Supporting your research with our capabilities

BD Accuri™ C6 Plus Personal Flow Cytometer

BD FACSCelesta™ Cell Analyzer

BD LSRFortessa™ X-20 Cell Analyzer

BD FACSMelody™ Cell Sorter

One of the largest portfolios of reagents

Learn more>

BD

IJC
International Journal of Cancer

Elevated expression of the centromere protein-A(CENP-A)-encoding gene as a prognostic and predictive biomarker in human cancers

Xia Sun[1], Pier-Luc Clermont[2,3,4], Wenlin Jiao[1], Cheryl D. Helgason[2,3,5], Peter W. Gout[2], Yuzhuo Wang[2,3,6,7] and Sifeng Qu[2,3,6]

[1] Department of Pharmacology, School of Medicine, Shandong University, Jinan, Shandong, China
[2] Department of Experimental Therapeutics, BC Cancer Research Centre, Vancouver, BC, Canada
[3] Interdisciplinary Oncology Program, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada
[4] Faculty of Medicine, Université Laval, Québec, QC, Canada
[5] Department of Surgery, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada
[6] Vancouver Prostate Center, Vancouver, BC, Canada
[7] Department of Urologic Sciences, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

Centromere protein-A (CENP-A), a histone-H3 variant, plays an essential role in cell division by ensuring proper formation and function of centromeres and kinetochores. Elevated CENP-A expression has been associated with cancer development. This study aimed to establish whether elevated *CENP-A* expression can be used as a prognostic and predictive cancer biomarker. Molecular profiling of *CENP-A* in human cancers was investigated using genomic, transcriptomic and patient information from databases, including COSMIC, Oncomine, Kaplan–Meier plotter and cBioPortal. A network of *CENP-A* co-expressed genes was derived from cBioPortal and analyzed using Ingenuity Pathway Analysis (IPA) and Oncomine protocols to explore the function of CENP-A and its predictive potential. Transcriptional and post-transcriptional regulation of *CENP-A* expression was analyzed *in silico*. It was found that *CENP-A* expression was elevated in 20 types of solid cancer compared with normal counterparts. Elevated *CENP-A* expression highly correlated with cancer progression and poor patient outcome. Genomic analysis indicated that the elevated *CENP-A* expression was not due to alterations in the sequence or copy number of the CENP-A gene. Furthermore, *CENP-A* can be regulated by key oncogenic proteins and tumor-suppressive microRNAs. *CENP-A* co-expression network analysis indicated that CENP-A function is associated with cell cycle progression. Oncomine analysis showed a strong correlation between elevated *CENP-A* expression and oncolytic response of breast cancer patients to taxane-based chemotherapy. In conclusion, elevated *CENP-A* expression is coupled to malignant progression of numerous types of cancer. It may be useful as a biomarker of poor patient prognosis and as a predictive biomarker for taxane-based chemotherapy.

Introduction

Mitosis is characterized by segregation of chromatids to opposite poles of the dividing cell. This complex process involves binding of microtubule fibers to kinetochores (protein structures located on centromeres of the chromosomes) and subsequent shortening of the fibers resulting in the separation of the chromosomes. The presence of a single, functional and properly located centromere on each chromosome is essential for

Key words: centromere protein-A, CENP-A, co-expression network, prognostic biomarker, predictive biomarker

Abbreviations: CENP-A: centromere protein-A; CNV: copy number variation; CNS: central nervous system; ER: estrogen receptor; HER2: human epidermal growth factor receptor 2; IPA: ingenuity pathway analysis; PR: progesterone receptor; TCGA: the Cancer Genome Atlas; ; TNBC: triple negative breast cancer

Additional Supporting Information may be found in the online version of this article.

This article was published online on 21 April 2016. An error was subsequently identified. This notice is included in the online and print versions to indicate that both have been corrected 27 April 2016.

Grant sponsor: National Natural Science Foundation of China; **Grant number:** 81302770; **Grant sponsor:** Postdoctoral Science Foundation of China; **Grant number:** 2014M551917; **Grant sponsor:** Innovation Foundation of Shandong Province; **Grant number:** 201402029.

DOI: 10.1002/ijc.30133

History: Received 9 Dec 2015; Accepted 23 Mar 2016; Online 23 Mar 2016

Correspondence to: Xia Sun, Room 8308, Building 8, Department of Pharmacology, School of Medicine, Shandong University, 44 Wen Hua Xi Rd, Jinan, Shandong, China, 250012, Tel.: +86 (531) 88382605, Fax: +86 (531) 88382502, E-mail: sunxia@sdu.edu.cn or Sifeng Qu, Room 540A, Jack Bell Research Centre, 2660 Oak Street, Vancouver, BC, Canada, V6H 3Z6, Tel.: +1 604-675-8000 ext. 7013, Fax: +1 604-675-8019, E-mail: siqu@bccrc.ca

What's new?

Elevated expression of Centromere protein-A (CENP-A), a histone-H3 variant with a regulatory role in cell division, has been associated with cancer progression. Using publicly available databases, this study demonstrates that elevated *CENP-A* expression is coupled to malignant progression of numerous types of cancer. In particular, it can be used as a critically needed biomarker for (*i*) poor patient prognosis and (*ii*) predicting the response of breast cancer patients to taxane-based chemotherapy.

accurate chromosome segregation and hence normal cell division.[1,2] The proper formation, structure, function and number of the centromeres (one per chromosome) is thought to be maintained by centromere protein-A (CENP-A), a histone H3 variant which is part of the DNA–protein structure of the centromere.[3,4] CENP-A also has an essential role in establishing the assembly of the kinetochore.[3] As such, it is a key protein for maintaining accurate cell division.

