MAPPING THE 5' END OF THE *XYLT1* GENE IN SEARCH OF GENETIC AND EPIGENETIC CAUSATIVE MUTATIONS IN BARATELA-SCOTT SYNDROME

by

Rebecca Sahraoui

A thesis submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Master of Science in Biological Sciences

Spring 2016

MAPPING THE 5' END OF THE *XYLT1* GENE IN SEARCH OF GENETIC AND EPIGENETIC CAUSATIVE MUTATIONS IN BARATELA-SCOTT SYNDROME

by

Rebecca Sahraoui

Approved: _____
Erica Selva, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____
Katia Sol-Church, Ph.D.
Professor in charge of thesis on behalf of the Advisory Committee

Approved: _____
Robin W. Morgan, Ph.D.
Chairperson of the Department of Biological Sciences

Approved: _____
George H. Watson, Ph.D.
Dean of the College of Arts and Sciences

Approved: _____
Ann L. Ardis, Ph.D.
Senior Vice Provost for Graduate and Professional Education

ACKNOWLEDGMENTS

A very wise woman once told me that we stand on the shoulders of those who have come before us. So this is my futile attempt to honor those who have gone before me, as well as those who have been on this difficult journey with me, although we know that no words will ever be sufficient. How do you thank those who have literally brought you back to life and given you a reason to wake up every morning? No matter how difficult or disheartening my days were, I knew I had my family that believed in me and loved me no matter what. I cannot imagine going through this process without their unconditional love. I would also like to thank the laboratory members, Deborah Stabley, without whom I could not have done this, Jennifer Holbrook, the voice of calm in the storm on many a day, and Katherine Robbins, always my ray of sunshine on a cloudy day. I would also like to thank my advisor, Dr. Katia Sol-Church, for taking me in without reservation, putting up with my senior moments (of which I had many), and for listening to me when I felt the need to talk about what I was feeling, without judgment. I could not dream of a better support system or more knowledgeable people to find myself amongst. Lastly, I would like to dedicate this to the mothers of children who are sick with rare and often incurable disorders. Nothing we will ever do in the laboratory compares to the sacrifice and courage these families show every day. What a blessing to be able to be a small part of their journey, even for

such a short time, and stand up for the women who will never feel their children's arms wrapped around them or hear them whisper "I love you". At the end of the day, it is about a mother's unconditional love for her child.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

According to the National Institute of Health (NIH) a disease or disorder is categorized as a rare disorder (or orphan disease) when it affects less than 200,000 people within the US in any given year. Genetic disorders are often characterized by mutations or DNA variants in a given gene. The human genome contains approximately twenty to twenty-five thousand genes, making up the "exome" of an individual. With the availability of the "complete sequence" of the human genome, the study of the structure and function, as well as gene interactions has expanded the ability to improve the diagnosis and treatment of genetic disorders. Sanger sequencing is still considered the "gold standard" in molecular diagnostics, and is often used to confirm mutations found with other technologies. Next generation sequencing (NGS), also called massively paralleled sequencing, has given researchers the ability to overcome some of the issues with traditional Sanger sequencing. While the focus for many years has been genomic mutations causing disease, we know now that epigenetics plays a major role not only in disease, but in normal regulation of the human genome. My project is based on the study of the rare genetic disorder Baratela-Scott Syndrome (BSS), a rare autosomal recessive disorder characterized by common phenotypes including skeletal dysplasia, distinct facial features such as a flattened midface and wide nasal bridge, as well as developmental delay with pre-school age onset. Prior to my joining the laboratory, there was parentally inherited homozygous

variants in the *XYLT1* gene found in one patient out of a total of 9 in the cohort. Other enrolled patients had either no mutation in *XYLT1* or only one allele of this gene was found to carry a deleterious mutation. The reference genome at this locus was revealed to be incomplete, and was shown to be resistant to amplification. With the use of molecular techniques, we were eventually able to identify inherited homozygous causative variants in one newly enrolled patient and most importantly, discover parentally inherited *XYLT1* CpG methylation (mCpG) in patients carrying a heterozygous mutation. One patient carried two alleles with hypermethylated CpG in the promoter region and exon 1 of the gene. *XYLT1* is not an imprinted gene, and controls as well as unaffected siblings do not show any methylation in the 5'end of this gene.

Thus we are now able to conclude BSS can be caused by different events affecting the function of the XYLT1 protein: homozygous loss of function mutations (point mutations or InDels), parentally inherited homozygous methylation defects (mCpG) or a combination of mutation and methylation defects inherited through the germline by unaffected heterozygous carrier parents.

Chapter 1

INTRODUCTION

1.1 Overview of Rare Genetic Diseases

According to the National Institute of Health (NIH) a disease or disorder is categorized as a rare disorder (or orphan disease) when it affects less than 200,000 people within the US in any given year (www.genome.gov). As of 2015, there were more than 7,000 rare disorders characterized, and upwards of 80% of these are genetically based, with fully half of those affected being children. The statistics for these pediatric disorders are sobering, as approximately one in three of these children will not live to their fifth year; with 35% of them not surviving beyond their first year. With such a large number of individuals affected by rare disorders, only 350 rare diseases make up 80% of the affected population (www.globalgenes.org).

Genetic disorders are often characterized by point mutations or other DNA variants in a given gene. The human genome contains approximately twenty to twenty-five thousand genes, making up the "exome" of an individual. In a monogenetic disease, a single gene with a mutation is the cause of the disorder. These can be further characterized as dominant, recessive, X-linked, etc. A dominant trait/disorder requires that the individual have only one (dominant) mutated allele to display the phenotype. Conversely, a recessive trait/genetic disorder generally requires

that the individual have inherited a mutated allele from each of their carrier parents in order to display the phenotype associated with the disorder (www.nih.gov/inheritancepatterns).

With the availability of the "complete sequence" of the human genome, the study of the structure and function, as well as gene interactions has expanded the ability to improve the diagnosis and treatment of genetic disorders. Clinicians are now able to screen for known disease causing genes rapidly. In many cases, this provides a timely diagnosis; however, there are also cases where traditional Sanger screening finds no known causative mutation. This is often the case for rare genetic disorders (American College of Preventive Medicine, 2010), that may be caused by a mutation in a single gene, mutations in multiple genes, or be caused by a combination of genetic and epigenetic alterations (www.genome.gov). With the advent of Sanger sequencing, microarrays, and other more recent high throughput technology, such as next generation sequencing (NGS), molecular biology has transformed, and given researchers the ability to see in depth the causative alterations at the genomic level that result in allele loss-of-function in the patient (Braunholz, 2015).

1.2 Hypothesis driven – Screening Known Genes

Sanger sequencing is based on a prioi knowledge of the gene causing a specific disease. The technique relies on the incorporation of chain-terminating dideoxynucleotides with fluorescent tagging during *in vitro* DNA replication. Sanger

sequencing is still considered the "gold standard" in molecular diagnostics, and is often used to confirm mutations found with other technologies. However, there are some limitations to Sanger, as is the case of mosaicism which can be missed due to low allelic representation. Indeed, Sanger sequencing is also what is referred to as a modal population technique, meaning, the sequencing takes into account the average of all the cell populations in the sample.

1.3 Non-hypothesis Driven Approach - Next Generation Sequencing (NGS)

Next generation sequencing (NGS), also called massively paralleled sequencing, has given researchers the ability to overcome some of the issues with traditional Sanger sequencing. For instance, while Sanger sequencing requires a priori knowledge of the target gene, NGS is completely hypothesis free, and enables the discovery of new genes or variants not previously known to associate with the disorder. NGS outputs millions of short reads at a time, and this is powerful technology is able to identify mutations that may appear only in a small population of cells (mosaicism), as well as give an accurate assessment of the copy number in different regions of the genome (Reis-Filho 2009). The two most common forms of NGS are whole genome sequencing (WGS) and whole exome sequencing (WES). With WGS, it is possible to investigate single nucleotide polymorphisms (SNPs) and copy number variants (CNVs), as well as structural variants (SVs) for not only the less than 2% of the genome that codes for protein, but also the other 98% of the genome that is non-coding sequence. Many of the noncoding regions contain important

regulatory regions like enhancers and promoters that control transcription. However, the amount of data generated (data load) in WGS can be overwhelming, and the technology is very costly to run.

Whole exome sequencing is the most common type of NGS used in clinical molecular laboratories. In WES, the portion of the genome which codes for protein is captured. While less than 2% of the human genome is exome, the vast majority of disease causing mutations occur within the exome. Therefore, there is a good chance that in the process of exome sequencing, the clinical investigators will be able to find the causative mutation without the expense of WGS (Rabbani et al. 2014).

Even with all the new technology available, there is a large amount of information that is still not known about the human genome. Only a small fraction of the genome is characterized, with up to 85% of the human genome still uncharted (Botstein et al. 2003, Rabbani et al. 2014). However, there are still a number of drawbacks to relying solely on NGS, and more specifically, WES for diagnostic purposes when dealing with rare genetic disorders. As was mentioned earlier, there is still a great lack of understanding of the importance of certain regions of the genome, such as the enhancer and promoter elements, etc. Another important aspect of genetic disorders is being able to see what is happening at the epigenetic level. In traditional WES, the epigenetic markers are not visible. For these reasons alone, it is important to use WES as a tool, while still performing some of the standard molecular procedures such as traditional PCR, Sanger sequencing, and bisulfite sequencing, as well as other

methods, in conjunction with NGS. Another major pitfall found with both NGS and Sanger sequencing when researching rare genetic disorders is that it depends on a complete reference genome in order to correctly call mutations, and as we discovered during this study, there are regions of the genome which are incomplete or missing in the reference for the simple fact that they are refractory to analysis using current technologies.

1.4 Epigenetics and Rare Diseases

While the focus for many years has been genomic mutations causing disease, we know now that epigenetics plays a major role not only in disease state, but also in normal structural and functional regulation of the human genome. Notably, CpG methylation is directly involved in transcriptional regulation, X chromosome inactivation, and chromatin structure, among other roles have been shown to play a role in gene regulation (Robertson 2002). In mammals, methylation of DNA and post-translational modifications of histone proteins are the only known epigenetic modifications (Robertson 2005). Rather than being evenly dispersed across the genome, methylation is highly compartmentalized (Robertson 2002). There are two stages to the methylation process. Setting up and maintaining methylation patterns is performed by DNA methyltransferases (DNMTs), while methyl-CpG binding domain proteins (MBD) are responsible for reading the methylation marks (Robertson 2005). The methylated CpG site acts as a ligand for the MBD proteins to bind. The binding of

MBD proteins to the CpG site recruits other chromatin remodeling complexes to the site and cause changes to chromatin structure. These changes ultimately lead to the nucleosome becoming compacted and the silencing of transcription (Brenet et al. 2011). During embryonic development, epigenetic modifications are needed in determining the final cell identity through transcriptional control, and are known to be essential for development and differentiation (Li et al. 2014). Patterns of methylation vary widely among different cell types, and cells go through waves of methylation and demethylation during different phases of development (Quon et al. 2012). During germ cell and embryonic development, these waves of methylation and demethylation occur, with the final status of methylation usually being established during gametogenesis. Another period which includes large changes in methylation status is just after fertilization, where rapid demethylation of both the maternal and paternal genome occurs (Robertson 2005). These methylation patterns are different than genomic imprinting, which is tied directly to a specific parental chromosome. Imprinting leads to differential expression in the offspring (Robertson 2005). While there is still much that is not known about the process of methylation, research has shown that the genetic sequence plays a critical role. Global methylation of the human genome is standard, with CpG islands making up a high percentage of the small areas of the genome which are not methylated. Unmethylated CpGs represent about 1% of the human genome, and are thought to be exclusively located in promoter regions (Hendrich et al. 2003). Recent research however, has shown that genomic methylation into the first exon of a gene plays an important role in transcriptional silencing (Brenet

et al 2011). Rennet et al found that, as in the promoter region, the density of CpGs within the gene was important to the outcome of methylation on transcription, as well as the location within the gene. The 5' end of the gene, when methylated, was much more tightly associated with transcriptional control than further downstream (2011). There are a number of rare genetic disorders that are caused or exacerbated by aberrant DNA methylation. Methylation abnormalities due to any number of reasons (UPD, loss of imprinting genes, chromosomal loss, etc.) is involved in both Prader-Willi and Angelman syndrome (Fairbrother et al. 2015). Methylation is also believed to be involved in some autism spectrum disorders (Behnia et al. 2015). Abnormal methylation patterns are also involved in Fragile X syndrome (Biancalana et al. 2015).

Chapter 2

BARATELA-SCOTT SYNDROME (BSS)

My project is based on the study of the rare genetic disorder Baratela-Scott Syndrome.

