AD_____

Award Number: W81XWH-09-1-0109

TITLE: Histone Code Modulation by Oncogenic PWWP-Domain Protein in Breast Cancers

PRINCIPAL INVESTIGATOR: Zeng-Quan Yang

CONTRACTING ORGANIZATION: WAYNE STATE UNIVERSITY, Detroit, MI 48201

REPORT DATE: June, 2010

TYPE OF REPORT: annual report

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT:

☒ Approved for public release; distribution unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-06-2010	Annual	1 Jun 2009 - 31 May 2010

4. TITLE AND SUBTITLE
Histone Code Modulation by Oncogenic PWWP-Domain Protein

in Breast Cancers

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-09-1-0109

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Zeng-Quan Yang

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Wayne State University
Detroit, Michigan 48201

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Earlier, we used genomic analysis of copy number and gene expression to perform a detailed analysis of the 8p11-12 amplicon to identify candidate oncogenes in breast cancer. We identified 21 candidate genes and provided evidence that three genes; LSM-1, TC-1 and BAG4 have transforming properties when over expressed. Recently, we systematically investigated the transforming properties of thirteen newly identified 8p11-12 candidate oncogenes in vitro. We found that WHSC1L1, DDHD2 and ERLIN2 are the most potently transforming oncogenes we tested from the 8p11-12 region based on the number of altered phenotypes expressed by the cells. WHSC1L1 contains a PWWP-domain that is a methyl-lysine recognition motif involved in histone code modification and epigenetic regulation of gene expression. Knockdown of WHSC1L1 in 8p11-12 amplified breast cancer cells resulted in profound loss of growth and survival of these cells. Further, we identified several WHSC1L1 target genes, one of which is iroquois homeobox 3 gene (IRX3), a member of the Iroquois homeobox transcription factor family.

15. SUBJECT TERMS
Gene amplification, PWWP-domain, histone modification

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	35	19b. TELEPHONE NUMBER *(include area code)*

Introduction

Amplification of the 8p11-12 region has been found in about 10-15% of human breast cancer and may be associated with poor prognosis. Earlier, we used genomic analysis of copy number and gene expression to perform a detailed analysis of the 8p11-12 amplicon for identifying candidate oncogenes in breast cancer. We identified Wolf-Hirschhorn syndrome candidate 1-like 1 (WHSC1L1) as a candidate oncogene based on statistical analysis of copy number increase and over expression. The WHSC1L1 gene encodes a PWWP domain protein that regulates histone methylation marks, and regulates the transcription profile and the differentiated function of cells. To address the question of whether WHSC1L1 possesses transforming properties, we established lentiviral expression constructs containing the WHSC1L1 as well as other 20 candidate genes of 8p11-12 region. MCF10A cells were infected with these lentiviral vectors and assayed for alterations in growth rates, growth factor-independent proliferation, anchorage-independent growth and for three-dimensional morphogenesis assays in Matrigel culture. From these experiments, we found that the WHSC1L1 is the most potently transforming oncogene we tested from the 8p11-12 region. In this DOD Idea award, we have proposed to elucidate fundamental mechanisms by which amplification and over expression of WHSC1L1 results in histone code modulation and neoplastic transformation of luminal breast cancer cells, and to develop therapeutic strategies to target the involvement of WHSC1L1 in hormone refractory breast cancers.

Body

1. Specific Aims

This project consists of 3 specific aims:
Aim 1: To investigate the molecular mechanism, including the structural details, of WHSC1L1 that are involved in their transforming function through the alteration of the epigenetic histone code in human breast cancer cells.
Aim 2: To determine whether the histone modulation function of WHSC1L1 is linked to cancer stem cell phenotypes.
Aim : To examine the potential of WHSC1L1 as a therapeutic target in aggressive, ER-postive breast cancers that harbor the 8p11 amplicon.

2. Studies and Results

***Task 1*. To investigate the molecular mechanism, including the structural details, of WHSC1L1 that are involved in their transforming function through the alteration of the epigenetic histone code in human breast cancer cells**. *Month 1-16*

Previously, our group identified 21 candidate oncogenes within 8p11-12 amplicon in breast cancer based on statistical analysis of copy number increase and gene over expression. We tested eight of the 21 candidate oncogenes for transforming function *in vitro* and identified three genes, LSM1, BAG4 and C8orf4 (TC-1), that could induce transformed phenotypes. Recently, we expanded our analysis to the remaining 13 candidate oncogenes including WHSC1L1. We established lentiviral expression constructs containing the WHSC1L1 as well as other 12 candidate genes of 8p11-12 region. MCF10A cells were infected with these lentiviral vectors and assayed for alterations in growth rates, growth factor-independent proliferation, anchorage-independent growth and for three-dimensional morphogenesis assays in Matrigel culture. From these experiments, we found that the WHSC1L1 is the most potently transforming oncogene we tested from the 8p11-12 region (Append Figure 1 and 2).

Expression of the WHSC1L1 gene results in two alternatively spliced variants, a long isoform and a short isoform that are derived from alternative splicing of exon 10. The WHSC1L1 long isoform encodes a 1437 amino acid protein containing 2 PWWP domains, 2 PHD-type zinc finger motifs, a TANG2 domain, an AWS domain and a SET domain. The short isoform encodes a 645 amino acid protein containing a PWWP domain only. Our western blot assays demonstrates that both SUM-44 and SUM-52 cells have amplifications of the full-length gene, but at the protein level expression of the short isoform predominates. That expression of the short isoform of WHSC1L1 is highly transforming in MCF10A cells, and the observation that, at the protein level, expression of the short isoform predominantes in breast cancer cells suggests an important role for the short isoform of the protein in breast cancer development. Interestingly, we identified one primary breast cancer specimen (10173A) with the 8p11-12 amplicon in which array CGH demonstrated genomic loss of the C-terminal region of the WHSC1L1 long isoform but with amplification of exons 1-10 (Append Figure 3). We validated this finding in that particular breast cancer specimen by genomic PCR using primers specific for the short isoform exon 10 (S-10) and the long isoform exon 20 (L-20) as shown in Append figure 3. This result provides further evidence for the importance of the short isoform of WHSC1L1 that contains only PWWP domain in cell transformation when over expressed.

Because WHSC1L1 encodes a PWWP domain nuclear protein, it has been postulated that it can promote malignant transformation by altering the histone code and hence expression of specific target genes. To identify genes that may be altered in their expression by over expression of the short isoform of WHSC1L1, we performed expression profiling of MCF10A cells, MCF10A-WHSC1L1 cells, and SUM-44 cells. To identify genes most likely to be regulated by over expression of WHSC1L1 and relevant to human breast cancer, we determined which genes are differentially expressed in MCF10A-WHSC1L1 cells relative to parental MCF10A cells, and then determined which of those genes are also differentially expressed in SUM-44 cells compared to MCF10A cells. This orthogonal analysis resulted in the identification of 184 genes differentially expressed in both SUM-44 cells and MCF10A-WHSC1L1 cells, relative to MCF10A cells. Of the 184 differentially expressed genes, 36 are coordinately up-regulated in MCF10A-WHSC1L1 cells and SUM-44 cells. Append figure 5A shows the four up-regulated genes (TBL1X, IRX3, RAG1AP1 and RAPGEF3), and two down-regulated genes (TFBI and SFRP1) in both SUM-44 cells and MCF10A-WHSC1L1 cells, relative to MCF10A cells. Thus, these genes could be directly altered in their expression by over expression of WHSC1L1-short isoform. To directly validate some of these array-based observations, we chose three up-regulated genes and one down-regulated gene to examine by Q-RT-PCR (Append figure 5B). Figure 5 shows that IRX3, RAPGEF3, and TBLX1 are significantly over expressed at the mRNA level in SUM-44 cells compared to MCF10A cells. Furthermore, knock-down of WHSC1L1 in SUM-44 cells using the shRNA constructs described in task 3 resulted in significant down regulation of these three putative target genes. These results support the array-based analysis and indicate that WHSC1L1 regulates the expression of these target genes. Similarly, examination of one down-regulated gene, TGFBI also confirmed reduced expression in SUM-44 compared to MCF10A cells, and the expression of this gene was increased in SUM-44 cells bearing the WHSC1L1 shRNA constructs.

To further elucidate the relationship between the transforming function and specific structural motifs, particularly the PWWP domain, of the WHSC1L1 gene, we have established a series of WHSC1L1 deletion constructs that delete PWWP or SET domains. Each of these deletion constructs along with wild-type full-length constructs has been incorporated into the lentiviral expression system. In next year, we will generate lentivirus for all of these constructs and infect MCF10A cells. Next we will determine whether MCF10A cells expressing different truncated forms of WHSC1L1 express transformed phenotypes similar to the MCF10A cells over expressing the wild-type WHSC1L1 short isoform. These model cells will also be used to determine the genome-wide distribution of the histone modifying protein WHSC1L1 in mammary epithelial cells by ChIP-on-chip assays.

***Task2*. To determine whether the histone modulation function of WHSC1L1 is linked to cancer stem cell phenotypes**. Months 12-30

Recent studies have demonstrated that mammary stem and progenitor cells can be cultured in vitro as floating spherical colonies, termed "mammospheres". To examine whether over expression of WHSC1L1 modulate mammary stem self-renewal, *in vitro* mammosphere formation assay will be used in this task. We have successfully established the assay for mammosphere formation. Further, we tested several specific stem cell surface markers including aldehyde dehydrogenase (ALDH) by using flow cytometry in breast cancer cells. Next, we will used these methods to determine if altered expression of histone-modifying protein WHSC1L1 results in expansion/contraction of cancer stem cell pools.

