CHAPTER 2
RESEARCH METHODOLOGY

Materials

DNA Samples

DNA samples of unrelated males with normal CGG alleles (19-56 repeats) and FXS patients (>200 repeats) were obtained from the Human Genetics Units at Songklanagarind Hospital and Ramathibodi Hospital. The samples had been diagnosed previously through a routine FXS analysis at laboratories using either PCR amplification of the CGG repeat region or Southern blot techniques. Allele and haplotype frequencies of the polymorphic markers were derived from a total of 133 normal X chromosomes and 50 FXS chromosomes. The control groups were divided into 6 subgroups corresponding to CGG repeat lengths of 19-28, 29, 30, 31-35, 36 and 37-56 repeats. The numbers of DNA samples in the common CGG alleles (29, 30 and 36 repeats) were randomly selected from plentiful sample pools of these alleles according to the CGG allele distributions in normal Thai subjects (Appendix B, Figure 10), whereas all of the DNA samples of the rare, intermediate and full mutation alleles were collected from both laboratories due to the limited sample numbers. In addition, a total of 133 normal individuals and 50 FXS patients, with 111 normal X chromosomes (83.5%) and 25 FXS chromosomes (50.0%) had been previously analyzed for 2 microsatellites (DXS548 and FRAXAC1) and 2 SNPs (ATL1 and IVS10) by Limprasert (2001) were considered in the study sample. The total amount of DNA sample from both prior samples and new cases in each group is detailed in Table 1.

We incorporated the DNA samples of two affected families from the Human Genetics Unit at Songklanagarind Hospital. The members of these families had at least 3 affected generations, including a grandmother, a mother who carried the premutation allele and her affected children. These samples also had been diagnosed for the number of CGG repeats by routine FXS analysis. All subjects had their haplotype established using standard polymorphic markers.

Table 1. The subgroups and numbers of DNA samples.

CGG repeat numbers	Number of DNA samples		
	Prior	New	Total
Normal			
19-28	7	9	16
29	32	-	32
30	29	2	31
31-35	4	9	13
36	28	-	28
37-56	11	2	13
Total	111	22	133
FXS			
>200	25	25	50

Primers for PCR amplification

PCR primers of two microsatellites were constructed following the methods of Riggins (1992) for DXS548, Richards (1991) for FRAXAC1 forward primer and Zhong (1993) for FRAXAC1 reverse primer. ATL1 SNP primers were imitated from Brightwell (2002a) for ATL1F, ATL1R and ATL1AF and Curlis (2005) for ATL1GR. Novel primers of four SNPs (WEX5, IVS10, rs25702 and rs25723) were designed specifically for use with the Biallelic-ARMS PCR technique, along with a SNP locus of rs25731 which was generally designed for PCR-RFLP techniques. In the study of AGG interruption, published primers (primer A and 571R) according to Chong (1994) were used to amplify across the CGG repeat region. Due to the difficulty of working with higher CGG lengths (≥37 repeats), we also devised a new sequencing forward primer, FXS-SEQF, about 70 bp distal to primer A. All primer sequences are listed in Table 2.

Table 2. The oligonucleotide sequences of primers.

Locus	Name	Sequence (5′ to 3′)	Reference (year)
Microsatellites			
DXS548	DXS548F	GTACATTAGAGTCACCTGTGGTGC	Riggins (1992)
	DXS548R	AGAGCTTCACTATGCAATGGAATC	
FRAXAC1	FRAXAC1F	GATCTAATCAACATCTATAGACTTTATT	Richards (1991)
	FRAXAC1R	GATGAGAGTCACTTGAAGCTGG	Zhong (1993)
SNPs			
WEX5	WEX5F	GAATGTGGCCCTAGATCCAC	Newly designed
	WEX5R	GTGCTAACGAGAAATCGGTG	
	WEX5CF	CTTATCACAGCTGCAACTACAC	
	WEX5GR	CAAATTGTCAGACAAGTAAACC	
ATL1	ATL1F	ACCCTGATGAAGAACTTGTATCTCT	Brightwell (2002a)
	ATL1R	GAAATTACACACATAGGTGGCACT	
	ATL1AF	TGTACATTTTCCAAATGCAAAGA	
	ATL1GR	AGAGACACAGAATCATAAATGC	Curlis (2005)
rs25731	731F	AGATTCCCACCTCCTGTAGG	Newly designed
	731R	CATGCTCTGAGTACTGCTC	
IVS10	IVS10F	AAAGCTGATTCAGGAGATTGTG	Newly designed
	IVS10R	ACTGCATTAGAGGACAGAGA	Xu (1999)
	IVS10TF	CAAGAAGAGGTATGTTACAGTAT	Newly designed
	IVS10CR	TTATTATATGTGCCACAAAATATTGG	
rs25702	702F	ACTCAGTTTAGGCAATCCTG	Newly designed
	702R	CACAGCTAGTTCATTTGCTG	
	702AF	TCAGTTTAGTTAGTGTGATGTA	
	702GR	GAAATTTTAAGGAGGCATAATC	
rs25723	723F	GAGCGAGACTGTCTGGGAA	Newly designed
	723R	TGGAAGGACTGGAATCCTAG	
	723AF	ACATTTAAAACACATGCACATATA	
	723CR	TTTCAAAGTATGTTTAAGTAGTAG	
	Primer A	GGAACAGCGTTGATCACGTGACGTGGTTTC	Chong (1994)
	Primer 571R	GGGGCCTGCCCTAGAGCCAAGTACCTTGT	
	FXS-SEQF	TCTGAGCGGGCGGCGGGCCGA	Newly designed

