



# EBV Associated Breast Cancer Whole Methylome Analysis Reveals Viral and Developmental Enriched Pathways

Mohammad O. E. Abdallah<sup>1</sup>, Ubai K. Algizouli<sup>1</sup>, Maram A. Suliman<sup>1</sup>, Rawya A. Abdulrahman<sup>1</sup>, Mahmoud Koko<sup>1</sup>, Ghimja Fessahaye<sup>1</sup>, Jamal H. Shakir<sup>2</sup>, Ahmed H. Fahal<sup>3</sup>, Ahmed M. Elhassan<sup>1</sup>, Muntaser E. Ibrahim<sup>1</sup> and Hiba S. Mohamed<sup>1,4\*</sup>

<sup>1</sup> Department of Molecular Biology, Institute of Endemic Disease, University of Khartoum, Khartoum, Sudan, <sup>2</sup> Department of Surgery, Khartoum Teaching Hospital, Khartoum, Sudan, <sup>3</sup> Department of Surgery, Faculty of Medicine, University of Khartoum, Khartoum, Sudan, <sup>4</sup> Department of Biology, Taibah University, Medina, Saudi Arabia

## OPEN ACCESS

### Edited by:

Ala-Eddin Al Moustafa,  
Qatar University, Qatar

### Reviewed by:

Saima Wajid,  
Jamia Hamdard University, India  
Said Dermime,  
National Center for Cancer Care and  
Research, Qatar

### \*Correspondence:

Hiba S. Mohamed  
hibasalah@iend.org

### Specialty section:

This article was submitted to  
Cancer Epidemiology and Prevention,  
a section of the journal  
Frontiers in Oncology

**Received:** 22 March 2018

**Accepted:** 24 July 2018

**Published:** 13 August 2018

### Citation:

Abdallah MOE, Algizouli UK, Suliman MA, Abdulrahman RA, Koko M, Fessahaye G, Shakir JH, Fahal AH, Elhassan AM, Ibrahim ME and Mohamed HS (2018) EBV Associated Breast Cancer Whole Methylome Analysis Reveals Viral and Developmental Enriched Pathways. *Front. Oncol.* 8:316. doi: 10.3389/fonc.2018.00316

**Background:** Breast cancer (BC) ranks among the most common cancers in Sudan and worldwide with hefty toll on female health and human resources. Recent studies have uncovered a common BC signature characterized by low frequency of oncogenic mutations and high frequency of epigenetic silencing of major BC tumor suppressor genes. Therefore, we conducted a pilot genome-wide methylome study to characterize aberrant DNA methylation in breast cancer.

**Results:** Differential methylation analysis between primary tumor samples and normal samples from healthy adjacent tissues yielded 20,188 differentially methylated positions (DMPs), which is further divided into 13,633 hypermethylated sites corresponding to 5339 genes and 6,555 hypomethylated sites corresponding to 2811 genes. Moreover, bioinformatics analysis revealed epigenetic dysregulation of major developmental pathways including hippo signaling pathway. We also uncovered many clues to a possible role for EBV infection in BC.

**Conclusion:** Our results clearly show the utility of epigenetic assays in interrogating breast cancer tumorigenesis, and pinpointing specific developmental and viral pathways dysregulation that might serve as potential biomarkers or targets for therapeutic interventions.

**Keywords:** methylome, breast cancer, epigenetics, DNA methylation, HM450, epigenome reference, EBV

## BACKGROUND

Breast cancer (BC) is the most common cancer among females in Sudan (1–3), and is still a leading cause of high morbidity and mortality across the world. According to a recent report from the national cancer registry (2), BC had an incidence rate of 25.1 per 100,000, more than twice the incidence rate of the second commonest cancer. Furthermore, Sudanese BC patients tend to present at young age, at late stage, and with advanced disease compared to their counterparts in other countries (4). Another study (5) reported a young age of presentation for locally advanced BC. Therefore, there is an urgent need for serious epidemiologic and molecular studies in order to trace the underlying mechanisms behind BC, and for developing better early detection methods as well as a nationwide educational effort to tackle this ravaging disease.

Epigenetics has emerged as a new, rapidly growing field in biology, with significant implications for cancer research. Epigenetic modifications include DNA methylation, and histone modifications, although they both do not alter DNA sequence *per se*, they influence chromatin remodeling and thus offer a dynamic and flexible way of controlling gene expression.

DNA methylation of cytosine residues occurs predominantly at CpG sites, and is mediated by three DNA methyltransferases (DNMTs). DNMT1, which maintains DNA methylation during cell replication, and a pair of DNMT3s–DNMT3a and DNMT3b—which is responsible for *de novo* DNA methylation. Epigenetic reprogramming through genome-wide alteration of DNA methylation (methylome) is critical for control of development and differentiation in normal cells and tissues, however, faulty epigenetic reprogramming, as in aberrant DNA methylation, can be a major driver of multiple types of cancer including BC (6, 7).

Methylome analysis has proved to be very pertinent to the study of the different aspects of cancer tumorigenesis. The vast majority of methylation changes occur in a tissue-specific manner (8), which makes methylome profiling a very sensitive and specific method for delineating dysregulated epigenetic pathways at the tissue level, as in cancer, which usually arises from a single tissue. Moreover, DNA methylation is a stable epigenetic mark that is ideal for development of biomarker assays, which can offer a rapid, cost effective, and minimally invasive diagnostic/prognostic tests (9, 10). Additionally, methylome analysis has been effectively used in tumor subtype classification (11–15). Furthermore, genome-wide methylome assays have also proved to be very useful in detecting and profiling viral epigenetic signature in cancer (16–18).

The aim of the present study is to investigate genome-wide DNA methylation profile of breast cancer in Sudanese patients utilizing Illumina Infinium HumanMethylation450 BeadChips (HM450) methylation assay. This array-based assay is widely used in epigenetics studies, and is a reliable, cost effective, high throughput method. We conducted methylome analysis comparing primary BC tissue samples against normal samples from adjacent healthy tissues. The results of this study provide a valuable insight into the epigenetics of BC in Sudanese patients.

## RESULTS

### Genome-Wide DNA Differential Methylation Analysis

Each of three approaches—listed in Materials and Methods—produced a list of differentially methylated sites: Limma, 39,940; Wilcoxon, 34,099; Nimbl, 22,251 (0.2 median beta value difference, Benjamini-Hochberg adjusted  $p$ -value  $\leq 0.05$ ). Here we only report the results for final set obtained from Nimbl-compare module, which represents the intersection of the three methods. The final set consisted of 20,188 differentially

methylated CpG sites, which is further divided into 13,633 hypermethylated sites corresponding to 5339 genes and 6555 hypomethylated sites corresponding to 2811 genes. Nimbl unique approach ensured detection of differentially methylated positions (DMPs) that have the largest effect size as illustrated in **Figure 1A**, a volcano plot showing the demarcation of differentially methylated sites by both statistical significance and effect size is shown in **Figure 1B**. Hierarchical clustering of the top 250 differentially methylated sites sorted by  $F$  value (low intragroup variability and higher intergroup variability) is shown in **Figure 2**. The resulting heatmap and dendrogram showed clear separation of tumor samples from normal samples.

