Monday 17 lecture

We have learnt that several small sequences in cds and 3'UTR of mRNA may regulate the turn-over rate.
We have also examined a paper in which scientists sopught to identify all mRNAs bound "in vivo" by an RBP (PUM) that possibly regulate mRNA translation, osing RBP immunoprecipitation and microarray analysis.

One of the main features of RNA cis-elements is that very often they are very small sequences with low conservation (PUM element actully represents an exception, rather than a rule).

Conversely, very oftene RBPs contain multiple RNA-binding domain, suggesting cooperative action of repeated small sequences.

Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP

Markus Hafner,[1,5] Markus Landthaler,[1,4,5] Lukas Burger,[2] Mohsen Khorshid,[2] Jean Hausser,[2] Philipp Berninger,[2] Andrea Rothballer,[1] Manuel Ascano, Jr.,[1] Anna-Carina Jungkamp,[1,4] Mathias Munschauer,[1] Alexander Ulrich,[1] Greg S. Wardle,[1] Scott Dewell,[3] Mihaela Zavolan,[2,*] and Thomas Tuschl[1,*]

[1]Howard Hughes Medical Institute, Laboratory of RNA Molecular Biology, The Rockefeller University, 1230 York Avenue, Box 186, New York, NY 10065, USA

[2]Biozentrum der Universität Basel and Swiss Institute of Bioinformatics (SIB), Klingelbergstr. 50-70, CH-4056 Basel, Switzerland

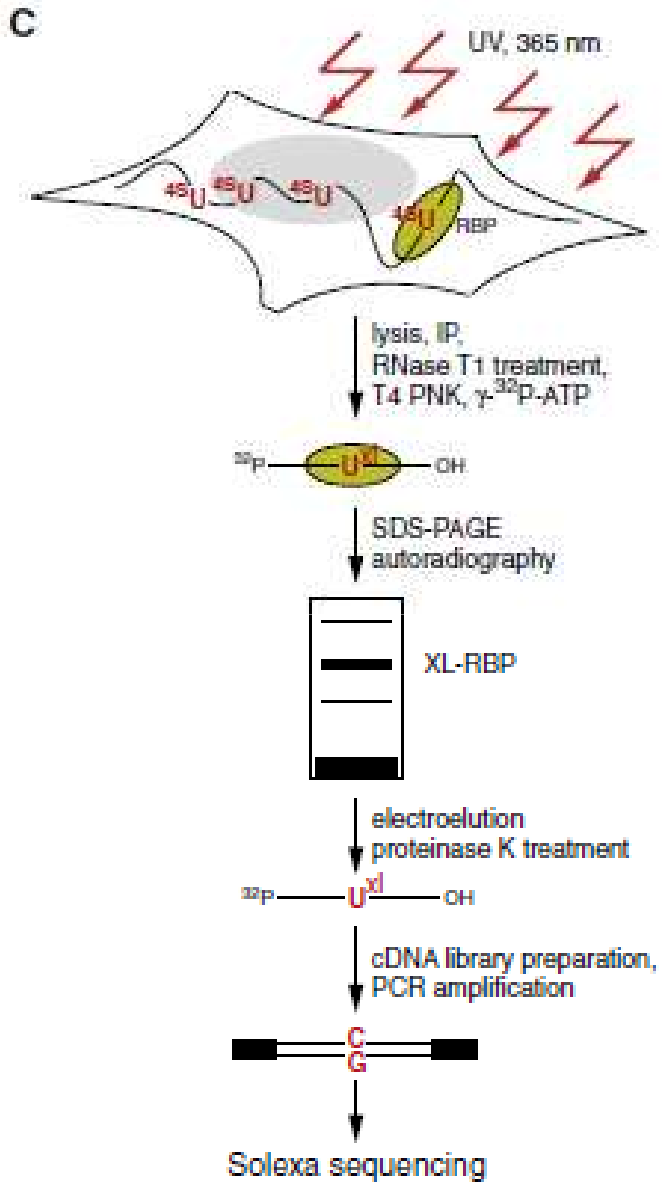
[3]Genomics Resource Center, The Rockefeller University, 1230 York Avenue, Box 241, New York, NY 10065, USA

[4]Present address: Berlin Institute for Medical Systems Biology, Max-Delbrück-Center for Molecular Medicine, 13125 Berlin, Germany

[5]These authors contributed equally to this work

*Correspondence: mihaela.zavolan@unibas.ch (M.Z.), ttuschl@rockefeller.edu (T.T.)

RNA transcripts are subject to posttranscriptional gene regulation involving hundreds of RNA-binding proteins (RBPs) and microRNA-containing ribonucleoprotein complexes (miRNPs) expressed in a cell-type dependent fashion. We developed a cell-based crosslinking approach to determine at high resolution and transcriptome-wide the binding sites of cellular RBPs and miRNPs. The crosslinked sites are revealed by thymidine to cytidine transitions in the cDNAs prepared from immunopurified RNPs of 4-thiouridine-treated cells. We determined the binding sites and regulatory consequences for several intensely studied RBPs and miRNPs, including PUM2, QKI, IGF2BP1-3, AGO/EIF2C1-4 and TNRC6A-C. Our study revealed that these factors bind thousands of sites containing defined sequence motifs and have distinct preferences for exonic versus intronic or coding versus untranslated transcript regions. The precise mapping of binding sites across the transcriptome will be critical to the interpretation of the rapidly emerging data on genetic variation between individuals and how these variations contribute to complex genetic diseases.



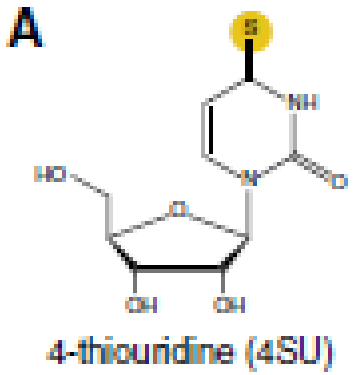


Figure 1. PAR-CLIP Methodology.

(A) Structure of photoactivatable nucleosides.

(C) Illustration of PAR-CLIP. 4SU-labeled transcripts were crosslinked to RBPs and partially RNase-digested RNA-protein complexes were immunopurified and size-fractionated. RNA molecules were recovered and converted into a cDNA library and deep sequenced.

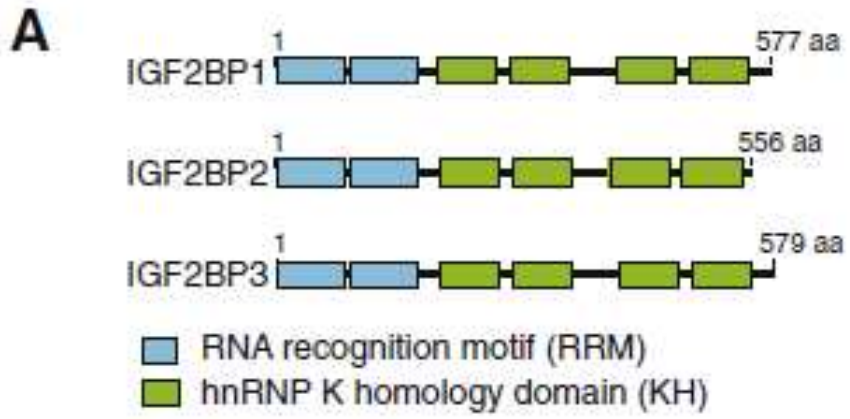


Figure 4. RNA Recognition by the IGF2BP Protein Family

(A) Domain structure of IGF2BP1-3 proteins.

(C) Alignments of IGF2BP1 PAR-CLIP cDNA sequence reads to the corresponding regions of the 30UTRs of EEF2 and MRPL9 transcripts. Red bars indicate the 4-nt IGF2BP1 recognition motif and nucleotides marked in red indicate T to C sequence changes.

(D) Sequence logo of the IGF2BP1-3 RRE generated by PhyloGibbs analysis of the top 100 sequence read clusters.

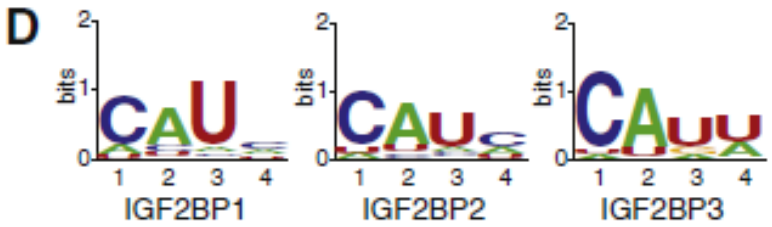
C

3'UTR of EEF2

```
GGGGCCCGCTGCGTGCCATCACTCAACCATAACACTTGATGCCGTTTCTTTC # reads error
---------------CCATCACTCAACCACAACACTTG-------------- 40      1
---------------CCATCACTCAACCATAACACTTG-------------- 21      0
---------------CCATCACCCAACCATAACACTTG-------------- 14      1
---------------CCACCACTCAACCATAACACTTG-------------- 11      1
---------------CCATCACTCAACCATAACACCTG-------------- 7       1
```

3'UTR of MRPL9

```
TGTCTCCAGTACTTGCCTCATTCTCATCATCCAAACTGAACATTTGTATCCC # reads error
---------------CCTCATTCTCATCACCCAAACTG-------------- 18      1
---------------CCTCACTCTCATCATCCAAACTG-------------- 15      1
---------------CCTCATTCTCATCATCCAAACTG-------------- 11      0
---------------CCTCATTCTCACCATCCAAACTG-------------- 9       1
---------------CCCATTCTCATCATCCAAACTG-------------- 9       1
```



Monday 17 lecture

We have learnt that several small sequences in cds and 3'UTR of mRNA may regulate the turn-over rate.
We have also examined a paper in which scientists sopught to identify all mRNAs bound "in vivo" by an RBP (PUM) that possibly regulate mRNA translation, osing RBP immunoprecipitation and microarray analysis.

One of the main features of RNA cis-elements is that very often they are very small sequences with low conservation (PUM element actully represents an exception, rather than a rule).

Conversely, very oftene RBPs contain multiple RNA-binding domain, suggesting cooperative action of repeated small sequences.

In addition, we have learnt that microRNA target sequences and RBP-binding sequences may interact to determine regulatory result. For example, in C. elegans Puf2 is required for miR-let7 action, and in the next slide, another RNA-binding protein Dnd1 actually inhibits function of several miRNA by binding to sequences close to miRNA atrget and inhibiting access. (see next slide), demostrating that a network of interaction exist among regulators.

RNA-Binding Protein Dnd1 Inhibits MicroRNA Access to Target mRNA

Martijn Kedde,[1] Markus J. Strasser,[2] Bijan Boldajipour,[2] Joachim A.F. Oude Vrielink,[1] Krasimir Slanchev,[2,5] Carlos le Sage,[1] Remco Nagel,[1] P. Mathijs Voorhoeve,[1] Josyanne van Duijse,[1] Ulf Andersson Ørom,[3] Anders H. Lund,[3] Anastassis Perrakis,[4] Erez Raz,[2,*] and Reuven Agami[1,*]

[1]The Netherlands Cancer Institute, Division of Tumor Biology, Plesmanlaan 121, 1066CX, Amsterdam, The Netherlands
[2]Max-Planck-Institute for Biophysical Chemistry, Germ Cell Development, Am Fassberg 11, 37070 Goettingen, and Institute for Cell Biology, ZMBE, Center for Molecular Biology of Inflammation, University of Münster, Münster 48149, Germany
[3]Biotech Research and Innovation Centre, University of Copenhagen, Ole Maaløes Vej 5, 2200N, Copenhagen, Denmark
[4]The Netherlands Cancer Institute, Division of Molecular Carcinogenesis, Plesmanlaan 121, 1066CX, Amsterdam, The Netherlands
[5]Present address: Max-Planck-Institute for Immunology, Stuebeweg 51, 79108 Freiburg, Germany.

Cell *131*, 1273–1286, December 28, 2007 ©2007 Elsevier Inc.

SUMMARY

MicroRNAs (miRNAs) are inhibitors of gene expression capable of controlling processes in normal development and cancer. In mammals, miRNAs use a seed sequence of 6–8 nucleotides (nt) to associate with 30 untranslated regions (30UTRs) of mRNAs and inhibit their expression. Intriguingly, occasionally not only the miRNA-targeting site but also sequences in its vicinity are highly conserved throughout evolution. We therefore hypothesized that conserved regions in mRNAs may serve as docking platforms for modulators of miRNA activity. Here we demonstrate that the expression of dead end 1 (Dnd1), an evolutionary conserved RNAbinding protein (RBP), counteracts the function of several miRNAs in human cells and in primordial germ cells of zebrafish by binding mRNAs and prohibiting miRNAs from associating with their target sites. These effects of Dnd1 are mediated through uridine-rich regions present in the miRNA-targeted mRNAs. Thus, our data unravel a novel role of Dnd1 in protecting certain mRNAs from miRNA-mediated repression.

The other important class of regulators of mRNA turn-over and translation are microRNAs

General aspects of microRNA (=miRNA) were discussed in lecture 26 (BMG06-03/12/10) and lecture 30 (BMG10-15/12/10). Students will refer to supporting materials in these lectures for basic information.

