

MINIREVIEW

Barkur S. Shastry**SNP alleles in human disease and evolution**

Received: June 12, 2002 / Accepted: July 26, 2002

Abstract In two randomly selected human genomes, 99.9% of the DNA sequence is identical. The remaining 0.1% of DNA contains sequence variations. The most common type of such variation is called a single-nucleotide polymorphism, or SNP. SNPs are highly abundant, stable, and distributed throughout the genome. These variations are associated with diversity in the population, individuality, susceptibility to diseases, and individual response to medicine. Recently, it has been suggested that SNPs can be used for homogeneity testing and pharmacogenetic studies and to identify and map complex, common diseases such as high blood pressure, diabetes, and heart disease. Consistent with this proposal is the identification of the patterns of SNPs in conditions such as diabetes, schizophrenia, and blood-pressure homeostasis. Although these studies have provided insight into the nature of human sequence variation, it is not known at present whether these variations are truly significant toxicologically and pharmacologically. Moreover, it is possible that most complex, common disorders are caused by the combined effects of multigenes and nongenetic environmental factors (multifactorial). Therefore, it is likely that sequence variation alone is not sufficient to predict the risk of disease susceptibility, particularly in homeostatic organisms like humans. Nevertheless, these variants may provide a starting point for further inquiry.

Key words Polymorphism · Allele · Disease · Association study · Pharmacogenetics

Introduction

It is now clear that mapping the entire human genome sequence is a reality and will be completed by the year 2003. Many investigators now have to decide what to do with this

information and how to take advantage of this effort. The majority of them believe that one possible benefit of this information is to use it to understand the genetic basis of the most common familial traits, evolutionary processes, and complex and common diseases such as hypertension, diabetes, obesity, and psychiatric disorders. These common diseases are likely to be caused by multiple genes and multiple nongenetic factors (environmental factors), each contributing a modest effect. Their cumulative effect results in the condition or trait. The traditional methods of identifying disease-related genes are not readily applicable to the detection of genes responsible for these multifactorial diseases. The availability of the entire genome sequence, therefore, may speed up the gene-hunting efforts in the near future, but what approaches are to be taken to accomplish this task and what are their limitations?

The human genome and the discovery of single-nucleotide polymorphisms (SNPs) as genetic markers

In two randomly selected human genomes, 99.9% of the DNA sequence is identical. The remaining 0.1% is thought to include some differences or variations in the genome between individuals. This variation, called polymorphism, arises because of mutations. The simplest form of these variations is the substitution of one single nucleotide for another (Fig. 1A), termed SNP. SNPs are more common than other types of polymorphisms and occur at a frequency of approximately 1 in 1000 base pairs (Brookes 1999) throughout the genome (promoter region, coding sequences, and intronic sequences). These simple changes in DNA sequence, most of which are probably located in intergenic spacers, are believed to be stable and not deleterious to organisms. SNPs that do not change encoded amino acids are called synonymous and are not subject to natural selection (Kimura 1983, snp.cshl.org). On the other hand, nonsynonymous SNPs alter amino acids and might be subject to natural selection. SNPs can be observed between individuals in a population, may influence promoter activity

B.S. Shastry (✉)
Department of Biological Sciences, Oakland University, Rochester,
MI 48309, USA
Tel. +1-248-370-3577; Fax +1-248-370-4225
e-mail: Barkur@aol.com

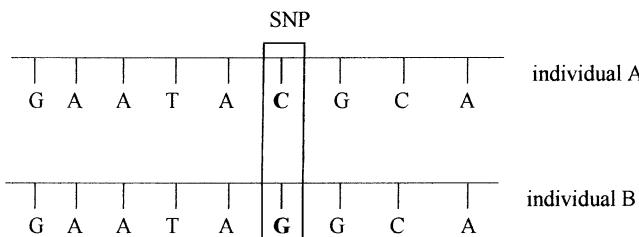
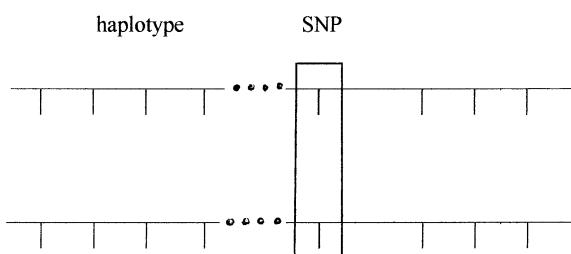
a**b**

