Noncoding RNA and Rnomics

Runsheng CHEN

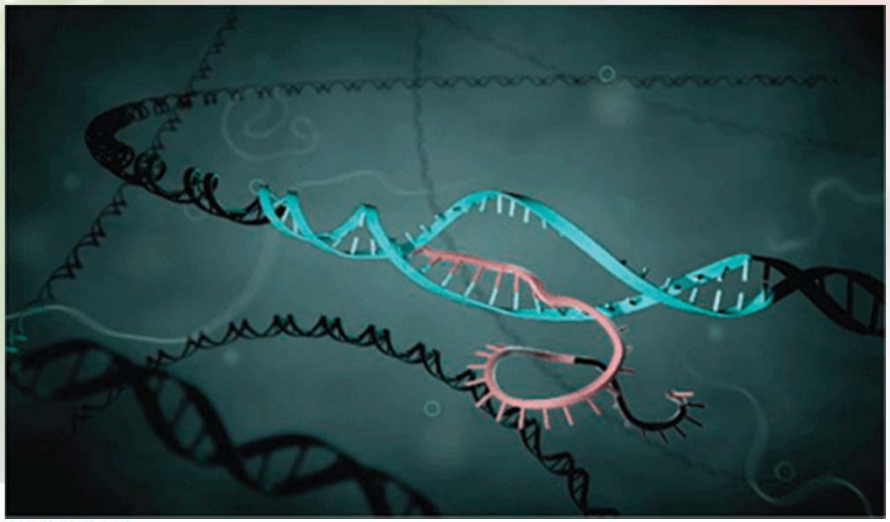
June. 19, 2014

The 17 December 2010 issue of *Science* includes special sections highlighting the **Breakthrough of the Year** and **Insights of the Decade**.

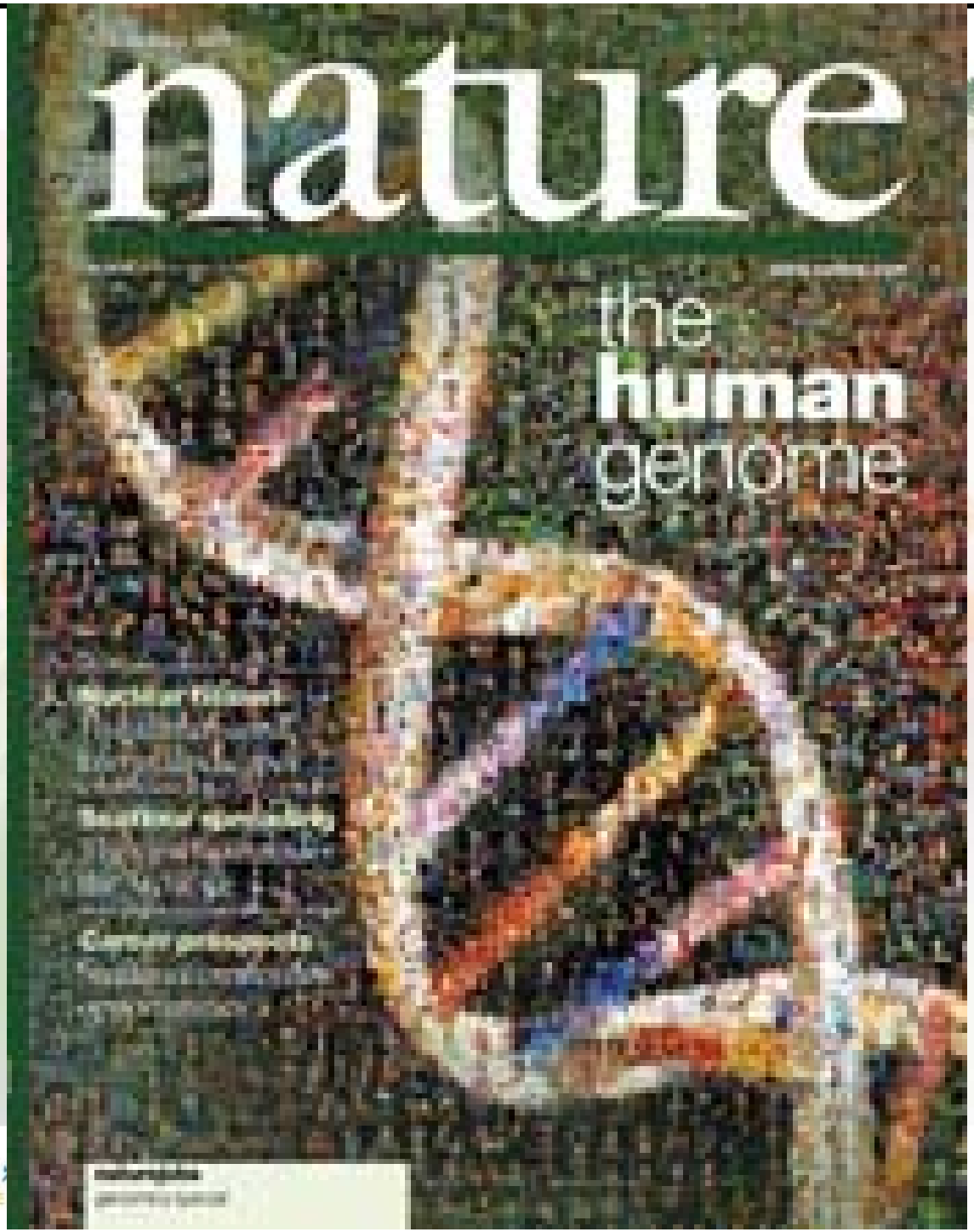
Insights of the Decade

Shining a Light on the Genome's 'Dark Matter'

The scope of this “dark genome” became apparent in 2001, when the human genome was first published. Scientists expected to find as many as 100,000 genes packed into the 3 billion bases of human DNA; they were startled to learn that there were fewer than 35,000. (The current count is 21,000.) Protein-coding regions accounted for just 1.5% of the genome. Could the rest of our DNA really just be junk?



Comparative Genome Analysis

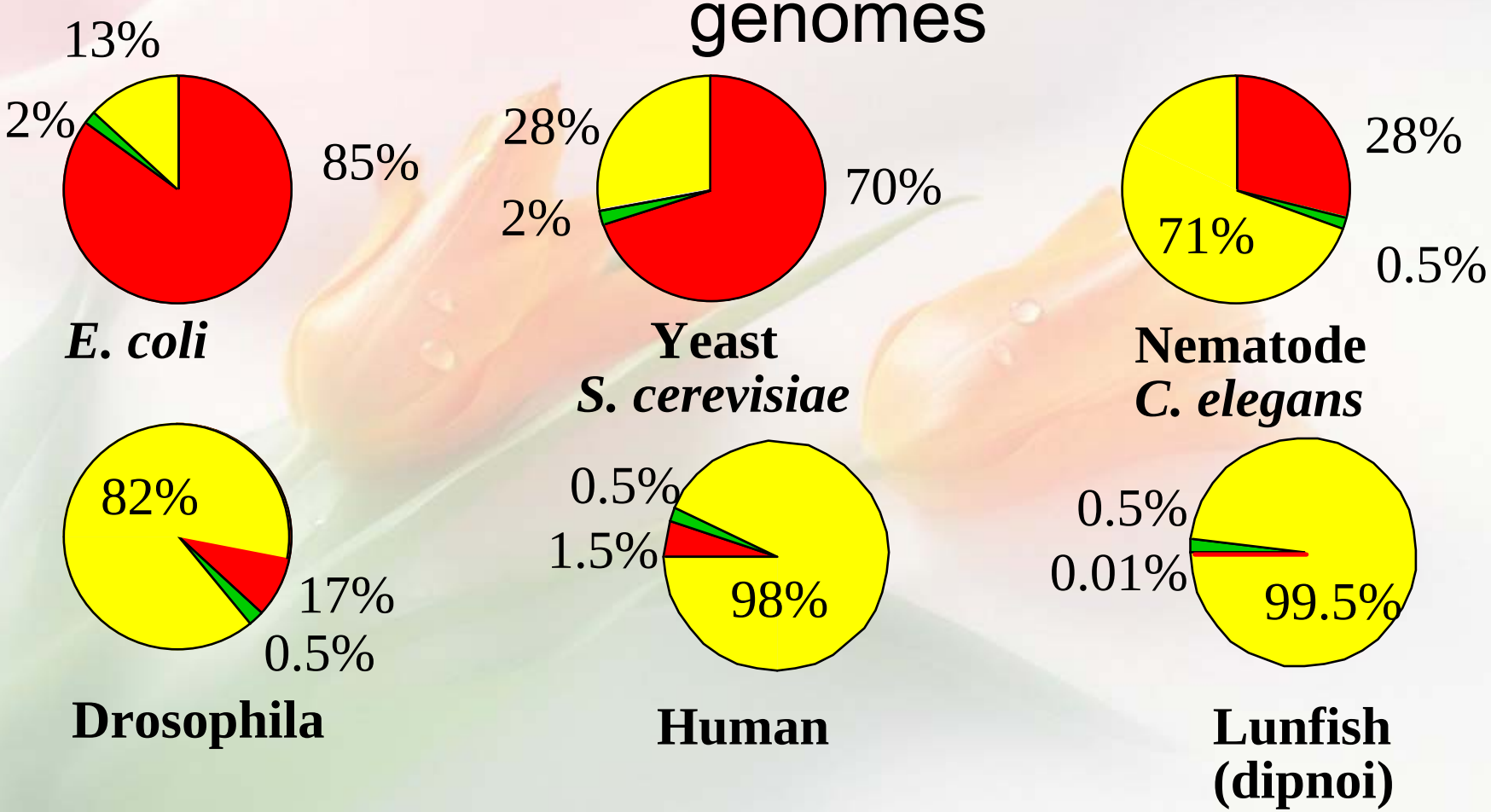

nature
the human genome

Noncoding sequences: Sequences in genome, which are not coding for any proteins.

How many persent of the human genome are noncoding sequences?

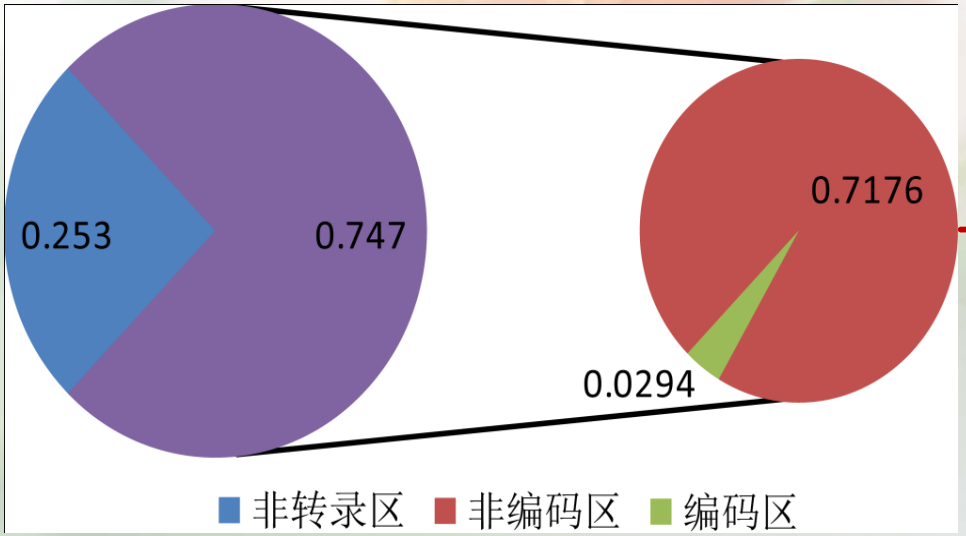
More than 97%!!!

Proportion of functional elements within genomes



Transcriptome Analysis

Transcriptional activity



ENCODE 2012

- Long ncRNA
- microRNA, piRNA
- tRNA,rRNA,snoRNA
- …….

Transcriptional output/complexity

基因组的转录情况

Transcriptional activity *(tiling array data)*基因组的转录水平

–Human 人 基因组的 ≧ 80 % (40-50X)

–C.elegans 线虫 基因组的 ~70 % (2-3X)

Protein coding sequence 编码蛋白序列

--Human 人 基因组的 ~2-3 %

–C.elegans 线虫 基因组的 ~25 %

The majority of transcripts are non-coding RNAs

绝大部分的转录产物是 非编码RNA

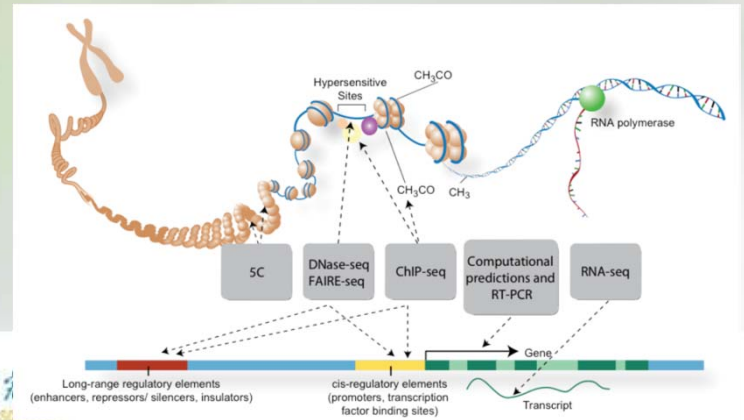
The major differences among different organisms are ncRNAs

物种间最主要的差别也是 非编码RNA

The Encyclopedia of DNA Elements(ENCODE) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute (NHGRI). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

5 September 2012 - ENCODE results published in Nature, Science and other journals

The results of the ENCODE project were published today in a coordinated set of 30 papers published in multiple journals. These publications are the result of cross-consortium integrative analysis, covering more than 4 million regulatory regions in the human genome mapped as part of ENCODE. The coordinated publication set includes one main integrative paper and five other papers in the journal *Nature*; 18 papers in*Genome Research*; and six papers in *Genome Biology*. The ENCODE data are so complex that the three journals have developed a pioneering way to present the information in an integrated form they term "threads." Since the same topics were addressed in different ways in different papers, the*Nature* ENCODE websitewas developed to allow readers to follow a topic through all of the papers in the ENCODE publication set. In addition to these publications, six review articles are being published in the *Journal of Biological Chemistry*, and other affiliated papers in*Science*, *Cell*, and other journals. The newIntegrative Analysis page on this portal provides links and descriptive material for these publications and related analysis resources.



Function of Noncoding RNA

X-inactivation is the mammalian dosage compensation mechanism, used to equalize X-linked gene dosage between male and female cells.

Xist encodes a large, spliced, polyadenylated, noncoding RNA that is expressed exclusively from the otherwise inactive X chromosome.

http://www.ucsf.edu/pibs/faculty/panning.html
Barbara Panning Lab

Small RNA and RNA Interference（RNAi）

In RNAi, dsRNA introduced into susceptible organisms is processed into ~22 nucleotide (nt) siRNAs. These 22 nt siRNAs subsequently bind to the homologous region of their target transcript and tag it for nuclease cleavage. Thus gene silencing is effected by destruction of the target mRNA.

A novel class of small RNAs bind to MILI protein in mouse testes

NATURE|Vol 442|13 July 2006

The small RNAs are 26–31 nucleotides (nt) in length—clearly distinct from the 21–23 nt of microRNAs (miRNAs) or short interfering RNAs (siRNAs)—and we refer to them as ‘Piwi-interacting RNAs’ or piRNAs.

Noncoding RNA in Plants

http://www.prl.msu.edu/PLANTncRNAs/

Prions 'borrow' RNA

Nature 16OCT 2003

An experiment in a cell-free amplification system shows that unidentified host RNA molecules are required for efficient conversion of normal prion protein into its pathogenic form.

Elevated expression of **PCGEM1**, a prostate-specific gene with cell growth-promoting function, is associated with high-risk prostate cancer patients

Gyorgy Petrovics*,1, Wei Zhang2, Mazen Makarem1, Jesse P Street1, Roger Connelly1, Leon Sun1, Isabell A Sesterhenn2, Vasantha Srikantan1, Judd W Moul1,3 and Shiv Srivastava1

Oncogene (2004) 23, 605–611

PCGEM1 appears to be a noncoding functional RNA gene (Srikantan et al., 2000).

His-1: A noncoding RNA implicated in mouse leukemogenesis

Fan Xu, Molly McFarland and David S. Askew*

Histol. Histopathol., 1999, 14, 235–241.

The *His-1* gene is highly conserved among vertebrate species and is transcribed as a single spliced and polyadenylated cytoplasmic RNA that shares several features in common with the emerging class of untranslated RNAs.

MALAT-1, a novel noncoding RNA, and thymosin b4 predict metastasis and survival in early-stage non-small cell lung cancer

Ping Ji1,5, Sven Diederichs1,5, Wenbing Wang1, Sebastian Bo¨ ing1, Ralf Metzger2, Paul M Schneider2, Nicola Tidow3, Burkhard Brandt3, Horst Buerger4, Etmar Bulk1, Michael Thomas1,Wolfgang E Berdel1, Hubert Serve*,1 and Carsten Mu¨ ller-Tidow*,1

Oncogene (2003) 22, 8031–8041

Evidence for evolutionarily conserved secondary structure in the *H19* tumor suppressor RNA

Veronica Juan, Chad Crain and Charles Wilson

Nucleic Acids Research 2000 Vol. 28 1221-1227

Comparison between Protein and ncRNA Functions in the Cell
(From Prof. Xiangdong Fu)

Function	**Protein**		**ncRNA**
Enzymes	Catalytic centers	>	RNase P, Group II introns
Structural components	Structural modules	=	Scaffold in the nucleus
Molecular interactions	Inter/intra-mol interactions	=	Nucleation of interactions
Intracellular targeting	Specific localization/function	=	Partitioning nuclear domains
Chaperons	Assisted protein folding	=	Allosteric inducers
Small molecule channels	Transport across membrane	0	Unlikely to have similar role
Secreted signals/hormones	Long distance regulation	=	Secreted miRs
Modifications	Various PTMs	=	Multiple base modifications

Biological Dark Matter

Newfound RNA suggests a hidden complexity inside cells

John Travis

In the early 1990s, Victor Ambros and his colleagues were conducting a gene hunt. In particular, they were searching for the gene that was mutated in a perplexing strain of *Caenorhabditis elegans*, the small nematode whose development many biologists study. Unlike most genes, the one identified by Ambros' group doesn't encode a protein. It spawns a small molecule of RNA—a chemical relative of DNA—that somehow turns off other genes that play a role in worm development. Several groups, including one led by Eddy, Ambros' team and two other research groups reported that *Escherichia coli* , worms, flies, and people contain dozens of previously undetected genes that spawn RNA instead of protein.

The RNA genes found so far are "just the tip of a huge iceberg," says Ruvkun.

Prospect and List of Notable Aspect

一、Long Noncoding RNA should be the Major Parts of Noncoding RNA

>200nt　John Mattick

>100nt　my group

Number: Large

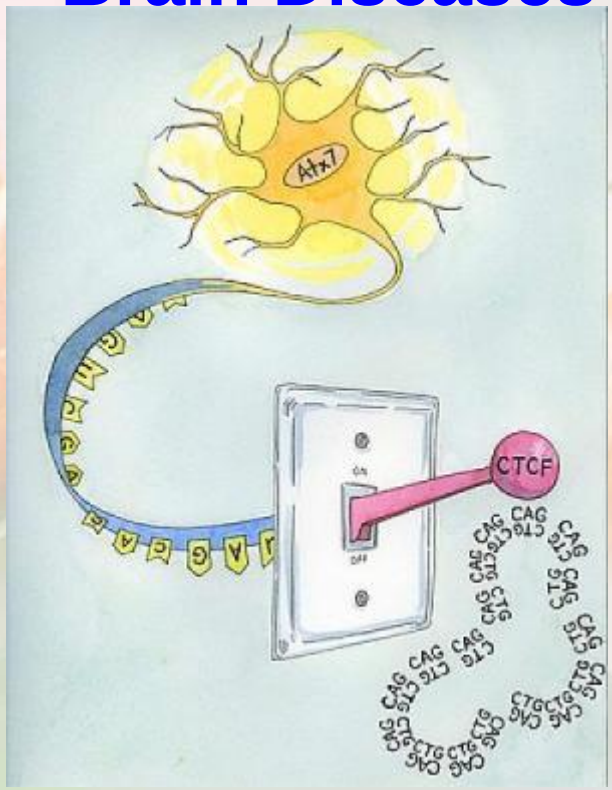
Capacity forming 3D-structure: Better

The way for biological function: Different

short NcRNA: base match

long NcRNA: 3D-structure interaction

Non-Coding RNA Has Role in Inherited Neurological Disorder, and Maybe Other Brain Diseases Too

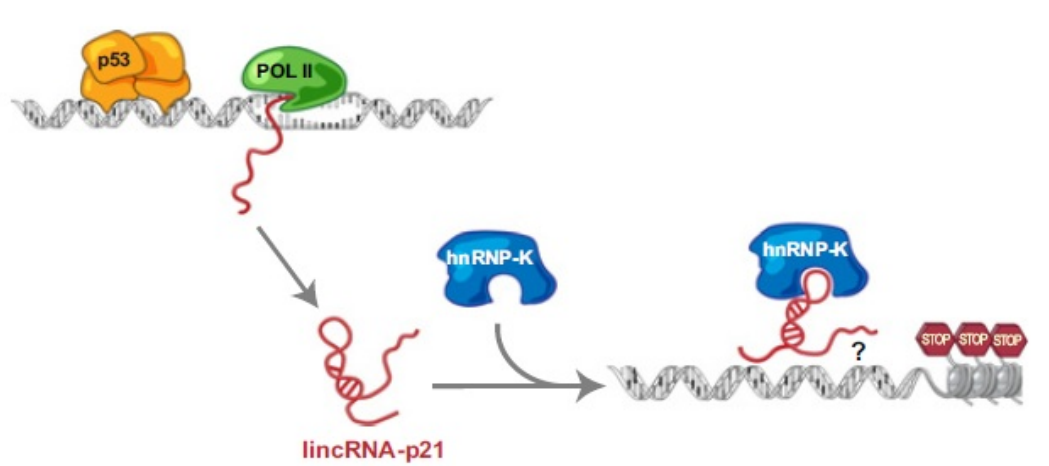


shiver

Researchers have discovered that expression of the ataxin-7 gene -- the cause of the neurological disorder spinocerebellar ataxia type 7 -- has two regulators: a highly conserved, multi-tasking protein called CTCF and, surprisingly, an adjacent promoter containing non-coding RNA. (Credit: Illustration courtesy of ChristinaTakamatsu-Butler, UC San Diego, published in the June 22 issue of the journal *Neuron.)*

A Large Intergenic Noncoding RNA Induced by p53 Mediates Global Gene Repression in the p53 Response

Cell 142, 409–419, August 6, 2010



在这篇文章，作者发现P53可以直接提高lincRNA-P21的表达，然后lincRNA-P21与蛋白hnRNP-K结合，再调节其他基因的表达（如图）。

For example, the ncRNA Evf-2 functions as a co-activator for the homeobox transcription factor Dlx2, which plays important roles in forebrain development and neurogenesis (Feng 2006; Panganiban 2002). Sonic hedgehog induces transcription of Evf-2 from an ultra-conserved element located between the Dlx5 and Dlx6 genes during forebrain development (Feng 2006). Evf-2 then recruits the Dlx2 transcription factor to the same ultra-conserved element whereby Dlx2 subsequently induces expression of Dlx5. The existence of other similar ultra- or highly conserved elements within the mammalian genome that are both transcribed and fulfil enhancer functions suggest Evf-2 may be illustrative of a generalised mechanism that tightly regulates important developmental genes with complex expression patterns during vertebrate growth (Pennacchio 2006; Visel 2008).