### Genomic Distribution of Differentially Methylated CpG Sites

Differentially hypermethylated and hypomethylated sites displayed similar distribution with regard to gene elements as defined by HM45–TSS1500, TSS200, First Exon, gene body, and 3UTR—**Figure 3A**. However, they showed an asymmetric distribution with regard to CpG island relation with most of the hypermethylated sites mapping to CpG islands, whereas most of the hypomethylated sites mapped to open sea areas **Figure 3B**.

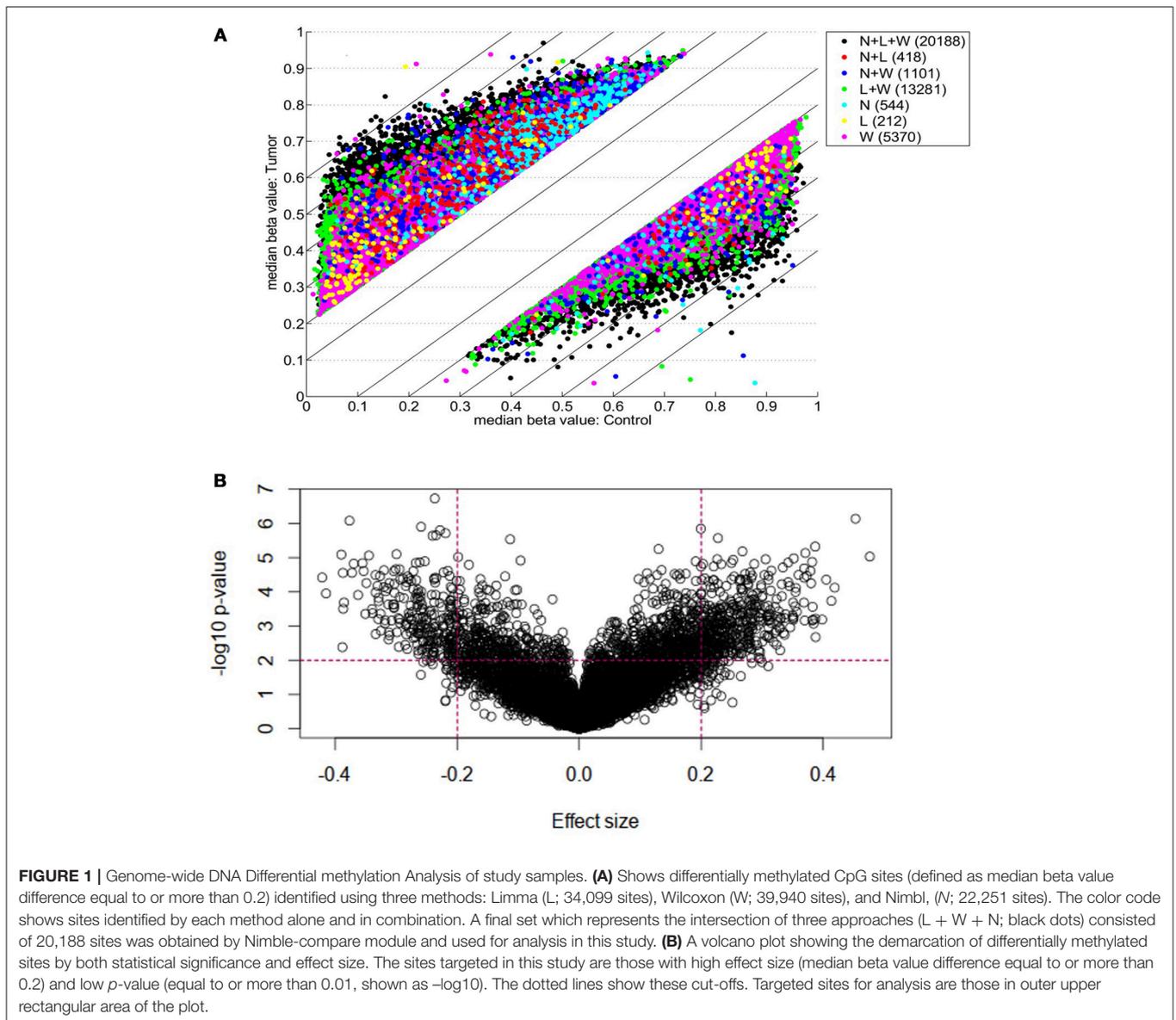
Of the 13,633 hypermethylated sites, 24.37% ( $N = 3,323$ ) mapped to Dnase hypersensitive sites compared with only 8.67% ( $N = 568$ ) of hypomethylated sites. Interestingly, while a greater percentage of hypermethylated compared to hypomethylated sites overlapped differentially methylated regions (DMR), [54.83% ( $N = 1,612$ ), 11.47% ( $N = 46$ )], respectively, hypomethylated sites were more concentrated in cancer DMR (CDMR), with 49.63% compared with 14.66% in hypermethylated sites, hypomethylated sites were more concentrated in cancer DMR (CDMR), with 49.63% compared with 14.66% in hypermethylated sites. The genomic distribution of hypermethylation and hypomethylation sites at each chromosome is shown in **Figures S1, S2**.

### Comparison to Reference Epigenome

We utilized data from the recently released Human epigenome reference data (19) to annotate the set of differentially methylated CpG sites. We mapped hyper and hypo DMPs in the promoter region from our data against two reference epigenome breast cell lines: HMEC (Human mammary primary epithelial cells), and vHMEC (Human mammary primary epithelial cell variant) (20, 21). We examined the change in chromatin states—from the 15-chromatin states model (19)—that accompany the acquisition or loss of DNA methylation in the context of transitioning from normal to tumor states. Our results revealed a noticeable gain of repressive marks for the hypermethylated DMPs, which increased from 55.5% in HMEC cells to 78.7% in vHMEC cells. Interestingly, we also found a slight increase in the percentage of repressive marks in the hypomethylated DMPs, which increased from 54.3 to 61.6%. Notably, in both cases, most of the upsurge in repressive regions were concentrated in Polycomb-repressed regions **Figures 3C,D**.

In addition, we observed a marked drop of all active chromatin states except for weak transcription and distal enhancer activity between the HMEC and vHMEC cells for the hypermethylated

**Abbreviations:** BC, breast cancer; DMP, differentially methylated position; DMR, differentially methylated region; CDMR, cancer differentially methylated site; TSS, transcription start site; UTR, untranslated region; MSig, mutation signature.



group. On the other hand, the hypomethylated group showed multiple notable shifts: From quiescent to Polycomb repression, from weak transcription to strong transcription, and from distal enhancers to genic enhancer (intronic enhancers).

