8 April 2018

Submitted: Molecular Systems Biology

The Translation Machinery Is Immune from miRNA Perturbations: A Cell-Based Probabilistic Approach

Shelly Mahlab-Aviv[1], Nathan Linial[1] and Michal Linial[2*]

[1]The Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem, Israel; [2]Department of Biological Chemistry, Institute of Life Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel

*Corresponding author: ML

The Hebrew University of Jerusalem,

Edmond J. Safra Campus, Givat Ram, Jerusalem 91904, ISRAEL

Telephone: +972-2-6585448; +972-54-8820035; FAX: 972-2-65623429

LM: michall@cc.huji.ac.il

Running title: Modeling cellular miRNA regulation

Figures: 1-7

Expanded View:

Appendix Text EV1, Appendix Figures S1-S6; Table EV1, Dataset EV1-EV10

Abstract

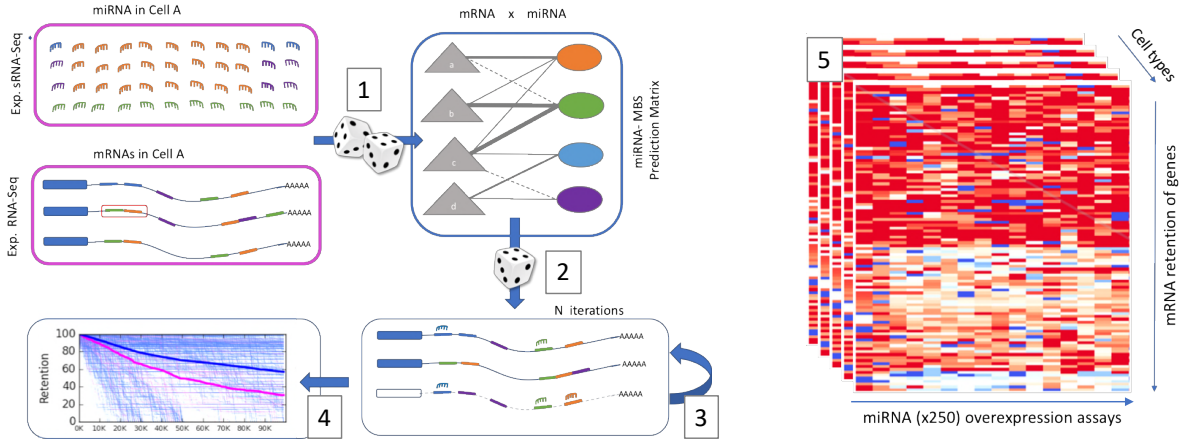
Mature microRNAs (miRNAs) are non-coding RNA that regulate most human genes through base-pairing with their targets. Under a condition of transcriptional arrest, cells were manipulated by overexpressing miRNAs. We observed global time-dependent changes in mRNA retention which are not restricted to the overexpressed miRNA targets. We developed COMICS (Competition of MiRNA Interactions in Cell Systems), a stochastic computational iterative framework for identifying general principles in miRNA regulation. We show that altering the composition of miRNAs governs cell identity. We identified gene sets that exhibit a coordinated behavior with respect to an exhaustive overexpression of all miRNAs. Among the stable genes that exhibit high mRNA retention levels, many participate in translation and belong to the translation machinery. The stable genes are shared among all tested cells, in contrast to the sensitive, low retention genes that are cell-type specific. We conclude that the stochastic nature of miRNA action imparts an unexpected robustness to living cells. The use of a systematic probabilistic approach exposes design principles of miRNAs regulation toward cell states, cell identity, and the translational machinery.

Keywords: Actinomycin D/ CLIP-Seq / miRNA-target prediction/ Stochastic model/ TargetScan

Synopsis

This study models the miRNA regulatory network in different cell-lines under a transcriptional arrest paradigm. The probabilistic model is implemented using a stochastic computational framework called COMICS. The molecular outcome under miRNA regulation scheme is revealed from thousands of simulations under exhaustive miRNA manipulations.

- Transcription arrest emphasizes the impact of miRNA manipulations on gene expression. The composition of miRNA, and primarily the most abundant ones dominate mRNA attenuation in living cells.
- Cell identity can be shifted by the cell-specific composition of the miRNAs but not the mRNAs.
- An exceptional immunity of genes vis-a-vis any manipulations of miRNAs is a property of the translational machinery. It suggests a common signature with respect to the nature of the miRNA binding sites.
- Changing the parameters of the miRNAs probabilistic model affects the dynamics of gene expression but not its steady state or the cell identity.



Introduction

Mature microRNAs (miRNAs) are small, non-coding RNA molecules (~22 nucleotides) that regulate genes through base-pairing with their cognate mRNAs, mostly at the 3′ untranslated region (3'-UTR) (Ameres & Zamore, 2013; Moore et al, 2015; Pasquinelli, 2012). In multicellular organisms, miRNAs act post-transcriptionally by affecting the destabilization and degradation of mRNAs, as well as interfering with the translation machinery (Chekulaeva & Filipowicz, 2009; Eichhorn et al, 2014; Filipowicz et al, 2008). Transition between cell states are accompanied by alterations in the profile of miRNAs (Pelaez & Carthew, 2012). This applies, e.g., to cells that undergo quiescence (Cheung et al, 2012), differentiation (Yang et al, 2013), viral infection (Zhang et al, 2012) and cancer transformation (Bertoli et al, 2015; Lu et al, 2005).

In humans there are ~2500 mature miRNAs that are derived from ~1900 genes (Ameres & Zamore, 2013). The corresponding miRNA post-transcriptional regulatory network can be modeled as an edge-weighted bipartite graph with nodes corresponding to ~20,000 coding gene transcripts, and ~2500 mature miRNAs. Edge weights represent the efficiency of the relevant miRNA-mRNA interactions. Studies of the miRNA-mRNA regulatory networks reveal that almost all coding genes have multiple putative miRNA binding sites (MBS) at their 3'UTR (Landgraf et al, 2007; Liang et al, 2007; Stark et al, 2005), and many miRNAs can possibly target hundreds of transcripts (Balaga et al, 2012; Rajewsky, 2006). The abundance of miRNAs in cells is extremely unbalanced, with a few highly expressed miRNAs, and hundreds of low abundance (Gaur et al, 2007). This miRNA-mRNA "many to many" relation underlies the capacity for noise reduction (Ebert & Sharp, 2012; Herranz & Cohen, 2010; Schmiedel et al, 2015) and robustness against environmental fluctuation (Li et al, 2009).

Most of our current knowledge on the specificity of miRNA-mRNA regulatory network is based on computational prediction tools (Peterson et al, 2014). However, available miRNA-target prediction tools suffer from a large number of false positives (Pinzon et al, 2017). In-vitro studies in which gene expression is monitored following individual miRNA overexpression or knockdown (Hausser & Zavolan, 2014) validated hundreds of miRNA-mRNA pairs. Furthermore, large-scale experimental

settings such as PAR-CLIP (Hafner et al, 2010), AGO-CLIP (Wen et al, 2011), IMPACT-seq (Tan et al, 2014) and other CLIP-based protocols provide more direct measures for the miRNAs-mRNAs interactions in living cells (Li et al, 2014a). The generation of miRNA-mRNA chimeras by CLASH (Helwak et al, 2013) has exposed the abundance of non-canonical binding sites and a large number of coding-region interactions. Unfortunately, many of the above protocols suffer from low coverage and poor consistency (discussed in (Chi et al, 2012)). Hundreds of novel miRNAs (Kozomara & Griffiths-Jones, 2013) have led to further expansion in the number of miRNA-targets predicted pairs. Current estimates suggest that in reality only ~60% of the human coding genes are regulated by miRNAs in a cellular context (Ha & Kim, 2014; Jonas & Izaurralde, 2015). Thus, the potential of many predicted miRNA-mRNA pairs to effectively regulate gene expression remains questionable (Betel et al, 2010; Helwak et al, 2013; Seok et al, 2016). A picture emerges in which poorly expressed miRNAs are not part of the post-translational regulation of gene expression (Hausser & Zavolan, 2014).

The specifics of the miRNA-mRNA regulation in a particular cell type depend on the amounts and concentration of miRNAs and their stoichiometry. A key role in this regulatory mechanism is due to the availability of AGO protein, the catalytic component of the RNA silencing complex (RISC) (Janas et al, 2012; Wen et al, 2011). From the mRNA perspective, the number of molecules, the specific expressed variants and the positions of MBS along the relevant transcript (Jens & Rajewsky, 2015) dictate the potential of miRNA interaction, but not necessarily the potential for successful gene expression attenuation (Agarwal et al, 2015). In vivo, the miRNA regulatory network and its dynamics is part of larger circuits that include master regulators and transcription factors (Bisognin et al, 2012; Friard et al, 2010). The properties of the network call for an unbiased probabilistic model to the end of improving predictions (Nam et al, 2005).

A quantitative view of the miRNAs network in living cells challenges the level of competition, as formulated by the ceRNA hypothesis (Denzler et al, 2014; Salmena et al, 2011). Accordingly, an overexpression of MBS-rich molecules of RNA may displace miRNAs from their primary authentic targets (Denzler et al, 2016; Tay et al, 2014), resulting in an attenuation relief of specific mRNAs. As such, the availability of free and occupied MBS (Arvey et al, 2010), concentrations and binding

affinities are major determinants in governing the new steady state in a cells. The result of such a competition is an interplay between direct and indirect effects on gene expression (Yuan et al, 2015a). Under this scheme, the functional potency of pairing of miRNAs to their targets is driven by the degree of complementarity (Seok et al, 2016). It was furthermore postulated that many weak sites contribute to target-site competition without imparting repression (Denzler et al, 2016). The dynamics of the miRNA-target regulatory network in view of direct and distal ceRNA regulation had been modeled (Nitzan et al, 2014).

In this paper, we describe a quantitative stochastic model that challenges the cell steady-state in view of alteration in miRNAs' abundance. The model operates at the cellular level and compares the overall trend of miRNA regulation in various cell lines. We have tested the validity of an iterative cell simulator to correlate with results extracted from experimental data, under the condition of transcriptional arrest. Based on a probabilistic exhaustive list of all miRNA-target pairings, we configure an iterative stochastic model and test it in view of an exhaustive set of miRNA overexpression manipulations. We show that genes that belong to the translational machinery and function in the translation process are in general unaffected by overexpression of the majority of miRNAs, while a small group of cell-specific gene set is extremely sensitive to the regulation across most miRNAs. We systematically analyzed the design principles of the miRNA-mRNA regulation in various cell types. We confirm that the stochastic nature of miRNA regulation imparts an unexpected robustness of the miRNA regulation in living cells.

Results

miRNAs stability and decay rate of mRNAs upon transcriptional arrest

The nature and extent of miRNA regulation in living cells is depicted by the absolute quantities, composition and stoichiometry of the main players of the network, i.e., the miRNAs and mRNAs (Arvey et al, 2010). The goal of this study is to model the outcome of the miRNA-mRNA network under a simplified setting of transcriptional arrest, where synthesis of new transcripts (miRNA and

mRNAs) is prevented. To this end, we first tested the relative changes in the quantities of miRNAs and mRNAs in HeLa and HEK293 cell-lines in the presence of the transcriptional inhibitor Actinomycin D (ActD, Fig 1A). Overall, we mapped 539 and 594 different miRNAs expressed at time 0 in HeLa and HEK293 cells, respectively (Fig 1A). In addition, prior to ActD treatment, 16,236 and 16,463 different expressed mRNAs (not including miRNAs) were mapped from HeLa (Dataset EV1) and HEK293 cells (Dataset EV2), respectively.

We tested the composition of miRNAs and mRNAs along 24 hrs post ActD treatment. The amount of miRNAs after 24 hrs from the application of the drug is practically unchanged. High correlations between miRNA expression of the untreated (t = 0 hr) and ActD treated for 24 hrs (Fig 1B) are evident for HeLa (Spearman rank correlation, r= 0.94) and HEK293 cells (Spearman rank correlation, r= 0.97). In contrast to the stability of miRNAs, the number of mRNAs types monotonically declines, in accordance with the effect of ActD on the bulk of short lived mRNAs (Fig 1A). A maximal variability in the profile of mRNAs is measured between 0 hr and 24 hrs for HeLa (Fig 1B, Spearman rank correlation, r= 0.84, top right) and HEK293 cells (Spearman rank correlation, r= 0.88, bottom right, right). The correlations of the pairs of all time points for HeLa are shown in Appendix Fig S1, and for HEK293 in Appendix Fig S2.

Fig 1C follows the change in the expression level of individual genes in HeLa cells along 24 hrs from ActD treatment. All expressed genes were normalized according to the basal condition (i.e., 100% expression at 0 hr). The change in the abundance of each mRNA, at each time point was quantified relative to the abundance at the starting time point. As >50% of the identified mRNAs are expressed at a very low level (Dataset EV1), we only report on the percentage of retention for genes that are expressed above a predetermined threshold (total 860 genes, Fig 1C). We illustrate how the distribution of the retention level (in %) of these 860 genes vary between two different time points (Fig 1D). The calculated average retention rate 8 hrs after ActD treatment is ~83%, and decreases to 53% after 24 hrs. These results validate that the decay rate for most mRNAs is a gradual process that continues for 24 hrs. However, the trend for the attenuation in individual gene is already achieved by 8 hrs from the initiation of transcription inhibition.

>>>Figure 1<<<

Overexpression of specific miRNAs impacts targets and non-target mRNAs

Fig 2 shows the results of direct and indirect effects of overexpressing hsa-mir-155 under the setting of transcriptional arrest. HeLa cells were transfected with the individual miRNA, and the effect of miRNA and mRNA composition was tested 24 hrs after cell exposure to ActD. We quantified the effect of hsa-mir-155 by considering its predicted targets. Specifically, for each miRNA, we split the list of all expressed genes to targets and non-targets using the high-quality TargetScan 7.1 prediction table (Agarwal et al, 2015) (see Materials and Methods).

Retention rates of all genes relative to their starting point (i.e., overexpressed miRNAs prior to the transcriptional arrest) are shown (Fig. 2A). The average decay rates of the direct target genes (Fig 2A, pink thick line) and non-target genes (Fig 2A, blue thick line) for hsa-mir-155 shows that the decay of direct targets is slightly faster compared with the rest of the non-target set (Fig 2A, right panels).

The significance of the differences in the decay rate after 24 hrs of ActD treatment on HeLa cells for cells overexpressing hsa-mir-155 indicates faster degradation and an overall lower retention for hsa-mir-155 targets in transfected vs. naïve cells (Fig 2B, upper panel). The observed shift in the relative mean statistics (Fig 2B) of direct targets is below significance (p-value = 0.12). More significant is the shift in the higher retention rates for the non-target genes (p-value 0.002, Fig 2B, compare solid and dashed line). This implies a certain degree of indirect stabilization of hsa-mir-155 non-target genes as a result of overexpression of hsa-mir-155. The same trend in the retention profiles was observed by analyzing HeLa cells overexpressing hsa-mir-124a in HeLa cell (Appendix Fig S3).