Elevated CENP-A expression can lead to mislocalization of CENP-A followed by the formation of ectopic neocentromeres and kinetochores in chromosome arms.[5,6] These ectopic structures disturb the normal segregation of sister chromatids, leading to aneuploidy,[5] a phenomenon thought to predispose cells to tumor development and malignant progression.[7] A direct link has been established between elevated CENP-A levels and genomic instability, a condition that can initiate cancer and enhance disease progression.[8]

There is a critical need for prognostic and predictive biomarkers in oncology. To date, elevated CENP-A expression in breast and ovarian cancer patients has been associated with poor patient survival.[9,10] In the present study, we have investigated whether elevated expression of *CENP-A* could be used as a prognostic and predictive factor in a variety of human cancers. To this end, we used publicly available clinical information to generate a comprehensive molecular profile of *CENP-A* in human cancers at the genomic and transcriptomic levels. Our findings indicate that elevated *CENP-A* expression represents a prognostic biomarker for a large number of cancers. In addition we have found that elevated *CENP-A* expression may also be used to predict response of breast cancer patients to taxane-based chemotherapy.

Material and Methods

Catalogue of somatic mutations in cancer (COSMIC) database analysis

The COSMIC database[11] is an online resource providing information regarding gene mutations, gene fusions, genomic rearrangements, and copy number variations in human cancers. Based on this authoritative resource, we constructed a summary of alterations affecting *CENP-A*. All data were extracted on September 9, 2015 (COSMIC v74 version).

Oncomine database analysis

Oncomine[12] is a cancer transcriptomic database and web-based discovery platform with genome-wide expression analyses of 729 patient cohorts containing a total of 91,866 cancer and normal tissue specimens (Oncomine v4.5 version). Gene expression analyses for a single gene or a set of genes can be conducted across various types of cancer and include comparisons relative to normal tissues, other cancer subtypes and various clinicopathological features.

Kaplan–Meier plotter analysis

The Kaplan–Meier plotter[13,14] is a database that can be used to assess the effect of 22,277 genes on patient survival using 10,188 cancer samples (breast, ovarian, lung and gastric cancer). Data were obtained before September 9, 2015. Patients with higher and lower expression of *CENP-A* (Probe ID: 210821_x_at, Jetset best probe[15]) were segregated and analyzed using the log-rank test. The hazard ratio with 95% confidence intervals and log-rank *p* values were noted.

cBio cancer genomics portal (cBioPortal) analysis

The cBioPortal[16] provides a resource with multidimensional cancer genomics datasets for cancer research. Among these datasets, The Cancer Genome Atlas (TCGA) contains a comprehensive atlas of cancer genomic profiles. By applying high-throughput technologies based on Agilent Microarray, it serves details on the cancer genomics. With TCGA breast-invasive carcinoma microarray data published in Nature 2012,[17] we downloaded the *CENP-A* co-expression genes in this cohort.

Ingenuity pathway analysis (IPA)

IPA[18] software is used to provide functional properties of specific gene sets. Through IPA core analysis (which is used to interpret data by biological processes, pathways, upstream regulators and molecular networks in the IPA library), cellular functions and canonical pathways associated with *CENP-A* co-expressed genes can be inferred. Upstream analysis is used to identify potential upstream gene regulators based on changes in expression of given genes.

Analysis of transcription factor binding sites

The JASPAR[19] and MAPPER databases[20] are commonly used platforms for predicting potential transcription factor binding sites. Here, they were used to predict transcription factors that could bind to the CENP-A promoter.

MicroRNAs screening

The microRNA.org[21] and TargetScan databases[22] are comprehensive online resources for microRNA and target predictions. They were used to predict microRNAs with a role in *CENP-A* post-transcriptional regulation.

Table 1. Genetic alterations affecting *CENP-A* in 22,696 cancer samples (COSMIC database)

Genetic alteration	Number	Percentage (%)
Substitution nonsense	3	0.01
Substitution missense	16	0.07
Substitution synonymous	6	0.03
Copy number gain	18	0.08
Copy number loss	2	0.01
Insertion	0	0
Deletion	0	0
Translocation	0	0
Loss of heterozygosity	0	0

Statistical analysis

Unless otherwise mentioned, all analyses were conducted with $p < 0.05$ as the threshold for statistical significance. For Oncomine analysis, the Student's *t* test was used for two class differential expression analyses (*e.g.*, cancer tissues versus normal tissues) and Pearson's correlation was used for multiclass ordinal analyses (*e.g.*, Grade II and III breast cancer). The *p* values are corrected for multiple hypotheses testing using the false discovery rate method.[12] With the Kaplan–Meier plotter, the log-rank test was used to analyze the data. In the IPA analysis, right-tailed Fisher's exact test was used to determine *p* values for gene networks. The Benjamini–Hochberg (B-H) multiple-test correction method was used to correct for occurrence of false positives in IPA canonical pathways analysis[23] and, in this analysis, pathways with a B-H $p < 0.05$ were considered significant.

Results

CENP-A DNA mutations

The CENP-A gene in cancers of patients was assessed for mutations using the COSMIC database (COSMIC v74 version), a comprehensive resource for exploring somatic mutations in human cancer.[11] Prior to September 9, 2015, the CENP-A gene was tested in 22,696 patients' specimens spanning 37 different types of cancer. While in 26 of these cancer types there were no mutations, 25 point mutations were identified (overall frequency = 0.11%) in the 11 remaining types of cancer; 20 copy number variations (CNV) were determined (overall frequency = 0.09%) in 13 of the cancer types. No translocations, insertions, deletions, or loss of heterozygosity were identified (Table 1). Of the 25 point mutations in the CENP-A gene, 3 were nonsense and 16 were missense; the other 6 mutations were synonymous and did not affect the corresponding amino acids (Table 1). The frequency of point mutations was not high in any of the cancer types, with the highest mutation frequency being 0.61%, found for cancers of the colon (Supporting Information Table S1). As shown in Supporting Information Table S2, two, three and five CNVs were found in upper aerodigestive tract cancer, lung cancer and endometrial cancer, and one CNV in each of 10 other malignancies. Taken together, the results indicate that there were no major alterations in the sequence or copy number of the CENP-A gene that could account for the development of the malignancies.

