

Article

High WHSC1L1 Expression Reduces Survival Rates in Operated Breast Cancer Patients with Decreased CD8+ T Cells: Machine Learning Approach

Hyung-Suk Kim ¹, Kyueng-Whan Min ^{2,*}, Dong-Hoon Kim ^{3,*}, Byoung-Kwan Son ⁴, Mi-Jung Kwon ⁵
and Sang-Mo Hong ⁶

- ¹ Division of Breast Surgery, Department of Surgery, Hanyang University Guri Hospital, Hanyang University College of Medicine, Guri 15588, Korea; hyung6960@naver.com
 - ² Department of Pathology, Hanyang University Guri Hospital, Hanyang University College of Medicine, Guri 15588, Korea
 - ³ Department of Pathology, Kangbuk Samsung Hospital, Sungkyunkwan University School of Medicine, 29 Saemunanro, Seoul 03181, Korea
 - ⁴ Uijeongbu Eulji Medical Center, Department of Internal Medicine, Eulji University School of Medicine, Daejeon 34824, Korea; sbk1026@eulji.ac.kr
 - ⁵ Department of Pathology, Hallym University Sacred Heart Hospital, Hallym University College of Medicine, Anyang 24252, Korea; mulank@hanmail.net
 - ⁶ Division of Endocrinology, Department of Internal Medicine, Hanyang University Guri Hospital, Hanyang University College of Medicine, Guri 15588, Korea; lanugo@hanyang.ac.kr
- * Correspondence: kyueng@gmail.com (K.-W.M.); idavid.kim@samsung.com (D.-H.K.);
Tel.: +82-31-560-2346 (K.-W.M.); +82-2-2001-2392 (D.-H.K.); Fax: +82-2-31-560-2402 (K.-W.M.);
+82-2-2001-2398 (D.-H.K.)



Citation: Kim, H.-S.; Min, K.-W.; Kim, D.-H.; Son, B.-K.; Kwon, M.-J.; Hong, S.-M. High WHSC1L1 Expression Reduces Survival Rates in Operated Breast Cancer Patients with Decreased CD8+ T Cells: Machine Learning Approach. *J. Pers. Med.* **2021**, *11*, 636. <https://doi.org/10.3390/jpm11070636>

Academic Editors: Amabile Maria Ida and Salvatore Sorrenti

Received: 11 June 2021
Accepted: 1 July 2021
Published: 5 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Nuclear receptor-binding SET domain protein (NSD), a histone methyltransferase, is known to play an important role in cancer pathogenesis. The WHSC1L1 (Wolf-Hirschhorn syndrome candidate 1-like 1) gene, encoding NSD3, is highly expressed in breast cancer, but its role in the development of breast cancer is still unknown. The purpose of this study was to analyze the survival rates and immune responses of breast cancer patients with high WHSC1L1 expression and to validate the results using gradient boosting machine (GBM) in breast cancer. We investigated the clinicopathologic parameters, proportions of immune cells, pathway networks and in vitro drug responses according to WHSC1L1 expression in 456, 1500 and 776 breast cancer patients from the Hanyang University Guri Hospital, METABRIC and TCGA, respectively. High WHSC1L1 expression was associated with poor prognosis, decreased CD8+ T cells and high CD274 expression (encoding PD-L1). In the pathway networks, WHSC1L1 was indirectly linked to the regulation of the lymphocyte apoptotic process. The GBM model with WHSC1L1 showed improved prognostic performance compared with the model without WHSC1L1. We found that VX-11e, CZC24832, LY2109761, oxaliplatin and erlotinib were effective in inhibiting breast cancer cell lines with high WHSC1L1 expression. High WHSC1L1 expression could play potential roles in the progression of breast cancer and targeting WHSC1L1 could be a potential strategy for the treatment of breast cancer.

Keywords: WHSC1L1; breast neoplasm; prognosis; tumor infiltrating lymphocyte; PD-L1

1. Introduction

Epigenetic modifications of histones, including acetylation, methylation, phosphorylation and ubiquitination, are known to play an important role in many cellular events linked to development and disease pathogenesis, including malignancy [1,2]. Previous studies investigated the aberrant expression of histone methylation and histone methyltransferases that regulate the epigenetic modification of histones and are closely associated with various cancers [3]. Therefore, targeting histone methyltransferase (HMTase) overexpression is an important part of the epigenetic treatment of cancers. The nuclear receptor-binding SET

domain protein (NSD) family of HMTases is composed of NSD1, NSD2 and NSD3. These proteins are known to regulate chromatin integrity and gene expression primarily through the methylation of lysine 36 of histone H3 (H3K36), which is known as an indicator of transcriptional elongation [4,5]. Therefore, NSD family carcinogenic proteins could affect cell proliferation and cancer progression.

NSD3 is located on the chromosome in the 8p11-p12 locus known as Wolf-Hirschhorn syndrome candidate 1-like 1 (WHSC1L1), which is amplified in breast cancer cell lines [6]. In previous studies, high WHSC1L1 expression has been shown to be common in lung cancer and pancreatic cancer [7,8]. Other studies have shown that WHSC1L1 plays roles as an oncogene and a transforming gene, driving the development and progression of breast cancer [9,10]. In studies of breast cancer, high WHSC1L1 expression was associated with a poor prognosis [11,12]. Nevertheless, the biological functions of WHSC1L1 in the carcinogenesis of breast cancer are not well understood.

Cancer cell survival could depend on the interactions between cancer cells and immune cells that make up the tumor microenvironment (TME). The TME plays an important role in cancer progression and the response to treatment, thus affecting the patient's outcome. Marked lymphocytic infiltrates, known as tumor-infiltrating lymphocytes (TILs), at the invasive front of the tumor could play a pivotal role in anticancer immunity and be beneficial prognostic factors in various cancers. Cytotoxic T lymphocytes (CTLs) that express CD8 on the cell surface play a major role in anticancer immunity [13]. The identification of different types of immune cells in the TME helps predict cancer prognosis [14]. Previous studies have reported an association between NSD1 and immune cells, but the association between WHSC1L1 and immune cells has never been reported to the best of our knowledge [15].

In this study, we assessed the clinicopathological parameters and survival rate according to WHSC1L1 expression in breast cancer cohorts from Hanyang University Guri Hospital (HYGH), Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) and The Cancer Genome Atlas (TCGA) database [16]. We analyzed the effect of WHSC1L1 on the survival of breast cancer patients using the gradient boosting machine (GBM) algorithm [17]. In addition, we investigated gene sets related to WHSC1L1 using a pathway-based network [18,19]. Anticancer immune responses were analyzed by assessing the distributions of CD8+ T cells and CD4+ T cells. Using the Genomics of Drug Sensitivity in Cancer (GDSC) database as an *in vitro* drug screening platform, we found sensitive drugs in breast cancer cell lines with high WHSC1L1 expression [20,21].

2. Materials and Methods

2.1. Patient Selection

This study enrolled 456 invasive ductal carcinoma (IDC) patients with tissue samples obtained at HYGH in Korea between 2005 and 2015. The Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK) criteria were followed throughout this study [22]. The inclusion criteria were as follows: (1) stage I–III breast cancer patients who underwent curative surgery; (2) patients who did not receive neoadjuvant chemotherapy; and (3) patients with available paraffin blocks of tumor tissues or complete clinical outcome data and follow-up data. We assessed the survival rate and clinicopathologic characteristics of the tumors, including age, T stage, N stage, histopathological grade, lymphatic invasion, vascular invasion, perineural invasion, hormonal receptors, human epidermal growth factor receptor 2 (HER2) status, Ki-67, P53, and anti-programmed death-ligand 1 (PD-L1) and patient follow-up information.

