In silico analysis of coding SNPs and 3'-UTR associated miRNAs in DCAF17 gene that may affect the regulation and pathogenesis of Woodhouse-Sakati Syndrome

Abdelrahman H. Abdelmoneim1*, Asia M. Elrashied2, Alaa I. Mohammed3, Sara A. Mirghani 1, Rania E. Osman1,4. Esraa O. Gadim1, Mohamed A. Hassan1.

1 Department of Biotechnology, Africa City of Technology, Sudan.
2 Faculty of Botany, University of Khartoum, Sudan.
3 Department of Hematology & Immunohematology, Faculty of Medical Laboratory Sciences, University of Khartoum, Sudan.
4 Alfarabi Medical Laboratory, Sudan.

* Corresponding author: Email: (abduhamza009@gmail.com).

Abstract:

Background: Woodhouse-Sakati Syndrome refers to a group of inherited disorders characterized by alopecia, hypogonadism, diabetes mellitus, hypothyroidism and progressive extrapyramidal signs. The aim of this study is to identify the pathogenic SNPs in the DCAF17 gene with their related microRNAs and their effect on the structure and function of the protein.

Material and Methods: We used different bioinformatics tools to predict the effect of each SNP on the structure and function of the protein. After that we defined the miRNAs founded in the 3'-UTR region on the DCAF17 gene and studied the annotations relative to it.

Results: Ten deleterious SNPs out of 339 were found to have a damaging effect on the protein structure and function of the protein. With one significant microRNA in the 3'-UTR region.

Conclusion: This was the first in silico analysis of DCAF17 gene, in which 10 novel mutations were found using different bioinformatics tools that could be used as a diagnostic markers for Woodhouse-Sakati syndrome, with one relevant microRNA that can regulate the function of the protein.

Keywords: Woodhouse-Sakati Syndrome, DCAF17 gene, SNPs, 3’-UTR, in silico analysis, diagnostic markers.
1. Introduction:

Woodhouse-Sakati Syndrome (WSS, MIM: 241080) is a rare autosomal-recessive multi systemic disorder which is characterized by a combination of hypogonadism, alopecia, Diabetes Mellitus (DM), mental retardation and extrapyramidal signs. (1-10)

It was originally described in a number of consanguineous Saudi families in the Middle East, but has recently been reported in other ethnicities as well. (11, 12) Since its original description in 1983, approximately 50 cases have been reported until now. (13) Exogenous hormone therapy promoting secondary sex characteristic development and insulin are among the recommended treatments for the disease. (14, 15)

The availability of vast amounts of sequence data, coupled to advances in computational biology in the recent years provides an ideal framework for in silico gene expression analysis. Single Nucleotide Polymorphisms (SNPs) make up about 90% of DNA sequence variations, making it the most common type of genetic variation. The SNPs that are most likely to have a direct impact on the protein product of a gene are those that change the amino acid sequence and variants in gene regulatory regions, which control protein expression levels. (16-18)

The disease was first described in 2008 by al Alazami et al. (4, 11) Mutations in the DCAF17 gene are the cause of Woodhouse-Sakati syndrome. (3, 19, 20) It is located on chromosome 2q31.1. (4, 21) It encodes a nucleolar protein with poorly understood function, adding to that the pathogenic mechanism underlying WSS is also unknown. (13)

The aim of this study was to identify the pathogenic SNPs in DCAF17 gene located in the coding region and analyzing the microRNA(miRNA) in the 3'-untranslated regions (3'-UTR) using in silico prediction software’s that determine the structure, function and regulation of their respective proteins. This is the first in silico analysis of it is kind in DCAF17 gene to address this matter.
2. Material and Methods:

2.1. Retrieving nsSNPs:
SNPs associated with DCAF17 gene were obtained from the Single Nucleotide Polymorphism database (dbSNP) in the National Center for Biotechnology Information (NCBI). (http://www.ncbi.nlm.nih.gov/snp/).
The sequence and natural variants of DCAF17 protein were obtained from the UniProt database as it considered as the most reliable and unambiguous database for protein sequences. (22) (https://www.uniprot.org/).
A total number of 339 nonsynonymous Single Nucleotide Polymorphisms (ns SNPs) were found from NCBI database, all of them were subjected to in silico analysis using ten different algorithms and softwares; SIFT, PROVEAN, PolyPhen-2, SNPs&GO, PhD-SNP, PMUT, I-mutant, Project Hope, GeneMANIA and Chimera.

2.2. Identifying the most damaging nsSNPs and disease related mutations:
Five different bioinformatics tools were utilized to predict functional effects of nsSNPs obtained from dbSNP database on the protein. These algorithmic include: SIFT, PROVEAN, PolyPhen-2, SNAP2 and PhD-SNP.
The SNPs predicted deleterious by at least four softwares were considered high risk nsSNPs and investigated further.

2.2.1. SIFT Server:
Phenotypic effects of amino acid substitution on protein function were predicted by using Sorting Intolerant From Tolerant (SIFT) server, which is a powerful tool used to fulfill this purpose. A list of nsSNPs (rsIDs) from NCBI's dbSNP database was submitted as a query (original) sequence to SIFT to predict tolerated and deleterious substitutions for every position of a protein sequence. The server divides the results into “Deleterious” and “Tolerated”, nsSNPs with SIFT score \( \leq 0.05 \) were classified as deleterious and are further analyzed to identify the damaging ones, and those \( > 0.05 \) were classified as tolerated and are not further analyzed. (Available at: http://sift.bii.a-star.edu.sg/). (23-26)

