Whole Exome Sequencing and Analysis to Rule Out Three Bone-Expressed Genes as Candidates for Split Hand Foot Malformation

C. Ronnie Funk,[1] Melanie M. May,[2] Anna V. Blenda, Ph.D,[3] & Charles E. Schwartz, Ph.D.[2]

1 Erskine College, 2 Greenwood Genetic Center, & 3 USC School of Medicine Greenville

Abstract

Family studies are useful in the genetic analysis of rare conditions. In this study, a rare autosomal dominant disease, called split hand foot malformation (SHFM), was analyzed in a family. Mutations in the known SHFM genes are thought to abrogate development of the central rays of the hands and feet through various mechanisms. A typical SHFM phenotype manifests as syndactyly, median clefts of the hands and feet, and/or aplasia or hypoplasia of the phalanges, metacarpals, and metatarsals. Although family members presented with the SHFM phenotype, mutations were absent in the nine loci associated with nonsyndromal SHFM, while patient history and presentation eliminated syndromal SHFM as a potential diagnosis. In order to determine a hypothesized genetic cause of SHFM in this family, genomic DNA of four family members was subjected to whole exome sequencing (WES) to screen all protein-encoding regions of the DNA. The sequencing data were filtered to identify gene alterations that were novel, and not present in databases of sequence variants. Based on the analysis, mutations in three genes, each expressed in bone, MAP1b, NEK1, and CHD6 were selected for further investigation based on known biological functions. While this work was underway, twins – one of whom was clearly affected – were born to an affected father. The extension of the pedigree enabled segregation analysis based on DNA sequencing to observe whether mutations in MAP1b, NEK1, and CHD6 were present in the affected newborn twin. Through this methodology, none of the three gene alterations were associated with the SHFM phenotype in the family. This exposition conveys a methodology for analyzing a rare genetic disorder with the dual goals of strengthening future study designs and communicating the value of confirming WES data in order to prevent incorrectly counseling patients and their families on rare genetic conditions.

INTRODUCTION

Patients with Split Hand Foot Malformation (SHFM) may present with polydactyly, syndactyly, median clefts, hypoplasia, or hyperplasia of the digits of the hands and feet.[1] Abrogation of central-ray formation in developing limbs is thought to cause SHFM.[2] Multiple genetic mutations can cause SHFM through putative alteration of gene expression in the proliferating zone of the apical ectodermal ridge in a developing limb bud.[1]

In some patients, SHFM is present as an isolated condition without other abnormalities, and these patients are properly diagnosed with nonsyndromal SHFM.[3] In other patients, SHFM occurs along with additional abnormalities such as long-bone deficiency or cleft lip/palate, and these cases are known as syndromal SHFM.[4,5] In both nonsyndromal and syndromal SHFM, penetrance and expressivity of the malformation varies.[3,6] Thus, penetrance and expressivity need to be considered to accurately interpret segregation analyses.

Family K6821 presented a nonsyndromal SHFM phenotype. The family elected to undergo genetic testing to determine the cause of their condition. Through a variety of screening and diagnostic procedures, this family was found to lack mutations in any of the six unique chromosomal loci associated

