Heritable DNA methylation marks associated with susceptibility to breast cancer

Jihoon E. Joo1,2, James G. Dowty3, Roger L. Milne3,4, Ee Ming Wong1,2, Pierre-Antoine Dugué3,4, kConFab, Dallas English3,4, John L. Hopper3, David E. Goldgar1,5, Graham G. Giles3,4 & Melissa C. Southey1,2

Mendelian-like inheritance of germline DNA methylation in cancer susceptibility genes has been previously reported. We aimed to scan the genome for heritable methylation marks associated with breast cancer susceptibility by studying 25 Australian multiple-case breast cancer families. Here we report genome-wide DNA methylation measured in 210 peripheral blood DNA samples provided by family members using the Infinium HumanMethylation450. We develop and apply a new statistical method to identify heritable methylation marks based on complex segregation analysis. We estimate carrier probabilities for the 1000 most heritable methylation marks based on family structure, and we use Cox proportional hazards survival analysis to identify 24 methylation marks with corresponding carrier probabilities significantly associated with breast cancer. We replicate an association with breast cancer risk for four of the 24 marks using an independent nested case–control study. Here, we report a novel approach for identifying heritable DNA methylation marks associated with breast cancer risk.
DNA methylation is a breast cancer risk factor. Several genome-wide studies of DNA methylation have found evidence that global methylation levels measured in blood-derived DNA is associated with breast cancer risk for women in the general population, and for women from families at high genetic risk. While increased global methylation is associated with a reduced risk, increased methylation levels within functional promoters have been associated with an increased risk of breast cancer.

Candidate gene approaches have been used to assess whether methylation at CpG islands of breast cancer susceptibility genes is associated with breast cancer risk. Women carrying germline mutations in BRCA1 have a substantially elevated risk of breast cancer and their tumours typically have distinctive histological features. We found that peripheral blood DNA methylation at the BRCA1 promoter was associated with an estimated 3.5-fold (95% CI, 1.4–10.5) increased risk of breast cancer diagnosed before the age of 40 years. Hansmann et al. reported that 1.4% of 600 women from the German Consortium for Hereditary Breast and Ovarian Cancer had constitutive BRCA1 hypermethylation confined to one of the two alleles.

Women carrying specific rare germline mutations in ATM are also at substantially elevated risk of breast cancer. Flanagan et al. performed methylation microarray analyses of peripheral blood DNA across several genes including BRCA1, BRCA2, CHEK2, ATM, TP53, CDH1, and MLH1, and demonstrated that gene body hypermethylation of ATM was associated with an estimated threefold increased risk of breast cancer. Brennan et al. combined two nested case-control studies of women at high risk of breast cancer and found evidence that methylation at an intragenic locus in ATM (ATMmvp2a) was associated with increased risk of breast cancer.

Potapova et al. described promoter region methylation of PALB2 was evident in ~7% of breast and ovarian cancers, including those with germline mutations in BRCA2, using methylation-specific PCR and bisulfite sequencing. In contrast, Mikeska et al. found little evidence of PALB2 methylation in high-grade serous ovarian cancers using a methylation-sensitive high-resolution melting assay.

The terminology being used to describe these observations is variable and vulnerable to misuse and misinterpretation. The term ‘epimutation’ is strictly defined as a heritable change in gene activity that is not associated with a DNA mutation but rather with gain or loss of DNA methylation or other heritable modifications of chromatin. Changes in gene expression through altered DNA methylation or histone modifications induced from cis- or trans-acting genetic factors known as methylation Quantitative Trait Loci, (mQTL) are therefore not epimutations in this strict sense.

Epimutations and mQTLs can mimic germline mutations in their effect on cancer predisposition and it is likely that their contribution has been largely underestimated due to limited research beyond the candidate gene approaches described above. These phenomena could therefore account for some of the familial risk of breast cancer that is not yet identified.

Intergenerational transmission of epimutations (as described by the authors in the initial reports) has been observed in MLH1 and MSH2 in the context of Lynch Syndrome (LS), a hereditary condition in which genetic mutations in key mismatch repair genes predispose individuals to colorectal, endometrial, and other cancers. While two thirds of LS cases carry germline genetic mutations at the DNA mismatch genes, a small proportion of LS has been associated with epimutations. It has since been demonstrated that some methylation marks at MLH1 and MSH2 that are transmitted transgenerationally are in fact linked to nearby cis-acting genetic variants and consequently follow Mendelian inheritance patterns, and are thus not strictly epimutations. Other MLH1 epimutations occur sporadically and have not been linked to underlying genetic variations; while these epimutations are often observed in a familial context, they do not follow complete Mendelian inheritance patterns.

We hypothesised that breast cancers in multiple-case breast cancer families with no known genetic susceptibility mutations are in part due to the contribution of heritable DNA methylation marks (including true epimutations and mQTLs). To test this, we assessed genome-wide DNA methylation for 25 multiple-case breast cancer families using the Infinium HumanMethylation450 K BeadArray. One or more women with breast cancer in these families had been previously screened for, and found not to carry germline mutations in known breast cancer susceptibility genes. In this study, we report a new analytic approach to identify CpG sites with Mendelian-like inheritance patterns and a set of 24 heritable methylation sites associated with breast cancer risk.

Results
DNA methylation within families. After removing 3949 poorly performing CpG probes (detection p-value < 0.05), β-values and M-values were obtained from a total of 481,563 analysable CpG probes across DNA samples from 210 individuals in 25 families (20 families participating in kConFab and 5 families participating in the ABCFR). β-values denote % methylation levels obtained from the HM450K platform, where 0 indicates 0% methylation and 1 indicates 100% methylation. Due to the heteroscedastic nature of β-values, the log2 ratio of methylation intensity, known as M-values, are also calculated and used for all statistical analyses.

DNA samples were collected from 87 breast cancer cases (one third of the cases had blood collected prior to diagnosis) and 123 unaffected controls. In order to examine the overall genome-wide methylation similarities between samples and families, a hierarchical clustering analysis was performed according to M-values across 481,563 probes. No distinct clustering by case-control status was observed but some families shared similar overall methylation patterns (Supplementary Fig. 1).

Heritable methylation sites. The proportion of probes within 10 bp of known single-nucleotide polymorphisms (SNPs) increased significantly with Δl (p < 0.0001, and see Fig. 1). We then removed all probes within 10 bp of known SNPs and those located on sex chromosomes (see Methods). We screened the remaining 365,169 sites for those most consistent with having a Mendelian pattern of inheritance using the statistic Δl (Supplementary Fig. 2A). The 1000 most Mendelian methylation marks (those with the highest values of Δl) are listed in Supplementary Data 1. These marks all have values of Δl above 77, which suggests that they are highly heritable. We estimated carrier probabilities for the 1000 most heritable methylation marks using family structure alone.

