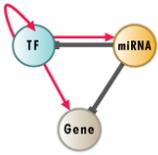


**RegNetwork: Regulatory Network Repository**

*of Transcription Factor and microRNA Mediated Gene Regulations*

# RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse

丁怡雯  
籍若言



## ■ 数据库概述

➤ 数据库建立背景

➤ 数据库功能

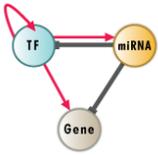
■ Data sources

■ Regulatory relationship curation and prediction

■ Database implementation and web user interface design

■ Network analysis

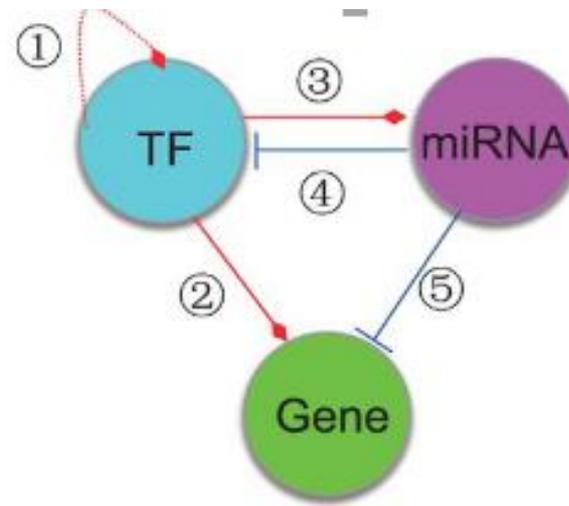
■ 数据库研究进展

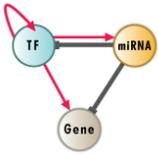


**RegNetwork: Regulatory Network Repository**  
*of Transcription Factor and microRNA Mediated Gene Regulations*

- RegNetwork 是人类和小鼠五种转录和转录后调控关系的数据储存库:

- 1、TF→TF
- 2、TF→gene
- 3、TF→miRNA
- 4、miRNA→TF
- 5、miRNA→gene

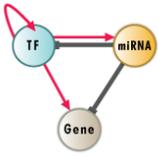




**RegNetwork: Regulatory Network Repository**

*of Transcription Factor and microRNA Mediated Gene Regulations*

- RegNetwork 的目的是建立一个平台，收集实验验证的调控关系和预测的调控关系。从 RegNetwork，你可以查询并识别 TFs，miRNA 和基因之间的组合和协同的调控关系。
- 网址：<http://www.regnetworkweb.org>



## ■ 数据库概述

## ➤ 数据库建立背景

## ➤ 数据库功能

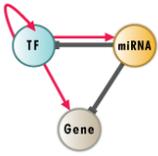
## ■ Data sources

## ■ Regulatory relationship curation and prediction

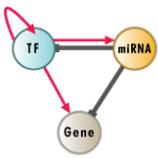
## ■ Database implementation and web user interface design

## ■ Network analysis

## ■ 数据库研究进展



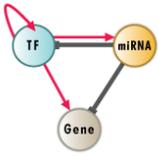
- 基因表达的转录和转录后调控具有根本重要性在众多生物过程中。如今，越来越多的大量基因调控关系已经出现在各种数据库和文献中。然而，为了更有效地利用这种知识为生物医学研究和应用，构建全基因组调控网络数据库在基因调控关系上整合广泛分散在多个不同位置的信息是必要的。
- 先验知识对来自多个源的（如基因组上下文、保守的基因共表达、基因敲除或高通量实验）基因调控关系在各种数据库和文献中展开。因此开发一个统一的数据库,为用户提供信息访问或检索工具是必要的。



**RegNetwork: Regulatory Network Repository**

*of Transcription Factor and microRNA Mediated Gene Regulations*

然而，之前向这一目标迈进只是有限的，如 RegulonDB 为埃希氏大肠杆菌 *Escherichia coli*，并对其他物种的全基因组调控关系工程到目前为止仍缺乏。考虑到人类和小鼠在生物医学研究中压倒性的重要性，我们为这两个物种建立了全基因组调控关系的数据库。



## ■ 数据库概述

### ➤ 数据库建立背景

### ➤ 数据库功能

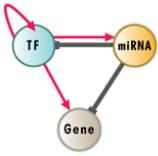
## ■ Data sources

## ■ Regulatory relationship curation and prediction

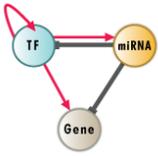
## ■ Database implementation and web user interface design

