

Correspondence

EMT-Regulome: a database for EMT-related regulatory interactions, motifs and network

Cell Death and Disease (2017) 8, e2872; doi:10.1038/cddis.2017.267; published online 15 June 2017

Dear Editor,

The Epithelial-mesenchymal transition (EMT) is a fundamental cellular process that governs biological processes, such as embryonic development, tumor cell metastasis, organ fibrosis and tissue regeneration.¹ EMT allows polarized, immotile epithelial cells to transform into motile mesenchymal cells.² This switch is mediated by regulatory factors such as transcription factor (TF), microRNA (miRNA) and long non-coding RNA (lncRNA) at transcriptional and post-transcriptional levels.^{3,4} These regulatory factors play co-regulatory roles on the EMT process and the regulatory loops act as motifs, which are the basic units in the regulatory network.⁵ However, a database with comprehensive EMT regulatory relationships is still lacking. EMT-Regulome comprehensively collected the interactions between regulators and targets, including TF-target, miRNA-target, miRNA-lncRNA, TF-lncRNA and TF-miRNA and identified 10 types of co-regulation motifs for EMT process (Figure 1). EMT-Regulome provides the option for searching EMT-related regulatory interactions, motifs and networks, which facilitates the investigations of the mechanisms of EMT involved in diseases, such as tumor cell invasion. The database is freely available at www.medsysbio.org/EMTRegulome.

In current version, EMT-Regulome contains 1740 miRNAs, 8783 lncRNAs and 13 025 coding genes, which includes 370 TF coding genes and 350 EMT-related genes (Supplementary Figure 1A). We collected regulatory relationships, including TF-targets, miRNA-targets, miRNA-lncRNA, TF-lncRNA and TF-miRNA, from public databases (Figure 1). Then, we integrated the regulation into a comprehensive network and searched 10 types of motifs that are the basic units in the regulatory network (Figure 1). Especially, for the fifth motif type, known as competing endogenous RNA (ceRNA), we used hypergeometric test to measure whether two RNA components significantly shared miRNAs. The motifs containing at least one EMT-related gene from dbEMT,⁶ which collected experimentally verified EMT-related genes from literatures, were defined as EMT-related motifs in the EMT-Regulome database. The statistics of motifs are shown in Figure 1. The EMT-Regulome database provides a user-friendly interface that allows users to conveniently search, browse and download data (Supplementary Figure 2). In the 'Search' page, the query function can be performed using

Symbol or ID of coding gene/miRNA/lncRNA with specified motif types. The search results are composed of three parts: (i) information box of gene, miRNA or lncRNA, which is offered as a search condition, including brief functional description, location in chromosome and external link to NCBI, miRBase or Ensembl for more detail; (ii) motifs are grouped by types, and source databases of their regulation are provided. (iii) returned motifs could be visualized as a network and are allowed to be exported as PNG files in various layouts by the Cytoscape web tool (<http://js.cytoscape.org>).

As a case analysis, the EMT-Regulome database was applied to identify potential activated EMT motifs in the mesenchymal subtype of ovarian cancer by employing expression profiles from The Cancer Genome Atlas (TCGA). Yang *et al.*⁷ have classified the TCGA ovarian cancer samples into integrated epithelial (iE) subtype and integrated mesenchymal (iM) subtype labels. Here, the activated motifs were defined as the nodes in motifs that were differentially expressed in iM compared with the iE ovarian cancer samples (T-test) and the edges in motifs were differentially co-expressed in iM type (Pearson correlation test). All the activated EMT-related motifs in ovarian cancer are available in the download page or by searching specific nodes with statistical control in search page.

In conclusion, EMT-Regulome provides a comprehensive landscape of EMT-related regulation motifs to facilitate the investigation of the mechanism of EMT involved in the pathological and physiological processes. The features of EMT-Regulome include, (i) comprehensively integrated different EMT-related regulatory relationships from multiple databases; (ii) systematically provided EMT-related motifs, such as feed-forward loop and ceRNA with statistical control; (iii) all the EMT-related regulatory relationships can be visualized as network and be downloaded.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements. The authors acknowledge the efforts of all of researchers who have contributed the data to the public databases of TargetScan, miRanda, miRBase, miRTarBase, StarBase, miRcode, MsigDB, TRANSFAC, Transmir, ChIPBase, dbEMT, NCBI, Gencode, TCGA and TANRIC. This work was supported by the National Natural Science Foundation of China [grant numbers 61673143, 31671187, 81572935 and 81372213] and Postdoctoral Scientific Research Developmental Fund [grant number: LBH-Q16166].

Zhangxiang Zhao^{1,2,6}, Wenbin Zhou^{1,6}, Yue Han¹, Fuduan Peng¹, Ruiping Wang¹, Ruihan Yu¹, Chengyu Wang^{1,2}, Haihai Liang³, Zheng Guo^{*,1,4,5} and Yunyan Gu^{*,1,2}

¹College of Bioinformatics Science and Technology, Department of Systems Biology, Harbin Medical University, Harbin 150086, China; ²Training Center for Students Innovation and Entrepreneurship Education, Harbin Medical University, Harbin 150086, China; ³Department of Pharmacology, Harbin Medical University, Harbin 150086, China; ⁴Key Laboratory of Ministry of Education for Gastro-intestinal Cancer, Department of Bioinformatics, Fujian Medical University, Fuzhou 350004, China and ⁵Fujian Key Laboratory of Tumor Microbiology, Fujian Medical University, Fuzhou 350004, China

*Corresponding author: Y Gu or Z Guo, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150086, China. Tel: +86-0451-86699584; Fax: +13936325420; E-mail: guyunyan@ems.hrbmu.edu.cn or guoz@ems.hrbmu.edu.cn

⁶These authors contributed equally to this work.

