INNSPUB

International Journal of Biosciences | IJB |
ISSN: 2220-6655 (Print), 2222-5234 (Online)
http://www.innspub.net
Vol. 10, No. 4, p. 41-48, 2017

RESEARCH PAPER OPEN ACCESS

A computational report on the variants of *ZSCAN4* gene in treating down syndrome

Pratichi Singh, Rashi Nagotra, Arthi Venkatesan, J. Febin Prabhu Dass*

Department of Integrative Biology, School of Biosciences and Technology, VIT University, Vellore, Tamil Nadu, India

Key words: Down syndrome, Mutation, SNP, *ZSCAN4* gene.

http://dx.doi.org/10.12692/ijb/10.4.41-48 Article published on April 14, 2017

Abstract

Down syndrome (DS), also known as trisomy 21 is the most commonly known aneuploidy condition. It causes severe problems in human growth, function and its development. Recent reports suggest Zinc finger and SCAN domain containing 4 gene *(ZSCAN4)* as a new therapy for chromosome abnormalities, hence helps in treating the down syndrome. The expression of *ZSCAN4* gene located on chromosome 19 increases the telomere length in human adult cells. This study deals with Insilico approach to find the deleterious SNPs in *ZSCAN4* that are linked with this disease condition. Single Nucleotide Polymorphism (SNPs) in *ZSCAN4*is retrieved to predict the harmful effect in protein using computational tools like SNAP2, PolyPhen 2,I Mutant 2 and SIFT. As a result, two common SNPs are found to be highly deleterious with rs-id377104601 (R151I) and rs-id545052223 (I154T). Further, the structural analysis was performed and the result shows no similarity between the native and mutant protein. Therefore, these reported mutations(R151I and I154T) may alter the function and expression of *ZSCAN4* gene and may perhaps not be helpful in treating Down syndrome.

* **Corresponding Author:** J. Febin Prabhu Dass ✉ jfebinprabhudass@vit.ac.in

Introduction

Down syndrome (DS) is caused by an inaccuracy in chromosome number i.e aneuploidy. Aneuploidy is a condition in which either one chromosome is added or deleted from a normal diploid number(Torres *et al.,* 2005; Epstein, 1986).Down syndrome occurs due to presence of one extra chromosome at chromosome 21, also known as trisomy 21, represented by (47,XX,+21) in female and (47,XY,+21) in males (Akinci, 2009). This syndrome was first explained by 'John Langdon Down' in 1866.

The prevalence rate differs among the population, it may be between 1 in 319 and 1in 1000 live births (Frances *et al.,* 2009). According to WHO, 3,000 to 5,000 children are born with DS and about 25,000 families have affected once in USA (Morris and alberman, 2009). The researchers tried to find out the incidence of DS in England during the year 1998 to 2008 by collecting the data from The National Down Syndrome Cytogenetic Register1 which has the data of more than 26,000 families in England and leads to only 1% decrease in DS within these years (Megarbane *et al.,* 2009). The affected live births in the year 1998 were 752 and in the year 2008 to 2009 was 743, and also about 90% of the women decided to terminate their pregnancy due to DS (Megarbane *et al.,* 2009).

DS is the most common disease, the reason behind the fact is it does not produce severe effects as our body can bear an extra genetic material than the deficit (Liu *et al.,* 2015). The main factor considered for this disease is maternal age (DSR) and the reason observed for the occurrence of disorder is non-disjunction, means the failure of chromosome to disjoin properly during meiosis I and meiosis II at chromosome 21. Due to this disease several growth and developmental disorder has been seen, such as, Alzheimer's disease, cognitive impairment, hematopoietic disorders, premature ageing, mental retardation, and congenital heart defects (Tecklin *et al.,* 2015; Geppino *et al.,* 2007). Children affected from DS have a high risk of pulmonary disease (Zalzman *et al.,* 2010) and also suffer from muscle weakness and adolescent (Amano *et al.,* 2013).

