***In silico* approach to identify non-synonymous SNPs with highest predicted deleterious effect on protein function in human obesity related gene, neuronal growth regulator 1 (NEGR1)**

Permendra Kumar¹ · Kulandaivelu Mahalingam¹

Received: 28 July 2017 / Accepted: 24 September 2018 / Published online: 1 November 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

Neuronal growth regulator 1 (NEGR1) is a candidate gene for human obesity, which encodes the neural cell adhesion and growth molecule. The aim of the current study was to recognize the non-synonymous SNPs (nsSNPs) with the highest predicted deleterious effect on protein function of the *NEGR1* gene. We have used five computational tools, namely, PolyPhen, SIFT, PROVEAN, MutPred and M-CAP, to predict the deleterious and pathogenic nsSNPs of the *NEGR1* gene. Homology modeling approach was used to model the native and mutant *NEGR1* protein models. Furthermore, structural validation was performed by the PROCHECK server to interpret the stability of the predicted models. We have predicted four potential deleterious nsSNPs, i.e., rs145524630 (Ala70Thr), rs267598710 (Pro168Leu), rs373419972 (Arg239Cys) and rs375352213 (Leu158Phe), which might be involved in causing obesity phenotypes. The predicted mutant models showed higher root mean square deviation and free energy values under the PyMoL and SWISS-PDB viewer, respectively. Additionally, the FTSite server predicted one nsSNP, i.e., rs145524630 (Ala70Thr) out of four identified nsSNPs found in the *NEGR1* protein-binding site. There were four potential deleterious and pathogenic nsSNPs, i.e., rs145524630, rs267598710, rs373419972 and rs375352213, identified from the above-mentioned tools. In future, further functional in vitro and in vivo analysis could lead to better knowledge about these nsSNPs on the influence of the *NEGR1* gene in causing human obesity. Hence, the present computational examination suggest that predicated nsSNPs may feasibly be a drug target and play an important role in contributing to human obesity.

Keywords Obesity · nsSNPs · *NEGR1* · Computational tools · Homology modeling

Introduction

Obesity is a significant problem occurring in different parts of the world. Many young children and adolescents are affected by this epidemic in some or the other way, such as certain cardiovascular disorders, skeletal problems, diabetes mellitus and many more (Guh et al. 2009). Obesity can be clinically measured by the measure of

body mass index (BMI), which can be calculated as the weight of an individual divided by the square of height. Individuals who have BMI ≥ 25 kg/m² are considered overweight, while individuals with BMI ≥ 30 kg/m² are categorized as obese (Frayling et al. 2007). The hypothalamus is involved in many neuronal and hormonal signals of satiety and nutritional status, which takes control of nutrients, regulates their function and maintains glucose homeostasis and peripheral lipid metabolism (Nogueiras et al. 2007). If the hypothalamus does not function properly, obesity develops in some individuals (Williams 2012). Type 2 diabetes (T2D) is one of the primary consequence of obesity, which is characterized by increased secretion of insulin by the pancreatic beta cells leading to hyperglycemia and increased blood pressure (Nolan et al. 2011). Many genes are reported to cause obesity in humans. Till now, 127 genetic loci have been identified in the human genome for their association with obesity

In memory of Prof. K. Mahalingam

✉ Kulandaivelu Mahalingam
kmahalingam@vit.ac.in
Permendra Kumar
permendrak@gmail.com

¹ Department of Bio-Medical Sciences, School of Biosciences and Technology, Vellore Institute of Technology, Vellore 632014, Tamilnadu, India

phenotypes through genome-wide association studies (GWAS) (Alonso et al. 2016). The *NEGR1* is a newly recognized raft-associated protein, which has recently been featured as a new locus related to human obesity (Kim et al. 2017). Niemann–Pick disease type C2 (NPC2) is a fatal human neurodegenerative disease which stimulates intracellular cholesterol trafficking; in this case, *NEGR1* interacts with NPC2 and increases its protein stability. Thereby, *NEGR1* plays a role in intracellular cholesterol homeostasis and is possibly linked with human obesity (Kim et al. 2017). The *NEGR1* gene is also known as kilon or neurotractin and is the most common variant associated with BMI (Thorleifsson et al. 2009; Willer et al. 2009; Speliotes et al. 2010). The *NEGR1* protein is encoded by the *NEGR1* gene. The *NEGR1* is a 354 amino acid *Ig*LON protein, which belongs to the immunoglobulin superfamily of cell adhesion molecules (Miyata et al. 2003). This gene is located on chromosome 1p31.1, has nine exons and functions as a trans-neural growth-promoting factor in regenerative exon sprouting in the mammalian brain. It also plays an important role in cell to cell adhesion and neurite outgrowth in hypothalamic neurons and physical activity, which is associated with obesity (Lee et al. 2012, 2015). It is located at the post-synaptic sites of dendritic and somatic synapses and highly expressed in the cerebral cortex, hippocampus, and cerebellum during postnatal development (Schäfer et al. 2005). It has also been shown that *NEGR1* is involved in neurite outgrowth and also determines synapse number in the hippocampus during neuronal development (Pischedda et al. 2014; Hashimoto et al. 2008; Marg et al. 1999). There are no biochemical functional studies reported on human *NEGR1* so far. However GWAS analysis reported that *NEGR1* was associated with human body weight regulation and human dyslexia (Veerappa et al. 2013; Willer et al. 2009).

Single nucleotide polymorphism (SNP) plays a vital role in identifying genetic variants and a potential biomarker for investigating the deleterious and neutral effects on protein function associated with obesity (Lyon and Hirschhorn 2005). The study of the SNPs on the living system is complicated. Since single nucleotide change might control the protein function by altering the transcription activity, the efficiency of the splicing mechanism, m-RNA expression and in the protein coding region, it causes missense or nonsense mutations (Vazquez-Martinez and Gasman 2014; Garg and Wood 2013; Komar 2007). A recent study also supports the genome-wide prioritization of sequence variants with the most deleterious effect on protein function which was performed in nine vertebrate species including humans (Rozman and Kunej 2018). Computational studies are cost-effective, less time consuming and more reliable as compared to experimental processes (Singh and Mahalingam 2017; Singh et al. 2016). Thus, the present study

aims to recognize and explore the deleterious nsSNPs of the *NEGR1* gene, which is associated with human obesity, computationally.

Materials and methods

Collection of nsSNPs

The information about nsSNPs of the *NEGR1* gene was collected from dbSNP (<http://www.ncbi.nlm.nih.gov/snp/>) and GeneCard (<http://www.genecards.org/cgi-bin/snp>) databases. The information about the missense mutation of the human *NEGR1* gene was obtained from the Human Gene Mutation Database (HGMD) (<http://www.hgmd.cf.ac.uk/>).

Prediction of functional impact and deleterious nature of the collected nsSNPs

There are various computational methods and tools widely available to predict deleterious nsSNPs. In this study, we used five computational tools, i.e., PolyPhen, SIFT, PROVEAN, MutPred and M-CAP, to predict the structural, functional and evolutionary information about the *NEGR1* protein (*Homo sapiens*; UniProt ID Q7Z3B1). The PolyPhen and PROVEAN use structural and functional features of the protein, whereas SIFT and MutPred use evolutionary sequence relationship to identify the tolerance or intolerance (Singh and Mahalingam 2017; Singh et al. 2016; Kumar et al. 2009, 2015; Goswami 2015; Choi et al. 2012; Adzhubei et al. 2010; Li et al. 2009; Ng and Henikoff 2003).

