Developing and analysing novel tools to study endogenous WNT signalling in mice

van de Moosdijk, A.A.A.

Publication date
2021

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 5

Design, construction and validation of a novel multicolour reporter allele for \textit{in vivo} lineage tracing

Anoeska van de Moosdijk$^1$, Britt van der Swaan$^1$, Lona Kroese$^2$, Ivo Huijbers$^2$ and Renée van Amerongen$^1$

$^1$ Swammerdam Institute for Life Sciences, Section of Molecular Cytology, University of Amsterdam, The Netherlands
$^2$ Mouse Clinic for Cancer and Aging (MCCA) Transgenic Facility, The Netherlands Cancer Institute, Amsterdam, The Netherlands
Abstract

Stem cells are essential for development and tissue homeostasis. They show both differentiation and self-renewal potential and are important to maintain tissue integrity in all multicellular animals. Model organisms are key to studying the role of stem cells in building and maintaining tissue architecture in a complex 3D environment. Genetically engineered mouse models have become the golden standard to do so in mammals. Over the years, multiple genetic tools have been developed that allow the labelling of stem cells in situ as well as the long-term tracking of their developmental potential. Such lineage tracing is highly informative to investigate both the developmental fate and cell division dynamics of stem cells in vivo over time. Here we describe the design, construction and first characterization of a novel, Cre/Lox-inducible fluorescent multi-colour reporter allele. It combines aspects of several existing multi-colour mouse models, incorporates new bright fluorescent proteins and is designed to allow optimal imaging of whole-mount tissues. We call it Rosa26PRIME, for PRimary colours In the MEbrane.

Introduction

The term ‘lineage tracing’ covers multiple different approaches to identify the progeny of any living cell (Kretzschmar and Watt, 2012). It is a powerful tool to understand the emergence of spatial patterns and the growth of tissues and organs, out of a single fertilized egg or another stem cell. Developmental biologists have been trying to trace lineages since the emergence of the cell theory early 19th century, when it became clear that all cells arise from already existing cells (Conklin, 1905). By knowing the lineage of all cells, development and its defects can be understood. In addition, it is also the basis for understanding diseases such as cancer, and processes of regeneration such as wound healing.

The first lineage tree of an entire multicellular organism from fertilized egg to adult was constructed by Sulston and colleagues, who identified the lineage of all 959 (hermaphrodite) and 1033 (male) cells in the C. elegans embryo by observation and manual tracking via light microscopy (Sulston et al., 1983).

In more complex multicellular animals, such as the mouse, resolving the entire cell lineage by light microscopy is not possible because the organism is simply too large and not transparent, creating the need for labelling techniques to track the progeny of specific cells of interest. A solution to this problem was found by introducing radioactive dyes and later fluorescent dyes, such as fluorescein, to cells.
A novel multicolour reporter allele for *in vivo* lineage tracing

However, these are diluted upon cell division and were found to spread to neighbouring cells not directly part of the lineage, as reviewed by (Kretzschmar and Watt, 2012).

Since these initial labelling techniques, many methods have been developed to provide insight on lineage relationships and on the contribution of different cell populations to maintain tissue homeostasis. One experimental approach is to transplant tissue from one organism or embryo to another, but this can influence the developmental potential of cells, leading to different behaviour as compared to the normal physiological situation (van Keymeulen et al., 2011; Watt and Jensen, 2009). Therefore, methods allowing *in situ* analysis, where cells are left in their natural context, are preferred.

**Genetic labelling of cell lineages with fluorescent proteins**

Robust, non-invasive, *in situ* labelling was made possible by using genetically encoded markers, as was first demonstrated via retroviral infection to track the fate of mouse hematopoietic stem cells (Lemischka et al., 1986). By stably introducing the genetic code for a specific marker in the DNA of a cell, the label became permanent and thus heritable.

A breakthrough came with the use of green fluorescent protein (GFP), which could be used as a genetic marker *in vivo* in both the prokaryote *E. coli* and the eukaryote *C. elegans* (Chalfie et al., 1994). Unfortunately, the original GFP protein – originally isolated from the jellyfish *Aequorea victoria* - turned out to be very dim in mouse cells (Prasher et al., 1992). Optimization only took a few years and resulted in the creation of the MmGFP variant, which shows better maturation at 37°C due to the introduction of several mutations that improved folding and altered the chromophore structure (Zernicka-Goetz et al., 1997). It has four mutations introduced for this goal: F64L, S65T, V163A and S175G, and was codon-optimized for use in the mouse. This MmGFP was bright enough to make imaging of green fluorescent proteins in mouse embryos possible. Not long after, the first genetic fluorescent mouse reporter strain was published, containing another improved GFP variant that was well expressed in mice and that is still used widely in the mouse field today: EGFP (Hadjantonakis et al., 1998). This variant has 3 mutations, F64L, S65T and H231L, and is codon-optimized according to human codon usage. Since the generation of this first fluorescent reporter mouse strain, more have been created in various colours of the available fluorescent rainbow, from blue to green, yellow and red (Abe and Fujimori, 2013; Li et al., 2018).

Lineage tracing requires experimental control of the induction of reporter gene activity, such that it is ‘switched on’ at a specific developmental timepoint or within a subset of cells. The possibility for such inducible genetic labelling emerged with the use of genetic recombination technology in the 1990s. In mice, the CreERT2/Lox
system is the most used recombinase system for lineage tracing, realizing spatio-temporal control over the cells to trace.