Previously, it was considered that cure for this disease is not possible. But the discovery of a new gene may change the perspective (Hung *et al.,* 2013). *ZSCAN4*(Zinc finger and SCAN domain containing 4)gene, which produces *ZSCAN4* protein considered to be helpful in treating DS. It is a mammal gene and is found in the mouse preimplantation embryo i.e. at 2-celled stage and well thought-out in the development of the blastocyst. It is expressed in embryonic stem cell (Teer *et al.,* 2012). *ZSCAN4* is located on chromosome number 19q13.43. It also plays a vital role in reprogramming the somatic cells as induced pluripotent stem cells (iPSCs). The study suggests that *ZSCAN4* rapidly reactivates embryonic genes during the production of induced pluripotent stem cell which promoted scientists to forcefully express this gene in somatic cells to generate iPSCs (Robinson *et al.,* 2011). Expression of this gene causes an increase in telomere length in human adult cells (human fibroblast cell of a patient suffering from Fanconianemia)(Li *et al.,* 2009). It is also stated that *ZSCAN4*may also be an epigenetic regulator or a chromatin remodeling factor (Robinson *et al.,* 2011). Furthermore, it is explained that this gene combines with Yamanaka factors (Oct3/4, Sox2, Klf4, c-Myc). These factors when transferred to the donor cell causes DNA damage response (DDR) and lower down DDR as a result genomic stability is maintained during the process of somatic reprogramming (Kormann *et al.,* 2011).

The Scientist tried to sort out the problem of the DS with *XIST* gene by silencing of the X chromosome of female performing genetic engineering (Yonemitsu *et al.,* 2013). But due to iPSCs property of *ZSCAN4*, the researcher found it more relevant for the treatment of Down syndrome. The reason behind it is *ZSCAN4* act as biologic and does not lead to any genetic alteration (Seki *et al.,* 2013). It has been reported that forced expression of *ZSCAN4* causes an increase in telomere length in *in-vitro* (Thittgen *et al.,* 2012). It has also observed that after treating this novel gene by using genetic engineering tools such as a Vector Sendai virus in human fibroblast cells of affected DS, there is a major increase in normal cells in the culture (Kiefer *et al.,* 2009).

This result triggers the researchers to initiate the work on *ZSCAN4* which can be helpful in treating the chromosomal abnormalities in future generation (Laskowski *et al.,* 1996).

Therefore, in this study by knowing the function and effect of *ZSCAN4* on the certain irremediable disorder, we tried to find out the mutation which may alter the function of *ZSCAN4* protein. We did polymorphism in a single nucleotide to find out the deleterious effect of the protein by using various computational tools.

Materials and methods

The schematic representation of methodology used for this work is illustrated in Figure 1.

Data mining

The SNP data were retrieved from 1000 Genome project database on 7 April 2016. Our search was narrowed down toHomo sapiens, coding non-synonymous, introns, coding synonymous,5'UTR (Un-Translated Region) and 3' UTR.

Sequence retrieval

For collecting the sequences of protein, Uniprot database (www.uniprot.org) was used. Uniprot database is a universal resource of protein where all the protein sequence data are available. The protein sequence is downloaded in FASTA format for Homo sapiens. The *ZSCAN4*Uniprot id is Q8NAM6, having 433 amino acid sequences.

Prediction of deleterious SNP

All the non-synonymous SNPs (ns-SNPs) of ZSCAN4 were subjected to various computational tools like SNAP2, PolyPhen 2, I Mutant 2, and SIFT Blink to find its deleterious effect.

SNAP2

SNAP2 (www.rostlab.org/services/snap) is an online tool which gives us the result based upon their secondary structure. It shows the deviation of the mutated and native structure depends on their solvent accessibility and find out the effect as deleterious (+100,strongly predicted) and neutral (-100, strongly predicted) (Laskowski *et al.,* 2001).

Protein sequences in FASTA format were given as input to predict the result. As a result, It gives a map with neutral and possible effect, indicating normal and highly deleterious SNPs respectively.

PolyPhen 2

The Polymorphism Phenotyping v2 (www.genetics.bwh.harvard.edu/pph2/) finds the damaging effect by an iterative algorithm. This is structure and format based online tool. Suitable inquiries are given as input such as the FASTA format of sequences and the substitution of native amino acid and mutated amino acids (Porollo and Meller, 2007). It gives the output as score, specificity, sensitivity, and calculate the PSIC (position-specific independent count) and clarify the result as probably or possibly damaging, or benign (Porollo and Meller, 2007).

I Mutant 2

I Mutant2.0 is a sustain vector machine-based web server (www.folding.uib.es/cgi-bin/i-mutant2.0.cgi) which performs the automatic guess of protein stability changes, upon single-site mutations (Adamczak *et al.,* 2004). As input the FASTA sequences are given with deviations in residues and as output protein stability is obtained. Results are obtained in the form of energy which is calculated as DDG in kcal/Mol.

