Supplementary Information for

FAM46B is a prokaryotic-like cytoplasmic poly(A) polymerase essential in human embryonic stem cells

Jia-Li Hu[1,8,†], He Liang[2,†], Hong Zhang[1,†], Ming-Zhu Yang[3,†], Wei Sun[4,5,10], Peng Zhang[2], Li Luo[1], Jian-Xiong Feng[1], Huajun Bai[2], Fang Liu[3], Tianpeng Zhang[6], Jin-Yu Yang[1], Qingsong Gao[5], Yongkang Long[4], Xiao-Yan Ma[1], Yang Chen[1], Qian Zhong[1], Bing Yu[1], Shuang Liao[1], Yongbo Wang[7], Yong Zhao[6], Mu-Sheng Zeng[1], Nan Cao[3], Jichang Wang[3], Wei Chen[4,*], Huang-Tian Yang[2,*], Song Gao[1,9,*]

1. State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, 510060, China.

2. CAS Key Laboratory of Tissue Microenvironment and Tumor, Laboratory of Molecular Cardiology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, 200031, China.

3. MOE Key Laboratory for Stem Cells and Tissue Engineering, Department of Histology and Embryology, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, 510080, China

4. Department of Biology, Southern University of Science and Technology, Shenzhen, 518055, P.R. China

5. Laboratory for Functional Genomics and Systems Biology, The Berlin Institute for Medical Systems Biology, 13092 Berlin, Germany

6. MOE Key Laboratory of Gene Function and Regulation, State Key Laboratory for Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, 510006, China.

7. Department of Cellular and Genetic Medicine, School of Basic Medical Sciences, Fudan University, Shanghai, 200032, China.

8. Department of Oncology, The Second Affiliated Hospital of Nanchang University, Nanchang, Jiangxi, 330006, P.R. China.

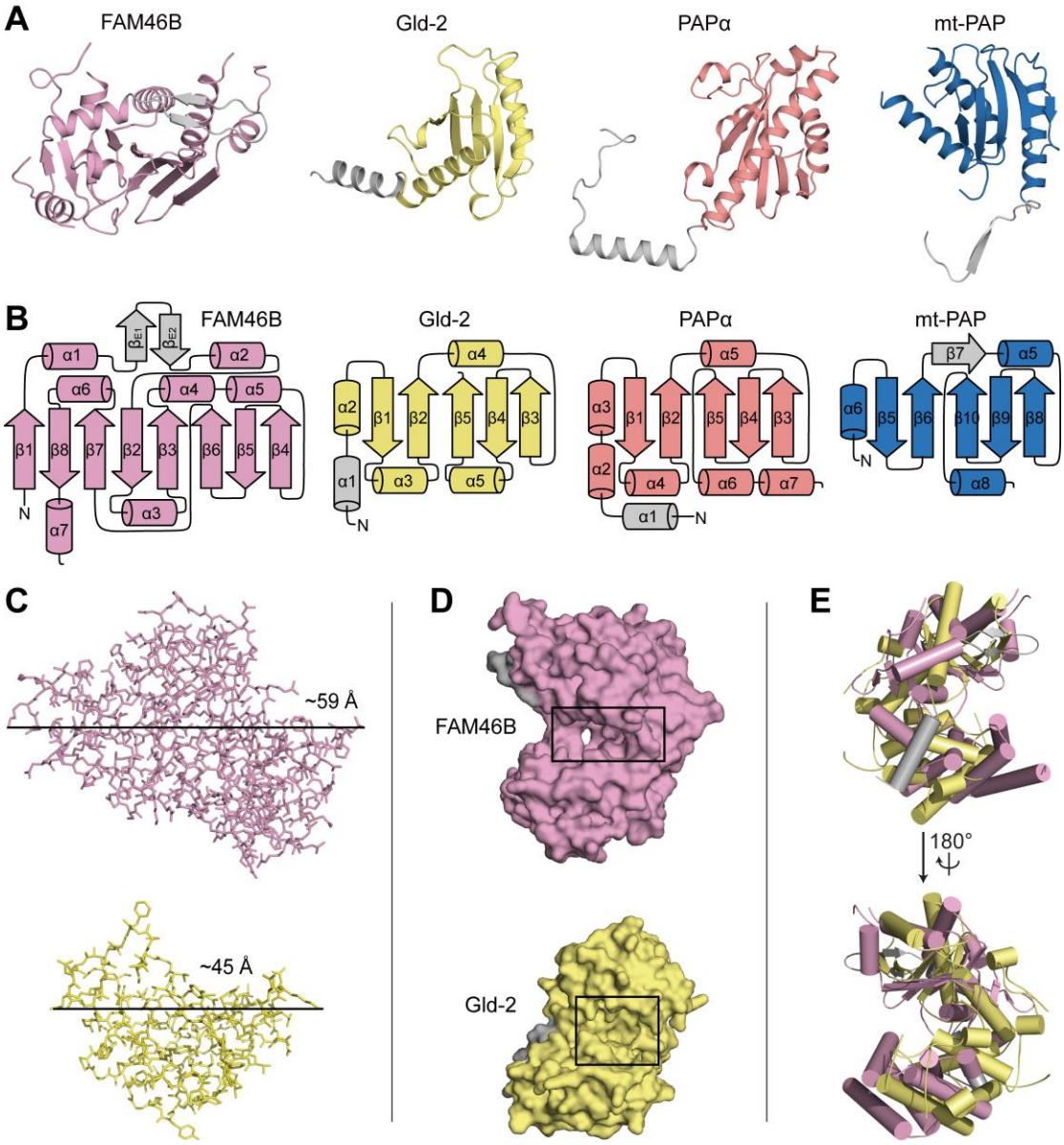
9. Guangzhou Regenerative Medicine and Health Guangdong Laboratory, Guangzhou, 510530, China.