These results argue that under the described experimental settings, indirect effects are anticipated presumably due to a competition on MBS and the continuous changes in the stoichiometry of the key players which affect the probabilities of miRNA-target interactions.

>>>Figure 2<<<

Assessing the probabilistic approach for miRNA - mRNA interactions

The experimental results (Figs 1-2) emphasize the need for a systematic analysis of the miRNA-mRNA interaction network acting under quantitative and stoichiometric constrains in living cells. It is evident that even under transcriptional arrest paradigm the high levels of overexpressed miRNA do not solely affect their direct target genes.

Evidently, the molecular interactions of miRNA and mRNA within the cell is a stochastic process. The specific composition of miRNA and mRNA in cells, and their binding probabilities dictate the effectiveness of attenuation of gene expression. The miRNA-target prediction of TargetScan provides a sparse table of miRNA-MBS interactions and reports on 1,183,166 such pairs (see Materials and Methods). Each miRNA-MBS interaction is associated with a probabilistic score that is a proxy for the level of confidence for that interaction, and can be considered the probability of effective binding for any specific pair.

To further investigate the properties and the design principles of the miRNA-mRNA interaction network, we developed an iterative simulator called COMICS (Competition of miRNAs Interactions in Cell Systems). Fig 3A illustrates the main flow in a single iterative cycle in COMICS. The probabilistic framework relies on a constant update of the cell state which is defined by the amounts and composition of miRNAs and mRNA types, and the balance of occupied and free molecules. COMICS iterations capture the stochastic process that takes place in living cells that are subjected to miRNA regulation.

Fig 3B is a breakdown of COMICS process according to the fate of the regulated mRNAs. Specifically, it shows the sampling process driven by the composition of miRNAs and mRNAs and their measured amounts (Fig 3B, pink frames). Recall that the measured expression profiles of miRNAs and mRNAs is cell-type specific as evident from the list of mRNAs and miRNA from HeLa (Dataset EV1) and HEK293 (Dataset EV2). Each mRNA is characterized by the types and positioning of its MBS at the 3'-UTS of the transcript. The interaction prediction table is associated with a probability-based scores for any specific pairs of miRNA and MBS in the context of a specific mRNA. In each iteration, a miRNA is sampled randomly, according to the cell's miRNA abundance and composition. Next, one of its target genes is chosen randomly according to the measured

expressed mRNAs distribution. In the following randomized step, the chosen miRNA and its target may possibly bind according to the reported probability. Following an actual binding event, the distribution of the miRNAs and the mRNAs are updated accordingly (Figs 3A-3B). The status of the mRNA following a successful pairing is changed (i.e., marked as prone to degradation). The new status of mRNA as 'occupied' does not prevent it from engaging in additional subsequent miRNA binding. For MBS that are in close proximity to each other, a minimal spacing is required, otherwise, the interaction will be prevented. The occupied mRNA is marked for degradation with some delay that mimics the likely instance of a cooperative binding on a target by multiple miRNAs prior to its degradation. Based on the validated stability of miRNAs (Fig 1B), once the occupied mRNA is removed, all miRNAs that were bound to it are relieved and return to the free miRNA pool. As a result, the stoichiometry of miRNA to mRNA is gradually changing with an increase in the apparent ratio of miRNAs to free mRNAs in the cell.

Fig 3C shows the result of COMICS following one million iterations on HeLa cells. Note that following 1M iterations, the average retention of mRNAs is 43.5%, similar to the decay rate observed in living cells (Figs 1-2). Fig 3C shows the decay rate of 755 genes (with expression threshold (>0.02%). The output of the mRNAs along the 1M iteration run is reported in Dataset EV3.

We challenged the validity of COMICS to capture the pattern of mRNA downregulation upon transcriptional arrest using the results from living cells. We applied two complementary tests: (i) Detecting a correlation between genes that were unoccupied by miRNAs (i.e., all genes that remain available following 24 hrs of ActD treatment, Fig 3D). (ii) Scoring the correspondence of genes that were occupied along the COMICS iterative run (100k iterations, Fig 3E) and their correspondence with published results from CLASH performed on HEK293 (Helwak et al, 2013). For the first test, we compare the COMICS performance with the experimental results considering a large set of genes that display high level of retention (>85%). This high retention set includes 122 genes in HeLa and 158 genes in HEK293, respectively. We found that overall, the corresponding score (Jaccard score) is higher for HEK293 cells (Fig 3D). However, the statistical significance is maximal for 100k iteration in the case of HeLa cells (hypergeometric p-value of 0.00064) with high significance for matching results from COMICS simulator and the experimental results (Fig 3D).

Using the cell state data for HEK293 along the simulation process (up to 100k iterations), the overlap with the results presented by CLASH (Helwak et al, 2013) is highly significant. We compared all the pairs that were identified in CLASH experiment (Helwak et al, 2013) and expressed above a pre-determined threshold in HEK293. The analysis on HEK293 indicates that as the number of the iterations increases, the overlap of the validated set of pairs increases, and remain highly significant throughout the process (Fig 3E, hypergeometric p-value 0.0014). Most importantly, the correspondence of the results of COMICS to the pairing observed by the CLASH methodology, is strongly dependent of the use of TargetScan miRNA-MBS probabilistic interaction table. Applying two randomization modes (see Materials and Methods and Appendix Text EV1) for the miRNA-MBS interaction table, has led to a drop in the correspondence of the simulator and CLASH reported results (Fig 3E).

We further tested COMICS for mimicking the competition of miRNAs and the trend of mRNAs regulation in living cells. This was performed by testing the sensitivity and the robustness of the simulation results with respect to changes in a broad range of parameters (Appendix Text EV1). Importantly, changing the binding probabilities from the TargetScan interaction table (see Materials and Methods) by two randomization procedures has drastically reduced the correlations at the end of the simulation run. For details see Appendix Text EV1. Additionally, extensive changes in the sets of the operational parameters used by COMICS were tested. Overall, the output of the different simulation runs was consistent and implies the robustness of the system to a wide range of parameters (see Appendix Text EV1).

To account for the variation in the result due to the stochastic nature of the samplings (Fig 3B) along the simulation, we repeated 3 simulation runs. We show that without changing any of the parameters, the final retentions of the genes after 100k iterations in each simulation run are highly correlated (Pearson correlations 0.86-0.88, Appendix Fig. S4).

>>>Figure 3<<<

Alteration in miRNA profiles but not the mRNAs determines cell identity

The gene expression profiles for miRNAs and mRNAs of the 3 tested cell-lines are quite different (Fig. 4A and Dataset EV5). HeLa, HEK293 and MCF-7 cells are representatives for fibroblasts, cells from kidney and breast origin, respectively. We compared the dominant miRNA profiles as processed by COMICS for these cell-types (Fig 4A). The difference in evident in comparing the fraction and composition of the miRNAs that occupy 90% of the total amounts of miRNA in each cell (i.e. 45k out of 50k molecules of miRNA per cell). The unified set for all three cell line miRNAs is shown (total 28 miRNAs). Differences among cells is evident from the presence or absence of specific miRNAs (hsa-let-7, hsa-mir-16, Fig 4A). More significant is the large deviation in expression of miRNAs in a specific cell (e.g., hsa-mir-21), and the fraction that is occupied by a small set of dominating miRNAs.

To further test the impact of the unique miRNA profile for the cell identity, we applied COMICS but activated the simulation process using input extracted from the miRNA profile of one cell type and tested it in the context of mRNA composition of the other cell. We repeated such artificial settings for all 9 combinations (of which 6 are artificial) as shown in Fig 4B. In all the settings, the outcome of the retention profile fully resembles the profile of the cell providing the miRNA profile. For example, by including in COMICS input the data of miRNA profile from HeLa (according to Dataset EV5), the correlation of the retention profile in the end of the simulation run is in the range of 0.9 for input of the mRNA of any of the other cell types. On the other hand, the input of miRNA of MCF-7 or HEK293 using the authentic mRNA profile resulted is a Pearson correlation r= 0.53 and r=0.46, respectively. A similar trend is observed for the HEK293 that shows a high correlation of r=0.93 and r=0.9 when applied on the background of mRNA from HeLa or MCF-7, respectively (Fig 4B).

>>>Figure 4<<<

Simulating miRNA overexpression by COMICS reveals stabilization of non-target genes

The COMICS system was used to simulate exhaustive overexpression experiments while testing explicitly the effect of miRNA-mRNA probabilities on the mRNA decay. We activated COMICS by manipulating the abundance of hsa-mir-155 from the naive cell state (x1, no overexpression) to

another 7 levels for hsa-mir-155 amounts (x0.5, x3, x9, x18, x90, x300 and x1000). The addition of miRNA molecules is considered in a probabilistic framework. Thus, in practical terms, overexpression of a single specific miRNA changes the distribution of all other expressed miRNAs (Fig 5A). The probability of miRNA and target to engage in a successful interaction is dependent on the interaction strength converted to the miRNA-MBS scoring (Fig 3B). The scheme in Fig 5A shows that at an overexpressed factor x300 for hsa-mir-155, almost 20% of all miRNAs are occupied by this miRNA and it reaches almost 50% at the extreme overexpression level (x1000). It is important to note that the calculated fraction of miRNAs in the cell following overexpression is an immediate reflection of its original abundance in the naïve cell. These amounts are characteristic for the different cell types.

Fig 5B shows the gradual change in the retention of each gene (above a predetermined threshold) along 100k iterations of COMICS simulation process. Each panel shows the results of the simulation that starts from the abundance of hsa-mir-155 in the naïve cell (control). Fig 5B show the gradual alteration in the dynamic of the gene retention following the increase in different overexpression levels (determined by the multiplication factor).

We found that the final retention level is sensitive to the overexpression factor (Fig 5C). In the case of hsa-mir-155 in HeLa cells, elevating the miRNA from x18 to x90 caused a drop in the average retention of its targets followed by a steep drop in retentions at overexpressing factor of x1000. A minor but consistent increase in the retention of non-target genes is observed along the elevation in the overexpression level for of hsa-mir-155.

>>>Figure 5<<<

A unified pattern of mRNA retention is associated with overexpression of miRNAs

To determine whether the composition and stoichiometry of miRNAs and mRNA are dominant features in determining the design principles of many cell states, we performed an exhaustive and systematic manipulations of cellular miRNAs. We applied COMICS simulations on overexpression experiments for miRNA families. For HeLa cells, 248 miRNAs were compiled and match their representation in the miRNA-MBS TargetScan prediction table (some miRNA are identified as

miRNA family). We multiply the basal abundance (x1) of each of these miRNA families by several factors: x3, x9, x18, x90, x300 and x1000. For each such factor (f), a final retention table was computed. The result for HeLa cells is a table with 773 genes (rows) whose initial expression level exceeds a pre-determined threshold, and 248 miRNA (columns). Each of these miRNA was overexpressed by the tested factor. Therefore, each cell in the matrix Mf_{ij} is the final retention of gene *i* after 100k iterations of COMICS for the overexpressed experiment of miRNA *j* (Fig 6A, Dataset EV6). For unexpressed miRNA, a standard minimal level of expression is defined as x1 level (see Materials and Methods).

Inspecting Mf_{ij} for each overexpression condition reveals the presence of a substantial set of genes that are characterized by final retention above 85% for ≥90% of the tested miRNAs in the overexpression experiments (i.e., high retention criterion satisfied by at least 225 or 248 tested miRNAs). We refer to them as cross-miRNAs stable genes. We found such stable genes for Hela (185 genes), HEK293 (176 genes) and MCF-7 (124 genes). For a full detailed analysis of cross-miRNAs stable genes see Dataset EV7. These results imply that a set of genes in each cell type are resistant to gene expression attenuation regardless of which miRNA is overexpressed.

The matrix M*f* also reveals a small defined gene set that is globally sensitive to miRNA regulation. Specifically, these are genes with retention rate below 50% for ≥90% of the tested miRNAs among all overexpression experiments. These genes are referred to as cross-miRNAs sensitive genes. We report on these sensitive genes for Hela (23 genes), HEK293 (34 genes) and MCF-7 (22 genes). For a full detailed analysis of cross-miRNAs sensitive genes see Dataset EV8. These results imply that a set of genes in each cell type are sensitive to gene expression attenuation regardless of which miRNA is overexpressed.

The retention matrices [Mfij (x300)] for HeLa (Fig 6B) and HEK293 (Fig 6C) are colored by high retention (red) and low retention (blue) levels. The matrices represent clustering by genes and miRNAs. Analyzing the matrices by the clustering dendrogram shows that the set of sensitive genes (cluster of blue rows) in both cell types is distinguished from all other genes. However, the richness of the retention pattern is demonstrated by zooming at any section of the Mf_{ij} (x300). It is also

evident that some miRNAs are naturally clustered by a similarly profile of their columns. We tested whether the type of the cell dictates the characteristic of the retention profiles, and whether cells that were manipulated by overexpressing miRNAs established a new cell states that are different from their naïve states. Fig 6D and 6E compare the average retention observed for each of the shared genes in the HeLa and HEK293. Large difference is observed in the distribution of genes for HeLa and HEK29 when the profile in Mf_{ij} (x1) and Mf_{ij} (x300) are compared. This global view suggests that each manipulated cell converges to it unique cell state and potentially reflects the different sensitivity of the analyzed genes to the unique cell specific miRNA composition.

We analyzed the specificity of the unified pattern for Mf_{ij} that was found for HeLa, HEK-293 and MCF-7. We tested the overlap in the resulting sets of cross-miRNAs stable genes and cross-miRNAs sensitive genes from these cells (Fig. 7A). In the present analysis we consider only genes that are common to all three cell types and are expressed at >0.04%. Of those, the numbers of cross-miRNAs stable genes that appear in all three cells are 78 (MCF-7), 102 (HeLa) and 110 genes (HEK-293). Of these, 48 genes are common to all three types. The overlap of this number of genes is statistically very significant (chi-square test p-value 1.35e-08, Fig 7A). It argues that these stable genes are designed to resist miRNA regulation under a wide range of overexpression settings and across cell types. The list of these 48 genes that are shared by all three cell lines is shown in Dataset EV9.

Cross-miRNAs sensitive genes are cell specific

We next carry out a similar analysis for cross-miRNAs sensitive (Dataset EV8). These sets are much smaller with 6 (MCF-7), 12 (HeLa) and 17 genes (HEK293) and no sensitive genes are common to all three cell types (Fig 7B). We conclude that these cell-specific sensitive genes are exhibit an exceptional sensitivity to miRNA regulation.

We analyzed the cross-miRNA sensitive lists from the different cell lines using annotation statistics tools (see Materials and Methods). This analysis shows only a moderate enrichment for terms associated with mRNA regulation and processing. For example, among the 23 sensitive genes from HeLa (Fig. 7B) numerous genes are related to transcription processing and modifications, other annotations are for RNA poly-adenylation, and RNA-related metabolic processes (Dataset EV10).

