

Shortening of 3' UTRs Correlates with Poor Prognosis in Breast and Lung Cancer

Antonio Lembo, Ferdinando Di Cunto, Paolo Provero*

Department of Genetics, Biology and Biochemistry and Molecular Biotechnology Center, University of Turin, Turin, Italy

Abstract

A major part of the post-transcriptional regulation of gene expression is affected by trans-acting elements, such as microRNAs, binding the 3' untranslated region (UTR) of their target mRNAs. Proliferating cells partly escape this type of negative regulation by expressing shorter 3' UTRs, depleted of microRNA binding sites, compared to non-proliferating cells. Using large-scale gene expression datasets, we show that a similar phenomenon takes place in breast and lung cancer: tumors expressing shorter 3' UTRs tend to be more aggressive and to result in shorter patient survival. Moreover, we show that a gene expression signature based only on the expression ratio of alternative 3' UTRs is a strong predictor of survival in both tumors. Genes undergoing 3' UTR shortening in aggressive tumors of the two tissues significantly overlap, and several of them are known to be involved in tumor progression. However the pattern of 3' UTR shortening in aggressive tumors *in vivo* is clearly distinct from analogous patterns involved in proliferation and transformation.

Citation: Lembo A, Di Cunto F, Provero P (2012) Shortening of 3' UTRs Correlates with Poor Prognosis in Breast and Lung Cancer. PLoS ONE 7(2): e31129. doi:10.1371/journal.pone.0031129

Editor: Jun Li, Sun Yat-sen University Medical School, China

Received: August 12, 2011; **Accepted:** January 3, 2012; **Published:** February 8, 2012

Copyright: © 2012 Lembo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: PP and AL are supported by the Italian Association for Cancer Research under grant IG-9408. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: paolo.provero@unito.it

Introduction

MicroRNAs and other trans-acting factors that bind mRNA regulate protein levels by preventing mRNA translation and inducing its degradation. The key role played by this mechanism of post-transcriptional regulation in development [1] and disease [2] is emerging from an ever growing body of experimental results.

Many individual microRNAs have been shown to be aberrantly expressed in several types of cancer [2]. While some microRNAs act as oncogenes, in most reported cases it is the loss of microRNA expression that correlates with the onset or the aggressiveness of the tumor. A picture emerges in which microRNAs keep cells from inappropriately proliferating and invading the surrounding tissues [3,4] by repressing the expression of their targets.

In a heterogeneous cell population, selection will favor the cells that are able to escape such control. This can be done either by suppressing the expression of the individual microRNAs or, perhaps more efficiently, by interfering globally with the mechanism of regulation by microRNAs, for example by decreasing expression of genes involved in the microRNA pathway such as *DICER1* [5,6] or *TARBP2* [7].

Another way of escaping microRNA control is to remove their binding sites by expressing shorter transcripts: often this can be achieved without altering the gene product since the microRNA binding sites are mostly located within 3' untranslated regions. Many mRNAs contain 3' UTRs with alternative polyadenylation (APA) sites that can be used to produce transcripts of different length (see e.g. the review [8]).

Indeed it was shown that proliferating cells [9], cancer cells [10,11], early-stage embryonic cells [12,13] and induced pluripotent stem cells [14] express shortened 3' UTRs with fewer

microRNA binding sites. This way of escaping microRNA control by removing binding sites in *cis* has the advantage of being more fine-tunable than the removal in *trans* of the microRNA, as specific mRNA targets can be selectively released from regulation.

We therefore asked whether the same phenomenon exists in tumors, and in particular whether shorter 3' UTRs correlate with tumor aggressiveness and shorter survival times. Such correlation was previously shown in a mouse model of lymphoma in [15]. To address this question we took advantage of the extensive cancer gene expression datasets available in the public domain produced with Affymetrix chips of the U133 generation, and including patient survival and other clinical data.

While these chips were not originally designed for discriminating between alternative transcripts, they can be used [12] to answer our question, at least partially. Indeed, on one hand, most probes match the 3' region of the target transcripts and, on the other, probes can be reorganized into user-defined probesets. Exon arrays would be the obvious choice, but the number of large-scale cancer datasets obtained with this platform is still limited.

We thus defined custom probesets matching 3' UTR regions separated by APA sites: this turned out to be possible for thousands of transcripts, for which we determined the relative expression of the long and short forms of the 3' UTR in each patient. The correlation between these data and patient survival could then be studied to determine whether the expression of longer or shorter 3' UTRs correlates with cancer aggressiveness.

Results

Affymetrix chips of the HG-U133 series have been extensively used in studies of gene expression in cancer. Several datasets are

available containing hundreds of patients together with survival data and other clinical information. Most probes are located at the 3' end of the target transcript. Even though the single probes were organized by the manufacturer in standard probesets, they can also be reorganized into custom probesets [16], either to take advantage of improvements in the annotation of the genome or to perform special-purpose analyses. These features of the Affymetrix chips were previously exploited in [12,14] to study 3' UTR shortening in mouse development and in induced pluripotent stem cells.

We created a custom Chip Description File (CDF), *i.e.* a custom set of probesets, for the HG-U133A, HG-U133B and HG-U133_Plus_2 chips, matching alternative transcripts that differ in the portion of the 3'UTR they express. Alternative transcripts of this type are generated using APA sites, whose location can be predicted by sequence analysis and comparison with cDNA/EST sequences. We used the polyA_DB database [17,18] to obtain a collection of predicted APA sites.

Given a transcript with an APA site, we divided the Affymetrix probes matching the transcript into two sets, 5' and 3' probes, relative to the APA site. We then created separate probesets from the 5' and 3' probes. The software package we used [16] recommends creating probesets containing at least three probes. With this criterion, we could build separate 5' and 3' probesets for 2,706 transcripts represented in the HG-U133A and HG-U133B chips. For 685 transcripts it was possible to create 5' and 3' probes for more than one APA sites.

The long version of a transcript will hybridize with both the 5' and 3' probes, while the short version generated by the APA site will hybridize only with the 5' probes. Denoting with e_L (e_S) the expression level of the long (short) form, the signal from the 5' probeset will be

$$S_{5'} = \alpha_L e_L + \alpha_S e_S \quad (1)$$

where $\alpha_{L,S}$ represent the affinities of the long and short forms for the 5' probeset. Similarly the signal from the 3' probeset will be

$$S_{3'} = \beta_L e_L \quad (2)$$

so that the Expression Ratio Index (ERI) defined as their ratio

$$ERI \equiv \frac{S_{5'}}{S_{3'}} = \frac{\alpha_L}{\beta_L} + \frac{\alpha_S}{\beta_L} \frac{e_S}{e_L} \quad (3)$$

is a linear function of the expression ratio between the two alternative transcripts, and can be used to study their relative prevalence. On the other hand an overall change of expression of the transcript, which does not alter the relative amounts of the two forms, does not change the ERI value.

We can thus compare the ERI values in two biological samples to determine whether the long and short forms of the transcript are expressed in different ratios. A higher ERI corresponds to a higher prevalence of the short form.

Comparison with PCR-based results

We validated the use of the ERI to infer relative expression of long and short 3' UTR forms by comparing the PCR-based data reported in Ref. [10] in various cancer cell lines to microarray data obtained on HG-U133_Plus_2 array on some of the same cell lines and available from GEO repository accession GSE10843.

