Prediction of Four Novel SNPs V17M, R11H, A66T, and F57S in the SBDS Gene Possibly Associated with Formation of Shwachman-Diamond Syndrome, using an Insilico Approach

Anfal Osama Mohamed Sati^{*1}, Sara Ali Abdalla Ali¹, Rouaa Babikir Ahmed Abdulllah¹, Manal Satti Awad Elsied¹, Mohamed A. Hassan¹

¹ Department of Applied Bioinformatics - Africa City of Technology, Biotechnology Park, Khartoum- Sudan

*corresponding author:

Anfal Osama Mohamed Sati

E-mail: Anfalos93@hotmail.com

Abstract:

Shwachman-Diamond syndrome SDS (MIM 260400) is an autosomal recessive disorder. Characterized by exocrine insufficiency of the pancreas, bone marrow hypoplasia resulting in cytopenias, especially neutropenia, variable degree of skeletal abnormalities, failure to thrive, and increasing risk of developing myelodysplasia or transformation to leukemia. SDS is mainly caused by Shwachman Bodian Diamond Syndrome gene (MIM ID 607444). SBDS gene is 14899 bp in length, located in chromosome seven in the eleventh region of the long arm, and is composed of five exons. It's a ribosomal maturation factor which encodes a highly conservative protein that has widely unknown functions despite of its abundance in the nucleolus. A total number of 53 SNPs of homo sapiens SBDS gene were obtained from the national center for biotechnology information (NCBI) analyzed using translational tools, 7 of which were deleterious according to SIFT server and were further analyzed using several software's (Polyhen-2, SNPs&Go, I-Mutant 2.0, Mutpred2, structural analysis software's and multiple sequence alignment software). Four SNPs (rs11557408 (V17M), rs11557409 (R11H), rs367842164 (A66T), and rs376960114 (F57S)) were predicted to be disease causing, localized at highly conservative regions of the SBDS protein and were not reported in any previous study. In addition, the study predicts new functions of the SBDS protein DNA related, and suggests explanations for Patients developing cytopenias and failure to thrive through genetic co-expression and physical interaction with RBF1 and EXOSC3, respectively.

Keywords:

Shwachman Diamond Bodian, Shwachman Diamond Syndrome, SDS, SBDS gene, Single Nucleotide Polymorphisms, SNPs, mutations, Bioinformatics, insilico, bioinsilico, single nucleotide variations.

Introduction

Shwachman-Diamond syndrome (SDS; MIM [260400](#)) is an autosomal recessive disorder. Characterized by exocrine insufficiency of the pancreas, bone marrow hypoplasia resulting in cytopenias, especially neutropenia, variable degree of skeletal abnormalities, failure to thrive, and increasing risk of developing myelodysplasia or transformation to leukemia. Other common features were found in younger patients. Such as, hepatomegaly, abnormal liver and biochemical tests findings.^{1–13} SDS is a rare disorder which counts for 1:77000 individuals¹⁴. The hematological abnormalities was a constituent in 88% of the cases. Neutropenia in 95%, anemia in 50%, and thrombocytopenia in 70% of cases. 50% of the cases presented with short stature and severe infections occurred in 85% of cases. In addition, some degree of mental retardation and learning disabilities was reported in some of the studied cases.^{10,15–18}

SDS is mainly caused by Shwachman Bodian Diamond Syndrome gene (*SBDS*, gene bank ID 51119).^{12,15,16,19,20} *SBDS* gene is 14899 bp in length, located in chromosome seven in the eleventh region of the long arm, and is composed of five exons^{21,22}. *SBDS* gene is a ribosomal maturation factor encodes a highly conservative protein which has widely unknown functions despite of its abundance in the nucleolus. However, in *SBDS* orthologs it's known to play a vital role in ribosomal biogenesis.^{20,21,23–27} *SBDS* along with the 60S large ribosomal subunit migrate and precipitates with 28S ribosomal RNA (rRNA). However, loss of *SBDS* Protein has not been reported to be associated directly with an individual block in rRNA maturation. Yet, there's a notable increase in immature 60S ribosomal subunits.^{5,26} Additionally, *SBDS* forms a protein complex with nucleophosmin, a protein involved in maintenance of genomic integrity, ribosomal biogenesis and generation of leukemia.^{26,28,29}

Shwachman-Diamond syndrome is an interesting disease with a high probability of transitioning to leukaemia and myelodysplasia and a better understanding of the clinical aspects of this rare disorder would assist in providing further information about its pathology. Many previous studies had various approaches to understand the pathophysiology of the *SBDS* gene and the possible impact of its chromosomal aberrations in the formation of the disease.^{3,20,21,23,26,30–32} However, the key answer to these questions remains unknown. Therefore, the aim of this study, is to analyze the single nucleotide polymorphisms (SNPs) of the *SBDS* gene to unveil any possible relationships between single amino acid substitution mutations (missense mutations) and development of SDS.

