DNA methylation profile of triple negative breast cancer-specific genes comparing lymph node positive patients to lymph node negative patients

Andrea Mathe\textsuperscript{1,2}, Michelle Wong-Brown\textsuperscript{1,2}, Warwick J. Locke\textsuperscript{3,4}, Clare Stirzaker\textsuperscript{3,4}, Stephen G. Braye\textsuperscript{5}, John F. Forbes\textsuperscript{2,6}, Susan J. Clark\textsuperscript{3,4}, Kelly A. Avery-Kiejda\textsuperscript{1,2,*} & Rodney J. Scott\textsuperscript{1,2,5,*}

Triple negative breast cancer (TNBC) is the most aggressive breast cancer subtype with no targeted treatment available. Our previous study identified 38 TNBC-specific genes with altered expression comparing tumour to normal samples. This study aimed to establish whether DNA methylation contributed to these expression changes in the same cohort as well as disease progression from primary breast tumour to lymph node metastasis associated with changes in the epigenome. We obtained DNA from 23 primary TNBC samples, 12 matched lymph node metastases, and 11 matched normal adjacent tissues and assayed for differential methylation profiles using Illumina HumanMethylation450 BeadChips. The results were validated in an independent cohort of 70 primary TNBC samples. The expression of 16/38 TNBC-specific genes was associated with alteration in DNA methylation. Novel methylation changes between primary tumours and lymph node metastases, as well as those associated with survival were identified. Altered methylation of 18 genes associated with lymph node metastasis were identified and validated. This study reveals the important role DNA methylation plays in altered gene expression of TNBC-specific genes and lymph node metastases. The novel insights into progression of TNBC to secondary disease may provide potential prognostic indicators for this hard-to-treat breast cancer subtype.

Breast cancer is responsible for the highest cancer incidence in women worldwide and its rates are increasing. Out of all breast cancer subtypes, triple negative breast cancer (TNBC) is the most aggressive, accounting for somewhere between 10 and 20\% of all diagnosed breast cancers\textsuperscript{1–3}. TNBC is recognised for its more frequent and rapid progression to metastasis, high rate of BRCA1 mutations\textsuperscript{4}, chromosome instability\textsuperscript{5} and is more commonly diagnosed in younger women (pre-menopausal) as well as women of African-American descent\textsuperscript{6}. It lacks the expression of estrogen and progesterone receptors (ER and PR) as well as the human epidermal growth factor receptor 2 (HER2). Treatment which target these receptors are ineffective in TNBC and the aggressive nature of TNBC drives the urgent need for new treatment targets.

The development and progression of cancer is due to the accumulation of multiple genetic and epigenetic change. DNA methylation is the most well-studied epigenetic change, which culminates in altered gene

\textsuperscript{1}Centre for Information Based Medicine, Hunter Medical Research Institute, NSW, 2305, Australia. \textsuperscript{2}Priority Research Centre for Cancer, School of Biomedical Sciences and Pharmacy, Faculty of Health, University of Newcastle, NSW, 2308, Australia. \textsuperscript{3}Epigenetics Research Program, Genomics and Epigenetics Division, Garvan Institute of Medical Research, Sydney, NSW, 2010, Australia. \textsuperscript{4}Faculty of Medicine, St Vincent’s Clinical School, UNSW, NSW 2052 & St Vincent’s Hospital, Sydney, NSW, 2010, Australia. \textsuperscript{5}Pathology North, John Hunter Hospital, Newcastle, NSW, 2305, Australia. \textsuperscript{6}Department of Surgical Oncology, Calvary Mater Newcastle Hospital, Australian New Zealand Breast Cancer Trials Group, and School of Medicine and Public Health, University of Newcastle, NSW, 2308, Australia. \textsuperscript{*}These authors contributed equally to this work. Correspondence and requests for materials should be addressed to R.J.S. (email: rodney.scott@newcastle.edu.au)
expression\(^7\text{,}^8\). DNA methylation is mediated by DNA-methyltransferase (DNMT), which leads to a conformational change of the nucleosomes, where histones are drawn tighter together thereby excluding transcription factors access such that genes cannot be transcribed and expressed. Generally, there is a global decrease of DNA methylation (hypomethylation) in cancer cells, which leads to increased genomic instability\(^9\). Nevertheless, an increase in DNA methylation (hypermethylation) has been identified at tumour suppressor genes in multiple cancers\(^10\).

DNA methylation loci are attractive candidates as biomarkers for TNBC as they are more stable than RNA or proteins and are readily detectable in tissue samples and blood\(^11\). In TNBC, the methylation pattern of a number of cancer-related genes has been analysed\(^12\). Additionally, methylation patterns have been used to differentiate breast cancer subtypes\(^13\). Sharma et al. discovered that the methylation of the BRCA1 promoter region is associated with worse overall survival and relapse-free survival in TNBC\(^14\). A recent study by Stirzaker et al. used whole genome DNA methylation analysis to identify a signature which divided TNBC into three prognostic subgroups and identified differentially methylated regions (DMRs) associated with overall survival\(^15\). However, there have been no studies to date that have addressed genome-wide methylation change during disease progression from the primary tumour to lymph node metastasis in TNBC.

In this study, we aimed to identify whether DNA methylation contributed to the altered expression of 38 genes we identified previously in TNBC\(^14\). By performing whole genome methylation analysis of 23 grade 3 primary invasive ductal carcinomas (IDC) and 11 matched normal adjacent tissues (NAT), we determined that 42\% of our TNBC specific genes had significantly altered methylation. Furthermore, by comparing IDC to NAT and IDC to 12 matched lymph node metastases (LN), we identified a set of DNA methylation aberrations associated with the progression of TNBC from primary tumour to LN metastases. We validated the methylation changes of 18 genes associated with LN metastasis with a regional DNA methylation analysis in an independent cohort. Additionally, we identified nine methylation probes, that have significantly altered methylation in LN samples to be associated with survival in TNBC. This is the first whole genome methylation study in TNBC including matched IDC, LN, and NAT samples. We were able correlate the findings of this study with the gene expression results of our previous report using the same sample cohort. A number of previously identified genes show differential methylation suggesting the potential functional relevance of these changes.

Results

Methylation profiles are altered in TNBC. We performed 450 K DNA methylation BeadChip array analysis (Illumina) in two independent TNBC cohorts. The study cohort contained 23 TNBC primary IDCs and the validation cohort contained 70 TNBC primary IDCs. All primary tumour samples were compared to three pooled NAT samples and one singular NAT (11 samples in total).

By comparing the methylation of the IDC samples to the NAT samples we identified 44,005 differentially methylated probes in the study cohort (52.2\% hypomethylated, 47.8\% hypermethylated) and 45,263 probes in the validation cohort (40.2\% hypomethylated, 59.8\% hypermethylated). We validated 29,612 probes (=67.29\%) in the independent validation cohort. Of the 29,612 validated probes, 41.4\% were hypomethylated and 58.6\% were hypermethylated. Within the validated probes there were 1,849 promoter-associated probes (52.8\% hypomethylated, 47.2\% hypermethylated) and 9,161 probes within enhancer regions (44.9\% hypomethylated, 55.1\% hypermethylated). In total, there were over 10,000 probes within enhancer or promoter regions that were significantly methylated. Performing a pathway enrichment analysis of these probes identified many important cancer pathways influenced by the respective genes. The pathways with the highest enrichment scores for hypomethylated genes were: Axon guidance, Rap1 signaling pathway, Platelet activation, Mucin type O-Glycan biosynthesis, and MAPK signaling pathway. The pathways with the highest enrichment scores for hypermethylated genes were: ECM-receptor interaction, Pathways in cancer, PI3K-Akt signaling pathway, focal adhesion, and signaling pathways regulating pluripotency of stem cells (Supplementary Table 1). We next identified differentially methylated regions (DMRs, a minimum of three significant consecutive probes). In the study cohort we identified 2,373 DMRs (10,082 probes) and in the validation cohort we identified 2,932 DMRs (12,938 probes), 72.62\% (1,756 DMRs/7,523 probes) were common in both cohorts (Supplementary Table 2). The results of this analysis are shown in Fig. 1.

Genes differentially expressed in TNBC compared to NAT are associated with altered methylation patterns. We have previously identified 66 genes to be differentially expressed in primary tumour samples compared to normal adjacent tissue in TNBC\(^16\). Therefore, in this study we aimed to identify the contribution of DNA methylation aberration to these gene expression changes. We determined that 26 of the 66 genes had significantly altered methylation of single loci (a total of 63 probes, 47.6\% hypomethylated, 52.4\% hypermethylated) in both the study and the validation cohorts (Supplementary Table 3). Of the significant probes there were nine within enhancers and four within promoter regions. Additionally, eight of the 66 genes had significantly altered regional methylation (40 probes, 30\% hypomethylated, 70\% hypermethylated) in both cohorts (Table 1). Of the eight genes with significantly altered regional methylation, one of these (EGRI) was significantly associated with overall survival in one probe (cg07336840). High DNA methylation of this probe was associated with better survival as shown in Fig. 2.