PolyPhen 2.0 (Polymorphism Phenotyping 2.0)

The PolyPhen 2.0 (<http://genetics.bwh.harvard.edu/pph2/>) tool predicts the effect of amino acid variations on the structure and function of a human protein using straightforward physical and comparative considerations (Adzhubei et al. 2010). This type of prediction is based on features including sequence, phylogenetic and structural information characterizing the substitution. We used the *NEGR1* protein sequence and specific amino acid variants as an input option. This tool classifies the nsSNPs as benign, probably damaging and possibly damaging, based on the position-specific independent count (PSIC) score (Adzhubei et al. 2010).

Sorting intolerant from tolerant (SIFT)

The SIFT (<http://sift.jcvi.org/>) tool predicts whether an amino acid substitution affects the protein function or not. It is a sequence homology-based tool which premises that protein evolution is correlated with protein function (Kumar et al. 2009). We have submitted all nsSNPs of *NEGR1* gene

in this tool as an input option to find the deleterious nsSNPs. This tool shows the impact as deleterious if SIFT score is < 0.05 and tolerated if SIFT score is ≥ 0.05 (Ng and Henikoff 2003).

Protein variation effect analyzer (PROVEAN)

PROVEAN (<http://provean.jcvi.org/index.php>) is a freely available tool which predicts the functional effect of the protein upon amino acid substitution (Choi et al. 2012). This tool gives a PROVEAN score for all types of human proteins and variants. If the PROVEAN score is < -2.5 , then it depicts a deleterious protein variant, while for neutral protein variants the score is > -2.5 . We have used batch protein sequence variants as an input option for this tool.

MutPred

MutPred (<http://mutpred.mutdb.org/>) is also a Web-based freely available tool which categorizes an amino acid substitution as a disease associated or neutral in humans. This tool predicts the molecular cause of disease, based on amino acid substitution (Li et al. 2009). It predicts the functional and structural characteristics of a protein upon amino acid modification. It uses SIFT, PSI-BLAST and Pfam profiles for functional attributes, whereas TMHMM, MARCOIL and DisProt are used for structural properties (Punta et al. 2012; Kumar et al. 2009; Sickmeier et al. 2007; Delorenzi and Speed 2002; Krogh et al. 2001; Altschul et al. 1997). The functional and structural characteristic predictions make this tool more efficient and accurate in finding the deleterious nsSNPs. This tool gives the top five features and classifies three hypotheses based on certain combinations of high values of general or probability of deleterious mutation scores (g) and low values of property scores (p) reported as follows: (1) scores with $g > 0.5$ and $p < 0.05$ are referred to as actionable hypotheses; (2) scores with $g > 0.75$ and $p < 0.05$ are referred to as confident hypotheses; (3) scores with $g > 0.75$ and $p < 0.01$ are referred to as very confident hypotheses (Li et al. 2009). We have used *NEGR1* protein sequence and amino acid substitutions as an input option for this tool.

Mendelian clinically applicable pathogenicity (M-CAP)

M-CAP (<http://bejerano.stanford.edu/mcap/>) is the first pathogenicity classifier for rare missense variants in the human genome and freely available Web-based tool. It is highly sensitive (Jagadeesh et al. 2016). It outperforms the various existing tools at all thresholds and accurately dismisses 60% of rare missense variants of tentative significance in the distinctive genome at 95% sensitivity. We have submitted the coordinates as chromosome position of corresponding rs ids. The

tool shows the M-CAP score and pathogenic prediction with 95% sensitivity.

Homology modeling and structural prediction

The homology modeling approach was performed by SWISS-MODEL to analyze the quality of the *NEGR1* protein and predicted nsSNPs (obtained by PolyPhen, SIFT, PROVEAN, MutPred and M-CAP tools) to find out their possible 3D structures (Arnold et al. 2006; Biasini et al. 2014). We used the *NEGR1* protein sequence (UniProt Id: Q7Z3B1) for homology modeling, where we used PDB Id: IQZ1.1A (*Rattus norvegicus*) as the best template with 22.22% identity, since there was no crystal structure present in the Protein Data Bank (PDB). After homology modeling, PDB files were downloaded for native and mutant models for further structural analysis. The PROCHECK server was used for the structural validation, stereochemical properties and Ramachandran plot analysis of the predicted models (Laskowski et al. 1996). The analysis of other structural properties such as RMSD, free energy and energy minimization was carried out and visualized under the PyMOL (Version 1.7.4 Schrodinger, LLC) viewer, Swiss-PDB viewer and NOMAD-Ref server (Lindahl et al. 2006).

Prediction of ligand-binding site with FTSite server

FTSite (<http://ftsites.bu.edu>) is an online server. It predicts the ligand-binding sites of proteins with very high accuracy (Ngan et al. 2012). It was performed to check whether the identified nsSNPs were present in the *NEGR1* protein-binding region.

Results

Distribution, selection, and prediction of SNPs

The *NEGR1* gene (*Homo sapiens*) had a total of 44,614 SNPs, out of which 101 were nsSNPs (missense), 185 were 3'UTR, 31 were 5'UTR, 65 were coding synonymous, 4 were frameshift, 3 were nonsense, 2 were stop gained and 44,169 were in the intronic region (searched on 22.11.2016 at dbSNP and Genecard database). In this investigation, we considered all 101 coding nsSNPs for the prediction of potential deleterious nsSNPs by five different computational tools, i.e., PolyPhen, SIFT, PROVEAN, MutPred and M-CAP (Fig. 1).

Identification of deleterious nsSNPs with PolyPhen 2.0 tool

The PolyPhen 2.0 tool identifies nsSNPs based on amino acid substitution with a PSIC score. Based on the PSIC

Fig. 1 Screening of deleterious nsSNPs in the *NEGR1* gene: this bar graph depicts screening and prediction of potential deleterious nsSNPs by five different computational tools as follows: PolyPhen (predicted deleterious nsSNPs=9), SIFT (predicted deleterious nsSNPs=12), PROVEAN (predicted deleterious nsSNPs=31), MutPred (predicted deleterious nsSNPs=21) and M-CAP (predicted pathogenic nsSNPs=21)

score, this computational tool classifies the nsSNPs as probably damaging, possibly damaging and benign. In this report, we considered all 101 nsSNPs for our analysis

and 9 nsSNPs were found to be probably deleterious with a PSIC score of ~ 1.0 (Table 1).

Identification of deleterious nsSNPs with SIFT tool

The SIFT tool predicts the nsSNPs based on amino acid substitution and gives details whether the substitution affects the protein function or not. This tool classifies the nsSNPs based on SIFT score as tolerated and deleterious. In this study, we have taken all 101 nsSNPs for our analysis and 12 nsSNPs were found to be deleterious with SIFT score of <0.05 (Table 2).

Identification of deleterious nsSNPs with PROVEAN tool

The PROVEAN tool predicts the deleterious nsSNPs and functional effect of the protein upon amino acid substitution. This tool classifies the nsSNPs as deleterious and neutral based on PROVEAN score. In our study, we have taken all 101 nsSNPs for the analysis and 31 nsSNPs were found to be deleterious (Table 3).

