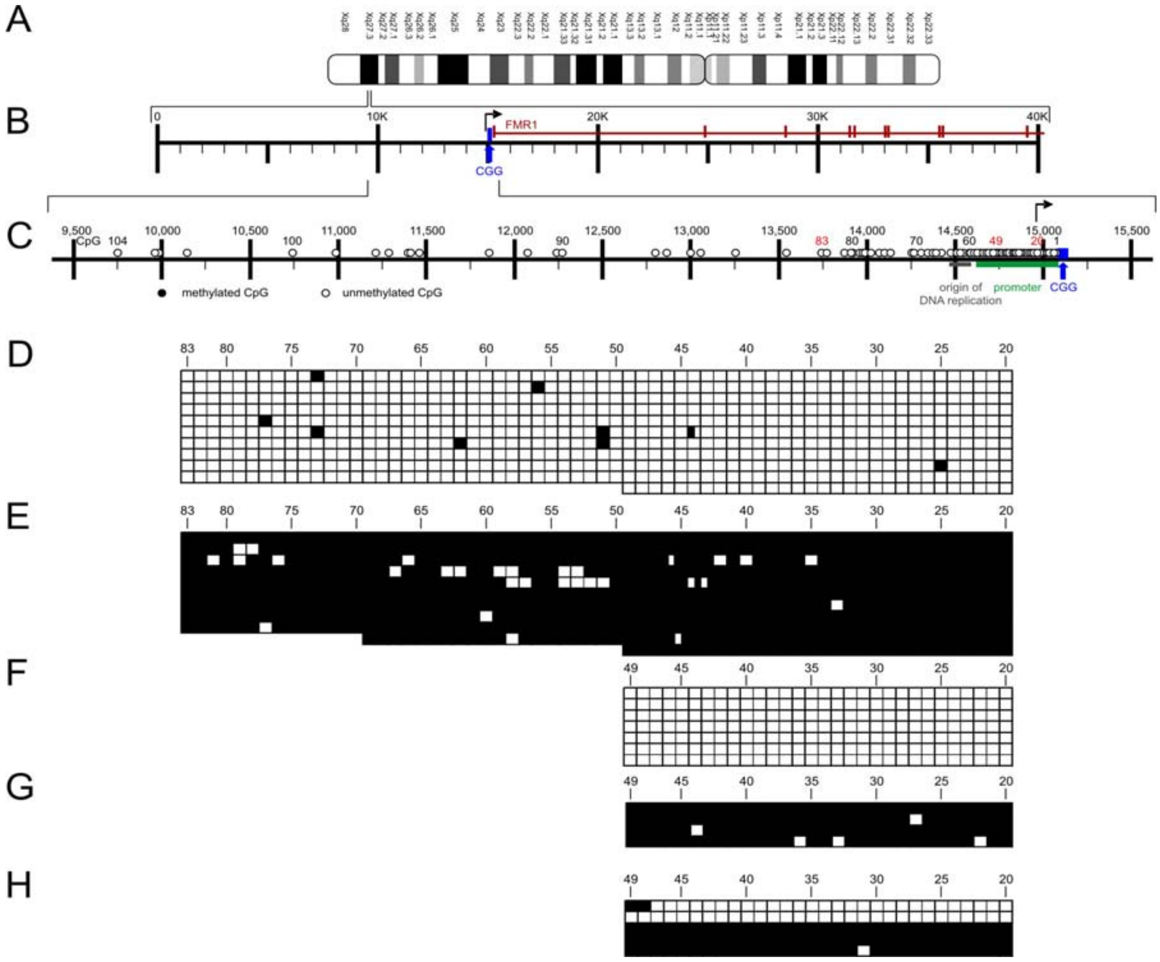
Supplemental Data

AJHG*, Volume *85

A Distinct DNA-Methylation Boundary in the 5′- Upstream Sequence of the *FMR1* Promoter Binds Nuclear Proteins and Is Lost in Fragile X Syndrome

Anja Naumann, Norbert Hochstein, Stefanie Weber, Ellen Fanning, and Walter Doerfler

Figure S1. Partial Map of the *FMR1* Gene and Its 5´-Upstream Region – Control Experiments



(A) Ideogram of the human X chromosome.