## ■ Network analysis

## ■ 数据库研究进展



- 通过结合的高通量转录组分析数据与在各种数据库和文献中可用的已知或预测调控关系的先验知识。例如，通过适当合并转录组分析数据的先验知识，可以显著提高调控关系识别的准确性。另外，从其他几个独立的研究结果表明对应于某些特定的表型，先验知识合并可有助于更好地确定特定于上下文的调控互作。因此，为了未来的生物医学研究和实践，收集、组织和与相关的组织共享等先验信息是首要任务。



## ■ 数据库概述

➤ 数据库建立背景

➤ 数据库功能

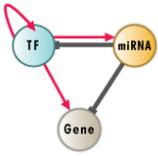
## ■ Data sources

■ Regulatory relationship curation and prediction

■ Database implementation and web user interface design

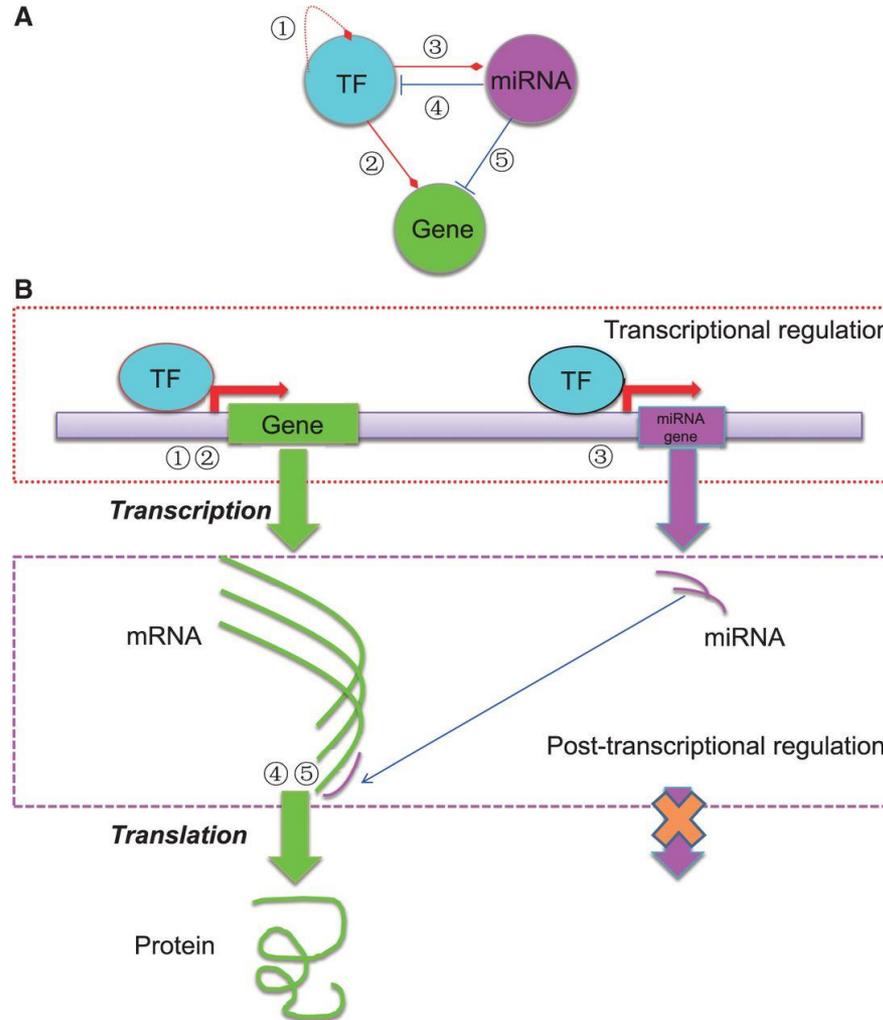
■ Network analysis

■ 数据库研究进展

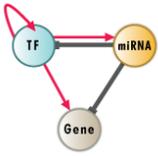


- 我们把 TFs 和 miRNAs 作为调控子。Figure 1 显示了涉及 TF, miRNA 和靶基因基本的调控通路, 以及基因表达的转录和转录后调控的重要步骤。请注意, 在以往大部分的研究基因调控网络的逆向工程中 miRNA 组件通常是缺失的。然而, 鉴于其在转录后的调控过程重要的作用, 我们认为在 RegNetwork 中包含 miRNA 是必要的。

