Identification of Novel Key Biomarkers in Simpson-Golabi-Behmel Syndrome (SGBS): Evidence from Bioinformatics Analysis

Mujahed I. Mustafa1*, Abdelrahman H. Abdelmoneim1, Nafisa M. Elfadol1, Naseem S. Murshed1, Zainab O. Mohammed2, Mohamed A. Hassan1

1 Department of Biotechnology, Africa City of Technology, Sudan.
2 Hematology Department, Ribat University Hospital, Sudan.

* Corresponding author email: mujahedmustafa@gmail.com

Received: 26 May 2019 / Revised: 10 July 2019 / Accepted: 12 July 2019 / Published: 21 July 2019

ABSTRACT

The Simpson-Golabi-Behmel Syndrome (SGBS) or overgrowth Syndrome is an uncommon genetic X-linked disorder highlighted by macrosomia, renal defects, cardiac weaknesses and skeletal abnormalities. The purpose of the work was to classify the functional nsSNPs of GPC3 to serve as genetic biomarkers for overgrowth syndrome. The raw data of GPC3 gene were retrieved from dbSNP database and used to examine the most damaging effect using eight functional analysis tools, while we used I-mutant and MUPro to examine the effect of SNPs on GPC3 protein structure; The 3D structure of GPC3 protein is not found in the PDB, so RaptorX was used to create a 3D structural prototype to visualize the amino acids alterations by UCSF Chimera. For biophysical validation we used project HOPE; Lastly we run conservational analysis by BioEdit and Consurf web server respectively. Our results revealed three novel missense mutations (rs1460413167, rs1295603457 and rs757475450) that are that are more likely to be responsible for disturbance in the function and structure of GPC3. This work provides new insight into the molecular basis of overgrowth Syndrome by evidence from bioinformatics analysis. Three novel missense mutations (rs757475450, rs1295603457 and rs1460413167) are more likely to be responsible for disturbance in the function and structure of GPC3; therefore, they may be assisting as genetic biomarkers for overgrowth syndrome. As well as these SNPs can be used for the larger population-based studies of overgrowth syndrome.

Keywords: Bioinformatics analysis; Diagnostic markers; GPC; nsSNPs; Overgrowth syndrome.

1 Introduction

The Simpson-Golabi-Behmel Syndrome (SGBS) or Overgrowth Syndrome is an uncommon genetic disorder characterized by macrosomia, renal defects, cardiac weaknesses and skeletal abnormalities.[1-4] the first case was reported around 1940.[5] So far, two unlike types of overgrowth Syndrome have been defined. The typical SBGS type one [2-8] and a fatal and rare system, possibly 10 conditions defined known as SGBS type two.[9-11] Furthermore, these cases could rapidly develop Wilms’ malignancy.[12] Early passing is more common.[13] Different mutations have been described in SGBS type one.[14-22] Overgrowth Syndrome caused by mutations in glypican 3 (GPC3) gene is localized on Xq26.1 [23, 24] which encrypts glypican-3. [17, 19, 20, 25-29] that seemingly acting a bad part in growth control by an anonymous fate. However, outcomes from an exhaustive qualified study of growth forms in dual mutants missing GPC3 provided conclusive genetic evidence inconsistent with the theory that GPC3 performances as a growth suppressor.[29] Such a proteoglycan is contingent to show a vital part in regulate and diagnosis in mesodermal tissues and in tumors predisposition.[30, 31] Some studies show association between GPC3 gene and some types of human cancers.[32-36]
The aim of this work was to detect the most deleterious SNPs in GPC3 that may cause overgrowth Syndrome type one by using of different bioinformatics tools. Furthermore and to be used as genetic biomarkers. nsSNP is an alteration that occurs in a single base pair of amino acid which leads to disturb or change the corresponding protein’s function, if the second possibility happened, it may cause a severe phenotypic impact and in return responsible for the pathology of the disease [37, 38] Clinical testing for deleterious SNPs frequently discloses alterations that are not easily considered as deleterious, for that reason a great effort has been done by translational bioinformatics tools for analysis of nsSNPs which have improved significantly in recent years and thus become more reliable for SNPs analysis,[39] Translational analysis has been considered as an essential science in the field of personalized medicine which aims to fill the gap between clinical and academic research by prioritizing the most pathogenic nsSNPs for further studies.[40-44] This is the first computational analysis of GPC3 gene that classify nsSNPs for the larger population-based studies of overgrowth syndrome.

