Acta Biochim Biophys Sin 2024, 56(11): 1687–1698
https://doi.org/10.3724/abbs.2024081
Advance Access Publication Date: 15 May 2024
Original Article

Original Article

Comparative study on the expression characteristics of transgenes inserted into the *Gt(ROSA)26Sor* and *H11* loci in mice

Zhilan Yu[1], Jiahao Shi[1], Jingyu Zhang[1], Youbing Wu[3], Ruling Shen[2,*], and Jian Fei[1,3,*]

[1]School of Life Science and Technology, Tongji University, Shanghai 200092, China, [2]Shanghai Laboratory Animal Research Center, Shanghai 201203, China, and [3]Shanghai Engineering Research Center for Model Organisms, SMOC, Shanghai 201203, China
*Correspondence address. Tel: +86-18916198076; E-mail: jfei@tongji.edu.cn (J.F.) / E-mail: shenruling@slarc.org.cn (R.S.)

Received 16 December 2023 Accepted 22 April 2024

Abstract

The *Gt(ROSA)26Sor* (*ROSA26*) and *H11* loci are commonly used as safe harbors for the construction of targeted transgenic mice. However, it remains unclear whether these two loci have distinct effects on transgene expression. In this study, we insert three differently colored fluorescent protein expression cassettes (EGFP, tdTomato and mTagBFP2) driven by the CAG promoter into the *ROSA26* and *H11* loci. We generate five single-transgenic mouse models and a triple-transgenic mouse model expressing three distinct fluorescent proteins simultaneously. Our results reveal that the efficiency of transgene expression is greater at the *ROSA26* locus with a reverse orientation relative to the transcription of the *ROSA26* gene. In most tissues examined, the efficiency of transgene expression at the *ROSA26* locus exceeds that at the *H11* locus. Moreover, the expression profiles of identical transgenes display discrepancies across various tissues, and notably, substantial heterogeneity in transgene expression is discernible within cells of the same tissue. Our findings offer a valuable reference for the selection of safe harbors and strategies for the construction of transgenic mouse models.

Key words transgenic mouse, knockin, *Gt(ROSA)26Sor*, *H11*, gene expression

Introduction

In the process of constructing transgenic mice, site-directed knock-in at a "safe harbor" is commonly employed to avoid the uncertainty caused by random insertion. The "safe harbor" refers to regions in the mouse genome that lack genes influencing physiological functions while ensuring that the exogenously inserted genes function properly. The *ROSA26* locus and the *H11* locus are the two most frequently utilized safe harbors thus far.

The *Gt(ROSA)26Sor* locus, also known as *ROSA26*, is located on mouse chromosome 6 and spans from 113,044,389 to 113,054,205 (GRCm39). It was identified by G. Friedrich and P. Soriano in their research on promoter trapping in mouse embryonic stem cells [1]. The *ROSA26* gene encodes a widely expressed noncoding RNA with an unknown function [2]. Mice exhibiting insertion mutations at this locus do not demonstrate significant phenotypic changes, making it a safe harbor. The chromosomal localization, promoter regions, and transcripts have been characterized [2]. By utilizing the endogenous promoter, the inserted transgene achieves ubiquitous expression in mouse embryos and newborn pups [3–5]. Furthermore, the homologues have been applied in the development of targeted transgenic models in rat, pig, and human embryonic stem cell lines, among others [6–10].

The *Hipp11* locus, also known as *H11* locus, was discovered by Hippenmeyer *et al.* [11] in their research on neuronal migration. It is situated between the *Eif4enif1* gene and the *Drg1* gene on chromosome 11, with no endogenous genes identified within this region. Several transgenic models inserted at this locus have confirmed its safety as a harbor [12–14].

The targeted insertion of transgenes at these two sites has been utilized in the construction of transgenic mice for more than a decade. However, there have been few systematic comparative studies examining the effects of these sites on transgene expression patterns. In this study, we constructed five mouse models expressing distinct fluorescent proteins by knocking in transgenes into the *ROSA26* or *H11* locus. By crossing these strains, we obtained a mouse line that simultaneously expresses three distinct fluorescent proteins.

We demonstrated that greater exogenous gene expression can be achieved by inserting the transgene in the opposite direction of *ROSA26* gene transcription at the *ROSA26* locus compared with the normal direction. In cardiac, splenic, pulmonary, and renal tissues, the same transgene expression strategy resulted in higher expression levels at the *ROSA26* locus than at the *H11* locus. Our study in tricolor transgenic mice revealed heterogeneity in the expression of the same transgenes within tissue cells. These findings provide valuable information and an interpretative basis for selecting transgene expression strategies and understanding the phenotypic changes induced by transgene expression.

Materials and Methods

Gain of transgenic mice and animal husbandry

The transgenic mouse lines used in this study, including the *ROSA26-CAG-EGFP-WPRE-polyA* (abbreviated as *R26-EGFP*) line, the *ROSA26-SA-2×polyA-CAG-EGFP-WPRE-polyA* (abbreviated as *R26-stop-EGFP*) line, the *ROSA26-exon2-CAG-EGFP-WPRE-polyA-exon1* (abbreviated as *R26-reverse-EGFP*) line, the *H11-CAG-EGFP-WPRE-polyA* (abbreviated as *H11-EGFP*) line, the *ROSA26-CAG-tdTomato-WPRE-polyA* (abbreviated as *R26-tdTomato*) line, and the *ROSA26-CAG-mTagBFP2-WPRE-polyA* (abbreviated as *R26-mTagBFP2*) line, were constructed and provided by the Shanghai Model Organisms Center (Shanghai, China). The full names, abbreviations and corresponding schematic maps of the edited genomes for each strain are summarized in Table 1. The models were constructed on the basis of the C57BL/6 mouse line. The CRISPR/Cas9 system was applied to mouse zygotes to target the *ROSA26* locus and the *H11* locus [15]. The sgRNA targeting the *ROSA26* locus was GGGGACACACTAAGGGAGCT, which corresponds to the genomic position 113,050,181 to 113,050,200 (with 113,050,178 to 113,050,180 being the PAM code) of chromosome 6 (Figure 1A). The sgRNA targeting the *H11* locus was AGCTCATTAGATGCCATCAT, which corresponds to the genomic position 3,195,257 to 3,195,276 (with 3,195,254 to 3,195,256 being the PAM code) of chromosome 11 (Figure 1G). DNA templates featuring homologous sequences corresponding to the targeted loci were utilized to facilitate homologous repair and integrate the transgene expression cassettes. The 5′ and 3′ homology arms flanking *ROSA26* cleavage were both 3.3 kb long. The expression cassette in the *R26-reverse-EGFP* line was assembled in an inverted manner in contrast to the other lines integrated at *ROSA26*, allowing it to attain the opposite transcriptional orientation. The 5′ and 3′ homology arms flanking *H11* cleavage were 5 kb and 3.7 kb long, respectively. The identical and differential components of the transgene expression cassettes for each strain are summarized in Table 2. The expression cassettes of each line shared the same CAG promoter, WPRE and polyA tail. The *R26-EGFP* line, the *R26-stop-EGFP* line, the *R26-reverse-EGFP* line and the *H11-EGFP* line shared the same reporter gene (*EGFP*). Compared to the *R26-EGFP* line, the *R26-stop-EGFP* line contained a splice acceptor (SA) sequence and a two-copy polyA tail upstream of the CAG promoter; the *R26-reverse-EGFP* line was designed to transcribe in the opposite orientation; the *H11-EGFP* line had the same cassette integrated at the different site; and the *R26-tdTomato* line and the *R26-mTagBFP2* line had different reporter genes.

Cas9 was delivered in the form of mRNA (Novoprotein Scientific, Suzhou, China). The sgRNA was delivered in the form of chemically synthesized RNA (GenScript Biotech, Nanjing, China). The targeting vector was delivered in the form of a circular plasmid assembled with the pBR322 plasmid backbone (Shanghai Model Organisms Center). The components were delivered into zygotes of C57BL/6 mice by microinjection at concentrations of 100–1000 ng of Cas9 mRNA, 50–250 ng of sgRNA and 500–2000 ng of targeting vector in 100 μL of fluid.