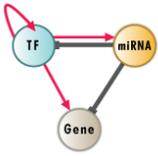
基本的调控通路涉及了TF、miRNA 和目标基因（A），基因表达转录和转录后调控的机制示意图(B)。TF、miRNA 和目标基因间共有五种类型的调控关系。



Zhi-Ping Liu et al. Database 2015;2015:bav095



- 如图1所示，TFs，miRNA和目标基因调控网络的五种类型的。更具体地说，转录调控的关系，‘TF-TF’ (①) 和 ‘TF-基因’ (②) ;转录后的调控关系的整合和预测 ‘miRNA- 基因’的相互作用 (⑤);对于调控子之间的互作，还包括 ‘TF -miRNA’ (③) 和 ‘miRNA-TF’ (④)。大量的数据库包含人类和小鼠的调控关系，从中我们收集相关信息和数据（例如 TFs，miRNS，TFBS域、基因及其注释）。



## ■ 数据库概述

➤ 数据库建立背景

➤ 数据库功能

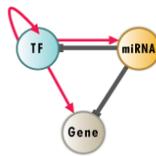
■ Data sources

■ Regulatory relationship curation and prediction

■ Database implementation and web user interface design

■ Network analysis

■ 数据库研究进展



## RegNetwork: Regulatory Network Repository

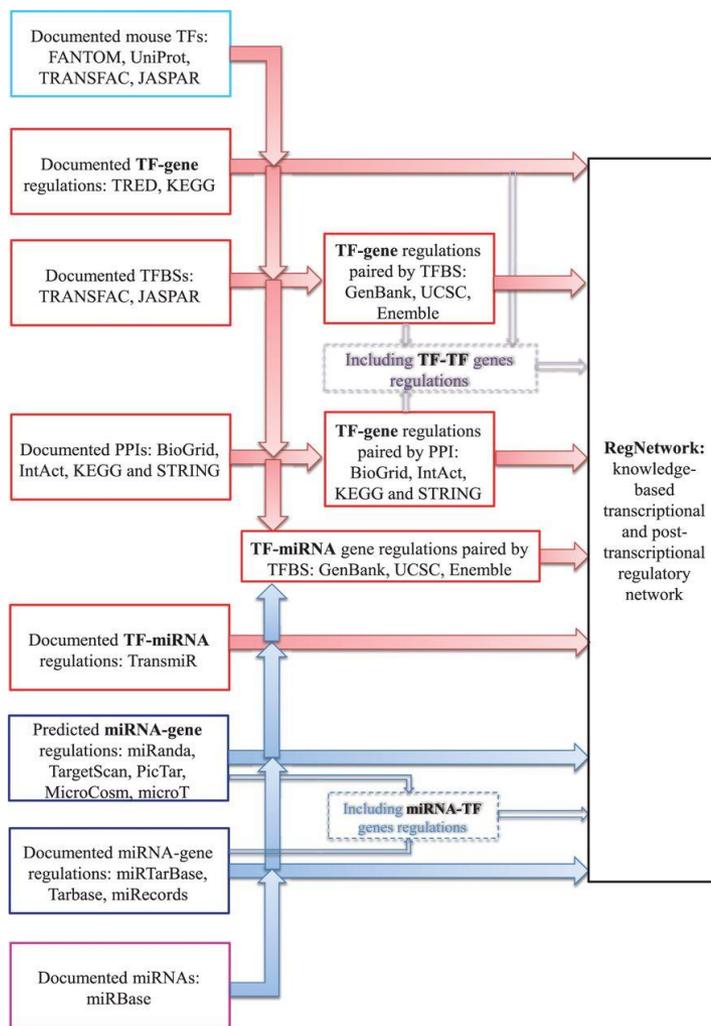
of Transcription Factor and microRNA Mediated Gene Regulations

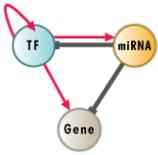
- 该表列出我们用于生成 RegNetwork 的数据库

**Table 1.**  
The databases used to build the RegNetwork database by collecting knowledge on gene regulatory relationships in human and mouse