Relationship of methylation changes to altered expression of TNBC-specific genes. Previously we identified 38 genes that had altered expression in the TNBC subtype but not in other breast cancer subtypes, using two independent cohorts\(^16\). In the first cohort 28 TNBC-specific genes and in the second 14 TNBC specific genes were identified. There were four genes common to both cohorts (ANKRD30A, ANP32E, DSC2, and IL6ST). Here, we sought to investigate the DNA methylation changes of these 38 TNBC-specific genes.
We found that 16 of the 38 TNBC-specific genes were associated with differentially methylated probes in the study and validation cohorts (41 probes) (Table 2 and Fig. 3). A set of five genes (ANKRD30B, COL14A1, IGF1, IL6ST, MEG3) exhibited regional methylation differences (28 probes) in both cohorts. Three of which showed very strong methylation changes in both cohorts (>20% methylation change), these were ANKRD30B (7 hyper-methylated probes), COL14A1 (6 hyper-methylated probes), and MEG3 (8 hyper-methylated probes).

Of the four TNBC-specific genes that were common in both analyses in our previous study, there was one (IL6ST) that showed significantly altered methylation in 3 probes in both cohorts (Table 2). The methylation change for IL6ST can be classed as a DMR. However, no probes within these regions were significantly associated with survival (data not shown).

**Methylation changes associated with lymph node metastases.** In our previous study\(^1\) we identified 83 genes that showed altered expression both in primary tumours with LN metastases and in their matched LN metastasis, but were unaltered in lymph node negative tumours. This led to the rationale that the expression of these genes may be affected by DNA methylation and is a contributing factor to tumour progression. To interrogate this, we performed 450 K DNA methylation arrays on 12 LN metastasis samples from the same cohort used in our previous study\(^1\) and compared the DNA methylation from these samples to that of the NAT samples. A total of 51,563 probes had significantly altered methylation when comparing LN to NAT samples (46.7% hyper-methylated, 53.3% hypomethylated) of which, there were 2,350 significant DMRs (these DMRs contain 11,218 probes). Furthermore, 38 of the 83 LN metastasis-associated genes had significantly altered DNA methylation in 107 probes. Of these, 14 genes were present in DMRs (74 probes). The gene expression and DNA methylation changes of these 14 genes is shown in Supplementary Table 4.

It was not possible to validate the methylation changes in the validation cohort due to the lack of LN samples. However, we hypothesised that due to their altered expression in LN metastases, they would be associated with survival outcome. We performed survival analysis on the 70 tumour samples from our validation cohort using the methylation analysis of the 74 probes that comprised the 14 DMRs. Nine of the 74 probes were significantly associated with survival in the TNBC validation cohort (cg18108818, cg20464151, cg09933058, cg24173596, cg07336840, cg08500417, cg20066782, cg04028606, cg00185066) (Fig. 4). Eight of these probes (cg18108818, cg20464151, cg24173596, cg07336840, cg08500417, cg20066782, cg04028606, cg00185066) were associated with improved survival when they were highly methylated and one probe (cg09933058) was associated with worse survival when it was highly methylated.
Due to the lack of LN samples in the validation cohort, Methyl-Binding-Domain-Capture (MBDcap) sequencing in an independent cohort of 7 LN and 4 NAT samples was used to validate the direction of DNA methylation of the 38 genes that showed significantly altered methylation when comparing LN to NAT samples in the study cohort. The MBDcap sequencing data analysed 1 kb regions covering all 38 genes associated with LN metastasis (identified in our previous gene expression analyses) starting 2 kb upstream from the first transcription start site to the end of the gene. Of the 38 genes that showed significant altered methylation in 107 probes, 18 of these had significant regional DNA methylation changes in the same direction (hyper/hypomethylation) using MBDcap sequencing data. The results of this analysis can be seen in Table 3.

Table 1. DNA methylation of 8 validated DMRs comparing IDC vs NAT. These 8 DMRs are within 8 of 66 validated genes from previous study. [1] Every CpG island consist of N- and S Shores next to it, which are neighboured by N- and S-shelfs. First N-shelf, N-shore, CpG island, S-shore, and last S-shelf [2] Methylation difference between IDC and NAT shows % of methylation change/100.

| DMR # | Column ID | CHR | UCSC_REFGENE | ENHANCER (E)/PROMOTER(P) | RELATION_TO_UCSC_CPG_ISLAND[1] | UCSC_REFGENE_NAME | Gene expression | DNA methylation | DNA methylation |
|-------|-----------|-----|--------------|--------------------------|-------------------------------|-------------------|----------------|----------------|----------------|
|       |           |     |              |                          |                               |                   | Fold change    | p-value         | p-value         |
|       |           |     |              |                          |                               |                   |                | (IDC vs NAT)   | (IDC vs NAT)   |
|       |           |     |              |                          |                               |                   |                | Methylation    | Methylation    |
|       |           |     |              |                          |                               |                   |                | difference (IDC | difference (IDC |
|       |           |     |              |                          |                               |                   |                | vs NAT)        | vs NAT)        |
|       |           |     |              |                          |                               |                   |                | (IDC vs NAT)   | (IDC vs NAT)   |
|       |           |     |              |                          |                               |                   |                | Methylation    | Methylation    |
|       |           |     |              |                          |                               |                   |                | difference (IDC | difference (IDC |
|       |           |     |              |                          |                               |                   |                | vs NAT)        | vs NAT)        |

1
---
| cg25824760 18 | TSS200 | Island | ANKRD30B | 0.003 | 0.127 | 0.02 | 0.121 |
| cg23703062 18 | TSS200 | Island |         | 0.035 | 0.183 | 0.027 | 0.191 |
| cg13266435 18 | TSS200 | Island |         | 0.008 | 0.237 | 0.037 | 0.179 |
| cg21293934 18 | TSS200 | Island |         | 0.017 | 0.21 | 0.03 | 0.168 |
| cg24061208 18 | 1st Exon, 5′UTR | S_Shore |         | 0.004 | 0.299 | 0.017 | 0.271 |
| cg03014226 18 | 1st Exon, 5′UTR | S_Shore |         | 0.17 | 0.441 | 0.003 | 0.431 |
| cg21281009 18 | 1st Exon, 5′UTR | S_Shore |         | 0.013 | 0.206 | 0.045 | 0.136 |

2
---
| cg23196831 8 | 1st Exon, 5′UTR | Island | COL1A1 | 0.049 | 0.229 | 0.034 | 0.261 |
| cg26626663 8 | 5′UTR | S_Shore |         | 0.015 | 0.366 | 0.011 | 0.349 |
| cg23586322 8 | 5′UTR | S_Shore |         | 0.014 | 0.14 | 0.04 | 0.276 |
| cg23281803 8 | 5′UTR | S_Shore |         | 0.014 | 0.303 | 0.004 | 0.296 |

3
---
| cg09102257 5 | gene body | Island | EGR1 | 0.002 | 0.238 | 0.005 | 0.232 |
| cg07336840 5 | gene body | Island |         | 0.017 | 0.146 | 0.008 | 0.152 |
| cg12065840 5 | gene body | Island |         | 0.038 | 0.116 | 0.011 | 0.135 |
| cg23586322 5 | gene body | Island |         | 0.002 | 0.267 | 0.003 | 0.257 |

4
---
| cg23046919 12 | 3′UTR | S_Shore | IGF1 | 0.017 | 0.216 | 0.035 | 0.178 |
| cg12742178 12 | gene body | E |         | 0.038 | 0.113 | 0.011 | 0.135 |
| cg00264799 12 | gene body | E |         | 0.017 | 0.146 | 0.008 | 0.152 |

5
---
| cg24729879 5 | gene body | Island | IL6ST | 0.001 | 0.225 | 0.001 | 0.218 |
| cg16375820 5 | gene body | Island |         | 0.002 | 0.167 | 0.004 | 0.142 |
| cg15219433 5 | TSS200 | P | Island | 0.01 | 0.122 | 0.004 | 0.144 |

6
---
| cg12922492 7 | gene body | E | INHBA | 0.024 | 0.2 | 0.025 | 0.184 |
| cg01412469 7 | TSS1500 | Island |         | 0.007 | 0.13 | 0.011 | 0.136 |
| cg12929057 7 | gene body | E |         | 0.015 | 0.195 | 0.001 | 0.252 |