Task 3. **To examine the potential of WHSC1L1 as a therapeutic target in aggressive, ER-postive breast cancers that harbor the 8p11 amplicon.** Months 18-36

To directly assess the contribution of endogenous WHSC1L1 over expression on the transformation of human breast cancer, we examined the effects of knock down of WHSC1L1 in SUM-44 and SUM-52 cells where WHSC1L1 is amplified and over expressed, and in the control cell line MCF10A. To perform RNAi knock-down experiments, we obtained eight pGIPZ-WHSC1L1 shRNA expression constructs from OpenBiosystems. (http://www.openbiosystems.com/). In this vector, TurboGFP and shRNA are part of a bicistronic transcript allowing the visual marking of the shRNA expressing cells. SUM-44, SUM-52 and control MCF10A cells were infected with these 8-shRNA lentivirus supernatants poolled or individually to determine which gave the best knock-down of WHSC1L1. Non-silencing shRNAmir lentiviral control, at the same titer as WHSC1L1 shRNA, was used in parallel as the negative control. First, the consequence of knock-down of WHSC1L1 using all eight shRNAs combined, on colony formation was evaluated in all three cell lines. WHSC1L1 knock-down suppressed proliferation of SUM-44 and SUM-52 cells, while WHSC1L1 shRNAs had no effect on the growth of MCF10A cells. Next, we identified the two most efficient shRNAs with respect to knock-down of WHSC1L1 expression levels in SUM-44 and SUM-52 cells. Q-RT-PCR and western blot data revealed that the WHSC1L1-shRNAs #2 and #6 resulted in decreases in mRNA and protein levels to approximately 20-30% of the level seen in the non-silencing control-infected cells (Append Figure 4). As shown in Figure 4B and C , WHSC1L1 knock-down with both shRNA constructs slowed cell growth of SUM-44 and SUM-52 cells. The results were most striking for SUM-44 cells in which WHSC1L1 knock-down inhibited cell proliferation by ~ 90% (Figure 4C). WHSC1L1 knock-down with these shRNA#2 and #6 had an undetectable effect on the cell growth of MCF10A cells (data no shown). Thus, knockdown of WHSC1L1 inhibits cell proliferation in breast cancer cells with WHSC1L1 gene amplification.

To our knowledge, several Pharmaceutical companies already initiated screening of small molecule libraries for identification of inhibitors targeting histone modifying protein WHSC1L1. Once available, WHSC1L1 inhibitors will be tested to investigate the contribution of endogenous WHSC1L1 over expression on the expression of transformed phenotypes in the luminal breast cancer cells with 8p11-12 amplification.

Key Research Accomplishments

In the present study, we systematically investigated the transforming properties of thirteen newly identified 8p11-12 candidate oncogenes *in vitro*. We found that *WHSC1L1* is the most potently transforming oncogene we tested from the 8p11-12 region based on the number of altered phenotypes expressed by the cells. Knockdown of *WHSC1L1* in 8p11-12 amplified breast cancer cells resulted in profound loss of growth and survival of these cells. Further, we identified the *iroquois homeobox 3* gene (*IRX3*), a member of the Iroquois homeobox transcription factor family, as a WHSC1L1 target gene.

Reportable Outcomes

Manuscript:
Two month ago, we submitted our manuscript "**Transforming properties of 8p11-12 amplified genes in human breast cancer**" to *Cancer Research*. We were pleased by the reviewer's comments that, "This manuscript provides important functional information needed to understand how amplification of 8p11-12 contributes to the pathophysiology of a subset of breast cancers." (Reviewer 1) and "the manuscript addresses an important issue in breast cancer genetics." (Reviewer 2). We will submit the revision of the manuscript in this month.

Abstract
"**Oncogenic PWWP-domain protein WHSC1L1 links the homeobox transcription factor IRX3 in breast cancer**" American Association for Cancer Research Annual Meeting in Washington, DC, April 17-21, 2010

Conclusion

We have made significant progress in the past year in characterizing the PWWP-domain protein WHSC1l1 in human breast cancer. As reported in our manuscript, we find WHSC1L1 to be the most potently transforming of all the 8p11 oncogenes we have tested. knockdown of this gene in WHSC1L1 amplified breast cancer cells resulted in profound loss of growth and survival potential in these cells. The PWWP-domain is a methyl-lysine recognition motif involved in histone code modification and epigenetic regulation of gene expression. To identify genes that may be altered in their expression by over expression of the WHSC1L1, we performed expression profiling of exogenous WHSC1L1 expressing MCF10A cells, and SUM-44 breast cancer cells. We identified several WHSC1L1 target genes, one of which is iroquois homeobox 3 gene (IRX3), a member of the Iroquois homeobox transcription factor family. In the next year, we will continue to investigate how WHSC1L1 contributes to its transformation through the alternation of epigenetic histone marks and acquisition of stem cell-like properties in breast cancer cells.

Appendices

Copy of the Manuscript: "Transforming properties of 8p11-12 amplified genes in human breast cancer"

Transforming properties of 8p11-12 amplified genes in human breast cancer

Zeng-Quan Yang[1], Gang Liu[1], Aliccia Bollig-Fischer and Stephen P. Ethier

Breast Cancer Program, Karmanos Cancer Institute, Department of Pathology, Karmanos Cancer Institute, Wayne State University, Detroit, MI 48201

Corresponding author: Zeng-Quan Yang, Ph.D. Karmanos Cancer Institute, 4100 John R HWCRC 815, Detroit, MI 48201. yangz@karmanos.org . ph: 313-576-8339 FAX: 313-576-8029.

[1] ZQY and GL contributed equally to this work

Running Title: Transforming properties of 8p11-12 genes

Key Words: gene amplification, breast cancer, WHSC1L1,

Abstract

Amplification of the 8p11-12 region has been found in about 15% of human breast cancer and is associated with poor prognosis. Earlier, we used genomic analysis of copy number and gene expression to perform a detailed analysis of the 8p11-12 amplicon to identify candidate oncogenes in breast cancer. We identified 21 candidate genes and provided evidence that three genes; LSM-1, TC-1 and BAG4 have transforming properties when over expressed. In the present study, we systematically investigated the transforming properties of thirteen newly identified 8p11-12 candidate oncogenes *in vitro*. We found that *WHSC1L1*, *DDHD2* and *ERLIN2* are the most potently transforming oncogenes we tested from the 8p11-12 region based on the number of altered phenotypes expressed by the cells. WHSC1L1 contains a PWWP-domain that is a methyl-lysine recognition motif involved in histone code modification and epigenetic regulation of gene expression. Knockdown of *WHSC1L1* in 8p11-12 amplified breast cancer cells resulted in profound loss of growth and survival of these cells. Further, we identified several WHSC1L1 target genes, one of which is *iroquois homeobox 3* gene (*IRX3*), a member of the Iroquois homeobox transcription factor family, and we obtained evidence that WHSC1L1 and IRX3 are genetically linked in subset of breast cancer.

Introduction

An important mechanism for the activation of oncogenes in human cancers is gene amplification, which results in gene over expression at both the message and protein levels (1, 2). Oncogenes, such as *ERBB2* at 17q12, *CCND1* at 11q13 and *C-MYC* at 8p24, have previously been identified as amplification targets linked to the development, progression, or metastasis of human cancers, including breast, prostate, lung and other cancers (2, 3). *ERBB2* is the most frequently amplified oncogene in breast cancer, and its over expression is associated with poor clinical outcomes. The prognostic and predictive values of ERBB2 amplification and over expression have been used to guide treatment decisions for patients with both lymph node-positive and negative diseases. More significantly, recognition of the mechanistic roles of ERBB2 in breast cancer has led to the development of ERBB2-targeting drugs such as Herceptin to treat breast cancer (4-6).

Amplification of 8p11-12 occurs in approximately 15% of human breast cancer (HBC), and this region of amplification is significantly associated with disease-specific survival and distant recurrence in breast cancer patients (7-11). These groups also identified several candidate oncogenes in this region. Chin *et al*. performed an analysis of the association of gene amplification and disease-free survival and distant relapse in human breast cancer specimens (12). They identified 23 genes from the 8p11-12 region as being correlated with progression. Recently, our laboratory published results of a detailed analysis of copy number and gene expression in the 8p11-12 region in a panel of breast cancer cell lines and primary human breast cancers (13). We identified 21 genes in this region that are over expressed when their copy number is increased (10). Furthermore, we directly tested the transforming function of eight 8p11-12 amplified genes in human mammary epithelial cells. From these experiments, we identified several genes including *LSM1*, *BAG4* and *C8orf4* (*TC-1*) as having the transforming properties *in vitro* (10, 14, 15). Accumulating evidence suggests that the 8p11-12 amplicon contains multiple candidate oncogenes that could play a role in breast cancer development (7-11).

Recent extensive genomic analyses and siRNA knock-down studies have identified the *Wolf-Hirschhorn syndrome candidate 1-like 1* gene (*WHSC1L1*, also known as *NSD3*) as one of major candidate oncogenes of the 8p11-12 amplicon in breast cancer (7-11). *WHSC1L1* is the third member of a gene family including *NSD1* and *WHSC1* (*NSD2*) (16, 17). *De novo* translocation of *NSD1* genes causes the childhood overgrowth syndrome, Sotos syndrome that is associated with elevated risks of cancer; while *de novo* deletion of *NSD2*/*WHSC1* causes the Wolf-Hirschhorn syndrome that displays growth retardation (18, 19). WHSC1L1/NSD3, NSD1 and WHSC1/NSD2 show strong sequence similarity, and share multiple functional domains (16). In addition, WHSC1L1 has two isoforms that are derived from alternative splicing of exon 10. The short isoform encodes a 645 amino acid protein containing a single PWWP domain. The PWWP domain belongs to the royal superfamily which includes chromodomain, tudor, malignant brain tumor (MBT) and plant agent motifs, and these domains exist in multiple histone modifying proteins. The N-terminal half of the PWWP domain exhibits a beta-barrel structure that resembles a SAND domain, while the C-terminal portion is made up of a five-helix bundle. Both the crystal and MMR solution structures of the superfamily complexes show that the beta-barrel structure recognizes and binds the histone lysine pocket (20, 21). A study on PWWP function in the DNA methyltransferase DNMT3B demonstrated that the PWWP domain binds methylated DNA (22). Recently, Wang *et al.* demonstrated that a PWWP domain protein binds to histone lysine *in vitro* and *in vivo*, and regulates Set9-mediated H4K20 methylation (23). Their results demonstrated that the PWWP domain is a methyl-lysine recognition motif that plays important roles in epigenetic regulation.