Heritable methylation sites associated with breast cancer. Of the 1000 most Mendelian methylation marks, 24 of them had carrier probabilities that were associated with breast cancer at the Bonferroni-adjusted p-value threshold of 5 × 10−5 (all p-values between 2 × 10−3 and 7.4 × 10−10, see Table 1 and Supplementary Fig. 2B). Notably, five of the heritable methylation marks were clustered together at VTRNA2-1. For all 24 marks, the methylation (β) differences were substantial (Δβ > 0.30) between individuals, with most of these marks showing methylation patterns distinctly falling into hypermethylated (β > 0.80), hypomethylated (β < 0.20), or hemimethylated (β ~ 0.50) groups, indicating...
potential allele-specific methylation pattern at these sites (Supplementary Fig. 3 and Supplementary Table 1). While unbiased hazard ratios could not be calculated (see Statistical Methods), the risk of breast cancer increased with carrier probabilities for all 24 sites (Supplementary Table 2, where the low carrier probabilities for some CpGs reflect the very low prior carrier probability), and the estimated effect of the hypothetical genetic variant on the M-values of each site can be seen from Supplementary Fig. 3 and Supplementary Table 2. For example, for cg06536614, ‘carriers’ are hemimethylated and ‘non-carriers’ are hypomethylated. In contrast, for cg18584561, ‘carriers’ are hypomethylated and ‘non-carriers’ have generally higher methylation levels but these are spread over a range of methylation.

The five probes associated with the VTRNA2-1 locus (previously known as mir886) encompass a ~50 bp region 150 bp upstream from the transcription start site and overlapping a CpG island. Although these probes target 5 independent proximally located CpG sites, the 50mer probes largely overlapped with each other (Fig. 2). In addition to these 5 CpG sites, DNA methylation at other proximal CpG probes showed similar patterns, although not meeting statistical significance. Within each individual, the methylation patterns at all CpGs across this VTRNA2-1 promoter region were consistent, suggesting allelic methylation at this locus (Fig. 2b). Two recent studies have suggested that this region might be maternally imprinted25,26. We have tested this in eight trios (father, mother, and child) and included additional siblings when possible by performing clonal bisulfite sequencing. We observed strong hypermethylation of the maternally inherited allele, confirming the maternal imprinting of this locus. We found complete loss of methylation in one child whose three other siblings retained the methylation in the maternal allele (Supplementary Fig. 4).

One of the heritable methylation marks associated with breast cancer risk was located close to the 5’ end of the gene Growth Regulation by Estrogen Breast Cancer 1 (GREB1). The methylation patterns of all samples at this methylation site grouped clearly into hypomethylated, hemimethylated, or hypermethylated. Only 13 of our 210 samples were hypermethylated at this methylation site. We found three of the other methylation marks overlapping promoter regions of DUSP22, TMC3, and PPP2R5C. Nine other heritable methylation marks were located in gene body regions of MMP27, ANO10, CIGN, ZZZF1, PNKD, XYLTI1, C7orf50, RASA3, and IL10RB, while six heritable methylation marks were not known to be associated with any gene (Table 1). The ZZZF1, PNKD, C7orf50, RASA3, and IL10RB probes overlapped CpG island shores or shelves. The MMP27, ANO10, XYLTI1, and GREB1 probes encompassed enhancer regions.

Breast cancer risk association in the general population.
Altogether, 433 invasive breast cancer cases and their matched controls were included in the analysis2. The median follow-up time was 9.5 years, interquartile range (IQR): 5.0 to 13.1 years. Supplementary Fig. 5 shows β-methylation value distribution for MCCS cases and controls for the 24 methylation sites showing heritable methylation patterns and associated with breast cancer in the family-based analyses. Of the 24 sites, four showed linear association with risk of breast cancer in the MCCS at the nominal significance threshold p < 0.05 (Table 2). The significant CpG probes were cg18584561 (GREB1; OR per standard deviation (s. d.): 1.18, 95% CI: 1.03–1.36), cg01741999 (PNKD; OR per 1 s.d.: 1.26, 95% CI: 1.03–1.54), cg03916490 (C7orf50; OR per 1 s.d.: 0.83, 95% CI: 0.72–0.96) and cg27639199 (TMC3; OR per 1 s.d.: 1.19, 95% CI: 1.03–1.36).

When comparing values belonging to the smaller vs. larger ‘peak’ of the methylation variable distribution, the results were consistent and more significant (Table 3). At cg18584561 (GREB1), which was trimodal, both the hypomethylated and hypermethylated peaks were associated with decreased breast cancer risk (OR = 0.60 (95% CI: 0.45–0.80), and OR = 0.56, (95% CI: 0.34–0.95), respectively). The methylation pattern at cg27639199 (TMC3) was also trimodal where the hypermethylated peak was strongly associated with breast cancer risk (OR = 2.16 (95% CI: 1.26–3.72)). At cg03916490 (C7orf50), reduced methylation was associated with the breast cancer risk (OR = 1.61 (95% CI: 1.16–2.24)). An annotated CpG probe (cg18514595) was associated with breast cancer risk when categorised into three methylation peaks.

These associations were robust to further adjustment for Houseman’s white blood cell composition, and to further adjustment for additional breast cancer risk factors (parity, hormonal replacement therapy use, age at menarche and menopausal status). Similar results were also found when restricting the analyses to DNA that was extracted from dried blood spots (Supplementary Table 3) and when repeating the analyses with carrier probabilities in place of M-values (Supplementary Table 4).

Associations between genetic variants and DNA methylation.
Genotyped and imputed variants from the iCOGS (± 1 kb of cg18584561, GREB1) representing 251 MCCS participants was included in the analysis. This region had eight common variants (in linkage disequilibrium) nominally associated with breast cancer risk. We found a very strong linear association between methylation at cg18584561 and the genotypes at this region (p = 1 × 10−65–1 × 10−72). The association between these genetic variants and the corresponding methylation β-value is presented graphically in Supplementary Fig. 6.

