EJMO

DOI: 10.14744/ejmo.2023.80452
EJMO 2023;7(4):371–387

Research Article

Identification of Potentially Therapeutic Target Genes in Metastatic Breast Cancer via Integrative Network Analysis

Hazel Jing Yi Leong,[1] Hao Dong Tan,[1] Wei Hsum Yap,[1] Adeline Yoke Yin Chia,[1,2] Serena Zacchigna,[3] Yin-Quan Tang[1,2]

[1]School of Biosciences, Faculty of Health and Medical Sciences Taylor's University, Subang Jaya, Malaysia
[2]Medical Advancement for Better Quality of Life Impact Lab, Taylor's University, Subang Jaya, Selangor Darul Ehsan, Malaysia
[3]Department of Cardiovascular Biology, International Centre for Genetic Engineering and Biotechnology, Trieste, Italy

Abstract

Objectives: Metastatic breast cancer (MBC) represents a significant cause of morbidity and mortality in patients. However, the molecular mechanism of the disease among MBC patients remains elusive.

Methods: An integrated characterization was performed using two independent datasets of normal and malignant breast tissues (GSE29431 and GSE12276), and the differentially expressed genes (DEGs) had been analysed. Network system biology based on the protein–protein interaction (PPI) network was performed using STRING, identified hub genes were confirmed by the Cytoscape and Kaplan–Meier survival analysis to study DEGs of overall survival.

Results: The study identified 159 DEGs which includes 54 up-expressed and 105 down-expressed genes, and network analysis indicate that nucleosome assemble, and cell cycle regulation mediate breast cancer metastasis. *HIST1H2BD* and *KMT2A* were recognized as key hub genes regulate cell fate transitions to promote tumor progression and metastasis. Another key hub gene, *ITGB1* could drive metastasis by modulating the cell cycle processes through FAK and AKT pathways. KM survival analysis revealed these three hub genes were closely correlated with the overall survival of patients.

Conclusion: This study provides a new deeper insight into better understanding of these hub genes (*HIST1H2BD*, *KMT2A* and *ITGB1*) can potentially be used in novel therapeutic strategies for MBC.

Keywords: Breast cancer, biomarkers, metastasis, Protein-Protein Interaction Network, therapeutic targets

Cite This Article: *Leong HJY, Tan HD, Yap WH, Chia AYY, Zacchigna S, Tang YQ. Identification of Potentially Therapeutic Target Genes in Metastatic Breast Cancer via Integrative Network Analysis. EJMO 2023;7(4):371–387.*

Breast cancer (BC) is commonly diagnosed among the women population, and it has been a matter of concern as a consequence of its high incidence and mortality rate. Despite the remarkable improvement in disease prevention, diagnosis, and treatment over the past decades, metastasis—a process in which cancer cells disseminate throughout the body and colonise distant organs after detaching from their tumour origin, still accounts for the vast majority of deaths in victims.[1-4] Metastatic disease could be manifested *de novo*, in which the metastasis process commenced at the primary detection and the cancer has already developed before diagnosis. Nonetheless, the metastatic disease is oftentimes due to relapse (recurrence), where metastases happen after distinct treatment.[5] It has been reported that the 5-year survival rate for affected patients accounts approximately 26% and patient's health can decline due to the metastasis lesions' invasion to vital organs which can lead to formation of multiple foci that are hard to surgically remove and resistant development to the systemic therapies that are presently accessible. Inevitably,

Address for correspondence: Yin-Quan Tang, MD. School of Biosciences, Faculty of Health and Medical Sciences Taylor's University, Subang Jaya, Malaysia; Medical Advancement for Better Quality of Life Impact Lab, Taylor's University, Subang Jaya, Selangor Darul Ehsan, Malaysia
Phone: +60356295000 **E-mail:** yinquan.tang@taylors.edu.my

Submitted Date: October 07, 2023 **Accepted Date:** November 18, 2023 **Available Online Date:** December 29, 2023

impeding metastasis is of core importance in combating against breast cancer.[1,6]

Recent bioinformatics studies have been trying to pinpoint markers affiliated to metastatic progression by employing gene expression data. However, these gene expression-based markers frequently have low reproducibility across dissimilar datasets. Small sample sizes, disparate experimental platforms, individual differences in gene expression that do not affect metastatic progression, and the limitation of microarray technology's inability to identify changes above the transcription level may all contribute to this.[7] In order to address the instability issue with markers acknowledged by previous studies, additional genomic information, such as pathway-based or various networks have been applied. As genes interconnected with metastasis are commonly biologically connected to one another, network construction and analyses are critical tools to provide a powerful abstraction of intracellular complicated interactions. It has been revealed that genes with that potential to be intricated with a myriad of diseases are disclosed through the inclusion of functional information of protein and protein interactions (PPI) network.[8] Based on the crucial insight that biomarkers may provide functional connection to differentially expressed genes (DEGs) in PPI network, the goal is to identify a group of genes that provide connectivity to DEGs in a PPI network.[7,9] PPIs are vital for biological processes including gene expression, cell proliferation, growth, and apoptosis.[10] A considerable number of studies have implied that aberrant PPIs are the cause of certain aggregation-related diseases, including those that contribute to the advancement and occurrence of cancer. In addition, the PPI network has a portion of highly connected regions with high likelihood of taking part in crucial biological regulation, while on the contrary those nodes with weaker connections do not represent a crucial role in the integrity on the entire network.[11,12] This will be a potent technique for locating hub genes with clinical relevance.[13]

Therefore, this research is performed mainly to discover potential therapeutic targets and further unveil the understanding of the probable key genes and crucial pathways in breast cancer metastasis employing bioinformatic analysis. This is to hopefully provide more accurate therapeutic targets for individualized prevention measures, and improvement of therapeutic efficacy of MBC.

Methods

The experimental flowchart presenting the overall study is depicted graphically in Figure 1.

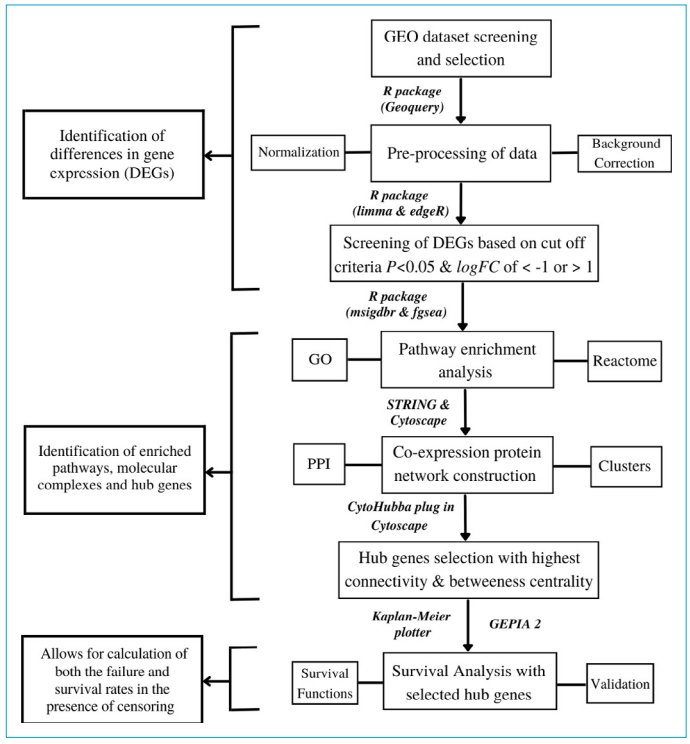

Figure 1. Experimental flow chart presenting the overall study.

Microarray Data Retrieval and Processing

The GEO database was searched using the following criteria: Search terms, 'Breast Cancer' and 'Metastatic Breast Cancer'; study type, 'Expression profiling by array'; sample count, >190. In this study, the microarray dataset of normal and malignant breast tissue (GSE29431) was retrieved from the NCBI Gene Expression Omnibus (GEO) online database (https://www.ncbi.nlm.nih.gov/geo/). On the other hand, the dataset GSE12276 retrieved from the same online database, comprising 204 primary breast cancer tumors with known sites of relapse was used for Kaplan Meier survival analysis due to the provided patients' survival time. Detailed information of the datasets was presented in Table S1.

The R software (version 4.2.1) and Bioconductor packages were used to analyse the microarray dataset. Prior to DEG analysis, pre-processing of the selected dataset which consists of background correction and normalization as well as subsequent gene annotation and unsupervised clustering were conducted in R software using the applicable packages, namely, GEOquery,[14] Tidyverse,[15] and Limma.[16]

Differentially Expressed Genes (DEG) Analysis

The linear models for microarray data (LIMMA) R package in Bioconductor (http://www.bioconductor.org/) were used to perform differential expression gene (DEG) analysis between groups of samples involving breast cancer and normal breast specimens. Significant DEGs were acquired

using the principal standards of |log fold change (FC)| > 1 and p-value ≤ 0.05, whereas the upregulated DEGs were considered if the logFC ≥ 1 and logFC ≤ -1 for downregulated DEGs. The results can be visualized in the volcano plot generated using the ggplot2 package of R.[17]

Functional Enrichment Analysis

To provide a comprehensive grasp of the biological information of the DEGs, proteins and their associated pathways in breast cancer metastasis, Gene Ontology (GO) enrichment results of biological process (BP), molecular function (MF), and cellular component (CC) were obtained using the cluster Profile (version 3.14.3) R package. Reactome pathway analysis of the DEGs was also performed using the R package. P-value of ≤ 0.05 was considered statistically significant to present GO and Reactome pathway enrichment analyses. Moreover, the ggplot2 R package (version 1.26.0) was utilized to the top enrichment terms of GO analysis and Reactome pathway analysis.[17]

Protein-Protein Interaction (PPI) Network Construction and Validation

Search Tool for the Retrieval of Interacting Genes (STRING) database (version 11.0) is a biological database that evaluates the relationship among the DEGs from the obtained dataset and identifies PPIs with a selected confident score.[18] In general, there are three types of confidence scores for PPIs: 1) low confidence: score < 0.4, 2) medium confidence: 0.4 < score < 0.7, and 3) high confidence: score > 0.7.[19] The minimum required interaction score was set to medium confidence 0.4 and the organism to Homo sapiens (hsa). However, in this study, a high confidence score of ≥ 0.7 was employed to annihilate PPIs with low probability/significance and obtain more reliable results. Network centrality values for the nodes in the PPI network were determined using the Network Analyzer app.

