The American Journal of Human Genetics, Volume *104*

Supplemental Data

GGC Repeat Expansion and Exon 1 Methylation of *XYLT1* Is a Common Pathogenic Variant in Baratela-Scott Syndrome

Amy J. LaCroix, Deborah Stabley, Rebecca Sahraoui, Margaret P. Adam, Michele Mehaffey, Kelly Kernan, Candace T. Myers, Carrie Fagerstrom, George Anadiotis, Yassmine M. Akkari, Katherine M. Robbins, Karen W. Gripp, Wagner A.R. Baratela, Michael B. Bober, Angela L. Duker, Dan Doherty, Jennifer C. Dempsey, Daniel G. Miller, Martin Kircher, Michael J. Bamshad, Deborah A. Nickerson, University of Washington Center for Mendelian Genomics, Heather C. Mefford, and Katia Sol-Church

Figure S1

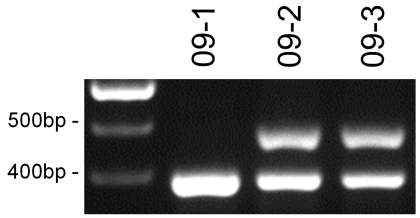

Figure S1: Splice variant deletes *XYLT1* exon 6 in Family 09. RT-PCR of WBC cDNA from Family 09. The proband 09-1 is homozygous for the splice site variant c.1290-1G>A, while each parent (09-2 and 09-3) is heterozygous. Primers spanning from exon 5 to exon 7 of *XYLT1* show a smaller product in addition to the 470bp wild-type product. Sanger sequencing (not shown) confirms that the variant causes exon 6 (81bp) to be skipped.

Figure S2

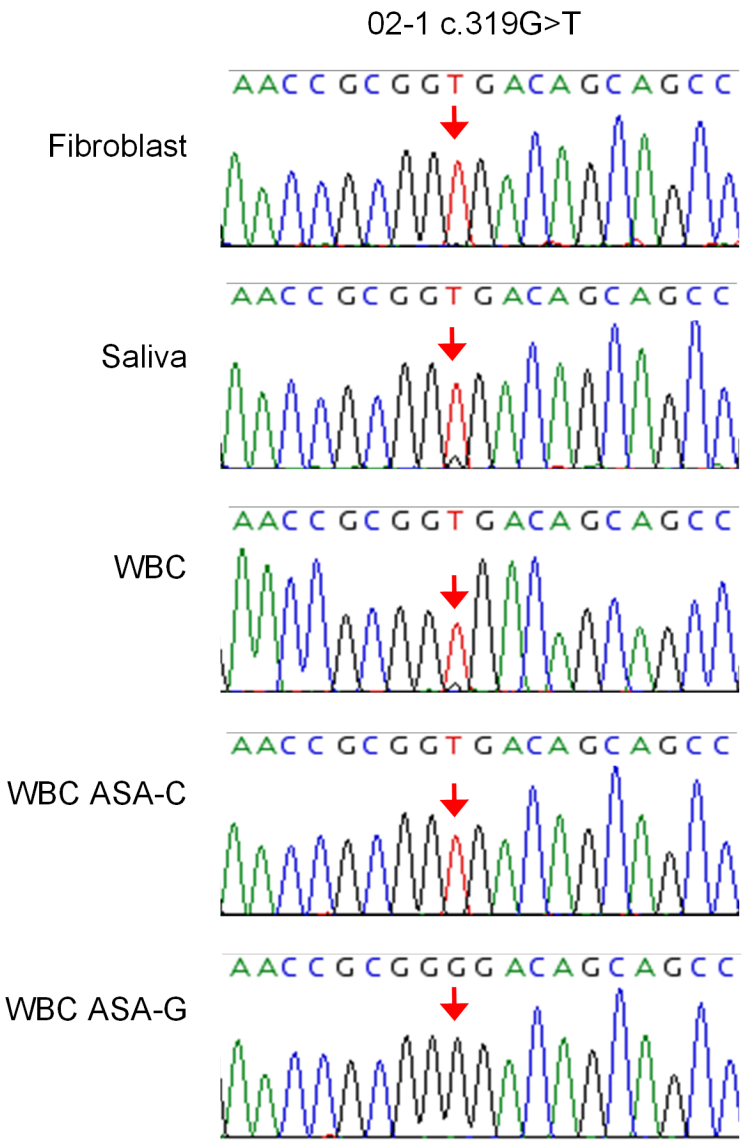

Figure S2: Allele specific amplification for 02-1. Sequencing of genomic DNA from multiple tissue types of proband 02-1 shows the small wild-type peak seen in saliva and WBC at position c.319G>T, suggesting allele drop-out. Allele specific amplification against SNP rs117041807 C/G in WBC DNA confirms variant c.319G>T is in fact heterozygous.