The animals were bred, and the genotypes were identified for successful transgene knockin. Genomic qPCR assays were performed on each line to confirm the copy number of the knockin alleles. The *Actb* gene was used as the 2-copy control, and the knockin allele copy numbers were calculated as $2\times2^{-\Delta Ct}$ (Supplementary Material S1).

All animals were maintained in a specific pathogen-free (SPF) environment with a 12-h light/dark cycle, temperature of 20–26°C,

Table 1. Transgenic mouse strain names, abbreviations, and correspondence of genome editing patterns

Strain full name	Abbreviation	Genome edit pattern
ROSA26-CAG-EGFP-WPRE-polyA	*R26-EGFP*	Figure 1B
ROSA26-SA-2×polyA-CAG-EGFP-WPRE-polyA	*R26-stop-EGFP*	Figure 1C
ROSA26-exon2-CAG-EGFP-WPRE-polyA-exon1	*R26-reverse-EGFP*	Figure 1D
H11-CAG-EGFP-WPRE-polyA	*H11-EGFP*	Figure 1H
ROSA26-CAG-tdTomato-WPRE-polyA	*R26-tdTomato*	Figure 1F
ROSA26-CAG-mTagBFP2-WPRE-polyA	*R26-mTagBFP2*	Figure 1E

Table 2. Components of the transgenic mouse strains

Component / Strain	SA-2×polyA (~0.6 kb)	CAG (~1.6 kb)	EGFP (~0.7 kb)	tdTomato (~1.4 kb)	mTagBFP2 (~0.7 kb)	WPRE (~0.6 kb)	polyA (~0.2 kb)
R26-EGFP		√	√			√	√
R26-stop-EGFP	√	√	√			√	√
R26-reverse-EGFP		√	√			√	√
R26-tdTomato		√		√		√	√
R26-mTagBFP2		√			√	√	√
H11-EGFP		√	√			√	√

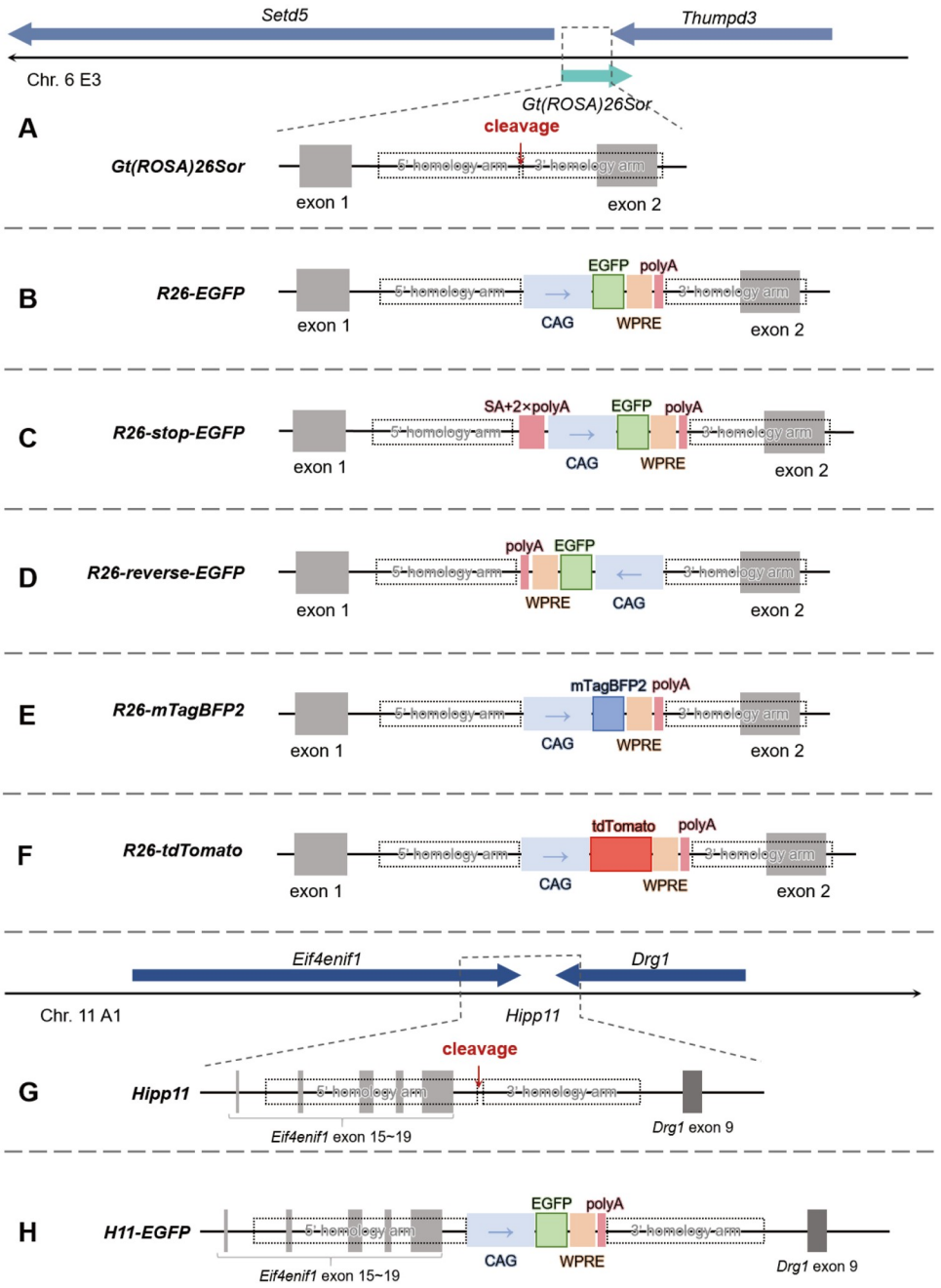

Figure 1. Structures of the *ROSA26* locus and the *H11* locus after the insertion of transgene expression cassettes (A) Location and structure of the wild-type *ROSA26* locus with a cleavage site. (B–F) Structures of the *ROSA26* locus in the transgenic models. (G) Location and structure of the wild-type *H11* locus with a cleavage site. (H) Structures of the *H11* locus in the transgenic model.

and humidity of 40%–70%, with *ad libitum* access to food and water. All experimental procedures involving mice were approved by the Institutional Animal Care and Use Committee (IACUC) of the Shanghai Engineering Research Center for Model Organisms, SMOC (authorization number: 2020-0021-1) and were conducted in accordance with relevant national and international guidelines. Mice were euthanized by CO_2 gas asphyxia or intraperitoneal injection of an overdose of phenobarbital (150 mg/kg).

Animal usage arrangement

To compare the expression of fluorescent proteins inserted at the *ROSA26* locus via different strategies, 1-month-old and 4-month-old male heterozygous genotype *R26-EGFP* mice, *R26-stop-EGFP* mice and *R26-reverse-EGFP* mice were used. A group of 3 mice for each line of each age was used, with the exception of 4-month-old *R26-EGFP* mice being a group of 2.

For comparative analysis of transgene expression at the *ROSA26* locus and the *H11* locus, 10- to 12-week-old male heterozygous genotype *R26-EGFP* mice and *H11-EGFP* mice were used, with a group of 3 mice for each line.

For the analysis of transgene expression in the trifluorescent transgenic mouse model, three 10- to 12-week-old male hetero-

zygous genotype mice were used.

Tissue sample preparation and treatment

The animals used for RNA and protein sampling were dissected, and heart, liver, spleen, lung, kidney, and stomach tissues were obtained. The tissues were cut and placed in TRNzol Universal solution (TIANGEN Biotech, Beijing, China) or RIPA solution with phosphatase inhibitor cocktail (MeilunBio, Dalian, China). Total RNA and total protein were extracted according to the manufacturers' instructions. The RNA samples were quantified according to the OD_{260} and reverse transcribed to cDNA (TransGen Biotech, Beijing, China). Protein samples were quantified using a BCA assay kit (Beyotime Biotech, Shanghai, China) and treated with loading buffer (Beyotime Biotech) for western blot analysis.

