Breast Cancer Res Treat
DOI 10.1007/s10549-016-4013-7

CrossMark

PRECLINICAL STUDY

miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2178 breast cancer patients

András Lánczky[1] · Ádám Nagy[1,2] · Giulia Bottai[3] · Gyöngyi Munkácsy[1,4] · András Szabó[2] · Libero Santarpia[3] · Balázs Győrffy[1,2]

Received: 8 July 2016 / Accepted: 8 October 2016

Abstract
Purpose The proper validation of prognostic biomarkers is an important clinical issue in breast cancer research. MicroRNAs (miRNAs) have emerged as a new class of promising breast cancer biomarkers. In the present work, we developed an integrated online bioinformatic tool to validate the prognostic relevance of miRNAs in breast cancer.
Methods A database was set up by searching the GEO, EGA, TCGA, and PubMed repositories to identify datasets with published miRNA expression and clinical data. Kaplan–Meier survival analysis was performed to validate the prognostic value of a set of 41 previously published survival-associated miRNAs.
Results All together 2178 samples from four independent datasets were integrated into the system including the expression of 1052 distinct human miRNAs. In addition, the web-tool allows for the selection of patients, which can be filtered by receptors status, lymph node involvement, histological grade, and treatments. The complete analysis tool can be accessed online at: www.kmplot.com/mirpower. We used this tool to analyze a large number of deregulated miRNAs associated with breast cancer features and outcome, and confirmed the prognostic value of 26 miRNAs. A significant correlation in three out of four datasets was validated only for *miR-29c* and *miR-101*.
Conclusions In summary, we established an integrated platform capable to mine all available miRNA data to perform a survival analysis for the identification and validation of prognostic miRNA markers in breast cancer.

Keywords Breast cancer · Biomarkers · MicroRNAs · Gene expression · Prognosis · Survival

Introduction

Breast cancer represents the most frequent malignancy, and is still a leading cause of cancer-related death in women worldwide [1]. Distinct histopathological features are routinely used as prognostic and predictive markers, ultimately driving clinical treatment decisions [2]. However, these characteristics are not able to capture the heterogeneity of breast cancer and to accurately predict patient outcome [3]. Data derived from genome-wide studies provided novel insights into breast cancer complexity, leading to the refinement of the breast cancer molecular classification, and enabling a deeper understanding of the clinical course of the disease.

MicroRNAs (miRNAs) are small non-coding RNA molecules regulating gene expression and widely influencing pathways associated with tumor development, progression, and response to therapy [4]. Several studies have

Electronic supplementary material The online version of this article (doi:10.1007/s10549-016-4013-7) contains supplementary material, which is available to authorized users.

✉ Libero Santarpia
libero.santarpia@humanitasresearch.it;
liberosantarpia@yahoo.it

✉ Balázs Győrffy
gyorffy.balazs@ttk.mta.hu

[1] MTA TTK Lendület Cancer Biomarker Research Group, Magyar Tudósok körútja 2, Budapest 1117, Hungary

[2] Department of Pediatrics, Semmelweis University, Budapest, Hungary

[3] Oncology Experimental Therapeutics Unit, Humanitas Clinical and Research Institute, Via Manzoni 113, 20089 Rozzano-Milan, Italy

[4] MTA-SE Pediatrics and Nephrology Research Group, Budapest, Hungary

Published online: 15 October 2016

demonstrated that miRNAs expression profiles could accurately classify molecular breast cancer subtypes and identify patients with different clinical outcome [5–8]. A large number of prognostic miRNAs for breast cancer have been described so far [9, 10]. However, there is an evident imbalance between the large amount of published candidate biomarkers and the reduced number of marker that have actually impacted clinical practice. The clinical and molecular heterogeneity of the breast cancer cohorts used in different studies, as well as methodological biases regarding reproducibility and standardization, have limited the identification of specific miRNAs as robust predictors of breast cancer patient outcome. Thus, to improve risk stratification and clinical decision making, the validation of the prognostic value of miRNAs in breast cancer is imperative.

In the present work, we aimed to structure a novel analytical web-tool based on the integration of miRNA expression and clinical data from different breast cancer datasets, and directed to validate the prognostic clinical relevance of miRNAs. In summary, this bioinformatic tool is able to perform a real-time analysis of published miRNA datasets in order to measure the power of miRNAs as predictor of survival in breast cancer patients.

Methods

MiRNA gene expression database

A database was established using miRNA expression data downloaded from gene expression omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/), the cancer genome Atlas (TCGA) (http://cancergenome.nih.gov/), European genome-phenome archive (EGA) (https://ega.crg.eu/), and PubMed (http://www.pubmed.com). For the database, the keywords "breast cancer," and "miRNA" or "microRNA" were used as the search terms. Only publications with available expression data, clinical survival information, and at least 50 breast cancer patients were included. All samples were checked using the ranked expression of all genes to identify repeatedly published microarrays. Four studies were identified that met our criteria [11–14]. Replicates were removed, and the published normalized expression data without a renormalization were used in the statistical computations. Detailed characteristics for each dataset are given in Online Resource 1. Only overall survival (OS) data were published for each of these datasets. Each dataset was processed separately.

