BRIEF REPORT

Human Mutation

OFFICIAL JOURNAL
HGVS
HUMAN GENOME VARIATION SOCIETY
www.hgvs.org

Protein–Protein Interaction Sites are Hot Spots for Disease-Associated Nonsynonymous SNPs

Alessia David, Rozami Razali, Mark N. Wass*,†, and Michael J.E. Sternberg*,†

Centre for Integrative Systems Biology and Bioinformatics, Division of Molecular Biosciences, Department of Life Sciences, Imperial College London, London SW7 2AZ, UK

Communicated by Mauno Vihinen

Received 16 August 2011; accepted revised manuscript 31 October 2011.
Published online 9 November 2011 in Wiley Online Library (www.wiley.com/humanmutation).DOI: 10.1002/humu.21656

ABSTRACT: **Many nonsynonymous single nucleotide polymorphisms (nsSNPs) are disease causing due to effects at protein-protein interfaces. We have integrated a database of the three-dimensional (3D) structures of human protein/protein complexes and the humsavar database of nsSNPs. We analyzed the location of nsSNPS in terms of their location in the protein core, at protein-protein interfaces, and on the surface when not at an interface. Disease-causing nsSNPs that do not occur in the protein core are preferentially located at protein-protein interfaces rather than surface noninterface regions when compared to random segregation. The disruption of the protein-protein interaction can be explained by a range of structural effects including the loss of an electrostatic salt bridge, the destabilization due to reduction of the hydrophobic effect, the formation of a steric clash, and the introduction of a proline altering the main-chain conformation.**
Hum Mutat 33:359–363, 2012. © 2011 Wiley Periodicals, Inc.

KEY WORDS: **nonsynonymous SNPs; protein structure; interactome; bioinformatics.**

The development of affordable techniques for sequencing the human genome and international efforts, such as the 1000genome and HAPmap projects [Stranger et al., 2011] are generating vast amounts of new data on common single nucleotide polymorphisms (SNPs). This information is being used to provide insight into disease susceptibility, particularly via genome wide association studies (GWAS). A major challenge is to identify the mechanisms by which SNPs influence the functions of genes and proteins. Particularly problematic is the understanding of complex disorders, such as cancer and cardiovascular diseases, which generally result from the cumulative effect of multiple common SNPs and the environment.

Here we focus on genetic variations causing amino acid changes (nonsynonymous SNPs, nsSNPs) and their effect on protein structure and function. Knowledge of the 3D structure of the protein in which the nsSNP occurs can provide valuable insights into the effects of the mutation on the conformation and function of the protein [Chambers et al., 2009]. Consequently, several groups have analyzed the relationships between nsSNPs and their location in a folded protein (e.g., [Burke et al., 2007; Wang and Moult, 2001; Yue and Moult, 2006]. However, proteins do not function in isolation and therefore it is vital to follow a systems biology approach and consider the effect of any nsSNP in terms of the mechanism by which the mutated protein interacts with others [Gong et al., 2011; Schuster-Bockler and Bateman, 2008; Teng et al., 2009; Zhang et al., 2010]. We now have experimental data on protein interactions both in terms of identifying the interacting partners (the interactome) and having 3D structures for complexes. Our analysis shows that disease-causing nsSNPs that do not occur in the protein core are located at protein-protein interfaces rather than surface noninterface regions compared to chance segregation between these two regions.

We started with the manually curated Human Interactome Database developed by the Structural Bioinformatics and Network Biology group at the Institute for Research in Biomedicine, Barcelona, Spain kindly provided by Dr. P. Aloy [Stein et al., 2011]. In this database, the interacting partners were obtained from experimentally derived protein-protein interactions from the major databases (see Supporting Information). Additional human protein-protein interactions are inferred from experimentally observed interaction between orthologous proteins in another species [Yu et al., 2004]. Structures were obtained from human complexes in the RCSB protein data bank (PDB) [Berman et al., 2000]. In addition, structures of homologous complexes are included where the human structure is not available. To ensure any model for a complex based on a homologous structure was reliable, we required that the sequence identity between the sequence of the homologue in the PDB and UniProt sequence of the human protein was >30%. In addition, to ensure that we have a model for most of the protein chain, for example, we are not ignoring other domains in the chain, we require that (1) at least 80% of the sequence in the PDB is covered by the UniProt sequence and (2) 80% of the UniProt sequence is covered by the PDB. This resulted in a set of 1,027 human proteins for which there is at least one structure for an interaction with another protein.

