microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer

Francesca M. Buffa\textsuperscript{a}, Carme Camps\textsuperscript{b}, Laura Winchester\textsuperscript{b}, Cameron E. Snell\textsuperscript{a}, Harriet E. Gee\textsuperscript{a}, Helen Sheldon\textsuperscript{a}, Marian Taylor\textsuperscript{a}, Adrian L. Harris\textsuperscript{a,1}, Jiannis Ragoussis\textsuperscript{b,1}

\textsuperscript{a}Oncology Department, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, OX3 9DS, United Kingdom; \textsuperscript{b}Genomics Research, Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN, United Kingdom

\textsuperscript{1}Senior co-leads

Running title: Integrated mRNA and microRNA analysis in breast cancer

Keywords: microRNA, gene expression, breast cancer, data-mining, target

Financial Support: Cancer Research UK; EU 6\textsuperscript{th}/7\textsuperscript{th} Framework Programme; Wellcome Trust.

Corresponding authors address:

Francesca M. Buffa: The Weatherall Institute of Molecular Medicine, University of Oxford, OX3 9DS, United Kingdom. E-mail: francesca.buffa@imm.ox.ac.uk. Phone: +44 (0)1865 222443.

Jiannis Ragoussis: The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford, OX3 7BN, United Kingdom. E-mail: ioannis.ragoussis@well.ox.ac.uk. Phone: +44(0)1865287526.

Conflicts of interest: none

Word count (including abstract, manuscript text, figure legends, but references): 4990

Total number of figures and tables: 7
ABSTRACT

microRNA expression profiling plays an emerging role in cancer classification and identification of therapeutic strategies. In this study, we have evaluated the benefits of a joint microRNA-mRNA analysis in breast cancer.

Matched mRNA and microRNA global expression profiling was performed in a well-annotated cohort of 207 cases with complete 10-year follow-up. Penalized Cox regression including microRNA expression, mRNA expression and clinical covariates was used to identify microRNAs associated with distant relapse-free survival (DRFS) that provide independent prognostic information, and are not simply surrogates of previously identified prognostic covariates. Penalized regression was chosen to prevent over-fitting. Furthermore, microRNA-mRNA relationships were explored by global expression analysis, and exploited to validate results in several published cohorts (N=592 with DRFS, N=1050 with recurrence-free survival).

Four microRNAs were independently associated with DRFS in oestrogen receptor (ER)-positive (3 novel and 1 known - miR-128a) and 6 in ER-negative (5 novel and 1 known - miR-210) cases. Of the latter, miR-342, -27b and -150 were prognostic also in triple receptor negative tumours. Coordinated inhibition of target expression by prognostic microRNAs strengthened these results; the most significant being miR-210, -128a and -27b, whose targets were prognostic in meta-analysis of several cohorts. In addition, miR-210 and -128a showed coordinated expression with their cognate pri-microRNAs, which were themselves prognostic in independent cohorts.

Our integrated microRNA-mRNA global profiling approach has identified microRNAs independently associated with prognosis in breast cancer. Furthermore it has validated known and predicted microRNA-target interactions, and elucidated their association with key pathways that could represent novel therapeutic targets.
INTRODUCTION

Breast cancer is a heterogeneous disease, with treatment resistance depending upon pathological and genomic characteristics. Messenger RNA (mRNA) expression profiling in clinical cohorts has led to identification of functional pathways with roles in tumor progression and collections of genes (gene signatures) associated with disease outcome (1), some of which are now FDA-approved for clinical use, such as MammaPrint and OncotypeDX.

MicroRNAs are small non-coding RNA molecules regulating cell function both at transcriptional and post-transcriptional levels, which thus open up a new area of prognostic marker research complementary to established transcriptional gene signature or traditional marker studies (2). A study by Blenkiron et al. of 93 breast cancers identified several human microRNAs associated with intrinsic breast cancer subtype (3). Another recent study of 38 breast cancers (4) selected 12 microRNAs associated with clinico-pathological variables for analysis in a cohort of 261 breast cancers. Amongst these, four (miR-7, -128a, -210, and -516–3p) were prognostic of which one, miR-210, had been previously identified (4).

However, integrated analysis of microRNA and mRNA global expression profiles has yet to be explored in prognostic studies. Such analyses have the potential to identify not only microRNAs that are independent prognostic factors, but also to elucidate microRNA function in-vivo, and identify interactions between microRNAs and targeted mRNAs for enhanced marker and therapeutic target discovery. Thus, we performed comprehensive microRNA and mRNA expression profiling in a large, well-annotated cohort of 207 early invasive breast cancers.

To identify microRNAs that provide independent information, and are not simply surrogates of previously identified prognostic covariates, a Cox regression for distant relapse-free survival (DRFS) was performed, including all microRNAs, clinical covariates and gene signatures. Pe-
nalized least-square minimization with variable selection and regularization was used to prevent over-fitting.

To investigate the functional role of prognostic microRNAs, relationships between host genes (pri-microRNAs), mature microRNAs and cognate target mRNAs were examined by global expression analysis. Finally, findings were confirmed using two further independent analyses. Firstly, pri-microRNA (up-stream of microRNA expression) and genes in the microRNA processing machinery were analyzed. Thus, expression of target genes, reflecting the functional effect downstream of microRNA expression, was considered. Prognostic significance of both analyses was assessed in published cohorts (N>1000 cases).

