Chapter 2
MicroRNA as an Integral Part of Cell Communication: Regularized Target Prediction and Network Prediction

Rolf Backofen, Fabrizio Costa, Fabian Theis, Carsten Marr, Martin Preusse, Claude Becker, Sita Saunders, Klaus Palme and Oleksandr Dovzhenko

Abstract MicroRNAs, gene encoded small RNA molecules, play an integral part in gene regulation by binding to target mRNAs and preventing their translation. The prediction of microRNA–mRNA-binding sites and the resulting interaction network are essential to understand, and thus influence, regulation of a genetic information flow inside the living organism. Numerous algorithms have been proposed based on

R. Backofen (✉) · F. Costa · S. Saunders
Institute of Computer Science, Freiburg University, George-Köhler-Allee, Geb. 106, 79110 Freiburg, Germany
e-mail: backofen@informatik.uni-freiburg.de

F. Costa
e-mail: costa@informatik.uni-freiburg.de

S. Saunders
e-mail: sita@informatik.uni-freiburg.de

F. Theis · C. Marr · M. Preusse
Institute of Computational Biology, Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany
e-mail: fabian.theis@helmholtz-muenchen.de

C. Marr
e-mail: carsten.marr@helmholtz-muenchen.de

M. Preusse
e-mail: martin.preusse@helmholtz-muenchen.de

K. Palme
Institute of Biology/Botany, Freiburg University, Schänzlestr. 1, 79104 Freiburg, Germany
e-mail: klaus.palme@biologie.uni-freiburg.de

O. Dovzhenko
Institute of Biology/Molecular Plant Physiology, Freiburg University, Sonnenstr. 5, 79104 Freiburg, Germany
e-mail: alexander.dovzhenko@biologie.uni-freiburg.de

C. Becker
Gregor Mendel Institute of Molecular Plant Biology GmbH, Dr. Bohr-Gasse 3, 1030 Vienna, Austria
e-mail: claude.becker@gmi.oeaw.ac.at

M. Bossert (ed.), *Information- and Communication Theory in Molecular Biology*, Lecture Notes in Bioengineering,
DOI 10.1007/978-3-319-54729-9_2

various heuristics; however the predictions often vary considerably. In this proposal we will extend a physical model for the binding of microRNAs to the corresponding target and establish an extended set of features influencing binding probabilities. We will be faced with the challenge of (i) too many features and (ii) few known interactions on which to train any prediction algorithm. This problem will be solved using (i) information-theoretical criteria for feature reduction, (ii) regularization, (iii) application of the Infomax approach to guarantee minimal loss of information after dimension reduction, and (iv) experimental validation of theoretical predictions using a novel test system. This strategy will allow (i) statistical analysis of the predicted microRNA–mRNA hypergraph, (ii) characterization of network motives and hierarchies, (iii) identification of missing links, and (iv) removal of false interactions.

DFG Grant No.
BA 2168/4, TH 900/3, and PA 279/21

Applicants

Rolf Backofen	University of Freiburg (ALU/LBI)
Fabian Theis	Helmholtz Zentrum Munich and Technical University of Munich (HMGU/TUM)
Klaus Palme	University of Freiburg (ALU/MPB)
Oleksandr Dovzhenko	University of Freiburg (ALU/MPB)

Scientific staff

Fabrizio Costa	ALU/LBI
Sita Saunders	ALU/LBI
Milad Miladi	ALU/LBI
Martin Preusse	HMGU/TUM
Michael Schwarzfischer	HMGU/TUM
Carsten Marr	HMGU/TUM
Fugang Ren	ALU/MPB
Bingshan Wang	ALU/MPB

Funding duration 01/2010 – 12/2017

Cooperation

- Prof. Ivo Große, Institut für Informatik, Martin-Luther-Universität
- Prof. Dr. Martin Bossert, Ulm University
- Prof. Dr. Daniel Keim, University of Konstanz
- Prof. Marie-France Sagot, INRIA Grenoble Rhône-Alpes
- Dr. Chris Brown, University of Otago
- Prof. Dr. Stefan Engelhardt, Technical University of Munich (TUM)
- Prof. Markus Landthaler, Berlin Institute for Medical Systems Biology at the Max-Delbrück-Center for Molecular Medicine (MDC Berlin)

Number of publications

Journal publications	10
Conference contributions	1

2.1 Conclusion

2.1.1 Summary of the Essential Results

In this project, we investigated in detail the information necessary for a functional miRNA-binding site on its target mRNA. In the first part of the project, we worked on an RNA-centric view. Besides sequence properties, miRNA function is critically influenced by the RNA secondary structure, which has to be predicted. However, prediction of mRNA structure is inherently error-prone. Since we need only local structure information, we first improved current approaches for local secondary structure prediction. We then improved the detection of conserved secondary structure motifs using an advanced graph kernel that extends the idea of k-mers with gaps to graphs, which in this case represent the RNA structures. We used this graph kernel further to improve the prediction of AGO-binding sites from CLIP data to improve the prediction of miRNA-binding site using this very recent type of data.

