

Linkage disequilibrium and the mapping of human disease genes

Nahid Askari*, Amin Baghizadeh

Department of Biotechnology, Institute of Science and High Technology and Environmental Sciences, Graduate University of Advanced Technology, Kerman-Iran

Email address:

nahid.askari@yahoo.com (N. Askari)

To cite this article:

Nahid Askari, Amin Baghizadeh. Linkage Disequilibrium and the Mapping of Human Disease Genes. *International Journal of Genetics and Genomics*. Vol. 2, No. 4, 2014, pp. 68-76. doi: 10.11648/j.ijgg.20140204.14

Abstract: Identification of human disease genes can be accomplished by two strategies: functional cloning and positional cloning. Genetic mapping is the localization of genes underlying phenotypes on the basis of correlation with DNA variation, without the need for prior hypotheses about biological function and the simplest form, called linkage analysis. The ability to clone and sequence DNA made it possible to tie genetic linkage maps in model organisms to the underlying DNA sequence. In conclusion particular alleles at neighboring loci tend to be co-inherited. For tightly linked loci, this might lead to associations between alleles known as linkage disequilibrium (LD). Considerable effort and expense have been expended in whole-genome screens aimed at detection of genetic loci contributing to the susceptibility to complex human diseases.

Keywords: Human Disease Genes, Functional Cloning, Positional Cloning, Genetic Mapping, Mendelian Trait, Linkage Disequilibrium (LD)

1. Introduction

For most of the 20th century, genome-wide linkage mapping was impractical in humans: Family sizes are small crosses are not by design, and there were too few classical genetic markers to systematically trace inheritance. Progress in identifying the genes contributing to human traits was initially limited to studies of biological candidates such as blood-type antigens [1] et al, and hemoglobin β protein in sickle-cell anemia [2-4].

Marker loci near a disease gene are often observed to be in linkage disequilibrium with the disease; that is, the relative frequencies of marker alleles in affected individuals differ from those in the general population. Linkage disequilibrium occurs because each new disease-predisposing mutation originally appears on a single chromosome. Individuals who inherit a disease mutation are likely to also inherit the alleles of the original chromosome, at neighboring marker loci. As generations pass, recombination or mutation can disrupt the joint transmission of disease mutation and marker allele. Because recombination with the disease gene happens less often for nearby marker loci, markers in the immediate vicinity of the gene should remain in greater disequilibrium than more distant marker loci. The potential value of

haplotypes defined by several single nucleotide polymorphisms has attracted recent interest. With sufficient linkage disequilibrium (LD), haplotypes could be used in association studies to map common alleles that might influence the susceptibility to common diseases, as well as for reconstructing the evolution of the genome. It has been proposed that a globally useful resource need only be based on high frequency variants, identified from a few modest samples. Rapid progress has been made in quantifying the pattern of human LD and haplotypes defined by such common variants within and among populations. However, the quality and utility of the proposed LD-based resource could be seriously compromised if important sampling and analytical factors are overlooked in its design. The LD map should be based on adequately justified criteria defined by sound population genetic principles. Identification of genes that harbor variation associated with interindividual differences in risk of complex diseases remains one of the most challenging and important problems in human genetics. For genetic variants that are sufficiently common and have sufficiently large effects, direct tests of association through linkage disequilibrium with anonymous SNPs may prove effective. But the two critical

parameters—the frequency of risk inflating alleles and the magnitudes of their effect on risk — remain largely unknown [5, 6]

2. Material and Methods

2.1. Factors that Influence Linkage Disequilibrium

Mutation (Figure 1) and recombination (Figure 2) might have the most evident impact on linkage disequilibrium (LD), but there are additional contributors to the extent and distribution of disequilibrium. Most of these involve demographic aspects of a population, and tend to sever the relationship between LD strength and the physical distance between loci

- *Genetic drift.* This phenomenon describes the change in gene and haplotype frequency in a population every generation owing to the random sampling of gametes that occurs during the production of a finite number of offspring. Frequency changes are accentuated in small populations. In general, the increased drift of small, stable (not growing) populations tends to increase LD, as haplotypes are lost from the population. Such populations might be suitable for disease-gene mapping, with the idea that genetic drift will accentuate disease and marker allele frequency differences between cases and controls [7-9]. However, the applicability of this phenomenon to gene mapping has not been well characterized.

- *Population growth.* Rapid population growth decreases LD by reducing genetic drift.

- *Admixture or migration.* LD can be created by ADMIXTURE, or by migration (gene flow), between populations.

Initially, LD is proportional to the allele frequency differences between the populations, and is unrelated to the distance between markers. In subsequent generations, the ‘spurious’ LD between unlinked markers quickly dissipates, while LD between nearby markers is more slowly broken down by recombination. In theory, this would allow the mapping of disease genes in hybrid populations without using many genetic markers [10, 11]. Several admixed populations, such as African Americans and Hispanic Americans, have been characterized with this application in mind [12- 15], but the success of this approach will depend heavily on the time since admixture occurred, the frequency differences of the disease of interest in the parental populations and the allele frequency differences. So, the diseases and circumstances for which this mapping approach will be feasible might turn out to be quite rare and exceptional.

