Molecular Biology and Pharmacogenetic Association Studies

Saurabh Singh Rathore

Research Scholar

Department of Medical Genetics

Sanjay Gandhi Post Graduate Institute of Medical Sciences, Lucknow, India

Abstract : Rapid advancements in molecular biology and genomic technologies have increased the hopes for personalized medicine. Pharmacogenetic association studies are the best example of scientific ways to prove the utility of genetic testing in medicine. To better understand the biology of human diseases one should be clear about basic molecular biology concepts. This article presents the molecular biology concepts in the background of human genetics. Pharmacogenetic studies have also been discussed at the end to give the reader an idea about their use in translational research.

Introduction

Clinicians have always dreamt of predicting patient responses to a specific drug. In particular, medical care would be greatly facilitated if clinicians have prior information about the most suitable drug in optimum amounts for individual patients. Some drugs cause adverse reactions in only some patients but not all. This kind of knowledge is significant as shown by meta-analysis study in US hospitals from 1966 to 1996. This study reported 6.7% incidence of adverse drug reactions. The adverse drug reactions were defined for those with death, permanent disability or prolonged hospital stay. The adverse drug reactions due to errors in dosing (e.g. overdosing) and drug selection were excluded.[1]

Generally, adverse drug reactions cause 1-4 days lengthened stays at hospitals with a cost increase amounting to around $2,300–5,600.[2] These factors make prediction of patient responses to specific drugs very important, particularly in terms of medical care and socioeconomics. In recent times, molecular biology has made it possible associate genetic profile of individual patients with drug response. Pharmacogenetics and pharmacogenomics are the two specialties which study drug responses in the background of genetic information. Pharmacogenetics studies the drug response and inherited variation in metabolism of drugs whereas pharmacogenomics broadly studies many genes associated with drug response.

To have a better understanding of disease physiology and genetics, one should have clear concepts of molecular biology. The structure of DNA, its replication and translation into specific proteins and how this information system is important in human disease manifestations are some of the topics which must be thoroughly understood by clinicians and translational scientists. Keeping this in mind one can have a better understanding and application ability to design new experiments.

Molecular Biology: Basic Concepts

Deoxyribonucleic acid (DNA) is the molecule coding for the molecular structure of every protein. DNA is made up sugar phosphate backbone and four purine/pyrimidine bases: adenine (A), guanine (G), thymine (T), or cytosine (C). The genetic information exists in the cell nucleus in the form of double helix DNA. Purines (A and G) pair with pyrimidines (T and C) and this base pairing occur between the two sugar phosphate backbone strands. Huge amount of variation within the sequence and order of the base pairs constitute the genetic code for different organisms.[3]

The two DNA strands serve as templates for messenger RNA (mRNA) transcription. The mRNA transcription is catalyzed by RNA polymerase in cell nucleus. DNA strand consists of a promoter region, 25-200 base pairs proximal to the transcription initiation site. This promoter region confirms the transcriptional specificity of the particular DNA strand, i.e., it orients the RNA polymerase so that only one of the two DNA

strands gets transcribed. The process of transcription is regulated by many regulatory proteins, *transcription factors*, which bind to the promoter site or to the specific DNA sequences at distant enhancer site. Amount of mRNA transcribed and thus protein production is regulated by promoter and enhancer regions present on the DNA template. mRNA produced from transcription is processed in cell nucleus through its splicing. mRNA is made up of regions, namely introns (non-coding) and exons (coding). The intronic regions of mRNA are spliced by enzymes known as *spliceosomes*. The exons remaining after intron splicing are joined together to form mature mRNA which is transported to cellular cytoplasm. To form various kinds of protein molecules, the mature mRNA undergoes ribosomal translation. *Codons* are the sequential triplet of bases present on mRNA. Each of these *codons* encodes an amino acid.