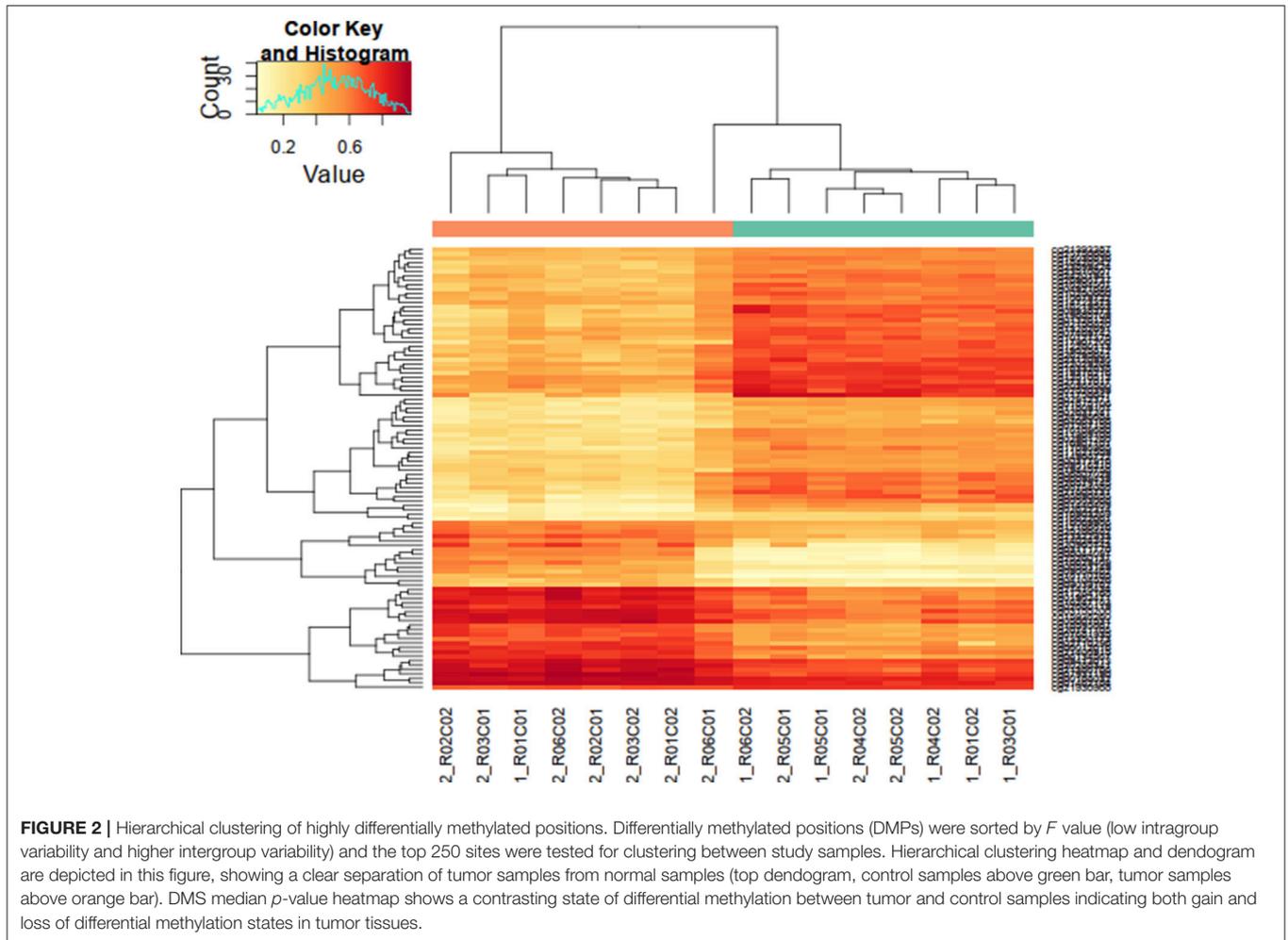
### Candidate Biomarkers Discovery

Nimbl method was used for detection and prioritization of candidate biomarkers with greatest inter-group variability, and lowest intra-group variability (22). Using this approach, we were able to identify a number of new as well as previously well-known BC biomarkers. Among the genes that showed significant promoter hypermethylation, we identified PAX6 (23, 24), WT1 (25), SOX1 (26), and TP73 (27, 28), all of them have been previously associated with BC. We also identified a set of previously uncharacterized biomarkers like PCDHGA1, HOXC4, and TBX15. To validate our candidate genes we interrogated

our candidate gene list against BC methylome data from the Cancer Genome Atlas Network: <http://cancergenome.nih.gov/> as compiled by MethHC (29) web portal. All the genes from our data were also significantly hypermethylated in the TCGA dataset. **Figure 4** shows promoter hypermethylation of the TP73 gene.

### Pathway and Network Analysis

Results from the ReactomeFI for the EDG network uncovered a massive network of 1310 nodes (genes) and 5097 edges (interactions), while the EUG list produced a smaller network of 763 nodes and 2265 edges. Furthermore, loading the NCI (National Cancer Institute) cancer gene index identified 781, and 470, neoplasia related genes from the EDG, and EUG networks, respectively, of which 332 EDG genes, and 222 EUG genes were associated with breast cancer in the cancer gene index.



Pathway enrichment analysis on the EUG network. Identified hippo signaling, Wnt signaling, and many extracellular matrix and metastasis promoting pathways as summarized in **Table 1**. Performing the pathway enrichment analysis on the breast cancer EUG subnetwork also identified hippo signaling and pathways of extracellular matrix in addition to pathways involved in immune response against viruses **Table 2**. Interestingly, breast cancer subnetwork showed significant enrichment for Epstein-Barr virus infection (FDR < 0.001).

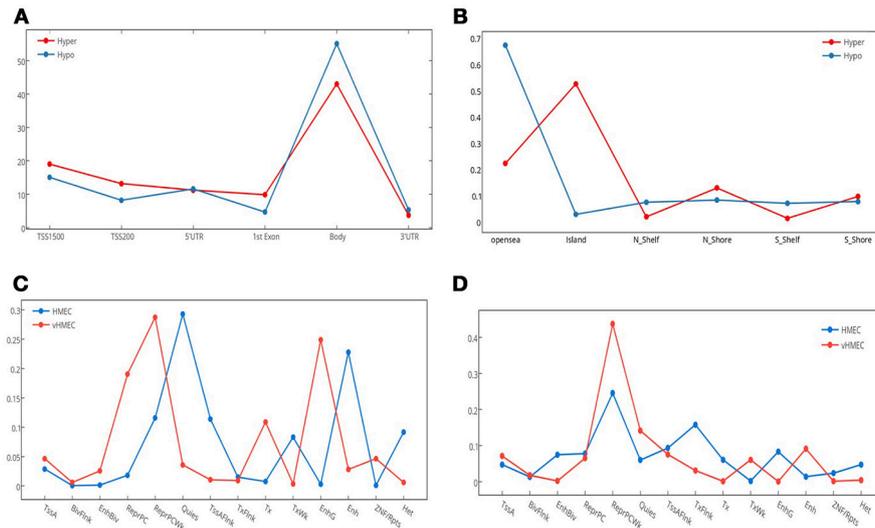
Pathway analysis on the EDG network identified Neuroactive ligand-receptor interactions, G-protein signaling, RAS signaling, RAS signaling, Potassium channel signaling, and many other pathways as summarized in **Table 3**. While the smaller EDG breast cancer subnetwork showed significant enrichment for a multitude of pathways including all the pathways that were enriched in the EDG network in addition to many cancer related and immune response pathways. Interestingly, the EDG sub network was also significant for direct p53 effectors. The complete list of enriched pathways for the EDG breast cancer subnetwork is shown in **Table S1**.

## DISCUSSION

### Leveraging the Reference Human Epigenome

The recent release of the human reference epigenome data by the Roadmap project ushered in a new era of epigenomics. The current study utilized this new wealth of information to interpret methylome data in the context of the human reference epigenome. We successfully mapped hyper and hypo DMPs to chromatin states from normal and premalignant reference breast cells (HMEC and vHMEC, respectively). Chromatin states reflect a concise and condensed representation of the epigenetic context, and are increasingly utilized to decipher genetic and epigenetic variability. Despite the fact that vHMEC is a premalignant and not a primary tumor cell, we argue that vHMEC is a suitable model for the epigenetic changes that accompany BC tumorigenesis because the vast majority of epigenetic changes tend to occur early during BC tumorigenesis (30–33).