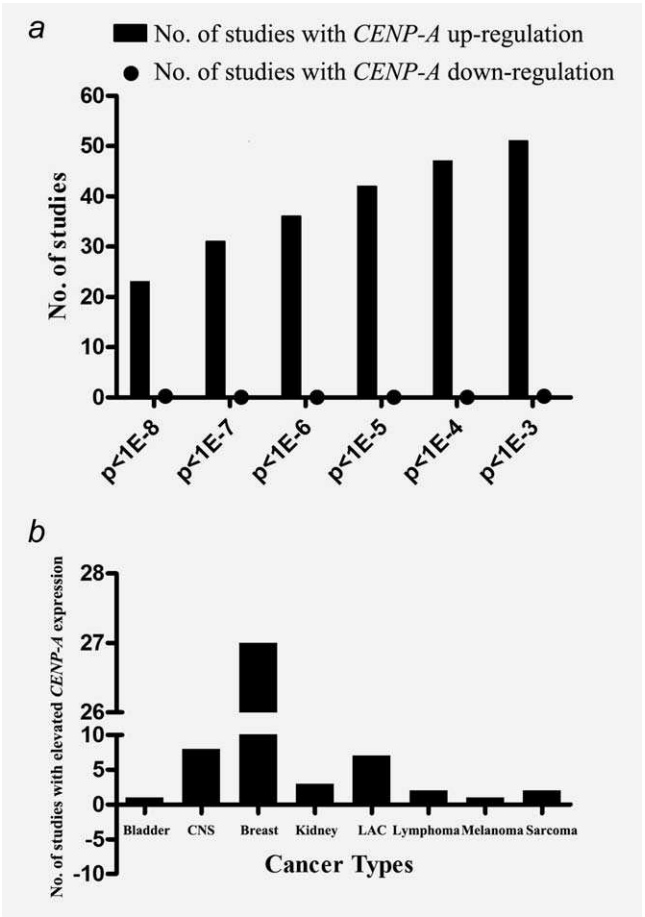

Figure 1. Elevated *CENP-A* expression is significantly correlated with malignancy development and poor patient outcome as indicated by Oncomine analysis. (***a***) The number of studies that significantly correlated with *CENP-A* up-regulation and down-regulation in cancer tissues versus normal tissues, is shown at different *p* values (fold change > 3, gene rank: top 5%). (***b***) The number of studies regarding clinical outcome significantly correlated with elevated *CENP-A* expression in cancer tissues of patients (fold change > 1.5, $p < 0.05$, Gene rank: top 10%). Black columns stand for studies showing positive correlation between elevated *CENP-A* expression and poor clinical outcome. CNS, central nervous system; LAC, lung adenocarcinoma.

Elevated *CENP-A* expression in human cancers

Using Oncomine analysis,[12] we investigated whether expression of the CENP-A gene was altered in human cancers. It was found that *CENP-A* expression was significantly elevated in 20 types of solid tumor compared with normal tissues (51 studies; Fig. 1*a* and Supporting Information Table S3; fold change > 3, $p <$ 1E-3, gene rank: top 5%). In contrast, there was not a single study, using the same criteria, where *CENP-A* was down-regulated in the cancers (Fig. 1*a*). The 51 studies involved 3,738 specimens, with a fold change of 3.01–22.05, and a *p* values of 4.76E-04–1.26E-41 (Supporting

Information Table S3). The most common cancer types showing elevated *CENP-A* expression in these studies were cancers of the breast (21.9%), central nervous system (CNS) (19.8%), ovary (15.9%) and lung (14.8%).

Next, we determined whether elevated *CENP-A* expression had an effect on clinical outcome. Using the Oncomine database (fold change > 1.5, $p < 0.05$, gene rank: top 10%), we found that in another group of 51 studies (2,166 patients) elevated *CENP-A* expression was associated with decreased time of cancer recurrence and time to death (Fig. 1*b*), showing the following distribution: breast cancer (52.9%), CNS cancer (15.7%), lung adenocarcinomas (13.7%) and other cancers (17.7%). While in 6 other studies elevated *CENP-A* expression was associated with increased time of recurrence/death, the relatively low number of patients in these studies ($n = 30$) makes this finding less definitive (Supporting Information Table S4).

We validated the link between elevated *CENP-A* expression and clinical outcome using a Kaplan–Meier plotter.[13,14] Elevated CENP-A mRNA levels significantly correlated with lower overall survival of breast, lung and gastric cancer patients (Figs. 2*a-c*). Furthermore, elevated CENP-A mRNA levels were found to be associated with shorter relapse-free survival of breast cancer patients, as well as shorter first progression time of lung and gastric cancer patients (Supporting Information Fig. S1). Together, the results indicate that elevated *CENP-A* expression is prognostic of poor clinical outcome for a large variety of human cancers.

Association of elevated *CENP-A* expression with metastatic and hormone receptor-negative breast cancer

Using the Kaplan–Meier plotter,[13] we found that elevated *CENP-A* expression in a large group of metastatic breast cancer patients (1,609 patients) was associated with a shorter Distant Metastasis-Free Survival (Fig. 2*d*). This is also confirmed by Oncomine analysis of a clinical cohort of a limited number of breast cancer patients indicating that metastatic breast cancer could be associated with elevated *CENP-A* expression (Fig. 2*e*) (fold change > 1.5, $p < 0.05$, gene rank: top 10%). Furthermore, patients with *BRCA1* mutations, who are prone to early metastatic dissemination,[24] displayed elevated CENP-A mRNA levels compared with patients expressing wild type *BRCA1* (Fig. 2*f*).

