Epigenome-wide methylation in DNA from peripheral blood as a marker of risk for breast cancer

Gianluca Severi, PhD ^{1,2,3}; Melissa C Southey, PhD ⁴; Dallas R English, PhD ^{1,2}; Chol-hee Jung, PhD ⁵; Andrew Lonie, PhD ⁵; Catriona McLean, MD ^{6,7}; Helen Tsimiklis, MSc ⁴; John L Hopper, PhD ^{2,8}; Graham G Giles, PhD ^{1,2,9}; Laura Baglietto PhD ^{1,2}.

¹ Cancer Epidemiology Centre, Cancer Council of Victoria, Melbourne, Australia.

² Centre for Epidemiology and Biostatistics, School of Population and Global Health, University of Melbourne, Australia.

³ Human Genetic Foundation, Torino, Italy

⁴ Genetic Epidemiology Laboratory, Department of Pathology, University of Melbourne

⁵ Life Sciences Computation Centre, Victorian Life Sciences Computation Initiative, Carlton, Victoria, 3010, Australia

⁶ Department of Pathology, Alfred Hospital, Melbourne, Australia.

⁷ Department of Medicine, Monash University, Melbourne, Australia.

⁸ Seoul National University, Seoul, Korea.

⁹ Department of Epidemiology and Preventive Medicine, Monash University, Melbourne, Australia.

Corresponding author: Assoc Prof Laura Baglietto

Cancer Epidemiology Centre

Cancer Council Victoria

615 St Kilda Road, Melbourne

Victoria 3004, AUSTRALIA

Phone: +39 388 658 5450

e-mail: laura.baglietto @cancervic.org.au

ABSTRACT

Purpose

Aberrant DNA methylation is a key feature of breast carcinoma. We aimed to test the association between breast cancer risk and epigenome-wide methylation in DNA from peripheral blood.

Methods

Nested case-control study within the prospective Melbourne Collaborative Cohort Study. DNA was extracted from before-diagnosis blood samples (420 incident cases and matched controls). Methylation was measured with the Illumina Infinium HumanMethylation 450 BeadChip array. Odds ratio (OR) for epigenome-wide methylation, quantified as the mean beta values across the CpGs, in relation to breast cancer risk were estimated using conditional logistic regression.

Results

Overall, the OR for breast cancer was 0.42 (95% CI 0.20-0.90) for the top versus bottom quartile of epigenome-wide DNA methylation and the OR for a one standard deviation (SD) increment was 0.69 (95% CI 0.50-0.95; test for linear trend, $p=0.02$). Epigenome-wide DNA methylation of CpGs within functional promoters was associated with an increased risk whereas epigenome-wide DNA methylation of genomic regions outside promoters was associated with decreased risk (test for heterogeneity, $p=0.0002$). The increased risk associated with epigenome-wide DNA methylation in functional promoters did not vary by time between blood collection and diagnosis, whereas the inverse association with epigenome-wide DNA methylation outside functional promoters was strongest when the interval from blood collection to diagnosis was less than 5 years and weakest for the longest interval.

Conclusions

Epigenome-wide methylation in DNA extracted from peripheral blood collected before diagnosis may have potential utility as markers of breast cancer risk and for early detection.

INTRODUCTION

DNA methylation is a key process involved in the regulation of gene expression. Aberrant methylation can disrupt normal gene regulation and expression and is involved in many human diseases including cancer [1-3]. Possible mechanisms leading to cancer development and progression include hypermethylation of CpG islands within the promoters of tumour suppressor genes causing the loss of expression [3], and hypomethylation causing chromosomal instability and increased genetic mutation events [4].

Breast cancer tissue is known to present lower levels of methylation on a epigenome-wide scale than normal breast tissue, but few specific genes have been reported to be hypomethylated in breast cancer cells [5, 6].

A small number of studies have attempted to investigate the role of DNA methylation as a marker of risk, or for early detection, classification and prognosis of breast cancer [7-13]. For instance, methylation of selected genes in DNA from breast ductal fluids and from plasma has been shown to be a potential marker for early detection [14, 15]. DNA methylation is of particular interest because it is potentially modifiable and is related to age, the strongest breast cancer risk predictor [16]. Various measures of global DNA methylation in peripheral blood have been investigated, mainly by case-control studies, in relation to colorectal [17, 18], head and neck [19] bladder [20], and breast cancer [21-24] with some promising results, though with limitations in interpretation given the samples for cases were taken after diagnosis. Another limitation of some of these studies is that they used measures of methylation in specific genetic elements (e.g. LINE-1) as surrogates of epigenome-wide methylation.

In this study, we hypothesised that epigenome-wide methylation, measured as the mean methylation level across the genome and assessed in DNA extracted from peripheral blood using the Illumina Infinium HumanMethylation 450 BeadChip array (Infinium HM450) [25], is associated with the risk of breast cancer.

