

**A STUDY ON COFFIN SIRIS SYNDROME – COMPUTATIONAL ANALYSIS OF  
ITS GENETIC VARIANTS**

**Zainab Khan<sup>1</sup>, Ayesha Humera<sup>2</sup>, Qasim Bee Ayesha<sup>3</sup>,  
Khadija Faiqa Ahmed<sup>4</sup>**

<sup>1</sup>Student, Microbiology dept, Shadan Degree College for Women, India

<sup>2,4</sup>Student, Genetics dept, Shadan Degree College for Women, India

<sup>3</sup>Student, Chemistry dept, Shadan Degree College for Women, India

## Table of Contents

<b>INTRODUCTION.....</b>	<b>9</b>
<b>GENETICS.....</b>	<b>9</b>
<b>BIOINFORMATICS TOOLS.....</b>	<b>12</b>
<b>UNIPORT .....</b>	<b>12</b>
<b>ENSEMBL .....</b>	<b>12</b>
<b>ALPHAFOLD .....</b>	<b>13</b>
<b>POLYPHEN 2.....</b>	<b>13</b>
<b>SIFT.....</b>	<b>14</b>
<b>MUTATION TASTER.....</b>	<b>15</b>
<b>NCBI.....</b>	<b>15</b>
<b>COFFIN SIRIS SYNDROME .....</b>	<b>16</b>
<b>GENES.....</b>	<b>21</b>
<b>ARID1A .....</b>	<b>21</b>
<b>ARID2 .....</b>	<b>21</b>
<b>SOX11 .....</b>	<b>22</b>
<b>DPF2.....</b>	<b>23</b>
<b>SMARCA4.....</b>	<b>24</b>
<b>SMARCC2.....</b>	<b>25</b>
<b>BICRA.....</b>	<b>27</b>
<b>METHODOLOGY .....</b>	<b>29</b>
<b>SOX11 .....</b>	<b>29</b>
<b>VARIANT c.178T&gt;C (p.Ser60Pro).....</b>	<b>29</b>
<b>VARIANT c.355C&gt;T (p.Arg119Trp).....</b>	<b>31</b>
<b>VARIANT c.170T&gt;C (p.Met57Thr).....</b>	<b>34</b>
<b>VARIANT c.250G&gt;C (p.Gly84Arg).....</b>	<b>36</b>
<b>DPF2.....</b>	<b>38</b>
<b>VARIANT :c.1066T&gt;G (p.Cys356Gly) .....</b>	<b>38</b>
<b>VARIANT c.1045G&gt;A (p.Asp349Asn).....</b>	<b>40</b>
<b>VARIANT c.868G&gt;A (p.Glu290Lys) .....</b>	<b>42</b>
<b>VARIANT c.827G&gt;T (p.Cys276Phe) .....</b>	<b>44</b>
<b>VARIANT c.990C&gt;G (p.Cys330Trp) .....</b>	<b>46</b>

<b>SMARCA4</b> .....	48
<b>VARIANT c.1645C&gt;T (p.Arg549Cys)</b> .....	48
<b>SMARCC2</b> .....	51
<b>VARIANT :c.1919T&gt;C (p.Leu640Pro)</b> .....	51
<b>BICRA</b> .....	54
<b>VARIANT :c.192G&gt;C (p.Glu64Asp)</b> .....	54
<b>RESULT</b> .....	57
<b>DISCUSSION</b> .....	95
<b>CONCLUSION</b> .....	96

## LIST OF FIGURES

Figure 1	11
Figure 2 The figures represent the 3D structures of the protein at position SER 60, obtained from AlphaFold.	33
Figure 3 The figures represent the 3D structures of the protein at position ARG119, obtained from AlphaFold.	35
Figure 4 The figures represent the 3D structures of the protein at position MET 57, obtained from AlphaFold.	37
Figure 5 The figures represent the 3D structures of the protein at position GLY 84, obtained from AlphaFold.	39
Figure 6 The figures represent the 3D structures of the protein at position CYS 356, obtained from AlphaFold.	41
Figure 7 The figures represent the 3D structures of the protein at position ASP 349, obtained from AlphaFold.	43
Figure 8 The figures represent the 3D structures of the protein at position GLU 290, obtained from AlphaFold.	45
Figure 9 The figures represent the 3D structures of the protein at position CYS 276, obtained from AlphaFold.	47
Figure 10 The figures represent the 3D structures of the protein at position CYS 330, obtained from AlphaFold.	49
Figure 11 The figures represent the 3D structures of the protein at position ARG 549, obtained from AlphaFold.	51
Figure 12 The figures represent the 3D structures of the protein at position LEU 640, obtained from AlphaFold.	53
Figure 13 The figures represent the 3D structures of the protein at position GLU 64, obtained from AlphaFold.	55

## ABSTRACT

Coffin-Siris Syndrome (CSS) is a rare congenital disorder characterized by developmental delays, intellectual disabilities, coarse facial features, and hypoplasia of the fifth fingernail or toe. It is a genetically heterogeneous condition primarily caused by mutations in genes associated with chromatin remodeling complexes, including **ARID1A**, **ARID1B**, **SMARCA4**, and **SMARCC2**. These genes play a crucial role in regulating gene expression and cellular differentiation. Identifying pathogenic variants in these genes is essential for understanding the molecular basis of CSS and improving diagnostic accuracy.

In this study, we conducted a comprehensive bioinformatics analysis of genetic variants associated with Coffin-Siris Syndrome. Publicly available genomic datasets were utilized to retrieve reported and novel variants in selected CSS-associated genes. Various in-silico prediction tools, including **SIFT**, **PolyPhen-2**, and **MutationTaster**, were employed to assess the pathogenicity of these variants.

Our findings highlight several potentially deleterious variants that may contribute to CSS pathogenesis. The results emphasize the importance of integrating computational approaches for variant prioritization, which can aid in genetic diagnosis and counseling. Additionally, this study provides insights into the functional consequences of specific mutations, paving the way for future experimental validation and therapeutic advancements. The application of bioinformatics tools in rare genetic disorders like CSS enhances our ability to decipher disease mechanisms and supports precision medicine efforts in clinical genetics.

## **OBJECTIVE**

- To provide an overview of Coffin-Siris Syndrome (CSS), including its clinical manifestations, genetic basis, and inheritance patterns.
- To identify and analyze genetic variants in CSS-associated genes using bioinformatics tools.
- To assess the pathogenicity of identified genetic variants using in-silico prediction tools.

# INTRODUCTION

## GENETICS

Genetics is the study of how genes and how traits are passed down from one generation to the next. genetics, study of [heredity](#) in general and of [genes](#) in particular. Genetics forms one of the central pillars of [biology](#) and overlaps with many other areas, such as agriculture, [medicine](#), and [biotechnology](#). Since the dawn of civilization, humankind has recognized the influence of heredity and applied its principles to the improvement of [cultivated](#) crops and domestic animals. A Babylonian tablet more than 6,000 years old, for example, shows [pedigrees](#) of horses and indicates possible inherited characteristics. Other old carvings show cross-pollination of [date palm](#) trees. Most of the mechanisms of heredity, however, remained a mystery until the 19th century, when genetics as a systematic [science](#) began. Genetics arose out of the identification of genes, the fundamental units responsible for heredity. Genetics may be defined as the study of [genes](#) at all levels, including the ways in which they act in the [cell](#) and the ways in which they are transmitted from parents to offspring. Modern genetics focuses on the chemical substance that genes are made of, called deoxyribonucleic acid, or [DNA](#), and the ways in which it affects the chemical reactions that [constitute](#) the living processes within the cell. Gene action depends on interaction with the [environment](#). (*Genetics / History, Biology, Timeline, & Facts / Britannica*)

Genetics as a scientific [discipline](#) stemmed from the work of [Gregor Mendel](#) in the middle of the 19th century. Mendel suspected that traits were inherited as discrete units, and, although he knew nothing of the physical or chemical nature of genes at the time, his units became the basis for the development of the present understanding of heredity. All present research in genetics can be traced back to Mendel's discovery of the laws governing the inheritance of traits. The word *genetics* was introduced in 1905 by English biologist [William Bateson](#), who was one of the discoverers of Mendel's work and who became a champion of Mendel's principles of inheritance. (*Genetics / History, Biology, Timeline, & Facts / Britannica*)

A gene is the basic physical and functional unit of heredity. Genes are made up of DNA. Some genes act as [instructions](#) to make molecules called proteins, which are needed for the body to function. However, many genes [do not code for proteins](#), instead they help control other genes. The information in DNA is encoded in genetic building blocks called base pairs. In humans,

genes vary in size from a few hundred DNA base pairs to more than 2 million base pairs. Between 1990 and 2003, an international research effort called the Human Genome Project worked to sequence all of the DNA in a human (known as the human genome). The project estimated that humans have between 20,000 and 25,000 genes that provide instructions for making proteins. Later studies sought to build on the work of the Human Genome Project and have provided additional details on the genome sequence. We now know that the human genome contains about 19,900 genes used to produce proteins. (*What Is a Gene?*)

Typically, people have two copies of each gene, one inherited from each parent. Most genes are the same in all people, but a small number of genes (less than 1 percent of the total) are slightly different between people. Forms of the same gene with small differences in their sequence of DNA bases are called alleles. These small differences contribute to each person's unique physical features. Scientists keep track of genes by giving them unique names. Because gene names can be long, genes are also assigned symbols, which are short combinations of letters (and sometimes numbers) that represent an abbreviated version of the gene name. For example, one of the genes on [chromosome 7](#) is called the cystic fibrosis transmembrane conductance regulator because variants in this gene can cause [cystic fibrosis](#). The symbol for this gene is [CFTR](#). (*What Is a Gene?*)

Genes are made up of DNA. Each chromosome contains many genes. (*What Is a Gene?*)

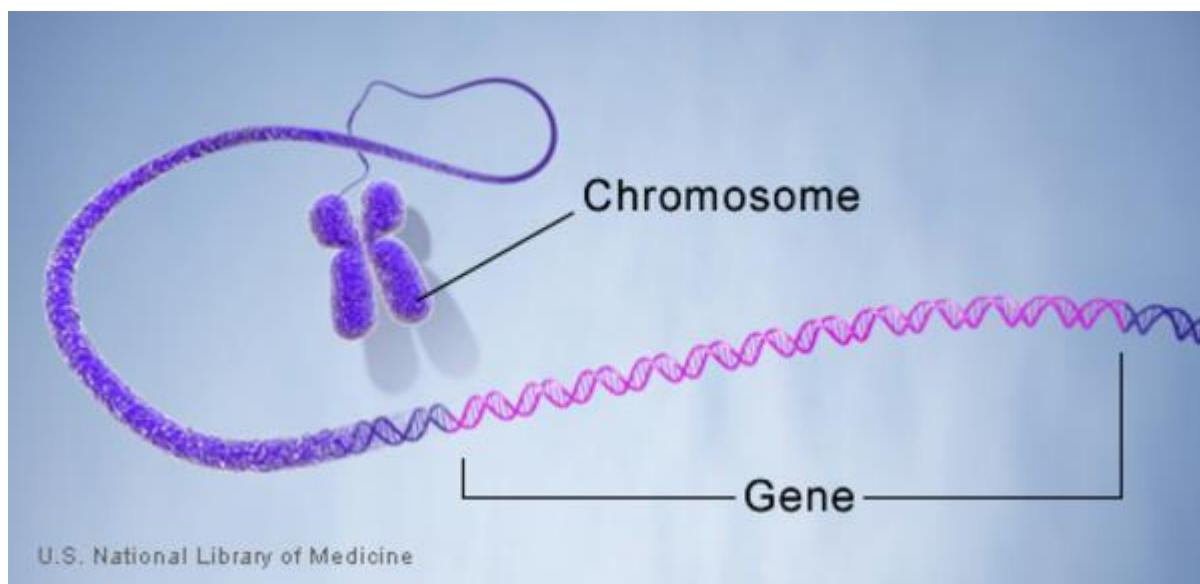


Figure 1

Humans typically have 46 chromosomes in each cell of their body, made up of 22 paired chromosomes and two sex chromosomes. These chromosomes contain between 20,000 and

25,000 genes. New genes are being identified all the time. The paired chromosomes are numbered from 1 to 22 according to size. (Chromosome number 1 is the biggest.) These non-sex chromosomes are called autosomes. People usually have two copies of each chromosome. One copy is inherited from their mother (via the egg) and the other from their father (via the sperm). A sperm and an egg each contain one set of 23 chromosomes. When the sperm fertilises the egg, two copies of each chromosome are present (and therefore two copies of each gene), and so an embryo forms. (Alliance and Screening Services)

The basic laws of inheritance are useful in understanding patterns of disease transmission. Single-gene diseases are usually inherited in one of several patterns, depending on the location of the gene (e.g., chromosomes 1-22 or X and Y) and whether one or two normal copies of the gene are needed for normal protein activity. Five basic modes of inheritance for single-gene diseases exist: autosomal dominant, autosomal recessive, X-linked dominant, X-linked recessive, and mitochondria. (Alliance and Screening Services)

Changes in the DNA sequence of single genes, also known as mutations, cause thousands of diseases. A gene can mutate in many ways, resulting in an altered protein product that is unable to perform its normal function. The most common gene mutation involves a change or “misspelling” in a single base in the DNA. Other mutations include the loss (deletion) or gain (duplication or insertion) of a single or multiple base(s). The altered protein product may still retain some normal function, but at a reduced capacity. In other cases, the protein may be totally disabled by the mutation or gain an entirely new, but damaging, function. The outcome of a particular mutation depends not only on how it alters a protein’s function, but also on how vital that particular protein is to survival. Other mutations, called polymorphisms, are natural variations in DNA sequence that have no adverse effects and are simply differences among individuals. (Alliance and Screening Services)

Rare diseases are characterised by a wide diversity of symptoms and signs that vary not only from disease to disease but also from patient to patient suffering from the same disease. (“What Is a Rare Disease?”).

Genes and variants of interest in rare diseases often benefit from modelling in cellular assays or genetic models to aid in understanding molecular and cellular mechanisms of dysfunction. Model organisms are useful for the discovery of new genetic diseases and key to understanding variant effects, and modelling a disease gene in a genetic model means that researchers can perform an in-depth exploration of gene or variant function. These studies can pinpoint disease

mechanisms, reveal unanticipated gene functions, or elucidate specific existing pathways. Unravelling the mechanism(s) by which variants act can highlight potential therapeutic strategies to help patients <https://academic.oup.com/gsajournals/pages/genetic-models-of-rare-diseases>

## BIOINFORMATICS TOOLS

Bioinformatics, a hybrid science that links biological data with techniques for information storage, distribution, and analysis to support multiple areas of scientific research, including biomedicine. Bioinformatics is fed by high-throughput data-generating experiments, including genomic sequence determinations and measurements of gene expression patterns. Database projects curate and annotate the data and then distribute it via the World Wide Web. Mining these data leads to scientific discoveries and to the identification of new clinical applications. In the field of medicine in particular, a number of important applications for bioinformatics have been discovered. For example, it is used to identify correlations between gene sequences and diseases, to predict protein structures from amino acid sequences, to aid in the design of novel drugs, and to tailor treatments to individual patients based on their DNA sequences (pharmacogenomics). (*Bioinformatics / Genomics, Proteomics & Data Analysis / Britannica*)

## UNIPORT

The UniProt databases enable the research community to explore the diversity of life as described by the complement of proteins expressed by each organism. The UniProt Knowledgebase (UniProtKB) comprises of the reviewed protein set (UniProtKB/Swiss-Prot), where each protein entry is linked to a summary of the experimentally verified, or computationally predicted, functional information added by our expert biocuration team, and the unreviewed(UniProtKB/TrEMBL), in which entries are computationally annotated by automated systems. The UniRef databases cluster sequence sets at various levels of sequence identity and the UniProt Archive (UniParc) delivers a complete set of known unique sequences, including historical obsolete sequences. Data from selected resources are additionally integrated into UniProtKB records to add biological knowledge and associated metadata enabling the database to act as a central hub from which users can link out to 183 other resources. Community functional annotation adds further value to the entry annotations. The integration of these data and the manual curation of protein features, such as functional domains and active sites, amino acid variants, ligand binding sites and post-translational modifications (PTMs) in the UniProt record, provide our users with mechanistic insights into how .  
<https://academic.oup.com/nar/article/51/D1/D523/6835362>

## **ENSEMBL**

Ensembl' is a joint' project' between' the' EBI' (European' Bioinformatics' Institute)' and' the' Wellcome Trust' Sanger' Institute . Ensembl provides' genes' and' other' annotation such' as' regulatory' regions,' conserved' base' pairs' across' species,' and' sequence' variations. Ensembl accelerates worldwide genomic research by integrating, harmonizing and annotating genome data and disseminating it via a coherent and consistent set of interfaces and tools. We import primary data from archive resources such as INSDC (1), dbSNP (2) and the European Variation Archive (EVA, <https://www.ebi.ac.uk/eva>), and add value via detailed and comprehensive annotation of transcript structures (3), genomic variants (4) and regulatory regions (5). We also enable the study of evolution by large-scale comparison of genomes and gene products across many species (6). These data can be accessed via our website, programmatically via a number of application programming interfaces (APIs) (7,8), and downloaded in numerous standard file formats. We develop and make available a variety of tools for genomic analysis, including the Ensembl Variant Effect Predictor (VEP) (9). Our software, database and tools infrastructure is freely available and is used to power the nonvertebrate genome resources provided by the clade-specific Ensembl Genomes websites (10), collaborating resources such as WormBase (11), and community-oriented databases focused on branches of the taxonomy, such as AvianBase (12) and LepBase (<http://lepbase.org>)

<https://academic.oup.com/nar/article/49/D1/D884/5952199>

## **ALPHAFOLD**

AlphaFold2 is a multicomponent artificial intelligence (AI) system that uses machine learning to predict a protein's 3D structure based on its primary amino acid sequence. Proteins are essential to life, and understanding their structure can facilitate a mechanistic understanding of their function. Through an enormous experimental effort<sup>1,2,3,4</sup>, the structures of around 100,000 unique proteins have been determined<sup>5</sup>, but this represents a small fraction of the billions of known protein sequences<sup>6,7</sup>. Structural coverage is bottlenecked by the months to years of painstaking effort required to determine a single protein structure. Accurate computational approaches are needed to address this gap and to enable large-scale structural bioinformatics. Predicting the three-dimensional structure that a protein will adopt based solely on its amino acid sequence—the structure prediction component of the ‘protein folding problem’<sup>8</sup>—has been an important open research problem for more than 50 years. AlphaFold is not a homology modelling tool: it can successfully operate without using any template structures and even predict previously unknown protein folds. (EMBL-EBI)

## POLYPHEN 2

PolyPhen-2 (Polymorphism Phenotyping v2) is a tool which predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations. (*PolyPhen-2: Prediction of Functional Effects of Human nsSNPs*) . PolyPhen-2 (Polymorphism Phenotyping v2), available as software and via a Web server, predicts the possible impact of amino acid substitutions on the stability and function of human proteins using structural and comparative evolutionary considerations. It performs functional annotation of single-nucleotide polymorphisms (SNPs), maps coding SNPs to gene transcripts, extracts protein sequence annotations and structural attributes, and builds conservation profiles. It then estimates the probability of the missense mutation being damaging based on a combination of all these properties. PolyPhen-2 features include a high-quality multiple protein sequence alignment pipeline and a prediction method employing machine-learning classification. The software also integrates the UCSC Genome Browser's human genome annotations and MultiZ multiple alignments of vertebrate genomes with the human genome. PolyPhen-2 is capable of analyzing large volumes of data produced by next-generation sequencing projects, thanks to built-in support for high-performance computing environments like Grid Engine and Platform LSF. (Adzhubei et al.)

## SIFT

Sorting Intolerant from Tolerant (SIFT) is an algorithm that predicts the potential impact of amino acid substitutions on protein function. We have recently extended SIFT to predict on frameshifting indels (6). For amino acid substitutions, SIFT has been used actively in human genetic research (7–9) [e.g. cancer (10,11) Mendelian diseases (12) and infectious diseases (13)]. We emphasize that SIFT's utility extends beyond research on humans and human disease studies. SIFT has been used to study the effects of missense mutations on agricultural plants (14,15), and model organisms like rats (16,17), canines (18) and Arabidopsis (19). In general, SIFT is useful in cases where research work involves filtering through a plethora of SNVs and indels to identify causal variants. An individual's genome contains approximately 3.7 million single nucleotide variants (SNVs) which can be identified by whole-genome sequencing (1). The challenge for geneticists is to identify what are the causal variants for the phenotype or disease being studied. The majority of SNVs found in a human are common among the population, but disease-causing variants are typically private or rare, and tend to occur in protein coding regions which constitute only 1% (30 megabases) of the total genome (2,3). Databases like dbSNP (4) and 1000 Genomes (5) are useful for filtering out common variants,

but the remaining variants need to be sorted and prioritized to identify those that may potentially affect protein function. Algorithms like SIFT can help in this respect. <https://academic.oup.com/nar/article/40/W1/W452/1751364> SIFT predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids. SIFT can be applied to naturally occurring nonsynonymous polymorphisms and laboratory-induced missense mutations. (*SIFT - Predict Effects of Nonsynonomous / Missense Variants*)

## MUTATION TASTER

MutationTaster, our DNA variant effect prediction tool. The new version uses a different prediction model and attains higher accuracy than its predecessor, especially for rare benign variants. In addition, we have integrated many sources of data that only became available after the last release (such as gnomAD and ExAC pLI scores) and changed the splice site prediction model. To more easily assess the relevance of detected known disease mutations to the clinical phenotype of the patient, MutationTaster now provides information on the diseases they cause. Further changes represent a major overhaul of the interfaces to increase user-friendliness whilst many changes under the hood have been designed to accelerate the processing of uploaded VCF files. We also offer an API for the rapid automated query of smaller numbers of variants from within other software. MutationTaster2021 integrates our disease mutation search engine, MutationDistiller, to prioritise variants from VCF files using the patient's clinical phenotype. (Steinhaus et al.)

## NCBI

The National Center for Biotechnology Information (NCBI) formed in 1988 as a division of the National Library of Medicine (NLM) at the National Institutes of Health (NIH). Among other responsibilities, the NCBI facilitates the use of databases and software and performs research on advanced methods of computer-based information processing for analyzing the structure and function of biologically important molecules including proteins. (*National Center for Biotechnology Information - an Overview / ScienceDirect Topics*) The National Center for Biotechnology Information (NCBI) at the National Institutes of Health was created in 1988 to develop information systems for molecular biology. In addition to maintaining the GenBank® (1) nucleic acid sequence database, which receives data through an international collaboration with the DNA Data Bank of Japan (DDBJ) and the European Nucleotide Archive (ENA) as well as from the scientific community, NCBI provides many other kinds of biological data as

well as retrieval systems and computational resources for the analysis of GenBank and other data. This article provides a summary of recent developments, including both new and updated resources, followed by an introduction to the Entrez system and a brief review of the suite of NCBI resources. All resources discussed are available through the NCBI home page at [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov) and can also be located using the NCBI Web Site database available in Entrez search menus. In most cases, the data underlying these resources and executables for the software described are available for download at [ftp.ncbi.nlm.nih.gov](http://ftp.ncbi.nlm.nih.gov).  
<https://academic.oup.com/nar/article/44/D1/D7/2503096>

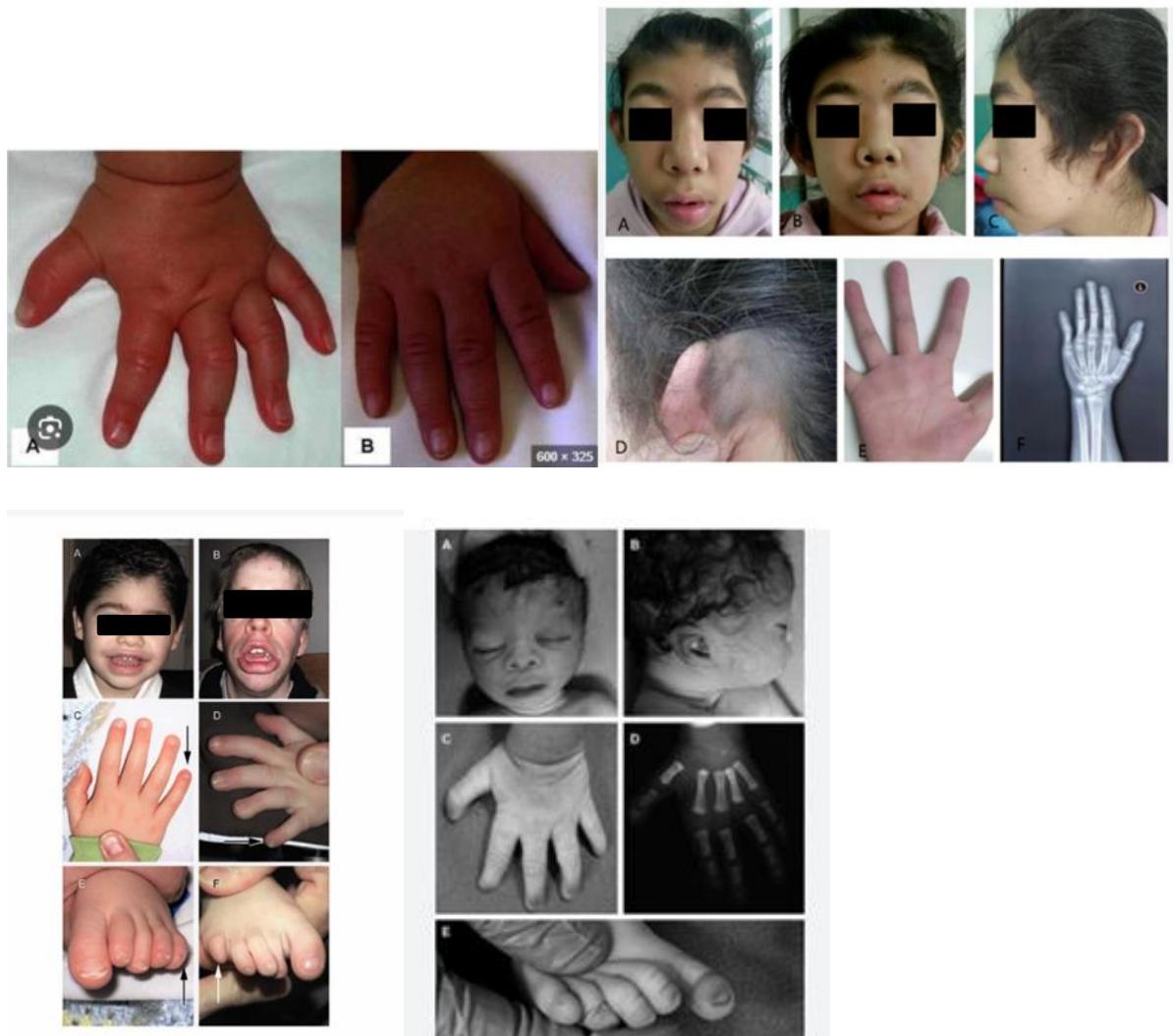
## COFFIN SIRIS SYNDROME

Coffin-Siris syndrome is a genetic condition that causes variable degrees of learning disability, developmental delays, underdeveloped 'pinky' toenails or fingernails, and distinct facial features. It can be caused by a change in any of several genes including the ARID1A, ARID1B, SMARCA4, SMARCB1, DPF2 or SMARCE1 genes. Coffin-Siris syndrome follows an autosomal dominant pattern of inheritance, however it usually occurs for the first time in a family due to a new genetic change. (*Coffin-Siris Syndrome*)

Coffin-Siris syndrome is a condition that affects several body systems. Although there are many variable signs and symptoms, hallmarks of this condition include developmental disability, abnormalities of the fifth (pinky) fingers or toes, and characteristic facial features. Most affected individuals have mild to severe intellectual disability or delayed development of speech and motor skills such as sitting and walking. Another feature of Coffin-Siris syndrome is underdevelopment (hypoplasia) of the tips of the fingers or toes, or hypoplasia or absence of the nails. These abnormalities are most common on the fifth fingers or toes. (*Coffin-Siris Syndrome*)

In addition, most people with Coffin-Siris syndrome have facial features described as coarse. These features typically include a wide nose with a flat nasal bridge, a wide mouth with thick lips, and thick eyebrows and eyelashes. Affected individuals can have excess hair on other parts of the face and body (hirsutism), but scalp hair is often sparse. People with Coffin-Siris syndrome can have a range of facial features, and not all affected individuals have the typical features. In addition, people with this condition may have an abnormally small head (microcephaly). (*Coffin-Siris Syndrome*)

Additionally, some infants and children with Coffin-Siris syndrome have frequent respiratory infections, difficulty feeding, and an inability to gain weight at the expected rate (failure to thrive). Other signs and symptoms that may occur in people with this condition include short stature, low muscle tone (hypotonia), and abnormally loose (lax) joints. Abnormalities of the eyes, brain, heart, and kidneys may also be present. (*Coffin-Siris Syndrome*)



#### ❖ PHENOTYPES OF PATIENTS WITH COFFIN SIRIS SYNDROME

### Causes

Coffin-Siris syndrome is caused by variants (also known as mutations) in one of several genes. Variants in the ARID1B gene are the most common known cause of the condition. Variants in the ARID1A, SMARCA4, SMARCB1, SMARCE1, or SOX11 gene can also cause the condition. In addition, variants in a few other genes (listed below) have been found to each cause a very small number of cases. In some of these cases, it is unclear if the condition should be considered Coffin-Siris syndrome or a similar but separate disorder. In many cases of Coffin-Siris syndrome, the genetic cause is unknown.