Interface residues were identified using the biological unit PDB files of the protein complexes. Residues were classed as part of the interface if one of their atoms was within 5 Å of an atom of a residue from the other protein [Mendez et al., 2003]. Interfaces consisting of fewer than five residues were excluded from the analysis. This cut off is based on the observation that biological protein-protein interactions are often mediated by a surprisingly few contacts [Kinjo and Nakamura, 2010]. Residue accessible surface was calculated using DSSP (Define Secondary Structure of proteins) [Kabsch and

Additional Supporting Information may be found in the online version of this article.

†These authors contributed equally to this work and should be considered joint last authors

*Correspondence to: Michael J.E. Sternberg. E-mail: m.sternberg@imperial.ac.uk or Mark N. Wass. E-mail: mark.wass04@imperial.ac.uk

Contract grant sponsor: MRC on the Advanced Course Masters Award (to A.D.); Majlis Amanah Rakyat (MARA) Malaysia (to R.R.); BBSRC (BB/F02048/1 to M.N.W.).

Sander, 1983]. A residue was defined as "buried" (core residue) when its solvent accessible surface area was <5 Å^2.

A list of nsSNPs identified in human proteins was retrieved from the Swiss-Prot/Uniprot humsavar database (http://www.uniprot.org/docs/humsavar, version released on October 8, 2010) [Yip et al., 2008]. nsSNPs are classified according to the Swiss-Prot/Uniprot humsavar database entry, which can have misclassifications: The three types of nsSNPs are:

1) "Disease causing", based on the disease classification in Uniprot. Apart from a very few exceptions, each disease variant in humsavar has an entry in the Online Mendelian Inheritance in Man Database (OMIM) [Amberger et al., 2011]. These nsSNPs were checked further against the Single Nucleotide Polymorphism Database (dbSNP) database (http://www.ncbi.nlm.nih.gov/projects/SNP/) [Sherry et al., 2001]. nsSNPs with a dbSNP id and a known minor allele frequency (MAF) were further investigated through a literature search in order to assess their disease-causing effect. This was done to avoid inclusion in the disease category of nsSNPs statistically associated with a particular disease but not representing a real cause of the disease, such as tag nsSNPs identified in GWAS. SNPs causing monogenic disorders are, in fact, expected to be predominantly at low frequencies in the population.
2) "Unclassified", when an nsSNP was identified in a pathological sample, but a statistical or experimental evidence was lacking to establish a disease-causing effect.
3) "Neutral polymorphisms" which are nsSNPs that are not present in OMIM. However, OMIM does not provide a complete list of disease-causing variants so some nsSNPs classed as "neutral" or "unclassified" may actually be disease causing. However, in the version of humsavar analyzed 36,549 nsSNPs were classed as polymorphism of which 32,939 have dbSNP entries showing that most "neutral" nsSNPs are not rare disease-causing mutations. The impact of misclassification of nsSNPs is discussed later.

nsSNPs were mapped onto the protein structures. Within our dataset, 537 proteins had at least one nsSNP (median number of nsSNPs = 2, range: 1–257). The total number of nsSNPs for our protein dataset was 4315, of which 1071 (24.8%) occurred in the protein core, 1264 (29.3%) in interfaces, and 1980 (45.9%) at the surface. The median number of interacting partners is 2 (range 1–37). Supp. Figure S1 shows the distribution of the fraction of noncore accessible surface occupied by interfaces; the median being 36% (range 4%–100%). For many proteins, we do not have structures or models for all the protein-protein interactions it is involved with. Thus for several proteins, we expect that some residues identified as surface are in fact at an interface.

Wild-type protein structures were visualized using the Pymol visualization program (http://www.pymol.org/) and PDB structural files. When multiple PDB entries were available for the same protein, the PDB entry with the highest sequence identity, coverage, and PDB-template/UniProt-target score was used.

