

RESEARCH

Open Access



Genomic instability-derived plasma extracellular vesicle-microRNA signature as a minimally invasive predictor of risk and unfavorable prognosis in breast cancer

Siqi Bao^{1†}, Ting Hu^{2†}, Jiaqi Liu^{3†}, Jianzhong Su¹, Jie Sun¹, Yue Ming⁴, Jiaxin Li³, Nan Wu^{5*}, Hongyan Chen^{2*} and Meng Zhou^{1*} 

Abstract

Background: Breast cancer (BC) is the most frequently diagnosed cancer and the leading cause of cancer-associated deaths in women. Recent studies have indicated that microRNA (miRNA) regulation in genomic instability (GI) is associated with disease risk and clinical outcome. Herein, we aimed to identify the GI-derived miRNA signature in extracellular vesicles (EVs) as a minimally invasive biomarker for early diagnosis and prognostic risk stratification.

Experimental design: Integrative analysis of miRNA expression and somatic mutation profiles was performed to identify GI-associated miRNAs. Then, we constructed a discovery and validation study with multicenter prospective cohorts. The GI-derived miRNA signature (miGISig) was developed in the TCGA discovery cohort (n = 261), and was subsequently independently validated in internal TCGA validation (n = 261) and GSE22220 (n = 210) cohorts for prognosis prediction, and in GSE73002 (n = 3966), GSE41922 (n = 54), and in-house clinical exosome (n = 30) cohorts for diagnostic performance.

Results: We identified a GI-derived three miRNA signature (*MIR421*, *MIR128-1* and *MIR128-2*) in the serum extracellular vesicles of BC patients, which was significantly associated with poor prognosis in all the cohorts tested and remained as an independent prognostic factor using multivariate analyses. When integrated with the clinical characteristics,

*Correspondence: dr.wunan@pumch.cn; chenhongyan@picamc.ac.cn; zhoumeng@wmu.edu.cn

[†]Siqi Bao, Ting Hu and Jiaqi Liu contribute equally to this work as first authors

¹ School of Biomedical Engineering, School of Ophthalmology & Optometry and Eye Hospital, Wenzhou Medical University, Wenzhou 325027, People's Republic of China

² State Key Laboratory of Molecular Oncology, National Cancer Center/ National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, People's Republic of China

⁵ Department of Orthopedic Surgery, Beijing Key Laboratory for Genetic Research of Skeletal Deformity & Key Laboratory of Big Data for Spinal Deformities, State Key Laboratory of Complex Severe and Rare Diseases, All at Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, People's Republic of China

Full list of author information is available at the end of the article



© The Author(s) 2021. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

the composite miRNA-clinical prognostic indicator showed improved prognostic performance. The miGISig also showed high accuracy in differentiating BC from healthy controls with the area under the receiver operating characteristics curve (ROC) with 0.915, 0.794 and 0.772 in GSE73002, GSE41922 and TCGA cohorts, respectively. Furthermore, circulating EVs from BC patients in the in-house cohort harbored elevated levels of miGISig, with effective diagnostic accuracy.

Conclusions: We report a novel GI-derived three miRNA signature in EVs, as an excellent minimally invasive biomarker for the early diagnosis and unfavorable prognosis in BC.

Keywords: Breast cancer, Genomic instability, Extracellular vesicle, Exosomes, microRNA

Background

Breast cancer (BC) is one of the most commonly diagnosed cancer types, accounting for 30% of all new cancer diagnoses in women. The incidence rate of BC has remained generally stable over the past few decades [1]. Despite the recent improvements in detection methods and therapeutic choice, BC remains a significant public health issue worldwide, and the clinical outcome of patients varies markedly due to the stage at diagnosis and molecular subtype [2]. Classic clinicopathological features, including tumor size, histological subtypes and grades, lymphatic invasion, lymph node metastasis, estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) have effective value in guiding clinical decisions; Serum tumor markers, such as carcinoembryonic antigen (CEA), cancer antigen 19-9 (CA19-9), cancer antigen 125 (CA125), cancer antigen 15-3 (CA15-3), have been typically used for follow-up monitoring [3, 4]. However, BC patients are still faced with variable biological and clinical behaviors, due to high molecular and cellular heterogeneity [5, 6]. Therefore, it is urgently needed to identify reliable and robust biomarkers to enhance early risk detection and prognosis prediction with minimal invasion for improving clinical management and treatment decision-making for BC patients.

Genomic instability (GI) is defined as a process in which genomic changes are prone to increase and can influence the phenotype [7, 8], and has been recognized as an evolving hallmark or characteristic of most types of cancer [9]. GI promotes inter- and intra-tumor heterogeneity and is a major driving force for cancer cells to survive, proliferate, and disseminate [10]. Frequent GI was commonly observed in BC cells, including numerical and structural genomic changes [11, 12]. The molecular basis of GI in BC has not been fully elucidated, and BC can be characterized by vast GI-derived heterogeneity; therefore, understanding the molecular basis of GI in BC would not only enable the molecular etiology and pathobiology of BC to be determined, but also to develop

improved cancer prevention, diagnosis, and prognosis methods [13, 14].

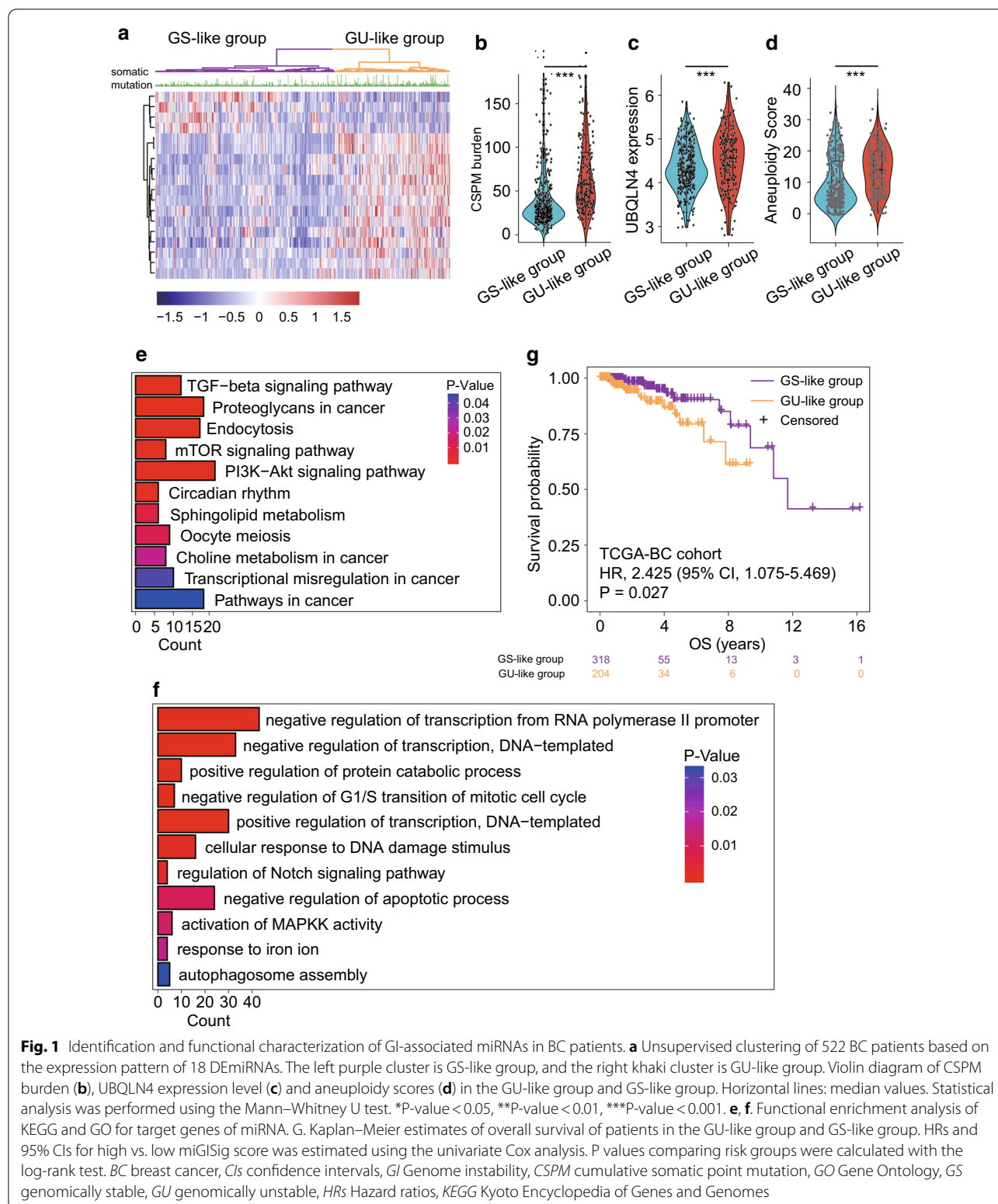
MicroRNA (miRNA) is a type of small, evolutionarily conserved, single-stranded, non-coding RNA, that induces mRNA degradation or translational repression by completely or incompletely binding to the target mRNA [15–17]. Accumulating evidence has highlighted the important roles of miRNAs, as crucial master regulators of various biological processes, including cell differentiation, development, and homeostasis [18, 19]. Deregulation of miRNA function plays a vital role in cancer pathogenesis [20–23]. Previous studies have revealed that several miRNAs are involved in enhancing GI by impairing DNA repair or preventing GI by enhancing the response to DNA damage [24]. In addition, GI may be mediated by horizontal transfer of tumor-derived macromolecules, such as miRNAs, via EVs [25]. Whether these alterations of GI-associated miRNAs can be detected in EVs, and therefore have clinical utility as promising minimally invasive biomarkers in BC, however, has not been investigated.

