Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Vedant Daga April 9, 2021

Identifying the causal variant of a rare skeletal disorder using whole-genome sequencing

by

Vedant Daga

Michael Zwick
Adviser

Biology

Michael Zwick

Adviser

Ron Calabrese

Committee Member

Steven La Fleur

Committee Member

2021

Identifying the causal variant of a rare skeletal disorder using whole-genome sequencing

By

Vedant Daga

Michael Zwick

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Biology

2021

Abstract

Identifying the causal variant of a rare skeletal disorder using whole-genome sequencing

By Vedant Daga

Desbuquois dysplasia (DBQD) is a severe skeletal disorder characterized by joint dislocation and stunted natal growth. We obtained and sequenced the DNA of two first cousins who clinically presented with a lethal form of DBQD. Our objective was to determine the causal variant for this disorder. Our first hypothesis was the disorder was autosomal recessive, and the individuals were homozygous for the same allele. The second hypothesis was the individuals were compound heterozygous, caused by two different alleles in the same gene. We performed whole-genome sequencing (WGS) of the two cousins, followed by computational analysis of the WGS data to identify a list of genetic variants. We then comprehensively sifted through the list of variants to identify putative mutations meeting the two hypotheses' criteria. The results of this study can provide insight into the disease affecting the two individuals.

Identifying the causal variant of a rare skeletal disorder using whole-genome sequencing

By

Vedant Daga

Michael Zwick

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Biology

2021

Acknowledgements

We would like to thank the Zwick Lab for help with the analyses. Additionally, we would like to thank the Cutler Lab (Dave Cutler, Rich Johnston) for assistance with scripts to go through various data files. We would also like to thank Lora Bean for assistance with the next steps of the homozygous analysis. We would like to thank Pankaj Chopra for assistance with running modified scripts. We would like to thank the TYE Team, without whom this project would not be possible. Lastly, we would like to acknowledge the following funding source, the Emory Treasure Your Exceptions Program.

Table of Contents

Abstract……1

Introduction……..2

Materials and Methods……3

Results……6

Discussion……10

Acknowledgements……..12

References……..13

Figures/Appendix……16

Identifying the causal variant of a rare skeletal disorder using whole-genome sequencing

Authors: Vedant Daga (VD), Trenell Mosley (TM), Michael E. Zwick (MEZ)

Author Contributions: VD executed most experiments, analyzed the data, and wrote the original draft. TM obtained samples, developed the quality control pipeline, and assisted with experiments. MEZ will review and edit the drafts until it is complete.

Abstract

Desbuquois dysplasia (DBQD) is a severe skeletal disorder characterized by joint dislocation and stunted natal growth. We obtained and sequenced the DNA of two first cousins who clinically presented with a lethal form of DBQD. Our objective was to determine the causal variant for this disorder. Our first hypothesis was the disorder was autosomal recessive, and the individuals were homozygous for the same allele. The second hypothesis was the individuals were compound heterozygous, caused by two different alleles in the same gene. We performed whole-genome sequencing (WGS) of the two cousins, followed by computational analysis of the WGS data to identify a list of genetic variants. We then comprehensively sifted through the list of variants to identify putative mutations meeting the two hypotheses' criteria. The results of this study can provide insight into the disease affecting the two individuals.

Introduction

Uncovering the causes of rare diseases has revealed important information about human development. For rare skeletal disorders, over 372 different conditions have been categorized and placed in 37 groups (Superti-Furga and Unger, 2007). Yet, the underlying genetic causes of many of these skeletal conditions remain to be discovered.

For our study, we obtained the DNA samples of two first cousins (Figure 1) that presented with short stature, joint laxity, distinct facial characteristics, and inclusions in chondrocytes. Both individuals died suddenly and other affected siblings died within a couple of months into infancy. Neither of the parents of either cousin were affected by the condition. We hypothesized the two cousins exhibited a rare autosomal recessive bone disorder. We suspected the mutation is likely homozygous at a single locus, or it was compound heterozygous, with two different mutant alleles at the same locus.

