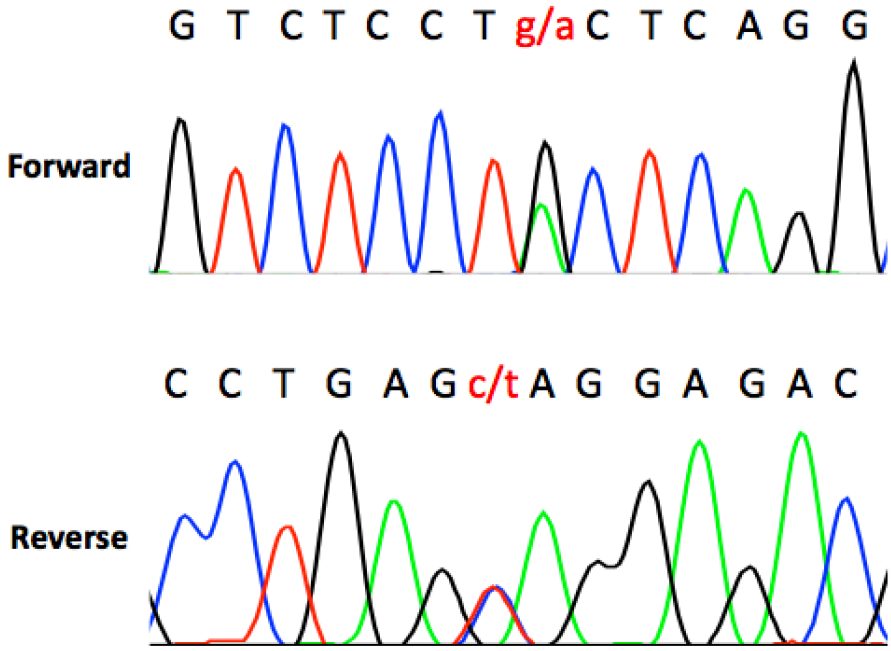
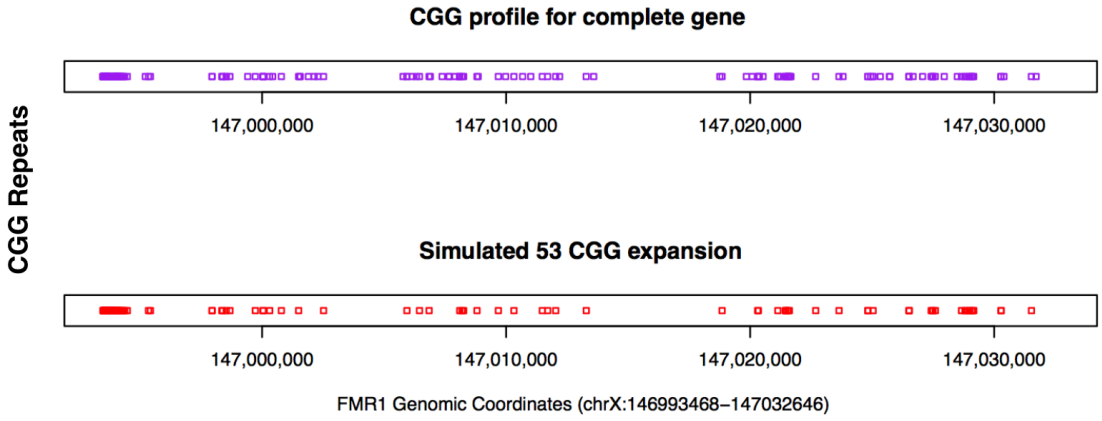
Supplemental Figures

Supplemental Figure 1.