MATERIALS AND METHODS

Subjects and samples

To estimate the association between breast cancer risk and aberrant DNA methylation from samples collected at baseline, we conducted a nested case-control study within the Melbourne Collaborative Cohort Study and used the Infinium HM450 produced by Illumina to quantify DNA methylation in more than 480,000 CpG sites covering up to 96% of CpG islands as well as a large number of non-island CpG sites (approximately 2% of all CpG sites in the genome) [25, 26].

Cases were women with a first diagnosis of invasive adenocarcinoma of the breast (International Classification of Diseases for Oncology, C50) occurring between baseline interview and 31 December 2007. Cases were ascertained by record linkage to the population-based Victorian Cancer Registry (VCR), and to the Australian Cancer Database. Controls were selected through density sampling and matched with cases on year of birth, year of baseline attendance, country of origin and, when possible, type of baseline blood specimen (dried blood spot (DBS), buffy coat, or lymphocyte). Tumour characteristics included tumour grade, size, nodal status, oestrogen receptor status (ER), progesterone receptor status (PR) and epidermal growth factor 2 status (Her2).

Samples were distributed into 96-well plates and processed by Infinium HM450 in chips of 12 arrays (8 chips per plate) with case-control pairs arranged consecutively on the same chip.

Further details about study design, DNA extraction, methylation measures and data processing are provided in Online Resource.

Statistical analysis

The dataset consisted of 420 cases and 420 matched controls and included methylation measures for 481,273 CpGs. For each sample, we quantified epigenome-wide methylation as the mean beta values across CpGs after removing the CpGs that were missing for at least one member of the matched pair.

We calculated the within-pair difference of epigenome-wide methylation between cases and controls and applied linear mixed effect models with plate and chip as random effects to estimate the difference between cases and controls overall, by type of biospecimen and by time from blood collection.

We estimated odds ratios for breast cancer risk by conducting a conditional logistic regression analysis using methylation quartiles based on the distribution of the mean beta levels in controls. The quartile cut points were as follows: <0.5123 , $0.5123-0.5196$, $0.5197-0.5264$, >0.5264 . In addition, we estimated the odds ratio for a one standard deviation (SD) increment in epigenome-wide methylation by fitting a model including the pseudocontinuous variable calculated for each woman as the median within the corresponding quartile. Odds ratios were estimated overall, by time from blood collection (less than 5 years, 5 to 9 years, 10 to 15 years) and by tumour characteristics.

Using the manifest file provided by Illumina, we classified CpGs according to their position relative to CpGs islands (Island/Shore or Shelf/Other) or relative to repetitive elements (within or outside repetitive elements); or their genomic regulatory feature (Promoter or Other). For each of these characteristics, we calculated the mean beta for the CpGs in each category and estimated the odds ratio for of breast cancer for each category fitting conditional logistic regression with and without adjustment for the other category.

Fitted conditional logistic models were evaluated using the Likelihood Ratio Test (LRT) and, for analysis of global methylation as a continuous variable, the Area Under the Receiver Operating Characteristic Curve (AUC). An AUC of 0.5 indicates discrimination no better than chance, while a value of 1.0 indicates perfect discrimination.

Sensitivity analyses were conducted (*i*) using only samples with DNA extracted from DBS; (*ii*) using only samples with DNA extracted from lymphocytes or buffy coat; (*iii*) excluding women with a diagnosis of any invasive cancer before baseline; (*iv*) adjusting for breast cancer risk factors (alcohol; energy from diet; body mass index; age at menarche; parity; hormone replacement therapy use; oral contraceptive use; menopausal status; (*v*) adjusting for comorbidity indicators (diagnosis of cancer other than breast cancer before and within 5 years since baseline, any diagnosis of asthma or diabetes before baseline, any diagnosis of angina, hypertension, heart attack or stroke before baseline); (*vi*) adjusting the analysis of DBS samples for the predicted cell-type proportion [27]; (*vii*) calculating epigenome-wide methylation after excluding cross-reactive probes or polymorphic probes [28].

We performed principal component analysis (PCA), estimated odds ratios for the first 10 principal components (PCs) and their correlation with epigenome-wide DNA methylation overall and by CpG genomic feature.

Statistical analyses were conducted with the software R version 3.0.1.

RESULTS

At blood collection, women were between 38 and 70 years of age (mean=56, SD=8, Online Resource Table 1). Seventy-seven percent of the biospecimens were DBS and 23% buffy coats or lymphocytes. Age at diagnosis for the cases ranged between 44 and 83 years (mean=64, SD=8), with 31% of the cases diagnosed within 5 years from blood collection, 35% between 5 and 10 years and 34% after 10 years.