The animals used for frozen sectioning were perfused with 4% PFA. The brain, heart, liver, spleen, lung, kidney, and testis tissues were dissected and postfixed in 4% PFA overnight. The fixed tissues were dehydrated twice with 15% and 30% sucrose/PBS solutions.

qPCR assay and western blot analysis

qPCR primers were designed for the *EGFP* gene and the *β-actin* gene (*EGFP*-F: 5′-ACGACGGCAACTACAAGACC-3′, *EGFP*-R: 5′-TTGTACTC CAGCTTGTGCCC-3′, *β-actin*-F: 5′-GGAGATTACTGCCCTGGCTCC TA-3′, *β-actin*-R: 5′-GACTCATCGTACTCCTGCTTGCTG-3′). The primers were synthesized by Sangon Biotech (Shanghai, China). The relative expression of transgenic mRNA was detected by qPCR (TransGen Biotech, Beijing, China). The cDNA sample corresponding to each RNA sample was assayed in 3 replicates for each gene.

Relative protein expression was detected by western blot analysis. After separation by 10% SDS-PAGE (Epizyme Biotech, Shanghai, China), the samples were transferred to nitrocellulose (NC) membranes (Amersham Biosciences, Braunschweig, Germany), blocked with Protein Free Rapid Blocking Buffer (Epizyme Biotech), incubated with anti-GFP antibody (1:2500 dilution, ab290; Abcam, Cambridge, UK) and anti-β-actin antibody (1:5000 dilution, LF201; Epizyme Biotech), and incubated with the corresponding HRP-conjugated secondary antibodies (1:5000 dilution, LF101, LF102; Epizyme Biotech), after which the luminescence signals were detected using *EasySee*® Western Blot Kit (TransGen Biotech).

Frozen sectioning and imaging

The fixed and dehydrated tissues were embedded in OCT reagent, rapidly frozen in liquid nitrogen for 20 s, and then cut into 10-μm tissue sections at –20°C. The sections were washed with PBS three times, sealed with Antifade Mounting Medium (Beyotime Biotech), and stored at 4°C in the dark. Fluorescence images of EGFP, tdTomato, and mTagBFP2 were captured using a confocal fluorescence microscope (Carl Zeiss Microscopy GmbH, Jena, Germany), and the excitation wavelength and reception wavelength were adjusted to prevent fluorescence signal interference.

Statistical analysis

The Ct results of the qPCR assay were used to calculate the relative mRNA expressions of the genes using the $2^{-\Delta Ct}$ method. The western blot bands were quantified using ImageJ (1.52a; NIH, Bethesda, USA) software to calculate the relative protein expressions of the genes. The fluorescence images of the sections were quantified by circling the area of interest to obtain signal values using the image processing software provided with the microscope (ZEN blue edition, Carl Zeiss Microscopy GmbH). 3D scatter plots were generated using OriginLab (Northampton, USA) software, while other images and statistical analyses were performed using GraphPad Prism (La Jolla, USA) software. All results are expressed as the mean±standard error (SEM). The Shapiro-Wilk test was used to assess whether the data sets conform to a normal distribution. For normally distributed data sets, Student's t test was used for intergroup comparisons, or the Pearson correlation coefficient was calculated. For nonnormally distributed data sets, the Mann-Whitney nonparametric test and Spearman correlation coefficient calculation were used. P values less than 0.05 were considered to indicate statistical significance, and $|r|$ values greater than 0.3 were considered to indicate a linear correlation.

Results

Comparison of the expression levels of fluorescent proteins inserted at the *ROSA26* locus via different strategies

Transgenic mouse models incorporating *EGFP* at the *ROSA26* locus were developed. The composition of the transgene expression cassettes is depicted in Figure 1B–D. The *EGFP* gene is driven by the CAG promoter, enhanced by the WPRE, and the polyA tail serves as a terminal transcription and mRNA stability element. The transgene cassettes were inserted into the intron between exon 1 and exon 2 of the *ROSA26* gene. Three distinct insertion strategies were employed.

In the first strategy, the transcription of the EGFP transgene proceeded in the same direction as the endogenous *ROSA26* gene, which was referred to as the *R26-EGFP* group (Figure 1B). The second strategy involved the addition of a splice acceptor (SA) sequence with a two-copy polyA tail upstream of the CAG promoter based on the *R26-EGFP* strategy to inhibit the transcription of the endogenous *ROSA26* gene, labelled the *R26-stop-EGFP* group (Figure 1C). The third strategy was to insert the transgene cassette in the reverse direction of the *ROSA26* gene, utilizing a cassette similar to the *R26-EGFP* strategy, which was labelled the *R26-reverse-EGFP* group (Figure 1D). All three mouse models were successfully generated and exhibited ubiquitous EGFP fluorescence in newborns (data not shown).

Differences in EGFP mRNA and protein expression in multiple tissues among the three models were assessed by qPCR and western blot (Figure 2). qPCR revealed that EGFP mRNA expression in the *R26-reverse-EGFP* group was significantly greater than that in the *R26-EGFP* group and the *R26-stop-EGFP* group in the heart, liver, spleen, lung, kidney, and stomach (Figure 2A,C,E,G,I,K). The *R26-EGFP* group exhibited greater expression than the *R26-stop-EGFP* group in the liver, while the *R26-stop-EGFP* group demonstrated greater expression than the *R26-EGFP* group in the spleen. According to the western blot analysis results, the levels of EGFP protein in the liver and stomach were significantly greater in the *R26-reverse-EGFP* group than in the other two groups (Figure 2D, L). In the spleen and kidney, the levels of EGFP protein in the *R26-reverse-EGFP* group were significantly greater than in the *R26-stop-EGFP* group (Figure 2F,J). In the heart and lung, the differences in EGFP expression between the groups were not significant, but the *R26-reverse-EGFP* group still exhibited the highest mean value (Figure 2B,H).