2 Methods
2.1 Data Mining
The raw data of GPC3 gene were retrieved from National Center for Biotechnology Information (NCBI) website.[45, 46] The reference sequence of the protein was retrieved from Uniprot database.[47]

2.2 Functional Analysis
2.2.1 SIFT
It is the first functional analysis online tool which designed to predict whether a SNP is damaging or not by specific algorithm have a score <0.05 are predicted to be damaging SNP, otherwise it reflected to be not damaging.[48, 49]

2.2.2 PolyPhen
It is a functional analysis online tool to examine potential influences of a SNP on functional and structural characteristics of our protein of interest.[50, 51]

2.2.3 PROVEAN
It is a functional analysis online tool which we used to calculate if a SNP has an impression on the physical role of our protein of interest. PROVEAN probability has two possibilities, deleterious or neutral with cutoff -2.5.[52]

2.2.4 SNAP
It is a functional analysis tool with an artificial intelligence machine device called "neural network"; It distinguishes between effect and neutral variants/non-synonymous SNPs by taking a variety of sequence and variant features into account. [53, 54]

2.2.5 SNPs&GO
It is a functional analysis tool which distinguishes between the damaging SNPs from the neutral ones. The other methods were used too (PHD-SNP and PANTHER).[55, 56]

2.2.6 P-Mut
It is an online functional analysis tool for the clarification of amino acid alternates on proteins, permits the swift and accurate intention (80%) of the obsessive characteristics of each SNP stranded on the preparation of neural systems.[57, 58]

2.3 Stability Analysis
2.3.1 I-Mutant 3.0
It is SVM-based (Support Vector Machine) tool for the automatic prediction of protein stability changes upon single point mutations. The predictions are performed starting either from the protein structure or, more importantly, from the protein sequence.[59, 60]

2.3.2 MUPro
It is an online tool we used; it runs by the same concept of I-Mutant 3.0 but it’s more accurate than I-Mutant 3.0 by 84.2%.[61, 62]
2.4 Biophysical and Visualization Analysis

2.4.1 Project Hope
It is an online web-server for biophysical validation which brings together a series of related protein data to form a model if there are enough 3D structural data; also to run this data to predict if the amino acid alteration may affect in the protein function or not.[63]

2.4.2 RaptorX
The 3D structure of the protein of GPC3 it is not found at protein data bank (PDB), so RaptorX was used to perform a 3D structure model for GPC3 protein.[64, 65]

2.4.3 UCSF Chimera
It is a visualization analysis program of 3D structure model, docking analysis and so many related analyses; the predicted model was used to visualize and compare the amino acid alterations by UCSF Chimera [66, 67].

2.5 Conservational Analysis

2.5.1 BioEdit
It is a program package created to stream a distinct program that can run approximately any sequences operation, demonstrating, as well as a few basic alignment studies.[68]

2.5.2 ConSurf Server
It is proposing evolutionary conservation outlines for proteins of known structure in the PDB. ConSurf red flag the similar amino acid sequences and run multi alignment approaches. The conserved regions of amino acids identify its site by using particular system.[69, 70]

3 Results
The effect of each SNP has been studied regarding to function and stability of the protein by different computational analysis tools with different considerations and features, in order to decrease the error to the lowest ratio possible (Figure 1).

![Figure 1: Illustrative Workflow used for SNPs analysis](image-url)
The total number of SNPs regarding to \textit{GPC3} gene is 765 SNPs, out of 256 nsSNPs were submitted to SIFT, PolyPhen-2, PROVEAN and SNAP2 respectively. SIFT predicted 109 damaging mutations, PolyPhen-2 predicted 115 deleterious mutations (50 possibly damaging (less confident prediction) and 65 probably damaging (more confident prediction)), PROVEAN predicted 82 deleterious mutations and SNAP2 predicted 127 damaging mutations. Once we filtered the four positive deleterious mutations, the number of SNPs reduced to 7. (Table 1) after that, the same 7 mutations were submitted to SNPs&GO, PHD-SNP, PANTHER and P-Mut for further study to examine their influence on the function of \textit{GPC3}; 7 deleterious mutations were predicted by PHD-SNP and P-mut, SNP&GO predicted 5, while PANTHER predicted 5 deleterious mutations. Once we filtered the four positive deleterious mutations the number reduced to 4 SNPs (Table 2) after that, we submitted them to I-Mutant and MUPro to investigate their effect on the stability; The two online tools revealed that, All the mutations decreased the protein stability, except for one SNP (G257D) was predicted by I-Mutant to increase the stability of the protein (Table 3).