In this study, we aimed to identify miRNAs involved in GI based on the mutator hypothesis by integrating expression and somatic mutation profiles, and develop a GI-derived miRNA signature (miGISig) for early diagnosis and prognosis prediction of BC patients. In addition, we also uncovered the functional roles of the newly identified three miRNAs in the regulation of GI and finally, we assessed the use of the miGISig, as a minimally invasive biomarker in circulating exosomes to identify BC patients from asymptomatic controls.

Results

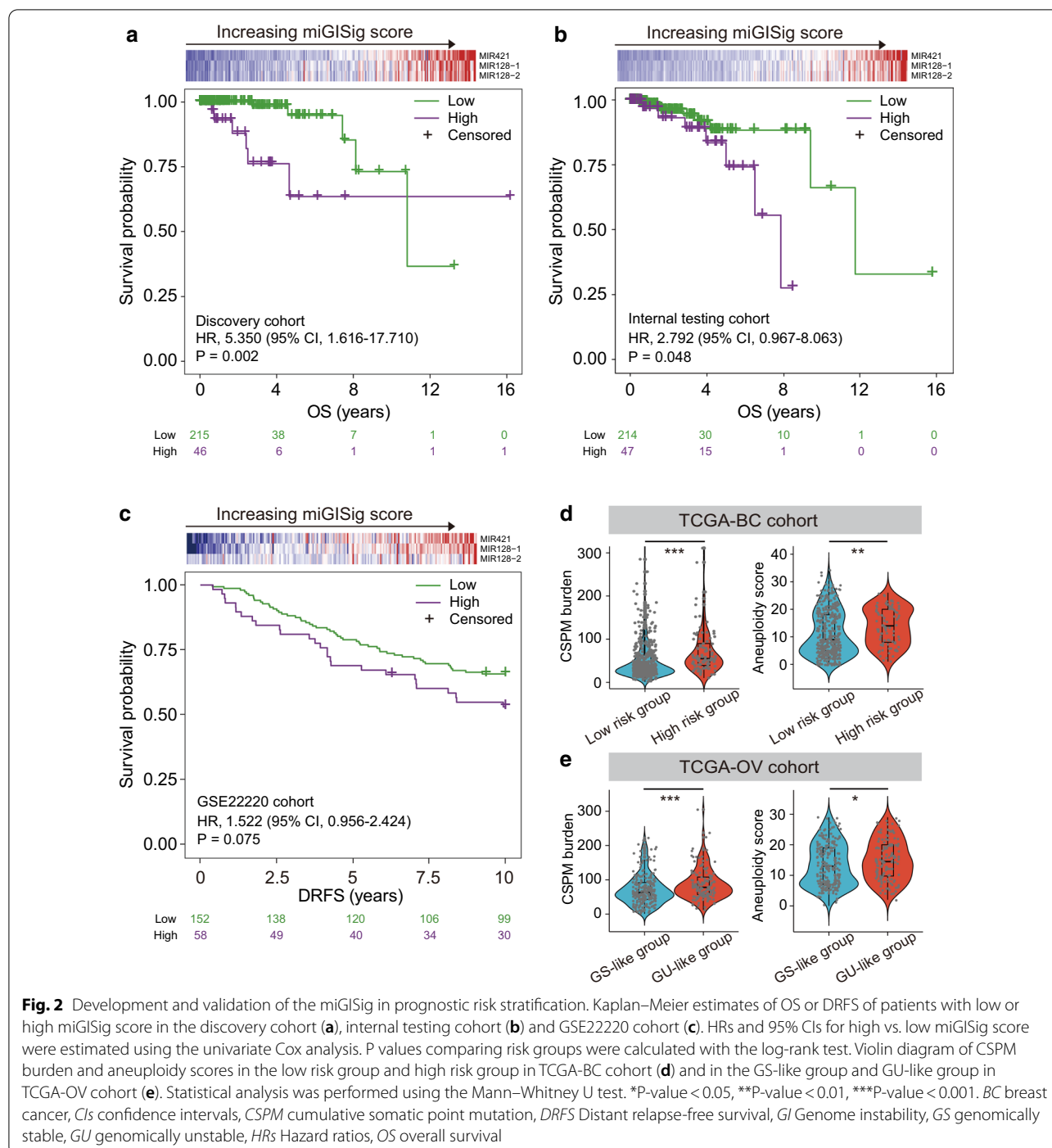
Identification of GI-associated miRNAs in BC

Differential analysis of miRNA expression was performed between the GU-like and GS-like groups using TCGA-BC cohort, in which, 18 differentially expressed miRNAs (DEmiRNAs) were identified (Additional file 1: Table S4). Among them, 14 and 4 miRNAs were found to be up- and downregulated, respectively, in the GS-like tumors compared with that in the GU-like tumors. The unsupervised hierarchical clustering



analysis based on the expression level of the 18 DEMiRNAs produced two patient clusters: a GS-like cluster (n = 318) and a GU-like cluster (n = 204), respectively

(Fig. 1a). There was a significantly high frequency of somatic mutations in the GU-like group compared with that in the GS-like group (median value 57.5 vs.



29; $P < 0.001$, Mann–Whitney U test; Fig. 1b). Meanwhile, the expression level of the *UBQLN4* gene and aneuploidy score in the GU-like group was significantly higher compared with that in the GS-like group ($P < 0.001$, Mann–Whitney U test; Figs. 1c and 1d).

Functional enrichment analysis of the 18 DE miRNAs-specific mRNA targets identified several enriched biological processes and pathways (Figs. 1e and 1f), including transforming growth factor (TGF)- β , regulation of transcription, negative regulation of G_1/S transition of mitotic cell cycle and cellular response to DNA damage stimulus, and the Notch signaling pathway,

which are all known to be GI-related biological pathways. Furthermore, there was a significant difference in OS time between the GU-like and the GS-like groups (median OS, 11.7 vs. > 16 years; $P=0.027$, log-rank test; Fig. 1g). These results suggested that the 18 DEmiRNAs were involved in GI and associated with BC patients' prognosis.

Development and validation of a GI-derived miRNA signature for prognostic risk stratification

We first analyzed the panel of 18 GI-related miRNAs using the univariate and multivariate Cox proportional hazard regression analysis in the TCGA discovery cohort and identified three-miRNA signature (hereafter referred to as miGISig) as follows: $\text{miGISig score} = (0.501 \times \text{expression level of } MIR421) + (0.6808 \times \text{expression level of } MIR128-1) + (-0.3617 \times \text{expression level of } MIR128-2)$ (Additional file 1: Table S5). When stratified by the optimal cut-off value (cut-off, 0.516), the miGISig demonstrated significant prognostic value, as shown by the Kaplan–Meier curves of the risk groups, with high miGISig demonstrating poor prognosis (HR=5.350, 95% CI 1.616–17.710; $P=0.002$, log-rank test; Fig. 2a). The 5-year OS rates for miGISig-driven low-risk and high-risk groups were 94.8 and 63.5%, respectively. The prognostic value of the miGISig was subsequently validated in multiple independent datasets, which were in agreement with the findings driven from the initial TCGA discovery cohort. The miGISig achieved significant or marginally significant discrimination for the survival time (internal testing cohort, $P=0.048$, and GSE22220 cohort, $P=0.075$, log-rank test; Fig. 2b, c). These results demonstrated the robust performance of the miGISig in predicting unfavorable prognosis in BC patients.

