



RESEARCH

Open Access

Coordinated epigenetic remodelling of transcriptional networks occurs during early breast carcinogenesis

Warwick J Locke^{1,2}, Elena Zotenko^{1,2}, Clare Stirzaker^{1,2}, Mark D Robinson³, Rebecca A Hinshelwood^{1,4}, Andrew Stone^{1,2}, Roger R Reddel^{5,6}, Lily I Huschtscha⁵ and Susan J Clark^{1,2*}

Abstract

Background: Dysregulation of the epigenome is a common event in malignancy; however, deciphering the earliest cancer-associated epigenetic events remains a challenge. Cancer epigenome studies to date have primarily utilised cancer cell lines or clinical samples, where it is difficult to identify the initial epigenetic lesions from those that occur over time. Here, we analysed the epigenome of human mammary epithelial cells (HMEC) and a matched variant cell population (vHMEC) that have spontaneously escaped senescence and undergone partial carcinogenic transformation. Using this model of basal-like breast carcinogenesis, we provide striking new insights into the very first epigenetic changes that occur during the initial stages of malignancy.

Results: The first phase of malignancy is defined by coordinated changes in the epigenome. At the chromatin level, this is embodied in long-range epigenetic deregulation, which involves the concomitant but atypical acquisition or loss of active and repressive histone modifications across large regional blocks. Changes in DNA methylation also occurs in a highly coordinated manner. We identified differentially methylated regions (DMRs) in the very earliest passages of vHMECs. Notably, we find that differential methylation targets loci regulated by key transcription factors including p53, AHR and E2F family members suggesting that epigenetic deregulation of transcription factor binding is a key event in breast carcinogenesis. Interestingly, DMRs identified in vHMEC are extensively methylated in breast cancer, with hypermethylation frequently encroaching into neighbouring regions. A subset of vHMEC DMRs exhibited a strong basal-like cancer specific hypermethylation.

Conclusions: Here, we generated epigenome-wide maps of the earliest phase of breast malignancy and show long-range epigenetic deregulation and coordinated DNA hypermethylation targets loci regulated by key transcription factors. These findings support a model where induction of breast cancer occurs through epigenetic disruption of transcription factor binding leading to deregulation of cancer-associated transcriptional networks. With their stability and very early occurrence, vHMECs hypermethylated loci could serve as excellent biomarkers for the initial detection of basal breast cancer.

Keywords: DNA methylation, Methylome, Epigenetics, Basal breast cancer, Epigenome sequencing, Biomarker

* Correspondence: s.clark@garvan.org.au

¹Epigenetic Research Laboratory, Genomics and Epigenetic Division, The Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, NSW 2010, Australia

²St. Vincent's Clinical School, Faculty of Medicine, University of New South Wales, Level 5 deLacy Building, St Vincent's Hospital, Victoria Street, Darlinghurst, NSW 2010, Australia

Full list of author information is available at the end of the article

Background

The epigenetic control of gene expression plays a critical role in the development of normal cells and is extensively deregulated in all cancers, including breast cancer (reviewed in Baylin and Jones [1]). Hypermethylation frequently occurs in high-density clusters termed CpG islands and is often associated with gene promoters and transcriptional repression in cancer, whereas decreased methylation at regulatory elements is associated with gene activation [1,2]. Alterations in DNA methylation contribute to carcinogenesis via a variety of mechanisms including the silencing of tumour suppressors [1,2] and reactivation of transposons [3], genomic instability [4,5] and blocking differentiation [6]. Furthermore, the cancer DNA methylation profile can be associated with specific molecular subtypes [7-9], various clinicopathological parameters (reviewed in Locke and Clark [10]) and response to endocrine therapy [11,12].

While epigenetic dysregulation in cancer is well established, the timing and extent of epigenetic change during carcinogenesis remains largely unknown. Most studies utilise comparisons of cancer cell lines or cancer tissue samples with normal tissue to identify cancer-associated epigenetic aberrations that occur during tumour development. However, these studies cannot identify the early events associated with epigenetic remodelling during transformation. To address this issue, we utilised the human mammary epithelial cell (HMEC) model system [10,13]. HMECs are derived from normal breast tissue removed from healthy women during breast reduction mammoplasty [14]. In culture HMEC undergo two distinct phases of growth separated by a temporary period of growth arrest similar to senescence, termed 'selection'. Cells from the two growth phases are termed HMEC prior to selection and then, after overcoming senescence barriers during selection, are termed variant HMEC (vHMEC) during the second growth phase [15,16]. When compared to HMECs, vHMECs exhibit multiple cancer-associated gene expression and epigenetic changes and are considered to represent a partially transformed pre-malignant breast cell (reviewed in Locke and Clark [10] and Hinshelwood and Clark [13]). Such changes include silencing of the p16^{ink4a} tumour suppressor associated with promoter hypermethylation [15,17-19], silencing of the transforming growth factor beta pathway via acquisition of repressive chromatin [20] and increased expression of the cancer-associated chromatin methyltransferase *EZH2* [20]. Due to its extended lifespan and multiple cancer-associated expression changes, the HMEC system provides a model of partial carcinogenic transformation from normal to pre-malignancy. Therefore, the HMEC system is an ideal tool for the identification of the first epigenomic events occurring during early breast carcinogenesis.

In order to understand the role of epigenomic deregulation in breast carcinogenesis, we performed detailed expression, DNA methylation and chromatin modification profiling of a set of HMECs and isogenic vHMECs. We show that epigenomic aberrations in key regulatory pathways and across domains occur during the very earliest stages of breast carcinogenesis. Furthermore, comparison to The Cancer Genome Atlas Breast invasive CArcinoma (TCGA-BRCA) cohort demonstrates that the methylation aberrations we identified in vHMEC are common in basal-like breast tumours suggesting that epigenetic lesions occurring early in carcinogenesis are derived by similar reprogramming events.

Results

vHMEC is a model of early basal-like breast carcinogenesis

To gain a more detailed understanding of the early epigenetic changes that occur in the first stages of carcinogenesis, we performed epigenome-wide profiling (gene expression, DNA methylation and chromatin modifications [GEO:GSE58882]) of four isogenic HMEC/vHMEC lines (Bre12, Bre38, Bre67 and Bre98). Given their basal culture conditions [14], it is likely that vHMECs resemble the basal-like molecular subtype of breast cancer. To confirm this, we first classified the vHMEC lines into the intrinsic molecular subtypes of breast cancer using Affymetrix GeneChip expression data with the PAM50 classifier [21]. As highlighted by Sorlie and colleagues [22], it is important that expression array data are gene-centred in addition to standard normalisation procedures prior to PAM50 classification. To ensure our findings were reproducible, we performed the gene-centred analysis with two independent publicly available datasets ([GEO:GSE2034] [23] and [GEO:GSE3494] [24]). After clustering, we found that the vHMEC lines from all four donors classified into the basal-like breast cancer subtype in both data sets, supporting the use of these cells as a model to study breast cancer (Figure 1A [GEO:GSE2034] and Additional file 1: Figure S1 [GEO:GSE3494]).

Next we aimed to identify the gene expression changes that potentially drive the earliest steps in development of breast cancer and/or basal breast cancer. First, we performed multi-dimensional scaling (MDS) analysis on the 500 most variably expressed genes across all donors and time points (Additional file 1: Figure S2A) which separated HMEC and vHMEC into two distinct clusters on dimension 1, indicating that selection (that is, the escape from senescence) is the largest source of variation in our dataset rather than inter-individual variation or extended time in culture.

Subsequent limma analysis identified 2,121 and 1,972 genes differentially expressed in early and late vHMEC, respectively, when compared to HMEC (adjusted $P < 0.05$,