MATERIALS AND METHODS

Patient characteristics

A retrospective series of 219 patients with early primary breast cancer was considered (5); extended demographics are provided in Table S1 and Suppl. Information. Informed consent was obtained for each subject. Clinical investigations were conducted after approval by the local research ethics committee and in accordance with the ethical principles expressed in the Declaration of Helsinki. Main endpoint was distant relapse-free survival (DRFS).

mRNA and microRNA profiling

Matched microRNA and mRNA profiling was successfully obtained from 207 of 219 samples using Illumina Human RefSeq-8 and miRNAv1 arrays were used (see Suppl. Information). Data have been submitted to GEO (6), superSeries GSE22220. In addition, Affymetrix U133A-B/plus2 array data from previously published breast cancer cohorts were analysed (Table S2, N=1050 with RFS, N=592 with RFS outcome).

Statistical methods
Workflow with study design is provided in Fig. S1. Analyses were implemented and performed in R v2.9.0, and Sweave (7) was used for Automatic Generation of Reports.

*Penalized regression* allows identification of microRNAs whose expression is prognostic, or associated with a clinico-pathological factor, independently from other covariates. Genomic datasets are characterized by a greater number of variables than samples and high structure; therefore we used L1- and L2-penalized regression to enable efficient variable selection and encourage a grouping effect (8). Penalization parameters were optimized using cross-validation, and leave-one-out was used to test the models. *Penalized Cox regression* was performed in two steps:

1- microRNAs associated with DRFS were selected using penalized Cox regression including all microRNAs on the array (Suppl. Information Eq. SM1).

2- Selected microRNAs were analysed further to assess whether they were prognostic independently from known clinico-pathological factors. Covariates considered were: microRNA expression, clinical factors, treatment and gene signatures of biological processes (Suppl. Information Eq. SM2).

Only microRNAs selected by both steps were considered independently prognostic of DRFS. *Penalized linear regression analysis* was used to study association of microRNA expression with clinico-pathological factors (Suppl. Information Eq. SM3).

Meta-analysis of published cohorts was performed as previously described (9); datasets are summarized in Table S2 and selection criteria are provided in Suppl. Information.

*Global expression analysis of microRNA-target relationship*

An extremely small number of microRNA targets have been validated previously, thus analyses of microRNA-target relationship relies on in-silico predictions. Simultaneous use of multiple prediction algorithms has been suggested (10) and to avoid bias resulting from underlying cor-
relations, we considered the union of all predicted targets from 6 major algorithms (further details in Suppl. Information). We expected that if a microRNA were functional, a significant proportion of its predicted targets would be downregulated. Due to the challenges involved, three methods were compared:

1) **Cumulative Relative Risk (RR) plots.** All transcripts present on the array were ranked based on the correlation of their expression with that of the microRNA under study. The Relative Risk (RR) was defined as the probability of observing, at each correlation level, a given number of predicted targets amongst genes whose expression is inversely correlated with that of the microRNA, with respect to genes with positive correlation (Suppl. Information Eq. SM4). If targets are regulated by the microRNA, RR>1. These plots were also used to study association of microRNA targets with clinical factors (Suppl. Information, Eqs. SM5-6).

2) **Predicted target signature (PTSign) score.** In each sample, the inverse of the weighted average expression of all predicted targets was calculated. Weights were the in-silico prediction scores (Suppl. Information, Eq. SM7).

3) **Regulatory effect (RE) score.** For each sample, transcripts were ranked by their expression level. The difference of the average rank between microRNA target and non-target transcripts was calculated as described in (11).

High PTSign or RE-scores indicate global target downregulation.

**RESULTS**

**microRNAs independently prognostic of distant relapse in breast cancer**

A two-step Cox analysis accounting for clinical, pathological and molecular features was used (Fig. 1 and S2). Specifically, prognostic microRNAs were first selected in a penalized Cox analysis where all microRNAs were tested simultaneously within the same model. Selected
microRNAs were then assessed to test prognostic capability independent from clinical covariates (patient age, tumor size and grade, nodal involvement, ER status, Tamoxifen and Chemotherapy treatment) and key biological processes as measured by gene expression signatures derived from large cancer cohort studies. These were: proliferation, ESR1 and HER2 signaling (12, 13), hypoxia (9), stem cell (14), invasion, immune-response and apoptosis (13). This process identified 3 novel and 1 known (miR-128a) prognostic microRNA in ER⁺ breast cancer, and 5 novel and 1 known (miR-210) in ER⁻ breast cancer (Fig. 1). Specifically, high miR-767-3p, -128a and -769-3p expression was associated with poor prognosis, and high miR-135a with good prognosis in ER⁺ cases; high miR-27b, -144, -210 were associated with poor prognosis, and high miR-342, -150 and -30c with good prognosis in ER⁻ cases. Full results are reported in Fig. S2A-G. Three further microRNAs, miR-29c, -642 (high values, good prognosis) and -548d (high values, poor prognosis) were identified in an analysis including all samples (Fig. S2C). An analysis including intrinsic subtype classification (Table S1) produced similar results (data not shown); however, since the stability of this classification has been recently questioned (15) we have used the HER2/ER/proliferation signature classification for the main analysis (Fig. S2C-E).

The above analysis confirmed two known prognostic microRNAs from previous studies (-210 and -128a); however miR-7 and miR-516-3p, also prognostic in a previous study (4), could not be confirmed. In our hands, miR-7 was associated with DRFS in univariate Cox analysis but not in a model including all microRNAs, suggesting that this microRNA is not an independent prognostic factor; whilst miR-516-3p was not significantly associated with DRFS.

