International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2019): 7.583

Lot Depends on our SNPs for our Existence

Raghavendra Krishnappa

Life Science, Healthcare and Pharma Vertical, Mphasis Limited, Bengaluru, India

Abstract: ***SNPs present in coding, non-coding or in intergenic region of human DNA sequence can impact in transcription, translation and deciding structure of the protein. When SNPs occur within a gene or in a regulatory region near a gene, they may play a more direct role in disease by affecting the gene's function or by aberrant proteins. Altered cellular metabolism caused by a specific SNP can be useful for diagnosing its related disease. Hence, SNPs can act as biological markers, helping scientists locate genes that are associated with disease and so, it is important to prioritize and study those SNPs that falls in the regulatory region of the gene.***

Keywords: SNPs - Single Nucleotide Polymorphisms (pronounced "snips"), UTR - Untranslated region, rs- Reference SNP (RefSNP), ss- submitted SNP, dbSNP –Public Domain SNP database, NCBI – National Centre for Biotechnology Information. DNA – Deoxyribose Nucleic Acid, RNA – Ribose Nucleic Acid, mRNA – Messenger RNA

1. Introduction

SNPs, variations in the DNA sequence that make each individual unique, are the genetic determinants of health and disease. Many SNPs act as biological markers enabling molecular biologist to trace elusive links between genes and disease, particularly the common diseases to which many genes contribute. All types of SNPs falling in coding, non-coding or in intergenic regions can have an observable phenotype or can result in disease. They may play a more direct role in disease by affecting the gene's function. Changes in many genes, each with a small effect, may underlie susceptibility to many common diseases, including cancer, obesity, diabetes, heart disease, and mental illness. SNPs in non-coding regions can manifest in a higher risk of cancer and may affect mRNA structure and disease susceptibility. Non-coding SNPs can also alter the level of expression of a gene. Though synonymous SNPs of the coding region do not affect the protein sequence, nonsynonymous SNPs (missense and nonsense) of the coding region change the amino acid sequence of protein and SNPs which are termed as eSNP (expression SNP) present upstream or downstream from the gene though not in protein-coding regions may still affect gene splicing, transcription factor binding, messenger RNA degradation, or the sequence of noncoding RNA.

2. Background

NCBI Entrez Gene database has almost all human genes that all together make human genome and NCBI dbSNP consists of all SNPs related to human genes. The number of SNPs data in public databases is increasing dramatically on every release. The number of unique human SNPs in the current dbSNP release:

(https://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi) (build-151) is more than 660million (in that 113.862 million SNPs are validated) and 381 million (0.38 billion) of them reside in gene. Last two dbSNP releases (150 and 151) has seen more than 100% increase in number of rs#'s in human genome compared with its previous release(Table 01).

Table 1: SNPs statistics on every release.(https://www.ncbi.nlm.nih.gov/projects/SNP/snp_summary.cgi)

Release Date	dbSNP Build	# of rs#'s in Genome	Millions	Billions	# of rs#'s in Gene	Millions	Billions
Oct-2017	151	660773127	660.773127	0.660773127	381785470	381.78547	0.38178547
Feb-2017	150	325658303	325.658303	0.325658303	191585061	191.585061	0.191585061
Nov-2016	149	154206854	154.206854	0.154206854	89404961	89.404961	0.089404961
Apr-2016	147	153953962	153.953962	0.153953962	89232381	89.232381	0.089232381
Nov-2015	146	150482731	150.482731	0.150482731	87339846	87.339846	0.087339846
Oct-2014	142	112743739	112.743739	0.112743739	53983919	53.983919	0.053983919
May-2014	141	62387983	62.387983	0.062387983	29901117	29.901117	0.029901117
Apr-2013	138	62676337	62.676337	0.062676337	27608151	27.608151	0.027608151
Jun-2012	137	53567890	53.56789	0.05356789	22450743	22.450743	0.022450743

3. Method

The SNPs detection method varies, as does the reliability of the SNPs in dbSNP. In build#151, there are totally 1.8billion (1, 803 million) SNP submission (ss#'s) made, in that, only 363 million (20.13%) SNPs have allele frequency information and only 113.862 (17.12%) million SNPs (rs#s) are validated compared with existing 660million SNP's (rs#'s). In addition to the validation and allele frequency information, it is also important to select SNPs based on their genomic location and their proposed functional significance that include 5' end of the gene, 3' end of the gene, 5' UTR, 3' UTR, transcription start site, transcription initiation site, translation start site, translation initiation site, exons, introns, promoters, etc.. (Figure-01)

