Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics

of the Ruperto Carola University Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

M.Sc. Janine Jung

Born in: Rottweil, Germany

Oral examination: 30.10.2020

Identifying and characterizing functionally relevant microRNAs and 5'isomiRs in triple-negative breast cancer

Referees: Prof. Dr. Stefan Wiemann

Prof. Dr. Peter Angel

Declaration of authorship

I hereby declare that the work presented in this thesis was carried out between November 2016 and August 2020 under the supervision of Prof. Dr. Stefan Wiemann in the group 'Molecular Genome Analysis' at the German Cancer Research Center (Heidelberg, Germany).

If not stated otherwise and referenced accordingly in the text, the data described in my dissertation is original, has been gathered by myself and has not yet been presented as part of a university examination. The main sources as well as the work of joint cooperation have been referenced appropriately. I, as author, herewith declare no potential conflict of interest.

Heidelberg, _____

Janine Jung

ACKNOWLEDGEMENTS

First of all, I would like to thank Prof. Dr. Stefan Wiemann for giving me the opportunity to conduct the work for this thesis in the Division of Molecular Genome Analysis and for providing a scientific environment that allowed me to develop my own project and to explore my ideas. I really enjoyed working at MGA.

Second, I would like to thank Dr. Cindy Körner for designing this project with me and for the supervision during the past years. I really appreciated your support, especially in difficult times.

I would also like to thank my Thesis Advisory Committee members Prof. Dr. Peter Angel, Prof. Dr. Claudia Scholl and Prof. Dr. Dirk Grimm for their valuable advice. Moreover, I want to thank my examiners Prof. Dr. Stefan Wiemann, Prof. Dr. Peter Angel, Prof. Dr. Claudia Scholl and Prof. Dr. Joachim Wittbrodt for evaluating my thesis.

Furthermore, I would like to thank Subarna Palit, Susanne Ibing and Shashwat Sahay for the bioinformatic support.

I am also very grateful to Prof. Dr. Peter Sinn and Dr. Martina Kirchner for allowing me to use their NanoString nCounter system and for the experimental help provided by Martina.

Many thanks to all members of MGA for the warm and positive atmosphere in the lab. I am especially thankful for the nice and familial atmosphere that Angelika and Heike spread in 'our' lab wing. Thank you, Angelika, for your help with the last-minute experiments.

A huge 'thank you' goes to my 'girls office'. Thank you so much for everything Xiaoya and Zhivka. I am grateful for all the scientific and non-scientific conversations we had and all the help you offered when I needed it. And most importantly, I enjoyed our shared love for food and us actually sharing a lot of food.

Furthermore, I want to thank Ana, Anni and Philipp from the R&D career day marketing team. I enjoyed working and not-working with you guys. I loved our coffee breaks!

I am also very grateful for the support my parents and my sister showed me throughout my entire education. In addition, I want to thank my parents-in-law for their support and for welcoming me into their family.

I owe more than words can say to my friends. Most of you have been my support system for many years and for you this is just the third thesis in a row, but to me it means so much more that you came this long way with me. We shared all the bad and mainly good and crazy times and not taking life or us too serious helped with everything. Thank you, Verena, Arthur, Dan, Jens, Bryan and

Björn, for all the parties, food, more food and all the funny and insane conversations (not to forget all the awesome business ideas). And Eddy, although we don't manage to meet that often, it is always like our time as Bachelor students never ended whenever we see each other. Marcel, thank you for introducing us to the world of handball and enlightening the gym every single day (mostly with your good mood, but also with the color of your head and hair :D). A huge thank you goes to my best friend Michael. There are really no words that describe the past three years and I am grateful for all the funny and awesome moments and (partially insane) conversations.

My deepest gratitude, however, goes to my husband Michael for his constant support in literally everything. Without you this thesis and many more things would not have been possible or turned out very differently. This journey would have simply not been the same without you.

Thanks to all of you for a memorable journey and for supporting me in making it to the finish line!

TABLE OF CONTENT

1. SUMMARY / ZUSAMMENFASSUNG

1.1 Summary

Triple-negative breast cancer is a highly aggressive breast cancer subtype and the treatment options are mainly limited to chemotherapy, however, the patients frequently develop resistance. As endogenous regulators of gene expression, microRNAs are involved in tumor development, progression and treatment resistance. microRNA sequence variants with a shifted seed sequence are termed 5'isomiRs and extend the complexity and impact of the miRNome in cancer. A shift in the seed sequence by only one nucleotide can drastically alter the target spectrum of a 5'isomiR compared to its canonical microRNA. Hence, this study aims at identifying microRNAs and 5'isomiRs with a potential role in tumorigenesis and chemoresistance and focuses on characterizing their functional differences in triple-negative breast cancer.

I selected microRNAs and 5'isomiRs that were differentially expressed between tumor and normal tissue of patients from the TCGA cohort and, thus, potentially involved in tumorigenesis and chemoresistance. Growing mammospheres from MDA-MB-231, HCC1806 and SUM-159 cells that overexpressed the selected microRNAs as pooled library enriched for cells with increased stemness and chemoresistance. Read-out of the library composition by NanoString after several sphere generations revealed strong enrichment of pre-miR-103a-1. In validation experiments, pre-miR-103a-1 overexpression did not influence stemness or chemoresistance.

In the second part of the project, I focused on the functional characterization of miR-1307-3p I0 and its 5'isomiR miR-1307-3p I1. Both were selected from the list of differentially expressed microRNAs based on their similar expression levels. Phenotypic assays in triple-negative breast cancer cell lines showed that both microRNAs reduce migration, miR-1307-3p I0 in a cell line-specific manner and less pronounced than miR-1307-3p I1. miR-1307-3p I1 repressed proliferation in a cell line-dependent context. Target predictions identified genes that might contribute to these phenotypes and explain differences between cell lines. The putative targets suggested that miR-1307-3p I0 plays a role in autophagy.

In summary, I showed that miR-1307-3p I0 and I1 influence different and similar phenotypes in a partially cell line-dependent manner by targeting specific as well as shared putative target subsets. This study underlines how complex and context-dependent microRNAs and their 5'isomiRs modulate gene expression and that they are of biological relevance. Consequently, diagnostic, prognostic and therapeutic approaches should discriminate between 5'isomiRs.

1.2 Zusammenfassung

Triple-negativer Brustkrebs ist eine sehr aggressive Form von Brustkrebs und die Behandlungsmöglichkeiten beschränken sich hauptsächlich auf Chemotherapie, gegen welche die Patienten jedoch häufig eine Resistenz entwickeln. Als endogene Regulatoren von Genexpression sind microRNAs in Entwicklung, Progression und Behandlungsresistenz von Tumoren involviert. microRNA Sequenzvarianten mit einer verschobenen Seedsequenz werden 5'isomiRs genannt und erweitern die Komplexität und den Einfluss des miRNomes in Krebs. Eine Verschiebung der Seedsequenz um nur ein Nukleotid kann das Targetspektrum einer 5'isomiR im Vergleich zur kanonischen microRNA drastisch verändern. Daher zielt diese Studie darauf ab microRNAs und 5'isomiRs mit einer möglichen Beteiligung an Tumorigenese und Chemoresistenz zu identifizieren und konzentriert sich darauf die funktionellen Unterschiede in triple-negativem Brustkrebs zu charakterisieren.

Ich wählte microRNAs und 5'isomiRs aus, die in Tumor- und Normalgewebe von Patienten aus der TCGA Kohorte differenziell exprimiert werden und daher wahrscheinlich in Tumorigenese und Chemoresistenz involviert sind. Die 3D-Kultivierung von MDA-MB-231, HCC1806 and SUM-159 Zellen, welche die ausgewählten microRNAs als gepoolte Library überexprimierten, reicherte Zellen mit Stammzell-Charakter und erhöhter Chemoresistenz an. Die Bestimmung der Library Zusammensetzung mit NanoString nach mehreren 3D-Generationen zeigte eine starke Anreicherung von pre-miR-103a-1. In Validierungsexperimenten beeinflusste die Überexpression von pre-miR-103a-1 Stammzell-Charakter und Chemoresistenz nicht.

Im zweiten Teil des Projekts konzentrierte ich mich auf die funktionelle Charakterisierung von miR-1307-3p I0 und ihrer 5'isomiR miR-1307-3p I1. Beide wurden aus der Liste differenziell exprimierter microRNAs aufgrund ihrer ähnlich starken Expression ausgewählt. Phäntotypische Experimente in triple-negativen Brustkrebszelllinien zeigten, dass beide microRNAs Migration reduzieren, miR-1307-3p I0 auf zelllinienspezifische Weise und weniger ausgeprägt als miR-1307-3p I1. miR-1307-3p I1 unterdrückte die Proliferation in zelllinien-abhängigem Kontext. Mit Hilfe von Targetvorhersagen wurden Gene identifiziert, die wahrscheinlich zu diesen Phänotypen beitragen und die Unterschiede zwischen den Zelllinien erklären. Zudem deuteten die putativen Targets darauf hin, dass miR-1307-3p I0 eine Rolle in Autophagie spielt.

Zusammenfassend zeigte ich, dass miR-1307-3p I0 und I1 unterschiedliche und ähnliche Phänotypen in teilweise zelllinienspezifischer Weise beeinflussen. Dies geschieht über spezifische und gemeinsame Teilmengen des Targetspektrums. Diese Studie unterstreicht wie komplex und

kontextabhängig microRNAs und ihre 5’isomiRs die Genexpression modulieren und dass sie von biologischer Relevanz sind. Infolgedessen sollten diagnostische, prognostische und therapeutische Ansätze zwischen 5’isomiRs unterscheiden.

2. INTRODUCTION

2.1 Breast cancer

Breast cancer is a common disease in women[1]. In 2019, the American Cancer Society reported that 268,600 women were diagnosed with breast cancer and 41,760 died from breast cancer[2]. These numbers render breast cancer the most common cancer entity in women and the second most frequent cancer-related cause of death. The heterogeneity of breast cancer requires the discrimination between subtypes in order to recommend a suitable therapeutic strategy. In the past, breast cancer was classified mainly based on immunohistological characteristics, for instance, the presence of estrogen receptor (ER), progesterone receptor (PR) and human epidermal receptor 2 (HER2)[3]. ER+ breast cancer is the most common receptor status of breast tumors and is treated with endocrine therapy, while patients with HER2+ breast cancer receive HER2-targeted therapy, for instance, Trastuzumab (Figure 1)[4]. Both patient groups have a more favorable prognosis than triple-negative breast cancer (TNBC) patients that lack ER, PR and HER2 expression. The lack of receptors that can be targeted explains why there is currently no targeted therapy available for the 15-20 % of patients that are diagnosed with TNBC[5,6]. Approximately 75 % of all breast cancer patients, however, have a good prognosis since they are ER+ and, thus, benefit from endocrine therapy[7,8].

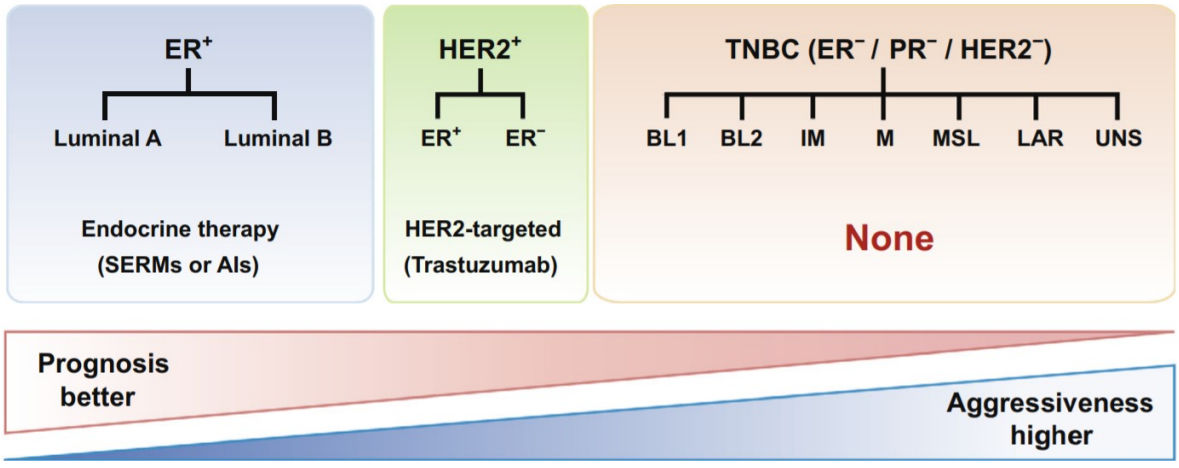


Figure 1: Breast cancer subtypes and targeted therapy options. Breast cancer classification based on receptor status: ER+, HER2+ and TNBC. These three groups strongly differ in their prognosis and different targeted therapy options are available. Each breast cancer type is divided into further subtypes. SERM = selective estrogen receptor modulator, AI = aromatase inhibitor. The figure was modified from Ma et al., 2018[4].

Nowadays, gene expression profiling and RNA sequencing allow to refine this classification by taking gene signatures into account and, thus, improve prognostic and therapeutic approaches. The PAM50 intrinsic subtype classifier discriminates between five breast cancer subtypes based on the mRNA expression pattern of 50 genes: Luminal A, Luminal B, HER2-enriched, Basal-like and Normal-like[9–11]. Luminal A is the most frequent PAM50 subtype (50-60 %), followed by Luminal B (15-20 %), HER2 (15-20 %) and Normal-like (5-10 %)[12]. The abundance of Basal breast cancer (8-37 %) varies a lot based on the prevalence of poorly differentiated grade 3 cases in the evaluated population. Basal and HER2+ breast cancer are the most aggressive subtypes and associated with poor prognosis[13,14]. Luminal A patients have the best prognosis, followed by Luminal B patients[12]. The Normal-like subtype, however, is poorly characterized.

2.1.1 Triple-negative and Basal breast cancer

In 70-80 % of all cases, Basal breast tumors are also classified as TNBC[11]. While TNBC lacks ER and PR expression as well as HER2 amplification, approximately 20 % of the Basal breast tumors overexpress either ER or HER2[5]. Moreover, several other markers including EGFR, c-Kit as well as cytokeratins 5, 6, 14 and 17 are associated with Basal tumors[15]. Basal patients overexpress genes related to proliferation, cell cycle and DNA damage response[16]. TNBC reveals a high frequency of TP53 mutations (80 %), while *PIK3CA* mutations present in 8 % of all cases[17,18]. The tumor-suppressors *PTEN*, *RB1* and *BRCA1* are frequently lost in TNBC, whereas *MYC* is commonly amplified[19]. Despite several differences, TNBC and Basal breast cancer also have common features: both are highly aggressive, frequently metastasize to lungs and brain and patients with residual disease after chemotherapy have a poor overall prognosis since they are more prone to relapse than other subtypes[5,20–22].

2.1.2 Triple-negative breast cancer subtypes

Based on gene expression and pathway activities TNBC was originally classified into six subtypes: Basal-Like 1 (BL1), Basal-Like 2 (BL2), Immunomodulatory (IM), Luminal Androgen Receptor (LAR), Mesenchymal (M) and Mesenchymal Stem-Like (MSL) (Figure 1)[23]. However, Lehman et al. showed that the subtypes IM and MSL originated from tumor-infiltrating lymphocytes and stromal cells, respectively[24]. Thus, the classification was simplified to the four subtypes BL1, BL2, LAR and M. BL1 is characterized by an increase in gene expression related to cell cycle and DNA damage response, moreover, this subtype represents the most prevalent one and has the best prognosis[24].

The TNBC subtype M is driven by enhanced EMT and growth factor signaling. The subtypes with the worst prognosis, BL2 and LAR, are enriched in growth factor signaling and myoepithelial markers or luminal gene expression, respectively. The highly proliferative subtype BL1 responds well to therapies targeting mitosis, while M patients might benefit from agents targeting angiogenesis and PI3K/mTOR inhibitors based on their enrichment in EMT signatures[25]. The dependence on the androgen receptor renders anti-androgen therapies highly relevant for LAR patients[26]. Since the BL2 subtype is characterized by upregulated growth factor signaling, angiogenetic factors, glycolysis and gluconeogenesis, inhibitors targeting the receptors of VEGF, PDGF and FGF show some promise[27]. Overall, the classification of TNBC into subtypes allows a more refined treatment strategy and is of high interest to patients who do not respond to chemotherapy or relapse after treatment. The distinct molecular signature allows identifying potential targets and therapeutically exploiting pathway vulnerabilities. The respective targeted therapies, however, still need to be tested with clinical trials.

2.1.3 Chemotherapy as standard care for triple-negative breast cancer patients

Due to the lack of targeted therapy options, neoadjuvant chemotherapy is the standard of care for patients with TNBC[28]. In 50 % of the patients, neoadjuvant chemotherapy does not result in pathological complete response (pCR). The lack of complete response is associated with a high rate of recurrence (40-60 %)[29]. Overall, TNBC patients with less progressed disease respond more likely to chemotherapy[30]. The pCR in the subtype BL1 is around 41-52 %[24,28,31], whereas the other subtypes have a worse response to chemotherapy. BL2 and LAR show a pCR of 0-18 % and 10-29 %, respectively[24,31]. In the subtype M, almost 40 % of the patient have a complete response to chemotherapy[24].

Anthracyclines and taxanes are administered as standard chemotherapy regimen in the neoadjuvant setting[32]. Platinum-based regimen have been proposed, but their effectiveness is still under research and they are not part of the standard treatment protocol[32]. Adding cisplatin or carboplatin to the chemotherapy regimen increased pCR rates, the improvement of survival, however, needs to be determined[33]. Anthracyclines and platinum agents target DNA synthesis and integrity, while taxanes affect cytokinesis[34]. Patients that receive anthracycline-based chemotherapy and relapse can be treated with taxanes and vice versa. For patients with recurrent disease after treatment with anthracyclines and taxanes, fluorouracil/capecitabine, eribulin, gemcitabine, vinorelbine or ixabepilone represent alternative treatments[35].

While TNBC patients with residual disease after neoadjuvant chemotherapy but without a relapse are characterized by a luminal-like gene signature, patients with recurrent disease reveal a stem cell-like gene signature[36]. Stem cells and stemness-related pathways strongly contribute to chemoresistance in TNBC[37]. However, several other molecular mechanisms play a major role in the development of chemoresistance: senescence and autophagy, for instance, circumvent apoptosis and thereby allow TNBC cells to escape chemotherapy[38]. The upregulation of ABC transporters helps the cells to get rid of the chemotherapeutic drugs[37]. In conclusion, insight into the mechanisms of chemoresistance of TNBC patients with recurrent disease after chemotherapy is urgently required to develop targeted therapies or strategies to overcome treatment resistance.

2.2 microRNAs

2.2.1 Biogenesis and function of microRNAs

MicroRNAs are major regulators of endogenous gene expression and typically regulate hundreds of targets[39,40]. Mammalian microRNA genes are mainly located within introns of protein-coding or non-coding genes, a smaller fraction overlaps with exons of protein-coding or non-coding transcripts[41,42]. In several cases it is determined by the splicing pattern whether the microRNA gene is located within an exonic or intronic sequence[42]. microRNA biogenesis starts with the primary microRNA (pri-microRNA) being transcribed from the microRNA gene by RNA polymerase II[43] (Figure 2). The pri-microRNA is composed of 500-3,000 nucleotides, a 7-methylguanosine cap at the 5'terminus and a polyadenylated 3'tail[44]. In the nucleus, the pri-microRNA is cleaved by a complex formed by Drosha and DGCR8, which results in a precursor microRNA (pre-microRNA) of 70-80 nucleotides length[45,46]. Export of the pre-microRNA into the cytoplasm by Exportin-5 allows further processing by the Dicer-TRBP complex, which forms a microRNA duplex consisting of a guide strand and a passenger strand[43]. Usually, the guide strand is characterized by lower stability at the 5'terminus or the base uracil at the 5'end[47]. After the microRNA duplex is loaded into the RNA-Induced Silencing Complex (RISC) formed by Argonaute proteins, both strands are unwound and the passenger strand is degraded[48]. The microRNA guide strand, however, remains incorporated into the RISC (miRISC) and governs the complex to the respective mRNA targets in order to repress their translation[49]. Alternatively, the target mRNA is degraded [40,50]. microRNA biogenesis can also occur via Drosha- or Dicer-independent pathways, however, these alternative pathways are less well studied[51].

Mature microRNA are comprised of 21-25 nucleotides, of which mainly the seed sequence (nucleotides 2-8) determines specific binding to a complementary region within the 3'UTR of mRNA targets[52,53]. The 3'UTRs of microRNA targets often harbor several binding sites for the same or other microRNAs[39,54], which amplifies the repression of the mRNA target. Moreover, mRNA targets may influence the levels of the microRNAs regulating them[55], which shows the complexity of these bidirectional microRNA-target relations. Overall, more than 60% of all protein-coding genes were found to be conserved microRNA targets[56]. This shows that microRNA-regulated gene expression is a global phenomenon and that their deregulation exerts a great impact on cellular fate by affecting multiple signaling pathways[57], which makes microRNAs crucial players in cancer development, therapy resistance and relapse[58].

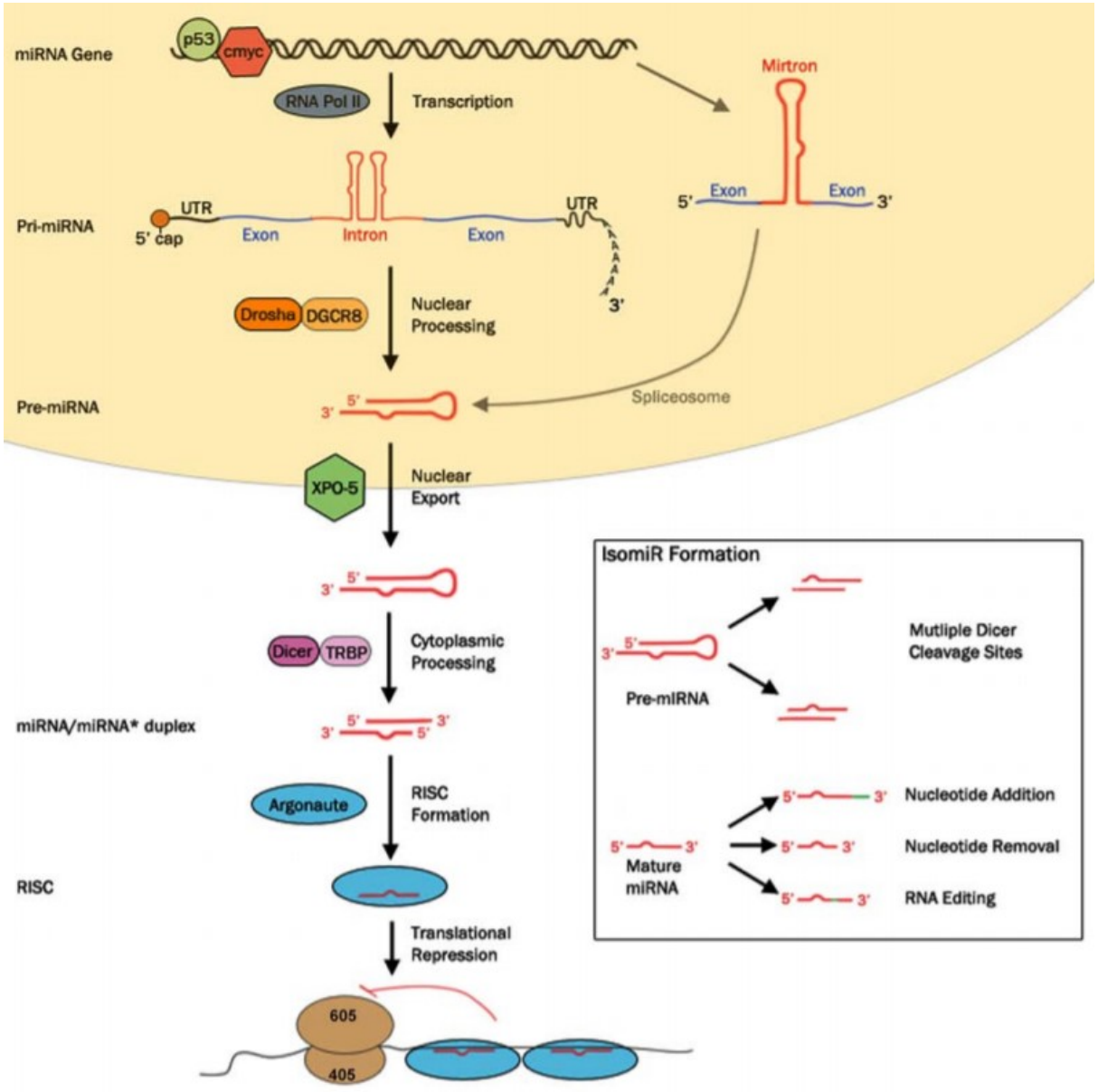


Figure 2: microRNA biogenesis. Mammalian microRNA biogenesis starts with pri-microRNA transcription by RNA polymerase II from the genome. Cleavage of the pri-microRNA by the microprocessor complex consisting of Drosha and DGCR8 results in the pre-microRNA. Subsequently, the pre-microRNA is exported to the cytoplasm by Exportin-5. In the cytoplasm, the Dicer-TRBP complex cleaves the pre-miRNA, which results in a microRNA duplex. The microRNA guide strand is bound by the RNA-Induced Silencing Complex (RISC) formed by Argonaute proteins. The mature miRNA guides the RISC to the mRNA targets, resulting in translational repression or degradation of the bound transcripts. During microRNA biogenesis, alternative Dicer cleavage sites, post-transcriptional modifications and RNA editing produce microRNA sequence variants, so-called isomiRs. The figure was modified from Bajan et al., 2014[43].

Different types of binding sites within the 3'UTR determine the impact that microRNAs have on their mRNA targets. Canonical binding sites are the most common type and differ in the number of bases that are complementary to the seed sequence of the microRNA (Figure 3). The extent to which a microRNA represses its target is largely determined by the amount of bases within the 3'UTR that are complementary to the seed sequence: 8mer binding sites in the 3'UTR consist of more base pairs that are complementary to the seed sequence than 7mer-m8, 7mer-A1 or 6mer binding sites, which results in a higher binding affinity and, thus, stronger repression of the mRNA target[59,60]. Besides the canonical binding sites, microRNAs can bind and regulate their targets via 3'supplementary and 3'compensatory sites. Both target sites provide additional base pairs complementary to parts of the microRNA: the 3'supplementary site is characterized by additional pairing of 3-4 nucleotides at position 13-16, while the 3'compensatory site reveals additional pairing at position 13-16 in order to compensate for discontinuous complementary between 3'UTR and seed sequence[61,62].

Although microRNAs predominantly target 3'UTRs, they repress mRNA targets by binding to the 5'UTRs or coding sequences as well[63]. microRNA-532-5p, for instance, downregulated RUNX3 by targeting the 5'UTR[64] and let-7 repressed Dicer by binding to the coding sequence[65]. Moreover, targeting of promoters upregulates gene expression. miR-324-3p, for instance, induced expression of RelA by binding within the promoter sequence[66].

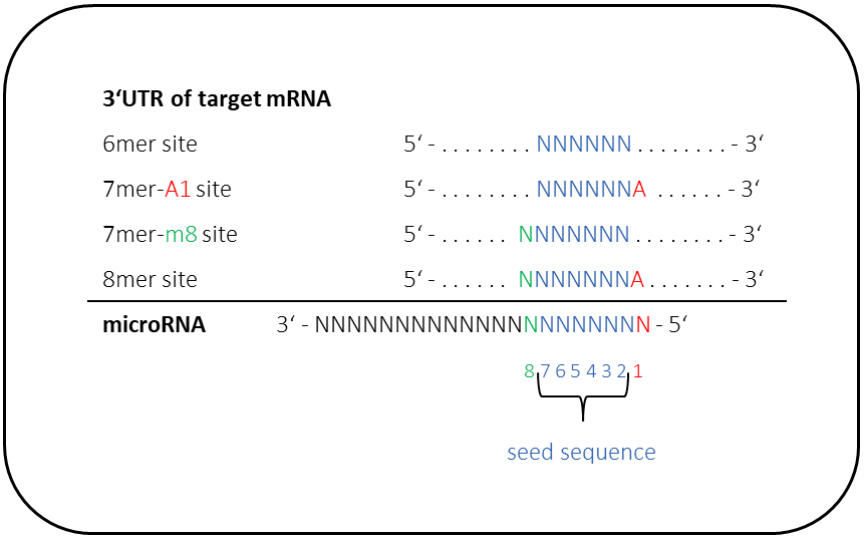


Figure 3: Different types of canonical microRNA binding sites. Canonical microRNA binding sites are differentiated based on the extent to which the seed sequence of the microRNA is complementary to a sequence within the 3'UTR of a mRNA target. The more bases of the 3'UTR are complementary to the microRNA seed sequence, the stronger the microRNA:target interaction and the resulting downregulation of gene expression.

2.2.2 Studying microRNA:target interactions

microRNA targets are usually identified and validated in a process consisting of multiple steps: target prediction tools such as PicTar, MiRanda or TargetScan are employed to detect targets with conserved microRNA binding sites[60]. A microRNA may target around 200 mRNAs, the number of predictions, however, often comprises thousands of potential targets containing numerous false positives[40]. To exclude false positives, the predicted targets are overlapped with experimental perturbation data that is generated by microRNA overexpression. Typical formats for the experimental validation are sequencing or MicroArray-based profiling of all expressed genes, whereas qRT-PCR-based validation is carried out for a smaller number of targets that has been preselected by literature, for instance. To confirm direct binding of validated mRNA targets by the microRNA, the 3'UTRs of the respective mRNA targets are cloned into a luciferase reporter construct. Cells are transfected with microRNA mimics together with the luciferase reporter and direct binding of the 3'UTR is detected by a reduction in luciferase activity compared to the non-targeting control[67]. Mutation of the binding sites within the reporter should abolish the effect of the microRNA on luciferase activity, which provides another layer of validation[67].

More advanced approaches to identify microRNA targets focus on determining the targets while the miRISC is interacting with them. Cross-linking and immunoprecipitation (CLIP) strategies employ UV light for crosslinking of mRNA targets bound by miRISC and immunoprecipitation of Ago to pull-down the miRISC complex[68,69]. Subsequent high-throughput sequencing allows identifying the immunoprecipitated targets and microRNAs that were incorporated into the RISC (HITS-CLIP)[68,70]. For photoactivatable-ribonucleoside-enhanced CLIP (PAR-CLIP), cells are fed with medium containing the nucleoside analog 4-Thiouridine and RNA:protein interactions are cross-linked at a different wave length than HITS-CLIP[70]. Further development of CLIP allows profiling of cross-linked sites at individual-nucleotide resolution (iCLIP)[71]. Although the mentioned techniques detect only RNA:protein interactions, intersecting the generated data with predicted targets allows to narrow down the number of putative microRNA:target interactions. Transfection of biotinylated microRNA mimics allows pull-down of mRNA targets regulated by the respective microRNA with streptavidin-coated beads[72]. Sequencing the precipitated mRNA targets and overlapping them with target predictions is another approach to obtain valid target candidates.

One method that does not depend on intersecting the data with predicted targets to unravel putative microRNA:target interactions is crosslinking, ligation and sequencing of hybrids (CLASH). CLASH directly identifies the respective microRNA that represses the mRNA target[68]. While the initial steps are similar to CLIP approaches, mRNA 5'ends are modified after immunoprecipitation

of Ago-RNA complexes. The modification is required to ligate microRNAs and their mRNA targets for subsequent sequencing[73]. A major benefit of CLASH is the detection of targets harboring canonical as well as non-canonical binding sites[74].

2.2.3 isomiRs - microRNA sequence variants

Currently 2675 mature canonical microRNAs are annotated in miRbase[75]. In recent years, however, microRNA sequencing revealed that one locus gives rise to multiple sequence variants, which are termed isomiRs[76,77]. Partially, the variation in microRNA sequence length can be explained by alternative Drosha- or Dicer-mediated cleavage during miRNA biogenesis[78–81] (Figure 2). Alternative processing by Drosha or Dicer is a templated process since the isomiR sequence still matches the pre-microRNA sequence from which the mature isoform is derived[82]. Moreover, microRNA sequence variants can result from 3'trimming by 3'-to-5' exoribonucleases[83,84]. A non-templated mechanism that generates isomiRs, is the post-transcriptional addition of nucleotides at the 5' or 3' terminus of the microRNA by nucleotidyl transferases[85]. The microRNA sequence variants generated by alternative Drosha/Dicer cleavage, 3'trimming or post-transcriptional 5'/3' nucleotide addition differ in their lengths and/or sequence at the 5' and 3'end (Figure 4) and are termed 5'isomiRs or 3'isomiRs, respectively[82]. Polymorphic isomiRs form the third class of isomiRs and differ in their sequence composition by harboring single nucleotide mismatches compared to the canonical microRNA[82]. Literature suggests that polymorphic isomiRs are a result of RNA editing[86].

The expression of isomiRs varies across cell and tissue types or developmental stages and reveals race- and gender-specific patterns[76,87–90]. Moreover, the abundance of isomiRs is regulated in a dynamic manner and can be modulated by various biological stimuli and different conditions, for instance, hypoxia, ischemia or interferon β stimulation[91–93]. While 3'variants affect microRNA stability as well as stability of the microRNA:target duplex by additional pairing with the mRNA target at the 3'end[59,94], 5'isomiRs are of greater functional relevance since their seed sequence differs from the canonical form[87]. A shift in the seed sequence of 5'isomiRs can drastically alter the mRNA target spectrum[95], which adds another layer of complexity to gene expression regulated by the miRNome. However, many 5'isomiRs still share a large subset of their targets with the respective canonical microRNA and thereby might have synergistic functions[96].

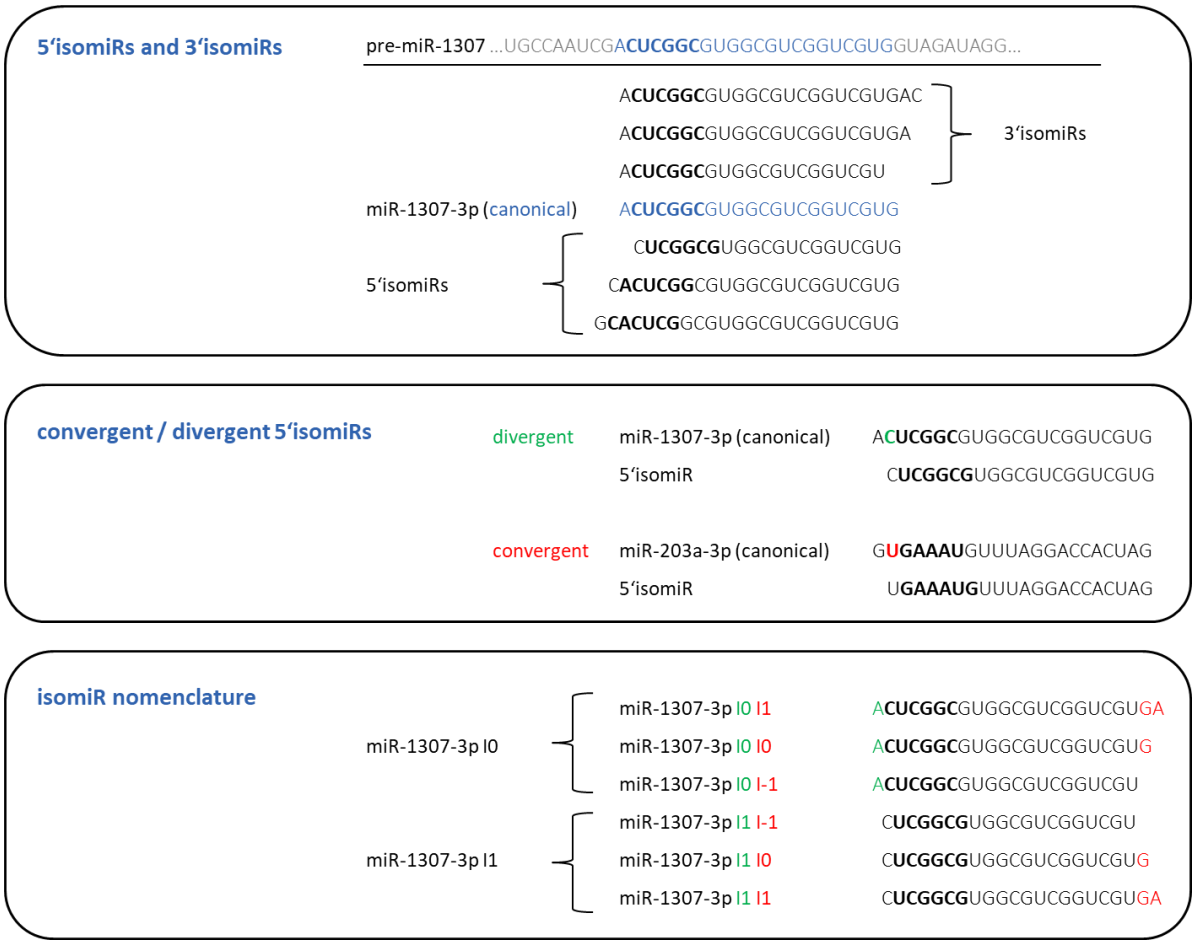


Figure 4: 5'isomiRs and 3'isomiRs - microRNA sequence variants. 5'isomiRs and 3'isomiRs differ in their length and/or sequence at the respective end from the canonical microRNA. While 3'variants affect the stability of the microRNA and the microRNA:target duplex, 5'isomiRs are shifted in their seed sequence and, thus, are of functional relevance. 5'isomiRs are discriminated based on their target spectrum: looking at the longer sequence variant when comparing two microRNAs, a U at nucleotide 2 indicates a convergent seed sequence. A, C or G at position 2 of the longer microRNA identifies a 5'isomiR with a divergent seed sequence. Divergent seed sequences require more extensive seed pairing for both microRNAs to share the binding site within a target and, thus, have more likely discrete target spectra. In this project, all microRNA sequences with the same 5'ends were summed up and considered as one 5'isomiR. The nomenclature that was used in this thesis refers to the 5'end of the sequence only and indicates how many bases the 5'end is shifted compared to the canonical microRNA. I1 indicates that the 5'isomiR is shifted by one nucleotide in 3'direction, I-1 would indicate a shift towards the 5'end.

5'isomiRs are divided into two categories: 5'isomiRs with a convergent seed sequence and 5'isomiRs with a divergent seed sequence. When comparing two 5'isomiRs, the second nucleotide of the longer sequence variant decides whether both 5'isomiRs share a majority of their targets or whether they rather regulate distinct target subsets. The nucleotide U at position 2 of the microRNA sequence indicates a convergent seed sequence, while A, C or G at position 2 identifies a 5'isomiR with a divergent seed sequence[96] (Figure 4). A divergent seed sequence of the longer

5'isomiR variant requires more nucleotides of the target 3'UTR to be complementary to the seed sequence of the microRNA. As a result, 5'isomiRs with a divergent seed sequence have more likely less overlapping target spectra.

To annotate isomiRs, Loher et al. proposed a system that uses the 5'end and 3'end of the canonical microRNA annotated in miRbase as reference[89]. A shift of the 5'end or 3'end in 3'direction by one nucleotide is annotated as I+1, whereas I-1 indicates a shift by one nucleotide in 5'direction. In this study, microRNA sequences with the same 5'ends were summed up and considered as one 5'isomiR disregarding the 3'ends (Figure 4). To simplify the nomenclature, a shift by one nucleotide in 5'direction is referred to as I1 instead of I+1.

2.2.4 Quantification of isomiR expression

Commercial qRT-PCR assays do not allow isomiR-specifc quantification of microRNA expression levels[97], neither do MicroArrays. While microRNA sequencing allows to unravel the isomiR expression landscape of the entire cell, the availability of custom-made, isomiR-specific detection assays for individual isomiRs of interest is limited. For detection of individual isomiRs and reliable discrimination from other isoforms, the specificity of adapters and probes is the limiting factor. One method that describes isomiR-specific detection is the Dumbbell-PCR, which employs adapters that are ligated to the 5'end or the 3'end of the isomiR of interest by T4 RNA ligase 2[98,99]. High specificity of the T4 RNA ligase 2 and subsequent qRT-PCR with TaqMan probes targeting the ligation product allows to determine isomiR expression levels.

Another method uses DNA probes as detection switch for isomiRs. The DNA probe contains a RNA hybridization module for the isomiR, a switching module and a restriction site for the nicking endonuclease Nt.BstNBI[100]. The DNA switch is activated upon binding of the respective isomiR to the DNA probe and reverse transcription of the assembled molecule. The DNA probe forms a hairpin only upon incorporation of nucleotides complementary to the 5'end of the bound isomiR. Subsequently, the signal is amplified by multiple cycles of DNA nicking and polymerization in a non-linear reaction.

A two-tailed qRT-PCR approach aims at isomiR detection by using two probes incorporated into a primer for reverse transcription[101]. One probe is located at the 5'end of the primer, the other probe is located at the 3'terminus. Binding of both probes to the target isomiR forms a stable complex for reverse transcription. In a second step, qRT-PCR primers bind to the target for amplification and SYBRGreen-based detection. Overall, assays for quantifying isomiR-specific

expression are still under development and there is no straight-forward solution yet. Moreover, optimization might be required for each individual isomiR in order to avoid unspecific detection of other isoforms or similar microRNA sequences.

2.2.5 microRNAs in triple-negative breast cancer

microRNAs have been implicated in a wide range of cancer-associated phenotypes and signaling pathways involved in tumorigenesis and therapy response. Features modulated by microRNAs include proliferation and cell cycle, motility and metastasis, apoptosis, autophagy, metabolism, stemness and resistance[102–106]. Cancer-associated phenotypes can be driven by various pathways in different cancer entities and subtypes. Exploiting pathway dependencies of cancer subtypes for targeted therapy is a common strategy and of high interest for entities with limited therapeutic options, such as TNBC.

The Notch and Wnt pathways, for instance, have been identified as highly relevant in TNBC, especially for the stem cell fraction of TNBC[107,108]. This renders Notch and Wnt signaling particularly relevant considering the association of stem cells and chemoresistance and the fact that only 60-70 % of the patients respond to chemotherapy[109]. And as for the majority of signaling pathways, microRNAs strongly affect these pathways: miR-124-3p was shown to promote TNBC cell growth via Wnt signaling[110], whereas miR-125b increased proliferation as well as motility via the Wnt pathway[111]. MiR-6838-5p on the other hand repressed Wnt signaling and, thus, motility in TNBC[112]. miR-105/93-3p affected TNBC stemness and chemoresistance by promoting Wnt signaling[113]. So far, only two microRNAs were associated with Notch signaling in TNBC: miR-106b-25 enhanced tumorigenesis via Notch signaling[114], whereas miR-34a targeted the Notch pathway to mediate tumor-suppressive effects[115].

Resulting from the heterogeneity of TNBC and its subtypes a large variety of signaling pathways contributes to tumorigenesis, disease progression and patient survival. This provides a lot of possibilities for targeted therapies, however, there is still a long way to go considering the complexity that microRNAs and isomiRs add to this context.

2.2.6 Clinical application of microRNAs and isomiRs

Since especially 5'isomiRs are of functional importance, discrimination between isomiRs is highly relevant in the context of biomarker discovery and clinical research. While microRNA-based

therapeutic approaches are still in the early phases of clinical approval, profiling of isomiRs opens extensive new possibilities for biomarker research that are exploited already[116,117]. 3'isomiRs of miR-574-3p and miR-205-5p, for instance, were identified as diagnostic biomarkers for esophageal squamous cell carcinoma[118]. On a larger scale, isomiRs were employed to distinguish between 32 cancer entities using 11,000 samples from the TCGA patient data[117] and to differentiate between breast cancer subtypes[119] or discriminate breast cancer tissue from normal tissue[76]. Even in patient serum or extracellular vesicles from cell culture supernatant deregulation of specific isomiRs allowed to detect breast cancer[120].

microRNA-based therapy works in two different ways: either by microRNA replacement therapy using microRNA mimics / precursors or by suppressing microRNAs with inhibitors or sponges[121]. microRNA replacement is a suitable approach for tumor-suppressive microRNAs that are downregulated in the tumor, for instance. microRNA inhibition on the other hand, aims at quenching oncogenic microRNAs that are highly abundant in the tumor. To date, there are no FDA-approved therapies based on microRNAs available, whereas some microRNA candidates are in clinical development or phase 1 / phase 2 trials. miR-16, for instance, completed a clinical phase 1 trial as second line or third line treatment for patients with lung cancer or recurrent thoracic cancer[122]. The clinical trial intravenously administered drug delivery vehicles that contained miR-16 mimics and were tailored to the EGF receptor with an antibody[123]. Moreover, several phase 2 trials subcutaneously injected an antisense oligonucleotide targeting miR-122 for the treatment of hepatitis C virus infection and were successfully completed[122,124,125].

Overall, there are a lot of promising microRNAs that could be exploited for therapeutic approaches. The major problem with their use as a drug, however, is their short half-life that is determined by the presence of nucleases and the necessary delivery to the cell via a carrier that allows passing membranes[121]. Two different classes of carriers are employed for microRNA delivery: viral vectors and non-viral carriers[121]. Carrier-free approaches are under development as well, for instance, the coupling of microRNAs to folate which allows the uptake into cancer cells that overexpress the folate receptor[126]. This method was refined recently to enhance endosomal escape of the microRNAs that were successfully delivered to the cell[127]. The microRNAs were coupled to nigericin in addition to folate, which promoted the swelling and bursting of endosomes and released the microRNA into the cytosol. However, folate receptor-mediated uptake of microRNAs strongly tailors this therapy to very specific subsets of cells or cancer types and cannot be applied as a general therapy concept.

microRNA-based therapies provide a lot of potential, especially for cancer entities and subtypes that lack targeted therapy options, such as TNBC. miR-708, for instance, was coupled to nanoparticles and administered to mice and reduced lung metastasis derived from the TNBC cell line MDA-MB-231[128]. Another study delivered nanoparticles tailored to the CD133 receptor and coated with a miR-21 inhibitor to TNBC cells and breast cancer stem cells[129]. Despite some success stories, microRNA-based therapies still have a long way to go until FDA-approval and general use in the clinics. Moreover, 5'isomiRs are currently not exploited for clinical studies, which might result from the fact that a multitude of studies does not discriminate between isoforms when profiling the miRNome in tumor tissue or patient serum.

