CME General

Unraveling henetics of complex diseases

S N T De Silva*, D J S Fernando, S D Jayaratne***, J J Taylor******

Journal of the Ceylon College of Physicians, 2004, **37**, 23-29

Disease results from interaction between genes, proteins and the environment. Improved understanding of genetic and environmental contributions to complex diseases would shift the focus of medicine from the disease itself to **genetic characteristics of individuals** and to **their** experiences, habits and environmental conditions. The environment has a strong influence on the expression of genetically determined diseases.

A focus on the genetic, developmental and environmental components of disease and their unique combination in a given individual **would facilitate more effective** methods of prevention. Once **the genetic basis for** susceptibility to a complex disease is established, **the health care profession can target and monitor potential sufferers, help them avoid the environmental factors that can provoke disease and encourage early self-detection of disease signs and symptoms.**

The risk for common complex disorders such as cardiovascular diseases, diabetes, obesity, hypertension, asthma, etc are believed to determined by a mixture of multiple genes and environment factors. There is a massive increase in the global prevalence of these conditions and in Sri Lanka alone, deaths from cardiovascular diseases accounts for nearly 15% and hypertensive heart diseases accounts for 10% of cardiovascular diseases[1].

Diabetes mellitus too has become a major social burden due to increase rate of mortality and morbidity, which raises health care costs in patient management[2]. Recent findings suggested that the prevalence rate of Type II diabetes in Sri Lanka is 5.02% in urban areas and 2% in rural areas and 5.07% of the adult population have impaired glucose tolerance[3]. There have been many attempts to elucidate both the genetic and environment components but aetiology remains still unclear for common complex diseases. This review briefly describes the DNA variations, analytical methods and technology currently available to progress towards finding the genetic mechanisms of common complex diseases.

* *Research Assistant,*
** *Professor of Medicine,*
*** *Senior Lecturer in Medicine,*
Department of Medicine, University of Sri Jayewardenepura, Gangodawila, Nugegoda.

**** *Lecturer in Molecular Biology, School of Dental Sciences, Department of Oral Biology, University of Newcastle Upon Tyne.*

DNA variations

The cells of all organisms contain basic DNA complement that is unique to species, called genome. DNA variation is the raw material for genetic studies. Variations in DNA have arisen during evolution as a consequence of mutations. New mutations arise in single individuals in somatic cells or in the germline. If a germline mutation **does not** impair an individuals ability to have **offspring** who can transmit the mutation it can **become established in** a population. They are frequently called polymorphisms (Greek *polymorphos*, which literally means multiform)[4-6]. In genetics, the term DNA polymorphism is used to indicate that a particular position in the DNA has more than one form in the population[4,6,7]. The term "allele" is used to refer to a particular form occurring at a given position and the position is called a "locus". A genotype is a description of the two alleles at a particular locus. A distinction has been made between "variations" and "polymorphisms" in that a locus is only called polymorphic if the most common variant occurs in less than 99% of the population and consequently all other variants occur with a total frequency of 1% or more[7].

There are different types of variations in DNA. The simplest is known as a single nucleotide polymorphism (SNP) in which a single nucleotide (and consequently the complementary nucleotide of the base pair) has two, or sometimes three, forms common in the population[5-8-9,10].

SNPs are classified according to their position in or around genes into coding and noncoding polymorphisms[5]. Noncoding polymorphisms may occur in the promoter region of the gene, within introns,

5'- and 3'- untranslated regions (regions of the gene that are present in the mRNA but do not code for protein), and intergenic regions[4]. Noncoding polymorphisms do not alter the base pair sequence in the part of the DMA that codes for the protein but may alter the rate of transcription, the processing, or the stability of the mRNA. They may change the amount of protein generated by translation. Coding SNPs are classified into synonymous and nonsynonymous[5,8-10]. Synonymous polymorphisms change the **codon** into another that codes for the same amino acid. Therefore, there is no change in the structure of the protein. Nonsynonymous SNPs change the codon to one specifying for a different **amino acid** and therefore may change protein structure.