*Feng J, Bi C, Clark BS, Mady R, Shah P, Kohtz JD (June 2006). "The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator". Genes & Development **20** (11): 1470–84. doi:10.1101/gad.1416106. PMC 1475760. PMID 16705037.*

transcriptional regulation

The formation of RNA duplexes between complementary ncRNA and mRNA may mask key elements within the mRNA required to bind trans-acting factors, potentially effecting any step in post-transcriptional gene expression including pre-mRNA processing and splicing, transport, translation, and degradation.

The splicing of mRNA can induce its translation and functionally diversify the repertoire of proteins it encodes. The Zeb2 mRNA, which has a particularly long 5'UTR, requires the retention of a 5'UTR intron that contains an internal ribosome entry site for efficient translation (Beltran 2008). However, retention of the intron is dependent on the expression of an antisense transcript that complements the intronic 5' splice site (Beltran 2008). Therefore, the ectopic expression of the antisense transcript represses splicing and induces translation of the Zeb2 mRNA during mesenchymal development.

Post-transcriptional regulation

Indeed it was recently shown that BC1 is associated with translational repression in dendrites to control the efficiency of dopamine D2 receptor-mediated transmission in the striatum (Centonze 2007) and BC1 RNA-deleted mice exhibit behavioural changes with reduced exploration and increased anxiety (Lewejohann 2004).

Centonze D, Rossi S, Napoli I, et al. (August 2007). "The brain cytoplasmic RNA BC1 regulates dopamine D2 receptor-mediated transmission in the striatum". The Journal of Neuroscience ***27*** *(33): 8885–92. doi:10.1523/JNEUROSCI.0548-07.2007. PMID 17699670.*

Lewejohann L, Skryabin BV, Sachser N, et al. (September 2004). "Role of a neuronal small non-messenger RNA: behavioural alterations in BC1 RNA-deleted mice". Behavioural Brain Research ***154*** *(1): 273–89. doi:10.1016/j.bbr.2004.02.015. PMID 15302134.*

translational regulation

Epigenetic modifications, including histone and DNA methylation, histone acetylation and sumoylation, affect many aspects of chromosomal biology, primarily including regulation of large numbers of genes by remodeling broad chromatin domains. While it has been known for some time that RNA is an integral component of chromatin, it is only recently that we are beginning to appreciate the means by which RNA is involved in pathways of chromatin modification (Chen 2008; Rinn 2007; Sanchez-Elsner 2006)

Chen X, Xu H, Yuan P, et al. (June 2008). "Integration of external signaling pathways with the core transcriptional network in embryonic stem cells". Cell ***133*** *(6): 1106–17. doi:10.1016/j.cell.2008.04.043. PMID 18555785.*

Rinn JL, Kertesz M, Wang JK, et al. (June 2007). "Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs". Cell ***129*** *(7): 1311–23. doi:10.1016/j.cell.2007.05.022. PMC 2084369. PMID 17604720.*

Sanchez-Elsner T, Gou D, Kremmer E, Sauer F (February 2006). "Noncoding RNAs of trithorax response elements recruit Drosophila Ash1 to Ultrabithorax". Science ***311*** *(5764): 1118–23. Bibcode:2006Sci...311.1118S. doi:10.1126/science.1117705. PMID 16497925.*

epigenetic regulation

Many emergent themes of ncRNA-directed chromatin modification were first apparent within the phenomenon of imprinting, whereby only one allele of a gene is expressed from either the maternal or the paternal chromosome. In general, imprinted genes are clustered together on chromosomes, suggesting the imprinting mechanism acts upon local chromosome domains rather than individual genes. These clusters are also often associated with long ncRNAs whose expression is correlated with the repression of the linked protein-coding gene on the same allele. Indeed, detailed analysis has revealed a crucial role for the ncRNAs Kcnqot1 and Igf2r/Air in directing imprinting (Braidotti 2004). Almost all the genes at the Kcnq1 loci are maternally inherited, except the paternally expressed antisense ncRNA Kcnqot1 (Mitsuya 1999).

Braidotti G, Baubec T, Pauler F, et al. (2004). "The Air noncoding RNA: an imprinted cis-silencing transcript". Cold Spring Harbor Symposia on Quantitative Biology ***69****: 55–66. doi:10.1101/sqb.2004.69.55. PMID 16117633.*

Mitsuya K, Meguro M, Lee MP, et al. (July 1999). "LIT1, an imprinted antisense RNA in the human KvLQT1 locus identified by screening for differentially expressed transcripts using monochromosomal hybrids". Human Molecular Genetics ***8*** *(7): 1209–17. doi:10.1093/hmg/8.7.1209. PMID 10369866.*

Imprinting

Telomeres have been long considered transcriptionally inert DNA-protein complexes until it was recently shown that telomeric repeats may be transcribed as telomeric RNAs (TelRNAs) (Schoeftner 2008) or telomeric repeat-containing RNAs (Azzalin 2007). These ncRNAs are heterogeneous in length, transcribed from several sub-telomeric loci and physically localise to telomeres. Their association with chromatin, which suggests an involvement in regulating telomere specific heterochromatin modifications, is repressed by SMG proteins that protect chromosome ends from telomere loss (Azzalin 2007). In addition, TelRNAs block telomerase activity in vitro and may therefore regulate telomerase activity (Schoeftner 2008).

*Azzalin CM, Reichenbach P, Khoriauli L, Giulotto E, Lingner J (November 2007). "Telomeric repeat containing RNA and RNA surveillance factors at mammalian chromosome ends". Science **318** (5851): 798–801. Bibcode:2007Sci...318..798A. doi:10.1126/science.1147182. PMID 17916692.*

*Schoeftner S, Blasco MA (February 2008). "Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II". Nature Cell Biology **10** (2): 228–36. doi:10.1038/ncb1685. PMID 18157120.*

Telomeric non-coding RNAs

Long noncoding RNAs: functional surprises from the RNA world

Genes Dev. 2009 23: 1494-1504
Access the most recent version at doi:10.1101/gad.1800909

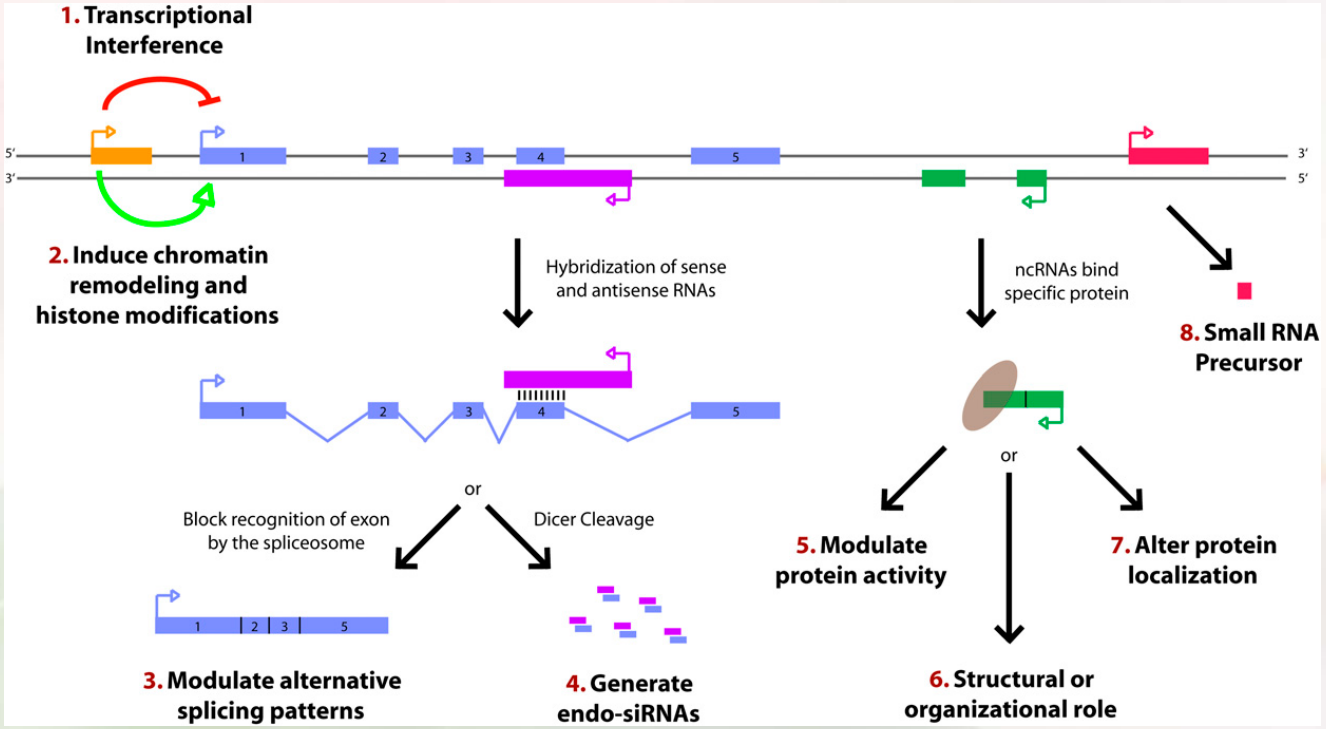


Figure 1. Paradigms for how long ncRNAs function. Recent studies have identified a variety of regulatory paradigms for how long ncRNAs function, many of which are highlighted here. Transcription from an upstream noncoding promoter (orange) can negatively (1) or positively (2) affect expression of the downstream gene (blue) by inhibiting RNA polymerase II recruitment or inducing chromatin remodeling, respectively. (3) An antisense transcript (purple) is able to hybridize to the overlapping sense transcript (blue) and block recognition of the splice sites by the spliceosome, thus resulting in an alternatively spliced transcript. (4) Alternatively, hybridization of the sense and antisense transcripts can allow Dicer to generate endogenous siRNAs. By binding to specific protein partners, a noncoding transcript (green) can modulate the activity of the protein (5), serve as a structural component that allows a larger RNA–proteincomplex to form (6), or alter where the protein localizes in the cell (7). (8) Long ncRNAs (pink) can be processed to yield small RNAs, such as miRNAs, piRNAs, and other less well-characterized classes of small transcripts.

A ceRNA Hypothesis: The Rosetta Stone of a Hidden RNA Language?

Leonardo Salmena, Laura Poliseno, Yvonne Tay, Lev Kats, Pier Paolo Pandolfi

Cell, Volume 146, Issue 3, 353-358, 28 July 2011

竞争性内源RNA（competing endogenous RNA，ceRNA)

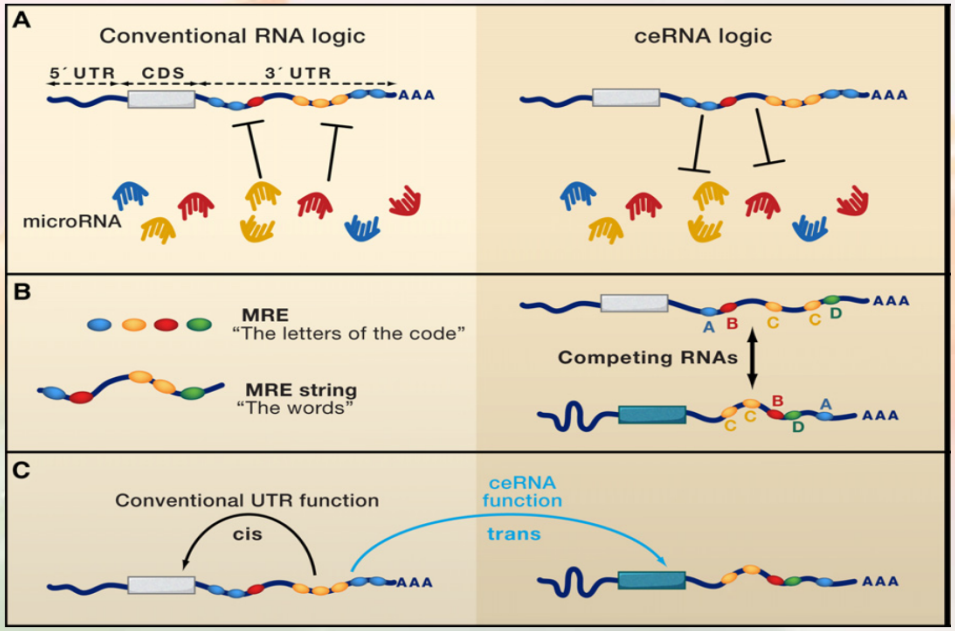


Figure 1. The Basis of the ceRNA Language

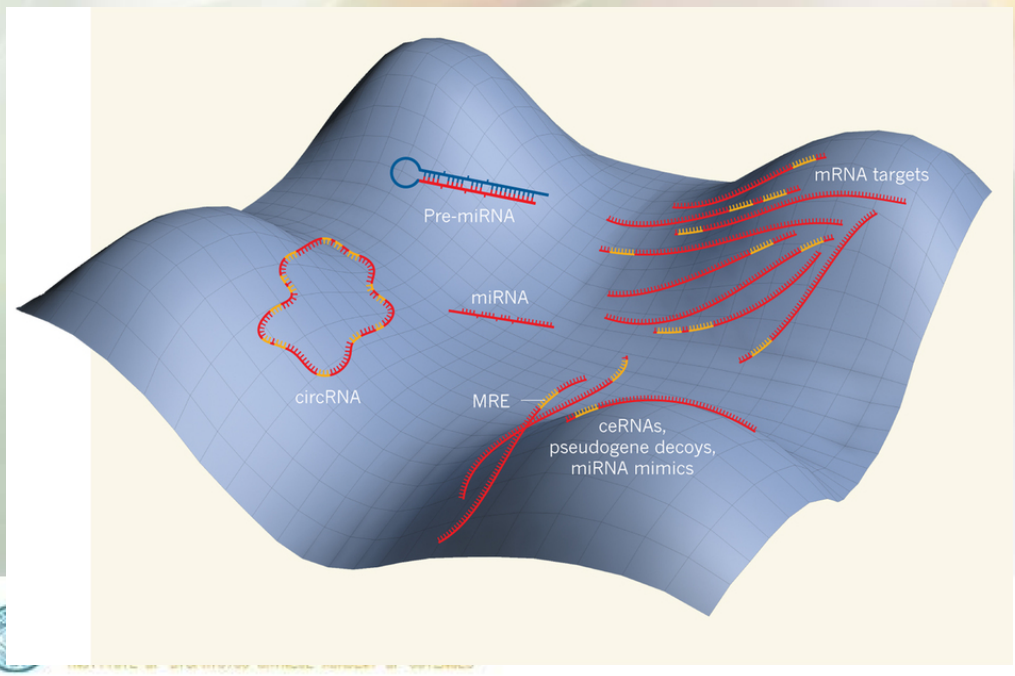
How mRNAs affect microRNAs is less well characterized than how microRNAs affect mRNAs. (A) The relationship between mRNAs and microRNAs could be reciprocal (Seitz, 2009), causing the level of one mRNA to influence the level and activity of another mRNA. (B) Thus, RNA molecules could communicate with each other through microRNA and microRNA response sequences (MREs). The greater the number of shared MREs, the greater the level of "communication" and thus coregulation. (C) The 30 UTRs of RNA molecules contain MREs, which can function in cis to regulate the RNA molecule itself but also possibly in trans to regulate levels of microRNAs and consequently other RNAs.

环RNA怎样像海绵一样吸收微RNA？

环RNA (circRNA)已在哺乳动物细胞中被发现，但它们的功能一直不清楚。现在，来自Nikolaus Rajewsky实验室和Jørgen Kjems实验室的两篇论文确定了与微RNA miR-7相结合的一个环RNA的一种功能。他们发现，这个环RNA充满了微RNA结合点，可起“海绵”的作用，能在每个环RNA分子上结合大量微RNA。这些研究说明环RNA在转录后调控中扮演一个角色。（Link to Article p. 333; Letter p. 384; News & Views p. 322）

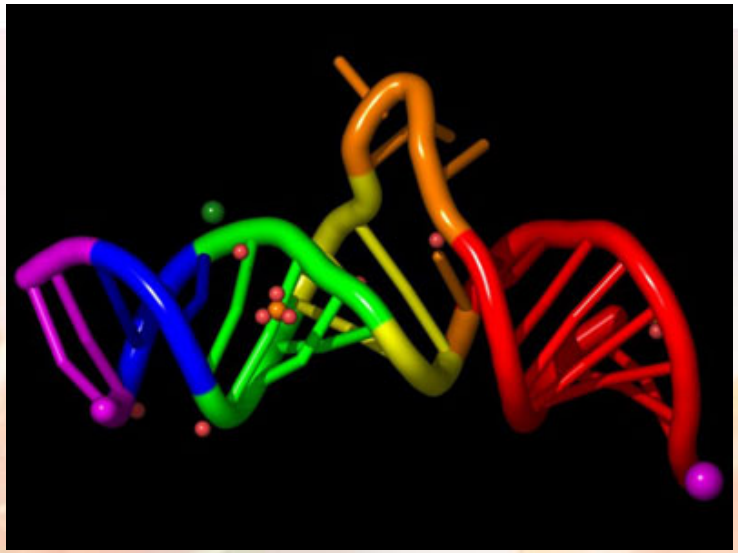
Sebastian Memczak1*, Marvin Jens1*, Antigoni Elefsinioti1*, Francesca Torti1*, Janna Krueger2, Agnieszka Rybak1, Luisa Maier1, Sebastian D. Mackowiak1, Lea H. Gregersen3, Mathias Munschauer3, Alexander Loewer4, Ulrike Ziebold1, Markus Landthaler3, Christine Kocks1, Ferdinand le Noble2 & Nikolaus Rajewsky1, **Circular RNAs are a large class of animal RNAs with regulatory potency,** ***21 MARCH 2013 | VOL 495 | NATURE | 333***

Thomas B. Hansen1, Trine I. Jensen1, Bettina H. Clausen2, Jesper B. Bramsen1,3, Bente Finsen2, Christian K. Damgaard1 & Jorgen Kjems1,3, **Natural RNA circles function as efficient microRNA sponges,** ***384 | NATURE | VOL 495 | 21 MARCH 2013***



MicroRNAs (miRNAs) lie in a fitness valley constrained by their numerous interactions, which include those with the hairpin structure of the precursor miRNA (pre-miRNA), the many target mRNAs and other RNAs that terminate or modulate miRNA binding to target sequences by competing against them. The latter category includes competing endogenous RNAs (ceRNAs), pseudogene decoys and miRNA mimics. Two studies[1, 2] introduce circular RNAs (circRNAs) as another constraining factor. MRE, miRNA-response element.

eRNA(enhancer RNAs, enhancer-directed RNAs)



Michael T. Y. Lam, Han Cho, Hanna P. Lesch, David Gosselin, Sven Heinz, Yumiko Tanaka-Oishi, Christopher Benner, Minna U. Kaikkonen, Aneeza S. Kim, Mika Kosaka, Cindy Y. Lee, Andy Watt, Tamar R. Grossman, Michael G. Rosenfeld, Ronald M. Evans, Christopher K. Glass. **Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription**. Nature, 02 June 2013; DOI: 10.1038/nature12209

Wenbo Li, Dimple Notani, Qi Ma, Bogdan Tanasa, Esperanza Nunez, Aaron Yun Chen, Daria Merkurjev, Jie Zhang, Kenneth Ohgi, Xiaoyuan Song, Soohwan Oh, Hong-Sook Kim, Christopher K. Glass, Michael G. Rosenfeld. **Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation**. Nature, 02 June 2013; DOI:10.1038/nature12210

Tae-Kyung Kim,1, 9, 10， Martin Hemberg,2, 9， Jesse M. Gray,1, 9， Allen M. Costa,1， Daniel M.，Bear,1， Jing Wu,3， David A. Harmin,1, 4， Mike Laptewicz,1， Kellie Barbara-Haley,5， Scott Kuersten,6， Eirene Markenscoff-Papadimitriou,1, 10，Dietmar Kuhl,7 ，Haruhiko Bito,8， Paul F. Worley,3， Gabriel Kreiman2 & Michael E. Greenberg1，**Widespread transcription at neuronal activity-regulated enhancers**。Nature, 13 May 2010 DOI:doi:10.1038/nature09033

Noncoding RNA Research in our Lab

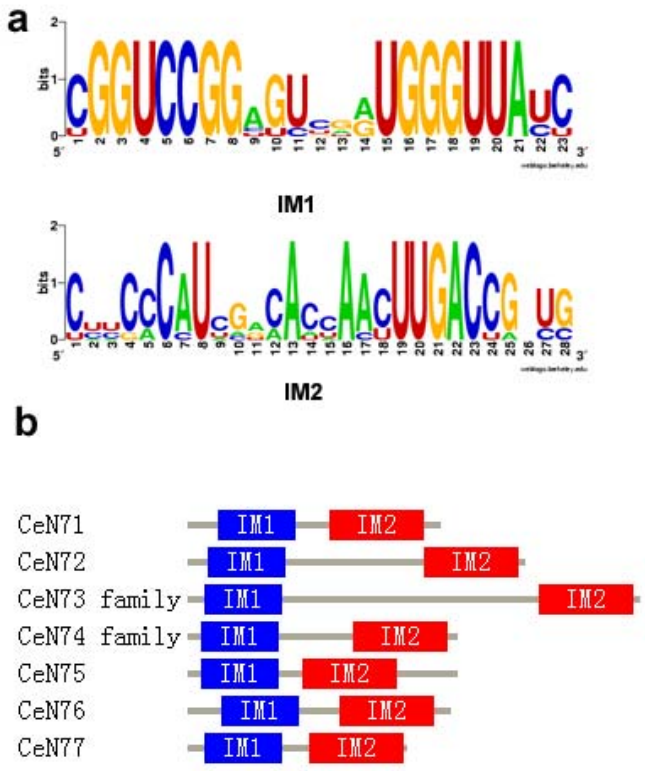
1、Found 164 novel noncoding RNAs and their genes in C.elegans