2.3 Aim of the project

For the highly aggressive breast cancer subtype TNBC, there are currently no targeted therapies available and patients treated with chemotherapy frequently develop resistance towards the treatment. Thus, unraveling the molecular mechanism of chemoresistance and identifying targets for therapeutic approaches is of high interest. Since microRNAs modulate gene expression which affects the majority of signaling pathways, they play a crucial role in tumorigenesis and influence the cellular response towards chemotherapy. In the past years, our knowledge of the miRNome gained more complexity with the discovery of 5'isomiRs, microRNA sequence variants with a shifted seed sequence that can affect the target spectrum drastically. The majority of studies, however, does not take the functional divergence of microRNAs and their 5'isomiRs into account. In conclusion, this study aimed at identifying microRNAs with a key role in tumorigenesis and chemoresistance and focused on characterizing the functional differences of particular microRNAs and their 5'isomiRs in this context.

To achieve this aim, this study focused on:

1) Identifying microRNAs and 5'isomiRs of high relevance for breast cancer tumorigenesis and with a potential impact on response to chemotherapy.
2) Establishing an experimental system to identify microRNAs and 5'isomiRs with a major role in chemoresistance.
3) Characterizing the effect of selected microRNAs and their 5'isomiRs on chemoresistance and other cancer-associated phenotypes.
4) Analyzing the functional and mechanistic differences between selected microRNAs and their 5'isomiRs with a focus on their direct targets.

3. MATERIAL AND METHODS

3.1 Material

3.1.1 Laboratory equipment

Bacterial incubator (37°C)	Memmert
Bacterial shaking incubator (37°C)	HT INFOS Minitron
Balance	Kern
CASY cell counter	Roche Innovatis
Cell culture hood HERA Safe	Thermo Fisher Scientific
Cell culture incubator (37°C)	Heraeus
Centrifuges	Eppendorf, Heraeus
DNA gel apparatus	Bio-Rad
Electrophoresis power supply	Pharmacia
Freezer (-20°C)	Liebherr
Freezer (-80°C)	Eppendorf
Fridge (4°C)	Liebherr
Gel documentation system	Herolab
Glomax Microplate Reader	Promega
Light microscope	Hund
Micropipettes	Gilson
Microwave	Panasonic
ImageXpress Micro Confocal Microscope	Molecular Devices
ImageXpress Micro XLS Widefield Microscope	Molecular Devices
Multichannel pipette	Eppendorf
Multipette plus	Eppendorf
NanoDrop nd 1000 spectrophotometer	Thermo Fisher Scientific
nCounter FLEX Analysis System	NanoString
Pipetboy	Integra Biosciences
Qubit Fluorometer	Thermo Fisher Scientific
Real-time PCR Thermocycler	Applied Biosystems
Rocker Platform	NeoLab
Thermocycler	Applied Biosystems

Thermomixer	Eppendorf
Tube Rotator	VWR
Vacuum Aspirator	Integra Biosciences
Vortex Mixer	NeoLab
xCELLigence Real-time Cell Analyzer	Roche

3.1.2 Consumables

10 cm Ø petri dish	TPP
24-well transwell plates (5.0 µm, 8.0 µm)	Corning
6-well plate, flat bottom, transparent	Greiner Bio-One
96-well deep well plate, 2.2 mL	Fischer Scientific
96-well plate, flat bottom, Black	Greiner Bio-One
96-well plate, flat bottom, transparent	Greiner Bio-One
96-well plate, flat bottom, white	Greiner Bio-One
Adhesive optically clear plate seals	Thermo Fisher Scientific
Cell culture flasks (25 cm^2, T75 cm^2)	TPP
Cell culture flasks (175 cm^2)	Greiner Bio-One
Cell scraper	Corning
CIM Plate 16	OLS OMNI Life Science
Combitips advanced (1 mL, 5 mL, 10 mL)	Eppendorf
Conical tubes (15 mL, 50 mL)	Greiner Bio-One
Costar Ultra-low attachment plates (24-well, 6-well)	Corning
Cryovials (1.8 mL)	Nunc
E Plate 16 (PET)	OLS OMNI Life Science
Filter tips (10 µL, 20 µL, 200 µL, 1000 µL)	Neptune Scientific
Inoculation loops (10 µL)	Copan
Matrigel invasion chamber (8.0 µm)	Corning
MicroAmp optical 384-well reaction plate	Applied Biosystems
Microcentrifuge tube (1.5 mL, 2.0 mL)	Eppendorf
PCR strips	Steinbrenner
Pasteur capillary pipettes (230 mm)	Waltham
Serological Pipettes (2.5 mL, 5 mL, 10 mL, 25 mL)	Corning
Sterile filters (0.45 µM)	Sigma

Sterile syringes	Sigma

3.1.3 Chemicals and reagents

Agar	Sigma
Agarose	Roth
Ampicillin	Sigma
B27 supplement (1x)	Gibco
Bacto Trypton	Difco
CASYton	Roche Innovatis
Chloroform	Sigma
Complete Mini Protease Inhibitor Cocktail	Roche
DMEM	Gibco
DMEM/F-12	Gibco
DMSO	Sigma
Doxycycline	Takara
D-PBS	Gibco
EDTA	Sigma
EGF, recombinant	Corning
Epirubicin	Biomol
Ethanol	Sigma
Ethidium bromide	Sigma
Fetal Bovine Serum (FBS)	Gibco
FGF, recombinant human basic	R&D
Geniticin	Sigma
Glycerol	Roth
Glycine	Gerbu
Ham's F-12 Nutrient Mix	Gibco
Heparin sodium salt	Sigma
HEPES buffer solution	Gibco
Hoechst 33342	Thermo Fisher Scientific
Insulin	Sigma
Isopropanol	Sigma
Lipofectamine 2000	Invitrogen

Methylcellulose	Sigma
NEAA (non-essential amino acids, 100x)	Gibco
Nuclease-free water	Ambion
OptiMEM	Gibco
Paclitaxel	Biomol
PeqGold ladder (1 kb)	Thermo Fisher Scientific
PhosSTOP	Roche
Polybrene	Merck Millipore
Poly-L-lysine	Sigma
primaQUANT qPCR Probe Master Mix	Steinbrenner
Propidium iodide	Sigma
Puromycin	Gibco
Restriction enzymes and buffer	New England Biolabs
RNase	Qiagen
RPMI 1640	Gibco
SOC medium	Invitrogen
Tris HCl	Sigma
Tris base	Sigma
Trypsin-EDTA (0.05 %, 0.25 %)	Sigma
Tween 20	Sigma
Yeast extract	Gerbu

3.1.4 Commercial kits

DNeasy Blood & Tissue Kit	Qiagen
miRNeasy Kit	Qiagen
miScript precursor assays	Qiagen
miScript primer assays	Qiagen
miScript RT Kit	Qiagen
miScript SYBR Green PCR Kit	Qiagen
NEBuilder HiFi DNA Assembly Cloning Kit	New England Biolabs
NucleoBond Xtra Midi Kit	Macherey-Nagel
QIAprep Spin Miniprep Kit	Qiagen
RevertAid H Minus First Strand cDNA Synthesis Kit	Thermo Scientific

RNase-free DNase Set	Qiagen
RNeasy Mini Kit	Qiagen
Titanium Taq DNA polymerase	CLONTECH
Universal Probe Library	Roche
Wizard SV Clean Up System	Promega
XT Elements Master Kit	NanoString
XT Elements TagSet-84	NanoString

3.1.5 Solutions and buffers

LB Medium	10 g Bacto Trypton 5 g yeast extract 10 g NaCl dissolve in ddH2O up to 1 L, autoclave
LB-Agar	15 g in 1 L of LB medium, autoclave
50x TAE (Tris-acetate-EDTA)	242 g Tris base 57.1 mL acetic acid 100 mL 0.5 M EDTA (pH 8.0) ad 1 L ddH2O
TE-Tween	10 mM Tris pH 7.5 1 mM EDTA 0.1 % Tween 20
DNA precipitation buffer	9 mL ethanol (absolute) 300 µL 3 M NaAc (pH 5.2) 1800 µL ddH2O

3.1.6 Cell lines and growth medium

The parental cell lines used in this thesis are listed in Table 1. All parental cell lines were authenticated prior to and in the end of the study (Multiplexion, Heidelberg). Parental cell lines and stable cell lines derived from the parental cell lines were cultivated in the growth medium specified below and tested for potential mycoplasma contamination on a regular basis.

Table 1: Cell lines and growth medium used in this thesis.

cell line	obtained from	derived from	full growth medium
HEK293-FT	ATCC (PTA-5077)	embryonic kidney cells, human, transformed with SV40 large T-antigen	DMEM, 10 % FBS, 1 % NEAA, 1 % Geneticin
MDA-MB-231	ATCC (HTB-26)	breast adenocarcinoma (metastasis), human	RPMI-1640, 10 % FBS
HCC1806	ATCC (CRL-2335)	breast squamous cell carcinoma, human	RPMI-1640, 10 % FBS
SUM-159	Roberto Würth (A010, DKFZ)	pleomorphic breast carcinoma, human	Ham's F-12, 5 % FBS, 10 mM HEPES, 1 ug/mL Hydrocortisone, 5 ug/mL Insulin

3.1.7 Bacterial strains

Competent MACH1 (*E.coli*) Thermo Fisher Scientific

3.1.8 Mouse lines

NSG mice the mice were bred at the DKFZ mouse facility

3.1.9 Primers and oligos

All primers and oligos used for cloning, sequencing, TaqMan or as NanoString probes were purchased from Sigma-Aldrich. All sequences are given in 5' - 3' direction.

Sequencing primers

The miRseq5 primer (tgtttgaatgaggcttcagtac) published by Fellmann et al[130] was used for sequencing the pre-microRNA plasmids for the library.

TaqMan primers

TaqMan primers were designed using the online tool 'Assay Design Center' for the 'Universal Probe Library' (UPL) of Roche. The primers and the respective UPL probes can be obtained from Table 2.

Table 2: Primers used for TaqMan assays.

target	fw primer	rev primer	probe
House-keeping genes			
ACTB	ccaaccgcgagaagatga	ccagaggcgtacagggatag	64
GAPDH	gcccaatacgaccaaatcc	agccacatcgctcagacac	60
HPRT1	tgaccttgatttattttgcatacc	cgagcaagacgttcagtcct	73
Drug efflux pumps and detoxification enzymes			
ABCC1	aatgcgccaagactaggaag	ttctgtggggacttgacga	10
ABCC2	cttttcctggatcacctcca	ccatcatcaaggctgaaaaga	1
CAT	ctccggaacaacagccttc	atagaatgcccgcacctg	1
GPX1	caaccagtttgggcatcag	gttcacctcgcacttctcg	77
SOD2	aatcaggatccactgcaagg	taagcgtgctcccacacat	3
Breast cancer stem cell markers			
ALDH1A1	ccaaagacattgataaagccataa	cacgccatagcaattcacc	82
CD24	atgggcagagcaatggtg	ccagttgttgtttcactggaat	23
CD44	gacaccatggacaagttttgg	cggcaggttatattcaaatcg	13
ITGA6	tggcctcttcatttggctat	aaaatactgtggggctccaat	77
ITGB3	catccacgaccgaaaagaa	tgaaggtagacgtggcctct	76
PROM1	ggaaactaagaagtatgggagaaca	cgatgccactttctcactgat	86
NANOG	tctccaacatcctgaacctca	ttgctattcttcggccagtt	87
EMT marker			
CDH1	cccgggacaacgtttattac	gctggctcaagtcaaagtcc	35
FN1	gggagaataagctgtaccatcg	tccattaccaagacacacacact	25
CDH2	ggtggaggagaagaagaccag	ggcatcaggctccacagt	66
VIM	gaccagctaaccaacgacaaa	gaagcatctcctcctgcaat	39
SNAI1	tacagcgagctgcaggact	atctccggaggtgggatg	11
SNAI2	tggttgcttcaaggacacat	gcaaatgctctgttgcagtg	7
Others			
ATP5MD	ctccagctgtgaaagcaaca	ttatcacatgatgagttggcatt	80
PDCD11	gagagggcccttaagacca	cacccacacgttcagcttc	68
MYC	caccagcagcgactctga	gatccagactctgacctttgc	34

Primers for pre-amplification of the NanoString samples

The following primer pair was used to pre-amplify the pre-microRNA barcodes that were integrated into the genomic DNA. The primers bound to a sequence within the retroviral backbone RT3GEPIR[130], which was used to generate the parental cell lines with the pre-microRNA library.

fw aacgagaagcgcgatcacatggt

rev gggaacttcctgactaggggagga

Oligos for cloning of the pre-microRNAs and probes for subsequent detection by NanoString

Each pre-microRNA was designed with two or three partially complementary oligos that covered the pre-microRNA sequence obtained from miRBase. After annealing, the pre-microRNA construct had sticky ends that matched with the *Eco*RI- and *Xho*I-digested vector RT3GEPIR. The probe oligos that were used to detect the retrovirally integrated pre-microRNAs are listed in Table 3. The probes were designed by NanoString.

Table 3: NanoString probes used for pre-microRNA detection.

target	probe	sequence
pre-let-7c	A	GCTCCAAGGAAAGCTAGAAGGTTGTACAGTTAACTCCCAGGGTGTAACTTCCTTCCTGTGTTCCAGCTACAAACTTAGAAAC
pre-miR-100	A	CCTAACAGACACATACCTATAGATACAAGCTTGTGCGGACTAATACCACACATAAAATTGGTTTTGCCTTTCAGCAATTCAACTT
pre-miR-103a-1	A	CAATGCCTTCATAGCCCTGTACAATGCTGCTTGATCCATATGCAACAACTGGTCAAGACTTGCATGAGGACCCGCAAATTCCT
pre-miR-103a-2	A	TGGTTCTTTCATAGCCCTGTACAATGCTGCTTGACCTGAATGCTACCTTTCGTTGGGACGCTTGAAGCGCAAGTAGAAAAC
pre-miR-106b	A	CCTGCTGGAGCAGCAAGTACCCACAGTGCGGTAGCCCAGCAGACCTGCAATATCAAAGTTATAAGCGCGT
pre-miR-10a	A	AGAGCGGAGTGTTTATGTCAACTACATATTCCCCTAGATACGAATTTGTGCCTGCCAATGCACTCGATCTTGTCATTTTTTTGCG
pre-miR-10b	A	TGAAGTTTTTGCATCGACCATATATTCCCCTAGAATCGAATCTGTGACTACAAACTGGAGAGAGAAGTGAAGACGATTTAACCCA
pre-miR-125b-1	A	AGCACGACTCGCAGCTCCCAAGAGCCTAACCCGTGCGATTGCTGCATTCCGCTCAACGCTTGAGGAAGTA
pre-miR-125b-2	A	TCCCCTCCGCCTAGGTCCCAAGAGCCTGACTTGTGCTGAGGCTGTTAAAGCTGTAGCAACTCTTCCACGA
pre-miR-126	A	TGCCGTGGACGGCGCATTATTACTCACGGTACGAGCTAGGACGCAAATCACTTGAAGAAGTGAAAGCGAG
pre-miR-127	A	GATGATGAGACTTCCGACCAGCCAAGCTCAGACGGATCCACGCGATGACGTTCGTCAAGAGTCGCATAATCT
pre-miR-1307	A	TGCATGACCGCCTATCTACCACGACCGACGCCACGCATTTGGAATGATGTGTACTGGGAATAAGACGACG

pre-miR-130b	A	GACCTGACCGATGCCCTTTCATCATTGCACTGCTTCCACAAGAATCCCTGCTAGCTGAAGGAGGGTCAAAC
pre-miR-139	A	GTTACTCCAACAGGGCCGCGTCTCCAGCCTCCGAGCTTGACGTAGATTGCTATCAGGTTACGATGACTGC
pre-miR-140	A	GGTGCCCCGGTATCCTGTCCGTGGTTCTACCCTGTCTTACAGATCGTGTGCTCATGACTTCCACAGACGT
pre-miR-141	A	GAACCCACCCGGGAGCCATCTTTACCAGACAGTGTCTTGGAGGAGTTGATAGTGGTAAAACAACATTAGC
pre-miR-142	A	CACAGTACACTCATCCATAAAGTAGGAAACACTACACCCTCCAGTGCTGTCCTACGTATATATCCAAGTGGTTATGTCCGACGGC
pre-miR-143	A	GCTGCAGAACAACTTCTCTCTTCCTGAGCTACAGTGCTTCATCTCCAGCAAGAAGGAGTATGGAACTTATAGCAAGAGAG
pre-miR-144	A	GGCGGTGCCCGGACTAGTACATCATCTATACTGTAGTGTCTCATCCACCCCTCCAAACGCATTCTTATTGGCAAATGGAA
pre-miR-145	A	AACCATGACCTCAAGAACAGTATTTCCAGGAATCCCCATCTTAGCATCTACCCGAAGCAATACTGTCGTCACTCTGTATGTCCGT
pre-miR-148b	A	TAGAAAGCTTTCGAGACAAAGTTCTGTGATGCACTGACTTTCAGAGAGCCCCGGGAATCGGCATTTCGCATTCTTAGGATCTAAA
pre-miR-151a	A	GAGGTGAGTATGACCATCCCTGTCCTCAAGGAGCTTCAGTCTAGTACCGATCTTCATAACGGACAAACTGAACGGGCCATT
pre-miR-155	A	CTGTTAATGCTAATATGTAGGAGTCAGTTGGAGGCAAAAACCCCTATCACCGCTATGCAGACGAGCTGGCAGAGGAGAGAAATCA
pre-miR-16-1	A	GTCAACCTTACTTCAGCAGCACAGTTAATACTGGAGATAATTTTAGAATCCATTCGCAACCATGTGAAGTAATGTGAGCGTACTT
pre-miR-16-2	A	GTCACACTAAAGCAGCACAGTAATATTGGTGTTTAATATATATTTCACTACACCAGTTAGCGTGGCGTATACCATGTTGTTAACA
pre-miR-17	A	GTCACCATAATGCTACAAGTGCCTTCACTGCAGTAGATGCACATATCACTCCTGAATCAATAGAACAATATCAGTTATGGCGGTG
pre-miR-182	A	GTGCCGGCTGAGTCCTCGCCCCATAGTTGGCAAGTCGGTTGTTAATATGACAGGCCGCTAAAGACGTTCT
pre-miR-183	A	TCGTGGATCTGTCTCTGCTCTGTTTATGGCCCTTCGGTAATTCCCGTCTCAGATGAGTGGGTTAATCAATCAAGTATG
pre-miR-190b	A	CTGCTGCTGTAAGAATATGTTTGACATTTAGTTGGTTCCTAATTAAACAACTGACACATTAGTAACGTCGGCAAGCACTTAGTCG
pre-miR-191	A	AGGCAGGAGAGCAGGGGACGAAATCCAAGCGCAGCCGTGAACCAGATTATGTATGGACGCGCAATAGATA
pre-miR-192	A	GCTGGCATTGAGGCGAACATACCTGTGACCTATGGAATTGCATACGAAATTTGAGCAAGCAATTGAAGGCTTAGA
pre-miR-200a	A	GCGGGTCACCTTTGAACATCGTTACCAGACAGTGTTAGAGTCAAGCTATCAGCTAATAGGGTCGGCTCAACAGTGTATCC
pre-miR-200c	A	CCTCCATCATTACCCGGCAGTATTAGAGACTCCCAACCGCTATCAATTCGTGACCCCGATCATCCAGTCCAGAA
pre-miR-203a	A	TCGCTGTCGCCGCGCCCGCCGGGTCTAGTGGTCCTCTTGAGCTCTAGGCCCAAAACGACCTTAATGGTCA
pre-miR-204	A	GCCAGTGATGACAATTGAACGTCCCTTTGCCTTCCCACTAGCCCAGATCCTACGAGATGAGCTACGTAACTA
pre-miR-21	A	TGTCAGACAGCCCATCGACTGGTGTTGCCATGAGATTCAAATGCACTCTATATGGAGGGAGAGTAGCTGGAT
pre-miR-210	A	GGGTCGCGCTGCCCAGGCACAGATCAGCCGCTGTCCCTGGTCTAGGTATCTAATTCGTGGGTCGGGTACT

pre-miR-22	A	GGCAGAGGGCAACAGTTCTTCAACTGGCAGCTTTAGCATTAGCTCGGATGCTATCAGCTTGCGCCTATTAT
pre-miR-29a	A	ATAACCGATTTCAGATGGTGCTAGAAAATTATATTGACTCTGAACACCAACACGATCTGTATTTTGCACCTTTCGCTATGCTGAG
pre-miR-29b-1	A	CCCCCAAGAACACTGATTTCAAATGGTGCTAGACAATCACTATTTAAATCCTGTGTCCGTCTATACGCATACTGGTCCACATATA
pre-miR-29b-2	A	CTCCTAAAACACTGATTTCAAATGGTGCTAGATACAAAGATGGAAAAATCCATGTTGGAGTTAACGGAGACCCGCCATCGTTTAC
pre-miR-3065	A	CTGTCCTCTCCAACAATATCCTGGTGCTGAGTGATGACTCAGGCGCTCATTTTGAACATACGATTGCGATTACGGAAA
pre-miR-326	A	TGAATCCGCCTCGGGGCTGGAGGAAGGGCCCAGAGCCTATGCATCATGTGCCTCACTAGGACATCATGCT
pre-miR-337	A	TTGAAGGGGATGAAGAAAGGCATCATATAGGAGCTGGATAACTGTGCATCCCTAAATTGGGAAAAAAGGTTTTAGCTATTGATGG
pre-miR-342	A	TAAGTAGGCCAAGGTGACGGGTGCGATTTCTGTGTGAGCTTCAGTTAAAGGCTATCTTGCTCCGCTCGTTCTC
pre-miR-365a	A	TGCAAGAGCAATAAGGATTTTTAGGGGCATTATGATAGTGGAATGGAAACCTTAAAGCTATCCACGAATGTCAAAAATGTGGTTT
pre-miR-375	A	GCCTCACGCGAGCCGAACGAACAAAACGCTCAGGTCCCGAATGTATAATGCTGACGTTCTTGCTTTTGGC
pre-miR-378a	A	AGGCCTTCTGACTCCAAGTCCAGTGCTATTTCTAGGTAACACACAGCCTATTGAAGCAATCCTCTCCCCAATACTTAAAAA
pre-miR-379	A	AGAGTTAGTGGACCATGTTACATAGGTCAGAAATCATAACGCCTACGTTCCTACGGTTACCGTCTTTATAAGTGAACAAAACCGG
pre-miR-381	A	TACTCACAGAGAGCTTGCCCTTGTATATTCCATGTCAATAAACCGAATATCTCTGTGAACTGTCATCGGTCCGATCAATTAGTCT
pre-miR-425	A	GAAAGAGCACTGGGCGGACACGACATTCCCGATGGCTCCCCTTTCCCAAGTAAATGTACGGGAATTATCG
pre-miR-451a	A	TCTGGGTATAGCAAGAGAACCATTACCATTACTAAACTCAGTAATGGTAACGCTTTATTATGTGTTCGTCTAACTCTGTTTCTGT
pre-miR-452	A	GCAAAGCACTTACTTCTTTGCAGATGAGACTGAGACATAGTTACAAAGTCCCGAGTGCATGAGCTGTCTTTCACATGATACATCG
pre-miR-455	A	GATGACATAGGCCTTGAGGCAAGTGTATATGCCCATGGACTGCATGGTGCCTATTTCTGTTCACGGATGAAGGCCTATATCAATG
pre-miR-486-1	A	GTATCCTGTACTGAGCTGCCCCGAGCTGGGCAGCACCATCCACTTTCATGGAAACAATAAGAGCAGGGAA
pre-miR-486-2	A	CATCCTGTACTGAGCTGCCCCGAGGCCCTTCATGCCACAAACTCACTACTACCAACAACCTCACCAAAAA
pre-miR-497	A	CCTCGGCGGTGCCTCCCCCACCCTCGCTCTAACACCTCATGTCCTCTGTTAATCCAGCCTGAATATGCCA
pre-miR-551b	A	TTATTCTCACAGCCTCTGAAACCAAGTATGGGTCGCCTTCCCAGAAATGTCACTCCCATGGTGGCTGATATAGAAA
pre-miR-7-1	A	CTGTAGAGGCATGGCCTGTGCCATATGGCAGACTGCATGTCGAACCTTGGATAGGAGCGACCGATTACGT
pre-miR-92b	A	GGGCCGGGCGGGCCGGAGGCCGGGACGAGTGCAATCTCAGGTTGTTACTTGAAGGGTTCAACACGAGCTC
pre-miR-93	A	CCGGCGGCTCGGGAAGTGCTAGCTCAGCAGTAGGTCAGAAGATCAAAAAACGATCCCTGTCCATCAATAC
pre-miR-99a	A	CACACTGACACAGACCCATAGAAGCGAGCTTGTGCCTTAGGCTACCAAATGAATTTAAAGCCAGCTGAAA

pre-miR-99b	A	GACACGGACCCACAGACACGAGCTTGTGTGCGGCGCCAATGCTTGCAG TATGTATCCTGATCGTGCGTGC
ctrl0_RT3GEPIR	A	GCATAGGAATTATAATGCTTATCTATACATCTGTGGCTTCACTATAGATA CCTGCATTCTCATGGAAATGCAATGGATTCATTCC
ctrl1_cel-pre-miR-67	A	AAGTTTTAAAATCGATCTACTCTTTCTAGGAGGTTGTGATGCTTAATCTG CCTGTTGCAGTATCACGTAAATACCTACTTCGATA
ctrl2_cel-pre-miR-239b	A	AGATAAAAGCAACTTGCCATTTTTGCACACCACAAAAGTGCTGAGCCTA GCTGTTATGGCTATTGCTGAAACAGCAAAATT
ctrl3_cel-pre-miR-1022	A	AGCCTTGAACAGCTGGATCATCATTGGACTATCATCTTTATATTGCTTCA CCTTACGACTTCACTGCAATTGACGATTCAGTTAA
ctrl4_cel-pre-miR-254	A	AAAAACTGCATGTTCGCCGCCTACAGTCGCGAAAGATTTGCCTCATACC AATGTAAAGTATAGTTAACGCCCTGT
ctrl5_cel-pre-miR-36	A	TCCGCGTCGGGGACCCATGCGAATTTTCACCCGGTCATCTCCATGACTG CTTGAGCGGCTGGAGAATCTG
ctrl6_cel-pre-miR-71	A	TTCCAGGTCACGATCCCGACGGCGAAAAACAGAATAGTGATACCTTTC GCCACCCATATAAACCCCACTTCGTCCTCA
ctrl7_cel-pre-miR-800	A	ACGGCGGCAGACAATTTCCGAGTTTGGCCACTGATTATAACAAGGCAG AGCAAATGTGACACTGTCTATCAGTAC
ctrl8_cel-pre-miR-90	A	TGGCATCCAATTCAAGGGGCATTCAAACAACATATCAACACGCAAAAG TGCCTACATATATAGGAAAAGGGAAGGTAGAAGAGCT
all constructs	B	CGAAAGCCATGACCTCCGATCACTCTAAACAAGATAATTGCTCGAATTC TAGCCCCTTGAAGTCCGAGGCAGTAG

3.1.8 siRNAs and microRNA mimics

All siRNAs and microRNA mimics as well as the respective non-targeting controls used in this thesis are listed in Table 4. All sequences are given in 5'-3' direction. The siRNA pools obtained from siTools Biotech contained 30 different siRNAs targeting the gene of interest.

Table 4: siRNAs and microRNA mimics used in this study.

mimic	**sequence**		**company**
siAllStars	-		Qiagen
mimic ctrl2	-		Dharmacon
miR-1307-3p l0	A**CUCGGC**GUGGCGUCGGUCGUG		Dharmacon
miR-1307-3p l1	C**UCGGCG**UGGCGUCGGUCGUG		Dharmacon
siRNA	**sequence**		**company**
sictrl1	-		siTOOLs Biotech
sictrl2	-		siTOOLs Biotech
siMyc	GAGAACAGTTGAAACACAA	GGACTTGTTGCGGAAACGA	siTOOLs Biotech
	GCCATAATGTAAACTGCCT	GAGGAGCAAAAGCTCATTT	
	GGTACTATAAACCCTAATT	CAGCATACATCCTGTCCGT	
	GGAAAACGATTCCTTCTAA	GAGCTAAAACGGAGCTTTT	
	GGCGAACACACAACGTCTT	CTGAAAGATTTAGCCATAA	
	CCCTGGTGCTCCATGAGGA	CCTAGTATTATAGGTACTA	
	CTCACAACCTTGGCTGAGT	GGGTCAAGTTGGACAGTGT	
	GCATGATCAAATGCAACCT	CTCCTACGTTGCGGTCACA	
	CCCAAGGTAGTTATCCTTA	CCCTACCCTCTCAACGACA	
	CTGCCTCAAATTGGACTTT	GCCACAGCAAACCTCCTCA	
	GCCACGTCTCCACACATCA	CAGATCCCGGAGTTGGAAA	
	GACTATCCTGCTGCCAAGA	CAGAGGAGGAACGAGCTAA	
	CGGTGCAGCCGTATTTCTA	GACATGGTGAACCAGAGTT	
	CCTATGAACTTGTTTCAAA	CGACGAGACCTTCATCAAA	
	GTCCTGAGCAATCACCTAT	CTGCTCTCCTCGACGGAGT	

3.1.9 Plasmids

pHIT60	kindly provided by Yuko Soneoka[131]
pMD2.G	Addgene
RT3GEPIR	kindly provided by Christof Fellmann[130]

3.1.10 Databases and software

cBioPortal	http://www.cbioportal.org/
Cellosaurus	https://web.expasy.org/cellosaurus/
COSMIC	https://cancer.sanger.ac.uk/cosmic/
GraphPad Prism 5	http://www.graphpad.com/
MiRanda	http://www.microrna.org/microrna/home.do
miRBase	http://www.mirbase.org/
Molecular Devices Analysis Software	Molecular Devices
Molecular Signature Database	Broad Institute
NCBI	http://www.ncbi.nlm.nih.gov/
nSolver Software	NanoString
QuantStudio Software	Thermo Fisher Scientific
Roche UPL Design Center	Roche
UCSC Genome Browser	https://genome.ucsc.edu/
SDS 2.2	Applied Biosystems
TargetScan	http://www.targetscan.org/
TCGA	http://cancergenome.nih.gov/

3.2 Methods

3.2.1 Cloning of the pre-microRNA library

Preparation of the vector

1-2 µg of the vector RT3GEPIR were digested with *Eco*RI and *Xho*I in parallel. For the digest, 5 µL NEB CutSmart Buffer (10x) were added to the vector and 1 µL of each restriction enzyme (20 units). The reaction volume was adjusted to 50 µL with ddH2O and incubated over night at 37°C. This step was followed by heat inactivation at 65°C for ten minutes. The linearized plasmid was purified with the Wizard SV Clean Up System according to the manufacturer's instructions.

Preparation of the pre-microRNAs

To generate pre-microRNA constructs, two or three oligos that were partially complementary to each other (15-80 base pairs overlap) were designed. For annealing of individual pre-microRNAs, 5 µL of each oligo (100 µM) were combined in a well of a 96-well plate. 2 µL NEB ligation buffer (10x) were added and the reaction was filled up to 20 µL with ddH2O. The mix was boiled at 95°C for five minutes and allowed to cool down slowly at room temperature.

Ligation via NEBuilder reaction

The linearized vector RT3GEPIR was ligated with the annealed pre-microRNA oligos using the NEBuilder HiFi DNA assembly Master Mix. Since 72 pre-microRNAs were selected for the library, the constructs were cloned in pools. Ten or eleven of the annealed pre-microRNA oligos from the previous step were combined and diluted to a final concentration of 0.5 µM per pre-microRNA. The reaction mix was set up as described below and was incubated at 50°C for one hour.

NEBuilder reaction:	30 ng vector (linearized)
	2 µL annealed oligos (0.5 µM, to use 1 µmol per pre-microRNA)
	10 µL NEBuilder HiFi DNA assembly Master Mix (2x)
	up to 20 µL with ddH2O

Bacterial transformation and expansion

The plasmids containing the individual pre-microRNAs for the library were transformed into competent Mach1 *E.coli* cells. Competent Mach1 cells were thawed on ice and 2 µL of NEBuilder reaction (containing up to ten different pre-microRNA constructs) from the previous step were added to 50 µL of competent cells and incubated on ice for 30 minutes. The mixture received a heat pulse at 42°C for 90 seconds and was incubated on ice for three to five minutes. Subsequently, 350 µL SOC medium were added to the transformation mix and the bacteria were incubated in a shaker at 37°C for one hour. 200 µL of the transformed bacteria were plated on LB agar plates supplemented with 100 µg/mL Ampicillin and incubated over night at 37°C. The next day, colonies were picked and cultivated in 5 mL LB medium containing 100 µg/mL Ampicillin. The cultures were grown over night at 37°C and the bacteria were harvested for plasmid isolation (Miniprep) the next day. After the correct sequences were confirmed by Sanger sequencing (GATC, Konstanz), larger over-night cultures were inoculated: 500 µL of the 5 mL bacteria suspension grown over night were added to 100 mL LB medium supplemented with 100 µg/mL Ampicillin. Plasmid isolation was performed (Midiprep) to obtain large plasmid quantities for generation of the stable cells harboring the pre-microRNA library. The correct sequence was confirmed again by Sanger sequencing.

Plasmid isolation

For **minipreps**, the QIAprep Spin Miniprep Kit was used. 4 mL from the over-night culture were pelleted in a 2 mL microcentrifuge tube at 3,000x g for ten minutes. The bacteria pellets were resuspended in 250 µL P1 buffer and lysed with 250 µL P2 buffer for five minutes. 350 µL of neutralization buffer were added and the mix was inverted five times. After centrifugation at 6,000x g for ten minutes, the supernatant was applied to a column and spun down at 6000x g for one minute. After the flow-through was discarded, the column was washed with 500 µL PB buffer and centrifuged at 6000x g for one minute. The flow-through was discarded and the column was washed with 750 µL PE buffer and then centrifuged at 6,000x g for one minute. The flow-through was discarded and the columns were placed into a new collection tube. Residual ethanol was removed by centrifuging the tubes at 6,000x g for one minute. To elute the plasmids, the columns were placed into 1.5 mL microcentrifuge tubes and 35 µL ddH2O were added. After one minute of incubation, the plasmid DNA was collected by spinning at 6000x g for one minute. Plasmid concentration and quality were measured by NanoDrop.

Midipreps were performed with the NucleoBond Xtra plasmid purification kit. Bacteria were pelleted at 4,000x g for five to ten minutes. The cell pellet was resuspended completely in 8 mL resuspension buffer containing RNase A by pipetting the cells up and down or vortexing. Then, 8 mL lysis buffer were added and the mix was inverted five times and incubated for five minutes. In the meantime, filters were inserted into the NucleoBond Xtra Columns and the columns were equilibrated with 12 mL equilibration buffer. 8 mL neutralization buffer were added to the lysate and invert immediately 10-15 times. The sample was centrifuged at 5,000x g for ten minutes and the supernatant was applied to the filter in the column. Subsequent to the lysate, 5 mL equilibration buffer were applied to the filter. In the next step, the filter was discarded and the column was washed with 8 mL washing buffer. The plasmids were eluted with 5 mL elution buffer, the flow-through was collected in a 50 mL tube. To precipitate the plasmid DNA, 3.5 mL isopropanol (room temperature) were added and the sample was vortexed properly. The plasmid DNA was pelleted at 5,000x g for 15 minutes. After the supernatant was discarded carefully, the pellet was washed with 2 mL 70 % ethanol and spun down at 5,000x g for five minutes. The supernatant was discarded and the remaining liquid was carefully removed with a pipette. The pellet was dried at room temperature for five to ten minutes with the lid open. The plasmid DNA pellet was dissolved in 100-150 µL TE buffer. Remaining ethanol evaporated at 60°C for five minutes (lid open). Plasmid concentration and quality were measured by NanoDrop.

Sanger sequencing

To verify the correct pre-microRNA sequence, the purified plasmid harboring the pre-microRNA was Sanger sequenced by GATC using the miRseq5 primer.

3.2.2 General cell culture

Passaging, freezing and thawing

All cell lines used in this thesis were adherent and were cultivated in 75 cm^2 or 175 cm^2 flasks at 37°C with 5 % CO2 in a humidified incubator. In order to provide enough space and nutrients, all cell lines were passaged approximately every three to four days and received fresh full growth medium (10-15 mL for 75 cm^2 and 25-30 mL for 175 cm^2 flasks). To allow the trypsin (0.25 %) to detach the cells for passaging, they were washed with PBS after the growth medium was

aspirated. The cells were incubated with 1.5 mL trypsin at 37°C. The trypsin was inactivated with 8.5 mL full growth medium after five minutes. For all assays, 50 µL cell suspension were diluted with 10 mL CASYton and cells were counted with a CASY cell counter. To have the cells at 70-80 % confluency after three or four days, the cell numbers indicated in Table 5 were seeded for the individual cell lines. All cell lines were kept in culture no longer than three months, then new vials were thawed.

Table 5: Cell numbers seeded for different cell culture flasks.

cell line	75 cm^2 (3 days)	75 cm^2 (4 days)	175 cm^2 (3 days)	175 cm^2 (4 days)
HEK293-FT	1,000,000	800,000	-	-
MDA-MB-231	800,000	500,000	2,300,000	2,000,000
HCC1806	600,000	400,000	2,000,000	1,700,000
SUM-159	250,000	150,000	300,000	175,000

Each cell line was expanded at a low passage and aliquots were frozen for long-term storage in liquid nitrogen. Per vial, 1-1.5 million cells were pelleted (0.5 million for SUM-159) at 1,500 rpm for five minutes. Each cell pellet was resuspended in 1 mL of freezing medium (full growth medium supplemented with 10 % FBS and 10 % DMSO) and transferred into 1.8 mL cryotubes. The aliquots were slowly frozen in an isopropanol bath at -80°C. After at least 24 hours, the frozen cell stocks were transferred to a liquid nitrogen tank for long-term storage. To thaw the cells, the vials were put in a 37°C water and diluted with 10 mL full growth medium. As soon as the suspension was thawed, the cell suspension was centrifuged at 1,500 rpm for five minutes. The supernatant was removed and the cell pellet was resuspended in 15 mL full growth medium, then cells were seeded into a 75 cm^2 flask. As soon as the cells had attached or latest the next day, the medium was replaced by fresh growth medium to remove dead cells and remaining DMSO.

Transient transfection

For transient transfection, all cell lines were seeded at 70-80 % confluency the day before transfection. microRNA mimics or siRNAs were transfected with Lipofectamine 2000 in OptiMEM medium. The respective amounts and volumes used for transfection in different cell culture plates are listed in Table 6. First, Lipofectamine 2000 was added to OptiMEM and incubated for 15 minutes. The mix was inverted occasionally. In the meantime, the siRNAs or microRNA mimics

were diluted in OptiMEM. The amount of transfected microRNA or mimic controls had a final concentration of 30 nM in the well. For the siRNAs and controls from siPools the final concentration was 2 nM. After incubation, both mixes were combined and incubated for 20 minutes. The mix was inverted several times during the incubation. To transfect cells in 96-well plates with the mix, the cell culture medium was aspirated and replaced by 100 µL transfection mix. For transfections in 6-wells, the medium was aspirated and replaced by 800 µL OptiMEM. Then, 200 µL transfection mix were added dropwise. The transfected cells were incubated for five hours at 37°C and 5 % CO2 in a humidified atmosphere. The transfection mix was aspirated and replaced by 100 µL or 2 mL regular growth medium for 96-well plates or 6-wells plates, respectively. For RNA extraction, cells were harvested 48 hours after transfection. The reseeding for different functional assays was performed 48 hours after transfection as well.

Table 6: Transfection mixtures for different plate formats.

	plate/dish	**96-well**	**6-well**
mix 1	Lipofectamine 2000 (µl)	0.4	4
	in OptiMEM (µL)	50	100
mix 2	siRNA/microRNA mimics (nM)	2 / 30	2 / 30
	in OptiMEM (µL)	50	100
volumes	final transfection mix (µL)	100	200
	OptiMEM (µL)	0	800
	finale volume per well (µL)	**100**	**1000**

3.2.3 Stable cell lines

Virus production

Transient transfection of HEK293-FT for virus production was performed similar to the microRNA mimics and siRNA transfections described above. 10 cm cell culture dishes were coated with poly-L-lysine solution (0.1 mg/mL) since HEK293-FT cells easily detach after transfection. The plates were coated for 30 minutes and washed twice with ddH2O before use. The day before transfection, 2 million HEK293-FT cells were seeded in 10 mL medium per dish. To produce virus, the cells were transfected with an equimolar plasmid pool of the pre-microRNA library and with a retroviral packaging system. The respective amounts and volumes are listed in Table 7.

Table 7: Transfection of HEK293-FT for virus production.

	plate/dish	10 cm dish
mix 1	Lipofectamine 2000 (µl)	10
	in OptiMEM (µL)	250
mix 2	plasmid pMD2.G (ng)	800
	plasmid pHIT60 (ng)	3500
	plasmid of interest (ng)	3500
	in OptiMEM (µL)	250
volumes	final transfection mix (µL)	500
	OptiMEM (µL)	4500
	finale volume per well (µL)	**5000**

One day after transfection, the medium was changed to the full growth medium of the target cell lines in order to produce the virus into the correct medium since the target cells are not resistant to the Geneticin in the medium of the HEK293-FT cells. The second day after transfection, the medium in the dish was collected and fresh full growth medium of the target cell line was added. The collected medium was centrifuged at 1,500 rpm for five minutes to remove dead HEK293-FT cells and cell debris. The supernatant was transferred to a fresh tube and stored at 4°C. The supernatant of the virus-producing HEK293-FT cells was collected again three days post transfection. After centrifugation at 1,500 rpm for five minutes, the supernatant was mixed with the supernatant collected two days after transfection. The virus-containing media was then frozen at -80°C. The HEK293-FT cells were trashed at this point.

Stable transduction of the target cell lines

The target cell lines were seeded in 75 cm^2 flasks at a density that allowed them to be confluent after three or four days. Moreover, target cells were seeded into 6-well plates to check the multiplicity of infection (MOI). The target cells were transduced with the retrovirus one day after seeding. For transduction, the virus-containing supernatant was thawed as fast as possible at 37°C in the water bath and resuspended. The medium of the cells was aspirated and replaced by 10 mL fresh growth medium per 75 cm^2 flask. Each dish received 5 mL of the virus-containing supernatant supplemented with 12 µL Polybrene (stock concentration: 10 mg/mL). The resulting Polybrene concentration in the diluted virus-containing supernatant (1:3) was 8 µg/mL. Target cells seeded in 6-wells for MOI evaluation, received 2 mL of the final transduction mix supplemented with 2 µg/mL Doxycycline to induce expression of the pre-microRNA library and GFP. Two days after transduction, the cells transduced in 6-wells were stained with Hoechst. By

microscope, the total cell number (Hoechst) and the number of transduced cells (GFP upon Doxycycline-induction) was determined. If the MOI was low enough (<0.1), Puromycin selection of the stable cells was performed.

Selection and cultivation of stable cell lines

For Puromycin selection of the successfully transduced cells, the cells were washed several times with PBS to remove dead cells and residual viral particles 48 hours post transduction. Then, fresh full growth medium containing 1 µg/mL Puromycin was added. The stable cell lines were expanded under Puromycin selection and stocks were generated and stored in liquid nitrogen. The stable cell lines were frozen and thawed in medium without Puromycin. Only when cells started to proliferate after thawing, Puromycin was added again. For functional assays, Puromycin-free medium was used.

3.2.4 Expression analysis

mRNA and microRNA isolation

Isolation of **mRNA** was performed with the RNeasy Mini Kit. A maximum of 10 million cells was lysed in 350 µL RLT buffer at room temperature for five minutes and was then collected in a 1.5 mL microcentrifuge tube. 350 µL of 70 % ethanol were added and mixed with the sample by pipetting. The entire sample was loaded on a column and spun down at 13,000 rpm for 30 seconds. After discarding the flow-through, 350 µL RW1 buffer were added to the column and spun down at 13,000 rpm for 30 seconds. The flow through was discarded and 80 µL DNase I dilution (10 µL DNase I stock solution and 70 µL RDD buffer) were applied to the column and incubated at room temperature for 15 minutes. Afterwards, the column was washed with 350 µL RW1 buffer at 13,000 rpm for 30 seconds. The flow-through was discarded and 500 µL RPE buffer were added to the column. The column was centrifuged at 13,000 rpm for 30 seconds. Another 500 µL of RPE buffer were added and centrifuged at 13,000 rpm for two minutes. The column was placed into a fresh collection tube and centrifuged dry 13,000 rpm for one minute. To elute the mRNA, the column was placed into a 1.5 mL microcentrifuge tube and 35 µL of ddH2O were added. After one minute of incubation, the eluted mRNA was collected by centrifugation at 13,000 rpm for one minute. The mRNA concentration was determined by NanoDrop.

For **microRNA** isolation, the miRNeasy Mini Kit was used. Cells were lysed in 700 µL Qiazol Lysis Reagent at room temperature for five minutes. Then, 140 µL chloroform were added and the samples were vortexed thoroughly. After incubation at room temperature for 2-3 minutes, the samples were centrifuged with 13,000 rpm at 4°C for 15 minutes. The upper aqueous phase (roughly 350 µL) was transferred into a clean 1.5 mL microcentrifuge tube and 1.5 volumes of 100 % ethanol were added. The samples were mixed by vortexing, 700 µL of the mix were loaded on a column and centrifuged at 13,000 rpm for 30 seconds. The following steps were identical to the mRNA isolation protocol, starting with washing the column with 350 µL RW1 buffer and the DNase digest.

cDNA Synthesis

For mRNA, reverse transcription into cDNA was performed with the RevertAid RT Reverse Transcription Kit. 2-3 µL of mRNA containing approximately 1-1.5 µg were incubated at 72°C for two minutes and put on ice afterwards. Then, RT reaction master mix was added and cDNA synthesis was performed as described below.

RT reaction:	1.2 µL reaction buffer (5x)
	0.25 µL of oligo-dT-primer (100 pmol/µL)
	0.6 µL dNTPs (10 mM)
	0.3 µL Prime RNase-Inhibitor (20 U/µL)
	0.2 µL RevertAid Reverse Transcriptase (200 U/µL)
	0.55-1.55 µL ddH2O (depending on mRNA input)

RT protocol:	37°C	5 min
	42°C	60 min
	70°C	10 min

The cDNAs were diluted to a final concentration of 2 ng/µL for gene expression analysis by TaqMan and were stored at -20°C.

