Chia Wan Ni Geraldine, Ng Gee Ling, Tay Kai Lin Victoria

Investigating the Role of Single Nucleotide Polymorphisms (SNPs) in Interethnic Differences in Statin Drug Response Pathways

Chia Wan Ni Geraldine, Ng Gee Ling, Tay Kai Lin Victoria

Hwa Chong Institution (College Section)

Abstract

An investigation on the genetic factor that is believed to influence interethnic differences in drug response is carried out. The focus is on the statin pathway looking at Single Nucleotide Polymorphims (SNPs). A combination of strong evidence from literature review and data mining, the investigation comprises of four perspectives namely polymorphism of the statin pathway through median SNP density values, population differentiated SNPs through fixation index values (F_{ST}), potentially functionality of SNPs and SNP associated with expressions. One candidate gene *SLCO1B3*, an organic anion transporter that is greatly involved in the statin related drug response pathways (DRP), population differentiated structural SNP, *rs3764006* and SNP rs60140950 which is associated with expression are highlighted in our investigations.

Introduction

Statins, also known as HMG-CoA reductase inhibitors, are a class of drugs that inhibit HMG-CoA reductase which plays a central role in synthesizing cholesterol. They are administered to lower cholesterol levels to prevent cardiovascular diseases. Statins enter the human's systemic circulation through passive and active transport facilitated by the ATP-binding cassette transporter (ABC) and solute carrier transporter (SLC) gene family transporters. The metabolism of this pathway is catalysed by cytochrom P450 (CYP) and UDP-glucuronosyltransferase (UGT) gene family enzymes. The ABC transporter is also responsible for mediated biliary excretion. [3]

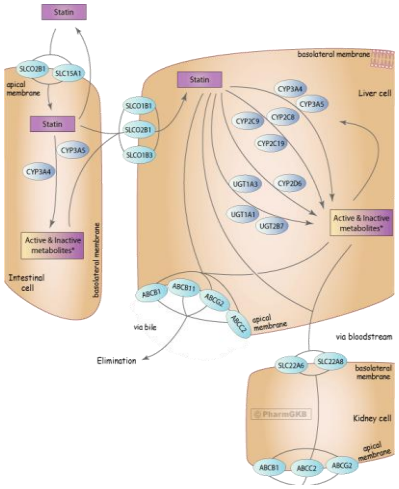

Figure 1 Pharmacokinetics of Statin Pathway. Adapted from: http://www.pharmgkb.org/pathway/PA145011108

Literatures suggest that specific genetic variation among populations contributes appreciably to differences in gene expression phenotypes after analyzing data from microarray on different global populations [5,6,7]. As such, this research paper proposes genetic polymorphisms e.g. single nucleotide polymorphisms (SNPs) may account for some population differences in responses to statin therapy. SNPs make up 90% of all human genetic variation and are believed to affect human's responses to drugs [14]. Hence, this investigation aims to give a comprehensive review on the role of genetics in drug responses in the statin pathway, addressing gaps between previous studies on this pathway. The hypothesis is that population differentiated regulatory SNPs play a major role in different statin drug responses by means of altering levels of expressions and functions of proteins essential in the

pharmacokinetics of the pathway.

Materials and Methods

The main databases for the 715 genes, 66 drug pathways and known single nucleotide polymorphisms investigated are the Pharmacogenomics Knowledge Database (pharmGKB) [8], National Centre for Biotechnology Information (NCBI) [9] Entrez Gene database, SNP (dbSNP) [10] respectively. These databases are used in relating the occurrence of SNPs to their respective genes and the genes that are involved in their respective drug pathways. The investigation of associating the genetic factor to differences across populations in statin drug responses is in four perspectives.

Polymorphism through SNP Density Calculation

Polymorphism contributes to pharmacokinetic variability of which includes the delivery of a drug or metabolites to the target molecules that are responsible in drug disposition. The relative polymorphisms of the drug response pathways were determined by the calculation of median SNP density values at the different genic regions, namely promoter (5kbp upstream and 500bp downstream transcription start site), 5'untranslated region, coding, intronic–and 3' untranslated region. The formula used in calculating the SNP density is given by: $SNP\ Density = \frac{total\ number\ of\ SNPs\ per\ region\ per\ gene}{length\ of\ respective\ region\ per\ gene}$. The median SNP density in different genic regions for all genes in each DRP was calculated and used to determine the polymorphism of each pathway.

