Dissertation

submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree
of Doctor of Natural Sciences

Presented by
Chiara Giacomelli, M.Sc. in Medical, Molecular, and Cellular Biotechnologies
Born in: Milano, Italy

Oral examination: 5th February 2018

Dissecting the miRNome impact on metastatic pathways in Triple Negative Breast Cancer

Referees: Prof. Dr. Stefan Wiemann
Prof. Dr. Michael Boutros

Declaration of Authorship

I hereby declare that the work presented in my dissertation was carried out between November 2012 and November 2017 under the supervision of Prof. Dr. Stefan Wiemann in the group of Molecular Genome Analysis at the German Cancer Research Center (DKFZ, Heidelberg, Germany).

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

Table of Contents

Summary

Metastatic disease at distant sites is the main cause of death for patients of breast cancer. Among them, patients diagnosed with the subtype termed Triple Negative (TNBC) are affected the most by both residual disease after treatment and distant metastases. It is therefore of utmost importance to investigate more in depth the role of pathways related to TNBC development and metastasization, possibly identifying novel therapeutic or diagnostic targets. In this PhD project, the effect of the miRNome (i.e. known miRNAs encoded in the genome) was analysed on three pathways relevant for TNBC metastatic processes. A high-throughput proteomic approach termed Reverse Phase Protein Arrays (RPPA) was exploited to quantify the effect of miRNAs on the expression of 62 proteins belonging to c-Met, Integrin, and WNT/β-catenin signalling pathways. Then, a novel network analysis method was developed, taking into consideration the role of the proteins within the pathway of interest, and therefore contextualising the miRNA:target interactions. miRNAs identified as modulators of WNT/β-catenin signalling were tested functionally by means of a reporter gene assay, characterising miR-193b, -409, -494-, and -92b as strong repressors of the pathway. Their ability to regulate proliferation was assessed in a context of pathway overactivation, revealing that miR-193b, -409, and -494 reduce it. Analysis of stem-associated surface markers CD44 and CD24 showed that miR-193b is able to decrease the CD44+ population. Additionally, direct molecular interactions with predicted target sequences in 3'UTRs of mRNAs corresponding to the proteins probed were validated for miR-193b and miR-494.

In summary, the findings of this dissertation describe the role of miRNAs as putative repressors or activators of metastatic pathways in TNBC, validating miR-193b, miR-409, and miR-494 as highly integrated within the network of WNT/β-catenin signalling. Furthermore, the RPPA results provide a useful resource available to the scientific community for further studies related to miRNA function in TNBC.

Zusammenfassung

Metastasen in entfernten Organen sind die Haupttodesursache bei Bruskrebspatienten. Patienten, die mit dem triple-negativen Subtyp (TNBC) diagnostiziert wurden, sind nach der Behandlung am häufigsten von dem übriggebliebenen Tumor und von Fernmetastasen betroffen. Daher ist es von essentieller Bedeutung, die Rolle von Signalwegen im Zusammenhang mit der Entstehung von TNBC und dessen Metastasierung eingehend zu untersuchen und möglicherweise neue therapeutische oder diagnostische Möglichkeiten zu identifizieren. In dieser Dissertation wurde der Effekt vom miRNome (d.h. allen bekannten miRNAs) auf drei für TNBC-Metastasen relevante Signalwege untersucht. Die hochdurchsatz Proteom-Methode der sogenannten Reverse Phase Protein Arrays (RPPA) wurde genutzt um die Wirkung von miRNAs auf die Expression von 62 Proteinen, die zu den c-Met, Integrin und WNT/β-Catenin-Signalwegen gehören, zu quantifizieren. Weiterhin wurde eine neue Methode für eine Netzwerkanalyse entwickelt, welche die Rolle der analysierten Proteine innerhalb des jeweilig fokussierten Signalweges berücksichtigt und somit miRNA-Protein Interaktionen in den funktionellen Kontext setzt. miRNAs, welche als Modulatoren des WNT/β-Catenin-Signalweges eingeordnet worden sind, wurden durch einen Reportergen-Assay funktionell getestet, wodurch miR-193b, -409, -494- und -92b als starke Repressoren des Signalweges charakterisiert worden sind. Ihre Fähigkeit die Proliferation zu regulieren wurde im Zusammenhang mit der Überaktivierung des Signalweges untersucht. Dies zeigte, dass miR-193b, -409 und -494 die induzierte Überaktivierung des WNT/β-Catenin-Signalweges reduzieren. Die Analyse von Stammzell-assoziierten Oberflächenmarkern CD44 und CD24 ergab, dass miR-193b in der Lage ist, die CD44+ Population zu verringern. Zusätzlich wurden direkte molekulare Wechselwirkungen von miR-193b und miR-494 mit vorhergesagten Zielsequenzen in 3'UTRs von mRNAs validiert.

Zusammengefasst beschreiben die Ergebnisse dieser Dissertation die funktionellen Rollen von miRNAs als putative Repressoren oder Aktivatoren von metastatischen Signalwegen in TNBC. miR-193b, miR-409 und miR-494 wurden als hochintegriert innerhalb des WNT/β-Catenin-Signalweges identifiziert. Darüber hinaus stellen die Ergebnisse der RPPA-Analyse einen umfangreichen Datensatz dar, der von der wissenschaftlichen Gemeinschaft für weitere Studien über miRNA-Funktionen im TNBC genutzt werden kann.

Abbreviations

AU	Arbitrary Units
BC	Breast Cancer
BCA	bicinchoninic acid assay
BL2	Basal-Like 2
BSA	Bovine Serum Albumine
DNA	deoxyribonucleic acid
ds	double stranded
EMT	Epithelial to Mesenchymal Transition
ER	Estrogen Receptor
EtOH	Ethanol
FBS	Fetal Bovine Serum
FC	Fold Change
FWD	forward
KD	Knock Down
KO	Knock Out
LF	lipofectamine
LumA	Luminal A
LumB	Luminal B
MBS	miRNA Binding Site
miRNA	microRNA
MSL	Mesenchymal Stem-Like
NF-kB	nuclear factor kappa-light-chain-enhancer of activated B cells
OE	Over Expression

ON	Over-night
ORF	Open Reading Frame
PBS	Phosphate-buffered saline
PCR	Polymerase Chain reaction
PFA	Paraformaldehyde
PR	Progesterone Receptor
q-val	q-value
REV	reverse
RNA	ribonucleic acid
RPPA	Reverse Phase Protein Array
ss	single stranded
TGFβ	Transforming Growth Factor β
TNBC	Triple Negative Breast Cancer
TS	TargetScan
UTR	Untranslated Region
β-cat	β-catenin

1. Introduction

1.1. Breast cancer

Breast cancer (BC) is the most diagnosed and the deadliest type of cancer in the female population worldwide, accounting for more than 1.6 million new cases and more than 500'000 deaths per year (Torre, Bray et al. 2015). Tumours of the breast tissue are highly heterogeneous and different subtypes with specific molecular and clinical features have now been identified. BC patients can be classified by immunohistological characteristics, such as expression of hormonal receptors (Oestrogen and/or Progesterone, ER/PR) or human epidermal growth factor 2 (*ERBB2*). Tumours which do not express any of these receptors are defined as Triple Negative Breast Cancers (TNBC) (Foulkes, Smith et al. 2010). With the advent of genomics, gene expression profiles of BC patients have been analysed and molecular 'intrinsic' subtypes defined by clusters of genes concordantly regulated. The five classes first proposed by Perou and colleagues are Luminal A, Luminal B, ERBB2+, Basal, and Normal Breast-like (Perou, Sorlie et al. 2000). These signatures could discriminate subtypes with distinct clinical outcomes, describing for the first time intrinsic differences in the biology of breast cancers (Sorlie, Perou et al. 2001) (*fig. 1.1*).

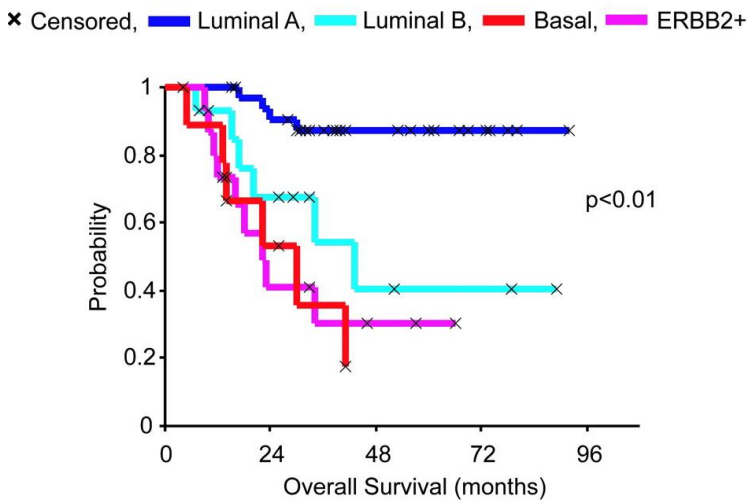


Figure 1-1 Survival of patients classified by intrinsic molecular subtypes. Kaplan-Meier curves describe the overall survival of a cohort of 72 patients with locally advanced cancer (Sorlie, Tibshirani et al. 2003).

Although not with a complete overlap in classification, intrinsic molecular subtypes are correlated with the receptor status observed by

immunohistochemistry. Indeed, breast cancers of the luminal types display higher levels of ER and PR among the different tumour types, with the A type showing the highest expression. Luminal B subtype is instead characterized by a lower expression of ER and PR, and an increased expression of cell cycle-associated genes (*CCNE1* and *CCNB1*), as well as proliferation-associated genes (*MKI67*) (Ades, Zardavas et al. 2014). The ERBB2+ subtype is characterized by amplification and/or overexpression of *ERBB2*, together with other genes found in the genomic area termed *ERBB2* amplicon (17q22.24). The majority of basal-like tumours are triple negative at the histological level, as well as being positive for the Epidermal Growth Factor Receptor (*EGFR*) and cytokeratin 5 or 17 (*KRT5/17*) (Sorlie 2004).

The American National Cancer Institute releases every year statistics regarding breast cancer epidemiology in the United States. From this broad datasets it is possible to appreciate how in the US the 5-year survival of BC patients has dramatically increased since 1975, incrementing from 75.2% to a staggering 89.7% for the years 2007-2013. The majority of patients greatly benefit from the early detection strategies that are now enacted in developed countries. This is not the case in low- and middle-income countries, indicating that easily accessible diagnostic strategies are needed to improve the survival rates. Despite the increase in survival in recent years, relative survival of women with distant metastases at diagnosis is still very low (26.9%) (SEER Cancer Stat Facts: Female Breast Cancer. National Cancer Institute. Bethesda, MD). Metastatic recurrent disease at distance sites remains the main cause of death for all breast cancer patients (Weigelt, Peterse et al. 2005), an extremely relevant issue for TNBC patients in particular. Indeed, they present the highest percentages in both local and distant recurrences, with metastases more common in brain and lungs (Lehmann, Jovanovic et al. 2016). As well, median duration of survival with distant metastasis is the lowest for TNBC (0.5 years) compared to other subtypes (2.2 for LumA, 1.6 LumB, 0.7 Her2+) (Kennecke, Yerushalmi et al. 2010).

Patients affected by Luminal A/B or Her2+ tumours greatly benefit from an array of targeted therapies that have been developed in recent years. ER+ cancer

patients can be treated by endocrine therapies such as Tamoxifen or Fulvestrant, which affect the ER-dependent growth. For ERBB2+ patients, monoclonal antibodies aimed at the Her2 receptor are available (Trastuzumab and Pertuzumab), as well as dual kinase inhibitors that affect the downstream signalling (Lapatinib). Few (~5%) of TNBC patients present mutations in *BRCA1* or *BRCA2* genes and can be effectively treated with PARP inhibitors (Robson, Im et al. 2017). However, for the majority of TNBC patients the only available treatments remain chemotherapy, either in adjuvant or neoadjuvant regimens. A promising novel target for TNBC tumours is EGFR, being overexpressed in 45-70% of patients. However, recent reports of clinical trials with molecules available to inhibit EGFR at various levels of the signalling cascade have shown very poor results (Nakai, Hung et al. 2016). While ongoing trials test additional therapeutic combination to confirm or refute the importance of targeting EGFR in TNBC treatment, researchers approached additional pathways which could be active in parallel to EGFR to sustain tumour growth. The Hepatocyte Growth Factor (HGF) receptor c-Met is another RTK found overexpressed in a portion of TNBC patients, where it was additionally found associated to worse disease free survival when co-expressed with EGFR (Kim, Lee et al. 2013). Additionally, its overexpression and/or its activity (measured as phospho-c-Met) have been associated multiple times to worse overall survival and disease free survival regardless of the breast cancer subtype, hinting perhaps at a broader relevance for possible treatments options (Minuti and Landi 2015).

1.2. Triple Negative Breast Cancer subtypes

As already mentioned, TNBCs are defined histologically by the absence of expression of ER/PR/Her2 receptors, therefore representing a highly heterogeneous group of tumours within itself (Denkert, Liedtke et al. 2017). Indeed, patients classified as TNBC present different responses to chemotherapies, ranging from a pathological complete response associated with high rates of survival (30 to 40% of patients), to residual disease after neoadjuvant treatment, a prognostic factor of extremely poor survival. While few patients will not progress to metastatic disease within the first 3 year from diagnosis, the majority will instead recur during this time and have the worse

survival rate throughout the different breast cancer subtypes (Liedtke, Mazouni et al. 2008).

Various attempts have been made to classify TNBC patients into distinct subtypes with clinical and intrinsic molecular characteristics, possibly stratifying them into groups which benefit or not from chemotherapy. In a study from the Pietenpol lab, data from 587 TNBC patients was extracted from 21 breast cancer publicly available datasets. Then, by unsupervised hierarchically clustering, six discrete subtypes were identified, each characterized by its specific gene expression profile. The resulting profiles were then analysed for pathway signatures, therefore identifying enrichment of specific signalling pathways in each subtype. The subtypes identified were Basal-Like1 and 2 (BL1 and BL2), Immunomodulatory (IM), mesenchymal (M), Luminal Androgen Receptor (LAR), and Mesenchymal Stem-Like (MSL). Additionally, by gene expression analysis, they classified the cell lines currently in use as TNBC models into the same six clinical subtypes, thus allowing researchers to choose experimental systems which hold a translational value. Indeed, this was proven in the study by targeted pharmacological inhibition of the pathways enriched in the subtypes (Lehmann, Bauer et al. 2011). The classification was refined in 2016 based on gene expression of micro-dissected tissues in order the remove the stromal and immune components. Indeed, the author showed that for some subtypes gene expression results were influenced by the non-tumour tissue, and therefore confounded the interpretation of the pathways relevant for the disease biology. Thanks to this analysis, IM and MSL signatures were identified as strongly dependent on the stromal cells gene expression. Therefore, the six subtypes were refined to four (BL1, BL2, M, and LAR) and patients previously classified as IM or MSL were reassigned. Further analyses of pathological complete response in the subtypes showed how BL2 and LAR display the lowest rates (0% and 10%, respectively). While LAR patients could benefit from anti-androgen receptor therapies which are under evaluation, BL2 patients face once again only the chemotherapeutic possibilities, with clear poor outcomes (Denkert, Liedtke et al. 2017).

In summary, TNBC as a heterogeneous disease and the BL2 subtype specifically require deeper biological investigations to fully understand pathological mechanisms which underlie its clinical aggressiveness. In particular, due to the highest percentages in both local and distant recurrences and the lowest median survival with distant metastasis, these processes remain seminal to examine.

1.3. microRNAs: small non-coding regulators of gene expression

microRNAs (miRNAs) are small (~22 nt. long), single stranded, non-coding RNA molecules that predominantly repress gene expression post-transcriptionally upon binding to 3' untranslated regions (UTRs) of target messenger RNAs (mRNAs). Since their first discovery in *C.elegans*, they have now been found encoded in eukaryotic genomes, such as the ones of plants, animals, and protists, as well as in viral genomes. Published miRNA sequences and annotation are listed in the database miRBase. Its latest release (v. 21, from June 2014) contains information for 28'645 genomic miRNA loci which giving rise to almost 36'000 mature sequences across all the organisms characterized (Kozomara and Griffiths-Jones 2014).

In *Homo sapiens* (GRCh38 genome assembly), 1881 precursors and 2588 mature miRNA sequences have been described, either experimentally or predicted by high homology with miRNAs discovered in other species. Analysis of conserved miRNA binding sites on 3'UTRs of human protein-coding transcript estimated that more than 60% of them are under selective pressure to maintain miRNA binding sites (Friedman, Farh et al. 2009). This widespread effect places miRNAs as regulators of virtually any biological process present in human development, physiology, as well as pathology (Ameres and Zamore 2013).

1.3.1. miRNA biogenesis in animals

miRNAs have been found located in diverse genomic contexts, encoded as single or polycistronic units. The majority of miRNA genes are positioned in intronic portions of both coding and non-coding transcripts, while few have been characterized in exonic regions as well. They are transcribed as long sequences (often more than 1kB) termed primary miRNA (pri-miRNA), containing local stem-loop structures that comprise the mature miRNAs. Their transcription is mainly mediated by RNA polymerase II (RNA Pol II) under the control of RNA pol II-associated transcriptional regulators. As for other RNA pol II-dependent genes, their transcription can be controlled by epigenetic modifications and specific transcription factors.

The pri-miRNA is processed in the nucleus by the microprocessor complex, comprised of the enzyme Drosha, an RNase III-type endonuclease, and its essential co-factor DGCR8. The complex recognizes the pri-miRNA and cleaves the stem-loop roughly 11 nt. away from the basal junction between ssRNA and dsRNA portion, releasing a hairpin ~65 nt. long, termed precursor miRNA (pre-miRNA) (*fig 1.2*).

The pre-miRNA is transported via the protein exportin 5 (EXP5) from the nucleus into the cytoplasm, where it undergoes another round of endonucleoytic cleavage. The RNase Dicer is the main effector of this process. Thanks to its "molecular ruler" structure, it is able to cut the pre-miRNA terminal loop, generating a double stranded RNA molecule which is roughly 22 nt. long. This step is frequently facilitated by dsRNA biding proteins (dsRBP) such as Tar RNA Binding Protein (TRBP) or PACT (also termed PRKRA). However, their presence doesn't seem necessary for pre-miRNA processing, since purified Dicer was found to be functional in vitro even when uncoupled from these dsRBPs.

This short RNA duplex is subsequently loaded onto an Argonaute protein (AGO1-4), where the two strands are separated. Upon unwinding, only one of the two strands is retained onto Ago to form the miRNA-Induced Silencing Complex (miRISC). Analysis of preferential strand loading has demonstrated that the strand with the relatively least stable 5' end is selected as a 'guide' and remains loaded, while the other strand, termed 'passenger' is promptly degraded. In few cases the selection seems dependent on the context, as both arms of the miRNA can be found expressed in different cellular settings (*fig 1.3*).

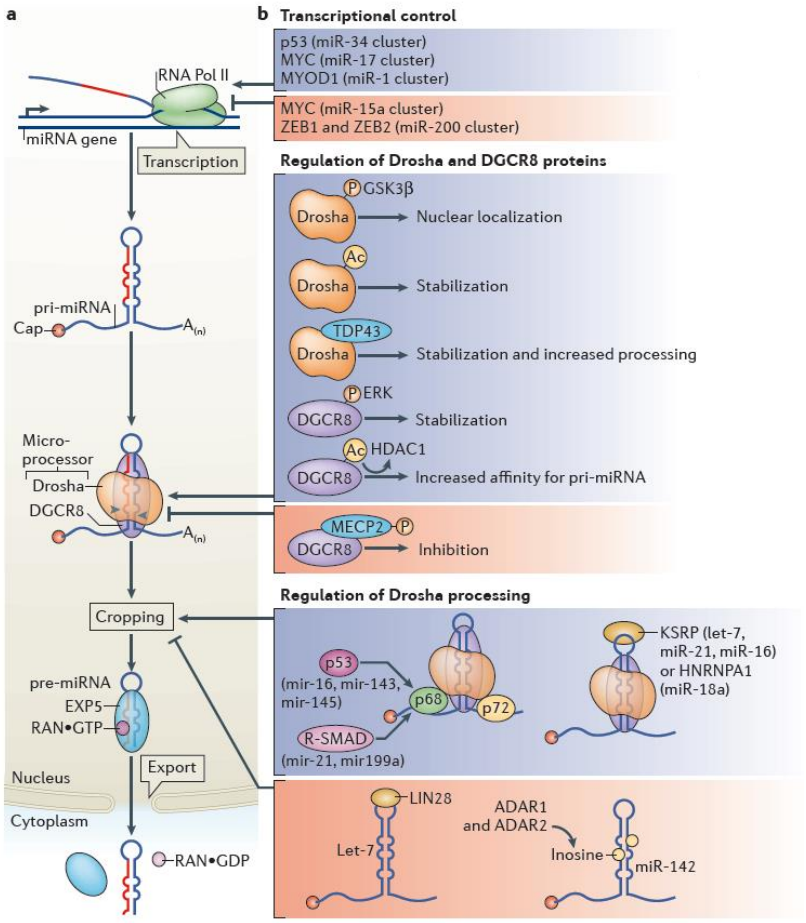


Figure 1-2 microRNA transcription an nuclear processing. A) miRNA genes are transcribed by RNA pol II into long pri-miRNA transcripts which are processed by the microprocessor into pre-miRNAs, which in turn are exported into the cytoplasm. B) Known mechanisms and modifications that regulate nuclear processing. (Figure from Ha and Kim, 2014, Nat Rev Mol Cell Biol)

Both nuclear and cytoplasmic biogenesis steps have been studied in depth to identify possible regulators of these processes. Multiple post-translational modifications of the enzymes, as well as pri-/pre-/mature miRNA sequence alterations have been identified and are outlined in *figures 1.2* (nuclear) and *1.3* (cytoplasmic). This plethora of mechanisms indicates how the biogenesis of miRNAs is a tightly controlled process, potentially coupling external stimuli as well as intrinsic cellular conditions to the abundance of miRNAs, in general, as well as for specific ones. (Ha and Kim 2014)

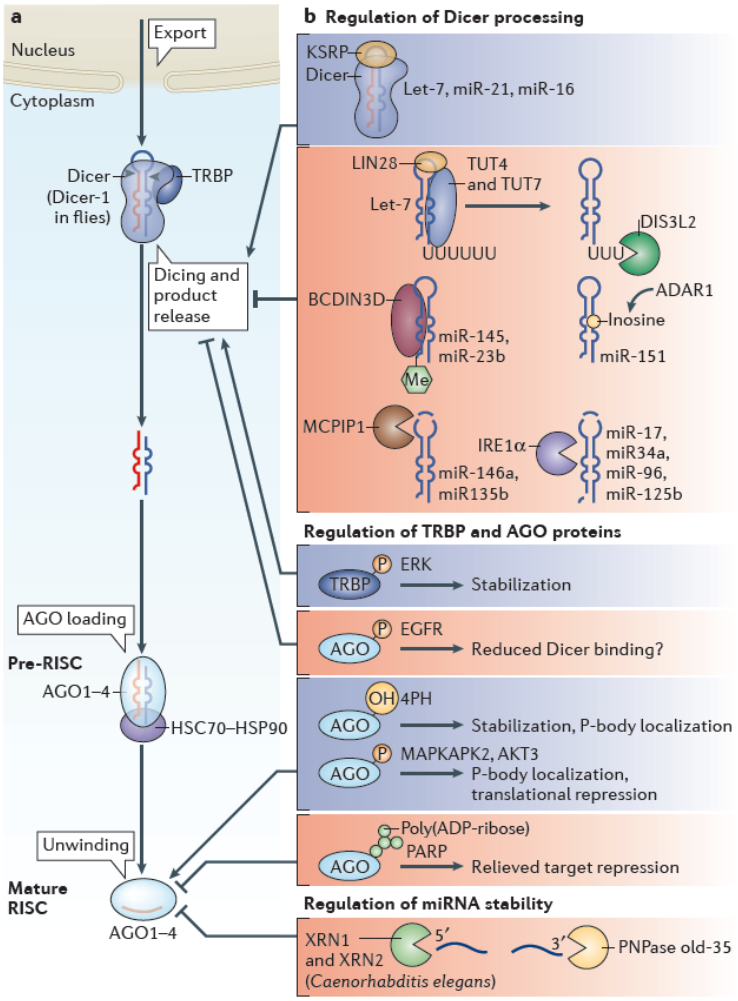


Figure 1-3 Cytoplasmic processing of miRNAs. A) exported pre-miRNA is processed by Dicer and helper proteins such as TRBP. The small dsRNA is released and loaded onto an AGO protein which unwinds the duplex with the help of HSP90/HSC70. The mature miRISC is ready to regulate target genes. B) Known mechanisms and modifications that regulate cytoplasmic miRNA processing. (Figure from Ha and Kim, 2014, Nat Rev Mol Cell Biol)

1.3.2. miRNA function in mammals

Upon incorporation in the RISC, the mature miRNA guides the silencing complex to the target mRNA by base pairing with sequences present in its 3'UTR. The majority of binding sites is characterized by a full complementarity to the miRNA 'seed' only (nt. 2-7). Various miRNA binding sites (MBS) with different affinities and efficacies have been described in literature. The most common ones are presented in *figure 1.4* and are termed canonical (Baek, Villen et al. 2008). microRNAs that share the same seed sequence are considered members of a miRNA family, typically named after the first miRNA discovered which displays

that specific seed. Due to the fact that the seed sequence is the major determinant of the interactions with a target, miRNAs within a family share a subset of target genes. The context of the miRNA binding site in the 3'UTR has been described to influence the repressing capabilities of a miRNA-target interaction, further increasing the complexity of gene expression regulation possibilities (Bartel 2009).

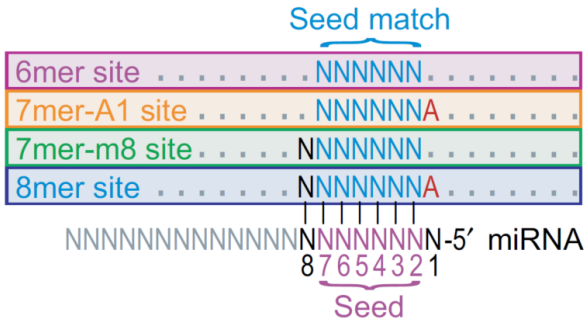


Figure 1-4 Schematic representation of the most commonly predicted miRNA binding sites. miRNA is depicted on the bottom in 3' to 5' direction with its seed sequence in purple. Predicted miRNA binding sites (seed matches) are depicted 5' to 3' sequences, color-coded according to seed match type (Baek, Villen et al. 2008).

miRISC-mediated silencing of the target has been widely investigated in the past 20 years with both *in vitro* and *in vivo* approaches to identify all the proteins that contribute to this mechanism. However, the precise spatial and temporal resolution of the events that lead to silencing is still under examination. Upon miRISC binding, silencing occurs as a combination of translational repression and mRNA processing and degradation. The predominant effect, as observed in multiple steady state *in vitro* experiments, is the degradation of the message, accounting for 2/3 and up to more than 90% of the final outcome on protein production. To achieve mRNA destabilization and subsequent exonucleolytic decay, AGOs proteins recruit GW182 proteins which further interact with downstream effectors such as the poly(A)-binding protein complex (PABPC) and the cytoplasmic de-adenylase complexes PAN2-3 and CCR4-NOT. Upon de-adenylation, target mRNAs are decapped and readily degraded by the major cytoplasmic 5'-to-3' exoribonuclease XRN1 (*fig. 1.5*).

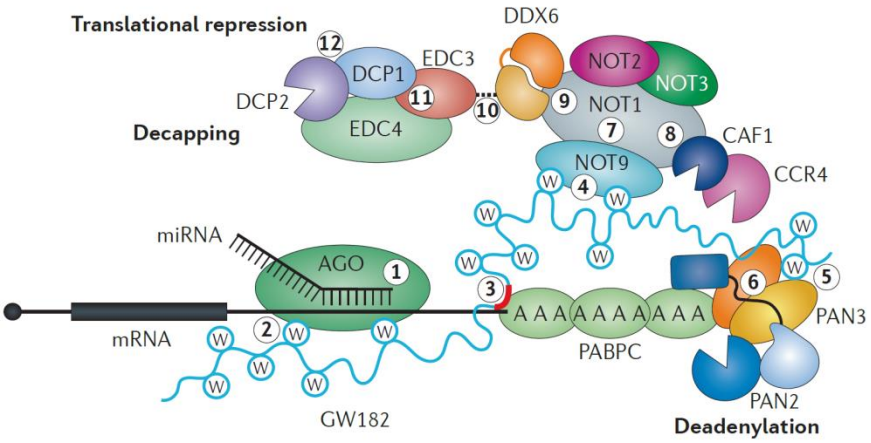


Figure 1-5 Proposed mechanisms of miRNA-mediated target gene silencing. The miRISC binds by partial complementarity to BS in the 3'UTR of target genes (1) and recruits GW182 proteins (2), which in turn interact with proteins inducing deadenylation of target mRNA (3-9). DDX6 might interact (10) with decapping enzymes as well (11-12) which remove the 5' cap of target mRNA, allowing the initiation of 5'-to-3' mRNA exonucelolytic process. All the numbered interactions are supported by experimental structures identified in X-ray crystallography or nuclear magnetic resonance (NMR) experiments. The scheme does not represent the translational inhibition effect caused by interactions with 5'cap-associated translation initiation factors (Jonas and Izaurralde 2015)

A proposed method by which miRISC induces translational repression is through interfering with the assembly and activity of the cap-associated eukaryotic initiation factor 4F (eIF4F), thus affecting translation at initiation. This process is currently considered as a minor effect that happens at early stage of the silencing, accounting only for 6-26% of the effect on endogenous targets in mammalian cells. However, it is important to highlight that the temporal resolution could be a key factor in understanding the flow of events which lead to an effective silencing of targets (Jonas and Izaurralde 2015).

Cell culture-based analyses of miRNA-mediated effects have repeatedly shown very mild effects, both at the transcriptomic and proteomic level. These effects have been quantified in gain- and loss-of-function studies, where miRNAs were overexpressed (OE) or knocked-down (KD). Indeed, in a seminal study from the Bartel group, miR-124 OE followed by analysis of protein and mRNA expression revealed an average repression of -0.18 and -0.3 log2FC for transcript that contained at least one MBS in their 3'UTR. As well, genetic inactivation of specific miRNAs (e.g. miR-223 in mouse neutrophils) caused de-repression of

targets at a similar range, with mRNAs and proteins upregulated +0.15 and +0.2 log2FC, respectively (Baek, Villen et al. 2008). These variations have been considered to fall within the range of stochastic fluctuations which can be found within a population of genetically identical cells, putting into question how a single microRNA can exert a significant biological function when it affects so mildly the expression of its targets (Ebert and Sharp 2012).

1.3.3. miRNA involved in control of networks

A possible answer to this conundrum lays in the presence of multiple targets of the same miRNA, for the concomitant repression (or de-repression) of many effectors which are functionally integrated in a pathway can have a profound phenotypical change. Additionally, a recent report analysed the changes of mRNA levels, as well as post-transcriptional and epigenetic modifications upon loss of miRNA activity by Dicer KO. The results suggested that the majority of the effects were observable already at the transcription level, suggesting that the main effect of miRNAs is on transcription factors which, in turn, affect downstream expression of target genes (Gosline, Gurtan et al. 2016).

Various networks of interactions between signalling pathway components and microRNAs, as well as feedback and feedforward loops with transcription factors have now been characterized, putting miRNAs in the spotlight as strong players in the gene expression arena (Bracken, Scott et al. 2016). Due to their dose-sensitivity nature, biological pathways require fine-tuned control of the signalling cascade. In this contexts, miRNAs may become pivotal regulators, thanks to their ability to direct the expression of multiple targets concomitantly (Inui, Martello et al. 2010).

The role of miRNAs as combinatorial regulators of specific phenotypes has been described multiple times *in vivo* for the miR-17-92 cluster, a group of six miRNAs concordantly regulated in development and cancer. Dissection of the contribution of each miRNA to specific phenotypes revealed that within the cluster both specialization of functions and cooperation are present (Han, Vidigal et al. 2015).