Among the sensitive genes are PCBP1 and PCBP2, single-stranded nucleic acid binding proteins that were identified in an mRNP complex, and play a role in viral internal ribosome entry site (IRES). Another small set of proteins include major pre-mRNA-binding proteins from the heterogeneous nuclear ribonucleoproteins family (HNRNPK and HNRNPA1). HNRNPA1, involved in the packaging and transport of poly(A) mRNA from the nucleus to the cytoplasm. It was recently proposed that HNRNPA1 binds specific miRNA hairpins and acts in miRNA biogenesis (Treiber et al, 2017). Inspection of the sensitive gene set from MCF-7 (22 genes) shows borderline statistical significance for regulation of transcription from RNA polymerase II promoter. HEK293 (34 genes) resulted in enrichment for translational processes (including mitochondria and viral translation). We conclude that in each cell type a small set (22-34 genes) that is extremely sensitive, but there is no unified functional coherence among these cell specific sets.

Cross-miRNAs stable genes are enriched in the translation machinery

We applied annotation enrichment tools (see Materials and Methods) to the set of cross-miRNA stable genes from HeLa cells (185 stable genes, Dataset EV7). This analysis reveals that they are extremely enriched by terms associated with many aspects of translation, including translational elongation (GO:0006414), mitochondrial translation (GO:0032543), SRP-dependent cotranslational protein targeting to membrane, translational termination (GO:0006415) and more. These annotations are associated with highly significant statistics (corrected FDR p-value = ~1*e-77). Clusters of enriched annotations were compiled using the DAVID gene expression tool. The score of the top clusters is 53, depicting e-53 as the average enrichment for all annotations in that cluster. The top clusters are scored by tens of coherent annotations, all related to the structure of the ribosome, elongation machinery and fidelity of translation. For example, the annotation of translational elongation (GO:0006414) shows an enrichment of 43.5 with FDR p-value of 5.53e-87. The results of the most significant annotation clusters are shown in Dataset EV10. A similar significant enrichment remains as we apply the statistical tests using more specific background lists (e.g., the set of HeLa expressing genes from Dataset EV1 and the set of genes that appears in the analyzed Mfij (Dataset EV6). Significant enrichment for clusters of annotation for protein translation and the translation machinery was duplicated for any of the cross-miRNA stable gene lists derived from HEK293 (176

genes) and MCF-7 (123 genes), Dataset EV8), with a DAVID clustering annotation scores of the top cluster of 56.5 and 39.9, respectively. The strongest enrichment of the annotations terms for tested cell-types, extends to genes participating in translation initiation, elongation and co-translation to the ER membranes.

>>>>**Figure 6**<<<<

Fig 7C shows the partition of 48 shared genes that are common to all three cell lines (Dataset EV9) by their protein function. The dominant role of translational machinery (DAVID clustering enrichment score 49.4) is shown by the enrichment corrected score for translation elongation and cytosolic ribosome (FDR p-value of 1.18e-67 and 9.36e-60, respectively, Dataset EV9). Translational machinery component with small and large subunits (35 genes), elongation factors (EIF4A1, EEF1D and EEF1B2) and nucleolin (NCL) account for 79% of this list and are key factors in ribosomal production and its function. The list includes also actin, myosin and tubulin (total 4 genes) that are major cytoskeletal components essential to cell shape and physiology.

We conclude that the cross-miRNA stable gene set signify the translational machinery. Thus, the translational machinery highlights a functional gene set that are immune to the regulatory layer of miRNAs. This observation applies to all tested cells.

Finally, we tested the properties that characterize genes associated with the cross-miRNA stable and sensitive genes in view of all genes that participate in the simulation process (Dataset EV7 and Dataset EV8). Four properties were tested: (i) the number of targeting miRNAs (Fig 7D), (ii) the number of MBS (Fig 7E). The other two features are: (iii) the initial expression level (Fig 7F) and (iv) the binding potential according to the expression of the most dominant miRNAs (Fig 7G). Features (iii-iv) are cell-type specific. The detailed t-test statistics of the results for all three cells are provided in Table EV1. The genes in the stable set are characterized by having fewer MBS and fewer targeting miRNA relative to the other (i.e., not stable and not sensitive) genes (t-test 6.48E-21 and 5.64E-15, respectively). The significance of the statistics for the initial expression of these gene is marginal in all cell types. However, the most significant feature that differentiate the stable from the sensitive set (t-test 2.52E-22, Fig. 7G) is associated with the immunity of the stable set to be

targeted by the most abundant miRNA (based on the miRNA list in Fig 4A). Therefore, the stable genes are unlikely to be effectively regulated by any of the most abundant miRNAs. This trend is evident from the statistics for the three analyzed cell lines (Table EV1).

>>>>**Figure 7**<<<<

The analysis for the 4 quantitative features was replicated using the sensitive and stable dataset from HEK293 and MCF-7 (Appendix Fig. S6). For example, the average number of MBS for a stable gene is 24.2 (vs. 62.5 for the sensitive gene set) in HEK293 cells. The most abundant miRNA expressed in HEK293 is hsa-mir-7 (25% of all miRNAs) targets only 3.5% of the stable genes but 94% of the genes among the cross-miRNA sensitive gene set have a MBS.

We conclude that a comparison between the stable and sensitive genes reveals a signal for MBS evolution. Despite the great difference in overall miRNA composition of cell lines (Fig 4A), several of the miRNAs are shared across many cell-types (e.g., hsa-mir-21, hsa-let-7 and hsa-mir-92). The stable genes are critical gene of the translational machinery, are mostly excluded from having MBS that could engage in binding of a dominant miRNA players in multiple cells.

Discussion

miRNAs stability as a major determinant in cell regulation

Cells' behavior cannot be trivially predicted from direct measurement of their composition of miRNAs and mRNAs (Arvey et al, 2010; Landgraf et al, 2007). Most insights on the regulation of gene expression by miRNAs are based on global observations (e.g., CLIP and CLASH (Li et al, 2013)), or knockdown or overexpression of a specific miRNA, in a specific cell type or tissue (Thomas et al, 2010). Based on many such studies, it was concluded that detailed quantitative considerations of miRNA and mRNA govern the dynamics and the steady state of gene expressed (Bosson et al, 2014; Hausser & Zavolan, 2014). Nevertheless, the underlying rules for post-transcriptional regulation by miRNAs are still missing (Erhard et al, 2014).

We studied cellular outcome following miRNA regulation under a simplified condition of transcriptional arrest, using ActD and focusing on mRNA retention profiles. The post-translation

regulation of miRNAs is not limited to mRNA attenuation. However, it was shown that under such condition of transcriptional arrest the dominant effect of miRNAs is via their influence on mRNA stability and not on translation repression (Bethune et al, 2012). As most miRNAs are transcribed by RNA Polymerase II, it was essential to assess the effect of transcription arrest on the abundance of miRNAs. The results shown in Fig 1B substantiate the notion of an extreme stability of miRNAs at least during 24 hrs from cells' exposure to the drug. In the time frame of our experiments, AGO-2 is insensitive to transcriptional or translational arrest (Olejniczak et al, 2013).

It is expected that miRNA stability is mostly attributed to AGO-2 that stabilizes the leading strand in the cytoplasm (Winter & Diederichs, 2011). The number of miRNA (and AGO proteins) relative to the number of MBS had been estimated from experimental data (Bosson et al, 2014; Janas et al, 2012). These estimates suggest that the number of AGOs proteins limit the pool of RISC loaded miRNAs which plays a role in gene expression attenuation. AGO-2 is estimated to account for 60% of the AGOs in cells. While the number of AGO-2 molecules is unknown, it was estimated at around ~15k in HeLa cells (Janas et al, 2012) and ~100k in skin tissue (Wang et al, 2012). We consider the miRNA pool that is constrained by the amount of AGOs to be at the order of 50k molecules in a cell. COMICS simulations are started with a molecular ratio of 2:1 ratio between miRNA to mRNA, accounting for the AGO-bound active miRNAs. For the probabilistic formulation of cells under varying levels of miRNA overexpression, the sampling is governed by the recalculated composition of miRNAs, assuming that loading of the leading strand of miRNA on AGO proteins is driven by such distribution of that miRNAs (Fig 5A).

miRNA composition is a major determinant in establishing cell identity

We developed the COMICS platform to handle cells that undergo a wide range of miRNA manipulations, and to monitor mRNA retention profiles at the simulation's endpoint (100k iterations, Fig 2B, Fig 3A). Altogether, thousands of simulation processes were completed to test the impact of altering the expression of hundreds of miRNA families. For example, in HeLa cells, 250 miRNAs (Fig 5) were considered, with each miRNA being altered by increasing factors of overexpression, from the base default level (x1) to x1000 (Fig 5, 8 factors). The same protocol was applied to test

miRNA regulation in the other cell types (HEK293 and MCF-7). General trends could be drawn from the results of thousands of COMICS simulations. The main observations hold across most (>90%) of the expressed miRNAs. Specifically, we show that there is a relatively larger set (about 20% of the reported genes) that are exceptionally stable (Fig 7, Dataset EV7). The overlap of these sets of genes in the three cell types is statistically very significant, thus suggesting that the identity of these genes reflects a strong evolutionary signal of stability in the face of extreme increase in almost all miRNAs.

The other phenomenon that COMICS revealed is the presence of a small set of genes (about 3% of the reported genes) that are extreme sensitivity to down-regulation by almost every miRNA. However, the identity of these genes is cell-type specific. Fig 7A shows the minimal overlap that is found between any pair of cells for the sensitive genes. The strong signal of cross-miRNA behavior suggests that it is the miRNA composition of each cell that dominates the identity of the most sensitive genes, many of which participate in transcription regulation. To further test the impact of the actual miRNA profile on cell identity using COMICS, we ran the simulation on artificial combinations of miRNA and mRNA profiles that come from different cell types. In all these artificial settings, the outcome of the retention profile was more similar to the cell providing the miRNA profile (Fig 4B) and not dictated by the mRNA composition. Our results are in accord with the notion of miRNA profiles as a major determinant for cell identity (He et al, 2012). Indeed, specific profiles of miRNAs are associated with varying malignancy states (Bockmeyer et al, 2011; Volinia et al, 2006). A transition among cell types is attributed to the expression of a specific miRNAs (e.g. miR-34a (Bu et al, 2013)). It was shown that miRNA profile is carefully regulated to promote and stabilize cell fate choices (Shenoy & Blelloch, 2014). In this study we discuss genes that belong to cross-miRNA stable or sensitive sets, however the retention profiles for most genes and along the different overexpression levels provides a rich picture and is far more complex as illustrated (Figs 6A, 6B zoom-in panels).

COMICS accounts for the effect of the most abundant miRNA

The stable and sensitive gene sets differ in a statistically significant manner in almost every relevant quantitative feature (Fig 7). As shown for HeLa cells (Figs 6D-6G), the numbers of MBS at the 3'-UTR and the number of different targeting miRNAs are much higher in the sensitive set compared with stable genes. However, the most significant difference between the sensitive and stable sets (p-value = 1.57E-23) concerns the sensitivity to miRNA expression level (Fig. 6F). As shown in Fig. 4A, a small number of miRNAs together comprises 90% of all miRNA molecules in any cell-type. Based on the mass conservation principle, we can expect that highly expressed miRNAs would play a key role in miRNA-mRNA interactions. This principle is indeed implemented in COMICS sampling protocol, and dominant miRNAs are sampled more frequently to interact with their possible targets. Evidently, the probability of a successful interaction for mRNA molecule grows the more MBS that it carries to which some highly abundant miRNAs can bind. Remarkably, and perhaps counterintuitively, it is the stable set has higher initial counts (Fig 6E). This provides additional support to the claim that miRNAs dominate the system, whereas the initial abundance of mRNAs plays only a minor role. The statistical observations regarding the dominant role of the highly expressed miRNAs shown in Fig 7G apply to the other tested cell types. For example, targeting by hsa-mir-21 is prevalent among the cross-miRNA sensitive genes. The hsa-mir-21 occupies 27% and 32% of total miRNAs in the naïve MCF-7 and HeLa cells, respectively (Dataset EV5). The hsa-mir-21 occupies 58 and 73 MBS among the 22 and 34 sensitive genes identified in MCF-7 and HEK293, respectively (Dataset EV8).

Even in settings with high levels of overexpression, the impact of miRNAs which are highly expressed in the naïve cells, prior to the overexpression manipulation (Fig 4A) remains substantial. Such miRNAs would still constitute a significant fraction of the miRNA cellular pool and are therefore likely to be selected by the probabilistic sampling protocol (see Fig 3A).

Fig 7 offers insights on the properties of the stable genes in the various cell lines (Appendix Fig. S6). Combining the statistical observations (Fig 7D-6G, Table EV1) with the functional annotation of the cross-miRNA stable genes suggests an evolutionary signature for these genes. This evolutionary signal could be detected thanks to the unbiased approach taken by COMICS and the exhaustive testing of different cells under many perturbations by all conserved miRNAs. In all the tested cells, a

strong enrichment in translation annotations was shown (including elongation, rRNA metabolic process co-translation to the ER and related functions, Datasets EV9 and EV10). We propose that an evolutionary signal unifies many of the ribosomal proteins and the observed components of the translation machinery. The driving force of evolution acting on targets and their MBS underlying the structure of miRNA network in many organisms (Berezikov, 2011). This evolution signal combines high expression with low density of MBS at the 3'-UTR (Fig. 7E). Even more critically, it relies on the low sensitivity of these mRNAs to the effect of highly abundant miRNAs (Fig. 7G). In cancer cells, the most abundant miRNA genes (e.g., hsa-mir-21, hsa-mir-30, hsa-mir-15/16) are involved in cancer development and even minor changes in their expression govern transition in cell states (Volinia et al, 2010).

We propose that the translation machinery has evolved to maximal robustness vis-a-vis miRNA regulation. The stable genes, enriched with the translation components, are resistant to changes in the presence of abundant miRNAs. The immunity of the translation system to miRNA regulation suggests that it may be part of a global cell strategy (Lopez-Maury et al, 2008). Accordingly, there is a fundamental difference between transcription and translation processes. While transcription system can quickly respond to the needs dictated by changes in the environment (e.g., post-transcription regulation of the sensitive set), the translational machinery is stable and much less prone to variations. The immunity from miRNA regulations of the translational machinery is achieved by the relatively high expression levels (Fig 7F), but most significantly the insensitivity to most abundant miRNA in the cell.

Alteration in COMICS simulation parameters only slightly change cells' steady state

We tested the reliability of the picture provided by COMICS of the miRNA-mRNA competition in living cells. This testing was done by rigorously varying the simulator's operational parameters and checking the sensitivity and robustness of the results (Appendix Text EV1).

Our simulations operate with a total of 50k miRNAs and 25k mRNAs. These numbers are at the lower range of the amounts estimated in most living cells. We showed that changing the number of molecules, while keeping the stoichiometry only slows the dynamics, with a minimal impact on the

endpoint (Appendix Fig S5).