A more complex type of DNA variation is the microsatellite repeat. Also known as simple tandom repeats. It consists of short nucleotide sequences such as di-, tri-, tetra-, or pentanucleotides repeated several to hundreds of times along DNA. Microsatellite repeats occur in all individuals. There are also minisatellite repeats which are similar to repeating units but are few tens in nudeotide length (up to 25 base pairs). These two types of repeat unit of polymorphisms are categorised together as variable number of tandem repeats (VNTRs).

Other types of polymorphisms include short insertion/deletion variants called "indel", large insertions and deletions of transposabte elements[4-6]. They are unstable DNA elements, which can migrate to different regions of the genome. Defective gene expression due to DNA transposition is extremely rare.

DNA variations and *their* significance

DNA replication process is error prone. If the error is not corrected or repaired, a mutation may occur[4,7]. One cell contains approximately 9×10^9 base pairs of nuclear DNA; 10^{17} cell divisions are required to transform a zygote into an adult individual with 100 trillion cells (10^{14}). Thus, 9×10^{26} nudeotide incorporations take place during the lifetime of a human. The error rate of DNA polymerase is only 10^{-10}. Many of these misincorporations are repaired by the proof reading segment of the DNA replication machinery. Despite this there is plenty of room for error and generations of new mutations and polymorphisms. Although much emphasis has been placed on the role of exogenous factors (*i.e.* toxins and other environmental factors) in the generation of DNA variation, the most mutagenic factors are endogenous. **Mutations that are not repaired** in reproductive cells, called germ cell mutations, may be passed on to the next generation. It has been estimated that the human genome contains one sequence variant in every 200 to 1000 base pairs of sequence[5].

The mutation refers to a rare variant that is the primary cause of a clinical phenotype or a disease, whereas the term polymorphism is used to denote a variant present in the population in a relatively high frequency. Polymorphism, itself is not sufficient to cause a disease, although it may contribute to susceptibility to a disease or variation in functional properties of a protein[11].

Mapping human disease

DNA makers for genetic mapping

When biochemical basis of a disease is not known it is necessary to locate the gene in a suspected region. Genes were the initial makers but are widely spaced out with large gaps between them. Mapped features which are not genes called DNA markers. A marker is any **Mendelian** character, which is sufficiently polymorphic such that a randomly selected person has a good chance of being heterozygous. ie a DNA marker should exist in two allelic forms. It helps if the character can be scored easily and cheaply using readily available material.

Technological developments in the early 1990s making possible high-throughput genotyping of informative markers, led whole genome screens in search of complex disease genes. Mapping requires makers spaced at intervals no greater than about 20cM across the genome. Recombination fraction is a measure of distance between two loci, which differ to the physical distance. Two loci which shows 1% recombination is define as being 1cM (one centimorgan apart on the genetic map. Thus, a minimum of 150 markers **is required** for a full genosome scan but **300 markers would be more informative**. DNA polymorphisms are the basis of all **currently used** genetic markers and the Human Genome Project so far has generated 10,000 highly polymorphic markers that were sufficiently numerous and spaced across the entire genome.

RFLP (Restriction Fragment Length Polymorphism)

The first generation of DNA markers were RFLPs. They were initially typed by preparing southern blots from restriction digest of DNA and hybridization with radio labeled probes **to identify common structural variants in DNA termed RFLPs**. Nowadays RFLPs can easily be typed by PCR. A sequence including the variable restriction site is amplified; the **PCR** product

is incubated with the appropriate restriction enzyme and then run on a gel to **analyse the digestion products**. A more fundamental limitation is their limited informativeness. Because RFLPs have only two alleles, the sight is present or not.

Minisatellites (Variable number of tandem repeats) VNTRs

The repeat unit is few tens in length. These are of great improvement. They have many alleles and high heterozygosrty. However, the technical problems of **Southern** blotting and radioactive probes were still an obstacle to easy mapping. **Furthermore**, VNTRs are not evenly spread across the genome.