10. Current address: Department of Pharmaceutical Chemistry and Cardiovascular Research Institute, University of California San Francisco, CA 94158, USA

* To whom correspondence should be addressed. Tel: +86 02087343168; Fax: +86 02087343180; Email: gaosong@sysucc.org.cn. Correspondence may also be addressed to Huang-Tian Yang. Tel: + 86 021 54923280; Fax: + 86 021 54923280; Email: htyang@sibs.ac.cn. Correspondence may also be addressed to Wei Chen. Tel: + 86 0755 88010849; Email: chenw@sustech.edu.cn.

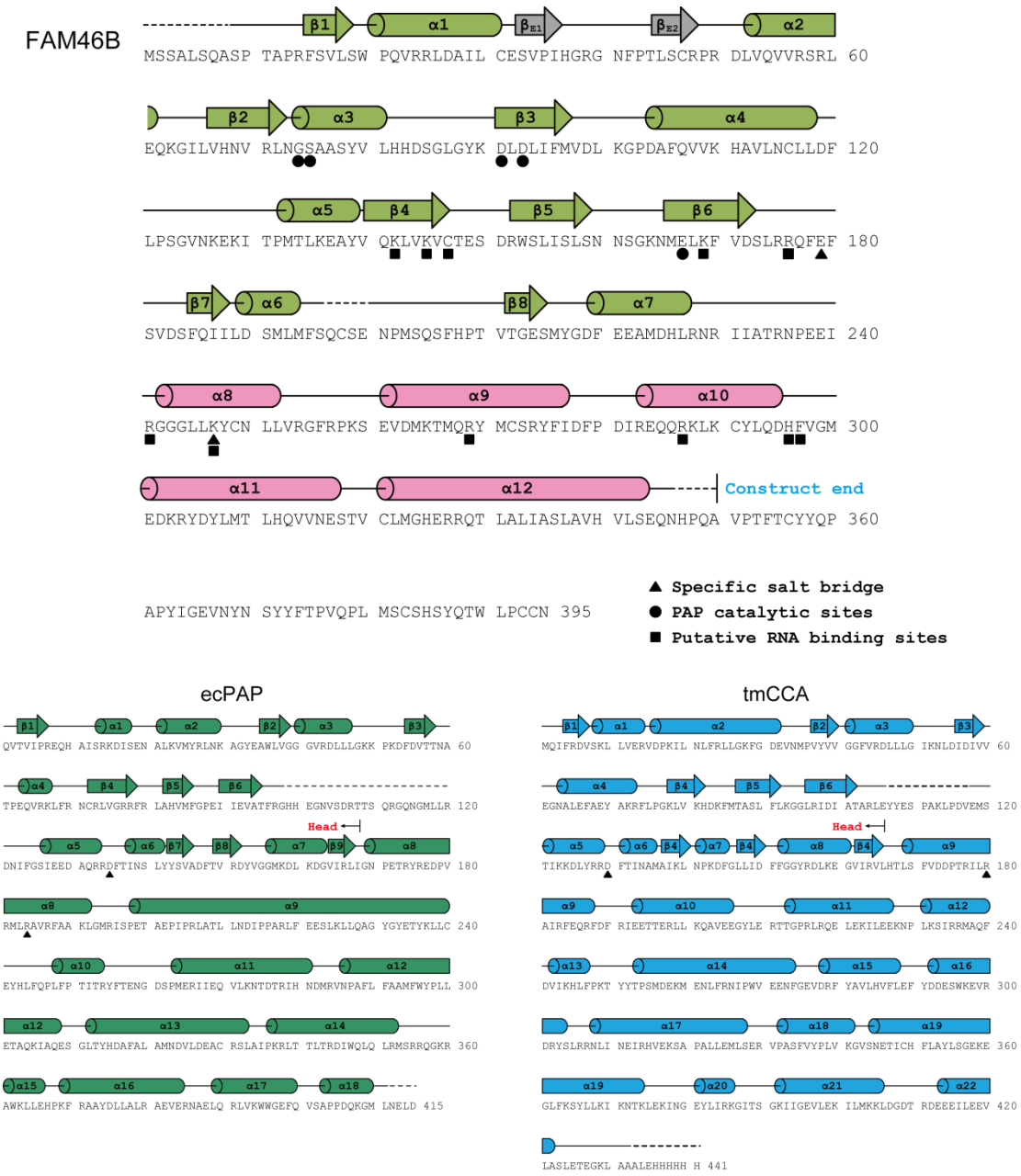
† The authors wish it to be known that, in their opinion, the first four authors should be regarded as Joint First Authors.

Supplementary Figures 1-8



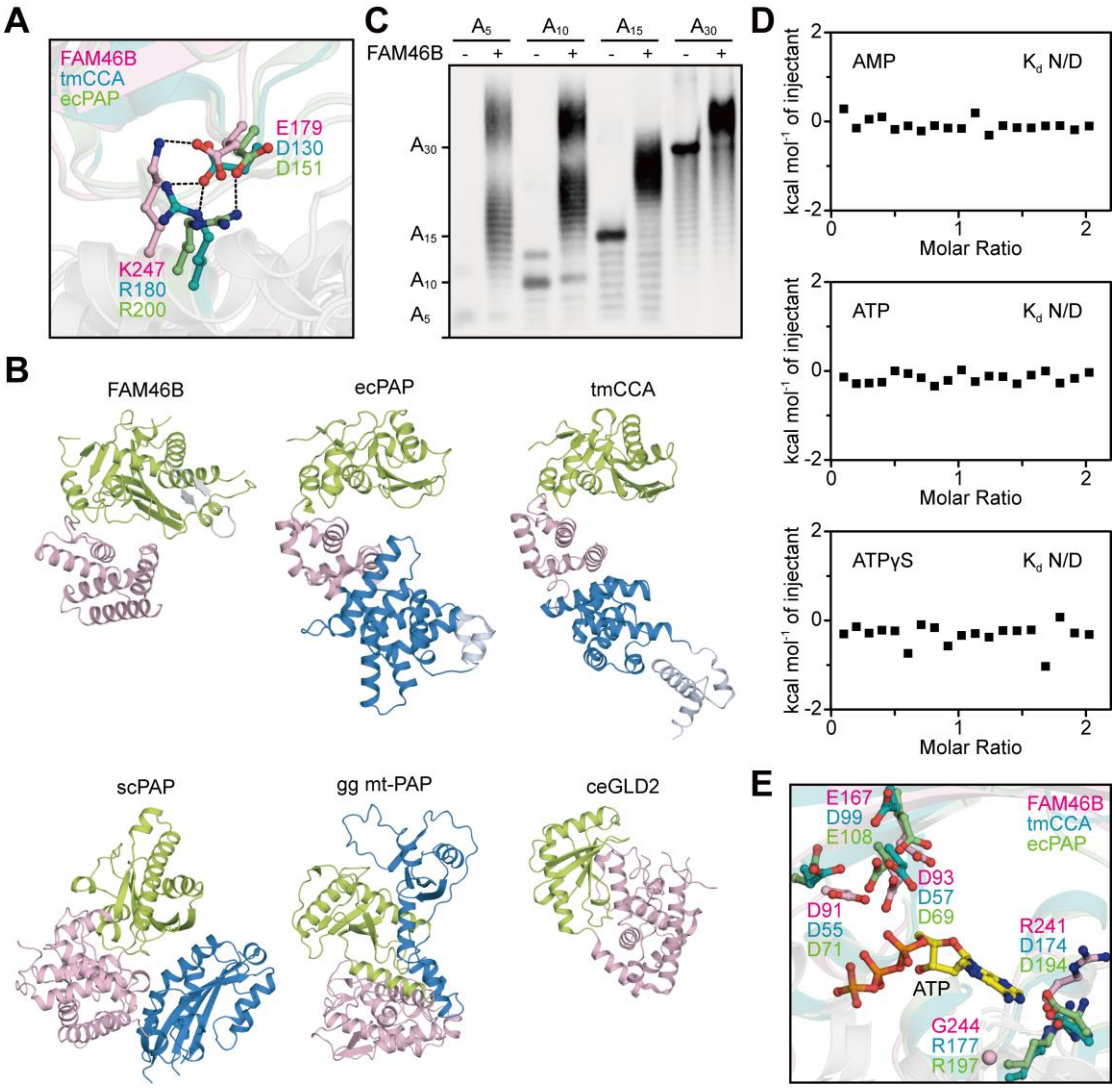
Supplementary Figure 1, Structure comparison between FAM46B and other eukaryotic PAPs

(A) Structural comparison between the FAM46B NCD and the catalytic domains of *Caenorhabditis elegans* (ce)Gld-2 (Protein Data Bank code 4ZRL), *Saccharomyces cerevisiae* (sc)PAPα (2Q66) and *Gallus gallus* (gg) mitochondrial (mt-)PAP (5A2V). The FAM46B featured β-hairpin, and the extra helices of ecGld-2 and scPAPα are colored gray.
(B) Structural topology of the catalytic domains of FAM46B, Gld-2, PAPα and mt-PAP. Color as in **A**.
(C) The difference between the widths of FAM46B NCD and the catalytic domain of Gld-2.
(D) Surface representation showing the central clefts of FAM46B and Gld-2.
(E) Comparison between FAM46B and Gld-2 in the relative positions of NCD and HD.



Supplementary Figure 2, Secondary structure topology of FAM46B and its prokaryotic homologs

Comparison in secondary structure topology for FAM46B, ecPAP (3AQM) and tmCCA (3H39). α-helices are shown as cylinders and β-strands as arrows for three proteins above the sequences. The secondary structure signs are colored and labeled as in Fig 1a and Fig 3b. Regions not resolved in the crystal structure are indicated by dashed lines. The end of the crystallized xtFAM46B, as well as the border of the head domain for ecPAP and tmCCA, are indicated. Important residues that are discussed in the paper are specified by different symbols according to their structural and functional roles.



Supplementary Figure 3, Structure comparison between FAM46B and other CCA/PAPs

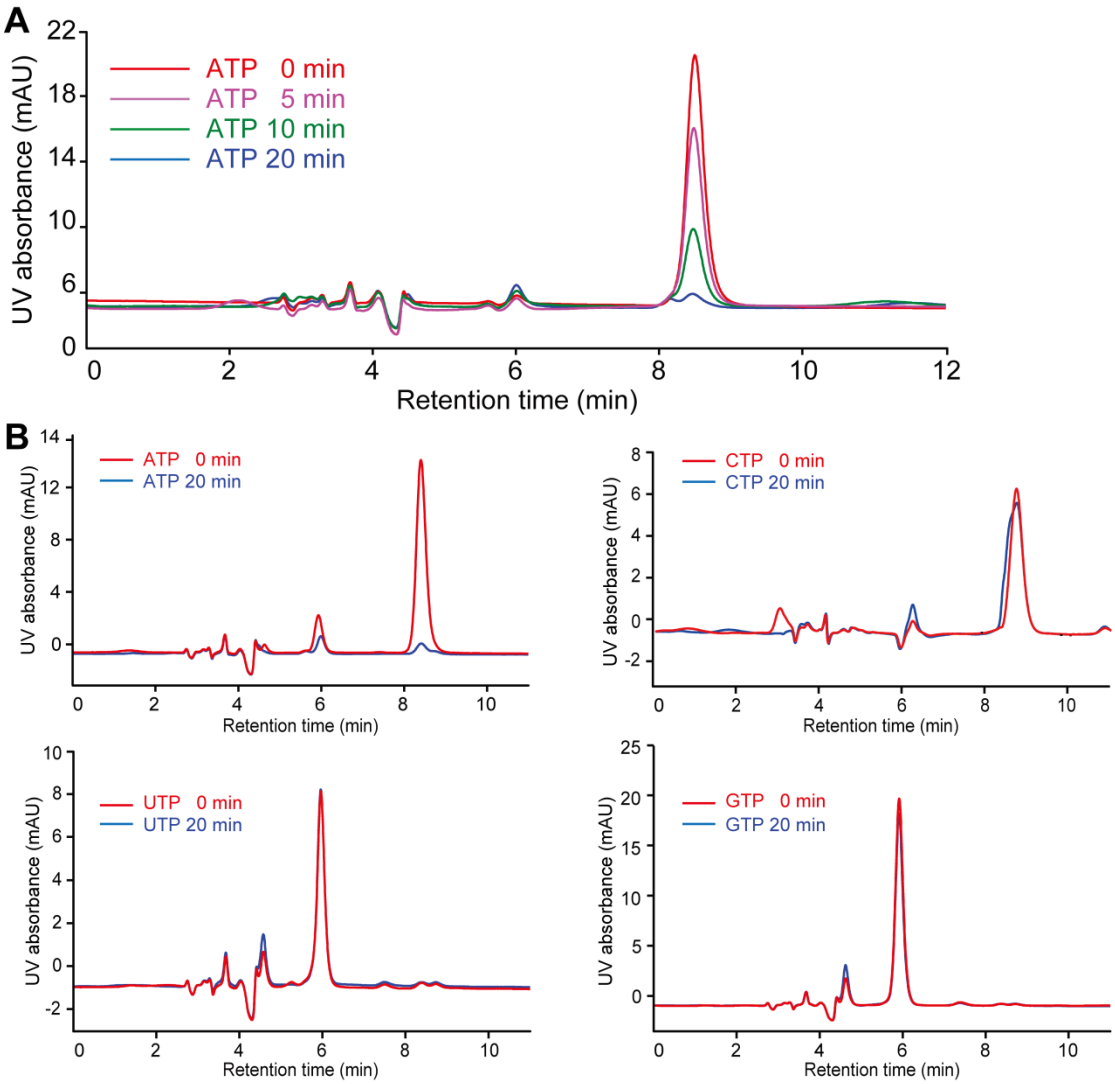
(A) The conserved inter-domain salt bridges for FAM46B (magenta), ecPAP (3AQM, pale cyan) and tmCCA (3H39, teal).

(B) FAM46B, two bacterial PAP/CCA (ecPAP and tmCCA), and three eukaryotic PAPs (PAPα, 2Q66; mt-PAP, 5A2V; Gld-2, 4ZRL) are shown as cartoon representation. The NCD or equivalent domains are shown in lemon, HD and equivalent domains in pink, additional domains in marine and light blue.

(C) Polyadenylation activity of human FAM46B on oligoA substrates with different lengths.

(D) Binding affinities (dissociation constant, K_d) to AMP, ATP and ATPγS for xtFAM46B were measured by ITC. FAM46B lacks binding affinity to these nucleotides. N/D, not determined.

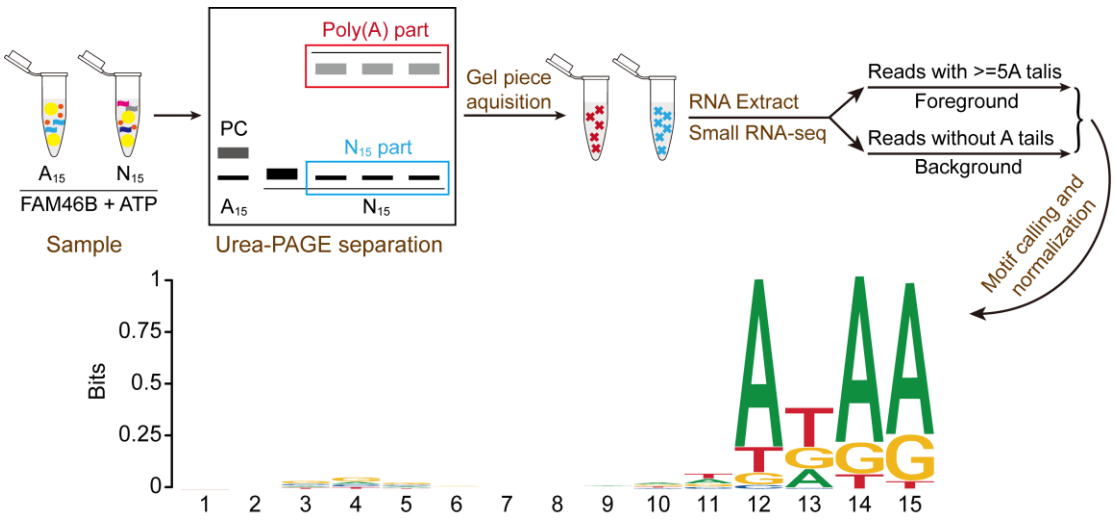
(E) Comparison of putative ATP recognition residues between FAM46B and ecPAP/ tmCCA. The conserved three catalytic residues are also indicated. Color as in **A**.



Supplementary Figure 4, NTP consumption of FAM46B measured by HPLC

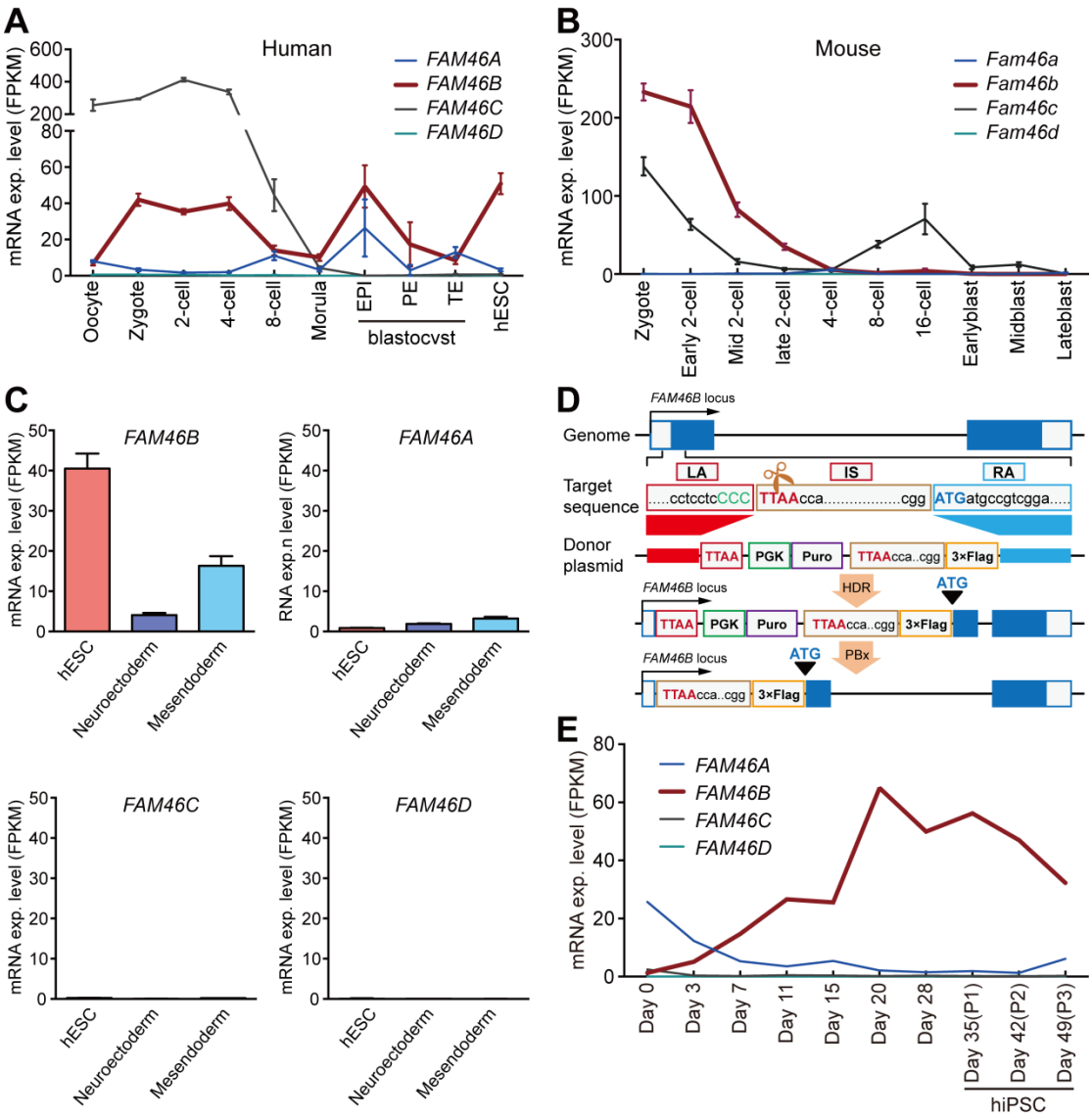
(A) Overlay of HPLC results for ATP consumption experiments at different time points. Note the time-dependent progressive decrease of the ATP peak area, indicating that ATP is being consumed.

(B) The HPLC data of NTP selectivity assay for FAM46B in the presence of A_{15} RNA primer. Only the ATP peak shows sharp shrink, suggesting that FAM46B prefers to consume ATP to extend the A_{15} primer.



Supplementary Figure 5, Experimental process and result of A-tailing assay

Schematic drawing showing experimental design and protocol of in vitro A-tailing assay. After normalization, the position-dependent sequence frequency indicates that FAM46B has an AxAA preference for the last 4 bases of the RNA substrate.

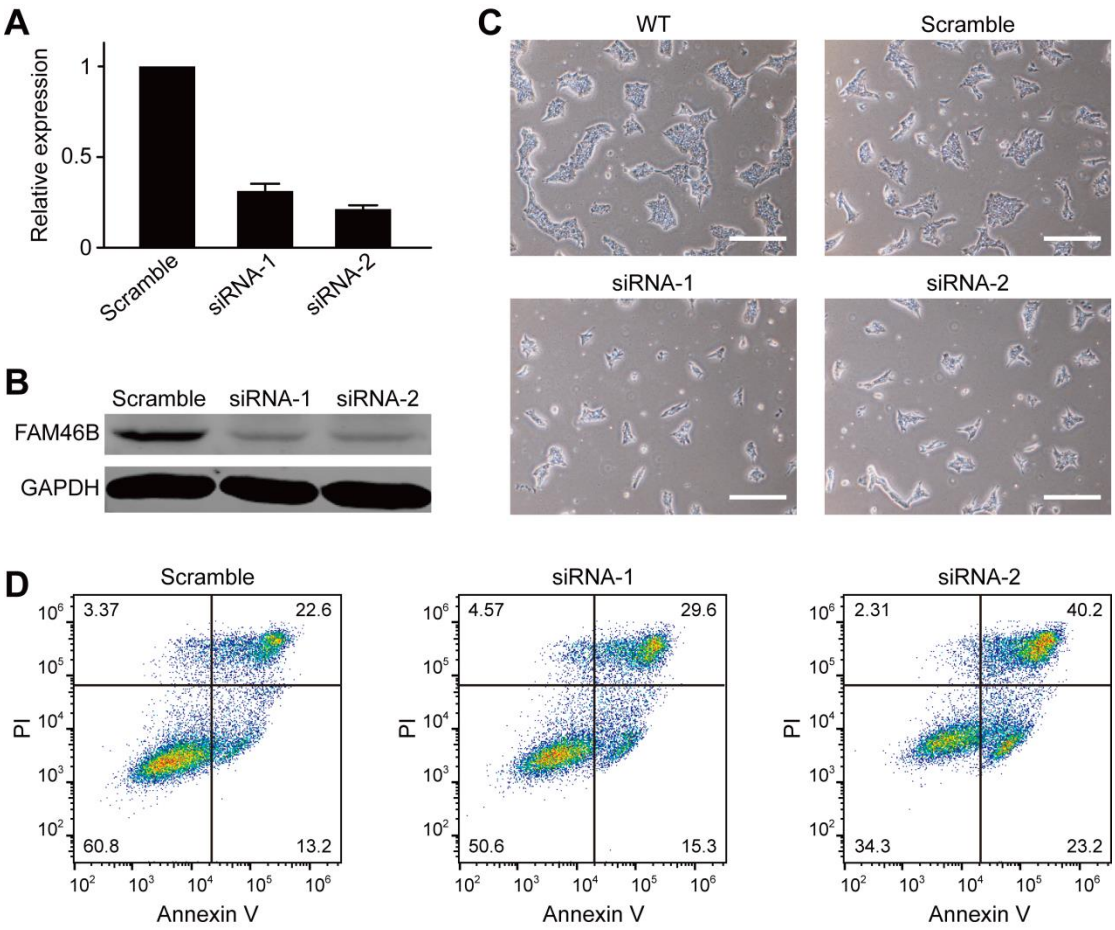


Supplementary Figure 6, Transcription of *FAM46* genes in early embryogenesis and pluripotent cell lines

(A) Transcriptional dynamics of *FAM46* genes during human early embryogenesis and cultured hESCs. RNA expression level was shown as FPKM (fragment per kilobase per million reads). RNA-seq data from public dataset was analyzed (49).
(B) Transcriptional dynamics of *FAM46* genes during mouse early embryogenesis. RNA-seq data from public dataset was analyzed (50).
(C) Transcription of *FAM46* genes during neuroectoderm and mesendoderm differentiation from hESCs. Only *FAM46B* mRNA level shows sharp decrease. *FAM46A* transcription is slightly increased whereas *FAM46C* and *FAM46D* are both undetectable. The RNA-seq data from public dataset was analyzed (51).
(D) Schematic drawing showing the design and process of inserting a 5'- 3×Flag to the *FAM46B* locus. The two-step strategy includes (i) integration of PGK-Puro-IS-3×Flag in the 5' flanking region of the start codon, and (ii) excision of the sequence between TTAA by using an excision-only PiggyBac Transposase PBx,

leaving behind 3×Flag sequence that is linked to the start codon of *FAM46B*. LA denotes left recombinant arm; IS, intermediate sequence; RA, right recombinant arm. The PiggyBac insertion site (TTAA) is shown in red and the start codon in pale-blue. Scissor stands for the CRISPR cutting site.

(E) Transcriptional dynamics of *FAM46* genes during human somatic cell reprogramming into induced pluripotent stem cells. The RNA-seq data from public dataset was analyzed (52). RNA expression level was shown as FPKM.



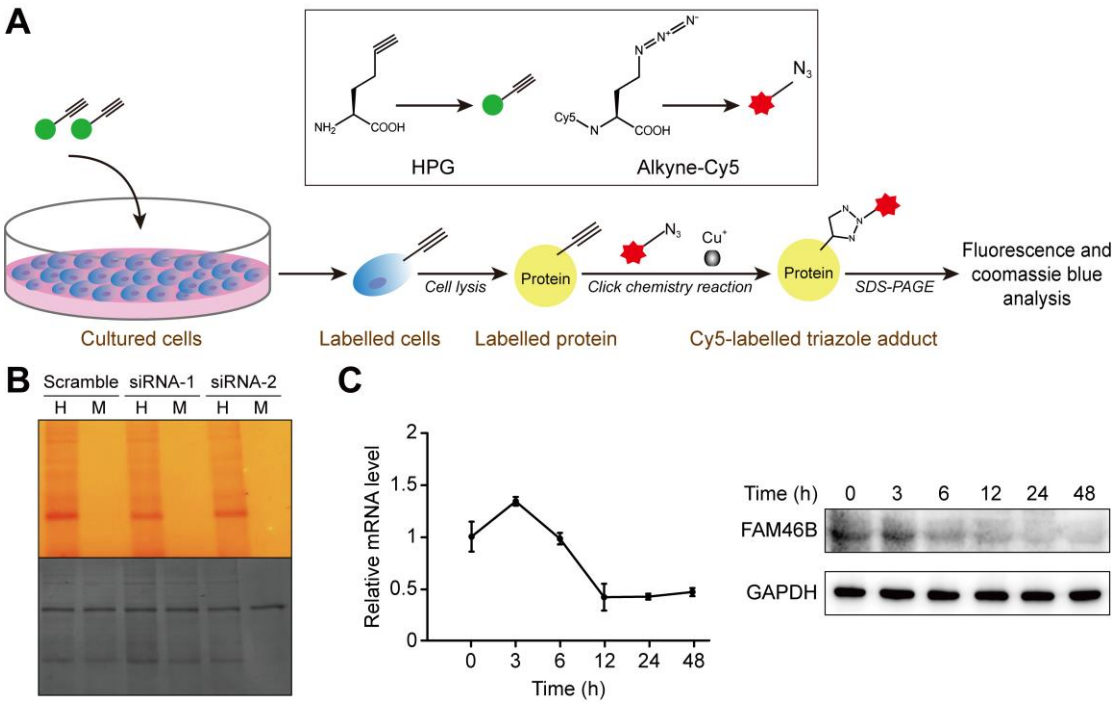
Supplementary Figure 7, *FAM46B* knock-down results in hESC apoptosis

(A) qPCR results showing the decreased *FAM46B* mRNA levels 24 hours after siRNA treatment.

(B) Western blot showing the decreased FAM46B protein levels 48 hours after siRNA treatment.

(C) FAM46B knock-down led to hESC death as observed 72 hours after transfection of siRNAs. Scale bar, 100 μm.

(D) Flow-cytometry result showing that *FAM46B* knock-down promoted hESC apoptosis.



Supplementary Figure 8, Experimental procedure of the FUNCAT assay

(A) Illustration of the FUNCAT assay for checking the newly synthesized protein. The cells are pulsed with methionine-substituting HPG after starvation. The newly synthesized proteins are crosslinked with fluorescent dye via click chemistry reaction, and then chased by fluorescence.

(B) FUNCAT assay showing knock-down of FAM46B inhibits protein synthesis. Left: Knock-down of FAM46B reduced total HPG labeled newly synthesized protein of hESCs. Right: Coomassie blue staining of the same gel after fluorescent imaging showing that the amount of total protein for each sample is consistent. H: HPG; M: Met

(C) The mRNA and protein levels of FAM46B at various time points after induction with doxycycline in Cas9 *FAM46B*-knock-out hESCs.