7
---
| cg14119337 14 | TSS1500 | N_Shore | MEG3 | 0.003 | 0.179 | 0.006 | 0.14 |
| cg12967319 14 | TSS1500 | N_Shore |         | 0.026 | 0.137 | 0.012 | 0.15 |
| cg04304932 14 | TSS1500 | Island |         | 0.002 | 0.203 | 0.031 | 0.127 |
| cg15419911 14 | gene body | Island |         | 0.004 | 0.195 | 0.017 | 0.177 |
| cg14232427 14 | gene body | Island |         | 0.001 | 0.21 | 0.012 | 0.183 |

8
---
| cg04212239 3 | gene body | P | S_Shore | 0.001 | 0.198 | 0.006 | 0.23 |
| cg13783238 3 | TSS1500 | S_Shelf | SMC4 | 3.15 | 1.08E−37 | 0 | 0.155 |
| cg12785694 3 | TSS1500 | S_Shelf |         | 0.015 | 0.213 | 0.009 | 0.212 |
| cg09100593 3 | TSS200 | S_Shelf |         | 0.003 | 0.28 | 0.001 | 0.329 |
| cg00475229 3 | gene body | P | S_Shore | 0.001 | 0.3 | 0 | 0.312 |
were expected. There has only been one other whole genome DNA methylation study in TNBC15, which focused on the prognostic value of DNA methylation patterns. The study by Stirzaker et al. identified 308 hypermethylated genes by comparing IDC versus NAT samples using Methylation-Binding-Domain Capture sequencing data15. We identified 227 (73.7%) of these genes in our study cohort to be significantly hypermethylated. The global analysis of DNA methylation revealed a higher number of hypermethylated probes and DMRs (366 probes) showed significantly altered methylation. Over 2,000 of these significantly associated probes were located within enhancer or promoter regions. Pathway enrichment analysis revealed the following significant pathways for hypomethylated genes: Inflammatory mediator regulation of TRP channels, Fructose and mannose metabolism, Regulation of lipolysis in adipocytes, Estrogen signaling pathway, and Platelet activation. Pathway enrichment analysis revealed the following significant pathways for hypermethylated genes: Glycosaminoglycan biosynthesis - heparan sulfate/heparin, Axon guidance, Gastric acid secretion, GABAergic synapse, and Glycosaminoglycan biosynthesis - chondroitin sulfate/dermatan sulfate (Supplementary Table 5).

Tumour versus lymph node. We have previously reported that miRNA and gene expression patterns are highly similar in LN metastases and the primary tumour within our study cohort16,17. We next compared whether the methylation changes occurring in LN metastases were similar to those occurring in the primary tumour. The two comparisons of IDC vs NAT and LN vs NAT yielded an overlap of 88.75% (=39,057 probes) or 89.93% (2134 DMRs), indicating that IDC and matched LN metastases DNA methylation alterations are highly similar.

We next compared the DNA methylation of all IDC samples with the DNA methylation of all LN samples to determine methylation changes specifically associated with metastases that were not present in the primary tumour. This comparison revealed that 5,221 probes (58.1% hypomethylated, 41.9% hypermethylated) and 104 DMRs (=366 probes) showed significantly altered methylation. Over 2,000 of these significantly associated probes were located within enhancer or promoter regions. Pathway enrichment analysis revealed the following significant pathways for hypomethylated genes: Inflammatory mediator regulation of TRP channels, Fructose and mannose metabolism, Regulation of lipolysis in adipocytes, Estrogen signaling pathway, and Platelet activation. Pathway enrichment analysis revealed the following significant pathways for hypermethylated genes: Glycosaminoglycan biosynthesis - heparan sulfate/heparin, Axon guidance, Gastric acid secretion, GABAergic synapse, and Glycosaminoglycan biosynthesis - chondroitin sulfate/dermatan sulfate (Supplementary Table 5).

Discussion
We have previously identified gene expression changes in TNBC primary tumours compared to matched NAT as well as gene expression changes associated with the progression of TNBC from primary tumour to lymph node metastasis. In this study we performed whole genome DNA profiling to determine the contribution of DNA methylation to these gene expression changes as DNA methylation is known to gene silencing events. Here we used the same sample cohort to identify DNA methylation changes that were associated with the previously observed alterations in gene expression by: (1) comparing tumour and matched normal samples and; (2) that were associated with lymph node metastasis.

First, we aimed to identify a global methylation profile of TNBC primary tumour samples (IDC) compared to matched normal adjacent tissue (NAT) to provide information about tumour specific differences. We identified and validated global hypermethylation and hypomethylation that included single loci as well as differentially methylated regions (DMRs) (a minimum of three significant consecutive probes). Since DNA methylation contributes to gene expression changes4,16, the differences in the methylation profiles between these two tissue types were expected. There has only been one whole genome DNA methylation study in TNBC15, which focused on the prognostic value of DNA methylation patterns. The study by Stirzaker et al. identified 308 hypermethylated genes by comparing IDC versus NAT samples using Methylation-Binding-Domain Capture sequencing data15. We identified 227 (73.7%) of these genes in our study cohort to be significantly hypermethylated. The global analysis of DNA methylation revealed a higher number of hypermethylated probes and DMRs (>17,000 probes, >1,300 DMRs) than hypomethylated probes and DMRs (>12,000 probes, 307 DMRs) comparing the two tissue types, which has also been reported by Stirzaker et al.15. Interestingly over 10,000 of these probes are located within enhancer or promoter regions. Activation/inactivation of enhancers can affect the transcription of the host gene so that they can act as alternative promoters19–21. Our study identified a number of pathways that were associated with altered methylation within enhancer/promoter regions, including the estrogen signalling pathway. In particular, DNA methylation has been associated with hormone receptor status of breast cancer patients22,23 and our current study suggests that DNA methylation may be involved in the downregulation of the estrogen receptor in TNBC patients, which was also shown in ref. 15.

Next, we focused more specifically on the genes identified in our previous study16. There we identified 66 genes to be differentially expressed in tumour versus normal samples. Our current study revealed that 26 of the 66 genes showed significantly altered DNA methylation, which was verified in two independent cohorts. The majority of these (19 genes = 73%) were negatively correlated, such that where gene expression was upregulated, DNA methylation was decreased and vice versa. However, there were genes whose DNA methylation profile was positively correlated with their gene expression (gene expression was upregulated when DNA methylation was increased, and vice versa). This does not mean that the DNA methylation of these genes does not contribute to their gene expression levels as described recently24. There Wan et al.24 hypothesised two mechanisms of DNA methylation-dependent gene regulation: (1) the dogma of gene repression due to DNA methylation; and (2) gene