Classification of breast cancers according to aggressiveness is currently based on the presence or absence of the estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor 2 (HER2). It has been shown that the triple negative breast cancer subtype ($ER^-/PR^-/HER2^-$), characterized by the absence in the cancers of all three receptors, has the worst overall and disease-free survival.[25] Using the Oncomine database, we identified 36 studies in which elevated *CENP-A* expression in breast cancers significantly correlated with ER or PR negativity, compared with ER or PR positivity (Supporting Information Table S5; fold change-> 1.5, $p < 0.01$, gene rank: top 10%). Furthermore, as found in 11 studies, elevated *CENP-A* expression correlated with triple negative breast cancer (TNBC) (Table 2) (fold change-> 1.5, $p < 0.01$, gene rank: top 10%), further supporting the idea that cancers displaying elevated expression of *CENP-A* represent an aggressive phenotype. In addition, Oncomine analysis revealed that elevated *CENP-A* expression was associated with other clinical parameters indicative of poor prognosis of breast cancer, such as high tumor histologic grade, advanced pathologic tumor stage and lymph node metastasis (Supporting Information Fig. S2). Taken together, the above findings indicate that elevated *CENP-A* levels in breast cancer are indicative of poor patient prognosis.

Gene co-expression network potentially associated with *CENP-A* and potential *CENP-A* regulatory mechanisms

Gene co-expression networks are increasingly used to explore the functionality of genes at the system level.[26] To identify a co-expression network for elevated *CENP-A* expression, we used a mRNA expression array of the Breast Invasive Carcinoma dataset (TCGA, Nature 2012)[17] available at the cBioPortal database. We identified 564 genes with a co-expression score ≥ 0.5 and ≤ -0.5 (Supporting Information Table S6). Using IPA core analysis,[18] we identified five potential top molecular functions of *CENP-A* and co-expressed genes, as well as the top 10 Canonical Pathways associated with the function of these genes (Table 3). As expected, this potential *CENP-A* co-expression gene network consists of genes primarily related to mitosis and cell cycle progression.

We also investigated what transcriptional regulators might act upstream of the *CENP-A* co-expression gene network. Using IPA upstream analysis,[18] the top five potential upstream transcription regulators of this network with high *CENP-A* co-expression scores were identified. They include the tumor growth-promoting genes *FOXM1, PTTG1, E2F8, CCNE1* and *E2F2* (Supporting Information Table S7). Use of JASPAR[19] and MAPPER[20] databases indicated that only FOXM1, E2F8 and E2F2 could directly bind to the CENP-A promoter (Supporting Information Table S7).

In addition, we investigated whether elevated *CENP-A* expression could stem from post-transcriptional regulation of *CENP-A*, for example, by microRNAs. Using microRNA.org[21] and TargetScan[22] databases, we determined whether seed sequences of microRNAs could be targeting CENP-A mRNA. Using stringent criteria (mirSVR score < -1, PhastCons score > 0.5 in microRNA.org; context++ score < -0.4, context++ score percentile $= 99$ in TargetScan), six microRNAs were identified as potentially targeting CENP-A mRNA, that is, hsa-miR-219a-5p, hsa-miR-3198, hsa-miR-4309, hsa-miR-629-3p, hsa-miR-519e-5p and hsa-miR-515-5p (Supporting Information Table S8).

Elevated *CENP-A* expression as a potential predictive biomarker

To investigate whether elevated *CENP-A* expression in pretreatment breast cancer biopsy specimens could be used to