2.2.2. Provean Server:
(Protein Variation Effect Analyzer) is the second software tool used. It also predicts the effect of an amino acid substitution on the biological function of a protein. It predicts the damaging effects of any type of protein sequence variations to not only single amino acid substitutions but also in-frame insertions, deletions, and multiple amino acid substitutions. The results are obtained as either “Deleterious” if the prediction score was \( \leq -2.5 \), while score \( > -2.5 \) indicates that the variant is predicted to have a “Neutral” effect. (Available at: http://provean.jcvi.org/index.php). (27, 28)
2.2.3. Polyphen-2 Server:
Polymorphism Phenotyping v2.0 (PolyPhen-2) is another online tool that predicts the possible effects of an amino acid substitution on the structure and function of the protein. The results are classified into “PROBABLY DAMAGING” that is the most disease causing with a score near to 1 (0.7-1), “POSSIBLY DAMAGING” with a less disease causing ability with a score of 0.5-0.8 and “BENIGN” which does not alter protein functions with a score closer to zero; (0-0.4). (Available at: http://genetics.bwh.harvard.edu/pph2/). (22, 29, 30)

2.2.4. SNPs & Go server:
An online web server that used to ensure the disease relationship with the studied single nucleotide polymorphisms SNPs. It gives three different results based on three different analytical algorithms; Panther result, PHD-SNP result, SNPs&GO result. Each one of these results is composed of three parts, the prediction which decides whether the mutation is neutral or disease related, reliability index (RI), and disease probability (if >0.5 mutation is considered as disease causing nsSNP). (Available at: http://snps-and-go.biocomp.unibo.it/snps-and-go/). (31)

2.2.5. PMUT Server:
PMUT is a powerful web-based tool used for the prediction of pathological variants on proteins. The prediction results are classified as “Neutral” or “Disease”. It is available at (http://mmb.irbbarcelona.org/PMut). (32)

2.3. Protein Structural analysis:
I-Mutant:
A online web-tool that is used for the prediction of protein stability changes upon single point mutations, determining whether the mutation increases or decreases the protein stability. (Available at: http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant2.0/I-Mutant2.0.cgi). (33)

2.4. Project HOPE:
It is an online web-server used to analyze the structural and functional variations on protein sequence that have been resulted from single amino acid substitution. It searches protein 3D structures by collecting structural information from a series of sources, including calculations on the 3D coordinates of the protein and sequence annotations from the UniProt database. Protein sequences are submitted to project HOPE server then HOPE builds a complete report with text, figures, and animations. It is available at (http://www.cmbi.ru.nl/hope). (34)

2.5. GeneMANIA:
A user-friendly web interface tool approaches to know protein function, analyzing submitted gene lists and prioritizing genes for functional assays. It provides lists with functionally similar
genes that it identifies and recognizes using available genomics and proteomics data. (Available at: [http://www.genemania.org/].) (35)

2.6. Modeling nsSNP locations on protein structure:

Chimera:

Chimera 1.8 software is used for visualization and editing of the three dimensional structure of protein. It visualizes the original amino acid with the mutated one to see the impact that can be produced. It is a highly extensible program for interactive visualization and analysis of molecular structures and related data, including density maps, supramolecular assemblies and sequence alignments. High-quality images and animations can be generated. (Available at: [http://www.cgl.ucsf.edu/chimera].) (36)

2.7. 3’-UTR analysis:

2.7.1. Ensemble 95:

RNA sequence were retrieved from ensemble, which is website that makes key genomic data sets available to the entire scientific community without restrictions. (37, 38) (Available at: [https://www.ensembl.org/index.html]).

2.7.2. RegRNA2.0:

Then the RNA sequences were inserted to RegRNA 2.0 (which is an integrated web server for identifying functional RNA motifs and sites. (39, 40) (Available at: [http://regrna2.mbc.nctu.edu.tw/] to find the related microRNA (miRNA).

2.7.3. miRmap22.1:

Then the miRNA sequence were further inserted in to miRmap website to predict the miRNA targets and study the repression strength using thermodynamic, probabilistic, evolutionary and sequence-based approaches. (41) (Available at: [https://mirmap.ezlab.org/]).

2.7.4. miRBase:

After that we use mirbase website to study its annotations (miRBase website provides a wide-range of information on published microRNAs, including their sequences, their biogenesis precursors, genome coordinates, and community driven annotation). (42-44) (Available at: [http://www.mirbase.org/]).
Result:

339 missense SNPs were retrieved from National Center for Biotechnology Information (NCBI) website and submitted to SIFT, Proven, Polyphen-2, SNAP2 respectively. After analysis, 60 mutations out of 339 missense mutations were found to be deleterious in all four servers as found in table (1).

Furthermore, the above result were submitted to SNP&GO, PHD, and Pmut to study the relation between the SNPs and the disease. Out of 60 SNPs 10 were found to be related to disease. As indicated in table (2). Then I mutant was used to detect the stability of the protein and the result indicated that all the SNPs except one (rs1279189246, H478L) lead to decrease in the stability which lead to unstable amino acid interaction. As found in table (3).

One significant microRNA was found and investigated as shown in table (4) and (5).
Figure. (1): Diagrammatic representation of in silico work flow for DCAF17 gene in coding and 3'-UTR regions.

