Human Molecular Genetics, 2009, Vol. 18, No. 7 ***1332–1342***
doi:10.1093/hmg/ddp033
Advance Access published on January 19, 2009

Gene-body hypermethylation of ATM in peripheral blood DNA of bilateral breast cancer patients

James M. Flanagan[1,*], Marta Munoz-Alegre[1], Stephen Henderson[1], Thomas Tang[2], Ping Sun[3], Nichola Johnson[4], Olivia Fletcher[4], Isabel dos Santos Silva[5], Julian Peto[5,6], Chris Boshoff[1], Steven Narod[3] and Arturas Petronis[2]

[1]UCL Cancer Institute, London, UK, [2]Centre for Addiction and Mental Health, Toronto, Ontario, Canada, [3]Centre for Research in Women's Health, University of Toronto, Toronto, Canada, [4]Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, London, UK, [5]Cancer Research UK Epidemiology and Genetics Group, London School of Hygiene and Tropical Medicine, London, UK and [6]The Institute of Cancer Research, Sutton, Surrey, UK

Received December 22, 2008; Revised December 22, 2008; Accepted January 15, 2009

Bilaterality of breast cancer is an indicator of constitutional cancer susceptibility; however, the molecular causes underlying this predisposition in the majority of cases is not known. We hypothesize that epigenetic misregulation of cancer-related genes could partially account for this predisposition. We have performed methylation microarray analysis of peripheral blood DNA from 14 women with bilateral breast cancer compared with 14 unaffected matched controls throughout 17 candidate breast cancer susceptibility genes including *BRCA1*, *BRCA2*, *CHEK2, ATM, ESR1*, *SFN*, *CDKN2A*, *TP53*, *GSTP1*, *CDH1*, *CDH13*, *HIC1*, *PGR*, *SFRP1, MLH1, RARB* and *HSD17B4*. We show that the majority of methylation variability is associated with intragenic repetitive elements. Detailed validation of the tiled region around *ATM* was performed by bisulphite modification and pyrosequencing of the same samples and in a second set of peripheral blood DNA from 190 bilateral breast cancer patients compared with 190 controls. We show significant hypermethylation of one intragenic repetitive element in breast cancer cases compared with controls ($P = 0.0017$), with the highest quartile of methylation associated with a 3-fold increased risk of breast cancer (OR 3.20, 95% CI 1.78–5.86, $P = 0.000083$). Increased methylation of this locus is associated with lower steady-state *ATM* mRNA level and correlates with age of cancer patients but not controls, suggesting a combined age–phenotype-related association. This research demonstrates the potential for gene-body epigenetic misregulation of *ATM* and other cancer-related genes in peripheral blood DNA that may be useful as a novel marker to estimate breast cancer risk.

Accession numbers: The microarray data and associated .BED and .WIG files can be accessed through Gene Expression Omnibus accession number: GSE14603.

INTRODUCTION

Breast cancer is one of the most common cancers in the Western world affecting one in 10 women during their lifetime (1). Mutations in the breast cancer susceptibility genes, *BRCA1* and *BRCA2*, and, to a lesser extent, *CHEK2*, *TP53* and *ATM*, account for ~25% of familial breast tumours (2). The remaining 75% of familial breast cancers (BRCAx tumours) and the majority of sporadic tumours are not attributable to known mutations in any of these genes. In bilateral cases of breast cancer (a second primary tumour in the contra-lateral breast), it is believed that the underlying 'mutation' is not limited to the epithelial cells of one breast,

*To whom correspondence should be addressed at: CR-UK Viral Oncology Group, UCL Cancer Institute, Paul O'Gorman Building, 74 Huntley Street, London WC1E 6BT, UK. Tel: +44 2076796749; Fax: +44 2076796851; Email: j.flanagan@ucl.ac.uk

but rather is systemic (3). The importance of epigenetic changes in cancer development is now well established; however, the role of epigenetic changes as a mechanism for increased cancer risk is yet to be fully explored. Therefore, we hypothesized that some of the systemic epigenetic changes are either inherited or acquired in the early developmental stages and therefore should be detectable in the tissues other than breast, such as peripheral blood DNA (4,5). Since carcinogenesis induces numerous genetic and epigenetic changes, the study of tumour cells alone cannot distinguish whether epigenetic inactivation of tumour suppressor genes is a cause or a consequence of the neoplastic process in breast cancer (6). Identification of epimutations in the tissues and cells that are not affected by the disease process would favour causal association. Therefore, we aimed to identify epigenetic misregulation of candidate genes in the normal peripheral blood cells of cancer patients compared with controls.

Epigenetics is the investigation of 'any information that is carried by the genome that is not coded by DNA' and includes DNA methylation, histone modifications, chromatin structure and non-coding RNA-mediated regulation of gene expression and various other genomic functions (7). DNA methylation studies in cancer revealed two main types of changes: hypermethylation of promoter CpG islands and hypomethylation of repetitive DNA sequences (8). DNA methylation is heritable through mitosis and is copied to the new strand by DNA methylatransferase 1 during DNA replication. Meiotic heritability of DNA methylation states, particularly transposon-associated methylation, has been shown in mice and in plants; however, it is not yet clear whether DNA methylation is heritable through meiosis in humans (9,10). Recent evidence suggests that epimutations are likely to arise somatically (11). Apart from trans-generational inheritance of epigenetic states, single-generation germline epigenetic effects are also possible. During germ cell development, the epigenome is reprogrammed with two waves of demethylation and re-establishment of DNA methylation marks allowing the possibility of errors that could persist in the germline (12). We have previously shown that the male germline exhibits locus-, cell- and age-dependent DNA methylation differences and that DNA methylation variation is significant across unrelated individuals, at a level that, by far, exceeds DNA sequence variation (13). This has allowed us to hypothesize that epigenetic variation in normal somatic cells, which need not be transgenerational, could be a predisposing factor for cancer.

DNA methylation alterations have been studied extensively in breast tumour tissues most recently with genome wide analyses revealing hypermethylated as well as hypomethylated loci compared with matched adjacent tissues (14). Approximately 100 candidate genes have been reported throughout the literature as promoters hypermethylated at varying frequencies in breast cancers (Pubmeth http://matrix.ugent.be/pubmeth/search.html). For our study, we have selected genes that have either previously been identified as breast cancer susceptibility genes with known mutations in familial cases (*BRCA1*, *BRCA2, ATM, CHEK2, TP53*) or genes that are frequently hypermethylated in sporadic breast cancers (*ESR1*, *SFN*, *CDKN2A*, *GSTP1*, *CDH1*, *CDH13*, *HIC1*, *PGR*, *SFRP1, MLH1, RARB* and *HSD17B4*). We have used a two-stage design to firstly identify methylation variable positions (MVPs) in peripheral blood DNA and then secondly to test selected sites for an association with breast cancer in a larger sample set.

