THE USE OF WHOLE EXOME SEQUENCING TO DETECT NOVEL GENETIC DISORDERS: TWO CASES AND AN ASSESSMENT OF THE TECHNOLOGY

by

Lauren Westerfield

BS Psychology, Florida State University, 2011

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2013

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This thesis was presented

by

Lauren Westerfield

It was defended on

March 26, 2013

and approved by

Robin Grubs, Ph.D., Assistant Professor, Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh

M. Michael Barmada, Ph.D., Associate Professor, Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh

Thesis Director: Gerard Vockley, M.D., Ph.D., Professor, Department of Pediatrics, School of Medicine, University of Pittsburgh

THE USE OF WHOLE EXOME SEQUENCING TO DETECT NOVEL GENETIC DISORDERS: TWO CASES AND AN ASSESSMENT OF THE TECHNOLOGY

Lauren Westerfield, MS

University of Pittsburgh, 2013

ABSTRACT

Over the past several years whole exome sequencing (WES) by high-throughput sequencing of target-enriched genomic DNA has become both technically feasible and financially practical as a means of studying Mendelian disorders. It is also entering the clinical realm as a powerful diagnostic tool for cases that have eluded answers and a cost effective one for cases with a suspected genetically heterogeneous disorder or set of differentials. This thesis examines the strategies for use and impact of such a technology in both the research and clinical setting. It presents an analysis of two cases in which WES was used to determine the causative mutation in the phenotype of an unknown/undiagnosed genetic disorder. The results demonstrate the strengths and limitations of variant filtering strategies, the need for co-segregating familial samples when possible, the value of a detailed phenotypic picture and family history, and value of functional studies in confirming the pathogenicity of candidate variants. In the first case report, WES succeeded in narrowing the candidate list to a manageable size for two sibs affected in the neonatal period with seizures, encephalopathy, and thrombocytopenia, and who died at a few months of age. Sequencing data on the parents and unaffected sibling is needed to elucidate the pathogenic mutations. In the second case report, WES detected a strong candidate mutation in *NDUFAF6,* a complex 1 assembly factor. Given the patient's presentation with multi-organ dysfunction, dramatic skeletal myopathy, and degenerative course suggestive of a mitochondrial disorder, complex 1 deficiency was suspected but Sanger sequencing failed to confirm the

mutation. This thesis also examined the ethical and practical considerations involved in incorporating WES into clinical practice and its impact on public health, namely improved treatment options for patients and an improved knowledge of the relationship between genetics and disease phenotypes.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

It takes a village to raise a child, and similarly this project is the culmination of the efforts and attentions of many people beyond myself. Naturally, I would like to thank Dr. Jerry Vockley for allowing me to be a part of such a fascinating project at such a crucial stage in the life of this technology. Dr. Lina Gonzalez, for working side-by-side with me on these patients and without whose knowledge of resources none of the analysis would have been possible. Dr. You Li of Dr. Cynthia Lo's lab and Rahil of the Greater Pittsburgh Cytogenetics Lab for getting what must have been extremely nasty raw data into a legible, analyzable form. Importantly, I would like to acknowledge the patients and their families presented in this thesis, for putting their trust in us to find them the answers they so dearly want in the face and aftermath of such suffering as a result of something as deceptively simple as a mistake in a DNA sequence. I would like to thank Robin Grubs and Betsy Gettig for their faith that I would be a worthwhile contributor to, and participant in, the University of Pittsburgh's Genetic Counseling Program. I truly believe that I found my way to the place where I needed to be at the time that I needed to be here and I cannot imagine any other program giving me what I found in Pittsburgh. On that note, I would like to thank Professor Murray Krantz for just happening to mention genetic counseling as a career option in passing "that one day" in class. I would also like to thank the lovely ladies of "the Herd" for sharing these two wonderful, crazy years with me. I could not ask for a better group of girls in the class of 2013. Finally, and most importantly, I would like to thank my Mom and Dad for the

love, care, and support they have provided me with throughout my life – and the good genes they gave me! I wouldn't be who I am today without such amazing parents like you and words can't express how grateful I am for all that you have done and continue to do for me, including knowing when to take a leap of faith and let me stand of my own two feet. I love you both so much.

1.0 INTRODUCTION

In 2001, after 13 years of effort, the first human genome was sequenced by the Human Genome Project at an estimated cost of $2.7 billion.[1] Since that time, the development of massively parallel pyrosequencing platforms has allowed the adoption of high-throughput genomic analysis known as next-generation sequencing, NGS, increasing the capacity to generate and analyze larger quantities of genotypic and phenotypic information than ever before. Over time, improvements to technology have also lowered the cost and improved the time required to sequence a human exome. Though until recently confined to a research setting, whole-exome sequencing (WES) is now available on a clinical basis through laboratories such as Baylor Genetics, Ambry Genetics, and GeneDx. WES has so far been shown to be a powerful tool in elucidating the causes behind many Mendelian diseases and in the clinical setting promises to provide a more effective method of providing patients with answers. Improved knowledge of pathogenic variants and their disease associations can allow one to prepare for, avoid, or treat the negative impacts that they can have on health, lifespan, and offspring. As with any new technology there are often many aspects to consider and challenges to overcome in its use. This project aimed to explore these issues, including an evaluation of the technology itself, its utility in discovering candidate genes for novel genetic syndromes and the data analysis process, and its impact on the field of genetics using two case studies of individuals affected by an unknown disorder.

1.1 THE GENOME

The human genome is composed of roughly 3 billion nucleotide base pairs arranged into approximately 30,000 genes. Each gene contains both protein-coding and non-coding regions. Coding regions (exons) contain information for the construction of the amino-acid sequence of the protein product and structural or regulatory RNA species. Non-coding regions include introns and the 3'- and 5' regions of each gene; their function is unknown at this time. Most variation between humans occurs in the non-coding DNA regions and in degenerate positions in amino acid codons that do not change the intended identity of the amino acid. Humans vary on average ever 1 out of 100 nucleotides and most of these variations occur frequently in the population with little or no effect on protein function. As such, they are called polymorphisms. Mutations in the genetic sequence are more likely to have detrimental effects if they result in a shift of the reading frame, non-synonymous substitution of one amino acid for another (particularly amino acids with vastly different chemical properties), insertion of a premature stop codon resulting in a truncation of the protein product, or loss of a stop codon. Though protein-coding genes comprise only about 1% of the genome, they harbor about 85% of the mutations with large effects on disease-related traits.[1]

1.2 THE DEVELOPMENT OF WES

After the completion of the Human Genome project it was still too expensive to sequence large numbers of human genomes. Researchers instead demonstrated that it was possible to capture and sequence the protein-coding exons from human genomes, leading then to the analysis of the

complete set of exons in the genome, labeled the exome. By 2009 researchers had used WES to discover the genetic basis for Bartter syndrome, Miller syndrome, and Kabuki syndrome.[2]. In 2011, exome sequencing was used to determine the basis for a previously undescribed and idiopathic disorder later named Ogden syndrome, which was shown to be located on the X chromosome and result from a defect in the amino-terminal acetylation of proteins. [3] Since 2009, more than 20 causative genes have been identified and the number is only expected to grow exponentially.[1]

1.3 IDENTIFYING DISEASE GENES

Until the advent of WES, most studies aiming to identify new genetic causes of disease used linkage analysis (positional cloning). These usually identified a genomic interval spanning 0.5-10 cM which could contain up to 300 genes. By 2009 that strategy identified less than 2000 genes responsible for less than 4000 diseases, with some genes being linked to multiple conditions.[4] We know also that genome of the human species as a whole is subject to numerous new pathogenic mutations each year. The number of known mutations in the human nuclear genes that either cause or are associated with heritable diseases exceeds 100,000 in more than 3700 different genes.[1] Even so, a large number of genes responsible for the approximately 7000 Mendelian diseases still remain unidentified and there are undoubtedly more Mendelian disorders that have not yet been named or discovered. When it comes to the process of using WES to discover candidate genes for such disorders, there are a number of factors to take into consideration.

1.3.1 DNA Capture and Enrichment Technology

The three major next-generation sequencing platforms are Illumina, Nimblegen, and Agilent. Each of these platforms is compatible with the major commercial options for the first step of WES, which is enriching the exonic sequences. The sequencing platform kits tend to contain exons from the consensus coding sequence project, which currently comprises 176,266 exons from 18,409 genes, as well as additional sequences. [4] Each company also has developed its own exome enrichment platform (Agilent's SureSelect Human All Exon 50Mb, Roche/Nimblegen's SeqCap EZ Exome Library v.2.0, and Illumina's TruSeq Exome Enrichment), which differ in design and experimental parameters that can affect variant discovery. Clark et al. 2011 performed a systematic analysis of these differences. [5]

1.3.1.1 Nimblegen

Uses DNA for capture of targeted genomic sequences. The platform contains overlapping baits that cover target bases multiple times, resulting in the highest density coverage of the three platforms. It covers a greater portion of miRNAs compared to other enrichment platforms.

1.3.1.2 Agilent

Uses RNA for capture of targeted genomic sequences, where baits reside immediately adjacent to each other across target exon intervals. It provides better coverage of genes in the Ensembl database.

1.3.1.3 Illumina

Uses DNA for capture of targeted genomic sequences and relies on paired-end reads to extend outside bait sequences and fill gaps. The majority of targets unique to this platform cover untranslated regions (UTRs).

1.3.1.4 Broad observations

Each platform contains 4.4-28Mb of unique target region. Nimblegen and Agilent share more with each other (38.8Mb) than either does with Illumina (30.3 Mb and 33.3 Mb respectively). 29.45Mb were found to be targeted by all three platforms. Coverage of mRNA coding exons in both RefSeq and Ensemble were similar between all platforms. Nimblegen enriched a higher percentage of targeted bases, which Illumina and Agilent enriched a higher total number of bases at higher read counts. A higher density design, targeting a smaller genomic interval, results in higher efficiency. Lower density designs required substantially larger amounts of sequencing, as efficient baits became saturated at 40M (Nimblegen) versus 50M (Agilent) and 60M (Illumina) reads. The percentage of off-target enrichments correlated strongly with this trend.

A potential source of inefficiency comes from areas with high GC or AT content, as low coverage in these areas has been observed [6]. All three platforms showed a sharp drop in read depth as GC content increased from 60% to 80%. As GC content dropped from 40% to 20%, Illumina and Nimblegen diminished with lower read depth over those targets, where the Agilent platform displayed only a slight reduction in read depth. This was felt to be due to its lower number of PCR cycles, longer baits, and/or the use of RNA probes.

In the detection of singly nucleotide variants (SNVs), concordance rates for a normalized 80M read exome data set compared to the SNP Chip were 99.3% for Agilent, 99.5% for Nimblegen, and 99.2% for Illumina. Allelic balance (AB) was calculated by determining the

ratio of reference base calls over the total number of calls at every SNV with a quality score of 30 or better (99.9% probability of an accurate call). For Agilent, AB=0.55, and 0.53 for both Nimblegen and Illumina. Therefore the reference biases were not strong, but explained some of the discordance with the SNP Chip data set. No significant different in the ratio of heterozygous to homozygous variants were observed between platforms. In shared regions, Nimblegen captured the most SNVs with the lowest number of reads, followed by Agilent and then Illumina. This demonstrates a correlation between bait density and sensitivity to SNV detection, and Nimblegen was also more effective at detecting SNVs in low-complexity, hard-to-target regions. Agilent detected unique SNVs most often in introns, as its baits sometimes extend farther outside of exon targets than the other platforms.

Coverage of regions containing insertions and deletions (indels) largely match coverage in other targeted regions. Small insertions and deletions ranging from -84bp to +18 bp were detected at a frequency of 12.5-14.5% that of SNVs. At lower read counts, more indels were detected after Agilent enrichment than Illumina. Past 50M reads, the reverse was true. In shared and RefSeq regions, Nimblegen had the highest sensitivity to detecting indels at lower read counts, while Agilent enrichment led to the largest number of detected indels at every read count in Ensembl CDS exons.

1.3.2 Sequencing Platforms

1.3.2.1 Applied Biosystems[7]

Applied Biosystems™ by Life Technologies offers the SOLiD sequencing platform, which stands for Sequencing by Oligonucleotide Ligation and Detection. Four fluorescently labeled di-base probes compete for ligation to the sequencing primer. Specificity for the di-base probe is

done by interrogating every first and second base in each ligation reaction, and the eventual read length is determined over multiple rounds of ligation, detection, and cleavage. Following a series of these ligation cycles, the extension product is removed and the template is reset with a primer complementary to the n-1 position for a second round of ligation cycles. Five rounds of primer reset are completed for each sequence tag. This allows nearly every base to be queried in two different ligation reactions by two different primers, improving the accuracy of nucleotide base calls. Variations from the reference sequence display as a fluorescent color change; sequencing errors would therefore show as one change while accurate calls would show two.

1.3.2.2 Illumina[8]

Illumina's sequencing platform uses sequencing by synthesis (SBS) technology to generate exome data. The technology is able to detect single bases as they are added to DNA strands using a reversible terminator-based method. The fluorescent terminator is imaged as deoxyribonucleotide triphosphate (dNTP) is added, and then cleaved so that the next base can be added and imaged. Incorporation bias is minimized by competition, as all four reversible terminator-bound dNTPs are present during each sequencing cycle. SBS supports both single read and paired end libraries. The platform combines short-insert paired-end capabilities as well as long-insert paired-end reads to fully characterize the genome being sequenced.

1.3.2.3 IonTorrent™ [9-11]

Ion Torrent is a long-read high-density semiconductor sequencing platform developed by Roche 454 Life Sciences in partnership with DNA Electronics. It is based on the detection of hydrogen ions that are released during the polymerization of DNA and represents another method of SBS. As the dNTP is incorporated into the DNA strand complementary to the template, the release of

a hydrogen ion triggers an ISFET ion sensor and records that a reaction has occurred. In this sequencing technology, unlike the others, no modified nucleotides or optics are used. Instead, only a single species of dNTP is used at a time compared to the simultaneous presence of all four dNTPs in other platforms. If the dNTP is not complementary to the template nucleotide, there is no reaction. The per base accuracy was 99.6% based on 50 base reads with 100Mb per run, with read lengths of 100 base pairs. One of the strengths of this technology is a rapid sequencing speed and low cost possible by avoiding the modified nucleotides and optical measurements. With this system it is difficult to enumerate long repeats, as multiple ions will be released as multiple nucleotides are incorporated and it is difficult to distinguish signals from a high repeat sequence from ones of a similar but different number (such as 7 repeats instead of 9). It also has a shorter read length and lower throughput than other sequencing technologies, though increasing the density of the chip might change this.