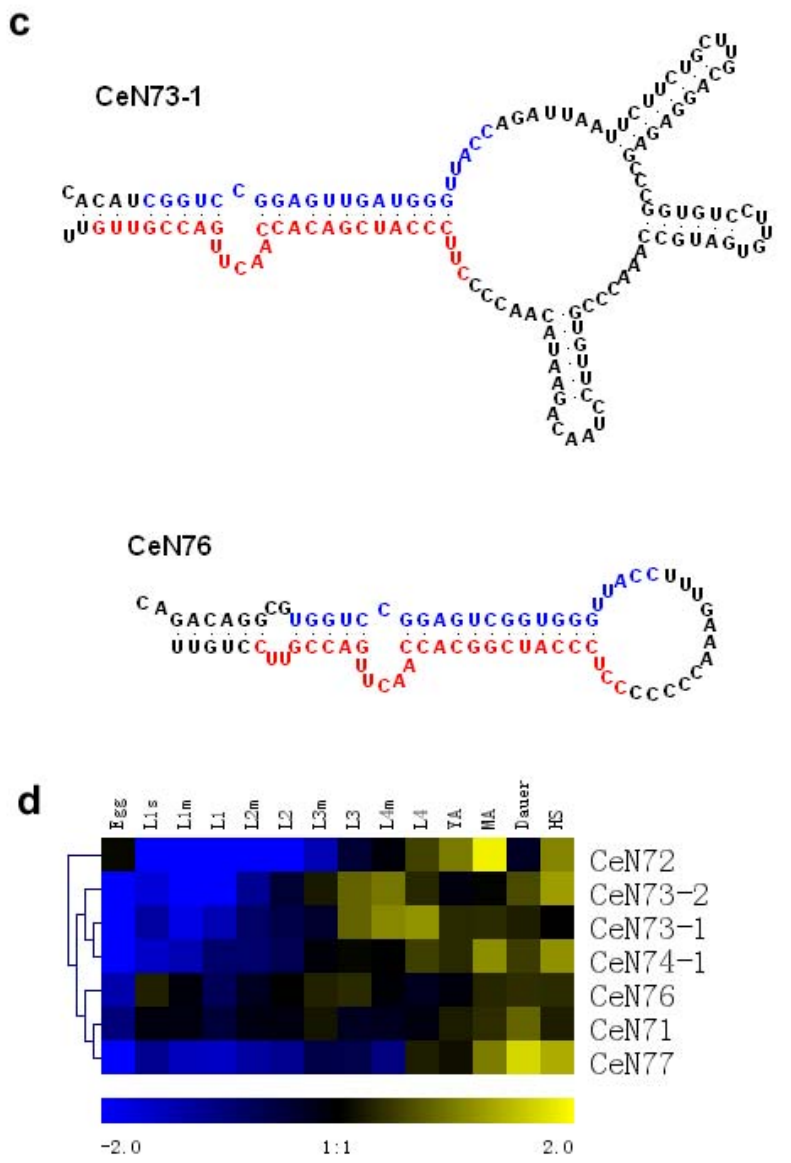
通过建立新的实验方案，在C.elegans基因组中发现了100个迄今国际上未见报导的全新的非编码RNA基因。这些基因不同于已知的各种类的NcRNA，也不同于**microRNA**。这是新类型的RNA基因。论文已发表在

Genome Research 16: 20-29, 2006;

NCBI accession number: AY948555-- AY948719

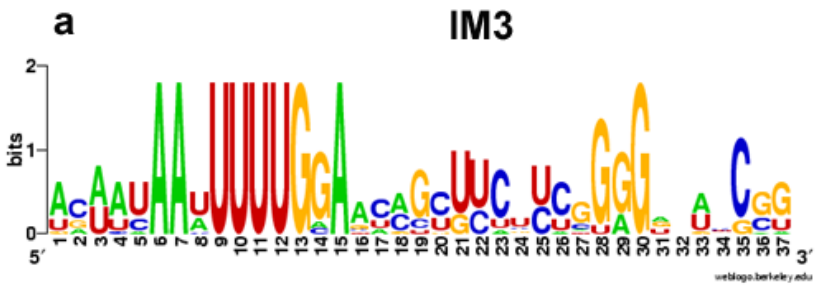
发现两个新非编码RNA基因家族
(To classify two new categories)

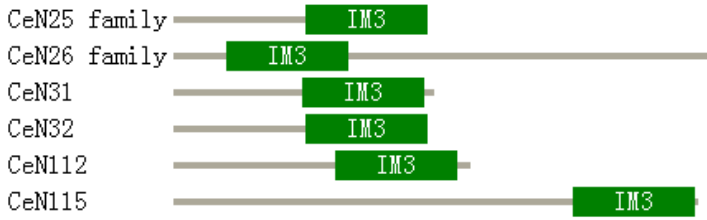


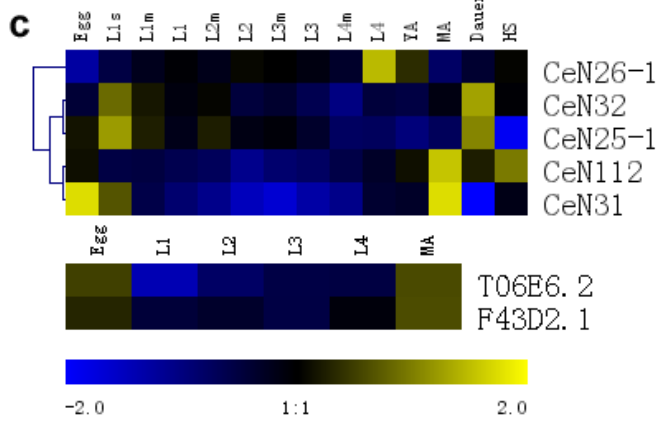


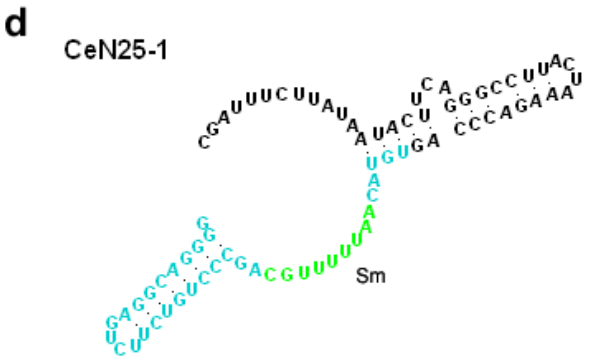
The stem-bulge RNAs of *C. elegans*

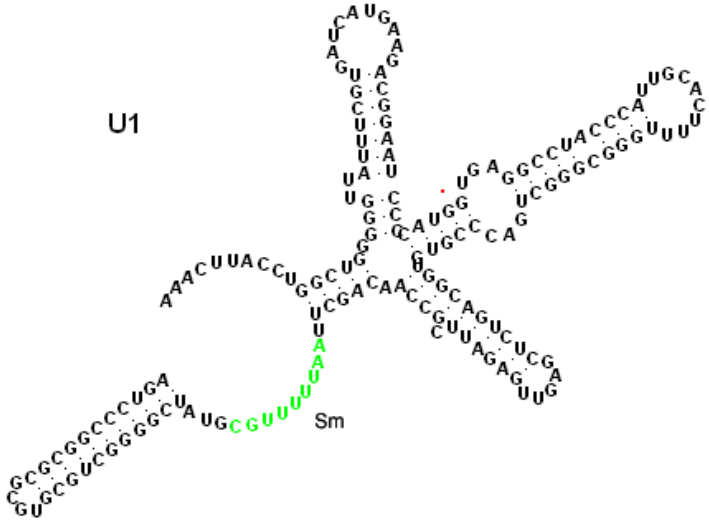
The snRNA like RNAs of *C. elegans*



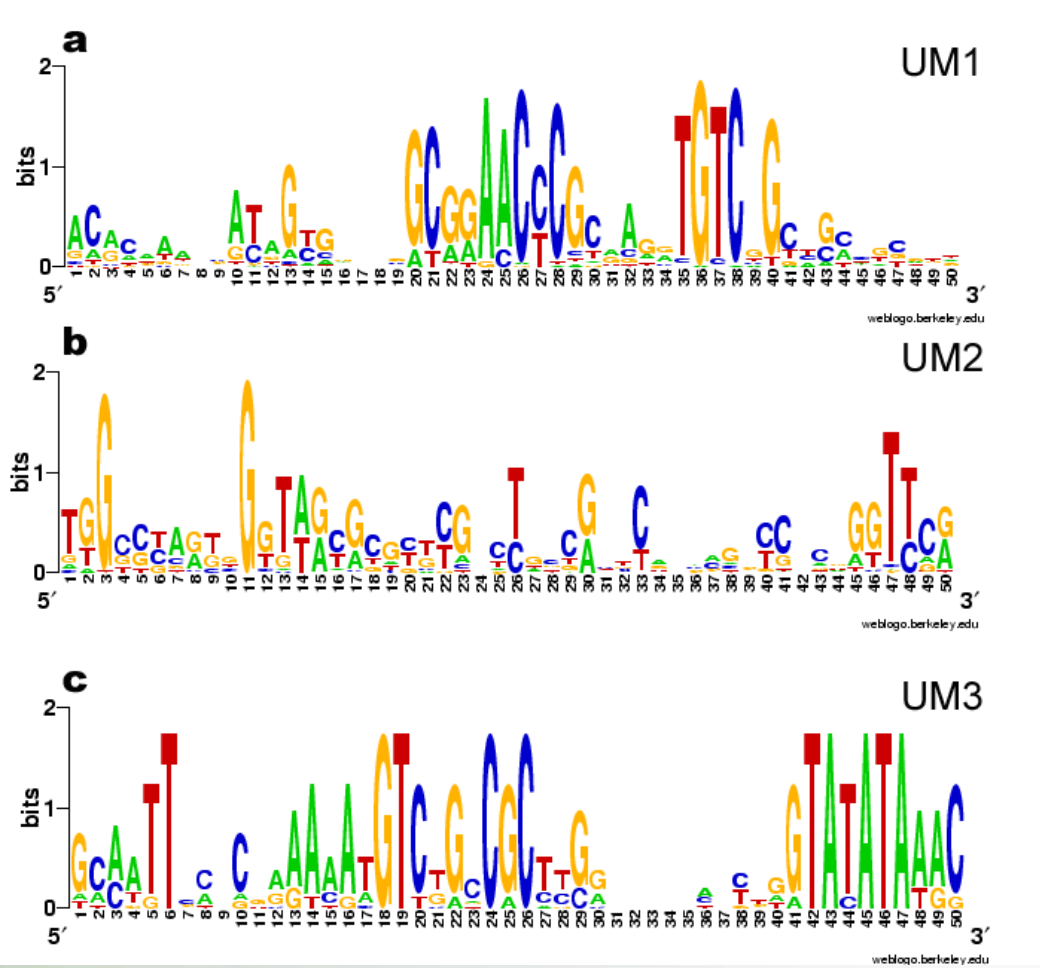








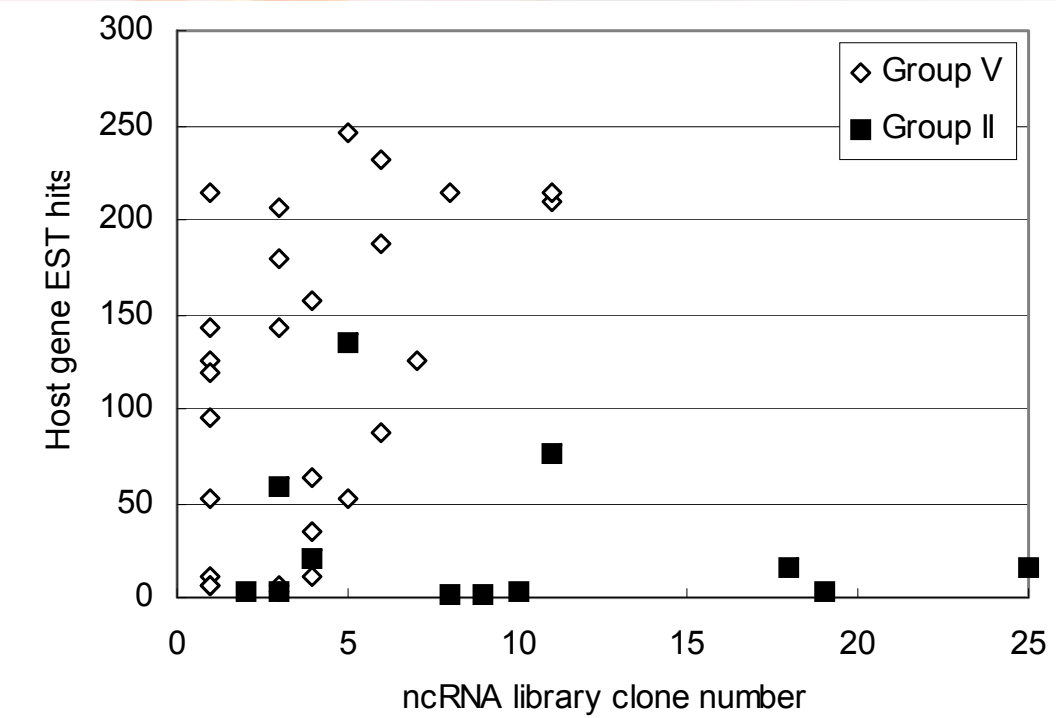
发现了三个独特的非编码基因上游motif (UM1-3)
(confirm three special upstream motifs of noncoding genes)



located within 40-80 bp upstream of the transcription initiation sites of the ncRNA loci were further revealed by MEME *(Bailey and Elkan, 1995).*

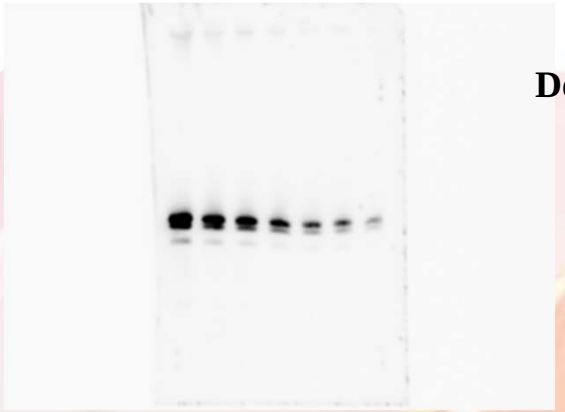
提出了非编码基因是一个独立的系统并给出了其基本调控模式

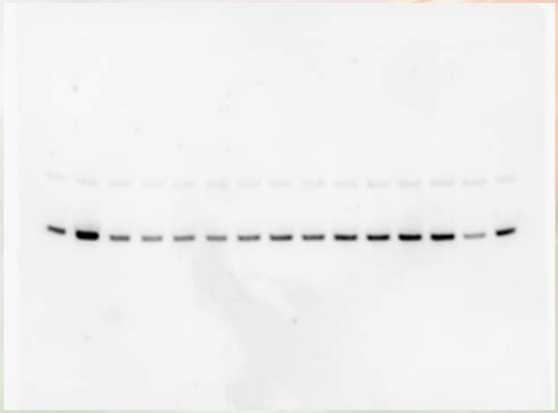
(found that many of the ncRNA genes are located in the introns of host protein-coding genes and are under the control of independent promoter elements.)

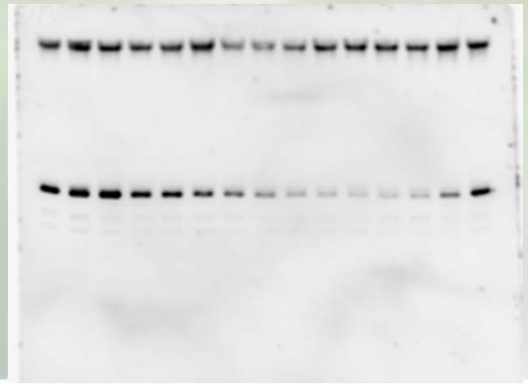


The expression levels of non-motif snoRNAs with the frequencies of ESTs corresponding to exons of their host genes, produced a distinct positive correlation not found for motif-containing loci.

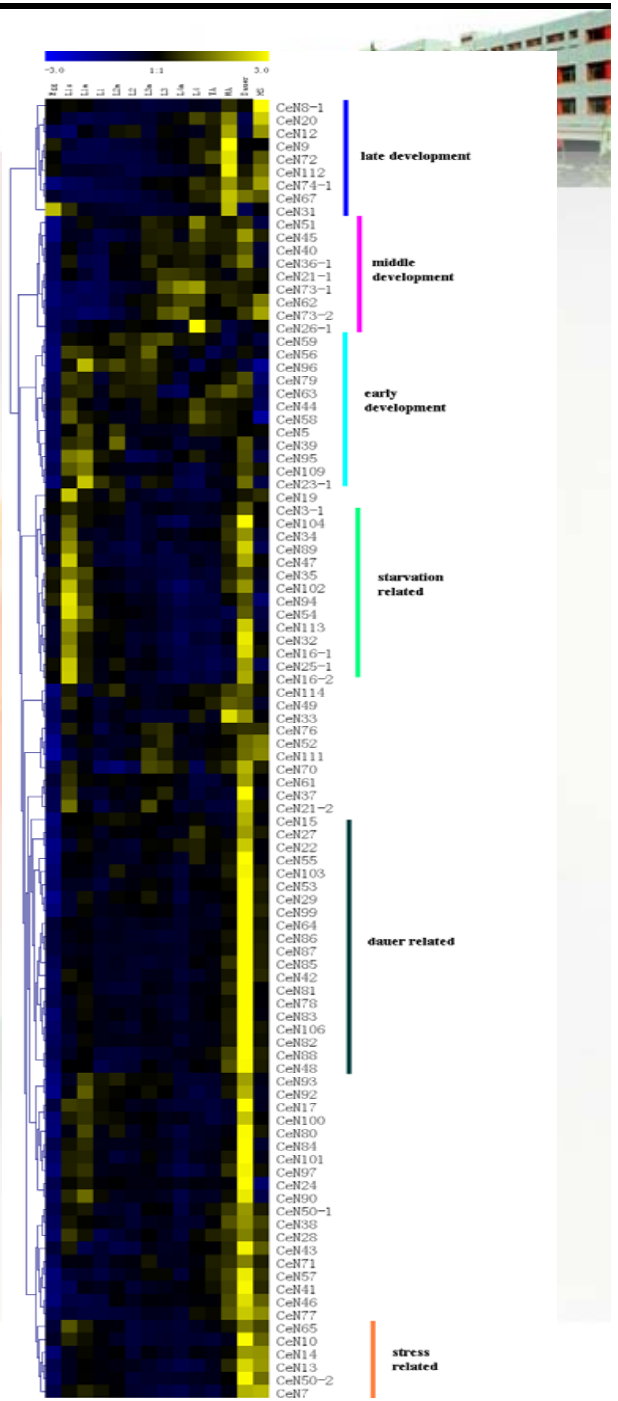
Developmentally regulated ncRNAs







Analysis of transcription levels of 106 ncRNA families were carried out with Northern blot. 61 showed variation exceeding two standard variation, composed of 6 distinct expression clusters.



EurekAlert! AAAS

Public release date: 9-Jan-2006

Contact: Maria Smit
smit@cshl.edu
516-422-4013
Cold Spring Harbor Laboratory

'Pregnant' protein-coding genes carry RNA 'babies'

Scientists characterize large numbers of independently expressed, non-protein-coding RNA genes in the introns of protein-coding genes

BEIJING, China Scientists from the Chinese Academy of Sciences have performed a comprehensive analysis of small, non-protein-coding RNAs in the model nematode, *C. elegans*. They characterize 100 heretofore-undescribed transcripts, including two novel classes; they provide insights into the genomic structure and transcriptional regulation of non-coding RNAs; and they underscore the importance of non-coding RNAs in nematode development. Their work appears this month in the journal *Genome Research*.

中国科学院生物物理研究所
INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

*"The significance of non-protein-coding RNAs as central components of various cellular processes has risen sharply over the recent years," explains Prof. Runsheng Chen, principal investigator on the study. Excluding microRNAs (miRNAs), or small transcripts that have recently received widespread attention and are known to play important roles in transcriptional regulation, small non-coding RNAs (or ncRNAs) in *C. elegans* have not been extensively investigated until now.

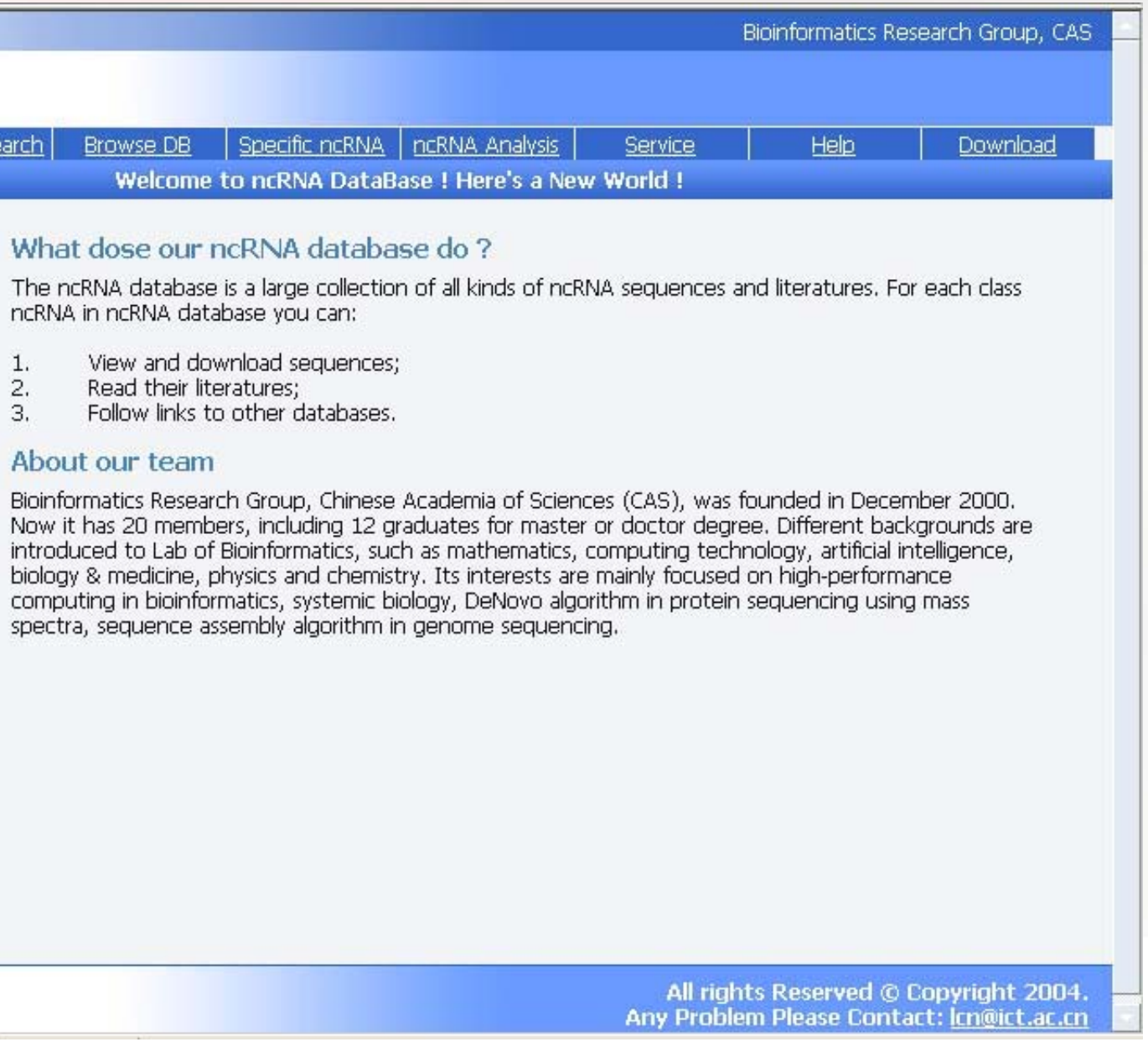
Using a new, high-throughput procedure to clone small, full-length ncRNAs, Chen's laboratory isolated and characterized 161 unique transcripts. A major advantage of the new cloning procedure is that it achieves an extraordinarily high detection rate for ncRNAs by current standards. "Studies published over recent years have only been able to reach a detection rate of about 3%, but our method reached a detection rate of 30% a 10-fold increase in cloning efficiency," explains Chen. "It's like going from a Model T Ford to a Ferrari in one fell swoop!"

Of the 161 transcripts detected by Chen's group, 100 were novel and 61 were previously known or predicted. Among the 100 novel genes, 30 had no known function, whereas 70 belonged to the ubiquitous class of small nucleolar RNAs (snoRNAs). Based on sequence and structural features, Chen and his colleagues were able to classify more than half of the 30 unknown RNAs into two new categories: stem-bulge RNAs (sbRNAs) and small nuclear-like RNAs (snlRNAs). Both classes of transcripts exhibited enhanced expression during the later stages of worm development, indicating a functional role for these transcripts in developmental processes.