- *Population structure.* Various aspects of population structure are thought to influence LD. Population

subdivision is likely to have been an important factor in establishing the patterns of LD in humans, but most of our limited knowledge comes from the study of model organisms. An interesting recent study of *Arabidopsis* indicated that extreme inbreeding can produce high levels of LD without a substantial reduction in levels of variation [16- 18]. This neglected area would benefit from intensified study in humans.

- *Natural selection.* There are two primary routes by which selection can affect the extent of disequilibrium. The first is a hitchhiking effect, in which an entire haplotype that flanks a favoured variant can be rapidly swept to high frequency or even fixation [19, 20]. Although the effect is generally milder, selection against deleterious variants can also inflate LD, as the deleterious haplotypes are swept from the population [21- 23]. The second way in which selection can affect LD is through epistatic selection for combinations of alleles at two or more loci on the same chromosome [24, 25]. This form of selection leads to the association of particular alleles at different loci. Although this has provided a major motivation for historical studies of LD in *Drosophila* genetics, as a means of detecting the action of [epistatic] natural selection [26, 27], it has not yet been shown to alter LD in humans.

- *Variable recombination rates.* Recombination rates are known to vary by more than an order of magnitude across the genome. Because breakdown of LD is primarily driven by recombination; the extent of LD is expected to vary in inverse relation to the local recombination rate. It is even possible that recombination is largely confined to highly localized recombination hot spots, with little recombination elsewhere. According to this view, LD will be strong across the non-recombining regions and break down at hot spots. Although there are intriguing indications that this reflects the situation for some regions [28, 29], the generality of the hot-spot phenomenon, the strength of recombination in and outside hot spots, and the length distributions of these regions remain to be determined.

- *Variable mutation rates.* Some single-nucleotide polymorphisms, such as those at CpG dinucleotides, might have high mutation rates and therefore show little or no LD with nearby markers, even in the absence of historical recombination.

- *Gene conversion.* In a gene conversion event, a short stretch of one copy of a chromosome is transferred to the other copy during meiosis. The effect is equivalent to two very closely spaced recombination events, and can break down LD in a manner similar to recombination or recurrent mutation. It has recently been shown that rates of gene conversion in humans are high and are important in LD between very tightly linked markers [30- 32].

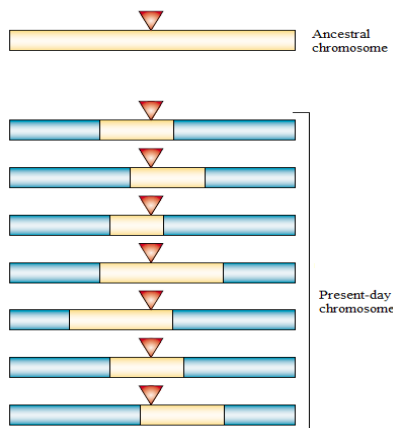


Figure 1. Linkage disequilibrium around an ancestral mutation. The mutation is indicated by a red triangle. Chromosomal stretches derived from the common ancestor of all mutant chromosomes are shown in yellow; and new stretches introduced by recombination are shown in blue. Markers that are physically close (that is, in the yellow regions of present-day chromosomes) tend to remain associated with the ancestral mutation even as recombination limits the extent of the region of association over time.

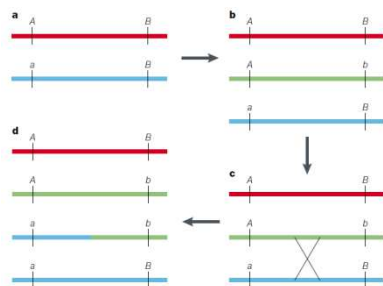


Figure 2. The erosion of linkage disequilibrium by recombination. a | At the outset, there is a polymorphic locus with alleles A and a. b | When a mutation occurs at a nearby locus, changing an allele B to b, this occurs on a single chromosome bearing either allele A or a at the first locus (A in this example). So, early in the lifetime of the mutation, only three out of the four possible haplotypes will be observed in the population. The b allele will always be found on a chromosome with the A allele at the adjacent locus. c | The association between alleles at the two loci will gradually be disrupted by recombination between the loci. d | This will result in the creation of the fourth possible haplotype and an eventual decline in LD among the markers in the population as the recombinant chromosome (a, b) increases in frequency.

2.1.1. Genetic Linkage Analysis

Today the identification of human disease genes rarely relies on the function, but on their locations in the human genome (positional cloning or reverse genetics). An obstacle in early human disease gene mapping was the lack of genetic markers, and blood groups and serum protein polymorphisms were the only available genetic markers. These markers were technically difficult to type and they only covered a small part of the human genome. This was overcome by the introduction of DNA restriction fragment length polymorphisms (RFLPs) in 1980 [33- 36]. RFLPs are based on single base pairs alterations that involve a cleavage site of a restriction endonuclease. The next generation of DNA markers constitutes the variable number of tandem repeats [VNTRs] [37, 38]. These highly

informative markers are scattered throughout the genome and usually reflects variations in repeat motifs of 10-20 bases per unit. A drawback of the RFLP and VNTR markers is that their typing involves the relatively time consuming Southern blot analysis and there is a requirement of large amounts of DNA for the analyses. Detection of susceptibility genes in indirect association studies depends not only on the degree of linkage disequilibrium between the disease variant and the SNP marker but also on the difference in their allele frequencies. Little is known about how variations in these parameters may affect the power of indirect association studies among related populations [39- 41].

