J Hum Genet (2007) 52:871–880
DOI 10.1007/s10038-007-0200-z

MINIREVIEW

SNPs in disease gene mapping, medicinal drug development and evolution

Barkur S. Shastry

Received: 28 July 2007 / Accepted: 18 September 2007 / Published online: 11 October 2007

Abstract Single nucleotide polymorphism (SNP) technologies can be used to identify disease-causing genes in humans and to understand the inter-individual variation in drug response. These areas of research have major medical benefits. By establishing an association between the genetic make-up of an individual and drug response it may be possible to develop a genome-based diet and medicines that are more effective and safer for each individual. Additionally, SNPs can be used to understand the molecular mechanisms of sequence evolution. It has been found that throughout the given gene, the rate, type and site of nucleotide substitutions as well as the selection pressure on codons is not uniform. The residues that evolve under strong selective pressures are found to be significantly associated with human disease. Deleterious mutations that affect biological function of proteins are effectively being rejected by natural selection from the gene pool. If substituted nucleotides are fixed during evolution then they may have selection advantages, they may be neutral, or they may be deleterious and cause pathology. Therefore, it is possible that disease-associated SNPs (or pathology) and evolution can be related to one another.

Keywords Evolution · Genomics · Medicine · Pharmacology · Polymorphism

B. S. Shastry (✉)
Department of Biological Sciences,
Oakland University, Rochester, MI, USA
e-mail: shastry@oakland.edu

Introduction

The completion of the human genome project has generated new enthusiasm and opportunities in life sciences. It has provided the necessary tools to understand the genetic basis of diversity among individuals, the most common familial traits, evolutionary processes, complex and common diseases such as diabetes, obesity, hypertension and psychiatric disorders, and to develop genome-based medicinal drugs (Emilien et al. 2000). Scientists generally think that the genomes between two randomly selected individuals contain approximately 0.1% differences or variations. This variation is called polymorphism, and it arises because of mutations. Several comparative studies on identical and fraternal twins (Martin et al. 1997) and siblings suggest that DNA polymorphism is one of the factors associated with susceptibility to many common diseases (Table 1), every human trait such as curly hair, individuality and inter-individual difference in drug response. DNA sequence variation is also considered to be responsible for genome evolution. Based on these observations, it has been proposed that by cataloging the DNA polymorphisms in different populations and in different species, it may be possible (a) to develop genome-based knowledge on the susceptibility of an individual to many common diseases, (b) to manufacture safer and more effective individualized diet and medications for patients, and (c) to understand evolutionary processes. However, many experts believe that this single nucleotide polymorphism (SNP) technology has to face several challenges before it makes its impact on medicine. What follows is a brief discussion of the above three aspects with emphasis placed on evolution.

Table 1 A partial list of diseases associated with single nucleotide polymorphisms

Disease	Gene	Disease	Gene
Asthma	EDN1 and NOS1	Lung cancer	MMP1
	Chemokine		p53
Arrhythmia	KCN1	Myocardial	TSP
		Infraction	PCS
Blood pressure	TAF1	Migraine	IR
Biliary cirrhosis	MBL	Obesity	PAI1
Bipolar affective disorder	HRT 3A	Ossification	Npps
Colorectal cancer	Cyclin D1	Oxalate stone	E-Cad
Crohn's Disease	MDR1	POAG	Myocilin
Dyslipidemia	Lipase	Rheumatoid arthritis	MIF
Eating disorder	Melanocortin	Systemic sclerosis	Fibrillin1
Esophagel adenocarcinoma	Cyclin D1	Severe sepsis	TNF-α
Hyperbilirubinemia	UGT1A1	Type II diabetes	Syntaxin1A
Idiopathic arthritis	MIF	Ulcerative colitis	MDR1
Idiopathic PD and FTD	Tau	Urinary bladder cancer	Cyclin D1
Knee and hip osteoarthritis	Collagen	Autism	CNP

EDN1 endothelin 1, *NOS1* neuronal nitric oxide synthetase 1, *KCN1* potassium channel protein, *TAF1* thrombin-activatable fibrinolysis inhibitor, *MBL* mannose binding protein, *UGT1A1* UDP glucoronosyl transferase, *MIF* macrophage migration inhibitory factor, *PD* Parkinson disease, *FTD* frontotemporal dementia, *MMP1* matrix metalloproteinase 1, *PAI* plasminogen activator inhibitor, *Npps* nucleotide pyrophosphatase, *Cad* cadherin, *POAG* primary open angle glaucoma, *TNF* tumor necrosis factor, *MDR1* p-glycoprotein (multiple-drug-resistant), *TSP* thrombospondin, *PCS* prostacyclin synthase, *IR* insulin receptor, *CNP* copy number polymorphism

Detection and analysis of DNA polymorphism

The simplest form of DNA variation among individuals is the substitution of one single nucleotide for another. This type of change (Fig. 1A) is called SNP. It is estimated that SNPs occur at a frequency of 1 in 1,000 bp throughout the genome. These simple changes can be of transition or transversion type. According to one report (Halushka et al. 1999), approximately 50% of SNPs are in the noncoding regions, 25% lead to missense mutations (coding SNPs or cSNPs), and the remaining 25% are silent mutations (they do not change encoded amino acids). These silent SNPs are called synonymous SNPs, and it is most likely that they are not subject to natural selection (but see below). On the other hand, nonsynonymous SNPs (nSNPs, change-encoded amino acids) may produce pathology and may be subject to natural selection. SNPs (both synonymous and nonsynonymous) influence promoter activity and pre-mRNA conformation (or stability). They also alter the ability of a protein to bind its substrate or inhibitors (Kimchi-Sarfaty et al. 2007) and change the subcellular localization of proteins (nSNPs). Therefore, they may be responsible for disease susceptibility, medicinal drug deposition and genome evolution. Although several of them affect the functions of genes, many of them are not deleterious to organisms and must have escaped selection pressure. For the purpose of identifying SNPs, several private and public organizations