The symptoms the two cousins displayed closely resemble those of Desbuquois dysplasia (DBQD). However, unlike standard DBQD, the individuals died in utero, and other affected siblings also died in utero or suddenly passed away in early infancy. DBQD Type I is already associated with two genes, *CANT1* and *XYLT1* (Bui et al., 2014; Laccone et al., 2011). The gene *XYLT1* is also associated with DBQD Type II and Baratela-Scott Syndrome; the gene's promotor regions contains a pathogenic GGC repeat expansion, and the gene itself is hypermethylated at exon 1 (LaCroix et al., 2019). In addition to the two hypotheses above, we also sought to test the hypothesis that the two probands might have novel coding sequence variants or compound heterozygous mutations at the *XYLT1* or CANT1 genes.

To identify potential candidate variants for the disorder, we used whole-genome sequencing and performed a computational analysis of the data on the two affected individuals'

genomes. We mapped the WGS data using the PEMapper software and identified variants using the PECaller software (Johnston et al., 2017). To ensure we analyzed high-quality data, the list of variants was filtered and went through a quality control check. In order to interpret the list of variants, we used Bystro, an online annotation program that uses information from various online genomic databases, to compile an annotation file (Kotlar et al., 2018). We created different search queries to narrow down our results to find variants that best matched the causal mutation criteria. This procedure was carried out for both the homozygous hypothesis as well as the compound heterozygous hypothesis. Identifying the mutation behind this disorder can help us understand the mechanisms behind fetal skeletal development.

Materials and Methods

DNA Library Prep and Whole-Genome Sequencing

DNA samples were normalized to 1,000ng of DNA in 50ul of water. Following normalization, samples were acoustically sheared via Covaris LE-220 instrument to a final fragment size of ~350-400bp. The sheared DNA was then transformed into a standard Illumina paired-end sequencing library via standard methods. The sheared DNA was end-repaired and A-tailed using New England Biolabs End-Repair and A-Tailing kits, respectively using the manufacturer's recommended conditions. Following each step, the library was purified via Agencourt AMPure XP beads and eluted in water. Standard Illumina paired-end adaptors were ligated to the A-tailed DNA via New England BioLabs Rapid Ligation kit. Following ligation, the reactions were purified using AMPure XP beads. The purified ligated DNA was amplified via PCR using KAPA Biosystems HIFI PCR kit using 6 cycles of PCR. The primers were standard Illumina primers with a custom 7-base sample barcode in the i7 position to allow

sample identification/de-multiplexing following sequencing. The final library was quality controlled using size verification via PerkinElmer LabChip GX and real-time PCR using the KAPA SYBR FAST qPCR Master Mix, primers and standards according to the manufacturer's directions. Libraries were normalized to 2.5 nM stocks for use in clustering and sequencing.

Sequence Alignment using PEMapper

We used PEMapper software to align the reads from our two DNA sample, SL158185 and SL313876, against the reference genome Hg38 as reported by the University of California at Santa Cruz (UCSC) Genome Browser and in Johnston et al. 2017 (Johnston et al., 2017). PEMapper reads FastQ files containing our sequenced data and the software maps the reads to their corresponding position in the reference genome and produces several output files (pileup and indel) to capture the results. The output summary file displayed mapping statistics which indicate the success of the mapping experiment. The pileup file contains a summary of the bases present at each position in the chromosomes' genes, while the indel file identifies the insertions and deletions in the genomes.