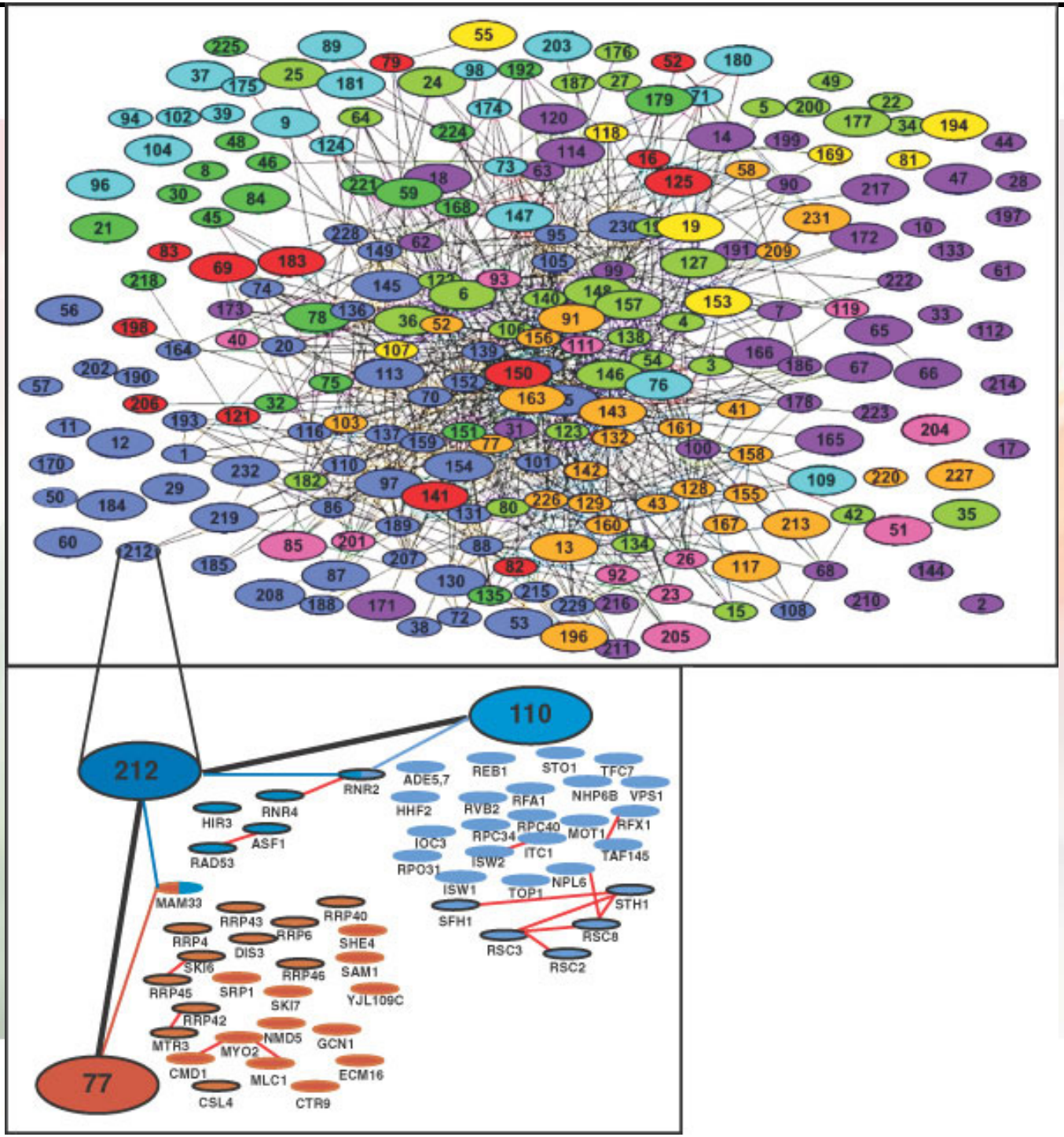
"The interesting thing about nematodes is that their genomic organization of both snoRNAs and other ncRNAs is quite different from other animals," says Chen. In contrast to the genomes of other metazoans, where most snoRNAs are found in introns and are under the control of independent promoters, nematode snoRNA loci are both intergenic and intronic (with and without promoters). Interestingly, plant snoRNAs are primarily located in intergenic regions. Other ncRNA genes (i.e., non-snoRNA genes) are mainly located in intergenic regions in both plants and animals. But in nematodes, Chen's team found that many of these other ncRNA genes are located in the introns of host protein-coding genes and are under the control of independent promoter elements.

Finally, Chen and his colleagues estimated that 2700 ncRNA genes are present in the *C. elegans* genome. "One particularly intriguing aspect of the non-coding transcriptome is its potential to fill the regulatory gap created by the surprisingly low number of protein-coding genes in higher organisms," says Chen. "Between one-celled yeast, thousand-celled nematodes, and trillion-celled mammals, there is a difference of a mere 6,000 to 19,000 to 25,000 in protein-coding gene numbers. We think that regulation by non-coding RNA accounts for this discrepancy and helps to explain the additional biological complexity of higher organisms."

2、The noncoding RNA database—NONCODE

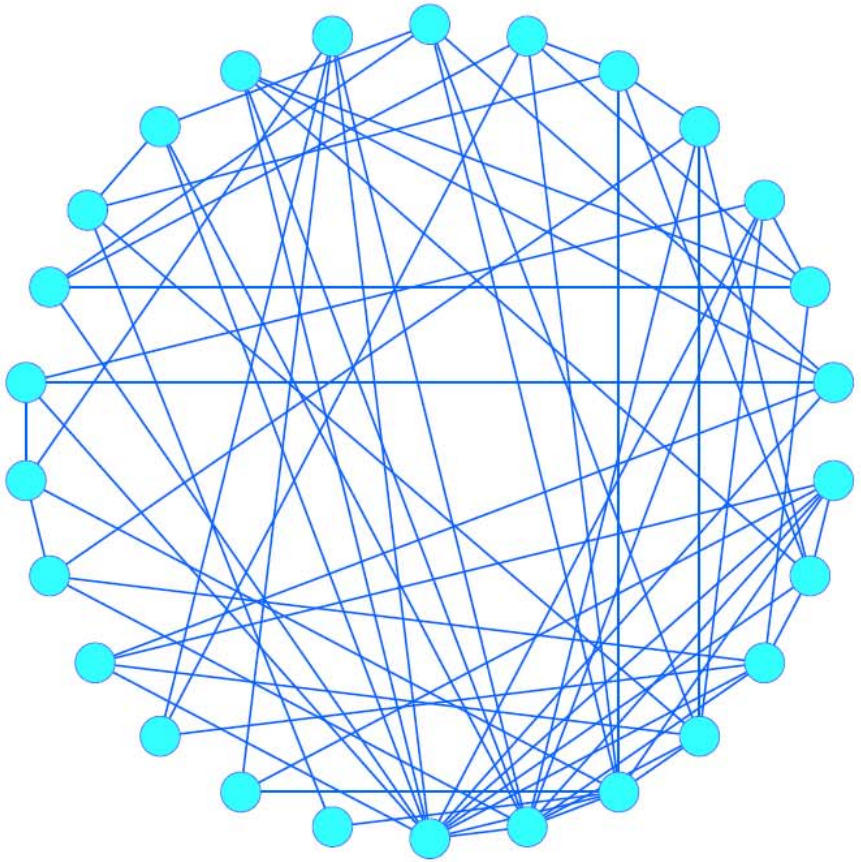


二、Noncoding RNA would be the Elements of Biological Network

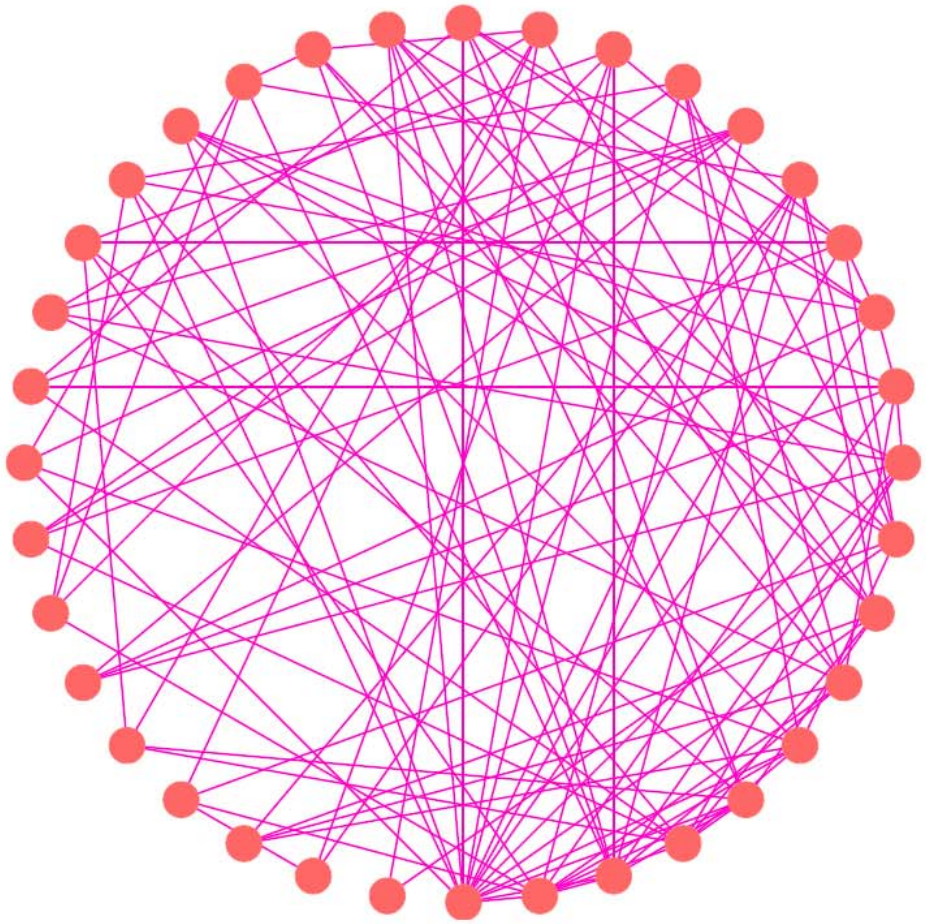


Nature **415**, 141 - 147 (2002)

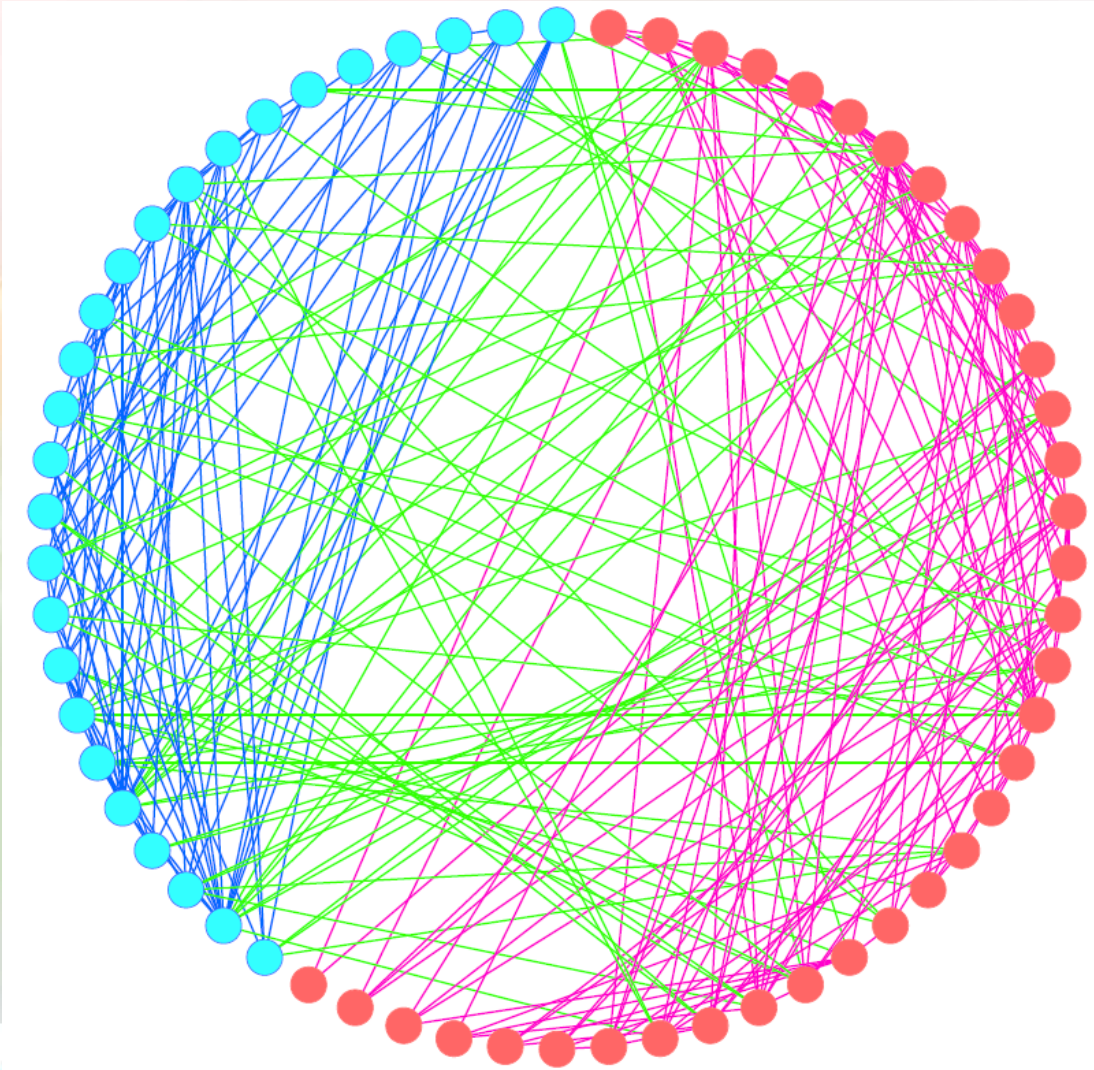
Protein Network



RNA Network



Bicolor (Protein NcRNA) Network



3、 Bi-colored network

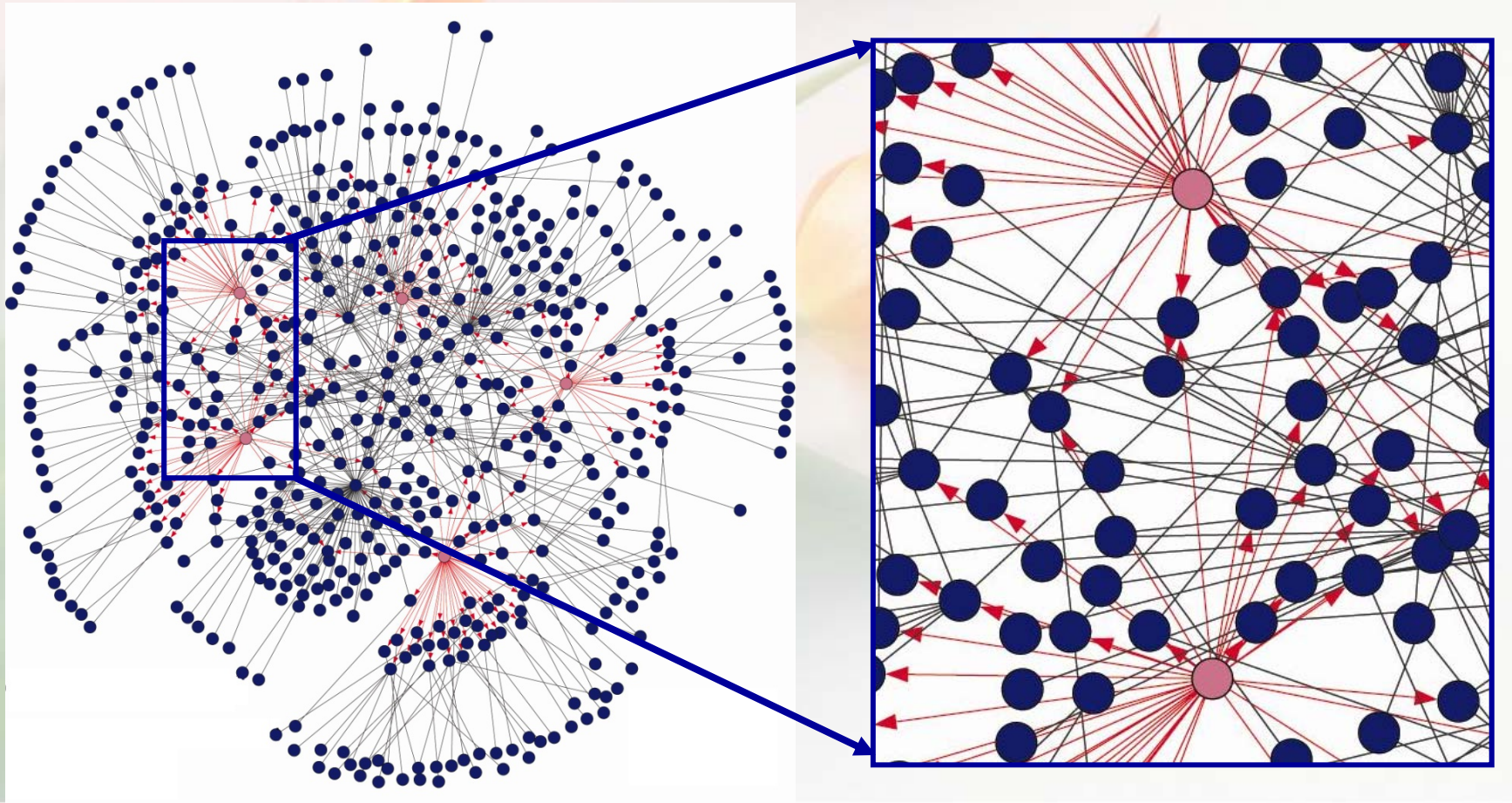
MicroRNAs regulate microRNAs
— a network of mutual microRNA control

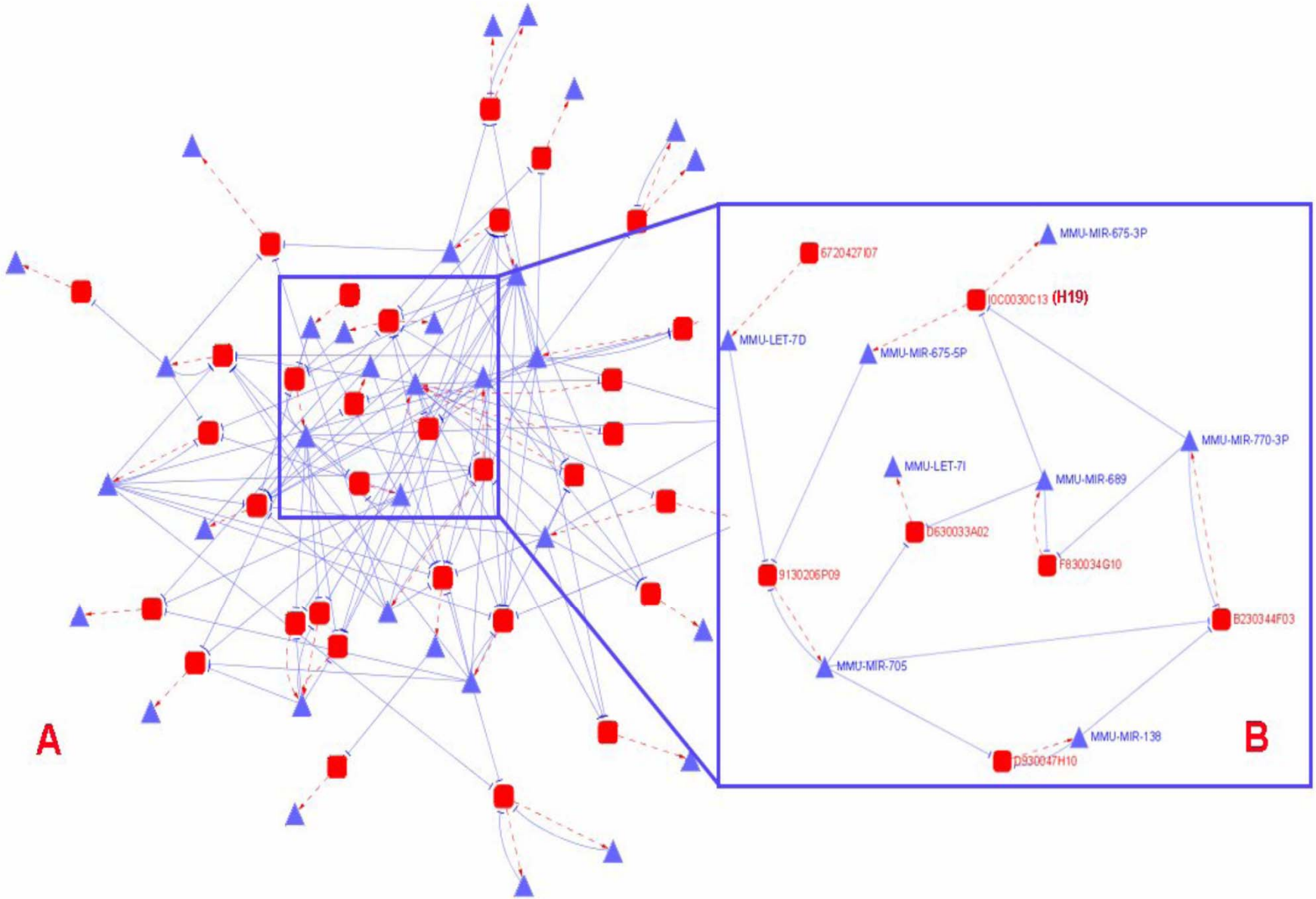
Trans in Genetics 2008, **24**:323

1、Messenger-like ncRNAs show miRNA-related reductions in expression, that means they would be regulated by microRNAs (miRNAs) like mRNA targets.

2、The mRNA-like-ncRNAs serve as vectors or storage forms for short ncRNAs （MicroRNA).

