Supplementary Information

Principles of microRNA regulation of a human cellular signaling network

Qinghua Cui[1+], Zhenbao Yu[2+], Enrico O. Purisima[1], and Edwin Wang[1*]

[1]*Computational Chemistry & Biology Group and* [2]*Mammalian Cell Genetics Group, Biotechnology Research Institute, National Research Council Canada, Montreal, Quebec, H4P 2R2, Canada*

[+]These authors contributed equally to this work

[*]To whom correspondence should be addressed. Email: edwin.wang@cnrc-nrc.gc.ca

Table of Contents

Supplementary Text

Genome-wide analysis of miRNA target abundance of ligands, receptors and transcription factors

We extracted all the ligands from human genome using a NCBI gene information file (ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/) and obtained 94 ligands (Supplementary Text file 17). We extracted all the receptors from human genome based on Gene Ontology (GO) terms (ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/) and obtained 1,511 receptors (Supplementary Text file 18). We obtained 1,795 transcription factors (TFs, Supplementary Text file 19) from human genome using the file downloaded from the paper published by Messina et al. [Messina et al., 2004]. We found that miRNA targets are significantly enriched in the TFs of the genome (35.7%, 629/1795, $P < 2x10^{-4}$). To identify human signaling TFs, we intersected the 1,795 TFs and a comprehensive manually curated human signaling pathway dataset (Supplementary Text file 20). Here we defined the TFs that are documented as components of signaling pathways as signaling TFs. To generate Supplementary Text file 20, we downloaded a file containing manually curated pathway proteins from a recent report [Subramanian et al., 2005]. Since this dataset contains many other pathways, such as metabolic pathways, we manually checked to keep the signaling pathway proteins only and removed other types of pathway proteins. We extracted TFs (Supplementary Text file 21) and intracellular signaling proteins (Supplementary Text file 22) from Supplementary Text file 20. The TFs found in Supplementary Text file 20 are considered as signaling TFs. We mapped miRNA targets to these ligands, receptors, intracellular signaling proteins and signaling TFs and

calculated the fractions of miRNA targets in each group (Supplementary Text files 17, 18, 22, 23).

We found that miRNA targets represent 9.6%, 19.1% and 52.5% of ligands, receptors and signaling TFs, respectively ($P < 2x10^{-4}$). These results are similar to those found in the network. The finding of the high abundance of miRNA targets in signaling TFs triggered us to ask if miRNAs more frequently target signaling TFs than other TFs. To answer this question, we mapped miRNA targets onto the 276 signaling TFs and the total TFs. We found that the abundance of miRNA targets in signaling TFs (52.5%) is 1.5 times more than that in the total TFs (35.7%, $P < 2x10^{-4}$). This result suggests that miRNAs preferentially regulate signaling TFs over other TFs. Signaling TFs are important in responding to extracellular signals and play a vital role in cellular phenotypic changes. Therefore, tight regulation of signaling TFs by miRNAs may facilitate more robust signal transitions.

Network themes formed from the network motifs regulated by a single miRNA

To find network themes that are regulated by a single miRNA, we collected the network motifs that contain at least one target of the miRNA and examined whether these motifs could aggregate into clusters using Pajek software (http://vlado.fmf.uni-lj.si/pub/networks/pajek/). Supplementary Text file 24 summarized the number and size of themes regulated by each miRNA. In general, the network motifs regulated by each miRNA formed 1 or 2 network themes. The sizes of the network themes range from 4 to 145 nodes. Most of network themes contain more than 20 nodes. For example, the motifs

that contain the targets of miR-1 form a single theme with 111 nodes (Supplementary Figure 3)

In the network, there are five output cellular machines: transcription machinery, translation machinery, secretion apparatus, motility machinery and electrical response. Each cellular machine contains 12 to 38 output nodes. Since network themes can be tied to specific biological processes [Zhang et al., 2005], we attempted to associate the network themes that are regulated by each miRNA to the output cellular machines of the signaling network. Toward this end, we first identified the overlapping nodes between each cellular machine and the network themes that are regulated by each miRNA. We then calculated of the percentage of the overlapping nodes to each cellular machine. The associations between cellular machines and the network themes regulated by a single miRNA are presented by these percentages. The statistical significance of the associations is inferred by randomization tests. We defined the test is to be statistically significant, when a P value is less or equal to 0.10. The results are summarized in Supplementary Text file 25. Nearly 60% of miRNAs in this study could associate with one or more cellular machines of the signaling network. We further constructed a "heatmap" to represent the associations between the network themes that are regulated by each miRNA and the cellular machines of the signaling outputs (Supplementary Figure 4). These results also infer associations between miRNAs and signaling cellular machines.