Today, we will deal about two main questions:

1) recognition of target mRNA by miRNAs

2) regulation of miRNA

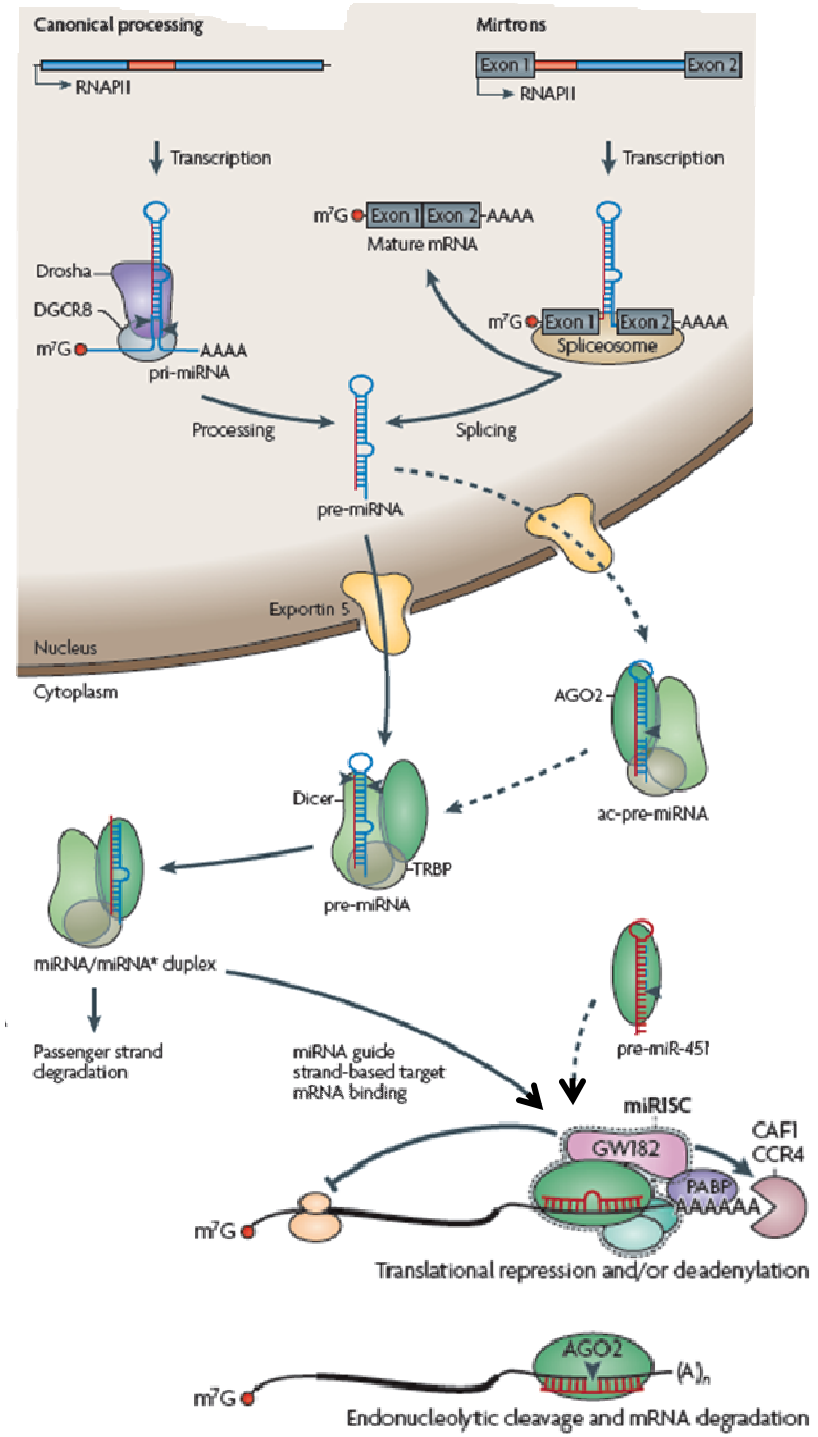
The widespread regulation of microRNA biogenesis, function and decay

Review

Jacek Krol, Inga Loedige* and Witold Filipowicz*

Abstract | MicroRNAs (miRNAs) are a large family of post-transcriptional regulators of gene expression that are ~21 nucleotides in length and control many developmental and cellular processes in eukaryotic organisms. Research during the past decade has identified major factors participating in miRNA biogenesis and has established basic principles of miRNA function. More recently, it has become apparent that miRNA regulators themselves are subject to sophisticated control. Many reports over the past few years have reported the regulation of miRNA metabolism and function by a range of mechanisms involving numerous protein–protein and protein–RNA interactions. Such regulation has an important role in the context-specific functions of miRNAs.

Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, PO BOX 2543, 4002 Basel, Switzerland.
**These authors contributed equally to this work.*
Correspondence to W.F.
e-mail:
witold.filipowicz@fmi.ch



MicroRNAs (miRNAs) are processed from RNA polymerase II (RNAPII)-specific transcripts of independent genes or from introns of protein-coding genes. In the canonical pathway, primary precursor (pri-miRNA) processing occurs in two steps, catalysed by two members of the RNase III family of enzymes, Drosha and Dicer, operating in complexes with dsRNA-binding proteins (dsRBPs), for example DGCR8 and transactivation-responsive (TAR) RNA-binding protein (TRBP) in mammals. In the first nuclear step, the Drosha–DGCR8 complex processes pri-miRNA into an ~70-nucleotide precursor hairpin (pre-miRNA), which is exported to the cytoplasm. Some pre-miRNAs are produced from very short introns (mirtrons) as a result of splicing and debranching, thereby bypassing the Drosha–DGCR8 step. In either case, cleavage by Dicer, assisted by TRBP, in the cytoplasm yields an ~20-bp miRNA/miRNA* duplex. In mammals, argonaute 2 (AGO2), which has robust RNaseH-like endonuclease activity, can support Dicer processing by cleaving the 3′ arm of some pre-miRNAs, thus forming an additional processing intermediate called AGO2-cleaved precursor miRNA (ac-pre-miRNA)70. Processing of pre-miR-451 also requires cleavage by AGO2, but is independent of Dicer and the 3′ end is generated by exonucleolytic trimming. Following processing, one strand of the miRNA/miRNA* duplex (the guide strand) is preferentially incorporated into an miRNA-induced silencing complex (miRISC), whereas the other strand (passenger or miRNA*) is released and degraded (not shown). Generally, the retained strand is the one that has the less stably base-paired 5′ end in the miRNA/miRNA* duplex. miRNA* strands are not always by-products of miRNA biogenesis and can also be loaded into miRISC to function as miRNAs. See BOX 2 for details of miRISC function. GW182, glycine-tryptophan protein of 182 kDa; m7G, 7-methylguanosine-cap; PABP, poly(A) binding protein.
(*From Krol 2010 NRG, modified)*

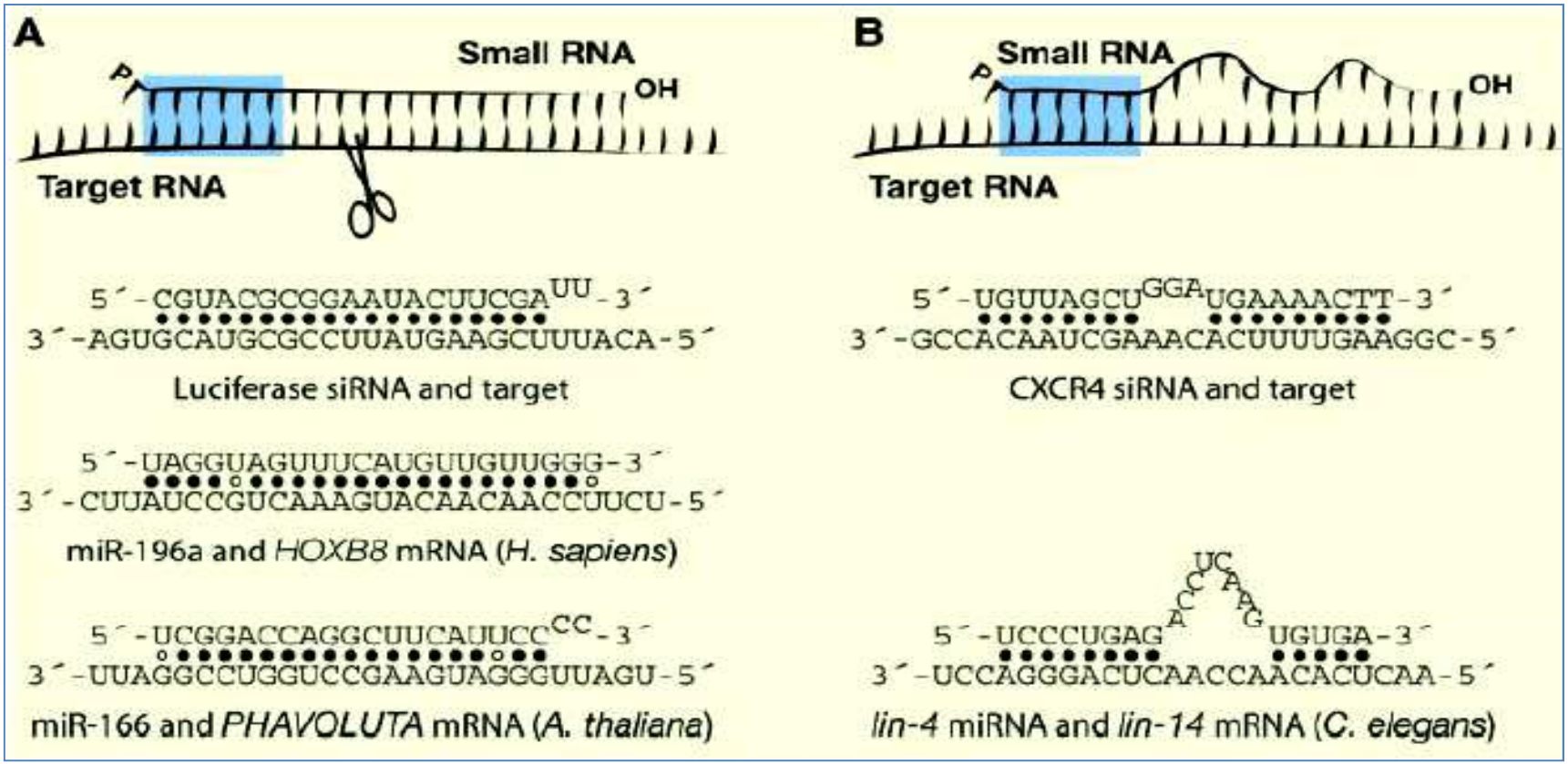


Fig. 3. Small RNA binding modes. (A) Extensive pairing of a small RNA to an mRNA allows the Piwi domain of a catalytically active Argonaute protein (e.g., Ago2 in humans or flies) to cut a single phosphodiester bond in the mRNA, triggering its destruction. Synthetic siRNAs typically exploit this mechanism, but some mammalian miRNAs (such as miR-196a) and most, if not all, plant miRNAs direct an Argonaute protein to cut their mRNA targets. (B) Partial pairing between the target RNA and the small RNA, especially through the ‘‘seed’’ sequence—roughly nucleotides 2 to 7 of the small RNA—tethers an Argonaute protein to its mRNA target. Binding of the miRNA and Argonaute protein prevents translation of the mRNA into protein. siRNAs can be designed to trigger such ‘‘translational repression’’ by including central mismatches with their target mRNAs; animal miRNAs such as lin-4, the first miRNA discovered, typically act by this mode because they are only partially complementary to their mRNA targets. The seed sequence of the small RNA guide is highlighted in blue.

Initial knowledge of miRNA-mRNA recognition came from anedoctical studies.

It has became clear soon that specificity is based on pairing of 6-8 positions starting from miRNA 5’ position 1 or 2, whereas pairing thereafer is less relevant.

This fact prospected that each miRNA could potentially regulate more than one target mRNA and experimental vlaidatin showed that this is the case, indeed. More strickingly, mRNAs that are co-regulated by a given miRNA often delineate specific GO classes.

When a number of cases were clear, people started to seek for algorithms capable of predicting the mRNAs targeted by the diverse known microRNAs.

In the web, there are several tools that join miRNA database with predicting algorithms. Of course, each of them is based on specific “assumptions” that may differ significantly from one site to the other.

Each predicting algorithm gives a list of the possible targets associated with a “score” that reflect the goodness of the fit (on the model employed).

They are quite good in general, but there is perplexing variability of results when considering unconventional binding sites...

PERSPECTIVE

nature structural & molecular biology

Desperately seeking microRNA targets

Marshall Thomas[1], Judy Lieberman[1] & Ashish Lal[1,2]

Review

MicroRNAs (miRNAs) suppress gene expression by inhibiting translation, promoting mRNA decay or both. Each miRNA may regulate hundreds of genes to control the cell's response to developmental and other environmental cues. The best way to understand the function of a miRNA is to identify the genes that it regulates. Target gene identification is challenging because miRNAs bind to their target mRNAs by partial complementarity over a short sequence, suppression of an individual target gene is often small, and the rules of targeting are not completely understood. Here we review computational and experimental approaches to the identification of miRNA-regulated genes. The examination of changes in gene expression that occur when miRNA expression is altered and biochemical isolation of miRNA-associated transcripts complement target prediction algorithms. Bioinformatic analysis of over-represented pathways and nodes in protein-DNA interactomes formed from experimental candidate miRNA gene target lists can focus attention on biologically significant target genes.

Table 1 miRNA target prediction algorithms differ in their ability to identify unconventional miRNA recognition elements (MREs)

MRE feature	Example		Target prediction algorithm[a]				
			TargetScan	PicTar	miRanda	PITA	rna22
Perfect seed	HMGA2 5′ CCGACAUUCAAUUUCUACCUCA 3′ let-7a 3′ UUGAUAUGUUGGAUGAUGGAGU 5′	Ref. 36					
G:U wobble seed	NF2 5′ UACAAGAGAUUCUCCUGCCUCA 3′ let-7a 3′ UUGAUAUGUUGGAUGAUGGAGU 5′	Ref. 37	–				
Imperfect seed/seedless	E2F2 5′ GUGGGUGCU-CUGGGCUGAACCA 3′ miR-24 3′ GACA-AGGACGACUUGACUCGGU 5′	Ref. 17	–				
Outside the 3′ UTR	DNMT3B 5′ UGGCAAAGAAGAUGUUUUGUGGUGCACUGAG 3′ miR-148 3′ -UGUUU.....CAAGACAUCACGUGACU- 5′	Ref. 38	–	–	–	+	+

[a]The algorithms are listed in the rough order by which they weigh perfect seed pairing and location within the 3′ UTR. The algorithms that allow identification of noncanonical recognition sites generally are more inclusive, but have much higher false prediction rates. +, algorithm predicts targets of this class; –, algorithm does not predict targets of this class.

Therefore, it has became very important to develop methods to understand, at the genome-wide level, the logic of miRNA-mRNA recognition.