Notably, our data revealed a strong Polycomb repression in both hypermethylated and hypomethylated CpG sites. These findings are in accordance with the emerging evidence that DMPs are enriched for Polycomb repression in primary breast



**FIGURE 3 |** Genomic and epigenomic distribution of differentially methylated positions (DMPs). This figure details the number of DMPs in relation to gene elements, CpG islands and chromatin states. **(A)** Distribution of hyper and hypo methylated CpG sites in relation to gene elements. TSS, transcription start site; UTR, untranslated region. **(B)** Distribution of hyper and hypo methylated CpG sites in relation to CpG Islands. N<sub>+</sub>, north; S<sub>-</sub>, south. **(C)** Distribution of Hypomethylated CpG sites in relation to chromatin states. **(D)** Distribution of Hypermethylated CpG sites in relation to chromatin states. Fourteen chromatin states are shown.

tumors (34) and triple negative BC (35). Moreover, various elements of the Polycomb repressive complexes are well-known to be overexpressed in BC (36, 37) and are required for stem cell state in mammary tumors (38, 39). Interestingly, Reyngold et al. found that unlike primary tumors, genes methylated in metastatic lesions seem to lack Polycomb repressive marks (40). Interestingly, an important mechanism for tumorigenesis such as Polycomb repression was only revealed by context dependent genome-wide comparison and not from any other method that interrogates hyper or hypomethylated region in isolation, without the paying attention to the broader epigenomic context.

## Network-Based Pathway Enrichment Analysis

Network-based pathway enrichment results for the EUG network revealed many upregulated pathways that have been previously associated with BC tumorigenesis. Hippo signaling, which appeared as the top significantly enriched pathway in our results, has recently emerged as an important regulator of BC growth, migration, invasiveness, stemness, as well as drug resistance (41). Wang et al. demonstrated that overexpression of YAP enhanced BC formation and growth. Hiemer et al. found that both TAZ and YAP-key effectors of the Hippo pathway are crucial to promote and maintain TGF $\beta$ -induced tumorigenic phenotypes in breast cancer cells (42). In addition, YAP was demonstrated to mediate drug resistance to RAF and MEK targeted cancer therapy (43, 44). Interestingly, we also reported an upregulated Wnt signaling pathway, which has been linked to BC growth and malignant behavior (45). Xu et al. found that Wnt signaling pathway is required for triple-negative breast cancer development (46). Recent studies have suggested long lasting reduced Wnt signaling

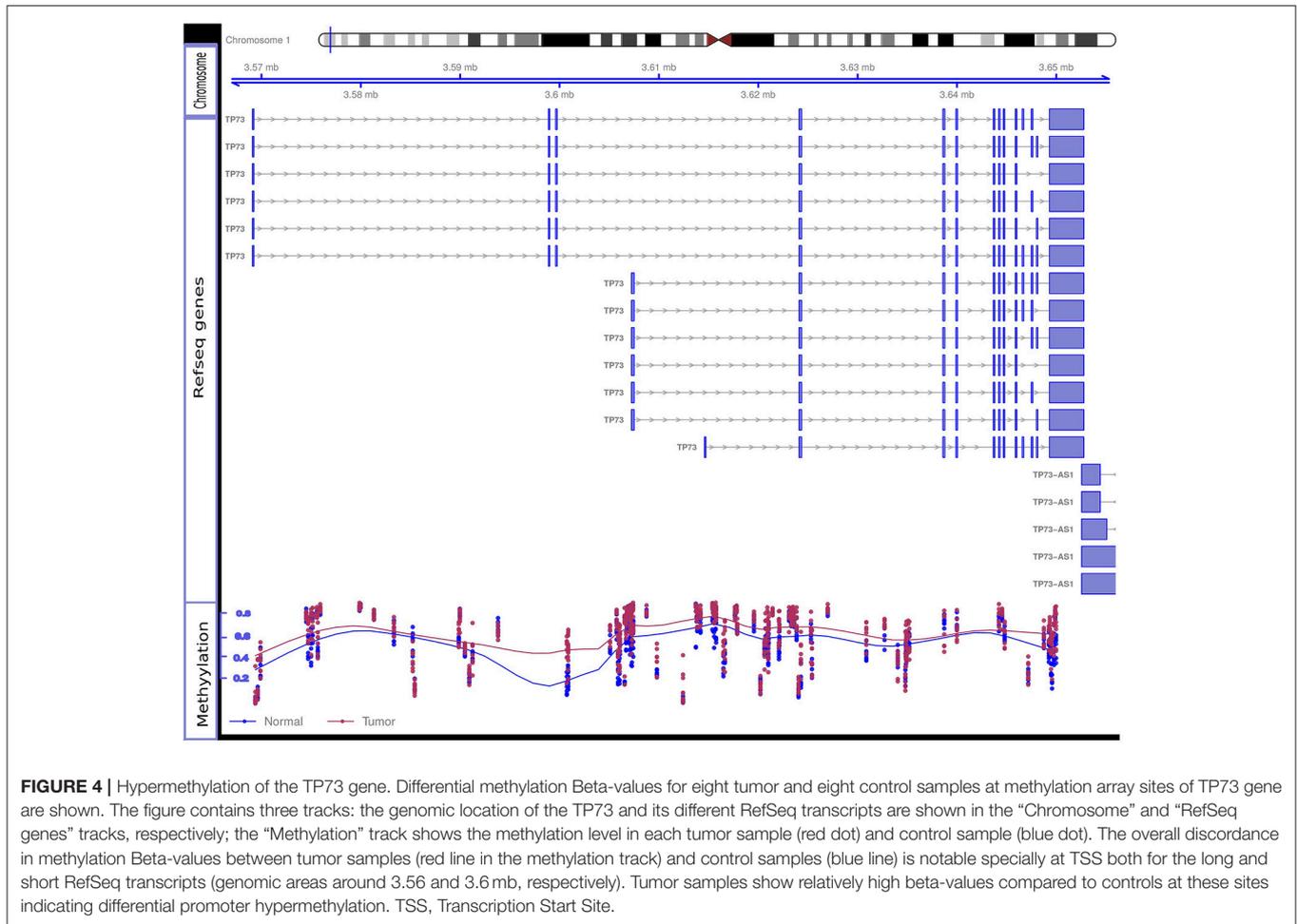
as the mechanism by which early pregnancy protects against BC (47).

Regarding the EDG network, Neuroactive ligand-receptor interaction, in addition to GPCR, RAS and Rap1 signaling were among the most significantly enriched pathways. Recent studies have found Neuroactive ligand-receptor interaction related genes to be hypermethylated in colorectal and EBV associated gastric cancers (20, 21, 48). Elements of RAS signaling like RASSF1 has been frequently found to be hypermethylated in BC (49), moreover, Qin et al. has demonstrated that resveratrol is able to demethylate RASSF1 promoter through decreased DNMT1 and DNMT3b in mammary tumors (50, 51). Notably, we reported the apparent silencing of multiple pro-tumor pathways in our results like GPCR and RAP1 signaling, the precise significance of this findings remains unclear. In addition, we also noticed the bivalent enrichment of multiple pathways (where different elements of the same pathway are both up and down regulated). Interpreting such perturbations is tricky, and predicting the net outcome of those perturbations might not be readily obvious given the crosstalk between different pathways.