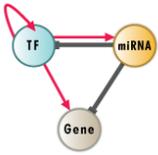
Database	Description	Species	Website	Reference	Version/access date
BioGrid	BioGRID is an online interaction repository with data compiled through comprehensive curation efforts	Mouse	<a href="http://thebiogrid.org/">http://thebiogrid.org/</a>	(21)	Version 3.2.100
Ensembl	Ensembl is to provide a centralized resource for geneticists, molecular biologists and other researchers studying the genomes of our own species and other vertebrates and model organisms	Human and mouse	<a href="http://www.ensembl.org">http://www.ensembl.org</a>	(22)	Release 71 (March 2013)
FANTOM	Functional Annotation Of Mammalian genome and is an international research consortium to assign functional annotations to the full-length complementary DNAs (cDNAs)	Human and mouse	<a href="http://fantom.gsc.riken.jp/">http://fantom.gsc.riken.jp/</a>	(23)	5 March 2010
GenBank	A comprehensive database developed by NCBI, NIH, which contains publicly available nucleotide sequences for more than 250 00 formally described species	Human and mouse	<a href="http://www.ncbi.nlm.nih.gov/genbank/">http://www.ncbi.nlm.nih.gov/genbank/</a>	(24)	14 August 2012
HPRD	HPRD is a curated human protein-protein interaction database	Human	<a href="http://www.hprd.org">http://www.hprd.org</a>	(25)	Release 9
IntAct	IntAct is a database system of molecular interaction data. All interactions are derived from	Mouse	<a href="http://www.ebi.ac.uk/intact/">http://www.ebi.ac.uk/intact/</a>	(26)	16 October 2012

图 2 说明了如何使用表 1 中列出的数据库构建的 RegNetwork，和分别对人类和小鼠执行同样计算机代码程序形式的测试。





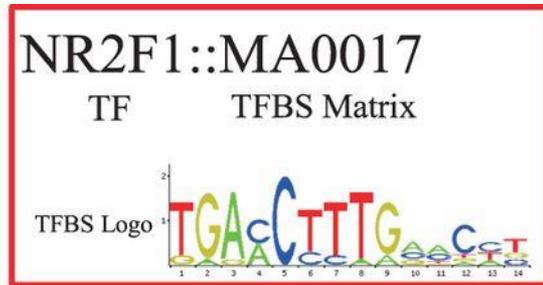
- 为描述转录调控的关系，我们首先列出人类和老鼠的TFs列表，分别从 FANTOM、UniProt、TRANSFAC 和 JASPAR。然后，把记录在 TRED 和 KEGG 的 'TF-gene' 互作直接存放在 RegNetwork。此外，我们从记录在 TRANSFAC 和 JASPAR 的 TF 结合位点 (TFBS) 域预测潜在的 '**TF-gene**' 互作。因为 TFs 通过结合到这些实验识别的 TFBS 上 调节靶基因，我们通过搜索转录起始位点 (TSS) 参考序列基因的上游 5 kb 到下游 1 kb 的转录启动子区对 TFs 和 gene 进行比对。



# example

- TF 'NR2F1'有一个已知的TFBS 'MA0017', 由一个位置加权矩阵表示, 图 3 中的左上角的序列标志显示其核苷酸组成。在人和小鼠的整个基因组启动子区筛查这个TFBS, 其启动子区域的基因包含 'MA0017'则被确定为 'NR2F1' 的潜在目标。一般情况下, 我们从 UCSC 基因组浏览器和 Ensembl数据库中检索 TFBS 保守轨迹的信息。具体来说, UCSC 的 tfbsConsSites 表中包含人类/小鼠TFSS保守的位置和得分的比对结果。如果比对结果的得分不低于阈值, 结合位点就被认为是保守的。

由TFBS的TF和基因配对的示意图。当记录的TFBS 'MA0017' 发现启动子区的 'Gene2' 和 'Gene 5' 时，TF NR2F1 预计将有可能相应地调节这两个基因。



NR2F1 regulates Gene2, Gene5, ...

Gene1 acttagttaagctcgtgcattgagactcttttctgagagcggaccgggtaattctcgggtgctcgtcgtgtct

Gene2 tccaacaacacgacgacagcattgttac**TGACCTTTGAACCT**tgaggcgcagggatcaggtacctgcttgat

Gene3 aacgcgcgggcgggtaatgtgagtgctcgggttgccctcgagagtagtcggatacagtgggctcctgtcata

Gene4 ggagtcaaagctgaggggtcgttggcatcatgaacaatttaaaggggaggtcggcggagtgcgttgtccagg

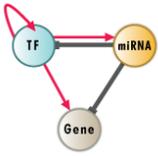
Gene5 **ggcgcgTAAACTTTGAACCT**atagaagtctctcccgtctcgtaaacatatcttcgtaaagaatcgatgcctt

Gene6 cttacgctcgttgtctactttgcggtgagactagacttactttactttgaactcatcacatttgataatggg

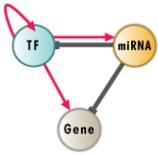
- agatatggatccttgaacctaatgtaaggatccggacaagtacggctgctggaaatgcctagcttaaagttt
- atgagggggaatggccgaaacacctttctggaacactaccaagagatcgtgtgcgcggtgga ttggaaagc
- cataaggtggatcattccgatagctctggggcgtagaagcgcgatgggtccaccagaaacatgggtggtagga
- Agtccgtgatcagcctatcttctcttcatgcacttatgtgtaagccatgtcagactaccgagggtgactaat

.....