Genetic linkage analysis is a statistical method that is used to associate functionality of genes to their location on chromosomes. Neighboring genes on the chromosome have a tendency to stick together when passed on to offsprings. Therefore, if some disease is often passed to offsprings along with specific marker-genes, then it can be concluded that the gene(s) which are responsible for the disease are located close on the chromosome to these markers. Accurate genetic maps are required for successful and efficient linkage mapping of disease genes. However, most available genome-wide genetic maps were built using only small collections of pedigrees, and therefore have large sampling errors. A large set of genetic studies genotyped by the NHLBI Mammalian Genotyping Service (MGS) provide appropriate data for generating more accurate maps [42- 46]. Genomic screening to map disease loci by association requires automation, pooling of DNA samples, and 3,000–6,000 highly polymorphic, evenly spaced microsatellite markers. Case-control samples can be used in an initial screen, followed by family-based data to confirm marker associations. Association mapping is relevant to genetic studies of complex diseases in which linkage analysis maybe less effective and to cases in which multi-generational data are difficult to obtain, including rare or late-onset conditions and infectious diseases. The method can also be used effectively to follow up and confirm regions identified in linkage studies or to investigate candidate disease loci. Study designs can incorporate disease heterogeneity and interaction effects by appropriate subdivision of samples before screening. Scientists reported, use of pooled DNA amplifications—the accurate determination of marker-disease associations for both case-control and nuclear family-based data—including application of correction methods for stutter artifact and preferential amplification. They combined with a discussion of both statistical power and experimental design to define the necessary requirements for detecting of disease loci while virtually eliminating false positives, suggest the feasibility and efficiency of association mapping using pooled DNA screening.

2.1.2. Mapping Human Disease Genes by Linkage Analysis

The International HapMap Project was designed to

create a genome-wide database of patterns of human genetic variation, with the expectation that these patterns would be useful for genetic association studies of common diseases. This expectation has been amply fulfilled with just the initial output of genome-wide association studies, identifying nearly 100 loci for nearly 40 common diseases and traits. These associations provided new insights into pathophysiology, suggesting previously unsuspected etiologic pathways for common diseases that will be of use in identifying new therapeutic targets and developing targeted interventions based on genetically defined risk. In addition, HapMap-based discoveries have shed new light on the impact of evolutionary pressures on the human genome, suggesting multiple loci important for adapting to disease-causing pathogens and new environments [47, 48].

2.1.3. Genomewide Scans of Complex Human Diseases

Many “complex” human diseases, which involve multiple genetic and environmental determinants, have increased in incidence during the past 2 decades. During the same time period, considerable effort and expense have been expended in whole-genome screens aimed at detection of genetic loci contributing to the susceptibility to complex human diseases. However, the success of positional cloning attempts based on whole-genome screens has been limited, and many of the fundamental questions relating to the genetic epidemiology of complex human disease remain unanswered. Multivariate analysis suggests that the only factors independently associated with increased study success are (a) an increase in the number of individuals studied and (b) study of a sample drawn from only one ethnic group. Positional cloning based on whole-genome screens in complex human disease has proved more difficult than originally had been envisioned; detection of linkage and positional cloning of specific disease-susceptibility loci remains elusive [49, 50].

2.2. Assessment of Linkage Disequilibrium by the Decay of Haplotype Sharing, with Application to Fine-Scale Genetic Mapping

Linkage disequilibrium (LD) is of great interest for gene mapping and the study of population history. We propose a multilocus model for LD, based on the decay of haplotype sharing (DHS). The DHS model is most appropriate when the LD in which one is interested is due to the introduction of a variant on an ancestral haplotype, with recombinations in succeeding generations resulting in preservation of only a small region of the ancestral haplotype around the variant. This is generally the scenario of interest for gene mapping by LD. The DHS parameter is a measure of LD that can be interpreted as the expected genetic distance to which the ancestral haplotype is preserved, or, equivalently, $1/(\text{time in generations to the ancestral haplotype})$. The method allows for multiple origins of alleles and for mutations, and it takes into account missing observations and ambiguities in haplotype determination, via a hidden Markov model. Whereas most commonly used measures of LD apply to

pairs of loci, the DHS measure is designed for application to the densely mapped haplotype data that are increasingly available. The DHS method explicitly models the dependence among multiple tightly linked loci on a chromosome. When the assumptions about population structure are sufficiently tractable, the estimate of LD is obtained by maximum likelihood. For more-complicated models of population history, Mary Sara McPeck and Andrew Strahs (1999) found means and covariances based on the model and solve a quasi-score estimating equation. Simulations showed that this approach works extremely well both for estimation of LD and for fine mapping. They applied the DHS method to published data sets for cystic fibrosis and progressive myoclonus epilepsy [51, 52].