2.2. Tissue Microarray Construction and Immunohistochemistry in Our Cohort

In HYGH samples, tissue microarray (TMA) blocks were assembled using a tissue array instrument (AccuMax Array; ISU ABXIS Co., Ltd., Seoul, Korea). We used duplicate 3 mm diameter tissue cores (tumor components in a tissue core > 70%) from each donor block. Four-micrometer sections were cut from the TMA blocks using routine techniques.

Immunostaining for WHSC1L1 (1:100, 11345-1-AP; Proteintech, Manchester, UK), estrogen receptor (ER) (1:200, Lab Vision Corporation, Fremont, CA, USA), progesterone receptor (PR) (1:200, Dako, Glostrup, Denmark), HER2 (1:1, Ventana Medical Systems Tucson, Oro Valley, AZ, USA), anti-CD8 (clone 4B11, Leica Biosystems, Newcastle, UK), anti-CD4 (clone 4B12, Leica Biosystems, Newcastle, UK), P53 (1:5000, Cell Marque, Hot Springs, AR, USA) and Ki67 (1:200; MIB-1, Dako, Glostrup, Denmark) was performed using the Bond Polymer Refine Detection System (Leica Biosystems Newcastle Ltd., Newcastle, UK) according to the manufacturer’s instructions and the Dako Autostainer Universal Staining System (Dako, Carpinteria, CA, USA) with the ChemMate DAKO EnVision™ Detection Kit (Dako) [23]. PD-L1 (clone SP142, Ventana Medical Systems, Roche, Tucson, AZ, USA) staining was performed. The intensity of immunostaining was recorded as follows: 0 (no staining), 1 (weak staining), 2 (moderate staining), and 3 (strong staining) (Figure 1A). The proportion of staining was graded as follows: 0 (0–5%), 1 (6–25%), 2 (26–50%), 3 (51–75%), and 4 (>75%). The immunoreactive score (IRS) was calculated (intensity × proportion), and WHSC1L1 expression was determined to be either low (IRS ≤ 3) or high (IRS > 3).

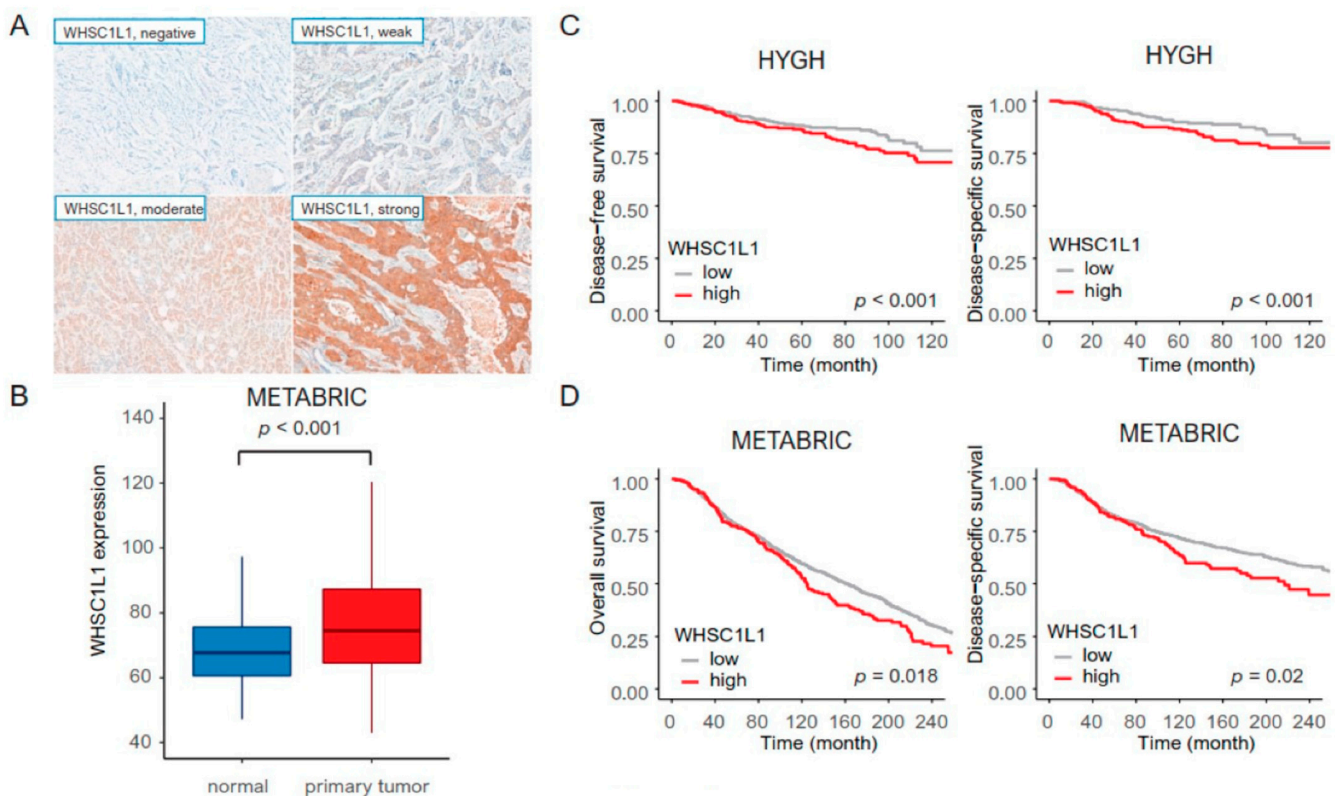


Figure 1. (A) The staining intensity was scored as negative (top left), weak (top right), moderate (bottom left), or strong (bottom right) from HYGH cohort (original magnification ×200). (B) METABRIC: High WHSC1L1 expression in primary tumors compared to that in normal tissues. (C) HYGH cohort: High WHSC1L1 expression was associated with poor disease-free survival (average survival time: high WHSC1L1 expression group, 67.2 months; low WHSC1L1 expression group, 87.6 months) (median survival time: high WHSC1L1 expression group, 63.5 months; low WHSC1L1 expression group, 95 months) and disease-specific survival (average survival time: high WHSC1L1 expression group, 74.6 months; low WHSC1L1 expression group, 95.3 months) (median survival time: high WHSC1L1 expression group, 71.5 months; low WHSC1L1 expression group, 98 months) ($p < 0.001$ and 0.001 , respectively). (D) METABRIC: High WHSC1L1 expression was associated with poor overall survival (average survival time: high WHSC1L1 expression group, 123.8 months; low WHSC1L1 expression group, 125.3 months) (median survival time: high WHSC1L1 expression group, 116.4 months; low WHSC1L1 expression group, 115.5 months) and disease-specific survival (average survival time: high WHSC1L1 expression group, 126.3 months; low WHSC1L1 expression group, 136.1 months) (median survival time: high WHSC1L1 expression group, 117.2 months; low WHSC1L1 expression group, 119.5 months) ($p = 0.018$ and 0.02 , respectively).

2.3. Analysis Based on the METABRIC Database and TCGA Database

We obtained 1500 IDC cases from the METABRIC database with gene data (cDNA microarray profiling, Illumina HT-12 v3 platform) (<https://www.cbioportal.org/>) (accessed on 1 June 2021) [24]. The microarray from METABRIC (log-transformed data) was assessed, and WHSC1L1 expression was determined to be either low (log-scale scores < 7.14402497) or high (log-scale scores > 7.14402497). WHSC1L1 expression, overall survival (OS) and disease-specific survival (DSS) were extracted using the R package (<http://www.r-project.org/>) (accessed on 1 June 2021). Normal samples and tumor samples with missing data were excluded from the analysis.