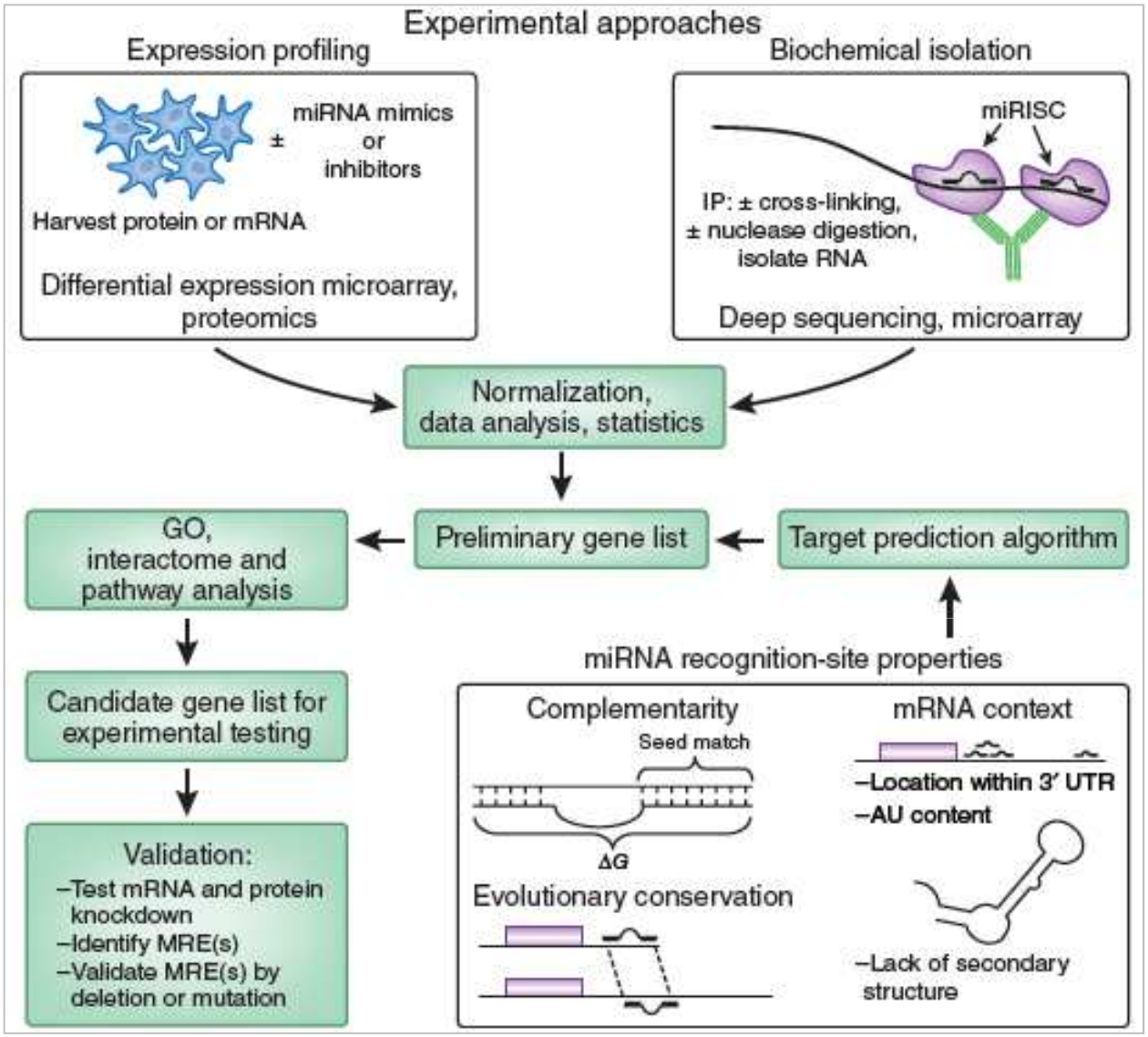


Figure 1 Methods for identifying miRNA targets. Putative target genes can be identified by expression profiling of cells in which the miRNA is overexpressed or antagonized, by biochemical isolation of the miRISC or by target prediction algorithms. These methods generally identify hundreds of candidate genes or more. Bioinformatic analysis of these large candidate gene lists for over-represented Gene Ontology (GO) terms, enriched biological pathways or gene interaction networks can then help researchers to select candidate genes to evaluate experimentally.

example of the clip method:

Vol 460|23 July 2009|doi:10.1038/nature08170 nature

ARTICLES

Argonaute HITS-CLIP decodes microRNA–mRNA interaction maps

Sung Wook Chi[1], Julie B. Zang[1], Aldo Mele[1] & Robert B. Darnell[1]

MicroRNAs (miRNAs) have critical roles in the regulation of gene expression; however, as miRNA activity requires base pairing with only 6–8 nucleotides of messenger RNA, predicting target mRNAs is a major challenge. Recently, high-throughput sequencing of RNAs isolated by crosslinking immunoprecipitation (HITS-CLIP) has identified functional protein–RNA interaction sites. Here we use HITS-CLIP to covalently crosslink native argonaute (Ago, also called Eif2c) protein–RNA complexes in mouse brain. This produced two simultaneous data sets—Ago–miRNA and Ago–mRNA binding sites—that were combined with bioinformatic analysis to identify interaction sites between miRNA and target mRNA. We validated genome-wide interaction maps for miR-124, and generated additional maps for the 20 most abundant miRNAs present in P13 mouse brain. Ago HITS-CLIP provides a general platform for exploring the specificity and range of miRNA action *in vivo*, and identifies precise sequences for targeting clinically relevant miRNA–mRNA interactions.

MicroRNA Targeting Specificity in Mammals: Determinants beyond Seed Pairing

Andrew Grimson,[1,2,4,6] Kyle Kai-How Farh,[1,2,3,4,6] Wendy K. Johnston,[1,2,4] Philip Garrett-Engele,[5] Lee P. Lim,[5,*] and David P. Bartel[1,2,4,*]

[1]Howard Hughes Medical Institute
[2]Department of Biology
[3]Division of Health Sciences and Technology
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge,
[5]Rosetta Inpharmatics, 401 Terry Avenue N, Seattle, WA 98109, USA
[6]These authors contributed equally to this work.
*Correspondence: lee_lim@merck.com (L.P.L.), dbartel@wi.mit.edu (D.P.B.)
DOI 10.1016/j.molcel.2007.06.017

Mammalian microRNAs (miRNAs) pair to 3′UTRs of mRNAs to direct their posttranscriptional repression. Important for target recognition are ~7 nt sites that match the seed region of the miRNA. However, these seed matches are not always sufficient for repression, indicating that other characteristics help specify targeting. By combining computational and experimental approaches, we uncovered five general features of site context that boost site efficacy: AU-rich nucleotide composition near the site, proximity to sites for coexpressed miRNAs (which leads to cooperative action), proximity to residues pairing to miRNA nucleotides 13–16, positioning within the 3′UTR at least 15 nt from the stop codon, and positioning away from the center of long UTRs. A model combining these context determinants quantitatively predicts site performance both for exogenously added miRNAs and for endogenous miRNA-message interactions. Because it predicts site efficacy without recourse to evolutionary conservation, the model also identifies effective nonconserved sites and siRNA off-targets.

HeLa cells transfected with 11 synthetic double-stranded miRNA

mRNA extracted from cells and analyzed on expression microarrays

Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature 433, 769–773.

miRNA targets were then identified in the regulated mRNAs and classified.

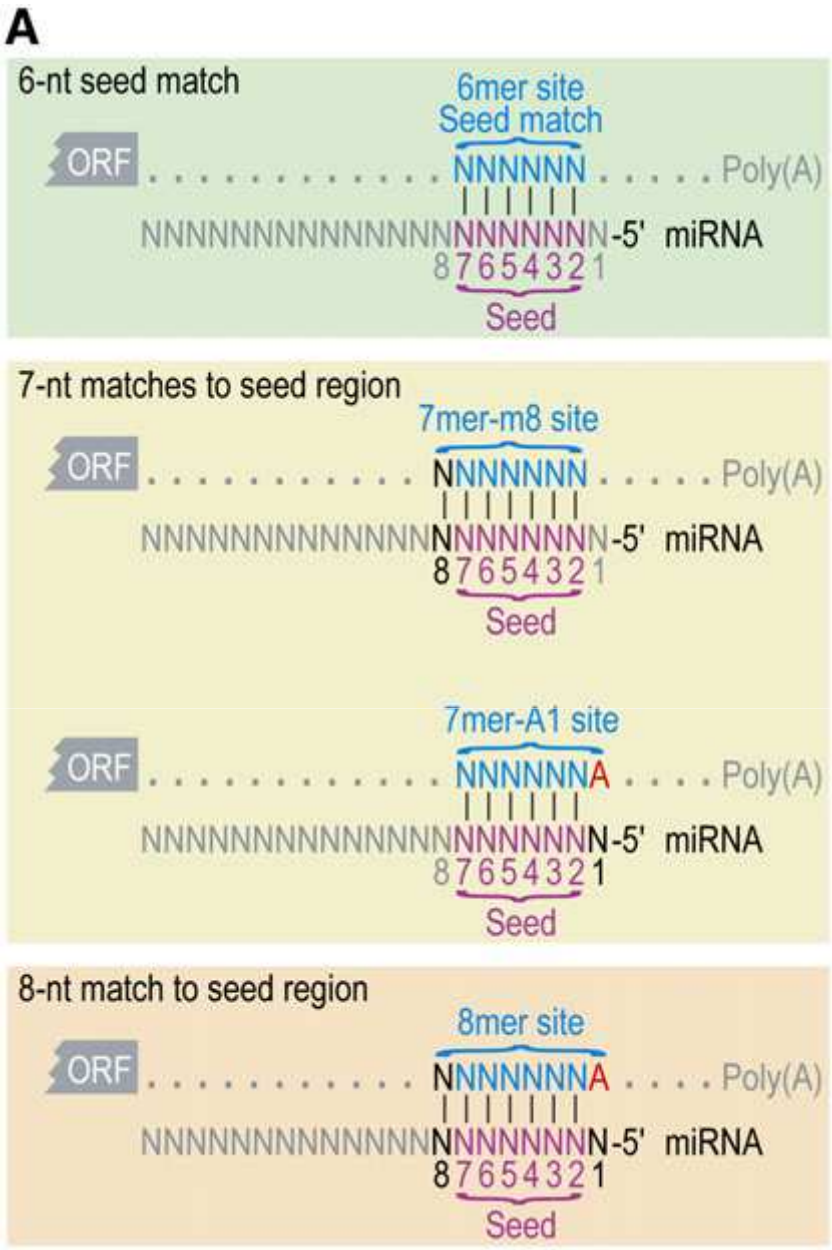
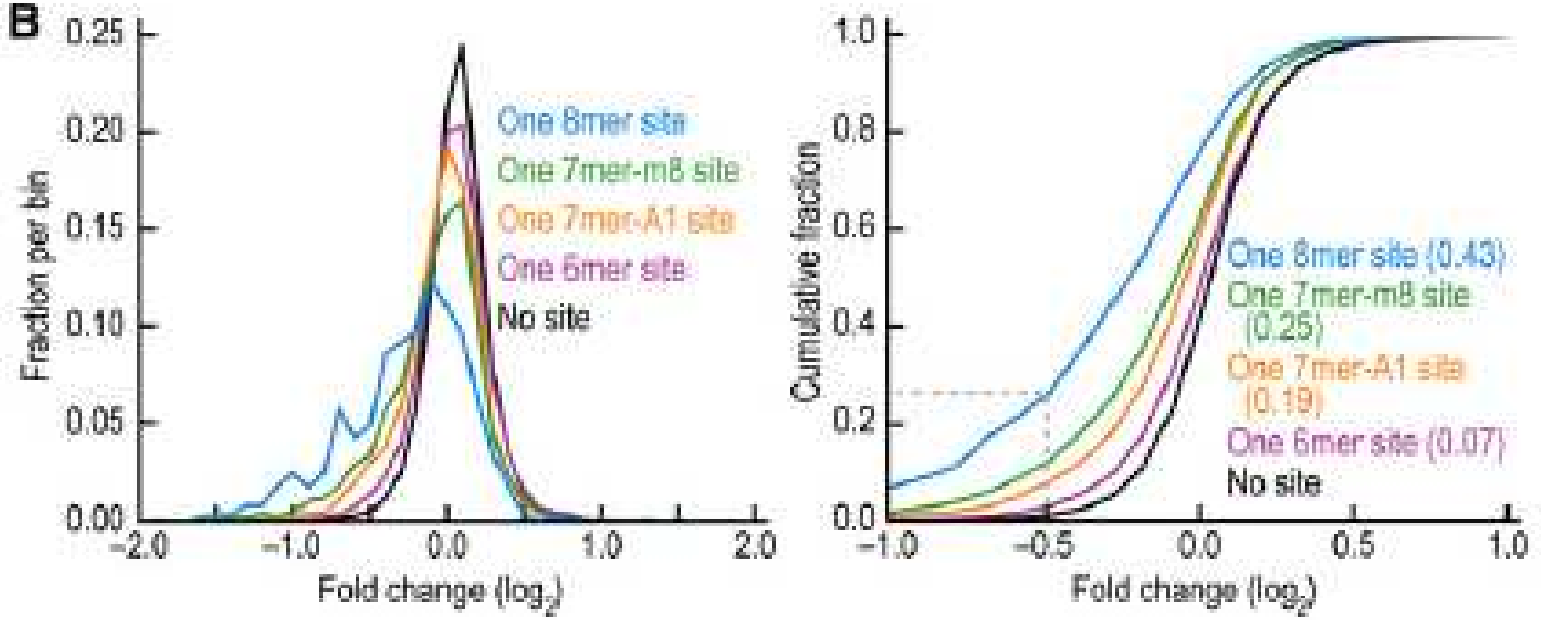
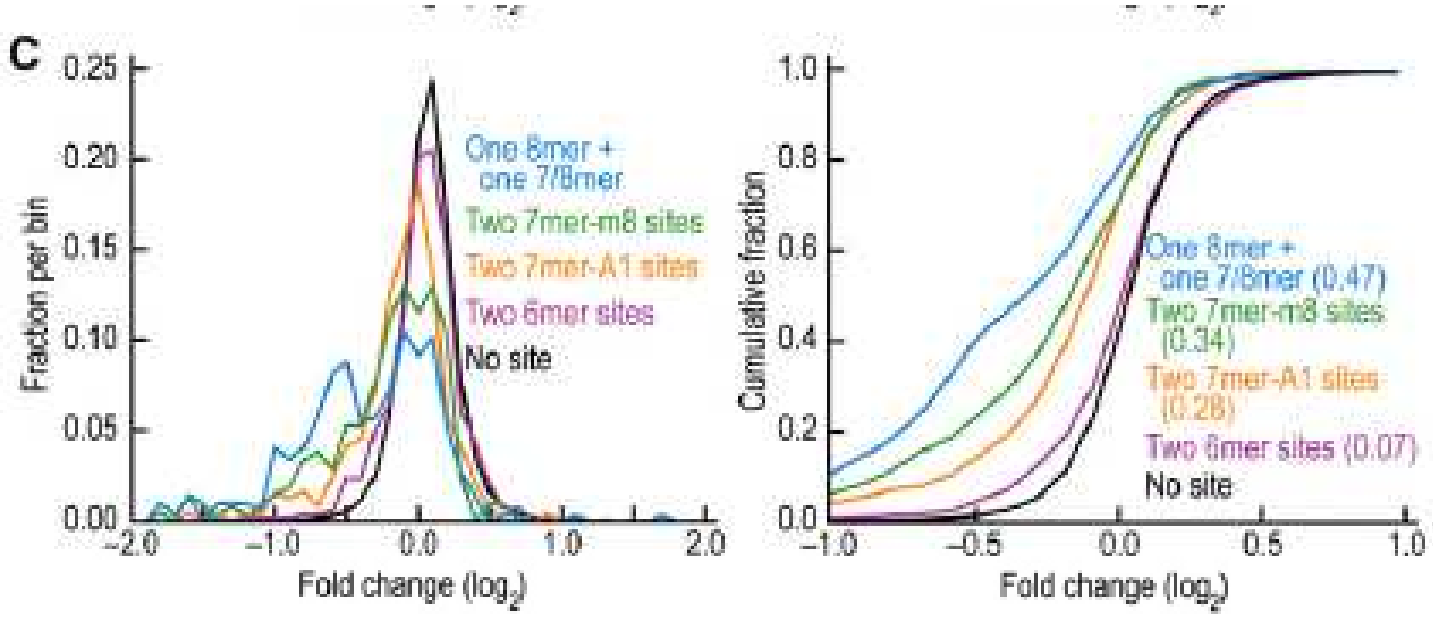