Sampling of miRNAs and mRNA and assessing their pairing is done according to a table of probability scores that represents a rich computational-experimental body of knowledge (Agarwal et al, 2015). Once two molecules are successfully paired, our tables of probabilities get appropriately updated (Fig 3A). We show that the mRNA retention statistics at the endpoint of the simulation runs of COMICS is surprisingly robust under a wide range of parameters (Appendix Fig S4 and Fig 5). From computational considerations and based on numerous empiric observations, each pair of miRNA-target is assigned a score that reflects the calculated degree of downregulation (e.g., (Betel et al, 2010). Correlating the profile of overexpressed miRNAs from in-vitro experiments with the outcome of downregulated genes allows a refinement of the miRNA-target prediction scores (Bloom et al, 2014; Li et al, 2014b), and inferring a probabilistic measure for the effectiveness for each prediction. Along this line, TargetScan prediction tool reports on the probability of a given MBS and its combination to effectively interact with the miRNA at hand (Agarwal et al, 2015). The underlying scoring method takes into account 14 sequence features and integrates them into a probabilistic framework of repression effectiveness. We show that the values calculated in the TargetScan interaction table ((Agarwal et al, 2015)) are the most sensitive parameters for the implementation of COMICS (Appendix Fig S4). If we run the simulation after the MBS interaction values undergo a constrained randomization modification, then our self-correlations drop drastically (Fig 3E, Appendix Text EV1). In all our simulations we use the same TargetScan probabilities table. The approach offered by COMICS allows us to develop more refined models for different cell types.

A dynamic view of miRNA regulation

The dynamics of AGO-miRNA search complex is driven by the need to accelerate the search, while maintaining high specificity (Jo et al, 2015; Klein et al, 2017). Under the described protocol, we consider AGO-miRNA to be sampled according to its relative abundance, followed by a second random sampling, for any of the mRNA potential candidates (based on their probability). According to the sparse interaction affinity matrix from TargetScan (see Methods), most sampling interaction are futile, and on average only 3% of iterations end in a successful binding. This model mimics the

search process of miRNAs is living cells as studied by single molecule microscopy analyses (Chandradoss et al, 2015).

Another aspect of the miRNA-mRNA dynamics concerns the ceRNA paradigm (Denzler et al, 2016). Activation of ceRNA in cells is strongly dependent on an induction of specific transcripts (e.g. circular RNA, pseudogene (Thomson & Dinger, 2016)) that may alter the availability and composition of MBS (Tay et al, 2014). Accordingly, the MBS accessibility may result in redistribution of already occupied MBS by competition. In practice, the extent of de-repression due to MBS shuffling is limited (Bosson et al, 2014). A quantitative test on the direct and indirect effects of manipulating miR-122 in liver cells (Denzler et al, 2014) shows that the abundance of miRNAs and their binding sites precludes the proposed ceRNA effects under all physiological contexts (Broderick & Zamore, 2014). The situation in which miRNAs are competing by reshuffling their binding is supported in living cells only in extreme cases and under artificial extreme parameters. Based on this careful quantitation, it was proposed that low-abundance miRNAs, even with targets of highest-affinity sites are unlikely to experience any meaningful repression (Denzler et al, 2014). Indeed, the simulator explicitly guarantees that low-abundance miRNAs are sampled with low probability, and are thus unlikely to dictate the repression and the outcome of the mRNA retention level. The probabilistic nature of COMICS simulation makes it an ideal tool for testing this paradigm.

There are several improvements and extensions to COMICS that we intend to explore in the future. This includes a synergistic cooperativity of miRNA binding at different non-overlapping MBS (Balaga et al, 2012; Friedman et al, 2014), and the possible induction of preselected mRNA, for assessing the quantitative parameters associated with the ceRNA paradigm. An exhaustive application of the COMICS framework can be used for testing unresolved questions and emerging principles of miRNA regulation in vivo (Denzler et al, 2014; Hausser & Zavolan, 2014; Yuan et al, 2015b).

Materials and Methods

Cell culture

Human cell line of HeLa (cervix epithelial, # CCL-2) and HEK-293 (embryonic kidney, # CRL-1573) were purchased from the cell-line collection of ATCC. Cells were cultured at 37^0C, 5% CO_2 in Dulbecco's Modified Eagle Media (DMEM, Sigma), supplemented with 10% FBS (Life Technologies), and 1% antibiotics mixture (Sigma-Aldrich, Cat # P4333). Cells were maintained for 2 weeks and passing and splitting cells was carried out at 70-80% confluence.

Transcription arrest and miRNA overexpression

Overexpression of miRNAs was performed by transfected HeLa cells and HEK-293 with miRNA expression vectors that are based on the miR-Vec system, under the control of CMV promotor (Origene). Cell transfection was done using Lipofectamine 3000 (Invitrogen) as described by the manufacturer. Cells at 70% to 80% confluency were transfected with 1.5μg purified plasmid DNA containing hsa-mir-155 and hsa-mir-124a (kindly contributed by Noam Shomron, Tel Aviv University). Medium was changed 6 hrs post-transfection and fresh media was added 24 hrs post-transfection. Control empty vector expressing GFP (0.15μg) was mixed with the CMV-miR expressing vectors. Cells were monitored by fluorescent microscopy in a parallel culture at 36 and 48 hrs post transfection. The efficiency of cell transfection was >75% of the HeLa cells and ~100% of the HEK293 according to the GFP expression at 48 hrs post transfection. Transcription inhibition was achieved by adding to cultured HeLa and HEK293 cells media containing Actinomycin D (ActD, 10 μg/ml in DMSO), or the appropriate control (i.e. DMSO). Cells were treated with ActD (10 μg/mL, Sigma) 24 hrs post-transfection. Cells were cultured in 6-well plates and following treatment were lysed in 1 ml TRIzol (Invitrogen) at the indicated time points (0 hrs, 2 hrs, 8 hrs, 24 hrs).

RNA library preparations

At the indicated time points, HeLa cells and HEK293 cells were harvested using a cell-scraper. Purification of total RNA containing miRNA extracted from ~10^6 cells using QIAzol Lysis Reagent RNeasy plus Universal Mini Kit (QIAGEN, GmbH, Hilden, Germany). To ensure homogenization a QIAshredder (QIAGEN, GmbH, Hilden, Germany) mini-spin column has been used. To the upper

aqueous phase 1.5 volumes of 100% ethanol added, and mix thoroughly. Sample has been transferred up to an RNeasy Mini spin column and centrifuge for 15s at ≥8000g at room temperature, and the mixture was processed according to the manufacturer's standard protocol. Samples with an RNA Integrity Number (RIN) >8.5, as measured by Agilent 2100 Bioanalyzer, were considered for further analysis. mRNA libraries were generated using the Illumina Truseq RNA V2 library Seq protocols.

miRNA library preparation

High quality RNA was determined by Agilent 2100 bioAnalyzer. For small RNA library construction, ~1 µg of RNA was used. RNA was ethanol precipitated to enrich for small RNA. Small RNA libraries were prepared according to NEBNext Small RNA Library Prep Set for Illumina (Multiplex Compatible) Library Preparation Manual. Adaptors were then ligated to the 5' and 3' ends of the RNA, and cDNA was prepared from the ligated RNA and amplified to prepare the sequencing library. The amplified sequences were purified on 4% E-Gel Agarose gels (ThermoFisher # G401004), and sequences representing RNA <200 nt were extracted. Data used are derived from at least two biological duplicates. The average values of the two independent sets is reported.

TargetScan probabilistic miRNA-mRNA pairing

The probabilistic framework interaction table was adapted from the scores provided by TargetScan (Agarwal et al, 2015). Accordingly, high probability of successful interactions is calculated from a combination of strongly supported miRNA-mRNA pairs that comply with many features from sequence, secondary structure and evolution conservation. The complete miRNA-mRNA table include 8.22 M pairs that covers also poorly conserved interactions. We compiled the version of TargetScanHuman (Release 7.1) that reports on 19,475 genes (28,353 transcripts). We extracted the TargetScan mRNA CWCS scores (cumulative weighted context++ score), which is proxy for the predicted repression based on the different properties of the MBS sites. The CWCW estimates the score by compiling the contribution of multiple MBS according to a miRNA family and the relative positioning at the 3'-UTR of the transcript. The predicted repression scores range from 0-1, and are

identical for all representation of the relevant miRNA family members (Agarwal et al, 2015). We used a compressed version of the table that report only on pairs that are supported by conserved miRNAs with 1,183,166 pairs, covering 18953 genes and 289 miRNA families.

RNA deep sequencing analysis

RNA extracted from HeLa and HEK293 cells were taken from independent library preparations and were processed in the same sequencing slides according to standard Illumina Protocol (Trapnell et al, 2012). Next Generation sequencing was performed on small RNA (<200 nt) molecules and for mRNA by standard RNA-Seq Illumina Protocol. Each of the 48 RNA-sequencing samples covers the mRNA and miRNA sets (24 sets for the ActD treated on two cells types, at 4 times points with two sets of miRNA overexpression and a set for the control transfected by an empty vector). Each sample consisting of ~25M total reads of length 100 for each read for mRNA detection, and ~10M total reads for the miRNA detection.

The sequencing data, after removal the adaptors and filtering out low quality sequences, were aligned to mirBase (Release 21). In addition, the filtered high-quality fragments were mapped to the human transcriptome of hg19 gtf file from UCSC provided by Galaxy. Specifically, the sequenced small RNA were trimmed using Cutadapt ver. 1.13 and quality filtered using FASTX toolkit. Short reads (~30 nt long) were mapped to miRNA using mapped to miRNA genes using miRExpress 2.0 (Kim et al, 2009). Longer reads were aligned against human genome hg19 using TopHat 2.1.1, allowing 90% sequence identity and a maximum of two mismatches, with a limit of two genome alignment matches. For mRNA expression evaluation, mapped reads were submitted to Cufflinks toolkit version 2.2.1. Out of the mapped reads, only reads of length >= 17 were considered. miRNA sequences refer to mapped, high quality reads that are aligned to any of the pre-miRNA as defined by miRbase databases (ver. 21) (Kozomara & Griffiths-Jones, 2013)

Normalizations of mRNA expression and miRNA families

For analysis of all experimentally tested samples an estimation of mRNA molecules per cell was assigned to 25,000 molecules at time 0 (prior to activation of the transcription inhibition protocol).

Ten of the highly expressed genes were selected from the top ranked list of mRNA. These genes were selected as being stable throughout the 24 hrs of the ActD protocol, along all four time points. These genes were used as anchor genes. According to their quantification and the total quantity of the gene expression distribution a correction was implemented based on these anchored genes. For the rest of the analysis, reported genes are those with an overall expression which is above the threshold of 5 molecules after the quantification correction procedure (>0.02% expression). For miRNA normalization we estimated 50,000 molecules per cells and only miRNAs with more than 1 molecule after quantification were considered. The identified miRNAs were compiled to their families. Within a miRNA family, we combine the expression of miRNAs that are marked by either 5p or 3p, as well as duplicated miRNAs that are annotated according to their genomic positions. From original list of 303 miRNAs we compiled a list of 250 miRNA families. This transformation was applied to the TargetScan scoring tables and the most significant score of miRNA representative was assigned to its family.

Probabilistic based miRNA-mRNA simulator

The simulator input are the number of molecules for the expression profiles of miRNAs (total 50k molecules) and mRNAs (total 25k molecules) in the specific cell type, and a table of miRNA-mRNA interaction prediction extracted from TargetScan. In addition, the simulator, called COMICS (COmpetition of MiRNA Interactions in Cell Systems) supports a wide set of configurable parameters: (i) the number of total miRNA; (ii) the number of mRNA molecules in the cell; (iii) the number of iterations for completing the run; (iv) the number of iteration interval between miRNA-mRNA binding event and the mRNA removal; (v) a random removal of unbounded mRNAs according to predetermined decay rate of the mRNA as extracted and extrapolated from experimental data of mRNA half-life; (vi) addition of newly transcribed mRNAs during a configurable number of iterations interval; (vii) miRNAs or genes overexpression according to a selected multiplication factor for the degree of overexpression. (vii) incorporation of alternative miRNA-target mapping. It is also possible to activate the simulator by a set of random genes as an initial state of pre-existing iterations prior to the simulation run.

In each run, a random miRNA is chosen from the predetermined available miRNAs distribution. Next, a target is chosen randomly according to the available targets distribution. mRNA that is already bounded by other miRNA molecules can be a putative target for the chosen miRNA, if the relevant binding site is not overlapping other occupied MBS on the same molecule. Overlapping binding sites are considered for neighboring MBS that are <50 nucleotides apart. Note that MBS that physically overlap in their sequence are already removed by TargetScan with the notion that overlapping sites cannot be occupied at the same time. A binding event will occur according to the miRNA-mRNA binding probability as extracted from TargetScan interaction table (or other prediction tables). The conversion of the interaction scores to the binding probabilities was done according to TargetScan score: $p = 1 - 2^{score}$. Upon a binding event, the free miRNA and mRNA distributions are updated, and the bounded mRNA molecules are marked as being occupied. An occupied molecule is removed after 1000 iterations following a successful binding event (a tunable parameter for halting an instant mRNA degradation). For mRNA to be eliminated, at least one MBS must be reported as occupied. During those iterations it is still available to bind other miRNAs in any of its non-overlapping binding sites. After mRNA removal, the bound miRNAs are released and return to the free miRNA pool and becomes eligible to engage in further binding events.

Overexpression scheme is based on multiplication of the available miRNA amount by all 7 factors (from x1 to x1000). This addition of miRNA molecules calls for calculating a new miRNA distribution while remaining with the same amount of miRNA in the cell. In case the miRNA had not detected in naïve cell, an arbitrary starting minimal amount of 0.01% (equivalent of 5 molecule/ cell).

Statistics and Bioinformatics

P-values were calculated using a paired and unpaired t-test, Fisher exact test, Kolmogorov Smirnov (KS) test or Chi-square tests. For testing the correspondence of two sets of different sizes, we have used the Jaccard score (J-score) that is the size of the intersection divided by the size of the union of the sample sets and it is range from 0 to 1 (no correspondence to a complete overlap, respectively). Statistical values are that are based on correlations were performed using standard Python statistical package. For annotation enrichment statistics and visualization Enrich (Kuleshov et al, 2016) was

used. For testing the effect of different background gene lists for the enrichments statistics we applied DAVID (Huang et al, 2007) clustering enrichment score is based on one tail Fisher exact corrected for the number of gene ontology annotation that are used. GOrilla ranked list enrichment score that is based on Hyper Geometric statistics for any selected background (Eden et al, 2009). Enrichment was performed in view of genes that are potential candidates for our analysis and against the set of genes that express with a minimum of 0.02% of the mRNA overall expression.

Acknowledgements

The study was supported in part by Supported by ERC grant 339096 on High Dimensional Combinatorics (N.L) and ELIXIR-Excelerate grant (M.L). We thank Tsiona Eliyau for supporting the experimental data for HeLa and HEK293. We thank Noam Shomron (Tel Aviv University) for sharing the miRNA expression plasmids used in this study. We thank the members of the Linial's lab for useful comments and discussions.