In the present study, we systematically investigated the transforming properties of thirteen newly identified 8p11-12 candidate oncogenes *in vitro*. We found that *WHSC1L1*, *DDHD2* and *ERLIN2* are the most potently transforming oncogenes we tested from the 8p11-12 region based on the number of altered phenotypes expressed by the cells. Knockdown of *WHSC1L1* in 8p11-12 amplified breast cancer cells resulted in profound loss of growth and survival of these cells. Further, we identified the *iroquois homeobox 3* gene (*IRX3*), a member of the Iroquois homeobox transcription factor family, as a

WHSC1L1 target gene, and we provide the first evidence that WHSC1L1 and IRX3 are genetically linked in subset of breast cancer.

Materials and Methods

Genomic array CGH

The isolation and culture of the SUM series of human breast cancer cell lines and MCF10A cells have been described in detail previously (10, 24). Genomic array CGH experiments were performed using the Agilent 44K human genome CGH microarray chip (Agilent Technologies, Palo Alto, CA). Agilent's CGH Analytics software was used to calculate various measurement parameters, including log2 ratio of total integrated Cy-5 and Cy-3 intensities for each probe.

Semiquantitative RT-PCR reactions

Total RNA was prepared from human breast cancer cell lines and the MCF10A cell line by standard methods (10, 25). For RT-PCR reactions, RNA was converted into cDNA via a reverse transcription reaction using random hexamer primers. Primers were ordered from Invitrogen (Carlsbad, CA) and all the relevant primer sequences are available on request. A GAPDH primer set was used as a control. Semiquantitative RT-PCR was done using the iQSYBR Green Supermix (Bio-Rad, Hercules, CA).

Lentivirus construction and transduction of cells

The lentiviral expression constructs containing the 13 genes tested in the present experiments, stared in Table 1, were established as previously described (10). Briefly,we first created entry clones from cDNA of SUM-44 cells using the pENTR directional TOPO cloning kit and then performed the LR recombination reaction to transfer the gene into the Gateway destination vector, pLenti6/V5-DEST. Specifically, the pLenti-WHSC1L1 construct was established from the full-length short isoform which only contained PWWP domain. The lentivirus for each construct was generated and used to infect the immortalized, nontransformed mammary epithelial MCF10A cells. Control infections with pLenti-LacZ virus were performed in parallel. Selection began 48 hours

after infection in growth medium with 10 µg/mL blasticidin in the absence of insulin. Upon confluence, selected cells were passaged and serially cultured.

Growth in soft agar and Matrigel

Soft agar assays were performed as previously described (10). For three-dimensional morphogenesis assays in Matrigel, cells grown in monolayer culture were detached by trypsin/EDTA treatment and seeded in Matrigel (BD Biosciences, San Jose, CA) precoated 8-well chamber slides. The appropriate volume of medium were added and maintained in culture for 10-18 days. Phase-contrast images and immunostaining images were photographed with bright-field and confocal microscopy (26).

Lentivirus-mediated shRNA knockdown of gene expression

We knocked down the expression of the human *WHSC1* gene in breast cancer cell lines, SUM-44 and SUM-52, and in the MCF10A cell line using the Expression Arrest GIPZ lentiviral shRNAmir system (OpenBiosystems, Huntsville, AL). Lentivirus was produced by transfecting 293FT cells with the combination of the lentiviral expression plasmid DNA and Trans-Lentiviral packaging mix (OpenBiosystems. Huntsville, AL). For cell infection, viral supernatants were supplemented with 6 µg/mL polybrene and incubated with cells for 24 h. Cells expressing shRNA were selected with puromycin for 2-3 wk for functional studies (cell proliferation and colony formation assays) and for 4 to 10 d after infection for RNA extraction.

Results

The effect of different 8p11-12 genes on growth factor independent proliferation

Recently, our group identified 21 candidate oncogenes within 8p11-12 amplicon in breast cancer based on statistical analysis of copy number increase and gene over expression. We tested eight of the 21 candidate oncogenes for transforming function *in vitro* and identified three genes, LSM1, BAG4 and C8orf4 (TC-1), that could induce transformed phenotypes (10). In the present report, we expanded our analysis to the

remaining 13 candidate oncogenes. Table 1 shows the original 21 gene list with the 13 genes tested in the present experiments highlighted with stars. Details on the origins and sequence validations of each clone are given in Materials and Methods, and in supplementary data.

To systematically investigate the transforming properties of thirteen 8p11-12 candidate breast cancer oncogenes, we transduced MCF10A cells, which are highly growth factor dependent, with individual lentiviral expression vectors for each gene. Growth factor independent proliferation of MCF10A cells transduced with each candidate gene was investigated. We first tested transduced cells for their ability to grow in the absence of insulin-like growth factors. Each lentiviral vector was packaged and tittered before infecting 3×10^4 MCF10A cells in 6-well plates. Forty-eight hours after infection, MCF10A cells were switched to media with blasticidin and devoid of insulin. After 3 weeks, colonies that emerged in growth factor-free media were stained to visualize colony formation in insulin-free medium. In parallel, cells from additional wells cells were pooled and propagated continuously in the same insulin-free media. RT-PCR was performed to confirm the expression of the gene using primers specific for the gene and for the vector. Over expression of WHSC1L1 protein in MCF10A-WHSC1L1 cells was further confirmed by western blot (Supplementary Figure S1). As shown in Figure 1, MCF10A cells expressing 6 genes WHSC1L1, DDHD2, PROSC, BRF2, ASH2L and ERLIN2 formed expanding colonies in insulin-free medium, and then grew continuously in the absence of insulin-like growth factors. Colony formation assays in MCF10A cells with equalized viral titer of the tested genes indicated that over expression of WHSC1L1 and DDHD2 resulted in the highest number of insulin-independent colonies. Growth curves of MCF10A cells over expressing the 5 genes (WHSC1L1, DDHD2, PROSC BRF2, and ERLIN2) were performed within two passages of isolation in insulin-free medium. Data in Figure 1B shows that over expression of WHSC1L1 not only resulted in the largest number of colonies emerging in insulin-free medium, but also gave rise to cells with the most rapid proliferation rate under these conditions. These results extend our previous findings and indicate that a total of 9 genes from the 8p11-12 have ability to

induce insulin-like growth factor independent proliferation when over expressed in MCF10A cells.

Transforming properties of 8p11-12 candidate genes

To assess the expression of other transformed phenotypes of MCF10A cells over expressing the newly identified candidate oncogenes, we evaluated each of them for their ability to form colonies in soft agar and for altered morphogenesis in Matrigel. Figure 2 shows that after 3 weeks in culture, MCF10A cells over expressing WHSC1L1, DDHD2, and ERLIN2 formed colonies in soft agar. MCF10A cells over expressing WHSC1L1 and DDHD2 had the highest soft-agar colony forming efficiency (Figure 2B). By contrast, MCF10A cells over expressing ASH2, BRF2, and PROSC did not form soft agar colonies. We also examined whether these six genes affect the growth or morphology of colonies in 3D Matrigel culture, as aberrant behavior in this environment is frequently associated with transformation and/or tumor progression (27). In 3D basement membrane cultures, the immortalized, nontransformed mammary epithelial cells, MCF10A, formed acinar-like structures consisting of a single cell layer of polarized, growth-arrested mammary epithelial cells surrounding a hollow lumen (Figure 2A). As shown in Figure 2A, MCF10A cells over expressing WHSC1L1 formed strikingly abnormal acini that were enlarged, disorganized, and contained filled lumens. In contract, MCF10A cells over expressing DDHD2 formed disorganized, small abnormal acini. MCF10A cells over expressing ERLIN2 also formed large, highly-proliferative colonies, while insulin-independent MCF10A cells over expressing the other three candidate oncogenes formed polarized, growth-arrested acinar structures with hollow lumens similar to MCF10A parental cells (Data not shown). These experiments demonstrate that three of the transforming genes, PROSC, ASH2L and BRF2, induced insulin-independent growth and no other altered phenotypes. By contrast, WHSC1L1, DDHD2, and ERLIN2 were the most transforming oncogenes based on the number of altered phenotypes expressed by the cells.

Amplification of WHSC1L1 isoforms in breast cancer

Expression of the WHSC1L1 gene results in two alternatively spliced variants, a long isoform and a short isoform that are derived from alternative splicing of exon 10. The WHSC1L1 long isoform encodes a 1437 amino acid protein containing 2 PWWP domains, 2 PHD-type zinc finger motifs, a TANG2 domain, an AWS domain and a SET domain. The short isoform encodes a 645 amino acid protein containing a PWWP domain only (Supplement Figure S2). Data shown in figure 3 demonstrates that both SUM-44 and SUM-52 cells have amplifications of the full-length gene, but at the protein level expression of the short isoform predominates. The transformation data for WHSC1L1 over expressing MCF10A cells shown above was obtained using an expression construct coding for the short isoform, and similar results were obtained when we transduced MCF10A cells with a vector coding for full length WHSC1L1 (data not shown). That expression of the short isoform of WHSC1L1 is highly transforming in MCF10A cells, and the observation that, at the protein level, expression of the short isoform predominantes in breast cancer cells suggests an important role for the short isoform of the protein in breast cancer development. Interestingly, we identified one primary breast cancer specimen (10173A) with the 8p11-12 amplicon in which array CGH demonstrated genomic loss of the C-terminal region of the WHSC1L1 long isoform but with amplification of exons 1-10. We validated this finding in that particular breast cancer specimen by genomic PCR using primers specific for the short isoform exon 10 (S-10) and the long isoform exon 20 (L-20) as shown in figure 3C and supplemental figure S2. This result provides further evidence for the importance of the short isoform of WHSC1L1 in cell transformation when over expressed. To further determine whether the WHSC1L1 short isoform protein, which only contains a PWWP domain, is also localized in the nucleus, we generated expression constructs containing the short isoform WHSC1L1 coding sequences fused to the EGFP epitope at the C-terminus. The constructs were transfected into MCF10A and COS7 cells, and localization of the proteins was examined by fluorescence microscopy. The WHSC1L1 short isoform was localized to the nucleus as expected (Supplementary Figure S3). These results implicate the short isoform of WHSC1L1 as the transforming oncoprotein that is over expressed in breast cancer cells bearing the 8p11p12 amplicon.