Hub Genes Identification

Genes showing a significant connection in candidate modules are referred to as hub genes and high connectivity denotes a top 10% ranking for connectivity. Hub genes generally represent an essential role in a biological system as a result of their high interactions with other genes. With their significance in providing information regarding key pathways related to a certain disease, it can be considered important in the search for the prognosis, diagnosis, or treatment applications.[20,21] In this present study, hub genes selected are based on their consistently high centrality values across the three parameters which take in the local scale (degree connectivity) and global scale (closeness centrality and betweenness centrality).[22]

Survival Analysis

Kaplan-Meier plotter was used to plot survival analyses of the top module genes, followed by the employment of Gene Expression Profiling Interactive Analysis 2.0 (GEPIA2) web server for the confirmation of the survival analyses outcomes. In this study, the dataset GSE12276 was utilized to perform survival analysis. Patients were split into groups determined by the expression level of a particular hub gene, and from which, Kaplan Meier estimation, log-rank tests and Cox proportional hazard models using the Survival package in R software were used to quantify the effects of that gene on overall patient survival.[23] A log-rank p-value ≤ 0.05 indicates significant differences. To validate the survival influence of hub genes, the analyses was performed on hub genes that correlated with patient survival in GSE12276. The survival analyses using GEPIA2 online service provided a total of five datasets for breast cancer which includes the following: general dataset, basal-like or triple negative (135 samples), HER2+ non luminal (66 samples), luminal A (415 samples) and luminal B (192 samples).

Results

Identification of DEGs

The results of unsupervised clustering using Multidimensional Scaling (MDS) clustering are shown in Figure S1. The plotting presented two distinctive groups, whereby tumor and non-tumor samples were well-differentiated against one another, while metastatic and non-metastatic samples were less well-differentiated with one another. DEG analysis performed between the samples using the limma microarray analysis pipeline revealed a total of 159 DEGs, of which 54 were up-regulated and 105 were down-regulated. A volcano plot is presented in Figure 2A. The red colour represents high expression, while the blue colour represents low expression.

Functional Enrichment Analysis on DEGs

To understand the functions and pathways in which the identified DEGs are involved, GO and Reactome functional and pathway enrichment analysis were performed. The results were sorted out by a p-value ≤ 0.05, obtaining in a total of 143 pathways, 37 of which from GO: BP, 10 of which from GO: CC, 17 of which from GO: MF, 79 of which from Reactome. The top seven pathways from each database can be visualized in Figure 2B.

Protein-Protein Interaction (PPI) Network Construction

PPI analysis was performed to understand the system-level of functional interactions of the identified DEGs based on

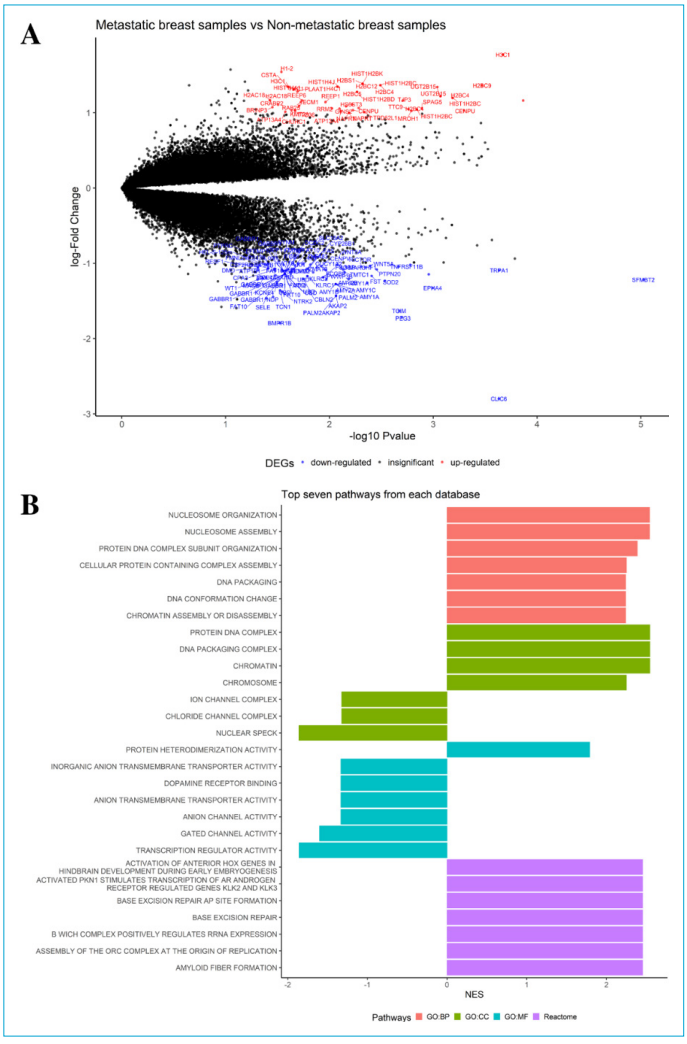

Figure 2. (a) Volcano plotting. Volcano plots were generated using the R package 'ggplot2'. Significant DEG were annotated using the DecideTests function, with "global" setting. **(b)** Functional enrichment analysis. GO biological processes, GO cellular components, and GO molecular functions, and Reactome functional were performed on the DEGs. The results were filtered with p-value ≤ 0.05.

information in the STRING database. The DEGs were uploaded to the STRING database (version 11.0) to produce a differential gene PPI network. Figure 3 presents a network with 51 nodes, in which the node size corresponds to the degree of connectivity and the colour corresponds to the logFC values. The network map data obtained from the STRING database was uploaded to Cytoscape software (version 3.9.1) to further identify the key genes.

Hub Genes Selection and Identification

Parameters which include high degree connectivity, closeness and betweenness centrality were criteria considered while determining the hub genes. Genes that score consistently high values for the mentioned three centrality parameters will be considered as a candidate gene. In

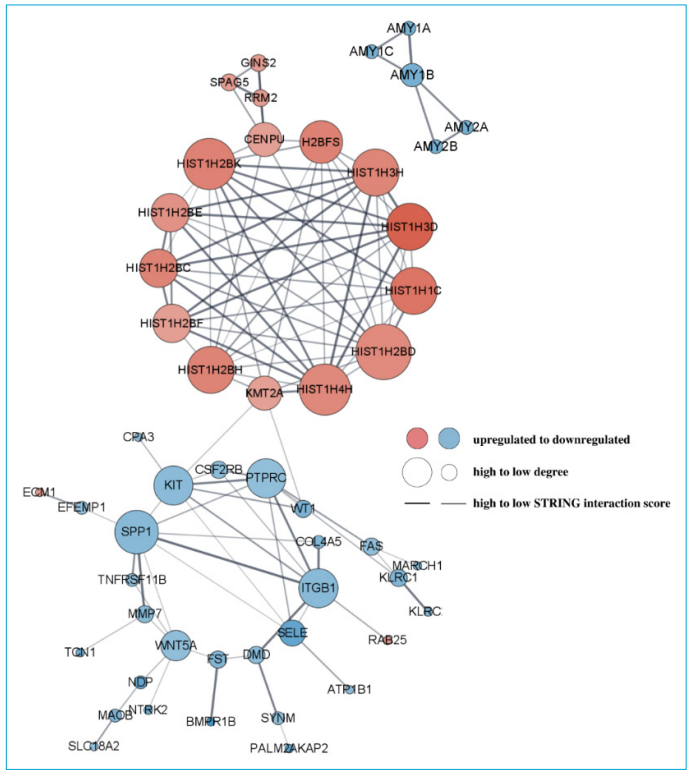

Figure 3. PPI network constructed using STRING. The network presents nodes of genes discovered to be differentially expressed in tumor samples. Small clusters of nodes that were disconnected from the biggest cluster, with two or a smaller number of nodes were removed. The colour of nodes was in accordance with their respective logFC values, in which positive logFC values are indicated in red and negative logFC values are indicated in blue. The nodes' sizes correspond to its respective degree of connectivity.

this study, a total of six hub genes were selected (Table 1). These top six hub genes were then subjected to survival analysis using microarray expression data provided by dataset GSE12276.