Population Differentiation: F_{ST} Calculation

Population differentiated SNPs were used in determining the factor of race or ethnicity in the association of the allele with DRP. Fixation index (F_{ST}) was calculated for the SNPs under investigation with the standard formula: $F_{ST} = \frac{MSP-MSG}{MSP+(n_c-1)MSG}$ where n_c refers to the sample size in a population, MSG denotes the observed mean square errors for loci within the populations and MSP denotes the observed mean square errors for between populations [12]. F_{ST} is a fixation indexes that measures population differentiation, genetic distance [15], based on genetic polymorphism data which is single-nucleotide polymorphisms (SNPs) in our investigation.

The genotype data was obtained from HAPMAP [11] and Singapore Genome Variation Project (SGVP) [2]. There are a total of nine populations further organized into four population groups. The population groups are the *East Asians* which is made up of Chinese from Beijing (CHB) and Singapore (CHS) and the Japanese (JPT), the *South Asians* which is made up of Gujarati Indians from Houston (GIH) and Indians from Singapore (INS), the *Caucasians* which is made up of Caucasians from Utah (CEU) and Tuscan, Italy (TSI) and lastly, the *Africans* which is made up of Luhya, Kenya (LWK) and Yoruban, Nigeria (YRI).

Functional SNPs

As alterations in DNA sequence that affect RNA stability, protein function, or other cellular mechanisms, the investigation also looked at the potential functionality of SNPs to verify causality of SNPs for the drug response phenotype. The data were obtained from the pfSNP database. [13] The functional SNPs were categorized into three categories namely regulatory, structural and protein or coding. Regulatory SNPs affect transcription factor or microRNA bindings thus resulting in changes in expression. Structural SNPs affect splicing during post transcription hence, leading to changes in the RNA structure while protein or coding SNPs change the amino acid sequence which consequentially cause deleterious effects on protein structure and/or functions.

SNPs associated with Gene Expression

Lastly, we tested associations of SNPs for gene expressions. Expression-associated SNPs affect the expression of phenotypes, potentially that of a protein synthesized in the case of our study on drug responses. In the investigation of SNPs associated with expressions, the raw genotype data from individuals (43 Caucasians, European (CEU),59 Chinese Han, Beijing and Japanese, Tokyo (CHBJPT), 42 Yoruban, Nigeria(YRI)) were obtained from 1000 Genomes Project [19] and the raw gene expression data (in the same group individuals) was obtained from Gene Expression Omnibus (GEO) database under accession ID GSE6536. Both the genotype and gene expression data were correlated using linear regression tests and SNPs associated with expression showing $p<0.05$ after multiple test correction were identified as expression-associated SNP. Data mining was done via Microsoft Office Access and statistical tests were done in the investigation via the R software and Minitab.

Result

Relative Polymorphism of the Statin Pathway based on Genic Regions

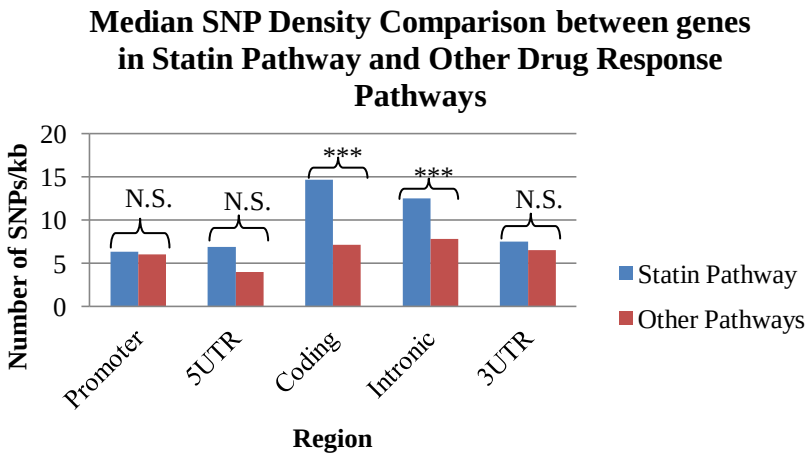

Figure 2: Genes involved in statin pathways showed significantly higher SNP densities in coding regions and introns as compared with all the other genes in the drug response pathways. ***indicates $p<0.001$ while N.S indicates a $p>0.05$ by the Mann-Whitney test after Bonferroni correction

In the investigation of the polymorphism of the statin pathway with other genes in DRPs, the SNP density of genes that are involved in each drug response pathway was calculated. The median was then calculated for the statin pathway based on the 46 genes that are involved in it. This was also done similarly for the remaining 669 genes, investigated in the project, for the calculation of the median.