The above genes are involved in controlling the activity (expression) of other genes. The ARID1A, ARID1B, SMARCA4, SMARCB1, and SMARCE1 genes, as well as some of the genes involved in rare cases of Coffin-Siris syndrome, provide instructions for making single

pieces (subunits) of several different SWI/SNF protein complexes. SWI/SNF complexes regulate gene expression by a process known as chromatin remodeling .

Although it is unclear what effect variants in the ARID1A, ARID1B, SMARCA4, SMARCB1, or SMARCE1 gene have on SWI/SNF complexes, researchers suggest that the variants result in abnormal chromatin remodeling. Disturbance of this process alters the activity of many genes and disrupts several cellular processes, which could explain the diverse signs and symptoms of Coffin-Siris syndrome. This disease is caused by a change in the genetic material (DNA).

[https://medlineplus.gov/download/genetics/condition/coffin-siris-syndrome.pdf?utm\\_source=chatgpt.com](https://medlineplus.gov/download/genetics/condition/coffin-siris-syndrome.pdf?utm_source=chatgpt.com)

Coffin-Siris syndrome is caused by genetic mutations, also known as pathogenic variants. Genetic mutations can be hereditary, when parents pass them down to their children, or they may occur randomly when cells are dividing. Genetic mutations may also result from contracted viruses, environmental factors, such as UV radiation from sunlight exposure, or a combination of any of these. (Committee on Diagnostic Error in Health Care et al.)

## Inheritance

Coffin-Siris syndrome appears to follow an autosomal dominant pattern of inheritance, which means one copy of the altered gene in each cell is sufficient to cause the disorder. However, the condition is not usually inherited from an affected parent , but occurs from new (de novo) variants in the gene that likely occur during early embryonic development.

[https://medlineplus.gov/download/genetics/condition/coffin-siris-syndrome.pdf?utm\\_source=chatgpt.com](https://medlineplus.gov/download/genetics/condition/coffin-siris-syndrome.pdf?utm_source=chatgpt.com)

Autosomal dominant inheritance with a new (de novo) mutation.

## Symptoms

May start to appear as a Newborn. A 2021 study reports that most individuals with Coffin-Siris syndrome experience language delays. The researchers associate the condition with autism spectrum disorder and attention deficit hyperactivity disorder. frequent symptoms of Coffin-Siris syndrome

- coarse facial features
- a thick lower lip
- a wide mouth

- increased bodily hair
- feeding difficulties

Other common symptoms include:

- heart abnormalities
- urogenital abnormalities
- recurrent infections
- delayed skeletal maturation and growth
- underdevelopment of the fifth toe or finger or the fifth toenails or fingernails
- loose joints
- low muscle tone
- sparse scalp hair
- hearing difficulties
- strabismus, when the eyes point in different directions
- seizures
- behavioural abnormalities (*Coffin-Siris Syndrome*)

## **Diagnosis**

Diagnostic teams for coffin siris syndrome may include

- GENETICS
- NEUROLOGY
- ORTHOPEDICS

CSS should be suspected in newborns with underdeveloped nails and short fifth fingers and distinctive facial features. The facial features may become more apparent as the child grows. A diagnosis is based upon a thorough clinical evaluation and characteristic physical findings. However, physical features of CSS may be more variable as more individuals are diagnosed. Specialized testing may be conducted to detect certain findings that may be associated with the disorder. Diagnostic criteria were proposed in 2012 noting that most affected individuals have short fifth fingers with absent or underdeveloped nails, developmental and/or cognitive delays, and facial features such as a wide mouth and broad nose. Given the recent discovery of the genetic mutations causing CSS, diagnostic criteria will likely evolve to include clinical evaluations and molecular testing.

Diagnosis of this condition currently relies on a combination of clinical features, including major and minor signs observed in the patient, as there are no definitive diagnostic criteria; however, the most accurate confirmation comes from molecular genetic testing of the responsible genes, particularly looking for microdeletions within the ARID18 gene, despite recent findings indicating that absence of the fifth fingernail/distal phalanx is not a necessary diagnostic marker (*Coffin-Siris Syndrome - Symptoms, Causes, Treatment / NORD*)

## Treatment

Treatment for Coffin-Siris Syndrome (CSS) is symptom-based and may involve therapy, surgery, medications, and educational services. Various therapies play a crucial role in managing the condition, including physical therapy to enhance movement and strength, occupational therapy to improve motor skills and overall development, speech therapy to address communication challenges, and feeding therapy to ensure proper nutrition.

Surgical interventions may be necessary to correct craniofacial, skeletal, cardiac, or other structural abnormalities, including eye surgeries when required. Medications are often prescribed to manage seizures or other associated health conditions. Educational services are also essential in addressing the learning needs of children as they grow.

Additional interventions include genetic counseling to support families, regular vision and hearing tests, and follow-ups with gastroenterologists and feeding specialists to monitor and manage feeding difficulties. Ongoing ophthalmologic and audiologic evaluations, nutritional supplementation, and gastrostomy tube placement may also be required to support the individual's health.

Comprehensive care involves regular follow-ups with hospital and community pediatricians, as well as a multidisciplinary team of specialists, including neurologists, cardiologists, and orthopedists, to address the diverse medical needs associated with CSS. Since this rare genetic disorder affects multiple body systems, a coordinated approach is essential to improving the quality of life for those affected.

## Management

After a diagnosis of Coffin-Siris Syndrome (CSS), several evaluations are recommended to determine the extent of the disease and the individual's specific needs. These include consultations with a clinical geneticist or genetic counselor, as well as neurologic and

developmental assessments to track milestones and identify any deficits. Occupational, speech, and physical therapy evaluations may be necessary, along with gastrointestinal and dietary assessments for feeding difficulties or poor growth. Additionally, ophthalmologic and audiology evaluations help detect vision and hearing issues, while echocardiograms and renal ultrasounds assess for potential cardiac or kidney abnormalities.

Treatment focuses on managing symptoms and improving quality of life. Developmental therapies, including occupational, physical, and speech therapy, are essential for optimizing progress. Nutritional support, such as feeding therapy or gastrostomy tube placement, may be needed for those with significant feeding difficulties. Vision and hearing impairments can be addressed with glasses, surgery for conditions like strabismus, or hearing aids if required.

Preventing complications involves ongoing medical surveillance tailored to the individual's needs. Regular developmental, cardiac, gastrointestinal, and neurological evaluations help manage potential issues. Continuous monitoring of vision and hearing ensures early intervention when necessary.

Surveillance recommendations include annual assessments by a developmental pediatrician to track progress and adjust therapies. Gastrointestinal follow-ups help manage feeding and weight gain concerns, while routine ophthalmologic and audiologic monitoring ensures early detection of abnormalities. Due to the rarity of tumors in CSS, routine tumor surveillance is not currently recommended.

Other considerations include genetic counseling for at-risk relatives, which can help guide family planning and testing decisions. Since no females with CSS have been reported to reproduce, pregnancy-related risks remain unknown. While there are currently no established treatments for CSS, individuals can explore potential clinical trials for emerging therapies.

# **GENES**

We Found Multiple Genes Associated With COFFIN SIRIS SYNDROME .

## **ARID1A**

This gene encodes a member of the SWI/SNF family, whose members have helicase and ATPase activities and are thought to regulate transcription of certain genes by altering the chromatin structure around those genes. The encoded protein is part of the large ATP-dependent chromatin remodeling complex SNF/SWI, which is required for transcriptional activation of genes normally repressed by chromatin. It possesses at least two conserved domains that could be important for its function. First, it has a DNA-binding domain that can specifically bind an AT-rich DNA sequence known to be recognized by a SNF/SWI complex at the beta-globin locus. Second, the C-terminus of the protein can stimulate glucocorticoid receptor-dependent transcriptional activation. It is thought that the protein encoded by this gene confers specificity to the SNF/SWI complex and may recruit the complex to its targets through either protein-DNA or protein-protein interactions. Two transcript variants encoding different isoforms have been found for this gene. (*ARID1A AT-Rich Interaction Domain 1A [Homo Sapiens (Human)] - Gene - NCBI*)

## **ARID2**

This gene encodes a member of the AT-rich interactive domain (ARID)-containing family of DNA-binding proteins. Members of the ARID family have roles in embryonic patterning, cell lineage gene regulation, cell cycle control, transcriptional regulation and chromatin structure modification. This protein functions as a subunit of the polybromo- and BRG1-associated factor or PBAF (SWI/SNF-B) chromatin remodeling complex which facilitates ligand-dependent transcriptional activation by nuclear receptors. Mutations in this gene are associated with hepatocellular carcinomas. A pseudogene of this gene is found on chromosome1. (*ARID2 AT-Rich Interaction Domain 2 [Homo Sapiens (Human)] - Gene - NCBI*)

- These Genes Are Strongly Associated WITH COFFIN SIRIS SYNDROME But Has No Variants .

## **SOX11**

The SOX11 gene provides instructions for making a protein that plays a critical role in the development of the brain and nerve cells (neurons). This protein is a transcription factor, which means it attaches (binds) to specific regions of DNA and coordinates with other proteins to turn on particular genes.

Studies suggest that the SOX11 protein may also be able to regulate the activity of genes involved in brain and nerve cell development through a different process known as chromatin remodeling. Chromatin is the network of DNA and protein that packages DNA into chromosomes. The structure of chromatin can be changed (remodeled) to alter how tightly regions of DNA are packaged. Chromatin remodeling is one way gene expression is regulated during development; when DNA is tightly packed, gene expression is often lower than when DNA is loosely packed.

The activity (expression) of the SOX11 gene is controlled by chromatin remodeling by special protein groups called SWI/SNF complexes. (*SOX11 Gene*)

Proteins encoded by SOX genes are a family of transcription factors, which play crucial roles in multiple developmental processes.<sup>8</sup> All SOX protein family members contain an HMG box, which is the hallmark of these proteins. The HMG box binds to and regulates target genes. The HMG box also controls protein–protein interactions and trafficking of SOX proteins between cytoplasm and nucleus. Variants in a variety of SOX genes are associated with human developmental disorders (termed SOXopathies); examples include Waardenburg syndrome<sup>9</sup> (caused by pathogenic variants in SOX10) and SOX2-anophthalmia syndrome.<sup>10</sup> SOXopathies share some common features such as ocular malformations, ID, hypogonadotropic hypogonadism, and genital malformations. SOX11 forms a peripheral component of the SWI/SNF complex, and thus, SOX11-associated syndrome (SOX11 syndrome) may lie in the CSS spectrum. However, given the similarity in protein sequence and function between SOX11 and other SOX gene family members, it is possible that the phenotypes associated with SOX11 variants may be more congruent with SOXopathies.

## **DPF2**

"BAF (SWI-SNF) complexes are essential for controlling gene expression. They modify chromatin, allowing transcription factors (TFs) to access DNA. These complexes are built around an ATPase (BRG1 or BRM) and vary in composition, leading to three main types:

cBAF, PBAF, and ncBAF. cBAF primarily works at distant gene regulatory regions (enhancers), while ncBAF targets gene promoters and CTCF sites. Subunits from all three BAF complexes, such as BAF53a, BAF45A, ARID1A, ARID2, BRM-SMARCA2, and BAF180, are known to be important for hematopoietic stem cell (HSC) function and immune responses. However, it remains unclear whether these complexes also play a role in emergency blood cell production or in preventing chronic inflammatory diseases.

DPF2/BAF45D is a defining subunit of the cBAF complex in hematopoietic cells. Loss-of-function *Dpf2* mutations are found in cancer and in patients with Coffin-Siris syndrome. Work from our laboratory and others has shown that DPF2 regulates myelopoiesis. DPF2 also interacts with NF-κB to control immune response genes in cancer cell lines. Here, we report that hematopoiesis-specific *Dpf2*-KO mice developed a lethal inflammatory disease involving dysfunctional HSCs, macrophages, and Th cells — phenotypes that mirror NRF2 deficiency. Mechanistically, NRF2 binding to active enhancers in HSCs depends on DPF2, and pharmacological reactivation of NRF2 overcomes the inflammatory defects driven by DPF2 loss, thereby prolonging survival. Our work uncovers the multilineage control of inflammation by DPF2, mediated by NRF2, establishing a role for the BAF complex in modulating inflammation.

The protein encoded by this gene is a member of the d4 domain family, characterized by a zinc finger-like structural motif. This protein functions as a transcription factor which is necessary for the apoptotic response following deprivation of survival factors. It likely serves a regulatory role in rapid hematopoietic cell growth and turnover. This gene is considered a candidate gene for multiple endocrine neoplasia type I, an inherited cancer syndrome involving multiple parathyroid, enteropancreatic, and pituitary tumors.

## **SMARCA4**

One piece (subunit) of several different protein groupings called SWI/SNF protein complexes. SWI/SNF complexes regulate gene activity (expression) by a process known as chromatin remodeling. Chromatin is the network of DNA and protein that packages DNA into chromosomes. The structure of chromatin can be changed (remodeled) to alter how tightly DNA is packaged. Chromatin remodeling is one way gene expression is regulated during development; when DNA is tightly packed, gene expression is lower than when DNA is loosely packed. The BRG1 protein uses a molecule called ATP to provide energy for chromatin remodeling, although the protein's specific role in remodeling is unclear.

Through their ability to regulate gene activity, SWI/SNF complexes are involved in many processes, including repairing damaged DNA; copying (replicating) DNA; and controlling

the growth, division, and maturation (differentiation) of cells. Through these processes, the BRG1 protein and other SWI/SNF subunits are thought to act as tumor suppressors, which keep cells from growing and dividing too rapidly or in an uncontrolled way. The SMARCA4 gene provides instructions for making a protein called BRG1, which forms. (*SMARCA4 Gene*)

ATP-dependent helicase SMARCA4, also known as BAF190, brahma protein-like 1, BRG1, BRG1-associated factor 190A, BRM/SWI2-related gene 1, protein BRG-1, RTPS2, SMCA4\_HUMAN, SNF2, and SWI2, is a crucial protein involved in chromatin remodeling and gene regulation.

## **SMARCC2**

This protein functions as a critical regulator of gene expression, exerting its influence through the intricate process of chromatin remodeling. Chromatin, the complex of DNA and proteins that comprises chromosomes, is not static; its structure can be dynamically altered to control which genes are accessible for transcription. This protein is a key component of the SWI/SNF chromatin remodeling complexes, which are molecular machines that precisely manipulate chromatin architecture.

Specifically, these complexes utilize the energy derived from ATP hydrolysis to physically alter the interactions between DNA and histone proteins within nucleosomes, the fundamental repeating units of chromatin. By disrupting or modifying these interactions, the SWI/SNF complexes can either enhance or repress the binding of transcription factors and other regulatory proteins to DNA, thereby controlling gene activity. This protein, in particular, has been shown to stimulate the ATPase activity of the catalytic subunit within these complexes, further amplifying their chromatin remodeling capabilities.

Beyond its general role in transcriptional regulation, this protein also participates in specialized regulatory mechanisms. For instance, it may contribute to the CoREST-dependent repression of neuronal-specific gene promoters in non-neuronal cells. This suggests that it plays a role in establishing and maintaining cell-type-specific gene expression patterns, preventing the inappropriate activation of genes in cells where they are not needed.

During neural development, this protein is involved in the dynamic transition from neural progenitor cells to mature neurons. This process involves a switch in the composition of chromatin remodeling complexes. In proliferating neural stem/progenitor cells, the protein is found in the neural progenitor-specific BAF (npBAF) complex, which contains subunits such

as ACTL6A/BAF53A and PHF10/BAF45A. As these progenitors exit the cell cycle and differentiate into neurons, the npBAF complex is replaced by the neuron-specific BAF (nBAF) complex, which contains homologous alternative subunits such as ACTL6B/BAF53B and DPF1/BAF45B or DPF3/BAF45C. This switch in subunit composition reflects a shift in the functional properties of the complexes, with the npBAF complex being essential for the self-renewal and proliferative capacity of neural stem cells, while the nBAF complex, in cooperation with CREST, plays a role in regulating genes crucial for dendrite growth and neuronal maturation.

Furthermore, this protein is a key determinant of myeloid differentiation, the process by which multipotent hematopoietic stem cells give rise to various types of blood cells. Specifically, it exerts control over granulocytopoiesis, the formation of granulocytes, and the expression of genes required for the assembly of neutrophil granules, which are essential for the antimicrobial functions of neutrophils. This highlights its role in the development and function of the innate immune system.

## BICRA

SWI/SNF-related intellectual disability disorders (SSRIDDs) are rare neurodevelopmental conditions characterized by developmental delays, distinct facial features, and underdevelopment of the fifth fingers or nails. These disorders arise from harmful genetic changes (pathogenic variants) in genes that encode for proteins belonging to the SWI/SNF (also known as BAF) family, a group of complexes that remodel chromatin.

Rare genetic variations within the BICRA gene, also known as GLTSCR1, have been identified in individuals with neurodevelopmental phenotypes. Ten of these variations resulted in a loss of function, meaning the BICRA protein could no longer work correctly, while two were missense mutations, altering the protein's structure. BICRA encodes a subunit of the non-canonical BAF (ncBAF) complex, a specific type of chromatin remodeling complex.

These individuals displayed a range of neurodevelopmental symptoms, including delays in development, intellectual disability, autism spectrum disorder, and behavioral abnormalities. They also exhibited distinct facial features (dysmorphic features). Notably, a significant difference from typical SSRIDDs was that most of these individuals did not show hypoplasia (underdevelopment) of the fifth fingers or nails, a common characteristic of other SSRIDDs.

To validate BICRA's role in these observed phenotypes, functional studies using zebrafish and Drosophila (fruit flies), which have corresponding genes (orthologs) to the human BICRA gene, were conducted. In zebrafish, a mutation in the *bicra* gene that mirrored one of the loss-of-function variants found in humans led to craniofacial defects, suggesting a link to the dysmorphic facial features observed in humans with BICRA variations.

Furthermore, it has been demonstrated that the *Bicra* protein physically interacts with other members of the non-canonical ncBAF complex, including the fly ortholog of BRD9/7, known as CG7154. This confirms that *Bicra* is a key defining component of the ncBAF complex in flies.

Consistent with the known functions of other SWI/SNF complex members, loss of *Bicra* function in flies acted as a dominant enhancer of position effect variegation (PEV). PEV is a phenomenon where gene expression is altered depending on the gene's position in the chromosome. However, in this case, the effect of *Bicra* loss was more context-specific, indicating a nuanced role in gene regulation.

In conclusion, haploinsufficiency (having only one functional copy of the gene) of BICRA leads to a distinct form of SSRIDD in humans. This disorder shares some features with previously described SSRIDDs, but also exhibits unique characteristics, particularly the absence of fifth digit/nail hypoplasia in most cases. This highlights the diverse roles of SWI/SNF complex subunits in neurodevelopment and the importance of BICRA in proper brain function." (Barish et al.)

# **METHODOLOGY**

## **IN-SILICO ANALYSIS**

### **THE SEQUENCES**

The transcript sequence and the protein sequence are retrieved using the database Ensembl. Specifically for the protein sequence the protein database UniProt is used. The protein structure is retrieved using the AlphaFold.

### **SOX11**

The transcript sequence are downloaded from Ensembl. The protein sequence is retrieved from the UniProt. The location of the mutation is highlighted in the whole sequence or the sequence snippets shown below.

# **SOX11**

## **VARIANT c.178T>C (p.Ser60Pro)**

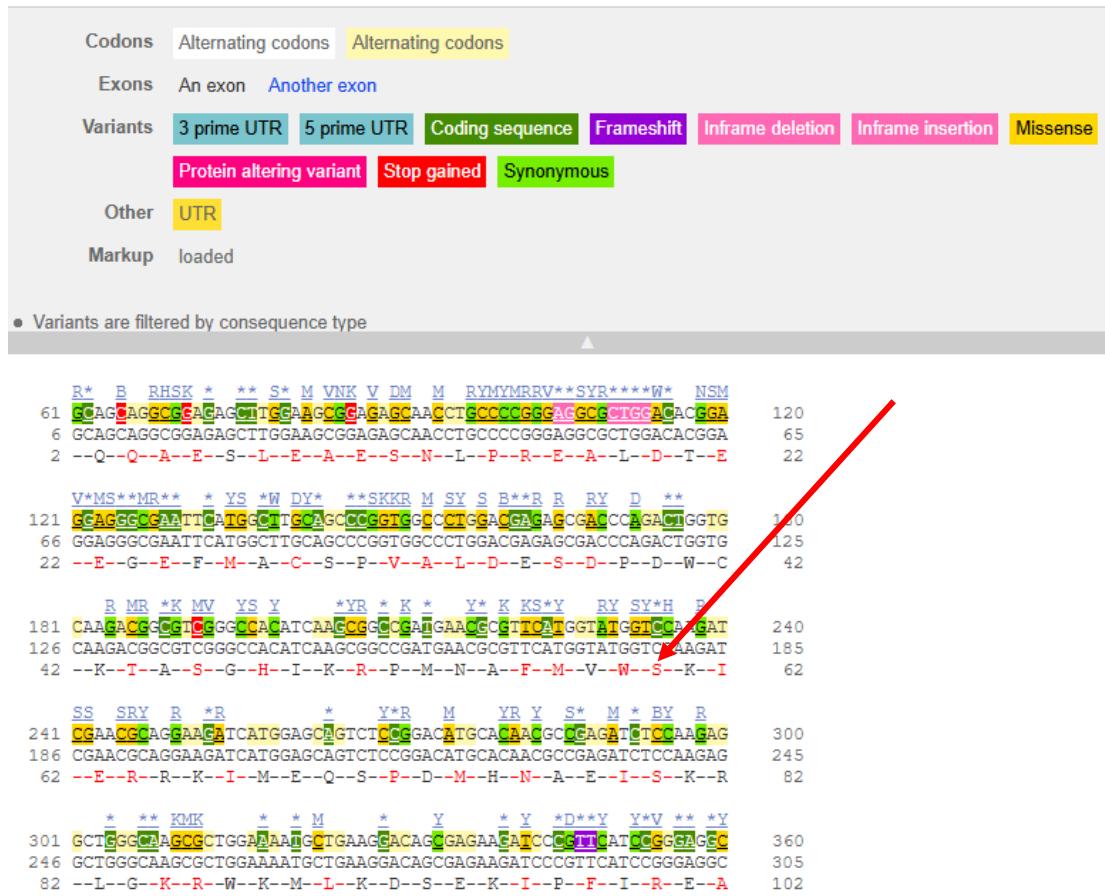
### **PROTEIN SEQUENCE**

The protein sequence is retrieved from the protein database, UniProt and the mutation is highlighted.

```
1 MVQQAESLEAESNLPREALDTEEgefMACSPVALDESDPDWCKTASGHIKRPMNAFMVWS 60
61 KIERRKIMEQSPDMHNAEISKRLGKRWKMVKDSEKIPFIREAERLRLKHMADYPDYKYRP 120
121 RKKPKMDPSAKPSASQSPEKSAAGGGGGSAGGGAGGGAKTSKGSSKKCGKLKAPAAAGAKA 180
181 GAGKAAQSGDYGGAGDDYVLGSLRVSGSGGGAGKTVKCVFLDEDDDDDDDELQLQIK 240
241 QEPDEEDEEPPHQQLLQPPGQQPSQLRRYNVAKVPASPTLSSSAESPEGASLYDEVRAAG 300
301 ATSGAGGGSRLYYSFKNITKQHPPPLAQPALSPASSRSVSTSSSSSSGSSSGGEDADD 360
361 LMFDLISLNFSQSAHSASEQQLGGGAAAGNLSLSLVDKLDLDSFSEGSLGSHFEFPDYCTPE 420
421 LSEMIAGDWLEANFSDLVFTY 441
```

## TRANSCRIPT SEQUENCE

The arrow in the image below indicates the position of the variant within the cDNA sequence.



## PROTEIN STRUCTURE

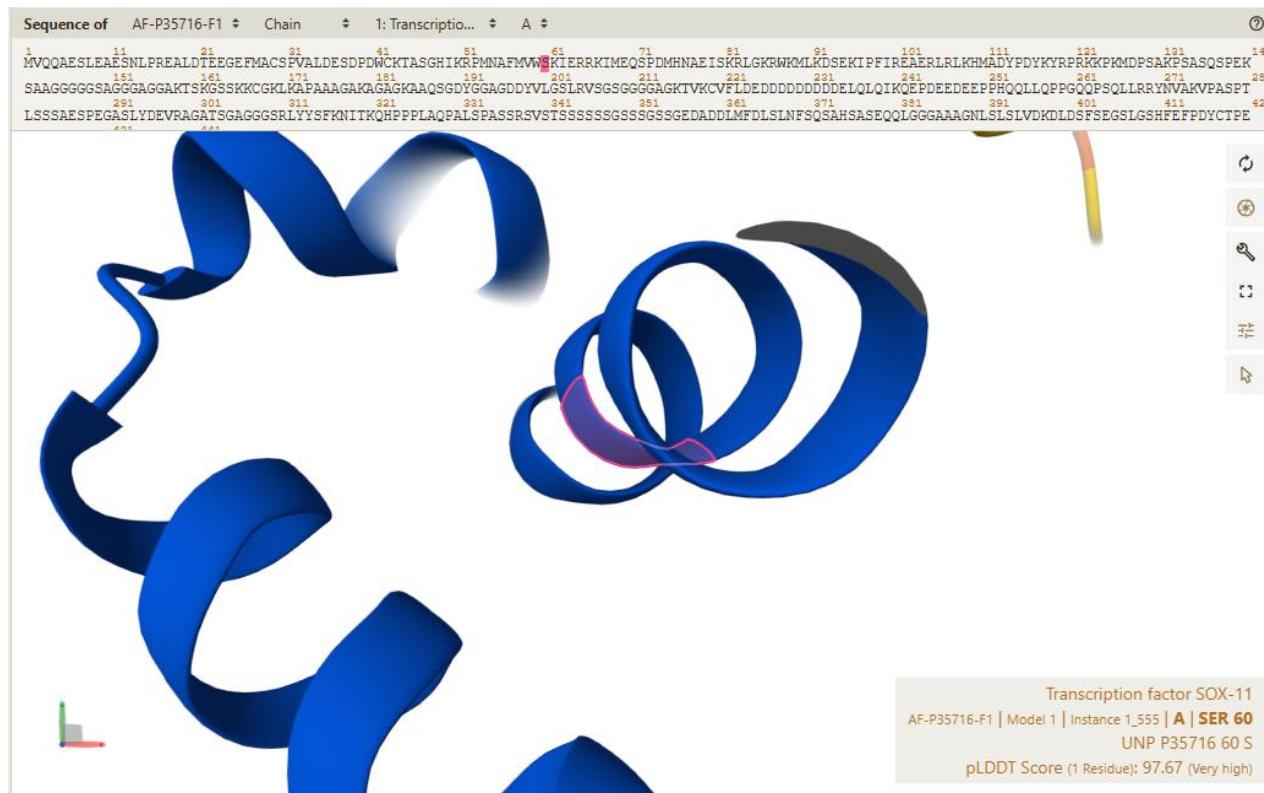


Figure 2 The figures represent the 3D structures of the protein at position SER 60, obtained from AlphaFold.