Reagents

Acetic acid, glacial

Acrylamide/Bis solution

Agarose

Ammonium hydroxide

Ammonium persulfate

AmpliTaq Gold DNA polymerase (AB)

BigDye Terminator v1.1 (AB)

Boric acid

Bovine serum albumin

Bromophenol blue

Citric acid

dATP, dTTP, dCTP, dGTP, 7-Deaza dGTP

DMSO

DyeEx 2.0 Spin kit (Qiagen)

EDTA

Ethidium bromide

Formaldehyde

Formamide

Glutaraldehyde

Glycerol

Immolase DNA polymerase (Bioline)

Methanol, absolute

QIAquick PCR purification kit (Qiagen)

Restriction enzyme, *Dra* I (NEB)

Silver nitrate

Sodium hydroxide

Taq DNA polymerase (Invitrogen)

TEMED

Tris

Template suppressor reagent (AB)

Urea

Xylene cyanol FF

Equipment

Genetic analyzer (ABI PRISM 310, AB)

Autopipettes: P2, P10, P20, P200 and P1000

Beakers: 50, 100, 250, 500 and 1000 ml

Bottles: 50, 100, 250, 500 and 1000 ml

Centrifuge tubes: 15 and 50 ml

Cooling box

Cylinders: 50, 100 and 500 ml

Freezer

Gel dryer

Gel image systems with UV transilluminator

Gloves

Glass pipettes: 1, 5, 10 and 25 ml

Horizontal mini gel electrophoresis (GelMate2000, TOYOBO)

Hot plate

Incubator

Magnetic stirrer

Microcentrifuge

Microcentrifuge tubes: 0.2, 0.5 and 1.5 ml

Microtips: 0.1-10, 2-200 and 1000 μl

PCR Thermal cycler (PTC200, MJ Research)

PCR Thermal cycler (GeneAmp PCR System 9700, AB)

Plastic boxes

Racks

Refrigerator

Sample boxes

Shaker

Spectrophotometer

SpeedVac systems

Syringes: 5 and 25 ml

Vertical gel electrophoresis with power supply (Protean II Xi cell, 20X20 cm, Biorad)

Vortexer

Required solutions

The components and quantities of chemicals in each solution are detailed in Appendix C.

8% Acrylamide solution

25% Ammonium persulfate

2.3 M Citric acid

Developer solution

2.5 mM dNTP Mix

2.5 mM dNTP Mix with 50% 7-Deaza dGTP

6X Gel loading buffer

10% Glutaraldehyde solution

Loading dye

10% Methanol-Acetic acid solution

8% Polyacrylamide gel

20% Silver nitrate

Silver solution

0.36% Sodium hydroxide

50X TAE

10X TBE

Methods

The selection of polymorphic markers

We selected 4 prior investigated markers (2 microsatellites: DXS548, FRAXAC1 and 2 SNPs: ATL1, IVS10) in Thai subjects reported by Limprasert (2001) along with 4 novel SNPs (WEX5, rs25731, rs25702 and rs25723) from a SNP map for the FRAX region in Brightwell (2002b) as well as from the dbSNP (database SNP) of the NCBI web site (www.ncbi.nlm.nih.gov/SNP) and the International HapMap Project web site (www.hapmap.org). These SNPs had high heterozygosity (close to the value of 0.5), near the frequencies of both alleles in Asian populations and spread throughout the *FMR1* gene (Table 3). These markers were located both proximal and distal to the CGG repeat region of the *FMR1* gene (Figure 1).