Knockdown of WHSC1L1 inhibits cell proliferation in breast cancer cells

To directly assess the contribution of endogenous WHSC1L1 over expression on the transformation of HBC, we examined the effects of knock down of WHSC1L1 in SUM-44 and SUM-52 cells where WHSC1L1 is amplified and over expressed, and in the control cell line MCF10A. To perform RNAi knock-down experiments, we obtained eight pGIPZ-WHSC1L1 shRNA expression constructs from OpenBiosystems. (http://www.openbiosystems.com/). In this vector, TurboGFP and shRNA are part of a bicistronic transcript allowing the visual marking of the shRNA expressing cells. SUM-44, SUM-52 and control MCF10A cells were infected with these 8-shRNA lentivirus supernatants poolled or individually to determine which gave the best knock-down of WHSC1L1. Non-silencing shRNAmir lentiviral control, at the same titer as WHSC1L1 shRNA, was used in parallel as the negative control. First, the consequence of knock-down of WHSC1L1 using all eight shRNAs combined, on colony formation was evaluated in all three cell lines. WHSC1L1 knock-down suppressed proliferation of SUM-44 and SUM-52 cells, while WHSC1L1 shRNAs had no effect on the growth of MCF10A cells (Supplementary Figure S4) . Next, we identified the two most efficient shRNAs with respect to knock-down of WHSC1L1 expression levels in SUM-44 and SUM-52 cells. Q-RT-PCR and western blot data revealed that the WHSC1L1-shRNAs #2 and #6 resulted in decreases in mRNA and protein levels to approximately 20-30% of the level seen in the non-silencing control-infected cells (Figure 4A). As shown in Figure 4B and C , WHSC1L1 knock-down with both shRNA constructs slowed cell growth of SUM-44 and SUM-52 cells. The results were most striking for SUM-44 cells in which WHSC1L1 knock-down inhibited cell proliferation by ~ 90% (Figure 4C). WHSC1L1 knock-down with these shRNA#2 and #6 had an undetectable effect on the cell growth of MCF10A cells (data no shown). Thus, knockdown of WHSC1L1 inhibits cell proliferation in breast cancer cells with WHSC1L1 gene amplification.

IRX3 is a novel target gene of WHSC1L1

Because WHSC1L1 encodes a PWWP domain nuclear protein that has histone methyl transferase activity, it has been postulated that it can promote malignant transformation by altering the histone code and hence expression of specific target genes.

To identify genes that may be altered in their expression by over expression of the short isoform of WHSC1L1, we performed expression profiling of MCF10A cells, MCF10A-WHSC1L1 cells, and SUM-44 cells. To identify genes most likely to be regulated by over expression of WHSC1L1 and relevant to human breast cancer, we determined which genes are differentially expressed in MCF10A-WHSC1L1 cells relative to parental MCF10A cells, and then determined which of those genes are also differentially expressed in SUM-44 cells compared to MCF10A cells. This orthogonal analysis resulted in the identification of 184 genes differentially expressed in both SUM-44 cells and MCF10A-WHSC1L1 cells, relative to MCF10A cells. Of the 184 differentially expressed genes, 36 are coordinately up-regulated in MCF10A-WHSC1L1 cells and SUM-44 cells. Figure 5A shows the four up-regulated genes (TBL1X, IRX3, RAG1AP1 and RAPGEF3), and two down-regulated genes (TFBI and SFRP1) in both SUM-44 cells and MCF10A-WHSC1L1 cells, relative to MCF10A cells. Thus, these genes could be directly altered in their expression by over expression of WHSC1L1-short isoform. To directly validate some of these array-based observations, we chose three up-regulated genes and one down-regulated gene to examine by Q-RT-PCR (Figure 5B). Figure 5 shows that IRX3, RAPGEF3, and TBLX1 are significantly over expressed at the mRNA level in SUM-44 cells compared to MCF10A cells. Furthermore, knock-down of WHSC1L1 in SUM-44 cells using the shRNA constructs described previously resulted in significant down regulation of these three putative target genes. These results support the array-based analysis and indicate that WHSC1L1 regulates the expression of these target genes. Similarly, examination of one down-regulated gene, TGFBI also confirmed reduced expression in SUM-44 compared to MCF10A cells, and the expression of this gene was increased in SUM-44 cells bearing the WHSC1L1 shRNA constructs.

Since WHSC1L1 regulates the expression of IRX3 in SUM-44 cells and MCF10A-WHSC1L1 cells, we examined the genomic state of this potentially important target gene. Interestingly, we found that SUM-44 cells have an amplification at the IRX3 locus of chromosome 16q12. In addition, SUM-225 cells, which also have the 8p11-12 amplicon have an overlapping region of copy number increase in chromosome 16 (Supplementary Figure S5). Thus, these findings indicate that in SUM-44 cells, over

expression of an amplified oncogene on chromosome 8p11-12 drives the expression of another amplified gene on a different chromosome. This genetic interaction explains the very high level of expression of IRX3 in SUM-44 cells compared to MCF10A-WHSC1L1 cells. This result indicates that not only is WHSC1L1 a major driving oncogene in the 8p11-12 amplicon, it is genetically linked to IRX3 in a subset of breast cancer.

With respect to down regulated genes, the finding that over expression of WHSC1L1 resulted in down regulation of the negative regulator of WNT signaling, SFRP1 in MCF10A cells is intriguing. We have recently demonstrated that even though the SFRP1 gene is part of the 8p11-12 amplicon and is increased in copy number in SUM-44 cells, it is highly methylated and not expressed in these cells (28). Another down regulated gene is TGFBI, that encodes a secreted protein induced by transforming growth factor-β (Figure 5). Recent studies with TGFBI-null mice demonstrated that TGFBI loss promotes cell proliferation and predisposes mice to spontaneous tumor development (29). Thus, PWWP-protein WHSC1L1 may regulate a subset of genes involving in various functional pathways in breast cancer.

Discussion

The 8p11-12 amplicon has been the subject of a number of studies using high-resolution genomic analysis of copy number and gene expression in HBC (7-11, 30). Our first studies in this area demonstrated that the 8p11-12 amplicon has a complex genomic structure and the size of the amplicon is variable in three HBC lines, SUM-44, SUM-52 and SUM-225 (9, 31). In this work we showed that *FGFR1* was only one of several candidate oncogenes in the amplicon, and we provided evidence that FGFR1 is not the driving oncogene in every breast cancer with the 8p11 amplicon (31). In addition, our correlative evidence suggested that other genes in the region, including *LSM-1*, *C8orf4* (*TC-1*), *RAB11FIP1*, *WHSC1L1* and *ERLIN2* were good candidate oncogenes based on their over expression associated with gene amplification (10). Our findings are consistent with those of other laboratories. Gelsi-Boyer *et al*. performed a

comprehensive study combining genomic, expression, and chromosome break analyses of the 8p11-12 region in 37 HBC lines and 134 primary breast cancer specimens. (8). They identified four overlapping amplicon cores at 8p11-12 and 14 candidate oncogenes that are significantly over expressed in relation to amplification. Based on elegant and rigorous statistical analysis, they provided evidence that several genes within each core were over expressed as clusters because of their genomic proximity. In core A1, five over expressed genes correlated with amplification including *ZNF703*, *ERLIN2*, *PROSC*, *BRF2* and *RAB11FIP1*. In core A2, six over expressed genes correlated with amplification including; *LSM1*, *DDHD2*, *PPAPDC1B*, *WHSC1L1*, *FGFR1*, and *TM2D2*. Core A3 did not contain any correlated genes, but core A4 included *AP3M2* and *IKBKB*. In subsequent work, Bernard-Pierrot *et al*. carried out BAC-array CGH on 21 HBC lines and 152 ductal breast carcinomas and identified five genes (*LSM1*, *BAG4*, *DDHD2*, *PPAPDC1B* and *WHSC1L1*) within the 8p11-12 amplified region as consistently over expressed due to an increased gene copy number. Finally, Chin *et al*. published an analysis of the association of 8p11-12 gene amplification and disease-free survival and distant relapse in human breast cancer specimens and identified 23 genes from the 8p11-12 region as being correlated with progression, all of which have been named already (12). Thus, several groups have performed extensive analyses of the 8p11-12 genomic region in human breast cancer and there is substantial agreement on the candidate oncogenes present in this region. The candidate oncogenes consistently identified by all groups include, *FGFR1*, *WHSC1L1*, *RAB11FIP1*, *LSM1*, *BAG4* and *ERLIN2*.

In addition to the correlative studies, there are now several papers in the literature reporting experimental analysis of the transforming function of the candidate oncogenes from this region. We reported that *BAG4*, *LSM1*, and *C8orf4* (*TC-1*) are transforming when over expressed in MCF10A cells (10, 14, 15). In the present report, we provide evidence that three additional genes *WHSC1L*1, *ERLIN2* and *DDHD2* are transforming based on their ability to induce growth factor-independent proliferation, anchorage independent growth, and altered morphogenesis in Matrigel culture. As reported in this paper, we find *WHSC1L1* to be the most potently transforming of all the 8p11 oncogenes we have tested. It is noteworthy that FGFR1 consistently fails to score in our

transformation assays. Our results with FGFR1 are in contrast to those reported recently by Xian *et al*. who reported transforming activity of FGFR1 in MCF-10A cells (32). However, in their study, MCF-10A cells were transfected with an artificial construct containing the FGFR1 tyrosine kinase domain fused with the F36V mutant FKBP12, which allows for high level receptor dimerization in the presence of the bivalent compound AP20187. This discrepancy is significant because we have previously shown that FGFR2 is a driving oncogene in SUM-52 cells (chromosome 10q26 amplification) which results in over expression of highly transforming isoforms of FGFR2. Transduction of these naturally occurring FGFR2 isoforms into immortalized human mammary epithelial cells induces transformation in a ligand independent manner (33).

Our results are consistent with those reported earlier by Bernard-Pierrot *et al*. who performed RNAi experiments to knock-down the expression of candidate genes, in two cell lines (CAMA-1 and ZR-75-1) with 8p11-12 amplification. Their siRNA results suggest that *PPAPDC1B*, and *WHSC1L1* are two driving oncogenes from this amplicon. Further, microarray experiments on *PPAPDC1B* knockdown indicated that this gene interfered with multiple cell signaling pathways, including the Janus-activated kinase, signal transducer and activator of transcription, mitogen-activated protein kinase, and protein kinase C pathways (34). Zhang and colleagues applied a novel algorithm, termed TRIAGE (triangulating oncogenes through clinico-genomic intersects), to a collection of microarray expression profiles of primary human breast cancers in an effort to identify candidate genes in amplicons that could contribute to patient outcome (35). They identified RAP11FIP1 and also identified WHSC1L1 as being strongly associated with breast cancer subtype and outcome. They selected RAP11FIP1 for further transfection and knockdown studies and found that RAP11FIP1 is not sufficient to transform naive cells. However, in breast cancer cell lines, over expression of RAP11FIP1 decreased growth factor dependence, increased survival under anoikis conditions, induced motility, invasion, an epithelial-mesenchymal transition *in vitro*, and increased tumor growth *in vivo (35)*. In other studies, Luscher-Firzlaff *et a*l. reported that ASH2L encodes the trithorax protein and cooperates with H-Ras to transform primary rat embryo fibroblasts (36).