We obtained 776 IDC cases with RNA-Seq data from the TCGA database [16]. The RNA-Seq data from TCGA was assessed, and WHSC1L1 expression was determined to be either low (scores < 1472.755) or high (scores > 1472.755). We applied in silico cytometry known as CIBERSORT (<https://cibersort.stanford.edu/>) (accessed on 1 June 2021) to determine the proportions of 22 subsets of immune cells using 547 genes [25]. For grouping of networks based on functionally enriched Gene Ontology (GO) terms and pathways, pathway network analyses were visualized using Cytoscape software (version 3.8.2). We observed which genes had the closest relationship with high WHSC1L1 expression using the kappa value and elucidated the functionally grouped Gene Ontology and pathway annotation networks using the ClueGO application (version 2.5.6), an app for Gene Ontology analysis [18,19].

2.4. Machine Learning Algorithm for Validation

We integrated WHSC1L1 with clinical risk factors (T stage, N stage, histological grade, lymphatic invasion, perineural invasion and ER) to construct prognostic models for survival prediction by applying machine learning (ML) algorithms in 456 cases (HYGH) (randomization: training set, 70%; validation set, 30%). A learning algorithm was independently applied to select and combine multiple covariates from GBM based on multivariate Gaussian models. In this step, the “forward” search method, which initiates on a prototype set and selects a feature if and only if the addition of the feature could increase the performance of the prognostic model, was adopted to select optimal features sequentially. The hyperparameters of the ML algorithms, such as the learning rate in GBM, were optimized for each combination of selected covariates and learning algorithm by grid search cross-validation through a predefined range. We searched 81 models with varying learning rates and tree depths. The final optimal models were trained based on the selected covariates and the optimized hyperparameters [17]. To explore the performance outcomes of the GBM method, the receiver operator characteristic (ROC) curve was used.

2.5. GDSC Database

We analyzed the relationship between anticancer drug sensitivity and WHSC1L1 expression based on the Genomics of Drug Sensitivity in Cancer (GDSC version 2.0) dataset, which contains data on the drug responses of approximately 19 breast cancer cell lines to 172 anticancer drugs [26]. We measured anticancer drug sensitivity in 19 breast cancer cell lines with the natural log-half-maximal inhibitory concentration (LN IC₅₀). A drug was identified as an effective drug when the calculated LN IC₅₀ value was decreased in cell lines with high WHSC1L1 expression and increased in those with low NSD3 expression, i.e., when an inverse correlation was observed. Pearson’s correlation and Student’s *t*-test were used to assess the comparisons between the LN IC₅₀ values and WHSC1L1 expression [20,21].

2.6. Statistical Analysis

Correlations between several clinicopathological variables and WHSC1L1 expression were analyzed using the χ^2 test. Student’s *t*-test and Pearson’s correlation were used to examine the differences between continuous variables. Disease-free survival (DFS) was defined as the time from the date of diagnosis to recurrence/new distant metastasis, and

DSS was defined as the time from the date of diagnosis to cancer-related death. OS was defined as the time from the date of diagnosis to all-cause death. The Kaplan–Meier method was used to determine the probability of survival, and survival rates were compared using the log-rank test and Cox regression analyses. A two-tailed p -value of <0.05 was considered statistically significant. All data were analyzed using R software packages and SPSS statistics (version 25.0, SPSS Inc., Chicago, IL, USA).

3. Results

3.1. Clinicopathological Parameters and Survival Rate

We investigated a total of 1956 patients with WHSC1L1 expression data and survival data in the HYGH cohort and METABRIC. In the HYGH cohort, high WHSC1L1 expression was related to ER negativity and PR negativity ($p = 0.04$ and 0.017 , respectively). High WHSC1L1 expression was frequently observed in tumors with PD-L1 negativity, high p53 expression and high Ki-67 ($p = 0.001$, 0.005 and <0.001 , respectively) (Table 1). In METABRIC, WHSC1L1 expression was increased in primary tumors compared to normal tissues ($p = 0.046$) (Figure 1B).

Table 1. Clinicopathological parameters of WHSC1L1 (Wolf-Hirschhorn syndrome candidate 1-like protein 1) in HYGH cohort.

Parameter	WHSC1L1 (HYGH Cohort)		p -Value
	Low (n = 224), n (%)	High (n = 232), n (%)	
Age	49.5 ± 9.3	49.5 ± 10.5	0.99 ¹
T stage			
1	109 (48.7)	99 (42.7)	0.06 ³
2	109 (48.7)	117 (50.4)	
3	6 (2.7)	16 (6.9)	
N stage			
0	115 (51.3)	118 (50.9)	0.763 ³
1	64 (28.6)	71 (30.6)	
2	31 (13.8)	29 (12.5)	
3	14 (6.2)	14 (6.0)	
Histological grade			
1	39 (17.4)	48 (20.7)	0.335 ³
2	115 (51.3)	101 (43.5)	
3	70 (31.2)	83 (35.8)	
Lymphatic invasion			
Negative	123 (54.9)	107 (46.1)	0.075 ¹
Positive	101 (45.1)	125 (53.9)	
Vascular invasion			
Negative	213 (95.1)	213 (91.8)	0.221 ¹
Positive	11 (4.9)	19 (8.2)	
Perineural invasion			
Negative	184 (82.1)	178 (76.7)	0.189 ¹
Positive	40 (17.9)	54 (23.3)	
ER			
Negative	53 (23.7)	76 (32.8)	0.04 ¹
Positive	171 (76.3)	156 (67.2)	
PR			
Negative	75 (33.5)	104 (44.8)	0.017 ¹
Positive	149 (66.5)	128 (55.2)	

Table 1. Cont.

Parameter	WHSC1L1 (HYGH Cohort)		p-Value
	Low (n = 224), n (%)	High (n = 232), n (%)	
HER2			
Negative	161 (71.9)	157 (67.7)	0.382 ¹
Positive	63 (28.1)	75 (32.3)	
PD-L1			
Negative	155 (62.9)	127 (54.7)	0.001 ¹
Positive	69 (30.8)	105 (45.3)	
P53 percentage	8.4 ± 11.9	11.8 ± 13.7	0.005 ²
Ki-67 index	21.5 ± 32.8	34.0 ± 38.5	<0.001 ²

T or N stage, 8th edition; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; PD-L1, programmed death-ligand 1; ¹ Chi-square test; ² Student's *t*-test; ³ T stage: 1, 2 vs. 3; N stage: 0, 1 vs. 2, 3; Histological grade: 1, 2 vs. 3; *p* < 0.05 is shown in bold.

In the HYGH cohort, patients with high WHSC1L1 expression had significantly worse DFS and DSS than those with low WHSC1L1 expression (high WHSC1L1 expression group, 354 patients; low WHSC1L1 expression group, 102 patients) (all *p* < 0.001) (Figure 1C). After adjustment for confounders, such as T stage, N stage, histological grade, lymphatic invasion, perineural invasion and ER, the significance remained (Table 2). Regarding molecular subtypes, such as luminal A, luminal B, HER2, and triple-negative, high WHSC1L1 expression was associated with short DSS and DFS in the luminal A (*p* < 0.001 and 0.001, respectively) and HER2 subtypes (*p* = 0.031 and 0.01, respectively).

Table 2. Disease-free and disease-specific survival analyses according to WHSC1L1 (Wolf-Hirschhorn syndrome candidate 1-like protein 1) in HYGH cohort.

