Genomics Data 7 (2016) 73–75

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in Brief

RNA sequencing analysis to demonstrate Erk dependent and independent functions of Mek

Chang Liu, Haixia Chen, Lin Liu *, Lingyi Chen *

State Key Laboratory of Medicinal Chemical Biology, Collaborative Innovation Center for Biotherapy, 2011 Collaborative Innovation Center of Tianjin for Medical Epigenetics, Tianjin Key Laboratory of Protein Sciences and College of Life Sciences, Nankai University, Tianjin 300071, China

ARTICLE INFO

Article history:
Received 17 November 2015
Accepted 30 November 2015
Available online 1 December 2015

Keywords:
Erk
Mek
Mouse embryonic stem cells

ABSTRACT

Mek inhibition and *Erk* knockout (KO) have quite distinct effects on pluripotency maintenance in mouse embryonic stem cells (ESCs). To test whether there is an Erk-independent function of Mek, RNA-sequencing (RNA-seq) is carried out on six samples, WT KH2 ESCs treated with or without PD0325901 (PD) for 48 h (KH2_PD and KH2, respectively), *iErk1*; *Erk KO* ESCs cultured in the presence of Dox (P0), 48 and 96 h after Dox withdrawal (P1 and P2, respectively), and *iErk1*; *Erk KO* ESCs cultured without Dox for 96 h, and treated with PD in the last 48 h (P2_PD). These RNA-seq data demonstrate that Mek inhibition has quite different effect on the transcriptional profile of mouse ESCs, compared to *Erk* KO. Moreover, a significant fraction of genes is regulated by Mek inhibition, regardless of the presence or absence of Erk, indicating an Erk-independent function of Mek. RNA-seq data are deposited in Gene Expression Omnibus (GEO) datasets under accession number GSE70304.

Specifications	
Organism/cell line/tissue	KH2 mouse ESC line
Sex	Male
Sequencer or array type	Illumina HiSeq 2000
Data format	Raw and analyzed
Experimental factors	Mek inhibition in WT KH2 ESCs; Mek inhibition in *Erk KO* ESCs; Erk knockout in *iErk1*; *Erk KO* ESCs.
Experimental features	Comparison between WT KH2 ESCs treated with or without PD0325901 for 48 h (KH2_PD and KH2, respectively) shows the Mek inhibition effect in WT ESCs. Comparison between *iErk1*; *Erk KO* ESCs without Dox for 96 h, with or without PD0325901 treatment (P2_PD and P2, respectively) reveals the Mek inhibition effect in *Erk KO* ESCs. The *Erk* KO effect in ESCs is revealed by comparing *iErk1*; *Erk KO* ESCs with or without Dox for 48 h (P0 and P1, respectively).
Consent	N/A
Sample source location	Tianjin, China

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE70304

2. Experimental design

2.1. Cell lines

KH2 mouse embryonic stem cell (ESC) line is used as wild type cells [1]. *iErk1*; *Erk KO* cell lines are derived from KH2 ESCs as described previously [2]. Briefly, two endogenous *Erk1* alleles are disrupted by TALENs. Next, an exogenous *Erk1* gene is integrated into the engineered *ColA1* locus through FLPe recombinase-mediated recombination, resulting in a doxycycline (Dox)-inducible *Erk1* transgene (*iErk1*; *Erk1*$^{-/-}$ ESCs). In the presence of Dox, both endogenous *Erk2* alleles in *iErk1*; *Erk1*$^{-/-}$ cells are disrupted by Cas9. The resulting cell line is named *iErk1*; *Erk KO* ESCs. When cultured in medium supplemented with Dox, the exogenous *Erk1* transgene is expressed. Forty-eight hours after Dox withdrawal, no Erk protein can be detected. Thus, *iErk1*; *Erk KO* ESCs cultured in the absence of Dox for longer than 48 h are considered as *Erk KO* ESCs.

2.2. Samples prepared for RNA-sequencing

To gain insights into the role of Erk signaling in Mek inhibition of ESCs, six samples, WT KH2 ESCs treated with or without PD for 48 h (KH2_PD and KH2, respectively), iErk1; Erk KO ESCs cultured with Dox (P0), 48 and 96 h after Dox withdrawal (P1 and P2, respectively),

* Corresponding authors.
E-mail addresses: liulin@nankai.edu.cn (L. Liu), lingyichen@nankai.edu.cn (L. Chen).

http://dx.doi.org/10.1016/j.gdata.2015.11.029

Table 1
Summary of the sequencing and mapping result.