Table 1 List of deleterious nsSNPs identified by the PolyPhen tool

S. no.	SNP IDs	Amino acid substitution	Nucleotide substitution	PSIC score	Predicted impact
1	rs145524630	Ala70Thr	C/T	0.997	Probably damaging
2	rs267598710	Pro168Leu	A/G	0.998	Probably damaging
3	rs373419972	Arg239Cys	A/G	0.901	Possibly damaging
4	rs375352213	Leu158Phe	A/G	0.998	Probably damaging
5	rs552625764	Gly238Arg	C/T	0.746	Probably damaging
6	rs751003323	Ser280Phe	C/G	0.969	Probably damaging
7	rs759773075	Val135Ala	A/G	0.989	Probably damaging
8	rs761710178	Asp83Ala	G/T	0.995	Probably damaging
9	rs778158263	Tyr101Cys	C/T	0.999	Probably damaging

Table 2 List of deleterious nsSNPs identified by the SIFT tool

S. no.	SNP IDs	Amino acid substitution	Nucleotide substitution	SIFT score	Predicted impact
1	rs145524630	Ala70Thr	C/T	0	Deleterious
2	rs267598710	Pro168Leu	A/G	0	Deleterious
3	rs373419972	Arg239Cys	A/G	0	Deleterious
4	rs375352213	Leu158Phe	A/G	0	Deleterious
5	rs202121169	Arg195Thr	C/G	0	Deleterious
6	rs267598711	Asp46His	C/G	0.004	Deleterious
7	rs368350763	Gln8His	C/G	0.001	Deleterious
8	rs371044840	Asn47Asp	C/T	0.02	Deleterious
9	rs371109100	Ala308Val	A/C/G	0.06/0	Tolerated/deleterious
10	rs376699190	Arg74Gly	C/G	0	Deleterious
11	rs41289154	Tyr347Ser	G/T	0.001	Deleterious
12	rs183107213	Val20Gly	A/C	0.002	Deleterious

Identification of deleterious nsSNPs with MutPred tool

The MutPred tool identifies nsSNPs and categorizes an amino acid substitution as a disease associated or neutral in humans. It predicts the molecular cause of a disease and also classifies three hypotheses as actionable, confident and very confident. It also gives a score of the probability of deleterious mutation, and based on the score predicts the top five features associated with functional and structural changes in a protein upon amino acid modifications. In this study, we have taken all 101 nsSNPs for our analysis and 21 nsSNPs were found to be deleterious with the top five features (Table 4).

Identification of pathogenic nsSNPs with M-CAP score

The M-CAP tools identify nsSNPs alternate alleles for its pathogenic effect over 95% sensitivity by M-CAP score. Out of total analyzed 101 nsSNPs, 21 nsSNPs were predicted to be possibly pathogenic over one or more alleles from M-CAP (Table 5).

Comparison between PolyPhen, SIFT, PROVEAN, MutPred and M-CAP tools

The computational tools used in this study had their predictive capability to find deleterious nsSNPs based on the

Table 3 List of deleterious nsSNPs identified with PROVEAN tool

S. no.	SNP IDs	Amino acid substitution	Nucleotide substitution	PROVEAN score	Predicted impact
1	rs145524630	Ala70Thr	C/T	−3.608	Deleterious
2	rs267598710	Pro168Leu	A/G	−9.483	Deleterious
3	rs373419972	Arg239Cys	A/G	−4.690	Deleterious
		Arg195Cys		−7.417	Deleterious
4	rs375352213	Leu158Phe	A/G	−3.837	Deleterious
5	rs202121169	Arg195Thr	C/G	−5.529	Deleterious
6	rs267598711	Asp46His	C/G	−4.908	Deleterious
7	rs371044840	Asn47Asp	C/T	−3.244	Deleterious
8	rs371109100	Ala308Val	A/C/G	−2.725	Deleterious
9	rs376699190	Arg74Gly	C/G	−6.359	Deleterious
10	rs370438310	Asn151Ser	C/T	−4.797	Deleterious
11	rs535136498	Ala299Val	A/G	−3.587	Deleterious
12	rs540095646	Trp16Arg	A/G	−2.973	Deleterious
13	rs548564946	Glu202Asp	A/T	−2.750	Deleterious
		Glu246Asp		−2.527	Deleterious
14	rs576347795	Gly238Arg	C/T	−7.281	Deleterious
15	rs747740530	Pro315Leu	A/G	−5.103	Deleterious
16	rs751156857	Ser280Phe	A/G	−5.108	Deleterious
17	rs751848901	Asn222Ser	C/T	−4.480	Deleterious
18	rs752419911	Arg173Gln	C/T	−3.171	Deleterious
19	rs756093179	Gly324Arg	C/T	−2.562	Deleterious
20	rs756257324	Pro177Leu	A/G	−6.715	Deleterious
21	rs759773075	Val135Ala	A/G	−3.756	Deleterious
22	rs761710178	Asp83Ala	G/T	−7.158	Deleterious
23	rs762264715	Asp147Asn	C/T	−3.838	Deleterious
24	rs763636088	Arg244Ser	G/T	−2.768	Deleterious
25	rs765126599	Pro177Ala	C/G	−5.009	Deleterious
26	rs767134486	Asp208Gly	A/C/T	−3.914	Deleterious
		Asp208Val		−6.755	Deleterious
27	rs767198961	Arg195His	C/T	−4.630	Deleterious
28	rs768784952	Asn222Ser	C/T	−4.480	Deleterious
29	rs773893499	Gly241Val	A/C	−5.284	Deleterious
30	rs774857815	Asp147Val	A/T	−7.265	Deleterious
31	rs778158263	Tyr101Cys	C/T	−7.337	Deleterious