Next, we further examined the relationship between miGISig and CSPM burden, and aneuploidy score. In the TCGA-BC cohort, the CSPM burden and aneuploidy score of patients in the high-risk group was significantly higher compared with that in patients in the low risk-group (median somatic mutations 56 vs. 34, $P<0.001$; and median aneuploidy score 14 vs. 9, $P=0.002$; Mann–Whitney U test; Fig. 2d). One of the core features of OV is GI. The miGISig score of TCGA-OV patients was calculated, the same used in the TCGA-BC cohort. These patients were then classified into the GU-like and GS-like groups based on the median score of miGISig. There was a similar trend as in BC, with significantly higher CSPM burden and aneuploidy score in the GU-like group compared with that in the GS-like group (median somatic mutations 77 vs. 59.5, $P<0.001$; and median aneuploidy score 14.5 vs. 13, $P=0.033$; Mann–Whitney U test;

Fig. 2e). These results highlighted the association of the miGISig with GI.

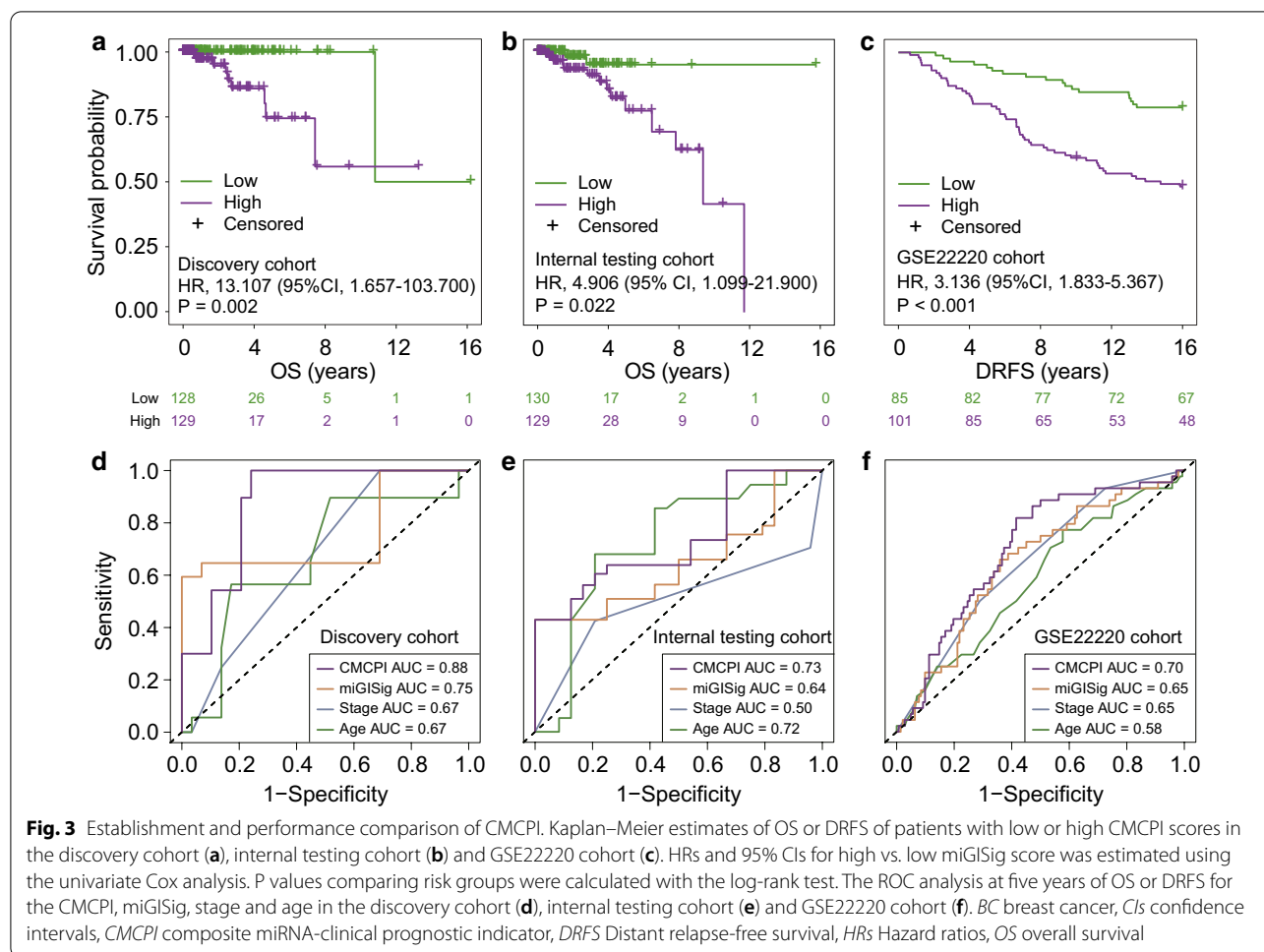
Association between the miGISig and clinical characteristics

We first compared the clinical characteristics between the miGISig-derived high-risk and low-risk groups, and found that ER, PR and TP53 mutations were significantly different (Additional file 1: Table S6). Patients with high miGISig were more likely to be characterized as ER-/PR-negative and high TP53 mutation, whereas patients with ER-/PR-positive and low TP53 mutation rates were enriched in the miGISig-derived low-risk group, suggesting that the miGISig was additionally associated with known prognostic factors. Therefore, to further examine whether the prognostic value of the miGISig was independent of these common clinicopathological factors, the performance of miGISig was tested compared with common clinical variables, using multivariate analysis. The results demonstrated a significant association between miGISig and poor prognosis when adjusted for the various clinical factors in all BC cohorts, indicating that the miGISig was an independent predictor of poor prognosis (Additional file 1: Figure S1).

Establishment of a composite miRNA-clinical prognostic indicator

To further improve the prognostic performance, we combined the miGISig with stage and age to fit a multivariate cox regression model in TCGA discovery cohort and established a composite miRNA-clinical prognostic indicator (CMCPI) calculated as $(0.069 \times \text{age}) + (0.527 \times \text{stage}) + (1.363 \times \text{miGISig score})$. The median CMCPI score of the discovery cohort was used as a cut-off value for stratifying patients. Remarkably, the CMCPI was significantly associated with prognosis in all BC cohorts, and the CMCPI-derived high-risk group had significantly shorter survival times compared with patients in the CMCPI-derived low-risk group (HR=13.107, 95% CI 1.657–103.700, $P=0.002$ for TCGA discovery cohort; HR=4.906, 95% CI 1.099–21.900, $P=0.022$ for the internal testing cohort, and HR=3.136, 95% CI 1.833–5.367, $P<0.001$ for the GSE22220 cohort; Fig. 3a–c).

We further compared the CMCPI with the miGISig, age and stage in all BC cohorts. The CMCPI achieved significantly improved survival estimation with 5-year ROC AUC of 0.88, 0.73 and 0.70 compared with age, stage and miGISig alone in all BC cohorts, respectively (Fig. 3d–f). These results further strengthened the clinical significance of the miGISig in combination with additional important clinical features, with effective predictive performance in predicting poor prognosis in BC patients.



Minimally invasive diagnostic value of the miGISig for BC

In addition, we next examined the performance of the miGISig for risk assessment of BC. We first performed comparative expression analysis for *MIR421*, *MIR128-1* and *MIR128-2* between tumors and healthy controls. The three miRNAs had significantly higher expression in breast tumors compared with that in healthy controls in TCGA, GSE73002 and GSE41922 cohorts (Fig. 4a–c), which indicated their oncogenic roles in BC development and highlighted the potential of the miGISig for the risk assessment of BC. Therefore, we assessed the diagnostic accuracy of the miGISig in detecting BC patients. The miGISig could effectively differentiate BC from healthy controls in TCGA (AUC=0.772), GSE73002 (AUC=0.915) and GSE41922 (AUC=0.794) cohorts, despite the absence of *MIR-128-2* in the GSE41922 cohort (Fig. 4d–f), demonstrating the reliable and robust performance of the miGISig in BC risk prediction.

To further investigate the clinical adaptability of the miGISig as a potential minimally invasive diagnostic biomarker, we next analyzed the expression level of these

miRNAs in circulating exosomes from our in-house clinical cohort of 30 samples using small RNA-seq. Compared with that in healthy individuals, the expression level of exosomal *miR-128* and *miR-421* were significantly higher in the serum of BC patients (P=0.003 for *miR-128* and P=0.002 for *miR-421*, Mann–Whitney U test; Fig. 4g). In addition, we evaluated the diagnostic value of exosomal *miR-128* and *miR-421* using ROC analysis, which revealed that both exosomal *miR-128* (AUC=0.825) and *miR-421* (AUC=0.835) were significant predictors of BC risk (Fig. 4h). Furthermore, the expression levels were also significantly higher in BC patients with stage I for all the miRNAs tested (Figure S2A). BC patients with stage I were distinguished from healthy individuals with an AUC of 0.809 and 0.773 for *miR-421* and *miR-128*, respectively (Additional file 1: Figure S2B). Taken together, our results validated and confirmed the in silico findings from public BC datasets and highlighted that the miGISig could serve as a minimally invasive diagnostic biomarker for early risk assessment of BC in the clinic.

