technique is used in pre-natal diagnosis, a much larger sample size should be analysed for the consistency of these methylation patterns in CVS and amniocytes of FM carriers.

Another potential limitation is that the FREE analysis may not differentiate a proportion of FM carrier females from controls due to severely skewed inactivation, favouring the normal allele on the active X. These FM females may be as common as 30% of all FM female carriers (21) and are potentially indistinguishable from control females when using methylation or FMRP analysis, without complementary CGG-based tests (52,53). These females are also phenotypically relatively normal (54). However, this limitation can be overcome if instead of Southern blot our tests are used in combination with CGG-based analysis in targeted FXS diagnostics.

Furthermore, although the inability to detect female FM carriers with completely active normal allele may be a limitation from the viewpoint of calculating reproductive risks in adult population, it is far less relevant or may in fact be an advantage for newborn screening, where the focus is on early detection of abnormal phenotype and use of this information for improvement in the prognosis for FXS children through early treatment intervention (55). In addition, if our tests fail to identify girls who may carry FM alleles early in life, but otherwise may be clinically normal, the lack of this information can potentially avoid stigma related to ethical issues.

In summary, we have shown that the FREE MALDI-TOF MS can accurately identify all carriers of pathogenic methylated CGG repeat expansion mutations of FMR1, while addressing all the main limitations of the existing PCR-based FXS tests, associated with differentiation of unmethylated large PM from methylated FM alleles from males and females as well as detection of unmethylated and/or partially methylated FM alleles (23,24,37–39). Although applications of this technology in fragile X-related disorders have not been characterized until now, if used in combination with previously described PCR-based analyses that can differentiate between normal size, GZ and PM alleles (24,56) in FXS molecular diagnostics, the FREE MOLDI-TOF MS approach can replace Southern blot testing altogether. In contrast, if used alone, the novel test cannot differentiate between GZ and PM and normal range alleles. This avoids ethical issues and high counselling costs associated with the detection of prevalent GZ and PM alleles if used for population screening. Thus, due to its ability to rapidly examine large stretches of DNA for methylation with minimal DNA quantity and quality requirements and at low cost, the FREE MALDI-TOF MS analysis can be added to the existing set of conditions currently tested for in newborns. Early detection through newborn screening may lead to significant improvement in the prognosis for FXS children through early treatment intervention (55). It may also facilitate detection, through family cascade testing, of other affected or normal transmitting carriers, which can be provided with counselling regarding the ongoing reproductive risks of recurrence for this major cause of intellectual disability.

MATERIALS AND METHODS

Patient samples

The study was approved by the Royal Children’s Hospital Ethics Committee and by the Institutional Review Board of the University of California at Davis. FM carrier and control Epstein–Barr virus (EBV)-transformed lymphoblast cell lines were obtained from the Murdoch Childrens Research Institute tissue culture storage repository or purchased from Coriell Cell Repositories. Blood from controls, GZ and PM carriers aged between 55 and 80 was obtained from a number of centres across Melbourne, including The Movement Disorders Services unit in the Department of Neurology at the Royal Melbourne Hospital, Neuropsychiatry Unit, RMH; Melbourne Neuropsychiatry Centre, University of Melbourne; Department of Clinical Neuroscience and Neurological Research, St Vincent’s Hospital, University of Melbourne and Department of Neurosciences, Alfred Hospital, Monash University.
An aliquot of 10 ml of whole blood from each individual was collected in ethylene diamine tetraacetic acid (EDTA)-coated tubes for DNA extraction and EBV lymphoblast transformation. The DNA aliquots from blood of 21 high functioning males including mosaic individuals were taken from the samples described previously by Tassone et al. (33). The Southern blot methylation analysis values, FMRP expression by lymphocytes and standardized IQ values for some of these samples used in this manuscript were described previously by Tassone et al. (33).