Disease-Free Survival	Univariate ¹	Multivariate ²	HR	95% CI	
WHSC1L1 (low vs. high)	<0.001	<0.001	2.265	1.451	3.537
T stage (1, 2 vs. 3)	0.001	0.001	2.820	1.507	5.275
N stage (0, 1 vs. 2, 3)	<0.001	0.001	2.124	1.334	3.381
Histological grade (1, 2 vs. 3)	0.01	0.05	1.529	0.988	2.368
Lymphatic invasion (absence vs. presence)	<0.001	0.234	1.353	0.822	2.225
Perineural invasion (absence vs. presence)	<0.001	<0.001	2.192	1.415	3.394
Estrogen receptor (negative vs. positive)	0.045	0.05	0.647	0.414	1.011
Disease-specific survival	Univariate ¹	Multivariate ²	HR	95% CI	
WHSC1L1 (low vs. high)	<0.001	<0.001	2.505	1.567	4.005
T stage (1, 2 vs. 3)	<0.001	0.022	2.251	1.122	4.516
N stage (0, 1 vs. 2, 3)	<0.001	<0.001	2.494	1.541	4.037
Histological grade (1, 2 vs. 3)	0.001	0.3	1.274	0.806	2.015
Lymphatic invasion (absence vs. presence)	<0.001	0.31	1.311	0.777	2.210
Perineural invasion (absence vs. presence)	<0.001	<0.001	2.477	1.586	3.868
Estrogen receptor (negative vs. positive)	0.008	0.076	0.655	0.410	1.046

p < 0.05 is shown in bold. ¹ Log-rank test; ² Cox proportional hazard model.

In METABRIC, we investigated 1500 IDC patients to validate the relationship between WHSC1L1 and survival. High WHSC1L1 expression was significantly correlated with poor DSS and OS (high WHSC1L1 expression group, 168 patients; low WHSC1L1 expression group, 1332 patients) (*p* = 0.02 and *p* = 0.018, respectively) compared to low WHSC1L1 expression (Figure 1D).

3.2. Anticancer Immune Response and Pathway Network Analysis

We analyzed the relationship between WHSC1L1 expression and immune cells using the HYGH and TCGA cohorts. In the HYGH cohort, high WHSC1L1 expression was associated with decreased CD8+ T cell count, increased CD4+ T cell count and high PD-L1 expression ($p = 0.017$, 0.024 and 0.012 , respectively) (Figure 2A,B). In the TCGA cohort, high WHSC1L1 expression was also correlated with a low fraction of CD8+ T cells, a high fraction of CD4+ T cells and high CD274 (encoding PD-L1) expression ($p = 0.046$, < 0.001 and $= 0.002$, respectively) (Figure 2C). In pathway network analysis, WHSC1L1 was linked to the regulation of translation initiation, ERBB2 signaling pathway, positive regulation of the DNA metabolic process, regulation of chromosome organization and regulation of the lymphocyte apoptotic process (Figure 3).

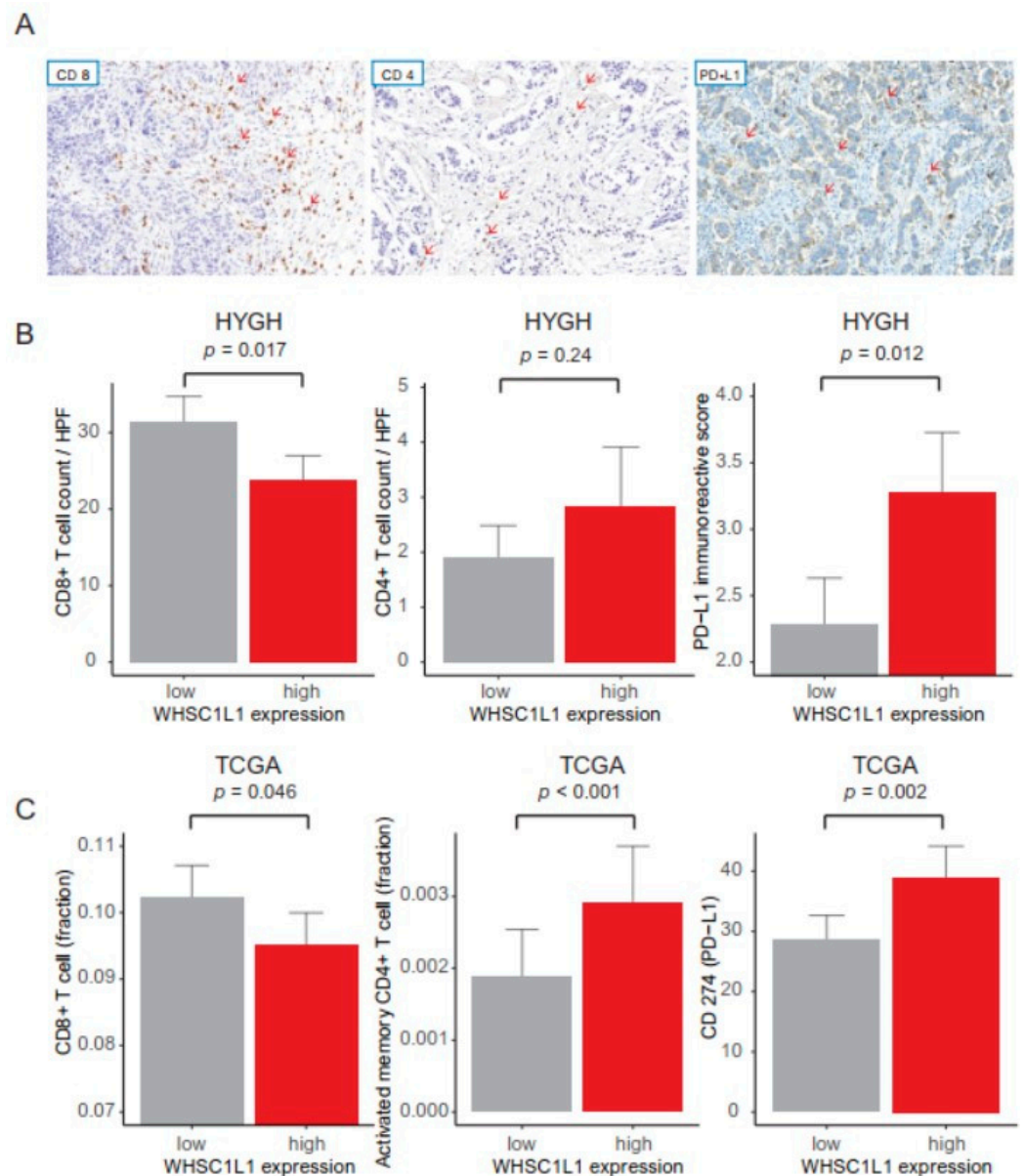


Figure 2. (A) Representative microphotographs showing CD8+ T cells, CD4+ T cells and PD-L1 expression from HYGH cohort (red arrows, positive stain (brown color)). (B) Bar plot of CD8+ T cell count, CD4+ T cell count per high-power field (HPF, $\times 400$) and PD-L1 expression in HYGH cells. (C) Bar plot of CD8+ T cell fraction, activated memory CD4+ T cell fraction and CD274 (encoding PD-L1) expression in TCGA.