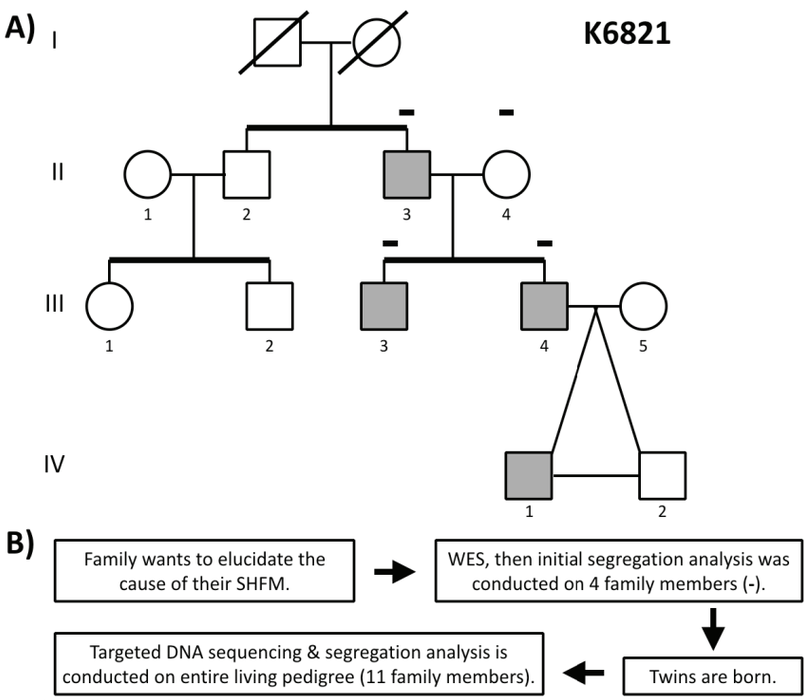

Figure 1: A) Pedigree of family K6821. Individuals with the SHFM phenotype are shaded gray. Square = male; circle = female, triangle = identical twins, and a bar over square or circle denotes Whole Exome Sequencing (WES).

B) Chronology of events in this study.

Table 1: RNA primer sequences used to amplify MAP1b Exon 4, NEK1 Exon 31, and CHD6 Exon 23.

Gene	[1]Forward Primer (5' to 3')	[2]Reverse Primer (5' to 3')
M13 Tag	TTGTAAAACGACGGCCAGTG	CACACAGGAAACAGCTATGACCATG
MAP1b Exon 4	M13F-CTAGCCCTGTCCTGAAGGTG	M13R-TTCTTTTGTCTCCCCCAGTG
NEK1 Exon 31	M13F-TTGTGTTACATACTGTGCCTTATCAA	M13R-TGCAAAGGATTCTTCTGGTG
CHD6 Exon 23	M13F-CCACTTGCCCTATGCTGTCT	M13R-CAGAAGTAGCACAAGAGCATGAA

[1]Forward primers target intronic DNA upstream of target genes.

[2]Reverse primers target intronic DNA downstream of target genes.

with nonsyndromal SHFM (7q21.2-q21.3, Xq26, 10q24, 3q27, 2q31, and 12q31).[1] Furthermore, the family lacked a rearrangement at the three chromosomal loci associated with SHFM or SHFM with long-bone deficiency (1q42.2-q43, 6q14.1, and 17p13.3).[1] In summation, these data indicated the family lacked a chromosomal rearrangement association of known nonsyndromal SHFM genes.

An investigation was designed and initiated with the goal of elucidating the unknown molecular etiology of SHFM in this family. Due to the family history of SHFM, the underlying etiology was hypothesized to involve genetics, rather than being solely environmental. A genetic mutation in known exonic DNA was hypothesized to underlie SHFM in this family. Thus, whole exome sequencing (WES) was selected as the genetic test of choice. WES enables analysis of all the variation present in known exonic regions of DNA. From the wealth of variants detected by WES, bioinformatic analysis of the WES data was conducted with the goal of identifying novel candidate SHFM genes for further investigation based on the putative biologic functions of the encoded proteins. These candidate genes were then studied further using DNA sequencing and segregation analyses. In this exposition, the methods and thought processes used to systematically search for a novel cause of a rare genetic condition are discussed, with emphasis upon the value of an extended pedigree when studying a genetic condition.

MATERIALS AND METHODS

Whole Exome Sequencing

Initial WES was conducted by Beijing Genomics Institute on four affected individuals (Figure 1A). Genomic DNA from all members of family K6821 was isolated from blood samples after informed consent was obtained. The chronology of events is shown in Figure 1B to highlight that WES and an initial segregation analysis were conducted before the birth of a set of identical twins.

Bioinformatic Analysis

The WES data from four individuals in family K6821, consisting of approximately 100,000 single nucleotide variants, were imported into Ingenuity Variant Analysis (IVA). IVA was used to analyze the variants and calculate allele frequencies using the 1000 Genomes database. IVA was also utilized to predict whether the variants were likely to be pathogenic. The screened data were imported to Excel and selected for variants that were heterozygous in the three affected individuals (II-3, III-3, III-4) but not present in the single unaffected individual (II-4). Rare variants (allele frequency <1%) were then screened for nonsynonymous variants. Using these data, the Sorting Intolerant from Tolerant (SIFT) program, which sorts intolerant from tolerant amino acid substitutions; and Polyphen, a program that predicts effects of amino acid substitution, were used to predict pathogenic variants for further analyses. The data were then filtered to only include alterations identified in genes with expression in bone. From this analysis, heterozygous mutation in three bone-expressed genes was identified: *MAP1b, NEK1,* and *CHD6*.