1.3.3 Data Analysis Strategies

The first major hurdle to overcome when analyzing a set of exome sequencing data is the sheer number of variants that are present compared to the reference sequences. Based on the literature, a researcher can expect to be confronted with anywhere from 20-30,000 variants in a single exome sequence. Of these, approximately 10,000 will be predicted to result in nonsynonymous amino acid substitutions, splice-site alterations, insertions, or deletions.[4] Filtering these results further requires a set of assumptions about which variants are more likely to be deleterious.

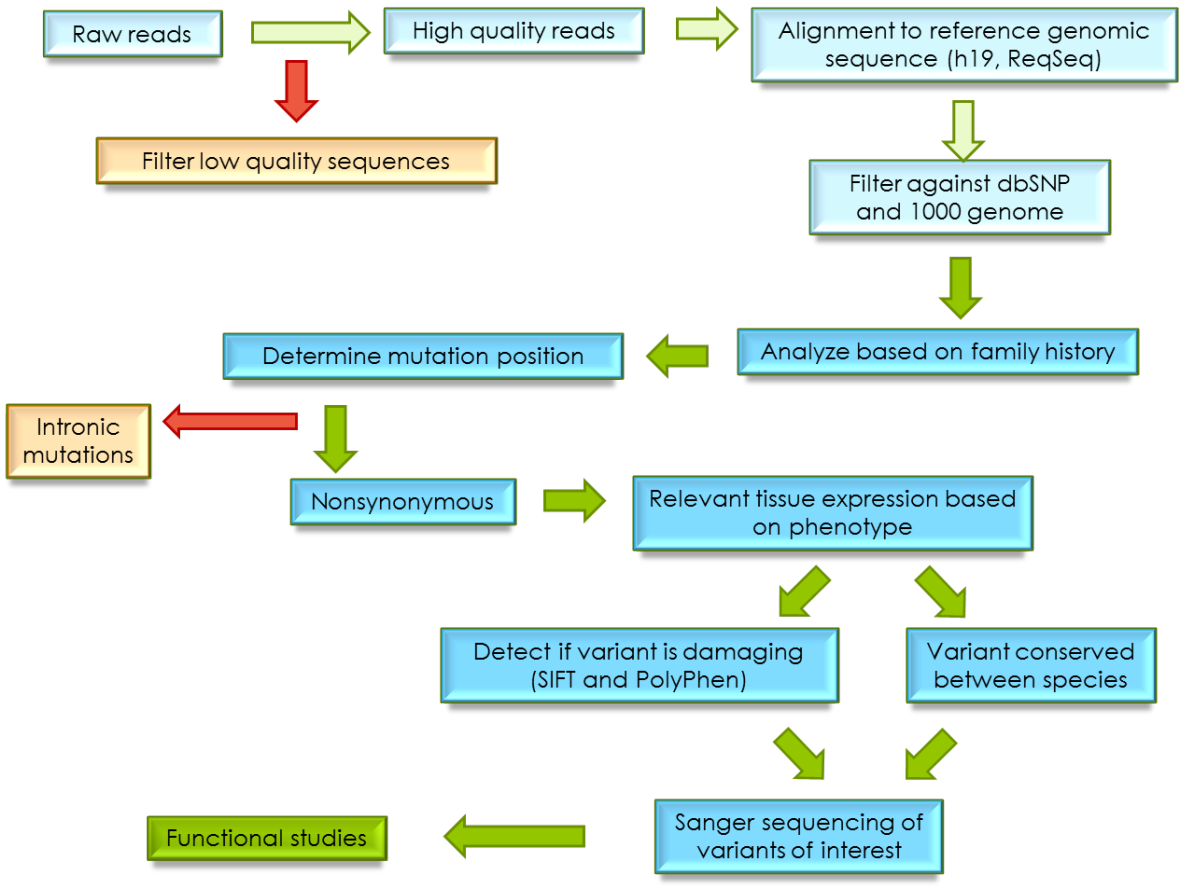

Figure 1 Sample pipeline for variant analysis

Variants reported to be common in the general population are not likely to be responsible for Mendelian disease. Such variants can be found in databases such as dpSNP, the 1000 Genome Project, and in-house exome databases. The caveat for using these databases is that there is a change that information on certain variants is mislabeled, though databases make efforts to correct such errors when they encounter them. For example, of the more than 17 million SNPs in the human genome documented in dbSNP, the false-positive rate is estimated at 15-17% [12]. Computational algorithms are available online that can predict the pathogenicity of variants and can therefore allow variants that are predicted to be benign to be removed. Two examples of these databases are SIFT and PolyPhen. However, computational algorithms in general have been shown to have high false-positive and false-negative rates, likely at least 20%

for WES data.[4] Therefore, this kind of filtering is more useful once other filters have already been applied to narrow the list of candidate genes to a manageable size. The NHLBI Exome Variant Server, composed of the data from the NHLBI Grand Opportunity Exome Sequencing Project, allows researchers to check variants that they have found against a database of 6503 exomes in the current version. The goal of the ESP data set is to the frequency of counts of specific variants without regard to phenotype. The data set was selected to contain controls, as well as extremes of specific traits (LDL levels and blood pressure) and specific diseases (early-onset myocardial infarction, early-onset stroke, and lung disease). Once variants of interest have been identified, it can also be useful to determine whether the gene is one that is conserved across evolution, and therefore a more functionally important gene, using the UCSC Genome Browser. The mutation(s) of greatest interest can then be confirmed using Sanger sequencing, particularly if the read coverage is relatively low, and, if possible, functional studies can be performed on tissue samples to confirm the physiologic effects of the mutation, such as reduced enzyme activity.

In the event that multiple unrelated individuals with the same phenotype are available for sequencing, comparison of their common variants can be extremely useful as a filter. The assumption is that sequence variants unrelated to the disease of interest will be randomly distributed in the exome; thus the likelihood of these individuals sharing the same variants by random chance becomes extremely low. This strategy cannot be used in a blanket approach, however, as it neglects the possibility of genetic heterogeneity. When determining the genetic basis of Kabuki syndrome in 10 unrelated individuals, only one gene was found to have at least 1 non-synonymous/splice site/indel mutation in every individual. The gene, *MUC16*, codes for a

protein that provides a lubricating barrier against particles at mucosal surfaces and was clearly a false-positive result.[13].

Whole-exome sequencing does not negate the need to consider the suspected mode of inheritance in a patient, especially when parent or other family samples are available for comparison. If there is enough medical history data to theorize a mode of inheritance, or an etiological diagnosis can be made from the phenotype of the patient, this can provide another filter by which to narrow candidate genes. Autosomal recessive conditions would manifest as a set of homozygous or compound heterozygous mutations in the proband and each parent would be expected to be a carrier of one or the other mutation. Autosomal dominant conditions would be present in a heterozygous form in the proband and the mutation may or may not be carried by the parent. Additional considerations for dominant conditions include reduced penetrance and variable expressivity, making a detailed examination of the parents for their offspring's traits extremely useful, if it is possible to gather such information. If an X-linked condition is suspected, the mutation would be expected to be present hemizygously in the proband and in a heterozygous state in the proband's mother. Lack of presence in the mother should not be immediate cause for discarding of the variant, as some of these conditions have high rates of de-novo cases. In Duchenne Muscular Dystrophy, a woman has only a 2/3 chance to be a carrier when she has a single affected son.

There is research to suggest that the role of de novo mutations in certain situations is underappreciated. Often medical genetics professionals encounter isolated cases of mental retardation, multiple congenital anomalies, or other diseases. Unless adiagnosis can be made, the underlying basis of the condition is unclear and can possible be autosomal recessive, multifactorial, due to environmental factors, oligogenic, or the result of a spontaneous mutation.

The per-generation mutation rate has been estimated at $7.6x10^{9} – 2.2x10^{8}$, roughly 1/100 million positions in the haploid genome. This would translate into a rate of 0.86 amino-acid-altering de novo mutations per person.[14] In situations such as intellectual disability where there is such genetic heterogeneity, analysis strategies can use parent-child trios to examine potentially pathogenic de novo mutations. In one such study, WES was obtained on 10 trios after ruling out CNVs by microarray. Exclusion of common, predicted non-pathogenic, and non-de novo mutations led to the identification of convincing candidate mutations in 7 of the 10 patients.[15]

1.4 COMPARISON TO OTHER TESTS

Previously, physicians were restricted to single-gene diagnostic odysseys or multi-gene panels, which could in some cases cost more than $100,000 and stretch over several years or many more [2], depending on if a causative mutation was ever identified. Exome, or even whole-genome sequencing, can examine all of the genes in the genome at various levels for a fraction of the price.

Compared to WGS, single-nucleotide variants found WES average greater Phred-based quality scores. Phred scores were originally developed by the computer program Phred to assist in the automation of DNA sequencing in the Human Genome Project. They are a measure of the probability of a variant base call being incorrect. The higher the quality score, the lower the probability of an incorrect call.[16] There are some regions (and therefore variants) missed by a typical WGS but observed in WES due to the higher coverage achieved by the target-enriched sequencing of specific regions. Similarly, there are some targeted regions and variants missed by WES that are detectable by WGS[5], and WGS can, by using a paired-end approach, detect large

structural variations such as insertions, deletions, inversions, and translocations.[1] However, to detect those one much be prepared to receive variant data on many genomic regions in which there is little evidence to be concerned about disease loci and, if insertions and deletions are the main variant of interest, high-coverage array CGH can perform the same function for a lower price and less extraneous information. Repetitive regions, exonic and other, are difficult to align in either case and can result in either missed variants or an excess of variant calls, and WGS is not immune to the drawbacks of WES including variation in coverage and efficiency of sequencing across the genome.[1] Though in the future WGS is predicted to be more economical than WES as it bypasses the need for the capture process, the amount of data generated by WGS is 100x more than the already overwhelming amount of data obtained through WES that is proving a challenge for data storage, bioinformatics filtering capabilities, and hardware and software for analysis.[1] Ultimately, unless analysis is to be focused on non-coding regions or structural variation, WES provides most of the benefits of WGS at a lower cost.

At this time, WES is not efficient as a first-line approach, and this is recognized by the American College of Medical Genetics and Genomics. Currently, it is recommended that WES be considered in the clinical diagnostic setting of an affected individual in one of three situations: if the phenotype or family history suggests a genetic cause, but the phenotype does not correspond to a specific disorder for which a targeted genetic test is clinically available; if the patient presents with a defined disorder that is known to have a high degree of genetic heterogeneity and thus WES is more practical and cost-effective; or, if the patient presents with a likely genetic disorder but specific genetics tests for the phenotype have failed to yield a diagnosis.[17] It is possible to perform WES prenatally, in the event that a fetus with a likely genetic disorder has failed to be diagnosed by other means, but ACMG counsels caution as WES

has several limitations in this setting, including a long turn-around time, and significantly higher rates of false-positives, false-negatives, and uncertain variants than seen with other prenatal technologies such as array CGH.[17] If the parents have decided to carry the pregnancy to term, it may be just as timely to undergo sequencing neonatally. Current research is looking at new WGS protocols that use automated bioinformatics analysis to develop a differential diagnosis within 50 hours[18] for use in neonatal intensive care units, as more than 20% of infant deaths are caused by congenital malformations, deformations, and chromosomal anomalies.[18] Hopefully, such an approach will continue to be refined and prove clinically useful at providing faster diagnoses and targeted treatment options families dealing with traumatic experience of having new baby with health issues.

1.5 DRAWBACKS TO WES

It is possible that mutations could be located in exons that are poorly covered by current targeting technologies and thus the candidate gene could falsely be removed from consideration. Currently, reasonable coverage can be achieved for approximately 90% of the sequenced exome.[4]. It has been found that 5-50% of RefSeq exons (approximately 3% of RefSeq coding exons) have less than 5x coverage in current commercial capture kits [1]. In addition, the ability to interpret results is also only as good as our current knowledge of the genes, their functions, expression, and possibly associated conditions.

Relevant variants might be predicted to be deleterious by algorithms such as SIFT or PolyPhen, but if little or nothing is known about the gene that they are located in such variants might have been falsely removed from consideration at earlier filtering stages. A relevant

mutation might also be falsely removed for not falling into the typical nonsynonymous/splice site/indel categories. A mutation that induces exon skipping can cause Mendelian disease, such as the silent mutation c.6354C>T in exon 51 of the fibrillin-1 gene in Marfan syndrome [19], and yet would not be detectable based on current filtering strategies.

Though most point mutation in inherited diseases so far have been located in or near exons, mutations in distant enhancers and regulatory elements have been implicated in hereditary conditions and would not typically be detectable using current enrichment strategies. Point mutations in the ZRS region, the long-range limb-specific cis-regulator of the sonic hedgehog (SHH) gene, were shown to cause pre-axial polydactyly in cat models.[20] As mentioned above in the case of Kabuki syndrome, genetically heterogeneous disorders can be missed in study groups of unrelated individuals, as individual patients could have varying gene involvement.

Researchers and clinicians should take care when considering dominant conditions, as potentially relevant mutations may be falsely discarded if they are present in an unaffected parent, even though reduced penetrance is a common feature of many such disorders.

There are logistical drawbacks to WES as well. The comprehensive nature of WES demands a longer turnaround time compared to traditional singe-gene tests or multi-gene panel. Labs currently offering clinical WES quote turnaround times of anywhere from 15-28 weeks, which can be an agonizing wait for patients who want to discover the cause of their condition, receive a diagnosis, and make use of available treatment or management guidelines. Currently next-generation technologies also have difficulty accurately calling insertions, deletions, tri-nucleotide repeats, and copy number variations, so a second testing method is usually required to identify these with a good degree of confidence, adding to the cost.[21] Whole-exome sequencing currently costs approximately $8,000 and has varying levels of insurance coverage.

2.0 CASE SUMMARY 1

Whole-exome sequencing was performed on the five members of family A. The mother and father were healthy individuals who were not consanguineous but were both from Iraq. They had one son who was unaffected. The mother was noted to have low levels of protein Z on a blood test dated February 25, 2011.A blood sample taken in April of 2012 showed abnormal levels of protein Z (0.68, reference range 0.70-2.61) and factor X (120, reference range 0.60-140).