**Inducible labelling of cell lineages by Cre/lox recombination**

After its initial discovery in bacteriophage P1 (Sternberg and Hamilton, 1981), the Cre/Lox system was first implemented in mice to delete the promoter and first exon of the DNA polymerase beta gene in T cells, which is lethal if deleted in the mouse germline (Gu et al., 1994). Since then, Cre/Lox has become widely adopted to achieve selective (in)activation or mutation of genes of interest. For lineage tracing purposes, the Cre recombinase is typically expressed from a tissue- or lineage-specific promoter. This so-called ‘Cre driver’ mouse strain is crossed with a second mouse strain carrying an inducible reporter, the expression of which is conditional upon the Cre-mediated recombination (i.e. inversion or deletion) of a cassette flanked by Lox sites. Lox sites can flank any genetic cassette, such as the often-used translational roadblock (referred to as ‘Lox-STOP-Lox’).

A versatile reporter strain should allow lineage tracing of different cell types in multiple settings. The Rosa26 locus is most often used for this purpose since it allows efficient targeting and ubiquitous expression of the integrated cassette. The locus was first used in combination with a β-galactosidase reporter, showing expression in all embryonic tissues when driven by the endogenous Rosa26 promoter (Soriano, 1999). The first Rosa26 fluorescent reporter mouse strain was Rosa26EGFP (Mao et al., 2001). In the same year, reporter strains Rosa26EYFP and Rosa26ECFP also became available (Srinivas et al., 2001). The next step came by introducing not one, but two fluorescent markers in the form of the Rosa26mTmG reporter mouse, which expresses a red fluorescent protein (tdTomato) prior to Cre-mediated recombination and a green fluorescent protein (EGFP) afterwards (Muzumdar et al., 2007). In this case, expression is no longer driven by the endogenous Rosa26 promoter, but by a chicken beta-actin core promoter with a CMV enhancer (pCAG) to achieve stronger expression. In addition, the membrane localization of both fluorescent reporter proteins facilitates imaging of single cell architecture in complex tissues.

While functional, the use of only a single reporter colour to label all switched cells makes it difficult, if not impossible, to distinguish individual clones amidst labelled patches of cells. Although this can be solved by performing all labelling experiments at clonal density (Lloyd-Lewis et al., 2018), another solution lies in the use of multicolour reporter strains. The presence of multiple fluorescent colours, each of which is induced stochastically, provides enhanced resolution since it allows the offspring of individually recombined (‘switched’) cells presenting as different colours to be confidently discriminated, thus providing clonal information even at higher initial labelling frequencies.
Building a brighter and better multicolour reporter

Two well-known multicolour reporters are the Thy1-Brainbow (Cai et al., 2013; Livet et al., 2007) and Rosa26\(^{Confetti}\) (Snippert et al., 2010a) mouse reporter strains. Consecutive generations of Brainbow have been updated with brighter fluorescent proteins (mOrange2, EGFP and mKate2 in the Brainbow 3 reporters) (Cai et al., 2013), but all are expressed under the control of the Thy1 promotor, enabling marking of neuronal cell lineages only. For studying other cell types or tissues in the embryo or postnatally, the Rosa26\(^{Confetti}\) (Snippert et al., 2010a) and Rosa26\(^{Rainbow}\) (Rinkevich et al., 2011) mouse strains have proven useful. For instance, the Rosa26\(^{Confetti}\) reporter strain has allowed elucidation of the stem cell division properties in the intestinal crypt (Snippert 2010). One disadvantage of this reporter is that cells do not express a label prior to switching, which limits information on tissue architecture in imaging experiments.

The aim of this study was to combine the best features of existing multicolour fluorescent reporter strains (depicted in Figure 1a): First, to use bright fluorescent proteins that can be expressed conditionally in all cell types. Second, to ensure labelling of cells both before and after switching. Third, to provide information about tissue architecture and clone characteristics by targeting the proteins to the plasma membrane in switched cells. Finally, we used this opportunity to test two novel fluorescent proteins, mScarlet and iRFP670, for use in vivo for the first time. Taking these variables into consideration, we designed a multicolour reporter allele that shows ubiquitous expression of a nuclear label before Cre-mediated recombination, and membrane labelling of irreversibly switched cells to increase positional information by essentially outlining the cells after recombination. The resulting reporter strain will allow lineage tracing of cell clones in blue, yellow and red colours and is therefore named PRIME: for Primary colours in the membrane.

Results

Selecting the colours

To allow multicolour, wholemount imaging of intact tissues, we need fluorescent proteins (FPs) that are bright, have minimal spectral overlap, display good photostability, and which survive standard histological fixation procedures with minimal loss or alteration of fluorescence emission. Ultimately, we opted to find the best possible combination of a cyan, yellow, red and far-red fluorescent protein. A set of these fluorescent proteins will have minimal spectral overlap, while providing the option to combine the reporter with the most widely used FP: any green fluorescent protein (GFP). GFP does show some overlap with cyan and yellow FPs,
but despite this they can be imaged together and distinguished in tissues on most confocal setups, as shown by the widely used *Rosa26Confetti* strain that combines green (hriGFP), yellow (EYFP), and cyan (mCerulean) fluorescent proteins (Livet et al., 2007). In fact, the *Rosa26PRIME* allele was designed to be used in combination with the *Axin2P2A-rtTA3-T2A-3xNLS-EGFP2* strain from chapter 4 (van de Moosdijk et al., 2020).