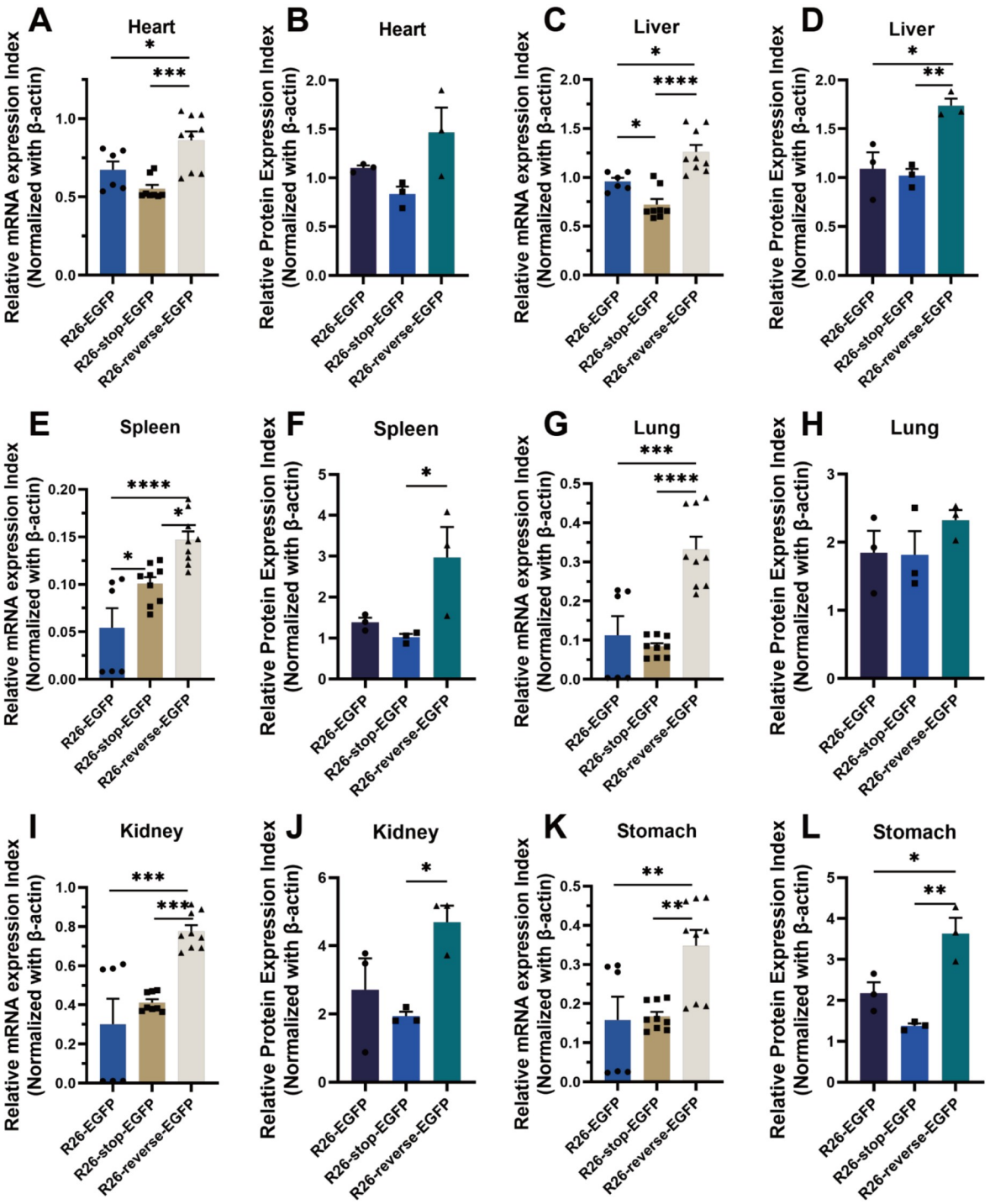

Figure 2. Relative mRNA and protein expressions of transgenic EGFP in various tissues from the three groups of mice (A,C,E,G,I,K) Relative mRNA expression of transgenic EGFP quantified by qPCR in cardiac, hepatic, splenic, pulmonary, renal, and gastric tissues from the three groups of mice. (B,D,F,H,J,L) Relative protein expression of transgenic EGFP detected by western blot analysis in cardiac, hepatic, splenic, pulmonary, renal, and gastric tissues from the three groups of mice. Band images are provided in Supplementary Material S2. n=3. *P<0.05, **P<0.01, ***P<0.001, and ****P<0.0001.

From the aforementioned results, it was observed that the mRNA and corresponding protein expressions of the transgene did not consistently show identical trends, consistent with the known view. Statistically significant differences between mRNAs could be detected in several tissues, whereas not all differences between relevant proteins were detected. In the spleen, lung, kidney, and stomach, the mRNAs and proteins exhibited opposite mean value differences between the *R26-EGFP* and *R26-stop-EGFP* groups. This suggests that detecting different forms of transgene expression may lead to diverse statistical conclusions.

Comparative analysis of transgene expression at the *ROSA26* locus and the *H11* locus

The CAG-EGFP-WPRE-polyA cassette was inserted into the *H11* locus, which is located between exon 19 of the *Eif4enif1* gene and exon 9 of the *Drg1* gene on chromosome 11, with the transcription direction parallel to that of the *Eif4enif1* gene (Figure 1H). This model was labeled the *H11-EGFP* model.

The expressions of EGFP mRNA and protein in various tissues of the *R26-EGFP* model and the *H11-EGFP* model were assessed via qPCR and western blot analysis. In the heart, spleen, lung, and kidney, the mean relative transgene expression levels of mRNA and protein in the *R26-EGFP* model group were greater than those in the *H11-EGFP* model group (Figure 3A,B,E–J). In the liver, no significant differences in mRNA or protein expression were observed between the two groups (Figure 3C,D). However, a notable discrepancy in mRNA levels was observed in the stomach, whereas no such

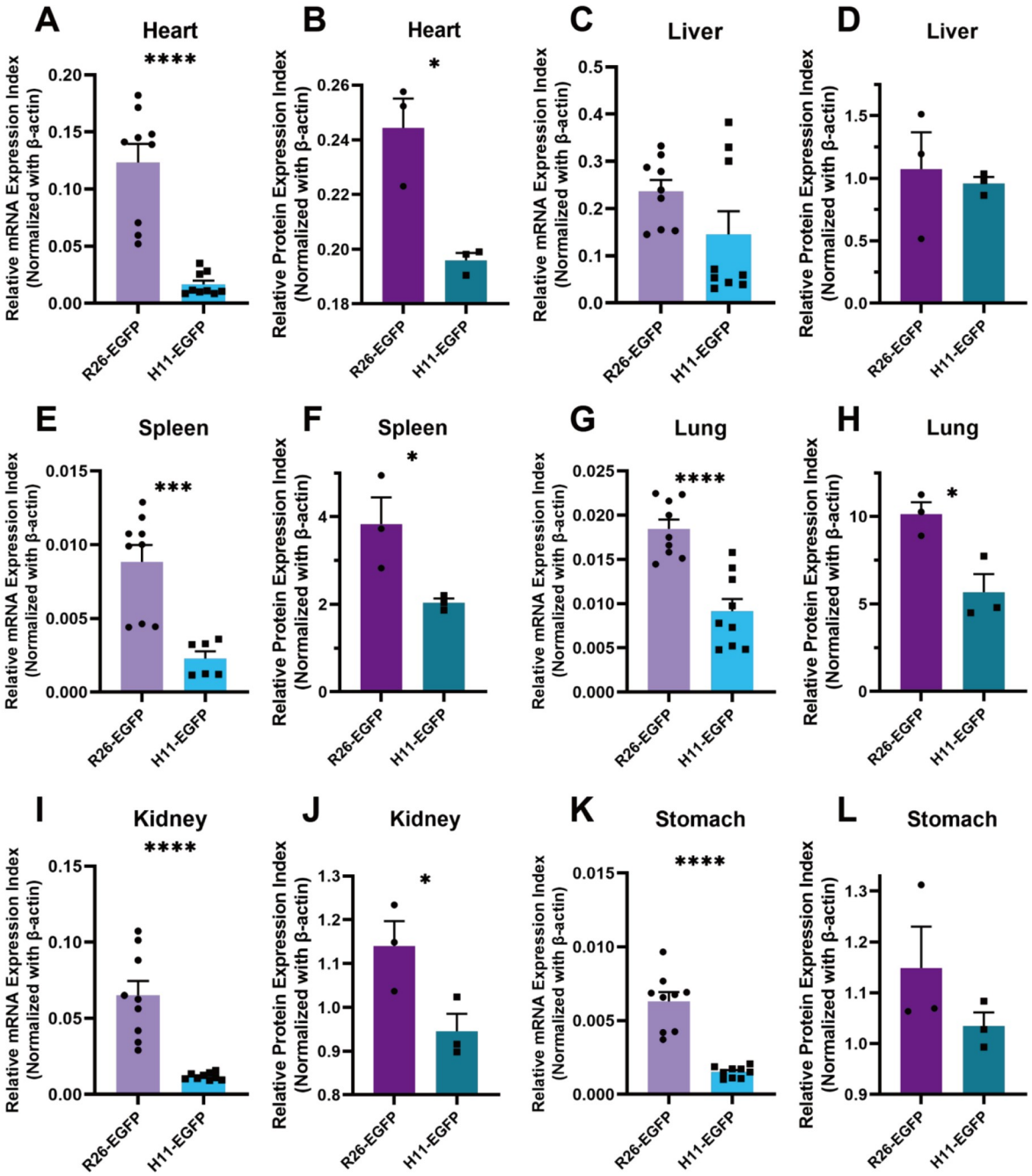

Figure 3. Relative mRNA and protein expressions of transgenic EGFP in various tissues of the *R26-EGFP* model and the *H11-EGFP* model (A,C,E,G,I,K) Relative mRNA expression of transgenic EGFP detected by qPCR in cardiac, hepatic, splenic, pulmonary, renal, and gastric tissues of the *R26-EGFP* model and the *H11-EGFP* model. (B,D,F,H,J,L) Relative protein expression of transgenic EGFP detected by western blot analysis in cardiac, hepatic, splenic, pulmonary, renal, and gastric tissues of the *R26-EGFP* model and the *H11-EGFP* model. Band images are provided in Supplementary Material S3. n=3. *P<0.05, **P<0.01, and ****P<0.0001.

difference was observed in protein levels (Figure 3K,L). These results suggest that identical exogenous ubiquitous promoters exhibit varying expression efficiencies at different safe harbors.