Fig. 1. **a** Single-nucleotide polymorphisms (SNPs) are single-base variations among different people. Figure shows strings of nucleotides at which individuals A and B differ by just one base. Individual A may respond to the drug positively, whereas individual B may show an adverse reaction or moderate response or no response at all. **b** A long stretch of DNA (e.g., 100 000 bases) with a distinctive pattern of SNPs at a given location of a chromosome is called haplotype. Haplotype diversity may be generated by new SNP alleles, which can arise because of mutations at different loci

or DNA and pre-mRNA conformation, and play a direct or indirect role in phenotypic expression (Krawezak et al. 1992; Lohrer and Tangen 2000; Pitarque et al. 2001; Spicker et al. 2001; LeVan et al. 2001). Because some SNPs are functional, comparative studies on identical twins, fraternal twins, and siblings suggest that genetic variation is one of the factors associated with susceptibility to many common diseases (Table 1) as well as every human trait such as tallness, curly hair, and individuality (Martin et al. 1997). Diversity in the population is also associated with these variations. Therefore, it may be possible to understand why some individuals are susceptible to common disorders by using the human genome sequence and the variations between individuals. However, there are limitations, and practical and ethical issues must be considered before undertaking such analyses (Chanock 2001).

Mapping and characterization of SNPs

Because most sequence variants are SNPs, a massive effort has been undertaken by several private and public organizations, such as Celera Genomics, Incyte Genomics, the Wellcome Trust Sanger Institute in the United Kingdom, and Washington University in the United States, to generate a high-density SNP map of the genome (Marth et al. 2001; Irizarry et al. 2000; Altshuler et al. 2000; Mullikin et al. 2000; also see the International SNP Map Working Group

Table 1. A partial list of disorders associated with SNPs

Disorder	Gene	Reference
Asthma	EDN 1 and NOS 1	Immervoll et al. 2001
POAG	Mycocillin	Colomb et al. 2001
Systemic sclerosis	Fibrillin 1	Tan et al. 2001
Lung cancer	MMP-1	Zhu et al. 2001
Arrhythmias	KCNQ1	Kubota et al. 2001
Idiopathic arthritis	MIF	Donn et al. 2001
Blood pressure	TAF1	Koschinsky et al. 2001
Biliary cirrhosis	MBL	Matsushita et al. 2001
Type II diabetes	Syntaxin 1A	Tsunoda et al. 2001
Systemic lupus erythematosus	Prolactin	Stevens et al. 2001
Eating disorder	Melanocortin	Adan and Vink 2001
Migraine	Insulin receptor	McCarthy et al. 2001
Ossification	Npps	Koshizuka et al. 2002
Lung cancer	p53	Biros et al. 2001
Late-onset PD	tau	Martin et al. 2001

SNPs, Single-nucleotide polymorphisms; POAG, primary open-angle glaucoma; MMP-1, matrix metalloproteinase 1; EDN 1, endothelin 1; NOS 1, neuronal nitric oxide synthetase 1; MBL, mannose binding protein; Npps, nucleotide pyrophosphatase; TAF1, thrombin activatable fibrinolysis inhibitor; KCNQ1, potassium channel protein; MIF, macrophage inhibitory factor; PD, Parkinson disease

2001 and <http://www.ncbi.nlm.nih.gov/SNP/>). As a result of this worldwide intensive effort, more than 2.8 million SNPs have been identified and a high-density map has been constructed in some cases (Iida et al. 2001a-d; Iwasaki et al. 2001; Saito et al. 2001; Osier et al. 2001). Although several mapping methods (Table 2), such as single-strand conformational polymorphism (Orita et al. 1989), denaturing gradient gel electrophoresis (DGGE), enzymatic mutation detection (Youil et al. 1995), microarray or variant detector arrays (Wang et al. 1998; Marshall and Hodgson 1998; Ramsay 1998; Hacia et al. 1999; Hacia and Collins 1999; Dong et al. 2001; Qi et al. 2001; Yoshino et al. 2001), and heteroduplex analysis (Lichten and Fox 1983) are available, so far none of them has supplanted DNA sequencing as the method of choice. SNP mapping requires a tremendous amount of time and resources (hundreds and thousands of individuals must be studied to eliminate false-positive and false-negative results). Additionally, many of the methods detect only about 80% of mutations. In some methods, mutations in GC-rich regions may not be detected (e.g., DGGE), and other methods involve expensive instruments and kits or toxic chemicals (Table 2). Therefore, new, faster methods must be developed and old methods refined so that, at least, cost is not a key consideration. In this respect, recently developed high-throughput SNP genotyping (Jenkins and Gibson 2002; McClay et al. 2002) and molecular beacon methods (Mhlanga and Malmberg 2001) are highly valuable. According to some estimates, 50% of mutations are likely to be in a noncoding sequence, 25% lead to amino acid substitution, and 25% are silent (Cargill et al. 1999; Halushka et al. 1999). However, it should be noted that the above estimates are only for cSNPs (SNPs found in protein-coding regions) and SNPs in untranslated regions and depend on sample size because they have different allele frequency distributions.