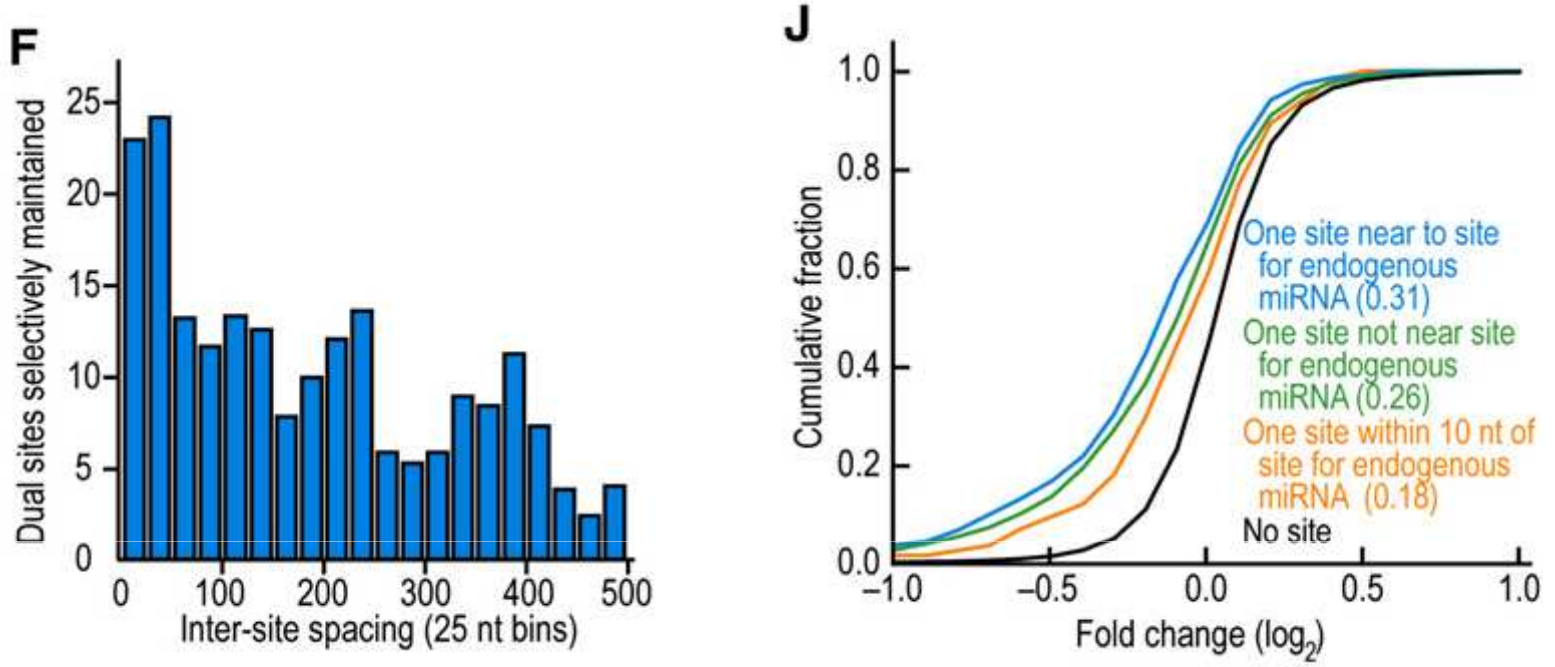
Figure 1. Downregulation of Messages with 6-8Mer Sites (A) Canonical miRNA complementary sites.



(B) Effectiveness of single canonical sites. Changes in abundance of mRNAs after miRNA transfection were monitored with microarrays. Distributions of changes (0.1 unit bins) for messages containing the indicated single sites in their UTRs are shown (left), together with the cumulative distributions (right). The dashed line in the cumulative distributions indicates that 27% of mRNAs with UTRs containing a single 8mer were downregulated at least 29% (20.5 = 0.71). Results of 11 experiments, each performed in duplicate and each transfecting a duplex for a different miRNA (Table S2), were consolidated. Results shown were an amalgam of the data from all 11 miRNAs; the relative strengths of the different sites were consistent when examining each transfection individually. For the cumulative plots, the minimal fraction of downregulated genes in that distribution is reported (parentheses), based on comparison with the no-site distribution. Repression from UTRs containing an 8mer site was significantly more than that from UTRs with a 7mer-m1 site ($p < 1020$, one-sided K-S test); similar comparisons between UTRs containing a 7mer-m8 site versus a 7mer-A1 site, a 7mer-A1 versus a 6mer, and a 6mer versus no site were also significant ($p < 10\text{-}6$, $p < 10\text{-}20$, and $p < 10\text{-}31$, respectively).



(C) Increased effectiveness of dual sites. Changes in mRNA abundance after miRNA transfection, represented as in (B), except mRNAs with 30UTRs containing the indicated pairs of sites were monitored. Repression from UTRs containing both an 8mer and either a 7mer or 8mer site was significantly more than that from UTRs with two 7mer-m8 sites ($p < 10\text{-}3$, one-sided K-S test); similar comparisons between UTRs containing two 7mer-m8 sites versus two 7mer-A1 sites, two 7mer-A1 sites versus two 6mer sites, and two 6mer sites versus no site were also significant ($p = 0.034$, $p < 10\text{-}11$, and $p < 10\text{-}6$, respectively).



(F) Selective maintenance of dual sites spaced at different intervals. Human 30UTRs with exactly two 7mer sites to the same miRNA were binned based on intersite distance (counting the number of nucleotides between the 30 nt of the first site and the 50 nt of the second site). The number of conserved dual sites exceeding the background (as estimated from the average of control cohorts) was plotted after performing for each bin siteconservation analysis analogous to that in Lewis et al. (2005), using the miRNA families conserved broadly among vertebrates (Table S1).

(J) Cooperativity between sites to transfected and endogenous miRNAs in HeLa cells. Endogenous sites considered were those for let-7 RNA, miR-16, miR-21, miR-23, miR-24, miR-27, and miR-30 (Landgraf et al., 2007). 7mer-m8 sites at a cooperative distance (>7 and <40 nt) from an endogenous miRNA 7-8mer site were significantly more downregulated than sites that were either too close to an endogenous miRNA (%7 nt, including overlapping sites; p = 0.0054, one-sided K-S test) or not close to an endogenous site (R40 nt, or no endogenous site; p = 0.036, one-sided K-S test).

five general features of site context that boost site efficacy:

1) AU-rich nucleotide composition near the site,
2) proximity to sites for coexpressed miRNAs (which leads to cooperative action),
3) proximity of seed to residues pairing to miRNA nucleotides 13–16,
4) positioning within the 3'UTR at least 15 nt from the stop codon,
5) positioning away from the center of long UTRs.

A model combining these context determinants quantitatively predicts site performance both for exogenously added miRNAs and for endogenous miRNAmessage interactions.

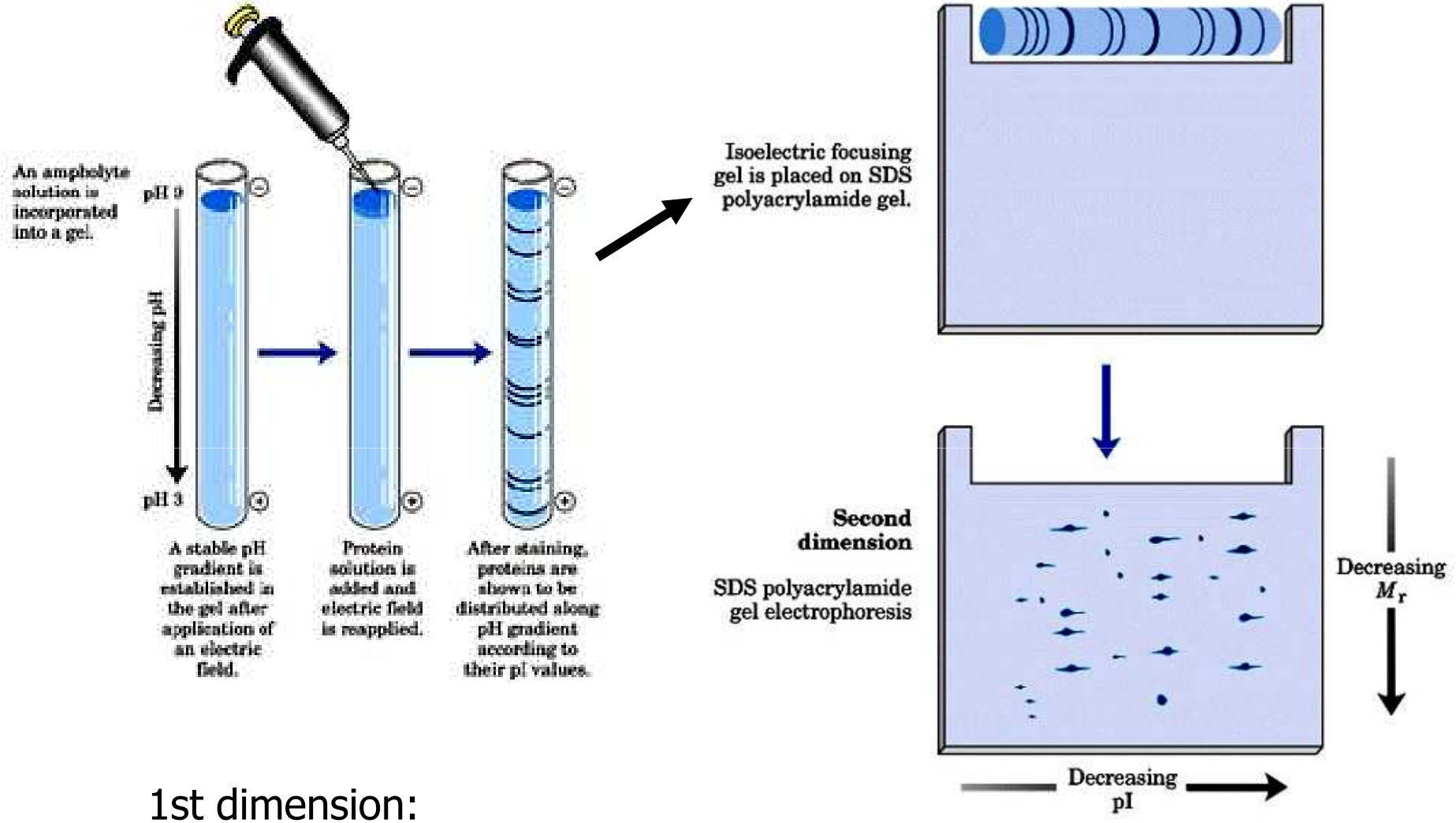
Because it predicts site efficacy without recourse to evolutionary conservation, the model also identifies effective nonconserved sites and siRNA off-targets.

Next question: degradation of translational inhibition?

The only way to answer this point is to run proteomic experiments.

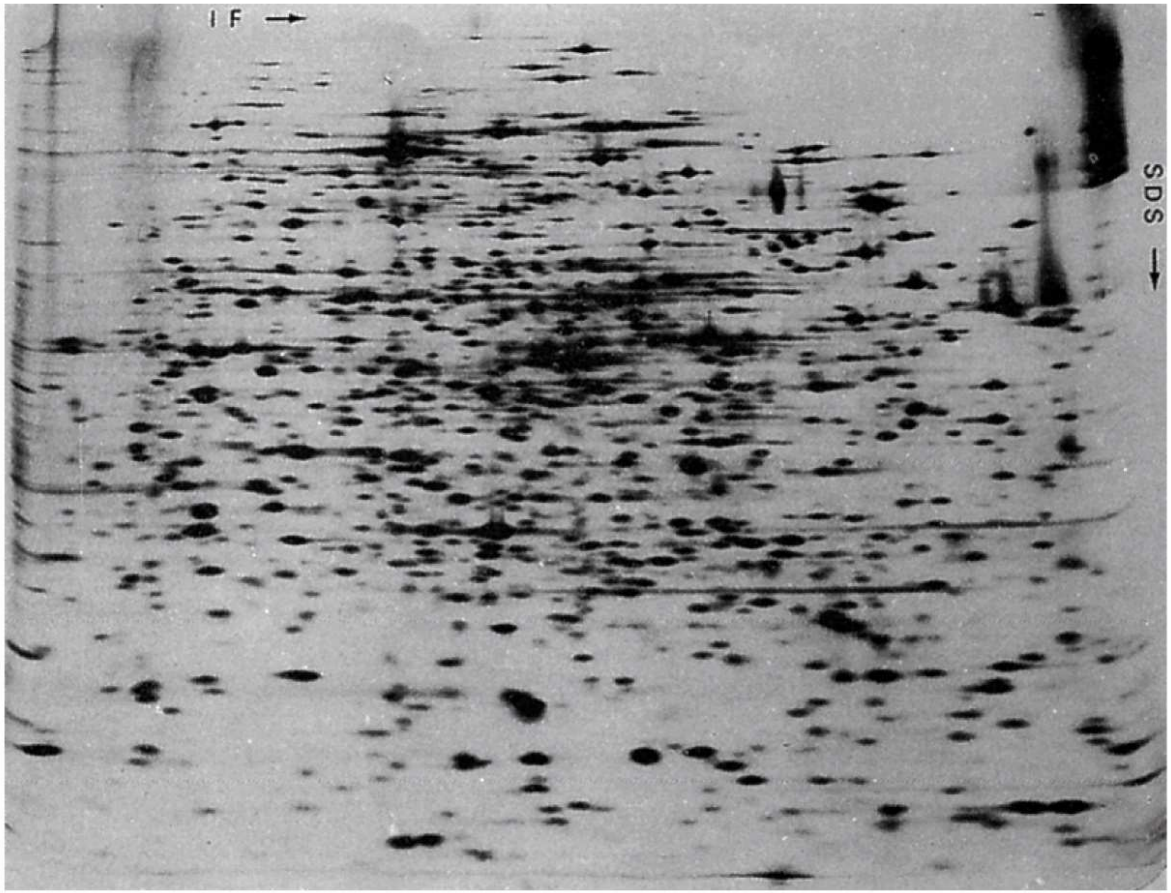
How to evaluate all the proteins in a cell or tissue?