Zhi-Ping Liu et al. Database 2015;2015:bav095



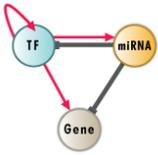
- 我们还考虑将**蛋白质-蛋白质**相互作用（PPI）包含在 RegNetwork 中。我们从 HPRD, BioGrid, IntAct, KEGG 和 STRING 检索包含至少一个 TF 的 PPI 对。
- TF 和其相互作用的目标之间的功能联系预示假定基因调控。很明显，当 TF 调节自身基因的表达，'**TF-TF**' 自我调控也能被识别。
- 转录后调控，miRTarBase, TarBase 和 miRecords 实验验证人类和老鼠的 'miRNA-gene' 对直接存放在 RegNetwork。然后，由五个代表性算法预测 '**miRNA-gene**' 相互作用，即 miRanda、TargetScan、PicTar、MicroCosm 和 micorT。同样，记录 '**miRNA-TF**' 基因的调控关系直接存入 RegNetwork。



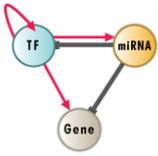
**RegNetwork: Regulatory Network Repository**

*of Transcription Factor and microRNA Mediated Gene Regulations*

- TransmiR中记录 '**TF-miRNA**' 的调控关系也直接导入 RegNetwork。然后，基于TFBS信息，使用如上文所述的潜在 '**TF-gene**' 相互作用类似的方法，预测TFs 和miRNA 编码基因之间潜在的相互作用。



- 以这样一种方式，记录和推测的调节对两者都被包含在了 RegNetwork 里。对于某些成对的基因，一个数据库中调节子（或目标）基因在另一个数据库中被标记为目标（或调节）基因。我们合并这样的结果，因此这些成对的基因之间的相互作用可以双向。那就是，我们用包容性的原则来处理数据库之间不一致。同时，我们还提供一个链接到原始数据库，供用户在特定的情况下检查不一致的问题的详细资料，并决定哪些结果他们会相信和使用。最后，我们通过使用3个水平标签（如'高'、'中'或'低'信度）的方法，为每个调控关系添加置信度。更具体地说，实验验证的调控被标记'高信度'标签，只有一个算法/方法所作的预测标记为'低信度'，其余标记为'中信度'。



## ■ 数据库概述

➤ 数据库建立背景

➤ 数据库功能

■ Data sources

■ Regulatory relationship curation and prediction

■ Database implementation and web user interface design

■ Network analysis

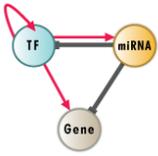
■ 数据库研究进展

# The web user interface of RegNetwork.

The screenshot shows the RegNetwork web user interface. At the top, there is a logo with 'TF', 'miRNA', and 'Gene' nodes. Below it is the title 'RegNetwork: Regulatory Network Repository of Transcription Factor and microRNA Mediated Gene Regulations'. A navigation bar includes 'HOME', 'SEARCH', 'ABOUT', 'SOURCE', 'DOWNLOAD', 'CONTRIBUTION', and 'LINKS'. The search section has a search box, filters for Type (All), Organism (Human), Evidence (All), Confidence (All), Databases (All, biogrid, ensembl, hprd, intact, kegg), and Records per Page (30). There are 'Search' and 'Reset' buttons. Below the search section, it says 'Found 368573 records, in 12286 pages'. There are 'Export as: CSV' and 'Export' buttons, and 'Jump to page: 1' with a 'Go' button. A table shows the first two results:

NO.	REGULATOR SYMBOL	REGULATOR ID	TARGET SYMBOL	TARGET ID	DATABASES	EVIDENCE	CONFIDENCE
1	AATF	26574	SP1	6667	hprd	Predicted	Low
2	AATF	26574	CHEK2	11200	hprd	Predicted	Low

