International Journal of Biology Sciences
www.biologyjournal.net
Online ISSN: 2664-9934, Print ISSN: 2664-9926
Received: 02-07-2021; Accepted: 16-07-2021; Published: 30-07-2021
Volume 3, Issue 2, 2021, Page No. 01-05

Bioinformatics methods for identifying human disease genes

Sudheer Menon
Department of Bioinformatics, Bharathiar University, Coimbatore, Tamil Nadu, India

Abstract
With the explosion in genomic and practical genomics information, techniques for disorder gene identification are unexpectedly evolving. Databases are now quintessential to the manner of choosing candidate sickness genes. Combining positional data with disorder traits and useful records is the typical method through which candidate disorder genes are selected. Enrichment for candidate disorder genes, however, relies upon on the competencies of the running researcher. Over the previous few years, a wide variety of bioinformatics strategies that enrich for the most probably candidate sickness genes have been developed. Such in silico prioritization techniques might also in addition enhance by way of completion of datasets, by using improvement of standardized ontologies throughout databases and species and, ultimately, via the integration of distinctive strategies.

Keywords: bioinformatics methods, human disease genes

Introduction

Currently, with the extend in handy facts and the improvement of novel molecular biology techniques, new strategies for the identification of sickness genes are evolving. Linkage research and mutation screening are turning into less difficult and the wide variety of recognized (disease) genes is growing rapidly. 2003 noticed the completion of the human genome sequence and the quantity of genes is now set to 20,000-25,000 [1, 2]. With all the genetics science in place, identification of disease-related mutations in Mendelian single-gene problems ordinarily relies upon on having the proper sufferers and families. The genetic evaluation of complicated ailments nonetheless stays a tough task, however, and most genes for multifactorial sickness continue to be to be discovered. Genetic mapping via linkage is a mainstay of human genetics research. While positional data reduces the wide variety of genes that are candidates for inflicting the disease, this discount is frequently now not adequate for speedy sickness gene identification. The goal of candidate gene prioritization strategies is to select these genes for special mutation evaluation that are most probable to be the purpose of the disease. This is specifically applicable given that positional techniques might also depart up to a hundred special genes as candidates. Hence extra facts to be used for prioritization is essential.

Databases have come to be a core supply for contemporary gene hunters. Retrieval structures such as the National Center for Biotechnology Information's Entrez [3], the Sequence Retrieval System [4] and Maarten's Retrieval System [5] grant handy and quickly get admission to a series of regularly used databases. The essential center of attention of these retrieval structures is to fetch a set of database entries that meet the person query. Identification of a disorder gene is most in all likelihood to be profitable when positional and purposeful routes are integrated. Integration of facts primarily based on genomic context, such as in the University of California, Santa Cruz genome browser and Ensembl [6, 7], resulted in step via step interfaces (e.g EnsMart [8]) which extract facts based totally on chromosomal position, gene expression [9] and gene ontology (GO) [10]. Enrichment for ailment candidate genes the usage of these database interfaces, however, relies upon closely on the operation capabilities of the researcher. Alternative strategies have been developed systematically to discover datasets for the most possibly candidate sickness genes. This paper provides an overview of such strategies and their accessibility.

Human disease gene identification methods

Disease gene identification strategies frequently comply with the identical average procedure. DNA is first accumulated from quite a few sufferers who are believed to have the equal genetic disease. Then, their DNA samples are analyzed and screened to decide in all likelihood areas the place the mutation should probably reside. These methods are referred to below.

These probably areas are then lined-up with one some other and the overlapping vicinity need to include the mutant gene.

If ample of the genome sequence is known, that area is searched for candidate genes.

Coding areas of these genes are then sequenced till a mutation is determined or any other affected person is discovered, in which case the evaluation can be repeated, doubtlessly narrowing down the area of interest.

The variations between most disorder gene identification techniques are in the 2d step (where DNA samples are analyzed and screened to decide areas in which the mutation may want to reside).