NcRNA-NcRNA Interaction Network

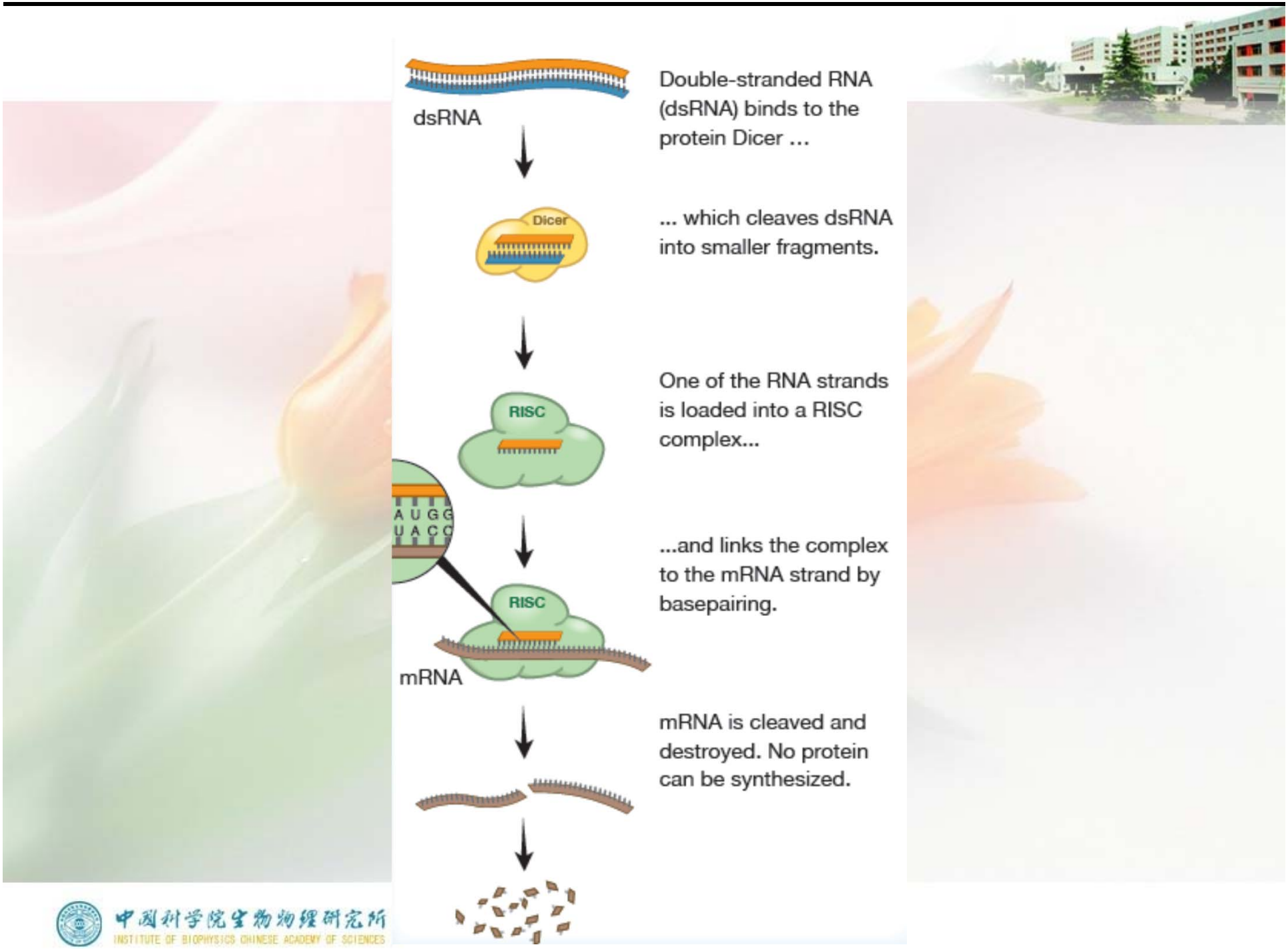


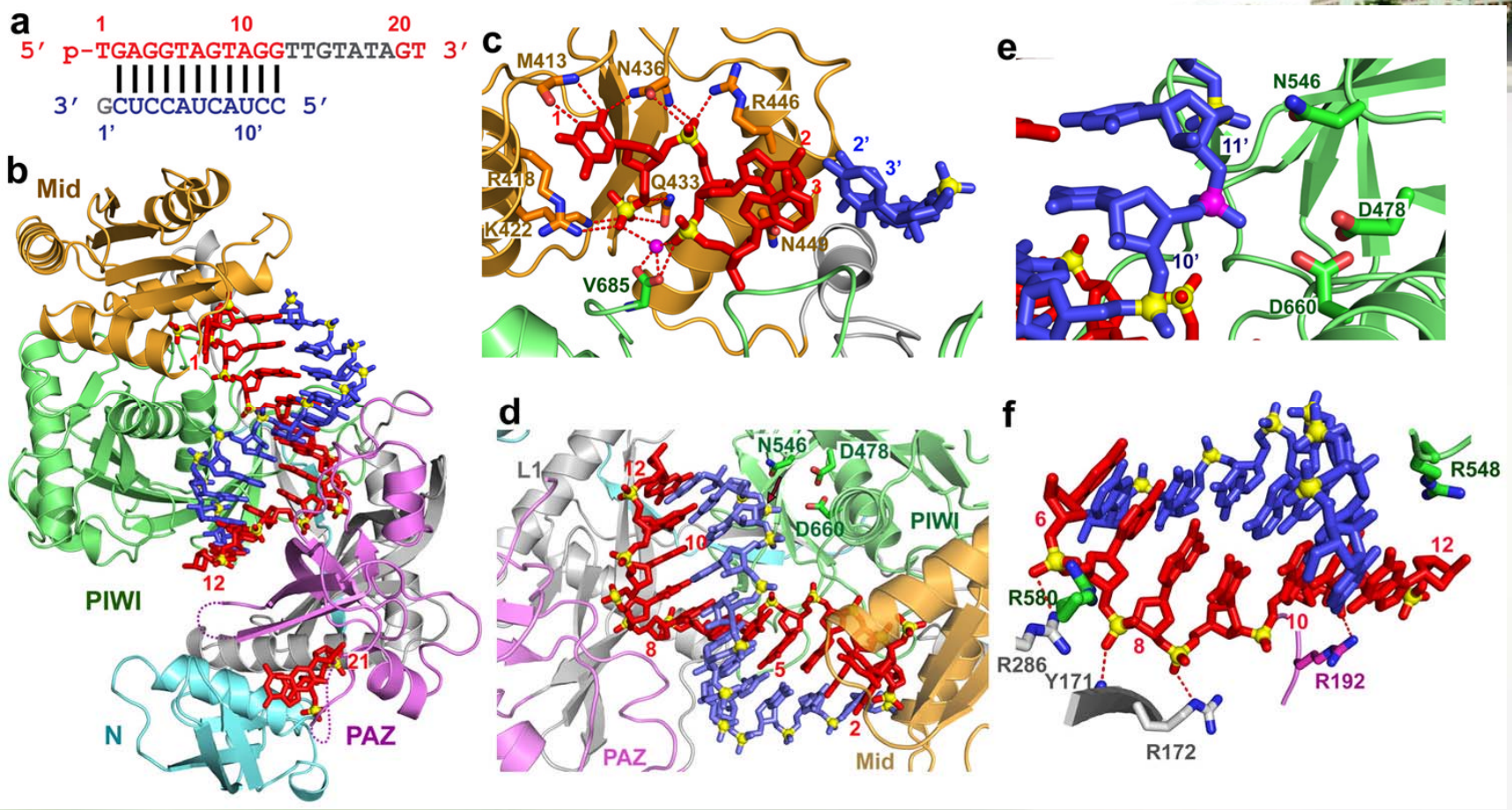

A
B
6720427I07
MMU-MIR-675-3P
0C0030C13 (H19)
MMU-LET-7D
MMU-MIR-675-5P
MMU-MIR-770-3P
MMU-LET-7I
MMU-MIR-689
D630033A02
F830034G10
9130206P09
B230344F03
MMU-MIR-705
MMU-MIR-138
p530047H10

And they hypothesized a regulatory relationship between microRNAs. If this model is true, it will further greatly stimulate the community's interest in the regulatory role of microRNA, and represents a **milestone** in exploring microRNA's functionality.

三、Variety of NcRNA regulation

miRNA-induced transcriptional inhibition


dsRNA
Double-stranded RNA (dsRNA) binds to the protein Dicer ...
Dicer
... which cleaves dsRNA into smaller fragments.
RISC
One of the RNA strands is loaded into a RISC complex...
A U G G
U A C C
...and links the complex to the mRNA strand by basepairing.
RISC
mRNA
mRNA is cleaved and destroyed. No protein can be synthesized.



- **Wang Y**, ……, Patel DJ.
- ***Nature***. 2008, 456:209-213； ***Nature***. 2008, 456:921-926；
- ***Nature.***2009, 461:754-761； ***Nat Struct Mol Biol.*** 2010, 17：781-787.

miRNA-induced transcriptional gene silencing

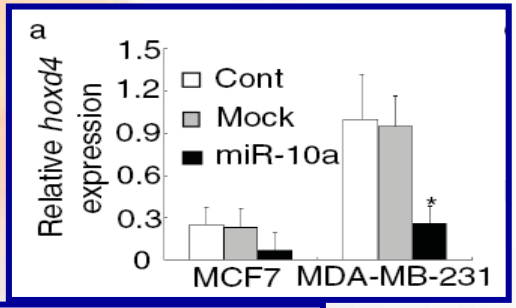
Promotor asscoiated RNA

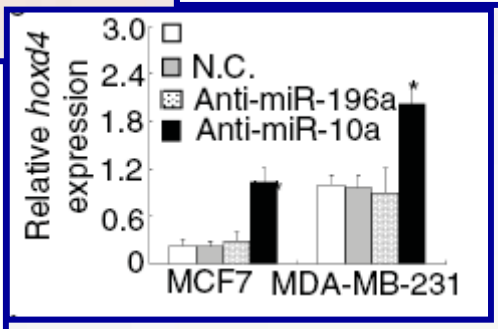
Promoter-associated RNA is required for RNA-directed transcriptional gene silencing in human cells

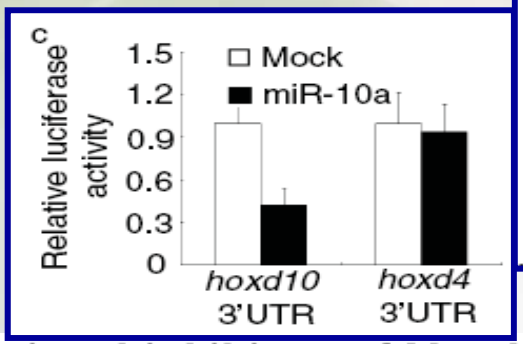
Jiang Han*, Daniel Kim†, and Kevin V. Morris*‡

PNAS 2007

- Transcriptional activity in the *promoter region* of a target gene is required for siRNA-induced transcriptional silencing

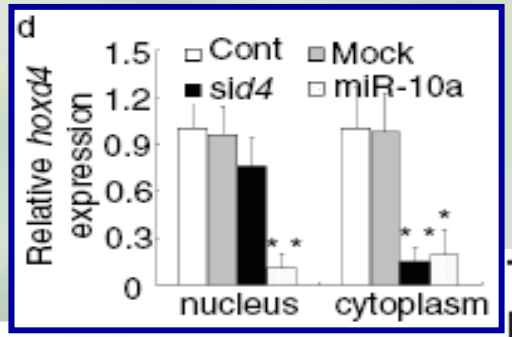






miR-10a inhibits *hoxd4* transcription

- Transfection studies
  - *Hoxd4* expression is inhibited by miR-10a
  - Anti-miR-10a increases *hoxd4* expression
  - Inhibition not mediated by the 3'UTR
- Nuclear run-on assay confirms transcriptional inhibition



Transcriptional inhibiton of Hoxd4 expression by miRNA-10a in human breast cancer cells

BMC Molecular Biology 2009, **10**:12

中国科学院生物物理研究所 INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

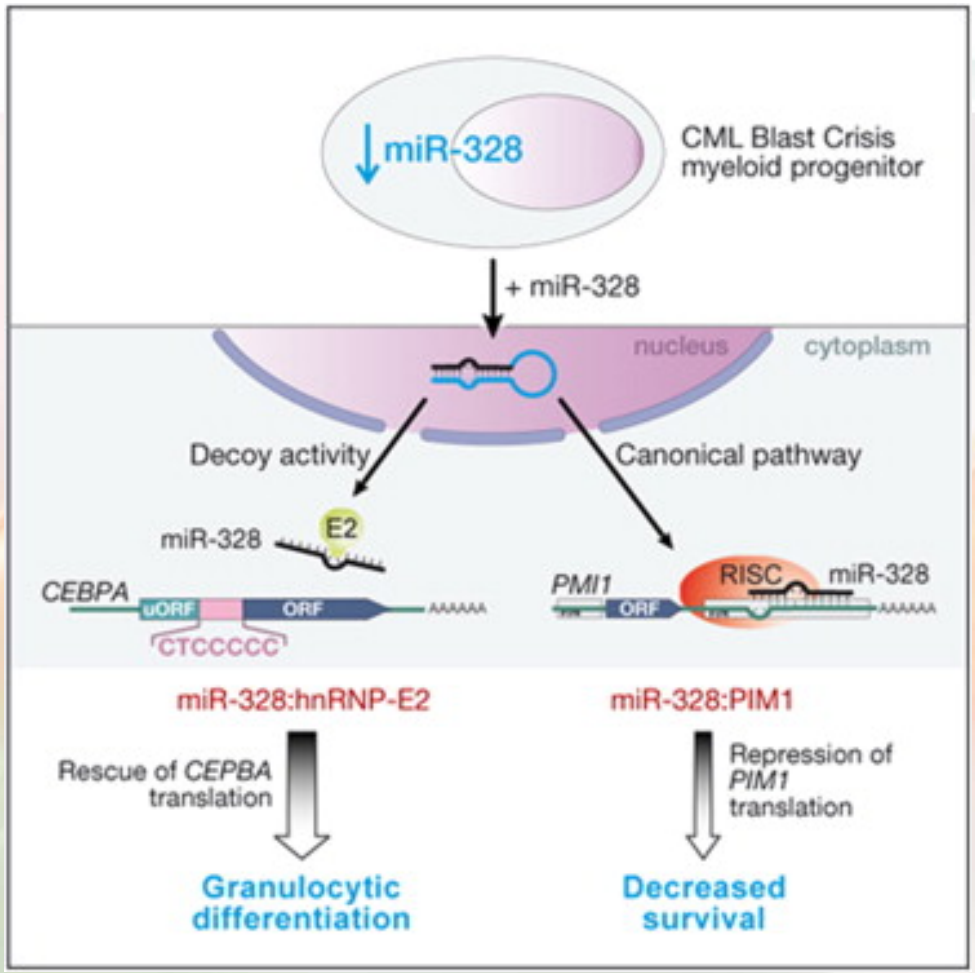
Yuliang Tan[1,2], Bo Zhang[1,2], Tao Wu[1,2], Geir Skogerbø[1], Xiaopeng Zhu[1], Xiangqian Guo[1,2], Shunmin He[1,2] and Runsheng Chen*[1]

Switching from Repression to Activation: MicroRNAs Can Up-Regulate Translation.

Vasudevan S, Tong Y, Steitz J A.

Science, 2007, 318(5858): 1931—1934.

AU-rich elements (AREs) and microRNA target sites are conserved sequences in messenger RNA (mRNA) 3′ untranslated regions (3′UTRs) that control gene expression posttranscriptionally. Upon cell cycle arrest, the ARE in tumor necrosis factor–a (TNFa) mRNA is transformed into a translation activation signal, recruiting Argonaute (AGO) and fragile X mental retardation–related protein 1 (FXR1), factors associated with micro-ribonucleoproteins (microRNPs). We show that human microRNA miR369-3 directs association of these proteins with the AREs to activate translation. Furthermore, we document that two well-studied microRNAs—Let-7 and the synthetic microRNA miRcxcr4—likewise induce translation upregulation of target mRNAs on cell cycle arrest, yet they repress translation in proliferating cells. Thus, activation is a common function of microRNPs on cell cycle arrest. We propose that translation regulation by microRNPs oscillates between repression and activation during the cell cycle.



miR-328 Functions as an RNA Decoy to Modulate hnRNP E2 Regulation of mRNA Translation in Leukemic Blasts.

Cell, Volume 140, 652-665, 5 March 2010

四、NcRNA Mutation can also lead to Disease

Examples of lncRNAs correlated with diseases/disorders（Kannanganattu V. Prasanth and David L. Spector， *Genes Dev. 2007 21: 11-42*）

NcRNAs	Disease/disorder	Reference
<u>NcRNAs with altered expression levels in cancer</u>		
Antisense intronic ncRNAs	Prostate cancer	Reis et al. 2004
BC1	Overexpressed in several cancers	Chen et al. 1997b
BC200	Overexpressed in breast, cervix, esophagus, lung, ovary, parotid, and tongue cancer（乳腺、宫颈、食道、肺、卵巢、腮腺、舌癌）	Chen et al. 1997a; Iacoangeli et al. 2004
BCMS	B-cell neoplasia （B-细胞瘤）	Wolf et al. 2001
CMPD	Campomyelic displasia （短指發育不良）	Ninomiya et al. 1996
DD3	Overexpressed in prostate cancer	Bussemakers et al. 1999
H19	Overexpressed in liver and breast cancer	Looijenga et al. 1997; Lottin et al. 2002
HIS-1	Overexpressed in myeloid leukemia	Askew et al. 1994
HOST2	Expressed in ovarian cancer cells	Rangel et al. 2003
MALAT-1	NSCLC, endometrial sarcoma, and hepatocellular carcinoma （子宫内膜肉瘤、肝细胞癌）	Ji et al. 2003; Lin et al. 2006; Yamada et al. 2006
NC612	Prostate cancer	A.P. Silva et al. 2003
NCRMS	Elevated expression in alveolar rhabdomyosarcoma （横纹肌肉瘤）	Chan et al. 2002
OCC1	Overexpressed in colon carcinoma	Pibouin et al. 2002
PCGEM1	Overexpressed in prostate cancer	Srikantan et al. 2000
PEG8/IGF2AS	Fetal tumors	Okutsu et al. 2000
SRA	Steroid receptor activated RNA isoform expressed in breast cancer	Lanz et al. 1999
TRNG10	Various cancers	Roberts et al. 1998
U50HG	snoRNA host gene; located at the chromosomal breakpoint involved in human B-cell lymphoma	Tanaka et al. 2000
<u>NcRNAs correlated with neurological diseases/disorders</u>		
BC200	Alzheimer’s	Lukiw et al. 1992
DISC2	Schizophrenia and bipolar affective disorder	Millar et al. 2000, 2004; Blackwood et al.2001
IPW	Prader-Willi syndrome	Wevrick et al. 1994
Prion-associated RNAs	Prion pathologies	Deleault et al. 2003; Supattapone 2004
PSZA11q14	Reduced expression in brains of patients with schizophrenia （精神分裂症）	Polesskaya et al. 2003
RAY1/ST7	Autistic disorder	Vincent et al. 2002
SCA8 (KLHL1 antisense)	Spinocerebellar ataxia type 8	Nemes et al. 2000; Mutsuddi et al. 2004
UBE3A-AS	Angelman syndrome	Chamberlain and Brannan 2001
ZNF127AS	Prader-Willi syndrome	Jong et al. 1999

NcRNAs correlated with other diseases/disorders

22k48	HIRA intronic transcript deleted in DiGeorge syndrome	Pizzuti et al. 1999
C6orf37OS	Antisense transcript from C6orf37 locus within diffuse panbronchiolitis critical region	Matsuzaka et al. 2002
COPG2IT1	Russell-Silver syndrome	Yamasaki et al. 2000
DGCR5	Disrupted in DiGeorge syndrome	Sutherland et al. 1996
H19	Beckwith-Wiedemann syndrome	Sparago et al. 2004
LIT1	Beckwith-Wiedemann syndrome	Niemitz et al. 2004
LIT1	Romano-Ward, Jervell and Lange-Nielsen syndromes	Horike et al.2000
MESTIT 1	Russell-Silver syndrome	T. Li et al. 2002; Nakabayashi et al. 2002
PRINS	Psoriasis	Sonkoly et al. 2005

此外：

一种名为HOTAIR的lincRNA在原发乳腺癌和其迁移过程中表达量上升。 HOTAIR表达水平可以作为最终迁移和死亡预测指标。上皮细胞癌中HOTAIR表达上升，导致PRC2重新定位且类似于胚胎成纤维细胞，致使组蛋白H3的27位赖氨酸甲基化、基因表达改变，癌症的侵入和迁移增加。相反，HOTAIR缺失可以抑制癌症的迁移，特别是在有过量PRC2的细胞中。这些发现表明lincRNA在调控癌症表观组时有积极作用，可能作为癌症诊断和治疗的重要靶点。（Chang, Howard Y. Nature Volume 464 April 2010）

Cancer-Associated Long Non-Coding RNA Regulates Pre-mRNA Splicing

ScienceDaily (Sep. 27, 2010) — Researchers report this month that MALAT1, a long non-coding RNA that is implicated in certain cancers, regulates pre-mRNA splicing -- a critical step in the earliest stage of protein production. Their study appears in the journal *Molecular Cell*.

Disease	diagnosis & therapy
Drug	design
Animal & plant character	breeding new variety &

Difficulty of RNA Research

1、Big data

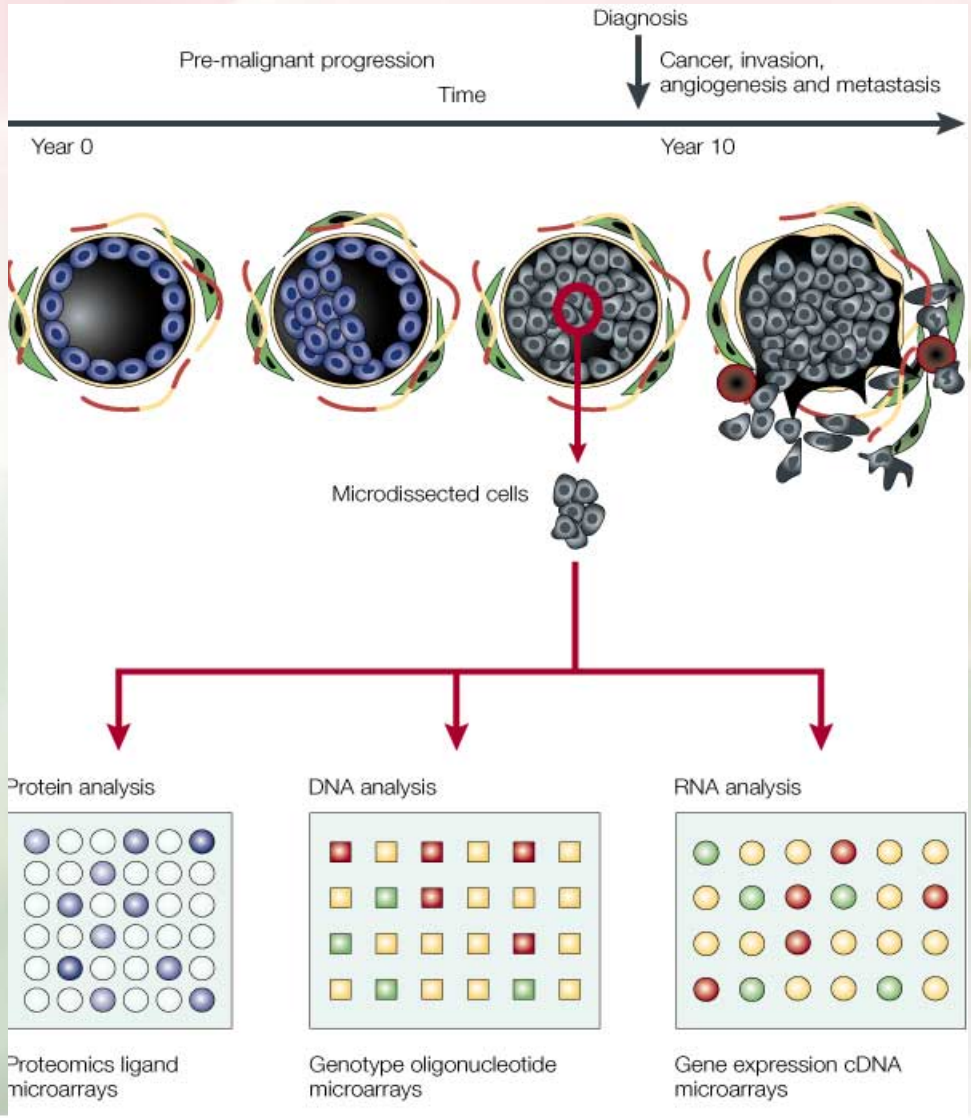
Volume　　Variety
Velocity　　Value

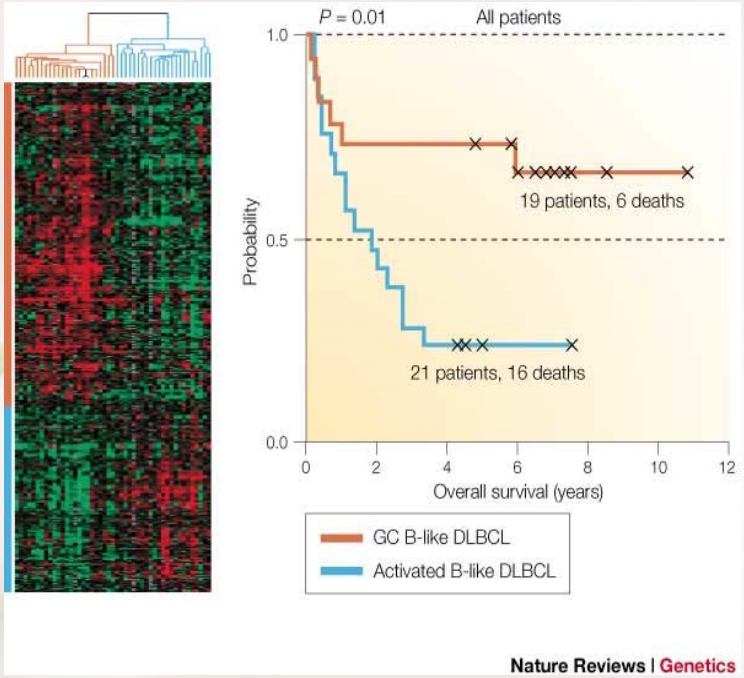
genome
Transcriptome
proteinome
……

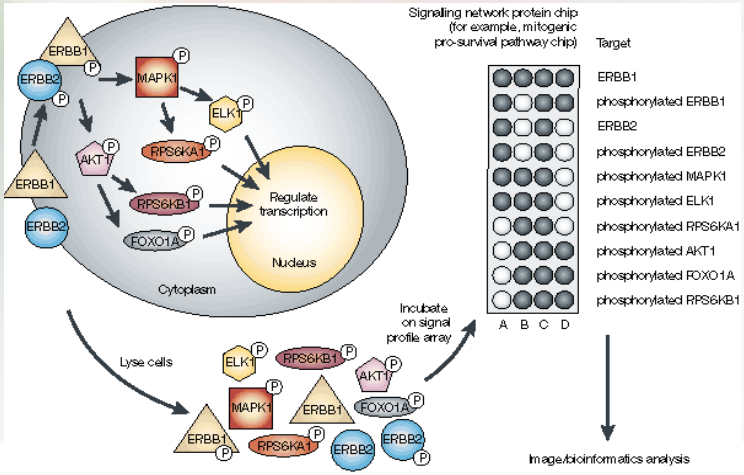
中国科学院生物物理研究所
INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