2.1 BSS Overview

Baratela-Scott Syndrome (BSS) (OMIM: 300881) is a rare autosomal recessive disorder that was first recognized by Dr. Wagner Baratela and Dr. Charles Scott, Jr., two clinicians on staff at the Alfred I. duPont Hospital for Children. Dr. Scott is renowned in the field of achondroplasia research and has written definitive guides on dwarfism. In our cohort, we currently have nine children from eight unrelated families, although the last patient was enrolled very recently, and is not included in my research. Figure 1 shows photographs of seven of our original patients. As you can see in figure 1, our patients share common phenotypes which include skeletal dysplasia, distinct facial features which include a flattened midface, wide nasal bridge, as well as cleft palate in some cases (though not pictured). All patient have shown a developmental delay, which becomes apparent at about pre-school age. Skeletal changes include a tendency for dislocations, shortened long bones and other changes in the bones of the extremities.

Figure 1. BSS cohort at the Alfred I. duPont Hospital for Children

Fig. 1 Seven of our original patients in the cohort are pictured here. Phenotypic similarities are apparent. (from Baratela et al 2012)

Figure 2. Pedigrees of the BSS cohort and mutation status at the start of thesis

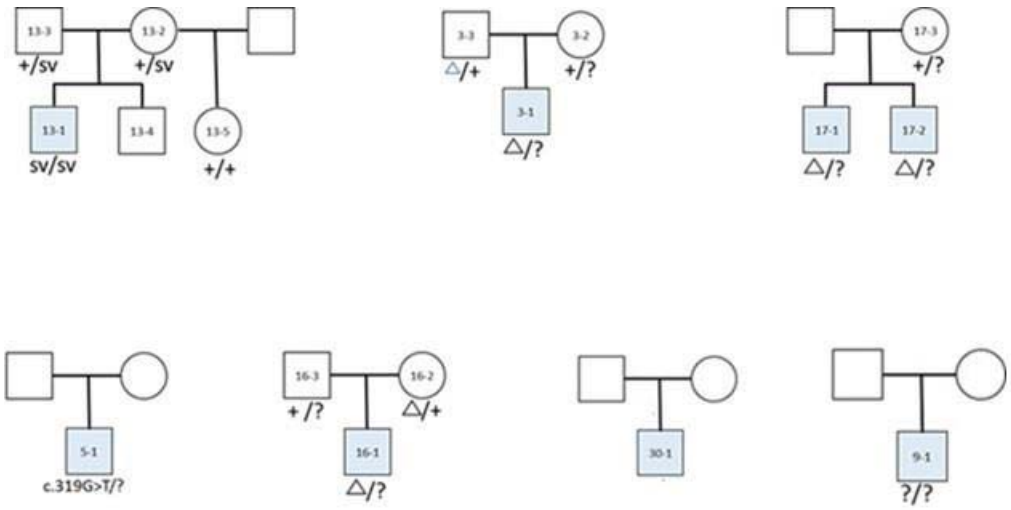


Fig. 2 Pedigrees of families of 8 to 9 patients in our BSS cohort with found mutations from preliminary data.

Figure 2 shows the pedigrees, with our BSS patients shaded in blue, as well as unaffected parents and siblings. Circles represent females and squares represent males. Individuals with no ID number were not enrolled in the study. Underneath each individual is the known genotype of subjects at the beginning of my research. Plus signs represent a wild type allele, SV is a splice variant, and the Delta indicates a known deletion. Because BSS is a rare recessive disorder which requires the presence of a mutation on both alleles of a given gene, the question marks represent the unknown mutation. The causative genetic changes/lesions resulting in the unknown allele is the focus of my research.

Due to phenotypic overlap, all the children were initially screened for a mutation in the *CANT1* gene, which is known to cause Desbuquois dysplasia (DBQD) (OMIM: 251450). In 2014, Bui et al published findings of a Desbuquois type 2 cohort, which differed from previous findings in DBQD patients in that this was the first instance of developmental delay being associated with children diagnosed with DBQD. It was significant to our research because the causative mutations were found to be in *XYLT1*, with no mutation in *CANT1* (Bui et al 2014). Due to the location of the causative mutations, as well as the phenotype including developmental delay in the entire cohort, Dr. Michael Bober, an expert in skeletal dysplasias, found that they are likely BSS as opposed to DBQD patients (personal communication).

2.2 *XYLT1* (xylosyltransferase 1)

The gene for xylosyltransferase 1 (*XYLT1*), which encodes a type 2 transmembrane protein, is located on 16p12.3, encodes for a 959 amino acid protein, and has a predicted molecular weight of 108 kDa. There are two isoforms of xylosyltransferase in higher organisms, XYLT1 and XYLT2, and it is the only enzyme in vertebrates with the ability to transfer xylose to serine residues in target proteins, thereby beginning the synthesis of glycosaminoglycan chains (GAG). The *XYLT2* gene codes for an 865 amino acid protein and has a predicted molecular weight of 97kDa (Hinsdale 2014). The presence of two differentially expressed xylosyltransferase genes encoding proteoglycan xylosyltransferases in vertebrates with similar function is believed to be a redundancy which allows for tighter regulation of post-translational GAG biosynthesis (Gotting 2007). While *XYLT1* is expressed predominantly in the early phase of chondrogenic stem cell differentiation, *XYLT2* is upregulated 7 days after induction (Gotting 2007). Exons 1 and 2 of *XYLT1* contain the cytoplasmic tail, transmembrane domain and parts of stem region. There are four types of GAG chains XYLT1 is responsible for synthesizing, heparin, heparan sulfate, chondroitin sulfate, and dermatan sulfate. Figure 3 shows a basic schematic of the addition of the tetrasaccraride chain linker which begins GAG biosynthesis. These GAG chains differ in the repeating structural units that are added after the initial four sugar linker chain. Altogether, the modifications make each GAG chain unique and determine its overall structure and function.

Figure 3. GAG chain synthesis image

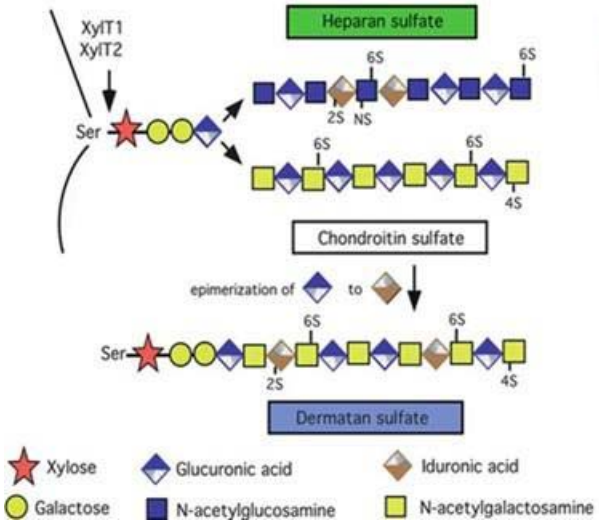


Fig. 3 XYLT1 beings the process of GAG chain synthesis with a tetrasacchraride linker.

These chains affect, among other things, cell proliferation, interactions with growth factors and other cytokines, as well as tissue morphogenesis (Mizumoto et al. 2015). Chondroitin sulfate (CS) is a major element in cartilage, a specialized connective tissue essential for extracellular maxtrix (ECM) formation and maintenance. CS side chains also regulate BMP (bone morphogenic protein) and TGF-B (transforming growth factor beta) signaling (Mizumoto et al. 2015). Heparin, heparan sulfate, dermatan, and hyaluronin are also involved in ECM patterning and stability, each interacting with their respective growth and signaling molecules (Knudson 2011). Research has shown that the first 260 amino acids of XYLT1 can be lost with no difference in enzymatic activity in vivo. However, the deletion of the first 272 amino acids results in complete loss of enzymatic activity. This has led to the identification of a crucial motif in XYLT1 (Gly^{261}-Lys-Glu-Ala-Iso-Ser-Ala-Leu-Ser-

Arg-Ala-Lys^{272}) (Gotting 2007). Unfortunately, there is still a great lack of information on the *XYLT1* gene, as well as the importance of different domains, including the promoter region. In 2014, Faust et al reported that they had finally uncovered the complete *XYLT1* promoter region. However, as we will show in the course of this document, while they did successfully sequence 238 bases of what we will be referring as the "newly discovered sequence", the sequence published does not contain the correct start to exon 1.

2.3 Preliminary Findings

2.3.1 Whole Exome sequencing

Although we currently have nine patients from eight families, as seen in figure 2, in our cohort, the project began with seven patients from six families. Of these original seven patients, six (3-1, 5-1, 13-1, 16-1, 17-2, 17-2) were subjected to WES at the University of Washington Center for Mendelian Genomics (CMG). Figure 4 is a schematic representation of the exonic structure of *XYLT1*. Mutations identified by WES and by Sanger, prior to my joining the laboratory are as indicated. The only patient found by WES to carry a homozygous mutation is patient 13-1. The proband carries a c.1290-1 G>A splice site mutation that was inherited from his two heterozygous parents. No other mutations were found using WES sequence analysis, however, after running Conifer, software able to detect large structural variants, a 3Mb deletion including the XYLT1 locus was found in patients 3-1, 17-1, 17-2, as well as 3-3, the father of patient 3-1. The deletions were confirmed using microarray (data not shown). Hence, all four individuals had a heterozygous deletion in *XYLT1*. However, BSS being a recessive disorder we knew a second mutation in *XYLT1* was not

uncovered. We then decided to screen the other WES patients for mutations in *XYLT1* using Sanger sequencing. Sanger sequencing uncovered a point mutation in exon 1 of *XYLT1* in patient 5-1 (c.319G>T [p.Gly107Ter]), which results in a change from amino acid glycine to a terminator, as well as a 26 base pair deletion in exon 1 of *XYLT1* in patient 16-1 (c.281-306del[p.Gln94Argfs59Ter]), which results in a frameshift and addition of arginine instead of glutamine with a terminator 59 amino acids further on.

Figure 4. Map of mutations found in XYLT1 with WES and Sanger sequencing

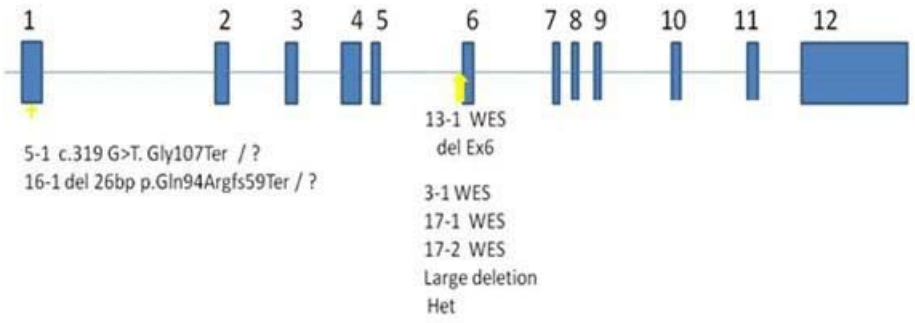
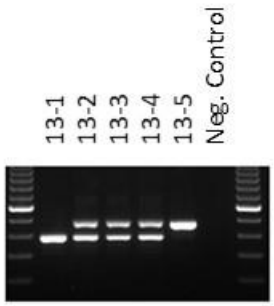


Fig.4 Mutations found prior to my joining the lab are mapped on a simple representation of the *XYLT1* gene.

For patient 13-1, the gel image in figure 5 shows the confirmation of WES data which uncovered the causative homozygous mutation in intron 5, causing the splicing out of exon 6. We can also see that both the mother (13-2) and the father (13-3) are

heterozygous for the deletion, as well as his brother (13-4). The maternal half-sister (13-5) does not have the mutation, as we can see by the single band of a larger size.

Figure 5. Gel image of family 13 confirming homozygous deletion



unpublished data D. Stabley

Fig. 5 Gel electrophoresis of cDNA amplification for family 13

2.3.2 Search for the missing second mutation

Sanger sequencing performed in the laboratory using primers placed in the 5' region of *XYLT1* which had not been covered with WES showed poor amplification. The coding minus five position (c.-5), a known SNP (C>G r.s.117041807) for patients 3-1, 17-1 and 17-2 showed a 'G' while the mother in both family 3 and family 17 amplified only a 'C' allele, underlined in figure 6.

Sanger sequencing of the 5' region of *XYLT1* in control samples identified a 232 base pair DNA sequence that was not in the reference genome. The new region was also found in both the patients and unaffected family members. Figure 6 displays the new sequence in red, with the reference genome in black and the published start of

exon 1 in purple. The newly discovered sequence is located 40 bases 5' of the published start to exon 1. The sequence contains unique sequences and features a GGC trinucleotide repeat region. The number of short trinucleotide repeats was shown to vary not only among individuals, but within an individual's alleles as well.

Figure 6. New sequence found in 5' region of XYLT1

Fig. 6 The black sequence is found in the reference genome. The red underlined portion is the sequence not seen in reference genome. The purple sequence indicates the published start of exon 1 found in the reference genome for XYLT1.

While there were some clues for the other patients as to at least one of their mutations, for patient 9-1, there were no indications of any causative mutation in *XYLT1*. Unfortunately, the laboratory received prepared DNA, so no exome sequencing was done on this patient due to an insufficient amount of genomic DNA. We also had no parental DNA for this patient (figure 2). PCR amplification followed by Sanger sequencing of exons 1 through 12 showed no mutations.