For **microRNAs** expression analysis, cDNA was synthesized with the miScript RT kit as indicated below. The HiFlex buffer was used for pre-microRNA assays, the HiSpec buffer for mature microRNA. Subsequent to reverse transcription, cDNAs derived from pre-microRNAs were diluted appropriately and stored at -20°C.

RT reaction:	12 µL total RNA (10-2000 ng, max. 500 ng for pre-microRNA assays)
	4 µL miScript HiFlex or HiSpec buffer (5x)
	2 µL miScript Nucleics Mix (10x)
	2 µL miScript RT Mix

RT protocol:	37°C	60 min
	95°C	5 min

Quantitative real-time PCR

mRNA expression analysis was performed with TaqMan assays, probes were obtained from the Roche universal probe library (UPL) and primers were designed with the Roche UPL Design Center. Each sample was analyzed in technical triplicates in a 384-well plate. The composition of the reaction mix and the qRT-PCR protocol are listed below.

qRT-PCR reaction:	5 µL cDNA (2 ng/µL)
	5.5 µL primaQUANT qPCR Probe Master Mix (2x)
	0.11 µL of 20 µM forward primer
	0.11 µL of 20 µM reverse primer
	0.11 µL UPL probe
	0.17 µL ddH2O

qRT-PCR protocol:	1x	50°C	2 min
	1x	95°C	15 min
	45x	95°C	15 sec
		60°C	1 min

ACTB, *GAPDH* and *HPRT1* were included as house-keeping genes. Raw data was evaluated with SDS or QuantStudio and the cycle threshold was adjusted manually if necessary. For analysis, the median of the technical triplicates per sample was used. The fold change in gene expression between two conditions was calculated by normalizing the gene of interest to the mean of the three house-keeping genes using the $2^{-\Delta\Delta CT}$ method[132].

For **microRNA** quantification, the miScript SYBR Green PCR kit was used in combination with miScript precursor assays or miScript primer assays. The miScript precursor assays or miScript primer assays contain primers for the amplification of a specific pre-microRNA or mature microRNA, respectively. Quantification of (pre-)microRNAs was performed as indicated below.

qRT-PCR reaction:	3 µL cDNA (20 ng for pre-microRNA detection, 3 ng for microRNAs)
	5 µL SYBR Green Master Mix (2x)
	1 µL miScript primer assay (10x) or miScript precursor assay (10x)
	1 µL miScript universal assay (10x) for (pre-)microRNAs
	up to 7 µL with ddH2O

qRT-PCR protocol:	1x	95°C	15 min	
	45x	94°C	15 sec	
		55°C	30 sec	
		70°C	30 sec	
	melt curve:	95°C	15 sec	1.6°C/s
		60°C	1 min	1.6°C/s
		95°C	15 sec	0.075°C/s

SNORD61, SNORD72 and SNORD95 were included as house-keeping genes. The data was analyzed as described previously for mRNA-based expression analysis. In addition, the primer specificity was evaluated based on the melt curves.

3.2.5 Phenotypic assays

The phenotypic assays for the pre-microRNA-1307 overexpression cell lines were performed by Xiaoya Li and are described in her thesis. This includes cell cycle, apoptosis and transwell-based migration/invasion assays.

Reseeding for various phenotypic assays

To seed defined cell numbers of successfully transfected cells for phenotypic assays, the cells were reseeded 48 hours after transfection. Transient transfection was performed in 6-wells as previously described and the cells were starved (growth medium without FBS) for 16-20 hours prior reseeding. The cells were reseeded for proliferation, chemoresistance and migration assays. The cell numbers seeded for transfection in 6-wells and reseeded for the different assays can be obtained from Table 8. Proliferation, migration and mammosphere assays compared different conditions (e.g. microRNA mimics compared to non-targeting control). To account for differences in the cell concentration of the individual cell suspensions (e.g. due to differences in pipetting or counting), a seeding control was included. For the seeding control, cells were seeded into black 96-well plates with clear bottom and counted by microscope shortly after they attached. The data obtained from the individual assays was normalized to the seeding control. For the chemoresistance assay, the response of one condition to different drug concentrations was evaluated and, therefore, a seeding control was not necessary. For all assays, each condition was analyzed in three to six technical replicates for each biological replicate.

For phenotypic assays with the stable cell lines, the same cell numbers as for reseeding after transfection were used (listed in Table 8). All stable cell lines were induced with 2 µg/mL Doxycycline 72 hours before the assay to overexpress the respective pre-microRNAs. Moreover, Doxycycline was added to the medium during the assay to maintain pre-microRNA overexpression.

Table 8: Cell numbers seeded for different phenotypic assays.

cell line	6-wells (transfection)	cell count (96 h)	migration (RTCA)	attachment (RTCA)
MDA-MB-231	400,000 (coated)	2,000	100,000	10,000
HCC1806	200,000	1,500	-	-
SUM-159	70,000	400	50,000	5,000

Proliferation assay

For cell proliferation, the read-out was performed 96 hours after reseeding of the transfected cells. Proliferation was analyzed by counting the nuclei after 30 minutes of Hoechst 33342 staining (20 mM stock, 1:1000 diluted). The plates were scanned with the Molecular Devices Microscope IXM XLS using the 4x S Fluor objective and nuclei were counted with the Molecular Devices analysis software.

Chemoresistance assay

The chemoresistance assays were performed similar to the proliferation assays. Cells were reseeded after transfection and treated with different drug concentrations of Epirubicin and Paclitaxel after they attached. Epirubicin was dissolved in water to a stock concentration of 5 mM, whereas Paclitaxel was adjusted to 10 mM in DMSO. The stocks of both drugs were aliquoted and stored at -80°C. After 96 hours, the nuclei were stained with Hoechst and counted by microscope as described for the proliferation assay. Fitted drug response curves and IC50 concentrations were calculated with GraphPad Prism.

Transwell-based migration assay

For transwell-based migration, the xCELLigence Real-Time Cell Analyzer (RTCA) system was used. The RTCA contains golden electrodes that measure electrical impedance. The impedance increases upon attachment of cells to the plate surface containing the electrodes. E plates have the electrodes at the bottom of the well, whereas CIM plates contain a transwell and the electrodes are on a surface right below the transwell. For migration, 100 µL cell suspension in serum-free medium was seeded into the upper chamber of a CIM plate. The lower chamber was filled with 175 µL of full growth medium, which worked as a chemoattractant. After blanking, the device measures the increase in cell index (arbitrary unit for the impedance) every 15 minutes. For the CIM plates, this means that the impedance increases for every cell that migrates through the transwell pore. To make sure that the transfection of different microRNA mimics or siRNAs did not alter attachment strength of the cells compared to the non-targeting controls, I seeded the cells also in E plates (in 100 µL full growth medium) to evaluate whether all conditions attached to the well bottom with similar strength. The RTCA device was kept in a cell culture incubator (37°C, 5 % CO2, humidified) through-out the assay.

Mammosphere assay

To grow mammospheres, transfected cells or Doxycycline-induced stable cells were reseeded into ultra-low attachment plates. The cells were cultivated in serum-free growth medium that was supplemented with specific growth factors required for sphere formation. The medium composition and the seeded cell numbers can be obtained from Table 9.

Table 9: Cell numbers and medium composition for mammosphere assays.

cell line	cells/6-well	cells/24-well	sphere medium
MDA-MB-231	20,000/2 mL	250/0.5 mL	DMEM/F-12, 20 ng/mL EGF, 20 ng/mL bFGF, 1xB27
HCC1806	20,000/2 mL	250/0.5 mL	4 ug/mL Heparin, 2 µg/mL Doxy., 1 µg/mL Puromycin
SUM-159	20,000/2 mL	250/0.5 mL	see above + 1 µg/mL Hydrocortisone, 5 µg/mL Insulin

For the mammosphere screen, cells were seeded into 6-well ultra-low attachment plates. Prior to seeding the first sphere generation, the trypsinized cells were washed with PBS at 1,500 rpm for five minutes to remove remaining FBS. The cells were resuspended in sphere medium and counted by CASY. Moreover, the cell suspension was checked by microscope to ensure a single cell suspension. Each sphere generation was grown in ultra-low attachment plates for seven days, then the spheres were dissociated and reseeded for the next generation. For passaging spheres from one generation to the next, the spheres were collected by centrifugation at 1,200 rpm for five minutes. The spheres were resuspended in 500 µL trypsin to dissociate them to single cells. The trypsin was stopped with 500 µL FBS and cells were subsequently washed with PBS to remove the FBS again (centrifugation at 1,500 rpm for five minutes). After washing, the cells were resuspended in 1 mL mammosphere medium and the cell number was determined. Then, the same number of cells as for the previous generation was seeded for the next sphere generation.

For the validation of pre-microRNA-103a-1, sphere formation was evaluated in 24-well ultra-low attachment plates. The cells were seeded according as described for the mammosphere screen. The spheres were grown without Puromycin in the medium. Sphere size and number was quantified after seven days of sphere formation with the Molecular Devices ImageXpress Micro Confocal Microscope. Cell aggregates larger than 70 µM in diameter where counted as spheres.

***In vivo* experiments**

For *in vivo* experiments, stable MDA-MB-231 cell lines were expanded to a sufficient amount and grown without Puromycin in the medium prior injection into NSG mice. At the day of injection, cells in exponential growth were trypsinized and washed several times with 50 mL PBS at 1,500 rpm for five minutes. The cells were counted by CASY and adjusted to the required concentration in PBS. The mouse experiments including the cell injection were performed by the lab of Karin Müller-Decker. 3 million stable MDA-MB-231 cells were injected into the mammary fat pad of each NSG mouse (six mice per group). Doxycycline was administered via the drinking water (1 mg/mL Doxycycline, 5 % Saccharose), starting seven days post injection. The tumor size was determined on a regular basis, the animals were sacrificed when the tumor had reached a diameter of 1 cm (i.e. the ethical limit).

3.2.6 Bioinformatic and statistical analysis

TCGA data analysis

For analysis of patient data, the sequencing data that is available for the TCGA cohort was used since this data set allowed resolving microRNA expression by 5'isomiRs. The microRNA and mRNA sequencing data from TCGA was processed by Susanne Ibing, who obtained the data from the following website: https://portal.gdc.cancer.gov/projects/TCGA-BRCA. Briefly, the GDC microRNA analysis workflow was based on the British Columbia Genome Sciences Centre microRNA profiling pipeline[133]. Re-annotation of isomiR features was performed using an adaptation of miRBase version 22.1[134]. To discriminate only between microRNA sequence variants that differ at their respective 5'ends, all isomiRs with the same 5'end were summed regardless of their 3'end. Batch effects were corrected with the R package 'ComBat'.

The selection of pre-microRNAs for the library was based on differentially expressed 5'isomiRs in TCGA breast cancer patients. Initially, the differential expression analysis was performed by Subarna Palit. After Cindy Körner had discovered batch effects that severely biased the TCGA data, she ran Subarna Palit's code for the differential expression analysis on the batch-corrected data from Susanne Ibing.

Statistical analysis with GraphPad Prism

Data are presented as mean ± standard deviation unless indicated otherwise. All samples were analyzed with unpaired two-tailed Student's t-test using GraphPad Prism version 5.01. For matched tumor-normal patients from TCGA, a paired two-tailed Student's t-test was applied. Correlation coefficients were obtained based on Spearman's correlation and significant differences in survival were determined using log-rank test. P-values < 0.05, < 0.01 and < 0.001 are indicated with one, two and three asterisks, respectively.

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) is applied to identify gene signatures that are over- or under-represented in a large data set (indicated by a positive or negative enrichment score, respectively). GSEA for Basal breast cancer patients from the TCGA cohort was performed by Cindy Körner. As input, the batch-corrected microRNA or mRNA TCGA data was used. To identify cancer-related gene sets that are potentially modulated, the Hallmark gene collection was used[135]. Spearman correlation coefficients between the microRNA or mRNA of interest and the individual genes from the gene signature were calculated and used for ranking. Permutation of the data was performed by phenotype. The GSEA-P software was executed in the default mode[136,137].

3.2.7 Establishment of a customized NanoString assay

NanoString is a hybridization-based molecule count. Two highly specific probes are designed to bind the region of interest (e.g. mRNA or genomic DNA). One probe is coupled to a biotin tag, the other probe is labeled with a unique fluorescence tag. After probe hybridization, the target complex is immobilized, excess probe is removed by washing and the fluorescence tags are used for read-out of the samples. For this project, I developed a customized NanoString assay that targeted the region of interest on the genomic level similar to a copy number variation (CNV) assay. For my assay, the biotin-coupled probe was universal and targeted a part of the retroviral backbone used for generation of the stable cell lines. This allowed discriminating between the retrovirally integrated pre-microRNAs and the endogenous pre-microRNAs. Unique probes were designed for the individual pre-microRNAs of the library. All probes were designed by NanoString and are listed in 3.1.7.

Isolation of genomic DNA

For analysis by NanoString, the genomic DNA of my samples was isolated with the DNeasy Blood & Tissue kit. Frozen cell pellets of maximum 5 million cells each were thawed at room temperature and were then resuspended in 200 µL PBS. Subsequently, 20 µL proteinase K were added. Since NanoString recommends RNA-free genomic DNA, 4 µL RNase A (100 mg/mL) were added to the cell suspension. The mix was vortexed and incubated at room temperature for two minutes. After adding 200 µL AL buffer, the mix was vortexed again and incubated at 56°C for ten minutes. Subsequently, 200 µL ethanol (96-100 %) were added to the sample and vortexed thoroughly. The entire mix was pipetted into a column and spun at 8,000 rpm for 1 minute. The flow-through was discarded and 500 µL AW1 buffer were added to the column. The sample was centrifuged at 8,000 rpm for 1 minute and the flow-through was discarded. Then, 500 µL AW2 buffer were applied to the column and centrifuged at 13,000 rpm for three minutes. The collection tube containing the column was replaced by a fresh one and centrifuged at 13,000 rpm for one minute to remove residual ethanol. For elution, the column was placed into a clean 1.5 mL microcentrifuge tube and 100 µL AE buffer (prewarmed to 56°C) were added to the column. After five minutes of incubation, the column was spun at 8,000 rpm for one minute. 100 µL fresh, prewarmed AE buffer were added to the column and the centrifugation step was repeated after five minutes incubation. DNA concentration was analyzed by NanoDrop. All genomic DNA samples were of high quality as recommended by NanoString: the A260/280 and A260/230 ratios of the genomic DNA were in the range of 1.7-1.9 and 1.3-2.0, respectively.

PCR pre-amplification

For NanoString analysis, the genomic region of interest was pre-amplified with the PCR protocol below and yielded a fragment size of approximately 660 base pairs, depending on the length of the individual pre-microRNAs. The primer sequences are listed in 3.1.7. For evaluation of the PCR specificity, the PCR was also performed on genomic DNA from parental cell lines. Moreover, a water control was included to test for contaminated PCR reagents.

PCR reaction:	10 µL Titanium Taq PCR buffer (10x)
	2 µL 50x dNTPs (10 mM each)
	2 µL primer mix (10 µM each)
	4 µL DMSO

2 µL of Titanium Taq DNA polymerase (50x)
gDNA (400 ng)
up to 100 µL ddH2O

PCR protocol:	1x	94°C	3 min
	30x	94°C	30 sec
		68°C	30 sec
	1x	68°C	3 min

PCR product purification

The PCR products were purified with the Wizard SV Clean Up System. The entire PCR volume was combined with the same volume of membrane binding solution. The solution was applied to a column and spun at maximum speed for one minute. The column was washed with 700 µL of membrane wash solution. The wash was removed by spinning at maximum speed for one minute. The wash step was repeated with 500 µL membrane wash solution. To remove remaining liquid, the column was inserted into a fresh collection tube and was spun empty at maximum speed for one minute. The PCR product was eluted in 75 µL prewarmed (56°C) elution buffer from the DNeasy Blood & Tissue kit for ten minutes. This step was repeated with 75 µL fresh elution buffer.

To concentrate the PCR product, 450 µL DNA precipitation buffer were added to 150 µL purified PCR product. The solution was vortexed briefly and centrifuged at maximum speed for 15 minutes. The supernatant was decanted immediately and the pellet was washed once with 100 µL 80 % ethanol at maximum speed for 15 minutes. The supernatant was decanted again and the pellet was air-dried until the remaining ethanol evaporated. The pellet was resuspended in 30 µL prewarmed (56°C) elution buffer from the DNeasy Blood & Tissue kit.

DNA quantification

The concentration of the purified PCR products containing the region of interest was determined with a Qubit fluorometer according to the manufacturer's instructions. Qubit measurements were also performed for the unpurified PCR products obtained from the parental cell lines as well as the water control. This allowed to evaluate whether the PCR primers discriminated the retrovirally integrated pre-microRNA from the endogenous pre-microRNAs.

Agarose gel electrophoresis

The dilutions that were prepared for DNA quantification with the Qubit fluorometer were also used for agarose gel electrophoroesis. The agarose gel electrophoresis allowed visualizing whether the amplified region had the correct size and whether the primers were specific to the region of interest. 1 % agarose gels were prepared with TAE buffer and 20-80 ng of the PCR product were loaded on the gel. Images of the gel were taken after running the gel at 100 Volt for 35-40 minutes. The correct size of the PCR product was determined by comparing it to the DNA ladder (2 µL of 1kb PeqGold ladder were loaded). For the PCR products from the parental cell lines and from the water control, 6-7 µL of the undiluted PCR product were loaded.

Sample input titration for NanoString

For CNV assays, NanoString recommends to use 150-300 ng of genomic DNA or up to 600 ng for small CNVs. In the case of CNV assays, however, at least one copy of the molecule of interest is expected per cell. In the stable cell lines harboring the pre-microRNA library, approximately only every 72nd cell contained the same pre-microRNA integrate, at least before any selection. I determined the ideal input amount by titration and found 7 µg of genomic DNA to be ideal. Using 7 µg genomic DNA provided a strong signal and did not saturate the NanoString cartridge. Since I pre-amplified the region of interest, the input for the customized NanoString assay had to be adjusted considering the length of the PCR product. Hence, 1.5 pg of the purified PCR products were used (or 20 pg of the pre-microRNA plasmids for the evaluation of a potential PCR bias).

Pooling and dilution of the NanoString probes

Each probe A was specific for one pre-microRNA integrate and partially complementary to a fluorescently labeled reporter tag. Probe B was universal for all pre-microRNAs and hybridized to a biotinylated universal capture tag that allowed to anchor the targets to streptavidin-coated cartridges. Pooling of the probes was performed by Sigma-Aldrich, the manufacturer of the probe oligos. Probe pool A (each probe at 5 nM) and probe pool B (each probe at 25 nM) were stored in aliquots at -80°C. For the NanoString assay, working dilution of the pools were generated directly before they were used. The working dilutions consisted of 4 µL of the probe stock A or B and 29 µL TE-Tween. This resulted in a concentration of 0.6 nM per probe in pool A and 3 nM per probe in pool B.

Probe hybridization

For the hybridization, an aliquot of the probe stocks A and B was thawn and working dilution were prepared as described before. Moreover, an aliquot of the tag set was thawn and carefully mixed by flicking to avoid shearing of the tags. The hybridization buffer and the probe pool A were added to the tag set and the mix was inverted and spun down. Subsequently, the probe pool B was added. The mix was inverted and spun down, then 8 µL were added to each reaction tube of the tube stripe provided by NanoString. The dilutions of the PCR products were denatured at 95°C for five minutes and cooled down on ice for two minutes. 7 µL of each sample were added to 8 µL of the master mix. The final reaction mix was inverted again and spun down briefly. Hybridization was performed at 67°C for 16 hours. The thermal cycler lid was set 5°C above the block temperature to minimize evaporation.

Hybridization reaction:	5 µL hybridization buffer
	2 µL tag set
	0.5 µL working dilution of probe pool A (0.6 nM per probe)
	0.5 µL working dilution of probe pool B (3 nM per probe)
	7 µL of sample dilution (containing 1.5 pg purified PCR product)
	up to 15 µL ddH2O

Sample preparation and read-out with the NanoString nCounter FLEX Analysis System

After hybridization, excess probes were removed with a two-step purification protocol using magnetic beads. The washing and immobilization process as well as the data acquisition was done automatically by the nCounter FLEX Analysis System. The machine was operated by Martina Kirchner.

3.2.8 Analysis of NanoString data

The nSolver 4.0 software was used for analysis of the data generated with the nCounter FLEX Analysis System. Analysis was performed as suggested by the manufacturer with a few modifications taking the customized assay design into account. The analysis process is briefly described in the following.

Quality control

For each analyzed NanoString cartridge, the nCounter FLEX Analysis System provides four scores that allow evaluating the quality of the run:

1) The **imaging QC** describes how much of the scanned cartridge lane the instrument was able to scan. To obtain robust data this value should be at least 75% and was ≥ 98% for all my cartridges.
2) The **binding density score** indicates how much of the cartridge surface per lane was covered by probe signal and should range between 0.1 and 2.25 spots/micron². For my experiments, the value ranged from 0.2 to 0.37.
3) The **positive control linearity QC** represents the correlation between the concentrations and the absolute counts for the positive control that is spiked into the tag set. A correlation of less than 0.95 indicates an issue with the hybridization reaction or assay performance. All my cartridges revealed a correlation of at least 0.95.
4) The **positive control limit of detection QC** shows whether the absolute counts for the positive control with the lowest concentration are above the background. Since this would be the parental cell lines in my system, this was not applicable due to the PCR pre-amplification.

Data normalization and background subtraction

After the parameters for the quality control are checked, the absolute counts from the parental cell lines are normally averaged and subtracted as background from the other samples. The pre-amplification by PCR made this redundant since there was no background resulting from other regions than the target region. Moreover, the data is usually normalized to the invariant region controls (INVs) that target invariant regions of the genome. This allows to account for differences in sample input. Due to the PCR-based pre-amplification of my assay, this normalization step was not possible.

Relative abundance

For each sample, NanoString detected absolute counts for the individual pre-microRNA targets. In order to compare between samples, relative counts were calculated (absolute counts of a particular pre-microRNA divided by the absolute counts of all pre-microRNAs in the respective sample).

4. RESULTS

4.1 microRNAs and 5'isomiRs in chemoresistance of triple-negative breast cancer

Breast cancer is one of the most common cancer entities in women worldwide[138] and chemotherapy is the standard of care for patients with the aggressive subtype TNBC[28]. A substantial fraction of TNBC patients responds only partially to the treatment and, consequently, has a high rate of recurrence[29]. Understanding the molecular mechanism and exploiting it for the development of biomarkers and targeted therapy is, therefore, highly relevant for this patient population. microRNAs modulate the majority of cellular signaling pathways[139] and play a key role in chemoresistance, especially in TNBC[37,140]. The multitude of microRNAs and the increase in complexity by functionally relevant sequence variants, however, requires more research on this topic. Many microRNAs are still poorly characterized and the majority of studies does not discriminate between effects from 5'isomiRs when gathering sequencing data. Along this line, the aim of this study was to identify and characterize microRNAs and 5'isomiRs that exert a key role in promoting or preventing chemoresistance in TNBC cell lines.

4.1.1 Establishment of an in vitro system to study microRNAs in chemoresistance

Selection of microRNAs with a potential role in chemoresistance

To identify microRNAs that play a key role in chemoresistance of TNBC, I established a functional screening for microRNAs and 5'isomiRs modulating chemoresistance of TNBC cell lines. Specifically, I hypothesized that microRNAs and 5'isomiRs that were differentially expressed in tumor compared to normal breast tissue of patients from the TCGA cohort are functionally connected to tumorigenesis and, therefore, might play a role in chemoresistance and patient outcome. Due to the limited number of samples from normal breast tissue in the TCGA cohort, the selection of differentially expressed 5'isomiRs was performed using data from patients of all subtypes (n=100). The annotation of 5'isomiRs in the TCGA data was performed by Subarna Palit, differential expression analysis was done by Subarna Palit and Cindy Körner. Based on these data, I selected microRNAs for further analysis based on two criteria. First, 5'isomiRs had to be differentially expressed in tumor compared to normal samples from the same patients (adjusted p value ≤ 0.05, log2 fold change ≥ 1). Second, only microRNAs with a mean expression of at least 10 reads per million (rpm) across tumor and normal samples were selected. These criteria

prioritized the most deregulated microRNAs in breast cancer patients and these microRNAs were selected for the microRNA library.

Design and cloning of a pooled pre-microRNA library

Selection of microRNAs and their 5'isomiRs differentially expressed in tumor tissue compared to normal tissue yielded 86 microRNAs that might have tumor-promoting or tumor-suppressive features, for instance, in the context of chemotherapy. Since several of the selected microRNAs were 5'sequence variants or derived from the other arm of the same pre-microRNA, the library complexity was reduced to 63 pre-microRNAs. The selected pre-microRNAs were cloned into a Doxycycline-inducible, retroviral vector RT3GEPIR that was kindly provided by Fellmann et al[130]. This vector system allowed embedding the pre-microRNAs of interest into a pri-microRNA-30 backbone which resembled the endogenous microRNA context.

All pre-microRNAs were cloned using two or three overlapping oligos (Figure 5A) to create the respective pre-microRNA sequence obtained from miRbase. Sanger sequencing of all plasmids was performed to confirm that the correct pre-microRNA sequence was inserted into the RT3GEPIR vector and did not harbor any mismatches. The pooled microRNA library was generated by equimolar pooling of all 63 pre-microRNAs in RT3GEPIR, the unmodified RT3GEPIR vector, and eight non-conserved *C. elegans* pre-microRNAs in the RT3GEPIR backbone. Retroviral integration of the pre-microRNA library into the genome allowed biogenesis of mature microRNAs and their sequence variants using the endogenous microRNA processing machinery in the cells.

Moreover, the pre-microRNAs served as barcodes to detect the enrichment or depletion of pre-microRNA integrates from the cell pool. Mature microRNAs mediating the effect on chemoresistance could be validated by qRT-PCR-based approaches. One possibility to validate the individual microRNAs is to determine expression of all 5'isomiRs from the 3p arm or from the 5p arm with miScript assays from Qiagen. The second option is to optimize 5'isomiR-specific detection for the candidate microRNAs by using isoform-specific adapters and probes for the detection as described for Dumbbell-PCR[98,99].

Establishment of triple-negative breast cancer cell lines harboring the pre-microRNA library

Since the aim of this project was to identify microRNAs with a role in chemoresistance of TNBC, I chose TNBC cell lines that are frequently used for studies in TNBC and belong to the BL2 subtype, which has the lowest pCR rate[24]. Based on these criteria the cell lines MDA-MB-231, HCC1806 and

SUM-159 were selected. In order to generate stable TNBC cell lines harboring the microRNA library, a retroviral system was used for transduction (Figure 5A). To ensure that each cell received and integrated maximum one pre-microRNA construct, I used a MOI of 1-3%. For each cell line, three transduction replicates (#A, #B and #C) were generated by preparing independent master mixes and transducing independent flasks of the same cell line. The cell lines were named accordingly: replicate #A of MDA-MB-231, HCC1806 and SUM-159 received the same transduction mix.

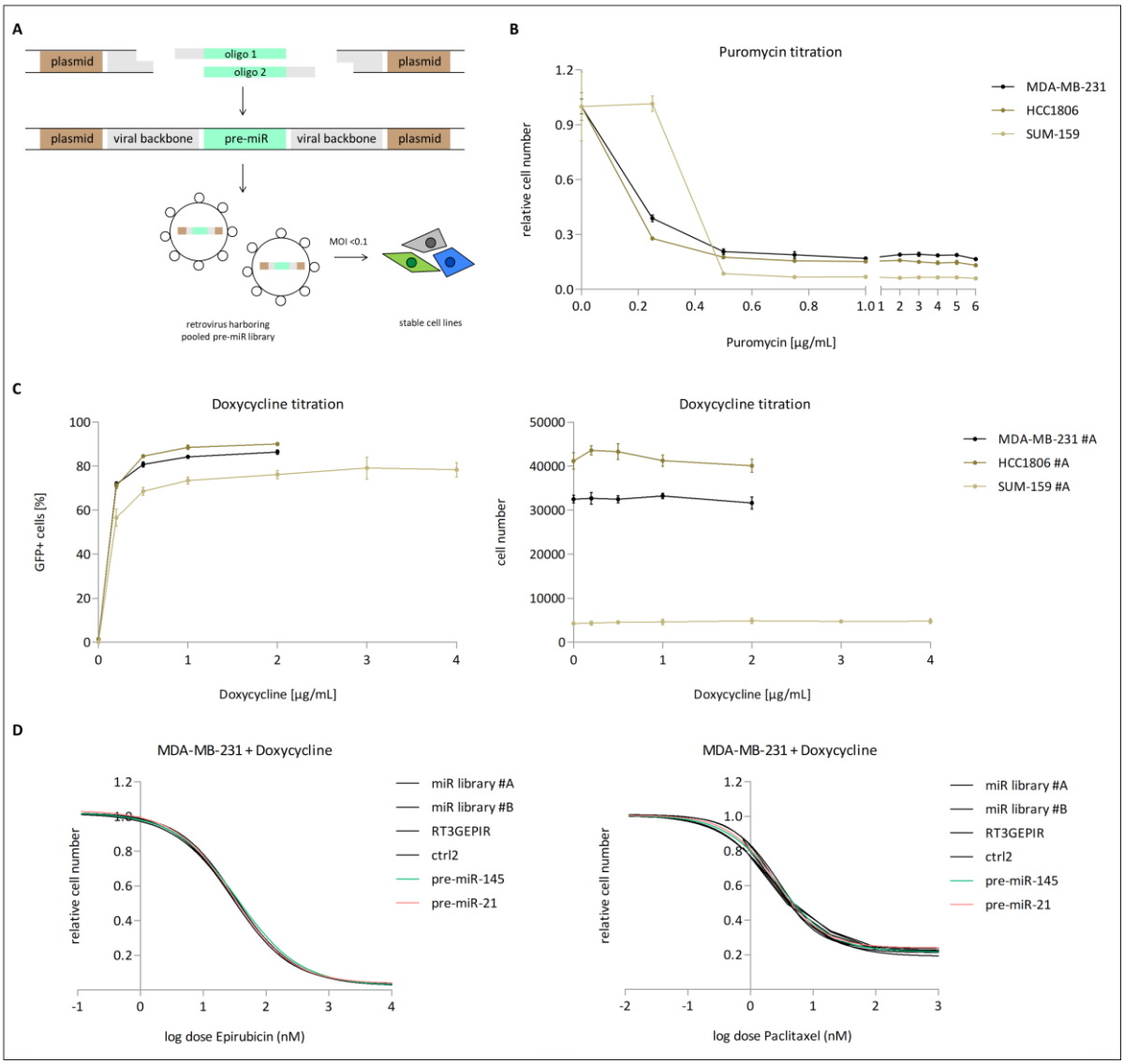


Figure 5: Establishment of an *in vitro* system to study microRNAs in chemoresistance. A) The scheme illustrates how the microRNA library was generated. **B)** Wildtype MDA-MB-231, HCC1806 and SUM-159 cells were treated with different concentrations of Puromycin for 72h to identify the ideal concentration to select for cells successfully transduced with the microRNA library. The cell number was counted by microscope after Hoechst staining. **C)** Stable MDA-MB-231, HCC1806 and SUM-159 cells harboring the microRNA library were treated with different Doxycycline concentrations for 72h to induce expression of GFP and the pre-microRNAs. The total cell number was determined by microscope after Hoechst staining. GFP-positive cells were counted and normalized to the total cell number to select a concentration suitable for induction of the

microRNA library. The total cell number was plotted to evaluate potentially cytotoxic effects of different Doxycycline concentrations on the cells. **D)** The stable MDA-MB-231 cell lines harboring different constructs were induced with Doxycycline for 72h and treated with Epirubicin or Paclitaxel in the presence of Doxycycline for another 72h. Each condition is displayed as mean of six technical replicates +/- standard deviation (B-C) or as fitted curve based on the mean of six technical replicates (D).

Cells with an integrated pre-microRNA construct were selected with Puromycin. The ideal concentration of Puromycin to deplete wildtype cells from the cell pool was determined with a titration curve (Figure 5B). 0.5 - 1.0 µg/mL Puromycin was the minimum effective concentration range that killed the majority of wildtype cells, higher concentrations did not increase potency. In conclusion, 1 µg/mL Puromycin was chosen for selection of all stable TNBC cell lines. After selection with Puromycin, the stable cell lines were induced with different Doxycycline concentrations to find the optimal concentration to induce pre-microRNA expression. Doxycycline induction of the pre-microRNA integrate induced also GFP expression, which was used as proxy to evaluate the fraction of cells in which expression was induced at the respective concentrations (Figure 5C). 2 µg/mL Doxycycline were chosen for induction of all three TNBC cell lines since this concentration yielded 80-90% GFP-positive cells for MDA-MB-231 and HCC1806. For SUM-159, only 70-80% GFP-positive cells were counted after induction with 2 µg/mL Doxycycline. Inducing SUM-159 cells with higher Doxycycline concentrations, however, failed to increase the number of GFP-positive cells. To exclude potential toxic effects of Doxycycline, the total cell number was evaluated upon incubation with the different concentrations of Doxycycline. No effects on cell proliferation were observed at the tested Doxycycline concentrations (Figure 5C). In conclusion, the pre-microRNA library was induced with 2 µg/mL Doxycycline since this concentration resulted in the highest percentage of GFP+ cells for all three cell lines without having cytotoxic effects.

Determining a suitable drug concentration to study microRNAs in chemoresistance

To study the role of the selected microRNAs in chemoresistance, I chose the taxane Paclitaxel and the anthracycline Epirubicin to challenge the TNBC cells harboring the microRNA library. Both of these drugs belong to the most commonly administered neoadjuvant regimen in TNBC[141] and target cytokinesis (taxanes) or DNA synthesis and integrity (anthracyclines)[34]. I hypothesized that drug response curves for a negative and a positive control would provide a concentration range suitable for studying the enrichment or depletion of those integrated pre-microRNAs with a role in chemoresistance. For this purpose, I selected pre-miR-21 and pre-miR-145 from the library. miR-21 was previously shown to promote chemoresistance[142,143], whereas miR-145 sensitized towards

chemotherapeutic drugs [144,145]. MDA-MB-231 cells stably overexpressing pre-miR-21 or pre-miR-145 were compared to two replicates containing the entire microRNA library (#A, #B), the empty vector RT3GEPIR or the non-targeting *C. elegans* pre-microRNA control 2 (Figure 5D). Two transduction replicates of the pre-microRNA library were included to evaluate potential differences in drug response resulting from a strong variation in the pre-microRNA representation. Strong differences in the pre-microRNA composition after transduction might reflect in a different drug response. However, there was no difference in resistance between any of the tested conditions.

Establishment of an alternative *in vitro* system to study microRNAs in chemoresistance

Since the tested cell lines, including the positive and the negative control, showed no difference in their response to Epirubicin and Paclitaxel, I chose an alternative approach to identify pre-microRNAs influencing chemoresistance. It is known that stem cells play an important role in chemoresistance and therapy failure: not only are stem cells often resistant to chemotherapy, the treatment also increases stemness features in other cancer cells[146]. In Basal breast cancer, the percentage of stem cells is particularly high compared to other subtypes, especially in tumors resistant to chemotherapy[147]. Therefore, I hypothesized that a mammosphere assay should enrich for chemoresistant breast cancer stem cells (BCSCs) since stem cells are driving chemoresistance. To enrich for potentially chemoresistant cells, I grew mammospheres from the TNBC cell lines harboring the pre-microRNA library. Subsequently, the mammospheres were profiled to identify pre-microRNAs with a role in chemoresistance.

The setup for the mammosphere assay consisted of nine cell lines including the three transduction replicates #A, #B and #C of the TNBC cell lines MDA-MB-231, HCC1806 and SUM-159. These cell lines received Doxycycline to induce the pre-microRNA library and were grown as mammospheres for seven days. Ultra-low attachment plates in combination with media containing specific growth factors allowed sphere formation (Figure 6A). After seven days, the spheres were dissociated with Trypsin and a defined number of cells was reseeded into ultra-low attachment plates for the next sphere generation, the rest was harvested for analysis. Passaging of the spheres was repeated several times, which allowed profiling the enrichment or depletion of pre-microRNAs over multiple sphere generations. I assumed that tracing enrichment over multiple passages should identify pre-microRNAs being associated with a very strong stemness phenotype. To account for pre-microRNAs that have a strong effect on proliferation or act cytotoxic upon overexpression, all Doxycycline-induced cell lines were also grown in 2D under normal growth conditions. Strong

enrichment in 3D and 2D, for instance, would be indicative of a proliferation phenotype rather than a stemness phenotype.

Enrichment of chemoresistant breast cancer stem cells with a mammosphere assay

Based on the involvement of BCSCs in chemoresistance, I expected that mammospheres enrich for cells that are resistant to chemotherapeutic drugs. As a consequence, growing mammospheres from the TNBC cell lines harboring the pre-microRNA library would allow identifying pre-microRNAs with a role in chemoresistance. I hypothesized that the enrichment of potentially chemoresistant stem cells in spheres reflects in enrichment of BCSC and resistance markers. To test this hypothesis, the mRNA level of several BCSC and chemoresistance markers was measured in time-matched 3D and 2D samples (t6) by TaqMan (Figure 6B). While five sphere generations (t1-t5) and their time-matched 2D samples were cultivated to profile pre-microRNA enrichment by NanoString, one more time point was collected only for expression analysis (t6). I evaluated the most commonly used markers to identify BCSCs: *CD44*, *CD24*, *ALDH1A*, *PROM1*, *ITGA6*, *ITGB3* and *NANOG*[148,149]. Tested markers for chemoresistance included *ABCC1*, *ABCC2*, *SOD2*, *CAT* and *GPX1*. *ABCC1* and *ABCC2* were analyzed since they are considered to encode the main ABC transporters involved in multidrug resistance development in breast cancer[150,151]. Besides ABC transporters, detoxification enzymes are involved in stemness and chemoresistance. BCSCs upregulate SOD2, CAT and GPX1, for instance, to scavenge reactive oxygen species (ROS) more efficiently[152]. High levels of SOD2, moreover, were associated with increased chemoresistance in breast cancer cell lines[153].

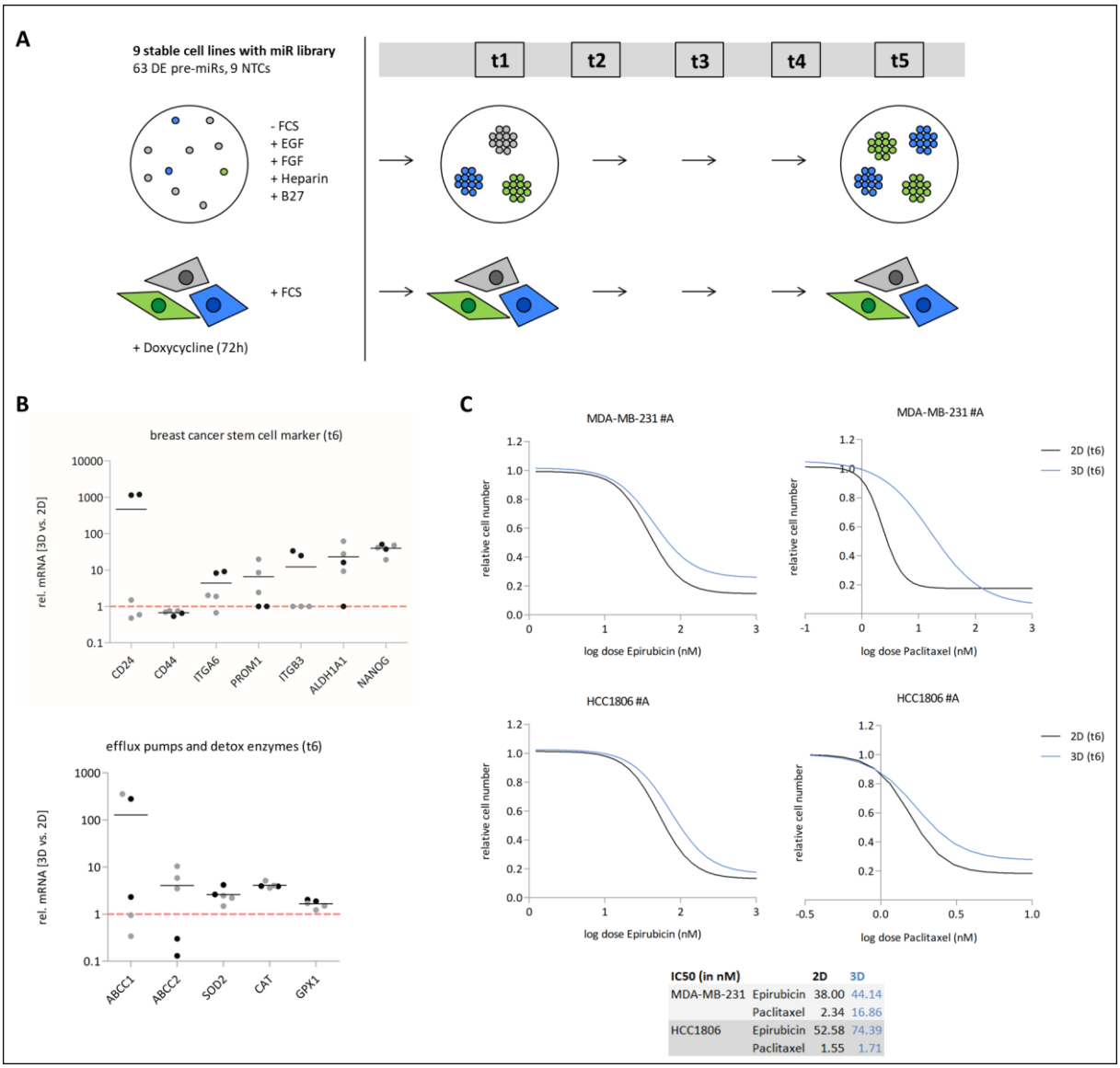


Figure 6: Mammosphere assay design and proof of principle. A) Three different transduction replicates (#A, #B and #C) of MDA-MB-231, HCC1806 and SUM-159 were grown in 3D or 2D after the pre-microRNA library had been induced with Doxycycline. Mammospheres were grown in ultra-low attachment plates and received serum-free medium supplemented with EGF, FGF, Heparin and B27, while cells in 2D received regular growth medium with serum. Spheres were grown for seven days (t1), dissolved with Trypsin and reseeded for the next sphere generation (t2-t5). Excess cells were harvested for analysis. Cells in 2D were passaged twice per week and samples for analysis were taken the same day as the sphere samples. **B)** Various breast cancer stem cells and chemoresistance markers were analyzed by TaqMan. To evaluate the changes in relevant markers, mRNA levels in 3D (t6) were normalized to time-matched 2D samples (t6). Each dot represents one biological replicate derived from MDA-MB-231 #B or #C (black dots), HCC1806 #A, #B or #C (grey dots). For MDA-MB-231 #A the spheres did not yield enough material for analysis. The mean fold change (3D/2D) for the five analyzed cell lines is indicated by a horizontal line. **C)** Resistance towards Paclitaxel and Epirubicin of MDA-MB-231 #A and HCC1806 #A cells grown in 3D or 2D (both t6) was evaluated after 72h of drug and Doxycycline exposure in 2D conditions. The cell number was determined by microscope after Hoechst staining. Drug response curves are displayed as fitted curve based on the mean of six technical replicates from one biological replicate.

Comparing the mRNA levels of the tested markers in 3D and 2D, showed similar fold changes across all tested replicates for some BCSC markers (*CD44* and *NANOG*), while others were regulated more heterogeneously (*CD24*, *ITGA6*, *PROM1*, *ITGB3* and *ALDH1A1*). In some cell line replicates the tested markers were not enriched at all in mammospheres. Most striking was *CD44*, where mRNA levels were even slightly decreased in 3D compared to 2D for all of the tested samples. *NANOG* was the only BCSC marker showing a strong and consistent increase in 3D of all tested cell lines (19-fold to 51-fold). The main drug efflux pumps in breast cancer, *ABCC1* and *ABCC2*, were strongly upregulated in 3D conditions of some replicates, but decreased compared to 2D in other replicates. For the three tested detoxification enzymes *SOD2*, *CAT* and *GPX1*, the mRNA levels were upregulated 1.5 - 5-fold in 3D for all tested cell line replicates. Overall, expression analysis showed that there was a rather a cell line-specific enrichment of certain markers in 3D conditions effects (*CD24*, *ITGA6*, *PROM1*, *ITGB3*, *ABCC2*), however, this did not apply to all tested markers (*ALDH1A1*, *ABCC1*).

To test whether the enrichment of markers associated with resistance would correspond to increased drug tolerance, I treated MDA-MB-231 #A and HCC1806 #A with different Paclitaxel and Epirubicin concentrations and generated drug response curves from the obtained data (Figure 6C). MDA-MB-231 #A cells from mammospheres indicated an increase in resistance to Epirubicin and a strong increase in resistance to Paclitaxel compared to cells cultivated in 2D. The changes in resistance of HCC1806 #A cells derived from mammospheres compared to 2D were stronger for Epirubicin, whereas the increase in resistance to Paclitaxel was minor. In summary, the performed experiments showed that the mammosphere assay indeed enriched for chemoresistant BCSCs as characterized by different stem cell markers.

4.1.2 Establishment of a NanoString assay to detect microRNAs modulating chemoresistance

Mammospheres grown from the TNBC cell lines harboring the pre-microRNA library had a higher fraction of stem cells and were more resistant to chemotherapeutic drugs than cells grown in 2D. Subsequently, I profiled the library composition in the spheres since enrichment or depletion of pre-microRNAs would indicate a potential role of the respective pre-microRNA in stemness and chemoresistance. For profiling, I chose NanoString based on the high sensitivity, multiplexing capability and simple analysis of this assay. This assay combines the benefits of qRT-PCR-based approaches and MicroArrays / sequencing and is applied in a variety of settings for research or diagnostic purposes. NanoString is a hybridization-based molecule count that uses two specific probes binding next to each on the target molecule of interest (Figure 7A). One probe is coupled

to a biotinylated universal tag, the other probe is uniquely labeled with a fluorescent tag. After hybridization of the probes, the biotinylated tag is used to purify, immobilize and analyze the complex of interest. The read-out is performed by scanning the unique reporter tags and calculating the read counts for each target. For this project, Prof. Dr. Peter Sinn (Institute of Pathology, University Clinics Heidelberg) kindly allowed me to use the NanoString nCounter FLEX Analysis System in his lab. The machine purifies the samples after the probe hybridization and acquires the data. The nCounter FLEX Analysis System was operated by Martina Kirchner.