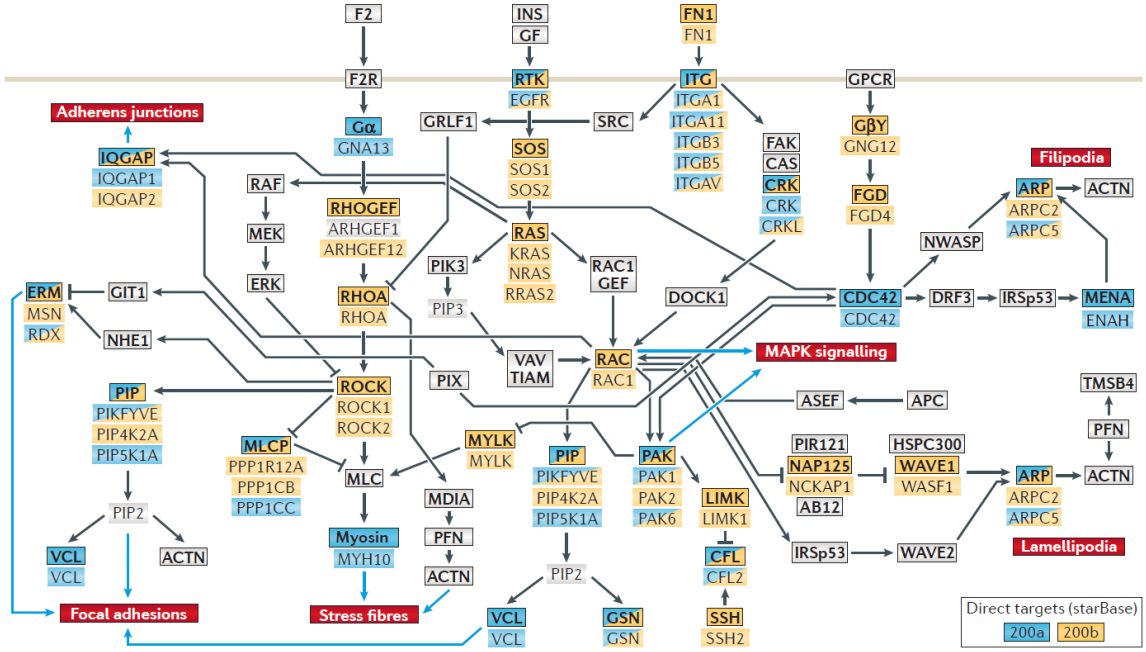


Figure 1-6 miR-200 family members regulate actin cytoskeleton. Proteins belonging to the actin cytoskeleton KEGG map (hsa04810) are coloured based on the verified interactions found in starBase database (Yang et al, 2011, NAR). As indicated by the blue lines, the phenotypes regulated by this multiple targeting strategy include various processes downstream the Integrin and RTK signalling cascades (Bracken, Scott et al. 2016).

Many interesting examples have shown cooperative features of a miRNA on multiple targets, both in physiological and cancer related settings. In adult mouse neurons, miR-128 controls expression of various ion channels and ERK2 signalling. By controlling this wide network, miR-128 regulates neuronal excitability and its loss causes intensified motor activity and fatal epilepsy (Tan, Plotkin et al. 2013). The miR-200 family members are found expressed preferentially in the epithelial compartment compared to the mesenchymal one. In cancer, their loss has been widely associated with epithelial to mesenchymal transition (EMT), mostly by de-repressing the transcription factors ZEB1 and ZEB2. However, a broader investigation of direct targets of the five miRNAs belonging to this family revealed how they cumulatively target proteins involved in actin cytoskeleton remodelling, affecting invasion and invadopodia formation in MDA-MB-231 breast cancer cells (*fig 1.6*) (Bracken, Li et al. 2014).

1.3.4. miRNAs in cancer

microRNA expression is globally dysregulated in cancer tissues compared to their normal counterpart, an aberration further enhanced when tumours are poorly differentiated (Lu, Getz et al. 2005). While few of the deregulated miRNAs are overexpressed and might act as *bona fide* oncogenes, the majority is broadly downregulated. Indeed, miRNA genes have been mapped to genomic areas that are frequently found altered in the cancer, for instance fragile sites, minimal regions for loss of heterozygosity, minimal amplicons, and common chromosomal breakpoint loci (Calin, Sevignani et al. 2004). These genetic factors might contribute to the general miRNA expression repression characteristic of tumour tissues. Additional mechanisms that have been characterized as potential causative agents are broad epigenetic alterations such as aberrant promoter DNA methylation (Lujambio, Calin et al. 2008) and loss of specific histone acetylation markers (Saito, Liang et al. 2006). In recent years, special attention has been paid to understand the contribution of enzymes and critical cofactors involved in miRNA biogenesis in this context. Patient data analysis showed both upregulation and downregulation of the enzymes, as well as somatic and germline mutations, being correlated with clinical parameters only occasionally. A high variability in prognosis and other clinical features was found depending on the type of cancer, thus suggesting that the expression of enzymes involved in miRNA biogenesis is not enough to explain the global expression repression. In depth analyses of other key players in the biogenesis processing, showed the involvement of DEAD box helicases DDX5 (p68) and DDX17 (p72) in nuclear processing deregulation, upon interacting with BRCA1 as well as p53 tumour suppressor. SMADs transducers of TGFβ and BMP signalling have also been shown to interact with DDX5 to promote miRNA processing. miRNA processing modulation via interaction with DDX17 was instead discovered for the Hippo pathway, correlating cell density to mature miRNA levels (Lin and Gregory 2015). To conclude, core components of the miRNA biogenesis machinery have also been investigated for post-translational modifications such as phosphorylation and acetylation (*fig 1.1 and 1.2*) (Ha and Kim 2014).

Of note, the widespread repression of miRNA expression in cancer is similarly found in cell types which are maintaining a pluripotent or dedifferentiated state. In

developmental contexts, extensive repression occurs concomitantly with high expression levels of specific miRNAs, which then represent the majority of the population present in the cell milieu. For instance, in mouse ESCs, more than 75% of the miRNA population expressed was found to be encoding miRNA belonging to only six genomic loci, with a clear preference for the miR-290-295 cluster (~60% of the population expressed) (Marson, Levine et al. 2008). It has therefore been speculated that miRNA increased expression during development is necessary to canalise phenotypic decision, promoting lineage-specific functions or suppressing alternative fates. On the contrary, in the tumour context, the widespread reduction in miRNA expression causes derepression of multiple targets and therefore an increase in transcriptional noise. In turn, this allows cancer cells to be highly adaptable to a constantly changing tumour environment, therefore conferring them an “evolutionary” advantage (Ebert and Sharp 2012).

1.3.5. Functional analysis of miRNA-mediated network regulation

Common methods to investigate the functional role of a miRNA within a specific context are gain- or loss-of-function experiments, followed by quantitative approaches such reporter gene or phenotypic assays, as well as gene expression analysis. Previous members of the Molecular Genome Analysis division have exploited these kinds of methodologies to investigate miRNAs regulating NF-kB and EGFR signalling pathways in the context of breast cancer (Keklikoglou, Koerner et al. 2012, Uhlmann, Mannsperger et al. 2012). In order to identify which miRNAs control NF-kB signalling, a library of constructs was co-transfected with a plasmid containing 3 consensuses NF-κB-binding sites upstream the firefly luciferase gene, functioning as reporter for the pathway activity. MicroRNA-520/373 family was identified as a potent repressor of NF-kB signalling and of the phenotypes dependent on its activation. Additionally, miR-520/373 was characterized to repress TGFβ signalling, providing a bridge between the two pathways (Keklikoglou, Koerner et al. 2012). A different approach has been exploited to identify miRNAs affecting cell cycle through EGFR signalling regulation. Upon miRNA OE, protein expression was quantified with Reverse Phase Protein Arrays (RPPA). The data was used to build a regulatory network which provided the basis to infer which miRNAs are

modulating the cell cycle. Three miRNAs (-124, -147, and -193a-3p) were validated for their role as strong repressors of cell cycle (Uhlmann, Mannsperger et al. 2012). The advantages of techniques such as RPPA are the quantification at the protein level and the high sample content. While the latter is necessary to understand the function of many miRNAs within the same assay, the former allows the identification interactions more likely to be functional. Despite having been described that the majority of miRNA-mediated effects can already be evaluated at the mRNA level (Guo, Ingolia et al. 2010), the possibility to investigate the protein output should still be preferred. Indeed, various reports still show that the correlation between mRNA and protein expression is poor, with the ability of the former to explain roughly 40% of the variation of the latter (Vogel and Marcotte 2012).

A very informative version of such assays combines both the gain- and the loss-of-function to identify molecules that may act as phenotypic switches. This approach was recently exploited to characterize miRNAs regulating the stem cell phenotype. By quantifying the cells positive for aldehyde dehydrogenase (ALDH) enzymatic activity upon miRNA OE and KD, researchers identified miR-600 as a bimodal switch that regulates the stem cell fate (El Helou, Pinna et al. 2017). Another method to identify miRNA that holds a functional value which was broadly exploited in the past years encompasses the analysis of miRNA differential expression between experimental conditions. *In silico* target predictions of the differentially expressed miRNAs is then performed, and predicted targets are listed. Then an enrichment for pathways based on gene ontology repositories is performed and the function of the miRNA is inferred (Gusev, Schmittgen et al. 2007). It is important to note that this method and its derivatives have been criticized due to the intrinsic bias in the sets of genes which are predicted to be targets of miRNAs (Bleazard, Lamb et al. 2015). However, gene set enrichment analysis becomes extremely helpful when performed on genes that are differentially expressed upon miRNA OE or KD to identify pathways concordantly regulated or *bona fide* direct targets that present predicted MBS in their 3’UTR (van Dongen, Abreu-Goodger et al. 2008).

Phenotypic screenings and analyses of reporter gene activity upon miRNA OE or KD have the advantage of defining directly which ones have a function. However, to fully understand the mechanisms underlying phenotypes identified, they require the dissection of the intricate network of miRNA-target regulations. On the other hand, while giving more insight into the molecular mechanism, gene expression or protein quantifications analyses necessitate computational approaches to infer miRNA function, therefore relying on pathway enrichment tools or other modelling methods.

In silico prediction of target genes and online repositories of validated miRNA-target interaction come to aid to sharpen the focus on interesting candidates that might arise from high-throughput functional or genomic screens. However, it is important to highlight that predictions are unable to identify all miRNA:mRNA interactions. This is mostly due to the fact that the knowledge that we hold on target identification mechanisms is still incomplete. Therefore, false-positives and –negatives are to be expected among target prediction results, with one or the others predominantly present depending on the algorithm exploited (Riffo-Campos, Riquelme et al. 2016). The most recent release (v7) of TargetScan, one of the predominant target prediction algorithm, is powered by the context++ model which is based on 14 variables pertinent to the miRNA:mRNA interaction under investigation (Agarwal, Bell et al. 2015). Its performance, evaluated on gene expression data (i.e. mRNA level), recognizes it as the most precise tool currently available. However, its high specificity might hinder non-conserved, rare, or non-canonical interactions that are anyway biologically relevant (Riffo-Campos, Riquelme et al. 2016).
This reinforces the idea that while prediction algorithms come in great aid in the possible identification of direct targets, and therefore of the precise molecular mechanism that underlies miRNA function, their results require experimental validation.

1.4. Aims of the PhD project

Regardless of the advancements in recent years in breast cancer biology, the current predominant treatment option for TNBC patients remains chemotherapy, with dismal chances of survival upon recurrence. Additional studies of the biological mechanisms underlying the metastatic processes are therefore needed to improve the clinical management of this disease.

microRNAs, characterised in recent years as modulators of gene expression, have been implicated in the metastasis formation process in various types of cancer. However, a systematic approach to assess their ability to regulate metastatic pathways had not been undertaken before.

The **main aim** of this project was **to identify miRNAs able to repress pathways related to metastasis formation** in a cell model of Triple Negative Breast Cancer.

To achieve this aim, the project focused on:

a) **Systematically characterise the effect of the miRNome** on proteins belonging to the pathways of interest by means of a targeted high throughput proteomic approach (RPPA).
b) Developing **a novel network analysis method** to infer the effect of miRNAs on the pathways from their ability to regulate the single proteins.
c) Focusing on a single pathway, **validate the miRNAs for their capability to repress** its activity and phenotypes dependent on it.
d) Analyse interactions between miRNAs and target mRNAs to **validate the direct integration of miRNAs in the pathway.**

2. Materials and methods

2.1. Materials

2.1.1. Instruments

2470 microarray printing platform	Aushon (Billerica, MA, USA)
ABI Prism 7900HT Sequence Detection System	Applied Biosystems (Foster City, USA)
Bacterial incubator (37°C)	Memmert (Schwabach, Germany)
Cell culture hood HERA Safe	Thermo Fisher Scientific (Waltham, USA)
Cell culture incubator CO_2 controlled	Binder (Tuttlingen, Germany)
Cell Observer	Zeiss (Jena, Germany)
Centrifuges	Eppendorf AG (Hamburg, Germany) Heraeus Instruments (Hanau, Germany)
DNA gel apparatus	Renner GmbH (Germany)
Flow Cytometer FACS Canto II	Becton Dickinson (New Jersey, USA)
Gel documentation system	Herolab GmbH (Wiesloch, Germany)
Infinite M200 microplate reader	Tecan Group (Männedorf, Schweiz)
LSM 780 confocal laser scanning microscope	Zeiss (Jena, Germany)
Nanodrop ND-1000 spectrophotometer	Thermo Fisher Scientific (Waltham, USA)
Odyssey Infrared Imaging System	Li-Cor Biosciences GmbH (Bad Homburg, Germany)
Olympus microscope	Olympus (Tokyo, Japan)
Protein Gel Apparatus MiniProtean II	Bio-Rad (Hercules, USA)
Shaking bacterial incubator (37°C)	INFORS HT (Eisenbach, Germany)
Spectrophotometer Nano Drop nd-1000	Thermo Fisher Scientific (Waltham, USA)
Thermocycler	Applied Biosystems (Foster City, USA)
Thermomixer	Eppendorf AG (Hamburg, Germany)
Trans-Blot SD Semi-Dry Electrophoretic Transfer	Bio-Rad (Hercules, USA)
UV imager	Vilber (Eberhardzell, Germany)
Vortex mixer	neoLab (Heidelberg, Germany)

2.1.2. Disposables

1.5 mL micro centrifuge tube	Eppendorf AG (Hamburg, Germany)
2 mL micro centrifuge tube	Eppendorf AG (Hamburg, Germany)
4–20% Mini-PROTEAN® TGX™	Bio-Rad (Hercules, USA)

12 mm round coverslips	VWR International
10cm Ø Petri dish	Techno Plastic Products (TPP) AG (Trasadingen, Switzerland)
15mL conical tube	Becton Dickinson (New Jersey, USA)
50mL conical tube	Becton Dickinson (New Jersey, USA)
6-well plate, flat bottom, transparent	Nunc, Thermo Fisher Scientific (Waltham, USA)
6-well plate, Ultra Low Attachment	Costar (Corning, USA)
12-well plate, flat bottom, transparent	Nunc, Thermo Fisher Scientific (Waltham, USA)
96-well plate, flat bottom, transparent	Becton Dickinson (New Jersey, USA)
96-well plate, flat bottom, white	Greiner Bio-One International GmbH (Kremsmünster, Austria)
96-well plate, flat bottom, µCLEAR®, black	Greiner Bio-One International GmbH (Kremsmünster, Austria)
Adhesive Optically Clear Plate Seal	Thermo Fisher Scientific (Waltham, USA)
Cell Culture Flasks, T-25, T-75, T-175	Greiner Bio-One International GmbH (Kremsmünster, Austria)
Cell Scraper	Corning (Corning, USA)
Cryo vials 1.8mL	Nunc, Thermo Fisher Scientific (Waltham, USA)
Filter tips, 10µL, 20µL, 100µL, 200µL, 1000µL	Neptune Scientific (San Diego, USA)
Glass slides	Thermo Fisher Scientific (Waltham, USA)
Incubation chambers	Metecon (Mannheim, Germany)
Inoculation loops 10 µL	Copan (California, USA)
ONCYTE® AVID nitrocellulose coated sldies	Grace Biolabs (Bend, OR, USA)
PCR strips	Steinbrenner Laborsysteme GmbH (Wiesenbach, Germany)
PVDF membrane Immobilon-P	Merck Millipore (Darmstadt, Germany)
Serological pipettes 2.5mL, 5mL, 10mL, 25mL, 50ml	Becton Dickinson (New Jersey, USA)
Whatman 3 MM filter paper	GE Healthcare (Little Chalfont, United Kingdom)

2.1.3. Chemicals and reagents

4',6-diamidino-2-phenylindole dihydrochloride (DAPI)	Sigma-Aldrich (Saint-Louis, USA)

4× Roti Load	Carl Roth (Karlsruhe, Germany)
Acetic acid	Sigma-Aldrich (Saint-Louis, USA)
Acrylamide/bisacrylamide 37.5:1	Carl Roth (Karlsruhe, Germany)
Agarose	Carl Roth (Karlsruhe, Germany)
Ammoniumperoxodisulfate (APS)	Sigma-Aldrich (Saint-Louis, USA)
Ampicillin	Sigma-Aldrich (Saint-Louis, USA)
Bacto-Agar	Roche Diagnostics (Mannheim, Germany)
Biozym Sieve GP Agarose	Biozym (Hessisch Oldendorf, Germany)
BSA	Sigma-Aldrich (Saint-Louis, USA)
Calcein, AM	Invitrogen (Carlsbad, USA)
Coelenterazine	Sigma-Aldrich (Saint-Louis, USA)
CoenzymeA	AppliChem (Darmstadt, Germany)
Complete Mini Protease Inhibitor	Roche (Penzberg, Germany)
Dimethylsulfoxide (DMSO)	Sigma-Aldrich (Saint-Louis, USA)
DL-Dithiothreitol (DTT)	Sigma-Aldrich (Saint-Louis, USA)
D-luciferin	AppliChem (Darmstadt, Germany)
EDTA	Sigma-Aldrich (Saint-Louis, USA)
Ethanol	Sigma-Aldrich (Saint-Louis, USA)
Ethidium bromide	Sigma-Aldrich (Saint-Louis, USA)
Gelatin From Pig Skin, Oregon Green® 488 Conjugate	Invitrogen (Carlsbad, USA)
Gelatin from porcine skin (clear)	Sigma-Aldrich (Saint-Louis, USA)
Glutaraldehyde	Sigma-Aldrich (Saint-Louis, USA)
Hoechst 33342, Trihydrochloride, Trihydrate	Thermo Fisher Scientific (Waltham, USA)
Isopropanol	Greiner Bio-One International GmbH (Kremsmünster, Austria)
Lithium Chloride (LiCl)	Sigma-Aldrich (Saint-Louis, USA)
Methanol	Greiner Bio-One International GmbH (Kremsmünster, Austria)
MgSO4	Sigma-Aldrich (Saint-Louis, USA)
Midori Green Advance	Nippon Genetics (Düren, Germany)
NaOH	Sigma-Aldrich (Saint-Louis, USA)
non-DEPC treated nuclease-free water	Ambion, Thermo Fisher Scientific (Waltham, USA)
PFA 16% W/V AQ	VWR International (Darmstadt, Germany)

PhosSTOP	Roche (Penzberg, Germany)
poly-L-lysine	Sigma-Aldrich (Saint-Louis, USA)
ProLong® Gold antifade reagent	Invitrogen (Carlsbad, USA)
Rhodamin Phalloidin	Thermo Fisher Scientific (Waltham, USA)
Rockland Blocking Buffer for Fluorescent Western Blotting	Rockland (Limerick, PA, USA)
SDS	Carl Roth (Karlsruhe, Germany)
SOC medium	Invitrogen (Carlsbad, USA)
Sodium borohydride	Sigma-Aldrich (Saint-Louis, USA)
Sodium Chloride (NaCl)	VWR International (Darmstadt, Germany)
Sodium Citrate Tribasic Dihydrate	AppliChem (Darmstadt, Germany)
Sodium Fluoride (NaF) 0.5M	Sigma-Aldrich (Saint-Louis, USA)
Sodium Hydrogen Carbonate for cell culture ($NaHCO_3$)	AppliChem (Darmstadt, Germany)
Sodium orthovanadate (Na_3VO_4)	Sigma-Aldrich (Saint-Louis, USA)
TEMED	Carl Roth (Karlsruhe, Germany)
Tricine	Carl Roth (Karlsruhe, Germany)
Tris HCl	Sigma-Aldrich (Saint-Louis, USA)
Tris-base	Sigma-Aldrich (Saint-Louis, USA)
Triton X-100	Sigma-Aldrich (Saint-Louis, USA)
Tryptone	Carl Roth (Karlsruhe, Germany)
Tween® 20	Sigma-Aldrich (Saint-Louis, USA)
Yeast extract	Gerbu (Wieblingen, Germany)

2.1.4. Buffers

FCF staining solution	0.005% FCF 30% EtOH 10% Glacial Acetic Acid in ddH2O

FCF destaining solution	30% EtOH 10% Glacial Acetic Acid in ddH2O

Blocking buffer	1:1 Rockland Blocking buffer: TBS 5 mM Sodium Fluoride 1 mM Sodium Vanadate

TBS-T (washing buffer)	50mM Tris 150mM NaCl 0.1% Tween20 (v/v) in ddH2O

4x running buffer	1.5 M Tris-HCl, pH to 8.8

4x stacking buffer	0.5 M Tris-HCl, pH to 6.8

Anode I	300 mM Tris 20% EtOH (v/v) in ddH2O

Anode II	25 mM Tris 20% EtOH (v/v) in ddH2O

Cathode	40 mM 6-aminocaproic acid 20% EtOH (v/v) in ddH2O

SDS page running buffer	25 mM Tris 14.41 g/l glycine 1% SDS (w/v) in ddH2O

1x Tris-acetate-EDTA (TAE)	40 mM Tris 20 mM acetic acid 1 mM EDTA in ddH2O

FACS assay buffer	2% FBS in PBS

2.1.5. Molecular biology reagents

Deoxynucleoside Triphosphate (dNTPS)	PeqLab (Darmstadt, Germany)
Lipofectamine2000®	Invitrogen (Carlsbad, USA)
M-PER Mammalian Protein Extraction Reagent	Thermo Fisher Scientific (Waltham, USA)
SsoAdvanced™ Universal SYBR® Green Supermix (2x)	Bio-Rad (Hercules, USA)
TaqMan® Fast Universal PCR Master Mix (2x)	Applied Biosystems (Foster City, USA)
TaqMan® MicroRNA Assays	
TaqMan™ Universal PCR Master Mix, no AmpErase™ UNG	
Universal Probe Library (UPL)	Roche Diagnostics (Mannheim, Germany)

2.1.6. Kits

Unless otherwise stated, all kits were used according to manufacturer's instruction.

BCA Protein Assay Kit Pierce™	Thermo Fisher Scientific (Waltham, USA)
miRNeasy Mini kit	Qiagen (Hilden, Germany)
QiaPrep Spin Miniprep Kit	
NucleoBond® Xtra Midi	Macherey-Nagel GmbH & Co. (Düren, Germany)
QuikChange Lightning Site-Directed Mutagenesis Kit	Agilent Technologies (Waldbronn, Germany)
QuikChangeLightning MultiSite-Directed Mutagenesis kit	
RevertAid™ H Minus First Strand cDNA synthesis kit	Fermentas, Thermo Fisher Scientific (Waltham, USA)
TA-Cloning Kit	Invitrogen (Carlsbad, USA)
TaqMan® MicroRNA Reverse Transcription Kit	Applied Biosystems (Foster City, USA)
Wizard® SV Gel and PCR Clean-Up System	Promega (Mannheim, Germany)

2.1.7. Enzymes

Phusion® Hot Start High-Fidelity DNA Polymerase	Finnzymes, Thermo Fisher Scientific (Waltham, USA)
T4 DNA ligase	NEB (Frankfurt, Germany)

AmpliTaq DNA Polymerase	Applied Biosystems (Foster City, USA)
DreamTaq Polymerase	Thermo Fisher Scientific (Waltham, USA)
XhoI / NotI-HF / PmeI	NEB (Frankfurt, Germany)

2.1.8. Size markers and loading dyes

Orange DNA Loading Dye (6X)	Thermo Fisher Scientific (Waltham, USA)
O'Range Ruler 200 bp	
O'Range Ruler 100+500 bp	
MassRuler Mix	
4× Roti Load	Carl Roth (Karlsruhe, Germany)
Precision Plus Protein Standards Dual Color	Bio-Rad (Hercules, USA)

2.1.9. Cell culture: cell lines, media, and other reagents

MDA-MB-231 (ATCC® HTB-26™)	American Type Culture Collection (Manassas, VA, USA)
MDA-MB-468 (ATCC® HTB-132™)	

Standard Medium	Leibovitz-L15 10 % Fetal Bovine Serum	Gibco BRL (New York, USA)
	3 g/lt of NaHCO3	Applichem (Darmstadt, Germany)

Mammosphere Growth Medium	DMEM-F12 1x B27 supplement	Gibco BRL (New York, USA)
	20 ng/ml recombinant human basic FGF	R&D systems (Minneapolis, MN, USA)
	20 ng/ml recombinant EGF	Corning (Corning, USA)
	4 ug/ml Heparin sodium salt	Sigma-Aldrich (Saint-Louis, USA)

Mammosphere formation medium was prepared according to the work of Dontu and colleagues (Dontu, Abdallah et al. 2003).

PBS	Gibco BRL (New York, USA)
OptiMEM	
0.25% Trypsin EDTA Solution	
Cell Dissociation Buffer, enzyme-free	

WNT-7a recombinant human	Peprotech (Rocky Hill, NJ, USA)
WNT-1 recombinant human	
WNT-3a recombinant mous	
iCRT 3	CalbioChem Merck (Darmstadt, Germany)
iCRT 14	SCB

2.1.10. Bacterial culture: strains and broth

Subcloning Efficiency™ DH5α™ Competent Cells	Thermo Fisher Scientific (Waltham, USA)
XL10-Gold® Ultracompetent Cells	Agilent Technologies (Waldbronn, Germany)

Lysogeny Broth (LB)	10g Bacto-Trypton 10g NaCl 5g yeast extract to 1 lt with ddH_2O, then autoclave 100 µg/ml Ampicillin

For agar plates preparation, 6 gr of Bacto-Agar are dissolved in 400 ml of LB broth, and autoclaved. Prior to pouring, Amp is addedd to a final concentration of 100 µg/ml

2.1.11. Antibodies

RPPA and Western Blot

Primary antibodies used for target protein detection in RPPA and WB. For RPPA, except when noted, antibodies were used in dilution 1:300 in Blocking Buffer Rockland:TBS. For WB, the last column indicates the dilution used.

Companies from which the Ab were purchased are listed abbreviated as subsequent.

Abcam	Abcam (Cambridge, UK)
BD	BD Biosciences (Heidelberg, Germany)
CST	Cell Signaling Technology (Danvers, MA, USA)
Epit	Epitomics (Abcam - Cambridge, UK)
HPA	Human Protein Atlas (Sigma-Aldrich, Munich)
Merck	Millipore Merck (Darmstadt, Germany)
RnD	RnD systems Inc. (Minneapolis, MN, USA)
SCB	Santa Cruz Biotechnology (Santa Cruz, CA, USA)

Sigma	Sigma-Aldrich (Munich, Germany)

Protein Abbreviation	Gene ID	AB ID	company (abbr.)	WB dilution
AKT/PRKBA	AKT1	BD 610860	BD	
PRKBB	AKT2	CST 3063	CST	
APC	APC	CST 2504	CST	1:500
B-raf	BRAF	CST 9434	CST	
CCND1	CCND1	CST 2926	CST	
CCND3	CCND3	CST 2936	CST	
E-Cadherin	CDH1	CST 4065	CST	1:1000
CDK2	CDK2	C5223	Sigma	
CDK4	CDK4	2341-1 (ab68266)	Epit	
p27/Kip1	CDKN1B	BD 610241	BD	
CrkII	CRK	CST 3492	CST	
CKI alpha	CSNK1A1	CST 2655	CST	1:500
CKI epsilon	CSNK1E	CST 12448	CST	1:1000
CKII alpha1	CSNK2A1	CST 2656	CST	1:1000
CKII β	CSNK2B	sc-12739	SCB	1:200
β-CATENIN	CTNNB1	CST 9562	CST	1:1000
ICAT	CTNNBIP1	ab129011	Abcam	1:1000
DIAPH1	DIAPH1	CST 5486	CST	
DUSP6	DUSP6	2138-1 (ab76310)	Epit	
Dishevelld 3	DVL3	CST 3218	CST	1:1000
EGFR	EGFR	CST 2646	CST	
c-Fos	FOS	sc-7202	SCB	
GRB2	GRB2	CST 3972	CST	
GSK3B	GSK3B	CST 9315	CST	1:1000
HRAS	HRAS	05-516	Merck	
ILK	ILK	CST 3862	CST	
JUN	JUN	CST 9165	CST	
LRP5	LRP5	CST 5731	CST	1:1000
LRP6	LRP6	CST 2560	CST	1:200

MEK1	MAP2K1	BD 610122	BD	
ERK2 (1:200 dil.)	MAPK1	sc-81458	SCB	
p38	MAPK14	CST 9212	CST	
ERK1	MAPK3	AF1575	RnD	
JNK1	MAPK8	ab7949	Abcam	
HGFR	MET	CST 3148	CST	
FRAP1	MTOR	CST 2972	CST	
PAK1	PAK1	CST 2602	CST	
PAK2	PAK2	2247-1 (ab76293)	Epit	
PDK1	PDPK1	sc-7686	SCB	
PI3K alpha	PIK3CA	CST 4249	CST	
PI3Kβ	PIK3CB	ab32569	Abcam	
PLCG1	PLCG1	ab41433	Abcam	
MLCP	PPP1CA	sc-7482	SCB	
PKC alpha	PRKCA	CST 2056	CST	
PKC delta	PRKCD	CST 9616	CST	
PS-1	PSEN1	CST 3622	CST	1:200
PTEN	PTEN	CST 9552	CST	
FAK1	PTK2	CST 3285	CST	
SHP-2	PTPN11	1609-1 (ab32159)	Epit	
Paxillin	PXN	CST 2542	CST	
RAC1	RAC1	CST 2465	CST	
Rap1 alpha	RAP1A	ab124963	Abcam	
pRB	RB1	CST 9309	CST	
RhoA	RHOA	CST 2117	CST	
p160ROCK	ROCK1	CST 4035	CST	
p164 ROCK-2	ROCK2	HPA007459	HPA	
Axam2	SENP2	HPA029247	HPA	1:250
SHC1	SHC1	CST 2432	CST	
SOS	SOS1	CST 12409	CST	
SRC	SRC	CST 2123	CST	
STAT3	STAT3	BD 610190	BD	
p53	TP53	sc-126	SCB	

For WB, an antibody against Actin (MP Biomed 69100) was used diluted 1:10'000 as normalizer.

For RPPA target detection, secondary antibodies were diluted 1:8'000 in TBS-T, except for the Donkey α-Goat which was diluted 1:20'000. For WB, secondary antibodies were diluted 1:10'000.

