Profiles of miRNAs matched to biology in aromatase inhibitor resistant breast cancer

Supplementary Material

MATERIALS AND METHODS

Statistical and Bioinformatic Analyses

Extended statistical analyses were performed with R-3.2.2 including additional packages gplots-2.17.0 [1], partykit-1.0-3 [2], survival-2.38-3 [3] and edgeR-3.10.2 [4]. Open-access miRNA-Seq level 3 isoform quantification data (version 3.1.17.0) and clinical information (version 2.0.28.0) from The Cancer Genome Atlas (TCGA) breast cancer cohort (BRCA) were downloaded from TCGA data portal (https://tcga-data.nci.nih.gov/tcga/) on May 11, 2015. For 746 tumor samples both clinical data and miRNA expression data were available. edgeR was used to calculate the logarithmized counts per million (CPM) values as a measure of expression. To adjust for varying sequencing depth between miRNA-Seq libraries, raw read counts were normalized by edgeR using the trimmed mean of M values approach. By means of sequence identity, miRNAs covered by the Affymetrix miRNA2.0 chip and miRNAs measured by TCGA miRNA-Seq were associated to each other. TCGA breast cancer samples were previously classified using the PAM50 model [5]. PAM50 subtype assignments of 513 tumors were obtained from the Cancer Browser platform (https://genome-cancer.ucsc.edu/) (dataset: TCGA_BRCA_exp_HiSeqV2, version 2015-02-24). KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway data were downloaded from http://www.kegg.jp/kegg/rest/ on June 11, 2015. Full and slim set of Gene Ontology annotations for human were downloaded from http://geneontology.org/page/downloads on September 17, 2015. RefSeq mRNA sequences (genome assembly hg19) were downloaded from UCSC Genome Browser (https://genome.ucsc.edu/cgi-bin/hgTables) on April 29, 2014. Processed CLIP-Seq data provided by the starBase v2.0 web-server (http://starbase.sysu.edu.cn/) were downloaded on April 30, 2014. The miRBase database (version 20) was downloaded on December 14, 2014.
miRNA expression in PAM50 subtypes: Based on raw read counts, miRNAs significantly differentially expressed between PAM50 subtypes were determined using non-parametric Kruskal-Wallis tests as well as pairwise Wilcoxon-Mann-Whitney post-tests.

Survival analyses: Overall survival was used as endpoint in all survival analyses. Associations between survival and miRNA expression was explored using the conditional inference tree framework from the R package partykit. The method recursively partitions the samples in subsets with significantly differing survival functions. The $P$-value cut-off to accept a split was set to 0.1. The significance of the association between groups of samples and overall survival was analyzed by log-rank tests. Holm correction procedure [6] was applied to adjust resulting $P$-values for multiple testing.

Identification of miRNA sets: Putative functions of miRNA sets were identified via in silico functional enrichment analysis of the respective miRNA targets. The population of miRNAs consisted of 1,067 distinct miRNA that were both represented on the array and registered in miRBase version 20. We conducted a transcriptome-wide search on RefSeq mRNA sequences for miRNA targets using the TargetScan software (version 6.2). In order to increase the reliability of target predictions, predicted target sites were overlapped with experimentally identified miRNA-target interaction sites. The starBase v2.0 web-server provides a library of RNA-binding protein interaction sites collected from different CLIP-Seq experiments. Here, we used the subset of 36 CLIP-Seq experiments conducted within 11 different AGO-RNA interaction studies. Inclusion criteria for the definition of target sites were: (A) The target site had to be confirmed by at least one CLIP-Seq study and (B) the percentile value calculated by TargetScan to facilitate comparison of target lists of different miRNAs had to be above 75. A gene was considered to be a target of a certain miRNA if at least one of its associated transcripts contained a minimum of one accepted target site of that miRNA.