Based on the findings from several laboratories, *WHSC1L1* is clearly emerging as an important transforming gene within the 8p11-12 amplicon in breast and other cancer types. *WHSC1L1* is involved in a chromosomal translocation in acute myeloid leukemia, t(8;11)(p11.2;p15) (37). Amplification and over expression of WHSC1L1 has also been found in lung cancer (10, 17, 38). Integrated DNA-RNA analyses identified WHSC1L1 as the most likely candidate target for the 8p11-12 amplicon in lung cancer (38). siRNA-mediated knockdown of WHSC1L1 resulted in a 50% reduction in the number of soft agar colonies in a lung cancer cell line (H1703) with *WHSC1L1* gene amplification and over expression (38). Furthermore, deep sequencing of a primary human breast cancer identified a deletion within the *WHSC1L1* gene (39).

A finding of particular interest from our study is that *IRX3*, a member of the homeobox gene family, and *TBLX1* are target genes of WHSC1L1. Interestingly, *IRX3* is also amplified in SUM-44 cells and in SUM-225 cells. This is of interest because in embryonic stem cells IRX3 and TBLX1 are linked in a gene expression network that regulates WNT signaling (40). In addition, we have previously shown that in breast cancers with the 8p11 amplicon, SFRP1, a negative regulator of WNT signaling, is silenced by promoter methylation, despite being present on the 8p11 amplicon and increased in copy number (28). These results suggest that the over expression of WHSC1L1 and the silencing of SFRP1 result in potent activation of a transcriptional network linked to WNT signaling and expression of stem cell phenotypes.

Despite a significant level of work to understand the basis for selection of the 8p11-12 amplicon in ER+, luminal B type breast cancers, there continues to be uncertainty regarding the role of FGFR1 amplification in these breast cancers. While we and others have failed to find compelling evidence for FGFR1 as a dominant driving oncogene from this region, Turner *et al*. recently published a paper re-affirming it's role in 8p11-12 amplified breast cancers (41). Many of the results reported by Turner *et al*. are consistent with our previously published results, which show that in most breast cancer cell lines with 8p11-12 amplification and FGFR1 over expression, knockdown of

FGFR1 or inhibition of FGFR kinase activity has little or no effect on the growth and survival of the cells. However, one cell line, MDA-134 does appear to be sensitive to FGFR1 knock-down or inhibition, suggesting that this gene could be a driving oncogene in some breast cancers with the amplicon. Of greater interest, however, is their finding that over expression of FGFR1 increases the sensitivity and responsiveness of cells to FGF ligands, and this influences the response of the cells to 4-OH Tamoxifen. That this occurs even in SUM-44 cells that do not require FGFR1 expression or activation for growth in culture suggests that genes that are over expressed as part of an amplicon need not always be driving oncogenes in order to be important in tumor progression. In this regard, we have long considered the 8p11-12 amplicon to be an oncogenic unit based on the consistent co-expression of many candidate oncogenes when the amplicon is present in ER+, luminal B type breast cancers. Within this oncogenic unit are two genes that can regulate the histone code (*WHSC1L1*, *ASH2L*), a gene that regulates RNA metabolism (*LSM1*), a receptor tyrosine kinase (*FGFR1*), a gene that regulates endoplasmic reticulum stress pathway (ERLIN2), and a gene that influences receptor trafficking (RAB11FIP1). The remarkably consistent coordinate over expression of these genes in breast cancers with the amplicon predicts that expression of any one of them would correlate with survival parameters, as FGFR1 does, and this was reported in the Chin paper. Thus, one must be cautions when using data to correlate over expression of any one gene from this amplicon with survival parameters.

The 8p11-12 amplicon is now known to play an important role in the development of ER+, luminal B type breast cancers. Given that eventual acquisition of hormone independence is a critical clinical problem for this disease subset, understanding how the genes in this region influence the survival of these breast cancer cells offers new possible strategies for attacking breast cancers that have become resistant to anti-estrogen therapies.

Figure Legends

Table 1. List of the 21 candidate genes of 8p11-12 region

Figure 1. (A) MCF10A cells expressing 6 genes WHSC1L1, BRF2, DDHD2, PROSC, ERLIN2 and ASH2L formed expanding colonies in insulin-free medium, while MCF10A cells expressing HOOK3 and control LacZ did not form colonies. (B) *In vitro* growth rate of the MCF10A cells that stably over express the 5 genes (ERLIN2, WHSC1L1, DDHD2, PROSC and BRF2) relative to MCF10A-LacZ control cells in insulin deficient media. Cells were seeded into 35-mm culture wells and grown in the absence of insulin-like growth factors.

Figure 2. (A) Top: Representative pictures of MCF10A cells that stably over express DDHD2 and WHSC1L1 genes and control cell soft-agar colonies. Cells were grown for 3 weeks in soft agar and stained with the vital dye p-iodonitrotetrazolium violet. Bottom: Effects of DDHD2 and WHSC1L1 on mammary acinar morphogenesis. MCF10A-DDHD2, WHSC1L1 and control cells were cultured on a bed of Matrigel as described in Materials and Methods. Representative images of structures with staining for actin with phalloidin conjugated to Alexa Fluor-568 (red), and DAPI as a marker of nuclei (blue). (B) Soft agar colony forming efficiency in MCF10A cells stably over expressing the 6 genes (WHSC1L1, BRF2, DDHD2, PROSC, ERLIN2 and ASH2L) and control cell soft-agar colonies. Data represent the mean number of colonies per low power field three weeks after seeding 10^5 cells per well.

Figure 3. (A) Genomic copy number profiles of the WHSC1L1 region analyzed on the Agilent oligonucleotide array CGH in two SUM breast cancer cell lines (SUM-44 and SUM-52) and one primary breast cancer specimen (10173A). Array probes and genes are displayed horizontally by genome position. Log2 ratio in each sample is relative to normal female DNA. S represents WHSC1L1 short isoform and L represents long isoform. (B) WHSC1l1 protein levels were analyzed by western blot in two breast cancer

cell lines, SUM-44, SUM-52 and control MCF10A line. (C) genomic PCR using primers specific for the short isoform exon 10 (S-10) and the long isoform exon 20 (L-20) of WHSC1L1 were used to validate array-CGH data in breast cancer specimen 10173A.

Figure 4. (A) WHSC1L1 expression in SUM-44 cells was analyzed by semiquantitative RT-PCR and western blot after infection with non-silencing control shRNA or WHSC1L1 specific shRNA (shRNA#2 and #6). (B). The images showed the TurboGFP fluorescence of pGIPZ-WHSC1L1 shRNAs in SUM-44 cells after 3 weeks. (C) shRNA-mediated knock-down of WHSC1L1 inhibits cell growth in breast cancer cells SUM-44 and SUM-52 with WHSC1L1 amplification.

Figure 5. (A) Six genes differentially expressed in both SUM-44 cells and MCF10A-WHSC1L1 cells, relative to MCF10A cells with the Illumine expression Beadarray. (B) TBL1X , IRX3, RAPGEF3 and TGFB1 expression level was measured by semiquantitative RT-PCR in MCF10A, SUM-44 cell (left panels), and WHSC1L1 knock-down SUM-44 cells (right panels).

Achnowledgements

This work was supported by a grant from the Department of Defense Breast Cancer Program (BC083945) to Zeng-Quan Yang and a grant from the National Institutes of Health (RO1 CA100724) to Stephen P. Ethier. We thank Michele Dziubinski and Katie L. Streicher1 for technical assistance on the cell culture and three-dimensional morphogenesis assays. The array CGH work was facilitated by the Microarray and Bioinformatics Core Facility of the Wayne State University Environmental Health Sciences Center, NIEHS P30 ES06639.