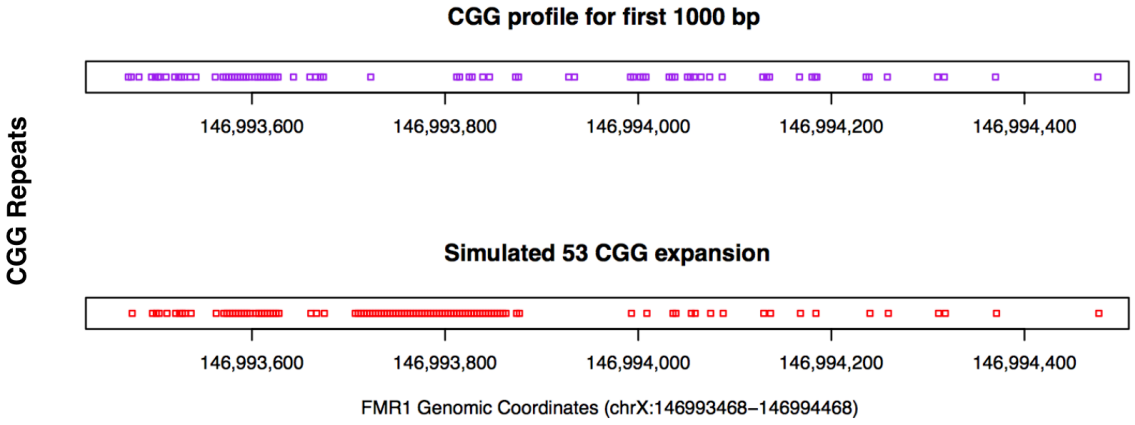


Supplemental Figure 2.

(a)

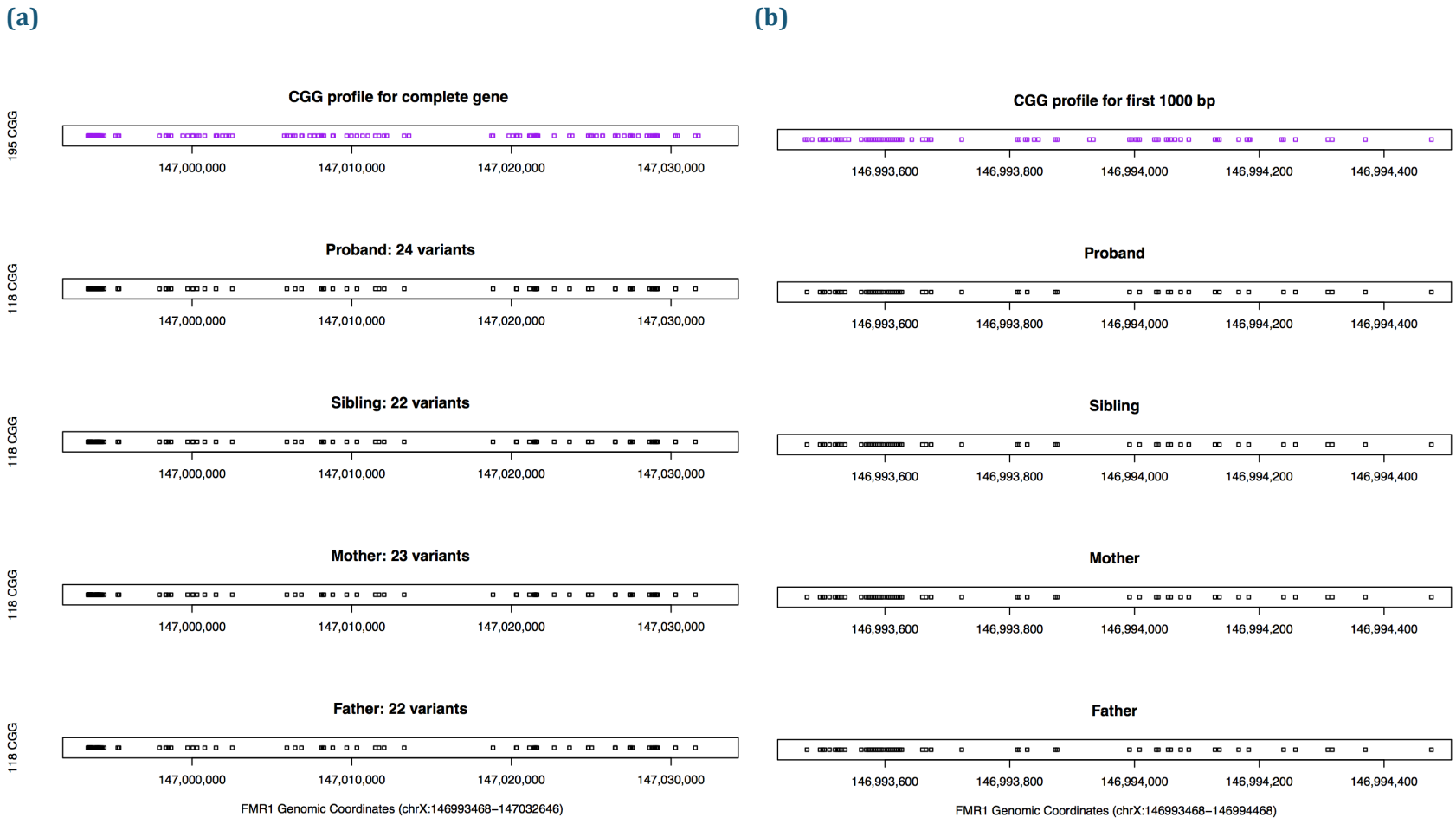


(b)