---

**Figure 2. Survival analysis of the EGR1 probe cg07336840.** The y-axis shows the percent of survival of patients within the validation cohort. The x-axis shows the number of months of survival since diagnosis. The blue line represents patients with low DNA methylation of this probe, whereas the red line represents patients with high DNA methylation of this probe.
| Column ID | CHR | UCSC_REFGENE_GROUP | ENHANCER(E)/Promoter(P) | RELATION_TO_UCSC_CPG_ISLAND[1] | UCSC_REFGENE_NAME | Gene expression | DNA methylation | Validation cohort | Study cohort |
|-----------|-----|--------------------|-------------------------|--------------------------------|------------------|----------------|----------------|----------------|-------------|
| cg12565635 | 7   | TSS1500            |                         |                                | AGR2             | 8.72E-12 -3.01 | 0.038 0.137 0.003 0.161 |                 |
| cg25824760 | 18  | TSS200             | Island                  |                                |                  |                | 0.003 0.127 0.02 0.121 |                 |
| cg23703062 | 18  | TSS200             | Island                  |                                |                  |                | 0.035 0.183 0.027 0.191 |                 |
| cg13266435 | 18  | TSS200             | Island                  |                                |                  |                | 0.008 0.237 0.037 0.179 |                 |
| cg21293934 | 18  | TSS200             | Island                  |                                |                  |                | 0.017 0.21 0.03 0.168 |                 |
| cg24061208 | 18  | 1st Exon, 5′UTR    | S_Shore                 |                                |                  |                | 0.004 0.299 0.017 0.271 |                 |
| cg03014326 | 18  | 1st Exon, 5′UTR    | S_Shore                 |                                |                  |                | 0 0.441 0.003 0.431 |                 |
| cg21281009 | 18  | 1st Exon, 5′UTR    | S_Shore                 |                                |                  |                | 0.013 0.206 0.045 0.136 |                 |
| cg23196831 | 8   | 1st Exon, 5′UTR    | Island                  |                                |                  |                | 0.049 0.229 0.034 0.261 |                 |
| cg26626663 | 8   | 5′UTR              | S_Shore                 |                                |                  |                | 0.015 0.366 0.011 0.349 |                 |
| cg12065840 | 8   | 5′UTR              | S_Shore                 |                                |                  |                | 0.032 0.284 0.085 0.223 |                 |
| cg23586322 | 8   | 5′UTR              | S_Shore                 |                                |                  |                | 0.014 0.36 0.04 0.276  |                 |
| cg23281803 | 8   | 5′UTR              | S_Shore                 |                                |                  |                | 0.014 0.303 0.004 0.296 |                 |
| cg25448355 | 8   | gene body          | E                       |                                |                  |                | 0 0.189 0 0.204 |                 |
| cg04239786 | 5   | gene body          | E                       |                                |                  |                | 0.001 0.231 0.001 0.22 |                 |
| cg16204420 | 5   | gene body          | E                       |                                |                  |                | 0.001 0.17 0.002 0.219 |                 |
| cg20041567 | 6   | 1st Exon, TSS1500  | Island                  |                                |                  | 0.012 0.17 0.002 0.219 |                 |
| cg23046919 | 12  | 3′UTR              |                         |                                |                  | 0.02 0.163 0.008 0.152 |                 |
| cg12742178 | 12  | gene body          | E                       |                                |                  | 0.017 0.216 0.035 0.178 |                 |
| cg09267499 | 12  | gene body          | E                       |                                |                  | 0.038 0.116 0.011 0.135 |                 |
| cg24729879 | 5   | 5′UTR              | N_Shore                 |                                |                  | 0.001 0.225 0.001 0.218 |                 |
| cg16375820 | 5   | 5′UTR              | N_Shore                 |                                |                  | 0.002 0.167 0.004 0.142 |                 |
| cg15219433 | 5   | TSS200             | P Island                |                                |                  | 0.01 0.122 0.004 0.144 |                 |
| cg23143313 | 6   | gene body          | Island                  |                                |                  | 0.019 0.136 0.005 0.161 |                 |
| cg08884591 | 11  | TSS1500            |                         |                                |                  | 0.008 0.292 0.003 0.301 |                 |
| cg06372223 | 16  | gene body          | N_Shelf                 |                                |                  | 0.007 0.197 0.001 0.245 |                 |
| cg08005102 | 4   | TSS1500            |                         |                                |                  | 0.009 0.194 0.034 0.141 |                 |
| cg14119337 | 14  | TSS1500            | N_Shore                 |                                |                  | 0.003 0.179 0.006 0.14 |                 |
| cg12967319 | 14  | TSS1500            | N_Shore                 |                                |                  | 0.026 0.137 0.012 0.15 |                 |
| cg09403493 | 14  | TSS1500            | Island                  |                                |                  | 0.002 0.203 0.031 0.127 |                 |
| cg15419911 | 14  | gene body          | Island                  |                                |                  | 0.004 0.195 0.017 0.177 |                 |
| cg14123427 | 14  | gene body          | Island                  |                                |                  | 0.001 0.21 0.012 0.183 |                 |
| cg26374305 | 14  | gene body          | Island                  |                                |                  | 0 0.337 0 0.329 |                 |
| cg98669721 | 14  | gene body          | Island                  |                                |                  | 0 0.19 0.006 0.168 |                 |
| cg24183098 | 14  | gene body          | S_Shore                 |                                |                  | 0.039 -0.143 0.034 -0.151 |                 |
| cg03099990 | 14  | TSS1500            | N_Shore                 |                                |                  | 0.049 -0.151 0.006 -0.23 |                 |
| cg08451113 | 6   | gene body          | N_Shore                 |                                |                  | 0.007 -0.118 0.007 -0.133 |                 |
| cg06845268 | 2   | TSS1500            | S_Shore                 |                                |                  | 0.028 -0.187 0.007 -0.242 |                 |
| cg26212229 | 18  | TSS1500            | N_Shore                 |                                |                  | 0.026 -0.18 0.011 -0.183 |                 |
| cg23433370 | 12  | TSS200             | P N_Shore               |                                |                  | 0 -0.336 0 0 -0.35 |                 |
| cg07240557 | 12  | TSS200             | P N_Shore               |                                |                  | 0 -0.215 0 0 -0.212 |                 |

Table 2. Validated DNA methylation of 41 probes that are significantly different in the comparison of IDC vs NAT. These are associated with 16 of the 38 TNBC specific genes identified in our previous study [1]. Every CpG island consist of N- and S Shores next to it, which are neighboured by N- and S-shelves. First N-shelf, N-shore, CpG island, S-shore, and last S-shelf. [2] Methylation difference between IDC and NAT shows % of methylation change/100.

activation through DNA methylation. They found positively correlated gene-methylation relationships to be in more conserved regions and mainly in promoter regions, which suggests that these positive correlations have a regulatory role and do not just happen by chance. It may also indicate differential promoter usage as seen in ref. 25.

Our previous study found that EGR1 gene expression is downregulated comparing tumour versus normal. The methylation of EGR1 was found to be negatively correlated with its gene expression in our TNBC cohort, where there were five significant CpG methylation probes that were hypermethylated in both cohorts comparing these two tissue types. This gene has also been identified by other studies comparing IDC versus NAT in TNBC[20].
is a zinc finger protein that acts as a transcriptional regulator functioning as a tumour suppressor by regulating other tumour suppressors including TGFβ1, PTEN, p53, and fibronectin27. We were able to correlate the DNA methylation of one of the significant probes within EGR1 (cg07336840) with overall survival (p < 0.05). High DNA methylation at this locus is significantly associated with better overall survival. Interestingly, the study by Stirzaker et al. identified a DMR located in the WT-1 gene to be associated with poorer overall survival in TNBC patients15. EGR1 and WT-1 are members of the same family (early growth response – zinc-finger family) but with mostly opposing functions, EGR1 activates the transcription of genes that WT-1 represses28,29.

We previously identified that the gene expression of 38 genes is specific to TNBC16. Here we investigated the DNA methylation levels of these genes to determine the contribution of DNA methylation to gene expression. We showed that half (16/38) of the TNBC specific genes showed significantly altered DNA methylation at 41 probes; and of these there are five genes classed as DMRs (28 probes). These five genes are ANKRD30B, COL14A1, IGF1, IL6ST, and MEG3. ANKRD30B has been shown to be expressed in breast, brain, and testicular tumours30 but it has not been studied in TNBC to date. Stirzaker et al.15 also identified significant hypermethylation of COL14A1 when comparing tumour versus normal tissue15. However, there is a need for functional analysis of COL14A1 in TNBC. IGF1 has been the focus of multiple TNBC studies and is known to regulate cell proliferation and survival, and has been suggested as a potential treatment target for TNBC31–33. We have previously identified IL6ST as a TNBC-specific gene and its gene expression to be associated with overall survival (increased gene expression → better survival)16. Our previous study showed MEG3 is associated with lymph node metastasis16. It is a long non-coding RNA that is known to be down-regulated in multiple cancers and to regulate cell proliferation through the p53-tumour suppressor pathway34,35.

Due to its aggressive nature and increased number of metastasis, TNBC patients have much poorer outcomes relative to other subtypes. Therefore, we aimed to identify differences during early cancer progression from the primary tumour site to lymph node metastasis (LN). Eighty three genes were previously shown to be associated with LN metastasis16. Here we revealed that the expression of 38 of these genes may be influenced by methylation changes at single DNA methylation loci. Of these there are 14 genes that were differentially methylated in DMRs (over 74 probes). The survival analysis on our validation cohort using these 74 probes (Fig. 4) showed that nine probes were significantly associated with survival. Seven probes associated with five genes (SPRY2, EGR1, GREB1, ITIH5, LRRC17) were associated with better survival having higher methylation, whereas one probe for AMIGO2 (cg09933058) shows better survival with low methylation. Interestingly, of the significant DMRs comparing IDC versus NAT only one probe showed significant association with survival (EGR1 (cg07336840)) which is also in a DMR comparing LN versus NAT. SPRY2 and AMIGO2 have not previously been studied in TNBC. However, SPRY2 is a known tumour suppressor that regulates the RAS-ERK pathway36,37. AMIGO2 has been shown to be differentially expressed in other cancers including gastric adenocarcinomas, and it is known to effect ploidy, chromosomal stability, cell adhesion/migration, and tumourigenicity38. It also controls cell survival and angiogenesis via Akt activation39. The DMR of LRRC17 is associated with survival in three probes (high methylation/ better survival). Interestingly, this gene has been identified as a TNBC specific gene previously40. However, no functional analysis of LRRC17 in TNBC has been done. Promoter methylation of the tumour suppressor ITIH5 has been suggested as early breast cancer detection biomarker41. Finally, GREB1 is a key estrogen regulator42 and is expressed in hormone responsive breast cancers43 but not in TNBC.16.
Due to the lack of LN samples in our validation cohort, we utilised Methylation-Binding-Domain Capture sequencing (MBDcap seq) data from 7 LN and 4 NAT samples to validate the direction of methylation change of the genes that are associated with LN metastasis. Differential methylation was validated for 18 of the 38 genes associated with LN metastasis. The MBDcap seq provided regional methylation analysis covering 2 kb upstream from the first transcription start site to the end of the gene of interest, these regions were broken up into 1 kb tiles. The majority of genes (12 of 14) show a negative correlation between gene expression and DNA methylation. However, there are four genes that show negative and positive correlation in different probes. These are TSHZ2 (known to be down-regulated in breast and prostate cancer\(^4\)), ITIH5 (promoter methylation is an early breast cancer detection biomarker\(^4\)), GREB1 (estrogen regulator\(^42,43\)), MEG3 (long non-coding RNA known to