Sanger sequencing also confirmed the mother of patient 16-1 had two alleles, one wild type and one with the 26 base pair deletion. Sequencing of the father showed no deletion or apparent mutation. Cytoscan results on the father confirmed there are two alleles. However, sequencing of a 5' SNP site (r.s.118030014), with the patient showing a 'T', the mother showing 'C/T' and the father showing 'C', which did not work genetically by parental transmission. For patient 5-1, the laboratory did not receive samples from the parents. Figure 7 is a chromatogram image for family 3, showing patient 3-1 has a G at a the c.-5 SNP, both the mother (3-2) and father (3-3) show a homozygous C at that same position.

Figure 7. Chromatogram showing preferential amplification of alleles

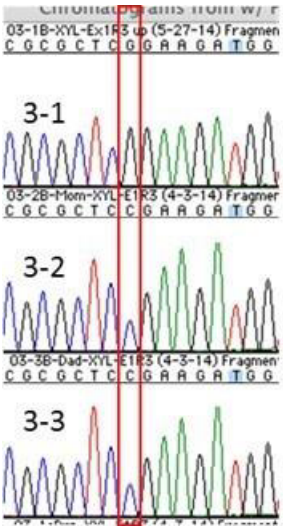


Fig 7. Chromatogram showing preferential amplification of the maternal allele (3-2). 3-1 and 3-3 appear homozygous because they contain the large deletion.

My specific aims were a direct consequence of all the data gleaned from exome and Sanger sequencing. The group had established that the patients had an autosomal

recessive disorder caused by mutations in the *XYLT1* gene. They discovered a poorly characterized region at the 5' end of *XYLT1* that was missing from the reference genome. Amplification of the 5' end was proving to be very difficult, with the promoter and exon 1 region was not amenable to standard molecular protocols. In addition, there was a 'hidden' allele in the parents as well as their offspring who inherited the hidden allele. Even in our controls, there were differences in amplification based on the region the primers were designated to amplify. With the identification of *XYLT1* as the causative gene, the focus was narrowed down to find the mutations leading to the phenotypes seen in the children (3-1, 5-1, 9-1, 16-1, 17-1 & 17-2).

2.4 Hypothesis

Baratela-Scott syndrome is a rare autosomal recessive disorder caused by variants in *XYLT1*.

Specific aim 1: to correctly map the 5' end of the *XYLT1* gene with emphasis on the promoter region; (a) to identify features in the promoter and Exon 1 of *XYLT1* that are missing from the reference genome found within the recently identified 232 base pair region and (b) to characterize the exon structure of the gene.

Specific aim 2: to find the second causative mutation/variant in our patients and unaffected carrier parents and identify any genetic and epigenetic changes unique to our patients.

Chapter 3

MATERIALS AND METHODS

3.1 Materials

All patients as well as family members were enrolled in an IRB approved research protocol and all samples, including but not limited to biological samples and photos, were provided with consent and assent. Whole blood and buccal samples were processed prior to my joining the Nemours Biomolecular Core Laboratory, using Qiagen/Gentra Puregene Blood Kit and Qiagen/Gentra Puregene Buccal Cell Kit per standard kit protocol under standard operating procedures (SOP). DNA and RNA were stored at -20°C and -80°C respectively until needed. Our study cohort consists presently of eight patients, and unaffected relatives (mothers, fathers, siblings). Of the eight patients, six were exome sequenced (3-1, 5-1, 13-1, 16-1, 17-1, 17-2), while two were only screened with Sanger sequencing (9-1, 30-1). All findings were validated using PCR and Sanger sequencing. For patient 5-1 there is a fibroblast cell line. Table 1 below is a complete list of all primers used throughout the course of this research, both by the BCL prior to my coming into the laboratory, as well as throughout the course of my project. Red indicates M13 tailed primers.

Table 1: Complete list of primers for BSS project

Name	Sequence	Name	Sequence
CH16-F1	AGCCCTGGATACCTTTGGAC	XYLT1-I2F	CAGGCACAGGGTAAGTATGAATAG
CH16-F2	GCTTGCCTTCCTATTCTTCG	XYLT1-5'UTRF4	GAGCGGGAGCCCGAGCGGCA
CH16-F3	TGGGTTACAGAGCAAGACT	XYLT1-ExIF3	AGCGCGGGGGCGGCCCGGAGCGT
CH16-F4	ATCCACATCAGGCTTCTCTA	XYLT1-ExIF2	AGGCGGCGGCGGAGGAGGAGGAGGAC
XYLT1-5UTRF1	AACTCCAGAAACTCACCCAAAGCC	XYLT1-Ex2F	CCACACCCAAGTCCGCTCATCA
XYLT1-5UTRF2	GCCTCCTCCCCACCCCTTCAA	XYLT1-Ex3F	GCTCCCCCGAGACCAAGTAT
XYLT1-5UTRF3	AGGGGTGTCTGTCTTCTCTGAT	XYLT1-Ex4F	CCGTGGAGTACATGCCAGCCAA
XYLT1-M13F-ASA-C-F	TGTAAAACGACGGCCAGTCCCCGGCGCCTTCCCaATC	XYLT1-Ex5F	CTCCAGGTCTCCAGGCAGTA
XYLT1-PromInsertF	CCGCTCCCGCCTCCCGCCTTCT	BISULFITE-F1	TTTTTTAGYGGGGATAGGGGTGTGGGG
XYLT1-PromInsertF2	TCCCTGCCCGTCTGAAAACTCC	BISULFITE-F2	GTTGGTYGTGTGGAATTTTAGTAG
XYLT1-I(1)F	CATCACCCTCCCCTCCAG	BISULFITE-F3	TTTTTTTTTGTYGTTATATTATTTGTTAATTGGGTTTG
XYLT1-5'UTRF4	GAGCGGGAGCCCGAGCGGCA	BISULFITE-F4	TTTAGGATGAAGGGATTAGATTGTGG
XYLT1-ExIF3	AGCGCGGGGGCGGCCCGGAGCGT	CH16-R1	CTGACACATAGTAAGTGCTCC
XYLT1_ASAc.-5C-F_MM_Tail	ATATACCTCGCTCGCCGGCCGCGATCC	CH16-R2	CGCTCTCTTAGTGTCTCAGCa
XYLT1_ASAc.-5G-F_MM_Tail	ATATACCTCGCTCGCCGGCCGCGTTTCG	CH16-R3	TTACTCCTGGTTAGGGGTGC
XYLT1-ExIF2	AGGCGGCGGCGGAGGAGGAGGAGGAC	5UTR-R1	TTCCCTGTGGGCCGCTGGAAGGA
XYLT1-IntIF	GGTTGTCGCAGCAGCATCTCTC	5UTR-R2	GAGGCGGAGAGGGCGGGGCCT
XYLT1-IntIF2	TGGGTAGCAGGGAGAAGTCT	XYLT1-PromInsertR	GCAGGGACCCGGACGTCACCAG
XYLT1-IntIF3	GGTCTGTCCTACTCTGACTG	XYLT1-ExIR	TTCCACACGACCAGCGTCTGC
XYLT1-2F	CAACACGAGAGGGTGAA	XYLT1-IntIR	AGTCTGATGACCGAGTTCAAGG
XYLT1-IntIF4	AAGATGCTGGGGTAGTAA	XYLT1-IntIR2	AGGCGATGTGGAGTCGGTAG
XYLT1-3(1)F	CAAGTCCGTATGACTCACTCTTC	XYLT1-ExIR3	CTGCAGCCGGCTCGGCGGGCAGGTC
XYLT1-4F	AGAAGCACCAGAGAAACTTCC	XYLT1_ASAc.-5C-R_MM_Tail	CACGGCGCCGCCACCATGTTCG
XYLT1-5F	TACAGAAAGACACTCCAGTCCAG	XYLT1_ASAc.-5G-R_MM_Tail	CACGGCGCCGCCACCATGTTCC
XYLT1-Int5F	TGCCCTCACAGAGTCTCCAG	XYLT1-M13F-ASAc.-5C-R_MM	TGTAAAACGACGGCCAGTCACGGCGCCGCCACCATGTTCG
XYLT1-7F	TTTATTTGGACTCTGGCTGG	XYLT1-M13F-ASAc.-5G-R_MM	TGTAAAACGACGGCCAGTCACGGCGCCGCCACCATGTTCC
XYLT1-8F	CCAGTTCTTACATGAACTCAGAGC	XYLT1-I(1)R	GAAAGGGCATCTTACCAGAGC
XYLT1-9F	CTAACCACAGGAATATGCCCTTAG	XYLT1-IntIR3	CCCTGACTGTTCACACCCACGA
XYLT1-10F	TAGATACAGAGGACGCAAGG	XYLT1-IntIR4	GGGACCTATGAAGAAATCAGC
XYLT1-11F	TGGGTATATGAAAGCTAACAAAGG	XYLT1-IntIR5	CAAATGCTCTGAGGCTGTGG
5'RACE Anchor	GACCACGCGTATCGATGTCGAC	5'RACE Anchor dt16V	GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTTV

3.2 Methods

3.2.1 Polymerase chain reaction (PCR)

The purpose of performing PCR is to amplify select regions of genomic DNA. PCR is an example of exponential amplification, so only a trace amount of genomic DNA is sufficient/required to produce enough of the desired amplicon to perform further molecular techniques/analysis. PCR was used for multiple purposes, namely mapping the 5' features of *XYLT1* with slowdown PCR and allele specific amplification, exon and exon-intron junction screening for mutations, cDNA amplification during

5'RACE protocol, colony amplification, and bisulfite treated DNA amplification. Table 2 shows only the primer pairs I used for my project for each application listed above. The size of each amplicon as well as annealing temperature (Ta) used for successful PCR are indicated.

Table 2: Primer Pairs used in the course of my research

F Name	R Name	Size	Ta	Methodology
CH16-F2	XYLT1-PromInsertR	1522	56	Screening
CH16-F4	XYLT1-Ex1R	2108	56	Mapping 5' region
XYLT1-5UTRF2	XYLT1-Ex1R	979	63	Mapping 5' region
XYLT1-5UTRF3	XYLT1-Ex1R	641	63	Mapping 5' region
XYLT1-5UTRF3	XYLT1-Ex1R3	769	63	Mapping 5' region
XYLT1-PromInsertF-M13	XYLT1_ASAc.-5G-R	320	62	Mapping 5' region
XYLT1-PromInsertF2	XYLT1_ASAc.-5C-R_MM_Tail	259(264)	62	Mapping 5' region
XYLT1-PromInsertF2	XYLT1_ASAc.-5G-R_MM_Tail	259(264)	62	Mapping 5' region
XYLT1-5'UTRF4	XYLT1-Int1R2	601	63	Mapping 5' region
CH16-F2	XYLT1-PromInsertR	1522	56	Screening
XYLT1-5'UTRF4	XYLT1-Ex1R	210	63	Screening
XYLT1-5'UTRF4	XYLT1-Ex1R3	338	63	Screening
XYLT1-Ex1F3	XYLT1-Ex1R	165	63	Screening
XYLT1-Ex1F2	XYLT1-Int1R2	247	63	Screening
XYLT1-Int1F	XYLT1-Int1R3	422	62	Screening
XYLT1-Int1F4	XYLT1-Ex2R	282	58	Screening
XYLT1-3(1)F	XYLT1-3(2)R	762	60	Screening
XYLT1-4F	XYLT1-4R	440	60	Screening
XYLT1-5F	XYLT1-5R	503	60	Screening
XYLT1-Int5F	XYLT1-Int6R	461	58	Screening
XYLT1-7F	XYLT1-7R	407	56	Screening
XYLT1-8F	XYLT1-8R	442	59	Screening
XYLT1-9F	XYLT1-9R	561	60	Screening
XYLT1-10F	XYLT1-10R	424	58	Screening
XYLT1-11F	XYLT1-11R	608	59	Screening
XYLT1-12F	XYLT1-12R	543	60	Screening
Anchor-dt16V	XYLT1-Ex7R	1500	60	5' RACE
XYLT1-Ex2F	XYLT1-Ex4R	650	55	5' RACE
Anchor-dt16V	Ex5R	1500	55	5' RACE
Anchor-dt16V	Ex4R	1275	55	5' RACE
Anchor-dt16V	Ex3R2	1100	55	5' RACE
Anchor-dt16V	Ex2R	600	55	5' RACE
Anchor-dt16V	Ex1R	400	55	5' RACE
BISULFITE-F1	BISULFITE-R1	491	55	Bisulfite/Colony OCR
BISULFITE-F2	BISULFITE-R2	398	55	Bisulfite/Colony OCR
BISULFITE-F3	BISULFITE-R3	339	57	Bisulfite/Colony OCR
BISULFITE-F4	BISULFITE-R4	350	58	Bisulfite/Colony OCR

PCR amplification for mutation analysis, promoter region mapping and colony amplification was performed using Qiagen Taq DNA polymerase kit (QIAGEN). All

reactions were done in 0.2mL tubes. Standard reaction volumes were 25µL, with some variation depending on the need to optimize. Standard reaction included 5µL Q solution, 2.5µl 10X Qiagen coral load buffer, 2.5µL of a mixture of 10µM forward and reverse primers (Table 1), 1.5µL 25mM MgCl, 0.5µL 25mM dNTP Mix (Promega), 0.15µL Qiagen Taq, 100ng to 500ng gDNA, nuclease free water up to 25µL. All cycling was done on either Applied Biosystems Veriti 96 Well Thermocycler or Applied Biosystems GeneAmp® PCR System 9700. Cycling conditions varied depending on experimental design and specific primer requirements. A general description is found in Table 3.