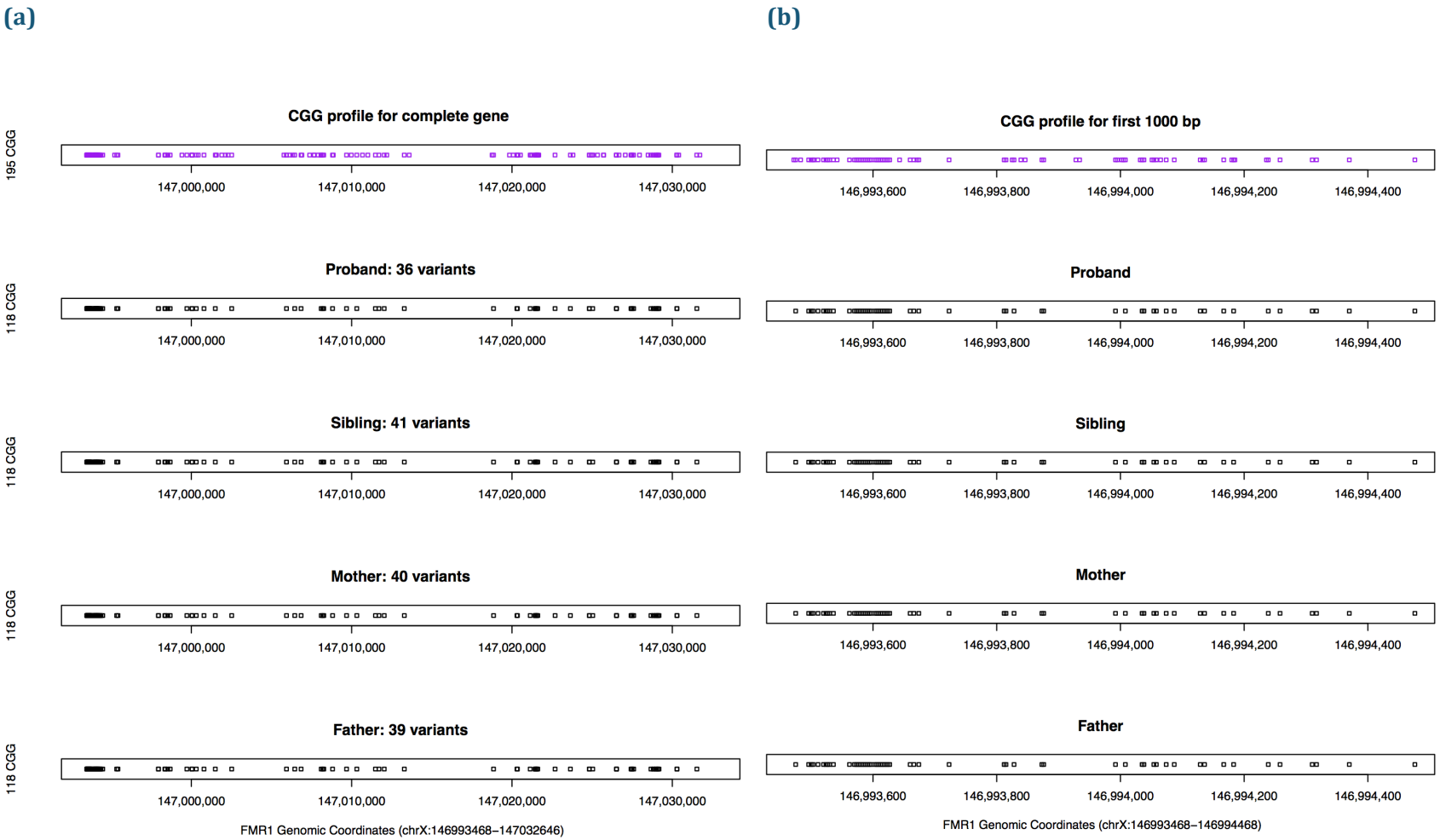

CGG profile for first 1000 bp
CGG Repeats
146,993,600
146,993,800
146,994,000
146,994,200
146,994,400
Simulated 53 CGG expansion
146,993,600
146,993,800
146,994,000
146,994,200
146,994,400
FMR1 Genomic Coordinates (chrX:146993468–146994468)

Supplemental Figure 3.

