

UNIVERSITY OF CALIFORNIA,
IRVINE

Diagnostic Odyssey: Investigating the Role of a Mutation in the CUL4B Gene and its Effect on
the CUL4B Protein Complex in a Patient with Seizures and Developmental Delay

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Genetic Counseling

by

Desiree Basila

Thesis Committee:

Professor Moyra Smith, Chair
Adjunct Professor Pamela Flodman
Professor Virginia Kimonis

2015

DEDICATION

To

My Family - my Mother whose compassionate heart blessed all her knew her, my Father whose creativity and love of learning was unending to his last breath and to my amazing sisters and brothers, especially my sister, Sheree, whose love, courage and optimism have always been a source of grace in my life.

To

My amazing healthcare team – Shelley Hwang, B.J Miller, Anand Dhruva, Sandra Chu, Diane Shumay, Sharon Kman, Beverley Burns, and so many others at UCSF who kept me alive body and mind, heart and soul. Know that I strive to give back in kind, if only in small ways, anywhere I can.

To

My wonderful friends from so many walks in life, all ages, old and new. You have all brought so much warmth, beauty, laughter and hope into my life. I am so grateful for you all. I thank most especially Ara Norwood and Hector Lee, life-long friends, without whom this work simply would not have been possible.

“Your work is your love made visible.”

Khalil Gibran

The Prophet

TABLE OF CONTENTS

	Page
LIST OF FIGURES	iv
LIST OF TABLES	v
ACKNOWLEDGMENTS	vi
ABSTRACT OF THE THESIS	vii
INTRODUCTION	1
The Diagnostic Odyssey	1
Overview of Intellectual Disability	3
Overview of Variant Classification	4
Overview of Splice Site Variants	14
Role of CUL4B: Ubiquitination	19
Structure of the CUL4B Protein and Protein Complex	20
Review of Patient Reports in Literature	27
METHODS	30
Patient Profile	30
Exome Sequencing	31
mRNA Studies	32
RESULTS	34
DISCUSSION	37
CONCLUSION	49
REFERENCES	51

LIST OF FIGURES

		Page
FIGURE 1	RNA Molecular Structure of RNA	15
FIGURE 2	Patterns of mRNA splicing	18
FIGURE 3	The Cullin Family Protein Complexes	24
FIGURE 4	Alternative Splicing and Location of PCR Primers	33
FIGURE 5	PCR Products Resulting From Primer Sets	36
FIGURE 6	Blast Analysis of PCR Product Obtained in Experiment Three	37
FIGURE 7	Interactions Impacting Stability of CUL4B	39
FIGURE 8	CUL4B Gene Expression Studies	45

LIST OF TABLES

	Page
Table 1 Summary of Variants Identified in Selected Genes	41
Table 2 Summary of Variants Identified in CUL4A	43
Table 3 Comparison of Patients With Splice Site Mutation in c.1906+1	48

ACKNOWLEDGMENTS

I wish to express my gratitude

To

Moyra Smith whose lively curiosity, love of learning and sense of wonder encouraged the same in me, reminding me that nature is infinitely beautiful,

Virginia Kimonis whose caring commitment to her patients and enthusiasm for creating new understanding in medicine has been and will continue to be an inspiration,

Pam Flodman who taught me there is more than one way to skin a cat – even if the “cat” is the human heart. Her skill as a counselor, educator and thinker, and her warm humanism have challenged me to deepen my own humanity,

Our patient and his family whose courage and generosity have helped create new understanding of the CUL4B gene. May they know that their efforts are what make advances in medicine possible.....and what make hope possible.

I would also like to thank Elizabeth Chao for her help in obtaining permissions to use data from Ambry Genetics, Robert Huether for the generous use of his figures and analysis, and Eric Bennett for permission to use figure 3.

ABSTRACT OF THE THESIS

Diagnostic Odyssey: Investigating the Role of a Mutation in the CUL4B Gene and its Effect on the CUL4B Protein Complex in a Patient with Seizures and Developmental Delay

By

Desiree Basila

Master of Science in Genetic Counseling

University of California, Irvine, 2015

Professor Moyra Smith, Chair

For many patients with complex diseases the time from onset of symptoms to diagnosis may involve many years. If a diagnosis can be reached, it often comes at great financial and emotional cost. With the advent of sequencing technologies, the ability to identify gene variants has increased exponentially. Interpreting these findings, however, remains challenging. The purpose of this study was to analyze a mutation in the CUL4B gene revealed by whole exome sequencing (WES) in a patient with seizures and developmental delay. Recognizing that complex disease may involve the interaction of many genes and their protein products, an analysis of variants in selected CUL4B associated proteins and CUL4A was also undertaken with the ultimate goal of determining their potential clinical relevancy.

WES was performed by Ambry Genetics and mRNA studies were performed by University of California CLIA laboratory MitoMed. WES identified a variant in CUL4B located on the X chromosome. Analysis revealed a splice site mutation resulting in skipping of exon 14

or inclusion of intron 14. Examination of the biological structure indicated that in either case no functional protein would be found. Fifteen intronic variants and one synonymous exonic variant were identified in CUL4B related genes. Fourteen intronic variants in CUL4A were also identified. Current understanding of these types of variants suggests they would not be contributory, however further study is required. Taken together, these results contribute to the growing body of evidence that mutations in the CUL4B gene may be causal in cases of X linked intellectual disability.

INTRODUCTION

“The burden which those who are undiagnosed and misdiagnosed carry is a huge one to bear. Genomics promises little in the way of outright cures at the current time, but what it does have is the power to lift a weighty burden and provide those who are mysteriously ill with one of the most valuable and soul-soothing gifts they can ever be given...

...a definitive answer.” Patient (Maisel, 2014)

THE DIAGNOSTIC ODDISEY

For patients with rare or complex diseases the time between when a patient first becomes concerned about a health issue and when a diagnosis is received may be many years. These patients may undergo a long series of investigations, tests, and referrals to find answers. The term “diagnostic odyssey” has become a common expression in describing this tortuous journey. For patients and their families who are embroiled in such a journey the challenges are many, and these challenges are especially difficult when they involve children. Access to support services is often dependent on a diagnosis. Parents find that caring for their child becomes more demanding with age rather than less so (McCann et al., 2012). Importantly, the uncertainty of having no explanation for a child’s illness creates an emotional burden stemming from simply not knowing what to expect. Thus the lack of validation and prognosis compounds the struggle to care for the child (Carmichael et al., 2014).

With the advent of next generation sequencing (NGS) it has become increasingly common to find a genetic explanation for rare or complex disorders. Once a diagnosis is made, there are many practical and psychosocial benefits to a family. Parents are able to obtain therapies and make necessary modifications to the living environment. The family as a whole

can begin the process of adapting to a different lifestyle. Parents can begin the process of creating support networks by contacting other parents, local support groups and online organizations. In the exchange of stories and shared experiences families learn from each other (Carmichael et al., 2014).

Perhaps the most significant benefit is the validation that comes with a unifying diagnosis (Carmichael et al., 2014). In a 2009 study, Makela et al. identified seven categories for which parents felt a diagnosis was of value for their children with intellectual disability: (1) validation, (2) information, (3) obtaining services or therapies in the school system, (4) early intervention, (5) support, (6) need to know (curiosity), and (7) prenatal diagnosis. Of these seven, validation was the reason most commonly put forth for wanting a diagnosis. “Parents felt a diagnosis established proof of a credible problem, providing them mental and emotional relief that would, in turn, give them confidence and empowerment to advocate for their child.” (Makela et al, 2009).

The aim of this work is to examine the changes observed in the exome of a patient with intellectual disability to determine if these findings are relevant to the phenotype of the patient and can be put forth as a diagnosis. Mutations in the CUL4B gene are associated with a syndromic form of X-linked mental retardation characterized by a range of features including severe intellectual deficit, short stature, hypogonadism and abnormal gait. Neurological traits include speech delay, tremor, seizures, hyperactivity and decreased attention span. The protein complex is also thought to play a role in cell proliferation and DNA damage repair and as such may have a role in human cancers (UniProt, 2015).

A second related aim of this project is to better understand the relationship between gene mutations which result in changes in a particular protein domain and how these changes manifest

in the phenotype by analyzing the known splice site mutation and its product in a patient with disease manifestations. Further analysis will include mutations reported in the literature, the types of mutations (missense, frameshift, nonsense) and how these mutations effect phenotypic outcome given the domain in which they occur. These findings will be compared to the mutations revealed in the exome of the patient in this study with specific focus on CUL4B

In order to achieve these aims several lines of evidence will be explored. These include the identification of a candidate mutation, annotation of the gene and its resulting protein product, literature review of other reported patients with similar mutations and phenotypes, functional studies, and familial cosegregation.

OVERVIEW OF MENTAL RETARDATION/ INTELLECTUAL DISABILITY

“Mental retardation is a condition of arrested or incomplete development of the mind, which is especially characterized by impairment of skills manifested during the developmental period, which contribute to the overall level of intelligence, i.e. cognitive, language, motor, and social abilities. Retardation can occur with or without any other mental or physical disorder.”(WHO ICD-10). This definition, put forth by the World Health Organization, describes a condition which in current practice is referred to as intellectual disability, although some papers may use the term mental retardation.

The etiology of intellectual disability has been a focus of many studies. It may be caused by constitutional genetic abnormalities in chromosomes such as large deletions, duplications or aneuploidies that affect multiple genes. A second cause is mutations in individual genes. These diseases may have an autosomal dominant, autosomal recessive or X-linked inheritance pattern.

Although there has been some success in identifying these genes, many encode proteins with multiple functions or unknown roles (Tarpey et al., 2009). Additionally, gene-gene interactions, protein complexes and protein networks make it difficult to determine the etiology of the condition. Intellectual disability illustrates the challenges of understanding the etiology of complex conditions in which many genes and their protein products may be involved. These conditions may not be caused by a single gene variant, but the cumulative effects of multiple genes and/or their roles within the larger biological construct.

OVERVIEW OF VARIANT CLASSIFICATION

While powerful new technologies are making it possible to detect changes in DNA with ever increasing accuracy and resolution, interpreting these changes to achieve a meaningful diagnosis remains a challenge. Historically, mutations in a significant portion of the gene such as an exon, splice site or promoter region that directly affected the expression and function of the associated protein were considered as possible explanations for a condition based on Mendelian principles. With a deepening understanding of the underlying biological mechanisms, it is becoming increasingly clear that the relationship between a variant and clinical manifestation may be more complex. A given DNA mutation and its resulting protein may act as part of a protein complex or within a protein network to ultimately affect phenotype. Mutations in genes with predisposing, triggering, modifying, and protective functions may all play a role in causing disease. Digenic or polygenic disorders may require more in depth analysis in order to establish the underlying mechanism by which a variant in a given gene may be contributing to the disease process relative to other genes and their protein products (Quintans et al., 2014). Thus, while

technologies such as next generation sequencing (NGS) have increased access to sequencing data, interpretation of that data for use in making clinical decisions remains complex. One possible approach for establishing clinical relevancy put forth by Quintans et al. is to identify a mutation and then analyze it using a process that includes annotation, literature review, functional studies, experimental studies and familial cosegregation (Quintans et al., 2014).

Genetic testing is one avenue for identifying the presence of variants that may be disease causing. Exome sequencing of the human genome with coverage of at least 100x is estimated to reveal an average of between 20,000 and 30,000 single nucleotide variants (SNV) and small insertions or deletions (indels). While the term SNV and single nucleotide polymorphisms (SNP) may sometimes be used interchangeably, a SNP is defined as occurring in >1% of the population (Nature/Scitable, 2015). By contrast, SNV is a general term for any single nucleotide variant that is identified as different from the consensus. It has no frequency specification and may be rare or common. In order to identify those variants most likely to be clinically significant, less reliable variant calls are eliminated in a process referred to as filtering. One means of filtering is to remove variants with low coverage or low quality. Other filtering strategies may include eliminating variants located in SNV clusters, or excluding those with poor read alignment which indicate possible error in sequencing. A second means of evaluation is through comparison of the variants detected in family members. For example a variant that presents in a child but not in either parent is not consistent with Mendelian laws. Such a variant could be further evaluated to determine if it represents a sequencing error or is a *de novo* mutation in the child (Bao et al., 2014). However, germline mosaicism or false paternity might also be considered.

While most analysis tends to focus on the non-synonymous SNVs, (those variants in the DNA which would result in an amino acid change in the protein product), and indels (insertions and deletions) found in the protein coding regions, important information can be obtained from synonymous SNVs as well (Bao et al., 2014). For example, some missense variations may have no effect on protein function while a synonymous variant which introduces a cryptic splice site can be pathogenic (Tarpey et al., 2009).

Examination of existing data bases may also reveal the frequency with which a given variant sequence occurs in random individuals relative to the population as a whole. If the variant identified in an individual is also present above a certain frequency threshold in the general population it is less likely to be disease causing or may not be highly penetrant (Quintans et al., 2014). This type of information can, in turn, be used to identify a mutation that is clinically relevant rather than a DNA change which reflects polymorphisms within a population.

Care should be taken when performing this type of analysis, however. Quintans et al. point out that some potentially pathogenic mutations may occur in the general population by chance which may lead to misinterpretation of these variants as benign. Alternatively, given that the frequency of most DNA variants found in the human genome are less than 1%, a variant that is harmless may not be found in the general population. Thus, the fact that a variant under investigation is not detected in a control may reflect the fact that it is rare, rather than pathogenic. Importantly, a benign variant that appears with some frequency in one population may be deleterious in another population with differing haplotype, epigenetic or environmental considerations (Quintans et al., 2014).

Once a variant has been identified several steps may be taken to determine if it is significant to the observed phenotype and to what degree. Annotation is the process by which

the variant is described by comparison to a standard reference sequence with the goal of determining the variant position and functional classification (Quintans et al., 2014). Annotation characteristics may include genomic features, potential role in exon function, and the nature of the amino acid change in the resulting protein (Bao et al., 2014). Establishing the relationship between the variant or gene and disease provides further evidence that may indicate a role for the variant. Quintans et al. point out that this may include examining sources that describe cellular pathways, gene expression, protein domain structure and modeling. For example, a gene that is highly conserved across species may suggest that its product is essential to life processes. Additionally if it can be established that the variant is present in other patients with a similar phenotype it may support a role for that variant in causing the condition (Quintans et al., 2014)

Understanding the various interrelationships of sequence variation and phenotypic outcome requires interpretation of an ever-increasing body of data. Key to successful interpretation and annotation is the use of bioinformatics tools. Luscombe et al. describe bioinformatics as “conceptualizing biology in terms of macromolecules (in the sense of physical-chemistry) and then applying ‘informatics’ techniques (derived from disciplines such as applied maths, computer science, and statistics) to understand and organize the information associated with these molecules, on a large-scale.” (Luscombe et al., 2001). Bioinformatics tools are essential to identifying a mutation and establishing the function of a sequence whether it is coding or has some other function. This process is the basis for determining the pathogenicity of a variant and a possible role in diagnosis.