Epigenome-wide methylation was lower for cases than controls (difference between cases and controls, -0.0011 (0.11%); $p=0.006$). The difference in epigenome-wide methylation between cases and controls was similar for lymphocytes/buffy coat and DBS biospecimens ($p=0.98$). Mean and SD of global methylation was 0.5190 (52%) and 0.0101 (1%) for all samples, 0.5186 (52%) and 0.0100 (1%) for cases and 0.5195 (52%) and 0.0101 (1%) for controls, respectively.

The difference in epigenome-wide methylation between cases and controls decreased with time from blood collection (test for linear trend, $p=0.005$, Figure 1) and this trend with time was similar for DNA extracted from DBS and from lymphocytes/buffy coat (test for heterogeneity of the trend with time by type of biospecimen, $p=0.40$). The difference in epigenome-wide methylation between cases and controls was -0.0028 (0.28%) ($p=5 \times 10^{-4}$) when the interval between blood collection and diagnosis was less than 5 years and negligible (i.e. -0.0005 (0.05%), $p=0.36$) when the interval was 5 years or more (Figure 1, bottom panels).

The odds ratio for breast cancer was 0.42 (95% CI 0.20, 0.90) for the top versus bottom quartile of epigenome-wide DNA methylation and 0.69 (95% CI 0.50-0.95) for an increment of one SD (0.01;1%) in epigenome-wide methylation (test for linear trend, $p=0.02$, Table 1). The odds ratio for breast cancer for a one SD increment in epigenome-wide methylation was 0.49 (95% CI 0.28, 0.86) when the interval between blood collection and diagnosis was less than 5 years, 0.59 (95%CI: 0.33, 1.06) for 5 to 9 years and 1.08 (0.64-1.84) for 10 to 15 years (test for heterogeneity by time from blood collection, $p=0.05$). There was evidence of heterogeneity for the association of epigenome-wide methylation with breast cancer risk by HER2 status (test for heterogeneity, $p=0.01$) whereas the odds ratios were similar for all categories of tumour grade and size, nodal status, or hormone receptor status (all p -values from test for heterogeneity >0.05) (Table 1).

There was evidence of heterogeneity for the association of epigenome-wide methylation with risk of breast cancer by CpG genomic location (Table 2). Epigenome-wide methylation in CpGs far from CpG islands was lower for cases than for controls, with an odds ratio for a one SD increment of 0.63 (95% CI 0.48, 0.82) whereas there was no association with breast cancer risk for epigenome-wide methylation in CpGs within Islands or Shores (test for heterogeneity, $p=0.03$). Epigenome-wide methylation in CpGs within functional promoters was associated with an increased risk of breast cancer (OR: 1.82; 95% CI 1.20, 2.75) whereas epigenome-wide methylation in genomic regions outside promoters was associated with a decreased breast cancer risk (OR: 0.63, 95% CI 0.46, 0.85; test for heterogeneity, $p=0.0002$). When CpGs within functional promoters were split into those on or around CpG islands (island or shore) and those far from CpG islands (shelf or none), the odd ratios were similar (test for heterogeneity, $p=0.92$). When CpG sites within TSS200 and TSS1500 regions were divided into those on or around CpG islands (island or shore) and those far from CpG islands (shelf or none), the odds ratios were close to unity for islands and

shores and inverse for those far from CpG islands (shelf or none) OR: 0.62, 95% CI 0.47, 0.82, test for heterogeneity, $p=0.0041$). The odds ratios were similar to that for epigenome-wide methylation within and outside repetitive elements (test for heterogeneity, $p=0.08$).

All conditional logistic regression models including epigenome-wide DNA methylation overall or by CpG genomic feature fitted significantly better than the null model ($p<0.05$ for all LRT). However, all of them had low discriminatory ability as suggested by their AUC of around 0.6.

The epigenome-wide methylation in functional promoters was slightly, but not statistically significantly, higher for cases than for controls and did not vary with the interval from blood collection to diagnosis (test for linear trend, $p=0.48$; Figure 2); the odds ratio for an increment of 1 SD of epigenome-wide methylation in functional promoters was 1.77 (95% CI 1.20, 2.75) and did not vary by time to diagnosis (test for heterogeneity, $p=0.52$). Conversely, epigenome-wide methylation outside functional promoters was lower for cases than for controls and the difference decreased with increasing time between blood collection and diagnosis (test for linear trend, $p=0.0002$; Figure 2) as decreased the strength of its inverse association with breast cancer risk (test heterogeneity by time to diagnosis, $p=0.10$; Table 3). Epigenome-wide DNA methylation in CpG islands was similar for cases and controls irrespective of time to diagnosis (test for trend, $p=0.62$) and it was not associated with an increased risk (OR for 1 SD, 1.02; 95% CI 0.50, 2.05; test for heterogeneity by time to diagnosis, $p=0.51$). Outside of CpG islands, epigenome-wide DNA methylation was lower for cases than controls with the difference linearly decreasing with time to diagnosis (test for trend, $p=0.002$); also the strength of the inverse association with breast cancer risk decreased with increasing time since blood collection (test for heterogeneity by time to diagnosis, $p=0.10$; Table 3). This was also observed for CpG sites limited to those with TSS200 and TSS1500 regions for less than 10 years prior to diagnosis (OR for 1SD, 0.47; CI 95% 0.29,