Overall Survival Analysis on Hub Genes

The expression levels of the selected hub genes were extracted from the microarray dataset GSE12267 downloaded from the Gene Expression Omnibus (GEO) database. Two groups were formed based on the median gene expression level, with expression that is greater than the median expression value assigned to the high expression group and expression that is less than the median expression value assigned to the low expression group. The correlation of patients' overall survival of the selected six hub genes were analysed using the univariate cox proportional hazard regression (Table 2). Kaplan Meier analysis indicated that three of the six hub genes (*HIST1H2BD*, *ITGB1*, *KMT2A*) were significantly correlated with patients' overall survival (Figure 4).

Table 1. Hub genes selection and identification. The hub genes are selected by determining nodes with consistently high centrality values across degree, closeness and betweenness centrality

	ENSEMBL	Symbol	Degree	Closeness	Betweenness
1	ENS2P00000289316	HIST1H2BD	12	0.3103	0.0924
2	ENSP00000378517	SPP1	9	0.3814	0.3753
3	ENSP00000411355	PTPCR	8	0.3600	0.1811
4	ENSP00000288135	KIT	8	0.4018	0.4559
5	ENSP00000379350	ITGB1	8	0.3659	0.1907
6	ENSP00000436786	KMT2A	7	0.3600	0.4545

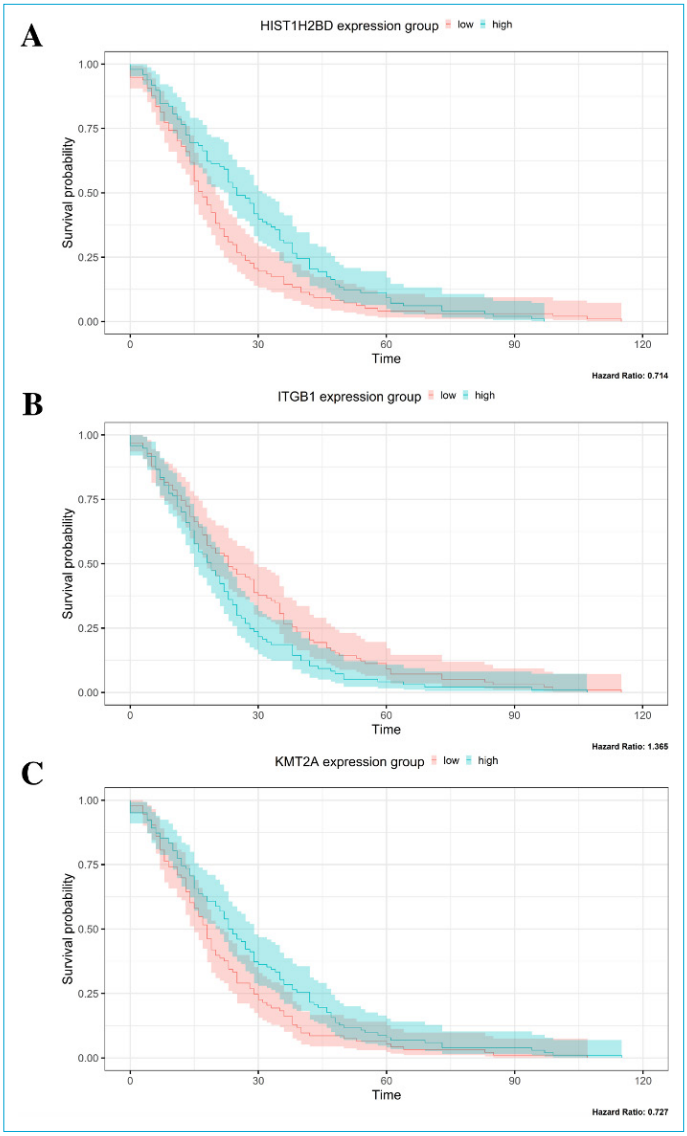

Figure 4. Survival analysis using Kaplan-Meier plotting. The expression levels of HIST1H2BD **(a)**, ITGB1 **(b)**, and KMT2A **(c)** were extracted from GSE12267 and was differentiated into two groups: higher than median expression level indicated in blue; lower than median expression level indicated in red.

Table 2. Log-rank p-values of the hub genes. Selected hub genes were subjected to univariate cox proportional hazard regression, with the log-rank p-values determined. Hub genes indicated with an asterisk* are significant with a log-rank p-value ≤ 0.05.

	Gene	logFC	p
1	*HIST1H2BD**	1.278	0.0194
2	*SPP1*	-1.102	0.0659
3	*PTPCR*	-1.012	0.3530
4	*KIT*	-1.073	0.8629
5	*ITGB1**	-1.096	0.0583
6	*KMT2A**	1.029	0.0269

The three hub genes that were found significantly correlated with patients' overall survival (log-rank p-value ≤0.05) were further subjected to a more thorough survival analysis. Table 3 present approximated hazard ratio using the Univariate Cox proportional hazard regression analysis. *HIST1H2BD* revealed a hazard ratio of 0.712 in which this implies that patients with higher than median levels of *HIST1H2BD* expression are 0.712 times likely to survive as compared with patients with lower than median levels of *HIST1H2BD*. This is similar for *KMT2A* that presents a hazard ratio of 0.727, implying patients with higher than median levels of the gene have 0.727 times more likely to survive as compared with patients with lower than median levels of the gene. In contrast, patients with lower than median levels of *ITGB1* have 1.365 times the possibility to survive as compared to patients with higher than median levels of *ITGB1*.

Table 3. Hazard Ratio and its p-value of the three selected hub genes. Three hub genes with a log-rank p-value ≤ 0.05 were subjected to further survival analysis by calculating their hazard ratio and the ratio's significance.

	Gene	Hazard Ratio	p-value for Hazard Ratio
1	HIST1H2BD	0.712	0.0211
2	ITGB1	1.365	0.0319
3	KMT2A	0.727	0.0276

Validation of Hub Genes using GEPIA2

Validation of the survival analysis conducted on the selected hub genes are also conducted using the online GEPIA2 database (Fig. S2-S4). The analyses for the three hub genes majority showed insignificant values with a log rank p-value of > 0.05 for all the breast cancer datasets. However, only the validation of survival analysis performed on *KMT2A* using the luminal A dataset was found to be significant at a log rank p-value of 0.0024. The hazard ratio from the analysis presented at 2.3, indicating that MBC patients who have a lower than median expression level of *KMT2A* were 2.3 times more likely to survive compared to those who had a higher than median expression level of *KMT2A*. Though this finding was comparable to the survival analysis conducted in this study, the hazard ratio generated using GEPIA2 luminal A dataset showed to be larger than the hazard ratio generated using GSE12267 microarray dataset which is at 0.727 (Table 3). This is suggestive that *KMT2A* might be important in driving metastasis and affect the survival outcome of individuals diagnosed with luminal A breast cancer subtype as to other subtypes.

The validation analyses of hub genes using GEPIA2 in this study presented mostly insignificant values may be due to either the presence of a specific subtype in the selected dataset leading to biasness or censored data present in some breast cancer subtype (eg: HER2+ non luminal) in the database.

Discussion

Metastatic breast cancer (MBC) is a significant cause of mortality and morbidity in patients owing to its fatal outcome and absence of definite cure for the disease.[6,24] Metastasis disease persists the primary culprit in the majority of breast cancer individuals who succumb to their death as it is known to be a complicated pathological mechanism that involves many steps and is governed by various genes as well as signalling pathways.[25] Studies proposed that key dysregulated networks are considerably enriched in crucial breast cancer-related pathways and driver genes. This indicates that the key dysregulated genes may function as driver genes, therapeutic targets, or prognostic indicators. Consequently, it is anticipated that the identification of the key genes and dysregulated pathways will be useful to unveil the mechanism of metastasis in breast cancer.[26] In the present study, three genes namely *HIST1H2BD*, *ITGB1* and *KMT2A* were recognized as potential hub genes.

HIST1H2BD is identified as a hub gene in this study with a significant upregulated expression. This is similar to past studies that revealed the expression of *HIST1H2BD* was remarkably raised in breast tumor cells relative to normal breast cells.[27-29] In patients with luminal A, HER2+, and normal-like subtypes of breast cancer, primary tumor expression of *HIST1H2BD* was associated with recurrence-free survival. In contrast, in TNBC patients, primary tumor expression of *HIST1H2BD* was correlated with distant metastasis-free survival. This suggests that TNBC's genesis, maintenance, or advancement may be affected by *HIST1H2BD*.[30] Additionally, in a study conducted by Li and colleagues (2017),[31] the expression levels of *HIST1H2BD* along with *HIST1H2BJ* were found significantly linked to the overall survival of cervical squamous cell cancer patients. As a result, longer patient life was associated with elevated expression of *HIST1H2BD* and *HIST1H2BJ*. From GO biological analysis, 49 pathways (p-value ≤0.05) were found upregulated with positive normalized enrichment score (NES) values ranging from 2.694 to 1.666 (see supplementary Table S2). The constructed PPI network from the study showed subsets of histones family genes were expressed together and functioned together with *HIST1H2BD* as the seed gene in the submodule.[31] These observations were similar to this current study in which *HIST1H2BD* presented the highest degree of connectivity, having strong connections with other histone genes such as *HIST1H2BK* and *HIST1H2BH* (Fig. 3). *HIST1H2BK* overexpression in breast cancer cells were reported to activate the LIFR-JAK1-STAT3 signalling pathway in which this results in the development of aggressiveness in breast cancer.[32] Singh et al. (2018) revealed the gene's direct correlation to metastasis and its elevated levels in highly metastatic cell lines.[33] As for *HIST1H2BH*, it also appeared to be remarkably higher in more aggressive breast cancer subtypes such as TNBC compared to normal breast cells.[30] Taken together, it may be deduced that *HIST1H2BD* might drive metastasis in breast cancer jointly with other highly connected histones genes that are not hub genes.