For 4 of the 6 genes shown in Figure 2 of [10] we could build separate probesets for the regions upstream and downstream of

the relevant APA site (*CCND1*, *CCND2*, *IGF2BP1*, *RAB10*). For *FGF2* and *DICER1* this was not possible, because all Affymetrix probes lie distal to the APA site. Incidentally, this shows that expression data for these genes obtained with HG-U133 platforms must be interpreted with great caution, because the microarray cannot hybridize to the transcript expressing the shorter 3' UTR form, which might very well be the one associated with higher protein expression. The relevant probes are shown in Figure 1.

As explained above, we expect the ERI to be a linear function of the expression ratio of short vs. long 3'UTR. Therefore we expect the ERI as computed from microarray data to correlate positively with direct measurements of expression ratios (in the absence of all experimental variability and noise we would expect a Pearson correlation coefficient equal to 1). The authors of [10] report in their Figure 2 the expression ratios of short vs. long 3'UTR for several cell lines originating from breast, colon and lung tumors. Each ratio is normalized to the same ratio in the corresponding normal tissue: therefore we computed, separately for each tissue of origin and each of the four genes, the Pearson correlation coefficient between the ERI derived from the microarray data and the normalized ratio reported by [10], computing a total of 6 correlation coefficients. All of them turned out to be positive ($P=0.016$, binomial test), varying between 0.34 for *CCND1* in lung cancer cells and >0.99 for *CCND1* in breast cancer cells. These results convinced us that the ERI we defined can indeed be used to probe the differential expression of alternative 3' UTR forms.

Comparison with high-throughput sequencing results

To validate the method also at a more global scale we compared its results to those obtained on breast cancer cell lines using high-throughput sequencing in a recent publication [11]. Here the authors studied differential APA usage between two breast cancer cell lines (MDA-MB-231 and MCF7) and a cultured mammary epithelial cell line (MCF10A).

We used microarray data for the three cell lines available from the GEO database under accession GSE10890, and selected untreated samples of the three cell lines of interest (7 samples for each cancer cell line and 2 samples of MCF10A). The platform used here is the HG-U133_Plus_2, for which we could build separate 5' and 3' probesets for 6,045 APA sites in 3,542 unique genes. For each APA site and each pair of cell lines we identified differential APA usage with a *t*-test on the ERI values followed by multiple testing correction with the Benjamini-Hochberg procedure and a False Discovery Rate (FDR) cutoff of 5%.

We thus obtained lists of genes with significantly shortened or lengthened 3' UTR for the three possible comparison, thus a total of 6 lists, to be compared with the analogous lists compiled from high-throughput sequencing in [11]. The overlap turned out to be statistically significant ($P < 0.05$) in 5 out of 6 cases, the most significant one corresponding to 3' UTRs that are shortened in MDA-MB-231 cells compared to MCF7 cells (318 genes concordantly detected by microarray and sequencing, compared to 227 expected by chance, $P = 2.7 \cdot 10^{-17}$, exact Fisher test). The only case in which the overlap was not statistically significant was 3' UTRs that are lengthened in MCF7 cells compared to MCF10A, which is also the least common case according to both approaches. We concluded that our method, when used on a genome-wide scale, gives results that are in good agreement with unbiased approaches such as high-throughput sequencing.

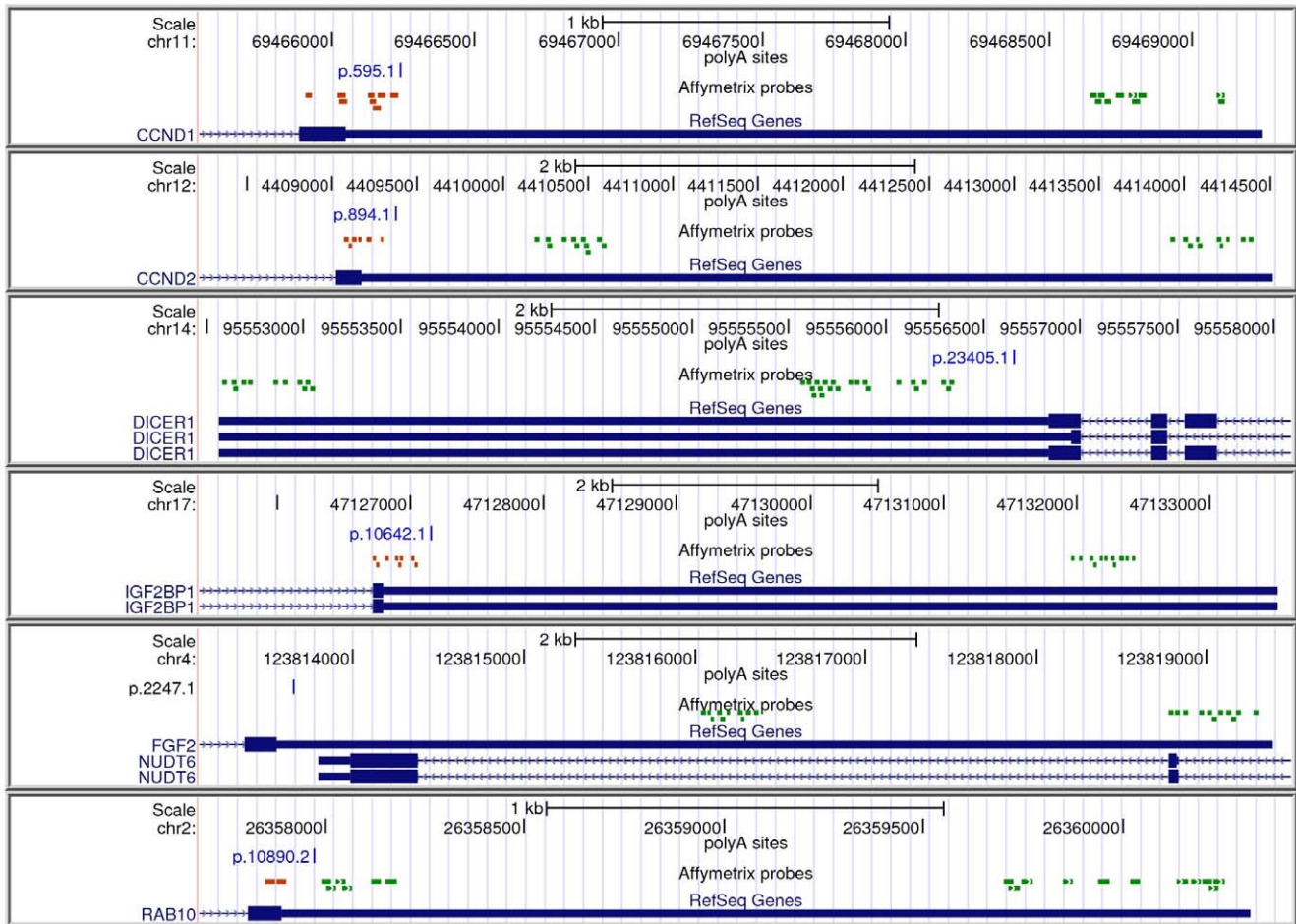


Figure 1. Affymetrix probes and APA sites in the 3' UTR of six genes studied in Ref. [10]. The probes shown in red are used, possibly together with probes located in the coding region, to build the 5' probeset. The probes shown in green are used to build the 3' probeset. For DICER1 and FGF2 it is not possible to study the APA site since all probes lie beyond it.
doi:10.1371/journal.pone.0031129.g001