Author contributions

SMA, NL and ML contributed to the experimental design and leading the research for this study. All authors provided critical review. SMA developed the cell simulator, executed the results and conducted the data modeling. SMA wrote the computational code for the implementation of COMICS platform, SMA, NL and ML wrote the paper.

Conflict of interest

None. The authors declare that they have no conflict of interest.

Expanded View

Data availability

The mapped mRNAs and miRNAs for HeLa and HEK293 cells are listed in Datasets EV1 and EV2, respectively. The output of the % retention along 1M COMICS iterations at a 10k resolution from

HeLa as input is shown in Datasets EV3. The output of the % retention along 100k COMICS iterations at a 1k resolution is shown in Datasets EV4. List of the cross miRNA stable genes for three cell types is in Datasets EV4. The lists of miRNA expression levels for 3 cells as input for COMICS are in Dataset EV5. The Matrix of the miRNA retention levels for 248 miRNAs and 773 genes are in Dataset EV6. Lists of the stable genes for the three cells are in Dataset EV7. Lists of the stable genes for the three cells are in Dataset EV8. The results from enrichment tests for functional annotation for stable genes that are shared among 3 cell lines (48 genes) are in Dataset EV9. The results from enrichment tests for functional annotation for stable genes for 3 cell lines are in Dataset EV10. Appendix Text EV1 presents the alteration of the operational parameters of COMICS. Table EV1 summarize the statistics of characteristics for sensitive and stable sets for 3 cell lines. Appendix Fig S1 and Fig S2 show the time point correlations of the expression of miRNAs and mRNAs from HeLa and HEK293, respectively. Appendix Fig S3 shows the retention profile for hsa-mir-124a in HeLa cells. Appendix Fig S4 shows the results of repeated runs and randomization protocol for the probabilistic miRNA-mRNA interaction table. Appendix Fig S5 shows the results of alteration of COMICS parameters. Appendix Fig S6 shows the boxplots of four quantitative characteristics of the sensitive and stable sets for HEK293 and MCF-7 cell lines.

Figure Legends

Figure 1. Expression profiles of miRNA and mRNA under transcription arrest

A. Counting of miRNA miRNAs (left) and mRNAs (right) for the 4 different time points for HeLa (top) and HEK293 (bottom). The samples were collected at 0 hr, 2 hrs,8 hrs and 24 hrs following transcription inhibition by ActD. Source data is available in Dataset EV1 (HeLa) and Dataset EV2 (HEK293).
B. Expression of miRNAs (left) and mRNAs (right) in pairs of 4 different time points for HeLa (top) and HEK293 (bottom). The samples were collected at 0 hr, 2 hrs,8 hrs and 24 hrs following transcription inhibition by ActD. The expression is presented by

logarithmic scale (log_{10}). Spearman correlation (r) is listed for each pair along with the p-value of the significance. Source data is available in Dataset EV1 (HeLa) and Dataset EV2 (HEK293).

C. Relative abundance of each expressed mRNAs at four different time points for HeLa cells. At time 0, the relative abundance is set to 100%, and at each proceeding time points the abundance relative to time 0 is reported. Each line in the graph represent one gene (mRNA). Only genes with a minimal expression level of 0.02% expression are listed (equivalent to 97 fpkm, total of 860 genes). The bolded blue line represents the average of all reported genes at each time point.

D. Compilation of mRNA retention distribution (PDF) of all the reported genes after 8 hrs and 24 hrs from initiation of transcription inhibition by ActD. All genes with a retention level ≥100 are combined (at 100% retention).

Figure 2: Retention profile of mRNAs following overexpressing miRNAs in HeLa cells

A. Relative mRNA retention in HeLa cells that were transfected and overexpressed with hsa-mir-155. Measurement were taken at 4 time points as indicated. The retention plots are partitioned to target genes (i) (pink, left panels) and (ii) non-target genes (blue, middle panels). In (iii), the average retention patterns for of hsa-mir-155 targets (pink line) and non-targets (blue line) are shown.

B. Distribution of genes retention after 24 hrs from ActD treatment, according to their labels as targets (upper panel, pink) and non-targets (lower panel, blue). The plots compare the retention of genes from the control (smooth line), and from hsa-mir-155 overexpressed condition (dashed line). The number of genes that are included in the analyses are shown in parentheses. Target genes are marked by pink lines (top) and the non-target genes by blue lines (bottom). Note the shift in the distribution in the non-target genes towards the genes with higher retention level. All genes with a retention level ≥100 are shown as 100% retention.

Figure 3: Scheme of the COMICS platform and performance of the simulation process

A. A schematic view for a single iteration of COMICS simulation. The iteration step is repeated 100k times (unless indicated differently). After each successful interaction step (i.e. a valid binding of miRNA with MBS on mRNA molecule), the entire distributions of the miRNA and mRNA in the cells are updated. Therefore, the next iteration is slightly changed due to the refinement of mRNA composition and the availability of the free pool of miRNAs. On average, only 3% successful iterations occur from a total of 100k sampling events. The input for the simulator varies according to the profiles of the cell-type under study.

B. The outline of the major steps of COMICS operation from the mRNA perspective. The composition of the miRNA in the cells is obtained from the experimental measurement (at time point of 0 hr) but normalized to have 50k miRNAs and total of 25k mRNA. For HeLa cells, these are 3666 types of mRNAs that are included in the analysis (i.e. above a minimal threshold, see Materials and Methods). Sampling of the miRNA and mRNAs are done according to their distribution and the probability of the interaction is driven from the score of TargetScan MBS interaction scores (with 1.2 M values). A dashed mRNA shown after N iterations signifies an occupied transcript that is still halted (e.g., delay of 1k iterations) prior to its degradation, and releasing the bound miRNAs to the pool. In the end of the simulation, two sets of mRNAs are considered, an occupied set that is validated against miRNA-MBS pairs from CLASH data and the unoccupied mRNA set that is tested in view of the experimental results under transcriptional inhibition scheme (Fig 1-2).

C. The retention of HeLa expressed genes along COMICS simulation run. In this setting, 1M iteration steps were performed. For this input of COMICS simulation on HeLa cells, 3666 types of mRNA and 110 of miRNAs are included in the input. These numbers account for the 50k and 25k molecules of miRNAs and mRNAs, respectively. Each grey line represents the retention profile of a single type on mRNA. The blue line shows the average retention profile. After 1M iterations, the average retention level is 43.5%, Altogether, for 1M iterations, 11k successful events had been reported. For clarity of

presentation, only mRNA above a predefined expression are presented (see Materials and Methods). Total of 755 genes with a minimal expression of 0.02% are shown.

D. Validation of COMICS performance in view of the results from transcription arrest in HeLa (grey) and HEK293 cells (orange). At each of the indicated steps of the COMICS simulation run, the overlap in gene retention for the set of genes that remain stable (defined as >85% retention) was measured by calculating the Jaccard score. The statistical significance associated with the correspondence of the results (measured by p-value of the fisher exact test) are indicated by asterisks * <0.05 and **, <0.005.

E. Testing COMIC performance and dependency on the information in TargetScan interaction matrix. COMICS simulation performance in HEK293 was compared to the bounded pairs as reported from CLASH data on HEK293 (see Materials and Methods). The histogram shows the performance in term of the significant of the overlap of the reported COMICS results (100k iterations) using TargetScan probabilistic converted matrix (grey), and two versions of randomization for the interaction table (see Materials and Methods). The statistical test was based on the 251 genes that are reported as pairs miRNA-mRNA pairs by CLASH and expressed above the minimal expression threshold used for COMICS simulation protocol. The use of the TargetScan matrix shows significant results versus CLASH data (at the significant range p-value of e-4 to e-6). Applying any of the randomization for the miRNA-MBS interaction table, caused a drop in the performance to non-significant values.

Figure 4. Abundant miRNA and shuffling cell-specific miRNA profiles

A. Heatmap of miRNA from HeLa, HEK293 and MCF-7 cell-lines in view of the abundance of each miRNA (color in log scale) in each cell-type. The joint list of miRNAs includes the most abundant miRNAs that occupies 90% of the miRNA molecules (i.e. 45k out of 50k) in each cell-type. The fraction occupied by each of the listed miRNA, for each of the cell-type is available in Database EV5.

B. Pairs of miRNA and mRNA are shown according to their origin. Pearson correlation of the endpoint of the genes following 100k iterations and testing any of the genes that are above the minimal expression (5 molecules, 0.02% of mRNA). The number of mRNAs that are considered in the analyses are the 516 to the pair of HeLa and HEK293; 285 for the pair of HeLa and MCF-7 and 305 to the pair of HEK293 and MCF-7. The p-value of the correlations are very significant for all pairs. All p-values correlation values are < 1*e-15.

Figure 5. miRNA overexpression paradigm using COMICS platform

A. Schematic The relative percentage of each miRNA abundance is separated by vertical line. The miRNAs percentage is sorted from lowest (left side) to highest (right side). Seven different hsa-mir-155 over expression simulations are shown (x1, x3, x9, x18, x90, x300 and x1000) from bottom to top. The percentage of hsa-mir-155 is marked in pink.
B. Over expression simulation of hsa-mir-155 in eight different factors, where hsa-mir-155 target genes are marked in pink and all other non-target genes are marked in blue. The average retentions of both gene groups are plotted in bold lines.
C. Average final retention of the different simulation runs, using different overexpression factors of hsa-mir-155 as shown in B, of hsa-mir-155 target and non-target genes.

Figure 6: A scheme of the over expression matrix by a single factor

A. The different columns stand for the different pre-miRNA over-expressed by factor f, and the rows stand for the different genes.
B. Heatmap of the range of the retention for genes that were overexpressed at a factor x300. Each row is associated with a miRNA. The clustering is performed by the row (i.e. genes). The matrix includes 250 expressed miRNAs in HeLa cells.

C. Zoom in of a small section of the heatmap of the range of the retention for genes that were overexpressed at a factor x300. Each row is associated with a miRNA. The clustering is performed by the row (i.e. genes). The matrix includes 250 expressed miRNAs in HeLa cells

D. HeLa and HEK293 average final retention comparison in control simulation (no over expression). Each point stand for each gene

E. HeLa and HEK293 average final retention comparison. Each point stand for each gene in all 248 overexpression condition (each row in the heatmap presented in B, C) using over expression factor x300.

Figure 7: Comparison of sensitive and stable gene sets in different cell type

A. Overlap of the cross-miRNA stable genes in HeLa, HEK293 and MCF7 cells. Only genes that are expressed in at least two cells are listed. The gene list of the stable genes is available in Dataset EV6

B. Overlap of the cross-miRNA sensitive genes in HeLa, HEK293 and MCF7 cells. Only genes that are expressed in at least two cells are listed. The gene list of the stable genes is available in Dataset EV7

C. Partition of the stable genes expressed in HeLa, HEK293 and MCF-7 cells to their functional annotations: (i) small ribosomal subunit (18 genes), (ii) large ribosomal subunit (17 genes), (iii) cytoskeleton (5 genes), (iv) translation elongation (3 genes) (v) 4 aditional genes – NCL, ATP5J2, CALR and MIF. For detailed list see Dataset EV8.

D. Comparison of the number of targeting miRNA of sensitive genes, stable genes, and other (not sensitive and not stable in HeLa cells. Statistics of the comparisons are significant for the comparison of stable genes set and both sensitive and other gene sets (ttest p-values of 7.53e-11 and 6.48e-21, respectively), and no significant difference between sensitive and other gene sets. Full statistics are shown in Table EV1.

E. Comparison of the number of MBS of sensitive genes, stable genes, and other (not sensitive and not stable in HeLa cells. Statistics of the comparisons are significant for the comparison of stable genes set and both sensitive and other gene sets (ttest p-values of 2.07e-9 and 5.6e-

15, respectively), and no significant difference between sensitive and other gene sets. Full statistics are shown in Table EV1

F. Comparison of the initial abundance of sensitive genes, stable genes, and other (not sensitive and not stable in HeLa cells. Statistics of the comparisons are significant for the comparison of stable genes set and both sensitive and other gene sets (ttest p-values of 0.017 and 0.015, respectively), and no significant difference between sensitive and other gene sets. Full statistics are shown in Table EV1

G. Comparison of the of the average expression of the targeting miRNA of each gene, of sensitive genes, stable genes, and other (not sensitive and not stable in HeLa cells. Significant differences between all three gene sets were found (ttest p-values 2.52e-22, 3.07e-9 and 8.7e-11 for the comparison of stable-sensitive, stable-other and sensitive-other, respectively.

References

Agarwal V, Bell GW, Nam JW, Bartel DP (2015) Predicting effective microRNA target sites in mammalian mRNAs. *Elife* **4**

Ameres SL, Zamore PD (2013) Diversifying microRNA sequence and function. *Nat Rev Mol Cell Biol* **14:** 475-488

Arvey A, Larsson E, Sander C, Leslie CS, Marks DS (2010) Target mRNA abundance dilutes microRNA and siRNA activity. *Mol Syst Biol* **6:** 363

Balaga O, Friedman Y, Linial M (2012) Toward a combinatorial nature of microRNA regulation in human cells. *Nucleic Acids Res* **40:** 9404-9416

Berezikov E (2011) Evolution of microRNA diversity and regulation in animals. *Nat Rev Genet* **12:** 846-860

Bertoli G, Cava C, Castiglioni I (2015) MicroRNAs: New Biomarkers for Diagnosis, Prognosis, Therapy Prediction and Therapeutic Tools for Breast Cancer. *Theranostics* **5:** 1122-1143

Betel D, Koppal A, Agius P, Sander C, Leslie C (2010) Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol* **11:** R90

Bethune J, Artus-Revel CG, Filipowicz W (2012) Kinetic analysis reveals successive steps leading to miRNA-mediated silencing in mammalian cells. *EMBO Rep* **13:** 716-723

Bisognin A, Sales G, Coppe A, Bortoluzzi S, Romualdi C (2012) MAGIA(2): from miRNA and genes expression data integrative analysis to microRNA-transcription factor mixed regulatory circuits (2012 update). *Nucleic Acids Res* **40:** W13-21

Bloom RJ, Winkler SM, Smolke CD (2014) A quantitative framework for the forward design of synthetic miRNA circuits. *Nat Methods* **11:** 1147-1153

Bockmeyer CL, Christgen M, Muller M, Fischer S, Ahrens P, Langer F, Kreipe H, Lehmann U (2011) MicroRNA profiles of healthy basal and luminal mammary epithelial cells are distinct and reflected in different breast cancer subtypes. *Breast Cancer Res Treat* **130:** 735-745

Bosson AD, Zamudio JR, Sharp PA (2014) Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. *Molecular cell* **56:** 347-359

Broderick JA, Zamore PD (2014) Competitive endogenous RNAs cannot alter microRNA function in vivo. *Mol Cell* **54:** 711-713

Bu P, Chen K-Y, Chen JH, Wang L, Walters J, Shin YJ, Goerger JP, Sun J, Witherspoon M, Rakhilin N (2013) A microRNA miR-34a-regulated bimodal switch targets Notch in colon cancer stem cells. *Cell stem cell* **12:** 602-615