Over-represented functional categories are usually identified by comparing the observed overlap between category and miRNA target set with the expected overlap estimated from the background gene set. Notably, the definition of the background gene set is of crucial importance. It is well known, that miRNA targets are enriched for certain functions compared to non-miRNA targets [7,8] and therefore, the
background gene set was compiled in such a way to include only predicted miRNA targets. This is critical in order to avoid a bias in the calling of functional terms. The resulting miRNA-gene interaction map included 1,067 miRNAs and 11,795 genes connected via 263,451 interactions.

**KEGG pathway analyses:** We based our pathway analyses on the number of interactions between miRNAs and pathway genes. The interaction map was filtered to include genes for which KEGG pathway annotations were available. This resulted in a background interaction map of 1,065 miRNAs, 4,330 genes and 99,246 interactions. Fisher's exact test was used to test if the number of observed interactions between genes of a certain pathway and a miRNA set significantly exceeded the number of expected interactions as estimated from the background interaction map. Additionally, we applied an approach in line with Bleazard et al. [9] using permutation tests for the identification of significantly associated pathways. Here, the number of interactions between a miRNA set and a pathway was compared with the number of interactions obtained from a random set of miRNAs of the same size as the input set sampled (without replacement) from the background miRNA set. Sampling was repeated 10,000 times and a permutation $P$-value was calculated using the proportion of random miRNA sets that exhibited equal or more interactions with the pathway genes. The overlap between gene sets from enriched KEGG pathways ($P < 0.05$) were illustrated as graph diagrams executed with yED Graph Editor (http://www.yworks.com/products/yed). Two nodes, representing two different KEGG pathways were connected if their intersection was significantly greater than expected (one-sided Fisher's exact test, correction for multiple testing by Holm's method). Importantly, non-miRNA targets among the pathway genes were not considered for the network construction. The node size refers to the pathway size in terms of the number of miRNA targets. The edge weight illustrates the degree of similarity between two pathways as measured by the Jaccard index. General KEGG pathways of the category "global and overview maps" and "pathways in cancer" were neglected.

**Gene Ontology (GO) term enrichment:** Over-representation of GO terms were analyzed for the sub-category ‘Biological process’. Target gene lists generated from the overlap between gene sets from significantly enriched KEGG pathways were searched for enriched GO terms using the GORILLA tool
The background gene set was defined as all genes with a GO annotation. GO terms were summarized by removing redundant terms using REVIGO (http://revigo.irb.hr/) using default parameters.

REFERENCES

1. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, Maechler M, Magnusson A, Moeller S, Schwartz M, Venables B. gplots: Various R programming tools for plotting data: R package version 2.17.0. http://CRAN.R-project.org/package=gplots. 2015.

2. Hothorn T, Zeileis A. partykit: A modular toolkit for recursive partytioning in R. Journal of Machine Learning Research. URL http://EconPapers.RePEc.org/RePEc:inn:wpaper:2014-10. 2015.

3. Therneau TM, Grambsch PM. Modeling survival data: Extending the cox model. New York: Springer; 2000.

4. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26: 139-40.

5. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012; 490: 61-70.

6. Holm S. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics. 1979; 6: 65-70.

7. Cui Q, Yu Z, Pan Y, Purisima EO, Wang E. MicroRNAs preferentially target the genes with high transcriptional regulation complexity. Biochem Biophys Res Commun. 2007; 352: 733-8.

8. Godard P, van EJ. Pathway analysis from lists of microRNAs: common pitfalls and alternative strategy. Nucleic Acids Res. 2015; 43: 3490-7.

9. Bleazard T, Lamb JA, Griffiths-Jones S. Bias in microRNA functional enrichment analysis. Bioinformatics. 2015; 31: 1592-8.
SUPPLEMENTARY TABLES

Supplementary Table 1: miRNAs differentially expressed in MCF-7:2A vs. MCF-7:WS8 cells

Supplementary Table 2: miRNAs differentially expressed in MCF-7:5C vs. MCF-7:WS8 cells

Supplementary Table 3: miRNAs differentially expressed in MCF-7:5C vs. MCF-7:2A cells

Supplementary Table 4: miRNAs differentially expressed in MCF-7:WS8 on E2 treatment (72h). MCF-7:WS8 vs MCF-7:WS8_E2 cells