RESULTS

Microarray analysis

We performed differential DNA methylation analysis using a methylation-sensitive enzyme-based approach to compare the methylation status of peripheral blood DNA of 14 bilateral breast cancer cases with 14 matched controls. We have designed a custom-tiled microarray covering a total of 4 Mb to completely cover 17 breast cancer susceptibility genes (Supplementary Material, Fig. S1A) and an additional 34 genes also captured in the flanking regions (Supplementary Material, Table S1). Using the Model-based Analysis of 2-Color Arrays (MA2C) algorithm, we were able to detect 181 significantly variable regions ($P < 0.001$) across the 28 individuals, which we have termed inter-individual MVPs. These included 143 intragenic, or gene-body, MVPs and 38 intergenic regions not associated with known gene promoters (Supplementary Material, Table S2). We observed intragenic methylation variability in each of the remaining targeted genes *ATM*, *PGR*, *CDH1*, *CDH13*, *CHEK2*, *MLH1*, *RARB*, *HSD17B4*, *ESR1*, *SFRP1* and *CDKN2A*. Of all of the genes tested, we detected no significant intragenic methylation variability in six of the targeted genes including *BRCA1*, *BRCA2*, *TP53*, *SFN*, *HIC1* and *GSTP1* (Supplementary Material, Fig. S2).

Methylation variability is located in intragenic repetitive elements

Although the repetitive sequences themselves are not tiled, the enrichment of unmethylated genome can assess methylation status of the repetitive sequences by probing the unique sequences flanking the repeats (gaps in the tiling microarrays) (Supplementary Material, Fig. S1B). To confirm the association of MVPs with repetitive elements, we have calculated the distance from the middle of each peak to the nearest repetitive element for each of the 181 identified loci. This analysis identified 80/181 (44%) of MVPs within 100 bp of the nearest repetitive element and 60% within 200 bp. This distribution is significantly closer to repetitive elements than can be expected by random chance ($P = 7.39e{-}07$, Kolmogorov–Smirnov test) (Fig. 1A). The distribution of repeat element type is not significantly different to the distribution throughout the genome. By aligning each gene at the transcription start sites (TSSs) and averaging the MATScores, we were able to generate a general picture of methylation variability over the first 10 kb of genes longer than 10 kb or across the first 5 kb of shorter genes. This reveals very low methylation variability over all gene promoters up to the TSS and a sharp increase in variability over the first 1 kb of intragenic sequence followed by periodic increase and decrease in variability every 2 kb (Fig. 1B).

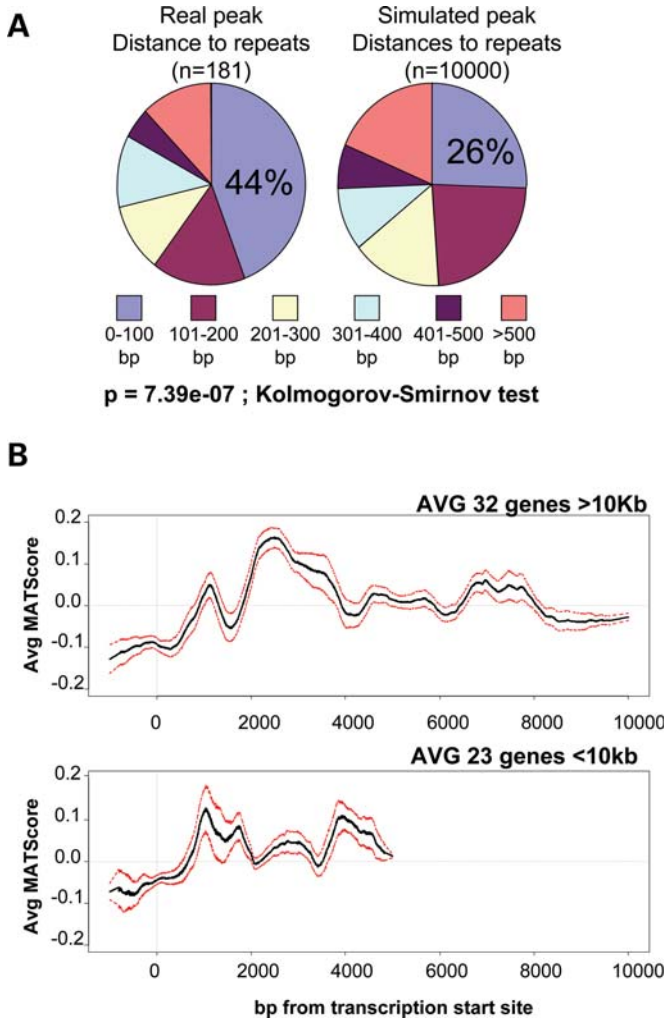

Figure 1. Methylation microarray analysis of 17 genes reveals methylation variability in repetitive elements. (**A**) The distance from each of the MVP to the nearest repetitive element was calculated for each of the 181 distinct MVPs revealing 44% of peaks within 100 bp of the nearest repetitive element. Random placement of 181 simulated MVP peaks reveals a distribution containing only 26% of peaks within 100 bp of the nearest repetitive element. Kolmogorov–Smirnov test was used to determine the significance of the difference in distribution. (**B**) The smoothed average of MATScores (± 2 SD in red) for genes aligned at the TSS reveals periodic increase in methylation variability across the first 10 kb (of genes >10 kb) or across the first 5 kb (of shorter genes, median 4.5 kb).

Investigation of DNA methylation of ATM

We have performed a detailed validation of the MVPs within the targeted region surrounding one gene, *ATM*, which we have selected because it is known to contain germline mutations in familial breast cancer cases with a clear pathogenic role and it contains several intragenic MVPs associated with repetitive elements (Fig. 2A). We used bisulphite modification coupled with pyrosequencing assays to map methylated cytosines in the repetitive elements nearest to six MVPs in the targeted region surrounding *ATM*. We first analysed the 14 matched case–control paired samples that we performed the microarray analysis on. Pair-wise comparison of methylation percentages confirmed the significant inter-individual variability in each of the loci identified by the microarray as exemplified by ATMmvp2a with 8/14 pairs significantly different ($P < 0.05$, Wilcoxon signed rank sum test) (Fig. 2B). Combined data from all of the 14 cases and controls revealed no statistically significant differences overall between patients and controls; however, the highest range of methylation across individuals were observed in the intragenic regions nearest the TSSs. The two most variable loci, ATMmvp1 and ATMmvp2, are both 4 kb downstream of the start sites of the genes *NPAT* and *ATM*, respectively (Fig. 2C). We detected no significant difference in genome-wide methylation as assessed by the pyrosequencing assay for LINE1; therefore, any differences detected are unlikely to be due to overall genome-wide methylation differences. Furthermore, in each of the 14 cases and controls, we detected no methylation of the *ATM* promoter CpG island in the peripheral blood DNA.

We performed a second stage of validation for the most variable loci, MVP2, within the *ATM* gene as well as the *ATM* CpG island and LINE1 assays in peripheral blood DNA from 190 bilateral breast cancer patients compared with 190 controls (Fig. 3). Again, we detected no significant difference between cases and controls in the LINE1 assay or the *ATM* CpG island. The LINE1 assay did, however, show inter-individual variability with genome-wide methylation values ranging from the lowest individual at 48% to the highest at 74% (median 56%). We did detect a significant increase in methylation of ATMmvp2b in the bilateral breast cancer patients (range 72.8–98.4%, mean 91.4%) compared with control individuals (range 53–98%, mean 89.8%) ($P = 0.001686$, Wilcoxon rank sum test) (Supplementary Material, Fig. S3). We have used an inter-quartile analysis of the ATM mvp2b methylation data to reveal a significant increase in methylation in cases in the highest quartile ($P = 0.0011$ using χ^2 test), which is associated with a 3-fold increased risk of breast cancer (odds ratio—OR 3.20, 95% CI 1.78–5.86, $P = 0.000083$), compared with the lowest quartile (Table 1). We have analysed these data to identify whether the methylation status of this locus is associated with any phenotypic characteristics. We have investigated the age at blood draw (median 6.6 years after second diagnosis), age at first and second diagnoses and the time between diagnoses (Table 2). There was no association between the methylation level and tumour morphology, age at first full-term pregnancy, age at menarche, menopausal status, body mass index, weight, parity and the number of affected first-degree relatives (data not shown). Only the age and the age of diagnoses in the breast cancer patients were significantly associated with the methylation status (adjusted for multiple testing, $P = 0.01035$). Interestingly, even though the controls were age-matched with patients, age was not correlated with the methylation status in the controls ($r = -0.1077$, $P = 0.07$), only in the cases at blood draw ($r = 0.242$, $P = 0.0004$), at first diagnosis ($r = 0.189$, $P = 0.0045$) and at second diagnosis ($r = 0.251$, $P = 0.0002$) (Supplementary Material, Fig. S4).