Chandradoss SD, Schirle NT, Szczepaniak M, MacRae IJ, Joo C (2015) A Dynamic Search Process Underlies MicroRNA Targeting. *Cell* **162:** 96-107

Chekulaeva M, Filipowicz W (2009) Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. *Curr Opin Cell Biol* **21:** 452-460

Cheung TH, Quach NL, Charville GW, Liu L, Park L, Edalati A, Yoo B, Hoang P, Rando TA (2012) Maintenance of muscle stem-cell quiescence by microRNA-489. *Nature* **482:** 524-528

Chi SW, Hannon GJ, Darnell RB (2012) An alternative mode of microRNA target recognition. *Nat Struct Mol Biol* **19:** 321-327

Denzler R, Agarwal V, Stefano J, Bartel DP, Stoffel M (2014) Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Molecular cell* **54:** 766-776

Denzler R, McGeary SE, Title AC, Agarwal V, Bartel DP, Stoffel M (2016) Impact of MicroRNA Levels, Target-Site Complementarity, and Cooperativity on Competing Endogenous RNA-Regulated Gene Expression. *Mol Cell* **64:** 565-579

Ebert MS, Sharp PA (2012) Roles for microRNAs in conferring robustness to biological processes. *Cell* **149:** 515-524

Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z (2009) GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10:** 48

Eichhorn SW, Guo H, McGeary SE, Rodriguez-Mias RA, Shin C, Baek D, Hsu SH, Ghoshal K, Villen J, Bartel DP (2014) mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Mol Cell* **56:** 104-115

Erhard F, Haas J, Lieber D, Malterer G, Jaskiewicz L, Zavolan M, Dolken L, Zimmer R (2014) Widespread context dependency of microRNA-mediated regulation. *Genome Res* **24:** 906-919

Filipowicz W, Bhattacharyya SN, Sonenberg N (2008) Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature reviews genetics* **9:** 102

Friard O, Re A, Taverna D, De Bortoli M, Cora D (2010) CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse. *BMC Bioinformatics* **11:** 435

Friedman Y, Karsenty S, Linial M (2014) miRror-Suite: decoding coordinated regulation by microRNAs. *Database (Oxford)* **2014**

Gaur A, Jewell DA, Liang Y, Ridzon D, Moore JH, Chen C, Ambros VR, Israel MA (2007) Characterization of microRNA expression levels and their biological correlates in human cancer cell lines. *Cancer research* **67:** 2456-2468

Ha M, Kim VN (2014) Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol* **15:** 509-524

Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jr., Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141:** 129-141

Hausser J, Zavolan M (2014) Identification and consequences of miRNA-target interactions--beyond repression of gene expression. *Nature Reviews Genetics* **15:** 599

He M, Liu Y, Wang X, Zhang MQ, Hannon GJ, Huang ZJ (2012) Cell-type-based analysis of microRNA profiles in the mouse brain. *Neuron* **73:** 35-48

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

Herranz H, Cohen SM (2010) MicroRNAs and gene regulatory networks: managing the impact of noise in biological systems. *Genes Dev* **24:** 1339-1344

Huang DW, Sherman BT, Tan Q, Collins JR, Alvord WG, Roayaei J, Stephens R, Baseler MW, Lane HC, Lempicki RA (2007) The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. *Genome Biol* **8:** R183

Janas MM, Wang B, Harris AS, Aguiar M, Shaffer JM, Subrahmanyam YV, Behlke MA, Wucherpfennig KW, Gygi SP, Gagnon E, Novina CD (2012) Alternative RISC assembly: binding and repression of microRNA-mRNA duplexes by human Ago proteins. *RNA* **18:** 2041-2055

Jens M, Rajewsky N (2015) Competition between target sites of regulators shapes post-transcriptional gene regulation. *Nature Reviews Genetics* **16:** 113

Jo MH, Shin S, Jung SR, Kim E, Song JJ, Hohng S (2015) Human Argonaute 2 Has Diverse Reaction Pathways on Target RNAs. *Mol Cell* **59:** 117-124

Jonas S, Izaurralde E (2015) Towards a molecular understanding of microRNA-mediated gene silencing. *Nature reviews Genetics* **16:** 421

Kim S, Hwang DW, Lee DS (2009) A study of microRNAs in silico and in vivo: bioimaging of microRNA biogenesis and regulation. *FEBS J* **276:** 2165-2174

Klein M, Chandradoss SD, Depken M, Joo C (2017) Why Argonaute is needed to make microRNA target search fast and reliable. *Semin Cell Dev Biol* **65:** 20-28

Kozomara A, Griffiths-Jones S (2013) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic acids research* **42:** D68-D73

Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44:** W90-97

Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129:** 1401-1414

Li J-H, Liu S, Zhou H, Qu L-H, Yang J-H (2013) starBase v2. 0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. *Nucleic acids research* **42:** D92-D97

Li JH, Liu S, Zhou H, Qu LH, Yang JH (2014a) starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Res* **42:** D92-97

Li X, Cassidy JJ, Reinke CA, Fischboeck S, Carthew RW (2009) A microRNA imparts robustness against environmental fluctuation during development. *Cell* **137:** 273-282

Li Y, Liang C, Wong KC, Jin K, Zhang Z (2014b) Inferring probabilistic miRNA-mRNA interaction signatures in cancers: a role-switch approach. *Nucleic Acids Res* **42:** e76

Liang Y, Ridzon D, Wong L, Chen C (2007) Characterization of microRNA expression profiles in normal human tissues. *BMC genomics* **8:** 166

Lopez-Maury L, Marguerat S, Bahler J (2008) Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat Rev Genet* **9:** 583-593

Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, Sweet-Cordero A, Ebert BL, Mak RH, Ferrando AA, Downing JR, Jacks T, Horvitz HR, Golub TR (2005) MicroRNA expression profiles classify human cancers. *Nature* **435:** 834-838

Moore MJ, Scheel TK, Luna JM, Park CY, Fak JJ, Nishiuchi E, Rice CM, Darnell RB (2015) miRNA-target chimeras reveal miRNA 3'-end pairing as a major determinant of Argonaute target specificity. *Nat Commun* **6:** 8864

Nam JW, Shin KR, Han J, Lee Y, Kim VN, Zhang BT (2005) Human microRNA prediction through a probabilistic co-learning model of sequence and structure. *Nucleic Acids Res* **33:** 3570-3581

Nitzan M, Steiman-Shimony A, Altuvia Y, Biham O, Margalit H (2014) Interactions between distant ceRNAs in regulatory networks. *Biophys J* **106:** 2254-2266

Olejniczak SH, La Rocca G, Gruber JJ, Thompson CB (2013) Long-lived microRNA-Argonaute complexes in quiescent cells can be activated to regulate mitogenic responses. *Proc Natl Acad Sci U S A* **110:** 157-162

Pasquinelli AE (2012) MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat Rev Genet* **13:** 271-282

Pelaez N, Carthew RW (2012) Biological robustness and the role of microRNAs: a network perspective. *Curr Top Dev Biol* **99:** 237-255

Peterson SM, Thompson JA, Ufkin ML, Sathyanarayana P, Liaw L, Congdon CB (2014) Common features of microRNA target prediction tools. *Frontiers in genetics* **5**

Pinzon N, Li B, Martinez L, Sergeeva A, Presumey J, Apparailly F, Seitz H (2017) microRNA target prediction programs predict many false positives. *Genome Res* **27:** 234-245

Rajewsky N (2006) microRNA target predictions in animals. *Nature genetics* **38:** S8

Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP (2011) A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* **146:** 353-358

Schmiedel JM, Klemm SL, Zheng Y, Sahay A, Bluthgen N, Marks DS, van Oudenaarden A (2015) Gene expression. MicroRNA control of protein expression noise. *Science* **348:** 128-132

Seok H, Ham J, Jang ES, Chi SW (2016) MicroRNA Target Recognition: Insights from Transcriptome-Wide Non-Canonical Interactions. *Mol Cells* **39:** 375-381

Shenoy A, Blelloch RH (2014) Regulation of microRNA function in somatic stem cell proliferation and differentiation. *Nature Reviews Molecular Cell Biology* **15:** 565

Stark A, Brennecke J, Bushati N, Russell RB, Cohen SM (2005) Animal MicroRNAs confer robustness to gene expression and have a significant impact on 3′ UTR evolution. *Cell* **123:** 1133-1146

Tan SM, Kirchner R, Jin J, Hofmann O, McReynolds L, Hide W, Lieberman J (2014) Sequencing of captive target transcripts identifies the network of regulated genes and functions of primate-specific miR-522. *Cell Rep* **8:** 1225-1239

Tay Y, Rinn J, Pandolfi PP (2014) The multilayered complexity of ceRNA crosstalk and competition. *Nature* **505:** 344-352

Thomas M, Lieberman J, Lal A (2010) Desperately seeking microRNA targets. *Nature structural and molecular biology* **17:** 1169

Thomson DW, Dinger ME (2016) Endogenous microRNA sponges: evidence and controversy. *Nat Rev Genet* **17:** 272-283

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7:** 562-578

Treiber T, Treiber N, Plessmann U, Harlander S, Daiß J-L, Eichner N, Lehmann G, Schall K, Urlaub H, Meister G (2017) A compendium of RNA-binding proteins that regulate MicroRNA biogenesis. *Molecular cell* **66:** 270-284. e213

Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, Visone R, Iorio M, Roldo C, Ferracin M, Prueitt RL, Yanaihara N, Lanza G, Scarpa A, Vecchione A, Negrini M, Harris CC, Croce CM (2006) A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc Natl Acad Sci U S A* **103:** 2257-2261

Volinia S, Galasso M, Costinean S, Tagliavini L, Gamberoni G, Drusco A, Marchesini J, Mascellani N, Sana ME, Abu Jarour R, Desponts C, Teitell M, Baffa R, Aqeilan R, Iorio MV, Taccioli C, Garzon R, Di Leva G, Fabbri M, Catozzi M et al (2010) Reprogramming of miRNA networks in cancer and leukemia. *Genome Res* **20:** 589-599

Wang D, Zhang Z, O'Loughlin E, Lee T, Houel S, O'Carroll D, Tarakhovsky A, Ahn NG, Yi R (2012) Quantitative functions of Argonaute proteins in mammalian development. *Genes Dev* **26:** 693-704

Wen J, Parker BJ, Jacobsen A, Krogh A (2011) MicroRNA transfection and AGO-bound CLIP-seq data sets reveal distinct determinants of miRNA action. *Rna* **17:** 820-834

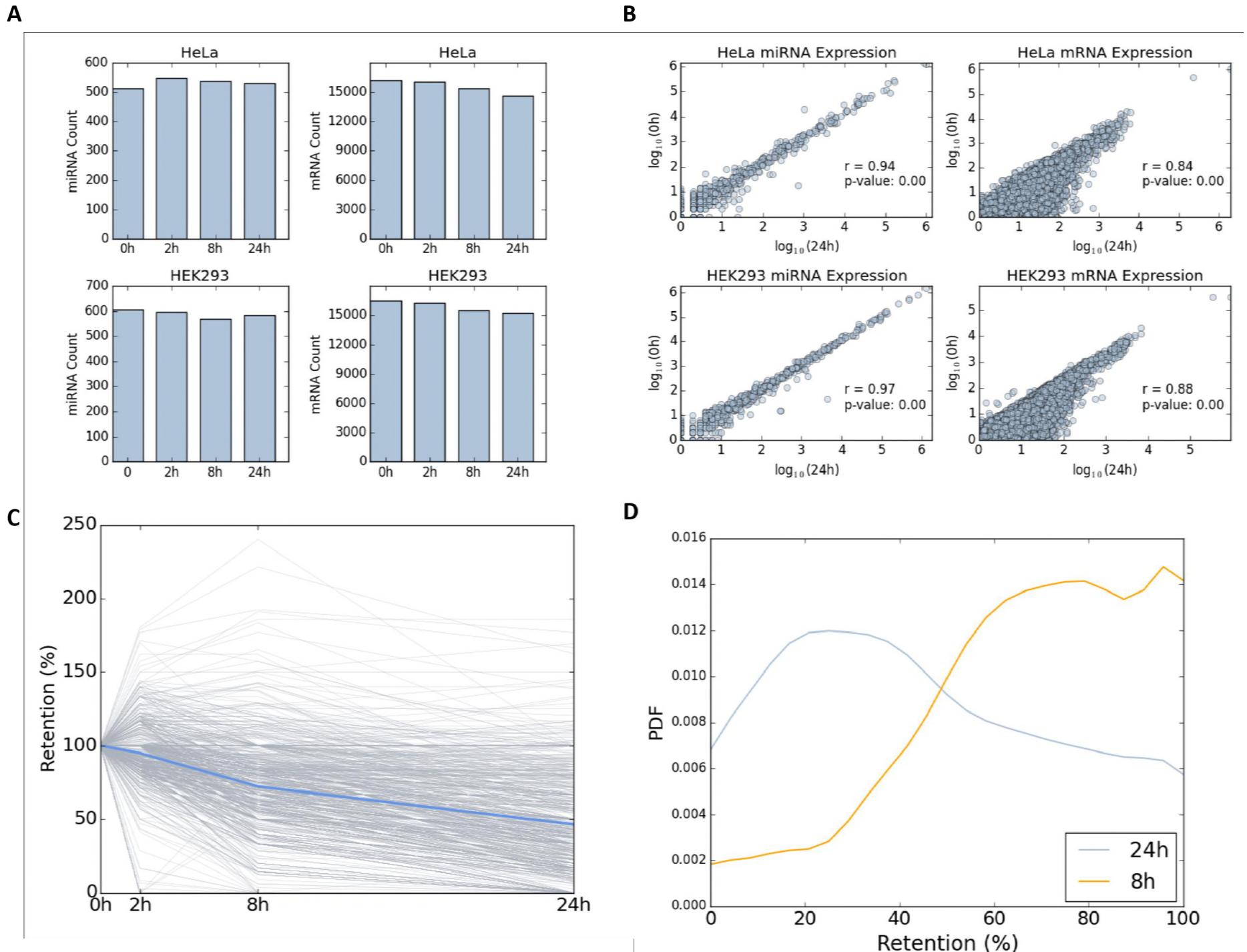
Winter J, Diederichs S (2011) Argonaute proteins regulate microRNA stability: Increased microRNA abundance by Argonaute proteins is due to microRNA stabilization. *RNA Biol* **8:** 1149-1157