**Figure 1: Multi-colour fluorescent mouse reporter strains.** a) Schematic representation of existing multi-colour fluorescent mouse reporter strains. *Rosa26mTmG* (Muzumdar et al., 2007): cells are labelled with membrane-bound tandem dimer Tomato (M-tdTom). Upon Cre/Lox recombination, cells switch to a membrane-localized EGFP (M-EGFP). *Thy1-Brainbow3.2* (Cai et al., 2013): All cells with an active Thy1 promoter, mostly differentiated neurons, express a nuclear non-fluorescent YFP (N-ØNFP). Cre/Lox recombination leads to stochastic expression of one of three fluorescent proteins (FPs): mOrange2, EGFP or mKate2. *Rosa26Confetti* (Snippert et al., 2010a): Cells do not express any FP before switching. Upon Cre/Lox recombination cells switch stochastically to a nuclear GFP (N-hriGFP), yellow FP (EYFP), red FP (tdimer2) or the membrane-localized cyan FP mCerulean (M-mCer). *Rosa26Rainbow* (Rinkevich et al., 2011): no fluorescent protein is expressed. After Cre/Lox recombination cells stochastically express a red FP (mCherry), orange FP (mOrange), cyan FP (mCerulean) or green FP (EGFP). b) Cartoon illustrating cells expressing nuclear infra-red fluorescent protein (iRFP670) c) and the three possible recombination outcomes (red, cyan, yellow) with membrane-bound fluorescent protein (mScarlet-I, mTurquoise2, mVenus).
To meet the demand of marking all cells prior to switching, we decided to search for a far- or infra-red FP. This choice was made because wavelengths toward the infrared part of the spectrum penetrate deeper into tissues. We reasoned that this would not only be of use when visualizing wholemount preparations but might also open new opportunities for intravital imaging studies. Unfortunately, currently available far- or infra-red FPs are relatively dim compared to the shorter wavelength FPs. Therefore, we decided to add a nuclear localization signal to concentrate the signal in the nucleus. We used a triple Nuclear Localization Signal (3xNLS), identical to the one used in chapter 4, since it shows less cytoplasmic retention than a 1xNLS sequence (Chertkova et al., 2017). The 3xNLS signals also shows enhanced labelling of nucleoli, which comes in handy when one needs to discriminate specific signal from autofluorescence.

For the FPs expressed after switching, we decided to add a membrane tag similar to the strategy used in the Rosa26mTmG mouse to easily distinguish individual cells and outline their shape. The membrane localization domain MARCKS works well in this mouse and therefore we decided to use the same sequence (Muzumdar et al., 2007). This membrane localization sequence is the lipid anchor domain from the myristoylated alanine-rich C kinase substrate (MARCKS) protein (Ohmori et al., 2000; De Paola et al., 2003). Only the short N-terminal membrane localization domain of eight amino acids is needed. This membrane localization works best with monomeric fluorescent proteins, while many bright RFPs are tetramers or dimers (Chudakov et al., 2010). Thus, it is important that the FPs combined with the MARCKS localization marker in our reporter are highly monomeric to prevent aggregates in the cells. The pre- and post-recombination status of our reporter design is depicted in Figure 1b,c: all cells will show nuclear far-red labelling prior to Cre-mediated recombination (Figure 1b), but upon switching cells will lose expression of the nuclear marker and stochastically express a membrane-localized red, cyan or yellow FP (Figure 1c). In theory, recombination of the different fluorescent proteins should occur in a 1:1:1 ratio.

Since it is challenging and time-consuming to create mouse models, we selected candidate FPs based on prior performance in existing fluorescent mouse lines as described in the literature. Where these were not available, we looked for the performance of FPs in mammalian cell lines in vitro, to make sure that the selected FPs would perform well at 37°C. Candidates were further selected to fit available laser and filter settings at our and other commonly used confocal microscopes, to make sure that imaging conditions would allow optimal detection of the FPs in different labs.

**Finding cyan and yellow FPs with high photostability**

FP performance depends on its inherent brightness as well as its photostability. No matter how bright the FP is, if it isn’t photostable, it will bleach before a good image
can be taken. Therefore, we explicitly considered bleaching characteristics when choosing our candidate proteins.

It is possible to compare the practical photostability of FPs by calculating the relative photostability, a normalized value that states how long an FP will take to lose 50% of an initial emission rate of 1,000 photons/s (Shaner et al., 2005). However, FPs in the same spectral class can show different bleaching rates at different laser powers or imaging conditions (Cranfill et al., 2016). Because of that, a single calculated value does not always represent what will happen in given experimental setup, since the light dose can differ. Many studies have tested photostability in vitro or in bacteria, for example as in the overview by (Shaner et al., 2005), but their behaviour in mammalian cells can be different, as shown for orange and red fluorescent proteins (Bindels et al., 2016; Shaner et al., 2008). Therefore, we checked for photostability under conditions resembling our conditions for imaging wholemount tissues.

The performance of cyan fluorescent proteins in mammalian cells was tested in HeLa cells during the design and characterization of mTurquoise2 (mTq2) (Goedhart et al., 2012). mTq2 stands out for not only being the brightest cyan monomeric fluorescent protein available, but also for being highly photostable compared to the mCerulean family of fluorescent proteins (Goedhart et al., 2012; Kremers et al., 2006). Therefore, mTq2 seemed an obvious choice for our cyan coloured reporter protein and should result in a brighter cyan reporter signal than Rosa26Confetti and Rosa26Rainbow, which use mCerulean.