SIFT blink

SIFT is an online server (www.sift.jcvi.org/SIFTBLink_submit.html). It provides the result on the basis of homology and physical status of the protein. The input required is the rs-id of the protein, and the output is served as tolerated or damaging. The mutation can be judged on the basis of the score as Tolerant or deleterious when the probability score is ≥0.05 or<0.05 respectively (Rajamanikandan *et al.,* 2012).

Structure modelling

The structure of *ZSCAN4* protein was modelled, as the native structure of all the residues was not available in the PDB database (protein data bank) only structure of 90 amino acids was present which does not fit in the criteria to perform further analysis.

Therefore, we need to model the structure for 433 amino acid using ITasser, an online server (www.zhanglab.ccmb.med.umich.edu/ITASSER/).
This server provides us the three dimensional structure of the protein(Cheng *et al.,* 2005). All important information such as the sequence of native protein and mutated protein was provided as input to the software. The output gives the best models along with their TM score and RMSD(Root Mean Square Deviation) values.

Trajectory analysis

SRide is used to discover the stabilizing residues of native and mutated protein structures. It is an online database (www.sride.enzim.hu) provides stabilizing residues on the basis of hydrophobicity, long-range interactions (LHO),and conservation of amino acid residues. The solvent accessibility and secondary structure of the molecules are gathered from ITasser itself. It tells us the change and confirmations of the structure of both native and mutated ones. The solvent accessibility is predicted based on0-9 values, where 0 is fully buried and 9 are fully exposed.

Result and discussion

Non synonymous SNPs analysis has been developed as a new diagnostic method for diseases in the current year (Smith *et al.,* 2012).

Table 1. Deleterious SNPs report in ns-SNPs of ZSCAN4 gene.

rs-id	Allele	AA Pos	Wild AA	Mut AA	PolyPhen2	I Mut 2	SIFT	SNAP 2
rs11668570	G/A	387	E	K	1	-0.55	0.18	94
rs116138022	G/C	284	E	D	0.0053	-0.33	0.43	26
rs200305142	G/A	227	G	S	0.029	-1	0.54	-40
rs201878902	A/G	415	M	K	0.9	-1.19	0.009	14
rs377104601	G/T	151	R	I	0.7	-0.01	0.021	20
rs528877759	G/A	30	G	R	0.003	-0.91	0.06	1
rs530730923	C/A	169	A	E	0	0.3	1	17
rs534675684	C/A	228	P	H	0.025	-1.28	0.045	-32
rs536885944	A/C	301	H	P	0.006	0.29	0.06	77
rs544189666	A/G	425	S	G	0	-2.16	0.37	-66
rs545052223	T/C	154	I	T	0.997	-2.39	0.03	63
rs553085243	C/T	329	A	V	0.493	-0.15	0.84	-73
rs557480976	C/A	296	S	F	0.008	0.39	0.02	26

Deleterious variants by all the four tools (highlighted in bold letters).

As mentioned above, on 7thApril 2016,a total number of 885 SNPS were noted in 1000 genome project for the ZSCAN4 gene in Homo-sapiens. Then again, it is filtered for consideration of non-synonymous (missense) which is 13, introns 764, coding synonymous 17, 5'UTR 40 and 3' UTR is 6 represented in Figure 2.

Prediction of possibly deleterious SNPs

SNP analysis was performed usingSNAP2, PolyPhen 2, IMUTANT 2 and SIFT Blink. According to recent studies for considering the better results, the output of all the software should be combined and analysed (Singh and Dass, 2016). Diagnosis of output was done in which only 13 SNPs was found to be deleterious out of 885 SNPs, which was further depicted and analyzed on the basis of the most deleterious among 13 SNPs.

Table 2. Total energy and stabilizing residues of mutant models of ZSCAN4.

Residues	Total Energy (KJ/Mol)	Stabilizing residues	No. of Stabilizing residues
Native	-12515.26	LEU89, GLU90, GLN91, PHE92, PRO129, VAL133, HIS134, PHE350, LYS365, SER378, ARG406	11
R151I	-5410.330	LEU89, GLU90, GLN91, PHE92, PRO129, VAL133, HIS134, VAL153, VAL155, PHE350, LYS365, SER378, ARG406	13
I154T	-13215.78	LEU89, GLU90, GLN91, PHE92, PRO129, PRO130, VAL133, PHE350, LYS365, SER378, ARG406	11

Table 3. Solvent accessibility and secondary structural changes in ZSCAN4.

