

Global microRNA level regulation of EGFR-driven cell-cycle protein network in breast cancer

Stefan Uhlmann¹, Heiko Mannsperger¹, Jitao David Zhang^{1,3}, Emöke-Ágnes Horvat², Christian Schmidt¹, Moritz Küblbeck¹, Frauke Henjes¹, Aoife Ward¹, Ulrich Tschulena¹, Katharina Zweig², Ulrike Korf¹, Stefan Wiemann¹ and Özgür Sahin^{1,*}

¹ Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Heidelberg, Germany and ² Interdisciplinary Center for Scientific Computing (IWR), University of Heidelberg, Heidelberg, Germany

³ Present address: Computational Biology, Translational Research Sciences, F. Hoffmann-La Roche AG, Pharmaceutical Research and Early Development (pRED), 4070 Basel, Switzerland

* Corresponding author. Division of Molecular Genome Analysis, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 580, Heidelberg, Germany. Tel.: + 49 6221 424716; Fax: + 49 6221 423454; E-mail: sahinogur@gmail.com

Received 11.7.11; accepted 12.12.11

The EGFR-driven cell-cycle pathway has been extensively studied due to its pivotal role in breast cancer proliferation and pathogenesis. Although several studies reported regulation of individual pathway components by microRNAs (miRNAs), little is known about how miRNAs coordinate the EGFR protein network on a global miRNA (miRNome) level. Here, we combined a large-scale miRNA screening approach with a high-throughput proteomic readout and network-based data analysis to identify which miRNAs are involved, and to uncover potential regulatory patterns. Our results indicated that the regulation of proteins by miRNAs is dominated by the nucleotide matching mechanism between seed sequences of the miRNAs and 3'-UTR of target genes. Furthermore, the novel network-analysis methodology we developed implied the existence of consistent intrinsic regulatory patterns where miRNAs simultaneously co-regulate several proteins acting in the same functional module. Finally, our approach led us to identify and validate three miRNAs (miR-124, miR-147 and miR-193a-3p) as novel tumor suppressors that co-target EGFR-driven cell-cycle network proteins and inhibit cell-cycle progression and proliferation in breast cancer.

Molecular Systems Biology 8: 5; published online 14 February 2012; doi:10.1038/msb.2011.100

Subject Categories: cell cycle; molecular biology of disease

Keywords: breast cancer; EGFR signaling; microRNA; miRNA-protein interaction network; network analysis

Introduction

Signaling pathways are prime candidates for regulation by microRNAs (miRNAs) due to their dose-sensitive nature and the fine-tuning role of miRNAs (Inui *et al*, 2010). Indeed, EGFR signaling and cell-cycle components are known to be regulated by miRNAs in the cancer context (Bueno *et al*, 2008; Barker *et al*, 2010). EGFR itself is regulated by miR-7, which can induce cell-cycle arrest and cell death in cancer cell lines (Webster *et al*, 2009). In addition, let-7 regulates the oncogene Ras, which is the activator of the MAPK pathway upon EGFR activation (Johnson *et al*, 2005). Similarly, several cell-cycle genes have been identified as direct targets of miRNAs, including the miR-16 family and miR-17/20 that target Cyclin D1 and have important roles in regulating G1 arrest and S-phase entry (Liu *et al*, 2008; Yu *et al*, 2008). A systematic analysis, however, of miRNAs regulating the EGFR pathway and downstream cell-cycle proteins has not been performed.

The identification of targets of individual miRNAs is commonly achieved by combining transient transfection of miRNA mimics or inhibitors, target prediction algorithms and gene expression profiling (Wellner *et al*, 2009; Uhlmann *et al*,

2010). This approach has two main drawbacks: first, it only captures regulatory effects at the transcriptional level, which might deviate from the outputs at the protein level; second, it cannot distinguish direct and indirect effects induced by miRNAs. A more direct, biochemical approach to identify targets of miRNAs is RISC-IP, where miRNA and target mRNAs are co-immunoprecipitated with Ago2 (component of RISC) and then analyzed either by array-based hybridization (Karginov *et al*, 2007) or by deep sequencing (Hanina *et al*, 2010). This approach has identified several genes that are regulated both by mRNA degradation and by translational repression (Thomson *et al*, 2011), and some rare cases of 5'-UTR targeting (Grey *et al*, 2010).

Even with the knowledge of direct targeting, measurements of miRNA effects directly at the protein level are needed to fully comprehend the complexity of their activities. Recent advances in high-throughput proteomics have allowed researchers to begin addressing this issue. For example, two reports used quantitative mass spectrometry to assess changes in protein levels after overexpressing or knocking down selected miRNAs (Baek *et al*, 2008; Selbach *et al*, 2008). Both reached a similar conclusion: the miRNAs they studied affected the expression of hundreds of proteins, but in a rather mild

manner, fine-tuning protein synthesis. Another study used a combination of a luciferase-based 3'-UTR reporter assay and label-free shotgun proteomics to identify targets of a liver-specific miRNA, miR-122, revealing a network of miR-122-regulated genes that was enriched for proliferation, cell cycle and apoptosis, all of which are relevant to liver metabolism, liver diseases and cancer-related processes (Boutz *et al*, 2011). Researchers have also employed protein lysate arrays, which allow higher throughput compared with mass spectrometry methods, in one case screening >300 miRNAs for regulatory effects on the expression of the estrogen receptor alpha (ER α) protein (Leivonen *et al*, 2009).

Despite these advances, understanding how miRNAs regulate an oncogenic signaling system requires that we simultaneously investigate the interactions between a body of miRNAs and a network of proteins. In this work, we aimed to systematically identify miRNAs that regulate the EGFR pathway on the protein level, and to unravel principles of these regulations. To this end, we screened a genome-wide library of miRNA mimics for effects on the expression of 26 proteins in the EGFR-driven cell-cycle pathway using reverse phase protein arrays (RPPAs). This approach identified a series of miRNAs that downregulate one or more proteins of the EGFR-driven cell-cycle pathway. For most of the proteins with reduced expression levels, sequence-based prediction of targeting miRNAs can explain a significant portion of the regulation. In order to uncover the organizational principles of a mild, yet potentially very dense regulatory network, we developed a novel network-based analysis approach to identify those miRNAs that regulate multiple proteins in the studied system. Network analysis implies that the proteins controlling EGFR-driven G1/S transition are co-regulated by several miRNAs. This principle is supported by the discovery of three novel tumor-suppressor miRNAs that regulate the EGFR-driven cell-cycle pathway.

Results

miRNome screen to unravel the miRNA–protein interaction network of the EGFR pathway

To elucidate how the EGFR-driven cell-cycle protein network is regulated at the genome-wide miRNA level, we performed a gain-of-function screen using a mimic library containing 810 mature miRNAs (miRBase version 10.0). As readout, we measured and quantified the abundances of 26 proteins in this pathway. We employed the MDA-MB-231 breast cancer cell line as our experimental system, where a large body of known human miRNAs is expressed (429 detected with microarray, and 598 by next-generation sequencing data, see Supplementary information, Supplementary Tables S1 and S2 and Supplementary Figure S1 for details). Cells were transfected using an automated robotics system with individual miRNA mimics, and total protein lysates in miRNA overexpressing cells were spotted on RPPAs. Next, we incubated the protein arrays with thoroughly validated antibodies in an automated pipeline (Figure 1A).

We chose the proteins to be examined from the EGFR signaling/cell-cycle network (Table 1) based on the two criteria. First, the expression of the gene must be detectable

in the given cell line. We analyzed published RNA sequencing data (Sun *et al*, 2011) and chose those genes with at least one transcript detectable in the MDA-MB-231 cell line (Supplementary Table S3). Second, a validated antibody for the RPPAs must be available. To determine the specificity and sensitivity of the antibodies, we validated each antibody using the RNAi-based antibody validation method that we have previously published (Mannsperger *et al*, 2010). Knockdowns with siRNAs resulted in strong reductions of targeted proteins, confirming the antibody specificity/sensitivity for all proteins analyzed (Figure 1B, raw data provided in Supplementary Table S4).

We performed the miRNome screen with two biological replicates, which showed high correlations with an average Pearson's correlation coefficient of 0.78 (Figure 1C; Supplementary Figure S2). We included miR-Control (scramble), known EGFR-targeting microRNA miR-7 (Webster *et al*, 2009), si-Control (scramble) and si-EGFR in all 34 screening runs (each run involved transfection of 24 different miRNA mimics). These quality control measures assured the robustness of the screen (Supplementary Figure S3). We present the global effect of whole-genome miRNAs on the EGFR/cell-cycle network proteins, summarized as a matrix of $810 \times 26 = 21\,060$ data points, as a heatmap in Figure 1D (the screen results are provided in Supplementary Table S5). Overall, this experimental platform combining a robust whole-genome miRNA screen with quantitative proteomics allowed us to identify miRNA–protein regulations in the MDA-MB-231 breast cancer cell system.

Fine-tuning of EGFR pathway proteins by miRNAs and the resulting miRNA–protein interaction network

To build an miRNome–protein interaction network, we first identified miRNA–protein pairs where the regulation was statistically significant. For each protein, we used the z-score method to quantify the effects of all miRNAs on its expression (see Materials and methods). While a positive z-score suggests upregulation of protein expression, a negative value indicated a downregulation upon miRNA expression. For most proteins analyzed, the global effect of miRNAs followed a normal distribution with relatively short tails. This pattern suggests that the effects of miRNAs are rather mild, thereby fine-tuning the protein expression (in Figure 2A, PLCG1 is shown as an example. Histograms for all other proteins are given in Supplementary Figure S4).