Association with breast cancer estrogen receptor status.
We tested whether methylation levels at any of these 24 CpG sites were influenced by ER status in our nested case–control study and found evidence for three methylation marks cg06536614 (ER−; OR = 1.02 (95% CI: 0.86–1.21) vs. ER+; OR = 0.71 (95% CI: 0.53–0.96), p-value (heterogeneity) = 0.03), cg01074083 (ER−; OR
| CpG site   | ΔI     | P-value for association with breast cancer | Chromosome | Position (hg19) | UCSC reference gene |
|------------|--------|------------------------------------------|------------|----------------|---------------------|
| cg06536614 | 143.6285 | 7.23 × 10^{-09}                            | 5          | 135416381       | VTRNA2-1 (MIR886)   |
| cg10306192 | 109.4419 | 3.46 × 10^{-05}                            | 11         | 102576374       | MMP27               |
| cg18110333 | 108.7894 | 4.13 × 10^{-10}                            | 6          | 292329         | DUSP22              |
| cg00124993 | 107.9848 | 1.71 × 10^{-08}                            | 5          | 135416412       | VTRNA2-1 (MIR886)   |
| cg26328633 | 107.4759 | 1.64 × 10^{-08}                            | 5          | 135416394       | VTRNA2-1 (MIR886)   |
| cg25340688 | 105.9031 | 2.73 × 10^{-08}                            | 5          | 135416398       | VTRNA2-1 (MIR886)   |
| cg18514595 | 95.2513  | 1.67 × 10^{-07}                            | 22         | 49579968        | unannotated         |
| cg26896946 | 92.0795  | 1.50 × 10^{-09}                            | 5          | 135416405       | VTRNA2-1 (MIR886)   |
| cg11035303 | 90.9039  | 1.74 × 10^{-10}                            | 3          | 43465503        | ANO10               |
| cg23012654 | 89.7585  | 3.85 × 10^{-05}                            | 14         | 97493395        | unannotated         |
| cg26773954 | 88.7692  | 1.12 × 10^{-06}                            | 13         | 111969980       | unannotated         |
| cg22901919 | 87.5936  | 1.85 × 10^{-06}                            | 4          | 141317067       | ClGN                |
| cg04417708 | 85.0287  | 1.28 × 10^{-08}                            | 17         | 4043867         | ZZEF1               |
| cg18584561 | 85.0000  | 9.30 × 10^{-06}                            | 2          | 11682017        | GREB1               |
| cg1608150  | 82.6156  | 5.21 × 10^{-07}                            | 5          | 135415948       | unannotated         |
| cg01741999 | 81.7762  | 3.28 × 10^{-09}                            | 2          | 219137824       | PNKO                |
| cg01074083 | 80.4167  | 1.58 × 10^{-05}                            | 16         | 17516862        | XYL1                |
| cg02096220 | 80.3509  | 3.64 × 10^{-07}                            | 4          | 129212777       | unannotated         |
| cg03916490 | 79.7094  | 2.07 × 10^{-06}                            | 7          | 1080558         | C7orf50             |
| cg27639199 | 79.5279  | 5.37 × 10^{-06}                            | 15         | 81666528        | TMC3                |
| cg25188166 | 79.4058  | 4.90 × 10^{-08}                            | 3          | 119420208       | unannotated         |
| cg05865327 | 78.9441  | 1.65 × 10^{-06}                            | 14         | 102274741       | PPP2R5C             |
| cg23947138 | 77.3448  | 7.47 × 10^{-10}                            | 13         | 114782778       | RASA3               |
| cg05187003 | 77.2261  | 1.50 × 10^{-08}                            | 21         | 34641507        | IL10RB              |

Table 1 The methylation marks associated with breast cancer

**Fig. 2** DNA methylation at the VTRNA2-1 promoter. a Genomic locations of 6 HM450K probes associated with VTRNA2-1 promoter region. b DNA methylation levels (β-values) of these 6 probes labelled by breast cancer status. β-values for each individual are shown on y axis for 6 VTRNA2-1 probes. c Average DNA methylation levels across all six probes shown separately for individual families and labelled by breast cancer status. β-values are shown on y axis for members from each family (y axis).
Inheritance patterns and tested some of the most heritable in familial breast cancer. We demonstrated that genome-wide methylation patterns were likely to be in fl overlap SNPs (the methylation measurements at these probes are similar within some families, indicating that shared genetics demonstrated that genome-wide methylation patterns were maximisation algorithm and genetic segregation analysis, ‘value’ = p -value. Odds ratio from conditional logistic regression of the risk of breast cancer on M-values (per 1 s.d.), adjusting for body mass index, tobacco smoking, alcohol drinking, time between blood collection and cancer diagnosis, and sample type (dried blood spots, peripheral blood mononuclear cells, and buffy coats). Cases and controls were individually matched on year of birth, year of blood draw, country of birth, and sample type for the vast majority of them (97%).

Table 2 Associations between heritable DNA methylation marks (associated with breast cancer in multiple-case families) and risk of breast cancer in the general population (Melbourne Collaborative Cohort Study)

| Site     | Chr. | Position | Gene name        | ORa | 95% CI   | p   |
|----------|------|----------|------------------|-----|----------|-----|
| cg06536614 | 5    | 135416381 | VTRNA2-1 (MIR886) | 0.95 | 0.83-1.10 | 0.497 |
| cg10306192  | 11   | 102576374 | MMP27            | 1.09 | 0.94-1.27 | 0.235 |
| cg18110333  | 6    | 292329   | DUSP22           | 0.96 | 0.83-1.11 | 0.588 |
| cg00124993  | 5    | 135416412 | VTRNA2-1 (MIR886) | 0.97 | 0.84-1.12 | 0.667 |
| cg26328633  | 5    | 135416394 | VTRNA2-1 (MIR886) | 0.98 | 0.85-1.13 | 0.761 |
| cg25340688  | 5    | 135416398 | VTRNA2-1 (MIR886) | 0.95 | 0.82-1.09 | 0.441 |
| cg18514595  | 22   | 49579968 | unannotated      | 1.14 | 0.99-1.31 | 0.077 |
| cg26896946  | 5    | 135416405 | VTRNA2-1 (MIR886) | 0.94 | 0.82-1.09 | 0.426 |
| cg10353033  | 3    | 129456503 | ANO10            | 1.01 | 0.88-1.16 | 0.894 |
| cg23012654  | 14   | 97493995 | unannotated      | 0.95 | 0.83-1.10 | 0.503 |
| cg26773954  | 11   | 11969980 | unannotated      | 1.10 | 0.88-1.17 | 0.813 |
| cg22909109  | 4    | 141317067 | CLGN             | 0.91 | 0.78-1.06 | 0.224 |
| cg04417708  | 17   | 4043867  | ZEEF1            | 1.00 | 0.87-1.15 | 0.989 |
| cg18584561  | 2    | 11682017 | GREB1            | 1.18b | 1.03-1.36 | 0.015 |
| cg11608150  | 5    | 135415948 | unannotated      | 0.93 | 0.80-1.07 | 0.311 |
| cg01741999  | 2    | 219137824 | PNKD             | 1.26 | 1.03-1.54 | 0.027 |
| cg01074083  | 16   | 17516862 | XYL1T            | 0.98 | 0.84-1.13 | 0.749 |
| cg02963220  | 4    | 129212177 | unannotated      | 1.02 | 0.89-1.18 | 0.743 |
| cg03964490  | 7    | 10805558 | C7orf50          | 0.83 | 0.72-0.96 | 0.012 |
| cg27639919  | 15   | 81666528 | TM3C             | 1.19 | 1.03-1.36 | 0.018 |
| cg25188166  | 3    | 119420208 | unannotated      | 0.96 | 0.83-1.10 | 0.551 |
| cg05865327  | 14   | 102274741 | PPP2R5C          | 1.04 | 0.90-1.20 | 0.589 |
| cg23947138  | 13   | 114782778 | RASA3            | 0.89 | 0.78-1.02 | 0.091 |
| cg05187003  | 21   | 34641507 | IL10RB           | 1.00 | 0.86-1.15 | 0.950 |

Discussion

Genome-wide studies of heritable DNA methylation studies in the context of familial breast cancer have not been conducted previously, although ~50% of familial breast cancer cases cannot be explained by what we currently know about genetic risk. In this study, we tested whether heritable DNA methylation marks are associated with breast cancer risk in multiple-case breast cancer families that do not carry pathogenic mutations in known breast cancer susceptibility genes.