Table 3: Standard PCR program for thermocycler

Phase	Duration	Temperature °C	# of cycles
Initial Denaturation	5-10 minutes	95-98	1
Standard Denaturation	30 seconds	95-98	25-35
Annealing	30 seconds	See Table 2	
Extension	1-4 minutes	72	
Final Extension	5-10 minutes	72	1
Hold	indeterminate	4	1

At the end of PCR, typically 1/10th of each reaction was run on a 2% agarose gel in TAE running buffer with Invitrogen 100bp ladder marker. Bands were visualized using 3µL 1% ethidium bromide on Alpha Imager software.

Slowdown PCR amplification with 7-Deaza-2'-deoxy-guanosine-5'-triphosphate (7-Deaza) was performed as follows. All reactions were done in 0.2mL tubes. Standard reaction volume was 50µL and included 10µL Q solution, 5µl 10X Qiagen buffer, 0.5µL 10µM primers (Table 1), 3µL 25mM MgCl, 0.5µL 10mM 7-

Deaza-dGTP Mix (Promega), 0.5µL Qiagen Taq, 200ng gDNA, nuclease free H_2O up to 50µL. When second round amplification was needed, 1µL of the first round PCR reaction was used as template under the same original cycling conditions. 2% agarose gel was used as described above. Cycling conditions are in Table 4 as follows:

Table 4: Slowdown PCR cycling conditions

Phase	Duration	mperature	# Cycles
Initial Denaturation	5 minutes	95	1
Denaturation	30 seconds	95	**48**
Annealing	30 seconds	70**	
Extension	1 minute	72	
Denaturation	30 seconds	95	
Annealing	30 seconds	70**	
Extension	1 minute	72	
Denaturation	30 seconds	95	
Annealing	30 seconds	69**	
Extension	1 minute	72	
Denaturation	30 seconds	95	**15-25**
Annealing	30 seconds	58	
Extension	1 minute	72	
Final Extension	7 minutes	72	1
Hold	Indetermina	4	1

3.2.2 PCR Cleanup

After verification via gel electrophoresis that the PCR amplifications were successful and specific for the targeted regions, all PCR products were cleaned up for downstream analysis.

3.2.2.1 Qiagen Column PCR purification

For standard and slowdown PCR: amplicons were cleaned up with Qiagen Qiaquick PCR Purification®. Five volumes of binding buffer was added to each PCR reaction in a 1.5mL tube, vortexed briefly, and then transferred into a Qiaquick column placed in a clean labelled 1.5mL tube. Unless otherwise mentioned, spins were done at 16,000x g at room temperature. Binding buffer spin was done for 30 seconds. Column was removed, flow through poured off and column was placed back into tube. 750µL ethanol wash buffer was added to each column, tubes were spun 30 seconds and flow through was removed. An additional two minute dry spin was done on all tubes. Columns were then transferred to clean 1.5mL labelled tube and 30-50µL elution buffer was added to elution column. A final spin was done at 11,000x g for 1 minute. DNA recovery was visualized by agarose gel.

3.2.2.2 Exo-SAP IT PCR Purification

Before sequencing of bisulfite PCR products, they were cleaned using ExoSAP-IT to remove single stranded gDNA, primers and dNTPs. Five microliters of post-PCR reaction were added to two microliters of ExoSAP in a clean 0.2µL tube. The thermocycler program was as follows: 15 minutes at 37°C followed by 15 minutes at 80°C. Processed samples were not run on a gel or visualized, but were sequenced directly.

3.2.3 Sequencing of Standard and Bisulfite Amplicons

Sanger sequencing was done using ABI 3.1 Chemistry. Sequencing reactions were 10µL total as follows: 1µLTerminator Chemistry (2.5x), 1.5µL dilution buffer (5x), 1µL 3.2µM primer (either forward or reverse), 0.5 µL DMSO, variable DNA template, H_2O up to 10µL total volume. All cycling was done on either Applied Biosystems Veriti 96 Well Thermocycler or Applied Biosystems GeneAmp® PCR System 9700. Standard cycle sequencing conditions includes an initial denaturation for two minutes @ 96°C followed by 25 cycles of denaturation @ 96°C for 10 seconds, annealing @ 50°C for 5 seconds, and extension @ 60°C for three to four minutes (depending on size of product), followed by a final indeterminate hold @ 4°C.

3.2.4 Sequencing Purification SDS Treatment

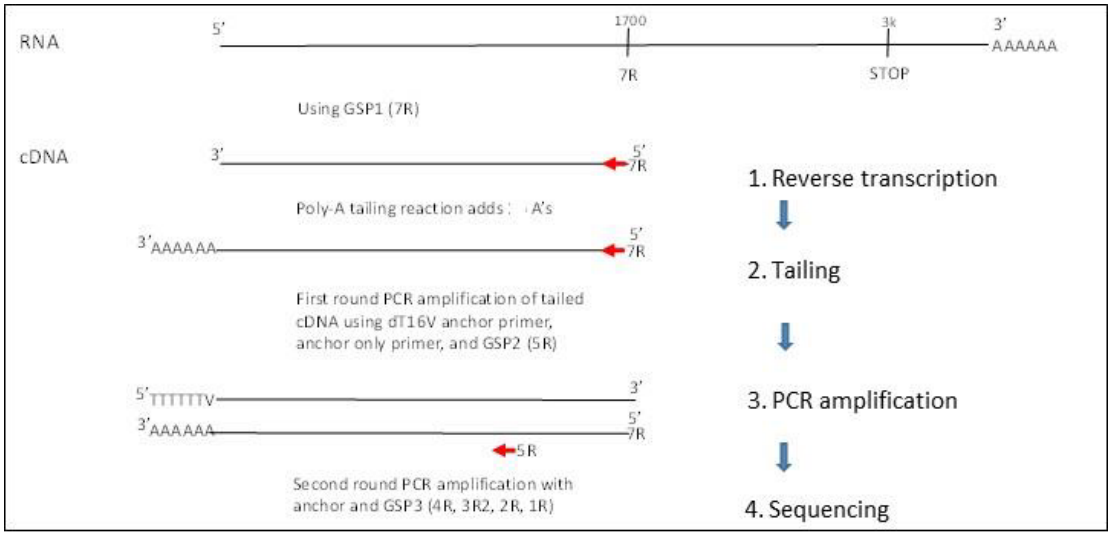
Sequenced samples were brought to 20µL with nuclease free water, and two microliters of 2.2% sodium dodecyl sulfate (SDS) was added to each tube. Tubes were vortexed, spun and heated to 98°C for five minutes in a thermocycler. Samples were column purified using Edge Biosystem Performa DTR Gel Filtration kit. Columns were centrifuged for three minutes at 800 rcf. Spin cartridge was removed and placed in clean 1.5mL tube. SDS treated samples were loaded into the packed column in center of the resin bed. Tubes were centrifuged three minutes at 800 rcf. Samples were then transferred to 0.2mL tubes and run through capillary electrophoresis. Capillary electrophoresis through a polyacrylamide gel was used to separate sequenced

fragments and generate a chromatogram for analysis using the BCL 3130xl Sequencing Protocol. Chromatograms were printed for visual analysis using Applied Biosystems Sequencing Analysis Software 6.

3.2.5 Rapid Amplification of cDNA Ends (5' RACE)

5' RACE is broken down into a number of different steps, as indicated in figure 8. RNA isolated from the control fibroblast line GM8447, as well as from 5-1 fibroblast were used.

Figure 8. Schematic of 5' RACE protocol



3.2.5.1 Reverse transcription (RT-PCR)

Applied Biosystems High Capacity cDNA RT® kit was used for reverse transcription on all samples to produce a single stranded cDNA copy of the targeted region of an RNA transcript. Prior to RT, all areas and materials needed, including but not limited to pipettes, outside of pipette boxes, pipette racks, etc. were cleaned thoroughly with RNase Zap, followed by RNase free water, and finally with 70% ethanol. Pipettes, pipette tips and racks were then placed in the UV box for fifteen minutes to ensure sterility. For the RT reaction, approximately 1µg RNA was added to 0.2mL tube with nuclease free water to ten microliters. Tube was heated in thermocycler (ABI Veriti or PE9700) five minutes at 65°C to denature RNA. Sample was immediately placed on ice after five minute denaturation. Master mix was as follows: 2µL 10x buffer, 0.8µL dNTP, 1µL RNase inhibitor, 1µL Multiscribe RT enzyme, 1µL gene specific primer (GSP – Ex7R 12.5µM), 4.2µL RNase free H_2O. Cycling was done on the ABI Veriti or PE9700 as follows: 10 minutes @ 25°C, two hours @ 37°C, five minutes @ 85°C, hold @ 4°C.

Qiagen Qiaquick PCR purification kit was used to clean RT samples. 160 µL of binding buffer with pH indicator was added to sample in 0.2mL tube, vortexed, spun and transferred to Qiaquick column placed in clean 1.5mL tube. Tubes were spun @ 16,000x g at RT (room temperature) 30 seconds. Flow through was removed and 750µL ethanol wash buffer was added. Tubes were incubated five minutes @ RT, spun 30 seconds. Flow through was removed, and an additional two minute dry spin was done. Columns were removed and placed in clean 1.5mL tubes. Forty microliters

of elution buffer) was added, spun @ 11,000x g for one minute. Remainder of sample not used immediately in poly-A tailing reaction was frozen at -20°C.

3.2.5.2 Poly A tailing reaction

The poly-A tailing reaction protocol was a hybrid of Roche and NEB protocols. Into a 0.2mL tube 17µL of cDNA sample was incubated 3 minutes in the 9700 thermocycler @ 94°C in the presence of 2.5 µL 10x TdT reaction buffer, 1µL dATP (5mM), 2.5µL CoCl2 (10X). Tubes were then placed immediately on ice and 2µL terminal transferase was added. Tubes were placed in PE9700 thermocycler and incubated 30 minutes @ 37°C, followed by 10 minutes @ 70°C and 4°C hold.

3.2.5.3 Gel extraction for sequencing

Gel extraction was done using Qiagen Qiaquick Gel Extraction® kit. The PCR amplified RACE sample was run on a 1% agarose gel to separate by size (range from 500 bp to 1.5kB). Bands were cut out of the gel and placed in pre-weighed 1.5mL tube. Tubes were then reweighed. Three volumes Buffer QG was added to one volume of gel in each tube. Tubes were incubated 10 minutes in water bath at 50°C until gel completely dissolved, with intermittent vortexing every two to three minutes. Once gel was completely dissolved, one volume of original gel slice weight of isopropanol was added and mixed. The sample was transferred to Qiaquick spin column and spun at 16,000x g for one minute. The flow through was discarded and 500µL Buffer QG was again added to column. Tubes were spun again @16,000x g for one minute with flow through being discarded. 750µL of ethanol wash buffer was added to each tube, after

which tubes were centrifuged @ 16,000x g for thirty seconds with an additional 2 minute dry spin. The column was placed in a clean 1.5mL tube and 30 to 50 µL Buffer EB was added to each and tubes were centrifuged for one minute @ 16,000x g.

3.2.6 Bisulfite Conversion

Bisulfite conversion was performed using Zymo Research EZ DNA Methylation-Lightning™ Kit. Specimens include DNA extracted from saliva or blood. Approximately 500ng of DNA was brought up to 20µL with dH_20 prior to the addition of 130µL Lightning conversion reagent. Tubes were placed in a thermocycler and incubated at 98°C for eight minutes followed by a one hour incubation at 54°C and the final hold was 4°C. Tubes were then removed and 600µL M-Binding buffer was added to a Zymo-Spin™ column placed in a collection tube. Samples were loaded into column containing the binding buffer, inverted repeatedly, and centrifuged @ 16,000 g for 30 seconds. Flow through was discarded and 100µL of M-Wash buffer was added to the column, followed by another 30 second centrifugation at 16,000 g. Flow through was discarded. 200µL of L-Desulfonation buffer was added to columns and tubes which were incubated at room temperature for 20 minutes, followed by centrifugation and removal of flow through. An additional 200µL of M-Wash buffer was then added to the column and tubes were centrifuged 30 seconds at 16,000g. An additional wash step with 200µL M-Wash buffer was done under the same conditions as above. Columns were then placed in a clean 1.5mL tube and 10µL of Elution buffer was placed into the column. Tubes were centrifuged @ 16,000g for 30 seconds and

column was removed. Samples were stored at -20°C short term until needed. For long term storage, tubes were placed in -80°C.