Separation of complex protein mixtures by 2D gel electrophoresis



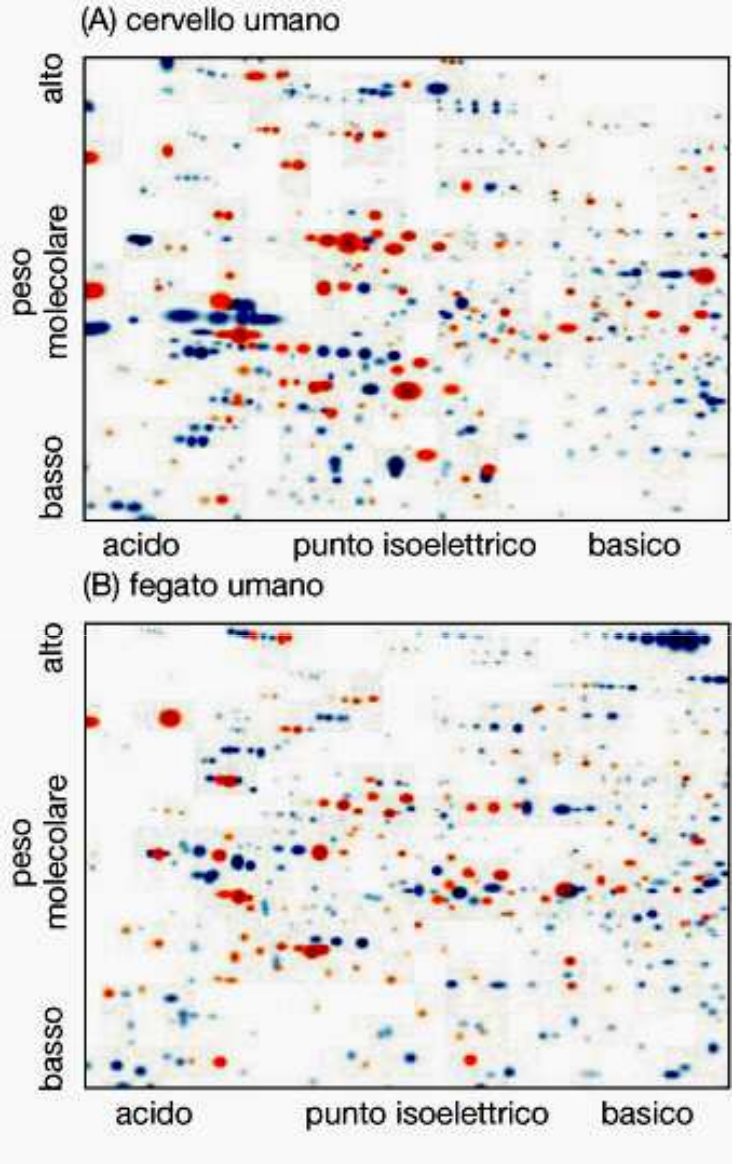
1st dimension: separation by isoelectric point

2nd dimension: separation by size



(b)

A 2D protein gel

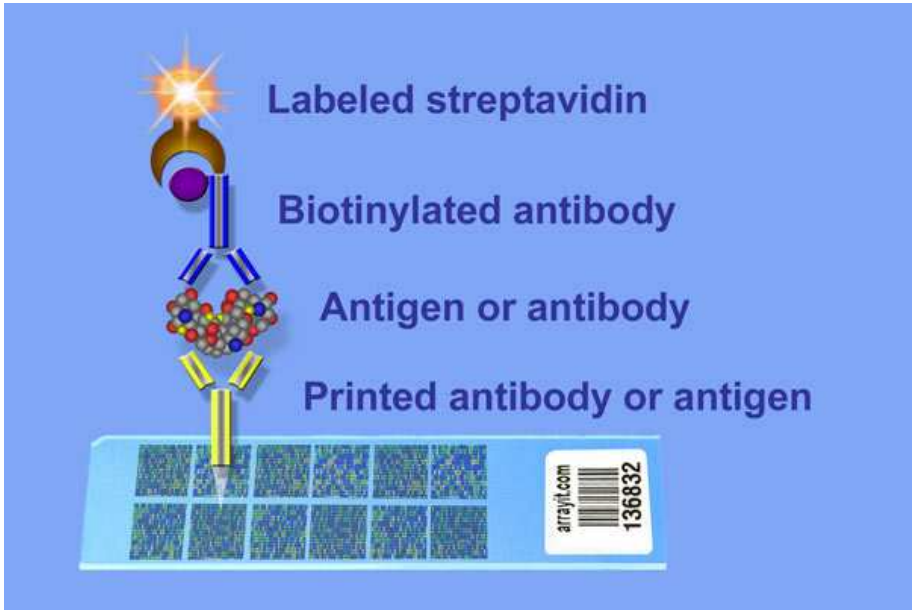


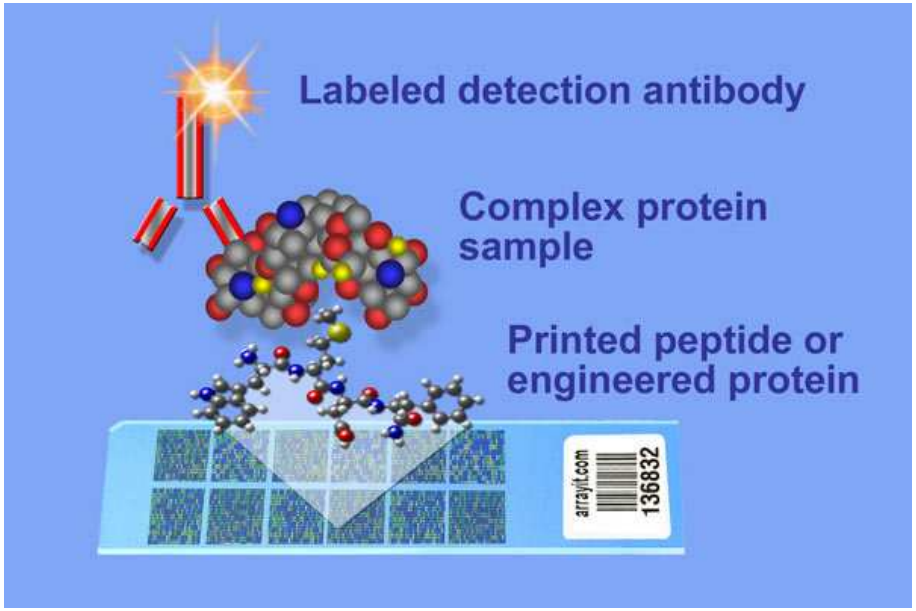
In red: common proteins

In blue: proteins expressed in either of the samples

Individual spots can be identified by mass spectrometry

Mass spectrometry identification of proteins





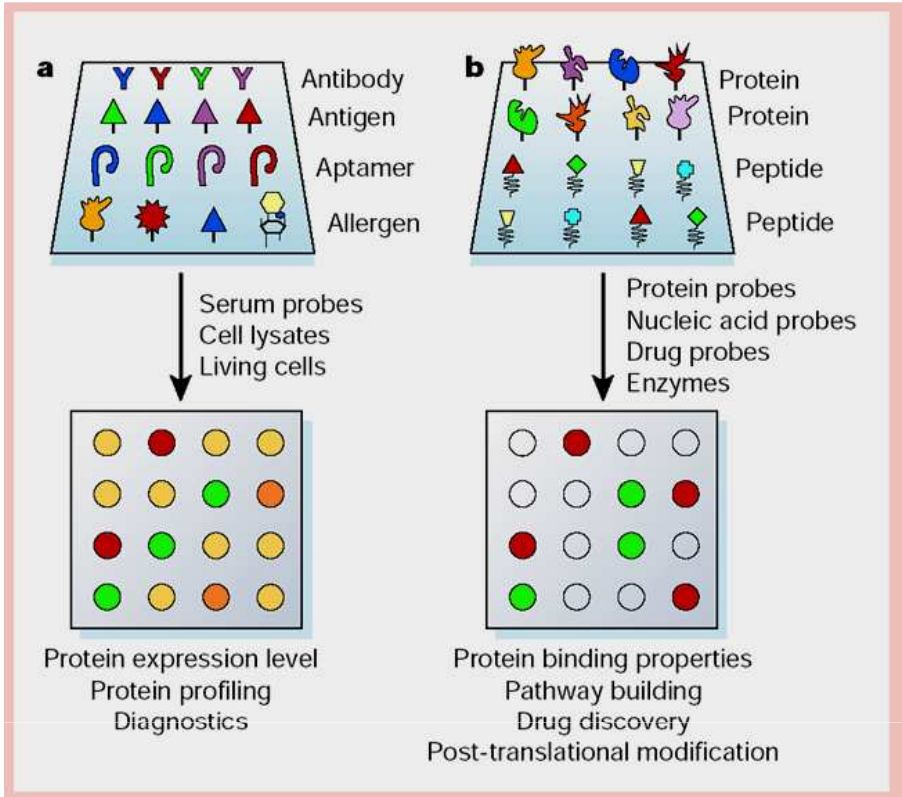


Figure 1 Analytical versus functional protein microarrays. **a**, Analytical protein microarray. Different types of ligands, including antibodies, antigens, DNA or RNA aptamers, carbohydrates or small molecules, with high affinity and specificity, are spotted down onto a derivatized surface. These chips can be used for monitoring protein expression level, protein profiling and clinical diagnostics. Similar to the procedure in DNA microarray experiments, protein samples from two biological states to be compared are separately labelled with red or green fluorescent dyes, mixed, and incubated with the chips. Spots in red or green colour identify an excess of proteins from one state over the other. **b**, Functional protein microarray. Native proteins or peptides are individually purified or synthesized using high-throughput approaches and arrayed onto a suitable surface to form the functional protein microarrays. These chips are used to analyse protein activities, binding properties and post-translational modifications. With the proper detection method, functional protein microarrays can be used to identify the substrates of enzymes of interest. Consequently, this class of chips is particularly useful in drug and drug-target identification and in building biological networks.

Review

Applying mass spectrometry-based proteomics to genetics, genomics and network biology

Matthias Gstaiger[*‡] *and Ruedi Aebersold*[*‡§||]

Abstract | The systematic and quantitative molecular analysis of mutant organisms that has been pioneered by studies on mutant metabolomes and transcriptomes holds great promise for improving our understanding of how phenotypes emerge. Unfortunately, owing to the limitations of classical biochemical analysis, proteins have previously been excluded from such studies. Here we review how technical advances in mass spectrometry-based proteomics can be applied to measure changes in protein abundance, posttranslational modifications and protein–protein interactions in mutants at the scale of the proteome. We finally discuss examples that integrate proteomics data with genomic and phenomic information to build network-centred models, which provide a promising route for understanding how phenotypes emerge.

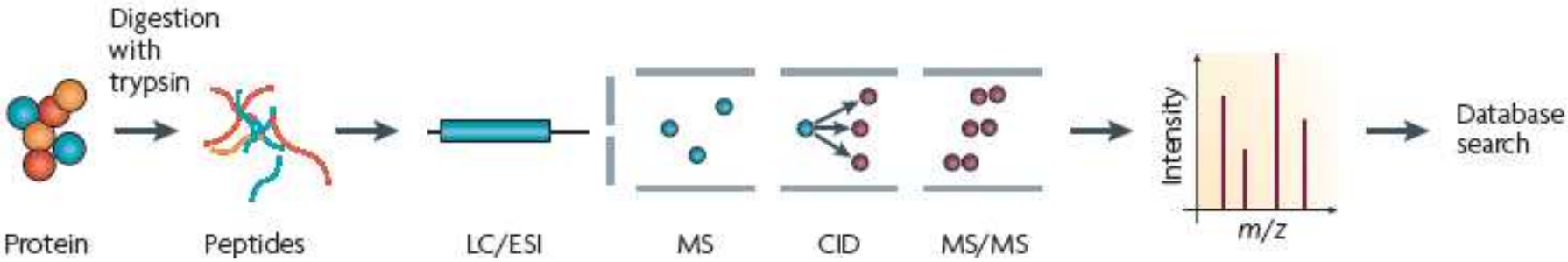


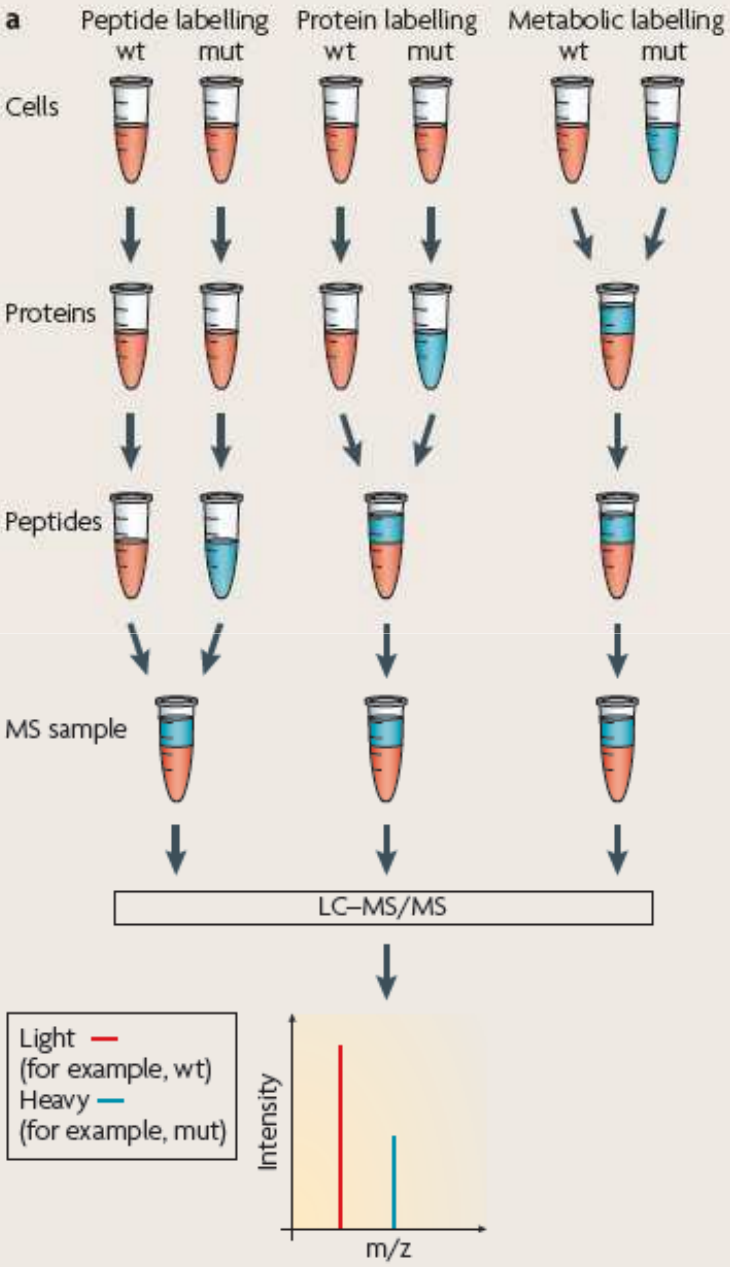
Figure 1 | **Mass spectrometry-based protein identification.** Protein samples are digested with trypsin and the resulting peptide mixtures are fractionated by reversed-phase liquid chromatography (LC). The fractionated peptide solution is subjected to an electric potential, which causes a spray to be formed, leading to the desolvation and ionization of the peptides (electrospray ionization; ESI). Mass to charge (m/z) ratios are measured from peptide ions that pass the collision cell without fragmentation in the mass spectrometer (MS). Specific ions are randomly selected for collision-induced dissociation (CID) with neutral gas molecules (for example, helium) and the resulting fragment ions are measured in the second mass analyser in tandem mass spectrometry (MS/MS). The MS precursor ion intensities obtained in the first stage can be used for peptide quantification, whereas MS/MS fragment ion information from the second stage contains sequence information that can be compared with sequences from *in silico* digested protein sequence databases for peptide and subsequent protein identification.

Isotope-labelling approaches

As shown in part **a** in the figure, differential labelling of proteins or peptides with heavy or light isotopes (indicated in red or blue) can be done *in vitro* or by the incorporation of isotope-labelled amino acids by metabolic labelling *in vivo*. For *in vitro* labelling, wild-type (wt) and mutant (mut) samples are prepared separately and isolated proteins or peptides are differentially labelled with heavy or light versions of isotope-tagging reagents, mainly through their sulphhydryl (for example, isotope-coded affinity tags)[71] or amine groups (for example, isotope-coded protein labels)[73]. Differential labelling introduces a characteristic mass shift, which can be used to determine the MS1 peptide ratios between pairs of heavy and light peptides. Peptide labelling with recently introduced isobaric tags for relative and absolute quantitation, which as the name indicates, keep the mass of the differentially labelled precursor ions of a given peptide constant but allow quantification after tandem mass spectrometry (MS/MS) analysis on the basis of sample-specific reporter ion intensities from up to eight different samples in a single liquid chromatography–tandem mass spectrometry (LC–MS/MS) experiment[74,75].