The hierarchical clustering analysis of all detected probes demonstrated that genome-wide methylation patterns were similar within some families, indicating that shared genetics might have an influence on DNA methylation, as shown in previous studies. However, overall genome-wide methylation did not appear to segregate with affected status in any families (Supplementary Fig. 1).

We developed a new statistical methodology, based on an expectation–maximisation algorithm and genetic segregation analysis, to identify heritable DNA methylation marks using the HM450 K platform (see Methods). We validated this analytic approach by showing that it identifies probes that are known to overlap SNPs (the methylation measurements at these probes are likely to be influenced by the underlying SNPs). We then removed all SNP-overlapping probes from the analysis, screened the remaining probes for those with the most Mendelian-like inheritance patterns and tested some of the most heritable methylation marks for association with breast cancer. Note that our screening for probes with Mendelian-like inheritance patterns removed many probes that cannot be associated with familial breast cancer, so this screening step greatly increased our statistical power for detecting probes associated with familial breast cancer.

We found 24 probes associated with breast cancer risk after adjusting for multiple testing (Table 1). Five of these 24 CpG probes were adjacent to the promoter region of a vault RNA, VTRNA2-1 (previously known as nc886 or miiR886; Fig. 2). This vault RNA has been shown to be involved in the inhibition of protein kinase R (PKR) activity and acts as a tumour suppressor in several cancer types. It is located at chromosome 5q13, which is often associated with cancer-associated LOH including basal-like breast cancers. Hypomethylation at this promoter, suggestive of loss of imprinting, occurs systematically in specific individuals in diverse populations, at least partially due to periconceptual environment and is stable for at least 10 years. Silver et al. (2015) also noted that VTRNA2-1 exhibits all the hallmarks of ‘metabolic imprinting’ and is likely to be a determinant of cancer risk. Here we have shown that methylation at the VTRNA2-1 promoter is also associated with heritable breast cancer risk that is measurable in DNA extracted from blood.

All 210 DNAs included in this study had hemi- or hypomethylation across all CpG probes at the VTRNA2-1 locus (Fig. 2) indicating potential allele-specific DNA methylation (ASM). ASM at this locus has been reported previously by studies utilising clonal bisulphite sequencing of multiple tissue types. However, these studies did not explore nearby genetic variation that could be superimposed on imprinting to influence the allelic methylation pattern. Hemimethylation patterns generally associated with genomic imprinting were only observed in 170 of the 210 DNAs (~80%) included in our study (Fig. 2). Genomic

\[ OR = 0.72 (0.53-0.98), p-value = 0.02 \]

\[ OR = 0.80 (0.68-0.94) vs. OR = 1.24 (0.93-1.65), p-value = 0.01 \]
imprinting is usually highly effective and loss-of-imprinting is often associated with growth retardation syndromes or tumour development. In reference to other typically imprinted region (e.g., H19/IGF2), the methylation at VTTNA2-1 seemed exceptionally variable in the families included in our study. Romani et al. also report an atypical imprinting pattern at this locus and concluded this region was a polymorphically imprinted differential methylation region. By performing clonal bisulitase sequencing within families, we confirmed the polymorphic imprinting of this locus as reported by above studies (Supplementary Table 2).

We hypothesised that breast cancer arising in multiple-case breast cancer families with no known genetic mutation might be in part due to the contribution of heritable DNA methylation marks (including epimutations and mQTLs). As discussed above, methylation at the VTTNA2-1 promoter is a strong epimutation candidate but many of the other identified heritable methylation marks are likely to be mQTLs. More work is required to characterise these marks further. It is not likely that common genetic variation currently recognised to be associated with breast cancer risk (already identified via genome-wide-association-studies) underlies these methylation marks. The currently published risk-associated SNP closest to any of the identified heritable methylation marks is located ~1.5 Mb from cg18584561 (GREB1). We found a strong linear association between the DNA methylation pattern at cg18584561 (GREB1) and 8 proximal common genetic variants (Supplementary Fig. 6). The genotypes of all 8 SNPs strongly correlated with the methylation pattern (e.g., DNAs hypermethylated at cg18584561 were homozygous across this region). This suggests a potential mQTL at this locus.

The other 19 CpG probes were all located in different genomic regions. We showed that a single CpG overlapping the transcription start region of the GREB1 gene is associated with heritable breast cancer risk. This gene codes for the protein ‘growth regulation by oestrogen in breast cancer 1’ and has been shown to play a critical role in hormone dependent breast cancer. There is currently no direct evidence of epigenetic regulation of this gene.

Four of the 24 methylation marks were associated with breast cancer risk in an independent nested case-control study of methylation and breast cancer risk (Table 2). This outcome provides information with which one could use to hypothesise further about the relative frequency of the 24 methylation marks. For some marks, such as the one at GREB1, approximately half of the families appear to be methylated which is consistent with replication being possible in a population-based sample, another fraction (~10%) of the population are hypermethylated at this CpG. Interestingly, both the hypomethylated and hypermethylated profiles were associated with a decreased breast cancer risk, with similar estimated risk reduction of 40–45% at cg03916490.

### Table 3 Associations between heritable DNA methylation marks (associated with breast cancer in 608 multiple-case families) and risk of breast cancer in the general population (Melbourne Collaborative Cohort Study), M-values categorised into 2 or 3 groups according to observed bimodal or trimodal 610 distribution (i.e., peaks)

| Site   | Chr. | Position | Gene name     | Smaller peak definition | N cases/controls in peak | OR     | 95% CI  | p       |
|--------|------|----------|---------------|-------------------------|--------------------------|--------|---------|---------|
| cg06536614 | 5    | 135416381| Unannotated   | M < -1.8                | 98/90                    | 1.12   | 0.80-1.57| 0.503   |
| cg10306192 | 11   | 102576374| Unannotated   | M > -2.5                | 186/172                  | 1.13   | 0.84-1.51| 0.420   |
| cg18110333 | 6    | 292329   | Unannotated   | M < -2                  | 97/92                    | 1.09   | 0.77-1.54| 0.626   |
| cg00124993 | 5    | 135416412| VTTNA2-1      | M < -2.8                | 96/85                    | 1.17   | 0.83-1.64| 0.379   |
| cg26328633 | 5    | 135416394| VTTNA2-1      | M < -2                  | 100/91                   | 1.15   | 0.82-1.61| 0.414   |
| cg25340688 | 5    | 135416398| VTTNA2-1      | M < -2                  | 100/83                   | 1.12   | 0.80-1.57| 0.521   |

* Odds ratio from conditional logistic regression of the risk of breast cancer on M-values (per 1 s.d.), adjusting for body mass index, tobacco smoking, alcohol drinking, time between blood collection and cancer diagnosis; and sample type for the vast majority of them (97%). Bold text indicates statistically significant associations (Table 2).
because all germline genetic variants are heritable by design. The set of marks to test for association with breast cancer risk, but the marks for heritability is an effective way of greatly reducing the incorporated into previous studies. Second, screening methylation studies. First, our approach utilises DNA methylation levels, several years before breast cancer diagnosis.