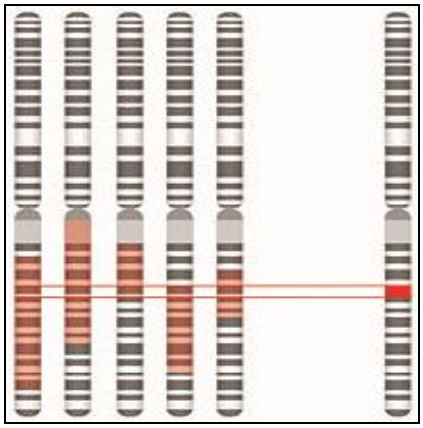

Fig 1

Pre-genomics techniques

Without the resource of the whole-genome sequences, pre-genomics investigations appeared at pick out areas of the genome, regularly with solely minimal know-how of the gene sequences they had been searching at. Genetic methods successful of presenting this kind of data consist of Restriction Fragment Length Polymorphism (RFLP) evaluation and microsatellite analysis.

Loss of heterozygosity (LOH)

Loss of heterozygosity (LOH) is method that can solely be used to evaluate two samples from the identical individual. LOH evaluation is frequently used when figuring out cancer-causing oncogenes in that one pattern consists of (mutant) tumor DNA and the different (control) pattern consists of genomic DNA from non-cancerous cells from the identical individual. RFLPs and microsatellite markers grant patterns of DNA polymorphisms, which can be interpreted as living in a heterozygous vicinity or a homozygous location of the genome. Provided that all persons are affected with the equal ailment ensuing from a manifestation of a deletion of a single replica of the equal gene, all humans will include one vicinity the place their manage pattern is heterozygous however the mutant pattern is homozygous - this location will incorporate the disorder gene [1, 2].

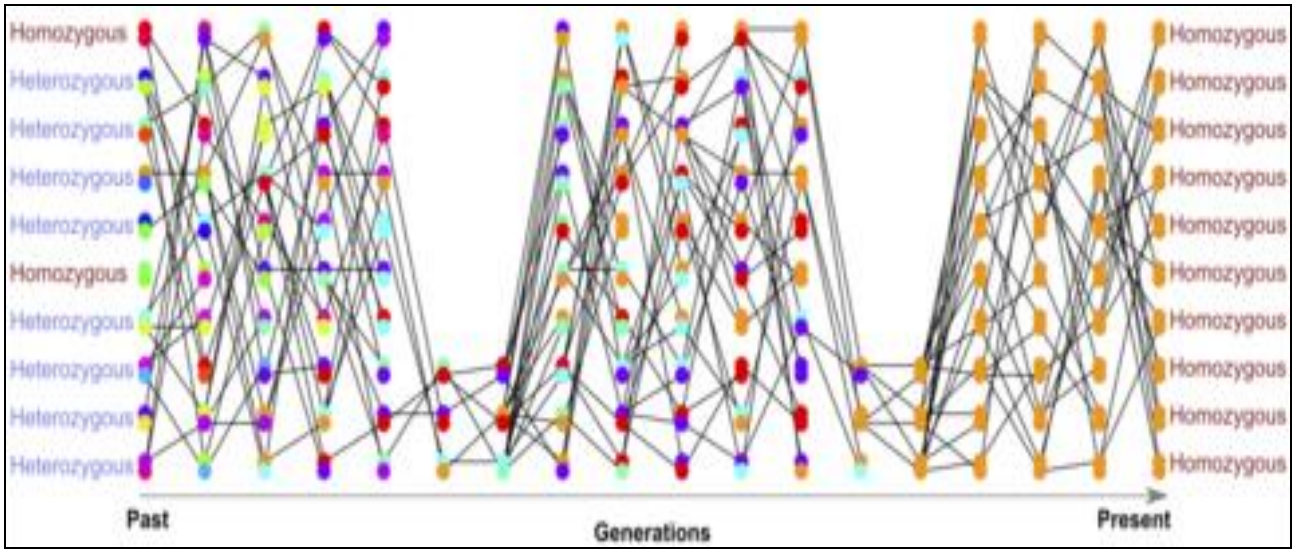

Fig 2: Loss of heterozygosity (LOH) is a cross chromosomal event that results in loss of the entire gene and the surrounding chromosomal region.

Post-genomics techniques

With the introduction of modern-day laboratory methods such as High-throughput sequencing and software program successful of genome-wide analysis, sequence acquisition has come to be increasingly more much less luxurious and time-consuming, as a result presenting extensive advantages to science in the structure of extra environment friendly disorder gene identification techniques.