Microsatellites

Microsatellites are simple repetitive DNA sequences highly polymorphic between individuals in terms of the number of repeats. The advent of PCR mapping relatively quick and easy. Minisatellites are too long to amplify **easily by PCR** and so the standard tools for PCR linkage analysis employ **analysis of microsatellite poymorphisms**. These are mostly (CA)n repeats.

Microsatellites are more popular than minisatellite as DNA makers. Minisatellites are not evenly spread across the genome and found near the ends of chromosomes. Therefore, finding an association would limit to the ends of the chromosomes. However, microsatellites are more conveniently spaced through out the genome making it possible to genotype the entire genome. Easiest way to type length polymorphism is by PCR. However, PCR typing is much quicker and accurate for short lengths preferably less than 300 base pairs. Most of the minisatellites repeating units are more than 10 base pairs in length exceed this amount in a single array but in microsatellites which consist of 10-30 copies of simple di-tri-or tetra nucleotides units are more suitable for PCR typing.

Single *nucleotide* polymorphisms (SNPs)

The newest generation of markers are two allele single nucleotide polymorphisms. They include RFLPs but also polymorphisms that do not happen to create or abolish a restriction site. The advantage of SNPs is that they can be scored on solid-state arrays without recourse to gel electrophoresis. Typically the test DNA is amplified in very large multiples and hybridized to an array comprising a series of anchored oligonucleotide primers each terminating with a polymorphic nucleotide. A single primer extension step is carried out on the array using a mixture of **fluorescently** labeled **dideoxynucleotides**. Label is **incorporated into** primers that perfectly match the test DNA but not to those with a 3' mismatch. Reading the cells of the array for presence of fluorescent allows the types for every SNP on the array to be read off [4].

Mapping genes in complex diseases

There are three main approaches to mapping the genetic variants involved in a disease:

1. Functional cloning,
2. Candidate gene strategy
3. Positional cloning[12].

Functional cloning

This approach is used only when the biological basis was known. The identification of underlying protein defect leads to localization of responsible gene. When biological basis was known it is possible to purify and partially characterize some of the gene product. Specific oligonudeotides or specific antibodies can be generated from this and can be used to identify the gene. Genes in sickle cell anemia- β-globin chain of hemoglobin, in phenylketonuria the enzyme phenylalanine hydroxylase, in lesch Nyhan syndrome hypoxanthine phosphoribosyltransferase and in hemophilia blood clotting factor viii swere identified by this approach.

Positional cloning

Positional cloning is used when the biochemical nature of a disease is unknown. With positional cloning for complex diseases, marker genes and two complementary analytical methods, linkage analyses and association mapping, are used to detect the specific genetic regions and genes involved in the disease process[12]. Linkage analyses test for the co-segregation of a marker and disease phenotype within a pedigree[13], whereas association studies test for differences in marker allele frequencies at the population level between patients and a control population[13,14].

Marker genes not related to disease physiology and genome-wide screens are the starting point for mapping the genetic components of the disease. This will first identify the genetic region where a disease-predisposing gene lies and, once this has been found, will localize the gene and **the** functional and biological role in relation to disease will **be determined**[9]. This was used for example, for cloning the cystic fibrosis gene on chromosome 7 in 1989. Microsatellites and

single-nucleotide polymorphisms (SNPs) are the two most commonly used DNA polymorphisms employed with positional cloning[15,16].

Candidate gene approach

In the candidate gene approach, genes with a known or proposed function with potential to the disease phenotype are investigated for a direct role in disease. Here a candidate gene encodes a protein of known function which if mutated could account in part for the pathogenesis of the disorder. The success of this approach depends of course on knowing enough about the possible molecular pathogenesis of the disorder to make an educated guess as to candidates. These approaches can be applied without any prior knowledge of the biological basis of the disease using genome-wide studies, combined with the candidate gene approach and comparative analyses using animal models of disease. The most successful application of the candidate gene approach to mapping complex diseases in humans has been with the HLA region. Genes in the HLA region have been implicated in the aetiology of over 200 diseases[17,18]. These include, complex autoimmune diseases such as type 1 diabetes, rheumatoid arthritis and multiple sclerosis; cancers such as Hodgkin's disease; infectious diseases such as malaria, tuberculosis and AIDS; and such other diseases as narcolepsy.