Figure S3

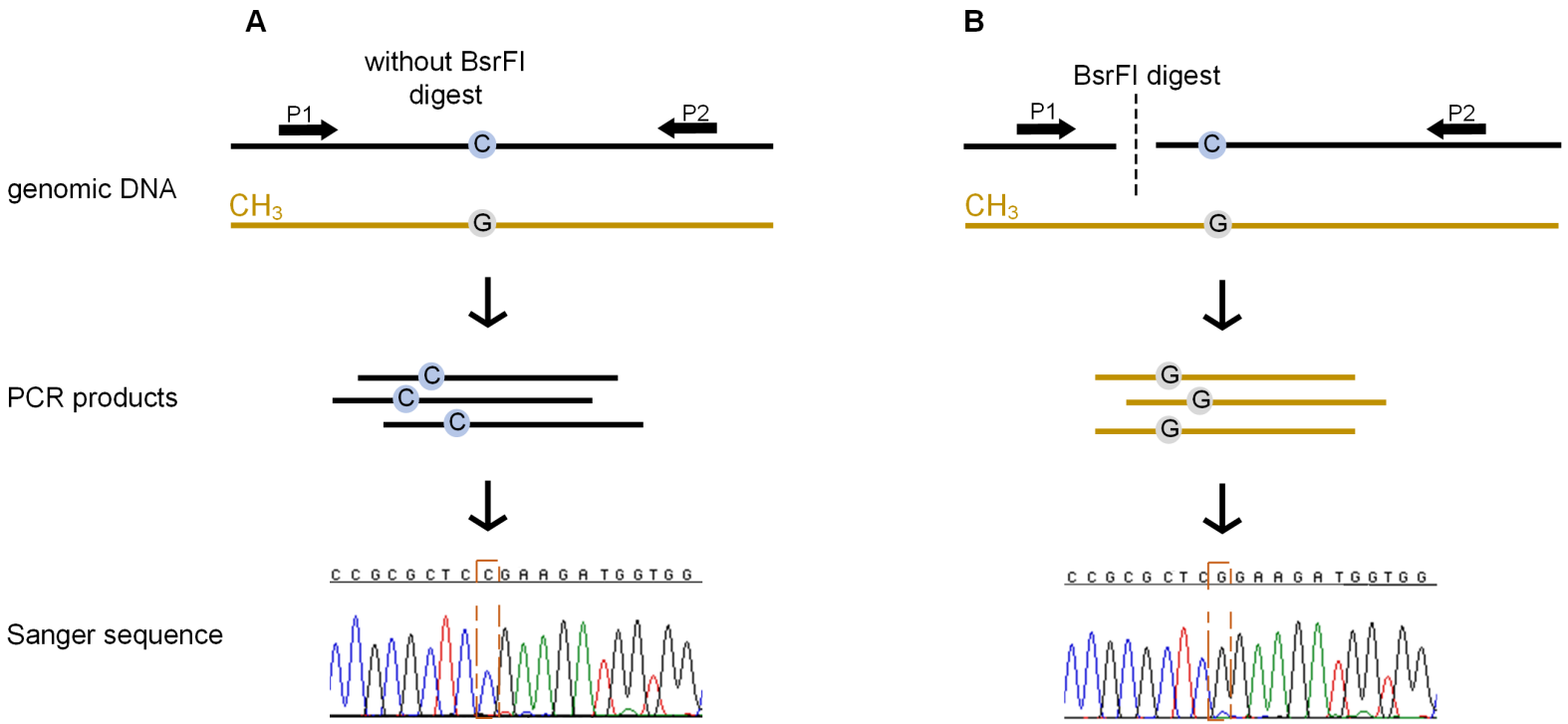

Figure S3: BsrFI digest allows detection of methylated allele
Schematic of PCR to amplify *XYLT1* SNP rs117041807 (NM_022166.3:c.-5C>G) for individual 05-2. (A) PCR of undigested genomic DNA preferentially amplified the unmethylated allele, showing only a single C allele at the site of the SNP. (B) Following digestion with BsrFI, a methylation sensitive restriction enzyme, a BsrFI site between the primers disrupted amplification of the unmethylated allele, allowing detection of the methylated G allele that was previously "hidden".