Identity by descent mapping

Identity via descent (IBD) mapping typically makes use of single nucleotide polymorphism (SNP) arrays to survey regarded polymorphic websites at some stage in the genome of affected humans and their mother and father and/or siblings, each affected and unaffected. While these SNPs likely do no longer reason the disease, they furnish precious perception into the make-up of the genomes in question. A area of the genome is viewed equal by way of descent if contiguous SNPs share the equal genotype. When evaluating an affected man or woman to his/her affected sibling, all same areas are recorded (ex. Shaded in crimson in above figure). Given that an affected sibling and an unaffected sibling do no longer have the identical disorder phenotype, their DNA have to through definition be distinctive (barring the presence of a genetic or environmental modifier). Thus, the IBD mapping effects can be similarly supplemented with the aid of getting rid of any areas that are equal in each affected humans and unaffected siblings [3].

This is then repeated for a couple of families, as a consequence producing a small, overlapping fragment, which theoretically consists of the sickness gene.

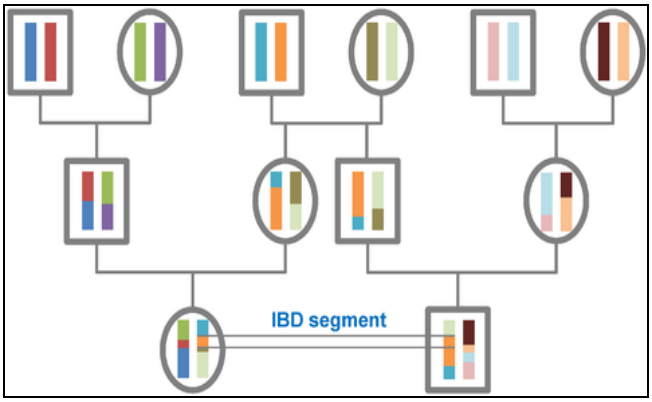

Fig 3

A DNA segment is identical by state (IBS) in two or more individuals if they have identical nucleotide sequences in this segment. An IBS segment is identical by descent (IBD) in two or more individuals if they have inherited it from a common ancestor without recombination, that is, the segment has the same ancestral origin in these individuals. DNA segments that are IBD are IBS per definition, but segments that are not IBD can still be IBS due to the same mutations in different individuals or recombinations that do not alter the segment.

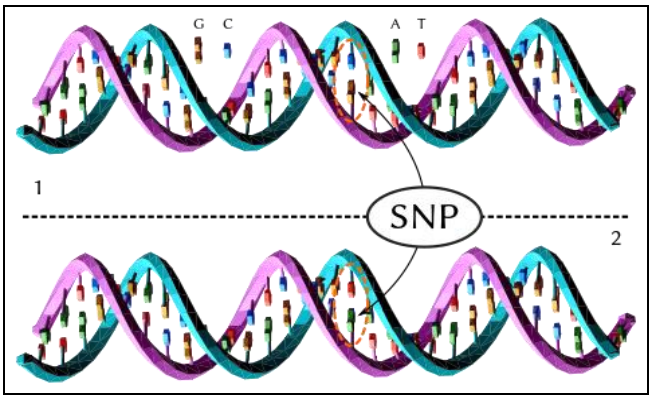

Fig 4

In genetics, a single-nucleotide polymorphism is a germline substitution of a single nucleotide at a specific position in the genome. Although certain definitions require the substitution to be present in a sufficiently large fraction of the population, many publications do not apply such a frequency threshold.

Homozygosity/autozygosity mapping

Homozygosity/Autozygosity mapping is a effective technique, however is solely legitimate when looking for a mutation segregating inside a small, closed population. Such a small population, perhaps created by using the founder effect, will have a constrained gene pool, and consequently any inherited ailment will possibly be a end result of two copies of the equal mutation segregating on the same haplotype. Since affected folks will likely be homozygous in the regions, searching at SNPs in a place is an sufficient marker of areas of homozygosity and heterozygosity. Modern day SNP arrays are used to survey the genome and perceive giant areas of homozygosity. Homozygous blocks in the genomes of affected men and women can then be laid on pinnacle of every other, and the overlapping area have to comprise the sickness gene [4].