The statin pathway is significantly

more polymorphic in these regions as compared to the other drug response pathways investigated. Polymorphism in the coding region offers much to investigate as it is likely to result in changes in the amino acid sequence which eventually affects the protein structure and thus, function. Proteins affected could potentially be enzymes or transporters that play a role in the statin pathway, particularly the pharmacokinetics of it since all of the statin related pathways under our study are involved in the pharmacokinetics.

Population Differentiation Genes that Occur in the Statin Pathway

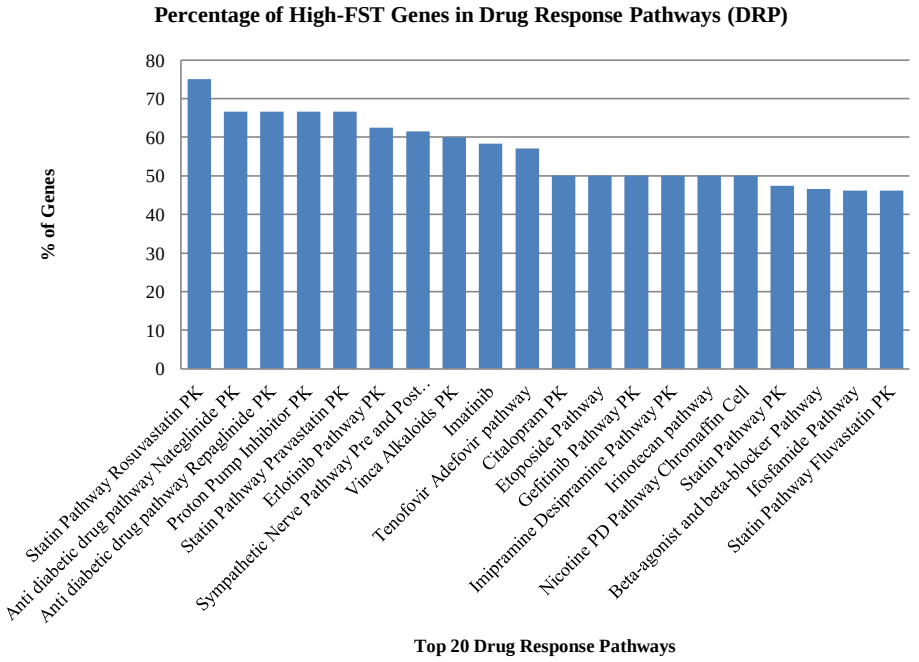

Figure 3: Stain-related pathways (Stain Pathway Rosuvastatin PK and Statin Pathway Provastatin Pathway PK) contain the highest percentage of high-FST genes among all the drug response pathways.

From the F_{ST} scores amongst all the genotyped SNPs in the human genome, we selected the SNPs with top 5% FST scores. The threshold was set at 0.3 which is a 95 percentile cut off at genomic level and they were termed as high-F_{ST} SNPs. The distribution of these population-differentiated SNPs in drug response genes was investigated. Given that genes containing at least one high-F_{ST} SNPs showed evidence of population differentiation, the distribution of these high-F_{ST} genes in the 66 drug response pathways were studied. Statin-related pathways were found to contain the highest proportions of high-F_{ST} genes. As seen from Figure 3, Statin Pathway Rosuvastatin PK and Statin Pathway Provastatin Pathway PK are amongst the 66 pathways with the highest percentage. These two pathways are specific drug response pathways of statin.

The genes involved in statin pathways that are highly differentiated amongst populations were identified and the functionality of SNPs in these genes were further studied. 68 unique SNPs occur in all the statin related pathways. Of which, there are 6 high F_{ST} SNPs involved in the statin pathways. There are 27 regulatory SNPs, 4 protein or coding SNPs and 1 structural SNP. In addition, there is only 1 SNP associated with expression in the statin pathways out of the 159 identified in all the drug-response genes. It is found to occur in SLCO1B3, which is a common gene involved in all the statin-related pathways.