Figure S4

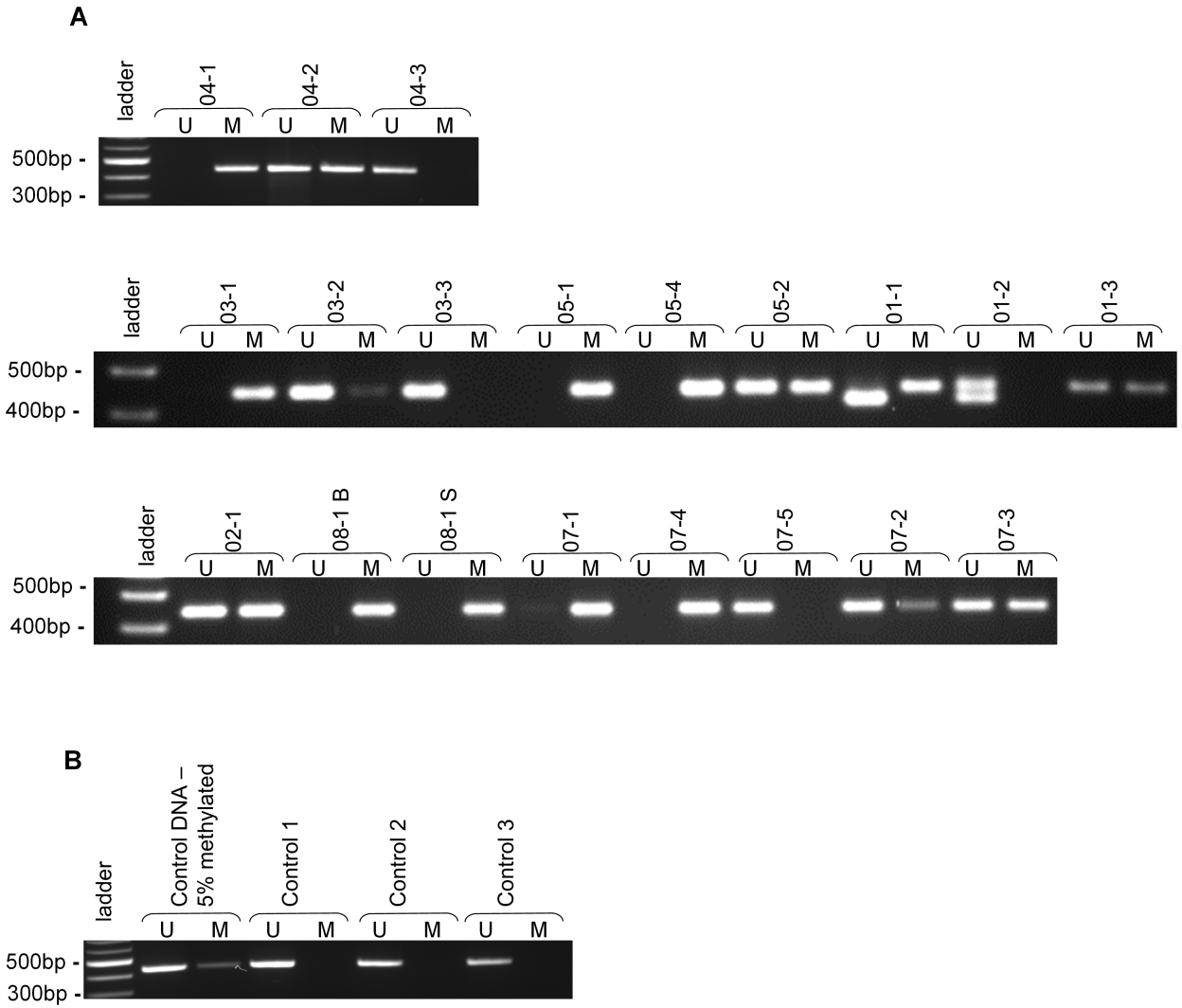

Figure S4: Representative results of MS-PCR. MS-PCR was performed on bisulfite treated genomic DNA using primers specific for unmethylated (U) or methylated (M) DNA encompassing the majority of exon 1. **(A)** MS-PCR results for members of our BSS cohort. DNA from both whole blood (B) and saliva (S) were tested for proband 08-1. Similar findings were seen in family 06 (not pictured). The smaller size of the unmethylated product in individual 01-1 (and 01-2) is due to a 26-bp deletion in exon 1. **(B)** Representative results for unaffected controls, where only unmethylated DNA is detected; an internal control of commercially purchased fully methylated genomic DNA diluted to 5% with fully unmethylated genomic DNA was included in each experiment.

