Biological Sciences

Article

Computational Analysis of Functional Coding/Noncoding Single Nucleotide Polymorphisms (SNPs/Indels) in Human *NEUROG1* gene

Shimaa M. S. Mahalah[1]*, Zhoor A. A Hamid[2], Samah M Ibrahim[3], Shazalia k Babiker[4]

1 Faculty of Medicine, Department of Biochemistry, Sudan University of Science and Technology, Sudan
2 Faculty of Pharmacy, Al-Neelain University, Sudan
3 Faculty of Medical Laboratory, University of Medical Sciences and Technology, Sudan
4 Faculty of Medical Laboratory, University of Medical Sciences and technology, Sudan
* Corresponding author: Shimaa.mahalah@gmail.com

Article History
Received: 30 July, 2022
Accepted: 29 August, 2022
Published: 03 September, 2022

Citation
Mahalah, S.M.S., Hamid, Z.A.A., Ibrahim, S.M., & Babiker, S.K (2022) . Computational Analysis of Functional Coding/Noncoding Single Nucleotide Polymorphisms (SNPs/Indels) in Human *NEUROG1* gene. *Journal of Biological Sciences 1(1),* 1-9.
https://doi.org/10.56388/bio220903

Abstract
Human NEUROG1 gene encodes a very important protein Neurogenin1, which has been demonstrated to have an essential role as a transcription factor in the process of neurogenesis and neural repair. Mutations in *NEUROG1* gene have been linked to many congenital diseases and some neurodegenerative diseases. This study used bioinformatics tools to evaluate the effects of mutations along the sequence of the gene. The 3D structural modeling of *NEUROG1* was generated through I-TASSER and validated by different software. Genomic data has been retrieved from NCBI, GenBank, and Ensembl databases. Of the 617 SNPs and INDels reported in the dbSNP spanning coding and noncoding regions, only 4 of the 193 SNPs affecting coding regions were predicted to be deleterious with bioinformatics software. Analysis of 3' UTR (3′-untranslated regions) revealed that eight SNPs were found to affect microRNA binding sites either by creating or disturbing them, and other three INDels had no effect This study is a proposed computational analysis for the possible effects of Non- synonymous SNPs on the functionality of *NEUROG1* gene and subsequently its protein as an important cofactor in new neurons formation. The results suggested that, *NEUROG1* gene can serve as a candidate for gene therapy at genetics and epigenetics levels and/or drug design for neurodegenerative and other neurological diseases.

Keywords: Computational analysis, *NEUROG1*, Single Nucleotide Polymorphisms (SNPs)

Introduction

Human NEUROG1 gene encodes a very important protein Neurogenin1 and is located in chromosome 5 (q31.1) of the human genome from 135, 534,282 base pairs to 135, 535, 949 base pairs. The *NEUROG1* involved in transcription factor activity, sequence-specific DNA binding and protein dimerization activity. It is conserved in mice, chicken, zebra fish, and lizards. An important paralog of this gene is *NEUROG2* [1]. *NEUROG1* translates into Neurogenin1 protein**,** a member of the basic Helix Loop Helix (bHLH) proteins which is a group of neural transcription factors including, Neurogenin 2 (Ngn2), Mash1, and Math1. Each of these factors has a unique and transient expression pattern in neural progenitor cells during nervous system development. Neurogenin1 is expressed exclusively in the ventricular zone during development of the cerebral cortex [2-4]. It act as transcriptional regulator involving in the initiation of neuronal differentiation, activates transcription by binding to the Ebox (5- CANNTG-3), and associates with chromatin to enhance regulatory elements in genes encoding key transcriptional regulators of neurogenesis [1]. Many studies suggested that bHLH transcription factors have an essential role in embryonic development and human neurogenesis. Mutations in the *NEUROG1* gene are associated with various diseases including moebius syndrome and cerebellar liponeurocytoma [1,5]. Literature also states, mutation in *NEUROG1 has*

association with and increased susceptibility to schizophrenia [6]. In addition, *NEUROG1* gene has been regulated at the epigenetics level, Herbst A and his colleagues reported that Methylation of *NEUROG1* in serum is a sensitive marker for the detection of early colorectal cancer and this finding is so interesting [7]. Furthermore, in a recent study in 2021 done by Juliette Dupont , Ana Berta Sousa and their colleagues mentioned that their findings support the growing compelling evidence that loss of *NEUROG1* leads to a very distinctive disorder of cranial nerves development [8].

