INNSPUB

International Journal of Biosciences | IJB |
ISSN: 2220-6655 (Print) 2222-5234 (Online)
http://www.innspub.net
Vol. 4, No. 8, p. 222-230, 2014

RESEARCH PAPER OPEN ACCESS

Identifying and predicting the impact of functional single nucleotide polymorphisms (SNPs) in human *CAV-3* gene

Waseem El-Huneidi*

Department of Basic Sciences, College of Science and Health Professions, King Saud bin Abdulaziz University for Health Sciences, Riyadh, Saudi Arabia

Key words: *CAV3* gene, SNPs, In Silico, SIFT, PolyPhen 2, HOPE Project.

http://dx.doi.org/10.12692/ijb/4.8.222-230 Article published on April 22, 2014

Abstract

Genetic mutations, such as Single nucleotide polymorphisms (SNPs) are known to be the most reported type of genetic variations in individuals and thought to have a crucial impact on the understanding of many genetic related diseases. One of the major challenges is to find the deleterious SNPs in disease-associated genes. In the current study, computational methods were used to analyze the SNPs which may effect the expression and function of CAV 3 gene. Of the total 377 SNPs, 288(76.8%) were found to be non-coding SNPs, while 53 (14.1%) were non synonymous (ns)SNPs and 4(1.1%) were in the mRNA UTRs. Forty nine nsSNPs encoding missense mutation were tested for the parable effect on protein structure and functions using SIFT, I-Mutant 2, PolyPhen 2 and PANTHER soft wares. Moreover, one nsSNP (R27Q) was further analyzed using HOPE project, and this mutation predicted to substantially affect the structure along with the function of caveolin 3 protein.

* **Corresponding Author:** Waseem El-Huneidi ✉ huneidiw@ksau-hs.edu.sa

Introduction

Genetic mutations, such as Single nucleotide polymorphisms (SNPs) are known to be the most reported type of genetic variations. In Human Genome, more than 500,000 SNPs have been reported in the exonic regions(Collins *et al.*, 1998). SNPs are classified into synonymous and nonsynonymous SNPs (nsSNPs); the latter can amino acid substitution. nsSNPs may cause a crucial influence on the function and structure of proteins encoded in the human genome(Lander, 1996). Many genetic regulatory pathways could be affected by different SNPs, for instance, the nsSNPs could affect the DNA structure leading to impair its ability to bind to the corresponding transcriptional (Barroso *et al.*, 1999) and the structural integrity of cells can also be affected by nsSNP (Thomas *et al.*, 1999). Furthermore, nsSNPs were reported to be deleteriously to many signal transduction pathway proteins within the cell . (Dryja *et al.*, 1990; Smith *et al.*, 1994).

Caveolae are invaginations of the cell membrane which were revealed by electron microscopy. Caveolins are the main component of caveolae, which are a class of proteins including three isoforms, caveolin-1 (Cav-1), -2 and -3. Those proteins are encoded in evolutionarily conserved genes, specially Cav-1 and Cav-3 which exhibit the highest degree of homology in their amino acid sequences (Parton, 1996; Scherer *et al.*, 1996; Williams and Lisanti, 2004).

Initially, in vitro and in vivo studies were used to elucidate the role of caveolin 3 in the physiology of the muscle cell which was further confirmed by finding that many caveolin-3 gene (CAV3) mutations are associated to different cardiac and neuromuscular diseases such as Limb Girdle Muscular Dystrophy (LGMD) 1-C, idiopathic persistent elevation of serum creatine kinase (H-CK), inherited rippling muscle disease (RMD), distal myopathy (MD) and familial hypertrophic cardiomyopathy (HCM) (Galbiati *et al.*, 2001; Williams and Lisanti, 2004; Woodman *et al.*, 2004). Moreover, mutations in the CAV3 gene were revealed to be associated to arrhythmogenic long QT syndrome (LQTS) and sudden infant death syndrome (SIDS).

The present study was undertaken primarily to analyze (*in silico)* the SNPs in the *CAV3* gene, in order to identify and evaluate possible effect of important nsSNPs on the structure as well as the function of caveolin 3 protein

Materials and methods

Data mining for SNPs

To be used in computational analysis, the SNP information for *CAV3* and amino acid sequence were obtained from National Centre for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov/projects/SNP) and Ensembl Genome Browser (http://asia.ensembl.org/index.html).