The use of synthetic isotope-labelled reference peptides for absolute quantification that was pioneered by Desiderio *et al.*[76] has been extended to proteomic studies by Steve Gygi and colleagues[77]. In this approach, known amounts of synthetic isotope-labelled reference peptides, which correspond to proteotypic peptides of the proteins to be analysed, are added to the samples before LC–MS/MS analysis for absolute quantification of proteins.

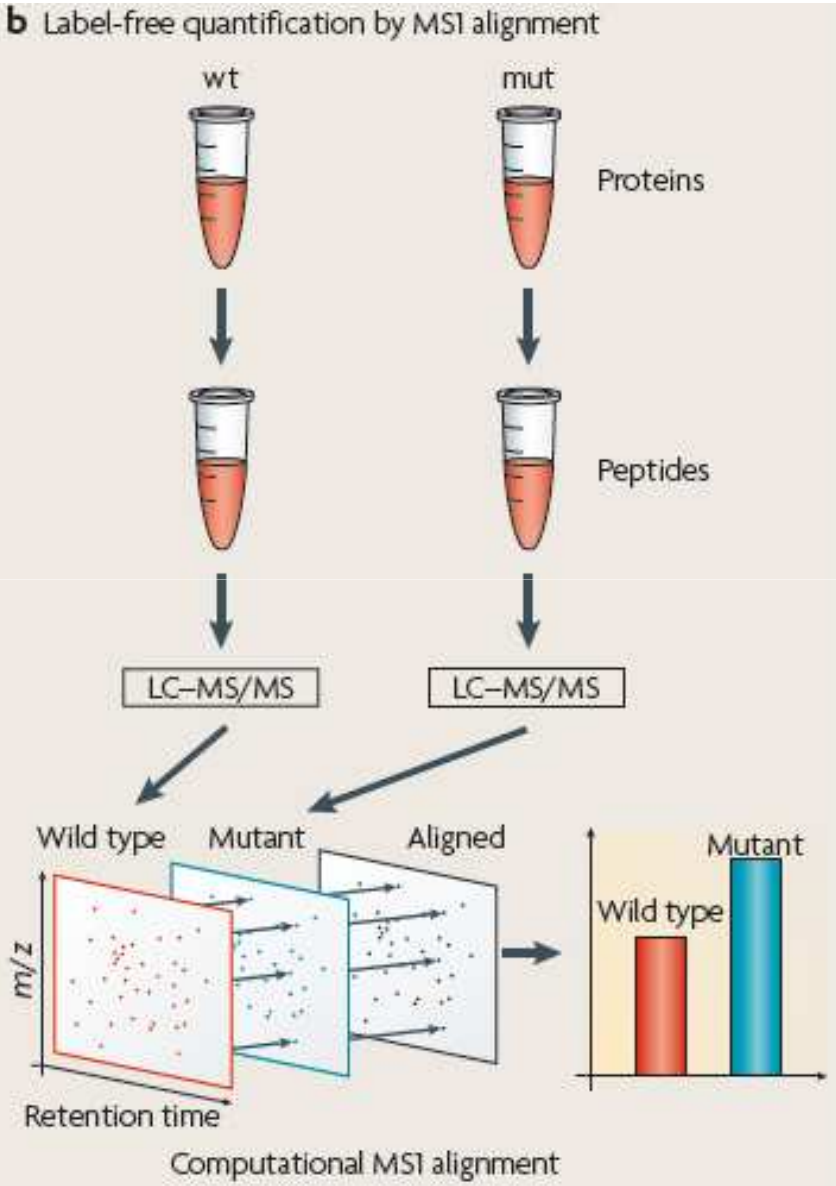
Stable isotope labelling with amino acids in cell culture is an *in vivo* isotope labelling method that is becoming increasingly popular[78]. Wild-type and mutant cells are grown in media that contains either light or heavy isotope versions of lysine or arginine, which yield differentially labelled proteomes. The entire labelling process occurs



at the beginning of the experiment, which has the advantage that samples can be combined at early steps to avoid errors that can be introduced when samples are separately processed. As the method is limited to cells or organisms that can be metabolically labelled, it is not generally applicable to human tissues and body fluids.

Label-free quantification from aligned MS1 spectra

For label-free quantification (part **b** in the figure) wild-type and mutant proteomes are analysed by separate LC–MS/MS experiments and the MS1 spectra are computationally aligned to calculate the relative protein abundance changes on the basis of the signal intensities of extracted ion chromatograms from aligned peptide features. This reduces the undersampling problem that is known to occur with MS/MS-based approaches and results in a dynamic range of three to four orders of magnitude[79]. Newer hybrid MS instruments (LTQ FT and LTQ Orbitrap) offer the option to simultaneously record MS signal intensities and identify peptides using MS/MS. These two types of information can be combined by recently developed computational approaches[80–82]. The number of peptides that can be mapped across different LC–MS/MS experiments therefore depends on the accuracy of the peptide masses that are determined by the mass analyser and reproducibility of the LC system. Strategies for signal normalization and for correcting variations in LC performances have been developed and are now integrated in automated computational platforms for label-free MS analysis[83].



nature Vol 455|4 September 2008|doi:10.1038/nature07228

ARTICLES

Widespread changes in protein synthesis induced by microRNAs

Matthias Selbach[1], Björn Schwanhäusser[1]*, Nadine Thierfelder[1]*, Zhuo Fang[1], Raya Khanin[2] & Nikolaus Rajewsky[1]

Animal microRNAs (miRNAs) regulate gene expression by inhibiting translation and/or by inducing degradation of target messenger RNAs. It is unknown how much translational control is exerted by miRNAs on a genome-wide scale. We used a new proteomic approach to measure changes in synthesis of several thousand proteins in response to miRNA transfection or endogenous miRNA knockdown. In parallel, we quantified mRNA levels using microarrays. Here we show that a single miRNA can repress the production of hundreds of proteins, but that this repression is typically relatively mild. A number of known features of the miRNA-binding site such as the seed sequence also govern repression of human protein synthesis, and we report additional target sequence characteristics. We demonstrate that, in addition to downregulating mRNA levels, miRNAs also directly repress translation of hundreds of genes. Finally, our data suggest that a miRNA can, by direct or indirect effects, tune protein synthesis from thousands of genes.

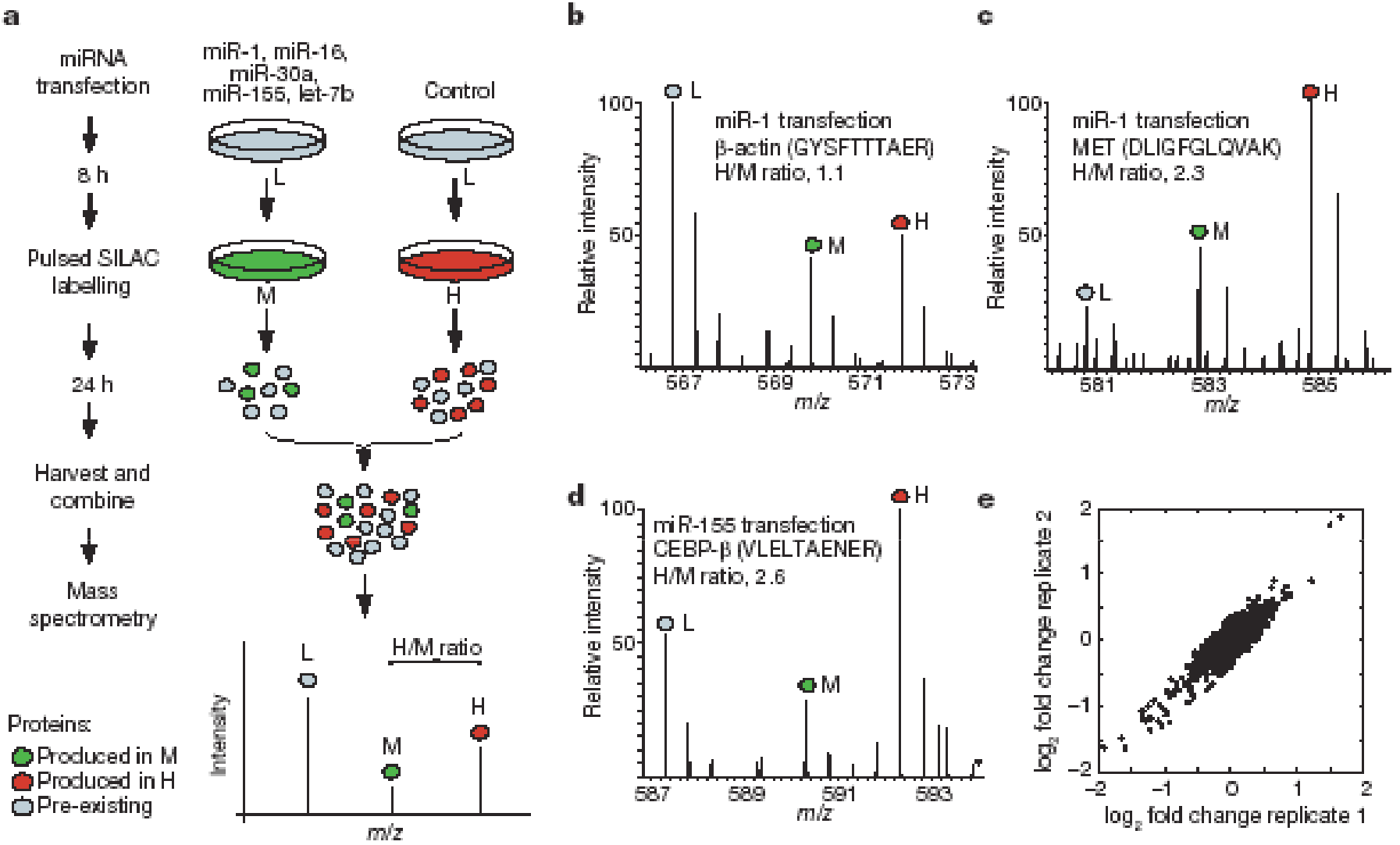


Figure 1 | Global analysis of changes in protein production induced by microRNAs. a, HeLa cells cultivated in normal light (L) medium were either transfected with a miRNA or mock transfected. After 8 h, transfected and control cells were pulse-labelled by transferring them to culture medium containing medium-heavy (M) or heavy (H) isotope-labelled amino acids, respectively (pSILAC). All newly synthesized proteins will appear in theH or M form. Samples were combined after 24 h and analysed by mass spectrometry. Intensity peak ratios between heavy and medium-heavy peptides (H/M ratio) reflect changes in protein production. RNA from the same samples was analysed by microarrays. b–d, Exemplary peptide mass spectra (sequences are in parentheses). The production of most proteins is unaltered, as shown for a b-actin peptide. In contrast, synthesis of MET and CEBPb is reduced by miR-1 or miR-155 overexpression. e, Reproducibility of pSILAC (biological replicate, see Supplementary Methods).

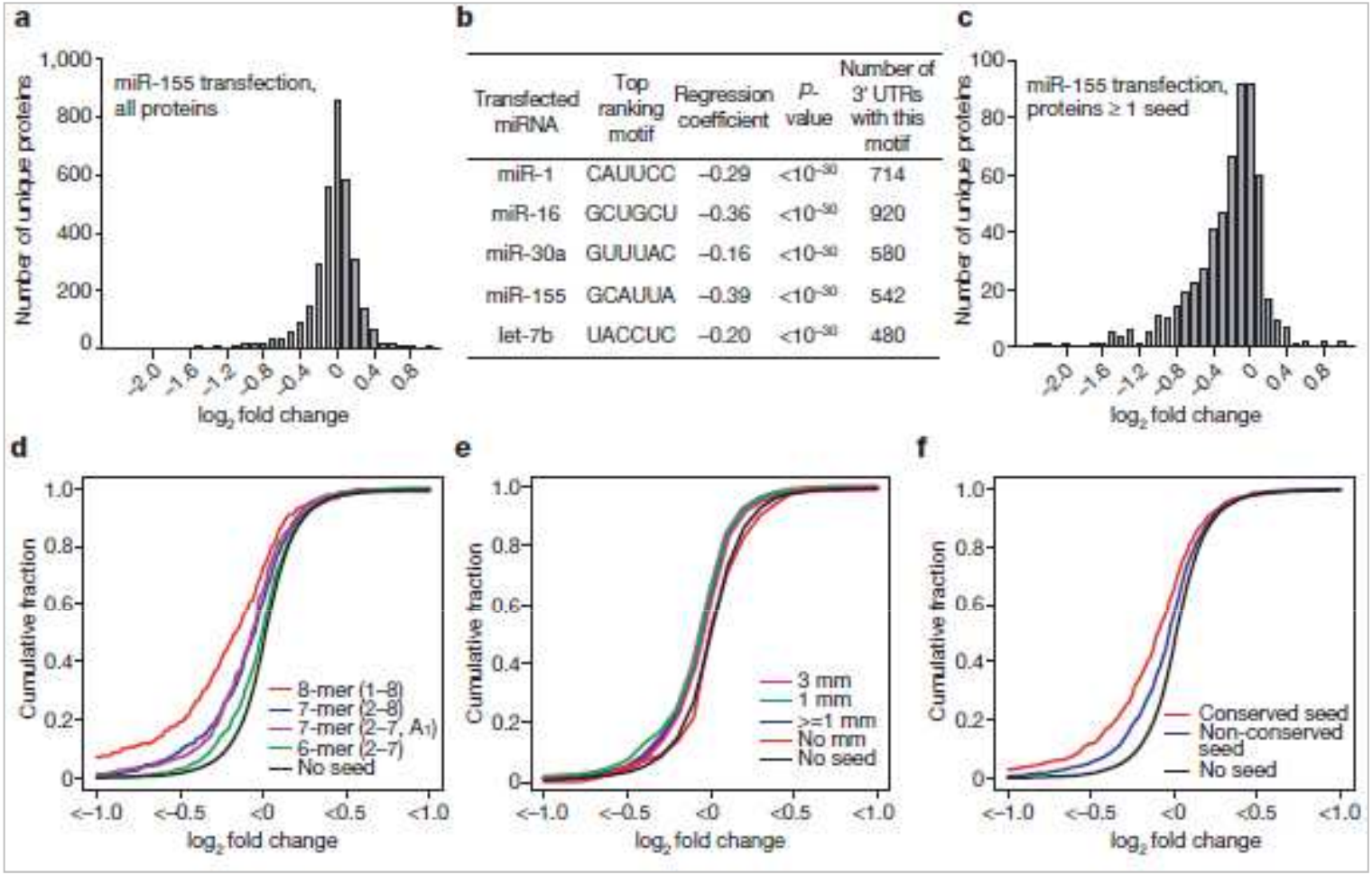


Figure 2 | miRNAs downregulate protein synthesis of hundreds of genes.
a, Histogram of changes in production of 3,299 proteins in HeLa cells after miR-155 overexpression. b, An unbiased search for 39UTR motifs that correlate with pSILAC fold changes yielded precisely the miRNA seed sequences. c, Proteins with miR-155 seeds tend to be downregulated by miR-155 overexpression. d, Cumulative distributions of different seed classes (matches to positions 1–8 (8-mer), 2–8 (7-mers), 2–7 with adenosine in position 1 (2–7, A1) and 2–7 (6-mer)). e, Mismatches (mm) between positions 9 and 11 of the miRNA and target mRNAs with a seed correlate with downregulation. Protein synthesis from mRNAs with perfect complementarity at positions 9–11 (red) and synthesis from mRNAs without seeds (black) is indistinguishable. f, Conserved seeds mediate more downregulation than non-conserved seeds. Results are shown for pooled data based on messages with one seed only (d–f).