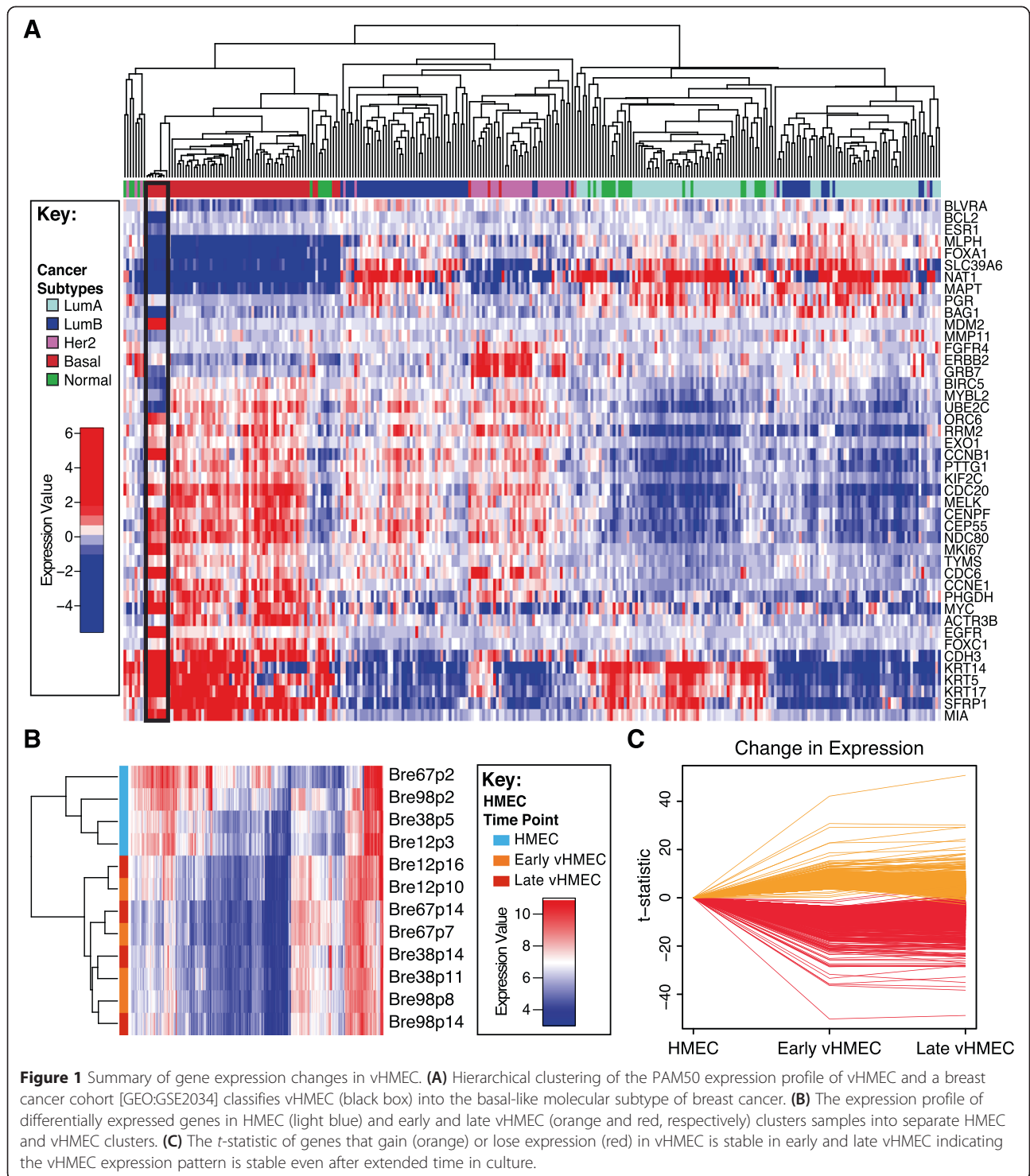


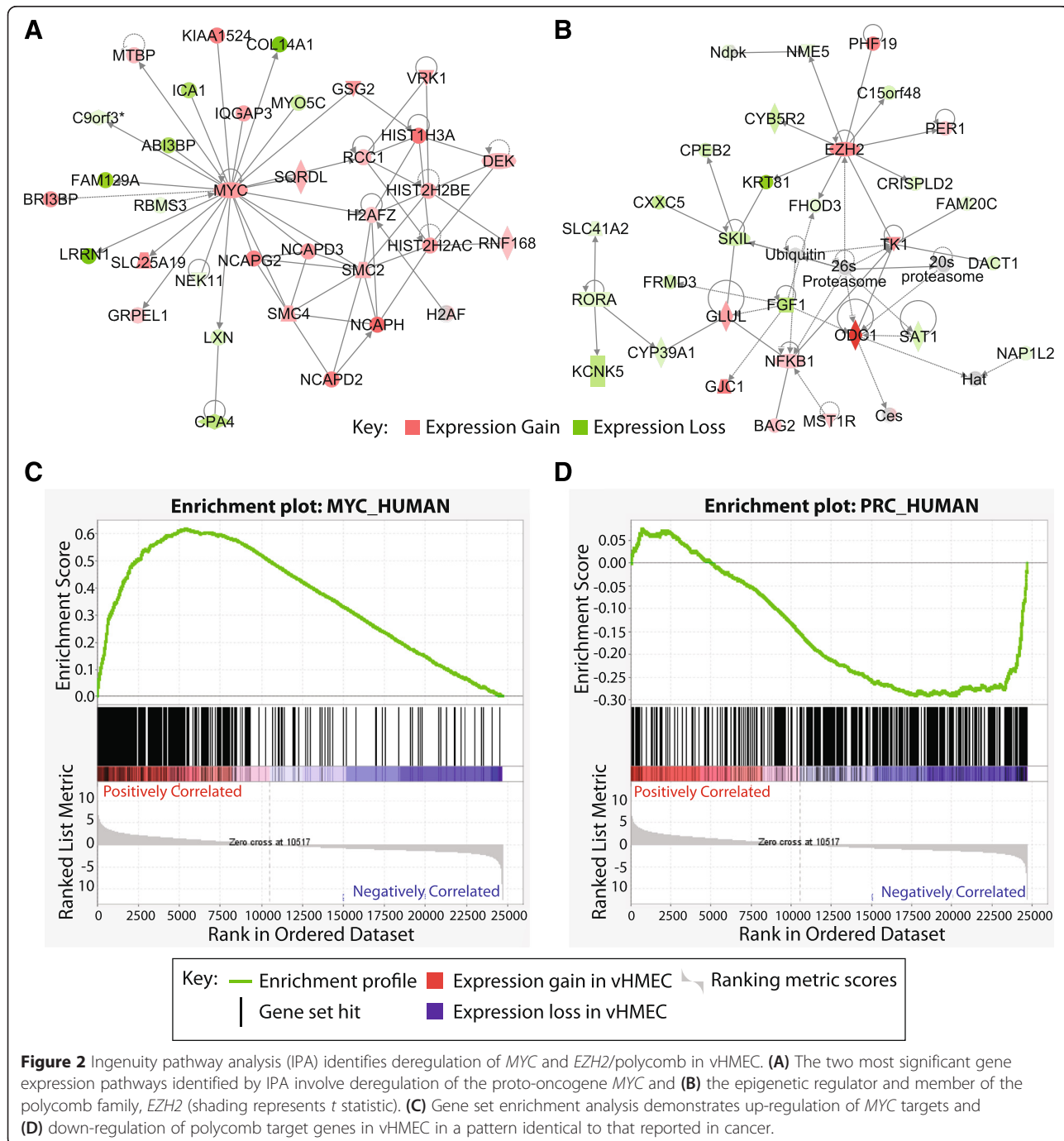
Figure 1B). Validation of differential expression was performed on a subset of eight genes by quantitative polymerase chain reaction (qPCR) and exhibited high correlation with array profiling (Additional file 1: Figure S2B). The expression level of these approximately 2,000 differentially expressed genes was consistent across both

the early and late vHMEC time points (Figure 1C), further indicating that the initial escape from senescence (selection) is the greatest contributor to differential gene expression. Given the high similarity between the vHMEC time points, further expression analyses were performed on the early vHMEC time point only.

Deregulation of polycomb, MYC and the p53 pathway drive early carcinogenesis

To understand the impact of gene expression change on the biology of vHMEC and its ability to overcome senescence, we performed Ingenuity Pathway Analysis (IPA). IPA identified that genes differentially expressed in vHMEC were significantly enriched for functions related to cancer ($P < 0.001$, Additional file 1: Figure S2C), including cellular growth and proliferation, cellular movement,

cell cycle and, DNA replication, recombination and repair ($P < 0.001$). In addition, genes associated with development were enriched, indicating a role for development-related genes in the vHMEC's escape from senescence. The two most significant pathways identified by IPA contained known breast oncogenes, specifically *MYC* (Figure 2B) and the histone methyltransferase *EZH2* (Figure 2B). IPA transcription factor (TF) analysis predicted activation of both *MYC* and *EZH2* in vHMEC



($P = 1.33 \times 10^{-13}$ and 2.12×10^{-03} , respectively, Additional file 2: Table S1). In addition to the identification of gene expression pathways and enrichment of biological functions, IPA can predict alterations in the apparent activity of transcription factors (TF). IPA TF analysis predicted activation of *EZH2* and *MYC*, in agreement with their increased expression levels, as well as several other cancer-associated TFs. These included activation of the NF- κ B (including pathway members *NFKB1* and *RELA*), and E2F (including pathway members *E2F1*, *E2F2* and *E2F3*) signalling pathways, activation of the stem cell/cancer-associated transcription factor *NANOG* and activation of the ligand-activated transcription factor *AHR*. Inhibited TFs included the tumour suppressing p53 and Rb pathways (including pathway members *RBI* and *CDKN2A* (p16)).

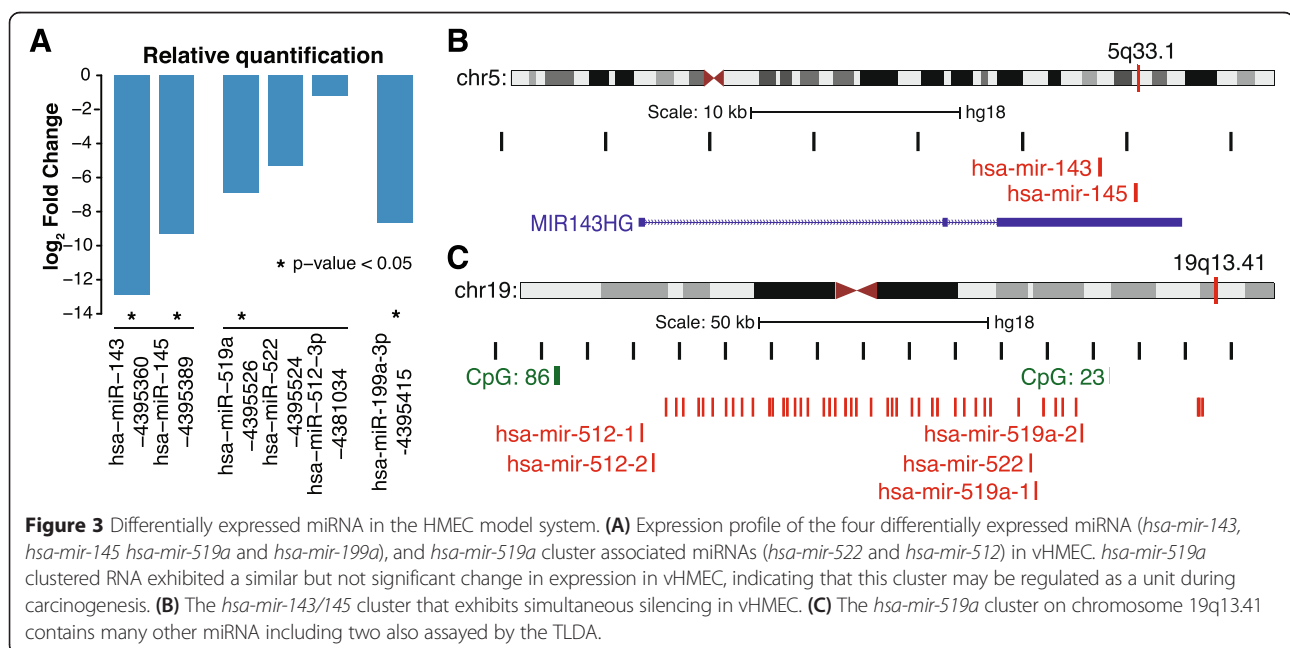
Western blot analysis of AHR in HMEC and vHMEC was performed (Additional file 1: Figure S3A). When active, AHR/ARNT form a complex, and this state was observed in vHMEC and not in HMEC (Additional file 1: Figure S3B). Intriguingly, *ANRT* was identified as inhibited in vHMEC by IPA, and the AHR/ARNT target genes *CYP1A1* and *CYP1B1* [25] are both down regulated at the mRNA level relative to HMEC (adjusted $P = 0.03$ and 0.000057 , respectively). The predicted inhibition of *ARNT* and the apparent silencing of xenobiotic metabolising AHR target genes may suggest that activation of *AHR* in vHMEC is in relation to its putative roles in development or other novel processes rather than xenobiotic response.