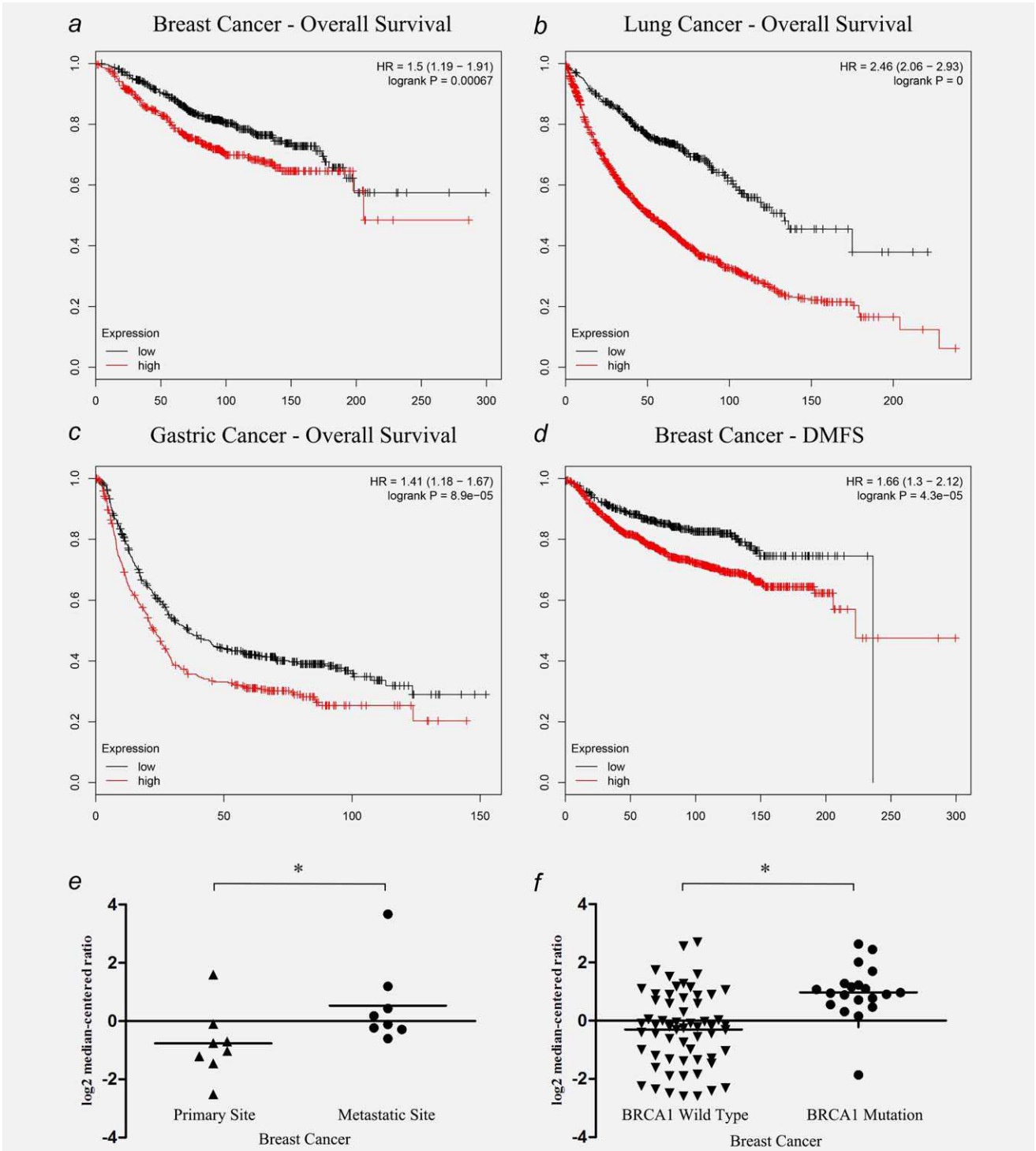

Figure 2. Elevated *CENP-A* expression significantly correlated with poor overall survival of cancer patients, breast cancer metastasis and *BRCA1* mutation. **(*a–c*)** The breast, lung and gastric cancer patients with relatively higher expression of *CENP-A* show poor overall survival compared with patients with relatively lower *CENP-A* expression as indicated by Kaplan–Meier plotter analysis. **(*d*)** Breast cancer patients with even higher *CENP-A* expression take shorter time to develop distant metastasis as indicated by Kaplan–Meier plotter analysis. DMFS, distant metastasis-free survival. **(*e*)** Metastatic breast cancer tissues show higher expression of *CENP-A* compared with primary breast cancer tissues, with $p < 0.05$ and a 2.5-fold change (Weigelt Breast cohort, Oncomine database). **(*f*)** Breast cancer patients with *BRCA1* mutations are positively associated with elevated *CENP-A* expression, with $p < 0.05$ and a 2.4-fold change (Waddell Breast cohort, Oncomine database).

predict an oncolytic response of the patients to chemotherapy, we used Oncomine in search of correlations between the oncolytic response of patients and elevated *CENP-A* expression. A strong correlation was found between elevated *CENP-A* expression and positive oncolytic response to chemotherapy (fold change > 1.5, $p < 0.05$, gene rank: top 10%); furthermore, the effective chemotherapeutic regimens all included use of taxanes (*e.g.*, paclitaxel and docetaxel) (Fig. 3). These results indicate that elevated *CENP-A* expression may be used as a predictive biomarker for positive outcome of taxane-based therapy of breast cancer. Evidence supporting this conclusion comes from a similar correlation found between expressions of a *CENP-A* co-expression network and positive oncolytic response obtained with taxane-based therapy (Supporting Information Table S9; odds ratio > 2, $p <$ 1E-4).

Table 2. Elevated *CENP-A* expression in TNBC (Oncomine database)

Study name	TNBC patient number	Non-TNBC patient number	Fold change (TNBC *vs.* Non-TNBC)	*p* values
Curtis breast	211	1,340	2.14	7.02E-57
TCGA breast	46	250	3.04	1.84E-20
Bittner breast	39	129	2.40	4.33E-8
Hatzis breast	178	320	1.59	8.22E-17
Zhao breast	5	28	3.63	5.90E-4
Waddell breast	22	44	2.15	1.01E-5
Chin breast	19	87	2.24	1.17E-4
Minn breast 2	25	71	1.77	3.03E-4
Kao breast	32	295	1.99	1.13E-5
Richardson breast 2	18	19	2.09	5.80E-4
Stickeler breast	8	24	3.08	3.42E-5

Discussion

While elevated levels of CENP-A have been reported for breast cancer in particular,[27] the present study, based on Oncomine analysis, shows that elevated *CENP-A* expression can be found in as many as 20 types of cancer (Supporting Information Table S3). In the case of breast cancer, we have further confirmed strong correlations between elevated *CENP-A* expression and metastatic ability or lack of hormone receptor expression (TNBC),[9] and discovered strong correlations between elevated *CENP-A* expression and advanced stages of the disease including the presence of *BRCA1* mutations (Fig. 2*f* and Supporting Information Fig. S2). The correlations are consistent with *BRCA1* mutations generally being associated with metastatic and aggressive disease,[24] and with TNBC patients having the worst overall and disease-free survival[25] and poor distant metastasis-free time.[28] Furthermore, we found statistically significant correlations between elevated *CENP-A* expression and poor outcome of patients with breast, bladder, CNS, kidney and gastric cancer, as well as lung adenocarcinoma, lymphoma, melanoma and sarcoma (Fig. 1*b*; Supporting Information Fig. S1). Taken together, the

Table 3. Molecular functions and canonical pathways associated with *CENP-A* and co-expressed genes in TCGA Breast Cancer, Nature 2012 (IPA)

Molecular functions	*p* values range	# genes
Cell cycle progression	6.66E-04–2.98E-40	200
Cellular assembly and organization	4.79E-04–2.15E-38	165
DNA replication, recombination and repair	6.66E-04–2.15E-38	188
Cellular movement	6.73E-05–2.01E-17	32
Cell death and survival	6.38E-04–1.78E-14	191
Canonical pathways	***p* values[1]**	**Ratio[2]**
Cell cycle control of chromosomal replication	1.65E-13	14/27 (0.519)
Role of BRCA1 in DNA damage response	6.83E-13	20/78 (0.256)
Mitotic roles of polo-like kinase	4.63E-11	17/66 (0.258)
Cell cycle: G2/M DNA damage checkpoint regulation	4.95E-11	15/49 (0.306)
Mismatch repair in eukaryotes	9.10E-08	8/16 (0.5)
Estrogen-mediated S-phase entry	1.66E-07	9/24 (0.375)
Hereditary breast cancer signaling	1.97E-07	18/129 (0.14)
Role of CHK proteins in cell cycle checkpoint control	3.51E-07	12/55 (0.218)
ATM signaling	7.04E-06	11/59 (0.186)
Cyclins and cell cycle regulation	1.62E-05	12/78 (0.154)

[1]The *p* values is Benjamini–Hochberg (B-H) multiple-test correction *p* values, a B-H $p < 0.05$ was considered significant.
[2]Ratio of the genes from co-expression list to the total genes in the pathway.