Gene-Specific Sequencing of MAP1b, NEK1, & CHD6

Amplifications of *MAP1b* exon 4, *NEK1* exon 31, and *CHD6* exon 23 were conducted using intronic RNA primers conjugated to M13 tags (Table 1).

To amplify *MAP1b* exon 4, *NEK1* exon 31, and *CHD6* exon 23, annealing temperatures were optimized by gradient PCR. Using primers shown in Table 1, each exon was amplified by initially denaturing the sample for 5 minutes at 95°C, then conducting 35 cycles of denaturation (95°C, 30s), annealing (optimized temperature, 30s), and adding extension (72°C, 40s), with a final extension of 72°C for 5 minutes. GoTaq® DNA Polymerase and 5X Green GoTaq® Reaction Buffer (Promega, Madison, WI) were used in all DNA amplifications. Electrophoresis using 2% agarose gels was then conducted to qualitatively assess DNA amplification.

ExoSAP-IT (Affemytrex, Santa Clara, CA) enzyme cocktail was added to the amplified DNA product, and the solution was heated to 37°C to degrade the primers and dNTPs. Then the reaction mixture was heated to 80°C to denature the enzyme without reaching the T_m of the duplex DNA products. This purified sample, devoid of primers, dNTPs, and other short DNA sequences, was labeled with BigDye® (Thermo Fisher Scientific, Grand Island, NY) by first disrupting the hydrogen bonds in the duplex DNA at 96°C for 1 minute, then subjecting the samples to 25 cycles consisting of denaturation at 96°C for 10 seconds, with 5 seconds of labelling at 50°C and 1 minute 15 seconds of extension at 60°C. BigDye® Terminator v1.1 and v3.1 5X Sequencing Buffer (Thermo Fisher Scientific, Grand Island, NY) were used to stabilize the sequencing reagents. The DNA product labelled with BigDye® was purified using Qiagen DYE-EX 96-well processing plates (Quiagen, Venlo, Limbug). Capillary electrophoresis using a 3730 DNA Analyzer (ThermoFischer Scientific, Grand Island, NY) was performed.

RESULTS

WES, followed by bioinformatic analysis, identified single nucleotide polymorphisms (SNPs) in *MAP1b, NEK1,* and *CHD6* as candidate mutations causing the SHFM phenotype in family K6821. Sequencing of exon 4 of *MAP1b* identified an adenosine to guanine change, c.437 A>G, in three affected individuals: II-3, III-3, III-4 (Figure 2). This DNA variation was predicted to encode an amino acid change from asparagine to serine at residue 146 (p.A146S). Each family member with SHFM was heterozygous for the c.437G>A change except for the affected newborn twin IV-1(Figure 2). These data indicated that the missense mutation in *MAP1b* does not co-segregate with SHFM in K6821.

DNA sequencing data for *NEK1* (Figure 3) revealed that individuals II-3, III-3, and III-4 were heterozygous for the mutation at c.2975-2 A>T. Furthermore, the polymorphism was not present in individuals lacking the SHFM phenotype. Analysis of the newborn twins revealed lack of segregation of this alteration in the affected twin, IV-1, thereby demonstrating that this mutation in *NEK1* is not associated with SHFM in this family.