0.77; OR for 1SD, 0.51; CI 95% 0.30, 0.85) for <5 year and 5-9 years respectively (test for heterogeneity by time to diagnosis, $p=0.05$).

Results did not materially change in any of the sensitivity analyses conducted (results not shown).

The first 10 PCs explained up to 42% of the total variance of the CpGs, with the first PC alone explaining the 27% (Online Resource Table 2 and Online Resource Figure 2). Only the first, second, fifth and eighth PCs were associated either with breast cancer risk or with a linear trend in the odds ratio by time since blood collection (Online Resource Table 3). The first and second PCs, which were directly correlated with epigenome-wide methylation ($\rho=0.56$ and $\rho=0.67$ respectively), were associated with a reduced risk of breast cancer that did not vary by time since blood collection. The fifth PC was inversely correlated with epigenome-wide methylation in the functional promoters ($\rho=-0.65$) and was associated with a reduced risk of breast cancer for large intervals between blood collection and diagnosis, consistently with the increased risk observed for epigenome-wide methylation in functional promoters. The eighth PC was associated with an increased risk of breast cancer in the first 5 years since blood collection; this PC is not readily interpreted in terms of CpG genomic features.

DISCUSSION

Using a nested case-control design within our prospective cohort study, we have observed a relatively strong inverse association between epigenome-wide DNA methylation (quantified using data from a methylation array) and breast cancer risk. This inverse association was limited to short intervals between blood collection and diagnosis and to CpG

sites outside functional promoters, while epigenome-wide methylation calculated across CpG sites within functional promoters was associated with an increased risk of breast cancer independent of time from blood collection. For analyses limited to CpG sites outside of islands and shores, the inverse association was observed for the TSS200 and TSS1500 regions.

The strengths of our study include its prospective design, relatively large sample size, matched design, plating of case-control pairs within the same chip to significantly reduce the impact of the technical variability of the beta values on the estimates of association, and the availability of detailed data on tumour characteristics. Also, we quantified methylation using the Infinium HM450, a high-density array with good coverage of the most important classes of CpGs [25]. The limitations of our study include the heterogeneity in the type of biospecimens from which DNA was extracted and its lack of power to detect moderate differences in association by tumour characteristics.

A variety of methods to quantify global DNA methylation are available [29] and some have been used to investigate the association between global methylation in DNA from whole blood or white blood cells and breast cancer risk [30]. Consistent with our findings, a retrospective case-control study found that global DNA methylation measured by mass spectroscopy of leukocyte DNA as the percentage of 5-methyldeoxycytosine (%5-mdC) with respect to the total cytosine content was lower for breast cancer cases than for controls [22]. A similar measure of global DNA methylation in leukocyte DNA was found to be inversely associated with bladder cancer risk in a large retrospective case-control study [20]. Other studies have measured global DNA methylation in repetitive elements (e.g. LINE-1) and found hypomethylation in many cancers [19, 31-34]. No association has been found between leukocyte DNA methylation in LINE-1 and breast cancer risk [22, 24] but different measures of global DNA methylation obtained with different techniques might give different results. In

the study of Choi *et al.*, LINE-1 methylation and %5-mdC levels were not correlated, and only hypomethylation quantified as %5-mdc level was significantly associated with breast cancer risk [22]. In another study that measured global DNA methylation with both the luminometric methylation assay (LUMA) and with a LINE-1 assay, Xu *et al.* found that higher LUMA methylation measures were associated with breast cancer risk whereas no association was observed between LINE-1 measures and breast cancer risk [24]. The LUMA assay measures levels of 5-^mC in the C^mCGG motif, which is over-represented in gene promoters in the genome, and the result reported by Xu *et al.* would therefore be consistent with our observation of an increased risk associated with epigenome-wide DNA methylation within promoters. Measures by the LUMA assay were not found to be associated with breast cancer risk by another retrospective study that used a sibling design [23].