2、Small samples

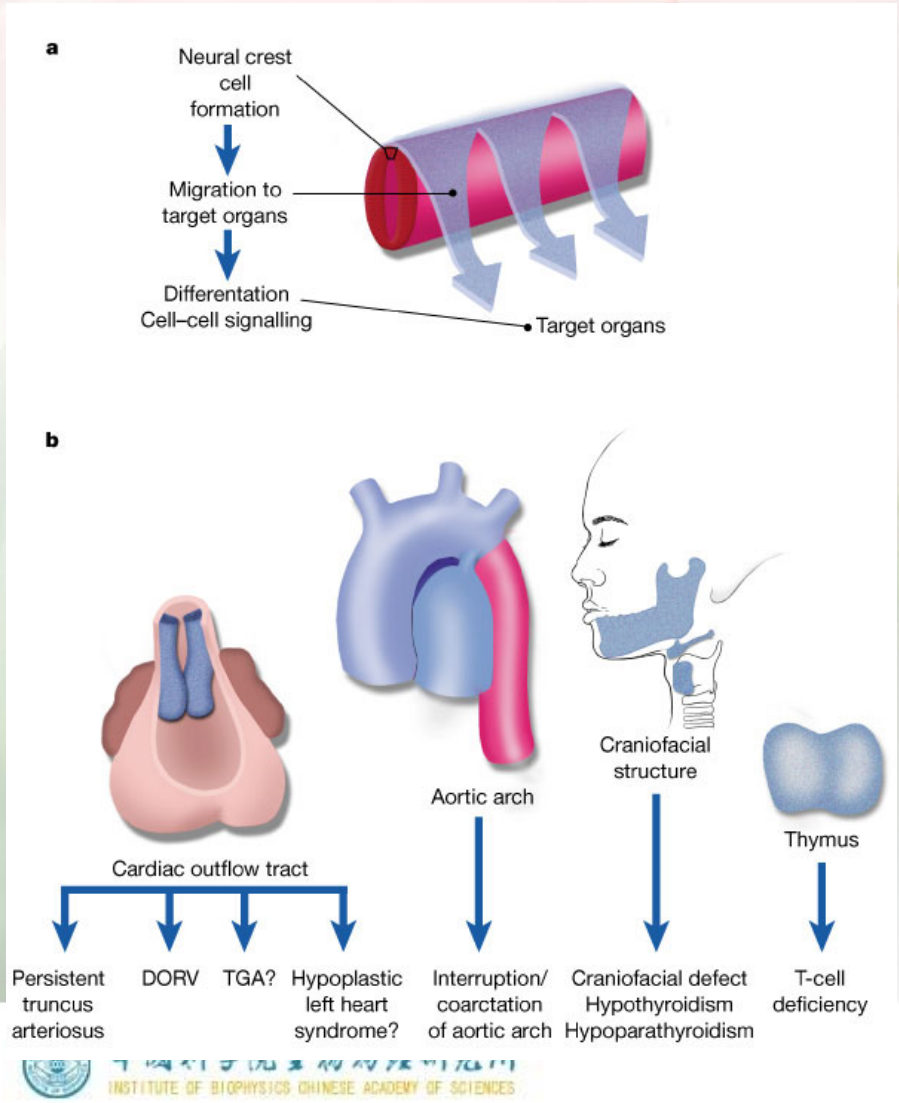
Cancer research

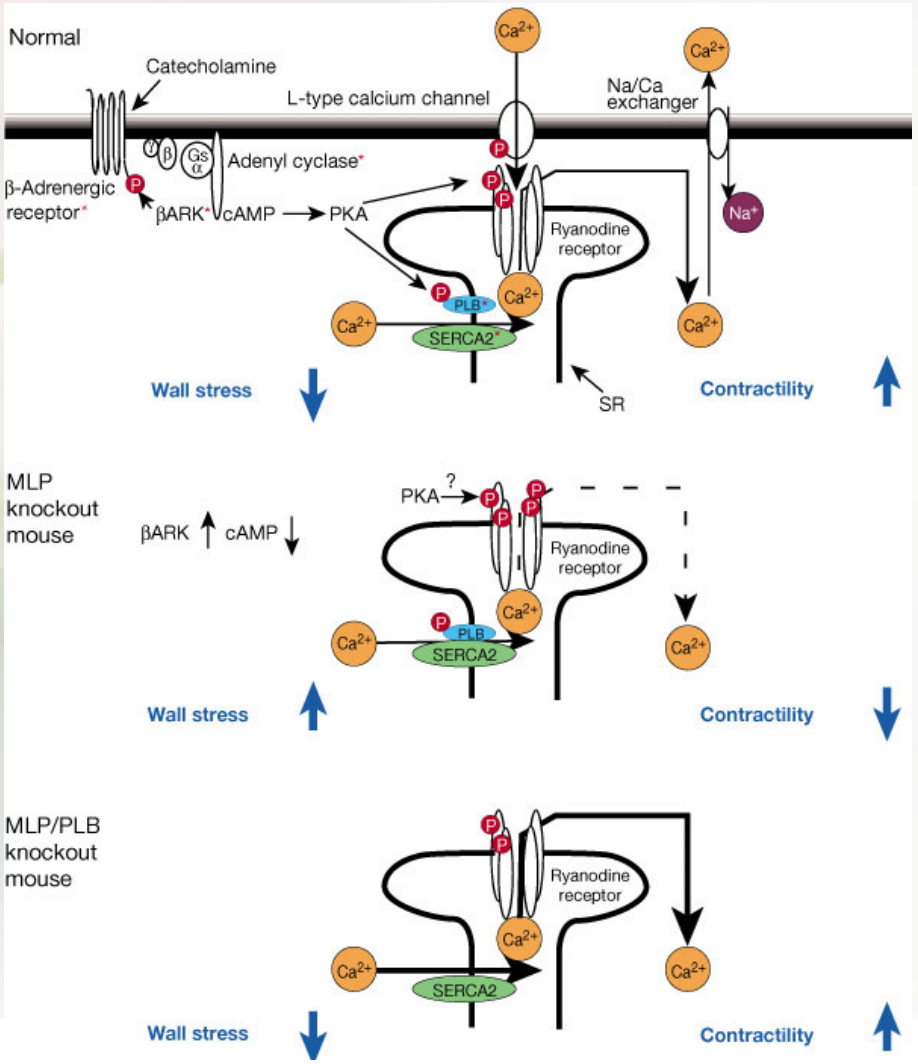






CVD(cardiovascular disease)





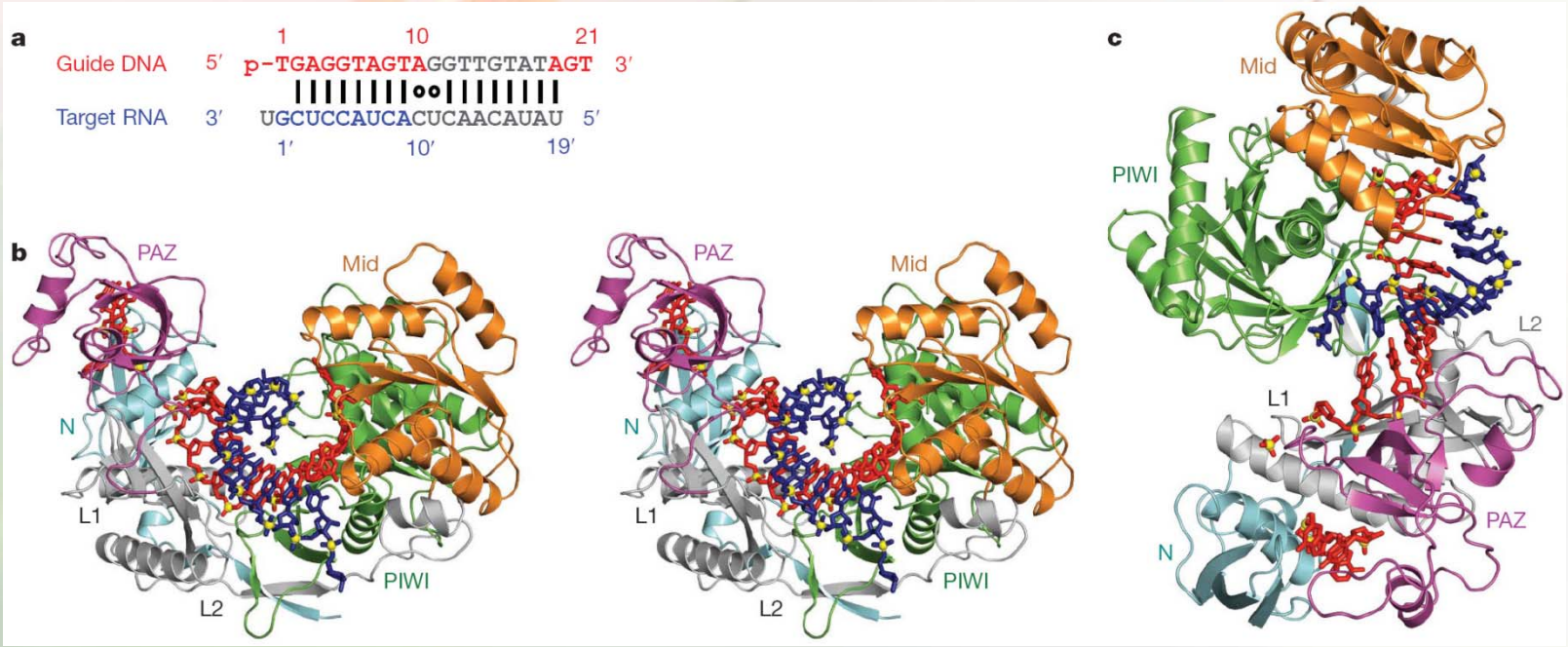
3、Low frequence of effective events

different tissue

period

state

Argonaute 蛋白与不同长度的靶RNA的结构测定

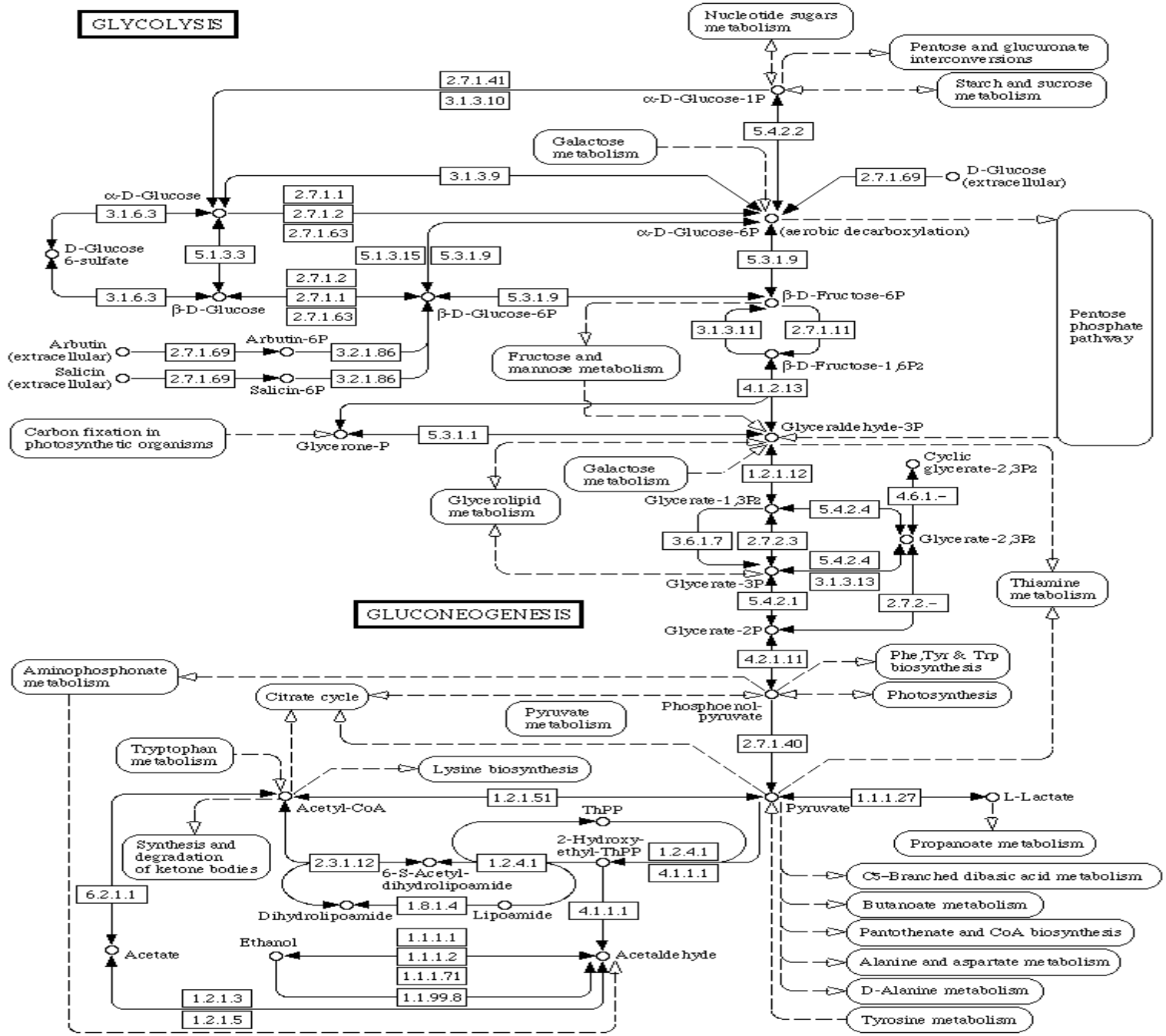


Wang Y, ……, Patel DJ.

Nature. 2008, 456:209-213； ***Nature***. 2008, 456:921-926；

Nature.2009, 461:754-761； ***Nat Struct Mol Biol.*** 2010, 17： 781-787.

4、Complex network


GLYCOLYSIS
Nucleotide sugars metabolism
Pentose and glucuronate interconversions
Starch and sucrose metabolism
2.7.1.41
3.1.3.10
α-D-Glucose-1P
5.4.2.2
Galactose metabolism
3.1.3.9
2.7.1.69
D-Glucose (extracellular)
α-D-Glucose
3.1.6.3
2.7.1.1
2.7.1.2
2.7.1.63
α-D-Glucose-6P
(aerobic decarboxylation)
D-Glucose 6-sulfate
5.1.3.3
5.1.3.15
5.3.1.9
5.3.1.9
3.1.6.3
β-D-Glucose
2.7.1.2
2.7.1.1
2.7.1.63
5.3.1.9
β-D-Glucose-6P
β-D-Fructose-6P
Pentose phosphate pathway
3.1.3.11
2.7.1.11
Arbutin (extracellular)
2.7.1.69
Arbutin-6P
3.2.1.86
Salicin (extracellular)
2.7.1.69
Salicin-6P
3.2.1.86
Fructose and mannose metabolism
β-D-Fructose-1,6P2
4.1.2.13
Carbon fixation in photosynthetic organisms
Glycerone-P
5.3.1.1
Glyceraldehyde-3P
Galactose metabolism
1.2.1.12
Cyclic glycerate-2,3P2
4.6.1.-
Glycerolipid metabolism
Glycerate-1,3P2
5.4.2.4
3.6.1.7
2.7.2.3
Glycerate-2,3P2
5.4.2.4
Glycerate-3P
3.1.3.13
Thiamine metabolism
5.4.2.1
2.7.2.-
GLUCONEOGENESIS
Glycerate-2P
4.2.1.11
Phe, Tyr & Trp biosynthesis
Aminophosphonate metabolism
Citrate cycle
Phosphoenol-pyruvate
Photosynthesis
Pyruvate metabolism
2.7.1.40
Tryptophan metabolism
Lysine biosynthesis
1.2.1.51
Acetyl-CoA
ThPP
Pyruvate
1.1.1.27
L-Lactate
Propanoate metabolism
Synthesis and degradation of ketone bodies
2-Hydroxy-ethyl-ThPP
1.2.4.1
2.3.1.12
6-S-Acetyl-dihydrolipoamide
1.2.4.1
4.1.1.1
C5-Branched dibasic acid metabolism
6.2.1.1
Butanoate metabolism
Dihydrolipoamide
1.8.1.4
Lipoamide
4.1.1.1
Pantothenate and CoA biosynthesis
Acetate
Ethanol
1.1.1.1
1.1.1.2
1.1.1.71
Acetaldehyde
Alanine and aspartate metabolism
D-Alanine metabolism
1.2.1.3
1.2.1.5
1.1.99.8
Tyrosine metabolism
00010 2/5/04