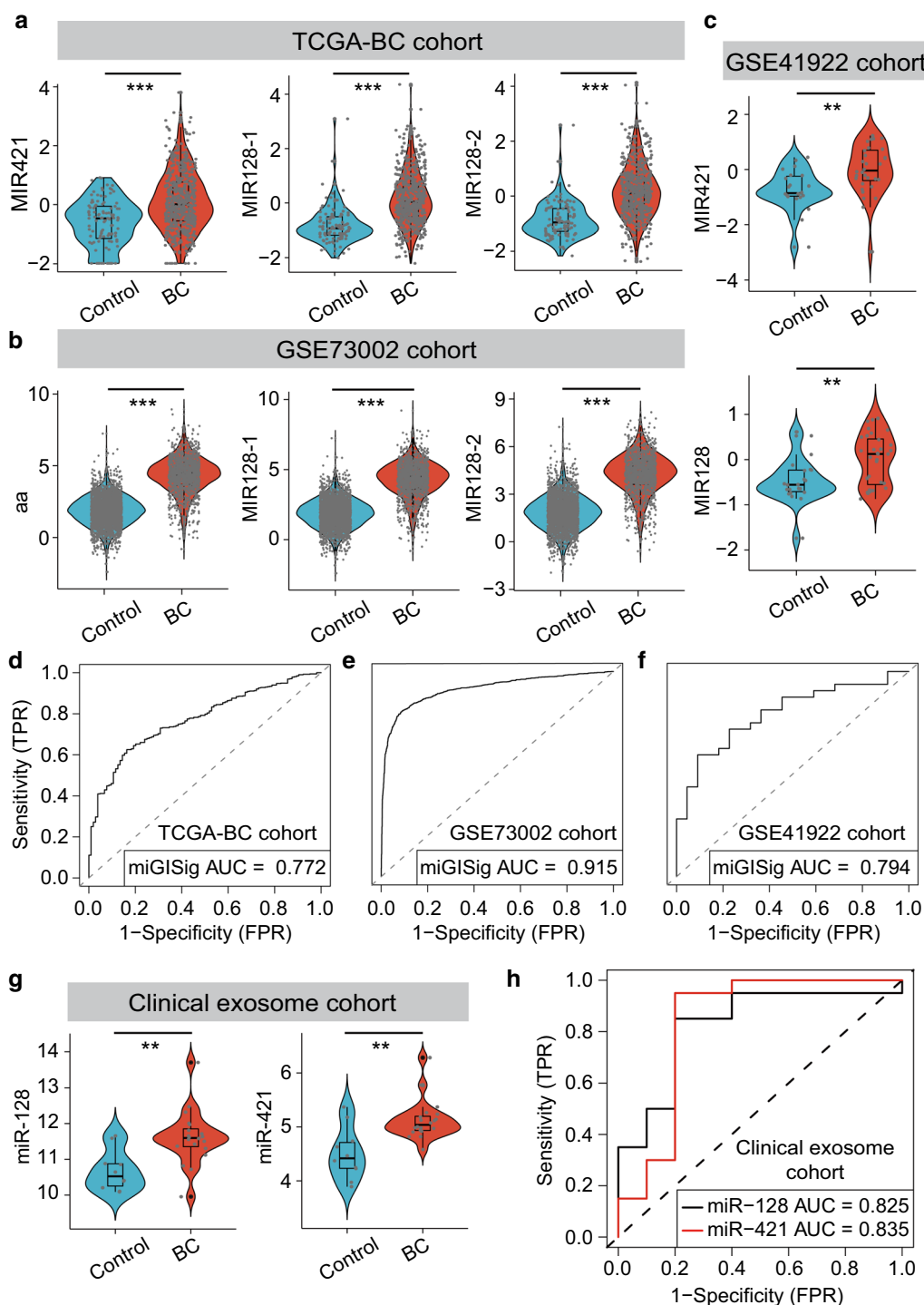


Fig. 4 Value of miGSig for the diagnosis of BC. Violin diagram of the miGSig miRNAs expression level in healthy controls and BC patients in TCGA-BC cohort (**a**), GSE73002 cohort (**b**) and GSE41922 cohort (**c**). Statistical analysis was performed using the Mann–Whitney U test. ROC curve for the performance of the miGSig in the diagnosis in the TCGA-BC cohort (**d**), GSE73002 cohort (**e**) and GSE41922 cohort (**f**). **g** Violin diagram of the miGSig miRNAs expression level in healthy controls and BC patients in the in-house clinical exosome cohort. Statistical analysis was performed using the Mann–Whitney U test. **h** ROC curve for the performance of *miR-128* and *miR-421* in the clinical exosome cohort. AUC the area under the curve, BC breast cancer, ROC area under the receiver operating characteristic curve. *P-value < 0.05, **P-value < 0.01, ***P-value < 0.001

Overexpression of the miGSig increases genomic instability by inducing an S-phase arrest and promotes the growth of breast cancer cells.

Since multi-nuclei and micro-nuclei are biomarkers of genotoxic and chromosomal instability [26, 27], we detected the frequency of multi-nuclei and micro-nuclei following overexpression of the three miRNAs. We first detected the expression level of *miR-128-1*, *miR-128-2*, and *miR-421* in BC cells and healthy human mammary epithelial cells, 184A1 and MCF-10A, and then selected the genomically unstable aneuploid, MDA-MB-231 cell

line with a low-level expression of the three miRNAs to perform the following experiments (Fig. 5a). miRNAs mimics were transfected into MDA-MB-231 cells and cultured for 48 h, and the expression level of *miR-128-1*, *miR-128-2* and *miR-421* was markedly increased (Fig. 5b). Using a high-content system, we showed that overexpression of *miR-128-1*, *miR-128-2* and *miR-421* spontaneously increased the proportion of multi-nuclei and micro-nuclei in the MDA-MB-231 cell line (25, 31 and 22% in the *miR-128-1*, *miR-128-2* and *miR-421* groups, respectively, vs. 18% in the control group)