Candidate Gene, SLCO1B3 with Candidate Population Differentiated Structural SNP and SNP associated with expression

Gene *SLCO1B3* (GeneID: 28234) was chosen as a candidate gene in our investigation as it has been found to be highly involved in all of the statin-related pathways, including Statin Pathway Atorvastatin, Lovastatin and Simvastatin PK, Statin Pathway Fluvastatin PK, Statin Pathway PK, Statin Pathway Pravastatin PK and Statin Pathway Rosuvastatin PK. This gene belongs to an organic anion transporter (OATP) family that is involved in the membrane transport of bile acids, conjugated steroids, thyroid hormone, eicosanoids, peptides, and numerous drugs in many tissues [16]. Within gene *SLCO1B3*, we have also chosen two candidate SNPs that are interesting in the statin pathway from the stated perspectives.

Structural SNP1 (*rs3764006)* has a high F_{ST} score of 0.422 across all populations. SNP1 resides 1833 base pairs downstream of the transcription start site of gene *SLCO1B3* in the coding region and is predicted to alter a splicing regulatory element (enhancer) and result in alternative splicing patterns. In addition to its potential functionality, SNP1 also showed evidence of population differentiation. The greatest pairwise score is between the South Asians and African Blacks with 0.704 (to 3 decimal places). The genetic variant here may play a role in the observed variation in differential drug responses potentially across populations [17].

Second candidate SNP2 (*rs60140950),* associates with SLCO1B3 gene expression in the CEU population. It's a non-synonymous SNP, leading change from amino acid glysine to alanine. It was predicted by Polyphen (Polymorphism Phenotyping) to be deleterious to protein function. Due to its effects on protein structure and expression, SNP2 is thus worth further investigating upon.

Discussion

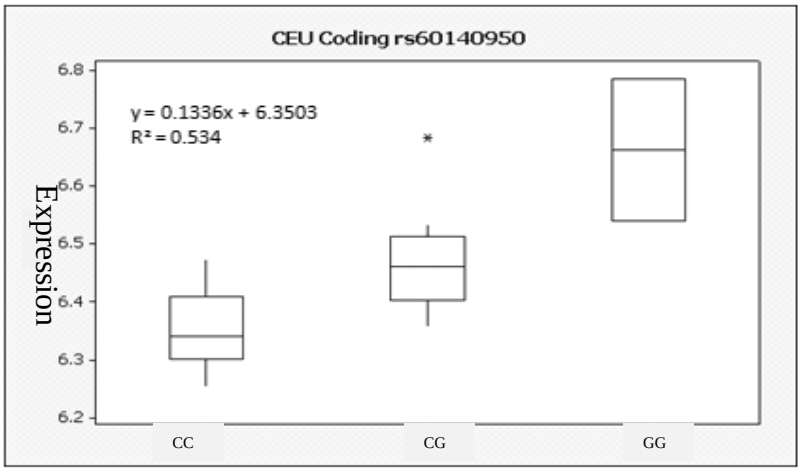

Figure 4: By the linear regression statistical test, it can be concluded that rs60140950 shows significant association with gene *SLCO1B3*

In a study by Suet N. Chen, it was verified that a polymorphism, E670G cSNP is a significant determinant of plasma low density lipoproteins – cholesterol (LDL-C) levels in response to statin therapy [20]. Our investigations have also showed that the statin pathway is relatively polymorphic as well, in the coding and intronic region particularly. Polymorphism in the coding region gives much interest to further studies. Genetic variants present in the coding region affect the protein structure or function. A dysfunctional protein, such as an enzyme or transporter, produced may then affect the

CC CG

pharmacokinetics of the statin pathway leading to varying drug response. Hence, further conclusive studies could be greatly emphasized on the specific mechanisms and functions affected by the polymorphisms in the statin pathway. In another study, it stated that different frequencies of a SNP *rs4363657,* within gene *SLCO1B1*, C variant allele and homozygous genotype CC were observed between the Americans and African descendents. Gene *SLCO1B3* investigated in this project belongs to the same family of organic anion transporter with *SLCO1B1* gene in the past study. Both genes belong to the organic anion transporter (OATP) family that is involved in the membrane transport of bile acids, conjugated steroids, thyroid hormone, eicosanoids, peptides, and numerous drugs in many tissues. [16] As seen from figure 1, the transporter gene family is heavily involved in the pharmacokinetics of the statin pathway. It plays a crucial role in transporting the statin related drugs after absorption in the intestines across the membrane of the liver cells for assimilation. Given that drug uptake is now commonly known to be transporter-mediated, an uptake transporter (such as the candidate gene in our study) thus plays a major role in drug response. [21] For instance, drug-membrane interaction can change the rate of entrance of the drug to reach a specific target cell organelle or system [18], further emphasizing the importance of genes in this family such as *SLCO1B3*.