Also, real-time PCR was performed to measure expressions of four prognostic microRNAs, miR-210, miR-342, miR-144 and miR-27b relative to 3 RNU controls (Suppl. Information). Strength of correlation between real-time PCR and arrays results varied but was significant for
all four microRNAs (p<0.0001 in all cases); analysis using non-normalized CT values produced similar results (see Suppl. Information for discussion on PCR normalization).

microRNA signatures divided patients efficiently into good and poor prognosis groups when either a simple median split (Fig. 1E, G), or clustering with Bayesian Information Criterion (Fig. S2F) were considered, and importantly they performed well when tested on the left-out cases (Fig. S2B, G). Amongst microRNAs prognostic in ER− samples, five (miR-150, -342 good prognosis; and miR-210, -144, -27b, poor prognosis) were prognostic also in univariate analysis of triple negative (TRN) breast cancers (N=37, see legend of Table S1); and miR-342, 27b and 150 were also significant in a Cox analysis of this group including clinic-pathological factors and gene signatures (Fig. S2H).

Other independent prognostic factors were number of positive nodes (Fig. S2C-E) and tumor grade (Fig. S2C). In agreement with previous studies (9, 13), proliferation and hypoxia signatures were prognostic in ER+ cases (Fig. S2E); hypoxia, invasion and immune response signatures in ER− cases (Fig. S2D). Crucially, microRNAs identified by our analysis were prognostically independent of these signatures (Fig. S2C-E).

**Prognostic microRNA clusters**

miR-767-3p (prognostic in ER+ cases), and miR-27b and miR-144 (prognostic in ER− cases) are part of microRNA clusters. Analysis of clustered microRNAs revealed that miR-451, clustered with miR-144, was significantly associated with good prognosis in ER+ cases (Table S3); suggesting an independent role for these two clustered microRNAs in ER+ and ER− tumors. Conversely, the miR-24/27/23 cluster was consistently associated with poor prognosis, with all-but-one microRNA of this family (miR-27a/b, -23a/b, -24; but not -189) significantly associated with DRFS in univariate analysis of all cases, and all microRNAs but one (-23b) significant in analysis of ER− cases (Table S3). However, none of them were associated with prog-
nosis in ER\(^+\) cases, providing evidence that this family plays a specific role in ER\(^-\) breast cancer.

**Transcriptional regulation of microRNAs and analysis of pri-microRNA in independent cohorts**

We investigated whether prognostic microRNAs and their pri-microRNAs showed coordinated expression, and studied the prognostic potential of pri-microRNAs in independent cohorts (N=1050 with RFS, N=592 with DRFS information; for further details on these cohorts see Suppl. Information and Table S2). Transcripts containing pri-microRNA for 6 prognostic micro-RNAs could be mapped to the arrays and their expression found to be significantly correlated with that of the mature microRNA in our series (Spearman Rank Correlation Test p<0.05, microRNAs are listed in Table 1). This suggests transcriptional regulation of these microRNAs in breast cancer, rather than modulation of the mature microRNA due to changes in processing.

Amongst these, pri-miR-128a, 210, -29c, -342 and -548d showed significant association with prognosis in meta-analysis of these cohorts and a concordant effect with respect to the micro-RNA prognostic analysis in Fig. 1; specifically, pri-miR-128a, -210, -342 and -548d were significant for DRFS, and -128a, -210, -27b, -342 and -548d were significant for RFS (Table 1).

**Association of prognostic microRNAs with processes critical for breast cancer biology**

microRNA association with clinico-pathological factors was studied by multiple penalized linear regression (Fig. 2A-B). This confirmed, for example, the previously reported hypoxia regulation of miR-210 (Fig. 2C), but suggested that miR-210 expression is associated also with proliferation, ER positivity, grade and nodal invasion (Fig. 2B, C). Interestingly, association with hypoxia and proliferation was significant after accounting for grade and ER status (Fig. 2B and C). This highlights the usefulness of testing associations simultaneously by including clinical
and molecular covariates in the same regression model. miR-128a was found to be associated with ER positivity, both as measured by IHC and ESR1 expression signature (Fig. 2A).

Amongst the newly discovered prognostic microRNAs, a reported role for miR-150 in immunity (16) agrees with an association with immune response signature in our dataset (Fig. 2C). miR-135a was inversely correlated with the proliferation signature, in agreement with reports of negative regulation of cell growth in Hodgkin’s lymphoma (17). Finally, miR-27b was associated with the invasion signature (Fig. 2B), in agreement with its reported ability to promote invasion and angiogenesis (18).

A recent study in 20 inflammatory breast cancers (IBCs) (19), one of the most aggressive forms of locally advanced breast cancer, identified microRNAs belonging to families -29 (-29a), -30 (-30b) and -342 as down-regulated in IBC. This agrees with our findings of expression of microRNAs in these families (-29c, -30c, -342) being associated with good prognosis. In the same study, miR-548a-5p was found up-regulated in IBC, in agreement with high expression of the -548 family being associated with poor prognosis (-548d).