Single Nucleotide Polymorphisms (SNPs) are the single base change in coding or non-coding DNA sequence and are presented in every 200–300 bp in the human genome. It accounts for the more common form of human genetic variation. About 500,000 SNPs fall in the coding regions of the human genome, among these, the non-synonymous SNPs (nsSNPs) lead to changes in the amino acid residues, which are considered to be an important factor contributing to the functional diversity of the encoded proteins in the human population [9,10,11]. The nsSNPs affect gene regulation by altering DNA and transcriptional binding factors and the maintenance of the structural integrity of cells and tissues [12,13]. Also, they affect the functional roles of proteins in the signal transduction of visual, hormonal, and other stimulants [14,15].

Polymorphism in non coding region (3′-UTR, 3′-untranslated regions), such as SNPs in microRNAs binding sites, called mirSNPs, can affect miRNAs function and gene expression [16). MicroRNAs (miRNAs) are small 22 nt long non coding RNAs that usually act as posttranscriptional regulators by binding to the 3′-UTR) of mRNAs [17,18]. The miRNA targeting process depends on the binding of the miRNA to complementary target sites. This binding is affected by sequence polymorphism in either miRNA or their target site, which affect miRNA function and gene expression. Prior to initiating such a kind of research in life sciences, especially molecular genetics, it is crucial to use computational bioinformatics tools to determine which SNPs to be tested or screened in the human population. The *NEUROG1* gene was selected for study not only due to its important role in Neurogenesis as found in animal models (Mice), but also because it might be a suggested candidate gene for gene therapy at genetic or epigenetics level. The main objective of this study is to perform computational analysis of SNPs in the *NEUROG1* gene, including the 3′-UTR, to identify the possible mutations and to propose a modeled structure of the mutant protein.

2. Material and Method

2.1 Data set

The data were selected from the National Center for Biological Information (NCBI) SNPs database[1] for the retrieval of human *NEUROG1* gene SNPs in September 2017, and the FASTA format of the protein was obtained using UniProt[2]. The Obtained SNPs were analyzed using computational softwares as follows.

2.2 Sorting Intolerant from Tolerant (SIFT)

SIFT software (Sorting Tolerant From Intolerant)[3] is an online bioinformatics tool used to predict whether an amino acid substitution will affect the protein function or not [1]. The main underlying principle of this program is that it generates alignments with a large number of homologous sequences, and assigns scores to each residue ranging from zero to one. Scores close to zero indicate evolutionary conservation of the genes and intolerance to substitution, while those close to one indicate tolerance to substitution only [19,20].

2.3 Polyphen

The predicted damages were confirmed by using PolyPhen-2, which is another online software that predicts possible impact of an amino acid substitution on the structure and function of a human protein [21]. We entered the protein ID retrieved from UniProt, the substitution of amino acid in the deleterious nsSNPs and its position. The prediction is based on a number of features comprising the sequence, phylogenetic and structural information characterizing the substitution. Basically, this program searches for 3D protein structures, multiple alignments of homologous sequences and amino acid contact information in several protein structure databases, then calculates Position-Specific Independent Count Scores (PSIC) for each of two variants, and then computes the PSIC scores difference between two variants. The higher a PSIC score difference, the higher the functional impact a particular amino acid substitution is likely to have. PolyPhen scores were assigned, probably damaging (2.00 or more), possibly damaging (1.40–1.90), potentially damaging (1.0–1.50), benign (0.00–0.90). Basically PolyPhen accepts input in the form of SNPs or protein sequences[4] [22].

2.4 SNAP2

SNAP2 is a trained classifier that is based on a machine learning 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. The most important input signal for the prediction is the evolutionary information taken from an automatically generated multiple sequence alignment. Also structural features such as predicted secondary structure and solvent accessibility are considered. If available, annotation (i.e. known functional residues, pattern, regions) of the sequence or close homologs are also introduced in. It predicts a score, ranging from -100 strong neutral prediction to +100 strong effect predictions, which reflects the

[1] Source: http://www.ncbi.nlm.nih.gov/projects/SNP
[2] Source: http://www.uniprot.org/
[3] Source: http://blocks.fhcrc.org/sift/SIFT.htm
[4] Source: http://genetics.bwh.harvard.edu/pph2/

likelihood of a specific mutation to alter the native protein function. The analysis suggests that the prediction score is to some extent correlated to the severity of effect[5] [23].

2.5 I mutant

I mutant software was used for the nsSNPs which were found to be deleterious by SIFT and confirmed by PolyPhen to predict the stability of protein. It predicts protein stability changes upon single point mutation from protein sequence [24]. We entered the protein sequence, position of mutation and the new residue resulted from mutation[6].

