

Heritable expansion of the genetic code in mouse and zebrafish

Cell Research (2017) 27:294–297. doi:10.1038/cr.2016.145; published online 9 December 2016

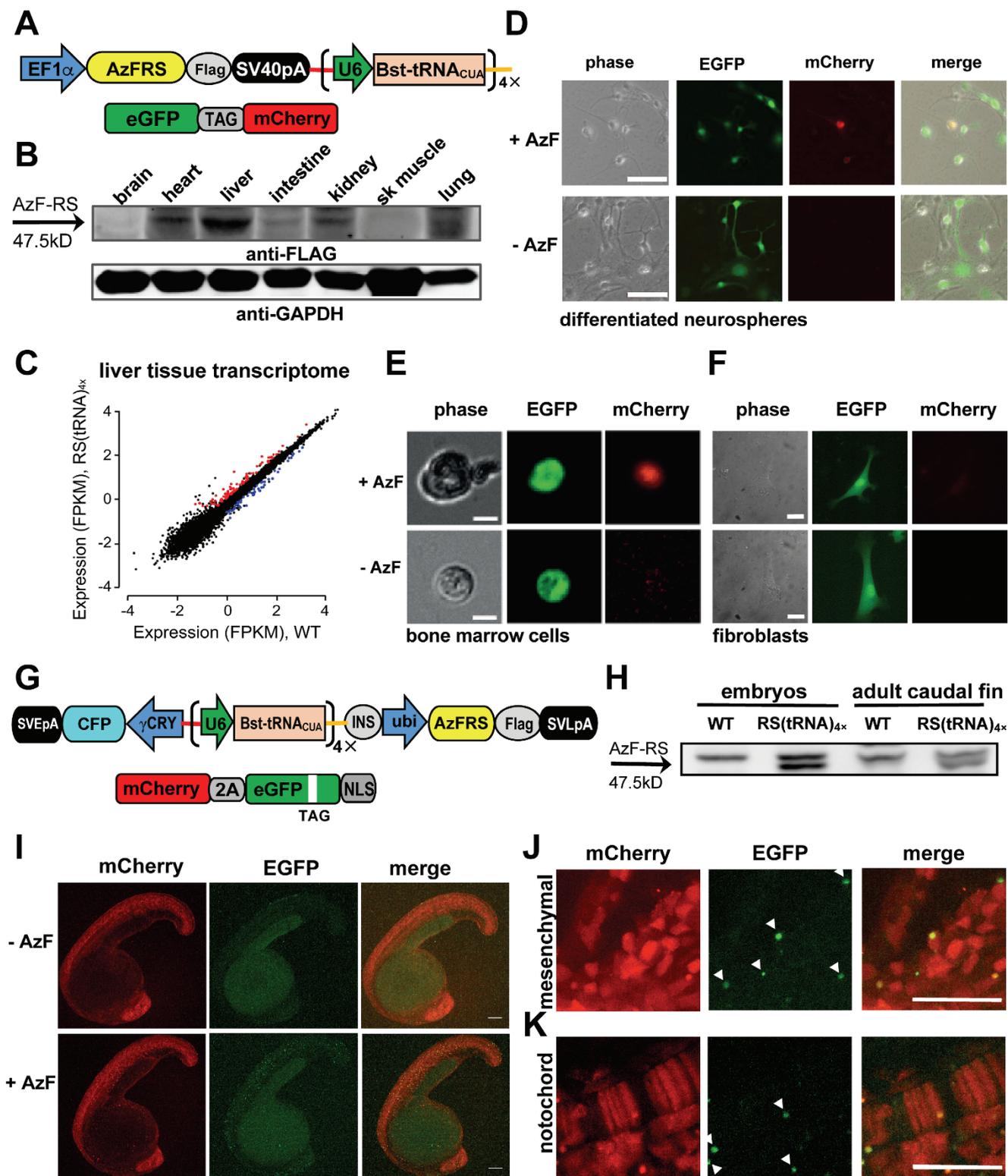
Dear Editor,

The genetic code can be expanded to encode unnatural amino acids (Uaas) by introducing an orthogonal tRNA/aminoacyl-tRNA synthetase (aaRS) pair to decode a stop codon. Initially developed in *Escherichia coli* [1], this strategy has been proven generally applicable in eukaryotic cells [2] and in generating transgenic invertebrates capable of Uaa incorporation, including *Caenorhabditis elegans* [3] and *Drosophila melanogaster* [4]. Although cells and tissues of mouse can be transduced transiently to incorporate Uaas [5], it remains unknown whether transgenic vertebrates with a heritable expanded genetic code system-wide can be generated, as it is unclear whether the biological complexity of vertebrates allows the introduction, maintenance and transmission of the newly introduced genetic material for code expansion. Here we for the first time report the generation of transgenic mice and zebrafish with an expanded genetic code, providing valuable vertebrate model animals for biological and biomedical research.

Laboratory mouse (*Mus musculus*) is a widely used mammalian model for developmental, physiological and pathological studies due to its high genetic homology with humans. We decided to integrate the orthogonal tRNA_{CUA}/AzFRS pair, which can decode the amber stop codon UAG as the Uaa *p*-azido-phenylalanine (AzF) [6], into the mouse genome. A construct containing FLAG-tagged AzFRS driven by the human elongation factor 1 alpha (EF1 α) promoter followed by four copies of the U6 promoter-driven tRNA was selected after optimization of AzF incorporation efficiency in mammalian cells (Figure 1A). This RS(tRNA)_{4x} construct was linearized and directly delivered into nuclei of 190 mouse zygotes, yielding 36 F0 pups, four of which showed positive genotyping results (Supplementary information, Figure S1A). RT-PCR revealed the presence of AzFRS mRNA in tail-tip fibroblasts isolated from two founders, F0-1 and F0-10 (Supplementary information, Figure S1B). The F0-1 colony was maintained by successive backcrossing with WT C57BL/6 mice. The resulting offspring were viable