One preliminary approach is to annotate the variant by comparing it to the sequences reported in dbSNP, a database which catalogues observed variation in the DNA sequence. Single nucleotide polymorphisms (SNPs) are the most common variations observed, occurring

approximately once every 500-1000 bases. The name, dbSNP is somewhat misleading however since in addition to SNPs, short deletion and insertions, microsatellite markers, and elements such as retrotransposons are also included in the database. A wide variety of contributors ranging from individual laboratories to private industry may submit to the database (Sherry et al., 1999). While dbSNP remains a valuable means of cataloguing the burgeoning information being created regarding the human genome, it is important to note that there may be errors in base-calling that affect the quality of the reported sequences. These artifacts which may be misinterpreted as SNPs are referred to as single nucleotide differences (SNDs). Arthur et al. have noted that the incidence of SNDs has dropped and they may represent about 2.2% of all biallelic SNPs. While improvements in detection and curation have improved quality generally, the presence of SNDs in the database may compromise research in human genomics (Arthur et al. 2014). This example underscores the importance of using a variety of databases and approaches when characterizing a variant and establishing pathogenicity.

The functional study of a genetic variant includes analysis of the genetic mutations and the possible outcomes in the protein product, including missense, splicing effect, and nonsense mutations, among others. Importantly, these functional classifications alone cannot establish a variant's pathogenicity. For example, a truncating mutation which is often considered to be deleterious may have no effect in a dominant disease typically caused by a mutation which results in an abnormal gain of function (Quintans et al., 2014). Thus a variant must be seen in relation to other factors that determine function and phenotypic manifestations.

Variants are often analyzed using in silico models which are designed to predict the likelihood that a given variant will have a deleterious effect on the encoded protein. Models designed to promote understanding of the structure of the particular amino acids in a protein give insight into

the biochemical properties and three dimensional structure of the protein and its function (Quintans et al., 2014). For example SIFT (Sorting Intolerant From Tolerant) is a program that predicts whether the substitution of an amino acid will have an effect on protein function based on the resulting protein amino acid sequence. (Ng et al. 2003; sift.jcvi.org). SIFT takes as its premise that amino acids essential to protein function will be conserved in a protein family. Thus changes in highly conserved positions are likely to be predicted as deleterious. If, for example, isoleucine is the only amino acid found at a particular position in an alignment, substitution of any other amino acid is presumed to have been selected against evolutionarily suggesting that isoleucine is essential to protein function and substitution at that position is likely deleterious. By similar reasoning, if the hydrophobic amino acids isoleucine, valine and leucine are found in a position in an alignment, SIFT assumes that the hydrophobic attribute is required at this position. A change to other hydrophobic amino acids at this position would likely be predicted to be tolerated, while changes resulting in a substitution of a charged or polar amino acid would be expected to have an effect on the protein function. (Ng et al. 2003).

While SIFT makes predictions based solely on amino acid sequence, PolyPhen (Polymorphism Phenotyping) also examines 3D-structure (<http://genetics.bwh.harvard.edu/pph2>). This model takes into account the specific site at which a substitution may occur, considering for example, active or binding sites, and globular versus non-globular regions such as a transmembrane site. Other considerations might include sites annotated as lipid, metal, or carbohydrate, as well signaling regions. Bonds, such as disulfide or thioesters, and molecular structures such as coil regions are also noted. The next step is to compare the variant sequence to a known protein 3D structure. This is achieved using the Basic Local Alignment Search Tool (BLAST) which identifies local regions of similarity between

sequences by comparing protein (or nucleotide) sequences to databases and then performing statistical analysis to identify matches. This comparison reveals what impact the amino acid substitution may have on important aspects of the protein such as destruction of the hydrophobic core, electrostatic interactions, solubility, or interaction with ligands (PolyPhen Help Text, 2015; Sunyaev et al., 2001).

Also important to the analysis are structural parameters and contacts. PolyPhen considers what changes an amino acid substitution might cause in the resulting protein with regard to the solubility, the potential effect of the alternate side chain, or changes in the ability of the polypeptide backbone to rotate (Phi-psi dihedral angles). An amino acid residue may also have specific spatial contacts which are central to its role in the protein. A contact may serve to determine the protein structure or the procedure by which that structure is achieved. A polymorphism may change the interaction between polypeptide chains which in turn affect quaternary structure. They may also interact directly with residues essential to the protein biological function such as the binding, active site, and lipid or metal sites identified from sequence annotation. PolyPhen integrates these various data points to predict the effect of a non-synonymous SNP on protein function which, in turn, may inform an interpretation of the mutation's clinical relevance (PolyPhen Help Text, 2015)

It is noteworthy that current understanding of sequence variation outside the coding regions of the gene remains limited. Regulation of gene expression varies between species and between tissues within a species. Variants in the splicing regions, untranslated regions (UTR), promoters and other regulatory regions may prove more difficult to interpret using predictive models (Quintans et al., 2014). Yet it is becoming increasingly clear that intronic regions may

play a significant role in gene regulation and expression. By extension mutations in these regions may be contributing to disease manifestations.

One system for studying and interpreting these regions is the ENCODE Project database. The Encyclopedia of DNA Elements (ENCODE) project has as its goal the delineation of all functional units encoded in the human genome. The ENCODE Project Consortium defines a functional unit or element as a particular DNA segment that encodes a specific product such as a protein or non-coding RNA or has an identifiable biochemical function such as protein binding or chromatin structure. Since 2007, the ENCODE project has used various approaches to map functional units in the human genome. Some of these methodologies and the elements mapped are described below:

1. RNA-seq: Isolation of RNA sequences
2. CAGE-Analysis of the RNA 5' methylated cap added at the beginning of transcription
3. RNA-PET Captures RNAs with both 5' methyl cap and poly A tail. Taken together this assays full length RNAs for further study
4. ChIP-seq – Uses chromatin immunoprecipitation determine regions of chromatin most often bound by proteins. Antibodies directed at transcription factors, chromatin binding proteins and chemical changes to histone proteins to identify genomic DNA bound by these proteins.
5. DNase seq- An assay used to identify regions that are hypersensitive to DNase I indicating open chromatin
6. FAIRE-seq – Isolates regulatory regions.
7. RRBS – Reduced representation bisulphite sequencing uses the fact that DNA treated with bisulphite will result in the conversion of unmethylated cytosines to uracil. Methylated cytosines are protected from conversion and can thus be identified.

(ENCODE Project Consortium)

The ENCODE Project Consortium has noted that the fraction of bases in non-coding DNA thought to be involved in direct gene regulation is significantly higher than the fraction of bases involved in protein coding exons. They suggest this may imply that more information in the human genome is focused on gene regulation rather than biological/biochemical function. (ENCODE Project Consortium). Thus, the role of functional units in non-coding regions of DNA takes on new significance and variants in these regions emerge as possible cause for disease.

Another avenue for establishing pathogenicity is experimental evidence. Patient samples may be used to evaluate abnormal gene expression by RNA or protein analysis. Additional studies can include introducing the variant into cells or animal models. If the variant causes changes analogous to the phenotype in the donor and if the observed abnormalities are abrogated by experimentally restoring wild type pathogenicity is supported (Quintans et al., 2014)

Familial co-segregation is yet another important consideration in variant interpretation. The causal role of a given variant is supported if it is present in all affected family members. However, when evaluating genetic variants in a clinical diagnostic setting, candidate genes are often evaluated one by one. While this may be useful for disorders that follow a Mendelian inheritance pattern, it may obscure the analysis given that other variants in the genome may be present an individual patient which also influence clinical presentation.

Synthesizing the lines of evidence in order to arrive at the most likely role for a given variant requires both quantitative and qualitative analysis to achieve a classification. One commonly used system describes variants as benign, probably benign, of uncertain significance, probably pathogenic, and pathogenic. This system is somewhat over simplified and is mostly applicable to Mendelian disorders. In more complex disease other categories are used to further

describe a genetic variant and its role in disease. These may include susceptibility variant, prognostic factor, trigger, modifier and protective. As the names suggest, these variants may not be directly causal, but rather have a role in protein networks or metabolic pathways that when disrupted, in turn, cause disease (Quintans et al., 2014).

Complex disease requires stratification of the genotype-phenotype relationship being examined whether it be at the cellular level, organ level or in the individual organism. A given variant may be deleterious at one of these levels, but not necessarily cause disease or have a deleterious effect in the others. Thus determining that a variant is deleterious to the function of an encoded protein does not, de facto, translate into clinical relevance (Quintans et al., 2014). When seeking a diagnosis for more complex disease, such as intellectual disability it is essential to consider the intricate biological interactions that may define the relationship between a variant and the manifestation of disease. This requires analyzing the variant from different perspectives in order to determine its clinical significance in diagnosis.

As research begins to reveal the mechanisms by which genes are regulated and expressed in addition to the relationship between protein structure and function, bioinformatics tools are being developed to help determine the roles mutations may have in disease. One example of a tool that combines annotation from multiple sources is ANNOVAR. Bao et al. note that this program includes three annotation platforms, gene-based, region-based, and filter-based, assimilating over 4,000 public databases for annotation. These include dbSNP, 1000Genomes, ESP6500, Complete Genomics personal genomes and NCI-60 human tumor cell line panel exome sequencing data which provides information to evaluate minor allele frequency (MAF). ANNOVAR includes seven programs from the LJB23 database (or dbNSFP, database for nonsynonymous SNPs' functional predictions) and Combined Annotation Dependent Depletion

(CADD) database to enable prediction of deleterious variants. Scores are included to assess the conservation of the mutated site of interest across 29 mammalian species. The program takes into account experimental evidence for pathogenesis by referring to disease variant databases such as COSMIC (Catalogue Of Somatic Mutations In Cancer) (Bioinformatics Primer, 2015; <http://cancer.sanger.ac.uk/cosmic>) and ClinVar (<http://www.ncbi.nlm.nih.gov/clinvar/>) which is a compilation of reported relationships among human variations and phenotypes (Bao et al., 2014). Programs such as ANNOVAR take advantage of multiple databases and their varied strengths to analyze and assess a variant from multiple perspectives, thus increasing the ability to use genetic information in the clinical setting.

OVERVIEW OF SPLICING MECHANISM

The Central Dogma of biology refers to the two step process, transcription and translation, by which information flows from DNA to mRNA to produce protein in the cells as described below:



It is well understood that DNA sequences encode specific products and also contain information specifying when, where and how these products are made. The DNA sequence is a guide for the construction of an intermediate molecule, mRNA, through the process of transcription. (Griffith et al., 2010) Unlike DNA, RNA is single stranded and as a result more flexible. It is capable of forming complex, three dimensional shapes. In some cases, this allows

the RNA strand to bend in such a way that its own bases may form base pairs with each other. A second important distinction of the RNA molecule is the presence of the pyrimidine base uracil instead of thymine. Like thymine, uracil forms two hydrogen bonds with adenine, but in addition uracil can also bond with guanine. This occurs only during RNA folding, however, not during transcription. This ability of U to pair with both A and G gives the RNA molecule its ability to form complex structures which are important in facilitating essential biological processes. RNA contains a hydroxyl group at the 2' carbon atom of ribose. This moiety plays a key role in cellular processes involving RNA including the splicing of the pre-mRNA molecule into the mature mRNA which ultimately determines the final protein product (Griffith et al., 2010).

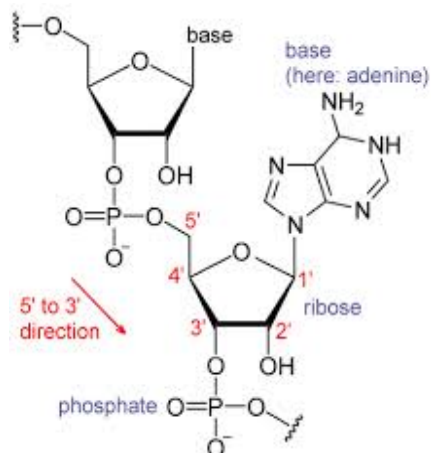


Figure 1. MOLECULAR STRUCTURE OF RNA Illustration of the RNA molecule indicating carbons involved in synthesis and the hydroxyl group at the 2' carbon important to the splicing mechanism. (Attribution: Narayanese at the English language Wikipedia. “RNA”; https://commons.wikimedia.org/wiki/File:RNA_chemical_structure.GIF)

In higher organisms the coding sequences that make up a gene are not continuous in the DNA molecule. Rather the genes are comprised of intermittent coding sequence interspersed with extended regions of DNA sequence that is non-coding. These sequences called introns are

transcribed but not translated. During transcription introns are removed from the primary transcript in a process referred to as splicing. The resulting mRNA contains a coding sequence that is completely parallel with the protein that it will ultimately encode (Griffiths et al, 2010).

While the human genome has only 20,000 genes, there are more than 100,000 proteins encoded. This suggests that any given gene may encode the information for more than one protein. One mechanism by which this is achieved is alternative splicing, a process by which exons are included or excluded in the final mature mRNA based on the splicing process. Proteins produced by alternative splicing, referred to as isoforms, are usually related because they share groups of exons present in the primary transcript (Griffiths et al., 2010).

The processing of a primary transcript into a mature mRNA transcript requires the precise removal of introns and the exact joining of exons. The exon-intron junctions are the sites at which splicing reactions take place and certain specific nucleotides are known to be highly conserved across individual genes and across species. Thus introns almost invariably have a GU at the 5' end and an AG at the 3' end. Also highly conserved is an adenine molecule that appears between 15 and 45 nucleotides upstream of the 3' splice site referred to as the branch point. The nucleotides flanking these sites are also highly conserved, but less rigorously (Griffiths et al., 2010).

The high degree of conservation at these sites reflects the fact that biological mechanisms must depend on these sites to facilitate the splicing process. A complex of proteins and one of 5 small nuclear RNAs (snRNA) is known as the small nuclear ribonucleoprotein (snRNPs). These 5 complexes associate in turn with more than 100 additional proteins to form the spliceosome. The snRNPs in the spliceosome complex are complementary to the consensus sequences at the intron-exon splice site junctions. The 5 snRNPs are referred to as U1, U2, U4, U5 and U6.

Hydrogen bonding between the snRNPs and the consensus sequences aligns the splice sites in order for accurate excision of the intron to occur. Thus the consensus sequence plays a central role in defining the phosphodiester bonds of the pre-mRNA that will be cleaved (Griffiths et al., 2010)

The process of nuclear mRNA splicing is intricate and still under investigation, however the essential biochemistry of the splicing process indicates the importance of the conserved sites. U1 attaches to the GU consensus sequence at the 5' end of the intron. U2 binds to the branch point and U4, U5 and U6 interact with the spliceosome to facilitate the configuration of the RNA in a position that will promote the splicing process. This process occurs in two steps. The first step involves the 2'-OH group of the adenosine at the branch point which initiates a nucleophilic attack on the phosphate 5' splice site of the intron creating a loop similar to a lariat. The second step is initiated when the 3'-OH of the 5' exon carries out nucleophilic attack on the phosphate of the 3' splice site. With this two-step process the intron is excised and the 3' and 5' exons are joined.