Table 3. The microsatellite and SNP markers.

Polymorphism	Allele	Location	Nucleotide position (L29074)	Position relative to CGG repeat region (~Kb)	Heterozygosity	Allele frequency	
						Chinese[c]	Japanese[d]
DXS548	AC repeat	Upstream	-	-150	-	-	-
FRAXAC1	AC repeat	5′ UTR	6613	-7.2	-	-	-
WEX5 or rs1805420	C/G	5′ UTR	12570	-1.3	0.352	C 0.523 G 0.477	C 0.523 G 0.477
FRAXA	CGG repeat	Exon1	13833	0	-	-	-
ATL1[a] or rs4949	A/G	Intron1	19445	+5.6	0.481	A 0.375 G 0.625	A 0.409 G 0.591
rs25731	A/T	Intron3	27484	+13.7	0.351	T 0.511 A 0.489	T 0.523 A 0.477
IVS10[b] or rs25714	C/T	Intron10	38410	+24.6	0.415	C 0.378 T 0.622	C 0.488 T 0.512
rs25702	A/G	Intron13	42950	+29.1	0.419	G 0.375 A 0.625	G 0.477 A 0.523
rs25723	A/C	Intron16	49567	+35.7	0.447	C 0.378 A 0.622	C 0.477 A 0.523

[a] Allele frequencies of normal chromosomes in Thai subjects studied by Limprasert (2001), A 0.312 and G 0.688.

[b] Allele frequencies of normal chromosomes in Thai subjects studied by Limprasert (2001), C 0.304 and T 0.696.

[c] Allele frequencies in Chinese from the International HapMap Project.

[d] Allele frequencies in Japanese from the International HapMap Project.