To better characterize the properties of the underlying miRNA–protein regulation network, we tested different significance thresholds and chose two commonly used ones, $P < 0.05$ and $P < 0.001$ (equivalent to $|z| > 1.96$ and $|z| > 3.29$, see Materials and methods for details). We observed that networks of different significance thresholds show distinct patterns with respect to the number of remaining edges (Figure 2B). The resulting miRNA–protein regulation network is very dense at the lower threshold ($|z| > 1.96$), as can be seen in Figure 2C. This high density is not entirely unexpected, since (1) it has been computationally predicted that one miRNA can regulate several genes, (2) one gene can be

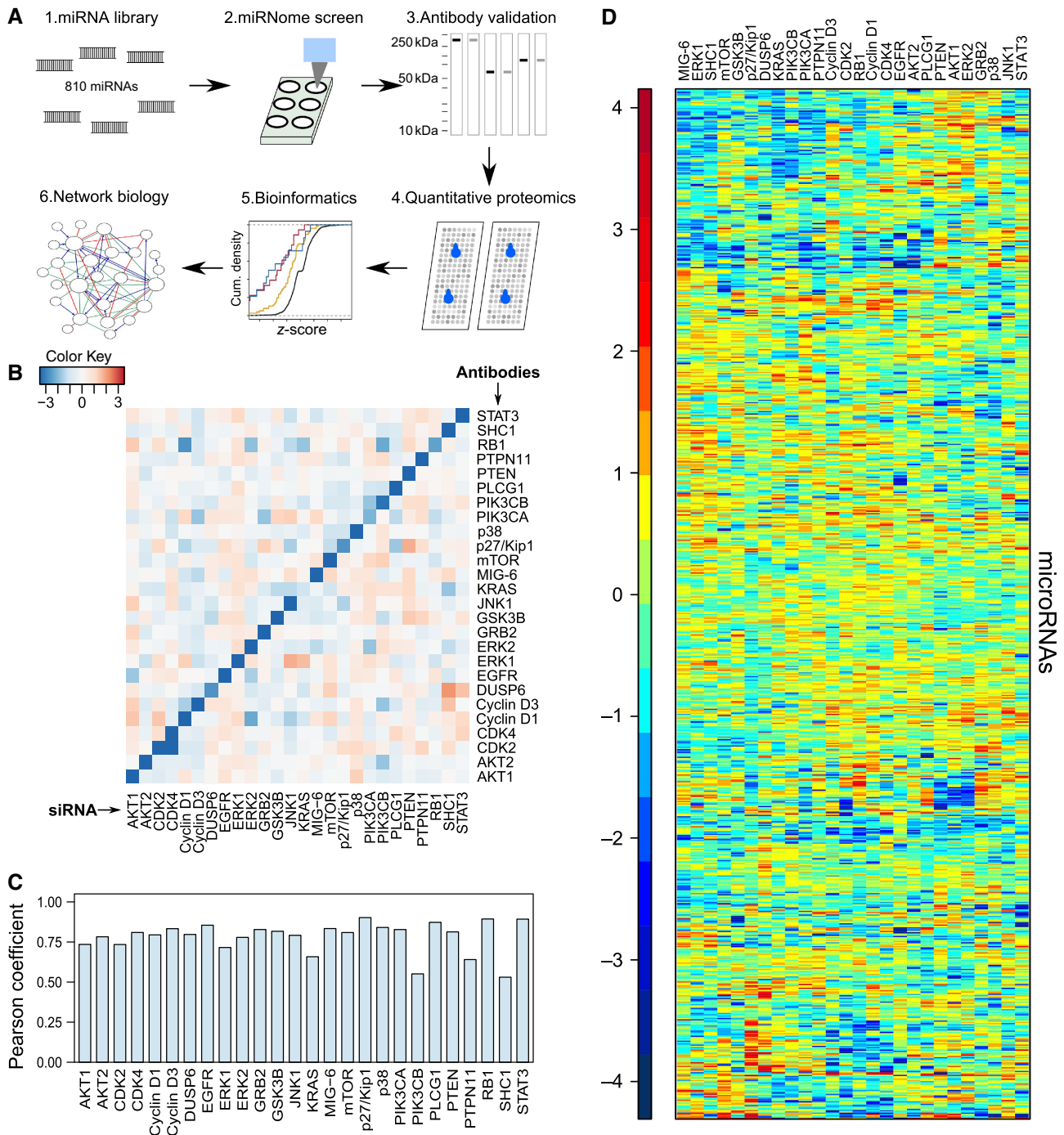


Figure 1 Whole-genome miRNA (miRNome) screen with a quantitative proteomics readout. **(A)** Flow chart of the study. MDA-MD-231 cells were transfected in 6-well format in an automated system with the miRNA library containing 810 mimics. Cells were lysed after 48 h of transfection and incubated with siRNA-validated antibodies for RPPAs. Subsequent to primary data analysis, screen results were compared with target prediction algorithms for each downregulated network protein, and network analysis of the EGFR/cell-cycle proteins at the miRNome level was done using the *more-than-random* co-regulation approach. **(B)** RNAi-based antibody validation. For each network proteins, siRNAs were used to validate the specificity/sensitivity of the antibodies before they were incubated with the lysates of the miRNA screen. The x axis of the heatmap shows siRNAs, and the y axis represents the antibodies used to quantify the abundance of proteins. **(C)** Pearson's correlation coefficients between the two biological replicates in the screen for each protein analyzed. **(D)** Heatmap showing the effect of whole-genome miRNAs on the EGFR/cell-cycle protein network. MicroRNAs are given in rows and proteins in columns. While red rectangles show upregulation, blue ones show downregulation of proteins for given miRNAs.

regulated by dozens of miRNAs and (3) indirect regulations are possible. Our data are well in line with these hypotheses, and provide the first large-scale experimental evidence, to the best of our knowledge, at the miRNome level. The number of links

decreased sharply with increasing stringency (Figure 2B) and we observed a relatively sparse graph with a low number of regulations when applying a high significance threshold ($|z| > 3.29$) (Figure 2D). These observations reveal potential

Table 1 Investigated proteins from the EGFR signaling/cell-cycle network

Gene symbol	Protein symbol	UniProtKB	Protein name
AKT1	AKT1	P31749	v-akt murine thymoma viral oncogene homolog 1
AKT2	AKT2	P31751	v-akt murine thymoma viral oncogene homolog 2
CCND1	Cyclin D1	P24385	Cyclin D1
CCND3	Cyclin D3	P30281	Cyclin D3
CDK2	CDK2	P24941	Cyclin-dependent kinase 2
CDK4	CDK4	P11802	Cyclin-dependent kinase 4
CDKN1B	p27/Kip1	P46527	Cyclin-dependent kinase inhibitor 1B
DUSP6	DUSP6	Q16828	Dual specificity phosphatase 6
EGFR	EGFR	P00533	Epidermal growth factor receptor
ERRF1	MIG-6	Q9UJM3	Mitogen-inducibile gene-6
FRAP1	mTOR	P42345	Mechanistic target of rapamycin
GRB2	GRB2	P62993	Growth factor receptor-bound protein 2
GSK3B	GSK3B	P49841	Glycogen synthase kinase 3 β
KRAS	KRAS	P01116	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
MAPK1	ERK2	P28482	Extracellular signal-regulated kinase-2
MAPK14	p38	Q16539	MAP kinase p38 α
MAPK3	ERK1	P27361	Extracellular signal-regulated kinase-1
MAPK8	JNK1	P45983	Jun N-terminal kinase
PIK3CA	PIK3CA	P42336	Phosphoinositide-3-kinase, catalytic, α polypeptide
PIK3CB	PIK3CB	P42338	Phosphoinositide-3-kinase, catalytic, β polypeptide
PLCG1	PLCG1	P19174	Phospholipase C, γ 1
PTEN	PTEN	P60484	Phosphatase and tensin homolog
PTPN11	PTPN11	Q06124	Protein tyrosine phosphatase, non-receptor type 11
RB1	RB1	P06400	Retinoblastoma 1
SHC1	SHC1	P29353	SHC (Src homology 2 domain containing) transforming protein 1
STAT3	STAT3	P40763	Signal transducer and activator of transcription 3 (acute-phase response factor)

challenges in understanding the complex network: if the significance threshold is set too low, regulatory relationships are complicated and are noisy; on the other hand, if it is set too high, too few edges will remain to reflect structural features of the network.

To overcome this dilemma, one possibility is to combine the low-stringency network with prior knowledge of miRNA-direct target predictions. In order to construct such a network, one can use pre-compiled miRNA-mRNA sequence mappings provided by the microRNA.org database (<http://www.microRNA.org/microRNA/getDownloads.do>, Release August 2010) by choosing different stringencies of the sequence-matching requirements (seed sequence mapping). We set the stringency filter so that all target predictions having a 6-mer or more seed sequence pairing are included. To construct the subnetwork, we filtered the whole network by keeping miRNA-protein edges that (1) has an absolute z -score larger than 1.96 and (2) the miRNA is mapped to gene's 3'-UTR with at least one 6-mer match or more. This approach led to 241 miRNAs and 25 genes (PIK3CB was not a predicted miRNA target), with 355 edges connecting them (Supplementary Figure S5; Supplementary Table S6). We could further reduce the complexity of the resulting network by using an evolutionary conserved sequence-matching filter. To this end, we compared the low-stringency network obtained from experimental data with miRNA-direct target predictions made by TargetScan (Friedman *et al*, 2009). We only kept those miRNA-protein links where (1) the miRNA significantly downregulates the protein with $|z| > 1.96$ and (2) the miRNA is predicted to directly target the 3'-UTR of the gene in an evolutionarily conserved manner. The resulting network, supported by both computational predictions and experimental evidence, ended up with 120

potential miRNA/protein target interactions (Figure 3; Supplementary Table S7).

In order to find out how many interactions in this network were validated by previously published independent studies, we manually searched the miWalk database (<http://www.ma.uni-heidelberg.de/apps/zmf/mirwalk/>) for predicted and validated miRNA targets. Indeed, some of the miRNA direct targeting interactions had been validated in other studies (16 validated interactions), e.g., the miR-143/KRAS (Chen *et al*, 2009), miR-200b/c/429/PLCG1 (Uhlmann *et al*, 2010) and miR-7/EGFR (Webster *et al*, 2009). However, our results identified many more new potential direct regulations. These results suggest that our experimental platform, aided by computational predictions to filter results, can indeed identify miRNA-protein interactions, even if their effects are merely moderate.

Seed/3'-UTR matches dominate the negative regulation of protein expression by miRNAs

To test whether the regulation by miRNAs is dominated by the nucleotide matching mechanism between the miRNA seed sequences and 3'-UTRs of target genes, we assessed whether protein expression differs in the presence of predicted targeting miRNAs compared with non-targeting ones for each protein in the network. For this purpose, we used two distinct target prediction algorithms (TargetScan, Friedman *et al*, 2009; and MicroCosm, Griffiths-Jones *et al*, 2008; see Bartel, 2009 for detailed discussions on the prediction algorithms) to compare predicted versus non-predicted miRNAs for 26 proteins in our network. We applied the Kolmogorov-