The average human gene contains eight exons and seven introns. Through the process of alternative splicing an average of three or more mRNA isoforms may be produced (Lee et al., 2015). Lee et al. also report that recent studies using high-throughput sequencing suggest that 100% of human genes produce at least two mRNA isoforms (Lee et al., 2015). These isoforms in turn encode distinct proteins. In the process of alternative splicing some sequences may be included as exons under given conditions resulting in one mRNA transcript, but under a different set of conditions this same sequence is spliced out, in which case it is treated as an intron, and does not appear in the mature mRNA (Edwards-Gilbert, 2010). Thus the same gene may encode two or more protein products with different or even opposing functions (Edwards-

Gilbert, 2010). Figure 2 shows possible types of alternative splicing. These include exon inclusion or skipping, mutually exclusive exons, alternative splice site selection, alternative promoters and poly (A) sites, and intron retentions. Any given pre-mRNA may show a combination of different types of alternative splicing (Black, 2003).

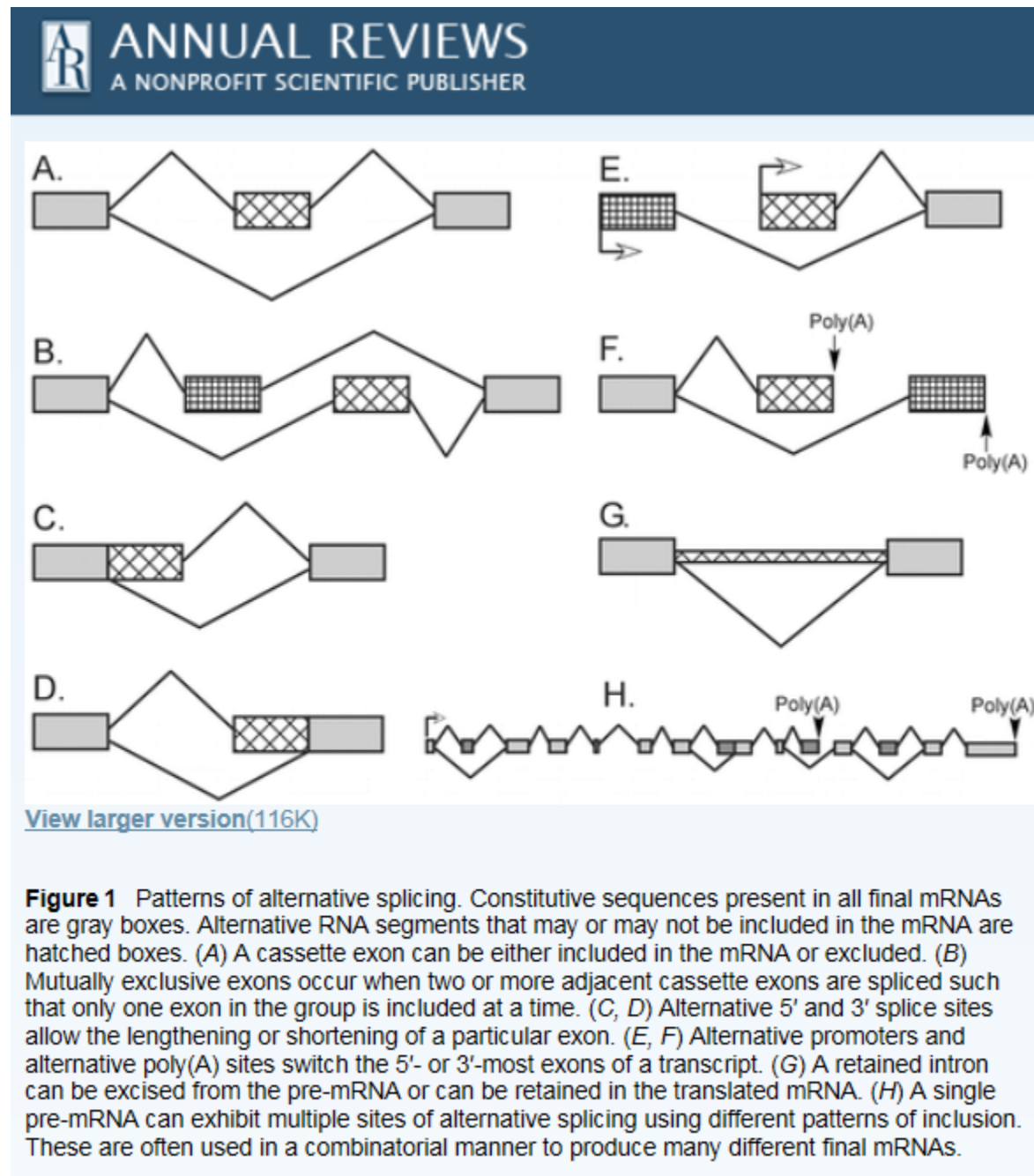


Figure 2. PATTERNS OF ALTERNATIVE SPLICING (Black, 2003)

It is estimated that one third of all deleterious mutations disrupt splicing (Vaz-Drago et al., 2015). This disruption may be at the splice site itself or through other alterations of the splicing mechanism such as splicing enhancers or inhibitory motifs. The phenotypic outcome is often a nonfunctional protein or the introduction of a premature termination codon which, in turn, targets the mRNA for degradation by nonsense mediated decay (Vaz-Drago et al., 2015). Given the role that splicing mechanisms play in protein function and dysfunction, it is an important consideration in establishing the clinical significance of a variant.

ROLE OF CUL4B: UBIQUITINATION

CUL4B plays a key role in an essential biological process known as ubiquitination. Ubiquitination is a chemical reaction whereby proteins are marked for degradation or transport to specific compartments within the cell. It is also involved in histone modification and by extension may play a role in gene expression. Ubiquitin (Ub) is a small protein consisting of 76 amino acid residues. It possesses seven lysine residues and an N-terminal methionine residue. Each of these can bind with the carboxyl terminus of another Ub molecule. This structure allows for unique signal diversity. Ub bonds covalently to the ε-amino group of lysine residue of the protein substrate. The protein may be modified with a single Ub molecule at one lysine site (monoUb), a single Ub molecule at multiple lysine sites (multi-monoUb) or a polymeric chain of Ub molecules attached at a single lysine site (polyUb) (Komander, 2009).

In eukaryotes the process of ubiquitination is one of the most predominant mechanisms for the regulation of protein function and stability. Strieter et al. suggest that sequence

annotations indicate as much as 5% of the human genome is involved in the addition and removal of ubiquitin to and from target proteins. Ubiquitination plays a pivotal role in the control of diverse biological processes. Misregulation has been associated with several human diseases, including cancer, immune disorders, neurodegenerative diseases, and congestive heart failure (Streiter et al., 2012).

The process of ubiquitination involves a progression of enzymatic events. Initially ubiquitin activating enzymes (E1s) activate free ubiquitin. The second step involves the transfer of the activated ubiquitin to ubiquitin conjugating enzymes, (E2s). These, in turn, coordinate with ubiquitin ligases (E3s), such as the CUL4B protein complex, to finally bind ubiquitin to targeted substrates (Streiter et al., 2012).

THE CUL4B PROTEIN PRODUCT AND PROTEIN COMPLEX

The CUL4B gene is located on the X chromosome at Xq23. The gene contains 22 exons and encodes a 913 amino acid protein that is a core component of cullin RING (Really Interesting New Gene) based E3 ubiquitin protein ligase complex (Zou et al., 2011). The SMART nrdb database estimates 60% conservation in the cullin homology domains (Sarikas et al., 2011).

The InterPro consortium which combines protein analysis from several databases suggests the following structural architecture for CUL4B (Figure 3). Six specific protein characteristics are summarized below:

1. Amino acid residues 209-555. The N-terminal cullin repeat-like-containing domain forms a multi helical 2-layered alpha/alpha structure, which is folded into a right-handed superhelix.

2. Amino acid residues 217-814. The N-terminal region encompasses several domains, including the cullin repeat domain, a bundle domain comprised of 4 helices, an alpha+beta domain, and a winged helix-like domain.
3. Amino acid residues 558-786. The cullin homology region includes three domains the 4 helix bundle domain, the alpha+beta domain, and winged helix-like domain.
4. Amino acid residues 746-813, 814-913. The winged-helix-turn-helix DNA-binding domain is found in many different proteins with widely varied biological functions. The winged helix is comprised of two wings, three alpha helices and three beta sheets. The wings are small beta-sheets which make contact with DNA at different points including the minor groove or the DNA backbone. The DNA recognition site is a helix that makes contacts at sequence specific sites with the major groove of DNA.
- 5, Amino acid residues 842-907. The neddylation site is common to the cullin proteins including Cul4B. Members of the cullin family are modified by Nedd8 which bonds covalently to a conserved cullin lysine residue. Cullins are the only known Nedd8 substrate.
6. Conserved region among the various cullin family proteins (InterPro, 2015), <http://www.ebi.ac.uk/interpro/protein/Q13620>)

The cullin family proteins, so named for their role in “culling, sorting or selecting” proteins in the cell for ubiquitination and subsequent degradation, contain an evolutionarily conserved cullin homology domain. There are seven cullins in mammals. Cullins 1 through 5 have in common a long amino-terminal domain composed of three cullin repeats (CR1-CR3). (Sarikas et al., 2011). The cullin domain is located between amino acid residues 217 and 815 (Tarpey et al, 2007). The carboxyl-terminal domain (CTD) has a globular structure and contains a highly conserved stretch of approximately 200 amino acids known as the cullin homology domain. The cullin protein is one of three components in the cullin protein complex. The other

two elements that comprise the complex are the RING finger protein which acts as a recruiter for the E2 enzyme carrying ubiquitin and a substrate adapter which brings substrates and the ubiquitin/E2 enzyme complex into close proximity.

The RING molecule, regulator of cullins (ROC1 or ROC2) also known as RING box protein (Rbx1 or Rbx2), binds to the cullin scaffold at the CTD. These RING molecules contain a domain rich in histidine and cysteine and coordinate two zinc atoms, the so called RING finger. They recruit the E2 enzymes carrying ubiquitin for eventual ligation to the substrate target (Sarikas et al., 2011). The transfer of ubiquitin from E2 to the substrate is achieved in part by an allosteric change of E2 promoted by the RING finger which results in the release of ubiquitin (Jackson et al., 2009; Kerzendorfer et al., 2011).

The cullin family proteins do not bind substrates directly. Instead substrate recruiting receptors containing a WD40 domain are linked to the CUL4 scaffold at the N-terminus through a linker protein which, in the case of the CUL 4 proteins, is DDB1 (damaged DNA binding 1). DDB1 is defined by three beta propellers (BPA, BPB, BPC). Higa et al. describe the relationship between BPA and BPC like two halves of an open clam shell with the opening representing a binding pocket. The third beta propeller, BPB, is connected to the BPA-BPC binding pocket by a hinge-like construction which provides relative flexibility. This allows for three different conformations of the BPA-BPC clamshell relative to the cullin-binding BPB (Higa et al. 2007).

The WD 40 domain which distinguishes the substrate recruiting receptors is defined by a Gly-His dipeptide and Trp-Asp (WD) separated by 20 to 30 residues and is approximately 40 residues in length. These substrate recruiting receptors may be referred to as DDB1-binding WD40 (DWD), DDB1-CUL4-associated factors (DCAF), or CUL4-DDB1-associated WDR (CDW) proteins (Jackson 2009). They link to DDB1 at the WD 40 domain, then associate with

the substrate specific to the recruiting protein which, in turn, will be ubiquitinated. To further characterize the WD40 domain found in DDB1-binding proteins, He et al. identified 16 residues referred to as the DWD box (DDB1-binding and WD40 repeat). The defining characteristics of the DWD box are the first 14 residues in the DWD box which correspond to the second half of a WD40 repeat that is more conserved when compared to other WD40 repeats. The Arg16 following the WD dipeptide is a key residue of the DWD box. Several of these conserved residues, among them Arg16, are positioned on the protein surface which further supports the role of these residues in binding with other proteins. He et al. note that a DWD protein may function either as a substrate itself or as a substrate receptor that aligns other proteins for ubiquitination by the CUL4 protein complex (He et al., 2006). This construction allows for the CUL4B complex to interact with a variety of specific targets. Figure 4 illustrates schematically the structure of the CUL4B complex.

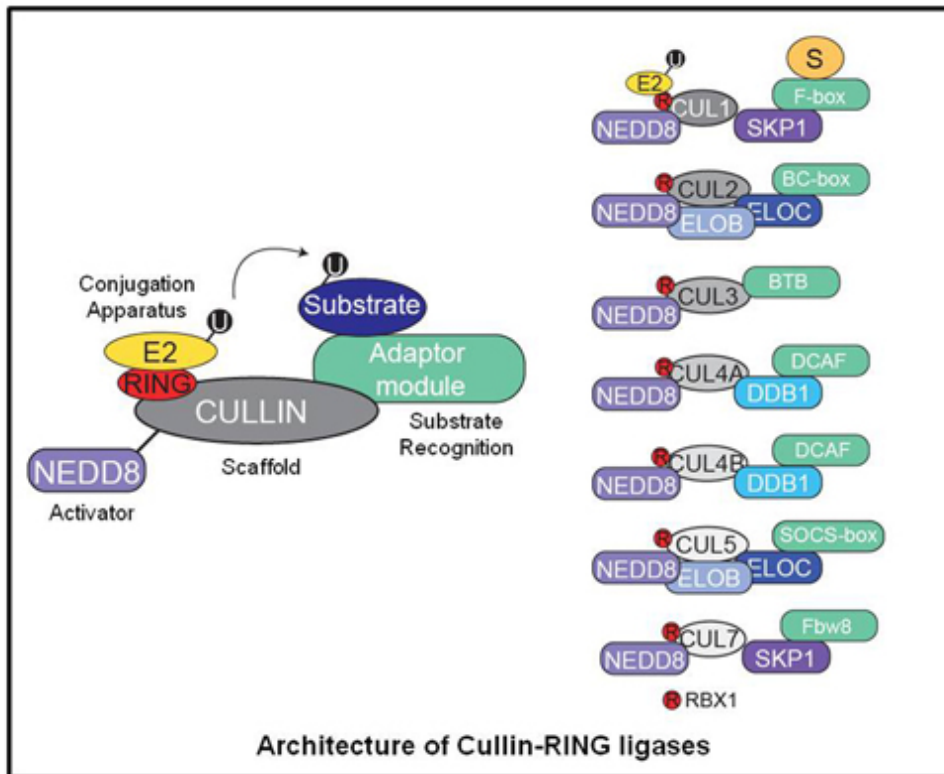


Figure 3. THE CULLIN FAMILY PROTEIN COMPLEXES. Schematic illustrating the general structure of the cullin protein complex and the seven homologs in the cullin family. Shown are the cullin scaffold proteins, the regulator/activator, the recruiter protein RING, the adaptor (linker) module including DDB1 and the regulator/activator NEDD8. The substrate recruiting receptors are also represented. (Reprinted with permission from <http://Biology.ucsd.edu/research/faculty/e1bennett>)

Examination of the CUL4B protein indicates that it is part of a protein complex and that its function is dependent not only on the integrity of the CUL4B structure, but also on that of other proteins it interacts with. In addition to the proteins described above NEDD8, CAND1, and COPS5 (subunit of COPS9) interact with CUL4B and are involved, in regulation of the ubiquitination activity. NEDD8 (Neural Precursor Cell Expressed, Developmentally Down-Regulated 8) is a ubiquitin-like protein which, like the RING protein, covalently modifies cullins (genecards.org). NEDD8 acts on the C-terminus between residues 842 to 907 at a conserved

lysine residue. This induces a conformational change that promotes the transfer of ubiquitin to the substrate (Liu et al., 2012).