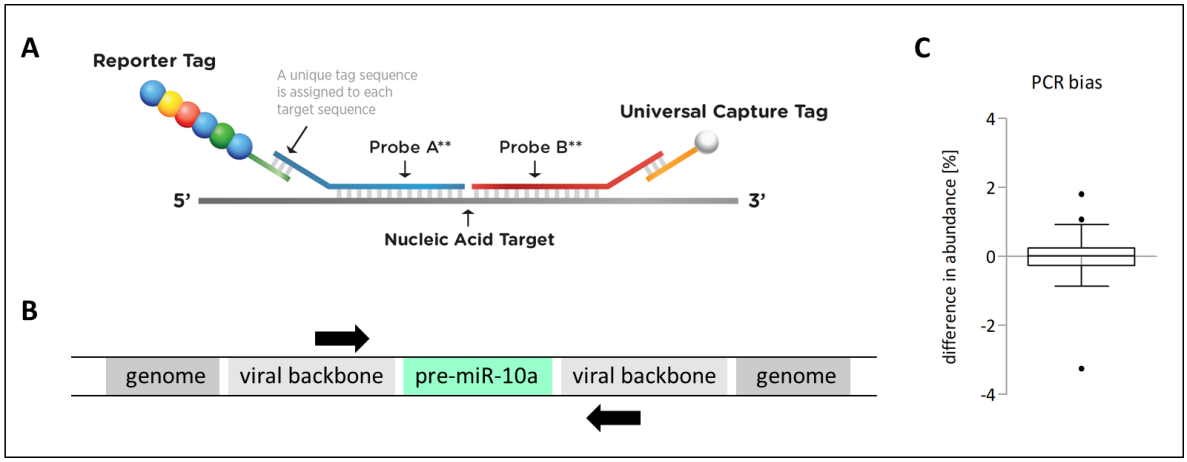


Figure 7: Establishment of a NanoString assay to evaluate the pre-microRNA library composition in spheres. A) The principle of NanoString assays is based on two specific probes hybridizing to the molecule of interest. One probe is linked to a biotinylated universal tag, the other one has a unique fluorescence tag. After hybridization, the complex is purified, immobilized and analyzed with the help of the biotin tag, which is used to pull-down the molecules of interest. The image was taken from www.nanostring.com. **B)** To determine which pre-microRNAs of the library were enriched in mammospheres, genomic DNA was isolated and the region of interest containing the pre-microRNA 'barcode' was pre-amplified with a PCR. The primer position for amplification of the pre-microRNA integrates is indicated by arrows. **C)** Evaluation of a potential PCR bias caused by pre-amplification of the pre-microRNA sequences was evaluated by NanoString. The pre-microRNA plasmid pool (n=72) used for generation of the pre-microRNA library was compared before and after PCR amplification. Differences in relative abundance of the individual pre-microRNAs are plotted.

The target in my setup was genomic DNA which harbored one individual pre-microRNA integrate per cell. To allow efficient probe binding, the genomic DNA needs to be fragmented with *Alu*I which yields an average fragment size of 250 base pairs. In my case, the probes had to target the specific pre-microRNA as well as a part of the retroviral plasmid backbone in order to discriminate the pre-microRNA library from the respective endogenous pre-microRNAs. *Alu*I fragmentation was not possible in this case since approximately one third of the pre-microRNA targets would have been cleaved by *Alu*I. To solve this problem, I established a PCR to pre-amplify the region of interest (Figure 7B). The PCR generated a product with an average size of 660 base pairs, which

could be used directly for the NanoString assay without pre-processing by *Alu*I. As another positive effect of the PCR step, the NanoString assay required less input material due to the pre-amplified DNA. This was an advantage since a comparably low amount of genomic DNA was isolated from the mammospheres.

To make sure that PCR amplification did not create a bias in the abundance of targets, I compared the pre-microRNA distribution in the plasmid pool (average size of 9 kb) containing all constructs for the pre-microRNA library before and after amplification using the customized NanoString assay (Figure 7C). Plasmids are much smaller than genomic DNA and have a less complex secondary structure, therefore, accessibility of the target by the probe is much better and *Alu*I digest is not necessary. Only for three pre-microRNAs the relative abundance before and after PCR amplification differed by more than 1 %. For pre-miR-551b, the relative counts were decreased by 3.25 % after PCR, while PCR overrepresented pre-miR-200a and pre-miR-125b-2 with a difference of 1.08 % and 1.81 %, respectively. In conclusion, NanoString detected no major PCR bias from pre-amplification of the region of interest.

4.1.3 NanoString assay detects strong enrichment of pre-miR-103a-1 in 3D

NanoString reveals heterogeneous enrichment of some pre-microRNAs across 3D samples

Subsequently, the established NanoString assay was used to determine the relative abundance of the pre-microRNAs from the library in selected sphere samples. Since the mammospheres were cultivated for five generations (t5), I decided to investigate pre-microRNA abundance in 3D (t5) and to compare this with data from time-matched 2D samples (t5). The tumor-suppressive candidates would most likely drop out quite early, especially if some oncogenic pre-microRNAs provided a growth advantage in 3D conditions. In order to evaluate enrichment in 3D in a time-resolved manner, 3D samples from the earliest time point available (t3) were profiled as well. Moreover, to distinguish enrichment based on stem cell features from enrichment due to a proliferation advantage, 2D samples at t1 were analyzed as well. Enrichment in 3D and 2D at t5 compared to 2D at t1 would suggest a proliferation phenotype.

Analysis of the selected samples showed heterogeneous enrichment of pre-microRNAs between different cell lines but also between replicates (Figure 8A). I hypothesized that highly oncogenic hits would outcompete other pre-microRNAs and show strong enrichment over multiple sphere generations. The majority of strongly enriched pre-microRNAs showed the high relative abundance only in one replicate. One striking example was pre-miR-93 (Figure 8B): the majority

of cells harvested from MDA-MB-231 #B mammospheres harbored a pre-miR-93 integrate (62 % after three sphere generations and 76 % after five generations). The 2D samples excluded the possibility that the enrichment was the result of a strong proliferation phenotype (difference of 5 % in t1 and 7 % in t5 compared to 62 - 76 % in 3D). None of the other sphere samples, however, showed even a slight enrichment of pre-miR-93, suggesting that pre-miR-93 conferred a strong potentially stemness-related selection advantage in 3D only for MDA-MB-231 #B cells.

NanoString finds pre-miR-103a-1 highly enriched in several mammosphere samples

Analyzing the pre-microRNA composition of the mammospheres with NanoString revealed heterogeneous enrichment across cell lines and replicates for most pre-microRNAs. For pre-miR-103a-1, however, the relative abundance in 3D was increased in all three cell lines tested (highlighted by boxes in Figure 8A and B): MDA-MB-231 #C, HCC1806 #B and SUM-159 #B. In all three cases, the 2D samples from t5 did not show a strong enrichment of pre-miR-103a-1 abundance and thereby suggested that the enrichment was not the result of a general growth advantage, but rather the result of a 3D-specific selection advantage (e.g. an increase in stemness). In MDA-MB-231 #C, pre-miR-103a-1 was more dominant after three sphere generations (67 %) than after five generations (46 %). Most likely, pre-miR-191 had a selection advantage in this replicate since this barcode enriched from 14 % relative abundance in t3 to 45 % in t5 and, thus, likely outcompeted pre-miR-103a-1. In HCC1806 #B, the relative abundance of pre-miR-103a-1 increased from 5 % in 3D (t3) to 16 % in 3D (t5). In mammospheres derived from SUM-159 #B, pre-miR-103a-1 representation was 8 % in t3 and 23 % in t5. Pre-miR-103a was selected as hit for further analysis based on the strong enrichment in spheres generated from three different transduction replicates that originated from three different parental cell lines, which suggests a stemness and chemoresistance promoting role.

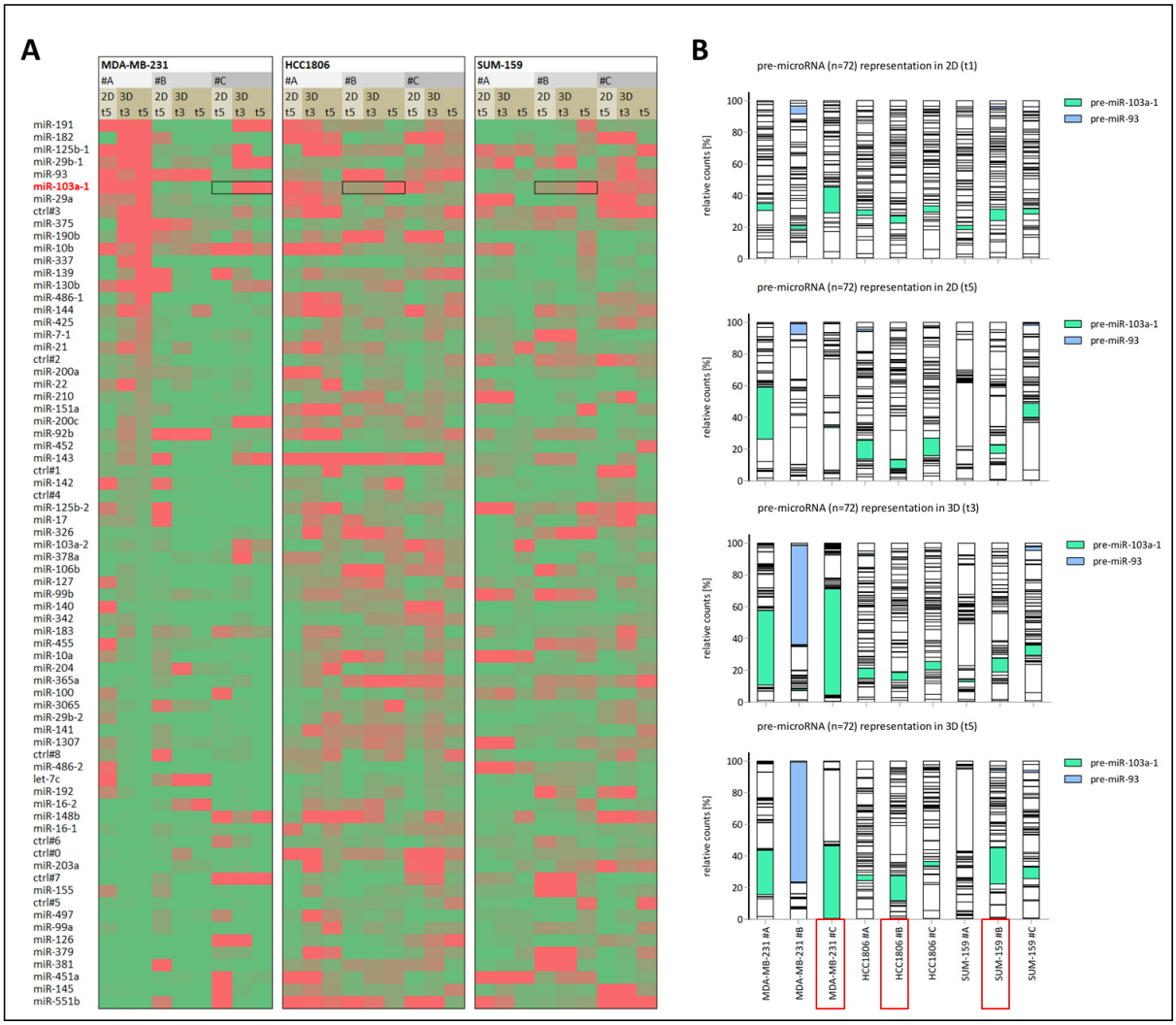


Figure 8: NanoString-based detection of pre-microRNA enrichment in mammospheres. A) Heatmap of NanoString results illustrating fold changes for the relative abundance of pre-microRNAs in 2D (t5) or 3D (t3, t5) compared to 2D (t1). Red indicates enrichment, green shows depletion. **B)** Relative abundance of the analyzed 72 pre-microRNAs in samples from nine cell lines harvested at different time points and in different conditions with pre-miR-103a-1 and pre-miR-93 highlighted in green and blue, respectively.

Evaluation of pre-miR-103a-1 in survival, stemness and chemoresistance

The pre-microRNAs for the library were selected based on differentially expressed microRNAs and 5'isomiRs in TCGA patients. I hypothesized that enrichment of a pre-microRNA in a mammosphere assay selecting for chemoresistant BCSCs should correspond with an oncogenic role of the microRNAs expressed from the respective pre-microRNA in patients. To evaluate this, expression analysis in tumor tissue compared to normal tissue was performed for those 5'isomiRs of pre-miR-103a-1 that were considered expressed in patients (mean >15 rpm). All expressed isoforms of pre-miR-103a-1 showed significantly higher expression in tumor tissue compared to normal tissue

(Figure 9A), which was in line with the putative oncogenic role of pre-miR-103a-1 that I observed in the mammosphere-based screen. Furthermore, survival analysis showed that high levels of all miR-103a-3p 5'isomiRs were associated with worse prognosis in patients (Figure 9A).

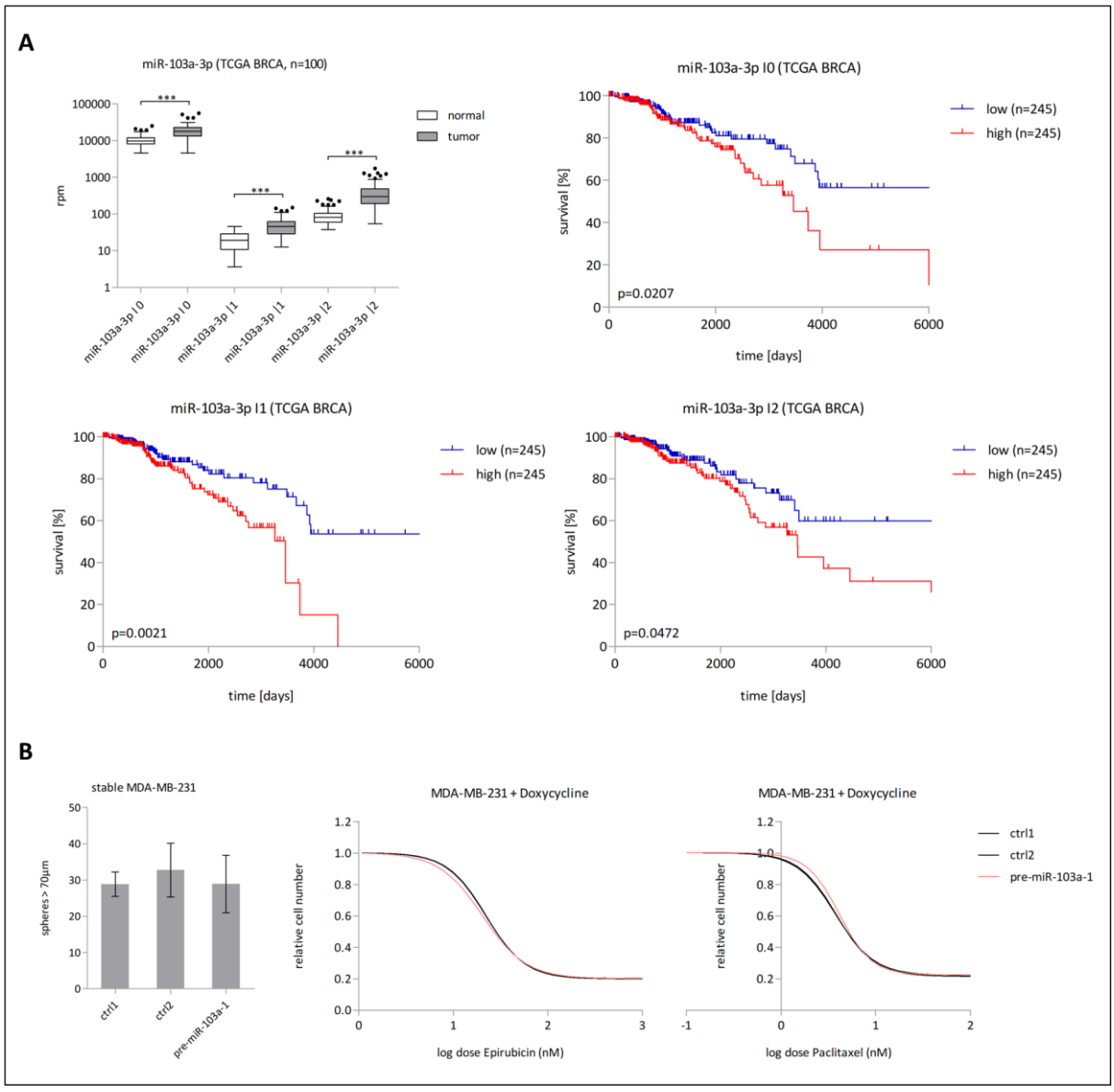


Figure 9: Validation of pre-miR-103a-1. A) Expression of miR-103a-3p 5'isomiRs in matched tumor and normal-like samples from breast cancer patients (TCGA). Statistical comparison was based on paired, two-tailed Student's t-tests. P-values are represented by asterisks (* ≤ 0.05, ** ≤ 0.01 and *** ≤ 0.001). For survival analysis, breast cancer patients (Basal, HER2, LumB and LumA) from TCGA were grouped into low (lower quartile) and high microRNA expression (upper quartile). Statistical significance of survival differences was determined using log-rank test. P-values are represented by asterisks (* ≤ 0.05, ** ≤ 0.01 and *** ≤ 0.001). **B)** Stable MDA-MB-231 cells were generated for functional validation of pre-miR-103a-1 and overexpression was induced with Doxycycline. Spheres >70 µM were counted by microscope after seven days of sphere formation. Error bars show the standard deviation of three biological replicates. Statistical comparison was based on unpaired, two-tailed Student's t-tests. Moreover, resistance towards Paclitaxel and Epirubicin was evaluated after 72h. The cell number was determined by microscope after Hoechst staining. Drug response curves are displayed as fitted curve based on the mean of six technical replicates.

In the next step, I wanted to validate the phenotype of pre-miR-103a-1. For this purpose, a stable MDA-MB-231 cell line for inducible overexpression of pre-miR-103a-1 was generated. I hypothesized that overexpression of pre-miR-103a-1 would promote sphere formation and chemoresistance since cells with the pre-miR-103a-1 integrate were strongly enriched in several 3D samples. To test this, sphere formation was assessed after seven days. Surprisingly, no significant differences in sphere numbers (Figure 9B) or size (data not shown) were detected in comparison to the negative controls. Even challenging the cells with Paclitaxel or Epirubicin (data not shown) during sphere formation or drug exposure in 2D (Figure 9B) did not show any differences for cells overexpressing pre-miR-103a-1.

Sanger sequencing-based validation of NanoString results

Despite the strong enrichment of pre-miR-103a in various mammosphere replicates, there was no effect of pre-miR-103a overexpression on sphere formation or chemoresistance of MDA-MB-231 cells. This prompted the question, whether the enrichment detected by the custom NanoString assay was reliable. To exclude technical error, I established a SYBR Green-based qRT-PCR approach to validate strongly enriched pre-microRNAs. For this purpose, I designed specific qRT-PCR primers for each construct of interest. To make sure that the primers did not detect the endogenous pre-microRNAs, they were designed to span the respective pre-microRNA sequences as well as the backbone of the viral vector. To ensure specificity of the primers, they were tested in the parental cells as well. For none of the tested pre-microRNAs, however, the primers were specific enough. The qRT-PCR always detected the endogenous construct as well (Figure 10A). Since amplification of the endogenous pre-microRNA would compete for primers with the region of interest, this would affect reliability of the quantification.

Another option to validate the NanoString results was to perform Sanger sequencing for samples in which one pre-microRNA was strongly enriched. Since the strongest enrichment was observed for pre-miR-93 in mammospheres formed by MDA-MB-231 #B (62 % in t3 and 76 % in t5), those samples were sent for Sanger sequencing. For comparison, the 2D samples in which pre-miR-93 showed rather low abundance (5 % in t1 and 7 % in t5) were also sent for sequencing. As expected, sequencing of the 3D samples resulted in the sequence of the viral pre-miR-93 integrate, whereas sequencing of the 2D samples failed starting from the position where the common sequence of the vector backbone ended (Figure 10B). In conclusion, Sanger sequencing confirmed strong enrichment of pre-miR-93 in the respective mammosphere samples and thereby showed that the NanoString assay had performed reliably in this case. As a consequence, the strong enrichment of

pre-miR-103a-1 in selected mammospheres was likely detected accurately by NanoString as well. Excluding a technical problem of the detection system, this implies that pre-miR-103a-1 was enriched, but had no effect on stemness and chemoresistance. Based on the assumption that the enrichment of pre-miR-103a-1 as well as other pre-microRNAs in 3D was not connected to stemness- and resistance-promoting features, I decided to stop working on this project.

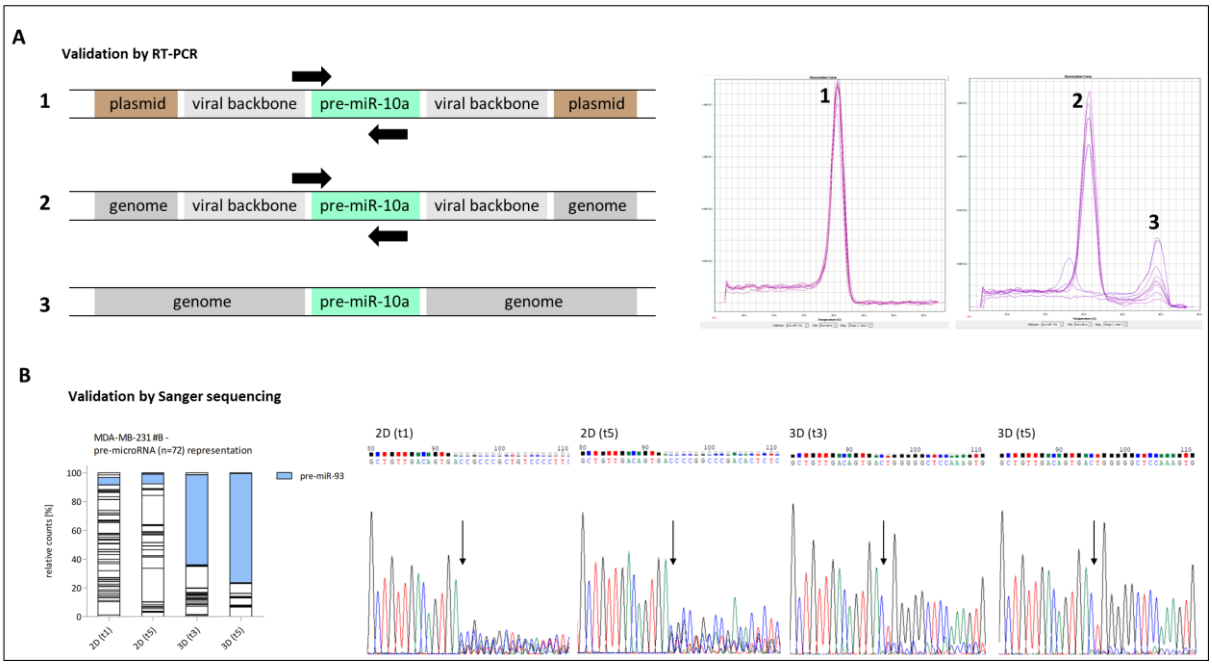


Figure 10: Validation of the NanoString results. A) Validation of the NanoString results by SYBRGreen-based qRT-PCR failed as shown exemplary for pre-miR-10a. The melt curves show that the designed primers were not specific enough to discriminate the retrovirally integrated pre-microRNA (1 and 2) from the endogenous pre-microRNA (3). Amplification of the pre-microRNA plasmid pool resulted in a specific melt curve showing only one peak (1). Performing the PCR on genomic DNA from stable cell lines harboring the pre-microRNA library resulted in a clear peak for the viral pre-microRNA integrates (2) and a smaller peak (3). Testing genomic DNA from the parental cell lines showed that the PCR partially amplified endogenous pre-microRNAs as well (3). **B)** Sanger sequencing confirmed the strong enrichment of pre-miR-93 in mammospheres from MDA-MB-231 #B (t3 and t5) that was detected by NanoString.

4.1.4 Heterogeneous overexpression and selection bias of pre-microRNAs in the library

NanoString detected strong enrichment of cells with pre-miR-103a-1 integrates in 3D conditions, which selected for stem cells and chemoresistant cells. Pre-miR-103a-1 overexpressing cell lines, however, did not modulate sphere formation capacity or drug response compared to the controls (Figure 9B). After I had confirmed that the NanoString assay performed reliably by Sanger sequencing (Figure 10), I hypothesized that the enrichment of pre-miR-103a-1 and other

microRNAs did not result from a selection advantage that overexpression of the respective pre-microRNAs mediated in 3D. To test this, I measured pre-miR-103a and miR-103a-3p levels in various SUM-159 samples from the screen. Expression of pre-miR-103a in 2D cell culture increased with time in some of the SUM-159 replicates (Figure 11A). In 3D, however, pre-miR-103a levels in t1 and t3 were higher than in t5, which does not correspond with the strong enrichment of the pre-miR-103a barcode in SUM-159#B spheres (Figure 8B). Moreover, the pre-miR-103a expression in 3D (t5) is even below the initial levels in 2D (t0). Subsequently, I wanted to know how pre-miR-103a overexpression affects microRNA expression of the mature form. For this purpose, miR-103a-3p levels were measured in SUM-159 #B, the sample in which 23 % of the cells in 3D (t5) carried the pre-miR-103a-1 barcode. Quantification of the mature microRNA levels confirmed the impression conveyed by the pre-miR-103a expression data: miR-103a-3p levels in 3D were only slightly more abundant than in 2D (Figure 11B). The fold changes for miR-103a-3p in 3D compared to 2D (t0) were highest at t1 with a 2.2-fold increase. From one to the next sphere generation, miR-103a-3p levels dropped and were lower than the expression in the time-matched 2D sample at t5, which is in line with the pre-miR-103a expression data.

To obtain a more detailed understanding of the overexpression levels, I reviewed data that had been generated in the context of another project. Xiaoya Li kindly provided the microRNA sequencing results from stable cell lines that overexpressed one pre-microRNA each and used the same retroviral backbone that I used for my pre-microRNA library. The data revealed that microRNA and 5'isomiR overexpression strongly varied between different pre-microRNAs (Figure 11C). The expression levels ranged from no overexpression for pre-miR-21, pre-miR-27a and pre-miR-320a to 10-fold, 100-fold or even 1000-fold overexpression for 5'isomiRs derived from other pre-microRNA constructs. The diverse overexpression patterns did not only affect the different pre-microRNAs, but also the mature microRNA level: the overexpression level of microRNAs and 5'isomiRs from the same pre-microRNA differed a lot from each other in several cases. In conclusion, qRT-PCR and microRNA sequencing confirmed the hypothesis that the enrichment of pre-microRNAs detected by NanoString did not result from a selection advantage mediated by overexpression of the individual pre-microRNAs, at least not in the case of pre-miR-103a-1.

Cindy Körner recently uncovered that the microRNA expression data from TCGA patients are strongly affected by batch effects. The R package 'ComBat' was applied by Susanne Ibing to correct for this bias. While my library was comprised of 63 pre-microRNAs, applying the same selection criteria to the batch-corrected TCGA data resulted in 85 pre-microRNAs (listed in Figure S1). After

batch correction, only 47 pre-microRNAs from my library were still included in the selection. The potential impact of the altered pre-microRNA library composition was not studied in the context of this project.

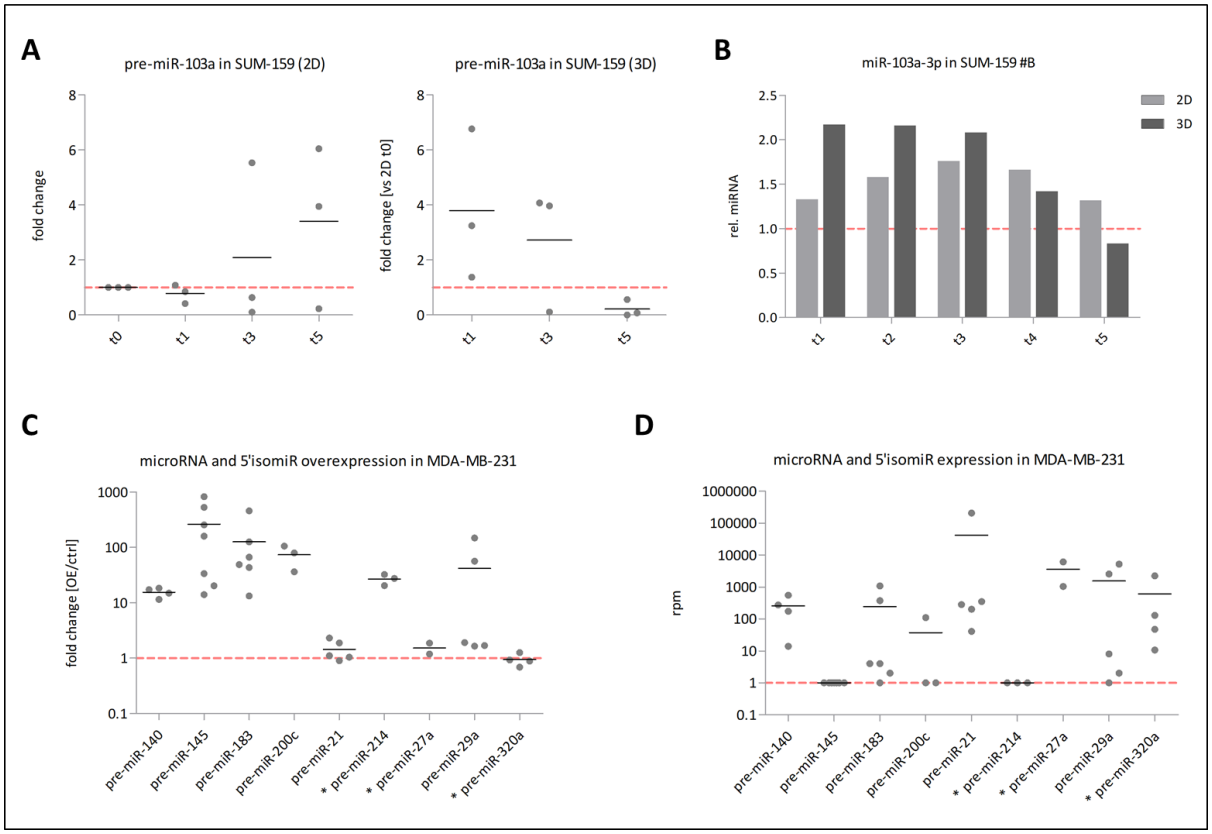


Figure 11: Pre-microRNA and microRNA overexpression levels in stable cell lines. A) Expression of pre-miR-103a in 2D and 3D samples of SUM-159 cells was determined by SYBRGreen-based qRT-PCR. microRNA levels were detected at different time points during the assay and normalized to the levels in 2D (t0). Each dot represents one of the SUM-159 transduction replicates #A, #B or #C. **B)** miR-103a-3p expression in SUM-159 #B, the sample in which cells with the pre-miR-103a-1 integrate strongly accumulated in the spheres. Expression at the different time points was detected with SYBRGreen-based qRT-PCR and normalized to the miR-103a-3p levels in 2D (t0). **C)** Overexpression of microRNAs and 5'isomiRs in stable MDA-MB-231 cells was analyzed by microRNA sequencing. microRNA levels in the cell line overexpressing the respective pre-microRNA were normalized to the cell lines overexpressing a non-targeting pre-microRNA control. Asterisks mark pre-microRNAs that were not included in the library generated for this project. Each dot represents one 5'isomiR. **D)** microRNA sequencing determined endogenous expression of microRNAs and 5'isomiRs in cell lines overexpressing a non-targeting pre-microRNA control. Asterisks mark pre-microRNAs that were not included in the library generated for this project. Each dot represents one 5'isomiR. Expression levels of 5'isomiRs with an expression between 0 and 1 rpm was set to 1 rpm to calculate reasonable fold changes in C). A), C), D) The horizontal line in each condition represents the mean.

4.2 Functional differences of divergent 5'isomiRs in triple-negative breast cancer

The overarching aim of my PhD project was to gain more insights into the role of the miRNome in chemoresistance of TNBC. Along these lines, I aimed at characterizing the functional differences that 5'isomiRs can mediate in this context. Several studies showed that 5'isomiRs may influence different phenotypes and mechanisms compared to their canonical microRNA[92,95,96]. Especially in the case of divergent 5'isomiRs, the target spectra and the resulting phenotypes may differ substantially despite a seed sequence that is shifted only by one nucleotide[95,96]. To study microRNAs and 5'isomiRs with an impact on chemoresistance, I described the enrichment of selected overexpressed pre-microRNAs in mammospheres with NanoString in the previous chapter. The pre-microRNAs were selected based on differentially expressed microRNAs and 5'isomiRs from TCGA breast cancer patients. The mammosphere assay enriched for chemoresistant BCSCs (Figure 6) and, therefore, was considered as ideal setup to enrich for pre-microRNAs that modulate resistance to chemotherapy. Despite the strong enrichment of pre-miR-103a-1 (Figure 8), however, I could not show a functional impact of this pre-microRNA on sphere formation or chemoresistance (Figure 9). As a result, I could not study potential differences between the microRNAs and 5'isomiRs transcribed from pre-miR-103a-1 on chemoresistance and, therefore, this project was not pursued further. Instead, I decided for an alternative approach to study the functional relevance of 5'isomiRs.

4.2.1 miR-1307-3p I0 and its divergent 5'isomiR are highly abundant in breast cancer

I selected 5'isomiR pairs that were differentially expressed in tumor compared to normal tissue in the TCGA breast cancer data set in order to identify and subsequently characterize 5'isomiRs that are relevant for tumor progression or prevention in breast cancer patients. Moreover, I focused on 5'isomiRs with divergent seed sequences that were expressed at similar levels and were not studied and published yet[95]. Applying these criteria prioritized three 5'isomiR pairs, which were upregulated in tumor compared to normal tissue (Table 10). Since all of them were upregulated significantly, I chose miR-1307-3p I0 and miR-1307-3p I1 due to the highest abundance in tumor tissue rather than selecting the 5'isomiR pair based on the highest fold change.

Table 10: Differentially expressed 5'isomiR pairs in TCGA breast cancer patients.

	normal (mean rpm)	tumor (mean rpm)	seed sequence	publication status
miR-1307-3p I0	250	575	divergent	unpublished
miR-1307-3p I1	151	416		
miR-21-5p I2	8	64	divergent	unpublished
miR-21-5p I3	5	47		
miR-425-5p I0	37	183	divergent	unpublished
miR-425-5p I1	47	160		
miR-140-3p I0	515	246	divergent	published
miR-140-3p I1	769	316		
miR-183-5p I0	1341	8827	divergent	in preparation
miR-183-5p I1	821	6582		
miR-192-5p I0	90	260	convergent	unpublished
miR-192-5p I1	47	180		
miR-203a-3p I0	8736	5079	convergent	unpublished
miR-203a-3p I1	4670	2758		

miR-1307-3p I0 and miR-1307-3p I1 were upregulated in tumor tissue and were found to be expressed at similar levels when looking at tumor or normal tissue separately (Figure 12A). Moreover, the expression of miR-1307-3p I0 and miR-1307-3p I1 was significantly higher in patients diagnosed with the aggressive subtype TNBC compared to patients with tumors of other subtypes (Figure 12B). This was also true for the aggressive PAM50 subtypes Basal and HER2 (Figure 12C). Expression levels in Luminal A, the subtype with the best prognosis, were significantly lower for miR-1307-3p I0 as well as for miR-1307-3p I1. Based on the high abundance of both isoforms in tumors, especially of the aggressive subtypes, I hypothesized that miR-1307-3p I0 and miR-1307-3p I1 have oncogenic functions. To evaluate the role of both isoforms in cancer and to identify potential differences between microRNA and 5'isomiR, I next investigated the phenotypic consequences of miR-1307-3p overexpression.

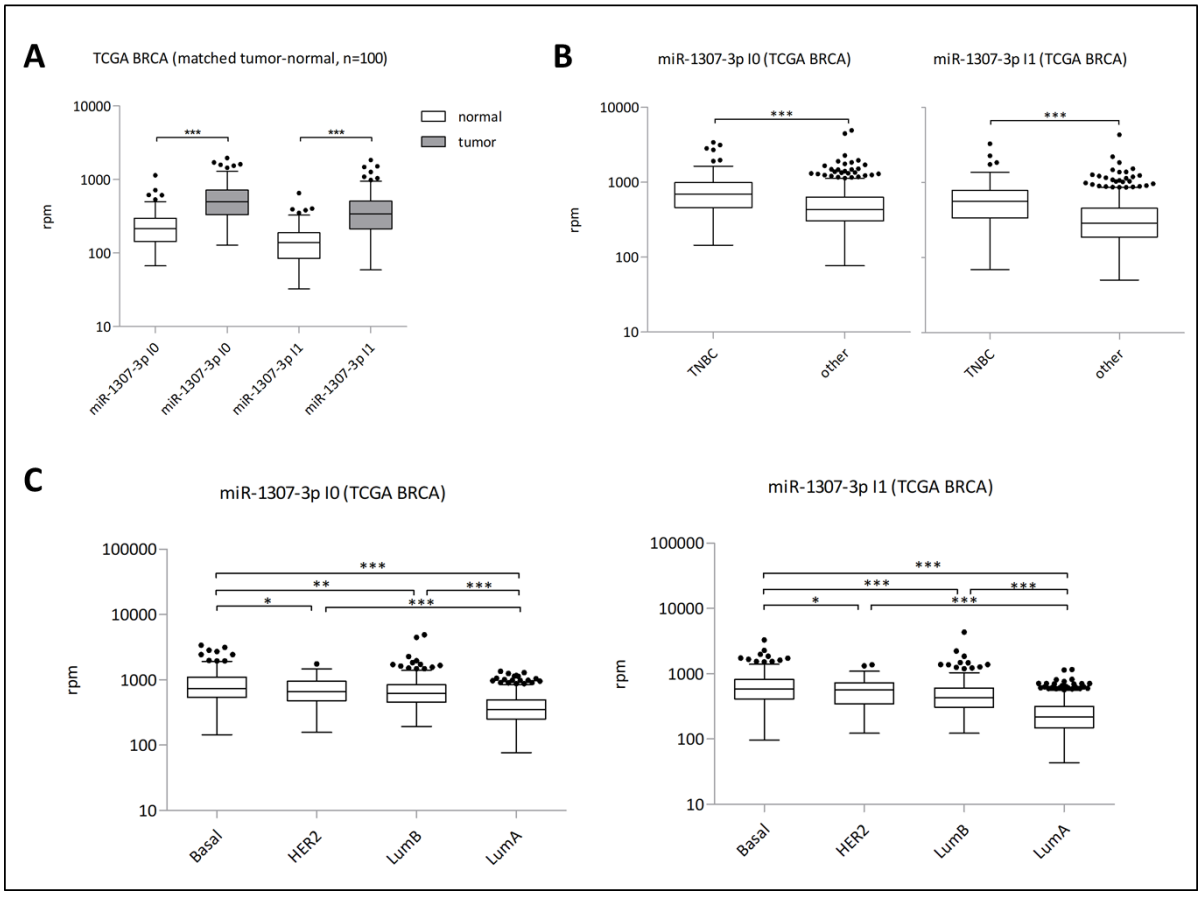


Figure 12: miR-1307-3p I0 and miR-1307-3p I1 expression in breast cancer. A) Expression of miR-1307-3p I0 and miR-1307-3p I1 in matched tumor-normal samples of TCGA breast cancer patients. Statistical comparison was based on paired, two-tailed Student's t-tests. P-values are represented by asterisks (* ≤ 0.05, ** ≤ 0.01 and *** ≤ 0.001). **B), C)** Expression of miR-1307-3p I0 and miR-1307-3p I1 in 108 TNBC versus 592 non-TNBC patients and in the PAM50 subtypes of breast cancer patients from TCGA. The PAM50 subtypes comprised Basal (n=169), HER2 (n=80), LumB (n=191) and LumA (n=540). Statistical comparison was based on unpaired, two-tailed Student's t-tests. P-values are represented by asterisks (* ≤ 0.05, ** ≤ 0.01 and *** ≤ 0.001).

4.2.2 pre-miR-1307 reduces mesenchymal traits of MDA-MB-231

After selection of miR-1307-3p I0 and miR-1307-3p I1, I aimed at identifying cancer-associated phenotypes that might be differentially modulated by the 5'isomiR pair with the divergent seed sequences. For that purpose, a Doxycycline-inducible pre-miR-1307 overexpression cell line was generated using MDA-MB-231 cells. The impact of pre-miR-1307 overexpression on various phenotypes was evaluated by me (stemness, chemoresistance) and by Xiaoya Li (proliferation, cell cycle, apoptosis, migration, invasion and tumor growth *in vivo*). Stemness and chemoresistance were not altered by pre-miR-1307 compared to two non-targeting pre-microRNA controls (Figure S2). Furthermore, pre-miR-1307 overexpression did not affect cell cycle or apoptotic

behavior, even when induced with Staurosporine (Figure S2). For migration and invasion, however, Xiaoya Li detected a tumor-suppressive effect upon pre-miR-1307 overexpression (Figure 13A). Although there was no difference for proliferation *in vitro*, tumor volume was decreased 35 days post injection of the stable pre-miR-1307 cells into the mammary fat pad of immunocompromised mice (Figure 13B). Mechanistically, I hypothesized that miR-1307-3p I0 and miR-1307-3p I1 can only have a differential impact on migration since pre-miR-1307 overexpression did not modulate any of the other tested phenotypes.

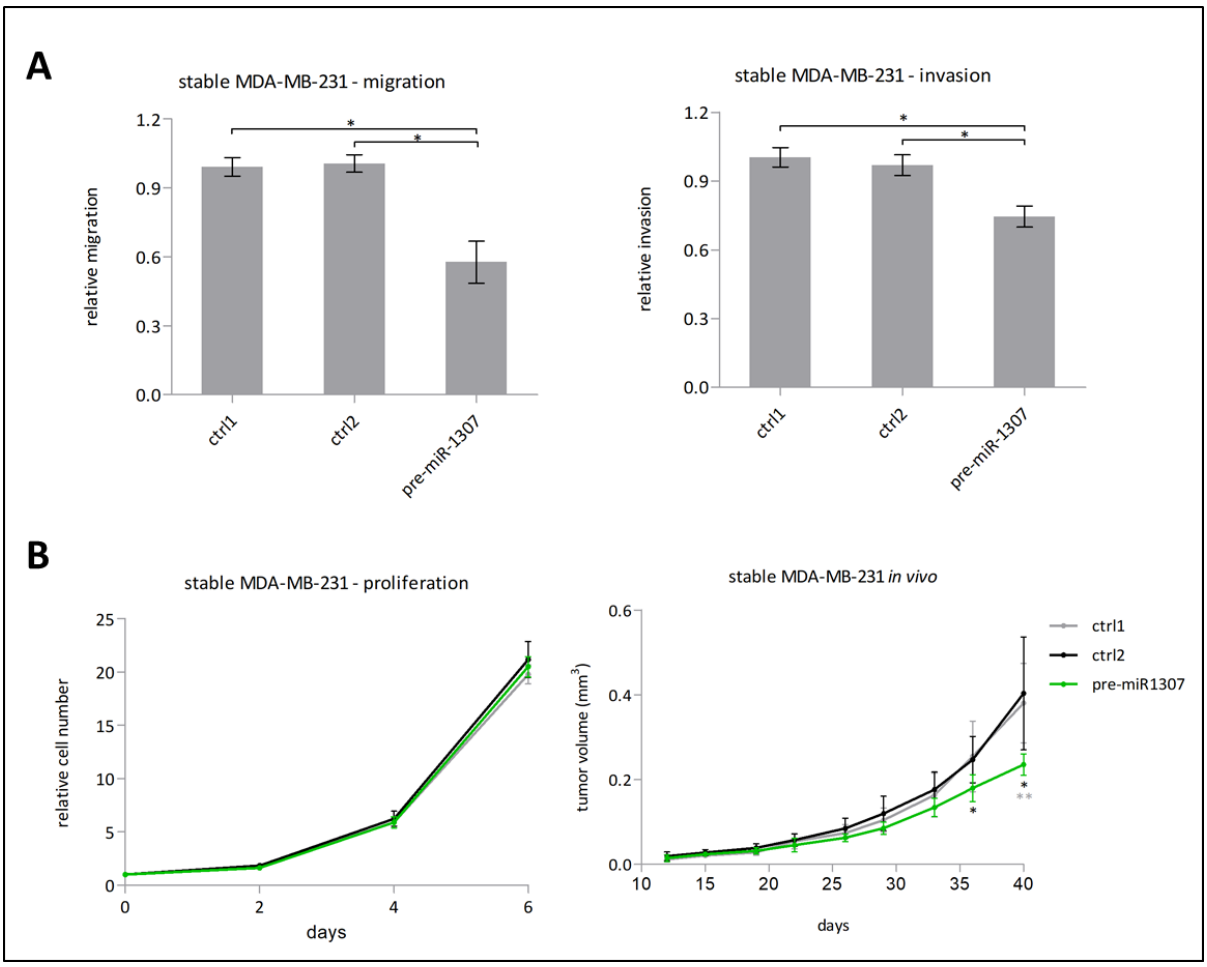


Figure 13: Cancer phenotypes modulated by pre-miR-1307 overexpression in MDA-MB-231. MDA-MB-231 cells were retrovirally transduced with pre-miR-1307 or two different non-targeting pre-microRNAs under control of a Doxycycline-inducible promoter. **A)** Migration and invasion were assessed in a transwell-based assay (n=3). **B)** Proliferation was quantified by counting the nuclei with a microscope after Hoechst staining (n=3). Moreover, the effect of pre-miR-1307 overexpression was evaluated *in vivo*. For this purpose, the stable cell lines were injected into the mammary fat pad of immunocompromised mice (n=6) and the tumor volume of each mouse was quantified at the indicated time points after engraftment. Error bars represent the standard deviation of biological replicates. Statistical significance was calculated with an unpaired two-tailed student's t-test. P-values are represented by asterisks (* ≤ 0.05, ** ≤ 0.01).