2.1 BABY AA

Baby female AA was born in 2009 at 31 weeks gestation by Cesarean section due to decelerations. The pregnancy was significant for gestational diabetes which was well-controlled, with insulin therapy beginning at 28 weeks. Ultrasounds revealed a small head circumference as well as a suspected head mass. The birth weight was 1357 g with APGAR scores of 7 and 9. There was thick meconium with a true knot in the placenta and a single nuchal cord. Though serum cytomegalovirus IgG was positive, there were no overt maternal signs of a CMV infection; parvovirus and toxoplasmosis testing were negative. Baby AA demonstrated low platelets after birth and was worked up for Neonatal Alloimmune Thrombocytopenia by NICU staff. AA had microcephaly and developed SIADH (syndrome of inappropriate antidiuretic hormone secretion) at day of life 10. A brain MRI showed multiple areas of intracranial

hemorrhage in the white matter, basal ganglia, and thalamus. The hemorrhage was from both old and new bleeds. TORCH testing was negative. Testing on factor XIII, factor VIII, the thrombophilic risk panel, protein Z, and the extended LAC panel was negative. AA died at a few months of life.

2.2 BABY MM

Baby female MM was born in 2010 at approximately 38 weeks gestation by (C-section vs. NSVD). The pregnancy was again significant for gestational diabetes. Birth length was in the 10^{th} percentile, birth weight was in the 3^{rd} percentile, and head circumference was 10^{th} percentile. Her status was normal until day of life 9 when she began to deteriorate, showing hypothermia, hypotonia, foot drop, and respiratory failure requiring intubation. MM was placed on a mechanical ventilator as well as total parenteral nutrition. Numerous raised blancheable pink papules were noted on her face and trunk. Pathology was consistent with PLEVA. She developed thrombocytopenia, anemia, seizures, and cerebral hemorrhage, and passed away at four months of age. The autopsy report revealed hepatomegaly with cholestasis, hemosiderosis, and fibrosis; congestive splenomegaly; mild bronchopulmonary dysplasia; and bilateral serosanguineous hydrothorax. MRI revealed global multicystic encephalopathy due to a prolonged continuous series of small discrete infarcts affecting the cortex, subcortical gray matter, and hindbrain structures. There appeared to be a relationship between small vessel vasculopathy and the infarcts. The vasculopathy was unusual with intimal foam cell accumulation, and interestingly limited to the central nervous system, appearing to even spare the spinal cord. A primary cause of the vasculopathy could not be determined and prior coagulopathy evaluation was normal.

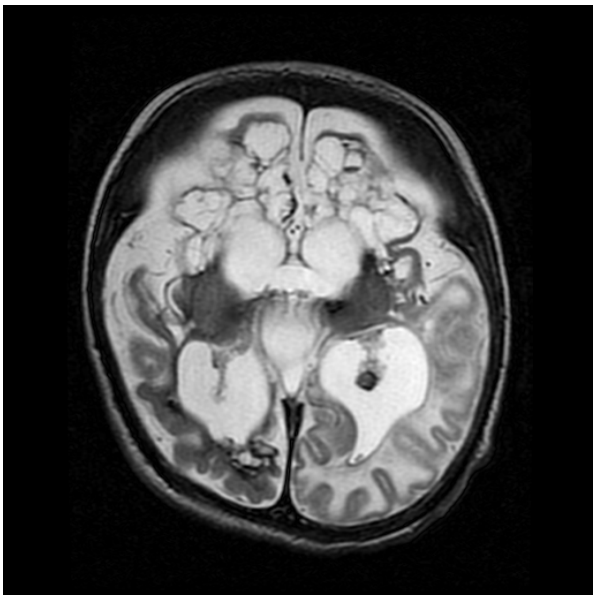

Figure 2. MRI of MM, showing multicystic encephalomalacia

2.3 CASE SUMMARY 2

Patient TW was a Caucasian woman in her 40's affected with multi-organ dysfunction, normal intelligence, and dramatic skeletal myopathy with normal heart function. Onset of muscle weakness was noted around in her early teens accompanied by episodes of mild hypoglycemia and hyperammonemia during periods of intercurrent illness. Over the years the disease has followed a neurodegenerative course. Extensive genetic testing for mitochondrial disease performed at another institution over many years was normal, including most recently Baylor's MItome400, a targeted gene sequencing panel for >400 nuclear encoded mitochondrial genes. A previous muscle biopsy for mitochondrial respiratory chain enzyme testing obtained only fibrotic tissue, and the analysis could not be performed. TW was adopted and there is no information about her biological family.

3.0 METHODS

WES was first undertaken on fibroblasts obtained post-mortem from Baby A, and on 20 ml whole-blood in two EDTA tubes from Baby M and TW. Whole blood was frozen immediately and shipped on dry ice to the Greater Pittsburgh Cytogenetics Laboratory (GPCL). Following sequencing and analysis of variants found in Baby A and Baby M, sequencing was performed on 20ml whole blood from their mother, father, and unaffected brother. WES was undertaken on DNA isolated from whole blood and initial analysis was performed by GPCL and Dr. Cecilia Lo's lab depending on the sample. Some samples were sequenced using either the SOLiD^{TM}5500xl system at GPCL or Illumina HiSeq2000 at the Beijing Genome Institute, taking advantage of a discount provided to Dr. Lo.

3.1 DNA CAPTURE AND AMPLIFICATION

DNA capture and amplification from WBC genomic DNA was performed using the SureSelect^{TM} Enrichment System. The 50 Mb SureSelect Human All Exon Kit is designed to target all human exons in a single tube, covering 99% of CCDS regions (http://www.ncbi.nlm.gov/projects/CCDS/CcdsBrowse.cgi), with additional Ensembl, Genebank, and RefSeq content. Genomic DNA was fragmented by sonication using Covaris^{TM} S2 (Covaris Inc, MA) and sequencing primers were ligated. SureSelect^{TM} baits, composed of biotinylated

RNA were hybridized to fragmented genomic DNA for 24-72 hours and the resulting heteroduplexes were enriched by binding to streptavidin beads. After magnetic extraction and several washes to remove non-targeted DNA the RNA was specifically degraded releasing single-stranded DNA for amplification and sequencing.

3.2 SEQUENCE BY LIGATION

The SOLiD^{TM}5500xl System (AB Life Technologies, CA) is designed to enable massively parallel sequencing by ligation of clonally amplified DNA fragments linked to beads. Sequential ligation of 8 base dye-labeled oligonucleotides allowed the query of two sequential bases (di-base encoding) with four-color competitive fluorescent detection. Following detection, unextended primers were capped and cleavage removed the last 3 bases and the fluorescent moiety. Repeated rounds of ligation and primer reset allowed each base to be read twice and color-space to deconcolute base-space sequences from the four possible dinucleotides coded by dye color. Each base was read in two independent rounds of ligation; therefore, a SNP resulted in two adjacent color changes. A measurement error results in a single color change, greatly reducing the number of false-positive SNPs and giving calling accuracies greater than 99%. Exact call chemistry (ECC) allowed for an additional round of primer ligation, which together with the two-base encoding formed an error-correction code providing highly accurate results in rare variant experiments such as this one. Each flowcell was divided into six lanes that could accommodate resequencing of two exomes.

3.3 INITIAL ANALYSIS

Once raw sequencing data was returned, low quality read sequences were discarded. The rest were then aligned to a reference genomic sequence using either the NCBI reference sequence (RefSeq), or human genome reference sequence (hg19 build) using the CLC Genomics Workbench.

3.3.1 Case 1

Initial analysis was performed by You Li of Dr. Cecilia Lo's lab. Coverage cutoff was set at 5x, with medium coverage of 80x. The cutoff for frequency of mutations was set at 0.25 (25%). Mutations with less than 25% were not recalled. Frequency designations were made as follows: 0% for wild-type, 50% for heterozygous, and 75-100% for homozygous mutations. Data from HGMD, dbSNP, and 1000 genome was used to designate novel vs. reported mutations. The cutoff for splicing mutations was designated at 5 bases. A list was then generated of all exonic mutations shared by both AA and MA.

3.3.2 Case 2

Initial analysis of the raw sequencing data was performed by GPCL, yielding separate lists of all homozygous, heterozygous, insertion, and deletion mutations detected in TW's exome. Frequency designations were made as follows: 0% for wild-type, 50% for heterozygous, and 75-100% for homozygous mutations. Data from dbSNP was used to designate novel vs. reported mutations.

3.4 VARIANT ANALYSIS

3.4.1 Case 1

Variants were narrowed down to candidates in a step-wise manner. Given the apparent pattern of inheritance, the first criteria was that SNPs, insertions, and deletions be either homozygous or compound heterozygous and shared between AA and MM. Based on previous negative genetic testing and clinical suspicion that a novel mutation was responsible for the clinical phenotype, mutations were excluded if they were known polymorphisms or had previously been reported in any or all of the HGMD, dbSNP, or 1000 genome databases. Mutations were then narrowed based on likelihood of functional impact on the protein, beginning with mutations that were designated as nonsynonymous, frameshift insertion or deletion, stop-loss or stop-gain, non-frameshift insertions or deletions, or splicing mutations. Tissue expression for the remaining genes was determined using the BioGPS database (http://biogps.org). Based upon the clinical presentation of AA and MM, genes with increased expression in brain, fetal brain, and/or immune system and blood cells were deemed to have “relevant” tissue expression. Genes lacking relevant expression, those with no information in the database, and those with even expression in all tissues were excluded. For compound heterozygous mutations, genes were excluded if they had more than 10 mutations, with the thought that these were likely to be polymorphisms and not relevant to the phenotype. If at any point the application of filters narrowed the list of mutations in a certain gene to 1, that gene was removed from further consideration. After further analysis, genes with more than 3 mutations were also excluded. Splicing mutations were excluded if they were greater than 3 bases away from an exon, with the thought that these would be less likely to have a functional impact on the RNA product and thus the protein. Compound heterozygous

mutations were also examined using the SIFT and PolyPhen algorithms to predict their effect on the gene. After this level of analysis, the resulting candidate genes were compared to sequencing results from the parents and unaffected brother of AA and MM. In order to continue to be a candidate, the homozygous or compound heterozygous mutation could not be shared with their unaffected brother. Given that the clinical picture seemed to suggest autosomal recessive inheritance, each parent had to be a carrier of the candidate mutation if homozygous, or one of the compound heterozygous mutations. Though the parents are not consanguineous, their similar geographic origin suggests that a homozygous mutation is more likely than a compound heterozygous one. If particular genes were known to be associated with any disorders in the Online Mendelian Inheritance in Man (OMIM) database, a note was made of this as well.

3.4.2 Case 2

Variants were again narrowed down in a stepwise manner in two separate analysis sets. Given that TW was adopted and no information was available about her biological family, little guidance was available to hypothesize which form of inheritance most likely characterized her disorder. Therefore, all homozygous, compound heterozygous, insertions, and deletion mutations were subjected to analysis. Based on her clinical presentation, a mitochondrial disorder was suspected and the first round of analysis focused on mutations found in the genes contained in the comprehensive mitochondrial and metabolic disease panel v2.8 from Baylor Laboratories. The panel includes approximately 351 genes associated with approximately 180 distinct disorders or recognized subtypes of disorders of the mitochondria or metabolism. Candidate mutations matching genes in the panel were narrowed down by position in the exon, relevant expression, nonsynonymous mutations, and mutations unreported in dbSNP. Relevant expression

was designated as genes noted in the BioGPS database as having elevated expression in the brain or skeletal muscle, or with approximately even expression in all tissues (as this would suggest mitochondrial expression). If at any point the application of filters to the list of compound heterozygotes narrowed the list of mutations in a certain gene to 1, that gene was removed from further consideration.

As analysis by the first method yielded only mildly interesting candidate mutations, a second round of analysis was undertaken. In this round, mutations in patient TW's exome were compared to a larger database called MitoCarta, an inventory compiled by The Eli and Edythe L. Broad Institute of Harvard and MIT of 1013 nuclear and mtDNA genes encoding proteins with strong support of mitochondrial localization based on homology to mouse MitoCarta genes. Genes from the database not already examined in Round 1 of analysis were then subjected to the same analysis parameters.

4.0 RESULTS

4.1 CASE 1

As shown in table 1, of the total number of mutations present in AA and MM, 20,849 were shared between the two siblings. 2558 were not listed in dbSNP or 1000Genome and 1647 were determined to be truly novel coding variants. 132 of the novel mutations were homozygous in both patients and 1445 were heterozygous. In the list of homozygous mutations, 86 were found to be the only mutation left for consideration in that gene and 43 of those were determined to have relevant expression patterns. After excluding splicing mutations more than 3 base pairs from an exon, a list of 15 candidate mutations remained. Of the 1445 heterozygous mutations, 1088 remained after excluding genes with only a single mutation (those that were therefore not compound heterozygotes). Excluding genes with more than 10 mutations reduced the list to 306 mutations, of which 124 had relevant expression. Splicing mutations more than 3 base pairs from an exon, genes with more than three mutations, cadherin genes, and zing finger genes narrowed the candidate list to 62 mutations.

Three candidate genes were determined to be of particular interest as they show high expression in the brain and relate to immune system function. The variants were homozygous in both siblings and were confirmed by Sanger sequencing and testing of the parents and unaffected sibling is in progress. The *DIXDC1* gene contained a c.813insC frameshift mutation in exon 10,

resulting in a p.S271fs mutation. The gene is expressed in the brain fetal brain, immune system, and blood. The *ITPR2* gene contained a frameshift substitution of ACTC in exon 3 at c.1408. The gene is expressed in the brain, immune system, and blood. The *NLRC3* gene contained a c.2320delC frameshift mutation in exon 10, resulting in a p.L774fs mutation. The gene is expressed in the immune system and the blood.