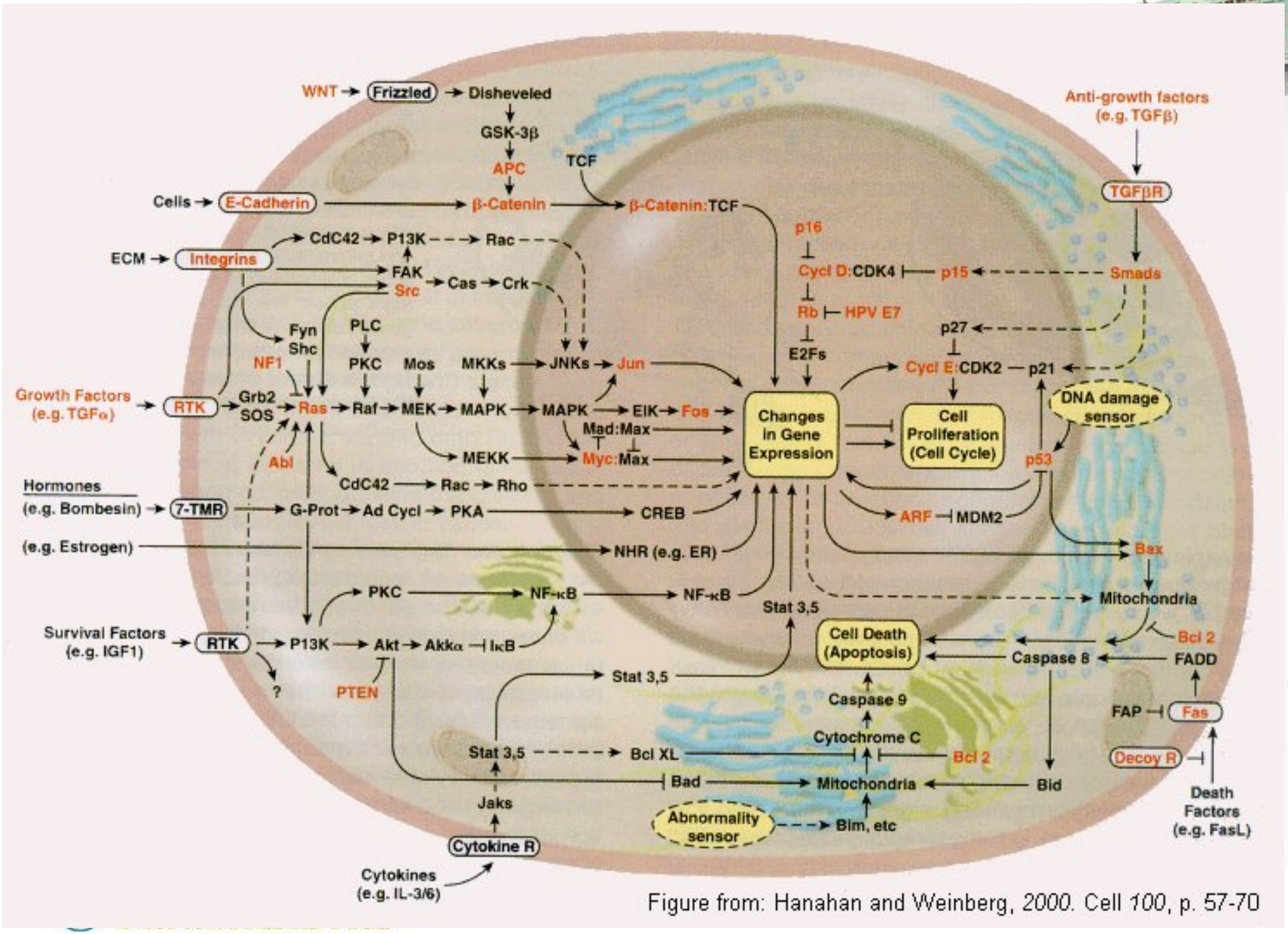
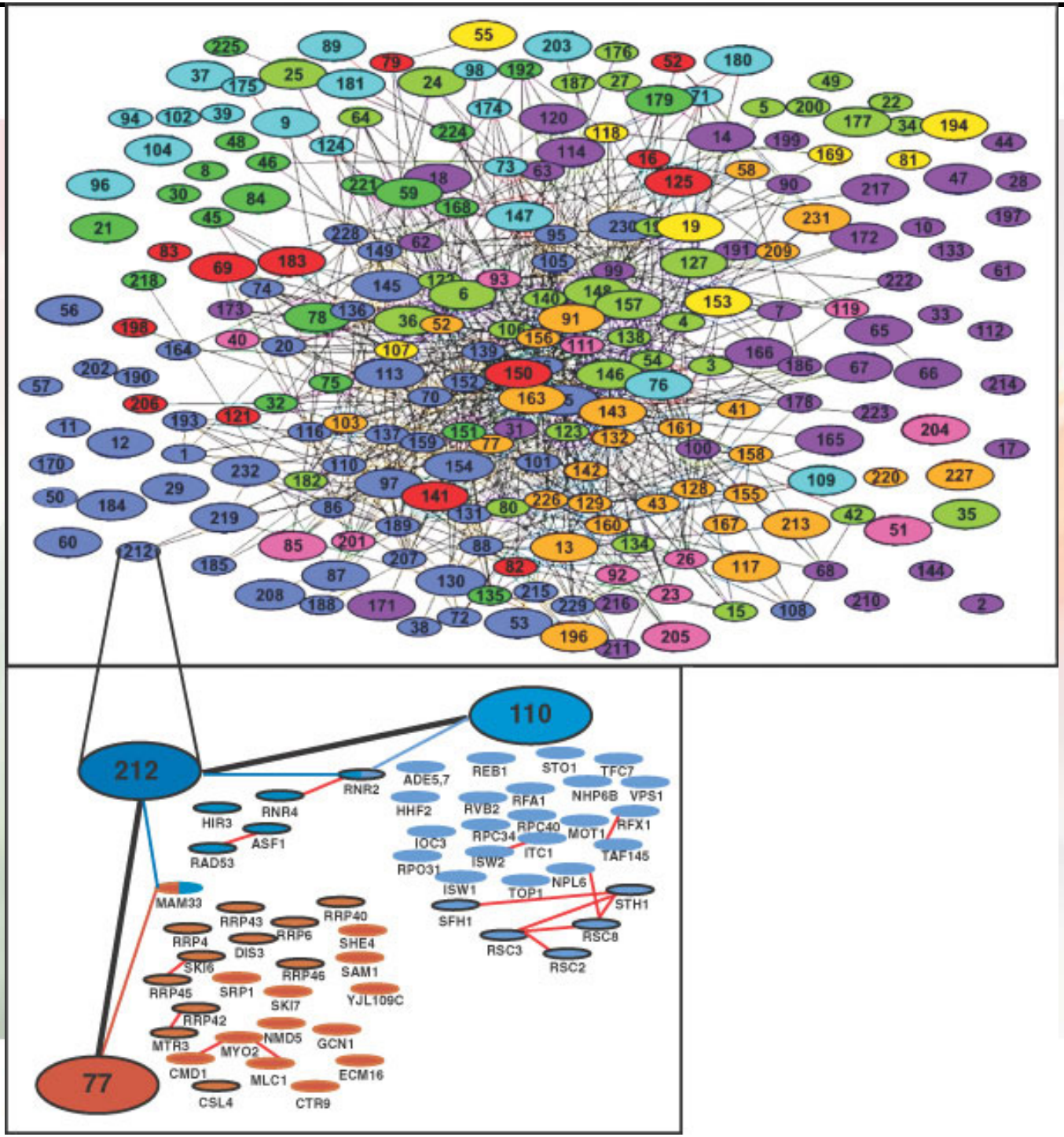
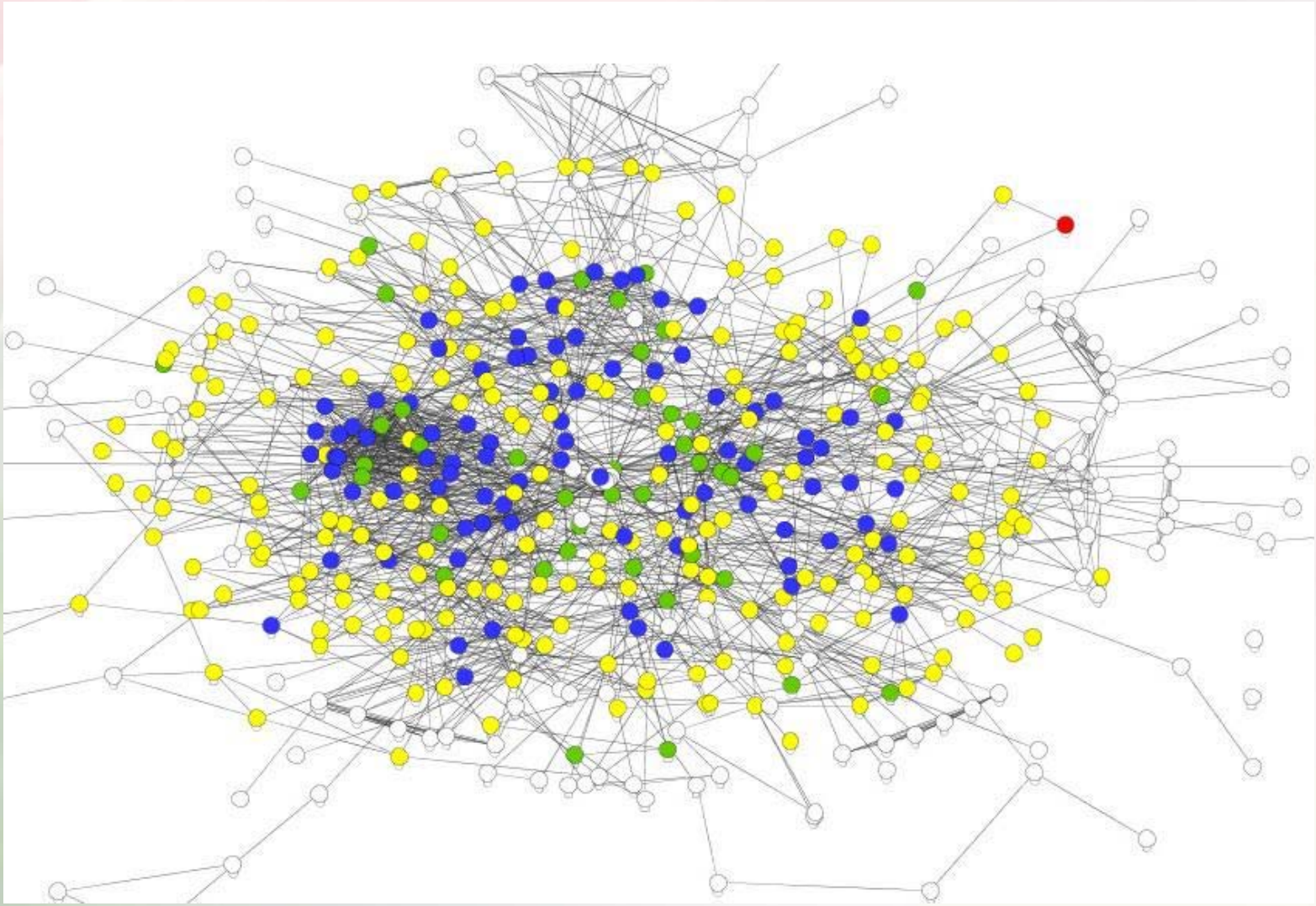


Figure from: Hanahan and Weinberg, *2000*. Cell *100*, p. 57-70



Nature **415**, 141 - 147 (2002)



Dynamic

Directed

Bicorourable

Bioinformatics in noncoding RNA Research

Predict:

- noncoding gene
- regulatory element of gene expression
- target of noncoding RNA
- interaction between noncoding RNA and another macromolecule
- 3D-structure of noncoding RNA

ethods & tools of noncoding RNA Research

Database

NONCODE v4: exploring the world of long noncoding RNA

Weimin Zhu, Yi Zhao, Runsheng Chen

Institute of Basic medical Science, C.A.M.S. ;
Institute of Computational Technology;

Institute of Biophysics, C.A.S, China

中国科学院生物物理研究所
INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

History of NONCODE database

D112–D115 *Nucleic Acids Research, 2005, Vol. 33, Database issue*
doi:10.1093/nar/gki041

NONCODE: an integrated knowledge database of non-coding RNAs

Changning Liu[1,2,3], Baoyan Bai[1,3], Geir Skogerbø[2], Lun Cai[2,3], Wei Deng[1], Yong Zhang[1,3], Dongbo Bu[2], Yi Zhao[2] and Runsheng Chen[1,2,*]

[1]Bioinformatics Laboratory, Institute of Biophysics, Chinese Academy of Sciences, Beijing, China, [2]Bioinformatics Research Group, Key Laboratory of Intelligent Information Processing, Institute of Computing Technology and [3]Graduate School of the Chinese Academy of Science, Beijing 100080, China

Received August 11, 2004; Revised and Accepted September 27, 2004

D170–D172 *Nucleic Acids Research, 2008, Vol. 36, Database issue* *Published online 13 November 2007*
doi:10.1093/nar/gkm1011

NONCODE v2.0: decoding the non-coding

Shunmin He[1,4], Changning Liu[2], Geir Skogerbø[1], Haitao Zhao[3], Jie Wang[1,4], Tao Liu[1], Baoyan Bai[1], Yi Zhao[2] and Runsheng Chen[1,2,*]

[1]Bioinformatics Laboratory and National Laboratory of Biomacromolecules, Institute of Biophysics, [2]Bioinformatics Research Group, Center for Advanced Computing Technology Research, Institute of Computing Technology, [3]Department of Liver Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, CAMS & PUMC and [4]Graduate School of the Chinese Academy of Sciences, Beijing, China

Received September 15, 2007; Revised October 23, 2007; Accepted October 24, 2007

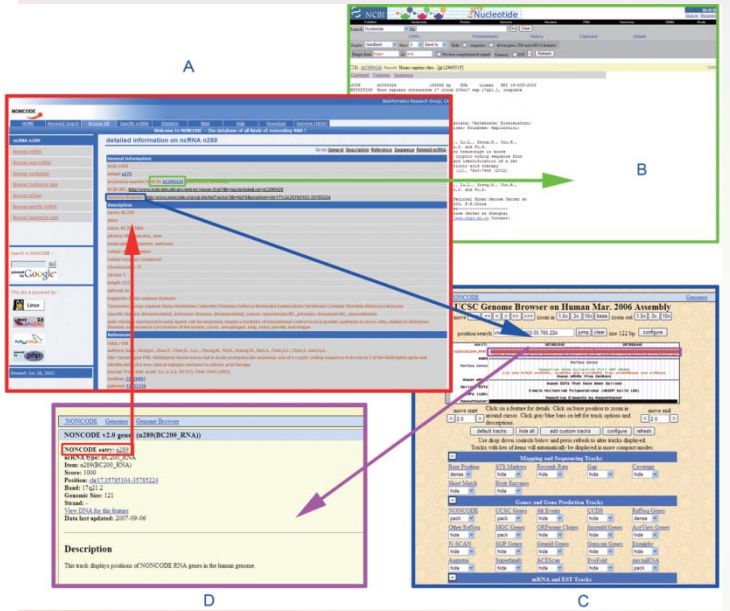
D210–D215 *Nucleic Acids Research, 2012, Vol. 40, Database issue* *Published online 1 December 2011*
doi:10.1093/nar/gkr1175

NONCODE v3.0: integrative annotation of long noncoding RNAs

Dechao Bu[1,2], Kuntao Yu[1,2], Silong Sun[1], Chaoyong Xie[1,2], Geir Skogerbø[3], Ruoyu Miao[1,4], Hui Xiao[1], Qi Liao[1], Haitao Luo[1], Guoguang Zhao[1,2], Haitao Zhao[4], Zhiyong Liu[1], Changning Liu[1], Runsheng Chen[3,*] and Yi Zhao[1,*]

[1]Bioinformatics Research Group, Key Laboratory of Intelligent Information Processing, Advanced Computer Research Center, Institute of Computing Technology, Chinese Academy of Sciences, Beijing, PR China, [2]Graduate School of the Chinese Academy of Sciences, Beijing, PR China, [3]Bioinformatics Laboratory and National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing, PR China and [4]Department of Liver Surgery, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, CAMS & PUMC, Beijing 100730, China

Received September 15, 2011; Revised and Accepted November 13, 2011

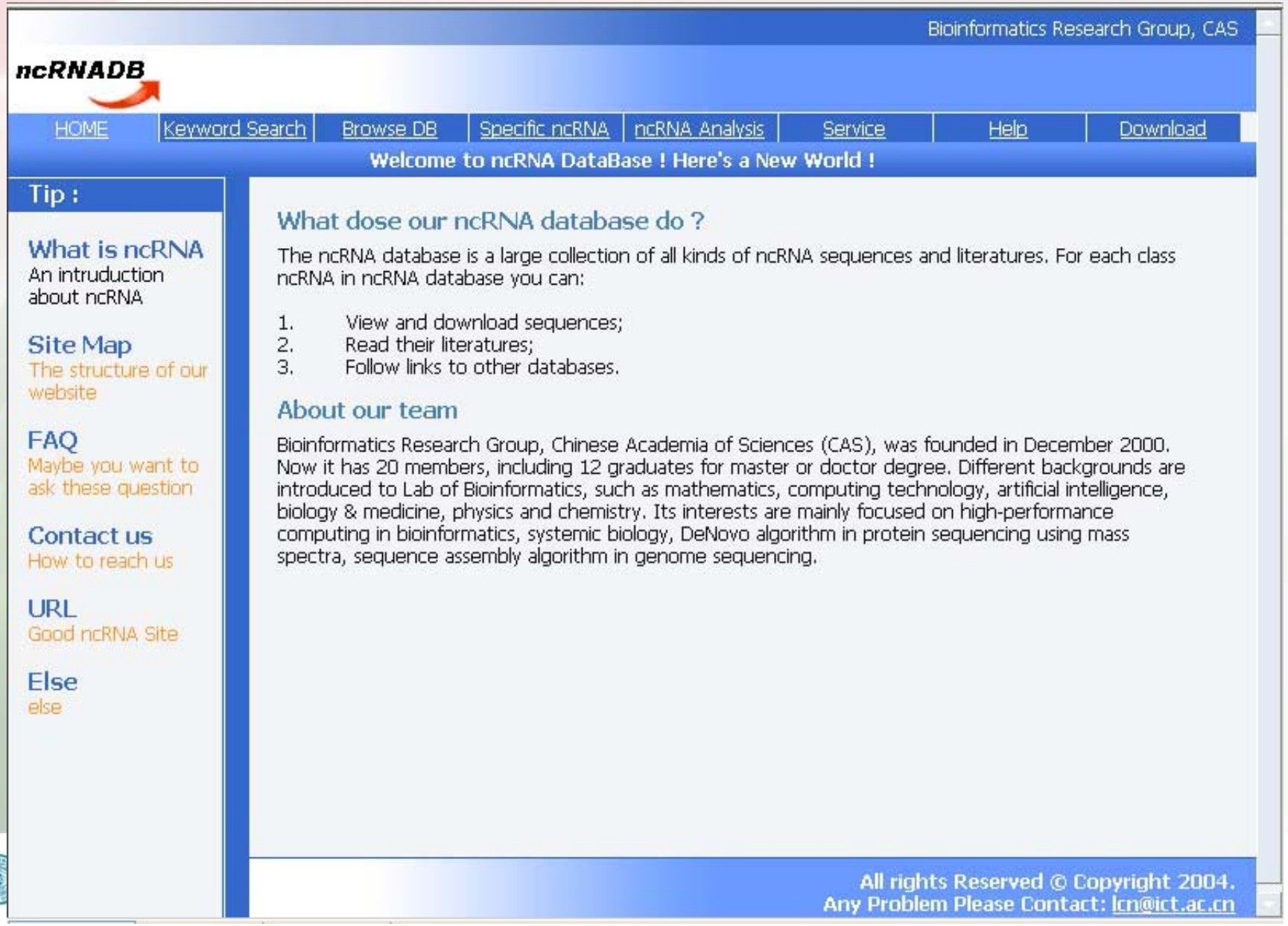


NETWATCH Science

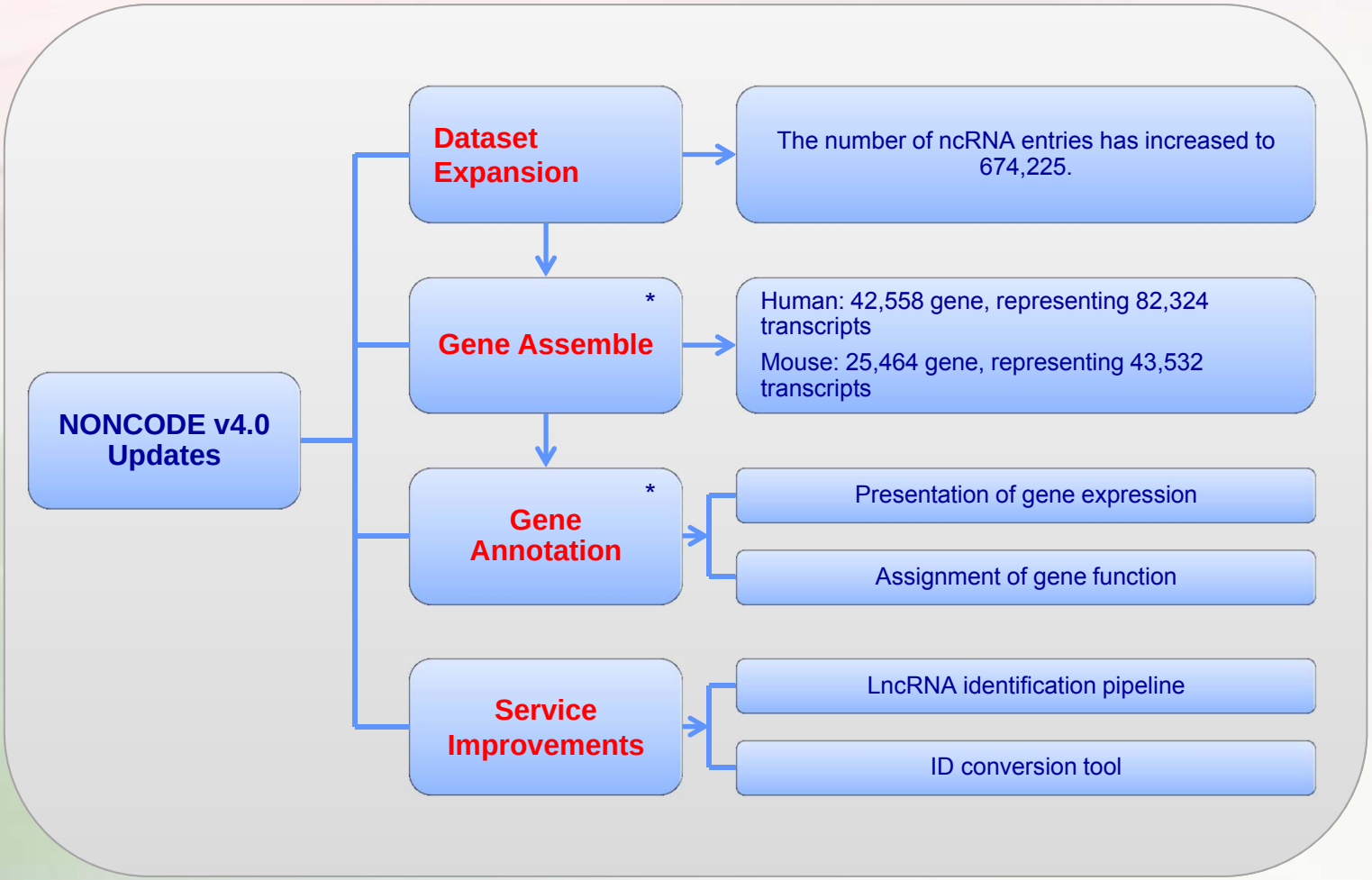
Decoding the NONCODE
Highlight by ***Science***, *Vol 307, Issue 5708, 329 ,21,2005*

中国科学院生物物理研究所
INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

ncRNA 数据库


Bioinformatics Research Group, CAS
ncRNADB
HOME
Keyword Search
Browse DB
Specific ncRNA
ncRNA Analysis
Service
Help
Download
Welcome to ncRNA DataBase ! Here's a New World !
Tip :
What is ncRNA
An intruduction about ncRNA
Site Map
The structure of our website
FAQ
Maybe you want to ask these question
Contact us
How to reach us
URL
Good ncRNA Site
Else
else
What dose our ncRNA database do ?
The ncRNA database is a large collection of all kinds of ncRNA sequences and literatures. For each class ncRNA in ncRNA database you can:
1. View and download sequences;
2. Read their literatures;
3. Follow links to other databases.
About our team
Bioinformatics Research Group, Chinese Academia of Sciences (CAS), was founded in December 2000. Now it has 20 members, including 12 graduates for master or doctor degree. Different backgrounds are introduced to Lab of Bioinformatics, such as mathematics, computing technology, artificial intelligence, biology & medicine, physics and chemistry. Its interests are mainly focused on high-performance computing in bioinformatics, systemic biology, DeNovo algorithm in protein sequencing using mass spectra, sequence assembly algorithm in genome sequencing.
All rights Reserved © Copyright 2004.
Any Problem Please Contact: lcn@ict.ac.cn

An overview of updates in NONCODE v4.0



* represents that this process is specified to lncRNAs in human and mouse.

The number of ncRNAs in 4 version of NONCODE

Version	Publication	Total ncRNA number	lncRNA number	Functional annotation, etc	Services
1.0	2005-NAR	5,339	1,557	Process, function, class	Browse, Search, Download
2.0	2008-NAR	206,226	35,805	Process, function, class	Browse, Search, Download, Blast, Genome Browser
3.0	2012-NAR	411,552	73,370	Process ,function, class Expression profile, predicted functions	Browse, Search, Download, Blast, Genome Browser, Soap API, DAS, On-line Submission
4.0	2014-NAR	523,976	14,744	lncRNA Gene, Expression profile, predicted functions	Browse, Search, Download, Blast, Genome Browser, On-line Submission, iLncRNA, ID conversion.

edited by Mitch Leslie

IMAGES

Sketching Out Past Worlds

For more than 200 years, drawings of fossils and extinct plants and animals have helped paleontologists share their findings with other scientists and the public. A new site from illustrator Mary Parrish of the Smithsonian Institution's National Museum of Natural History in Washington, D.C., explores this corner of paleobiology. An online gallery displays examples, such as this *Triceratops* from the dinosaur collection of late-19th-century paleontologist Othniel Charles Marsh. The site's primer on techniques describes how drawings provide what photos can't: reconstructing a jumble of fossilized bones, putting flesh on a skeleton, or illustrating an ancient landscape. A third section discusses the museum's efforts to preserve its 3500 illustrations, launched in 1995 after staffers discovered a crumbling cache of ink drawings.

www.nmnh.si.edu/paleo/PaleoArt

DATABASES

Decoding the Noncode

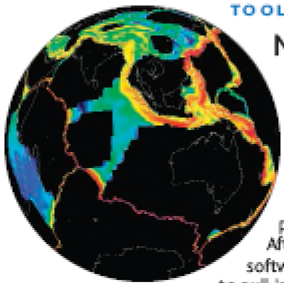
Researchers once paid little attention to RNA that doesn't code for or help manufacture proteins, but they now realize that strands of untranslated RNA perform all kinds of tasks that keep a cell humming. A new database called NONCODE,* from the Chinese Academy of Sciences, documents more than 5000 noncoding RNA sequences from hundreds of organisms. Curators pull sequences from GenBank and other sources, then annotate them by consulting the literature. Categories include disease and function, such as DNA repair or protein transport. NONCODE debuted this month in the annual database issue of open-access *Nucleic Acids Research*,† which lists 719 databases of note on everything from immune system genes to the silkworm genome.