HIST1H2BD gene belongs under the histone H2B family. It has been reported that histone H2B family of variants appear as potential mediators of drug sensitivity and resistance in cancer.[27,34] H2B forms a (H2A-H2B)-2 tetramer and comparatively to H3 and H4, this tetramer and its component dimers can interchange in and out of the nucleosome with ease in which this suggests that the changes on H2A and H2B are less likely to be preserved in chromatin. As a result, alterations on H2A/H2B have received less attention in the field of epigenetics than those on H3 and H4. However, numerous studies indicate that chromatin dynamics may be impacted by changes to H2B.[31,35] For instance, a few studies showed a strong correlation between a reduction in H2Bub1 levels and the development of breast cancer supporting H2Bub1's tumor-suppressor function by implicating H2Bub1 in carcinogenesis and DNA repair.[31,36] According to preliminary research, histone genes may be

involved in a variety of human cancers and there was currently no thorough examination of the gene family which may contain predictive biomarkers. This is also suggestive that *HIST1H2BD* may be important in breast cancer metastasis, but further research must be performed to explore its potentiality as a therapeutic target.

Ubiquitination plays a role as a protein degradation mechanism, but it is also required in DNA damage repair and NFκB inflammatory response activation in cells. It is not surprising that cancer cells take advantage of the components of the ubiquitination pathway to stabilize abnormal oncogenic signalling as most proteins undergo ubiquitination as a post-translational alteration in most cell types.[37] In the dynamic process of protein ubiquitination, there are two types of participants that include ubiquitin enzymes (writers), and deubiquitin enzymes (erasers). All histone molecules have the potential to be ubiquitinated, and they have also been linked to cancer. Most identified ubiquitin histone molecules thus far are of the H2A and H2B type.[38] To our dismay, there was limited literature on the role and mechanism of H2B variants in breast cancer progression compared to H2A variants.[27,38,39] However, as substantial amounts of studies have displayed H2B variants' significance involvement in breast cancer metastasis, this is suggestive that H2B histone variants including *HIST1H2BD* could be a potential focus and an area with much further research needed.

Integrin β1 (*ITGB1*) is the representative member of the integrin subfamily, and it has 12 α-subunits that can form heterodimers.[40] Integrins, which connect the extracellular matrix with the intracellular cytoskeleton to mediate cell adhesion, survival, differentiation, and migration by a variety of intracellular signalling pathways, are heterodimeric cell-surface receptors made up of the α and β subunits. β1 integrin can dimerize with other α integrins such as α2β1 which have been observed to promote tumor metastasis through epithelial-mesenchymal transition (EMT) and mesenchymal-epithelial transition (MET) regulation. Once a secondary metastasis has formed, breast cancer cells frequently go through a MET in order to establish and multiply at the secondary site.[41] Breast cancer cells frequently go through an EMT in order to migrate, invade, and spread from the primary site. The regulation of EMT and MET by α2 and β1 integrins and their downstream signalling components has been demonstrated to change the behaviour of cancer cells.[42] The kinases FAK and Src regulate integrin-mediated cell adhesion and migration during EMT; as epithelial cancer progresses, Src activation and FAK downregulation increase the migratory capacity and intercellular contact suppression.[43] Because it promotes cell-cell and cell-extracellular matrix contacts to mediate the survival, differentiation, angiogenesis, and invasion of cancer cells, *ITGB1* is recognized as the most significant member of the integrin family.[44,45]

ITGB1 has been reported by several past studies that its high expression is strongly correlated with breast cancer metastasis. As the gene plays vital roles in breast cancer cell motility and proliferation, its high expression has been seen in aggressive tumors such as TNBC to drive metastasis. From GO biological analysis, 4 pathways (p-value ≤0.05) were found upregulated with positive normalized enrichment score (NES) values ranging from 2.6150 to 1.666 (see supplementary Table S3). *ITGB1* functions as a signal transducer in the PI3K/Akt and p130Cas/paxillin/JNK signalling pathways, which control survival and proliferation.[46] Besides, studies have shown that *ITGB1* promotes tumor growth in breast cancer via enhancing EMT. In BT549 and Hs578T breast cancer cells, the knockdown of *ITGB1* somewhat raised the expression of E-cadherin and decreased that of N-cadherin, fibronectin, and vimentin.[47] The study conducted by Klahan et al., (2016) discovered that inhibiting *ITGB1* markedly decreased calcium (Ca^{2+}) influx through the stored-operated calcium (SOC) channel and the study hypothesized that *ITGB1* can reduce migration invasion in TNBC by regulating Ca^{2+} influx through the SOC channel.[48]

Hub genes are known to have many interactions with other genes to work hand in hand to elicit gene regulation and certain biological processes. For instance, the focal adhesion pathway is facilitated by *ITGB1* and *FN1*, and these genes mediate the interaction of the ECM by dysregulating the focal adhesion pathway as seen in the transformation of ductal carcinoma in situ (DCIS) to invasive breast cancer.[49] Similarly, *ITGB1* in this current study was observed to interact strongly with nearby non-hub genes in the PPI such as *SPP1* and *COL4A5* (Fig. 3). Zeng and colleagues (2018) discovered *SPP1* promotes progression in ovarian cancer through *ITGB1*/FAK/AKT pathway and silencing *SPP1* subsequently inhibits the particular pathway.[50] Additionally, a downregulated expression of *COL4A5* was reported in colorectal cancer and it was also shown to promote lung cancer progression.[51,52] It was proposed that the COL4As might influence integrin-mediated signalling pathways and adhesion-related pathways, thereby controlling the downstream of the Akt pathway. The proliferation and invasion of gastric cancer may be aided by the activation of the Akt pathway.[53] These similar findings were observed in this present study whereby *ITGB1* was noticed to have a strong correlation to the neighbour non-hub gene *COL4A5*. It can be hypothesized that *ITGB1* interacts with neighbouring genes to drive metastasis in breast cancer.

Accumulated findings displayed that high *ITGB1* expression was notably correlated with poor overall survival in several malignancies including breast cancer, gastric cancer, colorectal cancer, and prostate cancer to name a few.[46,54] However, interestingly, *ITGB1* was observed to be downregulated in this present study. This discovery is relatively similar to a few studies that have reported inconsistency in the prognostic significance of *ITGB1* expression in cancer patients. The study by Sun and colleagues (2018) have reported that downregulated *ITGB1* expression is affiliated with more aggressive breast cancer subtypes.[55] In addition, they also disclosed no correlation between higher *ITGB1* expression and breast cancer N (cancer spread to nearby lymph node) stage, tumor grade, T (tumor size) stage, ER, PR, or HER. Interpreting this outcome should be done with care. Depending on the cancer subtype, the prognostic significance and correlation between *ITGB1* expression and the clinicopathological attributes of breast may differ. Nonetheless, statistical bias may present due to the limited sample size used in the particular study. Evidence also showed bone metastatic tumor cells have a decreased *ITGB1* expression while expressing high ITGB3 levels, indicating that integrin switching is taking place in the bone microenvironment. According to studies, inactivation of *ITGB1* results in ITGB3 switching and TGF-β induced breast cancer progression.[56] These past literature findings may indicate that *ITGB1* expression level is probably influenced by various factors such as different cancer subtypes and site of metastasis. Nonetheless, survival analysis in this present study revealed an elevated *ITGB1* expression is significantly associated with a worse survival outcome in affected individuals. This discovery is aligned with many previous studies.[48,55]

There has been a lot of interest in *ITGB1*'s role in the malignant phenotypes of cancer. According to earlier research, *ITGB1* has been shown to influence the resistance to chemotherapy and radiation by promoting cell survival and inhibiting apoptosis in various human malignancies. The effectiveness of *ITGB1* inhibitors in the treatment of resistant malignancies and advanced metastatic illness has been demonstrated in a number of animal models. Thus, *ITGB1* could be considered as a vital therapeutic target and carries substantial significance in cancer patients.[46] However, *ITGB1*'s prognostic value remains controversial and this is suggestive that further investigations regarding the gene's expression and role in driving metastasis in breast cancer should be performed.

KMT2A, sometimes referred to as mixed-lineage leukemia (*MLL*), is a transcriptional coactivator that controls the expression of certain genes throughout haematopoiesis and early development.[57,58] KMT2 family proteins modify DNA accessibility and chromatin architecture by methylating lysine 4 on the histone H3 tail (H3K4) in significant regulatory areas of the genome. Acute leukemia is produced by recurrent chromosomal translocations and translocation-associated gene fusions, both of which include *KMT2A*.[59] Besides, *KMT2A* plays significant roles in solid tumors including breast, colon, lung, bladder and endometrial despite rearrangements in leukemia and lymphoma.[57,60] Several literatures disclosed that *KMT2A* upregulation plays a role in driving cancer advancement, which this aligned with the findings in this present study. Many studies have suggested *KMT2A* acts as a crucial factor in vasculogenesis, hypoxia signalling, and tumor growth. It was disclosed that *KMT2A* along with hypoxia-inducible factor α (HIF1α) were overexpressed in tumor areas with low oxygen levels and this resulted in angiogenesis and tumor progression. A knockdown in *KMT2A* in cervical and breast cancer studies observed a reduction in the expression of HIF1α and vascular growth factor (VEGF) which influenced angiogenesis and suppressed subsequent tumor growth.[60,61] Other studies involving cervical and melanoma cell lines also demonstrated the functional importance of *KMT2A* as the knockdown of the gene resulted in impediment of cell viability and cell migration as well as induced cell apoptosis.[62,63] Additionally, the elevated expression of KTM2A was discovered in solid tumors with gain of function (GOF) mutations of TP53.[64] In these tumors, p53 mutants bound to and up expressed the *KMT2A*, KMT2D, and acetyltransferase MOZ (*KAT6A*) genes, enhancing global H3K4 methylation and histone acetylation as well as upregulating the *KMT2A* target genes including the *HOXA* gene cluster.[60,61] *KMT2A* is revealed to be vital for the cancer phenotype of cells carrying GOF p53 mutants and the gene is shown responsible for promoting cancer growth. These indications were evidenced by the fact that the *KMT2A* knockdown or pharmacological inhibition of *KMT2A* was adequate to limit tumor development.