Table 1 Case 1 Homozygous Candidate Mutations

Gene	**Variant**	**Sequence Chance**	**Amino Acid Change**	**Expression**
ASCC3	Splicing	-	-	721 B lymphocytes, smooth muscle, bronchial epithelial
C12orf10	Nonframeshift substitution	CAA>CGC	-	testes, 721 B lymphocytes, even in: lung, liver, whole brain
C14orf169	Frameshift insertion	insC	p.A87fs	testis, thymus, immune/blood
COG3	Nonframeshift substitution	TTG>TCA	-	immune/blood, prostate, pancreas
DDX24	Nonframeshift substitution	CACGG	-	brain, especially pineal night/day, prefrontal cortex, hypothalamus
DIXC1	Frameshift insertion	insC	pS271fs	brain, especially retina, amygdala, hypothalamus, immune, fetal brain
EI24	Frameshift insertion	insC	pT282fs	bronchial epithelial, prostate, colon, liver, CD34+. CD105+ endothelial
ITPR2	Frameshift substitution	ACTC	-	immune/blood, brain, ganglions, blood coagulation, platelet activation
KRBA1	Frameshift insertion	insC	p.A552 fs	heart ventricle, brain, cervix, pituitary gland
LEPREL2	Frameshift insertion	insG	p.R140fs	pineal day

Table 1 continued

LIMD1	Splicing	-	-	BDCA4+ dendritic cells, 721 B lymphoblasts
NLRC3	Frameshift deletion	delG	p.L774fs	immune/blood, all tissues
PCDHB9	Frameshift insertion	insA	p.T381fs	spinal cord, occipital lobe, hypothalamus
SCAMP1	Frameshift insertion	insA	p.I244fs	trigeminal ganglion, pineal night/day
SON	Frameshift insertion	insA	p.G2412fs	immune/blood

4.2 CASE 2

4.2.1 Round 1 Analysis

The results of the first round of analysis are detailed in Appendix A, tables 4, 5, and 6. TW's exome was found to contain 16215 homozygous, 43999 heterozygous, 4605 insertion, and 39 deletion mutations. In the list of homozygous mutations, 232 were located in genes on the reference mito chip. The list was narrowed to 76 by removing intron-located mutations, then to 39 by removing mutations with non-relevant expression. Of that list, 22 were non-synonymous mutations. Removing mutations reported in dbSNP narrowed the list to 7 candidate mutations. In the list of heterozygous mutations, 542 were located in genes on the reference mito chip. The list was narrowed to only those that had two exonic mutations, for a total of 191. 64 of those had relevant expression patterns. Of that list, 27 were non-synonymous mutations and removing previously reported dbSNP mutations narrowed the candidate list to 19. Of the list of insertions, 1505 were located in exons. 23 of those were located in genes on the mito panel and 12 had relevant expression. Of the deletions, 7 were located in exons. Of the candidate mutations two, a

compound heterozygous mutation in *GFM1* and a homozygous mutation in *PET112L*, looked the most likely to be relevant but were not strikingly obvious as the causative mutations. They were confirmed using Sanger sequencing and were unreported in the NLHBI exome sequencing project database. They are also both conserved across species. They were deemed to warrant further consideration if nothing more significant was found upon a second round of analysis.

4.2.2 Round 2 Analysis

Four lists were generated in the second round of analysis of mutations matching genes in the MitoCarta database that had not been investigated in the first round. This criterion generated 433 homozygous and 1071 heterozygous mutations for further study, as detailed in table 2. In the list of homozygous mutations, 175 were located in exons. 102 of those 175 were non-synonymous mutations and 22 were also unreported in dbSNP. The list of candidate genes was then narrowed to 9 with relevant patterns of tissue expression. As shown in table 7 of Appendix A, 941 mutations were found to be compound heterozygotes. Removing genes with more than 5 mutations narrowed the list to 673 mutations. Mutations that were apparently sequencing errors, in which three sequential nucleotides were reported as mutation, were removed to narrow the list to 563. 241 of those were located in an exon. Removal of synonymous mutations narrowed the list to 141 mutations, of which 122 were unreported in dbSNP. Removing mutations in genes with non-relevant expression reduced the list of candidate compound heterozygous mutations to 42.

Table 2 Final Candidate Gene List, all homozygous mutations

Gene	Sequence change	Amino Acid Change	Tissue Expression	Sanger Sequencing	NHLBI
ADCK5	A>C	342K>Q	even expression in all	-	Not found
ABCB4	G>C	1163F>L	even expression all tissues	-	Not found
	A>T	1163F>Y		-	Not found
ATPAF2	C>G	129R>P	cerebellum peduncles, trigeminal ganglion, even expression in the rest	-	Not found
	G>A	129R>W		-	Not found
BCL2L2	T>G	87F>L	multiple brain elevations (incl. prefrontal cortex, whole brain, hypothalamus, amygdala, all others)	-	Not found
C8orf38 (NDUFAF6)	A>C	67K>T	no data	Not detected	Not found
CCDC109A	G>A	238G>S	even expression in all	-	Not found
GFM1	A>T	609F>Y	Even expression in all	confirmed	Not found
GATM	A>C	223Y>X	kidney, fetal liver/liver, slight elevations in amygdala, prefrontal cortex, spinal cord)	-	Not found
MTRR	C>G	213H>Q	CD4+ T cells, hypothalamus, pineal night/day	-	Not found
PET112L	T>A	397T>S	Heart, spinal cord	pending	Not found

Table 2 continued

RAB24	A>T	32F>L	no data	-	Not found
SLC25A30	C>A	196M>I	even expression in all	-	Not found
	A>G	196M>T		-	Not found
SLC25A40	C>G	86G>R	even expression in all	-	Not found

The second round of analysis yielded one mutation of particular interest. TW was found to have homozygous A>C mutations at chromosome position 96044225 of the NADH dehydrogenase (ubiquinone) complex 1, assembly factor 6 gene (*NDUFAF6*, also known as C8orf38). This mutation resulted in a p.67K>T amino acid change and was predicted to be "possibly damaging" by PolyPhen-2, with a score of 0.624 (sensitivity 0.87, specificity: 0.91). When the variant was analyzed using SIFT it was predicted to be tolerated, with a SIFT score of 0.17 and a median information content of 1.84. The gene is associated with mitochondrial complex-1 deficiency, which is the most common enzymatic defect of the oxidative phosphorylation disorders. It is characterized by a wide range of clinical disorders ranging from lethal neonatal disease to adult-onset neurodegenerative disorders.[22] This is felt to be consistent with TW's clinical presentation and disease course. Lysine is a positively charged polar molecule with a basic R group. As the side chain has three methylene groups, the side chain has significant hydrophobic characteristics even though the terminal amino group will be charged under physiological conditions. Threonine, by contrast, is a hydrophilic uncharged polar molecule with a non-aromatic hydroxyl as its R group and therefore the reported amino acid substitution would represent a change in the chemical composition and properties of the protein.

Fibroblast samples were also still available and are in the process of being stained for complex-1 activity levels and super-complex assembly. Unfortunately, when Sanger sequencing

was undertaken it did not validate the mutation. The mutation had a coverage depth of 5x, which is the exact cutoff for report and therefore it represents a false positive call. Two additional genes in the MitoCarta gene set have been identified with variations, one involved in mitochondrial chromosomal maintenance (*PET112L)* and the other in mitochondrial protein translation (*GFM1)*. Both variants are in the process of being further evaluated.

5.0 DISCUSSION

5.1 CASE 1

As has previously been discussed, the availability of family members for sequencing is an important factor in determining relevant candidate genes. In Case 1, analyzing the mutations shared between both sisters allowed for a significant reduction in the number of candidate genes being considered. However, due to the nature of the genes on that list and the lack of known associations with human disease we were unable to pinpoint a leading candidate by their sequencing alone. Reports in the literature have demonstrated that the more affected patients there are available to do sequencing on, the higher the chance of finding a causative mutation. However, the number of patients and/or unaffected family members necessary to discover the variant is going to depend on the nature of the phenotype being examined and so there is no recommended number. In one study of spinocerebellar ataxia in four generations of a Chinese family, four exome data sets were enough to determine the sole candidate gene responsible for the phenotype in the family.[23] Other studies have needed anywhere from a single patient to ten family members or unrelated patients to pinpoint candidates.[12] The hope is that the addition of three family samples will allow us to elucidate the most relevant of the candidate genes and then examine them further with functional studies.

5.2 DIFFERENTIAL DIAGNOSES FOR CASE 2

Phenotypes of mitochondrial complex 1 deficiency include macrocephaly with progressive leukodystrophy, nonspecific encephalopathy, hypertrophic cardiomyopathy, liver disease, Leigh syndrome, Leber hereditary optic neuropathy, and some forms of Parkinson disease. As a disorder it shows significant heterogeneity and can be caused by mutations in either nuclear-encoded genes or in mitochondrial encoded genes. The majority of cases are caused by mutations in nuclear encoded genes, which *NDUFAF6* is. There are no obvious genotype-phenotype correlations and inference of the underlying basis from the clinical or biochemical presentation is difficult, if not impossible. Inheritance can follow either X-linked, autosomal recessive, or mitochondrial patterns. Mutations in the nuclear-encoded genes NDUFS1, NDUFS4, NDUFS7, NDUFS8, and NDUFV1 result in neurologic diseases, mostly Leigh syndrome or Leigh-like syndrome. [22] Before Sanger sequencing, this was felt to be the most likely of the differentials for patient TW.

Given that the *NDUFAF6* mutation proved to be a false positive, the *GFM1* and *PET112L* mutations will be subjected to further consideration. *GFM1* codes for the mitochondrial elongation factor G1. In order to successfully complete the elongation phase of protein translation, mitochondria need three functional elongation factors: Tu, Ts, and G. The bacterial Efg catalyzes ribosome translocation during peptide elongation and mediates ribosomal disassembly during ribosome recycling. In humans, the same role is divided between EFG1 and EFG2, with EFG1 catalyzing the translocation component[24] as well as being involved in the GTP catabolic process. As such, it has a number of disease associations. Defects are the cause of combined oxidative phosphorylation deficiency type 1 that leads to early fatal progressive hepatoencephalopathy. Additionally, mutations have been linked to factor 7 deficiency,

hemophilia B, heart aneurysms, candidiasis, and chondrosarcoma.[25] The known phenotypic presentations of *GFM1*'s main associated disorder differ significantly from TW's phenotypic presentation. Onset occurs at or soon after birth, with features including growth retardation, microcephaly, hypertonicity, axial hypotonia, encephalopathy, cardiomyopathy, and liver dysfunction. Death usually occurs in the first few weeks or years of life, compared to TW's adult-onset presentation. [26]

PET112L is a homolog of the S. cerevisiae gene pet112. Mutations in this gene block the accumulation of cytochrome oxidase subunit II and disruption of the gene results in the destabilization of the mitochondrial genome. It is therefore suggested that the pet112 protein plays a major role in mitochondrial gene expression, most likely in translation. The *PET112L* protein shares 30% identity with the yeast version, contains a mitochondrial leader peptide, and is predominantly expressed in tissues with high rates of oxidative phosphorylation.[27]. Currently, the gene has not been associated with any genetic disorders [28], but its role in the stability of the mitochondrial genome suggests that dysfunction could cause mitochondrial depletion. It is therefore possible that such firm disease associations simply have not been made yet and the mutation is related to TW's symptoms, but also that another mutation is a stronger candidate. Though it can be postulated, and in some cases proven, that mutations in each of these two main candidate genes would have widespread detrimental effects on mitochondrial function, the known phenotypes associated with them seem to fit less with the patient's phenotype. However, recognizing that our level of genetic information about disease associations is incomplete, it is possible that the patient represents an atypical presentation of one of these disorders. More definite conclusions will have to wait for the completion of functional studies.

6.0 SIGNIFICANCE

As whole-exome sequencing moves from the research setting into more widespread clinical use it will greatly change the landscape of what genetic testing can offer to patients. This may include searching for novel mutations, attempting to find a diagnosis for a patient on a diagnostic odyssey, or it may include diagnosing a known condition as a one-shot test in lieu of reflexing through genes in a panel.

6.1 IMPACT ON GENETIC COUNSELING

6.1.1 Ethical Issues

The sheer quantity of information and range of possible results produced by WES raises a number of ethical and practical issues. Before undertaking WES with a patient the physician or genetic counselor needs to consider all of these aspects carefully so that they can navigate the process to the best advantage of the patient. In genetic counseling, the four main guiding ethical principles are beneficence, non-maleficence, justice, and autonomy.[29] In the context of WES, some of these principles may come into conflict with each other in specific situations. All need to be weighed carefully when consenting patients and deciding what level of results to disclose to them.

6.1.1.1 Informed Consent

Given the popularity of genomic technology in the media, it is always important to assess the patient's level of knowledge, concerns, and expectations about testing. With whole-exome sequencing in particular, patients may have an unrealistically high expectation of a test that "looks at all of the genes" to deliver an answer or diagnosis. The limitations of current knowledge and testing should therefore be addressed in the informed consent process. The nature of the discussion should also be determined in part by the age of the patient and the laboratory's policy regarding the return of results. It is historically not recommended for minors to undergo testing for conditions such as carrier status and adult onset conditions. Certain conditions may not be reported on at all, or reporting may be limited to adult patients only (see Appendix C for a comparison of clinical lab policies). Different labs might also have differing capabilities regarding the ability to obtain results on certain classes of mutations, such as mitochondrial mutations or X-linked carrier females.

WES is a complex test for patients to understand, and even after a thorough explanation by a genetic counselor or researcher patients can have a difficult time explaining their understanding of the test or explaining the test to family members.[30] Genetic counselors must also keep in mind the length of the discussion involved in the consent process and the amount of information the patient must process. Tabor et al. 2012 elicited feedback from Miller Syndrome patients and family members undergoing whole-genome sequencing and they reported that at some point the information "went in one ear and out the other". [30] It is not practical to break the informed consent process into smaller pieces in a clinical setting, so genetic counselors should consider ways to give patients preliminary information beforehand to read, as this would give them an opportunity to absorb some of the information at their own pace and think of some

questions ahead of their session. Given the possible categories of results to choose from unrelated to the presenting phenotype (carrier status, adult onset conditions, pharmacogenetics, mitochondrial conditions, etc), extra care will be needed to ensure that the patient has a good understanding of the classifications and impact, and has thought about how a positive result of different types would affect his/her life.

6.1.1.2 Return of Results

A limitation to the return of relevant results will be the level of current genetic knowledge, something that will be continuously changing. The WES results a patient receives today might have a completely different interpretation two or three years from now particularly when they involve variants of uncertain significance (VUS). VUS may be reclassified as pathogenic or benign as more data accumulates in the testing laboratories on the phenotypes of patients with that VUS. VUSs can also be reclassified if additional research is able to further elucidate the mutation's effect on the gene and protein expression, whether through avenues such as molecular studies or animal models. How then should genetic counselors provide updates in VUS status to their patients? Is it the patient's responsibility to contact the counselor to check for updates, as in the case in current BRCA1/2 testing? Or should the labs create a database of variants to allow them to easily contact all patients with a certain VUS if/when it gets reclassified? There are issues of logistics and confidentiality to consider in both cases and it is difficult to say at this point which would ultimately be more practical for WES. In part, it may depend on the volume of patients undergoing the testing and the willingness of patients to potentially have their information stored in such a database. As in cancer counseling, some patients may prefer to receive their results by phone and save themselves the trouble of having to arrange and attend a face-to-face appointment, but given the broad spectrum of possible results and implications in

WES it may not be practical to handle disclosure over the phone. If a genetic condition is diagnosed by WES, it will also likely necessitate that the patient return to the clinic for a follow-up appointment anyway to discuss the implications for their medical management.

