

## **Novel common low penetrance breast cancer predisposition genes.**

**Alison Dunning, University of Cambridge**

### **Background**

Breast cancer is a common disease with a genetic component. Approximately 75% of genes which contribute to familial risk are unidentified at present. This study was designed to identify common low penetrance alleles that contribute to breast cancer susceptibility.

### **Method**

A genome-wide association study using SNPs was applied. The aim was to identify as many as possible of the common variants that each explain greater than 1% of the genetic variance in breast cancer. A big advantage of this method is that there is no need for prior knowledge of gene function. The theoretical basis is that a disease-causing common variant should be in linkage disequilibrium with nearby markers (SNPs) so that SNPs with a different frequency between breast cancer cases and controls can indicate proximity to an allele that leads to an increased risk of breast cancer.

### **Identifying significant SNPs**

A tagging set of SNPs (266,722) representative of the whole genome was isolated. A phase study design was applied. In phase 1 440 cases with a strong family history but no mutation in BRCA1/2 were matched with 400 controls. The 266,722 SNPs were genotyped and those that showed the most variation between cases and controls were identified. The 13,023 SNPs which showed greatest variance ( $p < 0.05$ ) were taken forward into phase 2.

In phase 2 the 13,023 SNPs were genotyped in 4,000 breast cancer cases and controls. A higher level of significance ( $p < 0.0001$ ) was used to identify the best 30 SNPs for indicating variance. In phase 3 these 30 SNPs were genotyped in more than 30,000 cases and controls.

At the end of phase 3, 8 of the 30 SNPs were confirmed as significant and this represents 6 new breast cancer loci which were previously unidentified. Three of the top 6 loci are associated with having a family history of breast cancer and ductal carcinoma in situ (DCIS). None of the loci had previously been studied as candidate genes although 2 were plausible candidates. One of the loci is in a desert region, >100kb from any known gene.

The SNP that displayed the most significant variation ("locus A") is a G>A common polymorphism. GA heterozygotes had about a 1.22 relative risk of breast cancer and AA homozygotes 1.56. The p value ( $10^{-18}$  after phase 2) indicated this variation was significant. It was concluded that there are no common low penetrance alleles in the population that convey a relative risk much larger than this as the study design would have been expected to identify them.

### **Finding the causal variant**

The next problem is moving from the associated SNP to the causal variant. This is a difficult and expensive process. A haplotype block is defined around the SNP and then resequenced to find every possible polymorphism in the block. Association and linkage disequilibrium is then used to identify a minimum set of possible causal SNPs.

For example, at locus A: mapping of the locus identified the gene closest to locus A. Each of tagged SNPs from the original phase 1 scan that fall within this region were mapped. Examining the association significance for each one of these SNPs identified that the most significant region is present within intron 2. Resequencing of this haplotype block identified 30 potentially functional SNPs. Only 3 haplotypes for the 30 SNPs are present in the European population therefore it is impossible to isolate the causal variant further. Asian case-controls were used to eliminate all but 6 SNPs. Functional assays must now be performed to identify which is most likely to be the causal variant.

Data were also presented to show how 17 SNPs have been identified as the possible causal variant at locus C.

### **Conclusions**

- The genome wide association study was an effective method to find common low penetrance breast cancer alleles (even though the genetic contribution to the disease is not large)

- The next difficulty will be mapping the causal variant having found the association.
- Populations less closely related than Europeans may need to be identified to make it possible to determine the causal variant from the associated SNP.