Human SNP associations with genetically determined disorders

Avdyunina P.[1], Vlasov P.[2]

[1] *Lomonosov Moscow State University, Moscow, Russia*

E-mail: p.avdiunina@gmail.com

[2] *Centre for Genomic Regulation (CRG) and*

Universitat Pompeu Fabra (UPF),Barcelona, Spain

E-mail: peter.vlasov@crg.eu

Creating genome database and decoding individual genomes of people have been actively carried out all over the world since the decoding of the first human genome in 2000 [1] and up to the present time. In particular, one of the key research areas is the search for genomic variations and the identification of their clinical significance. Most genetic differences are represented by point mutations in the genome - single nucleotide polymorphisms (SNP), with some of them being able to determine susceptibility to specific diseases, individual immune responses to pathogens, therapeutic agents, metabolites, toxins, etc. Here we analyzed the effects of such polymorphisms arising in the sites of proteins-compounds binding, and also combine our predictions with ClinVar annotation records [2].

We obtained data on human SNPs from the 1000 genome project. After we selected all small molecular compound-protein structural complexes from PDB (Protein Data Bank) [3] in which the compound was identical or highly similar to FDA-registered drug and drug-like compounds and identified the binding sites of the protein in direct interaction with the ligand. We then combine data about 1000g SNPs and about protein-ligand binding sites to select polymorphisms with the potential impact on the binding properties. Using a docking approach we calculate the difference in the free energy of the drug and wild-type protein interaction and that of the drug with the protein sequencing incorporating the amino acid polymorphism. Finally, we combine our data with clinical annotation) using public archives of reports of the relationships among human variations and phenotypes. The search for SNP was based on the database of the project "1000 genomes"[4] and "The Cancer Genome Atlas"[5].

Among the 2504 available human genomes, our approach identified 53 SNPs that probably affect a drug-protein interaction with 30 FDA-approved drugs and 192 SNPs that possibly have this effect on 75 FDA-approved drugs, 390 SNPs probably affecting 165 FDA experimental drugs and 835 SNPs possibly having this effect on 357 FDA experimental drugs. For instance, among the SNPs from the binding sites with serious $\Delta\Delta G$ effect, mutations rs114468011, rs369382075 and rs558267822 were found in the protein Glycogen phosphorylase b (1GPB), which plays a key

role in the Glycogen storage disease, type V formation [6]. The results of the annotation require further analysis.

1. Venter J.C. [et al.]. The sequence of the human genome // Science. – 2001. – Vol. 291, №5507. – P. 1304–1351.

2. Landrum M, Lee JM, Riley G, Jang W, Rubinstein WS, Church DM, Maglott D. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014;42:D980–D985.

3. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H. et al. (2000). The Protein Data Bank. Nucleic Acids Res. 28, 235–242.

4. Birney E., Soranzo N. Human genomics: The end of the start for population sequencing // Nature. – 2015. – Vol. 526, №7571. – P. 52–53.

5. The Cancer Genome Atlas Research Network, Weinstein, J. N., Collisson, E. A., Mills, G. B., Mills Shaw, K. R., Ozenberger, B. A., Ellrott, K., Shmulevich, I., Sander, C., and Stuart, J. M. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. Nat. Genet. 45, 1113–1120.

6. Rubio JC et al. (2006) A proposed molecular diagnostic flowchart for myophosphorylase deficiency (McArdle disease) in blood samples from Spanish patients. Hum Mutat 28: 203–204