Sensitivity analysis

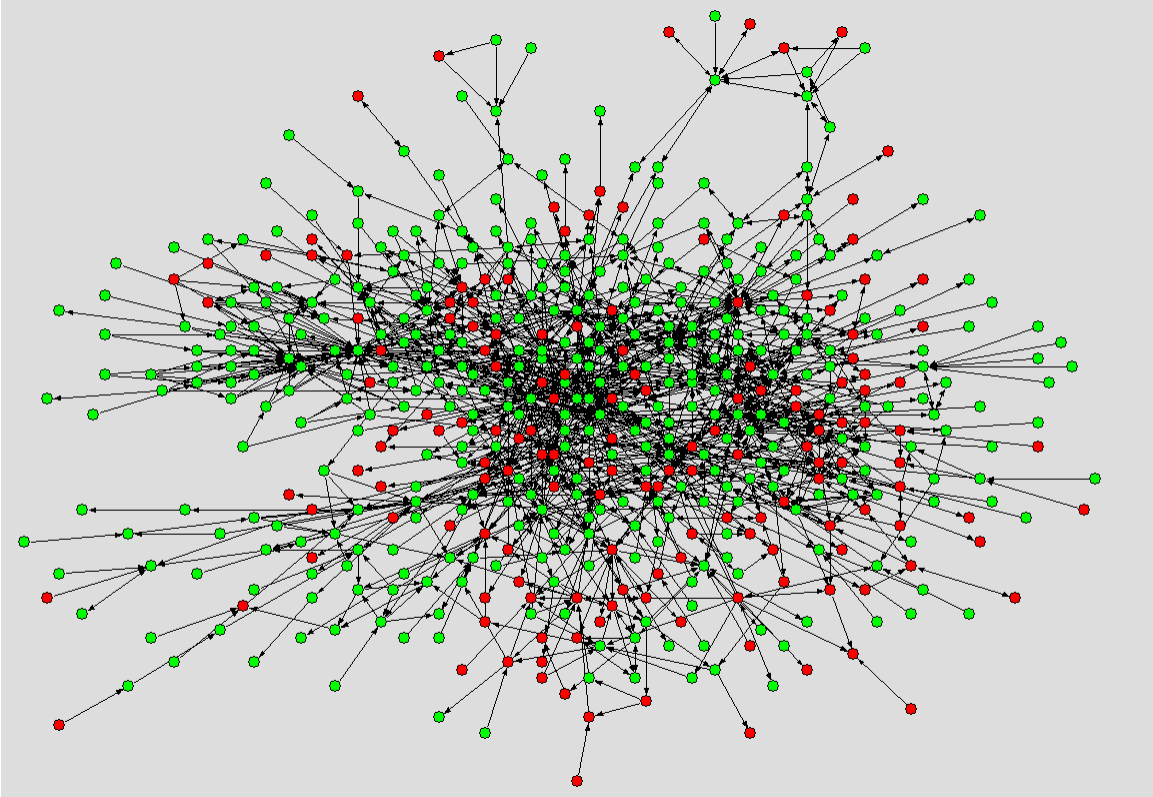
As approximately 12% of the predicted miRNA targets might not be real targets, we analyzed the potential effects of false positives and false negatives on the observations. To examine the effects of false positives, we randomly added 10% and 20% of the network proteins that are not predicted miRNA targets to the miRNA target list and performed the same analyses. To examine the effects of false negatives, we randomly removed 10% and 20% of the miRNA targets and performed the same analyses. The results show that, in general, the trends remained unchanged by the addition of either false positives or false negatives (Supplementary Tables IV, V, VI, VII). Except for the miRNA target fraction in the downstream components of adaptors, in which the P value remained significant when adding 10% of false positives ($P = 0.042$), but changed to 0.149 when adding 20% of false positives (Supplementary Table V), the P values remained significant for all other observations when adding 20% of false positives (Supplementary Tables IV, VI, VII). For example, addition of 20% of false positives caused an increase of the fraction of miRNA targets in the whole network from 29.4% to 35.4% and meanwhile an increase of the fraction of miRNA targets of the ligands and the nuclear signaling proteins from 9.1% to 15.2% and from 50.0% to 52.6%, respectively (Supplementary Table IV). Similarly, addition of false negatives did not have obvious effect on the observations and the P values remained significant (Supplementary Tables IV, V, VI and VII). Therefore, the observations are robust against substantial errors.