For the yellow part of the spectrum, several bright and monomeric yellow fluorescent protein (YFP) variants are available. They have been used for imaging in mammalian cells, but their photostability is relatively low (Nowotschin et al., 2009; Shaner et al., 2005) and in practice they bleach rapidly. This is a special problem when the goal is to image thick tissues, in which case large Z-stacks need to be imaged, requiring relatively long exposure to the laser. Therefore, we spent extra efforts to find the most stable candidate. Several YFPs were identified as suitable candidates, including Ypet (Nguyen and Daugherty, 2005), PhiYFP (Shagin et al., 2004), mVenus ((Kremers et al., 2006)) and SYFP2 (Kremers et al., 2006). Ypet is one of the brightest YFPs in HeLa cells (Joachim Goedhart; personal communication), but it is not monomeric. PhiYFP would be an antigenically distinct yellow variant distinct from the other FPs and should be relatively bright in vivo (Shagin et al., 2004). Unfortunately, it is a dimer and quenches upon PFA fixation (Cai et al., 2013). Of the monomeric YFPs, mVenus (mVen) is most frequently used in transgenic mouse models. An improved variant is available in the form of SYFP2 (Kremers et al., 2006). Therefore, we considered these to be our top candidates.

SYFP2 was reported to be slightly brighter than mVen in HeLa cells (1.4x the brightness of EYFP for SYFP2 versus 1.3x the brightness of EYFP for mVen (Kremers et al., 2006)). However, it bleaches faster: at a rate relative to EYFP of 0.7 for SYFP2
and 1 for mVen (a higher number means higher photostability) (Kremers et al., 2006). To test which of the two would perform best in longer imaging experiments in our hands, we transfected mouse mammary epithelial BC44 cells (Deugnier et al., 1999) with MARCKS-mVen (M-mVen) and MARCKS-SYFP2 (M-SYFP2) and imaged them continuously for 10 min (Figure 2a). We started with a relatively low laser power (5%) so that no saturated pixels were visible to allow quantification. At this low laser power, both M-mVen and M-SYFP2 lost about 20% of their brightness in 10 minutes (Figure 2a).

Based on our imaging experience with other fluorescent mouse models, however, we expected that for wholemount tissue imaging we would need higher laser power, due to thickness and optical complexity of the tissue, and the anticipated lower expression levels of the genetic reporter as compared to a monolayer of transfected cells. Next to that, when scanning big Z-stacks, averaging options can be limited due to the time required to scan all planes. Therefore, we repeated the experiment with a higher laser power (15%) in the expected range for whole-tissue imaging (Figure 2b). This resulted in increased bleaching for both YFPs. Of note, SYFP2 bleached until only 26% of the initial brightness was left, whereas mVen retained 41% brightness after 10 minutes. Even though SYFP2 was slightly brighter to start with, after 10 minutes of imaging the signal had become so dim that cells were barely visible (Figure 2c). Therefore, we decided to stick with the widely used mVen as our YFP of choice for the PRIME reporter construct.

**Far into the red fluorescence**

Far-red and infrared fluorescent proteins are often used for deep optical imaging because in this part of the spectrum the absorption by hemoglobin, melanin and water in mammalian tissues is minimal (Weissleder, 2001). Moreover, scatter and autofluorescence play a relatively small role at these higher wavelengths (Andersson et al., 1998).
Figure 2: bleaching characteristics of FPs in mammalian cells. a) Graph showing relative intensity of membrane-tagged yellow fluorescent proteins (M-YFPs) in BC44 cells when imaged continuously for 500 seconds at low laser power (5%; no saturated pixels in cell monolayer) and high laser power (15%; estimated power needed for imaging wholemount tissue). Two independent biological experiments were performed. The number of cells imaged are the combined cell numbers from both experiments. mVen low n=15, SYFP2 low n=11. mVen high n=15, SYFP2 high n=18. b) Confocal image showing representative M-mVen (top) and SYFP2 (bottom) signal at t=0 s and t=500 s imaged with high laser power (15%). c) Graph showing relative intensity of membrane-tagged red fluorescent protein (M-RFP) mSc-I in BC44 cells imaged continuously for 500 seconds at laser power similar to imaging RFPs in wholemount tissues (8%) based on imaging mTmG wholemount mammary gland tissue with same settings. Two independent biological experiments were performed. The number of cells imaged are the combined cell numbers from both experiments. N=28. d) Graph showing relative intensity of far-red fluorescent protein (FRFP) iRFP670 in BC44 cells imaged continuously for 500 seconds at low laser power (no saturated pixels in cell monolayer; 30%) and high laser power (estimated power needed for imaging wholemount tissue; 65%). Two independent biological experiments were performed. The number of cells imaged are the combined cell numbers from both experiments. iRFP670 low n=9 iRFP670 high n=14. e) Image showing representative M-mSc-I signal at t=0 s and t=500 s. f) Image showing representative iRFP670 signal at t=0 s and t=500 s imaged with high laser power (65%).
The biggest challenge in finding a suitable far-red FP is identifying one that has a high enough fluorescence intensity for wholemount imaging of tissues using a regular confocal microscopy set up as available in most labs. We initially checked the far-red fluorescent protein mCardinal (Chu et al., 2014), but this showed too much spectral overlap with RFPs. High bleed trough between the red and far-red channels therefore disqualified this and other far-red FPs, so we looked for near infra-red FPs. The near infra-red eqFP650 was tested as part of improvements to the Brainbow toolbox, but turned out to be too dim (Cai et al., 2013). In the same year, a set of bacterial phytochrome photoreceptors was used as template to engineer several near-infrared FPs specifically intended for multi-colour in vivo imaging (Shcherbakova and Verkhusha, 2013). The most promising of this set is iRFP670, since it shows the highest brightness in HeLa cells (Shcherbakova and Verkhusha, 2013). Moreover, it is easily distinguishable from an RFP, while fitting common laser lines and filter settings found in most microscopes, with optimal excitation around 640 nm.