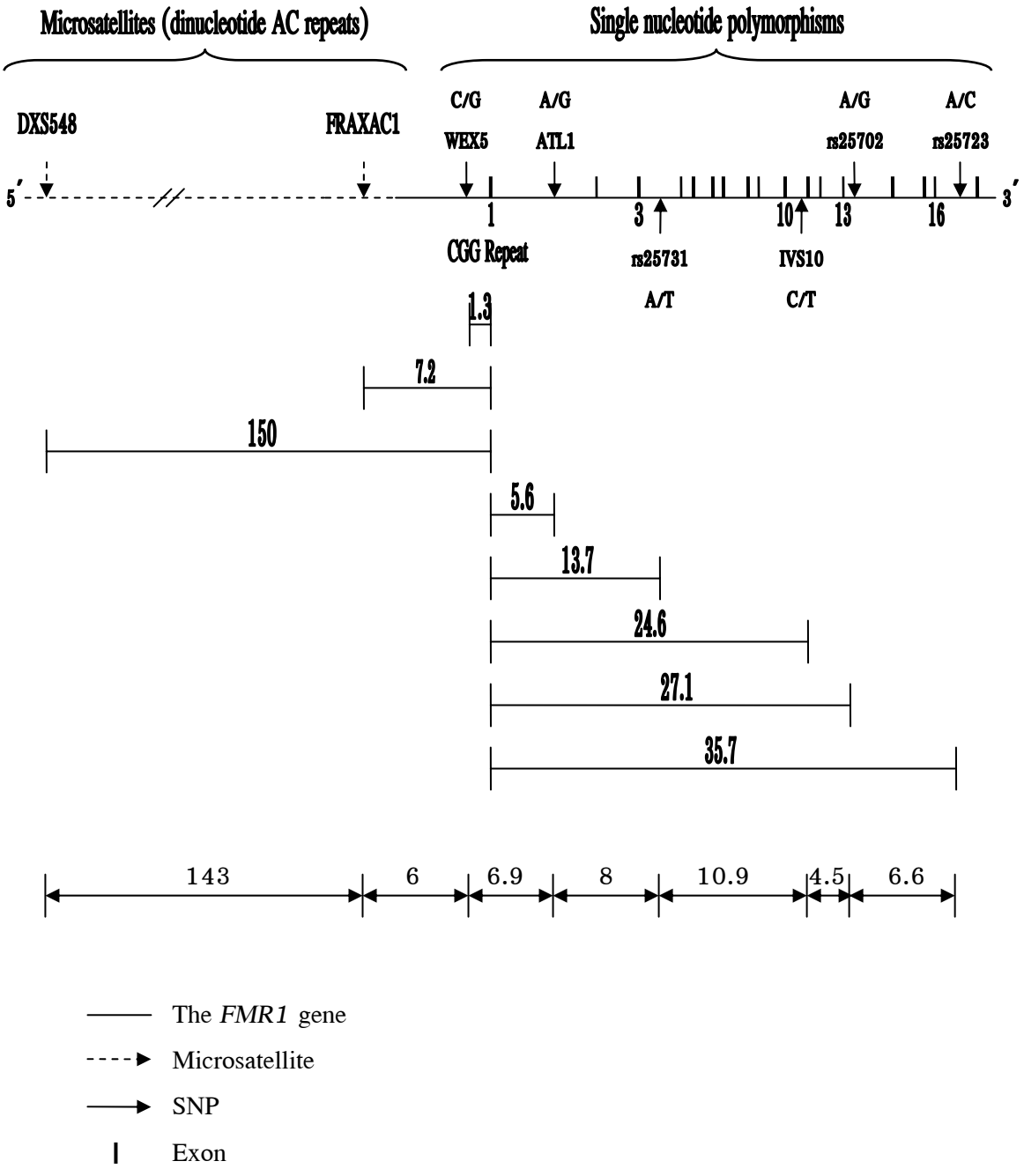

Figure 1. The locations of microsatellite and SNP markers. The polymorphic markers flank the CGG repeat region and the distances relative to the CGG region along with the distance between each marker in Kb are shown below.

Primer design

Each SNP locus was genotyped by the Biallelic-ARMS PCR technique which is able to analyze 2 different alleles in the same reaction. The single PCR reaction contained 4 primers comprised of 2 control primers (CP) and 2 allele specific primers (ASP). As seen in Figure 2, both ASPs were designed to prime in opposite directions at the site of the SNP to be genotyped so that one ASP was a forward strand (ASPF) and another ASP was a reverse strand (ASPR). The last 3′ base of each strand was complementary to one of the SNP alleles and the additional mismatch at position -1 from the 3′-terminus using the sterically unfavorable purine to purine and pyrimidine to pyrimidine mismatch as the general rule for a single base alteration was also intentionally incorporated in these primers to increase allelic specificity. However, two CPs were designed normally without any mismatch as the forward (CPF) and reverse strands (CPR) but their positions flanked the SNP at different distances. An allele specific product was generated by each ASP and its opposing CP, while the control product that served as the internal control in every reaction was amplified by the control forward and reverse primers. Therefore, both SNP alleles and control products were identified on the basis of distinct PCR product size. The reference sequence of the *FMR1* gene, accession number L29074, from GenBank, was downloaded to employ as the original sequence template for primer design. All primers were manually designed using several criteria of primer design. They were:

1. 18-30 nucleotides in length.
2. 40-60% GC content for control primers. Because of an unavoidable sequence encompassing the SNP site, the primer GC content of allele specific primer was no regard.
3. Primer sequences according to primer design principles, they were designed to:

3.1 Have random base distribution.

3.2 Avoid T, runs of three or more G or C and palindromic sequences at the 3′ end.

3.3 Contain G or C or GC or CG residue, called GC clamp, at the 3′ terminus.

3.4 Avoid homopolymeric runs (polypurines and polypyrimidines) and repeated sequences along with inverse repeat (internal complementarity).

3.5 Avoid the 3′ complementary sequence between the primer itself and primer pairs.

4. Similarity of melting temperature for each primer ranging 60-65 °C

It was also an important consideration that the product size is between 100-500 bp so it would be clearly distinguishable on 2.5% agarose gel. After many oligonucleotide sequences were constructed, their sequence properties were inspected for hairpins and dimers with the Vector NTI5 program, and they were subsequently aligned with the NCBI nucleotide-nucleotide BLAST (blastn) program to check the primer specificity. The best oligonucleotide primers with the fewest hairpins and dimers, as well as having high specificity (a low Expect value) were chosen.