Sample ID	Total reads	Total base pairs	Total mapped reads	Perfect match	≤2 bp mismatch	Unique match	Multi-position match	Total unmapped reads
P0	11,451,219 (100.00%)	561,109,731 (100.00%)	9,917,796 (86.61%)	8,844,577 (77.24%)	1,073,219 (9.37%)	9,420,231 (82.26%)	497,565 (4.35%)	1,533,423 (13.39%)
P1	11,817,535 (100.00%)	579,059,215 (100.00%)	10,238,164 (86.64%)	9,151,544 (77.44%)	1,086,620 (9.19%)	9,716,556 (82.22%)	521,608 (4.41%)	1,579,371 (13.36%)
P2	12,089,054 (100.00%)	592,363,646 (100.00%)	10,535,239 (87.15%)	9,321,739 (77.11%)	1,213,500 (10.04%)	9,969,006 (82.46%)	566,233 (4.68%)	1,553,815 (12.85%)
P2_PD	11,900,370 (100.00%)	583,118,130 (100.00%)	10,431,854 (87.66%)	9,250,291 (77.73%)	1,181,563 (9.93%)	9,845,925 (82.74%)	585,929 (4.92%)	1,468,516 (12.34%)
KH2	11,962,550 (100.00%)	586,164,950 (100.00%)	9,761,489 (81.60%)	8,473,892 (70.84%)	1,287,597 (10.76%)	9,105,242 (76.11%)	656,247 (5.49%)	2,201,061 (18.40%)
KH2_PD	11,997,433 (100.00%)	587,874,217 (100.00%)	9,709,863 (80.93%)	8,410,837 (70.11%)	1,299,026 (10.83%)	9,067,108 (75.58%)	642,755 (5.36%)	2,287,570 (19.07%)

and iErk1; Erk KO ESCs cultured without Dox for 96 h and treated with PD in the last 48 h (P2_PD), are subjected to RNA sequencing (RNA-seq) analysis.

Total RNA is extracted from cells using RNeasy Mini Kit (Qiagen). On-column DNase I digestion (DNase-Free DNase Set, Qiagen) is performed according to manufacturer protocols to eliminate genomic DNA contamination. Then the mRNA is enriched with the oligo(dT) magnetic beads (for eukaryotes), and is fragmented into short fragments (about 200 bp). With random hexamer-primer, the first strand of cDNA is synthesized, and then the second strand is synthesized. The double strand cDNA is purified with magnetic beads. The ends of the double strand cDNA are repaired, and a single nucleotide A (adenine) is added to the 3′-ends. Finally, sequencing adaptors are ligated to the fragments. The ligation products are amplified with PCR. For quality control, RNA and library preparation integrity are verified using Agilent 2100 BioAnalyzer system and ABI StepOnePlus Real-Time PCR System. Standard barcoded RNA-seq libraries are generated for sequencing with Illumina HiSeqTM 2000 (SE50). Construction of RNA sequencing library, sequencing with Illumina HiSeqTM 2000, and bioinformatic analysis are performed by BGI Tech (BGI, Shenzhen, China).

2.3. Data analysis

By base calling, the original image data produced by the sequencer is translated into sequences, which are defined as “raw reads”(or “raw data”) and saved as “.fastq” files. Raw data is filtered to remove sequences of low quality, and only high quality reads are retained as the clean reads (clean data) and used for the subsequent analysis. The filtering procedure includes the following three steps: (a) remove reads with adaptor sequences. (b) Remove reads in which the percentage of unknown bases (N) is greater than 10%. (c) Remove low quality reads. If the percentage of the low quality base (base with quality value ≤5) is greater than 50% in a read, we define this read as low quality. Clean reads are mapped to reference sequences (mm9) using SOAPaligner/SOAP2 [3]. No more than 2 mismatches are allowed in the alignment. The sequencing and alignment results are summarized in Table 1. Reads Per kilobase per Million reads (RPKM) is calculated to represent the gene expression level, and saved as “.Gene.rpkm.xls” files. The formula is $RPKM = 10^6 \times C/(N \times L/10^3)$ [4]. Here RPKM(A) is the expression level of gene A, C is the number of reads that uniquely aligned to gene A, N is the total number of reads that uniquely aligned to all genes, and L is the number of bases of gene A. If there is more than one transcript for a gene, the longest one is used to calculate its expression level and coverage.