We tested how iRFP (Filonov et al., 2011), iRFP670 (Shcherbakova and Verkhusha, 2013) and miRFP670 (Shcherbakova et al., 2016) performed in mammalian cells in our hands. We briefly considered to also include the near infrared smURFP (Rodriguez et al., 2016), but in HeLa cells the brightness of smURFP appears 53-fold lower than miRFP670 without addition of the co-factor biliverdin (Shemetov et al., 2017), so we decided against its use for this reason. Following transfection and confocal imaging in BC44 cells, iRFP was so dim that we could not detect the signal above background (data not shown). Both iRFP670 and miRFP670 could be detected. Because fusion to the 3xNLS signal peptide does not require a monomeric protein, we decided to continue with iRFP670, since it was originally reported to be brighter in HeLa cells than its monomeric variant miRFP670 (Shcherbakova et al., 2016).

To test the behaviour of iRFP670 upon prolonged imaging, we again measured the decrease in fluorescent signal during a 10-minute time frame using both a lower laser power (30%, no saturated pixels), and a higher laser power (65%) (Figure 2e). At lower power, 51% of the initial signal remained, while at higher laser power this was 27% (Figure 2g). Thus, bleaching of iRFP670 was in the same range as the bleaching of FPs in the yellow and red parts of the spectrum that we had tested (Figure 2b,d), so we inferred that iRFP670 should be stable enough to use for imaging in mouse tissues.
Chapter 5

**Figure a:**
- Graph showing fixation (min) vs. intensity (a.u.) for tdTom, T2A, and mTq2.

**Figure b:**
- Graph showing fixation (min) vs. ratio of tdTom/mTq2.

**Figure c:**
- Images showing fluorescence at 0 min, 5 min, 10 min, 15 min, 30 min, and 60 min for mTq2 and tdTom.

**Figure d:**
- Graph showing fixation (min) vs. intensity (a.u.) for mSc-I, T2A, and mTq2.

**Figure e:**
- Graph showing fixation (min) vs. ratio of mSc-I/mTq2.

**Figure f:**
- Images showing fluorescence at 0 min, 5 min, 10 min, 15 min, 30 min, and 60 min for mTq2 and mSc-I.

**Figure g:**
- Graph showing fixation (min) vs. intensity (a.u.) for 3xNLS-iRFP670 and iRFP670.

**Figure h:**
- Images showing fluorescence at 0 min, 5 min, 10 min, 15 min, 30 min, and 60 min for NLS-iRFP670.
Figure 3: fluorescent signal in cells after fixation with PFA. a) Graph showing median intensity signal of mTq2 and tdTom expressed together from one vector, cleaved via a T2A sequence between the FPs. HeLa cells were fixed with 4% PFA for the time stated and imaged using the same settings. Data is from two separate transfections, imaged on two different days. Number of cells measured at 0 min n=13; 5 min n=25; 10 min n=23; 15 min n=21; 30 min n=23; 60 min n=22. b) Graph showing ratio of tdTom/mTq2 per measured cell for different fixation times, with horizontal red lines showing the median. c) Representative confocal images showing mTq2 and tdTom signal in cells fixed for 0 – 60 minutes. d) Graph showing median intensity signal of mTq2 and mSc-I expressed together from one vector, cleaved via a T2A sequence between the FPs. HeLa cells were fixed with 4% PFA for the time stated and imaged using the same settings. Data is from two separate transfections, imaged on two different days. Number of cells measured at 0 min n=13; 5 min n=18; 10 min n=29; 15 min n=22; 30 min n=26; 60 min n=25. e) Graph showing ratio mSc-I/mTq2 per cell for different fixation times, with horizontal red lines showing the median. f) Representative confocal images showing mTq2 and mSc-I signal in cells fixated for 0 – 60 minutes. g) Graph showing intensity signal of 3xNLS-iRFP670 expressed via transfection into HeLa cells. Error bars represent SEM. Cells were fixed with 4% PFA for the time stated and imaged using the same settings. Number of cells measured at 0 min n=11; 5 min n=19; 10 min n=22; 15 min n=18; 30 min n=22; 60 min n=17. h) Representative confocal images showing median 3xNLS-iRFP670 signal in cells fixated for 0-60 minutes.

Preservation of fluorescent signal upon fixation
A major difference between imaging cells and tissues, is that cells can be kept alive relatively easily while imaging, whereas tissues are typically fixed to stop post-mortem changes and optimally preserve cellular details. Unfortunately, the fluorescence of an FP can be partially lost upon fixation (Schnell et al., 2012). Both mVen and mTq2 retain a bright fluorescent signal upon aldehyde fixation (Joosen et al., 2014 and Joachim Goedhart, personal communication). We therefore tested whether our RFP and iRFP candidate proteins would also retain their signal upon paraformaldehyde fixation.

To test the preservation of the fluorescent signal of mScarlet-I and iRFP670 upon fixation, we used a construct that allowed expression of the FP of interest and a reference FP (here: mTq2) from a single, bicistronic transcript. The two proteins are separated posttranslationally by a 2A cleaving sequence (Goedhart et al., 2011). Inclusion of a reference protein allows normalization of the fluorescent signal on a per cell basis, thus negating cell-to-cell heterogeneity in expression.

As a control, we first measured the effect of paraformaldehyde fixation on the fluorescent signal of tdTom, which we know retains sufficient fluorescent signal in our standard histological fixation and clearing protocols (Figure 3a). HeLa cells were transfected with a tdTom-T2A-mTq2 expression construct and fixed 48 hours later with fresh 4% Paraformaldehyde (PFA) for different durations. Fixation times were
chosen up to 60 minutes, as is regularly used for fixing mouse tissues. The intensity of both tdTom and the reference mTq2 decreases after 10 minutes fixation. We quantified this by measuring the total intensity in the images and the signal decreases by approximately 65% after 10 minutes, but longer timepoints do not seem to further decrease the brightness. Both show a similar pattern, so the ratio between tdTom and mTq2 appears relatively stable upon longer fixation (Figure 3b,c). This fits with our experience of using these proteins in wholemount tissue imaging experiments (chapter 3 and 6).