Our study has two advantages over previous genome-wide studies. First, our approach utilises DNA methylation levels, which are important intermediate biomarkers that have not been incorporated into previous studies. Second, screening methylation marks for heritability is an effective way of greatly reducing the set of marks to test for association with breast cancer risk, but because all germline genetic variants are heritable by definition, this screening step could not be applied to previous studies.

Fitting the Mendelian model (C7orf50), about 30% of the nested case–control participants were not strongly hypermethylated, which was associated with a 60% increase in risk. It is possible that the marks that did not validate in the nest case–control sample were either not present or at a very low frequency. Fig 3 graphically illustrates our analytical approach using 2 CpGs with different ‘carrier’ probabilities as examples (cg06536614 and cg18584561).

Two thirds of the bloods collected from affected members of the multiple-case breast cancer families were collected after breast cancer diagnosis. Reverse causation is therefore a potential reason for non-replication of some of the methylation marks in the nested case–control study where blood samples were collected several years before breast cancer diagnosis.

Our study has two advantages over previous genome-wide studies. First, our approach utilises DNA methylation levels, which are important intermediate biomarkers that have not been incorporated into previous studies. Second, screening methylation marks for heritability is an effective way of greatly reducing the set of marks to test for association with breast cancer risk, but because all germline genetic variants are heritable by definition, this screening step could not be applied to previous studies. Heritable methylation sites are interesting, regardless of whether or not they are associated with breast cancer susceptibility. We have devised a method for identifying heritable methylation sites and we have used this as a screening step to increase our power for detecting heritable methylation marks that are associated with breast cancer. This work could found a new area of exploration in the context of disease susceptibility. Specifically for breast cancer, this work provides new opportunities for increasing the precision of current risk prediction models, new strategies for cancer control (including screening) and new opportunities for the development of (or repurposing of) epigenetic therapeutics targeting these risk factors including chemo-prevention.

Methods
Study subjects. Multiple case breast cancer families. Subjects were members of 25 multi-generational families with multiple cases of breast cancer. The families were participants in the Kathleen Cunningham Foundation Consortium for Research into Familial Breast Cancer (kConFab) and the Australian Breast Cancer Family Registry (ABCReg) 39–41. The present study was based on samples and phenotypic data from a total of 210 family members (87 affected and 123 unaffected) from 25 families and phenotypic data on their relatives.
One or more members of these families had undergone previous genetic testing and were not found to carry a mutation in a known breast cancer susceptibility gene. Genomic DNA was isolated from blood samples or (if no blood specimen was available) from Epstein-Barr virus transformed cell lines (Supplementary Data 2). All participants provided signed informed consent to participate in the relevant research resources. This study was approved by the Human Research Ethics Committee of The University of Melbourne (144955) and meets the principles of the Declaration of Helsinki.

**Melbourne Collaborative Cohort Study (MCCS).** Data from an independent nested case–control study of methylation as a risk factor for breast cancer within the Melbourne Collaborative Cohort Study (MCCS) were used to test the findings from the family analysis. MCCS included all families with a case of invasive adenocarcinoma of the breast (International Classification of Diseases for Oncology, C50) occurring between blood collection and 31 December 2007 and ascertained by record linkage to the population-based Victorian Cancer Registry (VCR), and to the Australian Cancer Database. Controls were selected through incidence 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, buffy coat, or lymphocyte). The HM450k array was used to measure genome-wide methylation in DNA prepared from peripheral blood sample collected prior to cancer diagnosis of the cases as described by Severi et al. (2014).2 All participants provided signed informed consent to participate in the relevant research resources.

**Bisulfit conversion and the HM450K array.** A total of 500 ng of genomic DNA per sample was bisulfite converted using Zymo Gold EZ-DNA kit (Irvine, CA). Prior to processing the bisulfit converted samples on the Infinium HM450 K BeadArray, the conversion was confirmed using bisulfi-te-specific PCR designed in-house27. The Infinium HM450 K (San Diego, CA) was performed using the TECAN automated liquid handler (Männedorf, Switzerland) according to the manufacturer’s instruction.

**HM450K data processing.** All bioinformatic processing was performed with R version 3.2.0.20 Raw intensity signals were imported and processed using the minfi package. All samples had an average detection p-value < 0.001, indicating good quality data. Therefore, no sample was removed from the analysis. Wherever possible, all individuals in the same families were run on the same chips. Individual CpG probes with detection p-value greater than 0.03 (3949 CpG probes) were deemed unreliable and excluded from further analyses. All samples were Illumina and SWAN normalised to reduce technical bias between Type 1 and Type 2 probes25. p-values and M-values were calculated in minfi25,26,27, p-values denote relative methylation percentage calculated from the ratio of the methylated probe intensity and the overall intensity, where 0 indicates 0% methylation and 1 indicates 100% methylation.24 Due to the heteroscedastic nature of p-values and unsuitable for many statistical tests, M-values, which are the log2 of p-values, are also calculated.25 Methylation measures from twelve technical duplicates were used for testing the reproducibility of methylation measures and removed from subsequence analysis. No further batch correction method was performed.

**Clonal bisulfite sequencing.** Clonal bisulfite sequencing was performed to test for the parent-of-origin allelic methylation patterns of the VTRNA2-1 locus as previously described.25 Germline DNA provided by 8 families, including 16 children from the parent-of-origin allelic methylation patterns of the VTRNA2-1 gene. Genomic DNA was isolated from blood samples or (if no blood specimen was available) from Epstein-Barr virus transformed cell lines (Supplementary Data 2). All participants provided signed informed consent to participate in the relevant research resources. This study was approved by the Human Research Ethics Committee of The University of Melbourne (144955) and meets the principles of the Declaration of Helsinki.