**Target expression is consistent with regulatory effect of prognostic microRNAs**

Downregulation of predicted target transcript was observed using cumulative Relative Risk (RR) plots for all prognostic microRNAs, with miR-128a, -144 and -210 providing the most consistent and significant results (Fig. 3A, S3). Furthermore, microRNA association with clinico-pathological factors (Fig. 2) was reflected by coordinated downregulation of cognate target mRNAs (Fig. 3B and S3). The most significant were miR-210 and -144 targets, strongly under-expressed in highly proliferative tumors (Fig. 3). In three cases (miR-144, -210 and 769-3p), results could be confirmed consistently by both PTsign- and RE-score methods (Table S4); however, PTsign- and RE-score were strongly correlated for all microRNAs (p<<0.00001 for all
microRNAs, correlation coefficient range: [0.7-0.95]). Contrarily to RR plots, these measure inhibition of target expression irrespective of amount of downregulation.

**A global analysis of microRNA targets reveals pathways deregulated in cancer**

A pathway analysis revealed that a number of downregulated predicted targets, greater than that expected by chance, were implicated in pathways which have important roles in direct tumor growth and metastasis (Fig. S4); and also that microRNA target downregulation would lead to induction of these pathways (Fig. S4A). For example, downregulated targets included MAPK8 and MAPK14 ((miR-144 and miR-27b targets respectively) which are pro-apoptotic under stress conditions (20); SPRY2 (miR-128a target) in the FGF receptor signaling pathway, which has a major role in inhibiting tyrosine kinases (21); and tumor suppressor genes PTEN and FOXO1 (miR-144 and -128a targets). miR-27b predicted targets in the metabotropic glutamate receptor pathway, namely GRIN3A, GRM6 and GRIA4, and voltage-dependent L-type calcium channel subunit, were consistently downregulated. Their association with neuronal cell death under hypoxic stress (22) suggests a new potential mechanism by which cancer cells escape death under stress. Recently, mutations in this pathway were reported as the commonest mutations in melanoma (23). Wnt antagonists such as miR-144 predicted target SRFP1, and targets involved in angiogenesis such as Frizzled4 were also amongst downregulated predicted targets, where excessive Frizzled4 has been associated with disrupted embryonic vasculature (24).

In TRN cancers, several signaling pathways related with miR-150 predicted targets were found up-regulated in poor prognosis cases including Akt2, insulin receptor, ErbB3, S6 kinase, MAP kinase pathways and stress response kinase JNK, as well as downstream protease MMP-13 (Fig. S4B). miR-342 predicted target RRM2 was also up-regulated in this group (Fig. S4B).
RRM2 is a target for several inhibitors of proliferation that affect cells in S-phase and would be compatible with the high proliferation rate of TRN cancer (6). Similar pattern was observed for miR-342 target glucose 1,6-bisphosphate synthase, a critical component of glycolysis (25) to which inhibitors are currently being developed; and TLE-1, a downstream component of notch signaling (26), that is already recognized to be activated in TRN cancers.

Validation of prognostic microRNAs using target expression in the present and independent cohorts

We focused on prognostic microRNAs showing consistent and significant downregulation of cognate target transcripts. Amongst these, miR-210, -128 and -27b predicted targets were prognostic for DRFS and RFS, both when using cumulative RR (Fig. S5) and summary scores in meta-analysis of published cohort studies (Fig. 4A-F). Furthermore, tumours with concomitant microRNA overexpression and target underexpression had significantly worse prognosis, while the opposite was true for cases with low microRNA levels and high target expression (Fig. 4G-I).

We performed also a single target analysis for experimentally-derived miR-210 targets (Fig. S4E). Three out of five targets showing most significant downregulation, namely ISCU, CBX7 and IGF1R, were significantly associated with DRFS; specifically, low levels were associated with poor prognosis (Fig. S4E). This suggests that novel targets can be confirmed using this approach. To further assess this potential we assessed protein expression using immunohistochemistry (IHC) for a predicted target with suitable commercial antibody available. NFE2L2 (alias NRF2) was chosen as top ranking downregulated predicted target of miR-144, novel prognostic microRNAs showing evidence of coordinated downregulation of target mRNAs (Fig. 1 and Table S4). IHC could be performed on 137 cases (Suppl. Information); staining was predominantly cytoplasmic with minimal stromal positivity (Fig. S6). Tumors with weak protein and
mRNA expression showed significantly higher miR-144 expression than tumours with consistently high NFE2L2 expression (Fig. S6). This was true only in ER− cases (Fig. S6D, E), in agreement miR-144 being prognostic only in ER− tumors (Fig. 1). Low NFE2L2 IHC product score (IPS) was associated with worse DRFS (univariate Cox HR=0.64, p=0.18; IPS ranked and normalized between 0 and 1); in agreement with NFE2L2 acting as a tumor suppressor (23). Furthermore, tumors with concomitant miR-144 overexpression and NFE2L2 underexpression had significantly worse prognosis, while the opposite was true for cases with low miR-144 levels and high NFE2L2 expression (Fig. S6F).

**microRNAs are prognostic independent of expression of microRNA-processing genes**

We found that the expression of several, but not all, pri-microRNAs was correlated with that of the mature microRNAs. Thus, we investigated whether the microRNA-processing genes had a role in the regulation of mature microRNA levels. We tested the prognostic significance of several processing genes. High expression of transport gene exportin 5 (XPO5) and RISC complex genes **EIF2C2** and **EIF2C3** were significantly associated with poor prognosis, whilst high expression of DICER genes **DICER1** and **TARBP1**, and RISC complex gene **EIF2C4** were associated with good prognosis (Table 2). Of these, expression of XPO5 showed a significant correlation with mature microRNA expression both for the ER− and ER+ microRNA prognostic signatures (Fig. 5 A-B), whilst **EIF2C2** and **EIF2C3** showed correlation only with the ER− microRNA prognostic signature. However, in all cases fold expression changes were very small (Fig. 5 A-B). When samples were stratified by low and high risk based on microRNA signature and expression of processing genes, the effect of the latter was never significant (Fig. 5 C-F). A Cox analysis including single prognostic microRNAs and processing genes showed that in all cases, mature microRNA expression was significant independent of the expression of the
processing genes (Table S5). Overall these results suggest that the prognostic significance of these microRNAs is due to transcriptional regulation rather than differential processing. For miR-128a, -210, -342 and -548d, this is also confirmed by the prognostic significance of their pri-microRNA expression in independent cohorts (Table 1).