Sift

To test the potential functional effect of amino acid substitution on the corresponding protein; Sorting Intolerant From Tolerant (SIFT) program, developed by Ng and Henikoff, based on sequence homolog, was used. SIFT works on the bases that protein functions are evolutionary interconnected, indicating that any variations in conserved regions of a protein will most probably lead to a damaging effect on that protein. (Ng and Henikoff, 2003). SIFT algorithm uses a modified version of PSIBLAST (Altschul *et al.*, 1997); Thus a position having a tolerance index of less than 0.05 is predicted to be damaging otherwise it will be predicted to be tolerated.

PolyPhen-2

Polymorphism Phenotyping-2 (PolyPhen-2) was used to investigate the possible consequences of nsSNPs on protein structure and function. The predictions depend on physical and evolutionary comparative considerations and naïve Bayes classifier is used to calculate the deleterious effects of nonsynonymous variants according to three structure and eight sequences as a root features for predictive (Adzhubei *et al.*, 2010). PolyPhen-2 scores classify the effect into

benign, possibly damaging and probably damaging. (Adzhubei *et al.*, 2013).

Mutant 2.0

I-Mutant 2.0 is a web available support vector machine (SVM) tool used to test for any change in protein stability after single site mutation. The difference in energy change (DDG) between the unfolded mutated native protein (Kcal/mol) are used as the basis for prediction by I-Mutant 2.0. If DDG is negative, this will demonstrate a low stability of mutated protein and vice versa. (Bava *et al.*, 2004).

Panther

A web based tool PANTHER (Protein analysis through evolutionary relationships) is used to cluster proteins by molecular function, with evolutionarily related protein families and subfamilies exhibiting similar biological function, pathway or process(Mi *et al.*, 2007; Mi *et al.*, 2013). Thus subPSEC score was used and calculated according to evolutionary related proteins and it estimates the potential consequence of nsSNPs on the protein. If the SubPSEC score is less than or equal to – 3, the effect is predicted to be deleterious. (Thomas *et al.*, 2003).

I-Tasser

The webserver I-TASSER from Zang's lab was used to predict the structures of both native and mutated proteins. This tool predicts models through removing continuous fragments from threading alignments and rebuilds those using replica-exchanged Monte Carlo simulations. Throughout the simulation, SPICKER cluster decoys (low temperature replicas) and then the full length atomic models are formed by choosing the top five cluster centroids. The decoy's number in unit space of SPICKER cluster is used to define the cluster density. In I-TASSER, The reliability of structures obtained is assessed using confidence score (C-score), which could be in the range of – 5 to 2. The higher C-Score of a predicted model represent a higher prediction confidence. In addition, C-score could be affected by TM-score and RMSD. If the structure of the native protein is known, the reported TM-score and RMSD are used to assess the predicted model obtained , otherwise, the same parameters, by considering one model as a native according to C-score, are used to evaluate the reliability of other predicted models. The TM-score is a newly proposed scale to unravel the local error problem usually associated with RMSD. TM-score is less than 0.17 shows a random similarity while TM-score more than 0.5 reflects a model of correct topoly (Zhang and Skolnick, 2004; Zhang, 2008; Roy *et al.*, 2010).

Project hope

To analyze mutants in order to study the structural features of wild type protein and the mutated models, Project Have yOur Protein Explained (HOPE; http://www.cmbi.ru.nl/hope/home) server was used. Project HOPE uses the protein sequence in FASTA format in addition to a selected mutant variants as an input method. The output of the HOPE server predicts the structural differences between native and mutant amino acids (Venselaar *et al.*, 2010).

Results and discussion

SNPs of *CAV3* gene were selected from coding regions, mRNA UTR and intronic region. Out of total 377 SNPs, 288(76.8%) non coding SNPs, 53 (14.1%) nsSNPs and 4(1.1%) mRNA UTRs (Fig. 1). Forty nine SNPs from the nsSNPs were selected for further analysis in the study; these 49 SNPs were encoding missense mutations, while the remaining 4 nsSNPs were of nonsense and frameshift mutations.

49 nsSNPs were independently submitted to the SIFT software to check for tolerance index (TI); Out of the 49 nsSNPs 35(71.4%) predicted to be damaging while 14 (29.6%) predicted to be tolerated (Table 1).