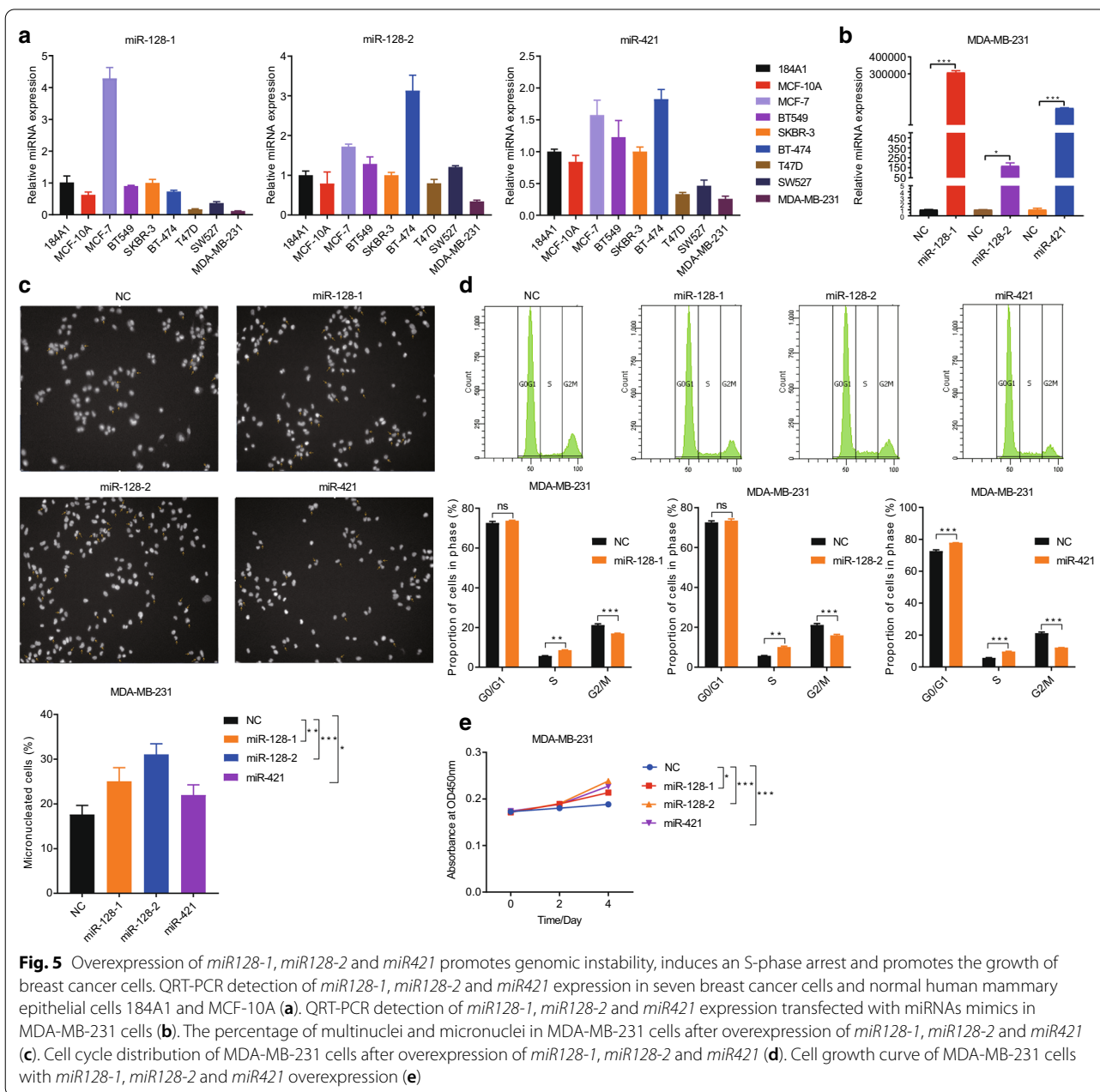
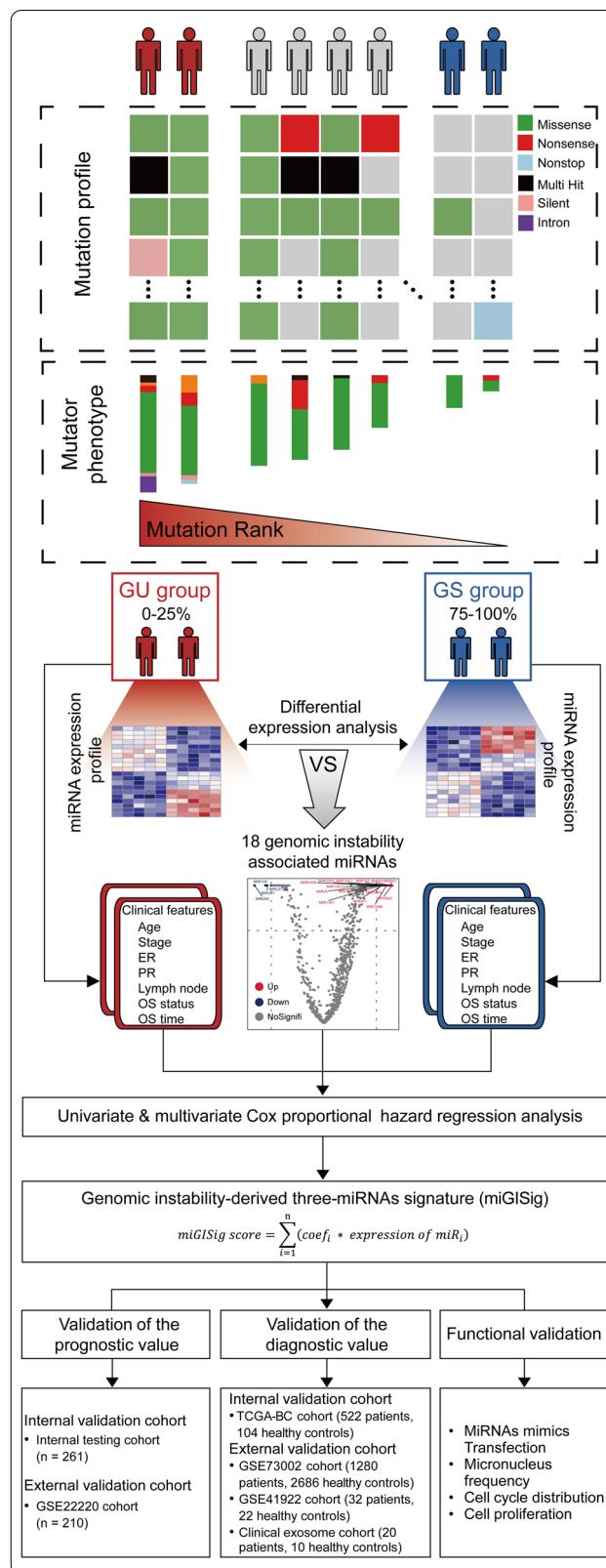


Fig. 6 Study flowchart. The study was performed in multicenter cohorts, including TCGA-BC, GSE22220, GSE73002, GSE41922, and in-house clinical exosome cohorts. Genome instability-derived miRNAs signature (miGISig) was identified in the discovery cohort. The miGISig was applied to an internal validation cohort and multiple external validation cohorts to verify its value in prognosis and diagnosis of BC. The effects of the miGISig on BC growth were investigated using the in vitro functional assays. BC breast cancer, GI Genome instability

(Fig. 5c). Since GI can result from defects in cell cycle regulation [28–33], we examined whether the overexpression *miR-128-1*, *miR-128-2* and *miR-421* lead to the aberrant cell cycle progression. FACS analysis showed that the proportion of cells in the S phase increased, while the number of cells in G₂/M phase decreased following the overexpression of the three miRNAs using mimics in the MDA-MB-231 and MCF-7 cell lines (Fig. 5d and Additional file 1: Figure S3A and S3B). Consistent with these results, the CCK-8 assay demonstrated that the proliferative ability of the MDA-MB-231 and MCF-7 cell lines was significantly increased following the overexpression of *miR-128-1*, *miR-128-2* and *miR-421* (Fig. 5e and Additional file 1: Figure S3C). These results demonstrated that the miGISig was associated with GI and could serve as oncogenes to promote the growth of breast cancer cells.

Discussion

BC is the most frequently diagnosed cancer and the leading cause of cancer-associated death in women. Considerable efforts have been made in the development of precision medicine for BC; however, its early diagnosis and prognosis risk stratification, using a minimally invasive method is still a clinical challenge. In this study, we focused on miRNAs involved in GI, as evaluated by expression alteration and tumor mutator phenotype and identified 18 candidate GI-related miRNAs. Target genes of the 18 candidate GI-related miRNAs were enriched in known GI pathways. For example, TGF-β maintains genomic stability by enhancing the DNA damage response [34]. Activation of the PI3K/Akt pathway can support tumor growth and progression, and can lead to inhibition of DNA repair, which may contribute to GI. Recent studies have indicated that sustained mitogen-activated protein kinase signaling relaxes the cell cycle checkpoints and allows cells to escape prolonged G₂ arrest by inducing the accumulation of the pro-mitotic kinase, thereby enhancing GI [35]. DNA damage response (DDR) assures the maintenance of GI, while a recent report has shown that Notch is a direct negative regulator of DDR.



Furthermore, the 18 candidate GI-related miRNAs can differentiate BC patients into two clusters with significantly different prognosis, *UBQLN4* expression and aneuploidy score. A recent study indicated that *UBQLN4* is an essential driver of GI and its overexpression represses homologous recombination-mediated double-strand break repair in aggressive tumors [36]. The aneuploidy score reflects the total burden of arm-level copy-number alterations and is an indicator of chromosome instability [37, 38]. These results provided evidence to support the relevance of the 18 miRNAs in GI. The identification of these miRNAs involved in GI will not only enhance our understanding of the biology of GI, but it will also provide new candidates for early diagnosis and prognosis of BC.