Candidate SNP1 within *SLCO1B3* is also highlighted in the study. SNP1 alters the DNA sequence of exon splicing enhancer (ESE). Moreover, one allele is prevalent in the African populations while the other is more common in the South Asian populations (pairwise F_{ST} score = 0.704, to 3 decimal places). Changes from one allele to another may lead to an alternative splicing isoform. ESEs are binding sites for a family of serine/argentine rich proteins (splicing factors) that influence differential splicing. This allows pre-mRNAs to be spliced into multiple forms of matured mRNA, thus giving rise to various protein isoforms. The presence of SNPs thus further increases the variation of splice sites giving rise to even varying matured mRNA molecules [1]. Consequentially, different forms of protein involved as transporters across the liver cell membrane would be synthesized. This could potentially explain for varying responses across populations in the statin drug pathways due to different drug uptake activities.

SNP2 also holds a great significance to further studies of the statin drug pathway. Being a protein or coding SNP that causes change in amino acid sequence, SNP2 is deleterious, and a dysfunctional protein could potentially be synthesised as a result. Moreover, SNP2 occurs in the coding region, suggesting that change potentially resides in a functionally important site of the three dimensional structure of the protein synthesized. Possible implications could be that the amino acid substitution damages the hydrophobic core of the protein or the electrostatic interactions [4] that the protein may have with the drug molecules when transporting them across the liver cell membranes.

One limitation in our study is that we have only investigated the individual effects of SNPs. Thus we have evidently ruled out the possibility that it could be a combination of SNPs that exert the effect on the

pharmacokinetics of the statin pathway hence giving rise to differential responses. Moreover, within our study, we have used a conservative and stringent multiple test correction on our statistical tests results. This may have not allowed us to capture other potential candidate SNPs that are regulatory SNPs, hence limiting our range of candidate SNPs. Also, given that the results from Gene Expression Omnibus (GEO) used in our investigation of SNPs associated with expressions measured gene expression on the lymphoblastoid cell line, LCL, this may give a certain level of ambiguity. From literature search, the target organ in the statin pathway is highly to be the liver. Hence, with a difference in the target organs, this indicate that the differences in gene expression shown for the LCL cell line may not necessarily be translated in the liver or even any other organs involved in the statin. Lastly, our F_{ST} results were taken from HAPMAP, which has only tested a fraction of the total genomic SNPs suggesting that not all SNP data has an F_{ST} score. Thus there may be some SNPs that were unable to be identified as candidate SNPs due to the lack of SNP genotype data to calculate F_{ST} scores.

The main future work that is proposed in this study is validation. This could be achieved via vector cloning or protein structure modelling to determine the mechanisms through which SNPs affect the protein structure and if the regions that they affect are functionally important. SNPs that have been validated to be functionally important is crucial to affirming all bioinformatics prediction. Hence, validation is necessary for continuation of the investigation on the candidate gene and SNPs.

Conclusion

In terms of polymorphism and population differentiation, it could be concluded that the statin pathway is highly polymorphic in the important coding region and both its polymorphism and population differentiation have been supported by many research papers. However, the expression associated SNPs were not well-established in this study and the significance lies in that there is only one SNP associated with expression that occurs in the statin pathway. Although it is hypothesized that regulatory genes would play a major role in statin drug efficacy, this paper narrowed the comprehensive study of the statin pathway to a population differentiated structural SNP and a SNP associated with expression that affects protein structure or function occurring in the gene *SLCO1B3.* As such, this study has provided a headstart to further investigation on the mechanisms behind the impact of different functional SNPs on gene function. This could further conclude the role of SNPs responsible for different statin drug responses across populations. This study has increased our understanding of the pharmacogenetics of the statin pathway before accurately administering the statin related drugs and has spearheaded the initial step of associating the genetic factor, which is unique in each population and even individuals, to varying responses to the statin pathway. This ultimately leads to the aim of administering patients with statin-related drugs that could maximize the benefits and minimize the harmful effects, benefiting mankind.