## VARIANT c.355C>T (p.Arg119Trp)

### PROTEIN SEQUENCE

The protein sequence is retrieved from the protein database, UniProt and the mutation is highlighted.

```
1 MVQQAESLEAESNLPREALDTEEgefMACSPVALDESDPDWCKTASGHIKRPMNAFMVWS 60
61 KIERRKIMEQSPDMHNAEISKRLGKRWMLKDSEKIPFIREAERLRLKHMADYPDYKYRP 120
121 RKKPKMDPSAKPSAQSQSPEKSAAAGGGGSAGGGAGGGAKTSKGSSKKCGKLKAPAAAGAKA 180
181 GAGKAAQSGDYGGAGDDYVLGSILRVSGSGGGGAGKTVKCVFLDEDAAAAAELQLQIK 240
241 QEPDEEDEEPPHQQLLQPPGQQPSQLRRYNVAKVPASPTLSSSAESPEGASLYDEVRAG 300
301 ATSGAGGGSRLLYYSFKNITKQHPPPLAQPALSPASSRSVSTSSSSSGSSSGGEDADD 360
361 LMFDLSLNFSQSAHSASEQQQLGGGAAAGNLSLSLVDKLDLDSFSEGSLGSHFEFPDYCTPE 420
421 LSEMIAGDWLEANFSDLVFTY 441
```

## TRANSCRIPT SEQUENCE

The arrow in the image below indicates the position of the variant within the cDNA sequence.

## PROTEIN STRUCTURE

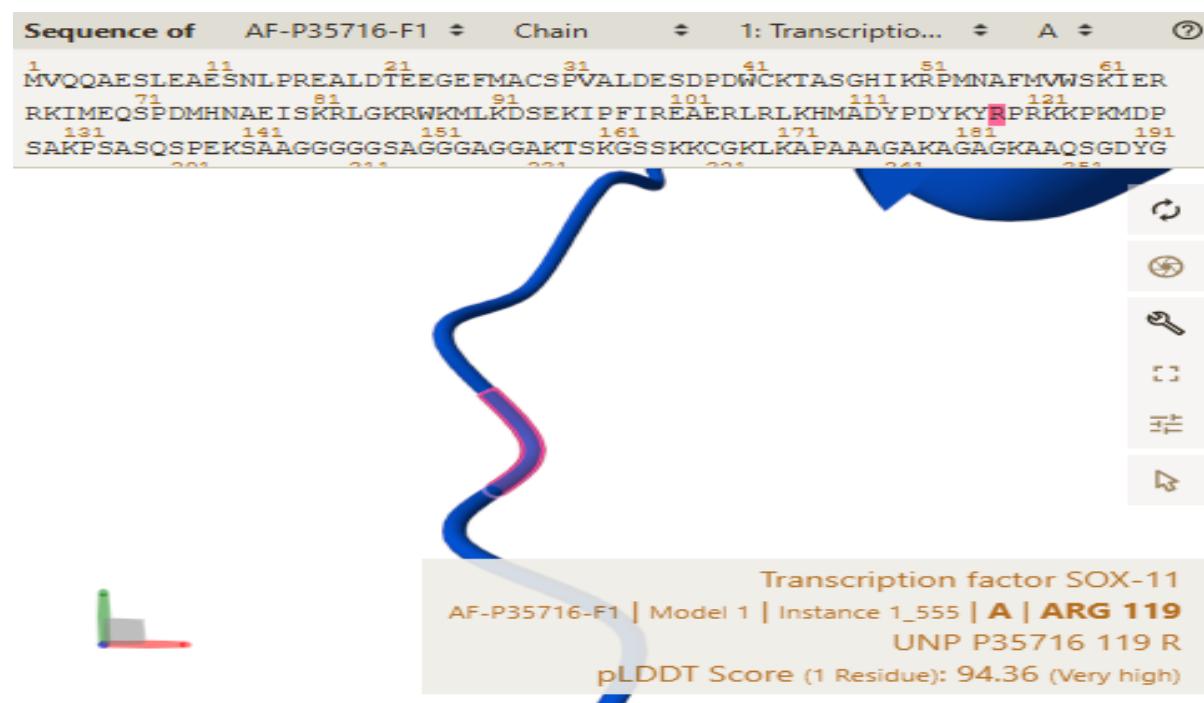


Figure 3 The figures represent the 3D structures of the protein at position ARG119, obtained from AlphaFold.

## VARIANT c.170T>C (p.Met57Thr)

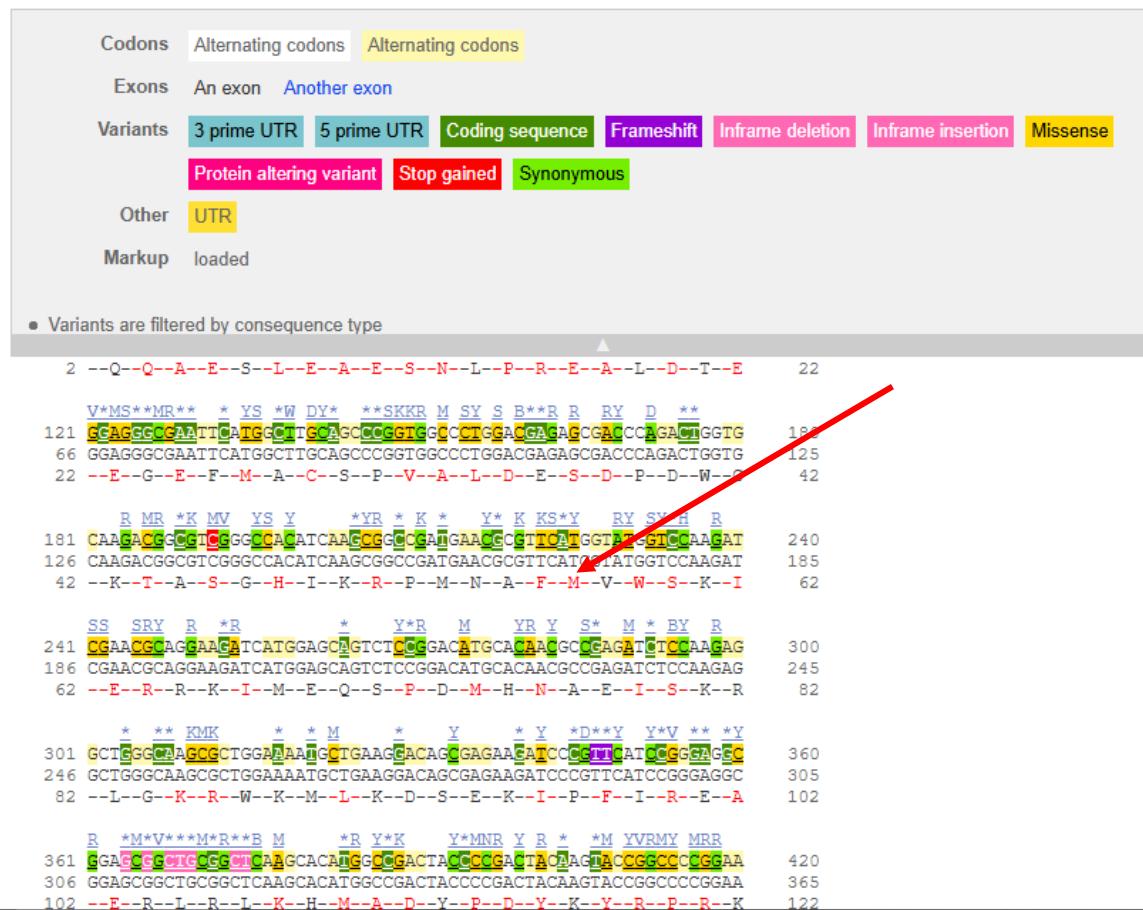
### PROTEIN SEQUENCE

The protein sequence is retrieved from the protein database, UniProt and the mutation is highlighted.

```
1 MVQQAESLEAESNLPREALDTEEGERMACSPVALDESDPDWCKTASGHIKRPMNAFMVWS 60
61 KIERRKIMEQSPDMHNAEISKRIGKRWIMLKDESEKIPFIREAERLRLKHMADYPDYKYRP 120
121 RKKPKMMDPSAKPSASQSPEKSAAGGGGGSAGGGAGGAKTSKGSSKKCGKLKAPAAAGAKA 180
181 GAGKAAQSGDYGGAGDDYVLGSLRVSGSGGGAGKTVKCVFLDEDDEDDDDDDDELQLQIK 240
241 QEPDEEDEEPPPHQQQLLQPPGQQPSQILLRYYNVAKVPASPTLSSAESPEGASLYDEVRAG 300
301 ATSGAGGGSRILYYSFKNITKQHPPPLAQPALSPASSRSVSTSSSSSGSSSGGEDADD 360
361 LMFDSLNSQSAHSASEQQLGGGAAAGNLSLSLVDKLDLDSFSEGSLGSHFEFPDYCTPE 420
421 LSEMIAGDWLEANFSQSLVFTY                                         441
```

## TRANSCRIPT SEQUENCE

The arrow in the image below indicates the position of the variant within the cDNA sequence.



## PROTEIN STRUCTURE



Figure 4 The figures represent the 3D structures of the protein at position MET 57, obtained from AlphaFold.

## VARIANT c.250G>C (p.Gly84Arg)

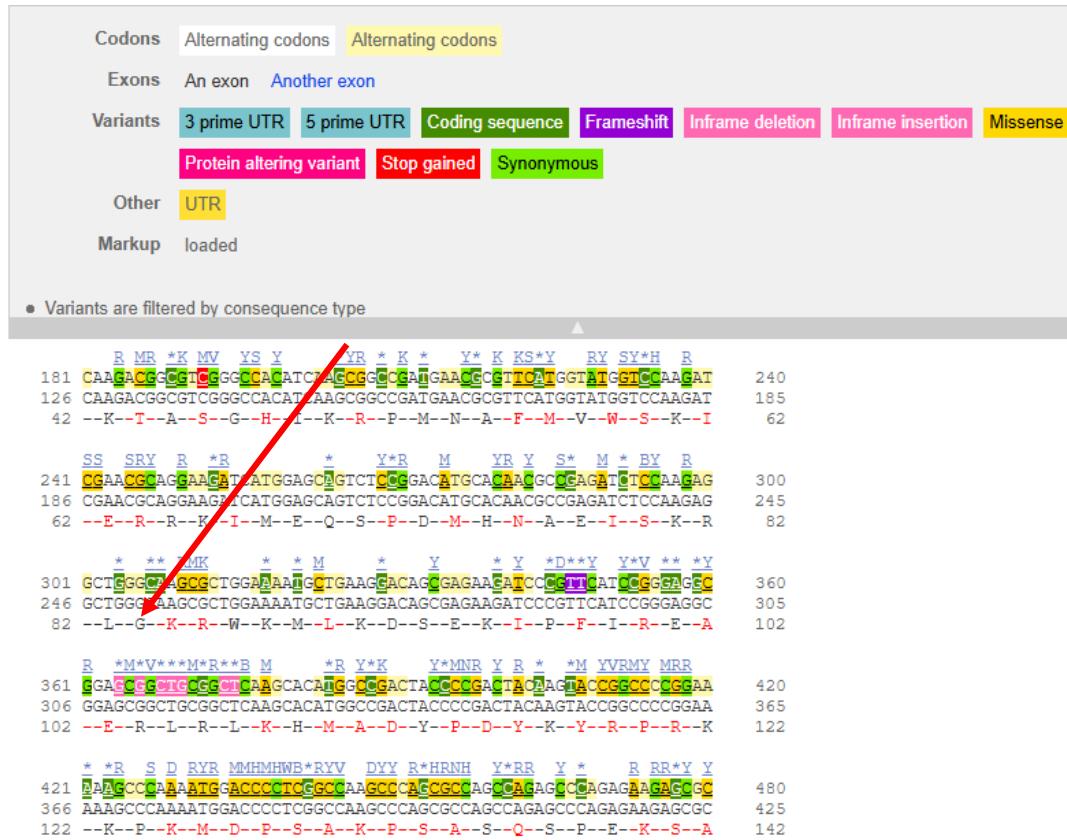
### PROTEIN SEQUENCE

The protein sequence is retrieved from the protein database, UniProt and the mutation is highlighted.

```
1 MVQQAESLEAESNLPREALDTEEGERMACSPVALDESDPDWCKTASGHIKRPMNAFMVWS 60
61 KIERRKIMEQSPDMHNAEISKRLGKRWKMLKDSEKIPFIREAERLRLKHMADYPDYKYRP 120
121 RKKPKMDPSAKPSAQSQSPEKSAAGGGGGSAGGGAGGAKTSGSSKKCGKLKAPAAAGAKA 180
181 GAGKAAQSGDYGGAGDDYVLGSLRVSGSGGGAGKTVKCVFLDEDDDDDDDDELQLQIK 240
241 QEPDEEDEEPPHQQLLQPPGQQPSQLRRYNVAKVPASPTLSSSAESPEGASILYDEVVRAG 300
301 ATSGAGGGSRILYYSFKNITKQHPPPLAQPALSPASSRSVSTSSSSSGSSSGGEDADD 360
361 LMFDSLNSQSAHSASEQQLGGGAAAGNLSLSLVDKLDLSFSEGSILGSHFEFPDYCTPE 420
421 LSEMIAGDWLEANFSQSLVFTY                                         441
```

## TRANSCRIPT SEQUENCE

The arrow in the image below indicates the position of the variant within the cDNA sequence.



## PROTEIN STRUCTURE

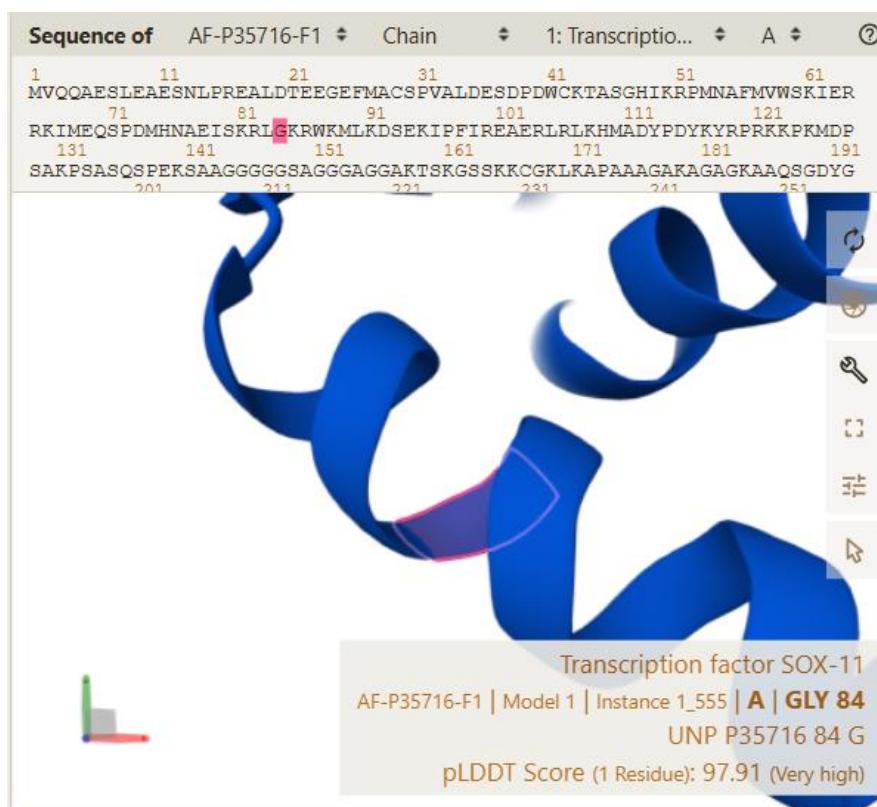


Figure 5 The figures represent the 3D structures of the protein at position GLY 84, obtained from AlphaFold.

## **DPF2**

### **VARIANT :c.1066T>G (p.Cys356Gly)**

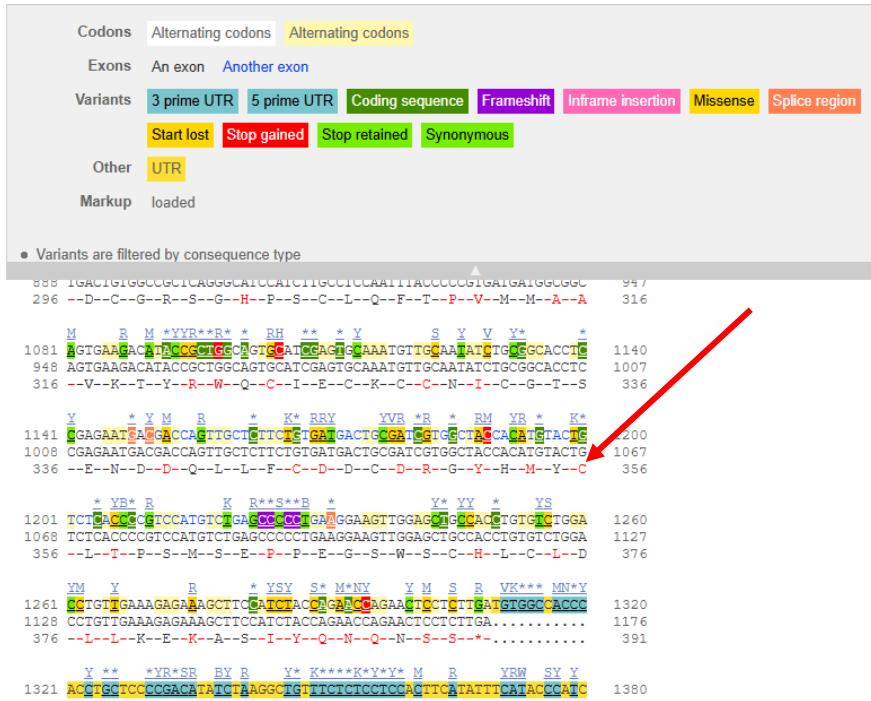
#### **PROTEIN SEQUENCE**

The protein sequence is retrieved from the protein database, UniProt and the mutation is highlighted.

```
1 MAAVENVVKLLGEQYYKDAMEQCHNYNARLCAERSVRLPFLDSQTGVAQSNCYIWMEKR 60
61 HRGPGLASGQLYSYPARRWRKKRRAHPPEDPRLSFPSIKPDTDQTLKKEGLISQDGSSLE 120
121 ALLRTDPLEKRGAPDPRVDDDSLGEGFPVTNSRARKRILEPDDFLDDLDEDYEEDTPKRR 180
181 GKGKSKGKGVGSARKKLDASILEDRDKPYACDICGKRYKNRPGLSYHYAHSHLAEEEGED 240
241 KEDSQPPTPVSQRSEEQKSKKGPGLALPNNYCDFCLGDSKINKKTGQPEELVSCSDCGR 300
301 SGHPSClQFTPVMMAAVKTYRWQCIECKCCNICGTSENDDQLLFCDDCDRGYHMYCLTPS 360
361 MSEPPEGSWSCHLCLDLLKEKASIYQNQNSS 391
```

## TRANSCRIPT SEQUENCE

The arrow in the image below indicates the position of the variant within the cDNA sequence.



## PROTEIN STRUCTURE



Figure 6 The figures represent the 3D structures of the protein at position CYS 356, obtained from AlphaFold.

## VARIANT c.1045G>A (p.Asp349Asn)

### PROTEIN SEQUENCE

The protein sequence is retrieved from the protein database, UniProt and the mutation is highlighted.

```
1 MAAVVENVVKLLGEQYYKDAMEQCHNYNARLCAERSVRLPFLDSQTGVAQSNCIYWMEKR 60
61 HRGPFGLASGQLSYPPARRWRKKRAHPPEPDPLRSFPSIKPDTDQTLKKEGLISQDGSSLE 120
121 ALLRTDPLEKRGAPDPRVDDDSLGEFPVTNSRARKRILEPDDFLDDDEDYEEDTPKRR 180
181 GKGKSKGKGVGSARKLDAISILEDRDKPYACDICGKRYKNRPGLSYHYAHSHLAEEGED 240
241 KEDSQPPTPVSRSEEQSKKGPDGLALPNNYCDFCLGDSKINKKTGQPEELVSCSDCGR 300
301 SGHPSCPQFTPVMMMAVKTYRWQCIECKCCNICGTSENDDQLLFCDDCDRGYHMYCLTPS 360
361 MSEPPREGSWSCHLCIDLLEKASIYQNQNSS 391
```

### TRANSCRIPT SEQUENCE

The arrow in the image below indicates the position of the variant within the cDNA sequence.

Codons   Alternating codons   Alternating codons

Exons   An exon   Another exon

Variants   3 prime UTR   5 prime UTR   Coding sequence   Frameshift   Inframe insertion   Missense   Splice region

Start lost   Stop gained   Stop retained   Synonymous

Other   UTR

Markup   loaded

• Variants are filtered by consequence type

Position	Sequence	Consequence Type
1021	<b>CTTC</b> AAAT <b>ATC</b> T <b>CG</b> AC <b>CTC</b> CGAGAA <b>AT</b> <b>AC</b> <b>CCA</b> <b>TTG</b> C <b>TTC</b> <b>AT</b> <b>GAT</b> <b>GACT</b> <b>CG</b>	Missense
986	GTGCAATA <b>TCTGC</b> CGAC <b>CTCCG</b> AGAA <b>AT</b> GAC <b>GACC</b> ATTG <b>GCT</b> CT <b>CTCTGT</b> GAT <b>GACT</b> <b>CG</b>	Missense
329	C--C--N--I--G--T--S--E--N--D--D--Q--L--F--C--D--D--C--	Missense
1080	1080	
1045	1045	
348	348	
1081	<b>A</b> <b>T</b> <b>CG</b> <b>GCT</b> <b>AC</b> <b>CA</b> <b>AT</b> <b>T</b> <b>AC</b> <b>T</b> <b>CT</b> <b>CG</b> <b>AC</b> <b>CC</b> <b>T</b> <b>CC</b> <b>AT</b> <b>GTC</b> <b>G</b> <b>A</b> <b>CC</b> <b>CC</b> <b>T</b> <b>G</b> <b>A</b> <b>G</b> <b>TT</b>	Missense
1046	A <b>T</b> <b>G</b> <b>GG</b> <b>CT</b> <b>TAC</b> <b>AC</b> <b>AT</b> <b>T</b> <b>AC</b> <b>T</b> <b>CT</b> <b>CG</b> <b>AC</b> <b>CC</b> <b>T</b> <b>CC</b> <b>AT</b> <b>GTC</b> <b>G</b> <b>A</b> <b>CC</b> <b>CC</b> <b>T</b> <b>G</b> <b>A</b> <b>G</b> <b>TT</b>	Missense
349	D--R--G--Y--H--M--Y--C--L--T--P--S--M--S--E--P--P--E--G--S--	Missense
1140	1140	
1105	1105	
368	368	
1141	<b>GG</b> <b>AC</b> <b>CT</b> <b>G</b> <b>CC</b> <b>AC</b> <b>CT</b> <b>GT</b> <b>GT</b> <b>GC</b> <b>CT</b> <b>GT</b> <b>T</b> <b>G</b> <b>AA</b> <b>AG</b> <b>AG</b> <b>AA</b> <b>AG</b> <b>CT</b> <b>TC</b> <b>CA</b> <b>T</b> <b>AC</b> <b>CA</b> <b>G</b> <b>AA</b> <b>CA</b>	Missense
1106	GGAGCTGCCAC <b>CTGTG</b> TGGAC <b>CTGTTG</b> AAAGAGAAAG <b>CTTCC</b> AT <b>CTAC</b> <b>CCAGA</b> <b>AC</b> <b>AGA</b>	Missense
369	W--S--C--H--L--C--L--D--L--L--K--E--K--A--S--I--Y--Q--N--Q--	Missense
1200	1200	
1165	1165	
388	388	
1201	<b>A</b> <b>C</b> <b>T</b> <b>CT</b> <b>CT</b> <b>GT</b> <b>AT</b> <b>G</b> <b>T</b> <b>GG</b> <b>CC</b> <b>AC</b> <b>CC</b> <b>AC</b> <b>T</b> <b>CT</b> <b>CC</b> <b>CG</b> <b>AC</b> <b>AT</b> <b>AT</b> <b>CT</b> <b>A</b> <b>GG</b> <b>GT</b> <b>T</b> <b>TC</b> <b>T</b> <b>CT</b> <b>CC</b> <b>TC</b>	Missense
1166	ACTCCCT <b>CTTGA</b> .....	Missense
389	N--S--S--*--.....	Missense
1260	1260	
1176	1176	
391	391	
1261	<b>C</b> <b>A</b> <b>C</b> <b>T</b> <b>TC</b> <b>AT</b> <b>AT</b> <b>TT</b> <b>C</b> <b>A</b> <b>T</b> <b>AC</b> <b>CC</b> <b>AT</b> <b>T</b> <b>CT</b> <b>CC</b> <b>CT</b> <b>C</b> <b>T</b> <b>CT</b> <b>CC</b> <b>TT</b> <b>C</b> <b>A</b> <b>C</b> <b>AA</b> <b>A</b> <b>T</b> <b>CC</b> <b>AG</b> <b>AG</b> <b>AA</b>	Missense
.....	.....	
1320	1320	

## PROTEIN STRUCTURE



Figure 7 The figures represent the 3D structures of the protein at position ASP 349, obtained from AlphaFold.

## VARIANT c.868G>A (p.Glu290Lys)

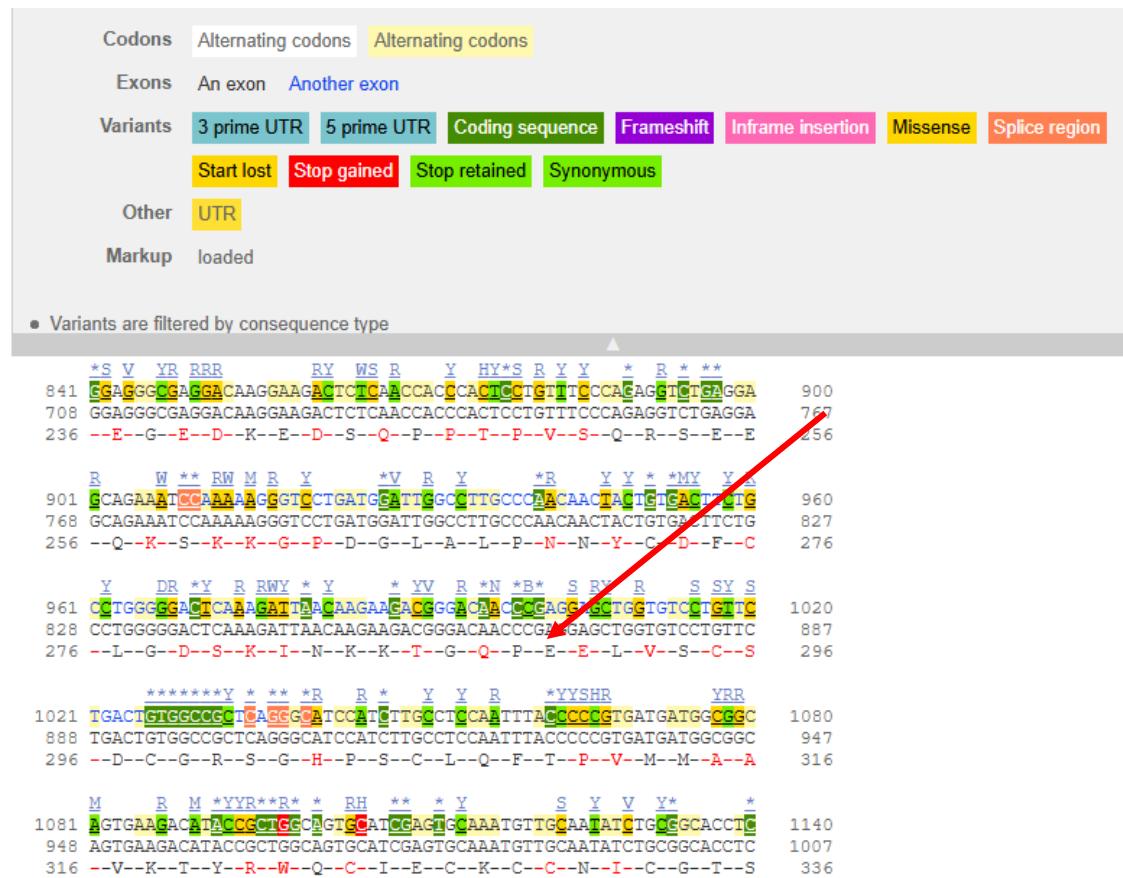
### PROTEIN SEQUENCE

The protein sequence is retrieved from the protein database, UniProt and the mutation is highlighted.

```
1 MAAVENVVKLLGEQYYKDAMEQCHHYNARLCAERSVRLPFLDSQTGVAQSNCYIWMERK 60
61 HRGPGLASGQLYSYPARRWRKKRRAHPPEPDPLSFPISKPDTDQTLKKEGLISQDGSSLE 120
121 ALLRTDPLEKRGAPDPRVDDDSLGEFPVTNSRARKRILEPDDFLDDLDEDYEEDTPKRR 180
181 GKGKSKGKGVGSARKKLDASILEDRDKPYACDICCGKRYKNRPGLSYHYAHSHLAEEGED 240
241 KEDSQPPPTPVQRSEEQKSKKGPDGLALPNYCDFCLGDSKINKKTGQEEELVSCSDCGR 300
301 SGHPSCLQFTPVMMAAVKTYRWQCIECKCCNICGTSENDDQLLFCDDCDRGYHMYCLTPS 360
361 MSEPPEGWSWCHLCLLLEKASIYQNQNS 391
```

### TRANSCRIPT SEQUENCE

The arrow in the image below indicates the position of the variant within the cDNA sequence.