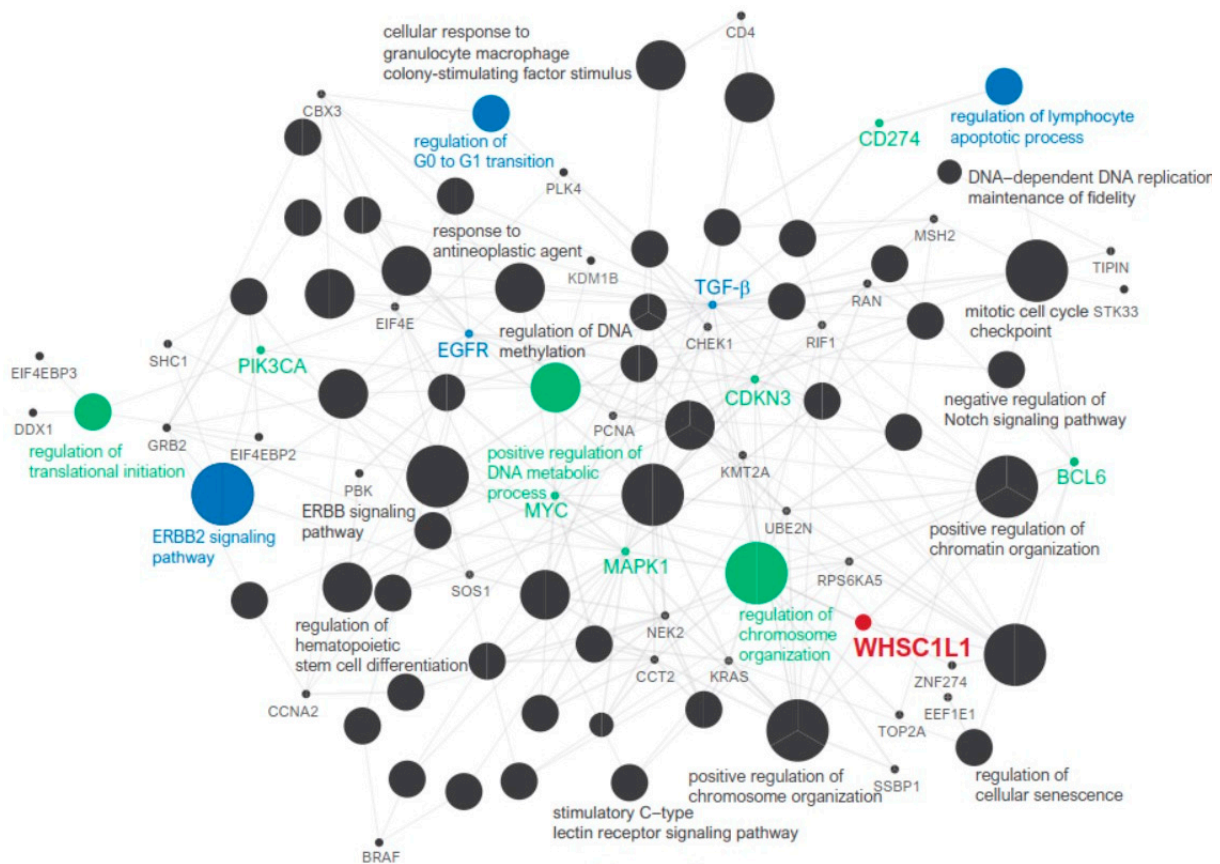


Figure 3. Functionally grouped networks are linked to their biological function, and only the most significant term of the group is labeled.

3.3. Machine Learning and Drug Screening

We compared the performance of the two GBM models in predicting survival rates (Model 1 (T stage, N stage, histological grade, lymphatic invasion, perineural invasion, and ER) vs. Model 2 (WHSC1L1, T stage, N stage, histological grade, lymphatic invasion, perineural invasion, and ER)). ROC curves were generated (area under the curve: Model 1, 0.771; Model 2, 0.823). We found that the GBM algorithm performed the best, while the addition of WHSC1L1 to the prediction model improved the prognostic performance (Figure 4A).

In the GDSC database, we analyzed the drug sensitivity of 50 breast cancer cell lines according to WHSC1L1 expression. We found five anticancer drugs that most effectively reduced the growth of breast cancer cells with high WHSC1L1 expression: VX-11e, CZC24832, LY2109761, oxaliplatin and erlotinib (Figure 4B).

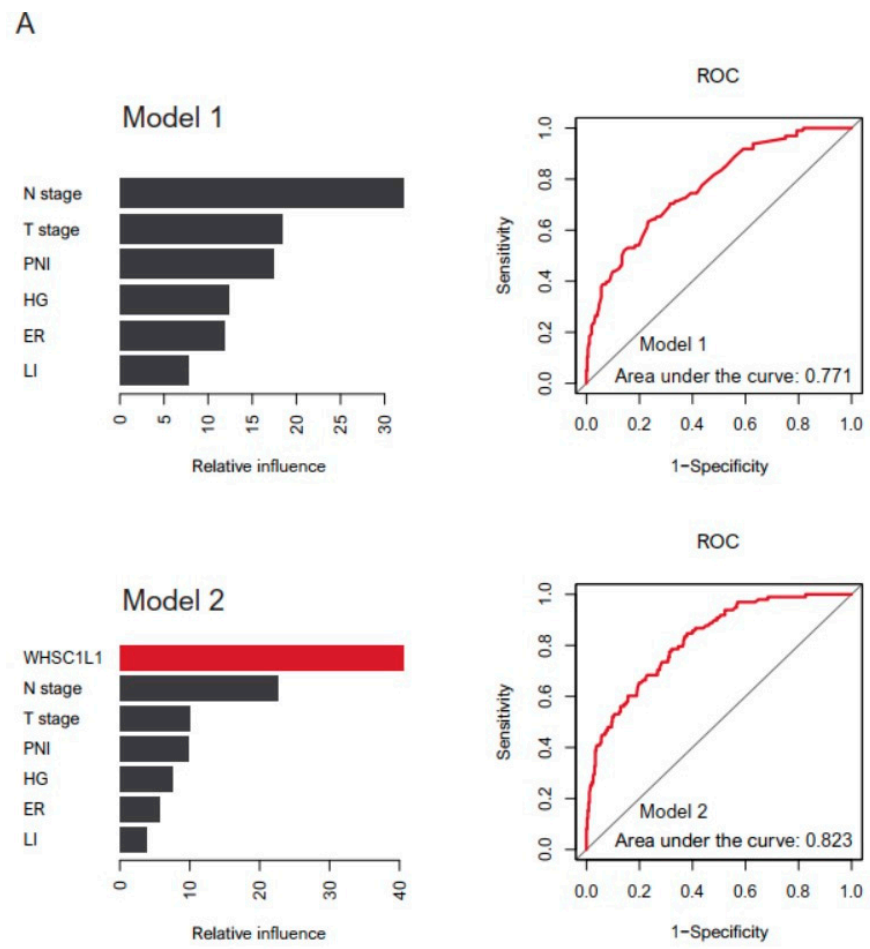


Figure 4. Cont.