Variant Calling using PECaller

PECaller is a program that uses the pileup and indel files from the PEMapper output and joint calls them with 57 other high-quality reference genomes. These 57 reference genomes serve as our controls (Johnston et al., 2017). PECaller then produces a list of variants in the sample and reference genomes as a .snp file, that contains the differing alleles found at each location of the SNPs. We merged the contents of the indel file from the PEMapper output with the .snp file to create a new "merged" .snp file. We used this list of variants as the starting point for

Quality Control Pipeline

The purpose of quality control is to verify the success of our mapping and calling experiments and indicate whether our dataset contains high quality data. The pipeline checks the transition-transversion ratio as well as silent/replacement calls for individual genomes in the dataset. Transitions are base changes from one purine or pyrimidine to the other, while transversions are base changes from a purine to a pyrimidine or vice versa (Johnston et al., 2017). The silent/replacement call determines the number of mutations that are "silent," meaning they do not alter the resulting amino acid, while replacements do modify the amino acid. The expected value for the transition/transversion ratio is 2.00-2.04 and the expected value for the silent/replacement call is between 1.05-1.15. If the actual values are not close to the expected values, it indicates something wrong with the sample or the dataset.

The next step was to run the two sample genomes and the control genome through the PLINK software (Purcell et al., 2007). By computing each variant site's missing rate and performing a Mendelian error check, PLINK validated the data quality. The missing rate determines how many samples in the dataset have the variant confidently called, and only variants with a missing call rate of less than 10% are kept in the dataset. Checking for Mendelian error ensures SL158185 and SL313876 follow standard inheritance patterns based on known family history.

Variant Annotation and Analysis

Using a web application called Bystro.io, we explored all the variants remaining in our merged .snp file. Once uploaded, Bystro created an annotation file that compiled information

about the genes gathered from various databases (Kotlar et al., 2018). To narrow down our search results, we used effective search queries to find variants meeting our hypotheses' criteria. Our main search query was “Homozygous: (SL158185 AND SL313876),” to return variants for which both SL158185 and SL313876 were homozygous. Once the results were returned, we downloaded the annotation file and reviewed other relevant information, such as function, type, and rarity, that may help us determine whether or not a specific variant is a putative causative mutation. Variants not meeting these parameters are ruled out of the search results. A similar procedure was used to create a list of variants matching the compound heterozygous hypothesis.

After creating a list of variants that matched the hypotheses, we also started searching for genes associated with the variants in the Online Mendelian in Man (OMIM) catalog. OMIM is an online catalog of all human genes and genetic phenotypes that contains information about all known Mendelian disorders and focuses on the relationship between genotype and phenotype.

Results

Mapping and Calling

The summary file statistics from the PEMapper output indicate that both SL158185 and SL313876 were successfully mapped. The average coverage for SL158185 was 28.5991 and the average coverage for SL313876 was 24.5948. After running PECaller, the program gave us an output of a .snp file. This file contained a total of 7,014,803 variants. This .snp file was the starting point for our quality control (QC) pipeline. After running the .snp file through the QC pipeline, we reduced the number of variants to 6,599,097.

Quality Control Pipeline

The quality control pipeline gave us useful statistics to determine the quality of our samples. The statistics confirmed that both individuals (SL158185 and SL313876) had met the expected values for QC. The expected value for the transition/transversion ratio was 2.00-2.04 and the expected value for the silent/replacement call is between 1.05-1.15. The transition/transversion ratio for SL158185 was 2.037948, while SL313876 had a ratio of 2.038211. Additionally, the silent/replacement ratio for SL158185 was 1.153698 and for SL313876 the ratio was 1.174695. Following these ratio checks, we used an F statistic test to confirm the sex of the individuals. The PLINK program provided a sex-check for the individuals (Purcell et al., 2007). SL158185 had an F statistic of 0.05796 and SL313876 had an F statistic of 0.96390. This indicates that SL158185 was female, as its F statistic was below 0.2, and SL313876 was male, as its F statistic was above 0.8.