## PROTEIN STRUCTURE

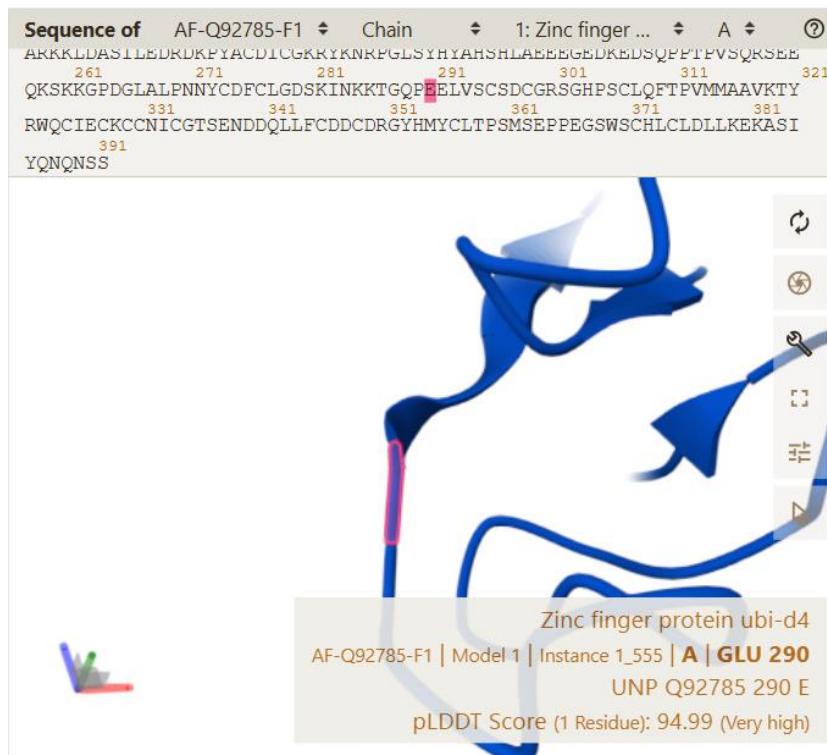


Figure 8 The figures represent the 3D structures of the protein at position GLU 290, obtained from AlphaFold.

## VARIANT c.827G>T (p.Cys276Phe)

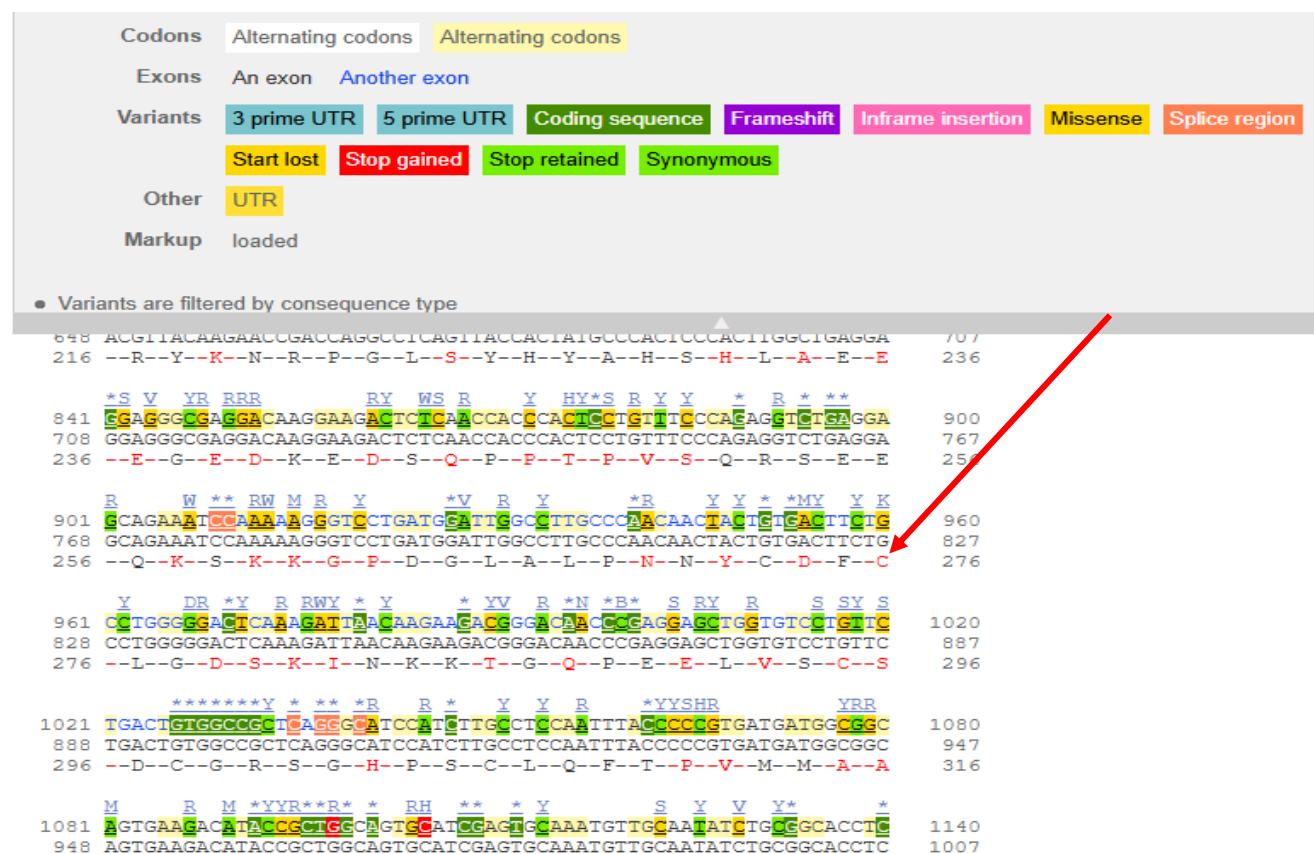
### PROTEIN SEQUENCE

The protein sequence is retrieved from the protein database, UniProt and the mutation is highlighted.

```
1 MAAVENVVKLLGEQYYKDAMEQCHNYNARLCAERSVRLPFLDSQTGVAQSNCYIWMEKR 60
61 HRGPGLASGQLYSYPARRWRKKRRAHPPEDPRLSFPSIKPDTDQTLKKEGLISQDGSSLE 120
121 ALLRTDPLEKRGAPDPRVDDDSLGEFPVTNSRARKRILEPDDFLDDLDDEDYEEDTPKRR 180
181 GKGKSKGKGVGSARKKLDASILEDRDKPYACDICGKRYKNRPGLSYHYAHSHLAEEEGED 240
241 KEDSQPPTPVSQRSEEQKSSKKGPDGLALPNNYCDFCLGDSKINKKTGQPEELVSCSDCGR 300
301 SGHPSCLQFTPVMMAAVKTYRWQCIECKCCNICGTSENDDQLLFCDDCDRGYHMYCLTPS 360
361 MSEPPEGSWSCHLCLDLLKEKASIYQNQNSS 391
```

### TRANSCRIPT SEQUENCE

The arrow in the image below indicates the position of the variant within the cDNA sequence.



## PROTEIN STRUCTURE

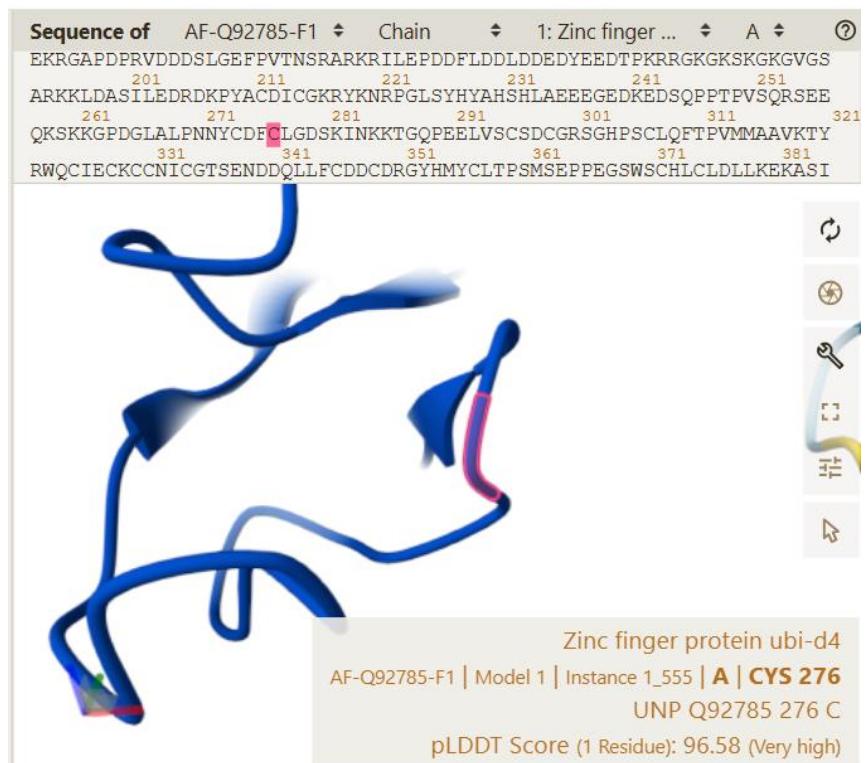


Figure 9 The figures represent the 3D structures of the protein at position CYS 276, obtained from AlphaFold.

## VARIANT c.990C>G (p.Cys330Trp)

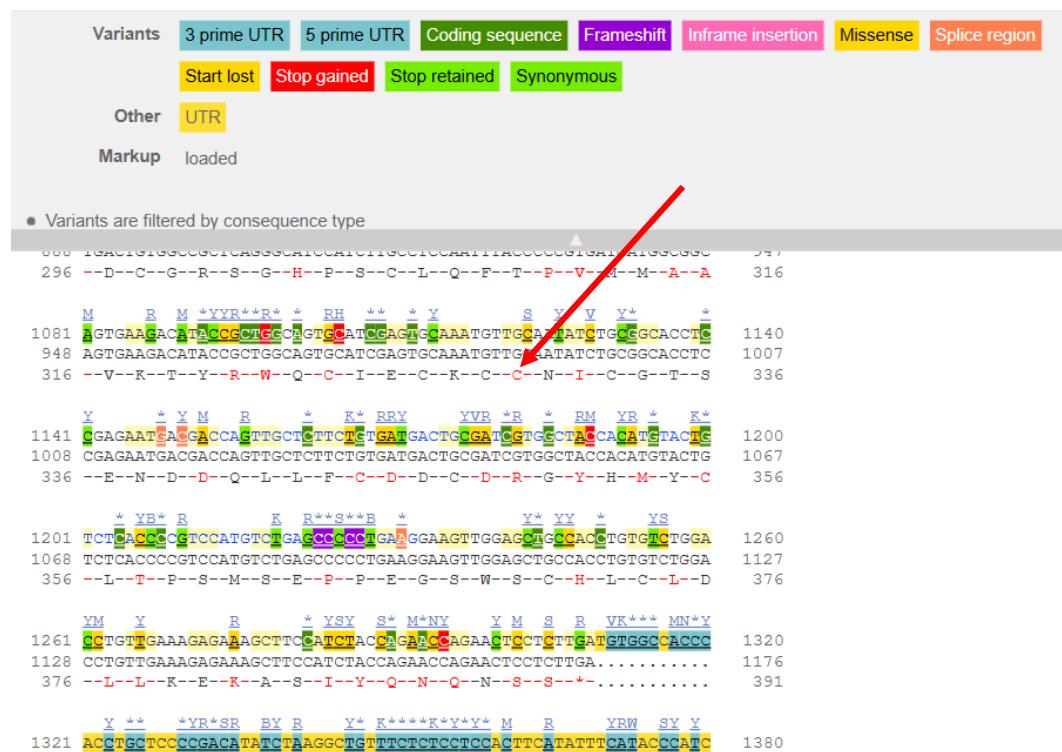
### PROTEIN SEQUENCE

The protein sequence is retrieved from the protein database, UniProt and the mutation is highlighted.

```
1 MAAVENVVKLGEQYYYKDAMEQCHNYNARLCAERSVRLPFLDSQTGVAQSNCYIWMEKR 60
61 HRGPGLASGQLYSYPARRWRKKRRAHPPEDPLSFPSIKPDTDQLKKEGLISQDGSSLE 120
121 ALLRTDPLEKRGAPDPRVDDDSLEFPVTSRARKRILEPDDFLDDLDDEDYEEDTPKRR 180
181 GKGSKGKGVGSARKKLDASILEDRDKPYACDIICGKRYKNRPGLSYHYAHSHLAEEEGED 240
241 KEDSQPPTPSQRSEEQKSKKGPDGLALPNYCDFCLGDSKINKKTGQPEELVSCSDCGR 300
301 SGHPSCLQFTPVMMAAVKTYRWQCIECKCCNICGTSENDDQLLFCDDCDRGYHMYCLTPS 360
361 MSEPPEGGWSWSCHLCLDLLKEKASIYQNQNS 391
```

### TRANSCRIPT SEQUENCE

The arrow in the image below indicates the position of the variant within the cDNA sequence.



## PROTEIN STRUCTURE

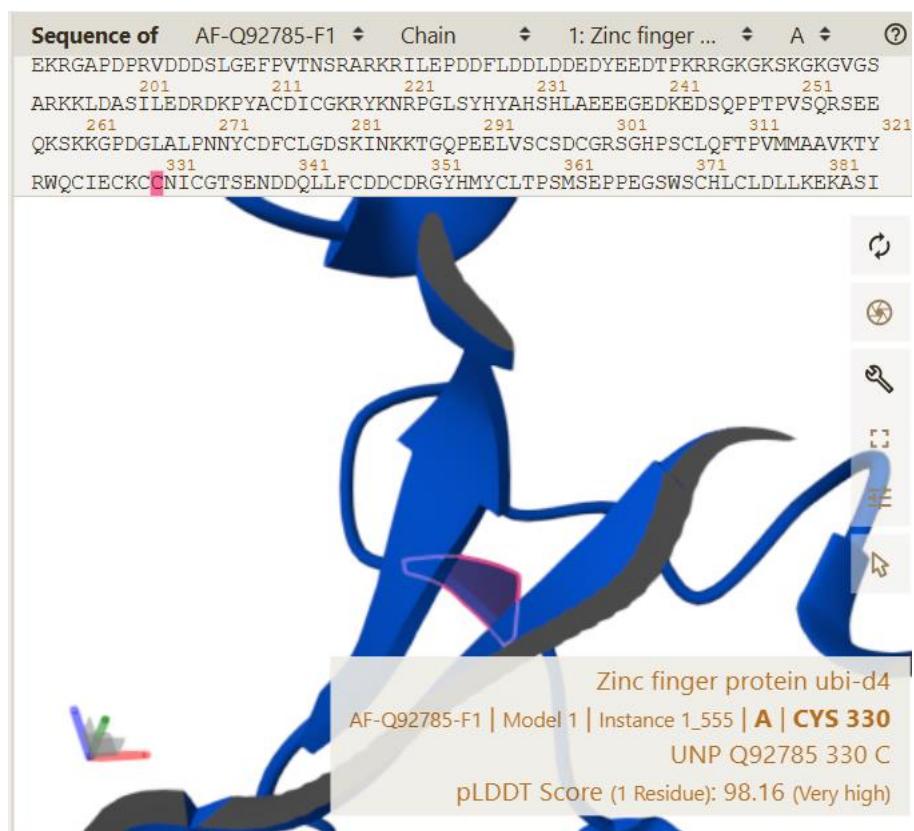


Figure 10 The figures represent the 3D structures of the protein at position CYS 330, obtained from AlphaFold.

## **SMARCA4**

### **VARIANT c.1645C>T (p.Arg549Cys)**

#### **PROTEIN SEQUENCE**

The protein sequence is retrieved from the protein database, UniProt and the mutation is highlighted.

```
1 MSTPDPLGGTPRPGPSPGPSPGAMLGSPSPGSAHSMMGPSPGPPSAGHPIPTQG 60
61 PGGYPQDNMHQMHKPMESMHEKGMSDDPRYNQMKGMRSCGHAGMPPPSPMDQHSQGY 120
121 PSPLGGSEASSPVPASGPSSGPQMSSGPGGAPLDGADPQALGQQNRRGPTPFNQNQLHQL 180
181 RAQIMAYKMLARGQPLPDHLQMAVQGKRPMPGMQQQMPTLPPPSVSATGPGPGPGPGP 240
241 GPGPAPPNYSRPHGMGGPNMPPPGPSGVPPGMPGQPPGGPKWPEGPMANAAAPTSTPQ 300
301 KLIPPQPTGRPSAPPAPPAASPVMPPQTQSPGQPAQPAVMVPLHQKQSRTPIQKPRG 360
361 LDPVEILQEREYRLQARIAHRIQELENLPGSLAGDLRTKATIELKALRLLNFQRQLRQEV 420
421 VVCMRRDTALETALNAKAYKRSKRQSLREARITEKLEKQQKIEQERKRRQKHQEYLNSIL 480
481 QHAKDFKEYHRSVTGKIQKLTKAVATYHANTEREQKKENERIEKERMRRLMAEDEEGYRK 540
541 LIDQKKDKRLAYLLQQTDEYVANLTELVRQHAAQVAKEKKKKKKKKKAENAEGQTPAIG 600
601 PDGEPLDETSQMSDLPVKVIHVESGKILTGTDAPKAGGLEAWLEMNPGYEVAPRSDSEES 660
661 GSEEEEEEEEEEQPQAAQPPTLPVEEKKKIPDPDSDDVSEVDARHIIENAKQDVDEYGV 720
721 SQALARGLQSYYAVAHAVTERVDKQSALMVNGVLKQYQIKGLEWLVLSLYNNNNLNGILADE 780
781 MGLGKTIQTIALITYLMEHKRINGPFLIIIVPLSTLSNWAEFDKWAPSVVKVSYKGSPAA 840
841 RRAFVPQLRSGFNVLLTTYEYIIKDHILAKIRWKYMIVDEGHRMKHHCKLTQVLNTH 900
901 YVAPRLLLTGTPLQNKLPELWALLNFLLPTIFKSCSTFEQWFNAFAMTGEKVDLNEEE 960
```

## TRANSCRIPT SEQUENCE

The arrow in the image below indicates the position of the variant within the cDNA sequence.

Codons	Alternating codons	Alternating codons
Exons	An exon	Another exon
Variants	3 prime UTR	5 prime UTR
	Coding sequence	Frameshift
	Splice donor	Inframe deletion
	Splice region	Inframe insertion
	Start lost	Missense
	Stop gained	
	Stop lost	
	Stop retained	
		Synonymous
Other	UTR	
Markup	loaded	
● Variants are filtered by consequence type		
1861	NNHNSDVNNNNN NNNHHNDI MRYVNNN NNNVNNBNNVDDNBHDDNN BNRWVRVD*NN TGCGGAGGCTCATGGCTAAGATGAGCAGGGCTACGGCAAGCTCATGCACCAAGACAGG	1920
1580	TGCGGAGGCTCATGGCTAAGATGAGGAGGGTACCCGAAGCTCATGCAGCAAAGAAGGG	1699
527	M--R--L--M-A-E--D--E--E-G-Y--R-K-L--I--D-Q--K--K--	546
1921	NNNDNRRNNNN NNNHHNNNNNSY NNRRNNDNDVNHDVNNDNNNN*NNNNKHNNHHNDYVNDVB ACAAGCCGCTGGCTAACCTCTGCAAGCAGACAGACAGACAGACTACGGCTAACCTCACGGCG	1980
1640	ACAAGCCGCTGGCTAACCTCTGCAAGCAGACAGACAGACTACGGCTAACCTCACGGAGC	1699
547	D--K--R--L--A--Y--L--L--Q--Q--T--D--E--Y--V--A--N--L--T--E--	566
1881	*NVNVBNVNNNDVNNMNNSN NBBNNBNNNNNNNWRRRDVNR*NNRDS*NWDD*D*WHDV*	
1700	TGGTGC CGCAGCACACAGCTGGCTGCCAGGTGCGCAAGGAGAAAAAGAAGAAAAAGAAGAAAAGA	1759
567	L--V--R--Q--H--K--A--A--Q--V--A--K--E--K--K--K--K--K--K--K--	586
2041	DN DVNNND DDDDDNNNDRD*RVNBNNMHNNNNVNSNNNDVNVRVNN VNVDDNHNNYVN AGAAGCCGAAATGCGAAGAAGACGCCAGTCGCAAGGAGAAAGCAGCTGGCGCGATGCGAGACCGCTCG	2100
1760	AGAAGGCAGAAAATCGAGAACAGACGCGCTGCCATTGGCCGGATGGCGAGCCTCTGG	1819
587	K--K--A--E--N--A--E--G--Q--T--P--A--I--G--P--D--G--E--P--L--	606
2101	WNVVV*NEWRNBDVWNNDVNNNNBKHNNNVDRWHNNNNVNNNNNDYV NRDRWVIVNNRDVH ACGGAGCAGCAGCATGAGCGACCTCCGGCTAAGGCTATCCACGTGGAGACTGGCAGAGA	2160
1820	ACGGAGCAGCAGCATGAGCGACCTCCGGCTAAGGCTATCCACGTGGAGACTGGCAGAGA	1879
607	D--E--T--S--Q--M--S--D--L--P--V--K--V--I--H--V--E--S--G--K--	626

## PROTEIN STRUCTURE

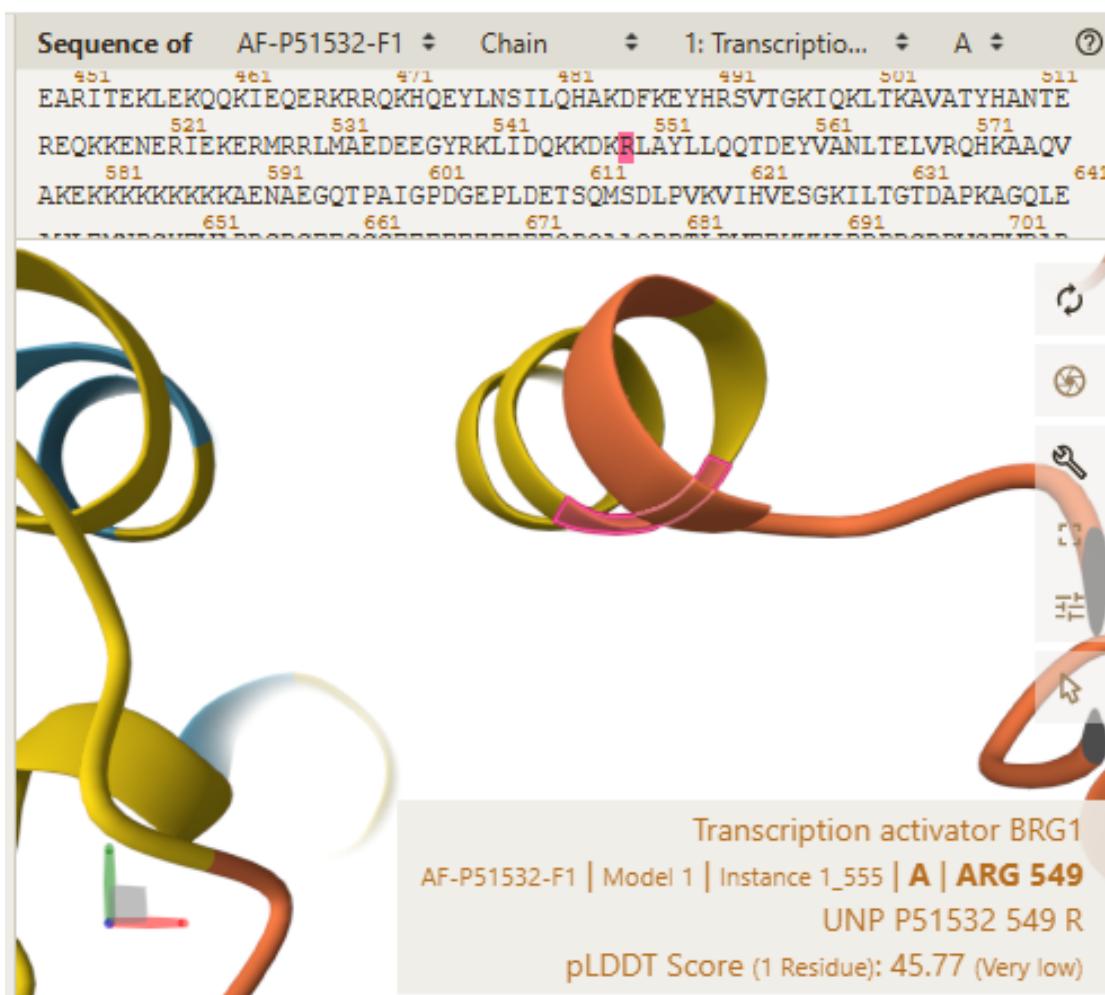


Figure 11 The figures represent the 3D structures of the protein at position ARG 549, obtained from AlphaFold.