and indistinguishable from WT littermates, and in general these crossing produced normal litter sizes along with a Mendelian inheritance pattern. The RS(tRNA)_{4x} gene was transmitted efficiently to subsequent generations (F1, F2 and F3; Supplementary information, Table S1A). On average 15 copies of transgenes was detected in an F2 mouse as determined by qPCR using GAPDH as the internal control (Supplementary information, Table S1B), indicating that the transgene was integrated into the F0 genome and stably transmitted. The expression of AzFRS proteins was detected in various tissues in F2 mice by western blotting (Figure 1B).

We next evaluated the potential impact of gene integration in the living mice. F2 transgenic mice and descendants were used in subsequent characterization. Through histology analysis, no detectable morphological differences were observed between transgenic mice and the littermate controls from analyzed tissues including brain, heart, liver, colon, kidney, skeletal muscle and lung (Supplementary information, Figure S1C). To further assess effects of the transgene on gene expression, we used RNA-seq to analyze the transcriptomes of the liver tissue, where AzFRS showed the highest expression (Figure 1B). In comparison with the WT mice, we identified minor changes in 97 upregulated and 47 downregulated transcripts ($P < 0.05$; fold change > 2) in the RS(tRNA)_{4x} transgenic mice (Figure 1C). Among those liver-specific genes, the upregulated genes fell into a variety of categories related to metabolism (Supplementary information, Figure S1D). Among the downregulated genes, there was a slight enrichment of primary metabolic enzymes associated with insulin signaling processes, suggesting that the transgene may have a repressive effect on primary metabolism in liver. To test whether these small expression changes impaired the liver function in the transgenic line, we measured blood chemical indexes (Supplementary information, Figure S1E). No significant difference was observed between WT and transgenic mice in levels of blood aspartate transaminase (AST), alanine transaminase (ALT), albumin and other indexes. In addition, we fed the transgenic mice with or without 30 mg/ml AzF



for 10 days, and no obvious defects (data not shown) and abnormalities in blood chemical indexes were observed (Supplementary information, Figure S1F). Taken together,

these results demonstrate that genome integration of RS(tRNA)_{4x} gene in mice did not cause significant physiological impairment.