Figure 4. Survival analysis of nine probes that show significant methylation changes comparing lymph node metastasis to matched normal adjacent tissue. The x-axis shows the number of months of survival since diagnosis. The green line represents patients with low DNA methylation of this probe, whereas the red line represents patients with high DNA methylation of this probe.
| Gene name | Methylation difference | 450K array | MBDCap seq | Gene expression |
|-----------|------------------------|------------|-------------|-----------------|
| GREB1     | −0.207                 | 11673928   | TSS1500     | 0.045           |
|           | −0.160                 | 11674057   | TSS200      |                 |
|           | −0.180                 | 11674557   | 5'UTR       |                 |
|           | −0.234                 | 11724901   | gene body   |                 |
|           | −0.173                 | 11734257   | gene body   |                 |
|           | −0.132                 | 11761803   | gene body   |                 |
| RBMS3     | −0.176                 | 29782270   | 3'UTR       | 0.032           |
|           | −0.191                 | 38535574   | 5'UTR       |                 |
|           | −0.284                 | 38595383   | 1st Exon 5'UTR |          |
| LIFR      | −0.140                 | 38535574   | 5'UTR       | 0.048           |
|           | −0.284                 | 38595383   | 1st Exon 5'UTR |          |
| IL20RA    | −0.193                 | 137362747  | gene body   | 0.025           |
| CD36      | −0.137                 | 80274687   | TSS1500     | 0.040           |
| LRRC17    | 0.122                  | 102574455  | gene body   | 0.047           |
|           | 0.110                  | 102574504  | gene body   |                 |
| AGR3      | −0.139                 | 16922492   | TSS1500     | 0.025           |
| RELN      | −0.148                 | 103279088  | gene body   | 0.013           |
|           | −0.213                 | 103301932  | gene body   | 0.037           |
| CDCA2     | −0.204                 | 25317638   | 5'UTR       | 0.044           |
| MMP16     | −0.144                 | 89053427   | 3'UTR       | 0.049           |
|           | −0.222                 | 890966     | E           |                 |
|           | 0.010                  | 102563144  | gene body   |                 |
|           | 0.013                  | 80285891   | gene body   |                 |
|           | 0.011                  | 102563144  | gene body   |                 |
|           | 0.010                  | 103216964  | gene body   |                 |
|           | 0.040                  | 103220964  | gene body   |                 |
|           | 0.040                  | 103250964  | gene body   |                 |
|           | 0.008                  | 103261964  | gene body   |                 |
|           | 0.006                  | 103404964  | gene body   |                 |
|           | 0.029                  | 103449964  | gene body   |                 |
|           | 0.034                  | 103450964  | gene body   |                 |
|           | 0.037                  | 103566964  | gene body   |                 |
|           | 0.044                  | 253220153  | gene body   |                 |
|           | 0.005                  | 253220153  | gene body   |                 |
|           | 0.006                  | 253220153  | gene body   |                 |
|           | 0.975                  | 3.65E-07   | gene body   |                 |
|           | 1.670                  | 6.38E-05   | gene body   |                 |
|           | 2.870                  | 0.000112   | gene body   |                 |

**Continued**
be down-regulated in cancer\textsuperscript{14,15}, and RELN (methylated and down-regulated in pancreatic cancer, where its expression has been associated with increased cell motility, invasiveness and colony-forming ability\textsuperscript{46} but has not been described in breast cancer). This could mean that some loci overpower others or potentially, that during tumour progression to lymph nodes, the methylation becomes tissue specific and changes. Further research into loci specific DNA methylation during cancer progression is urgently needed to explain these phenomena.

Additionally, we reviewed the connection of the 18 validated genes to epithelial-mesenchymal transition (EMT) (a process which cells undergo to travel to distant sites, leading to distant cancerous disease). The majority of these genes (12/18) have a known connection to EMT, which we summarised in Table 4. These findings support the importance of these genes during cancer progression.

In conclusion, this is the first whole genome DNA methylation analysis in a TNBC cohort including matched lymph node metastases. Here we identified and validated a global DNA methylation profile, which we correlated with our previously published gene expression findings in tumour, matched lymph node and matched normal

### Table 4. Validation of the direction of methylation comparing lymph node metastasis (LN) to matched normal adjacent tissue (NAT).

| Gene name | 450K array | MBDcap seq | Gene expression |
|-----------|------------|------------|-----------------|
| SVEP1     |            |            |                 |
| 0.005     | 0.243      | 113308063  | E               | gene body       |
| 0.005     |            |            | 0.005           | Fold change     |
| 0.005     |            |            | −3.650          | chr9:113134161-113135160 |
| 0.001     |            |            | −4.528          | chr9:113173161-113174160 |
| 0.001     |            |            | 4.632           | chr9:113179161-113180160 |
| 0.033     |            |            | −2.510          | chr9:113320161-1133202160 |
| 0.020     |            |            | 2.810           | chr9:1133266161-1133267160 |
| 0.005     |            |            | −3.681          | chr9:1133268161-1133269160 |
| 0.001     |            |            | −5.016          | chr9:113324161-113325160 |
| ITIH5     |            |            |                 |
| 0.045     | −0.147     | 7614151    | gene body       |
| 0.000     | −0.241     | 7614570    | gene body       |
| 0.001     |            |            | 0.029           | Fold change     |
| 0.001     |            |            | −2.592          | chr10:7628962-7629961 |
| 0.025     |            |            | −2.680          | chr10:7644962-7645961 |
| 0.005     |            |            | −5.016          | chr10:7649662-7650661 |
| 0.001     |            |            | −4.689          | chr10:7699662-7700661 |
| 0.027     | −0.115     | 7662761    | TSS1500         |
| 0.003     | −0.120     | 101291100  | TSS1500         |
| 0.003     |            |            | 0.005           | Fold change     |
| 0.003     |            |            | −3.724          | chr14:101318445-101319444 |
| 0.033     | −0.201     | 101296297  | gene body       |
| 0.020     | −0.223     | 101317622  | TSS1500         |
| 0.019     | 0.239      | 51592177   | gene body       |
| 0.048     |            |            | −2.287          | chr20:51601946-51602945 |
| 0.005     |            |            | −3.707          | chr20:5168496-51685945 |
| 0.001     |            |            | −4.547          | chr20:51752946-51753945 |
| 0.000     |            |            | −6.638          | chr20:51790946-51791945 |
| 0.038     |            |            | −2.422          | chr20:51819946-51820945 |
| 0.047     | −0.190     | 52029943   | 3’UTR           |
| 0.041     |            |            | −2.386          | chr20:51852946-51862945 |
| 0.035     |            |            | −2.488          | chr20:51879946-51880945 |
| 0.026     |            |            | −2.672          | chr20:51914946-51915945 |
| 0.043     |            |            | −2.353          | chr20:51928946-51929945 |
| 0.037     |            |            | 2.453           | chr20:51969646-51975945 |
| 0.041     | −0.119     | 52036204   | 3’UTR           |
| 0.021     |            |            | −2.803          | chr20:51995946-51996945 |
| 0.036     |            |            | −2.457          | chr20:52006946-52007945 |
| 0.050     |            |            | −2.268          | chr20:52049946-52050945 |
| 0.016     |            |            | 2.954           | chr20:52082946-52083945 |
| 0.046     |            |            | −2.311          | chr20:52103946-52104945 |

| Gene name | 450K array | MBDcap seq | Gene expression |
|-----------|------------|------------|-----------------|
| MEG3      |            |            |                 |
| 0.047     | −0.190     | 52029943   | 3’UTR           |
| 0.035     |            |            | −2.386          | chr20:51852946-51862945 |
| 0.026     |            |            | −2.672          | chr20:51914946-51915945 |
| 0.043     |            |            | −2.353          | chr20:51928946-51929945 |
| 0.041     |            |            | 2.453           | chr20:51959646-51975945 |
| 0.041     |            |            | −2.803          | chr20:51995946-51996945 |
| 0.036     |            |            | −2.457          | chr20:52006946-52007945 |
| 0.050     |            |            | −2.268          | chr20:52049946-52050945 |
| 0.016     |            |            | 2.954           | chr20:52082946-52083945 |
| 0.046     |            |            | −2.311          | chr20:52103946-52104945 |
adjacent tissue. Our findings show that DNA methylation contributes to the deregulation of gene expression changes and is associated with overall survival.