All 49 nsSNPs, were also submitted to PolyPhen 2.0 server, from a total of 49 nsSNPs submitted to PolyPhen 2.0, 39 (79.6%) predicted to be probably damaging while the remaining 10 (20.4%) expected to have a benign or possibly damaging effect.

The same nsSNPs were submitted to I-Mutant 2 tool to find possible influence on the stability of the

corresponding protein, 45 (91.8%) nsSNPs predicted to have a negative ΔΔG indicating that mutated protein is less stable than the wild type.

All the nsSNPs of *CAV 3* gene which were tested by SIFT, PolyPhen2 and I-Mutanty 2.0 were also analyzed by PANTHER. This further analysis provide an extra level of complexity to enhance SNP prediction, 34(69.4%) score less than -3 which was identified as a cutoff point for the functional significance.

Table 1. Results of missense encoding nsSNPs from SIFT, PolyPhen 2, PANTHER and I-Mutant-2.

rsID	AA substitution	SIFT (IT scores)	PolyPhen 2 (Confidence)	PANTHER (SubSPEC)	I-MUTANT 2 (DDG)
rs375301072	M2T	damaging (0.03)	possibly damaging (0.908)	-2.4721	-0.66
rs139786391	E10K	tolerated (0.13)	possibly damaging (0.511)	-2.23333	0.02
rs121909281	V14I	tolerated (0.72)	benign (0.011)	-0.34402	0.27
rs116840778	R27Q	damaging (0)	probably damaging(1)	-5.19899	-1.76
rs199476324	R27G	damaging(0)	probably damaging(1)	-5.46901	-0.72
rs116840782	D28E	damaging(0)	probably damaging(0.996)	-4.45566	-1.3
rs116840785	P29T	damaging(0)	probably damaging(1)	-5.60378	-0.21
rs116840786	P29H	damaging(0)	probably damaging(1)	-6.56414	-1.57
rs1008642	N33K	damaging (0.01)	probably damaging(0.997)	-5.11421	-0.72
rs199476325	E34K	damaging(0.04)	possibly damaging(0.593)	-1.69988	-1.77
rs374523166	V37L	tolerated(0.05)	benign(0.156)	-2.55024	-0.79
rs137901165	E42A	tolerated(0.09)	probably damaging (0.967)	-5.20256	-0.94
rs116840788	V44E	damaging(0)	probably damaging(1)	-4.2147	-1.32
rs116840773	A46E	damaging(0)	probably damaging(1)	-3.77965	-0.79
rs116840789	A46T	damaging(0)	probably damaging(1)	-3.06086	-0.73
rs116840793	E47K	damaging(0)	probably damaging(1)	-3.37526	-1.58
rs199476327	E47A	damaging(0)	probably damaging(1)	-3.52586	-1.56
rs199476328	E47D	damaging(0)	probably damaging(1)	-2.86875	-1.06
rs116840794	S53G	tolerable(0.08)	probably damaging(1)	-4.52524	-2.18
rs199476326	S53N	damaging(0)	probably damaging(1)	-5.13696	-0.57
rs72546667	G56S	tolerable(0.53)	probably damaging(0.999)	-3.78648	-1.68
rs116840795	V57M	damaging(0)	probably damaging(0.998)	-3.15197	-3.49
rs199476329	V57G	damaging(0)	probably damaging(0.99)	-3.43917	-5.73
rs199476330	W58R	damaging(0)	probably damaging(1)	-6.29299	-2.28
rs116840796	S61R	damaging(0)	probably damaging(1)	-5.21006	-1.2
rs116840798	T63P	tolerated(0.17)	probably damaging(0.983)	-2.87321	-2.04
rs121909280	T64S	tolerated(0.09)	probably damaging(0.967)	-2.94877	-0.95
rs199476332	T64P	damaging(0.01)	probably damaging(1)	-4.01182	-1.61
rs116840776	C72W	tolerated(0.09)	possibly damaging(0.496)	-5.16295	0.24
rs199476334	Y73C	damaging(0)	probably damaging(1)	-6.58105	-0.46
rs201893621	R74H	damaging(0.02)	probably damaging(1)	-4.15058	-1.45
rs72546668	T78K	tolerated(0.16)	possibly damaging(0.696)	-2.12892	-0.39
rs121909282	L79R	damaging(0.01)	probably damaging(0.961)	-3.31539	-1.32