To the best of our knowledge, no study so far has published the use of DNA methylation arrays to test the hypothesis that epigenome-wide methylation of DNA from blood is associated with cancer risk but we know that this is being actively investigated in other cohort studies such as EPIC [35] A recently published case-cohort study of breast cancer, within the prospective National Institute of Environmental Health Sciences Sister Study, used the previous version of the Infinium DNA methylation array interrogating ~27,000 CpG sites but focused only on the identification of differentially methylated CpG sites (dmCpG) [11]. Interestingly, for three quarters of the 250 dmCpGs identified in that study at a false discovery rate of less than 0.05, methylation levels for cases were lower than those for non-cases and the difference in methylation levels relative to non-cases was generally greater for cases diagnosed within a year after blood collection than for cases diagnosed after a longer follow-up. Recognising the need for replication, we have exchanged some of our findings with EPIC investigators and found that using the HM450K our estimates of global methylation and breast cancer risk are very similar [based on 164 case

control pairs, the EPIC OR per 1SD is 0.52, CI 95% 0.37, 0.72, $p=0.00009$] (Personal communication Professors van Veldhoven and Vineis, 10 November 2014).

The interpretation of our findings can only be speculative because epigenetic patterns are generally considered to be tissue specific and it is difficult to understand why methylation patterns measurable in DNA extracted from blood is associated with the risk of cancer in the breast. One possible explanation, consistent with the observed epigenome-wide hypomethylation of CpGs outside the promoter observed for biospecimens collected close to diagnosis, is that cancer developing in the breast alters body metabolisms regulating methyl transfer reactions, thus causing a systemic instability in the DNA methylation status. Another explanation, consistent with the hypermethylation of CpGs within promoters observed also in biospecimens collected many years before diagnosis, is the presence of a common mechanism, perhaps related to 1-carbon metabolisms or influencing the regulation of methyl transfer reactions, that influences both DNA methylation in blood cells and the development of breast cancer.

The strength of the associations we observed are of the same order of magnitude as those observed for the combination of all the single nucleotide polymorphisms (SNPs) identified for breast cancer so far [36]. For example, the female population in the top 5% of the polygenic risk distribution (i.e. carrying most of the risk alleles for the known breast cancer SNPs) have a risk that is ~2.3 fold higher than the population average [36], whereas according to our estimates the women in bottom 5% of the distribution of epigenome-wide methylation have a risk 1.9 fold higher than the population average.

The genome-wide and high density nature of the methylation measures acquired using the HM450K array draws some parallel with more advanced research examining common genetic variation (including single nucleotide polymorphisms, SNPs) and breast cancer risk.

Indeed, it is widely anticipated that both genetic and epigenetic marks (once adequately characterized) will be critical to improving models for personalized risk estimation and to further develop the emerging area of precision public health. Combining these data may be particularly beneficial as, with considerable potential for interrelationship, common genetic variation captures heritable genetic risk in contrast to epigenetic variation that is likely to capture marks that are associated with differing environmental and lifestyle exposures. Although this is an attractive scenario, if borne out, it does raise new issues related to how, when, and how often to make methylation measurements for use in breast cancer risk estimates for women, given that exposures will differ, change and accumulate over a lifetime. Predicting risk of breast cancer in the next five years would be an important component of a better-targeted screening program.

Further studies are needed to elucidate the mechanisms responsible for our findings but the strength of the observed associations and the prevalence of the exposure, indicate that our measures of epigenome-wide DNA methylation are associated with a large population attributable risk (i.e. 39% by categorizing genome-wide methylation in quartile and using the top quartile as “reference” category); although the concept of population attributable risk implies causation [37], which cannot be inferred from our observational study.

The cost of the Infinium HM450 array limits its use as a screening tool and future studies will need to identify a more cost-effective way of obtaining the epigenome-wide DNA methylation measures that we have shown to be associated with breast cancer risk.

FUNDING

This work was supported by grants from the National Health and Medical Research Council (grant number 1011618); and the Victorian Breast Cancer Research Consortium.

ACKNOWLEDGEMENTS

We would like to express our gratitude to the many thousands of Melbourne residents who continue to participate in the Melbourne Collaborative Cohort Study, the original investigators, program managers and the diligent team who recruited the participants and who continue working on follow-up. The methylation measures were conducted by the Australian Genome Research Facility (AGRF) in Melbourne.

DISCLOSURES

None.

REFERENCES

1. Das, P.M. and R. Singal, (2004).DNA methylation and cancer. *J Clin Oncol*, **22**(22): p. 4632-42.
2. Kulis, M. and M. Esteller, (2010).DNA methylation and cancer. *Adv Genet*, **70**: p. 27-56.
3. Phillips, T., (2008).The role of methylation in gene expression. *Nature Education*, **1**(1).
4. Wilson, A.S., B.E. Power and P.L. Molloy, (2007).DNA hypomethylation and human diseases. *Biochim Biophys Acta*, **1775**(1): p. 138-62.