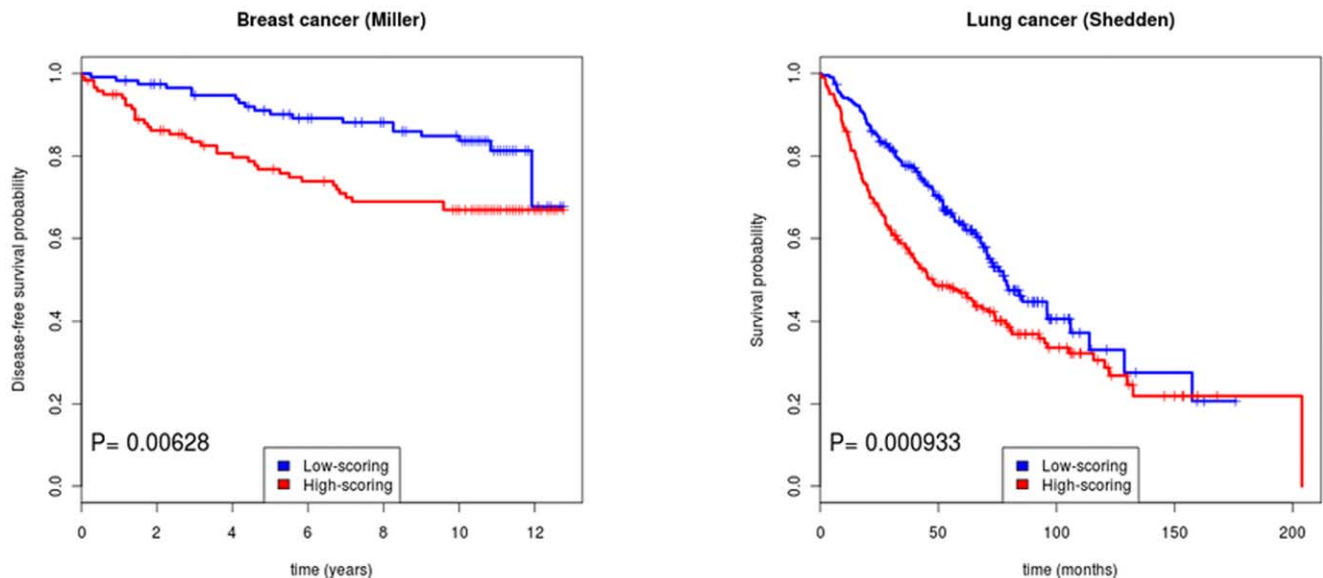


Figure 2. Predictive power of the prognostic score based on ERI in two dataset not used to derive the signature. (A) the Miller breast cancer dataset [20] and (B) the Shedden lung cancer dataset [32]. Patients are divided into groups using the median score as a cutoff.
doi:10.1371/journal.pone.0031129.g002

Aggressive breast cancers express transcripts with shorter 3' UTRs

Using the strategy described above, we evaluated the relative expression of long and short 3' UTR forms in a large, population-derived dataset of 159 cases of breast cancer in which gene expression profiles were measured using both HG-U133A and HG-U133B platforms [19], which we will refer to as the Pawitan dataset.

We divided the patients into two groups based on outcome (relapse) at 5 years. The large sample size allowed us to use a non-parametric test, so we replaced the *t*-test used for cell lines with a Mann-Whitney *U* test to find the transcripts for which the ERI is significantly different between the two groups, *i.e.* the transcripts for which the relative prevalence of long and short forms of the 3' UTR differs between the two groups of patients. After correcting for multiple testing using the Benjamini-Hochberg procedure we found, at 5% FDR, 35 APA sites in 28 different transcripts for which the relative expression of the short and the long form is significantly different between the two outcome groups (Table S1). Of these, 34 showed higher short/long expression ratio in the poor-outcome group and only one showed the opposite behavior: we conclude that, globally, the expression of shorter 3' UTRs correlates with poor outcome, in qualitative agreement with results obtained from cell lines [10].

A prognostic signature based on APA site usage

These results suggest that the relative prevalence of the short and long forms of the 35 APA sites can be used to stratify patients into risk classes. We thus developed a prognostic score based on the ERI of these APA sites. The prognostic score of a patient was defined as a linear combination of the ERI values of the 35 sites in the signature. The coefficients in the linear combination are determined from Cox univariate analysis as detailed in the Methods. Patients with high prognostic scores are characterized by relatively higher expression of the isoform associated to poor outcome (the short form for all but one APA site).

The prognostic score can be used to separate patients in two groups based on a cutoff. When applied to the Pawitan dataset, using the median score as a cutoff, the two groups of patients are significantly different in survival probability ($P = 1.99 \cdot 10^{-7}$, log-rank test). This is however expected, since the APA sites in the signature were chosen based on their differential ERI in the same dataset.

We thus used an independent, population-derived breast cancer dataset of 236 samples measured with the same platforms [20], the Miller dataset, to verify the robustness of the prognostic signature. The prognostic score of each patient of this dataset was computed as in the case of the Pawitan dataset, and the median score was used to divide the patients into two groups. Importantly, the coefficients used were the ones derived from the Pawitan dataset, so that the Miller dataset was not used at all in determining the signature. The two groups of patients differed significantly in disease-specific survival probability ($P = 6.28 \cdot 10^{-3}$, Figure 2A). Therefore the short/long expression ratio of the 35 APA sites identified in the Pawitan dataset can be used to create a robust outcome predictor based exclusively on the relative prevalence of alternative 3' UTRs.

Differential synthesis vs. differential degradation

The observed higher ERI in poor-outcome tumors could be due either to altered ratios in the synthesis of the two isoforms, or to altered degradation rates. As an example of the latter mechanism, assume a microRNA, able to bind the long isoform of a transcript

but not the short one, is expressed at higher levels in poor-outcome tumors. Then even if the two isoforms are synthesized at the same rate, the long one will be degraded faster in the poor-outcome tumors, leading to higher ERI. At least in the simplest case in which only one of the two mechanisms is at work it is possible to distinguish between them by looking at the differences in the total expression of the transcript between poor- and good-outcome cases (Figure 3).

First suppose differential synthesis is involved, *i.e.* relatively more of the short form is synthesized in poor-outcome tumors. Then the decay rate in this class of tumors will be lower, because of the loss of microRNA binding sites: hence we expect a higher total expression of the transcript in poor-prognosis tumors.

On the contrary, if the ratio of synthesized isoforms is the same in the two classes of tumors, but microRNAs able to degrade the transcript are more abundant in the poor-outcome cases, we expect higher degradation rates in the latter class, and thus lower total expression of the transcript in poor-prognosis tumors.

Since the probes located 5' of the APA site are able to hybridize to both isoforms, they can be used to evaluate the total expression of each transcript. Therefore we can distinguish between the two cases by looking at the correlation between differences in ERI and differences in signal from the 5' probesets: if differential synthesis (degradation) is the dominant mechanism, we expect positive (negative) correlation between these two quantities.

We thus computed, for each transcript, the difference in median expression of the 5' probeset between poor- and good-outcome tumors, and its correlation with the difference in median ERI. To avoid overcounting transcripts with more than one APA site we included only the 5'-most site for each transcript. We obtained a positive and highly significant Spearman correlation coefficient ($r_S = 0.30$, $P = 4.1 \cdot 10^{-58}$). While this analysis is based on a simplified model, it suggests differential synthesis is a significant mechanism leading to 3' UTR shortening in aggressive tumors.