Generation of the trifluorescent transgenic mouse model

Two mouse models, the *R26-tdTomato* model and the *R26-mTagBFP2* model, were developed with distinct fluorescent protein transgenes inserted at the *ROSA26* locus. Similar to the previously described *R26-EGFP* model, the transgene cassettes in these models incorporated a CAG promoter upstream and a WPRE sequence followed by a polyA tail downstream of the coding sequence. The coding sequences for the fluorescent proteins within these cassettes were different and consisted of tdTomato or mTagBFP2 (Figure 1E, F). These two models, along with the aforementioned *H11-EGFP* model, constituted a total of three monofluorescent protein transgenic lines as parents for subsequent crosses.

By crossbreeding, the first line expressing two transgenes, tdTomato and EGFP, was obtained. Subsequently, mTagBFP2 transgenic lines were crossed to generate a mouse model expressing trifluorescent transgenes (Figure 4). This model exhibited tdTomato (red), EGFP (green), and mTagBFP2 (blue) fluorescence, hereafter referred to as *RGB* mice. The varying proportions of the three fluorescent proteins in the cells resulted in distinct colors, allowing intuitive visualization of transgene expression at the tissue and cellular levels.

The *RGB* trifluorescent mouse model exhibited ubiquitous fluorescence of the three proteins. Frozen sections of the cerebellum, heart, liver, kidney, lung, spleen, and testis tissues were examined by confocal microscopy. Each tissue sample was observed to display a combination of three fluorescent signals. In the merged fluorescence images, the three fluorescent proteins could be visualized when combined in different ratios to form distinct colors, as red, green, and blue are the three primary colors in optics. For example, the molecular layer in the cerebellar cortex could be clearly distinguished from the granular layer, as well as

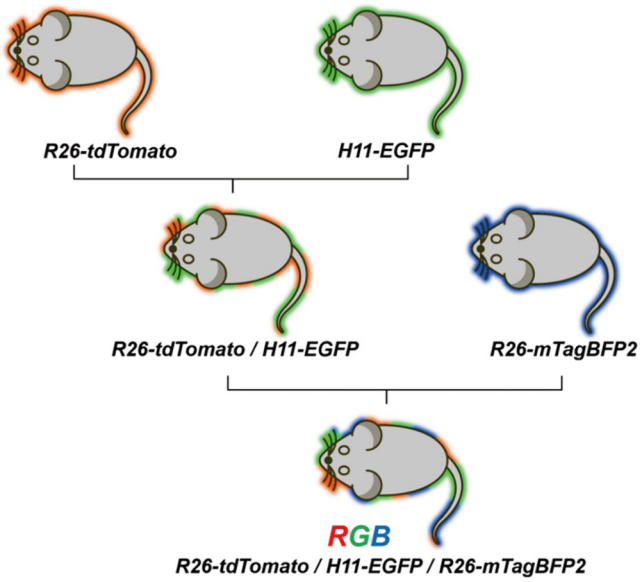

Figure 4. The crossbreeding strategy for generating the trifluorescent mouse line from three monofluorescent mouse lines The mice obtained and utilized thereafter were all heterozygous genotypes at the transgenic locus.

glomeruli and tubules in the kidney tissue, and the spleen and testis partially showed mosaic patterns of various fluorescences (Figure 5). However, without the addition of nucleus or membrane markers, only a few cell types with distinct morphologies could be recognized, such as hepatocytes.

Fluorescence variations among diverse cell types in the trifluorescent model

In the cerebellar cortex sections of the *RGB* mice, the Purkinje cell somas were observed to have a distinct shape, with fluorescence merged into a vibrant blue-green color. In the molecular layer, some neuron somas (stellate or basket cells) were visible with dimmer and reddish fluorescence (Figure 6A). To quantify the fluorescence in the images, the average values of the fluorescence signals per pixel in the circled area were calculated. Based on the quantification, the proportions of the three fluorescent types for cerebellar molecular layer neuron somas and Purkinje cell somas could be calculated, which differed between the two cell types and resulted in distinct colors on the merged images (Figure 6D). A 3D coordinate system was established based on the fluorescence proportions of the cells, with each dot representing a cell, indicating the separation of the two cell types (Figure 6E). Notably, the fluorescence per unit was greater in Purkinje cell somas than in molecular layer neuron somas (Figure 6B). There was no significant difference in tdTomato expression between the two cell types, whereas EGFP and mTagBFP2 expression was significantly greater in Purkinje cell somas than in molecular layer neuron somas (Figure 6C).

In the lung sections of the *RGB* mice, bronchial epithelial cells and type II alveolar cells were detected (Figure 6F). There was no significant difference in the sum values of fluorescence for the two cell types (Figure 6G) or in the individual fluorescence values (Figure 6H). However, calculation of the proportions of fluorescence resulted in significant differences in tdTomato and EGFP (Figure 6I), and the cells showed minor separation in the 3D scatter plot (Figure 6J), partially explaining the color differences observed in the fluorescence images. The type I alveolar cells exhibited more noticeable color differences in the images but were not quantified due to the difficulty in delineating the cell shapes. As demonstrated by the examples from the cerebellum and lung, the expressions of the three fluorescence signals and the proportions of each to the other varied significantly between different tissues. However, the summed signals of transgene expressions and the proportions of each fluorescence in the same tissue under the same sectioning and photographing conditions might also vary considerably between different cell types. This highlights the fact that CAG, an exogenous ubiquitous promoter inserted into the same safe harbor (at different chromosomes), can have different transcriptional activities in diverse tissues and cells.

Differences in fluorescence between identical cell types in the trifluorescent model

In the liver section images of the *RGB* mice, well-defined hepatocytes with minimal differences in merged fluorescence were observed (Figure 7A). The hepatocytes were circled for fluorescence intensity quantification, and a 3D scatter plot was generated based on the absolute values of fluorescence (Figure 7B). No evident clustering tendency was observed in the 3D scatter plot. A normality test was performed on the hepatocyte fluorescence values, revealing

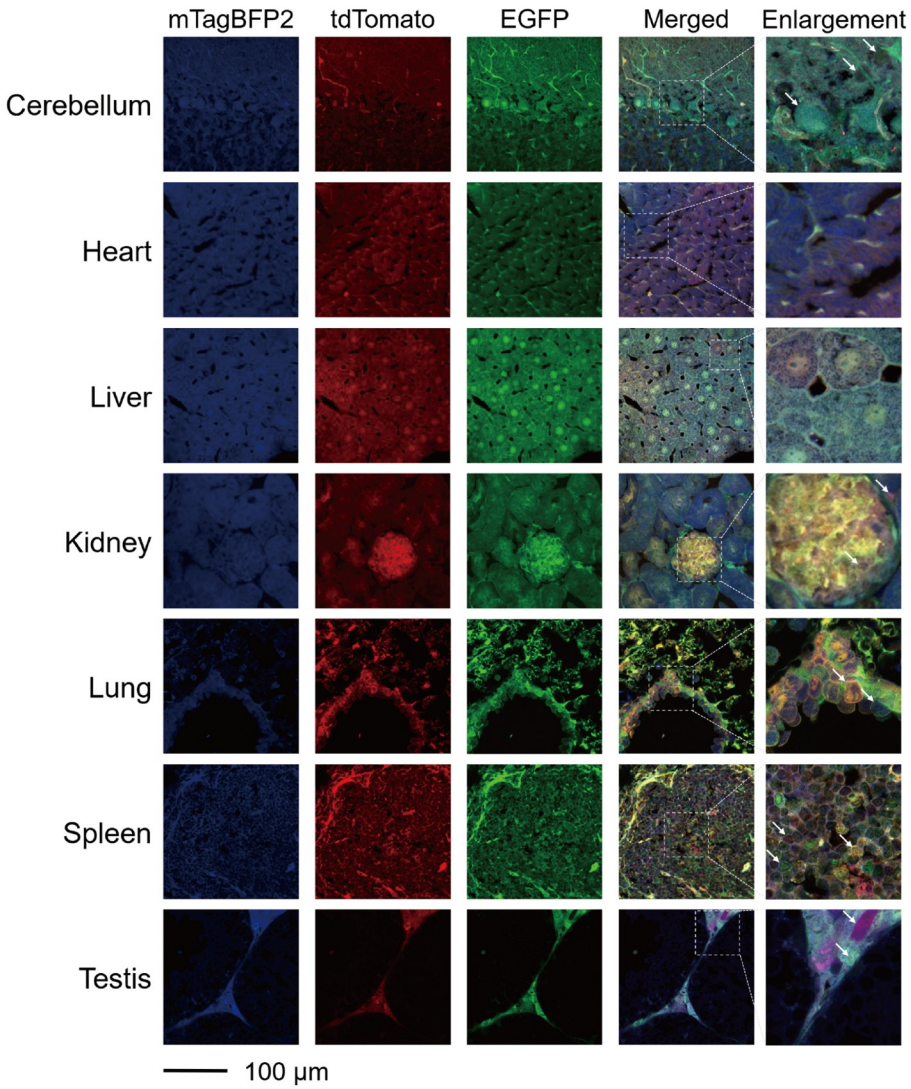

Figure 5. Images of frozen sections of the cerebellar cortex, heart, liver, kidney, lung, spleen, and testis of the *RGB* mice obtained using fluorescence confocal microscopy