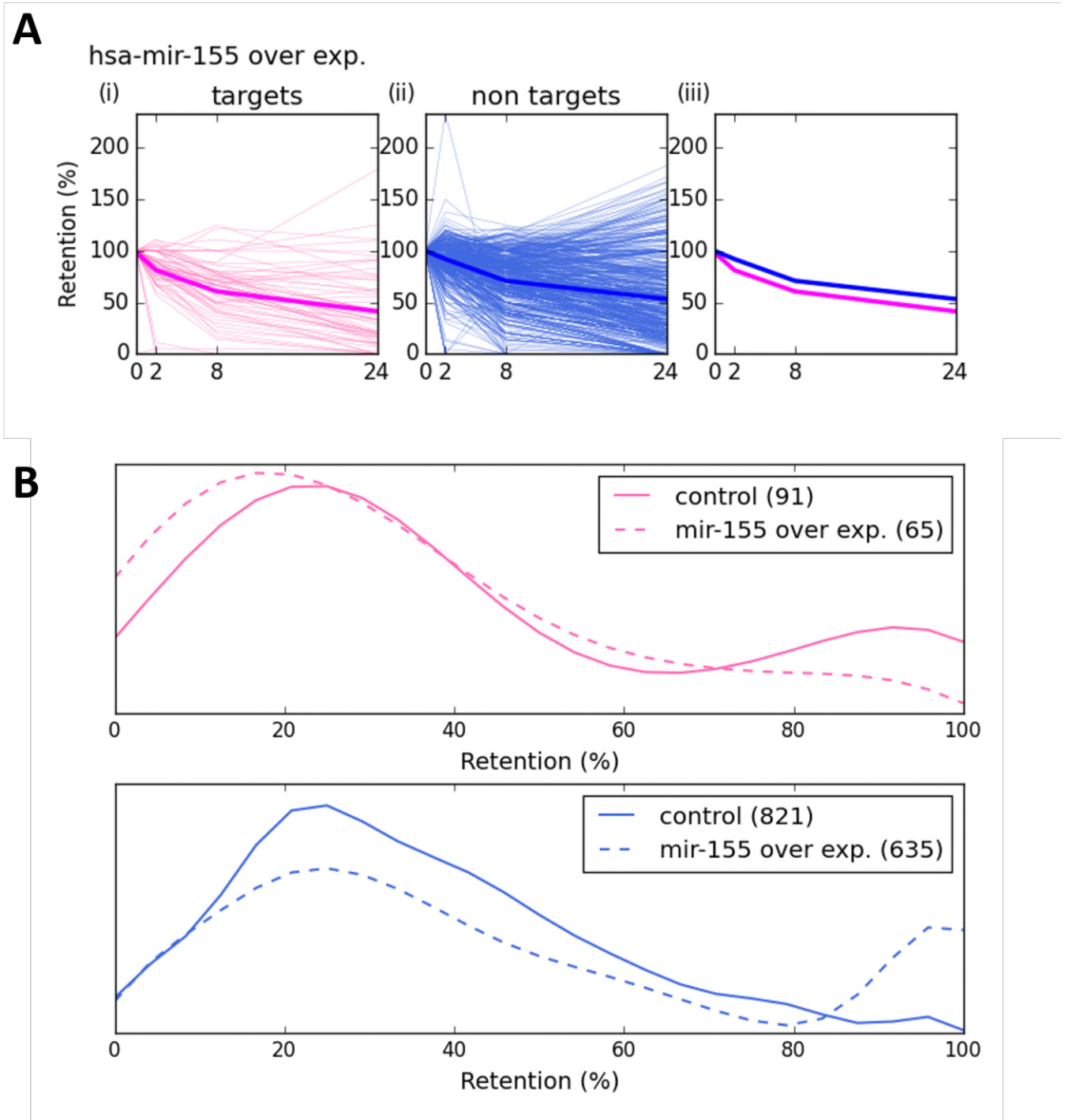
Yang QE, Racicot KE, Kaucher AV, Oatley MJ, Oatley JM (2013) MicroRNAs 221 and 222 regulate the undifferentiated state in mammalian male germ cells. *Development* **140:** 280-290

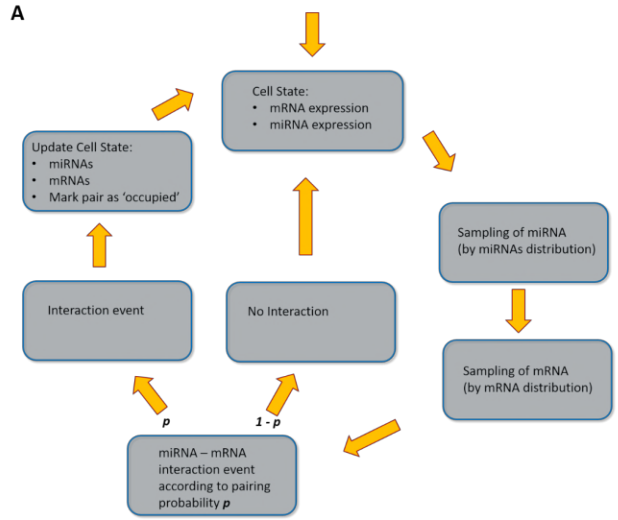
Yuan Y, Liu B, Xie P, Zhang MQ, Li Y, Xie Z, Wang X (2015a) Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. *Proceedings of the National Academy of Sciences* **112:** 3158-3163

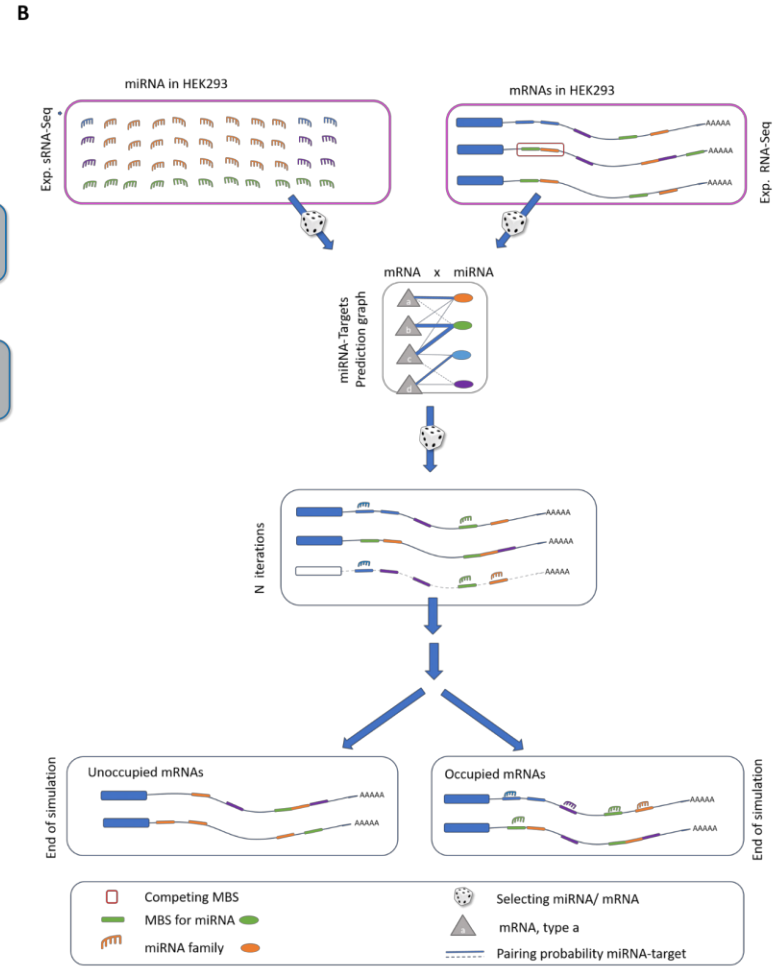
Yuan Y, Liu B, Xie P, Zhang MQ, Li Y, Xie Z, Wang X (2015b) Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. *Proc Natl Acad Sci U S A* **112:** 3158-3163

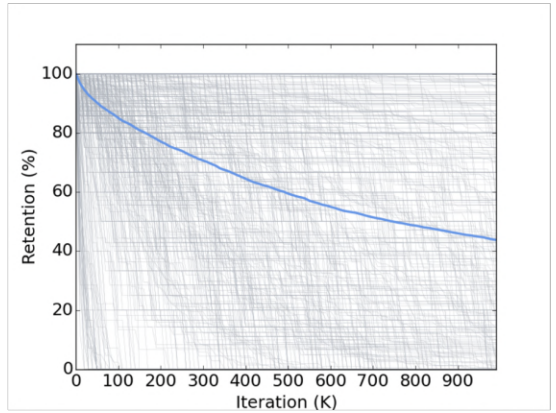
Zhang Y, Wei W, Cheng N, Wang K, Li B, Jiang X, Sun S (2012) Hepatitis C virus-induced up-regulation of microRNA-155 promotes hepatocarcinogenesis by activating Wnt signaling. *Hepatology* **56:** 1631-1640

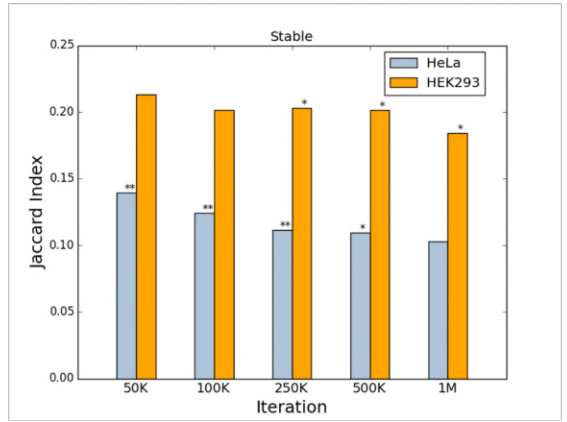

A
HeLa
miRNA Count
mRNA Count
HEK293
0h
2h
8h
24h
B
HeLa miRNA Expression
HeLa mRNA Expression
HEK293 miRNA Expression
HEK293 mRNA Expression
log10(0h)
log10(24h)
r = 0.94
p-value: 0.00
r = 0.84
p-value: 0.00
r = 0.97
p-value: 0.00
r = 0.88
p-value: 0.00
C
Retention (%)
D
PDF
Retention (%)
24h
8h

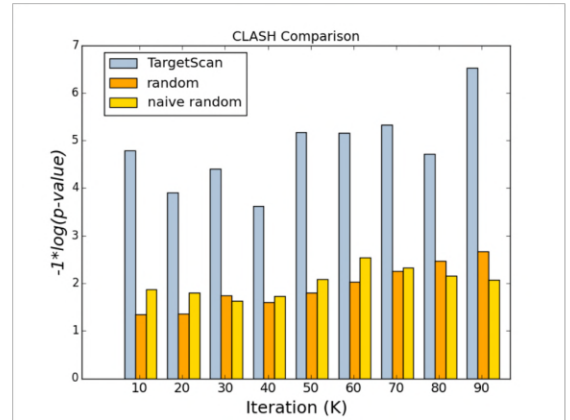

A
hsa-mir-155 over exp.
(i)
targets
(ii)
non targets
(iii)
Retention (%)
200
150
100
50
0
0 2
8
24
B
control (91)
mir-155 over exp. (65)
0
20
40
60
80
100
Retention (%)
control (821)
mir-155 over exp. (635)
Retention (%)


A
Cell State:
• mRNA expression
• miRNA expression
Update Cell State:
• miRNAs
• mRNAs
• Mark pair as 'occupied'
Sampling of miRNA
(by miRNAs distribution)
Sampling of mRNA
(by mRNA distribution)
Interaction event
No Interaction
p
1 - p
miRNA – mRNA
interaction event
according to pairing
probability p

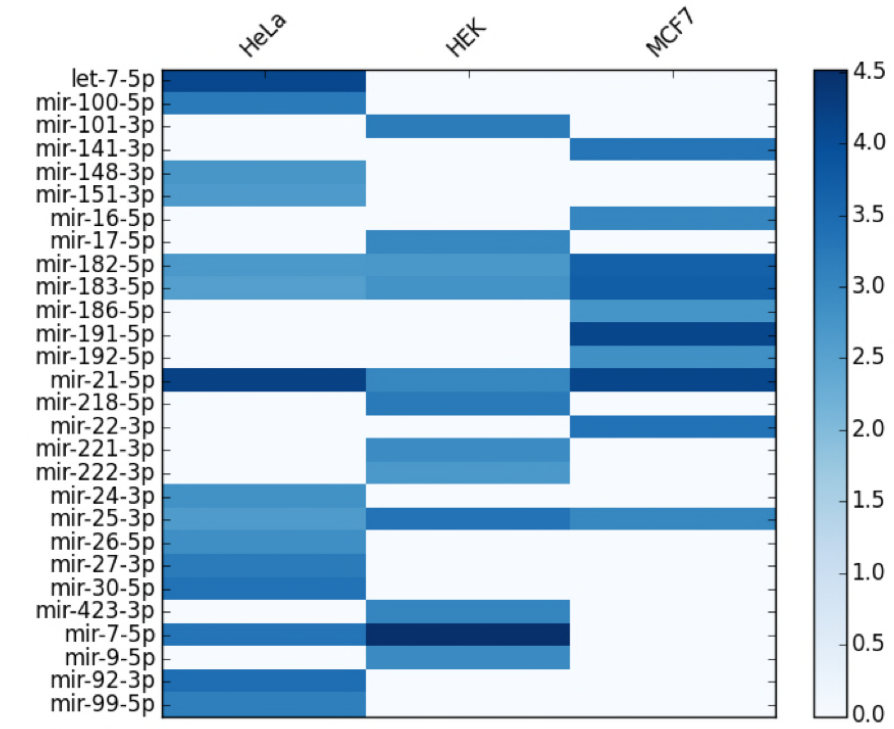

B
miRNA in HEK293
mRNAs in HEK293
Exp. sRNA-Seq
Exp. RNA-Seq
mRNA x miRNA
miRNA-Targets
Prediction graph
N iterations
Unoccupied mRNAs
Occupied mRNAs
End of simulation
End of simulation
Competing MBS
MBS for miRNA
miRNA family
Selecting miRNA/ mRNA
mRNA, type a
Pairing probability miRNA-target


C
Retention (%)
Iteration (K)

