

PublisherInfo		
PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

## Whole-genome scan for common breast cancer?

ArticleInfo		
ArticleID	:	3612
ArticleDOI	:	10.1186/bcr-1999-66590
ArticleCitationID	:	66590
ArticleSequenceNumber	:	32
ArticleCategory	:	Paper Report
ArticleFirstPage	:	1
ArticleLastPage	:	4
ArticleHistory	:	RegistrationDate : 1999-7-15 OnlineDate : 1999-7-15
ArticleCopyright	:	Current Science Ltd1999
ArticleGrants	:	
ArticleContext	:	1305811

## Keywords

association study, genome scan, linkage disequilibrium, SNP maps

---

## Introduction

Breast cancer is one of a class of common diseases with a genetic component: the first-degree relatives of breast cancer sufferers have approximately a two-fold increased risk of the disease over the general population. The linkage approach has proved very valuable for cloning the highly penetrant breast cancer susceptibility genes; BRCA1 and BRCA2. However, whole genome association studies that are likely to be more powerful for identifying the multiple genetic variants that are expected to increase breast cancer risk in the general population. Presently, association studies are carried out using the candidate gene (or direct) approach: polymorphisms in genes of known function, that might feasibly have a role in breast cancer development are identified, then the allele frequencies are compared between breast cancer cases and controls. However, almost a decade's work has yielded very few leads to date. Recently, dense maps of SNPs [single nucleotide polymorphisms - bi-allelic (eg C to T) markers] have begun to be published. SNPs are suitable for automated high-throughput detection methods, and we now have the prospect, therefore, of using an indirect approach-carrying out association studies for all common diseases using neutral markers spaced throughout the genome. The theory is that any neutral allele found more frequently in cases than in controls must be in linkage disequilibrium (LD) with, and therefore physically close to, a truly functional cause of breast cancer. A number of genes with involvement in cancer susceptibility could theoretically be identified in a single (albeit very large) experiment, without any prior knowledge about the biological role of the genes. The major issue in the planning of such studies is; how close together do these neutral markers need to be? This will determine how many SNPs will be needed to properly cover the whole genome and make sure that no susceptibility genes are missed.

## Aims

This paper is a theoretical attempt to address the feasibility of whole genome association studies to search for common disease alleles.

## Comments

This paper describes whole genome association studies to assess links between common disease alleles involved in breast cancer. The findings show that if 2000 breast cancer cases and 2000 controls were used in the study (probably the minimum needed to achieve significant results) it would be necessary to carry out two billion genotypings! Even with the high-throughput prospects of being able to genotype 100,000 samples per week, such a study would take 400 person-years and would currently cost at least one billion Euros/Dollars. However, as the author himself comments, everything is dependent on the assumptions made. Empirical studies have already shown that in certain genome regions significant LD is maintained over more than 100 kb. If this is generally the case, it may be quite possible to conduct a study using fewer than 50,000 SNPs - a 10-fold reduction in the work and cost - and a magnitude on the verge of practical feasibility.

## Methods

Kruglyak has carried out computer simulation experiments, incorporating a number of different assumptions into his models to investigate these questions. His assumptions are that all the polymorphic markers used are bi-allelic, have arisen only once, are common in the population (ie have allele frequencies between 25 and 75%) and are neutral to natural selection. He has used  $d^2$  as a measure of LD [this varies between 0 (no LD) and 1] and he has set a value of 0.1 as being a useful amount of LD. The recombination frequency is assumed to be a constant 1% per 1 Mb (million base pairs) across the genome. He has also assumed that historically human populations remained at a constant size (N) until exponential expansion began a number of generations (G) ago, and he has varied N and G in his simulations.

## Results

The first results presented are for the general out-bred human population. With an allele frequency of 50%, there would be strongest LD between a marker SNP and a disease-causing variant if they were 300 bp apart, some LD would be detectable at 10 kb, but there would be nothing if the two were as much as 300 kb apart. Thus, there would have to be a marker SNP at least every 6 kb, meaning half a million SNPs would be needed to cover the whole genome. It has generally been believed that fewer SNP markers would be required to carry out the same study in an isolated (founder) population because LD is expected to be greater in these populations. However, Kruglyak suggests that this may not be the case when common SNPs and disease markers are being examined. A common variant will be old - probably older than the isolated population - and will have been introduced by many members of the initial founding population, so LD may not be increased. However, this situation might improve if the isolated population remained small for a long time and has only expanded recently, since it would then behave in the same way as a population founded by only very few individuals.

# Discussion

Using his assumptions, Kruglyak estimates that a comprehensive scan in an out-bred population would need an SNP to be positioned every 3 kb, and would thus require between 0.5-1.0 million SNPs to cover the entire genome. It is expected that there will be one SNP every 1 kb in the genome, so there should be sufficient markers available to use and almost all of them will be required. Isolated populations derived from a very small number of founders might significantly reduce the number of SNPs required. These results are entirely dependent on the initial assumptions, given that human population history is very complex, these may well be erroneous. In particular, the probable migration ?out of Africa? may have generated the very small number of founders necessary to maintain LD between common alleles in most of the world?s populations. In addition, natural selection and the likely existence of hot and cold spots for recombination may also dramatically alter the distances over which significant LD can be detected. Thus, the estimate given here may well be the least optimistic scenario.

## References

1. Disis ML, Grabstein KH, Sleath PR, Cheever MA: Prospects for whole genome linkage disequilibrium mapping of common diseases. *Nat Genet.* 1999, 22: 139-144.