In this paper, translation technology was used to analyze SNPs of the *SBDS* gene and assess both gene and protein functions and the impact of these mutations on gene expression. This study is different than other studies for being the first to use computational tools to analyze *SBDS* gene SNPs.

Methodology:

A total number of 51 SNPs of Homo sapiens SBDS gene were obtained from the National Center for Biotechnology Information (NCBI) on October 2018, of which twenty-four SNPs were in the 5' and 3' un-translated regions, and twenty-seven non-synonymous SNPs (nsSNPs) were in coding regions. The latter, was analyzed using several softwares with different approaches and working algorithms, this is to ensure accuracy of results as insilico tools work consistently, and thus, one tool compensates for other tool's weaknesses.(Figure 1)³³

Figure (1): a hierarchy graph showing workflow of *SBDS* SNPs analysis

Sequence Databases:

National Center for Biotechnology Information (NCBI):

Is the largest database concerned in providing biomedical and genomic information, it was used in this study to obtain the single nucleotide polymorphisms (SNPs) of the *SBDS* gene from dbSNP, and reference protein sequences of *SBDS* family (*Homo sapiens*, *Mus Musculus*, *Rattus Norvegicus*, *Felis Catus*, *Sus Scrofa*, *Xenopus Laevis*) with accession numbers ([NP_057122.2], [NP_075737.1], [NP_001008290.1], [XP_003998651.1], [NP_001231322.1], and [NP_001092159.1], respectively). Adding to that, it was used for obtaining precise information concerning the gene of the study from genebank and nucleotide databases. (Available at: <https://www.ncbi.nlm.nih.gov/>)

ExPaSy:

(The Expert Protein Analysis System) is a proteomic database by the Swiss Institute of Bioinformatics (SIB). It provides access to a variety of protein databases and proteomic analytical tools by using protein ID. It has been used to obtain the UniProt entry.^{11,34,1} (Available at: <https://www.expasy.org/>)

UniProt:

A world wide database for protein sequences and related information. Data can be collected either by SNPs ID or protein ID which is provided by SIFT server, it provides full results on sequence in the shape of detailed information about protein structure, function, interactions and reported mutations, or simply as FASTA format to be used on other programs (e.g.: SNP&GO, and I-mutant 2.0). (Available at: <http://www.uniprot.org/>).

Functional analysis of SNPs:

SIFT (Sorting Intolerant from Tolerant) is an online tool to predict whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids. SIFT has been applied to human variant databases and was able to predict mutations (deleterious) involved in disease from (tolerant) polymorphisms which has no ability to alter protein. Thus, cause a disease .The cutoff value in the SIFT program is a tolerance index of ≥ 0.05 . The higher the tolerance index, the less functional impact a particular amino acid substitution is likely to have.³⁵ (table 1, 2) (Figure 2) (Available at: <http://sift.jcvi.org/>)

Polyphen-2:

PolyPhen-2 (Polymorphism phenotyping volume 2) is an online translational tool to predict the possible impact of single amino acid substitutions on the function of human proteins using structural and comparative evolutionary considerations, through multiple sequence

alignment and protein 3D structure analysis. It also calculates position-specific independent count scores (PSIC) for each of the two variants, and then calculates the PSIC scores difference between two variants for quantitative assessment of the severity of the effect on protein function. Prediction outcomes could be classified as [benign, possibly damaging or probably damaging] according to the value of PSIC with scores designated as (0.00–0.4), (0.5–0.8), and (0.9 –1) respectively. (Table 2)(Figure 3) (available at: <http://genetics.bwh.harvard.edu/pph2/>).^{7,96}

SNPs&Go:

It's an online software or program, It illustrate three results based on three different analytical algorithms, panther result, PHD-SNP result, and SNPs&GO result. All the results are contained in three main parts, the prediction which decides whether the mutation is neutral or disease related, reliability index (RI), and disease probability (if >0.5 mutation is disease causing).³⁶⁴(table 3)(figure 4) (Available at: <http://snps.biofold.org/snps-and-go/snps-and-go.html>)

PhD-SNP:

(Predictor of Human Deleterious Single Nucleotide Polymorphisms). An online support vector machine (SVM) based classifier, uses protein sequence information which can predict if the new phenotype produced after nsSNP can be related to a genetic disease or as a neutral polymorphism.³⁷ (table 3) (Figure 4) (Available at: <http://snps.biofold.org/phd-snp/phd-snp.html>)

I-mutant 2.0:

Is a support vector machine (SVM) tool for the prediction of protein stability free-energy change ($\Delta\Delta G$ or $\Delta\Delta G$) on a specific nsSNP. It predicts the free energy changes starting from either the protein structure or the protein sequence. A negative $\Delta\Delta G$ value means that the mutation decreases the stability of the protein, while a positive $\Delta\Delta G$ value indicates an increase in stability. The result show (prediction, reliability index RI, and a $\Delta\Delta G$ value³⁸(table 3) (Figure 5) (Available at: <http://folding.biofold.org/cgi-bin/i-mutant2.0.cgi>)

Structural analysis of SNPs:

MutPred2:

Is a tool that detects the effect of missense mutations on protein structure by providing it with the protein sequence in a FASTA format, mutation position, wild and mutant types. The result contains Mutpred2 score which is the probability that the amino acid substitution is pathogenic [>0.50 is considered pathogenic], affected PROSITE and ELM motifs, and a table with the structural changes occurred whether loss/gain of a helix, strand or loop. Also, gives other structural changes information if any and a P-value.^{39,40} (table 4) (Available at: <http://mutpred2.mutdb.org/index.html>)

Chimera 1.13.1:

Is a software program for analysis of molecular structures, visualizes protein in 3D structure of the protein and then change it between native and mutant amino acids with the candidate to display the impact that can be produced. Chimera accepts the input in the form of PDB ID or PDB file.⁴¹⁻⁴³ (Figure 6) (Available at: <http://www.cgl.ucsf.edu/chimera>.)

Multiple Sequence Alignment:

Using **Clustal W** package, **Bioedit** software which is a nucleic acid/protein alignment and BLAST tool. It was used to align six of SBDS protein family to illustrate areas of conservation and identities⁴⁴. Areas of high conservation are the areas associated with protein functions. Thus, mutations located at highly conservative regions (motifs) are more capable of altering protein structure and function and causing pathology than mutations at areas of low conservation.⁴⁵⁻⁴⁷ (Figure 7)

Prediction of gene functions and interactions:

Using **GeneMANIA version 3.6.0** which is an online server that finds genes which are related to an input gene, through functional association data. These data provide gene to gene (co-expression), gene to protein (physical interactions), pathways the gene is a part of and co-localization of the gene with a high accuracy.⁴⁸ The genes co-expressed and/or interacting with SBDS gene and their descriptions are shown in (table 5) (figures 8, 9). (Available at: <http://www.genemania.org>)

Results:

Table (1): Single nucleotide polymorphisms (SNPs) predicted to be tolerant using SIFT server

SNPs ID	Amino acid change	Protein ID	SIFT Score	SIFT Prediction
rs73151675	F217F	ENSP00000246868	1	Tolerated
rs113993989	L47L	ENSP00000246868	1	Tolerated
rs140297326	H54Y	ENSP00000246868	0.18	Tolerated
rs145800188	L12L	ENSP00000246868	1	Tolerated
rs146166884	L237L	ENSP00000246868	1	Tolerated
rs146780418	S143S	ENSP00000246868	1	Tolerated
rs147652512	V43L	ENSP00000246868	0.176	Tolerated
rs181546379	L239L	ENSP00000246868	1	Tolerated
rs188752805	D246N	ENSP00000246868	0.088	Tolerated
rs189441106	C119C	ENSP00000246868	1	Tolerated
rs200346194	S61S	ENSP00000246868	1	Tolerated
rs201070132	V130M	ENSP00000246868	0.092	Tolerated
rs201668348	T113T	ENSP00000246868	1	Tolerated
rs367957034	I111I	ENSP00000246868	1	Tolerated
rs371133001	E222Q	ENSP00000246868	0.116	Tolerated
rs371739884	G63G	ENSP00000246868	1	Tolerated

rs373073532	S199G	ENSP00000246868	0.392	Tolerated
rs373730800	K33R	ENSP00000246868	0.201	Tolerated
rs374083295	A116A	ENSP00000246868	1	Tolerated
rs377761960	K240R	ENSP00000246868	0.273	Tolerated
		ENSP00000246868		

Table (2): SNPs results using SIFT server and Polyphen-2 server.

SNP ID	SIFT			polyPhen-2	
	Amino acid change	score	Prediction	value	prediction
rs11557408	V17M	0.002	Deleterious	0.999	Probably damaging
rs11557409	R11H	0.012	Deleterious	0.988	Probably damaging
rs28942099	N8K	0.032	Deleterious	0.970	Probably damaging
rs79344818	I212T	0.017	Deleterious	0.145	Benign
rs113993995	R126T	0.003	Deleterious	0.990	Probably damaging
rs376960114	F57S	0	Deleterious	0.988	Probably damaging
rs367842164	A66T	0.002	Deleterious	0.858	Possibly damaging

Table (3): Single nucleotide polymorphisms SNPs analysis using SNPs&GO and I-mutant 2.0 server