It is well defined that different tissues show differential methylation particularly between different peripheral blood cell types (CD4+ and CD8+ T cells compared with B cells) (15). Therefore, the small effect (increase in methylation at ATM mvp2b) in patients compared with controls could be

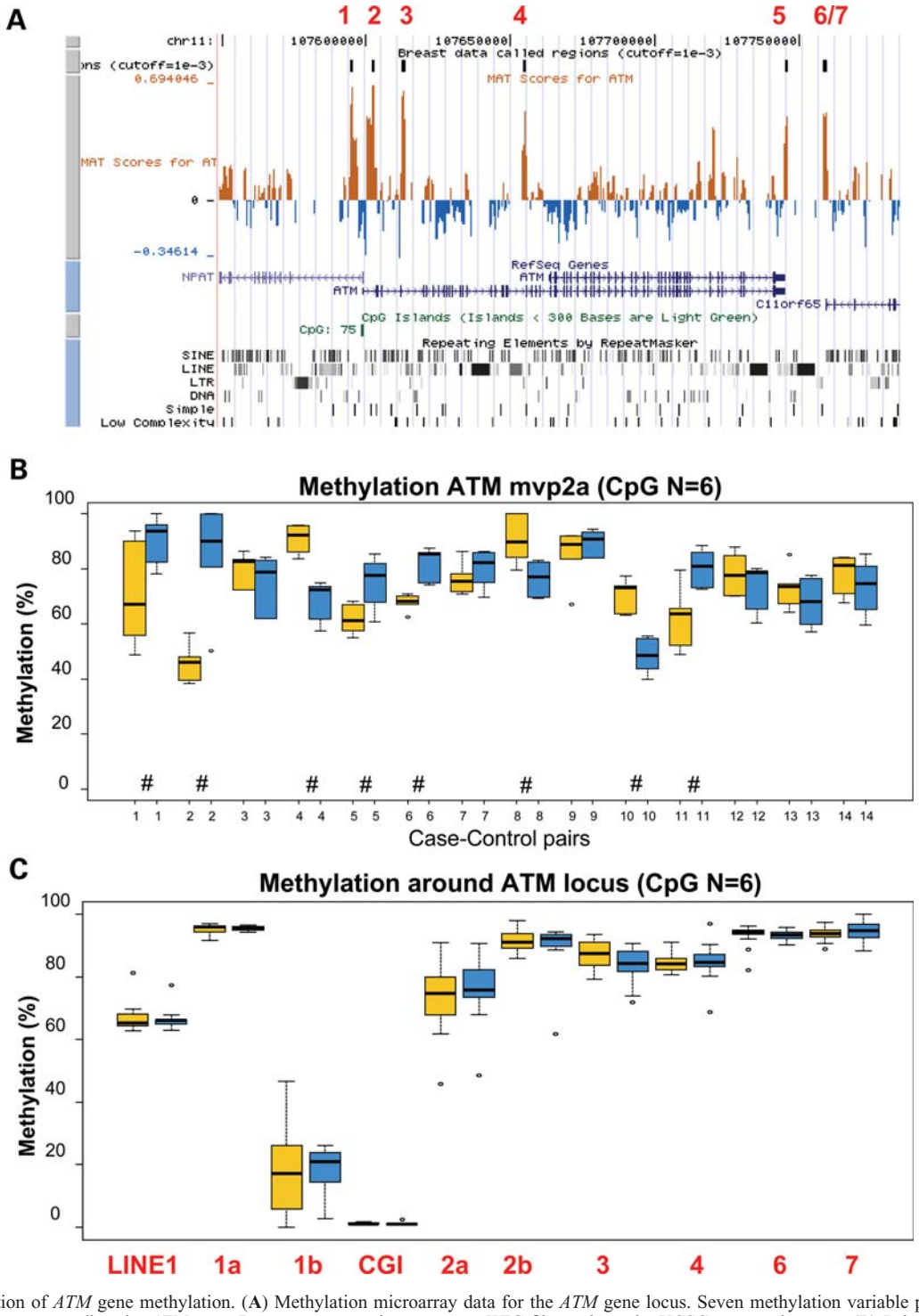

Figure 2. Investigation of *ATM* gene methylation. (**A**) Methylation microarray data for the *ATM* gene locus. Seven methylation variable peaks were detected across the tiled region surrounding the *ATM* gene. Data are presented as a custom WIG file track on the UCSC genome browser. (**B**) Pair-wise comparison of ATM mvp2a methylation reveals increased methylation in 5/14 bilateral breast cancer patients (blue) compared with matched controls (yellow). Methylation of six CpG dinucleotides within the repetitive element mvp2a was determined by pyrosequencing. Pair-wise Wilcoxon signed rank sum test was used to determine statistical significance (# indicates $P < 0.05$). (**C**) DNA methylation analysis of MVPs reveals significant variability detected within the *ATM* gene. LINE 1 assay shows no significant differences in methylation of LINE1 repetitive elements in peripheral blood DNA of cases (blue) compared with controls (yellow). Methylation variability was detected in mvp1b, mvp2a, mvp2b, mvp3 and mvp4. Box and whisker plots represent median (centre line), inter-quartile range (box) and 95th percentiles (whisker), and samples outwith this range are represented as points.