The increased activity of *MYC* and a related pathway identified by IPA (Figure 2A and Additional file 2: Table S1) suggests a key role for activating *MYC* in the capacity of vHMEC to overcome senescence barriers. To

investigate this further, we assessed the differential expression patterns of the MYC, PRC and CORE modules published by Kim *et al.* [26]. Together, these gene sets can distinguish the expression patterns of cancer cells from that of stem cells and are comprised of genes activated by *MYC* (MYC_MODULE) and genes silenced by PRC2 (PRC_MODULE) in cancer and stem cells and a set of genes activated in stem cells only by a core group of pluripotency factors (CORE_MODULE). In vHMEC, we observed enrichment of up- and down-regulated genes in the MYC and PRC modules, respectively (Figure 2C, D, respectively, family-wise error rate P value < 0.05) and no significant change in the CORE module genes ($P = 0.06$, Additional file 1: Figure S2D), similar to the expression pattern of cancer [26]. Furthermore, the deregulation of the MYC and PRC modules matches the observed increase in *MYC* and *EZH2* expression and their predicted increase in activity by IPA.

Tumour-suppressing miRNA are silenced in vHMEC

To assess the potential contribution of the deregulation of microRNA (miRNA) expression in the capacity of vHMEC to overcome senescence (and, therefore, in early carcinogenesis), we profiled miRNA expression in a subset of HMEC and early vHMEC (donors Bre12, Bre67 and Bre98) by Taqman® Low-Density Array (TLDA; Life Technologies, Carlsbad, USA). TLDA analysis identified four miRNA with a statistically significant ($P < 0.05$) loss of expression (Figure 3A). Notably, of the four statistically significantly differentially expressed miRNA, three fell into clusters (Figure 3B, C). Both *hsa-mir-143* and *hsa-mir-145* are encoded by the same gene, which



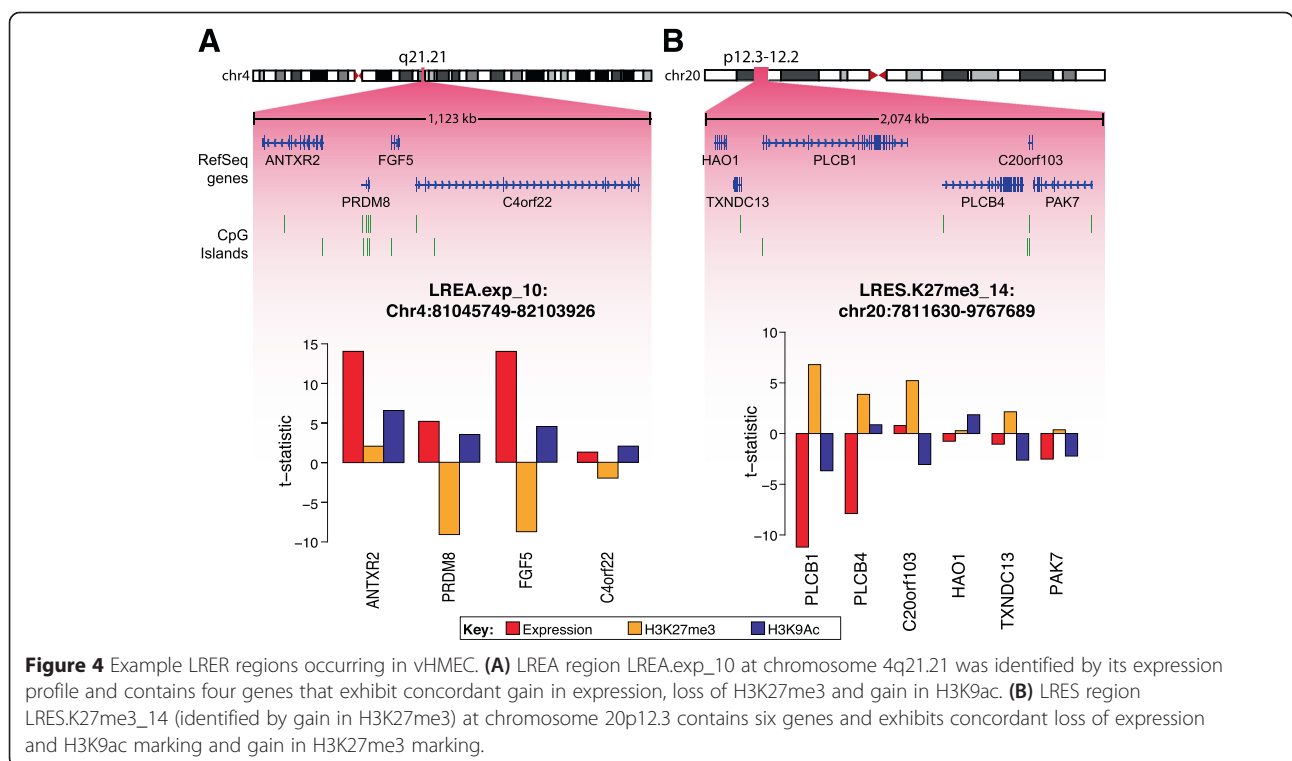
accounts for their similar expression profile. Two copies of *hsa-mir-519a* are located within a very large cluster of miRNA on chromosome 19. Two other miRNA from the *hsa-mir-519a* cluster were assayed by the TLDA platform and exhibited a similar, although non-significant, change in gene expression (*hsa-mir-522* and *hsa-mir-512*, Figure 3A) indicating that this cluster may be regulated as a single unit.

Next, we identified which experimentally validated and predicted target genes of the four vHMEC-silenced miRNA in the TarBase and microRNA.org databases [27,28] exhibited increased expression in vHMEC. This identified 470 interactions (196 unique genes, Additional file 2: Table S2) from the microRNA.org database and six validated targets from the TarBase database (Additional file 2: Table S3). Notably, several key cancer-associated genes were included in the list of deregulated targets, including *MYC* (*hsa-mir-145*), *EZH2* (*hsa-mir199a*) and members of the p53 pathway (*MDM2* and *CDKN1A* targeted by *hsa-mir-199a* and; *hsa-mir-519a* and *hsa-mir-145*, respectively). Additionally, hypergeometric testing using the MSigDB C2 Canonical Pathways database (Additional file 1: Figure S4) revealed that cell cycle regulation, proliferation and p53 regulated pathways were enriched within the predicted miRNA targets.

Long-range epigenetic regulation is an early cancer event

While the deregulation of individual oncogenes and tumour suppressors is an important event in cancer, gene

expression changes can also affect large domains of the genome containing multiple genes that exhibit a concordant change in expression in a phenomenon termed long-range epigenetic regulation (LRER) [29,30]. To study the role of long-range epigenetic activation (LREA) and silencing (LRES), we performed ChIP-chip analysis of H3K9ac on Bre12 and Bre38 and, H3K27me3 on Bre12, Bre38, Bre67 and Bre98 paired HMEC/vHMEC lines and combined with the expression data to look for concordant changes in chromatin-associated activation or repression. Analysis identified 33 regions of LREA and 41 regions of LRES based on change in expression, H3K9ac or H3K27me3 (Additional file 2: Table S4). On average, LREA regions were 0.5 Mb in length and contained an average of nine genes (Figure 4A) and LRES regions were 1.1 Mb and contained an average of eight genes (Figure 4B). Hypergeometric testing revealed that genes within LREA regions were enriched in gene sets associated with transcription and maintenance of the DNA (Additional file 1: Figure S5A), whereas LRES genes were enriched by gene sets associated with cancer and involved in the interaction with the extracellular environment (Additional file 1: Figure S5B). To investigate TF deregulation as a mechanism of LRER region formation, we interrogated LRER regions for TF binding sites (TFBS) associated with deregulated TF identified by IPA. TFBS were identified using a set of evolutionarily conserved, computationally derived TFBS available from the UCSC genome browser website [31,32]. The majority of TFBS tested exhibited no enrichment in



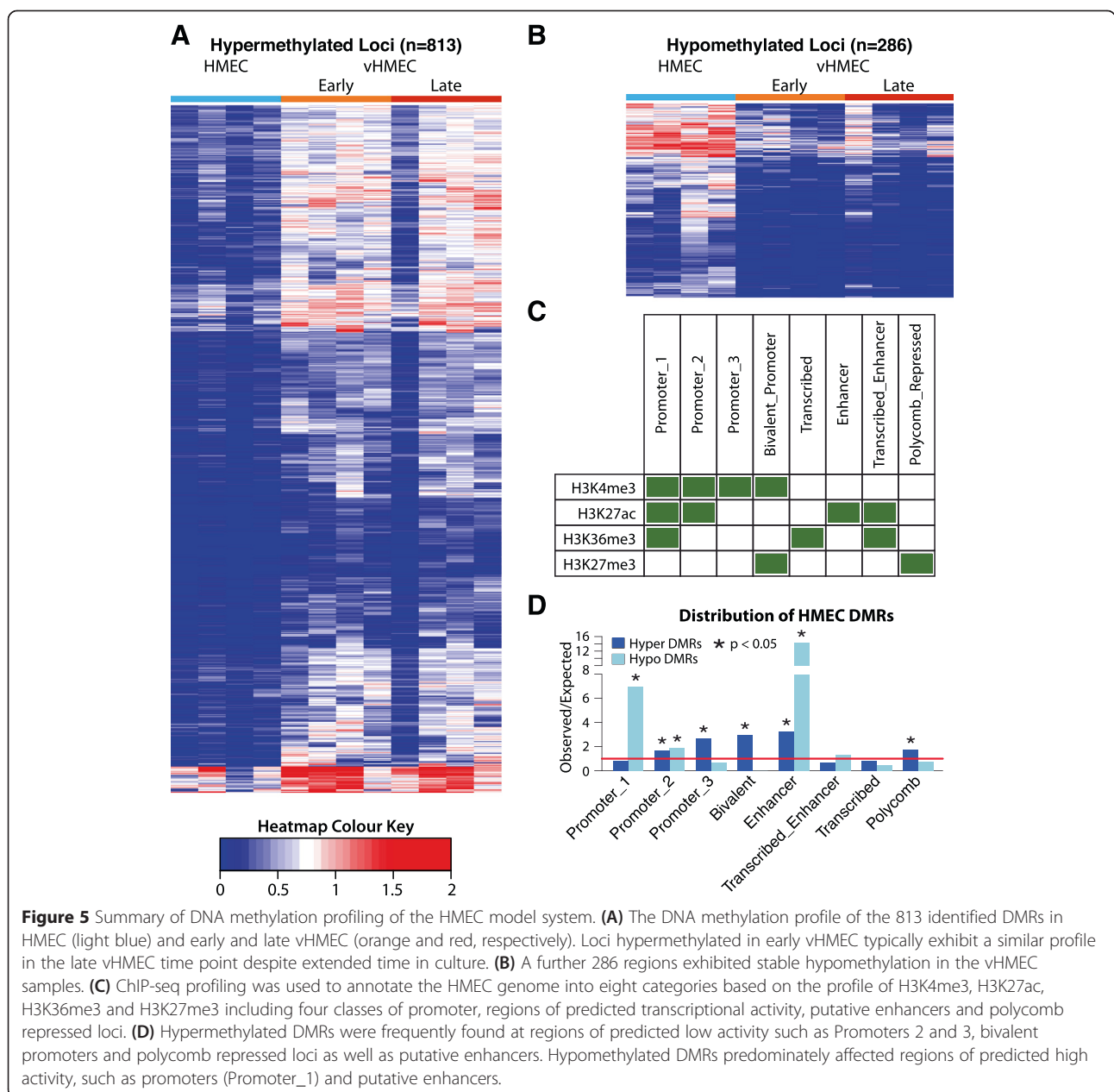
LRER. However, LRES regions were significantly depleted for sites associated with AHR, ARNT, E2F family members and MYC (Additional file 1: Figure S6).