In the second part of the project, we worked on a network-based approach. In order to circumvent limitations in current miRNA targeting data, we included systemic features of miRNA-mediated gene regulation into functional analysis methods. We thus aimed at capturing system-level effects of miRNA regulation to improve functional classification. Our genome-wide analyses indicate that distance-dependent cooperativity, tissue-specific pathway regulation, and co-regulation with RNA-binding proteins serve as predictors for novel regulatory circuits. We implemented our methods in web applications to share the improved functional analysis platform with the research community.

In the third part of the project, we aimed on using the information acquired by the first two parts of the project for design and testing design functional artificial microRNAs in plants. For this we developed fluorescence- and luminescence-based assays in single plant cells to assess sequential and context-specific features to design amiRNAs in silico with defined functionality. Target sequences and tested amiRNAs were transiently co-expressed in Arabidopsis protoplasts. Translational fusion of the target sequences with fluorescent or luminescent reporters were generated using specially developed molecular cloning tools. Subsequently, amiRNA-mediated changes in expression levels of the targeted sequences were quantitatively monitored using corresponding detectors. After normalization, the generated data sets were fitted into the prediction pipeline for identification of relevant features and their subsequent validation. The robustness of our approach was demonstrated by generating plant lines expressing amiRNAs with different functionalities (non-efficient,

partially functional, or efficiently functional) against *PIN1* (auxin transport protein) and *JMJ10* (JUMONJI-transcription factor) targets. A strong correlation of *ex vivo* assays with *in planta* phenotype studies proved high efficiency of the developed amiRNA evaluation platform and relevance of design features. We further applied site-specific mutagenesis to context environment of the target sites and demonstrated that RNA regions surrounding the target site influence amiRNA functionality. In these experiments originally non-functional amiRNAs showed functionality in some cases, thus indicating on the relevance of secondary mRNA structure and other context features relevant for amiRNA design. Therefore, PAR-CLIP approach (Hafner et al. 2010) was established in plant cells in order to further improve amiRNA design by integrating mRNA–protein interaction information.

2.1.2 Possible Future Work

Concerning the design of amiRNAs, we have so far exploited information from CLIP data for AGO–mRNA interactions. However, a full integration of this type of information in the amiRNA design pipeline will require an extension of the models learned by GraphProt from CLIP data using the actual interaction hybrid.

For the network-based approach, the functional miRNA analysis could be extended with data on genetic variation. Considering SNPs and CNVs would allow to identify mechanisms behind misregulation of miRNAs in a patient-specific disease context.

Concerning amiRNAs, our approach for validation of amiRNA functionality can be fully automated, therefore enabling considerable increase of experimental data points as well as generation of pathways—up to genome-wide amiRNA libraries for plant research. In concert with PAR-CLIP data this will lead to comprehensive understanding of relevant features and high accuracy of *in silico* amiRNA design. By achieving this goal and by combining with the precise genome editing, amiRNAs can be used as a valuable tool not only for fundamental studies in plants, but as well as a unique platform for rational crop design.

2.2 Working and Results Report

2.2.1 Starting Point

MicroRNAs (miRNAs), gene encoded small RNA molecules, are a key component of eukaryotic cell regulation and are pervasive in both plants and animals (Axtell et al. 2011). They play important roles in many different processes like cell differentiation, proliferation, apoptosis, or development. Plant miRNAs (Dezulian et al. 2006; Schubert et al. 2005; Zhang et al. 2005, 2006) have a higher complementarity to their target than animal miRNAs and often bind in the coding region, triggering the cleavage and successive degradation of their target mRNA. The prediction of

microRNA–mRNA binding sites and the resulting interaction network are essential to understand, and thus to influence, the regulation of a genetic information flow inside the living organism. Numerous algorithms have been proposed based on various heuristics. However, the predictions often vary considerably due to several challenges in the prediction task. Here, most critically is the fact that we are given too few known miRNA interactions and have to assess too many RNA-related features in the prediction task.

2.2.2 Work Performed

Influence of Secondary Structure on miRNA Function.

It is well known that miRNA function is critically influenced by RNA secondary structure. First, the processing of pre-microRNAs by DICER requires a specific secondary structure in form of a hairpin loop. Second, the binding of miRNA requires the binding site to be accessible. However, the prediction of secondary structures, especially for mRNAs, is error-prone. For that reason, we (1) improved the prediction of local structures, (2) developed a clustering approach to determine conserved secondary structures and applied it to detect putative miRNA genes, and (3) used secondary structure in an advanced machine learning approach to characterize AGO-binding sites from CLIP data:

1. **Determination of local mRNA structure**. The accessibility of binding site is determined by the structural properties of the mRNA, an effect that we already observed for the DNA-binding sites of transcription factors (Pudimat et al. 2005). However, due to several factors such RNA-binding proteins, secondary structure prediction for mRNA is inherently error-prone, especially for long-range interactions. The state-of-the-art is to use local folding, even combined with averaging over different windows to determine locally stable motifs. In Lange et al. (2012), we first assessed whether local folding is actually better suited to determine local RNA motifs then global folding. We then investigate different local folding approaches in more detail, finding that the popular windowing approaches suffer from a border effect, favoring long-range base pairs, which can be avoided by choosing appropriate combinations of window lengths and base pair spans (see Fig. 2.1).
2. **Detection of RNA motifs**. RNA motifs play a crucial role, not only in the processing of functional miRNAs, but also as possible elements regulated miRNA binding. MiRNA is an example of RNA elements that do not stem from a common ancestor, but share a similar secondary structure important for function. A collection of structurally similar RNA elements with the same function is termed RNA class, in contrast to RNA families, which share a common ancestor. Members of an RNA classes usually share a lower sequence identity, which implies that they cannot be found via sequence alignments. Instead, clustering with sequence structure alignments tools such as LocaRNA (citation missing) has to be used to detect