4.2.3 miR-1307-3p I1 reduces migration and proliferation

Phenotypic assays identified reduced migration and invasion upon pre-miR-1307 overexpression. Since miR-1307-3p I0 and miR-1307-3p I1 are highly abundant microRNAs processed from pre-miR-1307 and have divergent seed sequences, I assumed that the two isoforms differ in their functional impact on migration. To test this hypothesis, I transiently transfected MDA-MB-231 and SUM-159 cells with isoform-specific microRNA mimics. This time, migration was assessed with an RTCA system which allows time-resolved profiling of transwell-based migration. Indeed, miR-1307-3p I1 reduced migration significantly in both TNBC cell lines (Figure 14A). miR-1307-3p I0, however, reduced migration only in MDA-MB-231 cells and there much less compared to miR-1307-3p I1. Next, I assumed that the impact on migration might be accompanied by changes in EMT markers. To evaluate this, I quantified mRNA levels of *CDH1*, *CDH2*, *FN1*, *VIM*, *SNAI1* and *SNAI2* in both cell lines. Both, miR-1307-3p I0 and miR-1307-3p I1, increased *CDH1* and decreased *VIM* levels in MDA-MB-231 significantly (Figure 14B). This observation confirmed the more epithelial phenotype exerted by miR-1307-3p I0 and miR-1307-3p I1 in MDA-MB-231. For SUM-159, none of the EMT markers was significantly modulated and *CDH1* was not expressed (data not shown). Subsequently, I wanted to assess whether the connection between miR-1307-3p I0 / miR-1307-3p I1 and a more epithelial phenotype is also reflected in patients. For this purpose, Cindy Körner performed GSEA using gene expression data from Basal patients of the TCGA cohort. Patients with high levels of miR-1307-3p I0 / miR-1307-3p I1 did indeed show a depletion of an EMT-related gene signature (Figure 14C). Based on the strong correlation of miR-1307-3p I0 and miR-1307-3p I1 ($r>0.9$) in tumors of TCGA breast cancer patients (Figure 14D), the observed depletion of EMT-related genes in this data set could not be attributed to a particular 5'isomiR.

Since both, miR-1307-3p I0 and miR-1307-3p I1, modulated migration, I next tested a potential differential effect the 5'isomiRs might have on proliferation since pre-miR-1307 overexpression repressed tumor growth *in vivo* (Figure 13B). And indeed, proliferation of SUM-159 cells was significantly reduced by miR-1307-3p I1, whereas miR-1307-3p I0 did not have an impact (Figure 14E). For MDA-MB-231 on the other hand, both 5'isomiRs significantly reduced the cell number compared to the negative controls but the effects were minor and in line with the proliferation results for pre-miR-1307 (Figure 13B). In conclusion, the data illustrates that miR-1307-3p I1 has a functionally different impact than miR-1307-3p I0, at least in SUM-159 cells. In this cell line, only miR-1307-3p I1 reduces migration and proliferation. Consistently, miR-1307-3p I1 strongly reduced migration in MDA-MB-231, whereas miR-1307-3p I0 had only a mild effect. The differences on the phenotype of the two TNBC cell lines, indicate a context-dependent impact

of miR-1307-3p I0 and miR-1307-3p I1. The effect of the individual isoforms might be influenced by differentially expressed targets or activated pathways as well as by functionally relevant mutations that are present in either MDA-MB-231 or SUM-159 cells.

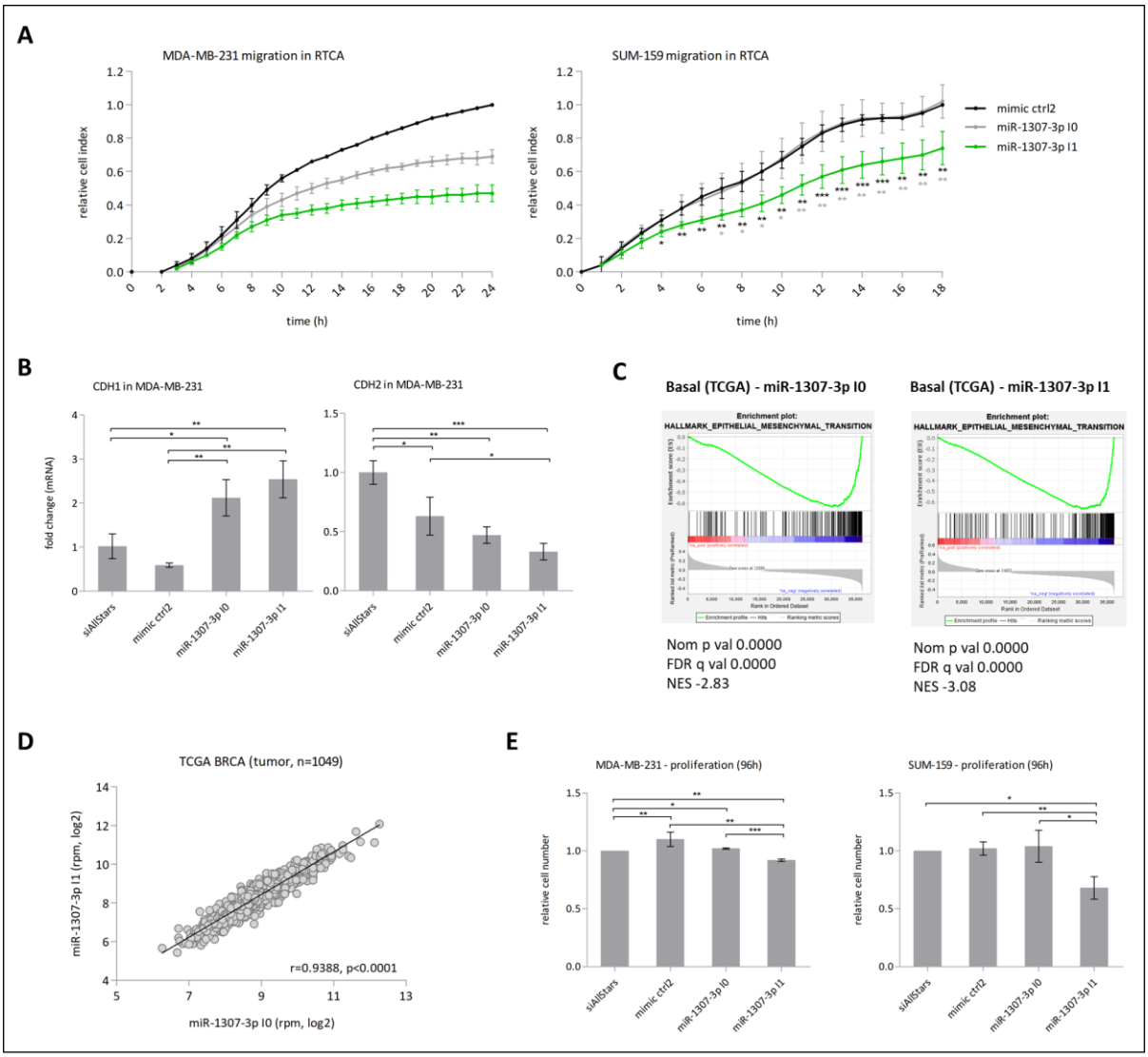


Figure 14: Different impact of miR-1307-3p I0 and miR-1307-3p I1 on migration and proliferation. Cells were transfected with miR-1307-3p I0, miR-1307-3p I1 or non-targeting control mimics. **A)** Transwell migration of MDA-MB-231 (n=2) and SUM-159 (n=3) was detected by RTCA. **B)** Expression of EMT markers in MDA-MB-231 (n=3) was measured with TaqMan. **C)** GSEA of Basal patients from TCGA ranked by their miR-1307-3p I0 or miR-1307-3p I1 expression was performed and revealed significant depletion of an EMT gene signature. NES represents the normalized enrichment score indicating negative enrichment. **D)** Correlation of miR-1307-3p I0 and miR-1307-3p I1 in tumors from the TCGA breast cancer cohort. Statistical comparison was based on Spearman's correlation. **E)** Proliferation assays were performed in MDA-MB-231 and SUM-159 (each n=3). Cell number was determined by microscope after Hoechst staining of the nuclei. A), B), E) Error bars represent the standard deviation of biological replicates. Statistical significance was calculated with an unpaired two-tailed student's t-test. One sample t-tests were performed for comparison where one condition lacked a standard deviation due to the way the data was normalized. P-values are represented by asterisks (* ≤ 0.05, ** ≤ 0.01, *** ≤ 0.001).

4.2.4 pre-miR-1307 is a mirtron of the highly expressed ATP synthase subunit *ATP5MD*

In vitro and *in vivo* experiments revealed a tumor-suppressive phenotype of pre-miR-1307 (Figure 13) as well as miR-1307-3p I0 or miR-1307-3p I1 (Figure 14) overexpression. These findings are surprising considering the high abundance of miR-1307-3p I0 or miR-1307-3p I1 in tumors, especially in the more aggressive subtypes (Figure 12). This contradictory data led me to evaluate the genomic location of pre-miR-1307. Pre-miR-1307 is located on the minus strand of chromosome 10 and a mirtron of the gene 'ATP Synthase Membrane Subunit DAPIT' (*ATP5MD*), which codes for a subunit of the mitochondrial ATP synthase (Figure 15A). ATP5MD is a highly conserved peptide with 58 amino acids, the particular function and relevance of that peptide is unknown. In breast cancer patients, *ATP5MD* is significantly higher expressed in the aggressive subtypes, Basal/TNBC and HER2 (Figure 15B) suggesting an oncogenic role.

Based on the high abundance of *ATP5MD* and the microRNAs that are expressed from its mirtron pre-miR-1307 in breast tumors, I hypothesized that *ATP5MD* and pre-miR-1307 are coregulated. To test this, I correlated the expression levels of *ATP5MD* with miR-1307-3p I0 (data not shown) or miR-1307-3p I1 (Figure 15C). Looking at all TCGA breast cancer patients, *ATP5MD* correlated moderately with miR-1307-3p I0 or miR-1307-3p I1 (r=0.41 and r=0.38, respectively). In Basal patients, the correlation of miR-1307-3p I0 or miR-1307-3p I1 with *ATP5MD* increased slightly (r=0.42 and r=0.41, respectively). Overall, the expression of *ATP5MD* and its mirtron correlates only mildly, suggesting that these transcripts are not strongly co-regulated.

Based on the mild correlation of *ATP5MD* and miR-1307-3p, I hypothesized that the encoded protein and the microRNAs might be involved in different cellular processes. To confirm this hypothesis, Cindy Körner performed GSEA on Basal TCGA patients and I compared gene signatures that were significantly enriched or depleted. This analysis, however, showed that gene sets related to OXPHOS, ROS and DNA repair were significantly enriched for *ATP5MD* (Figure 15D), miR-1307-3p I0 (data not shown) and miR-1307-3p I1 (Figure 15D). The genes enriched in the OXPHOS signature of patients with high ATP5MD levels contained many components of the ATP synthase, including the ATP Synthase F1 Subunit of which ATP5MD is a part. Moreover, several markers of mitochondrial biogenesis and translation, such as *MRPS15*, *MRPS22* and *TIMM9*, were enriched. In summary, ATP5MD seems to be highly relevant to Basal breast tumors in the context of OXPHOS signaling and related cellular functions. The exact role and relevance of this subunit, however, is not clear yet. Furthermore, no functional studies connect pre-miR-1307 to metabolism.

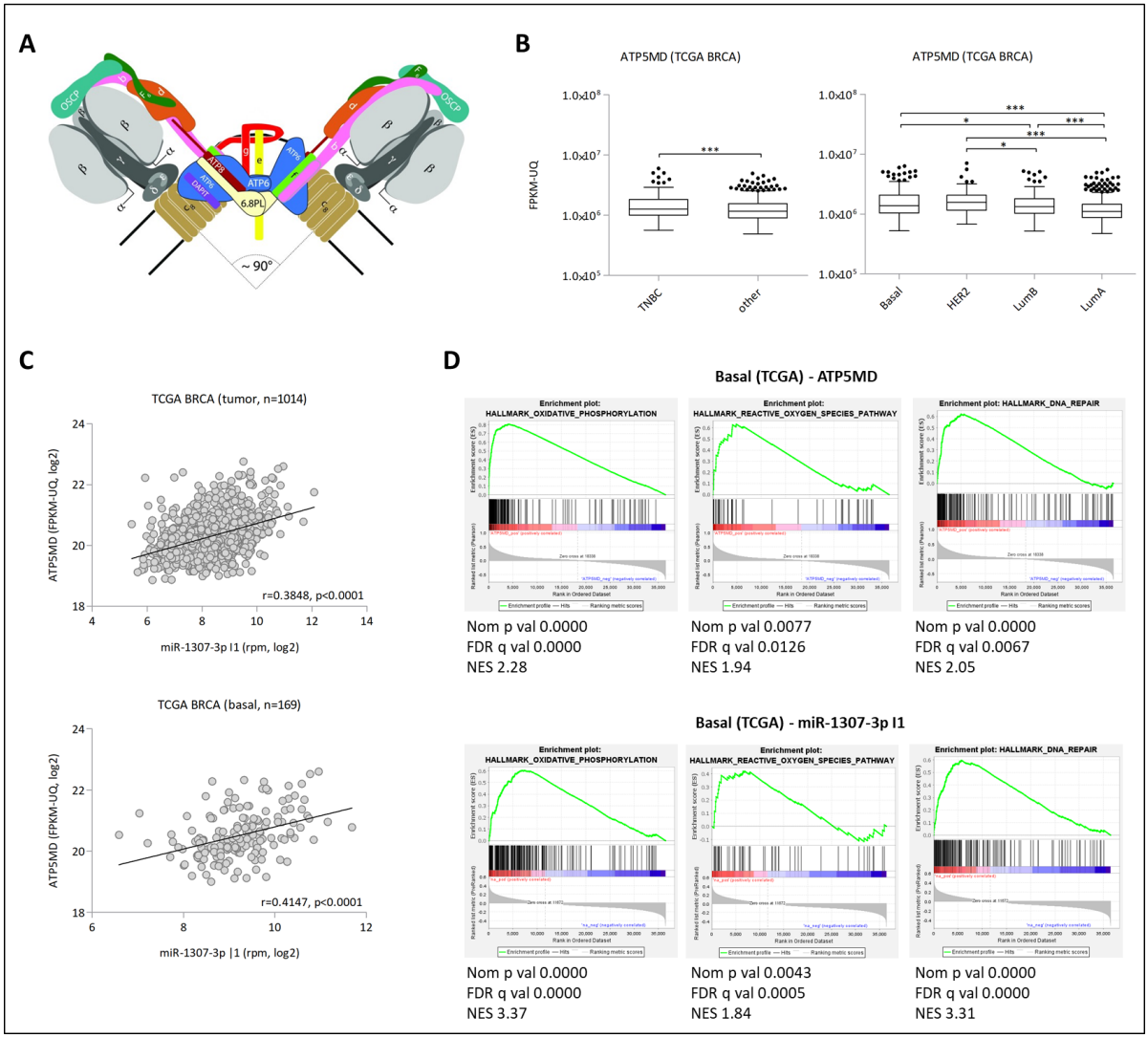


Figure 15: ATP synthase subunit ATP5MD and its mirtron pre-miR-1307. A) Structure of the mammalian ATP synthase and localization of ATP5MD (DAPIT). The image was taken from Carroll et al[154]. **B)** Expression of *ATP5MD* in 108 TNBC versus 592 non-TNBC and in the PAM50 subtypes of breast cancer patients from TCGA. The PAM50 subtypes comprised Basal (n=169), HER2 (n=80), LumB (n=191) and LumA (n=540). Statistical comparison was based on unpaired, two-tailed Student's t-tests. P-values are represented by asterisks (* ≤ 0.05, ** ≤ 0.01 and *** ≤ 0.001). **C)** Correlation of miR-1307-3p I1 and *ATP5MD* in the TCGA breast cancer cohort. Statistical comparison was based on Spearman's correlation. **D)** Gene signatures significantly enriched by GSEA in Basal TCGA patients ranked by their *ATP5MD* or miR-1307-3p I1 expression. NES represents the normalized enrichment score indicating positive enrichment

4.2.5 Myc as potential regulator of pre-miR-1307 and *ATP5MD*

The mapping of pre-miR-1307 within an *ATP5MD* intron raises the question of a potential biological reason. GSEA suggested some functional overlap, but neither the functional role of ATP5MD as subunit of the ATP synthase nor its relevance in breast cancer are well understood. I assumed that GSEA in breast cancer patients might identify transcription factors that could explain

the upregulation of *ATP5MD* and pre-miR-1307 in breast cancer, especially in the aggressive subtypes. Indeed, two different gene signatures related to Myc targets were significantly enriched in Basal patients with high levels of miR-1307-3p I0 (data not shown), miR-1307-3p I1 (Figure 16A) as well as *ATP5MD* (Figure 16B). Among the genes contributing to the enrichment of these gene sets were *MYC* itself and the Myc targets *NPM1* and *FBL*[155]. The amplification and overexpression of *MYC* in breast cancer, especially in the Basal/TNBC subtype[156,157], is well established.

The enrichment of Myc targets in patients with high miR-1307-3p I0 / miR-1307-3p I1 or *ATP5MD* levels, let me assume that *ATP5MD* and its mirtron pre-miR-1307 might also be regulated by Myc. To test this hypothesis, I searched for ChIP-Seq data in proximity to the genomic location of *ATP5MD* in breast cancer cell lines. The ENCODE data base contained ChIP-Seq data from the Luminal breast cancer cell line MCF7 (experiment ENCSR000DMM) for the region of interest and was visualized with the UCSC Genome Browser (Figure 16C). The results revealed a prominent Myc peak upstream of the first exon of *ATP5MD*, indicating a potential regulation of *ATP5MD* and also of pre-miR-1307 by Myc. To test this hypothesis experimentally, *MYC* was knocked-down with siRNAs in MDA-MB-231 and SUM-159 cells and miR-1307-3p levels as well as *ATP5MD* mRNA levels were measured by TaqMan. Moreover, I determined *PDCD11* levels since the Myc peak in MCF7 was also located at the beginning of this gene and suggested a potential regulation by Myc as well. Upon knock-down, *MYC* levels were only reduced by 60 % in MDA-MB-231 (Figure 16D) and by 25 % in SUM-159 (Figure 16E). Increasing the siRNA concentration from 2 nM to 5 nM reduced *MYC* levels by 30 % in SUM-159 (data not shown). These observations suggested that the siRNAs targeting *MYC* were not very efficient or that the siRNA concentration was not sufficient to downregulate the mRNA. miR-1307-3p and *ATP5MD* levels were not repressed by *MYC* knock-down, whereas *PDCD11* was reduced by 50 % in MDA-MB-231 and by 30 % in SUM-159. Increasing the siRNA concentration for SUM-159 did not repress miR-1307-3p or *ATP5MD* levels either, *PDCD11* levels, however, were reduced by 45 % (data not shown). In conclusion, the data suggests that miR-1307-3p and *ATP5MD* are not regulated by Myc, whereas *PDCD11* is repressed by *MYC* knock-down even with a knock-down efficiency of only 25-60 %.

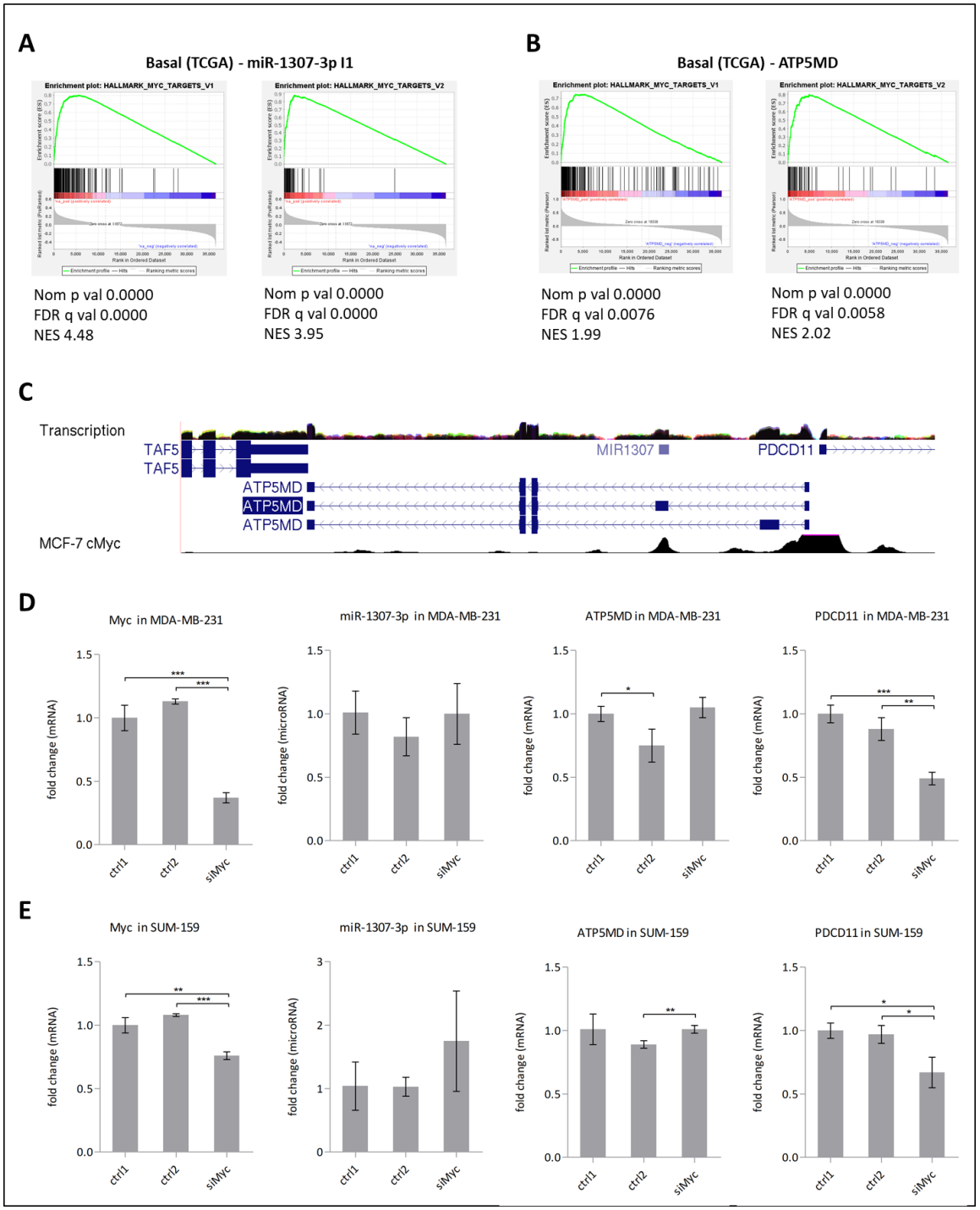


Figure 16: Potential regulation of *ATP5MD* and its mirtron pre-miR-1307 by Myc. A), B) Myc targets significantly enriched by GSEA in Basal TCGA patients ranked by miR-1307-3p I1 or *ATP5MD* expression. NES represents the normalized enrichment score indicating positive enrichment. **C)** Genomic location of pre-miR-1307 and *ATP5MD* in the UCSC genome browser and ChIP-Seq data in MCF7 for the region of interest (ENCODE experiment ENCSR000DMM). **D), E)** Potential Myc regulation of miR-1307-3p, *ATP5MD* and *PDCD11* in MDA-MB-231 (D) and SUM-159 (E). mRNA expression was determined by TaqMan, microRNA expression was analyzed by SYBRGreen-based qRT-PCR. Error bars represent the standard deviation of three biological replicates, except for *ATP5MD* in MDA-MB-231 (n=2). Statistical significance was calculated with an unpaired two-tailed student's t-test. P-values are represented by asterisks (* ≤ 0.05, ** ≤ 0.01, *** ≤ 0.001).

4.2.6 miR-1307-5p I0 as driver for high miR-1307-3p levels in breast cancer patients?

The *in vitro* and *in vivo* experiments performed in this study revealed a tumor-suppressive phenotype of pre-miR-1307, which contradicted the high abundance in tumor tissue of breast cancer patients. Since I could not show a co-regulation of miR-1307-3p and its host gene *ATP5MD* by Myc, I assumed that an oncogenic microRNA might be processed from the 5p arm of pre-miR-1307 and, thus, could explain the high expression levels in tumors. In TCGA breast cancer patients as well as in various breast cancer cell lines that were sequenced by the lab, miR-1307-3p I0, miR-1307-3p I1 and miR-1307-5p I0 were the only three expressed (> 15 rpm) microRNAs from pre-miR-1307 (Figure 12A, Figure 17B, data not shown). The reduced migratory potential that we observed upon miR-1307-3p I0 or miR-1307-3p I1 overexpression rendered miR-1307-5p I0 the only possible microRNA derived from pre-miR-1307 that could exert an oncogenic function. However, I did not investigate the functional role of miR-1307-5p I0 in the context of this project due to time constraints.

To evaluate whether it would make sense to look into a potentially oncogenic role of miR-1307-5p I0, I analyzed miR-1307-5p I0 expression levels in TCGA patient data and compared them to miR-1307-3p I0 and miR-1307-3p I1 abundance. miR-1307-5p I0 was well correlated with miR-1307-3p I0 and miR-1307-3p I1 (Figure 17A). Similar to miR-1307-3p I0 and miR-1307-3p I1, miR-1307-5p I0 was also significantly differentially expressed between tumor and normal breast tissue (Figure 17B) and displayed higher expression levels in aggressive breast cancer subtypes (Figure 17C). I hypothesized that there would be a stronger increase in expression levels when comparing tumor to normal tissue for miR-1307-5p I0 than for miR-1307-3p I0 and miR-1307-3p I1, if miR-1307-5p I0 is the main factor for high pre-miR-1307 abundance in tumors. When looking at fold change expression in tumor compared to normal tissue of the same patient, the mean fold change was similar for miR-1307-3p I0 and miR-1307-3p I1, but indeed much higher for miR-1307-5p I0 (Figure 17D). Nevertheless, the data was not completely conclusive since the differences in the fold change for all three microRNAs were significant and the spread of the fold change in miR-1307-5p I0 expression across patients was very high.

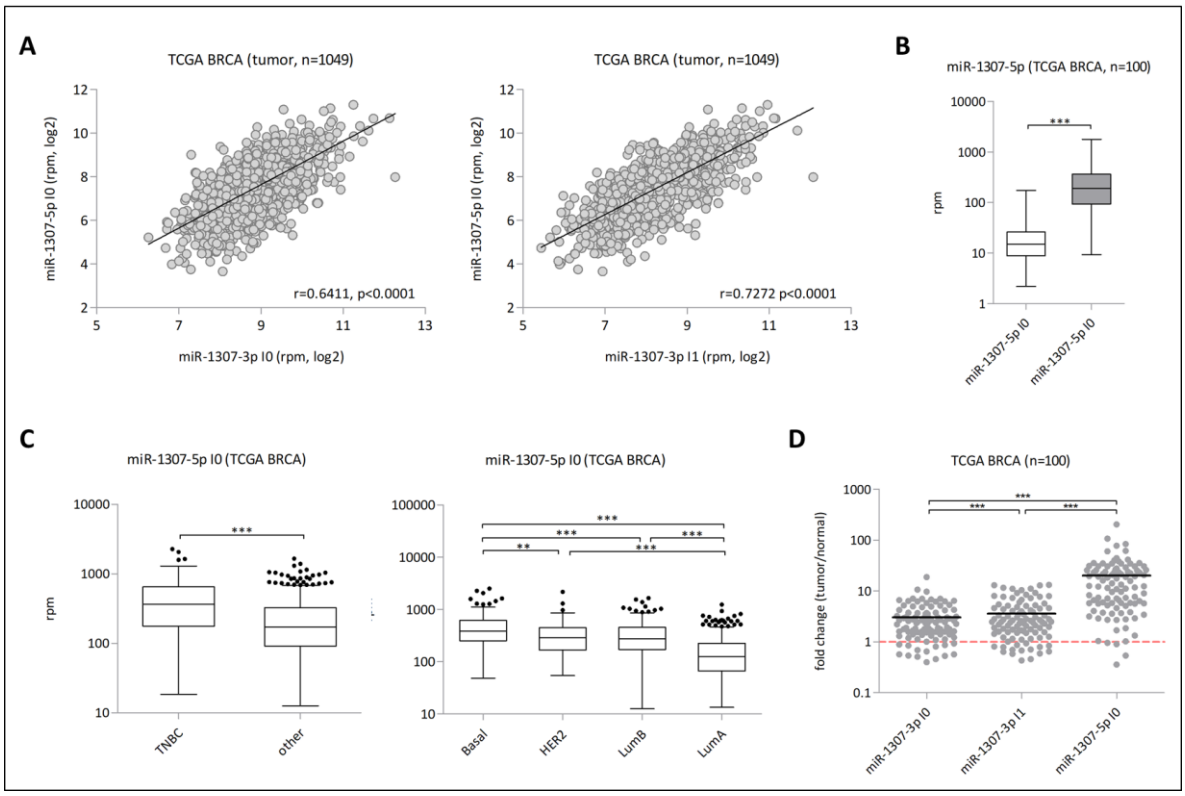


Figure 17: miR-1307-5p I0 expression in breast cancer. A) Correlation of miR-1307-3p I0 and miR-1307-3p I1 with miR-1307-5p I0 in tumors from the TCGA breast cancer cohort. Statistical comparison was based on Spearman's correlation. **B)** Expression of miR-1307-5p I0 in matched tumor-normal samples of TCGA breast cancer patients. Statistical comparison was based on paired, two-tailed Student's t-tests. P-values are represented by asterisks (* ≤ 0.05, ** ≤ 0.01 and *** ≤ 0.001). **C)** Expression of miR-1307-5p I0 in 108 TNBC versus 592 non-TNBC patients and in the PAM50 subtypes of breast cancer patients from TCGA. The PAM50 subtypes comprised Basal (n=169), HER2 (n=80), LumB (n=191) and LumA (n=540). Statistical comparison was based on unpaired, two-tailed Student's t-tests. P-values are represented by asterisks (* ≤ 0.05, ** ≤ 0.01 and *** ≤ 0.001). **D)** Fold changes of miR-1307-3p I0, miR-1307-3p I1 and miR-1307-5p I0 between tumor and normal tissue from the same patient. The mean fold change for each microRNA across all patients is indicated by a horizontal line. Statistical comparison was based on paired, two-tailed Student's t-tests. P-values are represented by asterisks (* ≤ 0.05, ** ≤ 0.01 and *** ≤ 0.001).

4.2.7 miR-1307-3p I0 and miR-1307-3p I1 have distinct and shared target subsets

Phenotypic assays showed partially cell line-dependent, tumor-suppressive effects of miR-1307-3p I0 or miR-1307-3p I1 overexpression on migration. Moreover, miR-1307-3p I1 reduced proliferation in a cell-line dependent manner. I hypothesized that the divergent seed sequences of miR-1307-3p I0 and miR-1307-3p I1 might have a substantial impact on the target spectrum and, thus, result in differentially modulated target genes and pathways that might explain the differences in cancer-associated phenotypes that I observed. This hypothesis was addressed with a MicroArray and target predictions. Shashwat Sahay generated target predictions for miR-1307-

3p I0 and miR-1307-3p I1 by using the target prediction algorithms from TargetScan and MiRanda. In addition to the seed matches that are predicted by TargetScan as well, the MiRanda algorithm takes secondary structure predictions in form of minimum free energy calculations into account and, therefore, is more stringent. The higher stringency of predictions generated with the MiRanda algorithm was reflected by a smaller number of predicted targets (Table 11). Consensus predictions from two different algorithms thus allowed limiting the number of false positive predictions. Since predicted microRNA binding sites are often located in 3'UTRs of mRNAs that are not expressed, Shashwat Sahay intersected the predictions with expression data from breast cancer cell lines.

Moreover, a MicroArray overexpressing miR-1307-3p I0 or miR-1307-3p I1 in MDA-MB-231 was performed by Neşe Erdem Borgoni in order to identify repressed genes ($p<0.05$, fold change<0.65) that contribute to the phenotypes mediated by miR-1307-3p I0 and miR-1307-3p I1. Regulation of genes upon miR-1307-3p I0 or miR-1307-3p I1 overexpression was analyzed in two biological replicates per condition. To obtain genes specifically repressed by miR-1307-3p I0 or miR-1307-3p I1, the respective condition was compared to cells transfected with the non-targeting control siAllStars. The MicroArray data needed to be interpreted with caution, however, since the Benjamini-Hochberg corrected p values for all calculated fold changes between two conditions were not significant and the uncorrected p values did not deviate from a random distribution. The insignificant p values suggest that the genes considered significantly repressed by miR-1307-3p I0 or miR-1307-3p I1 based on the uncorrected p value might represent artefacts and observed differences might be random. For this reason, I decided to focus on the target predictions for identifying potentially relevant targets that explain the phenotypic differences between miR-1307-3p I0 and miR-1307-3p I1. Overall, the target predictions and the MicroArray revealed shared as well as specific target subsets for miR-1307-3p I0 and miR-1307-3p I1.

Table 11: Targets predicted for or downregulated by miR-1307-3p I0 or miR-1307-3p I1

	shared	miR-1307-3p I0	miR-1307-3p I1
TargetScan	346	2,672	138
MiRanda	54	211	56
consensus	54	206	52
MicroArray ($p<0.05$, FC<0.65)	40	30	89

4.2.8 miR-1307-3p I1 might reduce migration and proliferation by targeting *NCS1*

To identify pathways that are differentially modulated by miR-1307-3p I0 and miR-1307-3p I1 and might explain the phenotypic differences, I used the target predictions to perform functional enrichment with WebGestalt. This analysis, however, did not yield any pathway enrichment with statistical significance. Most likely the reason for the lack in 5'isomiR-specific enrichment was due to the rather small number of predictions (compared to other microRNAs that the lab is working with). In order to determine potentially relevant targets nevertheless, I checked the literature for predicted targets from the consensus list. Literature search was performed using key words like migration and proliferation in combination with the consensus target predictions specific for miR-1307-3p I1. Another requirement was a tumor-promoting role of the respective targets matching the tumor-suppressive function of miR-1307-3p I1. 'Neuronal Calcium Sensor 1' (*NCS1*) met the defined criteria: the miR-1307-3p I1-specific target was shown to promote migration and metastasis in breast cancer *in vitro* and *in vivo*, respectively[158]. Moreover, NCS1 promoted proliferation and migration in breast cancer cells via enhancing AKT activity[159].

Based on the literature, downregulation of *NCS1* by miR-1307-3p I1 might explain the changes in proliferation and migratory behavior upon overexpression of the 5'isomiR. To confirm the gene regulation experimentally, I checked *NCS1* levels in the MicroArray and validated its expression in an independent set of samples generated in MDA-MB-231 by TaqMan. In the MicroArray, *NCS1* was downregulated by 35 % upon miR-1307-3p I1 overexpression, whereas miR-1307-3p I0 overexpression modulated *NCS1* levels only slightly (Figure 18A). Validation of the target repression by TaqMan confirmed the MicroArray results. TaqMan analysis revealed a significantly lowered mRNA level of *NCS1* upon miR-1307-3p I1 in MDA-MB-231 (Figure 18B). miR-1307-3p I0 overexpression reduced *NCS1* levels as well, but this was only significant in comparison to ctrl2, which might originate from the high standard deviation of siAllStars, the second non-targeting control. Nevertheless, the impact of miR-1307-3p I1 was stronger (approximately 30 % compared to 20 % for miR-1307-3p I0). Since *NCS1* was predicted to be a miR-1307-3p I1-specific target in the consensus list, but was downregulated upon miR-1307-3p I0 overexpression as well, I looked at the target predictions generated by TargetScan and MiRanda. TargetScan predicted NCS1 as a shared target, while the more stringent algorithm MiRanda predicted *NCS1* as a miR-1307-3p I1-specific target. As a result, the consensus list contained *NCS1* as miR-1307-3p I1-specific target. The type of binding sites that were predicted by TargetScan provided an explanation for the extent of gene regulation detected by TaqMan: The 3'UTR of *NCS1* harbored a 7mer-m8 binding site for miR-1307-3p I1 and a 6mer binding site for miR-1307-3p I0 (Figure 18C). A 7mer-m8 binding site

requires more base pairs between the seed sequence of the microRNA and the 3'UTR of the mRNA target to be complementary than a 6mer binding site. More complementary base pairs between a microRNA and its target result in a higher binding affinity that is reflected by a stronger downregulation of the target upon microRNA overexpression[59]. The obtained results suggested that for *NCS1* the TargetScan-based predictions were more accurate.

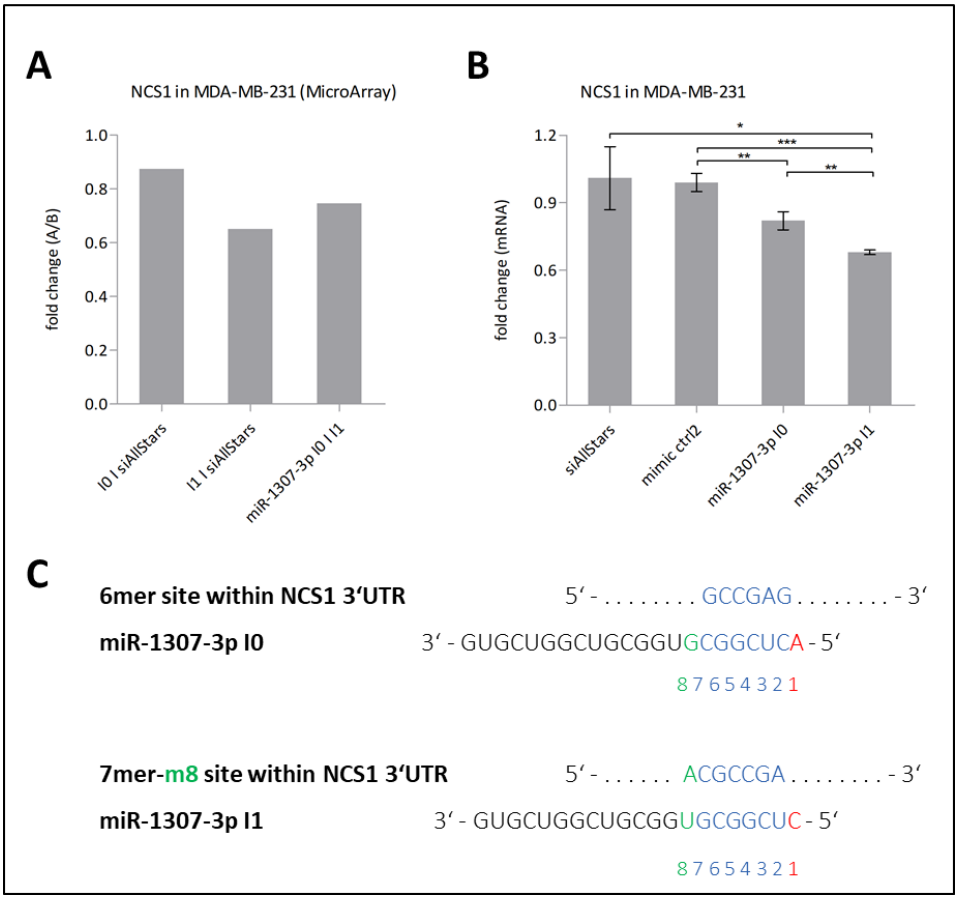


Figure 18: *NCS1* repression by miR-1307-3p I0 and miR-1307-3p I1. A) MicroArray-based expression profiling in MDA-MB-231 upon miR-1307-3p I0 and miR-1307-3p I1 overexpression. The fold changes in expression of *NCS1* between two conditions are plotted. **B)** *NCS1* mRNA levels in MDA-MB-231 were measured with TaqMan after miR-1307-3p I0 and miR-1307-3p I1 overexpression. Error bars represent the standard deviation of three biological replicates. Statistical significance was calculated with an unpaired two-tailed student's t-test. P-values are represented by asterisks (* ≤ 0.05, ** ≤ 0.01, *** ≤ 0.001). **C)** Scheme illustrating the binding sites that TargetScan predicted for miR-1307-3p I0 and miR-1307-3p I1 within the 3'UTR of *NCS1*.

4.2.9 miR-1307-3p I0 might reduce migration by targeting *LBH*

Although functional enrichment analysis based on the target predictions did not yield significant enrichment of any pathway or phenotype, I hypothesized that literature-based evaluation of the miR-1307-3p I0-specific targets in the consensus list (n=206) could identify phenotypes that might be specifically regulated by miR-1307-3p I0 and that we did not test. Moreover, I assumed that screening the literature for the miR-1307-3p I0-specific targets might identify targets that could play a role in mediating the effect that miR-1307-3p I0 had on migration of MDA-MB-231 cells. To narrow down the number of potential targets, I intersected the consensus predictions for miR-1307-3p I0 with the genes specifically downregulated by miR-1307-3p I0 in the MicroArray. The three remaining targets were *LBH*, *TOR1B* and *VPS37C*. There were no publications showing a connection between cancer and *TOR1B* or *VPS37C*. For 'Limb Bud And Heart Development' (LBH), however, literature showed an invasion-promoting role via PI3K/AKT signaling in gastric cancer cells[160]. In glioma, LBH enhanced expression and secretion of VEGF-A in hypoxic conditions, which promoted angiogenesis[161]. According to literature, downregulation of *LBH* by miR-1307-3p I0 could explain the less migratory phenotype that I observed in MDA-MB-231 and, moreover, miR-1307-3p I0 might repress angiogenesis via targeting of *LBH*. In line with the angiogenesis-promoting role of LBH, GSEA showed repression of an angiogenesis-related gene signature in Basal patients with high miR-1307-3p I0 or miR-1307-3p I1 levels (Figure 19A).

In the MicroArray, *LBH* was detected by two probes: one probe showed barely changes in *LBH* expression upon miR-1307-3p I0 or miR-1307-3p I1 transfection (Figure 19B). The other probe detected a reduction in *LBH* by 40 % and 25 % upon miR-1307-3p I0 and miR-1307-3p I1 overexpression, respectively. Although miR-1307-3p I1 did not repress *LBH* strong enough to meet the pre-defined cut-off (fold change<0.65), probe 1 showed a mild regulation. I assumed that this might result from different types of binding sites and, thus, evaluated the target predictions produced by TargetScan and MiRanda individually. While MiRanda predicted LBH as a miR-1307-3p I0-specifc target, TargetScan predicted LBH to be a shared target with an 8mer binding site for miR-1307-3p I0 and a 6mer for miR-1307-3p I1 (Figure 19C). The results obtained by probe 1 in the MicroArray supported the TargetScan predictions rather than the predictions based on MiRanda: strong repression of *LBH* by miR-1307-3p I0 and a mild reduction in *LBH* levels upon miR-1307-3p I1 overexpression. To confirm the regulation of *LBH* by miR-1307-3p I0 and miR-1307-3p I1, I performed expression analysis in an independent set of samples (MDA-MB-231). Validation of the results determined no significant changes of LBH abundance upon miR-1307-3p I0 or miR-1307-3p I1 overexpression (Figure 19D). The standard deviation between the biological replicates

of all conditions was high, which might mask a potential effect of miR-1307-3p I0 / miR-1307-3p I1. On the other hand, the TaqMan results could confirm the lack of regulation indicated by probe 2 in the MicroArray.