The preference for an nsSNP to be in region *i* rather than region *j* is expressed by an odds ratio (OR_{ij}), where

$$OR_{ij} = \frac{x_i/(1-x_i)}{x_j/(1-x_j)}$$

where x_i is the probability of observing an nsSNP in region *i* and is

$$x_i = n_i / N_i$$

where n_i is the number of nsSNPs observed in region *i* and N_i is the total number of residues in region *i*. A two-tailed *P*-value of less than 0.05 was considered indicative of statistical significance of a preference for nsSNPs to be in one region over another compared to a random distribution based on the number of residues in the regions. Statistical analysis was performed using the statistical packages in R version 2.12.1 (http://www.r-project.org/).

First, we examined disease-causing nsSNPs. Of the 2,420 disease nsSNPs, 781 are in the core, 620 in the interface, and 1,019 in the surface (Table 1). Thus many nsSNPs are disease-causing due to their effect on protein interactions [Schuster-Bockler and Bateman, 2008]. Furthermore, as our structural information on protein-protein interactions is far from complete, many of the nsSNPs assigned here to the surface may actually be at an interface. Thus, we are underestimating the importance of disrupting protein-protein interaction as an explanation of many of the mechanisms of disease causation.

We now compare the frequencies of the nsSNPs in the three regions to those expected by chance based on the total numbers of protein residues in these three regions. Preferences for nsSNPS to occur between any two regions are quantified by an OR. Compared to a chance distribution, nsSNPs are significantly more likely to

Table 1. Location of nsSNPs in Protein Complexes

	Total residues	Observed	Expected	O/E percent ratio		OR	95 percent CI	*P* value
Disease nsSNPs					Disease nsSNPs			
Core	42,342	781	516.1	1.5	Core vs. noncore	1.77	1.62–1.93	< 0.0001
Interface	43,381	620	528.8	1.2	Interface vs. surface noninterface	1.59	1.44–1.76	< 0.0001
Surface noninterface	112,805	1019	1375.1	0.7	Core vs. interface	1.30	1.17–1.44	< 0.0001
Total	198,528	2420						
Polymorphism nsSNPs					Polymorphism nsSNPs			
Core	42,342	175	241.6	0.7	Core vs. noncore	0.67	0.57–0.79	< 0.0001
Interface	43,381	294	247.6	1.2	Interface vs. surface noninterface	1.15	1.00–1.32	0.0436
Surface noninterface	112,805	664	643.8	1.0	Core vs. interface	0.61	0.50–0.73	< 0.0001
Total	198,528	1133						
Unclassified nsSNPs					Unclassified nsSNPs			
Core	42,342	115	162.5	0.7	Core vs. noncore	0.65	0.54–0.80	< 0.0001
Interface	43,381	350	166.5	2.1	Interface vs. surface noninterface	3.08	2.64–3.60	< 0.0001
Surface noninterface	112,805	297	433.0	0.7	Core vs. interface	0.33	0.27–0.41	< 0.0001
Total	198,528	762						

The left table gives: the total number of residues in each of the regions, the observed number of nsSNPs, the expected number based on a random distribution of residues, and the ratio of observed to expected. The right table gives the odds ratio (OR), 95% confidence interval (CI), and the *P*-value for a 2-tailed test that OR is different from 1.0.