Supplementary Discussion

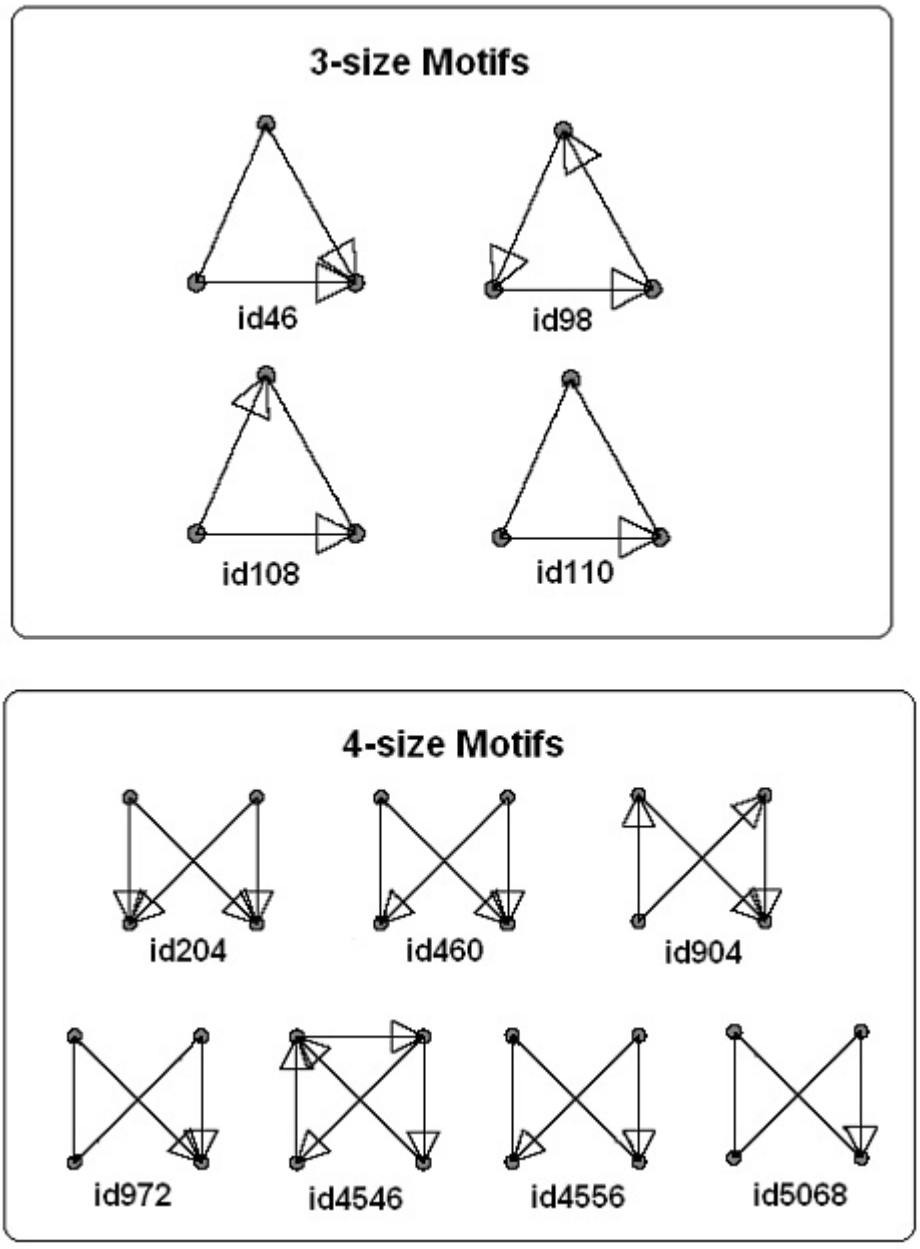
MiRNA regulation of signaling information flow