**Statistical methods.** Our method for identifying heritable methylation marks is based on a generalisation of the standard expectation–maximisation (EM) algorithm for Gaussian mixtures to allow for non-independent group memberships. These calculations were performed using custom code implemented in R version 3.1.1.43 Because existing general segregation analysis software was too slow to make the calculations feasible for almost half a million probes. For each methylation site (CpG probe), two statistical models were fitted to the site’s M-values: a mixture model, in which the M-values were modelled as a mixture of two normal distributions (with means and variances to be estimated); and a Mendelian model, which is the same as the mixture model except that group membership was modelled as the carrier status (e.g., for a rare variant) at an autosomal genetic locus. Therefore, group memberships are independent under the mixture model but not under the Mendelian model. The maximised log-likelihood, \( l_{\text{mix}} \), for the mixture EM algorithm, with \( l_{\text{mix}} \) obtained from the standard EM algorithm for Gaussian mixtures and \( l_{\text{Mendel}} \) calculated using the modification of this algorithm described in The EM algorithm for the Mendelian model, below. For each model, setting the means and variances for the two groups to be equal corresponds to a Gaussian model in which the M-values follow a normal distribution, so this Gaussian model is nested inside both the mixture and Mendelian models. Using the likelihood ratio test to compare these models to the Gaussian model is uninformative because many probes appear to have a bimodal distribution, so instead we computed \( \ln l_{\text{mix}} \) to \( l_{\text{Mendel}} \), the likelihood quantifies how well a model fits the observed data. Therefore, \( \Delta l = l_{\text{Mendel}} - l_{\text{mix}} \) is a measure of how Mendel’s the probe’s M-values are, and over and above how bimodal they are. Note also that since the mixture and Mendelian models have the same number of model parameters, \( \Delta l \) is the difference between the AIC for the mixture model and the AIC for the Mendelian model so we select the Mendel model in preference to the mixture model whenever \( \Delta l > 0 \) (and similarly for the BIC).47 To validate the ability of the \( \Delta l \) statistic to identify methylation sites with Mendelian-like inheritance patterns, we calculated \( \Delta l \) for all 481,563 methylation sites and used logistic regression and the likelihood ratio test to test whether or not the proportion of probes within 10 bp of a known SNP increases with \( \Delta l \). This is a test on the efficacy of our statistic \( \Delta l \), because the observed M-values of methylation probes with nearby SNPs are likely to have Mendelian-like inheritance patterns, just as an artefact of how the HM450 K array measures methylation.48 The HM450K probes are Stimer oligonucleotides in design with the interrogated target CpGs at the last base. A technical limitation of the platform is that a large proportion of probes overlap one or more known SNPs.48 As the accuracy of methylation measurements relies on the efficient hybridising of probes to target complementary DNA fragment, SNPs within probes potentially interfere with this binding and interrupt the action of the methylated probe and will tend to follow Mendel patterns of inheritance. We could therefore assess if \( \Delta l \) identified heritable sites by testing whether probes with higher values of \( \Delta l \) were more likely to have nearby SNPs. In addition to the formal test above, we also grouped probes by the number of bases of \( \Delta l \) and graphed the proportion of probes within 10 bp of a known SNP for each bin. Known SNPs were defined by Illumina’s HM450K Manifest v1.2 (see Web resources).

To identify heritable methylation marks associated with breast cancer, we first excluded all methylation probes on sex chromosomes or within 10 bp of known SNPs when we screened the remaining 365,169 probes for those most consistent with a Mendelian pattern of inheritance, using the statistic \( \Delta l \). Note that this screening was based on the structure of the 25 families and did not use any data on breast cancer-affected status or age. For each of the 1000 most Mendelian sites (those with the highest values of \( \Delta l \)), we calculated carrier probabilities for the hypothesised Mendelian genetic trait and the likelihood ratio test to test whether the relevant research resources.
she is a non-carrier. Note that $y_i$ and $y_j$ will generally be dependent random variables if individuals $i$ and $j$ belong to the same pedigree, though we still assume that $x_1, ..., x_n$ are conditionally independent given $y_1, ..., y_n$.

We will refer to these models as the mixture and Mendelian models, respectively. Setting $\sigma_n = \mu_1$ and $\sigma_2 = \mu_2$ in either of these models gives a third model for the $M$-values, in which $x_1, ..., x_n$ are independent and follow a univariate normal distribution, that we call the Gaussian model. The maximised log-likelihood $\ell_{\text{mix}}$ and $\ell_{\text{Gauss}}$ of these three models measure the goodness-of-fit of each model to the site’s $M$-values. Since the Gaussian model is nested inside the other two models, $\ell_{\text{mix}}$ and $\ell_{\text{Gauss}}$ can both be formally compared to $\ell_{\text{Gauss}}$, using a likelihood ratio test in order to determine if either of these models gives a more parsimonious fit to the data than the Gaussian model. However, the $M$-values of such a large number of the sites are bimodal, so these tests very often prefer both of the other models to the Gaussian model. To discover sites whose methylation patterns are Mendelian, we therefore compare $\ell_{\text{mix}}$ to $\ell_{\text{Mendel}}$, even though the mixture and Mendelian models are not nested. Since these models have the same number of parameters, $\Delta \ell = \ell_{\text{Mendel}} - \ell_{\text{mix}}$ is the carrier probability for person $i$ corresponding to $x$ and the parameter values $\theta$, when $l = 1$ (note that the $t$ in $\Delta \ell$ is a general superscript, not a power). Therefore, (2) is a weighted log-likelihood of normal distributions, so it can be maximised in exactly the same way as for the standard EM algorithm for Gaussian mixtures. This gives the following parameter values at iteration $t + 1$, for each $l = 0, 1$:

$$\mu_{l+1} = \frac{\sum_i y_{lj} x_i}{\sum_i x_i} \quad \text{and} \quad \sigma_{l+1} = \sqrt{\frac{\sum_i (x_i - \mu_{l+1})^2}{\sum_i x_i}}$$

To calculate $\Delta \ell$ from the right-hand side of (5), we note that, as before, $\Delta \ell = P(x_{lj} \mid y_{lj}) = \prod_{y_{lj}} P(x_{lj} \mid y_{lj}) / P(y_{lj} \mid \theta_l)$.

Also, $P(y_{lj} \mid \theta_l)$ can be calculated using standard techniques from segregation analysis, as described in more detail in Statistical methods, above. Finally, the denominator $P(y_{lj} \mid \theta_l)$ in the right-hand side of (5), which is just a normalising constant, can be obtained by summing the numerator overall values of $y_{lj}$. Therefore, $P(x_{lj} \mid y_{lj})$ can be calculated from (5), so substituting this into (3) gives $q_d^t$ by (4), gives the updated parameters for the EM algorithm.

Improving calculation speeds: our analyses of ~480,000 sites would not be feasible without a number of techniques to improve the speed of the EM algorithm for the Mendelian model, so we briefly describe two of these techniques now.

The Mendelian model is a segregation analysis model, and for such models the most time-consuming part of the calculation is summing over all possible genotype combinations for all family members in each family. However, this part of the calculation is essentially common to all methylation sites, so we obtain considerable improvements in speed by performing this calculation once and storing the results for later use.

More precisely, the update equations (4) for the EM algorithm depend on the carrier probabilities $P(y_{lj} \mid \theta_l)$ via (3) and (5), where we recall that $y_{lj}$ is a set of all sites over all members of family $f$ with epigenome-wide data. Using standard techniques from segregation analysis, $P(y_{lj} \mid \theta_l)$ can be expressed as a sum over carrier genotypes for the family which are consistent with the genotypes $y_l$. Evaluating these sums is usually very time-consuming, however $P(y_{lj} \mid \theta_l)$ depends on $\sigma_1^2$ in $\Delta \ell$ but not on $\mu_1^2$, $\mu_2^2$, $\sigma_1^2$, or $\sigma_2^2$, and $\sigma_1^2$ and $\sigma_2^2$ are held fixed for all $l$, so $P(y_{lj} \mid \theta_l)$ does not depend on $\sigma_1$ or $\sigma_2$ of the family. Therefore, we calculated $P(y_{lj} \mid \theta_l)$ once for every possible combination $y_l$ of genotypes, and stored these values of $P(y_{lj} \mid \theta_l)$ for later use in the update equations (4) (via (3) and (5)) for each family.