new RNA classes. However, the high complexity of these approaches hinders their large-scale application. This especially poses a problem with the detection hundreds of thousands putative noncoding RNAs. In Mendes et al. (2012), we used a feature vector representation of sequence and structure properties of candidates to determine putative miRNA on a genome-wide scale. In Heyne et al. (2012), extended this approach to general RNA classes by developing an alignment-free approach to efficiently cluster hundreds of thousands RNAs according to sequence and structure. It is based on a graph kernel, which extends the idea of k-mers with gaps to subgraphs.

3. **Determination of AGO-binding sites**. MiRNA targets are recognized by a complex consisting of an Argonaute protein (AGO) loaded with the mature miRNA. CLIP is an experimental protocol that uses a combination of crosslinking, an antibody against AGO to enriched miRNA-bound RNAs, and sequencing of these bound RNAs to determine AGO (and hence miRNA)-binding sites. However, this protocol highly depends on the used cell type, and results cannot be transferred to other cell types or conditions. We extended our graph kernel approach in Heyne et al. (2012) to determine binding models for RNA-binding proteins such as AGO. Using a positive set derived from the CLIP data, and a computationally derived negative set, sequential and structural features are learned using an SVM. The features are generated by a graph kernel (see Fig. 2.2). An important extension is the concept of a *viewpoint*, which restricts the set of features to those that

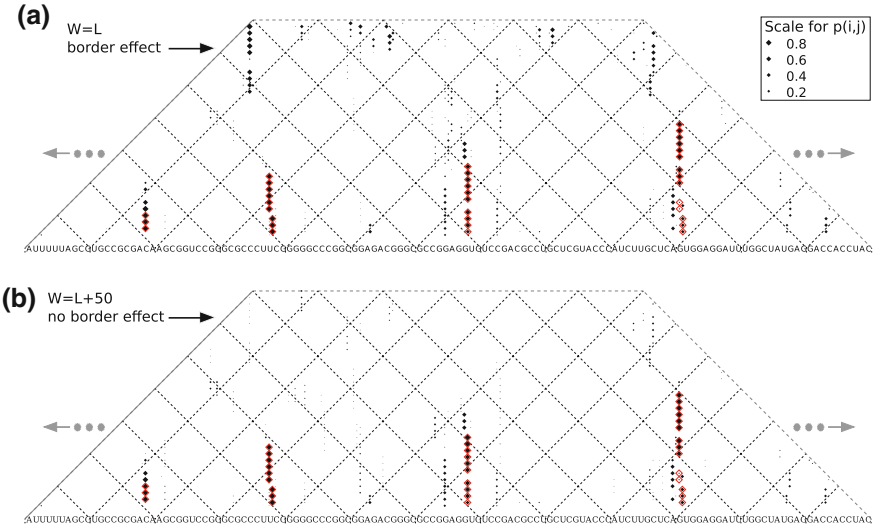

Fig. 2.1 Border effect in local secondary structure prediction. We display the base pair probability matrices for a heat shock gene expression (ROSE) element, with the base pairs of the target structure marked in *red*. The size of each dot corresponds to the probability of the associated base pair. If the window size is close to the considered base pair range (*top panel*), incorrect long-range base pairs are much more likely. Figure taken from Lange et al. (2012)

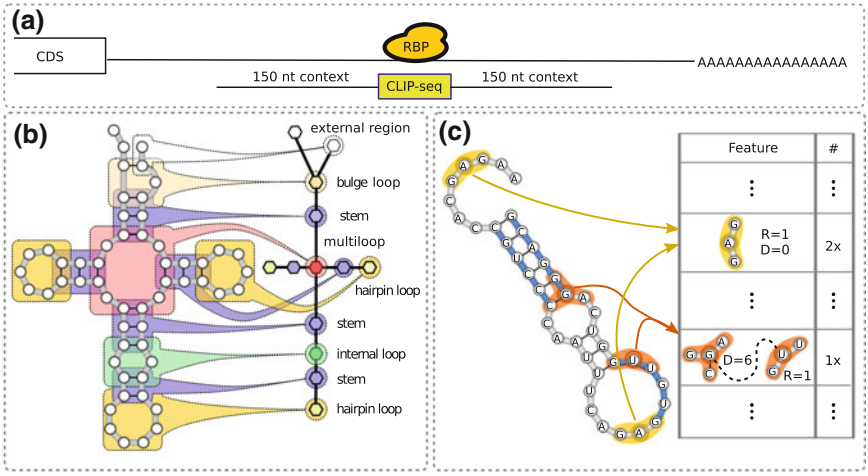

Fig. 2.2 Encoding of RNA sites bound by RNA-binding proteins sites. **a** We calculate the secondary structure from the region identified in the CLIP experiment, extended by a context of 150 nt. **b** The resulting secondary structure is represented by a graph, enriched by additional information about substructures. **c** From this, a very large set of features corresponding to small subgraphs or pair of subgraphs is extracted. Figure taken from Lange et al. (2012)

start withing the actual binding sites, thus greatly reducing the number of used features.