rs112626848	V82I	tolerated(1)	benign(0.008)	-0.38369	-0.79
rs137881434	P83S	damaging(0)	probably damaging(1)	-5.47056	-0.86
rs104893715	A85T	damaging(0)	probably damaging(0.998)	-2.59566	-1.01
rs116840801	L86P	damaging(0)	probably damaging(0.999)	-3.95687	-1.06
rs28936685	L87P	damaging(0.02)	probably damaging(0.987)	-3.95687	-1.06
rs376624103	G89S	damaging(0.01)	probably damaging(1)	-3.65892	-1.05
rs28936686	A93T	damaging(0)	probably damaging(1)	-3.171	-1.2
rs104893714	F97C	tolerated (0.14)	probably damaging(0.97)	-4.9388	-1.06
rs199476336	I100F	damaging(0)	probably damaging(1)	-4.70411	-2.31
rs199476337	W101R	damaging(0)	probably damaging(1)	-6.00595	-1.47
rs116840805	P105L	damaging(0)	probably damaging(1)	-6.2061	-0.97
rs116840777	R126H	tolerated(0.66)	benign(0.004)	-0.93677	-1.65
rs201267913	A134V	damaging(0)	possibly damaging(0.81)	-1.96738	0
rs104893713	S141R	damaging (0)	probably damaging(0.988)	-2.93498	-1.11
rs142475018	V145M	damaging(0.03)	probably damaging(0.997)	-3.15197	-1.3
rs140575619	R148Q	tolerated(0.41)	benign(0.083)	-2.94102	-1.26

Compiling different computational methods can enhance the reliability of the prediction process of deleterious nsSNPs. That's why, different sequence and structure homology based programs were collective used to significantly enhance the prediction performance as shown in Fig. 2; the results showed that I-Mutant 2.0 predicted 91.8% of nsSNPs as deleterious, while PolyPhen, SIFT and PANTHER predicted 79.6%, 71.42% and 69.4% of ns SNP to be deleterious respectively. The noted differences in the prediction outcomes by the used programs are mainly because of the differences in strategies and sequences for alignment utilized by each software. Collectively, the data indicates that combining information from different tools will substantial increase in the predictive power to determine those nsSNPs with a functional impact on the protein.

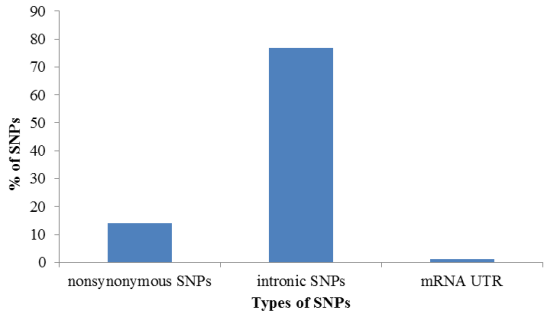

Fig. 1. Distribution of Intropnic, nonsynonymopus and mRNA UTR SNPs of *CAV* 3 gene.

For further analysis on the possible effect of nsSNPs on the structure and function of caveolin 3 protein, one nsSNP (rs116840778) (R27Q) was chosen out of the 49 nsSNPs tested. rs116840778 was chosen based on two points; first; it was one of the nsSNPs that showed the highest deleterious and damaging values with all prediction tools used (SIFT, PolyPhen, I-Mutant 2 and PANTHER) (see table 1). Second, it is the most sited *CAV* 3 nsSNPs in the literature showing high correlation with cavoelin 3 related diseases (Figarella-Branger *et al.*, 2003; Fee *et al.*, 2004; Woodman *et al.*, 2004).

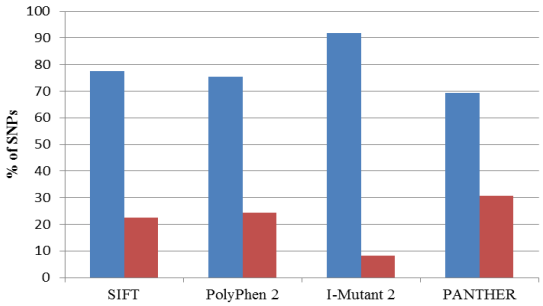

Fig. 2. Bar chart displaying the (%) of deleterious and benign nsSNPs predicted by SIFT, PolyPhen 2, I-Mutant 2 and PANTHER. Blue bar indicates (%) of nsSNPs found to be deleterious by SIFT and PANTHER, Probably damaging by PolyPhen 2 and decreased stability by I-Mutant 2, red bar indicates the (%) of nsSNPs tolerated by SIFT and PANTHER, benign or possible damaging by PolyPhen 2 and increase stability by I-Mutant 2.