Supplemental Figure 4.



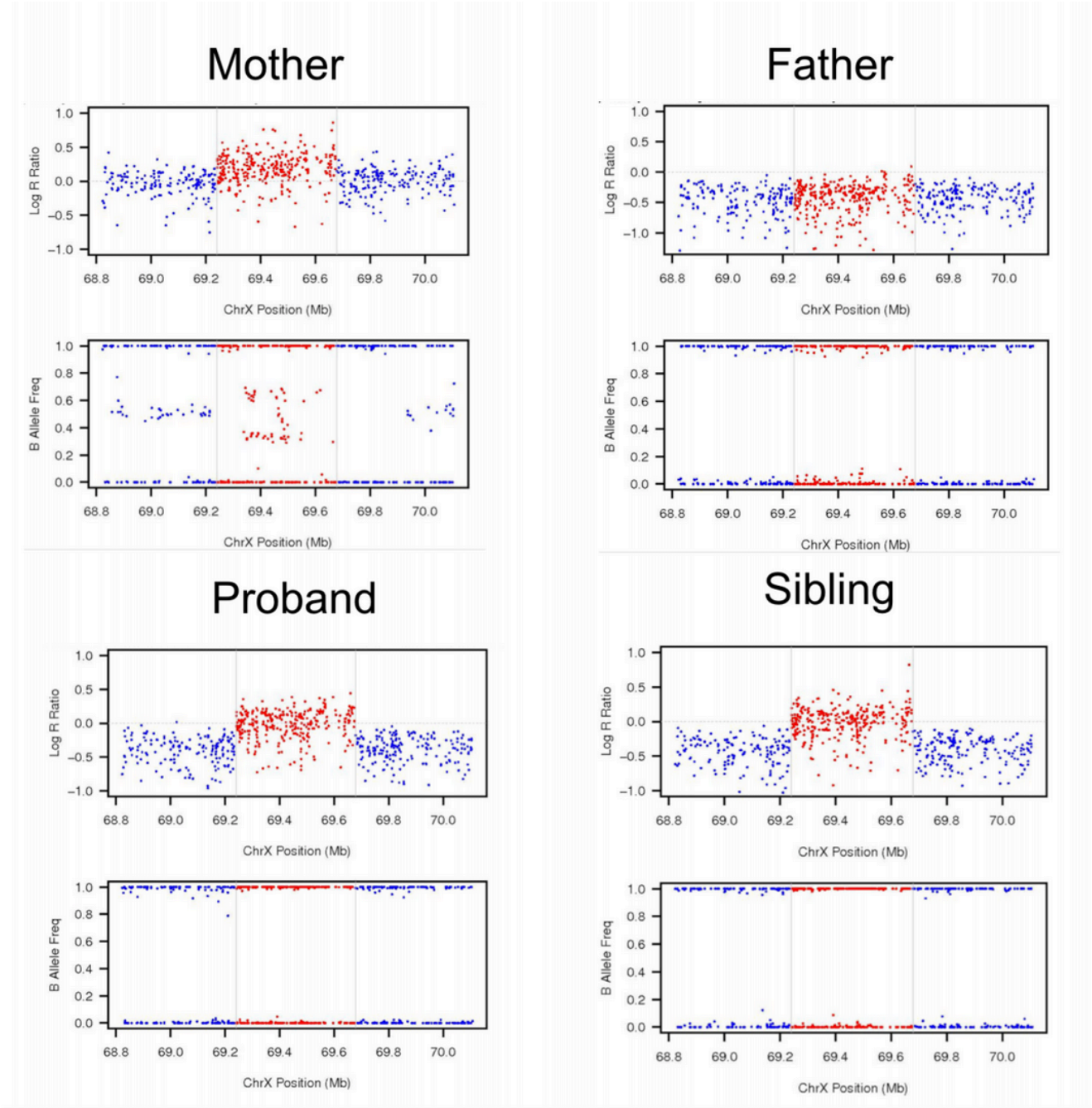
Supplemental Figure 5.