Table 4 List of deleterious nsSNPs identified with the MutPred tool

S. no.	SNP IDs	Amino acid substitution	Nucleotide substitution	Probability of deleterious mutation	Top 5 features predicted by MutPred
1	rs145524630	Ala70Thr	C/T	0.579	Loss of catalytic residue at A70 ($P=0.003$); gain of ubiquitination at K68 ($P=0.1162$); gain of methylation at K68 ($P=0.1221$); gain of MoRF binding ($P=0.2216$); gain of disorder ($P=0.2293$)
2	rs267598710	Pro168Leu	A/G	0.802	Loss of disorder ($P=0.0207$); gain of sheet ($P=0.0344$); loss of methylation at K165 ($P=0.0372$); loss of glycosylation at P168 ($P=0.0452$); loss of ubiquitination at K165 ($P=0.059$)
3	rs373419972	Arg239Cys	A/G	0.509	Loss of disorder ($P=0.0358$); loss of phosphorylation at T236 ($P=0.0885$); loss of glycosylation at S240 ($P=0.1059$); loss of methylation at R239 ($P=0.135$); loss of sheet ($P=0.1501$)
4	rs375352213	Leu158Phe	A/G	0.715	Loss of sheet ($P=0.1158$); gain of catalytic residue at L158 ($P=0.2353$); gain of phosphorylation at T159 ($P=0.2915$); loss of glycosylation at T163 ($P=0.3467$); loss of stability ($P=0.3644$)
5	rs183107213	Val20Gly	A/C	0.581	Loss of stability ($P=1e-04$); loss of helix ($P=0.0444$); loss of catalytic residue at V20 ($P=0.094$); gain of loop ($P=0.1069$); gain of glycosylation at S23 ($P=0.6173$)
6	rs371109100	Ala66Pro	A/C/G	0.503	Loss of catalytic residue at A66 ($P=0.0433$); loss of sheet ($P=0.0457$); gain of methylation at K68 ($P=0.0711$); loss of stability ($P=0.083$); gain of loop ($P=0.0851$)
7	rs576347795	Gly238Arg	C/T	0.795	Gain of solvent accessibility ($P=0.0306$); gain of glycosylation at T234 ($P=0.0712$); gain of MoRF binding ($P=0.0722$); gain of phosphorylation at S240 ($P=0.0829$); gain of disorder ($P=0.2211$)
8	rs746827197	Ser30Thr	A/T	0.583	Gain of glycosylation at S30 ($P=0.0265$); gain of helix ($P=0.132$); loss of catalytic residue at S30 ($P=0.3389$); loss of sheet ($P=0.5184$); loss of disorder ($P=0.5965$)
9	rs751848901	Asn222Ser	C/T	0.551	Loss of sheet ($P=0.0457$); gain of ubiquitination at K218 ($P=0.0913$); gain of loop ($P=0.1069$); gain of phosphorylation at N222 ($P=0.1582$); gain of disorder ($P=0.2171$)
10	rs768784952	Asn222Ser	C/T	0.551	Loss of sheet ($P=0.0457$); gain of ubiquitination at K218 ($P=0.0913$); gain of loop ($P=0.1069$); gain of phosphorylation at N222 ($P=0.1582$); gain of disorder ($P=0.2171$)
11	rs756289443	Pro41Arg	C/G	0.691	Loss of catalytic residue at W42 ($P=0.0098$); gain of MoRF binding ($P=0.086$); gain of sheet ($P=0.1208$); loss of loop ($P=0.2237$); loss of stability ($P=0.4276$)
12	rs757450838	Lys218Glu	C/T	0.567	Loss of ubiquitination at K218 ($P=0.027$); loss of MoRF binding ($P=0.066$); loss of sheet ($P=0.0817$); gain of relative solvent accessibility ($P=0.09$); gain of solvent accessibility ($P=0.0971$)
13	rs759429542	Ser23Gly	C/T	0.565	Loss of helix ($P=0.0444$); gain of catalytic residue at L24 ($P=0.116$); loss of stability ($P=0.1792$); gain of loop ($P=0.3485$); gain of sheet ($P=0.4423$)
14	rs759773075	Val135Ala	A/G	0.576	Loss of sheet ($P=0.0181$); gain of loop ($P=0.0851$); loss of methylation at K140 ($P=0.1864$); gain of disorder ($P=0.2177$); loss of stability ($P=0.218$)
15	rs761710178	Asp83Ala	G/T	0.615	Gain of catalytic residue at D83 ($P=0.0067$); gain of MoRF binding ($P=0.0512$); loss of solvent accessibility ($P=0.0595$); gain of ubiquitination at K84 ($P=0.0798$); loss of relative solvent accessibility ($P=0.1807$)
16	rs762264715	Asp147Asn	C/T	0.610	Loss of catalytic residue at D147 ($P=0.0171$); loss of stability ($P=0.0857$); loss of phosphorylation at Y142 ($P=0.1557$); loss of loop ($P=0.2237$); gain of disorder ($P=0.2971$)
17	rs762660335	Ala19Pro	C/G	0.639	Loss of helix ($P=0.0444$); loss of stability ($P=0.1234$); gain of loop ($P=0.2045$); gain of catalytic residue at A19 ($P=0.3791$); gain of sheet ($P=0.6509$)

Table 4 (continued)

S. no.	SNP IDs	Amino acid substitution	Nucleotide substitution	Probability of deleterious mutation	Top 5 features predicted by MutPred
18	rs765470454	Ile190Val	C/T	0.643	Loss of catalytic residue at I190 ($P=0.0269$); gain of phosphorylation at Y191 ($P=0.1369$); gain of sheet ($P=0.1451$); gain of loop ($P=0.2045$); gain of disorder ($P=0.4458$)
19	rs773720428	Phe342Ser	A/G	0.534	Loss of stability ($P=0.0039$); gain of glycosylation at F342 ($P=0.0224$); gain of sheet ($P=0.039$); loss of loop ($P=0.0986$); gain of disorder ($P=0.2845$)
20	rs773893499	Gly241Val	A/C	0.530	Loss of glycosylation at S240 ($P=0.0449$); loss of disorder ($P=0.0645$); loss of catalytic residue at P237 ($P=0.0698$); loss of methylation at R244 ($P=0.2221$); loss of phosphorylation at T236 ($P=0.2522$)
21	rs779210726	Asp2Asn	C/T	0.524	Loss of sheet ($P=0.0315$); gain of loop ($P=0.069$); gain of relative solvent accessibility ($P=0.09$); gain of catalytic residue at D2 ($P=0.1291$); gain of solvent accessibility ($P=0.1505$)

different mechanisms of action. In this investigation, we compared the commonly predicted deleterious nsSNPs by these five tools. After comparison, we were able to identify that four deleterious nsSNPs, i.e., rs145524630 (Ala70Thr), rs267598710 (Pro168Leu), rs373419972 (Arg239Cys) and rs375352213 (Leu158Phe), were functionally significant (Table 6).

Homology modeling and structural analysis of native and mutant (rs145524630, rs267598710, rs373419972, and rs375352213) NEGR1 protein

Homology modeling approach was carried out with the SWISS-MODEL server to predict the possible 3D structures of native and mutant (rs145524630, rs267598710, rs373419972, and rs375352213) *NEGR1* protein (Biasini et al. 2014). The 3D structure of the native and mutant models and selected template along with their alignment has been predicted (Fig. 2). The Swiss PDB Viewer with GROMOS 96 program and NOMAD-Ref server was used to perform the total energy calculations and energy minimizations for the native and mutated structures. In this study, the total energy of the native structure (*NEGR1*: PDB ID: 1QZ1.1.A) after energy minimization was -10083.506 KJ/mol and before energy minimization, 1617.995 KJ/mol (Table 7). After energy minimization, we found that the total energy of the mutant models was higher than that of the native model. This variation in the total energy after minimization depicted that the mutant models were less stable than the native model. The RMSD values were calculated by superimposing the mutant models over the native model under a PyMOL viewer (Version 1.7.4 Schrodinger, LLC.) (Fig. 3; Table 7). In this case, we found that rs373419972 had greater RMSD value than rs267598710, rs375352213 and rs145524630, which indicates the more deleterious

nature of rs373419972 than rs267598710, rs375352213 and rs145524630 (Table 7).

Furthermore, the stereochemical properties and structural validation of the predicted models of the *NEGR1* protein were investigated by the PROCHECK algorithm. The stereochemical quality of a protein and structure and analysis by its geometry were investigated by the PROCHECK algorithm. This program is beneficial for evaluating the quality of the existing protein structures as well as modeled structures (Laskowski et al. 1996). In this study, the PROCHECK algorithm was used to calculate the Ramachandran plot statistics. All the predicted structural models exhibited slightly less than 90 % of the residues in the most favoured region in regions (A, B, L) in the Ramachandran plot (Fig. 4; Table 8). Thus, this analysis supports that the predicted models generate an optimum quality structure. The PROCHECK also calculates the G factor which identifies the unusual structural property. In our investigation, we found the G factor value was optimum, i.e., it was > -0.5 , which shows that the structural properties of the predicted models were optimal (Table 8). Thus, the results of PROCHECK analysis suggest that the stereochemical and structural properties of the predicted models were good enough to proceed for further analysis.