Statistical analysis

For each analysis, the data were loaded into the R statistical environment, where calculations were performed. In case of missing data, the samples are excluded from the analysis (this is also the reason for the reduction in the sample number in case a filter is employed). The package "survival" is used to calculate and plot Kaplan–Meier survival curves, and the number-at-risk is indicated below the main plot. Hazard ratio (HR), 95 % confidence intervals (CI), and log-rank p values were calculated and displayed. Proportional hazard was computed by the "coxph" package. Statistical significance was set at $p \leq 0.05$. Bonferroni correction was executed for studies simultaneously publishing multiple miRNA biomarker candidates.

Online analysis interface

A web interface was set up to enable reproduction of computations in a platform-independent user interface. The entire dataset with clinical data is loaded into a PostgreSQL database which then enables immediate filtering of the data. The interactivity of the service is increased by the usage of JavaScript and Ajax technologies. The server is running on Debian Linux (www.debian.org) and is powered by Apache (www.apache.org). The server scripts were made in hypertext preprocessor (PHP), which controls both the analysis requests and delivers the results.

Identification of miRNAs associated with prognosis in breast cancer

A PubMed search was performed using the keywords "breast cancer", "miRNA", "microRNA", "overall survival", and "biomarker" to identify miRNAs described in the literature as potential prognostic biomarkers for breast cancer. Only studies published in English were considered. We uncovered 41 miRNAs associated with OS in breast cancer tissues (Online Resource 2). Then, using the original publications and PubMed gene, we added a unique gene symbol for each of the miRNAs and linked these to the corresponding probe IDs in each dataset. The capability of these genes to predict survival was measured by running the analysis in the online tool. The validation analysis was performed in each of the four cohorts separately. The median expression was used for splitting the patients into cohorts during the analysis.

Results

Database setup

We identified four studies meeting our criteria in the GEO, TCGA, EGA, and PubMed. These included 634 patients in the TCGA, 1262 patients in the Metabric, 181 patients in the GSE40267, and 101 patients in the GSE19783 [11–14]. Estrogen receptor (ER)-positive patients represent 72.1 % of

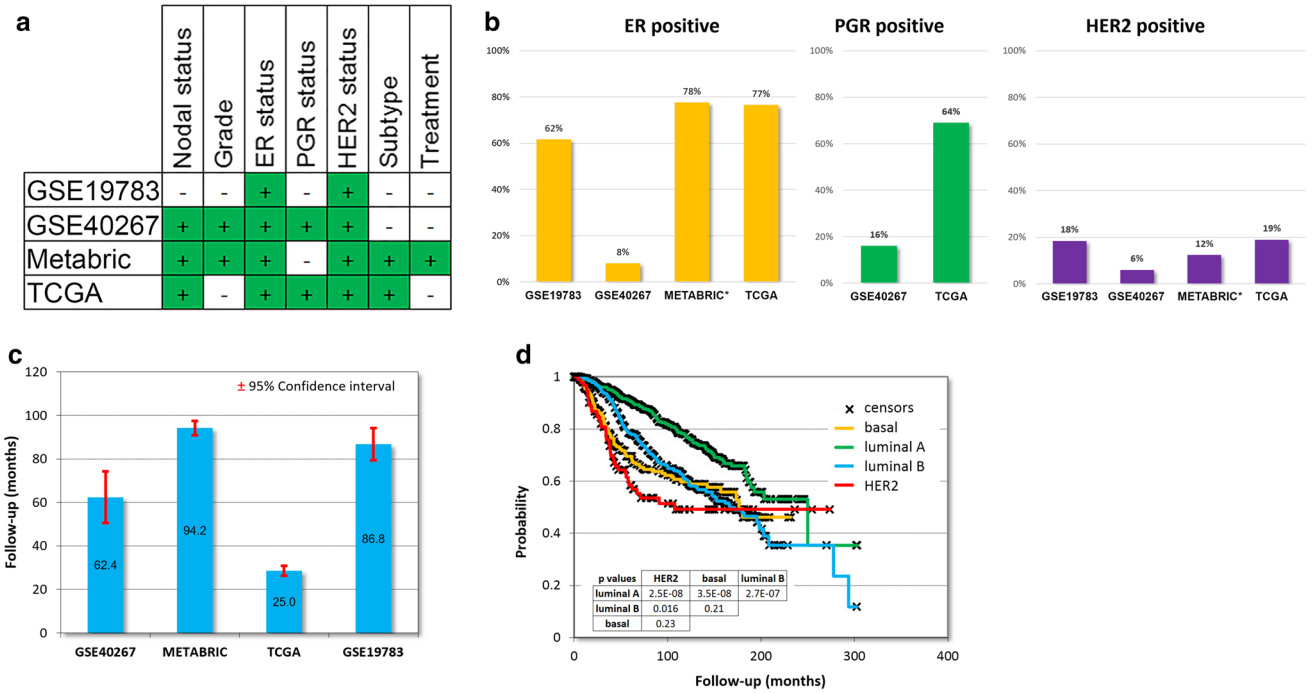

Fig. 1 Clinical characteristics of the samples included in the cohorts used in this study. **a** Availability of clinical data for each of the datasets. **b** Proportions of receptor status for ER, PGR, and HER2 in each dataset. Receptor status is based on immunohistochemistry, with the exception of the gene array-based Metabric (*). **c** Length of follow-up in each dataset. **d** Survival differences according to molecular subtype