## EBV Signature

We previously reported a strong association between EBV and BC in Sudanese patients (52), we also reported frequent epigenetic silencing of major tumor suppressor genes coupled with low frequency of known tumor associated mutations in the same population (53). In this study, we have demonstrated genome-wide epigenetic alterations consistent with our original proposition that epigenetic changes are the primary driver of BC tumorigenesis in Sudanese patients.

A myriad of recent studies point toward a common theme in EBV associated cancers characterized by genome-wide epigenetic



**TABLE 1 |** Pathway enrichment analysis results for epigenetically upregulated genes (EUG) interaction network.

Pathway	Number of genes in the geneset	Number of genes in the network	FDR
Hippo signaling pathway	154	31	<1.000e-03
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	74	20	<5.000e-04
L1CAM interactions	94	21	2.50E-04
Wnt signaling pathway	269	41	3.33E-04

*This pathway enrichment analysis and the interaction network were prepared using ReactomeFI Cytoscape app. The table shows the enriched pathways, the number of genes in the pathway from the total query gene set, and the number of genes in the pathway found in the interaction network. Results having p-values <0.01 and a False Detection Rate <0.001 are shown.*

changes coupled with a paucity of mutations. EBV infection is now known to play significant role in epithelial cancers like nasopharyngeal and gastric carcinomas mainly through genome-wide epigenetic changes (54–56). Li et al. observed a unique epiphenotype of EBV associated carcinomas suggesting

a predominant role for EBV infection in the ensuing epigenetic dysregulation of those cancers (17). Another study attributed the genome-wide promoter methylation in EBV driven gastric cancer to the induced expression of DNA methyltransferase-3b (DNMT3b) (57).

Our data mirrored the overall unique pattern of EBV infection characterized by sweeping epigenetic changes accompanied by low mutation frequency. Significantly, a major mechanism by which tumorigenic EBV virus avoids the Immune system is through manipulation of Polycomb proteins. Furthermore, we also showed that the EUG network was significantly enriched for EBV infection pathway **Figure 5**. In addition, results from MSig perturbations obtained from GREAT web tool (which predicts functions of cis regulatory elements) (58), showed significant enrichment for a set of downregulated genes which had been previously correlated with increased expression of EBV EBNA1 protein in NPC, in the hypermethylated CpG sites group, data not shown. For the hypomethylated CpG group, we found genes upregulated in B2264-19/3 cells (primary B lymphocytes) within 30–60 min after activation of LMP1 to be significantly enriched in MSig oncogenic signature. These findings taken together provide the first bioinformatics evidence of a possible active role for EBV infection in BC tumorigenesis in Sudanese patients.



**TABLE 2 |** Pathway enrichment results for breast cancer related epigenetically upregulated genes (EUG) subnetwork.

Pathway	Number of genes in the geneset	Number of genes in the network	FDR
CXCR4-mediated signaling events	79	10	2.50E-04
AP-1 transcription factor network	70	9	7.27E-04
HIF-1-alpha transcription factor network	66	9	4.00E-04
Viral myocarditis	59	9	2.50E-04
Pathways in cancer	327	22	<1.000e-03
HTLV-I infection	260	18	3.33E-04
Proteoglycans in cancer	225	17	2.00E-04
Epstein-Barr virus infection	202	16	1.67E-04
Hippo signaling pathway	154	14	3.33E-04
Natural killer cell mediated cytotoxicity	135	13	1.43E-04
Alzheimer disease-presenilin pathway	111	12	5.00E-04

*ReactomeFl cytoscape app was used to extract breast cancer related subnetworks from EUG set by loading NCI cancer index and performing pathway enrichment analysis on interaction networks. Nodes that corresponded to malignant breast cancer were selected. The table shows the enriched pathways, the number of genes in the pathway from the total query gene set, and the number of genes in the pathway found in the interaction network. Results having p-values <0.01 and a False Detection Rate <0.001 are shown.*

Coverage was targeted across gene regions with sites in the promoter region, 5'UTR, first exon, gene body, and 3'UTR.

### Data Preprocessing

For quality control, any array probes with *p* detection value <0.05 or missing beta values were removed. In addition, array sites corresponding to sex chromosomes or mapping to SNPs were filtered out. Peak-based correction (61) (PBC) was used to normalize the final dataset and to correct for probe type bias. Density plots of beta values for individual samples are shown in **Figure S3**.

### Genome-Wide DNA Differential Methylation Analysis

A trilateral approach consisting of two statistical methods augmented by one numerical method was used for the differential methylation analysis: Moderated *t*-test from R limma (62) package; Wilcoxon test (Non-Parametric test) from R stat package; and Nimbl (22) (Numerical Identification of Methylation Biomarker Lists) which is a Matlab package designed to identify and prioritize differentially methylated sites.

Nimbl core module identify potential biomarkers by calculating a score based on the inter-group and intra-group variability:

$$\text{Score} = \text{beta\_valdist} - (\text{mediandiff} - \text{beta\_valdist})$$

**TABLE 3 |** Pathway analysis on the epigenetically downregulated genes (EDG) interaction network.

Pathway	Number of genes in the geneset	Number of genes in the network	FDR
Neuroactive ligand-receptor interaction	275	81	<1.000e-03
GPCR ligand binding	433	107	<5.000e-04
PI3K-Akt signaling pathway	346	81	<3.333e-04
Extracellular matrix organization	263	65	<2.500e-04
Pathways in cancer	327	74	<2.000e-04
Rap1 signaling pathway	213	54	<1.667e-04
Regulation of actin cytoskeleton	215	54	<1.429e-04
Neurotransmitter receptor binding and downstream transmission in the postsynaptic cell	137	40	<1.250e-04
Potassium channels	86	30	<1.111e-04
Heterotrimeric G-protein signaling pathway-Gi alpha and Gs alpha mediated pathway	147	41	<1.000e-04
Proteoglycans in cancer	225	54	<9.091e-05
Ras signaling pathway	227	54	<8.333e-05
ECM-receptor interaction	86	28	1.54E-04
Calcium signaling pathway	181	45	2.14E-04
FGF signaling pathway	92	29	2.00E-04
Focal adhesion	206	48	2.50E-04
Gastrin-CREB signaling pathway via PKC and MAPK	207	48	2.35E-04
Cell adhesion molecules (CAMs)	143	37	2.78E-04
Wnt signaling pathway	269	57	4.21E-04
MAPK signaling pathway	259	55	4.00E-04
Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway	108	30	5.24E-04
IL4-mediated signaling events	63	21	7.73E-04
HTLV-I infection	260	54	7.83E-04
GABAergic synapse	90	26	7.92E-04
Signaling by Type 1 insulin-like growth factor 1 receptor (IGF1R)	86	25	8.40E-04
Retrograde endocannabinoid signaling	103	28	8.85E-04
Melanoma	71	22	9.26E-04

*The functional interaction network was constructed using ReactomeFl cytoscape app. The table shows the enriched pathways, the number of genes in the pathway from the total query gene set, and the number of genes in the pathway found in the interaction network. Results having p-values <0.01 and a False Detection Rate <0.001 are shown.*

Where beta\_valdist is the distance in beta values between non-overlapping groups and mediandiff is the absolute difference of the medians of each group (22). It then assigns high scores for CpG sites that achieve higher discrimination between groups while maintaining low within-group variability. Nimbl-compare module was also used to extract the final set of CpG sites that were identified by all three methods. Hierarchical clustering analysis was performed

using the top 250 differentially methylated sites sorted by  $F$  value.