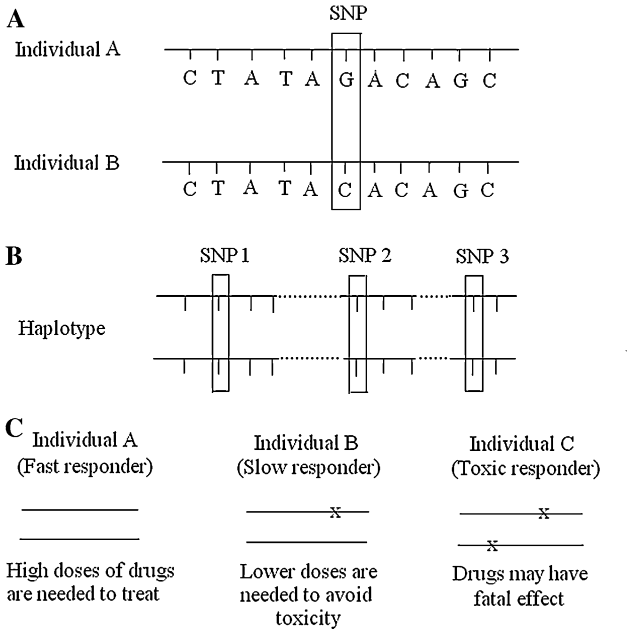


Fig. 1A–C A schematic representation of single nucleotide polymorphism (*SNP*) (**A**), a haplotype (**B**), and the relationship between the genotype and variation in drug response among three individuals (**C**). In **A**, strings of nucleotides at which individuals *A* and *B* differ are shown. In **B**, a long stretch of DNA with distinctive patterns of SNPs at a given location of a chromosome is shown. Haplotype diversity may be generated by new SNP alleles. In **C**, two *horizontal lines* denote a pair of homologous genes and the symbol *X* indicates polymorphism in the gene

have undertaken massive efforts to develop high-throughput SNP genotyping methods over the past 20 years (reviewed in Shastry 2002, 2005). As a result of these efforts, a large collection of SNPs is now available from the human genome project (http://www.ncbi.nlm.nih.gov/SNP/snp_summary.cgi).

SNPs in gene discovery

Because some diseases are hereditary, one immediate goal of the human genome project is to find out which genes predispose people to various disorders and how the sequence variation in a gene affects the functions of its product. As mentioned previously, SNPs occur frequently throughout the genome. Therefore, they can be used as markers to identify disease-causing genes by an association study (Gray et al. 2000). In such studies, it is assumed that two closely located alleles (gene and marker) are inherited together. Therefore, a simple comparison of patterns of genetic variations between patients and normal individuals may provide a method of identifying the loci responsible for disease susceptibility (Hirschhorn and Daly 2005). One advantage of this method is that it does not need a large family. However, several limitations such as population structure, different levels of linkage disequilibrium (LD) (see below) in loci, and epistatic interaction of alleles may impose difficulties. Despite these limitations, there has been some success in identifying the association between polymorphisms and diseases (Table 1).

Unfortunately, however, this type of whole-genome approach to mapping requires the genotyping of thousands of samples. Although there are several high-throughput methods that are available for these studies, they are expensive, laborious and cannot be undertaken by many laboratories. Therefore, a different procedure called haplotype (collection of SNPs on a single chromosome at a locus that is inherited in blocks, Fig. 1B) analysis has been used to identify common disease genes (Hirschhorn and Daly 2005). Because the genome undergoes recombination involving large stretches of DNA, there may be several SNPs linked together in this large region of DNA. These closely linked SNPs may then be cotransmitted from generation to generation in these large blocks (Reich et al. 2001). This phenomenon is called LD. In this type of analysis, multiple genotypes are reduced to haplotypes and hence only a small number of SNPs are required to map the disease gene. Therefore, this method can be more effective in gene mapping and can also provide substantial statistical power in association studies (McVean et al. 2005). However, for conducting genetic association studies, putative polymorphism must be validated. The deleterious effect of SNPs should be evaluated in the context of a relevant haplotype because it is more accurate than a single SNP (Fig. 2). Additionally, previous studies have shown that in the human genome many variants may be common to all populations and others may have a very restricted distribution (Salisbury et al. 2003). Hence its use in disease gene mapping requires additional research. In this regard, the recently characterized second type of DNA variation, called copy number polymorphisms (CNVs), which show marked variations among populations and individuals, may be helpful (Sharp et al. 2005; Locke et al. 2006).

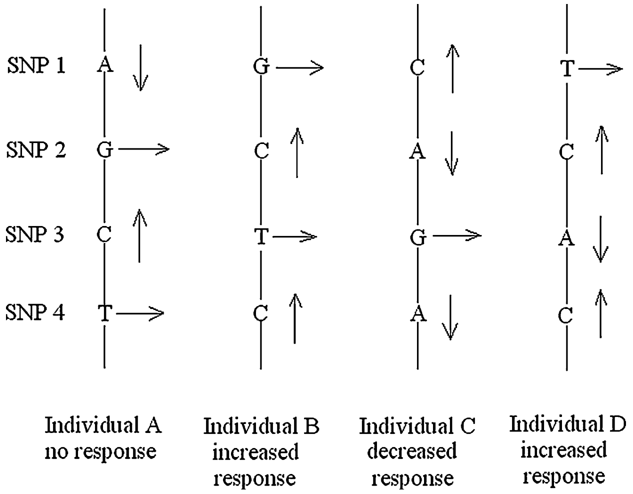

SNP 1
SNP 2
SNP 3
SNP 4
A
G
C
T
G
C
T
C
C
A
G
A
T
C
A
C
Individual A no response
Individual B increased response
Individual C decreased response
Individual D increased response

