Searching for Blood DNA Methylation Markers of Breast Cancer Risk and Early Detection

Montserrat García-Closas, Mitchell H. Gail, Karl T. Kelsey, Regina G. Ziegler

Correspondence to: Regina G. Ziegler, PhD, MPH, Epidemiology and Biostatistics Program, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Executive Plaza S 8098, Bethesda, MD, 20892-7246 (e-mail: regina.ziegler@nih.gov).

In this issue of the Journal, Xu et al. (1) used prospectively collected blood samples to determine whether epigenome-wide methylation profiles in blood DNA differ between women who subsequently develop breast cancer and those who do not. Potentially, such a profile could predict the risk of developing breast cancer or detect this disease days to several years before it appears clinically. Existing breast cancer risk prediction models have limited discriminatory accuracy for this common disease (2), and improved methods for early detection are also needed (3,4). In this context, the findings of Xu et al. are promising and exemplify the potential value of epigenome-wide association studies (5,6). However, several important considerations temper the initial enthusiasm and highlight challenges for future studies.

In the Sister Study, a cohort of women with a biologic sister with breast cancer, DNA methylation at 27 578 CpG sites was compared in blood samples that were stored at study entry from 298 women who developed breast cancer during follow-up and from a random sample of 612 women who remained cancer-free. This study is possibly the first cancer study to use prospectively collected blood and an epigenome-wide assay platform with single nucleotide resolution. Several epigenome-wide association studies have reported methylation profile differences in blood DNA from cancer case subjects and control subjects but relied on blood samples collected after cancer diagnosis and treatment (7). This retrospective design can induce bias if methylation levels are affected by disease progression or treatment.

As the authors discuss, a major limitation of the Xu et al. report is the lack of replication in independent study populations. In a small replication sample of the 25 case subjects and 56 non-case subjects excluded from the original analysis because of diverse ethnicity, the authors found promising indications of discriminatory accuracy for the five most statistically significant CpGs. However, what is required to confirm the promise of epigenome-wide association studies are independent studies of methylation markers using prospective epidemiological designs and larger sample sizes.

The mean time from baseline blood draw to diagnosis among the breast cancer cases was only 1.3 years. Thus the data cannot tell us whether epigenetic changes can predict risk years into the future or are, instead, a response to incipient disease. For 72% of the 250 differentially methylated CpG sites, mean methylation values for the case subjects with bloods collected more than 1 year before diagnosis were intermediate between the methylation values for the non-case subjects and the case subjects with blood collected within a year before diagnosis. For all CpGs in the array, methylation values were intermediate for case subjects with blood collected more than 1 year before diagnosis at statistically significantly fewer CpGs (34%). These findings suggest that the epigenetic changes are early markers of disease. However, cohort studies with longer follow-up and serial blood collections are needed to estimate lead times, clarify biology, and apply appropriate methods for evaluating predictive value (8).

By design, all women in the Sister Study cohort had a family history of breast cancer. Therefore, the results from this study may not generalize to populations with lower levels of genetic risk. In particular, as the authors discuss, the risk discrimination estimates in this study might not be comparable with estimates in the general population, and risk models that incorporate family history, such as the Gail model, would be expected to have diminished discriminatory accuracy.

Xu et al. assessed whether there was differential methylation at each of 27 578 CpG sites across the epigenome using case-cohort proportional hazard regression, adjusted only for age and laboratory parameters. After correction for multiple testing using a false discovery rate threshold of 0.05, Xu et al. identified 250 CpGs associated with breast cancer. The distribution of $P$ values (Supplementary Figure 1A, available online) indicates that many CpGs were differentially methylated. This broad signal could be a pervasive response to early, not-yet-diagnosed disease. It is also possible the signal could reflect bias or the influence on methylation of known or as-yet-undiscovered breast cancer risk factors, which were not controlled in the proportional hazard analyses.

To evaluate their ability to distinguish between case subjects and non-case subjects, Xu et al. used repeated internal cross-validation (9) to obtain unbiased estimates of the area under the receiver operating characteristic curve (AUC), which measures discriminatory accuracy. They used training data (two-thirds of the data) to both reselect promising CpG sites and fit classifiers and then used the remaining data to estimate AUC (Supplementary Methods, available online). This entire process was repeated 500 times to obtain stable estimates. By reselecting CpGs and developing new classifiers each time, they avoided overestimating AUC (10).

In this study, the AUC estimated for methylation markers (65.8%) was larger than for the Gail model (56.0%) or nine highly ranked single nucleotide polymorphisms from genome-wide association studies of breast cancer (58.8%). When Gail model predictors or single nucleotide polymorphisms were added to the methylation markers, the methylation AUC was only marginally improved (<1.0%). However, it would have been of interest to...
examine AUC improvement when adding methylation markers to the Gail and single nucleotide polymorphism models. Adding more predictors to AUC estimates generally leads to diminishing improvement. Furthermore, if the methylation markers are detecting early disease and the Gail and genetic models are identifying risk factors, then AUC models may not be directly comparable.

Ideally, all three models should be tested and compared for both short-term and long-term discriminatory accuracy in independent study populations. Internal validation, as was conducted for the methylation markers, is never as rigorous a test as external validation (11). However, Xu et al. did not present an explicit risk model or classifier based on the promising CpGs, and such a model is required for external validation in independent studies.