**Methods**

**Study design.** The study cohort comprised a total of 23 grade three invasive ductal carcinomas (IDC), 12 matched lymph node metastasis (LN), and 11 matched normal adjacent tissues (NAT), from which DNA was isolated and screened using the Illumina HumanMethylation450 BeadChips to reveal DNA methylation changes across the genome. All samples in the study cohort were obtained as formalin-fixed, paraffin-embedded (FFPE) blocks from the archives of the Hunter Area Pathology Service, John Hunter Hospital, Newcastle, Australia. This cohort has been described previously 17. A pathologist confirmed the triple negative phenotype, areas of NAT, invasive cancer and LN metastasis. As previously described, 1.5 mm punch biopsies were used to isolate areas of IDC, LN and NAT from these sections.

The validation cohort used in this study contained 70 IDC samples derived from the Australian Breast Cancer Tissue Bank (ABCTB) and the same 11 NAT samples used in the study cohort. All participants consented to the use of their tissue in this study. Details regarding this cohort are shown in Table 5.

To assess the direction of DNA methylation in LN metastasis samples, we utilised Methylation-Binding-Domain-Capture sequencing data from 7 LN samples and 4 NAT samples (Supplementary Table 6).

**Ethics statement.** All experiments were performed in accordance with approved guidelines and regulations. This study, including all experimental protocols, was granted a waiver of consent in accordance with the National Statement on Ethical Conduct in Research Involving Humans.

This study, including all experimental protocols, complies with the Helsinki Declaration with ethical approval from the Hunter New England Human Research Ethics Committee (Approval number: 09/05/20/5.02).

Written informed consent was obtained from all patients included in this study.

**DNA extraction.** The Gentra Puregene Tissue Kit (Qiagen, Venlo, Limburg, Netherlands) was used to isolate DNA from FFPE tissue following the manufacturers’ instruction with few alterations. The protocol including the alterations was as follows.

All biopsy samples were placed in 1.5 ml tubes. Five hundred microliters of Xylene were added in a fume hood and incubated with constant gentle mixing at 55 °C. This was followed by centrifugation at 16,000 g for 3 minutes. The supernatant was discarded and the Xylene wash step repeated twice. Five hundred microliters of 100% ethanol was added and incubated for 5 mins under constant mixing at room temperature. For the cell lysis, 300 μl of Cell Lysis Solution was added to each tube and incubated for 10 mins at 70°C. Twenty microliters Protease K (20 mg/ml) were added to each sample, mixed for 20 sec and incubated at 55°C overnight. On day two, a further 10 μl Protease K was added to each sample, mixed and again incubated overnight at 55 °C. On day three, 5 freeze-thaw cycles were performed, 5 mins on dry ice and 5 mins at 95 °C. All samples were brought

| Gene   | Connection to EMT                                                                 | Ref. |
|--------|----------------------------------------------------------------------------------|------|
| GREB1  | EMT is regulated by 17β-estradiol which can be regulated by GREB1                 | 46   |
| RB53M  | is up-regulated during TGF-β1-induced EMT. No direct connection to EMT in TNBC has been shown | 47   |
| MME    | there is no known connection to EMT                                              |      |
| PLSCR4 | there is no known connection to EMT                                              |      |
| GSGF10 | there is no known connection to EMT                                              |      |
| LIFR   | is targeted by miR-9 (EMT related microRNA), which is upregulated in breast cancer. LIFR acts as metastasis suppressor | 48   |
| IL20RA | there is no known connection to EMT                                              |      |
| CD36   | high level of CD36 (fatty acid transporter) promote EMT                          | 49   |
| LRRC17 | there is no known connection to EMT                                              |      |
| AGR3   | is regulated by ZEB1 (key regulator of EMT)                                     | 50   |
| RELN   | interacts with SNAIL (key regulator of EMT)                                     | 51   |
| CDCA2  | its gene expression can predict metastasis outcome in synovial sarcomas           | 52   |
| MMP16  | is well known to play a role in EMT; high expression leads to tumour invasion; is targeted by miR-200 (key microRNA in EMT) | 53   |
| SVEP1  | is activated by TNFα (a pro-inflammatory cytokine able to affect adhesion and migration, and to induce EMT) | 54   |
| ITIH5  | inactivation of ITIH5 promotes cancer progression in bladder cancer; no direct connection to EMT in TNBC to date | 55   |
| MEG3   | known regulator of EMT, targets TGF-β1 receptor genes (EMT regulators); inhibits cell proliferation and induces apoptosis (through p53) | 48   |
| TSHZ2  | down-regulation of TSHZ2 releases GLI1 (which can induce EMT)                    | 56   |

Table 4. Connection of 18 validated genes, which are associated with lymph node metastasis, to epithelial-mesenchymal transition (EMT).
| Patient Number | Tumour grade | Age at diagnosis | Month of follow up/to death | Follow up Status |
|----------------|--------------|------------------|-----------------------------|-----------------|
| 1              | 1            | 51.94            | 186                         | Died from other cause |
| 2              | 3            | 52.1             | 44                          | Died from disease |
| 3              | 3            | 37               | 16                          | Died from disease |
| 4              | 3            | 66               | 82                          | Died from disease |
| 5              | 3            | 76               | 113                         | Alive no disease |
| 6              | 3            | 45               | 40                          | Died from disease |
| 7              | 3            | 71               | 68                          | Alive no disease |
| 8              | 3            | 62               | 52                          | Alive no disease |
| 9              | 3            | 81               | 8                           | Died from other cause |
| 10             | 3            | 75               | 17                          | NA |
| 11             | 3            | 45               | 35                          | Died from other cause |
| 12             | 3            | 42               | 140                         | Alive no disease |
| 13             | 3            | 41               | 39                          | Died from disease |
| 14             | 3            | 37               | 172                         | Alive no disease |
| 15             | 3            | 52.55            | 61                          | Died from disease |
| 16             | 3            | 68.26            | NA                          | Died from other cause |
| 17             | 3            | 70               | 33                          | Died from disease |
| 18             | 3            | 37               | 7                           | Alive no disease |
| 19             | 3            | 71               | 17                          | Died from disease |
| 20             | 3            | 74               | 170                         | Alive no disease |
| 21             | 3            | 41               | 23                          | Died from disease |
| 22             | 3            | 39               | 66                          | Alive no disease |
| 23             | 3            | 59               | 21                          | Died from disease |
| 24             | 3            | 83               | 46                          | Alive no disease |
| 25             | 3            | 42               | 37                          | Died from other cause |
| 26             | 3            | 32               | 66                          | Alive no disease |
| 27             | 3            | 44               | 29                          | Died from other cause |
| 28             | 3            | 35               | 46                          | Alive no disease |
| 29             | 3            | 43               | 62                          | Died from other cause |
| 30             | 3            | 61               | 48                          | Alive no disease |
| 31             | 3            | 39               | 40                          | Died from other cause |
| 32             | 2            | 55.46            | 188                         | Died from other cause |
| 33             | 2            | 29.09            | NA                          | Died from disease |
| 34             | 3            | 32               | 49                          | Alive no disease |
| 35             | 3            | 74               | 24                          | Died from disease |
| 36             | 3            | 45               | 169                         | Alive no disease |
| 37             | 3            | 55.9             | 131                         | Died from other cause |
| 38             | 3            | 30               | 43                          | Died from disease |
| 39             | 3            | 62               | 7                           | Died from disease |
| 40             | 3            | 45.64            | 32                          | Died from disease |
| 41             | 3            | 44               | 38                          | Died from disease |
| 42             | 3            | NA               | NA                          | Alive no disease |
| 43             | 3            | 47               | 27                          | Died from disease |
| 44             | 3            | 62               | 35                          | Alive no disease |
| 45             | 3            | 67               | 28                          | Alive no disease |
| 46             | 3            | 56.59            | 30                          | Died from disease |
| 47             | 3            | 47               | 8                           | Died from disease |
| 48             | 3            | 79               | 103                         | Died from disease |
| 49             | 3            | 69               | 34                          | NA |
| 50             | 3            | 39               | 62                          | Died from other cause |
| 51             | 3            | 54.58            | 23                          | Died from disease |
| 52             | 3            | 78               | 49                          | Alive no disease |
| 53             | 3            | 60               | 24                          | Died from disease |
| 54             | 3            | NA               | NA                          | Alive no disease |
| 55             | 3            | 44               | 81                          | Died from disease |

Continued
samples were assayed in triplicates. Two microliters of genomic DNA (at 1 ng/assay according to the manufacturers' instructions (Applied Biosystems 7500 Fast Real-Time PCR system). All used to assess the integrity of the DNA used for methylation analysis. It was performed using a Real-time PCR Expression Omnibus (GEO) with Accession No. GSE78758.