Table. (1): Shows the most deleterious SNPs by four softwares.
| dbSNP rs#       | Substitution | sift   | Predicted | sift Score |  | Provean Prediction | Provean Score | POLYPHEN2 Prediction | POLYPHEN2 Score | SNAP2 Predicted Effect | SNA Score |
|----------------|--------------|--------|-----------|------------|-------------------------------|----------------|-----------------------|-------------------|-------------------------|------------|
| rs1229235381   | R5W          | AFFECT | Deleterious | -2.538     | possibly damaging              | 0.95           | Effect 50              |                   |                         |            |
| rs1330567400   | C10R         | AFFECT | Deleterious | -4.238     | probably damaging              | 0.997          | Effect 84              |                   |                         |            |
| rs1420986481   | L13P         | AFFECT | Deleterious | -3.429     | probably damaging              | 0.99           | Effect 63              |                   |                         |            |
| rs867160323    | G19C         | AFFECT | Deleterious | -4.108     | probably damaging              | 1              | Effect 85              |                   |                         |            |
| rs748586810    | N32T         | AFFECT | Deleterious | -2.666     | possibly damaging              | 0.996          | Effect 20              |                   |                         |            |
|                | N32I         | AFFECT | Deleterious | -4.378     | probably damaging              | 0.999          | Effect 57              |                   |                         |            |
| rs1379715347   | V40E         | AFFECT | Deleterious | -2.98      | possibly damaging              | 0.828          | Effect 35              |                   |                         |            |
| rs576105333    | Y61C         | AFFECT | Deleterious | -2.61      | probably damaging              | 1              | Effect 71              |                   |                         |            |
| rs1224937342   | F68C         | AFFECT | Deleterious | -2.561     | probably damaging              | 0.992          | Effect 31              |                   |                         |            |
| rs1356908272   | D69V         | AFFECT | Deleterious | -3.745     | probably damaging              | 0.514          | Effect 76              |                   |                         |            |
| rs886055106    | R72W         | AFFECT | Deleterious | -3.542     | probably damaging              | 0.999          | effect 85              |                   |                         |            |
| rs764744666    | Y86H         | AFFECT | Deleterious | -2.834     | probably damaging              | 0.999          | effect 68              |                   |                         |            |
| rs889604735    | P89R         | AFFECT | Deleterious | -3.725     | probably damaging              | 1              | effect 19              |                   |                         |            |
| rs1220794681   | E98G         | AFFECT | Deleterious | -3.449     | probably damaging              | 0.999          | effect 50              |                   |                         |            |
| rs754648178    | W129R        | AFFECT | Deleterious | -5.379     | probably damaging              | 0.999          | effect 69              |                   |                         |            |
| rs1302837695   | K167T        | AFFECT | Deleterious | -3.663     | probably damaging              | 0.994          | effect 67              |                   |                         |            |
| rs377364475    | R178W        | AFFECT | Deleterious | -3.983     | probably damaging              | 0.95           | effect 87              |                   |                         |            |
| rs1157407921   | G181C        | AFFECT | Deleterious | -2.715     | probably damaging              | 1              | effect 57              |                   |                         |            |
| rs750319630    | V186D        | AFFECT | Deleterious | -3.175     | probably damaging              | 0.969          | effect 56              |                   |                         |            |
| rs       | Amino Acid | AFFECT | Score 1 | Description    | Score 2 | Description    | Score 3 | Description    | Effect |
|----------|------------|--------|---------|---------------|---------|---------------|---------|---------------|--------|
| 118193053 | A191E      | Deleterious | -2.86   | possibly damaging | 0.948   | effect 56     |
| 778114154 | P197L      | Deleterious | -7.417  | probably damaging | 1       | effect 21     |
| 778975572 | L204P      | Deleterious | -4.144  | possibly damaging | 0.772   | effect 78     |
| 769602895 | G222R      | Deleterious | -4.367  | probably damaging | 0.979   | effect 84     |
| 1422292058 | L224P     | Deleterious | -4.894  | probably damaging | 0.999   | effect 77     |
| 1474908539 | S230P     | Deleterious | -2.73   | possibly damaging | 0.718   | effect 45     |
| 748887234 | R234I      | Deleterious | -4.45   | probably damaging | 0.997   | effect 68     |
| 369174398 | Y236C      | Deleterious | -5.723  | probably damaging | 0.927   | effect 19     |
| 1476963715 | G253E     | Deleterious | -3.683  | probably damaging | 1       | effect 76     |
| 1015369709 | C256R     | Deleterious | -3.356  | possibly damaging | 0.899   | effect 39     |
|         | C256G      | Deleterious | -3.232  | probably damaging | 0.718   | effect 37     |
| 778451530 | P269L      | Deleterious | -5.08   | probably damaging | 1       | effect 68     |
| 1439042141 | G271V     | Deleterious | -4.416  | probably damaging | 1       | effect 60     |
| 757839826 | P282S      | Deleterious | -5.005  | probably damaging | 1       | effect 50     |
| 1292011900 | P283L     | Deleterious | -4.623  | probably damaging | 0.959   | effect 38     |
| 746899310 | L285R      | Deleterious | -4.499  | probably damaging | 1       | effect 42     |
| 1463889939 | W302G     | Deleterious | -10.032 | probably damaging | 1       | effect 87     |
| 1412078281 | H318R     | Deleterious | -3.987  | probably damaging | 0.98    | effect 30     |
| 1319790255 | D324G     | Deleterious | -4.179  | probably damaging | 1       | effect 3      |
| 784888864 | W344R      | Deleterious | -3.279  | probably damaging | 0.999   | effect 26     |
| 1439891137 | W344C     | Deleterious | -2.985  | probably damaging | 1       | effect 39     |
| 752287637 | D350H      | Deleterious | -5.89   | probably damaging | 1       | effect 83     |
| 764182538 | H357Y      | Deleterious | -2.596  | probably damaging | 0.998   | effect 25     |
| 890180094 | H357P      | Deleterious | -6.354  | probably damaging | 1       | effect 90     |
| 764124723 | T402I      | Deleterious | -4.114  | probably damaging | 1       | effect 49     |
| SNP             | Amino Acid | AFFECT | Mutation Type | Predicted Effect | p.Value | Effect Size |
|-----------------|------------|--------|---------------|-----------------|---------|-------------|
| rs751732245     | R406W      | 0      | Deleterious   | -6.971          | 1       | 75          |
| rs1211965571    | D416H      | 0      | Deleterious   | -2.771          | 1       | 25          |
| rs1288356140    | D416G      | 0      | Deleterious   | -3.396          | 1       | 12          |
|                 | D416V      | 0      | Deleterious   | -4.712          | 1       | 8           |
| rs139061039     | D418G      | 0      | Deleterious   | -5.105          | 1       | 32          |
| rs759392662     | Y429C      | 0      | Deleterious   | -7.379          | 1       | 40          |
| rs1227872471    | D431H      | 0      | Deleterious   | -4.105          | 1       | 54          |
| rs1256723119    | L433S      | 0      | Deleterious   | -4.322          | 1       | 42          |
| rs766858148     | G448R      | 0      | Deleterious   | -3.596          | 1       | 48          |
| rs1409820369    | D473H      | 0      | Deleterious   | -4.896          | 1       | 52          |
| rs1279189246    | H478R      | 0      | Deleterious   | -3.431          | 1       | 68          |
|                 | H478L      | 0      | Deleterious   | -5.421          | 1       | 62          |
| rs1333368709    | D483G      | 0      | Deleterious   | -5.332          | 0.995   | 62          |
| rs753532618     | R484G      | 0      | Deleterious   | -3.653          | 0.996   | 76          |
| rs1387898294    | D485E      | 0      | Deleterious   | -2.773          | 0.971   | 3           |
| rs749237410     | F498V      | 0      | Deleterious   | -4.135          | 1       | 55          |
### Table (2): Disease effect nsSNPs associated variations predicted by SNPs&GO and PHD-SNP softwares.