Zhi-Ping Liu et al. Database 2015;2015:bav095



**RegNetwork: Regulatory Network Repository**

*of Transcription Factor and microRNA Mediated Gene Regulations*

## ■ 数据库概述

➤ 数据库建立背景

➤ 数据库功能

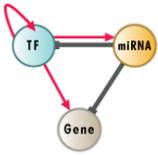
■ Data sources

■ Regulatory relationship curation and prediction

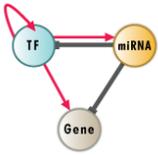
■ Database implementation and web user interface design

■ Network analysis

■ 数据库研究进展

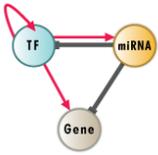


- 真实的生物网络，如基因调控网络和蛋白质-蛋白质互作网络是不同于随机网络的，在某些网络属性，像特征路径长度和节点度分布。因此，网络特征分析使我们能够评估网络是否是随机的。表 3 从 RegNetwork 总结了一些人类和小鼠已建立调控网络的网络功能的目录。已建立的人类和小鼠调控网络的聚类系数分别为 0.118 和 0.101，远高于同类规模随机网络( $\sim 1.5 \times 10^{-5}$ )的随机网络。此外，人类和小鼠调控网络的特征路径长度分别 3.200 和 3.229，比同类规模随机网络小，由此可以看出，调控信息的快速传播一个快速传播的调控信息是非随机的。所有其他的网络拓扑性质还表明人类和小鼠以建立的调控网络是不同于随机网络的



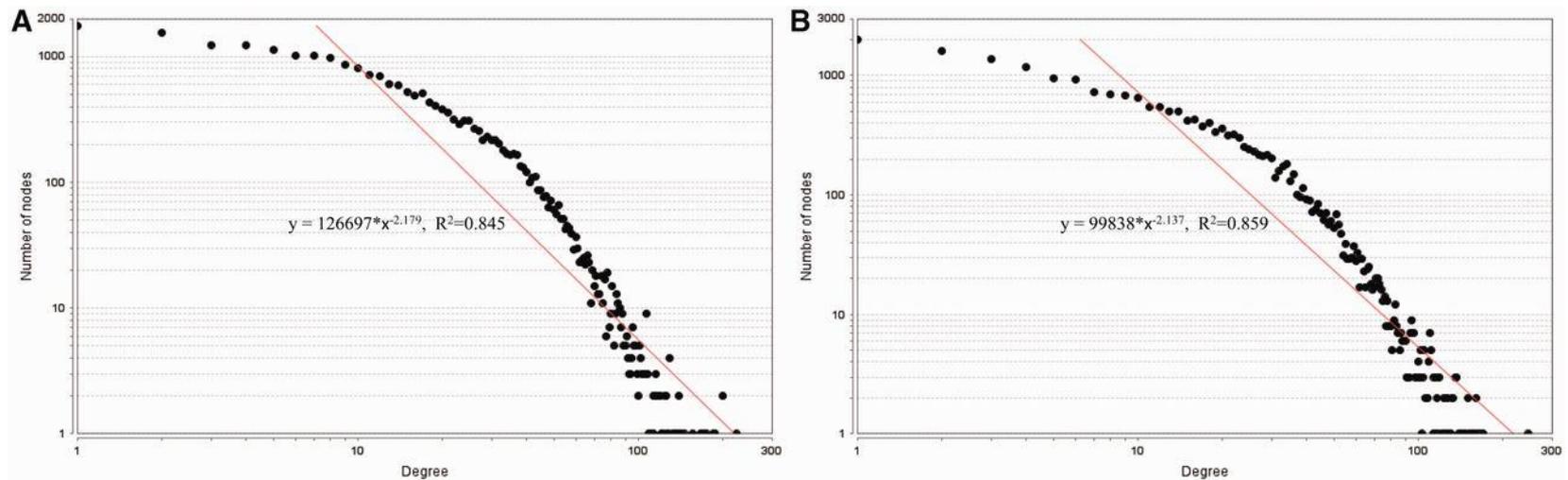
Parameter	Value	
	Human	Mouse
Clustering coefficient	0.118	0.101
Connected components	3	1
Network diameter	8	8
Shortest paths	42 727 382	36 743 196
Characteristic path length	3.200	3.229
Average number of neighbors	31.391	30.548

Table 3. Selected measures in the established regulatory networks for human and mouse

