

Correlation of SNP Distribution to Gene Function and Disease



Lily Chu
Montgomery High School
USA

Lily Chu* and Ruiyin Chu

Montgomery High School, New Jersey, USA

COLUMN ARTICLE

The genomic distribution of SNPs is not homogenous. Natural selection is acting and 'fixing' the allele (eliminating other variants) of the SNP so that it constitutes the most favorable genetic adaptation [1-3]. How does this uneven distribution relate to gene function? Whether the SNP susceptibility is linked to certain diseases?

In order to answer these questions, SNP data from the public dbSNP database (<http://www.ncbi.nlm.nih.gov/SNP/>) was extracted for the following gene families and genes:

1. Cytoskeleton protein actin gene family: ACTA1, ACTA2, ACTB, ACTC1, ACTG1, ACTG2
2. Stress protein HSP70 gene family: HSPA1A, HSPA2, HSPA4, HSPA5, HSPA6, HSPA8, HSPA12A, HSPA12B, HSPA13, HSPA14
3. Sodium channel SCN gene family: SCN1A, SCN2A, SCN3A, SCN4A, SCN5A, SCN8A, SCN9A, SCN10A, SCN11A
4. Nuclear protein histone gene family: HIST1H1A, HIST1H1C, HIST1H2AA, HIST2H2AA3, H1F0, H2AFJ, H2AFV, H2BFM, H3F3B, H3F3C
5. Breast cancer genes: BRCA1, BRCA2
6. Basic transcription factor GTF gene family: GTF2A1, GTF2A2, GTF2B, GTF2E1, GTF2H1, GTF2H5, GTF3A, GT-

F3C1, GTF3C6

7. Human leukocyte antigen HLA gene family: HLA-A,

Data analysis results of these 53 genes which belong to seven different gene families were shown in Figure 1. It was observed that SNP distribution within gene family was relatively similar; more distinctive differences were found between gene families. For example, the sodium channel (SCN) gene family possessed 28-fold higher total SNPs than the conserved histone gene family. The two breast cancer genes, however, not only had the second highest total SNP numbers, but also the highest nsSNP/SNP ratio (47%). The basic transcription factor gene family (GTF), on the other hand, had very low total SNP frequency and the lowest snSNP/SNP ratio (4.9%).

These results confirmed our hypothesis that if a gene/family is functionally more important to basic cellular function; it tends to have fewer SNPs, particularly sn-SNPs. The linkage of SNP with breast cancer was also well evidenced in this study: high total SNP frequency and high nsSNP/SNP ratio.

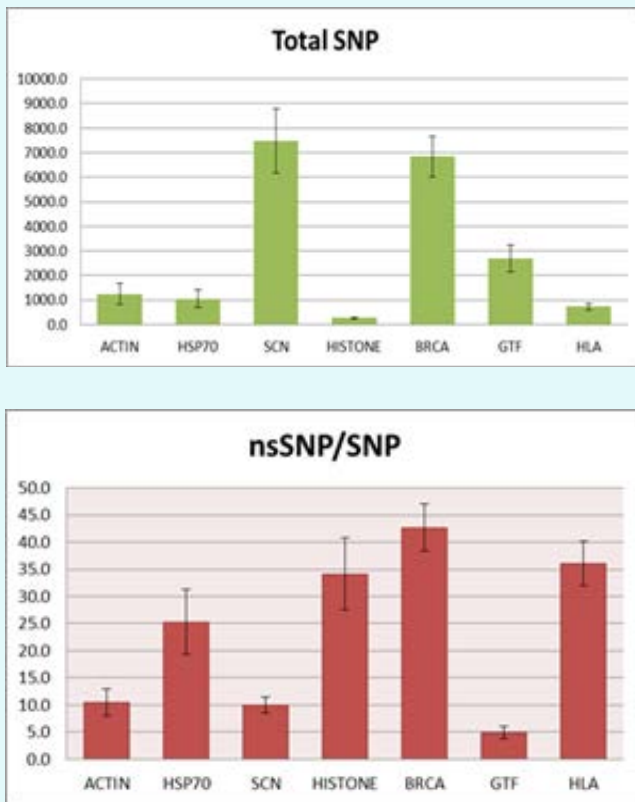


Figure 1: Total SNP distribution and nsSNP/SNP ratio (%) among different gene families.

BIBLIOGRAPHY

1. Manry J, *et al.* "A Genome-Wide Perspective of Human Diversity and Its Implications in Infectious Disease, Cold Spring". *Cold Spring Harbor Perspectives in Medicine* 3.1 (2013): a012450.
2. Brookes AJ. "The essence of SNPs". *Gene* 234.2 (1999): 177-186.
3. Datta A, *et al.* "Functional and Structural Consequences of Damaging Single Nucleotide Polymorphisms in Human Prostate Cancer Predisposition Gene RNASEL". *BioMed Research International* (2015): 271458.

©All rights reserved by Lily Chu and Ruiyin Chu.