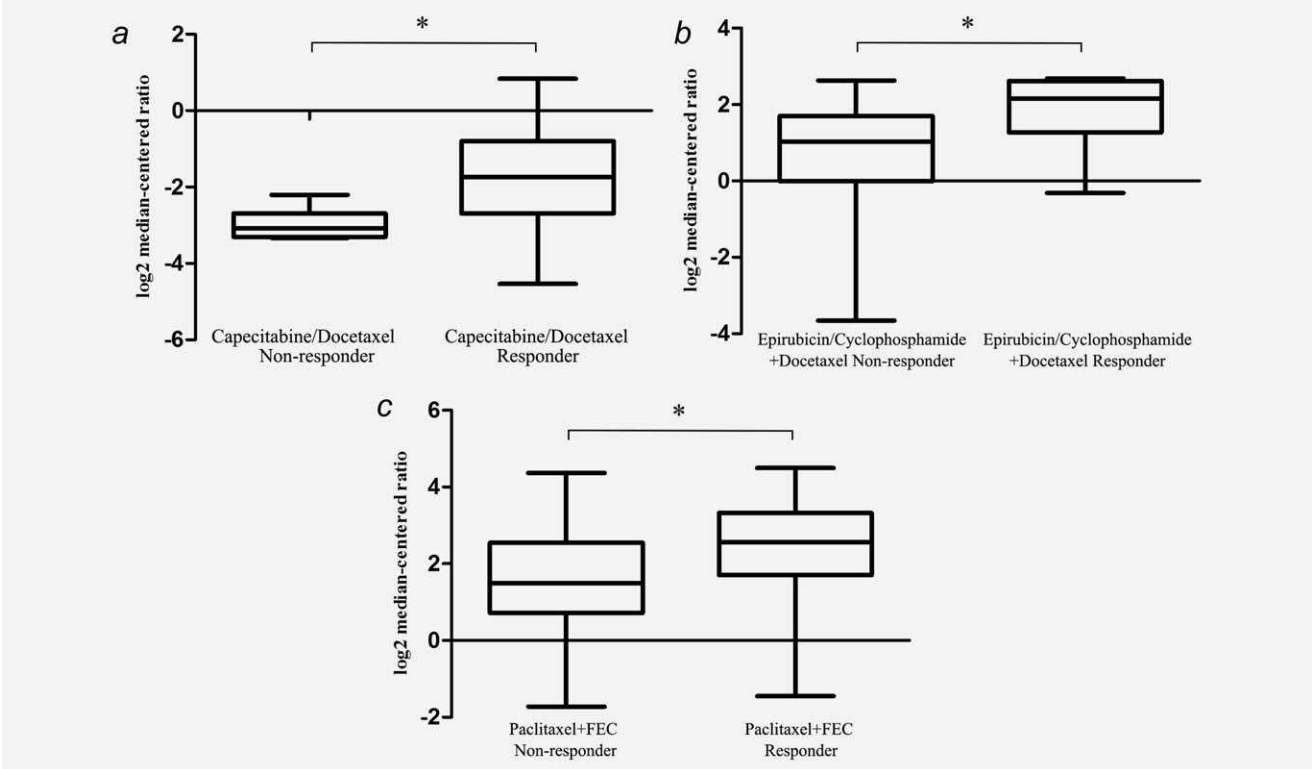

Figure 3. Response to taxane-based chemotherapy of breast cancer patients correlates with elevated *CENP-A* expression as indicated by Oncomine analysis. (***a***) Patients with elevated expression of *CENP-A* show better response to Capecitabine/Docetaxel treatment (Gluck Breast cohort, Oncomine database). (***b***) Patients with elevated expression of *CENP-A* show better response to Epirubicin/Cyclophosphamide plus Docetaxel treatment (Stickeler Breast cohort, Oncomine database). (***C***) Patients with elevated expression of *CENP-A* show better response to Paclitaxel plus FEC treatment (Miyake Breast cohort, Oncomine database). FEC, 5-fluorouracil/epirubicin/cyclophosphamide (fold change > 1.5, $p < 0.05$, Gene rank: top 10%).

data indicate that elevated *CENP-A* expression may be used as a prognostic biomarker of poor patient outcome in a great variety of cancers.

In trying to identify mechanisms underlying the elevated expression of *CENP-A*, genetic analysis has indicated that elevated *CENP-A* expression is not due to increases in the copy number of the gene, as the number of such alterations was small relative to common mutation frequencies in cancers (Table 1 and Supporting Information Table S2).[29] Elevated *CENP-A* expression may be related to a role of CENP-A protein in DNA repair. Thus upon induction of DNA breaks in cells, *CENP-A* expression is substantially increased followed by recruitment of the protein to the damaged sites.[30] A role for CENP-A in DNA repair is consistent with the finding that elevated *CENP-A* expression correlates with the presence of *BRCA1* mutations in breast cancer (Fig. 2*f*), as such mutations can lead to compromised repair of DNA double-strand breaks.[31] Furthermore, it has been reported that enhanced *CENP-A* expression can result from lack of Rb protein activity.[32] These findings suggest that elevated expression of *CENP-A* may result from a number of molecular events involved in the development of the disease.

Five genes were identified through IPA upstream regulator analysis as potential upstream regulators of *CENP-A* and its co-expression network, that is, *FOXM1, PTTG1, E2F8, CCNE1* and *E2F2,* genes that play important roles in cell cycle progression and tumorigenesis (Supporting Information Table S7).[33–37] Furthermore, FOXM1, E2F8 and E2F2 proteins were shown to have potential to bind to the CENP-A promoter sequence, a finding supporting a role for these genes in *CENP-A* transcription regulation (Supporting Information Table S7). It has been demonstrated by ChIP-seq that FOXM1 can directly bind to the CENP-A promoter in Chromosome 2,[38,39]—a finding suggesting that FOXM1 is directly involved in stimulating *CENP-A* expression.