## Reference Epigenome Annotations

Bed files of chromatin states for both HMEC and vHMEC cells were obtained from Roadmap web portal: [http://egg2.wustl.edu/roadmap/web\\_portal/](http://egg2.wustl.edu/roadmap/web_portal/), further analysis was performed in GALAXY web-based platform (63–65) and R statistical software.

## Network and Pathway Analysis

Differential methylation analysis produced two lists of differentially methylated genes (hyper and hypo) and their enrichment of differentially methylated sites in their gene regions, i.e., promoter region, gene body, 3'UTR, etc. The aggregated gene list was sorted by the count of methylated sites in the promoter area, first exon, and gene body regions. Subsequently all epigenetically upregulated genes (EUG) were combined in a single group, i.e., genes bearing methylation marks that promote gene expression–hypomethylation in the promoter area, and first exon or hypermethylation in the gene body region in a single group. Then we compiled a second group of epigenetically downregulated genes (EDG), i.e., genes bearing methylation marks that inhibit gene expression, i.e., hypermethylation of the promoter area, and the first exon or hypomethylation of the gene body region. We excluded other gene-based regions that are not well-correlated with gene expression from further analysis.

We utilized ReactomeFI (66), a Cytoscape (67) app to perform network and pathways analysis. Projecting the lists of EDG and EUG groups through the ReactomeFI functional network produced two corresponding networks. To extract breast cancer specific subnetworks from EUG and EDG groups we loaded NCI cancer index from within the ReactomeFI app, and we selected nodes that corresponded to malignant breast cancer.

## CONCLUSIONS

Interpreting cancer methylome is a complex process, as it is not easily correlated with cancer tumorigenesis as driver mutations or altered gene expression profiles. Other studies on Breast cancer, failed to correlate BC methylome with known and clear tumor subtypes that correlated with gene expression profiles. Gene lists of hyper and hypo methylated sites cannot be treated the same way we treat over and under expressed genes, and extreme caution should be exercised with such over simplistic approach. In this paper, we augmented old approaches with new enhanced analytic techniques that we think are more capable of deciphering methylome data than traditional methods. We are among the first studies to utilize chromatin states from the RODAMAP epigenome project to make sense of methylome data.

Utilizing the human reference epigenome, our study uncovered interesting epigenetic patterns characterized by increased acquisition of Polycomb repressive marks, as revealed

from comparison to human reference epigenome breast cells. We identified many potential BC biomarkers like TP73, and TBX15. Using pathway analysis over contextually aggregated methylome networks, we uncovered many significantly enriched developmental pathways including Hippo and Wnt signaling pathways. Additionally, our bioinformatics analysis indicated a possible role for EBV infection in BC tumorigenesis.

## AUTHOR CONTRIBUTIONS

HM conceived and design the study and contributed to manuscript writing and data interpretation. MI contributed to study design and manuscript writing. MA performed the data analysis, contributed to interpretation and prepared the manuscript draft. JS and AF recruited patients and provided samples. MK contributed to data analysis. AE performed the histopathology. UA, MS, RA, and GF contributed to sample collection, DNA extraction, and purification. All authors read and approved the final manuscript.

## FUNDING

This work received financial support from the international Centre for Genetic Engineering and Biotechnology (ICGEB) Project CRP/SUD/10-01.

## ACKNOWLEDGMENTS

We thank the breast cancer patients for their participation in this study. This work is dedicated to deceased Mohammed Abdelrazig. Senior Surgeon at Khartoum teaching hospital who facilitated samples collection.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fonc.2018.00316/full#supplementary-material>

**Figure S1** | Genomic distribution of hypermethylation marks shown at each chromosome. Black color indicates hypermethylation sites.

**Figure S2** | Genomic distribution of hypomethylation marks shown at each chromosome. Black color indicates hypermethylation sites.

**Figure S3** | Density plots of beta values for individual samples. Shades of red and yellow colors represent tumor samples, whereas shades of blue and green represent normal samples.

**Table S1** | Pathway enrichment results for breast cancer related Epigenetically Downregulated Genes (EDG) subnetwork. ReactomeFI cytoscape app was used to extract breast cancer related subnetworks from EUG set by loading NCI cancer index and performing pathway enrichment analysis on interaction networks. Nodes that corresponded to malignant breast cancer were selected. The table shows the enriched pathways, the number of genes in the pathway from the total query gene set, and the number of genes in the pathway found in the interaction network. Results having  $p$ -values  $<0.01$  and a False Detection Rate  $<0.001$  are shown.