The genetic information in living organisms is constituted by *genes*. A *gene* is a part of *chromosome*, i.e., a specific DNA sequence and its associated proteins. There are total 23 pairs of chromosomes in humans. The general structure of all the genes is common and consists of 5' untranslated region, exons, introns, and a 3'. A *gene* occurs in two or more alternative forms occupying the same chromosomal locus. The single-nucleotide polymorphism (SNP) is the position on DNA sequence at which to alternative nucleotides occurs. SNP is the most common type of human genetic or allelic variation. The other types of mutations including the insertion, deletion, translocation, or inversion of DNA segments are the primary allelic variations. Generally, mutation is the term used for the variation that has a frequency of less than 1% in the population and polymorphism is the term used for the variation occurring in more than 1% of the population. The words *mutation* and *polymorphism* are used interchangeably.

Most of the allelic variations are silent (i.e., they don't have any significant effect on the proteins they code). But some genetic variations may significantly affect the amount and composition of proteins and so the phenotype of organism. Metabolism and efficacy of drugs as well as their side effects are common phenomena that are altered by genetic variations. There are two copies of most genes present in the form of alleles in every organism and one of these two alleles is inherited from one parent and the other one from other parent. If mutation is present in both the inherited alleles then its clinical effects may get magnified and this condition is known as (homozygous expression). In contrast, if the mutation is present in a single allele the condition is heterozygous and the symptoms may be mild. The example of drug clearance can be taken to explain the effect of homozygous and heterozygous mutations. The activity of drug clearing enzyme coded in case of heterozygous individual is lesser than that in case of homozygous (normal copies of enzyme on both chromosomes) individual. But, if both copies of enzymes on both the chromosomes are abnormal then a mutated enzyme with very little or nil activity is produced which further causes significant reduction in drug clearance. So, it can be inferred that mutation effects may act in an additive manner. Mutations acting in a dominant fashion are sufficient enough to cause its effect. The mutations acting in a recessive manner need abnormal copies of the gene on both the chromosomes for disease to occur. For Mendelian traits, single or multiple genetic changes cause phenotypic change in "yes/no" fashion. For complex traits, epigenetic as well as genetic factors interact in a complex way to produce phenotypic effects.

Mendelian diseases are mostly unaffected by epigenetic factors. Postoperative analgesic requirements vary between individuals. This is a complex trait which is affected by the interaction of many genetic and environmental factors. All of these factors affect the drug metabolism, efficacy and drug side effects. Recent ingestion of a meal, previous alcohol or drug use, ways of preparing and administering drugs are some of the environmental factors that may affect their pharmacokinetics and pharmacodynamics. Mutation variability (heterogeneity means that many different genetic changes in same or different genes can cause the same disease), gene penetrance and gene expressivity are some of the issues that must be considered while doing genetic testing of patients for Mendelian and complex traits. Gene penetrance is the probability of expressing a mutation specific trait. An individual may not show the symptoms of disease in case of incomplete gene penetrance. The range of severity of a disease is known as gene expressivity. Gene expressivity occurs in varying degrees for Mendelian traits in addition to complex traits.

Nomenclature of Single Nucleotide Polymorphisms

Three different nomenclature systems are used for naming genetic polymorphisms. In the first system, a number is used to depict the gene locus at which a nucleotide change takes place. Two letters are placed adjacent to this number: one before and one after the number. The letter before the number is the one that occurs most commonly in the population and is known as the "major allele". The letter placed after the number is the one for the less likely allele that is known to be "minor allele". The other ways of representing single nucleotide polymorphisms are by writing the gene locus number followed by major and minor alleles separated by "_" or "/". Sometimes a single gene codon is affected by a polymorphism so that there is an amino acid switch. Such a change is shown by placing three letter abbreviations for the exchanged amino acids around the amino acid locus of the protein. The third way of polymorphism nomenclature is by numbering the alleles in the order of their discovery. This system of nomenclature provides the minimum description but it is flexible enough to accommodate single as well as multiple genetic mutations. Apart from these systematic nomenclature systems there exist many nonstandard nomenclature methods to depict genetic changes.

Pharmacogenetic Methods

There are two pharmacogenetic methodological approaches. First one is the "genetic linkage" and other one is "gene association". In genetic linkage, two or more alleles are so close on the same chromosome so that they tend to be inherited together. The close positioning prevents crossover between the alleles. A combination of alleles closely linked together on a single chromosome that tend to be inherited together is known as haplotype. Such SNPs are called to be in LD or linkage disequilibrium as these alleles miss the chance of random distribution in a population.