Supplemental Figure 6

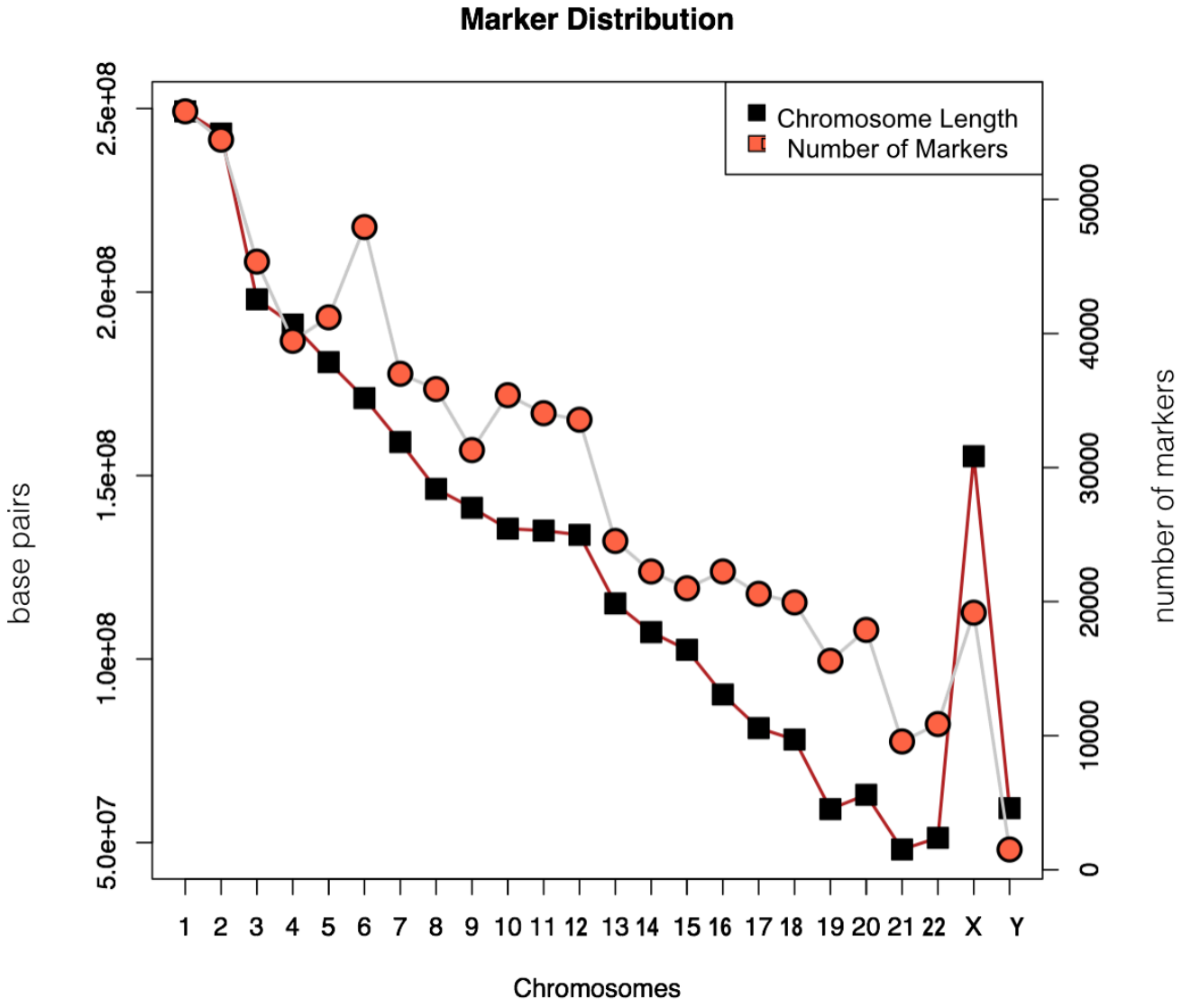
(a)