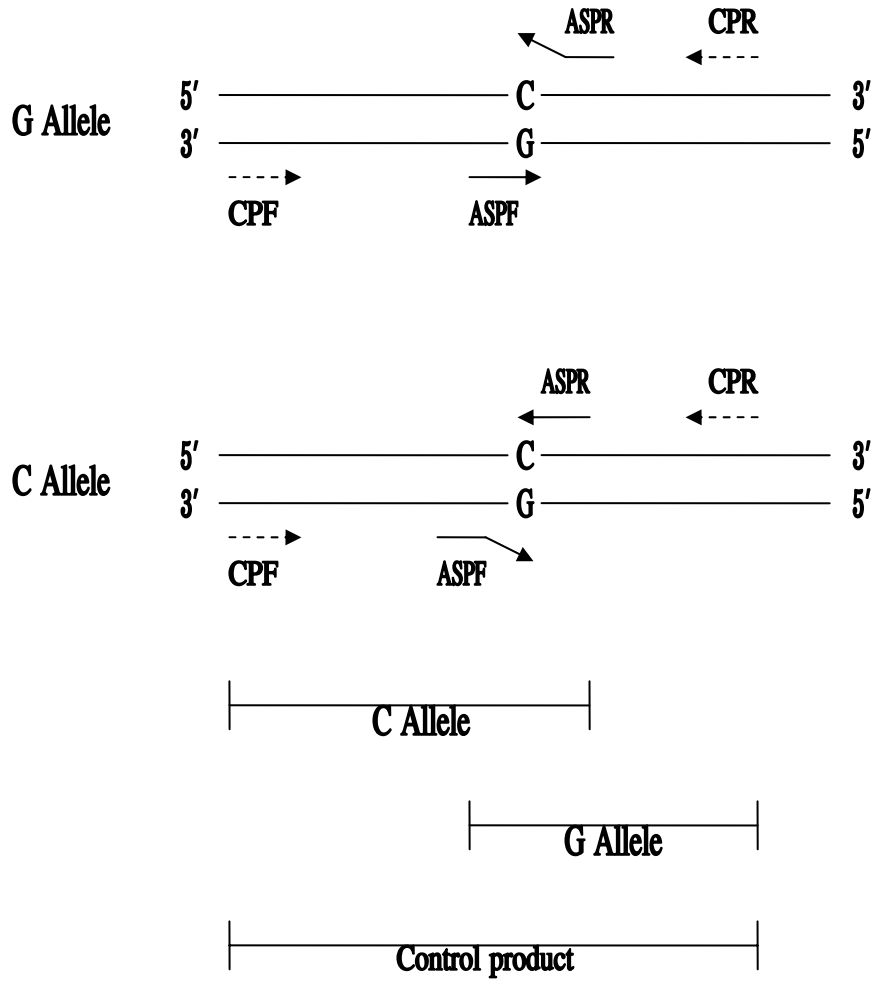

Figure 2. Schematic of the tetra-primer in Biallelic-ARMS PCR method.

Microsatellite analysis

We followed the technique of Limprasert (2001) with minor modification to amplify a singleplex PCR of the dinucleotide AC repeat region. A 20 µl reaction contained 1 X PCR buffer, 1 mM $MgCl_2$, 200 µM each of dNTP, 1 µM each of forward and reverse primer, 1 unit Taq DNA polymerase (Invitrogen) and 100 ng genomic DNA. The PCR reaction as programmed on a MJ thermal cycler was: 95 °C for 4 min, followed by 35 cycles of 95 °C, 30 sec; 55 °C, 30 sec; 72 °C, 30 sec and a final extension of 72 °C for 10 min. A 6 µl aliquots of the PCR reaction were blended with 9 µl of loading dye, heated in boiling water for 10 min, chilled on ice for 5 min and then subjected to 8% vertical denaturing polyacrylamide gel electrophoresis with 1 X TBE at 400 Volts in total running times of 7 and 6 hours for DXS548 and FRAXAC1, respectively. The bottom of the gel was cut into 8 cm-width gel and the PCR products were detected with the alkaline silver staining method as explained in Table 4. The AC alleles were compared with known allele sizes of loci DXS548 and FRAXAC1 as determined by sequencing.

Table 4. The protocol for silver staining of nucleic acids in polyacrylamide gels as used in the study (Nopparatana C., personal communication).

Step	Reagent	Volume (ml)	Time
1. Fixation	10% Methanol-Acetic acid solution	100	20 minutes
2. Sensitization	10% Glutaraldehyde solution	50	20 minutes
3. Rinse	Deionized distilled water	100	A few seconds
4. Wash	Deionized distilled water	400	Overnight
5. Rinse	Deionized distilled water	200	5-10 minutes
6. Silver impregnation	Silver solution	50	10 minutes
7. Wash (2X)	Deionized distilled water	200	10 minutes each
8. Image development	Developer solution	100	2-10 minutes

Table 4. (continued)

Step	Reagent	Volume (ml)	Time
9. Stop	10% Methanol-Acetic acid solution	100	5 minutes
10. Wash	Deionized distilled water	200	10 minutes

Note: All incubations were performed at room temperature with gentle agitation.