Interestingly, EIF2C2 (AGO2) in our samples, although mildly differentially regulated, was always very highly expressed (Fig. 5), thus allowing mature microRNAs to modulate target gene expression. This is particularly important in view of recent work (11) suggesting that AGO-expression is mandatory for an effect of microRNAs on the expression profiles of tumor samples.

**DISCUSSION**

In this study microRNAs that are independently prognostic for DRFS were identified and validated by integrated analysis of microRNA and mRNA profiling. Nine microRNAs were found prognostic in ER+ and ER– breast cancer, three of which remained significant in the clinically challenging TRN group. Our findings regarding mature miR-210 and miR-128a expression confirm previous findings of a smaller study (4). Multiple penalized linear regression revealed that prognostic microRNAs are associated with key biological processes in breast cancer, such as proliferation (e.g. miR-135a), hypoxia (e.g. miR-210, -342), invasion (e.g. miR-27b) and immune response (e.g. miR-150). However, these microRNAs are prognostic independent of gene expression signatures of these processes. This highlights the importance of considering both microRNA and transcript expression data in prognostic studies.

Several microRNAs previously linked with breast cancer subtype or progression in experimental models were not identified as independently prognostic. This agrees with previous results from prognostic studies (4, 27); reasons could include discrepancy of assays and/or the impor-
tance of microRNA expression in specific cell populations which would not appear in whole tumor sample analyses. However, it also demonstrates the need to differentiate between biomarkers of tumor presence and prognostic biomarkers, and identify factors carrying information independently rather than surrogates of known prognostic covariates.

To examine whether microRNAs might be prognostic due to regulatory effect on cognate targets, associations between microRNAs and predicted targets were studied by global expression analysis. This is challenging, as the microRNA-mRNA interaction network is complex and the effect on each individual mRNA is often small. Nevertheless, prognostic miR-128a, -27b and -210 showed evidence of cognate target downregulation, and expression of their targets was prognostic in meta-analysis of several cohorts (Fig. 4). Furthermore, combined use of microRNA and target expression identified cases with the poorest prognosis, suggesting that a combined score could reflect not only expression but also functionality of the microRNA.

Prognostic microRNAs could be used not only to select patients for specific interventions, but also to define therapeutic approaches. To this end, functional validation of microRNA targets is needed to elucidate the role of microRNAs in cancer. In this respect, analysis of microRNA-mRNA relationships in clinical datasets could assist in target prioritization, and cohorts such as the present one will be a useful resource for future validation studies. Prognostic microRNAs were found to regulate a wide range of previously poorly investigated pathways that may be related to cancer progression and potential candidates for therapy, such as RRM2 and TLE-1 in TRN breast cancers. A validated example was miR-210 target ISCU (iron-sulfur cluster scaffold homolog), which was downregulated in samples with high miR-210 levels, and whose underexpression was associated with both hypoxia and poor prognosis. ISCU plays an important role in mitochondrial respiration and DNA repair (28, 29). Tumors with high miR-210 might be for example more susceptible to DNA damaging agents combined with DNA repair inhibitors.
However, ISCU could not be identified in traditional supervised gene expression prognostic analyses.

An additional benefit of merging microRNA and mRNA data was that coordinate expression of microRNAs and precursor-containing pri-microRNAs, and expression of microRNA-processing genes could be explored. As this analysis suggested that several prognostic microRNAs are regulated at the transcriptional level rather than through changes in RNA processing mechanisms (Fig. 5), pri-microRNA data (Table 1) could be used for the validation of our findings in cohorts for which only mRNA profiling is available. The most striking results were for miR-210 and -128a, where coordinated expression of pri-microRNA (up-stream of microRNA expression), mature microRNA and predicted targets was observed. Expression of each one of these components was found to be prognostic.

In conclusion, this is the first large study integrating microRNA and mRNA global profiles in human breast cancer. Prognostic microRNAs in ER^+ and ER^- breast cancer were identified, and regulatory action on target transcripts demonstrated. This results not only elucidated potential novel therapeutic targets, but could also be exploited to validate findings in independent cohorts. Our approach consistently validated known microRNA-target interactions, and may therefore be broadly applicable to other biologically and clinically heterogeneous diseases.

Acknowledgements

We thank Helen Turley for assistance with immunohistochemistry, Dr. Russell Leek (tissue banking, Oxford) and the Computational Biology Research Group (hardware and services). This work was supported by:- Oxford NIHR Comprehensive Biomedical Research and Experimental Cancer Medicine Centres; Cancer Research UK (A.L.H., M.T., H.S., C.S.); EU 6th and
7th Framework Programs (F.M.B. and A.L.H.); Breast Cancer Research Foundation and Friends of Kennington Cancer Fund (A.L.H.); Rhodes Trust Scholarship to H.G.; Wellcome Trust (J.R., C.C, L.W.).