Extrapolation of ligand-binding sites of the *NEGR1* protein with the FTSite server

The FTSite server finds the ligand-binding regions of a particular protein and this algorithm does not depend on any evolutionary or statistical information. This server detects regions in a protein that have the maximum number of non-bonded interactions with overlapping low energy clusters of numerous small molecular probes which improve the robustness of FTSite (Ngan et al. 2012). In our investigation, we

Table 5 Analysis of nsSNPs by M-CAP for pathogenic effect

Sl. no.	SNP IDs, amino acid substitution, nucleotide substitution	Alt allele	M-CAP	95% sensitivity
1	rs267598711 Asp46His, C/G	C	0.026	Possibly pathogenic
2	rs138829805 Leu6Phe, C/G	A	0.028	Possibly pathogenic
		G	0.027	Possibly pathogenic
3	rs145524630 Ala70Thr, C/T	G	0.042	Possibly pathogenic
		T	0.026	Possibly pathogenic
4	rs183107213 Val20Gly, A/G	C	0.065	Possibly pathogenic
		G	0.035	Possibly pathogenic
5	rs267598710 P168L, A/G	A	0.102	Possibly pathogenic
		C	0.130	Possibly pathogenic
		T	0.112	Possibly pathogenic
6	rs267598711 Asp46His, C/G	A	0.067	Possibly pathogenic
		G	0.033	Possibly pathogenic
		T	0.033	Possibly pathogenic
7	Rs370438310 Asn151Ser, C/T	A	0.096	Possibly pathogenic
		G	0.043	Possibly pathogenic
		C	0.053	Possibly pathogenic
8	rs373419972 Arg239Cys, A/G	A	0.109	Possibly pathogenic
		T	0.036	Possibly pathogenic
9	rs375352213 Leu158Phe, A/G	G	0.014	Possibly pathogenic
10	rs535136498 Ala225Val, A/G	A	0.071	Possibly pathogenic
		C	0.071	Possibly pathogenic
		T	0.333	Possibly pathogenic
11	rs548564946 Glu202Asp, A/T	A	0.036	Possibly pathogenic
		G	0.035	Possibly pathogenic
12	rs576347795 Gly194Arg, C/T	G	0.347	Possibly pathogenic
		T	0.345	Possibly pathogenic
13	rs751848901 Asn222Ser, C/T	A	0.182	Possibly pathogenic
		C	0.069	Possibly pathogenic
		G	0.080	Possibly pathogenic
14	rs753502373 Tyr142Cys, C/T	C	0.030	Possibly pathogenic
15	rs757586998 Arg235Ser, G/T	A	0.032	Possibly pathogenic
		G	0.032	Possibly pathogenic
16	rs759773075 Val135Ala, A/G	C	0.134	Possibly pathogenic
		G	0.074	Possibly pathogenic
		T	0.117	Possibly pathogenic
17	rs761710178 Asp83Ala, G/T	A	0.032	Possibly pathogenic
		C	0.029	Possibly pathogenic
		G	0.034	Possibly pathogenic
18	rs763636088 Arg200Ser, G/T	A	0.036	Possibly pathogenic
		G	0.035	Possibly pathogenic
19	rs765470454 Ile190Val, C/T	A	0.079	Possibly pathogenic
		C	0.032	Possibly pathogenic
		G	0.029	Possibly pathogenic
20	rs767730810 Gln15His, A/C	C	0.065	Possibly pathogenic
		G	0.035	Possibly pathogenic
		T	0.066	Possibly pathogenic

Table 5 (continued)

Sl. no.	SNP IDs, amino acid substitution, nucleotide substitution	Alt allele	M-CAP	95% sensitivity
21	rs774857815 Asp147Val, A/T	A	0.105	Possibly pathogenic
		C	0.083	Possibly pathogenic
		G	0.056	Possibly pathogenic

used this server to predict whether the identified nsSNPs are present in the *NEGR1* protein-binding region or not. Afterward, we found three different ligand-binding sites with different amino acid residues in the *NEGR1* protein (Table 9). Interestingly, we found one nsSNP, i.e., rs145524630 (Ala-70Thr), out of four identified nsSNPs present in the second ligand-binding site of the *NEGR1* protein (Fig. 5). Hence, this nsSNP (rs145524630) of the *NEGR1* gene might be a potential drug target for the management of human obesity.

Discussion

Obesity, the most common disorder, has been affecting the human population worldwide. It has become a global epidemic which increases day by day. According to WHO, 1.46 billion adults are overweight, and over 200 million men and 300 million women are obese globally (Finucane et al. 2011). According to, NCD-RisC, 2016, if this trend continues till 2025 then global obesity prevalence will reach 18% in men and surpass 21% in women (NCD-RisC 2016). Obesity gives rise to several diseases such as type 2 diabetes, hypertension, cardiovascular disorders, osteoarthritis, Alzheimer's, prostate cancer, lower back pain and lumbar disc diseases (McCrindle 2015; Agalliu et al. 2015; Dario et al. 2015; Power et al. 2014; Visser et al. 2014; Luchsinger et al. 2012; Kotsis et al. 2010). Due to this epidemic and its related health complications, millions of people lose their lives every year. The genetic characteristics of obesity lead to an alteration in many genes responsible for regulating appetite and metabolism. Besides, more than 127 sites in the human genome have been reported to link with the development of obesity, based on GWAS studies (Alonso et al. 2016). Epigenetics, gene–environment interaction (GEI), obesity-related health complications (ORHC), energy homeostasis and genetic aspect of human obesity are important topics of modern research in the field of human obesity (Singh et al. 2017). The *NEGR1* is a newly recognized raft-associated protein, which has recently been featured as a new locus related to human obesity (Kim et al. 2017). Niemann–Pick disease type C2 (NPC2) is a fatal human neurodegenerative disease, which stimulates intracellular cholesterol trafficking, and in this case, *NEGR1* interacts with NPC2 and increases its protein stability. Thereby, *NEGR1* plays a role in intracellular cholesterol homeostasis and is possibly linked with

human obesity (Kim et al. 2017). The *NEGR1* gene plays a vital role in cell to cell adhesion and neurite outgrowth in hypothalamic neurons, which is associated with obesity and highly expressed in the cerebral cortex, hippocampus and cerebellum during postnatal development (Schäfer et al. 2005; Lee et al. 2012). It has also been shown that the *NEGR1* gene is involved in neurite outgrowth and also determines synapse number in the hippocampus during neuronal development (Pischedda et al. 2014; Hashimoto et al. 2008; Marg et al. 1999). Genetic variations in the *NEGR1* gene, strongly expressed in the brain, have been reported to affect the neuronal control of food intake that leads to inducing obesity. This genetic inconsistency may be associated with psychological traits usually displayed by eating disorder in the patients (Gamero-Villarreal et al. 2015). According to Gamero-Villarreal et al. (2015), there were 21 tag-SNPs identified in the coding sequence and adjacent regions of the *NEGR1* gene. SNPs are positioned in a genome where either of two different nucleotides can occur. In humans, polymorphisms are considered to be SNPs only if the least abundant allele has a frequency of 1%. These types of nucleotide variations may cause several diseases such as obesity, diabetes, hypertension and psychiatric disorders (Shastri 2009). SNPs play a vital role in identifying genetic variants and are used as potential biomarkers for exploring the deleterious and neutral effects on protein function associated with obesity (Lyon and Hirschhorn 2005). The extensive study of the SNPs on the biological system is not so easy because it resembles functional consequences such as amino acid changes, mRNA transcript stability changes and transcription factor-binding affinity changes (Griffith et al. 2007). Therefore, in this perspective, computational study may play a crucial role in identifying and predicting the deleterious SNPs which affect protein function. The computational study is easier and more reliable to carry out than biological experiments (Singh and Mahalingam 2017; Singh et al. 2016).