| dbSNP rs#       | SUB   | SNP & GO | RI* Probability | PHD   | RI* Probability | PMUT Score | Prediction |
|-----------------|-------|----------|-----------------|-------|-----------------|------------|------------|
| rs1356908272    | D69V  | Disease  | 3 0.648         | Disease 8 0.893 | 0.68 (85%) | Disease   |
| rs886055106     | R72W  | Disease  | 3 0.625         | Disease 7 0.868 | 0.53 (80%) | Disease   |
| rs750319630     | V186D | Disease  | 0 0.509         | Disease 7 0.832 | 0.64 (84%) | Disease   |
| rs778114154     | P197L | Disease  | 3 0.649         | Disease 8 0.921 | 0.84 (91%) | Disease   |
| rs778975572     | L204P | Disease  | 1 0.533         | Disease 7 0.825 | 0.68 (85%) | Disease   |
| rs1422292058    | L224P | Disease  | 1 0.535         | Disease 7 0.836 | 0.86 (91%) | Disease   |
| rs1256723119    | L433S | Disease  | 1 0.539         | Disease 7 0.828 | 0.86 (91%) | Disease   |
| rs1279189246    | H478R | Disease  | 1 0.564         | Disease 8 0.913 | 0.70 (86%) | Disease   |
| rs1333368709    | D483G | Disease  | 0 0.518         | Disease 8 0.885 | 0.86 (91%) | Disease   |
| rs749237410     | F498V | Disease  | 2 0.62          | Disease 8 0.886 | 0.69 (86%) | Disease   |

* RI: Reliability Index
Table. (3): Stability analysis predicted by I-Mutant version 3.0.

| dbSNP rs#  | Mutation | i-mutant prediction | RI | Score |
|------------|----------|---------------------|----|-------|
| rs1356908272 | D69V     | Decrease            | 5  | -0.6  |
| rs886055106    | R72W     | Decrease            | 5  | -0.86 |
| rs750319630    | V186D    | Decrease            | 7  | -1.22 |
| rs778114154    | P197L    | Decrease            | 8  | -0.84 |
| rs778975572    | L204P    | Decrease            | 7  | -1.72 |
| rs1422292058   | L224P    | Decrease            | 4  | -1.5  |
| rs1256723119   | L433S    | Decrease            | 9  | -2.29 |
| rs1279189246   | H478L    | Increase            | 5  | 0.4   |
| rs1333368709   | D483G    | Decrease            | 5  | -1.02 |
| rs749237410    | F498V    | Decrease            | 9  | -1.76 |

*RI: Reliability Index
### Table. (4): miRNAs target identified in 3'-UTR region of DCAF17 gene by miRmap website.

| miRNA         | Gene  | Position | No. of Nucleotides | Target Sequence | ΔG open | Probability exact | Conservation PhyloP | miRmap score |
|---------------|-------|----------|--------------------|-----------------|---------|------------------|---------------------|--------------|
| hsa-miR-4436b-3p | DCAF17 | 210
   ~2124       | 22     | acataaatgtcttcctgcccta | 12.32       | 0.02  | 0.71             | -0.55               |              |

* ΔG open: mRNA opening free energy – Accessibility

* PhyloP: evolutionary conservation predictor.