## REFERENCES

- Elamin A, Ibrahim ME, Abuidris D, Mohamed KEH, Mohammed SI. Part I. Cancer in Sudan—burden, distribution, and trends breast, gynecological, and prostate cancers. *Cancer Med.* (2015) 4:447–56. doi: 10.1002/cam4.378
- Saeed IE, Weng H-Y, Mohamed KH, Mohammed SI. Cancer incidence in Khartoum, Sudan. first results from the Cancer Registry, 2009–2010. *Cancer Med.* (2014) 3:1075–84. doi: 10.1002/cam4.254
- Elgaili EM, Abuidris DO, Rahman M, Michalek AM, Mohammed SI. Breast cancer burden in central Sudan. *Int J Womens Health* (2010) 2:77–82. doi: 10.2147/IJWH.S8447
- Awadelkarim KD, Arizzi C, Elamin EOM, Hamad HMA, De Blasio P, Mekki SO, et al. Pathological, clinical and prognostic characteristics of breast cancer in Central Sudan versus Northern Italy. Implications for breast cancer in Africa. *Histopathology* (2008) 52:445–56. doi: 10.1111/j.1365-2559.2008.02966.x
- Ahmed AAM. Clinicopathological profile of female sudanese patients with locally advanced breast cancer. *Breast Dis.* (2014) 34:131–4. doi: 10.3233/BD-140363
- Katz TA, Huang Y, Davidson NE, Jankowitz RC. Epigenetic reprogramming in breast cancer. From new targets to new therapies. *Ann Med.* (2014) 46:1–12. doi: 10.3109/07853890.2014.923740
- Liu H, Li X, Dong C. Epigenetic and metabolic regulation of breast cancer stem cells. *J Zhejiang Univ Sci B* (2015) 16:10–7. doi: 10.1631/jzus.B1400172
- Lokk K, Modhukur V, Rajashekar B, Mårtens K, Mägi R, Kolde R, et al. DNA methylome profiling of human tissues identifies global and tissue-specific methylation patterns. *Genome Biol.* (2014) 15:R54. doi: 10.1186/gb-2014-15-4-r54
- Fackler MJ, Lopez Bujanda Z, Umbricht C, Teo WW, Cho S, Zhang Z, et al. Novel methylated biomarkers and a robust assay to detect circulating tumor DNA in metastatic breast cancer. *Cancer Res.* (2014) 74:2160–70. doi: 10.1158/0008-5472.CAN-13-3392
- Bock C. Epigenetic biomarker development. *Epigenomics* (2009) 1:1–14. doi: 10.2217/epi.09.6
- Bediaga NG, Acha-Sagredo A, Guerra I, Viguri A, Albaina C, Ruiz Diaz I, et al. DNA methylation epigenotypes in breast cancer molecular subtypes. *Breast Cancer Res.* (2010) 12:R77. doi: 10.1186/bcr2721
- Rhee J-K, Kim K, Chae H, Evans J, Yan P, Zhang B-T, et al. Integrated analysis of genome-wide DNA methylation and gene expression profiles in molecular subtypes of breast cancer. *Nucleic Acids Res.* (2013) 41:8464–74. doi: 10.1093/nar/gkt643
- Park SY, Kwon HJ, Choi Y, Lee HE, Kim S-W, Kim JH, et al. Distinct patterns of promoter CpG island methylation of breast cancer subtypes are associated with stem cell phenotypes. *Mod Pathol.* (2011) 25:185–96. doi: 10.1038/modpathol.2011.160
- List M, Hauschild A, Tan Q, Kruse TA. Classification of breast cancer subtypes by combining gene expression and DNA methylation data. *J Integr Bioinform.* (2014) 11:236. doi: 10.1515/jib-2014-236
- Stefansson OA, Moran S, Gomez A, Sayols S, Arribas-Jorba C, Sandoval J, et al. A DNA methylation-based definition of biologically distinct breast cancer subtypes. *Mol Oncol.* (2014) 9:1–14. doi: 10.1016/j.molonc.2014.10.012
- Lechner M, Fenton T, West J, Wilson G, Feber A, Henderson S, et al. Identification and functional validation of HPV-mediated hypermethylation in head and neck squamous cell carcinoma. *Genome Med.* (2013) 5:15. doi: 10.1186/gm419
- Li L, Zhang Y, Guo BB, Chan FKLL, Tao Q. Oncogenic induction of cellular high CpG methylation by Epstein–Barr virus in malignant epithelial cells. *Chin J Cancer* (2014) 33:604–8. doi: 10.5732/cjc.014.10191
- Gulley ML. Genomic assays for Epstein–Barr virus-positive gastric adenocarcinoma. *Exp Mol Med.* (2015) 47:e134–12. doi: 10.1038/emmm.2014.93
- Consortium RE, Kundaje A, Meuleman W, Ernst J, Bilenyk M, Yen A, et al. Integrative analysis of 111 reference human epigenomes. *Nature* (2015) 518:317–30. doi: 10.1038/nature14248
- Berman H, Zhang J, Crawford YG, Gauthier ML, Fordyce CA, McDermott KM, et al. Genetic and epigenetic changes in mammary epithelial cells identify a subpopulation of cells involved in early carcinogenesis. *Cold Spring Harb Symp Quant Biol.* (2005) 70:317–27. doi: 10.1101/sqb.2005.70.051
- Dumont N, Crawford YG, Sigaroudinia M, Nagrani SS, Wilson MB, Buehring GC, et al. Research article Human mammary cancer progression model recapitulates methylation events associated with breast premalignancy. *Breast Cancer Res.* (2009) 11:1–17. doi: 10.1186/bcr2457
- Wessely F, Emes RD. Identification of DNA methylation biomarkers from Infinium arrays. *Front Genet.* (2012) 3:161. doi: 10.3389/fgene.2012.00161
- Urrutia G, Laurito S, Marzese DM, Gago F, Orozco J, Tello O, et al. Epigenetic variations in breast cancer progression to lymph node metastasis. *Clin Exp Metastasis* (2015) 32:99–110. doi: 10.1007/s10585-015-9695-4
- Wang D, Yang P-N, Chen J, Zhou X-Y, Liu Q-J, Li H-J, et al. Promoter hypermethylation may be an important mechanism of the transcriptional inactivation of ARRDC3, GATA5, and ELP3 in invasive ductal breast carcinoma. *Mol Cell Biochem.* (2014) 396:67–77. doi: 10.1007/s11010-014-2143-y
- Ruik Y, Imanaka Y, Sato F, Shimizu K, Tsujimoto G. Genome-wide analysis of aberrant methylation in human breast cancer cells using methyl-DNA immunoprecipitation combined with high-throughput sequencing. *BMC Genomics* (2010) 11:137. doi: 10.1186/1471-2164-11-137
- Conway K, Edmiston SN, May R, Kuan PF, Chu H, Bryant C, et al. DNA methylation profiling in the Carolina Breast Cancer Study defines cancer subclasses differing in clinicopathologic characteristics and survival. *Breast Cancer Res.* (2014) 16:450. doi: 10.1186/s13058-014-0450-6
- Marzese DM, Hoon DSB, Chong KK, Gago FE, Orozco JI, Tello OM, et al. DNA methylation index and methylation profile of invasive ductal breast tumors. *J Mol Diagn.* (2012) 14:613–22. doi: 10.1016/j.jmoldx.2012.07.001
- Moarii M, Pinheiro A, Sigal-Zafrani B, Fourquet A, Caly M, Servant N, et al. Epigenomic alterations in breast carcinoma from primary tumor to locoregional recurrences. *PLoS ONE* (2014) 9:e103986. doi: 10.1371/journal.pone.0103986
- Huang W-Y, Hsu S-D, Huang H-Y, Sun Y-M, Chou C-H, Weng S-L, et al. MethHC: a database of DNA methylation and gene expression in human cancer. *Nucleic Acids Res.* (2014) 43:D856–61. doi: 10.1093/nar/gku1151
- Fleischer T, Frigessi A, Johnson KC, Edvardsen H, Touleimat N, Klajic J, et al. Genome-wide DNA methylation profiles in progression to *in situ* and invasive carcinoma of the breast with impact on gene transcription and prognosis. *Genome Biol.* (2014) 15:1–13. doi: 10.1186/PREACCEPT-233349012841587
- Tommasi S, Karm DL, Wu X, Yen Y, Pfeifer GP. Methylation of homeobox genes is a frequent and early epigenetic event in breast cancer. *Breast Cancer Res.* (2009) 11:R14. doi: 10.1186/bcr2233
- Reinhardt D, Cruickshanks HA, Mjoseng HK, Rmcpersonedacuk RCM, Lentini A, Donnchadunicanigmedacuk DSD, et al. Rapid reprogramming of epigenetic and transcriptional profiles in mammalian culture systems. *Genome Biol.* (2015) 16:11. doi: 10.1186/s13059-014-0576-y
- Kalari S, Pfeifer GP. Identification of driver and passenger DNA methylation in cancer by epigenomic analysis. In: *Advances in Genetics [Internet]*. 1st ed. Vol. 70. Elsevier Inc. (2010) p.277–308. doi: 10.1016/B978-0-12-380866-0.00010-1
- Hon GC, Hawkins RD, Caballero OL, Lo C, Lister R, Pelizzola M, et al. Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* (2012) 22:246–58. doi: 10.1101/gr.125872.111
- Stirzaker C, Zotenko E, Song JZ, Qu W, Nair SS, Locke WJ, et al. Prognostic value. *Nat Commun.* (2015) 6:5899. doi: 10.1038/ncomms6899
- Gonzalez ME, Moore HM, Li X, Toy KA, Huang W, Sabel MS, et al. EZH2 expands breast stem cells through activation of NOTCH1 signaling. *Proc Natl Acad Sci USA.* (2014) 111:3098–103. doi: 10.1073/pnas.1308953111
- Cho J-H, Dimri M, Dimri GP. A positive feedback loop regulates the expression of polycomb group protein BMI1 via WNT signaling pathway. *J Biol Chem.* (2012) 288:3406–18. doi: 10.1074/jbc.M112.422931
- van Vlerken LE, Kiefer CM, Morehouse C, Li Y, Groves C, Wilson SD, et al. EZH2 is required for breast and pancreatic cancer stem cell maintenance and can be used as a functional cancer stem cell reporter. *Stem Cells Transl Med.* (2013) 2:43–52. doi: 10.5966/sctm.2012-0036
- Polytarchou C, Iliopoulos D, Struhl K. An integrated transcriptional regulatory circuit that reinforces the breast cancer stem cell state. *Proc Natl Acad Sci USA.* (2012) 109:14470–5. doi: 10.1073/pnas.1212811109
- Reynold M, Turcan S, Giri D, Kannan K, Walsh LA, Viale A, et al. Remodeling of the methylation landscape in breast cancer