To find out a possible effect of nsSNPs on caveolin 3 protein structure and function of, and since the cavoelin 3 protein structure is not available in PDB data base, the structure of cavoelin 3 was modeled using I-TASSER which uses the most effective five models optimize the C-score for each specific model. These processes were selected from the top 10 templates applied to the threading. The best C score obtained from our models was -3.42 for the wild type. The structure obtained is shown in Fig. 3. The C score for the structure obtained by I-TASSER are relatively low indicating a low confidence of prediction. Accordingly, the structure found not reliable to be used to predict the structure of the mutant.

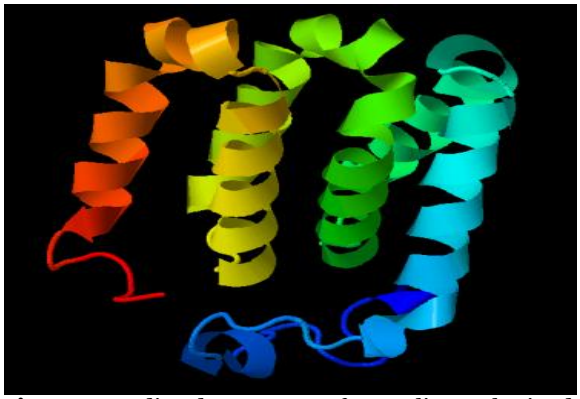

Fig. 3. Predicted structure of cavoelin 3 obtained from I-TASSER (C- Score is -3.42).

To find out what possible impact of (rs116840778) (R27Q) on cavoelin 3 structure, How yOur Protein Explained (HOPE)(Venselaar *et al.*, 2010) tool was used. This tool was developed at the Centre for molecular and bio-molecular informatics (CMBI), department of bioinformatics, Radboud University. HOPE evaluates the effect of the mutation on the following features: chemical interactions made by the mutated residue, the position of mutated residue in structural domains, the type of modifications on this residue (e.g. substituting polar amino acid with a nonpolar one) and available reported variants for this residue.

rs116840778 encodes a mutation in which Arginine is replaced by Glutamine at position 27. Each of them has its own properties such as charge, size, and hydrophobicity. Usually, the original wild-type amino acid and the introduced mutant amino acid show differences in these properties. The native amino acid was positively charged while the mutant amino acid is neutral, consequently, with this substitution the charge of the native amino acid will be lost, and this can lead to the loss of interactions between the mutated amino acid with other molecules or residues. In addition, both amino acids differ in size, the mutant amino acid is smaller, and this might drastically effect the interactions as well.

In addition, relying on HOPE project analysis.The wild-type amino acid is very conserved, however, a few other amino acids have been detected at the same position, but in other homologous sequences, the mutant residue nor any other amino acids with similar properties was reported at this position. Based on conservation scores this mutation is probably damaging to the protein. And the mutant amino acid is located in a highly conserved region.

In conclusion, the cav3 gene was investigated in this study by analyzing the effect of functional SNPs through computational methods, out of 377 SNPs, 49 SNPs were nonsynonymous encoding missense mutation. Those mutations were tested for their possible effect on the structure as well as the function on caveolin-3 and the results are summarized in table 1. Even more, one nsSNPs rs116840778 (R27Q) was tested for its possible effect on the protein structure and functions and it showed a high possible functional and structural influence on protein structure and function.

References

Adzhubei I, Jordan DM, Sunyaev SR. 2013. Predicting functional effect of human missense mutations using PolyPhen-2. Current Protocols in Human Genetics **Chapter 7**, Unit7 20. http://dx.doi.org/10.1002/0471142905.hg0720s76

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. Nature Methods **7**, *248-249*. http://dx.doi.org/10.1038/nmeth0410-248

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research **25**, 3389-3402.