2.6 SNPs & GO

It is a web server for the prediction of human disease-related single point protein mutations. It is an accurate method based on support vector machines, to predict disease related mutations from the protein sequence, scoring with accuracy=82% and Matthews correlation coefficient=0.63. It collects in a unique framework information derived from protein sequence, protein sequence profile, and protein function [25]. We entered UNIPROT accession number, mutation position, wild-type residue and substituting residue[7].

2.7 CPHmodels-3.0

It is a protein homology modeling prediction server, used to predict the 3D structure of proteins with an unknown 3D structure model[8]. Where the template recognition is based on profile-profile alignment, guided by secondary structure and exposure predictions [26]. Protein sequence requirements were submitted to the CPH server to get the model as a PDB file (for the structure that could not be predicted by the automated Project HOPE server). The resultant PDB files were opened using the Chimera program which was used to visualize the PDB structure.

2.8 Project HOPE

It is an easy-to-use web service that analyses the structural effects of a point mutation in a protein sequence. It collects structural information from a series of sources, including calculations on the 3D protein structure, sequence annotations in UniProt and prediction from the Reprof software. It combines this information to analyze the effect of a certain mutation on the protein structure. HOPE is an online web service where the user can submit a sequence and mutation[9] [27].

2.9 Chimera

UCSF Chimera is a highly extensible program for interactive visualization and analysis of molecular structures including, density maps, supramolecular assemblies, sequence alignments, docking results, trajectories, and conformational ensembles. This software was produced by University of California, San Francisco [28]. The PDB files obtained from the CPH model-3.0 were entered into chimera to get 3D structure of our protein[10].

2.10 I-TASSER software

I-TASSER (Iterative Threading ASSEmbly Refinement) is a hierarchical approach to protein structure and function prediction. First identify structural templates from the PDB by multiple threading approach LOMETS, then full length atomic models constructed by iterative template fragment assembly simulations. After that function insights of the target are derived by threading the 3D models through protein function database BioLiP. I-TASSER ('Zhang-Server') was ranked as number one server for protein structure prediction in recent community-wide experiments (in CASP7, CASP8, CASP9, CASP10, CASP11, and CASP12). It was also ranked as the best for function prediction in CASP9[11] [29].

2.11 PolymiRTS Database 3.0

It is an integrated platform for analyzing the functional impact of genetic polymorphisms in miRNA seed regions and miRNA target sites [16]. Single variants within 3′-UTR were selected from total variants and submitted to PolymiRTS server, to check if these variants could disrupt or create new miRNA binding sites or have no impact at all[12].

3. Result and discussion

NEUROG1 has an essential role in embryonic development and human neurogenesis, so it is rational to conduct an in-silico analysis using computational and bioinformatics tools before commence genetic or genomics research, in order to establish or generate a hypothesis that can be tested further, and of the concept and approaches were followed in this study. Data and SNPs of human *NEUROG1* gene were obtained from National Center of Biotechnology (NCBI) and other databases

[5] Source: https://rostlab.org/services/snap2web/
[6] Source: http://folding.biofold.org/i-mutant/i-mutant2.0.html
[7] Source: https://snps-and-go.biocomp.unibo.it/snps-and-go/
[8] Source: http://www.cbs.dtu.dk/services/CPHmodels/
[9] Source: http://www.cmbi.ru.nl/hope/
[10] Source: http://www.cgl.ucsf.edu/chimera/index.html
[11] Source: https://zhanglab.ccmb.med.umich.edu/I-TASSER/
[12] Source: http://compbio.uthsc.edu/miRSNP/

(dbSNP, UniProt, and GenBank). SNPs and inDels were categorized in coding and non-coding regions. As shown in **Table 1**, there is no reported SNP in the intron region.

Table 1. Distributions of SNPs and Indels.

Region	Coding				Non Coding					Total
Type of Variant	**Synonymous**	**Missense**	**Nonsense and stop gain**	**Frame shift**	**Near 3′-UTR**	**3 UTR**	**5'UTR**	**Near 5′-UTR**	**5' Splice**	
Numbers of SNPs	57	126	6	-	34	65	256	40	1	**585**
Numbers of Indels	-	4	-	2	12	7	4	3	-	**32**
Total	57	130	6	2	46	72	260	43	1	**617**

Non-Synonymous SNPs (nsSNPs) in the coding region include Missense, nonsense, stop gain and frameshift mutations. In total, 138 were submitted to SIFT, PolyPhen, and SNAP2 prediction tools, and 4 SNPs were predicted as damaging and having effect on structure and function of the protein (**Table 2).**

Table 2. Functional SNPs predicted by SIFT and POLYPHEN and SNAP2 prediction tools.