References

<table>
<thead>
<tr>
<th>pri-microRNA</th>
<th>host transcript (Illumina, Agilent and GenBank IDs)</th>
<th>Summary [95% CI]</th>
<th>HR 95% CI</th>
<th>p 2-sided</th>
<th>N 2-sided</th>
</tr>
</thead>
<tbody>
<tr>
<td>miR-128a</td>
<td>GI_31543534 (202754_at)</td>
<td>3.20</td>
<td>0.002</td>
<td>0.0117</td>
<td>1.117</td>
</tr>
<tr>
<td></td>
<td>BC041093 (233599_at)</td>
<td>3.08</td>
<td>0.026</td>
<td>0.04778</td>
<td>1.117</td>
</tr>
<tr>
<td>miR-210</td>
<td>GI_23071083 (228228_at)</td>
<td>2.94</td>
<td>0.0222</td>
<td>0.1162</td>
<td>3.06</td>
</tr>
<tr>
<td></td>
<td>GI_23071083 (228228_at)</td>
<td>2.89</td>
<td>0.0222</td>
<td>0.1162</td>
<td>3.06</td>
</tr>
<tr>
<td>miR-27b</td>
<td>GI_1145697 (22215_s_at)</td>
<td>2.49</td>
<td>0.0272</td>
<td>0.1195</td>
<td>0.133</td>
</tr>
<tr>
<td></td>
<td>GI_1145697 (22215_s_at)</td>
<td>2.43</td>
<td>0.026</td>
<td>0.04778</td>
<td>1.117</td>
</tr>
<tr>
<td>miR-29c</td>
<td>GI_211797_at</td>
<td>2.43</td>
<td>0.0272</td>
<td>0.1195</td>
<td>0.133</td>
</tr>
<tr>
<td></td>
<td>GI_211797_at</td>
<td>2.40</td>
<td>0.026</td>
<td>0.04778</td>
<td>1.117</td>
</tr>
<tr>
<td>miR-30c</td>
<td>GI_211797_at</td>
<td>2.40</td>
<td>0.026</td>
<td>0.04778</td>
<td>1.117</td>
</tr>
<tr>
<td></td>
<td>GI_211797_at</td>
<td>2.34</td>
<td>0.0234</td>
<td>0.0403</td>
<td>0.054</td>
</tr>
<tr>
<td>miR-30e-3p</td>
<td>GI_211797_at</td>
<td>2.34</td>
<td>0.0234</td>
<td>0.0403</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>GI_211797_at</td>
<td>2.34</td>
<td>0.0234</td>
<td>0.0403</td>
<td>0.054</td>
</tr>
<tr>
<td>miR-342</td>
<td>GI_211797_at</td>
<td>2.34</td>
<td>0.0234</td>
<td>0.0403</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>GI_211797_at</td>
<td>2.34</td>
<td>0.0234</td>
<td>0.0403</td>
<td>0.054</td>
</tr>
</tbody>
</table>

**Notes:**

- **pri-microRNAs** shown that could be mapped to the arrays and showed significant correlation with microRNA expression in the Oxf dataset or a subset of 69 samples where Affymetrix arrays were available (Spearman Rank Correlation Test p<0.05).
- **GSE6532 overlaps partially (N=69) with the present dataset. This was exploited to establish the correlation between microRNA and pri-microRNA expression as measured by Affymetrix arrays; however the common cases were taken out from GSE6532OXF in all prognostic analysis where this dataset is used and also in Fig. 4.
- **Univariate Cox analysis** Hazard Ratio (HR) and 2-sided significance (p) provided. Expression was considered as continuous variable (ranked, normalized between 0 - 1).
- **Summary HR and 95% confidence interval** for the meta-analysis. Significant results are in bold.

---

**Table 1. Analysis of microRNA host-genome (pri-microRNA) expression in published cohorts (N=1050 for RFS; N=592 for DRFS)**

<table>
<thead>
<tr>
<th>HR</th>
<th>95% CI</th>
<th>p 2-sided</th>
<th>N 2-sided</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.34</td>
<td>[1.43,4.32]</td>
<td>0.0117</td>
<td>1.117</td>
</tr>
<tr>
<td>2.34</td>
<td>[1.43,4.32]</td>
<td>0.0117</td>
<td>1.117</td>
</tr>
<tr>
<td>2.34</td>
<td>[1.43,4.32]</td>
<td>0.0117</td>
<td>1.117</td>
</tr>
<tr>
<td>2.34</td>
<td>[1.43,4.32]</td>
<td>0.0117</td>
<td>1.117</td>
</tr>
</tbody>
</table>

---

**Table 2. Analysis of miRNA expression in published cohorts (N=1050 for RFS; N=592 for DRFS)**

<table>
<thead>
<tr>
<th>miR</th>
<th>HR</th>
<th>95% CI</th>
<th>p 2-sided</th>
<th>N 2-sided</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.34</td>
<td>[1.43,4.32]</td>
<td>0.0117</td>
<td>1.117</td>
<td></td>
</tr>
<tr>
<td>2.34</td>
<td>[1.43,4.32]</td>
<td>0.0117</td>
<td>1.117</td>
<td></td>
</tr>
<tr>
<td>2.34</td>
<td>[1.43,4.32]</td>
<td>0.0117</td>
<td>1.117</td>
<td></td>
</tr>
<tr>
<td>2.34</td>
<td>[1.43,4.32]</td>
<td>0.0117</td>
<td>1.117</td>
<td></td>
</tr>
</tbody>
</table>
Table 2. Prognostic significance of microRNA processing genes.