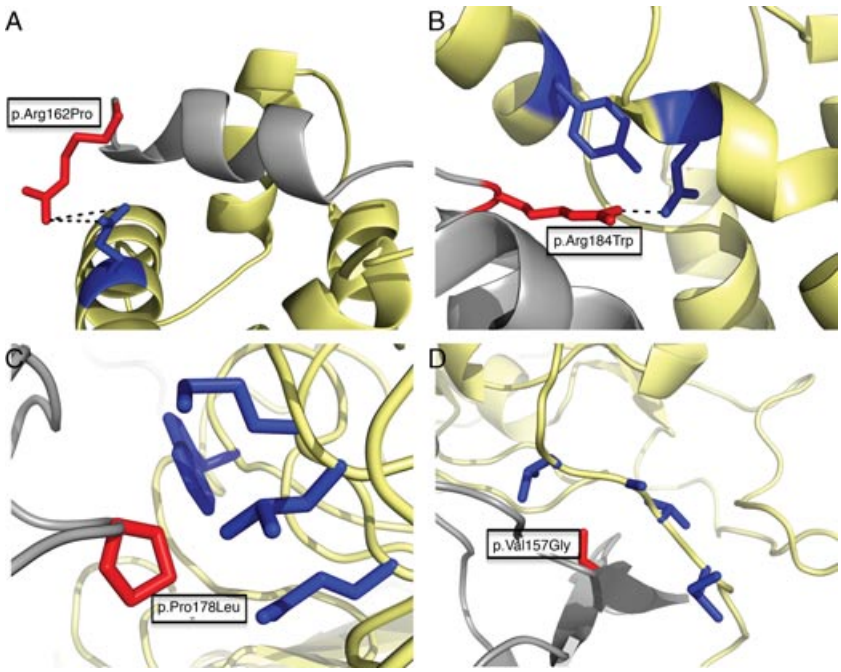

Figure 1. Structural changes induced by four disease-causing nsSNPs occurring at protein interfaces. The protein with the nsSNP is shown as a grey ribbon with the side chain that has the mutation shown as the wild-type in red. The interacting protein is shown in yellow with the side chains close to the variant residue in blue (see also the online Supporting Information for additional notes). Numbering corresponds to UniProt identifier. A: Cardiac troponin I protein (TnI) (Arg162Pro) interacting with troponin C (TnC). The Arg162Pro substitution would abolish a salt bridge with Glu 19 in Tnc. In addition, the φ,ψ main-chain angle of Arg126 (–145°, 360°) is not compatible with a proline and hence there will be a main-chain distortion. B: Connexin-26 (Cxs 26) (Arg184Trp) interacting with connexin 30 (Cxs 30). The Arg184Trp substitution causes loss of the salt bridge with Glu 47 in Cxs 30 and introduces a steric clash between Trp184 and the interacting residues on Cxs 30. C: Integrin β -2 protein (β-ITG, or CD18) (Pro178Leu) interacting with integrin α -X (CD11). One explanation is that the Pro178Leu substitution causes a steric clash between the side chain of Leu and interacting residues in CD11. Another possibility is that the favorable Pro-aromatic interaction is lost by the substitution to Leu. D: Electron transfer flavoprotein (α-ETF) (Val157Gly) interacting with β-ETF. The Val157Gly causes loss of the hydrophobic effect at the protein-protein interface, thus potentially resulting in energetic and conformational instability of the complex.

occur in the core rather than on the noncore (i.e., interface and surface, OR 1.77). This result accords with previous studies [Burke et al., 2007; Wang and Moult, 2001; Yue and Moult, 2006] and is expected since residues in a protein core generally are subject to more constraints than those not in the core. Since the majority of disease nsSNPs studied here are involved in the pathogenesis of monogenic disorders, these mutations are likely to have a marked effect on protein stability and/or function. Our study focuses on the nsSNPS at the interface and we found that compared to a chance distribution, disease SNPs are preferentially located at interfaces rather than on the surface (OR 1.59). There still remains a preference for nsSNPS to be in the core rather than at an interface (OR 1.29), but the OR is less than that for core versus the noncore.

Four disease-causing nsSNPs occurring at interfaces were analyzed in detail and their effects on protein interfaces are presented in Figure 1. In outline, we observe mutations that could lead to loss of a salt bridge (Fig. 1A and 1B), cause a steric clash or a loss of a favorable Pro-aromatic interaction (Fig. 1C) and reduction of the hydrophobic effect stabilizing the interface (Fig. 1D). However, conformational rearrangement of the protein interface could alleviate some of these adverse structural changes. Modeling of the energetics of disease-causing nsSNPs at interfaces using energy calculations suggests that often there is very little difference in the free energy of association between the wild-type and the disease-causing variant due to conformational rearrangements [Teng et al., 2009] [Zhang et al., 2010].

Second, we examined common polymorphisms. In contrast to disease nsSNPs, polymorphisms occur less frequently in the core than on the noncore (OR 0.67). This is consistent with the principle that the protein noncore accommodates changes in amino acid side chains more easily compared to the protein core and thus changes on the surface are less likely to be lethal for protein integrity and stability. There is a slight preference for common polymorphisms to be at the interface rather than the surface (OR 1.15) which is significant at the 5% but not the 1% level.

