Differential expression of the Wolf-Hirschhorn syndrome candidate 1, WHSC1 (NSD2/KMT3G), in cancers of the breast.

Shahan Mamoor, MS¹

¹shahanmamoor@gmail.com

East Islip, NY 11730

Breast cancer affects women at relatively high frequency¹. We mined published microarray datasets^{2,3} to determine in an unbiased fashion and at the systems level genes most differentially expressed in the primary tumors of patients with breast cancer. We report here significant differential expression of the gene encoding the H3K36 methyltransferase Wolf-Hirschhorn syndrome candidate 1, WHSC1, also known as Nsd2 or KMT3G, when comparing primary tumors of the breast to the tissue of origin, the normal breast. WHSC1 expression in primary tumors of the breast was significantly higher than in normal breast tissue, and relapse-free survival in breast cancer patients was significantly worse in patients with high tumor expression of WHSC1 than in patients with low tumor expression of WHSC1. WHSC1 may be of relevance to initiation, maintenance or progression of cancers of the female breast.

Keywords: breast cancer, WHSC1, KMT3G, NSD2, systems biology of breast cancer, targeted therapeutics in breast cancer.

Invasive breast cancer is diagnosed in over a quarter of a million women in the United States each year¹ and in 2018, breast cancer was the leading cause of cancer death in women worldwide⁴. While patients with localized breast cancer are provided a 99% 5-year survival rate, patients with regional breast cancer, cancer that has spread to lymph nodes or nearby structures, are provided an 86% 5-year survival rate^{5,6}. Patients with metastasis to distant sites are provided a 27% 5-year survival rate^{5,6}. Understanding how primary tumors are most transcriptionally different from the tissue from which they originate, the breast, can facilitate development of novel diagnostic and therapeutics to promote early detection and enhanced treatment, and contribute to efforts to prevent progression to metastatic stages. We mined published microarray data^{2,3} to understand at the transcriptome level and in an unbiased fashion genes most differentially expressed in primary tumors of the breast as compared to normal breast tissue. The histone methyltransferase KMT3G, also known as NSD2, or the Wolf-Hirschhorn syndrome candidate 1 (WSHC1) emerged as among the most differentially expressed genes in cancer of the female breast.

Methods

We utilized datasets GSE124648² and GSE109169³ and for this differential gene expression analysis of female breast cancer. GSE124648 was generated using Affymetrix Human Genome U133A Array technology with $n=10$ normal breast tissue and $n=126$ tumors of the breast and; analysis was performed using platform GPL96. GSE109169 was generated using Affymetrix Human Exon 1.0 ST Array technology with $n=25$ normal breast tissue and $n=25$ tumors of the breast and; analysis was performed using platform GPL5175. The tissues from this first dataset are paired tissues (25 tumors matching 25 breast tissues from 25 patients). The Benjamini and Hochberg method of p -value adjustment was used for ranking of differential

expression but raw p -values were used to assess statistical significance of global differential expression. Log-transformation of data was auto-detected, and the NCBI generated category of platform annotation was used. A statistical test was performed to evaluate whether WHSC1 expression was significantly between primary breast tumors and breast tissue using a two-tailed, unpaired t-test with Welch's correction. We used PRISM for all statistical analyses of differential gene expression in human breast cancer (Version 8.4.0)(455). For Kaplan-Meier survival analysis, we used the Kaplan-Meier plotter online tool⁷ for correlation of WHSC1 mRNA expression levels with relapse-free survival (RFS) in $n=3951$ breast cancer patients.

Results

We performed discovery of genes associated with breast cancer in females by mining two independently published microarray datasets^{2,3}.

WHSC1 is differentially expressed in primary tumors of the breast.

Comparison of 126 tumors of the breast to 10 normal breast tissues revealed that the gene encoding the histone methyltransferase, WHSC1, was among the genes most differentially expressed in tumors of the breast in female breast cancer (Table 1). When sorting each of the genes expressed in primary breast tumors based on significance of difference as compared to normal breast tissue, WHSC1 ranked 8 out of 22283 total transcripts, equating to 99.96% differential expression (Table 1). Differential expression of WHSC1 in female breast cancer was statistically significant (Table 1; $p=1.41E-23$).