DNA methylation changes occur early in carcinogenesis

The timing of DNA methylation changes during carcinogenic transformation remains largely unknown. To investigate the acquisition of differentially methylated regions (DMRs) during the earliest phases of carcinogenesis, we performed DNA methylation profiling using MBDCap-seq (see ‘Methods’) on HMEC lines (Bre12, Bre38, Bre67 and Bre98) and early and late passage vHMEC. Bioinformatic analysis identified 813 and 286 regions of hyper- and

hypomethylation, respectively, in vHMEC (Figure 5A, B) that were stably maintained between the early and late vHMEC time points. Five gene-associated hypermethylated loci were selected for validation by clonal bisulphite sequencing and all regions exhibited an increase in the density of heterogeneous methylation between HMEC and vHMEC (Additional file 1: Figure S7).

Next, we asked if DMRs observed in vHMEC were associated with functional genomic features. Using RefSeq gene annotations, hypermethylated DMRs were statistically significantly ($P < 0.05$) enriched at CpG island loci and gene promoters (Additional file 1: Figure S8). To identify potential functional non-genic regions, we performed ChIP-seq



profiling of H3K4me3, H3K27ac, H3K36me3 and H3K27me3 in Bre98 HMEC and vHMEC. Overlapping ChIP-seq peaks were used to classify 63,526 regions in HMEC and 61,096 in vHMEC into eight functional categories (designated Promoter_1, Promoter_2, Promoter_3, Bivalent_Promoter, Enhancer, Transcribed_Enhancer, Transcribed and Polycomb_Repressed, Figure 5C). Hypermethylated DMRs were enriched at *low* predicted activity promoters (Promoter_2 and 3), putative enhancer loci and polycomb regulated regions (Bivalent_promoter and Polycomb_Repressed). In contrast, hypomethylation was enriched at *high* activity promoters (Promoter_1) and *heavily* enriched (approximately 14-fold, $P < 0.05$) at the Enhancer category (Figure 5D).

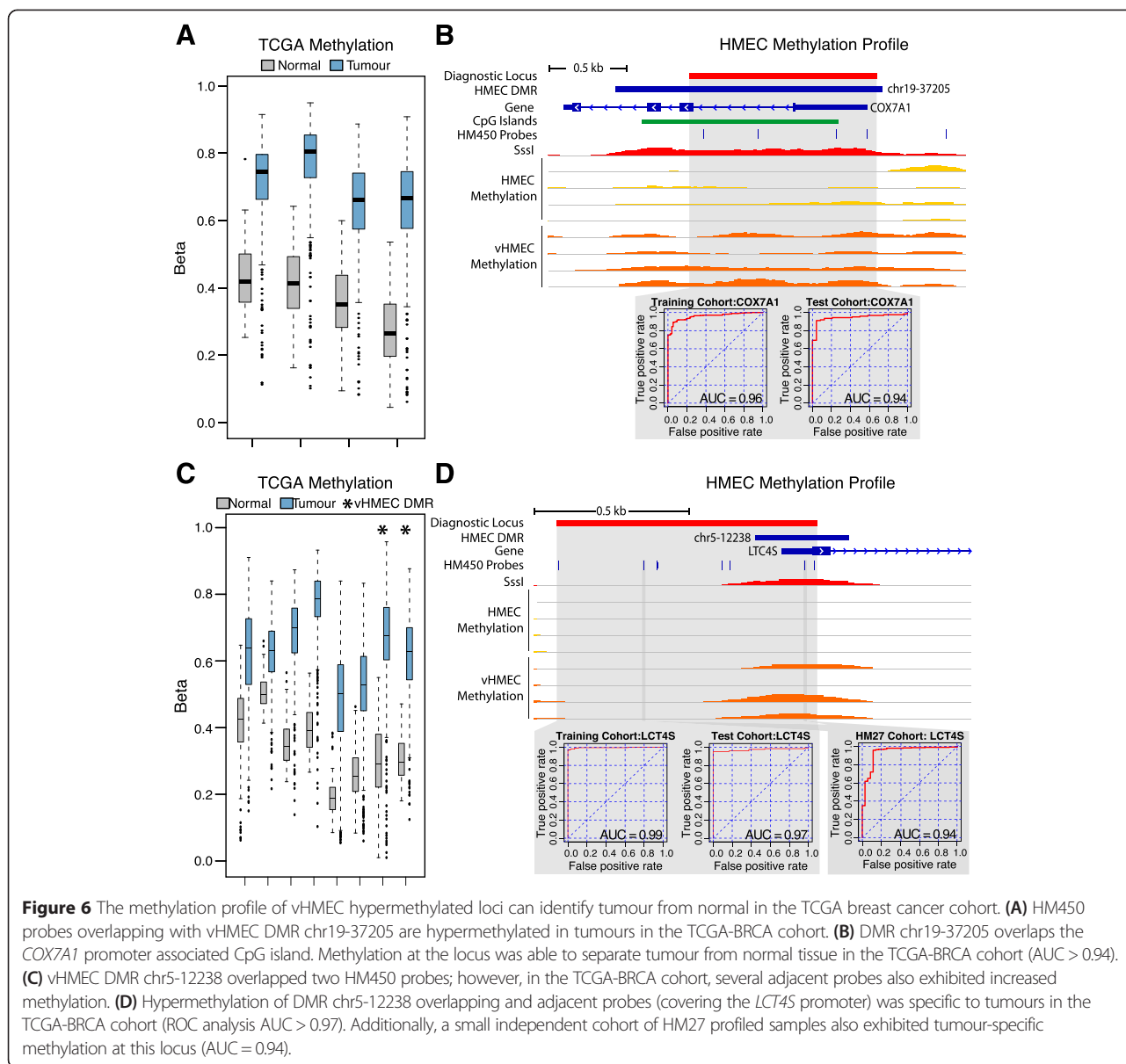
Next we tested the influence of promoter and TFBS hypermethylation on the biology of vHMEC. A total of 272 hypermethylated DMRs were associated with the promoters of 161 genes. We used these genes to perform Gene Set Enrichment Analysis (GSEA) in the MSigDB C2 curated gene set database [33]. Promoter hypermethylated genes in vHMEC were enriched in 48 gene sets that were predominately associated with cancer, polycomb regulation and p53 ($n = 20, 11$ and 3 for cancer, polycomb and p53 associated gene sets, respectively, Additional file 1: Figure S9A). Interestingly, several individual p53 and polycomb-related genes were identified as deregulated at the expression level by IPA, suggesting that p53 and polycomb deregulation involves multiple levels of genomic control (that is, both the transcriptome and methylome). At TFBS, hypermethylated DMRs were enriched for several TFs, most notable of which were the ligand activated and cancer-associated AHR, the development-associated PAX5 and cell cycle/cancer-related E2F family members (Additional file 1: Figure S9B). Overall, vHMEC DMRs impacted an independent set of genes with similar functions to those identified by gene expression analysis (that is, p53, polycomb, E2F and AHR regulation).

vHMEC differentially methylated regions are detected in breast cancer

Differential methylation that occurs very early during transformation may provide useful biomarkers to detect breast cancer. To determine if the early DMRs we identified in vHMECs were present in breast cancer, we analysed Infinium HumanMethylation450 BeadChip® (HM450) DNA methylation data obtained from TCGA-BRCA cohort (see 'Metho7ds'). Given the basal-like phenotype of vHMEC, we performed this analysis using all tumours and then separately with just the basal-like subtype. The 813 hypermethylated vHMEC DMRs overlapped with 2,430 HM450 probes, of which 1,560 or 961 were hypermethylated in all tumours or basal-like tumours, respectively (adjusted P values < 0.05 , Additional file 1: Figure S10A). This high frequency of hypermethylated probes in

vHMEC DMRs represents an approximately twofold enrichment of hypermethylation over what is expected to occur at random ($P < 0.05$). The enrichment of vHMEC DMRs for TCGA-BRCA hypermethylation suggests that these regions may be able to distinguish cancer from normal breast cells in both a general and basal-like specific manner. Therefore, we performed a receiver operating characteristic (ROC) analysis on TCGA_BRCA methylation to identify regions of potential diagnostic significance across all breast cancer subtypes. From the 813 vHMEC DMRs, 38 could sensitively and specifically separate tumours from normal in the training cohort (Additional file 2: Table S5). In the test cohort, all 38 regions exhibited an area under the curve (AUC) of > 0.91 , indicating the analysis is robust. Further validation was performed for a subset of ten loci that also overlapped probes from the Infinium HumanMethylation27 BeadChip® (HM27) DNA methylation platform. In the HM27 cohort, these regions exhibited an AUC of between 0.76 and 0.98 (Additional file 2: Table S5), confirming their diagnostic capacity in an independent dataset. Examples of diagnostic DMRs, associated with *LCT4S* (promoter), *COX7A1* (promoter), *TRIM62* (intronic) and *EPHB3* (promoter), are shown in Figure 6 and Additional file 1: Figure S11.