3' UTR portions containing microRNA binding sites are preferentially lost in aggressive tumors

If the expression of shorter 3' UTRs confers a selective advantage by allowing escape from microRNA-mediated repression, we expect the 3' UTR portions containing microRNA binding sites to be lost in aggressive tumors more often than those which do not contain them. To test this expectation we defined as "microRNA-relevant" the APA sites removing at least one predicted microRNA binding site from the transcript. For each APA site, we took the difference in mean ERI between poor- and good-prognosis tumors as a measure of 3' UTR shortening: microRNA-relevant sites showed significantly more shortening than irrelevant ones ($P = 3.3 \cdot 10^{-9}$, Mann-Whitney *U* test). Thus the shortening of 3' UTRs in aggressive tumors is more pronounced for those transcripts for which it leads to loss of microRNA binding sites.

The same analysis can be performed for individual microRNAs, by comparing APA sites that are relevant, in the sense defined above, for specific microRNAs to irrelevant ones. At 10% FDR we found 5 microRNAs significantly associated with 3' UTR shortening in aggressive tumors: miR-7, miR-96, miR-200bc/429, miR-299-5p and miR-496. No microRNA was significantly associated with 3' UTR shortening in good-prognosis tumors.

Since these are the microRNAs whose regulatory activity is partly suppressed by the removal of their binding sites in shortened 3' UTRs, we would expect the microRNA themselves to have tumor-suppressor activity. Indeed three of these microRNAs were previously shown to have such effect in breast cancer cell lines: the miR-200bc/429 family negatively regulates EGF-driven cell

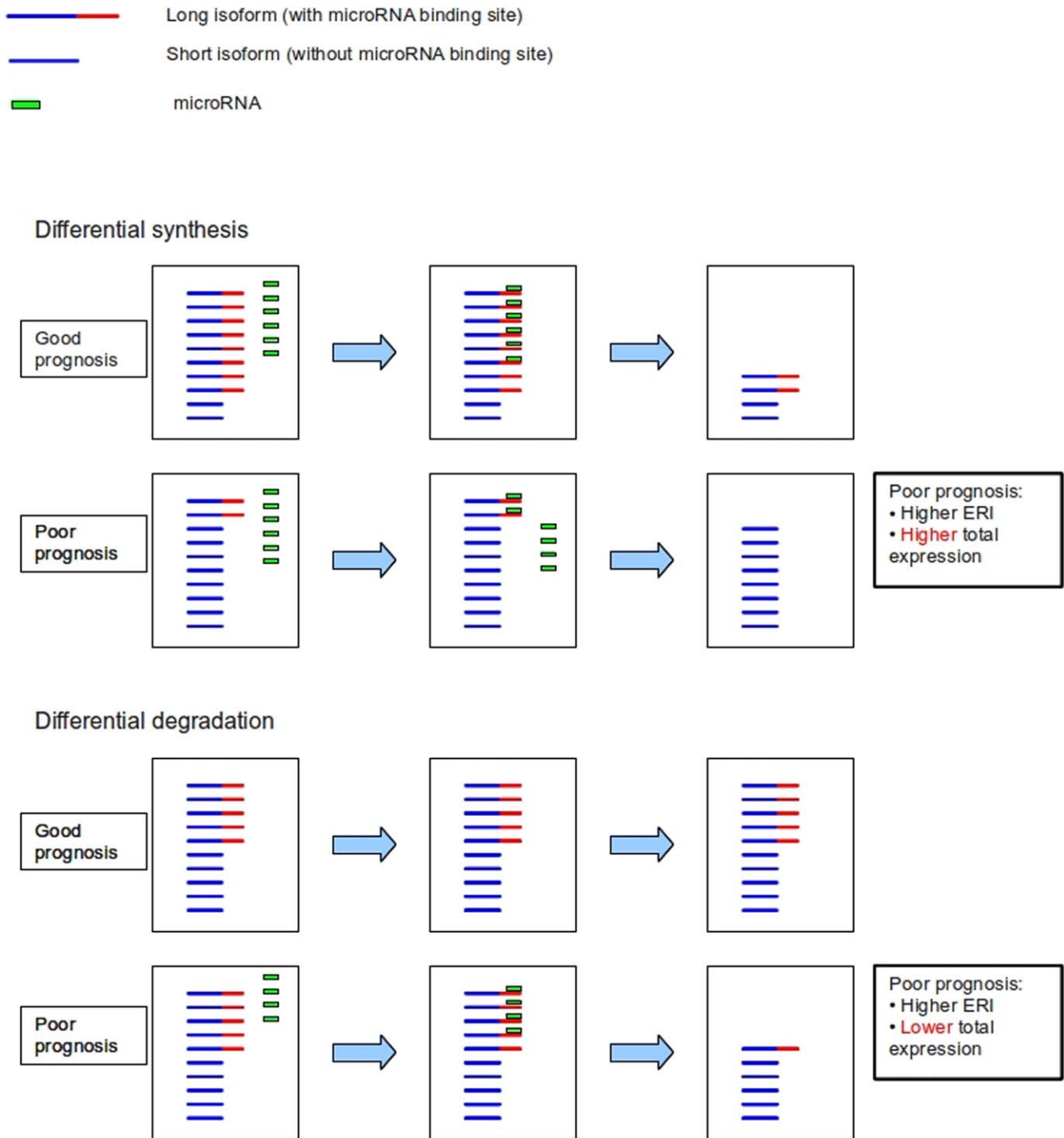


Figure 3. Schematic representation of two possible mechanisms leading to higher ERI in poor-prognosis tumors. (A) Differential synthesis: the shorter isoform is synthesized in a higher proportion in the poor-prognosis tumors, leading to less degradation by microRNAs. (B) Differential degradation: the isoforms are produced in the same proportion in the two cases, but a microRNA expressed exclusively in the poor-prognosis tumors selectively degrades the long isoform. In both cases we expect a higher ERI (relative prevalence of short form) in the poor-prognosis group. However in case (A) the overall expression of the two isoforms is expected to be higher in the poor-prognosis case, while the opposite is expected in case (B).
doi:10.1371/journal.pone.0031129.g003

invasion, viability and cell cycle progression [21], and suppresses the epithelial to mesenchymal transition [22]; tumor suppressor activity of miR-7 in breast cancer cell lines was shown in [23,24], and its role in preventing resistance to chemotherapy in [25,26]

(note, however, that miR-7 correlates with poor prognosis in a specific molecular class of breast cancer [27]); decreased expression of miR-299-5p was associated to increased tumorigenic potential [28]. Our results suggest that cancer cells can overcome

regulation by these microRNAs not only by decreasing their expression, but also by preventing them from binding to their targets.

More puzzling is the appearance of miR-96, which, on the contrary, was shown to have proliferation-promoting activity by repressing FOXO transcription factors [29,30]. However, as discussed above, the effect of removing selected targets from microRNA control can be vastly different from the effect of simply removing of repressing the microRNA itself.

Comparison with results in cell lines and normal tissues

Since 3UTR shortening has been previously associated to proliferation, one possibility to explain the specific effects that we detected could be the increased proliferation that characterizes aggressive tumors. Therefore we asked to what extent the transcripts showing differential APA usage in poor vs. good-prognosis patients coincide with those found in recent genome-wide studies of APA usage in normal and cancer cell lines. However, no significant overlap was found between the 28 transcripts involved in our breast cancer signature and the 980 mouse transcripts showing differential APA usage in mouse activated lymphocytes [9].