Bisulfite PCR amplification was performed using converted primers (Table 1), Qiagen HotStarTaq® DNA polymerase kit and had a reaction volume of 50µL which included 10µL Q solution, 5µl 10X buffer, 2.5µL of each 10µM primer (forward and reverse primers were stored and added separately), 3µL 25mM MgCl, 1.0µL 25mM dNTP Mix (Promega), 0.25µL Qiagen HotStar Taq, 2-4µL converted gDNA (may vary slightly), nuclease free H_2O up to 50µL. All cycling was done on either Applied Biosystems Veriti 96 Well Thermocycler or Applied Biosystems GeneAmp® PCR System 9700. Cycling conditions were as follows: initial denaturation 15 minutes @ 95°C, followed by 45 cycles of denaturation @ 94°C for one minute, annealing @ 55°C for one minute, extension @ 72°C for one minute, final extension @ 72°C for 10 minutes, and indeterminate hold @ 4°C. Bands were visualized as previously mentioned.

3.2.7 Cloning of Bisulfited DNA

Cloning of bisulfite PCR products was performed using Invitrogen™ TOPO® TA Cloning® Kit. Figure 9 includes a schematic of the TOPO 2.1 cloning vector plasmid used in all transformation, as well as the surrounding sequence of the cloning site, including the location of M13 forward/reverse primers for PCR and sequencing, as well as antibiotic resistance genes.

Figure 9. Map of 2.1-TOPO cloning vector

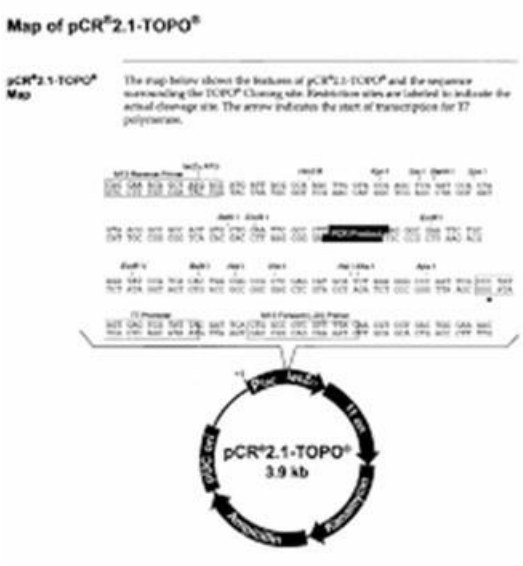


Fig. 9 Image of plasmid used for bisulfite cloning. Features include M13 forward and reverse sequences for primers, genes encoding ampicillin and kanamycin resistance for selection, as well as the reporter gene LacZ to allow for white/blue screening.

3.2.7.1 Preparation of LB ampicillin plates

LB plates were prepared using aseptic technique. LB/Agar (37g/L) was added to Milli-Q water in a round 500mL glass bottle and autoclaved. When LB/agar solution was cool enough to touch, 100µg/mL final concentration ampicillin was added to the LB/agar solution and swirled to mix. Approximately 20 mL was poured into each plate, cooled to room temperature to solidify and stored at 4°C until needed.

3.2.7.2 Cloning ligation and transformation

Four microliters of PCR product described in the previous section was added to one microliter of a buffer containing1.2M NaCl and 0.06 M $MgCl_2$, and one microliter of TOPO® Vector in a 0.2mL tube. Reactions were then incubated 20 minutes at room temperature. While reactions were incubating, one vial of One Shot®

chemically competent *E. coli* cells was thawed on ice for each ligation reaction, the hot water bath was equilibrated to 42°C, the proper number of LB/agar ampicillin plates were placed in a 37°C incubator and S.O.C. (super optimal broth with catabolite repression) medium was brought to room temperature.

After 20 minute incubation, two microliters of each ligation reaction was added to a vial of ice-thawed One Shot® cells and mixed gently via inverting. Vials were incubated on ice five minutes, followed by a 30 second heat-shock of cells in the warm water bath set to 42°C without shaking, and put immediately back on ice. 250µL room temperature S.O.C. medium was added to each vial of transformed cells and placed in a shaking incubator set to 37°C at 200 rpm for one hour.

3.2.7.3 Plating and screening

One half hour into the one hour incubation described in above section, selective amp plates were removed from the incubator, and 40µL of 40mg/mL X-gal in dimethylformamide (DMF) was spread on each plate. Plates were then placed back in the incubator for remainder of incubation.

After the incubation period, 50µL and 200µL of the transformed cells were spread onto two LB ampicillin plates. Plates were incubated upside down in an incubator @ 37°C overnight. After 24 hours, plates were inspected for the presence of white/blue colonies, and the white colonies were picked and gridded onto ampicillin plates to be grown out for an additional 24 hours in the 37°C incubator.

3.2.7.4 Cracking buffer gel electrophoresis/transformant analysis

Cracking buffer for cracking gel protocol was made as follows: 3g sucrose [10%], 300µL 10M NaOH [0.1M], 150µL 20% SDS [0.1%], 60µL orange dye (1:500 dilution). Solution was brought to 30mLs with water to a final concentration of 1X. White colonies were picked from the grid plate and added to a 1.5mL tube containing 30µL of cracking buffer by swirling and rapidly moving the tip up and down. Tubes were vortexed and centrifuged for five minutes @ 6,000 xg. 15µL of the supernatant was then run on a 0.8% agarose gel alongside a $1/4^{th}$ dilution of a 1kB ladder at 120v and a blue colony as a plasmid with no insert control comparison, for approximately two hours. Bands were visualized using 3µL 1% ethidium bromide on Alpha Imager software. Colonies which showed the recombinant plasmid in the gel were then subjected to colony PCR, cleaned and sequenced as previously explained in section 3.2.3.

3.2.7.5 QIAprep miniprep purification for plasmid purification

In rare cases, white non-satellite colonies were grown overnight in 1mL LB in 25mL falcon tubes and purified using the QIAprep Miniprep protocol. After overnight growth in 1mL LB/ampicillin broth, cells were pelleted via centrifugation at 11,000x g for 5 minutes. 250µL of resuspension buffer to lyse and neutralize the solution was added to cell pellet and repeatedly flicked and vortexed until homogenous. Cells were then transferred to a clean labelled 1.5mL tube and 250µL of an additional lysing buffer was added to sample to solubilize the cell wall and degrade proteins. Tubes

were inverted 4-6 times gently to mix. 350µL binding buffer was then added to allow reformation of hydrogen bonds. Upon addition of the buffer, tubes were immediately inverted 4-6 times to ensure proper mixing. All tubes were centrifuged for 10 minutes at 17,900xg. Standard Qiagen cleanup protocol was then done as described earlier. We were then left with purified plasmid, which was ready for sequencing.

Chapter 4

RESULTS

4.1 Specific Aim 1: To Map the 5' end of the *XYLT1* Gene with Emphasis on the Promoter Region

4.1.1 Characterize the structure of the new 232 base pair region

Two different methods were used to successfully sequence the 5' end of *XYLT1* gene in controls, patients as well as unaffected family members. In some cases, standard PCR followed by Sanger sequencing was unsuccessful in uncovering the unique features of the 5' region of *XYLT1* due to sequence complexity, and slowdown PCR with 7-deaza was necessary, especially for the "hidden" allele.

Figure 10 shows the specific features which were discovered with standard PCR, and in most cases, slowdown PCR with 7-deaza. Some of the features, such as the G quadruplex, the start of exon1 and the coding minus 5 SNP are indicated by colored boxes. While this regions showed variability with numbers of repeats, etc., the unique features were a valuable marker for mapping each individual's variation to see if a particular feature would explain why the area was so difficult to amplify with standard techniques.

Figure 10. Map of unique features in 5' region of XYLT1

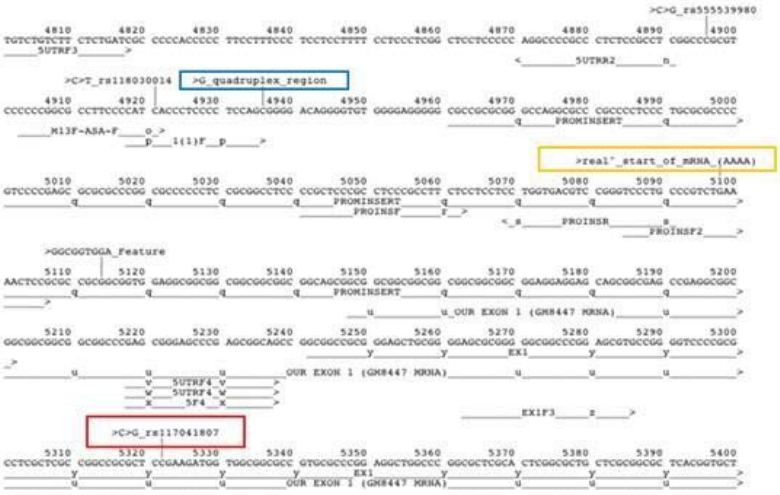
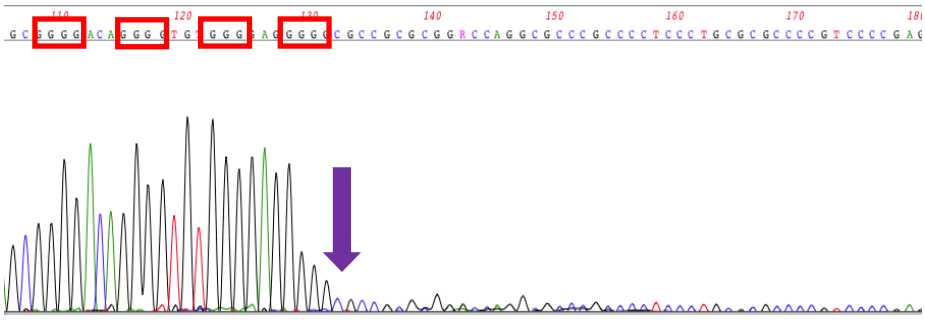


Fig. 10 A map of the unique features in the 5' region of XYLT1

While using slowdown with the addition of 7-deaza was more successful in amplifying through this region, I was still unable to completely make it through some of the features, as can be seen in figure 11, which shows a perfect example of the sudden drop in the chromatogram directly following the G quadruplex. A G

quadruplex is a secondary structure formed in the DNA when there are multiple stretches of four or more guanines. They bind to each other to form a guanine tetrad, and these tetrads then bind on top of each other to form the G-quadruplex (Huppert et al, 2007).

Figure 11. Chromatogram showing G quadruplex region



Chromatogram showing almost complete drop off in sequencing directly following the unique 5' features, the G quadruplex. This was seen in all of our patients for we were searching for the second mutation.

For patients 3-1, 5-1, 16-1, 17-1 and 17-2, slowdown-7 deaza was successful in mapping more of the unique features of the 5' region, but sequencing using forward primers, e.g. 5UTRF3 (Table 1), invariably fell apart right after a G quadruplex. In unaffected carrier parents, the wild type allele was more amenable to amplification. However, even in the parents, there were regions which were still not sequencable. Table 5 below details the specific features of the 5' region of *XYLT1*, as well as variation between patients and unaffected carrier parents. Our controls (non-family members) had the longest uninterrupted GGC repeats (up to 29), indicating it was not the expandable repeat region which was the second causative "mutation" in our patients. Interestingly, in table 5 patient 17-2 shows only one GGA repeat, while 17-1

has the two maternally inherited GGA repeats. Allele specific amplification in conjunction with slowdown PCR was also able to successfully amplify the 'G' allele (of the C>G SNP at c.-5 position) in more detail although there were regions still unamplified. Another feature in this area was the eight base pair deletion in the 3' sequence found only in our patients.

Table 5: Features of the 5' region of *XYLT1*

Sample	5' SNP C>T	5' Seq GGCGGTGGA	GG C	AG C	GG C	GG A	InDel	3' Seq GCAGCGGCGA	c.-5 SNP C>G
3-1				0	0	0	-	8bpdel……GA	G
3-1	Large 3Mb deletion including the *XYLT1* allele								
3-2	C	+	7	1	8	3	-	+	C
3-2	C			0	0	0	-	8bpdel……GA	G
3-3	Large 3Mb deletion including the *XYLT1* allele								
3-3	C	+	7	1	8	3	-	+	C
5-1							-		
5-1	C	+	7	1	8	3	-	+	C
13-1	C	+	15	0	0	3	-	+	C/C
13-1	C		15	0	0	3	-	+	C
13-2			15	0	0	3	-	+	C
13-2			9	1	5	3	-	+	C
16-1	T	+	9	1	5	3	-	+	C
16-1	C			0	0	0	-	8bpdel……GA	G
16-2	T	+	9	1	5	3	-	+	C
16-2	C	+	18	0	0	2	-	+	G
16-3	C	+	7	1	5	3	-	+	C
16-3	C			0	0	0	-	8bpdel……GA	G
17-1	C	+		0	0	2	-	+	G
17-1	Large 3Mb deletion including the *XYLT1* allele								
17-2	C	+				1	-	+	G
17-2	Large 3Mb deletion including the *XYLT1* allele								
17-3	T	+	9	1	5	2	GGC G	+	C
17-3	C	+		0	0	2	-	+	G
30-1	C	+	12	1	5	3	-	+	C
30-1	T	+	16	0	0	3	-	+	C

4.1.2 Mapping the start site of transcription

Rapid amplification of cDNA ends (5' RACE) was initially performed on RNA extracted from GM8447 fibroblasts to investigate and validate the correct start of exon 1 for *XYLT1* based on our findings of a 232 base region 5' of the gene which was not in the reference genome. In order to do this, the work needed to be done on unaffected individual's samples to ensure the results would be representative of the general population, and not an artifact of BSS. The results showed that the start of the mRNA for *XYLT1* is much further 5' of the published start of Exon 1. Figure 12 clearly shows the presence of the poly-A tail, indicative of an RNA start 94 bases upstream of the published start of exon 1. 5' RACE was then performed on 5-1 and results validated a new start for exon 1 upstream from the published site.

