Bioinformatics, 31(19), 2015, 3237–3239
doi: 10.1093/bioinformatics/btv331
Advance Access Publication Date: 29 May 2015
Applications Note

OXFORD

Database and ontologies

RPdb: a database of experimentally verified cellular reprogramming records

Yi Shen[1,†], Fan Gao[2,†], Minghui Wang[1,3] and Ao Li[1,3,*]

[1]School of Information Science and Technology, University of Science and Technology of China, Hefei, AH230027, China, [2]School of Life Science and [3]Centers for Biomedical Engineering, University of Science and Technology of China, Hefei, AH230037, China

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
[*]To whom correspondence should be addressed.

Associate Editor: Ivo Hofacker

Received on October 4, 2014; revised on May 20, 2015; accepted on May 25, 2015

Abstract

Summary: Many cell lines can be reprogrammed to other cell lines by forced expression of a few transcription factors or by specifically designed culture methods, which have attracted a great interest in the field of regenerative medicine and stem cell research. Plenty of cell lines have been used to generate induced pluripotent stem cells (IPSCs) by expressing a group of genes and microRNAs. These IPSCs can differentiate into somatic cells to promote tissue regeneration. Similarly, many somatic cells can be directly reprogrammed to other cells without a stem cell state. All these findings are helpful in searching for new reprogramming methods and understanding the biological mechanism inside. However, to the best of our knowledge, there is still no database dedicated to integrating the reprogramming records. We built RPdb (cellular reprogramming database) to collect cellular reprogramming information and make it easy to access. All entries in RPdb are manually extracted from more than 2000 published articles, which is helpful for researchers in regenerative medicine and cell biology.

Availability and Implementation: RPdb is freely available on the web at http://bioinformatics.ustc.edu.cn/rpdb with all major browsers supported.

Contact: aoli@ustc.edu.cn

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Cellular reprogramming is a process in which one cell type transforms into another cell type by expressing a group of genes and microRNAs, adding chemical molecules or changing cell culture environment (Hussein and Nagy, 2012). Cellular reprogramming technology, which mainly includes induced pluripotent stem cell (IPSC) generation, IPSC differentiation and direct reprogramming, is considered one of the most important advances in tissue replacement therapies (Tabar and Studer, 2014; Walmsley *et al.*, 2014). Somatic cells can be reprogramed to a stem cell-like state, which can differentiate to functional somatic cells. Multiple somatic cells, including blood cells (Giorgetti *et al.*, 2009), fibroblasts (Takahashi *et al.*, 2007) and hair follicle dermal papilla cells (Tsai *et al.*, 2011) have been successfully reprogrammed to IPSCs. There are also many functional somatic cells that have been generated from IPSCs, such as dopaminergic neurons (Theka *et al.*, 2013) and retinal cells (Mellough *et al.*, 2012). Multiple strategies have been used to achieve these goals, including chemical-inducible systems (Nie *et al.*, 2012), virus-free transfection, recombinant proteins and microRNAs (Walmsley *et al.*, 2014). Using similar strategies, various somatic cells have been directly converted into other cell types, which are named direct reprogramming or transdifferentiation.

Although there are several stem-cell-focused databases, such as SyStemCell (Yu *et al.*, 2012) and StemCellDB (Mallon *et al.*, 2013), to the best of our knowledge, there is no database that focuses on comprehensively collecting and analyzing cellular reprogramming

information, especially the reprogramming factors (genes or chemical molecules). To fill this gap, we collected experimentally verified reprogramming records from published articles and built RPdb (cellular reprogramming database), which is proposed to facilitate the new prescription screening and assist with investigating the mechanism of cellular reprogramming.

2 Methods

We collect reprogramming information from published literature included in the PubMed database and Web of Knowledge by using all the keywords 'induced pluripotent stem cell' AND 'generation', 'induced pluripotent stem cell' AND 'differentiation', 'direct reprogramming'. More than 2000 articles were obtained from PubMed and Web of Knowledge in this way. Based on the keywords, these articles are divided into three categories: IPSC generation, IPSCs differentiation and direct reprogramming. We manually extract reprogramming information from each article and the information is uploaded to RPdb after reviewing (supplementary material).

The flowchart of the construction of RPdb is illustrated in Figure 1. For IPSCs generation, we collect information regarding the cell lines used for reprogramming, reprogramming factors (genes, microRNAs and chemical molecules) and species. Based on the method used to deliver the reprogramming factors, we divide the entries into 'virus' and 'non-virus'. For manually extracted genes and microRNAs, we further map them to NCBI/EMBL Entrez IDs according to the symbols and description. Genes and microRNAs that cannot be found in public databases are removed (Supplement Material). A variety of biological databases, such as mirTarBase (Hsu *et al.*, 2014), STRING (Szklarczyk *et al.*, 2011) and AnimalTFDB (Zhang *et al.*, 2012), are integrated to provide functional and regulatory information about genes and microRNAs (Fig. 1A, B).

To help users find new reprogramming genes, we also provide an online tool for data analysis, which compares multiple tissues and identifies differentially expressed transcription factors and genes (Fig. 1C). We download the expression profiles of 27 human tissues and 6 mouse tissues from ATLAS (Petryszak *et al.*, 2014). The details of data processing and identification of differentially expressed genes can be found in supplementary file. Experimentally validated transcription factors are downloaded from AnimalTFDB and such information is used to identify transcription factors from the differentially expressed genes.