2.3. Use of Gene-Specific Oligonucleotides

This approach relies on the ability to isolate sufficient protein product to permit amino acid sequencing. Specific peptide bonds in the protein product can be cleaved using proteolytic enzymes such as trypsin (cuts at the carboxyl end of lysine or arginine residues) or reagents such as cyanogen bromide (cuts at the carboxyl end of methionine residues). The amino acid sequence of each resulting peptide can be determined by chemical sequencing. This involves a repeated series of chemical reactions in an automated amino acid sequencer. In each cycle, the peptide is exposed to a chemical that covalently bonds to the N-terminal amino acid and cleaves it off, allowing it to be identified by chromatography. Sequence overlaps identify overlapping peptides, enabling longer sequences to be assembled.

The resulting amino acid sequence is inspected to identify regions containing amino acids with minimal codon degeneracy (e.g. methionines and tryptophans) are uniquely oligonucleotides low so as to increase the chance of identifying the correct target. Once a suitable cDNA clone is isolated, it can be used to screen a genomic DNA library in order to isolate genomic DNA clones for full characterization of the gene [53-55].

In most of the Western world, age-related macular degeneration (AMD) remains the largest single cause of severe visual impairment, and its prevalence continues to increase. It is considered to be a complex disease, in which multiple genes and environment play a role in pathogenesis. Several environmental insults are implicated with smoking, serum cholesterol, hypertension, sunlight exposure, and many other factors being variously associated with disease pathogenesis. Until recently, there have been relatively few breakthroughs to further our understanding of the genetics of AMD, despite remarkable progress in molecular genetic techniques over the last 20 years, and the fact that many rare inherited macular diseases have had their causative genes mapped. Development of new tools such as high-density single-nucleotide polymorphism chips and microarrays have changed the face of genetic research, but have yet to directly translate into improved clinical outcomes in ophthalmology. However with the recent

finding of the Tyr402His polymorphism in the complement factor H gene being implicated in AMD, scientists are about to witness a new wave of research in this disease. Not only does the identification of a biologically plausible gene identify a new pathway, but it also identifies new biological mechanisms for disease, avenues to pursue treatment, and a better understanding of how the environment interacts with the genetic background to create disease [55- 58].

Familial clustering of a disease is a direct indicator of a possible heritable cause, provided that environmental sharing can be excluded. If the familial clustering is lacking, the likelihood of a heritable influence is also small. In the era of genome scans, the consideration of data on heritability should be important in the assessment of the likely success of the genome scan [59, 60].

2.4. Examples

2.4.1. Schizophrenia

Family-based linkage disequilibrium mapping using SNP markers is expected to be a major route to the identification of susceptibility alleles for complex diseases. However there are a number of methodological issues yet to be resolved, including the handling of extended haplotype data and analysis of haplotype transmission in sib-pair or family trio samples. Javier Costas et al (2005) analysed two dinucleotide repeat and six SNP markers at the COMT locus at chromosome 22q11, a region implicated in psychosis, for transmission distortion in 198 Chinese schizophrenic family trios. When individual markers were analysed, Using haplotypes of paired markers by the program TRANSMIT, The global P value for the haplotypes of all six SNP markers tested and found, which may represent a background haplotype for the transmission of a schizophrenia susceptibility allele at chromosome 22q11. Their results supported the hypotheses that either COMT is itself a susceptibility gene, or more likely that this region of chromosome 22 contains a susceptibility gene that is in linkage disequilibrium with COMT alleles [61- 63].

Detection of susceptibility genes in indirect association studies depends not only on the degree of linkage disequilibrium between the disease variant and the SNP marker but also on the difference in their allele frequencies. Little is known about how variations in these parameters may affect the power of indirect association studies among related populations. Therefore, these differences have to be an additional factor to consider when a replication study fails to confirm initial associations, especially if the replication is focused on very few markers [64, 65].

2.4.2. Fragile X

An origin of bidirectional DNA replication was mapped to the promoter of the FMR1 gene in human chromosome Xq27.3, which has been linked to the fragile X syndrome. This origin is adjacent to a CpG island and overlaps the site of expansion of the triplet repeat (CGG) at the fragile X

instability site, FRAXA. The promoter region of FMR2 in the FRAXE site (approximately 600 kb away, in chromosome band Xq28) also includes an origin of replication, as previously described [34] et al. FMR1 transcripts were detected in foreskin and male fetal lung fibroblasts, while FMR2 transcripts were not. However, both FMR1 and FMR2 were found to replicate late in S phase (approximately 6 h into the S phase of normal human fibroblasts). The position of the origin of replication relative to the CGG repeat, and perhaps the late replication of these genes, might be important factors in the susceptibility to triplet repeat amplification at the FRAXA and FRAXE sites [66, 67].