Host α-Antigen	type	label	
Goat α-Mouse	F(ab')2 of IgG	Alexa Fluor 680	Invitrogen (Carlsbad, USA)
Goat α-Rabbit	F(ab')2 of IgG	Alexa Fluor 680	
Donkey α-Goat IgG	IgG	Alexa Fluor 680	

FACS

blocking agent Human TruStain FcX™		BioLegend (San Diego, CA, USA)
anti-human CD24 Antibody	APC	
Mouse IgG2a, κ Isotype Ctrl Antibody	APC	
Anti-Human/Mouse CD44	PE	eBioscience (San Diego, CA, USA)
Rat IgG2b K Isotype Control	PE	

Invadopodia formation

Host α-Antigen	type		dilution	
Rabbit α-TKS5 (M-300)	poly IgG	sc-30122	1:300	SCB
Goat α-Rabbit	IgG	Alexa Fluor 633	1:400	Invitrogen (Carlsbad, USA)

2.1.12. Plasmids

pcDNA3.1 (+)	Invitrogen (Carlsbad, California)
psiCHECK2	Promega GmbH (Neckarau, Germany)

2.1.13. Oligonucleotides

miRNAs

Individual microRNA mimics as well as the miRIDIAN® miRNA mimic library (based on miRBase v.10) were purchased from Dharmacon (Lafayette, CO, USA). The negative controls miRIDIAN® miRNA mimic control #1 and #2 were used for the experiments.

siRNAs

siRNA against *CTNNB1* (ID M-003482-00) were picked from an siRNA library against whole human genome. ON-TARGETplus Non-targeting Pool (D-001810-10) was used as negative control in the experiments. Both products were purchased from Dharmacon (Lafayette, CO, USA)

Cloning primers

Primers	5' → 3' sequence	corresponding cloned area
APC_UTR_TS_fw_partI	TACTCGAGGAAGCGAGATTCCAAAACTG	APC_CA1
APC_UTR_TS_rev_partI	ATGCGGCCGCAATAGATAAGTGCCAACGCA	
APC_UTR_TS_fw_partII	TACTCGAGTGTAATAGCAATGCAAGCAG	APC_CA2
APC_UTR_TS_rev_partII	ATGCGGCCGCGGAATATTTGGCCTGCTATC	
APC_UTR_TS_fw_partIII	TACTCGAGTTAGTCAGGAGTTACGATGC	APC_CA3
APC_UTR_TS_rev_partIII	ATGCGGCCGCTTACAGGAGCCACAGTCTA	
CDH1_FWD_XhoI_CA1	TACTCGAGATGAAAAGTGGGCTTGGAG	CDH1
CDH1_REV_NotI_CA1	TAGCGGCCGCCACCCACACATGTATACACT	
CSNK1A1_FWD_XhoI_CA1	TACTCGAGCAAAAGGGCACAGGAGCCAC	CSNK1A1
CSNK1A1_REV_NotI_CA1	TAGCGGCCGCACTCAAGCCTGGGCGACAGT	
CTNNB1_TS_FWD_XhoI	TACTCGAGTCAGCTGGCCTGGTTTGATAC	CTNNB1
CTNNB1_TS_REV_NotI	TAGCGGCCGCTCAGCAACTCTACAGGCCAAT	
CTNNBIP1_FWD_XhoI_CA1	TACTCGAGTTCAGACTGGCCCTTAAGCCA	CTNNBIP1
CTNNBIP1_REV_NotI_CA1	TAGCGGCCGCCCCACAGCTGACTGTCAGCT	
DVL3_ENST00000313143_fwd	TACTCGAGCATTTTCTCACCTGGATAAAGC	DVL3
DVL3_ENST00000313143_rev2	ATGCGGCCGCTGTGATGAAGTAACCAGGAG	
GSK3B_ENST00000264235_fwd_partI	TACTCGAGCCAGCTGCACAGGAAAAACC	GSK3B_CA1
GSK3B_ENST00000264235_rev_partI	ATGCGGCCGCGGCAAACATAGTCCTTCAATTCTCC	
GSK3B_ENST00000264235_fwd_partII	TACTCGAGAAAACAGCAGATTCTGGGAG	GSK3B_CA2
GSK3B_ENST00000264235_rev_partII	ATGCGGCCGCAAAACTGTCCCTTTCAGAGTTG	
LRP6_ENST00000261349_fw_partI	TACTCGAGATTTGGTTGAGATCTGGAGG	LRP6_CA1
LRP6_ENST00000261349_rev_partI	ATGCGGCCGCTCACAAAAGAGGCTTCCATTA	

LRP6_ENST00000261349_fw_partII	TACTCGAGTCTTTCTGGGGTAGATGAGA	LRP6_CA2
LRP6_ENST00000261349_rev_partII	ATGCGGCCGCCCACACATTTATTAAGAGCTGAC	
LRP6_ENST00000261349_fw_partIII	TACTCGAGAGTATGGATATGAGTGTGTTGA	LRP6_CA3
LRP6_ENST00000261349_rev_partIII	ATGCGGCCGCAGGGAAGTGAATAAGTGGAC	
PSEN1_ENST00000324501_fw_partI	TACTCGAGTGGGCTTGTTTTCTACTTTG	PSEN_CA1
PSEN1_ENST00000324501_rev_partI	ATGCGGCCGCTTCAAATTCTGAGCTACTGTG	
PSEN1_ENST00000324501_fw_partIII	TACTCGAGTCTGACTCCAAATGACTCAA	PSEN_CA2
PSEN1_ENST00000324501_rev_partIII	ATGCGGCCGCCTGACAAATACAGAAAATCTGC	

qPCR primers

Target gene	5' → 3' sequence
UC.82- F	GTGTTTAAATCAATCAACTGCAAAC
UC.82- R	TTCAGTGTGGGATCACAGTAGG
CTNNB1 F	TGTTCTTGAGTGAAGGACTGA
CTNBB1 R	AAAATCCAGCGTGGACAATGG
GAPDH F	CGACAGTCAGCCGCATCTT
GAPDH R	CCCCATGGTGTCTGAGCG

Mutagenesis primers

Gene area	miRNA BS	5' → 3' sequence	FWD/ REV
LRP6_CA1	193b_NC	GCCAGATAGCCTTCAGTTAACTAACATTTCGCGGCCAACAAGTAAGA	F
		TCTTACTTGTTGGCCGCGAAATGTTAGTTAACTGAAGGCTATCTGGC	R
LRP6_CA2	494	CCAAGGGGGAAAATGAATAATGAGGCATTAGATT	F

	193b	CCTTATATGC GCATATAAGGAATCTAATGCCTCATTATTCATTTT CCCCCTTGG TTAAAGTACATTTTAATCTAATATAGGCTGCTAATT GGGGGAAGGGGCTTTGAGC GCTCAAAGCCCCTTCCCCCAATTAGCAGCCTATA TTAGATTAAAATGTACTTTAA	R F R
APC_CA2	193b position 1 position 2	 GATTTTTTTTAAAGCTAAAATGCTGCTAAATAAAA GTGCTATGACTTGAGC GCTCAAGTCATAGCACTTTTATTTAGCAGCATTTT AGCTTTAAAAAAAATC AGTGGGGGGAGACGGGCTGCTCTATACTACCCA TTG CAATGGGTAGTATAGAGCAGCCCGTCTCCCCCCA CT	 F R F R
PSEN1_CA1	494 193b position 1 position 2	GCGGTTAGAATCCCATGGATGAGGCTTCTTTGAC TATAACAAAATCTGGG CCCAGATTTTGTTATAGTCAAAGAAGCCTCATCCA TGGGATTCTAACCGC ATTGCCATTTCTTCCCAAGGCTGCTCTGAACCTG AGGTTGCT AGCAACCTCAGGTTCAGAGCAGCCTTGGGAAGA AATGGCAAT GCAGTCTTTTTCTACAGCTGCTAAGGCAGCTCTG TCGTGGTAG CTACCACGACAGAGCTGCCTTAGCAGCTGTAGAA AAAGACTGC	F R F R F R
DVL3	193b	ACAGGAGCCCACAGGCTGCTTGAGGTTGGGCAA GG CCTTGCCCAACCTCAAGCAGCCTGTGGGCTCCT GT	F R
CTNNB1	193b	GGTAAGAAGTTTTAAAAAGCTGCTTTGGGTAAAAT ACTTTTACTCTGCCTACAG CTGTAGGCAGAGTAAAAGTATTTTACCCAAAGCA	F R

		GCTTTTTAAAACTTCTTACC	
CTNNBIP1	193b	CTAGGTGCCCAGGGCTGCTGGGACCCTCTCAGA	F
		TCTGAGAGGGTCCCAGCAGCCCTGGGCACCTAG	R

miRNA TaqMan primers

mature miRNA	TaqMan assay ID	miRBase ID or target sequence
hsa-miR-103a-3p	000439	MIMAT0000101
hsa-miR-193b-3p	002367	MIMAT0002819
hsa-miR-409-3p	002332	MIMAT0001639
hsa-miR-494-3p	002365	MIMAT0002816
hsa-miR-889-3p	002202	MIMAT0004921
RNU44	001094	CCTGGATGATGATAGCAAATGCTGACTGAACAT GAAGGTCTTAATTAGCTCTAACTGACT
HY3	001214	CCAGTCACAGATTTCTTTGTTCCTTCTCCACTCC CACTGCATCACTTAACTAGCCTT

2.1.14. Databases and online resources

Online databases and resources are listed together with the version used and the respective release years, when available. KEGG Pathway is continuously updated and therefore the year listed corresponds to the time when the database was queried to determine the targets of interest for the project. The year indicated is the release year for the version indicated. If no version is indicated, the year corresponds to the time when data used in the thesis was accessed.

Database	Vers	Year	Link
miRBase	21	2014	http://www.miRBase.org
ENSEMBL	89	2017	http://www.ensembl.org
TargetScan	7	2015	http://www.targetscan.org/vert_71/
MicroCosm Targets	5	2009	http://www.ebi.ac.uk/enright-srv/microcosm/htdocs/targets/v5/#
KEGG Pathway		2013	http://www.genome.jp/kegg/pathway.html

Benchling		2016	https://benchling.com/
miRTar Base	6	2015	http://mirtarbase.mbc.nctu.edu.tw/php/index.php
TCGA data portal		2016	https://tcga-data.nci.nih.gov/docs/publications/tcga/

2.1.15. Software

Commercial	
FACSDIVA™	Becton Dickinson (New Jersey, USA)
GenePix Pro v7	Molecular Devices (Sunnyvale, USA)
GraphPad Prism 5	GraphPad Software, Inc. (La Jolla, USA)
ImageStudio v3.1	Li-Cor (Lincoln, USA)
Odyssey 2.1	Li-Cor (Lincoln, USA)
SDS 2.2	Applied Biosystems (Foster City, USA)
ZEN blue	Zeiss (Jena, Germany)

Open Source	
Fiji (ImageJ)	https://imagej.net/Fiji
R	https://cran.r-project.org/
TinnR v 2.4.1.7	https://sourceforge.net/projects/tinn-r/

2.2. Methods

2.2.1. Cell culture

The human triple negative breast cancer cell lines MDA-MB-231 (231) and MDA-MB-468 (468) were cultured in Leibovitz-L15 medium (Gibco, Thermo Fisher Scientific) supplemented with 10% Fetal Bovine Serum (Gibco). Leibovitz-L15 medium is formulated for culturing in environments without CO_2 control. Therefore, in order to culture 231 and 468 cells in CO_2 controlled incubators, the medium is supplied with 3 g/lt of $NaHCO_3$. Cells were grown at 37ºC and 5% CO_2. Cells were split every 3-4 days with a 1:3/1:4 ratio. Cells were obtained from ATCC (Manassas, VA, USA) at passage 2 (P2) and cultured to maximum P20 (231) and P10 (468).

2.2.2. Generation and maintenance of Stable Isogenic Recombinant Cells (SiR cells)

(performed by Dr. Rainer Will, Rita Schatten, and Birgit Kaiser of the Stable isogenic recombinant core facility, DKFZ, Heidelberg)

A) Generation of the isogenic recombinant cell line

MDA-MB-231 were generated as described in Wittig-Blaich S et al. (1) Briefly : A Mammalian expression vector (pPAR3) containing a Flp recombinase target site N-terminally fused to EGFP under control of an elongation factor 1-alpha (E1Fα) promoter and a neomycin selection marker, was stably integrated in the genome of MDA-MB-231 cells. Neomycin resistant and EGFP positive clones were isolated and further validated for single-copy integration of the FRT site by Southern blotting. Functionality of the MDA-MB-231-pPAR3 acceptor cell line was further evaluated by a control Flp-mediated control recombination with either a Doxcylin inducible hcRED expression vector for visual testing or a red firefly expression vector for quantitative expression analysis. The carefully examined MDA-MB-231-pPAR3 acceptor cell line served a platform for the generation of isogenic variants.

B) Generation of MDA-MB-231-pPAR3 WNT- Luciferase reporter cell lines and culturing

For the generation of MDA-MB-231-pPAR3 WNT-Pathway cell lines a promoterless FRT expression vector containing a RNA polymerase II transcriptional pause signal from the human α2 globin gene followed by 6 repeats of the TCF/LEF transcriptional response element (AGATCAAAGGGGGTA) joined to a minimal TATA-box promoter and destabilized firefly luciferase reporter was “flipped in” the MDA-MB-231-pPAR3 acceptor cell line by cotransfection with a Flp recombinase expression vector (pOG44 / Invitrogen). After selection for hygromycin resistance (expression vector) and loss of

EGFP expression (positive integration events), single colonies were picked and analyzed with for WNT / luciferase responsiveness in the presence of recombinant WNT.
Positive colonies were cultured in Leibovitz-L15/10%FBS/3gr/lt $NaHCO_3$ supplemented with 400µg/ml hygromycin.

2.2.3. Transfections

A) Transfections and protein extraction for RPPA screen (respectively performed by Stefan Uhlmann and Heiko Mannsperger, former PhD students, MGA Division, DKFZ, Heidelberg)

The miRIDIAN miRNA mimic library from Dharmacon (corresponding to the miRBase database v10) was transfected in MDA-MB-231 cells with the help of the pipetting robot Biomek FX (Beckman Coulter, Krefeld, Germany). The miRNAs were transiently transfected at a final concentration of 25 nM. To transfect the whole library, the process was divided into 34 rounds. Negative controls #1 and #2 (Dharmacon) were transfected in every round. Cells were harvested by trypsinization 48 hrs. after transfection. Cell pellets were stored at -80ºC.
Protein were extracted from the cell pellet with 25 ul of Mammalian Protein Extraction Reagent (M-PER, Pierce, Rockford, IL, USA), supplemented with Complete Mini Protase Inhibitor (Roche, Penzberg, Germany). Lysis was carried for 20' at 4ºC on a shaker. Protein lysate was cleared by centrifugation at 13'000 rpm for 5' and the supernatant was stored at -80ºC. 5 ul of the lysate were used to measure protein concentration using the BCA colorimetric kit. (Pierce, Rockford, IL, USA). Sample concentrations was adjusted (when possible) to similar concentration using M-PER buffer. Tween 20 (Sigma Aldrich, Munich, Germany) was added to each sample at a final concentration of 0.05% to reduce surface tension and render the RPPA spotting process smoother. Samples were loaded into 384-well plate for RPPA spotting, and subsequently stored at -80ºC.

B) 6-well transfections

231 and 468 cells were seeded at the density of 2x and $3x10^5$ per well (respectively). The subsequent day, cells were transfected with 25 nM miRNA mimics and mimic controls (Dharmacon) or 20 nM siRNA and OnTarget plus siRNA control (Dharmacon).
Transfection mixes per well were prepared as shown in the subsequent table.
Mixes were incubated separately for 15'. Then they are combined 1:1 in a single tube and incubated for 20'. During incubation time, the medium in each well was removed and replaced carefully with 1ml of fresh medium. Then the transfection mix was distributed drop-wise on cells. With this procedure, there is no need to change the media after 6 hrs.

	ul x1 well
LF 2000	4
OptiMEM	246
Tot vol mix1	250

OptiMEM	248.5
miRNA/siRNA (stock 25 uM)	1.5
Tot vol mix2	250

RNA and proteins were harvested 48 hrs post-transfection in at least 2 biological replicates to perform validation of RPPA results via Western Blot (WB), as well as to analyze gene and miRNA expression upon modulation of the WNT/β-catenin pathway.

C) 96-well transfections for proliferation assays and WNT-reporter luciferase assays.
231 cells were plated at a density of $5x10^4$ in 96-well plates. The subsequent day cells were transfected with 25 nM microRNA mimics or mimic controls with the subsequent mixes.

	ul x1 well
OptiMEM	49.6
Lipof2000	0.4
Tot vol mix 1	50

Optimem	49.9
miRNA (stock 25 uM)	0.1
Tot vol mix 2	50

Mix 1 and mix 2 were incubated separately for 10'. Mix 1 was prepared in amount to be used with all the different miRNA or control conditions. Subsequently the two mixes were combined 1:1 in a new tube and incubated for 20'. Afterwards, the medium was removed from the wells, and 100 ul of the transfection mix were distributed with a multichannel pipet. 6-7 hrs after transfection, the mix was removed and replaced by 150 ul of fresh full medium.

D) 96-well transfections for 3'UTR reporter assays

231 cells were plated at a density $1x10^5$ per well in 96-well plates.

The subsequent day cells were transfected with 25 nM microRNA mimics or mimic controls and 50 ng of reporter plasmid per well.

The mixes were prepared in order to have a unique Lipofectamine+OptiMEM mix (Mix1) for all the wells, as well as the same miRNA+OptiMEM mix (Mix2) for assay uniformity.

As a consequence, a single mix1 is then combined with appropriate plasmid (sub-mix1), which is then distributed in tubes containing the mix2.

Prior all assays, the reporter plasmids were diluted to 1ug/ul to streamline the process. In few cases the reporter plasmids were not concentrated enough after midiprep and therefore not diluted.

	ul x1 well	Sub-mix1 addition
OptiMEM	50	
Lipof2000	0.4	
Tot vol mix 1	50	
Plasmid (1000ng/ul)		0.05

Optimem	49.9
miRNA (stock 25 uM)	0.1
Tot vol mix 2	50

Mix 1 is prepared and incubated for 20'. In the meantime, mix 2 is prepared and distribute into a deepwell plate.

Then, the appropriate volume of mix1 is mixed with plasmid and then distributed into the deepwell plate containing the mix2. For each vector, transfection was performed with 2 miRNA controls, miRNAs of interest, and a mock mix. After 20' of incubation the medium was removed from the wells, and 100 ul of the transfection mix were distributed with a multichannel pipet. 6-7 hrs after transfection, the mix was removed and replaced by 150 ul of fresh full medium.

2.2.4. Reverse Phase Protein Array (RPPA)

Cell lysates frozen in 384-well plates prepared by Stefan Uhlmann and Heiko Mannsperger were thawn on ice. The lysates were spotted on nitrocellulose-coated glass slides (Grace Biolabs, Bend, OR, USA) using the Aushon 2470 contact printer (Aushon, MA, USA). The samples were divided into 4 groups, termed arrays (AI, AII, AIII, and AIV), which were spotted independently. In each of the group, two dilution series of untreated 231 cells and two series of mock-transfected 231 cell lysates were spotted to allow proper quality control.

Printed slides were allowed to dry overnight at room temperature, and then stored at -20ºC in sealed plastic bags containing silica gel beads to reduce at minimum the humidity and therefore decrease the formation of ice crystals which could damage the spotted proteins.

Printed slides were secured in 2-pad incubation chambers and blocked for 2 hrs at room temperature with 2 ml/pad of Blocking Buffer. Subsequently, they were incubated O.N. at 4ºC on a rocking platform with highly specific primary antibodies diluted in Blocking buffer at 1:300 (unless otherwise stated). The subsequent day, slides were washed 4x with 2.5 ml/pad of Washing buffer, then incubated for 1h at RT with Alexa Fluor 680 F(ab')2 fragments of goat anti-mouse immunoglobulin G (IgG) or anti-rabbit IgG (Life Technologies). Secondary antibodies were diluted 1:8'000 in Washing buffer (unless otherwise stated). Slides were then washed 4x with 2.5 ml/pad of washing buffer. The incubation chambers were then unmounted and slides were washed 2x with milliQ H_2O and allowed to dry for 30' at RT.

To confirm specificity of Ab detection, some pads were incubated only with the secondary Ab and were termed as blank.

One every 9 slides (corresponding to slide #5 for every tray) was stained with Fast Green FCF to detect the total protein content of each spot. Each slide was quickly made wet with PBS, which was then removed. Then, slides were stained for 45' at RT on a rocking platform with FCF staining buffer. Afterwards, slides were washed 2x for 15' with destaining buffer. Then, 2x for 5' with milliQ water, and allowed to dry at RT for at least 30'. Staining was performed in specific cassettes to avoid contaminating the incubation chambers with FCF which would lead to unspecific staining.

The slides were scanned at high resolution (21 um) using a 700 nm wavelength on an Odyssey scanner (LI-COR, NE, USA). The images were numbered according to the group of samples (AI, AII, AIII, and AIV), the slide printed (s1 to s9), the tray number they belong (t1 to t8).

Signal intensities were extracted from the images using the GenePix Pro 7 software (Molecular Devices, CA, USA) and saved as .gpr files with the same naming as the image and the definition of the top or bottom array with the _T or _B suffix.

2.2.5. RPPA analysis

The quality control prior to statistical analysis of miRNA gain-of-function effects on target protein was performed by me using the RPPanalyzer package (v. 1.4.3) (von der Heyde, Sonntag et al. 2014).

The remaining analyses described in this section were performed by the collaboration partner Astrid Wachter (Department of Medical Statistics led by Prof. Tim Beißbarth, University Medical Center Göttingen).

All of the analyses described here were performed in the statistical environment R.

2.2.6. Quality control

Quality control of the RPPA incubation includes two steps:

a) Dilution series control
b) Target vs blank control

The dilution series control plots the signal intensities of the four dilution curves against their concentration, for each separate target. In the same plot, it is also present the signal of the blank incubation performed with the same secondary Ab than the target-specific incubation. The aim of the dilution series control is to examine whether the Ab incubated is able to detect the target in a linear manner over a concentration range that exceeds in both directions the one of all the lysates present in the screening.

The target vs blank control plots the signal intensities of the target protein vs the ones of the corresponding blank for every sample present on the array. The aim is to detect if the signal is specific to the primary Ab used or if it derives from the secondary Ab incubation and is therefore experimental noise.

Out of the initial 76 antibodies probed, 62 passed the quality control steps (see 2.1.11 in materials)

2.2.7. miRNA-target interactions

Signal intensities were normalized spot-wise with the Fast Green FCF method (Loebke, Sueltmann et al. 2007). Potential block effects dependent on the spotting of the lysates were removed by shifting the median value of each block to the overall median. At this step, microRNAs which had been transfected as part of the library were checked for their

current status in the miRBase repository (v21). Of the initial 810, 10 miRNAs were defined as *dead* entries of the database, having being recognized through time as fragments of other transcripts or due to the fact that they had previously only been computationally predicted. Another 6 miRNAs were merged with other entries, since they had been recognized as multiple entries of the same miRNA. Therefore the total amount of miRNAs tested downstream was 794.

The package 'limma' (version 3.26.9) (Ritchie, Phipson et al. 2015) was used to identify differentially expressed proteins by comparing for each transfection round the signal intensities of the miRNA OE sample against the two mimic negative control values. Multiple testing correction was performed with the Benjamini-Hochberg method (Benjamini and Hochberg 1995). Visualization of moderated t-statistics (Ritchie, Phipson et al. 2015) was done using the package 'ComplexHeatmap' (version 1.6.0) (Gu, Eils et al. 2016), applying euclidean distance and the complete agglomeration method for unsupervised hierarchical clustering.

2.2.8. Target prediction enrichment analysis

Analysis concept by me and development by Astrid Wachter)

TargetScanHuman (Agarwal, Bell et al. 2015) (v7) and MicroCosm Targets (Enright, John et al. 2003) (v5) databases were inspected to retrieve predicted miRNA-target relationships. TargetScanHuman database information was individually analyzed in regard to conserved and non-conserved miRNA binding sites. Fisher's exact test was used for enrichment testing, individually addressing downregulated miRNA-target pairs and upregulated miRNA-target pairs. Differential protein expression was considered significant below a false discovery rate (FDR) threshold of 0.1%. Only those miRNAs that had at least one significant interaction in the pathway of interest were considered for the testing.

2.2.9. Network analyses

The network analyses were performed at a visual level and at a mathematical level.

a. Visual (analysis conceived and performed by Astrid Wachter)

   The visual analysis consisted in plotting only the most significant interactions (top 100, by statistical significance) between miRNAs and target proteins of a single pathway. microRNAs with the highest amount of edges present in the graph could have been investigated further. Pathway graphs of interactions were generated with the package 'igraph' (version 1.0.1) by Astrid Wachter and then subsequently manually adjusted by me to render them visually helpful.

b. PC score calculations (analysis conceived together with Astrid Wachter and performed by her. Specifically the idea of permuting the FC matrix into a putative miRNA-pathway effect was mine, as well as the suggestion to test the scores against randomization)

The mathematical analysis consisted into three steps.

a) Identification of significant interactions.
b) Permutation of the data matrix from miRNA-target into miRNA-pathway effect.
c) Calculation of the total effect on pathway and statistical testing of the score.

The results of the limma test consisted of two matrices: q-values (q-vals), which identify statistically significant interactions, and log2fold changes (log2FC), which identify to which extent the miRNA affects the protein. The interactions that had a q-value ≤0.001 were considered significant (a).

The log2FC were then permuted according to the known role of the target protein on the pathway. I assigned the role of the protein on the pathway based on the databases GeneCards® (http://www.genecards.org/) and KEGG pathway (http://www.genome.jp/kegg/pathway.html).

Astrid Wachter then permuted the statistically significant FC according to the subsequent rule which aims to define the putative role of the miRNAs on the pathway as a whole:

	miRNA downregulates target −	miRNA upregulates target +
Protein is activator of pathway +	−1	+1
Protein is repressor of pathway −	+1	−1

The output is a data matrix where the +1 and -1 values are indicative of a putative effect of each miRNA on the pathway, based on the modulation of a specific protein.

In this context, the extent of the FC is disregarded (b).

The third step consisted in generating a score, termed pathway coregulatory (PC) score, which was calculated for each miRNA as the sum of all miRNA-mediated effects on the pathway, weighted by the number of measured proteins in the pathway. Permutation testing was performed to assess the miRNA-wise probability distribution of PC scores by 10000x resampling the significant miRNA-protein interactions for each protein. PC scores were considered significant based on an alpha-level of 5%. (c)

The PC score calculation was the method of choice for proceeding with candidate microRNAs in downstream experiments.

2.2.10. WNT pathway stimulation and inhibition

The WNT/β-catenin pathway was stimulated with mouse recombinant WNT3a (Peprotech) at a final concentration of 100 ng/ml, unless otherwise stated. WNT Signaling was inhibited by treating cells with iCRT14 (Santa Cruz, CA, USA) at a final concentration of 10uM, unless otherwise stated. Control treatments were 0.1%BSA in PBS and DMSO, respectively.

2.2.11. Cell proliferation assay

231 cells were plated at a density of $5x10^4$ in 96-well, black plates with transparent bottom (Greiner). The subsequent day cells were transfected as in 2.2.2.3 or treated with the WNT-pathway inhibitor iCRT14. Cell nuclei were stained with Hoechst 33342 (Thermo Scientific) at a final concentration of 20 uM at 48, 72, and 96 hours post-transfection/post-treatment and imaged using an Olympus microscope with 4x objective (Olympus). Quantification of Hoechst+ nuclei was performed with in-built software. 6 technical replicates were performed for each experiment. Median for the 6 tech replicates was used for calculating the ratio of the treatment over the control. The data was loaded into GraphPad Prism 5 software and the ratios of 4 different biological replicates were used for statistical testing (Student's t-test, two tail).

2.2.12. FACS analysis of CD24 and CD44 surface markers expression

231 cells were plated at $2.5x10^5$ per well and then transfected with the same protocol as in 2.2.2.2. In alternative, cells were treated with iCRT14. Throughout the protocol, cells were always detached from plastic surfaces using a non-enzymatic dissociation buffer to prevent damaging the surface markers. Centrifugations were all performed at 4ºC and cells were kept on ice throughout the whole procedure to minimize cell death.

2 days post-transfection/-treatment, cells were split 1:2. 4 days post-transfection/-treatment, cells were detached, counted, and washed 1x with the Assay Buffer (2%FBS in PBS cold). Then 100'000 cells per condition were plated in a deep-well plate where they were resuspended in 100 ul of Assay Buffer containing the appropriate antibody mix.

	ng/well	ul vol 1x
Anti-CD24 APC-conj Biolegend	25	0.25

Isotype G2k APC-conj	25	0.25
Anti-CD44 PE-conj	120	0.6
Isotype PE-conj	120	0.6
Assay buffer		100 ul

Samples were stained for 45' on ice, protected from light. After 2x washing in Assay Buffer, cells were transferred to a FACS tube through a filtering cap and measured at FACS Canto II with the voltage settings listed.

Detection channel	Voltage
Forward Scatter	181
Side Scatter	410
PE	342
APC	460

6 separate biological replicates were performed, with 2 technical replicates each.
Analysis of cell population percentages was performed on FACS Diva software, with gating based on the isotype controls for each specific sample.
The data was loaded into GraphPad Prism 5 software tested with a two-tailed Student't t-test to compare percentages of treated vs. control-treated samples.

2.2.13. Mammosphere formation in ultra-low attachment conditions

231 cells were plated at $2x10^5$ per well and transfected the next day with the same protocol as in 2.2.2.3. In alternative, cells were treated with iCRT14. Two days after transfection, cells were detached, counted, re-suspended in mammosphere medium (see 2.1.9) and plated in ultra-low attachment 6-well plates at a final density of 530 cells/cm^2. For cells treated with iCRT14, the treatment was continued also during the mammosphere formation. 7 days after transfer, 1 ml of fresh mammosphere medium was added to each well. 14 days after transfer, cells were stained for 30' with Calcein AM (final conc. 1uM).
Imaging was performed on a Zeiss Cell Observer equipped with EC Plan-Neofluar 2.5x/0.075 objective.
The subsequent parameters were used for the image acquisition:

- Excitation source: LED Colibrí LED at 470 nm at 100% of power
- Filter block: 38 HE GFP

- Detector: gray scale CCD camera AxioCam.
- 90 msec exposition
- Tile function: 90% coverage of well with 5% tiles overlap

Images where exported as tiff files and processed using the Fiji software, using the analyze particle function. Particles bigger than 50 µm in diameter were counted as spheres.

Due to the possible subjectivity in the analysis of images deriving from the mammosphere formation assay, the experiment was performed blinded. A member of the division would re-label the tubes prior to plating the cells in ultra-low attachment conditions. After imaging and quantification, I was given the key to resolve the sample origin.
The percentage of sphere-forming cells was calculated for every well as (number of spheres/number of cells)*100.

2.2.14. Invadopodia formation on fluorescent gelatine

a. Fluorescent gelatin coating of glass coverslips

Glass round coverslips are pre-coated with Poly-L-lysine (final conc. 0.1 mg/ml in PBS) for 20' at RT. After 1x wash with PBS, they were primed with 1 ml glutaraldehyde 0.5% in PBS for 15' at RT. After 3x washes with PBS, coverslips were coated with a fluorescent gelatin mix (Oregon Green® 488 Conjugated Gelatin diluted in clear gelatin 1:8) for 15' protected from light. After 3x washes with PBS, the glutaraldehyde was reduced by a short and sharply timed 3' incubation with sodium borohydrate freshly prepared at 5mg/ml. The coverslips were then washed 6 to 8 times with PBS to get rid completely of the sodium borohydrate. Coverslips were stored at most 1 week at 4ºC prior to use.

b. Invadopodia formation assay

231 cells were plated at $2x10^5$ per well and transfected the next day with the same protocol as in 2.2.2.2. Two days later, 50'000 cells were transferred in a well of a 12-well plate containing a gelatin-coated coverslip. 10 hrs post-transfer, cells were fixed in 4% PFA, permeabilized in 0.3% Triton X-100, and subsequently blocked in 3% BSA in PBS.
Coverslips were then incubated O.N. at 4ºC with Ab anti-TKS5 (1:300 in 1% BSA-PBS). The next day, they were washed 3x in 1% BSA-PBS, and then stained with DAPI (1 ng/ml), Rhodamin phalloidin (1:400), and a goat anti-rabbit Alexa 633 conjugated secondary Ab (1:400) for 1h 30' at RT. After 3x washes in PBS and 1x in milliQ water,

the coverslips were mounted on glass slides with the ProLong Gold medium and allowed to dry overnight prior to imaging.