## SMARCC2

### VARIANT :c.1919T>C (p.Leu640Pro)

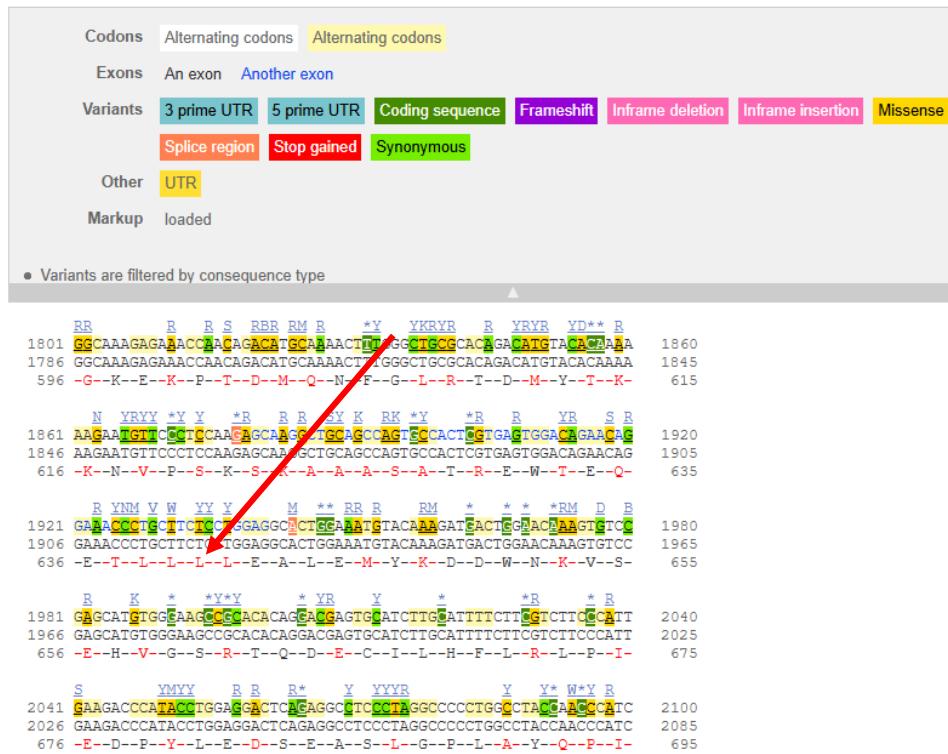
#### PROTEIN SEQUENCE

The protein sequence is retrieved from the protein database, UniProt and the mutation is highlighted.

```
241 WILD TDT FNEW MNEED YEV NDD KNP VSRR KKISAK TL DE VNS PDS DR RD KK GG NY KK R 300
301 RSP SP SPT PEAK KNAKK G P ST PY TKS KR GHREE E QED LT KDM D E P S P VP VN VEE VT LP KT 360
361 V NT KK DSE APV KGG TMD L DE QE DE SMET TG K D E D E N ST GN K GE QT KNP DL HED NV TE Q 420
421 THH III I P SY AAW FD YN S V HAI RRAL P E FF NG K NK S KTP E IY LAY RN FMID TY RL NP Q E Y 480
481 LT STAC RR NL A G DV C AIM RV HA FLE QW GL IN Y QV DAE S RPT PM GPP PT SHF VL AD TPS G 540
541 LV PL QP KTP Q GR QVD ADT KAG R KG KE L DD LV P ET AK G K P E L Q TS AS Q QML NF PD KG KE KP 600
601 TDM QNF GL RT DM YTK KN VP SK A A A S A T R E W T E Q E T L L I I E A L E M Y K D DW NK V S E H V G S 660
661 RT QD E C I L HFL RL P I E DP Y L E D S E A S L G P L A Y Q P I P F S Q S G N P V M S T V A F L A S V V D P R V A 720
721 SAA AKS A L E E F S K M K E E V P T A L V E A H V R K V E E A A K V T G K A D P A F G L E S S G I A G T T S D E P E 780
781 RIE E SG N D E A R V E G Q A T D E K K P K E P R E G G A I E E A K E K T S E A P K D E E K G K E G D S E K E 840
841 S E K S D G D P I V D P E K E K E P K E G Q E E V L K E V V E S E G E R K T K V E R D I G E G N I L S T A A A A L A A A A 900
901 AV KAK HLA AVE E R K I K S L V A L L V E T Q M K K L E I K L R H F E E L E T I M D R R E A L E Y Q R Q Q L L A 960
961 D R Q A F H M E Q L K Y A E M R A R Q Q H F Q Q M H Q Q Q Q Q P P A L P P G S Q P I P P T G A A G P P A V H G L A V A 1020
1021 P A S V V P A P A G S G A P P G S L G P S E Q I G Q A G S T A G P Q Q Q Q P A G A P Q P G A V P P G V P P P G P H G P S 1080
1081 D F D M O O P T P D S M M D C A T V D C S C H D G V A C N A D T C T D F C M D D D D D D D D A D S T T D F C S T A D S T S T M 1140
```

## TRANSCRIPT SEQUENCE

The arrow in the image below indicates the position of the variant within the cDNA sequence.



## PROTEIN STRUCTURE

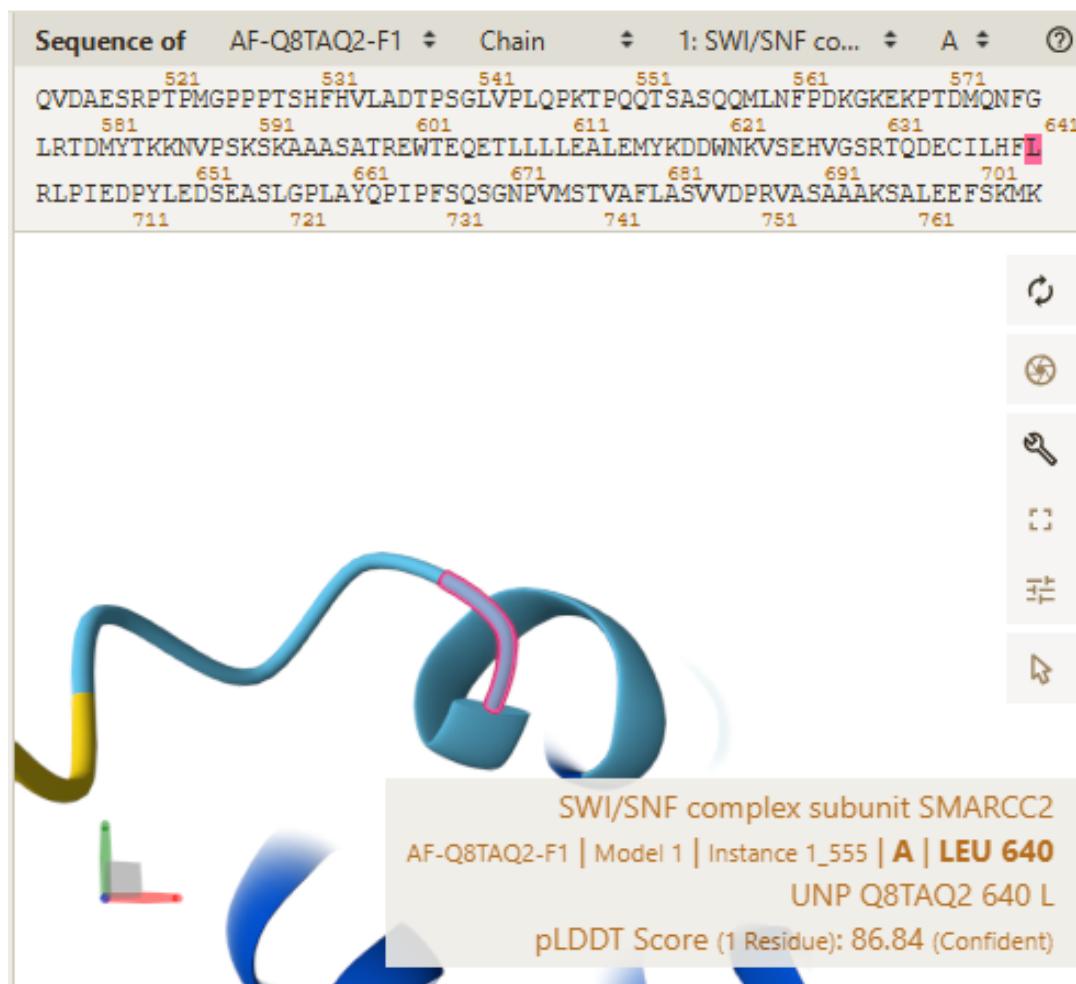


Figure 12 The figures represent the 3D structures of the protein at position LEU 640, obtained from AlphaFold.

BICRA

## VARIANT :c.192G>C (p.Glu64Asp)

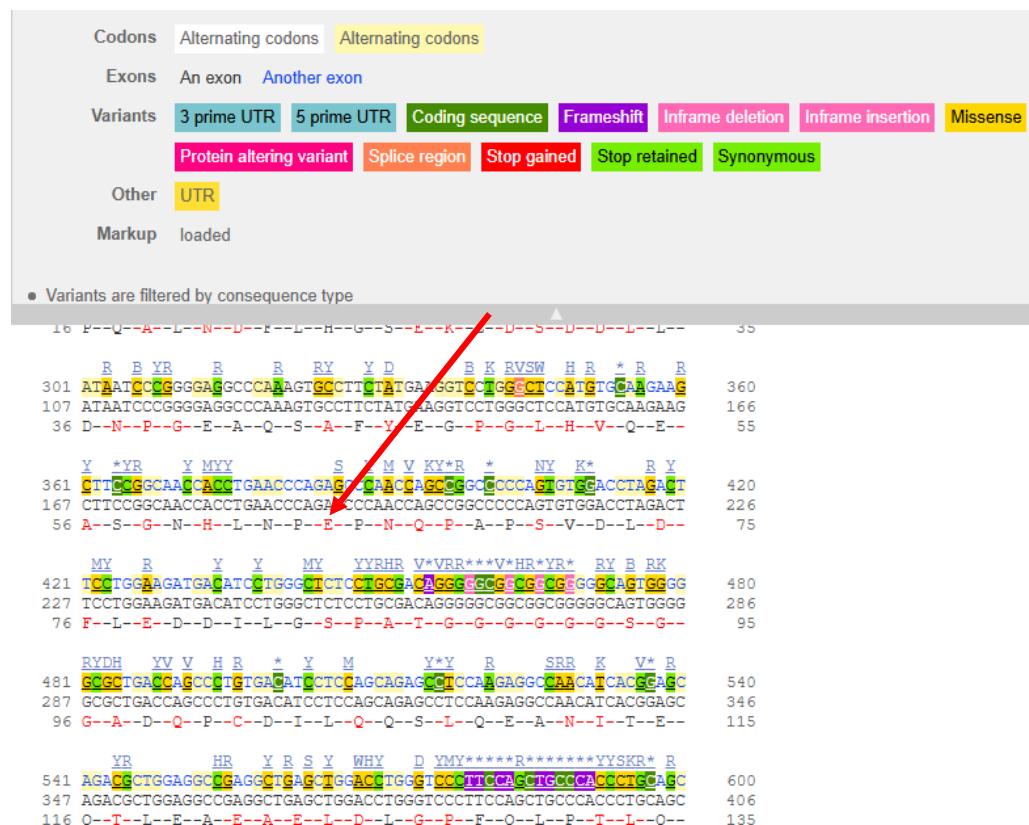
## PROTEIN SEQUENCE

The protein sequence is retrieved from the protein database, UniProt and the mutation is highlighted.

1	MDDDEDGRCLLDVVICDPQALNDFLHGSEKLDSDDLLDNPGEAQSAFYEGPGHLHVQEASGNH	60
61	LNP <span style="background-color: yellow;">EPNQPAPSVDLDFLEDDILGSPATGGGGGGSGGADQPCDILQQSLQEANITEQTLEA</span>	120
121	EAELDLGPQLPLQPADGGAGPTAGGAAA VAAGPQALFPGSTDLLGLQGPVTVLTHQA	180
181	LVPQDVNVKALSVPQFLQPVGLGNVTLQPIPLQGLPNGSPGGATAATLGLAPIQVVQG	240
241	FVMAINTPTSQLLAKQVPSVGSYLASAAGPSEPVTLASAGVSPQGAGLVIQKNLSAAVATT	300
301	LNGNSVFGGAGAA SAPTGT PSGQPLAVAPGLGSSPLVPAPNVLHRTPTPIQPKPAGVLP	360
361	PKLYQLTPKPFAPAGATLT I QGE PGAL PQPKAPQNLTFMAAGKAGQNVLSGFPAPALQ	420
421	ANVFQKPPATTGAPPQPPGALKSFKPMVS HLLNQGSSIVIPAQHMLPGQNQFLLPGAPAV	480
481	QLPQQLSALPANVGQ QILAAAAPHTGGQLIANPILTINQNLAGPLSLGPVLAPHSGAHSAH	540
541	IILSAAPIQVGQPA LFQMVPVSLAAGSLPTQSOPAPAGPAATTV LQGTVLPPSAVAMLNTPD	600
601	GLVQPATP AAATGEAA PVLTVQ PAPQAPPAV STPL LQGP QPA QQPP QAP TP PQAA APPQ	660
661	ATTPQPS PGLASS SPEKIVL GQPP SATPT TAITL QDSL QMFL PQERS SQQPL SAEGPH LSVP A	720
721	SVIVSAPP A QDP A PAT P V A K G A L G P Q A P D S Q A S P A P Q I P A A A P L K G P G P S S P L P	780
722	W O R L P C D R E P T P C E L T U O N G	840

## TRANSCRIPT SEQUENCE

The arrow in the image below indicates the position of the variant within the cDNA sequence.



## PROTEIN STRUCTURE

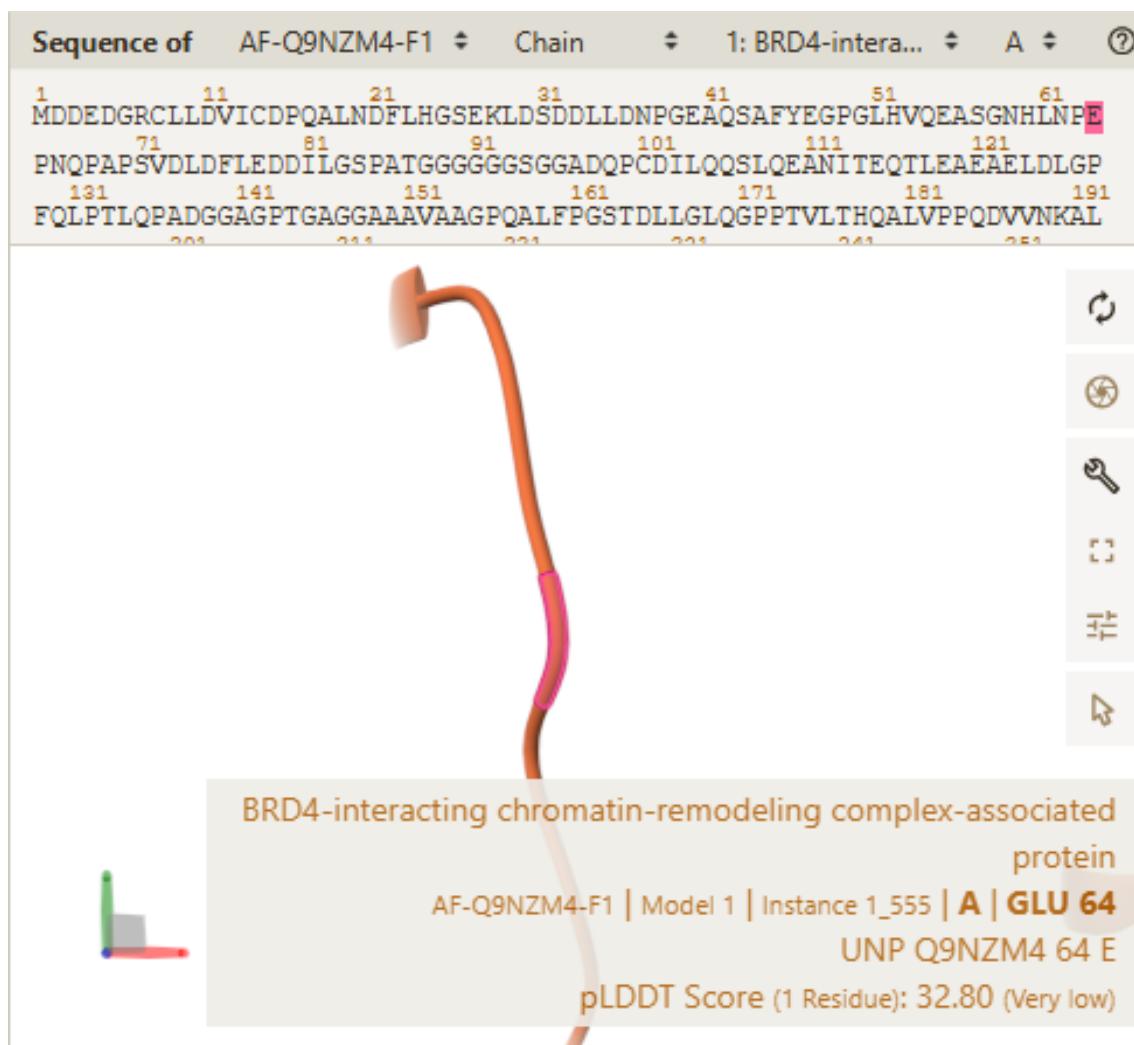


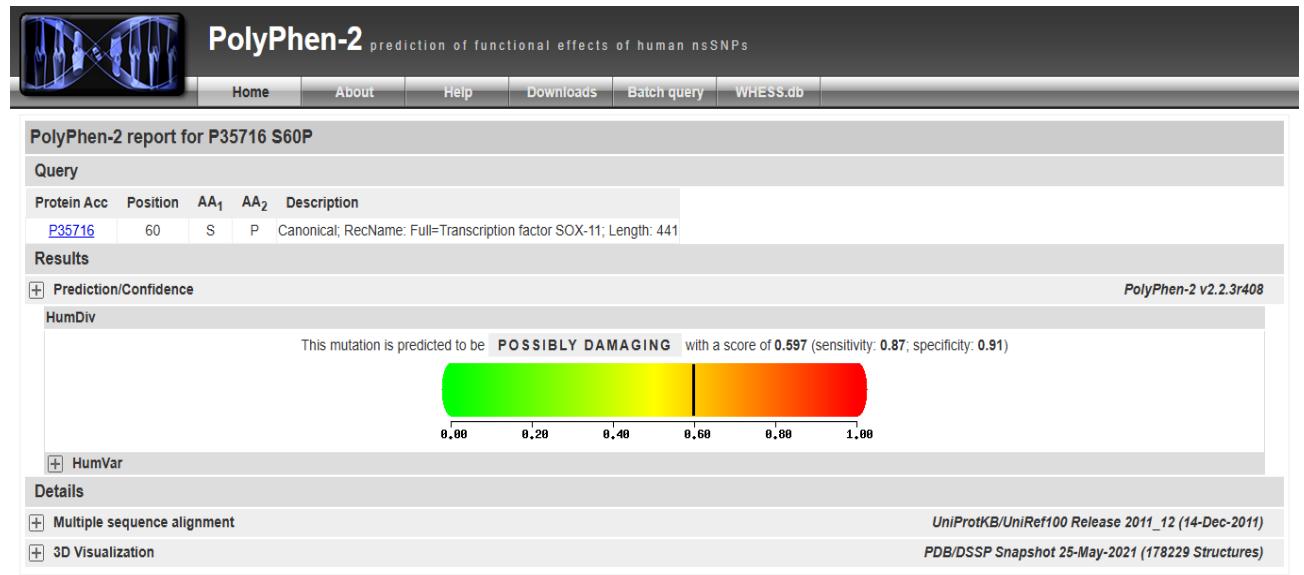
Figure 13 The figures represent the 3D structures of the protein at position GLU 64, obtained from AlphaFold.

# RESULT

## SOX11

**Variant – c.178T>C (p.Ser60Pro)**

**PolyPhen2** – Predicted the variant to be a possibly damaging kind with a score of 0.597



## SIFT- Sorting Intolerant From Tolerant

Predicted the variant under not tolerated

### Predictions for positions 1 through 100

Threshold for intolerance is 0.05.  
 Amino acid color code: nonpolar, uncharged polar, basic, acidic.  
 Capital letters indicate amino acids appearing in the alignment, lower case letters result from prediction.  
 'Seq Rep' is the fraction of sequences that contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information. Expect poor prediction at th

Predict Not Tolerated	Position	Seq	Rep	Predict Tolerated
cwdfmiyvgshnlateQP	50K	0.90	RK	
cwfmdiyyvshnatlpgeqK	51R	0.92	R	
ywvtsrqnmklkihgfedca	52P	0.93	P	
dhcwnyegsqrtafvikPL	53M	0.93	M	
mwifvclyrqhptkeagds	54N	0.93	N	
whyfimnqrdelkvtpSCG	55A	0.93	A	
kqhnrdrgepctsamvilWY	56F	0.93	F	
dhgneecswrkpyqtaVLF	57M	0.96	IM	
hdqnpecrgskytafmWI	58V	0.96	LV	
kqhnrdrgepctsamvil	59W	0.96	YFW	
whyfiqndelkvtpgCMR	60S	0.96	A5	
wfmiyvpghlatneCDS	61K	0.96	RQK	
wcyfM	62I	0.96	pLVHNGRqTSKADEI	
cwdpigstvlNKFYMA	63E	0.96	HRQE	
wdfmiyvgpsnlteACKQH	64R	0.96	R	
cwdfmiysnlteqGHVPA	65R	0.96	KR	
wcfypvgsNIDTHLMEQAR	66K	0.96	K	
hqpwdencrsatfKY	67I	0.96	MVL	I
wpDenGsCfQIyRTVHLK	68M	0.96	AM	
cwfymihgrnptV	69E	0.98	KASQDLE	
wycfhivpgnDMTRKSAL	70Q	1.00	EQ	

**MUTATION TASTER**– This software predicts the pathogenicity of the variant.

It predicted the variant as a deleterious variant.



Prediction:	<b>Deleterious</b>	<a href="#">Permalink</a>
<b>Summary:</b>		
	<ul style="list-style-type: none"> <li>Amino acid sequence changed</li> <li>Known disease mutation at this position (HGMD CM145914)</li> <li>Known disease mutation: ClinVar ID 139530 (pathogenic)</li> <li>Protein features (might be) affected</li> </ul>	<ul style="list-style-type: none"> <li>Model: simple_aae</li> <li>Tree vote: 9614 (del   benign) </li> <li>Automatic classification due to ClinVar</li> </ul>
<b>Analysed issue</b>	<b>Analysis result</b>	
Phys. location	chr2:5833031T>C <a href="#">show variant in all transcripts</a> <a href="#">IGV</a>	
Gene symbol	<a href="#">SOX11</a>	
ExAC LOF metrics	LOF: 0.34, missense: 4.25, synonymous: 2.64	
Ensembl transcript ID	<a href="#">ENST00000322002.3</a>	
Genbank transcript ID	<a href="#">NM_003108 (exact from MANE)</a>	
UniProt peptide	<a href="#">P35716</a>	
Variant type	Single base exchange	
Gene region	CDS	
DNA changes	c.178T>C g.233T>C	

Protein conservation	Species	Match	Gene	AA Alignment
Human mutated	not conserved			60 IKRPWNAFIMWSKIERRKIMEQS
Ptroglobytes	all identical		<a href="#">ENSPTRG00000045174</a>	60 IKRPWNAFIMWSKIERRKIMEQS
Mmulatta	all identical		<a href="#">ENSMMUG00000038367</a>	60 IKRPWNAFIMWSKIERRKIMEQS
Fcatus	all identical		<a href="#">ENSFCAG00000039092</a>	60 IKRPWNAFIMWSKIERRKIMEQS
Mmusculus	all identical		<a href="#">ENSMUSG00000036382</a>	60 IKRPWNAFIMWSKIERRKIMEQS
Ggalus	no homologue			
Trubripes	all identical		<a href="#">ENSTRUG0000003243</a>	60 IKRPWNAFIMWSKIERRKIMEQS
Drerio	all identical		<a href="#">ENSDARG00000077811</a>	60 IKRPWNAFIMWSKIERRKIMEQS
Dmelanogaster	no homologue			
Celegans	no homologue			
Xtrropicalis	all identical		<a href="#">ENSXETG00000038964</a>	60 IKRPWNAFIMWSKIERRKIMEQS

Original gDNA sequence snippet	TGAACCGCGTTATGGTATGG <b>T</b> CCAAGATCGAACGCAGGAAG
Altered gDNA sequence snippet	TGAACCGCGTTATGG <b>T</b> CCAAGATCGAACGCAGGAAG
Original cDNA sequence snippet	TGAACCGCGTTATGGTATGG <b>T</b> CCAAGATCGAACGCAGGAAG
Altered cDNA sequence snippet	TGAACCGCGTTATGG <b>T</b> CCAAGATCGAACGCAGGAAG
Wildtype AA sequence	MVQQAESLEA ESNLPREALD TEEGEFIMACS PVALDESPPD WCKTASGHIK RPMNAFIMWSKIERRKIMEQS SPDMHNAEIS KRLGKRWKLH KDSKIPFIR EAERRLKHM ADYPDYKYRP RKKPKMDPSA KPSSQSPEK SAAAGGGGSA GGGAGGAKTS KGSKKCGKL KAPAAAGAKA GAGKAAQSGD YGGAGDDYVL GSLRVSGSGG GGAGKTVKCV FLDEDDDDDD DDDDELQLQTK QEPDEEDEEP PHQQLLQPQPG QQPSSLRRY IVAKVPASTP LSSSAESPEG ASLYDEVRAZ ATSGAGGGSR LYYSFKNITK QHPPPLAQPA LSPASSRSVS TSSSSSSGSS SGSSGEDAD D LMFDLSLNFS QSAHSASEQQ LGGGAAAGNL SLSLVDKDLD SFSEGSLGH FEFPDYCTPE LSEMIAGDNL EANFSDLVFT Y*
Mutated AA sequence	MVQQAESLEA ESNLPREALD TEEGEFIMACS PVALDESPPD WCKTASGHIK RPMNAFIMWSKIERRKIMEQS SPDMHNAEIS KRLGKRWKLH KDSKIPFIR EAERRLKHM ADYPDYKYRP RKKPKMDPSA KPSSQSPEK SAAAGGGGSA GGGAGGAKTS KGSKKCGKL KAPAAAGAKA GAGKAAQSGD YGGAGDDYVL GSLRVSGSGG GGAGKTVKCV FLDEDDDDDD DDDDELQLQTK QEPDEEDEEP PHQQLLQPQPG QQPSSLRRY IVAKVPASTP LSSSAESPEG ASLYDEVRAZ ATSGAGGGSR LYYSFKNITK QHPPPLAQPA LSPASSRSVS TSSSSSSGSS SGSSGEDAD D LMFDLSLNFS QSAHSASEQQ LGGGAAAGNL SLSLVDKDLD SFSEGSLGH FEFPDYCTPE LSEMIAGDNL EANFSDLVFT Y*

## Variant - c.355C>T (p.Arg119Trp)

**PolyPhen 2** – Predicted the variant to be a probably damaging kind with a score of 1.000.



## SIFT - Sorting Intolerant From Tolerant

### Predictions for positions 101 through 200

Threshold for intolerance is 0.05.  
 Amino acid color code: nonpolar, uncharged polar, basic, acidic.  
 Capital letters indicate amino acids appearing in the alignment, lower case letters result from prediction.  
 'Seq Rep' is the fraction of sequences that  
 contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information. Expect poor prediction at these position

Predicted the variant under not tolerated

Predict Not Tolerated	Position	Seq	Rep	Predict Tolerated
..				
cwdpgvynhtIASFEQ	110M	0.97	LKRM	
wyfcmHpvIgLNTS	111A	0.97	QRDEKA	
cwfyivhgntspRMQAL	112D	0.97	EKD	
wcmpqeilrktdgsalFVN	113Y	0.96	HY	
whyfimnqrdekkcvtgSA	114P	0.96	P	
miwvflcyrpakQTSHLEG	115D	0.96	ND	
kqhnrdegttsamvilfCW	116Y	0.96	Y	
cdfmiygphnlateqVWSR	117K	0.96	K	
cmdhineeqptvlrsqFKWA	118Y	0.94	Y	
wdfmiygpsaeHNLTVC <span style="outline: 2px solid red;">K</span>	119R	0.94	OR	
wmifcliyvhqretagdNSK	120P	0.87	P	
cwdfmiygpsnalteHQVK	121R	0.79	R	
cwfmiyvgghslatneqPD	122K	0.76	RK	
cwfdfmiyvshgpnalteq	123K	0.68	RK	
wfymchdilegnQR	124P	0.47	SKVATP	
cwfdfmiyvgphlaeNTRSQ	125K	0.47	K	
wcfh	126M	0.30	yPlIdqeNAGKRSMTV	
cwfmyihvgtnRLEKS	127D	0.25	PAQD	
wfmcyihvrntgELSADK	128P	0.26	QP	
wcmhipvlqdFYRGGEN	129S	0.24	TKAS	
wmfiyckevHpnDQTG	130A	0.23	SLA	

**MUTATION TASTER**– This software predicts the pathogenicity of the variant. It predicted the variant as a deleterious variant.