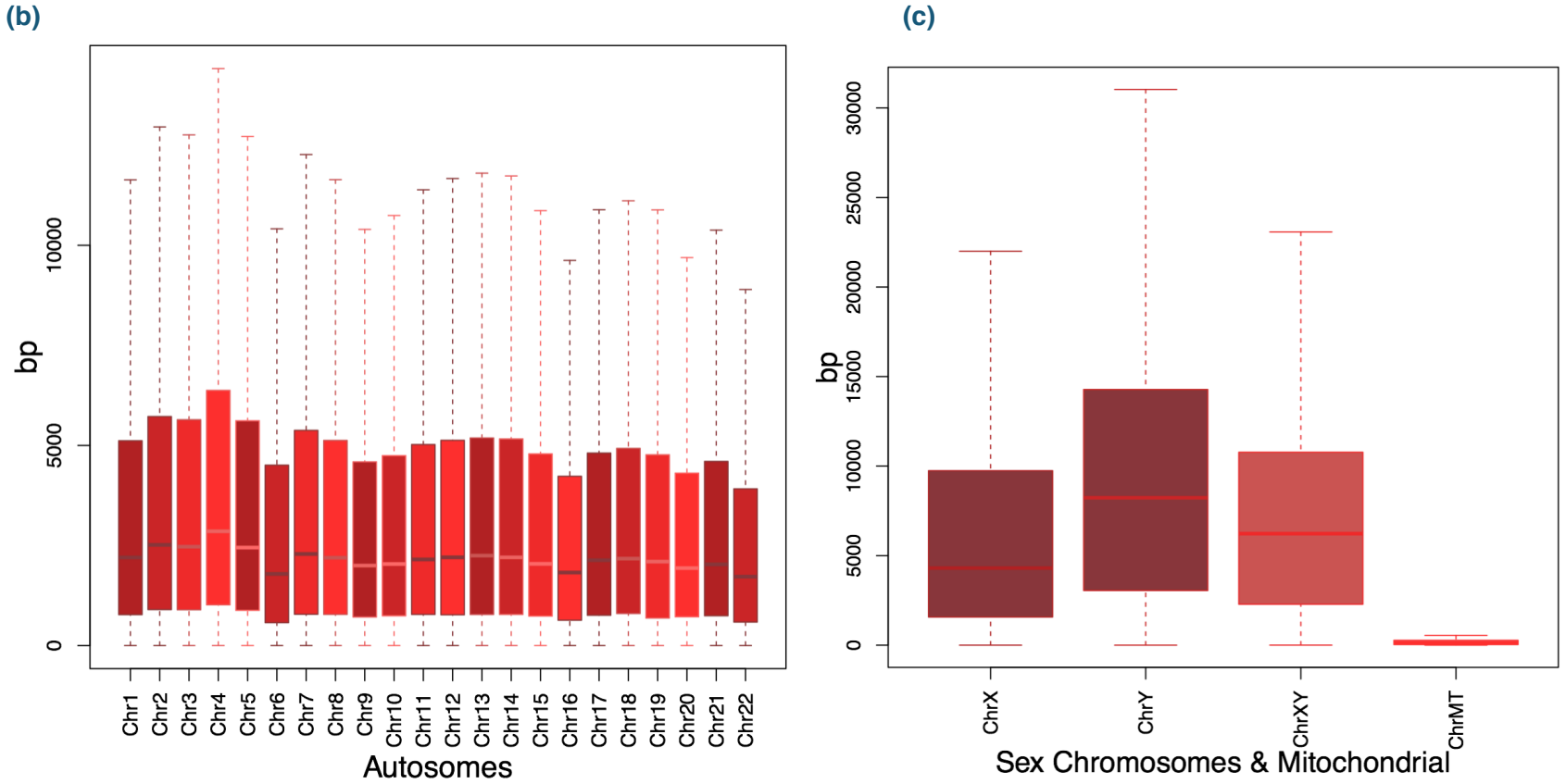
(b)


Mother
Log R Ratio
1.0
0.5
0.0
−0.5
−1.0
68.8
69.0
69.2
69.4
69.6
69.8
70.0
ChrX Position (Mb)
B Allele Freq
1.0
0.8
0.6
0.4
0.2
0.0
68.8
69.0
69.2
69.4
69.6
69.8
70.0
ChrX Position (Mb)
Father
Log R Ratio
1.0
0.5
0.0
−0.5
−1.0
68.8
69.0
69.2
69.4
69.6
69.8
70.0
ChrX Position (Mb)
B Allele Freq
1.0
0.8
0.6
0.4
0.2
0.0
68.8
69.0
69.2
69.4
69.6
69.8
70.0
ChrX Position (Mb)
Proband
Log R Ratio
1.0
0.5
0.0
−0.5
−1.0
68.8
69.0
69.2
69.4
69.6
69.8
70.0
ChrX Position (Mb)
B Allele Freq
1.0
0.8
0.6
0.4
0.2
0.0
68.8
69.0
69.2
69.4
69.6
69.8
70.0
ChrX Position (Mb)
Sibling
Log R Ratio
1.0
0.5
0.0
−0.5
−1.0
68.8
69.0
69.2
69.4
69.6
69.8
70.0
ChrX Position (Mb)
B Allele Freq
1.0
0.8
0.6
0.4
0.2
0.0
68.8
69.0
69.2
69.4
69.6
69.8
70.0
ChrX Position (Mb)