3. Discussion

3.1. Likelihood Analysis of Disequilibrium Mapping, and Related Problems

The sampling distribution provides a basis for maximum-likelihood estimation of the recombination rate, the mutation rate, or the age of the allele, provided that the two other parameters are known. This theory is applied to [1], the data of Ha⁻ stbacka et al., to estimate the recombination rate between a locus associated with diastrophic dysplasia and a linked RFLP marker; [49], the data of Risch et al., to estimate the age of a presumptive allele causing idiopathic distortion dystonia in Ashkenazi jews; and [39], the data of Tishkoff et al., to estimate the date at which, at the CD4 locus, non-African lineages diverged from African lineages. Bruce Rannala and Montgomery Slatkin found that the extent of linkage disequilibrium can lead to relatively accurate estimates of recombination and mutation rates and that those estimates are not very sensitive to parameters, such as the population age, whose values are not known with certainty. In contrast, they also concluded that, in many cases, linkage disequilibrium may not lead to useful estimates of allele age, because of the relatively large degree of uncertainty in those estimates [68, 69].

3.2. Disequilibrium Likelihoods for Fine-Scale Mapping of a Rare Allele

Genetic linkage studies based on pedigree data have limited resolution, because of the relatively small number of segregations. Disequilibrium mapping, which uses population associations to infer the location of a disease mutation, provides one possible strategy for narrowing the candidate region. The coalescent process provides a model for the ancestry of a sample of disease alleles, and recombination events between disease locus and marker may be placed on this ancestral phylogeny. These events define the recombinant classes, the sets of sampled disease copies descending from the meiosis at which a given recombination occurred Jinko Graham and Elizabeth A. Thompson showed (1998) that, how Monte Carlo generation of the recombinant classes leads to linkage likelihood for fine-scale mapping from disease haplotypes.

They compared single-marker disequilibrium mapping with interval-disequilibrium mapping and discuss how the approach may be extended to multipoint-disequilibrium mapping. The method and its properties are illustrated with an example of simulated data, constructed to be typical of fine-scale mapping of a rare disease in the Japanese population. The method can take into account known features of population history, such as changing patterns of population growth [70, 71].

4. Conclusions

The debate over the average extent of levels of LD those are useful for association mapping is becoming narrower as data become available for more genomic regions and populations. Perhaps more importantly, it has become clear that the average extent might not be a good guide for the design and feasibility of LD mapping approaches. This is true for at least two reasons. First, the tremendous variability in the extent of LD from one region of the genome to another means that the average will greatly overstate the useful range for some regions and understate it for others. Second, even in a region of high mean LD, some pairs of loci do not show useful levels of LD due to gene conversion, differences in allele frequency and perhaps other factors. An important and almost entirely unanswered question is whether the patterns of LD found in one population will be replicated in other populations with differing population histories. What little data that can be applied to answering this question are conflicting, with hints that patterns of LD are similar among different populations equaled by indications that each population is substantially different. Answering this question, and establishing the generality (or not) of haplotype maps constructed in one population, should be an urgent priority for research. It is also worth briefly noting that the use of LD for mapping relies on assumptions regarding the genetic architecture of common diseases that are open to question; this point is discussed extensively elsewhere [72-74]. A less practical but perhaps more interesting question is what forces have shaped the patterns of LD in humans. An increasingly persuasive case can be made that simple demographic models of population expansions and contractions are insufficient to explain the observed patterns. More complex historical models might do better, but molecular forces, such as gene conversion [75] and recombination hot spots [76], have also recently come to the fore. Selection — positive, negative or balancing — must also have had an influence, but its role has been difficult to show conclusively. Sorting out these factors might occupy students of LD long after its more utilitarian uses have played themselves out.

Acknowledgement

This research was supported financially by a grant for scientific research from Graduate University of Advanced Technology, Kerman, Iran. Grant number: 1385-1/2487.