To calculate the delta Ct ($\Delta$Ct) the average Ct of the QC template reagent was subtracted from the average Ct as well as the scanning (iScan) of the HumanMethylation 450 K BeadChip arrays was performed using the manufacturer’s instructions (Life Technologies, Carlsbad, CA, United States of America).

**Illumina Infinium HD FFPE methylation arrays.** The array results have been deposited in Gene Expression Omnibus (GEO) with Accession No. GSE78758.

The Infinium HD FFPE quality control (QC) Assay (Illumina, San Diego, CA, United States of America) was used to assess the integrity of the DNA used for methylation analysis. It was performed using a Real-time PCR assay according to the manufacturers’ instructions (Applied Biosystems 7500 Fast Real-Time PCR system). All samples were assayed in triplicates. Two microliters of genomic DNA (at 1 ng/µl) was used for each reaction. The threshold cycle (Ct) was calculated for each individual sample. Replicates where the Ct diverged by more than half a unit were excluded. The average Ct was calculated for each sample as well as the QC template reagent (Illumina). To calculate the delta Ct ($\Delta$Ct) the average Ct of the QC template reagent was subtracted from the average Ct of each sample. The $\Delta$Ct had to remain below 5 for the samples to pass the quality control test. Of the initial 37 IDC samples, 27 had enough DNA (>250 ng) for further analysis, and 23 of these passed the quality control ($\Delta$Ct < 5), all 12 LN samples passed the QC stage as well as the 3 NAT pooled samples and one of the single NAT samples passed the QC.

Bisulphite conversion was performed using the EZ-96 DNA Methylation Kit (Zymo Research, Irvine, CA, United States of America).

Next, FFPE Restoration was undertaken following the Infinium HD FFPE Restore Protocol (Illumina). Infinium HD FFPE Methylation Assay (Illumina) with the hybridisation, washing and staining of the arrays as well as the scanning (iScan) of the HumanMethylation 450 K BeadChip arrays was performed using the manufacturers’ instructions.

**Methylation array analysis.** The data from all samples was imported in form of the idat files into Genomic Suite 6.6 (Partek, St Louis, Missouri, United States of America) and Illumina normalisation was performed. ANOVA analysis was performed to detect differentially methylated probes between groups (IDC versus NAT, LN versus NAT, and IDC versus LN). Significance was granted if $p < 0.05$ and the estimated difference between groups ($\Delta$) was $< -0.1$ or $> 0.1$, signifying a methylation change of at least 10%.

These analyses were performed on single loci (=probes on the BeadChips) and on differentially methylated regions (DMR) – a minimum of three significant consecutive probes. The focus in this study was the 38 triple negative specific genes which we identified in our previous study.$^{16}$

All samples of the validation cohort were treated and analysed in the same way. Due to the lack of LN samples in this cohort, only the IDC versus NAT comparisons were performed.

Pathway enrichment analysis was performed using Genomic Suite 6.6 (Partek). All significant probes were filtered to include probes that are within enhancer and/or promoter regions. These were then used for pathway enrichment analysis, which is a tool within Genomic Suite 6.6 (Partek). The enrichment score is the negative natural log of the enrichment p-value derived from the Fisher's exact test of the pathway enrichment analysis.

### Table 5. Sample information from the validation cohort.

| Patient Number | Tumour grade | Age at diagnosis | Month of follow up/to death | Follow up Status |
|----------------|--------------|------------------|-----------------------------|------------------|
| 56             | 3            | 39               | 46                          | Alive no disease |
| 57             | 3            | 38.84            | 44                          | Died from disease|
| 58             | 3            | 46               | 43                          | Alive no disease |
| 59             | 3            | 45.38            | 194                         | Alive no disease |
| 60             | 3            | 66.18            | 69                          | Died from disease|
| 61             | 3            | 68               | 59                          | Died from other cause|
| 62             | 3            | 38               | 18                          | Died from disease|
| 63             | 3            | 39               | 168                         | Died from other cause|
| 64             | 3            | 62               | 9                           | Died from disease|
| 65             | 3            | 62               | 148                         | Died from other cause|
| 66             | 3            | 78               | 15                          | Died from disease|
| 67             | 3            | 69               | 34                          | NA               |
| 68             | 3            | 60               | 195                         | Died from disease|
| 69             | 3            | 62               | 102                         | Died from disease|
| 70             | 3            | 58               | 45                          | Died from disease|
**MDB-Cap Sequencing.** We compared our findings (comparing LN versus NAT) with MDBcap sequencing data provided by Dr. Clare Stirzaker and Prof. Susan Clark from the Garvan Institute Sydney. The methylation profiling was performed as previously described in ref. 15. This analysis provides regional DNA methylation information, therefore a strict validation with our CpG-specific analysis was not possible. The regions that were analysed using the MDBcap data included our genes of interest (83 genes that are associated with LN metastasis, identified in ref. 16) starting 2 kb upstream of the first transcription start site (TSS) to the 3′ end of the genes. These regions are broken up into 1 kb tiles. The methylation profiling was performed as described by Stirzaker, et al.15 where the same sample cohort had been used (excluding the LN samples) and methylation differences between LN and NAT samples were assessed using a Student's t-test (using the R base package) (p < 0.05).