This evaluation is regularly prolonged by means of examining autozygosity, an extension of homozygosity, in the genomes of affected individuals [5]. This can be completed with the aid of plotting a cumulative LOD rating alongside the overlaid blocks of homozygosity. By taking into consideration the populace allele frequencies for all SNPs by means of autozygosity mapping, the outcomes of homozygosity can be confirmed [5]. Furthermore, if two suspicious areas show up as a end result of homozygosity mapping, autozygosity mapping can also be capable to distinguish between the two (ex. If one block of homozygosity is a end result of a very non-diverse area of the genome, the LOD score will be very low).

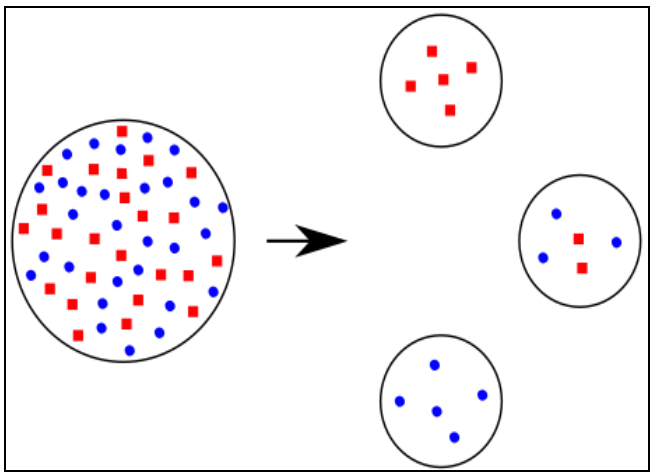

Fig 5

In population genetics, the founder effect is the loss of genetic variation that occurs when a new population is established by a very small number of individuals from a larger population. It was first fully outlined by Ernst Mayr in 1942, using existing theoretical work by those such as Sewall Wright. As a result of the loss of genetic variation, the new population may be distinctively different, both genotypically and phenotypically, from the parent population from which it is derived. In extreme cases, the founder effect is thought to lead to the speciation and subsequent evolution of new species.

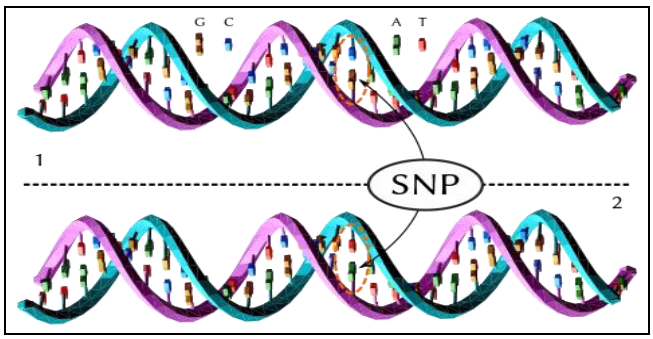

Fig 6

A haplotype is a group of alleles in an organism that are inherited together from a single parent.

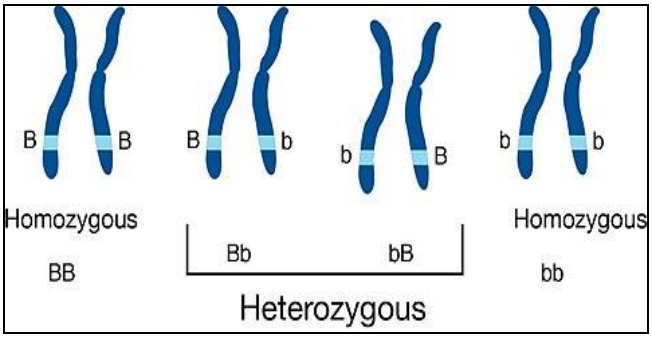

Fig 7

Zygosity is the degree to which both copies of a chromosome or gene have the same genetic sequence. In other words, it is the degree of similarity of the alleles in an organism.

Tools for Homozygosity Mapping

1. **HomSI:** a homozygous stretch identifier from next-generation sequencing data [6] A tool that identifies homozygous regions using deep sequence data.

Genome-wide knockdown studies

Genome-wide knockdown research are an instance of the reverse genetics made feasible with the aid of the acquisition of complete genome sequences, and the introduction of genomics and gene-silencing technologies, mostly siRNA and deletion mapping. Genome-wide knockdown research contain systematic knockdown or deletion of genes or segments of the genome [7].