SNP ID	Amino acid change	SNPs&GO			PHD-SNP			I-mutant 2.0	
		Value	prediction	RI	value	prediction	RI	prediction	RI
rs11557408	V17M	0.256	Neutral	5	0.427	Neutral	1	Increased	2
rs11557409	R11H	0.19	Neutral	6	0.536	Disease	1	Decreased	8
rs28942099	N8K	0.348	Neutral	3	0.667	Disease	3	Increased	4
rs113993995	R126T	0.291	Neutral	4	0.514	Disease	0	Increased	4
rs367842164	A66T	0.59	Disease	2	0.71	Disease	4	Decreased	7
rs376960114	F57S	0.812	Disease	6	0.871	Disease	7	Decreased	9

Table (4): Showing structural effects of mutation based on Mutpred2 tool.

SNPs ID	UnitProt ID	Amino acid change	Mutpred2 Score	PROSITE/ELM motifs
rs11557408	Q9Y3A5	V17M	0.645	—
rs11557409	Q9Y3A5	R11H	0.346	—

rs28942099	Q9Y3A5	N8K	0.857	ELME000052 ELME000063 ELME000153 ELME000159 ELME000233 ELME000358
rs113993995	Q9Y3A5	R126T	0.644	ELME000100 ELME000108 ELME000136 ELME000142 ELME000159 ELME000233 PS00005
rs367842164	Q9Y3A5	A66T	0.753	ELME000002 ELME000064 ELME000220 PS00005 PS00006
rs376960114	Q9Y3A5	F57S	0.912	ELME000052 ELME000053 PS01267

Table (5): list of genes co-expressed/physically interacting with *SBDS* gene and their description:

Gene	Description	Co-expression	Physical interaction
<i>EFLI</i>	elongation factor like GTPase 1	No	Yes
<i>CLN3</i>	ceroid-lipofuscinosis, neuronal 3	No	Yes
<i>TPT1</i>	tumor protein, translationally-controlled 1	Yes	No
<i>TYMS</i>	thymidylate synthetase	Yes	Yes/Strong
<i>CHAF1A</i>	chromatin assembly factor 1 subunit A	No	Yes/Strong
<i>GOT1</i>	lutamic-oxaloacetic transaminase 1	No	Yes/Strong
<i>PNP</i>	purine nucleoside phosphorylase	No	Yes/Strong
<i>EXOSC2</i>	exosome component 2	Yes	No
<i>EXOSC9</i>	exosome component 9	Yes	No
<i>EXOSC3</i>	exosome component 3	Yes	No
<i>PNKP</i>	polynucleotide kinase 3'-phosphatase	No	Yes
<i>RPL37A</i>	ribosomal protein L37a	Yes	No
<i>RPF1</i>	ribosome production factor 1 homolog	Yes	No

IDH1	isocitrate dehydrogenase (NADP(+)) 1, cytosolic	No	Yes
CHAMP1	chromosome alignment maintaining phosphoprotein 1	No	Yes

Figure (2): number of SNPs predicted as deleterious and tolerant using SIFT server

Figure (3): number of SNPs and their disease relationship predictions according PolyPhen-2 server

Figure (4): number and prediction of SNPs effect on SBDS protein using SNPs&Go and PHD-SNP softwares

Figure (5): numbers and percentages of SNPs and their effect on protein stability (disease causing mutations alter protein stability either by increasing or decreasing it) using I-Mutant 2.0 server.

Figure (6): 3D structure of SBDS protein showing the six disease causing mutations, wild types (Red) and mutant types (Yellow) using Chimera 1.13.1.

Figure (7): Multiple sequence alignment of the SBDS Protein family, shows mutations (A. rs28942099 (N8K), B. rs11557409 (R11H), C. rs11557408 (V17M), D. rs376960114 (F57S), E. rs367842164 (A66T), F. rs113993995 (R126T)) are located at areas of high conservation using Bioedit software.

Figure (8): GENEmania network showing genes co-expression patterns with *SBDS* gene, color codes indicates functions shared by genes.

Figure (9): GENEmania network showing genes physical interaction patterns with *SBDS* gene, color highlights indicates shared functions.

Discussion:

Six SNPs were found to be the most disease causing SNPs (rs11557408, rs11557409, rs28942099, rs113993995, rs367842164, rs376960114) ((V17M), (R11H), (N8K), (R126T), (A66T), (F57S)) indicating they have the ability to change the SBDS protein structure and functions; thus causing SDS.

The following SNPs with IDs (rs11557408, rs11557409, rs367842164, and rs376960114) and mutations of ((V17M), (R11H), (A66T), and (F57S), respectively) have been reported to be deleterious by SIFT server, and probably/ possibly damaging according to PolyPhen-2. Mutations of both (A66T) and (F57S) were predicted to be double disease causing by SNPS&GO and PHD-SNP and showed decreased protein stability, while (R11H) was disease causing by PHD-SNP, neutral by SNPs&Go and showed decreased protein stability, and (V17M) was neutral by both with elevated protein stability. However, the four SNPs were located on highly conservative region of the SBDS protein as previously shown in the multiple sequence alignment with SBDS protein family and were not reported by any previous study regarding this gene.