SNP	AMINO ACID CHANGE	FUNCTION	SIFT SCORE	SIFT PREDICTION	POLYPEN SCORE	PPLYPHEN PREDICTION	SNAP2	SNAP2 SCORE
rs150467271	T27I	MISSENSE	0.019	DELETERIOUS	0.995	PROBABLY DAMAGING	EFFECT	40
rs200689943	A183E	MISSENSE	0.036	DELETERIOUS	0.808	POSSIBLY DAMAGING	EFFECT	45
rs368232160	A203D	MISSENSE	0.031	DELETERIOUS	0.535	POSSIBLY DAMAGING	EFFECT	68
rs373314804	R90S	MISSENSE	0.006	DELETERIOUS	0.995	PROBABLY DAMAGING	EFFECT	64

Stability of the protein was measured by the I Mutant server for the 4 deleterious SNPS. Result is found as shown in **Table 3**. Stability increased in one SNP and decreased in three SNPs. SNPs&GO server's results show that one SNP (rs373314804) at position (90) was predicted to cause disease and the other three were neutral as in **Table 4**.

Table 3. Predicted results of I-Mutant software.

SNP	AMINO ACID CHANGE	PROTEIN ID	FUNCTION	I MUTANT STABILITY	iMUTANT RI
rs150467271	T27I	ENSP00000317580	MISSENSE	INCREASE	4
rs200689943	A183E	ENSP00000317580	MISSENSE	DECREASE	8
rs368232160	A203D	ENSP00000317580	MISSENSE	DECREASE	6
rs373314804	R90S	ENSP00000317580	MISSENSE	DECREASE	8

Table 4. Predicted results of SNP and GO server.

SNP Id	AMINO ACID CHANGE	SNPS&GO PREDICTION	SNPS&GO RI	PROBABILITY
rs150467271	T27I	NEUTRAL	2	0.388
rs200689943	A183E	NEUTRAL	6	0.213
s368232160	A203D	NEUTRAL	3	0.329
rs373314804	R90S	DISEASE	2	0.593

From this appears the importance of correlations between genotype and phenotype presentation of the most lethal mutations in this gene, as has been reported in moebius syndrome. To visualize the effect of these damaging SNPs, CPH modeler server has been used because there is no PDB 'ID' for neurogenin1 protein in the protein database, then Chimera software has been used to visualize the 3D structure of the query PDB 'ID' as illustrated in in **Figure 1.**

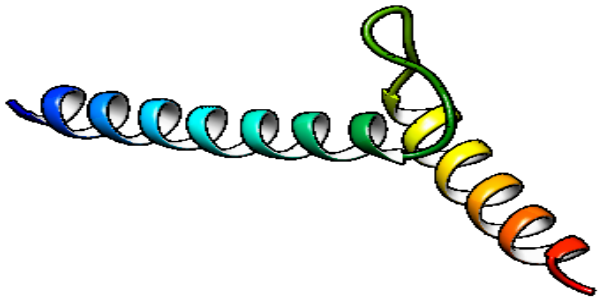

Figure 1. 3D structure visualized by Chimera software.

Human Neurogenin1 protein is 237 amino acids in length, and the predicted structure is composed of two long _ helices (helix loop helix domain) starting from position 92-149 connected by a short loop, which is 11 residues long. This region is conserved and is similar to other members of the bHLH superfamily, which have two highly conserved and functionally distinct domains that together constitute a region of approximately 60 amino-acid residues with two terminals. The amino terminus is the basic domain that binds transcription factors to DNA at a consensus hexanucleotide sequence known as E box. The carboxy terminus has HLH domain that facilitates interactions with other protein subunits to form homo- and hetero-dimeric complexes, since dimerization with another bHLH protein complexes is required to form efficient DNA binding.

Predicted deleterious SNPs didn't affect the conserved region of the protein, but it affected structure, function and stability of the protein. I-TASSER has been used, and I-TASSER gives 5 different structures with their C-score (confidence score). C-score is typically in the range of [-5, -2], where a C-score of higher value signifies a model with a high confidence and vice-versa. The model we chose had a C-score of -4.33. The I-TASSER model predicted the full length of the protein in 3D structure (**Figure 2) and** functional sites where protein binds to DNA (**Figure 3)**.

Figure 2. 3D structure predicted by I-TASSER server.