In addition to promoting ubiquitination, NEDD8 also is involved in regulating CRLs. When NEDD8 is bound to the cullin scaffold a CAND1 (cullin-associated NEDD8-dissociated protein 1) binding site is obscured (Duda et al., 2008). This prevents CAND1 from associating with the CUL4B protein. CAND1 is thought to play a role in regulating the CUL4B protein complex by binding unneddylated CUL4B at the exposed binding site and rendering the complex inactive in a process referred to as sequestration. (Bennett et al., 2010).

COPS5 is a subunit of the COPS9 (constitutive photomorphogenic 9) signalosome complex (CSN) which is made up of 8 subunits (Wei et al., 1999). It contains the catalytic site responsible for deneddylation. By removing NEDD8, it acts to disrupt ubiquitination while also keeping the CUL4B protein complex inactivated. (Bennett et al, 2010). While the regulatory mechanism of the cullin family in general and CUL4B in particular is not well understood, it is clear that it involves an interaction among proteins. The emerging picture suggests that disruption of any one of the proteins may result in dysregulation and by extension disruption of ubiquitination. This, in turn, affects many biological processes.

The functional roles of CUL4B and its homolog, CUL4A, are not completely characterized. Although they are highly identical, the fact that no human disorder is associated with CUL4A while CUL4B is associated with Cabezas syndrome suggests that CUL4A cannot fully compensate for CUL4B. However, in spite of several mutations in CUL4B reported in the literature that lead to functionally null mutations, complete loss of function of the CUL4B protein product is compatible with life in spite of significant, pleiotropic effects. Ravn et al.

suggest that although CUL4A cannot completely compensate for CUL4B this may indicate a possible redundancy between the two proteins (Ravn et al., 2012).

CUL4B is thought to play significant roles in cellular processes including ubiquitination of histones and control of the cell cycle through down regulation of cyclin E and b-catenin. It is also known that genetic mutations of the CUL4B gene are associated with X-linked intellectual disability (XLID) in males. Liu et al. examined the endogenous Cul4B isoforms in the brain in order to better understand what role the CUL4B protein product might have in the neuronal cells.

Both human and rodent tissues express three principal Cul4B protein isoforms. In the brain Cul4B-1 and Cul4B-2 are unneddylated (thus likely inactive due to sequestration) and more abundant. Cul4B-3 which is missing the N-terminus present in the other two isoforms is neddylated and thus more likely to be active. Studies by Liu et al. suggest that the N-terminus of Cul4B inhibits neddylation in the larger isoforms. The group's work further demonstrated that most Cul4B is unneddylated. They noted that unneddylated Cul4B accumulates during mitosis of neural progenitor cells (NPC) and that down regulation of Cul4B arrests cells in the G2/M phase in NPCs. Thus a higher concentration of Cul4B promotes movement through cell division although it is unneddylated and likely sequestered. The authors found excessive amounts of unneddylated Cul4B in human brain hippocampal NPCs. Taking into account the fact that accumulated Cul4B may play a role in regulating cell division the authors suggest a role for Cul4B in brain development (Liu et al., 2012).. This example further illustrates the complex nature of Cul4B regulation and mechanism of action.

LITERATURE REVIEW OF PATIENTS WITH CUL4B MUTATIONS

In 2000, Cabezas et al., reported a large family in which intellectual disability was tracked from four normal and one minimally affected female to seven males over three generations. No male to male transmission was observed. The observed pattern suggested X-linked inheritance. Upon examination, consistent clinical features were observed including short stature, prominent lip, small testes, muscle wasting of the lower legs, kyphosis, joint hyper extensibility, abnormal gait, tremor, and decreased fine motor coordination. (Cabezas et al., 2000). This report established a new XLMR syndrome which is often referred to in the literature as Cabezas syndrome.

In 2007 Tarpey et al. observed that intellectual disability had been associated with more than 60 (approximately 6%) of the X-linked genes annotated in the Vega Genome Browser. The authors noted, however, that each reported X-linked mental retardation (XLMR) gene was associated with only a few affected families. The two exceptions to this observation were mutations in the FMR1 gene which causes fragile X syndrome and is associated with ~20% of families, and ARX gene mutations which are seen in ~9.5% of affected families. The authors indicated that despite success in identifying genes associated with intellectual disability, the underlying gene defect in most affected families could not be attributed to any of the known genes (Tarpey et al., 2007). In an attempt to further identify the genes responsible for XLMR, Tarpey et al. performed mutational screening of the X chromosome in 250 families with multiple members who presented with intellectual disability, including those previously reported by Cabezas et al. The use of a direct screening approach revealed truncating and missense mutations in CUL4B in 8 of the 250 families studied for a total of 39 individuals.

The pattern of clinical features observed in these individuals included intellectual disability that was typically moderate, but widely variable. While some individuals were institutionalized, others were able to live with some degree of independence. Speech delay, which was observed in early childhood, was disproportionately severe relative to the degree of intellectual disability. Adults were capable of little or no speech, but could understand basic commands. They were also able to cooperate with others to complete a task. Aggressive behavior was reported to be common among affected individuals and also varied widely, ranging from shouting and stamping to attempted strangulation (Tarpey et al., 2007).

Neurological features included tremor observed in the hands, some compromised motor activities, and seizures. Seizures were typically single febrile fits and, while common in young children, were not often observed in adults. While severe ataxia was observed in one child, patients were most often described as clumsy particularly in adulthood. Wasting of the calf muscles, retained tendon reflexes, and pes cavus were also observed in some patients (Tarpey et al, 2007).

Physical features commonly observed included short stature, central obesity, and genital tract abnormalities such as undescended and/or small testes, hypospadias, and hypogonadism. It was noted that the average adult height was in the 5th percentile while the average head circumference was in the 97th percentile. Feet were reported to be small in some cases and abnormalities of the toes and sandal gap were also observed (Tarpey et al., 2007).

Also in 2007, Zou et al. reported a family with XLMR presenting in five affected males in four sibships. The authors were able to use polymorphic markers to identify a candidate interval in the 10.26 Mb region between XSTR3 and XSTR4, eventually finding a nonsense mutation in the CUL4B gene. A base substitution at 1564 in exon 9, c.1564 C->T resulted in the

conversion of a codon for arginine (CGA) to a termination codon (TGA). The mutation cosegregated with affected individuals in the family (Zou et al., 2007). The phenotypes of affected individuals had overlapping characteristics both among those in the family and with the phenotypes described in the patients of the Tarpey group.

Following these two seminal papers, other groups also reported findings of CUL4B mutations in patients with XLMR. According to a 2014 paper by Vulto-van Silfhout et al., 13 families with 12 discrete CUL4B mutations had been reported in the literature. In addition to the families studied in 2007 by Tarpey et al. and Zou et al., Badura-Stronka et al. reported a novel nonsense mutation in three brothers with XLMR in 2010. Also in 2010 Isidore et al. reported a patient with a deletion in CUL4B with Cabezas syndrome, followed by Ravn et al. who reported a monozygotic twin pair with a deletion in CUL4B leading to a concordant phenotype with features overlapping those of other patients with CUL4B mutations. In 2014, Londin et al. identified a donor splice-site mutation expected to result in abnormal or absent splicing of exon 6 in CUL4B. As with previous reports the affected individuals were all male and had features typical of Cabezas syndrome. Writing in 2012, Ravn et al. observed that of the 11 families identified at that time all but one were clinically diagnosed after the genetic mutation was established, suggesting that XLMR associated with CUL4B mutations may be underdiagnosed.

In 2015 Vulto-van Silfhout et al. described 11 families with variants in CUL4B, adding 25 new patients for a total of 85 patients known to have a CUL4B mutation and Cabezas syndrome. Nine of the variants identified were novel. Based on these patients and clinical data previously published, the Vulto-van Silfhout group proposed a guideline for use in identifying patients who might have a CUL4B mutation. These guidelines break down the observed

phenotypic characteristics into seven groups: growth, central nervous system, neurological manifestations, craniofacial, hypogonadism/genital abnormalities, and extremities. (Vulto-van Silfhout et al., 2015)

The CUL4B variants identified in the Vulto-van Silfhout study were distributed throughout the gene. They included one missense mutation, an in-frame deletion of three base pairs, an in-frame duplication of three base pairs, two splice site variants, and five truncating variants. One variant, a five base pair deletion had been previously reported by Tarpey et al. in an unrelated family (Vulto-van Silfhout et al., 2015). Taken together these findings indicate that Cabezas syndrome may constitute a definite syndromic form of XLMR for which mutations in the CUL4B gene are responsible.

METHODS

Parents of the patient were consented under IRB protocol HS2002-2608: Mitochondrial Inborn Errors of Metabolism and ANT Defects in Mitochondrial Disease. The Parents also provided consent for the participation of their child, the patient. Principal investigator is Virginia Kimonis. Co-investigators are Moyra Smith and Pam Flodman.

Patient Profile

The patient at the time of testing was a seven-year, one-month-old male of Philipino ancestry. At birth he weighed 5lbs. 10 ounces and had an Apgar, score as reported by the mother, of 9. The patient has a history of developmental delay including intellectual disability and speech delay. He began having seizures at age 2. Head size is increased in proportion to his

growth in height and weight. At seven years of age, his height and weight were near the 50th percentile, while head circumference was in the 98th percentile. Neurological examination at the age of 4 years noted unsteady gait, hypotonia, delayed gross and fine motor skills, and delayed social development. At seven years of age speech was absent.

Metabolic testing performed in the Philippines and in Chicago included organic acid analysis, amino acid analysis, lactic acid measurement, and acyl carnitine analysis. All results were normal.

Genetic testing including cytogenetic chromosome studies, microarray, gene studies on MTHFR and SCN1A, fragile X and methylation studies were all reported normal. PTEN studies performed subsequently by Ambry genetics using PCR were also normal. MRI studies were performed and revealed no structural abnormalities or ventricular enlargement, although peria-atrial periventricular bright signals were present.

The family history was non-contributory. The patient has a sister who is two years older. Her growth and development are reportedly normal.

Exome Sequencing

The parents of the patient were consented and blood samples from the patient and both parents were shipped from the Philippines. First Tier exome sequencing was carried out through Ambry Genetics in Orange County, California. First Tier sequencing includes approximately 4000 characterized genes in which disease causing mutations are most commonly found. The general protocol as described in the clinical testing report is as follows:

“Genomic deoxyribonucleic acid (gDNA) is isolated from the patient’s whole blood. Samples are prepared using the SeqCap EZ VCRome 2.0 (Roche NimbleGen) or the IDT xGen Exome Research Panel V1.0. Each DNA sample is sheared, adaptor ligated, PCR-amplified and incubated with the exome baits. Captured DNA is eluted and PCR amplified. Final quantified libraries are seeded onto an Illumina flow cell and sequenced using paired-end, 100 cycle chemistry on the Illumina HiSeq 2500. Initial data processing, base calling, alignments and variant calls are generated by various bioinformatics tools.”

mRNA Studies

Studies of mRNA were performed at University of California MitoMed CLIA laboratory. Blood samples were obtained from the patient and his mother and were shipped from the Philippines in PAX tubes. Shipping time was 2 days. A control blood sample was obtained at University of California, Irvine. RNA was obtained from the samples using Qiagen PAX RNA kits. RNA was then processed to derive cDNA. Polymerase chain reaction experiments were performed to amplify the derived cDNA sequence using forward and reverse primers as shown in figure 5.

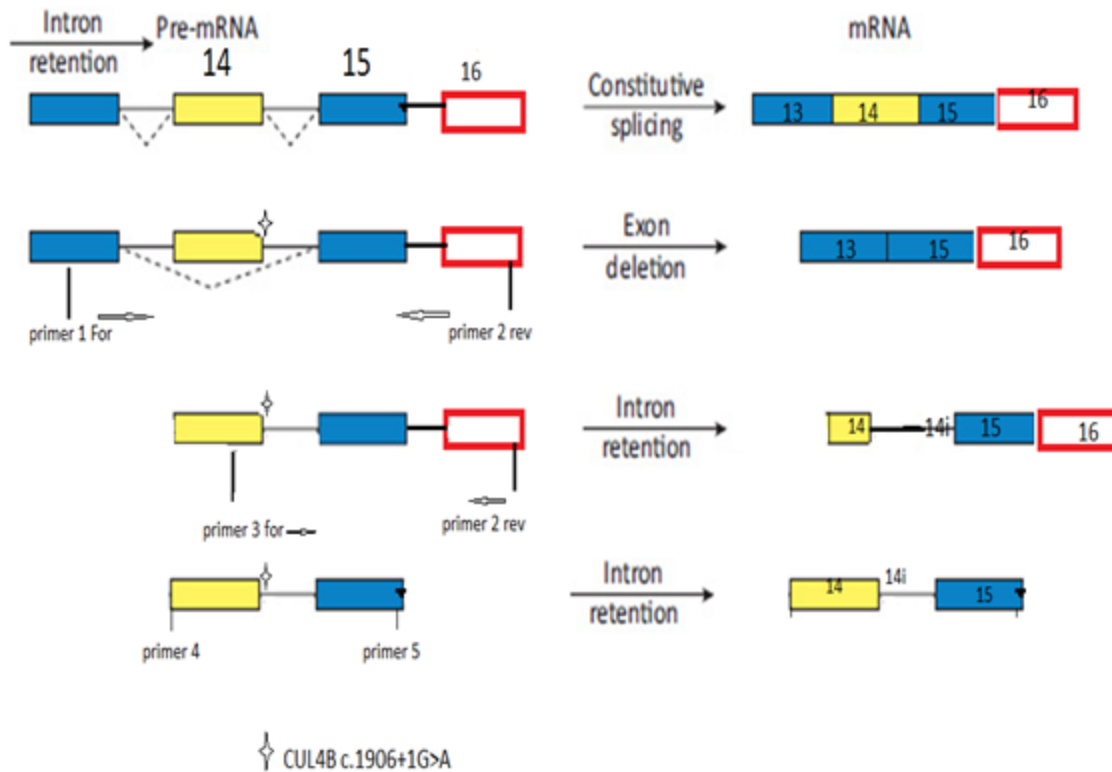


Figure 4. ALTERNATIVE SPLICING AND LOCATION OF PCR PRIMERS. Schematic showing the constitutive splicing and alternative splicing that may result from the splice site mutation at position c.1906. Positions of primers designed to determine the presence of alternate transcripts are also shown. (Moyna Smith, personal communication)

Three PCR experiments were performed. In experiment 1 (primers 1 and 2) the forward primer was located within exon 13 and the reverse primer was located at the end of exon 16. The resulting PCR product would be predicted to contain exons 13, 15 and 16. Consistent with the aberrant donor splice site at the 3' end of exon 14, this product would not include exon 14. In experiment 2 (primers 3 and 2) the forward primer was located within exon 14 and the reverse primer was positioned at the 3' end of exon 16. In this case, a subset of transcripts would be predicted that include the intron between exons 14 and 15. To further test for intron inclusion, in the final experiment (primers 4 and 5) a forward primer was located at the 5' end of exon 14 and a reverse primer was located at the 3' end of exon 15. The resulting PCR products were

Sanger sequenced and the DNA sequence derived was analyzed using BLAST N and BLAST X programs from NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch) (Moyra Smith, personal communication).