Fig. 2 A hypothetical haplotype and drug response. Individual *A* shows no response, individuals *B* and *D* show an increased response, whereas individual *C* shows a decreased response to a drug. The *horizontal arrows* denote no change in gene activity, whereas *upward* and *downward arrows* represent increased and decreased gene activity, respectively. Haplotype analysis may give a more accurate prediction of drug response

Pharmacogenetics and pharmacogenomics

Single nucleotide polymorphism technologies are also applicable in the development of individualized medicine. Over the past 20 years it has become increasingly clear that genetic polymorphism in genes encoding drug-metabolizing enzymes, drug transporters and receptors contribute, at least in part, to the inter-individual variability in drug response (reviewed in Shastry 2003, 2004; Evans and Johnson 2001). These factors affect drug absorption, distribution, metabolism and excretion. As a result, some drugs work better in some patients than others, and some drugs may be highly toxic to certain patients (Ansari and Krajinovic 2007). This type of anti-drug reaction has been observed in several diseases, such as pulmonary hypertension, epilepsy, cardiac arrhythmia, renal cell carcinoma, leukemia and liver cancer (reviewed in Shastry 2006a, 2006b; Roses 2000). In order to understand the relationship between heritable changes in genes and

Table 2 A partial list of genetic polymorphisms associated with drug response

Gene	Name	Variant	Phenotypic effect
CYP 3A5	Cytochrome P-450	Splicing	Severely reduced activity
CYP 3A4	Cytochrome P-450	F189S	Reduced activity
LIP C	Hepatic lipase	C-514T	C/C genotype shows increased response to statin, variation in HDL-C levels
MTHFR	Methylene tetra-hydrofolate reductase	C677T	T/T genotype shows increased toxicity
HTR2A	Serotonin receptor 2A	H452Y	Reduced response to clozapine
ABCB1	p-Glycoprotein 1	C3435T	T/T patients may have less drug resistance
IL-10	Interleukin 10	A1082G	G/G individuals have better response to prednisone
GSTP1	Glutathione *S*-transferase	I105V	Increased survival for 5-fluorouracil
AGT	Angiotensinogen	M235T	Reduction of blood pressure
ADD1	Adducin 1	G460W	Increased response to diuretic in hypertensive patients
ABCG2	Efflux transporter of gefitinib	C421A	Gefitinib accumulates in heterozygotes
CDA	Cytidine deaminase	G208A	Severe drug toxicity to gemcitabine

inter-individual variations to drug response, two related fields namely pharmacogenetics and pharmacogenomics (Dervieux and Bala 2006) emerged and gained popularity in the late 1990s. They have undertaken massive studies on the genetic personalization of drug response (McLeod and Evans 2001). As a result, there are several high-density SNP maps of genes encoding proteins of medical importance (Iida et al. 2002, 2003), and now there is strong evidence (Table 2) that links SNPs to inter-individual differences in drug response (Nothen and Cichon 2002; Ansari and Krajinovic 2007). Patients with more active drug-metabolizing enzymes may require higher doses of the drug, and those who do not have an active enzyme may exhibit toxicity (Fig. 1C).

However, it should be noted that there are many negative results regarding the association of gene polymorphism and drug efficacy and toxicity. In addition to these negative results there are other problems. For example, drug metabolism or variation in drug response includes dozens of genes, and many of these often have multiple polymorphisms. Moreover, there are inducible genes, signaling molecules and environmental factors that may also contribute to variable drug response. The greatest challenge for the future (Roden et al. 2006) is to understand the genotypic–environmental factor interaction, ethnicity, inheritance patterns in drug response and how genetic variance responds to medicine. If the goals of pharmacogenetics and pharmacogenomics are fulfilled, it may allow clinicians to genetically subdivide and profile individual patients and treat each patient according to their genetic make-up. This type of medical practice may gradually replace the current trial-and-error-based selection of medicine in the future. However, at present, studies do not unambiguously prove the clinical value of pharmacogenetic testing.

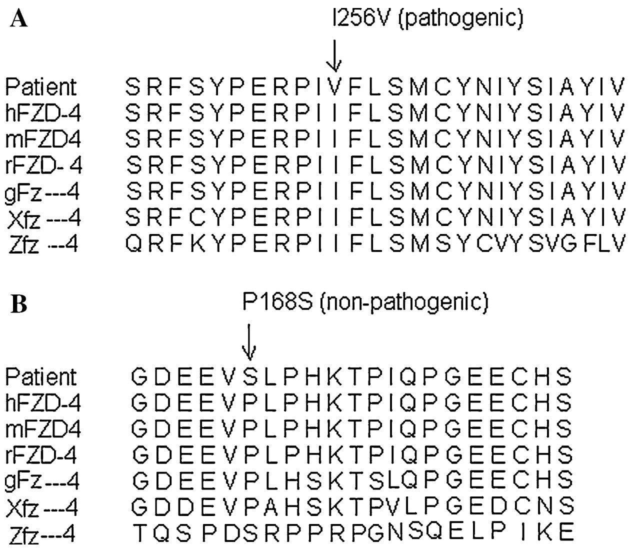
Nutrigenetics and nutrigenomics

It is well known that certain monogenic disorders are associated with the interaction between variant genes and nutrients. The best example is phenylketonuria. Several recent population-based and intervention studies (Subbiah 2007; Ordovas and Mooser 2004; Ordovas et al. 2002a) support this gene–nutrient interaction. Polymorphism on its own may not have an effect, but nutrients may modulate the expression of genes. For instance, a significant variation in the low-density lipoprotein cholesterol level is shown to be associated with A-204C variant in CYP7 (Couture et al. 1999) gene, and high-density cholesterol concentration is determined by polymorphism C-514T in the hepatic lipase gene (Zhang et al. 2005; Couture et al. 2000). Similarly, the high-density lipoprotein cholesterol level is modulated by apolipoprotein A1 (APOA1) genetic polymorphism (-75G/A) in the promoter region (Ordovas et al. 2002b). Therefore, an understanding of the genetic make-up of an individual may lead to the development of an individualized diet. This may reduce diet-related disease risk more efficiently in some common multifactorial disorders (Ordovas and Mooser 2004). Because of this important relationship between gene–nutrient interactions and human health, two recently developed multidisciplinary fields, namely nutrigenetics and nutrigenomics, are exploring the possibility of developing personalized diets based on the genetic make-up of an individual.