References

- [1] Aird IA, Vince GS, Bates MD, Johnson PM, Lewis-Jones ID. Leukocytes in semen from men with spinal cord injuries. *Fertil Steril* 1999; 72:97-103.
- [2] Haŕstbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver W, Lander ES. Linkage disequilibrium mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat Genet* 1992; 2:204-211.
- [3] Weiss, K. M. & Terwilliger, J. D. How many SNPs does it take to map a gene with SNPs? *Nature Genet* 2000; 26: 151-157.
- [4] Bennion A, Forshaw MJ. Insights from the experiences of older people with hearing impairment in the United Kingdom: recommendations for nurse-led rehabilitation. *Int J Older People Nurs.* 2012; 10: 1748-3743.
- [5] Kari H, Xinjun Li , Kristina S , Sundquist J. Familial risks for common diseases: Etiologic clues and guidance to gene identification, *Mutation Research* 2008; 658 247-258
- [6] Seymour R, Shirley BJ, Hogenesch H, Shultz LD, Sundberg JP. Loss of Function of the Mouse Sharpin Gene Results in Peyer's Patch Regression. *PLoS One.* 2013; 8:e55224.
- [7] Weiss, K. M. & Terwilliger, J. D. How many SNPs does it take to map a gene with SNPs? *Nature Genet* 2000; 26: 151-157.
- [8] Lambert, Brian W, Terwilliger, Joseph D, Weiss, Kenneth M. ForSim: a tool for exploring the genetic architecture of complex traits with controlled truth. *Bioinformatics* 2008; 24: 1821 -1822.
- [9] Wyart M, Botstein D, Wingreen NS. Evaluating gene expression dynamics using pairwise RNA FISH data. *PLoS Comput Biol.* 2010; 6:e1000979.
- [10] Chakraborty, R. & Weiss, K. M. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc. Natl Acad. Sci. USA* 1998; 85, 9119-9123.
- [11] McIsaac RS, Oakes BL, Wang X, Dummit KA, Botstein D, Noyes MB. Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res.* 2013; 41:e57.
- [12] Stamboliyska R, Parsch J. Dissecting gene expression in mosquito. *BMC Genomics.* 2011; 12 : 297.
- [13] Breuss M, Heng JI, Poirier K, Tian G, Jaglin XH, Qu Z, Braun A, Gstrein T, Ngo L, Haas M, Bahi-Buisson N, Moutard ML, Passemard S, Verloes A, Gressens P, Xie Y, Robson KJ, Rani DS, Thangaraj K, Clausen T, Chelly J, Cowan NJ, Keays DA. Mutations in the β -tubulin gene TUBB5 cause microcephaly with structural brain abnormalities. *Cell Rep.* 2012; 2: 1554-62.
- [14] Lambert, Brian W, Terwilliger, Joseph D, Weiss, Kenneth M. ForSim: a tool for exploring the genetic architecture of complex traits with controlled truth. *Bioinformatics* 2008; 24: 1821 -1822.
- [15] Wilson, J. F. & Goldstein, D. B. Consistent long-range linkage disequilibrium generated by admixture in a Bantu-Semitic hybrid population. *Am. J. Hum. Genet* 2000; 67: 926-935.

- [16] Nakamura Y, Leppert M, O'Connell P, Wolff R, Holm T, Culver M, Martin C, Fujimoto E, Hoff M, Kumlin E. Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* 1987; 235:1616-1622.
- [17] Deelen J, Beekman M, Capri M, Franceschi C, Slagboom PE. Bioessays. Identifying the genomic determinants of aging and longevity in human population studies: Progress and challenges. 2013; 10.1002/bies.201200148.
- [18] Luo L, Zhu Y, Xiong M. Quantitative trait locus analysis for next-generation sequencing with the functional linear models. *J Med Genet*. 2012; 49:513-24.
- [19] Nordborg M. The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genet* 2002; 30: 190–193.
- [20] Wang, W., Thornton, K., Berry, A. & Long, M. Nucleotide variation along the *Drosophila melanogaster* fourth chromosome. *Science* 2002; 295, 134–137.
- [21] Charlesworth, B., Morgan, M. T. & Charlesworth, D. The effect of deleterious mutations on neutral molecular variation. *Genetics* 1983; 134: 1289–1303.
- [22] Van den Berg SM, Service SK. Power of IRT in GWAS: Successful QTL Mapping of Sum Score Phenotypes Depends on Interplay Between Risk Allele Frequency, Variance Explained by the Risk Allele, and Test Characteristics. *Genet Epidemiol*. 2012; 10: 21680.
- [23] Melo C, Quintanilla R, Gallardo D, Zidi A, Jordana J, Díaz I, Pena RN, Amills M. Association analysis with lipid traits of two candidate genes (LRP12 and TRIB1) mapping to a SSC4 QTL for serum triglyceride concentration in pigs. *J Anim Sci*. 2013; [Epub ahead of print]