Coverslips were inspected on a confocal Zeiss LSM 780 microscope equipped with a Plan-Apochromat 63x/1.40 Oil Ph3 M27 objective.
The detection of different fluorophores was performed with the subsequent wavelengths.

Structure	fluorophore/stain	excitation	emission	laser
Nuclei	DAPI	405	415-478	UV diode
Gelatin	Oregon Gr 488	488	498-550	Argon Multiline
Actin	Rhodamin	561	571-620	DPSS
TKS5	Alexa 633 secondary Ab	633	643-660	HeNe 2

For each coverslip, at least 60 cells were counted and defined as positive for invadopodia formation if they presented actin foci concomitant with TKS5 staining and clear holes in the fluorescent gelatin layer. Representative images of invadopodia^{+} cells were acquired as z-stacks of 18 images for a total of 5.5 um.

The experiment was performed in 3 biological replicates, with 2 coverslips quantified for each experimental condition in each replicate. Due to the possible subjectivity in the quantification of cells from the fields, the experiment was performed blinded. A member of the division would re-label the tubes prior to transfection. After imaging and quantification, I was given the key to resolve the sample origin.

Percentages of positive and negative cells were plotted as mean and standard error of the mean (statistics: Student's t-test, two-tailed).

2.2.15. 3' UTR cloning

To prove direct interactions between microRNAs and target mRNAs, 3'UTR cloning downstream a luciferase reporter gene is a necessary step. The section describes all the steps taken to produce the plasmids used for the downstream luciferase assays (2.2.11).

a. Target prediction

The TargetScan (v 7.1) database was used as the main platform for target prediction. A total of 23 putative binding sites for miR-193b, miR-409, and miR-494, were identified on the 3'UTRs of 9 target genes. TargetScan also describes the most prevalent transcript

isoform, according to 3P-seq tags. The sequence of the suggested 3'UTR was downloaded by the ENSEMBL database and uploaded in the Benchling online tool to proceed with primer design. For APC and CTNNB1, some nucleotides of the ORF were included in the sequence since the first predicted binding site was too close to the end of the ORF to design optimal cloning primers.

b. Primer design

Primers were manually designed to span as many binding sites as possible, however it wasn't possible for all the genes to clone full UTRs and some were divided into smaller portions (termed cloning area 1, 2, etc)

A total of 15 fragments corresponding to the 9 UTRs were cloned.

Primers were designed to have a GC content between 35 and 55%, Tm 50ºC to 60ºC (matched FWD and REV with max 1ºC difference between each other), and a maximum of 2 kB of fragment. FWD and REV primers were designed to have XhoI and NotI restriction sites respectively for end-cloning into PsiCheck2 vector.

All primers are listed in section 2.1.13.

c. PCR to amplify fragments

cDNA from MDA-MB-231 was generated with the Reverd Aid H Minus M-MulV RT kit, using oligo-dT as RT primers

PCR reaction mix	x1
5x Phusion HF Buffer	10
10 mM dNTPs	1
primer A (10uM)	2
primer B (10uM)	2
template cDNA (~200ng)	1.3
Phusion Hot Start	0.5
H2O	32.7
50 mM MgCl	0.5
total vol	50

PCR thermic profile			
Step	temperature	time	cycles
Initial denaturation	98°C	5'	
Denaturation	98°C	10"	35 cycles

Annealing	Optimized	30”	
Extension	72°C	2’	
Final Extension	72°C	10’	

Annealing temperature was optimized for fragments amplified with primers that had similar Tm. The temperatures were ranging between 46ºC and 57ºC. CSNK1A1 amplification required a 2-step PCR where annealing was skipped and the profile proceeded directly to extension, due to the high Tm of the primers.

d. Cloning into pCR®2.1 vector to screen for fragment sequences

PCR products were run on a 2% Agarose gel and fragment of the correct size were excised and purified with the Wizard SV Gel and PCR Clean-up kit. After addition of an adenosine to each end of the purified products, the fragments were cloned into pCR®2.1 vector by an O.N. ligation at 18ºC. The subsequent day DH5α bacteria were transformed with the ligated vector and plated onto XGal coated LB-Agar+Amp plates.

This technique allows to quickly identify clones which had an integration event, since the self-ligated vector is able to process the XGal and therefore colonies are blue.

The next days, colonies which were white were picked, growth in 5 ml of LB+Amp broth, purified with QIAprep Spin Miniprep kit, and send for sequencing wit M13 FWD and M13 REV primers.

Plasmid that presented the correct sequence were digested O.N. at 37ºC with XhoI and NotI enzymes.

Digestion reaction	x1
NEB3 buffer	3
Xho I	1
Not I	1
BSA 100x	0.3
template (mini) 3 ug	24.5

The subsequent day, the digestion products were run on a 2% Agarose gel and bands of the correct size were excised and purifies as above stated.

e. Cloning into psiCheck2

The fragments were then subcloned into a psiCheck2 vector predigested with XhoI and NotI, as well as treated with Antarctic Phosphatase to help preventing self-ligation of the backbone. In the ligation reaction, PmeI restriction enzyme is added to reduce the chance of empty vectors since it would cut in the multiple cloning site of psiCheck2 in case it had previously been digested only with one of the two proper restriction enzymes.

Ligation reaction	x1
PmeI	1
Template (UTR)	to 6.6 fragment + H_2O
10X Ligation Buffer	1
psiCHECK_Xho-Not	0.46 (=50ng)
T4 DNA Ligase	1
Final volume	10

Ligation proceeded O.N. at 18ºC.

The subsequent day DH5α bacteria were transformed with the ligated vector and plated onto LB-Agar+Amp plates. The next days, colonies were picked, growth in 5 ml of LB+Amp and the plasmids were purified with QIAprep Spin Miniprep kit. After sequencing to confirm the correct insertion, bacterial clones were cultured in 50ml of LB+Amp broth and plasmids were purified with NucleoBond Xtra plasmid purification kit (Machery-Nagel).

f. Mutagenesis of miRNA binding sites

To prove that the miRNA affects the luciferase with a direct interaction with the 3'UTR it is necessary to perform rescue experiments by mutating the putative miRNA binding sites. Here, we aimed at mutating the 3 nucleotides of the 3'UTR which match the nt 3-5 of the miRNA seed.

Mutagenesis primers (listed in 2.1.13) were manually designed to anneal with perfect Watson-Crick base pairing at least 20 nt upstream and downstream the 3 nt to mutate.

Mutagenesis PCR was performed with the QuikChange Lightning Site-Directed Mutagenesis Kit.

Mutagenesis PCR mix	x1
dsDNA WT	100 ng
FWD mutagenesis primer	125 ng
REV mutagenesis primer	125 ng

dNTP mix	1
QuikSolution reagent	1.5
10x reaction buffer	5
QuikChange Lightning enzyme	1
NF-free H20	to 50

PCR thermic profile			
Step	temperature	time	cycles
Initial denaturation	95°C	2’	
Denaturation	95°C	20”	18 cycles
Annealing	60°C	10”	
Extension	68°C	4’ 30”	
Final Extension	68°C	5’	

For 3’UTRs containing multiple predicted binding sites, the mutagenesis of all was performed with the QuikChangeLightning MultiSite kit.

	x1
10× QuikChange Lightning Multi reaction buffer	2.5
QuikSolution (DMSO)	0.75
ds-DNA template (prediluted to 100 ng/ul)	100 ng
primers for each binding site	100 ng/each
dNTP mix	1
QuikChange Lightning Multi enzyme blend	1
NF-h20	to 25

PCR thermic profile			
Step	temperature	time	cycles
Initial denaturation	95°C	2’	
Denaturation	95°C	20”	30 cycles
Annealing	55°C	10”	
Extension	65°C	5’	
Final Extension	65°C	5’	

Immediately following each of the mutagenesis PCR, 2 ul of DntpI enzyme were added to each reaction and incubated at 37ºC for 5'. DH5α bacteria were then transformed with the PCR products and growth on LB-Agar+Amp plates. Colonies were picked and grown in 5 ml LB+Amp broth. Plasmids were harvested with QIAprep Spin Miniprep kit. After sequencing to confirm the correct mutations, bacterial clones were cultured in 50ml of LB+Amp broth and plasmids were purified with NucleoBond Xtra plasmid purification kit (Machery-Nagel).

Prior to using the plasmid in luciferase reporter assay, they were all sent for sequencing with psiCheck2 specific primers to confirm that the cloned portions represented the correct sequence, as well as the desired mutations.

2.2.16. Luciferase reporter assay

Reporter assays were exploited throughout the project to test two different hypotheses.

a) 231 cells with stably integrated firefly luciferase under the control of 6 responsive elements for β-catenin (see 2.2.1) were used to test whether miRNA O.E. affected responsiveness to WNT stimulation.
b) 3'UTR reporter assays were instead performed to test whether there is a direct interaction between the miRNA of interest and the target genes. For this second type of assay, the activity of both firefly and renilla luciferase were assessed in order to normalize for transfection efficiency.

For an overview of the timeline of experiments, see end of the chapter

Cells plated in white bottom 96-well plates and transfected/treated/stimulated were washed 1x with PBS and then lysed into 50 ul of M-PER buffer supplemented with DTT (final conc 2 mM). After 10' of lysis, the plates were read on a TECAN plate reader equipped with injectors primed with luciferase mixes. In the 3'UTR reporter assays, the lysates were supplemented with 50 ul of PBS, then divided into two different plates in order to read both firefly and renilla activities.

For β-catenin reporter experiments, firefly signal intensities were averaged through 6 technical replicates, and then a ratio between the miRNA transfection and the control transfection was calculated. Data representation and statistical testing were performed on GraphPad Prism 5 software. Ratios from 4 biological replicates were used for statistical testing (two-tailed, Student's t-test).

For 3'UTR reporter experiments, the data was first normalized for transfection efficiency by dividing renilla over firefly signal intensity. Ratios were then averaged through 4 technical replicates. Data was then normalized over an empty vector control to normalize for miRNA effects on the plasmid backbone. Eventually, a ratio between the miRNA transfection and the control transfections was calculated. Data representation and statistical testing were performed on GraphPad Prism 5 software. Ratios from 3 biological replicates were used for statistical testing (two-tailed, Student's t-test).

Firefly substrate/well	Fin Conc.(mM)	Stock Conc.(mM)	x1
Tricine	40	500	4
ATP	2	100	1
MgSO4	10	100	5
EDTA	0.5	500	0.05
DTT	10	1000	0.5
CoA	1	20	2.5
Luciferin	1	25	2
ddH2O	-	-	34.95

Renilla Substrate/well	Fin Conc.(mM)	Stock Conc.(mM)	x1
Coelenterazine	0.05	2	1.25
EDTA	8	500	0.8
DTT	10	1000	0.5
ddH2O	-	-	47.45

2.2.17. Gene and miRNA expression analyses

1. RNA extraction

Cells cultured in 6-well plates and transfected were lysed in 700 ul of QIAzol reagent. Total RNA extraction was then performed using the miRNeasy mini kit.

Samples belonging to WNT stimulation and inhibition experiments were additionally digested on column with DNaseI to deplete the residual genomic DNA from the preparation.

2. miRNA quantification

RT was performed using TaqMan® MicroRNA Reverse Transcription Kit.

miRNA-RT mix	x1
Nuclease Free Water	1.664
10x RT Buffer	0.6
100mM dNTPs	0.06
RNAse Inhibitor	0.076
Multiscribe	0.4
Total pre-mixed with gene specific primer	2.8
5x Primer	1.2
Total gene-specific	4

Thermic profile	
Time	Temp.
30’	16°C
30’	42°C
5’	85°C
∞	4°C

RT product was diluted 1:5 in NF-H_2O. Quantitative PCR was performed with TaqMan® Universal PCR Master Mix, no AmpErase™ UNG. Assays listed in 2.1.13.

Component	x1
20x primers mix	0.45
Universal PCR Master Mix	3.6
Nuclease Free Water	0.95
total	5

Thermic profile		
Time	Temp	
2'	50°C	
15'	95°C	
15''	95°C	x 45 cycles
60''	60°C	

3. mRNA

RT was performed using with ReverdAid H Minus First Strand cDNA Synthesis Kit.

An oligo(dT) annealing mix was added to 2 ul of diluted RNA (100ng to 5 ug)

oligo(dT) annealing	x1
oligo(dT) primers (100 pmol/µl)	1
RNase-free water	9
Total	10

Then incubated 5' at 70ºC, then cool on ice. Master mix was then added.

RT master mix	x1
5x reaction buffer	4
RiboLock Ribonucl Inhibitor (20u/µl)	1
10mM dNTP mix	2
Reverd Aid H Minus M-MulV RT (200u/µl)	1
Total	8

Thermic profile	
Time	Temp.
5'	37°C
60'	42°C
10'	70°C
∞	4°C

RT product was diluted in NF-H_2O and qPCR was performed with the 2x SsoAdvanced™ Universal SYBR® Green Supermix.

SYBR Master Mix	x1
SsoAdvanced™ Universal SYBR® Green (2x)	5
FWD primer (250nM final)	0.25
REV primer (250nM final)	0.25
cDNA	2
NF-H20	2.5
final	10

Thermic Profile		
95°C	30"	
95°C	10"	x 45 cycles
60°C	1'	
melting curve analysis		

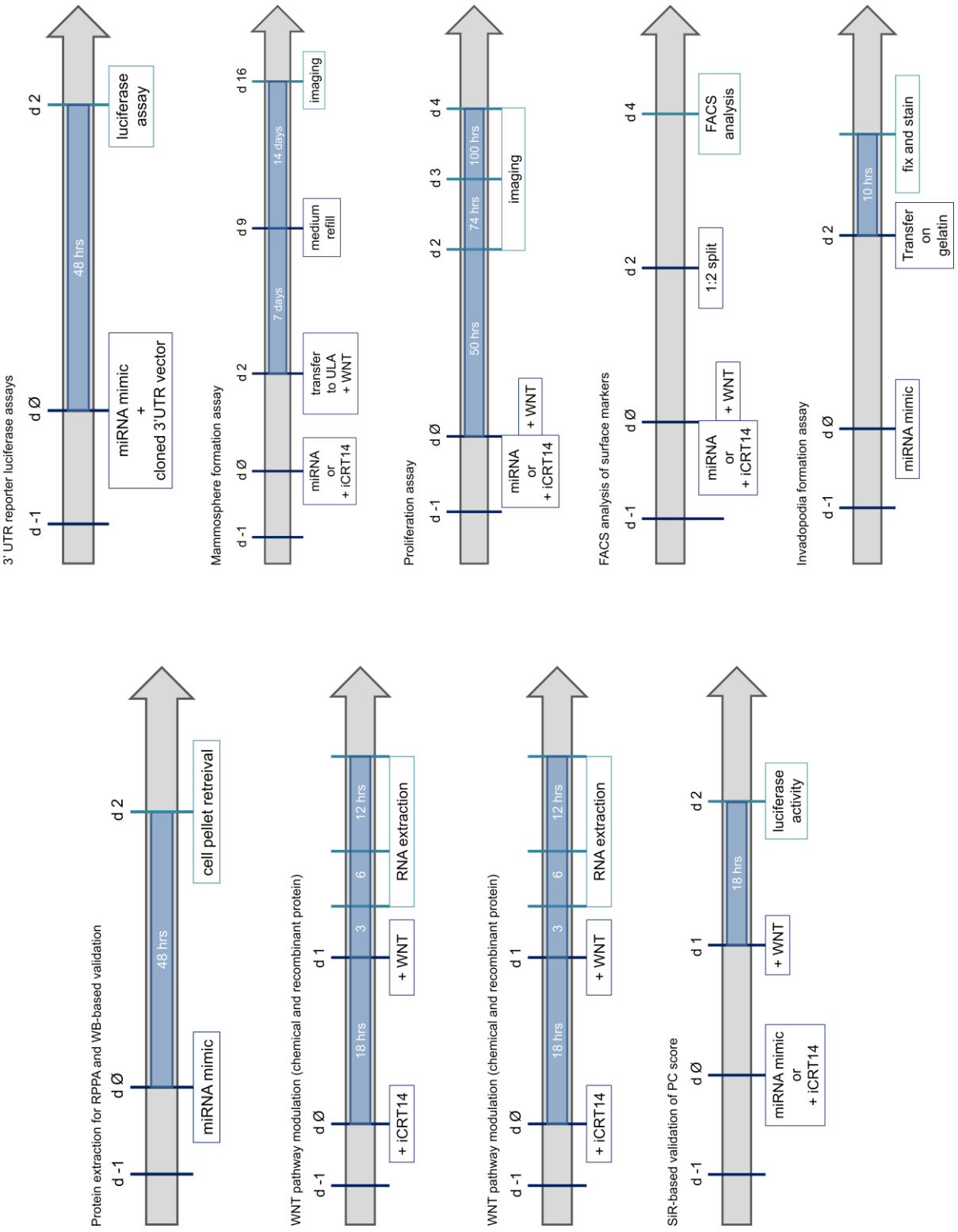
Data was analyzed with manual thresholding and ΔRn on SDS 2.2.2 software. Ct values were analyzed with the ΔΔCt method.

2.2.18. Western Blot

Two days after transfection (as 2.2.2.2), cells were pelleted using trypsin. Cell pellets were lysed in M-PER buffer (Perbio Science) and after 30' rotation in cold room, the lysates were cleared at 13'000 rpm, 15', 4ºC. Protein concentration was assayed with BCA kit. Prior to gel loading, samples were mixed with 6x Rotiload and boiled for 5' at 95C. 15 ug of protein were loaded per lane on 4–20% Mini-PROTEAN® TGX™ precasted gel and run 1h 30' at 120V. Proteins were then transferred onto a PVDF membrane using a semi-dry cassette for 1h at 25V. Membranes were blocked in 1:1 Rockland-TBS for 1h at RT, and then incubated overnight with primary Ab specific for the protein of interest. After washings in TBS-T, membranes were incubated for 1h at RT in light-protected condition with secondary Ab conjugated with 600 or 700 nm emitting fluorophores diluted in TBS-T. After washings in TBS-T, membranes were imaged using an Odyssey scanner. The same membranes were then incubated overnight in Rockland:TBS with primary antibody for normalizers. After washing and incubation with secondary antibody emitting in the different channel, membranes were imaged to normalize for uniform loading. Signal intensity from the band was extracted with the imageStudio software. Ratio over the normalizer data was performed to correct for gel loading.

Primary and secondary Ab were used with the dilutions listed in 2.1.11.

Timepoints of experiments



3. Results

Dissecting the miRNome impact on metastatic pathways in triple negative breast cancer

3.1. High throughput targeted proteomics

The workflow of the proteomic screening is outlined in *figure 3.1*. The first step corresponds to the generation of protein lysates of MDA-MB-231 cells transfected with a library of 800 miRNAs. This material was prepared by Stefan Uhlmann, a previous PhD student of the division of Molecular Genome Analysis (Uhlmann, Mannsperger et al. 2012). The samples had been optimized for assaying miRNA gain of function effects with reduced off-targets effects. Dr. Uhlmann had transfected cells with a range of mimic concentrations (1 to 100 nM) and evaluated via Western Blot the knock-down of proteins for miRNA-target interactions already validated in literature. He determined that 25 nM was delivering the best cost-benefit performance. Indeed, at higher concentrations (50 and 100 nM) only a marginal additive effect was detectable, accompanied by a higher chance of off-target effects.

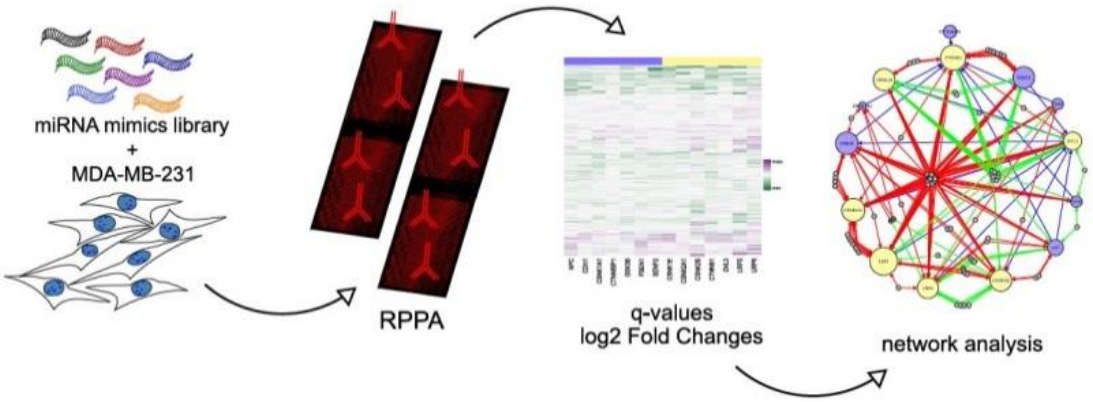


Figure 3.1 Workflow of the miRNome screen

At the beginning of my project, the three pathways to investigate (c-Met, integrin signalling, and WNT/β-catenin) were defined. RNA-sequencing data of MDA-MB-231 available *in house* was exploited to define whether the mRNAs corresponding to targets belonging to the three pathways of interest were expressed. Antibody validation was performed on a panel of cell lines as described (Golan-Lavi, Giacomelli et al. 2017). This antibody validation protocol characterizes the specificity of antibodies via Western Blot not only in the cell line

of interest, but in additional ones as well, increasing the confidence in RPPA signal. The antibodies of interest which passed the WB validation were probed on arrays spotted on nitrocellulose coated slides (RPPA). After data quality control, our collaborators Astrid Wachter and Tim Beißbarth (Department of Medical Statistics, Universitätsmedizin Göttingen) quantified the effects of single miRNAs on target protein expression (q-values and log2 fold changes). The data was then exploited to define a network of interaction which would be pathway-specific. To do so, we developed in collaboration with Astrid and Tim a score that summarize the effects taking into consideration the protein role on the networks (network analysis).

3.2. Identification of targets of interest

The targets belonging to the three pathways were defined based on the KEGG pathway database repository (Kanehisa, Sato et al. 2016). WNT/β cat canonical pathway (map04310) as well as focal adhesion signalling (map04510) were queried as provided, while for c-Met, the network was built based on map04151 and map05200, corresponding to PI3K-Akt signalling and pathways affected in cancer networks, respectively (Kanehisa, Sato et al. 2016).

Considering different isoforms for receptors or kinases present in the pathways, the total amount of target selected was 155. Of those, 34 were not expressed in the cell line of interest. Therefore, 121 candidate target proteins remained listed.

Antibodies to probe the corresponding target were searched. However, few targets proved very challenging and were therefore discarded *a priori*:

- Membrane proteins for which antibodies were available uniquely for the extracellular portion. This was an experimental constraint deriving from the fact that the cell had been trypsinized prior to protein isolation. As a consequence, the experimental procedure might have affected potential signal reduction quantified via antibody-detection.
- Proteins with multiple isoforms. In particular WNT ligands and Frizzled receptors, Integrins and myosin light chain kinases. For many of these targets, antibodies were not available for isoform-specific detection. Therefore, rather than risking a highly complex deconvolution analysis downstream, the targets were not probed.

The total amount of targets for which antibodies were searched and tested were 95. Of those, 76 passed quality criteria in WB analysis and were therefore probed in the HTS to quantify of miRNA-mediated effects on proteins. Of these, 62 passed the quality control of the RPPA pipeline and the data gathered from their incubation was used for downstream analyses.

3.3. miRNA regulate only mildly their targets

The effect of the miRNome on the expression of all our targets of interest was quantified with the limma method by Astrid Wachter. Every miRNA OE sample was tested against the 4 microRNA mimic negative controls that were transfected in the same round of cell culture (performed by Stefan Uhlmann). Subsequent to the limma test, the Benjamini-Hochberg method was used to correct for multiple testing. For the analysis, we intentionally discarded the z-score method, used in the previous miRNome screen (Uhlmann, Mannsperger et al. 2012). This choice was based on the fact that within sample groups, a variable amount of miRNAs belonging to the same family (i.e. putatively targeting a similar subset of targets). Since the z-score exploits the average and standard deviation of the whole group of samples to define the miRNA-dependent effect, the presence of miRNAs putatively affecting similar subgroups could create a bias that could potentially hinder significant interactions. Therefore, the limma test was preferred in order to control this form of bias.

The final output of the analysis corresponded to fold changes in log2 scale as well as q-values to describe the statistical significance of the variation of target protein expression compared to the mimic negative controls.
miRNA-target interactions are represented by means of heatmaps of the moderated t-statistics, a visualization suggested by Astrid Wachter as a form that combines both pieces of information (log2FC and q-value). Unsupervised hierarchical clustering was performed on the miRNA-mediated effects, applying euclidean distance and the complete agglomeration method.
The results are shown in *figure 3.2 A* for WNT/β-catenin pathway and in *supplementary figures 1* and *2* for the other two pathways.

As it can be noted from the heatmaps representing the t-statistics, within each pathway there seems to be many miRNAs affecting the expression of multiple targets probed. Therefore, to ensure that the interactions exploited in downstream analysis are specific, only interactions which pass a very stringent statistical significance cutoff (q-value ≤ 0.001) were taken into account for all the downstream analyses.

The average of negative significant log2 fold changes throughout the HTS is -0.52 (*fig. 3.2 B*), corresponding to roughly a 30% reduction in expression, consistent with the known mild effect that miRNAs generally induce on the majority of the targets (Baek, Villen et al. 2008, Selbach, Schwanhausser et al. 2008).

Target-wise analysis of the distribution of the log2 FC as well as the q-values by means of volcano plots shows that few targets have a small number of miRNAs which induce very strong effects. However, for the majority of the cases, statistically significant interactions are relatively little in fold change. This is the case for APC (*fig. 3.2.C*), for which we have significant repression of a target already at -0.17 log2FC (roughly 10% reduction in expression). On the other hand, β-catenin fold changes pass the significance threshold much later, with values of – 0.27 and +0.54. In contrast with APC, this is associated with overall stronger repression and overexpression dynamics which range between -1.4 and + 0.9. Volcano plots for all the targets analysed are available in *supplementary figure 3*.

This mild effect on multiple target proteins prompted us to analyse the pathways with a system biology approach. The hypothesis is whether by analysing many mild, but highly significant interactions induced by a single miRNA we are able to identify powerful modulators of the whole pathway.

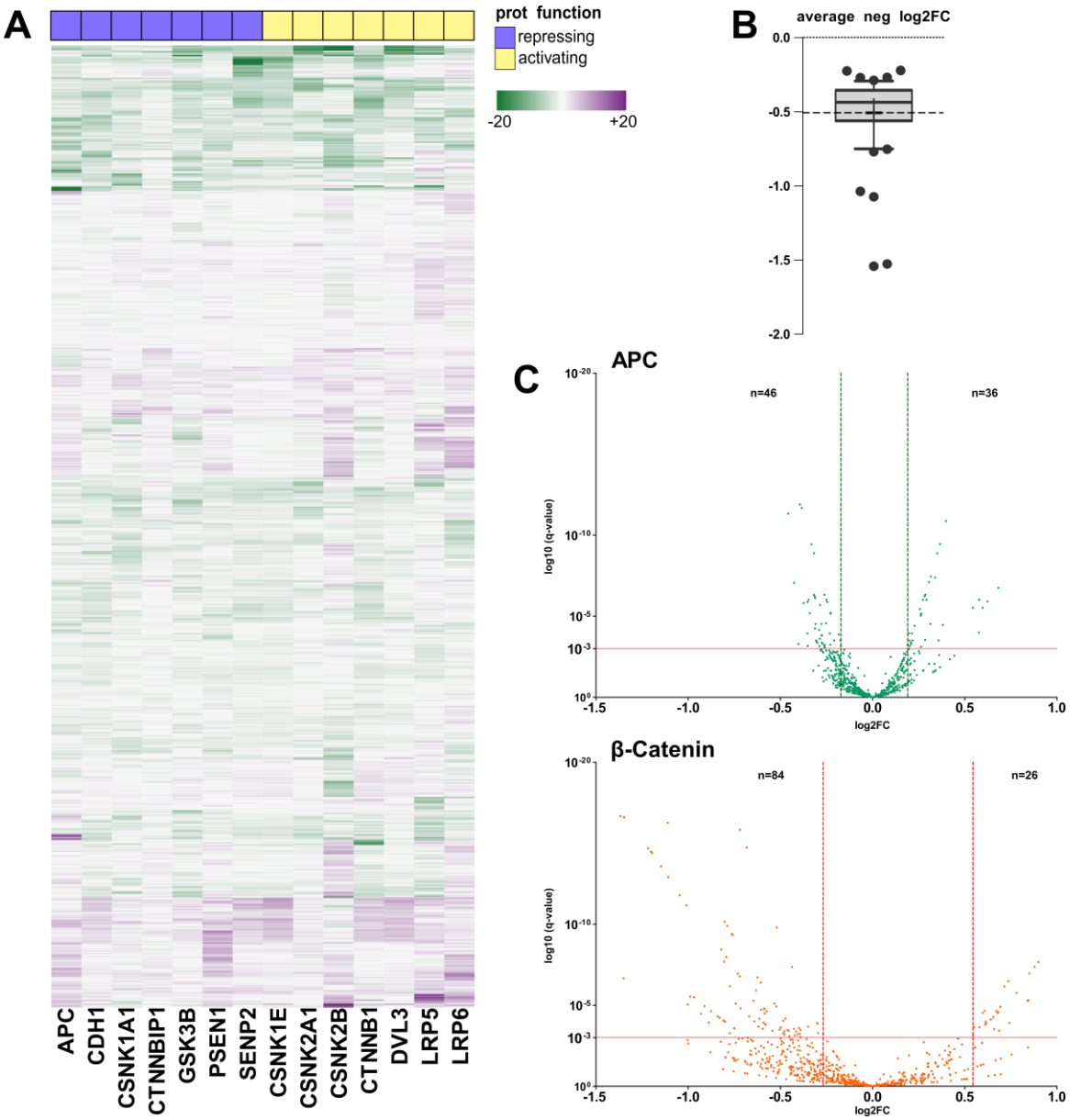


***Fig 3.2 Effect of the miRNome on target proteins*. A** The heatmap represents the moderated t-statistics for the miRNA-target interactions for the WNT/β-catenin pathway. Proteins are given as column and miRNAs as rows. Downregulations of targets is in green, while upregulation is represented in purple. The miRNA effects are clustered applying euclidean distance and the complete agglomeration method. Clustering tree is given in supplementary figure. For clarity, names of the target proteins are listed with the corresponding target gene name. In supplementary table 1 the corresponding protein names are listed. **B** The average negative effect of miRNAs on the targets through the whole screen is represented in a box plot. Whiskers represent the 10th and 90th percentile. Dotted line marks the average negative log2 FC. C The two volcano plots represent the distribution of miRNA-mediated effects on APC (top) and β-catenin (bottom). The x-axis represents the log2FC and the y-axis represents the q-value in log10 scale. The horizontal red line represents the q-value cutoff. The vertical lines show for each protein the minimum log2FC value that passes the q-vals cutoff, both for downregulations and upregulations.

3.4. The downregulation of target proteins is mostly mediated by predicted interactions in the 3'UTR of the corresponding mRNA

The predominant way a microRNA affects target mRNAs and protein levels, is via base pairing its 'seed sequence' (nt 2-7) with one or multiple MBSs positioned in the targets' 3'UTRs.

Therefore, in collaboration with Astrid, the results of the miRNome screen were tested for enrichment of significant interactions within MBS prediction databases. Two different databases were analysed: TargetScan (v7) and MicroCosm Targets (v5). Significant upregulations and significant downregulations were independently tested, with the former serving as an internal negative control. Indeed, upregulations induced by miRNA OE are expected not to be dependent on 3'UTR predicted MBSs. TargetScan (TS) additionally provides information on the conservation of the MBS. Thus, in our enrichment analysis, we kept distinct the MBS conserved among mammals and the non-conserved ones. Fisher's exact test was used to calculate the enrichments, recapitulated in *table 3.1*.