Analysis of a second microarray dataset³, here comparing 25 tumors of the breast to 25 normal breast tissues² again revealed significant differential expression of WHSC1 in tumors of the breast in female breast cancer (Table 2). When sorting each of the genes expressed in

tumors of the breast based on significance of difference as compared to normal breast tissue, WHSC1 ranked 1208 out of 19076 total transcripts, equating to 93.7% differential expression (Table 2). Differential expression of WHSC1 in female breast cancer was again statistically significant (Table 2; $p=2.01E-08$).

WHSC1 is expressed at significantly higher levels in breast tumors as compared to the breast.

We obtained exact mRNA expression levels for WHSC1 from the breast and from breast tumors to understand the magnitude and direction of WHSC1 expression change in human breast cancer. In both datasets, WHSC1 was expressed at higher levels in tumors of the breast as compared to normal breast tissue (Figure 1). Increased expression of WHSC1 in primary breast tumors was statistically significant (Figure 1: $p<0.0001$). We calculated a mean fold change of 1.7962 ± 0.2065 when comparing primary tumors of the breast to normal breast tissues (Table 1).

WHSC1 expression associates with survival outcomes in breast cancer.

We performed Kaplan-Meier survival analysis in an attempt to correlate WHSC1 mRNA expression levels with survival outcomes in 3951 breast cancer patients. We found a statistically significant correlation between WHSC1 expression and relapse-free survival (RFS) in patients with breast cancer (Figure 2; log rank p -value: $1.1e-09$ for relapse-free survival, hazard ratio: 1.4 (1.26-1.57) (Fig. 2)). High WHSC1 mRNA levels were a negative prognostic indicator in patients with breast cancer. Median RFS was 65.1 months for patients with low tumor expression of WHSC1 while median RFS was 37.51 months for patients with high tumor

1 expression of WHSC1 (Table 3).

2 Thus, through comparative transcriptome analysis of primary tumors of the breast and
3 normal breast tissue, we found that differential expression and up-regulation of WHSC1 was
4 among the most significant transcriptional features in primary tumors from patients with breast
5 cancer. WHSC1 expression in primary tumors of the breast demonstrated a significant
6 correlation with relapse-free survival in patients with breast cancer, with low expression of
7 WHSC1 correlating with improved survival.
8

9 **Discussion**

10 NSD2 functions in deposition of the H3K36 mark: di-methylation of histone H3 at lysine
11 36⁸. NSD2 has features of a transcriptional co-repressor, forming a complex with HDAC1,
12 HDAC2, mSin3a and LSD1 in multiple myelomas⁹ but can also positively function in
13 transcription mediated by the androgen receptor, interacting with the AR DNA-binding domain
14 through its HMG domain¹⁰. NSD2 contains a 2 PWWP domains, an HMG domain, the
15 methyltransferase SET domain, and a carboxy-terminal PHD finger¹⁰. NSD2 is required for
16 development of the follicular helper T-cell lineage¹¹. One relevant target of NSD2 methylation in
17 cancer is PTEN - more specifically, the nuclear functions of PTEN. Upon double-stranded DNA
18 breakage, ATM phosphorylates T/S398 of PTEN; PTEN is also di-methylated by NSD2 at lysine
19 349, which is then bound by the tudor domain of 53BP1 for recruitment of PTEN to sites of DNA
20 damage¹². Thus, by the combined action of NSD2 and ATM, DNA double-stranded breaks
21 promote interaction of the mediator of DNA damage checkpoint 1, MDC1, with PTEN¹².
22

23 In the context of cancer, NDS2 can support growth of multiple myeloma cells though
24 MYC, which is regulated by the microRNA miR-126* ¹³, and in this context NSD2 is recruited to
25 target genes through its PHD finger domain¹⁴. Interestingly, *in vivo*, NSD2 is required for the
26
27
28