All diagnostic whole blood and CVS taken for routine fragile X testing for The Victorian Clinical Genetics Services Pathology and used in this study were de-identified upon arrival. These bloods were taken from males as part of FXS cascade testing. CVS was taken trans-abdominally from women 10–20 weeks of gestation using ultrasound guidance from centres around Victoria, Australia. About 12–15 mg of each CVS was suspended in 1 ml Dulbecco’s phosphate-buffered saline (PBS)/Ca\(^{2+}\)/Mg\(^{2+}\) and transferred into 25 cm\(^2\) flasks containing 5 ml PBS, 10 ml penicillin (10 000 IU/ml)/streptomycin (10 000 IU/ml) and 10 ml heparin (1000 IU/ml) and stored at 4°C until DNA extraction.

**DNA extraction**

DNA for CGG repeat size PCR, Southern blot and MALDI-TOF MS analyses was extracted from 3 to 10 ml blood samples anticoagulated with EDTA or from EBV-transformed lymphoblast 5–10\(\times\)10\(^6\) cells per sample or freshly isolated CVS samples and extracted using a BIO ROBOT M48 DNA Extractor, as per manufacturer’s instructions (Qiagen Inc., Hilden, Germany).

**CGG repeat size PCR amplification**

The CGG repeat size for all samples was assessed initially using a fully validated PCR assay with precision of +/- one triplet repeat across the normal and GZ ranges, performed using a fragment analyser (MegaBACE, GE Healthcare), with the higher detection limit of 170 repeats, as described previously (56). Briefly, PCR amplifications were performed using primers c and f (57) in a total volume of 25 \(\mu\)l containing 50 ng of genomic DNA, 0.75 pmol of each primer, 8 \(\mu\)l of 5 \(\times\) Q-solution (Qiagen Inc.), 2.5 \(\mu\)l of 10 \(\times\) PCR buffer and 1 U of HotStar Taq Plus DNA polymerase (Qiagen Inc.) in a Gene Amp\(^{\text{®}}\) PCR System 9700. The PCR cycling profile was as follows: initial denaturation at 98°C for 5 min; 35 cycles at 98°C for 45 s, 70°C for 45 s and 72°C for 2 min and a final extension at 72°C for 10 min. Alleles were sized by capillary electrophoresis using an automatic sequencer (MegaBACE™ 1000—GE Healthcare Amersham), with size standards (HealthCare) and controls of lengths 10, 23, 29, 30, 52 and 74 repeats determined by sequencing in-house or obtained from Coriell Cell Repositories (http://www.phppo .cdc.gov/dls/genetics/qcmaterials/).

**CGG repeat size by Southern blot**

CGG sizes were assessed using a fully validated Southern blot procedure with appropriate normal and abnormal controls for samples, in which the products could not be amplified using PCR (57,58). Briefly, 5 \(\mu\)g of DNA was digested with PstI (Boehringer Mannheim, Castle Hill, Australia), separated on 1% agarose gels and analysed by Southern blot hybridization. The FMR-1 gene was detected using Southern blot analysis with probe Fxa3 and an X chromosome control probe pS8 (59). Probes were labelled using random oligonucleotide priming (Boehringer) with [a32-P]CTP (NEN Dupont, Boston, MA, USA). Autoradiography was performed at −80°C, with intensifying screens and Fuji Medical X-Ray film (Bedford, UK), and FMR1 methylation values were calculated for the expanded alleles, as described previously (21,33). The FMR1 activation ratios for female samples were calculated on the basis of the following formula: optically scanned density of the 2.8 kb band/combined densities of the 2.8 and 5.2 kb bands (where the 2.8 kb band represents the proportion of normal active X and the 5.2 kb band represents the proportion of normal inactive X), as described previously (21).