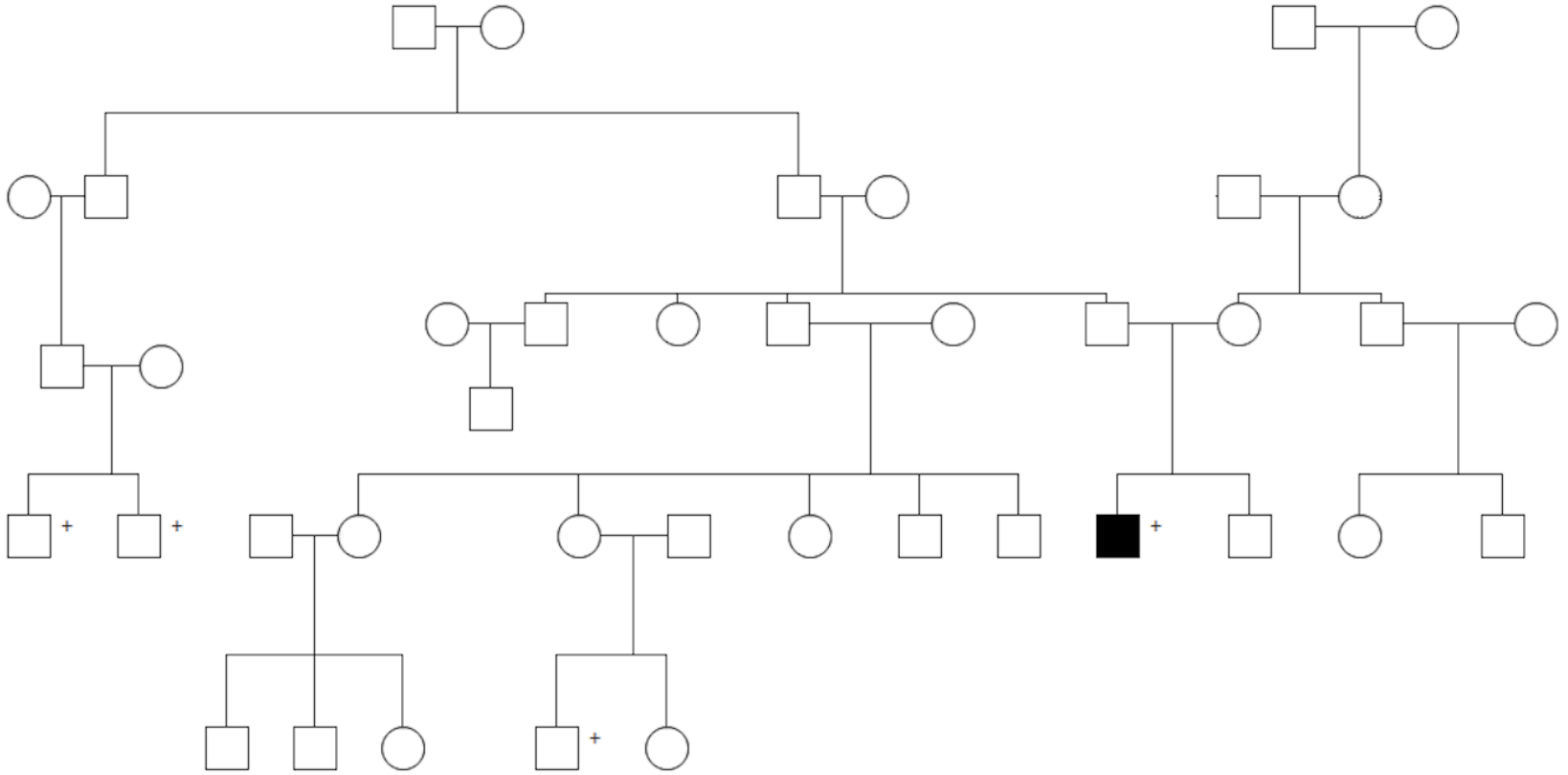
Supplemental Figure 7.