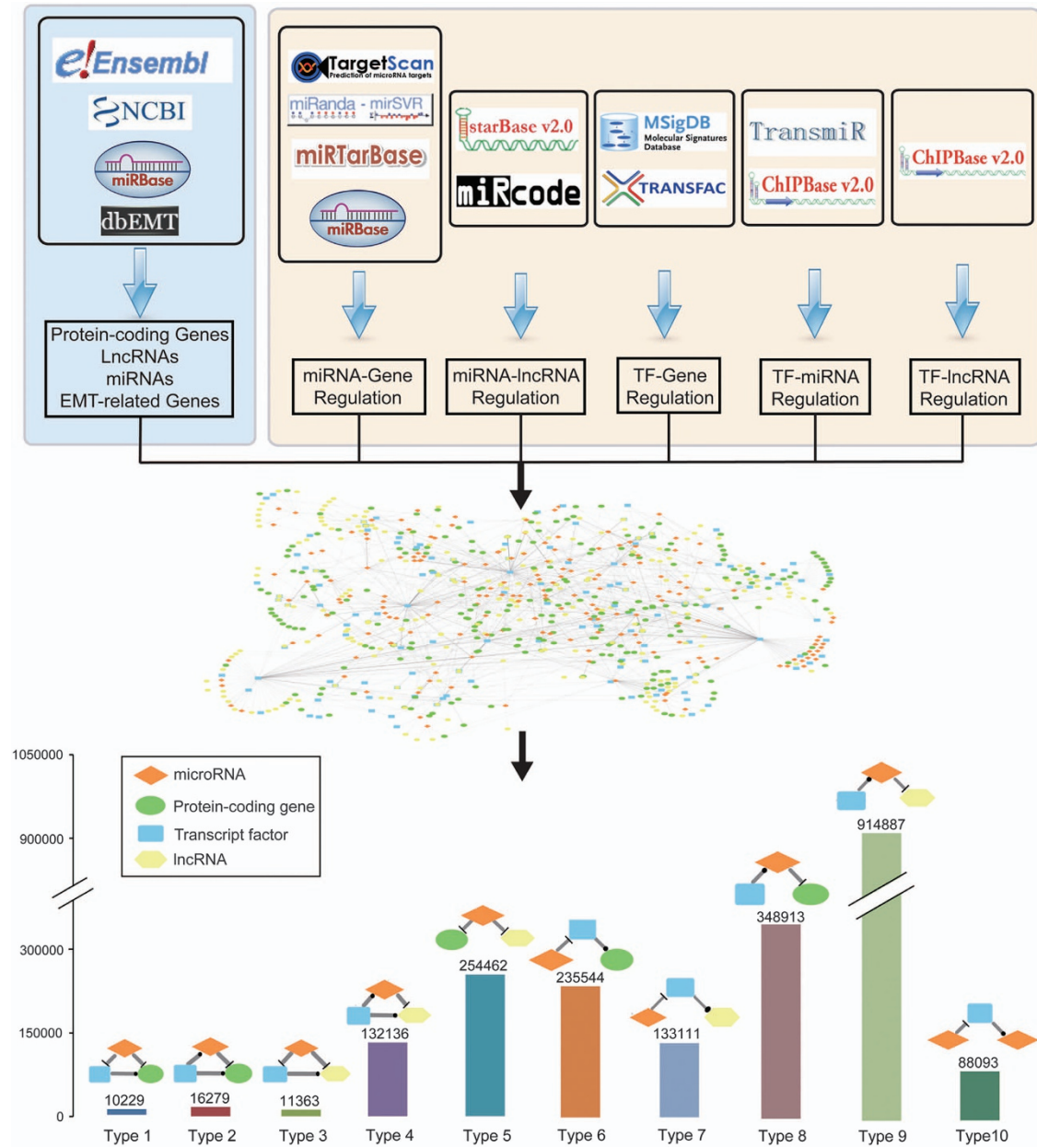


Figure 1 The flow chart of the EMT-Regulome database

1. Nieto MA *et al.* *Cell* 2016; **166**: 21–45.
2. Yang J, Weinberg RA. *Dev Cell* 2008; **14**: 818–829.
3. Wu CY *et al.* *Trends Genet. TIG* 2012; **28**: 454–463.
4. Richards EJ *et al.* *J Biol Chem* 2015; **290**: 6857–6867.
5. Zhang HM *et al.* *Brief Bioinformatics* 2015; **16**: 45–58.
6. Zhao M *et al.* *Sci Rep* 2015; **5**: 11459.
7. Yang D *et al.* *Cancer Cell* 2013; **23**: 186–199.

Supplementary Information accompanies this paper on Cell Death and Disease website (<http://www.nature.com/cddis>)



Cell Death and Disease is an open-access journal published by Nature Publishing Group. This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017