6.1.1.3 Pediatric versus Adult WES

Genetic counselors and clinicians need to be aware of the varying policies that laboratories have regarding what classes of mutations they are willing to report on in the pediatric setting. Adult patients have the autonomy and authority to decide for themselves what type of results they want disclosed to them, whereas for pediatric patients the decision lies with their parents, and they may have the opportunity to receive results not generally offered to minors with other types of genetic tests (carrier status and adult onset conditions that will not affect their current medical management). Care must therefore be exercised by the pediatric genetic counselor or geneticist in balancing both the rights of their patient and the patient's parents, recognizing the growing decision-making capabilities as adolescents mature, their right not to know certain information, and possibly their desire to keep certain kinds of information from their parents (specifically genetic information not pertinent to their current medical condition).[31; 32]

6.1.2 Diagnostic odysseys and negative results

At this point when WES is new it will likely be first used for those patients of a geneticist who have eluded a diagnosis through all other testing avenues. For the patient, this can mean years or even decades of knowing that they have a condition but not having any information on the name or the advantage of medical literature to guide them in anticipating what they might expect in the future, reproductive impact, or treatment options specifically for that condition. Unfortunately, it

is possible that even a powerful test such as WES may not provide an answer for them. Current labs are quoting diagnostic rates in the range of 34-51%[33]. In the research setting the rates have often been far lower, in the realm of 8-24%[34]. Regardless of which number is closer to the truth, it still remains a fact that they are statistically more likely not to find an answer. Families can be confused and frustrated by the inability of medical professionals to give them a diagnosis. The months, years, or decades of biochemical, genetic, and imaging tests can seem like a waste of time and a financial burden. Numerous feelings can be associated with such an odyssey, including hope, fear, depression, anger, and isolation. [35] Achieving a diagnosis for a rare disorder can provide many benefits for the patient and family in both the short and long term. Both qualitative and quantitative assessments of this impact represent an important area for future research as WES becomes utilized clinically, failing to receive an answer from WES may have a unique type or degree of impact on a patient or family's mental well-being compared to more traditional tests.

6.2 IMPROVED THERAPEUTIC OPTIONS

Gene discovery is an essential first step in the process of understanding the genetic and biochemical mechanisms of inherited diseases, and for providing clues to direct research into therapies. Gene-specific treatments are currently being undertaken worldwide and there have been a number of successful gene-therapy trials aimed at correcting the inborn errors causing immune deficiencies, metabolic disorders, and more recently thalassemia.[1] Local deliver of replacement genes is also being tested in human clinical trials for several forms of hereditary

blindness such as Leber congenital amaurosis and retinitis pigmentosa. The elucidation of a wider range of causative genes can only lead to a further broadening of the conditions eligible for this kind of research. Beyond gene therapy, understanding of the genetic and biological mechanisms of the conditions can lead to knowledge of what currently-existing treatments and therapies could be applied, or how current treatment techniques such as enzyme replacement therapy might be altered for the specific needs of affected individuals.

Worthey et al (2011) reported a case in which WES was successfully used to diagnose an infant with a severe gastro-intestinal presentation of inflammatory bowel disease requiring a cholectomy and ileostomy that, despite a thorough clinical evaluation and extensive genetic testing, was unable to be definitively given a diagnosis. WES detected a novel hemizygous mutation in the X-linked inhibitor of apoptosis gene, leading to a diagnosis of X-linked inhibitor of apoptosis deficiency and the identification of a novel cause of IBD. This diagnosis allowed the child to receive an allogenic hematopoietic progenitor cell transplant, which is the recommended treatment for the condition, to prevent the development of life-threatening hemophagocytic lymphohistiocytosis. The gastrointestinal disease also resolved post-transplant, suggesting that the mutation underlay those symptoms as well. [36]

WES can also be used to drive clinical care in instances of a known diagnosis of unknown genetic etiology. In the case of a set of male and female twins diagnosed clinically with Dopa (3,4-dihydroxyphenylalanine)-responsive dystonia (DRD), mutation analysis of the two primary genes (*TH* and *GRP*) yielded no results. Sequencing of a third gene, *SRP*, was not available clinically, so WES was performed, detecting compound heterozygous mutations in *SRP*. The clinical diagnosis of DRD was sufficient to start the twins on L-dopa treatment, but patients display a range of responses to such therapy and L-dopa alone may not completely

alleviate symptoms, such as in this case. As *SRP* encodes sepiapterin reductase, which synthesizes a cofactor needed for the action of enzymes that make both dopamine and serotonin, identification of the *SRP* mutation suggested that supplementing L-dopa with the serotonin precursor 5-hydroxytryptophan might further improve symptoms. This proved to be the case. Interestingly, the mutations also co-segregated in the family with a fibromyalgia phenotype. Fibromyalgia can respond to serotonin reuptake inhibitor drugs (SSRIs), suggesting that the disease is related to reduced serotonin, and the authors hypothesized that loss-of-function *SRP* mutations might therefore contribute to fibromyalgia susceptibility. [37]

Sometimes, WES can be used to rule-out treatments that would not ultimately improve the individual's state of health. An infant with acute liver failure was found to have a recessive disorder due to mutations in *C10orf2* (*TWINKLE)* that resulted in mitochondrial DNA depletion. In this case, the diagnosis allowed the parents to be counseled that the infant would not be an appropriate candidate for a liver transplant. Though liver transplantation is the treatment of choice for liver failure, it has been shown to be futile in one form of mtDNA depletion and to have similar poor long-term outcomes in others. Though this sadly was not an answer that led to a treatment, it allowed the parents to have knowledge of the progression and outcome that they could expect for their child, and to spare themselves and their child a major, invasive operation that would have been extremely expensive and ultimately not beneficial. [38]

6.3 LIMITATIONS

As with any type of research there exist opportunities for error in the analysis of these two case examples. As the analysis process is not an automated one, naturally there is the chance for

human error in the input, examination, and interpretation of the data and its relevance to the research question. The variant filtering strategy used follows closely with the pipeline suggested in the literature, but with some deviations in the order of the application of certain filters. This influenced the type of variants filtered out at each stage and lead to some variants making it to higher or lower levels of consideration that perhaps is warranted. The interpretation of data is also limited by the current level of knowledge about the genes studied and the information available in databases such as dbSNP, BioGPS, OMIM, and others. As previously discussed, there is a certain level of misclassification of variants in some databases and other databases reflect our incomplete knowledge about the function and disease associations of certain genes that may also lead to candidate variants being erroneously discarded. This project also utilized different labs for the sequencing and initial analysis of each case. Ideally, the same lab and analysis process should be used with each new patient to maintain the greatest concordance between the resulting data sets. Sequencing of the exome also achieved lower coverage of target regions than is possible and ideal by clinical standards, in the range of 4-136x for Case 2 with an average read depth of 14.12x and only 21% of calls at a read depth of 20x or greater. The lower the coverage level is, the higher the chances for both false positives and false negatives, as demonstrated by the red herring mutation in *NDUFAF6*. This is not always necessarily the case, as Dr. Vockley's lab has confirmed mutations by Sanger sequencing read at a coverage of 4x, but the general trend should be noted. A 30x median coverage of the target may be sufficient, but 100x coverage is better to ensure that variants can confidently be determined for a higher proportion of the exome.[1] Approximately 3.5-8% of variant calls in WES will be false positives with current technology[5], and the total percentage of false calls may be higher, in the range of 15-20%. False positive calls are most often due to incorrect mapping and systematic sequencing

errors such as certain combination of nucleotides being routinely misread by the sequencer. False negatives are typically the result of low overall coverage, poor capture efficiency of certain regions, and difficulty in unambiguously aligning repetitive regions. One strategy to reduce these errors is to compare each test sample against previously sequenced exomes, and missing regions are relatively easy to flag and report, to be followed up with more targeted sequencing if the researcher chooses.[1]

The cases studied here also demonstrate the challenges of sequencing single patients and the need to include other family members for comparison whenever possible. Not only is it useful for elucidating risks to other family members based on carrier status or possession of the same deleterious mutation, but it also provides a way to better classify certain variants as more likely to be benign based on the presence in unaffected family members. If a family member shares the same phenotype, such in Case 1, comparison to the affected individuals provides a good candidate list from which to then weed out variants based on the genetic status of unaffected family members. However, a researcher should use caution when ruling out variants by this method in the case of a suspected autosomal dominant condition, recognizing the limitations imposed by variable expressivity and incomplete penetrance that may lead to the premature discarding of the pathogenic mutation. Classification is easier when autosomal recessive inheritance is more likely, given the more penetrant and phenotypically noticeable effects of such conditions.

In this study, closer examination of the shared candidate mutations in the patients needed to wait until sequencing data was available from their parents, as samples had been collected at a later time from them. Potentially the siblings could have had a shorter list of candidate genes or more striking candidates, in which case delaying the collection and sequencing of parental

samples would have proven a cost-saving measure. As in this case, it instead resulted in a delay of results by several months, drawing out the family's search for answers to a traumatic situation and leaving then unsure whether to plan future pregnancies. It is therefore my opinion that in such a case, samples from the affected individuals and their parents should all be collected and analyzed together. Then if necessary, samples from other pertinent individuals in the pedigree can be collected.

As Case 2 demonstrated, using existing databases runs the risk of also discarding plausible candidate mutations when filters are too stringently applied. Some genes have no data on tissue expression available in BioGPS, a situation likely common to all such databases, and simply discarding such genes would have resulted in the candidate *NDUFAF6* mutation being overlooked. However, as Sanger sequencing ultimately discredited the finding of the mutation, it also highlights the need in both the research and clinical settings to confirm the effects of suspicious mutations by other means whenever possible. Particularly when computer algorithms of the pathogenicity of the mutation are not in concordance, functional studies of enzyme activity or other biomarkers can disprove or lend credence to the finding. Given that this would incur additional time and expense, such methods should only be used to examine the most likely relevant findings. In the clinical setting, laboratories may have more resources with which to examine candidate variations and draw firmer conclusions about their significance on the first round of analysis. In contrast, in the research setting there exists a greater freedom to run samples through alternate analysis filters, such as re-examining Case 2 using MitoCarta, and spend more time closely examining a single data set.

6.4 FUTURE DIRECTIONS

Whole exome sequencing is just now entering the phase of broader utilization of its capabilities and as such there exist many opportunities for further research and expansion of knowledge. Researchers need to continue to study individual genes and conditions to elucidate the genotype-phenotype correlations so essential to the accurate interpretation of WES data. WES results will themselves play a part, in determining the genetic bases of certain disorders. Further research is also needed into the clinical impact of WES, on the patients, genetic counselors, and physicians who are the beneficiaries and utilizers of the technology. We need to understand the perceptions, psychosocial impact, biases, and logistical impact that the test has in the clinical setting to a better degree than we do now as its use is increasing in clinical practice. Such research will allow us to apply this type of testing with the fullest degree of awareness, sensitivity, and efficacy, as possible.

7.0 CONCLUSION

In summary, recent advances in exome sequencing are accelerating the pace of gene discovery for Mendelian disorders. Studies such as this one demonstrate the degree of complexity that researchers encounter as they analyze such sequencing data on gene-discovery studies and the need for continued research to expand our knowledge of genes and their physiological impact so that new associations with human disease can be made. Even though WES represents a powerful new diagnostic tool for patients, there are still limitations to its capabilities, as demonstrated by the results of this study and the difficulty in finding causative mutations for these three patients.

As WES is enters the clinical practice, genetic counselors will play an important role in helping patients navigate the process and understand the impact of the results on their lives. It will take time to fully appreciate the logistical impact that such a test will have on the clinic, but clinicians will have to be aware of the financial, ethical, and psychosocial issues that arise in the context of such testing. As WES becomes a standard component of genetic diagnosis, it will also be extremely important to educate the public about how this technology can be used.

APPENDIX A

SUPPLEMENTARY TABLES

Table 3 Case 1 Compound Heterozygous Candidate Mutations

Gene	Mutation Type	Sequence change	Amino Acid Change	Tissue Expression
ANKRD36	nonframeshift substitution	c.11883_1185TGT	-	CD8+ T cells, pineal day
	nonsynonymous	c.G1186T	p.V396F	
	nonsynonymous	c.T1697C	p.I566T	
ANKRD38B	nonframeshift substitution	c.2731_2733ACG		pineal night/day, CD8+ and CD4+ T cells
	nonframeshift substitution	c.3124_3126TGT		
	nonsynonymous	c.C3127G	p.R1043G	
C8orf59	nonframeshift substitution	c.259_264AATGTT		immune/blood
	nonsynonymous	c.G265A	p.D89N	
CNN2	nonsynonymous	c.G630A	p.M210I	immune/blood
	nonsynonymous	c.G747A	p.M249I	
	nonsynonymous	c.T629C	p.M210T	

Table 3 continued

DCD5	nonsynonymous	c.C2041A	p.L681I	fetal brain, brain, spinal cord, ovary, hippocampus, bronchus, atrioventricular node, liver, lung
	nonsynonymous	c.T821G	p.F274C	
EXOG	stop loss	c.A956G	p.X319W	All tissues, 721 B lymphoblasts, multiple brain elevations (cerebellum, hypothalamus, whole brain, cerebellum peduncles)
	nonsynonymous	c.C953T	p.S318F	
FAM38A	nonframeshift substitution	c.5632_5634---		immune/blood, pineal night/day
	nonsynonymous	c.G6411C	p.M2137I	
FAM8A1	nonsynonymous	c.A395G	p.H132R	whole blood, CD56+ NK cells, many brain elevations (prefrontal cortex, amygdala, pineal night/day, fetal brain, hypothalamus, spinal cord)
	nonsynonymous	c.G401T	p.G134V	
	nonsynonymous	c.G418A	p.A140T	
HSPA9	nonsynonymous	c.A1900G	p.R634G	721 B lymphoblasts
	nonsynonymous	c.G1906A	p.A636T	
IBTK	nonframeshift substitution	c.4057_4059TAT		721 B lymphoblasts, CD33+ myeloid, pineal night/day
	nonsynonymous	c.C4013T	p.T1338I	
	nonsynonymous	c.G3970C	p.E1324Q	
LARP7	nonsynonymous	c.C1508T	p.A503V	immune/blood, pineal night/day, thyroid
	nonsynonymous	c.G1504A	p.D502N	
MEX3D	frameshift insertion	c.525_526insG	p.K175fs	relatively even in all, some elevations in brain
	nonframeshift	c.535_537ACC		