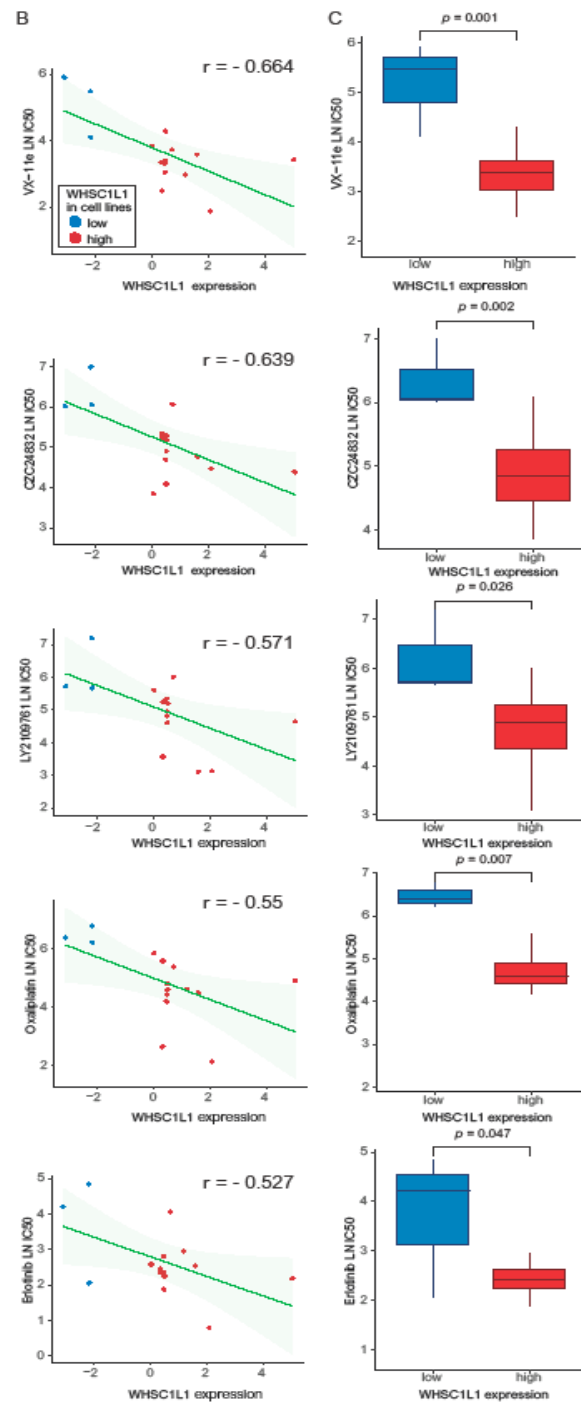


Figure 4. (A) We used the supervised machine learning model (gradient boosting machine) for prognosis prediction. Covariates were included as confounding factors (Model 1 (T stage, N stage, perineural invasion (PNI), histological grade (HG), estrogen receptor (ER) and lymphatic invasion (LI)) vs. Model 2 (WHSC1L1, T stage, N stage, PNI, HG, ER and LI)), and their relative importance was assessed using survival analysis. A receiver operator characteristic curve for GBM was used based on a multivariate Gaussian model. (B) Pearson’s correlation analysis showing the natural log of the half-maximal inhibitory concentration (LN IC50) values of VX-11e, CZC24832, LY2109761, oxaliplatin and erlotinib in breast cancer cells (blue, low WHSC1L1 expression; red, high WHSC1L1 expression). (C) Bar plot showing the LN IC50 values of VX-11e, CZC24832, LY2109761, oxaliplatin and erlotinib in breast cancer cells with low (blue) and high (red) WHSC1L1 expression ($p = 0.001$, 0.002 , 0.026 , 0.007 and 0.047 , respectively) (error bars: standard errors of the mean).

4. Discussion

This study found that WHSC1L1 was significantly overexpressed in breast cancer tissue compared to normal breast tissue. Higher WHSC1L1 expression was associated with worse DFS and DSS in breast cancer. To validate our results, we compared the relationship between the survival rate and WHSC1L1 in METABRIC. In this study, the inclusion of WHSC1L1 in the machine learning model increased the accuracy of predicting the survival rate. GBM, a type of machine learning, has the advantage of processing large amounts of predictors through simple prediction algorithms and combining the results in a non-linear and interactive way, which can improve the accuracy of predictions [17]. Therefore, our findings suggest that WHSC1L1 expression plays an important role in the development and progression of breast cancer as well as epigenetic regulation; thus, it is expected to contribute to effective treatments for breast cancer [27–30].

WHSC1L1, a histone methyltransferase, is an important driving oncogene of the amplification of 8p11-12 in breast cancer and is an epigenetic marker that regulates cell growth and differentiation [31–34]. WHSC1L1 expression is significantly elevated in various malignant tumors, such as breast cancer, bladder cancer, osteosarcoma, head and neck cancer, and colorectal cancer [7,35–38]. A previous study found that high WHSC1L1 expression was related to a high Ki-67 index and poor prognosis in breast cancer [5,31]. These results are consistent with our results.

The TME is composed of immune cells, fibroblasts, satellite cells, and blood vessels or lymphatic vessels. These factors play a pivotal role in tumor progression, treatment response, and clinical outcomes [13]. In the TME, cytotoxic T lymphocytes can induce the apoptosis of target cells through the cancer-immunity cycle [39]. In our study, high WHSC1L1 expression was related to decreased CD8+ T cell counts and high PD-L1 expression. We applied *in silico* flow cytometry to TCGA data and found that high WHSC1L1 expression was significantly associated with a decreased CD8+ T cell fraction and high CD274 (encoding PD-L1) expression. These results suggest that upregulated PD-L1 inhibits CD8+ T cells, thereby causing the immune escape of cancer cells. Considerably high WHSC1L1 expression indicates worse clinical outcomes by inhibiting antitumoral immune activity. In pathway network analysis, WHSC1L1 was indirectly linked to the regulation of the lymphocyte apoptosis process.

There are few reports on specific inhibitors suitable for H3K36 methylation of the NSD family, so pharmacological inhibition of WHSC1L1 is currently not available. We suggest candidate drugs related to WHSC1L1 expression using the GDSC database, not an experimental method. We investigated 175 anticancer drugs in 50 breast cancer cell lines from the GDSC database [21]. We identified the following five anticancer drugs that can most effectively reduce the growth of breast cancer cells with high WHSC1L1 expression: VX-11e, CZC24832, LY2109761, oxaliplatin and erlotinib. VX-11e is a potent and selective ERK2 inhibitor that reduces tumor growth, proliferation and viability in a variety of cancer cell lines. VX-11e affects G0/G1 cell cycle arrest and induces high expression of p21 cell cycle inhibitors [40]. Our study showed that WHSC1L1 was indirectly linked to the regulation of the G0 to G1 transition. In liver cancer, LY2109761, a TGF- β receptor inhibitor, was shown to reduce tumor cell growth and intravascular and metastatic dissemination [41]. WHSC1L1 was also indirectly related to TGF- β , indicating that LY2109761 may be effective in breast cancer cell lines with high WHSC1L1 expression. WHSC1L1 was indirectly related to the regulation of the lymphocyte apoptotic process. PI3K γ is related to lymphocyte activation, differentiation, and chemotaxis [42]. CZC24832, a PI3K γ inhibitor, may be effective in breast cancer cell lines with high WHSC1L1 expression. Oxaliplatin and trastuzumab have a synergistic antitumor effect in gastric cancer cells with Erb-B2 receptor tyrosine kinase 2 (ERBB2) [43]. Our results showed that WHSC1L1 was related to the ERBB2 signaling pathway and identified specific hub genes, such as epidermal growth factor receptor (EGFR). Erlotinib could inhibit the tyrosine kinase activity of EGFR and may be a candidate drug in breast cancers with high WHSC1L1 expression [44].

This study has several limitations. First, this retrospective study had potential selection bias. Second, our study analyzed the oncogenic role of high WHSC1L1 expression using a bioinformatic approach: *in silico* analyses. *In vivo* experimental studies are needed to identify the molecular mechanisms. Third, in patients with breast cancer, the drugs suggested in this study may be different depending on disease status, microenvironments, and immunities. Further studies are needed to evaluate the therapeutic utility of WHSC1L1 inhibition in patients with breast cancer.

5. Conclusions

Our study revealed that WHSC1L1 was highly expressed in breast cancer tissues compared to normal tissues. High WHSC1L1 expression was associated with decreased CD8+ T cells and increased PD-L1. Pathway-based network analysis revealed a significant relationship between WHSC1L1 and the regulation of the lymphocyte apoptotic process pathway. Thus, the result could be one of several factors that can explain the relationships between high WHSC1L1 expression and low survival in patients with breast cancer. Additionally, this study confirmed the importance of WHSC1L1 in predicting survival rates using machine learning. We identified five drugs that inhibited breast cancer cells with high WHSC1L1 expression. We believe that medical oncologists and researchers will be interested in the role of WHSC1L1 in breast cancer and that our results will facilitate further studies.

Author Contributions: Conceptualization, B.-K.S., M.-J.K. and S.M.-H.; Data curation, H.-S.K., K.-W.M. and D.-H.K.; Formal analysis, H.-S.K., K.-W.M. and D.-H.K.; Investigation, H.-S.K., K.-W.M., D.-H.K., B.-K.S., M.-J.K. and S.M.-H.; Methodology, H.-S.K.; Supervision, K.-W.M., D.-H.K., B.-K.S., M.-J.K. and S.M.-H.; Writing—original draft, H.-S.K.; Writing—review & editing, H.-S.K., K.-W.M. and D.-H.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the research fund of Hanyang University (HY-20210000001054).