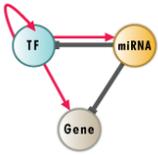


- 第二，已建立的网络节点度计算，并发现以满足power law分布，如图 5 所示。拟合 power law模型  $y = \alpha \cdot x^{-\gamma}$  表示节点的数目，而  $x$  表示节点的度，人类调控网络的  $\gamma = 2.179$ ，小鼠调控网络的  $\gamma = 2.137$ 。由于  $2 \leq \gamma \leq 3$ ，我们背景网络是无标度。网络参数提供的证据表明，我们综合的调控网络有别于随机生成网络。

# The node degree distributions of the established regulatory networks in human (A) and mouse (B).

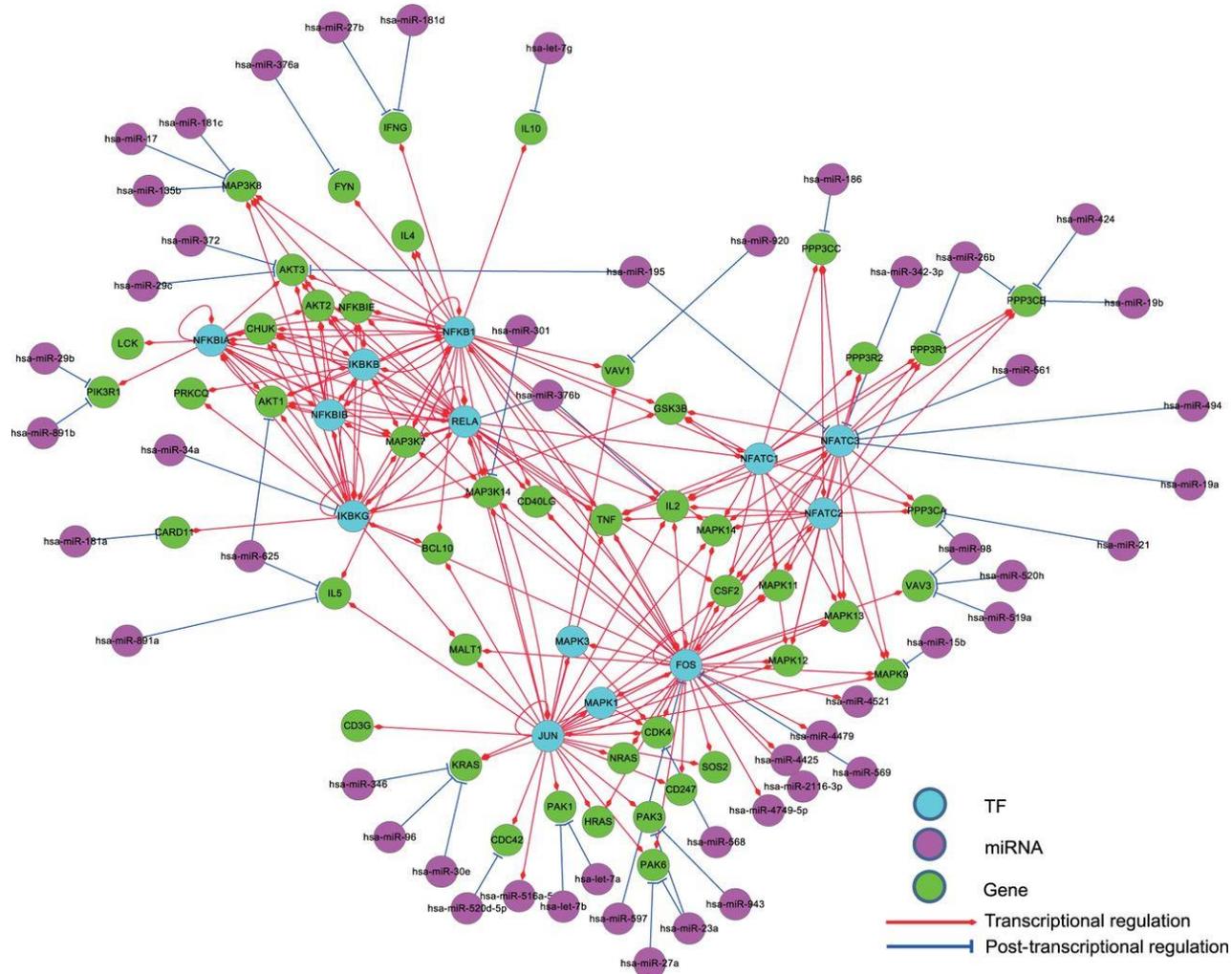


Zhi-Ping Liu et al. Database 2015;2015:bav095

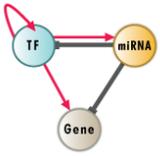


- 不同于现有的调控关系数据库，如 TRED，RegNetwork 包含转录和转录后调控相互作用，这使得我们能够调查调节子（即 TF 和 miRNA）和其靶基因之间更为复杂的互作关系。图 6 阐明了从KEGG基因集收集的涉及人类T细胞信号通路的相互作用，转录后调控关系用蓝线绘制。
- 例如，视觉上我们可以区分几个网络图案在图 6 中，如 'FOS'-'hsa-miR-569'-'MAPK12' 和 'JUN' - 'hsa-let-7a' - 'PAK1'，这些被认为是主要的网络建设组件和调控网络中的功能块。
- 通过这个简单的例子，我们展示了RegNetwork 是查询基于知识的转录和转录后组合调控关系的有用工具。

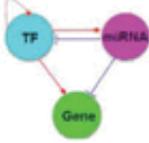
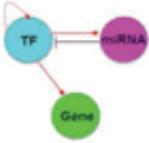
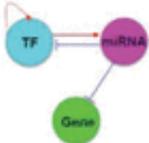
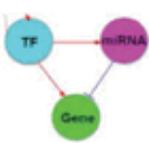
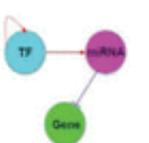
# The regulatory relationships of a KEGG gene set for the human T cell receptor signaling pathway in RegNetwork.



Zhi-Ping Liu et al. Database 2015;2015:bav095

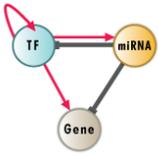


**Table 4.**  
 The three-node network motifs 'TF-miRNA-gene' in human and mouse regulatory networks