Systemic Analysis of miRNA Function.

Post-transcriptional regulation of target genes by miRNAs is a widespread phenomenon that influences most mammalian genes. In many cases, however, the actual function of miRNA-mediated regulation *in vivo* is not clear. Current functional analyses do not account for the complexity of miRNA regulation due to limitations of targeting data and enrichment methods. We thus included three systemic features of miRNA-mediated regulation into functional miRNA analysis in order to capture systemic effects (Fig. 2.3):

1. **Distance-dependent cooperativity of miRNAs**. It is generally believed that mammalian mRNAs carry multiple miRNA-binding sites and are in fact regulated by multiple miRNAs simultaneously. Experimental studies with reporter constructs suggest that binding sites in close proximity increase the down-regulation of target genes and produce cooperative effects, that is the repression of the target gene is higher than the additive effects of the individual binding sites. We thus performed a genome-wide analysis of binding site distributions. Our results demonstrated that distance-dependent miRNA cooperativity is a widespread phenomenon that is especially relevant for regulation by multiple different miRNAs. Based on these findings, we developed *miRco*, a web tool to predict cooperative

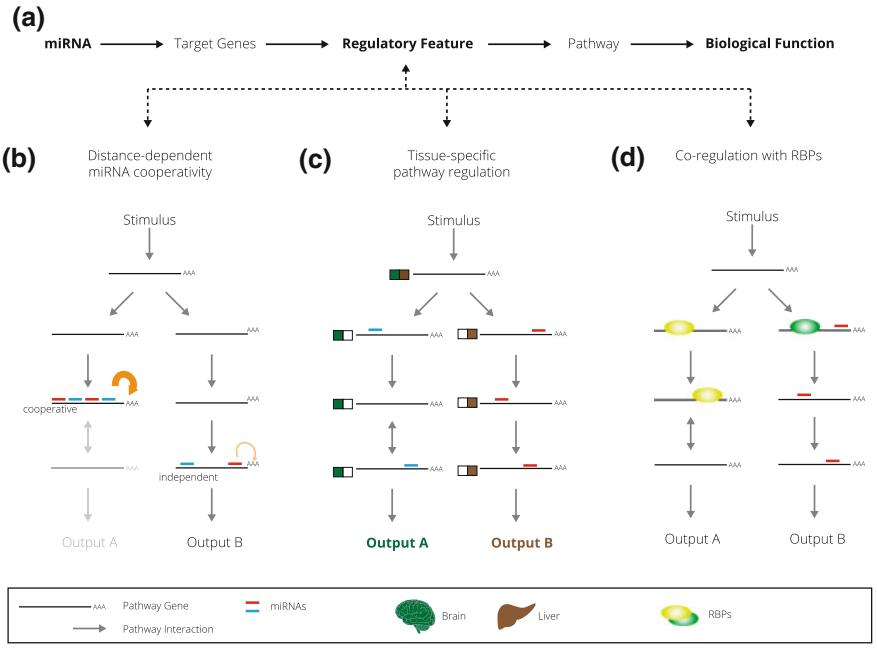

Fig. 2.3 Functional analysis of miRNAs and extended concepts of miRNA regulation. **a** Additional regulatory features to capture systemic effects of miRNA-mediated regulation in functional pathway analysis. **b** Neighboring miRNA-binding sites lead to an increased down-regulation of target genes (*red* and *blue* miRNA). **c** Genes are not uniformly expressed among tissues. MiRNA regulation of pathways is thus tissue-specific (*blue* miRNA in brain and *red* miRNA in liver). **d** miRNAs- and miRNA-independent RBPs have been reported to cooperate in regulation of gene expression

regulation of miRNAs (http://mips.helmholtz-muenchen.de/mirco/) (Rinck et al. 2013).

2. **Tissue-specific pathway regulation**. Protein coding genes are not uniformly expressed among different cell types and tissues. Consequently, miRNA-mediated regulation of biological pathways could be tissue-specific and may contribute to cell-type-specific modulation which has been reported for various signaling pathways. To capture tissue-specific effects, we developed a novel methodology for tissue-specific pathway analysis of miRNAs (Kowarsch et al. 2011). In an update to our method, we incorporated the most recent and highest quality miRNA targeting data (TargetScan and StarBase), RNA-seq-based gene expression data (EBI Expression Atlas) and multiple new pathway data sources to increase the biological relevance of the predicted miRNA-pathway associations (Preusse et al. 2016). We developed the web tool *miTALOS v2* to share our methodology with the research community (http://mips.helmholtz-muenchen.de/mitalos).
3. **Co-regulation with RNA-binding proteins**. RNAs are constantly bound by numerous RNA-binding proteins (RBPs). They participate in regulation of all