From GO biological analysis, 6 pathways (p-value ≤0.05) were found upregulated with positive normalized enrichment score (NES) values ranging from 2.459 to 1.987 (see supplementary Table S3). *KMT2A* was observed to interact strongly with other gene in the PPI such as *KIT* (Fig. 3). Receptor-Tyrosine Kinase (RTK) signalling is one of the crucial signal transduction pathways in the development of cancer. Cell activities such as proliferation, differentiation, survival, and angiogenesis all depend on this route. It has been reported that *KIT* gene is essential in this pathway.[65] The study by Rahimi and colleagues (2020) displayed *KIT* may serve as a possible target for cancer therapy as an inhibition of the gene are likely to suppress angiogenesis,

migration, and advancement n in metastatic tumors.[66] In addition, *KIT* was also observed to have strong interactions with non-hub gene *SPP1* and hub gene *ITGB1*.This ultimately suggests the identification hub genes will aid in combating meta progression through these crucial neighbouring gene and pathways interaction.

High *KMT2A* is related to a reduced possibility of recurrence-free survival and a worse overall survival compared to a lower KTM2A expression.[67] Interestingly, survival analysis of *KMT2A* in this present study displays a different finding in which up-expressed *KMT2A* was associated with better overall survival and vice versa. *KMT2A* has been studied extensively and there is growing proof suggesting it has a special role in the progression of cancer. Recurrent translocations in leukemias were found to influence the dominant cancer gene *KMT2A*.[68] However, according to recent investigations, *KMT2A* may have a recessive function in some solid tumors including gastric cancer. In some cancer types, *KMT2A* expression was observed to be up-or downregulated. While *KMT2A* expression showed association with a good prognosis in certain cancer types, *KMT2A* was shown to be decreased in tumors on the other hand.[69] Rabello et al. (2013) disclosed a subtle higher expression of *MLL* genes in less aggressive and no metastasis breast cancer cell line and suggest the decreased expression of *MLL* family genes as disease progresses.[70] Specifically, compared to normal samples, the expression of the *MLL* gene was slightly lower in tumor samples. *MLL* was downregulated in every cancer cell line compared to the collection of normal samples. Thus, *KMT2A*'s function as a tumor promoter should be further evaluated and considered.

Figure 5 encapsulates all of the pathways and its mediators, in which the hub genes may drive or impede metastasis. The figure contains all the mechanisms that were suggested by the results of the present study and had been proven by prior studies in regards to their relevance with cancer metastasis. It can be suggested that the respective hub genes drive metastasis with the assistance of other non-hub genes. The interactions between the genes collectively stimulate metastasis progression through mechanisms such as cell differentiation, proliferation, invasion, migration, resistance, and angiogenesis.

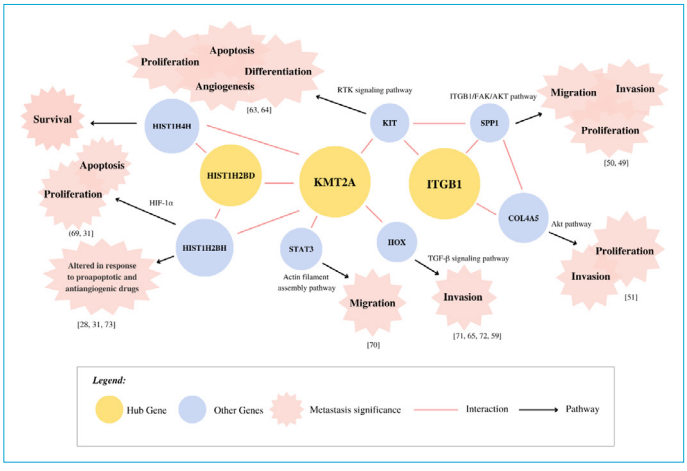

Figure 5. Schematic diagram presenting an overview mechanism of identified hub genes in promoting metastasis.

Conclusion

Based on integrated bioinformatic analysis, the current research has discovered three hub genes (*HIST1H2BD*, *ITGB1*, and *KMT2A*) that are associated with the development and advancement of breast cancer into its metastatic state. The under and over-expression of these hub genes in the metastatic breast cancer tissues as demonstrated in database analysis indicates poor clinical outcome in affected individuals. These findings suggest that thorough research into these hub genes will aid in enhancing the understanding of the pathophysiology and progression in breast cancer to its metastatic state.

Disclosures

Acknowledgements: This work was supported by the School of Biosciences, Faculty of Health and Medical Sciences, Taylor's University, Malaysia.

Research Involving Human Participants and/or Animals: This article does not contain any studies with human participants or animals performed by any of the authors. The microarray dataset used during the present study is available from the NCBI GEO (https://www.ncbi.nlm.nih.gov/geo/) repository, and the reference numbers (GSE29431 and GSE12276)) to access this data are given in the manuscript.

Ethics Committee Approval: The study was approved by the Local Ethics Committee.

Peer-review: Externally peer-reviewed.

Conflict of Interest: None declared.

Fundings: This research was supported by Ministry of Higher Education (MOHE) Malaysia through Fundamental Research Grant Scheme (FRGS/1/2020/SKK0/TAYLOR/02/2). This research was supported by Taylor's University through the Taylor's Internal Research Grant Scheme – Impact Lab Grant (TIRGS-ILG) (TIRGS-ILG/1/2023/SOB/001).

Authorship Contributions: Concept – Y.Q.T.; Design – H.J.Y.L., H.D.T., Y.Q.T.; Supervision – W.H.Y., A.Y.Y.C., S.Z., Y.Q.T.; Data collection &/or processing – H.D.T., H.J.Y.L.; Analysis and/or interpretation – H.J.Y.L., H.D.T., Y.Q.T.; Writing – H.J.Y.L., H.D.T., Y.Q.T.; Review and final revision approval – H.D.T., H.J.Y.L., W.H.Y., A.Y.Y.C., S.Z., Y.Q.T.