Table 3.1 List of targets analysed with corresponding p-value for enrichment of predicted miRNA binding sites. In yellow are highlighted the enrichments which pass the significance threshold of p-value ≤ 0.05

Gene name	Significant downregulations tested			Significant upregulations tested		
	TS Cons	TS NC	MCT	TS Cons	TS NC	MCT
AKT1	0.000	0.005	1.000	1.000	0.280	1.000
AKT2	0.131	0.884	0.060	1.000	1.000	1.000
APC	0.004	0.009	0.434	0.236	0.465	1.000
BRAF	1.000	0.668	1.000	1.000	1.000	1.000
CCND1	0.000	0.000	1.000	1.000	0.006	1.000
CCND3	0.002	0.000	0.169	1.000	0.045	0.389
CDH1	1.000	0.320	1.000	0.094	0.270	0.375
CDK2	0.000	0.032	0.004	1.000	1.000	1.000
CDK4	0.347	0.461	0.548	1.000	1.000	1.000
CDKN1B	0.000	0.006	0.370	0.202	0.033	0.350

CRK	0.045	0.526	0.524	0.006	0.300	1.000
CSNK1A1	0.013	0.018	1.000	1.000	0.747	1.000
CSNK1E	NA	1.000	0.142	NA	1.000	0.615
CSNK2A1	1.000	0.356	0.257	1.000	0.570	1.000
CSNK2B	1.000	0.079	0.565	1.000	0.600	1.000
CTNNB1	1.000	0.013	0.703	1.000	0.129	0.460
CTNNBIP1	1.000	1.000	1.000	1.000	1.000	0.352
DIAPH1	0.522	0.732	0.275	1.000	0.438	0.615
DUSP6	0.031	0.101	0.410	1.000	1.000	1.000
DVL3	0.516	0.205	0.373	0.184	1.000	0.354
EGFR	0.000	0.000	0.555	1.000	1.000	1.000
FOS	1.000	0.446	1.000	1.000	0.518	0.687
GRB2	0.000	0.000	0.124	1.000	0.717	1.000
GSK3B	0.003	0.001	0.133	1.000	1.000	1.000
HRAS	1.000	1.000	0.344	1.000	0.411	0.531
ILK	0.078	0.548	0.856	1.000	1.000	1.000
JUN	1.000	0.381	0.062	1.000	0.144	1.000
LRP5	1.000	0.641	0.730	1.000	1.000	0.236
LRP6	0.656	0.088	0.782	0.111	0.110	0.814
MAP2K1	0.001	0.166	0.521	1.000	0.711	1.000
MAPK1	0.434	1.000	1.000	1.000	0.233	1.000
MAPK14	0.019	0.000	1.000	0.026	0.026	1.000
MAPK3	0.212	0.015	0.711	1.000	0.713	0.569
MAPK8	0.000	0.447	1.000	0.040	1.000	0.525
MET	0.000	0.000	0.232	1.000	0.197	0.399
MTOR	0.003	0.002	1.000	1.000	0.416	1.000
PAK1	1.000	0.333	0.374	1.000	0.225	1.000
PAK2	1.000	1.000	1.000	0.411	0.463	1.000
PDPK1	0.315	1.000	0.123	1.000	0.426	1.000
PIK3CA	0.003	0.135	0.017	0.185	0.491	1.000
PIK3CB	0.005	0.163	0.253	0.377	0.465	0.597
PLCG1	0.000	0.011	0.010	1.000	0.173	1.000
PPP1CA	1.000	0.101	1.000	1.000	1.000	1.000

PRKCA	0.309	0.085	1.000	1.000	0.249	1.000
PSEN1	0.072	0.178	0.305	0.437	0.325	1.000
PTEN	0.000	0.019	1.000	0.245	0.457	1.000
PTK2	0.020	0.233	0.501	1.000	1.000	0.615
PTPN11	0.001	0.416	1.000	1.000	1.000	1.000
PXN	0.105	0.027	0.397	1.000	0.404	1.000
RAC	1.000	1.000	1.000	1.000	1.000	1.000
RAP1A	0.001	0.014	0.639	0.368	1.000	1.000
RB1	0.064	0.053	0.484	1.000	0.060	0.222
RHOA	0.223	1.000	0.367	1.000	0.132	1.000
ROCK1	0.000	0.593	0.069	1.000	0.260	0.182
ROCK2	0.003	0.000	0.153	1.000	0.412	1.000
SENP2	0.134	1.000	0.295	1.000	0.542	1.000
SHC1	0.013	0.307	1.000	1.000	0.137	0.364
SOS1	0.073	0.701	1.000	0.197	0.716	0.658
SRC	0.035	0.829	1.000	1.000	0.548	1.000
STAT3	0.105	0.006	0.012	0.323	0.507	1.000
TP53	0.053	0.382	1.000	0.288	1.000	1.000

Results show that roughly 45% of the targets probed have a significant p-value when tested for enrichment of predicted conserved binding sites (TS Cons). The percentage drops to 34% for the nonconserved predicted sites (TS NC). The MicroCosm Targets (MCT) returns only a 6.5% of targets with significant enrichment, possibly due to prediction algorithm which is less restrictive than the TargetScan one. For the upregulations, we also identify roughly 6% of targets which have a significant enrichment. This can be explained by the fact that the predictions are often redundant (i.e. false positive) or might be non-functional in specific cellular contexts.

A hypothesis formulated on the basis of the results of the screening performed by Stefan Uhlmann, was that the length of mRNAs' 3'UTRs could correlate with the amount of repressions identified. On the contrary, upregulations, being not dependent on 3'UTRs, should not correlate. Therefore, 3'UTRs sequences of the genes investigate in this screening were analysed to compare their length with

the amount of significant positive or negative interactions. As well, the length of 3'UTRs was compared to already validated miRNA:target interactions (miRTarBase MTI database) (Chou, Chang et al. 2016). All the data is available in supplementary table 1 found on CD due to size limitations.

For specific genes a trend seemed to be there, such as *CDKN1B* (1774 nt.) which shows a clear tendency to upregulation with 224 miRNAs, while only 28 downregulate its expression. However, correlation analysis between 3'UTR length and the amount of significant downregulation did not return a significant p-value. Pearson's r was -0.026 (*fig. 3.3*). The hypothesis that mRNAs with longer 3'UTRs would be under a stronger regulation by microRNAs was therefore refuted.

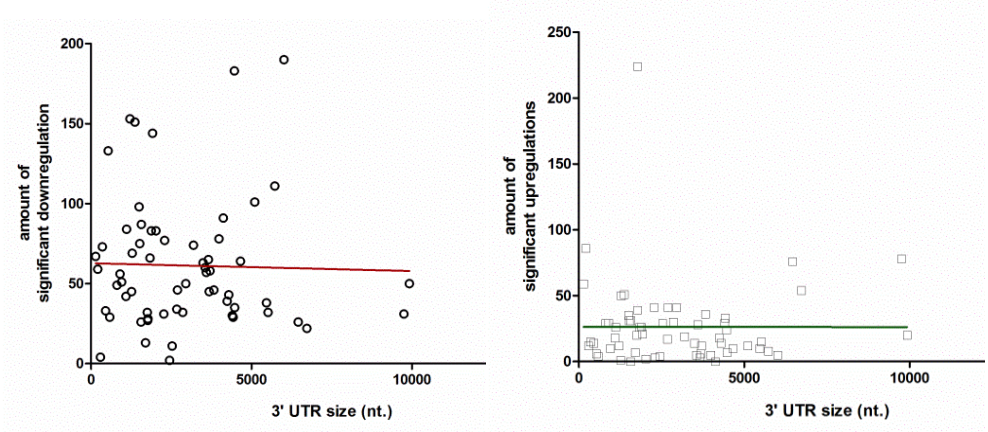


Figure 3.3. **3' UTR length does not significantly correlate with downregulations detected at the protein level.** 3'UTRs lengths (x-axis) are plotted against the amount of significant downregulations (y-axis, left plot) or against the significant upregulations (y-axis, right plot) detected at the protein level.

In summary, in collaboration with Astrid Wachter, the high throughput targeted proteomic approach successfully described the effects of 800 microRNAs on 62 proteins belonging to c-Met, Integrin, and WNT/beta-catenin signalling pathways. Additionally, more than 50% of the downregulations identified were found enriched for predictions of miRNA binding sites in the 3'UTRs of the target mRNAs. The mild effects detected at the protein level and the target prediction enrichment results were in line with previously known biological mechanisms of miRNA-mediated gene expression control, therefore reaffirming RPPA as an appropriate tool to investigate these regulations.

As previously suggested, assigning phenotypical functionality to a miRNA on the basis of a single strong effect on a target can be misleading and might not reflect the actual biological effects achieved intracellularly. Therefore, the HTS data was used to develop a novel network analysis approach to identify miRNAs that are able to affect the pathways of interest.

3.5. Network analysis to unravel the role of miRNAs on complete pathways

The network analyses here described were conceptualized by me and discussed with Astrid in depth. The analysis pipelines were entirely developed and ran by Astrid.

3.5.1. Score development

Two concepts proposed by me were the basis on which the analyses were developed in order to unravel the biological role of miRNAs:

a) The extent of the miRNA-target regulation is not relevant in the network context, therefore only the information of whether the interaction is positive or negative is retained.

b) The role that the targets investigated play within the pathway is fundamental information, as it contextualises the effect of the miRNA on the pathway. Therefore, the miRNA-dependent effect on the pathway as a whole is taken into account as subordinate to the role of the target on the pathway itself.

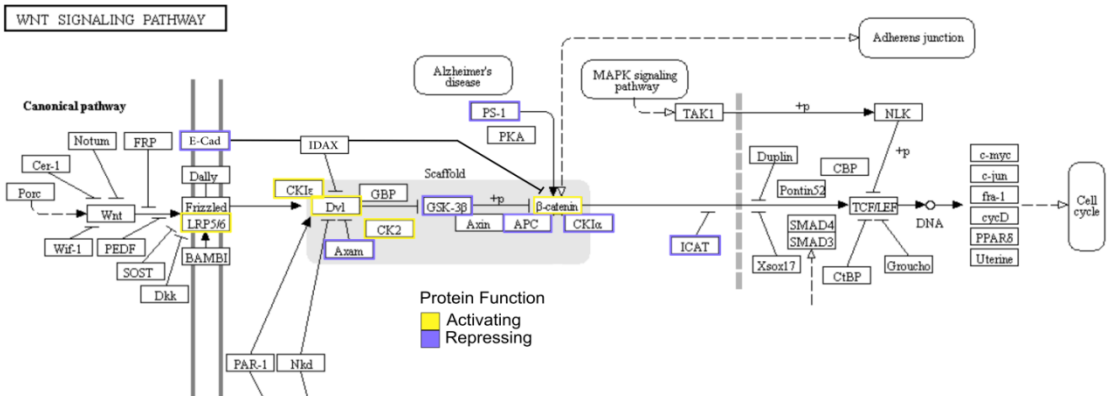


Figure 3.4. WNT/β-catenin signaling pathway and the role of probed targets. The WNT/β-catenin signaling pathway is represented as available from KEGG pathways (Kanehisa, Sato et al. 2016). The target proteins probed in the RPPA screen are colour coded according to their role in the pathway. Protein edges which are activators or repressors of the pathway are marked in yellow or purple, respectively.

The WNT/β-catenin pathway is presented in *figure 3.4* with a colour coding that represents the putative role of the proteins on the respective pathway, an information collected through the KEGG pathway database itself, as well as the GeneCards repository of the Weizmann Institute of Science. In *supplementary figure 4*, the KEGG pathways with the protein functions are shown for c-Met and Integrin signalling.

Based on the hypotheses provided, we permuted the data matrix of significant log2FC to represent the putative effect of the miRNA on the pathway. Specifically, the log2FCs are transformed into a matrix of +/- 1 which reflects the effect of the miRNA on a target and in the meantime integrate in the matrix the role of the protein on the pathway itself (*fig 3.5*). This matrix therefore describes the putative miRNA-network (miR-N) effects. The term miR-N will be used further to identify this matrix.

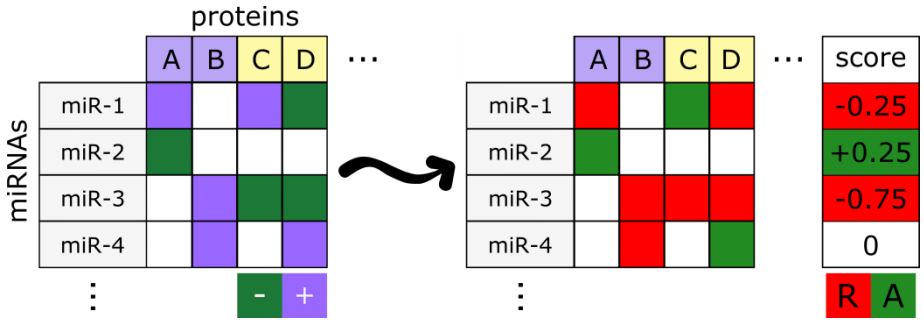


Figure 3.5. Data permutation. On the left, the heatmap of miRNA-mediated effects at the protein level is schematically represented (purple, significant upregulations; dark green, significant repressions). The colour of the protein columns (A-D) represents their role on the pathway investigated (light purple, repressing; light yellow, activating). The arrow represents the permutation which leads to the matrix of +1 (green) or -1 (bright red). The matrix as a whole is then exploited to compute the score which indicates a putative role for the miRNA in the same row (R, bright red, repressive; A, green, activator).

Starting from the permuted matrix, we computed a Pathway Coregulatory (PC) score which represents the sum of the permuted interactions, divided by the number of targets in the pathway (fig 3.5). This last step is carried out in order to normalize the score and therefore render it comparable between the different

pathways analysed. Hence, the score represents the cumulative putative effect of each miRNA on the whole pathway. In *figure 3.6* the distribution of the PC scores for the three pathways are shown.

Of note, the WNT pathway has a more symmetrical distribution of PC scores compared to the other two pathways. This is probably due to the skewing of the biological role of the protein we probed, with the WNT pathway having 50% activators and 50% repressors probed. For Integrin and c-Met, the repressors represent only 9-16% of the probed targets. Since the majority of the significant effects identified in the HTS are downregulations, the PC score is biased toward repression.

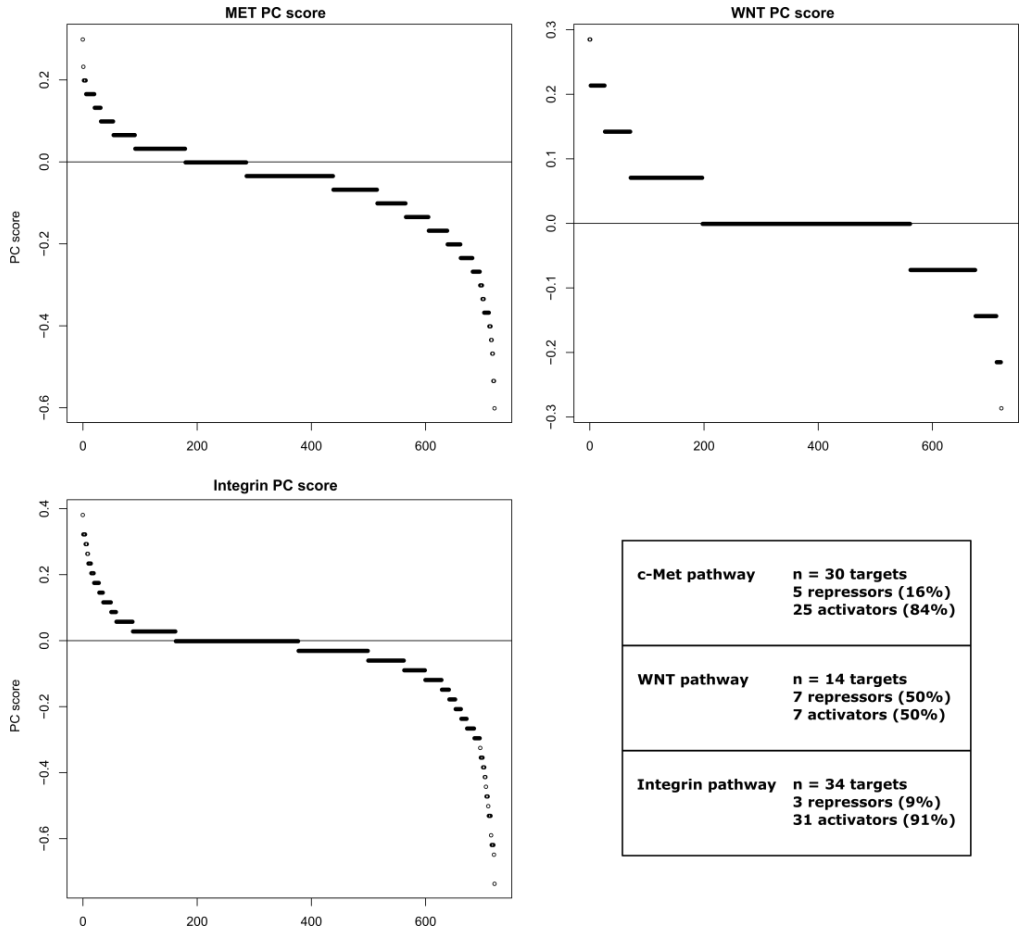


Figure 3.6. PC score distribution. The three panels show the distribution of the PC scores for the three separate pathways. In the x-axes the miRNAs are represented as ranked from the highest to the lowest PC score. On the y-axes the value of the respective PC score is represented. The ranking of the miRNAs is done pathway-wise. On the bottom right corner the box contains a brief description of the number of targets probed by pathway and the respective amount of activators or repressors of the pathway. The percentage in parenthesis specifies how many of the targets of the pathway belong to one or the other category.

Regarding the values of the PC score, the integrin pathway has longer tails with higher absolute values, compared to both the other two pathways. This is not simply due to the larger amount of proteins probed. Indeed, while in the integrin pathway there are more targets, the total amount of statistically significant interactions is lower than in the c-Met pathway, where only 30 were probed. By analysing the amount of interactions per protein, it is possible to see how within the integrin pathway, a smaller number of miRNAs affect single target. As a consequence, it seems that for the integrin pathway there is a higher coregulatory complexity which is reflected by higher PC score absolute values.

To identify which miRNAs with a strong PC score are functional regulators, we proceeded to statistically validate the scores.

3.5.2. Validation of the PC scores: Testing against a random distribution

The randomization testing was proposed by me and discussed in depth with Astrid, who then developed and ran the pipeline.

While the PC score represents still a putative effect on the pathway, there is the need to test for its robustness in statistical terms. Therefore, the matrix of permuted data is resampled 10’000 times/protein (*fig 3.7*). The actual PC score calculated with the data is then tested against the probability distribution of random PC scores and considered significant based on an alpha-level of 5% (p-value ≤ 0.05).

The distributions of random PC scores for two miRNAs which pass the significance threshold for the WNT/β-catenin pathway analysis are shown in *figure 3.8*. For miR-92b-3p it is possible to notice how the actual PC score (red vertical line) is lower than the negative threshold line, and therefore the miRNA is considered a negative regulator of the pathway. For miR-195, the actual score falls instead above the threshold, and therefore it is considered an upregulator of the pathway.

The statistical validation revealed 46 putative repressors of WNT pathway, while only 4 miRNAs as possible upregulators. For the Integrin signalling pathway, 93

miRNAs were listed as putative repressors and 59 as activators. C-Met signalling pathway returned 113 putative repressors and 54 activators.

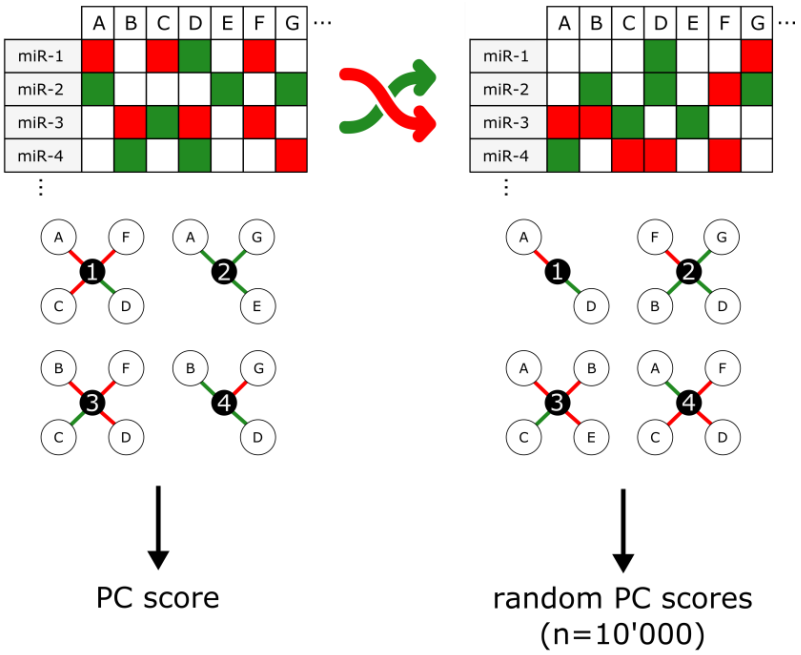


Figure 3.7. PC score randomization. The randomization testing is schematically represented here. On the left side the miR-N matrix from actual RPPA data is represented. This matrix is permuted 10'000 to generate random PC scores (right side). The lower portion represents visually miRNAs (1-4, in black) and the randomized interactions with proteins (A-G, in white).

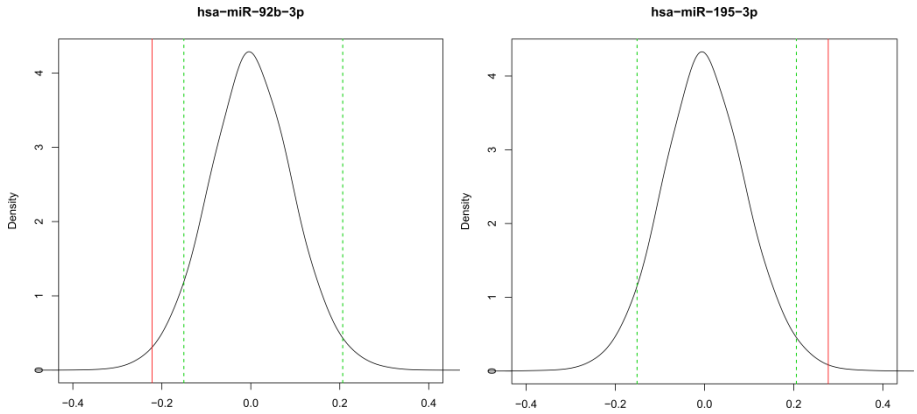


Figure 3.8. Examples of random PC score distribution. The two graphs represent the cumulative distribution of the randomized PC scores for miR-92b (left) and miR-195 (right). On the y-axis is depicted the density, on the x-axis the corresponding PC score. The vertical green dotted lines represent at which PC score value the miRNA would pass the alpha-level of 5%. The vertical red line depicts the actual PC score value for the two miRNA shown. In the graphs, the distribution is presented as smoothened.

The list of miRNAs passing the statistical testing is listed for each pathway in *Supplementary table 2*.

3.6. Experimental validation

The experimental validation of miRNAs ability to regulate whole networks was focused on the WNT/β-catenin pathway for four main reasons:

- By means of a reporter assay for TCF/LEF transcription factors, the activity of the WNT/β-catenin pathway could be quantitatively investigated.
- The RPPA HTS covered well both upregulators and downregulators targets, giving more confidence that the results would not be biased toward repressing miRNA-effects.
- MDA-MB-231 cells activate the pathway by autocrine secretion of WNT7B. As a consequence, we assumed that identified miRNAs might be more relevant regulators of a pathway which is constitutively active (Huguet, McMahon et al. 1994).
- In TNBC the β-catenin pathway activation is regarded as a driver of metastatic pathological features (Dey, Barwick et al. 2013). However, members of this pathway do not seem to be altered at the genomic level, compared to other tumour types, such as colorectal carcinoma. Therefore, investigating the pathway was considered interesting in view of possible clinical applications that might stem from harnessing microRNAs repressing WNT/β-catenin signalling.

To experimentally validate the interactions with a reporter assay, the DKFZ SiR core facility was requested to prepare a cell line which contains the firefly luciferase reporter gene under the control of six TCF/LEF responsive elements (CCTTTGWW, where W is an A or T) and therefore under the regulation of the WNT pathway activity. (*fig 3.9).*

Five different clones were validated by stimulating cells with recombinant WNT3a as well as lithium chloride (LiCl) treatment (*fig. 3.10*). LiCl has been described to inhibit GSK3B, a kinase that phosphorylates β-catenin and promotes its ubiquitination and proteasome-mediated degradation (Stambolic, Ruel et al. 1996).

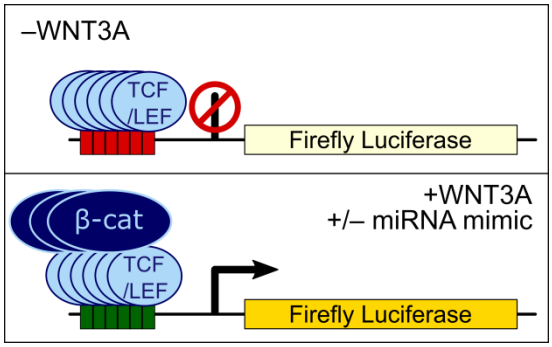


Figure 3.9. Schematic representation of the reporter gene assay. When the pathway is not activated (upper panel), the firefly luciferase is not expressed. When cells are stimulated with recombinant WNT3a, β-catenin translocates to the nucleus and activates transcription of the reporter gene. Upon stimulation and miRNA OE, the luciferase expression should be reduced, at the condition that the miRNA is actually able to repress the pathway.

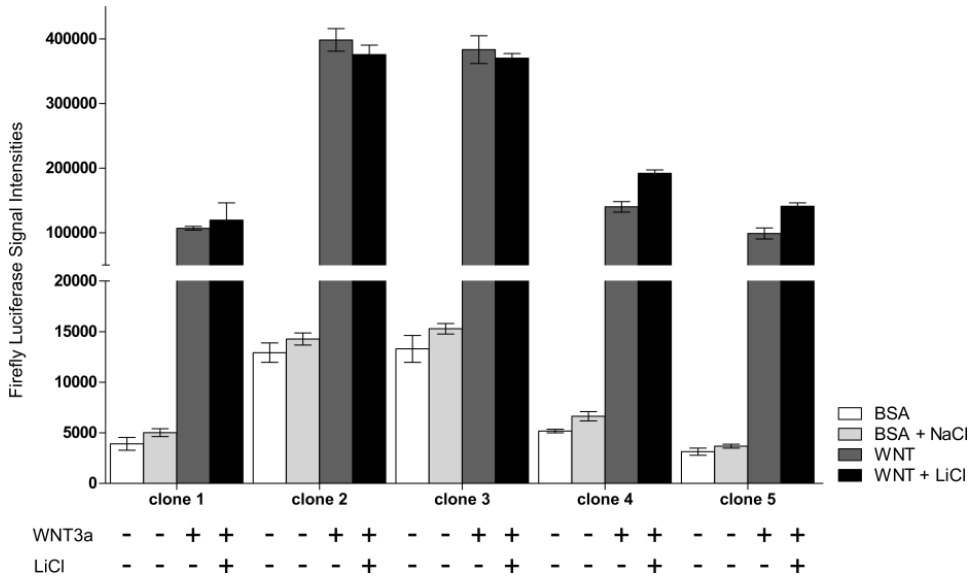


Figure 3.10. Firefly luciferase response upon modulation of the WNT pathway. The five clones of stable isogenic recombinant cells were tested for responsiveness to the WNT pathway by stimulating with recombinant WNT3a alone or in combination with LiCl. Y-axis represents the signal intensities of firefly-induced luminescence, representing the activity of the enzyme.

In the case of MDA-MB-231 it seems that inhibition of GSK3B does not give any additional benefit for activation of the pathway. The presence of a mild luciferase activity even when the cells are not stimulated is most likely due to the fact that MDA-MB-231 are able to activate the pathway by secreting their own ligand, as previously stated.

In order to avoid working with miRNAs which are likely irrelevant in the pathological context, their expression in breast cancer was analysed. miRNAs quantified at ≤10 RPKM in >95% of patients or with a median RPKM ≤ 5 in the TCGA Breast Cancer dataset were considered not expressed (TCGA data portal, see in materials>databases and online portals). These miRNAs have been marked in data tables with a suffix "_NE".
As second filtering step, the stability of the PC score was analysed by increasing thresholds for both q-values and log2FCs. PC values which remained stable in direction of regulation (i.e. positive or negative) indicated miRNAs considered more reliable in their effect on the pathway. This analysis was not further statistically tested; however, it was exploited as a guide when choosing which microRNAs to validate experimentally. In *supplementary table 3*, the PC scores calculated with increasing thresholds are listed (available on CD version due to size limitation).

After expression selection, 19 miRNAs remained listed. On the basis of known literature and the PC score stability, 12 miRNAs were selected for experimental validation exploiting the reporter gene assay.

Of the 12 miRNAs tested, six significantly downregulated reporter gene expression compared to a miRNA mimic control (*fig. 3.11*). On the contrary, two miRNAs strongly upregulated the signal.

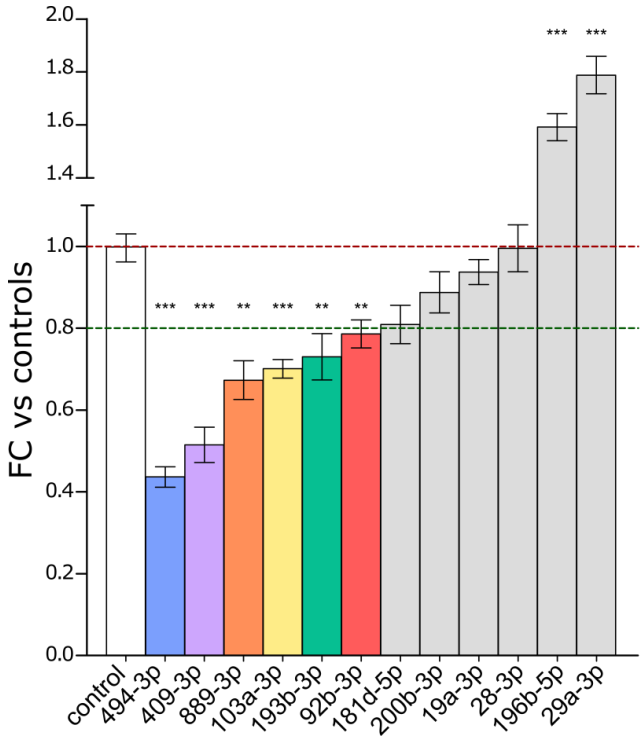


Figure 3.11. Relative luciferase activity of the WNT reporter assay upon miRNA OE. On the y-axis, the relative luciferase activity upon WNT3a stimulation is represented as a fold change compared to the miRNA mimic control condition. Luciferase activity was measured 48 hrs post-transfection and 18 hrs post-stimulation. The green horizontal dotted line represents a downregulation of 20% in luciferase activity. Statistical significance against the control condition (red horizontal dotted line) was calculated with a two-tailed student's t-test. P-values are represented with asterisks on top of the corresponding condition (** ≤ 0.01, *** ≤ 0.001).

To further validate that the miRNAs are not affecting the luciferase signal by reducing the proliferation of cells, the six top repressing miRNAs were tested also in three independent clones of SiR which harbour a renilla luciferase constitutively expressed (*fig. 3.12*).

As positive control for downstream experiments, iCRT14, a chemical inhibitor of nuclear translocation of β-catenin (Gonsalves, Klein et al. 2011), was tested on the reporter cell line to identify the concentration that would most likely resemble the change in the signal that was observed upon miRNA OE (*fig 3.13*).