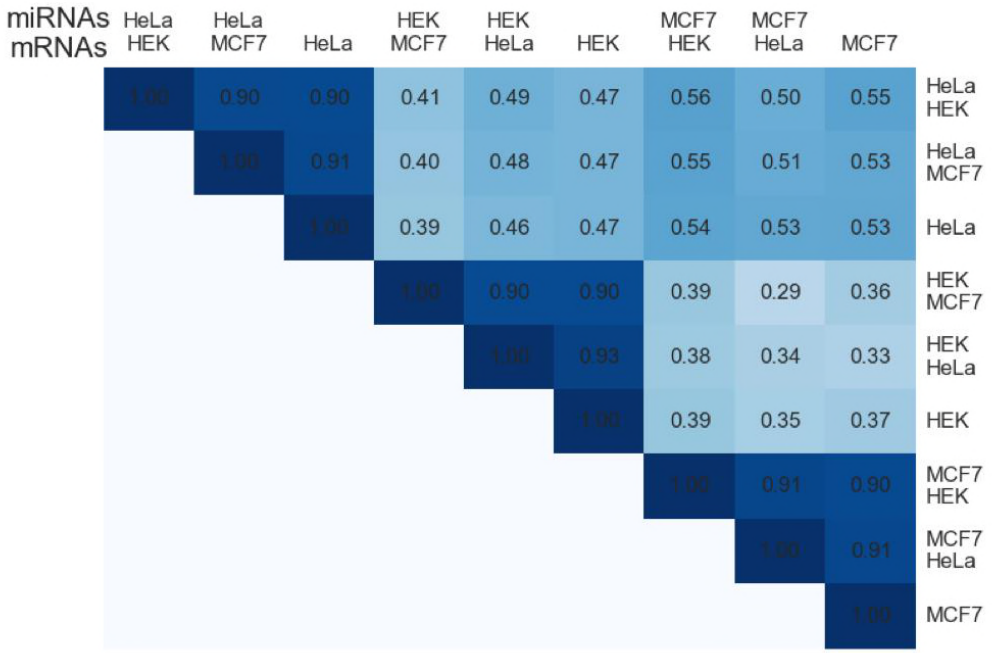

D
Stable
HeLa
HEK293
Jaccard Index
Iteration

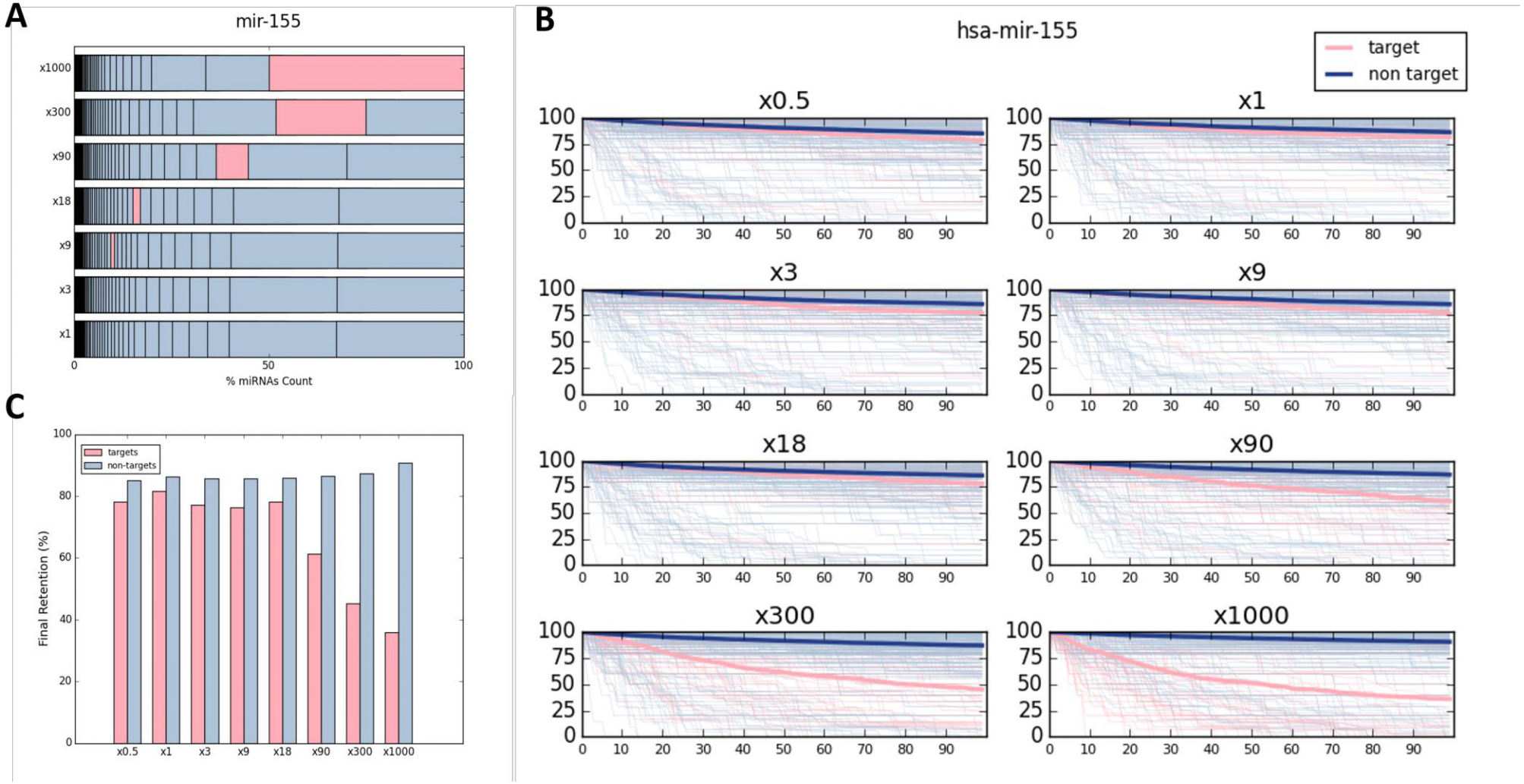

E
CLASH Comparison
TargetScan
random
naive random
-1*log(p-value)
Iteration (K)

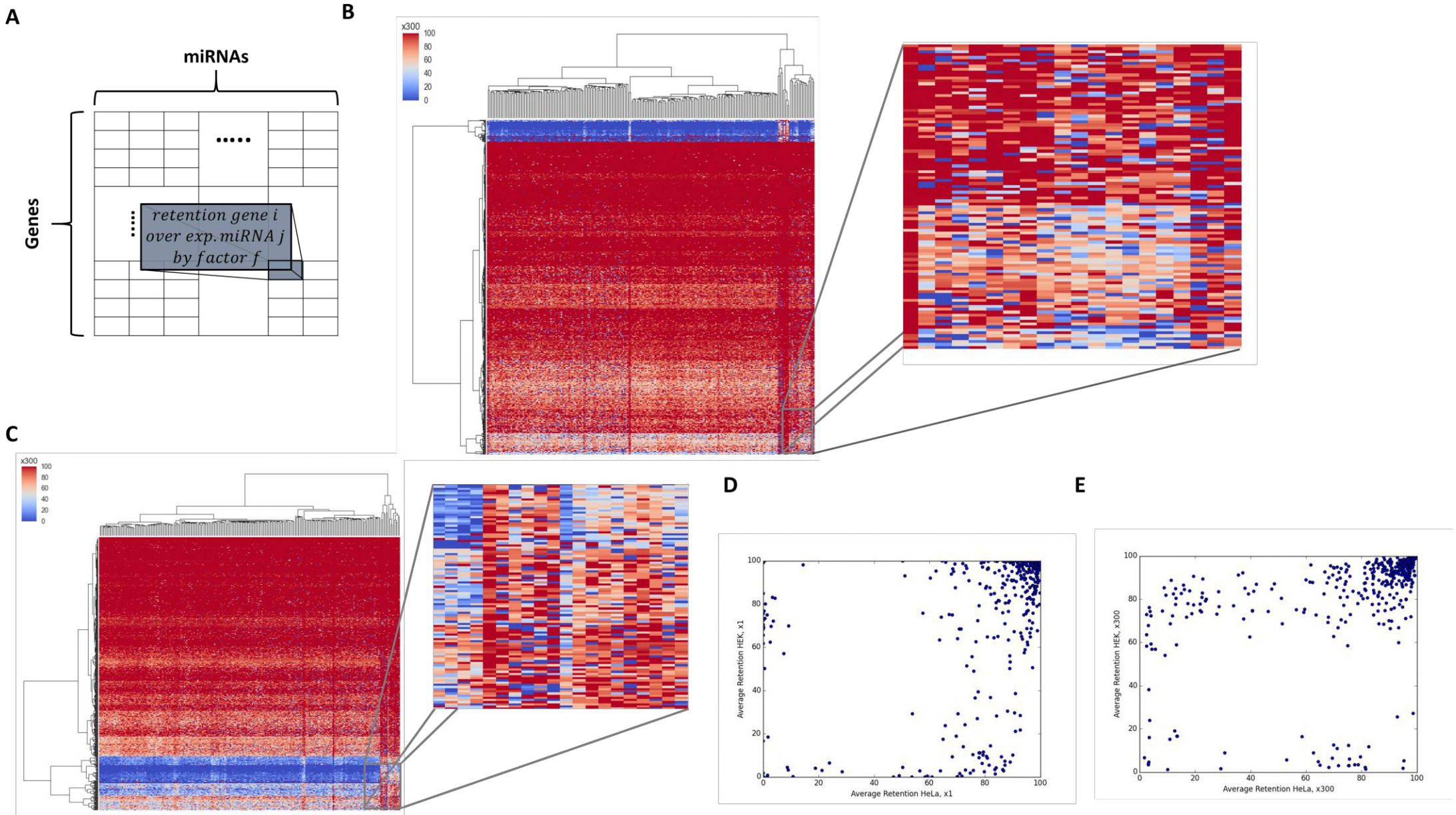
A

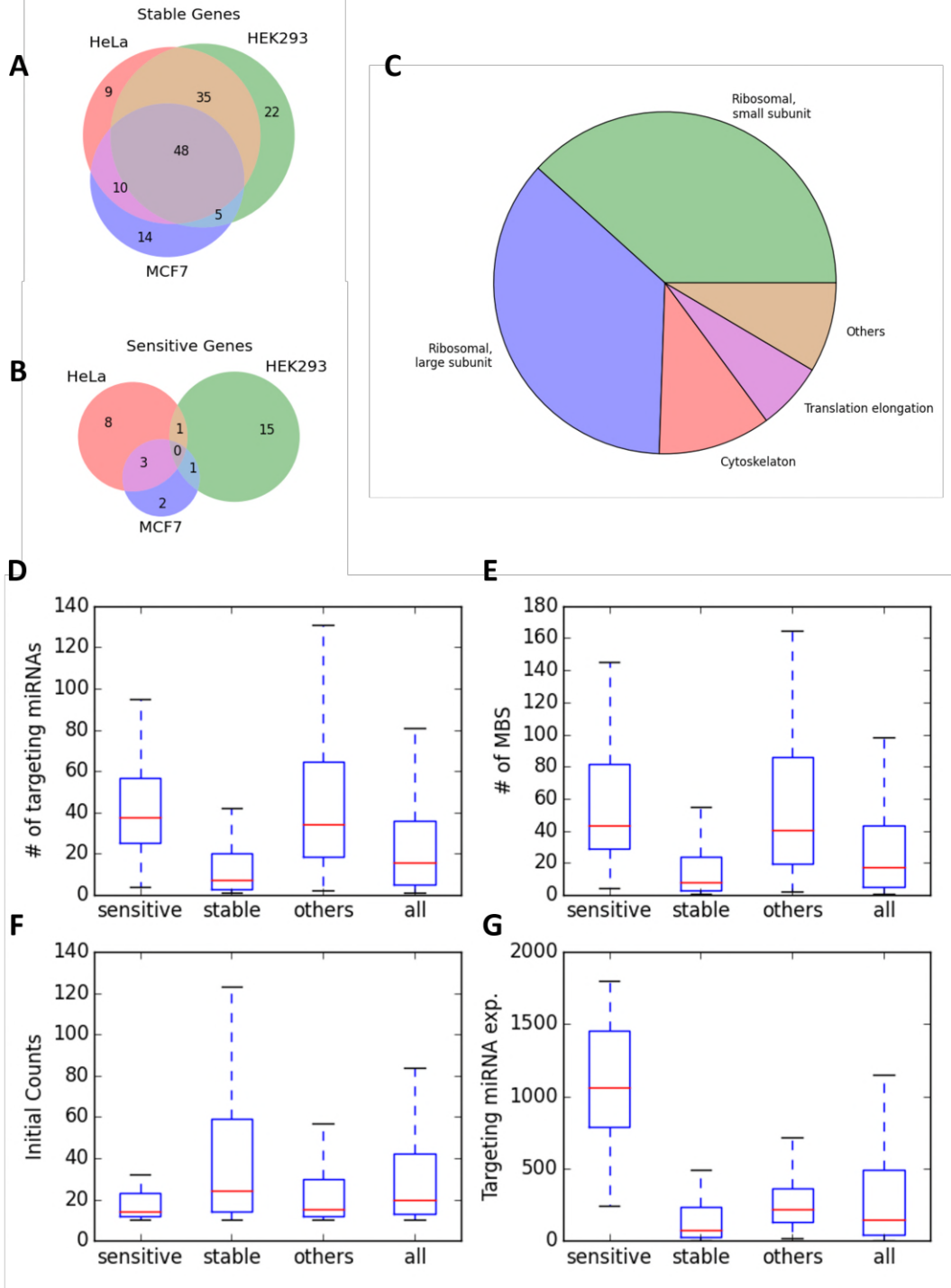

HeLa
HEK
MCF7
let-7-5p
mir-100-5p
mir-101-3p
mir-141-3p
mir-148-3p
mir-151-3p
mir-16-5p
mir-17-5p
mir-182-5p
mir-183-5p
mir-186-5p
mir-191-5p
mir-192-5p
mir-21-5p
mir-218-5p
mir-22-3p
mir-221-3p
mir-222-3p
mir-24-3p
mir-25-3p
mir-26-5p
mir-27-3p
mir-30-5p
mir-423-3p
mir-7-5p
mir-9-5p
mir-92-3p
mir-99-5p
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0

B


miRNAs
mRNAs
HeLa HEK
HeLa MCF7
HeLa
HEK MCF7
HEK HeLa
HEK
MCF7 HEK
MCF7 HeLa
MCF7
1.00 0.90 0.90 0.41 0.49 0.47 0.56 0.50 0.55 HeLa HEK
1.00 0.91 0.40 0.48 0.47 0.55 0.51 0.53 HeLa MCF7
1.00 0.39 0.46 0.47 0.54 0.53 0.53 HeLa
1.00 0.90 0.90 0.39 0.29 0.36 HEK MCF7
1.00 0.93 0.38 0.34 0.33 HEK HeLa
1.00 0.39 0.35 0.37 HEK
1.00 0.91 0.90 MCF7 HEK
1.00 0.91 MCF7 HeLa
1.00 MCF7


A
mir-155
x1000
x300
x90
x18
x9
x3
x1
0
50
100
% miRNAs Count
B
hsa-mir-155
target
non target
x0.5
x1
x3
x9
x18
x90
x300
x1000
C
targets
non-targets
Final Retention (%)
x0.5
x1
x3
x9
x18
x90
x300
x1000


A
miRNAs
Genes
retention gene i
over exp. miRNA j
by factor f
B
x300
100
80
60
40
20
0
C
x300
100
80
60
40
20
0
D
Average Retention HEK, x1
Average Retention HeLa, x1
0
20
40
60
80
100
E
Average Retention HEK, x300
Average Retention HeLa, x300
0
20
40
60
80
100


A
Stable Genes
HeLa
HEK293
MCF7
9
35
22
48
10
5
14
B
Sensitive Genes
HeLa
HEK293
MCF7
8
1
15
0
3
1
2
C
Ribosomal, small subunit
Ribosomal, large subunit
Others
Translation elongation
Cytoskelaton
D
# of targeting miRNAs
E
# of MBS
F
Initial Counts
G
Targeting miRNA exp.
sensitive
stable
others
all