References

1. Jin X, Mu P. Targeting breast cancer metastasis. Breast Cancer Res Treat 2015;9:23–34.
2. Tan C, Zuo F, Lu M, Chen S, Tian Z, Yong Hu. Identification of potential genes correlated with breast cancer metastasis and prognosis. All Life 2022;15:126–33.
3. Lamba M, Munjal G, Gigras Y. Computational studies on breast cancer analysis. J Stat Manag Syst 2020;23:999–1009.
4. Abbas N, Tan HD, Goh BH, Yap WH, Tang YQ. In-silico study of anticancer and antimicrobial peptides derived from cycloviolacin O2 (CyO2). Biointerface Res Appl Chem 2023;13:437.
5. Riggio AI, Varley KE, Welm AL. The lingering mysteries of metastatic recurrence in breast cancer. Br J Cancer 2021;124:13–26.
6. Peart O. Metastatic breast cancer. Radiol Technol 2017;88:519–39.
7. Jahid MJ, Ruan J. A Steiner tree-based method for biomarker discovery and classification in breast cancer metastasis. BMC Genomics 2012;13(Suppl 6):8.
8. Engin HB, Guney E, Keskin O, Oliva B, Gursoy A. Integrating structure to protein-protein interaction networks that drive metastasis to brain and lung in breast cancer. PLoS One 2013;8:e81035.
9. Tan HD, Leong HJY, Yap WH, Chia AYY, Tang YQ. Network analysis in the identification of genes conferring metastatic potential in hepatocellular carcinoma. Eurasian J Med Oncol 2022;6:364–80.
10. Xu J, Li Y. Discovering disease-genes by topological features in human protein-protein interaction network. Bioinformatics Oxford 2006;22:2800–5.
11. He X, Zhang J. Why do hubs tend to be essential in protein networks? PLoS Genet 2006;2:e88.
12. Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. Nature 2001;411:41–2.
13. Xiong Y, You W, Wang R, Peng L, Fu Z. Prediction and validation of hub genes associated with colorectal cancer by integrating PPI network and gene expression data. Biomed Res Int 2017;2017:2421459.
14. Davis S, Meltzer PS. GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics 2007;23:1846–7.
15. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Source Softw 2019;4:1686.
16. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
17. Ito K, Murphy D. Application of ggplot2 to pharmacometric graphics. CPT Pharmacometrics Syst Pharmacol 2013;2:e79.
18. Doncheva NT, Morris JH, Gorodkin J, Jensen LJ. Cytoscape StringApp: Network analysis and visualization of proteomics data. J Proteome Res 2019;18:623–32.
19. Assenov Y, Ramírez F, Schelhorn SE, Lengauer T, Albrecht M. Computing topological parameters of biological networks. Bioinformatics 2008;24:282–4.
20. Mahapatra S, Bhuyan R, Das J, Swarnkar T. Integrated multiplex network-based approach for hub gene identification in oral cancer. Heliyon 2021;7:e07418.
21. Yu D, Lim J, Wang X, Liang F, Xiao G. Enhanced construction of gene regulatory networks using hub gene information. BMC Bioinformatics 2017;18:186.
22. Liu J, Hua P, Hui L, Zhang LL, Hu Z, Zhu YW. Identification of hub genes and pathways associated with hepatocellular carcinoma based on network strategy. Exp Ther Med 2016;12:2109–19.
23. Therneau T. A package for survival analysis in R. Available at: https://cran.r-project.org/web/packages/survival/vignettes/survival.pdf. Accessed Dec 21, 2023.
24. Eng LG, Dawood S, Sopik V, Haaland B, Tan PS, Bhoo-Pathy N, et al. Ten-year survival in women with primary stage IV breast cancer. Breast Cancer Res Treat 2016;160:145–52.
25. Song X, Wei C, Li X. The signaling pathways associated with breast cancer bone metastasis. Front Oncol 2022;12:855609.
26. Huo Y, Li X, Xu P, Bao Z, Liu W. Analysis of breast cancer based on the dysregulated network. Front Genet 2022;13:856075.
27. Nayak SR, Harrington E, Boone D, Hartmaier R, Chen J, Pathiraja TN, et al. A role for histone H2B variants in endocrine-resistant breast cancer. Horm Cancer 2015;6:214–24.
28. Amjad E, Asnaashari S, Sokouti B, Dastmalchi S. Systems biology comprehensive analysis on breast cancer for identification of key gene modules and genes associated with TNM-based clinical stages. Sci Rep 2020;10:10816.
29. Arimura Y, Ikura M, Fujita R, Noda M, Kobayashi W, Horikoshi N, et al. Cancer-associated mutations of histones H2B, H3.1, and H2A.Z.1 affect the structure and stability of the nucleosome. Nucleic Acids Res 2018;46:10007–18.
30. Shahan M. Differential expression of histone cluster 1, H2bd in triple negative breast cancer. Available at: https://osf.io/preprints/osf/5fcwd. Accessed Dec 21, 2023.
31. Li X, Tian R, Gao H, Yang Y, Williams BRG, Gantier MP, et al. Identification of a histone family gene signature for predicting the prognosis of cervical cancer patients. Sci Rep 2017;7:16495.
32. Liu W, Xu Z, Zhou J, Xing S, Li Z, Gao X, et al. High levels of HIST1H2BK in low-grade glioma predicts poor prognosis: a study using CGGA and TCGA data. Front Oncol 2020;10:627.
33. Singh R, Bassett E, Chakravarti A, Parthun MR. Replication-dependent histone isoforms: A new source of complexity in chromatin structure and function. Nucleic Acids Res 2018;46:8665–78.
34. Braunstein M, Liao L, Lyttle N, Lobo N, Taylor KJ, Krzyzanowski PM, et al. Downregulation of histone H2A and H2B pathways is associated with anthracycline sensitivity in breast cancer.

Breast Cancer Res 2016;18:16.
35. Wu Y, Gu Y, Guo S, Dai Q, Zhang W. Expressing status and correlation of ARID1A and histone H2B on breast cancer. BioMed Res Int 2016;2016:7593787.
36. Cole AJ, Clifton-Bligh R, Marsh DJ. Histone H2B monoubiquitination: Roles to play in human malignancy. Endocr Relat Cancer 2015;22:T19–33.
37. Gallo LH, Ko J, Donoghue DJ. The importance of regulatory ubiquitination in cancer and metastasis. Cell Cycle 2017;16:634–48.
38. Salhia B, Kiefer J, Ross JT, Metapally R, Martinez RA, Johnson KN, et al. Integrated genomic and epigenomic analysis of breast cancer brain metastasis. PLoS One 2014;9:e85448.
39. Nandy D, Rajam SM, Dutta D. A three-layered histone epigenetics in breast cancer metastasis. Cell Biosci 2020;10:52.
40. Kawahara R, Niwa Y, Simizu S. Integrin β1 is an essential factor in vasculogenic mimicry of human cancer cells. Cancer Sci 2018;109:2490–6.
41. Thiery JP. Epithelial-mesenchymal transitions in tumour progression. Nat Rev Cancer 2002;2:442–54.
42. Wafai R, Williams ED, de Souza E, Simpson PT, McCart Reed AE, Kutasovic JR, et al. Integrin alpha-2 and beta-1 expression increases through multiple generations of the EDW01 patient-derived xenograft model of breast cancer-insight into their role in epithelial mesenchymal transition in vivo gained from an in vitro model system. Breast Cancer Res 2020;22:136.
43. Avizienyte E, Frame MC. Src and FAK signalling controls adhesion fate and the epithelial-to-mesenchymal transition. Curr Opin Cell Biol 2005;17:542–7.
44. Barnawi R, Al-Khaldi S, Colak D, Tulbah A, Al-Tweigeri T, Fallatah M, et al. β1 Integrin is essential for fascin-mediated breast cancer stem cell function and disease progression. Int J Cancer 2019;145:830–41.
45. Askari JA, Buckley PA, Mould AP, Humphries MJ. Linking integrin conformation to function. J Cell Sci 2009;122:165–70.
46. Liu QZ, Gao XH, Chang WJ, Gong HF, Fu CG, Zhang W, et al. Expression of ITGB1 predicts prognosis in colorectal cancer: A large prospective study based on tissue microarray. Int J Clin Exp Pathol 2015;8:12802–10.
47. Ren L, Mo W, Wang L, Wang X. Matrine suppresses breast cancer metastasis by targeting ITGB1 and inhibiting epithelial-to-mesenchymal transition. Exp Ther Med 2020;19:367–74.
48. Klahan S, Huang WC, Chang CM, Wong HS, Huang CC, Wu MS, et al. Gene expression profiling combined with functional analysis identify integrin beta1 (ITGB1) as a potential prognosis biomarker in triple negative breast cancer. Pharmacol Res 2016;104:31–7.
49. Pranavathiyani G, Thanmalagan RR, Leimarembi Devi N, Venkatesan A. Integrated transcriptome interactome study of oncogenes and tumor suppressor genes in breast cancer. Genes Dis 2019;6:78–87.
50. Zeng B, Zhou M, Wu H, Xiong Z. SPP1 promotes ovarian cancer progression via Integrin β1/FAK/AKT signaling pathway. Onco Targets Ther 2018;11:1333–43.
51. Xiao Q, Jiang Y, Liu Q, Yue J, Liu C, Zhao X, et al. Minor type IV collagen α5 chain promotes cancer progression through discoidin domain receptor-1. PLoS Genet 2015;11:e1005249.
52. Ikeda K, Iyama K, Ishikawa N, Egami H, Nakao M, Sado Y, et al. Loss of expression of type IV collagen alpha5 and alpha6 chains in colorectal cancer associated with the hypermethylation of their promoter region. Am J Pathol 2006;168:856–65.
53. Zeng X, Wang HY, Wang YP, Bai SY, Pu K, Zheng Y, et al. COL4A family: Potential prognostic biomarkers and therapeutic targets for gastric cancer. Transl Cancer Res 2020;9:5218–32.
54. Xie J, Guo T, Zhong Z, Wang N, Liang Y, Zeng W, et al. ITGB1 drives hepatocellular carcinoma progression by modulating cell cycle process through PXN/YWHAZ/AKT pathways. Front Cell Dev Biol 2021;9:711149.
55. Sun Q, Zhou C, Ma R, Guo Q, Huang H, Hao J, et al. Prognostic value of increased integrin-beta 1 expression in solid cancers: A meta-analysis. Onco Targets Ther 2018;11:1787–99.
56. Parvani JG, Galliher-Beckley AJ, Schiemann BJ, Schiemann WP. Targeted inactivation of β1 integrin induces β3 integrin switching, which drives breast cancer metastasis by TGF-β. Mol Biol Cell 2013;24:3449–59.
57. Rao RC, Dou Y. Hijacked in cancer: The KMT2 (MLL) family of methyltransferases. Nat Rev Cancer 2015;15:334–46.
58. Ford DJ, Dingwall AK. The cancer COMPASS: Navigating the functions of MLL complexes in cancer. Cancer Genet 2015;208:178–91.
59. Xu J, Li L, Xiong J, denDekker A, Ye A, Karatas H, et al. MLL1 and MLL1 fusion proteins have distinct functions in regulating leukemic transcription program. Cell Discov 2016;2:16008.
60. Ghanbari M, Hosseinpour-Feizi M, Safaralizadeh R, Aghazadeh A, Montazeri V. Study of KMT2B (MLL2) gene expression changes in patients with breast cancer. Breast Cancer Manag 2019;8:BMT24.
61. Poreba E, Lesniewicz K, Durzynska J. Aberrant activity of Histone-Lysine N-Methyltransferase 2 (KMT2) complexes in oncogenesis. Int J Mol Sci 2020;21:9340.
62. Zhang C, Hua Y, Qiu H, Liu T, Long Q, Liao W, et al. KMT2A regulates cervical cancer cell growth through targeting VDAC1. Aging 2020;12:9604–20.
63. Zhang C, Song C, Liu T, Tang R, Chen M, Gao F, et al. KMT2A promotes melanoma cell growth by targeting hTERT signaling pathway. Cell Death Dis 2017;8:e2940.
64. Zhu J, Sammons MA, Donahue G, Dou Z, Vedadi M, Getlik M, et al. Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. Nature 2015;525:206–11.
65. Rahimi M, Talebi Kakroodi S, Tajvidi M. The importance of RTK signaling genes and their inhibitors in breast cancer. J Obstet Gynecol Cancer Res 2022;7:258–71.