that EGFP, mTagBFP2, and the sum fluorescence values conformed to a normal distribution, indicating that they originated from a homogeneous set of samples, while tdTomato did not pass the normality test, suggesting a lack of uniformity (Table 3). The coefficient of variation calculation showed that tdTomato had the highest coefficient of dispersion (Table 3). These findings suggested that the random bias in the variation in the expression of tdTomato among hepatocytes was greater than that of EGFP and mTagBFP2. Considering that the promoters of the three transgenes were identical and that the insertion locus of tdTomato was the same as that of mTagBFP2, it was speculated that the greater expression heterogeneity of tdTomato might be related to the biochemical properties of the protein. Pearson and Spearman correlation coefficient analyses of the fluorescence values showed that the tdTomato level was slightly correlated with the EGFP level and that the EGFP level was moderately correlated with the mTagBFP2 level (Table 4). Another 3D scatter plot was constructed using the proportions of each fluorescence to the sum values of the merged fluorescence, which exhibited a rather discrete trend (Figure 7C).

In the spleen sections of the *RGB* mice, dense lymphocytes were visible in the white pulp (Figure 7D). Fluorescence quantification and construction of a 3D scatter plot showed that the fluorescence intensity varied among lymphocytes (Figure 7E). All three fluorescence values and the sum of the values failed the normality test and exhibited a strong tendency to deviate from the normal distribution, with EGFP having the highest coefficient of dispersion (Table 3). Spearman correlation coefficient analysis indicated that EGFP expression was weakly correlated with tdTomato and mTagBFP2 expressions (Table 4). No clear trend was observed in the 3D scatter plot of fluorescence proportions (Figure 7F).

In the kidney sections of the *RGB* mice, glomeruli and tubules could be observed (Figure 7G), but distinct cellular boundaries could not be discerned. The renal tubules were circled on the image for fluorescence quantification and 3D scatter plot construction (Figure 7H,I), and no evident clustering trend was presented in the scatter plot. The tdTomato signal did not pass the normality test and had a greater dispersion coefficient than the other fluorescence signals (Table 3). According to the correlation coefficient analysis,

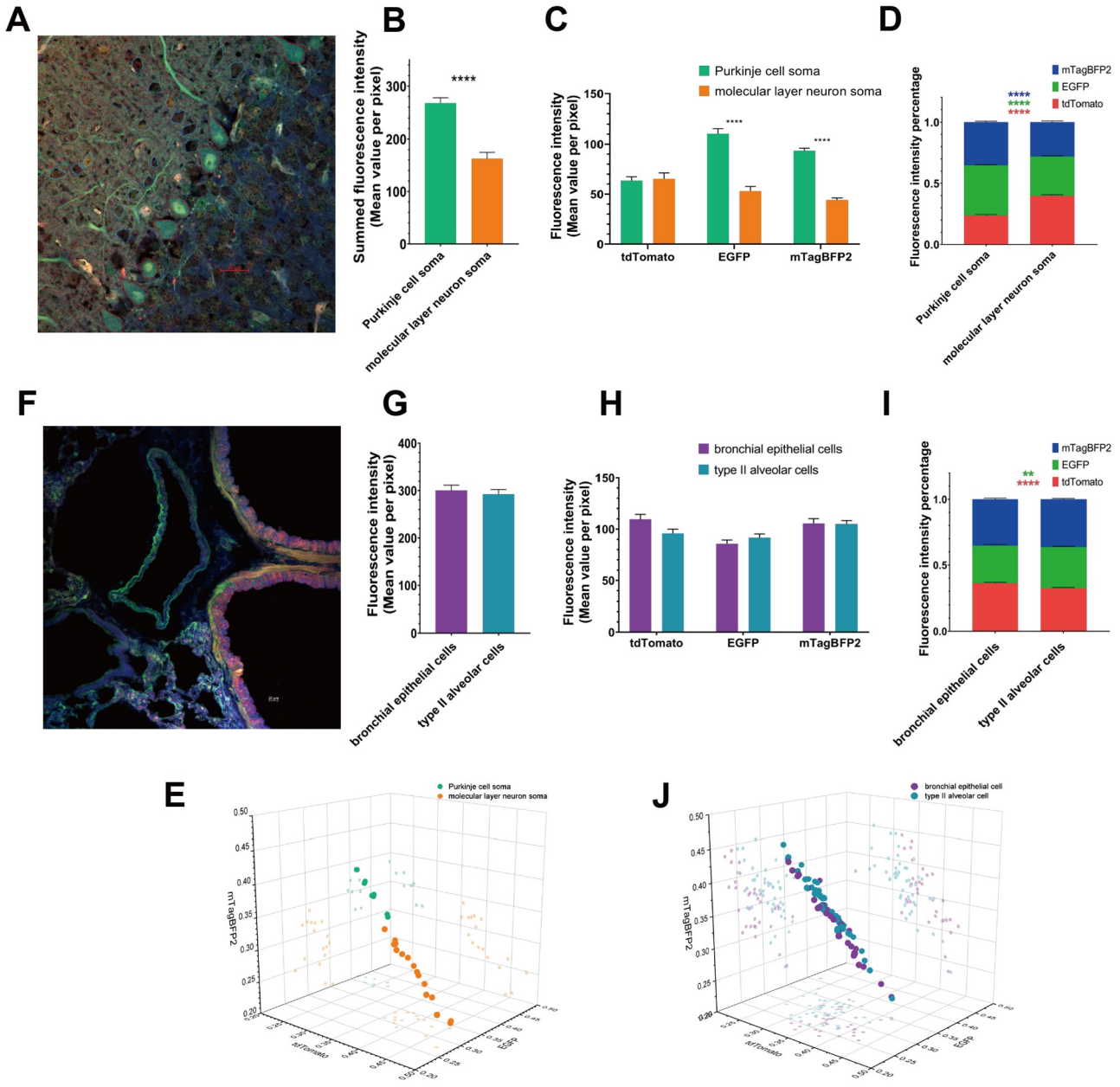

Figure 6. Images and graphs of the cerebellar cortex and lung tissue in the *RGB* mouse model (A–E) Cerebellar cortex tissue. (F–J) Lung tissue. (A,F) Fluorescence images of frozen sections of mouse tissues, with dashed lines outlining Purkinje cell somas and molecular layer neuron somas in the cerebellar cortex section (n=8, 17), as well as bronchial epithelial cells and type II alveolar cells in the lung section (n=34, 48). (B,G) The summed fluorescence values in cells. (C,H) The absolute fluorescence values in cells. (D,I) The proportion of each fluorescence value to the summed value in cells. (E,J) 3D scatter plot constructed using the fluorescence proportion of each cell. **P<0.01, ****P<0.0001.

the expression of tdTomato was strongly correlated with that of EGFP. The results in renal tubules were similar to those in hepatocytes, showing a tendency for tdTomato expression to deviate from the normal distribution model.