Mutations of Asparagine into Lysine at position 8 (N8K) (SNP ID rs28942099), Arginine into Threonine at position 126 (R126T) (SNP ID rs113993995), Isoleucine into Threonine at position 212 (I212T) (SNP ID rs79344818) were found by Boocock's et al 2002 study, these findings are in agreement with our findings, the mutations were predicted to be disease causing according to SIFT. (rs28942099, and 113993995, (N8K) and (R126T), respectively) were probably damaging by PolyPhen-2 servers, neutral according to SNPS&GO and disease causing according to PHD-SNP with an increased protein stability according to I-mutant 2.0, while (rs79344818) (I212T) was benign and had not further analyzed. Additionally, according to Mutpred2 prediction mutation of SNP (rs28942099) (N8k) led to structural changes including a gain of strand. Mutation of (R126T) (SNP ID rs113993995) was also reported by another study Finch et al. 2005 to be defective using protein expression and purification techniques which comes also in agreement to this study's findings.^{17,31}

Mutation at position 33 of the wild amino acid Lysine was reported in a study by Carvalho et al. 2014, using Array comparative genomic hybridization and Southern blotting detected mutation in a SDS patient, and study by Shammas et al. 2005 using western blotting and flowcytometry techniques.^{32,30} However, different mutant types occurred each time (Threonine (K33T), Glutamic acid (K33E), respectively). A similar mutation in the same position 33 and wild type Lysine converted into Arginine (K33R) was found in study by Boocock et al 2002 using southern hybridization and RT-PCR techniques in SDS affected individuals.¹⁷ This mutation is similar to the one obtained from NCBI database in this study (SNP ID rs373730800) (K33R) and had been predicted to be tolerated by SIFT server and had not been analyzed furthermore, which contradicts the previous studies results.

As previously mentioned in the introduction, *SBDS* gene plays an important role in ribosomal processes, most importantly, ribosomal biogenesis through (physical interaction with EFL1, and co-expression with EXOSC3, and EXOSC9), positive regulation of cell growth (co-expression with EXOSC2, EXOSC4, and EXOSC9), ribosomal assembly (through interaction with EFL1 only), and undergoing apoptotic and cellular clearance

processes through (co-expression with TPT1, Exosome components Two, three, nine and physical interaction with GOT1).^{49–52} According to GENeMania report, *SBDS* had strong physical interaction with nine genes, mainly (TYMS, CHAF1A, GOT1, and PNP). A lesser degree of physical interaction with both (RPP30, and CHAMP1). In addition, it had been co-expressed with eight genes (TPT1, TYMS, EXOSC2, EXOSC3, EXOSC5, EXOSC9, RPL37A, and RPF1).

This study encountered possible new functions of the *SBDS* gene, which is predicted to have new DNA related functions through GENeMania, as it has been both co-expressed and strongly interacting physically with (TYMS – a gene which plays an important role in pyrimidine metabolism and synthesis of dTTP and has polymorphisms associated with developing neoplasia-, CHAF1A –the chromatin assembly unit in DNA replication and repair-, and PKNP –responsible of DNA repair-)^{53–55}. It was also found that *SBDS* gene is co-expressed with RBF1 gene (a gene mainly expressed in the bone marrow which is responsible of ribosomal assembly, and RNA processes)⁵⁶. This suggests that, mutations in the *SBDS* gene could result in impaired expression of RBF1 gene resulting in patients developing variable levels of cytopenias. In addition to that, *SBDS* gene has a co-expression with EXOSC3 gene (which is partially responsible of muscle movement), impairment in expression of *SBDS* gene would affect the proper expression of EXOSC3 as well. Thus, this could also be used as an explanation of SDS patients developing skeletal abnormalities and failure to thrive.⁵⁷

The limitation of this study was using computational analysis only, we recommend wet lab analysis of the four novel mutations reported in this study, and clinical and laboratory analysis of the new predicted gene functions.

The most important findings in this study were four mutations predicted to be disease causing which have not been reported previously, (rs11557408 (V17M), rs11557409 (R11H), rs367842164 (A66T), and rs376960114 (F57S)) in addition, predicted functions of *SBDS* gene which are DNA processes related, RBF1 affected expression leading to cytopenias and EXOSC3 affected expression leading to failure to thrive.