Residues	Native	Mutant	Native	Mutant
R151I	C	C	5	4
I154T	C	C	2	3

C- Coil, Solvent accessibility value ranges from 0-9 scale (0 is completely buried, and 9 is fully exposed).

All the tools predict two common mutations withrs-id377104601, and rs-id545052223 as the most harmful mutation among the 13 ns-SNPs. The mutation occurs due to the change in a single nucleotide i.e at position151 R to I, and at the position 154 from I to T respectively. The values are depicted in the Table 1 showing the level of mutations. Figure 3 shows the polymorphism occurred in native and mutant structures.

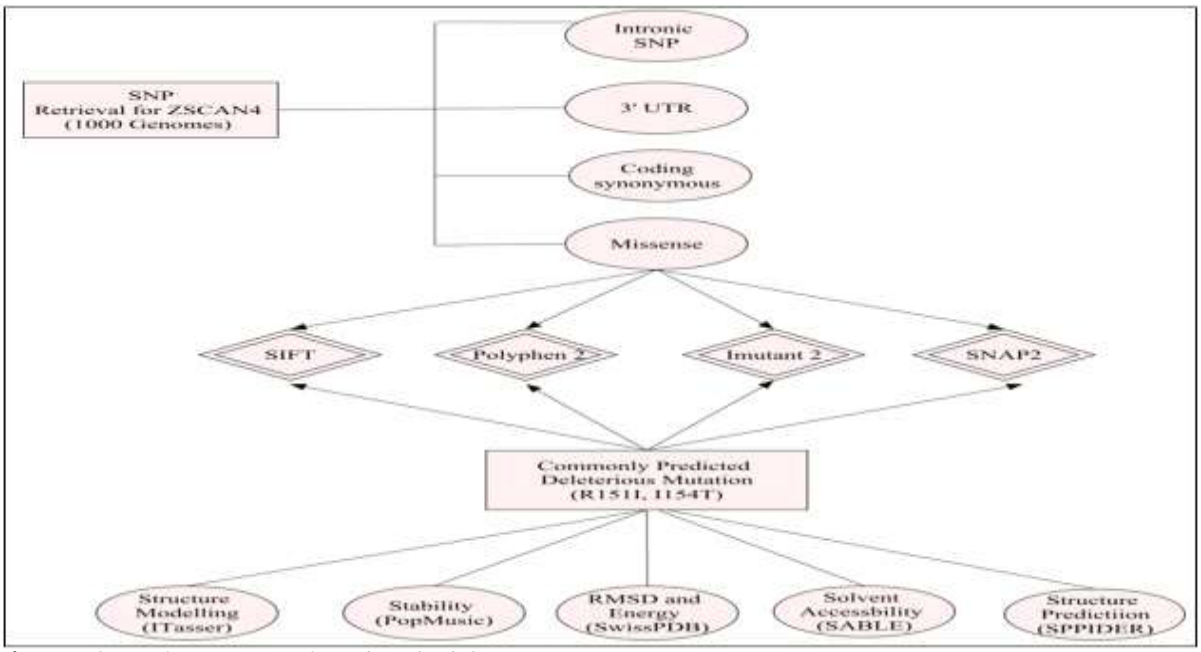

Fig. 1. Schematic representation of methodology

Structure modeling and its trajectory analysis

The 3D structure was modelled using ITasser server, and the best model was chosen based on their TM score and C- score. ITasser also gives us the information about the conformational change in mutated structure from native structure and its solvent accessibility.

As a result, no variation is found in secondary structure of native and mutated protein structures but variation in solvent accessibility is observed as shown in Table 2.

Furthermore, total energy value was also noted down shown in Table 3.

The stabilizing residues are predicted for the native and mutant structures from SRide database. SRide gives us the accurate idea of stabilizing and destabilizing residues of protein structures. All the stabilizing residues of native and mutant structure are shown in Table 3.

Result shows the native structure has 11 stabilizing residues, whereas the mutant structures of R151I has eight and I154T has three stabilizing residues respectively, which is less than the native protein. It confirms that native structure is more stabilized than the mutant structures.