SNPs in evolution

Genetic variants are not only considered to be responsible for disease risk and inter-individual differences, but also molecular evolution. Genetic evolution in part depends upon a balance between natural selection and environmentally driven mutation. The natural selection will maintain and retain the amino acid type and position among species because these amino acids are critical for the protein function. Therefore, in a given set of homologous genes, certain amino acids are highly conserved, even among distantly related species that diverged hundreds of million years ago (Fig. 3). These conserved residues are evolving under strong selective pressure. Deleterious mutations that affect the biological functions of proteins are effectively eliminated by natural selection from the gene pool. The selection pressure against deleterious SNPs depends upon the molecular functions of proteins and those genes that encode transcriptional regulatory proteins are generally found to be under the strongest selective pressure (Ramensky et al. 2002).

Because SNPs are present at all levels of evolution, including the branch point of speciation, they can be used to study sequence variation among species. Additionally, the rate, type and site of substitution as well as the selection pressure on codons are not uniform throughout the given gene. Therefore, if genetic variants are fixed during evolution, then they may have either selection advantages for the organism, they may be neutral regarding the fitness, or they may be deleterious and thus cause pathology. Hence, a comparative genomic study of disease-associated SNPs can be used to understand the relationship between the pathology and evolution.



```
A                    I256V (pathogenic)
                              ↓
Patient    S R F S Y P E R P I V F L S M C Y N I Y S I A Y I V
hFZD-4     S R F S Y P E R P I I F L S M C Y N I Y S I A Y I V
mFZD4      S R F S Y P E R P I I F L S M C Y N I Y S I A Y I V
rFZD- 4    S R F S Y P E R P I I F L S M C Y N I Y S I A Y I V
gFz---4    S R F S Y P E R P I I F L S M C Y N I Y S I A Y I V
Xfz ---4   S R F C Y P E R P I I F L S M C Y N I Y S I A Y I V
Zfz --4    Q R F K Y P E R P I I F L S M S Y C V Y S V G F L V

B               P168S (non-pathogenic)
                     ↓
Patient    G D E E V S L P H K T P I Q P G E E C H S
hFZD-4     G D E E V P L P H K T P I Q P G E E C H S
mFZD4      G D E E V P L P H K T P I Q P G E E C H S
rFZD-4     G D E E V P L P H K T P I Q P G E E C H S
gFz---4    G D E E V P L H S K T S L Q P G E E C H S
Xfz---4    G D D E V P A H S K T P V L P G E D C N S
Zfz---4    T Q S P D S R P P R P G N S Q E L P I K E
```

Fig. 3A–B Protein sequence alignment of the mutant part of the human frizzled-4 with that of other species. A conservative change in codon 256 causes pathology in the patient (**A**) whereas a radical change in a less conserved residue is nonpathogenic (**B**). *h*, human; *m*, mouse; *r*, rat; *X*, xenopus; *Z*, zebra fish; and *g*, chicken

Evolutionarily conserved regions more frequently contain disease-associated SNPs

The retention of variants by natural selection is considered to be an important step in evolution. According to the neutral theory of evolution (Kimura 1983), those amino acids that vary among species or those SNPs that do not occur in protein coding regions are either not subjected to natural selection or are under less selective pressure. This is because such amino acid changes can be tolerated and they only minimally affect protein functions. This may imply that such amino acids are more stable and less mutable. However, nSNPs in the coding regions of human genes may have phenotypic effects (Bao and Cui 2005) and may undergo natural selection. Hence, by comparing the rate ratio (omega) of nonsynonymous to synonymous changes (which is considered to be a measure of selective pressure on amino acid replacement mutations) in several proteins from several different species, raw evolutionary data can be generated.

In protein coding genes, patterns of selection can be inferred from amino acid substitution patterns (Jiang and Zhao 2006). One interesting example used to illustrate this is the patterns of distribution of disease-associated and nonpathogenic mutations in human genes. For instance, by using disease-associated mutation data and multiple species of phylogenetic lineage, it has been shown that disease-associated substitution (DAS) occurs more frequently in evolutionarily conserved positions (nonrandom distribution) than in positions that are undergoing variation (Subramanian and Kumar 2006; Miller and Kumar 2001). On the other hand, the opposite trend has been observed for silent and polymorphic mutations, and these are randomly distributed. These patterns are reinforcing the logic that the conserved region of the protein is under evolutionary pressure because these amino acids are critical to the proper functioning of a given gene. On the other hand, silent mutations have minimal affects on the organism because of their random distribution and may not be subjected to natural selection. However, polymorphic mutations of variable amino acids (nonconserved) may have moderate deleterious effects on the organism, and it is likely that such affects are tolerated and hence evolution may be more relaxed in these nucleotides. Similarly, a comparative study between human and chimpanzee genome indicates that some of the human specific traits could be due to positive selection, whereas loci for complex disorders could involve negative selection (Kehrer-Sawatzki and Cooper 2007; Patterson et al. 2006).