Project hope has been used to get more information about potentially destructive SNPs. rs150467271 Threonine into Isoleucine at position 27 (T27I): mutant residue is large in size and might lead to bumps. This occurs because mutations introduce more hydrophobic residue at this position, which result in loss of hydrogen bonds and/or disturb correct folding. rs373314804, Arginine into a Serine at position 90(R90S): There is a difference in charge between the wild-type and mutant amino acid. The charge of the wild-type residue will be lost, this can cause loss of interactions with other molecules or residues. The mutant residue is smaller, this might lead to loss of interactions. The mutation introduces more hydrophobic residue at this position. which result in loss of hydrogen bonds and/or disturb correct folding.

rs200689943, Alanine into Glutamic Acid at position 183(A183E): the mutation introduces a charge; this can cause repulsion of ligands or other residues with the same charge. The mutant residue is big and might lead to bumps, because 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. rs368232160 Alanine into Aspartic Acid at position 203 (A203D): The mutation introduces a charge, this can cause repulsion of ligands or other residues with the same charge. The mutant residues are larger, which might lead to bumps. Hydrophobic interactions at the protein core or surface will disappear. More research to be done on these SNPs to verify if it exists or is associated with other diseases in the brain.

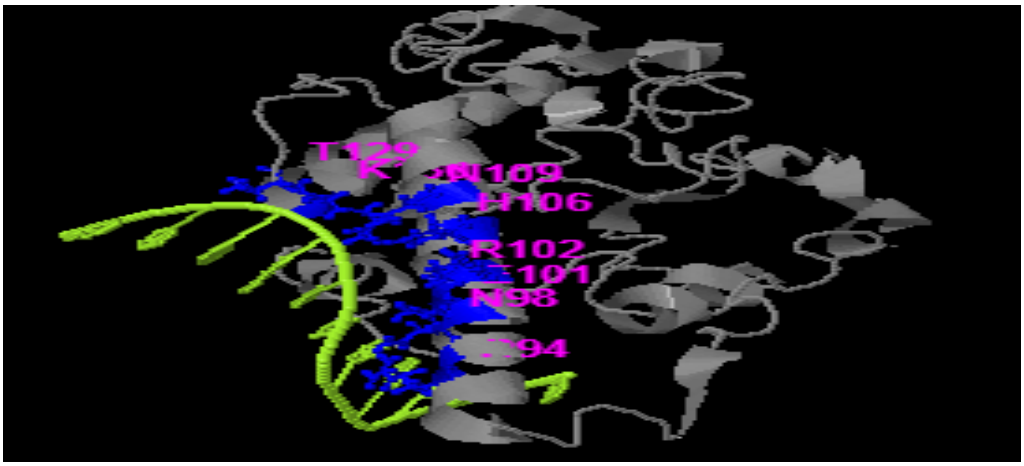

Figure 3. biological annotations of the target protein by COFACTOR and COACH based on the I-TASSER structure prediction showing Ligand Binding Site Residues: R94, N98, E101, R102, H106, N109, T129 and K130, bound to DNA.

As mentioned in one paper about methylation of this gene has been detected in colorectal cancer, we suggested this might be a new door to think about the involvement of brain - gut pathway in the pathogenesis of these diseases. PolymiRTS Database 3.0 has been used to analyze the 3′-UTR SNP. And it has to mention that, **Table 5** shows occurrence of the miRNA site in other vertebrate genomes in addition to the query genome.

From observations and literature related, it seems that sSNPs in MicroRNAs sites have been neglected recently and more studies are needed to link them with specific diseases. As shown in **Table 5,** there are eight SNPs that affect microRNAs binding sites by disturbing or creating new binding sites. Three Indels were observed with no change - a very interesting finding. These microRNAs in the *NEUROG1* gene were associated with other gene expression, which ultimately leads to a change in their expression level that might cause or be associated with other illnesses.

Table 5. Predicted results of functional 3′-UTR SNPs/Indels, using PolymiRTs server.