The 27K methylation array used in this study targets promoter regions of translated genes and has limited coverage of the methylome. Much denser arrays have been developed that interrogate methylation sites across the promoter and in the gene body, and outside of CpG islands (12). Thus future studies can be more comprehensive. Improved technologies, including bisulfite sequencing, will allow better characterization of the true nature and behavior of DNA methylation patterns identified as being predictive of disease. These patterns, precisely described and perhaps integrated with companion histone marks (13), will almost certainly advance our understanding of the biology.

In population studies, the extent of methylation at one specific CpG site in blood DNA does not vary substantially, partly because methylation status is averaged over a number of cells, a variety of leukocyte cell types, and many individuals. Indeed for the 250 promising CpGs identified in this study, the mean β values used to quantify methylation generally differed between case subjects and non-case subjects by only about 1%. To find reliable differences between case subjects and non-case subjects, laboratory variation in measurement must not overwhelm between-subject variation. Intraclass correlation coefficients estimate how much of total variation can be explained by between-subject variation (14). However, the intraclass correlation coefficients presented by Xu et al. for methylation markers are likely to be inflated because they were based on quality control samples created in the laboratory and not on patient samples. The study by Xu et al. illustrates the complexities of such work but also provides encouraging results for replication and extension. Prospective studies using the next generation of genome-wide methylation platforms and complementary epigenetic assays, collaborative efforts across studies to reach adequate statistical power and solid replication, further development of statistical methods, and improved understanding of the biology of DNA methylation and epigenetic control should advance our knowledge in this promising area of research.

References

The diagnosis and management of ductal carcinoma in situ (DCIS) is controversial (1). With widespread mammography screening, diagnosis of DCIS became more prevalent. Some are uncertain whether this has translated into a decrease in invasive cancer and a subsequent decline in breast cancer mortality. Part of the concern has been that frequently the treatments of DCIS are as extensive as for invasive cancer with a similar panoply of risks. A straightforward approach to selecting the optimum therapy—defined here as the minimum needed to avoid recurrence, particularly with an invasive component—is needed. Many solutions have been proposed, but none has gained wide acceptance. For example, the Van Nuys Prognostic Index has been in common use for decades (2). Several randomized clinical trials have compared lumpectomy alone to lumpectomy followed by radiation treatment, but no subset analysis of these results has found a group that does not benefit from radiation with a lower in-breast recurrence risk (3).

In this issue of the Journal, Solin et al. propose a measure, the 12-gene Oncotype DX DCIS Score (DCIS score) to assist clinicians and patients with the task of deciding optimal treatment (4). Genomic Health is currently marketing the assay (http://www.oncotypedx.com/en-US/Breast/HealthcareProfessionalsDCIS). The Oncotype DX score for invasive breast cancer developed by Genomic Health has found widespread acceptance. It also forms the basis for two clinically important ongoing trials. The first, TailorX, is to determine for women with early-stage, hormonal-positive, node-negative disease and an intermediate score whether adjuvant chemotherapy is needed in addition to hormonal therapy. This trial has closed to accrual and is in follow-up. The other study, spearheaded by Southwest Oncology Group (SWOG) and currently open, RxPonder (Rx for Positive Node, Endocrine Responsive Breast Cancer), will reveal whether chemotherapy benefits patients with node-positive breast cancer who have low to intermediate Oncotype DX recurrence scores.

Solin et al. collaborated with Genomic Health in a rigorous predefined manner to develop the DCIS Score. How well does it work? Should it be adopted? Is it needed for affordable, accountable care? Developmental work used DCIS specimens alone, invasive breast cancer specimens alone, and specimens in which DCIS was adjacent to invasive breast cancer. Initially, the Oncotype DX recurrence assay now in clinical use for invasive cancer was tested, and a wide range of recurrence scores was found, dovetailing with what has been observed about the heterogeneous nature of DCIS. The primary concern was the risk of an in-breast recurrence, scored as an ipsilateral breast event, which included invasive and noninvasive recurrence. The first task was to specifically assess recurrence risk in databases of invasive carcinoma for the 21 individual genes. Seven genes were found to be predictive of recurrence risk. One was progesterone receptor (PR), and a second was a glutathione S-transferase that functions in the detoxification of electrophilic compounds (GSTMI). Five of the genes clustered in a “proliferation group”—Ki67, STK15, Survivin, cyclin B1 (CCNB1), and MYBL2. Interestingly, for invasive cancer the proliferation genes were best categorized with a threshold, whereas for DCIS a continuous score was found to be best. The genetic drivers of progression from DCIS to invasive cancer are poorly understood; the drivers behind proliferation may serve as an avenue for better elucidating this phenomenon. Estrogen receptor (ER) was correlated with hormone therapy benefit, which is logical given that the mechanism of action of tamoxifen is through binding with the ER. A limitation of the Solin et al. study is that 97.2% of the participants in the validation set were ER positive and 29.4% received tamoxifen, so the investigators did not want to use a gene that was predictive of hormonal therapy benefit.

A set of pure intraductal carcinoma, 96 case subjects, from Marin General Hospital, Marin, California, were used for setting specific cutpoints for low, intermediate, and high risk of recurrence. This preset selection of genes and scale formed the DCIS Score that the researchers then validated in a completely