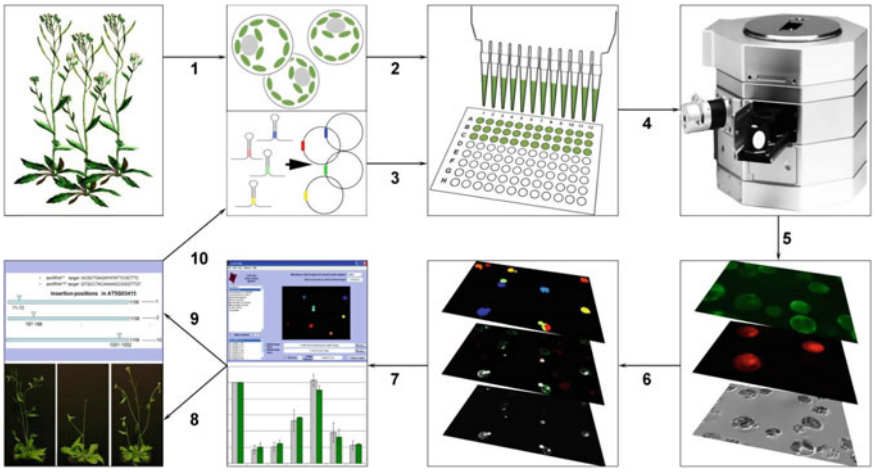

Fig. 2.4 Protoplast-based amiRNA screening and experimental validation pipeline. After isolation (1), protoplasts are transferred to microtitre plates (2) and transient transformation with GATEWAY-based amiRNA screening vectors (3) is performed. Plates are analyzed (4) using iMIC automated microscope. Images are acquired for corresponding emission channels (5) and processed using CellProfiler freeware (6). Data extraction and statistical analysis (7) are providing information on efficient amiRNAs to be transferred into plants (8) or for computational analysis of features (9). Iterative learning can be repeated according to needs (10)

steps of the mRNA life cycle from transcription to translation. It has been shown that miRNAs and other RBPs interact in regulation of gene expression. To identify microRNAs and RBPs with a similar functional context, we developed *simiRa*, a tool that compares enriched functional categories such as pathways and GO terms. By comparing the enriched categories for RBPs and miRNAs, we identified regulators with a similar biological function and demonstrated the usability of our approach in a case study identifying possible indirect cooperation of miRNAs and the nuclear RBP TAF15 (Preusse et al. 2015). *SimiRa* is available as a web tool (http://vsicb-simira.helmholtz-muenchen.de/).

Experimental Validation and Forward Genetics Search

Despite fragmented reports on successful exploitation of amiRNAs to specifically silence target genes, systematic studies of sequential and structural (context) features underpinning amiRNA/target sequence lacked simple and efficient experimental platform. Therefore, we developed single plant cell assays and corresponding molecular tools to deliver the datasets necessary for iterative learning processes allowing for feature analysis and their validation. Initially, an established fluorescent amiRNA screening pipeline (Fig. 2.4) was applied for validation of amiRNA efficiency in Arabidopsis protoplasts with subsequent conformation of their functional properties in planta. Upon functional assessment of 61 amiRNAs against *PIN1* as the target amiRNA were classified in three groups non-functional, partially functional and

efficiently functional. Remarkably, despite all amiRNAs were selected as functional accordingly to initial *in silico* prediction, about only 10% of them showed efficient silencing of *PIN1* target that was translationally fused to GFP reporter. Randomly selected representatives of each group were transformed in the Arabidopsis plants and evaluated. Phenotypic studies showed a direct correlation between the amiRNA activity in single-cell assays and in planta, thus suggesting that amiRNA validation pipeline in single cells can be considered as a reliable toolbox for further analysis of features determining amiRNA functionality.

The primary PIN1 screening datasets resulted in identification of features negatively influencing functional properties of amiRNAs. For validation of the determined sequential features negatively influencing amiRNA-mediated gene silencing, a set of 10 new amiRNAs against the JMJ10 was generated using the same design platform (WMD tool) and only candidates predicted as highly efficient were used. These 10 candidates were independently evaluated *in silico* and experimentally. Functionality of amiRNA sequences assessed experimentally and predicted by the new computational approach revealed only a single mistake in non-functional in screenings $amiR^{jmj10\text{-}10}$ was predicted as functional.

In order to further improve prediction, we adapted cloning vectors for mutagenesis studies and substituted fluorescent reporters with luminescent analogues to facilitate the data generation and processing (Fig. 2.5). To ensure functionality of new molecular components used in the screening vectors (luminescence reporters and 2A peptide) we developed an alternative approach for rapid luminescent ratiometric assessment, which was applicable and highly efficient in intact plant cells (Wend et al. 2013). For this, firefly (Fluc) and renilla (Rluc) luciferase enzymes were linked with 2A peptide and transiently expressed under CaMv 35S strong promoter. So-called 2A-like peptides greatly simplify expression of several proteins from a single open reading frame (ORF) by a self-processing mechanism (Tang et al. 2009). Further, we developed a strategy to specifically change the ratio between both enzymes in tests by translationally fusing the Fluc component with another peptide, sensitive to a common intracellular phytohormone auxin. Comparison of non-degrading control with the sensor constructs in transiently transformed protoplasts proved high sensitivity and robustness of ratiometric luminescence assessment and thereby was further implemented to develop luminescence-based amiRNA screenings.