There are currently 61, 731 gene entries for homo sapiens in NCBI Entrez Database (as of Feb2021), in that only 19, 710 genes are protein-coding. Gene sequences which are part of Promoter, 5' UTR, Intron and 3' UTR region play a vital role in governing expression of each human gene, SNPs falling in the genes functional elements like Transcription Start Site

(TSS), Transcription Factor Binding Sites (TFB's), CpG Islands (CG Islands) in promoter region. Kozak sequence, CpG Islands, Transcription Factor Binding Sites (TFB's), Cap signal, Transcription Start Site (TSS), CT signal (Cysteine Thymidine), Iron Responsive Element (IRE) in 5'UTR region. 5'splice site, 3' splice site, Branch Point Sequence (BPS), Polypyrimidine tract, CpG Island in Intron region. Poly A Signal, Poly A Site, Poly A Limiting Element (PLE) and AU rich Region (ARE) in 3' UTR determine phenotype of transcriptional products. (Table-01)

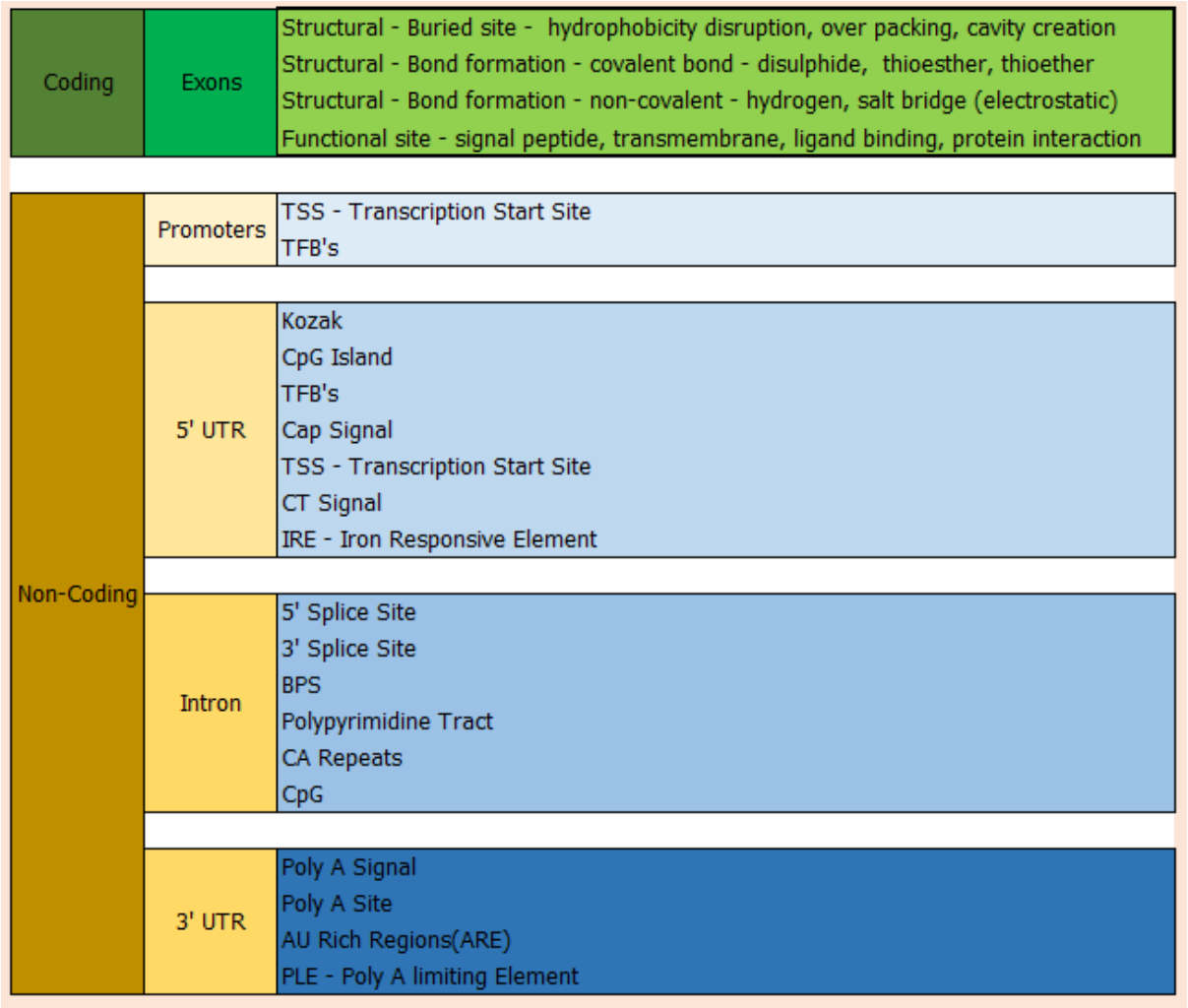

Figure 1: SNPs Selection – Gene functional element which determine phenotype of transcriptional products