Table 2. A comparison of selected mutation screening methods

Method	Fragment length (bp)	Advantage	Disadvantage	Efficiency (%)
Single-strand conformational polymorphism	~300	No expensive equipment	Small fragments, temperature variation	80
Heteroduplex	300–600	No expensive equipment	Conditions to be determined	80
Denaturing gradient gel electrophoresis	100–1000	Simple, long and short fragments	Gradient gel required, mutation in GC region may not be detected	100 with GC clamps
Enzymatic mismatch detection	300–1000	Long and short fragments	Identifies all kinds of mutations	100
Base excision sequence scanning	50–1000	Accurate	Expensive instruments	100
RNAase cleavage	1.6 kb	Longer fragment and fast	Requires special kit	100
Chemical cleavage	1–2 kb	Large fragment	Multistep, labor intensive, hazardous chemicals	
DNA sequencing	500	Rapid and easy, no additional sequencing	Labor intensive	100

SNPs in gene discovery

Once the map of these SNPs is confirmed, they can be used for evolutionary biology studies, gene discovery and mapping, prediction of drug and environmental response, diagnostic tests, heterogeneity testing, and association studies (Gray et al. 2000; Schork et al. 2000). For the purpose of gene discovery, SNPs are considered to be the most predominant segregating form of variation at the molecular level because of their frequent occurrence throughout the genome, and they can be useful in association studies. However, they are less informative, in the sense that humans have relatively low nucleotide diversity compared with *Drosophila* and maize, than another type of marker called a microsatellite. Microsatellites are simple sequence repeats, the most common classes being dinucleotide, trinucleotide, and tetranucleotide, and they occur at a rate of 1 in every 10 kb in a wide range of eukaryotic genomes. Human microsatellites are used for linkage studies and they average at least ten alleles with heterozygosity per locus over 80% (very high). Although they have a much higher mutation rate than the standard sequence, they are not densely distributed. Linkage mapping focuses on the small number of meiotic events within a family and association between marker alleles and traits. It does not require a very dense map of markers at the initial stage. On the other hand, linkage disequilibrium mapping explores family associations and requires a dense map of markers. In association studies, the marker is prevalent in patients versus those without the disease and this is considered evidence of association between the disease and the marker (SNP). Although there are many limitations (described following), association studies are perhaps the best for mapping of polygenic complex disease loci. However, this type of study requires a large number of patients and an adequate control group to achieve over 80% power to detect a locus. Once a significant site is identified, one can use either a pedigree-based transmission disequilibrium test (TDT), which measures the transmission of alleles from a heterozygous parent to the affected offspring (an unequal transmission of SNP alleles to affected and unaffected siblings), or a case-control

population sampling, which measures the association between SNPs and the disease in a large population (Keavney et al. 2000; Spielman et al. 1992). TDT detects both linkage and association, and SNPs are usually used. These methods are based on the assumption that SNP variants account for population susceptibility to certain disorders; however, it is unknown how many SNPs are needed. These methods have several limitations, such as difficulties associated with population structure, different levels of linkage disequilibrium in loci, allelic and nonallelic heterogeneity of phenotypes, and epistatic interaction of alleles, all of which have been previously discussed by others (Schork et al. 2000; Chakravarti 1999; Weiss and Terwilliger 2000). An allele frequency database for human polymorphic sites for multiple populations (Osier et al. 2001; Hirakawa et al. 2002) can be found on the Kidd lab home page (<http://info.med.yale.edu/genetics/kkidd/>). Additionally, a recently proposed haplotype (a distinct combination of single nucleotide types on a single chromosome at a locus) map (Fig. 1B) of the genome may speed up or simplify the hunt for the association between DNA variations and complex diseases (Helmuth 2001). This is because haplotype diversity, which is greater than SNP diversity, may be generated by new SNP alleles that can arise because of mutation at different loci, and hence can be studied by both linkage and association methods. In haplotype association studies, multitype genotypes are reduced to haplotypes and this has proved to be a more efficient mapping technique than that of SNP analysis.

