

Commentary

The promise and limitations of genome-wide association studies to elucidate the causes of breast cancer

Christine B Ambrosone

Roswell Park Cancer Institute, Elm & Carlton Streets, Buffalo, NY 14263, USA

Corresponding author: Christine B Ambrosone, Christine.ambrosone@roswellpark.org

Published: 5 December 2007

This article is online at <http://breast-cancer-research.com/content/9/6/114>

© 2007 BioMed Central Ltd

Breast Cancer Research 2007, **9**:114 (doi:10.1186/bcr1787)

Abstract

With the characterization of the human genome, as well as advances in technology to determine genetic variability across the genomes of populations, there has been focused effort on the identification of cancer susceptibility alleles through the use of genome-wide association studies. These efforts have recently resulted in identification of a susceptibility locus for breast cancer by several groups, although the increases in risk are modest. While genome-wide association studies will probably lead to discoveries of potentially important previously unstudied pathways in cancer etiology, the role of the environment, particularly gene-environment interactions, in breast cancer etiology should not be overlooked.

The 'known' risk factors for breast cancer – including ionizing radiation, breast cancer in a first-degree relative, reproductive and hormonal factors, alcohol consumption and physical activity – explain only a portion of the variability in breast cancer risk. Of these risk factors, other than ionizing radiation, a family history of breast cancer is responsible for the greatest increase in risk, with women with a first-degree relative with breast cancer having twice the risk of those who do not. A twin study indicated that up to 30% of breast cancer cases may be due to genetic factors [1]; however, the 'high-risk' genes that have been identified, such as *BRCA1*, *BRCA2*, *PTEN* and *p53*, explain 20–25% of familial breast cancer [2] and explain only 5% of total breast cancer [3]. There has thus been focused research on identification of additional genetic variants responsible for susceptibility to the disease. Because studies of familial breast cancer have failed to identify additional genes that infer high risk of breast cancer, it is thought the remaining genetic factors are likely to be numerous, with each genetic variant inferring a low to moderate risk.

There have thus been concerted efforts to identify these genes, facilitated by the characterization of the human genome and by rapid advances in technology, which now allow for interrogation across the genome for differences in

DNA sequence, or single nucleotide polymorphisms (SNPs), between cases and controls in search of common disease susceptibility genes. Taking this genome-wide association (GWA) approach, susceptibility loci for prostate cancer on chromosome 8q24 were recently identified, alleles which were also identified using family studies and linkage analysis (reviewed in [4]).

Most recently, several reports appeared from studies using a GWA study approach to breast cancer, all replicating findings for SNPs in fibroblast growth factor receptor 2, which may be involved in regulating gene expression. It is of interest that findings were replicated in three studies based in different types of populations. Using pooled data from more than 20,000 women with breast cancer and an equal number of control individuals for the final analysis, the authors of one study began a three-stage approach by comparing SNPs among cases with a strong family history of breast cancer with healthy control individuals to maximize the likelihood of identifying genes associated with inherited risk [5]. An additional study published at the same time found similar results in two data sets of postmenopausal women with sporadic, or nonfamilial, breast cancer [6], the first set of which was the Harvard Nurses' Health Study Nurses Health Study. A third study was conducted among an Icelandic population [7], although this study was limited to breast tumors positive for estrogen receptor expression. In the study enriched for familial breast cancer [5], it was estimated that the SNPs identified in fibroblast growth factor receptor 2 explained only 3.6% of familial risk, which would translate to much lower proportions in nonfamilial breast cancer. Although the estimates for increases in risk were slight to moderate (20–60%) in all studies, the variants are common in Caucasian populations; in the Nurses Health Study and the validation cohort [6], including primarily postmenopausal women with breast cancer, it was estimated that the population attributable risk associated with the variant was 16%.

GWA = genome-wide association; SNP = single nucleotide polymorphism.

These recent studies illustrate the power of GWA studies in large sample sizes to identify gene variants that may increase risk of breast cancer, although these are not high-penetrance genes. Perhaps the greater ramifications for these findings are that they identify pathways that have not been previously explored, and they open new doors for basic science and epidemiologic studies to identify additional causes of breast cancer.

There are obvious limitations to the GWA approach. First, although it is probable that the technology currently used captures the majority of common variants, based upon the concept of linkage disequilibrium (that blocks of chromosomes are inherited together and can be 'tagged' by a defined, limited set of SNPs within those blocks), it is possible that some genetic variants that may be important susceptibility alleles are not covered by the SNPs that are genotyped.

Furthermore, for the majority of genetic variants, it is probable that effects will only be noted in the context of exposures that either induce expression of that gene or are associated with increased or decreased risk in the context of pertinent exposures. Although we know that breast cancer in a first-degree relative increases risk of breast cancer, there are also substantial contributions to risk by reproductive and hormonal factors, as well as physical activity, alcohol consumption, radiation to the breast, and some dietary factors, probably the result of exposures over years or decades.

Failure to account for exposures when searching for genes that may increase cancer risk may obscure any associations between genetic variants and risk. For example, *N*-acetyltransferase is an enzyme that metabolizes aromatic amines, and a large proportion of Caucasians have variants resulting in slow activity or detoxification. In studies of bladder cancer, in which aromatic amines are a strong risk factor, there is generally no main effect of the genotype. It is only among those who are exposed to aromatic amines that *N*-acetyltransferase slow genotypes are associated with increased risk, with a recent large study finding elevated risk for the slow *N*-acetyltransferase genotype in smokers (odds ratio = 1.6, 95% confidence interval = 1.3–1.9) but no effect (odds ratio = 0.09, 95% confidence interval = 0.6–1.3) in non-smokers [8]. We noted similar associations for breast cancer in relation to smoking [9], which has been confirmed by recently pooled and meta-analysis [10]. Because of the complex systems of metabolism of numerous endogenous and exogenous compounds, it is likely that gene variants associated with increased risk only among populations with specific exposures will be overlooked when searching for the main effects of SNPs.