Linkage and linkage disequilibrium

Linkage disequilibrium refers to the nonrandom association of alleles at different loci, and random association is described as, 'linkage equilibrium'[19]. Linkage is the tendency of genes or DNA variants at different loci to be inherited together as a consequence of their physical proximity on a single chromosome and tests for co-segregation of a maker and a phenotype within a pedigree. An allele at a marker at one point in a **chromosome** will be physically connected with the alleles at all other markers on the same string. When a new mutation occurs at a particular locus, **it is** physically connected to every other site on the chromosome until it is subjected to recombination inducing an association in the population between the alleles that happen to be present at different loci on the mutated chromosome. It is this association that gives rise, over subsequent generations, to the observed linkage disequilibrium, the non-independence of alleles at different loci close together on the same chromosome. As the chromosome is passed from one generation to another, recombination breaks down the association, more so for loci further apart. The more loosely two loci are physically linked, the faster the decay of linkage disequilibrium under neutrality via recombination. That is, linkage disequilibrium decays over time (generations) so that many generations later, disequilibrium is localized to the region very close to the location of the original mutated locus. One distinction between linkage and linkage disequilibrium is that linkage is a within-family association between marker and disease alleles, whereas linkage disequilibrium is a population association between marker and disease alleles.

Parametric linkage analysis

This requires a precise genetic model, detailing the mode of inheritance, gene frequencies and penetrance of genotype. This is a powerful method to scan the genome in 20Mb segments to locate disease gene if a valid model is available.

Non-parametric linkage analysis

Non-parametric analysis does not require a specific genetic model. These methods ignore unaffected people and look for alleles or chromosomal segments that are shared by affected individuals. Shared segment methods can be used within nuclear families, within known extended families or in whole populations.

Identity by state and identity by descent

Identity by state (IBS) alleles look the same and may have the same DNA sequence but not derived from a known common ancestor. Identity by descent alleles (IBO) are demonstrably copies of the same ancestral (usually parental) allele.

LOD score

In linkage analysis the coinheritance of maker and gene are followed within a family. The probability that their observed inheritance pattern could occur by chance alone (that is if they are completely unlinked) is calculated. The calculation is then repeated assuming a particular degree of linkage and the ratio of two probabilities. (No linkage Vs a specified degree of linkage is determined. This ratio expresses the odd for (and against) that degree of linkage and because the logarithm of ratio is used it is known as the logarithm of odds score or LOD score. For practical purposes, a LOD score equal to or greater than three is taken to confirm that the gene and marker are linked. This represents 1000:1 odds that the two loci are linked. Negative LODs give evidence against linkage.

Phenometric analyses

A fundamental assumption of the different study designs described that the phenotypic definition is precise. These study designs are referred to as phenometric analyses[20]. One assumption of phenotypic precision is that all the cases and none of the controls have the disease under study.

A further implicit assumption is that the disease definition itself is precise, that is, that all cases have the same disease.

Genometric analysis

An alternative approach to dissect the genetic predisposition to these disorders is to characterize the genotype in a large population and then determine its relationship with the phenotype: a "genometric approach". This approach[21] is in early stages of development. The sample size required to test whether there is an association between a DNA variant and a disorder is a key consideration[22,25]. Estimation of the number of subjects required is influenced by the frequency of the DNA polymorphism in the particular population under investigation (ie, the frequency in a particular ethnic group[26,27].