Another simple example is the myostatin gene (a negative regulator of skeletal muscle growth). When two different human populations and other mammals are compared for the myostatin gene (Saunders et al. 2006), the number of highly conserved replacement mutations over the evolutionary time scale is greater (five) than the number of silent mutations (three). These data suggest a positive natural selection in the highly conserved region of the myostatin gene because, according to the neutral model of molecular evolution, the ratio of replacement to silent changes does not differ within and between species. However, at present it is not known what types of specific traits are associated with these five replacement changes and what kinds of selection advantages they may have for the species. Additional studies in the future may provide some answers to these questions.

Substitution patterns in the regulatory regions of DNA and noncoding RNA

Natural selection not only operates on protein coding genes but also at the RNA level and on the noncoding regions (regulatory regions) of the DNA. A similar distribution pattern to that discussed above has also been observed for DAS and SNPs in regulatory regions of genes and in RNAs that do not code for proteins (Keightley and Gaffney 2003). For example, it is estimated that at the genomic level, the deleterious point mutation rate is similar between noncoding and coding DNA. Moreover, deleterious mutations in noncoding DNA have quantitative effects, which means that these variations can produce complex genetic diseases (Keightley and Gaffney 2003). Similarly, using SNP genotypic data, it has been shown that negative selection in humans is stronger on conserved microRNA (miRNA binds to the target sites in the 3′-untranslated region of mRNA to repress the translation) binding sites than on other conserved sequence motifs in the 3′-untranslated regions (Saunders et al. 2007). This illustrates the importance of miRNAs to Darwinian fitness (Chen and Rajewsky 2006). Interestingly, a comparison between the miRNA and target sites shows a relatively low level of variation in the functional regions of miRNA and an appreciable level of variation in target sites. Some of these SNPs create novel target sites for miRNA and are found at relatively high frequencies in human populations. If some of these variants have functional effects, they may be involved in phenotypic differences and hence may undergo positive selection. Similarly, an evolutionary comparison using entire classes of mammalian sequences has provided other evidence for the relationship between the pathogenicity of RMRP (RNA component of the mitochondrial RNA processing ribonuclease) mutations and evolutionarily conserved sequences (Bonafe et al. 2005). Although this RNA does not code for a protein, some regions of the RNA are critical to protein binding. The encoding gene is remarkably conserved between species, but disease-causing mutations are once again found in highly conserved nucleotides whereas nonpathogenic variants are located in the nonconserved positions (evolution is more relaxed in these nucleotides). This is consistent with other examples discussed above for the protein coding genes.

Selection pressure is not uniform at amino acid sites

It should also be noted that there are differences in types of amino acid substitutions between species and diseases (Yang et al. 2000). For instance, among species glutamic acid is most commonly replaced by an aspartic acid (very similar) and phenylalanine is replaced by tyrosine. However, this trend has not been observed in disease. When a total of 4,236 mutations in 436 genes causing Mendelian disease (monogenic etiology) and 1,037 synonymous and nSNPs in 313 human genes are compared, a significantly larger contribution at arginine and glycine (also to some extent lysine) is observed in human genetic diseases (Vitkup et al. 2003). This is not the type of change accepted by natural selection. Additionally, a random mutation at tryptophan or cystein residues has the highest probability of causing a disease. This is in agreement with our understanding of their highest evolutionary contribution, which is nothing but their (trp and cys) involvement in determining the protein stability. Thus, selection pressure is not uniform among codons (Arbiza et al. 2006), and in many cases whenever a highly conserved codon is mutated it causes pathology (Fig. 3A).

Radical and less radical SNPs cause early- and late-onset diseases, respectively

Similar to the difference in types of amino acid substitutions between species and diseases, selection pressure also varies between species and disease depending on the properties of amino acids. Those amino acids that have larger chemical difference (radical) are more likely to produce disease phenotypes than those with smaller chemical properties (less radical). Amino acids that have smaller chemical properties are mostly observed among species. As mentioned above, it is the mutation with the larger chemical difference that is most likely to be removed from the population over a long period of time because they are likely to be deleterious. On the other hand, radical changes in variable positions (Fig. 3B) are more likely tolerated (they may not have large effects on protein functions) than in highly conserved positions. Hence these positions do not undergo strong selection. Interestingly, early-onset diseases (they are more damaging) are found to be associated with more radical amino acid mutations, and as a consequence these positions are expected to undergo strong selection. In the same way, late-onset diseases are associated with less radical amino acid mutations and they are not abundant in evolutionarily conserved positions. These less radical amino acid mutations are often associated with common diseases such as diabetes and hypertension. Because they are involved in late-onset diseases, they may have smaller effect on fertility and hence these positions may not undergo strong natural selection. In short, comparative genomic studies between homologous gene sequences from both closely and distantly related species predict that evolution and DAS (pathology) are interrelated. Those residues that evolve under strong selective pressures are likely to be significantly associated with human disease (Arbiza et al. 2006). These types of studies also give us some understanding of the types of variations that can be tolerated in a given gene over time.