(B) Partial map shows the first ten exons (vertical bars) and introns of the *FMR1* gene plus the upstream genome segment including the CGG repeat.

(C) The graph presents all CpG dinucleotides (1 to 104) in the region: Open circles designate unmethylated, closed circles methylated CpG dinucleotides. The arrow indicates the start site of transcription. The promoter and an origin of DNA replication are also indicated. Nucleotide numbering in this and the following graphs was adapted from the NCBI nucleotide nomenclature: NC_000023:146,786,201-146,840,303 *H. sapiens FMR1* gene region (nucleotide numbers on the human X = 23rd chromosome).

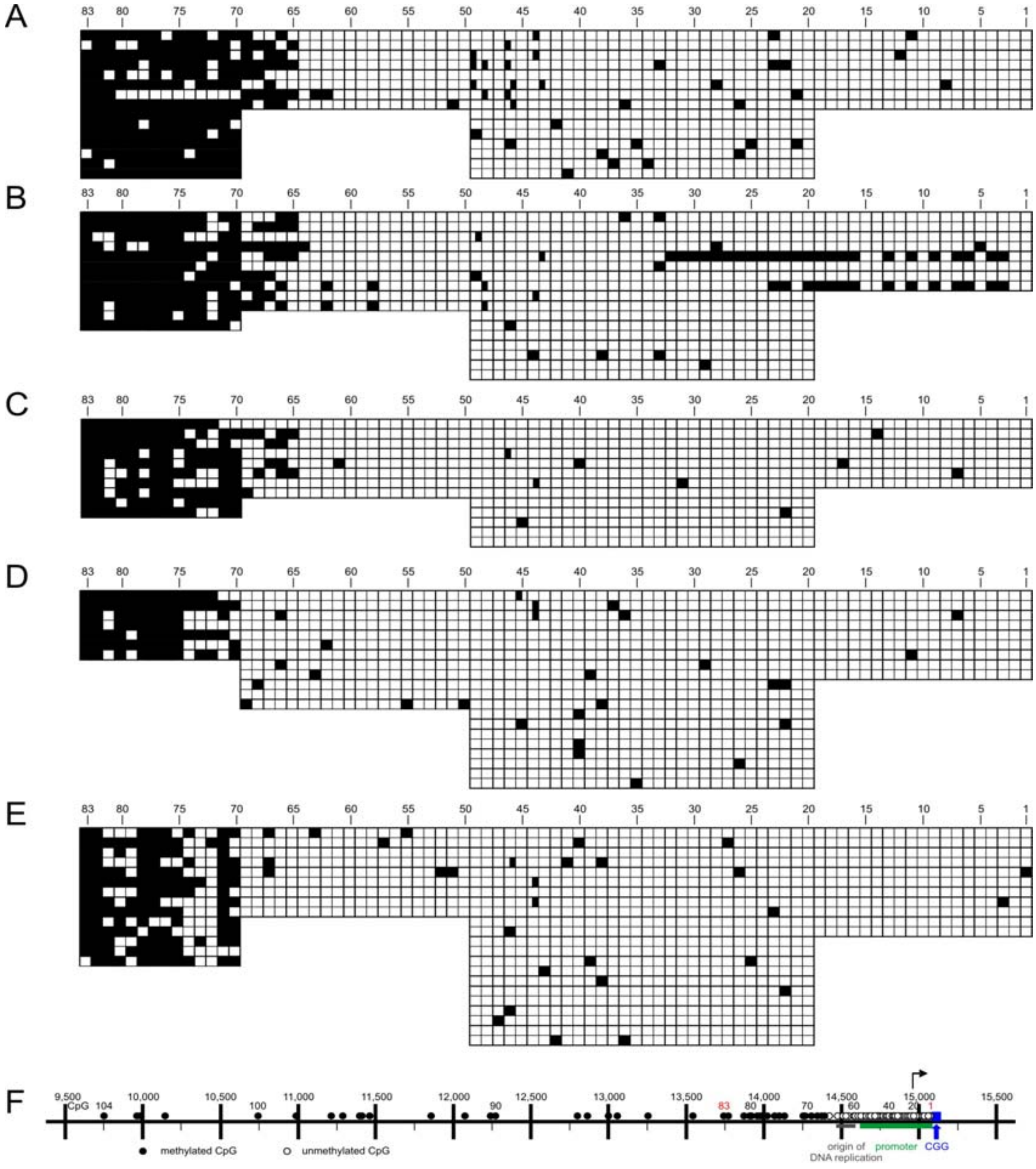
(D, E) Human DNA from a commercial source was obtained in the unmethylated (D) or the *M.Sss I*-premethylated form (E), subjected to denaturation, bisulfite conversion, and PCR amplification of the *FMR1* promoter and 5′-upstream region, cloning and sequencing. With the exception of spurious deviations, the un-methylated DNA (CpG dinucleotides 20 to 83) registered as such (D), the premethylated DNA was recognized as 5-mC methylated DNA (E). The M.Sss I DNA methyltransferase methylates all CpG dinucleotides in a DNA sequence.