Variant Search and Analysis

We first focused on the hypothesis that the disorder was homozygous for identical alleles at a single locus. To narrow the list of variants consistent with our hypothesis, we filtered our results on Bystro to include variants where SL158185 and SL313876 were both homozygous. Additional search/filter parameters included sorting variants by combined annotation-dependent depletion (CADD) score, minor allele frequency, the site type, and exonic allele function. Variants with a low minor allele frequency are uncommon in the population and can be considered rare. We considered a minor allele frequency of 0.01 to be our threshold, using frequencies listed in the GnomAD database (Lek et al., 2016). Another parameter used was the CADD score, which is a genome-wide measure of the predicted deleteriousness of a variant

(Richardson et al., 2016). We focused on mutations that had a CADD score greater than 15 when considering potential causal variants.

We were able to narrow down the results from over six million variants in the .snp file to a more manageable number of variants. Filtering variants that were homozygous in SL158185 and SL313876 left us with approximately 500,000 variants. To ensure variants would be in the coding sequence of genes, we filtered for variants that were "exonic" and "nonsynonymous" according to the NIH's Reference Sequence Database, leaving only 2,853 variants. Of those variants, those with a CADD score less than 15 were ruled out, so we were left with 413 variants. Since none of those genes had a minor allele frequency of less than 0.01, we focused on variants with minor allele frequencies that were unreported in GNOMAD for a total of 3 variants. These candidate genes include *RIMBP3C*, *KRTAP5-5*, and *HNRNPCL1*, and were homozygous for the same allele in both individuals. There were no variants observed in either *XYLT1* or *CANT1* meeting our criteria for either hypothesis.

The same search criteria were used for the compound heterozygous search, which included filtering for variants that are "exonic," "nonsynonymous," met a CADD score greater than 15, and had a minor allele frequency less than 0.01. This resulted in a total list of eight variants, or four pairs of genes with 2 variants each. The four genes in the results include *DNAH11*, *MELTF*, *TTN*, and *USP31*.

Candidate Genes – Homozygous Analysis

The first of the three candidate genes from the homozygous analysis was *RIMBP3C*, or RIMS Binding Protein 3C. It "expresses a single transcript" and consists of 1639 amino acid residues in humans (Mittelstaedt and Schoch, 2007). *RIMBP3C* is also encoded by a single exon

and levels of *RIMBP3C* are especially high outside the nervous system. *RIMBP3C* is known to play a role in spermiogenesis (Zhou et al., 2009). The variant that was observed changed the amino acid from tryptophan to arginine.

The second of the three candidate genes is *KRTAP5-5*, or Keratin Associated Protein 5-5. It is a high sulfur protein that is one of the major structural components of mammalian hair (Yahagi et al., 2004). The variant that was observed changed the amino acid from arginine to leucine.

The last candidate gene is *HNRNPCL1*, or Heterogenous Nuclear Ribonucleoprotein C-Like 1. It is predicted to play a role in nucleosome assembly as it neutralizes proteins (Swaminathan et al., 2011). One of *HNRNPCL1*'s target genes includes multiple genes from the *PRAME* Family, which has been associated with retinoic acid receptor binding (Fishilevich et al., 2017; Stelzer et al., 2016). Additionally, the observed variant changes the amino acid from aspartic acid to glycine.

Gene	Chromosome	Position	Base Change	Amino Acid Change	CADD Score	Minor Allele Frequency
RIMBP3C	22	21387042	T → C	Trp → Arg	17.7	No information
KRTAP5-5	11	1629890	G → T	Arg → Leu	16.1	No information
HNRNPCL1	1	12847526	A → G	Asp → Gly	16.9	No information

Table 1. This table contains variants filtered from our main search query along with additional criteria needed to meet the homozygous hypothesis.

Candidate Genes – Compound Heterozygous Analysis

The compound heterozygous analysis returned a total of four different genes, or four pairs of variants for a total of 8 variants. Based on the minor allele frequency and the respective CADD score, we looked into each of these genes to understand the function of each. *DNAH11* plays a role in generating respiratory cilia (Chapelin et al., 1997). *MELTF* is important to encoding a cell-surface glycoprotein found on melanoma cells. It is also involved in iron cellular

uptake. *TTN* encodes a large, abundant protein in different regions of striated muscle. *USP31* plays a role in the deubiquitinating TRAF proteins in certain pathways (Tzimas et al., 2006).