We also used a simplifying assumption. To reduce the number of genotype combinations $y_l$ for which we had to store values of $P(y_{lj} \mid \theta_l)$, we assumed that no more than 1 of the founders in each family is a carrier and that no founder is a homozygote carrier (as usually holds if the variant is rare). This assumption is not essential, however, and it can be weakened (e.g., to allow 2 variants or less among the alleles of the founders) or entirely dispensed with (if the families are not too large and not too many family members have epigenome-wide data).
Testing 24 methylation marks in the MCCS. For each of the 24 CpG sites of interest, we first estimated odds ratios (OR) for breast cancer risk using conditional logistic regression models, for a one standard deviation increase in the methylation M-values in blood HM450K data of set of 433 cases and their matched controls from the MCCS. The models were adjusted for body mass index, tobacco smoking, alcohol drinking, time between blood collection, and cancer diagnosis, and sample type (DNA extracted from dried blood spots, peripheral blood mononuclear cells, and buffy coats, although the vast majority (97%) of case–control pairs were successfully matched on sample type). For methylation marks exhibiting a bimodal or trimodal distribution, we categorised the methylation variables into groups corresponding to the observed ‘peaks’ of hypo-, hemimethylated or hypermethylated, based on visual inspection of the M-value distribution (Supplementary Fig. 4). We used the same models as for the continuous variable analyses. The larger peak was chosen as the reference category. Sensitivity analyses were conducted: (1) further adjusting the models for blood cell composition as estimated by the algorithm by Houseman et al.63; (2) further adjusting the models for age at menarche, menopausal status, number of live births, and use of hormonal replacement therapy; (3) restricting the analyses to DNA prepared from dried blood spots.

Associations between genetic variants and DNA methylation. Data for all variants with 1 kb of the GREB1 probe that were genotyped or imputed using the iCOGS array were retrieved for MCCS participants included in the Breast Cancer Association Consortium.46 A total of 251 participants (231 cases and 20 controls) had iCOGS and HM450K data available. Association between genotype and methylation was assessed using linear regression, with beta-value as the outcome variable and allele count as the explanatory variable. The allele count was estimated by rounding the allele dose to an integer value.

References
1. Delgado-Cruzata, L., Wu, H. C., Liao, Y., Santella, R. M. & Terry, M. B. Differences in DNA methylation by extent of breast cancer family history in unaffected women. Epigenetics. 9, 243–248 (2014).
2. Severi, G. et al. Epigenome-wide methylation in DNA from peripheral blood as a marker of risk for breast cancer. Breast Cancer Res. Treat. 148, 665–673 (2014).
3. van Veldhoven, K. et al. Epigenome-wide association study reveals decreased average methylation levels years before breast cancer diagnosis. Clin. Epigenet. 7, 67 (2015).
4. Armes, J. E. et al. The histologic phenotypes of breast carcinoma occurring before age 40 years in women with and without BRCA1 or BRCA2 germline mutations: a population-based study. Cancer 83, 2335–2345 (1998).
5. Lakhani, S. R. et al. Multifactorial analysis of differences between sporadic breast and ovarian cancer. Br. J. Cancer 104, 903–909 (2011).
6. Wong, E. M. et al. Constitutional methylation of the BRCA1 promoter is specifically associated with BRCA1 mutation-associated pathology in early-onset breast cancer. Cancer Prev. Res (Phila.) 4, 23–33 (2011).
7. Hansmann, T. et al. Constitutive promoter methylation of BRCA1 and RAD51C in patients with familial ovarian cancer and early-onset sporadic breast cancer. Hum. Mol. Genet. 21, 4669–4679 (2012).
8. Bernstein, J. L. et al. Population-based estimates of breast cancer risks associated with ATM gene variants c.727T>G and c.1066-6T>G (IVS10-6T>G) from the Breast Cancer Family Registry. Hum. Mutat. 27, 1122–1128 (2006).
9. Goldgar, D. E. et al. Rare variants in the ATM gene and risk of breast cancer. Breast Cancer Res. 13, R73 (2011).
10. Tavtigian, S. V. et al. Evolutionarily unlikely missense substitutions in ATM confer increased risk of breast cancer. Am. J. Hum. Genet. 85, 427–446 (2009).
11. Flanagan, J. M. et al. DNA methylation of familial breast cancer identifies distinct profiles defined by mutation status. Am. J. Hum. Genet. 86, 420–433 (2010).
12. Brennan, K. et al. Intragene ATM methylation in peripheral blood DNA as a biomarker of breast cancer risk. Cancer Res. 72, 2304–2313 (2012).
13. Potapaova, A., Hoffman, A. M., Godwin, A. K., Saleem, T. & Cairns, P. Promoter hypermethylation of the PALB2 susceptibility gene in inherited and sporadic breast and ovarian cancer. Cancer Res. 68, 998–1002 (2008).
14. Mikeska, T., Alsp, K., Australian Ovarian Cancer Study Group, Mitchell, G., Bowtell, D. D. & Dobrovic, A. No evidence for PALB2 methylation in high-grade serous ovarian cancer. J. Ovarian Res. 6, 26 (2013).
15. Oey, H. & Whitelaw, E. On the meaning of the word ‘epimutation’. Trends Genet. 30, 519–520 (2014).
16. Lynch, H. T. Hereditary nonpolyposis colorectal cancer (HNPPC). Cytogenet. Cell Genet. 66, 130–135 (1999).
17. Peltomaki, P. & de la Chapelle, A. Mutations predisposing to hereditary nonpolyposis colorectal cancer. Adv. Cancer Res. 71, 93–119 (1997).
18. Hitchins, M. P. et al. Inheritance of a cancer-associated MLH1 germ-line mutation. N. Engl. J. Med. 356, 897–905 (2007).
19. Suter, C. M., Martin, D. J. & Ward, R. L. Germline methylation of MLH1 in individuals with multiple cancers. Nat. Genet. 36, 497–501 (2004).
20. Hitchins, M. P. et al. Dominantly inherited constitutional epigenetic silencing of MLH1 in a cancer-affected family is linked to a single nucleotide variant within the 5′UTR. Cancer Cell 20, 260–273 (2011).
21. Ligtenberg, M. J. et al. Heritable somatic methylation and inactivation of MSH2 in families with Lynch syndrome due to deletion of the 3′ exons of TACSTD1. Nat. Genet. 41, 112–117 (2009).
22. Hitchins, M. P. The role of epigenetics in Lynch syndrome. Fam. Cancer 12, 189–205 (2013).
23. Di P. et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. BMC Bioinformatics 11, 587 (2010).
24. Paliwal, A. et al. Comparative anatomy of chromosomal domains with imprinted and non-imprinted allele-specific DNA methylation. PLoS Genet. 9, e1003622 (2013).
25. Romanelli, V. et al. Variable maternal methylation overlapping the nc886/eRNA2-1 locus is locked between hypermethylated repeats and is frequently altered in cancer. Epigenics. 9, 783–790 (2014).
26. Turnbull, C. & Rahman, N. Genetic predisposition to breast cancer: past, present, and future. Annu. Rev. Genom. Hum. Genet. 9, 321–345 (2008).
27. McRae, A. F. et al. Contribution of genetic variation to transgenerational inheritance of DNA methylation. Genome Biol. 15, R73 (2014).
28. Lee, K. et al. Precancer miR-866, a novel noncoding RNA repressed in cancer, associates with PKR and modulates its activity. RNA 17, 1076–1089 (2011).
29. Kunkeaw, N. et al. Cell death/proliferation fororn866, a non-coding RNA, in the protein kinase R pathway in cholangiocarcinoma. Oncogene 32, 3722–3731 (2013).
30. Lee, H. S. et al. Epigenetic silencing of the non-coding RNA nc868 provokes oncogenes during human esophageal tumorigenesis. Oncotarget 5, 3472–3481 (2014).
31. Lee, K. S. et al. nc868, a non-coding RNA of anti-proliferative role, is suppressed by CpG DNA methylation in human gastric cancer. Oncotarget 5, 3944–3955 (2014).
32. Johannsdottir, H. K. et al. Chromosome 5 imbalance mapping in breast tumors from BRCA1 and BRCA2 mutation carriers and sporadic breast tumors. Int. J. Cancer 119, 1052–1060 (2006).
33. Wang, Z. C. et al. Loss of heterozygosity and its correlation with expression profiles in subclasses of invasive breast cancers. Cancer Res. 64, 67–71 (2004).
34. Silver, M. I. et al. Independent genomewide screens identify the tumor suppressor VTRNA2-1 as a human epiallele responsive to periconceptional environment. Genome Biol. 16, 118 (2015).
35. Reik, W. & Murrell, A. Genomic imprinting. Silence across the border. Nature 405, 408–409 (2000).
36. Ghost, M. G., Thompson, D. A. & WeigelR. J. PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer. Cancer Res. 60, 6367–6375 (2000).
37. Rae, J. M. et al. GREB 1 is a critical regulator of hormone dependent breast cancer growth. Breast Cancer Res. Treat. 92, 141–149 (2005).
38. John, E. M. et al. The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer. Breast Cancer Res. BCR, 8, R375–R389 (2004).
39. Osborne, R. H. et al. KCoNFab: a research resource of Australasian breast cancer families. Kathleen Cunningham Foundation Consortium for Research into Familial Breast Cancer. Med. J. Aust. 172, 463–464 (2000).
40. Giles, G. G. & English, D. R. The Melbourne Collaborative Cohort Study. Iarc. Sci. Publ. 156, 69–70 (2008).
41. Wong, E. M. et al. Tools for translational epigenetic studies involving formalin-fixed paraffin-embedded human tissue: Applying the Infinum HumanMethylation450 Beadchip assay to large population-based studies. BMC Res. Notes 8, 543 (2015).
42. Noble, T. R. A language and environment for statistical computing. R Foundation for Statistical Computing. (R programming software, Vienna, Austria, 2015).
Acknowledgements