Table 3 continued

	substitution			
MLL3	nonsynonymous	c.C2578T	p.P860S	placenta, thyroid gland, blood, bone marrow
	nonsynonymous	c.G2573T	p.W858L	
MRPS17	nonsynonymous	c.G247A	p.E83K	CD105+ endothelial, 721 B lymphoblasts, bronchial epithelial
	nonsynonymous	c.T242A	p.L81Q	
	nonsynonymous	c.T298C	p.C100R	
MTCH2	nonframeshift substitution	c.235_237CAT		testes, 721 B lymphoblasts, CD105+ endothelial, CD34+
	stop gain	c.A229T	p.R77X	
NRBF2	nonsynonymous	c.A517T	p.T173S	whole blood
	nonsynonymous	c.T524C	p.I175T	
ONECUT3	nonframeshift substitution	c.564_566---		adrenal gland cortex, pancreas, accumbens, putamen
	nonsynonymous	c.C581G	p.A194G	
	nonsynonymous	c.G1333A	p.V445I	
PARG	nonsynonymous	c.C77T	p.S26L	superior cervical ganglion, testes, skeletal muscle, immune/blood
	nonsynonymous	c.A1720G	p.I574V	
PCSK6	nonsynonymous	c.C1454T	p.S485F	liver, spinal cord, other smaller brain elevations
	nonsynonymous	c.T1490C	p.I497T	
PDCD7	frameshift deletion	c.329delC	p.P110fs	pineal night/day
	stop gain	c.C193T	p.R65X	
	nonsynonymous	c.C197T	p.A66V	
ROCK2	nonframeshift substitution	c.487_489CGT		multiple brain areas, esp. prefrontal cortex, cingulate cortex, caudate nucleus
	nonframeshift substitution	c.538_540TTA		

Table 3 continued

RSL24D1	nonsynonymous	c.C430G	p.Q144E	immune/blood, bronchial epithelial, pineal night/day
	nonsynonymous	c.C487T	p.P163S	
	nonsynonymous	c.G463C	p.E155Q	
SALL2	nonsynonymous	c.C2140A	p.L714M	brain, esp. pineal night/day, cerebellum, cerebellum peduncles
	nonsynonymous	c.G2128A	p.V710I	
SRP19	stop gain	c.A124T	p.K42X	immune/blood
	nonframeshift substitution	c.277_279CGA		
	nonsynonymous	c.G261A	p.M87I	
TERF1	nonsynonymous	c.C250T	p.L84F	prefrontal cortex, pineal night/day, amygdala
	nonsynonymous	c.T240G	p.D80E	
USP6	nonsynonymous	c.C202T	p.R68W	testes, prefrontal cortex, fetal brain, appendix
	nonsynonymous	c.T362C	p.L121S	

Table 4 Case 2 Round 1 Compound Heterozygous Candidate Mutations

Gene	**Mutation type**	**Sequence change**	**Amino Acid Change**	**Tissue Expression**
ATPAF2	-	G>CG	161I>MI	cerebellum peduncles, trigeminal ganglion, even among rest
	-	A>AT	161I>IN	
	-	G>CG	128I>MI	
	-	A>AG	128I>IT	
GFM1	Non Synonymous	T>AT	512Y>NY	even expression in all tissues
	Non Synonymous	G>CG	513G>RG	
HLCS	-	T>CT	-	even expression in all tissues
	Non Synonymous	T>AT	-	

Table 4 continued

MUT	-	A>AT	-	pineal night/day, prefrontal cortex, CD34+
	Non Synonymous	T>CT	621K>RK	
PYGM	Non Synonymous	C>AC	790A>A	skeletal muscle, thyroid
	Non Synonymous	T>CT	236N>SN	
	Non Synonymous	G>AG	71T>MT	
RET	Non Synonymous	G>CG	106W>SW	skeletal muscle, uterus corpus, even expression in rest
	Non Synonymous	G>CG	106W>CW	
	Non Synonymous	G>CG	584D>D	
UBE3A	-	A>AC	-	brain, thyroid
	-	A>AT	-	
	-	T>AT	-	

Table 5 Case 2 Round 1 Insertion Candidates

Gene	Sequence Change	Amino Acid Change	Tissue Expression
COX4I1	insA	-	heart, multiple minor brain and immune/blood elevations
COX7A2L	insC	FS	pineal night/day, retina, immune/blood
FASTKD2	insA	FS	relatively even in all
GFM2	insT	FS	even in all, elevation in 721 B lymphoblasts
HADHB	insT	FS	immune/blood, spinal cord, hypothalamus, skeletal muscle, small intestine
LARS2	insA;AT	FS	relatively even in all
	insG;CT	FS	
MARS2	insG	FS	relatively even in all
NDUFB9	insG	-	heart, immune/blood, small brain elevations
PDHA1	insA	FS	pineal night/day, amygdala, spinal cord, lymphoma
SLC25A19	insG	FS	no data

Table 5 continued

VARS2	insT;CT	FS	relatively even in all

Table 6 Case 2 Round 1 Exonic Deletions

Gene	Sequence Change	Amino Acid Change
CCDC80	delT	FS
PIK3R1	delT	-
PCDHB9	delG	FS
BRPF3	delG	-
ABCC8	delG	FS
DDHD1	delAG	-
NLRP7	delG	FS
CCDC80	delT	FS
PIK3R1	delT	-
PCDHB9	delG	FS

Table 7 Case 2 Round 2 Compound Heterozygous Candidates

Gene	Sequence change	Amino Acid Change	Tissue Expression
ADCK4	T>AT	299Q>LQ	even expression in all
	G>GT	299Q>KQ	
ADHFE1	T>AT	309H>QH	no data
	A>AG	310M>MV	
AMACR	T>GT	310K>KQ	even expression in all
	C>AC	242E>XE	
	G>GT	241Y>XY	
ARMC4	G>CG	276T>RT	even expression in all
	T>AT	276T>ST	

Table 7 continued

DDX28	A>AT	-	even expression in all
	T>AT	-	
FOXRED1	A>AG	33K>KE	no data
	T>CT	130F>F	
GHITM	T>AT	327L>XL	multiple brain elevations (prefrontal cortex, pineal night/day, amygdala, hypothalamus, etc.)
	A>AT	327L>LF	
	A>AG	328N>NS	
GLS	T>GT	231I>IM	multiple brain elevations (occipital lobe, parietal lobe, prefrontal cortex, globus pallidus, etc.)
	G>GT	232D>YD	
GOT2	C>CG		721 B lymphoblasts, liver, whole brain, pineal night/day, retina
	G>CG	78P>AP	
HSPA9P	A>AC	-	no data
	A>AG	-	
	A>AC	-	
IDH3G	G>AG	354A>VA	heart, immune/blood, minor brain elevations (whole brain, amygdala, pineal night/day)
	C>CT	354A>AT	
IREB2	C>AC	-	superior cervical ganglion, trigeminal ganglion, skeletal muscle
	T>GT	-	
LDHD	A>AG	-	liver, even expression in rest
	A>AG	-	
	G>AG	218P>SP	
NFXL1	G>GT	248L>IL	no data
	C>AC	247W>CW	
PHYHIPL	A>AG	85I>IV	multiple brain elevations, including prefrontal cortex, amygdala, whole brain, occipital lobe
	T>CT	85I>TI	
PPTC7	G>CG	-	even expression in all
	A>AT	-	

Table 7 continued

SLC25A33	C>CG	87P>PA	no data
	A>AG	160Q>QR	
TIMM50	A>AC	-	even expression in all
	T>CT	-	
VAMP1	T>CT	-	multiple brain elevations, including prefrontal cortex, thalamus, spinal cord, parietal lobe, etc.
	T>CT	-	

APPENDIX B

CONSENT FORMS

Adult Consent Form

Division of Medical Genetics
Gerard Vockley, M.D., Ph.D.
Chief

One Children's Place
4401 Penn Avenue
Pittsburgh, PA 15224
Ph: (412) 692-5070
Fx: (412) 692-6472

CONSENT FOR AN ADULT TO ACT AS A PARTICIPANT IN A RESEARCH STUDY

TITLE: Use of whole exome and genome sequencing to identify new genetic disorders

Research Project Director:	Gerard Vockley, M.D., Ph.D., Chief of Medical Genetics University of Pittsburgh School of Medicine; Telephone: 412.692.7746

Other Key Research Team Members	
Lina Ghaloul Gonzalez, 412.692.5070	Stephanie DeWard 412-692-5232
M. Michael Barmada, 412.383.7959	

David Peters 412.624.5392	
Nancy Perrott, 412.692.3150	

SOURCE OF SUPPORT: Division discretionary funds

We are conducting research to understand the genetic basis for unknown genetic conditions. There are many genetic disorders which have already been identified. However, in some individuals, we may still suspect a genetic disorder even though a precise diagnosis is unknown. As a result of your/your family member's medical history and clinical testing, the genetic doctor thinks you/your family member (s) may have an unknown genetic disorder.

Your genetic material is a substance within the body, such as DNA and RNA, which is passed down from parents to children and can affect what types of diseases people have. DNA or deoxyribonucleic acid is the chemical inside the central part of a cell that carries the genetic instructions in humans and almost all organisms and makes the individual hereditary characteristics; RNA or ribonucleic acid is a chemical similar to single strand of DNA and determines the protein synthesis and the transmission of genetic information. In this study we will be studying genetic material from your blood. This research study will use new techniques to read all of the genetic information in your cells that might cause a health problem if it contained a mistake. These techniques are called whole exome and genome sequencing.

We are inviting you to participate in this study because the genetic doctor thinks you/your family member (s) may have a genetic disease. This study will allow us to test conditions to best sequence all the important DNA from individuals in your situation. We may also be able to identify the genetic cause of your/ your family's medical condition.

As part of this study, you will complete the following procedures in addition to tests ordered by your physician for routine care. Two tubes of blood will be drawn from you in the in the phlebotomy lab at Children's Hospital of UPMC by the lab personnel after the clinic visit and being seen by your medical genetics doctor The amount of blood in each tube will not exceed 10 ml total volume (about 2 teaspoons). The blood will then be frozen immediately and sent to the lab performing the DNA sequencing. No other procedures will be necessary for you to participate in this study.

There will be no need for specific follow-up appointments or outpatient visits related to this research study until we have the result of the research study which will be disclosed to you during a clinic visit regardless of being positive or negative. Positive results will be confirmed in a CLIA lab before being disclosed. Each subject will get his/her own result and not of the whole family.

We are also requesting your authorization or permission to review your medical records to record past, current, and future medical information from hospital and other medical facilities. We will obtain information concerning your diagnosis, health and family history, and results of any physical exams, tests of urine, blood, tissues, and any other tests, including results of genetic tests. We will use this information to determine whether you meet the conditions for participation in this study, and to help us understand the results of the genetic tests performed as part of this study. This identifiable information will be made available to members of the research team, for an indefinite period of time. The University of Pittsburgh Research Conduct may monitor this study and as the result of this monitoring may have access to your identifiable information.

We are also requesting your permission to re-contact you in the future regarding participation of your family/relatives in this study. You may refuse to be re-contacted in the future. Your decision will not affect your relationship with the University of Pittsburgh or the UPMC, nor will you lose any benefits that you might be eligible for because of this decision

Results of the research study will be disclosed to the subjects during a medical genetic clinic appointment with appropriate genetic counseling and plans for clinical follow up and testing. After the research study and verification studies are completed, your DNA sample will be stored indefinitely for future molecular studies related to the subject's condition and to compare to future planned whole exome/whole genome sequencing studies. This will be done by same researchers of this study. Upon participation in the research study and when stored, these samples will be given a case number and the code linking the name to this number will be maintained separately with very limited access to research team.

There are a number of possible risks, side effects, and discomforts associated with participation in this research study. The risks of each procedure are minimal and rare.

- **Blood draws**: Brief discomfort, bruising, slightly prolonged bleeding, infection at the site, scar noted at the site, the clotting of blood around the site, or fainting**.** Care will be taken to avoid these potential risks and discomforts. The blood draw will be obtained at the same time as other blood tests that your doctor will order for your routine care. If you are not having blood drawn for routine care, we will draw the blood during your study visit.
- **Because your genetic information is being used in this research study,** there is a rare risk that information could become accessible to people other than members of this research team. **Breaches in confidentiality** involving genetic information could impact future insurability, employability, or reproduction plans, or have a negative impact on family relationships, and/or result in paternity suits or stigmatization. To minimize these risks, genetic information (as well as your medical information) will only be recorded in files marked with case numbers, not your name.
- **There is also a possibility of learning life-altering results**. This will be managed by the appropriate counseling, support and provide the patient with the available treatment.

A new Federal law, called the Genetic Information Nondiscrimination Act (GINA), generally makes it illegal for health insurance companies and group health plans to use genetic information in making decisions regarding your eligibility or premiums. GINA also makes it illegal for employers with 15 or more employees to use your genetic information when making decisions regarding hiring, promoting, firing, or setting the terms of employment. This new Federal law does not protect you against genetic discrimination by companies that sell life, disability, or long-term care insurance.

If we learn of any new information about study risks that could cause you to change your mind about continuing to participate, we will notify you promptly.

Benefits of participation in this research study: There is no benefit to participation in this research study other than possibly increased knowledge about your/your family's disease. If a specific genetic disorder is identified, specific treatment may be available based on the information.

None of the procedures you receive during this research study (research blood draws or genetic analysis) will be billed to you or your health insurance. If you get a bill or believe your health insurance has been billed for something that is part of the study, notify a member of the research team. However, you or your insurer will be billed for all other usual care services, including routine surgery, blood draws for clinical/routine care, follow-up care, or testing done for clinical/routine purposes.

You will not be paid for your participation. Although it is possible that your biological samples may lead, in the future, to new inventions, discoveries or products that may be sold, licensed, or patented, there are currently no plans to share with you any money or other rewards that may result from the development of those new products.

If you believe that the research procedures have resulted in an injury to you, immediately contact Dr. Vockley or a member of the Research Team (see first page). Emergency medical treatment for injuries solely and directly related to your participation in this research study will be provided to you by the hospitals of UPMC. Your insurance provider may be billed for the costs of this emergency treatment, but none of those costs will be charged directly to you. If your research-related injury requires medical care beyond this emergency treatment, you will be responsible for the costs of this follow-up care. At this time, there is no plan for any additional financial compensation.