Conclusion:

This study uses insilico tools which are important diagnostic tools, they are necessary to guide molecular geneticists, they balance between the time consumption, low cost required and are more consistent. Not only that, but insilico technology proved its efficiency in the field of therapeutics as well. For conventional drug, vaccine designing and cell therapy methods consumes prolonged time, man power, cost and raises toxicity and safety issues, however, computational tools have been developed to design immunotherapy as well as peptide-based drugs and efficient regenerative medicine tools and proved to be a catalyst in drug and vaccine design with no safety issues at all.

References:

1. (PDF) The ExPASy proteome WWW server in 2003. https://www.researchgate.net/publication/228493982_The_ExPASy_proteome_WWW_server_in_2003. Accessed December 20, 2018.
2. Warde-Farley D1, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q. The GeneMANIA prediction server: biological network integration for gene p. <http://www.sciepub.com/reference/164545>. Accessed December 20, 2018.
3. Boocock GRB, Morrison JA, Popovic M, et al. Mutations in SBDS are associated with Shwachman–Diamond syndrome. *Nat Genet.* 2003;33(1):97-101. doi:10.1038/ng1062
4. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum Mutat.* 2009;30(8):1237-1244. doi:10.1002/humu.21047
5. Calamita P, Miluzio A, Russo A, et al. SBDS-Deficient Cells Have an Altered Homeostatic Equilibrium due to Translational Inefficiency Which Explains their Reduced Fitness and Provides a Logical Framework for Intervention. Copenhaver GP, ed. *PLOS Genet.* 2017;13(1):e1006552. doi:10.1371/journal.pgen.1006552
6. Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr Protoc Hum Genet.* 2013;76(1):7.20.1-7.20.41. doi:10.1002/0471142905.hg0720s76
7. Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr Protoc Hum Genet.* 2013;76(1):7.20.1-7.20.41. doi:10.1002/0471142905.hg0720s76
8. Adzhubei I, Jordan DM, Sunyaev SR. Predicting Functional Effect of Human Missense Mutations Using PolyPhen-2. *Curr Protoc Hum Genet.* 2013;707:Unit7.20. doi:10.1002/0471142905.HG0720S76
9. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet.* 2013;Chapter 7:Unit7.20. doi:10.1002/0471142905.hg0720s76
10. Aggett PJ, Cavanagh NP, Matthew DJ, Pincott JR, Sutcliffe J, Harries JT. Shwachman’s syndrome. A review of 21 cases. *Arch Dis Child.* 1980;55(5):331-347. doi:10.1136/ADC.55.5.331
11. Artimo P, Jonnalagedda M, Arnold K, et al. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 2012;40(W1):W597-W603. doi:10.1093/nar/gks400
12. Ball HL, Zhang B, Riches JJ, et al. Shwachman-Bodian Diamond syndrome is a multi-functional protein implicated in cellular stress responses. *Hum Mol Genet.* 2009;18(19):3684-3695. doi:10.1093/hmg/ddp316

13. BODIAN M, SHELDON W, LIGHTWOOD R. Congenital Hypoplasia of the Exocrine Pancreas. *Acta Paediatr.* 1964;53(3):282-293. doi:10.1111/j.1651-2227.1964.tb07237.x
14. Goobie S, Popovic M, Morrison J, et al. Shwachman-Diamond syndrome with exocrine pancreatic dysfunction and bone marrow failure maps to the centromeric region of chromosome 7. *Am J Hum Genet.* 2001;68(4):1048-1054. doi:10.1086/319505
15. Rawls AS, Gregory AD, Woloszynek JR, Liu F, Link DC. Lentiviral-mediated RNAi inhibition of Sbds in murine hematopoietic progenitors impairs their hematopoietic potential. *Blood.* 2007;110(7):2414-2422. doi:10.1182/BLOOD-2006-03-007112
16. Ginzberg H, Shin J, Ellis L, et al. Segregation analysis in Shwachman-Diamond syndrome: evidence for recessive inheritance. *Am J Hum Genet.* 2000;66(4):1413-1416. doi:10.1086/302856
17. Popovic M, Goobie S, Morrison J, et al. Fine mapping of the locus for Shwachman-Diamond syndrome at 7q11, identification of shared disease haplotypes and exclusion of TPST1 as a candidate gene. *Eur J Hum Genet.* 2002;10(4):250-258. doi:10.1038/sj.ejhg.5200798
18. Mack D, Forstner G, Wilschanski M, Freedman M, Durie P. Shwachman syndrome: Exocrine pancreatic dysfunction and variable phenotypic expression. *Gastroenterology.* 1996;111(6):1593-1602. doi:10.1016/S0016-5085(96)70022-7
19. Mäkitie O, Ellis L, Durie P, et al. Skeletal phenotype in patients with Shwachman-Diamond syndrome and mutations in SBDS. *Clin Genet.* 2004;65(2):101-112. doi:10.1111/j.0009-9163.2004.00198.x
20. Nihrane A, Sezgin G, Dsilva S, et al. Depletion of the Shwachman-Diamond syndrome gene product, SBDS, leads to growth inhibition and increased expression of OPG and VEGF-A. *Blood Cells, Mol Dis.* 2009;42(1):85-91. doi:10.1016/J.BCMD.2008.09.004
21. Yamaguchi M, Fujimura K, Toga H, Khwaja A, Okamura N, Chopra R. Shwachman-Diamond Syndrome is not necessary for the terminal maturation of neutrophils but is important for maintaining viability of granulocyte precursors. *Exp Hematol.* 2007;35(4):579-586. doi:10.1016/J.EXPHEM.2006.12.010
22. <https://NCBI.nlm.nih.gov>. Homo sapiens SBDS, ribosome maturation factor (SBDS), RefSeqGene (LRG_ - Nucleotide - NCBI. <https://www.ncbi.nlm.nih.gov/nuccore/163965402>. Accessed November 1, 2018.
23. NIHRANE A, SEZGIN G, DSILVA S, et al. Depletion of the Shwachman-Diamond syndrome gene product, SBDS, leads to growth inhibition and increased expression of OPG and VEGF-A. *Blood Cells, Mol Dis.* 2009;42(1):85-91. doi:10.1016/j.bcmd.2008.09.004
24. Minelli A, Nicolis E, Cannioto Z, et al. Incidence of Shwachman-Diamond