Discussion

In this study, we generated six mouse models with comparable transgene expression strategies involving fluorescent proteins inserted into the *ROSA26* and *H11* loci and compared transgene expressions across these models. Equal promoter and regulatory elements were employed in all transgene cassettes to exclude inconsistencies caused by different gene expression regulation elements. Given the widespread use of the *ROSA26* locus, we initially examined differences in transgene insertion orientations at this locus. Our results confirmed that in most tissues, inserting the transgene in the opposite direction of *ROSA26* gene transcription can achieve increased expression efficiency. Exceptions were also noted in Figure 2 and Supplementary Figure 2, such as the lung, where no differences in transgene protein expression were observed between different insertion strategies in 1-month-old mice (with the *R26-reverse-EGFP* group exhibiting slightly higher mean values compared to the other groups), and higher expression in the *R26-*

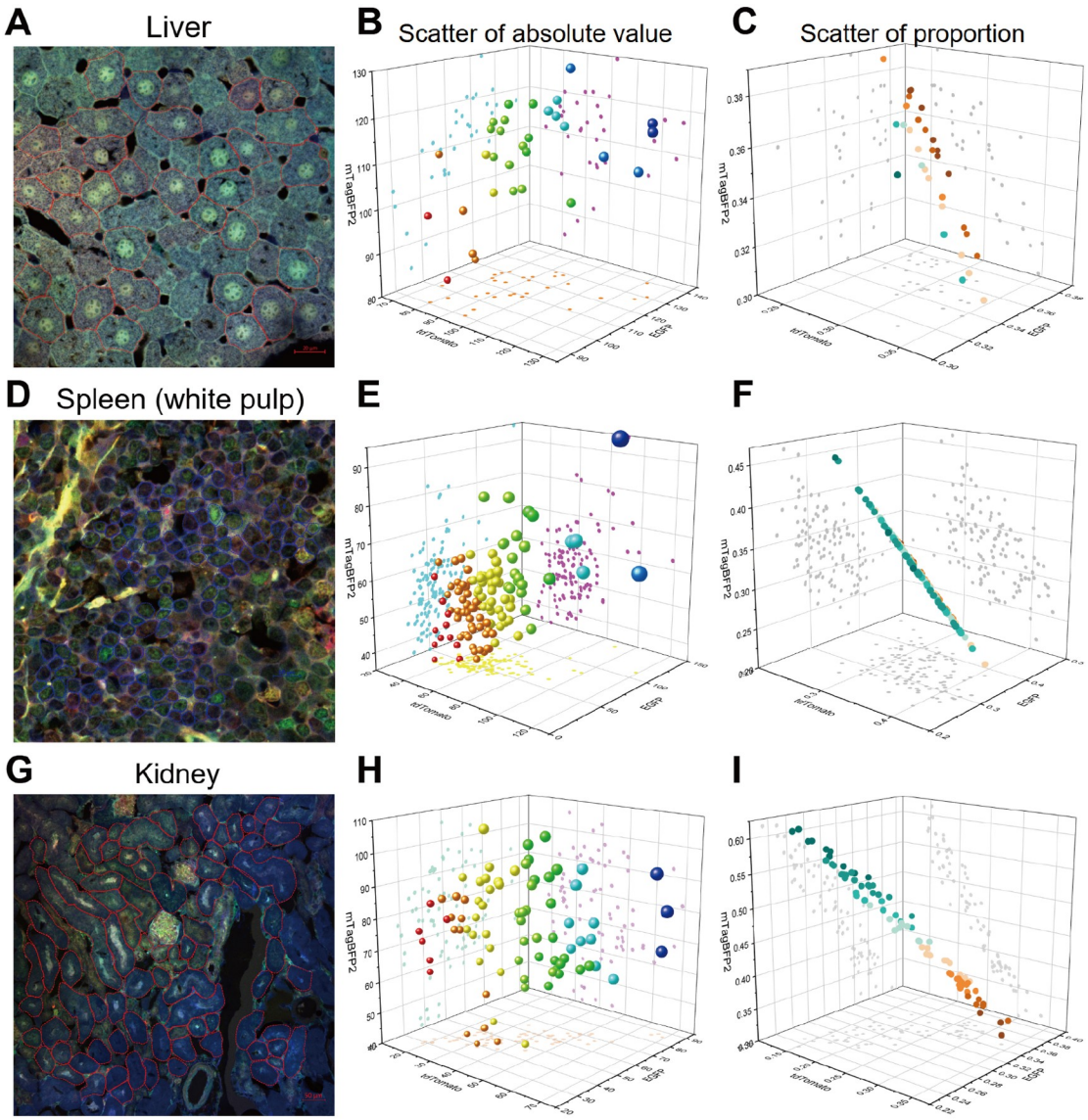

Figure 7. Images and plots of the liver, splenic white pulp and kidney tissues of the *RGB* mouse model Panels A–C derive from liver tissues (n=31), panels D–F from splenic white pulps (n=124), and panels G–I from kidney tissues (n=80). (A,D,G) Fluorescence images of frozen sections, with dashed lines outlining hepatocytes, lymphocytes, and renal tubules. (B,E,H) 3D scatter plots constructed using the absolute values of the fluorescence of each cell or renal tubule, where the dot size represents the sum value of the merged fluorescence. (C,F,I) 3D scatter plots constructed using the proportions of the fluorescence of each cell or renal tubule.

EGFP group observed in 4-month-old mice. Overall, differences in transgene expressions between insertion directions were less significant in the lung than in other tissues, such as the liver and kidney. Age-related comparisons revealed that differences in transgene expressions across different insertion strategies diminished as the mice aged. This suggests potential dynamic changes in transcription availability during various developmental periods or a top-touching effect in the high-expression group and protein accumulation in the low-expression group.

In two previous studies involving mouse embryonic stem cells, it was also found that inserting the transgene in different orientations at the *ROSA26* locus could impact expression efficiency [16,17]. However, the conclusions from these studies contradict each other. In Strathdee's study, the transgene cassette with antisense insertion at the *ROSA26* locus exhibited increased expression efficiency, while Chen's study reported slightly increased expression in the sense orientation. One common observation in both articles was that transcriptional read-through of the endogenous *ROSA26* gene may interfere with transgene expression. Our findings align more closely with those of Strathdee's study. We also detected variations in transgene expression across different tissues or in mice of different ages with different insertion orientation strategies. Our *in vivo* results provide a more direct and comprehensive reference for the construction and application of transgenic mouse models at the *ROSA26* locus.

By comparing the *R26-EGFP* model with the *H11-EGFP* model, we found that the same transgene cassette inserted into the *ROSA26* locus resulted in greater expression. Therefore, the *ROSA26* locus may be preferred for higher expression, while the *H11* locus is suitable for moderate expression. We also noticed that the

Table 3. Normality tests and coefficients of variation for fluorescence values in hepatocytes, splenic lymphocytes, and renal tubules of *RGB* mice

RGB hepatocytes				*n*=31
	tdTomato	EGFP	mTagBFP2	Sum
Shapiro-Wilk test *P* value	0.0060**	0.7733	0.0573	0.8703
Passed normality test?	No	Yes	Yes	Yes
Coefficient of variation	15.01%	10.70%	8.93%	9.09%
RGB splenic lymphocytes				*n*=124
	tdTomato	EGFP	mTagBFP2	Sum
Shapiro-Wilk test *P* value	<0.0001****	<0.0001****	0.0003***	<0.0001****
Passed normality test?	No	No	No	No
Coefficient of variation	24.94%	27.74%	18.56%	19.64%
RGB renal tubules				*n*=80
	tdTomato	EGFP	mTagBFP2	Sum
Shapiro-Wilk test *P* value	0.0314*	0.1614	0.1094	0.4340
Passed normality test?	No	Yes	Yes	Yes
Coefficient of variation	31.26%	23.82%	20.36%	17.28%

*P<0.05, **P<0.01, ***P<0.001, ****P<0.0001.

Table 4. Analysis of correlation coefficients for fluorescence values in hepatocytes, splenic lymphocytes, and renal tubules of *RGB* mice

Pearson *r*/Spearman *r* test in RGB hepatocytes				*n*=31
	tdTomato	EGFP	mTagBFP2	Sum
tdTomato	/	0.4161*	0.1121	0.6673**
EGFP	0.4161*	/	0.7476**	0.9175***
mTagBFP2	0.1121	0.7476**	/	0.7303**
Sum	0.6673**	0.9175***	0.7303**	/
Spearman *r* test in *RGB* splenic lymphocytes				*n*=124
	tdTomato	EGFP	mTagBFP2	Sum
tdTomato	/	0.4731*	0.2981	0.7688**
EGFP	0.4731*	/	0.3715*	0.8368***
mTagBFP2	0.2981	0.3715*	/	0.6445**
Sum	0.7688**	0.8368***	0.6445**	/
Pearson *r*/Spearman *r* test in RGB renal tubules				*n*=80
	tdTomato	EGFP	mTagBFP2	Sum
tdTomato	/	0.8802***	−0.1619	0.7322**
EGFP	0.8802***	/	0.1533	0.8931***
mTagBFP2	−0.1619	0.1533	/	0.5579**
Sum	0.7322**	0.8931***	0.5579**	/

Pearson *r* in blue, and Spearman *r* in red. *|*r*|>0.3, **|*r*|>0.5, ***|*r*|>0.8.

detectable differences in protein expression between the groups were less significant than those observed for mRNAs, which is consistent with the notion that mRNA levels do not always correspond to protein levels in cells.