SNPs in pharmacology

Another potential application of SNPs is the development of individualized medicine. Inherited genetic differences between individuals appear to determine each patient's response to medicine. For instance, some patients show response to the prescribed drug without any serious side effects, whereas others do not respond and experience adverse reactions. In fact, it is estimated that properly prescribed medications cause 2 million Americans to get sick and result in 100 000 deaths each year because of adverse

drug reactions (Pirmohamed and Park 2001; Bader 2001; Wolf et al. 2000; Lazarou et al. 1998). Antitumor agent 6-mercaptopurine, for example, has been used for children with leukemia. Those children who do not have its metabolizing enzyme — thiopurine methyltransferase — show a severe hematopoietic toxicity, whereas children with the more active enzyme require higher doses of the drug. Similar anti-drug reactions were also observed in asthma and other conditions (Drazen et al. 1999; Roses 2000). The most important of the several hundred thousand SNPs in the human genome for this purpose are SNPs in drug-metabolizing enzymes, particularly in the coding and the promoter regions, because changes there may have functional significance (Risch 2000). The SNP method can also be applied to crop (corn and soybean) genetics (Rafalski 2002). Haplotype analysis is possible for crop genetics because of the availability of many genes and expressed sequence tags, as well as their intraspecific nucleotide diversity.

SNPs and evolution

SNPs can be used to study DNA sequence variation among species. Because such variations are present at all levels of evolution, including branching points of speciation, they may provide an understanding of how the modern genome evolved. Most SNPs are not in protein-coding regions but are elsewhere in the genome; therefore, their distribution is not under selective pressure. Variations in the protein-coding region that affect the phenotype might be subject to natural selection, but if these variations were retained in the gene over time, then they must have some benefit for the individual for successful reproduction. The variants that are selected for retention by natural selection may represent an important step in evolution. Thus, by calculating a ratio between variants in noncoding and coding regions of a series of protein families found in different species, it may be possible to trace the branching point of an evolutionary tree. At this branch point, the variant must have become advantageous for the species and hence fixed in the gene pool (Osier et al. 2001; Stephens et al. 2001; Liberles 2001). While this gene pool has been continuously expanding during evolution, it might have resulted in the modern human genome. It is very well known from several studies that humans are similar to chimpanzees at the genomic level. However, there are differences between humans and chimpanzees. For instance, malaria; rheumatoid arthritis; and breast, colon, and lung cancers are extremely rare in chimpanzees but common in humans. In this regard, SNPs may provide important health clues.

Conclusion

There is no doubt that the identification of genes underlying polygenic and complex diseases such as psychiatric disorders, diabetes, hypertension, and asthma is of paramount interest for clinicians, geneticists, patients, and the public.

However, it is not clear at present how genetic variation alone determines the susceptibility of an individual to some diseases or to adverse drug interactions. This is because most common traits and phenotypes are the result of long-term interaction between genetic and nongenetic environmental factors. Factors such as lifestyle and diet may contribute to disease susceptibility by altering gene expression. Frequency of polymorphism may also vary among different populations (Wakeley et al. 2000; Nielsen 2000; Nielsen and Slatkin 2000). Therefore, even after complete sequencing and identification of SNPs in an individual's DNA, it is not simple to associate these variations to disease without knowing the functional significance of the identified SNPs. Additionally, researchers are struggling to understand disease heterogeneity even in monogenic disorders. In multigenic disorders, the contribution of susceptible individual genes to the disorder is very weak. Therefore, genotype alone is not sufficient to predict or of susceptibility to disease (Martin et al. 1997) nor is phenotype variation alone necessarily linked to DNA sequence variation (Lander and Schork 1994; Weeks and Lathrop 1995). On the other hand, finding out how SNPs affect an individual's health and transforming this knowledge into the development of new medicine, which requires the correlation of SNPs with specific diseases, will revolutionize the treatment of most common killer diseases. This is because such an understanding of the relationship between SNPs and the disease will allow clinicians to determine whether an individual will respond to a medicine or experience serious side effects. Drug companies will be able to design different drugs for each patient with similar clinical symptoms or disease phenotypes. In future, this knowledge may give clinicians more insight into the disease and change the definition of some disorders.