Third, we examined unclassified nsSNPs that are amino acid variants identified in pathological samples, nearly all of which were from tumor cells. However, it has yet to be clarified whether these mutations are disease causing. Unclassified nsSNPs were less likely to occur in the protein core compared to the noncore (OR 0.65). However, there was a strong preference for unclassified nsSNPs to be located at interfaces rather than on the surface (OR 3.08). This observation of a high OR for these nsSNPs to be at the interface while there is a low OR (0.65) for them to be in the core is surprising. Although unclassified nsSNPs represent a mixture of disease-causing and common variants, their extremely high occurrence in functionally important protein-protein interfaces suggests that many of them may be involved in the pathogenesis of disease. Indeed, since most of these unclassified nsSNPs are from cancer cells that are viable, the location of these nsSNPS at the interface might lead to modulation of function whereas if they were in the core they would be lethal. This provides a possible explanation why we find a greater tendency

for unclassified nsSNPs to be at interfaces as opposed to in the core or the surface.

The extent of stereochemical changes caused by the nsSNPs was evaluated by the change in the BLOSUM62 matrix core [Henikoff and Henikoff, 1993] for each mutation (Supp. Fig. S2). Almost half of the disease nsSNPs had scores below −1, suggesting that they were likely to have a damaging effect on the protein structure/function [Kumar et al., 2009]. Less than 6% of disease-causing nsSNPs had scores > 1. Irrespective of which region the nsSNP was located, disease-causing nsSNPs rarely have conservative changes. A far smaller fraction of polymorphisms had BLOSUM62 scores below −1, in keeping with the expectation that nearly all of these nsSNPs are not disease causing. The location of nsSNPs on the protein structure (core, interface, or surface) did not affect the distribution of BLOSUM62 scores ($P > 0.1$ using a 3×3 chi-squared test).

We have reported different distributions of nsSNPs at the core, interface, and surface and of BLOSUM62 scores for disease-causing variants, polymorphisms, and unclassified nsSNPs. However, some nsSNPs may have been erroneously excluded from being classified as disease variants. We expect that the frequency distribution of location and the frequency distribution of BLOSUM62 changes of nsSNPs that were erroneously omitted from the disease category will be similar to those of the correctly identified disease variants and different from those of the neutral and unclassified nsSNPs. Thus the effect of the misclassifications will be to reduce the differences in the distributions.

Our data show that, compared to a random distribution, nsSNPs on the surface of protein chain preferentially occur at protein-protein interfaces than on the remaining surface region. There are several structure-based approaches to predict the functional effect of nsSNPs considering information from protein structure (e.g., PolyPhen2 [Adzhubei et al., 2010], SNP3D [Yue et al., 2006], and SNAP [Bromberg and Rost, 2007] and reviewed in Jordan et al. [2010] and Teng et al. [2008]. Treatment of protein interactions via these approaches might well be improved by a consideration of the coordinates of the biological molecule rather than those deposited in the crystal structure that often does not identify the biological molecule [Xu et al., 2006]. In addition, one could include the prediction of protein-protein interactions from experimental structures of homologues. Accurate prediction of the effect of nsSNPS will prove a valuable tool for analyzing nsSNP data including selecting subsets of nsSNPs for retesting in GWAS.

References

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. Nat Methods 7:248–249.

Amberger J, Bocchini C, Hamosh A. 2011. A new face and new challenges for Online Mendelian Inheritance in Man (OMIM(R)). Hum Mutat 32:564–567.

Aranda B, Achuthan P, Alam-Faruque Y, Armean I, Bridge A, Derow C, Feuermann M, Ghanbarian AT, Kerrien S, Khadake J, Kerssemakers J, Leroy C, Menden M, Michaut M, Montecchi-palazzi L, Neuhauser SN, Orchard S, Perreau V, Roechert B, van Eijk K, Hermjakob H. 2010. The IntAct molecular interaction database in 2010. Nucleic Acids Res 38:D525–D531.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The Protein Data Bank. Nucleic Acids Res 28:235–242.