**References**

1. Carotenuto, P. et al. Triple negative breast cancer: from molecular portrait to therapeutic intervention. *Crit Rev Eukaryot Gene Expr* 20, 17–34 (2010).
2. Chacon, R. D. & Costanzo, M. V. Triple-negative breast cancer. *Breast Cancer Res* 12 Suppl 2, S3 (2010).
3. Chen, J. Q. & Russo, J. E. RALPHA-negative and triple negative breast cancer: molecular features and potential therapeutic approaches. *Biochem Biophys Acta* 1796, 162–175 (2009).
4. Wong-Brown, M. W. et al. Prevalence of BRCA1 and BRCA2 germline mutations in patients with triple-negative breast cancer. *Breast Cancer Res Treat* 150, 71–80 (2015).
5. Szasz, A. M. et al. The CIN4 chromosomal instability qPCR classifier defines tumor aneuploidy and stratifies outcome in grade 2 breast cancer. *PLoS One* 8, e56707 (2013).
6. Reis-Filho, J. S. & Tutt, A. N. Triple negative tumours: a critical review. *Histopathology* 52, 108–118 (2008).
7. Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nat Biotechnol* 28, 1057–1068 (2010).
8. Hirst, M. & Marra, M. A. Epigenetics and human disease. *Int J Biochem Cell Biol* 41, 136–146 (2009).
9. Feinberg, A. P. & Vogelstein, B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 301, 89–92 (1983).
10. Esteller, M. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene* 21, 5427–5440 (2002).
11. How Kit, A., Nielsen, H. M. & Tost, J. DNA methylation based biomarkers: practical considerations and applications. *Biochimie* 94, 2314–2337 (2012).
12. Branham, M. T. et al. Methylation profile of triple-negative breast carcinomas. *OncoGenesis* 1, e17 (2012).
13. Feng, W. et al. Correlation between CpG methylation profiles and hormone receptor status in breast cancers. *Breast Cancer Res* 9, R57 (2007).
14. Sharma, P. et al. The prognostic value of promoter methylation in early stage triple negative breast cancer. *J Cancer Ther Res* 3, 1–11 (2014).
15. Stirzaker, C. et al. Methyleme sequencing in triple-negative breast cancer reveals distinct methylation clusters with prognostic value. *Nat Commun* 6, 5899 (2015).
16. Mathe, A. et al. Novel genes associated with lymph node metastasis in triple negative breast cancer. *Sci Rep* 5, 15832 (2015).
17. Avery-Kiejda, K. A., Braye, S. G., Mathe, A., Forbes, J. F. & Scott, R. J. Decreased expression of key tumour suppressor microRNAs is associated with lymph node metastases in triple negative breast cancer. *BMC Cancer* 14, 51 (2014).
18. Mathe, A., Scott, R. J. & Avery-Kiejda, K. A. miRNAs and Other Epigenetic Changes as Biomarkers in Triple Negative Breast Cancer. *Int J Mol Sci* 16, 28347–28376 (2015).
19. Heintzman, N. D. et al. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108–112 (2009).
20. Kowalczyk, M. S. et al. Intragenic enhancers act as alternative promoters. *Mol Cell* 45, 447–458 (2012).
21. Stone, A. et al. DNA methylation of oestrogen-regulated enhancers defines endocrine sensitivity in breast cancer. *Nat Commun* 6, 7758 (2015).
22. Conway, K. et al. DNA methylation profiling in the Carolina Breast Cancer Study defines cancer subclasses differing in clinicopathologic characteristics and survival. *Breast Cancer Res* 16, 450 (2014).
23. Stefansson, O. A. et al. DNA methylation-based definition of biologically distinct breast cancer subtypes. *Mol Oncol* 9, 555–568 (2015).
24. Van, J. et al. Characterization of tissue-specific differential DNA methylation suggests distinct modes of positive and negative gene expression regulation. *BMC Genomics* 16, 19 (2015).
25. Bert, S. A. et al. Regional activation of the cancer genome by long-range epigenetic remodeling. *Cancer Cell* 23, 9–22 (2013).
26. Cascione, L. et al. Integrated microRNA and miRNA signatures associated with survival in triple negative breast cancer. *PLoS One* 8, e55910 (2013).
27. Baron, V., Adamson, E. D., Calogero, A., Ragona, G. & Mercola, D. The transcription factor Egr1 is a direct regulator of multiple tumor suppressors including TGFbeta1, PTEN, p53, and fibronectin. *Cancer Gene Ther* 20, 619–637 (2015).
28. Dey, B. R. et al. Repression of the transforming growth factor-beta 1 gene by the Wilms’ tumor suppressor WT1 gene product. *Mol Endocrinol* 8, 595–602 (1994).
29. Sukhatme, V. P. et al. A zinc finger-encoding gene coregulated with c-fos during growth and differentiation, and after cellular depolarization. *Cell* 53, 37–43 (1988).
30. Jager, D. et al. Identification of a tissue-specific putative transcription factor in breast tissue by serological screening of a breast cancer library. *Cancer Res* 61, 2055–2061 (2001).
31. Chandran, S., Harmey, J. & Toomey, S. Inhibition of IGFB signalling pathway in MDA-MB-231 triple negative breast cancer cells. *BMC Proc* 6, O14 (2012).
32. Davison, Z., de Blacquière, G. E., Westley, B. R. & May, F. E. B. Insulin-like Growth Factor-Dependent Proliferation and Survival of Triple-Negative Breast Cancer Cells: Implications for Therapy. *Neoplasia* 13, 504–515 (2011).
33. Lindner, R. et al. Molecular Phenotypes in Triple Negative Breast Cancer from African American Patients Suggest Targets for Therapy. *PLoS One* 8(2013).
34. Sun, M. et al. Downregulated long noncoding RNA MEG3 is associated with poor prognosis and promotes cell proliferation in gastric cancer. *Tumour Biol* 35, 1065–1073 (2014).
35. Wang, P., Ren, Z. & Sun, P. Overexpression of the long non-coding RNA MEG3 impairs in vitro glioma cell proliferation. *J Cell Biochem* 113, 1868–1874 (2012).
36. Kwik, H. J. et al. Downregulation of Spry2 by miR-21 triggers malignancy in human gliomas. *Oncogene* 30, 2433–2442 (2011).
37. Sharma, S. B. et al. MicroRNAs 206 and 21 cooperate to promote RAS-extracellular signal-regulated kinase signaling by suppressing the translation of RASA1 and SPRED1. *Mol Cell Biol* 34, 4143–4164 (2014).
38. Rabenau, K. E. et al. DEGA/AMIGO-2, a leucine-rich repeat family member, differentially expressed in human gastric adenocarcinoma: effects on ploidy, chromosomal stability, cell adhesion/migration and tumorigenecity. *Oncogene* 25, 5036–5047 (2006).
39. Park, H. et al. AMIGO2, a novel membrane anchor of PDK1, controls cell survival and angiogenesis via Akt activation. *J Cell Biol* 211, 619–637 (2015).
40. Zaka, M. U., Yonghong, P. & Sutton, C. W. Microarray big data integrated analysis to identify robust diagnostic signature for triple negative breast cancer. In Computational Intelligence in Big Data (CIBD), 2014 IEEE Symposium on 1–7 (2014).
41. Kloten, V. et al. Promoter hypermethylation of the tumor-suppressor genes ITIH5, DKK3, and RASSF1A as novel biomarkers for blood-based breast cancer screening. Breast Cancer Res 15, R4 (2013).
42. Mohammed, H. et al. Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. Cell Rep 3, 342–349 (2013).
43. Ghosh, M. G., Thompson, D. A. & Weigel, R. J. PDZK1 and GREB1 are estrogen-regulated genes expressed in hormone-responsive breast cancer. Cancer Res 66, 6367–6375 (2000).
44. Yamamoto, M., Cid, E., Bru, S. & Yamamoto, F. Rare and frequent promoter methylation, respectively, of TSHZ2 and 3 genes that are both downregulated in expression in breast and prostate cancers. PLoS One 6, e17149 (2011).
45. Sato, N., Fukushima, N., Chang, R., Matsubayashi, H. & Goggins, M. Differential and Epigenetic Gene Expression Profiling Identifies Frequent Disruption of the REL-N Pathway in Pancreatic Cancers. Gastroenterology 130, 548–565 (2006).
46. Laviolette, I. A., Hodgkinson, K. M., Minhas, N., Perez-Iratxeta, C. & Vanderhyden, B. C. 17beta-estradiol upregulates GREB1 and accelerates ovarian tumor progression in vivo. Int J Cancer 135, 1072–1084 (2014).
47. Johansson, J., Tabor, V., Wikell, A., Jalkanen, S. & Fuxe, J. TGF-beta1-Induced Epithelial-Mesenchymal Transition Promotes Monocyte/Macrophage Properties in Breast Cancer Cells. Front Oncol 5, 3 (2015).
48. Lin, C. W., Lin, P. Y. & Yang, P. C. Noncoding RNAs in Tumor Epithelial-to-Mesenchymal Transition. Stem Cells Int 2016, 2732705 (2016).
49. Nath, A., Li, I., Roberts, L. R. & Chan, C. Elevated free fatty acid uptake via CD36 promotes epithelial-mesenchymal transition in hepatocellular carcinoma. Sci Rep 5, 14752 (2015).
50. Obazac, J. et al. The role of AGR2 and AGR3 in cancer: similar but not identical. Eur J Cell Biol 94, 139–147 (2015).
51. Yuan, Y., Chen, H., Ma, G., Cao, X. & Liu, Z. Reelin is involved in transforming growth factor-beta1-induced cell migration in esophageal carcinoma cells. PLoS One 7, e31802 (2012).
52. Brodin, B. A. CINSARC: a new look into an old concept gives hope for new treatments for synovial sarcomas. Transl Pediatr 2, 70–72 (2013).
53. Xue, Z. et al. Mesenchymal stem cells promote epithelial to mesenchymal transition and metastasis in gastric cancer though paracrine cues and close physical contact. J Cell Biochem 116, 618–627 (2015).
54. Bonomi, S. et al. Oncogenic alternative splicing switches: role in cancer progression and prospects for therapy. Int J Cell Biol 2013, 962038 (2013).
55. Rose, M. et al. Epigenetic inactivation of ITIH5 promotes bladder cancer progression and predicts early relapse of pT1 high-grade urothelial tumours. Carcinogenesis 35, 727–736 (2014).
56. Riku, M. et al. Down-regulation of the zinc-finger homebox protein TSHZ2 releases GLI1 from the nuclear repressor complex to restore its transcriptional activity during mammary tumorigenesis. Oncotarget 7, 5690–5701 (2016).

Acknowledgements
The authors would like to thank the Pathologist Dr. Ricardo Vilain for the review of all tumour and normal tissue specimens used in this study, as well as Ms Sarah Nielsen and Dr. Mythily Mariasegaram for their assistance with archival specimens. We would also like to acknowledge, that the MBDCap data was kindly provided by Dr. Clare Stirzaker and Prof. Sue Clark from the Garvan Institute, Sydney. Dr. Andrea Mathe is supported by a University Postgraduate Award. This study was funded by the National Breast Cancer Foundation and the Hunter Medical Research Institute, through donations from Hunter Breast Cancer Foundation, PULSE, Brad Webb and Paul Craven along with the Wests Group Australia.

Author Contributions
A.M. study concept and design, carried out experiments, analysis and interpretation of data, drafting of manuscript. M.W.B. carried out experiments of the validation cohort with A.M. W.J.L. performed analysis of MBDCap seq data. C.S. provided data for MBDCap seq data. S.G.B. patient collection, material support. J.F.F. study and approved the final version of this manuscript.

Additional Information
Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Mathe, A. et al. DNA methylation profile of triple negative breast cancer-specific genes comparing lymph node positive patients to lymph node negative patients. Sci. Rep. 6, 33435; doi: 10.1038/srep33435 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

© The Author(s) 2016