This is commonly finished in prokaryotes or in a tissue lifestyle surroundings due to the huge variety of knockdowns that need to be performed [8]. After the systematic knockout is done (and perchance proven by way of mRNA expression analysis), the phenotypic effects of the knockdown/knockout can be observed. Observation parameters can be chosen to goal a distinctly particular phenotype [8]. The ensuing dataset is then be queried for samples which show off phenotypes matching the disorder in query – the gene(s) knocked down/out in stated samples can then be regarded candidate disorder genes for the man or woman in question.

Whole exome sequencing

Whole exome sequencing is a brute-force strategy that entails the use of contemporary day sequencing technological know-how and DNA sequence meeting equipment to piece collectively all coding parts of the genome. The sequence is then in contrast to a reference genome and any variations are noted. After filtering out all acknowledged benign polymorphisms, synonymous changes, and intronic adjustments (that do now not have an effect on splice sites), solely probably pathogenic variations will be left.

This approach can be mixed with different strategies to similarly knock out probably pathogenic versions have to extra than one be identified [9].

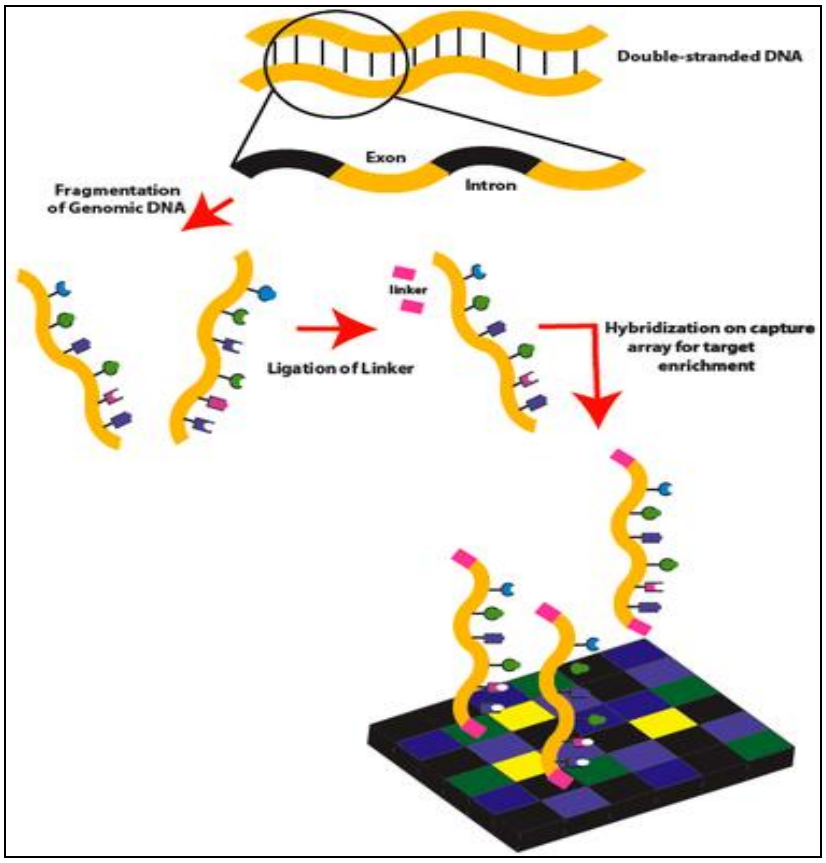

Fig 8

Exome sequencing, also known as whole exome sequencing (WES), is a genomic technique for sequencing all of the protein-coding regions of genes in a genome. It consists of two steps: the first step is to select only the subset of DNA that encodes proteins. These regions are known as exons – humans have about 180,000 exons, constituting about 1% of the human genome, or approximately 30 million base pairs. The second step is to sequence the exonic DNA using any high-throughput DNA sequencing technology.

Future: Integration and Standardization

The various methods for identifying my candidate disease genes in humans cover different concepts. They use functional and literature data, gene-specific characteristics, anatomy-based gene/protein expression data or phenotype comparison analyses. In light of the comparable enrichment levels achieved with the different methods, it is likely that they can complement each other.