ID	Motif	Species	Occurrence	Z-Score
M1		Human	2047	23.563
		Mouse	23 658	24.56
M2		Human	28 872	8.223
		Mouse	1 634 783	33.479
M3		Human	41 022	4.013
		Mouse	304 806	21.131
M4		Human	107 732	-3.773
		Mouse	165 549	-24.261
M5		Human	133 350	3.631
		Mouse	3 432 863	24.996

• 实分表在模异的FA如小'N同统谓

以。昇。差率。比。和。和法。



**RegNetwork: Regulatory Network Repository**

*of Transcription Factor and microRNA Mediated Gene Regulations*

## ■ 数据库概述

➤ 数据库建立背景

➤ 数据库功能

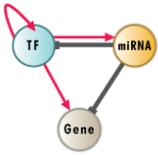
■ Data sources

■ Regulatory relationship curation and prediction

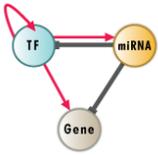
■ Database implementation and web user interface design

■ Network analysis

■ 数据库研究进展



- 在这篇文章中，通过整合不同的数据源，基于人类和小鼠全基因组调控网络知识，我们开发了一个数据库-RegNetwork。为公共访问收集和重组了一套全面的TFs, miRNAs 和目标基因之间的互作关系。RegNetwork 既定的调控网络提供全基因组调控互作的关系，奠定了最初的基础和建立一个事先背景网络，在通路或子网对应不同的表型中确定或验证分子和功能调节。并且，在特定的生理和发育条件（例如病毒感染）下结合高通量表达数据，可以识别RegNetwork背景网络下的微分子网和通路，这将导致在上下文特定的调控机制过程中新颖和有趣的新见解。



- 当前 RegNetwork was 版本的开发，在 ENCODE 项目高通量数据集中发布了涉及人类的成千上万的调控关系，其中包含 119 个 TFs、736 miRNA 与 15 131 个基因的 162 100 个调控关系。在 ENCODE 中大多数的 TFs，miRNA 和基因（96.5% 的调控机构和 89.9% 的目标）已经列入了我们的数据库。我们将继续跟踪和定期整理 ENCODE 调控关系到我们的数据库。我们也认识到文本挖掘工具用来识别和管理文库中调控关系，在另一个方向扩展 RegNetwork 的效用。我们也计划为每个调控关系扩大 RegNetwork 要包含附加信息如实验条件和原始的引用。
- 我们也会扩展 RegNetwork 包含其他的物种，如褐家鼠（鼠）、果蝇（果蝇）、秀丽线虫 (蠕虫)，埃希氏大肠杆菌（大肠杆菌）和酿酒酵母（酵母）。



**Have a rest**