Bhattacharyya R, Chakrabarti P. 2003. Stereospecific interactions of proline residues in protein structures and complexes. J Mol Biol 331:925–940.

Bromberg Y, Rost B. 2007. SNAP: predict effect of non-synonymous polymorphisms on function. Nucleic Acids Res 35:3823–3835.

Burke DF, Worth CL, Priego EM, Cheng T, Smink LJ, Todd JA, Blundell TL. 2007. Genome bioinformatic analysis of nonsynonymous SNPs. BMC Bioinformatics 8:301.

Ceol A, Chatr Aryamontri A, Licata L, Peluso D, Briganti L, Perfetto L, Castagnoli L, Cesareni G. 2010. MINT, the molecular interaction database: 2009 update. Nucleic Acids Res 38:D532–D539.

Chambers JC, Zhang W, Li Y, Sehmi J, Wass MN, Zabaneh D, Hoggart C, Bayele H, McCarthy MI, Peltonen L and others. 2009. Genome-wide association study identifies variants in TMPRSS6 associated with hemoglobin levels. Nat Genet 41:1170–1172.

Chautard E, Ballut L, Thierry-Mieg N, Ricard-Blum S. 2009. MatrixDB, a database focused on extracellular protein-protein and protein-carbohydrate interactions. Bioinformatics 25:690–691.

Goll J, Rajagopala SV, Shiau SC, Wu H, Lamb BT, Uetz P. 2008. MPIDB: the microbial protein interaction database. Bioinformatics 24:1743–1744.

Gong S, Worth CL, Cheng TM, Blundell TL. 2011. Meet me halfway: when genomics meets structural bioinformatics. J Cardiovasc Trans Res 4:281–303.

Hamelmann C, Amedofu GK, Albrecht K, Muntau B, Gelhaus A, Brobby GW, Horstmann RD. 2001. Pattern of connexin 26 (GJB2) mutations causing sensorineural hearing impairment in Ghana. Hum Mutat 18:84–85.

Henikoff S, Henikoff JG. 1993. Performance evaluation of amino acid substitution matrices. Proteins 17:49–61.

Huang GY, Zheng CH, Wu YX, Wang W. 2010. Involvement of connexin 43 in the acupuncture effect of improving rat blastocyst implantation. Fertil Steril 93:1715–1717.

Indo Y, Glassberg R, Yokota I, Tanaka K. 1991. Molecular characterization of variant alpha-subunit of electron transfer flavoprotein in three patients with glutaric acidemia type II–and identification of glycine substitution for valine-157 in the sequence of the precursor, producing an unstable mature protein in a patient. Am J Hum Genet 49:575–580.

Ingles J, Doolan A, Chiu C, Seidman J, Seidman C, Semsarian C. 2005. Compound and double mutations in patients with hypertrophic cardiomyopathy: implications for genetic testing and counselling. J Med Genet 42:e59.

Jordan DM, Ramensky VE, Sunyaev SR. 2010. Human allelic variation: perspective from protein function, structure, and evolution. Curr Opin Struc Biol 20:342–350.

Kabsch W, Sander C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22:2577–2637.

Kinjo AR, Nakamura H. 2010. Geometric similarities of protein–protein interfaces at atomic resolution are only observed within homologous families: an exhaustive structural classification study. J Mol Biol 399:526–540.

Krivov GG, Shapovalov MV, Dunbrack RL. 2009. Improved prediction of protein side-chain conformations with SCWRL4. Proteins: Structure, Function, and Bioinformatics 77:778–795.

Kumar S, Suleski MP, Markov GJ, Lawrence S, Marco A, Filipski AJ. 2009. Positional conservation and amino acids shape the correct diagnosis and population frequencies of benign and damaging personal amino acid mutations. Genome Res 19:1562–1569.

Maeda S, Nakagawa S, Suga M, Yamashita E, Oshima A, Fujiyoshi Y, Tsukihara T. 2009. Structure of the connexin 26 gap junction channel at 3.5 A resolution. Nature 458:597–602.

Mendez R, Leplae R, De Maria L, Wodak SJ. 2003. Assessment of blind predictions of protein-protein interactions: current status of docking methods. Proteins 52:51–67.