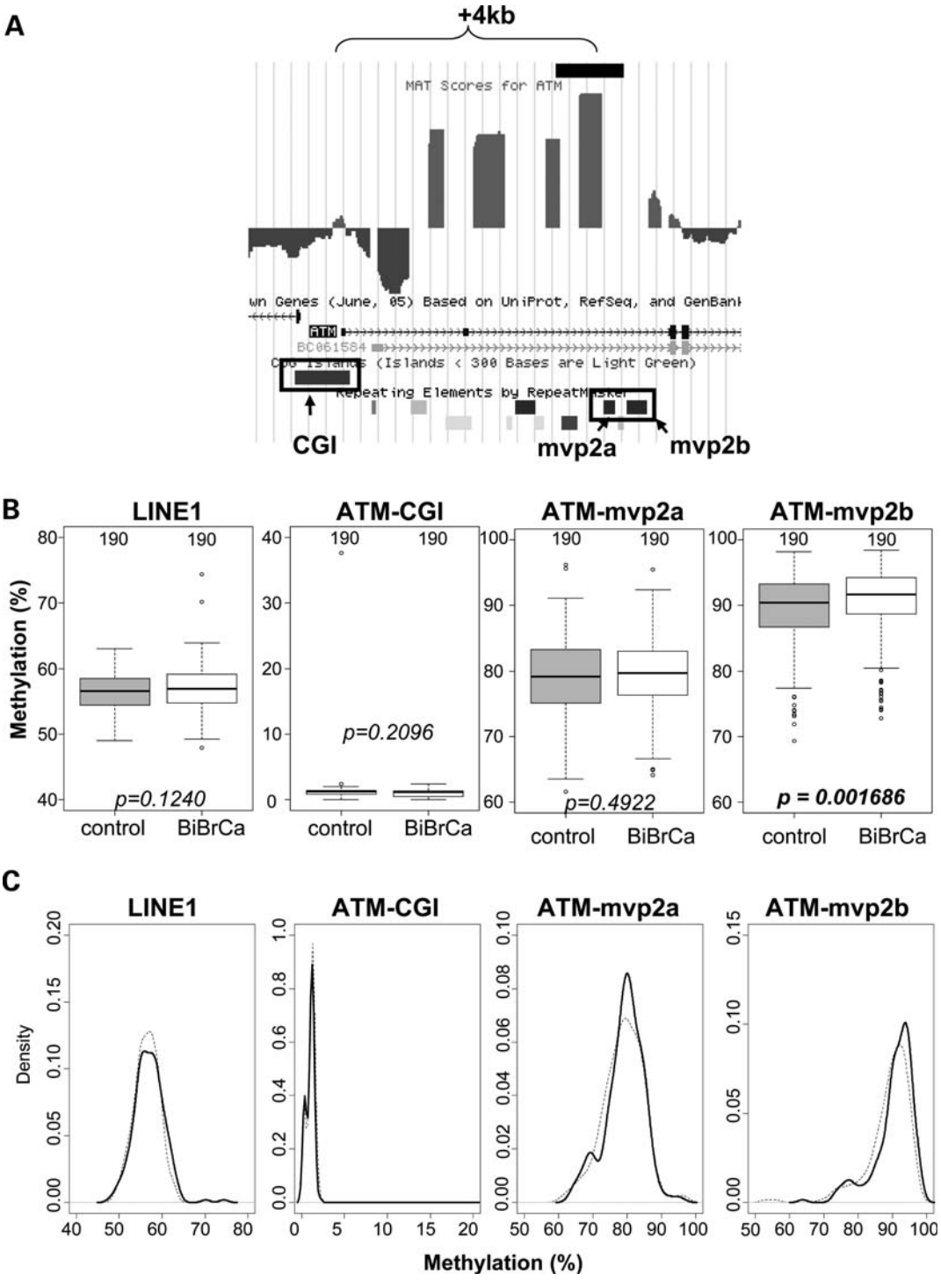

Figure 3. Methylation of LINE1, ATM CpG island, mvp2a and mvp2b in 190 bilateral breast cancer cases compared with 190 controls. (**A**) Schematic of the genomic location of the ATM CpG island (CGI) at the TSS and at mvp2a and mvp2b within the second intron of ATM. (**B**) Pyrosequencing-based methylation analysis of 190 bilateral breast cancer cases compared with 190 matched controls reveals no methylation in the promoter CpG island and significant inter-individual methylation variability in the intronic mvp2a and mvp2b. Significant hypermethylation of mvp2b is detected in bilateral breast cancer cases compared with controls ($P = 0.0017$, Wilcoxon signed rank sum test). Box and whisker plots represent median (centre line), inter-quartile range (box) and 95th percentiles (whisker), and samples outwith this range are represented as points. (**C**) Kernel density plot of methylation values in cases (solid line) compared with controls (dotted lines) showing overlapping distributions for LINE1, ATM CpG island and ATM mvp2a and a skewed distribution of methylation at ATM mvp2b in the cases.

Table 1. Inter-quartile analysis of ATM mvp2b methylation reveals increased risk of breast cancer in the highest quartile

ATM mvp2b	Control ($n = 189$)	Case ($n = 190$)	χ^2	OR	95% CI	*P*-value	*P*-trend
Q4 91.5–98.4%	32	63		3.20	1.78–5.86	0.000083	0.00032
Q3 88.6–91.5%	48	44		1.44	0.81–2.58	0.215	
Q2 88.6–91.5%	50	47		1.60	0.90–2.87	0.108	
Q1 53.0–88.6%	59	36	0.0011	1			

Table 2. Association between methylation of ATM mvp2b and phenotypic data from the bilateral breast cancer patients and controls reveals an association between methylation and age in cases but not in controls

	Cases, median (range)	Cases, *P*-value[a]	Cases, *P*-adjusted[b]	Controls, median (range)	Controls, *P*-value[a]	Controls, *P*-adjusted[b]
Age at blood draw	62.8 (37.5–79.6)	0.00109**	0.01035**	62.8 (37.1–79.4)	0.106	0.4028
Age at first diagnosis	47 (26–64)	0.0145**	0.0918	NA	NA	NA
Age at second diagnosis	56 (26–70)	0.000487**	0.0093**	NA	NA	NA
Time between diagnoses	6 years (0–26)	0.177	0.4204	NA	NA	NA

NA, not applicable.
[a]Linear model regression.
[b]FDR corrected for multiple testing.
**Signifies significant association ($P < 0.05$).

due to a larger effect in a small subpopulation of cells within the heterogeneous peripheral blood cell population. We have addressed this question by isolation of B cells, T cells and monocytes from peripheral blood mononuclear cells (PBMCs) from two healthy controls and investigating the methylation status of this locus. We show that the methylation of ATMmvp2b does not show cell-specific methylation differences between these three blood cell fractions, and all fractions have similar methylation levels to the whole PBMCs (Supplementary Material, Fig. S5).

In order to investigate the relationship between methylation at ATM mvp2b and expression of *ATM*, we screened a large panel of cancer cell lines and observed a similar methylation range as seen in peripheral blood (77–99%); however, the distribution of methylation percentages is significantly higher in the cancer cell lines than in the peripheral blood DNA of the patients and controls ($P < 1e{-}14$, Wilcoxon signed rank test) (Fig. 4A). We performed quantitative RT–PCR for *ATM* expression in five breast cancer cell lines (MCF7, T47D, SKBR3, MDA-MB-231 and BT549) and analysed pre-existing gene expression data for 18 mesenchymal tumour cell lines (16) (Fig. 4B and C). In both cases, we observed a correlation between the methylation of the intragenic repetitive element and expression of *ATM* (Spearmans rho -0.9 and -0.33, respectively).

DISCUSSION

This project is the first comprehensive DNA methylation study using high-density tiling arrays to look for methylation variation in white blood cells of bilateral breast cancer patients compared with controls. It is well established that identification of high-risk patients guiding the use of preventative treatment such as intensive surveillance (e.g. mammography), chemoprevention (e.g. tamoxifen) or prophylactic surgery (e.g. radical mastectomy) can significantly reduce the burden of breast cancer (17). Therefore, all additional tools with which we can predict a patient's risk of cancer, such as DNA methylation markers, would improve our ability to identify those at high risk. The difficulty in finding genetic markers of common breast cancer risk has been recently exemplified by a number of extraordinarily large case–control studies that have identified only minor increases in breast cancer risk (18,19). The overall goal of this current study was to screen candidate genes for epigenetic differences in peripheral blood DNA to identify markers of common breast cancer risk. We believe that DNA methylation markers could be more informative, as they are more stable than other candidate biomarkers such as RNA or protein-based markers, and can act as a surrogate for environmental exposures (20). The extent to which epigenetic markers can be used for risk assessment is yet to be fully explored.