By focusing on these miRNAs involved in GI, we identified three oncogenic miRNAs (*MIR421*, *MIR128-1* and *MIR128-2*), with the independent prognostic value from the list of 18 candidate GI-related miRNAs and produced a three-miRNA signature, termed miGISig, which is predictive of GI and clinical outcome. The predictive value of miGISig was validated by its association with clinical outcome and GI in TCGA and other public BC datasets across different platforms. Furthermore, the miGISig is not only an independent predictor, but also exerts superior or comparable performance to other clinical factors. We further leveraged the complementary value of molecular and clinical characteristics and showed that combining the miGISig with several complementary clinical factors could provide a more accurate estimation of prognosis in BC. For the miGISig identified here, *MIR421* has been reported to cause the failure of DNA repair by suppressing the expression of ataxia-telangiectasia mutated, a core component of the DNA repair system, thus enhancing GI [39]. Furthermore, the overexpression of *MIR421* has been previously shown to promote cancer cell proliferation in BC and in non-small cell lung cancer [40, 41]. *MIR128-1* and *MIR128-2*, members of the miRNA *MIR128* family, encoded the same mature *miR-128* whose expression pattern and roles in tumorigenesis and development vary in different types of cancer. However, little is known about their association with GI in BC. Our *in vitro* functional assays suggested that overexpression of the miGISig increases GI, induces an S-phase arrest and promotes the growth of breast cancer cells. Many studies have suggested that aberrant miRNA expression is not only involved in cancer prognosis, but is also an early event in tumorigenesis [42–44]. In addition, miRNA levels have high stability and activity, which enables their detection in the serum, plasma, and other body fluids, using RT-qPCR, miRNA microarrays and deep sequencing techniques [20], highlighting the superiority of circulating miRNAs, as minimally invasive biomarkers in cancers. However, the mechanisms and biological

impact of circulating miRNAs still remain unknown. Exosomes are cell-driven vesicles, which carry nucleic acids, proteins, lipids, and metabolites from their host cells [45]. Exosomes can be isolated from readily available biological fluids, such as blood and urine, and this minimally invasive advantage presents an attractive novel biomarker for diagnostic applications [46]. For the miRNA markers identified here, little is known with respect to their potential for minimally invasive diagnosis of BC. Therefore, we integrated circulating exosomal miRNA profiles from TCGA and GEO tissue datasets and found that the miRNAs in the miGISig were all upregulated in both BC tissues and circulating exosomes when compared with normal breast tissues or circulating exosomes from healthy subjects. Furthermore, the miGISig revealed higher efficacy and stability in distinguishing BC from healthy subjects. Collectively, these results suggested that the miGISig was a GI-derived oncogenic maker that represents a promising minimally invasive clinical genomic tool for BC detection and prognosis; however, additional cohorts are required to validate these findings. Furthermore, the miGISig identified in the EVs provided further evidence supporting exosome-mediated delivery of oncogenic miRNAs is associated with BC carcinogenesis and prognosis. From a therapeutic perspective, identification of oncogenic GI-associated miRNAs in EVs also implied potential therapeutic strategy via transfecting anti-miRNA compounds or miRNA-based agents into exosomes to interfere with the load or delivery of these oncogenic exo-miRNAs for targeting genomic instability as a multimodality treatment strategy of BC.

Conclusions

In summary, using genome-wide miRNA expression profiles from the tumor, healthy tissues and circulating exosomes, our study indicated the clinical value of GI-associated miRNAs and established a GI-derived three-miRNA signature allowing early detection and prognostic risk stratification with minimal invasion for BC. Following further investigation in prospective cohort studies, our study highlighted the potential clinical value of the EVs-derived GI-associated miRNAs, as a minimally invasive genomic tool to improve BC precision medicine.

Methods

Study design

The overall research design was illustrated in Fig. 6. The study was performed using multicenter prospective cohorts, including The Cancer Genome Atlas (TCGA)-BC, GSE22220, GSE73002, GSE41922, and in-house clinical exosome cohorts. miGISig was identified in the discovery cohort. The miGISig was then applied to an internal validation cohort and multiple external

validation cohorts to verify its value in the prognosis and diagnosis of BC. The effects of the miGISig on BC growth were investigated using in vitro functional assays.

Public RNA-sequencing (Seq), microarray and clinical data collection

Clinical characteristics, miRNA-seq [Illumina HiSeq reads per million (RPM) type] expression data, RNA-seq (Illumina Hiseq fragments per kilobase of transcript per million type) expression data and somatic mutation information from patients with breast tumors and ovarian cancers (OV) were obtained from TCGA Genomic Data Commons Data Portal (<https://portal.gdc.cancer.gov/>). Only data from female patients, including paired miRNA, mRNA expression profiles, survival information and somatic mutation information, were retained. A total of 522 BC samples and 355 OV samples, using the aforementioned information, were retained for further investigation. The non-TCGA BC miRNAs microarray datasets used in the validation phase were downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>), including GSE22220 (n=210) [47], GSE73002 (n=3966) [48] and GSE41922 (n=54) [49]. Public BC miRNA datasets used in this study were listed in Additional file 1: Tables S1 and S2.

Plasma exosomal dataset

A total of 30 subjects, including 20 BC patients and 10 age-matched healthy women were enrolled from the Cancer Hospital, Chinese Academy of Medical Sciences (CHCAMS) for this study. Peripheral blood samples (10 ml) from these 30 subjects were collected for exosome isolation, which was confirmed by nanoparticle tracking analysis (NTA), transmission electron microscopy (TEM), and western blot analysis according to previously reported protocols [50]. Small RNA-seq was performed using the Illumina HiSeq 2000 platform according to the manufacturer’s protocol and miRNA expression levels were calculated using the reads per million (RPM) values. This study was approved by the Ethics Committee of the CHCAMS and written informed consent was provided by all the participants.

Establishment of a miGISig

The GI for each TCGA sample was measured using the cumulative somatic point mutation (CSPM) burden within a cancer genome. Tumors with a high CSPM burden (ranked within the top 25%) were defined as the genomically unstable (GU)-like group and those with a low CSPM burden (within the last 25%) were defined as the genomically stable (GS)-like group. The differentially expressed miRNAs between GU-like and GS-like groups were defined as candidate GI-associated miRNAs

(GI miRs). Prognostic GI miRs were identified using the univariate Cox regression analysis for overall survival (OS) time. Finally, the miGISig was constructed as follows:

$$miGISig = \sum_{i=1}^N expression\ of\ miRNA_i * Coefficient_i$$

where *N* is the number of prognostic GI miRs, expression of miRNA_{*i*} is the expression value of prognostic GI miR_{*i*}, and Coefficient_{*i*} is the estimated regression coefficient of GI miR_{*i*} in the multivariable Cox regression analysis.

The optimal cut-off value for risk stratification was determined using the point representing the 100% true-positive rate and 0% false-positive rate in the receiver operating characteristics (ROC) curve, in the discovery cohort.

Cell culture and transfection

The MCF-7, BT549, SKBR-3, BT474, T47D, SW527 and MDA-MB-231 breast cancer cell lines, and the 184A1 and MCF-10A normal human mammary epithelial cell lines were purchased from the American Type Culture Collection (ATCC; Rockville, MD, USA). The 184A1 and MCF-10A cell lines were maintained in DMEM-F12 containing 2 mM L-glutamine, 20 ng/ml EGF, 100 ng/ml cholera toxin, 0.01 mg/ml insulin, 500 ng/ml hydrocortisone and 5% horse serum (HyClone, USA). The MCF-7 cells were maintained in DMEM containing 2 mM L-glutamine, 1 mM sodium pyruvate, 10 mM HEPES and 10% FBS (Gibco, USA). The BT549 and SKBR-3 cells were cultured in RPMI-1640 supplemented with 10% FBS, while the BT474 and T47D cells were maintained in RPMI-1640 containing 2 mM L-glutamine, 4.5 g/l glucose, 1.5 g/l sodium bicarbonate, 1 mM sodium pyruvate, 0.01 mg/ml insulin, 10 mM HEPES and 5% FBS. The SW527 cells were maintained in DMEM containing 2 mM L-glutamine, 4.5 g/l glucose, 1.5 g/l sodium bicarbonate, and 10% FBS. All the cell lines were maintained at 37 °C in a humidified atmosphere with 5% CO₂, while the MDA-MB-231 cells were maintained in Leibovitz’s L-15 medium with 10% FBS at 37 °C.

miRNA mimics transfection

Sequences of *miR-128-1-5p*, *miR-128-2-5p* and *miR-421* were obtained from the miRBase database. The miRNAs mimics and negative controls were designed and synthesized from RiboBio (Guangzhou, China). The primer sequences of the miRNAs and controls were listed in Additional file 1: Table S3. Transfections for miRNAs mimics were performed using HiperFect (Qiagen, Valencia, CA, USA) according to the manufacturers’ instructions.

Reverse transcription-quantitative PCR (RT-qPCR)

Total RNA was isolated using the TRIzol[®] reagent (Thermo Scientific, Grand Island, NY, USA) and then reverse transcribed into cDNA using a Quantscript RT kit (Tiangen, Beijing, China). The RT-qPCR analysis was performed using a StepOnePlus Real-Time PCR system (Applied Biosystems, Foster City, CA, USA) according to standard procedures. The relative expression levels of the miRNA were normalized to that of U6, which served as an endogenous control.

Cell proliferation assays

The MCF-7 and MDA-MB-231 cell lines were plated into 96-well microplates at 1×10^3 cells per well. At the indicated time points, the viability of the cells was determined using a Cell Counting Kit-8 (CCK8, Dojindo) and measured at 450 nm, with the BioTek Gen5 system (Bio-Tek, USA). The experiments were repeated three times, with the representative experiment shown in the figures.