Ohashi Y, Yambe T, Tsuchiya S, Kikuchi H, Konno T. 1993. Familial genetic defect in a case of leukocyte adhesion deficiency. Hum Mutat 2:458–467.

Prasad TS, Kandasamy K, Pandey A. 2009. Human protein reference database and human proteinpedia as discovery tools for systems biology. Methods Mol Biol 577:67–79.

Richard P, Charron P, Carrier L, Ledeuil C, Cheav T, Pichereau C, Benaiche A, Isnard R, Dubourg O, Burban M and others. 2003. Hypertrophic cardiomyopathy: distribution of disease genes, spectrum of mutations, and implications for a molecular diagnosis strategy. Circulation 107:2227–2232.

Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, Eisenberg D. 2004. The database of interacting proteins: 2004 update. Nucleic Acids Res 32:D449–D551.

Schuster-Bockler B, Bateman A. 2008. Protein interactions in human genetic diseases. Genome Biol 9:R9.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29:308–311.

Stark C, Breitkreutz BJ, Chatr-Aryamontri A, Boucher L, Oughtred R, Livstone MS, Nixon J, Van Auken K, Wang X, Shi X and others. 2011. The BioGRID interaction database: 2011 update. Nucleic Acids Res 39:D698–D704.

Stein A, Mosca R, Aloy P. 2011. Three-dimensional modeling of protein interactions and complexes is going 'omics. Curr Opin Struc Biol 21:200–208.

Stranger BE, Stahl EA, Raj T. 2011. Progress and promise of genome-wide association studies for human complex trait genetics. Genetics 187:367–383.

Takeda S, Yamashita A, Maeda K, Maeda Y. 2003. Structure of the core domain of human cardiac troponin in the Ca(2+)-saturated form. Nature 424:35–41.

Teng S, Madej T, Panchenko A, Alexov E. 2009. Modeling effects of human single nucleotide polymorphisms on protein-protein interactions. Biophys J 96:2178–2188.

Teng S, Michonova-Alexova E, Alexov E. 2008. Approaches and resources for prediction of the effects of non-synonymous single nucleotide polymorphism on protein function and interactions. Curr Pharm Biotechnol 9:123–133.

Toogood HS, van Thiel A, Basran J, Sutcliffe MJ, Scrutton NS, Leys D. 2004. Extensive domain motion and electron transfer in the human electron transferring flavoprotein.medium chain Acyl-CoA dehydrogenase complex. J Biol Chem 279:32904–32912.

Wang Z, Moult J. 2001. SNPs, protein structure, and disease. Hum Mut 17:263–270.

Wilcox SA, Saunders K, Osborn AH, Arnold A, Wunderlich J, Kelly T, Collins V, Wilcox LJ, McKinlay Gardner RJ, Kamarinos M and others. 2000. High frequency hearing loss correlated with mutations in the GJB2 gene. Hum Genet 106:399–405.

Xie C, Zhu J, Chen X, Mi L, Nishida N, Springer TA. 2010. Structure of an integrin with an alphaI domain, complement receptor type 4. EMBO J 29:666–679.

Xu Q, Canutescu A, Obradovic Z, Dunbrack RL. 2006. ProtBuD: a database of biological unit structures of protein families and superfamilies. Bioinformatics 22:2876–2882.

Yip YL, Famiglietti M, Gos A, Duek PD, David FP, Gateau A, Bairoch A. 2008. Annotating single amino acid polymorphisms in the UniProt/Swiss-Prot knowledgebase. Hum Mutat 29:361–366.

Yu H, Luscombe NM, Lu HX, Zhu X, Xia Y, Han JD, Bertin N, Chung S, Vidal M, Gerstein M. 2004. Annotation transfer between genomes: protein-protein interologs and protein-DNA regulogs. Genome Res 14:1107–1118.

Yue P, Melamud E, Moult J. 2006. SNPs3D: candidate gene and SNP selection for association studies. BMC Bioinformatics 7:166.

Yue P, Moult J. 2006. Identification and analysis of deleterious human SNPs. J Mol Biol 356:1263–1274.

Zhang Z, Teng S, Wang L, Schwartz CE, Alexov E. 2010. Computational analysis of missense mutations causing Snyder-Robinson syndrome. Hum Mutat 31:1043–1049.