Figure S5

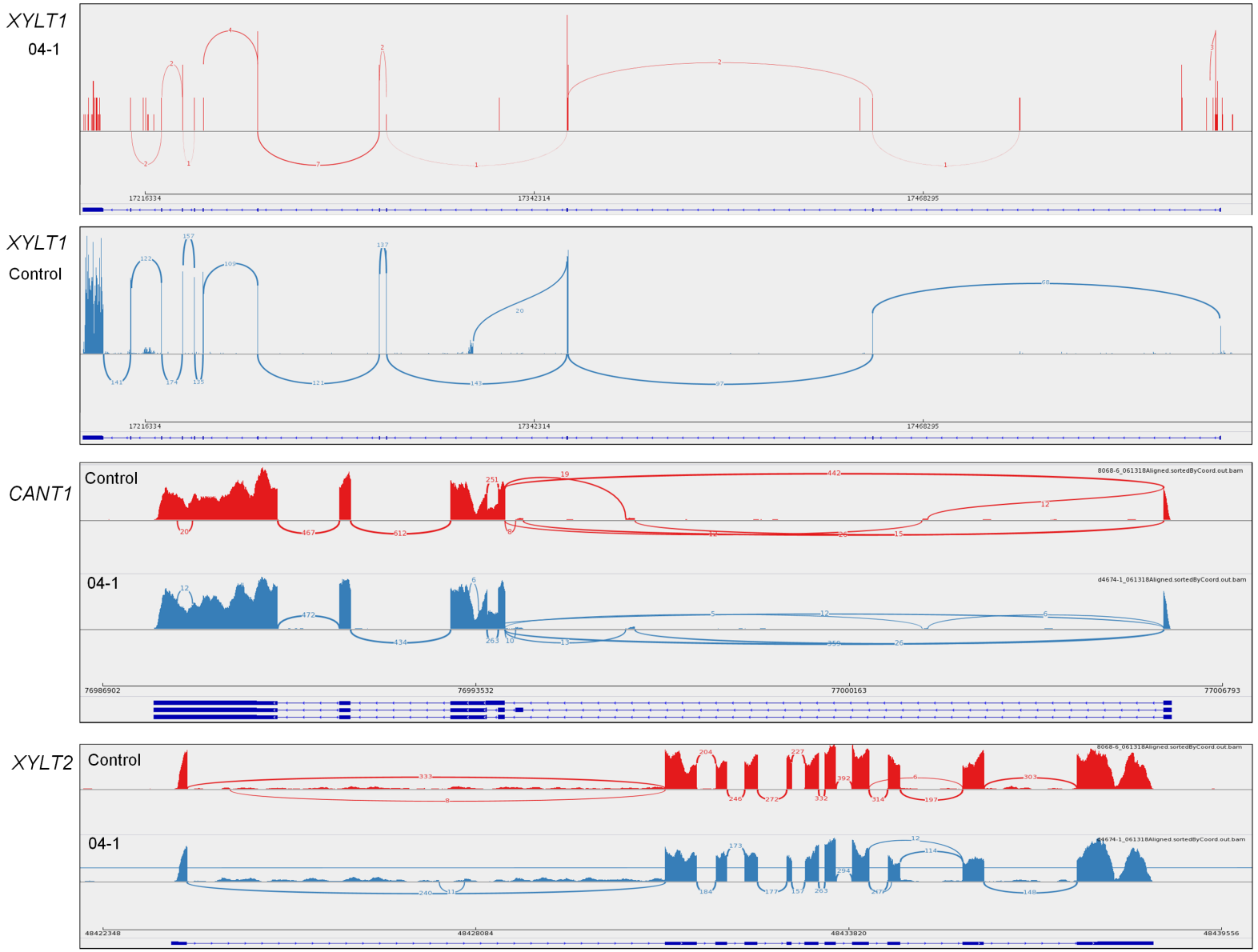

Figure S5: RNA-seq analysis of *XYLT1* expression in 04-1 fibroblasts. Sashimi Plot of RNA-seq data. *XYLT1* exon and exon junction coverage is greatly reduced for 04-1 and no full-length transcript is present as exon 1 is not transcribed. In comparison, coverage for *CANT1*, the causative gene for DBQD-I, and *XYLT2* is comparable over exons and exon junctions in both 04-1 and the control.