Volumes of the solutions used in each stage were sufficient for two 8X8-cm gels.

High quality deionized distilled water was required for both reagent preparation and washing.

Only the silver solution and developer solution were prepared fresh at the time needed; especially the unstable silver solution had to be used immediately and protected from light.

For preparation of silver solution, $AgNO_3$ solution was gradually added to the mixture drop by drop along with quick and thorough mixing to disrupt the formation of a brown precipitate in the clear solution.

The stained gel was dried 20 minutes in gel dryer for long term storage.

SNP analysis for Biallelic-ARMS PCR techniques

A single-tube multiplex PCR was amplified in a total volume 10 μl reaction containing 1 X PCR buffer, 200 μM each dNTP, 0.5 units Taq DNA polymerase (Invitrogen) and 50 ng genomic DNA. The presence or absence of 0.13 mg/ml BSA in the PCR reactions and the $MgCl_2$ concentration were optimized to obtain maximum yield of multiplexed PCR products. In order to enhance the efficiency of allele specific amplification, the concentration ratios of tetra-primer for each SNP assay were adjusted to produce similar band intensity of each PCR product after gel electrophoresis. All PCR reactions were carried out by a MJ thermal cycler under the following conditions: 5 min initial step at 95 $^{\circ}$C, 35 cycles of 30 sec in each 95 $^{\circ}$C denaturation, appropriate annealing temperature as described in Table 5, 72 $^{\circ}$C extension and a final extension for 10 min at 72 $^{\circ}$C. The total volumes of the PCR reactions were mixed with 2 μl of gel loading buffer. Multiple products of PCR were sized by horizontal electrophoresis on 2.5% mini-agarose gel with 1 X TAE at 100 Volts for 50-55 min, then stained in ethidium bromide solution

for 10 min and visualized on UV transilluminator. The SNP alleles of individual locus were genotyped according to the different size of the PCR products as shown in Table 5.

SNP analysis for PCR-RFLP techniques

We used the PCR-RFLP technique to create a SNP locus of rs25731 (Figure 3). In the 20 μl PCR reaction consisted of 1 X PCR buffer, 200 μM each dNTP, 1.5 mM $MgCl_2$, 0.25 μM of each forward and reverse primer, 1 unit of Taq DNA polymerase and 100 ng of DNA templates. PCR reactions were initiated using a MJ thermal cycler under the following conditions: 5 min initial step at 95 °C, 35 cycles of 30 sec in each 95 °C, 60 °C, 72 °C and a final extension for 10 min at 72 °C. After the success of the singleplex PCR amplification, 4 μl of direct PCR products were combined with 1 X NE buffer 4 and 4 units of the restriction endonuclease, *Dra* I, which recognizes the TTTAAA restriction site, in the 10 μl total reaction volume and incubated at 37 °C overnight. The total volumes of the RE reactions were mixed with 2 μl of gel loading buffer. Multiple bands of RFLP were sized by horizontal electrophoresis on 2.5% mini-agarose gel with 1 X TAE at 100 Volts for 50-55 min, then stained in ethidium bromide solution for 10 min and visualized on UV transilluminator. The SNP alleles were examined so they could be distinguished from the other SNP loci. There were 3 fragments (125, 119 and 25 bp) for the A allele and the pattern seemed to be one band on 2.5% gel (25-bp fragment was run out of gel). The distinct pattern of the T allele appeared clearly in two bands with 150-bp and 119-bp fragments (Table 5).