- [24] Cannon, G. B. The effects of natural selection on linkage disequilibrium and relative fitness in experimental populations of *Drosophila melanogaster*. *Genetics*. 1963; 48: 1201–1216.
- [25] Stadhouders R, Kolovos P, Brouwer R, Zuin J, van den Heuvel A, Kockx C, Palstra RJ, Wendt KS, Grosveld F, van Ijcken W, Soler E. Multiplexed chromosome conformation capture sequencing for rapid genome-scale high-resolution detection of long-range chromatin interactions. *Nat Protoc*. 2013; 8: 509-24.
- [26] Kruglyak, L. & Lander, E. S. High-resolution genetic mapping of complex traits. *Am. J. Hum. Genet* 1995; 56: 1212–1223.
- [27] Kobak S, Berdeli A. Fas/FasL promoter gene polymorphism in patients with rheumatoid arthritis. *Reumatismo*. 2012; 64:374-9.
- [28] Javier C, Torres M, Cristobo I, Christopher Ph, Carracedo. A Relative efficiency of the linkage disequilibrium mapping approach in detecting candidate genes for schizophrenia in different European populations, *Genomics* 2005; 86:280 – 286.
- [29] Fingerlin TE, Abecasis GR, Boehnke M. Using sex-averaged genetic maps in multipoint linkage analysis when identity-by-descent status is incompletely known. *Genet Epidemiol*. 2006; 30: 384–396.
- [30] Ardlie, K. Lower-than-expected linkage disequilibrium between tightly linked markers in humans suggests a role for gene conversion. *Am. J. Hum. Genet* 2001; 69: 582–589.
- [31] Collins F S. Positional cloning moves from the perdictional to the traditional. *Nature Genet* 1995; 9:347–350.
- [32] Meyer WK, Arbeithuber B, Ober C, Ebner T, Tiemann-Boege I, Hudson RR, Przeworski M. Evaluating the evidence for transmission distortion in human pedigrees. *Genetics*. 2012; 191: 215-32.
- [33] Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 1980;32: 314-331.
- [34] Stewart WC. Improving estimates of genetic maps: a meta-analysis-based approach. *Genet Epidemiol*. 2007;31:408–416
- [35] Tekin G, Ertan P, Horasan G, Berdeli A. SPP1 gene polymorphisms associated with nephrolithiasis in Turkish pediatric patients. *Urol J*. 2012; 9:640-7.
- [36] Septiningsih EM, Ignacio JC, Sendon PM, Sanchez DL, Ismail AM, Mackill DJ. QTL mapping and confirmation for tolerance of anaerobic conditions during germination derived from the rice landrace Ma-Zhan Red. *Theor Appl Genet*. 2013; [Epub ahead of print]
- [37] Jeffreys, A. J., Kauppi, L. & Neumann, R. Intensely punctuate meiotic recombination in the class II region of the major histocompatibility complex. *Nature Genet* 2001; 29: 217–222.
- [38] McPeck MS, Strahs A. Assessment of Linkage Disequilibrium by the Decay of Haplotype Sharing, with Application to Fine-Scale Genetic Mapping, *Am. J. Hum. Genet*. 1999; 65:858–875.
- [39] Andrew G Clark. Finding genes underlying risk of complex disease by linkage disequilibrium mapping. *Current Opinion in Genetics & Development* 2003; 13:296–302.
- [40] Stewart WC, Thompson EA. Improving estimates of genetic maps: a maximum likelihood approach. *Biometrics*. 2006;62: 728–734
- [41] Durmaz E, Flanagan S, Berdeli A, Semiz S, Akcurin S, Ellard S, Bircan I. Variability in the age at diagnosis of diabetes in two unrelated patients with a homozygous glucokinase gene mutation. *J Pediatr Endocrinol Metab*. 2012; 25:805-8.
- [42] Chunsheng H, Daniel E, Steven B, Goncalo R, William C, Tara C. Enhanced genetic maps from family-based disease studies: population-specific comparisons, The Enhanced Map Consortium, *BMC Medical Genetics* 2011; 12:15
- [43] Wilson, J. F. & Goldstein, D. B. Consistent long-range linkage disequilibrium generated by admixture in a Bantu–Semitic hybrid population. *Am. J. Hum. Genet* 2000; 67: 926–935.
- [44] Fingerlin TE, Abecasis GR, Boehnke M. Using sex-averaged genetic maps in multipoint linkage analysis when identity-by-descent status is incompletely known. *Genet Epidemiol* 2006; 30: 384–396.
- [45] Stewart WC, Subaran RL. Obtaining accurate p values from a dense SNP linkage scan. *Hum Hered*. 2012; 74: 12-6.
- [46] Stewart WC, Drill EN, Greenberg DA. Finding disease genes: a fast and flexible approach for analyzing high-throughput data. *Eur J Hum Genet*. 2011; 19: 1090-4.