Substitution patterns and rates at the chromosome level

Although a lengthy discussion on this subject is not intended in this article, it is relevant to add that the evolutionary rates across the human chromosome are also not constant (Prendergast et al. 2007). Previous studies have predicted a relatively constant mutation rate across mammalian genomes. However, a recent analysis of human–mouse alignment suggests an approximately threefold difference in substitution rates across chromosomes. One of the factors that are found to be associated with mutation rates is the chromatin structure. The human genome contains two types of chromatin structures—closed and open. The open regions of genome are gene-dense and closed regions are relatively gene-poor (Gilbert et al. 2004). Housekeeping and tissue-specific genes are generally found in the more open and most closed regions of the genome, respectively. According to a recent study, the density of SNPs is higher in the most closed regions of the human genome, and genes in these regions also show the highest level of selection at synonymous sites. In fact, the average rate of nonsynonymous changes (d*N*) observed in human–mouse alignments is much higher in the most closed chromatin region of the genome than in the most open regions. Similarly, the ratio of nonsynonymous to synonymous substitution rates (d*N*/d*S*) is also higher, which indicates a strong selection. On the other hand, genes in the regions of open chromatin display the lowest mutation rates and the least constraints at the synonymous sites. However, the average synonymous rate (d*S*) for genes in relatively open chromatin is higher than that for genes in a closed chromatin structure. One of the explanations suggested by researchers for the lesser constraint in the regions of open chromatin is that open regions may be more accessible to repair mechanisms. On the other hand, as mentioned earlier, changes at synonymous sites do not affect the encoded amino acids. Therefore, a synonymous site would have to undergo relatively strong selection to evolve in a non-neutral condition. It is also possible that synonymous sites may experience constraints because they may have a role in RNA stability or splicing.

Fitness, gene pool and functional redundancy

These types of SNPs studies (comparison of relative fixation rates of silent and nSNPs) may allow us to trace the branch point of an evolutionary tree. At this branch point, the variants must have become advantageous for the species and fixed in the gene pool (Zhang et al. 2006). According to comparative genomics, those sequences that contribute to the fitness of an organism evolve slowly. For example, selenoproteins play an important role in antioxidant defense. When polymorphisms of six genes, namely glutathione peroxidases (GPX1, GPX2, GPX3, GPX4), thioredoxine reductase 1 (TXNRD1) and selenoprotein P (SEPP1), were compared in 102 individual populations representing four major ethnic groups, evidence for positive selection was found at the GPX1 locus (Foster et al. 2006). However, in the remaining five genes there was no strong evidence for selection and hence they must have adopted the neutral equilibrium model of evolution. This may imply that they are functionally redundant. It is not clear at present whether this selective pressure on GPX1 is exerted to protect the genome from damaging oxidants or to reduce susceptibility to oxidative stress in erythrocytes, where it is mostly expressed, or both.

Similarly, the ability to digest lactose (present in milk) usually disappears in childhood in most human populations. However, in European-derived populations, lactase activity persists into adulthood. This type of lactase persistence could be due to multiple causes and it may also depend on the population under study. One interesting finding, however, is that, when a region of 3.2 Mbp around the lactase gene consisting of 101 SNPs were typed in northern European and African populations, two alleles were found to be tightly associated with lactase persistence (Trishkoff et al. 2007; Coelho et al. 2005; Bersaglieri et al. 2004). This association could be due to a strong positive selection because of animal domestication and adult milk consumption (advantages to the organism), and hence it is fixed in the gene pool. In contrast, the human mannose binding lectin (MBL-2) allele (a member of the collectin protein family that binds a broad range of microorganisms) occurs at a high frequency worldwide (Verdu et al. 2006). This allele produces little or no protein and was shown to result from human migration and genetic drifts. This evolutionary neutrality (with respect to fitness) of MBL-2 may also suggest that the MBL-2 allele is functionally redundant in the host human defense.

Additional factors that may also contribute to the evolution of the human genome may include DNA methylation, genome duplication, deletions, insertions and the presence of introns (Tang et al. 2006). Insertions and deletions are collectively known as indels and they are approximately 300 bp in length. Because of their high frequency and wide distribution, indels are considered to be the strong driving force of evolution. In addition, the distribution patterns of DAS and nSNPs also show that positions that have many indels in other species contain more nSNPs than DAS. This is not due to the mutation rate, because an excess of nSNPs would be expected in positions with many indels if it was, and that is not found to be the case. Future studies using a recently characterized second type of DNA variation, called CNVs, which show marked variations among populations and individuals, may be helpful (Sharp et al. 2005; Locke et al. 2006) in understanding genetic diversity and evolution.

Concluding remarks

After the *First International Meeting on SNPs* in 1998, it was realized that SNP technologies may have an impact on healthcare. There is no doubt that clinicians, geneticists, patients and the public will benefit from the identification of genes underlying polygenic diseases and adverse drug reactions. Over the past ten years, tremendous progress has been made in cataloging human sequence variations since this high-density map will provide the necessary tools to develop genetically based diagnostic and therapeutic tests. When more functional polymorphisms have been identified, it may be possible to develop useful genetic markers as well as personalized medicines. If the concept of individualized medicine becomes more realistic, every newborn child in the neonatal unit may be genotyped in a routine procedure (similar to a blood transfusion procedure) for improved treatment. The newly developed fields of toxicogenomics, pharmacogenetics and nutrigenetics are rapidly advancing to achieve their goals.

Another interesting aspect of SNPs is that they can also be used to understand the molecular mechanisms of sequence evolution. Natural selection will maintain the amino acid type and retain the amino acid position among species because these amino acids are critical to protein function. Deleterious mutations that affect the biological functions of proteins are effectively being eliminated by natural selection from the gene pool. As discussed above, there is a clear evolutionary relationship between the positions and types of neutral and DAS in the human genome. Residues that evolve under strong selective pressure are found to be significantly associated with human diseases. These patterns are clearly different among species. In short, nucleotide substitutions that are fixed during evolution are either in some way advantageous for the organism, remain neutral regarding fitness, or become deleterious and thus cause pathology. Therefore, evolution and disease-causing nucleotide substitutions can be considered to be related to one another. In the future, it is hoped that research will uncover methods of making SNP markers useful tags for medical testing. Finding out how SNPs affect the health of an individual and then transforming this knowledge into the development of new medicines will undoubtedly revolutionize the treatments of the most common devastating disorders. At the same time, this knowledge will also help us to uncover the secrets of human genome evolution.

Acknowledgments My apologies to those whose work or original publications could not be cited in this short article because of space limitations.