Supplementary Table 5: Basal expression levels of stem cell-like and EMT-like genes in MCF-7:5C and MCF-7:2A cells
**Supplementary Figure 1: miRNAs associated with E2-stimulated growth.** Differentially expressed miRNAs of the basal MCF-7:WS8 phenotype following 72h of E₂ exposure (WS8_E₂ versus WS8) are compared to pairwise comparisons of 2A versus WS8 cells and 5C versus WS8 cells to identify growth-related miRNAs. The Venn diagram intersections highlight the 71 WS8 specific miRNAs (14+33+24) subject to changes during E₂ exposure. Thirty-three miRNAs overlap with the basal AI resistance phenotypes 2A and 5C.
Supplementary Figure 2: Basal levels of expression of candidate miRNAs between the AI-resistant (5C, 2A) and WS8 cells. Relative miRNA expression was quantified by RT-PCR using RNU44 for normalization. Results are shown for miR-221/222, -31, -127-3p, -409-3p and -432-5p with experiments done in triplicates *: $P < 0.05$, **: $P < 0.001$. 
B.

eyellow, orange: 2A specific miRs

miR-29a  miR-29c*  miR-30c  miR-502-3p  miR-664*

miR-671-5p  miR-30a

blue, purple: 5C specific miRs

miR-210  miR-182  miR-148a*  miR-501-5p  miR-20a

miR-17  miR-3200  miR-27a*  miR-590-5p  miR-937

miR-497  miR-221  miR-218  miR-149  miR-222

miR-195  miR-342-3p  miR-1307  miR-140-3p  miR-199a-3p
Supplementary Figure 3: miRNA expression in PAM50-defined breast tumor subtypes from TCGA.

Box-Whisker plots indicate expression differences between the subgroups Basal-like (Basal, n=86), HER2 positive (Her2, n=43), Luminal A (LumA, n=253), Luminal B (LumB, n=115) and Normal breast-like (Normal, n=16). **A.** Expression levels of chromosome 14q32.31 miRNAs show differences for 19 miRNAs in the various subgroups (Holm-adjusted $P < 0.05$). **B.** Expression levels of the remainder miRNAs in the breast cancer subgroups. Plots are sorted according to the 2A-specific, 5C-specific and the
2A-5C-intersection miRNAs (see color code in Figure 2A and B). Y-axes: log2CPM (CPM: counts per million). *: $P < 0.05$, **: $P < 0.01$, ***: $P < 0.001$ and ****: $P < 0.0001$. 
Supplementary Figure 4: Kaplan-Meier curves of overall survival of breast cancer patients of TCGA stratified by miRNA expression. MiRNA expression cut-offs were determined by conditional inference tree models. High (red) and low (green) expression are indicated on the log2CPM scale (CPM: counts per million). **A.** Analyses across all 746 patients showed associations between high expression of miR-410, -381, -485-5p, -487a, -376c, -411, and -127-3p and favorable outcome as compared to low expression, respectively. **B.** Analyses across 253 patients with Luminal A tumors show associations of high expression of miR-431 and -654-3p with favorable outcome as compared to low expression, respectively. In contrast, low expression of miR-493* and -505* was associated with favorable outcome compared to high expression, respectively. Holm correction was applied to adjust P-values for multiple testing.
### miRs 5C_up

| Pathway                                                                 | MCF-7:5C | MCF-7:2A |
|-------------------------------------------------------------------------|----------|----------|
| Circadian rhythm                                                        |          |          |
| Choline metabolism in cancer                                            |          |          |
| mTOR signaling pathway                                                 |          |          |
| Oocyte meiosis                                                          |          |          |
| TGF-beta signaling pathway                                             |          |          |
| Fc epsilon RI signaling pathway                                         |          |          |
| Aldosterone-regulated sodium reabsorption                              |          |          |
| Natural killer cell mediated cytotoxicity                               |          |          |
| Rheumatoid arthritis                                                   |          |          |
| Endocytosis                                                             |          |          |
| Glycosphingolipid biosynthesis - ganglio series                         |          |          |
| Hippo signaling pathway                                                |          |          |
| GnRH signaling pathway                                                 |          |          |
| Complement and coagulation cascades                                    |          |          |
| Bladder cancer                                                         |          |          |
| Thyroid hormone signaling pathway                                       |          |          |
| Butirosin and neomycin biosynthesis                                    |          |          |
| Renin-angiotensin system                                               |          |          |
| PPAR signaling pathway                                                 |          |          |