In *RGB* trifluorescent mice, we used the principle of trinary color mixing to study the differences and proportional changes in the expression of three genes across distinct or identical tissue cells. Our research revealed substantial color disparities between transgenic fluorescent proteins in various tissue cells, reflecting varying amounts of fluorescent proteins expressed at each site in different cell types. This phenomenon can be attributed to potential differences in the chromatin environments at the insertion sites between distinct cell types. However, we observed differences in the expression ratios of the three fluorescent proteins even within cells of the same type. This cellular heterogeneity might arise from random fluctuations in gene expression or from the possibility that these cells can be further subdivided into more refined subtypes. Indeed, the existence of cellular heterogeneity has been corroborated by recent extensive single-cell sequencing data, and our results provide an intuitive spatial representation of cellular heterogeneity. Our mouse model serves as a valuable paradigm for studying such phenomena. Notably, the expressions of tdTomato and mTagBFP2, both of which are inserted at the *ROSA26* locus, did not significantly correlate across different cells. If allelic competition for expression exists, it would manifest a high negative correlation; if the alleles are co-expressed at high or low levels, it

would manifest a high positive correlation. However, our findings do not align with either of these predictions, suggesting random fluctuations in the expression efficiency of the two alleles at the *ROSA26* site. Halpern *et al*. [18] studied transcriptional bursts in mouse liver cells and suggested that expression variability can generate cell subsets that can quickly respond to environmental stimuli, which may be beneficial for rapid regulation of tissue function. The burst fraction and transcript degradation rates are tightly correlated. High transcription and degradation allow rapid changes in gene expression but can increase transcriptional noise stemming from promoter bursting. Tunnacliffe and Chubb [19] mentioned in their review about transcriptional bursts that the inactivation steps of synthetic promoters are fewer than those of endogenous promoters, resulting in more gene expression noise. The regulation of bursts is highly gene-specific for endogenous genes and is regulated by both enhancers and promoters. In the *RGB* model we studied, the inserted transgenic expression cassettes were assembled with the same promoter and enhancer elements. Therefore, the phenomenon of cellular heterogeneity may be related to the unclarified transcriptional burst dynamics of the CAG promoter in mice. This is a direction that we think has potential for further research.

In summary, in our study we utilized a widely expressed gene regulatory system and revealed that the efficacy of the *ROSA26* and *H11* loci in influencing the same expression cassette varies, with the degree of influence varying significantly across various tissues, the underlying mechanisms of which remain unclear. Further exploration is warranted to understand the effects of the *ROSA26* and *H11* loci on transgenes expressed by tissues and development-specific gene regulatory elements.

Supplementary Data

Supplementary data is available at *Acta Biochimica et Biphysica Sinica* online.

Acknowledgement

We thank Dr. Mengjie Zhang and Dr. Hua Yang for their discussions during the research program and Ms. Jiajuan Shen for her management of laboratory affairs in our lab.

Funding

This work was supported by the grants from the National Key R&D Program of China (No. 2019YFA0905900), the Science and Technology Commission of Shanghai Municipality (No. 18DZ2293500), the Shanghai Engineering Research Center for Model Organisms (No. 19DZ2280500), the Science and Technology Commission of Shanghai Municipality (Nos. 22140900102 and 201409001900) and the Shanghai Laboratory Animal Research Center (No. 2023NS01).

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Friedrich G, Soriano P. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev* 1991, 5: 1513–1523
2. Zambrowicz BP, Imamoto A, Fiering S, Herzenberg LA, Kerr WG, Soriano P. Disruption of overlapping transcripts in the ROSA β geo 26 gene trap strain leads to widespread expression of β-galactosidase in mouse embryos and hematopoietic cells. *Proc Natl Acad Sci USA* 1997, 94: 3789–3794
3. Soriano P. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat Genet* 1999, 21: 70–71
4. Mao X, Fujiwara Y, Orkin SH. Improved reporter strain for monitoring Cre recombinase-mediated DNA excisions in mice. *Proc Natl Acad Sci USA* 1999, 96: 5037–5042
5. Tchorz JS, Suply T, Ksiazek I, Giachino C, Cloëtta D, Danzer CP, Doll T, *et al.* A modified RMCE-compatible rosa26 locus for the expression of transgenes from exogenous promoters. *PLoS One* 2012, 7: e30011
6. Kobayashi T, Kato-Itoh M, Yamaguchi T, Tamura C, Sanbo M, Hirabayashi M, Nakauchi H. Identification of rat *rosa26* locus enables generation of knock-in rat lines ubiquitously expressing *tdTomato*. *Stem Cells Dev* 2012, 21: 2981–2986
7. Kong Q, Hai T, Ma J, Huang T, Jiang D, Xie B, Wu M, *et al.* Rosa26 locus supports tissue-specific promoter driving transgene expression specifically in pig. *PLoS One* 2014, 9: e107945
8. Ma Y, Yu L, Pan S, Gao S, Chen W, Zhang X, Dong W, *et al.* CRISPR/Cas9-mediated targeting of the *Rosa26* locus produces Cre reporter rat strains for monitoring Cre-*loxP*-mediated lineage tracing. *FEBS J* 2017, 284: 3262–3277
9. Ma X, Zeng W, Wang L, Cheng R, Zhao Z, Huang C, Sun Z, *et al.* Validation of reliable safe harbor locus for efficient porcine transgenesis. *Funct Integr Genomics* 2022, 22: 553–563
10. Irion S, Luche H, Gadue P, Fehling HJ, Kennedy M, Keller G. Identification and targeting of the ROSA26 locus in human embryonic stem cells. *Nat Biotechnol* 2007, 25: 1477–1482
11. Hippenmeyer S, Youn YH, Moon HM, Miyamichi K, Zong H, Wynshaw-Boris A, Luo L. Genetic mosaic dissection of Lis1 and Ndel1 in neuronal migration. *Neuron* 2010, 68: 695–709
12. Zhu F, Gamboa M, Farruggio AP, Hippenmeyer S, Tasic B, Schüle B, Chen-Tsai Y, *et al.* DICE, an efficient system for iterative genomic editing in human pluripotent stem cells. *Nucleic Acids Res* 2014, 42: e34
13. Li YS, Meng RR, Chen X, Shang CL, Li HB, Zhang TJ, Long HY, *et al.* Generation of *H11-albumin-rtTA* transgenic mice: a tool for inducible gene expression in the liver. *G3* 2019, 9: 591–599
14. Ruan J, Li H, Xu K, Wu T, Wei J, Zhou R, Liu Z, *et al.* Highly efficient CRISPR/Cas9-mediated transgene knockin at the H11 locus in pigs. *Sci Rep* 2015, 5: 14253
15. Yang H, Wang H, Jaenisch R. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nat Protoc* 2014, 9: 1956–1968
16. Strathdee D, Ibbotson H, Grant SGN, Fraser P. Expression of transgenes targeted to the *Gt(ROSA)26Sor* locus is orientation dependent. *PLoS One* 2006, 1: e4
17. Chen C, Krohn J, Bhattacharya S, Davies B, Van Wijnen A. A comparison of exogenous promoter activity at the ROSA26 locus using a PhiC31 integrase mediated cassette exchange approach in mouse ES cells. *PLoS One* 2011, 6: e23376
18. Bahar Halpern K, Tanami S, Landen S, Chapal M, Szlak L, Hutzler A, Nizhberg A, *et al.* Bursty gene expression in the intact mammalian liver. *Mol Cell* 2015, 58: 147–156
19. Tunnacliffe E, Chubb JR. What is a transcriptional burst? *Trends Genet* 2020, 36: 288–297