(a)




(b)
bp
0
5000
10000
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Chr8
Chr9
Chr10
Chr11
Chr12
Chr13
Chr14
Chr15
Chr16
Chr17
Chr18
Chr19
Chr20
Chr21
Chr22
Autosomes
(c)
bp
0
5000
10000
15000
20000
25000
30000
ChrX
ChrY
ChrXY
ChrMT
Sex Chromosomes & Mitochondrial

Supplemental Figure 8.



Supplemental Figure Legends

Supplemental Figure 1. *MYBBP1A* stop gain validation by Sanger sequencing. Sanger sequencing validation shows two overlapping peaks: one for C and one for T on the reverse strand.

Supplemental Figure 2. GGG repeats profile on the reference *FMR1* complete gene and a simulated expansion. (a) The x axis represent the coordinates of the reference *FMR1* gene, which includes the 5'UTR region where the CGG expansion occurs. **(b)** Here only the first 1000 closest nucleotides to the 5'UTR are plotted so a simulated expansion of randomly introduced CGG repeats is clearly appreciated.

Supplemental Figure 3. GGG repeats profile. Family K21 Complete *FMR1* region profile. (a)The complete FMR1 gene CGG profile for this family looks normal, the number of CCG repeats for the proband is even less than the reference **(b)** First 1000 closest nucleotides to the 5'UTR.

Supplemental Figure 4. GGG repeats profile for family SSC_12605. (a) The complete *FMR1* gene CGG profile for this family looks normal. **(b)** The first 1000 closest nucleotides to the 5'UTR look normal.

Supplemental Figure 5. GGG repeats profile for family SSC_12596. (a) The complete *FMR1* gene CGG profile for this family also looks normal. **(b)** The first 1000 closest nucleotides to the 5'UTR look normal.

Supplemental Figure 6 (a). Copy number variation on K21 detected by CMA, Illumina Omni2.5 array and WGS technologies. Here is the duplication region that was previously reported on the medical records which reported the following genes fully contained in the CNV: *OTUD6A/IGBP1/DGAT2L6/AWAT1/AWAT2/P2RY4/KIF4A/ARR3/GDPD2/RAB41/PDZD11* and the following genes partially contained in the CNV: *EDA/DLG3* which actually corresponds to the genes contained in the CNV region reported by ERDS. As the healthy male sibling also inherited. **(b)** Even though PennCNV did not detect the CNV, by plotting the LRR and BAF values for all the family, the CNV can be confirmed to be present in the mother, proband and healthy sibling.

Supp. Figure 7. Markers Number by chromosome and the inter-marker spacing. The recommended methods for filtering Copy Number Variants called from microarray data are too arbitrary in the sense they are not aware of some features array-specific that could make this general filtering criteria suitable for all microarray calls. The number of markers and the space between them was evaluated for each chromosome and considered into the CNVs filtering criteria. **(a)** Here, the size of the chromosomes in base pairs (black) and the number of markers (red) are plotted, a line is drown across the dots so it's easier to see that the greater the chromosome size, the greater the number of markers. However, the relation is not perfect and the number of markers for similar sized chromosomes, can vary largely. Because of this, the expected number of SNPs involved in a CNV call from one chromosome has to differ from those called in another chromosome. **(b)** However, not only the number of markers plays an important role but it was also important to make sure that the marker distribution across each chromosome was homogeneous without clusters of markers and large empty regions. If the size in base pairs of each chromosome was divided between the number of markers, the average inter-marker spacing should be 5Kb, to know if this was true, all the values for the spaces between markers (without outliers) are plotted here as quartiles, showing that 3/4 of the spacing values are around 5 or 6Kb and the other quarter having inter-marker spaces up to 10Kb, which

explains why some regions are more difficult to call. **(c)** The sex chromosomes are plotted separately as their upper quartile has greater values than the autosomes. On the other hand, the Mitochondrial chromosome inter-marker spacing is smaller, this make sense as its size is only of ~16Kb and the Illumina Omni2.5 microarray has 288 markers for it.

Supp. Figure 8. Extended Pedigree K21. The black square represents the K21 proband analyzed in this study. The individuals marked with a "+" sign have been diagnosed from left to right with: + Pervasive developmental disorder/overweight/blood pressure problems, + autism and + autism/ Hypophosphatasia/Chiari malformation.