The results discussed here suggest that the phenotype is a powerful source for revealing biological function and that special attention is needed for the standardization of the description of phenotypes. Various approaches to a more systematic description of phenotype data have been proposed and await further development. Essential to the improvement of the candidate disease gene identification methods will be the establishing of standard vocabularies that can be used across databases and species. A further challenge will be to develop, refine and integrate these methods into a system that aids in elucidation and understanding of the mechanisms of (complex) disease.

References

1. Baker SJ, Fearon ER, Nigro JM, Hamilton SR, Preisinger AC, Jessup JM *et al*. "Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas". Science,1989:244(4901):217-21.
2. Lee AS, Seo YC, Chang A, Tohari S, Eu KW, Seow-Choen F *et al*. "Detailed deletion mapping at chromosome 11q23 in colorectal carcinoma". Br. J. Cancer,2000:83(6):750-5.
3. Bell R, Herring SM, Gokul N, Monita M, Grove ML, Boerwinkle E *et al*. High-resolution identity by descent mapping uncovers the genetic basis for blood pressure differences between spontaneously hypertensive rat lines". Circ Cardiovasc Genet,2011:4(3):223-31.
4. Sherman EA, Strauss KA, Tortorelli S, Bennett MJ, Knerr I, Morton DH *et al*. "Genetic mapping of glutaric aciduria, type 3, to chromosome 7 and identification of mutations in c7orf10". Am. J. Hum. Genet,2008:83(5):604-9.
5. Broman KW, Weber JL. "Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain". Am. J. Hum. Genet,1999:65(6):1493-500.
6. Zeliha Görmez; Burcu Bakir-Gungor, Mahmut Şamil Sağıroğlu. "HomSI: a homozygous stretch identifier from next-generation sequencing data". Bioinformatics,2014:30(3):445-447.
7. Sudheer Menon. "Preparation and computational analysis of Bisulphite sequencing in Germfree Mice" International Journal for Science and Advance Research In Technology,2020:6(9):557-565.
8. Sudheer Menon, Shanmughavel Piramanayakam, Gopal Agarwal. "Computational identification of promoter regions in prokaryotes and Eukaryotes" EPRA International Journal of Agriculture and Rural Economic Research (ARER),2021:9(7):21-28.
9. Sudheer Menon. "Bioinformatics approaches to understand gene looping in human genome" EPRA International Journal of Research & Development (IJRD),2021:6(7):170-173.
10. Sudheer Menon. "Insilico analysis of terpenoids in Saccharomyces Cerevisiae"international Journal of Engineering Applied Sciences and Technology, ISSN No. 2455-2143,2021:6(1):43-52.
11. Sudheer Menon. "Computational analysis of Histone modification and TFBs that mediates gene looping" Bioinformatics, Pharmaceutical, and Chemical Sciences (RJLBPCS),2021:7(3):53-70.
12. Sudheer Menon Shanmughavel piramanayakam, Gopal Prasad Agarwal. "FPMD-Fungal promoter motif database: A database for the Promoter motifs regions in fungal genomes" EPRA International Journal of Multidisciplinary research,2021:7(7):620-623.
13. Sudheer Menon, Shanmughavel Piramanayakam, Gopal Agarwal. Computational Identification of promoter regions in fungal genomes, International Journal of Advance Research, Ideas and Innovations in Technology,2021:7(4):908-914.
14. Sudheer Menon, Vincent Chi Hang Lui, Paul Kwong Hang Tam. Bioinformatics methods for identifying hirschsprung disease genes, International Journal for Research in Applied Science & Engineering Technology (IJRASET),2021:9(7):2974-2978.
15. Luo J, Emanuele MJ, Li D, Creighton CJ, Schlabach MR, Westbrook TF, Wong KK *et al*. "A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene". Cell,2009:137(5):835-48.
16. Fortier S, Bilodeau M, Macrae T, Laverdure JP, Azcoitia V, Girard S *et al*. "Genome-wide interrogation of Mammalian stem cell fate determinants by nested chromosome deletions". PLOS Genet,2010:6(12):e1001241.
17. Chen WJ, Lin Y, Xiong ZQ, Wei W, Ni W, Tan GH *et al*. "Exome sequencing identifies truncating mutations in PRRT2 that cause paroxysmal kinesigenic dyskinesia". Nat. Genet,2011:43(12):1252-5.
18. Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G *et al*. "Haplotype tagging for the identification of common disease genes". Nat Genet,2001:29(2):233-7.