## mutation t@sting

Prediction: **Deleterious** [Permalink](#)

Summary: • Amino acid sequence changed

- Model: simple\_aae
- Tree vote: 64|36 (del | benign) [?](#)

Analysed issue	Analysis result
Phys. location	chr2:5833208C>T <a href="#">show variant in all transcripts</a> <a href="#">IGV</a>
Gene symbol	<a href="#">SOX11</a>
ExAC LOF metrics	LOF: 0.34, missense: 4.25, synonymous: 2.64
Ensembl transcript ID	<a href="#">ENST00000322002.3</a>
Genbank transcript ID	<a href="#">NM_003108 (exact from MANE)</a>
UniProt peptide	<a href="#">P35716</a>
Variant type	Single base exchange
Gene region	CDS
DNA changes	c.355C>T g.410C>T

Protein conservation	Species	Match	Gene	AA Alignment
Human	not conserved			119 KHADYDPYKVRPRKKPKIDPSAK
Pygmychimpanzee	all identical	<a href="#">ENSPTRG00000045174</a>		119 KHADYDPYKVRPRKKPKIDPSA
Chimpanzee	all identical	<a href="#">ENSMUG00000038367</a>		119 KHADYDPYKVRPRKKPKIDPSA
Orangutan	all identical	<a href="#">ENSFCAG00000039092</a>		119 KHADYDPYKVRPRKKPKIDPSA
Bornean orangutan	all identical	<a href="#">ENSMUSG00000063632</a>		119 KHADYDPYKVRPRKKPKTDPA
Gorilla	no homologue			
Trichloracanthus	all identical	<a href="#">ENSTRUG0000003243</a>		116 KHADYDPYKVRPKKKPKLDS-
Drepanocephalus	all identical	<a href="#">ENSDARG00000077811</a>		111 KHADYDPYKVRPKKKPKLDS
Drosophila melanogaster	no homologue			
Celegans	no homologue			
Xtropicalis	all identical	<a href="#">ENSXETG00000038954</a>		118 KHADYDPYKVRPRKKPKVDPSA

Original gDNA sequence snippet	ACTACCCGACTACAAGTA <b>C</b> GGCCCCGGAAAAAGCCCCAA
Altered gDNA sequence snippet	ACTACCCGACTACAAGTA <b>T</b> GGCCCCGGAAAAAGCCCCAA
Original cDNA sequence snippet	ACTACCCGACTACAAGTA <b>C</b> GGCCCCGGAAAAAGCCCCAA
Altered cDNA sequence snippet	ACTACCCGACTACAAGTA <b>T</b> GGCCCCGGAAAAAGCCCCAA
Wildtype AA sequence	MVQQAESLEA ESNLPREALD TEEGEFINAC SVALDESOPD WCKTAGSHIK RPINNAFIVMS KIERRKIMEQ SPDMHNAEIS KRLGRKRMML KDEKEIPFIR EAERLRLKHM ADYDPOYKRP RKKPKVIDPSA KPSASQSPEK SAAGGGGGSA GGGAGGAITS KGSSKKCGKL KAPAAAAGAKA GAGKAQSGD YGGAGQDGYVL GSLRVSGSGG GGAGKTVKCV FLDEDDDDDD DDDDELQLQIK QEPDQEDEEP PHQQLQPRG QQPQLLRV NVAKVPASPT LSSSAESPEG ASLYDEVRAG ATSGAGGGSR LYYSFKNITK QHPPPLAQPA LSPASSRSVS TSSSSSGSS SGSSGEDADD LMFDSLNSFS QSAHSASEQQ LGGGAAAGNL SLSLVDKLDL SFSEGSLGSH FEFPDYCTPE LSEMIAGWNL EAHFSDLVFT Y*
Mutated AA sequence	MVQQAESLEA ESNLPREALD TEEGEFINAC SVALDESOPD WCKTAGSHIK RPINNAFIVMS KIERRKIMEQ SPDMHNAEIS KRLGRKRMML KDEKEIPFIR EAERLRLKHM ADYDPOYK <b>T</b> P RKKPKVIDPSA KPSASQSPEK SAAGGGGGSA GGGAGGAITS KGSSKKCGKL KAPAAAAGAKA GAGKAQSGD YGGAGQDGYVL GSLRVSGSGG GGAGKTVKCV FLDEDDDDDD DDDDELQLQIK QEPDQEDEEP PHQQLQPRG QQPQLLRV NVAKVPASPT LSSSAESPEG ASLYDEVRAG ATSGAGGGSR LYYSFKNITK QHPPPLAQPA LSPASSRSVS TSSSSSGSS SGSSGEDADD LMFDSLNSFS QSAHSASEQQ LGGGAAAGNL SLSLVDKLDL SFSEGSLGSH FEFPDYCTPE LSEMIAGWNL EAHFSDLVFT Y*

## Variant - c.170T>C (p.Met57Thr)

**PolyPhen 2-** Predicted the variant to be a possibly damaging kind with a score of 0.996.



## SIFT Sorting Intolerant From Tolerant

Predicted the variant under not tolerated

### Predictions for positions 1 through 100

Threshold for intolerance is 0.05.  
Amino acid color code: nonpolar, uncharged polar, basic, acidic.  
Capital letters indicate amino acids appearing in the alignment, lower case letters result from prediction.  
'Seq Rep' is the fraction of sequences that  
contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information. Expect poor prediction at th

Predict Not Tolerated	Position	Seq	Rep	Predict Tolerated
cwdfmiyvgshnlateqP	50K	0.90	RK	
cwfmdiyyvshnatlpgeqK	51R	0.92	R	
ywvtsrqnmklkihgfedca	52P	0.93	P	
dhcwnyegsqrtafvikPL	53M	0.93	M	
mwifvclyrqhptkeagdS	54N	0.93	N	
whyfimnqrdelkvtpSCG	55A	0.93	A	
kqhnrdrgepctsamvilWY	56F	0.93	F	
<b>dhgneccswrkpyqtaVLF</b>	57M	0.96	<b>IM</b>	
hdqnpecrgskytafmWI	58V	0.96	LV	
kqhnrdrgepctsamvil	59W	0.96	YW	
whyfiqndelkvtpgCMR	60S	0.96	AS	
wfmiyvpghlatneCDS	61K	0.96	RQK	
wcyfM	62I	0.96	PLVHNGRqTSKADEI	
cwdpigstvlNKFYMA	63E	0.96	HRQE	
wdfmiyvgpsnlteACKQH	64R	0.96	R	
cwdfmiysnlteqGHVPA	65R	0.96	KR	
wcfypvgsNIDTHLMEQAR	66K	0.96	K	
hqpwdencrsgatfKY	67I	0.96	MVLI	
wpDenGsCfQIyRTVHLK	68M	0.96	AM	
cwfymihgrnptV	69E	0.98	KASQDLE	
wycfhivpgnDMTRKSAL	70Q	1.00	EQ	

**MUTATION TASTER**-This software predicts the pathogenicity of the variant. It predicted the variant as a deleterious variant.



## mutation t@sting

Prediction:
**Deleterious**
[Permalink](#)

Summary:

- Amino acid sequence changed
- Protein features (might be) affected

- Model: simple\_aae
- Tree vote: 92|8 (del | benign)

Analysed issue	Analysis result
Phys. location	chr2:5833023T>C <a href="#">show variant in all transcripts</a> <a href="#">IGV</a>
Gene symbol	<b>SOX11</b>
ExAC LOF metrics	LOF: 0.34, missense: 4.25, synonymous: 2.64
Ensembl transcript ID	<a href="#">ENST00000322002.3</a>
Genbank transcript ID	<a href="#">NM_003108 (exact from MANE)</a>
UniProt peptide	<a href="#">P35716</a>
Variant type	Single base exchange
Gene region	CDS
DNA changes	c.170T>C g.225T>C

Protein conservation	Species	Match	Gene	AA Alignment
	Human	not conserved		57 SGHIKRPNIAFVNWSKIERRKIME
	mutated			57 SGHIKRPNIAFTVWS
	Ptroglobutes	all identical	<a href="#">ENSPTRG00000045174</a>	57 SGHIKRPNIAFVNWS
	Mmulatta	all identical	<a href="#">ENSMMUG000000038367</a>	57 SGHIKRPNIAFVNWS
	Fcatus	all identical	<a href="#">ENSFCAG00000039092</a>	57 SGHIKRPNIAFVNWS
	Mmusculus	all identical	<a href="#">ENSMUSG00000063632</a>	57 SGHIKRPNIAFVNWS
	Ggallus	no homologue		
	Trubripes	all identical	<a href="#">ENSTRUG00000003243</a>	57 TGHIKRPNIAFVNWS
	Drerio	all identical	<a href="#">ENS DARG00000077811</a>	57 TGHIKRPNIAFVNWS
	Dimelanogaster	no homologue		
	Celegans	no homologue		
	Xtropicalis	all identical	<a href="#">ENS XETG00000038964</a>	57 TGHIKRPNIAFVNWS

Original gDNA sequence snippet	GGGGCGATGAACCGCTTC <b>A</b> TGGTATGGTCCAAGATCGAAC
Altered gDNA sequence snippet	GGGGCGATGAACCGCTTC <b>C</b> GGTATGGTCCAAGATCGAAC
Original cDNA sequence snippet	GGGGCGATGAACCGCTTC <b>A</b> TGGTATGGTCCAAGATCGAAC
Altered cDNA sequence snippet	GGGGCGATGAACCGCTTC <b>C</b> GGTATGGTCCAAGATCGAAC
Wildtype AA sequence	MVQQAESLEA ESNLPRLEALD TEEGEPMAC SVALDESDPW ICKTASGHIZK RPWNAFVNWS KIERRKIMEQ SPDNHHAELIS KRLGKRWKML KDEKIPFIR EAERRLKHM ADYPDVKYRP RKKPKVIDPSA KPSASQSEPK SAAGGGGSSA GGGAGGGAKTS KGSSKKGKL KAPAAAGAKA GAGKAAQSGD YGGAGDDYVL GSLRVSGSGG GGAGKTVKCV FLDDEDDDDO DDEDELQLQIK QEPQEEDEEP PHQQLLQPQPG QQPSSLRLRY IVAKVPASTP LSSSAESPEG ASLYDEVRAAG ATSGAGGGSR LYYSFKNITK QHPPPLAQPA LSPASSRSVS TSSSSSSSGSS SGSSGEDADD LMFDLSNFS QSAHSASEQQ LGGGAAAGNL SLSLVDKLD SFSEGSGLSH FEPPDYCTPE LSEMIAGDWL EANFSDLVFT Y*
Mutated AA sequence	MVQQAESLEA ESNLPRLEALD TEEGEPMAC SVALDESDPW ICKTASGHIZK RPWNAFTVWS KIERRKIMEQ SPDNHHAELIS KRLGKRWKML KDEKIPFIR EAERRLKHM ADYPDVKYRP RKKPKVIDPSA KPSASQSEPK SAAGGGGSSA GGGAGGGAKTS KGSSKKGKL KAPAAAGAKA GAGKAAQSGD YGGAGDDYVL GSLRVSGSGG GGAGKTVKCV FLDDEDDDDO DDEDELQLQIK QEPQEEDEEP PHQQLLQPQPG QQPSSLRLRY IVAKVPASTP LSSSAESPEG ASLYDEVRAAG ATSGAGGGSR LYYSFKNITK QHPPPLAQPA LSPASSRSVS TSSSSSSSGSS SGSSGEDADD LMFDLSNFS QSAHSASEQQ LGGGAAAGNL SLSLVDKLD SFSEGSGLSH FEPPDYCTPE LSEMIAGDWL EANFSDLVFT Y*

## Variant - c.250G>C (p.Gly84Arg)

**PolyPhen 2** -Predicted the variant to be a possibly damaging kind with a score of 1.000.



## SIFT Sorting Intolerant From Tolerant

Predicted the variant under not tolerated

### Predictions for positions 1 through 100

Threshold for intolerance is 0.05.  
 Amino acid color code: nonpolar, uncharged polar, basic, acidic.  
 Capital letters indicate amino acids appearing in the alignment, lower case letters result from prediction.  
 'Seq Rep' is the fraction of sequences that contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information. Expect poor prediction at th

Predict Not Tolerated	Position	Seq	Rep	Predict Tolerated
dwgcyhseprfvTNAQKI	74M	1.00	LM	
wmcfivlyagRKTSPEND	75H	1.00	QH	
mwfvclyrpqaektgDHIS	76N	1.00	N	
wfmyihclevkpdTQGNR	77A	1.00	SA	
cwfmyilhrgpkVSNQTA	78E	1.00	DE	
hwqdpnecrkgyatfMSV	79I	1.00	LI	
wfmyhirclqekvdptgAN	80S	1.00	S	
cwdfmiygpshnalETQVR	81K	1.00	K	
cfWdyvgphnALETKS	82R	1.00	MIQR	
gwhdynqeskfmaPRCIVT	83L	1.00	L	
wmfiyhrc1lqkvtpdnEas	84G	1.00	G	
wfmipHVLNTSGDC	85K	1.00	QAERYK	
cyfhvPnGTAIDWKSL	86R	1.00	MQER	
mhcqinvtekrdalypfgS	87W	1.00	W	
cwdfmivgpsaltEYNHQ	88K	1.00	RK	
wcfyhipVtQGDE	89M	1.00	SNKARLM	
dhgncswyrkpqtafvIME	90L	1.00	L	
wfcymihvlPRAe	91K	1.00	DGQSTNK	
wycfmhvgTAQSILRKN	92D	1.00	PED	
wfcmhYiprVLGQKDN	93S	1.00	TEAS	
cwfmiylhrnnpkGSTAVQ	94E	1.00	DE	

**MUTATION TASTER**-This software predicts the pathogenicity of the variant. It predicted the variant as a deleterious variant.

**Mutation Taster**

## mutation t@sting

Prediction: **Deleterious** [Permalink](#)

**Summary:**

- Amino acid sequence changed
- Protein features (might be) affected
- Model: simple\_aae
- Tree vote: 96|4 (del | benign) [?](#)

Analysed issue	Analysis result
Phys. location	chr2:5833103G>C <a href="#">show variant in all transcripts</a> <a href="#">IGV</a>
Gene symbol	<a href="#">SOX11</a>
EXAC LOF metrics	LOF: 0.34, missense: 4.25, synonymous: 2.64
Ensembl transcript ID	<a href="#">ENST00000322002</a>
Genbank transcript ID	<a href="#">NM_003108 (exact from MANE)</a>
UniProt peptide	<a href="#">P35716</a>
Variant type	Single base exchange
Gene region	CDS
DNA changes	c.250G>C g.305G>C

Protein conservation	Species	Match	Gene	AA Alignment
	Human	not conserved		84 DMHNAEISKRLKRKMLKDKSEK
	Pyroglodytes	all identical	<a href="#">ENSPTRG00000045174</a>	84 DMHNAEISKRLKRKMLKDKSEK
	Mmulatta	all identical	<a href="#">ENSMMUG00000038367</a>	84 DMHNAEISKRLKRKMLKDKSEK
	Fcatus	all identical	<a href="#">ENSFCAQ00000039092</a>	84 DMHNAEISKRLKRKMLKDKSEK
	Mmusculus	all identical	<a href="#">ENSMUSG00000063632</a>	84 DMHNAEISKRLKRKMLKDKSEK
	Ggallus	no homologue		
	Trubripes	all identical	<a href="#">ENSTRUG00000003243</a>	81 DMHNAEISKRLKRKMLKDKSEK
	Dreario	all identical	<a href="#">ENSDARG00000077811</a>	76 DMHNAEISKRLKRKMLKDKSEK
	Dmelanogaster	no homologue		
	Celegans	no homologue		
	Xtropicalis	all identical	<a href="#">ENSXETG00000038964</a>	83 DMHNAEISKRLKRKMLNDEK

Original gDNA sequence snippet	CCGAGATCTCCAAGAGGGCTGGCAAGCGCTGGAAAATGCTG
Altered gDNA sequence snippet	CCGAGATCTCCAAGAGGGCTGGCAAGCGCTGGAAAATGCTG
Original cDNA sequence snippet	CCGAGATCTCCAAGAGGGCTGGCAAGCGCTGGAAAATGCTG
Altered cDNA sequence snippet	CCGAGATCTCCAAGAGGGCTGGCAAGCGCTGGAAAATGCTG
Wildtype AA sequence	IVQQAESLEA ESNLPREALD TEEGEFIMACS PVALDESDDP WCKTASGHIK RPINNAFIWWS KIERRKJMEQ SPDMHNAEIS KRLGKRKMLK KOSKEKIPFIR EAERLRLKHM ADVPDYKYP RKKPKNDPSA KPSASQ\$PEK SAAGGGGGSA GGGAGGAKTS KGSSKKCGKL KAPAAAGAKA GAGKAAQSGD YGGAGDDYVL GSLRVSGSGG GGAGKTVKCV FLDEDODDD DDEDELQLQIK QEPDDEEDEP PHQQLQPPG QQPSQLLRRY IVAKVPASPT LSSSAESPEG ASLYDEVRAAG ATSGAGGGSR LYYSFKNITK QHPPLAOPA LSPASSRVS TSSSSSSGSS SGSSGEDADD LMFDLSLNSFS QSAHSASEQQ LGGGAAAGNL SLSLVDKDLD SFSEGSLGSH FFPDYCTPE LSEMIAGDWL EAIFSDLVFT Y*
Mutated AA sequence	IVQQAESLEA ESNLPREALD TEEGEFIMACS PVALDESDDP WCKTASGHIK RPINNAFIWWS KIERRKJMEQ SPDMHNAEIS KRLRKRMNL KOSKEKIPFIR EAERLRLKHM ADVPDYKYP RKKPKNDPSA KPSASQ\$PEK SAAGGGGGSA GGGAGGAKTS KGSSKKCGKL KAPAAAGAKA GAGKAAQSGD YGGAGDDYVL GSLRVSGSGG GGAGKTVKCV FLDEDODDD DDEDELQLQIK QEPDDEEDEP PHQQLQPPG QQPSQLLRRY IVAKVPASPT LSSSAESPEG ASLYDEVRAAG ATSGAGGGSR LYYSFKNITK QHPPLAOPA LSPASSRVS TSSSSSSGSS SGSSGEDADD LMFDLSLNSFS QSAHSASEQQ LGGGAAAGNL SLSLVDKDLD SFSEGSLGSH FFPDYCTPE LSEMIAGDWL EAIFSDLVFT Y*

## DPF2

**Variant** -c.1066T>G (p.Cys356Gly)

**PolyPhen 2** - Predicted the variant to be a possibly damaging kind with a score of 1.000.



## SIFT Sorting Intolerant From Tolerant

Predicted the variant under not tolerated.

### Predictions for positions 301 through 400

Threshold for intolerance is 0.05.  
Amino acid color code: nonpolar, uncharged polar, basic, acidic.  
Capital letters indicate amino acids appearing in the alignment, lower case letters result from prediction.  
'Seq Rep' is the fraction of sequences that contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information. Expect poor prediction at the end of the sequence.

Predict Not Tolerated	Position	Seq	Rep	Predict Tolerated
-----------------------	----------	-----	-----	-------------------

wcmfiyvlrhttpsakgqne	346D	0.99	D
wycfhiplMnraQVK	347D	0.99	ETGSD
ywvtsrqpnmlkihgfed	348C	0.99	C
mwivflcyrqhatkegSNP	349D	0.99	D
cwfyyvhpglntrqAMISDK	350R	0.99	R
whyfimqreldckTPVN	351G	0.99	ASG
wpmdeqkrngtisvalHC	352Y	0.99	FY
ywvtsrqpnmlkigfedca	353H	0.99	H
dghnwyeersckqpfvA	354M	0.99	ITLM
cwmpngrqitshK1DAV	355Y	0.99	EFY
ywvtsrqpnmlkihgfed	356C	0.99	C
dgwnyreskqptafmHC	357L	0.99	VIL
wyfcmh	358T	0.89	IpgLQVaREDKNTS
ywvtsrqpnmlkihgfedca	359P	0.89	P
wyfcmh1VnIqtDREK	360S	0.99	SGAP
dhgnecksrykpqtaf	361M	0.99	IMVL
wc	362S	0.99	MYHiFPVLGNRQDTKSEA
wfcYmhIplgVND	363E	0.99	SRTAQKE
gwyhdnraqsekfc	364P	0.99	MtLIAVP
wmyhficqnrllevkdtgSa	365P	0.99	P
wycfhplgInVtqMRA	366E	0.99	SKDE

**MUTATION TASTER**-This software predicts the pathogenicity of the variant. It predicted the variant as a deleterious variant.



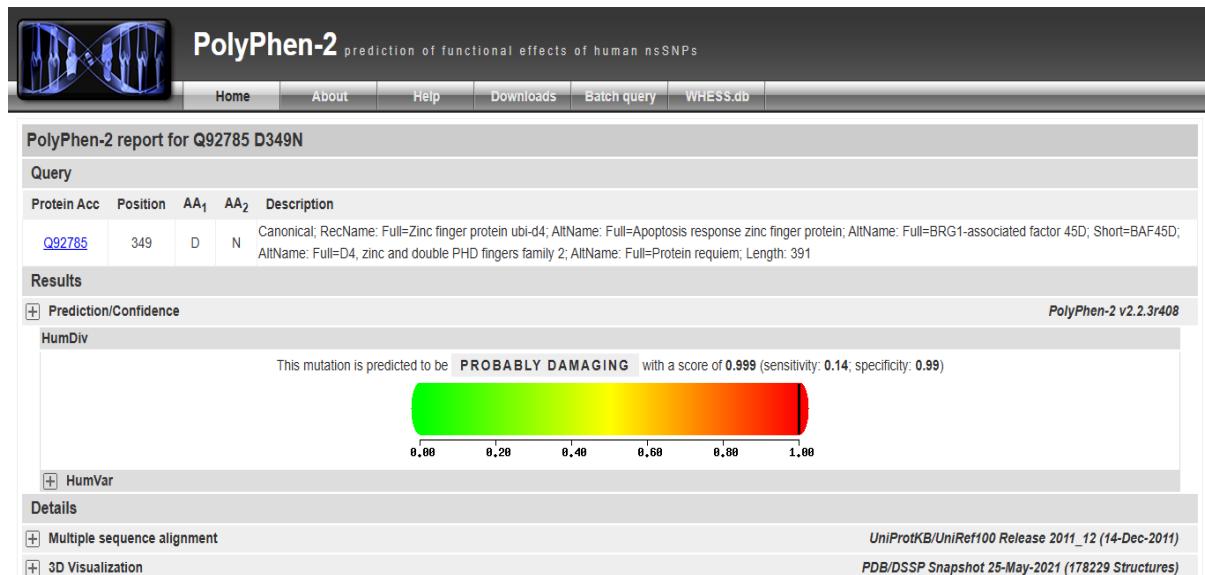
Prediction:	<b>Deleterious</b>	<a href="#">Permalink</a>
Summary:	<ul style="list-style-type: none"> <li>Amino acid sequence changed</li> <li>Protein features (might be) affected</li> </ul>	<ul style="list-style-type: none"> <li>Model: simple_aae</li> <li>Tree vote: 98 2 (del   benign) <a href="#">?</a></li> </ul>
<hr/>		
Analysed issue	Analysis result	
Phys. location	chr11:65116369T>G <a href="#">show variant in all transcripts</a> <a href="#">IGV</a>	
Gene symbol	<a href="#">DPF2</a>	
ExAC LOF metrics	LOF: 1.00, missense: 3.30, synonymous: -0.88	
Ensembl transcript ID	<a href="#">ENST00000528416.1</a>	
Genbank transcript ID	<a href="#">NM_006268 (exact from MANE)</a>	
UniProt peptide	<a href="#">Q92785</a>	
Variant type	Single base exchange	
Gene region	CDS	
DNA changes	c.1066T>G g.15145T>G	

Protein conservation	Species	Match	Gene	AA	Alignment
	Human	not conserved		356	CDCCDGYH <del>Y</del> CLTPSMSEPPEGS
mutated				356	CDCCDGYH <del>Y</del> MLTPSMSEPPEG
Ptroglyotes	all identical		<a href="#">ENSPTRG00000003872</a>	370	DCDGYH <del>Y</del> MLTPSMSEPPEG
Mmullata	no homologue				
Fcatus	all identical		<a href="#">ENSFCAG00000023858</a>	370	DCDGYH <del>Y</del> MLTPSMSEPPEG
Mmusculus	all identical		<a href="#">ENSMUSG00000024826</a>	356	CDCCDGYH <del>Y</del> MLTPS
Ggallus	no homologue				
Trubripes	all identical		<a href="#">ENSTRUG00000014692</a>	328	-----GYH <del>Y</del> MLSPPMTEPPEG
Dreiro	no alignment		<a href="#">ENSDARG00000012219</a>	n/a	
Dmelanogaster	all identical		<a href="#">FBgn0033015</a>	313	CDCCDGYH <del>Y</del> MLSPPLVTPPEG
Celegans	all identical		<a href="#">WBGene00016200</a>	340	CDCCDGYH <del>Y</del> MLPALEKAPDD
Xtropicalis	all identical		<a href="#">ENSXETG00000041314</a>	309	CDCCDGYH <del>Y</del> MLSPPMAEPEPPEG

Original gDNA sequence snippet	ATCGTGGCTTACCACTGTACTGTCTACCCCCGTCATGTCT
Altered gDNA sequence snippet	ATCGTGGCTTACCACTGTAC <del>G</del> GTCTACCCCCGTCATGTCT
Original cDNA sequence snippet	ATCGTGGCTTACCACTGTACTGTCTACCCCCGTCATGTCT
Altered cDNA sequence snippet	ATCGTGGCTTACCACTGTAC <del>G</del> GTCTACCCCCGTCATGTCT
Wildtype AA sequence	MAAWENVK LLGEQYYKDA MEQCHNINAR LCAERSVRLP FLDSTQVVAQ SNCYIWMERK HRGPGLASGQ LYSYPARRWR KKRAHPPED PRLSFPSIKP DTQDTLKKEG LISQDGSSLE ALLRTDPLEK RGADPDRVDD DSLGEFPVTN SRARKRILEP DDFLDDLODE DYEEOTPKRR GKKGSKKGKV GSARKLKDAS ILEDROKPYA C0ICGKRYKN RPGLSYHYAH SHLAEEEGED KEDSQPTPV SQSEEQKSK KGPDLALPN NYCDFCLGDS KINKKTGQPE ELVSCSDCRG SGHPSCLQFT PVWMAAVKY R1QJCIECKC NICGTSENDD QLLFCDCOR GYH <del>Y</del> MLTPS MSEPPEGWS CHLCLDLIKE KASIYQNQNS S*
Mutated AA sequence	MAAWENVK LLGEQYYKDA MEQCHNINAR LCAERSVRLP FLDSTQVVAQ SNCYIWMERK HRGPGLASGQ LYSYPARRWR KKRAHPPED PRLSFPSIKP DTQDTLKKEG LISQDGSSLE ALLRTDPLEK RGADPDRVDD DSLGEFPVTN SRARKRILEP DDFLDDLODE DYEEOTPKRR GKKGSKKGKV GSARKLKDAS ILEDROKPYA C0ICGKRYKN RPGLSYHYAH SHLAEEEGED KEDSQPTPV SQSEEQKSK KGPDLALPN NYCDFCLGDS KINKKTGQPE ELVSCSDCRG SGHPSCLQFT PVWMAAVKY R1QJCIECKC NICGTSENDD QLLFCDCOR GYH <del>Y</del> MLTPS MSEPPEGWS CHLCLDLIKE KASIYQNQNS S*

## Variant c.1045G>A (p.Asp349Asn)

**PolyPhen 2** -Predicted the variant to be a possibly damaging kind with a score of 0.999.



## SIFT Sorting Intolerant From Tolerant

### Predictions for positions 301 through 400

Threshold for intolerance is 0.05.  
Amino acid color code: nonpolar, uncharged polar, basic, acidic.  
Capital letters indicate amino acids appearing in the alignment, lower case letters result from prediction.  
'Seq Rep' is the fraction of sequences that  
contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information. Expect poor prediction at these positions.