Figure 12. Chromatogram of 5' RACE results showing start of exon 1

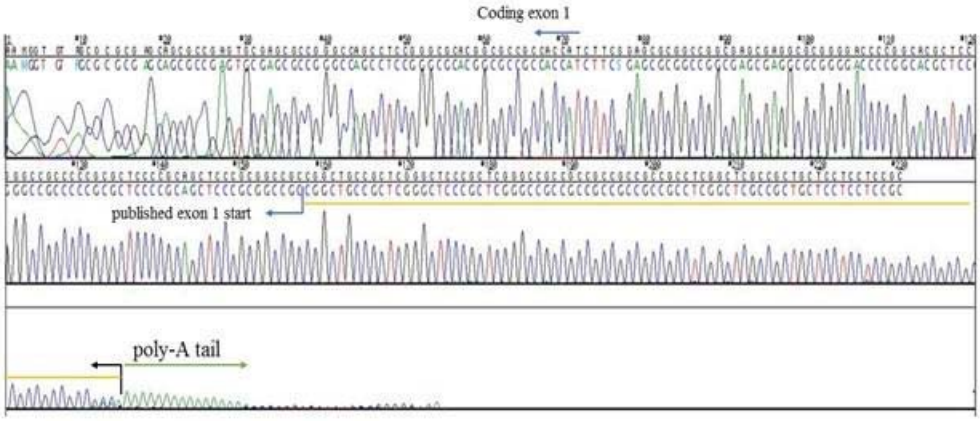


Fig. 12 The chromatogram from GM8447 shows the presence of the poly-a tail, indicating the protocol was successful, as well as the published start of exon1. However, there are bases between the published start and the poly-a tail, meaning the correct start to exon 1 is further into what was considered the promoter region than previously thought.

The sequencing in figure 12 was matched against the reference genome to confirm it was *XYLT1*. It also showed the presence of a portion of the newy found region not represented in the reference genome. However, the RT reaction stopped at the STR region which had shown previously to be difficult to amplify. Therefore, it was possible that the true start of exon 1 was even further 5' than what was being shown on the sequencing. Further mapping was able to show precisely the start of exon 1 at the "AAAA" feature in the newly discovered region, which is indicated in figure 10.

BsrF1 digestion was performed to see if, by cleaving the DNA into smaller segments, difficult regions would be more accessible to amplification. Amplification was done on digested samples across the 5' SNP and the coding minus five SNP across a BsrF1 site using the primers indicated in figure 13.

Figure 13. BsrF1 image showing location of digestion and amplification

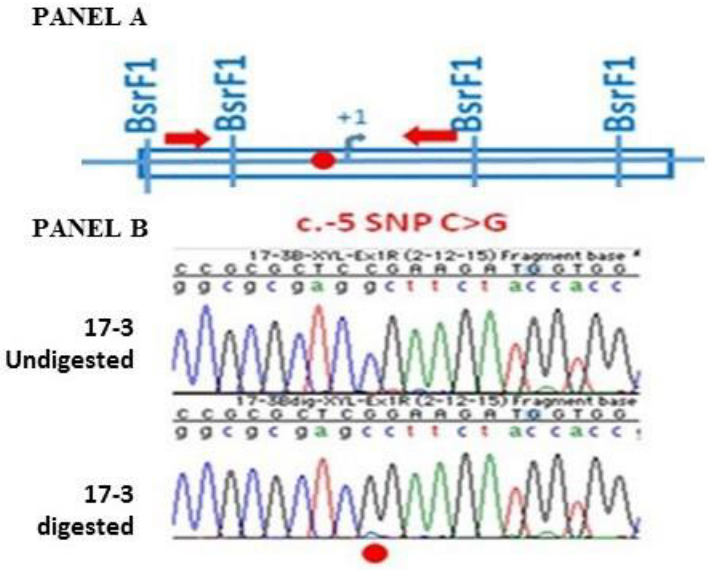


Fig. 13 Panel A shows a schematic of exon1 and identifies the BsrF1 sites as well as the start of translation (+1). A SNP identified in patients 17-1 and 17-2 at the c.-5 is also shown. The two affected individuals are G at the site while their mother (17-3) is C. Maternal DNA was digested with BsrF1 and amplified using primers spanning the SNP site. Panel B shows the sequencing results which clearly show the G allele.

Prior to enzyme digestion, only the wild type allele was visible when sequencing 17-3, the mother of affected siblings 17-1 and 17-2, and the allele she passed down to her sons remained elusive, as is shown in figure 13, with the allelic skewing between the undigested and digested chromatogram. We found the same phenomenon at the other SNP sites investigated. Results for patient 5-1 (for which no parental samples were made available for investigation), the mother of patient 3-1, and patient 16-1 were similar. In each case, the missing allele could be identified only after BsrF1 digestion. Where previously we had been amplifying one allele almost exclusively, after digestion, the allele which had remained hidden was the allele which amplified (meaning it was not digested), while the allele which had been amplifying with no problems was cleaved in the enzyme digestion. This meant that, for the first time in many cases, we were able to amplify the parental allele which had been inherited by the patients.

4.2 Specific Aim 2 results: Finding the Second Causative Mutation

Patient 30-1 was screened for all exons using Sanger sequencing. For this patient, the causative mutation is a homozygous four base pair insertion in Exon 8 (ATGA) c.1733_1734 ins ATGA, (p.D589Efs1), which causes the out of frame substitution of a glutamic acid for an aspartic acid leading to a frameshift and a terminator immediately thereafter, as can be seen in the chromatogram in figure 14. Panel A shows a control sequence through this region, and Panel B, taken from sequencing done on patient 30-1, clearly indicates a homozygous four base pair insertion. Using Sanger sequencing results, I was able to identify the cause of BSS for this patient.

Figure 14. Sequencing results for patient 30-1 shows homozygous mutation

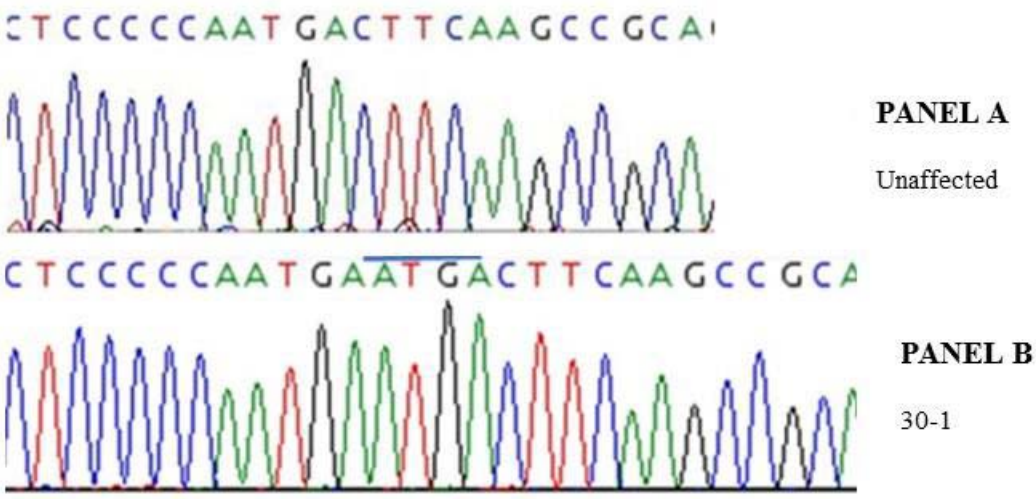


Fig. 14. The top panel is a chromatogram of an unaffected indivual containing the normal sequence. The bottom panel is the chromatogram from 30-1 showing the homozygous four base pair(ATGA) insertion in exon 8.

For patients with heterozygous mutations, 5-1 and 16-1, preliminary results with BsrF1 revealed the second parental allele had been difficult to amplify. Hence, this brought up the question as to why a segment of DNA which had been enzymatically cleaved and then amplified across the cleavage site with primers sitting on either side of the cut site would show a product upon analysis. Investigation led to the discovery that the enzyme was methylation sensitive, and thus revealed the second allele carried mCpG sites that could be investigated with bisulfite treatment of the genomic DNA, followed by PCR amplification and sequencing of bisulfite treated gDNA. All patients that were heterozygous or carried no detectable mutations (pt. 9-1) as well as the available unaffected carrier parents were subjected to bisulfite

sequencing.Figure 15 shows the chromatograms of one of the controls which was bisulfite sequenced compared to that of two patients with deleted alleles, 3-1, 17-2 and the mutation negative patient 9-1.

Figure 15. Bisulfite sequencing shows protection in patients CpG sites

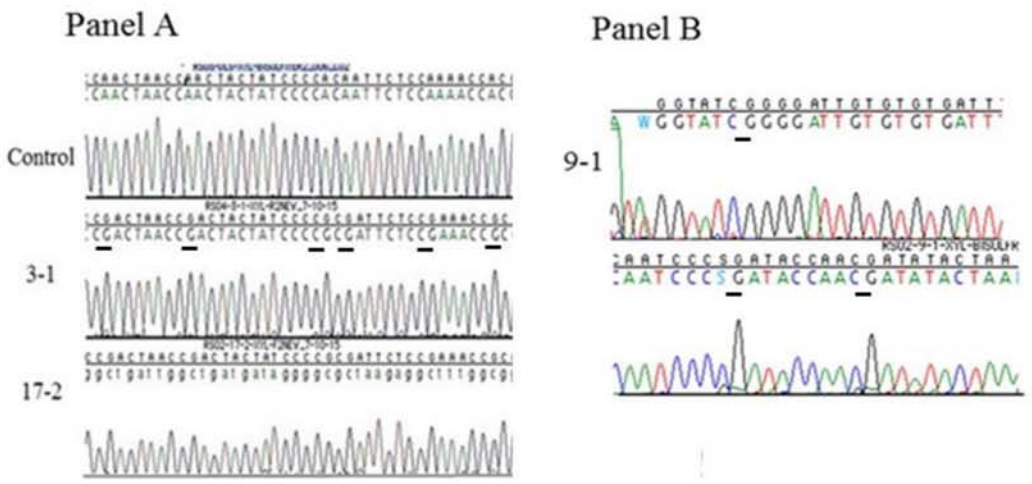


Fig 15. Chromatogram in Panel A shows patients 3-1 and 17-2 with complete conversion of CpG sites. Panel B is sequencing for patient 9-1, who shows homozygous methylation on both alleles in both forward and reverse sequencing.

While the bisulfite treatment successfully converted all the CpG sites in the control, it is clear the CpG sites in both 3-1 and 17-2 were protected from conversion, indicating methylation at those sites. This is representative of what was found in sequencing of bisulfite treated DNA for our patients as well as controls. Bisulfite sequencing was done in the promoter region as well as exon 1. Patients 3-1, 17-1 and 17-2, who only have one allele due to a large deletion, showed complete protection at every CpG site.

For patient 9-1, as seem in figure 15 panel B, both alleles are methylated. Both CpG regions in the promoter and in exon 1 showed the same methylation patterns in the affected patients that were absent from control samples.

While patients with one allele or both alleles methylated showed clean sequencing profiles (figure 15), patients with two alleles (16-1 and 5-1) as well as unaffected carrier parents showed heterozygous peaks at CpG sites, in both promoter and exon 1, indicating one allele was protected and one converted. These heterozygous bisulfite converted amplicons needed to be cloned to clarity the methylation pattern and ensure they were allele specific.

By cloning, I was also able to successfully correlate the presence of the point mutation in exon 1 for patient 5-1 to the non-methylated status of the CpG sites, with only the non-mutated allele showing methylation of CpGs (data not shown).

Figure 16 is of patient 16-1 and his father, 16-3, who passed along to his son the second mutated allele which had been so elusive.

Figure 16. Sequencing of clones shows methylation of patient 16-1

Fig. 16 Sequencing of clones for patient 16-1 show complete methylation at each CpG site in the exon1 region. 16-3 clones in the same region show conversion indicating there is no methylation of CpG sites in this particular clone.

While it is clear that patient 16-1 has CpG sites which are methylated, as indicated by the CG in sequencing on top, his father has CA in the same CpG sites (reverse sequencing) in the chromatogram below patient 16-1, indicating that in the case of this clone, it was from a plasmid which incorporated the non-methylated allele. Table 6 shows the summary of the cloning of bisulfite amplicons with the breakdown of the methylated vs. un-methylated clones found after cloning. This table is the result of collaboration with Deborah Stabley, so we may have a greater number of clones screened for both the promoter and exon CpG regions in all patients and unaffected parents.