**MALDI-TOF methylation analysis**

**Bisulphite treatment.** Bisulphite treatment of genomic DNA at 0.5 \(\mu\)g per sample was performed using XCEED kit from MethyEasy (Human Genetic Signatures, Sydney, Australia) for sample sets of \(n\) < 40. For sample sets \(n\) > 40, a 96-well Methylamp kit from Epigentek (Brookly, NY, USA) was used. Duplicate bisulphite reactions were made from each sample, and six of the same control DNA samples spiked with DNA from an FXS patient cell line at 0, 33.3, 50, 66.6 or 100% were included as standards within each run, as an indicator of the intra-run variation in the degree of bisulphate-related bias. Protocols were performed according to the manufacturer’s instructions. Briefly, for the MethylEasy conversion, 20 \(\mu\)l of genomic DNA (0.5 \(\mu\)g total) was mixed with 2.2 \(\mu\)l of 3 M NaOH, incubated at 37°C for 15 min and then denatured by 45 min incubation at 80°C. An aliquot of 240 \(\mu\)l of reagent 3 (XCEED kit, Human Genetic Signatures) was then added to the mixture, which was transferred into the purification column and spun down at 10 000g for 1 min. The captured DNA was then washed in reagent 4 (XCEED kit, Human Genetic Signatures), and DNA was eluted twice by placing 50 \(\mu\)l of the pre-warmed solution 5 (XCEED kit, Human Genetic Signatures) onto the column membrane, which was incubated for 1 min at room temperature, and
spun down at 10 000 g for 1 min. The eluted DNA was then incubated at 95 °C for 20 min, with the resulting final concentration at ~20 ng/μl per sample.

For the Methylation conversion, 7 μl of genomic DNA (0.5 μg total) was mixed with 5 μl of the CF3 (Methylamp kit, Epigentek) solution diluted 1:10 in distilled water in each well of the 96-well plate. The DNA was denatured by placing the plate at 65 °C for 90 min. It was then captured in the filter plate and washed in 150 μl of the CF5 solution (Methylamp kit, Epigentek) and then twice in 250 μl of 80% ethanol. The filter plate was then incubated in the CF3/90% ethanol solution and washed twice in 90% ethanol, as per manufacturer’s instructions. The modified and cleaned DNA was then eluted with 40 μl of the CF6 solution (Methylamp kit, Epigentek), with the resulting converted DNA final concentration at ~20 ng/μl per sample. For the short-term storage, the converted DNA was kept at ~20 °C, and for storage of more than 3 months, it was kept at ~80 °C.

**PCR and in vitro transcription.** The primers used to amplify the target regions and the annealing temperatures are listed in Table 2. Each bisulphate-converted sample was analysed in duplicate PCR reactions, carried out in a total volume of 5 μl using 1 pmol of each primer, 40 μM dNTP, 0.2 U Hot Start Taq DNA polymerase (Qiagen Inc.) and 1.5 mM MgCl₂ and buffer supplied with the enzyme (final concentration 1×). The reaction mix was pre-activated for 15 min at 95 °C, followed by 45 cycles of amplification at 94 °C for 20 s, primer-specific annealing (Table 2) for 30 s and 72 °C for 1 min followed by 72 °C for 3 min. The PCR products were run on 1.5% agarose gel to confirm successful PCR amplification and efficiency. The DNA was then cleaned up, and the T- or C-cleavage reactions were carried out (T-cleave for amplicons 1–5 and C-cleave for amplicon 5 only) as per manufacturer’s instructions (SEQUENOM). Briefly, unincorporated dNTPs were dephosphorylated by adding 1.7 μl H₂O and 0.3 U shrimp alkaline phosphatase (SAP) (SEQUENOM) to PCR products, which were incubated at 37 °C for 20 min, and 10 min at 85 °C to heat-inactivate the SAP. The transcription was performed on 2 μl of template DNA in the 6.5 μl reaction consisting of 20 U of T7 R&DNA polymerase (Epigentec, Madison, WI, USA) to incorporate either dCTP or dTTP; ribonucleotides at 1 nM and the dNTP substrate at 2.5 mM, with other components used as recommended (SEQUENOM). RNase A (SEQUENOM) was then added to the mix to cleave the in vitro transcript. The mix was diluted to 27 μl in H₂O, and 6 mg CLEAN resin (SEQUENOM) was added for conditioning the phosphate backbone prior to MALDI-TOF MS. The SEQUENOM Nanodispenser was then used to spot the samples onto a SpectroCHIP for subsequent analysis. MassARRAY mass spectrometer (Bruker-SEQUENOM) was then used to collect mass spectra, which were analysed using the Epityper software (Bruker-SEQUENOM). The calculation of the output methylation ratios for each CpG unit was based on the ratio of the signal intensities for a fragment from a methylated CpG unit (methylated + unmethylated CpG units). Further details are described in reference (26).