1 tumorigenesis of t(4;14)+ but not t(4;14)- multiple myelomas¹⁴. Another NSD family member,
2 NSD3, has been recently described to promote progression of breast cancer by catalyzing
3 methylation of lysine 36 on histone H3; remarkably, NSD3 functioned comparably to HRAS in
4 transformation of mammary epithelial cells *in vivo*¹⁵. High expression of NSD3 was associated
5 with worse survival and distant metastasis in patients with breast cancer¹⁵.
6

7 We report here that the H3K36 histone methyltransferase NSD2, also known as the
8 Wolf-Hirschhorn syndrome candidate or KMT3G, encoded by the WHSC1 gene is among the
9 genes most differentially expressed transcriptome-wide in human breast cancer when
10 comparing primary tumors of the breast to the tissue of origin, normal breast tissue. WHSC1
11 was expressed at significantly higher levels in primary breast tumors than in the breast, and low
12 tumor expression of WHSC1 was associated with enhanced relapse-free survival in patients
13 with breast cancer. WHSC1/NSD2/KMT3G may be relevant to the initiation or progression of
14 human breast cancer.
15
16
17
18
19
20
21
22
23
24
25
26
27
28

References

1. DeSantis, C.E., Ma, J., Goding Sauer, A., Newman, L.A. and Jemal, A., 2017. Breast cancer statistics, 2017, racial disparity in mortality by state. *CA: a cancer journal for clinicians*, 67(6), pp.439-448.
2. Sinn, B.V., Fu, C., Lau, R., Litton, J., Tsai, T.H., Murthy, R., Tam, A., Andreopoulou, E., Gong, Y., Murthy, R. and Gould, R., 2019. SET ER/PR: a robust 18-gene predictor for sensitivity to endocrine therapy for metastatic breast cancer. *NPJ breast cancer*, 5(1), pp. 1-8.
3. Chang, J.W., Kuo, W.H., Lin, C.M., Chen, W.L., Chan, S.H., Chiu, M.F., Chang, I.S., Jiang, S.S., Tsai, F.Y., Chen, C.H. and Huang, P.H., 2018. Wild-type p53 upregulates an early onset breast cancer-associated gene GAS7 to suppress metastasis via GAS7–CYFIP1-mediated signaling pathway. *Oncogene*, 37(30), pp.4137-4150.
4. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. and Jemal, A., 2018. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*, 68(6), pp.394-424.
5. ACS Cancer Facts & Figures 2019. <https://www.cancer.net/cancer-types/breast-cancer-metastatic/statistics>.
6. Survival Rates for Breast Cancer. <https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html>
7. Györfy, B., Lanczky, A., Eklund, A.C., Denkert, C., Budczies, J., Li, Q. and Szallasi, Z., 2010. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast cancer research and treatment*, 123(3), pp.725-731.
8. Kuo, A.J., Cheung, P., Chen, K., Zee, B.M., Kioi, M., Lauring, J., Xi, Y., Park, B.H., Shi, X., Garcia, B.A. and Li, W., 2011. NSD2 links dimethylation of histone H3 at lysine 36 to oncogenic programming. *Molecular cell*, 44(4), pp.609-620.
9. Marango, J., Shimoyama, M., Nishio, H., Meyer, J.A., Min, D.J., Sirulnik, A., Martinez-Martinez, Y., Chesi, M., Bergsagel, P.L., Zhou, M.M. and Waxman, S., 2008. The MMSET protein is a histone methyltransferase with characteristics of a transcriptional corepressor. *Blood, The Journal of the American Society of Hematology*, 111(6), pp. 3145-3154.