### Table. (5): Alignment of Query to mature miRNAs.

| Accession ID | Query start | Query end | Subject start | Subject end | Strand | Score | Value |
|--------------|-------------|-----------|---------------|-------------|--------|-------|-------|
| MIMAT0019941 | 10          | 21        | 2             | 13          | -      | 60    | 12    |
| MIMAT0025585 | 2           | 22        | 2             | 22          | -      | 60    | 12    |
| MIMAT0027651 | 10          | 21        | 3             | 14          | +      | 60    | 12    |
| MIMAT0017064 | 4           | 22        | 3             | 21          | -      | 59    | 15    |
| MIMAT0037309 | 4           | 22        | 4             | 22          | -      | 59    | 15    |
| MIMAT0015037 | 9           | 21        | 10            | 22          | -      | 56    | 27    |
| MIMAT0019772 | 11          | 21        | 6             | 16          | +      | 55    | 33    |
| MIMAT0027850 | 6           | 19        | 5             | 18          | -      | 52    | 58    |
| MIMAT0027973 | 9           | 22        | 4             | 17          | +      | 52    | 58    |
| MIMAT0028056 | 7           | 20        | 2             | 15          | -      | 52    | 58    |
| MIMAT0028057 | 7           | 20        | 6             | 19          | +      | 52    | 58    |
| MIMAT0019361 | 10          | 21        | 6             | 17          | -      | 51    | 70    |
| MIMAT0019940 | 10          | 21        | 8             | 19          | +      | 51    | 70    |
| MIMAT0027746 | 10          | 21        | 4             | 15          | -      | 51    | 70    |
| MIMAT0003339 | 3           | 12        | 7             | 16          | -      | 50    | 85    |
| MIMAT0004869 | 3           | 12        | 7             | 16          | -      | 50    | 85    |
Table. (6): The gene co-expressed, share domain and interaction with *DCAF17* gene network.

| Gene 1       | Gene 2       | Weight   | Network group   |
|--------------|--------------|----------|-----------------|
| GEMIN2       | C18orf25     | 0.020875 | Co-expression   |
| C18orf25     | DCAF17       | 0.01954  | Co-expression   |
| POT1         | CUL4B        | 0.015105 | Co-expression   |
| POT1         | TMEM267      | 0.010743 | Co-expression   |
| SKIV2L2      | DCAF17       | 0.004223 | Co-expression   |
| VIPAS39      | BBS7         | 0.006134 | Co-expression   |
| POLI         | DCAF17       | 0.01365  | Co-expression   |
| GEMIN2       | SKIV2L2      | 0.001957 | Co-expression   |
| RAD17        | CUL4B        | 0.009911 | Co-expression   |
| RAD17        | POT1         | 0.005801 | Co-expression   |
| RAD17        | VIPAS39      | 0.002685 | Co-expression   |
| EYA3         | POT1         | 0.02177  | Co-expression   |
| RAD17        | SKIV2L2      | 0.012174 | Co-expression   |
| TMEM267      | DCAF17       | 0.008714 | Co-expression   |
| POT1         | TMEM267      | 0.010902 | Co-expression   |
| SKIV2L2      | DCAF17       | 0.010967 | Co-expression   |
| SKIV2L2      | CUL4B        | 0.014126 | Co-expression   |
| SKIV2L2      | POT1         | 0.015277 | Co-expression   |
| BBS7         | DCAF17       | 0.016686 | Co-expression   |
| EXOC2        | BLMH         | 0.015184 | Co-expression   |
| GEMIN2       | CUL4B        | 0.011145 | Co-expression   |
| GEMIN2       | INTS7        | 0.011782 | Co-expression   |
| RAD17        | SKIV2L2      | 0.013417 | Co-expression   |
| CUL4B        | DCAF17       | 0.009401 | Co-expression   |
| Protein 1  | Protein 2   | P-value  | Co-expression |
|-----------|-------------|----------|---------------|
| TMEM267   | DCAF17      | 0.002878 |               |
| TMEM267   | CUL4B       | 0.005961 |               |
| TMEM267   | C18orf25    | 0.003443 |               |
| SKIV2L2   | DCAF17      | 0.010745 |               |
| POLI      | TMEM267     | 0.004744 |               |
| GEMIN2    | POT1        | 0.011051 |               |
| RAD17     | DCAF17      | 0.010133 |               |
| RAD17     | TMEM267     | 0.007081 |               |
| BLMH      | SKIV2L2     | 0.011372 |               |
| C18orf25  | DCAF17      | 0.026408 |               |
| TMEM267   | DCAF17      | 0.016632 |               |
| SKIV2L2   | DCAF17      | 0.01629  |               |
| NAA15     | DCAF17      | 0.022654 |               |
| GEMIN2    | DCAF17      | 0.019081 |               |
| GEMIN2    | TMEM267     | 0.015272 |               |
| SKIV2L2   | POT1        | 0.003939 |               |
| BBS7      | DCAF17      | 0.021653 |               |
| GEMIN2    | TMEM267     | 0.0039  |               |
| GEMIN2    | POT1        | 0.006586 |               |
| TMEM267   | DCAF17      | 0.017157 |               |
| POT1      | DCAF17      | 0.018457 |               |
| POT1      | CUL4B       | 0.016976 |               |
| INTS7     | PTTG1       | 0.009717 |               |
| POLI      | DCAF17      | 0.024595 |               |
| POLI      | POT1        | 0.015922 |               |
| GEMIN2    | POT1        | 0.014995 |               |
| Protein 1 | Protein 2 | Score       | Interaction   |
|----------|----------|-------------|---------------|
| POT1     | DCAF17   | 0.019193    | Co-expression |
| NAA15    | SKIV2L2  | 0.005278    | Co-expression |
| INTS7    | SKIV2L2  | 0.0093      | Co-expression |
| INTS7    | NAA15    | 0.011806    | Co-expression |
| GEMIN2   | CUL4B    | 0.014478    | Co-expression |
| GEMIN2   | POT1     | 0.017858    | Co-expression |
| RAD17    | CUL4B    | 0.018899    | Co-expression |
| RAD17    | GEMIN2   | 0.015564    | Co-expression |
| AGO1     | POT1     | 0.012863    | Co-expression |
| VIPAS39  | AGO1     | 0.006072    | Co-expression |
| EXOC2    | INTS7    | 0.01908     | Co-expression |
| RAD17    | POT1     | 0.027534    | Co-expression |
| RAD17    | VIPAS39  | 0.012628    | Co-expression |
| INTS7    | AGO1     | 0.019896    | Co-expression |
| GEMIN2   | PTTG1    | 0.002871    | Co-expression |
| RAD17    | CUL4B    | 0.013971    | Co-expression |
| RAD17    | GEMIN2   | 0.012185    | Co-expression |
| INTS7    | SKIV2L2  | 0.006518    | Co-expression |
| BLMH     | EYA3     | 0.014397    | Co-localization |
| POT1     | DCAF17   | 0.009539    | Co-localization |
| EYA3     | DCAF17   | 0.02755     | Co-localization |
| EYA3     | POT1     | 0.008808    | Co-localization |
| AGO1     | DCAF17   | 0.026643    | Co-localization |
| AGO1     | POT1     | 0.008327    | Co-localization |
| ARL15    | DCAF17   | 0.025506    | Co-localization |
| ARL15    | POT1     | 0.006414    | Co-localization |
| Protein 1 | Protein 2   | Score   | Interaction Type |
|-----------|------------|---------|-----------------|
| NAA15     | DCAF17     | 0.011778| Co-localization |
| NAA15     | POT1       | 0.003766| Co-localization |
| NAA15     | ARL15      | 0.008719| Co-localization |
| INTS7     | DCAF17     | 0.023841| Co-localization |
| INTS7     | POT1       | 0.008081| Co-localization |
| INTS7     | NAA15      | 0.010854| Co-localization |
| BLMH      | DCAF17     | 0.025363| Co-localization |
| BLMH      | POT1       | 0.009025| Co-localization |
| BLMH      | EYA3       | 0.02518 | Co-localization |
| BLMH      | NAA15      | 0.011865| Co-localization |
| VIPAS39   | DCAF17     | 0.023599| Co-localization |
| VIPAS39   | POT1       | 0.00637 | Co-localization |
| VIPAS39   | ARL15      | 0.025043| Co-localization |
| VIPAS39   | NAA15      | 0.008351| Co-localization |
| EXOC2     | DCAF17     | 0.024342| Co-localization |
| EXOC2     | POT1       | 0.007643| Co-localization |
| EXOC2     | ARL15      | 0.024647| Co-localization |
| EXOC2     | VIPAS39    | 0.028187| Co-localization |
| GEMIN2    | DCAF17     | 0.012483| Co-localization |
| GEMIN2    | POT1       | 0.004406| Co-localization |
| GEMIN2    | BLMH       | 0.01327 | Co-localization |
| RAD17     | DCAF17     | 0.015909| Co-localization |
| RAD17     | POT1       | 0.005233| Co-localization |
| RAD17     | EYA3       | 0.015112| Co-localization |
| PTTG1     | DCAF17     | 0.200927| Genetic Interactions |
| NAA15     | SKIV2L2    | 0.060909| Physical Interactions |
| Protein 1 | Protein 2 | Score  | Interaction Type   |
|----------|----------|--------|--------------------|
| CUL4B    | DCAF17   | 0.115469 | Physical Interactions |
| RAB40B   | DCAF17   | 0.353314 | Physical Interactions |
| ZMAT3    | DCAF17   | 0.300627 | Physical Interactions |
| BLMH     | CUL4B    | 0.065557 | Physical Interactions |
| POLI     | CUL4B    | 0.13731  | Predicted          |
Figure. (2): Interaction between *DCAF17* gene and related genes using GeneMANIA.
**Figure. (3):** D69V: amino acid change from Aspartic acid to Valine at position 69.