- metastasis. *PLoS ONE* (2014) 9:e103896. doi: 10.1371/journal.pone.0103896
41. Shi P, Feng J, Chen C. Hippo pathway in mammary gland development and breast cancer. *Acta Biochim Biophys Sin.* (2014) 47:53–59. doi: 10.1093/abbs/gmu114
  42. Hiemer SE, Szymaniak AD, Varelas X. The transcriptional regulators TAZ and YAP direct transforming growth factor  $\beta$ -induced tumorigenic phenotypes in breast cancer cells. *J Biol Chem.* (2014) 289:13461–74. doi: 10.1074/jbc.M113.529115
  43. Lin L, Sabnis AJ, Chan E, Olivás V, Cade L, Pazarentzos E, et al. The Hippo effector YAP promotes resistance to RAF- and MEK-targeted cancer therapies. *Nat Genet.* (2015) 47:250–6. doi: 10.1038/ng.3218
  44. Keren-Paz A, Emmanuel R, Samuels Y. YAP and the drug resistance highway. *Nat Genet.* (2015) 47:193–4. doi: 10.1038/ng.3228
  45. Chiang K-C, Yeh C-N, Chung L-C, Feng T-H, Sun C-C, Chen M-F, et al. WNT-1 inducible signaling pathway protein-1 enhances growth and tumorigenesis in human breast cancer. *Sci Rep.* (2015) 5:8686. doi: 10.1038/srep08686
  46. Xu J, Prospero JR, Choudhury N, Olopade OI, Goss KH.  $\beta$ -catenin is required for the tumorigenic behavior of triple-negative breast cancer cells. *PLoS ONE* (2015) 10:e0117097. doi: 10.1371/journal.pone.0117097
  47. Meier-abt F, Bentires-alj M, Rochlitz C. Breast cancer prevention. Lessons to be learned from mechanisms of early pregnancy-mediated breast cancer protection. *Cancer Res.* (2015) 75:803–8. doi: 10.1158/0008-5472.CAN-14-2717
  48. Naumov VA, Generozov EV, Zaharjevskaya NB, Matushkina DS, Larin AK, Chernyshov SV, et al. Genome-scale analysis of DNA methylation in colorectal cancer using Infinium HumanMethylation450 BeadChips. *Epigenetics* (2013) 8:921–34. doi: 10.4161/epi.25577
  49. Zhu W, Qin W, Hewett JE, Sauter ER. Quantitative evaluation of DNA hypermethylation in malignant and benign breast tissue and fluids. *Int J Cancer* (2010) 126:474–82. doi: 10.1002/ijc.24728
  50. Zhu W, Qin W, Zhang K, Rottinghaus GE, Chen Y-C, Kliethermes B, et al. Trans-resveratrol alters mammary promoter hypermethylation in women at increased risk for breast cancer. *Nutr Cancer* (2012) 64:393–400. doi: 10.1080/01635581.2012.654926
  51. Qin W, Zhang K, Clarke K, Weiland T, Sauter ER. Methylation and miRNA effects of resveratrol on mammary tumors vs. normal tissue. *Nutr Cancer* (2014) 66:270–7. doi: 10.1080/01635581.2014.868910
  52. Yahia ZA, Adam AA, Elgizouli M, Hussein A, Masri MA, Kamal M, et al. Epstein Barr virus. a prime candidate of breast cancer aetiology in Sudanese patients. *Infect Agent Cancer* (2014) 9:9. doi: 10.1186/1750-9378-9-9
  53. Masri MA, Abdel Seed NM, Fahal AH, Romano M, Baralle F, El Hassam AM, et al. Minor role for BRCA2 (exon11) and p53 (exon 5-9) among Sudanese breast cancer patients. *Breast Cancer Res Treat.* (2002) 71:145–7. doi: 10.1023/A:1013807830329
  54. Lin D-C, Meng X, Hazawa M, Nagata Y, Varela AM, Xu L, et al. The genomic landscape of nasopharyngeal carcinoma. *Nat Genet.* (2014) 46:866–71. doi: 10.1038/ng.3006
  55. Uozaki H, Fukayama M. Epstein-Barr virus and gastric carcinoma—viral carcinogenesis through epigenetic mechanisms. *Int J Clin Exp Pathol.* (2008) 1:198–216.
  56. Niller HH, Szenthe K, Minarovits J. Epstein-Barr virus-host cell interactions. an epigenetic dialog? *Front Genet.* (2014) 5:367. doi: 10.3389/fgene.2014.00367
  57. Zhao J, Liang Q, Cheung K-F, Kang W, Lung RWM, Tong JHM, et al. Genome-wide identification of Epstein-Barr virus-driven promoter methylation profiles of human genes in gastric cancer cells. *Cancer* (2012) 119:1–9. doi: 10.1002/cncr.27724
  58. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol.* (2010) 28:495–501. doi: 10.1038/nbt.1630
  59. Wizard<sup>®</sup> Genomic DNA Purification Kit Protocol. Available online at: <https://www.promega.com/resources/protocols/technical-manuals/0/wizard-genomic-dna-purification-kit-protocol/>
  60. Infinium HumanMethylation450 BeadChip Kit. Available online at: [http://www.illumina.com/products/methylation\\_450\\_beadchip\\_kits.html](http://www.illumina.com/products/methylation_450_beadchip_kits.html)
  61. Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the infinium methylation 450 K technology. *Epigenomics* (2011) 3:771–84. doi: 10.2217/epi.11.105
  62. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol.* (2004) 3. doi: 10.2202/1544-6115.1027
  63. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, et al. Galaxy. A platform for interactive large-scale genome analysis. *Genome Res.* (2005) 15:1451–5. doi: 10.1101/gr.4086505
  64. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, et al. Galaxy. a web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol.* (2010) Chapter 19:Unit 19.10.1-21. doi: 10.1002/0471142727.mb1910s89
  65. Goecks J, Nekrutenko A, Taylor J. Galaxy. a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* (2010) 11:R86. doi: 10.1186/gb-2010-11-8-r86
  66. Wu G, Feng X, Stein L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biol.* (2010) 11:R53. doi: 10.1186/gb-2010-11-5-r53
  67. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape. A software environment for integrated models of biomolecular interaction networks. *Genome Res.* (2003) 13:2498–504. doi: 10.1101/gr.1239303

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Abdallah, Algizouli, Suliman, Abdulrahman, Koko, Fessahaye, Shakir, Fahal, Elhassan, Ibrahim and Mohamed. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.