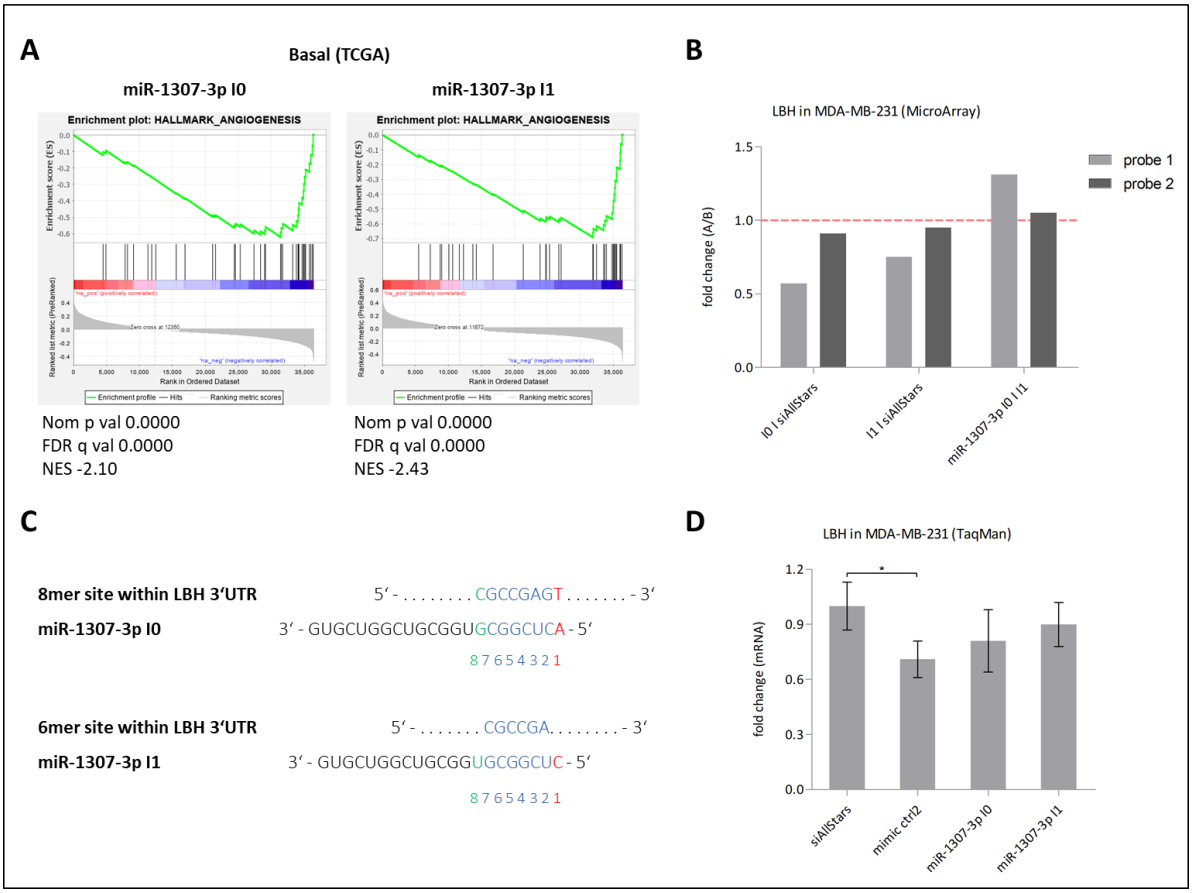
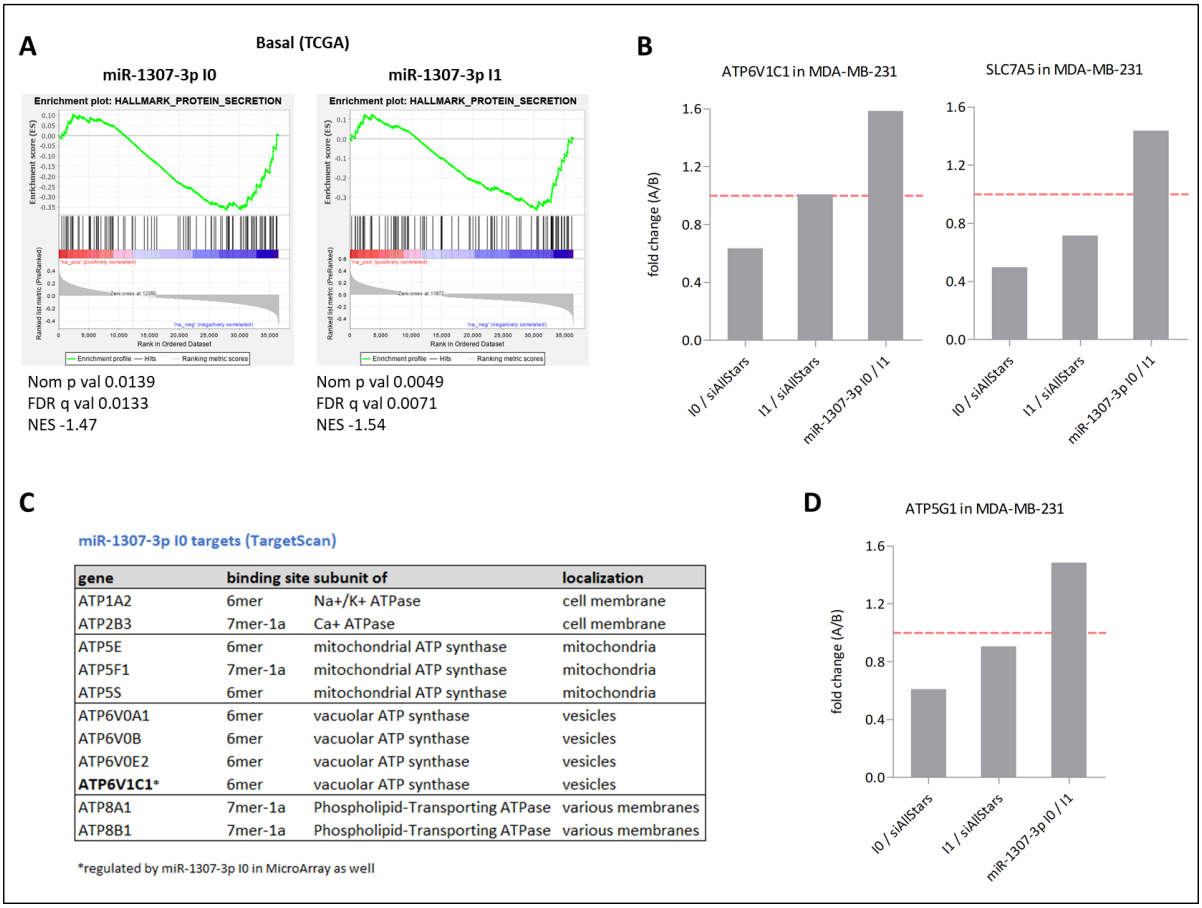


Figure 19: *LBH* repression by miR-1307-3p I0 and miR-1307-3p I1. A) GSEA of Basal patients from TCGA ranked by their miR-1307-3p I0 or miR-1307-3p I1 expression was performed and revealed significant depletion of an angiogenesis-related gene signature. NES represents the normalized enrichment score indicating negative enrichment. **B)** MicroArray-based expression profiling in MDA-MB-231 upon miR-1307-3p I0 and miR-1307-3p I1 overexpression. The fold changes in expression of *LBH* between two conditions are plotted. **C)** Scheme illustrating the binding sites that TargetScan predicted for miR-1307-3p I0 and miR-1307-3p I1 within the 3'UTR of *LBH*. **D)** *LBH* mRNA levels in MDA-MB-231 were measured with TaqMan after miR-1307-3p I0 and miR-1307-3p I1 overexpression. Error bars represent the standard deviation of three biological replicates. Statistical significance was calculated with an unpaired two-tailed student's t-test. P-values are represented by asterisks (* ≤ 0.05, ** ≤ 0.01, *** ≤ 0.001).

4.2.10 miR-1307-3p I0 targets multiple ATPase subunits and might play a role in autophagy

Overlapping the consensus target predictions for miR-1307-3p I0 with the genes specifically downregulated by miR-1307-3p I0 in the MicroArray did not suggest any phenotypes that might be specifically regulated by miR-1307-3p I0. The gene regulation that was determined by MicroArray and TaqMan experiments showed that the TargetScan predictions were more accurate for the targets *NCS1* and *LBH* than the predictions from MiRanda. Moreover, the TargetScan predictions comprised a lot more genes for miR-1307-3p I0 than predictions generated by MiRanda. Thus, I hypothesized that overlapping the TargetScan predictions for miR-1307-3p I0 with the MicroArray could provide more insight into pathways or phenotype specifically repressed by miR-1307-3p I0. However, only five genes overlapped between both lists: *ATP6V1C1*, *ELF4*, *RC3H2*, *SLC7A5* and *STK40*. Literature did not assign any cancer-related phenotypes with RC3H2. For ELF4 and STK40, a role in cancer was shown, however, both proteins were associated with phenotypes that we had tested.

The role of ATP6V1C1, a subunit of the vacuolar H+-ATPase, in cancer did not involve any potentially miR-1307-3p I0-specific phenotypes. Inhibition of ATP6V1A, another subunit of the vacuolar H+-ATPase, however, has been associated with induction of autophagy[162]. SLC7A5 is part of LAT1, a neutral amino acid transporter that was connected to an increase in autophagy upon inhibition[163]. Both, the vacuolar H+-ATPase and LAT1, play a crucial role in sensing nutrients via mTORC1 signaling at the lysosomal membrane[164]. LAT1 is recruited to the lysosomal membrane, where it induces mTORC1 signaling via the vacuolar H+-ATPase[165]. Thus, mTORC1 signaling remains active and blocks autophagy in the presence of nutrients[165]. Autophagy modulates protein secretion, the outcome, however, depends on the involved pathways[166]. As a result, protein secretion can be either increased or decreased upon induction of autophagy. GSEA had shown repression of the hallmark gene signature ‘protein secretion’ in Basal patients with high miR-1307-3p I0 or miR-1307-3p I1 levels (Figure 20A), which supported the assumption that miR-1307-3p I0 could play role in autophagy via repression of *ATP6V1C1* and *SLC7A5*, for instance. Since *ATP6V1C1* and *SLC7A5* were predicted to be miR-1307-3p I0-specific targets by TargetScan, I wanted to confirm this and looked into the MicroArray data. In line with the predictions, *ATP6V1C1* was specifically repressed by miR-1307-3p I0 (approximately 35 %) (Figure 20B). While miR-1307-3p I0 overexpression decreased *SLC7A5* levels even by 50 %, miR-1307-3p I1 downregulated *SLC7A5* as well (by almost 30 %).



gene	binding site	subunit of	localization
ATP1A2	6mer	Na+/K+ ATPase	cell membrane
ATP2B3	7mer-1a	Ca+ ATPase	cell membrane
ATP5E	6mer	mitochondrial ATP synthase	mitochondria
ATP5F1	7mer-1a	mitochondrial ATP synthase	mitochondria
ATP5S	6mer	mitochondrial ATP synthase	mitochondria
ATP6V0A1	6mer	vacuolar ATP synthase	vesicles
ATP6V0B	6mer	vacuolar ATP synthase	vesicles
ATP6V0E2	6mer	vacuolar ATP synthase	vesicles
ATP6V1C1*	6mer	vacuolar ATP synthase	vesicles
ATP8A1	7mer-1a	Phospholipid-Transporting ATPase	various membranes
ATP8B1	7mer-1a	Phospholipid-Transporting ATPase	various membranes

*regulated by miR-1307-3p I0 in MicroArray as well

Figure 20: miR-1307-3p I0 targets various ATPase subunits. A) GSEA of Basal patients from TCGA ranked by their miR-1307-3p I0 or miR-1307-3p I1 expression was performed and revealed significant depletion of a gene signature related to protein secretion. NES represents the normalized enrichment score indicating negative enrichment. **B)** MicroArray-based expression profiling in MDA-MB-231 upon miR-1307-3p I0 and miR-1307-3p I1 overexpression. The fold changes in expression of *ATP6V1C1* and *SLC7A5* between two conditions are plotted. **C)** A selection of targets that were predicted specifically for miR-1307-3p I0 with the TargetScan algorithm. **D)** MicroArray-based expression profiling in MDA-MB-231 upon miR-1307-3p I0 or miR-1307-3p I1 overexpression. The fold changes in expression of *ATP5G1* between two conditions are plotted.

Based on miR-1307-3p I0-mediated repression of *ATP6V1C1*, I hypothesized that more subunits of the vacuolar H+-ATPase could be regulated by miR-1307-3p I0. I thus screened the target predictions and the significantly downregulated genes in the MicroArray specifically for subunits of ATPase subunits. I found three more subunits for the vacuolar H+-ATPase to be predicted by TargetScan and, interestingly, subunits of several other ATPases (Figure 20C). Some of the potential miR-1307-3p I0 targets code for subunits of the mitochondrial ATP synthase, which contains a subunit that is encoded by *ATP5MD*, the host gene of pre-miR-1307. In the MicroArray, *ATP5G1*, another subunit of the mitochondrial ATP synthase was repressed by miR-1307-3p I0 overexpression (Figure 20D) but was not predicted as direct target of miR-1307-3p I0 or miR-1307-

3p I1. *ATP5G1* levels were mildly affected by miR-1307-3p I1. Overall, this suggests that miR-1307-3p I0 might target several ATPase subunits and *SLC7A5*, which might affect protein secretion and modulate autophagy.

Based on the connection of autophagy and the vacuolar H+-ATPase, which contains four subunits potentially targeted by miR-1307-3p I0, I hypothesized that miR-1307-3p I0 might target other parts of autophagy pathways as well. To test this, I checked whether the TargetScan predictions contained key players of autophagy. In fact, the TargetScan algorithm predicted many genes with a role in autophagy to be targeted by miR-1307-3p I0 (Figure S3A). Some of the targets that were predicted for miR-1307-3p I0 are involved in unfolded protein response (UPR)[167], which is not surprising considering that UPR is in crosstalk with autophagy and both processes are part of the ER stress response[168]. GSEA did not link high miR-1307-3p I0 / miR-1307-3p I1 levels in Basal breast cancer patients with a modulation of an autophagy-related gene signature, however, the analysis showed enrichment of an 'UPR' gene signature (Figure S3B). In summary, target predictions and regulated genes as well as GSEA in patients connected miR-1307-3p I0 with a potential role in autophagy and cellular stress response.

5. DISCUSSION

5.1 Mammosphere assay to identify microRNAs in chemoresistance

Patients diagnosed with the highly aggressive TNBC have limited treatment options: besides surgery and radiation, chemotherapy is the standard of care[28]. Still, TNBC patients frequently show only partial response to chemotherapy, which comes with a high risk of recurrence[29]. The wide-ranging influence of microRNAs on the cellular signaling landscape[139] renders them important hubs that guide the fate of a cancer cell. Identifying their targets and understanding their role in modulating the cellular phenotype allows to get insight into resistance development and to identify weaknesses that can be exploited for targeted therapy. In the past years our knowledge of the miRNome has gained increased complexity with the discovery and characterization of 5'isomiRs, functionally relevant sequence variants. Along this line, the aim of this project was to focus on microRNAs and 5'isomiRs differentially expressed in patients, to identify those which modulate chemoresistance of TNBC cell lines and to characterize how they modulate the cellular response to drugs.

5.1.1 Proof of principle - enrichment for chemoresistant BCSCs in mammospheres

After having generated a library of selected microRNA and 5'isomiR candidates on the pre-microRNA level, I determined the dynamic range of chemosensitivity by overexpressing a positive and a negative control (pre-miR-21 and pre-miR-145, respectively). As the dynamic range was limited when cells were grown in 2D conditions (Figure 5D), I implemented a mammosphere-based assay to enrich for stem cells, which are a main factor driving chemoresistance[146]. I could indeed show that the cultivation of mammospheres over six generations increased the mRNA levels of several BCSC markers, drug efflux pumps as well as detoxification enzymes (Figure 6B). Moreover, the sphere-derived cells were more resistant to Paclitaxel and Epirubicin than cells grown in 2D (Figure 6C). The drug response curves proved that the mammosphere assay does not only modulate markers associated with chemoresistance, but also enriches for cells that have a higher tolerance of chemotherapeutic drugs. Concluding, with the mammosphere assay I had established an experimental system that was suitable to study the role of microRNAs in chemoresistance.

While BCSC markers are commonly assessed at the protein level, the number of cells that I had left for analysis after the mammosphere screen was not sufficient to analyze individual BCSC markers by FACS and to isolate genomic DNA for detecting the enrichment of pre-microRNA

integrates by NanoString. Another factor rendering analysis at the mRNA level more suitable in my project, was the necessary use of Trypsin to dissociate the spheres to single cells prior to seeding for the next sphere generation. The protease Trypsin effectively dissociates cells attached to each other, however, this comes at the cost of degradation of the extracellular parts of surface proteins, including some of the stemness markers that I wanted to detect. Hence, the use of milder agents is required to keep the surface markers of interest intact. For cultivation of spheres over multiple generations, however, the harsh dissociation with Trypsin is necessary to obtain a single cell suspension than can be seeded for the next sphere generation. Some of my stable TNBC cell lines formed spheres that showed increased mRNA levels of the BCSC markers *ITGA6*, *PROM1*, *ITGB3* or *ALDH1A1*. For TNBC, CD44+/CD24-, ALDH1A1 and PROM1[169,170] have been described as stem cell markers. Furthermore, the transcription factors OCT4, SOX2 and Nanog have been found upregulated[171]. *NANOG* was indeed consistently upregulated in my samples, while it was not possible to design intron-spanning primers targeting *OCT4* and *SOX2* with the 'Assay Design Center' that I used for the TaqMan experiments. In my system, the CD44+/CD24- stem cell markers did not seem to be of high relevance. This is not surprising, at least for MDA-MB-231 as this cell line is comprised of more than 95 % cells that are CD44+/CD24-[172]. Additionally, the variety of BCSC markers described in the literature is even higher, and the plasticity of stem cells is influenced by multiple extrinsic and intrinsic factors[173]. This implies that stemness needs to be assessed at the functional level. Functional assays focus on quantification of sphere number or size *in vitro* or evaluate tumorigenicity and self-renewal *in vivo*[174].

Based on the project aim to study the impact of the selected pre-microRNAs on chemoresistance, I used the cultivation of the stable cell lines in 3D to enrich for potentially chemoresistant stem cells. To evaluate chemoresistance on the functional level, I compared the drug response of cells grown in 2D versus 3D. For the chemoresistance assay, cells from both conditions were cultivated in 2D to compare their response to the drugs in the same condition. As assumed, the mammospheres enriched for cells that tolerate higher drug concentrations. Nevertheless, it is possible that the highly concentrated growth factors EGF and FGF in the mammosphere medium conferred a strong growth advantage compared to 2D culture medium. Consequently, the bias generated by the growth factors in the sphere medium might promote a higher drug tolerance. The cells grown in 3D were washed with PBS, dissociated and then grown in 2D cell culture medium containing the drugs for 96 hours. This procedure should remove remaining growth factors from the sphere cultivation and strip the cells derived from 3D culture of any growth advantage coming from these factors. Assuming that the growth factors from the sphere medium partially remained, it is difficult to estimate how the effect compares to the effect of the growth

factors in the FBS that the 2D cell culture received. It is known that FBS contains various growth factors as well, for instance, EGF, FGF, Insulin and, most abundantly, PDGF[175]. The growth factor concentrations in FBS, however, are not defined as in case of the sphere medium. Ideally, one should determine whether the chemoresistance of cells grown in 3D is a long-term effect that can be still observed after several weeks in 2D conditions. It has been shown, however, that MCF7 spheres cultivated for five generations displayed a stable phenotype when cultured in 2D for an extended period[176]. This included increased chemoresistance, motility and *in vivo* tumorigenicity. Mechanistically, the stable phenotype of mammosphere-derived MCF7 cells in this study was accompanied by an EMT rendering the cells more mesenchymal. EMT was shown by evaluating the migratory potential as well as detection of EMT markers on the mRNA and protein level. In my study, I used the TNBC cell lines MDA-MB-231 and SUM-159 as experimental system. These cell lines already have a mesenchymal phenotype and are characterized by the expression of mesenchymal markers[177]. As a consequence, mammosphere culture might not further enhance EMT as strongly as it had been observed for MCF7 cells, which are classified as epithelial cell line[178]. Nevertheless, it makes sense to test this in particular for HCC1806, which has the weakest migration phenotype of the three TNBC cell lines that I used and is known to lack expression of epithelial and mesenchymal markers[177]. However, I employed 3D culture to enrich for chemoresistant cells, which is not necessarily connected to a more mesenchymal phenotype.

5.1.2 Enrichment of pre-miR-103a-1 in spheres is not reflected on the functional level

In mammospheres grown from TNBC cell lines harboring the library, several pre-microRNAs were strongly enriched compared to cells overexpressing the library and grown in 2D (Figure 8). However, the enrichment in 3D was limited to one replicate except for pre-miR-103a-1. I thus selected pre-miR-103a-1 for functional validation as this was supported by a strong fold-change in 3D compared to 2D in replicates from all three tested TNBC cell lines (Figure 8B). Interestingly, pre-miR-103a-2 did not show enrichment in any of the 3D samples, although both pre-microRNAs code for miR-103a-3p (Figure 21). The difference in enrichment between pre-miR-103a-1 and pre-miR-103a-2 could have been the result of different functions mediated by the microRNAs located on the 5p arm (Figure 21): miR-103a-1-5p and miR-103a-2-5p, respectively. miR-103a-2-5p could have mediated an effect that counteracts the function of miR-103a-3p, which then prevented enrichment of pre-miR-103a-2 in mammospheres. Another possible explanation for the strong enrichment of pre-miR-103a-1 is that the enrichment occurred due to a sphere-promoting effect of miR-103a-1-5p rather than miR-103a-3p. However, there are no studies supporting this idea.

Consequently, expression analysis as well as functional analysis of the microRNAs derived from the 5p arms of pre-miR-103a-1 and pre-miR-103a-2 is required to prove that this could have been the reason for the strong pre-miR-103a-1 enrichment.

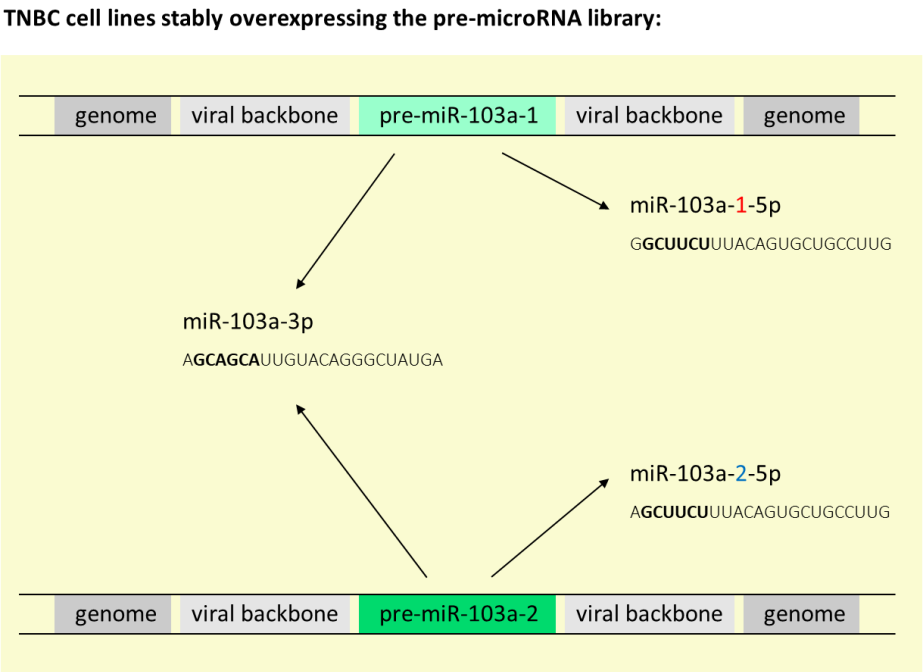


Figure 21: Scheme illustrating the possible microRNAs processed from pre-miR-103a. Pre-miR-103a-1 and pre-miR-103a-2 were included in my pre-microRNA library and were retrovirally transduced into the genome of different TNBC cell lines along with the other pre-microRNAs selected for the library. miR-103a-3p can be processed from either pre-miR-103a-1 or pre-miR-103a-2. Both pre-microRNAs, however, give rise to different microRNAs on the 5p arm: miR-103a-1-5p and miR-103a-2-5p.

Considering the variation in abundance of many pre-microRNA barcodes when comparing different 3D samples as well as different time points (Figure 8), however, the NanoString data rather suggests a random enrichment of pre-microRNA integrates. This notion was supported by the fact that I could not show an increase in sphere formation or chemoresistance upon pre-miR-103a-1 overexpression (Figure 9B). To confirm that pre-miR-103a-1 did not enrich in 3D due to high miR-103a-3p or miR-103a-5p levels that promote chemoresistance, the respective expression levels in the mammospheres should have been determined. Since pre-miR-103a-1 and pre-miR-103a-2 had been selected for the library based on the differential expression of miR-103a-3p I0 and miR-103a-3p I1, however, I quantified only pre-miR-103a and miR-103a-3p levels in my samples (Figure 11A and B). The comparison of 3D samples to 2D samples showed that pre-miR-103a and miR-103a-3p levels were similar in both conditions, partially even lower in 3D than in

2D. The low pre-miR-103a-1 abundance in 3D compared to 2D culture confirms that the enrichment of cells with a pre-miR-103a-1 barcode was not the result of a survival benefit mediated by pre-miR-103a-1 in 3D conditions. The enriched cells might have expressed genes that were involved in stemness and the pre-miR-103a-1 integrate could have served as a neutral barcode. Moreover, there is only one study that links miR-103a-3p to chemoresistance in lung cancer[179] and thereby supports the assumption that pre-miR-103a-1 does not give rise to mature microRNAs that play a major role in chemoresistance.

Based on the previously described findings, I decided not to continue with the project. As a consequence, I did not quantify endogenous expression or overexpression of pre-miR-103a, miR-103a-5p or miR-103a-3p in the stable cell lines. We considered sequencing the pre-microRNA library cell lines after the stable cell lines were generated to obtain the level of overexpression for the individual microRNAs and 5'isomiRs. Taking into account, however, that only every 72nd cell might harbor the same pre-microRNA integrate and breaking this down to the microRNA variants transcribed from each pre-microRNA, the sequencing depth would have not been sufficient. In the context of another project, the same retroviral backbone that I used for my pre-microRNA library was used to generate various stable cell lines overexpressing single pre-microRNAs. MicroRNA sequencing data of these cell lines confirmed that the magnitude of overexpression across different pre-microRNAs differed strongly between pre-microRNA constructs as well as between 5'isomiRs processed from these pre-microRNAs (Figure 11C), and most likely introduced a strong bias. The pre-miR-21 construct, for instance, lacked overexpression, but the 5'isomiRs produced from pre-miR-145 were increased 100-fold to 1000-fold. Looking at the endogenous expression levels, this is not very surprising: pre-miR-21 derived 5'isomiRs have already very high expression levels, while the 5'isomiRs processed from pre-miR-145 are endogenously expressed at low levels (Figure 11D). Since miR-21 is known to mediate chemoresistance[142,143], the lack of overexpression might have abolished a selection advantage for these cells compared to the control (Figure 5D). Moreover, the increase in expression for 5'isomiRs processed from the same pre-microRNA was not proportional for pre-miR-145, pre-miR-183 and pre-miR-29a. Considering that some of the generated 5'isomiRs might differ substantially from each other in their function and impact, the bias created by the overexpression might favor isoforms that are less relevant for chemoresistance. An increase in a particular microRNA without influence on chemoresistance would dilute the effect of other 5'isomiRs with a role in chemoresistance and might suppress enrichment of cells with the respective pre-microRNAs integrate in mammospheres.

The initial decision to work on the pre-microRNA levels was based on the advantage to use the pre-microRNAs as barcodes and to allow physiological biogenesis of microRNAs and 5'isomiRs. Previous cell line sequencing data generated in the lab showed that the relative microRNA and 5'isomiR levels in several breast cancer cell lines are similar to those in TCGA patient data suggesting conservation of the ratios, at least to some extent. Based on this, we assumed that pre-microRNA overexpression would maintain the ratios as well. Although overexpression constructs are an ideal tool in many cases, the situation turned out to be more difficult for 5'isomiRs than we anticipated considering the diversity in 5'isomiR overexpression for different pre-microRNA constructs.

With the knowledge of today, I would choose a completely different experimental design to study microRNAs in chemoresistance. I would use the parental TNBC cell lines and generate chemoresistant cell lines by constant drug exposure until they have acquired full resistance. Throughout the resistance development, I would harvest samples to profile the changes in microRNA and 5'isomiR expression shortly after start of the treatment and then roughly every three months. Using microRNA sequencing, I would then identify microRNAs that are potentially relevant for chemoresistance directly on the 5'isomiR level. Subsequently, I could validate the specific isoform without the need to go through the pre-microRNA level first and, moreover, could directly exclude that the effect comes from the microRNAs located on the other pre-microRNA arm. Another setup for future experiments in this direction could be to sequence the mammospheres and compare their endogenous microRNA and 5'isomir expression level to the input and to time-matched samples grown in 2D. In addition, sequencing-based approaches would have allowed profiling changes in the ratios of different 5'isomiRs to each other. This is of particular interest in cases where one isoform modulates chemoresistance, whereas other isoforms target different phenotypes. Overall, microRNA sequencing would have been the more straight-forward approach for any experimental setup. In theory, 5'isomiR-specific qRT-PCR assays would have been another strategy, however, so far they are not commercially available. Reproducing published methods for 5'isomiR-specific detection[98,99] and establishing our own 5'isomiR-specific methods was not successful. Even when 5'isomiR-specific quantification methods can be established, specific probes for a large number of 5'isomiRs are expensive and the development of specific assays is time-consuming and needs to be optimized for each microRNA and its 5'isomiRs.

5.1.3 Selection bias of microRNAs for the pre-microRNA library due to TCGA batch effects

Recently, Cindy Körner uncovered that the microRNA expression from the TCGA patient cohort was biased by strong batch effects. To understand the impact the selection bias had on my study, I compared the lists of differentially expressed microRNAs before and after batch correction of the TCGA data (Figure S1). Although there was quite some overlap between the pre-microRNAs, the biased selection missed many potentially relevant pre-microRNAs. Several of the pre-microRNAs that should have been included in the library based on the selection criteria I had defined in the beginning, give rise to mature microRNAs that have been connected to chemoresistance in various cancer entities: miR-1, miR-135b, miR-15b, miR-153, miR-184, miR-195, miR-30a, miR-330, miR-338, miR-429, miR-450b, miR-584, miR-193a, miR-19a, miR-223, miR-744, miR-9 and miR-96[180–198]. Other candidates that were missed because of the batch effects, modulate chemoresistance in breast cancer: miR-149, miR-200b, miR-205, miR-708, miR-30c and miR-423[199–204].

Some microRNA candidates derived from pre-microRNAs that should have been part of the library were shown to modulate stemness and, therefore, might influence chemoresistance as well. So far, stemness-modulating roles in breast cancer or other cancer entities have been shown for miR-1, miR-153, miR-195, miR-205, miR-450b and miR-708[185,190,205–208]. In many cases, including some of the referenced studies, stemness has been evaluated based on stemness markers rather than in functional assays. Different studies describing breast cancer stem cells with different markers emphasizes the dilemma evolving around the question which markers define stem cell populations. The enrichment of different BSCS markers in different studies is not surprising considering that even the same conditions cause variable enrichment of markers (Figure 6B). In this project I showed the heterogeneity of BCSC markers between cell lines, but also within cell line replicates. My findings underline how heterogeneous and dynamic the situation around BCSC markers is and that the reliability of markers seems to be very context-dependent. The homogenous upregulation of the transcription factor *NANOG* in my spheres, however, implies that evaluating stemness on the basis of transcription factors might be more reliable. The relevance of transcription factors such as OCT4, SOX2, KLF4 or NANOG as key players in stemness programs is widely accepted[209]. The so-called Yamanaka factors include OCT4, SOX2 and KLF4 and are sufficient for reprogramming terminally differentiated cells to induced pluripotent progenitor cells[210]. Nevertheless, stem cell plasticity underlies intrinsic and extrinsic stimuli[211], which may lead to a strong variation of BCSC markers as well as differences in their tumorigenic behavior between patients[212]. Overall, BCSCs as well as the cellular miRNome and its target spectrum are

highly complex and dynamic, which makes it difficult to enrich for the same microRNAs under different experimental conditions.

Whether the microRNAs that were missed by the TCGA batch effects would have played a role in my system is something that I can only speculate about. Those microRNAs that have been shown to modulate response to chemotherapy in other cancer entities than breast cancer might repress targets that are not expressed in breast cancer. Second, microRNAs can have different interaction partners in different cancer entities or under different circumstances, depending on which pathways are active and which genes are amplified or mutated. Variation in pathway activity and target availability renders cancer entities vulnerable to different therapeutic compounds. Thus, it is also not surprising that many studies about the phenotypic impact of microRNAs are conflicting. miR-181c, for instance, has been described to reduce chemoresistance in breast cancer[213], but to rather promote resistance in pancreatic cancer[214].

Looking at the pre-microRNAs that were included in the library and that remained in the selection after correction of the batch effects, reveals that my system failed to identify known modulators of chemoresistance. miR-139, miR-141, miR-200a, miR-29a, miR-381 have been shown to influence chemoresistance in breast cancer[215–219]. For these microRNAs, the lack of enrichment or depletion in my system could be explained by the assumption that the mammosphere assay enriched for microRNAs that modulate stem cell features, but not necessarily drug response. miR-142 and miR-200c, however, suppress stemness of breast cancer cells[220,221] and chemoresistance in the case of miR-200c[222]. The study on miR-142, for instance, showed modulation of the BCSC markers PROM1 and ALDH1, which were also among the regulated BCSC markers in my system. Based on this, a depletion of pre-miR-142 in my assay could be expected and low overexpression levels of the stable cells might explain the lack of enrichment in my system. Although overexpression for pre-miR-142 was not determined yet, we saw basically no overexpression for pre-miR-21, pre-miR-27a and pre-miR-320a (Figure 16C). Considering that pre-miR-145 had no effect on drug response in my hands (Figure 5D) despite strong overexpression (Figure 16C) and a published role in chemoresistance[144,145], suggests that the basal expression level matter. Endogenously, 5'isomiRs of miR-145 are almost not expressed and despite approximately 800-fold overexpression of miR-145-3p I3, the expression levels are still far below those of 5'isomiRs transcribed from pre-miR-21 and pre-miR-27a, pre-miR-29a and pre-miR-320a (Figure 16D). Based on this observation, it could make sense to focus only on microRNAs and 5'isomiRs that reach a certain expression level after upregulation and, therefore, are likely of higher physiological relevance.

Several of the pre-microRNAs that would have been excluded from the library after batch correction of the TCGA data, give rise to microRNAs associated with chemoresistance or stemness as well. miR-125b, miR-130b, miR-155, miR-16, miR-17, miR-29b-1 and miR-7[171,223–228] have been implicated in chemoresistance and partially in stemness of breast cancer or other entities. These examples show that my stringent selection of candidates for the pre-microRNA library is not the only approach to obtain potential modulators of chemoresistance and stemness. Different criteria might create a list of valid microRNA candidates as well. To study any of the discussed microRNAs, however, strong pre-microRNA overexpression is crucial to observe a potential selection advantage mediated by the respective pre-microRNA. The presented data, however, shows that the overexpression constructs used in this project failed to consistently overexpress pre-microRNAs and 5'isomiRs across different constructs and at sufficient levels. These unexpected findings suggest that future approaches to study the role of microRNAs in chemoresistance should focus on modulation of endogenous microRNAs upon treatment with chemotherapeutic drugs or at least ensure defined amounts of all studied microRNAs.

5.2 Divergent 5'isomiR miR-1307-3p I1 promotes a different phenotype than miR-1307-3p I0

For identification of microRNAs and 5'isomiRs playing a role in chemoresistance of TNBC, potentially chemoresistant BCSCs were enriched with a mammosphere assay. Profiling of the spheres with NanoString revealed pre-miR-103a-1 as strongly enriched candidate. Since I could not validate the role of miR-103a-3p in chemoresistance or stemness, the project was no longer pursued. In order to study the functional relevance of 5'isomiRs and characterize phenotypic and mechanistic differences, a 5'isomiR pair was selected from microRNAs differentially expressed in breast cancer patients. miR-1307-3p I0 and miR-1307-3p I1 were selected as 5'isomiR pair since both were differentially expressed in tumor compared to normal tissue of TCGA patients, expressed at similar levels in tumor or in normal tissue and had a divergent seed sequence. This project focused on a divergent 5'isomiR pair, because this subset of 5'isomiRs is known to differ largely in their target spectra and the resulting phenotypes[95,96]. As a result, this allows confirming the impact a seed sequence shifted by one nucleotide can have.

5.2.1 pre-miR-1307 reduces migration and invasion in TNBC

With a phenotypic screen, Xiaoya Li was able to identify migration and invasion as phenotypes repressed by pre-miR-1307 (Figure 13A). Moreover, pre-miR-1307 overexpression reduced tumor volume *in vivo* without influencing proliferation *in vitro* (Figure 13B). For miR-1307-3p I0, a few phenotypic studies have been published. Those publications studying the role of miR-1307-3p I0, however, are conflicting: while a tumor-suppressive role was described in colon cancer by one study [229], thus supporting our data, other studies have described an oncogenic impact of miR-1307-3p I0 in breast cancer as well as in other cancer entities[230–234]. The study in breast cancer showed that overexpression of miR-1307-3p I0 in MCF10A cells promoted proliferation, growth in soft agar and tumor formation in mice. The oncogenic effect in breast cancer is in line with the upregulation of miR-1307-3p I0 and miR-1307-3p I1 in tumor tissue, especially in tumors of aggressive subtypes. The study in colon cancer that assigned a tumor-suppressive role to miR-1307-3p I0 on the other hand, described exactly the opposite phenotype: reduced proliferation via G1 arrest and apoptosis[229]. These contradicting results could partially result from different mutation spectra and pathway activities present in the two cancer entities. With regard to EMT, miR-1307-3p I0 has been shown to promote migration and invasion in liver cancer cell lines[231,232]. To date, there are no studies addressing the functional role of miR-1307-3p I1. In conclusion,

literature assigns an oncogenic role to miR-1307-3p I0, but Xiaoya Li identified migration and invasion as phenotypes repressed by pre-miR-1307. Thus, migration was selected as phenotype to study the potentially different impact of miR-1307-3p I0 and miR-1307-3p I1.

5.2.2 miR-1307-3p I0 and I1 play divergent functional roles in a cell line-specific manner

To assess whether the migration phenotype in MDA-MB-231 cells was induced by miR-1307-3p I0 or miR-1307-3p I1 specifically, I next transfected 5'isomiR-specific mimics into MDA-MB-231 and SUM-159 cells to evaluate their respective impact on migration (Figure 14A). miR-1307-3p I1 repressed migration in both tested TNBC cell lines, while miR-1307-3p I0 reduced migration only in MDA-MB-231. Furthermore, the effect of miR-1307-3p I1 on migration in MDA-MB-231 was stronger compared to miR-1307-3p I0. The impact of both microRNAs on migration in MDA-MB-231 was supported by EMT marker changes on mRNA level: both isoforms significantly increased CDH1 and decreased *VIM* levels in MDA-MB-231 (Figure 14B), whereas *CDH1* was not expressed and *VIM* was not significantly regulated in SUM-159 (data not shown). In addition, the role of miR-1307-3p I0 and miR-1307-3p I1 in migration was recapitulated in Basal breast cancer patients: patients with high miR-1307-3p I0 / miR-1307-3p I1 levels were significantly depleted in an EMT gene signature, which is perfectly in line with my *in vitro* findings (Figure 14C). From the patient data it is not possible to specify which 5'isomiR plays the major role for the migration phenotype since both isoforms are highly correlated (Figure 14D). Moreover, miR-1307-3p I1 strongly repressed proliferation, but only in SUM-159 cells (Figure 14E).

The divergent modulation of phenotypes by miR-1307-3p I0 and miR-1307-3p I1 in a cell line-specific manner might originate from differences in mutation status, the availability of targets and/or pathway activities in MDA-MB-231 and SUM-159. Looking at differentially mutated genes, MDA-MB-231 cells are characterized by mutations in *BRAF*, *CDKN2A*, *KRAS* and *TP53*, while SUM-159 cells harbor mutations in *HRAS*, *PIK3CA* and *TP53* (information obtained from Cellosaurus). Although mutation of *KRAS* versus *HRAS* or different *TP53* mutations may also differ strongly in their impact on the cellular phenotype, the *PIK3CA* mutation stands out. PIK3CA, the α-catalytic subunit of PI3K, is constitutively active in SUM-159[235]. Knock-down experiments have shown that targeting PI3K and AKT strongly affects the growth of SUM-159 compared to MDA-MB-231 cells[236]. In breast cancer in general, activating *PIK3CA* mutations have been shown to promote proliferation[237]. These findings could explain the exclusive influence of miR-1307-3p I1 on proliferation in SUM-159, hypothesizing that the proliferation phenotype is mediated by a 5'isomiR target required for PI3K signaling. If this assumption holds true, this would mean that

miR-1307-3p I1 reduces migration and proliferation via different targets and mechanisms since the 5'isomiR affects migration in both TNBC cell lines. A cell line-specific pathway addiction or weakness could also be the reason for the effect of miR-1307-3p I0 on migration in MDA-MB-231 only. Without identifying and validating the respective targets, however, this is only speculation. Since the list of targets per microRNA is usually rather long, it would make more sense to use PIK3CA or mTOR inhibitors and then confirm that miR-1307-3p I1 overexpression affects the proliferation of SUM-159 cells less, for instance.

Overall, the results show that miR-1307-3p I0 and miR-1307-3p I1 differ in their phenotype as expected based on their divergent seed sequences. Despite the context-depending influence of both 5'isomiRs, I could show that miR-1307-3p I1 affects the phenotype of a cell differently than miR-1307-3p I0. The presented data highlights miR-1307-3p I1 as another example of a 5'isomiR with putative functional relevance.

5.2.3 Divergent target spectra provide target genes that might explain the phenotypic and mechanistic differences mediated by miR-1307-3p I0 and I1

The effect of miR-1307-3p I0 and miR-1307-3p I1 seems to be highly cell line-specific, at least for MDA-MB-231 and SUM-159. To identify regulated target genes that might mediate the phenotypic effects and could explain mechanistic differences, a MicroArray was performed by Neşe Erdem Borgoni and target predictions were generated by Shashwat Sahay. The number of potential targets differed a lot between the different prediction algorithms (Table 11). Moreover, the MicroArray data and the predictions showed that miR-1307-3p I0 and miR-1307-3p I1 share subsets of genes that they potentially modulate, but both 5'isomiRs seem to have specific target subsets as well.

Selecting promising targets from the generated data was complicated by several facts: First, the MicroArray data needed to be interpreted with caution since the p values suggested that differences between conditions might represent artefacts and fold changes might be random. Second, miR-1307-3p I0 and miR-1307-3p I1 both repressed migration in a cell line-specific manner, which might depend on differential target gene expression or mutational status (e.g. *PIK3CA*). Third, microRNAs are known to repress a large number of targets[39] and, thus, it is unlikely that the observed phenotypes were the result of only one repressed mRNA target per 5'isomiR. Fourth, targets shared by both isoforms might play a role in the modulated phenotypes as well. In combination with 5'isomiR-specific targets acting on the same phenotype, this generates a highly

complex situation. Ideally, predicted targets would be narrowed down by sequencing MDA-MB-231 and SUM-159 after miR-1307-3p I0 or miR-1307-3p I1 overexpression. Differentially expressed genes would allow verifying 5'isomiR-specific targets as well as cell line-specific differences. A more advanced approach that allows focusing on directly regulated targets, would be to transfect the cells with biotinylated microRNA mimics and pull down the microRNAs as well as the mRNA targets bound to them with streptavidin-coated beads. Subsequently, sequencing would identify the regulated targets. Timewise, the mentioned experiments were not possible any more in the context of this project, therefore, I performed functional enrichment analysis for the 5'isomiR-specific target predictions with WebGestalt to identify differentially modulated pathways. This analysis, however, did not yield any pathway enrichment with statistical significance.

miR-1307-3p I1 might reduce migration and proliferation by targeting the PI3K pathway via *NCS1*

The target predictions that were generated for miR-1307-3p I0 and miR-1307-3p I1 were based on two predictions algorithms: TargetScan and MiRanda. To limit the number of false positive predictions, both target lists were intersected and I used the consensus list to identify miR-1307-3p I1-specific targets that could explain the migration phenotype in MDA-MB-231 and SUM-159 and the effect on proliferation in SUM-159. Literature search for the predicted targets showed that the potential target *NCS1* might be highly relevant for the miR-1307-3p I1-mediated phenotypes: NCS1 has been shown to promote migration and proliferation via AKT signaling[159]. Moreover, literature showed that NCS1 is involved in the PI3K/AKT pathway also in the context of neurite sprouting, where *NCS1* overexpression enhanced AKT1 activity as well[238]. The connection of the potential miR-1307-3p I1 target NCS1 with PI3K/AKT signaling might provide an explanation for the cell line-specific proliferation phenotype that might be associated with a PIK3CA mutation in SUM-159 cells and that renders PI3K signaling in these cells constitutively active. Based on the assumption that SUM-159 strongly depend on the PI3K/AKT pathway for proliferation, inhibition of PI3K/AKT signaling could confirm that miR-1307-3p I1 overexpression or *NCS1* knock-down would affect proliferation less.

The MicroArray data as well as TaqMan-based validation of gene regulation by miR-1307-3p I1 showed that for *NCS1* the TargetScan prediction was more accurate than the predictions generated by MiRanda. MiRanda predicted *NCS1* to be a miR-1307-3p I1-specific target, while TargetScan detected *NCS1* as shared target due to a 7mer-m8 binding site for miR-1307-3p I1 and a 6mer binding site for miR-1307-3p I0 in the 3'UTR. The obtained results confirm the complexity of microRNA/target interactions and that they are rarely 'black' or 'white', but rather occur in a

dynamic manner. Whether the TargetScan predictions are more reliable in general, however, needs to be determined on a larger scale. To confirm NCS1 as miR-1307-3p I1 target, I would first knock-down *NCS1* to evaluate whether this mimics the effect of miR-1307-3p I1 overexpression on migration and proliferation. Next, I would show direct binding of the *NCS1* 3'UTR by miR-1307-3p I1 with luciferase reporter assays. This approach would also allow understanding whether *NCS1* levels are repressed by miR-1307-3p I0 as well and whether this is the result of a direct or an indirect effect.

miR-1307-3p I0 might reduce migration and angiogenesis by targeting *LBH*

I identified *LBH* as a potential miR-1307-3p I0-specific target by overlapping the consensus predictions generated from TargetScan and MiRanda with the significantly downregulated genes in the MicroArray. Literature showed that LBH promotes invasion of gastric cancer cells by modulating PI3K/AKT signaling[160]. Moreover, LBH promotes expression and secretion of VEGF-A and thereby increases angiogenesis[161]. Assuming that miR-1307-3p I0 directly targets *LBH*, the role of LBH could explain the migration phenotype that I observed in MDA-MB-231. Modulation of angiogenesis-stimulating factors via LBH repression could also explain our *in vivo* results (Figure 13B) since a less pronounced vasculature is known to limit tumor growth[239]. The smaller tumor volume in mice injected with MDA-MB-231 overexpressing pre-miR-1307 could result from reduced angiogenesis since at least *in vitro* proliferation of MDA-MB-231 was not affected, which would be in line with the depletion of the hallmark gene set 'angiogenesis' in Basal patients with high miR-1307-3p I0 levels (Figure 19A).

The MicroArray results differed a lot for different probes and ranged from 10 % to 40 % repression of LBH by miR-1307-3p I0 (Figure 19B). The TaqMan-based validation was inconclusive, mainly due to the high standard deviation of all conditions (Figure 19D). Based on the gene regulation detected by probe 2 in the MicroArray and the TargetScan predictions, however, *LBH* might be a shared target of miR-1307-3p I0 and miR-1307-3p I1. Repeating the TaqMan-based validation with another set of samples including four biological replicates might reveal what the actual situation is. Confirmation of *LBH* as miR-1307-3p I0 / miR-1307-3p I1 target is required before following up with phenocopy experiments and luciferase reporter assays.

In case *LBH* can be validated as miR-1307-3p I0 / miR-1307-3p I1 target, the following step would be to evaluate migration behavior of MDA-MB-231 and SUM-159 upon *LBH* knock-down. To prove that a miR-1307-3p I0 / miR-1307-3p I1-mediated *LBH* repression might have an impact on

angiogenesis via VEGF-A, I would determine *VEGF-A* mRNA levels by TaqMan and the amount of secreted protein by ELISA. Profiling the cell culture supernatant by mass spectrometry might allow to identify more factors that influence angiogenesis and are modulated by miR-1307-3p I0 / miR-1307-3p I1. In general, profiling changes might be especially relevant since GSEA identified that high levels of miR-1307-3p I0 / miR-1307-3p I1 in Basal breast cancer patients correspond with depletion of a gene set related to protein secretion (Figure 20A). LBH-mediated modulation of VEGF-A signaling might not only influence angiogenesis in a paracrine manner, but could also have an impact on migratory potential of tumor cells since VEGF-A has acts in an autocrine manner and thereby promote survival and migration of breast cancer cells[240]. Moreover, autocrine VEGF-A signaling in breast cancer seems to constitutively activate PI3K signaling[241]. Keeping in mind that the putative miR-1307-3p I1 target *NCS1* probably acts via PI3K signaling as well, the PI3K/AKT pathway could be targeted by miR-1307-3p I0 / miR-1307-3p I1 through multiple angles. In conclusion, screening the target predictions for genes that have been implicated in PI3K/AKT signaling might provide more insight into the mechanism of miR-1307-3p I0 / miR-1307-3p I1.