Next, we transfected cells with a mSci-T2A-mTq2 construct (Figure 3d). The initial intensity of mSc-I relatively to mTq2 was higher, but it decreased when cells were fixed for more than 5 minutes. The median ratio of mSc-i/mTq2 decreased from 1 to 0.54 over time due to the larger drop in signal intensity of mSc-I than mTq2 (Figure 3e). Even though mSc-I was less stable upon fixation than tdTom, sufficient fluorescence was preserved for imaging purposes (Figure 3f). We thus decided to continue with mSc-I rather than dTom, not only because it would reduce the size of the targeting construct, but also because it would be the first time this new red fluorescent protein could be tested in vivo.

We did not generate an iRFP670-T2A-mTq2 construct because we had no other near infra-red candidate to compare it to. Instead, we transfected HeLa cells with a 3xNLS-iRFP670 construct and measure the intensity of the signal after different fixation times (Figure 3g). In this experiment, the fluorescence signal remained well visible in the nuclei to detect above background even after 60 minutes of fixation (Figure 3h). It seems that at shorter fixation time points (<30 minutes), some of the nuclear signal may leak out of the nucleus into the cytoplasm.

At the moment of designing our reporter construct, to our knowledge none of the Scarlet and iRFP670 protein family members had been used as genetic reporter in mice, so their inclusion in our reporter construct was a bit of a gamble.

Construction and functional validation of the targeting vector

The goal for our reporter mouse is to have strong expression of the transgene cassette in all tissues. Several constitutive promoters are available, such as the X-linked phosphoglycerate kinase-1 locus (PGK) promoter, the human cytomegalovirus (CMV) immediate early promoter enhancers, and CAG: a combination of the minimal chicken beta-actin gene promoter (CAG) followed by the first exon and intron of beta-actin and a splice acceptor, fused to the CMV enhancer (Hitoshi et al., 1991).
A novel multicolour reporter allele for in vivo lineage tracing

Figure 4: Testing the genetic make-up of Rosa\textsuperscript{PRIME} mice. a) Schematic overview of PRIME genetic construct inserted into the Rosa26 locus. Three pairs of heterospecific Lox sites are inserted separating four fluorescent proteins: a nuclear near-infrared fluorescent protein (3xNLS-iRFP670), membrane-bound red fluorescent protein (M-mSc-I), membrane-bound cyan fluorescent protein (M-mTq2), membrane-bound yellow fluorescent protein (M-mVen). iRFP670 is constitutively expressed in all cells. Upon Cre activation the Lox sites recombine to create three mutually exclusive excision possibilities (1, 2, 3 in the figure). The fluorescent protein that is placed first after the CAG promotor is expressed. b) Excitation (top) and emission (bottom) spectra of the four fluorescent proteins of the PRIME construct. Spectral data from FPbase.org Spectra Viewer on 03-01-2020. c) Confocal images showing HeLa cells 24h after transient transfection with the PRIME construct and a vector expressing Cre, imaged for all four fluorescent proteins. The three cells depicted show several copies of the PRIME construct, and therefore all four fluorescent protein options (NLS-iRFP670, M-mSc-I, M-mTq2 and M-mVen) can be found in the cells.
Expression of PGK is widespread, but not uniform (McBurney et al., 1994). The CMV promoter has the tendency to become transcriptionally silenced in mammalian cells (Prösen et al., 1996). Therefore we followed the approach that has been successful in other fluorescent reporter mice, such as \textit{Rosa26}\textsuperscript{mTmG} (Muzumdar et al., 2007) and \textit{Rosa26}\textsuperscript{Confetti} (Snippert et al., 2010a). Both these strains use the CAG promoter, so we chose this promoter for our PRIME construct as well (Figure 4a). As the basis for our PRIME construct we used the pCAG-mTmG vector (Muzumdar et al., 2007), in which we essentially exchanged the tdTom-GFP cassette for our iRFP670-mScarlet-I-mTq2-mVenus cassette.

To ensure that only one fluorescent protein can be expressed at any one time, we cloned an SV40 polyadenylation signal downstream of each open reading frame. The different open reading frames are further separated by variant lox sites, analogous to Brainbow.

The final PRIME construct thus contains four fluorescent proteins and is designed to work as follows: 3xNLS-iRFP670 is constitutively expressed in all cells and should allow easy orientation when imaging tissues. For lineage tracing purposes, one out of the remaining three different fluorescent proteins (membrane-bound mSc-I, mTq2 or mVen) becomes stochastically expressed after Cre-mediated recombination of either the LoxP, Lox2272 or LoxN sites, resulting in the expression of mSc-I, mTq2 or mVen, respectively (Figure 4a; options 1,2,3). We based the order of Lox sites on the Brainbow3.2 construct (Cai et al., 2013). The excitation and emission spectra of the different fluorescent proteins of our PRIME construct show sufficient separation to reliably separate the different channels while imaging (Figure 4b).

The targeting vector for PRIME was cloned step-by-step (i.e. one FP at a time), which allowed us to check functionality of the individual Lox sites \textit{in vitro} as we were assembling the construct. To check if the four-colour PRIME reporter works as designed, the final pCAG-PRIME construct was introduced into HeLa cells together with a plasmid expressing Cre (Figure 4c). Since we transiently transfected the cells, multiple copies of the PRIME construct are present per cell, explaining why we could find cells with both switched and non-switched constructs, sometimes even expressing all four PRIME colours in one cell. From these \textit{in vitro} tests, we concluded that the construct was functional, allowing us to continue generation of the \textit{Rosa26}\textsuperscript{PRIME} mouse line.