5. Hinshelwood, R.A. and S.J. Clark, (2008).Breast cancer epigenetics: normal human mammary epithelial cells as a model system. *J Mol Med (Berl)*, **86**(12): p. 1315-28.
6. Jovanovic, J., J.A. Ronneberg, J. Tost and V. Kristensen, (2010).The epigenetics of breast cancer. *Mol Oncol*, **4**(3): p. 242-54.
7. Kristiansen, S., L.M. Jorgensen, P. Guldberg and G. Soletormos, (2013).Aberrantly methylated DNA as a biomarker in breast cancer. *Int J Biol Markers*, **28**(2): p. 141-50.
8. Szyf, M., (2012).DNA methylation signatures for breast cancer classification and prognosis. *Genome Med*, **4**(3): p. 26.
9. Fang, F., S. Turcan, A. Rimmer, A. Kaufman, D. Giri, L.G. Morris, et al., (2011).Breast cancer methylomes establish an epigenomic foundation for metastasis. *Sci Transl Med*, **3**(75): p. 75ra25.
10. Zhuang, J., A. Jones, S.H. Lee, E. Ng, H. Fiegl, M. Zikan, et al., (2012).The dynamics and prognostic potential of DNA methylation changes at stem cell gene loci in women's cancer. *PLoS Genet*, **8**(2): p. e1002517.
11. Xu, Z., S.C. Bolick, L.A. DeRoo, C.R. Weinberg, D.P. Sandler and J.A. Taylor, (2013).Epigenome-wide association study of breast cancer using prospectively collected sister study samples. *J Natl Cancer Inst*, **105**(10): p. 694-700.
12. Shenker, N.S., S. Polidoro, K. van Veldhoven, C. Sacerdote, F. Ricceri, M.A. Birrell, et al., (2012).Epigenome-wide association study in the European Prospective Investigation into Cancer and Nutrition (EPIC-Turin) identifies novel genetic loci associated with smoking. *Human molecular genetics*: p. dds488.
13. Anjum, S., E.-O. Fourkala, M. Zikan, A. Wong, A. Gentry-Maharaj, A. Jones, et al., (2014).A BRCA1-mutation associated DNA methylation signature in blood cells predicts sporadic breast cancer incidence and survival. *Genome Med*, **6**(6): p. 47

14. Fackler, M.J., K. Malone, Z. Zhang, E. Schilling, E. Garrett-Mayer, T. Swift-Scanlan, et al., (2006).Quantitative multiplex methylation-specific PCR analysis doubles detection of tumor cells in breast ductal fluid. *Clin Cancer Res*, **12**(11 Pt 1): p. 3306-10.
15. Hoque, M.O., Q. Feng, P. Toure, A. Dem, C.W. Critchlow, S.E. Hawes, et al., (2006).Detection of aberrant methylation of four genes in plasma DNA for the detection of breast cancer. *J Clin Oncol*, **24**(26): p. 4262-9.
16. Horvath, S., Y. Zhang, P. Langfelder, R.S. Kahn, M.P. Boks, K. van Eijk, et al., (2012).Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biol*, **13**(10): p. R97.
17. Lim, U., A. Flood, S.W. Choi, D. Albanes, A.J. Cross, A. Schatzkin, et al., (2008).Genomic methylation of leukocyte DNA in relation to colorectal adenoma among asymptomatic women. *Gastroenterology*, **134**(1): p. 47-55.
18. Pufulete, M., R. Al-Ghnaniem, A.J. Leather, P. Appleby, S. Gout, C. Terry, et al., (2003).Folate status, genomic DNA hypomethylation, and risk of colorectal adenoma and cancer: a case control study. *Gastroenterology*, **124**(5): p. 1240-8.
19. Hsiung, D.T., C.J. Marsit, E.A. Houseman, K. Eddy, C.S. Furniss, M.D. McClean, et al., (2007).Global DNA methylation level in whole blood as a biomarker in head and neck squamous cell carcinoma. *Cancer Epidemiol Biomarkers Prev*, **16**(1): p. 108-14.
20. Moore, L.E., R.M. Pfeiffer, C. Poscablo, F.X. Real, M. Kogevinas, D. Silverman, et al., (2008).Genomic DNA hypomethylation as a biomarker for bladder cancer susceptibility in the Spanish Bladder Cancer Study: a case-control study. *Lancet Oncol*, **9**(4): p. 359-66.