References

- Adan RAH, Vink T (2001) Drug target discovery by pharmacogenetics: mutation in the melanocortin system and eating disorders. *Eur Neuropsychopharmacol* 11:483–490
- Altshuler D, Pollara VJ, Cowies CR, Van Etten WJ, Balswin J, Linton L, Lander ES (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature* 407:513–516
- Bader JS (2001) The relative power of SNPs and haplotype as genetic markers for association tests. *Pharmacogenomics* 2:11–24
- Biros E, Kalina I, Biros J, Kohut A, Bogyiova E, Salagovic J, Stubna J (2001) Polymorphism of the p53 gene within the codon 72 in lung cancer patients. *Neoplasma* 48:407–411
- Brookes AJ (1999) The essence of SNPs. *Gene* 234:177–186
- Cargill M, Altshuler D, Ireland J, Skalr P, Ardile K, Patil N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES (1999) Characterization of single nucleotide polymorphism in coding regions of human genes. *Nat Genet* 22:231–238
- Chakravarti A (1999) Population genetics: making sense out of sequence. *Nat Genet* 21:56–60
- Chanock S (2001) Candidate genes and single nucleotide polymorphisms (SNPs) in the study of human disease. *Dis Markers* 17:89–98
- Colomb E, Nguyen TD, Bechettoille A, Dascotte JC, Valtot F, Brezin AP, Berkani M, Copin B, Gomez L, Polansky JR, Garchon HJ (2001) Association of a single nucleotide polymorphism in the