\textbf{ES cell targeting and testing of the PRIME allele}

The CAG-PRIME construct of Figure 4a was cloned into the standard Rosa26 targeting vector pROSA26-1 (Soriano, 1999), which contains 1 and 4kb homology arms. Finally, an FRT-Neo-FRT cassette was added for selection during the ES cell targeting procedure. The resulting PRIME targeting vector is 17,28bp (Supplemental figure 1).
The linearized targeting vector was introduced in B6_ColA_RMCE embryonic stem (ES) cells via electroporation (Figure 5a). Selection with neomycin and diphtheria toxin allowed positive selection for transfected cells and negative selection to eliminate clones with a random integration, respectively. Of 600 ES clones screened, 20 tested positive by PCR. Ten of these were further analysed, of which four showed a correct integration.

While generation of the Rosa26\textsuperscript{PRIME} mouse strain was in progress, we tested switching of the integrated reporter in targeted Rosa26\textsuperscript{PRIME} ES cells (ES\textsuperscript{PRIME}) cells by transfecting them with a Cre expression plasmid (Figure 5b). The MARKCS localization marker, which looked quite diffuse in the relatively flat BC44 cells (Figure 2), clearly outlines the cell membranes in the compact ES cell colonies (Figure 5b).

These Rosa26\textsuperscript{PRIME} ES cells only carry one targeted copy of the PRIME reporter and therefore independent ES cell colonies should express only one fluorescent colour. In the absence of Cre, all ES cells showed nuclear IRFP670 as expected (Figure 5b; colony 1). After Cre-mediated recombination, all three fluorescent colours could be found in separate colonies (Figure 5b; colony 2, 3 and 4).

When imaging the ES cell colonies, we got the impression that mSc-I colonies seemed more abundant than mTq2 and mVen colonies. To quantify the ratios of Cre/Lox recombination, we also analysed the Cre-transfected ES\textsuperscript{PRIME} cells by FACS to determine the prevalence of mSc-I, mTq2 and mVen signal in the total population (Figure 5c). With the gating strategy applied, almost all untransfected cells (99.7%) were scored as IRFP-positive. A small proportion of the cells transfected with Cre (5.7%) showed loss of the IRFP signal. This difference was detected in a total of 6 independent Cre transfections (from n=2 repeat experiments, Figure 5e). These IRFP negative cells should have switched to one of the other fluorescent colours. To test this, we analysed which proportion of the IRFP-negative cells showed either red, cyan or yellow fluorescence (Figure 5f). In agreement with our observations by confocal microscopy (5b), most cells showed red fluorescence, followed by cyan and then yellow. The ratio between the different colours show that following Cre transfection in ES cells recombination to mSc-I : mTq2 : mVen occurs at about 4 : 2 : 1.
Chapter 5

a

Homology arms

Introduce DNA into ES cells

Select for cells with correct integration

Blastocyst

Inject targeted ES cells

Mosaic blastocyst

Transfected

Pseudo-pregnant

Mosaic founders

Breed and check for germ line transmission

b Colony 1 Colony 2 Colony 3 Colony 4

10 μm

c Transfection

ESPROTE 

Cre plasmid

min 72 hours

FACS

d control + Cre

iRFP fluorescence

0 10^3 10^5 10^7 10^9

-10^3 -10^5 -10^7 -10^9

0 50K 100K 150K 200K 250K

FSC-A

iRFP- mScarlet+

iRFP- mTq2+

iRFP- mVenus+

e Colony 1 Colony 2 Colony 3 Colony 4

f % of this population

control + Cre

0 20 40 60 80 100

0 50 100

% of this population

control + Cre

0 20 40 60 80 100

0 50 100

Blastocyst Mosaic founders

Transfer

Pseudo-pregnant

Mosaic founders

Breed and check for germ line transmission
A novel multicolour reporter allele for in vivo lineage tracing

Figure 5: Generation and testing of the \textit{Rosa26}\textsuperscript{PRIME} allele. a) Strategy for the generation of \textit{Rosa}\textsuperscript{PRIME} transgenic mice. Linearized vector is introduced in mouse embryonic stem (ES) cells. Selection cassettes for Neomycin Resistance (NeoR) and Diphtheria Toxin A (DTA) allow selection for ES cells with a correctly integrated \textit{PRIME} cassette. Correctly targeted ES cell clones are introduced into blastocysts. These mosaic blastocysts are transferred into pseudo-pregnant females, where they develop into mosaic pups. Mosaic founders are used for breeding, offspring should be checked for transmission of the transgene. b) Confocal images of different live mESC colonies targeted with a single copy of the \textit{PRIME} construct. Some mESCs were successfully transfected with Cre and switched to red, cyan or yellow. c) Cartoon showing strategy to determine if the different recombinations occur in equal ratios. \textit{Rosa26} knock-in ES cells containing the \textit{PRIME} construct (\textit{ES}\textsuperscript{PRIME}) were transfected with a Cre expression plasmid and analysed by FACS at different timepoints (ranging from 72 hours to 16 days after transfection). d) FACS plots showing a reduction in IRFP670 signal in \textit{ES}\textsuperscript{PRIME} cells transfected with Cre and analysed 16 days after transfection. (left: untransfected control, right transfected with Cre). e) Graph showing the percentage of IRFP-negative (IRFP-, grey) and IRFP-positive (IRFP+, dark red) cells in the control and + Cre sample. Individual datapoints for two independent experiments are shown. See methods for details. Horizontal bars depict the mean. f) Graph showing the percentage of IRFP negative cells that have recombined the \textit{Rosa26}\textsuperscript{PRIME} allele to become positive for red (mScarlet-i+), cyan (mTq2+) or yellow (mVenus+) fluorescence. Individual datapoints for two independent experiments are shown. See methods for details. Horizontal bars depict the mean.