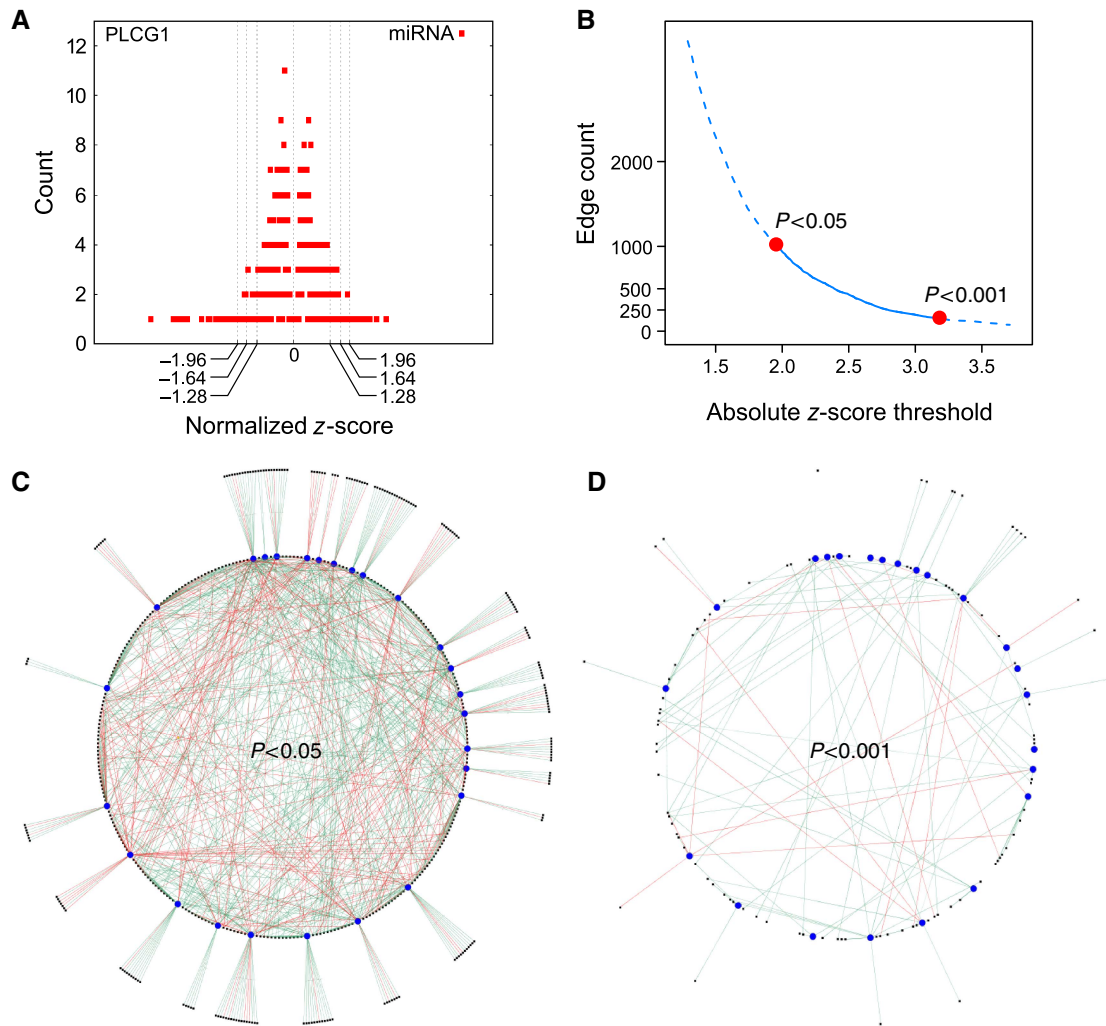


Figure 2 Whole-genome miRNA regulation of the EGFR/cell-cycle protein network. **(A)** Histogram of the effects of the whole-genome set of miRNAs on the given protein PLCG1 (for the histogram of the other proteins measured in this study, see Supplementary Figure S4). While the *x* axis demonstrates normalized *z*-scores of the expression change of PLCG1 upon miRNA transfections, *y* axis shows the frequency (count) of miRNAs. **(B)** Edge numbers for different *z*-score thresholds. With increasing stringencies of *z*-score, number of edges decreases rapidly. Two commonly used significance thresholds, $P < 0.05$ and $P < 0.001$ (equivalent to $|z| > 1.96$ and $|z| > 3.29$, respectively), were shown with red dots on the curve. **(C)** Dense miRNA–protein network at the absolute *z*-score threshold of 1.96. Blue circles represent the proteins and black circles indicate the miRNAs. While green edges between an miRNA and a protein show downregulation of that protein by miRNA, red edges show the upregulation of the protein by the given miRNA. miRNAs that regulate more than one protein are located on the circle and those miRNAs that regulate only one network protein are located outside of the circle. **(D)** miRNA regulation of the EGFR network at a more stringent threshold ($|z| > 3.29$). The resulting network is much less dense compared with the graph shown in (C).

Smirnov (K–S) test, a non-parametric, threshold-free statistical method (Smirnov, 1948), to assess whether distributions of *z*-scores of the two groups differ significantly (or three groups, in case of TargetScan with conserved and non-conserved predicted targets). For most of the proteins with predicted targeting miRNAs, we observed a significant negative aberration of expression distributions (18 proteins out of 25, with two-sided $P < 0.05$). This suggests that miRNAs, which are predicted to be direct targeting, are indeed more likely to negatively regulate the protein expression level. A primary motif to mediate this reduction was the match between the miRNA seed sequence and the 3'-UTR of the target gene (Figure 4).

This hypothesis has been supported by other studies in cases of single miRNAs at both mRNA and protein level (Baek *et al*,

2008; Selbach *et al*, 2008; Boutz *et al*, 2011). However, here we report the systematic expression shift caused by predicted miRNAs in dozens of proteins from the EGFR pathway at the miRNome level. We note that for some of the proteins, the shift caused by conserved seeds was more profound compared with the non-conserved ones. This potentially highlights the role of species conservation in miRNA targeting. Interestingly, the expression aberration was not significant for the rest of the proteins. We assume this could partially be caused by a lack of evolutionary conservation of seed-matching sequences, and consequent lack of accurate predictions. Overall, changes in the protein expression levels upon miRNA overexpression suggest that computationally predicted miRNAs are more likely to reduce expression of their targets, when compared with non-predicted miRNAs.

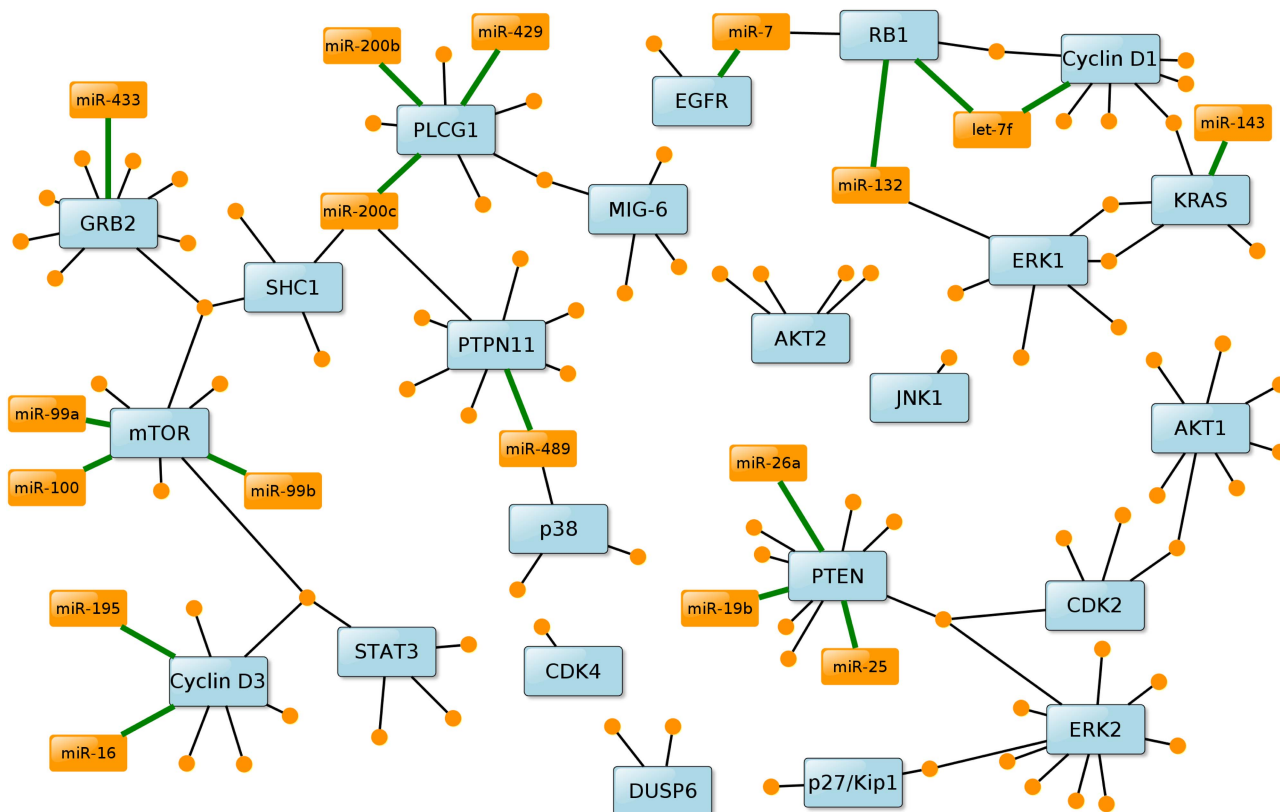


Figure 3 Regulatory network showing only the miRNAs that are both predicted to target the given gene and downregulating the protein in our screen (with absolute z-score > 1.96). This interaction is shown with a black edge between the miRNA and protein. The validated miRNA/protein interactions in other independent studies are shown with a green edge between the miRNA and the protein. The yellow circles indicate the miRNAs, which are identified to regulate EGFR-driven cell-cycle network in this study. The full list of miRNAs and proteins is given in Supplementary Table S7.

MiRNA co-regulation of proteins in the same functional modules

Having an integrated miRNA–protein regulation network is helpful to identify new miRNA targets. However, it does not address whether the regulation of protein expression by miRNAs follows any other rules besides the seed-matching mechanism. Moreover, challenges emerge in maximizing the signal-to-noise ratio for discoveries from the regulatory network, calling for new strategies to identify patterns of miRNAs regulating the EGFR-driven cell-cycle pathway.

In order to explicitly address these questions, we developed a novel framework in the context of graph theory and complex network analysis. First, we evaluated whether any two nodes (here: proteins) on one side of a given bipartite network would share a significant number of neighbors (here: miRNAs) on the other side. Based on the z-scores from the RPPA data, a *bipartite graph* containing all proteins and their significantly regulating miRNAs was built. This bipartite graph allowed for discovering statistically significant relationship patterns. Our focus was to find pairs of proteins that were either consistently co-upregulated or co-downregulated by a group of miRNAs. Furthermore, we were interested in two proteins where one was consistently upregulated and the other was downregulated by one or more miRNAs. We define any of these

three co-regulation patterns between proteins A and B to be *intrinsic*, if the observed pattern occurred statistically significantly more often than in a random model of the same bipartite graph (the *more-than-random* model). In essence, our method employed a bootstrapping approach in which the data were permuted by maintaining some of the properties (Figure 5A, see Materials and methods for details).

To this end and in order to detect the principles of miRNA regulations as well as potential co-regulation patterns, we built a *consensus graph*. This graph combined the information from different bipartite graphs by varying parameters in the network model (Figure 5B, see Materials and methods for details). Two proteins are connected if and only if they were significantly co-regulated by one or more miRNAs under all of the chosen parameter values. The upper bound FDR of the *consensus graph* was $q < 0.132$ determined with the Benjamini-Hochberg method (see Supplementary information and Supplementary Figure S6 for details).