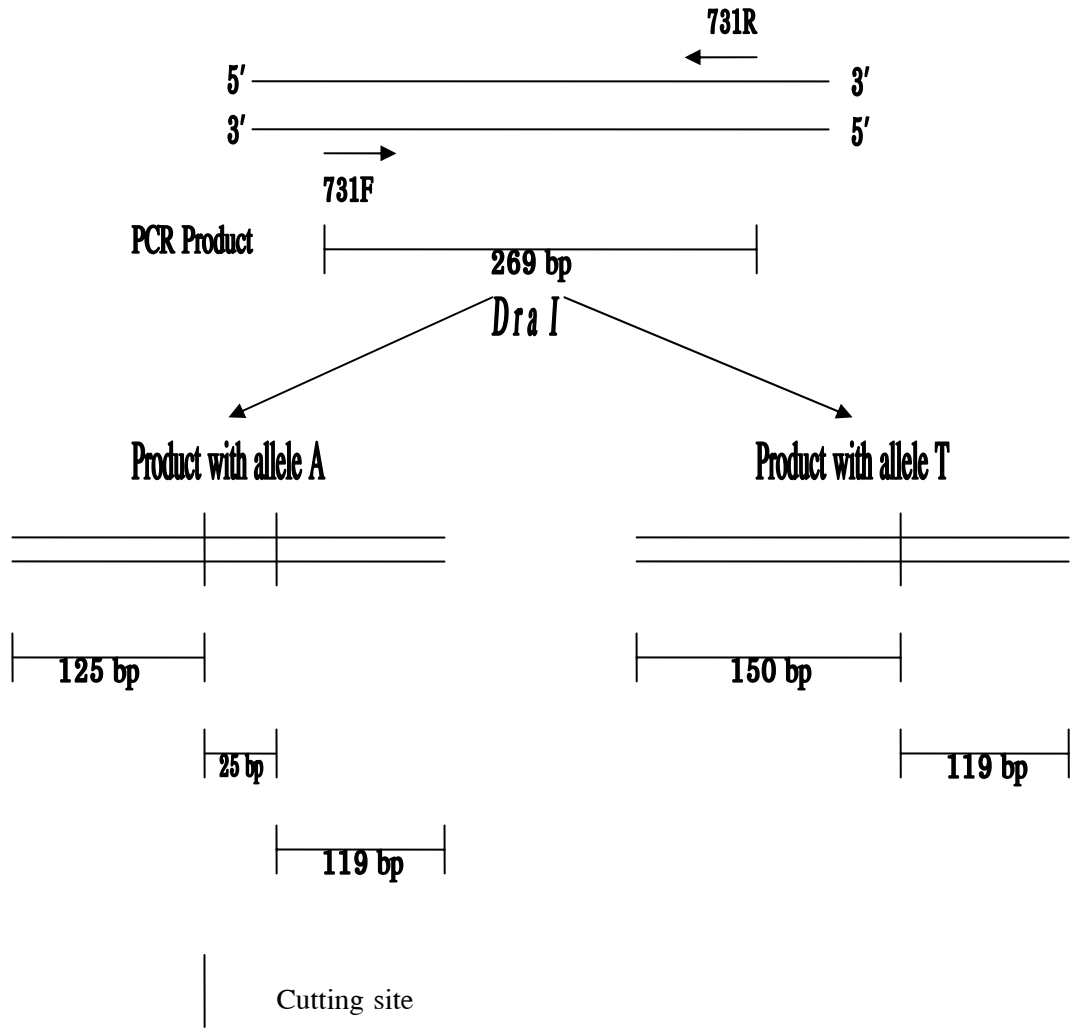

Figure 3. Schematic of PCR-RFLP for rs25731.

Table 5. The conditions and product sizes of PCR.

Locus	Primer		$MgCl_2$ Conc. (mM)	Adjuvant (Conc.)	Annealing temperature (°C)	Product size (bp)
	Name	Conc. (μM)				
WEX5	WEX5F	0.1	1.25	-	60	Control (361)
	WEX5R	0.1				
	WEX5CF	1				C allele (143)
	WEX5GR	3				G allele (261)
ATL1	ATL1F	0.1	1.5	BSA	53	Control (302)
	ATL1R	0.1		(0.13 mg/ml)		
	ATL1AF	3				A allele (107)
	ATL1GR	0.1				G allele (245)
rs25731*	731F	0.25	1.5	-	60	Product (269)
	731R	0.25				A allele (125, 119, 25)
						T allele (150, 119)
IVS10	IVS10F	0.1	1.5	BSA	53	Control (268)
	IVS10R	0.1		(0.13 mg/ml)		
	IVS10TF	3				T allele (189)
	IVS10CR	3				C allele (127)
rs25702	702F	0.15	1.5	BSA	55	Control (379)
	702R	0.15		(0.13 mg/ml)		
	702AF	3				A allele (148)
	702GR	2.8				G allele (274)
rs25723	723F	0.05	1.5	BSA	55	Control (330)
	723R	0.05		(0.13 mg/ml)		
	723AF	0.6				A allele (263)
	723CR	4				C allele (114)

* PCR-RFLP (*Dra* I restriction enzyme), other were determined by Biallelic-ARMS PCR.