all samples. GSE40267 has predominantly ER-negative samples (81.8 %; Online Resource 1). The TCGA dataset has a high proportion of progesterone receptor (PGR)-positive samples (64.0 %), while GSE40267 has only 16.0 % PGR-positive patients. HER2-positive patients account for 16.7 % of the entire database (5.5–16.8 % in the individual datasets; Online Resource 1). Only OS data were published for each of these datasets, and the mean follow-up is 71.6 months. However, individual datasets have a high difference in the length of follow-up: only 25.0 months for TCGA, 62.4 months for GSE40267, 86.8 months for GSE19783, and 94.2 months for the Metabric samples. An overview of the available clinical data is presented in Fig. 1. The Metabric dataset also had published detailed treatment information. Here, a set of patients were systemically untreated—these patients have neither received hormonal therapy nor chemotherapy nor radiotherapy ($n = 199$). When comparing clinical characteristics between the systemically untreated and treated patients, almost all the untreated patients were node negative (91.8 vs. 46.5 %, $p < 1E{-}16$).

The entire database contains 1052 distinct human miRNAs, of which 555 miRNAs are measured on one platform only, 141 miRNAs are measured on two, 148 miRNAs are measured on three, and 207 miRNAs are measured on each platform. Proportion of overlapping miRNAs among the four different platforms used ranged between 23.1 % (TCGA *v* GSE1973) and 75.7 % (GSE19673 *v* GSE40267) (Fig. 2; Online Resource 3).

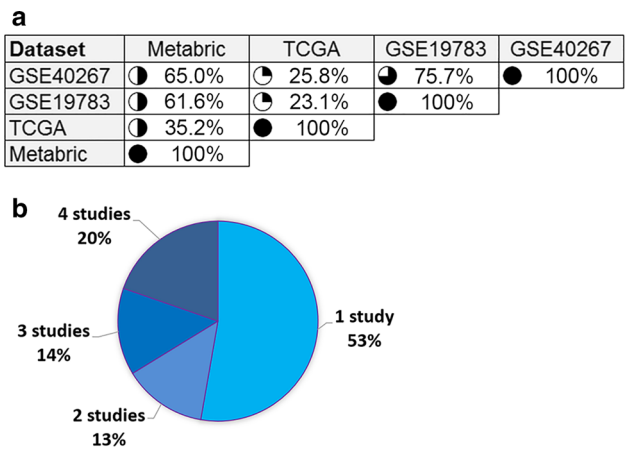

Fig. 2 Characteristics of overlapping miRNAs among different datasets. **a** The overlap of miRNAs measured in the four different studies. **b** Proportion of miRNAs measured by one, two, three, or all four studies

Online analysis interface

The entire computational pipeline with the associated databases is made accessible for reanalysis in an online accessible registration-free system. To measure the association between a queried miRNA and survival, the samples are grouped according to the median (or upper or lower quartile) expression of the selected miRNA, and then the two groups are compared by Cox proportional hazards regression, and a

Table 1 MiRNAs with validated prognostic values in breast cancer

MiRNA	Patients	Metabric					TCGA				
		n	HR	95 % CI	*p* value	*q* value	*n*	HR	95 % CI	*p* value	*q* value
let-7b	Het	1262	0.78	0.64–0.95	**1.3E−02**	7.5E−03	579	0.68	0.41–1.13	1.3E−01	0.10
let-7 g	Het	1262	0.79	0.65–0.96	**2.0E−02**	9.5E−03	579	0.92	0.56–1.53	7.6E−01	0.33
miR-10b	Het	1262	0.74	0.61–0.90	**3.0E−03**	3.2E−03	579	0.81	0.49–1.34	4.0E−01	0.22
miR-15a	TNBC	203	0.62	0.39–0.98	**3.9E−02**	1.6E−02	95	0.57	0.17–1.93	3.6E−01	0.22
miR-21	TNBC	203	1.03	0.66–1.61	8.9E−01	2.6E−01	95	0.73	0.22–2.42	6.0E−01	0.28
miR-22	Het	1262	1.3	1.0–1.5	**1.8E−02**	9.1E−03	579	2.05	1.22–3.45	**5.6E−03**	0.02
miR-27b	TNBC	203	0.95	0.61–1.48	8.2E−01	2.5E−01	95	1.64	0.48–5.65	4.2E−01	0.22
miR-29c	Het	1262	0.72	0.59–0.88	**1.2E−03**	1.5E−03	579	0.46	0.27–0.77	**2.5E−03**	0.01
miR-99a	LumA;ER+/HER2−	546	0.77	0.54–1.11	1.6E−01	5.7E−02	327	0.4	0.15–1.08	6.0E−02	0.05
	Het	1262	0.71	0.58–0.87	**8.0E−04**	1.2E−03	579	0.74	0.45–1.22	2.3E−01	0.16
miR-100	Het	1262	0.75	0.62–0.92	**5.2E−03**	4.3E−03	579	0.71	0.43–1.18	1.8E−01	0.13
miR-101	Het	1262	0.64	0.52–0.78	**1.0E−05**	4.1E−05	579	0.46	0.28–0.77	**2.4E−03**	0.01
miR-125b	LumA; ER+/HER2−	546	0.65	0.45-0.95	**2.4E−02**	1.1E−02	327	0.76	0.29–1.99	5.7E−01	0.28
miR-146a	TNBC	203	0.64	0.41–1.01	**5.1E−02**	1.9E−02	95	0.96	0.28–3.3	9.5E−01	0.37
miR-146	TNBC	203	0.61	0.39–0.96	**3.1E−02**	1.3E−02	95	0.88	0.27–2.93	8.4E−01	0.34
miR-155	TNBC	203	0.52	0.33–0.82	**4.7E−03**	4.3E−03	95	0.58	0.17–1.95	3.7E−01	0.22
miR-181a	TNBC	203	1.36	0.87-2.13	1.7E−01	6.0E−02	95	3.93	1.01–15.3	**3.3E−02**	0.03
miR-185	Het	1262	1.03	0.85–1.26	7.5E−01	2.4E−01	579	1.76	1.05–2.95	**3.0E−02**	0.03
miR-195	HER2−	1105	0.64	0.53–0.78	**1.1E−05**	4.1E−05	429	0.42	0.19–0.9	**2.1E−02**	0.03
miR-204	Het	1262	0.66	0.54–0.80	**2.9E−05**	7.3E−05	579	0.57	0.34–0.95	**2.7E−02**	0.03
miR-205	Het	1262	0.78	0.64–0.95	**1.3E−02**	7.6E−03	579	1.08	0.65–1.8	7.5E−01	0.33
miR-210	Het	1262	1.3	1.1–1.6	**9.5E−03**	6.5E−03	579	1.85	1.11–3.09	**1.7E−02**	0.03
miR-218	Het	1262	0.68	0.56–0.83	**2.0E−04**	3.8E−04	579	0.53	0.32–0.88	**1.2E−02**	0.03
miR-339-5p	Het	1262	0.82	0.67–0.99	**4.3E−02**	1.7E−02	–	–	–	–	–
miR-342-5p	Het	1262	0.76	0.62–0.92	**6.0E−03**	4.5E−03	–	–	–	–	–
miR-526b	Het	1262	1.03	0.84–1.25	7.9E−01	2.5E−01	579	1.59	0.93–2.73	8.8E−02	0.07
miR-1258	Het	–	–	–	–	–	579	0.55	0.32–0.9	**2.0E−02**	0.03