A third mutation identified by the bioinformatic analysis of the WES data was a

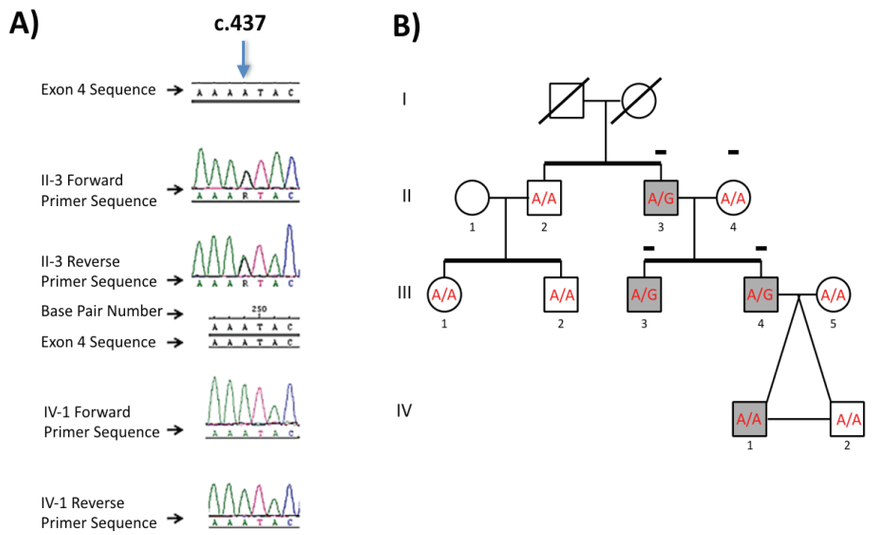

Figure 2: A) DNA sequence data for II-3 (top) and IV-1 (bottom) showing change c. 437 A>G (blue arrow). R denotes A or G. B) DNA sequence data generated for members of K6821 at c. 437 (shown in red). Individuals exhibiting the SHFM phenotype are shaded gray.

missense mutation in *CHD6*. Sequencing showed a change from cytosine to guanine, c.3675C>G. This was predicted to change the amino acid sequence at residue 1225 from histidine to glutamine (p.H1225Q) within exon 23 of *CHD6* isoform 1. Segregation analysis, shown in Figure 4, presented a potential case of variable penetrance in the family until DNA from the affected newborn twin was sequenced and found to lack the c.3675 C>G mutation. Additionally, his unaffected mother who married into the family had the alteration, thereby effectively eliminating the change in *CHD6* gene as pathogenic.

DISCUSSION

Once it became apparent that members of family K6821 lacked mutations in the known SHFM genes, whole exome sequencing (WES) was conducted on four individuals in the family. WES is a useful tool in the study of rare diseases following a Mendelian inheritance pattern because it is the most efficient means for detecting point mutations in all of an individual's genes. Once WES was complete, bioinformatic analysis was performed to analyze variants of unknown significance (VUS) detected by WES. The bioinformatic analysis systematically narrowed down the variants to form a list that could potentially explain the SHFM in family K6821. As part of this analysis, genes with expression in bone were selected, with *MAP1b*, *NEK1*, and *CHD6* hypothesized to be the most significant due to reported biologic functions of the encoded proteins. The normal chromosomal loci for *MAP1b*, *NEK1* and *CHD6*, are shown in Figure 5.

The reported biological functions of the three genes are summarized in Table 2.

A limitation of this study is that the SHFM causative gene in this family could be expressed in a different tissue other than bone. In fact, many of the mutations known to cause SHFM affect proteins expressed in the apical ectodermal ridge of a developing limb bud.[1]

After systematically searching the literature, it remained unclear as to whether *MAP1b*, *NEK1*, and *CHD6* are expressed in the AER during limb development. Furthermore, even if the causative gene in this family were expressed in bone, another limitation of this study is that genes expressed below detectable levels could have been excluded by the bioinformatic analysis. Despite the limitations of this analysis, the WES data was useful for distinguishing variants in the family's DNA for further study.

Confirmation of the variants identified by WES was also necessary, and Sanger sequencing was utilized to this accomplish this end. A possible limitation of WES is the potential for limited coverage of the sequence containing the causal variant.[12] Sanger sequencing confirmed the results of WES, and initial segregation analyses provided evidence suggesting the three variants could be associated with the SHFM phenotype. Once the twins (IV-1, IV-2, Figure 1) were born, the extension of the pedigree offered a valuable chance to see whether the mutations in *MAP1b*, *NEK1*, and *CHD6* contin-