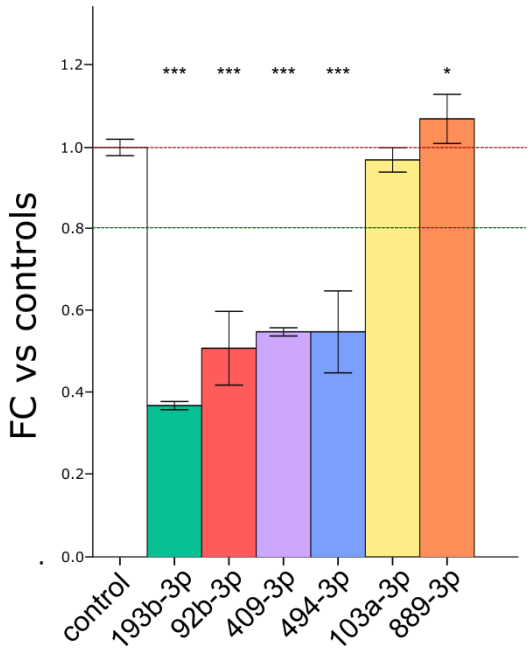


Figure 3.12. Validation of miRNA-mediated repression of the WNT/β-catenin pathway. On the y-axis, the relative luciferase activity upon WNT3a stimulation is represented as a fold change compared to the miRNA mimic control condition. Luciferase activity was measured 48 hrs post-transfection and 18 hrs post-stimulation. The green horizontal dotted line represents a downregulation of 20% in relative luciferase activity. Statistical significance against the control condition (red dotted line) was calculated with a two-tailed student's t-test. P-values are represented with asterisks on top of the corresponding condition (* ≤ 0.05, *** ≤ 0.001).

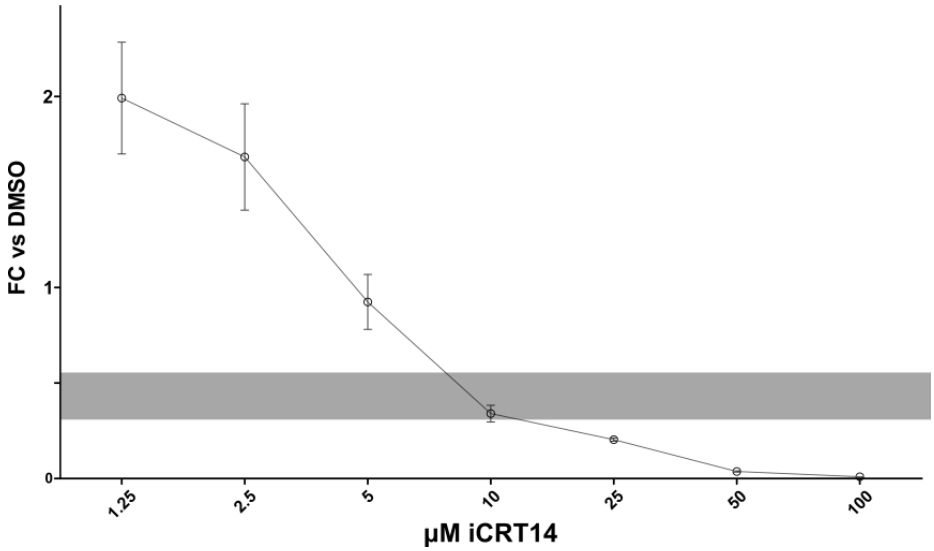


Figure 3.13. Evaluation of chemical inhibition of the pathway. On the y-axis, the relative luciferase activity upon WNT3a stimulation is represented as a fold change compared to the control treatment (DMSO). On the x-axis the different iCRT concentration tested are plotted. The grey rectangle indicates the repression of the pathway that we aimed for, according to the effect exerted by miRNAs in the previous assays.

Considering that the target downregulation aimed for is similar to the one exerted by the strongest miRNA (miR-193b), working concentration for downstream assays was chosen as 10 µM. This concentration is in line with the one used to assess cell proliferation or invasion effects in breast cancer cells (Bilir, Kucuk et al. 2013, Zhang, Liu et al. 2017) or beta-catenin dependent transcription in colon cancer cells (Gonsalves, Klein et al. 2011).
At low concentrations (<5 µM) the chemical inhibitor induced an upregulation of luciferase signal compared to the control-treated MDA-MB-231 (DMSO). This phenomenon wasn't further investigated. However, a possible mechanism underlying this effect could be a feedback loop dependent on a mild repression of the transcriptional activity of beta-catenin.

Concluding, of 12 miRNAs putatively affecting the WNT signalling pathway according to the network analysis, four were consistently able to repress the expression of a reporter gene under the control of six β-catenin responsive elements. Additionally, a chemical inhibitor (iCRT14) was tested to identify the appropriate concentration to mimic the effect of the miRNAs on the pathway-dependent transcription.

To visually summarize the miRNA:target interactions and their effect on the pathway, with the support of Astrid, a network was built (*fig. 3.14*). It is important to notice that the edges do not represent the direct miRNA:target effect. Indeed, the data used to display the network belongs to the miR-N matrix (*fig 3.9*). Therefore, the colours used are indicative of the putative effect of the miRNA on the pathway *via* interaction with the target protein.
It is possible to appreciate how the miRNAs which indeed affected the pathway have a high amount of significant interactions with the targets, supporting the hypothesis that a despite in a mild manner, many interactions are more relevant for the pathway regulation than a single, very strong, downregulation. However, it is imperative to recognize also that the four miRNAs that indeed affect the reporter gene activity (dark blue nodes) are connected to certain targets via activating edges (green).

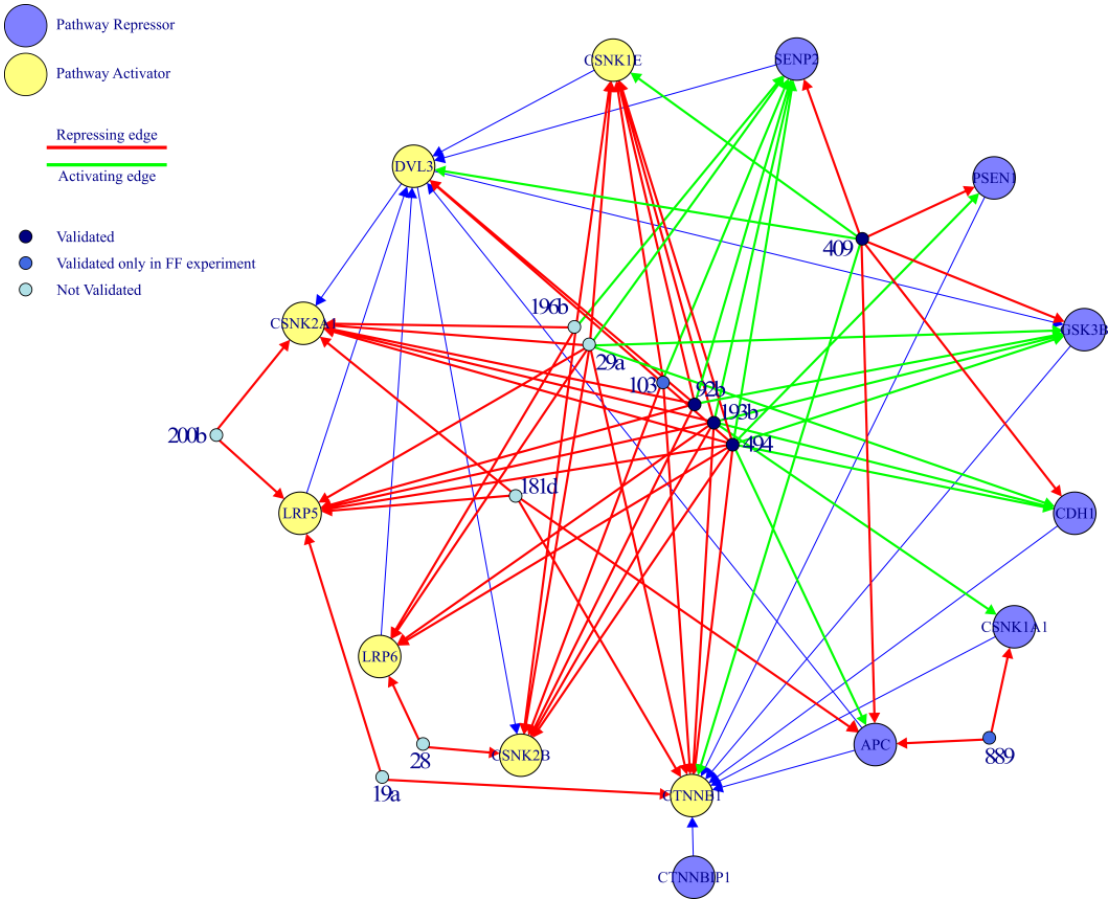


Figure 3.14. Network of significant miRNA-target interactions within the WNT/β-catenin signaling pathway. The network represents all the statistically significant interactions for those miRNAs which were tested for their ability to repress the reporter gene under β-catenin control. The legend on the upper left describes the colours for the edges and nodes shown in the network. The edges are colour coded based on the putative effect of the miRNA-target interaction that they are exerting, according to the permuted data table shown in figure 3.9.

3.7. Functional role of miRNAs downregulating the WNT/β-cat pathway

Of the 12 miRNA tested, miR-193b, miR-409, miR-494, and miR-92b were able to repress the expression of a WNT reporter gene, opening the question to whether they are able to affect also WNT-dependent phenotypes. One of the main phenotypical consequences of the activation of WNT/β-catenin pathway is proliferation, via induction of genes such as *MYC*, *JUN*, and cyclins. Moreover, in cancer the pathway has been frequently associated with EMT characteristics and maintenance of stem-like abilities. Therefore, phenotypes related to the pathway activation were hypothesized to be regulated by the four miRNAs identified via the reporter gene assay.

3.7.1. Proliferation

Upon miRNA O.E. and pathway stimulation with recombinant WNT3a, three of the four selected miRNAs are able to significantly downregulate proliferation. Treatment of cells with the chemical inhibitor at 10 µM shows to which extent the phenotype is dependent on WNT pathway-induced transcription.

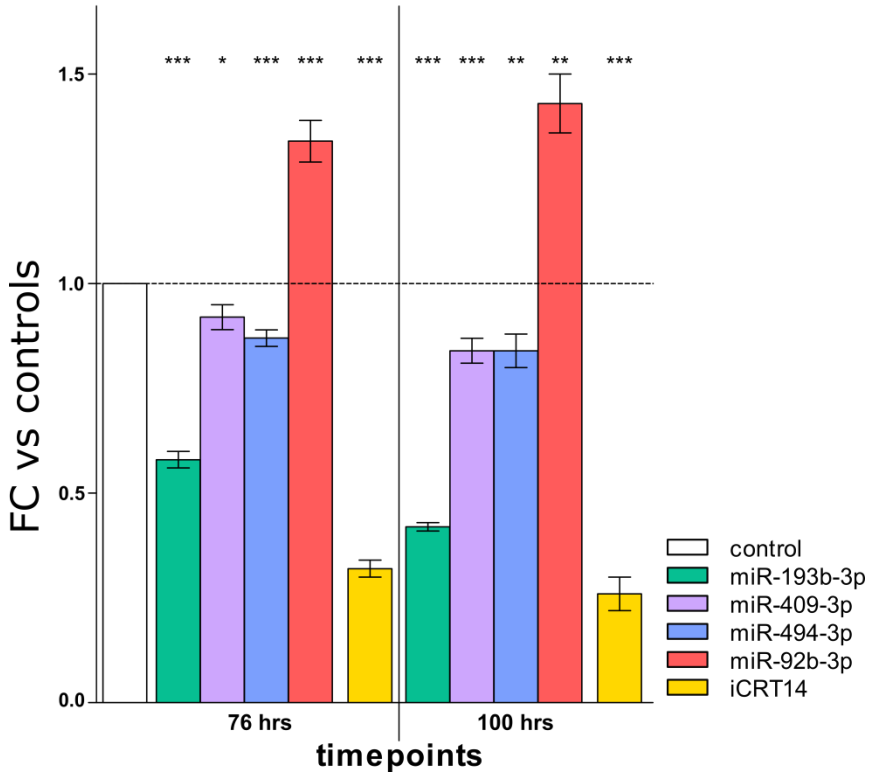


Figure 3.15. miRNA-mediated effects on proliferation upon WNT pathway induction. Proliferation was measured counting nuclei upon miRNA OE, or iCRT14 treatment by microscopy. The assay was performed at 76 hours and 100 hrs post-transfection. The y-axis represents the fold change in nuclei count compared to the control (mimic negative control for miRNAs/DMSO for iCRT14). Statistical significance against the control condition was calculated with a two-tailed student's t-test. P-values are represented with asterisks on top of the corresponding condition (* ≤ 0.05, ** ≤ 0.01, *** ≤ 0.001).

While iCRT14, at the selected concentration, is able to reduce proliferation by roughly 70%, the best-performing miRNA, miR-193b-3p, represses proliferation up to 60% at 100 hrs post-transfection (*fig. 3.15*). The presence of a major phenotype which is independent from the regulation of WNT pathway seems to be the case also for miR-92b which upregulates proliferation, regardless of its repressive role on the WNT pathway.

3.7.2. Stemness

To investigate the stem-like proprieties of MDA-MB-231 cells upon WNT pathway modulation, we proceeded to assay the expression of surface markers, as well as the ability of cells to overcome *anoikis* and grow in absence of attachment as spheres in miRNA gain-of-function assays. The hypothesis was that miRNAs downregulating the WNT pathway would be able to repress these stem-associated phenotypes.

3.7.3. Stem cell markers

Four days after miRNA OE, cells were detached to proceed with FACS analysis of CD24 and CD44 surface markers. This experimental procedure was optimised in order to deliver the least damage possible to the surface markers. Detachment procedures tested were: trypsinization with solutions at 0.25% and 0.05%, scraping with 2%FBS in PBS solution, treatment with TrypLE^{TM} Express (Gibco®), or treatment with Cell Dissociation Buffer enzyme-free (Gibco®). The latter provided the same results as scraping with 2%FBS in PBS, while requiring shorter handling time for the samples. Therefore, it was used as method to detach cells prior to surface markers quantification.

Quantification of the percentages of cells positively or negatively stained for the markers shows that miR-193b, as well as iCRT14, are able to significantly modulate the expression of CD24 and CD44. Indeed, both treatments significantly reduced the percentage of CD44+/CD24- compared to the respective control (fig. *3.16.A*). This result indicates that repression of WNT pathway is associated with the loss of stem-related surface markers.

However, analysing populations stained for a single marker, it becomes apparent how the miRNA tends to affect mostly CD44 expression, increasing the percentage of the CD44 negative cells. On the contrary, the chemical inhibitor rather increases the percentage of CD24+ cells. (*fig. 3.16.B*).

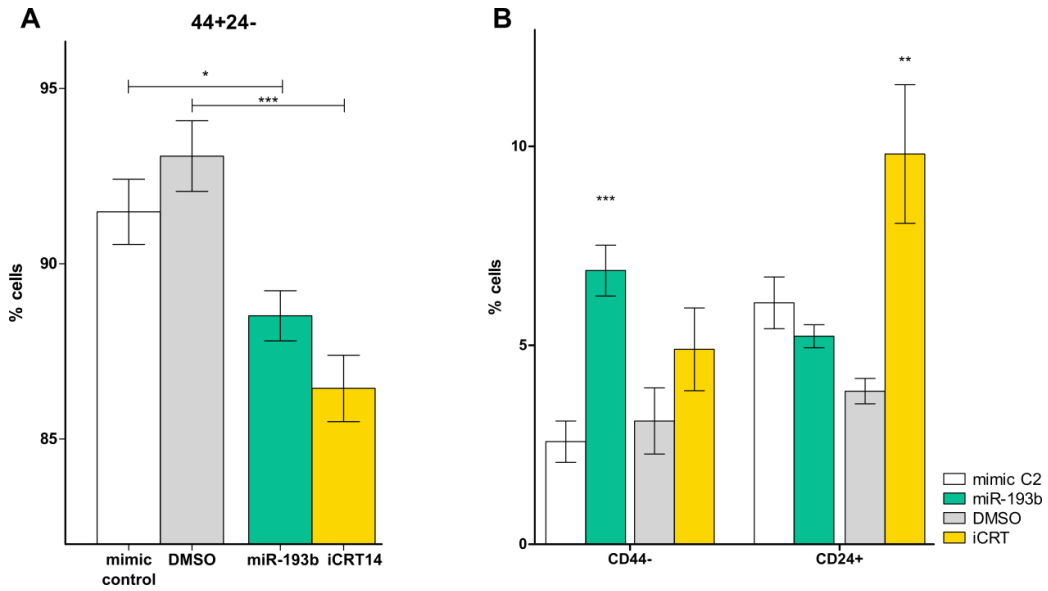


Figure 3.16. miRNA-mediated control of stem markers CD44 and CD24. A The percentages of CD44+/CD24- cells are shown for control conditions as well as treated. **B**. Deconvolution of the data to assess the effect of the treatments on the individual markers. Percentages of CD44- and CD24+ cells are shown separately for all the conditions. Statistical significance against the respective control conditions was calculated with a two-tailed student's t-test. P-values are represented with asterisks on top of the corresponding condition (* ≤ 0.05, ** ≤ 0.01, *** ≤ 0.001).

In conclusion, the results show that among the microRNAs which downregulate the expression of a WNT-dependent reporter gene, three are affecting proliferation of MDA-MB-231 cells, and one, miR-193b-3p is able to modulate the expression of CD44, a surface marker associated with stem-like properties. However, none of the miRNAs was able to regulate the formation of mammospheres. As well, iCRT14 was not affecting this phenotype. This possibly hints at the fact that while miR-193b and iCRT14 alone are able to modulate the surface markers, this is not enough to significantly reduce the stem-like content of the MDA-MB-231 cell population and therefore it does not affect the growth of spheres in ultra-low attachment conditions.

3.8. Validation of direct control of gene expression by microRNAs

Of the miRNAs tested for their effect on the WNT/β-catenin signalling pathway, three were shown to be strong regulators of proliferation in the context of pathway activation. The role of these miRNAs on the pathway was inferred by the target protein regulation. However, to fully understand the molecular mechanism

underlying the phenotype, it is necessary to investigate the direct interactions miRNA:mRNA.

This can tested *via* a reporter gene assay where the 3'UTR of the putative target genes is cloned downstream of a luciferase gene. In this manner, the luciferase activity will be a proxy to understand if the candidate miRNAs are binding to the 3'UTR of the putative target genes.
TargetScan v7 database was queried for putative MBS on the 3'UTR of the genes belonging to WNT pathway. Nine out of 14 genes contained predicted MBS in their 3'UTRs, therefore these sequences were cloned downstream a renilla luciferase reporter gene, either full, or in portions when longer than 1.5 kB (*fig. 3.17*). miRNAs and the vector harbouring the reporter genes were transiently transfected in MDA-MB-231 and 48 hrs later luciferase activities were assayed (*fig. 3.18*).

The results shown in *figure 3.18* clearly show that miR-193b is able to significantly downregulate the expression of the reporter gene in all cases, while miR-494 represses only of LRP6- and PSEN1-3'UTR harbouring reporter genes. Intriguingly, despite being predicted to have multiple MBS on various genes, miR-409 did not significantly affect any of the cloned 3'UTRs. This is highly consistent with the fact that in our RPPA data we found that the majority of the targets were upregulated upon miR-409 OE, hinting at the fact that this miRNA might affect the pathway indirectly through targets not yet identified. These results confirm that the interactions predicted for this miRNA are not functional in our cell system, as it is not able to repress reporter gene expression. As well, they tell a word of caution about using MBS prediction as definitive tool to estimate computationally the biological role of a miRNA: they always need to be proved within the context in use. Additionally, the presence of a functional 6mer which was not predicted (miR-193b on CTNNB1), indicates that in certain cases the available prediction tools are not enough to identify actual interactions.

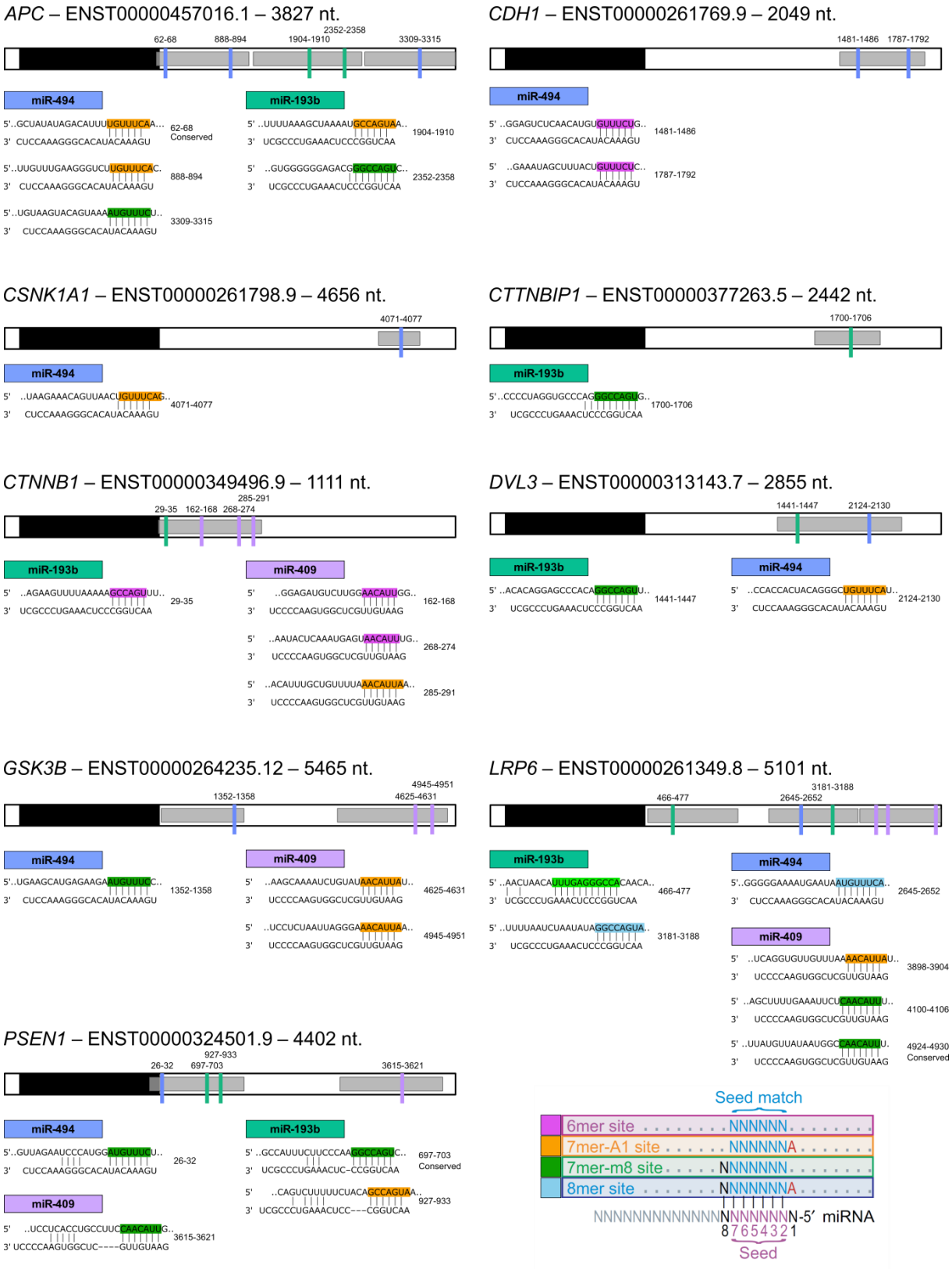


Figure 3.17. Schematic representation of genes of interest and cloned sites. Each gene of interest is shown, with the transcript ID used to retrieve the 3'UTR sequences as well as their length. Coding region is in black and the UTRs are in white, while the grey areas correspond to the cloned portions of the UTRs. The vertical bars represent the predicted miRNA binding sites, colour coded by miRNA. The predicted binding sites are further represented as nucleotide sequences with a colour coding that identifies the type

of binding site. In three instances, the binding sites are rated as conserved by TS v7. The first miR-193b site in LRP6 is a non-canonical binding site termed “centered site” (coded in light green). In the bottom right of the figure, the miRNA binding sites are represented as (Baek, Villen et al. 2008). The colour coding of the binding site type is used to mark the miRNA binding site in all the nucleotide alignment shown.

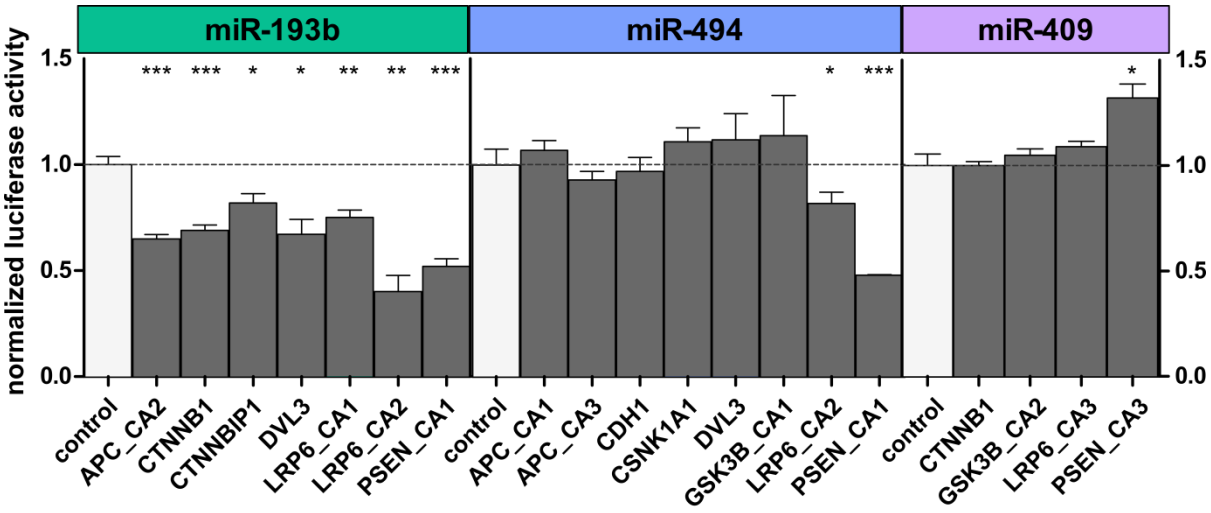


Figure 3.18. miRNA targeting 3’UTRs of interest. Normalized luciferase activity is shown for all the 3’UTRs that were predicted to be targeted. Statistical significance against the respective control conditions was calculated with a two-tailed student’s t-test. P-values are represented with asterisks on top of the corresponding condition (* ≤ 0.05, ** ≤ 0.01, *** ≤ 0.001).

To elucidate whether miRNA:mRNA interaction is directed specifically toward the predicted binding sites, the three nucleotides corresponding to the seed match positions 3 to 5 in each of the significantly downregulated 3’UTRs constructs were mutated. The mutation should disrupt the matching of the miRNA to its binding sites and rescue the expression of the reporter gene (*fig. 3.19*)

For miR-193b, mutations of the 3’UTR binding sites rescued the expression of the luciferase only for three genes out of seven interactions. In the case of PSEN1, it seems that the first binding site is the only functional one. However, when both were mutated the rescue is stronger, hinting at a possible cooperativity between the two sites. This was not the case for APC, where it seems that not the single mutations, nor the double, are able to rescue the expression of the reporter gene. Regarding the mutations of miR-494 predicted binding sites, there is a trend in re-establishing the expression of the reporter gene. However, these changes are not statistically significant.

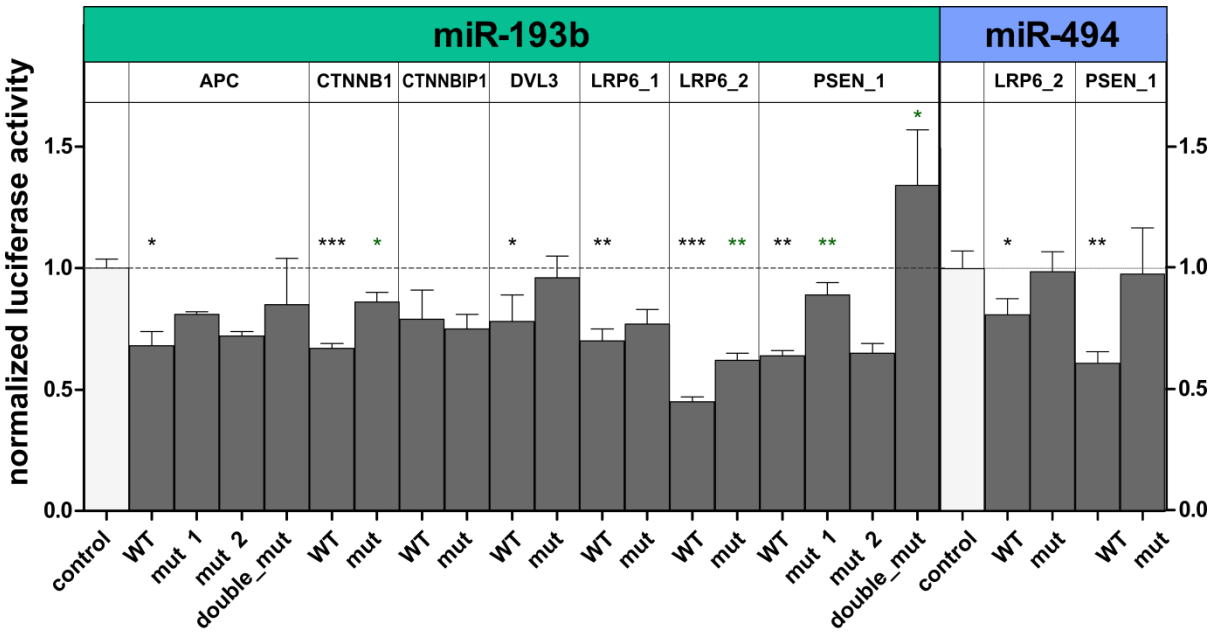


Figure 3.19. Validation of the reporter gene repression as direct at the 3'UTR. Normalized luciferase activity is shown for all the 3'UTRs that were predicted to be targeted and that were repressed in a WT condition. Statistical significance for the WT constructs was calculated against the control condition (black asterisks). Statistical significance for the mutated constructs was instead calculated against the WT counterpart (green asterisks). P-values are represented with asterisks on top of the corresponding condition (* ≤ 0.05, ** ≤ 0.01, *** ≤ 0.001).

The inability to rescue the expression of the reporter gene can be explained in different ways. First it could be that mutating the three nucleotides corresponding to 3 to 5 seed position is not enough to disrupt the miRNA-target interaction. Additionally, there could be functional but non-canonical miRNA binding sites in the 3'UTR of interest which have not been identified. Nevertheless, whether the interactions are direct on the 3'UTR or not, both miR-193b and miR-494 were shown to be highly integrated in the WNT/β-catenin signalling network.

In conclusion, in collaboration with Astrid, we described interactions between 800 microRNAs and 62 targets belonging to three pathways relevant for breast cancer metastasis. These interactions were evaluated at the protein level, therefore increasing the confidence in their biological function. By means of a novel network approach which takes into consideration the context of each miRNA-target interaction by integrating the information of the role of each protein on the pathway to which they are associated, Astrid and I were able to identify

miRNAs putatively affecting c-Met, Integrin signalling, and WNT β catenin pathways.

Considering the pathological relevance of WNT pathway and the possible clinical exploitation of results of miRNA that inhibit it, the focus of the project shifted to this pathway for the experimental validation. Four miRNAs were characterised as strong regulators of the pathway, significantly affecting β-catenin-dependent transcription. Of these, three were able to repress proliferation. One of these was also able to affect stem cell surface signatures. However, none of them was able to affect mammosphere formation in low attachment conditions. In this assay, the chemical inhibitor iCRT14 was also ineffective. This put into question whether the phenotype is indeed dependent on the WNT pathway in the cell system in use.

Of the three miRNAs phenotypically affecting proliferation, analysis of the direct binding at the 3'UTR of target genes was performed. Results at the reporter gene level recapitulated the extent (in FC) the repressions discovered by RPPA. However this did not happen for miR-409, hinting at the fact that the upregulation of proteins measured is indeed indirect and not mediated by interaction with the 3'UTRs of corresponding predicted genes. miR-409 inability to repress the reporter in the assay also point out at how predicted miRNA binding sites might actually not be functional upon testing.