Since the *NEGR1* gene has no missense mutations reported till date according to the HGMD database (<http://www.hgmd.cf.ac.uk>), the present study was undertaken to explore the nsSNPs of the *NEGR1* gene responsible for human obesity using different computational tools to predict the deleterious nature of the mutation and structural analysis. In this study, five computational tools, viz, PolyPhen, SIFT, PROVEAN, MutPred and M-CAP, were used

Table 6 List of common nsSNPs that were found to be deleterious by the PolyPhen, SIFT, PROVEAN, MutPred and M-CAP tools

SNP IDS	Amino acid substitution	Nucleotide substitution	PolyPhen		SIFT		PROVEAN		MutPred	Top 5 features	95% sensitivity
			PSIC score	Prediction	Score	Prediction	Score	Prediction			
rs145524630	Ala70Thr	C/T	0.998	Probably damaging	0	Deleterious	-3.608	Deleterious	0.579	Loss of catalytic residue at A70 ($P=0.003$); gain of ubiquitination at K68 ($P=0.1162$); gain of methylation at K68 ($P=0.1221$); gain of MoRF binding ($P=0.2216$); gain of disorder ($P=0.2293$)	Possibly pathogenic
rs267598710	Pro168Leu	A/G	0.998	Probably damaging	0	Deleterious	-9.483	Deleterious	0.802	Loss of disorder ($P=0.0207$); gain of sheet ($P=0.0344$); loss of methylation at K165 ($P=0.0372$); loss of glycosylation at P168 ($P=0.0452$); loss of ubiquitination at K165 ($P=0.059$)	Possibly pathogenic
rs373419972	Arg239Cys	A/G	0.901	Possibly damaging	0	Deleterious	-4.690	Deleterious	0.509	Loss of disorder ($P=0.0358$); loss of phosphorylation at T236 ($P=0.0885$); loss of glycosylation at S240 ($P=0.1059$); loss of methylation at R239 ($P=0.135$); loss of sheet ($P=0.1501$)	Possibly pathogenic
rs375352213	Leu158Phe	A/G	0.998	Probably damaging	0	Deleterious	-3.837	Deleterious	0.715	Loss of sheet ($P=0.1158$); gain of catalytic residue at L158 ($P=0.2353$); gain of phosphorylation at T159 ($P=0.2915$); loss of glycosylation at T163 ($P=0.3467$); loss of stability ($P=0.3644$)	Possibly pathogenic

Fig. 2 Homology-modeled 3D structures and model—template alignment of native and mutant *NEGR1* protein. **a** Native protein model, **b** mutant protein model, rs145524630 (Ala70Thr), **c** mutant protein model, rs267598710 (Pro168Leu), **d** mutant protein model,

rs373419972 (Arg239Cys), **e** mutant protein model, rs375352213 (Leu158Phe), **f** PDB ID 1QZ1.1.A (*Rattus norvegicus*) used as template for *NEGR1* protein modeling

to predict the deleterious nsSNPs of the *NEGR1* gene (Singh and Mahalingam 2017; Singh et al. 2016; Kumar et al. 2015, 2009; Goswami 2015; Ng and Henikoff 2003; Li et al. 2009; Adzhubei et al. 2010; Choi et al. 2012). After analysis of these tools, we found four potential deleterious and pathogenic nsSNPs, i.e., rs145524630 (Ala70Thr), rs267598710

(Pro168Leu), rs373419972 (Arg239Cys) and rs375352213 (Leu158Phe). However, the deleterious nature and in vivo and in vitro analysis of the identified nsSNPs have not been verified experimentally till date. Furthermore, we have taken all four identified nsSNPs for homology modeling and structural analysis to predict their functional impact on the

Table 7 Total energy and RMSD analysis of native and mutant models

Model	Total energy (after minimization) (KJ/mol) ^a	RMSD (Å)
rs145524630 (Ala70Thr); (C/T)	− 8707.066	1.916
rs267598710 (Pro168Leu); (A/G)	− 7996.684	1.601
rs373419972 (Arg239Cys); (A/G)	− 8388.164	0.919
rs375352213 (Leu158Phe); (A/G)	− 8890.926	0.771

^aThe total energy of native structure (*NEGR1*: PDB ID: 1QZ1.1.A) after energy minimization is − 10083.506 KJ/mol and before energy minimization, 1617.995 KJ/mol

protein function. There was no crystal structure found in the PDB database, so homology modeling was carried out by the SWISS-MODEL server to predict possible 3-D structures of native and mutant *NEGR1* proteins for further structural analysis (Biasini et al. 2014). The SWISS PDB viewer associated with the GROMOS 96 program and NOMAD-Ref server was used to perform the energy minimizations and to compare the native and mutated structures (Lindahl et al. 2006). The main aim behind energy minimization is to model a protein structure which is free from steric clings and potential energy, preferably lower than that of a crystal structure and lying as close to the native structure as possible in terms of RMSD (Ramachandran et al. 2011; Bruccoleri and Karplus 1990). In this study, the total energy of the native structure (*NEGR1*: PDB ID: 1QZ1.1.A) after energy minimization was − 10083.506 KJ/mol and before energy minimization, 1617.995 KJ/mol. We have analyzed that the total energy of mutant protein models was higher than that of the native model after energy minimization, which showed the mutant

model was less stable than the native model. However, an increase in total energy after energy minimization will lead to disrupting the molecular structure of the protein and this high energy is responsible for the higher motion in the particles of molecules. Moreover, disruption of the non-covalent contacts in the protein molecule will decrease the stability of the protein molecule (Lodish et al. 2000). The RMSD values were anticipated by superimposing the mutant models over the native model under a PyMOL viewer (Version 1.7.4 Schrodinger, LLC.). Herein, we identified that the variant rs373419972 had more RMSD value than rs267598710, rs375352213 and rs145524630, and this designates more deleterious nature of rs373419972 than rs267598710, rs375352213 and rs145524630. Moreover, structural validation and stereochemical properties of the predicted models of the *NEGR1* protein were identified by the PROCHECK algorithm (Laskowski et al. 1996). The PROCHECK algorithm helped in the analysis of the Ramachandran plot statistics and G factor values. In our analysis, we found that the Ramachandran plot statistics displayed the percentage of the most favored regions (A, B, L), and those were less than 90% in all the predicted structural models, showing that the predicted models had an optimum quality structure. The G factor value was also optimum, i.e., it was > − 0.5 which shows that the structural properties of the predicted models were usual. Thus, the structural analysis of *NEGR1* protein suggests that the structural properties of the predicted models were usual, valid and optimum. Furthermore, the FTSite server was used to find whether the predicted nsSNPs were present in the *NEGR1* protein-binding region or not (Ngan et al. 2012). Interestingly, we found one nsSNP, i.e., rs145524630 (Ala70Thr) out of four identified nsSNPs present in the second ligand-binding region of the *NEGR1* protein. Thus, this nsSNP (rs145524630) of the *NEGR1* gene might play an essential role in controlling human obesity.

Fig. 3 Superimposed structures of mutants and native model of *NEGR1* protein. **a** Native vs. rs145524630 (Ala70Thr), **b** native vs. rs267598710 (Pro168Leu), **c** native vs. rs373419972 (Arg239Cys), **d** native vs. rs375352213 (Leu158Phe). In the figure, green color signifies the native model, whereas red color is the mutant model

Fig. 4 Ramachandran plot analysis: **a** Ramachandran plot of the native model. **b** Ramachandran plot of the rs145524630 (Ala70Thr) model. **c** Ramachandran plot of the rs267598710 (Pro168Leu) model.

d Ramachandran plot of the rs373419972 (Arg239Cys) model. **e** Ramachandran plot of the rs375352213 (Leu158Phe) model

Table 8 PROCHECK investigation of the native and mutant models of the *NEGR1* protein

Model	Ramachandran plot statistics				G factor
	Most favored region (A, B, L) (%)	Additional allowed region [a, b, l, p] (%)	Generously allowed region [~a, ~b, ~l, ~p] (%)	Disallowed regions (xx) (%)	Overall average
Native	88.2	9.3	2.5	0.0	−0.44
rs145524630 (Ala70Thr); (C/T)	85.7	11.4	2.1	0.8	−0.45
rs267598710 (Pro168Leu); (A/G)	89.5	7.6	2.9	0.0	−0.45
rs373419972 (Arg239Cys); (A/G)	86.5	10.1	3.0	0.4	−0.41
rs375352213 (Leu158Phe); (A/G)	86.5	10.5	2.1	0.8	−0.50