66. Rahimi M, Behjati F, Hamid Reza KK, Karimlou M, Keyhani E. The relationship between KIT copy number variation, protein expression, and angiogenesis in sporadic breast cancer. Rep Biochem Mol Biol 2020;9:40–9.
67. Holmes AG. Potential efficacy of targeting MLL1 in breast cancer. Available at: https://ecommons.luc.edu/cgi/viewcontent.cgi?article=4677&context=luc_theses. Accessed Dec 21, 2023.
68. Brzezinka K, Nevedomskaya E, Lesche R, Haegebarth A, Ter Laak A, Fernández-Montalván AE, et al. Characterization of the Menin-MLL interaction as therapeutic cancer target. Cancers 2020;12:201.
69. Zhu J, Liu Z, Liang X, Wang L, Wu D, Mao W, et al. A pan-cancer study of KMT2 family as therapeutic targets in cancer. J Oncol 2022;2022:3982226.
70. Rabello Ddo A, de Moura CA, de Andrade RV, Motoyama AB, Silva FP. Altered expression of MLL methyltransferase family genes in breast cancer. Int J Oncol 2013;43:653–60.

Table S1. Details of datasets used in the present study

Dataset	Year	Platform	Tumor type	Sample	Country
GSE29431	2011	Affymetrix HG-U133_Plus_2	Breast cancer	12 normal breast tissues & 54 primary breast carcinomas	Spain
GSE12276	2009	Affymetrix HG-U133_Plus_2	Breast cancer	204 primary breast tumors	Netherlands

Table S2. Pathways involving HIST1H2BD. The pathways and the corresponding NES values extracted from the node table of the enrichment map network. Bolded genes indicate hub genes involved in the particular pathways.

	Pathways	NES	Genes
1	Activated PKN1 stimulates transcription of AR (androgen receptor) regulated genes KLK2 and KLK3	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
2	Activation of anterior HOX genes in hindbrain development during early embryogenesis	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
3	Amyloid fiber formation	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
4	B-WICH complex positively regulates rRNA expression	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
5	Cell Cycle Checkpoints	2.275	HIST1H2BD | HIST1H2BC | CENPU | HIST1H2BK | HIST1H2BH | H2BFS
6	Cell Cycle, Mitotic	2.275	HIST1H2BD | HIST1H2BC | CENPU | GINS2 | HIST1H2BK | HIST1H3D | HIST1H2BH | RRM2 | H2BFS
7	Cellular Senescence	2.694	HIST1H2BD | HIST1H2BC | HIST1H1C | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
8	Cellular protein-containing complex assembly	2.130	HIST1H2BD | HIST1H2BC | CENPU | HIST1H1C | HIST1H2BF | HIST1H2BK | HIST1H3H | HIST1H4H | HIST1H3D | HIST1H2BH | H2BFS | HIST1H2BE | DMD
9	Chromatin modifying enzymes	2.158	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | KMT2A
10	Chromatin organization	1.987	HIST1H2BD | HIST1H2BC | CENPU | HIST1H1C | HIST1H2BF | HIST1H2BK | HIST1H3H | HIST1H4H | HIST1H3D | HIST1H2BH | H2BFS | KMT2A | HIST1H2BE
11	Chromosome	2.171	HIST1H2BD | HIST1H2BC | CENPU | HIST1H1C | GINS2 | HIST1H2BF | HIST1H2BK | HIST1H3H | HIST1H4H | HIST1H3D | HIST1H2BH | H2BFS | HIST1H2BE | SPAG5
12	Condensation of Prophase Chromosomes	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
13	DNA Damage/Telomere Stress Induced Senescence	2.694	HIST1H2BD | HIST1H2BC | HIST1H1C | HIST1H2BK | HIST1H2BH | H2BFS
14	DNA conformation change	2.121	HIST1H2BD | HIST1H2BC | CENPU | HIST1H1C | GINS2 | HIST1H2BF | HIST1H2BK | HIST1H3H | HIST1H4H | HIST1H3D | HIST1H2BH | H2BFS | HIST1H2BE
15	DNA methylation	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
16	Deposition of new CENPA-containing nucleosomes at the centromere	2.275	HIST1H2BD | HIST1H2BC | CENPU | HIST1H2BK | HIST1H2BH | H2BFS
17	Developmental Biology	2.150	ITGB1 | HIST1H2BD | PTPRC | HIST1H2BC | WT1 | COL4A5 | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS | KMT2A
18	E3 ubiquitin ligases ubiquitinate target proteins	1.943	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H2BH
19	ERCC6 (CSB) and EHMT2 (G9a) positively regulate rRNA expression	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
20	ESR-mediated signaling	2.252	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS | MMP7
21	Estrogen-dependent gene expression	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
22	Formation of the beta-catenin:TCF transactivating complex	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS

Table S2. CONT.

	Pathways	NES	Genes
23	G2/M DNA damage checkpoint	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H2BH | H2BFS
24	HATs acetylate histones	2.158	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH
25	HCMV Early Events	1.905	ITGB1 | HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH
26	HCMV Late Events	2.158	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH
27	HDACs deacetylate histones	2.158	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH
28	Inhibition of DNA recombination at telomere	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H2BH | H2BFS
29	M Phase	2.275	HIST1H2BD | HIST1H2BC | CENPU | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
30	Meiotic recombination	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
31	Meiotic synapsis	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H2BH |H2BFS
32	Nonhomologous End-Joining (NHEJ)	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H2BH | H2BFS
33	Nucleosome assembly	2.458	HIST1H2BD | HIST1H2BC | CENPU |HIST1H1C | HIST1H2BF | HIST1H2BK | HIST1H3H | HIST1H4H | HIST1H3D | HIST1H2BH | H2BFS | HIST1H2BE
34	Oxidative Stress Induced Senescence	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
35	PRC2 methylates histones and DNA	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
36	Processing of DNA double-strand break ends	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H2BH | H2BFS
37	Protein heterodimerization activity	1.666	ITGB1 | HIST1H2BD | HIST1H2B | HIST1H2BF | HIST1H2BK | HIST1H3H | HIST1H4H | HIST1H3D | ATP1B1 | HIST1H2BH | H2BFS | HIST1H2BE
38	RHO GTPase Effectors	1.941	ITGB1 | HIST1H2BD | HIST1H2BC| CENPU | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
39	RNA Polymerase I Promoter Escape	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
40	RUNX1 regulates genes involved in megakaryocyte differentiation and platelet function	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS | KMT2A
41	RUNX1 regulates transcription of genes involved in differentiation of HSCs	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS | KMT2A
42	Recognition and association of DNA glycosylase with site containing an affected purine	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H2BH | H2BFS
43	SIRT1 negatively regulates rRNA expression	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
44	Senescence-Associated Secretory Phenotype (SASP)	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
45	Systemic lupus erythematosus	2.145	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H2BH | H2BFS
46	TCF dependent signaling in response to WNT	2.267	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS | WNT5A
47	Transcriptional regulation by small RNAs	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS
48	Transcriptional regulation of granulopoiesis	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS | KMT2A
49	Ub-specific processing proteases	1.955	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H2BH

Table S3. Pathways involving ITGB1. The pathways and the corresponding NES values extracted from the node table of the enrichment map network. Bolded genes indicate hub genes involved in the particular pathways.

	Pathways	NES	Genes
1	Developmental Biology	2.150	ITGB1 | HIST1H2BD | PTPRC | HIST1H2BC | WT1 | COL4A5 | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS | KMT2A
2	HCMV Early Events	1.905	ITGB1 | HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH
3	Protein heterodimerization activity	1.666	ITGB1 | HIST1H2BD | HIST1H2BC | HIST1H2BF | HIST1H2BK | HIST1H3H | HIST1H4H | HIST1H3D | ATP1B1 | HIST1H2BH | H2BFS | HIST1H2BE
4	RHO GTPase Effectors	1.941	ITGB1 | HIST1H2BD| HIST1H2BC | CENPU | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS

Table S4. Pathways involving KMT2A. The pathways and the corresponding NES values extracted from the node table of the enrichment map network. Bolded genes indicate hub genes involved in the particular pathways.