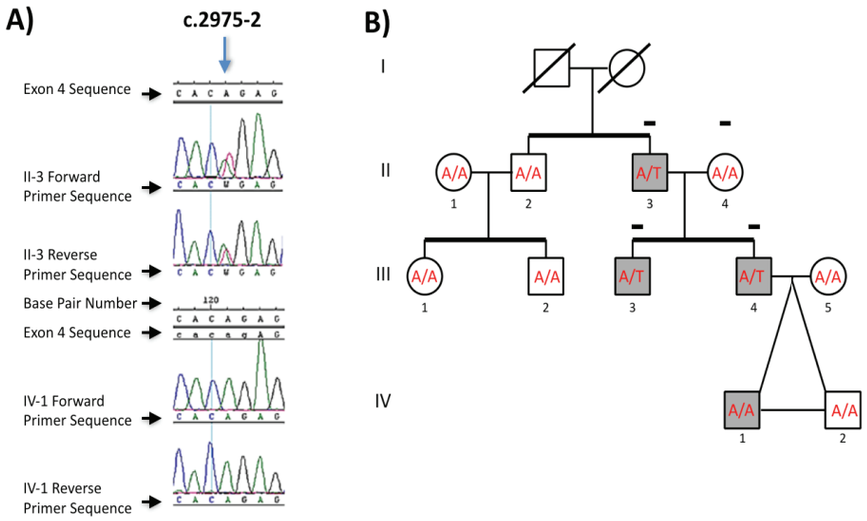

Figure 3: A) DNA sequence data for II-3 (top) and IV-1 (bottom) showing change c. 2975-2 A>T, w (blue arrow). W denotes T or A. B) DNA sequence data generated for members of K6821 at c. 2975-2 A>T (shown in red). Individuals exhibiting the SHFM phenotype are shaded gray.

Table 2: Reported functions of MAP1b, NEK1, and CHD6 proteins that were hypothesized to be significant to the etiology of SHFM.

Gene Product	Overview of Known Function
MAP1b	Regulation of cytoskeletal dynamics and outgrowth.[8]
NEK1	Regulation of mitotic and meiotic signal transduction cascades.[9,10]
CHD6	Alteration of chromatin structure to establish cell type-specific patterns of gene expression.[11]

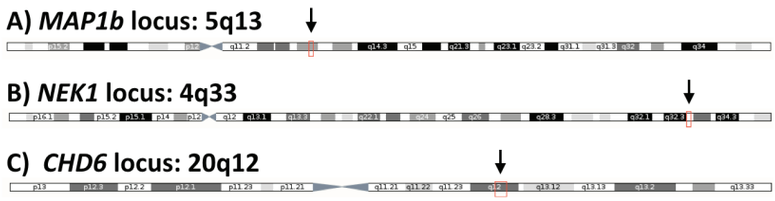

Figure 5: Red boxes, emphasized by arrows, denote the chromosomal loci of *MAP1b, NEK1*, and *CHD6* genes. Images obtained from Ensembl database[7]

ued to segregate with the SHFM phenotype. Discussion of each of these individual genes follows.

Before the birth of the twins, segregation analysis showed that the mutation in the *MAP1b* gene was present only in affected family members, which suggested this variant could underlie the SHFM phenotype. *MAP1b* is located on the forward strand of chromosome 5 at 5q13 (Figure 5). Microtubule-associated protein 1b (MAP1b) has a function which we hypothesized could be relevant to the SHFM phenotype. MAP1b regulates microtubule dynamics, which is important for axonal outgrowth and cytoskeletal growth.[13,14]

Additionally, MAP1b binds microtubules in a phosphorylation-state-dependent manner when phosphorylated at a serine/threonine kinase specific site.[8,15,16] We hypothesized that a mutation in a phosphorylated residue could alter MAP1b protein function. The missense mutation present in the MAP1b protein detected was p.N146S. This acquisition of a phosphorylatable serine residue could have enhanced the ability of MAP1b protein to phosphorylate substrates and/or bind microtubules leading to abrogation of central ray formation.

The presence of the missense mutation in the MAP1b protein in affected individuals would strengthen the gene's candidacy as a gene underlying SHFM. However, the data presented in Figure 2 show that the *MAP1b* mutation is not present in the affected newborn twin, IV-1. Therefore, the extended segregation analysis decreased the likelihood that the *MAP1b* missense mutation is a risk factor for SHFM in family K6821.

In a similar scenario, prior to the birth of the twins, segregation analysis of the mutation, c.2975-2A>T, in the *NEK1* gene showed that mutation was only present in affected family members. NEK1 protein also has a function we hypothesized could be relevant to the SHFM phenotype. Never in Mitosis Gene A related kinase-1 (NEK1) is located on the reverse strand of chromosome 4 at 4q33 (Figure 5) and is a kinase with dual-specificity, since it can phosphorylate either serine/threonine or tyrosine residues on target substrates after appropriate autophosphorylation.[17,18] This dual-specificity provides unique functions such as regulation of mitotic and meiotic signal transduction cascades.[9,10]

Recently, the NEK1 protein was demonstrated to mediate S-phase progression by assisting in the loading of replication factors onto chromatin.[19] In K6821, we hypothesized a guanine to cytosine variation two bases before a 5' splice site of intron 30 (c.2975-2 A>T) caused an alteration in a splice site between intron 30 and exon 31, ultimately altering protein structure and function. Although the mutation in *NEK1* was present in patients II-3, III-3, and III-4, it was not found in the affected newborn twin, IV-1(Figure 3B). These data suggest that the variation seen in *NEK1* is not responsible for the SHFM phenotype in family K6821.