Next, amiRNAs against Fluc were designed and selected highly functional $amiR^{Fluc}$ as a positive control for subsequent context perturbation assays. Afterward, evaluation of the context features was performed. For this, target sites (TSs) specific for either for efficiently functional $amiR^{PIN1\text{-}1}$ or non-functional $amiR^{PIN1\text{-}35}$ were inserted in various context positions (Fig. 2.6). Specific insertion sites (10 per a target sequence) were defined and inserted in different genes by overlap extension PCR. Generated target sequences were N-terminally linked with the Fluc reporter and co-expressed with corresponding amiRNAs in protoplasts. Knock-down efficiency of $amiR^{PIN1\text{-}1}$ or $amiR^{PIN1\text{-}35}$ against the corresponding TS in various context environment (inserted within coding sequences of GR2, NSF, CDC48B or DPB) was evaluated as described above for $amiR^{Fluc}$ mini-screening. $AmiR^{Fluc}$ (against Fluc)

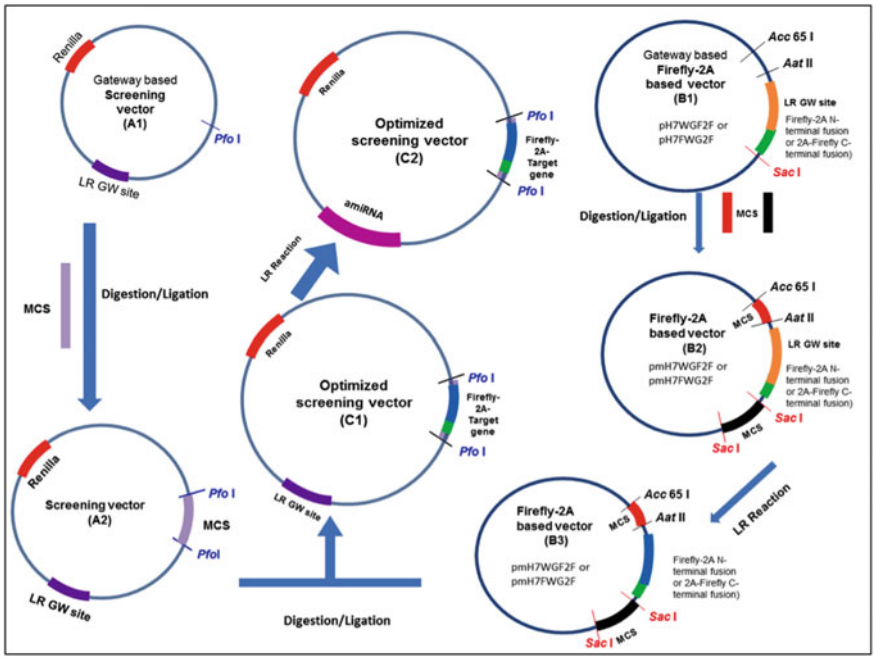

Fig. 2.5 Cloning strategy for generating luminescence-based pMIRL amiRNA screening vectors. The modified screening gateway vector (A1) was derived from previously generated pELWMS by substitution of mCherry reporter with renilla luciferase. The MCS was cloned into the vector A1 to generate intermediate vector A2. MCSs were also inserted into the gateway-based Firefly-2A fusion vectors (B1) to generate corresponding intermediated vectors B2. The vector B2 was used to transfer the target gene (in Gateway® pDONR™ vector) into vector B3. The optimized luminescence-based screening vector was generated by transferring the whole cassette from B3 into the A2 vector at MCS, resulting in vector C1. Candidate amiRNAs were cloned into C2 vector (pMIRL) by LR reaction with vector C1

was used as a positive control, while sequences missing either TS served as the negative control.

In order to evaluate effects of the sequence contexts features, a strategy using "relative response ratio" criterion to quantify amiRNA repression efficiency was developed. "Relative response ratio" was calculated as follows:

$$\text{Relative\_response\_ratio}(x) = \frac{x - PC}{NC - PC},$$

where x is the luminescence value of Fluc compared to Rluc. Average PC (positive control) and NC (negative control) values for every screening plate were used.

Comparison of the repression efficiencies of $\text{amiR}^{\text{PIN1-1}}$ and $\text{amiR}^{\text{PIN1-35}}$ for each of the tested sequences and context is summarized in Fig. 2.7. TS positions resulted in significantly different repression signals. This indicated that the TS context factors