4 Discussion

A significant interest in \textit{Homo sapiens} genome has been focused to classify the deleterious SNPs; those are more likely to be responsible for inherited disorders. Therefore, a good effort was dictated to identify the most deleterious SNPs that may cause overgrowth syndrome. Our analysis revealed three novel SNPs in \textit{GPC3} gene which were classified as highly deleterious SNPs, which as crucial impact at the functional level of the \textit{GPC3} gene, our analysis based on different sequence and structure-based algorithms, Figure (1).

There is a study that has been reported which shows a missense mutation that causes overgrowth syndrome; [19] which matches with
this study findings. Some studies show association between GPC3 gene and some types of liver cancer such as hepatocellular carcinoma.[30, 34, 71, 72] Therefore, this study can open the door for novel diagnostic biomarkers for hepatocellular carcinoma. Combination detection of serum GPC3 and pathogenic SNPs through clinical and genetic testing must be positively matched; this can enhance accuracy and efficiency of hepatocellular carcinoma diagnosis. In addition it confirms that (W296R) is pathogenic; this result matches with the result found previously in dbSNPs database. Furthermore, these mutations (P212H, C65Y, R39C) were recovered as untested, in this study were found to be all pathogenic.

At the functional level analysis, our results showed that all these nsSNPs substitutions (D500Y, G440R, W296R, P212H, F208L, C65Y, and R39C) were classified as likely pathogenic mutations, Table (1) the prediction efficacy has been increased by integrating the results of SIFT, PolyPhen-2, PROVEAN and SNAP2 based approaches, by combining the predictions of SNPs&GO, PhD-SNP, PANTHER, and P-Mut, Table (2) the output showed that all these nsSNPs (W296R, P212H, C65Y and R39C) are classified as highly pathogenic mutations. Therefore, our functional analysis suggested that these four nsSNPs might disrupt both the protein function and structure; while at the structural level analysis, MUPro results showed a decrease in stability for All these SNPs (W296R, P212H, C65Y and R39C) while I-Mutant results showed a decrease in stability for these SNPs (W296R, P212H and R39C), thus suggesting that these mutations could directly or indirectly destabilize the amino acid interactions triggering functional deviations of protein to some point. Table (3) The most four deleterious SNPs were submitted to project HOPE which shown that all they are located in a domain of GPC3 protein; therefore, they may have a dynamic alteration in the protein function; In (Figure 2): (R39C): Shows the schematic structures of the original amino acid (in the left) which is Arginine and the mutant one (in the right) which is Cysteine. The backbone, which is the same for each amino acid, is colored red (in the green and red boxes) and the side chain, unique for each amino acid, is colored black. In addition, figure shows Close-up angle of the mutation. The protein is colored white, wide type residue colored green and mutant one colored red in position 39. The mutant residue is smaller than the wild-type residue; the wild-type residue charge was positive, while the mutant residue charge is neutral, this can cause loss of interactions with other molecules or residues; the mutant residue is more hydrophobic than the wild-type residue, and this can result in loss of hydrogen bonds and/or disturb correct folding.

Figure 2: (rs757475450) (R39C) Arginine changes to Cysteine at position 39; illustrated by chimera (v 1.8) and project HOPE.

Figure 3: (rs1295603457): (C65Y) Cysteine changes to Tyrosine at position 65; illustrated by chimera (v 1.8) and project HOPE.
the green and red boxes) and the side chain, unique for each amino acid, is colored black. In addition, figure shows Close-up angle of the mutation. The protein is colored white, wide type residue colored green and mutant one colored red in position 65. The wild-type and mutant amino acids differ in size; the mutant residue is bigger, this might lead to bumps; the hydrophobicity of the wild-type and mutant residue differs; hydrophobic interactions, either in the core of the protein or on the surface, will be lost.

In (Figure 5): (W296R): Shows the schematic structures of the original amino acid (in the left) which is Tryptophan and the mutant one (in the right) which is Arginine. The backbone, which is the same for each amino acid, is colored red (in the green and red boxes) and the side chain, unique for each amino acid, is colored black. In addition, figure shows Close-up angle of the mutation. The protein is colored white, wide type residue colored green and mutant one colored red in position 296.