Institutional Review Board Statement: This study protocol was approved by the Institutional Review Board of Hanyang University Guri Hospital (IRB number: 2020-10-003) and was performed according to the ethical standards of the Declaration of Helsinki, as revised in 2008. The review conducted by our institutional review board confirmed that informed consent was not necessary for this study.

Informed Consent Statement: Patient consent was waived from the IRB due to the retrospective design.

Data Availability Statement: The data presented in this study can be available on request from the corresponding author. The data are not publicly available due to privacy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hyun, K.; Jeon, J.; Park, K.; Kim, J. Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.* **2017**, *49*. [[CrossRef](#)]
2. Zhao, Z.B.; Shilatifard, A. Epigenetic modifications of histones in cancer. *Genome Biol.* **2019**, *20*. [[CrossRef](#)]
3. Chen, Y.; Ren, B.; Yang, J.; Wang, H.; Yang, G.; Xu, R.; You, L.; Zhao, Y. The role of histone methylation in the development of digestive cancers: A potential direction for cancer management. *Signal. Transduct. Target Ther.* **2020**, *5*, 143. [[CrossRef](#)] [[PubMed](#)]
4. Wagner, E.J.; Carpenter, P.B. Understanding the language of Lys36 methylation at histone H3. *Nat. Rev. Mol. Cell Bio.* **2012**, *13*, 115–126. [[CrossRef](#)]
5. Liu, L.X.; Kimball, S.; Liu, H.; Holowatyj, A.; Yang, Z.Q. Genetic alterations of histone lysine methyltransferases and their significance in breast cancer. *Oncotarget* **2015**, *6*, 2466–2482. [[CrossRef](#)]
6. Angrand, P.O.; Apiou, F.; Stewart, A.F.; Dutrillaux, B.; Losson, R.; Chambon, P. NSD3, a new SET domain-containing gene, maps to 8p12 and is amplified in human breast cancer cell lines. *Genomics* **2001**, *74*, 79–88. [[CrossRef](#)]
7. Kang, D.; Cho, H.S.; Toyokawa, G.; Kogure, M.; Yamane, Y.; Iwai, Y.; Hayami, S.; Tsunoda, T.; Field, H.I.; Matsuda, K.; et al. The histone methyltransferase Wolf-Hirschhorn syndrome candidate 1-like 1 (WHSC1L1) is involved in human carcinogenesis. *Gene Chromosome Canc.* **2013**, *52*, 126–139. [[CrossRef](#)]

8. Mahmood, S.F.; Gruel, N.; Nicolle, R.; Chapeaublanc, E.; Delattre, O.; Radvanyi, F.; Bernard-Pierrot, I. PPAPDC1B and WHSC1L1 Are Common Drivers of the 8p11-12 Amplicon, Not Only in Breast Tumors But Also in Pancreatic Adenocarcinomas and Lung Tumors. *Am. J. Pathol.* **2013**, *183*, 1634–1644. [[CrossRef](#)]
9. Yang, Z.Q.; Liu, G.; Bollig-Fischer, A.; Giroux, C.N.; Ethier, S.P. Transforming properties of 8p11-12 amplified genes in human breast cancer. *Cancer Res.* **2010**, *70*, 8487–8497. [[CrossRef](#)] [[PubMed](#)]
10. Zhou, Z.; Thomsen, R.; Kahns, S.; Nielsen, A.L. The NSD3L histone methyltransferase regulates cell cycle and cell invasion in breast cancer cells. *Biochem. Biophys. Res. Commun.* **2010**, *398*, 565–570. [[CrossRef](#)] [[PubMed](#)]
11. Irish, J.C.; Mills, J.N.; Turner-Ivey, B.; Wilson, R.C.; Guest, S.T.; Rutkovsky, A.; Dombkowski, A.; Kappler, C.S.; Hardiman, G.; Ethier, S.P. Amplification of WHSC1L1 regulates expression and estrogen-independent activation of ER alpha in SUM-44 breast cancer cells and is associated with ER alpha over-expression in breast cancer. *Mol. Oncol.* **2016**, *10*, 850–865. [[CrossRef](#)]
12. Jeong, G.Y.; Park, M.K.; Choi, H.J.; An, H.W.; Park, Y.U.; Choi, H.J.; Park, J.; Kim, H.Y.; Son, T.; Lee, H.; et al. NSD3-Induced Methylation of H3K36 Activates NOTCH Signaling to Drive Breast Tumor Initiation and Metastatic Progression. *Cancer Res.* **2021**, *81*, 77–90. [[CrossRef](#)]
13. Katsuta, E.; Rashid, O.M.; Takabe, K. Clinical relevance of tumor microenvironment: Immune cells, vessels, and mouse models. *Hum. Cell* **2020**, *33*, 930–937. [[CrossRef](#)]
14. Fridman, W.H.; Pages, F.; Sautes-Fridman, C.; Galon, J. The immune contexture in human tumours: Impact on clinical outcome. *Nat. Rev. Cancer* **2012**, *12*, 298–306. [[CrossRef](#)]
15. Brennan, K.; Shin, J.H.; Tay, J.K.; Prunello, M.; Gentles, A.J.; Sunwoo, J.B.; Gevaert, O. NSD1 inactivation defines an immune cold, DNA hypomethylated subtype in squamous cell carcinoma. *Sci. Rep.* **2017**, *7*, 17064. [[CrossRef](#)]
16. Sanchez-Vega, F.; Mina, M.; Armenia, J.; Chatila, W.K.; Luna, A.; La, K.C.; Dimitriadoy, S.; Liu, D.L.; Kantheti, H.S.; Saghafinia, S.; et al. Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **2018**, *173*, 321–337. [[CrossRef](#)] [[PubMed](#)]
17. Ji, G.W.; Wang, K.; Xia, Y.X.; Wang, J.S.; Wang, X.H.; Li, X.C. Integrating Machine Learning and Tumor Immune Signature to Predict Oncologic Outcomes in Resected Biliary Tract Cancer. *Ann. Surg. Oncol.* **2020**. [[CrossRef](#)]
18. Bindea, G.; Galon, J.; Mlecnik, B. CluePedia Cytoscape plugin: Pathway insights using integrated experimental and in silico data. *Bioinformatics* **2013**, *29*, 661–663. [[CrossRef](#)]
19. Bindea, G.; Mlecnik, B.; Hackl, H.; Charoentong, P.; Tosolini, M.; Kirilovsky, A.; Fridman, W.H.; Pages, F.; Trajanoski, Z.; Galon, J. ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **2009**, *25*, 1091–1093. [[CrossRef](#)]
20. Garnett, M.J.; Edelman, E.J.; Heidorn, S.J.; Greenman, C.D.; Dastur, A.; Lau, K.W.; Greninger, P.; Thompson, I.R.; Luo, X.; Soares, J.; et al. Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **2012**, *483*, 570–575. [[CrossRef](#)]
21. Iorio, F.; Knijnenburg, T.A.; Vis, D.J.; Bignell, G.R.; Menden, M.P.; Schubert, M.; Aben, N.; Goncalves, E.; Barthorpe, S.; Lightfoot, H.; et al. A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* **2016**, *166*, 740–754. [[CrossRef](#)]
22. Sauerbrei, W.; Taube, S.E.; McShane, L.M.; Cavenagh, M.M.; Altman, D.G. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): An Abridged Explanation and Elaboration. *J. Natl. Cancer Inst.* **2018**, *110*, 803–811. [[CrossRef](#)] [[PubMed](#)]
23. Remmele, W.; Stegner, H.E. [Recommendation for uniform definition of an immunoreactive score (IRS) for immunohistochemical estrogen receptor detection (ER-ICA) in breast cancer tissue]. *Pathologe* **1987**, *8*, 138–140.
24. Curtis, C.; Shah, S.P.; Chin, S.F.; Turashvili, G.; Rueda, O.M.; Dunning, M.J.; Speed, D.; Lynch, A.G.; Samarajiwa, S.; Yuan, Y.; et al. The genomic and transcriptomic architecture of 2000 breast tumours reveals novel subgroups. *Nature* **2012**, *486*, 346–352. [[CrossRef](#)]
25. Newman, A.M.; Liu, C.L.; Green, M.R.; Gentles, A.J.; Feng, W.; Xu, Y.; Hoang, C.D.; Diehn, M.; Alizadeh, A.A. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **2015**, *12*, 453–457. [[CrossRef](#)]
26. Yang, W.; Soares, J.; Greninger, P.; Edelman, E.J.; Lightfoot, H.; Forbes, S.; Bindal, N.; Beare, D.; Smith, J.A.; Thompson, I.R.; et al. Genomics of Drug Sensitivity in Cancer (GDSC): A resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **2013**, *41*, D955–D961. [[CrossRef](#)]
27. El-Nachef, D.; Oyama, K.; Wu, Y.Y.; Freeman, M.; Zhang, Y.; MacLellan, W.R. Repressive histone methylation regulates cardiac myocyte cell cycle exit. *J. Mol. Cell. Cardiol.* **2018**, *121*, 1–12. [[CrossRef](#)]
28. Haydn, T.; Metzger, E.; Schuele, R.; Fulda, S. Concomitant epigenetic targeting of LSD1 and HDAC synergistically induces mitochondrial apoptosis in rhabdomyosarcoma cells. *Cell Death Dis.* **2017**, *8*, e2879. [[CrossRef](#)]
29. Higgs, M.R.; Sato, K.; Reynolds, J.J.; Begum, S.; Bayley, R.; Goula, A.; Vernet, A.; Paquin, K.L.; Skalnik, D.G.; Kobayashi, W.; et al. Histone Methylation by SETD1A Protects Nascent DNA through the Nucleosome Chaperone Activity of FANCD2. *Mol. Cell.* **2018**, *71*, 25–41.e26. [[CrossRef](#)]
30. Huang, Y.; Nayak, S.; Jankowitz, R.; Davidson, N.E.; Oesterreich, S. Epigenetics in breast cancer: what's new? *Breast Cancer Res.* **2011**, *13*, 225. [[CrossRef](#)]
31. Gelsi-Boyer, V.; Orsetti, B.; Cervera, N.; Finetti, P.; Siricoulomb, F.; Rouge, C.; Lasorsa, L.; Letessier, A.; Ginestier, C.; Monville, F.; et al. Comprehensive profiling of 8p11-12 amplification in breast cancer. *Mol. Cancer Res.* **2005**, *3*, 655–667. [[CrossRef](#)]
32. He, C.; Li, F.D.; Zhang, J.H.; Wu, J.H.; Shi, Y.Y. The Methyltransferase NSD3 Has Chromatin-binding Motifs, PHD5-C5HCH, That Are Distinct from Other NSD (Nuclear Receptor SET Domain) Family Members in Their Histone H3 Recognition. *J. Biol. Chem.* **2013**, *288*, 4692–4703. [[CrossRef](#)]