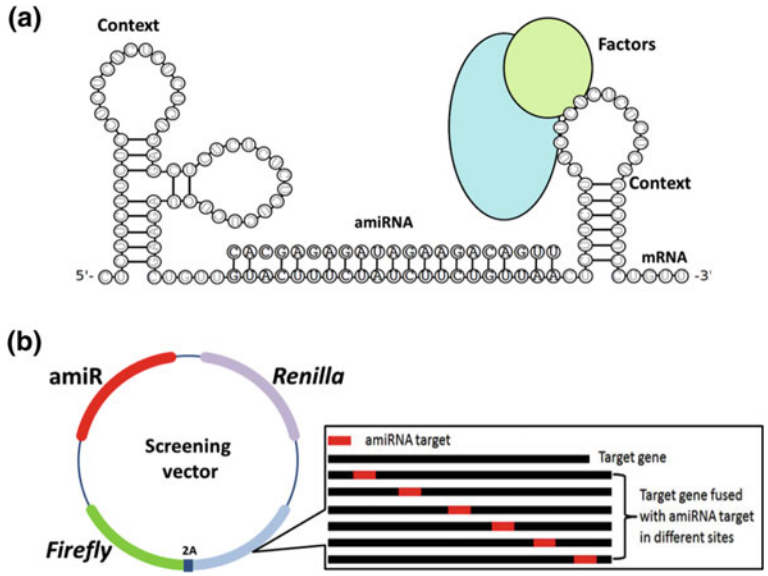

Fig. 2.6 Experimental strategy to study effect of context features on amiRNA functionality. **a** Structure of amiRNA and the target mRNA and factors (Appendix 2). **b** Vector for TS screening: it contains three independent cassete, the Renilla luciferase, amiRNA and the Firefly-2A-target gene fused with amiRNA target in different positions

influence repression efficiencies for both amiR^{PIN1-1} and amiR$^{PIN1-35}$. Although the target sites of amiR^{PIN1-1} and amiR$^{PIN1-35}$ were inserted at exactly the same locations, no correlation was observed between both repression efficiencies. Subsequent trend analysis revealed a weak, yet detectable trend for target repression efficiencies with amiR^{PIN1-1}. The repression efficiencies of the sequence contexts decreased toward the 3' end. No trend could be detected for the repression efficiencies of amiR$^{PIN1-35}$, which generally remained non-functional, although functionality was recovered in some cases. This indicates that significantly larger datasets need to be generated for identifying specific context features influencing miRNA/amiRNA functionality. However, our experimental results strongly suggest that contexts of amiRNA target sites play an important role and influence potential functionality of amiRNAs and the developed approach is optimal to fulfill evaluation requirements. To obtain informative insights about the context features determined by mRNA–RNA binding protein(RBP) interaction sites, we developed methods in plants enabling Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation (PAR-CLIP). For this we used FLAG-tagged AGO1 in the *ago1-36* mutant background to establish this methodology in Arabidopsis. Numerous individual steps (4SU treatment requirements, selection of optimal antibodies, UV-crosslinking settings, cell lysis and RNaseT1 digest and immunoprecipitation conditions, recovery of cross-linked target RNA fragments, and cDNA library preparation) were optimized.

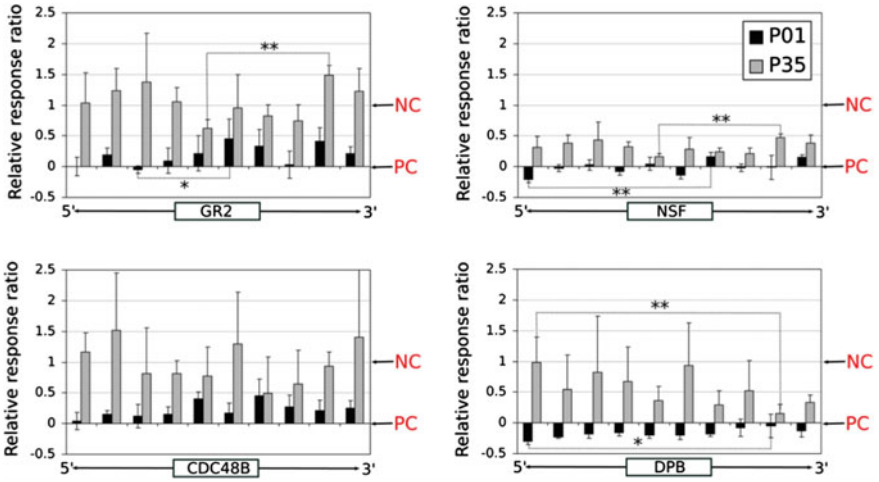

Fig. 2.7 Repression efficiencies of the context of target site locations for amiRPIN1-1(P01) and amiRPIN1-35(P35) in GR2, NSF, CDC48B, and DPB genes. The mean relative response ratio (y-axis) is plotted for each experiment where NC = 1 and PC = 0; The error bars correspond to the standard deviation of the replicates. * and ** indicate values significantly different from the control according to the Student's t test with $P \leq 0.05$ and $P \leq 0.01$, respectively

Next-generation sequencing technology was used to generate the initial dataset, bioinformatics analysis of which is currently in progress.

2.2.3 Future Work

It has been shown that miRNAs are misregulated in many diseases. However, both the cause and the functional impact of aberrant miRNA expression remain poorly understood. Our current methods analyze the functional role of miRNAs. Future projects will focus on genetic variation in this context. We will extend our approach by including single-nucleotide polymorphisms (SNPs) and copy-number variations (CNVs) into our analysis pipeline. This will allow us to identify genetic mechanisms behind patient-specific misregulation of miRNAs in diseases.

Concerning the design of amiRNAs, we have so far exploited information from CLIP data for AGO–mRNA interactions. However, a full integration of this type of information in the amiRNA design pipeline will require an extension of the models learned by GraphProt from CLIP data using the actual interaction hybrid. To do so we will incorporate the predicted interaction information materializing the duplexes and we will develop a novel graph kernel suitable for this more complex type of representations.