RESULTS

Exome Sequencing

First Tier sequencing revealed several variants in proteins associated with the CUL4B protein complex. The most relevant variant detected was the nucleotide change c.1906+1 G>A in the CUL4B gene. The variant is a hemizygous splice site mutation at genomic position X:119672514. Assembly GRCh37.p13(GCF_000001405.25)X: NC_000023.10 (119658444..119709684, complement).

Once the CUL4B mutation was identified, selected genes whose protein products are associated with CUL4B in the ubiquitination complex or are known substrates were reviewed for possible deleterious variants. These included DNMT3A, WDR48, COPS6, COPS5, DDB1, NEDD8, RBX1 (ROC1), and SUV39H. Of these, four had variants identified including DNMT3A, COPS6, SUV39H, and NEDD8. All of the variants identified were in intronic sequences with the exception of one variant in COPS6, located at position c. 429T>C resulting in a synonymous mutation in the protein product, p.C143C. The gene which encodes the CUL4B homolog, CUL4A, was also analyzed. In this case, fourteen variants in intronic sequences were identified.

Generally, fragments in the genome were well covered in the areas of interest. The majority of SNP variants identified had at least 10x coverage. However, two fragments, one from

CUL4A and one from SUV39H1 had no coverage. A second fragment for SUV39H1 had an average coverage of 2.21. Q scores, which indicate the likelihood that a given base may have been called incorrectly, ranged from 27 to 2531. Higher scores indicate better quality as determined by such criteria as signal strength, ratio of signal strength between the most common variant and the second most common variant for a given base, and observed background noise (Brockman et al., 2008)

Of the variants examined for CUL4B associated proteins, and CUL4A, one variant had an allele frequency when compared to the 1000 Genomes database of .08%, and eight variants were not reported in the 1000 Genome database. The remaining 20 variants reported had an MAF greater than 5% which is above the cutoff for a low frequency allele. Eight variants had no reference number in the dbSNP database.

mRNA studies

Given the location of the mutation at the splice site following the 3' end of exon 14 there are two possible outcomes. One possibility is that splicing will occur at the 5' end of exon 15 resulting in complete removal of exon 14. The second possibility is that no splicing will occur between exon 14 and exon 15 resulting in the inclusion of the intron in the final transcript. MRNA studies indicate that in this patient both mechanisms were occurring. Notably no functional CUL4B protein product would be predicted given the cDNA mutation identified. Figure _ shows the two smaller cDNA bands of the patient (columns 2 and 3) compared to the mother and a control indicating the absence of exon 14. In columns 5 and 6, the size of the two larger bands compared to that of the mother and a control indicate that intron 14 has been included in the patient's DNA.

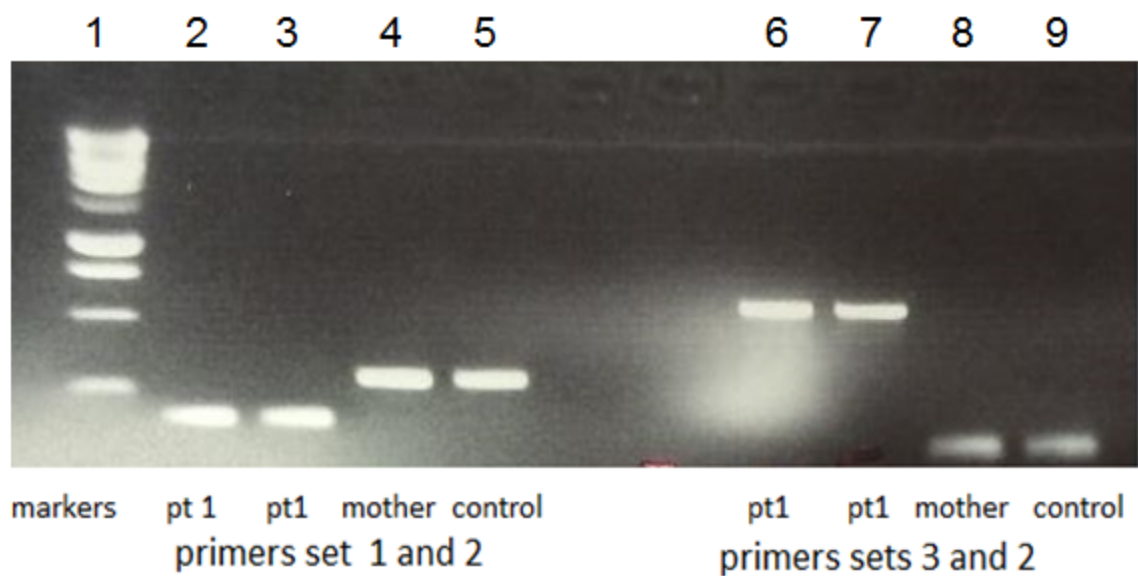


Figure 5. PCR PRODUCTS RESULTING FROM PRIMER SETs USED IN EXPERIMENTS ONE AND TWO. Photograph of gel showing PCR product from primer set 1 and 2 indicating the exclusion of exon 14 and the PCR product from primer set 3 and 2 indicating the inclusion of intron 14. (Moyra Smith, personal correspondence; mRNA analysis performed by Nhi Huong T Vu)

Blast analysis was performed on the PCR product of the third experiment (primers 4 and 5). Figure 7 below, shows the results of that analysis. It indicates that only the terminal sequences of the PCR product align with known exon sequences. These observations further support the results of experiment two which indicate that intron 14 has been included due to the abolished donor splice site at the end of exon 14.

15R-sequ-LB-546nt

```
AGTCGAAAGAGTGCATCTGTAGATGCTGAAAAATCAATGCTGTCCAACTTAAACATGATAGGTTTCTCATTCT
TTCTGTTCTTTTCTTTGATTCTGAAGGAGAATGTTAGCTTTTACCCTTTCAATCGATTTCTCTTTTGTCACTCA
TGCTATGTATATATGATGAAAACATAAATTGAGCAAGCTCCTGAGGTTTCCAACCTATTGTACATAATATCCAG
AGTTGTTTTTTTTAAGATAAAGCATTGTTCTTGGTTTTATTATTTGTATAAAAAATTTCCCATTTATTTTATACCTAAA
TAGTAGATAAGGTTACTACAGGATTAAGAGTTTGACTCTGAATTAATGGGGCATAACACTTTGGGGTAAGGATT
TCTTTTATTCTTATGAAAGTTACAGATCATACTTCATACTAAAGTTTTGTATTTTCGTGTTTCCCACTTTTAAACCTC
TTTTGTTTTTACCACCTTAAAATATTATACATGTACAAACATCTTTAATAGAATGCGGAGCTGCTTTCACCAGCAA
ACTTGAAG
```

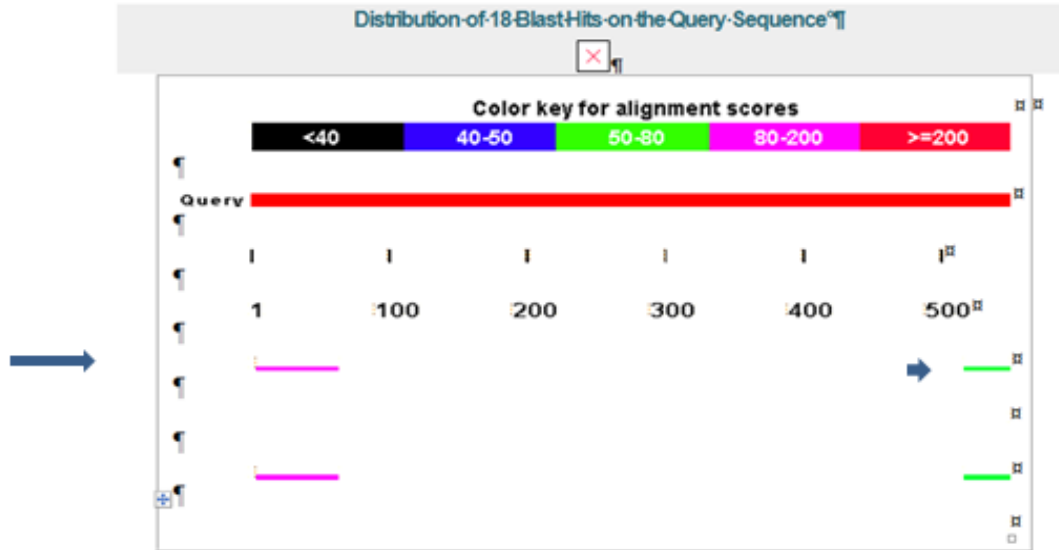


Figure 6. BLAST ANALYSIS OF PCR PRODUCT OBTAINED IN EXPERIMENT THREE. BLAST analysis of PCR product using primers at start of exon 14 and end of exon 15 (primers 4 and 5). Only terminal regions of the sequence indicated in pink and green match exon sequences. (Moyra Smith, personal correspondence)

DISCUSSION

CUL4B has emerged as a probable cause for XLMR. The aim of this work was to explore several lines of evidence for determining whether a specific mutation in the CUL4B gene may be the cause of intellectual disability in a patient with features of Cabezas syndrome. We also attempted to determine if variants associated with other proteins in the CUL4B protein complex might contribute to the condition or its severity in this patient.

Identification of the mutation was achieved using First Tier Exome Sequencing performed as a clinical test through Ambry Genetics. Previous clinical studies as reported in methods did not reveal any possible diagnosis for this patient. The variant identified occurs in the CUL4B gene, NM_003588 c.1906+1G>A. It is highly conserved among species indicating that it is likely to have an important biological function. The guanine nucleotide that occurs at this position is replaced by an adenine in the patient presented in this work. The mutation is expected to abolish the donor splice site and is predicted to result in the loss of exon 14 or the inclusion of intron 14. Exon 14 is 111 nucleotides long, representing 37 codons. Since this number is divisible by three the loss of these nucleotides would not result in a translational frameshift, however 37 amino acids would be missing from the final protein product. (Moyra Smith, personal communication)

An important consideration in assessing this mutation in the patient is the region of the protein in which it occurs. Given that many of the active sites and conserved domains in CUL4B are at the CTD and NTD, would the loss of 37 amino acids at this position be expected to have a deleterious effect? An analysis of the structural biology suggests that the stability of the C-terminal domain would be compromised, thus disrupting the CUL4B-RBX 1 complex. The amino acids encoded by exon 14 fold into two helices within the CTD/RBX1 strand anchoring region. This structure suggests a possible functional role in stabilizing the CTD. One of these helices (helix 1) is positioned completely within the globular C-terminal region and serves to connect the CTD with the first cullin-repeat domain in the central region of the protein. A second interaction occurs directly between the hydrophobic residues, CUL4B lysine L616 and RBX1 valine V30. In addition there is an indirect interaction between the helix dipole of CUL4B helix 1 and the positively charged histidine H729 of CUL4B as well as the polar

asparagine N28 of the RBX1 strand. Taken together these chemical interactions indicate a possible role for the 37 amino acids encoded by exon 14 in stabilizing the protein complex.

(Robert Huether, Moyra Smith, unpublished data/personal communication)

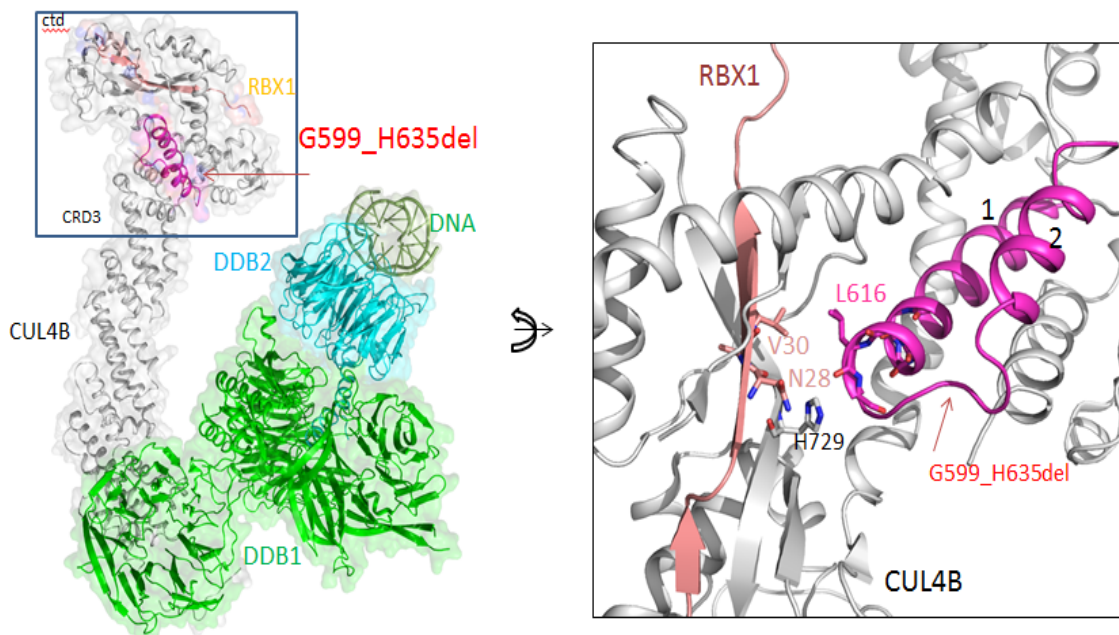


Figure 7. INTERACTIONS IMPACTING STABILITY OF CUL4B C-TERMINAL DOMAIN (CTD). CUL4B (white) and RBX1 (peach) form a complex which interacts with several adaptor molecules. In the left figure the adapter molecule DDB1 (green) and the substrate receptor DDB2 (teal) interacting with DNA (dark green) are shown. Amino acids encoded by exon 14 fold into two helices (magenta) located in the C-terminal domain. The structure shown in the right figure shows helix 1 positioned completely within the protein structure and acting to connect the CTD with the first cullin-repeat domain. Helix 1 has direct hydrophobic interactions with CUL4B lysine residue L616 and RBX1 valine V30 as well as indirect interactions involving the dipole of the helix with CUL4B histidine H729 and RBX1 asparagine N28 (see text). (Robert Huether and Moyra Smith unpublished data/personal communication)

In addition to destabilization of the CUL4B/RBX1 complex, deletion of exon 14 and the corresponding loss of the 37 amino acids it encodes may also affect interactions between CUL4B and the regulatory protein, cullin-associated NEDD8-dissociated protein 1 (CAND1). The structural biology indicates that one of the helices makes direct contact with residues of CAND1 via polar interactions. Thus the loss of exon 14 alters the interactions with CAND1. Given the

role of CAND1 in regulating the CUL4B protein complex, the disruption of this interaction can be expected to compromise the overall function of the CUL4B complex. (Robert Huether, Moyra Smith, unpublished data/personal correspondence)

The alternate possibility to deletion of exon 14 is inclusion of intron 14. In this case, the loss of the splice site at the 3' end of exon 14 would result in the disruption of the splicing mechanism such that the intron sequence between exons 14 and 15 was transcribed and therefore present in the mRNA transcript. Analysis indicates that this intron introduces translational stop codons and results in a truncated protein expected to have compromised or absent function (Moyra Smith, personal correspondence)

A review of the exome sequence data reveals that of the genes examined whose protein products are associated with CUL4B four, DNMT3A, COPS6, SUV39H, and NEDD8, had variants reported. Table 1 summarizes these mutations. All of the variants except one in COPS6 were in intronic regions. At first glance, it might be assumed that these variants would not be contributory since the final protein product would be unaffected. Although this is the likely scenario, there are several other factors worth considering. Several of the variants appeared quite deeply into the intronic region, the deepest being in the DNMT3A gene at c. 640-1663G>A. Current understanding of the intronic regions and how they contribute to gene expression is incomplete. Future studies may show that mutations in introns do affect the protein or its regulation and expression, and by extension, affect the function of the protein complex.