The extent and duration of activation of a given signaling process may determine if a cell maintains homeostatic behavior or undergoes cell phenotypic and state changes. To maintain tolerance and homeostasis, mammalian cells have the capability to filter out noisy stimuli from extracellular environments and precisely terminate on-going cellular signaling through negative regulation of signal transduction. To efficiently avoid unwanted and auto-activation of cell surface receptors, negative regulatory signaling events selectively occur in the vicinity of the plasma membrane (the upstream region of a signaling network). Indeed, previous work indicated that negative feedback loops are abundant in the upstream section of the signaling pathways, and with positive feedback loops becoming more abundant as the signal progresses along the information flow [Ma'ayan et al., 2005]. This indicates that the early steps of the information flows may have built-in controls to limit signal propagation. Therefore, weak or short-lived signals (noise) may not penetrate into the network because of the early barrier posed by the abundance of negative feedback loops. Cellular signaling can be also suppressed by down-regulation of the concentration of signaling components, for example, by miRNAs. The fact that miRNA targets are enriched in the later steps and negative feedback loops are enriched in the early steps of the signaling information flows indicates that the role of the negative regulation of intracellular signaling by miRNAs is different from that by network inhibitory links. One of the main functions of network inhibitory links is to maintain cell tolerance to noisy extracellular stimuli and spontaneous auto-activation by blocking the signaling in the upstream of the network, whereas the main function of miRNAs is probably to maintain cell homeostasis through terminating on-going signaling at the downstream and the late stages of signaling information flows.