'Gene body methylation' is a recent term describing the DNA methylation that occurs throughout the gene from the TSS, through exons and introns, to the 3′ untranslated region (21–23). In *Arabidopsis*, the level of gene-body methylation has been linked to the level of gene transcription such that gene-body methylation was absent in genes that were not transcribed, was low in genes that were transcribed at low levels, higher levels of methylation in the highest transcribed genes but even higher methylation levels in genes that had intermediate levels of transcription (22). Gene-body methylation is also more associated with genes on the active X chromosome rather than on the inactive X-chromosome (21). The association with the transcription of individual genes and gene-body methylation has not yet been investigated in cancer. We have recently suggested that gene-body methylation in the human genome may be associated with intragenic repetitive elements and that altered methylation in the gene-

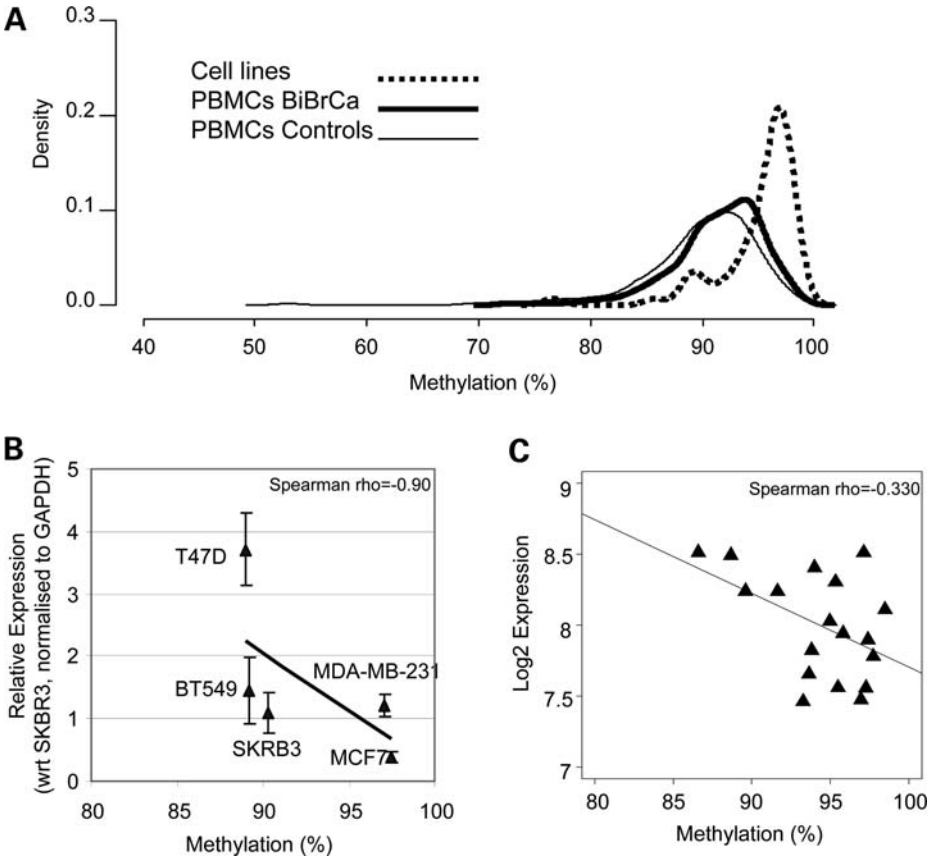

Figure 4. Correlation between methylation of ATM mvp2b and expression of ATM in cancer cell lines. (**A**) Methylation analysis in 62 cancer cell lines reveals a similar range of methylation as in PBMCs of breast cancer patients (77–99%); however, the distribution is skewed towards increased methylation in the cancer cell lines (green). Methylation distributions of breast cancer patients (red) and controls (black) are shown for comparison. (**B**) ATM expression was determined by qRT–PCR in the five breast cancer cell lines and by gene expression microarray data (**C**) for a panel of sarcoma cell lines. Error bars represent SEM from triplicate qRT–PCR experiments. Correlation between methylation and expression is shown using Spearmans rank correlation coefficient.

body may be a mechanism of modulating the level of transcription (24). In the present study, we show for the first time that the methylation of one intragenic repetitive element inversely correlates with the expression of the gene which supports this hypothesis. Although there are numerous examples in the literature showing that active genes have more gene-body methylation levels than inactive genes, all of these studies compare the gene body methylation levels of one gene with another and compare the expression levels (22). These studies make a distinction between 'On' expression and gene-body methylation compared with 'Off' expression and no gene-body methylation. We suggest from our data that the level of gene-body methylation may be modulating or fine-tuning the level of expression of the active genes. To date, there has been no report looking at the subtle differences in gene-body methylation across samples and associating that with the expression of the same gene in those samples. Therefore, our study is the first report of an association between different levels of gene-body methylation and expression of the same gene.

How gene-body methylation may alter the gene expression, or alternatively how expression level may affect the methylation level, is still not clear. The mvp within *ATM* that we have identified could be a reporter of a more extended regulatory epigenetic profile at *ATM*, similar to a tag SNP for a region of linkage disequilibrium in DNA sequence variation. Alternatively, this region could be involved in the regulation of unannotated non-coding transcript(s) or antisense transcripts that regulate the sense gene (22). Recent reports have shown a high level of antisense transcription throughout the genome which could be involved in gene regulation, which suggests new levels of understanding of the process of transcriptional regulation (25,26).

Our study of the breast cancer susceptibility genes has provided evidence that the regions of inter-individual methylation variability are located within repetitive elements, particularly SINEs (Alu sequences), and that the variability is highest at $\sim$1 kb downstream of the TSSs and increases periodically along the gene. We interpret this increased variability as the point in the genome where the methylation along the chromatin is changing from unmethylated at the promoters to fully methylated in the gene-body. We have identified one repetitive element, ATM mvp2b, which is significantly more methylated in the peripheral blood DNA of women affected with bilateral breast cancer compared with matched control individuals. On its own, this is not sufficient to be considered a potential diagnostic test, as there is considerable overlap between cases and controls and the receiver operating curve area under the curve

of 0.59 supports this lack of specificity as a diagnostic. However, this may improve in combination with similar markers from the other genes that we have identified, if such markers provide an additive effect. Interestingly, not all genes showed signs of inter-individual methylation variability. For some of the genes, this could be explained by the short length (e.g. *SFN*, *GSTP1* and *HIC1*); however, this is not the case for other longer genes which could represent genes with conserved methylation patterns (e.g. *BRCA1*, *BRCA2* and *TP53*) compared with other long genes that show high levels of methylation variability (e.g. *CDH13*).