- TIGR/MYOCILIN gene promoter with the severity of primary open-angle glaucoma. *Clin Genet* 60:220–225
- Dong SL, Wang E, Hsie L, Cao YX, Chen XG, Gingeras TR (2001) Flexible use of high density oligonucleotide arrays for single nucleotide polymorphism discovery and validation. *Genome Res* 11:1418–1424
- Donn RP, Shelley E, Olier WER, Thomson W (2001) A novel 5'-flanking region polymorphism of macrophage migration inhibitory factor is associated with systemic-onset juvenile idiopathic arthritis. *Arthritis Rheum* 44:1782–1785
- Drazen JM, Yaudava CN, Dube L, Szczerczak N, Hippenstein R, Pillari A, Israel E, Schork N, Silverman ES, Katz DA, Drazen J (1999) Pharmacogenetic association between ALOX 5 promoter genotype and the response to anti-asthma treatment. *Nat Genet* 22:168–170
- Gray IC, Campbell DA, Spurr NK (2000) Single nucleotide polymorphisms as tools in human genetics. *Hum Mol Genet* 9:2403–2408
- Hacia JG, Collins FS (1999) Mutational analysis using oligonucleotide microarrays. *J Med Genet* 36:730–736
- Hacia JG, Fan JB, Ryder O, Jin L, Edgemon K, Ghandour G, Mayer RA, Sun B, Hsie L, Robbins CM, Brody LC, Wang D, Lander ES, Lipshutz R, Fodor SPA, Collins FS (1999) Determination of ancestral alleles for human single nucleotide polymorphisms using high density oligonucleotide arrays. *Nat Genet* 22:164–167
- 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
- Helmuth L (2001) Map of the human genome 3.0. *Science* 293:583–584
- Hirakawa M, Tanaka T, Hashimoto Y, Kuroda M, Takagi T, Nakamura Y (2002) JSNP: a database of common gene variations in the Japanese population. *Nucleic Acids Res* 30:158–162
- Iida A, Saito S, Sekine A, Kitamoto T, Kitamura Y, Mishima C, Osawa S, Kondo K, Harigae S, Nakamura Y (2001a) Catalog of 434 single nucleotide polymorphisms (SNPs) in genes of the alcohol dehydrogenase, glutathione S-transferase and nicotinamide adenine dinucleotide reduced (NADH) ubiquinone oxidoreductase families. *J Hum Genet* 46:385–407
- Iida A, Saito S, Sekine A, Kitamura Y, Kondo K, Mishima C, Osawa S, Harigae S, Nakamura Y (2001b) High density single nucleotide polymorphisms (SNPs) map of the 150 kb region corresponding to the human ATP binding cassette transporter A1 (ABCA1) gene. *J Hum Genet* 46:522–528
- Iida A, Ohnishi Y, Ozaki K, Ariji Y, Nakamura Y, Tanaka T (2001c) High density single nucleotide polymorphism (SNP) map in the 96 kb region containing the entire human DiGeorge syndrome critical region 2 (*DGCR2*) gene at 22q11.2. *J Hum Genet* 46:604–608
- Iida A, Saito S, Sekine A, Harigae S, Osawa S, Mishima C, Kondo K, Kitamura Y, Nakamura S (2001d) Catalog of 46 single nucleotide polymorphisms (SNPs) in the microsomal glutathione S-transferase 1 (*MGST1*) gene. *J Hum Genet* 46:590–594
- Immervoll T, Loesgen S, Dutsch G, Gohlke H, Herbon N, Klugbauer S, Dempfle A, Bickeboller H, Becker-Follmann J, Ruschendorf F, Saar K, Reis A, Wichmann HE, Wjst M (2001) Fine mapping and single nucleotide polymorphism association results of candidate gene for asthma and related phenotypes. *Hum Mutat* 18:327–336
- Irizarry K, Kustanovich V, Li C, Brown N, Nelson S, Wong W, Lee CJ (2000) Genomewide analysis of single nucleotide polymorphisms in human expressed sequences. *Nat Genet* 26:233–236
- Iwasaki H, Shinohara Y, Ezura Y, Ishida R, Kodaira M, Kajita M, Nakajima T, Shiba T, Emi M (2001) Thirteen single nucleotide polymorphisms in the human osteopontin gene identified by sequencing of the entire gene in Japanese individuals. *J Hum Genet* 46:544–546
- Jenkins S, Gibson N (2002) High-throughput SNP genotyping. *Comp Funct Genomics* 3:57–66
- Keavney B, McKenzie C, Parish S, Palmer A, Clark S, Youngman L, Delapine M, Lathrop M, Peto R, Collins R (2000) Large scale test of hypothesized associations between the angiotensin-converting enzyme insertion/deletion polymorphisms and myocardial infarction in about 5000 cases and 6000 controls. *Lancet* 355:434–442
- Kimura M (1983) The neutral theory of molecular evolution. Cambridge, Cambridge University Press.