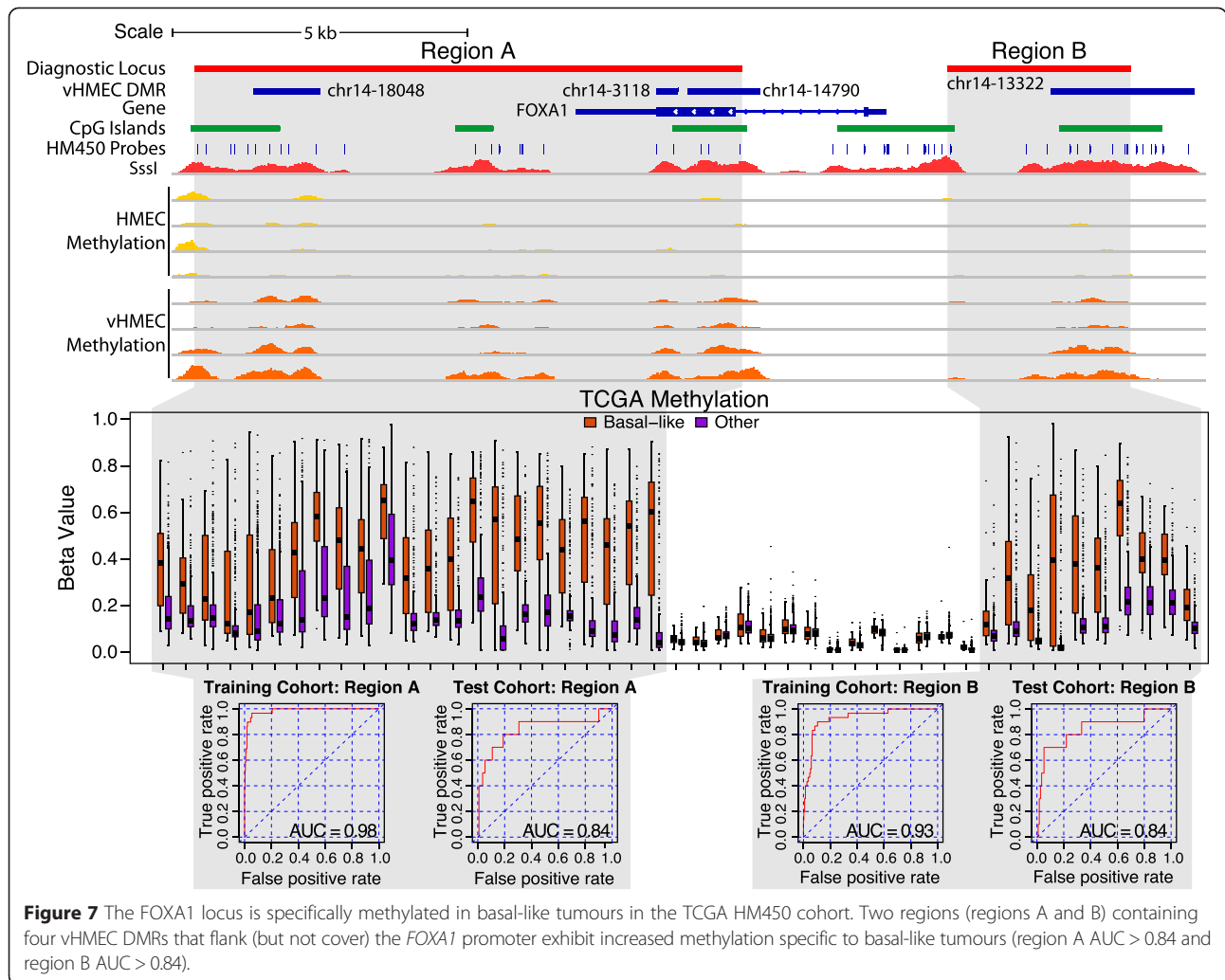
Given the early nature of methylation changes in vHMEC, it is possible that regions of hypermethylation are more extensive in cancer due to 'spread' of aberrant DNA methylation. Two candidate diagnostic vHMEC DMRs, associated with the genes *LCT4S* and *EPHB3*, exhibited a high AUC despite only overlapping two and three HM450 probes, respectively. In TCGA-BRCA cohort, several probes immediately adjacent to these regions also exhibited hypermethylation in cancer when compared to normal (Figure 6C and Additional file 1: Figure S11C) and could therefore improve the cancer specificity of these loci. With these additional probes included, ROC analysis revealed an increased AUC for both genes (Figure 6D AUC = 0.99 and Additional file 1: Figure S11D AUC = 0.99), supporting the possibility of methylation 'spread' in cancer. We also asked if DMRs were associated with altered gene expression in the same TCGA breast cancer samples. The correlation between methylation and expression was variable for each of the genes ($R^2 \leq 0.11, 0.3, 0.23$ and 0.6 for *LTC4S*, *TRIM62*, *EPHB3* and *COX7A1*, respectively); however, all diagnostic loci overlapped with the region where the correlation between methylation and expression was highest (Additional file 1: Figure S12), indicating that methylation at these loci may impact gene expression. Additionally, the DMRs in *EPHB3* and *TRIM62* overlapped with the binding sites of transcription factors identified as deregulated by IPA (Additional file 2: Table S6). Notably, the *EPHB3* methylated region, contained p53,



E2F and AHR sites and the TRIM62 methylated region contained p53 and PAX5 sites, suggesting that these specific loci may be important in epigenetic deregulation of these genes.

In addition to general cancer diagnosis, vHMEC DMRs may be useful in separating basal-like tumours from the other subtypes. To this end, we repeated the ROC analysis of vHMEC DMRs to determine if any could separate basal-like tumours from all other tumour subtypes. Given the lower levels of methylation in basal-like tumours when compared to other subtypes (Additional file 1: Figure S10A), to reduce the identification of hypomethylated regions, we only analysed vHMEC DMRs that overlapped one or more HM450 probe that was hypermethylated in basal-like tumours (313 vHMEC

DMRs overlapping 1,750 HM450 probes). Using an AUC cut off of >0.7, 14 candidate loci could separate basal-like from other tumours (Additional file 2: Table S7) although only four exhibited hypermethylation in basal-like when compared to other subtypes. For the four basal-like hypermethylated regions, two were associated with *FOXA1* (vHMEC DMRs chr14-3118 and chr14-14790), one with *LZTS1* (chr8-12256) and one within *CIRBP* (chr19-47460). We again inspected the regions for spread of hypermethylation into probes adjacent to the vHMEC DMR, and for the *FOXA1* and *LZTS1* regions, multiple adjacent probes did exhibit hypermethylation (Figure 7 and Additional file 1: Figure S13A, respectively) whereas *CIRBP* did not (Additional file 1: Figure S13B). *FOXA1* exhibited the most widespread methylation in basal-like



tumours, with hypermethylation encompassing two separate regions; a very large region covering much of the FOXA1 gene body and ending approximately 6 kb downstream of the transcription termination site (region A) and a shorter region approximately 2 kb upstream of the transcription start site (region B). ROC analysis revealed an AUC of 0.98 and 0.84 in the training and testing cohorts, respectively, for region A and 0.93 and 0.84 for region B (Figure 7). Similarly for LZTS1 (Additional file 1: Figure S13A), with adjacent probes included, methylation was highly specific to basal-like tumours with AUC of 0.95 and 0.86 in the training and test cohorts, respectively. Hypermethylation of the LZTS1 transcription start site did not exhibit an association with expression in TCGA breast tumours; however, FOXA1 ($R^2 \leq 0.75$ and 0.67 for regions A and B, respectively) and CIRBP ($R^2 \leq 0.212$) both exhibited loss of expression with increased methylation (Additional file 1: Figure S14). Additionally, the FOXA1 region A and CIRBP regions contained binding sites for p53 and AHR and, p53 and PAX5, respectively,

all of which have been identified as deregulated in prior analyses (Additional file 2: Table S6).

Discussion

The progression from a normal cell to a cancer cell involves genome-wide epigenetic deregulation and altered gene expression; however, the initial epigenetic lesions that occur during early carcinogenesis remain largely unknown. Here we utilised a model that recapitulates pre-malignant phase of basal-like breast carcinogenesis [34] and found significant alterations in DNA methylation and chromatin modification patterns that were associated with dysregulation of cancer-associated genes and pathways; specifically, genes and loci targeted by, EZH2/polycomb, MYC, AHR and the p53 pathway. Notably, these patterns of epigenetic deregulation were stable over an extended period of cell culture and occurred similarly in four independent HMEC lines, suggesting an orchestrated process of alteration in DNA methylation and chromatin modification.

The hypermethylation of key transcription factor binding sites occurs in vHMECs and may indicate a putative mechanism to establish genome-wide differential methylation in cancer. For example, activation of *AHR* may lead to a coordinated change in the epigenetic landscape by triggering the hypermethylation of its binding sites. *AHR* is a ligand activated transcription factor that has previously been implicated as having roles in differentiation and cancer [35-40] and we show that *AHR* is active in vHMEC. In MCF7 cells, *AHR* activation has been linked to the silencing of *BRCA1* by recruitment of DNA methyltransferases to the *BRCA1* promoter region [41-43]. Additionally, in keratinocytes *AHR* is reported to induce the silencing and hypermethylation of the *CDKN2A* (p16) tumour suppressor gene and the inhibition of p53 signalling [44]. Notably, the silencing of p16 and subsequent promoter hypermethylation is an essential component of the process of generating vHMEC [15,18]. Together, these findings may suggest that activation of *AHR* could contribute to the extensive epigenomic change observed in vHMEC (including hypermethylation of *CDKN2A*), with potentially far ranging effects on the biology of carcinogenesis, such as the inhibition of p53 in vHMEC [44]. The observed inhibition of p53 signalling in vHMEC is intriguing, given that vHMEC are reported to express high levels of wild type p53 protein [45]. However, instead of modulation or alteration of the p53 protein/gene, we observed DNA hypermethylation of its target genes and loci. Additionally, we identified loss of expression of the p53-interacting/tumour-suppressing miRNA *mir143/145* cluster. The *mir143/145* cluster silenced in vHMEC is reported to act downstream of p53 by promoting cell death [46], further suggesting that p53 activity is impeded via silencing of its targets and inhibition of downstream signalling. Similar to p53, the *PAX5* transcription factor also exhibited frequent hypermethylation of its target sites in vHMEC. Interestingly, *PAX5* itself is reported as being frequently methylated in breast cancer [47,48] and its overexpression in cell lines is reported to reduce the cancer phenotype by promoting normal epithelial characteristics [49]. In vHMEC, *PAX5* remains unmethylated, suggesting it is the disruption of *PAX5* targets rather than of the regulator itself that may facilitate carcinogenesis.