References

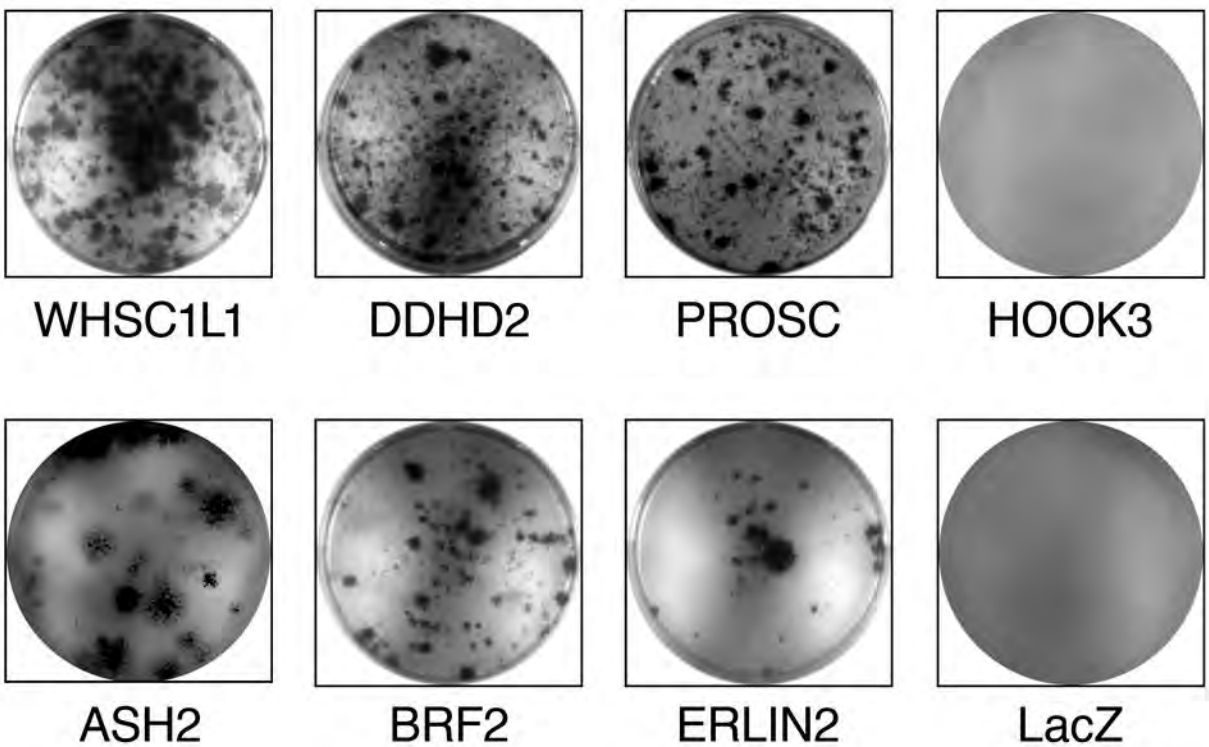
1. Albertson DG. Gene amplification in cancer. Trends Genet 2006; 22: 447-55.
2. Albertson DG, Collins C, McCormick F, Gray JW. Chromosome aberrations in solid tumors. Nature genetics 2003; 34: 369-76.
3. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. Nature medicine 2004; 10: 789-99.
4. Moasser MM. The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. Oncogene 2007.
5. Arteaga CL, Moulder SL, Yakes FM. HER (erbB) tyrosine kinase inhibitors in the treatment of breast cancer. Seminars in oncology 2002; 29: 4-10.
6. Badache A, Goncalves A. The ErbB2 signaling network as a target for breast cancer therapy. J Mammary Gland Biol Neoplasia 2006; 11: 13-25.
7. Garcia MJ, Pole JC, Chin SF, et al. A 1 Mb minimal amplicon at 8p11-12 in breast cancer identifies new candidate oncogenes. Oncogene 2005; 24: 5235-45.
8. Gelsi-Boyer V, Orsetti B, Cervera N, et al. Comprehensive profiling of 8p11-12 amplification in breast cancer. Mol Cancer Res 2005; 3: 655-67.
9. Yang ZQ, Albertson D, Ethier SP. Genomic organization of the 8p11-p12 amplicon in three breast cancer cell lines. Cancer genetics and cytogenetics 2004; 155: 57-62.
10. Yang ZQ, Streicher KL, Ray ME, Abrams J, Ethier SP. Multiple interacting oncogenes on the 8p11-p12 amplicon in human breast cancer. Cancer research 2006; 66: 11632-43.
11. Pole JC, Courtay-Cahen C, Garcia MJ, et al. High-resolution analysis of chromosome rearrangements on 8p in breast, colon and pancreatic cancer reveals a complex pattern of loss, gain and translocation. Oncogene 2006; 25: 5693-706.
12. Chin K, DeVries S, Fridlyand J, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 2006; 10: 529-41.
13. Yang ZQ, Streicher KL, Ray ME, Abrams J, Ethier SP. Multiple interacting oncogenes on the 8p11-p12 amplicon in human breast cancer. Cancer Res 2006; 66: 11632-43.
14. Yang ZQ, Moffa AB, Haddad R, Streicher KL, Ethier SP. Transforming properties of TC-1 in human breast cancer: interaction with FGFR2 and beta-catenin signaling pathways. International journal of cancer 2007; 121: 1265-73.
15. Streicher KL, Yang ZQ, Draghici S, Ethier SP. Transforming function of the LSM1 oncogene in human breast cancers with the 8p11-12 amplicon. Oncogene 2007; 26: 2104-14.
16. Stec I, van Ommen GJ, den Dunnen JT. WHSC1L1, on human chromosome 8p11.2, closely resembles WHSC1 and maps to a duplicated region shared with 4p16.3. Genomics 2001; 76: 5-8.
17. Angrand PO, Apiou F, Stewart AF, 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.
18. Bergemann AD, Cole F, Hirschhorn K. The etiology of Wolf-Hirschhorn syndrome. Trends Genet 2005; 21: 188-95.

19. Rahman N. Mechanisms predisposing to childhood overgrowth and cancer. Curr Opin Genet Dev 2005; 15: 227-33.
20. Ruthenburg AJ, Allis CD, Wysocka J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Molecular cell 2007; 25: 15-30.
21. Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol 2007; 14: 1025-40.
22. Chen T, Tsujimoto N, Li E. The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Mol Cell Biol 2004; 24: 9048-58.
23. Wang Y, Reddy B, Thompson J, et al. Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. Molecular cell 2009; 33: 428-37.
24. Forozan F, Veldman R, Ammerman CA, et al. Molecular cytogenetic analysis of 11 new breast cancer cell lines. British journal of cancer 1999; 81: 1328-34.
25. Yang ZQ, Imoto I, Fukuda Y, et al. Identification of a novel gene, GASC1, within an amplicon at 9p23-24 frequently detected in esophageal cancer cell lines. Cancer research 2000; 60: 4735-9.
26. Debnath J, Muthuswamy SK, Brugge JS. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. Methods (San Diego, Calif 2003; 30: 256-68.
27. Debnath J, Brugge JS. Modelling glandular epithelial cancers in three-dimensional cultures. Nature reviews 2005; 5: 675-88.
28. Yang ZQ, Liu G, Bollig-Fischer A, Haddad R, Tarca AL, Ethier SP. Methylation-associated silencing of SFRP1 with an 8p11-12 amplification inhibits canonical and non-canonical WNT pathways in breast cancers. International journal of cancer 2009; 125: 1613-21.
29. Zhang Y, Wen G, Shao G, et al. TGFBI deficiency predisposes mice to spontaneous tumor development. Cancer research 2009; 69: 37-44.
30. Haverty PM, Fridlyand J, Li L, et al. High-resolution genomic and expression analyses of copy number alterations in breast tumors. Genes Chromosomes Cancer 2008; 47: 530-42.
31. Ray ME, Yang ZQ, Albertson D, et al. Genomic and expression analysis of the 8p11-12 amplicon in human breast cancer cell lines. Cancer research 2004; 64: 40-7.
32. Xian W, Pappas L, Pandya D, et al. Fibroblast growth factor receptor 1-transformed mammary epithelial cells are dependent on RSK activity for growth and survival. Cancer research 2009; 69: 2244-51.
33. Moffa AB, Tannheimer SL, Ethier SP. Transforming potential of alternatively spliced variants of fibroblast growth factor receptor 2 in human mammary epithelial cells. Mol Cancer Res 2004; 2: 643-52.
34. Bernard-Pierrot I, Gruel N, Stransky N, et al. Characterization of the recurrent 8p11-12 amplicon identifies PPAPDC1B, a phosphatase protein, as a new therapeutic target in breast cancer. Cancer research 2008; 68: 7165-75.
35. Zhang J, Liu X, Datta A, et al. RCP is a human breast cancer-promoting gene with Ras-activating function. The Journal of clinical investigation 2009; 119: 2171-83.
36. Luscher-Firzlaff J, Gawlista I, Vervoorts J, et al. The human trithorax protein hASH2 functions as an oncoprotein. Cancer research 2008; 68: 749-58.

37. Rosati R, La Starza R, Veronese A, et al. NUP98 is fused to the NSD3 gene in acute myeloid leukemia associated with t(8;11)(p11.2;p15). Blood 2002; 99: 3857-60.
38. Tonon G, Wong KK, Maulik G, et al. High-resolution genomic profiles of human lung cancer. Proceedings of the National Academy of Sciences of the United States of America 2005; 102: 9625-30.
39. Guffanti A, Iacono M, Pelucchi P, et al. A transcriptional sketch of a primary human breast cancer by 454 deep sequencing. BMC genomics 2009; 10: 163.
40. Ferrell CM, Dorsam ST, Ohta H, et al. Activation of stem-cell specific genes by HOXA9 and HOXA10 homeodomain proteins in CD34+ human cord blood cells. Stem cells (Dayton, Ohio) 2005; 23: 644-55.
41. Turner N, Pearson A, Sharpe R, et al. FGFR1 Amplification Drives Endocrine Therapy Resistance and Is a Therapeutic Target in Breast Cancer. Cancer research; 70: 2085-94.