To functionally evaluate the integrated tRNA_{CUA}/AzFRS pair, AzF incorporation experiments were performed in adult transgenic mice-derived primary cells with transduction of an eGFP-amb-mCherry reporter, in which eGFP is fused to mCherry linked by the amber stop codon [7] (Figure 1A). Neurospheres isolated from the subventricular zone of transgenic mice were induced to differentiate into neurons and simultaneously transduced with the lentivirus carrying the eGFP-amb-mCherry reporter (Figure 1D). eGFP fluorescence appeared between day 4 and day 5 post transduction and differentiation, indicating the successful transduction of the neurons. Although we detected eGFP in neurons with or without AzF, mCherry signals colocalized with eGFP were detected only in the presence of AzF. In addition, through infecting with the eGFP-amb-mCherry reporter and adding AzF, amber suppression was also detected in primary bone marrow cells from transgenic mice (Figure 1E) but surprisingly absent in fibroblasts (Figure 1F and Supplementary information, Figure S1G). However, when we cotransfected additional tRNA_{CUA} with the eGFP-amb-mCherry reporter into fibroblasts derived from transgenic mice, AzF incorporation was detected (data not shown), suggesting that AzFRS was functional and tRNA_{CUA} expression may be limited in fibroblasts. Taken together, these data demonstrate the successful expression of functional tRNA_{CUA}/AzFRS in primary cells isolated from the transgenic mice, which directs AzF incorporation into the reporter in response to the amber stop codon. AzF incorporation was cell type-dependent,

in part influenced by the amount of tRNA_{CUA} expressed.

We also sought to expand the genetic code in zebrafish (*Danio rerio*), which is a popular vertebrate model for live imaging and shares many conserved molecular and cellular mechanisms with mammals. A construct for ubiquitous expression of tRNA_{CUA}/AzFRS (Figure 1G) was co-injected with the mRNA of Tol2 transposase into zebrafish embryos at 1- or 2-cell stage [8] to generate the transgenic fish. AzFRS expression was detected via its C-terminal FLAG tag in both larvae (2 days post fertilization (dpf)) and adult caudal fin in F2 generation (Figure 1H), revealing the successful and stable integration of the transgene. To assess AzF incorporation *in vivo*, we used an mRNA coding for a nuclear-targeted eGFP reporter, in which an amber stop codon is inserted at the position of eGFP-Y145 and the mCherry-coding sequence (as a translational control) followed by a “self-cleaving” 2A peptide is fused to the N-terminus of eGFP (Figure 1G). *In vitro*-transcribed mCherry-eGFP (Y145amb) mRNA was injected into F2 transgenic fish embryo at 1-cell stage, and the embryos were incubated in water containing 2.5 mM AzF, a concentration causing no detectable toxicity as determined in WT and transgenic zebrafish embryos (Supplementary information, Figure S2A). In the presence of AzF, there was a significant increase in the number of cells with eGFP signals 24 h post fertilization (hpf), indicating suppression of the amber stop codon (Figure 1I and Supplementary information, Figure S2B). Images with higher magnification showed that the eGFP signal in the nucleus could be observed in all cell

Figure 1 Generation and characterization of transgenic mice and zebrafish with an expanded genetic code. **(A)** Schematic diagram of the construct used to generate transgenic mice and the reporter construct. The transgene construct contains AzFRS and four copies (4×) of *Bacillus stearothermophilus* (Bst)-tRNA_{CUA}. pA, polyadenylation signal. The eGFP-TAG-mCherry reporter construct was generated by inserting an amber stop codon (TAG) between eGFP and mCherry genes. Verified in HEK293T and mouse embryonic fibroblast (MEF) cells, this reporter expresses mCherry only when tRNA_{CUA}, AzF-RS and AzF are all available to cells. **(B)** Western blot analysis of the indicated tissue samples from transgenic RS(tRNA)_{4x} mice of F2 generation and descendants. Arrow indicates the immunoreactive band of AzFRS-FLAG detected by an anti-FLAG antibody. GAPDH serves as an internal loading control. **(C)** RNA-seq analysis of liver tissues isolated from 6-week-old WT and F2 transgenic RS(tRNA)_{4x} mice. The plot shows whole-transcriptome fragments per kilobase of exon per million fragments mapped (FPKM). Significantly ($P < 0.05$; fold change > 2) up- and downregulated genes in the transgenic mice from three biological replicates are colored red and blue, respectively; other genes are in black. **(D-F)** Genome-integrated orthogonal tRNA_{CUA}/AzFRS pair-directed AzF incorporation in response to the amber codon in primary cells isolated from transgenic mice of F2-F3 generation. Representative images are shown for primary neuronal cells **(D)**, bone marrow cells **(E)** and fibroblasts **(F)** transduced with lentivirus carrying the reporter gene eGFP-amb-mCherry in the presence or absence of 1 mM AzF. Scale bar: 100 μm in **D** and **F**; 20 μm in **E**. **(G)** Schematic diagram of the construct used to generate transgenic zebrafish and the reporter construct. The transgene construct contains an orthogonal tRNA_{CUA}/AzFRS pair, SVEpA (SV40 early polyadenylation signal), SVLpA (SV40 late polyadenylation signal), γCRY (xenopus γ-crystallin promoter), U6 (human U6 pol III promoter), INS (SP-10 mouse insulator) and ubi (zebrafish ubiquitin promoter). **(H)** Western blot analysis of AzFRS-FLAG expression in embryos or adult caudal fin tissues in F2 generation using an anti-FLAG antibody (indicated by an arrow). The FLAG antibody also detects a non-specific band, migrating slower than the tagged AzFRS band. **(I-K)** Incorporation of AzF into the eGFP reporter in the transgenic zebrafish *in vivo*. 1-cell-stage embryos of F2-F3 generations of transgenic zebrafish were injected with the mRNA encoding the mCherry-eGFP (Y145amb) reporter. The embryos were incubated in water with or without 2.5 mM AzF for 24 h. Incorporation of AzF at the amber codon site allows full-length eGFP expression in the nucleus. The fluorescence images of whole embryos **(I)**, and mesenchymal **(J)** and notochord **(K)** tissues are shown. Arrowheads indicate cells with AzF incorporation. Scale bars: 50 μm.

types analyzed, i.e., mesenchymal, notochord and muscle cells (Figure 1J and 1K and Supplementary information, Figure S2C).

In conclusion, we have shown that by integrating an orthogonal tRNA/aaRS pair into the genome, an expanded genetic code can be established and inherited in two vertebrate species, mouse and zebrafish. This work thus demonstrates the successful code expansion in living beings of a higher biological complexity than reported, suggesting unexpectedly high malleability of the genetic code. Of note, these vertebrate models can be equipped with Uaas of tailored chemical, physical and biological properties, e.g., light-sensitive Uaas for optogenetics [9] and bio-reactive Uaas for *in situ* chemistry [10], which would greatly facilitate their applications in biological research. Uaa incorporation *in vivo* has the potential to impact developmental and behavior studies by precisely probing and manipulating target proteins in their native habitat. Moreover, with Uaa-incorporated cells, tissues and whole animal derived from the same transgenic animal, it now becomes possible to correlate Uaa incorporation-enabled results from single cells to organs and the intact organism in a cohesive manner. Further investigations of host tolerance to code expansion [11], site-specific integration of transgene via genome editing and tissue-dependent responses will guide optimization of the introduced components, ultimately leading to increased efficiency of Uaa encoding and extension of the strategy to other tRNA/aaRS pairs and vertebrates.

Acknowledgments

We thank Rodriguez P, Gautier A and Bensaude O for suggestions on the paper. We thank professor Thomas P Sakmar for generous gift of the tRNACUA/AzFRS pair and Dr Frank Kao from Flow Cytometry Support in Centenary Institute for bone marrow cell sorting. We also thank Jiwei Chen in East China Normal University for analyzing the RNA-Seq data. This work was supported the State Key Development Programs of China (2012CB910400 to ML), the National Natural Science Foundation of China (31371455 and 31171318 to DL, 81330049 to ML, and 31528007 to SY and CY), and the Science and Technology Commission of Shanghai Municipality (14140900300 to DL). Other financial supports were provided by the ECNU Fund (short-term visiting fellowship to YC), the Chinese Scholars Council

(CSC fellowship to MT), the Agence Nationale de la Recherche of France (ANR-JCJC ChemProbe to SY) and US National Institutes of Health (1R01GM118384-01 to LW).

Yuting Chen¹, Ji Ma², Wenqing Lu¹, Meilin Tian^{2,3}, Marion Thauvin⁴, Chonggang Yuan², Michel Volovitch^{4,5,6}, Qian Wang⁷, Jeff Holst⁷, Mingyao Liu¹, Sophie Vriz^{4,5,6,8}, Shixin Ye^{3,5,6,10}, Lei Wang⁹, Dali Li¹

¹Shanghai Key Laboratory of Regulatory Biology, Institute of Biomedical Sciences and School of Life Sciences, East China Normal University (ECNU), Shanghai, China; ²Shanghai Key Laboratory of Brain Functional Genomics, School of Life Sciences, ECNU, Shanghai, China; ³Institut de Biologie de l'Ecole Normale Supérieure, Ecole Normale Supérieure, Paris, France; ⁴Center for Interdisciplinary Research in Biology (CIRB), Collège de France, Paris, France; ⁵Institut National de la Santé et de la Recherche Médicale (INSERM) U1024, U1050, Paris, France; ⁶Centre National de la Recherche Scientifique (CNRS) UMR 7241, 8197, Paris, France; ⁷Origins of Cancer Program, Centenary Institute, Sydney Medical School, University of Sydney, Australia; ⁸Department of Biology, Université Paris Diderot, Sorbonne Paris Cité, Paris, France; ⁹Department of Pharmaceutical Chemistry and the Cardiovascular Research Institute, University of California, San Francisco, CA, USA

Correspondence: Shixin Ye^a, Lei Wang^b, Dali Li^c

^aE-mail: yelehman@biologie.ens.fr

^bE-mail: lei.wang2@ucsf.edu

^cE-mail: dllli@bio.ecnu.edu.cn

¹⁰Current address: Laboratory of Computational and Quantitative Biology (LCQB), Institute of Biology, Paris-Seine, University of Pierre and Marie Curie, Paris, France

References

- 1 Wang L, Brock A, Herberich B, Schultz PG. *Science* 2001; **292**:498-500.
- 2 Liu CC, Schultz PG. *Annu Rev Biochem* 2010; **79**:413-444.
- 3 Parrish AR, She X, Xiang Z, et al. *ACS Chem Biol* 2012; **7**:1292-1302.
- 4 Bianco A, Townsley FM, Greiss S, Lang K, Chin JW. *Nat Chem Biol* 2012; **8**:748-750.
- 5 Kang JY, Kawaguchi D, Coin I, et al. *Neuron* 2013; **80**:358-370.
- 6 Ye S, Zaitseva E, Calabiano G, et al. *Nature* 2010; **464**:1386-1389.
- 7 Shen B, Xiang Z, Miller B, et al. *Stem Cells* 2011; **29**:1231-1240.
- 8 Suster ML, Kikuta H, Urasaki A, Asakawa K, Kawakami K. *Methods Mol Biol* 2009; **561**:41-63.
- 9 Fehrentz T, Schonberger M, Trauner D. *Angew Chem Int Ed Engl* 2011; **50**:12156-12182.
- 10 Xiang Z, Ren H, Hu YS, et al. *Nat Methods* 2013; **10**:885-888.
- 11 Wang Q, Sun T, Xu J, et al. *Chem Bio Chem* 2014; **15**:1744-1749.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)