Interestingly, DMRs identified in vHMEC were extensively methylated in breast cancer, with hypermethylation frequently found to have extended into neighbouring regions. With their stability and very early occurrence, vHMEC hypermethylated loci could therefore serve as excellent biomarkers for the initial detection of cancer. Indeed, we found that a subset of the vHMEC DMRs exhibit a strong cancer-specific DNA methylation signature and additionally we determined a set of loci that could specifically identify basal-like tumours. Primarily, these loci were frequently associated with genes that have known

roles in cancer, for example, *FOXA1*, *LZTS1* and *CIRBP*. Increased methylation at two regions surrounding *FOXA1* also exhibited a strong correlation with decreased expression. Notably hypermethylation of these two regions has previously been reported in leukaemia [50], colorectal cancer [51] and pancreatic cancer [52], indicating that they are potential regulatory loci for *FOXA1*. Given that basal-like cancers and HMEC express low levels of *FOXA1*, hypermethylation at these regions may play a role in reducing the transcriptional plasticity of the *FOXA1* locus. Interestingly, several candidate biomarker loci also overlapped with predicted binding sites of the p53, *AHR* and *PAX5* transcription factors, further indicating a key role for hypermethylation at these transcription factor binding sites during carcinogenesis.

Conclusions

The HMEC model system provides an excellent tool to study the biology of early breast carcinogenesis with the potential to deliver clinically relevant biomarkers for the detection and classification of breast cancer. Additionally, our results indicate that the initial cancer driving events involve genome-wide epigenetic silencing of transcription factor targets that potentially disrupt gene networks. Finally, we show that in pre-malignancy dysregulation of the epigenome is extensive and can occur across large domains with wide-ranging impact on the process of carcinogenesis.

Methods

Tissue culture

HMEC lines (Table 1) were generated from tissue removed from healthy women during breast reduction mammoplasty with informed consent and approval from the Sydney Adventist Hospital Human Research Ethics Committee and prepared according to the protocol set out in Stampfer *et al.* [53]. HMEC lines were expanded in MBD170 serum-free basal medium (Life Technologies, Carlsbad CA, USA) until proliferation slowed and the cells entered the selection phase (passage 4 to 7). Selection phase HMEC were maintained with twice-weekly media changes until large colonies (1 to 2 cm) of growth were observed (2 to 4 weeks). Growing cells were trypsinised

Table 1 HMEC donor details

Donor	Age	Passage		
		HMEC	Early vHMEC	Late vHMEC
Bre12	23	3* (3 [#])	10* (7 [#])	16* (14 [#])
Bre38	29	5* (3 [#])	11* (7 [#])	14* (26 [#])
Bre67	23	2* (2 [#])	7* (7 [#])	14* (14 [#])
Bre98	30	2* (2 [#])	8* (7 [#])	14* (14 [#])

*Passage used for expression profiling. [#]Passage used for methylation profiling. HMEC, human mammary epithelial cells; vHMEC, variant HMEC.

and transferred to a new culture vessel, generating the vHMEC lines. At the time of surgery, all donors were aged between 23 and 43 years of age.

MBDCap-seq

Methylated DNA Binding Domain protein Capture followed by sequencing (MBDCap-seq) utilised 1 µg genomic DNA sonicated to mean fragment size of approximately 500 bp using a Bioruptor® water bath sonicator (Diagenode, Denville NJ, USA). Methylated DNA was captured using the Methylminer® Methylated DNA Enrichment Kit (Life Technologies, Carlsbad CA, USA) according to the manufacturer's protocol and total captured DNA was eluted with one high salt wash (2,000 mM). Sequencing was performed on 10 ng MBD captured DNA on the Illumina GAIIIX sequencing platform (Illumina, San Diego, CA, USA) at the Ramaciotti Centre for Gene Function Analysis (RCGFA) or The Beijing Genomics Institute (BGI) generating single end 36 bp reads at RCGFA and 50 bp at BGI.

Short sequence reads from Illumina sequencing were mapped to the human genome build HG18 with Bowtie [54]. Reads mapping to multiple locations, containing sequencing errors or duplicated reads were discarded from further analysis. The number of reads mapping at any given locus depends upon CpG density [55]. Therefore, to identify all genomic loci assayable by MBDCap-seq, an SssI treated 100% methylated sample (CpGenome Universally methylated DNA, Merck Millipore, Billerica MA, USA) was used to identify regions of the genome attracting a sufficient number of reads for the reliable detection of DNA methylation status. Specifically, the findPeaks function of the Homer peak-calling suite of programs [56] was applied to the SssI material (parameter settings of style = histone, size = 300, minDist = 300, tagThreshold = 18) which identified 230,655 assayable loci covering approximately 116 Mb across the genome. The edgeR Bioconductor package [57-59] was then used to model the distribution of reads in assayable loci to detect regions with statistically significant changes in read density.

Clonal bisulphite sequencing

Bisulphite conversion of DNA was performed according to a previously published protocol [60]. Briefly, genomic DNA (1 µg) was incubated in cell lysis solution (100 ng/µl tRNA, 280 ng/µl Proteinase K, 1% SDS) at 37°C for 1 h to ensure complete protein digestion. NaOH (Sigma-Aldrich, St Louis, MO, USA) was added to a final concentration 0.3 M and DNA was denatured at 90°C for 5 min, then rapidly cooled on ice. The denatured sample was combined with a saturated solution of sodium metabisulphite (pH 5) and 10 mM Quinol (0.5 mM final Quinol concentration) and incubated at 55°C for 16 h. Bisulphite treated DNA was purified (Wizard® DNA Clean-Up System,

Promega, Madison, WI, USA) according to the manufacturer's protocol and desulphonated by the addition of NaOH to a final concentration of 0.3 M followed by incubation at 37°C for 15 min then purified by ethanol precipitation.

PCR amplification of bisulphite converted DNA was performed using Platinum Taq Polymerase (Life Technologies) according to the manufacturer's protocol using the primers listed in Additional file 2: Table S8 under nested or fully nested conditions. PCR products were cloned into the pGEM T-easy cloning kit (Promega) and sequenced at Australian Cancer Research Foundation facility, located within the Garvan Institute of Medical Research, Sydney, Australia.

Taqman low-density miRNA array

High-quality RNA was converted to cDNA using the Megaplex Pools® for microRNA expression analysis cDNA synthesis kit (Life Technologies, Carlsbad CA, USA) according to the manufacturer's protocol, without pre-amplification. The total cDNA reaction was loaded onto a TLDA and qPCR was performed on an AbiPrism 7900HT (Life Technologies, Carlsbad CA, USA) sequence detection system. TLDA qPCR data was analysed using the HTqPCR R package [61] by the $\Delta\Delta C_t$ method (normalised to the included control genes). Genes either lowly or invariably expressed across all patients and time points were excluded, leaving 116 out of 379 miRNA for further analysis. Predicted miRNA targets were identified using the microrna.org database [62]. To remove spurious interactions, predictions were limited to those that had a mirSVR score of < -0.5.

Expression microarray

Gene expression profiling was performed by Affymetrix GeneChip® Human Gene 1.0 ST arrays (at one time point in the HMEC growth phase and two time points in the vHMEC phase referred to as early and late vHMEC, respectively), Taqman Low-Density miRNA arrays and RNA-seq.

RNA was extracted using TRIzol® reagent (Life Technologies, Carlsbad CA, USA) according to manufacturer's protocol and quality assessed on the Agilent Bioanalyzer RNA Nano chip (Agilent Technologies, Santa Clara CA, USA) to ensure an RNA integrity number (RIN) of >9. RNA labelling and hybridisation to Affymetrix GeneChip® 1.0ST expression arrays was performed at the Ramaciotti Centre for Gene Function Analysis (UNSW, Sydney, Australia). Data analysis was performed using aroma.affymetrix [63]. Robust Multi-Array (RMA) normalisation to summarise gene-probe intensities and differentially expressed genes were identified using limma [64].

TCGA data acquisition and analysis

Processed RNA-Seq expression data (level 3) and raw HM450 methylation data (level 1) were obtained from the TCGA data portal in January 2012 and clinical annotation from the 2012 TCGA-BRCA cohort primary publication [65]. Raw methylation data was pre-processed and background normalised with Bioconductor minfi package using `preprocessIllumina (... , bg.correct = TRUE, normalize = 'controls', reference = 1)` command. HM450 probes exhibiting differential methylation were identified using limma. Comparisons were carried out between all tumours and normal or basal-like tumours and all normal samples. For ROC analysis, the database was split randomly into two cohorts (75% and 25% of all samples, respectively) to allow training and testing of the models. ROC was performed using ROCR. For tumour-specific regions, regions exhibiting an AUC of >0.95 in the test cohort (after training) were selected as having potential utility as a diagnostic marker. For basal-like specific methylation, regions with an AUC of >0.7 were taken as having potential clinical utility.

ChIP-seq

ChIP was performed according to the manufacturer's protocol (17–295, Millipore, Billerica MA, USA). Briefly, approximately 2×10^6 cells were fixed in culture medium plus 1% (*w/v*) formaldehyde for 10 min at 37°C. Fixed chromatin was sonicated to a mean fragment size of approximately 500 bp and immunoprecipitation performed for the chromatin marks H3K4me3, H3K27ac, H3K36me3 and H3K27me3 (see Additional file 2: Table S9 for antibody conditions). H3K4me3 profiling was performed on HMEC and early vHMEC from donors Bre67 and Bre98 and H3K27ac, H3K27me3 and H3K36me3 ChIP-seq was performed on donor Bre98 only. Illumina 36 bp single end sequencing was performed on 5 to 12 ng of ChIP DNA at either the Ramaciotti Centre for Gene Function Analysis or the Beijing Genomics Institute (see Additional file 2: Table S10 for sequencing summary statistics).