- Koschinsky ML, Boffa MB, Nesheim ME, Zinman B, Hanley AJG, Harris SB, Cao H, Hegele RA (2001) Association of a single nucleotide polymorphism in CPB2 encoding the thrombin-activatable fibrinolysis inhibitor (TAF 1) with blood pressure. *Clin Genet* 60:345–349
- Koshizuka Y, Kawaguchi H, Ogata N, Ikeda T, Mabuchi A, Seichi A, Nakamura Y, Nakamura K, Ikegawa S (2002) Nucleotide pyrophosphatase gene polymorphism associated with ossification of the posterior longitudinal ligament of the spine. *J Bone Miner Res* 17:138–144
- Krawezak M, Reiss J, Cooper DN (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet* 90:41–54
- Kubota T, Horie M, Takano M, Yoshida H, Takenaka K, Watanabe E, Tsuchiya T, Otani H, Sasayama S (2001) Evidence for a single nucleotide polymorphism in the KCNQ1 potassium channel that underlies susceptibility to life-threatening arrhythmias. *J Cardiovasc Electrophysiol* 12:1223–1229
- Lander ES, Schork NJ (1994) Genetic dissection of complex traits. *Science* 265:2037–2048
- Lazarou J, Pomeranz BH, Corey PN (1998) Incidence of adverse drug reactions in hospitalized patients: a meta analysis of prospective studies. *JAMA* 279:1200–1205
- LeVan TD, Bloom JW, Bailey TJ, Karp CL, Halonen M, Martinez FD, Vercelli D (2001) A common single nucleotide polymorphism in the CD14 promoter decreases the affinity of Sp protein binding and enhances transcriptional activity. *J Immunol* 167:5838–5844
- Liberles DA (2001) The adaptive evolution database. *Genome Biol* 2:1
- Lichten MJ, Fox MS (1983) Detection of non-homology containing heteroduplex molecules. *Nucleic Acids Res* 11:3959–3971
- Lohrer HD, Tangen U (2000) Investigations into the molecular effects of single nucleotide polymorphism. *Pathobiology* 68:283–290
- Marshall A, Hodgson J (1998) DNA chips: an array of possibilities. *Nat Biotechnol* 16:27–31
- Marth G, Yeh R, Minton D, Donaldson R, Li Q, Duan S, Devenport R, Miller RD, Kwok P-Y (2001) Single nucleotide polymorphisms in the public domain: how useful are they? *Nat Genet* 27:371–372
- Martin N, Boomsma D, Machin G (1997) A twin-pronged attacks on complex traits. *Nat Genet* 17:387–392
- Martin ER, Scott WK, Nance MA, Watts RL, Hubble JP, Koller WC, Lyons K, Pahwa R, Stern MB, Colcher A, Hiner BC, Jankovic J, Ondo WG, Allen FH, Goetz CG, Small GW, Masterman D, Mastaglia F, Laing NG, Stajich JM, Ribble RC, Booze MW, Rogala A, Hauser MA, Zhang FY, Gibson RA, Middleton LT, Roses AD, Haines JL, Scott BL, Pericak-Vance MA, Vance JM (2001) Association of single nucleotide polymorphisms of the tau gene with late-onset Parkinson disease. *JAMA* 286:2245–2250
- Matsushita M, Miyakawa H, Tanaka A, Hijikata M, Kikuchi K, Fujikawa H, Arai J, Sainokami S, Hino K, Terai I, Mishiro S, Gershwin ME (2001) Single nucleotide polymorphism of the mannose-binding lectin are associated with susceptibility to primary biliary cirrhosis. *J Autoimmun* 17:251–257
- McCarthy MC, Hosford DA, Riley JH, Bird MI, White MJ, Hewett DR, Peroutka SJ, Griffiths LR, Boyd PR, Lea RA, Bhatti SM, Hosking LK, Hood CM, Jones KW, Handley AR, Rallan R, Lewis KF, Yeo AJM, Williams PM, Priest RC, Khan P, Donnelly C, Lumsden SM, O'Sullivan J, See CG, Smart DH, Shaw-Hawkins S, Patel J, Langrish TC, Feniuk W, Knowles RG, Thomas M, Libri V, Montgomery DS, Manasco PK, Xu CF, Dykes C, Humphrey PPA, Roses AD, Purvis U (2001) Single nucleotide polymorphism alleles in the insulin receptor gene are associated with typical migraine. *Genomics* 78:135–149
- McClay JL, Sugden K, Koch HG, Higuchi S, Craig IW (2002) High-throughput single nucleotide polymorphisms genotyping by fluorescent competitive allele-specific polymerase chain reaction (SNiPTag). *Anal Biochem* 301:200–206
- Mhlanga MM, Malmberg L (2001) Using molecular beacons to detect single nucleotide polymorphisms with real-time PCR. *Methods* 25:463–471
- Mullikin JC, Hunt SE, Cole CG, Moltimore BJ, Rice CM, Burton J, Matthews LH, Pavitt R, Plumb RW, Sims SK, Ainscough RMR, Attwood J, Bailey JM, Barlow K, Bruskiewich RMM, Butcher PN, Carter NP, Chen Y, Clee CM, Coggill PC, Davies J, Davies RM, Dawson E, Francis MD, Joy AA, Lambel RG, Langford CF,