The same lack of statistically significant overlap was found when comparing our results to the ones obtained comparing different breast cancer cell lines to a normal breast cell line in [11]. Such negative result holds also when using our own microarray-derived results for the same cell lines (from the same dataset GSE10890 used above to check our algorithm), suggesting that the lack of overlap is not due to different biases between sequencing- and microarray-based measurements.

On the contrary, a pattern of differential 3' UTR usage similar to the one we find in aggressive vs. less aggressive breast tumors is found when comparing breast tumor samples (independently of aggressiveness) to healthy tissue: using the data of Ref. [31] (accession GSE3744 in GEO) we compared 38 sporadic tumor samples to 7 normal breast samples. We observed, as expected, a large number of APAs displaying 3' UTR shortening in cancer vs normal tissues (1,222 APAs corresponding to 921 unique genes at 5% FDR) and only 212 APAs (189 genes) showing the opposite behavior. The platform used here is the HG-U133_Plus_2, which allows us to analyze a total of 4,933 APAs (3,452 genes). The genes undergoing 3' UTR shortening significantly overlap with the ones found above from the Pawitan dataset (21 genes in common compared to 6.98 expected by chance, $P=2.1 \cdot 10^{-8}$).

Therefore while widespread 3' UTR shortening is observed in many different contexts, all generically related to increased proliferation, the transcripts involved are significantly different in the different contexts, even within the same tissue of origin. In particular we observe similar patterns of 3' UTR shortening when comparing breast tumors to normal tissues and when comparing breast tumors of different aggressiveness: however this pattern is markedly different from what observed in breast cancer cell lines, suggesting that the use of cancer cell lines in studying this particular phenomenon might not be ideal.

3' UTR shortening in aggressive lung cancer

We asked whether the pattern of 3'UTR shortening found in aggressive breast tumors was specific of breast cancer or relevant to other tumors. Specifically we considered a large lung cancer dataset [32], containing expression profiling of more than 400 lung adenocarcinomas. The 35-APA signature found in breast cancer turned out to be predictive of survival also in these patients: when computing the prognostic score of each patient using the signature and dividing them into two groups using the median score as cutoff

we found a significant difference in patient survival ($P=9.33 \cdot 10^{-4}$, Figure 2B). This result suggests that the pattern of 3' UTR shortening in aggressive lung cancer is similar to the one observed in breast cancer.

Therefore we repeated the analysis leading to the breast cancer signature directly on the lung cancer dataset. To account for the shorter overall survival of lung cancer patients we defined the poor-prognosis group to include all patients who died of the disease within 30 months. At 5% FDR we found a total of 319 APA sites with differential usage between good- and poor-prognosis patients (Table S2): in 258 (81%) of these the shorter isoform is predominantly expressed in poor-prognosis patients.

The overlap between the breast and lung cancer results is highly significant: 11 genes (Table 1) show significant 3' UTR shortening in poor-prognosis patients of both cancers, compared to an overlap of 2.4 genes expected by chance ($P=6.7 \cdot 10^{-6}$). This result is surprising given the lack of significant overlap with results in cell lines discussed above, and suggests the existence of a specific program of 3' UTR shortening associated with tumor aggressiveness in different tissues. This program is, on one hand, largely shared between breast and lung tumors, and, on the other, clearly distinct from analogous programs associated with physiological proliferation [9] and cell transformation [11].

A first interesting observation about these genes is that many of them have been positively implicated in breast and/or lung cancer progression and resistance to therapy: *AURKA* [33,34], *CAV2* [35], *RAB27A* [36], *RAD51* [37–39], *TIMELESS* [40], *TIMM17A* [41]. On the other hand there is evidence of a tumor-suppressing role of *SPARC* [42,43].

The available information about the function of these genes suggests that one of the mechanisms by which they may promote cancer aggressiveness could be through increasing cell migration and invasiveness. Indeed, the overexpression of *AURKA* in breast cancer cells has been reported to increase cell migration through ADF/cofilin pathway [33]. A related action on cytoskeleton could be exerted by *CACYBP* [44], although the results for its implication in breast cancer are seemingly contradictory [45,46]. Caveolins are well known modulators of cell motility and migration [47] and are significantly overexpressed in breast cancer cells characterized by basal-like and triple-negative immunophenotype [48]. *RAB27A* was shown to confer invasive and metastatic phenotypes on breast cancer cells by promoting the

Table 1. Genes showing significant 3' UTR shortening in poor-prognosis patients of both breast and lung cancer.

Gene Symbol	Gene
<i>ARL1</i>	ADP-ribosylation factor-like 1
<i>AURKA</i>	aurora kinase A
<i>CACYBP</i>	calyculin binding protein
<i>CAV2</i>	caveolin 2
<i>CENPN</i>	centromere protein N
<i>KIAA0101</i>	KIAA0101
<i>RAB27A</i>	RAB27A, member RAS oncogene family
<i>RAD51</i>	RAD51 homolog (RecA homolog, E. coli) (<i>S. cerevisiae</i>)
<i>SPARC</i>	secreted protein, acidic, cysteine-rich (osteonectin)
<i>TIMELESS</i>	timeless homolog (<i>Drosophila</i>)
<i>TIMM17A</i>	translocase of inner mitochondrial membrane 17 homolog A (yeast)

doi:10.1371/journal.pone.0031129.t001

secretion of insulin-like growth factor-II [36]. Finally, SPARC influences the synthesis of extracellular matrix, elicits changes in cell shape, inhibits cell-cycle progression [49]. An even more striking observation is that three genes in the list, i.e. AURKA [50,51], KIAA0101 [52] and RAD51 [53–55] are localized to centrosomes, may affect centrosome duplication and are capable to physically or functionally interact with the tumor suppressor gene BRCA1. A strong role in control of genomic stability for the genes in our signature is further indicated by the facts that CENPN is a factor required for centromere assembly [56] and that TIMELESS plays a crucial role in the ATM/ATR growth arrest response induced by DNA-damage and in the G2/M checkpoint [57].

Extension to other datasets and tumor types

To further investigate the robustness and general applicability of the two signatures derived from breast and lung cancer datasets we applied them to several other datasets available from GEO and for which both CEL files and survival data are available. Using the breast cancer-derived signature of 35 APAs we systematically observe, with the only exception of the colon cancer dataset of Ref. [58], shorter survival for high-scoring patients, even if this difference reaches statistical significance only in some cases (Figure S1). Also in this case the signature was built using the APAs and the coefficients derived from the Pawitan breast cancer dataset. In particular for breast cancer we observe that statistical significance is reached for the datasets including lymph-node positive patients [19,20,59], but not for those including only lymph-node negative patients [60,61]. This might suggest that shortening of 3' UTRs becomes important in the advanced stages of the disease. The lung-cancer derived signature shows similar results, but appears to be significantly less robust (Figure S2).

Discussion

We have shown that 3'UTR shortening in specific mRNAs correlates with poor prognosis in both breast and lung cancer, and that a significant part of this phenomenon is attributable to differential synthesis of alternative isoforms rather than to their differential degradation. Several genes involved in tumor progression and resistance to therapy express shortened 3' UTRs in aggressive tumors, supporting the notion that tumor cells gain an evolutionary advantage by removing these genes from the control of trans-acting regulators acting on the 3' UTR.