**Figure. (4):** R72W: amino acid change from Arginine to Tryptophan at position 72.
Figure. (5): V186D: amino acid Valine change to Aspartic acid at position 186.

Figure. (6): P197L: amino acid Proline change to Leucine at position 197.
**Figure. (7):** L204P: aminoacid Leucine change to Proline at position 204.

**Figure. (8):** L224P: amino acid change from Leucine to Proline at position 224.
**Figure. (9):** L433S: amino acid Leucine change to Serine at position 4333.

**Figure. (10):** H478R: amino acid Histidine change to Arginine at position 478.
Figure. (11): D483G: amino acid Aspartic acid change to Glycine at position 483.

Figure. (12): F498V: amino acid Phenylalanine change to Valine at position 498.
Discussion:

Ten novel mutations have been found that affect the stability, structure and the function of *DCAF17* gene with one associated microRNA in the 3' -UTR region utilizing various *in silico* bioinformatics tools. These methods for prediction used standard tools and multiple disciplines to detect and characterize the missense variation. In order to ensure our results and increase the accuracy; eight web server *in silico* algorithm have been used: SIFT, Proven, Polyphen, SNAP 2, SNP&GO, PHD, Pmut and I mutant.

This gene considered to be with no function but it has relation to the disease Woodhouse-Sakati Syndrome with MIM no 241080 (from NCBI) \(^{(13, 45)}\). Because there is no similar mutation pathogenicity prediction study to this gene; this founding considered to be novel mutation \(^{(46), (13), (47)}\).

GeneMANIA did not indicated specific function for *DCAF17* gene, but there is interaction with another genes \(^{(47), (15), (48), (14)}\). The function of the gene is not clear until now \(^{(13)}\).

Project HOPE server was used to predict the effect of the amino acid substitutions on the molecular level of the protein. The ten SNPs were submitted to hope server individually and the obtained result indicated a change in the size, hydrophobicity and charge of the final protein, all the SNPs were found in Ddb1- And Cul4-Associated Factor 17 IPR031620 domain region in the gene, which further confirm its pathogenicity.

All the mutant amino acids except three ( R72W, V186D, P197L ) were found to be smaller in size in comparison to the wild type which may lead to loss of interactions with other molecules or residues. The hydrophobicity of the wild-type and mutant residue differs, there were 6 SNPs mentioned in project hope report ( D69V, R72W, V186D, L433P, H478L, D483G ) that has mutant amino acid with more hydrophobicity that could lead to loss of hydrogen bonds and/or disturb correct folding. There is a difference in charge between the wild-type and mutant amino acid, with three SNPS(D69V, R72W, D483G) mentioned in the report to have different neutral charge and one SNPs (V186D) with negative charge in the mutant amino acids which could cause a loss of interactions with other molecules or residues.