The wild-type residue charge was neutral, the mutant residue charge is positive, the mutation introduces a charge, and this can cause repulsion of ligands or other residues with the same charge; the wild-type and mutant amino acids differ in size, and the mutant residue is smaller, this might lead to loss of interactions; The hydrophobicity of the wild-type and mutant residue differs, hydrophobic interactions, either in the core of the protein or on the surface, will be lost.

We also observed that, all the four SNPs were located in conserve region. We believe that amino acids conserved across species are playing a crucial role at the functional level; therefore, the four SNPs that we have detected are more probable disease causing ones; (Figure 6) The same results were confirmed by ConSurf, which show the nsSNPs that they are located at extremely conserved sites; therefore, we have confidence that these SNPs have a tendency to
be the most deleterious SNPs that may cause overgrowth syndrome. (Figure 7)
This study is the first computational approach while all other earlier studies were in vitro, in vivo and whole exome sequencing. [73-76] It revealed three novel missense mutations that are more likely to be responsible for disturbance in the function and structure of GPC3; therefore, they could be used as diagnostic markers to Predict overgrowth syndrome.[77] Lastly, some appreciations of wet lab techniques are suggested to support our in silico analysis findings.

**Figure 6:** GPC3 Family of seven protein sequences representing that, the normal amino acids are expected to be altered (showed by red arrows) are evolutionarily conserved across species.
Figure 7: Shows the conserved amino acids across species in GPC3 protein were determined using Consurf.

(a) An exposed residue according to the neural-network algorithm via an orange letter.
(b) Residues predicted to be buried are demonstrated via a green letter.
(f) A predicted functional residue (highly conserved and exposed) are indicated with a red letter.
(s) A predicted structural residue (highly conserved and buried) that are demonstrated with a blue letter.
(I) Insufficient data- the calculation for this site was performed on less than 10% of the sequences are demonstrated via a yellow letter.
5 Conclusion

Functional and structural impact of SNPs in the GPC3 gene was found out by using computational prediction tools; Out of a total of 765 SNPs in the GPC3 gene, 256 were nsSNPs; out of 256 missense nsSNPs, 4 were found to be the most deleterious nsSNPs (three of them were novel R39C (rs757475450), C65Y (rs1295603457), and P212H (rs1460413167)) by eight functional analysis tools. Stability analysis results showed that the amino acid residue substitutions which had the greatest impact on the stability of the GPC3 protein were mutations R39C (rs757475450), C65Y (rs1295603457), P212H (rs1460413167) and W296R (rs104894854). This result helped us to characterize the impact of nsSNPs on GPC3 gene and should be considered important candidates in causing of overgrowth syndrome.

6 Declarations

6.1 Acknowledgment

The authors wish to acknowledge the enthusiastic cooperation of Africa City of Technology - Sudan.

6.2 Competing Interests

The authors declare that no conflict of interest exist in this publication.

7 How to Cite this Article:

Will be updated in the final version.

References


[46] "NCBI website.


[52] Y. Choi, G. E. Sims, S. Murphy, J. R. Miller, and A. P. Chan, "Predicting the functional effect of amino acid


[54] **SNAP2 server.** Available: https://rostlab.org/services/snap2web/


[58] **P-Mut server.** Available: http://tamb.irbbarcelona.org/PMut


[60] **I-Mutant server.** Available: http://gpcr2.biocomp.unibo.it/cgi/predictors/I-Mutant3.0/I-Mutant3.0.cgi


[63] **RaptorX server.** Available: (http://raptorx.uchicago.edu/)


[66] **UCSF Chimera.** Available: http://www.cgl.ucsf.edu/chimera/


---

**Publish your research article in AJJR journals**: 
- Online Submission and Tracking
- Peer-Reviewed
- Rapid decision
- Immediate Publication after acceptance
- Articles freely available online
- Retain full copyright of your article.

Submit your article at journals.aijr.in

---

**Publish your books with AJJR publisher**: 
- Publish with ISBN and DOI.
- Publish Thesis/Dissertation as Monograph.
- Publish Book Monograph.
- Publish Edited Volume/ Book.
- Publish Conference Proceedings
- Retain full copyright of your book.

Submit your manuscript at books.aijr.org

---

ISSN: 2456-7132
Available online at Journals.aijr.in