There is considerable evidence that microRNAs can also play a role in regulating gene expression by interacting with mRNA.[40] With regard to the six microRNAs that could play a role in post-transcriptional regulation of *CENP-A* (Supporting Information Table S8), it is of interest that hsa-miR-219a-5p, hsa-miR-515-5p, has-mir-629-3p and hsa-miR-519e-5p are down-regulated in cancers.[41–44] Their down-regulation would be consistent with elevated expression of *CENP-A*. Moreover, hsa-miR-219a-5p and hsa-miR-515-5p have been identified as tumor suppressors,[41,42] further supporting the idea that these microRNAs can function as CENP-A post-transcriptional regulators.

A role of *CENP-A* in cancer may be indicated by the profile of its co-expression gene network consisting of 564 genes

(Supporting Information Table S6). As suggested by IPA core analysis, the activities of *CENP-A* and its co-expressed genes are linked to cell proliferation (Table 3). Thus many of the co-expressed transcripts, in particular of the four top co-expressed genes, that is, *BUB1, KIF2C, CDCA8* and *TPX2*, have key roles in mitosis, notably in kinetochore and spindle assemblies, thereby facilitating cancer cell proliferation (Supporting Information Table S6).[45–48] Taken together, these data indicate that *CENP-A* has a growth-promoting role in cancer.

Ectopic location of CENP-A protein in chromosome arms can markedly induce aberrant transcription of genes in cancer cells, including oncogenes and tumor suppressor genes.[4] Various studies have shown that the ectopic location is not random, but is preferentially located near telomeres and that ectopic CENP-A can recruit telomerase.[4,6] Since telomerase can promote cancer cell immortality by maintaining the length of telomeres, CENP-A could also be involved in cancer cell immortalization. As such, *ectopic* CENP-A could serve as an oncogene promoting cancer survival and malignant progression.

While taxane-based chemotherapeutic regimens are widely used for treatment of breast cancer,[49] other non-taxane-based therapies are also included in the Oncomine database.[50] It is of great interest that in this study a strong correlation was found between elevated expression of *CENP-A* and a positive oncolytic response of breast cancers which was only obtained with taxane-based regimens (Fig. 3; Supporting Information Table S9). The reason for this is at present not clear.

In conclusion, the present study indicates that elevated expression of *CENP-A* can be useful as a biomarker of poor patient prognosis in a large variety of human cancers. As well, it may be used as a biomarker for predicting the response of breast cancer patients to taxane-based chemotherapy and possibly other treatments targeting cell division.

Authors' Contributions

Study concept and design: XS, SQ
Acquisition of data: XS, PLC, WJ, SQ
Analysis and interpretation of data: XS, SQ
Draft of the manuscript: XS, PLC, CDH, PWG, YW, SQ
Critical revision of the manuscript for important intellectual content: PWG, SQ, YW, XS
Obtained funding: XS
All authors read and approved the final manuscript.

References

1. Padeganeh A, Ryan J, Boisvert J, et al. Octameric CENP-A nucleosomes are present at human centromeres throughout the cell cycle. *Curr Biol* 2013;23:764–9.
2. Verdaasdonk JS, Bloom K. Centromeres: unique chromatin structures that drive chromosome segregation. *Nat Rev Mol Cell Biol* 2011;12:320–32.
3. Tomonaga T, Matsushita K, Yamaguchi S, et al. Overexpression and mistargeting of centromere protein-A in human primary colorectal cancer. *Cancer Res* 2003;63:3511–6.
4. Athwal RK, Walkiewicz MP, Baek S, et al. CENP-A nucleosomes localize to transcription factor hotspots and subtelomeric sites in human cancer cells. *Epigenetics Chromatin* 2015;8:2.
5. Heun P, Erhardt S, Blower MD, et al. Mislocalization of the Drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. *Dev Cell* 2006;10:303–15.
6. Olszak AM, van Essen D, Pereira AJ, et al. Heterochromatin boundaries are hotspots for de novo kinetochore formation. *Nat Cell Biol* 2011;13:799–808.
7. Schvartzman JM, Sotillo R, Benezra R. Mitotic chromosomal instability and cancer: mouse modelling of the human disease. *Nat Rev Cancer* 2010;10:102–15.
8. Ferguson LR, Chen H, Collins AR, et al. Genomic instability in human cancer: molecular insights and opportunities for therapeutic attack and prevention through diet and nutrition. *Semin Cancer Biol* 2015;35Suppl:S5–S24.
9. Zhang C, Han Y, Huang H, et al. Integrated analysis of expression profiling data identifies three genes in correlation with poor prognosis of triple-negative breast cancer. *Int J Oncol* 2014;44:2025–33.
10. Qiu JJ, Guo JJ, Lv TJ, et al. Prognostic value of centromere protein-A expression in patients with epithelial ovarian cancer. *Tumour Biol* 2013;34:2971–5.
11. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res* 2015;43:D805–11.
12. Rhodes DR, Kalyana-Sundaram S, Mahavisno V, et al. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* 2007;9:166–80.
13. Győrffy B, Lanczky A, Eklund AC, et al. 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 Res Treat* 2010;123:725–31.
14. Győrffy B, Surowiak P, Budczies J, et al. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One* 2013;8:e82241.
15. Li Q, Birkbak NJ, Gyorffy B, et al. Jetset: selecting the optimal microarray probe set to represent a gene. *BMC Bioinform* 2011;12:474.
16. Gao J, Aksoy BA, Dogrusoz U, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* 2013;6:pl1.
17. CGA Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;490:61–70.
18. Krämer A, Green J, Pollard J, et al. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* 2014;30:523–30.
19. Mathelier A, Fornes O, Arenillas DJ, et al. JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* 2016;44:D110–5.
20. Marinescu VD, Kohane IS, Riva A. The MAPPER database: a multi-genome catalog of putative transcription factor binding sites. *Nucleic Acids Res* 2005;33:D91–7.
21. Betel D, Wilson M, Gabow A, et al. The microRNA.org resource: targets and expression. *Nucleic Acids Res* 2008;36:D149–53.
22. Agarwal V, Bell GW, Nam JW, et al. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 2015;4:e05005.
23. Solskov L, Magnusson NE, Kristiansen SB, et al. Microarray expression analysis in delayed cardioprotection: the effect of exercise, AICAR, or metformin and the possible role of AMP-activated protein kinase (AMPK). *Mol Cell Biochem* 2012;360:353–62.
24. Bai F, Chan HL, Scott A, et al. BRCA1 suppresses epithelial-to-mesenchymal transition and stem cell dedifferentiation during mammary and tumor development. *Cancer Res* 2014;74:6161–72.
25. Onitilo AA, Engel JM, Greenlee RT, et al. Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. *Clin Med Res* 2009;7:4–13.
26. Villa-Vialaneix N, Liaubet L, Laurent T, et al. The structure of a gene co-expression network reveals biological functions underlying eQTLs. *PLoS One* 2013;8:e60045.
27. Rajput AB, Hu N, Varma S, et al. Immunohistochemical assessment of expression of centromere protein-a (cenpa) in human invasive breast cancer. *Cancers (Basel)* 2011;3:4212–27.
28. Haffty BG, Yang Q, Reiss M, et al. Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer. *J Clin Oncol* 2006;24:5652–7.
29. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 2014;505:495–501.