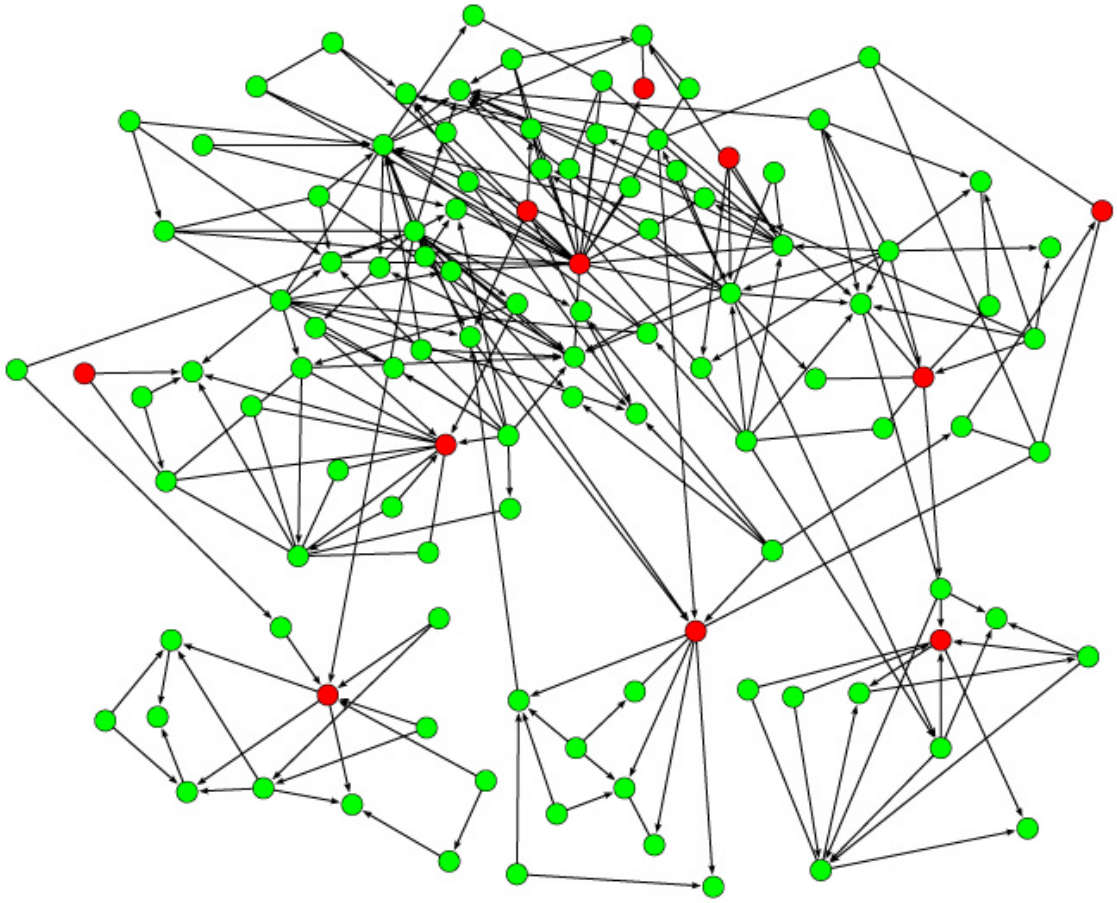
Supplementary Figures



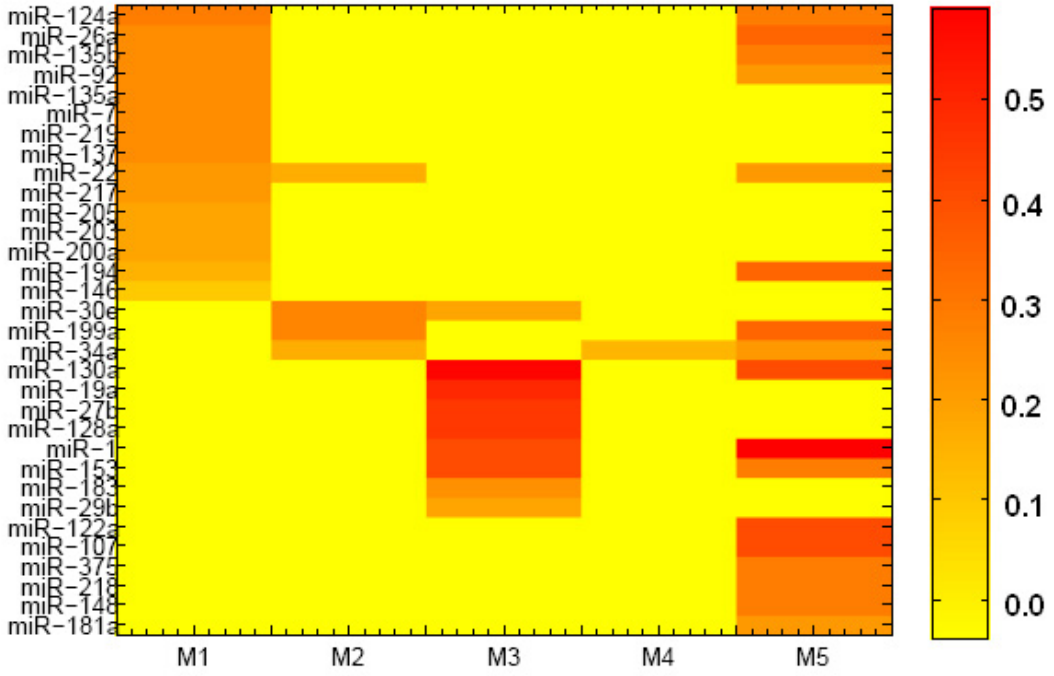
Supplementary Figure 1 Visualization of the signaling network using the Pajek software. The network is visualized by placing nodes as different colors based on whether they are miRNA targets. Red nodes represent miRNA targets and green nodes are not miRNA targets.



Supplementary Figure 2 Schematic diagram of the motifs identified within the signaling network. Network motifs were identified using Mfinder program. The number under each motif is the motif ID number of the corresponding motif. Non-directional links represent neutral links and directional links represent positive links or negative links. The network motif ID numbering system is from Alon's motif dictionary (http://www.weizmann.ac.il/mcb/UriAlon/NetworkMotifsSW/mfinder/motifDictionary.pdf).



Supplementary Figure 3 An example of network theme formed from all of the motifs containing the targets of a miRNA, miR-1. This theme is composed of 111 network components. The red nodes are miRNA targets.



Supplementary Figure 4 A heatmap illustrating the associations between cellular machines and the network themes regulated by each miRNA. Rows represent miRNAs while columns represent signaling cellular machines: M1 (Transcription), M2 (Translation), M3 (Secretion apparatus), M4 (Motility), and M5 (Electrical response).
To associate the network themes regulated by a single miRNA to cellular machines, we first identified the overlapping nodes between each cellular machine and the network themes that are regulated by the miRNA. The associations are presented by the percentage of the overlapping nodes to the output nodes of each cellular machine. The statistical significance of the associations is inferred by randomization tests. We defined the test is to be statistically significant, when a P value is less or equal to 0.10. A darker color represents a higher degree of association.