FIGURE 1

A


WHSC1L1
DDHD2
PROSC
HOOK3
ASH2
BRF2
ERLIN2
LacZ

B

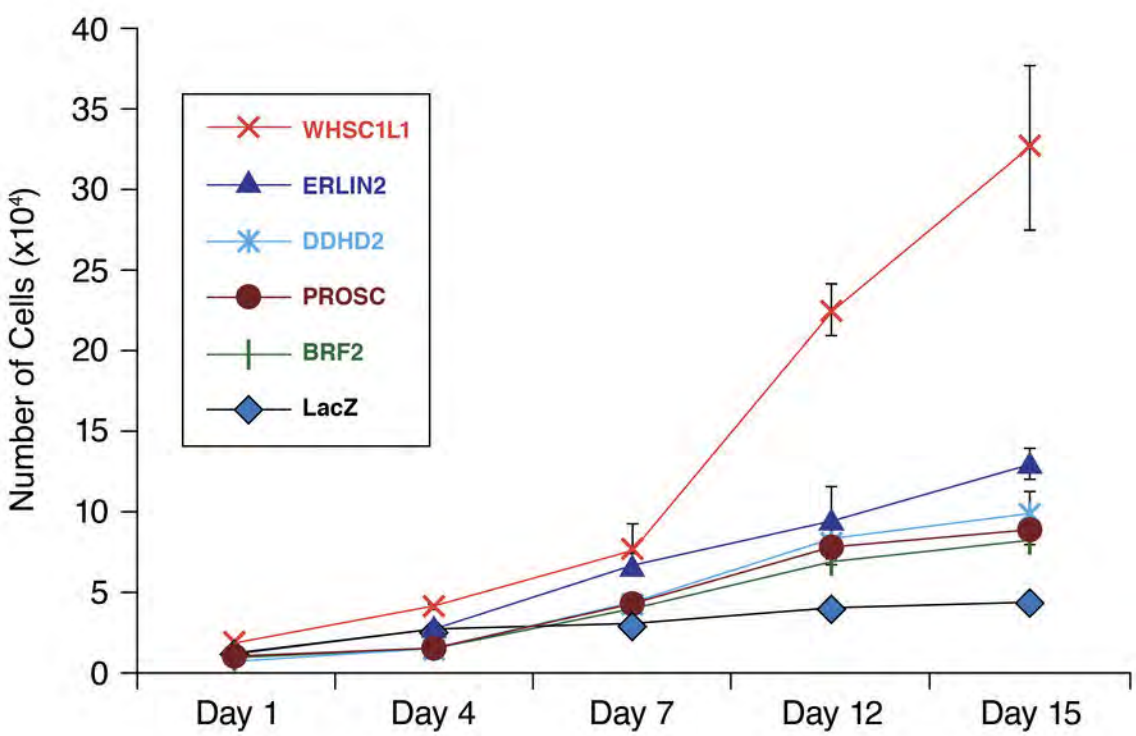
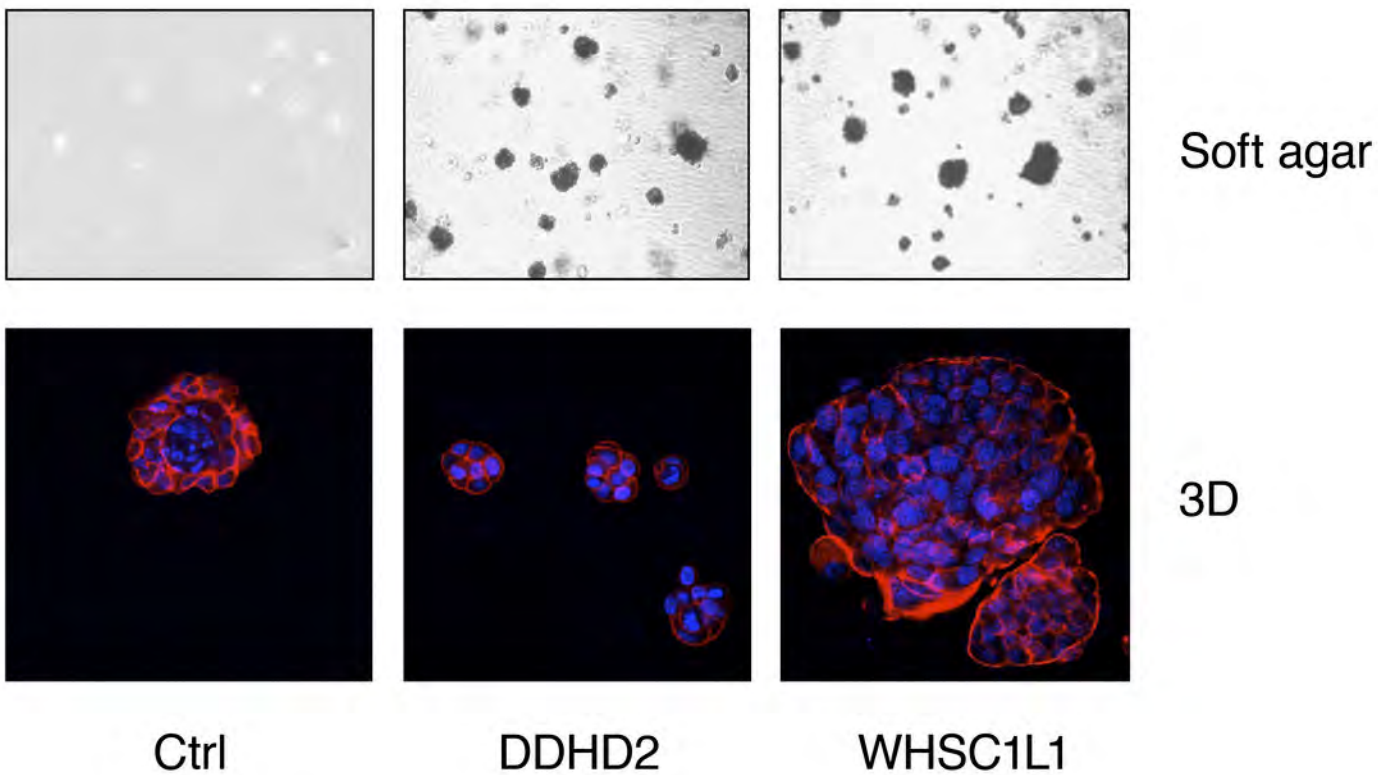

Number of Cells (x10⁴)
40
35
30
25
20
15
10
5
0
WHSC1L1
ERLIN2
DDHD2
PROSC
BRF2
LacZ
Day 1
Day 4
Day 7
Day 12
Day 15

FIGURE 2

A


Soft agar
3D
Ctrl
DDHD2
WHSC1L1

B

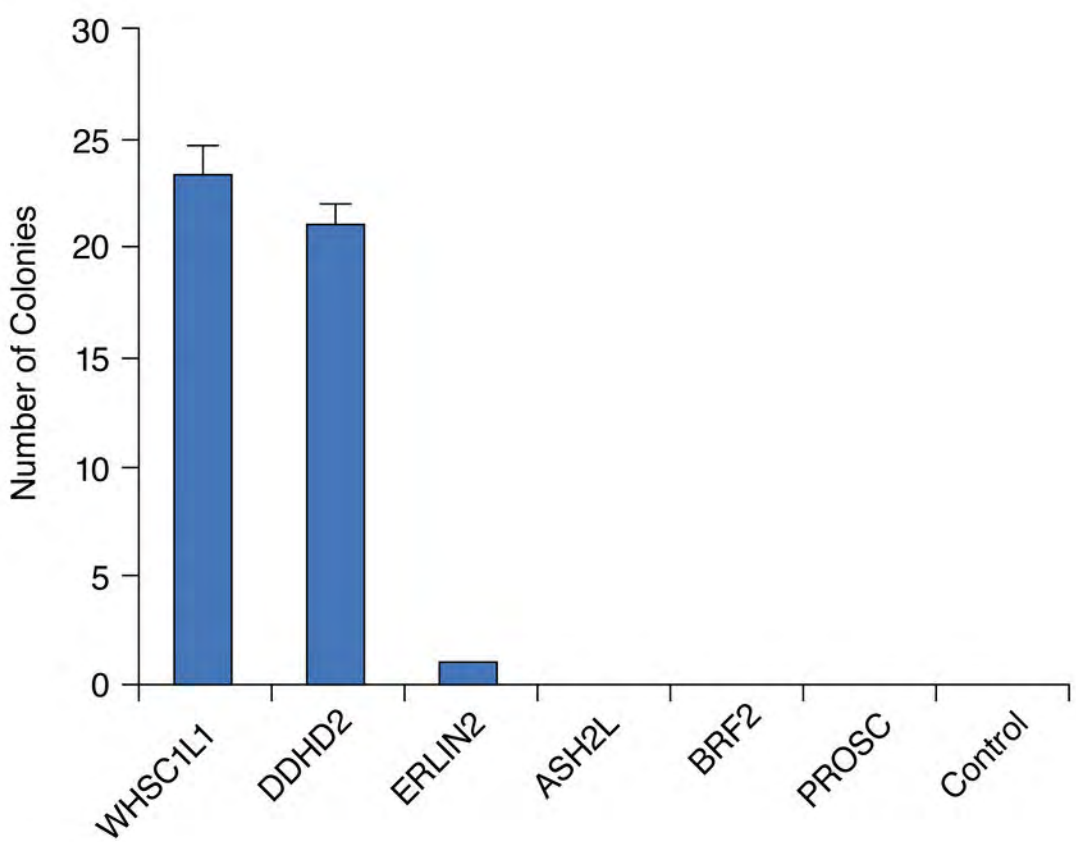
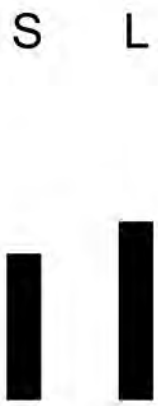

Number of Colonies
30
25
20
15
10
5
0
WHSC1L1
DDHD2
ERLIN2
ASH2L
BRF2
PROSC
Control

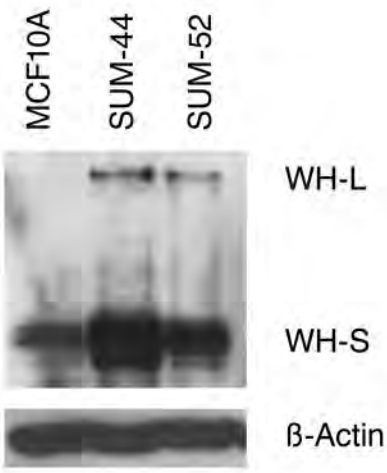
FIGURE 3

A

Probe	Gene	SUM-44	SUM-52	10173A
P134341	HTPAP	2.74	1.26	-0.94
P138390	HTPAP	2.91	1.23	-0.84
P133716	HTPAP	2.69	1.17	-0.73
P126567	WHSC1L1	3.05	1.30	-0.83
P138155	WHSC1L1	2.56	0.88	1.46
P107950	WHSC1L1	2.84	1.00	1.53
P139831	WHSC1L1	2.86	1.12	1.94
P109924	LETM2	2.92	1.55	1.84
P131121	LETM2	3.03	1.56	2.31
P139342	FGFR1	3.12	1.66	1.92
P117830	FGFR1	2.02	1.09	1.71