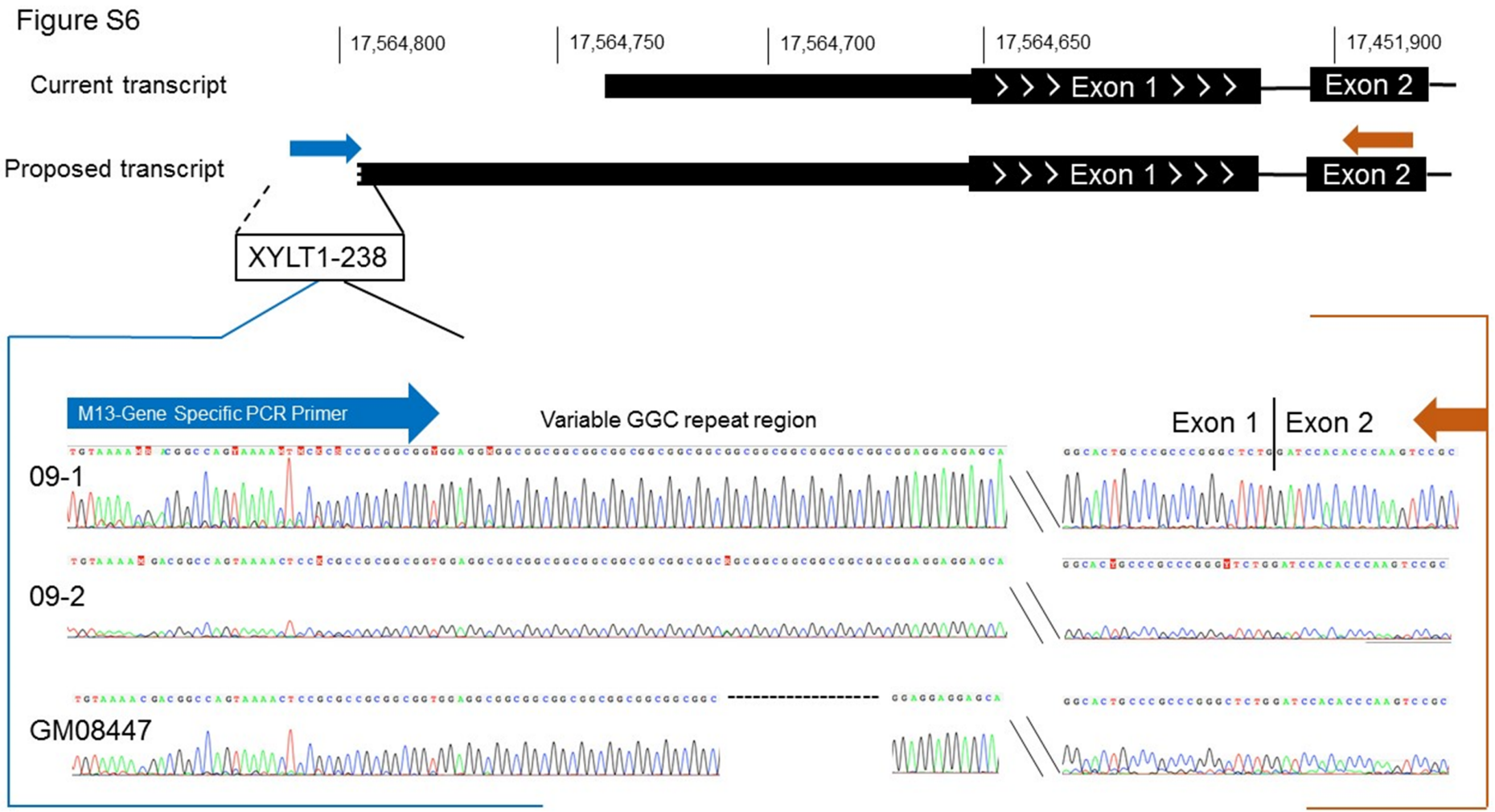

Figure S6: RT-PCR of mRNA suggests unannotated transcriptional start for *XYLT1*. RT-PCR was performed using RNA isolated from 09-1 (WBC), 09-2 (WBC) and a control (GM08447 fibroblasts). A primer pair spanning from upstream of the variable GGC repeat region, within XYLT1-238, into exon 2 produced a PCR product, suggesting a transcript start site further upstream than the current annotation. The current transcript and genomic coordinates are notated at the top (GRCh37/hg19, NM_022166). The proposed transcript is below with the approximate start of XYLT1-238 noted. Electropherograms show the M13 tailed gene specific primer upstream of the variable GGC repeat region and the exon 1/2 boundary, confirming the product is not genomic DNA. Sequencing of genomic DNA from all individuals (not shown) shows the GGC repeat region was within the normal range of 9-20 repeats previously reported.

