Editorial

Innumerable studies on single nucleotide polymorphisms: What could be its utility?

K. Ghosh, Ajit Gorakshakar

National Institute of Immunohaematology, Parel, Mumbai, Maharashtra, India

After the human genome map was made public and techniques became simpler and universally available to study single nucleotide polymorphisms (SNPs), it became apparent that SNP's are quite common and for every hundred nucleotides in human genome sequence one SNPs is found.[1] Hence out of 3 billion nucleotides sequence that spans human genome 30 million SNP's can be expected. They have been identified in codons, introns and promoter regions of the various genes. Biomedical researchers are now able to genotype biological samples for thousands of SNPs. Human hap map studies[2] and other studies have shown that certain SNP's of a particular gene are present either in a very high or in a very low number in different populations, i.e., SNP polymorphisms are not always random.

Whenever in biology one finds non-random events one has to explain these non-randomness in the light of evolutionary biology, i.e., mutation, natural selection, bottle necks, balanced polymorphisms, founder effects etc., These are different processes which drives the biological evolution of the genes.

Many of the SNP based studies which are spread across various journals over last one decade has one significant flaw, i.e., they are underpowered to detect the significance of the associations they claim. Because they are part of a larger region of linkage disequilibrium. Hence it is difficult to precisely identify the SNP or SNPs that have a biological link with the phenotype.

What is the nature of association of an SNP with either a biological trait or disease process? We must understand that <10% of 3 billion nucleotides are located in the exons which are responsible for the rate of synthesis, structure and function of the protein that a particular gene encodes for while 90% of the nucleotides are in the introns (so also 90% of SNP's in human genome). The functional significance of SNP's in introns is still shrouded in mystery except possibly those SNP's which are located close to intron- exon boundary and produce alternate intron-exon cleavage sites thus affecting the biosynthesis of the protein. Some of the β thalassemia mutations e.g. "IVS 1-5 (G-C)." Is a clear example of this kind of SNP and has clinical significance in antenatal diagnosis and in understanding the biology of β thalassemia syndrome.[3]

Recently discovered small interfering ribonucleic acid (RNA) based suppression of gene function[4] may also arise from intronic SNPs. Outside these two mechanisms some introns provide alternative initiation complex attachment site producing alternative version of a protein from the same gene, i.e., one gene may produce more than one proteins.[5] There could be many more other functions of intronic SNP's in altering the biology of the gene and this clearly needs to be discovered.

Finally the exonic SNP's understandably likely to have more and immediately visible biological significance. Apart from non-sense and missense changes, such SNP's could be synonymous, i.e., the change in SNP does not change the amino acid composition of the

Access this article online	
Quick Response Code:	Website: www.ijhg.com
	DOI: 10.4103/0971-6866.124354

Address for correspondence: Dr. K. Ghosh, National Institute of Immunohaematology, 13th Floor, Multistoreyed Building, K.E.M. Hospital Campus, Parel, Mumbai - 400 012, Maharashtra, India. E-mail: kanjakshaghosh@hotmail.com

protein and this phenomenon is due to degeneracy of genetic code, i.e., one particular amino acid may be could by more than one triplet code.[6] However, it has been demonstrated that some of these synonymous SNP changes could also be pathological and this is due to differential availability of RNAs for different triplet codons of the same amino acid.[7]

In addition, association of an SNP with a disease process could be direct, i.e., the SNP is situated in crucial part of the gene like in exons, 5'and 3' end of the gene, in intron - exon boundary. We have already referred to β thalassemia mutation. Similarly SNPs in the 5'end of protein C gene affects protein C levels in a population are lower protein C level can predefine to hereditary thrombophilia.[8]

Where SNPs have direct pathological connotation, detection of such SNPs can have diagnostic as well as prognostic significance and this is the reason why search for various SNPs and their association with various biological disease traits continues at breakneck speed.

Finally, there are SNPs which are not directly associated with the changes in the devised and altered gene but became of the close association of the SNP will the diseased gene such SNPs are said to be in linkage disequilibrium with the disease causing gene. Major problem in SNP studies arise from these group of SNPs where the linkage disequilibrium may exist in one population but may not exist in other population. Such SNPs also need statistically robust numbers to look for such associations.

In the present issue of the Journal several SNP based studies (restriction fragment length polymorphism [RFLP] or otherwise) have been presented.[9-14] Number of samples studied in each of the studies are very small, i.e., between 60 and 250 samples. However, number of RFLPs studied in these papers are also one or two, hence statistically these studies may not require a $P < 1 \times 10^{-4}$ or lower to confidently talk about its significance but a $P < 0.01$ should be acceptable. However, SNP studies also require a hypothesis generated set of populations (samples), which should be different from hypothesis testing sets. In none of the studies stated in this issue of the Journal such an approach has been used.