Barroso I, Gurnell M, Crowley VE, Agostini M, Schwabe JW, Soos MA, Maslen GL, Williams TD, Lewis H, Schafer AJ, Chatterjee VK, O'Rahilly S. 1999. Dominant negative mutations in human PPARgamma associated with severe insulin resistance, diabetes mellitus and hypertension. Nature **402**, 880-883. http://dx.doi.org/10.1038/47254

Bava KA, Gromiha MM, Uedaira H, Kitajima K, Sarai A. 2004. ProTherm, version 4.0: thermodynamic database for proteins and mutants. Nucleic Acids Research **32**, *D120-121*. http://dx.doi.org/10.1093/nar/gkh082

Collins FS, Brooks LD, Chakravarti A. 1998. A DNA polymorphism discovery resource for research

on human genetic variation. Genome Research **8**, *1229-1231*.

Dryja TP, McGee TL, Hahn LB, Cowley GS, Olsson JE, Reichel E, Sandberg MA, Berson EL. 1990. Mutations within the rhodopsin gene in patients with autosomal dominant retinitis pigmentosa. New England Journal of Medicine **323**, *1302-1307*.
http://dx.doi.org/10.1056/NEJM199011083231903

Fee DB, So YT, Barraza C, Figueroa KP, Pulst SM. 2004. Phenotypic variability associated with Arg26Gln mutation in caveolin3. Muscle & Nerve **30**, 375-378.
http://dx.doi.org/10.1002/mus.20092

Figarella-Branger D, Pouget J, Bernard R, Krahn M, Fernandez C, Levy N, Pellissier JF. 2003. Limb-girdle muscular dystrophy in a 71-year-old woman with an R27Q mutation in the CAV3 gene. Neurology **61**, 562-564.

Galbiati F, Razani B, Lisanti MP. 2001. Caveolae and caveolin-3 in muscular dystrophy. Trends in Molecular Medecine **7**, 435-441.

Lander ES. 1996. The new genomics: global views of biology. Science **274**, 536-539.

Mi H, Guo N, Kejariwal A, Thomas PD. 2007. PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. Nucleic Acids Research **35**, *D247-252*.
http://dx.doi.org/10.1093/nar/gkl869

Mi H, Muruganujan A, Thomas PD. 2013. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Research **41,** D377-386.
http://dx.doi.org/10.1093/nar/gks1118

Ng PC, Henikoff S. 2003. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Research **31**, 3812-3814.

Parton RG. 1996. Caveolae and caveolins. Current Opinion in Cell Biology **8**, 542-548.

Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols **5,** *725-738*.
http://dx.doi.org/10.1038/nprot.2010.5

Scherer PE, Okamoto T, Chun M, Nishimoto I, Lodish HF, Lisanti MP. 1996. Identification, sequence, and expression of caveolin-2 defines a caveolin gene family. Proceedings of the National Academy of Sciences of the United States of America **93**, 131-135.

Smith EP, Boyd J, Frank GR, Takahashi H, Cohen RM, Specker B, Williams TC, Lubahn DB, Korach KS. 1994. Estrogen resistance caused by a mutation in the estrogen-receptor gene in a man. New England Journal of Medicine **331**, *1056-1061*.
http://dx.doi.org/10.1056/NEJM199410203311604

Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A. 2003. PANTHER: a library of protein families and subfamilies indexed by function. Genome Research **13,** 2129-2141.
http://dx.doi.org/10.1101/gr.772403

Thomas R, McConnell R, Whittacker J, Kirkpatrick P, Bradley J, Sandford R. 1999. Identification of mutations in the repeated part of the autosomal dominant polycystic kidney disease type 1 gene, PKD1, by long-range PCR. American Journal of Human Genetics **65,** *39-49*.
http://dx.doi.org/10.1086/302460

Venselaar H, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. 2010. Protein structure analysis of mutations causing inheritable

diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics **11**, *548*. http://dx.doi.org/10.1186/1471-2105-11-548

Williams TM, Lisanti MP. 2004. The Caveolin genes: from cell biology to medicine. Annals of Medicine **36,** 584-595.

Woodman SE, Sotgia F, Galbiati F, Minetti C, Lisanti MP. 2004. Caveolinopathies: mutations in caveolin-3 cause four distinct autosomal dominant muscle diseases. Neurology **62,** 538-543.

Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics **9,** *40*. http://dx.doi.org/10.1186/1471-2105-9-40

Zhang Y, Skolnick J. 2004. Scoring function for automated assessment of protein structure template quality. Proteins **57**, 702-710. http://dx.doi.org/10.1002/prot.20264