The completion of nine large genome-wide association studies (1–9) has focused attention on single-nucleotide polymorphisms (SNPs) as risk factors for breast cancer. Despite considerable progress, the scientific yield, practical application, and commercial exploitation of SNP profiling for disease risk remain topics of debate (10–13).

In this context, the study by Milne et al. (14) offers an excellent example of both the promise and challenges of current genetic epidemiological approaches to SNP genotyping for breast cancer risk. Drawing on an international consortium of 25 centers contributing DNA from more than 30,000 breast cancer case patients and 35,000 healthy control subjects, these investigators sought to confirm if an SNP that was not detected in their initial scan was associated with risk for breast cancer. They confirmed that this SNP on the long arm of chromosome 2 was associated with a 1.12-fold risk for invasive breast cancer in white women of European origin. Unlike the initial report (2), they found an association in both estrogen receptor–negative and estrogen receptor–positive breast cancer. Although highly statistically significant ($P = 10^{-19}$), the clinical validity and biological implications of these findings, and those of many genome-wide association studies, require additional research.

For some readers, the first hurdle in interpreting this, and other, genome-wide association studies relates to the statistical indexes of risk that are used. For example, the aggregate “per-allele” risk of 1.12 cited by Milne et al. (14) is derived from a log additive model. Although it is the best way to compare risks of different SNPs in different studies (see Table 1), per-allele risk is not the measure of greatest interest to patients or clinicians. For women homozygous for rs13387042, the risk of breast cancer increases 28% from baseline (odds ratio [OR] = 1.28). For heterozygotes, the risk increases only 9% (OR = 1.09). The discriminatory value of the SNP may therefore be limited because the “at-risk” allele frequency of rs13387042 is so high (51%).

Other derived measures of risk in studies of SNPs are the “excess familial risk” and the “population attributable risk” (PAR%), also called the “population attributable fraction” (15). Both of these indexes reflect measures of the population frequency of the variant, as well as the relative risk (or odds ratio). Excess familial risk is the amount of hereditary cancer risk in a first-degree relative of a cancer patient that is due to the genetic variant in question. For very low penetrance SNPs, the amount of familial risk accounted for by the SNP is very low; only 0.6% for rs13387042 as reported by Milne et al. (14). Another measure of risk commonly used by epidemiologists, the PAR%, can be calculated to be 5.8% for...
rs13387042. This means that 5.8% of breast cancer in the population is associated with this low-risk variant. Citing the PAR% can be misleading in genome-wide association studies because of the high frequency of the variants and because a given cancer case may carry multiple SNPs in the germline that interact with other genetic variants or with nongenetic factors. The cumulative PAR% of the breast cancer SNPs in Table 1 could exceed 100% (16).

These comments underscore the special challenges in interpreting so-called “common variant, common disease” studies of polygenic susceptibility to cancer (15,17). Unlike the established rare, but “high-penetrant” mutations of genes such as BRCA1, BRCA2, and TP53, the more common variants discovered by genome-wide approaches are of the “low-penetrance” variety. The SNP-associated risks of breast cancer are only slightly greater than baseline, for example, homozygous carriers of rs13387042 have a risk increase comparable to that of delaying first pregnancy until after age 35 years (13). As shown in Table 1, the minor allele frequencies of these SNPs are high, varying from 25% for the 16q12 SNP to 76% for the 14q24.1 SNP (RAD51L1). The frequencies of the same SNP can also vary according to ancestry. Such population heterogeneity is well illustrated in the study by Milne et al. (14); rs13387042 is five times more prevalent in Caucasians than Asians. Similarly, different SNPs on the long arm of chromosome 6 have been noted in those of Asian ancestry compared with those of Eastern European Ashkenazi ancestry (7,8). Such variability in frequency and potential variability in risk associated with SNPs across populations remain major challenges for clinical translation of these findings.

In addition to the very low relative risks and population heterogeneity associated with currently available breast cancer SNPs, other scientific as well as clinical challenges will emerge as new variants are discovered. Most exciting among the scientific challenges is unraveling the functional biology of genome-based associations. As with other loci such as the 8q24, SNPs associated with risks for prostate, breast, and colon cancers, the SNP studied by Milne et al. (14) lies in a gene “desert.” Pending the documentation of mechanisms such as alteration of binding sites for transcription factors in the regions of implicated SNPs, pharmacological approaches to “targeted” prevention of breast cancer risk based on SNP profiling will be limited. Because current SNPs, including rs13387042, together account for only approximately 5% of the excess familial risk for breast cancer (17), ongoing studies are needed to address structural variation in the germline, as well as interactions of variants in common pathways.

As evident to clinicians, the modest magnitude of increased risk associated with rs13387042, and other individual SNPs to date, does not approach levels required to consider, for example, preventive surgeries, hormonal chemoprevention, or intensified breast surveillance. As with other SNPs, SNPs associated with risks for prostate, breast, and colon cancers, the SNP studied by Milne et al. (14) lies in a gene “desert.” Pending the documentation of mechanisms such as alteration of binding sites for transcription factors in the regions of implicated SNPs, pharmacological approaches to “targeted” prevention of breast cancer risk based on SNP profiling will be limited. Because current SNPs, including rs13387042, together account for only approximately 5% of the excess familial risk for breast cancer (17), ongoing studies are needed to address structural variation in the germline, as well as interactions of variants in common pathways.

Finally, there is a potential pitfall in the path ahead, which has little to do with the science of genomics and more to do with economics. This concern stems from the currently unregulated, for-profit marketing of SNP-based disease risk information directly to consumers. Professional societies as well as regulatory agencies should be encouraged to consider the most responsible means to prevent the false alarm and false reassurance that may result from premature dissemination of SNP-based risk profiling outside of a research context. (13)
References


Notes

The author thanks Peter Thom, MS, for Table 1; Mia Gaudet, PhD, Mark Robson, MD, for critical reading; and the Robert and Kate Niehaus Clinical Cancer Genetics Research Initiative.