The ataxia telangiectasia (A-T) mutated (*ATM*) gene is considered a breast cancer susceptibility gene, as female heterozygotes from A-T families have an increased risk of breast cancer and a number of breast cancer families have been identified with germline *ATM* mutations (27,28). Although one small study by Vo *et al*. identified increased *ATM* promoter methylation associated with decreased expression in locally advanced sporadic breast cancers, numerous other reports suggest no evidence of promoter methylation of this gene in breast cancers (29–32). One finding that is consistent is that *ATM* expression is often reduced in breast tumours (31–34). This could be explained by genetic haploinsufficiency (35); however, our data in peripheral blood DNA and cancer cell lines suggest a novel mechanism of decreased expression mediated by aberrant gene-body methylation, which warrants further investigation in tumour tissues. We propose that if normal tissues of the individual, including mammary epithelial cells, express lower levels of *ATM* owing to aberrantly increased gene-body methylation, then these individuals may be more susceptible to DNA mutations in those cells.

The biggest risk factor for breast cancer and indeed most cancers is increasing age. The incidence of breast cancer in women doubles for every 10 years until menopause with a relative risk of >10-fold (36). An association between DNA methylation and increasing age has also been reported, for example increased DNA methylation variability in older twins compared with younger twins and increasing variability with increasing age in familial clusters (37,38). However, we have found that the association between the level of methylation at ATM mvp2b and increasing age is only associated in the bilateral breast cancer patients and not in controls. This suggests that the relationship between age and DNA methylation may be more dependent on the phenotype of the individuals, in this case cancer, and may be due to other predisposing environmental exposures that accumulate with age.

One of the confounding factors which we cannot rule out in this study is the long-term effects of previous therapy in the patients that are not present in the controls. Future studies will need to investigate peripheral blood samples taken before the initiation of treatment preferably in prospectively collected cohort studies to rule out this possibility. However, if therapy accounted for an overall difference in genome-wide DNA methylation, the LINE1 assay for genome-wide methylation and the other unique sites tested (ATM mvp2a) would have detected this, which it did not. Another confounding factor is the effect of tissue-specific methylation. Although we have shown that in two healthy controls there is no evidence for cell-type-specific methylation of this locus, we cannot rule out this possibility in cancer patients. Therefore, this needs to be addressed in cancer patients.

In summary, we have shown that inter-individual variability in gene-body methylation is associated with repetitive elements and that it is possible to identify regions of significant differences between bilateral breast cancer patients and controls that may be useful as additional surrogate markers for breast cancer risk. This research also demonstrates the correlation between the level of gene-body methylation and mRNA level of a single gene and highlights the potential for gene-body epigenetic regulation of gene transcription. Finally, this work also supports the need to interrogate whole epigenomes, including repetitive elements, using tiling arrays or unbiased next-generation sequencing approaches to investigate the full complement of methylation variability across the genome.

MATERIALS AND METHODS

Clinical sample

For the microarray analysis, peripheral blood DNA samples from 14 women with bilateral breast cancer were compared with the DNA samples from 14 age- and ethnicity-matched controls. The cases were selected by being (i) *BRCA1* and *BRCA2* mutation negative, confirmed clinically; (ii) metachronous tumours (separate sites) (iii) and those meet the bilateral case criteria set by Chaudary *et al*. (39), namely (a) subsequent tumour contains *in situ* component; (b) distinct histological types; (c) subsequent tumour has greater degree of differentiation; (d) no evidence of metastatic spread from ipsilateral tumour. These samples were collected with ethics approval from the Institute for Womens Health, Toronto, Canada. Peripheral blood DNA samples from a second set of 190 bilateral cases and 190 controls were obtained from the British Breast Cancer study (40,41). Bilateral breast cancer patients were ascertained through the English and Scottish cancer registries, controls were friends or non-blood relatives of breast cancer patients. Controls were age-matched to patients at the age at blood draw (within ± 2 months) and all patients and controls were Caucasians resident in the UK. Written informed consent was obtained from all participants, and the study was approved by the South East Multicentre Research Ethics Committee.

DNA methylation profiling using microarrays

A custom oligonucleotide tiling array was designed to target the testable genomic regions (plus 100 kb flanking sequence) of genes, including breast cancer susceptibility genes *BRCA1* (82 kb), *BRCA2* (86 kb), *CHEK2* (57 kb) and *ATM* (143 kb) along with other genes that are known to be commonly methylated in breast cancer, including *ESR1* (296 kb), *SFN* (1.3 kb), *CDKN2A* (27 kb), *TP53* (19 kb), *GSTP1* (2.8 kb), *CDH1* (96 kb), *CDH13* (1169 kb), *HIC1* (4.0 kb), *PGR* (92 kb), *SFRP1* (47 kb), *MLH1* (57 kb), *RARB* (97 kb), *HSD17B4* (90 kb). Arrays were designed by and performed by Nimblegen Inc. Specified regions were from UCSC HG17 with overlapping 50mer probes every 21 bp excluding repeat masked regions, and the probe set was replicated four times on the chip and averaged data were used for analysis. In addition to the 17 targeted genes, we also captured 26

genes and 12 partial genes which were peripherally targeted in the flanking regions by these tiling arrays (Supplementary Material, Table S1).

The general principle of these DNA methylation arrays consists of hybridization of the unmethylated fraction of genomic DNA to the microarray containing oligonucleotides that represent the genomic region of interest. We performed the enrichment as described previously (42). Briefly, we used a cocktail of three methylation-sensitive enzymes (*Hpa*II, *Aci*I and *Hin*6I) to digest individuals genomic DNA and used ligation-mediated PCR to amplify products which were cleaned using Qiagen PCR cleanup kit (Qiagen, Canada), labelled with either Cy3 or Cy5 dyes and co-hybridized in matched pairs to the custom array. Hybridization intensity correlates with the DNA methylation status at the genomic locus homologous to each oligonucleotide on the array. MA2C was used for normalization, peak detection (using a cut-off of $P < 0.001$) and generation of MATScores for each probe (43).

Pyrosequencing

DNA samples were bisulphite-converted using EZ-96 DNA Methylation-Gold kit (Zymo Research, Orange, CA, USA) following the manufacturers protocol. Pyrosequencing of LINE1 elements was performed using the LINE1 assay (Biotage, Hertford, UK). All other pyrosequencing assays were designed using the PyroQ assay design software (primers in Supplementary Material, Table S3). All pyrosequencing assays for repetitive elements included at least one primer in a unique sequence outside of the repeat and followed by a unique nested primer pair for some amplicons to ensure specific amplification. A common tag was placed on either the forward or the reverse primer (depending on the strand to be sequenced), and a common universal biotinylated primer was used for all reactions as described previously (44). PCR was performed as described previously and cycling conditions included denaturation at 95°C for 4 min, followed by 10 cycles of 94°C for 15 s, touchdown from 60–50°C (−1 degree/cycle) for 15 s and 72°C for 20 s, followed by a further 30 cycles at 50°C annealing temperature. The second PCR used 2 μl of a 1:10 dilution of the first PCR as template and the same cycling conditions (13). All products were confirmed to be single bands by agarose gel electrophoresis. Methylation values were calculated as an average of all CpG sites within each assay as determined by the Pyro Q-CpG Software (Biotage, Uppsala, Sweden).

Blood cell fractionation

Peripheral blood was collected from two healthy female donors with informed consent. PBMCs were isolated using a standard Ficoll-paque protocol, and blood cell fractions were isolated using antibody-coated magnetic microbeads and MACS MS separation columns (cat: 130-042-201) following the manufacturers protocol (Miltenyi Biotech, Surrey, UK). Briefly, 5–10 million PBMCs were incubated with each of the antibody-coated beads for B-cells (CD19, cat: 130-050-301), T-cells (CD2, cat: 130-091-114) or monocytes (CD14, cat: 130-050-201). Genomic DNA was extracted from whole PBMCs and each cell-specific population. Bisulphite conversion and pyrosequencing were performed as described.