We observed that most edges between cell-cycle proteins were present in the *consensus graph*, indicating that their potential co-regulation by given miRNAs is robust. Furthermore, the miRNAs co-regulating the cell-cycle proteins showed an interesting pattern that is biologically relevant: they upregulated the expression of the CDK inhibitor protein p27/Kip1, while downregulating the Cyclin-dependent kinase-4

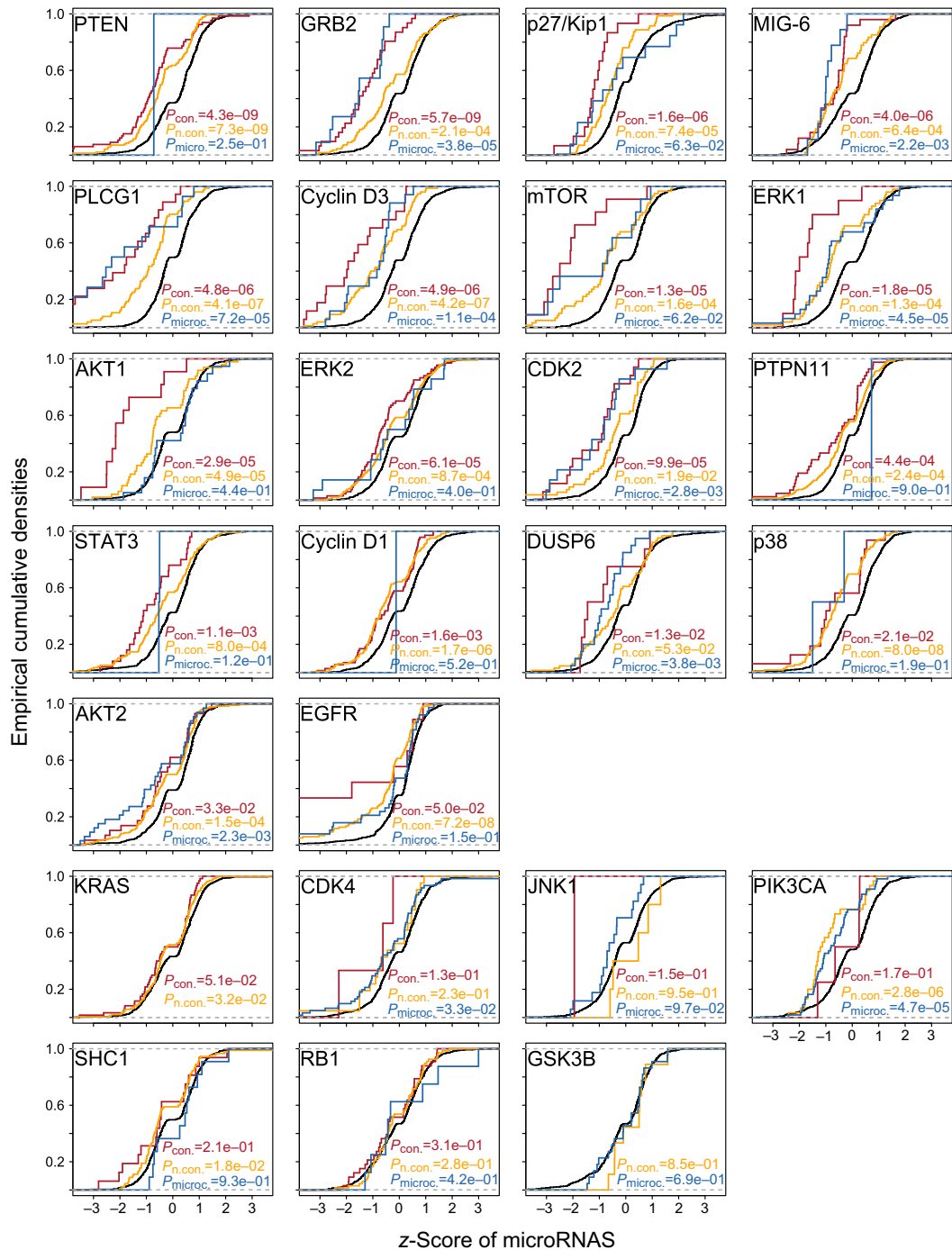


Figure 4 The match between miRNA seed sequences and 3'-UTR of the genes explains the reduction in the expression of most of the proteins in the EGFR/cell-cycle network. The Kolmogorov-Smirnov (K-S) test was used to assess whether distributions of z-scores of two groups (predicted versus non-predicted miRNAs) differ significantly. For most of the proteins with predicted targeting miRNAs, a significant negative aberration of expression distributions was observed (18 proteins out of 25, with two-sided $P < 0.05$). Black lines demonstrate the z-score distribution of the non-predicted miRNAs against that of predicted ones. Two target prediction algorithms were used shown in red (TargetScan, conserved), yellow (TargetScan, non-conserved) and blue (MicroCosm) lines.

(CDK4) and RB1 which would, in turn, inhibit cell-cycle progression (Figure 5B). Of note, PIK3CA, AKT2 and PLCG1 were co-downregulated with retinoblastoma protein RB1 (for PIK3CA and AKT2) and CDK4 (for PLCG1) by dozens of miRNAs. This observation might be linked to the key roles

these proteins likely have in the regulation of the G1/S transition in the MDA-MB-231 cell line model. Overall, these results indicate that proteins, which fall into the same known functional modules (here: EGFR-driven cell-cycle module), are co-regulated by miRNAs.

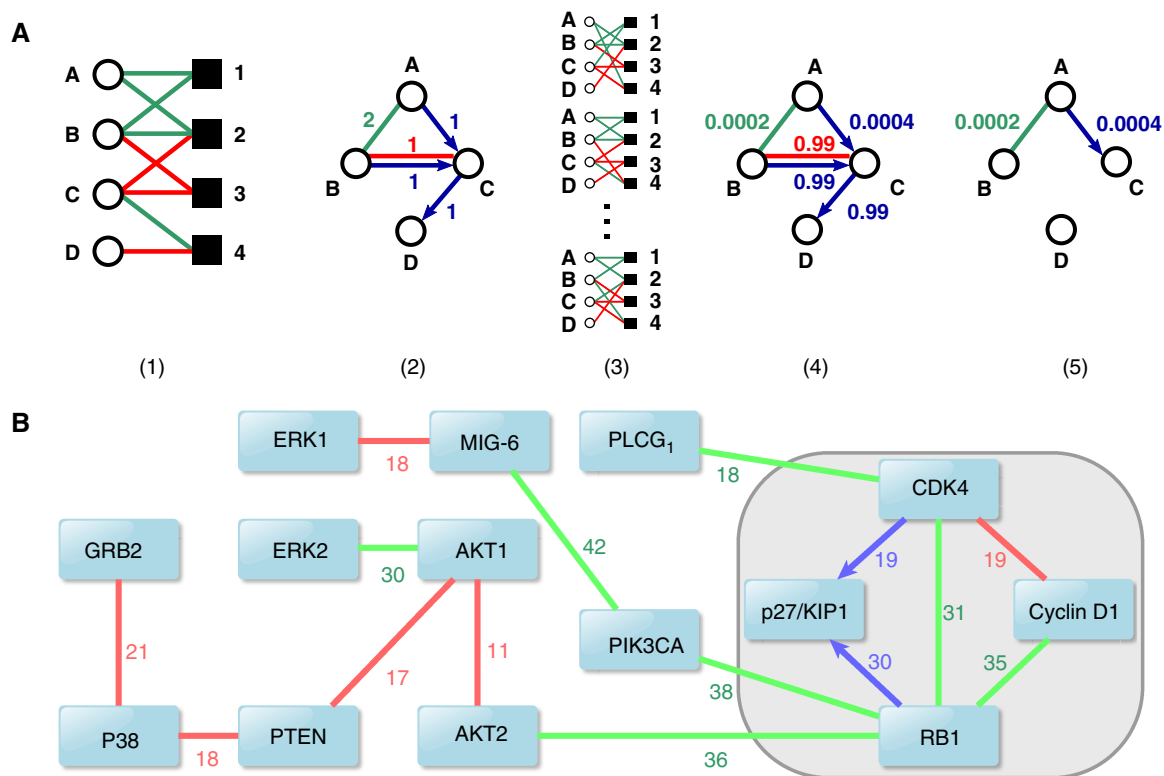


Figure 5 Co-regulation of EGFR/cell-cycle network proteins by whole-genome miRNAs. **(A)** Principles of the 'more-than-random co-regulation' approach. (1) Bipartite network consisting of proteins A, B, C, D and miRNAs 1, 2, 3, 4. Green edge between an miRNA and a protein indicates that expression of the protein is reduced by the miRNA whereas a red edge shows that expression of the protein is increased by the miRNA. (2) Protein co-regulation network where the numbers on the edges indicate the number of miRNAs which co-regulate the given protein pairs. Protein pairs A–B was co-downregulated two times, while protein pairs B–C were co-upregulated once. C–D, B–C, and A–C proteins were inversely regulated once by miRNAs, which is shown with a blue directed edge. (3) The significant co-regulations can be identified by the 'more-than-random' model. The model compares for each protein pair the number of miRNAs that co-regulate them in the experimentally observed data with the number of miRNAs co-regulating them in the random bipartite network graphs. (4) The statistical-significance test was performed by generating 10 000 randomized networks and computing for each protein pair the fraction of randomized networks in which their co-regulation appeared at least as often as in the observed bipartite network. The labels on the protein co-regulation network indicate the obtained *P*-values. (5) In a thresholding step, a co-regulation with $P < 0.05$ was considered statistically significant. **(B)** Consensus co-regulation graph of EGFR/cell-cycle proteins by miRNAs. Edges shown in the graph are obtained from the most stringent consensus graph. The green edges between the proteins show the co-downregulation of two proteins while the red edges show the co-upregulation. The directed blue edges indicate that the node where the edge originates from is downregulated whereas the node where the edge is directed to is upregulated. The numbers on the edges indicate the number of miRNAs having such regulation patterns in the most stringent consensus graph. Cell-cycle proteins that are regulated by miRNAs in a coordinated manner are marked by the gray box.

MiRNAs 124, 147 and 193a-3p inhibit G1/S transition and cell proliferation by targeting EGFR signaling and/or cell-cycle proteins

The network-analysis approach has the power to predict the outcome of overexpressing individual miRNAs in the context of the EGFR-driven cell cycle. For example, miR-124, 147 and 193a-3p were among the miRNAs with highest degrees in the consensus graph (Figure 6A, source data in Supplementary Table S8). Furthermore, the expression of several key proteins controlling G1/S transition was regulated in a tightly coordinated manner by these miRNAs (Figure 6B). Therefore, our analysis predicted that they should, when overexpressed, alter cell-cycle progression patterns by inhibiting the G1/S transition. We performed molecular and cellular assays to confirm the validity of these predictions.