Different biological processes are controlled by transcriptional regulation done by non-coding RNA molecules (miRNA) consisting of 18–24 nucleotides in length which could activate and/or suppress protein translation inside the cells at post-transcriptional level.\(^{(49-52)}\) it was found in gene card site (https://www.genecards.org/ ) that *DCAF17* gene may be dysregulated by miRNA, so we analyzed the microRNA in the 3'-UTR region of *DCAF17* gene using multiple softwares (ensemble, regRNA2, miRmap and mirbase). We found one relevant miRNA; (hsa-miR-4436b-3p) with PhyloP conservation score of 0.71 and after analyzing its annotation through miRmap using sequence from 5 species human,mouse,worm,fly and Arabidopsis it appeared to have a role in inhibiting adipose triglyceride lipase, carcinogenesis and other possible roles.
This study was limited as it depended in computerized software and it need further wet lab studies for confirmation.

We now appreciate the regulatory role of the 3'-UTR in the function of DCAF17 protein and with this 10 novel mutation we could develop a precise biomarker for Woodhouse-Sakati Syndrome in the future.

**Conclusion:**
Ten novel SNPs were found to be deleterious in this paper with possible role as future biomarker for Woodhouse-Sakati syndrome, and one microRNA was found in the 3'-UTR to regulate the function of the protein.

**Acknowledgment:**
We appreciate the kind help and assistance we had in Africa City of Technology.

**Conflict of Interest:**
No conflict of interest to declare.

**Funding:**
The author(s) received no financial support for the research, authorship, and/or publication of this article.