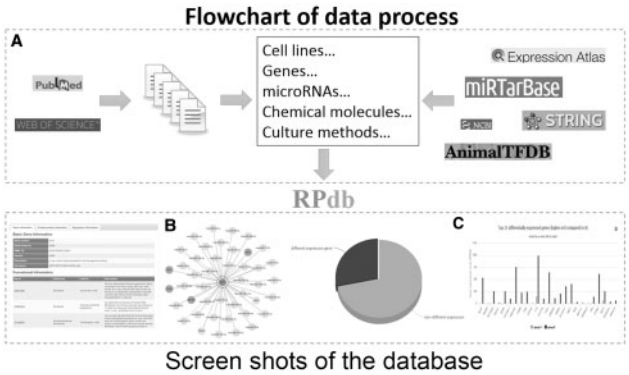

Fig. 1. The flowchart of RPdb construction. (**A**) The flowchart showing data processing and information integration. (**B**) The screen shots of the 'Gene' page with basic gene information and regulation information. (**C**) The results of comparing gene expression in two tissues

3 Results

The database includes reprogramming records of 250 cell lines, 197 genes and microRNAs, 144 chemical molecules and culture methods and 526 functional annotations. We classify the entries into two categories: 'Gene and microRNAs' and 'Molecules and culture methods'.

Users can explore all entries of the database through the 'Browse' page and compare gene expression levels of multiple tissues in the 'Analysis' page. All the results can be downloaded for further analysis. A tutorial for the database is available in the 'Help' page, including searching genes of interest, using the database to obtain reprogramming protocols and downloading the results for further analysis. Detailed information about genes and microRNAs involved in cellular reprogramming can be found in the 'Gene' page, which includes basic gene information, functional information, GO information, protein-protein interactions and regulatory relationships. All functional information is manually extracted from literature describing the role of genes in the reprogramming process. With RPdb, users can find genes, microRNAs and chemical molecules used for reprogramming processes. The 'Molecules and Culture Methods' page lists all chemical molecules and culture methods deposited in RPdb.

4 Discussion and conclusions

RPdb is an integrated database comprising plenty of experimentally validated reprogramming information. Our aim is to collect comprehensive information used for cellular reprogramming and make it freely accessible to users. RPdb provides detailed information about experimentally verified genes and microRNAs, including gene descriptions and functional annotations manually extracted from the literature. The development and expansion of RPdb will keep on going and more valuable resources will be integrated into the database in the future. As an efficient bioinformatics tool, RPdb will assist in the discovery of novel reprogramming factors in a broad spectrum of cell reprogramming research.

Acknowledgements

We thank Chuanchun Han (hanchuanchun@163.com) and Hao Gu (haogu@mail.ustc.edu.cn) for reviewing reprogramming information in the database and suggestions for database building.

Funding

National Natural Science Foundation of China (61471331, 31100955).

Conflict of Interest: none declared.

References

Giorgetti,A. *et al.* (2009) Generation of induced pluripotent stem cells from human cord blood using OCT4 and SOX2. *Cell Stem Cell*, **5**, 353–357.

Hsu,S.D. *et al.* (2014) miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. *Nucleic Acids Res.*, **42**, D78–D85.

Hussein,S.M.I. and Nagy,A.A. (2012) Progress made in the reprogramming field: new factors, new strategies and a new outlook. *Curr Opin. Genet. Dev.*, **22**, 435–443.

Mallon,B.S. *et al.* (2013) StemCellDB: the human pluripotent stem cell database at the National Institutes of Health. *Stem Cell Res.*, **10**, 57–66.

Mellough,C.B. *et al*. (2012) Efficient stage-specific differentiation of human pluripotent stem cells toward retinal photoreceptor cells. *Stem Cells*, **30**, 673–686.

Nie,B.M. *et al*. (2012) Cellular reprogramming: a small molecule perspective. *Curr. Opin. Cell Biol.*, **24**, 784–792.

Petryszak,R. *et al*. (2014) Expression Atlas update—a database of gene and transcript expression from microarray- and sequencing-based functional genomics experiments. *Nucleic Acids Res.*, **42**, D926–D932.

Szklarczyk,D. *et al*. (2011) The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.*, **39**, D561–D568.

Tabar,V. and Studer,L. (2014) Pluripotent stem cells in regenerative medicine: challenges and recent progress. *Nat. Rev. Genet.*, **15**, 82–92.

Takahashi,K. *et al*. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, **131**, 861–872.

Theka,I. *et al*. (2013) Rapid generation of functional dopaminergic neurons from human induced pluripotent stem cells through a single-step procedure using cell lineage transcription factors. *Stem Cells Transl. Med.*, **2**, 473–479.

Tsai,S.Y. *et al*. (2011) Single transcription factor reprogramming of hair follicle dermal papilla cells to induced pluripotent stem cells. *Stem Cells*, **29**, 964–971.

Walmsley,G.G. *et al*. (2014) Induced pluripotent stem cells in regenerative medicine and disease modeling. *Curr. Stem Cell Res. Therapy*, **9**, 73–81.

Yu,J. *et al*. (2012) SyStemCell: a database populated with multiple levels of experimental data from stem cell differentiation research. *PloS One*, **7**, e35230.

Zhang,H.M. *et al*. (2012) AnimalTFDB: a comprehensive animal transcription factor database. *Nucleic Acids Res.*, **40**, D144–D149.