- Macarthy J, Mall V, Moreland A, Overton-Larty EK, Ross MT, Smith LC, Steward CA, Sulston JE, Tinsley EJ, Turney KJ, Willey DL, Wilson GD, McMurray AA, Dunham I, Rogers J, Bentley DR (2000) An SNP map of human chromosome 22. *Nature* 407:516–520
- Nielsen R (2000) Estimation of population parameters and recombination rates from single nucleotide polymorphisms. *Genetics* 154:931–942
- Nielsen R, Slatkin M (2000) Likelihood analysis of ongoing gene flow and historical association. *Evolution* 54:44–50
- Orita M, Iwahana H, Hayashi K, Sekiya T (1989) Detection of polymorphism of human DNA by gel electrophoresis as single strand conformation polymorphisms. *Proc Natl Acad Sci USA* 86:2766–2770
- Osier MV, Cheung KH, Kidd JR, Pakstis AJ, Miller PL, Kidd KK (2001) ALFRED: an allele frequency database for diverse populations and DNA polymorphism — an update. *Nucleic Acids Res* 29:317–319
- Pirmohamed M, Park BK (2001) Genetic susceptibility to adverse drug reactions. *Trends Pharmacol Sci* 22:298–305
- Pitarque M, vonRichter O, Oke B, Berkkan H, Oscarson M, Ingelman-Sundberg M (2001) Identification of a single nucleotide polymorphism in the TATA box of the *CYP2A6* gene: impairment of its promoter activity. *Biochem Biophys Res Commun* 284:455–460
- Qi XQ, Bakht S, Devos KM, Gale MD, Osbourn A (2001) L-RCA (Ligation rolling circle amplification): a general method for genotyping of single nucleotide polymorphism (SNPs). *Nucleic Acids Res* 29:U68–U74
- Rafalski A (2002) Application of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol* 5:94–100
- Ramsay G (1998) DNA chips: state of the art. *Nat Biotechnol* 16:40–44
- Risch NJ (2000) Searching for genetic determinants in the new millennium. *Nature* 405:847–856
- Roses AD (2000) Pharmacogenetics and the practice of medicine. *Nature* 405:857–865
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok P-Y, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumentiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409:928–933
- Saito S, Iida A, Sekine A, Miura Y, Sakamoto T, Ogawa C, Kawauchi S, Higuchi S, Nakamura Y (2001) Identification of 197 genetic variations in six human methyltransferase genes in the Japanese population. *J Hum Genet* 46:529–537
- Schork NJ, Fallin D, Lanchbury JS (2000) Single nucleotide polymorphism and the future of genetic epidemiology. *Clin Genet* 58:250–264
- Spicker M, Darius H, Hankeln T, Soufi M, Friedl H, Schaefer JR, Schmidt ER, Zeldin DC, Liao JK (2001) Expression of vascular anti-inflammatory eicosanoids producing cytochrome P4502J2 (CYP2J2) is reduced by promoter single nucleotide polymorphism. *Circulation* 104(Suppl):314
- Spielman RS, McGinnis RE, Ewens WJ (1992) Transmission distortion test for linkage: the insulin gene and susceptibility to insulin dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506–516
- Stephens JC, Schneider JA, Tanagay DA, Choi J, Acharya T, Stanley SE, Jiang RH, Messer CJ, Chew A, Han JH, Duan JC, Carr JL, Lee MS, Koshy B, Kumar AM, Zhang G, Newell WR, Windemuth A, Xu CB, Kalbfleisch TS, Shaner SL, Arnold K, Schulz V, Drysdale CM, Nandabalan K, Judson RS, Ruano G, Vovis GF (2001) Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 293:489–493 and 1048
- Stevens A, Ray D, Alansari A, Hajeer A, Thomson W, Donn R, Ollier WER, Worthington J, Davis JRE (2001) Characterization of a prolactin gene polymorphism and its association with systemic lupus erythematosus. *Arthritis Rheum* 44:2358–2366
- Tan FK, Wang N, Kuwana M, Chakraborty R, Bona CA, Milewicz DM, Arnett FC (2001) Association of fibrillin 1 single nucleotide polymorphism haplotypes with systemic sclerosis in Choctaw and Japanese populations. *Arthritis Rheum* 44:893–901
- Tsunoda K, Sanke T, Nakagawa T, Furuta H, Nanjo K (2001) Single nucleotide polymorphism (D68D, T to C) in the syntaxin 1A gene correlates to age at onset and insulin requirement in type II diabetes patients. *Diabetologia* 44:2092–2097
- Wakeley J, Nielsen R, Liu-Cordero SN, Ardile K (2000) The discovery of single nucleotide polymorphisms and inference about human demographic history. *Am J Hum Genet* 69:1332–1347
- Wang DG, Fan J-B, Siao C-J, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglu T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen NP, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES (1998) Large-scale identification, mapping, and genotyping of single nucleotide polymorphisms in the human genome. *Science* 280:1077–1082
- Weeks DE, Lathrop GM (1995) Polygenic diseases: methods for mapping complex disease traits. *Trends Genet* 11:513–519
- Weiss KM, Terwilliger JD (2000) How many diseases does it take to map a gene with SNPs? *Nat Genet* 26:151–157
- Wolf CR, Smith G, Smith RL (2000) Pharmacogenetics. *BMJ* 320:987–990
- Yoshino T, Takeyama H, Matsunaga T (2001) Single nucleotide polymorphism analysis using a bacterial magnetic particle microarray. *Electrochemistry* 69:1008–1012
- Youil R, Kemper BW, Cotton RGH (1995) Screening for mutation by enzyme mismatch cleavage with T4 endonuclease VII. *Proc Natl Acad Sci USA* 92:87–91
- Zhu Y, Spitz MR, Lei L, Mills GB, Wu XF (2001) A single nucleotide polymorphism in the matrix metalloproteinase 1 promoter enhances lung cancer susceptibility. *Cancer Res* 61:7825–7829