### RNA extractions and quality assessment

Total RNA was extracted and purified using the RNeasy extraction kit, as per manufacturer’s instructions (Qiagen Inc.). RNA concentrations were measured in triplicate using a NanoDrop ND-1000 Spectrophotometer, with purity being determined by the A260/A280 ratio using the expected values between 1.8 and 2. Total RNA quality and the degree of DNA contamination were also assessed using capillary electrophoresis Standard Sens Kit (Bio-Rad), which involved descriptive comparison of chromatographic features based on previous publications using this system (60). Each RNA sample was then diluted to 30 ng/μl to be used in reverse transcription real-time PCR analysis, in which mRNA quality at the Xq27.3 region was assessed initially by examining the relationship between 5′ and 3′ levels of FMR1 mRNA, as described previously (61).

### Standard reverse transcription real-time PCR

Reverse transcription was performed one reaction per sample using the Multiscribe Reverse Transcription System, 50 U/μl (Applied Biosystems). The 7900HT Fast Real Time PCR (Applied Biosystems) was used to quantify FMR1-5′, FMR1-3′, GAPDH, B2M and GUS, using the relative standard
AmpliTaq Gold, 0.5 mM MgCl2, 1 mM buffer A (Applied Biosystems), 3.25 mM RNase-free water, 1.2 mM dNTPs, 0.01 U/µl of AmpliTaq Gold, 0.5 µl of TaqMan probe and 0.5 µl forward and 0.5 µl reverse primers and 1 µl of the reverse transcription (cDNA) reaction. The annealing temperature for thermal cycling protocol was 60°C for 40 cycles. The samples were quantified in arbitrary units in relation to the standard curves performed on each plate, standardized to the mean of three internal control genes (GUS, GAPDH and B2M).

Statistical analysis
A one-way analysis of variance was carried out for each CpG site. A P-value of less than 0.05 was considered as a statistically significant difference. Multiple testing was corrected for using the Benjamini and Hochberg method (63). This method controls the false discovery rate, the expected proportion of false discoveries among the rejected hypotheses. The relationships between methylation values and FMRP expression were assessed using a significance test in a linear regression. The analyses were conducted using the publicly available R statistical computing package (R Development Core Team 2007 R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, http://www.r-project.org/).

SUPPLEMENTARY MATERIAL
Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS
We thank the study participants and their families for their contribution. We would like to thank Louise Hills and Con Ngo from the VCGS Cytogenetics Laboratory, the Murdoch Childrens Research Institute for acquisition of patients’ CGG expansion size data, Mr Trent Peters from the Australian Genome Research Facility (AGRF) and Dr Benjamin Ong from the Sequenom Platform Facility at the Murdoch Childrens Research Institute for their assistance with MALDI-TOF MS applications.

Conflict of Interest statement. The authors declare that they have no conflicts of interest.

FUNDING
This work was supported by the March of Dimes Foundation, USA (6-FY06-318 to K.H.A.C.); Senior Principal Research Fellowship of the National Health and Medical Research Council of Australia (334300 to K.H.A.C.); National Institute of Child Health and Human Development grant, USA (HD36071 to D.Z.L.) and NHMRC project grant (no. 330400 to D.Z.L.).

REFERENCES