References

1. Humphrey MB, Bryan J, Cooper TA, Berget SM. A 32-nucleotide exon-splicing enhancer regulates usage of competing 5′ splice sites in a differential internal exon. *Mol. Cell Biol.* 1995;**15**:3979–3988 [PubMed]
2. Teo YY, et al. Singapore Genome Variation Project: A haplotype map of three South-East Asian populations. Genome Res. 2009b;19:2154-2162.
3. T.E. Klein, J.T. Chang, M.K. Cho, K.L. Easton, R. Fergerson, M. Hewett, Z. Lin, Y. Liu, S. Liu, D.E. Oliver, D.L. Rubin, F. Shafa, J.M. Stuart and R.B. Altman, "Integrating Genotype and Phenotype Information: An Overview of the PharmGKB Project", *The Pharmacogenomics Journal* (2001) 1, 167-170.
4. Nuchnoi P, Nantakomol D, Chumchua V, Plabplueng C, Isarankura- Na-Ayudhya C, et al. (2011) The Identification of Functional Non-Synonymous SNP in Human ATP-Binding Cassette (ABC), Subfamily Member 7 Gene: Application of Bioinformatics Tools in Biomedicine. J Bioanal Biomed 3: 026-031. doi:10.4172/1948-593X.1000039
5. Burkhardt R, Kenny EE, Lowe JK, Birkeland A, Josowitz R, Noel M, Salit J, Maller JB, Pe'er I, Daly MJ. et al. Common SNPs in HMGCR in micronesians and whites associated with LDL-cholesterol levels affect alternative splicing of exon13. *Arterioscler Thromb Vasc Biol.* 2008;**28**:2078–2084. [PubMed]
6. Kathiresan S, Melander O, Anevski D, Guiducci C, Burtt NP, et al. Polymorphisms associated with cholesterol and risk of cardiovascular events. *N Engl J Med.* 2008;358:1240–1249.[PubMed]
7. Medina MW, Krauss RM. The role of HMGCR alternative splicing in statin efficacy. Trends Cardiovasc Med. 2009;7:173-177.[PubMed]
8. Pharmacogenomics Knowledge Base. Available at: http://www.pharmgkb.org/
9. National Centre for Biotechnology Information. Available at: http://www.ncbi.nlm.nih.gov/
10. dbSNP Short Genetic Variations. National Centre for Biotechnology Information. Available at: http://www.ncbi.nlm.nih.gov/snp
11. Thorisson, G.A., Smith, A.V., Krishnan, L., and Stein, L.D. The International HapMap Project Web site. Genome Research,15:1591-1593. 2005. Available at: http://hapmap.ncbi.nlm.nih.gov/
12. Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res.* 2002;12(12):1805–1814 [PubMed]
13. Wang et al. pfSNP: An integrated potentially functional SNP resource that facilitates hypotheses generation through knowledge syntheses. Human Mutation, 2011 Jan;32(1):19-24.
14. SNP Fact Sheet. U.S. Department of Energy Office of Science, Office of Biological and Environmental Research. Available from: http://www.ornl.gov/sci/techresources/Human_Genome/faq/snps.shtml
15. Holsinger, Kent E.; Bruce S. Weir (2009). "Genetics in geographically structured populations: defining, estimating and interpreting FST". Nat Rev Genet 10 (9): 639–650. [PubMed]
16. SLCO1B3 solute carrier organic anion transporter family, member 1B3 [Homo sapiens], Available at: http://www.ncbi.nlm.nih.gov/gene/28234
17. Myles S, Davison D, Barrett J, Stoneking M, Timpson N. Worldwide population differentiation at disease-associated SNPs. *BMC Med Genomics.* 2008;**1**:22. [PubMed]
18. Pignatello R, Musumeci T, Basile L et al.: Biomembrane models and drug-biomembrane interaction studies: Involvement in drug design and development. 2011; 3(1):4-14. [PubMed]
19. 1000 Genomes. Available at: http://www.1000genomes.org/
20. Chen SN, Ballantyne CM, Gotto AM Jr, Tan Y, Willerson JT, Marian AJ. A common PCSK9 haplotype, encompassing the E670G coding single nucleotide polymorphism, is a novel genetic marker for plasma low-density lipoprotein cholesterol levels and severity of coronary atherosclerosis. *J Am Coll Cardiol.* 2005;**45**:1611–9 [PubMed]
21. Dobson PD, Lanthaler K, Oliver SG, Kell DB. Implications of the dominant role of transporters in drug uptake by cells. *Curr Top Med Chem.* 2009;9:163–81 [PubMed]