In recent times, research in molecular biology and completion of human genome project resulted in identification of millions of undiscovered polymorphisms. Many novel candidate genes have now been identified and this is facilitated by our increased knowledge about the enzymes, receptors, and other factors in pharmacokinetics and pharmacodynamics of drugs. Expression of the disease and adverse clinical events, in patients with specific genetic makeup, has become focus of many scientific studies. Human genetic variations affecting the variations in disease pathogenesis, expression and treatment response are being studied through genetic association studies. In these association studies, association analysis is done on patients and controls for particular identified SNPs to look for their occurrence as disease markers. Gene association studies are based on fundamental case-control methodology which is statistically more powerful than linkage studies. They have an additional advantage of non-requirement of family members. Therefore the case-control association study design is preferably favored by most of the current pharmacogenetic studies.

Limitations exists in genetic association studies too.[4] Firstly, a study population consists of patients originating from different genetic composition. Therefore an association study should account for this kind of genetic admixture and be sufficiently powered statistically for this purpose. This becomes particularly important for complex diseases because they can have multiple genetic origins. In case of complex diseases, association can be divided into multiple candidate loci in patients representing a population. Therefore, association studies, involving a few hundred samples, showing negative association for loci with a particular complex disease in a patient sample with high genetic admixture should be interpreted cautiously. The second limitation of association studies is about proper statistical analyses application and data interpretation. Two different statistical analyses show different results. So, it is advisable to select the statistical method very cautiously. There should be a logical way of data interpretation as there are multiple comparisons of different variables in different populations. Identification of spurious associations should be avoided. Additionally, a positive association of a clinical outcome with a specific genotype is no guarantee

of effect-causality relationship. In such situation, the associated genotype may be clinically silent or it might be linked to other variants to form a disease haplotype. Therefore, it is advisable to build a haplotype map of human genome so that disease predisposing genetic variants can be located in an easier, faster and cheaper manner. If haplotype maps are created then one has to search for few thousand base pairs of DNA instead of searching whole of human genome. Sensitivity and specificity of prediction of association of genotypic variation with clinical outcome may be improved by haplotype mapping.[5]

Conclusion

The genetic component of drug-related phenomena has been proved by growing pharmacogenetic data. These studies have shown significant effects of genetic variation on drug absorption, distribution, metabolism, excretion, and toxicity. Genetic testing might provide explanation for the unexplained medical conditions. Further identification of polymorphisms affecting the drug pharmacokinetics from those affecting drug pharmacodynamics. Pharmacogenetic studies may develop screening tests for genetic variants significantly affecting drug metabolism, efficacy and side effects. However, all these hopes rely much on the successful demonstration of pharmcogenetic prospective trials on fresh patient populations and the issue of cost to benefit ratio for performing genetic tests.

References

[1] Lazarou, J., B.H. Pomeranz, and P.N. Corey, *Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies.* JAMA, 1998. **279**(15): p. 1200-5.

[2] Rodriguez-Monguio, R., M.J. Otero, and J. Rovira, Assessing the economic impact of adverse drug effects. Pharmacoeconomics, 2003. 21(9): p. 623-50.

[3] Shah, S.A. and A.T. Brunger, The 1.8 A crystal structure of a statically disordered 17 base-pair RNA duplex: principles of RNA crystal packing and its effect on nucleic acid structure. J Mol Biol, 1999. 285(4): p. 1577-88.

[4] Ziegeler, S., B.E. Tsusaki, and C.D. Collard, Influence of genotype on perioperative risk and outcome. Anesthesiology, 2003. 99(1): p. 212-9.

[5] Hoehe, M.R., B. Timmermann, and H. Lehrach, Human inter-individual DNA sequence variation in candidate genes, drug targets, the importance of haplotypes and pharmacogenomics. Curr Pharm Biotechnol, 2003. **4**(6): p. 351-78.