### miR 5C_down

| Pathway                                                                 | MCF-7:5C | MCF-7:2A |
|-------------------------------------------------------------------------|----------|----------|
| Metabolic pathways                                                     |          |          |
| Parkinson's disease                                                     |          |          |
| Renal cell carcinoma                                                    |          |          |
| Pathways in cancer                                                      |          |          |
| Non-alcoholic fatty liver disease (NAFLD)                               |          |          |
| SNARE interactions in vesicular transport                              |          |          |
| Propanoate metabolism                                                  |          |          |
| Ras signaling pathway                                                  |          |          |
| Ether lipid metabolism                                                 |          |          |
| Insulin signaling pathway                                              |          |          |
| ErbB signaling pathway                                                 |          |          |
| Mucin type O-Glycan biosynthesis                                       |          |          |
| Glycerolipid metabolism                                                |          |          |
| Acute myeloid leukemia                                                 |          |          |
| Natural killer cell mediated cytotoxicity                              |          |          |
| Glycosaminoglycan biosynthesis - keratan...                             |          |          |
| Alanine, aspartate and glutamate metabolism                            |          |          |
| Taurine and hypotaurine metabolism                                     |          |          |
| AMPK signaling pathway                                                 |          |          |

- P < 0.05
Supplementary Figure 5: KEGG pathways identified in miRNA functional enrichment analyses. A. 5C cell-specific enriched pathways stratified by up (red)- and down (green)-regulated miRNA subsets. The yellow node represents pathways enriched for both up- and down-regulated miRNA subsets. B. 2A cell-specific enriched pathways stratified by up (red)- and down (green)-regulated miRNA subsets. Yellow nodes represent pathways enriched for both up- and down-regulated miRNA subsets. The node size reflects the number of pathway-related genes regulated by the specific miRNA set. The edge weight illustrates the degree of gene overlap between two pathways as measured by the Jaccard index. Pathways highlighted in bold are enriched in both AI-resistant cell models. The font size reflects the unadjusted
Fisher test P-values from the enrichment analysis (P < 0.001, large (24 pts); 0.001 ≤ P < 0.01, intermediate (18 pts); 0.01 ≤ P < 0.05, small (12 pts)). C. Enriched pathways of 5C- and 2A-specific up- and down-regulated miRNA subsets. X-axes represent -log_{10} value, Y-axes: enriched pathways sorted according to specific pathways enriched in 5C or 2A. D. Venn diagram illustrating pathway overlaps between up- and down-regulated miRNA sets of 5C and 2A cells (Table S1 and S2).

Supplementary Figure 6: Basal levels of gene and protein expression of key growth factor signal transduction mediators compared between AI resistance models 2A and 5C and their modulatory miRNAs. Gene expression levels were quantified by RT-PCR and protein expression levels were examined by Western blotting with β-action as loading control (WS8: E2 growth sensitive control). Up (red) and down (green) regulating miRNAs were obtained from specific miRNA sets given in Table 1. Putative miRNA-mRNA interaction thresholds were defined as CLIP confirmed Targetscan 7.0 in silico predictions (> 50 percentile). MiRNAs of the DLK-DIO3 cluster of chromosome 14 are underlined. Small down regulatory effects in 5C compared to 2A were observed for AKT1, MAPK1, KRAS, RAF-1. Small
up regulatory effects in 5C compared to 2A were observed for MTOR. *: $P < 0.05$, #: $P < 0.01$, **: $P < 0.001$ (comparisons to WS8).
Supplementary Figure 7: GO terms of the subcategory ‘Biological Process’ identified in Gene Set Enrichment Analyses with the GORILLA tool. Test set gene lists were obtained from KEGG pathway interaction networks (Figs. 5A, B). Resulting GO terms were summarized by removing redundant terms with REVIGO using default parameters. Terms were then filtered according to the following thresholds: frequency < 5% and $-\log_{10} P < 40$ (intersection gene set), $-\log_{10} P < 15$ (5C-, 2A-specific gene sets). Dark color bars represent major GO terms; light color bars specify minor sub-terms of major terms (bold). A. Intersection of 5C versus 2A gene sets. B. 5C-specific gene set. C. 2A-specific gene set. X-axes: adjusted $-\log_{10} P$-Value; Y-axes: GO terms.