One of the major questions which troubles all of us is what to do with large number of SNP studies with small number of samples? Are these studies baseless and be related to the doubting of the history of science or they can still be made useful. Recently Schaub *et al.*[15] using encode consortium identified functional SNPs which may be associated with the disease phenotypes in 80% of the previously reported "possible associations".

Experimental medical statistician can stratify these studies accepting to various attributes and can join them in a meta-analysis format to make the studies more robust. More over these small studies may pave the way for a bigger and more robust study. In that case such small studies are like mini pilot studies spread in time and space.

In the present issue of the Journal Larijani *et al.*[16] presented an extensive meta-analysis of X-ray repair cross complementing group I gene SNP frequencies. The study shows how meta-analysis can be refined to a degree and author talks about it in his conclusion though innumerable illustrations, forest plots, Galbraith plots and other plots are given.

However, we also have to understand that no amount of statistical jugglery can hide the weakness from the basic design of the study.

Hence, it is expected that all SNP based studies should be carefully crafted, well-designed from the beginning so that the utility of that particular study will exist at sent for some time.

References

1. Schneider JA, Pungliya MS, Choi JY, Jiang R, Sun XJ, Salisbury BA, *et al.* DNA variability of human genes. Mech Ageing Dev 2003;124:17-25.
2. Salisbury BA, Pungliya M, Choi JY, Jiang R, Sun XJ, Stephens JC. SNP and haplotype variation in the human genome. Mutat Res 2003;526:53-61.
3. Kazazian HH Jr, Orkin SH, Antonarakis SE, Sexton JP, Boehm CD, Goff SC, *et al.* Molecular characterization of seven beta-thalassemia mutations in Asian Indians. EMBO J 1984;3:593-6.
4. Dykxhoorn DM, Schlehuber LD, London IM, Lieberman J. Determinants of specific RNA interference-mediated silencing of human beta-globin alleles differing by a single nucleotide polymorphism. Proc Natl Acad Sci U S A 2006;103:5953-8.
5. Ast G. The alternative genome. Sci Am 2005;292:61-5.
6. Berg JM, Tymoczko JL, Stryer L. Biochemistry. 5th ed. New York: W H Freeman; 2002.
7. Smith AP. Nucleic acids to amino acids: DNA specifies

protein. Nat Educ 2008;1:126.
8. Lipe B, Ornstein DL. Deficiencies of natural anticoagulants, protein C, protein S, and antithrombin. Circulation 2011;124:e365-8.
9. Aliparasti MR, Almasi S, Majidi J, Zamani F, Khoramifar AR, Farshi Azari AR. Protein tyrosine phosphatase non receptor type 22 gene polymorphism C 1858 T is not associated with leprosy in Azerbaijan, Northwest Iran. Indian J Hum Genet 2013;19:403-7.
10. Pazarbasi A, Yilmaz MB, Alptekin D, Luleyap HU, Tansug Z, ye Ozpak L, *et al.* Genetic polymorphisms of estrogen receptor alpha and cate chol–O-methyltransferase genes in Turkish patients with familial prostate carcinoma. Indian J Hum Genet 2013;19:408-11.
11. Kaur A. Prevalence of methylenetetrahydrofolate reductase 677 C-T polymorphism among mothers of Down syndrome children. Indian J Hum Genet 2013;19:412-4.
12. Rooki H, Haerian MS, Azimzadeh P, Ebrahimi M, Mirhafez R, Ferns G, *et al.* Distribution and genotype frequency of the C1431T and Pro 12 Ala polymorphisms of the peroxisome proliferator activator receptor gamma gene in an Iranian population. Indian J Hum Genet 2013;19:423-9.
13. Bhanushali AA, Das BR. Promoter variants in interleukin 6 and tumor necrosis factor alpha and risk of coronary artery disease in a population from Western India. Indian J Hum Genet 2013;19:430-6.
14. Kumar Das D, Rahate SG, Mehta BP, Gawde HM, Tamhankar PM. Mutation analysis of mitogen activated protein kinase 1 gene in Indian cases of 46, XY disorder of sex development. Indian J Hum Genet 2013;19:437-42.
15. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations with regulatory information in the human genome. Genome Res 2012;22:1748-59.
16. Larijani B, Mohammadi Asl J, Keshtkar AA, Saki N, Larijani FA, Rahim F. Deoxyribonucleic acid repair gene X-ray repair cross complementing group I polymorphisms and non-carcinogenic disease risk in different populations: A meta-analysis. Indian J Hum Genet 2013;19:494-511.

Cite this article as: Ghosh K, Gorakshakar A. Innumerable studies on single nucleotide polymorphisms: What could be its utility?. Indian J Hum Genet 2013;19:381-3.

Source of Support: Nil, **Conflict of Interest:** None declared.