Complex diseases

Parametric linkage analysis, depends on following the inheritance of genetic markers in extended pedigrees to look for cosegregation of marker alleles in affected individuals, under a defined model of inheritance. The logic is that if a gene somewhere in the genome is responsible for the disease, affected family members are expected to have inherited the same disease predisposing allele at the locus, and markers that lie physically near this disease gene will be transmitted along with the disease allele[28]. Linkage methods with LOD score analysis requires parametric linkage analysis and suitable to identify genes when large multigenerational pedigrees are available. Monogenetic diseases like Huntington's disease and breast cancer genes and complex diseases such as schizophrenia and psorisis and maturity onset of diabetes of the young were identified by parametric linkage analysis method.

To resolve genetics of complex diseases do not lend themselves to follow parametric linkage analysis as the segregation of genes do not follow a **Mendalian** mode of inheritance pattern. Depending on a number of factors, such as mode of inheritance (dominant or recessive), penetrance (risk of being affected), and population under study, the number of different disease mutations represented in a collection of unrelated families will vary. Either all families in a collection will carry a different allele of the disease gene or all families possess the same ancestral mutation. In genetically isolated populations the families could be related to each other, although the relationship may be too distant to be recognizable therefore unique mutations can, by chance, become common. In such a group, the various collected families actually form one larger, extended pedigree and thus, in theory, provide much greater power for gene localization by linkage analysis. However, true connections between families cannot be **determined** owing to the fact that much of the generations that link the pedigree are missing. Without knowing the connections between families, conventional linkage analysis offers no further benefit. Alternative approach is linkage disequilibrium/ association studies performed using population samples of independent individuals or family-based association samples with specific family structures. Linkage disequilibrium can be used instead of linkage as an initial tool to map disease susceptibility genes in populations.

Association studies

The simplest definition of a "genetic association" is to determine whether there is a relationship between a specific DNA variation and the frequency or severity of a particular trait[19,27]. Association studies compare marker frequencies in unrelated cases and controls and test for the co-occurrence of a marker and disease at the population level, a significant association with disease might implicate a candidate gene in the etiology of a disease. An association can have many possible causes not all genetic[29].

Sampling strategies for genetic association studies

Case control study design

Different experimental designs can be used to conduct genetic association studies[20,27]. The case-control studies have been the most widely applied strategy for characterizing genetic contribution to a disease[29], and are identical to those used in clinical epidemiology and consist of determining the frequency of a DNA variant in individuals affected by a disease (cases) and in those not affected by the disorder (controls). The advantages of this approach are that cases are readily obtained and can be efficiently genotyped and compared with control populations. It is also economically and statistically efficient. However, **the results of case-control studies are often** inconsistent and **this** can be attributed to number of factors, including clinical heterogeneity (associations are often only shown with certain subsets of disease), genetic heterogeneity (it should not be unexpected to find ethnic differences in associations), and study

design (small, underpowered sample sizes, poor quality control of genotyping data, and inappropriate selection of controls will all contribute to inconsistent findings).

This approach has resulted in the identification of DNA variants that prove to be spuriously associated with disease. The selection of appropriately matched controls has also been the subject of discussion. Ethnically unmatched controls may lead to positive results due to population stratification. This is due to presence of multiple subgroups with different allele frequencies within a population. The different underlying allele frequencies in sampled subgroups might be independent of the disease within each group and can lead to erroneous conclusions of linkage disequlibrium or disease relevance. If two populations have subtle genetic differences and the cases come predominantly from one population, resulting evidence for association might be statistically highly significant, but the true association will be with the population rather than the disease.

Prospective cohort study design

This method requires prospectively assembling a group of individuals before the onset of disease followed to determine the frequency with which the disease develops. Then, the frequency of the DNA variant is determined in the entire cohort, allowing estimation of risk and predictive values. This excludes the bias for selection of control population. This approach requires sufficient number of cases to be adequately powered to study disease genetics.

Family based association studies

A study design unique to genetic epidemiology is the "family-based association" design[20,22,23]. Genetic association methods developed have focused on the need to account for population stratification. Family-based association studies use relatives of cases as controls, such as unaffected siblings or the **non-transmitted** alleles from parents.