MiRNA	Patients	GSE40267					GSE19783				
		n	HR	95 % CI	*p* value	*q* value	*n*	HR	95 % CI	*p* value	*q* value
let-7b	Het	85	0.81	0.48–1.37	4.3E−01	0.25	93	0.68	0.31–1.52	3.5E−01	0.83
let-7 g	Het	85	1.11	0.66–1.88	6.9E−01	0.29	93	0.67	0.30–1.51	3.3E−01	0.83
miR-10b	Het	85	1.45	0.86–2.46	1.6E−01	0.13	93	0.87	0.39–1.94	7.4E−01	0.88
miR-15a	TNBC	53	0.94	0.49–1.79	8.4E−01	0.32	–	–	–	–	–
miR-21	TNBC	53	1.9	0.98–3.70	**5.3E−02**	0.09	–	–	–	–	–
miR-22	Het	85	0.85	0.50–1.44	5.4E−01	0.27	93	0.79	0.36–1.77	5.7E−01	0.84
miR-27b	TNBC	53	2.1	1.1–4.1	**2.9E−02**	0.07	–	–	–	–	
miR-29c	Het	85	0.57	0.33–0.96	**3.1E−02**	0.07	93	0.47	0.21–1.07	6.6E−02	0.53
miR-99a	LumA;ER+/HER2−	–	–	–	–	–	47	0.19	0.041–0.918	**2.1E−02**	0.40
	Het	85	1.6	0.93–2.74	8.6E−02	0.10	93	0.82	0.37–1.82	6.2E−01	0.84
miR-100	Het	85	1.08	0.64–1.81	7.7E−01	0.31	93	0.78	0.35–1.73	5.3E−01	0.84
miR-101	Het	85	0.53	0.32–0.90	**1.6E−02**	0.07	93	0.54	0.24–1.22	1.3E−01	0.53
miR-125b	LumA; ER+/HER2−	–	–	–	–	–	47	0.35	0.09–1.35	1.1E−01	0.53
miR-146a	TNBC	53	0.66	0.35–1.27	2.1E−01	0.14	–	–	–	–	–
miR-146	TNBC	53	1.2	0.6–2.2	6.6E−01	0.29	–	–	–	–	–
miR-155	TNBC	53	0.65	0.34–1.24	1.9E−01	0.14	–	–	–	–	–
miR-181a	TNBC	53	1.74	0.91–3.34	9.2E−02	0.10	–	–	–	–	–

Table 1 continued

MiRNA	Patients	GSE40267					GSE19783				
		n	HR	95 % CI	*p* value	*q* value	*n*	HR	95 % CI	*p* value	*q* value
miR-185	Het	85	0.75	0.45–1.26	2.8E−01	0.18	93	1.25	0.56–2.82	5.9E−01	0.84
miR-195	HER2−	71	0.82	0.46–1.48	5.1E−01	0.27	68	0.48	0.18–1.3	1.4E−01	0.53
miR-204	Het	85	0.62	0.37–1.04	7.0E−02	0.10	93	1.03	0.46–2.30	9.4E−01	0.99
miR-205	Het	85	0.87	0.52–1.46	6.0E−01	0.28	93	1.4	0.6–3.1	4.6E−01	0.84
miR-210	Het	85	1.02	0.61–1.72	9.3E−01	0.34	93	1.16	0.52–2.59	7.2E−01	0.88
miR-218	Het	85	1.5	0.88–2.53	1.3E−01	0.12	93	0.97	0.43–2.16	9.4E−01	0.99
miR-339-5p	Het	85	0.86	0.51–1.45	5.6E−01	0.27	93	1	0.45–2.22	9.9E−01	0.99
miR-342-5p	Het	85	0.65	0.39–1.09	1.0E−01	0.10	93	0.62	0.27–1.39	2.4E−01	0.76
miR-526b	Het	85	2.1	1.2–3.5	**5.2E−03**	0.05	93	0.71	0.32–1.60	4.1E−01	0.84
miR-1258	Het	–	–	–	–	–	–	–	–	–	–