Gene	c. Variant	p. Variant	Zygoty	Coverage (Variant/Total reads)	1000Genome Allele Counts (Observed/Total)
DNMT3A	c.5-191G>A	NULL	HOMO	18/18	1420/2188
DNMT3A	c.5-74C>T	NULL	HET	64/126	N/A
DNMT3A	c.639+125_639+126ins	NULL	HET	14/33	N/A
DNMT3A	c.640-1663G>A	NULL	HOMO	18/18	1420/2188
DNMT3A	c.640-1546C>T	NULL	HET	64/126	N/A
DNMT3A	c.1606+26C>T	NULL	HET	93/180	843/2188
DNMT3A	c.2173+26C>T	NULL	HET	93/180	843/2188
COPS6	c.77-40G>T	NULL	HOMO	59/51	593/2188
COPS6	c.429T>C	p.C143C	HET	93/203	19/2188
SUV39H1	c.165+68T>A	NULL	HOMO	23/23	N/A
NEDD8-MDP1	c.149+856_149+858del	NULL	HET	53/204	N/A
NEDD8	c.150-54_150-52del	NULL	HET	53/204	N/A

KEY: Null=no variant; N/A=not available; HET=heterozygous; HOMO=homozygous;

Table 1: SUMMARY OF VARIANTS IDENTIFIED IN SELECTED GENES ASSOCIATED WITH THE CUL4B PROTEIN COMPLEX

Whole exome sequencing also identified variants in the CUL4B homolog, CUL4A, but again these were intronic. Although one region CUL4A (NM.001008895, coding DNA sequence 1 (CDS1) had no coverage and it cannot be presumed that no exonic mutation exists, in general this data indicates that the function of the CUL4A protein is not likely to be affected. As with other reports in the literature (Ravn et al., 2012), the high likelihood that this patient has functional CUL4A protein, further supports the position that CUL4A cannot replace CUL4B in spite of the high homology between the two protein products. Interestingly CUL4A differs from CUL4B primarily at the NTD. This, taken together with the structure of the CUL4B isoforms

which also differ primarily at the NTD, suggests that this domain may play a distinct role in the biological processes with which CUL4B is associated.

Two other variants, c.1339-5A>G and c.1639-5A>G, involving the CUL4B homolog, CUL4A, are close to splice sites and could be considered as potentially having some effect on the gene expression or splice site mechanism. However, allele frequency when compared to the 1000Genome database is approximately 30% indicating that these two variants are well-tolerated mutations in the population and are not likely to contribute to the patient's diagnosis. Table 2 summarizes the variants identified for CUL4A.

Gene	c. Variant	p. Variant	Zygoty	Coverage (Variant/ Total reads)	1000Genome Allele Counts (Observed/Total)
CUL4A	c.549-95T>C	NULL	HET	14/27	148/2188
CUL4A	c.549-62G>A	NULL	HOMO	47/47	1528/2188
CUL4A	c.617-122G>A	NULL	HET	41/90	688/2188
CUL4A	c.735+22C>T	NULL	HET	38/81	153/2188
CUL4A	c.849-95T>C	NULL	HET	14/27	148/2188
CUL4A	c.849-62G>A	NULL	HOMO	47/47	1528/2188
CUL4A	c.917-122G>A	NULL	HET	41/90	688/2188
CUL4A	c.1035+22C>T	NULL	HET	38/81	153/2188
CUL4A	c.1231-73G>A	NULL	HOMO	41/41	1619/2188
CUL4A	c.1339-206T>G	NULL	HOMO	10/10	2034/2188
CUL4A	c.1339-5A>G	NULL	HET	52/99	662/2188
CUL4A	c.1531-73G>A	NULL	HOMO	41/41	1619/2188
CUL4A	c.1639-206T>G	NULL	HOMO	10/10	2034/2188
CUL4A	c.1639-5A>G	NULL	HET	52/99	662/2188

KEY: Null=no variant; N/A=not available; HET=heterozygous; HOMO=homozygous;

Table 2. SUMMARY OF VARIANTS IDENTIFIED IN CUL4A (HOMOLOG OF CUL4B)

A second consideration is the depth of coverage for some regions. In addition to the CUL4A fragment previously discussed, one other fragment, SUV39H1 (NM_003173, (CDS1), had no coverage. Thus a variant may be present in this region. For a third fragment, SUV39H1 (NM_001281266, CDS1) the mean coverage was 2.21x. 100% of the base pairs in the region were covered at least one time; however none of the base pairs were covered 5 times. Calls at this low level may not be reliable and a variant might exist in the exon. SUV39H1 is a methyltransferase and its regulation may affect many processes. A missed deleterious variant may be important to changes in the overall effectiveness of the CUL4B protein complex.

Of the variants reported, only one had a minor allele frequency (MAF) of less than 1% based on comparison to the 1000 Genome database. This indicates a rare variant, which suggests a potential candidate for a pathological mutation. The variant, c.429T>C is a synonymous mutation in the COPS6 gene (p.C143C), and the only change in the data set reported in an exon. Although the mutation is rare, there are no changes in the resulting protein that would be expected to have a biological or biochemical effect and thus it is unlikely to contribute to the phenotypic manifestations in the patient.

An important consideration in understanding the role of CUL4B is the level of expression in specific tissues. One avenue of analysis is mRNA expression studies. These studies focus on understanding patterns of expression in specific tissues. Figure 9 shows the results of gene expression studies for CUL4B. Interestingly, the tissues where CUL4B is most highly expressed include those that are associated with the phenotypes described in the literature for individuals with known CUL4B mutations, notably those associated with the testes and dendritic cells. This further supports a causative role for CUL4 B in the manifestation of features associated with Cabezas syndrome.

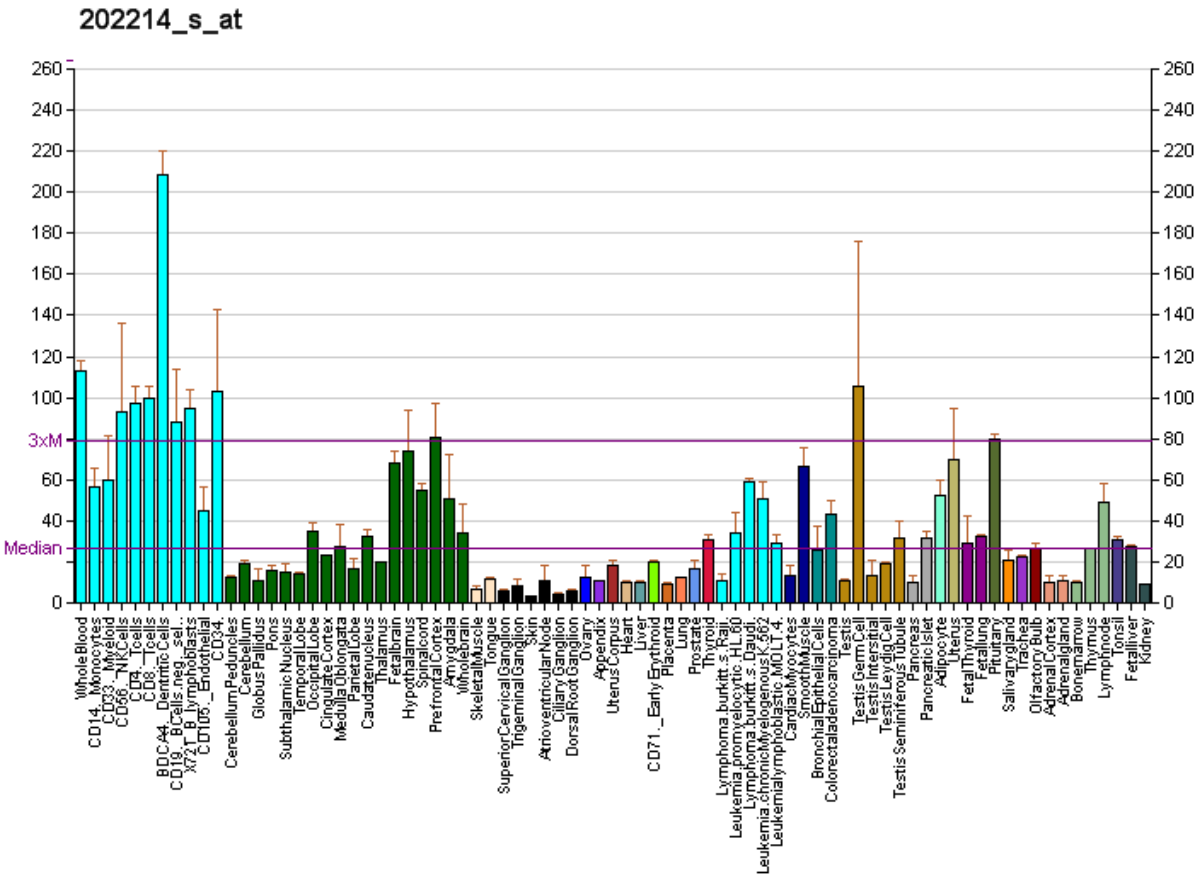


Figure 8. CUL4B GENE EXPRESSION STUDIES. Graph indicating the expression of CUL4B mRNA in various tissue types. Higher concentrations are noted in cell types associated with the testes and dendrites (Diagram created by AndrewGNF based on data from Su et al., 2004). ([https://commons.wikimedia.org/wiki/File:PBB GE_CUL4B_202214_s_at_fs.png](https://commons.wikimedia.org/wiki/File:PBB_GE_CUL4B_202214_s_at_fs.png))

A review of other patients reported in the literature shows overlap among them and with our patient. Comparing our patient with others who have a CUL4B mutation in general we find significant overlap in the phenotypic manifestations. Of particular interest are the two brothers (family 7) reported by Vulto-van Sulfhout et al. who have a splice site mutation in the same location as our patient. While these patients have a splice site mutation c. 1906+1G>T, our patient has splice site mutation c.1906+1G>A. As was the case for our patient, the c.1906+1G>T mutation was reported to abolish the donor splice site of exon 14 (reported as exon 15 in this paper) resulting in a loss of this exon and the deletion of 37 amino acids.

Importantly, the numbering of the exons reported in this paper differs from that of the case reported in the current study. However, biologically the exons and splice site mutation are the same. For the two patients reported in the Vulto-van Silfhout paper only exon skipping was reported. The alternate mechanism involving retention of intron 14 (or intron 15 using the numbering in that paper) was not reported. (Vulto-van Silfhout et al., 2014)

Table 3 summarizes the traits of our patient and the brothers from family 7 using the guidelines put forth by Vulto-van Silfhout et al. MRI studies for the brothers in family 7 were not performed and central nervous system (CNS) abnormalities were not reported. Common phenotypic manifestations between the brothers with the c. 1906+1 G>T mutation and our patient include severe intellectual disability, motor delay, and speech delay/absence.

Importantly, motor and speech delay are noted in all 85 of the patients reported in the literature at the time of the paper by Vulto-van Silfhout et al. Some behavioral problems were noted for both our patient and one of the patients with the G>T mutation. All three patients had prominent lower lips. Seizures and gait abnormalities were present in our patient, but not in either of the two brothers. Macrocephaly was also noted in our patient, but not in the other two patients. No genital abnormalities were reported for our patient, but one of the brothers from family 7 was noted to have this feature.

Generally, there is phenotypic heterogeneity among all patients with known CUL4B mutations. In the case of the three patients with mutations at c.1906+1, it is important to note that our patient at the time this comparison was made was 8 years and 5 months, while the two brothers previously reported were 39 and 41. Some traits such as behavioral problems and perhaps genital abnormalities may become more apparent over time. Complete records were not available for all three patients, but generally all three patients showed overlap with one another

and/or with other patients reported in the literature. It is not well understood at this time if there is a correlation between the domain where a mutation occurs and the extent or severity of manifestations and further study is needed in this area. However, the similarities between the patients with splice site mutations at CUL4B c.1906+1 further support this mutation as causal.

Feature	Patient with variant c.1906+1 G>A (n=1)	Patients with variant c.1906+1 G>T (n=2)
Growth		
Birth weight <p10	0/1	NR
Height <p10	0/1	1/1
Weight >p90	0/1	0/1
HC >p90	1/1	0/1
Central Nervous System (CNS)		
MRI performed	1/1	0/2
CNS abnormality	1/1	NR
Neurological		
ID level	1/1 severe	2/2 severe
Motor delay	1/1	2/2
Speech delay	1/1	2/2
Behavioral problems	+-/1	1/1
Tremor	0/1	0/1
Seizures	1/1	0/1
Gait abnormality	1/1	0/1
Craniofacial		
High prominent forehead	NR	NR
Malformed/abnormally positioned ears	NR	NR
HSR/deep-set eyes/narrow palpebral fissures	0/1 +-/1	0/2
Low nasal bridge/rounded tip	0/1	0/2
Prominent lower lip	1/1	2/2
Prognathia	0/1	NR
Hypogonadism/ genital abnormalities	NR	1/1
Extremities		
Brachydactyly/small hands/small feet	1/1 (feet)	0/1
Wasted lower leg muscles	1/1	1/1
Pes cavus	0/1	0/1

CNS, central nervous system; HC, head circumference; HSR, hyperplastic supraorbital ridges; ID, intellectual disability; MRI magnetic resonance imaging; +-, borderline; NR not reported.