References

Ansari M, Krajinovic M (2007) Pharmacogenomics in cancer treatment defining genetic bases for inter-individual differences in responses to chemotherapy. Curr Opin Pediatr 19:15–22

Arbiza L, Duchi S, Montaner D, Burguet J, Pantoga-Uceda D, Pineda-Lucena A, Dopazo J, Dopazo H (2006) Selective pressures at a codon level predict deleterious mutations in human disease genes. J Mol Biol 358:1390–1404

Bao L, Cui Y (2005) Prediction of the phenotypic effects of non-synonymous single nucleotide polymorphisms using structural and evolutionary information. Bioinformatics 21:2185–2190

Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, Rhodes M, Reich DE, Hirschhorn JN (2004) Genetic signatures of strong recent positive selection at the lactase gene. Am J Hum Genet 74:1111–1120

Bonafe L, Dermitzakis ET, Unger S, Greenberg CR, Campos-Xavier BA, Zankl A, Ucla C, Antonarakis SE, Superti-Furga A, Reymond A (2005) Evolutionary comparison provides evidence for pathogenicity of RMRP mutations. PLoS Genet 1:444–454

Chen K, Rajewsky N (2006) Natural selection on human microRNA binding sites inferred from SNP data. Nat Genet 38:1452–1456

Coelho M, Luiselli D, Bertorelle G, Lopes AT, Seixas S, Destro-Bisol G, Rocha J (2005) Microsatellite variation and evolution of human lactase persistence. Hum Genet 117:329–339

Couture P, Otvos JD, Cupples LA, Wilson PW, Schaefer EJ, Ordovas JM (1999) Association of the A-204C polymorphism in the cholesterol 7 alpha-hydroxylase gene with variations in plasma low-density lipoprotein cholesterol levels in the Framingham offspring. J Lipid Res 40:1883–1889

Couture P, Otvos JD, Cupples LA, Lahoz C, Wilson PW, Schaefer EJ, Ordovas JM (2000) Association of the C-514T polymorphism in the hepatic lipase gene with variations in lipoprotein subclass profiles: The Framingham Offspring Study. Arterioscler Thromb Vasc Biol 20:815–822

Dervieux T, Bala MV (2006) Overview of the pharmacoeconomics of pharmacogenetics. Pharmacogenomics 7:1175–1184

Emilien G, Ponchon M, Caldas C, Isacson O, Maloteaux J-M (2000) Impact of genomics on drug discovery and clinical medicine. Q J Med 93:391–423

Evans WE, Johnson JA (2001) Pharmacogenomics: the inherited basis for interindividual differences in drug response. Annu Rev Genomics Hum Genet 2:9–39

Foster CB, Aswath K, Chanock SJ, McKay HF, Peters U (2006) Polymorphism analysis of six selenoprotein genes: support for a selective sweep at the glutathione peroxidase 1 locus (3p21) in Asian populations. BMC Genet 7:56–63

Gilbert N, Boyle S, Fiegler H, Woodfine K, Carter NP, Bickmore WA (2004) Chromatin architecture of the human genome: gene rich domains are enriched in open chromatin fibers. Cell 118:555–566

Gray IC, Campbell DA, Spurr NK (2000) Single nucleotide polymorphism as tools in human genetics. Hum Mol Genet 9:2403–2408

Halushka MK, Fan JB, Bentley K, Hsie L, Shen NP, Weder A, Cooper R, Lipshutz R, Chakravarti A (1999) Patterns of single nucleotide polymorphisms in candidate genes for blood pressure homeostasis. Nat Genet 22:239–247

Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6:95–108

Iida A, Saito S, Sekine A, Mishima C, Kitamura Y, Kondo K, Harigae S, Osawa S, Nakamura Y (2002) Catalog of 605 single nucleotide polymorphisms (SNPs) among 13 genes encoding human ATP binding cassettes transporters: ABCA4, ABCA7, ABCA8, ABCD1, ABCD2, ABCD4, ABCE1, ABCF1, ABCG1, ABCG2, ABCG4, ABCG5 and ABCG8. J Hum Genet 47:285–310

Iida A, Saito S, Sekine A, Mishima C, Kitamura Y, Kondo K, Harigae S, Osawa S, Nakamura Y (2003) Catalog of 668 SNPs detected among 31 genes encoding potential drug targets on the cell surface. J Hum Genet 48:23–46

Jiang C, Zhao Z (2006) Mutational spectrum in the recent human genome inferred by single nucleotide polymorphisms. Genomics 88:527–534

Kehrer-Sawatzki H, Cooper DN (2007) Understanding the recent evolution of the human genome: insights from human–chimpanzee genome comparisons. Hum Mutat 28:99–130

Keightley PD, Gaffney DJ (2003) Functional constraints and frequency of deleterious mutations in non-coding DNA of rodents. Proc Natl Acad Sci USA 100:13402–13406

Kimchi-Sarfaty C, Oh JM, Kim I-W, Kim IW, Sauna ZE, Calcagno AM, Ambudkar SV, Gottesman MM (2007) A silent polymorphism in the MDR1 gene changes substrate specificity. Science 315:525–528

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge, UK

Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, Cheng Z, Schwartz S, Albertson DG, Pinkel D, Altshuler DM, Eichler EE (2006) Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. Am J Hum Genet 79:275–290

Martin N, Boomsma D, Machin G (1997) A twin pronged attacks on complex traits. Nat Genet 17:387–392

McLeod HL, Evans WE (2001) Pharmacogenomics: unlocking the human genome for better drug therapy. Annu Rev Pharmacol Toxicol 41:101–121

McVean G, Spencer CCA, Chaix R (2005) Perspectives on human genetic variation from the HapMap Project. PLoS Genet 1:413–418