10. Kang, H.B., Choi, Y., Lee, J.M., Choi, K.C., Kim, H.C., Yoo, J.Y., Lee, Y.H. and Yoon, H.G., 2009. The histone methyltransferase, NSD2, enhances androgen receptor-mediated transcription. *FEBS letters*, 583(12), pp.1880-1886.
11. Long, X., Zhang, L., Zhang, Y., Min, M., Lin, B., Chen, J., Ma, X., Zhai, S., Cai, Z., Liu, Y. and Lu, Y., 2020. Histone methyltransferase Nsd2 is required for follicular helper T cell differentiation. *The Journal of Experimental Medicine*, 217(1).
12. Zhang, J., Lee, Y.R., Dang, F., Gan, W., Menon, A.V., Katon, J.M., Hsu, C.H., Asara, J.M., Tibarewal, P., Leslie, N.R. and Shi, Y., 2019. PTEN methylation by NSD2 controls cellular sensitivity to DNA damage. *Cancer discovery*, 9(9), pp.1306-1323.
13. Min, D.J., Ezponda, T., Kim, M.K., Will, C.M., Martinez-Garcia, E., Popovic, R., Basrur, V., Elenitoba-Johnson, K.S. and Licht, J.D., 2013. MMSET stimulates myeloma cell growth through microRNA-mediated modulation of c-MYC. *Leukemia*, 27(3), pp.686-694.
14. Huang, Z., Wu, H., Chuai, S., Xu, F., Yan, F., Englund, N., Wang, Z., Zhang, H., Fang, M., Wang, Y. and Gu, J., 2013. NSD2 is recruited through its PHD domain to oncogenic gene loci to drive multiple myeloma. *Cancer research*, 73(20), pp.6277-6288.
15. 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. and Min, K.W., 2020. NSD3-induced methylation of H3K36 activates NOTCH signaling to drive breast tumor initiation and metastatic progression. *Cancer Research*.

Rank	ID	p-value	t	B	FC	Gene	Gene name
8	209053_s_at	1.41E-23	12.180231	42.93597	1.7962 ± 0.2065	WHSC1	Wolf-Hirschhorn syndrome candidate 1

Table 1: WHSC1 is differentially expressed in human breast cancer.

Rank of differential expression, probe ID, p-value of global differential expression, t, a moderated t-statistic, B, the log-odds of differential expression between the groups compared, fold change of WHSC1 in primary breast tumors as compared to normal breast, gene and gene name are listed in this chart.

Rank	ID	p-value	t	B	Gene	Gene name
1208	2715076	2.01E-08	6.64	9.026643	WHSC1	Wolf-Hirschhorn syndrome candidate 1

Table 2: WHSC1 is differentially expressed in human breast cancer.

Rank of differential expression, probe ID, *p*-value of global differential expression, *t*, a moderated *t*-statistic, *B*, the log-odds of differential expression between the groups compared, gene and gene name are listed in this chart.

Figure 1: WHSC1 is expressed at significantly higher levels in primary breast tumors as compared to normal breast tissue.

The mRNA expression level of WHSC1 in normal breast tissue (left) and in primary tumors of the breast (right) is graphically depicted with the result of a statistical test evaluating significance of difference in WHSC1 expression between normal breast tissue and primary tumors of the breast, a *p*-value, listed above.

Figure 2: WHSC1 expression correlates with relapse-free survival in patients with breast cancer.

Depicted in this Kaplan-Meier plot is the probability of relapse-free survival (RFS) for $n=3951$ total patients stratified into two groups, based on low or high expression of WHSC1 in patient primary tumors. The log rank p -value denoting statistical significance of difference in distant relapse-free survival when comparing the two groups, as well as hazard ratio for this comparison is listed above. Listed below is the number of patients at risk (number of patients alive) per interval, after stratification based on WHSC1 expression; in the first interval, number at risk is number of patients alive; in each subsequent interval, number at risk is the number at risk less those who have expired or are censored.

Low expression cohort (months)	High expression cohort (months)
65.1	37.51

Table 3: Median relapse-free survival is greater in breast cancer patients with low primary tumor expression of WHSC1.

The relapse-free survival of $n=3951$ breast cancer patients based on stratification into low or high tumor expression of WHSC1 is listed in this chart.