30. Zeitlin SG, Baker NM, Chapados BR, et al. Double-strand DNA breaks recruit the centromeric histone CENP-A. *Proc Natl Acad Sci U S A* 2009; 106:15762–7.
31. O'Donovan PJ, Livingston DM. BRCA1 and BRCA2: breast/ovarian cancer susceptibility gene products and participants in DNA double-strand break repair. *Carcinogenesis* 2010;31:961–7.
32. Amato A, Schillaci T, Lentini L, et al. CENPA overexpression promotes genome instability in pRb-depleted human cells. *Mol Cancer* 2009;8: 119.
33. Wang Z, Park HJ, Carr JR, et al. FoxM1 in tumorigenicity of the neuroblastoma cells and renewal of the neural progenitors. *Cancer Res* 2011;71:4292–302.
34. Wondergem B, Zhang Z, Huang D, et al. Expression of the PTTG1 oncogene is associated with aggressive clear cell renal cell carcinoma. *Cancer Res* 2012;72:4361–71.
35. Deng Q, Wang Q, Zong WY, et al. E2F8 contributes to human hepatocellular carcinoma via regulating cell proliferation. *Cancer Res* 2010; 70:782–91.
36. Nakayama N, Nakayama K, Shamima Y, et al. Gene amplification CCNE1 is related to poor survival and potential therapeutic target in ovarian cancer. *Cancer* 2010;116:2621–34.
37. Nakahata AM, Suzuki DE, Rodini CO, et al. RNAi-mediated knockdown of E2F2 inhibits tumorigenicity of human glioblastoma cells. *Oncol Lett* 2014;8:1487–91.
38. Yau C, Meyer L, Benz S, et al. FOXM1 cistrome predicts breast cancer metastatic outcome better than FOXM1 expression levels or tumor proliferation index. *Breast Cancer Res Treat* 2015;154: 23–32.
39. Grant GD, Brooks L, Zhang X, et al. Identification of cell cycle-regulated genes periodically expressed in U2OS cells and their regulation by FOXM1 and E2F transcription factors. *Mol Biol Cell* 2013;24:3634–50.
40. Zeng Y, Yi R, Cullen BR. MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc Natl Acad Sci U S A* 2003;100:9779–84.
41. Huang C, Cai Z, Huang M, et al. miR-219-5p modulates cell growth of papillary thyroid carcinoma by targeting estrogen receptor α. *J Clin Endocrinol Metab* 2015;100:E204–13.
42. Pinho FG, Frampton AE, Nunes J, et al. Downregulation of microRNA-515-5p by the estrogen receptor modulates sphingosine kinase 1 and breast cancer cell proliferation. *Cancer Res* 2013; 73:5936–48.
43. Shen J, Wang D, Gregory SR, et al. Evaluation of microRNA expression profiles and their associations with risk alleles in lymphoblastoid cell lines of familial ovarian cancer. *Carcinogenesis* 2012;33: 604–12.
44. Wang M, Zhao C, Shi H, et al. Deregulated microRNAs in gastric cancer tissue-derived mesenchymal stem cells: novel biomarkers and a mechanism for gastric cancer. *Br J Cancer* 2014; 110:1199–210.
45. Grabsch H, Takeno S, Parsons WJ, et al. Overexpression of the mitotic checkpoint genes BUB1, BUBR1, and BUB3 in gastric cancer–association with tumour cell proliferation. *J Pathol* 2003;200: 16–22.
46. Gwon MR, Cho JH, Kim JR. Mitotic centromere-associated kinase (MCAK/Kif2C) regulates cellular senescence in human primary cells through a p53-dependent pathway. *FEBS Lett* 2012;586:4148–56.
47. Hayama S, Daigo Y, Yamabuki T, et al. Phosphorylation and activation of cell division cycle associated 8 by aurora kinase B plays a significant role in human lung carcinogenesis. *Cancer Res* 2007;67:4113–22.
48. Pérez de Castro I, Malumbres M. Mitotic stress and chromosomal instability in cancer: the case for TPX2. *Genes Cancer* 2012;3:721–30.
49. Gradishar WJ. Taxanes for the treatment of metastatic breast cancer. *Breast Cancer (Auckl)* 2012; 6:159–71.
50. Bonnefoi H, Potti A, Delorenzi M, et al. Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: a substudy of the EORTC 10994/BIG 00-01 clinical trial. *Lancet Oncol* 2007;8:1071–8.