Figure S7

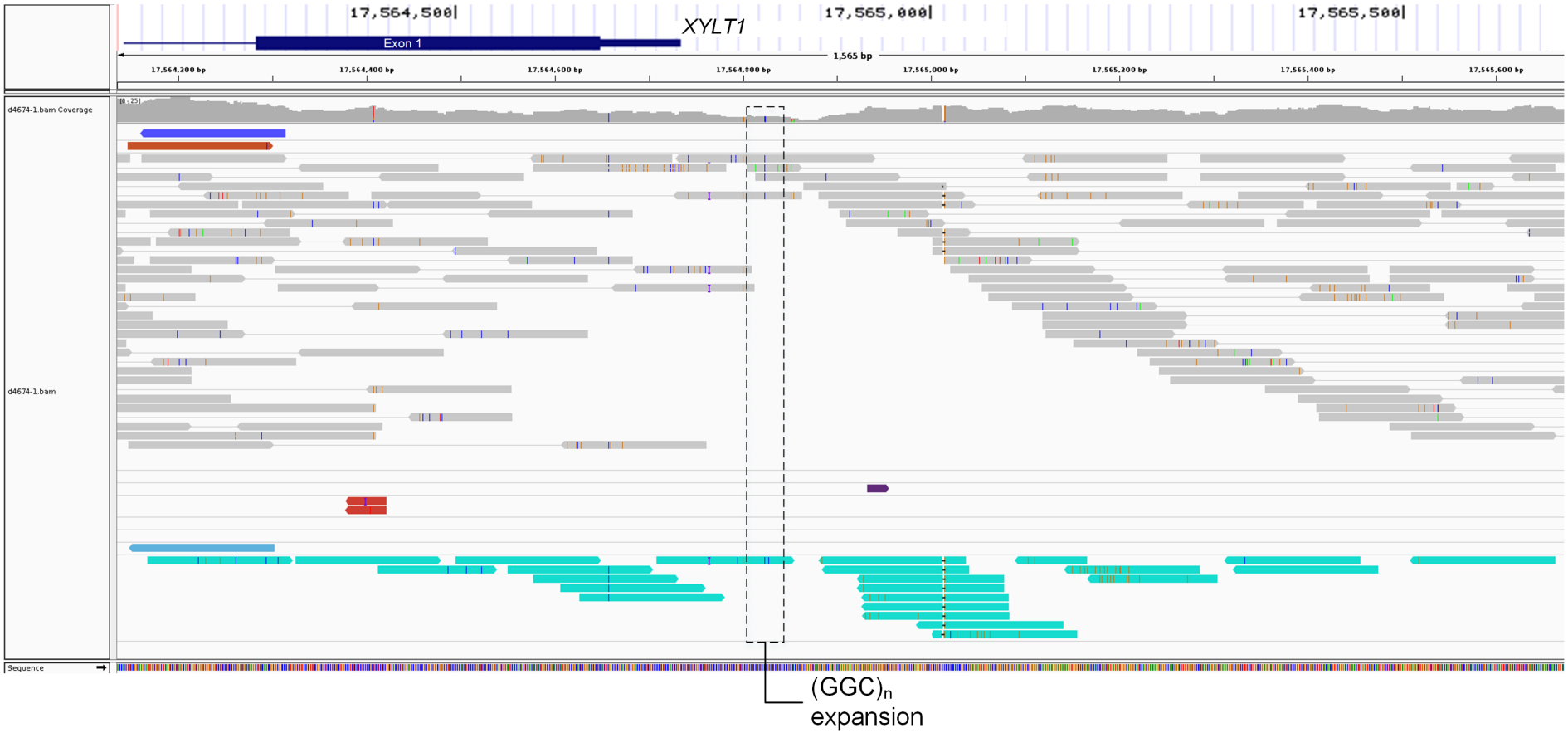

Figure S7: STRetch analysis of genome sequence. Image from Integrated Genomics Viewer (IGV) for 04-1 genome sequence data of exon 1 and the promoter region of *XYLT1*, including XYLT1-238. Connected paired-end reads that align to the reference genome are shown in grey. No read pair spans the area of the expansion site. Reads whose mate maps to a GGC repeat decoy sequence are shown in aquamarine, all of which are displayed. All such reads approach the expansion site, indicating that the missing sequence contains a large GGC repeat.