(F, G) In a similar experiment, part of the *FMR1* 5′-upstream region (CpGs 20 to 49) as a pUC19-cloned DNA fragment[25] was treated in the same way. (F) unmethylated, (G) *M.Sss I*-premethylated DNA.

(H) Analysis of a 1:1 mixture of DNA in (G) and (H). □ – unmethylated,

■ – methylated CpGs.

Figure S2. Methylation Boundaries in the *FMR1* Upstream Region in DNA from Fetal and Adult Human Organs



Human DNA from the following organs was analyzed: (A) fetal brain, (B) fetal liver; (C) adult brain, (D) adult liver, (E) adult muscle.

(F) *FMR1* map as in Figure 2C.

Table S1. Cells, Cell Lines and DNA Samples Analyzed in this Study

1. Human Cell Line				
HCT116 colon carcinoma	male		~ 30 5'-(CGG)$_n$-3' repeats	ATCC ' CCL-247
2. Primary Normal Human Cells				
HFF human foreskin fibroblasts	male			Pediatric clinic, Erlangen
PBMCs	male			Volunteers from Institute for Virology, Erlangen
PBMCs	female			Volunteers from Institute for Virology, Erlangen
3. Human DNA Samples				
Brain	male	adult		BioCat GmbH, Heidelberg
Liver	male	adult		BioCat GmbH, Heidelberg
Skeletal muscle	male	adult		BioCat GmbH, Heidelberg
Brain	male	fetal		BioCat GmbH, Heidelberg
Liver	male	fetal		BioCat GmbH, Heidelberg
4. FRAXA Cells				
GM07072 primary FRAXA lung fibroblast	male	22 FW	~ 450 5'-(CGG)$_n$-3' repeats	Coriell Cell Repositories (CCR), Camden, NJ
GM05848 primary FRAXA fibroblast	male	4 YR	~ 670 5'-(CGG)$_n$-3' repeats	Coriell Cell Repositories (CCR), Camden, NJ
14,451 PBMCs	male	9 YR	~ xxx 5'-(CGG)$_n$-3' repeats	Institute of Human Genetics, Erlangen
5. Mouse DNA Samples				
Brain	male	adult		BioCat GmbH, Heidelberg
Liver	male	adult		BioCat GmbH, Heidelberg
6. Controls				
Unmethylated and bisulfite converted human control DNA				Qiagen GmbH, Hilden
Premethylated and bisulfite converted human control DNA				Qiagen GmbH, Hilden
pUC19-FMR1: unmethylated, premethylated, 1:1 mixture				Vanderbildt University, Nashville, TE