Supplementary Tables

Supplementary Table I Abundances of miRNA targets in the upstream components of high- and low-link adaptors

	Number of upstream components	Number of miRNA targets	Fraction of miRNA targets	P value
High-link group	84	26	31.0%	0.33
Low-link group	76	26	34.2%	

Note: High-link group contains adaptors linking to more than four upstream components, while low-link group contains adaptors linking to four or less upstream components.

Supplementary Table II MiRNA-targeted scaffold proteins in the motif id110

Scaffold protein	miRNAs
CRK	miR-1, miR-10a, miR-126, miR-133a, miR-20, miR-93
PSD95	miR-1, miR-107, miR-122a
SAM68	miR-122a, miR-203, miR-204, miR-218
SHC	miR-10a, miR-204, miR-219, miR-9
SNAP25	miR-1, miR-128a, miR-130a, miR-153
YOTIAO	miR-133a

Supplementary Table III Fractions of miRNA targets in the shortest paths of each cellular machine

Cellular machines	Shortest paths from receptors to output nodes of cellular machines				
	Transcription	**Translation**	**Secretion apparatus**	**Motility**	**Electrical response**
Number of nodes	220	160	180	226	185
Number of targets	59	32	35	55	37
Fraction of miRNA targets	0.268	0.200	0.194	0.243	0.200
P value	0.116	0.0008	0.0002	0.009	0.0008

Note: P values were given by comparing the fraction of miRNA targets in each cellular machine to the fraction of miRNA targets (0.294) in the network using randomization tests.

Supplementary Table IV Effect of false positives and false negatives on the fractions of miRNA targets in different cellular locations

	False Negatives –10% targets	False Negatives –20% targets	Natural	False Positives +10% targets	False Positives +20% targets
Cellular Location	**Fraction of miRNA targets (P value)**	**Fraction of miRNA targets (P value)**	**Fraction of miRNA targets (P value)**	**Fraction of miRNA targets (P value)**	**Fraction of miRNA targets (P value)**
Whole Network	23.5%	26.5%	29.4 %	32.4%	35.4%
Ligands	9.1% (0.003)	9.1% (0.003)	9.1% (0.004)	9.1% (0.0006)	15.2% (0.0006)
Cell Surface Receptors	16.3% (0.006)	15.0% (0.015)	18.8% (0.016)	23.8% (0.022)	25% (0.010)
Intracellular Central Signaling Proteins	27.6% (0.139)	24.6% (0.131)	31.2% (0.08)	33.8% (0.111)	37.3% (0.052)
Nuclear Proteins	50.0% (0.0008)	42.1% (0.003)	50.0% (0.006)	55.3% (0.001)	52.6% (0.008)

Note: The whole network nodes were sorted into four subgroups, as shown in the table. False positives and false negatives of miRNA targets were randomly added and the fractions of miRNA targets and P values were recalculated. P values were given by comparing the fractions of miRNA targets in an individual subgroup with the fraction of miRNA targets in whole network using randomization tests.

Supplementary Table V Effect of false positives and false negatives on the fractions of miRNA targets in the downstream components of the high-link group and the low-link group of adaptors

	False Negatives – 20% targets	False Negatives –10% targets	Natural	False Positives +10% targets	False Positives +20% targets
Downstream components	**Fraction of miRNA targets**	**Fraction of miRNA targets**	**Fraction of miRNA targets**	**Fraction of miRNA targets**	**Fraction of miRNA targets**
High link	31.5%	31.5%	36.1%	38.0%	43.5%
Low link	23.1%	23.1%	24.2%	28.6%	37.4%
P value	0.069	0.057	0.015	0.042	0.149

Note: High-link group contains adaptors linking to more than four downstream components, while low-link group contains adaptors linking to four or less downstream components. False positives and false negatives of miRNA targets were randomly added and the fractions of miRNA targets and P values were recalculated.

Supplementary Table VI Effect of false positives and false negatives on the relations between the abundance of positive links and the abundance of miRNA targets in network motifs.