Table 9 Analysis of ligand-binding sites of the *NEGR1* protein with FTSite

Ligand-binding sites	Amino acid residues									
Ligand-binding site 1	Lys A259	Leu A264	Ile A272	Ile A273	Leu A282	Val A284	Glu A290	His A291	Tyr A295	
Ligand-binding site 2	Gly A69	Ala A70	Trp A71	Ile A77	Ile A78	Phe A79	Ala A80	Ile A93	Thr A95	Tyr A101
Ligand-binding site 3	AlaA44	Val A45	Met A48	Val A57	Leu A58	Arg A59				

Fig. 5 Ligand-binding sites of the *NEGR1* protein. This figure represents three ligand-binding sites (mesh loops). The first ligand-binding site (mesh loop 1) is shown in cyan. The second ligand-binding site (mesh loop 2) is shown in magenta. The third ligand-binding site (mesh loop 3) is shown in red. The nsSNP rs145524630 (Ala70Thr) is present in the second ligand-binding site of the *NEGR1* protein

Conclusion

In summary, based on the computational analysis of the *NEGR1* gene, it was concluded that this gene had a total of 101 nsSNPs (missense), out of which only four nsSNPs, i.e., rs145524630 (Ala70Thr), rs267598710 (Pro168Leu), rs373419972 (Arg239Cys) and rs375352213 (Leu158Phe), were predicted as potentially deleterious and pathogenic. There had been no missense mutations reported so far, according to the HGMD database. So, in the future, further functional in vitro and in vivo analysis will provide insight regarding these nsSNPs on the influence of *NEGR1* gene in causing human obesity. Moreover, we found one nsSNP, i.e., rs145524630 (Ala70Thr) out of four identified nsSNPs present in the second ligand-binding region of the

NEGR1 protein. Therefore, this nsSNP (rs145524630) of the *NEGR1* gene might be a potential drug target and play an important role in the treatment of human obesity. We can say that the computational study carried out here was cost-effective, easy to analyze and monitored the predicted deleterious nsSNPs of the *NEGR1* gene. It suggests that the predicted nsSNPs of the *NEGR1* gene would contribute to the drug discovery and development, as well as better understanding of human obesity.

Acknowledgements All authors are grateful to the Vellore Institute of Technology, Vellore, Tamilnadu, India, for providing the facility to carry out the presented work. Mr. Permendra Kumar is thankful to the University Grants Commission (Government of India) for providing the Rajiv Gandhi National Fellowship-Senior Research Fellowship (UGC-RGNF-SRF). The authors acknowledge Mr. Rajan Kumar Singh (Teaching cum Research Assistant, SBST, Vellore Institute of Technology, Vellore) for his technical help and manuscript preparation.

Compliance with ethical standards

Conflict of interest The authors declare no conflicts of interest for publishing this work.

References

- Adzhubei IA, Schmidt S, Peshkin L et al (2010) A method and server for predicting damaging missense mutations. *Nat Methods* 7:248–249. <https://doi.org/10.1038/nmeth0410-248>
- Agalliu I, Williams S, Adler B et al (2015) The impact of obesity on prostate cancer recurrence observed after exclusion of diabetics. *Cancer Causes Control* 26:821–830. <https://doi.org/10.1007/s10552-015-0554-z>
- Alonso R, Fariás M, Alvarez V, Cuevas A (2016) The genetics of obesity. In: *Translational cardiometabolic genomic medicine*. Elsevier, Amsterdam, pp 161–177
- Altschul SF, Madden TL, Schäffer AA et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 22:195–201. <https://doi.org/10.1093/bioinformatics/bti770>
- Biasini M, Bienert S, Waterhouse A et al (2014) SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Res* 42:W252–W258. <https://doi.org/10.1093/nar/gku340>
- Bruccoleri RE, Karplus M (1990) Conformational sampling using high-temperature molecular dynamics. *Biopolymers* 29:1847–1862. <https://doi.org/10.1002/bip.360291415>
- Choi Y, Sims GE, Murphy S et al (2012) Predicting the functional effect of amino acid substitutions and indels. *PLoS One* 7:e46688. <https://doi.org/10.1371/journal.pone.0046688>
- Dario AB, Ferreira ML, Refshauge KM et al (2015) The relationship between obesity, low back pain, and lumbar disc degeneration when genetics and the environment are considered: a systematic review of twin studies. *Spine J* 15:1106–1117. <https://doi.org/10.1016/j.spinee.2015.02.001>
- Delorenzi M, Speed T (2002) An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics* 18:617–625
- Finucane MM, Stevens GA, Cowan MJ et al (2011) National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9.1 million participants. *Lancet* 377:557–567. [https://doi.org/10.1016/S0140-6736\(10\)62037-5](https://doi.org/10.1016/S0140-6736(10)62037-5)
- Frayling TM, Timpson NJ, Weedon MN et al (2007) A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* 316:889–894. <https://doi.org/10.1126/science.1141634>
- Gamero-Villaruel C, González LM, Gordillo I et al (2015) Impact of NEGR1 genetic variability on psychological traits of patients with eating disorders. *Pharmacogenom J* 15:278–283. <https://doi.org/10.1038/tpj.2014.53>
- Garg ML, Wood LG (2013) *Nutrition and physical activity in inflammatory diseases*. CABI, Wallingford
- Goswami AM (2015) Structural modeling and in silico analysis of non-synonymous single nucleotide polymorphisms of human 3 β -hydroxysteroid dehydrogenase type 2. *Meta Gene* 5:162–172. <https://doi.org/10.1016/j.mgene.2015.07.007>
- Griffith OL, Montgomery SB, Bernier B et al (2007) ORegAnno: an open-access community-driven resource for regulatory annotation. *Nucleic Acids Res* 36:D107–D113. <https://doi.org/10.1093/nar/gkm967>
- Guh DP, Zhang W, Bansback N et al (2009) The incidence of comorbidities related to obesity and overweight: a systematic review and meta-analysis. *BMC Public Health*. <https://doi.org/10.1186/1471-2458-9-88>
- Hashimoto T, Yamada M, Maekawa S et al (2008) IgLON cell adhesion molecule Kilon is a crucial modulator for synapse number in hippocampal neurons. *Brain Res* 1224:1–11. <https://doi.org/10.1016/j.brainres.2008.05.069>
- Jagadeesh K, Wenger A, Berger M, Guturu H, Stenson P, Cooper D, Bernstein J, Bejerano G (2016) M-CAP eliminates a majority of variants with uncertain significance in clinical exomes at high sensitivity. *Nat Genet*. <https://doi.org/10.1038/ng.3703>
- Kim H, Chun Y, Che L et al (2017) The new obesity-associated protein, neuronal growth regulator 1 (NEGR1), is implicated in Niemann–Pick disease Type C (NPC2)-mediated cholesterol trafficking. *Biochem Biophys Res Commun* 482:1367–1374. <https://doi.org/10.1016/j.bbrc.2016.12.043>
- Komar AA (2007) GENETICS: SNPs, silent but not invisible. *Science* 315:466–467. <https://doi.org/10.1126/science.1138239>
- Kotsis V, Stabouli S, Papakatsika S et al (2010) Mechanisms of obesity-induced hypertension. *Hypertens Res* 33:386–393. <https://doi.org/10.1038/hr.2010.9>
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol* 305:567–580. <https://doi.org/10.1006/jmbi.2000.4315> (Edited by F. Cohen)
- Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4:1073–1081. <https://doi.org/10.1038/nprot.2009.86>
- Kumar P, Singh RK, Mahalingam K (2015) In silico analysis of fat mass obesity associated (FTO) gene using computational algorithms. *Int J Pharma Bio Sci* 6:B589–B599
- Laskowski RA, Rullmannn JA, MacArthur MW et al (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8:477–486
- Lee AWS, Hengstler H, Schwald K et al (2012) Functional inactivation of the genome-wide association study obesity gene neuronal growth regulator 1 in mice causes a body mass phenotype. *PLoS One* 7:e41537. <https://doi.org/10.1371/journal.pone.0041537>
- Lee H, Ash GI, Angelopoulos TJ et al (2015) Obesity-related genetic variants and their associations with physical activity. *Sports Med Open* 1:34. <https://doi.org/10.1186/s40798-015-0036-6>
- Li B, Krishnan VG, Mort ME et al (2009) Automated inference of molecular mechanisms of disease from amino acid substitutions. *Bioinformatics* 25:2744–2750. <https://doi.org/10.1093/bioinformatics/btp528>
- Lindahl E, Azuara C, Koehl P, Delarue M (2006) NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res* 34:W52–W56. <https://doi.org/10.1093/nar/gkl082>
- Lodish H, Berk A, Zipursky SL et al (2000) *Molecular cell biology*, 4th edn. W. H. Freeman, San Francisco
- Luchsinger JA, Cheng D, Tang MX et al (2012) Central obesity in the elderly is related to late-onset Alzheimer disease. *Alzheimer Dis Assoc Disord* 26:101–105. <https://doi.org/10.1097/WAD.0b013e318222f0d4>
- Lyon HN, Hirschhorn JN (2005) Genetics of common forms of obesity: a brief overview. *Am J Clin Nutr* 82:215S–217S
- Marg A, Sirim P, Spaltmann F et al (1999) Neurotractin, a novel neurite outgrowth-promoting Ig-like protein that interacts with CEPU-1 and LAMP. *J Cell Biol* 145:865–876