dbSNP ID	Variant type	Ancestral Allele	Allele	miR ID	Conservation	miRSite	Function Class	context+ score change
1-rs34991655	INDEL	-	T	hsa-miR-203a	17	cttCATTTCAtatg	O	-0.156
				hsa-miR-3646	16	cTTCATTTcatatg	O	-0.019
2-rs192260351	SNP	A	A	hsa-miR-3925-3p	5	ggCTGGAGAgtac	D	-0.125
				hsa-miR-6512-3p	7	GGCTGGAgagtac	D	-0.207
				hsa-miR-6720-5p	7	GGCTGGAgagtac	D	-0.198
				hsa-miR-6849-3p	7	GGCTGGAgagtac	D	-0.167
				hsa-miR-766-3p	5	gGCTGGAGAgtac	D	-0.369
			G	hsa-miR-2355-5p	5	ggCTGGGGAgtac	C	-0.187
3-rs34861604	INDEL	-	-	hsa-miR-483-3p	15	acaAGGAGTGggc	O	-0.18
			T	hsa-miR-3200-3p	14	aCAAGGTAgtgggc	O	-0.105
				hsa-miR-4301	17	acaaggTAGTGGGc	O	-0.179
				hsa-miR-4999-3p	15	acaaGGTAGTGggc	O	-0.186
				hsa-miR-6813-3p	14	aCAAGGTAgtgggc	O	-0.096
4-rs182771224	SNP	A	G	hsa-miR-4315	18	agaAAAGCGAgca	C	-0.14
				hsa-miR-4790-5p	18	agaAAAGCGAgca	C	-0.27
				hsa-miR-627-3p	21	AGAAAAGcgagca	C	-0.055

Note:
sequence context of the miRNA site: #bases complementary to the seed region are in capital letters and SNPs are highlighted in bold font.
Function class:

D:the derived allele **disrupts** a conserved miRNA site (ancestral allele with support>2);
C: the derived allele **creates** a new miRNA site;
O: the ancestral allele cannot be determined.

Context score: negative.
increase =increase in SNP functionality.

Table 5. Continued.

dbSNP ID	Variant type	Ancestral Allele	Allele	miR ID	Conservation	miRSite	Function Class	context+ score change
5-rs201015495	SNP	G	G	hsa-miR-204-5p	13	agaAAAGGGAcag	D	-0.127
				hsa-miR-211-5p	13	agaAAAGGGAcag	D	-0.127
				hsa-miR-627-3p	12	AGAAAAGggacag	D	-0.046
				hsa-miR-6832-3p	13	agAAAAGGGAcag	D	-0.166
			T	hsa-miR-1250-3p	13	aGAAAATGgacag	C	-0.027
				hsa-miR-3912-5p	14	agaaaATGGACAg	C	-0.153
6-rs185239448	SNP	C	C	hsa-miR-1226-5p	13	ccgGCCCTCAgcc	D	-0.104
				hsa-miR-3616-3p	13	ccgGCCCTCAgcc	D	-0.121
				hsa-miR-4721	13	ccgGCCCTCAgcc	D	-0.131
				hsa-miR-7160-5p	15	ccggcCCTCAGCc	D	-0.061
			T	hsa-miR-605-3p	13	ccgGCCTTCAgcc	C	-0.077
				hsa-miR-6126	13	ccgGCCTTCAgcc	C	-0.087
7-rs35758773	INDEL	-	G	hsa-miR-6840-5p	6	gtCGGGGGAtacct	O	-0.256
8-rs113601422	SNP	T	T	hsa-miR-1207-3p	6	cCCAGCTGccgcc	D	-0.079
				hsa-miR-6820-5p	2	cccagCTGCCGCc	D	-0.268
			A	hsa-miR-3692-5p	6	cCCAGCAGccgcc	C	-0.041
				hsa-miR-4731-5p	9	CCCAGCAgccgcc	C	-0.187
				hsa-miR-5589-5p	9	CCCAGCAgccgcc	C	-0.1
				hsa-miR-6762-3p	2	cccAGCAGCCgcc	C	-0.078
9-rs11960442	SNP	C	C	hsa-miR-149-3p	4	ctttccCCCTCCC	D	-0.19
				hsa-miR-3153	2	CTTTCCCcctccc	D	-0.149
				hsa-miR-4716-3p	4	cttTCCCCCTccc	D	-0.296
				hsa-miR-4728-5p	4	ctttccCCCTCCC	D	-0.199
				hsa-miR-625-5p	5	ctTTCCCCCtccc	D	-0.221
				hsa-miR-6733-5p	2	CTTTCCCcctccc	D	-0.159
				hsa-miR-6739-5p	2	CTTTCCCcctccc	D	-0.149
				hsa-miR-6785-5p	4	ctttccCCCTCCC	D	-0.199
				hsa-miR-6794-5p	4	cttTCCCCCTccc	D	-0.268
				hsa-miR-6883-5p	4	ctttccCCCTCCC	D	-0.218
			T	hsa-miR-7106-5p	4	ctttccTCCTCCC	C	-0.176
				hsa-miR-765	4	ctttcCTCCTCCc	C	-0.202
				hsa-miR-766-5p	5	ctTTCCTCCtccc	C	-0.18
10-rs200961598	SNP	C	C	hsa-miR-3153	2	taCTTTCCCcctc	D	-0.149
				hsa-miR-4716-3p	4	tacttTCCCCCTc	D	-0.296
				hsa-miR-625-5p	5	tactTTCCCCCtc	D	-0.221
				hsa-miR-6733-5p	2	taCTTTCCCcctc	D	-0.159
				hsa-miR-6739-5p	2	taCTTTCCCcctc	D	-0.149
				hsa-miR-6794-5p	4	tacttTCCCCCTc	D	-0.268
			T	hsa-miR-548n	4	TACTTTTcccctc	C	0.007
				hsa-miR-6124	2	taCTTTTCCcctc	C	-0.07
				hsa-miR-6783-5p	4	tacttTTCCCCTc	C	-0.199
11-rs199710381	SNP	A	A	hsa-miR-1283	2	cTTTGTAGAcact	D	No Change
				hsa-miR-3618	2	cttTGTAGACAct	D	No Change
				hsa-miR-934	4	ctttGTAGACAct	D	No Change