Table 1: Different phenotypic risks and putative functional effects which should be used for SNPs selection

SNPs Selection			
	Region	**Features**	**Effect**
Non-Coding	5' UTR	Kozak	Effects in Translation Initiation
		CpG Island	Effects Transcription Factor Binding Site
		TFB's	Transcription Factor Binding Site
		Cap Signal	Effect in transcription efficiency
		TSS - Transcription Start Site	Effect in 5' capping and Transcription Initiation Regulation
		CT Signal	Effect in transcription efficiency
		IRE - Iron Responsive Element	Effect in transcription efficiency
	3' UTR	Poly A Signal	Transcriptional termination, mRNA stability and degradation
		Poly A Site	mRNA stability
		AU Rich Regions (ARE)	Destabilizing Element
		PLE - Poly A limiting Element	Poly A limiting Element
	Introns	5' Splice Site	Effect in Splicing
		3' Splice Site	Effect in Splicing
		BPS	Effect in Splicing
		Polypyrimidine Tract	Effect in Splicing
		CA Repeats	Effect in Splicing
		CpG	Transcription protein binding region

	Promoters	TSS - Transcription Start Site	Effect in 5' caping and Transcription Initiation Regulation
		TFB's	Transcription Factor Binding Site
Coding	Exon	Structural - Buried site - hydrophobicity disruption, over packing, cavity creation	It is with high confidence supposed to affect protein function or structure
		Structural - bond formation - covalent bond - disulphide, thioester, thioether	It is with high confidence supposed to affect protein function or structure
		Structural - bond formation - non-covalent - hydrogen, salt bridge (electrostatic)	It is with high confidence supposed to affect protein function or structure
		Functional - functional site - signal peptide, transmembrane, ligand binding, protein interaction	It is with high confidence supposed to affect protein function or structure

4. Conclusion

Along with SNPs, dbSNP also consist of other polymorphisms like, DIP: deletion / insertion polymorphisms, Heterozygous: Variable but under defined at the nucleotide level, STR: Short tandem repeat (microsatellite) polymorphism, Named: Insertion / deletion polymorphism of named repetitive element, Mixed: cluster contains submissions from two or more allelic classes, MNP: multiple nucleotide polymorphisms with alleles of common length, but allele frequency of them is a must to give prominence to them.

Human transcription promoter sequences are difficult to identify and are poorly annotated in the public databases. It is reported that most of the human promoters are located within 2 kb upstream of the transcription start site (1). Therefore, the length of input sequences for selecting SNPs from each gene can be restricted to 3 kb upstream of the transcription start site.

Approximately 20K human genes produce around 80k to 400K proteins, some genes will be having more than one mRNA which goes down as splice variants, rating snips based on the selected splice variant is very important as the gene region gets changed for every selected mRNA, as the gene region varies SNP’s significance on loci also gets varied.

Although there are several efforts on rating snips with respect to functional annotation, the coverage of existing commercial or public domain efforts is far from complete. So far, these groups are each focusing on a narrow set of annotations, such as rating/annotating snips only in exon, promoter and in intron region (2, 3, 4). There is a need of classifying SNPs based on its sequence position, regulation, and its effect on gene function to trace elusive links between genes and disease. Once classified they can act as biological markers, helping scientists locate genes that are associated with disease. SNPs disrupting such target sites (Table-02) have the potential to change the expression pattern of a gene in regard to the amount of product expressed at a certain time, result in partly functional truncated polypeptide, will produce a completely inactive protein product or exhibit complete loss of normal protein structure and function. Single nucleotide polymorphisms (SNPs) are inherited from parents and they measure heritable events. In addition to an identified genetic change people with a genetic predisposition, the risk of disease can depend on the response to certain environmental signal and the level of expression in certain tissues.

Most SNPs have no effect on health or development but candidate SNPs pinpoint differences in our susceptibility to a wide range of diseases. The severity of illness and the way the body responds to treatments are also manifestations of genetic variations caused by SNPs.

References

[1] Kel, A.E., Gossling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V. and Wingender, E. (2003) MATCH: a tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res., 31, 3576–3579.
[2] SNP Function Portal: a web database for exploring the function implication of SNP alleles, Vol. 22 no. 14 2006
[3] FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization, Nucleic Acids Research, 2006, Vol. 34
[4] FESD: a Functional Element SNPs Database in human, D518–D522 Nucleic Acids Research, 2005, Vol. 33