First, we validated all regulations shown in Figure 6B for all three miRNAs with western blots (Figure 6C). We also validated the overexpression of miR-124, miR-147 and

miR-193a-3p after transfection with miRNA mimics by qRT-PCR (Supplementary Figure S7). All three miRNAs co-regulated cell-cycle components either by reducing RB1 level or by increasing p27/Kip1 level or affecting both (Figure 6B). For miR-124, we observed a core regulation of seven EGFR protein network components. The PTEN tumor suppressor had the highest number of connections with other proteins and was co-upregulated with p27/Kip1. In case of miR-147, AKT2 was co-downregulated with CDK4 and RB1, indicating the potential impact of AKT2 on cell-cycle progression. Finally, for miR-193a-3p, PIK3CA was co-downregulated with CDK4 and Cyclin D1 as well as with MIG-6.

We then tested the prediction that all three miRNAs inhibit EGFR-driven G1/S transition by targeting several components simultaneously at the phenotypic level. By staining cells with BrdU (for S-phase fraction) and 7-AAD (for total DNA content), we could show that all three miRNAs induced a very strong G1 arrest in the breast cancer cell line MDA-MB-231 (Figure 7A). Furthermore, viability of the cells was

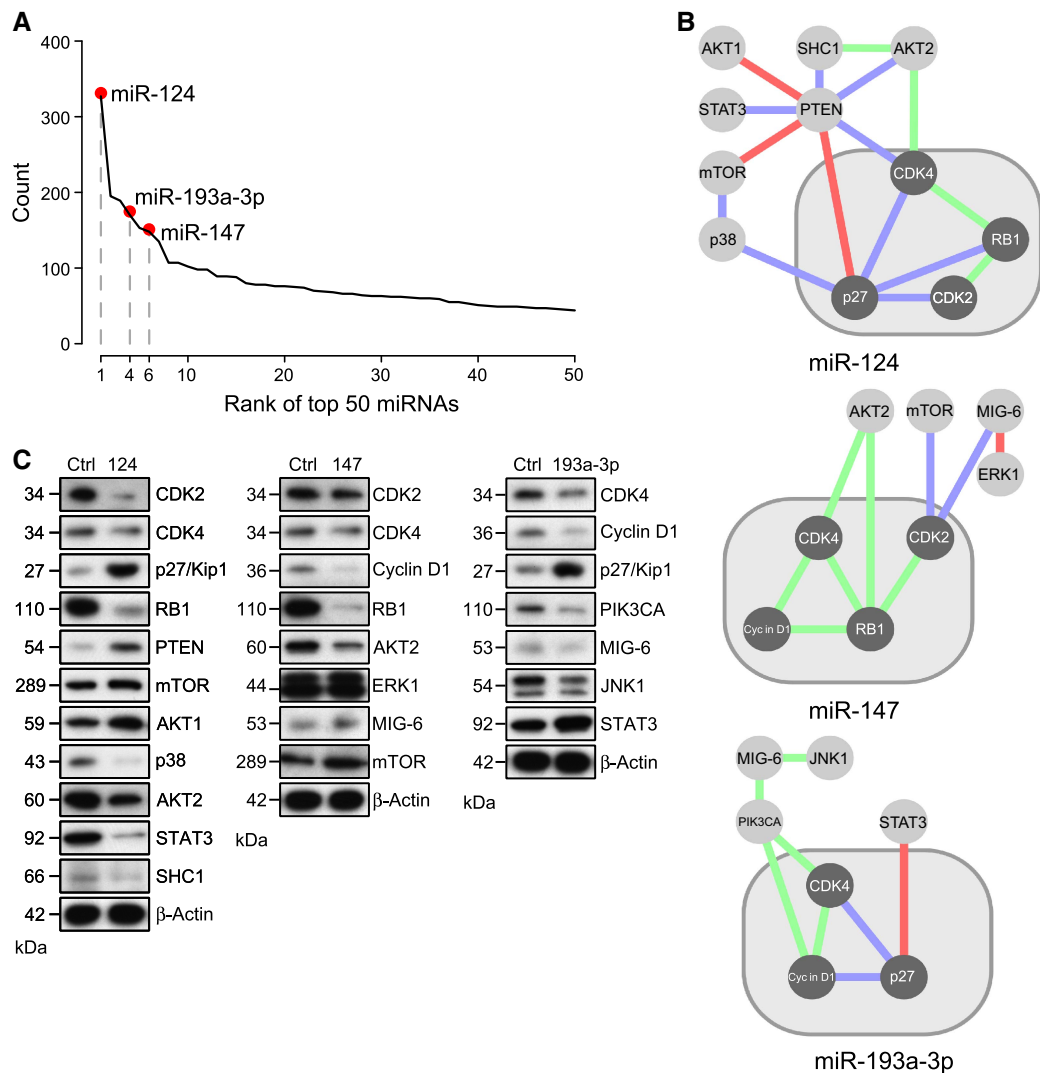


Figure 6 Regulatory graphs for miR-124, miR-147 and miR-193a-3p and Western blot validation of RPPA results. **(A)** Rank of miRNAs based on their frequency in the *consensus graphs*. Only the first top 50 miRNAs are drawn. The miRNAs (miR-124, miR-147 and miR-193a-3p) that were chosen for experimental validations are shown with red dots. **(B)** Co-regulation of EGFR/cell-cycle proteins is shown for three miRNAs, which were among the miRNAs having the highest number of *more-than-random* regulations. Cell-cycle proteins are drawn as dark gray nodes and shown in a light gray background. Color codes of the edges are as in Figure 5B. **(C)** Western blot validation of the effects of these three miRNAs on the protein expression the EGFR/cell-cycle components. β -Actin was used as loading control.

reduced after 72 and 96 h upon individual overexpression of these miRNAs (Figure 7B).

In order to test whether or not the regulation of proteins by miRNAs resulted from direct targeting, we cloned the 3'-UTRs of the downregulated genes into a luciferase reporter vector, and performed luciferase reporter assays. For genes with long and difficult to clone 3'-UTRs, we fragmented the 3'-UTR into two or three parts (denoted as P1, P2 and P3). AKT2, p38 MAPK and STAT3 were predicted to be conserved targets of miR-124 (Supplementary Figure S8A), and were also identified as potential direct targets in our study (Figure 7C). Although CDK4, CDK2 and SHC1 were poorly conserved predicted targets of miR-124 (Supplementary Figure S8A), they were not identified as direct targets. RB1 was neither predicted nor found to be a direct target. These data demonstrate that the inhibitory effect of miR-124 on cell-cycle progression could be,

at least in part, *via* targeting EGFR core proteins that were analyzed in this study. In the case of miR-147, AKT2 and Cyclin D1 were predicted to be poorly conserved direct targets (Supplementary Figure S8B), and identified here as direct targets. These data suggest that targeting of AKT2 and Cyclin D1 by miR-147 contributes to its strong inhibitory effect on G1/S transition (Figure 7C). Finally, JNK1 was identified as a direct target of miR-193a-3p, and could potentially contribute to the strong inhibitory effect miR-193a-3p had on cell-cycle progression (Figure 7C).

We further tested the effects of miRNAs on mRNA expression of their targets to examine if the targeting was mediated *via* mRNA cleavage or translational repression (Figure 7D). We found that three of six miRNA/target interactions (miR-124/AKT2, miR-147/AKT2 and miR-193a-3p/JNK1) reduced protein expression of their target genes