Table 3. COMPARISON OF PATIENTS WITH SPLICE SITE MUTATION IN c.1906+1. Phenotypes of patients with known splice site mutations in CUL4B at c.1906+1 (After Vulto-van Silfhout et al. 2014)

Familial cosegregation has been observed in families reported previously in the literature. Mutations in CUL4B cosegregate with affected males in an X-linked inheritance pattern. In the case of our patient, the mutation was not observed in the mother or father of the patient which

indicates that the mutation in the patient is likely *de novo*. The mother's negative test result is particularly relevant given that Cabezas syndrome is an X linked condition. However, the possibility of gonadal mosaicism cannot be ruled out and if this is the case a recurrence of the condition in a future pregnancy may be possible.

CONCLUSION

The lines of evidence examined in the case of this patient include identification of a candidate gene using First Tier Exome Sequencing, annotation, comparison to other patients reported in the literature, and functional studies. Taken together the results suggest a strong role for CUL4B in the etiology of this patient's phenotypic manifestations. While the variants observed in other proteins involved in the CUL4B ubiquitination complex did not seem to be contributory, our understanding of the functions of intronic regions is incomplete. Further study is needed to better understand the role of introns in all aspects of gene expression and regulation. Additionally, the role of CUL4B in the development of the brain and nervous system, as well as its role in other protein networks, remain poorly understood and are an area of ongoing study.

Because the CUL4B mutation observed in the patient did not appear in the mother's blood DNA, it is quite likely that the mutation is *de novo* in the patient and recurrence is unlikely. However, as indicated previously, there remains a possibility of gonadal mosaicism. The mother of the proband was advised that should this be the case, the mutation may be present in future offspring and risk of recurrence cannot reliably be predicted.

Writing in 2012, Ravn et al. observed that for the 11 families identified at that time, a definitive clinical diagnosis was put forth following confirmation of the genetic mutation. A fact

which they suggest indicates that CUL4B associated XLMR may be underdiagnosed. (Ravn et al., 2012). Writing in 2014, Vulto-van Silfhout et al. report 85 individuals with Cabezas or Cabezas-like features who have confirmed mutations in CUL4B. Several labs have begun to offer testing for mutations in the CUL4B gene. In spite of increasing awareness that CUL4B may play a role in XLMR, the patient in this study underwent many years of testing and evaluation before receiving a possible diagnosis. And this diagnosis would come from half way across the world with much arduous research and effort on the part of his family. The analysis put forth in this study contributes to the growing body of evidence in the literature supporting the role of CUL4B in the etiology of intellectual disability. While the likelihood that treatment can be offered to patients with intellectual disability at this time remains small, genomics and the emerging technologies for gene analysis may potentially be able to offer families a reason for the condition. It is also increasing the understanding of how these conditions may recur in a family and, in some cases how to best manage the condition. For many this may, in turn, serve to relieve the very real suffering that is the diagnostic odyssey.

REFERENCES

- "The 1000 Genomes Project Consortium. " A Map of Human Genome Variation from Population Scale Sequencing." *Nature* 467.731 (2010): 1061-073. Print. PMC.Web. 8 Nov. 2015
- Altschul, S., W. Gish, E. W. Myers, and D. J. Lipman. "Basic Local Alignment Search Tool." *Journal of Molecular Biology* 215.3 (1990): 403-10. Web.
- Arthur, Jonathan W., Florence S.g. Cheung, and Juergen K.v. Reichardt. "Single Nucleotide Differences (SNDs) Continue to Contaminate the DbSNP Database With Consequences for Human Genomics and Health." *Human Mutation* 36.2 (2015): 196-99. Web.
- Badura-Stronka, M., A. Jamsheer, A. Materna-Kiryluk, A. Sowiã,,Ska, K. Kiryluk, B. Budny, and A. Latos-Bieleã,,Ska. "A Novel Nonsense Mutation in CUL4B Gene in Three Brothers with X-linked Mental Retardation Syndrome." *Clinical Genetics* 77.2 (2010): 141-44. Web.
- Bao, R., L. Huang, J. Andrade, W. Tan, W. Kibbe, H. Jiang, and G. Feng. "Review of Current Methods, Applications, and Data Management for the Bioinformatics Analysis of Whole Exome Sequencing." *CIN Cancer Informatics* (2014): 67. Web.
- Barash, Yoseph, John A. Calarco, Weijun Gao, Qun Pan, Xinchun Wang, Ofer Shai, Benjamin J. Blencowe, and Brendan J. Frey. "Deciphering the Splicing Code." *Nature* 465.7294 (2010): 53-59. Web.
- Bennett, Eric J., John Rush, Steven P. Gygi, and J. Wade Harper. "Dynamics of Cullin-RING Ubiquitin Ligase Network Revealed by Systematic Quantitative Proteomics." *Cell* 143.6 (2010): 951-65. Web.
- "Bioinformatics Primer." *An Introduction to COSMIC, the Catalogue of Somatic Mutations in Cancer : : Pathway Interaction Database*. N.p., n.d. Web. Nov. 2015. <<http://pid.nci.nih.gov/2008/080708/full/pid.2008.3.shtml>>.
- Black, Douglas L. "Mechanisms of Alternative Pre-Messenger RNA Splicing." *Annual Revues. Biochem* (2003): n. pag. Web.
- Brockman, W., P. Alvarez, S. Young, M. Garber, G. Giannoukos, W. L. Lee, C. Russ, E. S. Lander, C. Nusbaum, and D. B. Jaffe. "Quality Scores and SNP Detection in Sequencing-by-synthesis Systems." *Genome Research* 18.5 (2008): 763-70. Web.
- Burton, Hilary. "Ending a Diagnostic Odyssey." *PHG Foundation*. N.p., 6 May 2014. Web. Winter 2015. <<http://www.phgfoundation.org/blog/16008/>>.
- Cabezas, D. A. "A New X Linked Mental Retardation (XLMR) Syndrome with Short Stature, Small Testes, Muscle Wasting, and Tremor Localises to Xq24-q25." *Journal of Medical Genetics* 37.9 (2000): 663-68. Web.
- Carmichael, Nikkola, Judith Tsipis, Gail Windmueller, Leslie Mandel, and Elicia Estrella. "'Is It Going to Hurt?': The Impact of the Diagnostic Odyssey on Children and Their Families." *Journal of Genetic Counseling J Genet Counsel* 24.2 (2014): 325-35. Web.
- Chen, C.-Y., M.-S. Tsai, C.-Y. Lin, I.-S. Yu, Y.-T. Chen, S.-R. Lin, L.-W. Juan, Y.-T. Chen, H.-M. Hsu, L.-J. Lee, and S.-W. Lin. "Rescue of the Genetically Engineered Cul4b Mutant

- Mouse as a Potential Model for Human X-linked Mental Retardation." *Human Molecular Genetics* 21.19 (2012): 4270-285. Web.
- Chua, Yee Shin, Boon Kim Boh, Wanpen Ponyeam, and Thilo Hagen. "Regulation of Cullin RING E3 Ubiquitin Ligases by CAND1 In Vivo." *PLoS ONE* 6.1 (2011): n. pag. Web.
- "Clinical Utility of Genetic and Genomic Services: A Position Statement of the American College of Medical Genetics and Genomics." *Genetics in Medicine Genet Med* 17.6 (2015): 505-07. Web.
- "CUL4B Gene." *Genetics Home Reference*. N.p., n.d. Web. 12 Dec. 2015.
- Duda, David M., Laura A. Borg, Daniel C. Scott, Harold W. Hunt, Michal Hammel, and Brenda A. Schulman. "Structural Insights into NEDD8 Activation of Cullin-RING Ligases: Conformational Control of Conjugation." *Cell* 134.6 (2008): 995-1006. Web.
- Eberle, Michael A. *Genotyping Accuracy of High Throughput Sequencing*. Illumina, 23 June 2010. Web. Nov. 2015.
<www.fda.gov/downloads/MedicalDevices/NewEvents/WorkshopsConferences/UCM261810PDF>.
- Edwalds-Gilbert, G. "Regulation of MRNA Splicing by Signal Transduction." *Nature Education* 3(9):43. Nature Publishing Group, n.d. Web. Feb.-Mar. 2015.
<<http://www.nature.com/scitable/topicpage/regulation-of-mrna-splicing-by-signal-transduction-14128469>>.
- "The ENCODE Project Consortium. An Integrated Encyclopedia of DNA Elements in the Human Genome." *Nature.com*. Nature Publishing Group, n.d. Web. Fall 2015.
<<http://www.nature.com/nature/journal/v489/n7414/full/nature11247.html>>.
- Fischer, Eric S., Andrea Scrima, Kerstin Böhm, Syota Matsumoto, Gondichatnahalli M. Lingaraju, Mahamadou Faty, Takeshi Yasuda, Simone Cavadini, Mitsuo Wakasugi, Fumio Hanaoka, Shigenori Iwai, Heinz Gut, Kaoru Sugasawa, and Nicolas H. Thomä. "The Molecular Basis of CRL4DDB2/CSA Ubiquitin Ligase Architecture, Targeting, and Activation." *Cell* 147.5 (2011): 1024-039. Web.
- Flanagan, Sarah E., Ann-Marie Patch, and Sian Ellard. "Using SIFT and PolyPhen to Predict Loss-of-Function and Gain-of-Function Mutations." *Genetic Testing and Molecular Biomarkers* 14.4 (2010): 533-37. Web.
- Gerstein, M. B., C. Bruce, J. S. Rozowsky, D. Zheng, J. Du, J. O. Korbel, O. Emanuelsson, Z. D. Zhang, S. Weissman, and M. Snyder. "What Is a Gene, Post-ENCODE? History and Updated Definition." *Genome Research* 17.6 (2007): 669-81. Web.
- Goldenberg, Seth J., Thomas C. Cascio, Stuart D. Shumway, Kenneth C. Garbutt, Jidong Liu, Yue Xiong, and Ning Zheng. "Structure of the Cand1-Cul1-Roc1 Complex Reveals Regulatory Mechanisms for the Assembly of the Multisubunit Cullin-Dependent Ubiquitin Ligases." *Cell* 119.4 (2004): 517-28. Web.
- Goode, D. L., G. M. Cooper, J. Schmutz, M. Dickson, E. Gonzales, M. Tsai, K. Karra, E. Davydov, S. Batzoglou, R. M. Myers, and A. Sidow. "Evolutionary Constraint Facilitates

- Interpretation of Genetic Variation in Resequenced Human Genomes." *Genome Research* 20.3 (2010): 301-10. Web.
- Griffiths, Anthony J. F., Susan R. Wessler, Sean B. Carroll, and John F. Doebley. "RNA: Transcription and Processing." *Introduction to Genetic Analysis*. 10th ed. N.p.: W. H. Freeman, n.d. N. pag. Print.
- He, Y. J., C. M. Mccall, J. Hu, Y. Zeng, and Y. Xiong. "DDB1 Functions as a Linker to Recruit Receptor WD40 Proteins to CUL4-ROC1 Ubiquitin Ligases." *Genes & Development* 20.21 (2006): 2949-954. Web.
- Hershko, Avram, and Aaron Ciechanover. "The Ubiquitin System." *Annu. Rev. Biochem. Annual Review of Biochemistry* 67.1 (1998): 425-79. Web.
- Higa, Leigh Ann, and Hui Zhang. ""Stealing the Spotlight: CUL4-DDB1 Ubiquitin Ligase Docks WD40-Repeat Proteins to Destroy."" *Cell Division* 2 5.PMC. (2007): n. pag. Web. 16 Nov. 2015.
- "InterPro." *Q13620*. N.p., n.d. Web. Oct. 2015.
<<http://www.ebi.ac.uk/interpro/protein/Q13620>>.
- Isidor, Bertrand, Olivier Pichon, Sabine Baron, Albert David, and Cédric Le Caignec. "Deletion of the CUL4B Gene in a Boy with Mental Retardation, Minor Facial Anomalies, Short Stature, Hypogonadism, and Ataxia." *Am. J. Med. Genet. American Journal of Medical Genetics Part A* 152A.1 (2009): 175-80. Web.
- Jackson, Sarah, and Yue Xiong. "CRL4s: The CUL4-RING E3 Ubiquitin Ligases." *Trends in Biochemical Sciences* 34.11 (2009): 562-70. Web.
- Kellis, M., B. Wold, M. P. Snyder, B. E. Bernstein, A. Kundaje, G. K. Marinov, L. D. Ward, E. Birney, G. E. Crawford, J. Dekker, I. Dunham, L. L. Elnitski, P. J. Farnham, E. A. Feingold, M. Gerstein, M. C. Giddings, D. M. Gilbert, T. R. Gingeras, E. D. Green, R. Guigo, T. Hubbard, J. Kent, J. D. Lieb, R. M. Myers, M. J. Pazin, B. Ren, J. A. Stamatoyannopoulos, Z. Weng, K. P. White, and R. C. Hardison. "Defining Functional DNA Elements in the Human Genome." *Proceedings of the National Academy of Sciences* 111.17 (2014): 6131-138. Web.
- Kerscher, Oliver, Rachael Felberbaum, and Mark Hochstrasser. "Modification of Proteins by Ubiquitin and Ubiquitin-Like Proteins." *Annual Review of Cell and Developmental Biology Annu. Rev. Cell Dev. Biol.* 22.1 (2006): 159-80. Web.
- Kerzendorfer, Claudia, Lesley Hart, Rita Colnaghi, Gillian Carpenter, Diana Alcantara, Emily Outwin, Antony M. Carr, and Mark O'Driscoll. "CUL4B-deficiency in Humans: Understanding the Clinical Consequences of Impaired Cullin 4-RING E3 Ubiquitin Ligase Function." *Mechanisms of Ageing and Development* 132.8-9 (2011): 366-73. Web.
- Kniffin, Cassandra L. "MENTAL RETARDATION, X-LINKED, SYNDROMIC, CABEZAS TYPE; MRXSC." *OMIM- Online Mendelian Inheritance in Man*. N.p., n.d. Web. Feb.-Mar. 2015.