The fact that 3' UTRs are globally shorter in aggressive tumors could be expected, since shortening of 3' UTRs in various proliferative contexts have been reported in the literature [9–15]. However our analysis shows that there are specific patterns of 3' UTR shortening for different proliferative contexts. In particular the pattern of 3' UTR shortening associated to aggressive breast and lung cancer does not overlap significantly with what is observed in proliferating T cells [9] or even in breast cancer cell lines [11]. Since the latter negative result is obtained also when cancer cell lines are analyzed with the same microarray-based method used for cancer patients, it is unlikely to be artefactual. These results suggest that cancer cell lines might not be a suitable environment to study alternative APA usage in cancer, in agreement with the results of [11].

It is therefore all the more striking that, on the contrary, there is a significant overlap between genes undergoing 3' UTR shortening in aggressive breast and lung cancer. We thus suggest the existence of a specific pattern of 3' UTR shortening associated with tumor aggressiveness, significantly shared between breast and

lung cancer, but clearly distinct from other such patterns associated with proliferation and transformation.

The available information about genes that compose this shared signature is very well consistent with their causal involvement in the increased aggressiveness of tumors. Indeed, the expression variation resulting from variation of their 3'-UTR length could, on one hand, affect their invasive properties and, on the other hand, affect their propensity to undergo centrosome amplification, aberrant mitosis and resistance to DNA damage. Considering that AURKA can play a role in all these processes [33,50,51], our results further underscore its importance as a critical pharmacological target for preventing cancer progression.

Individual *cis*-acting genetic hits involved in 3' UTR shortening in cancer have been documented. For example in 7 out of 15 cases of mantle cell lymphoma expressing a truncated *CCND1* mRNA [62] a genomic deletion in the 3' UTR region was detected, and in 3 other cases a point mutation led to the creation of premature polyadenylation sites. However the effect we observed seems to involve too many genes to be explained solely or predominantly in this way. Many regulatory programs involving alternative usage of APA sites in various physiological contexts have been demonstrated, so that it appears likely that cancer cells could exploit a similar program to gain an evolutionary advantage.

Such a mechanisms would involve the binding of trans-acting factors to specific sites in the vicinity of the polyA site, as shown for example in [63] where it was shown that differential polyA usage in different human tissues correlates with overrepresentation of specific *cis*-elements. However a similar analysis in our case did not reveal any statistically significant pattern. This could be due either to a lack of statistical power (possibly related to the limited number of transcripts we can study) or to the fact that the relevant *cis*-elements in the case of cancer recognize structural rather than sequence features.

The most significant limit of the approach we presented here is the use of 3'-based chips. This choice was due to the large availability of gene expression profiles with survival information, but severely limits the number of genes that can be analyzed (up to ~3,500 for the HG-U133_Plus_2 platform). We have shown that the alternative 3' UTR usage of these genes can be reliably inferred from the microarray data and that it is biologically and clinically relevant; however since we cannot assume these genes to be an unbiased subset of the whole genome, all inferences about the global pattern of alternative 3' UTR usage must be taken with great caution. Future work based on RNA-sequencing or exon arrays will be needed to study the overall usage of alternative 3' UTRs in aggressive tumors in a less biased way.

Moreover, the analysis of datasets combining expression profiling and cancer genome sequencing data will determine what proportion of the effect is due to individual genetic hits and to changes in the regulation of mRNA polyadenylation and cleavage. Identification of possible trans-acting factors that may mediate the tumor progression-specific 3' UTR shortening that we here described could potentially lead to novel therapeutic targets.

Methods

Generation of custom Chip Description Files

The sequence of all human mRNAs included in Refseq database was obtained from the UCSC genome browser. From the same source we obtained the location of the APA sites identified in the polyA_DB database [17,18]. Each APA site was used to divide a mRNA sequence into two fragments (5' and 3' of the APA site). These fragments were then used as input for the Bioconductor package “altcdfenvs” [16], which generates the

custom CDF. Affymetrix gene expression data can be re-analyzed with a custom CDF if the raw data (CEL files) are available.

Evaluation of 3' UTR shortening

Expression data for the 5' and 3' probesets included in the custom CDF were obtained using Robust Multi-Array Average (RMA) [64]. For each APA site and each biological sample, we computed the Expression Ratio Index (ERI) defined in Eq. 3 as the ratio of RMA expression values. A software package containing the custom CDF files and an R script to compute the ERIs is provided as Software S1).

A Mann-Whitney U test was used to determine which APA sites showed significantly different ERI between the two group of patients (defined on the basis of survival status at 5 years in breast cancer and 2.5 years in lung cancer). Multiple testing was controlled with the Benjamini-Hochberg procedure.

Prognostic score based on alternative 3' UTR usage

To define a prognostic score based on ERI values we first computed, for each of the 35 APA sites differentially used in the Pawitan dataset, its Cox univariate score z . This score measures the correlation between a continuous quantity (in our case the ERI) and patient survival. A positive (negative) z implies a negative (positive) correlation between ERI and survival.

We then defined the prognostic score of a patient as:

$$SCORE = \sum_A z_A \cdot ERI_A \quad (4)$$

where the sum is over the 35 APA sites, z_A is the Cox score and ERI_A is the ERI value of the APA. High-scoring patients are thus characterized by high ERI values (*i.e.* high prevalence of the short 3' UTR isoform) for those APA sites that are most strongly correlated to poor survival. The z_A values for the 35 APA sites can be found in Table S1.

The prognostic score of each patient in each dataset was computed using Eq. 4, in which the z_A coefficients are always the ones derived from the Pawitan datasets. The predictive power of the prognostic score was evaluated in each dataset by dividing the patients into two groups using the median score as cutoff, and using a log-rank test to compare the cumulative survival probabilities of the two groups.

Cancer datasets

The cancer datasets we used for developing and testing the prognostic signatures were all downloaded from the Gene Expression Omnibus (GEO) repository, with the exception of the lung cancer dataset of [32] which was obtained from the caArray repository (<https://array.nci.nih.gov/caarray/>). We used a total of 6 breast cancer datasets (GSE7390 [61], GSE20685 [65], GSE3494 [20], GSE1456 [19], GSE2990 [59], GSE2034 [60]), one colon cancer dataset (GSE14333 [58]), three lung cancer datasets (the Shedden dataset mentioned above plus GSE3141

[66] and GSE14814 [67]) and one melanoma dataset (GSE22138 [68]).

Prevalence of microRNA binding sites in shortened 3' UTRs

An APA site was defined to be relevant to a specific microRNA if the sequence located 3' of the site contained one or more binding site predicted by Targetscan [69], as reported in the corresponding UCSC track. An APA site was defined to be microRNA-relevant if it was relevant to one or more specific microRNAs. The ERI values of microRNA-relevant vs. irrelevant APA sites were compared using a Mann-Whitney U test, with multiple testing controlled with the Benjamini-Hochberg procedure.