Miller MP, Kumar S (2001) Understanding human disease mutations through the use of interspecific genetic variation. Hum Mol Genet 10:2319–2328

Nothen MM, Cichon S (2002) Linking single nucleotide polymorphisms. Pharmacogenetics 12:89–90

Ordovas JM, Mooser V (2004) Nutrigenomics and nutrigenetics. Curr Opin Lipidol 15:101–108

Ordovas JM, Corella D, Demissie S, Cupples LA, Couture P, Coltell O, Wilson PW, Schaefer EJ, Tucker KL (2002a) Dietary fat intake determines the effect of a common polymorphism in the hepatic lipase gene promoter on high-density lipoprotein metabolism: evidence of a strong dose effect in this gene–nutrient interaction in the Framingham study. Circulation 106:2315–2321

Ordovas JM, Corella D, Cupples LA, Demissie S, Kelleher A, Coltell O, Wilson PW, Schaefer EJ, Tucker K (2002b) Polyunsaturated fatty acids modulates the effects of the APOA1G-A polymorphism on HDL-cholesterol concentrations in a sex-specific manner: Framingham study. Am J Clin Nutr 75:38–46

Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D (2006) Genetic evidence for complex speciation of humans and chimpanzees. Nature 441:1103–1108

Prendergast JGD, Campbell H, Gilbert N, Dunlop MG, Biackmore WA, Semple CAM (2007) Chromatin structure and evolution in the human genome. BMC Evol Biol 7:72–84

Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. Nucleic Acids Res 30:3894–3900

Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumijian R, Farhadian SF, Ward R, Landers ES (2001) Linkage disequilibrium in the human genome. Nature 411:199–204

Roden DM, Altman RB, Benowitz NL, Flockhart DA, Giacomini KM, Johnson JA, Krauss RM, McLeod HL, Ratain MJU, Relling MV, Ring HZ, Shuldiner AR, Weinshilboum RM, Weiss ST (2006) Pharmacogenomics: challenges and opportunities. Ann Intern Med 145:749–757

Roses AD (2000) Pharmacogenetics and the practice of medicine. Nature 405:857–865

Salisbury BA, Pungliya M, Choi JY, Jiang RH, Sun XJ, Stephens JC (2003) SNP and haplotype variation in the human genome. Mutat Res-Fund Mol Mech Mutagenesis 526:53–61

Saunders MA, Good JM, Lawrence EC, Ferrell RE, Li W-H, Nachman MW (2006) Human adaptive evolution of myostatin (GDF8), a regulator of muscle growth. Am J Hum Genet 79:1089–1097

Saunders MA, Liang H, Li WH (2007) Human polymorphism at microRNAs and microRNA target sites. Proc Natl Acad Sci USA 104:3300–3305

Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, Oseroff VV, Albertson DG, Pinkel D, Eichler EE (2005) Segmental duplications and copy-number variation in the human genome. Am J Hum Genet 77:78–88

Shastry BS (2002) SNP alleles in human disease and evolution. J Hum Genet 47:561–566

Shastry BS (2003) SNPs and haplotypes: genetics markers for disease and drug response. Int J Mol Med 11:379–382

Shastry BS (2004) Role of SNP/haplotype map in gene discovery and drug development: an overview. Drug Dev Res 62:142–150

Shastry BS (2005) Genetic diversity and new therapeutic concepts. J Hum Genet 50:321–328

Shastry BS (2006a) Pharmacogenetics and the concept individualized medicine. Pharmacogenomics J 6:16–21

Shastry BS (2006b) Role of SNPs and haplotypes in human disease and drug development. In: Ozkan M, Heller MJ, Ferrari M (eds) Micro/nano technology in genomics and proteomics, vol II. Springer, New York, pp 447–458

Subbiah MT (2007) Nutrigenetics and nutriceuticals: the next wave riding on personalized medicine. Transl Res 149:55–61

Subramanian S, Kumar S (2006) Evolutionary anatomies of position and types of disease associated and neutral amino acid mutations in the human genome. BMC Genomics 7:306–312

Tang CS, Zhao YZ, Smith DK, Epstein RJ (2006) Intron length and accelerated 3′ gene evolution. Genomics 88:682–689

Trishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, Powell K, Mortensen HM, Hirbo JB, Osman M, Ibrahim M, Omar SA, Lema G, Nyambo TB, Ghori J, Bumpstead S, Pritchard JK, Wray GA, Deloukas P (2007) Convergent adaptation of human lactase persistent in Africa and Europe. Nat Genet 39:31–40

Verdu P, Barreiro LB, Patin E, Gessain A, Cassar O, Kidd JR, Kidd KK, Behar DM, Froment A, Heyer E, Sica L, Casanova JL, Abel L, Quintana-Murci L (2006) Evolutionary insights into the high worldwide prevalence of MBL2 deficiency alleles. Hum Mol Genet 15:2650–2658

Vitkup D, Sander C, Church GM (2003) The amino acid mutational spectrum of human genetic disease. Genome Biol 4: R72-R80

Yang Z, Nlelsen R, Goldman N, Pedersen AM (2000) Codon substitution models for heterogeneous selection pressure at amino acid sites. Genomics 155:431–449

Zhang C, Bailey DK, Awad T, Liu G, Xing G, Cao M, Valmeekam V, Retief J, Matsuzaki H, Taub M, Seielstad M, Kennedy GC (2006) A whole genome long-range haplotype (WGLRH) test for detecting imprints of positive selection in human populations. Bioinformatics 22:2122–2128

Zhang C, Lopez-Ridaura R, Rimm EB, Rifai N, Hunter DJ, Hu FB (2005) Interaction between the -514 C to T polymorphism of the hepatic lipase gene and life-style factors in relation to HDL concentrations among US diabetic men. Am J Clin Nutr 81:1429–1435