Given the importance of the *NEUROG1* gene in New neuron formation in previous research in animal models (Mice), it might be a candidate gene for gene therapy at genetic or epigenetic levels for neurodevelopmental and neurodegenerative diseases. The functions of these SNPs/Indels require further experimental verification.

4. Conclusion

Computational analysis using different bioinformatics tools is necessary before genomic and life science research. In this study, four damaging SNPs in human *NEUROG1* were predicted that did not affect DNA binding sites in its BLHL region, but affect structure, function and stability of the protein, thereby interfering with its role in Neurogenesis. *NEUROG1* has no database structure, so the 3D structure was made using I-TASSER software. Analysis of the 3′-UTR revealed that eight SNPs were found to affect microRNA binding sites by creating or disturbing them, and three INDels were not observed to do so. The *NEUROG1* gene can be used as a candidate of gene therapy at genetics and epigenetics levels and/or drug design for neurodegenerative and other neurological diseases.

5. Recommendation

As previous studies have revealed the crucial role of *NEUROG1* gene in new neuron formation, more research in neurodegenerative and neurodevelopmental diseases should be conducted to detect the deleterious SNPs identified in this study to provide more evidence and information on microRNA expression and regulation, and to explore the possibility of gene therapy at both genetics and epigenetics levels. X-ray crystallography and NMR must be performed for human Neurogenin1 protein structure in order to visualize the mutations and their effect on the protein, and to aid drug design and other treatment modalities in the future.