CI confidence interval, *LumA* luminal A, *Het* heterogeneous, *ER+* estrogen receptor-positive, *HER2−* human epidermal growth factor receptor 2-negative, *HR* hazard ratio, *TNBC* triple-negative breast cancer. *p* values ≤ 0.05 are given in bold

Kaplan–Meier plot is drawn. Importantly, the patients can be filtered using major clinical parameters, such as receptor status, molecular subtype, lymph node status, and histological grade. Additionally, the analysis can be performed filtering patients by the treatment received. Our tool allows the selection of systemically untreated patients, representing a real prognostic setting, and of patients treated with endocrine therapy or chemotherapy. The web address of the online service is available on the website: http://www.kmplot.com/mirpower.

Validation of previously published survival-associated miRNAs

MiRNAs reported to be associated with OS in breast cancer tissues were identified using literature search. We computed Kaplan–Meier plots for the selected 41 miRNAs to validate their prognostic relevance in breast cancer (Online Resource 4). Overall, we confirmed the previously found association with OS for only 26 miRNAs (Table 1). Of these, only two miRNAs (*miR-29c* and *miR-101*) were prognostic in three datasets (Fig. 3; Table 1), while 5 miRNAs (*miR-22*, *miR-195*, *miR-204*, *miR-210*, *miR-218*) and 19 miRNAs were associated with patient outcome in two and one datasets, respectively (Table 1). It is worth noting that the association with survival for *miR-10b* and *miR-22* were opposite to those reported in the literature.

Discussion

Oncogenic signaling pathways and gene expression profiles have been associated with breast tumor development and progression, and correlated with patient outcome [3, 15]. Even though novel technologies have accelerated the discovery of potential biomarkers, starting to reshape research directions, the identification of reliable prognostic biomarkers still represents an issue of great clinical interest in breast cancer research. Indeed, small sample size, clinical and molecular heterogeneity of the data, inconsistent performance of methodologies, and lack of technical standardization have limited the identification of definitely relevant prognostic markers in cancer patients. Thus, validation should be essential prior to translate the identified biomarkers into clinical practice. Recently, our group has developed an integrative data analysis tool for the preliminary assessment of prognostic biomarkers in breast, ovarian, and non-small-cell lung cancers [16–18]. Beside gene expression, miRNAs have emerged as a new class of promising breast cancer biomarkers, as well as novel molecular agents to be considered for different clinical applications [5–10, 19–23].

In our study, we established a database integrating miRNA expression data and clinical information derived from four independent datasets of breast cancer. Finally, we developed an online tool that allows a real-time analysis to evaluate the prognostic value of these miRNAs in breast cancer.

We demonstrated good performance capabilities of miRpower through the validation of a complete list of survival-associated miRNAs identified from a literature-based research. In particular, we confirmed the reliable association with OS for *miR-29c* and *miR-101*, which were prognostic in three datasets, and for *miR-22*, *miR-195*, *miR-204*, *miR-210*, *miR-218*, which were associated with patient outcome in two datasets. Association with survival for *miR-10b* and *miR-22* was found to be opposite to those reported in the literature [24, 25]. The differing results