Figure S8

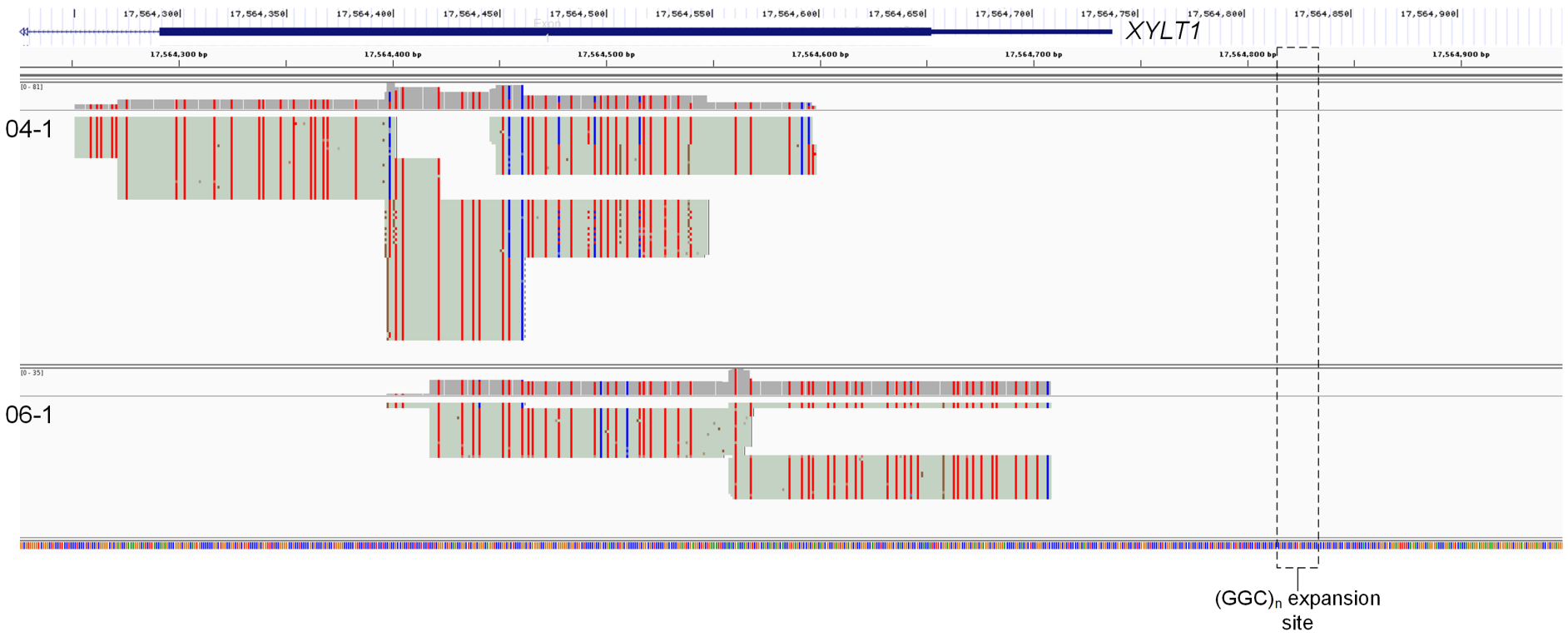

Figure S8: Whole genome bisulfite sequencing. Image from Integrated Genomics Viewer (IGV) for 04-1 and 06-1 whole genome bisulfite sequence data of exon 1 and the promoter region of *XYLT1* with tracks colored using 'CG bisulfite mode'. Reads have been collapsed for visualization. Shown in sage are reads that align to the altered reference genome containing XYLT1-238. A red nucleotide denotes a nonconverted cytosine (methyl protected), while a blue nucleotide denotes a bisulfite converted cytosine (unmethylated).

Table S1: Previously reported causal variants for DBQD-II in *XYLT1*. Reported variants for 16 individuals from 13 families with DBQD-II. All variants are found in or containing *XYLT1*, though only one case has a variant in exon 1, juxtaposed with our cohort. Abbreviations include: Homo, homozygous; Het, heterozygous; NR, not reported; Consang, consanguineous.

	Ethnicity	Consang	Sex	cDNA	Protein	Location	Status	Reference
1	Tunisian	+	Female	c.1792C>T	p.(Arg598Cys)	Exon 9	Homo	Bui et al., 2014 AJHG
			Male					
2	Mauritian	+	Female	c.439C>T	p.(Arg147Ter)	Exon 3	Homo	Bui et al., 2014 AJHG
3	Belgian	+	Male	c.276dupG	p.(Pro93AlafsTer69)	Exon 1	Homo	Bui et al., 2014 AJHG
4	Turkish	+	NR	c.1588-3C>T	Unknown	Intron 7	Homo	Bui et al., 2014 AJHG
5	Turkish	+	NR	c.1290-2A>C	Unknown	Intron 5	Homo	Bui et al., 2014 AJHG
6	Turkish	+	NR	c.1290-2A>C	Unknown	Intron 5	Homo	Bui et al., 2014 AJHG
7	Turkish	+	Female	c.1441C>T	p.(Arg481Trp)	Exon 7	Homo	Schreml et al., 2014 Hum Genet
			Male					
8	Polish	–	Male	c.595C>T	p.(Gln199Ter)	Exon 3	Het	Jamsheer et al., 2016 J Hum Genet
				c.1651C>T	p.(Arg551Cys)	Exon 8	Het	
9	Brazilian	+	Female	c.1651C>T	p.(Arg551Cys)	Exon 8	Homo	Silveira et al., 2016 AJMG
10	Turkish	+	Female	c.1792delC	p.(Arg598AlafsTer7)	Exon 9	Homo	Guo et al., 2017 J Hum Genet
			Female					
11	Turkish	–	Female	c.1290-2A>C	Unknown	Intron 5	Homo	Guo et al., 2017 J Hum Genet
12	Emirati	+	Male	c.2169dupA	p.(Val724SerfsTer10)	Exon 10	Homo	Al-Jezawi et al., 2017 AJMG
13	Dutch	-	Male	16p13 del (3.3 Mb)	Absent	Whole gene	Het	Van Koningsbruggen et al., 2016 AJMG
				c.1588-10_1595del	Unknown	Intron 7-Exon 8	Het	

Table S2: Southern blot probe generation primers. Primer sequences used to create the probe for Southern blot analysis, located in *XYLT1* Intron 1.