To protect your privacy and maintain the confidentiality of information we obtain from you and from your medical records, we will maintain all information about you in a secure location. This research study will involve the recording of current and/or future identifiable medical information from

your child's hospital and/or other (e.g., physician office) records. The information that will be recorded will be limited to information concerning your child's genetic disorder. All paper records that could identify you will be stored in locked file cabinets, and all electronic records will be stored in password-protected files. Your identity on these records will be indicated by a case number rather than by your name, and the code linking your name to this number will be maintained separately with very limited access to research team members. Although we will do everything in our power to protect your privacy and the confidentiality of your records, just as with the use of your medical information for health care purposes, we cannot guarantee the confidentiality of your research records, including information that we obtained from your medical records. However, **no third party, including relatives, personal physicians or insurance companies, or other researchers will have access to your identifiable information, with one exception**. Authorized representatives of the UPMC hospitals may have access to identifiable information only for the purpose of (1) filling orders made by the researchers for hospital and health care services (e.g., laboratory tests) associated with the research study, (2) addressing correct payment for tests and procedures ordered by the researchers, and/or (3) for internal hospital operations (e.g., quality assurance). Also, authorized representatives from the University of Pittsburgh Research Conduct and Compliance Office will have access to these files but only for the purpose of monitoring the conduct of the study.

Your doctor may also be involved as an investigator in this research study, but you are not under any obligation to participate in any research study offered by your doctor. Before agreeing to participate in this research study, or at any time thereafter, you may wish to discuss participation in this study with another health professional, to obtain a 'second opinion' about study participation. You may also contact the University 'Research Participant Advocate' 1-866-212-2668 for additional information.

Your participation in this research study is completely voluntary. Whether you participate/not participate in this research study will have no effect on your current or future relationship with the

University of Pittsburgh, UPMC or its affiliated health care providers or health care insurance providers. **If you decide you no longer wish to continue to participate** after you have signed the consent form, you should contact Dr. Vockley or his colleagues. Your blood samples and DNA will then be destroyed if they are not in the midst of being analyzed. You may also withdraw, at any time, your authorization to allow the research team to review your medical records, but if you do so, you will no longer be permitted to participate in this study. Any information obtained from you up to that point will, however, continue to be used by the research team. Your decision to withdraw from this study will have no effect on your current or future relationship with the University of Pittsburgh or with UPMC or its affiliated health care providers or health care insurance providers. However if withdrawal takes place, no information regarding results will be returned to you and your DNA sample will be destroyed so that no additional future testing can be performed. Results from the sequencing study obtained prior to withdrawal will still be analyzed to the extent possible.

VOLUNTARY CONSENT:

All of the above has been explained to me and all my current questions are answered. I understand I am encouraged to ask questions and voice concerns or complaints about any aspect of this research during the course of it, and that those questions, concerns, or complaints will be answered by the researchers listed on the first page of the form. I understand that I may always request that my concerns be addressed by a listed investigator. I understand that I may contact the Human Subjects Protection Advocate of the IRB Office, University of Pittsburgh (1-866-212-2668) to discuss any issues; obtain information; offer input; or discuss situations in the event that the research team is not available. By signing this form, I agree to participate in this research study. A copy of this consent form will be given to me.

_____

Participant's Name (Print)

_____ ----------

Participant's Signature Date

CERTIFICATION of INFORMED CONSENT:

I certify that I explained the nature and purpose of this research study to the above-named individual(s). I discussed the potential benefits and possible risks of study participation. Any questions the individual(s) have about this study have been answered, and we will always be available to address future issues that arise. I certify that no research component of this protocol was begun until after this consent form was signed.

_____ _____

Printed Name of Person Obtaining Consent Role in Research Study

_____ _____

Signature of Person Obtaining Consent Date

Pediatric Consent Form

Division of Medical Genetics
Gerard Vockley, M.D., Ph.D.
Chief

One Children's Place
4401 Penn Avenue
Pittsburgh, PA 15224
Ph: (412) 692-5070
Fx: (412) 692-6472

CONSENT FOR A CHILD TO ACT AS A PARTICIPANT IN A RESEARCH STUDY

TITLE: Use of whole exome and genome sequencing to identify new genetic disorders

Research Project Director:	Gerard Vockley, M.D., Ph.D., Chief of Medical Genetics University of Pittsburgh School of Medicine; Telephone: 412.692.7746

Other Key Research Team Members	
Lina Ghaloul Gonzalez, 412.692.5070	Stephanie DeWard 412-692-5232
M. Michael Barmada, 412.383.7959	Nancy Perrott, 412.692.3150
David Peters 412.624.5392	

SOURCE OF SUPPORT: Division discretionary funds

We are conducting research to understand the genetic basis for unknown genetic conditions. There are many genetic disorders which have already been identified. However, in some individuals, we may still suspect a genetic disorder even though a precise diagnosis is unknown. As a result of your medical history and clinical testing, your genetic doctor thinks your child/ your family member(s) may have an unknown genetic disorder.

Your genetic material is a substance within the body, such as DNA and RNA, which is passed down from parents to children and can affect what types of diseases people have. DNA or deoxyribonucleic acid is the chemical inside the central part of a cell that carries the genetic instructions

in humans and almost all organisms and makes the individual hereditary characteristics; RNA or ribonucleic acid is a chemical similar to single strand of DNA and determines the protein synthesis and the transmission of genetic information. In this study we will be studying genetic material from your child's blood. This research study will use new techniques to read all of the genetic information in his/her cells that might cause a health problem if it contained a mistake. These techniques are called whole exome and genome sequencing.

We are inviting your child to participate in this study because the genetic doctor thinks your child /your family member (s) may have a genetic disease. This study will allow us to test conditions to best sequence all the important DNA from individuals in your situation. We may also be able to identify the genetic cause of your/ your family's medical condition.

As part of this study, your child will complete the following procedure. Two tubes of blood will be drawn from your child in the in the phlebotomy lab at Children's Hospital of UPMC by the lab personnel after the clinic visit and being seen by your medical genetics doctor The amount of blood in each tube will not exceed 10 ml total volume (about 2 teaspoons). The blood will then be frozen immediately and sent to the lab performing the DNA sequencing. No other procedures will be necessary for you to participate in this study.

There will be no need for specific follow-up appointments or outpatient visits related to this research study until we have the result of the research study which will be disclosed to the parents/guardians during a clinic visit regardless of being positive or negative. Positive results will be confirmed in a CLIA lab before being disclosed.

We are also requesting your authorization or permission to review your child's medical records to record past, current, and future medical information from hospital and other medical facilities. We will obtain information concerning your child's diagnosis, health and family history, and results of any physical exams, tests of urine, blood, tissues, and any other tests, including results of

genetic tests. We will use this information to determine whether your child meets the conditions for participation in this study, and to help us understand the results of the genetic tests performed as part of this study. This identifiable information will be made available to members of the research team, for an indefinite period of time. The University of Pittsburgh Research Conduct may monitor this study and as the result of this monitoring may have access to your child's identifiable information.

We are also requesting your permission to re-contact you in the future regarding participation of your family/relatives in this study. You may refuse to be re-contacted in the future. Your decision will not affect your relationship with the University of Pittsburgh or the UPMC, nor will you lose any benefits that you might be eligible for because of this decision

Results of the research study will be disclosed to you during a medical genetic clinic appointment with appropriate genetic counseling and plans for clinical follow up and testing. After the research study and verification studies are completed, your child's DNA sample will be stored indefinitely for future molecular studies related to the subject's condition and to compare to future planned whole exome/whole genome sequencing studies. This will be done by same researchers of this study. Upon participation in the research study and when stored, these samples will be given a case number and the code linking the name to this number will be maintained separately with very limited access to research team.

There are a number of possible risks, side effects, and discomforts associated with participation in this research study. The risks of each procedure are minimal and rare, and occur in less than 1 time out of 100.

- **Blood draws**: Brief discomfort, bruising, slightly prolonged bleeding, infection at the site, scar noted at the site, the clotting of blood around the site, or fainting**.** Care will be taken to avoid these potential risks and discomforts. The blood draw will be obtained at the same time as other blood tests that your doctor will order for your child's routine care. If your child is not having blood drawn for routine care, we will draw the blood during their study visit.

- **Because your child's genetic information is being used in this research study,** there is a rare risk that that information could become accessible to people other than members of this research team. **Breaches in confidentiality** involving genetic information could impact future insurability, employability, or reproduction plans, or have a negative impact on family relationships, and/or result in paternity suits or stigmatization. To minimize these risks, genetic information (as well as medical information) will only be recorded in files marked with case numbers, not your child's name.
- **There is also a possibility of learning life-altering results**. This will be managed by the appropriate counseling, support and provide the patient with the available treatment
  A new Federal law, called the Genetic Information Nondiscrimination Act (GINA), generally

makes it illegal for health insurance companies and group health plans to use genetic information in making decisions regarding your eligibility or premiums. GINA also makes it illegal for employers with 15 or more employees to use your genetic information when making decisions regarding hiring, promoting, firing, or setting the terms of employment. This new Federal law does not protect you against genetic discrimination by companies that sell life, disability, or long-term care insurance.

If we learn of any new information about study risks that could cause you to change your mind about continuing to participate your child in the study, we will notify you promptly.

Benefits of participation in this research study: There is no benefit to participation in this research study other than possibly increased knowledge about your/your family's disease. If a specific genetic disorder is identified, specific treatment may be available based on the information.

None of the procedures you receive during this research study (research blood draws or genetic analysis) will be billed to you or your health insurance. If you get a bill or believe your health insurance has been billed for something that is part of the study, notify a member of the research team. However, you or your insurer will be billed for all other usual care services, including routine surgery, blood draws for clinical/routine care, follow-up care, or testing done for clinical/routine purposes.

You will not be paid for your child's participation. Although it is possible that your child's biological samples may lead, in the future, to new inventions, discoveries or products that may be sold, licensed, or patented, there are currently no plans to share with you any money or other rewards that may result from the development of those new products.

If you believe that the research procedures have resulted in an injury to your child, immediately contact Dr. Vockley or a member of the Research Team (see first page). Emergency medical treatment for injuries solely and directly related to your child's participation in this research study will be provided to your child by the hospitals of UPMC. Your insurance provider may be billed for the costs of this emergency treatment, but none of those costs will be charged directly to you. If his/her research-related injury requires medical care beyond this emergency treatment, you will be responsible for the costs of this follow-up care. At this time, there is no plan for any additional financial compensation.

To protect your child's privacy and maintain the confidentiality of information we obtain from your child and from his/her medical records, we will maintain all information about your child in a secure location. This research study will involve the recording of current and/or future identifiable medical information from your child's hospital and/or other (e.g., physician office) records. The information that will be recorded will be limited to information concerning your child's genetic disorder. All paper records that could identify your child will be stored in locked file cabinets, and all electronic records will be stored in password-protected files. Your child's identity on these records will be indicated by a case number rather than by his/her name, and the code linking his/her name to this number will be maintained separately with very limited access to research team members. Although we will do everything in our power to protect your privacy and the confidentiality of your child's records, just as with the use of his/her medical information for health care purposes, we cannot guarantee the confidentiality of your child's research records, including information that we obtained from medical records. However, **no third party, including relatives, personal physicians or insurance companies, or other researchers will have access to your child's identifiable information, with one exception**. Authorized representatives of the UPMC hospitals may have access to identifiable information only for the purpose of (1) filling orders made by the researchers for hospital and health care services (e.g.,

laboratory tests) associated with the research study, (2) addressing correct payment for tests and procedures ordered by the researchers, and/or (3) for internal hospital operations (e.g., quality assurance). Also, authorized representatives from the University of Pittsburgh Research Conduct and Compliance Office will have access to these files but only for the purpose of monitoring the conduct of the study.

Your child's doctor may also be involved as an investigator in this research study, but you are not under any obligation to give consent for your child to participate in any research study offered by your child's doctor. Before agreeing to participate in this research study, or at any time thereafter, you may wish to discuss participation in this study with another health professional, to obtain a 'second opinion' about study participation. You may also contact the University 'Research Participant Advocate' 1-866-212-2668 for additional information.

Your child's participation in this research study is completely voluntary. Whether you participate/not participate in this research study will have no effect on your/ your child's current or future relationship with the University of Pittsburgh, UPMC or its affiliated health care providers or health care insurance providers. **If you decide that your child no longer wishes to continue to participate** after you have signed the consent form, you should contact Dr. Vockley or his colleagues. Your child's blood samples and DNA will then be destroyed if they are not in the midst of being analyzed. You may also withdraw, at any time, your authorization to allow the research team to review your child's medical records, but if you do so, your child will no longer be permitted to participate in this study. Any information obtained from your child up to that point will, however, continue to be used by the research team. Your decision to withdraw from this study will have no effect on your/ your child's current or future relationship with the University of Pittsburgh or with UPMC or its affiliated health care providers or health care insurance providers. However if withdrawal takes place, no information regarding results will be returned to you and your child's DNA sample will be destroyed so that no

additional future testing can be performed. Results from the sequencing study obtained prior to withdrawal will still be analyzed to the extent possible.

VOLUNTARY CONSENT:

All of the above has been explained to me and all my current questions are answered. I understand I am encouraged to ask questions and voice concerns or complaints about any aspect of this research during the course of it, and that those questions, concerns, or complaints will be answered by the researchers listed on the first page of the form. I understand that I may always request that my concerns be addressed by a listed investigator. I understand that I may contact the Human Subjects Protection Advocate of the IRB Office, University of Pittsburgh (1-866-212-2668) to discuss any issues; obtain information; offer input; or discuss situations in the event that the research team is not available. By signing this form, I agree to participate in this research study. A copy of this consent form will be given to me.

_____

Child/Subject Name

"I understand that, as a minor (age less than 18 years), the above-named child is not permitted to participate in this research study without my consent. Therefore, by signing this form, I give my consent for his/her participation in this research study."

_____ _____

Parent's Name (Print) Relationship to Participant

_____ _____

Parent's Signature Date

CERTIFICATION of INFORMED CONSENT:

I certify that I explained the nature and purpose of this research study to the above-named individual(s). I discussed the potential benefits and possible risks of study participation. Any questions the individual(s) have about this study have been answered, and we will always be available to address future issues that arise. I certify that no research component of this protocol was begun until after this consent form was signed.

_____ _____

Printed Name of Person Obtaining Consent Role in Research Study

_____ _____

Signature of Person Obtaining Consent Date

ASSENT

This research has been explained to me and I agree to participate.