Supporting Information

Figure S1 Performance of the signature derived from the Pawitan breast cancer dataset as a predictor of survival in other cancer datasets. P-values are from log-rank tests. (TIFF)

Figure S2 Performance of the signature derived from the Shedden lung cancer dataset as a predictor of survival in other cancer datasets. P-values are from log-rank tests. (TIFF)

Table S1 APA sites showing significantly different Expression Ratio Index between bad- and good-prognosis breast cancer patients (Pawitan dataset). The columns are: (1) APA site, identified as in the PolyADB track of the UCSC genome browser (human genome release hg18); (2) Entrez gene id; (3) Official gene symbol; (4) gene name; (5) difference in logarithmic ERI between poor- and good-prognosis patients: a positive value indicates shortening of the 3' UTR in poor-prognosis patients; (6) P-value from Mann-Whitney U test; (7) Q-value from Benjamini-Hochberg correction; (8) z-score from Cox univariate analysis. (XLS)

Table S2 APA sites showing significantly different Expression Ratio Index between bad- and good-prognosis lung cancer patients (Shedden dataset). Columns as in Supporting Table S1. (XLS)

Software S1 An R package containing the custom CDF files and a script to compute the ERI values. (BZ2)

Acknowledgments

We thank N. Riggi, P. Shaw and I. Stamenkovic for stimulating discussions and suggestions.

Author Contributions

Conceived and designed the experiments: PP. Performed the experiments: AL PP. Analyzed the data: AL FDC PP. Wrote the paper: FDC PP.

References

- Bushati N, Cohen SM (2007) microRNA functions. *Annu Rev Cell Dev Biol* 23: 175–205.
- Croce CM (2009) Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet* 10: 704–714.
- Shi XB, Tepper CG, deVere White RW (2008) Cancerous miRNAs and their regulation. *Cell Cycle* 7: 1529–1538.
- Shi M, Liu D, Duan H, Shen B, Guo N (2010) Metastasis-related miRNAs, active players in breast cancer invasion, and metastasis. *Cancer Metastasis Rev* 29: 785–799.
- Kumar MS, Pester RE, Chen CY, Lane K, Chin C, et al. (2009) Dicer1 functions as a haploinsufficient tumor suppressor. *Genes Dev* 23: 2700–2704.
- Martello G, Rosato A, Ferrari F, Manfrin A, Cordenonsi M, et al. (2010) A microRNA targeting Dicer for metastasis control. *Cell* 141: 1195–1207.
- Melo SA, Ropero S, Moutinho C, Aaltonen LA, Yamamoto H, et al. (2009) A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function. *Nat Genet* 41: 365–370.
- Millevoi S, Vagner S (2010) Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic Acids Res* 38: 2757–2774.

9. Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB (2008) Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. *Science* 320: 1643–1647.
10. Mayr C, Bartel DP (2009) Widespread shortening of 3'UTRs by alternative cleavage and polyadenylation activates oncogenes in cancer cells. *Cell* 138: 673–684.
11. Fu Y, Sun Y, Li Y, Li J, Rao X, et al. (2011) Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing. *Genome Res* 21: 741–747.
12. Ji Z, Tian B (2009) Reprogramming of 3' untranslated regions of mRNAs by alternative polyadenylation in generation of pluripotent stem cells from different cell types. *PLoS One* 4: e8419.
13. Thomsen S, Azzam G, Kaschula R, Williams LS, Alonso CR (2010) Developmental RNA processing of 3'UTRs in Hox mRNAs as a context-dependent mechanism modulating visibility to microRNAs. *Development* 137: 2951–2960.
14. Ji Z, Lee JY, Pan Z, Jiang B, Tian B (2009) Progressive lengthening of 3' untranslated regions of mRNAs by alternative polyadenylation during mouse embryonic development. *Proc Natl Acad Sci U S A* 106: 7028–7033.
15. Singh P, Alley TL, Wright SM, Kamdar S, Schott W, et al. (2009) Global changes in processing of mRNA 3' untranslated regions characterize clinically distinct cancer subtypes. *Cancer Res* 69: 9422–9430.
16. Gautier L, Miller M, Friis-Hansen L, Knudsen S (2004) Alternative mapping of probes for Affymetrix chips. *BMC Bioinformatics* 5: 111.
17. Zhang H, Hu J, Recce M, Tian B (2005) PolyA DB: a database for mammalian mRNA polyadenylation. *Nucleic Acids Res* 33: D116–D120.
18. Lee JY, Yeh I, Park JY, Tian B (2007) PolyA DB 2: mRNA polyadenylation sites in vertebrate genes. *Nucleic Acids Res* 35: D165–D168.
19. Pawitan Y, Bjhlle J, Amler L, Borg AL, Eghazi S, et al. (2005) Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res* 7: R953–R964.
20. Miller LD, Smeds J, George J, Vega VB, Vergara L, et al. (2005) An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci U S A* 102: 13550–13555.
21. Uhlmann S, Zhang JD, Schwger A, Mannsperger H, Riazalhosseini Y, et al. (2010) miR-200bc/429 cluster targets PLCgamma1 and differentially regulates proliferation and EGF-driven invasion than miR-200a/141 in breast cancer. *Oncogene* 29: 4297–4306.
22. Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, et al. (2008) The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* 10: 593–601.
23. Webster RJ, Giles KM, Price KJ, Zhang PM, Mattick JS, et al. (2009) Regulation of epidermal growth factor receptor signaling in human cancer cells by microRNA-7. *J Biol Chem* 284: 5731–5741.
24. Reddy SDN, Ohshiro K, Rayala SK, Kumar R (2008) MicroRNA-7, a homeobox D10 target, inhibits p21-activated kinase 1 and regulates its functions. *Cancer Res* 68: 8195–8200.
25. Kastl L, Brown I, Schofield AC (2011) miRNA-34a is associated with docetaxel resistance in human breast cancer cells. *Breast Cancer Res Treat* doi: 10.1007/s10549-011-1424-3.
26. Pogribny IP, Filkowski JN, Tryndyak VP, Golubov A, Shpyleva SI, et al. (2010) Alterations of microRNAs and their targets are associated with acquired resistance of MCF-7 breast cancer cells to cisplatin. *Int J Cancer* 127: 1785–1794.
27. Foekens JA, Siewerts AM, Smid M, Look MP, de Weerd V, et al. (2008) Four miRNAs associated with aggressiveness of lymph node-negative, estrogen receptor-positive human breast cancer. *Proc Natl Acad Sci U S A* 105: 13021–13026.
28. Shevde LA, Metge BJ, Mitra A, Xi Y, Ju J, et al. (2010) Spheroid-forming subpopulation of breast cancer cells demonstrates vasculogenic mimicry via hsa-miR-299-5p regulated de novo expression of osteopontin. *J Cell Mol Med* 14: 1693–1706.
29. Guttilla IK, White BA (2009) Coordinate regulation of FOXO1 by miR-27a, miR-96, and miR-182 in breast cancer cells. *J Biol Chem* 284: 23204–23216.
30. Lin H, Dai T, Xiong H, Zhao X, Chen X, et al. (2010) Unregulated miR-96 induces cell proliferation in human breast cancer by downregulating transcriptional factor FOXO3a. *PLoS One* 5: e15797.
31. Richardson AL, Wang ZC, Nicolo AD, Lu X, Brown M, et al. (2006) X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell* 9: 121–132.
32. Shedden K, Taylor JMG, Enkemann SA, Tsao MS, Yeatman TJ, et al. (2008) Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* 14: 822–827.
33. hui Wang L, Xiang J, Yan M, Zhang Y, Zhao Y, et al. (2010) The mitotic kinase Aurora-A induces mammary cell migration and breast cancer metastasis by activating the Cofilin-F-actin pathway. *Cancer Res* 70: 9118–9128.
34. Ogawa E, Takenaka K, Katakura H, Adachi M, Otake Y, et al. (2008) Perimembrane Aurora-A expression is a significant prognostic factor in correlation with proliferative activity in non-small-cell lung cancer (NSCLC). *Ann Surg Oncol* 15: 547–554.
35. Savage K, Leung S, Todd SK, Brown LA, Jones RL, et al. (2008) Distribution and significance of caveolin 2 expression in normal breast and invasive breast cancer: an immunofluorescence and immunohistochemical analysis. *Breast Cancer Res Treat* 110: 245–256.
36. Wang JS, Wang FB, Zhang QG, Shen ZZ, Shao ZM (2008) Enhanced expression of Rab27A gene by breast cancer cells promoting invasiveness and the metastasis potential by secretion of insulin-like growth factor-II. *Mol Cancer Res* 6: 372–382.
37. Scodan RL, Cizeron-Clairac G, Fourme E, Meseure D, Vacher S, et al. (2010) DNA repair gene expression and risk of locoregional relapse in breast cancer patients. *Int J Radiat Oncol Biol Phys* 78: 328–336.
38. Barbano R, Copetti M, Perrone G, Paziienza V, Muscarella LA, et al. (2010) High RAD51 mRNA expression characterizes estrogen receptor-positive/progesterone receptor-negative breast cancer and is associated with patient's outcome. *Int J Cancer*.
39. Ko JC, Wang LH, Jhan JY, Ciou SC, Hong JH, et al. (2009) The role of celecoxib in Rad51 expression and cell survival affected by gefitinib in human non-small cell lung cancer cells. *Lung Cancer* 65: 290–298.
40. Tozlu-Kara S, Roux V, Andrieu C, Vendrell J, Vacher S, et al. (2007) Oligonucleotide microarray analysis of estrogen receptor alpha-positive postmenopausal breast carcinomas: identification of HRPAP20 and TIMELESS as outstanding candidate markers to predict the response to tamoxifen. *J Mol Endocrinol* 39: 305–318.
41. Xu X, Qiao M, Zhang Y, Jiang Y, Wei P, et al. (2010) Quantitative proteomics study of breast cancer cell lines isolated from a single patient: discovery of TIMM17A as a marker for breast cancer. *Proteomics* 10: 1374–1390.
42. Nagai MA, Gerhard R, Fregani JHTG, Nonogaki S, Rierger RB, et al. (2011) Prognostic value of NDRG1 and SPARC protein expression in breast cancer patients. *Breast Cancer Res Treat* 126: 1–14.
43. Suzuki M, Hao C, Takahashi T, Shigematsu H, Shivapurkar N, et al. (2005) Aberrant methylation of SPARC in human lung cancers. *Br J Cancer* 92: 942–948.
44. Schneider G, Nieznanski K, Jozwiak J, Slomnicki LP, Redowicz MJ, et al. (2010) Tubulin binding protein, CacyBP/SIP, induces actin polymerization and may link actin and tubulin cytoskeletons. *Biochim Biophys Acta* 1803: 1308–1317.
45. Wang N, Ma Q, Wang Y, Ma G, Zhai H (2010) CacyBP/SIP expression is involved in the clinical progression of breast cancer. *World J Surg* 34: 2545–2552.
46. Nie F, Yu XL, Wang XG, Tang YF, Wang LL, et al. (2010) Down-regulation of CacyBP is associated with poor prognosis and the effects on COX-2 expression in breast cancer. *Int J Oncol* 37: 1261–1269.
47. Grande-Garcia A, del Pozo MA (2008) Caveolin-1 in cell polarization and directional migration. *Eur J Cell Biol* 87: 641–647.
48. Elsheikh SE, Green AR, Rakha EA, Samama RM, Ammar AA, et al. (2008) Caveolin 1 and Caveolin 2 are associated with breast cancer basal-like and triple-negative immunophenotype. *Br J Cancer* 99: 327–334.
49. Bradshaw AD, Graves DC, Motamed K, Sage EH (2003) SPARC-null mice exhibit increased adiposity without significant differences in overall body weight. *Proc Natl Acad Sci U S A* 100: 6045–6050.
50. Sankaran S, Crone DE, Palazzo RE, Parvin JD (2007) Aurora-A kinase regulates breast cancer associated gene 1 inhibition of centrosome-dependent microtubule nucleation. *Cancer Res* 67: 11186–11194.
51. Dutertre S, Descamps S, Prigent C (2002) On the role of aurora-A in neurogenesis. *Oncogene* 21: 6175–6183.
52. Kais Z, Barsky SH, Mathsyaraja H, Zha A, Ransburgh DJR, et al. (2011) KIAA0101 interacts with BRCA1 and regulates centrosome number. *Mol Cancer Res*.
53. Dodson H, Bourke E, Jeffers LJ, Vagnarelli P, Sonoda E, et al. (2004) Centrosome amplification induced by DNA damage occurs during a prolonged G2 phase and involves ATM. *EMBO J* 23: 3864–3873.
54. Cappelli E, Townsend S, Griffin C, Thacker J (2011) Homologous recombination proteins are associated with centrosomes and are required for mitotic stability. *Exp Cell Res* 317: 1203–1213.
55. Sderlund K, Skoog L, Fornander T, Askmalms MS (2007) The BRCA1/BRCA2/Rad51 complex is a prognostic and predictive factor in early breast cancer. *Radiother Oncol* 84: 242–251.
56. Foltz DR, Jansen LET, Black BE, Bailey AO, Yates JR, et al. (2006) The human CENP-A centromeric nucleosome-associated complex. *Nat Cell Biol* 8: 458–469.
57. Yang X, Wood PA, Hrushesky WJM (2010) Mammalian TIMELESS is required for ATM-dependent CHK2 activation and G2/M checkpoint control. *J Biol Chem* 285: 3030–3034.
58. Jorissen RN, Gibbs P, Christie M, Prakash S, Lipton L, et al. (2009) Metastasis-Associated Gene Expression Changes Predict Poor Outcomes in Patients with Dukes Stage B and C Colorectal Cancer. *Clin Cancer Res* 15: 7642–7651.
59. Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, et al. (2006) Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* 98: 262–272.
60. Wang Y, Klijn JGM, Zhang Y, Siewerts AM, Look MP, et al. (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365: 671–679.
61. Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, et al. (2007) Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res* 13: 3207–3214.

62. Wiestner A, Tehrani M, Chiorazzi M, Wright G, Gibellini F, et al. (2007) Point mutations and genomic deletions in CCND1 create stable truncated cyclin D1 mRNAs that are associated with increased proliferation rate and shorter survival. *Blood* 109: 4599–4606.
63. Zhang H, Lee JY, Tian B (2005) Biased alternative polyadenylation in human tissues. *Genome Biol* 6: R100.
64. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, et al. (2003) Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 31: e15.
65. Kao KJ, Chang KM, Hsu HC, Huang AT (2011) Correlation of microarray-based breast cancer molecular subtypes and clinical outcomes: implications for treatment optimization. *BMC Cancer* 11: 143.
66. Bild AH, Yao G, Chang JT, Wang Q, Potti A, et al. (2006) Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 439: 353–357.
67. Zhu CQ, Ding K, Strumpf D, Weir BA, Meyerson M, et al. (2010) Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *J Clin Oncol* 28: 4417–4424.
68. Laurent C, Valet F, Planque N, Silveri L, Maacha S, et al. (2011) High PTP4A3 phosphatase expression correlates with metastatic risk in uveal melanoma patients. *Cancer Res* 71: 666–674.
69. Friedman RC, Farh KKH, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 19: 92–105.