B



C

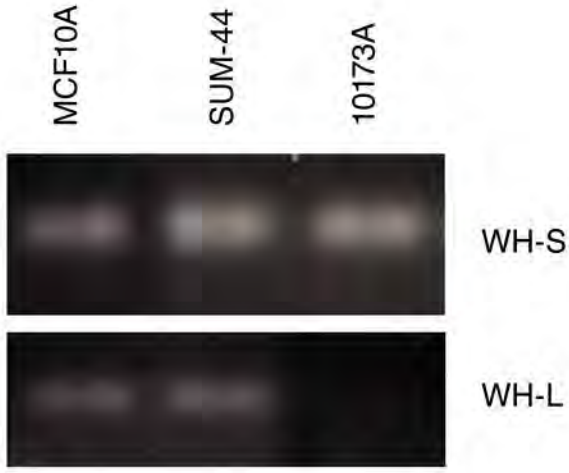


FIGURE 4

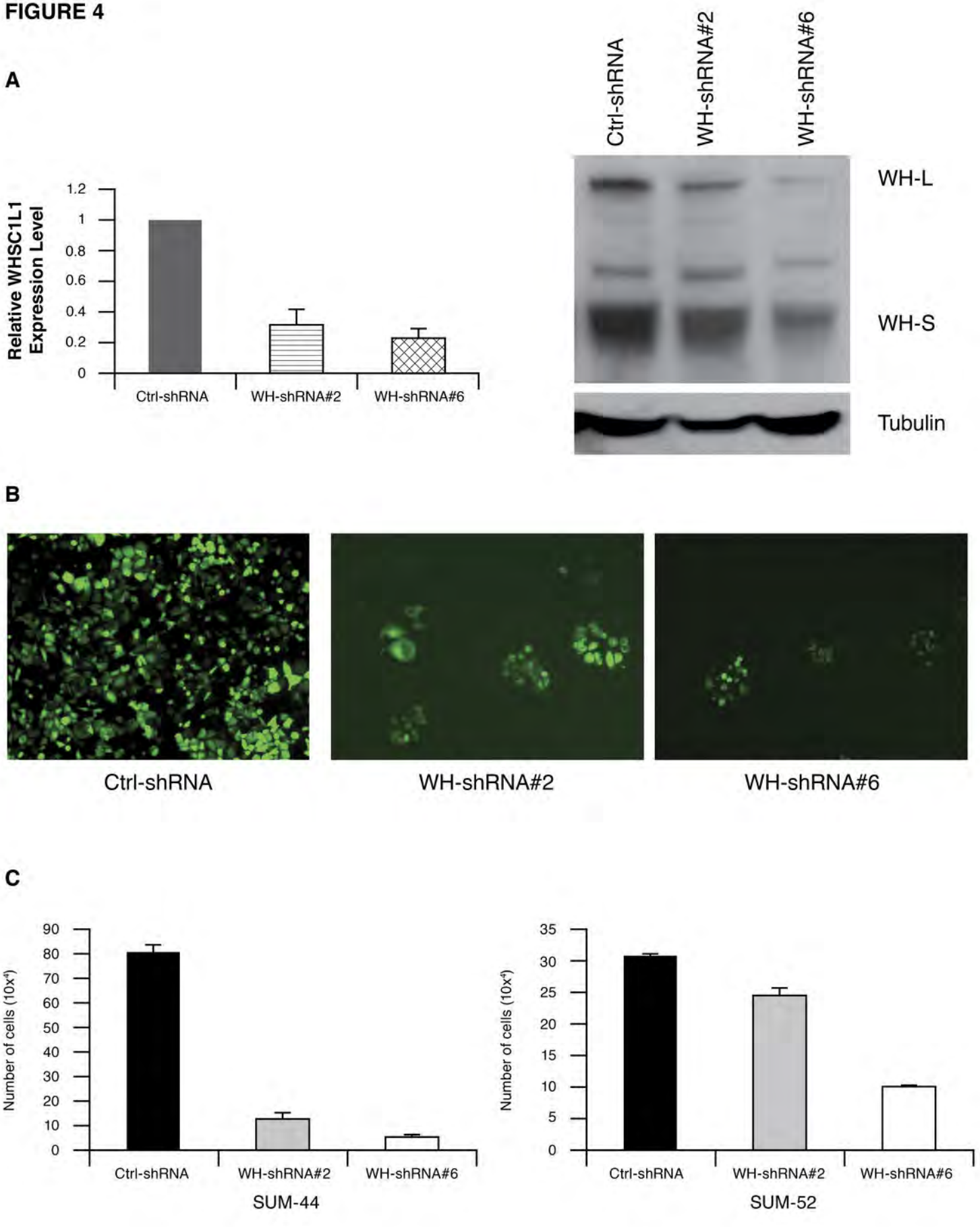

A
Ctrl-shRNA
WH-shRNA#2
WH-shRNA#6
Relative WHSC1L1 Expression Level
1.2
1
0.8
0.6
0.4
0.2
0
Ctrl-shRNA
WH-shRNA#2
WH-shRNA#6
WH-L
WH-S
Tubulin
B
Ctrl-shRNA
WH-shRNA#2
WH-shRNA#6
C
Number of cells (10x⁴)
90
80
70
60
50
40
30
20
10
0
Ctrl-shRNA
WH-shRNA#2
WH-shRNA#6
SUM-44
Number of cells (10x⁴)
35
30
25
20
15
10
5
0
Ctrl-shRNA
WH-shRNA#2
WH-shRNA#6
SUM-52

FIGURE 5

A

Gene	SUM-44	MCF10A-WH
TBL1X	15.15	3.63
IRX3	10.58	2.27
RAG1AP1	5.35	2.03
RAPGEF3	2.54	2.71
TGFBI	-119.93	-3.34
SFRP1	-40.52	-3.34

B

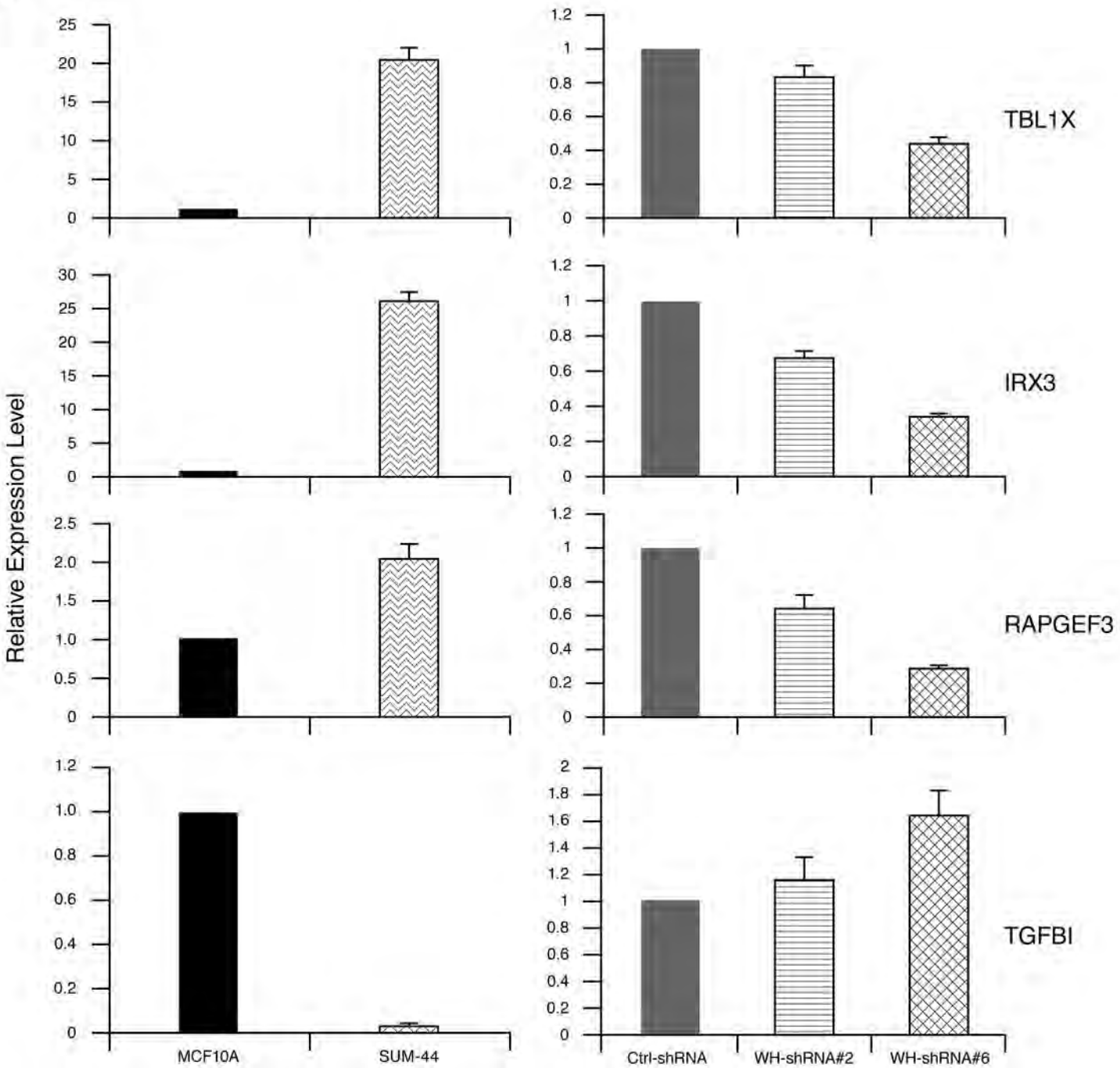


Table 1

	Gene	Description
	ZNF703	zinc finger protein 703
*	*ERLIN2*	ER lipid raft associated 2
*	*PROSC*	proline synthetase co-transcribed homolog (bacterial)
*	*BRF2*	BRF2, subunit of RNA polymerase III transcription initiation factor, BRF1-like
	RAB11FIP1	Rab coupling protein = RCP
	EIF4EBP1	elongation factor 4 binding protein 1
*	*ASH2L*	ash2 (absent, small, or homeotic)-like (Drosophila)
	LSM1	LSM1 homolog, U6 small nuclear RNA associated (S. cerevisiae)
	BAG4	BCL2-associated athanogene 4
*	*DDHD2*	DDHD domain containing 2
*	*PPAPDC1B*	phosphatidic acid phosphatase type 2 domain containing 1B
*	*WHSC1L1*	Wolf-Hirschhorn syndrome candidate 1-like 1
*	*LETM2*	leucine zipper-EF-hand containing transmembrane protein 2
	FGFR1	fibroblast growth factor receptor 1 (fms-related tyrosine kinase 2, Pfeiffer syndrome)
	TACC1	Transforming acidic coiled-coil
*	*TM2D2*	TM2 domain containing 2
	C8orf4	Chromosome 8 open reading frame 4 = TC-1
*	*AP3M2*	adaptor-related protein complex 3, mu 2 subunit
*	*POLB*	polymerase (DNA directed), beta
*	*VDAC3*	voltage-dependent anion channel 3
*	*HOOK3*	hook homolog 3 (Drosophila)