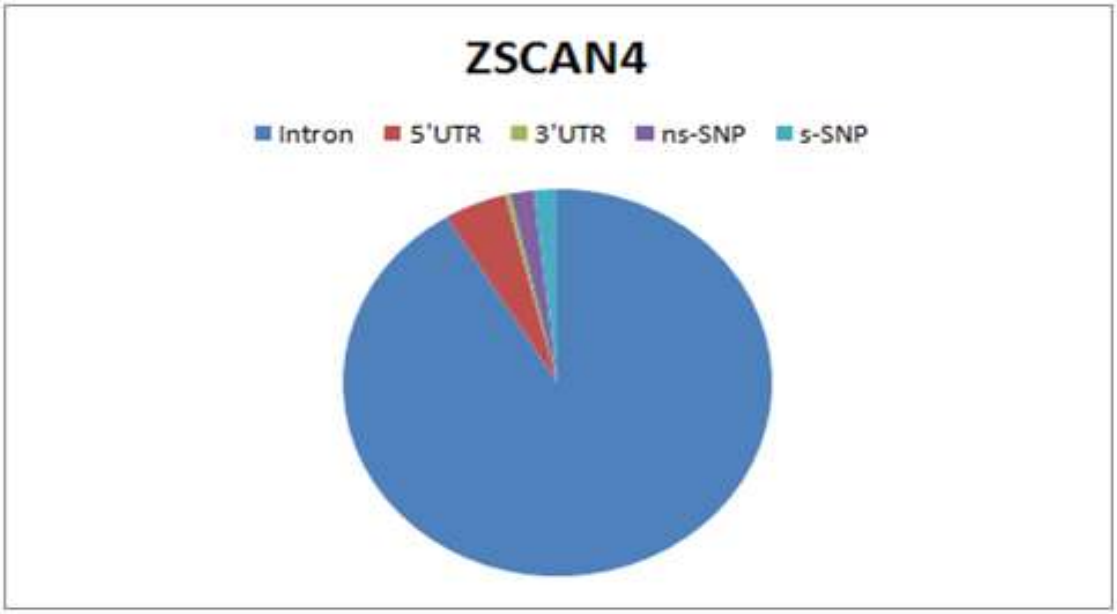

Fig. 2. Percentage distribution of regional SNPs in ZSCAN4.