The Australian site of Breast Cancer Family Registry was supported by grant UM1 CA164920 from the USA National Cancer Institute. The content of this manuscript does not necessarily reflect the views or policies of the National Cancer Institute or any of the collaborating centres in the Breast Cancer Family Registry (BCFR), nor does mention of trade names, commercial products, or organisations imply endorsement by the USA Government or the BCFR. We thank Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow-Up Study (which has received funding from the NHMRC, the National Breast Cancer Foundation, Cancer Australia, and the National Institute of Health (USA)) for their contributions to this resource, and the many families who contribute to kConFab. kConFab is supported by a grant from the National Breast Cancer Foundation, and previously by the National Health and Medical Research Council (NHMRC), the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. 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, programme managers and the diligent team who recruited the participants and who continue working on follow-up. The MCCS methylation work was supported by the National Health and Medical Research Council (Grant number 1011618); and the Victorian Breast Cancer Research Consortium. M.C.S. is a Senior Research Fellow and J.L.H. is a Senior Principal Research Fellow of the National Health and Medical Research Council of Australia. This work was supported by an Early Career Research Award to JEJ from The University of Melbourne.

Author contributions

This study was first conceived and designed by JEJ, E.M.W. and M.C.S. JEJ, performed all laboratory experiments. JEJ, J.G.D., R.L.M. and P.A.D. performed bioinformatics and statistical analyses. R.L.M. and G.G.G. facilitated the inclusion and interpretation of the data from the MCCS. Study materials were provided by kConFab, ABCER, and MCCS. The manuscript was first structured by JEJ, J.G.D., and M.C.S. D.E., J.L.H. and D.E.G. provided significant intellectual contributions. All authors reviewed the manuscript.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-03058-6.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

kConFab

Adrienne Sexton, Alice Christian, Alison Trainer, Allan Spigelman, Andrew Fellows, Andrew Shelling, Anna De Fazio, Anneke Blackburn, Ashley Crook, Bettina Meiser, Bronny Patterson, Christine Clarke, Christobel Saunders, Clare Hunt, Clare Scott, David Amor, Deborah Marsh, Edward Edkins, Elizabeth Salisbury, Eric Haan, Eveline Neidermayr, Finlay Macrae, Gelareh Farshid, Geoff Lindeman, Georgia Cheniveix-Trench, Graham Mann, Granthly Gill, Heather Thorne, Ian Campbell, Ian Hickie, Ingrid Winship, Jack Goldblatt, James Flanagan, James Kollias, Jane Visvader, Jennifer Stone, Jessica Taylor, Jo Burke, Jodi Saunus, John Forbes, Jonathan Beesley, Judy Kirk, Juliet French, Kathy Tucker, Kathy Wu, Kelly Phillips, Lara Lipton, Leslie Andrews, Elizabeth Lobb, Logan Walker, Maira Kentwell, Amanda Spurdle, Margaret Cummings, Margaret Gleeson, Marion Harris, Mark Jenkins, Mary Anne Young, Martin Delatycki, Mathew Wallis, Matthew Burgess, Melanie Price, Melissa Brown, Michael Bogwitz, Michael Field, Michael Friedlander, Michael Gattas, Mona Saleh, Nick Hayward, Nick Pachter, Paul Cohen, Pascal Duijf, Paul James, Peter Simpson, Peter Fong, Phyllis Butow, Rachael Williams, Richard Kefferd, Rodney Scott, Rosemary Balleine, Sarah-Jane Dawson, Sheau Lok, Shona O’Connell, Sian Greening, Sophie Nightingale, Stacey Edwards, Stephen Fox, Sue-Anne McLachlan, Sunil Lakhani, Susan Thomas & Yoland Antill.

© The Author(s) 2018
Author/s:
Joo, JE; Dowty, JG; Milne, RL; Wong, EM; Dugue, P-A; English, D; Hopper, JL; Goldgar, DE; Giles, GG; Southey, MC

Title:
Heritable DNA methylation marks associated with susceptibility to breast cancer

Date:
2018-02-28

Citation:
Joo, J. E., Dowty, J. G., Milne, R. L., Wong, E. M., Dugue, P.-A., English, D., Hopper, J. L., Goldgar, D. E., Giles, G. G. & Southey, M. C. (2018). Heritable DNA methylation marks associated with susceptibility to breast cancer. NATURE COMMUNICATIONS, 9 (1), https://doi.org/10.1038/s41467-018-03058-6.

Persistent Link:
http://hdl.handle.net/11343/220797

File Description:
Published version

License:
CC BY