Table S2. Primers Used in This Study

Primer	Position			Sequence 5' to 3'	Strand
NC_000023:146,786,201-146,840,303 *H. sapiens FMR1* gene region					
P33	13,659	f		TTA ATT TAG TGT TAT TAG TTA ATT TTT ATT	upper
P34	13,699	f	n	TAG TTG GAA TTA TTT TTT TTT TTT TAT TTT	upper
P35	14,288	r	n	AAA CCT TTC CTC AAA CCC AAC TTT AAC CCA	upper
P36	14,348	r		CCC TAT AAA AAC ACC CTA TAC CCT TTA AAA	upper
P29	14,169	f		GTA GAA ATG AAT TTT AAG TGT TTA ATA TAA	upper
P30	14,259	f	n	TGG GTT AAA GTT GGG TTT GAG GAA AGG TTT	upper
P31	14,788	r	n	AAA AAC CRA CCR AAA TAC CRA ATC RAA AAA	upper
P32	14,838	r		CCC CRC CCT CTC TCT TCA AAT AAC CTA AAA	upper
P25	14,664	f		GGG YGT TTT GGT TTT YGY GAG GTA GTG YGA TT	upper
P26	14,699	f	n	TAT YGT TTT TTA GTT TTT TYG TTT TTT ATT AAG	upper
P27	14,991	r	n	RCC CRC TCA AAA ACR ACC CTC CAC CRA AAA TAA A	upper
P28	15,056	r		ACC RCA CRC CCC CTA ACA ACR ACR CCT CCR T	upper
P21	14,589	f		AAT CCC AAA AAA ACC RAA CTA AAA TAA CCR	lower
P22	14,629	f	n	ATT TCC CAC RCC ACT AAA TAC ACC TCT ACA	lower
P23	14,838	r	n	TTT YGT TTT TTT TTT TTA AGT GGT TTG GGA	lower
P24	14,878	r		ATT AAY GTT GTT TTT TTT TTT YGY GGG TTT	lower
P92	9,560	f		ACA AAA CCC ATA AAT ACA AAA	lower
P93	9,670	f	n	AAT TCA ATA AAA AAA CAA TAA TC	lower
P94	10,225	r	n	TAA TAG GTG TAT AGT AGA AT	lower
P95	10,345	r		AGA TGA TTT TTT AAA GTG GTT G	lower
P96	10,530	f		AAT AAA CAA ACC ATA ATA TA	lower
P97	10,669	f	n	ACA TAC TAT ATA ATT CCA ACT	lower
P98	11,042	r	n	AAA ATT AAG TAG ATA GTG GAG AG	lower
P99	11,121	r		TGA TTT TTT ATT TAA ATG AG	lower
P100	11,109	f		AAA TAA AAA ATC ATA ATT AAA AAA AAT TAT	lower
P101	11,159	f	n	AAA TAT AAC TTT AAC TAA ATC CTA AAA AAT	lower
P102	11,508	r	n	TTT GTT TTT AGT GTT TAT TAT AGT GTT TGG	lower
P103	11,540	r		AGA ATA AAG ATT TTT ATT TGT TTT ATT TAT	lower
P40	11,531	f		TCT TTA TTC TCA CTA AAC TTA C	lower
P41	11,808	f	n	AAA CTA AAT AAA TTT CAT ACA C	lower
P42	12,304	r	n	TTA AAG TTT TTT TTT TTT TT	lower
P43	12,417	r		TTT GAT TTT GAG TTT TAG TTT	lower
P108	12,594	f		TAT TCT AAT ACA CTA ACC ATC ATA ATA AAA	lower
P109	12,659	f	n	AAC AAT ATA ATA AAA AAA CAA AAA ATA TAC	lower
P110	12,963	r	n	TGT TTA TGT TTT TTG TAT TTA TAG AGA TGG	lower
P111	13038	r		TTA TTT TTA TTT TTT TAG ATA TAA GTT GTA	lower
P112	12,869	f		ATA AAT ATA ACT CAA AAA CAA ACT ATT AAA	lower
P113	12,904	f	n	CAA TAA AAA ATA AAA ATT TAC AAT CCA ATA	lower
P114	13,308	r	n	TAT TAT AAG TTA TAT TTA ATT TGT GTT GTT	lower
P115	13,348	r		TTT AGG GTT TAA AAA TAT GTT AAT TTT TTA	lower
P116	13,359	f		AAC AAC TTA CAC TTA AAA AAA TAT AAT CAT	lower
P117	13,419	f	n	ACA TTA AAA TTA CAA CTA ATA CTC TAC TCC	lower