- syndrome. *Pediatr Blood Cancer*. 2012;59(7):1334-1335. doi:10.1002/pbc.24260
25. de Oliveira JF, Sforça ML, Blumenschein TMA, et al. Structure, Dynamics, and RNA Interaction Analysis of the Human SBDS Protein. *J Mol Biol*. 2010;396(4):1053-1069. doi:10.1016/J.JMB.2009.12.039
26. Ganapathi KA, Austin KM, Lee C-S, et al. The human Shwachman-Diamond syndrome protein, SBDS, associates with ribosomal RNA. *Blood*. 2007;110(5):1458-1465. doi:10.1182/blood-2007-02-075184
27. Hesling C, Oliveira CC, Castilho BA, Zanchin NIT. The Shwachman–Bodian–Diamond syndrome associated protein interacts with HsNip7 and its down-regulation affects gene expression at the transcriptional and translational levels. *Exp Cell Res*. 2007;313(20):4180-4195. doi:10.1016/j.yexcr.2007.06.024
28. Grisendi S, Bernardi R, Rossi M, et al. Role of nucleophosmin in embryonic development and tumorigenesis. *Nature*. 2005;437(7055):147-153. doi:10.1038/nature03915
29. Orelia C, Verkuijlen P, Geissler J, van den Berg TK, Kuijpers TW. SBDS expression and localization at the mitotic spindle in human myeloid progenitors. *PLoS One*. 2009;4(9):e7084. doi:10.1371/journal.pone.0007084
30. Shammas C, Menne TF, Hilcenko C, et al. Structural and Mutational Analysis of the SBDS Protein Family. *J Biol Chem*. 2005;280(19):19221-19229. doi:10.1074/jbc.M414656200
31. Finch AJ, Hilcenko C, Basse N, et al. Uncoupling of GTP hydrolysis from eIF6 release on the ribosome causes Shwachman-Diamond syndrome. *Genes Dev*. 2011;25(9):917-929. doi:10.1101/gad.623011
32. Carvalho CMB, Zuccherato LW, Williams CL, et al. Structural variation and missense mutation in SBDS associated with Shwachman-Diamond syndrome. *BMC Med Genet*. 2014;15(1):64. doi:10.1186/1471-2350-15-64
33. Houdayer C, Dehainault C, Mattler C, et al. Evaluation of in silico splice tools for decision-making in molecular diagnosis. *Hum Mutat*. 2008;29(7):975-982. doi:10.1002/humu.20765
34. Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. 2003;31(13):3784-3788. <http://www.ncbi.nlm.nih.gov/pubmed/12824418>. Accessed December 20, 2018.
35. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003;31(13):3812-3814. <http://www.ncbi.nlm.nih.gov/pubmed/12824425>. Accessed December 20, 2018.
36. Sunyaev SR, Eisenhaber F, Rodchenkov I V, Eisenhaber B, Tumanyan VG, Kuznetsov EN. PSIC: profile extraction from sequence alignments with position-specific counts of independent observations. *Protein Eng*. 1999;12(5):387-394. <http://www.ncbi.nlm.nih.gov/pubmed/10360979>. Accessed December 20, 2018.