| 5C | Proliferation / communication |
|----|--------------------------------|
|    | Signal transduction           |
|    | • Regulation                  |
|    | • Tyrosine/kinase signaling   |
|    | • Cellular response to stress |
|    | • Response to oxygen-containing compound |
|    | • Cell cycle regulation       |
| Metabolism | Generation of precursor metabolites and energy |
|            | Regulation of macromolecules  |
|            | • Kinase activity/phosphorylation |
|            | • Serine/threonine kinase activity |
|            | Catalytic activity            |
|            | • Transferase activity        |
|            | Protein localization          |
| Apoptosis  | Regulation of apoptosis       |
| Development | Regulation of cell differentiation |
|            | Wnt signalling pathway        |

| 5C, 2A | Proliferation / communication |
|--------|--------------------------------|
|        | Cell surface receptor signaling |
|        | • Signal transduction and its regulation |
|        | • Tyrosine/kinase receptor      |
|        | • Ras                          |
|        | • Phosphatidylinositol-mediated |
|        | • MAPK cascade                 |
| Metabolism | Generation of precursor metabolites and energy |
|            | Proteins                       |
|            | • Kinase activity/phosphorylation |
|            | • Transferase activity         |
|            | Regulation                     |
|            | • Response to stress and stimulus |
| Immune response | Regulation of immune system processes |
|                | • Innate immune response       |
|                | • Immune response regulation   |
|                | • Fc-receptor signaling        |
| Apoptosis | Regulation of apoptosis        |

| 2A | Metabolism |
|----|------------|
|    | Lipids     |
|    | • Glycerol-phospholipids |
|    | • Phosphatidylinositol |
|    | • Organophosphates |
|    | Coenzymes  |
|    | • Coenzymes |
|    | • Nicotinamide nucleotides |
|    | • Oxidoreduction coenzymes |
|    | Carbohydrates |
|    | Sulfur compounds |
|    | Amino acids |
|    | • Organic acids |
|    | • Carboxylic acids |
|    | • Glutamine |
|    | Nucleoside and nucleotides |
|    | Metabolic compound salvage |
|    | Protein glycosylation |
|    | • Macromolecules |
|    | • N-linked glycosylation |
|    | • Post-translational modification |
|    | Immune response |
|    | Cytokine-mediated signaling |

Supplementary Figure 8: Abbreviated GO terms of the subcategory ‘Biological process’ enriched in 5C and 2A models. Major (bold) and minor terms are listed for the AI resistance phenotype common to the 5C and 2A models (grey boxes), 5C-specific (blue boxes) and 2A-specific phenotype (green boxes).
Supplementary Figure 9: miR-134 mimic and inhibition affect EGFR expression in MCF-7:2A and MCF-7:5C cells, respectively. A. In 2A cells miR-134 expression levels are significantly increased at 3 and 5 days. Although EGFR mRNA levels are increased, the downregulation of EGFR protein correlates with relative miR-134 levels. B. In 5C cells miR-134 expression levels are significantly decreased at 3 and 5 day inhibition. EGFR mRNA level is significantly increased after 5 days which correlates with an increase of EGFR protein. β-actin was used as loading control; NC: negative control. qRT-PCR experiments were done in triplicates; *: P < 0.05, **: P < 0.001.