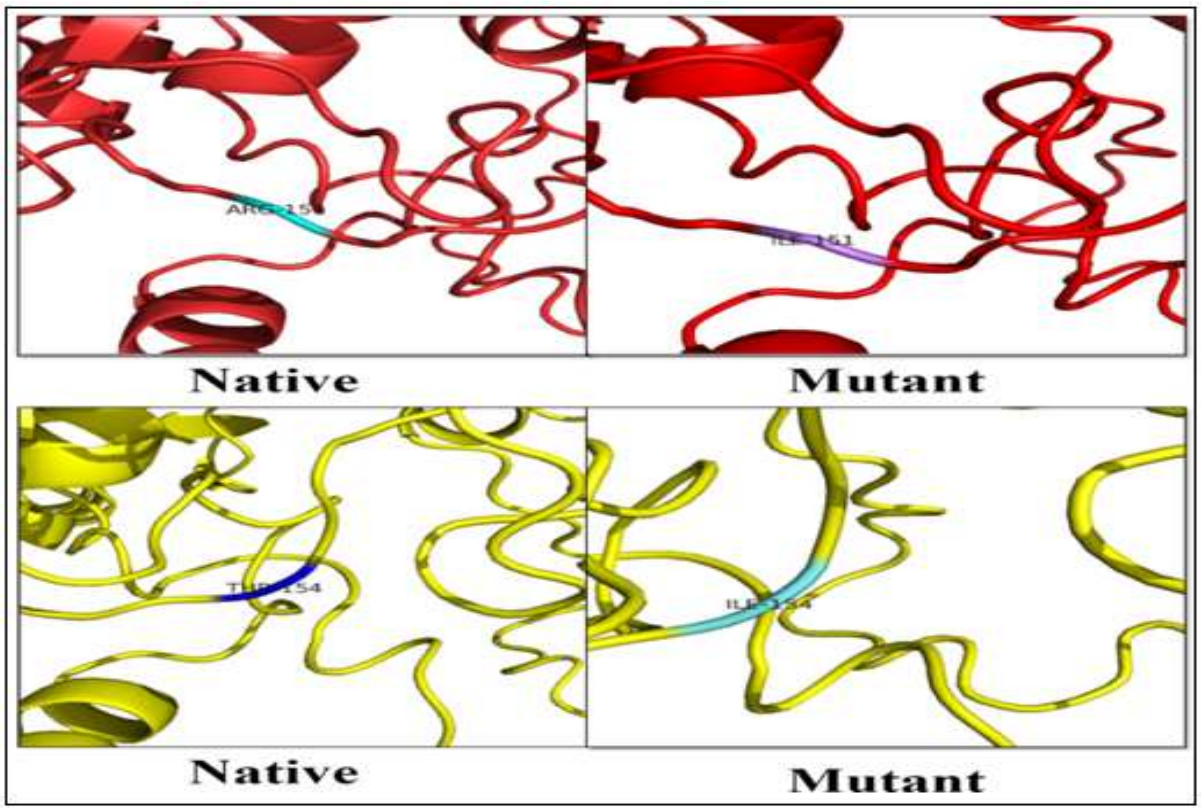

Fig. 3. Structures of native and mutant protein.

Conclusion

It is concluded that using multiple computational tools, out of 13 ns-SNPs, two SNPs rs377104601 (R151I) and rs545052223 (I154T) are the most deleterious SNPs. These mutations may be of high concern in ZSCAN4 gene and its associated diseases. These highly deleterious mutations reported by one or the other way may cause disruption in treating Down syndrome.

References

Adamczak R, Porollo A, Meller J. 2004. Accurate prediction of solvent accessibility using neural networks-based regression. Proteins **56(4),** 753-767.
https://doi.org/10.1002/prot.20176

Akinci A, Oner O, Bozkurt OH, Guven A, Degerliyurt A, Munir K. 2009. Refractive Errors and Strabismus in Children With Down Syndrome: A Controlled Study. Journal of Pediatric Ophthalmology and Strabismus **46(2),** 83-86.

Amano T, Hirata T, Falco G, Monti M, Sharova LV, Amano M, Ko MSH. 2013. Zscan4 restores the developmental potency of embryonic stem cells. Nature Communications **4**, 1966.
https://doi.org/10.1038/ncomms2966

Cheng J, Randall AZ, Sweredoski MJ,Baldi P. 2005. SCRATCH: a protein structure and structural feature prediction server. Nucleic Acids Research, **33**, 72-76.
https://doi.org/10.1093/nar/gki396

Evans HJ. 1987. The consequences of chromosome imbalance: Principles, mechanisms and models. Trends in Genetics **3,** 28-29.
https://doi.org/10.1016/0168-9525(87)90160-0

Falco G, Lee SL, Stanghellini I, Bassey UC, Hamatani T, Ko MSH. 2007. Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. Developmental Biology **307(2)**, 539-550.
https://doi.org/10.1016/j.ydbio.2007.05.003

Flanagan T, Russo N, Flores H, Burack J. 2008. The developmental approach to the study of Down syndrome: Contemporary issues in historical perspective.
https://doi.org/10.3104/reviews/2081

Hung SSC, Wong RCB, Sharov AA, Nakatake Y, Yu H, Ko MSH. 2013. Repression of global protein synthesis by Eif1a-like genes that are expressed specifically in the two-cell embryos and the transient Zscan4-positive state of embryonic stem cells. DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes, **20(4),** 391-402.
https://doi.org/10.1093/dnares/dst018

Kiefer F, Arnold K, Künzli M, Bordoli L, Schwede T. 2009.The SWISS-MODEL Repository and associated resources. Nucleic Acids Research **37,** 387-392.
https://doi.org/10.1093/nar/gkn750

Kormann MSD, Hasenpusch G, Aneja MK, Nica G, Flemmer AW, Herber-Jonat S, Rudolph C. 2011. Expression of therapeutic proteins after delivery of chemically modified mRNA in mice. Nature Biotechnology **29(2),** 154-157.
https://doi.org/10.1038/nbt.1733

Laskowski RA, Rullmannn JA, MacArthur MW, Kaptein R, Thornton JM. 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. Journal of Biomolecular NMR **8(4),** 477-486.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N. 2009.1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAM tools. Bioinformatics (Oxford, England) **25(16),** 2078-2079.
https://doi.org/10.1093/bioinformatics/btp352