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Function</th>
<th>Probeset used</th>
<th>Hazard Ratio(^a)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TARBP1</td>
<td>DICER</td>
<td>GI_19743835-S</td>
<td>0.32</td>
<td><strong>0.0041</strong></td>
</tr>
<tr>
<td>DICER1</td>
<td>DICER</td>
<td>GI_29294648-I</td>
<td>0.44</td>
<td><strong>0.0388</strong></td>
</tr>
<tr>
<td>EIF2C4 (AGO4)</td>
<td>RISC</td>
<td>GI_29029592-S</td>
<td>0.45</td>
<td><strong>0.0420</strong></td>
</tr>
<tr>
<td>ESR1</td>
<td>p68 and p72 interaction</td>
<td>GI_4503602-S</td>
<td>0.49</td>
<td>0.0657</td>
</tr>
<tr>
<td>ARS2</td>
<td>Regulation of microprocessor</td>
<td>GI_33383230-A</td>
<td>0.74</td>
<td>0.4313</td>
</tr>
<tr>
<td>PRKRA</td>
<td>DICER</td>
<td>GI_32261293-S</td>
<td>0.86</td>
<td>0.7036</td>
</tr>
<tr>
<td>DGCR8</td>
<td>Microprocessor</td>
<td>GI_38488719-S</td>
<td>0.89</td>
<td>0.7642</td>
</tr>
<tr>
<td>EIF2C1 (AGO1)</td>
<td>RISC</td>
<td>GI_29171732-S</td>
<td>0.90</td>
<td>0.7667</td>
</tr>
<tr>
<td>RNASEN</td>
<td>Microprocessor</td>
<td>GI_21359821-S</td>
<td>1.06</td>
<td>0.8757</td>
</tr>
<tr>
<td>TARBP2</td>
<td>DICER</td>
<td>GI_19743837-A</td>
<td>1.24</td>
<td>0.5725</td>
</tr>
<tr>
<td>ADARB1</td>
<td>Editing of specific microRNAs</td>
<td>GI_7669476-I</td>
<td>1.27</td>
<td>0.5206</td>
</tr>
<tr>
<td>RAN</td>
<td>Transport</td>
<td>GI_6042206-S</td>
<td>1.34</td>
<td>0.4535</td>
</tr>
<tr>
<td>ESR2</td>
<td>p68 and p72 interaction</td>
<td>GI_10835012-S</td>
<td>1.51</td>
<td>0.2767</td>
</tr>
<tr>
<td>TRIM32</td>
<td>Binding to miRISC, enhancing microRNA activity</td>
<td>GI_15208649-S</td>
<td>1.78</td>
<td>0.1326</td>
</tr>
<tr>
<td>EIF2C3 (AGO3)</td>
<td>RISC</td>
<td>GI_29337285-A</td>
<td>2.46</td>
<td><strong>0.0244</strong></td>
</tr>
<tr>
<td>EIF2C2 (AGO2)</td>
<td>RISC</td>
<td>GI_29171733-S</td>
<td>2.63</td>
<td><strong>0.0128</strong></td>
</tr>
<tr>
<td>XPO5</td>
<td>Transport</td>
<td>GI_22748936-S</td>
<td>3.11</td>
<td><strong>0.0035</strong></td>
</tr>
</tbody>
</table>

\(^{a}\)Cox Univariate analysis. Gene expression considered as continuous variable, ranked and normalized between 0 and 1. When more than one probeset mapped to the same gene, the probeset with the greater effect is shown.
Figure Legends

Figure 1. microRNAs independently prognostic of Distant Relapse-Free Survival (DRFS) in breast cancer (N=207). Prognostic microRNAs were selected by two-step penalized Cox regression (see Methods). A-C) First a model including all microRNAs (x-axis) was considered (full results in Fig. S2A). Heatmaps display the cross-validated model at each leave-one-out iteration (y-axis) for analyses in all 207 samples (A), 82 ER− (B) and 90 ER+ Tamoxifen-treated (C). Similar models are clustered (standard correlation). Colour reflects the Hazard Ratio (logged and rescaled per column, see bar). microRNAs that were selected in at least 5% of iterations are shown. Consistently selected microRNAs (> 95% iterations) were further analysed using a Cox model including clinico-pathological factors and gene expression signatures (D-G). microRNAs consistently significant also in this model were considered prognostic and are indicated with a star (red=not known, black=previously known to be prognostic). Summary results are shown for ER+ Tamoxifen-treated (D) and ER− (F) samples (full results in Fig. S2C-E). miR-768-3p is in parenthesis as it was retired from miRBase. Kaplan-Meier plots of prognostic microRNA signature in ER+ (E) and ER− (G) cases are also shown. Expression of each prognostic microRNA (stars from D and F respectively) was ranked. In each case, the mean rank of poor prognosis microRNAs was subtracted from that of good prognosis microRNAs to provide a non-parametric summary score. Cases were categorized as high and low risk by median value of this score. Boxplots of the microRNA ranked expression in the two groups show median, quartiles and range.