Figure 17. Pedigree with mutation/variant results for BSS cohort

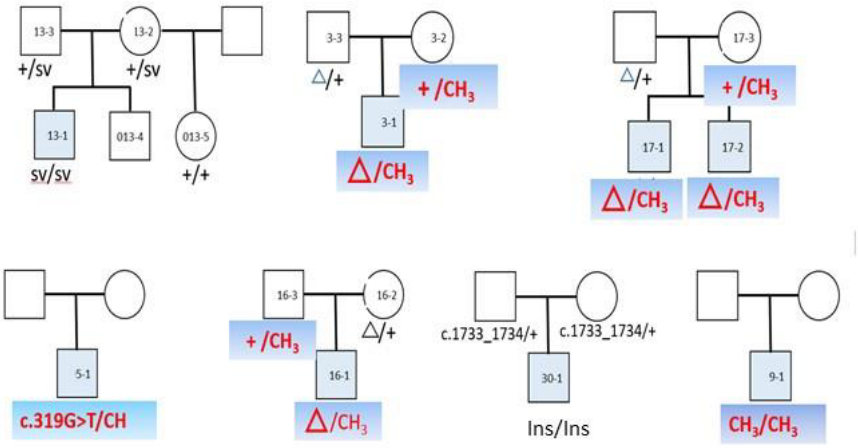


Fig 17. Pedigree of original BSS cohort with a complete listing of causative mutations/variants for each of our patients.

Table 6. Screening data from cloned bisulfited amplicons

Patient	Region	Expected	Total Done	Results
3-1	promoter	100% CH3	21	100% CH3
3-2	promoter	50% CH3 50% WT	39	36% CH3
5-1	promoter	50% CH3 50% MUT	28	100% CH3
	exon 1	50% CH3 50% MUT	19	75% CH3 25% MUT
9-1	promoter	100% CH3	12	100% CH3
16-1	promoter	50% CH3 50% MUT	17	67% CH3 33% MUT
16-3	promoter	50% CH3 50% WT	14	67% WT 33% CH3
17-1	promoter	100% CH3	21	100% CH3
17-2	exon 1	100% CH3	27	100% CH3
17-3	exon 1	50% CH3 50% WT	38	70% CH3

In collaboration with Deborah Stabley, we increased the number of clones for family 3 and family 17. The final breakdown for patient 3-1 was 21 clones showing complete

methylation. His mother, 3-2, had an additional 18 clones sequenced, with four of them showing protection and 14 methylation. The father, 3-3 had an additional 11 clones sequenced, none of which were methylated. Patient 17-1 had an additional 10 clones while his brother patient 17-2 had an additional 13 clones sequenced, all of which showed complete methylation. Their mother, 17-3, had an additional 20 clones sequenced, with 70% showing methylation. The additional screening gave a closer percentage of methylated vs. non-methylated to what was expected.

Chapter 5

DISCUSSION

5.1 Mutations in our Cohort

Within our cohort of BSS patients, we now have two patients with homozygous genetic mutations (patients 13-1 and 30-1), four (3-1, 5-1, 17-1, 17-2) with compound heterozygous mutations (InDel or mutation on one allele and aberrant methylation on the other allele), as well as one patient (9-1) with a homozygous epigenetic mutation in the *XYLT1* gene.

5.2 Why do Mutations in XYLT1 cause BSS?

When I first came into the lab to work on this project, what was known was Baratela-Scott Syndrome is a rare recessive disorder caused by a mutation in *XYLT1*, a gene that plays a vital role in bone formation and maturation, as well as cognitive ability. There are portions of *XYLT1* in the promoter region as well as into exon 1 which are highly resistant to amplification using standard molecular methods. The question then becomes, how does a mutation in the *XYLT1* gene cause BSS?

XYLT1 is responsible for the four sugar linker chain which begins the process of GAG chain synthesis. The four GAG chains which are dependent upon XYLT1 to begin synthesis are heparin, heparan sulfate, chondroitin sulfate, and dermatan sulfate. (Hinsdale 2014). Cartilage is an avascular connective tissue which provides the

flexible scaffolding and structure for the skeleton. What it does not possess, however, is the ability to easily regenerate due to low blood supply. The primary GAG chain in cartilage is chondroitin sulfate, which ultimately becomes the proteoglycan aggrecan, found in highest abundance in cartilage. Aggregan interacts with hyaluronan in the binding of other molecules to form a "supremolecule", as well as providing osmotic resistance in adult articular cartilage, and is vital in the process of chondrogenesis, the building of the cartilage matrix which will become bone during human development. Hyaluronan is also an important player in embryonic limb bud formation and maturation. Heparan sulfate proteoglycans have been shown to directly affect the function of growth factors involved in limb budding and formation such as hedgehog and bone morphogenic protein (BMP) during embryogenesis (Knudson 2001), as well as regulate the signaling proteins involved in growth (Huegl et al. 2013). So it is easy to see how a mutation or disturbance in the proper functioning of XYLT1 would adversely affect bone formation and growth.

5.3 What Have Others Found?

In 2014, Bui et al published findings of a cohort of patients with skeletal dysplasia and developmental delay with mutations in the *XYLT1* gene. After screening ruled out known genes for mutations, such as CANT1 which is known to be involved in DBQD, samples from consanguineous siblings were subjected to WES, where a homozygous mutation (c.1792C>T [p.Arg598Cys]) were identified (confirmed with Sanger sequencing), as well as the same mutation and the wild type allele in their

unaffected heterozygous carrier parents. With this information, the remaining individuals were screened for mutations in *XYLT1* using Sanger sequencing. They were able to uncover two splice site mutations in introns 5 (c.1290-2A>C) and 7 (c.1588-3C>T) just prior to the intron-exon junction, one missense mutation in exon 9 (c.1792C>T [p.Arg598Cys]), and two frameshift mutations resulting in a premature stop codon. One was in exon 3 (c.439C>T [p.Arg147*]) and the other was one base away from one of our patient's mutations in exon 1 [p.Pro93Alafs*69]. The intron 5 splice mutation (c.1290-2A>C) was only 1 base away from the homozygous splice found in our patient 13-1 (c.1290-1 G>A). No mutations were found in the rest of the patient samples and their disorder was classified as having an unknown molecular basis.

Schreml et al. also published in 2014 on two affected siblings, both with homozygous missense mutations in *XYLT1* (c.C1441T[p.R481W]), causing the replacement of the highly conserved arginine in that position with tryptophan. (Schreml et al. 2014). The researchers took advantage of WES and linkage analysis with unaffected family members to find the mutations, which were confirmed with Sanger sequencing. However, there were only two patients, and they were from a consanguineous family, which made it easier to confirm the causative mutation as they both shared the same alleles passed down from their parents. It is also important to note the location of the mutation was downstream in the protein sequence, within glycosylation family 14 domain. Further research indicated that while the mutation causes a change in the structural composition of extracellular matrix (ECM) in the patients, it did not lower the levels of XYLT1. Comparisons were made to other skeletal dysplasia disorders such as various Ehlers-Danlos subtypes, as well as Larsen-

like short stature syndrome, caused by the *B3GALT3*, *B3GALT6*, and *B4GALT7* gene, as well as research pointing to *XYLT1* mutations involved in osteoarthritis in a zebrafish model.

In both cases, the majority of the mutations were found in exons 2 through 12 or were intronic, with none found in the promoter region or the 5' end of exon 1. It is also important to note that in each case, the original mutations were found in a consanguineous family, significantly raising the probability that the mutation could be homozygous.

Faust et al published on the *XYLT1* promoter region in 2014, in which they also found the newly found region uncovered in our laboratory. Their consensus sequence was based on sequencing resulting from gDNA taken from healthy samples. Of these, close to half of their results indicated homozygosity throughout this region. They were also able to show no promoter activity dependence on the length of the variable trinucleotide repeat. Of note is the fact that they were also unable, even with slowdown and 7-deaza, to successfully amplify through the entire region in some instances, and were unable to identify the cause of the amplification issues. Within the different numbers of variable repeats, they also had a high percentage of homozygous samples (both alleles with the same number of GGC repeats). However, it is possible that what is being seen as a homozygous mutation is in fact the same phenomenon seen in our cohort, that of preferential amplification. When our laboratory first began investigating BSS, it was originally thought that the mutations found via Sanger sequencing were homozygous (patients 5-1 and 16-1). It was only due to the tenacity and skill of one of the BCL members, Deborah Stabley, that the second allele which showed almost no amplification was uncovered. It is possible they appeared as

homozygous mutations due to aberrant methylation patterns on one allele, with one allele overpowering the methylated allele during PCR amplification.

The latest publication to discuss mutations in *XYLT1* is from Koningsbruggen et al, and it discusses one patient. With the use of array comparative genomic hybridization (aCGH) which is a cytogenetic technique that detects genome-wide copy number variation), as well as with Sanger sequencing, the patient was found to harbor a compound heterozygous mutation. The maternally inherited allele contained a 3.3Mb deletion on chromosome 16 (including the *XYLT1* locus), while the paternally inherited allele contained an 18 base pair deletion in the intron 7/exon 8 junction c.1588-10_1595del, which results in the loss of the correct splice site. This is the first published case of a compound heterozygote in *XYLT1*. Three of our patients inherited an allele with a 3Mb deletion, patients 3-1, 17-1 and 17-2.

The common thread in all these findings is the large number of homozygous mutations and the problem of amplification. Based on the results of the research in our laboratory, it is quite likely that the amplification issues and appearance of homozygosity is actually the presence of aberrant methylation. The further downstream of exon 1 in *XYLT1*, the more likely it was that the researchers were able to find the causative mutations in their patients. Another important clue in the puzzle is when we look at the c.-5 C>G SNP (r.s.1107041807) and its representation in the genome databases. In 1000 Genomes, the typical allele count when giving frequencies is 2000 (2 alleles for each of the 1000 people). For this particular SNP, the allele count is only 408, significantly lower than it should be, with an allelic frequency of 8%. It is possible that, due to presence of the high number of GGC repeats, as well as the G

quadruplexes in this region, amplification in this region of the gene is much more refractory than expected even in a large control population.

In our cohort, we were eventually able to identify inherited causative variants for our patients. It was shown in the case of patients 3-1, 5-1, 16-1, 17-1 and 17-2, the presence of a genomic mutation on one allele, and aberrant CpG methylation on the other allele. We saw this as a unique feature in our patients and their unaffected carrier parents, while the controls showed no methylation, which is consistent with *XYLT1* not being an imprinted gene. With this information we now know that BSS can also be caused by two heterogeneous recessive alleles. In one patient, we were able to show, for the first time that we are aware of, the presence of two methylated alleles as the cause of BSS, or of any recessive disorder currently found in literature.

5.4 Future Directions – Questions which Need to be Answered

There are a number of questions which need to be answered. First, while we now know the presence of the allele with aberrant methylation was the second hit we had been searching for, we still do not know why this methylation is present in the patients of our cohort and their unaffected carrier parents. It is unlikely that there is a global methylation problem because this would likely be embryonically lethal. However, the methylation patterns are dictated throughout growth and development, by an exquisite control system. This would need to be delved into further. We know there are interactions between *XYLT1* and *XYLT2* which require further elucidation. Perhaps, as was suggested by Gotting, the presence of the second isoform in humans allows for a greater flexibility in levels of functional *XYLT1*.

Within our own cohort, there is a high probability of a genetic mutation located in an as yet unknown control region or gene that controls methylation of XYLT1.

Since global dysregulation of methylation would almost surely be lethal at an organismal level, the problem must be local methylation dysregulation. There are many clues which could lead to the next level of understanding. While it was mentioned briefly herein, the methylated allele, in each case, was associated with a G at the c.-5 position. While it may be a coincidence, it is certainly worth investigating further as it may reveal an ancestral allele. Based on our findings, it is likely the patients which were undiagnosed in the Bui cohort may be suffering from a methylation defect. This would explain the inability to find a molecular explanation for the disorder. It would be beneficial to investigate the methylation status of those patients with a disorder of unknown etiology.

REFERENCES

Ambrosius, M., Kleesiek, K., & Götting, C. (2009). The xylosyltransferase I gene polymorphism c. 343G> T (p. A115S) is associated with decreased serum glycosaminoglycan levels. *Clinical Biochemistry, 42*(1), 1-4.

Behnia, F., Parets, S. E., Kechichian, T., Yin, H., Dutta, E. H., Saade, G. R., . . . Menon, R. (2015). Fetal DNA methylation of autism spectrum disorders candidate genes: Association with spontaneous preterm birth. *American Journal of Obstetrics and Gynecology, 212*(4), 533. e1-533. e9.

Biancalana, V., Glaeser, D., McQuaid, S., & Steinbach, P. (2015). EMQN best practice guidelines for the molecular genetic testing and reporting of fragile X syndrome and other fragile X-associated disorders. *European Journal of Human Genetics, 23*(4), 417-425.

Botstein, D., & Risch, N. (2003). Discovering genotypes underlying human phenotypes: Past successes for mendelian disease, future approaches for complex disease. *Nature Genetics, 33*, 228-237.