A number of family-based association methods have been proposed, the most commonly used being the transmission disequilibrium test (TDT)[24]. The test examines the transmission of potential disease alleles from a parent who is heterozygous for the marker to an affected offspring. It is a test of association only in the presence of linkage, and because family members act as controls, spurious associations due to population differences do not arise. The original test uses a single affected offspring and both parents. A number of extensions to the original TDT have been proposed that allow both parents and **an** unaffected sibling to be analysed, making maximum use of incomplete nuclear family data.

Tests are available for both **SNPs** and multiallelic markers such as microsatellites. As the TDT is dependent upon the number of informative transmissions, microsatellite markers are often more useful. If 100 parents are genotyped, 80 informative transmissions will be expected for a microsatellite marker with 80% heterozygosity, whereas the maximum heterozygosity measure for a SNP is only 50%. However, at least two multilocus haplotype methods have been developed for the TDT, which should overcome the low information content of single SNPs by combining up to four SNPs in a single haplotype.

The TDT has become a very versatile methodology, allowing a range of family structures, marker types, and phenotype data to be analysed simultaneously.

This versatility, coupled with the increased power of an association-based method, will inevitably lead to increased use of the TDT in the search for complex disease susceptibility genes.

Affected-sibling-pair methods

Nonparametric (model-free) linkage analysis can be carried out on small, nuclear families, such as affected sibling pairs (ASPs), without making any assumptions about the genetics of diseases, provided sufficient number of families can be collected.

Pair of sibs are expected to share 0,1 or 2 parental haplotypes with frequency ¼, ½ and ¼ respectively. However, if both sibs are affected by a genetic disease then they are likely to share whichever segment of a chromosome carries the disease locus. **Taking the simplest assumption** that everybody with the disease carries a mutant allele at this locus, and if the disease is dominant, they will share at least one parental haplotypes and if the disease is recessive they will share both haplotypes[3].

Summary and conclusions

Epidemiological studies have shown that the prevalence rates are rising and Type 2 diabetes alone would afflict 80 million people worldwide by the year 2020[30]. Type 2 diabetes has been regarded as a common complex disease in which multiple genes interact with environmental factors to exceed the threshold to develop the diseases. The complication of **microvascular** and macrovascular diseases has been progressed by the time of diagnosis of the diseases hence, identifying the diseases before onset would enable the patients to minimize their exposures to environmental factors. Genetic association studies are becoming increasingly frequent in the literature and they are considered central to the deciphering of the

genetic basis of complex disease. However, research into genetic components of common complex diseases are in it's infancy in Sri Lanka. Investigating into the genetic component of Type 2 diabetes in Sri Lankan population will make it possible to identify possible gene mutations in relation to Sri Lankan population and plan for future **therapeutic** targets and lifestyle interventions since the prevalence is increasing tremendously throughout the world[30,3] and in Sri Lanka. Thus, it is advisable for clinicians and scientists working in the field of complex genetic disorders to be familiar with how genetic association studies are conducted and interpreted, their potential pitfalls, the strengths and weaknesses of genetic association studies and what these studies contribute to the understanding of predisposition to disease. In this context there is need for political, academic and economic will to promote genetic mechanisms into diseases in Sri Lanka.