Figure 2. Biological processes and clinico-pathological factors associated with prognostic microRNAs. Expression signatures derived from large-scale analysis of cancer
datasets were used as surrogate markers of biological processes; samples were ranked from lowest to highest as measured by summary scores of these signatures. microRNAs prognostic in ER+ (A) and ER- (B) cases are shown. Association between expression of each microRNA (y axis) with clinico-pathological variables and gene signatures (x axis) was obtained using penalized linear regression. Heatmaps illustrate the microRNA association with signature/clinical variable in the cross-validated model (see bar). Red = covariate levels high, microRNA expression high; blue = opposite. C) Scatter plots with examples of significant associations. MicroRNA expression (X-axis) and covariate (Y-axis) are plotted with a second covariate shown as colour stratification. Linear fits and Spearman Rank Correlation are shown in the two strata. HS up= Hypoxia Signature Score above median, HS down= below median. N+= Nodal involvement, N0= no involvement.

**Figure 3. Inhibition of cognate target expression by prognostic microRNAs.** Cumulative Relative Risk (RR) plots are shown for the most significant effects; further results in Fig. S3. A) Plots show enrichment in the number of predicted targets amongst transcripts whose downregulation is associated with microRNA overexpression (Fig. S3 for workflow of this analysis). At each correlation level r, the relative risk (RR) of observing a given number of predicted targets whose expression is anti-correlated with microRNA expression, with correlation of -r or lower, is plotted. Number of downregulated targets is shown on the plot. Dotted lines represent the mean RR for randomly sampled (x1000) set of transcripts, with 5% and 95% confidence intervals. B) RR plots of enrichment in the number of predicted targets amongst transcripts whose expression is anti-correlated with gene expression signatures of biological processes. Proliferation was considered
as biological process with strongest association with microRNA expression (Fig. 2 A).

RR plot statistics as described in A.

**Figure 4. Validation of prognostic microRNAs in independent breast cancer datasets by analysis of cognate targets. A-F)** Forest plots are shown for miR-210 (A, D), -27b (B, E) and -128a (C, F) predicted targets (PTSign summary score, see Methods) in the present (Oxf) and published BC datasets (GEO IDs provided, further details in Table S2). Dots represent Hazard Ratios (HR); dimensions are proportional to dataset size. Grey bars are 95% confidence intervals. DRFS= Distant Relapse-Free Survival (A-C); RFS= Recurrence-Free Survival (D-F). **G-I)** Kaplan-Meier curves for expression of mature miR-210 (G), -27b (H) and -128a (I), and respective PTSigns, in the Oxford breast cancer dataset (N=207). microRNA expression levels and PTSign score were split by median value (−=below, + =above). HR for deviation contrasts in a Cox regression comparing each category with the overall effect are shown for significant comparisons.

**Figure 5. Expression of microRNA-processing genes does not affect microRNA prognostic significance. A-B)** Boxplots of expression of microRNA-processing genes that were prognostic in Cox univariate analysis (Table 2) are shown for the Low Risk Profile (LRP) and High Risk Profile (HRP) in ER+ (A) and ER– (B) samples as defined in Fig. 1E, G legend. Fold changes in mean expression (FC) and non-parametric Mann-Whitney Test are shown for significant cases. **C-F)** Kaplan-Meier plots of distant relapse-free survival. Samples are stratified into the LRP and HRP, and also by median value of processing genes’ expression; [-] below, [+] above median. This value is calculated using all the samples, thus the expression cut-point is equal in both LRP and HRP.
group. The Hazard Ratio (HR) of LRP and HRP groups was always significant, whilst
the HR of [-] and [+\] groups was never significant (Cox model contrasts, threshold
p<0.05).
Figure 4

A

miR-210 targets

Ox
GSE6532GUY
GSE6532Ki
GSE6532OXF
GSE9195

Summary estimate [N=592]

0.1 1 5 10
Hazard Ratio for DRFS

B

miR-27b targets

Ox
GSE6532GUY
GSE6532Ki
GSE6532OXF
GSE9195

Summary estimate [N=592]

0.1 1 5 10
Hazard Ratio for DRFS

C

miR-128a targets

Ox
GSE6532GUY
GSE6532Ki
GSE6532OXF
GSE9195

Summary estimate [N=592]

0.1 1 5 10
Hazard Ratio for DRFS

D

miR-210 targets

Ox
GSE1456
GSE2034
GSE6532GUY
GSE6532Ki
GSE6532OXF
GSE9195

Summary estimate [N=1050]

0.1 1 5 10
Hazard Ratio for RFS

E

miR-27b targets

Ox
GSE1456
GSE2034
GSE6532GUY
GSE6532Ki
GSE6532OXF
GSE9195

Summary estimate [N=1050]

0.1 1 5 10
Hazard Ratio for RFS

F

miR-128a targets

Ox
GSE1456
GSE2034
GSE6532GUY
GSE6532Ki
GSE6532OXF
GSE9195

Summary estimate [N=1050]

0.1 1 5 10
Hazard Ratio for RFS

G

miR-210-PTSsign

miR-210+PTSsign

miR-210+/PTSsign-

miR-210+/PTSsign+

0 1 2 3 4 5 6 7 8 9 10 11
Time (years)

H

miR-27b-PTSsign

miR-27b+PTSsign

miR-27b+/PTSsign-

miR-27b+/PTSsign+

0 1 2 3 4 5 6 7 8 9 10 11
Time (years)

I

miR-128a-PTSsign

miR-128a+PTSsign

miR-128a+/PTSsign-

miR-128a+/PTSsign+

0 1 2 3 4 5 6 7 8 9 10 11
Time (years)
microRNA associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer

Francesca M Buffa, Carme Camps, Laura Winchester, et al.

Cancer Res  Published OnlineFirst July 7, 2011.