Sequencing analysis of AGG interruptions

For accurate AGG organization, direct sequencing across the CGG repeat regions were performed with primer A and primer 571R in 50 μl reaction comprised of 50.25 mM Tris-HCl, pH 8.8, 12.45 mM $(NH_4)_2SO_4$, 1 mM $MgCl_2$, 200 μM dATP, dCTP, dTTP, 100 μM dGTP and 100 μM 7-deaza dGTP, 0.25 μM each primer, 10% DMSO, 128 μg/ml BSA, 2.5 units of Immolase DNA polymerase (Bioline) and 250 ng genomic DNA. The PCR programs using a MJ thermal cycler first ran at 95 °C for 9 min, followed by 35 cycles of 1 min each at 95 °C, 64 °C and 72 °C and a final extension of 72 °C for 10 min. The total volumes of PCR product were purified by a QIAquick PCR purification kit following the manufacturer's protocol (Appendix D). The 2 μl of purified PCR products were measured by electrophoresis on 1% agarose gel with standard marker of known concentration to estimate their quantities. Sequencing was carried out in 10 μl reactions consisting of a 1X BigDye terminator v1.1 ready reaction premix, 1.6 μM of the internal sequencing primer FXS-SEQF for forward reactions of higher CGG numbers (≥37 repeats) or primer 571R for reverse reactions of low-to-middle CGG numbers (≤36 repeats) as well as 10-20 ng purified PCR product. Cycle sequencing reactions were performed in a GeneAmp PCR System 9700 thermal cycler with a temperature profile of 96 °C for 1 min followed by 25 cycles of 96 °C for 10 sec, and 60 °C for 4 min. After that the sequencing products were purified again to remove unincorporated fluorescent dye terminators using a DyeEx 2.0 spin kit according to the manufacturer's instructions (Appendix D). All sequencing pellets were dissolved with 15 μl TSR (template suppressor reagent) and separated by an ABI PRISM 310 Genetic Analyzer with 310 Collection Software. We carried through some portion of samples in all processes. The remaining samples were prepared as purified PCR products and sent to a commercial sequencing service (Macrogen, Korea). The resulting electropherograms of some samples were exhibited in Appendix B, Figures 11.1-11.6. Data analysis of the electropherograms was performed with Sequencing Analysis Software (to analyze the raw data), Autoassembler Software (to assemble the forward and reverse sequences together and build the consensus sequence) and Sequence Navigator Software (to compare the consensus sequence with the reference sequence). The numbers of AGG interruptions within the CGG repeat length were designated as the A character representing AGG interspersion and the number representing CGG repeats [i.e. 9A9A9 = $(CGG)_9AGG(CGG)_9AGG(CGG)_9$].

Sequencing confirmation of microsatellites and SNPs

In order to verify the precision of the microsatellite and SNP alleles, the AC repeat region of microsatellite and the nucleotide at SNP site were sequenced with the same set of microsatellites forward and reverse primers and the particular control forward and reverse primers of that SNP, respectively. The 50 µl PCR reactions contained 1X PCR buffer II, 1.5 mM $MgCl_2$, 200 µM dNTP mix, 0.25 µM of each primer, 2.5 units AmpliTaq Gold DNA polymerase and 250 ng genomic DNA. The reaction tubes were put in a GeneAmp PCR system 9700 thermal cycler programmed of 10 min at 95 °C, 35 cycles of 30 sec each at 95 °C, proper annealing temperature (Table 5) and 30 sec at 72 °C along with finally 10 min at 72 °C. The further processes were carried out the same as the sequencing procedure above, however, we reduced the final concentrations of a BigDye terminator v1.1 ready reaction premix to 0.5X with a 0.5X BigDye terminator sequencing buffer. For forward reaction, we used only DXS548F strand and each SNP control forward primer in the sequencing reactions, whereas only FRAXAC1R strand and each SNP control reverse primer were used in sequencing reverse reactions. The quantity of the purified PCR products was slightly changed to 3-10 ng. The additional annealing temperature of 50 °C for 5 sec was included in the temperature profiles of cycle sequencing. Data analysis of these electropherograms was carried out with the software that mentioned above.

Data analysis

The following steps were performed, and the resulting data entered into the Excel program.

- We collected the total amount of each allele in the individual markers.
- We combined these polymorphic markers to create 2 series of haplotype containing haplotype with only microsatellites markers as well as mixed haplotypes between microsatellite and SNP markers.
- We gathered the total number of possible haplotype in each haplotype set of normal and FXS group along with subgroups of controls.
- We calculated the percentage, the heterozygosity and chi-square test with the level of significance set at $P < 0.05$.
- We analyzed and compared the data between normal and FXS groups as well as between subgroups of controls and FXS subjects.