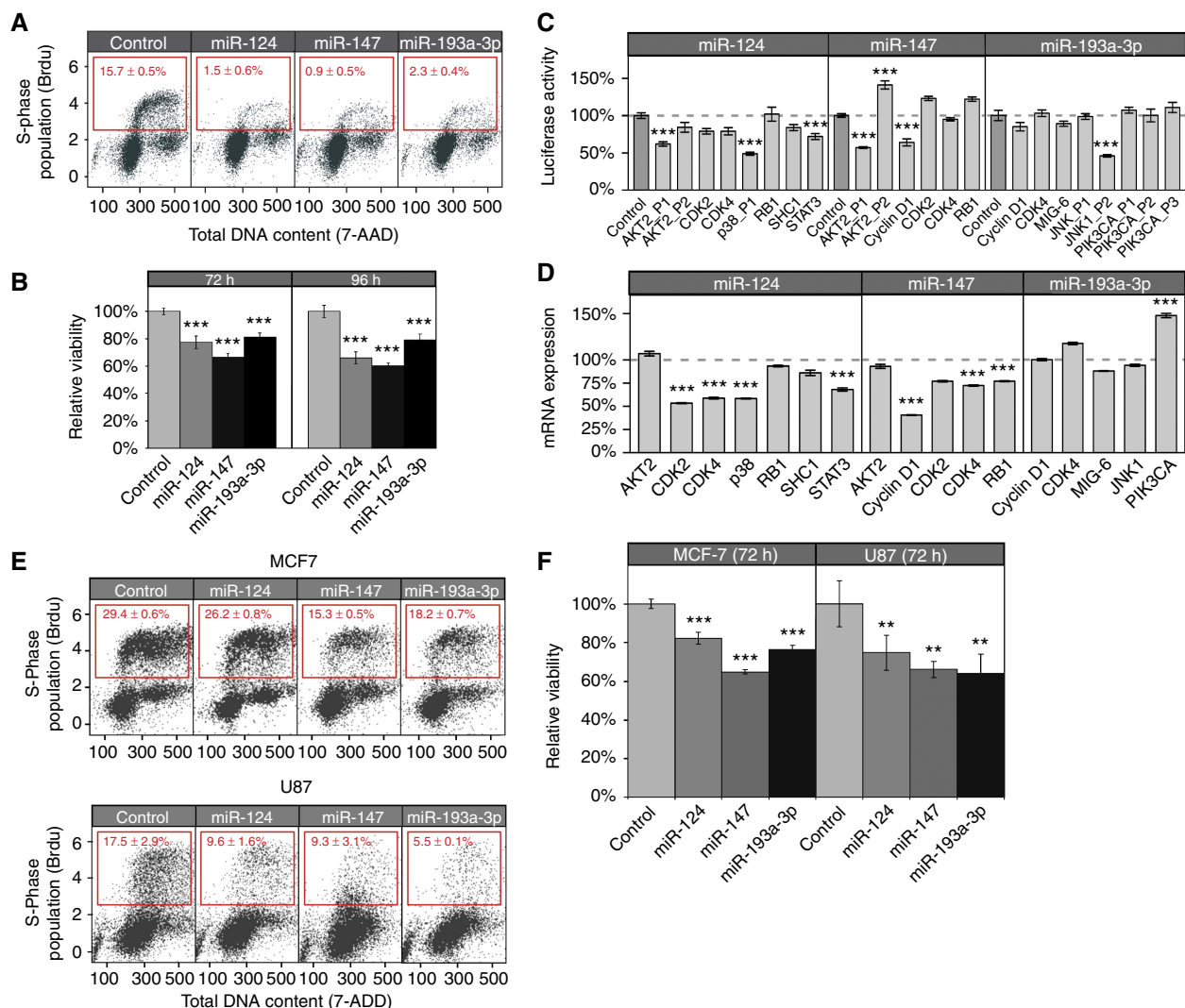


Figure 7 miR-124, miR-147 and miR-193a-3p are identified as novel tumor suppressors in breast cancer and glioblastoma cell lines by targeting EGFR/cell-cycle proteins. **(A)** The effects of miRNAs on the cell-cycle progression. After overexpressing the miRNAs in MDA-MB-231 cells, total DNA content (7-AAD) and S-phase population (BrdU) were determined by FACS after 48 h. The red rectangles show the percentage of proliferating cells in S-phase. The results are shown as an average of three biological and three technical replicates. **(B)** The effects of miRNAs on the viability of the cells. After overexpressing the miRNAs in MDA-MB-231 cells, viable cells were counted by a luciferase-based viability assay after 72 and 96 h. The results are shown as average of two biological and five technical replicates. **(C)** Validation of direct targeting of genes by miRNAs. A luciferase reporter assay was performed after co-transfecting MDA-MB-231 cells with the luciferase vector harboring the 3'-UTR of the gene of interest and miRNA mimics. Luciferase activities were measured after 48 h of transfection, and results are presented as average of four biological and three technical replicates. P1, P2 and P3 denote the parts of 3'-UTRs analyzed for the genes with very long or difficult to clone 3'-UTRs. **(D)** Quantification of effects of miR-124, miR-147 and miR-193a-3p on the mRNA level by qRT-PCR. The results are presented as average of two biological and three technical replicates. **(E)** The effects of miRNAs on the cell-cycle progression of MCF-7 and U87 cell lines as described in (A). **(F)** The effects of miRNAs on the viability of MCF-7 and U87 cell lines as described in (B). Two asterisks (**) denote the *P*-value < 0.01, three asterisks (***) denote the *P*-value < 0.001.

without significant changes at the mRNA level. These data consolidate the concept that the analysis of miRNA effects at the protein level is more informative and comprehensive than that only at the mRNA level.

Finally, in order to rule out that the observed results are cell line specific, but rather can be generalized in breast and other cancer types, we validated our results in another breast cancer cell line, MCF-7, as well as in the U87 glioblastoma cell line. The latter represents another tumor type where EGFR signaling has a critical role in the pathogenesis (Hatanpaa *et al*, 2010). Similarly to the results we had obtained in the MDA-MB-231 cell line, all three miRNAs (miR-124, miR-147

and miR-193a-3p) led to a reduction of the cell population in S-phase and also in reduced cell viability (Figure 7E and F). Furthermore, we recapitulated all the direct targeting relationships in these two cell lines (Supplementary Figure S9A). The only difference which we observed was that targeting of AKT2 by miR-124 or miR-147 was *via* translational repression in MDA-MB-231 cell line while it was *via* mRNA cleavage in MCF-7 (in case of miR-147) and in U87 (in case of both miR-124 and miR-147) (Supplementary Figure S9B).

Overall, our results indicate that the three miRNAs predicted to interfere with the cell-cycle subnetwork indeed block the G1/S transition of cancer cell lines by regulating multiple

proteins of the EGFR/cell-cycle network simultaneously. Therefore, we propose these miRNAs as potential tumor suppressors in breast cancer. Furthermore, their discovery underscores the value of the platform, which combines high-throughput proteomics and a network-analysis framework, in identifying regulatory patterns in a complex biological network.

Discussion

In this first proteomics screen investigating a mammalian oncogenic signaling network at the miRNome level, we systematically identified miRNAs that regulate the expression of 26 proteins in the EGFR-driven cell-cycle pathway, and developed a novel framework to unravel principles of these regulations. Our results indicate that a robust combination of miRNA screening with a proteomic readout and network-analysis tools helps to identify miRNAs regulating a given protein network and to discover (co-)regulation patterns of the proteins.

Our study led us to three key conclusions: (1) whole-genome miRNA effects on a given protein are generally mild supporting a fine-tuning role for miRNAs, and these effects are dominated by sequence-matching mechanisms. (2) A new network-analysis methodology identified co-regulated protein targets. Our results suggest that miRNAs simultaneously regulate several proteins belonging to the same functional module. (3) Three miRNAs (miR-124, miR-147 and miR-193a-3p) act as novel tumor suppressors by co-regulating EGFR-driven cell-cycle network proteins, thus qualifying as potential targets for breast cancer therapy.

As *in-silico* predictions for miRNA targets rely on Watson-Crick base pairing between the miRNA seed region and the 3'-UTR of target genes, we tested this hypothesis at the genome-wide miRNA level for 25 of the 26 proteins analyzed in this study. PIK3CB had to be excluded from this analysis as its 3'-UTR was not represented in the target prediction databases used. Our data showed a systematic expression shift caused by predicted miRNAs in dozens of proteins from the EGFR pathway. These observations are very much in line with two previous studies which examined the effects of a few miRNAs on the proteome (up to 5000 proteins) using mass spectrometry-based approaches (Baek *et al*, 2008; Selbach *et al*, 2008).

Compared with these two studies, we scaled up the tested miRNAs to the whole-genome level while scaling down the number of proteins to a single pathway. This trade-off between the number of miRNAs and proteins tested was indispensable due to the methodologies chosen and the biological questions addressed. Protein arrays have the capacity of examining thousands of samples, but are limited by antibody availability (Wilson *et al*, 2010). On the other hand, mass spectrometry can analyze only a limited number of samples, but has the capacity of analyzing thousands of proteins (Choudhary and Mann, 2010). Undoubtedly, our approach complements these two studies and suggests that the 3'-UTR-seed match contributes to the targeting of not only a few miRNAs, but for the whole-genome set of miRNAs. Further development in high-throughput technologies will eventually generate a global

picture by analyzing miRNome versus proteome in the future. For certain proteins, we did not observe a significant difference between expression changes caused by predicted miRNAs compared with those by non-predicted ones. Besides a possible lack of evolutionary conservation of seed-matching sequences other factors, e.g., shortened 3'-UTRs due to alternative polyadenylation (Fu *et al*, 2011), could contribute to this observation in the analyzed cell system. This raises the necessity for further investigations.

We next asked whether miRNA targets could also be identified by combining expression correlation analysis with computational target predictions and without performing a large-scale screen. Therefore, we tested if the miRNA-protein interaction network we obtained (Figure 3) could as well be simply inferred by performing computational target prediction based on expression correlation analyses. For this purpose, we employed two publicly available NCI60 cancer cell line panel data sets at miRNA-mRNA (Liu *et al*, 2010) and miRNA-protein levels (Shankavaram *et al*, 2007). *In-silico* prediction was indeed able to infer the potential regulatory relationships for a very small proportion of miRNA-gene or miRNA-protein pairs that we had identified with the proteomics screen approach (see Supplementary Table S9, Supplementary Figure S10 and Supplementary information for details). However, this method missed the major part of the miRNA-protein network identified in the screen. Therefore, we believe that correlation analysis combined with predictions might be useful as filters to prioritize miRNA-gene pairs for experimental validation. However, due to the high false-positive and false-negative rates, these computational methods cannot compensate for wet-lab proteomics screening.

Loss-of-function studies of miRNA targets (e.g., by siRNAs) partially mimics, in most cases, the molecular and phenotypic observations induced by miRNA expression (Uhlmann *et al*, 2010). This is likely to be caused by a very complex regulation of several targets by a single miRNA or simultaneous targeting of a protein by dozens of miRNAs. Furthermore, identifying targets of miRNAs with transcript-based array profiling is a useful approach that, however, has considerable drawbacks. As miRNAs might lead to translational inhibition without affecting the mRNA level, transcript-level approaches can miss certain targets. This latter assumption was proven in our study as three out of six examined miRNA/target pairs downregulated their targets only at the protein level without affecting the respective mRNA levels. Thus, it would have not been possible to identify these targets with an mRNA-based approach.

To circumvent these two problems, i.e., complexity of miRNA regulation and limitations of mRNA level study, and to get more insights into the miRNA biology, we devised a protein network level study. To this end, we combined a proteomic approach with a novel network-analysis framework to identify those proteins that are co-regulated by miRNAs in an intrinsic way. This approach helped us to filter noise from screen data while maintaining the ability to detect structural properties, i.e., co-regulated proteins, from a network containing mild and noisy relationships. In this way, we identified proteins that were significantly co-regulated by miRNAs. Several cell-cycle components that regulate the G1/S transition check point were found to be co-regulated. For example, while proteins driving

the G1/S transition (Cyclin D1 and CDK4) were co-down-regulated with RB1 by dozens of miRNAs (35 and 31 miRNAs, respectively), Cyclin D1 and CDK4 were co-upregulated by 19 miRNAs. Furthermore, RB1 and CDK4 were inversely co-regulated with p27/Kip1 CDK inhibitor (30 and 19 miRNAs, respectively). These data suggest that miRNAs might rather have tumor-suppressor than oncogenic functions. We also observed that AKT1 and ERK2 were co-downregulated by 30 miRNAs. These miRNAs may potentially lead to the simultaneous inhibition of two key pathways regulating cell survival and proliferation, highlighting their potential to be used as drugs themselves.

Based on the analysis framework, we selected three miRNAs, miR-124, miR-147 and miR-193a-3p, for their roles in the G1/S transition. Inhibitory effects of these three miRNAs on cell proliferation were potentially mediated by targeting EGFR network proteins (e.g., AKT2, STAT3, p38 and JNK1) as well as Cyclin D1, a key G1/S transition regulating protein. Supporting the tumor-suppressor function of miR-124, it was reported that its expression is reduced in glioblastoma and its ectopic expression inhibits tumor migration and invasion (Fowler *et al*, 2011). Furthermore, miR-124 was found to be epigenetically silenced in acute lymphoblastic leukemia (ALL; Agirre *et al*, 2009) and to regulate the CDK6/RB pathway by targeting CDK6 post-transcriptionally in ALL as well as in medullablastoma (Pierson *et al*, 2008). Similarly, miR-193 was epigenetically silenced in acute myeloid leukemia (AML) and targets the c-Kit oncogene leading to apoptosis (Gao *et al*, 2011). The role of miR-147 was reported to be in feedback regulation of inflammation where toll-like receptors (TLRs) induce miR-147 expression, which involves binding of NF- κ B and STAT1 α to the miR-147 promoter. This induction, in turn, inhibits cytokine expression and prevents excessive inflammatory responses (Liu *et al*, 2009).