Gene	Chromosome	Position	Base Change	Amino Acid Change	CADD Score	Minor Allele Frequency
DNAH11	7	21864563	C → G	Pro → Arg	23.1	0.00151779
DNAH11	7	21852546	C → T	Ala → Val	33	0.00174385
MELTF	3	197017123	C → T	Arg → Trp	33	0.00562124
MELTF	3	197008816	G → A	Ala → Thr	34	0.00905211
TTN	2	178571605	A → G	Asn → Asp	17.4	6.47E-05
TTN	2	178557385	C → T	Arg → Cys	24.2	0.00148646
USP31	16	23106225	T → C	Leu → Pro	30	0.000161363
USP31	16	23090706	G → T	Asp → Tyr	25.1	0.00584286

Table 2. This table contains variants filtered from our main search query along with additional criteria needed to meet the compound heterozygous hypothesis.

Discussion

The National Organization of Rare Disorders states that 1 in 10 Americans are affected by rare disorders. Additionally, rare skeletal disorders are a leading cause of disability. Not much is known about these disorders, but it is becoming increasingly feasible to identify their causes with improved genomics technology. To understand the lethal skeletal dysplasia both first cousins displayed, we used WGS to analyze their genomes and search for causal variants.

We hypothesize these abnormalities are caused by an autosomal recessive disorder caused by a homozygous mutation. As shown in Figure 1, none of the parents were affected by the skeletal dysplasia. Still, both sets of parents did have additional children that either died in utero or infancy and exhibited similar phenotypes to SL158185 and SL313876. Our alternate hypothesis is that a compound heterozygous mutation caused this disorder.

We began by pursuing the simplest hypothesis, that the disorder is caused by an identical mutation found at the same locus in both probands. Once the probands' DNA was sequenced and

ran through a series of computational tools, we processed the data with the quality control pipeline. Both of the samples passed most of the QC tests and we proceeded to variant analysis keeping the QC results in mind.

The filtering and searching in the homozygous analysis left us with three candidate genes. *KRTAP5-5* seems an unlikely candidate as it plays a role in the development of hair proteins and structure, quite far from skeletal development. Additionally, *RIMBP3C* plays a role in spermiogenesis, and is unlikely to cause a skeletal problem. However, *HNRNPCL1* is a likely candidate because some of its target genes are in the *PRAME* gene family, which are known to be associated with retinoic acid receptor binding. Seeing how retinoic acid has a crucial role in development, the *PRAMEF* genes are likely affected by *HNRNPCL1*. Retinoic acid and its receptors have an important role in craniofacial and skeletal development, and when there is a mutation, abnormalities are present (Lohnes et al., 1994).

The heterozygous analysis returned four different candidate genes. It is difficult to say with certainty whether any of these genes have an impact on skeletal development. *DNAH11* is an unlikely candidate as it plays a role in generating respiratory cilia (Chapelin et al., 1997). *MELTF* is important to encoding a cell-surface glycoprotein found on melanoma cells and is involved in the cellular uptake of iron, which may not be related to skeletal development. *TTN* encodes a large, abundant protein in different regions of striated muscle. *USP31* plays a role in the deubiquitinating TRAF proteins in certain pathways (Tzimas et al., 2006). Functional testing of variants may be required to fully understand if these variants caused the disorder observed in the probands.