	Pathways	NES	Gene
1	Chromatin modifying enzymes	2.158	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | KMT2A
2	Chromatin organization	1.987	HIST1H2BD | HIST1H2BC | CENPU | HIST1H1C | HIST1H2BF | HIST1H2BK | HIST1H3H | HIST1H4H | HIST1H3D | HIST1H2BH | H2BFS | KMT2A | HIST1H2BE
3	Developmental Biology	2.150	ITGB1 | HIST1H2BD | PTPRC | HIST1H2BC | WT1 | COL4A5 | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS | KMT2A
4	RUNX1 regulates genes involved in megakaryocyte differentiation and platelet function	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS | KMT2A
5	RUNX1 regulates transcription of genes involved in differentiation of HSCs	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS | KMT2A
6	Transcriptional regulation of granulopoiesis	2.459	HIST1H2BD | HIST1H2BC | HIST1H2BK | HIST1H3D | HIST1H2BH | H2BFS | KMT2A

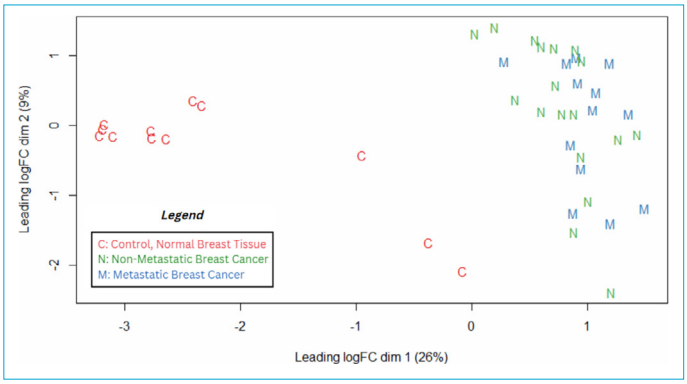

Figure S1. Unsupervised clustering of samples using Multidimensional Scaling (MDS).

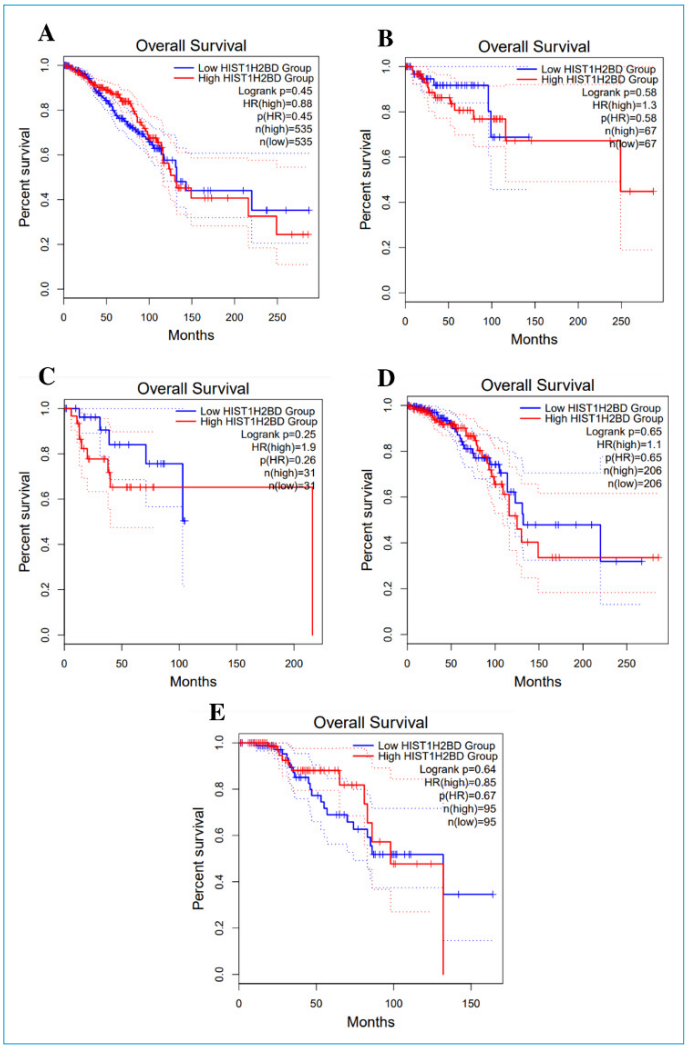

Figure S2. HIST1H2BD survival analysis using GEPIA2. The expression level was differentiated into two groups in which expression level higher than median expression level indicated in blue; expression level lower than median expression level indicated in red. **(a)** Survival analysis with general dataset. **(b)** Survival analysis with basal-like/triple negative dataset. **(c)** Survival analysis with HER2+ and non-luminal dataset. **(d)** Survival analysis with luminal A dataset **(e)** Survival analysis with luminal B dataset.

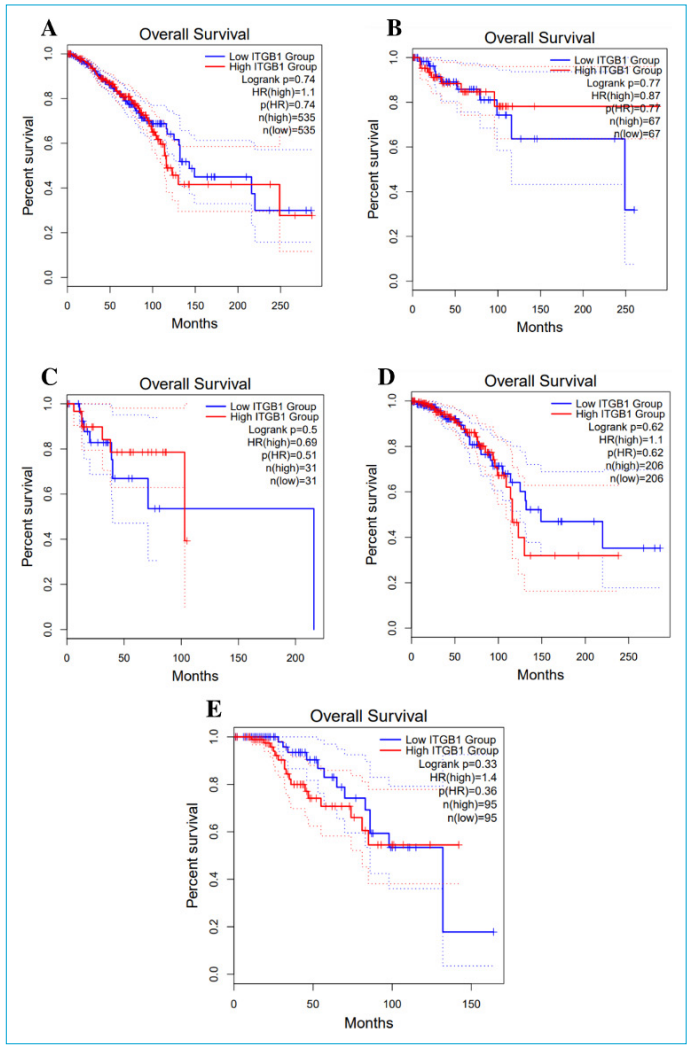

Figure S3. ITGB1 survival analysis using GEPIA2. The expression level was differentiated into two groups in which expression level higher than median expression level indicated in blue; expression level lower than median expression level indicated in red. **(a)** Survival analysis with general dataset. **(b)** Survival analysis with a basal-like/triple negative dataset. **(c)** Survival analysis with HER2+ and non-luminal dataset. **(d)** Survival analysis with luminal A dataset. **(e)** Survival analysis with luminal B dataset.

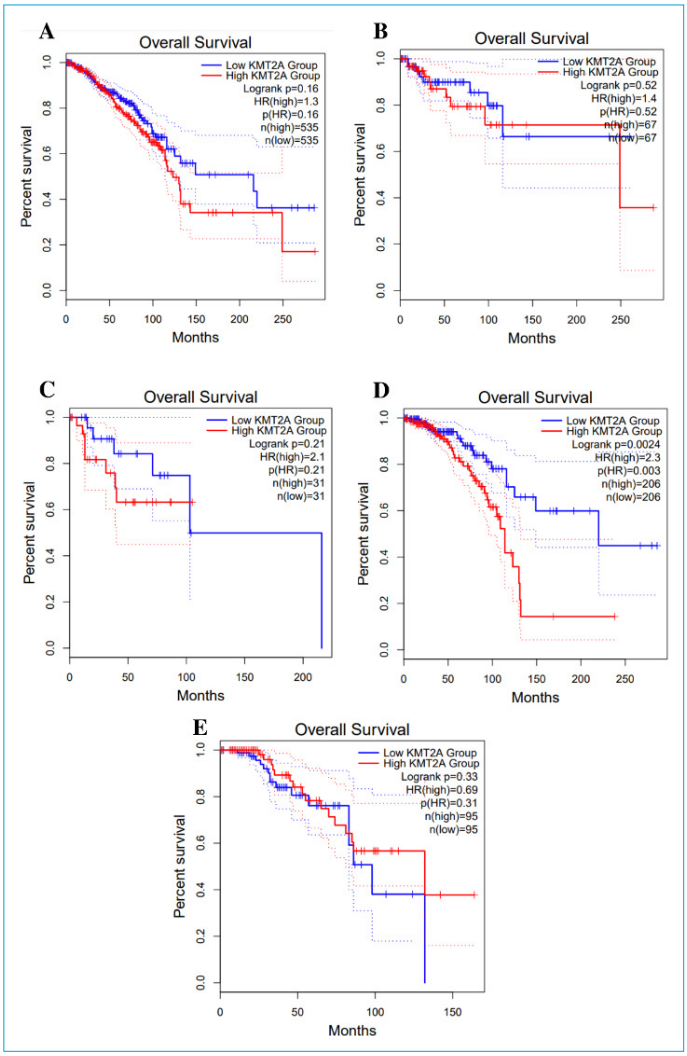

Figure S4. KMT2A survival analysis using GEPIA2. The expression level was differentiated into two groups in which expression level higher than median expression level indicated in blue; expression level lower than median expression level indicated in red. **(a)** Survival analysis with general dataset. **(b)** Survival analysis with basal-like/triple negative dataset. **(c)** Survival analysis with HER2+ and non-luminal dataset. **(d)** Survival analysis with luminal A dataset. **(e)** Survival analysis with luminal B dataset.