- McCordle BW (2015) Cardiovascular consequences of childhood obesity. *Can J Cardiol* 31:124–130. <https://doi.org/10.1016/j.cjca.2014.08.017>
- Miyata S, Matsumoto N, Taguchi K et al (2003) Biochemical and ultrastructural analyses of IgLON cell adhesion molecules, Kilon and OBCAM in the rat brain. *Neuroscience* 117:645–658
- NCD Risk Factor Collaboration (2016) Trends in adult body-mass index in 200 countries from 1975 to 2014: a pooled analysis of 1698 population-based measurement studies with 19.2 million participants. *Lancet* 387:1377–1396. [https://doi.org/10.1016/S0140-6736\(16\)30054-X](https://doi.org/10.1016/S0140-6736(16)30054-X)
- Ng PC, Henikoff S (2003) SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* 31:3812–3814
- Ngan C-H, Hall DR, Zerbe B et al (2012) FTSite: high accuracy detection of ligand binding sites on unbound protein structures. *Bioinformatics* 28:286–287. <https://doi.org/10.1093/bioinformatics/btr651>
- Nogueiras R, Wiedmer P, Perez-Tilve D et al (2007) The central melanocortin system directly controls peripheral lipid metabolism. *J Clin Invest* 117:3475–3488. <https://doi.org/10.1172/JCI31743>
- Nolan CJ, Damm P, Prentki M (2011) Type 2 diabetes across generations: from pathophysiology to prevention and management. *Lancet* 378:169–181. [https://doi.org/10.1016/S0140-6736\(11\)60614-4](https://doi.org/10.1016/S0140-6736(11)60614-4)
- Pischedda F, Szczurkowska J, Cîrnaru MD et al (2014) A cell surface biotinylation assay to reveal membrane-associated neuronal cues: Negr1 regulates dendritic arborization. *Mol Cell Proteom* 13:733–748. <https://doi.org/10.1074/mcp.M113.031716>
- Power C, Pinto Pereira SM, Law C, Ki M (2014) Obesity and risk factors for cardiovascular disease and type 2 diabetes: investigating the role of physical activity and sedentary behaviour in mid-life in the 1958 British cohort. *Atherosclerosis* 233:363–369. <https://doi.org/10.1016/j.atherosclerosis.2014.01.032>
- Punta M, Coghill PC, Eberhardt RY et al (2012) The Pfam protein families database. *Nucleic Acids Res* 40:D290–D301. <https://doi.org/10.1093/nar/gkr1065>
- Ramachandran S, Kota P, Ding F, Dokholyan NV (2011) Automated minimization of steric clashes in protein structures. *Proteins Struct Funct Bioinf* 79:261–270. <https://doi.org/10.1002/prot.22879>
- Rozman V, Kunej T (2018) Harnessing omics big data in nine vertebrate species by genome-wide prioritization of sequence variants with the highest predicted deleterious effect on protein function. *OMICS* 22:410–421. <https://doi.org/10.1089/omi.2018.0046>
- Schäfer M, Bräuer AU, Savaskan NE et al (2005) Neurotractin/kilon promotes neurite outgrowth and is expressed on reactive astrocytes after entorhinal cortex lesion. *Mol Cell Neurosci* 29:580–590. <https://doi.org/10.1016/j.mcn.2005.04.010>
- Shastri BS (2009) SNPs: impact on gene function and phenotype. In: Komar AA (ed) *Single nucleotide polymorphisms*. Humana Press, Totowa, pp 3–22
- Sickmeier M, Hamilton JA, LeGall T et al (2007) DisProt: the database of disordered proteins. *Nucleic Acids Res* 35:D786–D793. <https://doi.org/10.1093/nar/gkl893>
- Singh RK, Mahalingam K (2017) In silico approach to identify non-synonymous SNPs in human obesity related gene, MC3R (melanocortin-3-receptor). *Comput Biol Chem* 67:122–130. <https://doi.org/10.1016/j.compbiolchem.2016.12.009>
- Singh RK, Kumar P, Mahalingam K (2016) In silico analysis of SNPs in human obesity related gene, SLC6A14 (solute carrier family 6 (amino acid transporter), member 14). *Int J Pharm Bio Sci* 248–256
- Singh RK, Kumar P, Mahalingam K (2017) Molecular genetics of human obesity: a comprehensive review. *C R Biol* 340:87–108. <https://doi.org/10.1016/j.crv.2016.11.007>
- Speliotes EK, Willer CJ, Berndt SI et al (2010) Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat Genet* 42:937–948. <https://doi.org/10.1038/ng.686>
- Thorleifsson G, Walters GB, Gudbjartsson DF et al (2009) Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat Genet* 41:18–24. <https://doi.org/10.1038/ng.274>
- Vazquez-Martinez R, Gasman S (2014) The regulated secretory pathway in neuroendocrine cells. *Front Endocrinol*. <https://doi.org/10.3389/fendo.2014.00048>
- Veerappa AM, Saldanha M, Padakannaya P, Ramachandra NB (2013) Family-based genome-wide copy number scan identifies five new genes of dyslexia involved in dendritic spine plasticity. *J Hum Genet* 58:539–547. <https://doi.org/10.1038/jhg.2013.47>
- Visser A, Ioan-Facsinay A, de Mutsert R et al (2014) Adiposity and hand osteoarthritis: the Netherlands Epidemiology of Obesity study. *Arthr Res Ther* 16:R19. <https://doi.org/10.1186/ar4447>
- Willer CJ, Speliotes EK, Loos RJF et al (2009) Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat Genet* 41:25–34. <https://doi.org/10.1038/ng.287>
- Williams LM (2012) Hypothalamic dysfunction in obesity. *Proc Nutr Soc* 71:521–533. <https://doi.org/10.1017/S002966511200078X>