P118	13,615	r	n	AGG ATG TTG TTG TTG GTG GAA TTT TAG TTT	lower
P119	13,648	r		ATG ATT TGT TTT AGG TAG GTG TTT TAG AAT	lower
P13	13,659	f		CCA ACT CAA TAC TAT TAA TTA ACT CTC ACC	lower
P14	13,699	f	n	CAA CTA AAA TCA TCT CCC CTT CTC CAC CCC	lower
P15	14,288	r	n	GAG TTT TTT TTT AGA TTT AGT TTT GAT TTA	lower
P16	14,348	r		TTT TGT GGA GAT ATT TTG TGT TTT TTA AGG	lower
P9	14,169	f		ACA AAA ATA AAC CTC AAA TAC TTA ACA CAA	lower
P10	14,259	f	n	TAA ATC AAA ACT AAA TCT AAA AAA AAA CTC	lower
P11	14,788	r	n	GGG AAT YGG TYG GGG TGT YGG GTY GAA AGA	lower
P12	14,838	r		TTT YGT TTT TTT TTT TTA AGT GGT TTG GGA	lower
P1	14,664	f		AAA CRT TCT AAC CCT CRC RAA ACA AAT ACR ACC	lower
P2	14,699	f	n	CAC CRC CCT TCA ACC TTC CCR CCC TCC ACC AAG	lower
P3	14,991	r	n	GTT YGT TTA GAG GGY GGT TTT TAT YGG AAG TGA A	lower
P4	15,056	r		GTY GTA YGT TTT TTG GTA GYG GYG TTT TYG T	lower
P5	14,956	f		ATT TCA CTT CCR ATA AAA AAC CRC CTC TAA ACR AAC	lower
P6	15,029	f	n	AAA ACR CCR CTA CCA AAA AAC RTA CRA CAA C	lower
P7	15,155	r	n	TTT YGA GAG GTG GGT TGY GGG YGT TYG AGG TTT AG	lower
P8	15,195	r		TTT TAT TTT TTT TTT AGT TTT GTT AGY GTY GGG AG	lower
NC_000086:65,916,730-65,971,138 *M. musculus Fmr1* gene region					
M13	13,447	f		TAA AAC AAC TCA CCC TTA CTC TTT CAA ATA	lower
M14	13,497	f	n	TCT TAT CAC ATT CCA ATT AAA ATC ATC TTA	lower
M15	13,984	r	n	TAG GAG TAG ATG TGT AAA ATT TAA TTT TTT	lower
M16	14,030	r		GGT GAG GTA GTA ATA AAA AGA TTG TTA TTA	lower
M9	14,061	f		AAC ACA TAA TAA ACA TTC AAT AAA CTT CTC	lower
M10	14,096	f	n	AAT ACA AAC CTT CTA AAT TAA AAA AAC ATA	lower
M11	14,420	r	n	TTT TTT TTT TAG ATA GTT TTT TTT ATA TTT	lower
M12	14,460	r		TAA GGG TTA ATA GGG TGT GTT AGA ATT TTA	lower
M5	14,391	f		AAA TAT AAA AAA AAC TAT CTA AAA AAA AAA	lower
M6	14,431	f	n	TAA AAT TCT AAC ACA CCC TAT TAA CCC TTA	lower
M7	14,893	r	n	TTT TTT GGG TAT TTG TTT TAG GGT ATT AGG	lower
M8	14,936	r		TTY GTT TTT GTT TTT TTT YGG GTA GTG GAG	lower
M1	14,831	f		CCT TTC ACC AAA TAC RCR CRC CAA AAA ATA	lower
M2	14,864	f	n	CCT AAT ACC CTA AAA CAA ATA CCC AAA AAA	lower
M3	15,226	r	n	GGA GGT GGG TTG TAG GYG TTT GAG GTT TAG	lower
M4	15,295	r		GGA GTT TYG TAT TTT TAT TAT TAG TTT TTT	lower

Y = C or T; R = A or G; f = forward primer; r = reverse primer; n = nested PCR primer

The upper or lower strand was amplified during PCR.