Our next steps would be to conduct a Run of Homozygosity (ROH) analysis to identify regions within SL158185 and SL313876 where both genomes contain runs of homozygosity and

have the same allele. This could test the hypothesis that an unknown non-coding variant causes the disorder. We can also pursue a Copy Number Variants (CNV) analysis in order to find genes that are located within deletions and duplications shared by the probands. Additionally, we will continue to look into the OMIM catalog to find candidate genes for potentially associated diseases. Another option to consider would be using data files from intermediate steps and using various online Artificial Intelligence platforms to help understand and analyze variants and their relevance to the disease. The ultimate demonstration of causation of a specific mutation may require additional independently affected individuals from different families be identified in order to identify additional putative mutations.

Acknowledgements

We would like to thank the Zwick Lab for help with the analyses. Additionally, we would like to thank the Cutler Lab (Dave Cutler, Rich Johnston) for assistance with scripts to go through various data files. We would also like to thank Lora Bean for assistance with the next steps of the homozygous analysis. We would like to thank Pankaj Chopra for assistance with running modified scripts. We would like to thank the TYE Team, without whom this project would not be possible. Lastly, we would like to acknowledge the following funding source, the Emory Treasure Your Exceptions Program.

References

Bui, C., Huber, C., Tuysuz, B., Alanay, Y., Bole-Feysot, C., Leroy, J.G., Mortier, G., Nitschke, P., Munnich, A., and Cormier-Daire, V. (2014). XYLT1 mutations in Desbuquois dysplasia type 2. American journal of human genetics *94*, 405-414.

Chapelin, C., Duriez, B., Magnino, F., Goossens, M., Escudier, E., and Amselem, S. (1997). Isolation of several human axonemal dynein heavy chain genes: genomic structure of the catalytic site, phylogenetic analysis and chromosomal assignment. *412*, 325-330.

Fishilevich, S., Nudel, R., Rappaport, N., Hadar, R., Plaschkes, I., Iny Stein, T., Rosen, N., Kohn, A., Twik, M., Safran, M*., et al.* (2017). GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database *2017*, bax028-bax028.

Johnston, H.R., Chopra, P., Wingo, T.S., Patel, V., International Consortium on, B., Behavior in 22q11.2 Deletion, S., Epstein, M.P., Mulle, J.G., Warren, S.T., Zwick, M.E*., et al.* (2017). PEMapper and PECaller provide a simplified approach to whole-genome sequencing. Proc Natl Acad Sci U S A *114*, E1923-E1932.

Kotlar, A.V., Trevino, C.E., Zwick, M.E., Cutler, D.J., and Wingo, T.S.J.G.B. (2018). Bystro: rapid online variant annotation and natural-language filtering at whole-genome scale. *19*, 14.

Laccone, F., Schoner, K., Krabichler, B., Kluge, B., Schwerdtfeger, R., Schulze, B., Zschocke, J., and Rehder, H. (2011). Desbuquois dysplasia type I and fetal hydrops due to novel mutations in the CANT1 gene. European journal of human genetics : EJHG *19*, 1133-1137.

LaCroix, A.J., Stabley, D., Sahraoui, R., Adam, M.P., Mehaffey, M., Kernan, K., Myers, C.T., Fagerstrom, C., Anadiotis, G., Akkari, Y.M*., et al.* (2019). GGC Repeat Expansion and Exon 1 Methylation of XYLT1 Is a Common Pathogenic Variant in Baratela-Scott Syndrome. American journal of human genetics *104*, 35-44.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B.*, et al.* (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285.

Lohnes, D., Mark, M., Mendelsohn, C., Dolle, P., Dierich, A., Gorry, P., Gansmuller, A., and Chambon, P. (1994). Function of the retinoic acid receptors (RARs) during development (I). Craniofacial and skeletal abnormalities in RAR double mutants. *120*, 2723-2748.

Mittelstaedt, T., and Schoch, S. (2007). Structure and evolution of RIM-BP genes: Identification of a novel family member. Gene *403*, 70-79.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J.*, et al.* (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics *81*, 559-575.

Richardson, T.G., Campbell, C., Timpson, N.J., and Gaunt, T.R. (2016). Incorporating Non-Coding Annotations into Rare Variant Analysis. PLOS ONE *11*, e0154181.