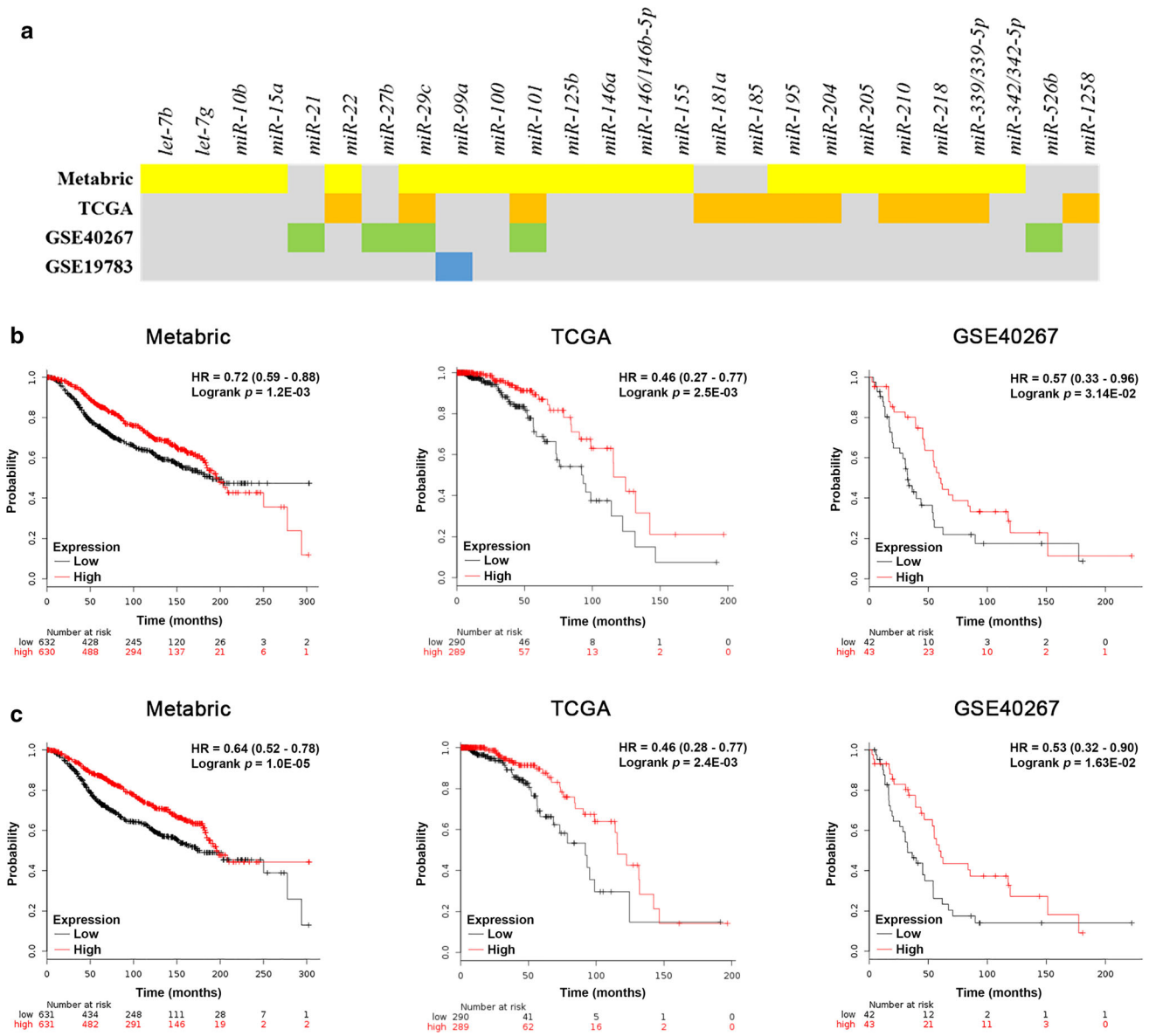

Fig. 3 Validation of previously published prognostic miRNAs. **a** Heatmap showing miRNAs with prognostic relevance in each dataset. **b** Kaplan–Meier plots for *miR-29c* in breast cancer cohorts. **c** Kaplan–Meier plots for *miR-101* in breast cancer cohorts. Log-rank *p* values and hazard ratios (HRs; 95 % confidence interval in parentheses) are shown

reported for *miR-10b* could be explained by differences in the cohort of tumors used, statistical analysis adopted, and an overall absence of consensus regarding the prognostic ability of this miRNA. For instance, in a previous study, the prognostic role of *miR-10b* was assessed using an expression value derived from a ratio between the *miR-10b* expression levels in cancer tissues and paired normal tissue [24]. Conversely, a different study demonstrated that *miR-10b* expression did not correlate with the development of distant metastases, relapse-free survival, and breast-cancer-specific survival, suggesting that *miR-10b* is unlikely to correlate with poor prognosis in breast cancer [26]. Our findings suggest that *miR-10b*, although involved in tumor invasion and metastasis, cannot be considered yet a reliable predictor of OS in breast cancer, given its significant association with outcome only in one dataset. Even though several studies have suggested a tumor suppressive role for *miR-22* and an association with better outcome, other studies demonstrated that *miR-22* acts as an oncogene, promoting epithelial-to-mesenchymal transition and an aggressive metastatic disease, and correlates with poor prognosis in breast cancer [25, 27, 28]. These data are in line with our results, which can be considered reliable as *miR-22* was confirmed a prognostic marker in two distinct

datasets. Thus, further analyses are needed to prove the prognostic value of the expression of these two miRNAs in breast cancer.

A second potential utility of miRpower is to generate *de novo* hypothesis using one or more of the datasets as a discovery set. There are certain caveats to keep in mind during such an analysis—in particular those related to unmeasured confounding variables and false discovery. A solution for this second issue is provided using a separate section of the online calculator enabling the computation of false discovery rate—this section of the homepage implements statistical features described in our previous paper [29].

Similar tools were recently developed to associate miRNAs expression with clinical outcome [30–32]. However, miRpower has several advantages in the evaluation of miRNAs as predictors of outcome in breast cancer patients. Using a large cohort composed of 2178 breast cancer patients, miRpower is the only tool acquiring data from multiple independent datasets for a single tumor entity. Having the highest number of patients increases statistical power and the robustness of the results. In addition, the selection of four studies meeting specific criteria provides the capability to cross-validate the selected genes in independent datasets. Furthermore, miRpower offers an unprecedented flexibility, allowing the user to select a priori group of patients to analyze. Indeed, before running the analysis, patients can be filtered using clinical parameters, including ER, PGR, HER2 statuses, lymph node involvement, and tumor grade. Importantly, our tool allows the selection of systemically untreated patients, and patients treated with endocrine therapy or chemotherapy, clearing the identification of miRNAs that could impact patient outcome in a specific clinical setting. We have to note that almost all patients in the untreated cohort were node negative and had a very good chance for a complete response using surgery only. Finally, miRpower was designed with a very intuitive interface, enabling also researchers with no bioinformatic expertise to perform survival analysis.