Predicted the variant under not tolerated

Predict	Not Tolerated	Position	Seq	Rep	Predict	Tolerated
---------	---------------	----------	-----	-----	---------	-----------

wcfmiyh1rtnVskqG	339D	0.92	APED
wcmfiylvrhhttpqnAKS	340D	0.95	GED
cwmyidvhpgtaFSL	341Q	0.95	NREKQ
dghnecswyrkpaTvFQI	342L	0.99	ML
dghnecswyrkpqt_afVI	343L	0.99	ML
hndkrqgqepstavCwiMY	344F	0.99	LF
ywvtsrqpnmlkihg_feda	345C	0.99	C
wcmfiyvlrhttpsakgqnE	346D	0.99	D
wycfhiplMnraQVK	347D	0.99	ETGSD
ywvtsrqpnmlkihg_feda	348C	0.99	C
<u>mwivflcyrqhatkegSNP</u>	349D	0.99	<u>D</u>
cwfyyvhpgln_teqAMISDK	350R	0.99	R
whyfimqreldckTPVN	351G	0.99	ASG
wpmdeqkrngtisvalHC	352Y	0.99	FY
ywvtsrqpnmlkihg_feda	353H	0.99	H
dghnwyersckqpfvA	354M	0.99	ITLM
cwmpngrqits_hK1DAV	355Y	0.99	EY
ywvtsrqpnmlkihg_feda	356C	0.99	C
dgwnyreskqptafmHC	357L	0.99	VIL
wyfcmh	358T	0.89	IpgLQVaREDKNTS
ywvtsrqpnmlkihg_feda	359P	0.89	P

**MUTATION TASTER**-This software predicts the pathogenicity of the variant. It predicted the variant as a deleterious variant.



## mutation t@sting

---

Prediction:
Deleterious
[Permalink](#)

Summary:

- Amino acid sequence changed
- Protein features (might be) affected

- Model: simple\_aae
- Tree vote: 52|48 (del | benign) ?

Analysed issue	Analysis result
Phys. location	chr11:65116348G>A <a href="#">show variant in all transcripts</a> <a href="#">IGV</a>
Gene symbol	<a href="#">DPF2</a>
ExAC LOF metrics	LOF: 1.00, missense: 3.30, synonymous: -0.88
Ensembl transcript ID	<a href="#">ENST00000528416.1</a>
Genbank transcript ID	<a href="#">NM_006268</a> (exact from MANE)
UniProt peptide	<a href="#">Q92785</a>
Variant type	Single base exchange
Gene region	CDS
DNA changes	c.1045G>A g.15124G>A

Protein conservation	Species	Match	Gene	AA Alignment
	Human	all conserved		349 NDDQLLFCDDCDRGHYYCLTPS
	Human mutated	all conserved		349 NDDQLLFCDDCDRGHYYCLTPS
	Proglodytes	all identical	<a href="#">ENSPTRG00000003872</a>	363 DCGYHYYCLTPS
	Mimula	no homologue		
	Fcatus	all identical	<a href="#">ENSFCAG00000023858</a>	363 DCGYHYYCLTPS
	Mmusculus	all identical	<a href="#">ENSMUSG00000024826</a>	349 NDDQLLFCDDCDRGHYYCLTPS
	Ggallus	no homologue		
	Trubipes	not conserved	<a href="#">ENSTRUG00000014692</a>	321 -----GYHYYCLSPP
	Dreio	no alignment	<a href="#">ENSDARG00000012219</a>	n/a
	Dmelanogaster	all identical	<a href="#">FBgn0033015</a>	306 NDDQLLFCDDCDRGHYYCLSPP
	Celegans	all identical	<a href="#">WBGene00016200</a>	333 KLLFCDDCDRGHYYCLTPA
	Xtropicalis	all identical	<a href="#">ENSXETG00000041314</a>	302 NDDQLLFCDDCDRGHYYCLSPP

Original gDNA sequence snippet	TGCTCTTCTGTGATGACTGCC <b>A</b> TCTGGCTTACACATGTAC
Altered gDNA sequence snippet	TGCTCTTCTGTGATGACTGC <b>A</b> ATCTGGCTTACACATGTAC
Original cDNA sequence snippet	TGCTCTTCTGTGATGACTGCC <b>A</b> TCTGGCTTACACATGTAC
Altered cDNA sequence snippet	TGCTCTTCTGTGATGACTGC <b>A</b> ATCTGGCTTACACATGTAC
Wildtype AA sequence	MAAVENWK LLGEQQYKDA MEQCHNINAR LCAERSVRLP FLDSTQVQA SNCYIWMERK HRPGGLASGQ LYSVPARRNR KKRRAHPPED PRLSFPSIKP TDQTLKKEG LISQDQSSLE ALLRTDPLEK RGADPDRVDD DSLGEFPVTIN SRARKRILEP DFDFLDDODE DYEDTPKRR GKGSKGKGKV GSARKKLDA SILEDROKPYA C0ICGKRYKN RPGLSYHYAH SHLAEEGED KEDSQPTPV SQRSSEEQSK KGPGLALPN NYCDFCLGDS KINKKTGQE ELVSCSDGR SGHPSCLQFT PVWMAAVKYI RWQCIECKCC NICGTSENDO QLLFCDCDR GYHYYCLTPS MSEPPGWS CHLC DLLKE KASIZQNQIS S*
Mutated AA sequence	MAAVENWK LLGEQQYKDA MEQCHNINAR LCAERSVRLP FLDSTQVQA SNCYIWMERK HRPGGLASGQ LYSVPARRNR KKRRAHPPED PRLSFPSIKP TDQTLKKEG LISQDQSSLE ALLRTDPLEK RGADPDRVDD DSLGEFPVTIN SRARKRILEP DFDFLDDODE DYEDTPKRR GKGSKGKGKV GSARKKLDA SILEDROKPYA C0ICGKRYKN RPGLSYHYAH SHLAEEGED KEDSQPTPV SQRSSEEQSK KGPGLALPN NYCDFCLGDS KINKKTGQE ELVSCSDGR SGHPSCLQFT PVWMAAVKYI RWQCIECKCC NICGTSENDO QLLFCDCDR GYHYYCLTPS MSEPPGWS CHLC DLLKE KASIZQNQIS S*

## Variant c.868G>A (p.Glu290Lys)

**PolyPhen 2-** Predicted the variant to be a possibly damaging kind with a score of 1.000.



## SIFT Sorting Intolerant From Tolerant

### Predictions for positions 201 through 300

Threshold for intolerance is 0.05.  
Amino acid color code: nonpolar, uncharged polar, **basic**, **acidic**.  
Capital letters indicate amino acids appearing in the alignment; lower case letters result from prediction.  
'Seq Rep' is the fraction of sequences that contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information. Expect poor prediction at these positions.

Predicted the variant under not tolerated

Predict Not Tolerated	Position	Seq	Rep	Predict Tolerated
-----------------------	----------	-----	-----	-------------------

whyfmicldnervtpQ	280S	0.33	GKAS	
	281K	0.32	wcmipFHVlrqytGasKDNE	
w	282I	0.29	cfyhplvgnIdQAREMTks	
wmifcvlyphqatekgdS	283N	0.29	RN	
cwfdfmiyvhpnglteqS	284K	0.31	ARK	
cwfmiydvpghsltq	285K	0.33	ERANK	
wfymchigdevpqna	286T	0.25	RKLST	
miwvflcyrpqathes	287G	0.27	DNKG	
	wdC	0.35	MfivygpsnHatLeORK	
whyfmirqlcknevtgD	289P	0.33	SAP	
<b>mwcfiyhvrltnspkqadG</b>	<b>290E</b>	<b>0.35</b>	<b>E</b>	
cwfmyilvhgtnkqRPS	291E	0.35	ADE	
dghnecswkrypqtaviF	292L	0.39	ML	
hqwpdnecrskgyatfM	293V	0.39	LIV	
wmfciydevnqpgtkLR	294S	0.39	HAS	
ywwtsrqpnmlkihgfed	295C	0.39	C	
	wmicflv	0.39	rqpYketgDnHAS	
wmcfilvyrhpakgsneT	297D	0.39	QD	
ywwtsrqpnmlkihgfed	298C	0.39	C	
wmifcvlyphakQEnSRT	299G	0.39	DG	
cwfmidyvpghlatesQK	300R	0.39	NR	

**MUTATION TASTER**-This software predicts the pathogenicity of the variant. It predicted the variant as a deleterious variant.



## mutation t@sting

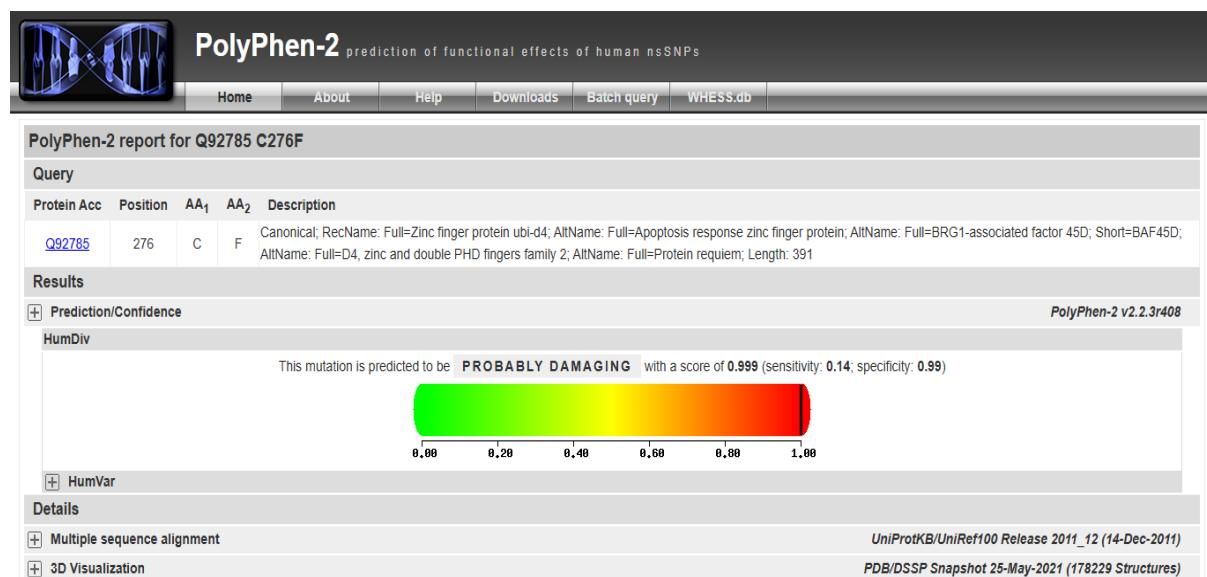
Prediction:	<b>Deleterious</b>	<a href="#">Permalink</a>
Summary:	<ul style="list-style-type: none"> <li>Amino acid sequence changed</li> <li>Protein features (might be) affected</li> </ul>	<ul style="list-style-type: none"> <li>Model: simple_aae</li> <li>Tree vote: 72 28 (del   benign) <a href="#">?</a></li> </ul>
Analysed issue	Analysis result	
Phys. location	chr11:65113493G>A <a href="#">show variant in all transcripts</a> <a href="#">IGV</a>	
Gene symbol	<a href="#">DPF2</a>	
ExAC LOF metrics	LOF: 1.00, missense: 3.30, synonymous: -0.88	
Ensembl transcript ID	<a href="#">ENST00000528416.1</a>	
Genbank transcript ID	<a href="#">NM_005268 (exact from MANE)</a>	
UniProt peptide	<a href="#">Q92785</a>	
Variant type	Single base exchange	
Gene region	CDS	
DNA changes	c.868G>A g.12269G>A	

Protein conservation	Species	Match	Gene	AA Alignment
	Human	all conserved		290 DSKINKKTGQPEELVSCSDCGRS
	mutated	all identical		290 DSKINKKTGQP <del>E</del> ELVSCSDGRS
	Ptroglydotes		<a href="#">ENSPTRG00000003872</a>	304 GQP <del>E</del> ELVSCSDGRS
	Mmulatta	no homologue		
	Fcatus	all identical	<a href="#">ENSFCAG00000023858</a>	304 GQP <del>E</del> ELVSCSDGRS
	Mmusculus	all identical	<a href="#">ENSMUSG00000024826</a>	290 DSKINKKTGQP <del>E</del> ELVSCSDGRS
	Ggalus	no homologue		
	Trubripes	all identical	<a href="#">ENSTRUG00000014692</a>	262 DSTLNQKTGQS <del>E</del> ELVSCSDGRS
	Drerio	no alignment	<a href="#">ENSDARG00000012219</a>	n/a
	Dmelanogaster	all identical	<a href="#">FBgn0033015</a>	247 DQRENKKTNMP <del>E</del> ELVSCSDGRS
	Celegans	all identical	<a href="#">WBGene00016200</a>	274 FMNKNTKLPE <del>D</del> LVSCHDCGRS
	Xtropicalis	all identical	<a href="#">ENSEXETG00000041314</a>	243 DSKINKKTNQ <del>E</del> ELVSCSDGRS

Original gDNA sequence snippet	ACAAGAAGACGGGACAACCC <b>G</b> AGGGAGCTGGTGTCTGTTCT
Altered gDNA sequence snippet	ACAAGAAGACGGGACAACCC <b>A</b> AGGGAGCTGGTGTCTGTTCT
Original cDNA sequence snippet	ACAAGAAGACGGGACAACCC <b>G</b> AGGGAGCTGGTGTCTGTTCT
Altered cDNA sequence snippet	ACAAGAAGACGGGACAACCC <b>A</b> AGGAAGCTGGTGTCTGTTCT
Wildtype AA sequence	MAAVVENVK LLGEQYYKDA MEQCHNNYAR LCAERSVRPL FLDQTGVHQ SNCYIWMERK HRGPGLASGQ LYSYPARRWR KKRRRAHPPED PRLSFPSIKP DTQDTLKGEG LISQDGSSLE ALLRTDPLEK RGAPDPRVDD DSLGEFPVTN SRARKRILEP DFFLDDLDE DYEEDTPKRR GKGSKGKGV GSARKKLDAS ILEDRDKPYA CDICGKRYKN RPGLSYHYAH SHLAEEGED KEDSQPPPTV SQRSEEQKSK KGPGLALPN NYCDFCLGDS KINKKTGQPK ELVSCSDGRS SGHPSCLQFT PVMMIAAVKTY RNQCIIECKCC NICGTSEND QLLFCDDCDR GHMYCLTPS MSEPPEGWS CHLCLDLLKE KASIYQNQNS S*
Mutated AA sequence	MAAVVENVK LLGEQYYKDA MEQCHNNYAR LCAERSVRPL FLDQTGVHQ SNCYIWMERK HRGPGLASGQ LYSYPARRWR KKRRRAHPPED PRLSFPSIKP DTQDTLKGEG LISQDGSSLE ALLRTDPLEK RGAPDPRVDD DSLGEFPVTN SRARKRILEP DFFLDDLDE DYEEDTPKRR GKGSKGKGV GSARKKLDAS ILEDRDKPYA CDICGKRYKN RPGLSYHYAH SHLAEEGED KEDSQPPPTV SQRSEEQKSK KGPGLALPN NYCDFCLGDS KINKKTGQPK ELVSCSDGRS SGHPSCLQFT PVMMIAAVKTY RNQCIIECKCC NICGTSEND QLLFCDDCDR GHMYCLTPS MSEPPEGWS CHLCLDLLKE KASIYQNQNS S*

## Variant c.827G>T (p.Cys276Phe)

**PolyPhen 2** -Predicted the variant to be a possibly damaging kind with a score of 0.999.



## SIFT Sorting Intolerant From Tolerant

### Predictions for positions 201 through 300

Threshold for intolerance is 0.05.  
Amino acid color code: nonpolar uncharged polar, **basic**, **acidic**.  
Capital letters indicate amino acids appearing in the alignment; lower case letters result from prediction.  
'Seq Rep' is the fraction of sequences that contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information. Expect poor prediction at these position

Predicted the variant under not tolerated.

Predict Not Tolerated	Position	Seq	Rep	Predict Tolerated
wghydrnq	266L	0.27	fecpmKSaLTIV	
whydngqfekc	267A	0.31	mpSRtLIAV	
wchgynpst	268L	0.31	qfamDkEvRLI	
whymfiqcrlekedvtga	269P	0.31	SP	
wcmfyvlhprtqaksgd	270N	0.31	IEN	
mwifvlc	271N	0.31	yrqhtkPeAsDGN	
dnghecrkqstaw	272Y	0.31	mvPfILY	
ywvtsrqpnmlkihgfed	273C	0.33	C	
wmiflvcy rqph tkean	274D	0.33	SGD	
hdngrkeqcpstawmylv	275F	0.33	IF	
ywvtsrqpnmlkihgfed	276C	0.33	C	
hdwyncergkpQt fam	277L	0.33	iVSL	
wcfmiyvhlpdtesnqaR	278G	0.32	KG	
wmifclvyrpqh	279D	0.33	eaKSNTDG	
whyfmicldnervtpQ	280S	0.33	GKAS	
	281K	0.32	wcmipFHVLrqytGasKDNE	
w	282I	0.29	cfyhplvgnIdQAREMTks	
wmifcvlyphqatekgds	283N	0.29	RN	
cwfdfmiyvhpnglteqs	284K	0.31	ARK	
cwfmiydvpghsltq	285K	0.33	ERANK	
wfymchigdevpqna	286T	0.25	RKLST	

**MUTATION TASTER**-This software predicts the pathogenicity of the variant. It predicted the variant as a deleterious variant.



## mutation t@sting

Prediction:	<b>Deleterious</b>	<a href="#">Permalink</a>
Summary:	<ul style="list-style-type: none"> <li>Amino acid sequence changed</li> <li>Known disease mutation: ClinVar ID 438643 (pathogenic)</li> <li>Protein features (might be) affected</li> </ul>	<ul style="list-style-type: none"> <li>Model: simple_aae</li> <li>Tree vote: 90 10 (del   benign) </li> <li>Automatic classification due to ClinVar</li> </ul>
Analysed issue	Analysis result	
Phys. location	chr11:65113452G>T <a href="#">show variant in all transcripts</a> <a href="#">IGV</a>	
Gene symbol	<a href="#">DPF2</a>	
ExAC LOF metrics	LOF: 1.00, missense: 3.30, synonymous: -0.88	
Ensembl transcript ID	<a href="#">ENST00000528416.1</a>	
Genbank transcript ID	<a href="#">NM_006268 (exact from MANE)</a>	
UniProt peptide	<a href="#">Q92785</a>	
Variant type	Single base exchange	
Gene region	CDS	
DNA changes	c.827G>T g.12228G>T	

Protein conservation	Species	Match	Gene	AA Alignment
Human	Human	not conserved		276 GLALPNYYCDFCLGDSKINKKTGQ
mutated				276 GLALPNYYCDFCLGDSKINKKTG
Ptroglobutes	all identical		<a href="#">ENSPTRG0000003872</a>	290 GLALPNYYCDFCLGDSKINKKT
Mmulatta	no homologue			
Fcatus	all identical		<a href="#">ENSFAG0000023858</a>	290 GLALPNYYCDFCLGDSKINKKT
Mmusculus	all identical		<a href="#">ENSMUSG00000024826</a>	276 GLALPNYYCDFCLGDSKINKKTG
Ggallus	no homologue			
Tribripes	all identical		<a href="#">ENSTRUG00000014692</a>	248 GLALPNYYCDFCLGDSLNIQKTG
Drerio	no alignment		<a href="#">ENSDARG00000012219</a>	n/a
Dmelanogaster	all identical		<a href="#">FBgn0030315</a>	233 IAQPSPYCDFLGQRENKTN
Celegans	all identical		<a href="#">WBGene00016200</a>	279 -----CDPSGTA
Xtrropicalis	all identical		<a href="#">ENSXETG00000041314</a>	229 GLALPNYYCDFCLGDSKINKTN

Original gDNA sequence snippet	CAACAACACTGTGACTTCT <b>T</b> GCCTGGGGACTCAAAGATTA
Altered gDNA sequence snippet	CAACAACACTGTGACTT <b>T</b> TCCTGGGGACTCAAAGATTA
Original cDNA sequence snippet	CAACAACACTGTGACTT <b>T</b> GCCTGGGGACTCAAAGATTA
Altered cDNA sequence snippet	CAACAACACTGTGACTT <b>T</b> CTCTGGGGACTCAAAGATTA
Wildtype AA sequence	MAAVVENVK LLGEQYYKDA MEQCHNYNAR LCAERSVRLP FLDSQTGVQA SNCYIWMKR HRGPGLASGQ LYSYPARRWR KKRRAHIPPED PRLSFPSPK DTQQLKKEG LISQDGSSLE ALLRTDPLEK RGAPPDRVDD DSLGEFPVTN SRARKRILEP DDFLDDLDDE DYEDTPKRR GKGSKGKGKV GSARKKLADAS ILEDRKPKYA CDICGKRYKN RPGLSYHYAH SHLAEEEGED KEDSQOPTPV SQRSSEEQSKS KGPDGLALPN NYCDFCLGDS KINKKTGQPE ELVSCSDGR SGHPSCLQFT PVMMIAAVKTY RWQCIECKCC NICGTSEND QLLFCDDCDR GHYHMYCLTPS MSEPPEGSWS CHLCCLLKE KASIYQNQNS S*
Mutated AA sequence	MAAVVENVK LLGEQYYKDA MEQCHNYNAR LCAERSVRLP FLDSQTGVQA SNCYIWMKR HRGPGLASGQ LYSYPARRWR KKRRAHIPPED PRLSFPSPK DTQQLKKEG LISQDGSSLE ALLRTDPLEK RGAPPDRVDD DSLGEFPVTN SRARKRILEP DDFLDDLDDE DYEDTPKRR GKGSKGKGKV GSARKKLADAS ILEDRKPKYA CDICGKRYKN RPGLSYHYAH SHLAEEEGED KEDSQOPTPV SQRSSEEQSKS KGPDGLALPN NYCDF <b>L</b> GDS KINKKTGQPE ELVSCSDGR SGHPSCLQFT PVMMIAAVKTY RWQCIECKCC NICGTSEND QLLFCDDCDR GHYHMYCLTPS MSEPPEGSWS CHLCCLLKE KASIYQNQNS S*

## Variant c.990C>G (p.Cys330Trp)

**PolyPhen 2-** Predicted the variant to be a possibly damaging kind with a score of 1.000.



## SIFT Sorting Intolerant From Tolerant

### Predictions for positions 301 through 400

Threshold for intolerance is 0.05.  
 Amino acid color code: nonpolar, uncharged polar, basic, acidic.  
 Capital letters indicate amino acids appearing in the alignment, lower case letters result from prediction.  
 'Seq Rep' is the fraction of sequences that  
 contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information. Expect poor prediction at these positions.

Predicted the variant under not tolerated.

Predict Not Tolerated	Position	Seq	Rep	Predict Tolerated
wcdmpeqnrtvha	320Y	0.45	GfKILSY	
wcfym	321R	0.45	ihv1PtNsDAeQGKR	
yvtsrqpnmlkihgfedca	322W	0.45	W	
cwfmiyvhlpgrtnadksRE	323Q	0.56	Q	
whymiqreldnvptgKaFS	324C	0.59	C	
gdhnyqekcftRSPA	325I	0.67	VMLI	
wcmfyvlrtpasqIKHGND	326E	0.67	E	
whydfneqrpitvgMaLSK	327C	0.69	C	
wcdfmgphnslaqIrTVEY	328K	0.67	K	
wmhfergqd	329C	0.71	klpIYAVNCTS	
wwvtsrqpnmlkihgfed	330C	0.97	C	
wcyf	331N	0.97	MHpvLgQdRIaKNTES	
wcdpqngrEhtMF	332I	0.97	SYKAVIL	
wmhyfiqerlndvtpsgKa	333C	0.99	C	
wmifcvlyphteAQNSKD	334G	0.92	RG	
wyfcmhipeLnGaDV	335T	0.92	REKQST	
wyfhmiClvndeQRK	336S	0.92	TPGAS	
wyfcmlivplaTNGRSQK	337E	0.96	DHE	
wmciflypVtksAGQRE	338N	0.96	DHN	
wcfmiyhlrtnvskqG	339D	0.92	APED	
wcmfiylvrhftpqnAKS	340D	0.95	GED	

**MUTATION TASTER**-This software predicts the pathogenicity of the variant. It predicted the variant as a deleterious variant.



mutation t@sting

Prediction:	<b>Deleterious</b>	<a href="#">Permalink</a>
Summary:	<ul style="list-style-type: none"> <li>Amino acid sequence changed</li> <li>Known disease mutation: ClinVar ID 545683 (pathogenic)</li> <li>Protein features (might be) affected</li> </ul>	<ul style="list-style-type: none"> <li>Model: simple_aae</li> <li>Tree vote: 96 4 (del   benign) </li> <li>Automatic classification due to ClinVar</li> </ul>
Analysed issue	Analysis result	
Phys. location	chr11:65113803C>G <a href="#">show variant in all transcripts</a> <a href="#">IGV</a>	
Gene symbol	<a href="#">DPF2</a>	
EXAC LOF metrics	LOF: 1.00, missense: 3.30, synonymous: -0.88	
Ensembl transcript ID	<a href="#">ENST00000528416.1</a>	
Genbank transcript ID	<a href="#">NM_006268</a> (exact from MANE)	
UniProt peptide	<a href="#">Q92785</a>	
Variant type	Single base exchange	
Gene region	CDS	
DNA changes	c 990C>G g.12579C>G	

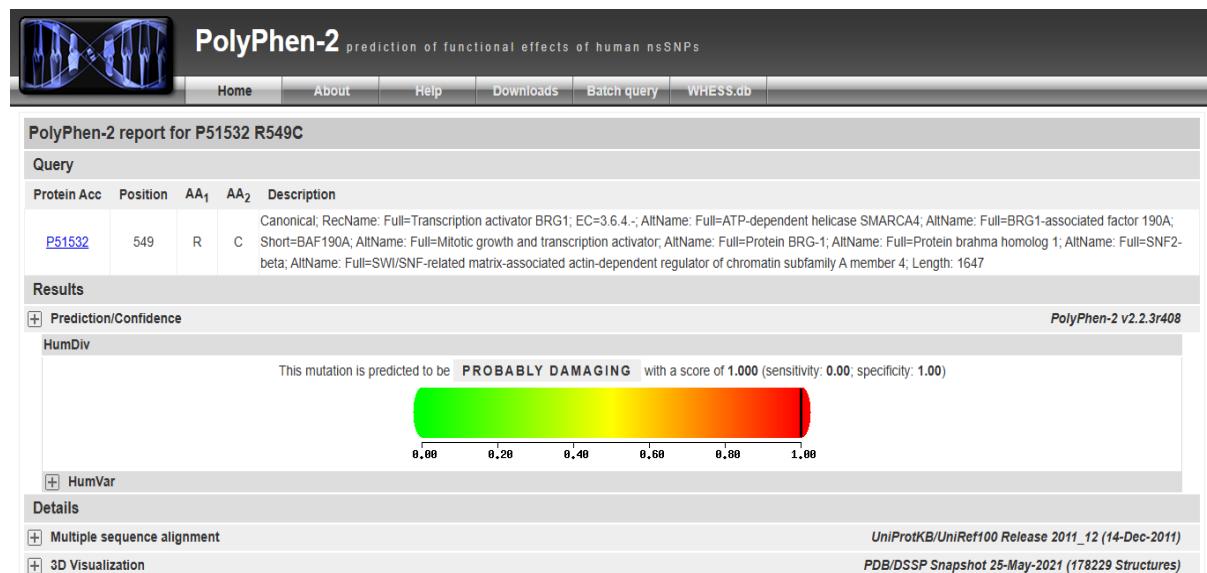
Protein conservation	Species	Match	Gene	AA	Alignment
	Human	not conserved		330	TYRWQ <del>CIECKCCNICGTSENDDQ</del>
	mutated			330	TYRWQ <del>CIECKCNICGTSENDDQ</del>
	Ptroglodytes	all identical	<a href="#">ENSPTRG0000003872</a>	344	TYRWQ <del>CIECKC<del>N</del>ICGTSENDDQ</del>
	Mmulatta	no homologue			
	Foatus	all identical	<a href="#">ENSFCAG00000023858</a>	344	TYRWQ <del>CIECKC<del>N</del>ICGTSENDDQ</del>
	Mmusculus	all identical	<a href="#">ENSMUSG00000024826</a>	330	TYRWQ <del>CIECKC<del>N</del>ICGTSENDDQ</del>
	Ggallus	no homologue			
	Tribripes	all identical	<a href="#">ENSTRUG00000014692</a>	302	YRWQ <del>CIECKC<del>N</del>FGC-----</del>
	Drerio	no alignment	<a href="#">ENSDAKG00000012219</a>	n/a	
	Dmelanogaster	all identical	<a href="#">FBgn0033015</a>	287	ECKY <del>S</del> ICGTSENDDQ
	Celegans	all identical	<a href="#">WBGene00016200</a>	314	RSGWQ <del>CLECKC<del>N</del>ICGTSENDDQ</del>
	Xtropicalis	all identical	<a href="#">ENSXETG00000041314</a>	283	TYRWQ <del>CIECKC<del>N</del>ICGTSENDDQ</del>

Original gDNA sequence snippet	TGCATCGAGTGCAAATGTTG <b>C</b> AATATCTGCGGCCACCTCCGA
Altered gDNA sequence snippet	TGCATCGAGTGCAAATGTT <b>G</b> AATATCTGCGGCCACCTCCGA
Original cDNA sequence snippet	TGCATCGAGTGCAAATGTT <b>G</b> AATATCTGCGGCCACCTCCGA
Altered cDNA sequence snippet	TGCATCGAGTGCAAATGTT <b>G</b> AATATCTGCGGCCACCTCCGA
Wildtype AA sequence	MAAVVENVK LLGEQYKDA MEQCHNLYNAR LCAERSVRLP FLDQTGVHQ SNCYIWMEKR HRGPGLASGQ LYSYPARRIR KKRRAHPPED PRLSFPSIKP DTDQTLKKEG LISQDGSSLE ALLRTDPLEK RGAPDPRVDD DSLGEFPVTN SRARKRILEP DDFLDDLODE DYEDTPKRR GIGKSIGKGV GSARKKLDA SILEDRKPYA CDTGCKRYKPN RPLGLSYHAYAH SHLAEEGED KEDSQPPTPV SQRSSEEQKSK KGPGLALPN NYCDFCLGDS KINKKTTGQE ELVSCSDGR SGHPSCLQFT PVWVAAVKTY RWQ <del>CIECKC<del>N</del>ICGTSENDDQ</del> MSEPPEGWSN CHLCLDLIKE KASIQYQNQS S*
Mutated AA sequence	MAAVVENVK LLGEQYKDA MEQCHNLYNAR LCAERSVRLP FLDQTGVHQ SNCYIWMEKR HRGPGLASGQ LYSYPARRIR KKRRAHPPED PRLSFPSIKP DTDQTLKKEG LISQDGSSLE ALLRTDPLEK RGAPDPRVDD DSLGEFPVTN SRARKRILEP DDFLDDLODE DYEDTPKRR GIGKSIGKGV GSARKKLDA SILEDRKPYA CDTGCKRYKPN RPLGLSYHAYAH SHLAEEGED KEDSQPPTPV SQRSSEEQKSK KGPGLALPN NYCDFCLGDS KINKKTTGQE ELVSCSDGR SGHPSCLQFT PVWVAAVKTY RWQ <del>CIECKC<del>N</del>ICGTSENDDQ</del> MSEPPEGWSN CHLCLDLIKE KASIQYQNQS S*

## SMARCA4

**Variant-** c.1645C>T (p.Arg549Cys)

**PolyPhen 2-** Predicted the variant to be a possibly damaging kind with a score of 1.000.



## SIFT Sorting Intolerant From Tolerant

### Predictions for positions 501 through 600

Threshold for intolerance is 0.05.  
Amino acid color code: nonpolar, uncharged polar, basic, acidic.  
Capital letters indicate amino acids appearing in the alignment, lower case letters result from prediction.  
'Seq Rep' is the fraction of sequences that contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information. Expect poor prediction at this position.