To demonstrate that the effects of miR-124, miR-147 and miR-193a-3p on EGFR signaling have a more general nature, we tested the miRNA-target interactions and tumor-suppressor functions in another breast cancer as well as in a glioblastoma cell line. The results potentially support the principal tumor-suppressor function of these miRNAs by targeting key components of the EGFR/cell-cycle pathway. To further confirm the miRNA-target interactions and tumor-suppressor roles of these miRNAs, large breast cancer or other cancer type collections have to be analyzed using miRNA/mRNA profiling, deep sequencing and/or immunohistochemistry (IHC) or protein array-based methods. The published data together with our results for breast cancer and glioblastoma cell lines indicate that these three miRNAs might have tumor-suppressor functions by regulating different processes, e.g., cell-cycle, invasion, and inflammation related to cancer, and could potentially be targets for cancer therapy.

In summary, we identified miRNAs that regulate the EGFR-driven cell-cycle pathway. Starting from experimental results and by combining computational analysis and a network-based analytical framework, we could show a parsimonious picture of how miRNAs regulate the protein network: by sequence matching, and by intrinsic co-regulation of proteins belonging to same functional module. Our results provide the research community with a unique resource to identify novel miRNA-target interactions. This paves the way toward

studying miRNA–protein interactions at a network level to elucidate the complex regulation of biological processes by miRNAs, which assure further detailed studies in future.

Materials and methods

Cell culture

The human breast cancer cell lines MDA-MB-231 and MCF-7 and the glioblastoma cell line U87 were obtained from ATCC (Manassas, VA, USA). Culturing media and supplements for the MDA-MB-231 and MCF-7 cells were described previously (Sahin *et al*, 2009). U87 was cultured in DMEM (Invitrogen, CA, USA) supplemented with 10% FBS and 1% Pen/Strep. Contamination tests of the cell lines were performed by the DKFZ Core Facility.

Screening with the miRNA mimic library

Cells were seeded in 6-well plates and were transfected with the miRIDIAN miRNA mimic library (Dharmacon, CO, USA), which comprises 810 miRNA mimics, using Lipofectamine 2000 (Invitrogen, CA, USA) with help of a Biomek FX pipetting robot (Beckman Coulter, CA, USA). Each miRNA mimic was transfected at a final concentration of 25 nM in two biological replicates. The miRNome screen was accomplished in 34 screening runs and each run included 24 miRNA mimics, two positive controls (miR-7 and si-EGFR; Dharmacon), three negative controls (miR-Control-1 and miR-Control-2; Dharmacon) and siRNA control siAllstar (Qiagen, Hilden, Germany). Forty-eight hours after the transfection, the cells were harvested and further used for RPPAs.

Reverse phase protein arrays

Lysates were prepared as previously described (Sahin *et al*, 2009). Two biological replicates and three technical replicates were spotted for each miRNA mimic on nitrocellulose-coated glass slides (Grace Biolabs, OR, USA) using an Aushon 2470 contact printer (Aushon, MA, USA). Target protein-specific primary antibodies (Supplementary Table S10) were validated using siRNAs (siRNA sequences are given in Supplementary Table S11) as previously described (Mannsperger *et al*, 2010) and incubated at a dilution of 1:300. After incubation with Alexa680-labeled secondary antibodies (Invitrogen, Karlsruhe, Germany), signals were visualized using an Odyssey scanner (LI-COR, NE, USA). Signal intensities were quantified using Gene Pix Pro 5.0 (Molecular Device, CA, USA) and background corrected. Data pre-processing and quality control was performed with the custom R package caUtils.

Immunoblot analysis

Preparation of protein lysates and western blotting was previously described (Sahin *et al*, 2009). In all, 10 μ g of protein was separated by SDS-PAGE and exposed to primary antibodies (Supplementary Table S10). HRP-conjugated secondary antibodies were purchased from Santa Cruz Biotechnology (Santa Cruz, CA, USA).

Plasmid constructs and luciferase reporter assay

The 3'-UTRs of potential miRNA target genes were amplified by PCR using human genomic or complementary DNA (primer sequences were given in Supplementary Table S12) of MDA-MB-231 cell line, and cloned into the psiCHECK2 vector downstream of the luciferase ORF using the *SgfI* or *XhoI* and *NotI* sites depending on the gene. For the luciferase reporter assays, cells were co-transfected with 25 nM of miRNAs and 15 ng of the luciferase vector. *Firefly* and *Renilla* luciferase activities were measured after 48 h using a luminometer (Tecan, Männedorf, Switzerland), according to the manufacturer's instructions. Relative luciferase activity was determined by the ratio of *Renilla* luciferase signal intensity to that of *firefly* for normalization. The

average and standard deviation of the ratio were estimated by the *Delta-method* (Bioconductor *ratioAssay*).

Cell-cycle analysis and viability assay

MDA-MB-231, MCF-7 or U87 cells were transfected with miRNA mimics (miR-124, miR-147 and miR-193a-3p) and control miRNA. Forty-eight hours after transfection, 7-AAD/BrdU cell-cycle assay was done according to the manufacturer's protocol (BD Biosciences, CA, USA). Stained cells were measured by flow cytometry (FACSCalibur, BD Bioscience, Heidelberg, Germany) using Cell Quest Pro software (BD Bioscience, Heidelberg, Germany). Cell viability was measured using Cell Titer Glo Luminescent Cell Viability assay (Promega, WI, USA) following the manufacturer's instructions. Transfections were carried out in a 96-well plate (8×10^3 cells per well) in a final miRNA mimic concentration of 25 nM per well with five replicates.

Quantitative RT-PCR of mRNAs

Total RNA was isolated using the RNeasy Mini Kit (Qiagen). cDNA was synthesized from 10 ng of RNA using the Revert Aid H Minus First Strand cDNA Synthesis Kit (Fermentas, St Leon-Rot, Germany) according to the manufacturer's instructions. For reverse transcriptase PCR reactions, the TaqMan Abgene universal mix (Thermo Scientific, Rockford, IL, USA) and probes from the Universal Probe Library (Roche, Penzberg, Germany) were used. Oligonucleotide primers were synthesized by Sigma-Aldrich (MO, USA). Sequences of primers and the respective UPL probe numbers are given (Supplementary Table S13). Data were analyzed according to the $\Delta\Delta C_t$ method (Bioconductor *ddCt*).

Data analysis

Images of RPPAs were captured by the GenePix software. Local background values were removed with the default option. The normalization against total protein concentration was performed with an additive linear model. Relative protein abundance of each miRNA was summarized as z-score (Malo et al, 2006). Two replicates were summarized by the mean of square sum.

The miRNA target predictions were queried in two target prediction databases: TargetScan (<http://www.targetscan.org>) and MicroCosm (<http://www.ebi.ac.uk/enright-srv/microcosm/cgi-bin/targets/v5/search.pl>) with default parameters. Kolmogorov-Smirnov tests were performed with R. Cell-based assay results were analyzed with R packages in the Bioconductor (Gentleman et al, 2004) community. All the other statistical analysis, if not explicitly stated, was performed using R scripts.

Network analysis of the miRNA-protein network

Construction of bipartite graph

The experimental data can be represented in an array A of dimensions (number of proteins) \times (number of miRNAs) $=810$. Let P be the set of proteins and M be the set of miRNAs. Given a threshold t_1 , the bipartite graph $G(t_1)=(MUP, E)$ contains an edge between protein A and miRNA X if the expression level of A upon transfection with X has a z-score whose absolute value is at least as large as t_1 . This edge is unweighted, but we differentiate it as upregulating or downregulating edge, which is represented by different colors.

The number of edges adjacent to a protein or miRNA is defined as its *degree*. For a given bipartite graph, we compute a sample of randomized bipartite graphs in which each node has the same number of edges of both regulation types, i.e., we maintain the so-called degree sequence of the two sides (Zweig, 2011). The method to compute these samples is explained in the following subsection. By this randomization, we accomplish a bootstrapping method that helps us to identify those co-regulations that are not merely caused by the structural constraints (*degrees*) of the system. For these samples and all of the three co-regulation patterns (co-upregulation, co-downregulation and antagonistic co-regulation), a P -value is then assigned to each pair of

proteins that describes how likely it is to see these two proteins at least as often in the given pattern just by random fluctuations. If these values are low, we claim that the pattern is *more-than-random*. The P -value then needs to be thresholded by a second threshold t_2 to result in the final co-regulation network between the proteins.

Definition of the *more-than-random* model

Given the bipartite graph $G(t_1)$ the degree of each protein and each miRNA is fixed, but the connections between them are perturbed by the following procedure: In each step, pick two edges (A,X) and (B,Y) uniformly at random (here: A and B are two proteins, and X and Y are two miRNAs). If both edges are of the same regulation type, try a so-called *edge swap*: if (A,Y) and (B,X) are not yet connected, remove edges (A,X) and (B,Y) and add edges (A,Y) and (B,X) . If at least one of the edges (A,Y) or (B,X) already exists, do nothing. The *edge swap* constitutes a random walk on a Markov chain (Brualdi, 2006). It is thus assured that, if the number of attempted and conducted edge swaps is large enough, the resulting graph is a random sample from all bipartite graphs with this fixed degree sequence. In our experiments, we computed 10 000 samples after $q \log q$ time steps, where q is the number of edges in the respective bipartite graph. The first random walk starts at the observed bipartite graph, all following ones start from the previously obtained bipartite graph.

Assessing the P -value of a given relationship between two proteins

For any two proteins A, B and any of the four relationships (I (co-upregulation), II (co-downregulation), and the two antagonistic patterns: IIIa and IIIb) the number of miRNAs constituting this relationship is computed in the data and denoted by $\text{obsI}(A,B)$, $\text{obsII}(A,B)$, $\text{obsIIIa}(A,B)$ and $\text{obsIIIb}(A,B)$. The observed value is compared with the distribution of these values in the samples. We are interested in the P -value, i.e., the percentage of samples in which the value is at least as large as the observed one. For each pair of proteins and each observed relationship that occurs at least once, this P -value is then computed for the given number of samples.

Computing consistent co-regulation relationships between proteins

After all P -values are computed, a second threshold t_2 is chosen and two proteins are connected if their P -value is at most as large as t_2 . The resulting graph is denoted by $G(t_1, t_2)$. To determine a set of suitable t_2 thresholds, we computed the so-called clustering coefficient for a fine-grained set of possible t_2 thresholds (Watts and Strogatz, 1998). The rationale is that if there is a substantial connectedness or even different levels of connectedness in the graph, then this curve will show one or more maxima or plateaus in the clustering coefficient. This will then enable us to identify structural changes between the graphs obtained with different thresholds t_2 . In our case, there were little jumps in which the average clustering coefficient increased (Supplementary Figure S11). A jump means a difference between the clustering coefficients for two consecutive P -values. We selected the five largest differences occurring at a still reasonable significance level. We chose the $0.01 < t_2 < 0.2$ domain and imposed that the difference between two consecutive chosen t_2 values should be > 0.01 .