* noncode.bioinfo.org.cn

† nar.oupjournals.org/content/vol33/suppl_1

TOOLS

Map-o-Matic



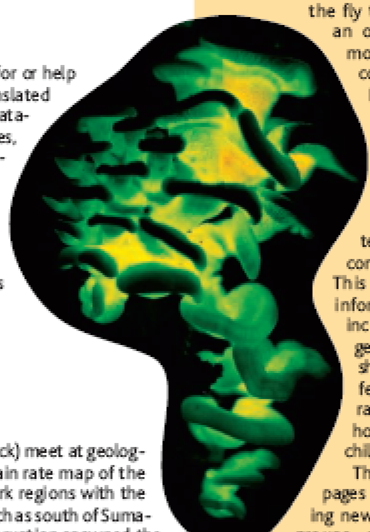
Earth's lithospheric plates (black) meet at geologically active zones in this strain rate map of the world. Red and magenta mark regions with the highest deformation rate, such as south of Sumatra, where a magnitude 9.0 eruption spawned the 26 December tsunami. The image was created with a handy mapping tool from UNAVCO Inc., a nonprofit earth science organization in Boulder, Colorado. After developing the tool 5 years ago for geophysicists, software developer Lou Estey realized it would be a snap to pull in public data sets on the planets, Earth's vegetation, and much more. Users can zoom in, pan out, or download high-resolution maps for printing. A junior version now used by some teachers makes it even easier to create a map of active volcanoes, say, or the world lit up at night. "I've sat down and showed 8-year-olds, and in 5 minutes they're having a blast," says Estey.

jules.unavco.org

RESOURCES

Growth Spurt at Tree of Life

The Tree of Life made a big splash when it debuted in 1994 in the Web's early days. But like many sites, it soon entered a dormant phase. Now the online phylogeny project has gotten new funding and a new educational mission and is seeking more contributors.



The revamped site retains the core of the original tree—now some 3000 pages on beetles, cephalopods, fish, flatworms, and other organisms—but it's now database-driven. That allows visitors to create custom pages on the fly that include, say, an online glossary or more images, notes co-creator David Maddison of the University of Arizona in Tucson. And the tree now invites visitors of all stripes to contribute material linked to the core scientific pages. This supplemental information might include a fruit fly geneticist's data, shots from a professional photographer, or "treehouses" created by children.

The tree's species pages have been sprouting new shoots, too, on groups such as angiosperms and fungi (above, a bioluminescent mushroom, *Panellus stypticus*). Other sections—such as those on mammals and birds—are still mostly blank. But with revisions to the site's architecture and tools now complete, says managing editor Katja Schulz, "this is the year we hope the content takes off."

tolweb.org

Send site suggestions to netwatch@aaas.org. Archive: www.sciencemag.org/netwatch

NPInter: a database of ncRNA interactions

http://www.bioinfo.org/NPInter/

中国科学院生物物理研究所
INSTITUTE OF BIOPHYSICS CHINESE ACADEMY OF SCIENCES

Application of NPInter v1.0

NPInter: the noncoding RNAs and protein related ...
nar.oxfordjournals.org/content/34/suppl_1/D150.full ▾ 翻译此页
作者：T Wu - 2006 - 被引用次数：38 - 相关文章
The data in NPInter are novel, in the sense that no earlier database has ... Therefore, NPInter sets up a bridge between the coding and the noncoding realms.

Predicting protein associations with long noncoding RNAs

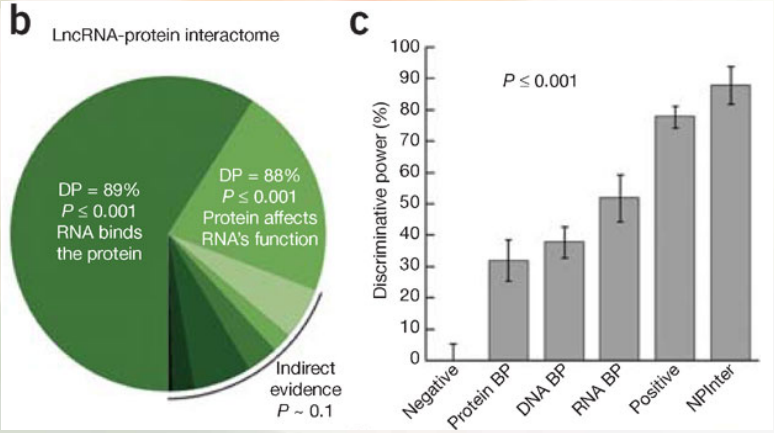
Matteo Bellucci, Federico Agostini, Marianela Masin & Gian Gaetano Tartaglia

Affiliations | **Corresponding author**

Nature Methods **8**, 444–445 (2011) | doi:10.1038/nmeth.1611
Published online 27 May 2011

catRAPID

NPInter Class	# Interactions
The ncRNA binds the protein	239
The protein as a factor affects the ncRNA's function	88
The ncRNA is regulated by the protein	22
Special linkages between the ncRNA and the Protein	8
Genetic interaction between the ncRNA gene and the protein	13
The ncRNA regulates the mRNA	24
The ncRNA indirectly regulates a gene (DNA)	9
The ncRNA as a factor affects the protein's function	2



Performance (dp) on NPInter dataset. The long noncoding database contains 405 interactions from 6 model organisms. Only for a subset of the NPInter database direct physical evidence for protein-RNA interactions is reported.

Performance (dp) on training and indicated sets. Positive and negative sets are lists of interacting and non-interacting protein-RNA pairs. Protein-binding, DNA-binding and RNA-binding proteins were obtained from the Uniprot database.

Data Collection for NPInter v2.0

Specialized databases

Literature mining

RNA binding sites of individual proteins identified using genome-wide techniques were screened against NONCODE. Binding sites correspond to NONCODE RNAs were extracted.

Extracted ncRNA interactions

Molecules in all interactions were annotated and assigned identifier codes from UniGene, NONCODE, miRBase, RefSeq, or UniProt.

Redundancy elimination according to General Information of interactions.

NPInter v2.0

Annotations of NPInter

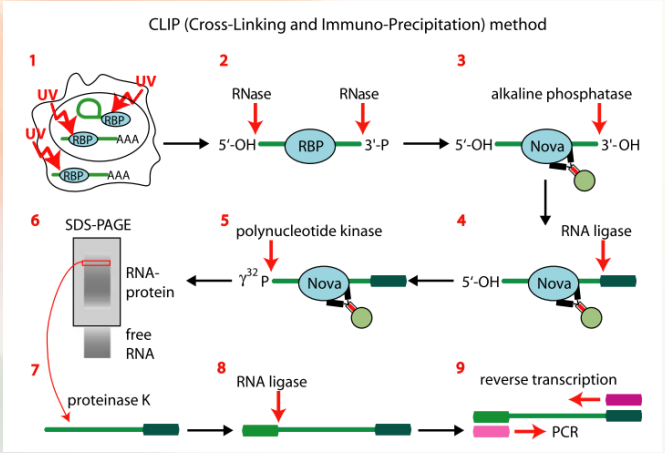
1. General Information
   Interaction ID, names of interacting molecules, level and class of interaction, organism, experimental information and a description of the interaction.
2. Molecule Information
   Molecular entity (based on NONCODE, miRBase and UniProt).
3. Reference
   Publication information.

Specialized databases:

LncRNADisease

Literature:

1. Interaction studies on individual ncRNAs
2. Binding sites generated by high-throughput sequencing



Detecting chips

Expression profiling array for noncoding RNA

◆Data collection of long noncoding RNA

lncRNA source	V1.0	V2.0	V3.0
GENCODE/ENSEMBL	-	12754	22444
Human LincRNA Catalog	-	8195	14353
RefSeq	4765	4765	4814
UCSC gene	13521	13521	5596
NRED	1289	1289	13701
H-InvDB	17203	17203	1038
Enhancer-like lncRNA	2975	2975	3019
RNAdb	-	-	1599
Antisense ncRNA pipeline	1053	1053	1053
ncRNAs from IBP	848	848	848
...	...	...	...
total	42283	63639	70856
After reduced redundancy	30,622	35,024	37,491

Flowsheet

GENCODE、RefSeq、UCSC gene、Human lincRNA catalog、lncRNAdb等13个主流lncRNA数据库及相关文献报道的序列。

陈润生院士实验室发现的未发表的848条lncRNA。

总计73,370条序列

序列ID交叉注释&基因组定位

序列相似性对比

37,581条lncRNA特异探针

Agilent Human 8x60K V2

34,235条mRNA特异探针

晶芯® lncRNA3.0芯片

Updata

v1.0
30,756 lncRNA

v2.0
35,024 lncRNA
32,205 mRNA

v3.0
37,491 lncRNA
34,235 mRNA

May 2011

March 2012

August 2013

Algorithm

Spectrum Analysis algorithm

The physical interactions are transformed into a graph, where each node represents a protein and each edge an interaction between two proteins. We apply the graph theory to analyze the complex protein-protein interaction network. A bi-directed graph G(V,E):

vertex set $V = \{P_1, P_2 \ldots P_n\}$

edge set $E = \{(P_i, P_j) \mid$ *there exists interaction between protein* P_i *and* P_i.

The symmetric $n \times n$ adjacent matrix is defined as $A = (a_{ij})$, where $a_{ij} = 1$, if $(P_i, P_j) \in E$, and $a_{ij} = 0$, if $(P_i, P_j) \notin E$.

Supposed there exists such a property X. A real number Xi would be used to measure the ith protein's intensity of property X, and Xi would be larger if the protein is intensive. If a protein has an intensive property, the neighbors interacting with it would also have the property through the interaction.

$\Delta X_i = c \times \sum_{j=1}^{n} a_{ij} \times X_j$, where c is a positive constant.

we could not compute out the real number Xi directly for its cyclic definition. The iteration method has been used to deal with such a cycle. It's interesting that Xi would converge to a fixed point, which is only concerned with the interaction network and independent of the original assignment.

Procedure

GetScore for ($A_{n\times n}$);

Initialize the score vector $X^{(0)}$ ($X=\{X_1, X_2 ... X_n\}$) with random assignment;

Repeat /* Iteration procedure */

for i = 1 to n do

$$X_j^{(k+1)} = \sum_{j=1}^{n} a_{ij} * X_j^{(k)}$$

$$X^{(k+1)} = \frac{X^{(k+1)}}{\left\| X^{(k+1)} \right\|} \text{ ; k++;}$$

Until $X^{(k)}$ is stable;

We could prove that the X would converge to a fixed point, which is one eigenvetor of the matrix. For matrix $A_{n\times n}$ is symmetrix, all of its eigenvectors constructs a spectrum.

Theorem 1. With arbitrary initial assignment to $X^{(0)}$, the iteration procedure would converge to an engienvector of the symmetric matrix A.

Proof: Assume matrix A's rank is r, and its eigenvectors are given by $\xi_1, \xi_2 \ldots\ldots \xi_r$ with corresponding eigenvalue $\lambda_1 \geq \lambda_2 \geq \ldots\ldots \geq \lambda_r$. For A is symmetric, so all the eigenvectors all mutually orthogonal and form a basis in r-dimensional space. So we have

$$X^{(0)} = c_1 * \xi_1 + c_2 * \xi_2 + \ldots\ldots + c_r * \xi_r$$

After a pass of iteration, we would get:

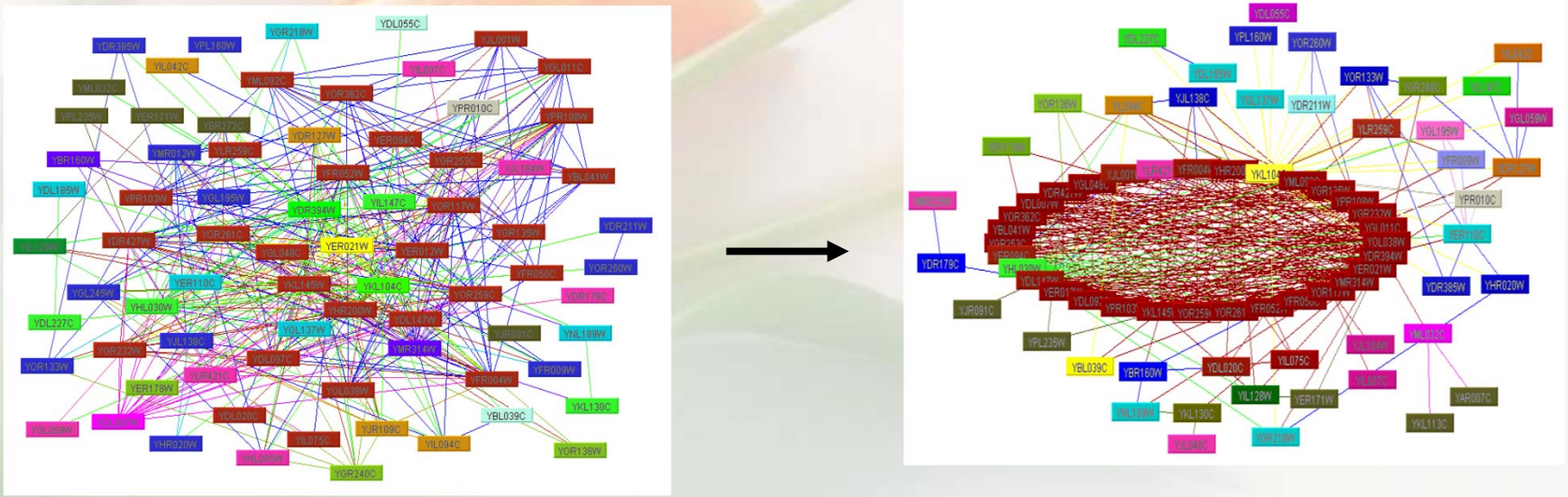
$$W^{(k+1)} = (A_{n\times n})^T \times W^{(k)}$$

$$= A*(c_1 * \chi_1 * \eta_1 + c_2 * \chi_2 * \eta_2 + ...c_m * \chi_m * \eta_m)$$

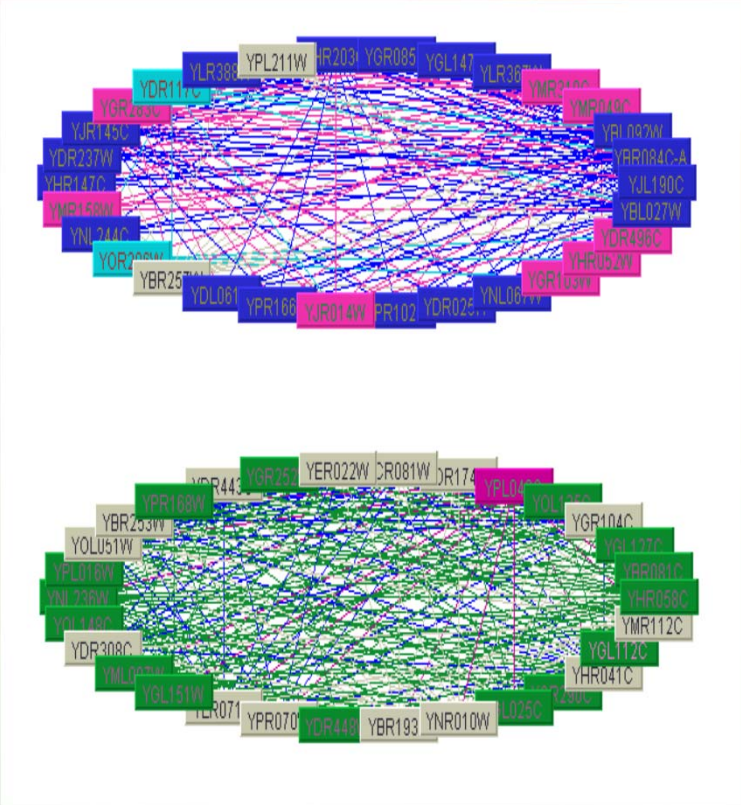
$$= c_1 * \chi_1{}^2 * \eta_1 + c_2 * \chi_2{}^2 * \eta_2 + ...c_m * \chi_m{}^2 * \eta_m$$

As $\left(\frac{\lambda_2}{\lambda_1}\right)<1, \left(\frac{\lambda_3}{\lambda_1}\right)<1 ..... \left(\frac{\lambda_m}{\lambda_1}\right)<1$ we would get $W^{(k)} \approx c_1 * \chi_1{}^k * \xi_1{}^k$ when k tends to infinite large. So X would converge to a fixed point independent of the original assignment.

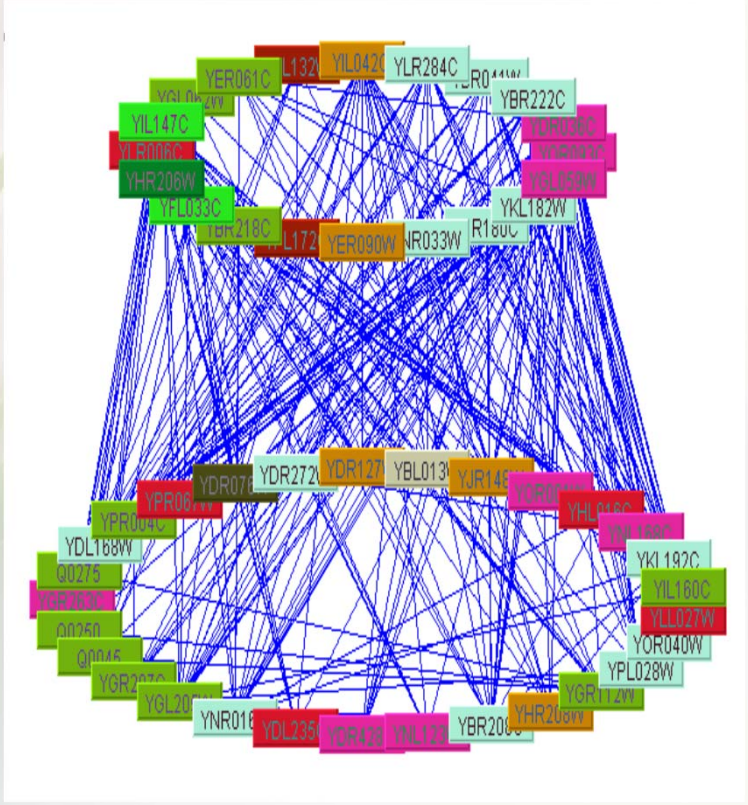
The protein-protein interaction network: before and after spectral analysis



The topological structure in protein-protein interaction network In clique, proteins connect quite tightly, almost interacting with each other. However, in each bipartite, proteins were divided into two parts, proteins seldom connect in same parts but connect tightly with proteins in counter part.

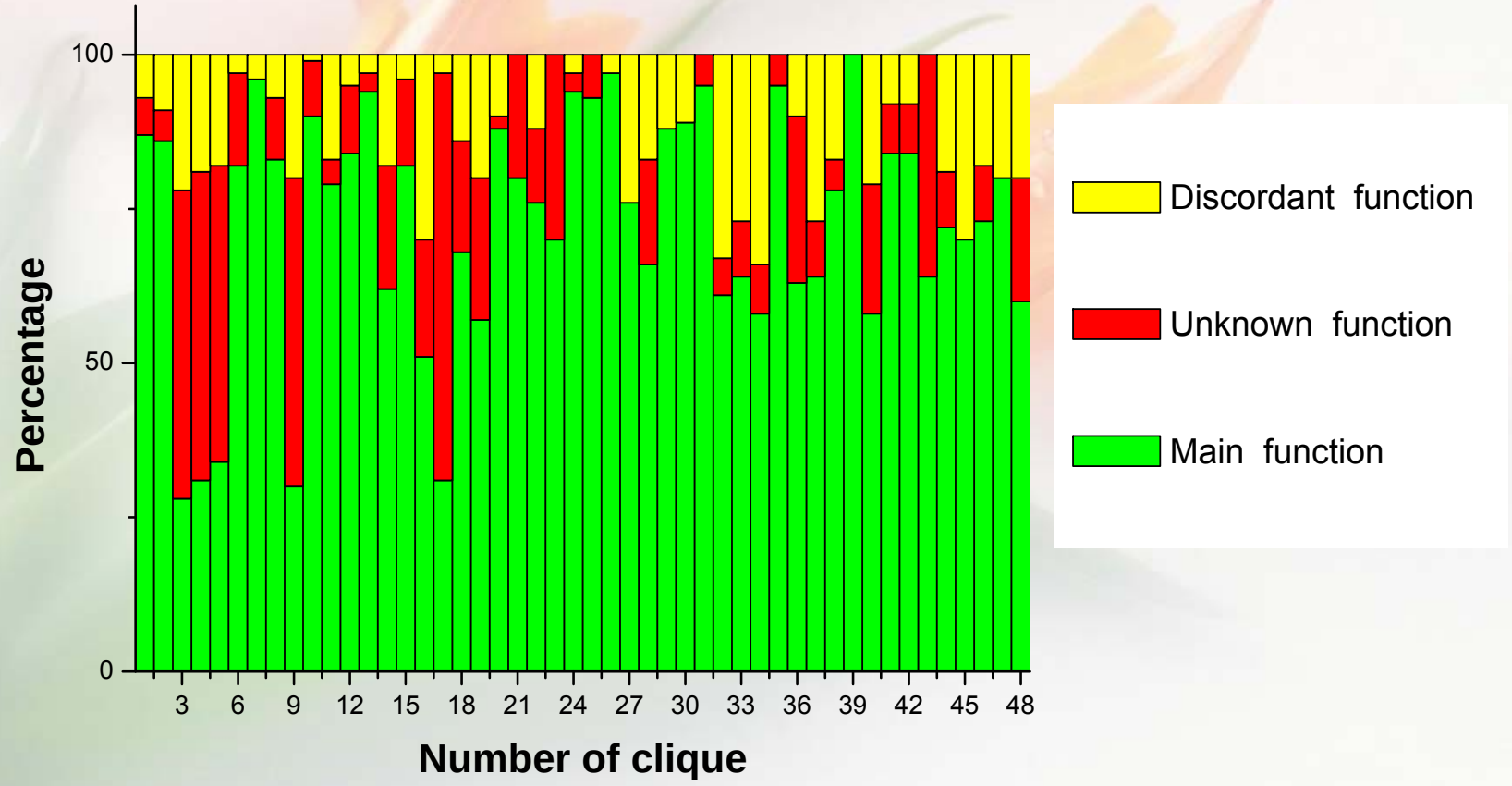


A Clique



b Bipartite

The percentage of function classes in every clique



Ma'ayan A, Blitzer RD, Iyengar R

Toward predictive models of mammalian cells

ANNUAL REVIEW OF BIOPHYSICS AND BIOMOLECULAR STRUCTURE 34: 319-349 2005

Papin JA, Hunter T, Palsson BO, et al.

Reconstruction of cellular signalling networks and analysis of their properties

NATURE REVIEWS MOLECULAR CELL BIOLOGY 6 (2): 99-111 FEB 2005

Xia Y, Yu HY, Jansen R, et al.

Analyzing cellular biochemistry in terms of molecular networks

ANNUAL REVIEW OF BIOCHEMISTRY 73: 1051-1087 2004

Environment of IBP

The Institute of Biophysics has a space of buildings more than 28000 square meters for the work, that includes laboratories, the library and auditorium. In addition, we have several courts for basketball, tennis and a gym center.

112


中国科学院生物物理研究所


Thank you!