Targeting of various ATPase subunits might connect miR-1307-3p I0 to an autophagy phenotype

I could show that migration and proliferation are repressed by miR-1307-3p I1 and that migration is repressed by miR-1307-3p I0 to some extent. However, the phenotypic assays I performed as well as overlapping the consensus target predictions with downregulated genes did not suggest any pathways or phenotypes exclusively modulated by miR-1307-3p I0. Since the TargetScan predictions were more accurate for *NCS1* and *LBH* when compared with the extent of gene regulation, I decided to overlap the miR-1307-3p I0-specific TargetScan predictions with the significantly downregulated genes detected by the MicroArray. I could identify two targets that have been connected with autophagy to some extent: *ATP6V1C1* as part of the vacuolar H+-ATPase and *SLC7A5* as part of the amino acid transporter LAT1[163,242]. In the MicroArray, *ATP6V1C1* was specifically regulated by miR-1307-3p I0, whereas *SLC7A5* was repressed by both 5'isomiRs (Figure 20B). The next step would be to confirm the gene regulation by TaqMan. Interestingly, literature connected *ATP6V1C1* not only with autophagy, but also with a migratory phenotype in breast cancer cell lines [243–245]. Considering the connection of autophagy and EMT in cancer cells, however, this is not surprising: autophagy can provide nutrients to fuel the migratory behavior of cancer cells[246].

To obtain more insight into the potential role of miR-1307-3p I0 in regulating the vacuolar ATPase and autophagy, I screened the literature for the TargetScan predictions and the genes regulated

in the MicroArray separately. The TargetScan predictions contained more subunits of the vacuolar H+-ATPase and other ATPases, including the mitochondrial ATP synthase (Figure 20C). The MicroArray detected significant repression of another subunit of the mitochondrial ATP synthase: *ATP5G1* (Figure 20D). Furthermore, the TargetScan predictions included a multitude of genes that have been linked with autophagy and UPR (Figure S3A), which are processes that are connected by crosstalk in the context of the cellular stress response[168].

Overall, the putative miR-1307-3p I0-specific targets including subunits of various ATPases, genes connected to autophagy/UPR and the migration/angiogenesis-modulator *LBH* were supported by GSEA: Basal breast cancer patients with high miR-1307-3p I0 levels were depleted in gene signatures related to angiogenesis (Figure 19A) and protein secretion gene set (Figure 20A) and showed enrichment for a UPR-related gene signature (Figure S3B). Protein secretion of particular factors like VEGF-A, IL-8, FGF-2 and MMPs is necessary to promote angiogenesis- and migration-related phenotypes in breast cancer[247]. Protein secretion, however, has also been linked to ATPases. In particular, the vacuolar ATPase as part of secretory vesicles filled with pro-invasive growth factors plays an important role in breast cancer aggressiveness[248]. This connection suggests that the ATPase subunits and autophagy genes potentially targeted by miR-1307-3p I0 could be associated with a reduction in protein secretion of migration-promoting and angiogenic factors that might be connected to autophagy.

To obtain a more general understanding of the potential role that miR-1307-3p I0 plays in migration, protein secretion and autophagy, I would collect the supernatant of miR-1307-3p I0-overexpressing cells and apply it to untransfected cells. Subsequently, I would evaluate a potential reduction in migration upon treatment with the supernatant. Once this has been shown, I would profile the secreted proteins by mass spectrometry in order to overlap the regulated proteins with the targets predicted for miR-1307-3p I0. The mass spectrometry could also assist in identifying proteins that are indirectly regulated by the microRNA, but play a crucial role for the miR-1307-3p I0-mediated phenotypes as well. After identifying secreted factors of interest, the autocrine function on migration needs to be shown experimentally. Moreover, the effect of miR-1307-3p I0 overexpression on autophagy needs to be assessed, for instance, by quantification of autophagosomes using immunofluorescence to stain for LC3 puncta[249].

5.2.4 Upregulation of a tumor-suppressive microRNA in breast cancer

All of the performed phenotypic assays revealed a likely tumor-suppressive role upon pre-miR-1307 as well as miR-1307-3p I0 / miR-1307-3p I1 overexpression (Figure 13 and 14). This raises the question why a tumor-suppressive (pre-) microRNA is highly abundant in tumor tissue, especially in tumors associated with aggressive subtypes (Figure 12). In TCGA breast cancer patients only three microRNAs processed from pre-miR-1307 were expressed (mean > 15 rpm): miR-1307-3p I0, miR-1307-3p I1 and miR-1307-5p I0. miR-1307-5p I0 was well correlated with miR-1307-3p I0 / miR-1307-3p I1 (Figure 17A), was significantly differentially expressed between tumor and normal breast tissue (Figure 17B) and revealed higher levels in the more aggressive subtypes (Figure 17C). The fold change between tumor and normal tissue was higher for miR-1307-5p I0 expression compared to miR-1307-3p I0 / miR-1307-3p I1 (Figure 17D), which might point in the direction that miR-1307-5p I0 has an oncogenic role and might explain the high levels of miR-1307-3p I0 and miR-1307-3p I1. Pre-miR-1307 overexpression in MDA-MB-231, however, did not mediate an oncogenic effect in our phenotypic screen. This could be explained by low miR-1307-5p I0 expression or miR-1307-5p I0 promotes phenotypes that we did not test (for instance, autophagy or metabolism-related phenotypes). Moreover, targets crucial for an oncogenic phenotype might not be expressed in the overexpression cell lines and, therefore, render even high levels of miR-1307-5p I0 ineffective.

Opposing roles of microRNAs transcribed from different arms of the pre-microRNA have been described already. miR-28-3p, for instance, promotes migration and invasion, whereas miR-28-5p represses both phenotypes and proliferation in colorectal as well as in nasopharyngeal cancer[250,251]. Other examples show contrasting roles for miR-34-3p and miR-34-5p, miR-514b-3p and miR-514b-5p as well as miR-574-3p and miR-574-5p[252–254]. However, miR-1307-5p I0 target predictions generated by Shashwat Sahay did not contain any well published tumor suppressors and there are no functional studies about miR-1307-5p I0 and its impact on cancer-related phenotypes. To exclude miR-1307-5p I0 as oncogenic driver of highly abundant, pre-miR-1307-derived microRNAs, expression levels would need to be determined in the overexpression cells and functional assays with miR-1307-5p I0 mimics would need to be performed.

The mapping of pre-miR-1307 within an intron of *ATP5MD* (Figure 16C) could be another possible explanation for high expression levels of the tumor-suppressive miR-1307-3p, since *ATP5MD* is also upregulated in (aggressive) tumors (Figure 15B). A modest correlation between *ATP5MD* and miR-1307-3p I0 / miR-1307-3p I1 (r=0.38-0.42, Figure 15C) in Basal breast cancer patients supports the notion that pre-miR-1307 might be co-regulated with *ATP5MD*. *ATP5MD* codes for a subunit

of the mitochondrial ATP synthase and, therefore, is involved in mitochondrial OXPHOS[154,255], the main source of ATP in aerobic organisms[256]. Besides colocalization with the mitochondrial ATP synthase, ATP5MD is also part of the vacuolar ATP synthase in lysosomes of HUVEC and HEK293T cells[257]. This is not surprising since the mitochondrial ATP synthase and the vacuolar ATP synthase are structurally and mechanistically related[258]. So far, however, studies about the functional importance of ATP5MD are contradicting or suggest a rather structural relevance of ATP5MD for alignment of the mitochondrial ATP synthase along the cristae[259].

The functional role of the mitochondrial ATP synthase matches the significant enrichment of gene signatures related to OXPHOS, ROS and DNA repair in Basal patients with high *ATP5MD* levels (Figure 15D). The enriched signatures underline the function of the ATP synthase: the crucial role of the ATP synthase for OXPHOS[154,255], the high ROS levels that come with this branch of the metabolism[260] and the importance of DNA repair to cope with ROS-induced DNA damage[261]. In TNBC cell lines, for instance, mitochondria are the main ROS source[262]. The genes enriched in the OXPHOS signature contained many components of the ATP synthase, including the ATP Synthase F1 Subunit of which ATP5MD is a part. Moreover, several markers of mitochondrial biogenesis and translation, such as *MRPS15*, *MRPS22* and *TIMM9*, were enriched. *MRPS15*, *MRPS22*, *TIMM9* as well as several ATP Synthase F1 Subunit are upregulated in epithelial breast cancer cells compared to stroma adjacent to the tumor[263]. Significant enrichment of exactly the same signatures for high levels of *ATP5MD*, miR-1307-3p I0 (data not shown) and miR-1307-3p I1 (Figure 15D) suggest that miR-1307-3p and ATP5MD might be involved in similar cellular processes, which would justify the location of pre-miR-1307 within an *ATP5MD* intron.

While the particular relevance of ATP5MD in cancer remains to be discovered, the relevance of the mitochondrial ATP synthase and OXPHOS has been studied extensively. In the past years, there has been a change in the notion that cancer cells would mainly rely on glycolysis over OXPHOS even in the presence of oxygen, the so-called Warburg effect. Now, more and more studies indicate that the decision between glycolysis and OXPHOS seems to be a rather delicate balance including mixed phenotypes[264,265]. Hybrid metabolic phenotypes have been shown for the two TNBC cell lines that I used, SUM-159 and MDA-MB-231[264]. Vacuolar ATP synthases, to which ATP5MD colocalized, acidify organelles like lysosomes or secreted vesicles[266]. Moreover, they are located within the plasma membrane of specific cell types like kidney cells, macrophages and osteoclasts for acidification of the urine, bone resorption and pH homeostasis, respectively[267].

Looking more at the role of ATPases in cancer than in their physiological context shows that several ATPases play a crucial role. Localization of vacuolar ATPase in the plasma membrane, for instance,

provides a higher tolerance of the cells towards extracellular acidity and, thus, renders tumor cells more aggressive and resistant to drugs[268]. Breast cancer cells with (high) abundance of vacuolar ATPase in the plasma membrane display a more migratory and invasive phenotype[269,270]. Both, the mitochondrial ATP synthase as well as the vacuolar ATPase were found in the cell membrane of MDA-MB-231[271]. The presence of ATPases like the mitochondrial ATP synthase in the plasma membrane (referred to as ecto-F-ATP synthase) of cancer cells expands the options in cancer therapy[272–274]. Targeting the ecto-F-ATP synthase with inhibitors affects not only the growth of endothelial cells[275], but also reduces proliferation of several breast cancer and leukemia cell lines[276–278]. Moreover, Angiostatin represses tumor angiogenesis involving ecto-ATP synthase metabolism and antibodies targeting subunits of the ATP synthase mimic this effect[279]. Targeting ecto-ATP5B, a subunit of the F-ATP synthase in the plasma membrane, exerts a cytotoxic effect on MDA-MB-231 cells injected into mice, while normal cells remain unaffected[280].

Considering the relevance of various ATPases and their subunits in the aggressiveness of tumor cells, I assume that the migration phenotype and the possible angiogenesis phenotype mediated by miR-1307-3p I0 could result from direct targeting of various ATPase subunits. In fact, inhibition of the vacuolar ATPase blocks EMT in MDA-MB-231[281]. Even overexpression of individual subunits, such as ATP6L, modulates EMT markers and enhances migration and invasion of colorectal cancer cells[282]. How targeting of the individual ATPase subunits by miR-1307-3p I0 would affect the entire complex and its activity needs to be determined. ATPase abundance in the plasma membrane upon miR-1307-3p I0 overexpression could be determined with immunofluorescence and the migratory potential of breast cancer cells upon knock-down of individual subunits should be evaluated.

miR-1307-3p I0-mediated targeting of ATPase subunits and genes involved in autophagy and UPR might form a negative feedback loop with *ATP5MD*, the host gene of pre-miR-1307 that is upregulated in tumor tissue. An initial test showed that miR-1307-3p I0 (and miR-1307-3p I1) overexpression does not repress the mRNA levels of *ATP5MD* (data not shown). This indicates that there is no negative feedback loop on this particular level, but possibly miR-1307-3p I0 overexpression reduces abundance or activity of the mitochondrial ATP synthase by targeting *ATP5E*, *ATP5F1*, *ATP5G1* and *ATP5S* (Figure 20C and D). Since the target predictions included subunits of other ATPases including the vacuolar ATPase as well, this suggests that miR-1307-3p I0 plays a crucial role in modulating mitochondrial and lysosomal ATPase activities and, thus, potentially maintains cellular stress response in balance. The answer to the question why a tumor-suppressive microRNA is upregulated in cancer indeed might be found in the context of cellular

stress response since autophagy is tightly linked to degradation of cellular compounds in lysosomes upon cellular stress such as ROS or DNA damage[283]. High ROS levels fuel metabolic reprogramming and promote cancer cell growth, on the downside, however, excessive ROS induces DNA damage and can result in cell death[284]. Thus, fine-tuned mechanisms such as mitophagy, autophagy of mitochondria, are required to maintain this delicate balance. miR-1307-3p I0 could play a crucial role in mediating this balance and since mitochondrial stress has been shown to promote invasion in breast cancer cells[284] the tumor-suppressive effects that I observed *in vitro* upon miR-1307-3p I0 overexpression might be a side effect of the role that miR-1307-3p I0 has in balancing cellular stress response.

5.2.5 Conclusion and outlook

Along the line of this project, the divergent 5'isomiR pair miR-1307-3p I0 and miR-1307-3p I1 was selected to characterize the functional importance of 5'isomiRs. I could indeed show that the divergent 5'isomiRs miR-1307-3p I0 and miR-1307-3p I1 mediate different phenotypes: miR-1307-3p I1 reduced proliferation in a cell line-dependent context. Nevertheless, this 5'isomiR pair shows how complex and context-dependent the role of microRNAs and their 5'isomiRs can be: miR-1307-3p I0 and miR-1307-3p I1 share a subset of putative targets and both downregulate migration, miR-1307-3p I0 even in a cell line-dependent manner. Although the mechanism of the phenotypes mediated by miR-1307-3p I0 and miR-1307-3p I1 is not clear, I identified promising putative targets that could contribute to the observed phenotypes. Moreover, target predictions propose autophagy as a miR-1307-3p I0-specific phenotype. Overall, target predictions, MicroArray data and GSEA with patient data suggest that miR-1307-3p I0 modulates migration through alterations in cellular stress response pathways such as autophagy and UPR via repression of various ATPases and *SLC7A5*. Consequently, this might affect protein secretion and could also impact on migration and angiogenesis by targeting of *LBH*, for instance, which might explain the reduced tumor size upon pre-miR-1307 overexpression.

The putative miR-1307-3p I1 target *NCS1* and the miR-1307-3p I0 / miR-1307-3p I1 target *LBH* have been implicated in PI3K/AKT signaling. Interestingly, PI3K signaling plays a crucial role in autophagy[285]. The cell line-dependent impact of miR-1307-3p I0 and miR-1307-3p I1 might be connected to cell line-specific differences in target availability, pathway activities as well as mutation status such as the *PIK3CA*-activating mutation in SUM-159. Considering the multitude of autophagy- and UPR-related targets predicted for miR-1307-3p I0, the cell line-specific effects

might also result from differences in copying with cellular stress, which might involve the PI3K pathway in the context of autophagy.

Future experiments should employ a systematic approach to profile the functional impact of miR-1307-3p I0 or miR-1307-3p I1 and to obtain a more detailed understanding of their mechanistic differences. Based on the data gathered in this study, protein secretion of potentially migration-/angiogenesis-promoting factors, autophagy and stress response seem to play an important role for the miR-1307-3p I0- and miR-1307-3p I1-mediated phenotypes. Secretome analysis and functional experiments with the cell culture supernatant upon miR-1307-3p I0 / miR-1307-3p I1 overexpression could be the first step to validate predicted targets that might contribute to the observed phenotypes. Moreover, experiments investigating autophagy might identify a miR-1307-3p I0-specific phenotype. Immunofluorescence of ATPases and autophagy-related targets could unravel the impact of miR-1307-3p I0 overexpression on lysosomal trafficking and protein secretion. In summary, gaining further insight into the phenotypic and mechanistic differences of miR-1307-3p I0 and miR-1307-3p I1 will add to the understanding of the miRNome and underline that it is crucial to distinguish between 5'isomiRs, for instance, in terms of diagnostic and prognostic purposes.

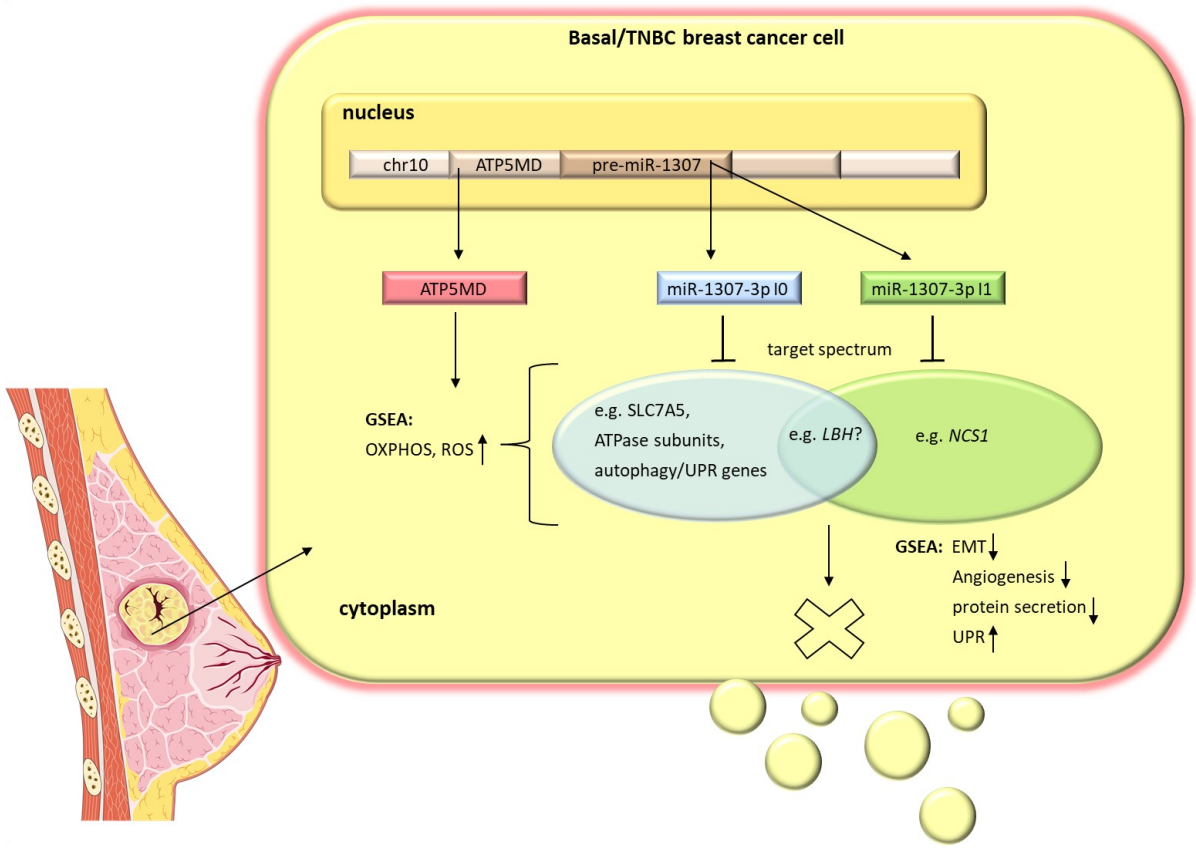


Figure 22: Schematic overview illustrating the potential role of miR-1307-3p I0 and I1 in breast cancer. pre-miR-1307 is located within an intron of *ATP5MD*, a subunit of the mitochondrial ATP synthase, and gives

rise to the mature microRNAs miR-1307-3p I0 and miR-1307-3p I1. miR-1307-3p I0 and miR-1307-3p I1 are divergent 5'isomiRs that mediate different tumor-suppressive effects *in vitro*. miR-1307-3p I1, for instance, reduces proliferation of SUM-159. Both microRNAs have been shown to repress migration in a TNBC cell line-dependent manner, which is in line with a depleted EMT gene set in Basal breast cancer patients (TCGA) with high miR-1307-3p I0 / miR-1307-3p I1 expression. Based on target predictions and preliminary MicroArray and TaqMan data, individual targets were identified as promising candidates that might explain the determined phenotypes. This includes *SLC7A5* and several ATPase subunits targeted by miR-1307-3p I0 as well as *NCS1*, a miR-1307-3p I1 target that could contribute to the potentially PI3K-dependent proliferation phenotype in SUM-159 cells. Moreover, *LBH* is likely involved in the potential angiogenesis phenotype and might contribute to the migration phenotype, although it is not certain whether this target is a common target of the two 5'isomiRs or a miR-1307-3p I0-specific target. GSEA indicated significant enrichment of gene signatures related to OXPHOS, ROS and UPR for Basal breast cancer patients with high levels of the 5'isomiRs. In combination with the multiple ATPase subunits and autophagy-/UPR-related genes potentially targeted by miR-1307-3p I0, this suggests that miR-1307-3p might play a crucial role in cellular stress response. Furthermore, ATPases have been linked to protein secretion, which could be associated with the secretion of migration-promoting and pro-angiogenic factors and would explain miR-1307-3p I0- and miR-1307-3p I1-mediated repression of migration in response to reduced protein secretion and targeting of ATPase subunits. This notion is supported by depletion of gene sets related to angiogenesis and protein secretion in patients with high levels of the 5'isomiRs and might explain reduced tumor size upon pre-miR-1307 overexpression *in vivo*. The illustration of the breast was obtained from https://smart.servier.com/.

6. SUPPLEMENTARY

my pre-microRNA library		library after batch correction	
pre-miR-141	pre-miR-125b-2*	pre-miR-141	pre-miR-193a
pre-miR-200a	pre-miR-130b	pre-miR-200a	pre-miR-19a
pre-miR-21	pre-miR-155*	pre-miR-21	pre-miR-223
pre-miR-29a	pre-miR-17	pre-miR-29a	pre-miR-30c-2
pre-miR-139	pre-miR-16-1	pre-miR-139	pre-miR-423
pre-miR-145	pre-miR-16-2	pre-miR-145	pre-miR-744
pre-miR-200c*	pre-miR-29b-1	pre-miR-200c*	pre-miR-9
pre-miR-381	pre-miR-7-1	pre-miR-381	pre-miR-96
pre-let-7c	pre-miR-10a	pre-let-7c	pre-miR-1
pre-miR-100	pre-miR-127	pre-miR-100	pre-miR-135b
pre-miR-103a-1	pre-miR-204	pre-miR-103a-1	pre-miR-149
pre-miR-103a-2	pre-miR-22	pre-miR-103a-2	pre-miR-153*
pre-miR-106b	pre-miR-29b-2	pre-miR-106b	pre-miR-15b
pre-miR-10b	pre-miR-326	pre-miR-10b	pre-miR-184
pre-miR-125b-1	pre-miR-342	pre-miR-125b-1	pre-miR-195*
pre-miR-126	pre-miR-551b	pre-miR-126	pre-miR-200b
pre-miR-1307		pre-miR-1307	pre-miR-205*
pre-miR-140		pre-miR-140	pre-miR-30a
pre-miR-142*		pre-miR-142*	pre-miR-330
pre-miR-143		pre-miR-143	pre-miR-338
pre-miR-144		pre-miR-144	pre-miR-429
pre-miR-148b		pre-miR-148b	pre-miR-450b*
pre-miR-151a		pre-miR-151a	pre-miR-584
pre-miR-182		pre-miR-182	pre-miR-708*
pre-miR-183		pre-miR-183	pre-let-7g
pre-miR-190b		pre-miR-190b	pre-miR-105
pre-miR-191		pre-miR-191	pre-miR-1269a
pre-miR-192		pre-miR-192	pre-miR-181c
pre-miR-203a		pre-miR-203a	pre-miR-187
pre-miR-210		pre-miR-210	pre-miR-199b
pre-miR-3065		pre-miR-3065	pre-miR-30b
pre-miR-337		pre-miR-337	pre-miR-30d
pre-miR-365a		pre-miR-365a	pre-miR-324
pre-miR-375		pre-miR-375	pre-miR-331
pre-miR-378a		pre-miR-378a	pre-miR-335
pre-miR-379		pre-miR-379	pre-miR-365b
pre-miR-425		pre-miR-425	pre-miR-628
pre-miR-451a		pre-miR-451a	pre-miR-629
pre-miR-452		pre-miR-452	
pre-miR-455		pre-miR-455	
pre-miR-486-1		pre-miR-486-1	
pre-miR-486-2		pre-miR-486-2	
pre-miR-497		pre-miR-497	
pre-miR-92b		pre-miR-92b	
pre-miR-93		pre-miR-93	
pre-miR-99a		pre-miR-99a	
pre-miR-99b		pre-miR-99b	

Figure S1: Comparison of the pre-microRNA selection before and after batch correction of the TCGA data. microRNA expression data from TCGA patients is strongly biased by batch effects. Comparing the selection of pre-microRNAs for the library based on TCGA data before and after batch correction with the R package 'ComBat' revealed a strong impact of the bias. Pre-microRNAs that remained in the selection after correction of the batch effects are highlighted in grey. Blue and red indicate pre-microRNAs that are processed to microRNAs which reduce or promote chemoresistance according to literature, respectively. Asterisks mark microRNAs that were shown to modulate stemness.

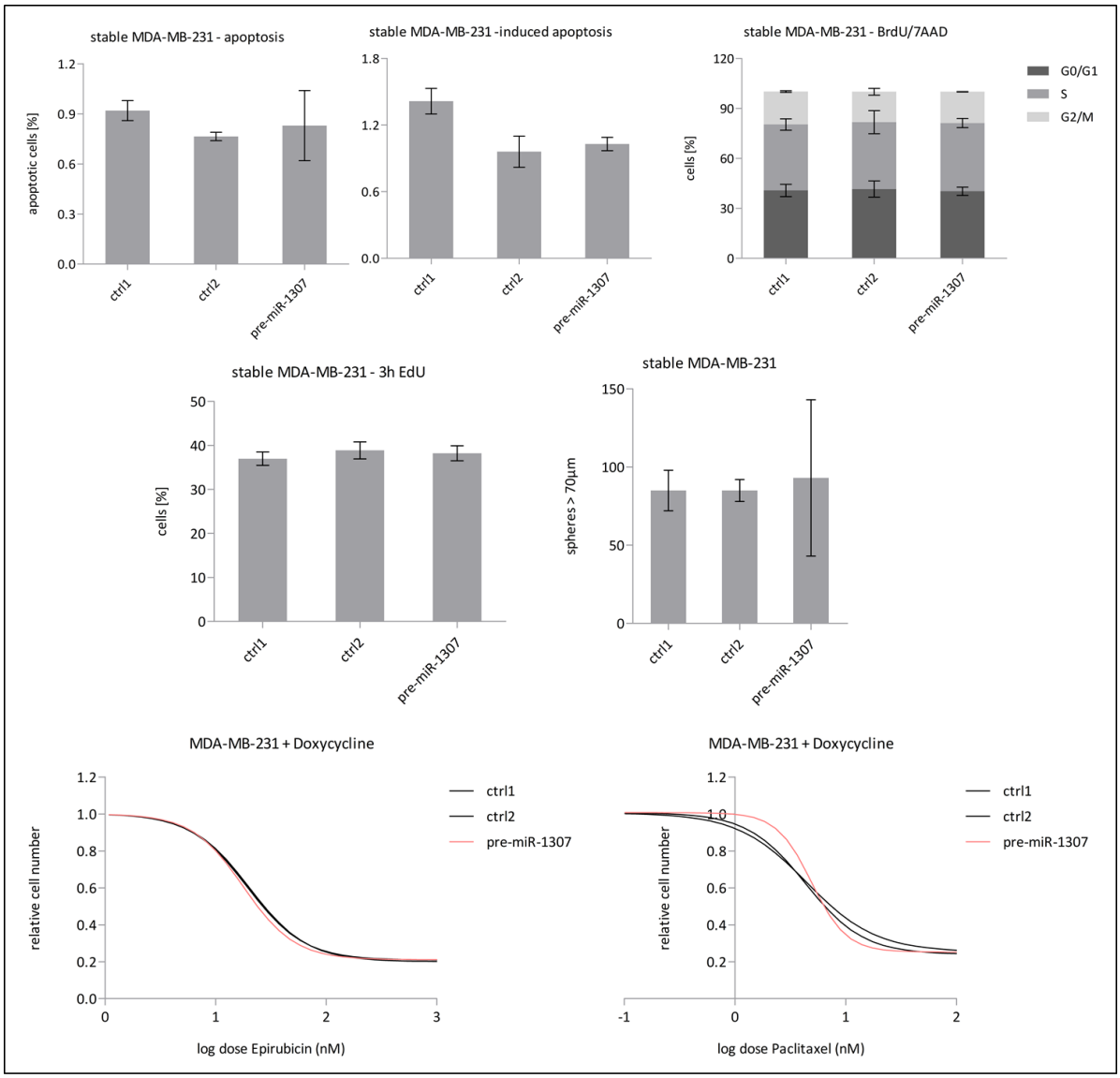


Figure S2: Evaluation of various cancer phenotypes upon pre-miR-1307 overexpression in MDA-MB-231. MDA-MB-231 cells were retrovirally transduced with pre-miR-1307 or non-targeting pre-microRNAs under control of a Doxycycline-inducible promoter. 48 hours post seeding, apoptosis (n=2) and apoptosis after Staurosporine-induction (n=2) was evaluated by FACS after Propidium Iodide staining. Cell cycle phases after BrdU/7AAD staining (n=2) were determined by FACS 48 hours post seeding as well. The fraction of cells in S phase after EdU labeling (n=3) was detected by microscope 48 hours post seeding. Hoechst staining was used to stain all cells and calculate the cell fraction in S phase. Spheres >70 µM were counted by microscope after seven days of sphere formation (n=2). Resistance towards Paclitaxel and Epirubicin was evaluated after 72h. The cell number was determined by microscope after Hoechst staining. Drug response curves are displayed as fitted curve based on the mean of six technical replicates. Error bars represent the standard deviation of biological replicates unless indicated otherwise. Statistical significance was calculated with an unpaired two-tailed student's t-test.

A

miR-1307-3p I0 targets (TargetScan)

gene	binding site	cellular process	reference
ATF5	6mer	autophagy	PMID 29137451
ATF6	6mer	UPR, autophagy	PMID 26622781
ATG14	6mer	autophagy	PMID 30622432
ATG4B	6mer	autophagy	PMID 30622432
ATG9B	6mer	autophagy	PMID 28740555
FOXO3	7mer-1a	autophagy	PMID 25775030
IRAK2	7mer-1a	UPR	PMID 23724040
IRF1	7mer-1a	autophagy	PMID 27308537
IRF8	7mer-1a	autophagy	PMID 25775030
MLST8	6mer	autophagy	PMID 21878654
TLR6	6mer	autophagy	PMID 25009542
TLR7	6mer	autophagy	PMID 25009542
TRAF2	6mer	UPR, autophagy	PMID 26622781
TRAF3	6mer	autophagy	PMID 29146913
TSC1	6mer	autophagy	PMID 26622781
ULK1	6mer	autophagy	PMID 30622432
XIAP	7mer-1a	autophagy	PMID 29743550

B

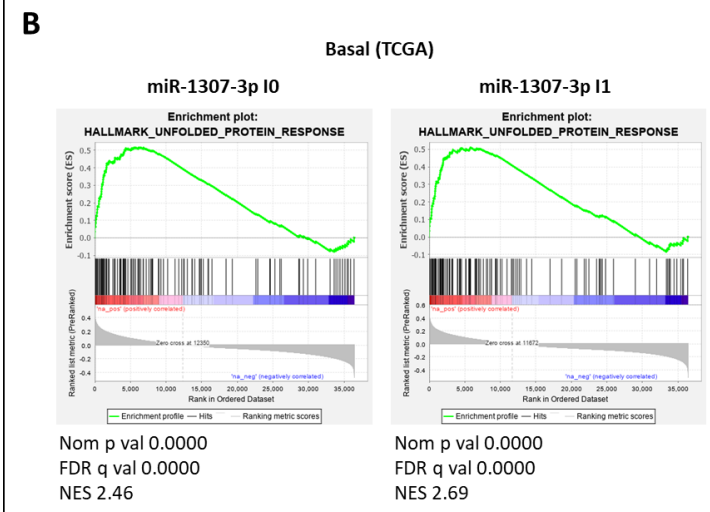


Figure S3: miR-1307-3p I0 targets various genes involved in autophagy and UPR. A) A selection of autophagy- and UPR-related targets that were specifically predicted for miR-1307-3p I0 with the TargetScan algorithm. **B)** GSEA of Basal patients from TCGA ranked by their miR-1307-3p I0 or miR-1307-3p I1 expression was performed and revealed significant enrichment of a gene signature related to UPR. NES represents the normalized enrichment score indicating positive enrichment.

7. LIST OF FIGURES

8. LIST OF TABLES

9. ABBREVIATIONS

3'UTR	3' untranslated region
ABCC1	ATP Binding Cassette Subfamily C Member 1
ABCC2	ATP Binding Cassette Subfamily C Member 2
ACTB	Actin Beta
AKT	AKT Serine/Threonine Kinase
ALDH1A1	Aldehyde Dehydrogenase 1 Family Member A1
ATP	Adenosine Triphosphate
ATP5B	ATP Synthase F1 Subunit Beta
ATP5G1	ATP Synthase Membrane Subunit C Locus 1
ATP6V1C1	ATPase H+ Transporting V1 Subunit C1
BRAF	B-Raf Proto-Oncogene, Serine/Threonine Kinase
BRCA	Breast Cancer
BRCA1	BRCA1 DNA Repair Associated
CAT	Catalase
ChIP	Chromatin Immunoprecipitation
CDH1	Cadherin 1, E-cadherin
CDH2	Cadherin 2, N-cadherin
CDKN2A	Cyclin Dependent Kinase Inhibitor 2A
cDNA	complementary DNA
c-Kit	KIT Proto-Oncogene
ddH2O	double-distilled water
DGCR8	DGCR8 Microprocessor Complex Subunit
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic Acid
dNTPs	Deoxynucleosidtriphosphate
Doxy	Doxycycline
EDTA	Ethylenediaminetetraacetic Acid
EGF	Epidermal Growth Factor
EGFR	EGFR receptor
ELF4	E74 Like ETS Transcription Factor 4
EMT	Epithelial to Mesenchymal Transition
ER	Endoplasmic Reticulum

FACS	Fluorescence-activated Cell Sorting
FBL	Fibrillarin
FC	fold change
FGF	Fibroblast Growth Factor
FGFR	FGF receptor
FN1	Fibronectin 1
GAPDH	Glyceraldehyde-3-Phosphate Dehydrogenase
GFP	Green Fluorescent Protein
GPX1	Glutathione Peroxidase 1
HER2	Human Epidermal Growth Factor Receptor 2
HPRT1	Hypoxanthine Phosphoribosyltransferase 1
HRAS	HRas Proto-Oncogene, GTPase
IL-8	C-X-C Motif Chemokine Ligand 8
ITGA6	Integrin Subunit Alpha 6
ITGB3	Integrin Subunit Beta 3
kb	kilobase = 1,000 base pairs
KLF4	Kruppel Like Factor 4
KRAS	KRAS Proto-Oncogene, GTPase
LAT1	L amino-acid transporter 1
LumA	Luminal A
LumB	Luminal B
MMP	Matrix Metallopeptidase
mRNA	messenger RNA
MRPS15	Mitochondrial Ribosomal Protein S15
MRPS22	Mitochondrial Ribosomal Protein S22
MYC	MYC Proto-Oncogene
NPM1	Nucleophosmin 1
OCT4	POU Class 5 Homeobox 1
PAM50	Prediction-Analysis-of-Microarray (based on 50 genes)
PBS	Phosphate-buffered Saline
PCR	Polymerase Chain Reaction
PDCD11	Programmed Cell Death 11
PDGF	Platelet-derived Growth Factor
PDGFR	PDGF receptor

PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
PTEN	Phosphatase And Tensin Homolog
pre-microRNA	precursor microRNA
PROM1	Prominin 1
QC	quality control
qRT-PCR	quantitative Real-Time PCR
RB1	RB Transcriptional Corepressor 1
RC3H2	Ring Finger And CCCH-Type Domains 2
RelA	RELA Proto-Oncogene
RNA	Ribonucleic Acid
RT	Reverse Transcription
RUNX3	RUNX Family Transcription Factor 3
siRNA	small interfering RNA
SLC7A5	Solute Carrier Family 7 Member 5
SNAI1	Snail Family Transcriptional Repressor 1
SNAI2	Snail Family Transcriptional Repressor 2
SNORD61	Small Nucleolar RNA, C/D Box 61
SNORD72	Small Nucleolar RNA, C/D Box 72
SNORD95	Small Nucleolar RNA, C/D Box 95
SOD2	Superoxid Dismutase 2
SOX2	SRY-Box Transcription Factor 2
STK40	Serine/Threonine Kinase 40
TCGA	The Cancer Genome Atlas
TIMM9	Translocase Of Inner Mitochondrial Membrane 9
TOR1B	Torsin Family 1 Member B
TP53	Tumor Protein P53
TRBP	TARBP2 Subunit Of RISC Loading Complex
UNC5B	Unc-5 Netrin Receptor B
UPR	unfolded protein response
VEGF-A	Vascular Endothelial Growth Factor A
VEGFR-1	VEGF receptor 1
VEGFR-2	VEGF receptor 2
VIM	Vimentin
VPS37C	VPS37C Subunit Of ESCRT-I

10. REFERENCES

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394–424 (2018).
2. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2019. *CA Cancer J Clin* **69**, 7–34 (2019).
3. Bertos, N. R. & Park, M. Breast cancer — one term, many entities? *J Clin Invest* **121**, 3789–3796 (2011).
4. Ma, L., Liang, Z., Zhou, H. & Qu, L. Applications of RNA Indexes for Precision Oncology in Breast Cancer. *Genomics Proteomics Bioinformatics* **16**, 108–119 (2018).
5. Lachapelle, J. & Foulkes, W. D. Triple-negative and basal-like breast cancer: implications for oncologists. *Curr Oncol* **18**, 161–164 (2011).
6. Yin, L., Duan, J.-J., Bian, X.-W. & Yu, S. Triple-negative breast cancer molecular subtyping and treatment progress. *Breast Cancer Research* **22**, 61 (2020).
7. Schettini, F. *et al.* Hormone Receptor/Human Epidermal Growth Factor Receptor 2-positive breast cancer: Where we are now and where we are going. *Cancer Treat. Rev.* **46**, 20–26 (2016).
8. Crowe, J. P. *et al.* Estrogen receptor determination and long term survival of patients with carcinoma of the breast. *Surg Gynecol Obstet* **173**, 273–278 (1991).
9. Bastien, R. R. L. *et al.* PAM50 breast cancer subtyping by RT-qPCR and concordance with standard clinical molecular markers. *BMC Med Genomics* **5**, 44 (2012).
10. Sweeney, C. *et al.* Intrinsic subtypes from PAM50 gene expression assay in a population-based breast cancer cohort: differences by age, race, and tumor characteristics. *Cancer Epidemiol. Biomarkers Prev.* **23**, 714–724 (2014).
11. Prat, A. & Perou, C. M. Deconstructing the molecular portraits of breast cancer. *Mol Oncol* **5**, 5–23 (2011).
12. Yersal, O. & Barutca, S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World J Clin Oncol* **5**, 412–424 (2014).

13. Wang, W.-J., Lei, Y.-Y., Mei, J.-H. & Wang, C.-L. Recent progress in HER2 associated breast cancer. *Asian Pac. J. Cancer Prev.* **16**, 2591–2600 (2015).

14. Rakha, E. A., Reis-Filho, J. S. & Ellis, I. O. Basal-like breast cancer: a critical review. *J. Clin. Oncol.* **26**, 2568–2581 (2008).

15. Badowska-Kozakiewicz, A. M. & Budzik, M. P. Immunohistochemical characteristics of basal-like breast cancer. *Contemp Oncol (Pozn)* **20**, 436–443 (2016).

16. Summa, S. D. *et al.* BRCAness: a deeper insight into basal-like breast tumors. *Annals of Oncology* **24**, viii13–viii21 (2013).

17. Darb-Esfahani, S. *et al.* Role of TP53 mutations in triple negative and HER2-positive breast cancer treated with neoadjuvant anthracycline/taxane-based chemotherapy. *Oncotarget* **7**, 67686–67698 (2016).

18. Santarpia, L. *et al.* Mutation profiling identifies numerous rare drug targets and distinct mutation patterns in different clinical subtypes of breast cancers. *Breast Cancer Res Treat* **134**, (2012).

19. Shi, Y., Jin, J., Ji, W. & Guan, X. Therapeutic landscape in mutational triple negative breast cancer. *Molecular Cancer* **17**, 99 (2018).

20. Liedtke, C. *et al.* Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J. Clin. Oncol.* **26**, 1275–1281 (2008).

21. Carey, L. A. *et al.* The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clin. Cancer Res.* **13**, 2329–2334 (2007).

22. Nahleh, Z., Sivasubramaniam, D., Dhaliwal, S., Sundarajan, V. & Komrokji, R. Residual cancer burden in locally advanced breast cancer: a superior tool. *Curr Oncol* **15**, 271–278 (2008).

23. Abramson, V. G., Lehmann, B. D., Ballinger, T. J. & Pietenpol, J. A. Subtyping of triple-negative breast cancer: implications for therapy. *Cancer* **121**, 8–16 (2015).

24. Lehmann, B. D. *et al.* Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. *PLoS ONE* **11**, e0157368 (2016).

25. Uscanga-Perales, G. I., Santuario-Facio, S. K. & Ortiz-López, R. Triple negative breast cancer: Deciphering the biology and heterogeneity. *Medicina Universitaria* **18**, 105–114 (2016).

26. Gerratana, L. *et al.* Androgen receptor in triple negative breast cancer: A potential target for the targetless subtype. *Cancer Treatment Reviews* **68**, 102–110 (2018).

27. Wathieu, H. *et al.* Differential prioritization of therapies to subtypes of triple negative breast cancer using a systems medicine method. *Oncotarget* **8**, 92926–92942 (2017).

28. Schneeweiss, A. *et al.* Diagnosis and Therapy of Triple-Negative Breast Cancer (TNBC) – Recommendations for Daily Routine Practice. *Geburtshilfe Frauenheilkd* **79**, 605–617 (2019).

29. Echeverria, G. V. *et al.* Resistance to neoadjuvant chemotherapy in triple-negative breast cancer mediated by a reversible drug-tolerant state. *Sci Transl Med* **11**, (2019).

30. Nakashoji, A. *et al.* Clinical predictors of pathological complete response to neoadjuvant chemotherapy in triple-negative breast cancer. *Oncol Lett* **14**, 4135–4141 (2017).

31. Masuda, H. *et al.* Differential response to neoadjuvant chemotherapy among 7 triple-negative breast cancer molecular subtypes. *Clin. Cancer Res.* **19**, 5533–5540 (2013).

32. Omarini, C. *et al.* Neoadjuvant treatments in triple-negative breast cancer patients: where we are now and where we are going. *Cancer Manag Res* **10**, 91–103 (2018).

33. Harbeck, N. & Gluz, O. Neoadjuvant therapy for triple negative and HER2-positive early breast cancer. *Breast* **34 Suppl 1**, S99–S103 (2017).

34. Yadav, B. S., Sharma, S. C., Chanana, P. & Jhamb, S. Systemic treatment strategies for triple-negative breast cancer. *World J Clin Oncol* **5**, 125–133 (2014).

35. Caparica, R., Lambertini, M. & de Azambuja, E. How I treat metastatic triple-negative breast cancer. *ESMO Open* **4**, e000504 (2019).

36. Yu, K.-D. *et al.* Identification of prognosis-relevant subgroups in patients with chemoresistant triple-negative breast cancer. *Clin. Cancer Res.* **19**, 2723–2733 (2013).

37. Nedeljković, M. & Damjanović, A. Mechanisms of Chemotherapy Resistance in Triple-Negative Breast Cancer-How We Can Rise to the Challenge. *Cells* **8**, (2019).

38. O'Reilly, E. A. *et al.* The fate of chemoresistance in triple negative breast cancer (TNBC). *BBA Clin* **3**, 257–275 (2015).

39. Zhang, F. & Wang, D. The Pattern of microRNA Binding Site Distribution. *Genes (Basel)* **8**, (2017).

40. Lee, I. *et al.* New class of microRNA targets containing simultaneous 5'-UTR and 3'-UTR interaction sites. *Genome Research* **19**, 1175 (2009).

41. Rodriguez, A., Griffiths-Jones, S., Ashurst, J. L. & Bradley, A. Identification of Mammalian microRNA Host Genes and Transcription Units. *Genome Res* **14**, 1902–1910 (2004).

42. Kim, Y.-K. & Kim, V. N. Processing of intronic microRNAs. *The EMBO Journal* **26**, 775–783 (2007).

43. Bajan, S. & Hutvagner, G. Regulation of miRNA processing and miRNA mediated gene repression in cancer. *Microrna* **3**, 10–17 (2014).

44. Ding, L. *et al.* MicroRNAs Involved in Carcinogenesis, Prognosis, Therapeutic Resistance, and Applications in Human Triple-Negative Breast Cancer. *Cells* **8**, (2019).

45. Cai, X., Hagedorn, C. H. & Cullen, B. R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* **10**, 1957–1966 (2004).

46. Tarver, J. E., Donoghue, P. C. J. & Peterson, K. J. Do miRNAs have a deep evolutionary history? *BioEssays* **34**, 857–866 (2012).

47. Meijer, H. A., Smith, E. M. & Bushell, M. Regulation of miRNA strand selection: follow the leader? *Biochem. Soc. Trans.* **42**, 1135–1140 (2014).

48. Kim, V. N. MicroRNA biogenesis: coordinated cropping and dicing. *Nature Reviews Molecular Cell Biology* **6**, 376–385 (2005).

49. O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Front Endocrinol (Lausanne)* **9**, (2018).

50. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).

51. Yang, J.-S. & Lai, E. C. Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. *Mol. Cell* **43**, 892–903 (2011).

52. Gebert, L. F. R. & MacRae, I. J. Regulation of microRNA function in animals. *Nat Rev Mol Cell Biol* **20**, 21–37 (2019).

53. Peters, L. & Meister, G. Argonaute Proteins: Mediators of RNA Silencing. *Molecular Cell* **26**, 611–623 (2007).

54. Ryan, B. M., Robles, A. I. & Harris, C. C. Genetic variation in microRNA networks: the implications for cancer research. *Nat. Rev. Cancer* **10**, 389–402 (2010).

55. Fuchs Wightman, F., Giono, L. E., Fededa, J. P. & de la Mata, M. Target RNAs Strike Back on MicroRNAs. *Front Genet* **9**, (2018).