Cancer cell lines

DNA was extracted from 62 cell lines from brain ($n = 7$), breast ($n = 5$), colon ($n = 5$), leukaemia ($n = 5$), lymphoma ($n = 9$), sarcoma ($n = 15$) and others ($n = 16$). A full list of cell line names is included in Supplementary Material, Table S4. Bisulphite conversion and pyrosequencing were performed as described.

Quantitative RT–PCR

Quantitative RT–PCR was performed for *ATM* using an Eppendorf Mastercycler with the following primer pairs: ATM_F (5′-tgctcatacagcaggccata-3′) with ATM_R (5′-aa ggctgaatgaaagggtaattc-3′) and GAPDH_F (5′-ggagtcaacgga tttggtcgta-3′) with GAPDH_R (5′-ggcaacaatatccactttaccaga gt-3′). The reaction mix contained 1× SYBR green master mix (Applied Biosystems, Foster City, CA, USA) and 0.5 μM of each forward and reverse primers in a volume of 30 μl. PCR cycling consisted of 95°C for 10 min, then 40 cycles of 95°C for 30 s, 60°C for 60 s, followed by a meltcurve analysis. Fold-change in expression was calculated by ΔΔCt normalized to *GAPDH* for each sample and normalizing ea ch of the cell lines to the cell line SKBR3.

Statistical analysis

The majority of statistical analysis in this study was performed using various R packages as described earlier. The simulated peak distributions were determined by randomly placing single peaks across the same tiled region and calculating the distances to the real repetitive elements and repeating 10 000 times. Kolmogorov–Smirnov test was used to test the significance of the difference in distributions (ks.test). We used Wilcoxon signed rank test for the analysis of methylation percentages obtained by pyrosequencing (wilcox.test). We performed further analysis for ATM mvp2a to confirm the significance of this difference using 1000 randomized simulations of the test statistic (W, from the Wilcoxon rank sum test) to calculate the real *P*-value of the observed test statistic. We used either generalized linear model (glm) regression or logistic regression analysis (lm) for further analysis of ATM MVP2b to assess the associations with phenotypic data. Interquartile analysis was performed by calculating the quartiles from all of the data (quantile) and counting the number of cases or controls in each quartile. We then used the χ^2 test (chisq.test), odds ratio (oddsratio) and Cochran–Armitage trend test (independence_test).

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Jon Mill for a critical review of the manuscript and Dr Matthew Trotter for additional bioinformatic assistance.

Conflict of Interest statement. None declared.

FUNDING

This work was funded in part by a pilot study grant from the Canadian Cancer Etiology Research Network (CCERN) to A.P., S.N. and J.M.F. and a grant from the University College London Hospital Charities to J.M.F. J.M.F. was funded in part by Cellcentric Ltd and by Cancer Research UK. The British Breast Cancer study is funded by Cancer Research UK and Breakthrough Breast Cancer Research Centre. We acknowledge NHS funding to the NIHR Biomedical Research Centre. Funding to pay the Open Access charge was provided by a grant from the University College London Hospital Charities.