Braunholz, D., Obieglo, C., Parenti, I., Pozojevic, J., Eckhold, J., Reiz, B., . . . Vodopiutz, J. (2015). Hidden mutations in cornelia de lange syndrome limitations of sanger sequencing in molecular diagnostics. *Human Mutation, 36*(1), 26-29.

Brenet, F., Moh, M., Funk, P., Feierstein, E., Viale, A. J., Socci, N. D., & Scandura, J. M. (2011). DNA methylation of the first exon is tightly linked to transcriptional silencing. *PloS One, 6*(1), e14524.

Breton, C., Bettler, E., Joziasse, D. H., Geremia, R. A., & Imberty, A. (1998). Sequence-function relationships of prokaryotic and eukaryotic galactosyltransferases. *Journal of Biochemistry, 123*(6), 1000-1009.

Bui, C., Huber, C., Tuysuz, B., Alanay, Y., Bole-Feysot, C., Leroy, J. G., . . . Cormier-Daire, V. (2014). XYLT1 mutations in desbuquois dysplasia type 2. *The American Journal of Human Genetics, 94*(3), 405-414.

Bui, C., Ouzzine, M., Talhaoui, I., Sharp, S., Prydz, K., Coughtrie, M. W., & Fournel-Gigleux, S. (2010). Epigenetics: Methylation-associated repression of heparan

sulfate 3-O-sulfotransferase gene expression contributes to the invasive phenotype of H-EMC-SS chondrosarcoma cells. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology, 24*(2), 436-450. doi:10.1096/fj.09-136291 [doi]

Conerly, M., & Grady, W. M. (2010). Insights into the role of DNA methylation in disease through the use of mouse models. *Disease Models & Mechanisms, 3*(5-6), 290-297. doi:10.1242/dmm.004812 [doi]

de Vooght, K. M., van Wijk, R., & van Solinge, W. W. (2009). Management of gene promoter mutations in molecular diagnostics. *Clinical Chemistry, 55*(4), 698-708. doi:10.1373/clinchem.2008.120931 [doi]

Eames, B. F., Yan, Y., Swartz, M. E., Levic, D. S., Knapik, E. W., Postlethwait, J. H., & Kimmel, C. B. (2011). Mutations in fam20b and xylt1 reveal that cartilage matrix controls timing of endochondral ossification by inhibiting chondrocyte maturation. *PLoS Genet, 7*(8), e1002246.

Edwards, J. R., O'Donnell, A. H., Rollins, R. A., Peckham, H. E., Lee, C., Milekic, M. H., . . . Bestor, T. H. (2010). Chromatin and sequence features that define the fine and gross structure of genomic methylation patterns. *Genome Research, 20*(7), 972-980. doi:10.1101/gr.101535.109 [doi]

Fairbrother, L. C., Cytrynbaum, C., Boutis, P., Buiting, K., Weksberg, R., & Williams, C. (2015). Mild angelman syndrome phenotype due to a mosaic methylation imprinting defect. *American Journal of Medical Genetics Part A,*

Faust, I., Roch, C., Kuhn, J., Prante, C., Knabbe, C., & Hendig, D. (2013). Human xylosyltransferase-I–a new marker for myofibroblast differentiation in skin fibrosis. *Biochemical and Biophysical Research Communications, 436*(3), 449-454.

Faust, I., Boker, K. O., Lichtenberg, C., Kuhn, J., Knabbe, C., & Hendig, D. (2014). First description of the complete human xylosyltransferase-I promoter region. *BMC Genetics, 15*, 129-014-0129-0. doi:10.1186/s12863-014-0129-0 [doi]

Feng, S., Cokus, S. J., Zhang, X., Chen, P. Y., Bostick, M., Goll, M. G., . . . Jacobsen, S. E. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proceedings of the National Academy of Sciences of the United States of America, 107*(19), 8689-8694. doi:10.1073/pnas.1002720107 [doi]

Frey, U. H., Bachmann, H. S., Peters, J., & Siffert, W. (2008). PCR-amplification of GC-rich regions:'slowdown PCR'. *Nature Protocols, 3*(8), 1312-1317.

Götting, C., Kuhn, J., & Kleesiek, K. (2007). Human xylosyltransferases in health and disease. *Cellular and Molecular Life Sciences, 64*(12), 1498-1517.

Gotting, C., Muller, S., Schottler, M., Schon, S., Prante, C., Brinkmann, T., . . . Kleesiek, K. (2004). Analysis of the DXD motifs in human xylosyltransferase I required for enzyme activity. *The Journal of Biological Chemistry, 279*(41), 42566-42573. doi:10.1074/jbc.M401340200 [doi]

Grebner, E. E., Hall, C. W., & Neufeld, E. F. (1966). Glycosylation of serine residues by a uridine diphosphate-xylose: Protein xylosyltransferase from mouse mastocytoma. *Archives of Biochemistry and Biophysics, 116*, 391-398.

Heard, E., & Martienssen, R. A. (2014). Transgenerational epigenetic inheritance: Myths and mechanisms. *Cell, 157*(1), 95-109.

Hendrich, B., & Tweedie, S. (2003). The methyl-CpG binding domain and the evolving role of DNA methylation in animals. *TRENDS in Genetics, 19*(5), 269-277.

Hinsdale, M. E. (2014). Xylosyltransferase I, II (XYLT1, 2). *Handbook of glycosyltransferases and related genes* (pp. 873-883) Springer.

Huegel, J., Sgariglia, F., Enomoto-Iwamoto, M., Koyama, E., Dormans, J. P., & Pacifici, M. (2013). Heparan sulfate in skeletal development, growth, and pathology: The case of hereditary multiple exostoses. *Developmental Dynamics,242*(9), 1021-1032.

Huppert, J. L., & Balasubramanian, S. (2007). G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Research, 35*(2), 406-413. doi:gkl1057 [pii]

Jochmann, K., Bachvarova, V., & Vortkamp, A. (2014). Heparan sulfate as a regulator of endochondral ossification and osteochondroma development. *Matrix Biology, 34*, 55-63.

Knudson, C. B., & Knudson, W. (2001). Cartilage proteoglycans. *Seminars in Cell & Developmental Biology, , 12*(2) 69-78.

Lelieveld, S. H., Spielmann, M., Mundlos, S., Veltman, J. A., & Gilissen, C. (2015). Comparison of exome and genome sequencing technologies for the complete capture of protein coding regions. *Human Mutation,*

Li, R., Mav, D., Grimm, S. A., Jothi, R., Shah, R., & Wade, P. A. (2014). Fine-tuning of epigenetic regulation with respect to promoter CpG content in a cell type-specific manner. *Epigenetics, 9*(5), 747-759.

Mis, E. K., Liem, K. F., Kong, Y., Schwartz, N. B., Domowicz, M., & Weatherbee, S. D. (2014). Forward genetics defines Xylt1 as a key, conserved regulator of early chondrocyte maturation and skeletal length. *Developmental Biology,385*(1), 67-82.

Mizumoto, S., Yamada, S., & Sugahara, K. (2015). Mutations in biosynthetic enzymes for the protein linker region of Chondroitin/Dermatan/Heparan sulfate cause skeletal and skin dysplasias. *BioMed Research International, 2015*

Muir, H. (1995). The chondrocyte, architect of cartilage. biomechanics, structure, function and molecular biology of cartilage matrix macromolecules. *Bioessays, 17*(12), 1039-1048.

Muller, B., Prante, C., Kleesiek, K., & Gotting, C. (2009). Identification and characterization of the human xylosyltransferase I gene promoter region. *The Journal of Biological Chemistry, 284*(45), 30775-30782. doi:10.1074/jbc.M109.016592 [doi]

Munns, C., Fahiminiya, S., Poudel, N., Munteanu, M., Majewski, J., Sillence, D., . . . Hinsdale, M. (2015). Homozygosity for frameshift mutations in XYLT2 result in a spondylo-ocular syndrome with bone fragility, cataracts, and hearing defects. *The American Journal of Human Genetics, 96*(6), 971-978. doi:http://dx.doi.org/10.1016/j.ajhg.2015.04.017

Okayama, H., Curiel, D. T., Brantly, M. L., Holmes, M. D., & Crystal, R. G. (1989). Rapid, nonradioactive detection of mutations in the human genome by allele-specific amplification. *J Lab Clin Med, 114*(2), 105-113.

Olson, S. K., Bishop, J. R., Yates, J. R., Oegema, K., & Esko, J. D. (2006). Identification of novel chondroitin proteoglycans in caenorhabditis elegans: Embryonic cell division depends on CPG-1 and CPG-2. *The Journal of Cell Biology,173*(6), 985-994. doi:jcb.200603003 [pii]

Quon, G., Lippert, C., Heckerman, D., & Listgarten, J. (2013). Patterns of methylation heritability in a genome-wide analysis of four brain regions. *Nucleic Acids Research, 41*(4), 2095-2104. doi:10.1093/nar/gks1449 [doi]

Rabbani, B., Tekin, M., & Mahdieh, N. (2014). The promise of whole-exome sequencing in medical genetics. *Journal of Human Genetics, 59*(1), 5-15.

Reis-Filho, J. S. (2009). Next-generation sequencing. *Breast Cancer Res, 11*(Suppl 3), S12.

Robertson, K. D. (2005). DNA methylation and human disease. *Nature Reviews Genetics, 6*(8), 597-610.

Robertson, K. D. (2002). DNA methylation and chromatin - unraveling the tangled web. *Oncogene, 21*(35), 5361-5379. doi:10.1038/sj.onc.1205609 [doi]

Schreml, J., Durmaz, B., Cogulu, O., Keupp, K., Beleggia, F., Pohl, E., . . . Nürnberg, G. (2014). The missing "link": An autosomal recessive short stature syndrome caused by a hypofunctional XYLT1 mutation. *Human Genetics,133*(1), 29-39.

Schuster, S. C. (2007). Next-generation sequencing transforms today's biology. *Nature, 200*(8), 16-18.

Schuster, S. C. (2007). Next-generation sequencing transforms today's biology. *Nature, 200*(8), 16-18.

Singal, R., & Ginder, G. D. (1999). DNA methylation. *Blood, 93*(12), 4059-4070.

Smith, Z. D., & Meissner, A. (2013). DNA methylation: Roles in mammalian development. *Nature Reviews Genetics, 14*(3), 204-220.

Sugahara, K., Mikami, T., Uyama, T., Mizuguchi, S., Nomura, K., & Kitagawa, H. (2003). Recent advances in the structural biology of chondroitin sulfate and dermatan sulfate. *Current Opinion in Structural Biology, 13*(5), 612-620.

van Koningsbruggen, S., Knoester, H., Bakx, R., Mook, O., Knegt, L., & Cobben, J. M. (2015). Complete and partial XYLT1 deletion in a patient with neonatal short limb skeletal dysplasia. *American Journal of Medical Genetics Part A*.

Appendix A

IRB APPROVAL FORM

Nemours Office of Human Subjects Protection
10140 Centurion Parkway North
Jacksonville, FL 32256
Phone: 904-697-4023 Fax: 904-697-4024

MEMORANDUM

DATE:	January 18, 2016
TO:	Karen Gripp, MD
FROM:	Nemours IRB 1
STUDY TITLE:	[212117-16] Delineation of the genetic etiology and biologic mechanism in distinctive physical and neurodevelopmental conditions
IRB #:	212117
SUBMISSION TYPE:	Continuing Review/Progress Report
ACTION:	APPROVED
APPROVAL DATE:	January 15, 2016
EXPIRATION DATE:	January 14, 2017

Thank you for your submission of Continuing Review/Progress Report materials for this research study. Your submission received Expedited Review based on the applicable federal regulation and meets all DHHS [and FDA] criteria for approval. The above-referenced research study is approved under the Expedited Review Categories 2, 5 and 7.

The IRB has determined that:

- This is "Research not involving greater than minimal risk per 45CFR46.404 and 21CFR50.51".
- Informed Consent or Parental Permission is required prior to initiation of any research procedures using only the most current IRB approved form(s) posted as a Board Document in IRBNet. All approved study documents can be accessed through "Designer" by clicking "Review Details" in IRBNet. **All approved documents can be found under "Board Documents".**
- The permission of one parent is sufficient. A person who is not a parent may not give permission without prior IRB review and approval.
- Assent of minors is required prior to initiation of any research procedures, using only the most current assent form(s) posted as a Board Document in IRBNet.
- A signed copy of the Parental Permission/Informed Consent form must be included in the Nemours' medical record. Research data may also be included into the Nemours medical record.
- The amendment application to the approved protocol was approved as submitted.
- To continue, the research requires IRB review and approval on an annual basis. January 14, 2017 is the last day that research may be conducted.
- The Principal Investigator is responsible for the timely submission of the continuing review application. Please post this date on your research calendar. For submissions that can be reviewed by expedited review, please be reminded that applications for continuing review need to be submitted at least 2 weeks ahead of the expiration date to give sufficient lead time for IRB review.

The IRB requires that a copy of the participant brochure: "Becoming A Research Volunteer" will be given to every individual enrolled in a research study. The PDF file for this document has been attached to this study as a Board Document.

Reviewed/approved documents in this submission: