nature protocols

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41596-019-0172-4

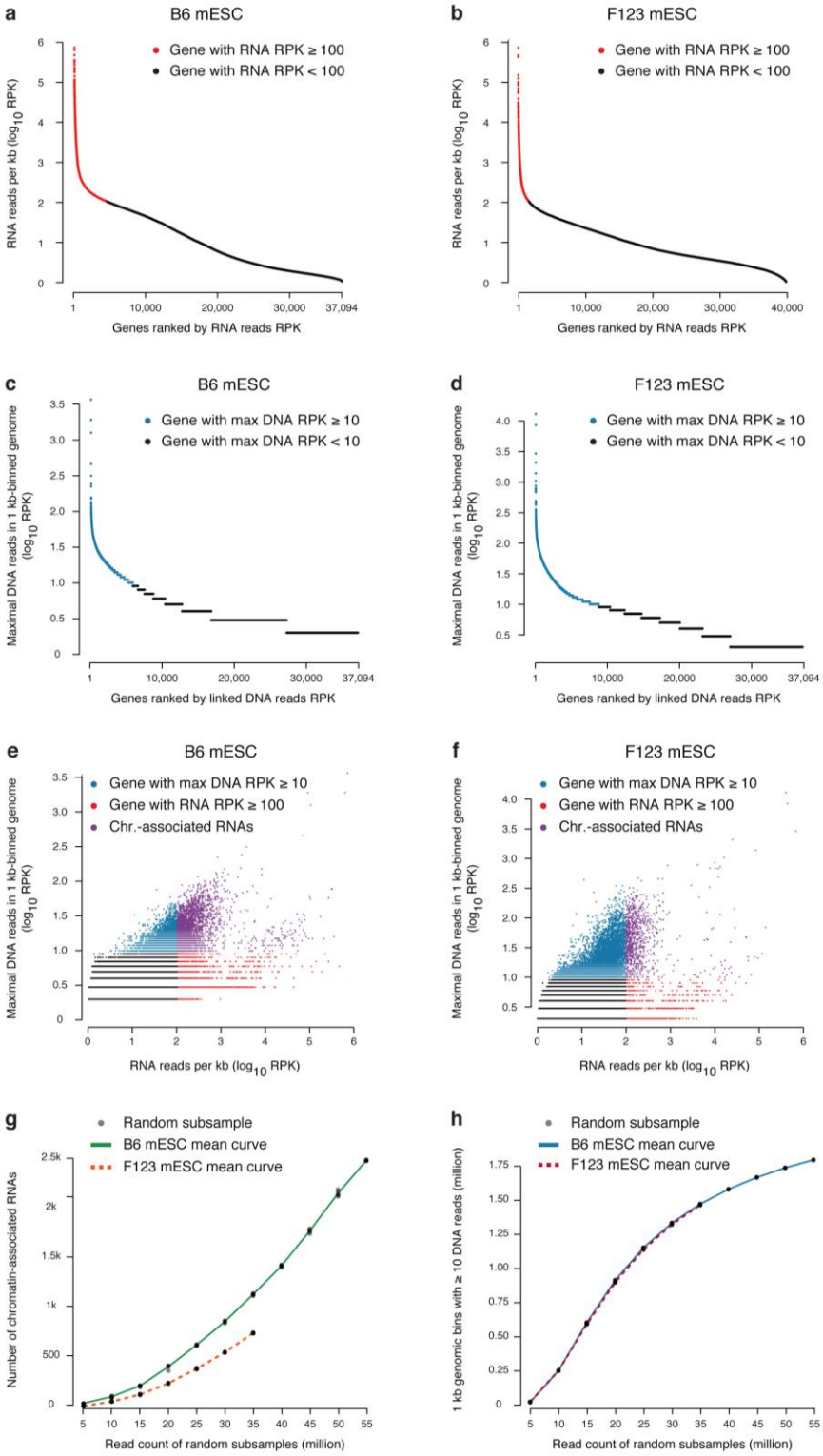
In the format provided by the authors and unedited.

GRID-seq for comprehensive analysis of global RNA–chromatin interactions

Bing Zhou[1,2,3,10], Xiao Li[3,4,10*], Daji Luo[3,5,6,10], Do-Hwan Lim[3], Yu Zhou[7,8] and Xiang-Dong Fu[3,9*]

[1]State Key Laboratory of Stem Cell and Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China. [2]Institute for Stem Cell and Regeneration, Chinese Academy of Sciences, Beijing, China. [3]Department of Cellular and Molecular Medicine, University of California, San Diego, La Jolla, CA, USA. [4]BGI-Shenzhen, Shenzhen, China. [5]School of Basic Medical Sciences, Wuhan University, Wuhan, China. [6]Hubei Provincial Key Laboratory of Developmentally Originated Disease, Wuhan University, Wuhan, China. [7]State Key Laboratory of Virology, Hubei Key Laboratory of Cell Homeostasis, College of Life Sciences, Wuhan University, Wuhan, China. [8]Institute for Advanced Studies, Wuhan University, Wuhan, China. [9]Institute of Genomic Medicine, University of California, San Diego, La Jolla, CA, USA. [10]These authors contributed equally: Bing Zhou, Xiao Li, Daji Luo. *e-mail: x4li@ucsd.edu; xdfu@ucsd.edu