21. Brennan, K., M. Garcia-Closas, N. Orr, O. Fletcher, M. Jones, A. Ashworth, et al., (2012). Intragenic ATM methylation in peripheral blood DNA as a biomarker of breast cancer risk. *Cancer Res*, **72**(9): p. 2304-13.
22. Choi, J.Y., S.R. James, P.A. Link, S.E. McCann, C.C. Hong, W. Davis, et al., (2009). Association between global DNA hypomethylation in leukocytes and risk of breast cancer. *Carcinogenesis*, **30**(11): p. 1889-97.
23. Delgado-Cruzata, L., H.C. Wu, M. Perrin, Y. Liao, M.A. Kappil, J.S. Ferris, et al., (2012). Global DNA methylation levels in white blood cell DNA from sisters discordant for breast cancer from the New York site of the Breast Cancer Family Registry. *Epigenetics*, **7**(8): p. 868-74.
24. Xu, X., M.D. Gammon, H. Hernandez-Vargas, Z. Herceg, J.G. Wetmur, S.L. Teitelbaum, et al., (2012). DNA methylation in peripheral blood measured by LUMA is associated with breast cancer in a population-based study. *FASEB J*, **26**(6): p. 2657-66.
25. Sandoval, J., H. Heyn, S. Moran, J. Serra-Musach, M.A. Pujana, M. Bibikova, et al., (2011). Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics*, **6**(6): p. 692-702.
26. Clark, C., P. Palta, C.J. Joyce, C. Scott, E. Grundberg, P. Deloukas, et al., (2012). A comparison of the whole genome approach of MeDIP-seq to the targeted approach of the Infinium HumanMethylation450 BeadChip((R)) for methylome profiling. *PLoS One*, **7**(11): p. e50233.
27. Houseman, E.A., W.P. Accomando, D.C. Koestler, B.C. Christensen, C.J. Marsit, H.H. Nelson, et al., (2012). DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*, **13**: p. 86.

28. Chen, Y.A., M. Lemire, S. Choufani, D.T. Butcher, D. Grafodatskaya, B.W. Zanke, et al., (2013).Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*, **8**(2): p. 203-9.
29. Laird, P.W., (2003).The power and the promise of DNA methylation markers. *Nat Rev Cancer*, **3**(4): p. 253-66.
30. Woo, H.D. and J. Kim, (2012).Global DNA hypomethylation in peripheral blood leukocytes as a biomarker for cancer risk: a meta-analysis. *PLoS One*, **7**(4): p. e34615.
31. Ostertag, E.M. and H.H. Kazazian, Jr., (2001).Biology of mammalian L1 retrotransposons. *Annu Rev Genet*, **35**: p. 501-38.
32. Hoffmann, M.J. and W.A. Schulz, (2005).Causes and consequences of DNA hypomethylation in human cancer. *Biochem Cell Biol*, **83**(3): p. 296-321.
33. Horard, B., A. Eymery, G. Fourel, N. Vassetzky, J. Puechberty, G. Roizes, et al., (2009).Global analysis of DNA methylation and transcription of human repetitive sequences. *Epigenetics*, **4**(5): p. 339-50.
34. Wu, H.C., Q. Wang, H.I. Yang, W.Y. Tsai, C.J. Chen and R.M. Santella, (2012).Global DNA methylation levels in white blood cells as a biomarker for hepatocellular carcinoma risk: a nested case-control study. *Carcinogenesis*, **33**(7): p. 1340-5.
35. Riboli, E. and R. Kaaks, (1997).The EPIC Project: rationale and study design. European Prospective Investigation into Cancer and Nutrition. *International Journal of Epidemiology*, **26**(suppl 1): p. S6.
36. Michailidou, K., P. Hall, A. Gonzalez-Neira, M. Ghoussaini, J. Dennis, R.L. Milne, et al., (2013).Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet*, **45**(4): p. 353-61, 361e1-2.

37. Severi, G., G.B. Byrnes and J.L. Hopper, (2008). Five genetic variants associated with prostate cancer. *N Engl J Med*, **358**(25): p. 2739-40; author reply 2741.

FIGURE LEGENDS

Fig. 1 Epigenome-wide methylation for cases and controls by time between blood collection and diagnosis (top) and density plots for samples collected less than 5 years before diagnosis (bottom left) and 5 years or more before diagnosis (bottom right)

Fig. 2 Difference between cases and controls in global methylation within (red) and outside (blue) regulatory promoter regions by time from blood collection. Continuous (dotted) lines correspond to the estimates (95% CI) from a linear mixed effect model allowing the difference outside the regulatory promoter regions to vary linearly with time from blood collection