Flow cytometry (FACS) analysis

The MCF-7 and MDA-MB-231 cell lines were plated into 6-well microplates, and cells at 40–50% confluency were transfected with miRNAs mimics and cultured for 48 h (h). The collected cells were incubated with the cell cycle kit (Beckman Coulter, Pasadena, CA, USA) according to the manufacturer's instructions and the results were analyzed using flow cytometry (Beckman Coulter, Pasadena, CA, USA).

Immunofluorescence staining

Cells were seeded into 6-well microplates, and cells at 40–50% confluency were transfected with miRNAs mimics and cultured for 24 h. Then, the cells were re-seeded in a plate and cultured for 24 h. Cells were fixed with 4% paraformaldehyde and permeabilized with Triton X-100 (Sigma-Aldrich, St. Louis, MO, USA), and stained with DAPI (Shanghai Yeasen Biotechnology Co. Ltd. Shanghai, China) for 5 min. Then, the cells were imaged using high content analysis (HCA, Thermo Scientific, USA).

Construction of an integrated prognostic index

Multivariate cox regression analysis was applied to construct an integrated prognostic index by combining age, stage and miGISig. The median score was used for the cut-off value for the integrated prognostic index. Time-dependent ROC curve was used to compare the prognostic performance among age, stage, miGISig and the integrated prognostic index.

Statistical analysis

All statistical analyses were performed using the R software (V.3.4.4) with the following packages: 'samr', 'ggpubr', 'survival', 'survminer', 'timeROC' and 'ROCit'. The Mann–Whitney U test was used to compare continuous variables. Survival analysis was conducted using the Kaplan–Meier method and the log-rank test. Univariate and multivariate analyses with Cox proportional hazards regression for OS time were performed on the individual variables by calculating hazard ratios (HR) and 95% confidence intervals (CI). Differential expression analysis was performed using the Significance Analysis of Microarrays method, and fold change > 1.5 or < 0.67 and false discovery rate adjusted $P < 0.05$ was considered significant. The time-dependent ROC curve was calculated with the nearest neighbor estimation method. The unsupervised hierarchical clustering analysis was performed with Euclidean distance for determining the similarity between patients and the ward's linkage method for merging similar objects.

Target genes and functional enrichment analysis of miRNAs

Experimentally verified miRNA-target relationships were obtained from The Encyclopedia of RNA Interactomes (ENCORI, <http://starbase.sysu.edu.cn/>) [51]. Pearson correlation coefficients were computed to measure the correlation between miRNAs and mRNAs. Experimentally verified miRNA targets with correlation coefficient < -0.3 and P-value < 0.05 were selected as BC-specific miRNA targets. Functional enrichment analysis was conducted for BC-specific miRNA targets to determine significantly enriched biological processes (BP) terms in Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway using DAVID Bioinformatics Resources 6.8 (version 6.8; <https://david.ncifcrf.gov/>) [52].

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12951-020-00767-3>.

Additional file 1: Table S1. Public miRNAs datasets used in this study. **Table S2.** Clinicopathological characteristics of BC patients used in this study. **Table S3.** Primer sequences of three genomic instability-related oncogenic miRNAs and controls in this study. **Table S4.** Lists of 18 genomic instability-related miRNAs in BC. **Table S5.** Univariate Cox regression analyses of three genomic instability-related miRNAs associated with overall survival in BC. **Table S6.** Comparison of clinical characteristics between miGISig-derived high-risk and low-risk groups. **Figure S1.** Multivariate analysis of the miGISig with clinical characteristics in different cohorts. **Figure S2.** Value of miGISig for early diagnosis of BC. **Figure S3.** Overexpression of *miR128-1*, *miR128-2* and *miR421* induces a S-phase arrest and promotes the proliferative ability of MCF-7 cells.

Abbreviations

BC: Breast cancer; CI: Confidence intervals; CSPM: Cumulative somatic point mutation; ENCORI: Encyclopedia of RNA interactomes; ER: Estrogen receptor; EVs: Extracellular vesicles; GEO: Gene Expression Omnibus; GI: Genomic instability; GO: Gene Ontology; GU: Genomically unstable; GS: Genomically stable; miRNA: MicroRNA; HER2: Human epidermal growth factor receptor 2; PR: Progesterone receptor; TCGA: The Cancer Genome Atlas; KEGG: Kyoto Encyclopedia of Genes and Genomes; OS: Overall survival; OV: Ovarian cancers; RPM: Reads per million; ROC: Receiver operating characteristics.

Authors' contributions

MZ, HYC and NW conceived and designed the study. SQB, MZ, JZS, JS and YM contributed to the data collection, statistical analysis and data interpretation. JQL, SQB and JXL contributed to exosome collection and interpretation. TH and HYC contributed to biological experiment and interpretation. MZ, SQB, HYC and JQL wrote this manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by Zhejiang Provincial Natural Science Foundation of China under Grant No. LY21C060004 to M.Z.; and Beijing Natural Science Foundation (JQ20032) to N.W., National Natural Science Foundation of China (81822030 and 82072391) to N.W.; and National Natural Science Foundation of China (81802669) to J.L., the CAMS Innovation Fund for Medical Sciences (2020-I2M-C&T-B-068) to J.L., and the CAMS Initiative Fund for Medical Sciences (2016-I2M-1-001); and CAMS Innovation Fund for Medical Sciences (CIFMS 2020-I2M-C&T-A-015) to Y.M., the PUMC Youth Fund & the Fundamental Research Funds for the Central Universities (No.3332019052) to Y.M.). The funders had no roles in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The dataset analyzed during the current study are available in the TCGA (<https://portal.gdc.cancer.gov/>) and GEO databases (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22220>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE73002>, <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE41922>).

Ethics approval and consent to participate

This study was approved by the Ethics Committee of the CHCAMS and written informed consent was provided by all the participants.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ School of Biomedical Engineering, School of Ophthalmology & Optometry and Eye Hospital, Wenzhou Medical University, Wenzhou 325027, People's Republic of China. ² State Key Laboratory of Molecular Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, People's Republic of China. ³ Department of Breast Surgical Oncology, National Cancer Center/National Clinical Research Center for Cancer/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, People's Republic of China. ⁴ PET-CT Center, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, People's Republic of China. ⁵ Department of Orthopedic Surgery, Beijing Key Laboratory for Genetic Research of Skeletal Deformity & Key Laboratory of Big Data for Spinal Deformities, State Key Laboratory of Complex Severe and Rare Diseases, All at Peking Union Medical College Hospital, Peking Union Medical College and Chinese Academy of Medical Sciences, Beijing 100730, People's Republic of China.

Received: 8 November 2020 Accepted: 30 December 2020
Published online: 12 January 2021