There are two limitations of miRpower we have to mention. Firstly, only OS data were available for each included dataset. In most settings, relapse-free survival is essential as it has higher relevance for selection of the optimal treatment. Secondly, only a fraction of about 20 % of all miRNAs were measured in each of the four datasets and more than half of all miRNAs were measured by only one platform. Cross-validation in at least two independent datasets is available for 498 miRNAs.

In conclusion, we have designed an easy-to-use bioinformatic tool capable of performing survival analyses for the identification and validation of prognostic miRNAs in breast cancer. We are planning to extend our platform by integrating future datasets with upcoming clinical data and providing additional statistical options. This resource represents a useful tool to support biomedical researchers in the evaluation of prognostic power of miRNA-based biomarkers.

Acknowledgments This study was supported by the Hungarian Scientific Research Fund (OTKA) K 108655 Grant (to B.G.), Associazione Italiana Ricerca sul Cancro (Grant 6251 to L.S.), and Fondazione Italiana Ricerca sul Cancro (FIRC fellowship 18328 to G.B.). The authors are grateful to Laura Paladini for her cooperation in data collection.

Author Contributions B.G. and L.S. conceived, designed, and supervised the study. B.G., A.L., A.N., and L.S. performed the analysis. G.B., B.G., G.M., and L.S. reviewed the literature. G.B., B.G., A.L., A.N., L.S., and A.S. participated in data interpretation. All authors were involved in writing and reviewing the manuscript, and approved the final manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 136:E359–E386. doi:10.1002/ijc.29210
2. Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thürlimann B, Senn HJ (2013) Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Ann Oncol 24:2206–2223. doi:10.1093/annonc/mdt303
3. Dowsett M, Dunbier AK (2008) Emerging biomarkers and new understanding of traditional markers in personalized therapy for breast cancer. Clin Cancer Res 14:8019–8026. doi:10.1158/1078-0432.CCR-08-0974
4. Iorio MV, Croce CM (2012) MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. EMBO Mol Med 4:143–159. doi:10.1002/emmm.201100209
5. Blenkiron C, Goldstein LD, Thorne NP, Spiteri I, Chin SF, Dunning MJ, Barbosa-Morais NL, Teschendorff AE, Green AR, Ellis IO, Tavaré S, Caldas C, Miska EA (2007) MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. Genome Biol 8:R214. doi:10.1186/gb-2007-8-10-r214
6. Iorio MV, Ferracin M, Liu CG, Veronese A, Spizzo R, Sabbioni S, Magri E, Pedriali M, Fabbri M, Campiglio M, Ménard S, Palazzo JP, Rosenberg A, Musiani P, Volinia S, Nenci I, Calin GA, Querzoli P, Negrini M, Croce CM (2005) MicroRNA gene expression deregulation in human breast cancer. Cancer Res 65:7065–7070. doi:10.1158/0008-5472.CAN-05-1783
7. Buffa FM, Camps C, Winchester L, Snell CE, Gee HE, Sheldon H, Taylor M, Harris AL, Ragoussis J (2011) microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression

profiling in breast cancer. Cancer Res 71:5635–5645. doi:10.1158/0008-5472.CAN-11-0489
8. Volinia S, Croce CM (2013) Prognostic microRNA/mRNA signature from the integrated analysis of patients with invasive breast cancer. Proc Natl Acad Sci USA 110:7413–7417. doi:10.1073/pnas.1304977110
9. van Schooneveld E, Wildiers H, Vergote I, Vermeulen PB, Dirix LY, Van Laere SJ (2015) Dysregulation of microRNAs in breast cancer and their potential role as prognostic and predictive biomarkers in patient management. Breast Cancer Res 17:21. doi:10.1186/s13058-015-0526-y
10. Bertoli G, Cava C, Castiglioni I (2015) MicroRNAs: new biomarkers for diagnosis, prognosis, therapy prediction and therapeutic tools for breast cancer. Theranostics 5:1122–1143. doi:10.7150/thno.11543
11. Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. Nature 490:61–70. doi:10.1038/nature11412
12. Dvinge H, Git A, Gräf S, Salmon-Divon M, Curtis C, Sottoriva A, Zhao Y, Hirst M, Armisen J, Miska EA, Chin SF, Provenzano E, Turashvili G, Green A, Ellis I, Aparicio S, Caldas C (2013) The shaping and functional consequences of the microRNA landscape in breast cancer. Nature 497:378–382. doi:10.1038/nature12108
13. de Rinaldis E, Gazinska P, Mera A, Modrusan Z, Fedorowicz GM, Burford B, Gillett C, Marra P, Grigoriadis A, Dornan D, Holmberg L, Pinder S, Tutt A (2013) Integrated genomic analysis of triple-negative breast cancers reveals novel microRNAs associated with clinical and molecular phenotypes and sheds light on the pathways they control. BMC Genom 23(14):643. doi:10.1186/1471-2164-14-643
14. Enerly E, Steinfeld I, Kleivi K, Leivonen SK, Aure MR, Russnes HG, Rønneberg JA, Johnsen H, Navon R, Rødland E, Mäkelä R, Naume B, Perälä M, Kallioniemi O, Kristensen VN, Yakhini Z, Børresen-Dale AL (2011) miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. PLoS One 6:e16915. doi:10.1371/journal.pone.0016915
15. Santarpia L, Bottai G, Kelly CM, Győrffy B, Székely B, Pusztai L (2016) Deciphering and targeting oncogenic mutations and pathways in breast cancer. Oncologist 21:1063–1078. doi:10.1634/theoncologist.2015-0369
16. Győrffy B, Lanczky A, Eklund AC, Denkert C, Budczies J, Li Q, Szallasi Z (2010) An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1809 patients. Breast Cancer Res Treat 123:725–731. doi:10.1007/s10549-009-0674-9
17. Gyorffy B, Lánczky A, Szállási Z (2012) Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients. Endocr Relat Cancer 19:197–208. doi:10.1530/ERC-11-0329
18. Győrffy B, Surowiak P, Budczies J, Lánczky A (2013) Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. PLoS One 8:e82241. doi:10.1371/journal.pone.0082241
19. Okada Y, Muramatsu T, Suita N, Kanai M, Kawakami E, Iotchkova V, Soranzo N, Inazawa J, Tanaka T (2016) Significant impact of miRNA-target gene networks on genetics of human complex traits. Sci Rep 6:22223. doi:10.1038/srep22223
20. Chen X, Yan CC, Zhang X, You ZH, Deng L, Liu Y, Zhang Y, Dai Q (2016) WBSMDA: within and between score for MiRNA-disease association prediction. Sci Rep 6:21106. doi:10.1038/srep21106
21. Meng F, Wang J, Dai E, Yang F, Chen X, Wang S, Yu X, Liu D, Jiang W (2016) Psmir: a database of potential associations between small molecules and miRNAs. Sci Rep 6:19264. doi:10.1038/srep19264
22. Kleivi Sahlberg K, Bottai G, Naume B, Burwinkel B, Calin GA, Børresen-Dale AL, Santarpia L (2015) A serum microRNA signature predicts tumor relapse and survival in triple-negative breast cancer patients. Clin Cancer Res 21:1207–1214. doi:10.1158/1078-0432.CCR-14-2011
23. De Mattos-Arruda L, Bottai G, Nuciforo PG, Di Tommaso L, Giovannetti E, Peg V, Losurdo A, Pérez-Garcia J, Masci G, Corsi F, Cortés J, Seoane J, Calin GA, Santarpia L (2015) MicroRNA-21 links epithelial-to-mesenchymal transition and inflammatory signals to confer resistance to neoadjuvant trastuzumab and chemotherapy in HER2-positive breast cancer patients. Oncotarget 6:37269–37280. doi:10.18632/oncotarget.5495
24. Parrella P, Barbano R, Pasculli B, Fontana A, Copetti M, Valori VM, Poeta ML, Perrone G, Righi D, Castelvetere M, Coco M, Balsamo T, Morritti M, Pellegrini F, Onetti-Muda A, Maiello E, Murgo R, Fazio VM (2014) Evaluation of microRNA-10b prognostic significance in a prospective cohort of breast cancer patients. Mol Cancer 13:142. doi:10.1186/1476-4598-13-142
25. Chen B, Tang H, Liu X, Liu P, Yang L, Xie X, Ye F, Song C, Xie X, Wei W (2015) miR-22 as a prognostic factor targets glucose transporter protein type 1 in breast cancer. Cancer Lett 356:410–417. doi:10.1016/j.canlet.2014.09.028
26. Gee HE, Camps C, Buffa FM, Colella S, Sheldon H, Gleadle JM, Ragoussis J, Harris AL (2008) MicroRNA-10b and breast cancer metastasis. Nature 455:E8–E9. doi:10.1038/nature07362
27. Pandey AK, Zhang Y, Zhang S, Li Y, Tucker-Kellogg G, Yang H, Jha S (2015) TIP60-miR-22 axis as a prognostic marker of breast cancer progression. Oncotarget 6:41290–41306. doi:10.18632/oncotarget.5636
28. Song SJ, Poliseno L, Song MS, Ala U, Webster K, Ng C, Beringer G, Brikbak NJ, Yuan X, Cantley LC, Richardson AL, Pandolfi PP (2013) MicroRNA-antagonism regulates breast cancer stemness and metastasis via TET-family-dependent chromatin remodeling. Cell 154:311–324. doi:10.1016/j.cell.2013.06.026
29. Gyorffy B, Gyorffy A, Tulassay Z (2005) The problem of multiple testing and solutions for genome-wide studies. Orv Hetil 146:559–563
30. Antonov AV, Knight RA, Melino G, Barlev NA, Tsvetkov PO (2013) MIRUMIR: an online tool to test microRNAs as biomarkers to predict survival in cancer using multiple clinical data sets. Cell Death Differ 20:367. doi:10.1038/cdd.2012.137
31. Goswami CP, Nakshatri H (2012) PROGmiR: a tool for identifying prognostic miRNA biomarkers in multiple cancers using publicly available data. J Clin Bioinform 2:23. doi:10.1186/2043-9113-2-23
32. Aguirre-Gamboa R, Trevino V (2014) SurvMicro: assessment of miRNA-based prognostic signatures for cancer clinical outcomes by multivariate survival analysis. Bioinformatics 30:1630–1632. doi:10.1093/bioinformatics/btu087