**References:**
1. Rachmiel M, Bistritzer T, Hershkoviz E, Khatil A, Epstein O, Parvari R. Woodhouse-Sakati syndrome in an Israeli-Arab family presenting with youth-onset diabetes mellitus and delayed puberty. Horm Res Paediatr. 2011;75(5):362-6.
2. Hdiji O, Turki E, Bouzidi N, Bouchhima I, Damak M, Bohlega S, et al. Woodhouse-Sakati Syndrome: Report of the First Tunisian Family with the C2orf37 Gene Mutation. J Mov Disord. 2016;9(2):120-3.
3. Abdulla MC, Alazami AM, Alungal J, Koya JM, Musambil M. Novel compound heterozygous frameshift mutations of C2orf37 in a familial Indian case of Woodhouse-Sakati syndrome. J Genet. 2015;94(3):489-92.
4. Agopiantz M, Corbonnois P, Sorlin A, Bonnet C, Klein M, Hubert N, et al. Endocrine disorders in Woodhouse-Sakati syndrome: a systematic review of the literature. J Endocrinol Invest. 2014;37(1):1-7.
5. Abusrair AH, Bohlega S, Al-Semari A, Al-Ajlan FS, Al-Ahmadi K, Mohamed B, et al. Brain MR Imaging Findings in Woodhouse-Sakati Syndrome. AJNR Am J Neuroradiol. 2018;39(12):2256-62.
6. Al-Swailem SA, Al-Assiri AA, Al-Torbak AA. Woodhouse Sakati syndrome associated with bilateral keratoconus. Br J Ophthalmol. 2006;90(1):116-7.
7. Ben-Omran T, Ali R, Almureikhi M, Alameer S, Al-Saffar M, Walsh CA, et al. Phenotypic heterogeneity in Woodhouse-Sakati syndrome: two new families with a mutation in the C2orf37 gene. Am J Med Genet A. 2011;155A(11):2647-53.
8. Gurbuz F, Desai S, Diao F, Turkkahraman D, Wranitz F, Wood-Trageser M, et al. Novel inactivating mutations of the DCAF17 gene in American and Turkish families cause male infertility and female subfertility in the mouse model. Clin Genet. 2018;93(4):853-9.
9. Koshy G, Danda S, Thomas N, Mathews V, Viswanathan V. Three siblings with Woodhouse-Sakati syndrome in an Indian family. Clin Dysmorphol. 2008;17(1):57-60.
10. Medica I, Sepcic J, Peterlin B. Woodhouse-Sakati syndrome: case report and symptoms review. Genet Couns. 2007;18(2):227-31.
11. Alazami AM, Al-Saif A, Al-Semari A, Bohlega S, Zlitni S, Alzahrani F, et al. Mutations in C2orf37, encoding a nucleolar protein, cause hypogonadism, alopecia, diabetes mellitus, mental retardation, and extrapyramidal syndrome. Am J Hum Genet. 2008;83(6):684-91.
12. Schneider SA, Bhatia KP. Dystonia in the Woodhouse Sakati syndrome: A new family and literature review. Mov Disord. 2008;23(4):592-6.
13. Almeqdadi M, Kemppainen JL, Pichurin PN, Gavrilova RH. Phenotypic Variability of c.436delC DCAF17 Gene Mutation in Woodhouse-Sakati Syndrome. The American journal of case reports. 2018;19:347-53.
14. Sheridan MB, Wohler E, Batista DA, Applegate C, Hoover-Fong J. The Use of High-Density SNP Array to Map Homozygosity in Consanguineous Families to Efficiently Identify Candidate Genes: Application to Woodhouse-Sakati Syndrome. Case Rep Genet. 2015;2015:169482.
15. Dastsooz H, Nemati H, Fard MAF, Fardae M, Faghihi MA. Novel mutations in PANK2 and PLA2G6 genes in patients with neurodegenerative disorders: two case reports. BMC Med Genet. 2017;18(1):87.
16. Nishizaki SS, Boyle AP. Mining the Unknown: Assigning Function to Noncoding Single Nucleotide Polymorphisms. Trends in genetics : TIG. 2017;33(1):34-45.
17. Li L, Chen Q, Wei DQ. Prediction and functional analysis of single nucleotide polymorphisms. Current drug metabolism. 2012;13(7):1012-23.
18. Murray D, Doran P, MacMathuna P, Moss AC.
19. Habib R, Basit S, Khan S, Khan MN, Ahmad W. A novel splice site mutation in gene C2orf37 underlying Woodhouse-Sakati syndrome (WSS) in a consanguineous family of Pakistani origin. Gene. 2011;490(1-2):26-31.
20. Nanda A, Pasterнак SM, Mahmoudi H, Ishorst N, Grimalt R, Betz RC. Alopecia and hypotrichosis as characteristic findings in Woodhouse-Sakati syndrome: report of a family with mutation in the C2orf37 gene. Pediatr Dermatol. 2014;31(1):83-7.
21. Wang T, Mao J, Liu MJ, Choy KW, Li HB, Cram DS, et al. A patient with five chromosomal rearrangements and a 2q31.1 microdeletion. Clin Chim Acta. 2014;430:129-33.
22. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;Chapter 7:Unit 7 20.
23. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. 2002;12(3):436-46.
24. Manickam M, Ravanam P, Singh P, Talwar P. In silico identification of genetic variants in glucocerebrosidase (GBA) gene involved in Gaucher's disease using multiple software tools. Front Genet. 2014;5:148.
25. Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function. 2006;7:61-80.
26. Schneider G, Hu J, Sim N-L, Kumar P, Henikoff S, Ng PC.
27. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. PLoS One. 2012;7(10):e46688.
28. Kumar P, Mahalingam K. In silico approach to identify non-synonymous SNPs with highest predicted deleterious effect on protein function in human obesity related gene, neuronal growth regulator 1 (NEGR1). 3 Biotech. 2018;8(11):466.
29. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic acids research. 2002 Sep 1;30(17):3894-900.
30. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nature methods. 2010;7(4):248-9.
31. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio RJHm. Functional annotations improve the predictive score of human disease-related mutations in proteins. 2009;30(8):1237-44.
32. López-Ferrando V, Gazzo A, de la Cruz X, Orozco M, Gelpí JLNar. PMut: a web-based tool for the annotation of pathological variants on proteins, 2017 update. 2017;45(W1):W222-W8.
33. Capriotti E, Fariselli P, Casadio RJNar. I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure. 2005;33(suppl_2):W306-W10.
34. Venselaar H, te Beek TA, Kuipers RK, Hekkelman ML, Vriend GJBB. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. 2010;11(1):548.
35. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, et al. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. 2010;38(suppl_2):W214-W20.
36. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. 2004;25(13):1605-12.
37. Cunningham F, Achuthan P, Akanni W, Allen J, Amode MR, Armean IM, et al. Ensembl 2019. Nucleic Acids Res. 2019;47(D1):D745-D51.
38. Frankish A, Vullo A, Zadissa A, Yates A, Thormann A, Parker A, et al.
39. Huang HY, Chien CH, Jen KH, Huang HD. RegRNA: an integrated web server for identifying regulatory RNA motifs and elements. Nucleic Acids Res. 2006;34(Web Server issue):W429-34.
40. Chang TH, Huang HY, Hsu JB, Weng SL, Horng JT, Huang HD. An enhanced computational platform for investigating the roles of regulatory RNA and for identifying functional RNA motifs. BMC Bioinformatics. 2013;14 Suppl 2:S4.
41. Vejnar CE, Zdobnov EM.
42. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. Nucleic Acids Res. 2019;47(D1):D155-D62.
43. Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res. 2014;42(Database issue):D68-73.
44. Kozomara A, Birgaoanu M, Griffiths-Jones S.
45. Ali A, Mistry BV, Ahmed HA, Abdulla R, Amer HA, Prince A, et al. Publisher Correction: Deletion of DDB1- and CUL4- associated factor-17 (Dcaf17) gene causes spermatogenesis defects and male infertility in mice. Scientific reports. 2018;8(1):11779.
46. Ali A, Mistry BV, Ahmed HA, Abdulla R, Amer HA, Prince A, et al. Deletion of DDB1- and CUL4- associated factor-17 (Dcaf17) gene causes spermatogenesis defects and male infertility in mice. Scientific reports. 2018;8(1):9202.

47. Arber CE, Li A, Houlden H, Wray S. Review: Insights into molecular mechanisms of disease in neurodegeneration with brain iron accumulation: unifying theories. Neuropathology and applied neurobiology. 2016;42(3):220-41.

48. Gispert S, Brehm N, Weil J, Seidel K, Rub U, Kern B, et al. Potentiation of neurotoxicity in double-mutant mice with Pink1 ablation and A53T-SNCA overexpression. Human molecular genetics. 2015;24(4):1061-76.

49. Kamaraj B, Gopalakrishnan C, Purohit R. In silico analysis of miRNA-mediated gene regulation in OCA and OA genes. Cell Biochem Biophys. 2014;70(3):1923-32.

50. Kristjansdottir K, Fogarty EA, Grimson A. Systematic analysis of the Hmga2 3' UTR identifies many independent regulatory sequences and a novel interaction between distal sites. RNA. 2015;21(7):1346-60.

51. Liang B, Zhao J, Wang X. A three-microRNA signature as a diagnostic and prognostic marker in clear cell renal cancer: An In Silico analysis. PLoS One. 2017;12(6):e0180660.

52. Yu L, Xiang L, Feng J, Li B, Zhou Z, Li J, et al.