Table 1 Associations between genome-wide methylation and breast cancer risk

		N _{Co}	N _{CA}	OR (95% CI)	<i>p</i> value
By quartiles#	Q1 [<0.5120]	105	122		
	Q2 [0.5120-0.5187]	105	108	0.73 (0.43-1.22)	0.23
	Q3 [0.5188-0.5253]	105	102	0.56 (0.30-1.02)	0.058
	Q4 [>0.5253]	105	88	0.42 (0.20-0.90)	0.026
	per 1 SD			0.69 (0.50-0.95)	0.02
Time from blood collection*	<5 years	131	131	0.49 (0.28-0.86)	0.01
	5-9 years	148	148	0.59 (0.33-1.06)	0.08
	10-14 years	141	141	1.08 (0.64-1.84)	0.77
				p het=0.052	
ER status	Positive	297	297	0.69 (0.47-1.01)	0.06
	Negative	98	98	0.60 (0.33-1.08)	0.09
				p het=0.69	
PR status	Positive	217	217	0.68 (0.44-1.03)	0.07
	Negative	173	173	0.64 (0.39-1.05)	0.08
				p het=0.88	
Her2 status	Positive	109	109	0.28 (0.13-0.64)	0.002
	Negative	281	281	0.83 (0.57-1.20)	0.32
				p het=0.01	
Grade	1	87	87	0.82 (0.43-1.57)	0.55
	2	179	179	0.45 (0.27-0.76)	0.003
	3	120	120	0.92 (0.53-1.59)	0.75
				p het=0.14	
Size	<2cm	286	286	0.73 (0.50-1.05)	0.09
	≥2cm	123	123	0.51 (0.26-0.99)	0.05
				p het=0.35	
Nodal Status	Negative	298	298	0.74 (0.50-1.10)	0.14
	Positive	86	86	0.62 (0.32-1.16)	0.14
				p het=0.61	

OR, odds ratios from the conditional logistic regression adjusted for age and type of biospecimen

Quartile cut points: <0.5120, 0.5120-0.5187, 0.5188-0.5253, >0.5253.

* difference between reference age and age at blood collection, including counts by specimen type in parentheses (dbs/lymphocyte/buffy coat).

Table 2 Associations between epigenome-wide methylation and breast cancer risk by CpG genomic feature

		# CpG loci	*Pearson's correlation	OR (95% CI) ¹	OR (95% CI) ²
CpG Island	Island or Shore	150070	0.48	0.9 (0.65-1.24)	1.06 (0.76-1.49)
	Shelf or None	219894		0.64 (0.5-0.82)	0.63 (0.48-0.82)
					$p^3=0.03$
Regulatory Feature Category	Promoter	97837	0.28	1.59 (1.07-2.36)	1.82 (1.20-2.75)
	Other	383436		0.69 (0.52-0.93)	0.63 (0.46-0.85)
					$p^3=0.0002$
Within Promoter	Island or Shore	89970	0.85	1.54 (1.04-2.29)	1.30 (0.81-2.09)
	Shelf or None	7867		1.4 (1.04-1.89)	1.26 (0.88-1.79)
					$p^3=0.92$
TSS1500 and TSS200	Island or Shore	108884	0.4	1.09 (0.79-1.51)	1.25 (0.89-1.74)
	Shelf or None	30427		0.65 (0.49-0.85)	0.62 (0.47-0.82)
					$p^3=0.0041$
Repetitive elements	Yes	99063	0.89	0.64 (0.48-0.85)	0.60 (0.41-0.86)
	No	382240		0.77 (0.56-1.06)	1.13 (0.74-1.73)
					$p^3=0.08$

OR, odds ratio from the conditional logistic regression adjusted for age and type of biospecimen

¹ For each CpG genomic feature, the two variables were fitted separately

² For each CpG genomic feature, the two variables were adjusted for each other

³ Test for heterogeneity

*Correlation between methylation values for the subcategories (Island/shores versus shelf/none).

Table 3 Associations between epigenome-wide methylation and breast cancer risk by time from blood collection to diagnosis by CpG genomic feature

		<5 years OR (95% CI)	5-9 years OR (95% CI)	10+ years OR (95% CI)	<i>Test for linear trend</i> <i>p</i>
CpG Island					
	Island or Shore	1.02 (0.50-2.05)	1.24 (0.70-2.18)	0.98 (0.58-1.65)	0.51
	Shelf or None	0.51 (0.32-0.82)	0.56 (0.35-0.88)	0.88 (0.56-1.36)	0.10
Regulatory Feature Category					
	Promoter	1.46 (0.70-3.03)	2.24 (1.07-4.69)	1.73 (0.87-3.48)	0.52
	Other	0.49 (0.29-0.82)	0.61 (0.35-1.05)	0.83 (0.49-1.38)	0.10
Within Promoter					
	Island or Shore	0.71 (0.32-1.61)	1.56 (0.68-3.58)	2.24 (0.94-5.33)	0.32
	Shelf or None	1.81 (0.97-3.37)	1.23 (0.65-2.30)	0.81 (0.43-1.53)	0.75
TSS1500 and TSS200					
	Island or Shore	1.42 (0.77-2.64)	1.27 (0.73-2.22)	1.18 (0.66-2.11)	0.75
	Shelf or None	0.47 (0.29-0.77)	0.51 (0.30-0.85)	0.97 (0.62-1.53)	0.05

OR, odds ratio from the conditional logistic regression adjusted for age and type of biospecimen