REFERENCES

1. Jemal, A., Murray, T., Ward, E., Samuels, A., Tiwari, R.C., Ghafoor, A., Feuer, E.J. and Thun, M.J. (2005) Cancer statistics, 2005. *CA Cancer J. Clin.*, **55**, 10–30.
2. Smith, P., McGuffog, L., Easton, D.F., Mann, G.J., Pupo, G.M., Newman, B., Chenevix-Trench, G., Szabo, C., Southey, M., Renard, H. *et al.* (2006) A genome wide linkage search for breast cancer susceptibility genes. *Genes, Chromosomes Cancer*, **45**, 646–655.
3. Weitzel, J.N., Robson, M., Pasini, B., Manoukian, S., Stoppa-Lyonnet, D., Lynch, H.T., McLennan, J., Foulkes, W.D., Wagner, T., Tung, N. *et al.* (2005) A comparison of bilateral breast cancers in BRCA carriers. *Cancer Epidemiol. Biomarkers Prev.*, **14**, 1534–1538.
4. Cui, H., Cruz-Correa, M., Giardiello, F.M., Hutcheon, D.F., Kafonek, D.R., Brandenburg, S., Wu, Y., He, X., Powe, N.R. and Feinberg, A.P. (2003) Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. *Science*, **299**, 1753–1755.
5. Suter, C.M., Martin, D.I. and Ward, R.L. (2004) Germline epimutation of MLH1 in individuals with multiple cancers. *Nat. Genet.*, **36**, 497–501.
6. Baylin, S. and Bestor, T.H. (2002) Altered methylation patterns in cancer cell genomes: cause or consequence? *Cancer Cell*, **1**, 299–305.
7. Kouzarides, T. (2007) Chromatin modifications and their function. *Cell*, **128**, 693–705.
8. Feinberg, A.P., Ohlsson, R. and Henikoff, S. (2006) The epigenetic progenitor origin of human cancer. *Nat. Rev. Genet.*, **7**, 21–33.
9. Richards, E.J. (2006) Inherited epigenetic variation—revisiting soft inheritance. *Nat. Rev. Genet.*, **7**, 395–401.
10. Morgan, D.K. and Whitelaw, E. (2008) The case for transgenerational epigenetic inheritance in humans. *Mamm. Genome*, **19**, 394–397.
11. Ito, Y., Koessler, T., Ibrahim, A.E., Rai, S., Vowler, S.L., Abu-Amero, S., Silva, A.L., Maia, A.T., Huddleston, J.E., Uribe-Lewis, S. *et al.* (2008) Somatically acquired hypomethylation of IGF2 in breast and colorectal cancer. *Hum. Mol. Genet.*, **17**, 2633–2643.
12. Reik, W., Dean, W. and Walter, J. (2001) Epigenetic reprogramming in mammalian development. *Science*, **293**, 1089–1093.
13. Flanagan, J.M., Popendikyte, V., Pozdniakovaite, N., Sobolev, M., Assadzadeh, A., Schumacher, A., Zangeneh, M., Lau, L., Virtanen, C., Wang, S.C. *et al.* (2006) Intra- and interindividual epigenetic variation in human germ cells. *Am. J. Hum. Genet.*, **79**, 67–84.
14. Ordway, J.M., Budiman, M.A., Korshunova, Y., Maloney, R.K., Bedell, J.A., Citek, R.W., Bacher, B., Peterson, S., Rohlfing, T., Hall, J. *et al.* (2007) Identification of novel high-frequency DNA methylation changes in breast cancer. *PLoS ONE*, **2**, e1314.
15. Rakyan, V.K., Down, T.A., Thorne, N.P., Flicek, P., Kulesha, E., Graf, S., Tomazou, E.M., Backdahl, L., Johnson, N., Herberth, M. *et al.* (2008) An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res.*, **18**, 1518–1529.
16. Henderson, S.R., Guiliano, D., Presneau, N., McLean, S., Frow, R., Vujovic, S., Anderson, J., Sebire, N., Whelan, J., Athanasou, N. *et al.* (2005) A molecular map of mesenchymal tumors. *Genome Biol.*, **6**, R76.
17. Sakorafas, G.H. (2003) The management of women at high risk for the development of breast cancer: risk estimation and preventative strategies. *Cancer Treat. Rev.*, **29**, 79–89.
18. Easton, D.F., Pooley, K.A., Dunning, A.M., Pharoah, P.D., Thompson, D., Ballinger, D.G., Struewing, J.P., Morrison, J., Field, H., Luben, R. *et al.* (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, **447**, 1087–1093.
19. Stacey, S.N., Manolescu, A., Sulem, P., Rafnar, T., Gudmundsson, J., Gudjonsson, S.A., Masson, G., Jakobsdottir, M., Thorlacius, S., Helgason, A. *et al.* (2007) Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat. Genet.*, **39**, 865–869.
20. Li, S., Hursting, S.D., Davis, B.J., McLachlan, J.A. and Barrett, J.C. (2003) Environmental exposure, DNA methylation, and gene regulation: lessons from diethylstilbesterol-induced cancers. *Ann. NY Acad. Sci.*, **983**, 161–169.
21. Hellman, A. and Chess, A. (2007) Gene body-specific methylation on the active X chromosome. *Science*, **315**, 1141–1143.
22. Zilberman, D., Gehring, M., Tran, R.K., Ballinger, T. and Henikoff, S. (2007) Genome-wide analysis of *Arabidopsis thaliana* DNA methylation uncovers an interdependence between methylation and transcription. *Nat. Genet.*, **39**, 61–69.
23. Suzuki, M.M. and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.*, **9**, 465–476.
24. Flanagan, J.M. and Wild, L. (2007) An epigenetic role for noncoding RNAs and intragenic DNA methylation. *Genome Biol.*, **8**, 307.
25. Core, L.J., Waterfall, J.J. and Lis, J.T. (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, **322**, 1845–1848.
26. He, Y., Vogelstein, B., Velculescu, V.E., Papadopoulos, N. and Kinzler, K.W. (2008) The antisense transcriptomes of human cells. *Science*, **322**, 1855–1857.
27. Chenevix-Trench, G., Spurdle, A.B., Gatei, M., Kelly, H., Marsh, A., Chen, X., Donn, K., Cummings, M., Nyholt, D., Jenkins, M.A. *et al.* (2002) Dominant negative ATM mutations in breast cancer families. *J. Natl. Cancer Inst.*, **94**, 205–215.
28. Swift, M., Morrell, D., Massey, R.B. and Chase, C.L. (1991) Incidence of cancer in 161 families affected by ataxia-telangiectasia. *N. Engl. J. Med.*, **325**, 1831–1836.
29. Allinen, M., Peri, L., Kujala, S., Lahti-Domenici, J., Outila, K., Karppinen, S.M., Launonen, V. and Winqvist, R. (2002) Analysis of 11q21–24 loss of heterozygosity candidate target genes in breast cancer: indications of TSLC1 promoter hypermethylation. *Genes, Chromosomes Cancer*, **34**, 384–389.
30. Kontorovich, T., Cohen, Y., Nir, U. and Friedman, E. (2008) Promoter methylation patterns of ATM, ATR, BRCA1, BRCA2 and P53 as putative cancer risk modifiers in Jewish BRCA1/BRCA2 mutation carriers. *Breast Cancer Res. Treat*. July 19 [Epub ahead of print].
31. Treilleux, I., Chapot, B., Goddard, S., Pisani, P., Angele, S. and Hall, J. (2007) The molecular causes of low ATM protein expression in breast carcinoma; promoter methylation and levels of the catalytic subunit of DNA-dependent protein kinase. *Histopathology*, **51**, 63–69.
32. Vo, Q.N., Kim, W.J., Cvitanovic, L., Boudreau, D.A., Ginzinger, D.G. and Brown, K.D. (2004) The ATM gene is a target for epigenetic silencing in locally advanced breast cancer. *Oncogene*, **23**, 9432–9437.
33. Tommiska, J., Bartkova, J., Heinonen, M., Hautala, L., Kilpivaara, O., Eerola, H., Aittomaki, K., Hofstetter, B., Lukas, J., von Smitten, K. *et al.* (2008) The DNA damage signalling kinase ATM is aberrantly reduced or lost in BRCA1/BRCA2-deficient and ER/PR/ERBB2-triple-negative breast cancer. *Oncogene*, **27**, 2501–2506.
34. Ye, C., Cai, Q., Dai, Q., Shu, X.O., Shin, A., Gao, Y.T. and Zheng, W. (2007) Expression patterns of the ATM gene in mammary tissues and their associations with breast cancer survival. *Cancer*, **109**, 1729–1735.
35. Lu, S., Shen, K., Wang, Y., Santner, S.J., Chen, J., Brooks, S.C. and Wang, Y.A. (2006) Atm-haploinsufficiency enhances susceptibility to carcinogen-induced mammary tumors. *Carcinogenesis*, **27**, 848–855.

36. McPherson, K., Steel, C.M. and Dixon, J.M. (2000) ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. *Br. Med. J. (Clin. Res. Ed.)*, **321**, 624–628.
37. Bjornsson, H.T., Sigurdsson, M.I., Fallin, M.D., Irizarry, R.A., Aspelund, T., Cui, H., Yu, W., Rongione, M.A., Ekstrom, T.J., Harris, T.B. *et al.* (2008) Intra-individual change over time in DNA methylation with familial clustering. *JAMA*, **299**, 2877–2883.
38. Fraga, M.F., Ballestar, E., Paz, M.F., Ropero, S., Setien, F., Ballestar, M.L., Heine-Suner, D., Cigudosa, J.C., Urioste, M., Benitez, J. *et al.* (2005) Epigenetic differences arise during the lifetime of monozygotic twins. *Proc. Natl Acad. Sci. USA*, **102**, 10604–10609.
39. Chaudary, M.A., Gravelle, I.H., Bulstrode, J.C., Wang, D.Y., Bulbrook, R.D., Millis, R.R. and Hayward, J.L. (1983) Breast parenchymal patterns in women with bilateral primary breast cancer. *Br. J. Radiol.*, **56**, 703–706.
40. Fletcher, O., Johnson, N., Gibson, L., Coupland, B., Fraser, A., Leonard, A., dos Santos Silva, I., Ashworth, A., Houlston, R. and Peto, J. (2008) Association of genetic variants at 8q24 with breast cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **17**, 702–705.
41. Johnson, N., Fletcher, O., Naceur-Lombardelli, C., dos Santos Silva, I., Ashworth, A. and Peto, J. (2005) Interaction between CHEK2*1100delC and other low-penetrance breast-cancer susceptibility genes: a familial study. *Lancet*, **366**, 1554–1557.
42. Schumacher, A., Kapranov, P., Kaminsky, Z., Flanagan, J., Assadzadeh, A., Yau, P., Virtanen, C., Winegarden, N., Cheng, J., Gingeras, T. *et al.* (2006) Microarray-based DNA methylation profiling: technology and applications. *Nucleic Acids Res.*, **34**, 528–542.
43. Song, J.S., Johnson, W.E., Zhu, X., Zhang, X., Li, W., Manrai, A.K., Liu, J.S., Chen, R. and Liu, X.S. (2007) Model-based analysis of two-color arrays (MA2C). *Genome Biol.*, **8**, R178.
44. Royo, J.L., Hidalgo, M. and Ruiz, A. (2007) Pyrosequencing protocol using a universal biotinylated primer for mutation detection and SNP genotyping. *Nat. Protoc.*, **2**, 1734–1739.