The initial segregation analysis conducted for the c.367C>G mutation in the *CHD6* gene presented the possibility of this variant causing SHFM, but with incomplete penetrance. The *CHD6* gene is located at 20q12 on the reverse strand (Figure 5). *CHD6* encodes chromodomain helicase DNA binding protein 6 (CHD6), which alters chromatin structure to establish cell type-specific patterns of gene expression.[11] CHD6 has multiple domains, including two N-terminal chromodomains, a SWItch/Sucrose Non-Fermentable (SWI2/SNF2)-like ATPase/helicase domain, and a DNA binding do-

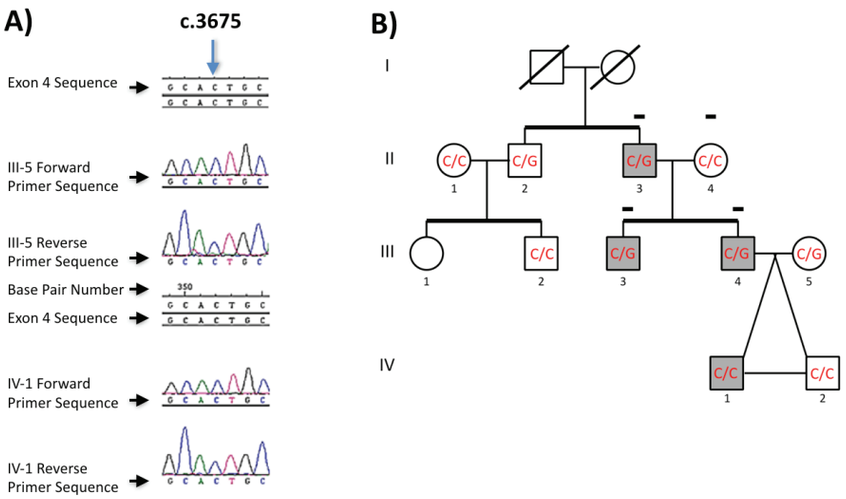

Figure 4: A) DNA sequence data for III-5(top) and IV-2 (bottom) that shows the presence of cytosine at c. 3675 (blue arrow), B) DNA sequence data generated for members of K6821 at c. 3675 (shown in red). Individuals exhibiting the SHFM phenotype are shaded gray.

main that enable the protein to influence gene expression in a cell type-dependent manner.[20,11,21] In K6821, the heterozygous mutation (c.3675G>C) was predicted to correspond to a change from histidine to glutamine at residue 1225 (p.H1225Q). This change was hypothesized to result in abrogation of regulation of cell type – specific gene expression. As shown in Figure 4, each affected patient except the affected newborn twin (IV-1) is heterozygous for the mutation. However, individuals who do not have the condition also have the mutation (e.g. individuals II-2 and III-5). Finally, the data shown in Figure 4 also shows that IV-1, an affected individual, is homozygous for the allele encoding a normal *CHD6* gene, which suggests that the locus does not play a role in the SHFM phenotype in family K6821.

While the analyses did not elucidate a novel cause of SHFM in family K6821, they do outline a scientific methodology that can be utilized and improved in future studies. Moving forward, it would be ideal to couple these methods along with a genome-wide association study of all patients with SHFM without a known genetic cause. Patients with similar phenotypes could be grouped together, and their exomes could be compared to controls. In summary, these data show that WES followed by confirmatory Sanger sequencing and segregation analysis constitutes a viable method for identifying and evaluating DNA sequence variants in families with a rare genetic disease. Whenever possible, extension of the analysis to newborn affected family members may aid in elucidation of the true etiology of the disease under investigation.

ACKNOWLEDGEMENTS

The authors extend gratitude to Sara Sarasua, PhD for conducting the bioinformatic analysis.

REFERENCES

[1]Gurrieri, F. and Everman, D. (2013) "Clinical, Genetic, and Molecular Aspects of Split-Hand/foot Malformation: An Update." *American journal of medical genetics. Part A*. 161A. 11. Pg 2860–72.