4. Discussion

The primary aim of the project was to identify novel miRNAs regulating signalling pathways relevant for the metastasization and stem-like abilities of triple negative breast cancer. The project began by selecting pathways of interest for the type of cancer and the cell line model used. Subsequently, target genes belonging to c-Met, Integrin, and WNT/β-catenin signalling pathways were chosen, and their expression analysed in the cell model. The effect of miRNAs on the corresponding proteins was analysed by means of RPPA. Selecting a very stringent statistical threshold, miRNAs were scored for their putative effect on the selected pathways with a novel network approach. In this analysis, the fundamental information regarding the role of the targets quantified was included to contextualise the effect of the miRNAs. The score was then tested against a random population upon resampling miRNA-target interactions, identifying multiple miRNAs that would regulate each pathway.

The WNT/β-catenin pathway was chosen to prove experimentally that the identified miRNAs indeed regulate its activity. Exploiting a reporter gene assay, four miRNAs were validated as as strong repressors of the pathway. Of these, three were able to reduce cell proliferation in a context of pathway overactivation. Analysing the direct miRNA-target interaction by means of a 3'UTR reporter assay, the molecular interactions were validated, proving that the identified miRNAs are remarkably highly integrated in the WNT/β-catenin pathway.

4.1. C-Met, Integrin, and WNT/β-catenin pathway choice

The choice of the pathways of interest started by assessing what type of triple negative breast cancer is modelled by MDA-MB-231 cells, since this was the system in which the miRNA library was overexpressed. To reach biologically relevant conclusions, the focus of the project was on pathways enriched in the subtype represented by the cell line in use. The classification of TNBC patients that was developed by Lehmann and colleagues was used as a basis for the pathway choice. In a seminal study from 2011, they classified TNBC patients in sub-subtypes on the basis of unsupervised clustering of their gene expression profiles. Then, gene expression data of TNBC cell line models was matched with

the profiles identified in patients. MDA-MB-231 cells fell into the category defined by Lehmann and colleagues as Mesenchymal Stem Like (MSL) (Lehmann et al, 2011, JCI). The six subtypes have been refined into four groups (BL1, BL2, M, and LAR) upon assessment of the stromal component of the tissues analysed. In the same study, TNBC cell line models have been reclassified, with MDA-MB-231 cells established as a model for the BL2 subtype. This subtype exhibits the lowest pathological complete response rates among all subtypes, making it an important biological entity to further investigate (Lehmann, Jovanovic et al. 2016).

Among the pathways enriched in the MSL subtype in the original paper, WNT/β-catenin, integrin mediated adhesion, and focal adhesions attracted my attention for their well-known association with metastatic phenotype and stem-like characteristics. Additionally, the increased evidence of the role of c-Met on basal breast cancers as well as its possible druggability made it an interesting subject of investigation. Notably, in various pathological and physiological conditions, the three pathways were found highly interconnected. In hepatocytes, HGF was able to induce Wnt-independent nuclear translocation of β-catenin (Monga, Mars et al. 2002). In colon cancer, WNT and MET pathways have been shown to cooperate to maintain the niche where cancer stem cells are residing (Vermeulen, Melo et al. 2010). In the same context, it was shown that *MET* is under the transcriptional control of TCF4 and therefore can be modulated by WNT signalling pathway (Boon, van der Neut et al. 2002). One of the canonical downstream arm of c-Met signalling is the CDC24/RAC, which in turn signals toward cadherins and integrins. Additionally, in breast cancer metastatic processes and in glioblastoma invasion, it was shown that c-Met dimerizes with integrin β1, displacing it from the complex with integrin α5 (Jahangiri, Nguyen et al. 2017). These findings, albeit mostly in tissues different than breast, might indicate a strong connectivity with functional relevance for the metastatic cascade and encourage further investigations.

4.2. RPPA identifies many mild miRNA-mediated protein changes

Many studies now have addressed the impact of microRNAs on protein levels both via mass spectrometry- (MS) or protein-based assays, however, this is one

of few in which the focus of the study are specific pathways of interest chosen *a priori*.

While MS techniques coupled to SILAC or pSILAC are unbiased and able to address general questions related to miRNA biology, they have a lower throughput when it comes to amount of conditions that can be simultaneously investigated (Baek, Villen et al. 2008, Selbach, Schwanhausser et al. 2008). Indeed, the majority of MS-based approaches have been used to define the basic biological rules that define direct miRNA-dependent effects. While confirming that the predominant way a miRNA identifies and regulates its targets is based on the seed sequence, as well as claiming that the MBS with higher complementarity are more efficiently exerting this role, no analysis regarding the networks of protein affected by the miRNA was performed. MS-based approaches have also shown that the average effect of miRNA gain-of-function on targets which contain at least a seed sequence binding site in their 3'UTRs is mild, ranging from around -0.3 to -0.5 log2FC (Baek, Villen et al. 2008, Selbach, Schwanhausser et al. 2008). In the context of ER+ breast cancer, mass spectrometry coupled with iTRAQ (isobaric tag for relative and absolute quantification) was exploited to identify in an unbiased fashion the targets of miR-193b at the protein level (Leivonen, Rokka et al. 2011). Analysing also gene expression upon miR-193b OE, they also showed that only a minority (13%) of the proteins identified had a matched repression of its mRNA molecule (Leivonen, Rokka et al. 2011).

In comparison to MS-based techniques, protein array-based technologies allow researchers to understand the impact on specific targets of many more miRNAs, thanks to the high sample-content and sensitivity. This approach has been now adopted with success numerous times to identify protein expression regulated by miRNAs in a "genome-wide" manner, by overexpressing libraries representing the known miRNome (Leivonen, Makela et al. 2009, Uhlmann, Mannsperger et al. 2012, Leivonen, Sahlberg et al. 2014, Moss, Luo et al. 2015). In most of the cases, the regulatory activity of a miRNA on a pathway was taken into account by probing few genes that should suffice to describe activation of the respective pathway, such as HER2, Ki67, cleaved PARP, phospho-AKT, and phospho-ERK1/2 as proxies for the entire HER2 pathway (Leivonen, Sahlberg et al. 2014). Few studies have taken into consideration larger numbers of targets, the most

numerous being a screening of 879 miRNAs on 127 targets in 3 cell lines (Moss, Luo et al. 2015). Based on this data, Moss and colleagues defined miRNAs to be functionally integrated in a pathway when they would modulate the expression of at least 50% of the probed protein from that pathway (Moss, Luo et al. 2015). However, except for the identification of miRNAs specifically affecting the EGFR-driven cell cycle (Uhlmann, Mannsperger et al. 2012), none of the currently available datasets specifically focuses *a priori* on a pathway as a whole. With the aim to identify as many miRNA as possible that could target specific pathways that are of clear clinical relevance, we proceeded by exploiting RPPA. This gave two main advantages: very high sample content, coupled with analysis at the proteomic level.

A main limitation of high-throughput screenings such as the one undertook in this PhD project and the ones of Uhlmann and colleagues, and Moss and colleagues, is the use of a library of miRNAs. Generally, libraries are constructed on available databases. Indeed, the library used in both this project and the EGFR pathway analysis (Uhlmann, Mannsperger et al. 2012) was developed on the miRBase database in version 10, corresponding to the release of 2007. This library contained 810 annotated miRNAs. Of these, 10 were considered “dead” (i.e. characterized later on not as miRNAs but as fragments of other longer RNA molecules) in the latest version of miRBase (v.21). While these were excluded from any analysis within this project, it cannot be ruled out that other of the annotated miRNAs will be, in the future, characterized as “dead”. Analysis of the library used for the presence of miRNAs identified as “High confidence” (Kozomara and Griffiths-Jones 2014) showed that more than 50% (448) are high confidence. Additionally, in the project, miRNAs not expressed in breast cancer tissues were excluded from validation, possibly preventing the inclusion of such “dead” molecules in downstream analyses.

To identify significant miRNA-target interactions, we did not utilize the z-score. The choice of the limma test for this purpose stems from the experimental setup. Indeed, the setup is roughly limited to 650 samples per slide printed. Due to the presence of control samples, and two biological replicates for each miRNA transfection, the total of samples required to be divided into four subgroups

(termed arrays) to be probed. In turn, the data produced in such manner would have required the z-score to be calculated within each subgroup, to be a statistically appropriated method.
However, distribution of the signal intensities for the same target varied between the different subgroups. While not tested, this phenomenon was hypothesized to be dependent on the non-random distribution of samples that would cause many miRNAs belonging to the same family to be spotted within the same array. Indeed, miRNAs of the same family (i.e. with the same seed sequence) should in principle share some targets. Therefore, they could affect *en masse* the signal of these targets and shift the distribution of the data for the whole array. Since the z-score depends on the average and standard deviation of the distribution to define the individual sample value, in the context of this screening, its use might mask true interactions. For this reason, Astrid Wachter chose the limma test to calculate the statistical significance and the log2 FC against the miRNA mimic negative controls, possibly preventing the bias that could otherwise have stemmed from the use of z-score. Additionally, the limma test calculates the values taking into account the two biological replicates which were spotted, making the results statistically more robust.

By means of the limma test, a high amount of miRNA-target interactions (5'435 out of 49'600, ~11%) were identified with strong statistical significance (q-value ≤ 0.001). Of these, 70% are downregulations (3804) and 30% upregulations (1631). Target-wise analysis revealed an average of 61 miRNAs repressing and 26 upregulating, with the extremes being EGFR repressed by 190 miRNA and CDKN1B upregulated by 224. The distribution of miRNAs regulating single target didn't seem dependent on whether the protein has an activating or repressive function on the pathway. It is tempting to speculate that proteins whose mRNA harbours a longer 3'UTR are more likely to be repressed by miRNAs. However, the correlation between length of 3'UTR (from TS v7) and amount of regulating miRNAs in the data gathered did not support this hypothesis. Indeed, there was no positive correlation between the miRNAs significantly downregulating a target and the length of the relative 3'UTRs. As for all analyses that exploit publicly available datasets, it is not possible to rule out whether, upon investigation of

3'UTR sizes specifically in MDA-MB-231 cells, results could instead support the hypothesis.

Similar to previously reported data, the average negative effect on the expression of target proteins was -0.52 log2FC (corresponding to roughly 30% in protein expression). Indirect positive effects exhibited an average of 0.45 log2FC. To understand whether the identified regulations are mediated by predicted binding sites within the 3'UTRs of target genes, an enrichment test was performed, interrogating separately the downregulations, as well as the upregulations, which would serve as in internal control. The test returned 45% of the targets to be enriched in predictions of conserved miRNA binding sites in their 3'UTR, as shown in table 3.1. Non conserved (NC) binding sites presented a lower percentage of protein with enrichment, consistently with previous reports (Baek, Villen et al. 2008, Uhlmann, Mannsperger et al. 2012), probably indicating how predicted NC sites are less likely to be functional. MicroCosm Targets (MCT) returned a much lower percentage of significant enrichments. Considering that MCT allows in its prediction algorithm seed mismatches as well as single GU wobble pairs, the results suggest that its inability to perform as well in the dataset indicates how these predicted interactions might be less functional in downregulating protein expression. In comparison to the results of the Uhlmann screening (Uhlmann, Mannsperger et al. 2012), a lower percentage of targets presented significant enrichment. Indeed, in the EGFR pathway screening, 18 out of 25 (72%) targets passed the significance threshold, computed with a Kolmogorov-Smirnov (KS) test. This test was not used in this screen due to the nature of the limma analysis method that we exploited to measure miRNA-dependent protein expression changes. Being the KS a threshold-free, rank-based test, Astrid Wachter deemed inappropriate to use it on q-values which were generated with multiple limma tests applied on the 34 rounds of transfections separately. While this difference in analysis could be a reason for the lower percentage of targets presenting enrichment, upon closer inspection we can recapitulate 15 of the 17 significant enrichments identified in the Uhlmann screen (MIG-6/ERRFI1 was not probed). When comparing the results, we should not forget that the limma method returned as significant also very small protein modulations (e.g. miR-526b-5p repressing CSNK2B -0.08 log2FC). Such small

variations were not considered significant when their z-scores were statistically tested in the EGFR-pathway screening.

A small percentage (9.7%) of targets presented significant p-values for enrichment also for the upregulations. This could be indicative of the fact that the prediction databases contain a number of false-positives that are not recapitulated in experimental conditions. In support to this hypothesis, albeit in a small scale which cannot explain effects on the whole screening, miR-409 might serve as example. Indeed, it presented many predicted interactions with 3'UTRs of genes of the WNT/β-catenin signalling pathway. However, luciferase assays to evaluate their functionality did not validate any of the interactions.
It remains an open biological question whether miRNAs target possible binding sites also in the coding sequence of genes or 5'UTRs, and whether they are functional. As well, it is important to keep in mind that the type of enrichment test performed is based on 3'UTR sequences retrieved in publicly available datasets (TS uses UCSC genes, MCT is based on the ENSEMBL gene repository). However, intrinsic variabilities of 3'UTRs (SNPs, alternative polyA sites) could be at play in this type of post-transcriptional regulation, affecting the results of enrichments, as well as the biological effect of the miRNAs of interest (Fu, Sun et al. 2011).

To conclude, the patterns of protein expression changes that we identified upon miRNAs OE are concordant to literature and to the known biological functions of miRNAs on their target. Additionally, we acknowledge that we are taking into account for downstream analysis, not only direct miRNA-target interactions, but indirect upregulations and downregulations as well. This fact is consistent with the purpose of the project to describe the role of miRNAs on complete pathways.

4.3. Computation of scores to predict the role of miRNAs on the whole pathway.

A main concept that was at the basis of the network analysis developed in collaboration with Astrid Wachter was that the context of a particular miRNA-target interaction is biologically more relevant than the extent of the regulation

caused by the same specific miRNA. Therefore, the analysis addressed the possibility that a miRNA affecting multiple targets, yet mildly, might still strongly regulate a pathway when analysed as a large system. Given that pathways are constituted of proteins that themselves have a functional role within the pathway itself, it seemed natural to consider this factor in the downstream analysis. For the purpose of modelling, for each protein a single function (either positive or negative) was taken into account. Hence, the role of the miRNAs on each target was permuted into a putative effect on the pathway, by substituting the log2FC with a +1/-1 value which would be dependent on the role of the target as well as the direction of the regulation exerted by the specific miRNA.

At this step of the analysis the extent of the regulation (log2FC) was arbitrarily disregarded and only strongly statistically significant interactions (q-val≤0.001) were taken into account. This choice was proposed by me and discussed together with Astrid Wachter to reach a conclusion we both would agree on. Two main reasons drove the decision:

a) To actually infer quantitatively the role of a miRNA on the pathway, it would have been required to first quantitatively evaluate also the role of the protein on it.
b) being the hypothesis that multiple small regulations are effective than few strong ones, the extent of the regulation of individual miRNA-target pairs was deemed not relevant.

This decision seems to have been appropriate, on the basis of experimental validation of miRNAs which affect the WNT/β-catenin signalling pathway. Inspecting miRNA-target interactions in RPPA data, we can appreciate that miR-19a-3p was strongly downregulating CTNNB1 and LRP5 (-0.93 and -0.42 log2FC respectively). This repression did not result in the ability of the miRNA to regulate the reporter gene response. On the contrary, miR-193b-3p influenced the expression of multiple targets in a mild manner (n = 10, -0.72 to -0.19, ave. -0.44) and was able to repress the reporter gene, as well as phenotypes downstream the WNT/β-catenin signalling pathway. Additionally supporting the decision, we can interrogate the PC scores calculated with increasing thresholds for the log2FC. Indeed, when we had taken into account a cutoff of $|0.5|$ log2FC, miR-

409-3p score indicated a role for this miRNA as positive regulator of the pathway, whilst the reporter gene assay confirmed it as a repressor. To take into consideration the log2FCs a possible solution could have been to multiply their values for -1 or +1, in case the target protein was a repressor or an activator, respectively. In this scenario, the PC score would have probably not represented once more the effect on the pathway. Indeed miR-181d would have been the strongest repressor with a value of -1.57, and miR-409 the weakest with -0.15. Instead, the first was unable to affect the reporter gene and the latter repressed the reporter gene consistently.

Therefore, the choice to disregard the log2FC information to compute the score was deemed appropriate. As well, it appeared that the randomization test was a feasible strategy to identify miRNAs modulating the pathway defined *a priori*.

Still, some miRNAs identified with the PC score randomization method did not affect the reporter gene assay, suggesting possibly that additional filtering steps to refine the list of candidate miRNAs are required. Reasonable strategies could include i.e. the exclusion of miRNAs which target less than 50% targets probed within the pathway, as applied by TJ Moss and colleagues (Moss, Luo et al. 2015). Higher pathway coverage via RPPA could also have provided a broader picture. Clearly, the availability of antibodies with enough specificity and sensitivity is a fundamental limit to the possibility of analysing consistently full pathways. Possibly, refined modelling strategies could have includes in the network analysis unknown variables corresponding to the unprobed (but expressed) proteins.

4.4. miRNAs modulate WNT/β-catenin signalling pathway activity.

The role of the miRNome on the WNT/β-catenin signalling pathway had been previously studied by Anton and colleagues upon transfecting single miRNA in HEK293T cells together with a TOP/Flash reporter (Anton, Chatterjee et al. 2011). Upon identification of the 12 miRNAs candidates from our screening, the results were compared with the ones identified by Anton and colleagues. miR-193b, miR-409, and miR-28 were all described as mild repressors of the reporter

gene. On the other hand, miR-181d was characterized as activator. Whilst in MDA-MB-231 miR-193b and miR-409 downregulate the signalling pathway as well, neither miR-28, nor miR-181d were able to regulate the reporter gene.
Some experimental differences could partially explain the results: e.g. Anton and colleagues transfected miRNA mimics at 40nM, possibly rendering the screening more prone to off-target and indirect effects when compared to the 25nM concentration used in the RPPA screening. However, an important message lies under these divergent results: within a cell system, the presence or absence of specific transcripts, as well as their abundance can deeply affect the role of a miRNA. Therefore, while supporting some of our findings, the screening from Anton and colleagues raises the issue that the effect of miRNAs should be considered in the cell type or tissue of interest.

4.5. miRNAs regulate phenotypes dependent on the WNT/β-catenin signalling pathway.

While validating the role of miRNAs on phenotypes that are dependent on the WNT/β-catenin pathway, the chemical inhibitor iCRT14, specifically designed to prevent CTNNB1 interaction with TCF7L2, thus blocking β-catenin dependent transcription, was exploited as positive control. First, iCRT14 was titrated using the reporter gene assay to define which concentration would resemble the most the effect seen with miRNA OE. This concentration (10 uM) was then used in all downstream assays.
The four miRNAs which were able to regulate the expression of the reporter gene in the dual reporter assay were tested for their ability to modulate proliferation in a cell context in which the WNT pathway is overstimulated with recombinant WNT3a. miR-193b repressed proliferation by 60%, almost as much as iCRT14 (75%). miR-409 and -494 repressed proliferation only by roughly 15%. The effect mediated by these miRNAs has been previously characterized in multiple tumour types, including triple negative breast cancer (for miR-193b and miR-494). Of note, miR-494 was described as repressor of WNT signalling in Pancreatic Ductal Adeno Carcinoma, where it was also able to repress proliferation and tumour growth *in vivo* (Li, Li et al. 2014). Additionally, the role of miR-494 on the WNT/β-catenin signalling pathway has been previously characterized in TNBC, however

it was explained as dependent on its ability to repress the expression of CXC4 (Song, Liu et al. 2015). Still, it has not been reported before that these miRNAs could repress cell proliferation under a condition of WNT/β-catenin pathway activation in triple negative breast cancer, such as in this case.

Of the four candidate miRNAs, miR-92b strongly induced proliferation (+45%). The role of this miRNA has been reported in various cellular contexts with controversial results. Indeed, it represses proliferation and induces apoptosis in glioma (Wang, Wang et al. 2013). As well, it is correlated with DFS in breast cancer, where its higher expression is associated to better prognosis (Smith, Baxter et al. 2015). On the other hand, it has been shown that in HCC miR-92b not only enhances proliferation, but it is able to induce β-catenin localization in the nucleus and activation of its target genes (Zhuang, Yang et al. 2016). It is therefore intriguing to see contradictory results in MDA-MB-231 cells, suggesting that in this case the repression of WNT pathway is uncoupled from the pro-proliferative effect.

Different effects of miR-193b and iCRT14 were also detected *via* FACS analysis of surface markers of MDA-MB-231 cells. Expression of CD44 and CD24 has been used to define a population of “cancer stem cells”, termed also “tumour initiating cells”. MDA-MB-231 cells displays at the basal level an extremely high CD44+/CD24- population (>95%) (figure 3.16 and (Ricardo, Vieira et al. 2011)).
Upon miRNA OE, the predominant effect detected was on the CD44+ population. In contrast, iCRT14 was able to increase the CD24+ population. Of note, CD44 is a direct transcriptional target of β-catenin. Therefore, it is tempting to speculate that in MDA-MB-231 miR-193b has a stronger ability than iCRT14 to regulate transcription dependent on β-catenin. However, to prove it, targeted experiments should be designed to compare transcriptional activity dependent on nuclear β-catenin. In alternative, it is possible that the phenotype induced by miR-193b is dependent not only by the repression of the WNT/β-catenin pathway activity.

In an attempt to define the effect of miR-193b, miR-409, and miR-494 on the stem-like phenotype of the cells, the ability of cells to survive *anoikis* and grow in clonal colonies was tested. First of all, no increase in sphere formation was

observed upon WNT3a stimulation of MDA-MB-231, as well as no reduction upon treatment with iCRT14. Additionally, none of the miRNA tested significantly affected the formation of spheres. Therefore, it seems that in this cell line the phenotype is not dependent on an additional stimulation of WNT pathway, possibly due to the already very high content of stem-like cells (Ricardo, Vieira et al. 2011). However, it was surprising not to see an effect caused by the WNT pathway inhibitor, suggesting perhaps that cells in which the pathway is inhibited cannot survive and are therefore negatively selected during the growth process. Experiments that allow for conditional tagging of cells (i.e. with a fluorescent protein under the control of WNT pathway) could perhaps support the hypothesis that only the cells retaining the pathway activity are the ones able survive low-attachment conditions.

Another method to assess the stem-like content of a population is the ALDEfluor assay, which identifies cells which express at high levels aldehyde dehydrogenase (ALDH), a feature characteristic of stem cells. The cell line model in use for this project, MDA-MB-231, was shown multiple times to exhibit a very low percentage (<2%) of of $ALDH^+$ cells (Charafe-Jauffret, Ginestier et al. 2009, Marcato, Dean et al. 2011, Ricardo, Vieira et al. 2011). This finding is highly in contrast with the percentage of stem-like population that is widely described for MDA-MB-231 when determined by surface markers (Ricardo). In the context of this project is therefore clear that the use of $ALDH^+$ cells as indicator of stemness might not be a feasible option. Indeed, being the hypothesis that the identified miRNAs reduce the stem capabilities of the cells they are transfected in, such assay in MDA-MB-231 would provide a dynamic range very ineffective. However, a recent publication dissected the role of miRNAs as bimodal switches of breast cancer stem cell fate by means of a miniaturized ALDEFLUOR assay in SUM159 cells (TNBC). In the screening, the authors focused on miR-600 as the top scoring miRNA which modulates WNT signalling via repression of SCD1. However, among the miRNAs identified in their screening, miR-193b was found as a very mild regulator of the phenotype (El Helou, Pinna et al. 2017), perhaps supporting its role as stem cell regulator in TNBC via WNT/β-catenin pathway modulation. Experiments to replicate the effects of miR-193b in additional TNBC

cell lines such as SUM159 should definitely be performed within this project to validate its role as suppressor of stem-like phenotypes.

4.6. The identified miRNAs regulate WNT/β-catenin directly and indirectly pathway.

In order to understand whether the interaction of a miRNA is direct, the golden standard experiment is to perform a reporter gene assay using the 3'UTR downstream of a reporter gene. This is ideally done maintaining the molecular context, i.e. by cloning the entire 3'UTR and not simply the predicted binding site. Indeed, it is now well understood that the position and the surrounding of the MBS can influence its functionality (Bartel 2009). Upon co-transfection of a miRNA mimic and the construct harbouring the reporter gene and the desired 3'UTR, the reporter gene expression should decrease in presence of a direct interaction. To prove that the interaction is strictly dependent on the binding site predicted, it is possible to mutate some of its nucleotides, in order to disrupt the interaction.

The three miRNAs that repress proliferation of MDA-MB-231 were analysed for their effect on the 3'UTRs of mRNAs encoding the targets that had been investigated via RPPA. The online prediction tool TS v7 was queried for MBS of the three miRNAs on the genes of the WNT signalling pathway. Additional 6-mer binding sites that are not predicted by the TS v7 were searched manually on the 3'UTRs (figure 3.15). Several of them were identified within the 3'UTRs of CTNNB1 and CDH1.

For miR-193b and miR-494, the reporter assays confirmed a strong molecular integration within the WNT/β-catenin signalling pathway. Indeed, CTNNB1 and LRP6 were identified as direct targets for both miRNAs. An interesting hypothesis to explore further is that the two miRNAs cooperate on the same 3'UTRs to repress in concert the expression of two of the pivotal players of the pathway. While the hypothesis was tested, the assay wasn't optimised yet to appropriately evaluate the effects. The 3'UTR reporter gene assay additionally defined

CTNNBIP1, APC, and PSEN1 as direct targets of miR-193b, indicating how this miRNA has indeed a very strong integration within the pathway.

With regards to the 6mers manually identified the binding sites on *CDH1* where not functional upon miR-494 OE, while the BS for miR-193b on CTNNB1 strongly repressed luciferase expression, proving their functionality. Despite the short complementarity between the BS and the miRNA, in this case it seems that the context is of outmost importance. Indeed, it is positioned very close to the end of the CDS, a location very well characterized for its efficacy (Bartel 2009). Therefore, manually searching 6-mers in 3'UTRs might identify functional sites, even when not predicted by sophisticated algorithms.

In the case of miR-409, the target prediction algorithms described binding sites that did not seem functional according to the results gathered with 3'UTR reporter assays (fig 3.16). In the cell context in use, it could be interesting to identify which direct targets of miR-409 are able to cause repression of the β-catenin reporter gene. Intriguingly, while being a repressor of the WNT/β-catenin signalling, miR-409 was described as pathway activator for integrin and c-Met within this screening.

With the interconnectivity between these pathways known, it could be extremely interesting to dissect further the reciprocal role and define whether the upregulation of the other two pathways is indeed related to the WNT signalling repression. A proxy for the pathways co-regulation by a single miRNA could be a significant PC score after the randomization. Indeed, the score values had previously been divided by the amount of proteins within each pathway in order to make them comparable. However, no statistical testing to prove that a miRNA is significantly co-regulating the three pathways concomitantly was performed, as it was deemed unfeasible upon discussion with the collaborating group in Göttingen.

These results open again the question on how reliable prediction algorithms are when the genome-wide role of a miRNA is only computationally extrapolated. The

absence of appropriate prediction of sites which are generally considered less functional (e.g. 6mers) could also influence the choice of relevant targets to investigate. On the other hand, due to their clear moderate functionality, to take into account minor predictions *a priori* could engulf the results with a myriad of false-positives. It is still a very open issue in miRNA biology how to appropriately predict miRNA binding sites and it will be interesting to see in the future the developments which will be put forward by the scientific community to build more reliable models and simplify the search for targets. However, what is still very clear is that experimental validation is fundamental to assess the functional integration of miRNAs in pathways of interest.

4.7. Is miR-193b involved in a feedback loop with β-catenin?

Of all miRNAs tested, miR-193b was causing the strongest effects on the WNT/β-catenin pathway, both in reporter gene and phenotypic assays. Additionally, many of the interactions with 3'UTRs of target genes were validated at the molecular level. To deepen our understanding of the mechanisms mediated by this miRNA, additional target predictions were checked on TS v7. Among the genes listed some stood out for their integration within the pathway: WNT receptors *FZD3* and *8*, *TCF7L2* (also known as TCF-4), and *DKK2*. Intriguingly, exploring the genomic locus of miR-193b-365a cluster for WNT responsive elements, one was found located between the two miRNAs. Therefore, the hypothesis that miR-193b-365a cluster could be under transcriptional control of the pathway was tested. Analysis of miRNAs expression upon pathway activation (recombinant WNT3a stimulation) or repression (iCRT14 treatment) did not show any detectable variation (data not shown). Therefore, it seemed that in the cell system in use the expression of miR-193b is not under WNT pathway control.

Still, according to a recent publication, in hepatobiliary cancers WNT induces an ultraconserved lncRNA (uc.158-) which is able to bind miR-193b, removing it from the available pool and therefore de-repressing its target (Carotenuto, Fassan et al. 2017). Due to the inability of WNT to modulate miR-193b expression itself, we sought to understand whether in TNBC the WNT pathway could regulate the expression of uc.158-, which in turn could affect the other miRNA targets. However, in MDA-MB-231 cells, uc.158- was undetectable by

qPCR (data not shown), neither in basal condition, nor upon WNT pathway activation. Therefore, it was concluded that the mechanism of miR-193b sequestration identified in hepatobiliary carcinoma is not established in the cell system in use. It remains open whether there are other mechanisms that allow the WNT pathway to modulate miR-193b expression or its downstream function, therefore creating a possible feedback loop between the two.

Bibliography

Ades, F., D. Zardavas, I. Bozovic-Spasojevic, L. Pugliano, D. Fumagalli, E. de Azambuja, G. Viale, C. Sotiriou and M. Piccart (2014). "Luminal B breast cancer: molecular characterization, clinical management, and future perspectives." J Clin Oncol **32**(25): 2794-2803.

Agarwal, V., G. W. Bell, J. W. Nam and D. P. Bartel (2015). "Predicting effective microRNA target sites in mammalian mRNAs." Elife **4**.

Ameres, S. L. and P. D. Zamore (2013). "Diversifying microRNA sequence and function." Nat Rev Mol Cell Biol **14**(8): 475-488.

Anton, R., S. S. Chatterjee, J. Simundza, P. Cowin and R. Dasgupta (2011). "A systematic screen for micro-RNAs regulating the canonical Wnt pathway." PLoS One **6**(10): e26257.

Baek, D., J. Villen, C. Shin, F. D. Camargo, S. P. Gygi and D. P. Bartel (2008). "The impact of microRNAs on protein output." Nature **455**(7209): 64-71.

Bartel, D. P. (2009). "MicroRNAs: target recognition and regulatory functions." Cell **136**(2): 215-233.

Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society Series B-Methodological **57**(1): 289-300.

Bilir, B., O. Kucuk and C. S. Moreno (2013). "Wnt signaling blockage inhibits cell proliferation and migration, and induces apoptosis in triple-negative breast cancer cells." J Transl Med **11**: 280.

Bleazard, T., J. A. Lamb and S. Griffiths-Jones (2015). "Bias in microRNA functional enrichment analysis." Bioinformatics **31**(10): 1592-1598.

Boon, E. M., R. van der Neut, M. van de Wetering, H. Clevers and S. T. Pals (2002). "Wnt signaling regulates expression of the receptor tyrosine kinase met in colorectal cancer." Cancer Res **62**(18): 5126-5128.

Bracken, C. P., X. Li, J. A. Wright, D. M. Lawrence, K. A. Pillman, M. Salmanidis, M. A. Anderson, B. K. Dredge, P. A. Gregory, A. Tsykin, C. Neilsen, D. W. Thomson, A. G. Bert, J. M. Leerberg, A. S. Yap, K. B. Jensen, Y. Khew-Goodall and G. J. Goodall (2014). "Genome-wide identification of miR-200 targets reveals a regulatory network controlling cell invasion." EMBO J **33**(18): 2040-2056.

Bracken, C. P., H. S. Scott and G. J. Goodall (2016). "A network-biology perspective of microRNA function and dysfunction in cancer." Nat Rev Genet **17**(12): 719-732.