	False Negatives –20% targets	False Negatives –10% targets	Natural	False Positives +10% targets	False Positives +20% targets
P value 1	0.007	0.007	0.004	0.008	0.005
P value 2	0.018	0.018	0.010	0.006	0.010

Note: False positives and false negatives of miRNA targets were randomly added and the fractions of miRNA targets and P values were recalculated. P value 1 represents the significance of less-enrichment of positive links in the subgroups with 0 miRNA targets. P value 2 represents the significance of enrichment of positive links in the subgroups whose nodes are all miRNA targets. The number of positive links and ratio were given in Supplementary Text File S26.

Supplementary Table VII Effect of false positives and false negatives on the fraction of miRNA targets of the common nodes that locate in the shortest paths of all five cellular machines

	False Negatives – 20% targets	False Negatives –10% targets	Natural	False Positives +10% targets	False Positives +20% targets
Fraction of miRNA targets	14.3%	14.3%	14.3%	15.7%	27.1%
P value	0.020	0.005	0.0002	0.0005	0.056

Note: The P values were given by comparing the fraction of miRNA targets of the common nodes to that of miRNA targets in the network using randomization tests. False positives and false negatives of miRNA targets were randomly added. The fractions of miRNA targets and P values were recalculated.

List of supporting data files and Java source code files*

File name	Description
Text file S1	List of network nodes
Text file S2	List of network links
Text file S3	Overlapped miRNA targets used in this study
Text file S4	Network nodes which are miRNA targets
Text file S5	Network nodes which are not miRNA targets
Text file S6	Pajek format text file that was used to visualize the network
Text file S7	List of each subgroup (classified according to cellular locations) nodes
Text file S8	Up- and down- stream components and link numbers for all the adaptors
Text file S9	High-link group and low-link group of the downstream proteins for all the adaptors
Text file S10	Input file of Mfinder program used in this study
Text file S11	Summary information of 3-size motifs outputted by MFinder
Text file S12	List of 3-size motifs outputted by MFinder
Text file S13	Summary information of 4-size motifs outputted by MFinder
Text file S14	List of 4-size motifs outputted by MFinder
Text file S15	Common proteins of five cellular machine shortest paths
Text file S16	MiRNA targets of common proteins listed in Text file S15
Text file S17	List of ligands extracted from human genome
Text file S18	List of receptors extracted from human genome
Text file S19	List of TFs extracted from human genome
Text file S20	List of signaling proteins from manually curated signaling pathways
Text file S21	List of signaling TFs
Text file S22	MiRNA targets in intracellular signaling proteins
Text file S23	MiRNA targets in signaling TFs
Text file S24	Number and size of the themes regulated by each miRNA
Text file S25	The statistical significance of the associations between cellular machines and the network themes regulated by each miRNA

Text file S26	The number and ratio of positive links in each subgroup of motifs upon false positive and negative miRNA target data
Zip file S1	MiRNA target and link information of 3-size motifs
Zip file S2	MiRNA target and link information of 4-size motifs
Zip file S3	Subgroup data (classified according to the number of miRNA targets) of 3-size motifs
Zip file S4	Subgroup data (classified according to the number of miRNA targets) of 4-size motifs
Zip file S5	Five cellular machine output node proteins
Zip file S6	Proteins (nodes) on the shortest paths (from receptors to the output nodes of each cellular machine)
Zip file S7	Shortest path nodes without input and output nodes
Java file S1	Java source code used to map miRNA targets to the network
Java file S2	Java source code used to classify motifs based on the number of miRNA targets
Java file S3	Java source code used to find shortest paths
Java file S4	Java source code used to perform randomization tests

*All data and Java source files can be downloaded at
http://www.bri.nrc.ca/wang/mirna2.html

Supplementary Reference List

Ma'ayan A, Jenkins SL, Neves S, Hasseldine A, Grace E, Dubin-Thaler B, Eungdamrong NJ, Weng G, Ram PT, Rice JJ, Kershenbaum A, Stolovitzky GA, Blitzer RD, Iyengar R (2005) Formation of regulatory patterns during signal propagation in a Mammalian cellular network. *Science* 309: 1078-1083

Messina DN, Glasscock J, Gish W, Lovett M (2004) An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. *Genome Res* 14: 2041-2047

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102: 15545-15550

Zhang LV, King OD, Wong SL, Goldberg DS, Tong AH, Lesage G, Andrews B, Bussey H, Boone C, Roth FP (2005) Motifs, themes and thematic maps of an integrated Saccharomyces cerevisiae interaction network. *J Biol* 4: 6