	Forward sequence 5' – 3'	**Reverse sequence 5' – 3'**	**Genomic coordinates**	**Product size**
Southern probe	GGGAGACGGCAAGGTTAGAG	CTCTGCAAGCCTACCGACTC	16: 17,563,659 - 17,564,191	533bp

Table S3: Methylation specific primers for *XYLT1*. Primer sequences used to perform Methylation-specific PCR. Primer sequences for both a region upstream (5') of the expansion and exon 2 are provided as an additional reference (data not shown).

Target	Methyl state	Forward sequence 5' – 3'	Reverse sequence 5' – 3'	Genomic coordinates	Product size
Exon 1	Unmethylated	TGTTGTTGTTGTAGATGTTGGTT	CCACCCAACTTCCCAAAACAAAAAACAATATAAAATCA	16:17,564,140 – 17,564,583	444bp
Exon 1	Methylated	TGTTGTTGTTGTAGACGTTGGTC	CCACCCGACTTCCCGAAACAAAAAACGATATAAAATCG	16:17,564,140 – 17,564,583	444bp
Upstream of expansion site	Unmethylated	TTGTGAAGGTTGAGTAATTTTATGGGTTTTTTTTTTGTT	CATAAAAAAAACTCCTAACCCCAAAATTCACA	16:17,565,129 – 17,565,486	358bp
Upstream of expansion site	Methylated	TTGCGAAGGTTGAGTAATTTTACGGGTTTTTTTTTGTC	CGTAAAAAAAACTCCTAACCCCGAAATTCGCG	16:17,565,129 – 17,565,486	358bp
Exon 2	Unmethylated	TTATGTTATTGTTTTTTAGTTTGGGTAATAAAGT	TCTCCACAAATAAAAAAAACACCAAACACCTAAATACT	16:17,451,624 – 17,452,051	428bp
Exon 2	Methylated	TTATGTTATTGTTTTTTAGTTTGGGTAATAAAGC	TCCCCACAAATAAAAAAAACACCAAACACCTAAATACC	16:17,451,624 – 17,452,051	428bp

Table S4: BSS Cohort Variants. Variant information corresponds to Homo sapiens xylosyltransferase I (XYLT1-001), GRCh37/hg19, NM_022166.3; CH3, methylated allele

Sample ID	Variant 1	Variant 2	Relationship	Status
04-1	16p13 del	CH3	proband	affected
04-2	-	CH3	mother	unaffected
04-3	-	-	father	unaffected
06-1	16p13 del	CH3	proband	affected
06-2	-	CH3	mother	unaffected
06-3	16p13 del	-	father	unaffected
06-4	16p13 del	-	sibling	unaffected
06-5	16p13 del	-	sibling	unaffected
03-1	16p13 del	CH3	proband	affected
03-2	-	CH3	mother	unaffected
03-3	16p13 del	-	father	unaffected
02-1	c.319G>T, p.Gly107Ter	CH3	proband	affected
08-1	CH3	CH3	proband	affected
01-1	c.281_306del, p.Gln94ArgfsTer59	CH3	proband	affected
01-2	c.281_306del, p.Gln94ArgfsTer59	-	mother	unaffected
01-3	-	CH3	father	unaffected
05-1	16p13 del	CH3	proband	affected
05-2	-	CH3	mother	unaffected
05-4	16p13 del	CH3	sibling	affected
07-1	CH3	CH3	proband	affected
07-2	-	CH3	mother	unaffected
07-3	-	CH3	father	unaffected
07-4	CH3	CH3	sibling	affected
07-5	-	-	sibling	unaffected
09-1	c.1290-1G>A, r.1375_1455del	c.1290-1G>A, r.1375_1455del	proband	affected
09-2	c.1290-1G>A,	-	mother	unaffected

	r.1375_1455del			
09-3	c.1290-1G>A, r.1375_1455del	-	father	unaffected
09-4	c.1290-1G>A, r.1375_1455del	-	sibling	unaffected
09-5	-	-	maternal half-sibling	unaffected
10-1	c.1730_1733dup, p.Asp578GlufsTer2	c.1730_1733dup, p.Asp578GlufsTer2	proband	affected

Table S5: RT-PCR primers used to examine the transcription start site of *XYLT1*. Primer sequences used in preliminary experiments to examine the true transcriptional start of *XYLT1*. Internal sequencing primers were used due to difficulty sequencing through the region.

Primer Name	**Sequence 5' – 3'**
M13 gene-specific primer; Forward	TGTAAAACGACGGCCAGTAAAACTCCGCGCCGCGGCGGT
XYLT1-Exon 2; Reverse	AGTCTCCAGGGTGATGAGCGGA
Internal seq; XYLT1-Exon 1; Reverse	TTCCACACGACCAGCGTCTGC
Internal seq; XYLT1-Exon 1; Forward	AGCGCGGGGGCGGCCCGGAGCGT
Internal seq; XYLT1-Exon 1; Reverse	CTGCAGCCGGCTCGGCGGGCAGGTC