It is therefore possible that findings from GWA studies for breast cancer will not result in identification of genes that explain a large proportion of breast cancer risk; for example, those with effects that would only be noted among sub-

groups with relevant exposures. A striking example of the importance of environment in breast cancer etiology is that of the effects of migration on breast cancer incidence. Japanese women have the lowest rates of breast cancer in the world, but their risk rises to that of Caucasians within one or two generations after migrating to the United States. Obviously, genetic make-up is not changing, and the preponderance of increased risk is probably due to external factors [11].

The study of gene–environment interactions in cancer etiology has always been difficult, due to the complex mixtures of multiple factors to which humans are exposed and due to the heterogeneity of carcinogenic pathways. The investigation of the effects of common variants on breast cancer risk, without consideration of epidemiologic factors and exposures, may therefore not be as productive as application of the GWA approach to other outcomes, such as predicting patient outcomes following radiation and/or chemotherapy for cancer. In these situations, the key exposures (therapeutic agents) are known and can be measured, and SNPs that may predict severe toxicity and or disease recurrence among patients who receive treatment may clearly reflect these gene–environment interactions. It is probable that, in the coming years, networks of genes that mediate treatment outcomes will be elucidated, and GWA studies in the context of clinical trials will lead to pathways to aid in personalized therapeutics.

Although GWA studies may reveal some previously unknown variants to be important in cancer risk, the further elucidation of the complex interactions between multiple genes and environmental exposures should not be overlooked. If GWA studies do not result in finding major genes that increase the risk of breast cancer, or that replicate those from candidate genes previously studied in the context of environmental exposures, there is a possibility for disillusionment with the study of genetic variability in relation to cancer risk. This could result in a movement away from hypothesis-driven evaluation of the effects of genetic variability on relationships between exposures and cancer risk, and could squelch potentially significant research into the causes of breast cancer, particularly those that are modifiable. It is therefore advisable to proceed cautiously with the conduct of and, particularly, the interpretation of studies designed to examine the effects of common SNPs on breast cancer risk, particularly in the absence of consideration of exposures. This is an exciting age of progress in and discovery of the role that the human genome plays in disease risk and outcome, and there is hope that the convergence of genomic approaches and molecular epidemiologic studies of gene–environment interactions will provide growing insight into the causes of and, importantly, the prevention of breast cancer.

Competing interests

The author declares that they have no competing interests.

References

1. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K: **Environmental and heritable factors in the causation of cancer – analyses of cohorts of twins from Sweden, Denmark, and Finland.** *N Engl J Med* 2000, **343**:78-85.
2. Peto J, Collins N, Barfoot R, Seal S, Warren W, Rahman N, Easton DF, Evans C, Deacon J, Stratton MR: **Prevalence of BRCA1 and BRCA2 gene mutations in patients with early-onset breast cancer.** *J Natl Cancer Inst* 1999, **91**:943-949.
3. Antoniou AC, Pharoah PD, McMullan G, Day NE, Stratton MR, Peto J, Ponder BJ, Easton DF: **A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes.** *Br J Cancer* 2002, **86**:76-83.
4. Witte JS: **Multiple prostate cancer risk variants on 8q24.** *Nat Genet* 2007, **39**:579-580.
5. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struwing JP, Morrison J, Field H, Luben R, *et al.*: **Genome-wide association study identifies novel breast cancer susceptibility loci.** *Nature* 2007, **28**:1087-1093.
6. Hunter DF, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, Wacholder S, Wang Z, Welch R, Hutchinson A, *et al.*: **A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer.** *Nat Genet* 2007, **39**:870-874.
7. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, Masson G, Jakobsdottir M, Thorlacius S, Helgason A, *et al.*: **Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer.** *Nat Genet* 2007, **39**:865-869.
8. Garcia-Closas M, Malats N, Silverman D, Dosemeci M, Kogevinas M, Hein DW, Tardon A, Serra C, Carrato A, Garcia-Closas R, *et al.*: **NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses.** *Lancet* 2005, **366**:610-612.
9. Ambrosone CB, Freudenheim JL, Graham S, Marshall JR, Vena JE, Brasure JR, Michalek AM, Laughlin R, Nemoto T, Gillenwater KA, *et al.*: **Cigarette smoking, N-acetyltransferase genetic polymorphisms, and breast cancer risk.** *JAMA* 1996, **276**: 1494-1501.
10. Ambrosone CB, Kropp S, Yang J, Yao S, Shields PG, Chang-Claude J: **Cigarette smoking, N-acetyltransferase 2 genotypes, and breast cancer risk: pooled analysis and meta-analysis.** *Cancer Epidemiol Biomarkers Prev* 2007, **16**:1-12.
11. Ziegler RG, Hoover RN, Pike MMC, Hildesheim A, Nomura AM, West DW, Wu-Williams AH, Kolonel LN, Horn-Ross PL, Rosenthal JF, *et al.*: **Migration patterns and breast cancer risk in Asian-American women.** *J Natl Cancer Inst* 1993, **85**:1819-1827.