37. Capriotti E, Calabrese R, Casadio R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. *Bioinformatics*. 2006;22(22):2729-2734. doi:10.1093/bioinformatics/btl423
38. Hepp D, Gonçalves GL, de Freitas TRO. Prediction of the damage-associated non-synonymous single nucleotide polymorphisms in the human MC1R gene. *PLoS One*. 2015;10(3):e0121812. doi:10.1371/journal.pone.0121812
39. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*. 2001;29(1):308-311. <http://www.ncbi.nlm.nih.gov/pubmed/11125122>. Accessed December 21, 2018.
40. Mottaz A, David FPA, Veuthey A-L, Yip YL. Easy retrieval of single amino-acid polymorphisms and phenotype information using SwissVar. *Bioinformatics*. 2010;26(6):851-852. doi:10.1093/bioinformatics/btq028
41. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera—A visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25(13):1605-1612. doi:10.1002/JCC.20084
42. Goddard TD, Ferrin TE. Visualization software for molecular assemblies. *Curr Opin Struct Biol*. 2007;17(5):587-595. doi:10.1016/j.sbi.2007.06.008
43. Goddard TD, Huang CC, Ferrin TE. Software Extensions to UCSF Chimera for Interactive Visualization of Large Molecular Assemblies. *Structure*. 2005;13(3):473-482. doi:10.1016/j.str.2005.01.006
44. Kirmani S. A User Friendly Approach for Design and Economic Analysis of Standalone SPV System. *Smart Grid Renew Energy*. 2015;06(04):67-74. doi:10.4236/sgre.2015.64007
45. Capra JA, Singh M. Predicting functionally important residues from sequence conservation. *Bioinformatics*. 2007;23(15):1875-1882. doi:10.1093/bioinformatics/btm270
46. Schueler-Furman O, Baker D. Conserved residue clustering and protein structure prediction. *Proteins Struct Funct Genet*. 2003;52(2):225-235. doi:10.1002/prot.10365
47. Lichtarge O, Bourne HR, Cohen FE. An Evolutionary Trace Method Defines Binding Surfaces Common to Protein Families. *J Mol Biol*. 1996;257(2):342-358. doi:10.1006/jmbi.1996.0167
48. Franz M, Rodriguez H, Lopes C, et al. GeneMANIA update 2018. *Nucleic Acids Res*. 2018;46(W1):W60-W64. doi:10.1093/nar/gky311
49. Stepensky P, Chacón-Flores M, Kim KH, et al. Mutations in EFL1, an SBDS partner, are associated with infantile pancytopenia, exocrine pancreatic insufficiency and skeletal anomalies in a Shwachman-Diamond like syndrome. *J Med Genet*. 2017;54(8):558-566. doi:10.1136/jmedgenet-2016-104366

50. Mukherjee D, Gao M, O'Connor JP, et al. The mammalian exosome mediates the efficient degradation of mRNAs that contain AU-rich elements. *EMBO J.* 2002;21(1):165-174. doi:10.1093/emboj/21.1.165
51. Preker P, Nielsen J, Kammler S, et al. RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters. *Science* (80-). 2008;322(5909):1851-1854. doi:10.1126/science.1164096
52. Panteghini M. Aspartate aminotransferase isoenzymes. *Clin Biochem.* 1990;23(4):311-319. <http://www.ncbi.nlm.nih.gov/pubmed/2225456>. Accessed January 12, 2019.
53. Anderson DD, Quintero CM, Stover PJ. Identification of a de novo thymidylate biosynthesis pathway in mammalian mitochondria. *Proc Natl Acad Sci.* 2011;108(37):15163-15168. doi:10.1073/pnas.1103623108
54. Sarraf SA, Stancheva I. Methyl-CpG Binding Protein MBD1 Couples Histone H3 Methylation at Lysine 9 by SETDB1 to DNA Replication and Chromatin Assembly. *Mol Cell.* 2004;15(4):595-605. doi:10.1016/j.molcel.2004.06.043
55. Jilani A, Ramotar D, Slack C, et al. Molecular cloning of the human gene, PNKP, encoding a polynucleotide kinase 3'-phosphatase and evidence for its role in repair of DNA strand breaks caused by oxidative damage. *J Biol Chem.* 1999;274(34):24176-24186. <http://www.ncbi.nlm.nih.gov/pubmed/10446192>. Accessed January 12, 2019.
56. Wehner KA, Baserga SJ. The sigma(70)-like motif: a eukaryotic RNA binding domain unique to a superfamily of proteins required for ribosome biogenesis. *Mol Cell.* 2002;9(2):329-339. <http://www.ncbi.nlm.nih.gov/pubmed/11864606>. Accessed January 12, 2019.
57. Rudnik-Schöneborn S, Senderek J, Jen JC, et al. Pontocerebellar hypoplasia type 1: clinical spectrum and relevance of EXOSC3 mutations. *Neurology.* 2013;80(5):438-446. doi:10.1212/WNL.0b013e31827f0f66