33. Li, Y.; Trojer, P.; Xu, C.F.; Cheung, P.; Kuo, A.; Drury, W.J., 3rd; Qiao, Q.; Neubert, T.A.; Xu, R.M.; Gozani, O.; et al. The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *J. Biol. Chem.* **2009**, *284*, 34283–34295. [[CrossRef](#)]
34. Morishita, M.; Mevius, D.; di Luccio, E. In vitro histone lysine methylation by NSD1, NSD2/MMSET/WHSC1 and NSD3/WHSC1L. *BMC Struct. Biol.* **2014**, *14*, 25. [[CrossRef](#)]
35. Turner-Ivey, B.; Smith, E.L.; Rutkovsky, A.C.; Spruill, L.S.; Mills, J.N.; Ethier, S.P. Development of mammary hyperplasia, dysplasia, and invasive ductal carcinoma in transgenic mice expressing the 8p11 amplicon oncogene NSD3. *Breast Cancer Res. Treat.* **2017**, *164*, 349–358. [[CrossRef](#)] [[PubMed](#)]
36. Liu, Z.; Piao, L.; Zhuang, M.; Qiu, X.; Xu, X.; Zhang, D.; Liu, M.; Ren, D. Silencing of histone methyltransferase NSD3 reduces cell viability in osteosarcoma with induction of apoptosis. *Oncol. Rep.* **2017**, *38*, 2796–2802. [[CrossRef](#)]
37. Saloura, V.; Vougiouklakis, T.; Zewde, M.; Deng, X.; Kiyotani, K.; Park, J.H.; Matsuo, Y.; Lingen, M.; Suzuki, T.; Dohmae, N.; et al. WHSC1L1-mediated EGFR mono-methylation enhances the cytoplasmic and nuclear oncogenic activity of EGFR in head and neck cancer. *Sci. Rep.* **2017**, *7*, 40664. [[CrossRef](#)]
38. Yi, L.; Yi, L.; Liu, Q.; Li, C. Downregulation of NSD3 (WHSC1L1) inhibits cell proliferation and migration via ERK1/2 deactivation and decreasing CAPG expression in colorectal cancer cells. *Onco. Targets Ther.* **2019**, *12*, 3933–3943. [[CrossRef](#)]
39. Chen, D.S.; Mellman, I. Oncology meets immunology: The cancer-immunity cycle. *Immunity* **2013**, *39*, 1–10. [[CrossRef](#)]
40. Jasek-Gajda, E.; Jurkowska, H.; Jasinska, M.; Litwin, J.A.; Lis, G.J. Combination of ERK2 inhibitor VX-11e and voreloxin synergistically enhances anti-proliferative and pro-apoptotic effects in leukemia cells. *Apoptosis* **2019**, *24*, 849–861. [[CrossRef](#)]
41. He, X.; Guo, X.; Zhang, H.; Kong, X.; Yang, F.; Zheng, C. Mechanism of action and efficacy of LY2109761, a TGF-beta receptor inhibitor, targeting tumor microenvironment in liver cancer after TACE. *Oncotarget* **2018**, *9*, 1130–1142. [[CrossRef](#)] [[PubMed](#)]
42. Ladygina, N.; Gottipati, S.; Ngo, K.; Castro, G.; Ma, J.Y.; Banie, H.; Rao, T.S.; Fung-Leung, W.P. PI3K gamma kinase activity is required for optimal T-cell activation and differentiation. *Eur. J. Immunol.* **2013**, *43*, 3183–3196. [[CrossRef](#)] [[PubMed](#)]
43. Ding, X.; Qu, X.; Fan, Y.; Che, X.; Qu, J.; Xu, L.; Liu, J.; Liu, Y. Trastuzumab and oxaliplatin exhibit a synergistic antitumor effect in HER2-positive gastric cancer cells. *Anticancer. Drugs* **2014**, *25*, 315–322. [[CrossRef](#)] [[PubMed](#)]
44. Gridelli, C.; Rossi, A.; Ciardiello, F.; De Marinis, F.; Crino, L.; Morabito, A.; Morgillo, F.; Montanino, A.; Daniele, G.; Piccirillo, M.C.; et al. BEVERLY: Rationale and Design of a Randomized Open-Label Phase III Trial Comparing Bevacizumab Plus Erlotinib Versus Erlotinib Alone as First-Line Treatment of Patients With EGFR-Mutated Advanced Nonsquamous Non-Small-Cell Lung Cancer. *Clin. Lung Cancer* **2016**, *17*, 461–465. [[CrossRef](#)] [[PubMed](#)]