2.4. Analysis of differentially expressed genes

The RPKM values are used for comparing the difference of gene expression among samples. The criteria of more than two fold change and false discovery rate (FDR) <0.001 are used to identify lists of differentially expressed genes (DEGs) in each compared group. Three lists of DEGs, Meki (*WT*), Meki (*Erk KO*), and *Erk* KO, are generated by paired comparisons of KH2_PD and KH2, P2_PD and P2, and P1 and P0, respectively (Fig. 1A). Further comparison of these three lists of DEGs reveals that *Erk* KO are quite different from Meki (*WT*) (Fig. 1B), consistent with our observation that *Erk* KO and Mek inhibition have distinct

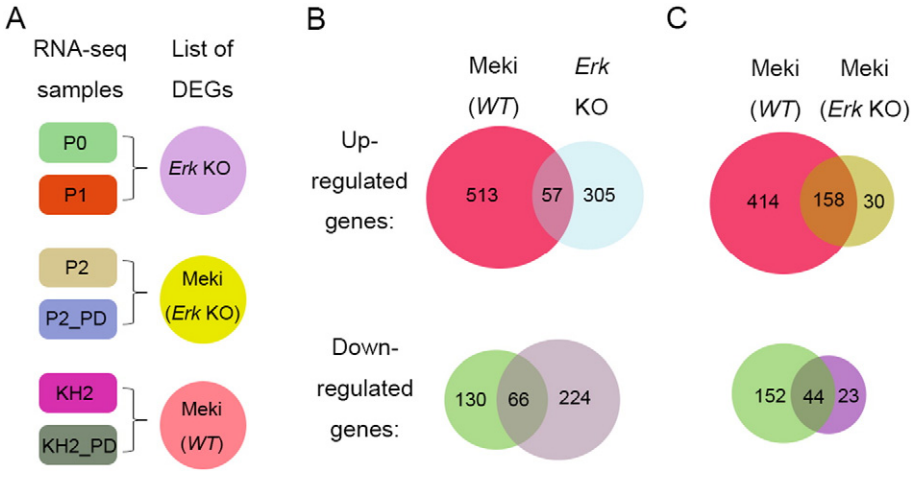

Fig. 1. RNA-seq data reveal an Erk-independent function of Mek. (A) Paired comparisons of KH2_PD and KH2, P2_PD and P2, and P1 and P0, generate three lists of DEGs, Meki (*WT*), Meki (*Erk KO*), and *Erk* KO. (B) The DEG lists of Meki (*WT*) and *Erk* KO are quite different. (C) There are significant overlaps between the DEG lists of Meki (*WT*) and Meki (*Erk KO*), suggesting Mek inhibition still regulates the overlapping genes even in the absence of Erk.

effect on the self-renewal and proliferation of mouse ESCs [2]. Moreover, Meki (*Erk KO*) and Meki (*WT*) share a significant portion of genes (Fig. 1C), suggesting that these genes are regulated by Mek inhibition, regardless of the presence or absence of Erk.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (grant nos. 31271547, 31470081 and 31271587), the Natural Science Foundation of Tianjin, China (no. 14JCYBJC23600), the Program for New Century Excellent Talents (NCET-13-0293), PCSIRT (no. IRT13023), China MOST National Key Basic Research Program (2012CB911202), and the 111 Project Grant (B08011).

References

[1] C. Beard, et al., Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. Genesis 44 (1) (2006) 23–28.
[2] H. Chen, et al., Erk signaling is indispensable for genomic stability and self-renewal of mouse embryonic stem cells. Proc. Natl. Acad. Sci. U. S. A. 112 (44) (2015) E5936–E5943.
[3] R. Li, et al., SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25 (15) (2009) 1966–1967.
[4] A. Mortazavi, et al., Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5 (7) (2008) 621–628.