Supplementary Figure 1

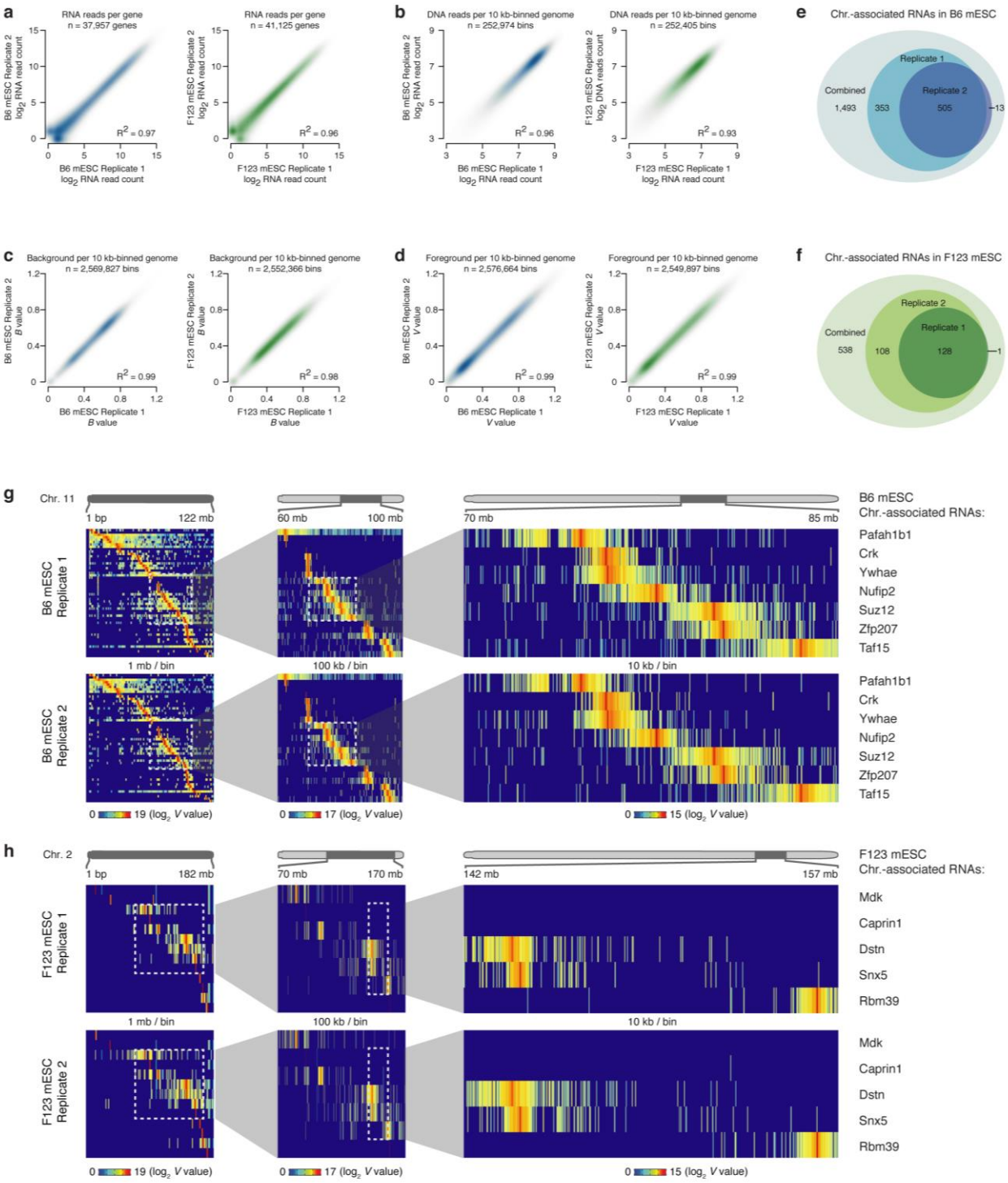


Supplementary Figure 1

Identification of chromatin-associated RNAs

(a) All mouse genes were ranked in a descending fashion by their normalized RNA read densities captured by GRID-seq in B6 mESC datasets. Each dot represents one gene and red ones highlight genes with the RNA read density of >100 RPK. RPK: reads per kb. **(b)** Similar plot based on F123 mESC datasets as in **a**. **(c)** All genes were ranked in a descending fashion by the maximal linked DNA read density on all genomic bins in B6 mESC datasets. Each dot represents one gene and blue ones emphasize genes with DNA read density of >10 RPK. **(d)** Similar plot based on F123 mESC datasets as in **c**. **(e)** Scatter plot showing the distribution of all genes based on their RNA read densities (x-axis) and maximal DNA read densities over genomic bins (y-axis) in B6 mESC datasets. Red dots correspond to genes marked with red dots in **a**; blue dots correspond to genes marked with blue dots in **c**; purple dots represent chromatin-associated RNA that survived both cutoffs. **(f)** Similar plot based on F123 mESC datasets as in **e**. **(g, h)** Saturation analyses performed by random subsampling 5 to 55 million uniquely mapped read mates in B6 and 5 to 35 million uniquely mapped read mates in F123 mESC datasets. Each level of random subsampling was independently repeated 5 times. Numbers of chromatin-associated RNAs identified were summarized and plotted in **g**. The green line represents the connected mean values in B6, and the orange dotted line represents those in the F123 mESC dataset. Numbers of 1 kb-binned genomic regions that have at least 10 linked DNA reads were summarized and plotted in **h**. The blue line represents the connected mean values in B6, and the red dotted line represents those in the F123 mESC dataset.

Supplementary Figure 2



Supplementary Figure 2

Reproducibility of GRID-seq libraries

(a) Smoothed scatter plots of uniquely mapped RNA read levels for all genes between replicated libraries of B6 (left) and F123 mESCs (right). **(b)** Smoothed scatter plots of uniquely mapped DNA read levels for all 10-kb genomic bins between replicated libraries of B6 (left) and F123 mESCs (right). **(c)** Smoothed scatter plots of background signals over all 10-kb genomic bins (*B* values) between replicated libraries of B6 (left) and F123 mESC (right). **(d)** Smoothed scatter plots of foreground signals over all 10-kb genomic bins (*V* values) between replicated libraries in B6 (left) and F123 mESC (right). **(e)** Overlapping of chromatin-associated RNAs identified in replicated and combined B6 mESC datasets. The numbers on the plot indicate the chromatin-associated RNA species in respectively coloured areas in the Venn diagram (for example: 505 common chromatin-associated RNAs are identified in the combined, replicate 1 and replicate 2 datasets; 13 chromatin-associated RNAs are identified in the combined and the replicate 2 data but not in the replicate 1 data). **(f)** Overlapping of chromatin-associated RNAs identified in replicated and combined F123 mESC datasets, displayed in a similar manner as in **e**. **(g)** RNA-chromatin interactome on chromosome 11 independently constructed from B6 mESC replicates. Boxed regions in each panel were enlarged with increasing resolution in the next panel on the right. A set of representative chromatin-associated RNAs are labeled on the right. **(h)** Interactome on chromosome 2 independently constructed from F123 mESC replicates similar to **g**.