Liu B, Filippi S, Roy A, Roberts I. (2015). Stem and progenitor cell dysfunction in human trisomies. EMBO Reports **16(1),** 44-62.
https://doi.org/10.15252/embr.201439583

Mégarbané A, Ravel A, Mircher C, Sturtz F, Grattau Y, Rethoré MO, Mobley WC. 2009. The 50th anniversary of the discovery of trisomy 21: the past, present, and future of research and treatment of Down syndrome. Genetics in Medicine: Official Journal of the American College of Medical Genetics **11(9),** 611-616.
https://doi.org/10.1097/GIM.0b013e3181b2e34c

Morris JK, Alberman E. 2009. Trends in Down's syndrome live births and antenatal diagnoses in England and Wales from 1989 to 2008: analysis of data from the National Down Syndrome Cytogenetic Register. BMJ (Clinical Research Ed.) **339**, 3794.

Porollo A, Meller J. 2007. Versatile annotation and publication quality visualization of protein complexes using POLYVIEW-3D.BMC Bioinformatics **8,** 316.
https://doi.org/10.1186/1471-2105-8-316

Rajamanika S, Vanajothi R, Sudha A, Rameshthan P, Srinivasan P. 2012. In silico Analysis of Deleterious SNPs of the FGFR2 Gene. Journal of Biological Science 12(2), 83-90.
https://doi.org/10.3923/jbs.2012.83.90

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative Genomics Viewer. Nature Biotechnology **29(1),** 24-26.
https://doi.org/10.1038/nbt.1754

Seki S, Matano T. 2013. Development of Vaccines Using SeV Vectors Against AIDS and Other Infectious Diseases. In Y. Nagai (Ed.), Sendai Virus Vector, 127-149. Springer Japan. Retrieved from http://link.springer.com/chapter/10.1007/978-4431-54556-9_5

Singh P, Dass JFP. 2016. A multifaceted computational report on the variants effect on KIR2DL3 and IFNL3 candidate gene in HCV clearance. Molecular Biology Reports **43(10),** 1101-1117.
https://doi.org/10.1007/s11033-016-4044-5

Smith N, Witham S, Sarkar S, Zhang J, Li L, Li C, Alexov E. 2012. DelPhi web server v2: incorporating atomic-style geometrical figures into the computational protocol. Bioinformatics (Oxford, England) **28(12),** 1655-1657.
https://doi.org/10.1093/bioinformatics/bts200

Song J, Tan H, Takemoto K, Akutsu T. 2008. HSEpred: predict half-sphere exposure from protein sequences. Bioinformatics **24(13),** 1489-1497.
https://doi.org/10.1093/bioinformatics/btn222

Teer J K, Green ED, Mullikin JC, Biesecker LG. 2012. VarSifter: visualizing and analyzingexome-scale sequence variation data on a desktop computer. Bioinformatics (Oxford, England) **28(4)**, 599-600.
https://doi.org/10.1093/bioinformatics/btr711

Thiltgen G, Goldstein RA. 2012. Assessing Predictors of Changes in Protein Stability upon Mutation Using Self-Consistency. PLOS ONE **7(10)**, e46084.
https://doi.org/10.1371/journal.pone.0046084

Torres EM, Williams BR, Amon A. 2008. Aneuploidy: cells losing their balance. Genetics **179(2),** 737-746.
https://doi.org/10.1534/genetics.108.090878

Wiseman FK, Alford KA, Tybulewicz VLJ, Fisher EMC. 2009. Down syndrome--recent progress and future prospects. Human Molecular Genetics **18**, 75-83.
https://doi.org/10.1093/hmg/ddp010

Yonemitsu Y, Matsumoto T, Itoh H, Okazaki J, Uchiyama M, Yoshida K, Maehara Y. 2013. DVC1-0101 to treat peripheral arterial disease: a Phase I/IIa open-label dose-escalation clinical trial. Molecular Therapy: The Journal of the American Society of Gene Therapy **21(3)**, 707-714.
https://doi.org/10.1038/mt.2012.279

Zalzman M, Falco G, Sharova LV, Nishiyama A, Thomas M, Lee SL, Ko MSH. 2010. Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature **464(7290)**, 858-863.
https://doi.org/10.1038/nature08882