References

1. Annual Health Bulletin, Ministry of Health, Sri Lanka. 1995.
2. Guidelines on management of diabetes mellitus in Sri Lanka. Sri Lanka Medical Association, 2000.
3. Fernando DJS. The prevalence of Complications of Diabetes in Patients Attending Diabetic Clinics in Sri Lanka and UK. [M.Sc Theses.] Manchester University of Manchester. 1992.
4. Strachan T, Read AP. Human molecular genetics. 1st Ed, BIOS Sceintific Publishers Ltd, 1999; 270-294.
5. Chakravarti A. Population genetics - making sense out of sequence. *Nat Genet* 1999; **21**: 56-60.
6. Gibson PS, Muse SV. A primer of genome science. Sunderland (MA): Sinauer Associates; 2002.
7. Harti DL, Clark AG. Principles of population genetics. Sunderland: Sinauer Associates; 1989.
8. Tabor HK, Risch NJ, Myers RM. Opinion: candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* 2002; **3**: 391-7.
9. Weiss KM, Terwilliger JD. How many diseases does it take to map a gene with SNPs? *Nat Genet* 2000; **26**: 151-7.
10. Fallin D, Cohen A, Essioux L, Chumakov I, Blumenfeld M, Cohen D, et al. Genetic analysis of case/control data using estimated haplotype frequencies, application to APOE locus variation and Alzheimer's disease. *Genome Res* 2001; **11**: 143-51.
11. Romero R, Kuivaniemi H, Tromp G, Olson JM. The design, execution, and interpretation of genetic association studies to decipher complex diseases. *Am J Obstet Gynecol* 2002; **187**:1299-312.
12. Thomson G. An overview of the genetic analysis of complex diseases, with reference to type 1 diabetes 2001; Vol.15, No. 3, 265-77.
13. Borecki IB, Suarez BK. Linkage and association: basic concepts. In Rao DC & Province MA Eds., Genetic Dissection of Complex Traits, San Diego: Academic Press 2001: 45-66.
14. Thomson G. Mapping disease genes: family-based association studies. *American Journal of Human Genetics* 1995; **57**: 487 - 498.
15. Thomson G. Mapping of disease loci. In Kalow W (ed.) Pharmacogenomics. New York: Marcel Dekker, 2001; 337-61.
16. Peltonen L, McKusick VA. Dissecting human disease in the post genomic era. *Science* 2001; **291**: 1224-9.
17. Thorsby E. Invited anniversary review: HLA associated diseases. *Human Immunology* 1997; **53**:1-11.
18. Svejgaard A. MHC and disease associations. In Herzenberg LA, Weir DM, Blackwell C, et al (eds) Weir's. Handbook of Experimental Immunology, 5th edn, pp 37.1-37.13. Cambridge, MA: Blackwell Science, 1996.
19. Hedrick PW. Genetics of Populations. Jones & Bartlett, 2000.
20. Collins-Schramm HE, Phillips CM, Operario DJ, Lee JS, Weber JL, Hanson RL, et al. Ethnic-difference markers for use in mapping by admixture linkage disequilibrium. *Am J Hum Genet* 2002; **70**: 737-50.
21. Elston RC. Linkage and association. *Genet Epidemiol* 1998; **15**: 565-76.
22. Garden LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001; **2**: 91-9.
23. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996; **273**:1516-7.
24. Dahlman I, Eaves IA, Kosoy R, Morrison VA, Heward J, Gough SC, et al. Parameters for reliable results in genetic association studies in common disease. *Nat Genet* 2002; **30**: 149-50.
25. Ioannidis JP, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic association studies. *Nat Genet* 2001; **29**: 306-9.
26. Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, et al. Estimating African American admixture proportions by use of population-specific alleles. *Am J Hum Genet* 1998; **63**: 1839-51.
27. Khoury MJ, Beary TH, Cohen BH. Fundamentals of genetic epidemiology. In: Kelsey J, Mamot M, Stoltey R Vessey M, editors. Monographs in epidemiology and biostatistics. Vol 22. New York: Oxford University Press 1993; 373-383.
28. Risch N. Searching for genes in complex diseases: lessons from systemic lupus erythematosus. *The Journal of Clinical Investigation* 2000; **105** (11): 1503-6.
29. Garden LR, Bell JI. Association study design for complex diseases. *Nature reviews /Genetics* 2001; **2**: 91-9.
30. Alberti KGMM, Zimmet PZ. Definition, Diagnosis and Classification of Diabetes Mellitus and its complications. Part 1: Diagnosis and Classification of Diabetes Mellitus, Provisional report of a WHO Consultation. *Diabetic Medicine* 1998; **15**: 539-53.