Note that A and B can, in principle, be connected by all three kinds of edges but, biologically, we expect that A and B are (a) in only one relationship, or (b) co-upregulated by some miRNAs (pattern I) and co-downregulated by others (pattern II) or (c) that they are always antagonistically co-regulated (pattern IIIa/b). Inconsistent mixtures of these patterns have not occurred with any combinations of t_1 and t_2 that we have tested so far. The edges can be assigned with the miRNAs that induce the relationship between the two proteins

The consensus graph

As we reasoned in the Results section, the choice of t_1 influences the resulting bipartite graph. It will also influence the identification of

those co-regulations that are statistically significant as edges are added to the system. To keep only those co-regulations that are assigned statistical significance under different t_1 thresholds, we built the *consensus graph*. For this purpose, we chose three relatively relaxed thresholds for t_1 , namely 1.28 ($P < 0.2$), 1.64 ($P < 0.1$) and 1.96 ($P < 0.05$). For those, we computed the P -values of all relationships and the resulting protein co-regulation graphs for a series of five t_2 values. The t_2 thresholds were chosen based on the change of the average clustering in function of the P -values. These graphs vary in their size and the number of edges. We concentrated on those edges that all of the three graphs (corresponding to one strictness level t_2) find statistically significant. This results in a *consensus graph*, in which each edge only exists if it is contained in all three graphs at the same t_2 value.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank Frederik Bethke, Maik Wosch and Corinna Becki for the excellent technical help. This work was supported within the National Genome Research Network (grant 01GS0864) and the Medical Systems Biology program (grant 0315396B) of the German Federal Ministry of Education and Research (BMBF) and Wilhelm Sanderstiftung (grant 2009.051.1). JDZ was supported by the DKFZ International PhD Program. EAH and KZ are supported by the Heidelberg Graduate School of Mathematical and Computational Methods for the Sciences, University of Heidelberg, Germany, which is funded by the German Excellence Initiative (GSC 220).

Author contributions: ÖS conceived, coordinated and supervised the study. SU, HM, UK and ÖS designed the experiments. SU, HM, CS, MK, FH and AW performed the experiments. JDZ, EAH and KZ analyzed the data. JDZ, UT, UK and SW contributed reagents/materials/analysis tools. SU, JDZ, EAH, KZ, SW and ÖS wrote the paper.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Agirre X, Vilas-Zornoza A, Jimenez-Velasco A, Martin-Subero JJ, Cordeu L, Garate L, San Jose-Eneriz E, Abizanda G, Rodriguez-Otero P, Fortes P, Rifon J, Bandres E, Calasanz MJ, Martin V, Heiniger A, Torres A, Siebert R, Roman-Gomez J, Prosper F (2009) Epigenetic silencing of the tumor suppressor microRNA Hsa-miR-124a regulates CDK6 expression and confers a poor prognosis in acute lymphoblastic leukemia. *Cancer Res* **69**: 4443–4453

Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP (2008) The impact of microRNAs on protein output. *Nature* **455**: 64–71

Barker A, Giles KM, Epis MR, Zhang PM, Kalinowski F, Leedman PJ (2010) Regulation of ErbB receptor signalling in cancer cells by microRNA. *Curr Opin Pharmacol* **10**: 655–661

Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. *Cell* **136**: 215–233

Boutz DR, Collins PJ, Suresh U, Lu M, Ramirez CM, Fernandez-Hernando C, Huang Y, Abreu Rde S, Le SY, Shapiro BA, Liu AM, Luk JM, Aldred SF, Trinklein ND, Marcotte EM, Penalva LO (2011) Two-tiered approach identifies a network of cancer and liver disease-related genes regulated by miR-122. *J Biol Chem* **286**: 18066–18078

Brualdi RA (2006) Algorithms for constructing (0,1)-matrices with prescribed row and column sum vectors. *Discrete Math* **306**: 3054–3062

Bueno MJ, Perez de Castro I, Malumbres M (2008) Control of cell proliferation pathways by microRNAs. *Cell Cycle* **7**: 3143–3148

Chen X, Guo X, Zhang H, Xiang Y, Chen J, Yin Y, Cai X, Wang K, Wang G, Ba Y, Zhu L, Wang J, Yang R, Zhang Y, Ren Z, Zen K, Zhang J, Zhang CY (2009) Role of miR-143 targeting KRAS in colorectal tumorigenesis. *Oncogene* **28**: 1385–1392

Choudhary C, Mann M (2010) Decoding signalling networks by mass spectrometry-based proteomics. *Nat Rev Mol Cell Biol* **11**: 427–439

Fowler A, Thomson D, Giles K, Maleki S, Mreich E, Wheeler H, Leedman P, Biggs M, Cook R, Little N, Robinson B, McDonald K (2011) miR-124a is frequently down-regulated in glioblastoma and is involved in migration and invasion. *Eur J Cancer* **47**: 953–963

Friedman RC, Farh KK, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**: 92–105

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

Gao XN, Lin J, Li YH, Gao L, Wang XR, Wang W, Kang HY, Yan GT, Wang LL, Yu L (2011) MicroRNA-193a represses c-kit expression and functions as a methylation-silenced tumor suppressor in acute myeloid leukemia. *Oncogene* **30**: 3416–3428

Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G *et al* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **5**: R80

Grey F, Tirabassi R, Meyers H, Wu G, McWeeney S, Hook L, Nelson JA (2010) A viral microRNA down-regulates multiple cell cycle genes through mRNA 5'UTRs. *PLoS Pathog* **6**: e1000967

Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res* **36**: D154–D158

Hanina SA, Mifsud W, Down TA, Hayashi K, O'Carroll D, Lao K, Miska EA, Surani MA (2010) Genome-wide identification of targets and function of individual MicroRNAs in mouse embryonic stem cells. *PLoS Genet* **6**: e1001163

Hatanpaa KJ, Burma S, Zhao D, Habib AA (2010) Epidermal growth factor receptor in glioma: signal transduction, neuropathology, imaging, and radioresistance. *Neoplasia* **12**: 675–684

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

Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, Labourier E, Reinert KL, Brown D, Slack FJ (2005) RAS is regulated by the let-7 microRNA family. *Cell* **120**: 635–647

Karginov FV, Conaco C, Xuan Z, Schmidt BH, Parker JS, Mandel G, Hannon GJ (2007) A biochemical approach to identifying microRNA targets. *Proc Natl Acad Sci USA* **104**: 19291–19296

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

Liu G, Friggeri A, Yang Y, Park YJ, Tsuruta Y, Abraham E (2009) miR-147, a microRNA that is induced upon Toll-like receptor stimulation, regulates murine macrophage inflammatory responses. *Proc Natl Acad Sci USA* **106**: 15819–15824

Liu H, D'Andrade P, Fulmer-Smentek S, Lorenzi P, Kohn KW, Weinstein JN, Pommier Y, Reinhold WC (2010) mRNA and microRNA expression profiles of the NCI-60 integrated with drug activities. *Mol Cancer Ther* **9**: 1080–1091

Liu Q, Fu H, Sun F, Zhang H, Tie Y, Zhu J, Xing R, Sun Z, Zheng X (2008) miR-16 family induces cell cycle arrest by regulating multiple cell cycle genes. *Nucleic Acids Res* **36**: 5391–5404

Malo N, Hanley JA, Cerquozzi S, Pelletier J, Nadon R (2006) Statistical practice in high-throughput screening data analysis. *Nat Biotechnol* **24**: 167–175

- Mannsperger HA, Uhlmann S, Schmidt C, Wiemann S, Sahin O, Korf U (2010) RNAi-based validation of antibodies for reverse phase protein arrays. *Proteome Sci* **8**: 69
- Pierson J, Hostager B, Fan R, Vibhakar R (2008) Regulation of cyclin dependent kinase 6 by microRNA 124 in medulloblastoma. *J Neurooncol* **90**: 1–7
- Sahin O, Frohlich H, Lobke C, Korf U, Burmester S, Majety M, Mattern J, Schupp I, Chaouiya C, Thieffry D, Poustka A, Wiemann S, Beissbarth T, Arlt D (2009) Modeling ERBB receptor-regulated G1/S transition to find novel targets for *de novo* trastuzumab resistance. *BMC Syst Biol* **3**: 1
- Selbach M, Schwanhauser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**: 58–63
- Shankavaram UT, Reinhold WC, Nishizuka S, Major S, Morita D, Chary KK, Reimers MA, Scherf U, Kahn A, Dolginow D, Cossman J, Kaldjian EP, Scudiero DA, Petricoin E, Liotta L, Lee JK, Weinstein JN (2007) Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study. *Mol Cancer Ther* **6**: 820–832
- Smirnov N (1948) Tables for estimating the goodness of fit of empirical distributions. *Ann Math Stat* **19**: 279–281
- Sun Z, Asmann YW, Kalari KR, Bot B, Eckel-Passow JE, Baker TR, Carr JM, Khrebtukova I, Luo S, Zhang L, Schroth GP, Perez EA, Thompson EA (2011) Integrated analysis of gene expression, CpG island methylation, and gene copy number in breast cancer cells by deep sequencing. *PLoS One* **6**: e17490
- Thomson DW, Bracken CP, Goodall GJ (2011) Experimental strategies for microRNA target identification. *Nucleic Acids Res* **39**: 6845–6853
- Uhlmann S, Zhang JD, Schwager A, Mannsperger H, Riazalhosseini Y, Burmester S, Ward A, Korf U, Wiemann S, Sahin O (2010) miR-200bc/429 cluster targets PLCgamma1 and differentially regulates proliferation and EGF-driven invasion than miR-200a/141 in breast cancer. *Oncogene* **29**: 4297–4306
- Watts DJ, Strogatz SH (1998) Collective dynamics of ‘small-world’ networks. *Nature* **393**: 440–442
- Webster RJ, Giles KM, Price KJ, Zhang PM, Mattick JS, Leedman PJ (2009) Regulation of epidermal growth factor receptor signaling in human cancer cells by microRNA-7. *J Biol Chem* **284**: 5731–5741
- Wellner U, Schubert J, Burk UC, Schmalhofer O, Zhu F, Sonntag A, Waldvogel B, Vannier C, Darling D, zur Hausen A, Brunton VG, Morton J, Sansom O, Schuler J, Stemmler MP, Herzberger C, Hopt U, Keck T, Brabletz S, Brabletz T (2009) The EMT-activator ZEB1 promotes tumorigenicity by repressing stemness-inhibiting microRNAs. *Nat Cell Biol* **11**: 1487–1495
- Wilson B, Liotta LA, Petricoin III E (2010) Monitoring proteins and protein networks using reverse phase protein arrays. *Dis Markers* **28**: 225–232
- Yu Z, Wang C, Wang M, Li Z, Casimiro MC, Liu M, Wu K, Whittle J, Ju X, Hyslop T, McCue P, Pestell RG (2008) A cyclin D1/microRNA 17/20 regulatory feedback loop in control of breast cancer cell proliferation. *J Cell Biol* **182**: 509–517
- Zweig K (2011) A systematic approach to the one-mode projection of bipartite graphs. *Social Netw Anal Mining* **1**: 187–218



Molecular Systems Biology is an open-access journal published by *European Molecular Biology Organization* and *Nature Publishing Group*. This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License.