In addition, we need to properly take into account the interdependent and non-identically distributed nature of the RNA interaction problem setting: here the joint

presence of multiple entities such as other miRNAs and RBPs in the neighborhood of an interaction site can significantly influence the properties and efficiency of the binding event. We will therefore develop collective classification schemes that can model the joint prediction of the RBP and the miRNA target sites in a more principled way.

In order to further improve identification of features essential for amiRNA design and their application in plants, the experimental pipeline will be adapted to 384-format and liquid-handling robotics. Such automated platform prototype was developed by the groups of Dovzhenko and Palme. Active machine learning approaches will be implemented to rationalize and minimize minimal number of experimental activities. The established PAR-CLIP using AGO1 protein will be extended to other ribosome-binding proteins to complement bioinformatics approaches with the experimental data. Based on currently developed tools and approaches amiRNA technology will be transferred for agriculture-relevant applications to suppress plant pathogens, particularly RNA viruses and viral vectors. Moreover, combination of amiRNA technology with precise genome editing and plant micropropagation techniques will provide a basis for the rational design of plant with enhanced resistance against biotic and abiotic stress factors.

2.2.4 Interdisciplinary Development

The research has been down in close collaboration between the experimental and computational groups. The computational groups used available data to improve the prediction of artificial miRNAs, which then have been tested by the experimental group. The results were used to further improve the computational pipeline.

Further, interdisciplinary scientific interactions were established with the Berlin Institute for Medical Systems Biology at the Max-Delbrück-Center for Molecular Medicine, Center for Biosystems Analysis and the Freiburg Institute of Advanced Studies.

2.2.5 Educational Qualification of Researchers

Doctoral Degrees

- Sita Saunders: Computational analyses of post-transcriptional regulatory mechanisms. ALU/LBI 2014.
- Martin Preusse (thesis to be handed in April 2016; HMGU/TUM): Analysis of microRNA function using systemic regulatory features and graph models
- Fugang Ren, Development of novel technologies for functional characterization and regulation of genes activity in plants. Dissertation University of Freiburg 2015.

Diploma and Master's Degrees

- Mariam Alshaikh (ALU/LBI): A graph kernel approach to the identification and characterisation of structured noncoding RNAs using multiple sequence alignment information, 2015
- Parastou Kohvaei (ALU/LBI): Reinforcement learning techniques in RNA inverse folding, 2015
- Stefan Mautner (ALU/LBI): Learning to design RNA polymers with graph kernels, 2016
- Martin Preusse (HMGU/TUM): Beyond enrichment: Measuring microRNA-pathway associations in signaling networks, 2011

Bachelor's Degrees and Student Research Projects

- Johannes Höffler (HMGU/TUM): Gene signatures of epithelial–mesenchymal transition in embryonic stem cells, 2015

Publications Within the Project

Heyne S et al (2012) GraphClust: alignment-free structural clustering of local RNA secondary structures. Bioinformatics 28(12):i224–i232

Kowarsch A et al (2011) miTALOS: analyzing the tissue-specific regulation of signaling pathways by human and mouse microRNAs. RNA (New York, N.Y.) 17(5):809–819

Lange SJ et al (2012) Global or local? predicting secondary structure and accessibility in mRNAs. Nucleic Acids Res 40(12):5215–5226

Mendes ND et al (2012) Navigating the unexplored seascape of pre-miRNA candidates in single-genome approaches. Bioinformatics 28(23):3034–3041

Preusse M et al (2015) SimiRa: a tool to identify coregulation between microRNAs and RNA-binding proteins. RNA Biol 12(9):998–1009

Preusse M, Theis FJ, Mueller NS (2016) miTALOS v2: analyzing tissue specific microRNA function. PloS one (accepted)

Rinck A et al (2013) The human transcriptome is enriched for miRNA-binding sites located in cooperativity-permitting distance. RNA Biol 10(6)

Wend S et al (2013) A quantitative ratiometric sensor for time-resolved analysis of auxin dynamics. Sci Rep 3:2052

Other Publications

Axtell MJ, Westholm JO, Lai EC (2011) Vive la difference: biogenesis and evolution of microRNAs in plants and animals. Genome Biol 12(4):221

Dezulian T et al (2006) Identification of plant microRNA homologs. Bioinformatics 22(3):359–360

Hafner M et al (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141(1):129–141

Pudimat R, Schukat-Talamazzini EG, Backofen R (2005) A multiplefeature framework for modelling and predicting transcription factor binding sites. Bioinformatics 21(14):3082–3088

Schubert S et al (2005) Local RNA target structure influences siRNA efficacy: systematic analysis of intentionally designed binding regions. J Mol Biol 348(4):883–893
Tang W et al (2009) Faithful expression of multiple proteins via 2A-peptide self-processing: a versatile and reliable method for manipulating brain circuits. J Neurosci 29(27):8621–8629
Zhang BH et al (2005) Identification and characterization of new plant microRNAs using EST analysis. Cell Res 15(5):336–360
Zhang B et al (2006) Conservation and divergence of plant microRNA genes. Plant J 46(2):243–259

http://www.springer.com/978-3-319-54728-2

Information- and Communication Theory in Molecular Biology

Bossert, M. (Ed.)

2018, XVI, 374 p. 118 illus., 95 illus. in color., Hardcover
ISBN: 978-3-319-54728-2