- [47] Keck GE, Poudel YB, Rudra A, Stephens JC, Kedei N, Lewin NE, Peach ML, Blumberg PM. Molecular modeling, total synthesis, and biological evaluations of C9-deoxy brystatin 1. *Angew Chem Int Ed Engl*. 2010; 49: 4580-4.
- [48] Matise TC, Chen F, Chen W, De La Vega FM, Hansen M, He C, Hyland FC, Kennedy GC, Kong X, Murray SS, Ziegler JS, Stewart WC, Buyske S. A second-generation combined linkage physical map of the human genome. *Genome Res*. 2007; 17: 1783-1786.
- [49] Altmüller J, Lyle J, Palmer A, Guido F, Hagen S, Matthias W. Genomewide Scans of Complex Human Diseases: True Linkage Is Hard to Find. *Am. J. Hum. Genet* 2001; 69:936-950.
- [50] Andrew G, Cohen A, Salgaonkar S, Patel V. The explanatory models of depression and anxiety in primary care: a qualitative study from India. *BMC Res Notes*. 2012; 10:1756-0500-5-499.
- [51] Berger MF, Hodis E, Heffernan TP, Deribe YL, Lawrence MS, Protopopov A, Ivanova E, Watson IR, Nickerson E, Ghosh P, Zhang H, Zeid R, Ren X, Cibulskis K, Sivachenko AY, Wagle N, Sucker A, Sougnez C, Onofrio R, Ambrogio L, Auclair D, Fennell T, Carter SL, Drier Y, Stojanov P, Singer MA, Voet D, Jing R, Saksena G, Barretina J, Ramos AH, Pugh TJ, Stransky N, Parkin M, Winckler W, Mahan S, Ardlie K, Baldwin J, Wargo J, Schadendorf D, Meyerson M, Gabriel SB, Golub TR, Wagner SN, Lander ES, Getz G, Chin L, Garraway LA. Melanoma genome sequencing reveals frequent PREX2 mutations. *Nature*. 2012; 485:502-6.
- [52] Matt C, Baird P, Dirani M, Ophth Sc, Guymer R. Unraveling A Complex Genetic Disease: Age-related Macular Degeneration. *Surv Ophthalmol* 2006; 51:576-586.
- [53] Pfaff, C. L. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. *Am. J. Hum. Genet*. 2001; 68: 198-207.
- [54] Botstein D. Why we need more basic biology research, not less. *Mol Biol Cell*. 2012; 21: 4160-1.
- [55] Yang C, Wang L, Zhang S, Zhao H. Accounting for Non-Genetic Factors by Low-Rank Representation and Sparse Regression for eQTL Mapping. *Bioinformatics*. 2013; [Epub ahead of print]
- [56] Li T, Ball D, Zhao J, Murray RM, Liu X, Sham PC, Collier DA. Family-based linkage disequilibrium mapping using SNP marker haplotypes: application to a potential locus for schizophrenia at chromosome 22q11, *Molecular Psychiatry* 2000; 5: 77-84.
- [57] Hu T, Andrew AS, Karagas MR, Moore JH. Statistical epistasis networks reduce the computational complexity of searching three-locus genetic models. *Pac Symp Biocomput*. 2013:397-408.
- [58] Courtney SM, Massett MP. Identification of exercise capacity QTL using association mapping in inbred mice. *Physiol Genomics*. 2012; 44:948-55.
- [59] Jeffreys AJ, Wilson V, Thein SL. Hypervariable 'minisatellite' regions in human DNA. *Nature* 1985; 314:67-73.
- [60] Chen C, Qiao R, Wei R, Guo Y, Ai H, Ma J, Ren J, Huang L. A comprehensive survey of copy number variation in 18 diverse pig populations and identification of candidate copy number variable genes associated with complex traits. *BMC Genomics*. 2012; 13:733.
- [61] Lewontin RC, Feldman MW. A general asymptotic property of two-locus selection models. *Theor Popul Biol*. 1988; 34: 177-93.
- [62] Lang GI, Botstein D, Desai MM. Genetic variation and the fate of beneficial mutations in asexual populations. *Genetics*. 2011; 188:647-61.
- [63] Xu Q, Mei G, Sun D, Zhang Q, Zhang Y, Yin C, Chen H, Ding X, Liu J. Detection of genetic association and functional polymorphisms of UGDH affecting milk production trait in Chinese Holstein cattle. *BMC Genomics*. 2012; 13:590.
- [64] Connes P, Machado R, Hue O, Reid H. Exercise limitation, exercise testing and exercise recommendations in sickle cell anemia. *Clin Hemorheol Microcirc*. 2011; 49: 151-63.
- [65] Kutukculer N, Gulez N, Karaca NE, Aksu G, Berdeli A. Novel mutations and diverse clinical phenotypes in recombination-activating gene 1 deficiency. *Ital J Pediatr*. 2012; 38:8.
- [66] Bruna P, Brylawski, Paul D. Chastain II, Stephanie M. Cohen, Marila Cordeiro-Stone, David G. Kaufman. Mapping of an origin of DNA replication in the promoter of fragile X gene FMR1. *Experimental and Molecular Pathology* 2007; 82: 190-196.
- [67] de la Peña P, Machín A, Fernández-Trillo J, Dominguez P, Barros F. Mapping of interactions between the amino and carboxy termini and the channel core in hERG K⁺ channels. *Biochem J*. 2013; [Epub ahead of print]
- [68] Quintana, P. J. E., Neuwirth, E. A. H. & Groszovsky, A. J. Interchromosomal gene conversion at an endogenous human cell locus. *Genetics* 2001; 158: 757-767.
- [69] Hulse AM, Cai JJ. Genetic variants contribute to gene expression variability in humans. *Genetics*. 2013; 193:95-108.
- [70] Graham J, Thompson E. Disequilibrium Likelihoods for Fine-Scale Mapping of a Rare Allele, *Am. J. Hum. Genet* 1998; 63:1517-1530.
- [71] Frisse, L. Gene conversion and different population histories may explain the contrast between polymorphism and linkage disequilibrium levels. *Am. J. Hum. Genet* 2001; 69: 831-843.
- [72] Weiss LA, Arking DE; Gene Discovery Project of Johns Hopkins & the Autism Consortium, Daly MJ, Chakravarti A. A genome-wide linkage and association scan reveals novel loci for autism *Nature* 2009; 61: 802-8.
- [73] Yang Z, Rannala B. Molecular phylogenetics: principles and practice. *Nat Rev Genet*. 2012; 28: 303-14.
- [74] Stover DA, Verrelli BC. Comparative vertebrate evolutionary analyses of type I collagen: potential of COL1a1 gene structure and intron variation for common bone-related diseases. *Mol Biol Evol*. 2011; 28: 533-42.
- [75] Pritchard, J. K. Are rare variants responsible for susceptibility to complex diseases? *Am. J. Hum. Genet*. 2001; 69: 124-137.

- [76] Suto J. Quantitative trait locus mapping of genes associated with vacuolation in the adrenal X-zone of the DDD/Sgn inbred mouse. *BMC Genet.* 2012; 13:95.