Short sequence reads from Illumina sequencing were aligned to the human genome build hg18 with bowtie [54] and regions of enrichment were identified using the Homer peak calling algorithm (Settings; H3K4me3: `-minTagThreshold 10 -size 600 -minDist 600 -F 0 -L 0 -C 0`, H3K27ac: `-region -size 2000 -minDist 4000 -localSize 1000000 -F 0 -L 0 -C 0`, H3K27me3 & H3K36me3; `-region -fdr 1e-04 -minTagThreshold 40 -size 2000 -minDist 4000 -localSize 1000000 -F 0 -L 0 -C 0`) [56]. Large regions of enrichment were broken into non-overlapping 1,000 bp tiles for the chromatin marks H3K27ac, H3K36me3 and H3K27me3 whereas, H3K4me3 exhibited smaller regions that did not need to be tiled. The density of sequence reads in regions of enrichment was assessed with edgeR [57–59] to establish regions of statistically significant

chromatin marking in each time. Where regions had to be tiled, adjacent statistically significant regions were combined into a single region and an aggregate statistic calculated using Fisher's exact test.

ChIP-chip profiling

ChIP-chip profiling for H3K9ac was performed on donors Bre12 and Bre38, and H3K27me3 ChIP-chip was performed on donors Bre12, Bre38, Bre67 and Bre98.

ChIP and input DNA for array hybridisation was amplified to ensure sufficient yield prior to fragmentation and labelling (recommended 7.5 µg per ChIP-chip experiment). H3K27me3 ChIP DNA was amplified once with the GenomePlex® Complete Whole Genome Amplification (WGA) Kit (WGA2, Sigma-Aldrich, St. Louis, MO, USA). WGA2 amplification gave insufficient yield when performed on H3K9ac ChIP DNA. Therefore, H3K9ac ChIP DNA was amplified once with WGA2 and reamplified with the GenomePlex® WGA Reamplification Kit (WGA3, Sigma-Aldrich, St. Louis, MO, USA).

WGA material was cleaned up with the GeneChip® Sample Cleanup Module (Affymetrix, Santa Clara, CA, USA) and 7.5 µg was then fragmented and labelled for array hybridisation with the GeneChip® WT Double-Stranded Target Assay (Affymetrix, Santa Clara, CA, USA) according to the manufacturer's protocols. Fragmented and labelled WGA ChIP and input DNA was hybridised to GeneChip® Human Promoter 1.0R arrays (Affymetrix, Santa Clara, CA, USA) at the RCGFA. Low level analysis and Model-based Analysis of Tiling arrays (MAT) normalisation of ChIP-chip data was performed using `aroma.affymetrix` [63]. Array signal across all promoters on the array was determined using Repitools [55,66,67]. ChIP data was then summarised to a *t*-statistic for a region of ±2 kb of all transcription start sites (TSS) with the function `blocksStats` [67]. Regions of LRER were identified by the methods set out in Bert *et al.* [30].

Availability of supporting data

All sequencing and array data supporting the results of this article are available through Gene Expression Omnibus super series GSE58882.

Additional files

Additional file 1: Supplementary Figures. Figures S1 to S14 and associated figure legends.

Additional file 2: Supplementary Tables. **Table S1** IPA TF deregulation predictions. **Table S2.** Predicted targets of deregulated miRNA (microrna.org). **Table S3.** TarBase validated targets of deregulated miRNA. **Table S4.** Coordinates and summary of vHMEC LRER. **Table S5.** Tumour specific vHMEC DMRs. **Table S6.** TFBS associated with candidate diagnostic DMRs. **Table S7.** Basal-like tumour specific vHMEC DMRs. **Table S8.** Bisulphite converted primer sequences, conditions and coordinates (HG18). **Table S9.** Antibodies used for ChIP-seq. **Table S10.** ChIP-seq validation primer sequences, conditions and coordinates (HG18).

Abbreviations

AUC: area under the curve; ChIP: chromatin immunoprecipitation; ChIP-seq: ChIP followed by next generation sequencing; DMR: differentially methylated region; GSEA: gene set enrichment analysis; HM450: Infinium Human Methylation 450 beadchip; HMEC: human mammary epithelial cells; IPA: Ingenuity Pathway Analysis; LREA: long-range epigenetic activation; LRER: long-range epigenetic regulation; LRES: long-range epigenetic silencing; MBDCap-seq: Methylated DNA Binding Domain protein Methylated DNA Capture followed by next generation sequencing; MDS: multi-dimensional scaling; ROC: receiver operator characteristics; TCGA: The Cancer Genome Atlas; TCGA-BRCA: The Cancer Genome Atlas Invasive Breast Carcinoma Cohort; TF: transcription factor; TFBS: transcription factor binding site; TLDA: Taqman Low-Density Array; vHMEC: variant human mammary epithelial cells.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

WJL performed all experiments, performed some bioinformatic analysis and writing of the manuscript. EZ designed and performed the bioinformatic analysis of MBDCap-seq and ChIP-seq data and MDR performed the analysis of the gene expression array data. CS and AS contributed to manuscript preparation. RAH contributed to initial experimentation. RRR and LIH provided organoids/HMECs and associated protocols. SJC conceived of the study design and writing of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

All work was performed using funding from the following grants, scholarships and fellowships; National Breast Cancer Foundation of Australia Postgraduate Research Scholarship #DS-09-06 (WJL, <http://www.nbcf.org.au/>), National Breast Cancer Foundation of Australia Collaborative Grant entitled 'Novel strategies for prediction and control of advanced breast cancer via nanoscaled epigenetic-based biosensors' (SJL, <http://www.nbcf.org.au/>), National Health and Medical Research Council Project grant #1011447 (SJL <https://www.nhmrc.gov.au/>), National Health and Medical Research Council Fellowship #1063559 (SJL <https://www.nhmrc.gov.au/>), National Health and Medical Research Council Project Grant #1070881 (LIH, RRR <https://www.nhmrc.gov.au/>), New South Wales Cancer Council Program Grant #PG11-08 (RRR <http://www.cancercouncil.com.au/>).

Author details

¹Epigenetic Research Laboratory, Genomics and Epigenetic Division, The Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, NSW 2010, Australia. ²St. Vincent's Clinical School, Faculty of Medicine, University of New South Wales, Level 5 deLacy Building, St Vincent's Hospital, Victoria Street, Darlinghurst, NSW 2010, Australia. ³Swiss Institute of Bioinformatics, University of Zurich, Zurich, and Institute of Molecular Life Sciences, University of Zurich, Winterthurerstrasse 190, Zurich CH-8057, Switzerland. ⁴Sydney West Cancer Trials Centre, Crown Princess Mary Cancer Centre Westmead, Westmead Hospital, Hawkesbury Road, Westmead, NSW 2145, Australia. ⁵Cancer Research Unit, Children's Medical Research Institute, 2145 Hawkesbury Road, Westmead, NSW 2145, Australia. ⁶Sydney Medical School, University of Sydney, Fisher Road, Sydney, NSW 2006, Australia.

Received: 24 November 2014 Accepted: 16 April 2015

Published online: 01 May 2015

References

- Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer*. 2011;11:726–34.
- Suva ML, Riggi N, Bernstein BE. Epigenetic reprogramming in cancer. *Science*. 2013;339:1567–70.
- Daskalos A, Nikolaidis G, Xinarianos G, Savvari P, Cassidy A, Zakopoulou R, et al. Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer. *Int J Cancer*. 2009;124:81–7.
- Narayan A, Ji W, Zhang XY, Marrogi A, Graff JR, Baylin SB, et al. Hypomethylation of pericentromeric DNA in breast adenocarcinomas. *Int J Cancer*. 1998;77:833–8.
- Prada D, Gonzalez R, Sanchez L, Castro C, Fabian E, Herrera LA. Satellite 2 demethylation induced by 5-azacytidine is associated with missegregation of chromosomes 1 and 16 in human somatic cells. *Mutat Res*. 2012;729:100–5.
- Jackson M, Krassowska A, Gilbert N, Chevassut T, Forrester L, Ansell J, et al. Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. *Mol Cell Biol*. 2004;24:8862–71.
- Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98:10869–74.
- Easwaran H, Johnstone SE, Van Neste L, Ohm J, Mosbrugger T, Wang Q, et al. A DNA hypermethylation module for the stem/progenitor cell signature of cancer. *Genome Res*. 2012;22:837–49.
- Holm K, Hegardt C, Staaf J, Vallon-Christersson J, Jonsson G, Olsson H, et al. Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. *Breast Cancer Res*. 2010;12:R36.
- Locke WJ, Clark SJ. Epigenome remodelling in breast cancer: insights from an early in vitro model of carcinogenesis. *Breast Cancer Res*. 2012;14:215.
- Stone A, Cowley MJ, Valdes-Mora F, McCloy RA, Sergio CM, Gallego-Ortega D, et al. BCL-2 hypermethylation is a potential biomarker of sensitivity to antimetabolic chemotherapy in endocrine-resistant breast cancer. *Mol Cancer Ther*. 2013;12:1874–85.
- Stone A, Valdes-Mora F, Gee JM, Farrow L, McClelland RA, Fiegl H, et al. Tamoxifen-induced epigenetic silencing of oestrogen-regulated genes in anti-hormone resistant breast cancer. *PLoS One*. 2012;7, e40466.
- Hinshelwood RA, Clark SJ. Breast cancer epigenetics: normal human mammary epithelial cells as a model system. *J Mol Med*. 2008;86:1315–28.
- Hammond SL, Ham RG, Stampfer MR. Serum-free growth of human mammary epithelial cells: rapid clonal growth in defined medium and extended serial passage with pituitary extract. *Proc Natl Acad Sci U S A*. 1984;81:5435–9.
- Huschtscha LI, Noble JR, Neumann AA, Moy EL, Barry P, Melki JR, et al. Loss of p16INK4 expression by methylation is associated with lifespan extension of human mammary epithelial cells. *Cancer Res*. 1998;58:3508–12.
- Tlsty TD, Crawford YG, Holst CR, Fordyce CA, Zhang J, McDermott K, et al. Genetic and epigenetic changes in mammary epithelial cells may mimic early events in carcinogenesis. *J Mammary Gland Biol Neoplasia*. 2004;9:263–74.
- Brenner AJ, Stampfer MR, Aldaz CM. Increased p16 expression with first senescence arrest in human mammary epithelial cells and extended growth capacity with p16 inactivation. *Oncogene*. 1998;17:199–205.
- Foster SA, Wong DJ, Barrett MT, Galloway DA. Inactivation of p16 in human mammary epithelial cells by CpG island methylation. *Mol Cell Biol*. 1998;18:1793–801.
- Novak P, Jensen TJ, Garbe JC, Stampfer MR, Futscher BW. Stepwise DNA methylation changes are linked to escape from defined proliferation barriers and mammary epithelial cell immortalization. *Cancer Res*. 2009;69:5251–8.
- Hinshelwood RA, Huschtscha LI, Melki J, Stizaker C, Abdipranoto A, Vissel B, et al. Concordant epigenetic silencing of transforming growth factor-beta signaling pathway genes occurs early in breast carcinogenesis. *Cancer Res*. 2007;67:11517–27.
- Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27:1160–7.
- Sorlie T, Borga E, Myhre S, Vollan HK, Russnes H, Zhao X, et al. The importance of gene-centring microarray data. *Lancet Oncol*. 2010;11:719–20. author reply 720–711.
- Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet*. 2005;365:671–9.
- Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci U S A*. 2005;102:13550–5.
- Tian Y, Ke S, Chen M, Sheng T. Interactions between the aryl hydrocarbon receptor and P-TEFb. Sequential recruitment of transcription factors and differential phosphorylation of C-terminal domain of RNA polymerase II at cyp1a1 promoter. *J Biol Chem*. 2003;278:44041–8.
- Kim J, Woo AJ, Chu J, Snow JW, Fujiwara Y, Kim CG, et al. A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. *Cell*. 2010;143:313–24.