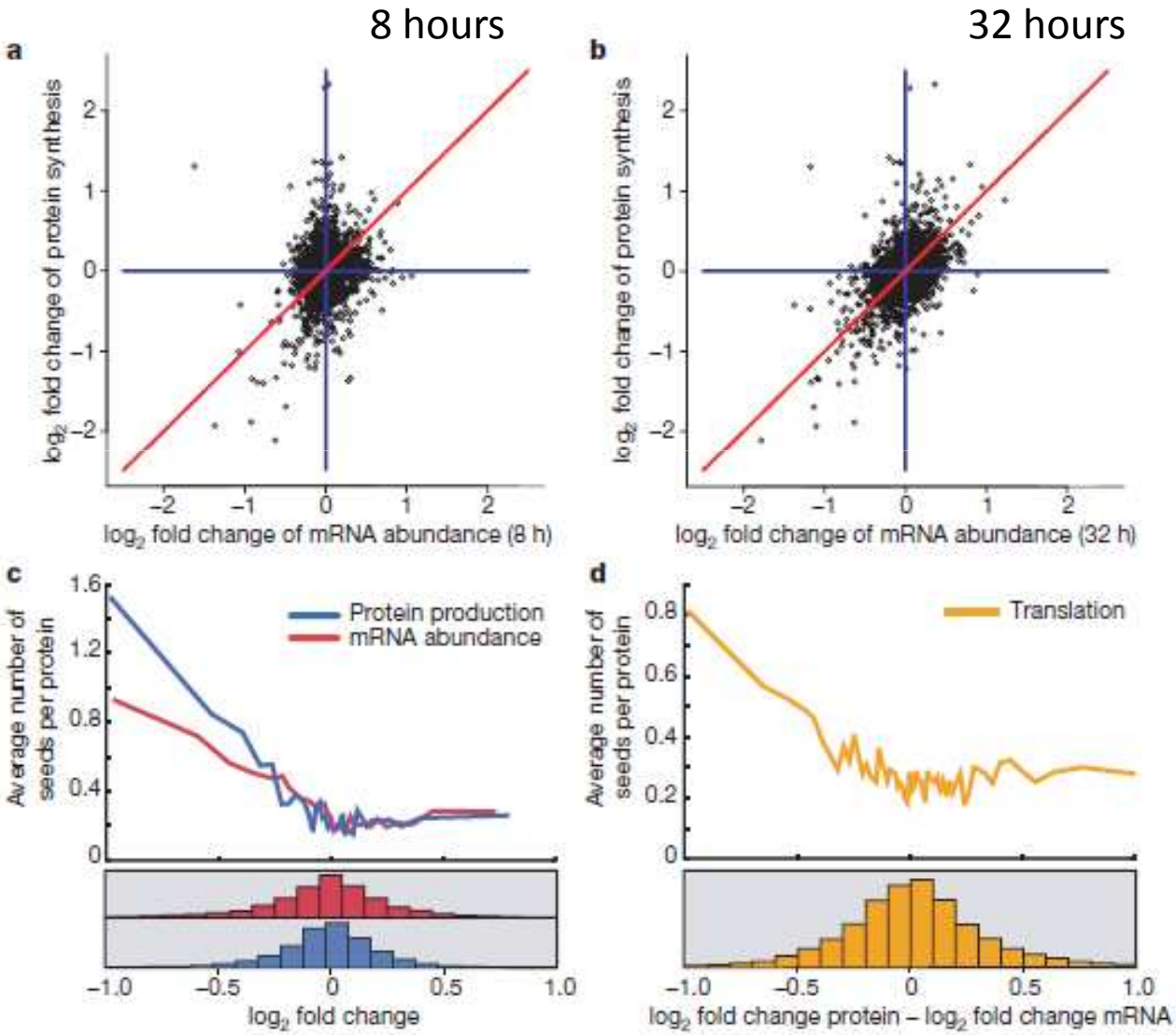


Figure 4 | miRNAs inhibit translation on a genome-wide scale. **a**, Changes in protein production between 8 h and 32 h after miR-1 transfection with mRNA fold changes at 8 h reveal poor overall correlation. **b**, mRNA levels at 32 h correlate remarkably well with changes in protein synthesis. **c**, Overall fold changes of mRNA and protein synthesis are similar (histograms). Reduced protein production and mRNA levels correlate with seed frequency (curves represent proteins ranked by fold change and grouped into bins of 250). **d**, Translational repression by miRNAs is revealed by subtracting mRNA log changes from log changes in protein production. Increased seed frequency, averaged as in **c**, correlates with translational repression. Results are shown for pooled data (**c**, **d**) after discarding genes with mRNA and pSILAC changes of unequal sign.

The other way around .... How is expression of miRNAs controlled ?

Review

Open Access

Regulation of MicroRNA Biogenesis: A miRiad of mechanisms

Brandi N Davis*[1,2] and Akiko Hata[1,2]

Address: [1]Department of Biochemistry, Tufts University School of Medicine, Boston MA 02111, USA and [2]Molecular Cardiology Research Institute, Tufts Medical Center, Boston, MA 02111, USA

Email: Brandi N Davis* - brandi.davis@tufts.edu; Akiko Hata - akiko.hata@tufts.edu

Review

Abstract

microRNAs are small, non-coding RNAs that influence diverse biological functions through the repression of target genes during normal development and pathological responses. Widespread use of microRNA arrays to profile microRNA expression has indicated that the levels of many microRNAs are altered during development and disease. These findings have prompted a great deal of investigation into the mechanism and function of microRNA-mediated repression. However, the mechanisms which govern the regulation of microRNA biogenesis and activity are just beginning to be uncovered. Following transcription, mature microRNA are generated through a series of coordinated processing events mediated by large protein complexes. It is increasingly clear that microRNA biogenesis does not proceed in a 'one-size-fits-all' manner. Rather, individual classes of microRNAs are differentially regulated through the association of regulatory factors with the core microRNA biogenesis machinery. Here, we review the regulation of microRNA biogenesis and activity, with particular focus on mechanisms of post-transcriptional control. Further understanding of the regulation of microRNA biogenesis and activity will undoubtedly provide important insights into normal development as well as pathological conditions such as cardiovascular disease and cancer.

Published: 10 August 2009

Cell Communication and Signaling 2009, 7:18 doi:10.1186/1478-811X-7-18

This article is available from: http://www.biosignaling.com/content/7/1/18

Expression of most miRNAs is spatio-temporal regulated

Expression of 50% miRNAs is tissue-specific

De-regulation of miRNA expression is observed in disease, such as cancer and heart failure.

Pattern of miRNA expression was found prognostic, predictive and predictive of response to treatment in various forms of cancer.

miRNA expression is also subjected to common gene regulation pathways

as examplified in the study shown in the next slide, demonstrating regulation of several miRNA expression by estrogenic hormones.

The American Journal of Pathology, Vol. 176, No. 5, May 2010

DOI: 10.2353/ajpath.2010.090837

Estrogen Receptor α Controls a Gene Network in Luminal-Like Breast Cancer Cells Comprising Multiple Transcription Factors and MicroRNAs

Luigi Cicatiello,* Margherita Mutarelli,*
Oli M.V. Grober,* Ornella Paris,* Lorenzo Ferraro,*
Maria Ravo,* Roberta Tarallo,* Shujun Luo,†
Gary P. Schroth,† Martin Seifert,‡
Christian Zinser,‡ Maria Luisa Chiusano,§
Alessandra Traini,§ Michele De Bortoli,¶
and Alessandro Weisz*‖

Luminal-like breast tumor cells express estrogen receptor α (ERα), a member of the nuclear receptor family of ligand-activated transcription factors that controls their proliferation, survival, and functional status. To identify the molecular determinants of this hormone-responsive tumor phenotype, a comprehensive genome-wide analysis was performed in estrogen stimulated MCF-7 and ZR-75.1 cells by integrating time-course mRNA expression profiling with global mapping of genomic ERα binding sites by chromatin immunoprecipitation coupled to massively parallel sequencing, microRNA expression profiling, and *in silico* analysis of transcription units and receptor binding regions identified. All 1270 genes that were found to respond to 17β-estradiol in both cell lines cluster in 33 highly concordant groups, each of which showed defined kinetics of RNA changes. This hormone-responsive gene set includes several direct targets of ERα and is organized in a gene regulation cascade, stemming from ligand-activated receptor and reaching a large number of downstream targets via AP-2γ, B-cell activating transcription factor, E2F1 and 2, E74-like factor 3, GTF2IRD1, hairy and enhancer of split homologue-1, MYB, SMAD3, RARα, and RXRα transcription factors. MicroRNAs are also integral components of this gene regulation network because miR-107, miR-424, miR-570, miR-618, and miR-760 are regulated by 17β-estradiol along with other microRNAs that can target a significant number of transcripts belonging to one or more estrogen-responsive gene clusters. *(Am J Pathol 2010, 176:2113–2130; DOI:*

Genomic organization and transcription of miRNA genes

miRNAs are encoded in diverse regions of the genome including both protein coding and non-coding transcription units.

Approximately 50% of miRNAs are derived from non-coding RNA transcripts, while an additional ~40% are located within the introns of protein coding genes.

The majority of miRNAs are transcribed by RNA polymerase (RNA pol) II and bear a 7-methyl guanylate cap at the 5' end and poly (A) tail at the 3' end.

RNA pol III has also been demonstrated to generate the transcripts of a subset of miRNAs.

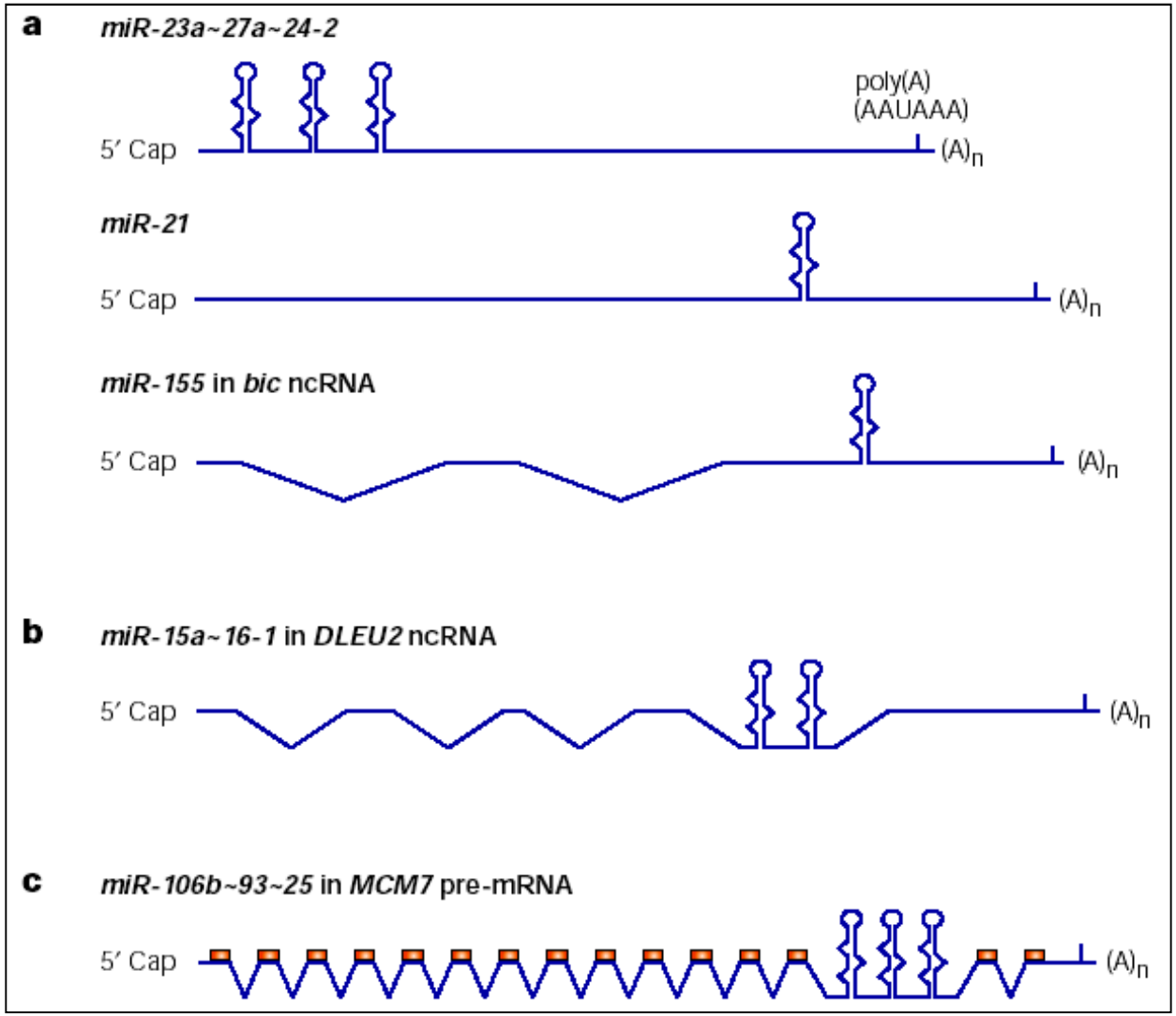


Figure 1 | **The structure of five pri-miRNAs.**

Primary transcripts that encode miRNAs, primiRNAs, contain 5′ cap structures as well as 3′ poly(A) tails. miRNAs can be categorized into three groups according to their genomic locations relative to their positions in an exon or intron.

a | Exonic miRNAs in non-coding transcripts such as an *miR-23a~27a~24-2* cluster, *miR-21* and *miR-155*. *miR-155* was found in a previously defined non-coding RNA (ncRNA) gene, *bic*17.

b | Intronic miRNAs in non-coding transcripts. For example, an *miR-15a~16-1* cluster was found in the fourth intron of a previously defined non-coding RNA gene, *DLEU2* (REF. 126). **c** | Intronic miRNAs in protein-coding transcripts. For example, an *miR-106b~93~25* cluster is embedded in the thirteenth intron of DNA replication licensing factor *MCM7* transcript (variant 1, which encodes isoform 1). The mouse *miR-06b~93~25* homologue is also found in the thirteenth intron of the mouse *MCM7* homologue gene15. The hairpins indicate the miRNA stem-loops. Orange boxes indicate the protein-coding region. This figure is not to scale.