Predicted the variant under not tolerated.

Predict Not Tolerated	Position	Seq	Rep	Predict Tolerated
cwfmidyvhspnltgaer	540	K	0.04	QK
dhgnecswrkypqtafmi	541	L	0.04	VL
hqwpdnecrskgyatfm	542	I	0.04	VLI
ywvtsrpqpnmlkihgfea	543	D	0.04	D
cwfmiyvdhs1tnpager	544	Q	0.04	KQ
wfymhcidlneqvprgs	545	K	0.04	TAK
cwfdfmiyvsgphnalteq	546	K	0.04	RK
miwvfclryrpqathkesg	547	D	0.04	ND
cwdfmfiyvgpshnlaeq	548	K	0.04	TRK
<b>mwcfiyvdstanlepqgk</b>	<b>549</b>	<b>R</b>	<b>0.04</b>	<b>HR</b>
dhgnecswrkypqtafvm	550	L	0.04	IL
whyfrqdenckilpvgs	551	A	0.04	MTA
cwmpdegitnskrval	552	Y	0.04	QFHY
dhgnecswrkypqtafmi	553	L	0.04	VL
dhgnecswkrypqtarfvm	554	L	0.04	IL
cwfdfmiyvphlante	555	Q	0.04	GSKRQ
ywvtsrpqpnmlkihgfedca	556	Q	0.04	Q
whyfrmqedckgnplias	557	T	0.04	VT
miwvfclryrpqathkesg	558	D	0.04	ND
cwfmiy1vrngntpkd	559	E	0.04	SHAQE
kqhnrdrgepctsamviwl	560	Y	0.04	FY

**MUTATION TASTER**-This software predicts the pathogenicity of the variant. It predicted the variant as a deleterious variant.

**Mutation Taster**

## mutation t@sting

Prediction: **Deleterious** [Permalink](#)

**Summary:**

- Amino acid sequence changed
- Protein features (might be) affected
- Model: simple\_aae
- Tree vote: 67|33 (del | benign) [?](#)

Analysed issue	Analysis result
Phys. location	chr15:11106940C>T <a href="#">show variant in all transcripts</a> <a href="#">IGV</a>
Gene symbol	<a href="#">SMARCA4</a>
ExAC LOF metrics	LOF: 1.00, missense: 8.36, synonymous: 0.55
Ensembl transcript ID	<a href="#">ENST00000429416.3</a>
Genbank transcript ID	<a href="#">NM_001128844 (by similarity)</a>
UniProt peptide	<a href="#">P51532</a>
Variant type	Single base exchange
Gene region	CDS
DNA changes	c.1645C>T g.35343C>T

Protein conservation	Species	Match	Gene	AA Alignment
Human	not conserved			549 YRKLIDQKKDKRLAYLLQQTDEY
Platypus	all identical	<a href="#">ENSPTRG00000010488</a>	549	LIDQKKDKCLAYLLQQTDEY
Mmulatta	all identical	<a href="#">ENSMUG00000012042</a>	549	LIDQKKDKCLAYLLQQTDEY
Fcatus	all identical	<a href="#">ENSCFAG00000004894</a>	549	LIDQKKDKCLAYLLQQTDEY
Mmusculus	all identical	<a href="#">ENSMUSG00000032187</a>	549	LIDQKKDKCLAYLLQQTDEY
Ggallus	all identical	<a href="#">ENSGALG00000047145</a>	546	LIDQKKDKCLAYLLQQTDEY
Tribripes	all identical	<a href="#">ENSTRUG00000004885</a>	561	YRKLIDQKKDKCLAYLLQQTDEY
Drerio	all identical	<a href="#">ENSDARG00000104339</a>	505	YRKLIDQKKDKCLAYLLQQT
Dmelanogaster	no homologue			
Celegans	no homologue			
Xtropicalis	all identical	<a href="#">ENSXETG0000009355</a>	534	LIDQKKDKCLAYLLQQTDEY

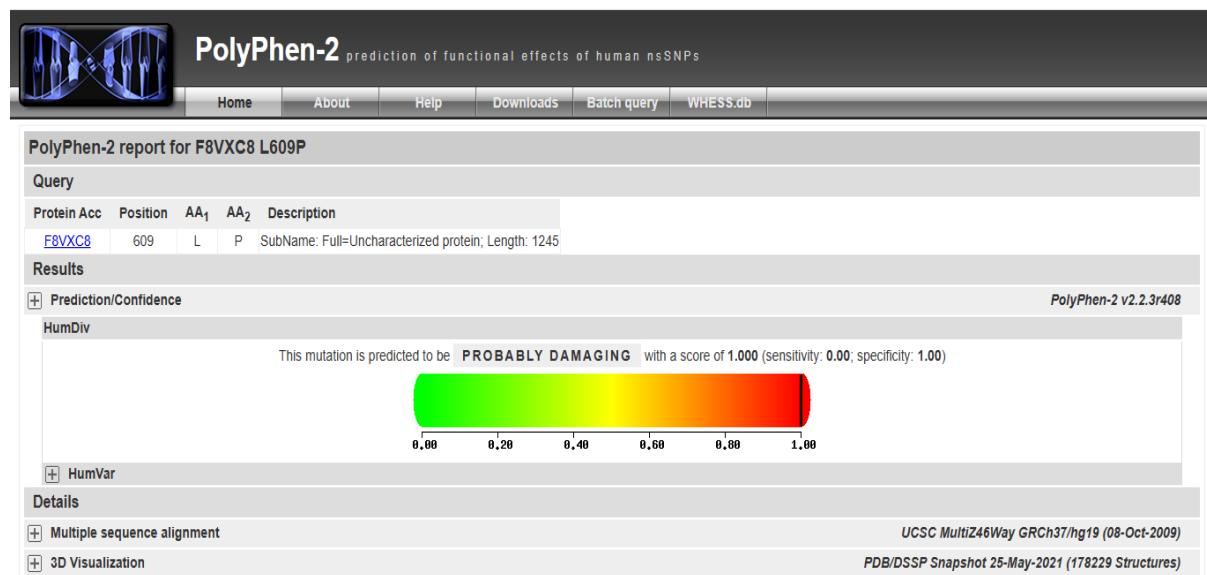
  

Original gDNA sequence snippet	TCGACCAGAAGAAGGACAAGCGCTGGCTACCTCTTGAG
Altered gDNA sequence snippet	TCGACCAGAAGAAGGACAAGTGCCTGGCTACCTCTTGAG
Original cDNA sequence snippet	TCGACCAGAAGAAGGACAAGCGCTGGCTACCTCTTGAG
Altered cDNA sequence snippet	TCGACCAGAAGAAGGACAAGTGCCTGGCTACCTCTTGAG

## SMARCC2

**VARIANT** c.1919T>C (p.Leu640Pro)

**PolyPhen 2-** Predicted the variant to be a possibly damaging kind with a score of 1.000.



## SIFT Sorting Intolerant From Tolerant

### Predictions for positions 601 through 700

Threshold for intolerance is 0.05.  
Amino acid color code: nonpolar, uncharged polar, basic, acidic.  
Capital letters indicate amino acids appearing in the alignment, lower case letters result from prediction.  
'Seq Rep' is the fraction of sequences that contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information. Expect poor prediction at these positions.

Predicted the variant under not tolerated .

Predict Not Tolerated	Position	Seq	Rep	Predict Tolerated
wcfmyihpldn	630R	0.50	tsqVGAEKR	
wcfymihlrgptsVaq	631E	0.56	KNDE	
yvtsrqpnmlkihgfedca	632W	0.67	W	
wfyhmcreliqgkdvvpna	633T	0.67	ST	
cwfmyihvlrgtpsq	634E	0.67	ANKDE	
cwfmyvhlgnpstIadr	635Q	0.67	KEQ	
cwmfiylvhrgtnspakq	636E	0.67	DE	
gdhwnyreckqpsf	637T	0.67	aMvILT	
dhgncweysrpktafmiV	638L	0.67	QL	
cwdhyngseptqarfvmvi	639L	0.67	KL	
ywvtsrqpnmkihgfedca	640L	0.67	L	
hwcdneryqkpgtfamvis	641L	0.67	L	
cwfmiyvlhgtnsrpaqd	642E	0.67	KE	
whyfimqrndelckvtp	643A	0.67	SGA	
hdwngecrspqkytafm	644L	0.67	VIL	
cwyfhigvntpsrakMLd	645E	0.67	QE	
wcpdggnq	646M	0.67	srfvaHyKTIELM	
cwpmdeqkngrtisvalH	647Y	0.67	FY	
wcfmyivhlp	648K	0.67	tqadRSENGK	
cwmfiylvhrtgpsnkq	649D	0.67	AED	
mwifvclyrphatksegQ	650D	0.67	ND	

**MUTATION TASTER**-This software predicts the pathogenicity of the variant. It predicted the variant as a deleterious variant.

**Mutation Taster**

## mutation t@sting

Prediction: **Deleterious** [Permalink](#)

**Summary:**

- Amino acid sequence changed
- Protein features (might be) affected

• Model: simple\_aae  
 • Tree vote: 99|1 (del | benign)

Analysed Issue	Analysis result
Phys. location	chr12:56566219A>G <a href="#">show variant in all transcripts</a> <a href="#">IGV</a>
Gene symbol	<a href="#">SMARCA2</a>
ExAC LOF metrics	LOF: 1.00, missense: 4.26, synonymous: 0.27
Ensembl transcript ID	<a href="#">ENST00000550164.1</a>
Genbank transcript ID	<a href="#">NM_001330288 (exact from MANE)</a>
UniProt peptide	<a href="#">Q8TAQ2</a>
Variant type	Single base exchange
Gene region	CDS
DNA changes	c.1919T>C g.17133T>C

Protein conservation	Species	Match	Gene	AA Alignment
	Human	not conserved		640 TREWTEQETLLLLEALEMYKDDW
	Pitheciamelanoptera	all identical	<a href="#">ENSPTRG00000005081</a>	640 TREWTEQETLLLLEALEMYKDDW
	Macaca mulatta	all identical	<a href="#">ENSMUG00000008273</a>	609 TREWTEQETLLLLEALEMYKDDW
	Felis catus	all identical	<a href="#">ENSCFAG00000014866</a>	640 TREWTEQETLLLLEALEMYKDDW
	Mus musculus	no alignment	<a href="#">ENSMUSG00000025369</a>	n/a
	Gallus gallus	all identical	<a href="#">ENSGALG00000041592</a>	681 TREWTEQETLLLLEALEMYKDDW
	Trichoplax adhaerens	all identical	<a href="#">ENSTRUG00000003962</a>	604 TRWTEQETLLLLEGLEMVKDW
	Danio rerio	all identical	<a href="#">ENSDARG00000077946</a>	606 TREWTDQETLLLLEGLEMVKDW
	Drosophila melanogaster	no homologue		
	Celegans	no homologue		
	Xenopus tropicalis	all identical	<a href="#">ENSXETG00000009486</a>	609 TREWTEQETLLLLEALEMYKDDW

Original gDNA sequence snippet	AGAACAGGAAACCTGCTTCTCCTGGAGGTAAATTGGGCAA
Altered gDNA sequence snippet	AGAACAGGAAACCTGCTTCCCTGGAGGTAAATTGGGCAA
Original cDNA sequence snippet	AGAACAGGAAACCTGCTTCTCCTGGAGGCCTGGAAATGT
Altered cDNA sequence snippet	AGAACAGGAAACCTGCTTCCCTGGAGGCCTGGAAATGT

## BICRA

**Variant -c.192G>C (p.Glu64Asp)**

**PolyPhen-2**- Predicted the variant to be a possibly damaging kind with a score of 1.000.



## SIFT Sorting Intolerant From Tolerant

Predicted the variant under tolerated.

### Predictions for positions 1 through 100

Threshold for intolerance is 0.05.  
Amino acid color code: nonpolar, uncharged polar, basic, acidic.  
Capital letters indicate amino acids appearing in the alignment, lower case letters result from prediction.  
'Seq Rep' is the fraction of sequences that  
contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information. Expect poor prediction at th

Predict Not Tolerated	Position	Seq	Rep	Predict Tolerated
w	54Q	0.75	cmfyhipvgndtsLaerQK	
wcfymhi	55E	0.75	vlpgnrtqadSKE	
whmfyicrlndvke	56A	0.75	tpsgA	
w	57S	0.75	hcmfyrqpdneilkgVtas	
wmfiychlrvq	58G	0.75	pketdnasG	
wmicfvlyprhqatk	59N	0.75	esgDN	
wcm	60H	0.75	ipfvgtlysanderkQH	
ywvtsrqpnmkihgfedca	61L	0.75	L	
wmicfv	62N	0.75	lyprhqtakesdGN	
wcmfy	63P	0.75	ihlvrqntskskadEP	
wcfm	64E	0.75	yihvlprtqnGskadE	
ywvtsrqnmklkihgfedca	65P	0.75	P	
wmcfiylvh	66N	0.75	rkqeagtdSN	
wcmf	67Q	0.75	yihvlprtgsnakDeQ	
w	68P	0.75	mfiyhclqvrtkendsPGa	
w	69A	0.75	hcynndqrpmfekgtsivLA	
ywvtsrqnmklkihgfedca	70P	0.75	P	
wfmhyccilrqvdnpke	71S	0.75	gtAS	
w	72V	0.75	hcmyfqrdnpekgltiasV	
wmcfiylvh	73D	0.75	prqktagenSD	
ywvtsrqpnmkihgfedca	74L	0.75	L	

**MUTATION TASTER**-This software predicts the pathogenicity of the variant. It predicted the variant as a benign variant.



## mutation t@sting

Prediction: **Benign** [Permalink](#)

Summary: • Amino acid sequence changed • Model: simple\_aae  
• Tree vote: 19|81 (del | benign) ?

Analysed issue	Analysis result
Phys. location	chr19:48182619G>C <a href="#">show variant in all transcripts</a> <a href="#">IGV</a>
Gene symbol	BICRA
ExAC LOF metrics	LOF: 1.00, missense: 3.57, synonymous: 2.04
Ensembl transcript ID	<a href="#">ENST00000396720.3</a>
Genbank transcript ID	<a href="#">NM_015711 (by similarity)</a>
UniProt peptide	<a href="#">QSN2M4</a>
Variant type	Single base exchange
Gene region	CDS
DNA changes	c.192G>C g.71167G>C

Protein conservation	Species	Match	Gene	AA Alignment
	Human			64 VQEASGNHLNPEPNQQPAPSVDLD
	mutated	all conserved		64 LNP <sup>T</sup> PNQPAPSVDLD
	Ptroglobutes	all identical	<a href="#">ENSPTRG00000011213</a>	64 LNP <sup>T</sup> PNQPAPSVDLD
	Mmulatta	all identical	<a href="#">ENSMUUG00000019356</a>	64 LNP <sup>T</sup> PNQPAPSVDLD
	Fcatus	all identical	<a href="#">ENSFCAQ00000027050</a>	13 SSGNHLSP <sup>T</sup> TOPAPSVDLD
	Mmusculus	all identical	<a href="#">ENSMUSG00000070808</a>	64 LNP <sup>T</sup> PSQPAPSVDLD
	Ggallus	no homologue		
	Tribolites	not conserved	<a href="#">ENSTRUG00000010033</a>	75 ANEPSGLPRVSVDLD
	Drerio	all identical	<a href="#">ENSDARG00000061159</a>	16 P <sup>T</sup> GLPRVSVDLD
	Dmelanogaster	no homologue		
	Celegans	no homologue		
	Xtropicalis	all identical	<a href="#">ENSXETG00000002787</a>	92 LT <sup>T</sup> SNPPTASVDLD

Original gDNA sequence snippet	Altered gDNA sequence snippet
GGCAACCACCTGAA <del>CC</del> AGA <del>G</del> CCCACCA <del>G</del> CGCCGGCCCCAG	GGCAACCACCTGAA <del>CC</del> AGA <del>G</del> CCCACCA <del>G</del> CGCCGGCCCCAG
Original cDNA sequence snippet	Altered cDNA sequence snippet
GGCAACCACCTGAA <del>CC</del> AGA <del>G</del> CCCACCA <del>G</del> CGCCGGCCCCAG	GGCAACCACCTGAA <del>CC</del> AGA <del>G</del> CCCACCA <del>G</del> CGCCGGCCCCAG

GENE	VARIANTS	PolyPhen2	<u>MutationTaster</u>	SIFT
SOX11	c.178T>C (p.Ser60Pro)	Possibly Damaging – 0.57	Deleterious	Not Tolerated
	c.355C>T (p.Arg119Trp)	Probably Damaging – 1.000	Deleterious	Not Tolerated
	c.170T>C (p.Met57Thr)	Probably Damaging – 0.996	Deleterious	Not Tolerated
	c.250G>C (p.Gly84Arg)	Probably Damaging – 1.000	Deleterious	Not Tolerated
DPF2	c.1066T>G(p.Cys356Gly)	Probably Damaging – 1.000	Deleterious	Not Tolerated
	c.1045G>A (p.Asp349Asn)	Probably Damaging – 0.999	Deleterious	Not Tolerated
	c.868G>A (p.Glu290Lys)	Probably Damaging – 1.000	Deleterious	Not Tolerated
	c.827G>T (p.Cys276Phe)	Probably Damaging – 0.999	Deleterious	Not Tolerated
	c.990C>G (p.Cys330Trp)	Probably Damaging – 1.000	Deleterious	Not Tolerated
SMARCA4	c.1645C>T(p.Arg549Cys)	Probably Damaging – 1.000	Deleterious	Not Tolerated
SMARCC2	c.1919T>C (p.Leu640Pro)	Probably Damaging – 1.000	Deleterious	Not Tolerated
BICRA	c.192G>C (p.Glu64Asp)	Probably Damaging – 1.000	Benign	Tolerated

- ❖ The table presented contains the results obtained from various software used for predicting the pathogenicity of the variants and assessing the potential protein stability changes caused by these variants in the two genes

## DISCUSSION

Coffin-Siris syndrome (CSS) is a rare genetic disorder characterized by developmental delay, intellectual disability, coarse facial features, and hypoplasia of the fifth digit. It is associated with mutations in genes encoding components of the BAF (BRG1/BRM-associated factor) chromatin remodeling complex, including ARID1A, ARID1B, SMARCA4, SMARCC2, and others. This study investigates genetic variants implicated in CSS using in-silico tools to assess their potential pathogenicity. We performed comprehensive bioinformatic analyses to evaluate the structural and functional impact of identified variants across five CSS-related genes: SOX11, DPF2, SMARCA4, SMARCC2, and BICRA.

For SOX11, four variants were identified, all predicted as deleterious and not tolerated by MutationTaster and SIFT. According to PolyPhen-2, one variant was possibly damaging (score 0.597), while the remaining three were predicted to be probably damaging with varying scores. For DPF2, four variants were identified, all predicted to be disease-causing. SMARCA4 and SMARCC2 each had one identified variant, both predicted to be deleterious, not tolerated, and probably damaging. For BICRA, PolyPhen-2 predicted the variant as probably damaging (score 1.000), whereas SIFT predicted it as tolerated, and MutationTaster classified it as benign.

These findings highlight the potential pathogenic impact of identified variants, with most exhibiting deleterious effects across multiple prediction tools. The variation in predicted pathogenicity for the BICRA variant suggests the need for further investigation using functional assays to confirm its biological impact. Our results emphasize the importance of integrating multiple in-silico tools for accurate variant interpretation in CSS. Further functional studies are necessary to validate these findings and elucidate the molecular mechanisms contributing to CSS pathology.

## **CONCLUSION**

In the present study, the in-silico analysis identified several potentially pathogenic variants in CSS-associated genes, including SOX11, DPF2, SMARCA4, SMARCC2, and BICRA. While most variants were consistently predicted as deleterious across multiple tools, the BICRA variant showed conflicting results, highlighting the need for further investigation. These findings provide valuable insights into the genetic landscape of Coffin-Siris syndrome and emphasize the importance of combining bioinformatic analysis with experimental validation to enhance diagnostic accuracy and improve our understanding of CSS pathogenesis.

## REFERENCES

*Genetics / History, Biology, Timeline, & Facts / Britannica.* 19 Mar. 2025, <https://www.britannica.com/science/genetics>.

*What Is a Gene?: MedlinePlus Genetics.*

<https://medlineplus.gov/genetics/understanding/basics/gene/>. Accessed 25 Mar. 2025.

Alliance, Genetic, and The New York-Mid-Atlantic Consortium for Genetic and Newborn Screening Services. “GENETICS 101.” *Understanding Genetics: A New York, Mid-Atlantic Guide for Patients and Health Professionals*, Genetic Alliance, 2009. [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov), <https://www.ncbi.nlm.nih.gov/books/NBK115568/>.

“What Is a Rare Disease?” *Rare Disease Day 2026*,

<https://www.rarediseaseday.org/what-is-a-rare-disease/>. Accessed 25 Mar. 2025.

<https://academic.oup.com/gsajournals/pages/genetic-models-of-rare-diseases>

*Bioinformatics / Genomics, Proteomics & Data Analysis / Britannica.* 9 Feb. 2025, <https://www.britannica.com/science/bioinformatics>.

<https://academic.oup.com/nar/article/51/D1/D523/6835362>

[https://www.ensembl.org/info/website/tutorials/module1\\_aug2011\\_ensembl\\_intro.pdf](https://www.ensembl.org/info/website/tutorials/module1_aug2011_ensembl_intro.pdf)

<https://academic.oup.com/nar/article/49/D1/D884/5952199>

EMBL-EBI. *What Is AlphaFold? / AlphaFold.*

<https://www.ebi.ac.uk/training/online/courses/alphafold/an-introductory-guide-to-its-strengths-and-limitations/what-is-alphafold/>. Accessed 25 Mar. 2025.

*PolyPhen-2: Prediction of Functional Effects of Human nsSNPs.*

<http://genetics.bwh.harvard.edu/pph2/>. Accessed 25 Mar. 2025.

Adzhubei, Ivan, et al. “Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2.” *Current Protocols in Human Genetics / Editorial Board, Jonathan L. Haines ... [et Al.]*, vol. 0 7, Jan. 2013, p. Unit7.20. *PubMed Central*, <https://doi.org/10.1002/0471142905.hg0720s76>.

<https://academic.oup.com/nar/article/40/W1/W452/1751364>

SIFT - Predict Effects of Nonsynonymous / Missense Variants. <https://sift.bii.a-star.edu.sg/>. Accessed 25 Mar. 2025.

Steinhaus, Robin, et al. "MutationTaster2021." *Nucleic Acids Research*, vol. 49, no. W1, July 2021, pp. W446–51. *PubMed*, <https://doi.org/10.1093/nar/gkab266>.

*National Center for Biotechnology Information - an Overview / ScienceDirect Topics*. <https://www.sciencedirect.com/topics/agricultural-and-biological-sciences/national-center-for-biotechnology-information>. Accessed 25 Mar. 2025.

<https://academic.oup.com/nar/article/44/D1/D7/2503096>

*Coffin-Siris Syndrome: MedlinePlus Genetics*.

<https://medlineplus.gov/genetics/condition/coffin-siris-syndrome/>. Accessed 25 Mar. 2025.

[https://medlineplus.gov/download/genetics/condition/coffin-siris-syndrome.pdf?utm\\_source=chatgpt.com](https://medlineplus.gov/download/genetics/condition/coffin-siris-syndrome.pdf?utm_source=chatgpt.com)

Committee on Diagnostic Error in Health Care, et al. *Improving Diagnosis in Health Care*. Edited by Erin P. Balogh et al., National Academies Press, 2015.

*DOI.org (Crossref)*, <https://doi.org/10.17226/21794>.

[https://medlineplus.gov/download/genetics/condition/coffin-siris-syndrome.pdf?utm\\_source=chatgpt.com](https://medlineplus.gov/download/genetics/condition/coffin-siris-syndrome.pdf?utm_source=chatgpt.com)

*Coffin-Siris Syndrome: Symptoms and Outlook*. 18 Jan. 2023,

<https://www.medicalnewstoday.com/articles/coffin-siris-syndrome>.

*Coffin-Siris Syndrome - Symptoms, Causes, Treatment / NORD*.

<https://rarediseases.org/rare-diseases/coffin-siris-syndrome/>. Accessed 26 Mar. 2025

*ARID1A AT-Rich Interaction Domain 1A [Homo Sapiens (Human)] - Gene - NCBI*. <https://www.ncbi.nlm.nih.gov/gene/8289>. Accessed 26 Mar. 2025.

*ARID2 AT-Rich Interaction Domain 2 [Homo Sapiens (Human)] - Gene - NCBI*. <https://www.ncbi.nlm.nih.gov/gene/196528>. Accessed 26 Mar. 2025.

*SOX11 Gene: MedlinePlus Genetics*.

<https://medlineplus.gov/genetics/gene/sox11/>. Accessed 26 Mar. 2025.

Barish, Scott, et al. "BICRA, a SWI/SNF Complex Member, Is Associated with BAF-Disorder Related Phenotypes in Humans and Model Organisms." *The American Journal of Human Genetics*, vol. 107, no. 6, Dec. 2020, pp. 1096–112. *ScienceDirect*, <https://doi.org/10.1016/j.ajhg.2020.11.003>.