- Komander, David. "The Emerging Complexity of Protein Ubiquitination." *Biochim. Soc. Trans. Biochemical Society Transactions* 37.5 (2009): 937-53. Web.
- Landrum, M. J., J. M. Lee, G. R. Riley, W. Jang, W. S. Rubinstein, D. M. Church, and D. R. Maglott. "ClinVar: Public Archive of Relationships among Sequence Variation and Human Phenotype." *Nucleic Acids Research* 42.D1 (2013): n. pag. Web.
- Lazaridis, Konstantinos N., Tammy M. Mcallister, Dusica Babovic-Vuksanovic, Scott A. Beck, Mitesh J. Borad, Alan H. Bryce, Asher A. Chanan-Khan, Matthew J. Ferber, Rafael Fonseca, Kiley J. Johnson, Eric W. Klee, Noralane M. Lindor, Jennifer B. McCormick, Robert R. McWilliams, Alexander S. Parker, Douglas L. Riegert-Johnson, Carolyn R. Rohrer Vitek, Kimberly A. Schahl, Cloann Schultz, Keith Stewart, George C. Then, Eric D. Wieben, and Gianrico Farrugia. "Implementing Individualized Medicine into the Medical Practice." *Am. J. Med. Genet. American Journal of Medical Genetics Part C: Seminars in Medical Genetics* 166.1 (2014): 15-23. Web.
- Lee, Yeon, and Donald C. Rio. "Mechanisms and Regulation of Alternative Pre-mRNA Splicing." *Annu. Rev. Biochem. Annual Review of Biochemistry* 84.1 (2015): 291-323. Web.
- Liu, Helio C., Grigori Enikolopov, and Yuzhi Chen. "Cul4B Regulates Neural Progenitor Cell Growth." *BMC Neuroscience BMC Neurosci* 13.1 (2012): 112. Web.
- Liu, Xiaoming, Xueqiu Jian, and Eric Boerwinkle. "DbNSFP: A Lightweight Database of Human Nonsynonymous SNPs and Their Functional Predictions." *Human Mutation Hum. Mutat.* 32.8 (2011): 894-99. Web.
- Lohmann, Katja, and Christine Klein. "Next Generation Sequencing and the Future of Genetic Diagnosis." *Neurotherapeutics* 11.4 (2014): 699-707. Web.
- Londin, Eric R., Jeffrey Adijanto, Nancy Philp, Antonio Novelli, Emilia Vitale, Chiara Perria, Gigliola Serra, Viola Alesi, Saul Surrey, and Paolo Fortina. "Donor Splice-site Mutation in CUL4B Is Likely Cause of X-linked Intellectual Disability." *Am. J. Med. Genet. American Journal of Medical Genetics Part A* 164.9 (2014): 2294-299. Web.
- Luscombe, N. M., D. Greenbaum, and M. Gerstein. "What Is Bioinformatics? A Proposed Definition and Overview of the Field." *National Center for Biotechnology Information. U.S. National Library of Medicine*, n.d. Web. 13 Dec. 2015. <<http://www.ncbi.nlm.nih.gov/pubmed/11552348>>.
- Maisel, Y. "Genomics and the Psychology of the 'Rare Disease' Diagnostic Odyssey." N.p., n.d. Web. 15 Feb. 2015. <<http://www.Primary-immunodeficiency.com/blog/genomics-the-psychology-of-the-rare-disease-diagnostic-odyssey>>.
- Makela, Nancy L., Patricia H. Birch, Jan M. Friedman, and Carlo A. Marra. "Parental Perceived Value of a Diagnosis for Intellectual Disability (ID): A Qualitative Comparison of Families with and without a Diagnosis for Their Child's ID." *Am. J. Med. Genet. American Journal of Medical Genetics Part A* 149A.11 (2009): 2393-402. Web.

- Mccann, D., R. Bull, and T. Winzenberg. "The Daily Patterns of Time Use for Parents of Children with Complex Needs: A Systematic Review." *Journal of Child Health Care* 16.1 (2012): 26-52. Web.
- Mills, Rachel, and Susanne B. Haga. "Genomic Counseling: Next Generation Counseling." *Journal of Genetic Counseling J Genet Counsel* 23.4 (2013): 689-92. Web.
- Muzzey, Dale, Eric A. Evans, and Caroline Lieber. "Understanding the Basics of NGS: From Mechanism to Variant Calling." *Curr Genet Med Rep Current Genetic Medicine Reports* 3.4 (2015): 158-65. Web.
- Nakagawa, Tadashi, and Yue Xiong. "X-Linked Mental Retardation Gene CUL4B Targets Ubiquitylation of H3K4 Methyltransferase Component WDR5 and Regulates Neuronal Gene Expression." *Molecular Cell* 43.3 (2011): 381-91. Web.
- Nature.com/scitable*. Nature Publishing Group, n.d. Web. 16 Dec. 2015.
<<http://www.nature.com/scitable/definition/single-nucleotide-polymorphism-snp-295>>.
- Ng, Pauline C., and Steven Henikoff. "Predicting the Effects of Amino Acid Substitutions on Protein Function." *Annual Review of Genomics and Human Genetics Annu. Rev. Genom. Human Genet.* 7.1 (2006): 61-80. Web.
- "PolyPhen Help Text." *PolyPhen Help Text*. N.p., n.d. Web. Nov. 2015.
<http://genetics.bwh.harvard.edu/pph/pph_help_text.html>.
- Quality Scores for Next-Generation Sequencing
http://www.illumina.com/Documents/products/technotes/technote_Q-Scores.pdf
- Quintáns, B., A. Ordóñez-Ugalde, P. Cacheiro, A. Carracedo, and M.j. Sobrido. "Medical Genomics: The Intricate Path from Genetic Variant Identification to Clinical Interpretation." *Applied & Translational Genomics* 3.3 (2014): 60-67. Web.
- Ramensky, V. "Human Non-synonymous SNPs: Server and Survey." *Nucleic Acids Research* 30.17 (2002): 3894-900. Web.
- Ravn, K., Sg Lindquist, K. Nielsen, Tl Dahm, and Z. Tümer. "Deletion of CUL4B Leads to Concordant Phenotype in a Monozygotic Twin Pair." *Clinical Genetics* 82.3 (2012): 292-94. Web.
- Sarikas, Antonio, Thomas Hartmann, and Zhen-Qiang Pan. "The Cullin Protein Family." *Genome Biol Genome Biology* 12.4 (2011): 220. Web.
- Shen, E., H. Shulha, Z. Weng, and S. Akbarian. "Regulation of Histone H3K4 Methylation in Brain Development and Disease." *Philosophical Transactions of the Royal Society B: Biological Sciences* 369.1652 (2014): 20130514. Web.
- Sherry, Stephen T., Minhong Ward, and Karl Sirotkin. "dbSNP-Database for Single Nucleotide Polymorphisms and Other Classes of Minor Genetic Variation." (n.d.): n. pag. Web. Nov. 2015. <genome.cshlp.org>.
- Sims, David, Ian Sudbery, Nicholas E. Illott, Andreas Heger, and Chris P. Ponting. "Sequencing Depth and Coverage: Key Considerations in Genomic Analyses." *Nat Rev Genet Nature Reviews Genetics* 15.2 (2014): 121-32. Web.

- Strieter, Eric R., and David A. Korasick. "Unraveling the Complexity of Ubiquitin Signaling." *ACS Chem. Biol. ACS Chemical Biology* 7.1 (2012): 52-63. Web.
- Su, A. I., T. Wiltshire, S. Batalov, H. Lapp, K. A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, M. P. Cooke, J. R. Walker, and J. B. Hogenesch. "A Gene Atlas of the Mouse and Human Protein-encoding Transcriptomes." *Proceedings of the National Academy of Sciences* 101.16 (2004): 6062-067. Web.
- Sunyaev, S., V. Ramensky, I. Koch, W. Lathe, III, A. S. Kondrashov, and P. Bork. "Prediction of Deleterious Human Alleles." *Human Molecular Genetics* 10.6 (2001): 591-97. Web.
- Tarpey, Patrick S., F. Lucy Raymond, Sarah O'Meara, Sarah Edkins, Jon Teague, Adam Butler, Ed Dicks, Claire Stevens, Calli Tofts, Tim Avis, Syd Barthorpe, Gemma Buck, Jennifer Cole, Kristian Gray, Kelly Halliday, Rachel Harrison, Katy Hills, Andrew Jenkinson, David Jones, Andrew Menzies, Tatiana Mironenko, Janet Perry, Keiran Raine, David Richardson, Rebecca Shepherd, Alexandra Small, Jennifer Varian, Sofie West, Sara Widaa, Uma Mallya, Jenny Moon, Ying Luo, Susan Holder, Sarah F. Smithson, Jane A. Hurst, Jill Clayton-Smith, Bronwyn Kerr, Jackie Boyle, Marie Shaw, Lucianne Vandeleur, Jayson Rodriguez, Rachel Slaugh, Douglas F. Easton, Richard Wooster, Martin Bobrow, Anand K. Srivastava, Roger E. Stevenson, Charles E. Schwartz, Gillian Turner, Jozef Gecz, P. Andrew Futreal, Michael R. Stratton, and Michael Partington. "Mutations in CUL4B, Which Encodes a Ubiquitin E3 Ligase Subunit, Cause an X-linked Mental Retardation Syndrome Associated with Aggressive Outbursts, Seizures, Relative Macrocephaly, Central Obesity, Hypogonadism, Pes Cavus, and Tremor." *The American Journal of Human Genetics* 80.2 (2007): 345-52. Web.
- Tarpey, Patrick S., Raffaella Smith, Erin Pleasance, Annabel Whibley, Sarah Edkins, Claire Hardy, Sarah O'meara, Calli Latimer, Ed Dicks, Andrew Menzies, Phil Stephens, Matt Blow, Chris Greenman, Yali Xue, Chris Tyler-Smith, Deborah Thompson, Kristian Gray, Jenny Andrews, Syd Barthorpe, Gemma Buck, Jennifer Cole, Rebecca Dunmore, David Jones, Mark Maddison, Tatiana Mironenko, Rachel Turner, Kelly Turrell, Jennifer Varian, Sofie West, Sara Widaa, Paul Wray, Jon Teague, Adam Butler, Andrew Jenkinson, Mingming Jia, David Richardson, Rebecca Shepherd, Richard Wooster, M. Isabel Tejada, Francisco Martinez, Gemma Carvill, Rene Goliath, Arjan P M De Brouwer, Hans Van Bokhoven, Hilde Van Esch, Jamel Chelly, Martine Raynaud, Hans-Hilger Ropers, Fatima E. Abidi, Anand K. Srivastava, James Cox, Ying Luo, Uma Mallya, Jenny Moon, Josef Parnau, Shehla Mohammed, John L. Tolmie, Cheryl Shoubridge, Mark Corbett, Alison Gardner, Eric Haan, Sinitdhorn Rujirabanjerd, Marie Shaw, Lucianne Vandeleur, Tod Fullston, Douglas F. Easton, Jackie Boyle, Michael Partington, Anna Hackett, Michael Field, Cindy Skinner, Roger E. Stevenson, Martin Bobrow, Gillian Turner, Charles E. Schwartz, Jozef Gecz, F. Lucy Raymond, P. Andrew Futreal, and Michael R. Stratton. "A Systematic, Large-scale Resequencing Screen of X-chromosome Coding Exons in Mental Retardation." *Nature Genetics Nat Genet* 41.5 (2009): 535-43. Web.

- The UniProt Consortium UniProt: a Hub for Protein Information *Nucleic Acids Res.* 43: D204-D212 (2015)
- Varshavsky, Alexander. "The Ubiquitin System, an Immense Realm." *Annu. Rev. Biochem. Annual Review of Biochemistry* 81.1 (2012): 167-76. Web.
- Vaz-Drago, R., M. T. Pinheiro, S. Martins, F. J. Enguita, M. Carmo-Fonseca, and N. Custodio. "Transcription-coupled RNA Surveillance in Human Genetic Diseases Caused by Splice Site Mutations." *Human Molecular Genetics* 24.10 (2015): 2784-795. Web.
- Vulto-van Silfhout, Anneke T., Tadashi Nakagawa, Nadia Bahi-Buisson, Stefan A. Haas, Hao Hu, Melanie Bienek, Lisenka E.I.m. Vissers, Christian Gilissen, Andreas Tzschach, Andreas Busche, Jörg Müsebeck, Patrick Rump, Inge B. Mathijssen, Kristiina Avela, Mirja Somer, Fatma Doagu, Anju K. Philips, Anita Rauch, Alessandra Baumer, Krysta Voesenek, Karine Poirier, Jacqueline Vigneron, Daniel Amram, Sylvie Odent, Magdalena Nawara, Ewa Obersztyń, Jacek Lenart, Agnieszka Charzewska, Nicolas Lebrun, Ute Fischer, Willy M. Nillesen, Helger G. Yntema, Irma Järvelä, Hans-Hilger Ropers, Bert B.a. De Vries, Han G. Brunner, Hans Van Bokhoven, F. Lucy Raymond, Michèl A.a.p. Willemsen, Jamel Chelly, Yue Xiong, A. James Barkovich, Vera M. Kalscheuer, Tjitske Kleefstra, and Arjan P.m. De Brouwer. "Variants in CUL4B Are Associated with Cerebral Malformations." *Human Mutation* 36.1 (2014): 106-17. Web.
- Wei, Ning, and Xing-Wang Deng. "Making Sense of the COP9 Signalosome: A Regulatory Protein Complex Conserved from Arabidopsis to Human." *Trends in Genetics* 15.3 (1999): 98-103. Web.
- Woolfe, Adam, James C. Mullikin, and Laura Elnitski. "Genomic Features Defining Exonic Variants That Modulate Splicing." *Genome Biol Genome Biology* 11.2 (2010): n. pag. Web.
- World Health Organization. "Intellectual Level." *Multiaxial Classification of Child and Adolescent Psychiatric Disorders: The ICD-10 Classification of Mental and Behavioural Disorders in Children and Adolescents*. Cambridge: Cambridge UP, 2002. 197. *Google Books*. Web. Feb.-Mar. 2015.
<https://books.google.com/books/about/Multiaxial_Classification_of_Child_and_A.html?id=ddWQFy0iG64C>.
- Zhao, Wei, Baichun Jiang, Huili Hu, Shuqian Zhang, Shuaishuai Lv, Jupeng Yuan, Yanyan Qian, Yongxin Zou, Xi Li, Hong Jiang, Fang Liu, Changshun Shao, and Yaoqin Gong. "Lack of CUL4B Leads to Increased Abundance of GFAP-positive Cells That Is Mediated by PTGDS in Mouse Brain." *Hum. Mol. Genet. Human Molecular Genetics* 24.16 (2015): 4686-697. Web.
- Zou, Yongxin, Qiji Liu, Bingxi Chen, Xiyu Zhang, Chenhong Guo, Haibin Zhou, Jiangxia Li, Guimin Gao, Yishou Guo, Chuanzhu Yan, Jianjun Wei, Changshun Shao, and Yaoqin Gong. "Mutation in CUL4B, Which Encodes a Member of Cullin-RING Ubiquitin Ligase

Complex, Causes X-Linked Mental Retardation." *The American Journal of Human Genetics* 80.3 (2007): 561-66. Web.