References

- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin.* 2019;69(1):7–34 (Epub 2019/01/09).
- DeSantis CE, Ma J, Gaudet MM, Newman LA, Miller KD, Goding Sauer A, et al. Breast cancer statistics, 2019. *CA Cancer J Clin.* 2019;69(6):438–51 (Epub 2019/10/03).
- Dawson SJ, Tsui DW, Murtaza M, Biggs H, Rueda OM, Chin SF, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N Engl J Med.* 2013;368(13):199–209 (Epub 2013/03/15).
- Bidard FC, Hajage D, Bachelot T, Delaloge S, Brain E, Campone M, et al. Assessment of circulating tumor cells and serum markers for progression-free survival prediction in metastatic breast cancer: a prospective observational study. *Breast Cancer Res.* 2012;14(1):R29 (Epub 2012/02/15).
- Harris L, Fritzsche H, Mennel R, Norton L, Ravdin P, Taube S, et al. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol.* 2007;25(33):5287–312 (Epub 2007/10/24).
- Waks AG, Winer EP. Breast cancer treatment: a review. *JAMA.* 2019;321(3):288–300 (Epub 2019/01/23).
- Ahmad SS, Ahmed K, Venkataraman AR. Science in focus: genomic instability and its implications for clinical cancer care. *Clin Oncol (R Coll Radiol).* 2018;30(12):751–5 (Epub 2018/10/03).
- Shen Z. Genomic instability and cancer: an introduction. *J Mol Cell Biol.* 2011;3(1):1–3 (Epub 2011/02/01).
- Negrini S, Gorgoulis VG, Halazonetis TD. Genomic instability—an evolving hallmark of cancer. *Nat Rev Mol Cell Biol.* 2010;11(3):220–8 (Epub 2010/02/24).
- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011;144(5):646–74 (Epub 2011/03/08).
- Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012;490(7418):61–70 (Epub 2012/09/25).
- Duijff PH, Schultz N, Benezra R. Cancer cells preferentially lose small chromosomes. *Int J Cancer.* 2013;132(10):2316–26 (Epub 2012/11/06).
- Duijff PHG, Nanayakkara D, Nones K, Srihari S, Kalimutho M, Khanna KK. Mechanisms of genomic instability in breast cancer. *Trends Mol Med.* 2019;25(7):595–611 (Epub 2019/05/13).
- Kalimutho M, Nones K, Srihari S, Duijff PHG, Waddell N, Khanna KK. Patterns of genomic instability in breast cancer. *Trends Pharmacol Sci.* 2019;40(3):198–211 (Epub 2019/02/10).
- Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell.* 2004;116(2):281–97 (Epub 2004/01/28).
- Lee RC, Feinbaum RL, Ambros V. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell.* 1993;75(5):843–54 (Epub 1993/12/03).
- Wightman B, Ha I, Ruvkun G. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in *C. elegans*. *Cell.* 1993;75(5):855–62 (Epub 1993/12/03).
- Krol J, Loedige I, Filipowicz W. The widespread regulation of microRNA biogenesis, function and decay. *Nat Rev Genet.* 2010;11(9):597–610 (Epub 2010/07/28).
- Gebert LFR, MacRae IJ. Regulation of microRNA function in animals. *Nat Rev Mol Cell Biol.* 2019;20(1):21–37 (Epub 2018/08/16).
- Cortez MA, Bueso-Ramos C, Ferdin J, Lopez-Berestein G, Sood AK, Calin GA. MicroRNAs in body fluids—the mix of hormones and biomarkers. *Nat Rev Clin Oncol.* 2011;8(8):467–77 (Epub 2011/06/08).
- Ling H, Fabbri M, Calin GA. MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nat Rev Drug Discov.* 2013;12(11):847–65 (Epub 2013/11/01).
- Ling H, Zhang W, Calin GA. Principles of microRNA involvement in human cancers. *Chin J Cancer.* 2011;30(11):739–48 (Epub 2011/11/01).
- Pichler M, Ress AL, Winter E, Stiegelbauer V, Karbiener M, Schwarzenbacher D, et al. miR-200a regulates epithelial to mesenchymal transition-related gene expression and determines prognosis in colorectal cancer patients. *Br J Cancer.* 2014;110(6):1614–21 (Epub 2014/02/08).
- Zhang X, Wan G, Mlotshwa S, Vance V, Berger FG, Chen H, et al. Oncogenic Wip1 phosphatase is inhibited by miR-16 in the DNA damage signaling pathway. *Cancer Res.* 2010;70(18):7176–86 (Epub 2010/07/30).
- Taylor DD, Gercel-Taylor C. The origin, function, and diagnostic potential of RNA within extracellular vesicles present in human biological fluids. *Front Genet.* 2013;4:142 (Epub 2013/08/03).

26. Holland AJ, Cleveland DW. Chromoanagenesis and cancer: mechanisms and consequences of localized, complex chromosomal rearrangements. *Nat Med*. 2012;18(11):1630–8 (Epub 2012/11/09).
27. Podrimaj-Bytyqi A, Borovecki A, Selimi Q, Manxhuka-Kerliu S, Gashi G, Elezaj IR. The frequencies of micronuclei, nucleoplasmic bridges and nuclear buds as biomarkers of genomic instability in patients with urothelial cell carcinoma. *Sci Rep*. 2018;8(1):173 (Epub 2018/12/16).
28. Sieber OM, Heinimann K, Tomlinson IP. Genomic instability—the engine of tumorigenesis? *Nat Rev Cancer*. 2003;3(9):701–8 (Epub 2003/09/03).
29. Vincent K, Pichler M, Lee GW, Ling H. MicroRNAs, genomic instability and cancer. *Int J Mol Sci*. 2014;15(8):14475–91 (Epub 2014/08/21).
30. Malumbres M, Barbacid M. Cell cycle, CDKs and cancer: a changing paradigm. *Nat Rev Cancer*. 2009;9(3):153–66 (Epub 2009/02/25).
31. Tian RQ, Wang XH, Hou LJ, Jia WH, Yang Q, Li YX, et al. MicroRNA-372 is down-regulated and targets cyclin-dependent kinase 2 (CDK2) and cyclin A1 in human cervical cancer, which may contribute to tumorigenesis. *J Biol Chem*. 2011;286(29):25556–63 (Epub 2011/06/08).
32. Voorhoeve PM, le Sage C, Schrier M, Gillis AJ, Stoop H, Nagel R, et al. A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. *Adv Exp Med Biol*. 2007;604:17–46 (Epub 2007/08/19).
33. Lobrich M, Jeggo PA. The impact of a negligent G2/M checkpoint on genomic instability and cancer induction. *Nat Rev Cancer*. 2007;7(11):861–9 (Epub 2007/10/19).
34. Barcellos-Hoff MH, Cucinotta FA. New tricks for an old fox: impact of TGFbeta on the DNA damage response and genomic stability. *Sci Signal*. 2014;7(341):5 (Epub 2014/09/04).
35. De S, Campbell C, Venkitaraman AR, Esposito A. Pulsatile MAPK Signaling Modulates p53 activity to control cell fate decisions at the G2 checkpoint for DNA damage. *Cell Rep*. 2020;30(7):2083–93 e5 (Epub 2020/02/23).
36. Jachimowicz RD, Beleggia F, Isensee J, Velpula BB, Goergens J, Bustos MA, et al. UBQLN4 represses homologous recombination and is overexpressed in aggressive tumors. *Cell*. 2019;176(3):505–19 e22 (Epub 2019/01/08).
37. Ben-David U, Amon A. Context is everything: aneuploidy in cancer. *Nat Rev Genet*. 2020;21(1):44–62 (Epub 2019/09/25).
38. Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell*. 2018;33(4):676–89 e3 (Epub 2018/04/07).
39. Hu H, Du L, Nagabayashi G, Seeger RC, Gatti RA. ATM is down-regulated by N-Myc-regulated microRNA-421. *Proc Natl Acad Sci U S A*. 2010;107(4):1506–11 (Epub 2010/01/19).
40. Wang Y, Liu Z, Shen J. MicroRNA-421-targeted PDCD4 regulates breast cancer cell proliferation. *Int J Mol Med*. 2019;43(1):267–75 (Epub 2018/10/27).
41. Li X, Chen SH, Zeng JW. MiR-421 Is Overexpressed and promotes cell proliferation in non-small cell lung cancer. *Med Princ Pract*. 2020;29(1):80–9 (Epub 2019/09/02).
42. Rupaimoole R, Slack FJ. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. *Nat Rev Drug Discov*. 2017;16(3):203–22 (Epub 2017/02/18).
43. Lin S, Gregory RI. MicroRNA biogenesis pathways in cancer. *Nat Rev Cancer*. 2015;15(6):321–33 (Epub 2015/05/23).
44. Peng Y, Croce CM. The role of MicroRNAs in human cancer. *Signal Transduct Target Ther*. 2016;1:15004 (Epub 2016/01/28).
45. Kalluri R, LeBleu VS. The biology, function, and biomedical applications of exosomes. *Science*. 2020. <https://doi.org/10.1126/science.aau6977> (Epub 2020/02/08).
46. Sun Z, Shi K, Yang S, Liu J, Zhou Q, Wang G, et al. Effect of exosomal miRNA on cancer biology and clinical applications. *Mol Cancer*. 2018;17(1):147 (Epub 2018/10/13).
47. Buffa FM, Camps C, Winchester L, Snell CE, Gee HE, Sheldon H, et al. microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer. *Cancer Res*. 2011;71(17):5635–45 (Epub 2011/07/09).
48. Shimomura A, Shiino S, Kawauchi J, Takizawa S, Sakamoto H, Matsuzaki J, et al. Novel combination of serum microRNA for detecting breast cancer in the early stage. *Cancer Sci*. 2016;107(3):326–34 (Epub 2016/01/11).
49. Chan M, Liaw CS, Ji SM, Tan HH, Wong CY, Thike AA, et al. Identification of circulating microRNA signatures for breast cancer detection. *Clin Cancer Res*. 2013;19(16):4477–87 (Epub 2013/06/26).
50. Min L, Zhu S, Chen L, Liu X, Wei R, Zhao L, et al. Evaluation of circulating small extracellular vesicles derived miRNAs as biomarkers of early colon cancer: a comparison with plasma total miRNAs. *J Extracell Vesicles*. 2019;8(1):1643670 (Epub 2019/08/27).
51. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res*. 2014;42(Database issue):D92–7 (Epub 2013/12/04).
52. da Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44–57 (Epub 2009/01/10).

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