[2]Elliott, A. and Evans, J. (2006) "Genotype-Phenotype Correlations in Mapped Split Hand Foot Malformation (SHFM) Patients." *American journal of medical genetics. Part A*. 140. 13. Pg 1419–27.

[3]Zlotogora, J. (1994) "On the Inheritance of the Split Hand/split Foot Malformation." *American journal of medical genetics*. 53. 1. Pg 29–32.

[4]Birnbaum, R. et al. (2012) "Coding Exons Function as Tissue-Specific Enhancers of Nearby Genes." *Genome research*. 22. 6. Pg 1059–68.

[5]Tackels-Horne, D. et al.(2001) "Split Hand/split Foot Malformation with Hearing Loss: First Report of Families Linked to the SHFM1 Locus in 7q21." *Clinical genetics*. 59. 1. Pg 28–36.

[6]Temtamy, S. and McKusick, V. (1978) "The Genetics of Hand Malformations." *Birth defects original article series*. 14. 3, i – xviii. Pg 1–619.

[7]Aken, B. et al. (2016) "Database Update The Ensembl Gene Annotation System." Pg 1–19.

[8]Aletta, J. and Lewis, S. (1988) "Nerve Growth Factor Regulates Both the Phosphorylation and Steady-State Levels of Microtubule-Associated Protein 1.2 (MAP1. 2)." *The Journal of cell* …. 106. Pg 1573–81.

[9]Gould, K. and Nurse, P. (1989) "Tyrosine Phosphorylation of the Fission Yeast cdc2+ Protein Kinase Regulates Entry into Mitosis." *Nature*. 342. 6245. Pg 39–45.

[10]Lundgren, K. et al. (1991) "mik1 and wee1 Cooperate in the Inhibitory Tyrosine Phosphorylation of cdc2." *Cell*. 64. 6. Pg 1111–22.

[11]Schuster, E. and Stöger, R. (2002) "CHD5 Defines a New Subfamily of Chromodomain-SWI2/SNF2-like Helicases." *Mammalian genome : official journal of the International Mammalian Genome Society*. 13. 2. Pg 117–19.

[12]Bamshad, M. et al. (2011) "Exome Sequencing as a Tool for Mendelian Disease Gene Discovery." *Nature reviews. Genetics*. 12. 11. Pg 745–55.

[13]Avila, J., Dominguez, J. and Diaz-nido, J. (1994) "Regulation of Microtubule Dynamics by Microtubule- Associated Protein Expression and Phosphorylation during Neuronal Development." 25. Pg 13–25.

[14]Mandell, J. and Banker, G. (1995) "The Microtubule Cytoskeleton and the Development of Neuronal Polarity." *Neurobiology of Aging*. 16. 3. Pg 229–37.

[15]Brugg, B. and Matus, A. (1988) "PC12 Cells Express Juvenile Microtubule-Associated Proteins during Nerve Growth Factor-Induced Neurite Outgrowth." *The Journal of cell biology*. 107. Pg 643–50.

[16]Diaz-nido, J., Mendez, S. and Avila, J. (1988) "A Casein Kinase II-Related Activity Is Involved in Phosphorylation of Microtubule-Associated Protein MAP-1B during Neuroblasoma Cell Differentiation." *Journal of Cell Biology*. 106. Pg 2057–65.

[17]Letwinl, K. et al. (1992) "A Mammalian Dual Specificity Protein Kinase, Nek1, Is Related to the NIMA Cell Cycle Regulator and Highly Expressed in Meiotic Germ Cells." *The EMBO journal*. 11. 10. Pg 3521–31.

[18]Lindberg, R., Quinn, A. and Hunter, T. (1992) "Dual-Specificity Protein Kinases: Will Any Hydroxyl Do?" *Trends in Biochemical Sciences*. 17. 3. Pg 114–19.

[19]Patil, M., Pabla, N., Ding, H. and Dong, Z. (2013) "Nek1 Interacts with Ku80 to Assist Chromatin Loading of Replication Factors and S-Phase Progression." *Cell cycle*. 12. 16. Pg 2608–16.

[20]Delmas, V., Stokes, D. and Perry, R. (1993) "A Mammalian DNA-Binding Protein That Contains a Chromodomain and an SNF2/SWI2-like Helicase Domain." *Proceedings of the* …. 90. Pg 2414–18.

[21]Woodage, T. and Basrai, M. (1997) "Characterization of the CHD Family of Proteins." *Proceedings of the National Academy of Sciences*. 94. Pg 11472–77.