Regulation of miRNA expression: Transcription

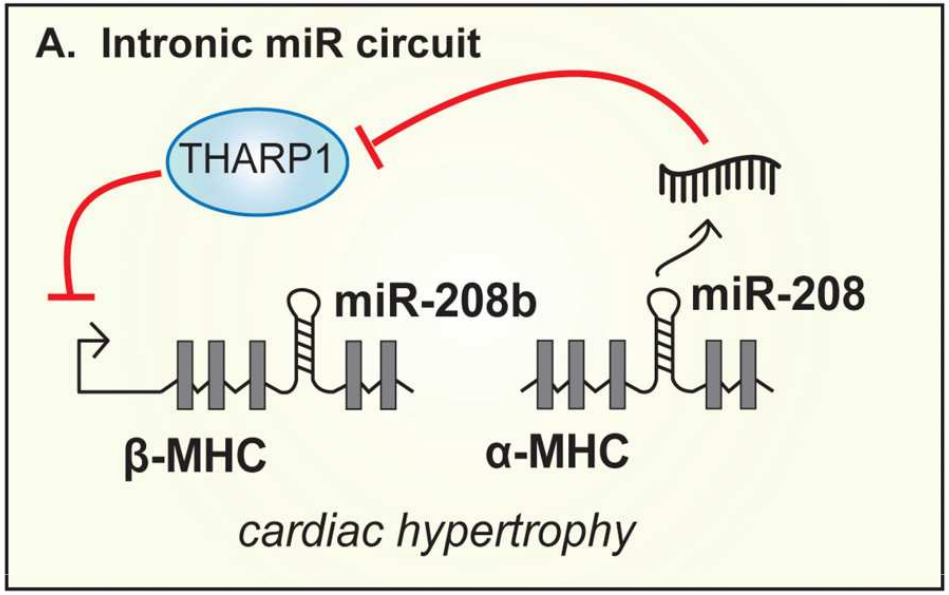
A recent large scale mapping of 175 human miRNA promoters through nucleosome positioning and chromatin immunoprecipitationon-genomic DNA microarray chip (or ChIP-onchip) analysis suggests that

the promoter structure of miRNA genes, including the relative frequencies of CpG islands, TATA box, TFIIB recognition, initiator elements, and histone modifications, is indistinguishable between the promoters of miRNA and mRNA.

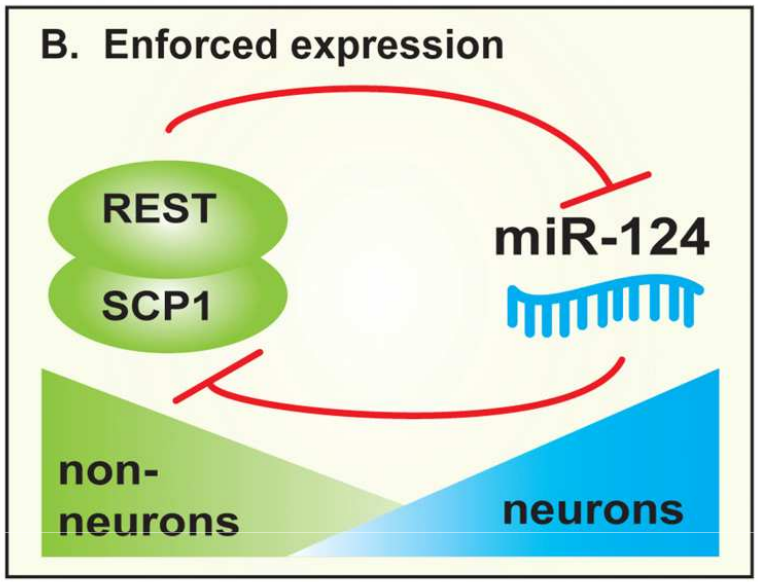
Furthermore, DNA binding factors that regulate miRNA transcription largely overlap with those that control protein coding genes.

It is intersting to note that mRNAs that encode for Transcription factors are very often controlled by miRNA.

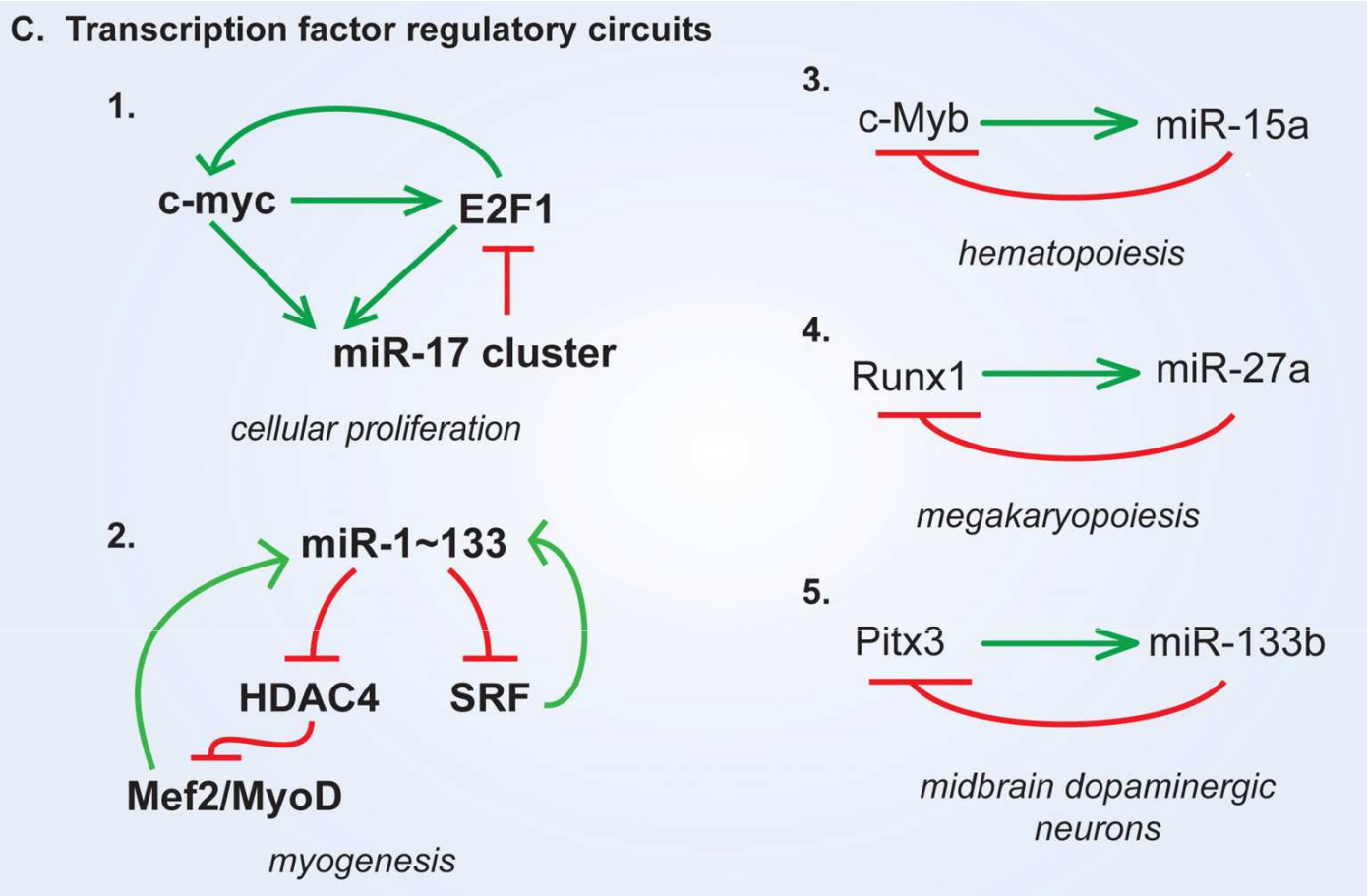
This suggests loop of feedback control.



miRNA regulatory circuits. A. The cardiac specific miR-208 family is encoded within the introns of myosin heavy chain (MHC) genes. miR-208 targets THARP1, which then down regulates the expression of β-MHC gene



B. Expression of miR-124
is negatively regulated by the binding of the RE1 silencing transcription (REST) factor to the promoter in non-neuronal cells



Examples of feed-back regulation of microRNA transcription through the repression of transcription factors.

Additional regulations by various signalling pathways were also noticed on various steps of miRNA biogenesis

Controlling the activity of Drosha, Dicer, Exportin 5, components of the RISC complex.

In addition, some miRNA get edited ! (By the adenosine deaminase enzyme: A→I)

Friard et al. BMC Bioinformatics 2010, 11:435
http://www.biomedcentral.com/1471-2105/11/435

BMC Bioinformatics

DATABASE Open Access

CircuitsDB: a database of mixed microRNA/transcription factor feed-forward regulatory circuits in human and mouse

Olivier Friard[1], Angela Re[2], Daniela Taverna[1,3,4], Michele De Bortoli[1,3], Davide Corá[1,5*]

Abstract

Background: Transcription Factors (TFs) and microRNAs (miRNAs) are key players for gene expression regulation in higher eukaryotes. In the last years, a large amount of bioinformatic studies were devoted to the elucidation of transcriptional and post-transcriptional (mostly miRNA-mediated) regulatory interactions, but little is known about the interplay between them.

Description: Here we describe a dynamic web-accessible database, `CircuitsDB`, supporting a genome-wide transcriptional and post-transcriptional regulatory network integration, for the human and mouse genomes, based on a bioinformatic sequence-analysis approach. In particular, `CircuitsDB` is currently focused on the study of mixed miRNA/TF Feed-Forward regulatory Loops (FFLs), i.e. elementary circuits in which a master TF regulates an miRNA and together with it a set of Joint Target protein-coding genes. The database was constructed using an ab-initio oligo analysis procedure for the identification of the transcriptional and post-transcriptional interactions. Several external sources of information were then pooled together to obtain the functional annotation of the proposed interactions. Results for human and mouse genomes are presented in an integrated web tool, that allows users to explore the circuits, investigate their sequence and functional properties and thus suggest possible biological experiments.

Conclusions: We present `CircuitsDB`, a web-server devoted to the study of human and mouse mixed miRNA/TF Feed-Forward regulatory circuits, freely available at: http://biocluster.di.unito.it/circuits/

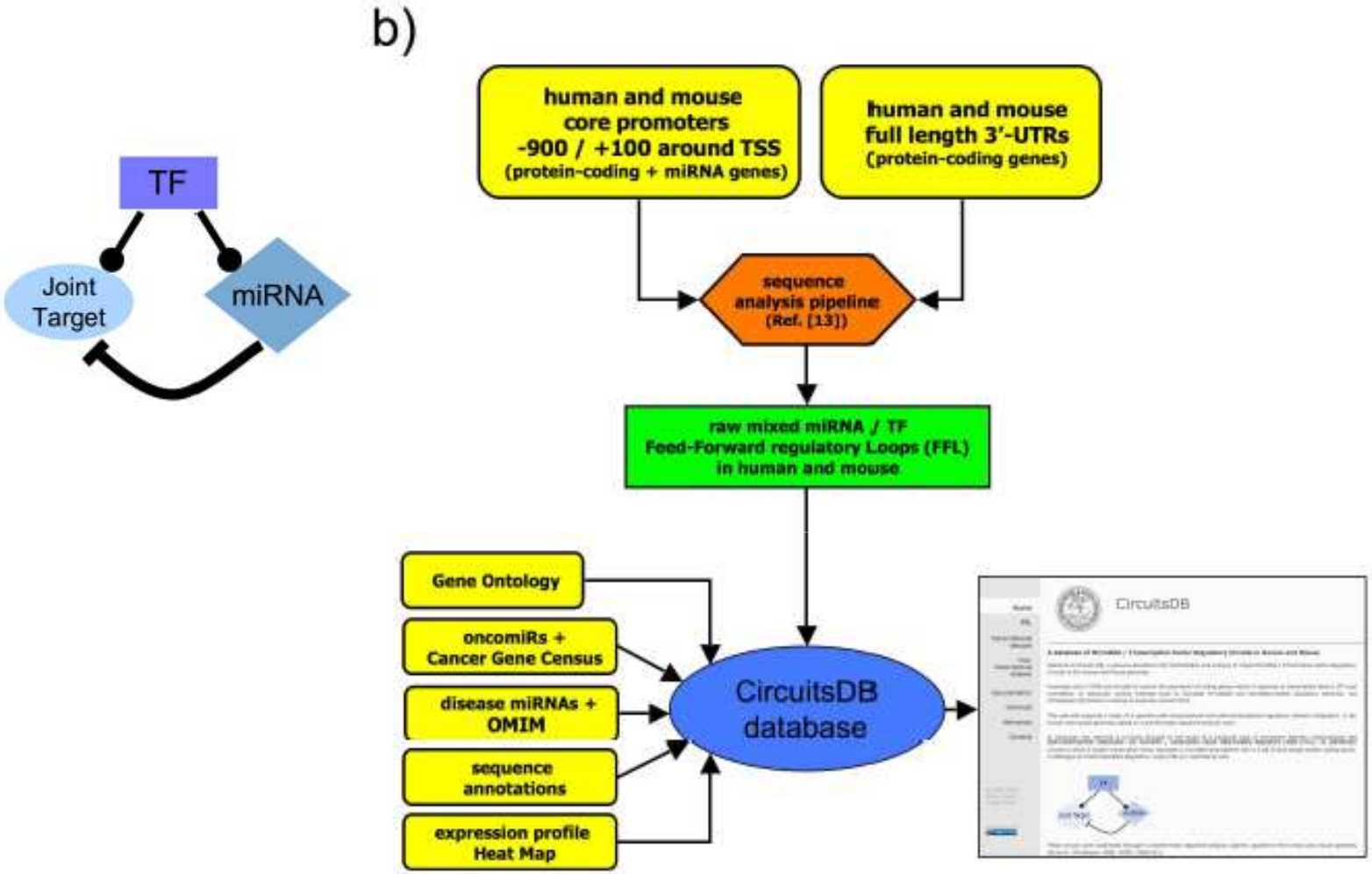


Figure 1 Mixed miRNA/TF Feed-Forward Loops and `CircuitsDB` construction pipeline. a) Representation of a typical mixed Feed-Forward regulatory Loop (FFL) included in `CircuitsDB`. Circuits are composed by a master Transcription Factor (TF, square box) that regulates a microRNA (miRNA, diamond-shaped box) and together with it a Joint Target protein-coding gene (round box). Inside each circuit, → indicates transcriptional activation/repression, whilst ⊣ post-transcriptional repression. b) Flow-chart of the strategy used for the construction of `CircuitsDB`. We first built a genome-wide catalogue of putative core promoter regions for protein-coding and miRNA genes, plus a set of 3'-UTRs for protein-coding genes, in human and mouse. Then we used the analysis pipeline developed in our previous work [13] to infer a dataset of mixed regulatory FFLs, in human and mouse. Finally we integrated different kinds of biological annotations to support the circuits' properties; the `CircuitsDB` web-site allows a dynamic exploration of such properties.

A very complex network, involving transcription factors, epigenetic factors, miRNA, RNA binding proteins and perhaps endosiRNA, together with more classical proteins, is more likely in place to control gene expression.

Understanding and modeling all this is

Systems Biology