Calin, G. A., C. Sevignani, C. D. Dumitru, T. Hyslop, E. Noch, S. Yendamuri, M. Shimizu, S. Rattan, F. Bullrich, M. Negrini and C. M. Croce (2004). "Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers." Proc Natl Acad Sci U S A **101**(9): 2999-3004.

Carotenuto, P., M. Fassan, R. Pandolfo, A. Lampis, C. Vicentini, L. Cascione, V. Paulus-Hock, L. Boulter, R. Guest, L. Quagliata, J. C. Hahne, R. Ridgway, T. Jamieson, D. Athineos, A. Veronese, R. Visone, C. Murgia, G. Ferrari, V. Guzzardo, T. R. J. Evans, M. MacLeod, G. J. Feng, T. Dale, M. Negrini, S. J. Forbes, L. Terracciano, A. Scarpa, T. Patel, N. Valeri, P. Workman, O. Sansom and C. Braconi (2017). "Wnt signalling modulates transcribed-ultraconserved regions in hepatobiliary cancers." Gut **66**(7): 1268-1277.

Charafe-Jauffret, E., C. Ginestier, F. Iovino, J. Wicinski, N. Cervera, P. Finetti, M. H. Hur, M. E. Diebel, F. Monville, J. Dutcher, M. Brown, P. Viens, L. Xerri, F. Bertucci, G. Stassi, G. Dontu, D. Birnbaum and M. S. Wicha (2009). "Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature." Cancer Res **69**(4): 1302-1313.

Chou, C. H., N. W. Chang, S. Shrestha, S. D. Hsu, Y. L. Lin, W. H. Lee, C. D. Yang, H. C. Hong, T. Y. Wei, S. J. Tu, T. R. Tsai, S. Y. Ho, T. Y. Jian, H. Y. Wu, P. R. Chen, N. C. Lin, H. T. Huang, T. L. Yang, C. Y. Pai, C. S. Tai, W. L. Chen, C. Y. Huang, C. C. Liu, S. L. Weng, K. W. Liao, W. L. Hsu and H. D. Huang (2016). "miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database." Nucleic Acids Res **44**(D1): D239-247.

Denkert, C., C. Liedtke, A. Tutt and G. von Minckwitz (2017). "Molecular alterations in triple-negative breast cancer-the road to new treatment strategies." Lancet **389**(10087): 2430-2442.

Dey, N., B. G. Barwick, C. S. Moreno, M. Ordanic-Kodani, Z. Chen, G. Oprea-Ilies, W. Tang, C. Catzavelos, K. F. Kerstann, G. W. Sledge, Jr., M. Abramovitz, M. Bouzyk, P. De and B. R. Leyland-Jones (2013). "Wnt signaling in triple negative breast cancer is associated with metastasis." BMC Cancer **13**: 537.

Dontu, G., W. M. Abdallah, J. M. Foley, K. W. Jackson, M. F. Clarke, M. J. Kawamura and M. S. Wicha (2003). "In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells." Genes Dev **17**(10): 1253-1270.

Ebert, M. S. and P. A. Sharp (2012). "Roles for microRNAs in conferring robustness to biological processes." Cell **149**(3): 515-524.

El Helou, R., G. Pinna, O. Cabaud, J. Wicinski, R. Bhajun, L. Guyon, C. Rioualen, P. Finetti, A. Gros, B. Mari, P. Barbry, F. Bertucci, G. Bidaut, A. Harel-Bellan, D. Birnbaum, E. Charafe-Jauffret and C. Ginestier (2017). "miR-600 Acts as a Bimodal Switch that Regulates Breast Cancer Stem Cell Fate through WNT Signaling." Cell Reports **18**(9): 2256-2268.

Enright, A. J., B. John, U. Gaul, T. Tuschl, C. Sander and D. S. Marks (2003). "MicroRNA targets in Drosophila." Genome Biol **5**(1): R1.

Foulkes, W. D., I. E. Smith and J. S. Reis-Filho (2010). "Triple-negative breast cancer." N Engl J Med **363**(20): 1938-1948.

Friedman, R. C., K. K. Farh, C. B. Burge and D. P. Bartel (2009). "Most mammalian mRNAs are conserved targets of microRNAs." Genome Res **19**(1): 92-105.

Fu, Y., Y. Sun, Y. Li, J. Li, X. Rao, C. Chen and A. Xu (2011). "Differential genome-wide profiling of tandem 3' UTRs among human breast cancer and normal cells by high-throughput sequencing." Genome Res **21**(5): 741-747.

Golan-Lavi, R., C. Giacomelli, G. Fuks, A. Zeisel, J. Sonntag, S. Sinha, W. Kostler, S. Wiemann, U. Korf, Y. Yarden and E. Domany (2017). "Coordinated Pulses of mRNA and of Protein Translation or Degradation Produce EGF-Induced Protein Bursts." Cell Rep **18**(13): 3129-3142.

Gonsalves, F. C., K. Klein, B. B. Carson, S. Katz, L. A. Ekas, S. Evans, R. Nagourney, T. Cardozo, A. M. Brown and R. DasGupta (2011). "An RNAi-based chemical genetic screen identifies three small-molecule inhibitors of the Wnt/wingless signaling pathway." Proc Natl Acad Sci U S A **108**(15): 5954-5963.

Gosline, S. J., A. M. Gurtan, C. K. JnBaptiste, A. Bosson, P. Milani, S. Dalin, B. J. Matthews, Y. S. Yap, P. A. Sharp and E. Fraenkel (2016). "Elucidating MicroRNA

Regulatory Networks Using Transcriptional, Post-transcriptional, and Histone Modification Measurements." Cell Rep **14**(2): 310-319.

Gu, Z. G., R. Eils and M. Schlesner (2016). "Complex heatmaps reveal patterns and correlations in multidimensional genomic data." Bioinformatics **32**(18): 2847-2849.

Guo, H., N. T. Ingolia, J. S. Weissman and D. P. Bartel (2010). "Mammalian microRNAs predominantly act to decrease target mRNA levels." Nature **466**(7308): 835-840.

Gusev, Y., T. D. Schmittgen, M. Lerner, R. Postier and D. Brackett (2007). "Computational analysis of biological functions and pathways collectively targeted by co-expressed microRNAs in cancer." BMC Bioinformatics **8 Suppl 7**: S16.

Ha, M. and V. N. Kim (2014). "Regulation of microRNA biogenesis." Nat Rev Mol Cell Biol **15**(8): 509-524.

Han, Y. C., J. A. Vidigal, P. Mu, E. Yao, I. Singh, A. J. Gonzalez, C. P. Concepcion, C. Bonetti, P. Ogrodowski, B. Carver, L. Selleri, D. Betel, C. Leslie and A. Ventura (2015). "An allelic series of miR-17 approximately 92-mutant mice uncovers functional specialization and cooperation among members of a microRNA polycistron." Nat Genet **47**(7): 766-775.

Huguet, E. L., J. A. McMahon, A. P. McMahon, R. Bicknell and A. L. Harris (1994). "Differential expression of human Wnt genes 2, 3, 4, and 7B in human breast cell lines and normal and disease states of human breast tissue." Cancer Res **54**(10): 2615-2621.

Inui, M., G. Martello and S. Piccolo (2010). "MicroRNA control of signal transduction." Nat Rev Mol Cell Biol **11**(4): 252-263.

Jahangiri, A., A. Nguyen, A. Chandra, M. K. Sidorov, G. Yagnik, J. Rick, S. W. Han, W. Chen, P. M. Flanigan, D. Schneidman-Duhovny, S. Mascharak, M. De Lay, B. Imber, C. C. Park, K. Matsumoto, K. Lu, G. Bergers, A. Sali, W. A. Weiss and M. K. Aghi (2017). "Cross-activating c-Met/beta1 integrin complex drives metastasis and invasive resistance in cancer." Proc Natl Acad Sci U S A.

Jonas, S. and E. Izaurralde (2015). "Towards a molecular understanding of microRNA-mediated gene silencing." Nat Rev Genet **16**(7): 421-433.

Kanehisa, M., Y. Sato, M. Kawashima, M. Furumichi and M. Tanabe (2016). "KEGG as a reference resource for gene and protein annotation." Nucleic Acids Research **44**(D1): D457-D462.

Keklikoglou, I., C. Koerner, C. Schmidt, J. D. Zhang, D. Heckmann, A. Shavinskaya, H. Allgayer, B. Guckel, T. Fehm, A. Schneeweiss, O. Sahin, S. Wiemann and U. Tschulena (2012). "MicroRNA-520/373 family functions as a tumor suppressor in estrogen receptor negative breast cancer by targeting NF-kappaB and TGF-beta signaling pathways." Oncogene **31**(37): 4150-4163.

Kennecke, H., R. Yerushalmi, R. Woods, M. C. Cheang, D. Voduc, C. H. Speers, T. O. Nielsen and K. Gelmon (2010). "Metastatic behavior of breast cancer subtypes." J Clin Oncol **28**(20): 3271-3277.

Kim, Y. H., J. K. Lee, B. Kim, J. P. DeWitt, J. E. Lee, J. H. Han, S. K. Kim, C. W. Oh and C. Y. Kim (2013). "Combination therapy of cilengitide with belotecan against experimental glioblastoma." Int J Cancer **133**(3): 749-756.

Kozomara, A. and S. Griffiths-Jones (2014). "miRBase: annotating high confidence microRNAs using deep sequencing data." Nucleic Acids Res **42**(Database issue): D68-73.

Lehmann, B. D., J. A. Bauer, X. Chen, M. E. Sanders, A. B. Chakravarthy, Y. Shyr and J. A. Pietenpol (2011). "Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies." J Clin Invest **121**(7): 2750-2767.

Lehmann, B. D., B. Jovanovic, X. Chen, M. V. Estrada, K. N. Johnson, Y. Shyr, H. L. Moses, M. E. Sanders and J. A. Pietenpol (2016). "Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection." PLoS One **11**(6): e0157368.

Leivonen, S. K., R. Makela, P. Ostling, P. Kohonen, S. Haapa-Paananen, K. Kleivi, E. Enerly, A. Aakula, K. Hellstrom, N. Sahlberg, V. N. Kristensen, A. L. Borresen-Dale, P. Saviranta, M. Perala and O. Kallioniemi (2009). "Protein lysate microarray analysis to identify microRNAs regulating estrogen receptor signaling in breast cancer cell lines." Oncogene **28**(44): 3926-3936.

Leivonen, S. K., A. Rokka, P. Ostling, P. Kohonen, G. L. Corthals, O. Kallioniemi and M. Perala (2011). "Identification of miR-193b targets in breast cancer cells and systems biological analysis of their functional impact." Mol Cell Proteomics **10**(7): M110 005322.

Leivonen, S. K., K. K. Sahlberg, R. Makela, E. U. Due, O. Kallioniemi, A. L. Borresen-Dale and M. Perala (2014). "High-throughput screens identify microRNAs essential for HER2 positive breast cancer cell growth." Mol Oncol **8**(1): 93-104.

Li, L., Z. Li, X. Kong, D. Xie, Z. Jia, W. Jiang, J. Cui, Y. Du, D. Wei, S. Huang and K. Xie (2014). "Down-regulation of microRNA-494 via loss of SMAD4 increases FOXM1 and beta-catenin signaling in pancreatic ductal adenocarcinoma cells." Gastroenterology **147**(2): 485-497 e418.

Liedtke, C., C. Mazouni, K. R. Hess, F. Andre, A. Tordai, J. A. Mejia, W. F. Symmans, A. M. Gonzalez-Angulo, B. Hennessy, M. Green, M. Cristofanilli, G. N. Hortobagyi and L. Pusztai (2008). "Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer." J Clin Oncol **26**(8): 1275-1281.

Lin, S. and R. I. Gregory (2015). "MicroRNA biogenesis pathways in cancer." Nat Rev Cancer **15**(6): 321-333.

Loebke, C., H. Sueltmann, C. Schmidt, F. Henjes, S. Wiemann, A. Poustka and U. Korf (2007). "Infrared-based protein detection arrays for quantitative proteomics." Proteomics **7**(4): 558-564.

Lu, J., G. Getz, E. A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B. L. Ebert, R. H. Mak, A. A. Ferrando, J. R. Downing, T. Jacks, H. R. Horvitz and T. R. Golub (2005). "MicroRNA expression profiles classify human cancers." Nature **435**(7043): 834-838.

Lujambio, A., G. A. Calin, A. Villanueva, S. Ropero, M. Sanchez-Cespedes, D. Blanco, L. M. Montuenga, S. Rossi, M. S. Nicoloso, W. J. Faller, W. M. Gallagher, S. A. Eccles, C. M. Croce and M. Esteller (2008). "A microRNA DNA methylation signature for human cancer metastasis." Proc Natl Acad Sci U S A **105**(36): 13556-13561.

Marcato, P., C. A. Dean, D. Pan, R. Araslanova, M. Gillis, M. Joshi, L. Helyer, L. Pan, A. Leidal, S. Gujar, C. A. Giacomantonio and P. W. Lee (2011). "Aldehyde dehydrogenase activity of breast cancer stem cells is primarily due to isoform ALDH1A3 and its expression is predictive of metastasis." Stem Cells **29**(1): 32-45.

Marson, A., S. S. Levine, M. F. Cole, G. M. Frampton, T. Brambrink, S. Johnstone, M. G. Guenther, W. K. Johnston, M. Wernig, J. Newman, J. M. Calabrese, L. M. Dennis, T. L. Volkert, S. Gupta, J. Love, N. Hannett, P. A. Sharp, D. P. Bartel, R. Jaenisch and R. A. Young (2008). "Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells." Cell **134**(3): 521-533.

Minuti, G. and L. Landi (2015). "MET deregulation in breast cancer." Ann Transl Med **3**(13): 181.

Monga, S. P., W. M. Mars, P. Pediaditakis, A. Bell, K. Mule, W. C. Bowen, X. Wang, R. Zarnegar and G. K. Michalopoulos (2002). "Hepatocyte growth factor induces Wnt-independent nuclear translocation of beta-catenin after Met-beta-catenin dissociation in hepatocytes." Cancer Res **62**(7): 2064-2071.

Moss, T. J., Z. Luo, E. G. Seviour, V. Sehgal, Y. Lu, S. M. Hill, R. Rupaimoole, J. S. Lee, C. Rodriguez-Aguayo, G. Lopez-Berestein, A. K. Sood, R. Azencott, J. W. Gray, S. Mukherjee, G. B. Mills and P. T. Ram (2015). "Genome-wide perturbations by miRNAs map onto functional cellular pathways, identifying regulators of chromatin modifiers." NPJ Syst Biol Appl **1**: 15001.

Nakai, K., M. C. Hung and H. Yamaguchi (2016). "A perspective on anti-EGFR therapies targeting triple-negative breast cancer." Am J Cancer Res **6**(8): 1609-1623.

Perou, C. M., T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lonning, A. L. Borresen-Dale, P. O. Brown and D. Botstein (2000). "Molecular portraits of human breast tumours." Nature.

Ricardo, S., A. F. Vieira, R. Gerhard, D. Leitao, R. Pinto, J. F. Cameselle-Teijeiro, F. Milanezi, F. Schmitt and J. Paredes (2011). "Breast cancer stem cell markers CD44, CD24 and ALDH1: expression distribution within intrinsic molecular subtype." J Clin Pathol **64**(11): 937-946.

Riffo-Campos, A. L., I. Riquelme and P. Brebi-Mieville (2016). "Tools for Sequence-Based miRNA Target Prediction: What to Choose?" Int J Mol Sci **17**(12).

Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi and G. K. Smyth (2015). "limma powers differential expression analyses for RNA-sequencing and microarray studies." Nucleic Acids Res **43**(7): e47.

Robson, M., S. A. Im, E. Senkus, B. Xu, S. M. Domchek, N. Masuda, S. Delaloge, W. Li, N. Tung, A. Armstrong, W. Wu, C. Goessl, S. Runswick and P. Conte (2017). "Olaparib for Metastatic Breast Cancer in Patients with a Germline BRCA Mutation." N Engl J Med **377**(6): 523-533.

Saito, Y., G. Liang, G. Egger, J. M. Friedman, J. C. Chuang, G. A. Coetzee and P. A. Jones (2006). "Specific activation of microRNA-127 with downregulation of the proto-oncogene BCL6 by chromatin-modifying drugs in human cancer cells." Cancer Cell **9**(6): 435-443.

Selbach, M., B. Schwanhausser, N. Thierfelder, Z. Fang, R. Khanin and N. Rajewsky (2008). "Widespread changes in protein synthesis induced by microRNAs." Nature **455**(7209): 58-63.

Smith, L., E. W. Baxter, P. A. Chambers, C. A. Green, A. M. Hanby, T. A. Hughes, C. E. Nash, R. A. Millican-Slater, L. F. Stead, E. T. Verghese and V. Speirs (2015). "Down-Regulation of miR-92 in Breast Epithelial Cells and in Normal but Not Tumour Fibroblasts Contributes to Breast Carcinogenesis." PLoS One **10**(10): e0139698.

Song, L., D. Liu, B. Wang, J. He, S. Zhang, Z. Dai, X. Ma and X. Wang (2015). "miR-494 suppresses the progression of breast cancer in vitro by targeting CXCR4 through the Wnt/beta-catenin signaling pathway." Oncol Rep **34**(1): 525-531.

Sorlie, T. (2004). "Molecular portraits of breast cancer: tumour subtypes as distinct disease entities." Eur J Cancer **40**(18): 2667-2675.

Sorlie, T., C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, T. Thorsen, H. Quist, J. C. Matese, P. O. Brown, D. Botstein, P. E. Lonning and A. L. Borresen-Dale (2001). "Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications." Proc Natl Acad Sci U S A **98**(19): 10869-10874.

Sorlie, T., R. Tibshirani, J. Parker, T. Hastie, J. S. Marron, A. Nobel, S. Deng, H. Johnsen, R. Pesich, S. Geisler, J. Demeter, C. M. Perou, P. E. Lonning, P. O. Brown, A. L. Borresen-Dale and D. Botstein (2003). "Repeated observation of breast tumor subtypes in independent gene expression data sets." Proc Natl Acad Sci U S A **100**(14): 8418-8423.

Stambolic, V., L. Ruel and J. R. Woodgett (1996). "Lithium inhibits glycogen synthase kinase-3 activity and mimics wingless signalling in intact cells." Curr Biol **6**(12): 1664-1668.

Tan, C. L., J. L. Plotkin, M. T. Veno, M. von Schimmelmann, P. Feinberg, S. Mann, A. Handler, J. Kjems, D. J. Surmeier, D. O'Carroll, P. Greengard and A. Schaefer (2013). "MicroRNA-128 governs neuronal excitability and motor behavior in mice." Science **342**(6163): 1254-1258.

Torre, L. A., F. Bray, R. L. Siegel, J. Ferlay, J. Lortet-Tieulent and A. Jemal (2015). "Global cancer statistics, 2012." CA Cancer J Clin **65**(2): 87-108.

Uhlmann, S., H. Mannsperger, J. D. Zhang, E. A. Horvat, C. Schmidt, M. Kublbeck, F. Henjes, A. Ward, U. Tschulena, K. Zweig, U. Korf, S. Wiemann and O. Sahin (2012). "Global microRNA level regulation of EGFR-driven cell-cycle protein network in breast cancer." Mol Syst Biol **8**: 570.

van Dongen, S., C. Abreu-Goodger and A. J. Enright (2008). "Detecting microRNA binding and siRNA off-target effects from expression data." Nat Methods **5**(12): 1023-1025.

Vermeulen, L., F. D. S. E. Melo, M. van der Heijden, K. Cameron, J. H. de Jong, T. Borovski, J. B. Tuynman, M. Todaro, C. Merz, H. Rodermond, M. R. Sprick, K. Kemper, D. J. Richel, G. Stassi and J. P. Medema (2010). "Wnt activity defines colon cancer stem cells and is regulated by the microenvironment." Nature Cell Biology **12**(5): 468-U121.

Vogel, C. and E. M. Marcotte (2012). "Insights into the regulation of protein abundance from proteomic and transcriptomic analyses." Nat Rev Genet **13**(4): 227-232.

von der Heyde, S., J. Sonntag, D. Kaschek, C. Bender, J. Bues, A. Wachter, J. Timmer, U. Korf and T. Beissbarth (2014). "RPPanalyzer toolbox: an improved R package for analysis of reverse phase protein array data." Biotechniques **57**(3): 125-135.

Wang, K., X. Wang, J. Zou, A. Zhang, Y. Wan, P. Pu, Z. Song, C. Qian, Y. Chen, S. Yang and Y. Wang (2013). "miR-92b controls glioma proliferation and invasion through regulating Wnt/beta-catenin signaling via Nemo-like kinase." Neuro Oncol **15**(5): 578-588.

Weigelt, B., J. L. Peterse and L. J. van 't Veer (2005). "Breast cancer metastasis: markers and models." Nat Rev Cancer **5**(8): 591-602.

Zhang, J. L., Y. Liu, H. Yang, H. Q. Zhang, X. X. Tian and W. G. Fang (2017). "ATP-P2Y2-beta-catenin axis promotes cell invasion in breast cancer cells." Cancer Sci **108**(7): 1318-1327.

Zhuang, L. K., Y. T. Yang, X. Ma, B. Han, Z. S. Wang, Q. Y. Zhao, L. Q. Wu and Z. Q. Qu (2016). "MicroRNA-92b promotes hepatocellular carcinoma progression by targeting Smad7 and is mediated by long non-coding RNA XIST." Cell Death Dis **7**: e2203.

Supplementary tables and figures
Due to size limitations, these files are available on the CD version of the thesis

Table 1

Target protein and corresponding gene names matched with the average negative and positive significant FC, amount of significant negative and positive interactions, 3'UTR length from TSv7, enrichment results, and validated miRTarBase miRNA numbers.

Table 2

miRNAs passing the randomisation test. One tab for each pathway. Vec_sign values indicates if the miRNA is a repressor (-1) or activator (+1) for the pathway. d-values represent the distance between the PC score of the miRNA and the average random PC score for the same miRNA after 10'000 randomizations.

Table 3

PC score calculated with increasing thresholds in q-value and FC (M value on the table). This testing was performed only for the WNT/beta-catenin signalling pathway to help for the identification of miRNAs to focus on for functional validation.

Table 4

q-values and log2FC for the entire RPPA screening, analysed with the limma testing by Astrid Wachter.

Figure 1 and 2

Heatmaps of moderate t-statistics for Integrin (1) and c-Met (2), including the tree of euclidean distance. The name of the targets is listed as the corresponding gene name. Please refer to supplementary table 1 for the protein target identifier.

Figure 3

Volcano plots of q-values (y-axis) and log2FC (x-axis) for all the targets probed within the RPPA screening. Values represented correspond to the ones reported in supplementary table 4.

Figure 4

c-Met and Integrin signalling pathways are represented as from KEGG pathway with colour coding indicative of putative repressive or activating function for each protein.

Acknowledgements

My time at DKFZ was particularly long and constellated by multiple encounters with wonderful people with whom I spent lots of time, experienced many events, and created memories which I will always hold in my heart. As a consequence, these acknowledgments are going to be long, all because I like to walk down the memory lane. I apologise in advance to those whom I didn't personally cite in these acknowledgments. Be sure: if I met you in these years, some sort of impact on my life and PhD you surely made.

I would like to thank first my supervisor Prof. Dr. Stefan Wiemann for giving me the opportunity to work in the Division of Molecular Genome Analysis at the German Cancer Research Center (DKFZ) in Heidelberg. Throughout these five years, I was allowed to pursue openly and freely my scientific interests and I will always be grateful for this.

I am extremely thankful for the guidance I had from Dr. Ulrike Korf. She was instrumental for the success of my project, thanks to the support throughout the RPPA screenings. Her loss surely impacted many of us, also at a personal level.

I am very grateful to Prof. Dr. Sven Diederichs and Prof. Dr. Yosef Yarden for their expertise and suggestions, not only through thesis advisory committee meetings but also when I requested additional help. I am particularly thankful to Prof. Dr. Yosef Yarden for hosting me in his lab in 2015 allowing me to learn new techniques and to experience new approaches to science, opening up to novel opportunities. I shall never forget that the sky is the limit, and that my PhD is like a key that can open (as well as close) doors to my future.

I would like to thank our collaborators. First and foremost, I am grateful to Dr. Astrid Wachter and Prof. Dr. Tim Beissbarth who contributed invaluably to the miRNome project. It required many years of long distance commitment, but I'd like to say that we were a successful partnership. Additionally, I am thankful to Dr. Rainer Will and to Rita and Brigit for the production the Stable isogenic recombinant (Sir) cell lines I used for validating the miRNA regulating the WNT pathway.

Collaborating with Dr. Roni Golan-Lavi, Dr. Gari Fuks, and Professors Eytan Domany and Yosef Yarden was also a great experience and I am very thankful for it and our final success.

All the MGA members had been wonderfully supportive through these years and I am more than happy to have met you all. Undoubtedly fundamental for the fruitful completion of this thesis as well as for the collaboration with Israel were Corinna and Daniela, the technicians “of the proteomic side” of MGA. Heike, Angelika, Moritz, Sara, and Sabine were also extremely helpful throughout these five years. I would like to specifically thank Ewald. Your positivity is dearly missed and I wish I had learned more cloning tricks from you. Surely great times were spent with all the other PhD and master students and postdoc that passed by the division or stayed through the years. For all these, I will forever be thankful.

I am grateful beyond words for the pack of PhD fellows that I met in the MGA group. You made this experience a brilliant time. I will forever carry with me great memories of solardraisine rides, canoeing on the Neckar, Sherlock Holmes board game nights (no, sorry, it is Scotland Yard!), discussions about movies and tv series, Exploding Kittens, retreats of any kind, and of course every single coffee break! So, thank you all for making these years unforgettable.
Special mentions go to Dr. Bott, my doctoral “brother in arms”, with whom I shared particularly though times, especially at the beginnings. I still think you should open a Bott-Brot bakery. And, of course, thank you Alex also for having introduced to me my future husband.
Neşe was the first person who ever confronted me after a particularly bad burst of mine toward her. It goes back to a time when I thought that a master student could not know more than me. Clearly, I was wrong. It might be that for you it’s not such a wonderful or particularly important memory. For me it was the moment I realized how much I liked you.
Devina came in when none of us had seen her before for interviews, nor had really seen applications circulating. She was a mystery. We only knew she would work on metabolism, which was a total mystery as well. She’s our living and walking dictionary “Devina how do you say…?” The only mistake you ever made in English will forever be attached to you, my dear, as I will always ask you once per year whether you managed to get a special hood for cytostatistics.

The epigenetics brothers, Simo and Emre, made a wonderful addition to our ever growing group. Tho, dudes, please make sure next time that you check your cell lines properly.

Another group of people I enjoyed spending my time with were the fellow German-Israeli Helmholtz Research School students. I experienced with you guys seriously wonderful moments, both in Israel and here in Germany.
So thank you Moritz, Laura, Mareike, Michael, Yonatan, and Alex, for all these moments we shared through the years.
I am particularly thankful for the time, talks, support, fun, sport activities, (what did I forget? Drinks?) that I spent with Kathrin. I didn't really know you until I picked you up on January 2015, when we were both heading to Israel for our first stay. We were wearing the same shirt from H&M, except yours had green stripes and mine had pink ones. We were dropped by our taxi driver in front of the wrong building, when we arrived at the Weizmann. And under pouring rain we had to find where the hell we would be sleeping in the next months. It started as an adventure, and I was happy to be with you to experience all of this.

While in Israel, I met additional students which were there at the Weizmann studying or visiting. We had great times together: Purim, visiting other cities, trespassing (by mistake) kibbutz, walking on dangerous collapsing grounds near the Dead Sea, playing beach volleyball in Tel Aviv, movie nights and pasta nights. I'm particularly thankful to Alexey and Mira, as the time we spent together truly helped me not to sink in the situations I had put myself in.

When I decided I wanted to be in the PhD student council, I expected to become great friends with my fellows, as I had seen Aoife doing with her own co-council members (more about Aoife later). I guess not all the groups are made with the same paste, so while I survived one year organizing interesting things with interesting people, I never truly bond with any of them. On the other hand, organizing the Heidelberg Forum for Young Life Scientists put me in contact with people I will never forget. First and foremost: my co-boss Patricia. We juggled our duties across countries when I was in Israel and then you were in Switzerland, even if not always super smoothly I like to think that we have been harmonised in achieving the same level of perfection in this conference organization. I'm happy that you now have Harry, take good care of him! You're the most caring person I've ever met, and I will never forget all you did for me throughout this time. Antonino also deserves special mentioning, as he had to withstand my constant

pressure. Nevertheless, he managed to befriend me and had also been very caring and supportive through time. To all the other HFYLS organisers: thank you for being there and for building all together this conference. I am still proud of what we achieved.

During my time in Heidelberg, I also tried some acting with the “the Big Thing” group led by Veli. I wasn’t particularly talented, but enjoyed rehearsing and trying things with all of you guys. It was great fun and brought on me a lot of thoughts about how and who I am. Definitely, not a Paris. Thank you for all the time together and the stage sharing.

When I first arrived, I was quite lost. However, I was very lucky: a special person took great care of me. Aoife had been with me in every baby step I had to take. She left the lab (way too soon for my taste!) but still kept me close. Even if we don’t hear each other very often, I am grateful beyond words for all she has done for me. Her greatest lesson came for me when she delivered the congratulation addresses at the graduation ceremony at DKFZ in 2013. She remembered everyone that, while it is easy to compare yourself to others, especially in science where the achievements are always showcased, the most important thing is to look back at the past and compare us to that. My own growth was strongly inspired by you, Aoife.
Thanks to Aoife I met also other great people with whom I spent interesting time. Among them, I am particularly thankful to Ansam. As she also organised two conferences while at DKFZ, the connection with her was very fruitful for that purpose. However, Annie, I prefer to remember all the times we spent dancing and having fun. A brilliant memory for me will always be at Aoife’s bachelorette in Budapest, when we went to the Dracula cave and you tried to divert my attention from something extremely scary by talking about asking Lyko not to be your second examiner anymore. Except it didn’t work, we run away together from the dark foggy corridor because you know, even if we are scientist, we believe in vampires.

My time in Heidelberg corresponded to the first time I left home, Milano, “for good”. My best friends, my family members, everyone I “left behind” have all been supportive in ways I never thought possible. Time and distances didn’t break our bonds and I hope they never will.
I am particularly thankful to my two best friends Francesca and Giulia because we managed to still meet and have our brunches, and we remained ourselves. I knew you’d be there for me every time I’d come back. This certainty, made it always very sweet to take the car on Friday evenings to drive home, even at impossible hours.

Old lab buddies from my time in Milano have been there every time to remind me that yeah, it's a PhD, it's supposed to be so hard/confusing/stressful/fun/mindbreaking. Lucia, Simona, Chiara, Matteo, Marilena, (and relative +1s) have been a stronghold where I could find support all the time. I am not sure I am allowed to, but I consider my ex-boss Francesco also a lab buddy. I missed you all guys, and sometimes I thought that it would have been much better had I stayed behind. Of course nobody knows if this is true.
Angelo always stayed close, always cared, even if sometimes a lot of time passed between the chances we had to see each other. Your friendship is an immense gift to me.
All my cousins and aunts and uncles were there to discuss and confront throughout these years. The increasing amount of babies in our family brings me happiness (and scares me sometimes), but I am overjoyed to have all of you next to me in this life.
My Grandpa Sandro still writes me e-mails and supports me with unconditional love. Nonno, sei un sostegno enorme, con 92 anni di serenita' e saggezza che spero solo di poter raggiungere in un forse domani.
My brother Alessandro and my sister Giulia and my dad are the persons that I miss the most. Though not very often, within these years, our talks have always been important for me. I wish mum would still be with us, to share the great moments we had together, the birth of a new life, the Christmases, the holidays, the departures, the returns, the family gatherings.

The last person I would like to thank is Gideon, whom I met one day in a corridor at DKFZ, when Alex introduced us. You've been next to me throughout all this, supporting me and challenging me to be better than ever. And if life is like a mountain, I want to climb it with you.