56. Friedman, R. C., Farh, K. K.-H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**, 92–105 (2009).

57. Dhawan, A., Scott, J. G., Harris, A. L. & Buffa, F. M. Pan-cancer characterisation of microRNA across cancer hallmarks reveals microRNA-mediated downregulation of tumour suppressors. *Nature Communications* **9**, 1–13 (2018).

58. Loh, H.-Y. *et al.* The Regulatory Role of MicroRNAs in Breast Cancer. *Int J Mol Sci* **20**, (2019).

59. Grimson, A. *et al.* MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* **27**, 91–105 (2007).

60. Baek, D. *et al.* The impact of microRNAs on protein output. *Nature* **455**, 64–71 (2008).

61. Witkos, T. M., Koscianska, E. & Krzyzosiak, W. J. Practical Aspects of microRNA Target Prediction. *Curr Mol Med* **11**, 93–109 (2011).

62. Bartel, D. P. MicroRNAs: target recognition and regulatory functions. *Cell* **136**, 215–233 (2009).

63. Lytle, J. R., Yario, T. A. & Steitz, J. A. Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR. *Proc Natl Acad Sci U S A* **104**, 9667–9672 (2007).

64. Zhang, J. *et al.* Oncogenic role of microRNA-532-5p in human colorectal cancer via targeting of the 5′UTR of RUNX3. *Oncol Lett* **15**, 7215–7220 (2018).

65. Forman, J. J., Legesse-Miller, A. & Coller, H. A. A search for conserved sequences in coding regions reveals that the let-7 microRNA targets Dicer within its coding sequence. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 14879–14884 (2008).

66. Dharap, A., Pokrzywa, C., Murali, S., Pandi, G. & Vemuganti, R. MicroRNA miR-324-3p induces promoter-mediated expression of RelA gene. *PLoS ONE* **8**, e79467 (2013).

67. Jin, Y., Chen, Z., Liu, X. & Zhou, X. Evaluating the MicroRNA Targeting Sites by Luciferase Reporter Gene Assay. *Methods Mol Biol* **936**, 117–127 (2013).

68. Helwak, A. & Tollervey, D. Mapping the miRNA interactome by crosslinking ligation and sequencing of hybrids (CLASH). *Nat Protoc* **9**, 711–728 (2014).

69. Ekimler, S. & Sahin, K. Computational Methods for MicroRNA Target Prediction. *Genes (Basel)* **5**, 671–683 (2014).

70. Haecker, I. & Renne, R. HITS-CLIP and PAR-CLIP advance viral miRNA targetome analysis. *Crit Rev Eukaryot Gene Expr* **24**, 101–116 (2014).

71. Huppertz, I. *et al.* iCLIP: Protein–RNA interactions at nucleotide resolution. *Methods* **65**, 274–287 (2014).

72. Dash, S., Balasubramaniam, M., Dash, C. & Pandhare, J. Biotin-based Pulldown Assay to Validate mRNA Targets of Cellular miRNAs. *J Vis Exp* (2018) doi:10.3791/57786.

73. Mockly, S. & Seitz, H. Inconsistencies and Limitations of Current MicroRNA Target Identification Methods. *Methods Mol. Biol.* **1970**, 291–314 (2019).

74. Helwak, A., Kudla, G., Dudnakova, T. & Tollervey, D. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**, 654–665 (2013).

75. Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39**, D152-157 (2011).

76. Telonis, A. G., Loher, P., Jing, Y., Londin, E. & Rigoutsos, I. Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity. *Nucleic Acids Res.* **43**, 9158–9175 (2015).

77. Desvignes, T. *et al.* miRNA Nomenclature: A View Incorporating Genetic Origins, Biosynthetic Pathways, and Sequence Variants. *Trends Genet.* **31**, 613–626 (2015).

78. Bofill-De Ros, X. *et al.* Structural Differences between Pri-miRNA Paralogs Promote Alternative Drosha Cleavage and Expand Target Repertoires. *Cell Rep* **26**, 447-459.e4 (2019).

79. Kim, B., Jeong, K. & Kim, V. N. Genome-wide Mapping of DROSHA Cleavage Sites on Primary MicroRNAs and Noncanonical Substrates. *Mol. Cell* **66**, 258-269.e5 (2017).

80. Ma, H. *et al.* A sliding-bulge structure at the Dicer processing site of pre-miRNAs regulates alternative Dicer processing to generate 5'-isomiRs. *Heliyon* **2**, (2016).

81. Starega-Roslan, J., Witkos, T. M., Galka-Marciniak, P. & Krzyzosiak, W. J. Sequence Features of Drosha and Dicer Cleavage Sites Affect the Complexity of IsomiRs. *Int J Mol Sci* **16**, 8110–8127 (2015).

82. Neilsen, C. T., Goodall, G. J. & Bracken, C. P. IsomiRs--the overlooked repertoire in the dynamic microRNAome. *Trends Genet.* **28**, 544–549 (2012).

83. Han, B. W., Hung, J.-H., Weng, Z., Zamore, P. D. & Ameres, S. L. The 3'-to-5' exoribonuclease Nibbler shapes the 3' ends of microRNAs bound to Drosophila Argonaute1. *Curr. Biol.* **21**, 1878–1887 (2011).

84. Westholm, J. O., Ladewig, E., Okamura, K., Robine, N. & Lai, E. C. Common and distinct patterns of terminal modifications to mirtrons and canonical microRNAs. *RNA* **18**, 177–192 (2012).

85. Wyman, S. K. *et al.* Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. *Genome Res.* **21**, 1450–1461 (2011).

86. Nishikura, K. Functions and regulation of RNA editing by ADAR deaminases. *Annu. Rev. Biochem.* **79**, 321–349 (2010).

87. Tan, G. C. *et al.* 5' isomiR variation is of functional and evolutionary importance. *Nucleic Acids Res.* **42**, 9424–9435 (2014).

88. Fernandez-Valverde, S. L., Taft, R. J. & Mattick, J. S. Dynamic isomiR regulation in Drosophila development. *RNA* **16**, 1881–1888 (2010).

89. Loher, P., Londin, E. R. & Rigoutsos, I. IsomiR expression profiles in human lymphoblastoid cell lines exhibit population and gender dependencies. *Oncotarget* **5**, 8790–8802 (2014).

90. Telonis, A. G. & Rigoutsos, I. Race Disparities in the Contribution of miRNA Isoforms and tRNA-Derived Fragments to Triple-Negative Breast Cancer. *Cancer Res.* **78**, 1140–1154 (2018).

91. Nigita, G. *et al.* microRNA editing in seed region aligns with cellular changes in hypoxic conditions. *Nucleic Acids Res.* **44**, 6298–6308 (2016).

92. van der Kwast, R. V. C. T., Woudenberg, T., Quax, P. H. A. & Nossent, A. Y. MicroRNA-411 and Its 5'-IsomiR Have Distinct Targets and Functions and Are Differentially Regulated in the Vasculature under Ischemia. *Mol. Ther.* **28**, 157–170 (2020).

93. Nejad, C. *et al.* miR-222 isoforms are differentially regulated by type-I interferon. *RNA* **24**, 332–341 (2018).

94. Llorens, F. *et al.* A highly expressed miR-101 isomiR is a functional silencing small RNA. *BMC Genomics* **14**, 104 (2013).

95. Salem, O. *et al.* The highly expressed 5'isomiR of hsa-miR-140-3p contributes to the tumor-suppressive effects of miR-140 by reducing breast cancer proliferation and migration. *BMC Genomics* **17**, 566 (2016).

96. Manzano, M., Forte, E., Raja, A. N., Schipma, M. J. & Gottwein, E. Divergent target recognition by coexpressed 5'-isomiRs of miR-142-3p and selective viral mimicry. *RNA* **21**, 1606–1620 (2015).

97. Magee, R., Telonis, A. G., Cherlin, T., Rigoutsos, I. & Londin, E. Assessment of isomiR Discrimination Using Commercial qPCR Methods. *Noncoding RNA* **3**, (2017).

98. Shigematsu, M., Honda, S. & Kirino, Y. Dumbbell-PCR for Discriminative Quantification of a Small RNA Variant. *Methods Mol. Biol.* **1680**, 65–73 (2018).

99. Honda, S. & Kirino, Y. Dumbbell-PCR: a method to quantify specific small RNA variants with a single nucleotide resolution at terminal sequences. *Nucleic Acids Res.* **43**, e77 (2015).

100. Connolly, A. R., Hirani, R., Ellis, A. V. & Trau, M. A DNA Circuit for IsomiR Detection. *Chembiochem* **17**, 2172–2178 (2016).

101. Androvic, P., Valihrach, L., Elling, J., Sjoback, R. & Kubista, M. Two-tailed RT-qPCR: a novel method for highly accurate miRNA quantification. *Nucleic Acids Res.* **45**, e144 (2017).

102. Pedroza-Torres, A. *et al.* MicroRNAs in Tumor Cell Metabolism: Roles and Therapeutic Opportunities. *Front Oncol* **9**, (2019).

103. Baranwal, S. & Alahari, S. K. miRNA control of tumor cell invasion and metastasis. *Int J Cancer* **126**, 1283–1290 (2010).

104. Silva Rodrigues, D. V. *et al.* MicroRNAs in cell cycle progression and proliferation: molecular mechanisms and pathways. *Non-coding RNA Investigation* **2**, (2018).

105. Garg, M. Emerging role of microRNAs in cancer stem cells: Implications in cancer therapy. *World J Stem Cells* **7**, 1078–1089 (2015).

106. Gozuacik, D., Akkoc, Y., Ozturk, D. G. & Kocak, M. Autophagy-Regulating microRNAs and Cancer. *Front Oncol* **7**, (2017).

107. Giuli, M. V., Giuliani, E., Screpanti, I., Bellavia, D. & Checquolo, S. Notch Signaling Activation as a Hallmark for Triple-Negative Breast Cancer Subtype. *J Oncol* **2019**, 8707053 (2019).

108. Pohl, S.-G. *et al.* Wnt signaling in triple-negative breast cancer. *Oncogenesis* **6**, e310 (2017).

109. Bianchini, G., Balko, J. M., Mayer, I. A., Sanders, M. E. & Gianni, L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nature Reviews Clinical Oncology* **13**, 674–690 (2016).

110. Yang, W., Cui, G., Ding, M., Yang, M. & Dai, D. MicroRNA-124-3p.1 promotes cell proliferation through Axin1-dependent Wnt signaling pathway and predicts a poor prognosis of triple-negative breast cancer. *J. Clin. Lab. Anal.* e23266 (2020) doi:10.1002/jcla.23266.

111. Nie, J. *et al.* MiR-125b regulates the proliferation and metastasis of triple negative breast cancer cells via the Wnt/β-catenin pathway and EMT. *Biosci. Biotechnol. Biochem.* **83**, 1062–1071 (2019).

112. Liu, G., Wang, P. & Zhang, H. MiR-6838-5p suppresses cell metastasis and the EMT process in triple-negative breast cancer by targeting WNT3A to inhibit the Wnt pathway. *J Gene Med* **21**, e3129 (2019).

113. Li, H.-Y. *et al.* miR-105/93-3p promotes chemoresistance and circulating miR-105/93-3p acts as a diagnostic biomarker for triple negative breast cancer. *Breast Cancer Res.* **19**, 133 (2017).

114. Guarnieri, A. L. *et al.* The miR-106b-25 cluster mediates breast tumor initiation through activation of NOTCH1 via direct repression of NEDD4L. *Oncogene* **37**, 3879–3893 (2018).

115. Kapadia, C. H., Ioele, S. A. & Day, E. S. Layer-by-layer assembled PLGA nanoparticles carrying miR-34a cargo inhibit the proliferation and cell cycle progression of triple-negative breast cancer cells. *J Biomed Mater Res A* **108**, 601–613 (2020).

116. Karlsen, T. A., Aae, T. F. & Brinchmann, J. E. Robust profiling of microRNAs and isomiRs in human plasma exosomes across 46 individuals. *Sci Rep* **9**, 19999 (2019).

117. Telonis, A. G. *et al.* Knowledge about the presence or absence of miRNA isoforms (isomiRs) can successfully discriminate amongst 32 TCGA cancer types. *Nucleic Acids Res.* **45**, 2973–2985 (2017).

118. Ibuki, Y. *et al.* Circulating microRNA/isomiRs as novel biomarkers of esophageal squamous cell carcinoma. *PLoS ONE* **15**, e0231116 (2020).

119. Lan, C., Peng, H., McGowan, E. M., Hutvagner, G. & Li, J. An isomiR expression panel based novel breast cancer classification approach using improved mutual information. *BMC Med Genomics* **11**, (2018).

120. Koi, Y. *et al.* Predicting the presence of breast cancer using circulating small RNAs, including those in the extracellular vesicles. *Cancer Sci.* **111**, 2104–2115 (2020).

121. Fu, Y., Chen, J. & Huang, Z. Recent progress in microRNA-based delivery systems for the treatment of human disease. *ExRNA* **1**, 24 (2019).

122. Hanna, J., Hossain, G. S. & Kocerha, J. The Potential for microRNA Therapeutics and Clinical Research. *Front. Genet.* **10**, (2019).

123. van Zandwijk, N. *et al.* Safety and activity of microRNA-loaded minicells in patients with recurrent malignant pleural mesothelioma: a first-in-man, phase 1, open-label, dose-escalation study. *Lancet Oncol.* **18**, 1386–1396 (2017).

124. Lindow, M. & Kauppinen, S. Discovering the first microRNA-targeted drug. *J Cell Biol* **199**, 407–412 (2012).

125. van der Ree, M. H. *et al.* Miravirsen dosing in chronic hepatitis C patients results in decreased microRNA-122 levels without affecting other microRNAs in plasma. *Aliment. Pharmacol. Ther.* **43**, 102–113 (2016).

126. Orellana, E. A. *et al.* FolamiRs: Ligand-targeted, vehicle-free delivery of microRNAs for the treatment of cancer. *Sci Transl Med* **9**, (2017).

127. Orellana, E. A. *et al.* Enhancing MicroRNA Activity through Increased Endosomal Release Mediated by Nigericin. *Molecular Therapy - Nucleic Acids* **16**, 505–518 (2019).

128. Ramchandani, D. *et al.* Nanoparticle Delivery of miR-708 Mimetic Impairs Breast Cancer Metastasis. *Mol Cancer Ther* **18**, 579–591 (2019).

129. Yin, H. *et al.* Delivery of Anti-miRNA for Triple-Negative Breast Cancer Therapy Using RNA Nanoparticles Targeting Stem Cell Marker CD133. *Molecular Therapy* **27**, 1252–1261 (2019).

130. Fellmann, C. *et al.* An optimized microRNA backbone for effective single-copy RNAi. *Cell Rep* **5**, 1704–1713 (2013).

131. Soneoka, Y. *et al.* A transient three-plasmid expression system for the production of high titer retroviral vectors. *Nucleic Acids Res.* **23**, 628–633 (1995).

132. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402–408 (2001).

133. Chu, L.-F. *et al.* Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. *Genome Biol.* **17**, 173 (2016).

134. Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* **47**, D155–D162 (2019).

135. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417–425 (2015).

136. Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).

137. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).

138. Miller, K. D. *et al.* Cancer treatment and survivorship statistics, 2019. *CA Cancer J Clin* **69**, 363–385 (2019).

139. Avraham, R. & Yarden, Y. Regulation of signalling by microRNAs. *Biochem. Soc. Trans.* **40**, 26–30 (2012).

140. Wang, J., Yang, M., Li, Y. & Han, B. The Role of MicroRNAs in the Chemoresistance of Breast Cancer. *Drug Dev. Res.* **76**, 368–374 (2015).

141. Rapoport, B. L., Demetriou, G. S., Moodley, S. D. & Benn, C. A. When and how do I use neoadjuvant chemotherapy for breast cancer? *Curr Treat Options Oncol* **15**, 86–98 (2014).

142. Ouyang, M. *et al.* MicroRNA profiling implies new markers of chemoresistance of triple-negative breast cancer. *PLoS ONE* **9**, e96228 (2014).

143. Bourguignon, L. Y. W., Spevak, C. C., Wong, G., Xia, W. & Gilad, E. Hyaluronan-CD44 interaction with protein kinase C(epsilon) promotes oncogenic signaling by the stem cell marker Nanog and the Production of microRNA-21, leading to down-regulation of the tumor suppressor protein PDCD4, anti-apoptosis, and chemotherapy resistance in breast tumor cells. *J. Biol. Chem.* **284**, 26533–26546 (2009).

144. Gao, M. *et al.* miR-145 sensitizes breast cancer to doxorubicin by targeting multidrug resistance-associated protein-1. *Oncotarget* (2016) doi:10.18632/oncotarget.10845.

145. Guan, X. & Guan, Y. miR-145-5p attenuates paclitaxel resistance and suppresses the progression in drug-resistant breast cancer cell lines. *Neoplasma* (2020) doi:10.4149/neo_2020_190622N536.

146. Chang, J. C. Cancer stem cells: Role in tumor growth, recurrence, metastasis, and treatment resistance. *Medicine (Baltimore)* **95**, S20-25 (2016).

147. Chekhun, V. F. *et al.* Association of CD44+CD24-/low with markers of aggressiveness and plasticity of cell lines and tumors of patients with breast cancer. *Exp. Oncol.* **39**, 203–211 (2017).

148. Sin, W. C. & Lim, C. L. Breast cancer stem cells-from origins to targeted therapy. *Stem Cell Investig* **4**, 96 (2017).

149. Sjöström, M. *et al.* Stem cell biomarker ALDH1A1 in breast cancer shows an association with prognosis and clinicopathological variables that is highly cut-off dependent. *J. Clin. Pathol.* **68**, 1012–1019 (2015).

150. Wind, N. S. & Holen, I. Multidrug resistance in breast cancer: from in vitro models to clinical studies. *Int J Breast Cancer* **2011**, 967419 (2011).

151. Chen, Z. *et al.* Mammalian drug efflux transporters of the ATP binding cassette (ABC) family in multidrug resistance: A review of the past decade. *Cancer Lett.* **370**, 153–164 (2016).

152. Bozorgi, A., Khazaei, M. & Khazaei, M. R. New Findings on Breast Cancer Stem Cells: A Review. *J Breast Cancer* **18**, 303–312 (2015).

153. Kumar, A. P. *et al.* Manganese superoxide dismutase is a promising target for enhancing chemosensitivity of basal-like breast carcinoma. *Antioxid. Redox Signal.* **20**, 2326–2346 (2014).

154. Carroll, J., He, J., Ding, S., Fearnley, I. M. & Walker, J. E. Persistence of the permeability transition pore in human mitochondria devoid of an assembled ATP synthase. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 12816–12821 (2019).

155. Chandriani, S. *et al.* A core MYC gene expression signature is prominent in basal-like breast cancer but only partially overlaps the core serum response. *PLoS ONE* **4**, e6693 (2009).

156. Alexandrou, S. *et al.* The Proliferative and Apoptotic Landscape of Basal-like Breast Cancer. *Int J Mol Sci* **20**, (2019).

157. Bareche, Y. *et al.* Unravelling triple-negative breast cancer molecular heterogeneity using an integrative multiomic analysis. *Ann. Oncol.* **29**, 895–902 (2018).

158. Apasu, J. E. *et al.* Neuronal calcium sensor 1 (NCS1) promotes motility and metastatic spread of breast cancer cells in vitro and in vivo. *FASEB J.* **33**, 4802–4813 (2019).

159. Grosshans, H. K., Fischer, T. T., Steinle, J. A., Brill, A. L. & Ehrlich, B. E. Neuronal Calcium Sensor 1 is up-regulated in response to stress to promote cell survival and motility in cancer cells. *Mol Oncol* **14**, 1134–1151 (2020).

160. Yu, R. *et al.* Elevated limb-bud and heart development (LBH) expression indicates poor prognosis and promotes gastric cancer cell proliferation and invasion via upregulating Integrin/FAK/Akt pathway. *PeerJ* **7**, e6885 (2019).

161. Jiang, Y. *et al.* Overexpression of Limb-Bud and Heart (LBH) promotes angiogenesis in human glioma via VEGFA-mediated ERK signalling under hypoxia. *EBioMedicine* **48**, 36–48 (2019).

162. Chung, C. Y.-S. *et al.* Covalent targeting of the vacuolar H + -ATPase activates autophagy via mTORC1 inhibition. *Nature Chemical Biology* **15**, 776–785 (2019).

163. Rosilio, C. *et al.* L-type amino-acid transporter 1 (LAT1): a therapeutic target supporting growth and survival of T-cell lymphoblastic lymphoma/T-cell acute lymphoblastic leukemia. *Leukemia* **29**, 1253–1266 (2015).

164. Rabanal-Ruiz, Y. & Korolchuk, V. I. mTORC1 and Nutrient Homeostasis: The Central Role of the Lysosome. *Int J Mol Sci* **19**, (2018).

165. Rabanal-Ruiz, Y., Otten, E. G. & Korolchuk, V. I. mTORC1 as the main gateway to autophagy. *Essays Biochem* **61**, 565–584 (2017).

166. Cotzomi-Ortega, I., Aguilar-Alonso, P., Reyes-Leyva, J. & Maycotte, P. Autophagy and Its Role in Protein Secretion: Implications for Cancer Therapy. *Mediators Inflamm* **2018**, (2018).

167. YAN, M.-M., NI, J.-D., SONG, D., DING, M. & HUANG, J. Interplay between unfolded protein response and autophagy promotes tumor drug resistance. *Oncol Lett* **10**, 1959–1969 (2015).

168. Senft, D. & Ronai, Z. A. UPR, autophagy, and mitochondria crosstalk underlies the ER stress response. *Trends Biochem. Sci.* **40**, 141–148 (2015).

169. O'Conor, C. J., Chen, T., González, I., Cao, D. & Peng, Y. Cancer stem cells in triple-negative breast cancer: a potential target and prognostic marker. *Biomark Med* **12**, 813–820 (2018).

170. Yang, F. *et al.* Evaluation of Breast Cancer Stem Cells and Intratumor Stemness Heterogeneity in Triple-negative Breast Cancer as Prognostic Factors. *Int. J. Biol. Sci.* **12**, 1568–1577 (2016).

171. Drago-Ferrante, R. *et al.* Suppressive role exerted by microRNA-29b-1-5p in triple negative breast cancer through SPIN1 regulation. *Oncotarget* **8**, 28939–28958 (2017).

172. Ricardo, S. *et al.* Breast cancer stem cell markers CD44, CD24 and ALDH1: expression distribution within intrinsic molecular subtype. *J. Clin. Pathol.* **64**, 937–946 (2011).

173. Kong, D., Hughes, C. J. & Ford, H. L. Cellular Plasticity in Breast Cancer Progression and Therapy. *Front Mol Biosci* **7**, 72 (2020).

174. Charafe-Jauffret, E., Ginestier, C. & Birnbaum, D. Breast cancer stem cells: tools and models to rely on. *BMC Cancer* **9**, 202 (2009).

175. Kwon, D. *et al.* The Effect of Fetal Bovine Serum (FBS) on Efficacy of Cellular Reprogramming for Induced Pluripotent Stem Cell (iPSC) Generation. *Cell Transplant* **25**, 1025–1042 (2016).

176. Guttilla, I. K. *et al.* Prolonged mammosphere culture of MCF-7 cells induces an EMT and repression of the estrogen receptor by microRNAs. *Breast Cancer Res. Treat.* **132**, 75–85 (2012).

177. Volk-Draper, L. D., Rajput, S., Hall, K. L., Wilber, A. & Ran, S. Novel Model for Basaloid Triple-negative Breast Cancer: Behavior In Vivo and Response to Therapy. *Neoplasia* **14**, 926–942 (2012).

178. Pallegar, N. K., Garland, C. J., Mahendralingam, M., Viloria-Petit, A. M. & Christian, S. L. A Novel 3-Dimensional Co-culture Method Reveals a Partial Mesenchymal to Epithelial Transition in Breast Cancer Cells Induced by Adipocytes. *J Mammary Gland Biol Neoplasia* **24**, 85–97 (2019).

179. Zhu, H., Yang, J. & Yang, S. MicroRNA-103a-3p potentiates chemoresistance to cisplatin in non-small cell lung carcinoma by targeting neurofibromatosis 1. *Exp Ther Med* **19**, 1797–1805 (2020).

180. Hua, L., Zhu, G. & Wei, J. MicroRNA-1 overexpression increases chemosensitivity of non-small cell lung cancer cells by inhibiting autophagy related 3-mediated autophagy. *Cell Biol. Int.* **42**, 1240–1249 (2018).

181. He, Y. *et al.* MicroRNA-135b regulates apoptosis and chemoresistance in colorectal cancer by targeting large tumor suppressor kinase 2. *Am J Cancer Res* **5**, 1382–1395 (2015).

182. Zhao, Z., Zhang, L., Yao, Q. & Tao, Z. miR-15b regulates cisplatin resistance and metastasis by targeting PEBP4 in human lung adenocarcinoma cells. *Cancer Gene Ther.* **22**, 108–114 (2015).

183. He, Y. *et al.* MiR-153-5p promotes sensibility of colorectal cancer cells to oxaliplatin via targeting Bcl-2-mediated autophagy pathway. *Biosci. Biotechnol. Biochem.* 1–7 (2020) doi:10.1080/09168451.2020.1760784.

184. He, T.-G. *et al.* Tumor Suppressor miR-184 Enhances Chemosensitivity by Directly Inhibiting SLC7A5 in Retinoblastoma. *Front Oncol* **9**, 1163 (2019).

185. Jin, Y. *et al.* Overcoming stemness and chemoresistance in colorectal cancer through miR-195-5p-modulated inhibition of notch signaling. *Int. J. Biol. Macromol.* **117**, 445–453 (2018).

186. Zhou, L. *et al.* Down-regulation of miR-30a-5p is Associated with Poor Prognosis and Promotes Chemoresistance of Gemcitabine in Pancreatic Ductal Adenocarcinoma. *J Cancer* **10**, 5031–5040 (2019).

187. Xu, W., Jiang, H., Zhang, F., Gao, J. & Hou, J. MicroRNA-330 inhibited cell proliferation and enhanced chemosensitivity to 5-fluorouracil in colorectal cancer by directly targeting thymidylate synthase. *Oncol Lett* **13**, 3387–3394 (2017).

188. Niu, Q., Liu, Z., Gao, J. & Wang, Q. MiR-338-3p Enhances Ovarian Cancer Cell Sensitivity to Cisplatin by Downregulating WNT2B. *Yonsei Med. J.* **60**, 1146–1156 (2019).

189. Yu, G. *et al.* MicroRNA-429 sensitizes pancreatic cancer cells to gemcitabine through regulation of PDCD4. *Am J Transl Res* **9**, 5048–5055 (2017).

190. Jin, Y. *et al.* miR-450b-5p Suppresses Stemness and the Development of Chemoresistance by Targeting SOX2 in Colorectal Cancer. *DNA Cell Biol.* **35**, 249–256 (2016).

191. Li, L. *et al.* MicroRNA-584 Impairs Cellular Proliferation and Sensitizes Osteosarcoma Cells to Cisplatin and Taxanes by Targeting CCN2. *Cancer Manag Res* **12**, 2577–2587 (2020).

192. Lv, L. *et al.* MiR-193a-3p promotes the multi-chemoresistance of bladder cancer by targeting the HOXC9 gene. *Cancer Lett.* **357**, 105–113 (2015).

193. Li, Q. *et al.* Circulating miR-19a and miR-205 in serum may predict the sensitivity of luminal A subtype of breast cancer patients to neoadjuvant chemotherapy with epirubicin plus paclitaxel. *PLoS ONE* **9**, e104870 (2014).

194. Ding, J., Zhao, Z., Song, J., Luo, B. & Huang, L. MiR-223 promotes the doxorubicin resistance of colorectal cancer cells via regulating epithelial-mesenchymal transition by targeting FBXW7. *Acta Biochim. Biophys. Sin. (Shanghai)* **50**, 597–604 (2018).

195. Zhou, Y., He, A., Zhang, L. & Yi, G. MiR-744 mediates the Oxaliplatin chemoresistance in colorectal cancer through inhibiting BIN1. *Neoplasma* **67**, 296–303 (2020).

196. Munoz, J. L., Rodriguez-Cruz, V. & Rameshwar, P. High expression of miR-9 in CD133+ glioblastoma cells in chemoresistance to temozolomide. *J Cancer Stem Cell Res* **3**, (2015).

197. Jiang, X.-M. *et al.* microRNA-19a-3p promotes tumor metastasis and chemoresistance through the PTEN/Akt pathway in hepatocellular carcinoma. *Biomed. Pharmacother.* **105**, 1147–1154 (2018).

198. Wu, L. *et al.* miR-96 induces cisplatin chemoresistance in non-small cell lung cancer cells by downregulating SAMD9. *Oncol Lett* **11**, 945–952 (2016).

199. He, D.-X. *et al.* Methylation-regulated miR-149 modulates chemoresistance by targeting GlcNAc N-deacetylase/N-sulfotransferase-1 in human breast cancer. *FEBS J.* **281**, 4718–4730 (2014).

200. Yang, X. *et al.* miR-200b regulates epithelial-mesenchymal transition of chemo-resistant breast cancer cells by targeting FN1. *Discov Med* **24**, 75–85 (2017).

201. Hu, Y. *et al.* miRNA-205 targets VEGFA and FGF2 and regulates resistance to chemotherapeutics in breast cancer. *Cell Death Dis* **7**, e2291 (2016).

202. Tan, W. *et al.* Metformin mediates induction of miR-708 to inhibit self-renewal and chemoresistance of breast cancer stem cells through targeting CD47. *J. Cell. Mol. Med.* **23**, 5994–6004 (2019).

203. Liang, Z., Feng, A. & Shim, H. MicroRNA-30c-regulated HDAC9 mediates chemoresistance of breast cancer. *Cancer Chemother. Pharmacol.* **85**, 413–423 (2020).

204. Xia, W. *et al.* MicroRNA-423 Drug Resistance and Proliferation of Breast Cancer Cells by Targeting ZFP36. *Onco Targets Ther* **13**, 769–782 (2020).

205. Liu, T. *et al.* MicroRNA-1 down-regulates proliferation and migration of breast cancer stem cells by inhibiting the Wnt/β-catenin pathway. *Oncotarget* **6**, 41638–41649 (2015).

206. Xiao, Y. *et al.* Integrin α5 down-regulation by miR-205 suppresses triple negative breast cancer stemness and metastasis by inhibiting the Src/Vav2/Rac1 pathway. *Cancer Lett.* **433**, 199–209 (2018).

207. Zhao, G. *et al.* MiR-153 reduces stem cell-like phenotype and tumor growth of lung adenocarcinoma by targeting Jagged1. *Stem Cell Res Ther* **11**, 170 (2020).

208. Liu, T., Wu, X., Chen, T., Luo, Z. & Hu, X. Downregulation of DNMT3A by miR-708-5p Inhibits Lung Cancer Stem Cell-like Phenotypes through Repressing Wnt/β-catenin Signaling. *Clin. Cancer Res.* **24**, 1748–1760 (2018).

209. Thankamony, A. P., Saxena, K., Murali, R., Jolly, M. K. & Nair, R. Cancer Stem Cell Plasticity - A Deadly Deal. *Front Mol Biosci* **7**, 79 (2020).

210. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).

211. Poli, V., Fagnocchi, L. & Zippo, A. Tumorigenic Cell Reprogramming and Cancer Plasticity: Interplay between Signaling, Microenvironment, and Epigenetics. *Stem Cells Int* **2018**, 4598195 (2018).

212. Tang, D. G. Understanding cancer stem cell heterogeneity and plasticity. *Cell Res.* **22**, 457–472 (2012).

213. The microRNA miR-181c Enhances Chemosensitivity and Reduces Chemoresistance in Breast Cancer Cells via Down-Regulating Osteopontin - PubMed. https://pubmed.ncbi.nlm.nih.gov/30537505/.

214. Chen, M., Wang, M., Xu, S., Guo, X. & Jiang, J. Upregulation of miR-181c contributes to chemoresistance in pancreatic cancer by inactivating the Hippo signaling pathway. *Oncotarget* **6**, 44466–44479 (2015).

215. Zhang, H. *et al.* MiR-139-5p inhibits the biological function of breast cancer cells by targeting Notch1 and mediates chemosensitivity to docetaxel. *Biochem. Biophys. Res. Commun.* **465**, 702–713 (2015).

216. Yi, D., Xu, L., Wang, R., Lu, X. & Sang, J. miR-381 overcomes cisplatin resistance in breast cancer by targeting MDR1. *Cell Biol. Int.* **43**, 12–21 (2019).

217. Yao, Y.-S. *et al.* miR-141 confers docetaxel chemoresistance of breast cancer cells via regulation of EIF4E expression. *Oncol. Rep.* **33**, 2504–2512 (2015).

218. Yu, S.-J. *et al.* MicroRNA-200a confers chemoresistance by antagonizing TP53INP1 and YAP1 in human breast cancer. *BMC Cancer* **18**, 74 (2018).

219. Shen, H. *et al.* MicroRNA-29a contributes to drug-resistance of breast cancer cells to adriamycin through PTEN/AKT/GSK3β signaling pathway. *Gene* **593**, 84–90 (2016).

220. Troschel, F. M. *et al.* miR-142-3p attenuates breast cancer stem cell characteristics and decreases radioresistance in vitro. *Tumour Biol.* **40**, 1010428318791887 (2018).

221. Feng, Z.-M. *et al.* Essential role of miR-200c in regulating self-renewal of breast cancer stem cells and their counterparts of mammary epithelium. *BMC Cancer* **15**, 645 (2015).

222. Kopp, F., Oak, P. S., Wagner, E. & Roidl, A. miR-200c sensitizes breast cancer cells to doxorubicin treatment by decreasing TrkB and Bmi1 expression. *PLoS ONE* **7**, e50469 (2012).

223. Wang, H.-J. *et al.* miR-125b regulates side population in breast cancer and confers a chemoresistant phenotype. *J. Cell. Biochem.* **114**, 2248–2257 (2013).

224. Yang, L.-W. *et al.* miR-155 increases stemness and decitabine resistance in triple-negative breast cancer cells by inhibiting TSPAN5. *Mol. Carcinog.* **59**, 447–461 (2020).

225. Miao, Y. *et al.* MicroRNA-130b targets PTEN to mediate drug resistance and proliferation of breast cancer cells via the PI3K/Akt signaling pathway. *Sci Rep* **7**, 41942 (2017).

226. Ao, X. *et al.* Decreased expression of microRNA-17 and microRNA-20b promotes breast cancer resistance to taxol therapy by upregulation of NCOA3. *Cell Death Dis* **7**, e2463 (2016).

227. Tang, X. *et al.* MicroRNA-16 sensitizes breast cancer cells to paclitaxel through suppression of IKBKB expression. *Oncotarget* **7**, 23668–23683 (2016).

228. Hong, T., Ding, J. & Li, W. miR-7 Reverses Breast Cancer Resistance To Chemotherapy By Targeting MRP1 And BCL2. *Onco Targets Ther* **12**, 11097–11105 (2019).

229. Zheng, Y., Zheng, Y., Lei, W., Xiang, L. & Chen, M. miR-1307-3p overexpression inhibits cell proliferation and promotes cell apoptosis by targeting ISM1 in colon cancer. *Mol. Cell. Probes* **48**, 101445 (2019).

230. Han, S. *et al.* miR-1307-3p Stimulates Breast Cancer Development and Progression by Targeting SMYD4. *J Cancer* **10**, 441–448 (2019).

231. Chen, S., Wang, L., Yao, B., Liu, Q. & Guo, C. miR-1307-3p promotes tumor growth and metastasis of hepatocellular carcinoma by repressing DAB2 interacting protein. *Biomed. Pharmacother.* **117**, 109055 (2019).

232. Guan, L. *et al.* MEIS2C and MEIS2D promote tumor progression via Wnt/β-catenin and hippo/YAP signaling in hepatocellular carcinoma. *J. Exp. Clin. Cancer Res.* **38**, 417 (2019).

233. Chen, W.-T. *et al.* MiR-1307 promotes ovarian cancer cell chemoresistance by targeting the ING5 expression. *J Ovarian Res* **10**, 1 (2017).

234. Qiu, X. & Dou, Y. miR-1307 promotes the proliferation of prostate cancer by targeting FOXO3A. *Biomed. Pharmacother.* **88**, 430–435 (2017).

235. Barnabas, N. & Cohen, D. Phenotypic and Molecular Characterization of MCF10DCIS and SUM Breast Cancer Cell Lines. *Int J Breast Cancer* **2013**, (2013).

236. Duncan, J. S. *et al.* Dynamic reprogramming of the kinome in response to targeted MEK inhibition in triple-negative breast cancer. *Cell* **149**, 307–321 (2012).

237. Isakoff, S. J. *et al.* Breast Cancer–Associated PIK3CA Mutations Are Oncogenic in Mammary Epithelial Cells. *Cancer Res* **65**, 10992–11000 (2005).

238. Yip, P. K., Wong, L.-F., Sears, T. A., Yáñez-Muñoz, R. J. & McMahon, S. B. Cortical overexpression of neuronal calcium sensor-1 induces functional plasticity in spinal cord following unilateral pyramidal tract injury in rat. *PLoS Biol.* **8**, e1000399 (2010).

239. Wang, Z. *et al.* Broad targeting of angiogenesis for cancer prevention and therapy. *Semin Cancer Biol* **35**, S224–S243 (2015).

240. Perrot-Applanat, M. & Di Benedetto, M. Autocrine functions of VEGF in breast tumor cells: adhesion, survival, migration and invasion. *Cell Adh Migr* **6**, 547–553 (2012).

241. Mercurio, A. M., Lipscomb, E. A. & Bachelder, R. E. Non-Angiogenic Functions of VEGF in Breast Cancer. *J Mammary Gland Biol Neoplasia* **10**, 283–290 (2005).

242. Stransky, L., Cotter, K. & Forgac, M. The Function of V-ATPases in Cancer. *Physiol Rev* **96**, 1071–1091 (2016).

243. Feng, S. *et al.* Silencing of atp6v1c1 prevents breast cancer growth and bone metastasis. *Int. J. Biol. Sci.* **9**, 853–862 (2013).

244. Cai, M. *et al.* Atp6v1c1 may regulate filament actin arrangement in breast cancer cells. *PLoS ONE* **9**, e84833 (2014).

245. McConnell, M. *et al.* Osteoclast proton pump regulator Atp6v1c1 enhances breast cancer growth by activating the mTORC1 pathway and bone metastasis by increasing V-ATPase activity. *Oncotarget* **8**, 47675–47690 (2017).

246. Chen, H.-T. *et al.* Crosstalk between autophagy and epithelial-mesenchymal transition and its application in cancer therapy. *Molecular Cancer* **18**, 101 (2019).

247. Castañeda-Gill, J. M. & Vishwanatha, J. K. Antiangiogenic mechanisms and factors in breast cancer treatment. *J Carcinog* **15**, (2016).

248. Hendrix, A. *et al.* Vacuolar H+ ATPase expression and activity is required for Rab27B-dependent invasive growth and metastasis of breast cancer. *Int. J. Cancer* **133**, 843–854 (2013).

249. Zhang, Z., Singh, R. & Aschner, M. Methods for the Detection of Autophagy in Mammalian Cells. *Curr Protoc Toxicol* **69**, 20.12.1-20.12.26 (2016).

250. Almeida, M. I. *et al.* Strand-specific miR-28-5p and miR-28-3p have distinct effects in colorectal cancer cells. *Gastroenterology* **142**, 886-896.e9 (2012).

251. Lv, Y., Yang, H., Ma, X. & Wu, G. Strand-specific miR-28-3p and miR-28-5p have differential effects on nasopharyngeal cancer cells proliferation, apoptosis, migration and invasion. *Cancer Cell Int.* **19**, 187 (2019).

252. Córdova-Rivas, S. *et al.* 5p and 3p Strands of miR-34 Family Members Have Differential Effects in Cell Proliferation, Migration, and Invasion in Cervical Cancer Cells. *Int J Mol Sci* **20**, (2019).

253. Ren, L.-L. *et al.* The distinct role of strand-specific miR-514b-3p and miR-514b-5p in colorectal cancer metastasis. *Cell Death Dis* **9**, 687 (2018).

254. Zhang, Z. *et al.* microRNA arm-imbalance in part from complementary targets mediated decay promotes gastric cancer progression. *Nature Communications* **10**, 4397 (2019).

255. Issartel, J. P. *et al.* The ATP synthase (F0-F1) complex in oxidative phosphorylation. *Experientia* **48**, 351–362 (1992).

256. Nagao, A., Kobayashi, M., Koyasu, S., Chow, C. C. T. & Harada, H. HIF-1-Dependent Reprogramming of Glucose Metabolic Pathway of Cancer Cells and Its Therapeutic Significance. *Int J Mol Sci* **20**, (2019).

257. Kontro, H., Hulmi, J. J., Rahkila, P. & Kainulainen, H. Cellular and tissue expression of DAPIT, a phylogenetically conserved peptide. *Eur J Histochem* **56**, e18 (2012).

258. Imamura, H. *et al.* Evidence for rotation of V1-ATPase. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 2312–2315 (2003).

259. He, J. *et al.* Assembly of the membrane domain of ATP synthase in human mitochondria. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 2988–2993 (2018).

260. Ippolito, L., Giannoni, E., Chiarugi, P. & Parri, M. Mitochondrial Redox Hubs as Promising Targets for Anticancer Therapy. *Front Oncol* **10**, 256 (2020).

261. Van Houten, B., Santa-Gonzalez, G. A. & Camargo, M. DNA repair after oxidative stress: current challenges. *Curr Opin Toxicol* **7**, 9–16 (2018).

262. Sarmiento-Salinas, F. L. *et al.* Breast Cancer Subtypes Present a Differential Production of Reactive Oxygen Species (ROS) and Susceptibility to Antioxidant Treatment. *Front Oncol* **9**, 480 (2019).

263. Sotgia, F. *et al.* Mitochondria 'fuel' breast cancer metabolism: fifteen markers of mitochondrial biogenesis label epithelial cancer cells, but are excluded from adjacent stromal cells. *Cell Cycle* **11**, 4390–4401 (2012).

264. Jia, D. *et al.* Elucidating cancer metabolic plasticity by coupling gene regulation with metabolic pathways. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 3909–3918 (2019).

265. Yu, L. *et al.* Modeling the Genetic Regulation of Cancer Metabolism: Interplay between Glycolysis and Oxidative Phosphorylation. *Cancer Res.* **77**, 1564–1574 (2017).

266. Maxson, M. E. & Grinstein, S. The vacuolar-type H+-ATPase at a glance – more than a proton pump. *J Cell Sci* **127**, 4987–4993 (2014).

267. Cipriano, D. J. *et al.* Structure and regulation of the vacuolar ATPases. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1777**, 599–604 (2008).

268. Huang, Y. & Sadée, W. Membrane transporters and channels in chemoresistance and -sensitivity of tumor cells. *Cancer Letters* **239**, 168–182 (2006).

269. Sennoune, S. R. *et al.* Vacuolar H+-ATPase in human breast cancer cells with distinct metastatic potential: distribution and functional activity. *Am. J. Physiol., Cell Physiol.* **286**, C1443-1452 (2004).

270. Hinton, A. *et al.* Function of a Subunit Isoforms of the V-ATPase in pH Homeostasis and in Vitro Invasion of MDA-MB231 Human Breast Cancer Cells. *J Biol Chem* **284**, 16400–16408 (2009).

271. Kawai, Y., Kaidoh, M., Yokoyama, Y. & Ohhashi, T. Cell surface F1/FO ATP synthase contributes to interstitial flow-mediated development of the acidic microenvironment in tumor tissues. *Am. J. Physiol., Cell Physiol.* **305**, C1139-1150 (2013).

272. Zhang, X. *et al.* Dual functions of a monoclonal antibody against cell surface F1F0 ATP synthase on both HUVEC and tumor cells. *Acta Pharmacol. Sin.* **29**, 942–950 (2008).

273. Chi, S. L. & Pizzo, S. V. Angiostatin is directly cytotoxic to tumor cells at low extracellular pH: a mechanism dependent on cell surface-associated ATP synthase. *Cancer Res.* **66**, 875–882 (2006).

274. Scotet, E. *et al.* Tumor recognition following Vgamma9Vdelta2 T cell receptor interactions with a surface F1-ATPase-related structure and apolipoprotein A-I. *Immunity* **22**, 71–80 (2005).

275. Arakaki, N. *et al.* Possible role of cell surface H+ -ATP synthase in the extracellular ATP synthesis and proliferation of human umbilical vein endothelial cells. *Mol. Cancer Res.* **1**, 931–939 (2003).

276. Chang, H.-Y., Huang, T.-C., Chen, N.-N., Huang, H.-C. & Juan, H.-F. Combination therapy targeting ectopic ATP synthase and 26S proteasome induces ER stress in breast cancer cells. *Cell Death & Disease* **5**, e1540–e1540 (2014).

277. Huang, T.-C. *et al.* Targeting therapy for breast carcinoma by ATP synthase inhibitor aurovertin B. *J. Proteome Res.* **7**, 1433–1444 (2008).

278. Wen-Li, Z. *et al.* Inhibition of the ecto-beta subunit of F1F0-ATPase inhibits proliferation and induces apoptosis in acute myeloid leukemia cell lines. *J Exp Clin Cancer Res* **31**, 92 (2012).

279. Moser, T. L. *et al.* Endothelial cell surface F1-F0 ATP synthase is active in ATP synthesis and is inhibited by angiostatin. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6656–6661 (2001).

280. Speransky, S. *et al.* A novel RNA aptamer identifies plasma membrane ATP synthase beta subunit as an early marker and therapeutic target in aggressive cancer. *Breast Cancer Res Treat* **176**, 271–289 (2019).

281. Merk, H. *et al.* Inhibition of the V-ATPase by Archazolid A: A New Strategy to Inhibit EMT. *Mol Cancer Ther* **16**, 2329–2339 (2017).

282. Wang, J. *et al.* ATP6L promotes metastasis of colorectal cancer by inducing epithelial-mesenchymal transition. *Cancer Sci* **111**, 477–488 (2020).

283. Kroemer, G., Mariño, G. & Levine, B. Autophagy and the integrated stress response. *Mol Cell* **40**, 280–293 (2010).

284. Kenny, T. C., Craig, A. J., Villanueva, A. & Germain, D. Mitohormesis Primes Tumor Invasion and Metastasis. *Cell Rep* **27**, 2292-2303.e6 (2019).

285. Xu, Z. *et al.* Targeting PI3K/AKT/mTOR-mediated autophagy for tumor therapy. *Appl Microbiol Biotechnol* **104**, 575–587 (2020).