27. Betel D, Wilson M, Gabow A, Marks DS, Sander C. The microRNA.org resource: targets and expression. *Nucleic Acids Res.* 2008;36:D149–153.
28. Vergoulis T, Vlachos IS, Alexiou P, Georgakilas G, Maragkakis M, Reczko M, et al. TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. *Nucleic Acids Res.* 2012;40:D222–229.
29. Coolen MW, Stirzaker C, Song JZ, Statham AL, Kassir Z, Moreno CS, et al. Consolidation of the cancer genome into domains of repressive chromatin by long-range epigenetic silencing (LRES) reduces transcriptional plasticity. *Nat Cell Biol.* 2010;12:235–46.
30. Bert SA, Robinson MD, Strbenac D, Statham AL, Song JZ, Hulf T, et al. Regional activation of the cancer genome by long-range epigenetic remodeling. *Cancer Cell.* 2013;23:9–22.
31. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The human genome browser at UCSC. *Genome Res.* 2002;12:996–1006.
32. Meyer LR, Zweig AS, Hinrichs AS, Karolchik D, Kuhn RM, Wong M, et al. The UCSC genome browser database: extensions and updates 2013. *Nucleic Acids Res.* 2013;41:D64–69.
33. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Biogeosciences.* 2011;7:1739–40.
34. Dumont N, Crawford YG, Sigaroudinia M, Nagrani SS, Wilson MB, Buehring GC, et al. Human mammary cancer progression model recapitulates methylation events associated with breast premalignancy. *Breast Cancer Res.* 2009;11:R87.
35. Wang C, Xu CX, Bu Y, Bottum KM, Tischkau SA. Beta-naphthoflavone (DB06732) mediates estrogen receptor-positive breast cancer cell cycle arrest through AhR-dependent regulation of PI3K/AKT and MAPK/ERK signaling. *Carcinogenesis.* 2014;35:703–13.
36. Saito R, Miki Y, Hata S, Takagi K, Iida S, Oba Y, et al. Aryl hydrocarbon receptor in breast cancer—a newly defined prognostic marker. *Hormones Cancer.* 2014;5:11–21.
37. Prigent L, Robineau M, Jouneau S, Morzadec C, Louarn L, Vernhet L, Fardel O, Sparfel L: The aryl hydrocarbon receptor is functionally upregulated early in the course of human T-cell activation. *European J Immunol* 2014.
38. O'Donnell EF, Koch DC, Bisson WH, Jang HS, Kolluri SK. The aryl hydrocarbon receptor mediates raloxifene-induced apoptosis in estrogen receptor-negative hepatoma and breast cancer cells. *Cell Death Disease.* 2014;5:e1038.
39. Wang K, Li Y, Jiang YZ, Dai CF, Patankar MS, Song JS, et al. An endogenous aryl hydrocarbon receptor ligand inhibits proliferation and migration of human ovarian cancer cells. *Cancer Lett.* 2013;340(1):63–71.
40. Hayashi S, Okabe-Kado J, Honma Y, Kawajiri K. Expression of Ah receptor (TCDD receptor) during human monocytic differentiation. *Carcinogenesis.* 1995;16:1403–9.
41. Papoutsis AJ, Selmin OI, Borg JL, Romagnolo DF. Gestational exposure to the AhR agonist 2,3,7,8-tetrachlorodibenzo-p-dioxin induces BRCA-1 promoter hypermethylation and reduces BRCA-1 expression in mammary tissue of rat offspring: Preventive effects of resveratrol. *Molecul Carcinogenesis.* 2013;54(4):261–9.
42. Frauenstein K, Sydlik U, Tigges J, Majora M, Wiek C, Hanenberg H, et al. Evidence for a novel anti-apoptotic pathway in human keratinocytes involving the aryl hydrocarbon receptor, E2F1, and checkpoint kinase 1. *Cell Death Differ.* 2013;20(10):63–71.
43. Papoutsis AJ, Borg JL, Selmin OI, Romagnolo DF. BRCA-1 promoter hypermethylation and silencing induced by the aromatic hydrocarbon receptor-ligand TCDD are prevented by resveratrol in MCF-7 cells. *J Nutr Biochem.* 2012;23:1324–32.
44. Ray SS, Swanson HI. Dioxin-induced immortalization of normal human keratinocytes and silencing of p53 and p16INK4a. *J Biol Chem.* 2004;279:27187–93.
45. Huschtscha LI, Moore JD, Noble JR, Campbell HG, Royds JA, Braithwaite AW, et al. Normal human mammary epithelial cells proliferate rapidly in the presence of elevated levels of the tumor suppressors p53 and p21 (WAF1/CIP1). *J Cell Sci.* 2009;122:2989–95.
46. Spizzo R, Nicoloso MS, Lupini L, Lu Y, Fogarty J, Rossi S, et al. miR-145 participates with TP53 in a death-promoting regulatory loop and targets estrogen receptor-alpha in human breast cancer cells. *Cell Death Differ.* 2010;17:246–54.
47. Kornegoor R, Moelans CB, Verschuur-Maes AH, Hogenes MC, de Bruin PC, Oudejans JJ, et al. Promoter hypermethylation in male breast cancer: analysis by multiplex ligation-dependent probe amplification. *Breast Cancer Res.* 2012;14:R101.
48. Moelans CB, Verschuur-Maes AH, van Diest PJ. Frequent promoter hypermethylation of BRCA2, CDH13, MSH6, PAX5, PAX6 and WT1 in ductal carcinoma in situ and invasive breast cancer. *J Pathol.* 2011;225:222–31.
49. Vidal LJ, Perry JK, Vouyovitch CM, Pandey V, Brunet-Dunand SE, Mertani HC, et al. PAX5alpha enhances the epithelial behavior of human mammary carcinoma cells. *Mol Cancer Res.* 2010;8:444–56.
50. Pei L, Choi JH, Liu J, Lee EJ, McCarthy B, Wilson JM, et al. Genome-wide DNA methylation analysis reveals novel epigenetic changes in chronic lymphocytic leukemia. *Epigenetics.* 2012;7:567–78.
51. van Roon EH, Boot A, Dihal AA, Ernst RF, van Wezel T, Morreau H, et al. BRAF mutation-specific promoter methylation of FOX genes in colorectal cancer. *Clin Epigenet.* 2013;5:2.
52. Vincent A, Omura N, Hong SM, Jaffe A, Eshleman J, Goggins M. Genome-wide analysis of promoter methylation associated with gene expression profile in pancreatic adenocarcinoma. *Clin Cancer Res.* 2011;17:4341–54.
53. Stampfer M. Isolation and growth of human mammary epithelial cells. *J Tissue Cult Methods.* 1985;9:107–15.
54. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 2009;10:R25.
55. Robinson MD, Stirzaker C, Statham AL, Coolen MW, Song JZ, Nair SS, et al. Evaluation of affinity-based genome-wide DNA methylation data: effects of CpG density, amplification bias, and copy number variation. *Genome Res.* 2010;20:1719–29.
56. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell.* 2010;38:576–89.
57. Robinson MD, Smyth GK. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics.* 2008;9:321–32.
58. Robinson MD, Smyth GK. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics.* 2007;23:2881–7.
59. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics.* 2010;26:139–40.
60. Clark SJ, Statham A, Stirzaker C, Molloy PL, Frommer M. DNA methylation: bisulphite modification and analysis. *Nat Protoc.* 2006;1:2353–64.
61. Dvinge H, Bertone P. HTqPCR: high-throughput analysis and visualization of quantitative real-time PCR data in R. *Bioinformatics.* 2009;25:3325–6.
62. Betel D, Koppal A, Agius P, Sander C, Leslie C. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* 2010;11:R90.
63. Bengtsson H, Simpson K, Bullard J, Hansen K: Aroma.affymetrix: a generic framework in R for analyzing small to very large Affymetrix data sets in bounded memory. Tech Report #745, Department of Statistics, University of California, Berkeley, February 2008.
64. Smyth G: limma: linear models for microarray data bioinformatics and computational biology solutions using R and bioconductor. In: Edited by Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S: Springer New York; 2005: 397–420: Statistics for Biology and Health].
65. Cancer Genome Atlas N: Comprehensive molecular portraits of human breast tumours. *Nature* 2012, 490:61–70.
66. Robinson MD, Statham AL, Speed TP, Clark SJ. Protocol matters: which methylome are you actually studying? *Epigenomics.* 2010;2:587–98.
67. Statham AL, Strbenac D, Coolen MW, Stirzaker C, Clark SJ, Robinson MD. Repitools: an R package for the analysis of enrichment-based epigenomic data. *Bioinformatics.* 2010;26:1662–3.