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T.I., Nudel, R., Lieder, I., Mazor, Y.*, et al.* (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *54*, 1.30.31-31.30.33.

Superti-Furga, A., and Unger, S. (2007). Nosology and classification of genetic skeletal disorders: 2006 revision. Am J Med Genet A *143A*, 1-18.

Swaminathan, S., Sungeun, K., Li, S., Risacher, S.L., Foroud, T., Pankratz, N., Potkin, S.G., Huentelman, M.J., Craig, D.W., Weiner, M.W.*, et al.* (2011). Genomic Copy Number Analysis in Alzheimer's Disease and Mild Cognitive Impairment: An ADNI Study. International Journal of Alzheimer's Disease *2011*, 1-10.

Tzimas, C., Michailidou, G., Arsenakis, M., Kieff, E., Mosialos, G., and Hatzivassiliou, E.G. (2006). Human ubiquitin specific protease 31 is a deubiquitinating enzyme implicated in activation of nuclear factor-κB. Cellular Signalling *18*, 83-92.

Yahagi, S., Shibuya, K., Obayashi, I., Masaki, H., Kurata, Y., Kudoh, J., and Shimizu, N. (2004). Identification of two novel clusters of ultrahigh-sulfur keratin-associated protein genes on human chromosome 11. Biochemical and Biophysical Research Communications *318*, 655-664.

Zhou, J., Du, Y.-R., Qin, W.-H., Hu, Y.-G., Huang, Y.-N., Bao, L., Han, D., Mansouri, A., and Xu, G.-L. (2009). RIM-BP3 is a manchette-associated protein essential for spermiogenesis. Development *136*, 373.

Figures/Appendix

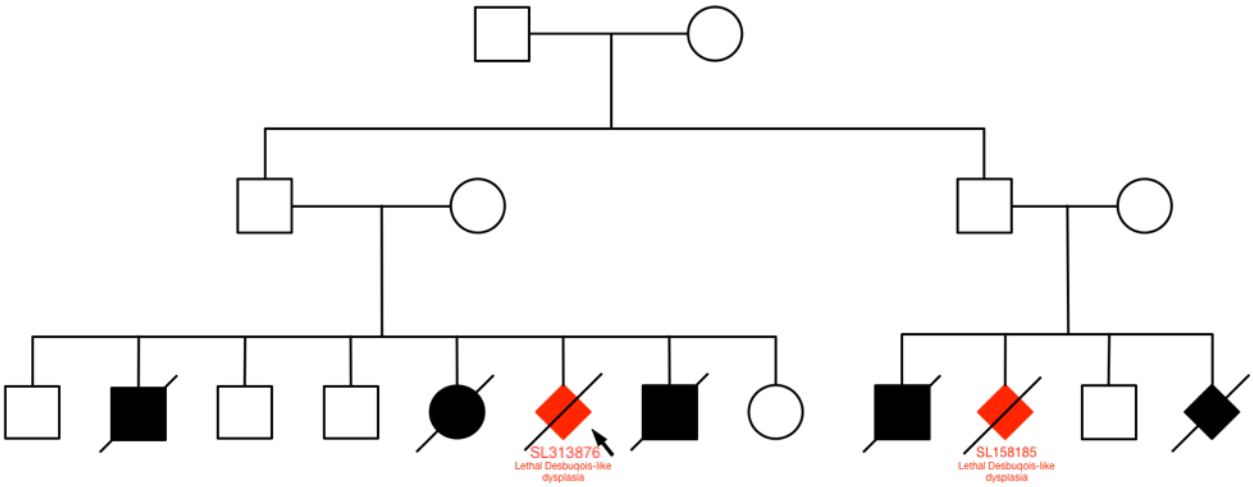

Figure 1. A pedigree of the case family of our study. The affected individuals are indicated in black, while the affected individuals we have samples for are indicated in red.