Author Contributions:. Research design Mahalah, S.M.S, data preparation Mahalah, S.M.S, methodology Mahalah, S.M.S, Hamid, Z.A.A., Ibrahim, S.M., & Babiker, S.K; ,data processing, Mahalah, S.M.S. Hamid, Z.A.A., Ibrahim, S.M., & Babiker, S.K, software, Mahalah, S.M.S Hamid, Z.A.A., Ibrahim, S.M., & Babiker, S.K.; validation, Mahalah, S.M.S, Hamid, Z.A.A., Ibrahim, S.M., writing—l draft preparation, Mahalah, S.M.S, Hamid, Z.A.A., Ibrahim, S.M.,writing—review and editing, Mahalah, S.M.S,; visualization, Mahalah, S.M.S, Hamid, Z.A.A; supervision, Mahalah, S.M.S. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data is public and available for download at the National Center for Biological Information (NCBI) SNPs database (Source: http://www.ncbi.nlm.nih.gov/projects/SNP).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. GeneCards: The Human Gene Database. http://www.genecards.org/.
2. Lee JE. Basic helix-loop-helix genes in neural development. *Current Opinion in Neurobiology*. 1997 Feb 28;7(1):13-20.
3. Bertrand N, Castro DS, Guillemot F. Proneural genes and the specification of neural cell types. *Nature Reviews Neuroscience*. 2002 Jul 1;3(7):517-530.
4. Sun Y, Nadal-Vicens M, Misono S, Lin MZ, Zubiaga A, Hua X, Fan G, Greenberg ME (2001) Neurogenin promotes neurogenesis and inhibits glial differentiation by independent mechanisms. *Cell*. *2001*,104:365–376.
5. Schröder JC, Läßig AK, Galetzka D, Peters A, Castle JC, Diederich S, Zechner U, Müller-Forell W, Keilmann A, Bartsch O. A boy with homozygous microdeletion of NEUROG1 presents with a congenital cranial dysinnervation disorder [Moebius syndrome variant]. *Behavioral and Brain Functions*. 2013 Feb 18;9(1):7.
6. Ho BC, Epping E, Wang K, Andreasen NC, Librant A, Wassink TH. Basic helix–loop–helix transcription factor NEUROG1 and schizophrenia: Effects on illness susceptibility, MRI brain morphometry and cognitive abilities. *Schizophrenia Research*. 2008 Dec 31;106(2):192-199.
7. Herbst A, Rahmig K, Stieber P, Philipp A, Jung A, Ofner A, Crispin A, Neumann J, Lamerz R, Kolligs FT. Methylation of NEUROG1 in serum is a sensitive marker for the detection of early colorectal cancer. *The American Journal of Gastroenterology*. 2011 Jun 1;106(6):1110-1118.
8. Dupont J, Vieira JP, Tavares ALT, Conceição CR, Khan S, Bertoli-Avella AM, Sousa AB. Adding evidence to the role of NEUROG1 in congenital cranial dysinnervation disorders. *Clin Genet*. 2021 Apr;99(4):588-593. doi: 10.1111/cge.13922. Epub 2021 Jan 20. PMID: 33439489.
9. Lee JE, Choi JH, Lee JH, Lee MG. Gene SNPs and mutations in clinical genetic testing: haplotype-based testing and analysis. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 2005 Jun 3;573(1):195-204.
10. Collins FS, Brooks LD, Chakravarti A. A DNA polymorphism discovery resource for research on human genetic variation. *Genome Research*. 1998 Dec 1;8(12):1229-1231.
11. Lander ES. The new genomics: global views of biology. *Science*. 1996 Oct 25;274(5287):536.
12. Barroso I, Gurnell M, Crowley VE, Agostini M, Schwabe JW, Soos MA, Maslen GL, Williams TD, Lewis H, Schafer AJ, Chatterjee VK. Dominant negative mutations in human PPARγ associated with severe insulin resistance, diabetes mellitus and hypertension. *Nature*. 1999 Dec 23;402(6764):880-883.
13. Thomas R, McConnell R, Whittacker J, Kirkpatrick P, Bradley J, Sandford R. Identification of mutations in the repeated part of the autosomal dominant polycystic kidney disease type 1 gene, PKD1, by long-range PCR. *The American Journal of Human Genetics*. 1999 Jul 31;65(1):39-49.
14. Dryja TP, McGee TL, Hahn LB, Cowley GS, Olsson JE, Reichel E, Sandberg MA, Berson EL. Mutations within the rhodopsin gene in patients with autosomal dominant retinitis pigmentosa. *New England Journal of Medicine*. 1990 Nov 8;323(19):1302-1307.
15. Smith EP, Boyd J, Frank GR, Takahashi H, Cohen RM, Specker B, Williams TC, Lubahn DB, Korach KS. Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. *New England Journal of Medicine*. 1994 Oct 20;331(16):1056-1061.
16. Bhattacharya A, Ziebarth JD, Cui Y. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Research*. 2013 Oct 24;42(D1):D86-91.

17. Ambros V. The functions of animal microRNAs. *Nature*. 2004 Sep 16;431(7006):350-355.
18. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004 Jan 23;116(2):281-297.
19. PC. Ng, S. Henikoff. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*. 2003 Jul;31(13):3812-3814. DOI= http://doi.org/10.1093/ nar/gkg509.
20. Kumar P., Henikoff S., Pauline C. Ng.Predicting the effects of coding nonsynonymous variants on protein function using the SIFT algorithm. *Nature Protocols*. 2009 June 4;1073-1081. DOI= http:// doi.org/10.1038/nprot.2009.86.
21. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nature Methods*. 2010 Apr 1;7(4):248-249.
22. González-Pérez A, López-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *The American Journal of Human Genetics*. 2011 Apr 8;88(4):440-449.
23. Hecht M, Bromberg Y, Rost B. News from the protein mutability landscape. *Journal of Molecular Biology*. 2013 Nov 1;425(21):3937-3948.
24. Capriotti E, Fariselli P, Casadio R. I-Mutant2. 0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Research*. 2005 Jul 1;33(suppl_2):W306-10.
25. Calabrese R, Capriotti E, Fariselli P, Martelli PL, Casadio R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Human Mutation*. 2009 Aug 1;30(8):1237-1244.
26. Nielsen M, Lundegaard C, Lund O, Petersen TN. CPHmodels-3.0—remote homology modeling using structure-guided sequence profiles. *Nucleic Acids Research*. 2010 Jun 11;38(suppl_2):W576-581.
27. Venselaar H, te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics*. 2010 Nov 8;11(1):548.
28. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*. 2004 Oct 1;25(13):1605-1612.
29. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nature Methods*. 2015 Jan 1;12(1):7-8.