_____ _____

Signature of Child-Subject Date

VERIFICATION OF EXPLANATION:

I certify that I have carefully explained the purpose and nature of this research study to the child-subject in age appropriate language. He/she has had an opportunity to discuss it with me in detail. I have answered all his/her questions and he/she has provided affirmative agreement (i.e., assent) to participate in this study

_____ _____

Investigator's Signature Date

APPENDIX C

COMPARISON OF COMMERCIALLY AVAILABLE WES

Table 8 Comparison of Commercially Available WES by Laboratory

	Ambry	Baylor	GeneDx
Platform: Illumina	• WES on 3 relatives • Trios can be medically relevant people in pedigree, not just parents and proband. • Can call to determine who in pedigree they'd want samples from • Best scenario: send as many samples as possible. Other relations can be coseg	• WES with CGH-SNP for proband only • Targeted Sanger sequencing for parents on select variants (AR cis vs trans, or to determine de novo status VUS)	• Specimens on family members required (both parents, in most cases) • Additional affected and unaffected members may be required (call to check)
Testing	• Novel genes, VUS • Initial eval: no SIFT/PolyPhen, remove SNPs based on pop frequency, degree of conservation • Collect samples on all 1st degree relatives (parents/sibs) and "affected" cousins • Whole exome proband- if geno/pheno	• Novel genes, VUS • Extended family members can be tested for specific mutations for a fee ($410 if billing insurance, $385 if self-pay) • Moving towards trios but no launch date yet	• Novel genes/VUS? • Targeted testing of family members for specific mutations for $350-500 • Can do other testing in conjunction to exome (full sequence analysis and deletion testing of mtGenome, SNP array, whole genome array CGH)

Table 8 continued

	correlation found, stop there. • If not, reflex to exome 2 family members, identify mutation of interest. Then do coseg on other family members and confirm utilizing sanger sequencing on exome candidates		
Sample Collection	• Blood sample: 6-10cc in EDTA (purple top) or citric acetate (yellow top) tube per patient • Saliva acceptable for co seg • Cultured skin fibroblasts: 3 T25 cell flasks at 80% confluence • Kits provided by lab	• Blood: 5cc (child) or 10cc (adult) in EDTA (purple top) tube • Saliva acceptable for parents, but blood preferred • Cultured skin fibroblasts: 2 T25 flasks at 80-100% confluence • Kits provided by lab	• 5-10cc Blood, EDTA tube. • Specimens may be refrigerated for 7 days prior to shipping • High quality extracted DNA can be accepted (>15ug with minimum concentration of 50ng/ul) • Other tissue types – call to discuss
Reporting	• PROBAND ONLY • Related to phenotype • Early-onset or childhood onset HGMD genes/mutations (no VUS) • No formal report on WES candidates. Single site testing is available for an additional $250 • VUS unrelated to the phenotype will not be interpreted by lab but sheet can be requested by CHP	<u>FOCUSED REPORT</u> • Deleterious related to phenotype • VUS related to phenotype • Mito related to phenotype (>20% heteroplasmy, no VUS or rare variants) • Symptomatic XLR females • Unrelated but medically actionable- only known/deleterious (i.e. adult onset cancer [HNPCC, BRCA1/2], cardiomyopathies [DSP], coagulpoathies, collagen mutations, G6PD, undiagnosed secondary conditions [NF1/2, Marfan] relevant for parents) • AR carrier status • Pharmacogenetics (only warfarin and Plavix metabolism) • CNV >1 Mb (will report <1 Mb if known syndrome) • Family can opt out of AR carrier status and pharmaco <u>EXPANDED REPORT</u>	• PROBAND ONLY • ONLY related to phenotype • 3 classes of reporting: • Cat.1: positive result, known disease association • Cat.2: mutations possibly representing a diagnosis, felt to be likely but lack of definitive info to say for sure • Cat.3: candidate genes, no human disease associations, based on function, expression, and model organisms • Identification of a medically actionable mutation will result in call to ordering physician to see if they want the information

Table 8 continued

		• Deleterious unrelated to phenotype • Mito <20% heteroplasmy • Includes: disease causing, VUS, deleterious w/ no known assoc with human • Asymptomatic XLR females • heterozygous VUS assoc with AR d/o will only be reported only if deleterious or 2nd VUS in the same gene is detected	(examples so far: hereditary arrhythmia, an anemia unsure if was caught on NBS) • Otherwise, not reporting unrelated secondary diagnoses
Not Reported	• AR carrier status, predisposition to late onset disorder or cancer risk • Exception: childhood cancer (VHL/NF…), with some exceptions, and even then only HGMD genes are reported (no VUS) • drug metabolism or common disease	• Adult onset dementia syndromes with no treatment will not be reported in focused or expanded. Reportable but needs to be specifically requested. May report if suspicious hit that is "not certain but likely to be involved w/patient's disease", but would discuss w/doctor • no other loci/conditions blacked out	• variations in genes unrelated to phenotype • includes: benign variants, common diseases, carrier status, and adult onset
Turnaround Time	• CDE: 28 weeks	• 15 weeks • Expanded report can be ordered (with parent consent) up to 6 months afterwards for free, with 4 week turnaround	• 12-16 weeks
Limitations	• CNV, large rearrangements, tri-allelic, mito genome, epigenetic, trinucleotide repeat expansion, XLR females	• Large rearrangements, tri-allelic, epigenetic, trinucleotide repeat expansion,	• CNV, large rearrangements, tri-allelic, mito genome, epigenetic, trinucleotide repeat expansion, XLR females
Follow-up VUS	• Review of VUS are the responsibility of ordering physician/counselor	• Review of VUS are the responsibility of ordering physician/counselor	• Review of VUS are the responsibility of ordering physician/counselor
Price	• $7900 for trio	• $7000 for proband	• Proband only - $5000 • 2 parents and proband - $9000 • Each additional family member - $2500

Table 8 continued

Insurance	• pre-auth service provided- >$300 OOP=family notified	• pre-auth service provided	• Pre-auth services provided, credit card info required at time of sample submission • OOP costs will be limited to $1000 per trio (except in FL and CO) • Does not accept Medicare, Medicaid, or Tricare
Patented Genes	• report gene(not specific mutation) to Athena/Myriad and CHP simultaneously- CHP pursue seq w/ Athena/Myriad	• report very specific mutation/gene to CHP simultaneously- suggest pursue testing by licensed lab	• Will report results, including specific mutation, just not confirm with Sanger seq
Stored Data	• Keep data for two years • DNA is held until the proband is 18, or for 1 year if proband >18 at time of testing • If patient <18 yr, suggest re-sequencing	• Keep data for 10 years • Can re-interpret existing data after 1 year, for a small fee (amount currently undetermined) • Will eventually (after enough years/improved tech) recommend resequencing (like with arrays)	• Full sequence data held for at least 1 yr • Variant calls held indefinitely • Offers annual re-analysis (first one at no charge) or re-eval of other ones (cost not yet determined for subsequent re-analysis)
Cancellation of Testing	• If sequencing has not started, extracted DNA is held for 1 year unless instructions to do otherwise. • If sequencing has started, dealt with on case-by-case basis depending on where in the process it is.	• If sequencing has not started, purple-top tubes will be discarded. Extracted DNA will be held and possibly used for research under CLIA guidelines. • If sequencing has started, a report will still be generated	• If sequencing has started, can request that the data not be analyzed, but sample will still be sequenced and patient billed • Have5 business days to cancel before sequencing starts

BIBLIOGRAPHY

1. Majewski J, S.J., Lalonde E, Montpetit A, Jabado N. (2011). What can exome sequencing do for you? Journal of Medical Genetics 48, 580-589.
2. Lyon GJ, W.K. (2012). Identifying disease mutations in genomic medicine settings: current challenges and how to accelerate progress. Genome Medicine 4.
3. Rope AF, W.K., Evjenth R, Xing J, Johnston JJ, Sewnsen JJ, et al. . (2011). Using VAAST to identify an X-linked disorder resulting in lethality in male infants due to N-terminal acyltransferase deficiency. American Journal of Human Genetics 89, 28-43.
4. Robinson PN, K.P., Mundlos S. (2011). Strategies for exome and genome sequence data analysis in disease-gene discovery projects. Clinical Genetics 80, 127-132.
5. Clark MJ, C.R., Lam HYK, Karczewski KJ, Chen R, Euskirchen G, et al. (2011). Performance comparison of exome DNA sequencing technologies. Nature Biotechnology 29, 908-914.
6. Dohm JC, L.C., Borodina T, HimmelBauer H. (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res 36.
7. Applied Biosystems. (2012). "Overview of SOLiD™ Sequencing Chemistry". In (http://www.appliedbiosystems.com/absite/us/en/home/applications-technologies/solid-next-generation-sequencing/next-generation-systems/solid-sequencing-chemistry.html. Retrieved March 27, 2013
8. Illumina. (2013). "Sequencing Technology". In (http://www.illumina.com/technology/sequencing_technology.ilmn.) Retrieved March 27, 2013
9. Rusk, N. (2011). Torrents of sequence. Nat Meth 8, 44-44.
10. Life Technologies. (2013). "IonProton System". In (http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Sequencing/Semiconductor-Sequencing/proton.html.) Retrieved March 27, 2013
11. Purushothaman S, T.C., Ou CP. (2006). Protons and single nucleotide polymorphism detection: A simple use for the Ion Sensitive Field Effect Transistor. Sensors and Actuators B: Chemical 114, 964-968.
12. Ku CS, N.N., Pawitan Y. (2011). Revisiting Mendelian disorders through exome sequencing. Human Genetics 129, 351-370.
13. Ng SB, B.A., Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, et al. . (2010). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nature Genetics 42, 790-793.
14. Lynch, M. (2010). Rate, molecular spectrum, and consequences of human mutation. Proceedings of the National Academy of Sciences of the United States of America 107, 961-968.

15. Vissers LELM, d.L.J., Gilissen C, Janssen I, Steehouwer M, de Vries P, et al. . (2010). A de novo paradigm for mental retardation. Nature Genetics 42, 1109-1112.
16. Illumina. (2010). "Quality Scores Overview". In. (http://www.illumina.com/truseq/quality_101/quality_scores.ilmn.) Retrieved March 27, 2013
17. Genomics, T.A.C.o.M.G.a. (2012). Policy Statement: Points to Consider in the Clinical Application of Genomic Sequencing. In. (The American College of Medical Genetics and Genomics Website.
18. Saunders CJ, M.N., Soden SE, Dinwiddie DL, Moll A, Abu Alnadi N, et al. (2012). Rapid whole-genome sequencing for genetic disease diagnosis in neonatal intensive care units. Science Translational Medicine 4, 1-12.
19. Liu W, Q.C., Francke U. (1997). Silent mutation induces exon skipping of fibrillin-1 gene in Marfan syndrome. Nature Genetics 16, 328-329.
20. Lettice LA, H.A., Devenney PS, Hill PE. (2008). Point mutations in a distant sonic hedgehog cis-regulator generate a variable regulatory output responsible for preaxial polydactyly. Human Molecular Genetics 17, 978-985.
21. O'Daniel JM, L.K. (2012). Whole-Genome and Whole-Exome Sequencing in Hereditary Cancer: Impact on Genetic Testing and Counseling. The Cancer Journal 18, 287-292.
22. 252010: Mitochondrial Complex 1 Deficiency. In. (Online Mendelian Inheritance in Man), Johns Hopkins University. Retrieved January 27, 2013 from http://omim.org/entry/252010#reference33
23. Wang JL, Y.X., Xia K, Hu ZM, Weng L, Jin X, et al. (2010). TGM6 identified as a novel causative gene of spinocerebellar ataxias using exome sequencing. Brain 133, 3510-3518.
24. 606639: Mitochondrial elongation factor G1; GFM1. In. (Online Mendelian Inheritance in Man, Johns Hopkins University. Retrieved February 25, 2013 from http://omim.org/entry/606639
25. G elongation factor, mitochondrial 1. In. (GeneCards, Weizmann Institute of Science.) Retrieved February 25, 2013 from http://www.genecards.org/cgi-bin/carddisp.pl?gene=GFM1&search=GFM1
26. 609060: Combined oxidative phosphorylation deficiency 1; COXPD1. In. (Online Mendelian Inheritance in Man. Johns Hopkins University. Retrieved February 25, 2013 from http://omim.org/entry/609060
27. 603645: PET112, S. cerevisae, homolog-like; PET112L. In. (Online Mendelian Inheritance in Man). Johns Hopkins University. Retrieved February 25, 2013 from http://omim.org/entry/603645?search=PET112L&highlight=pet112l
28. PET112 homolog (yeast). In. (GeneCards, Weizmann Institute of Science. Retrieved February 25, 2013 from http://www.genecards.org/cgi-bin/carddisp.pl?gene=PET112&search=PET112L
29. National Society of Genetic Counselors. (2006). "Code of Ethics". In. (http://www.nsgc.org/Advocacy/NSGCCodeofEthics/tabid/155/Default.aspx). Retrieved March 27, 2013
30. Tabor HK, S.J., Brazg T, McMillin MJ, Dent KM, Yu JH, et al. (2012). Informed consent for whole genome sequencing: A qualitative analysis of participant expectations and perceptions of risks, benefits, and harms. American Journal of Medical Genetics Part A 158A, 1310-1319.

31. Wilfond BS, C.K. (2008). Incidental findings in pediatric research. Journal of Law, Medicine, & Ethics, 332-340.
32. Wilfond BS, D.D. (2012). Engaging pediatric patients in genomic research. Genetics in Medicine 14, 437-443.
33. Ambry Genetics. (2012). Clinical Diagnostic Exome Detection Rates. In. (http://www.ambrygen.com/sites/default/files/Exome%20detection%20rates%20%28nov%202012%29.pdf) Retrieved March 3, 2013.
34. Yu Y, W.B., Wu J, Shen Y. (2008). Exome and Whole-Genome Sequencing as Clinical Tests: A Transformative Practice in Molecular Diagnostics. Clin Chem 58, 1507-1509.
35. Knott M, L.H., Downs J. (2011). The diagnostic odyssey to Rett Syndrome: The experience of an Australian family. American Journal of Medical Genetics Part A, 10-12.
36. Worthey EA, M.A., Syverson GD, Helbling D, Bonacci BB, Decker B, et al. . (2011). Making a definitive diagnosis: Successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. Genetics in Medicine 13, 255-262.
37. Bainbridge MN, W.W., Murdock DR, Friedman J, Gonzaga-Jauregui C, Newsham I, et al. . (2011). Whole-Genome Sequencing for Optimized Patient Management. Science Translational Medicine 3, 1-5.
38. Goh V, H.D., Biank V, Jarzembowski J, Dimmock D. (2012). Next generation sequencing facilitates the diagnosis in a child with acute cholestatic liver failure. Journal of Pediatric Gastroenterology 54, 291-296.