Discussion

Here we describe the design, cloning and functional validation of a novel reporter allele suitable for multicolour, fluorescent lineage tracing. \textit{Rosa26}\textsuperscript{PRIME} was designed to hold multiple improvements over existing reporters, including the fact that all cells are labelled before and after switching and the fact that we used membrane-localized bright fluorescent proteins for outlining the shape and size of labelled cell clones for quantitative lineage tracing purposes.

Designing a fluorescent multicolour reporter allele

We set out to create a multicolour reporter expressing fluorescent proteins that are easy to image using confocal microscope or fluorescent sorting setups available in most labs with minimal spectral overlap. Based on this, we selected four different fluorescent proteins: a cyan, yellow, red and near-infrared FP (Figure 1, 4b). To allow for easy detection of tissue structure, all cells of our fluorescent reporter should express a nuclear signal. We chose the near-infrared colour because they are still relatively dim compared to the other proteins (Rodriguez et al., 2017), and therefore
less well suited for outlining recombined cells after Cre-mediated recombination. However, concentrating the signal in the nucleus should render the signal sufficiently concentrated to mark all cells prior to switching, although this remains to be determined empirically in vivo.

The remaining three FPs carry a membrane localization sequence (Figure 1c). The different colours allow us to distinguish between different, independent recombination events, whereas membrane localization of the fluorescent signal allows for easy recognition of cell architecture – and clone shape, size and cellular composition – in recombined tissues. Both of these features are compatible with quantitative lineage tracing, which allows in depth analysis of stem cell division characteristics (Wuidart et al., 2016).

**Testing fluorescent proteins for the PRIME allele**

We tested the performance of different FPs, in anticipation of the demands that would be placed on the reporter in wholemount confocal imaging experiments for in vivo lineage tracing purposes. We ultimately selected the bright monomeric mTq2 and mVen, based on the fact that it bleached slightly less than SYFP2 (Figure 2b).

The analysis of YFP performance in fixed cells was done on a confocal system, since this is the microscopic setup that will be used to image tissues. In hindsight, a widefield analysis might have been better suited to quantitatively measure the total fluorescent signal per cell, thereby removing any potential focal plane differences, although that would mean we cannot test it on the setup that will be used to image mouse tissues.

To tackle bleaching, optimization should be dedicated to optimizing imaging conditions. This includes finding the right balance between laser power, scanning speed and averaging specific for the combination of the imaged tissue and FPs to be visualized. In addition, techniques such as spinning disk confocal microscopy or light-sheet microscopy can severely reduce bleaching. How the PRIME allele performs in vivo in these systems remains to be empirically determined.

Slightly riskier were the inclusion of mScarlet-I (newly developed by colleagues in our institute) and iRFP670, since it was the first time these new proteins would be used in a single copy fashion for confocal imaging of mouse tissues. IRFP670 has been successfully imaged in vivo, but only via whole-body In Vivo Imaging System (IVIS) imaging of injected cells or tumours for IRFP670 (Shcherbakova and Verkhusha, 2013; Shimizu et al., 2020). Often the brightest cells are picked for injection, containing several copies of the gene encoding the FP. However, IRFP720, closely related to IRFP670, recently has been inserted as knock-in genetic reporter, but again this reporter was only used for IVIS imaging and not for confocal imaging (Fukuda et al., 2019). Unfortunately, IVIS does not provide the resolution needed for studying individual cells in a tissue (Scarpelli et al., 2020). mScarlet (mSc), the close family member of mSc-I, was also used for whole-body fluorescent imaging in
A novel multicolour reporter allele for *in vivo* lineage tracing

xenograft tumour models (Min et al., 2019). A more recent publication presents NLS-mScarlet-I in a TGF-β signalling reporter, that can be imaged with a confocal microscope in skin tumour (squamous cell carcinoma) sections (Taniguchi et al., 2020). Because viral transduction of the gene generally leads to integration in a highly active gene region, and can lead to multiple integrations per cell, it cannot be directly compared to the expression level of a targeted, single-copy integration as designed for our reporter.

iRFP670 stems from bacterial phytochromes. To become fluorescent, phytochromes need biliverdin, an intermediate of heme metabolism that is ubiquitously expressed in mammalian tissues. Importantly, not all far-red FPs need the same amount of biliverdin. For instance, smURFP, initially a promising candidate for our PRIME allele, needs a high amount of biliverdin to become fluorescent (Rodriguez et al., 2016). Therefore, it’s brightness in HeLa cells appears 53-fold lower than miRFP670 (a monomeric variant of iRFP670) when no biliverdin is added (Shemetov et al., 2017). Adding biliverdin increased fluorescence of smURFP up to 24 times. Even though biliverdin should be present in most mouse tissues, the variances and effects thereof are not known and should thus be thoroughly tested.

We also tested another important feature: how our selected FPs would perform upon fixation. Many FPs are quenched by fixation with PFA, but most tissue imaging protocols cannot do without. We knew from previous experience with the *Rosa26mTmG* mouse that both the red and green fluorescent signals are still very bright after fixation and easy to image in wholemount tissues such as the mammary gland or intestine after PFA fixation and glycerol or methylsalicylate clearing. To directly compare TdTom with mSci-I in HeLa cells, they were co-expressed from a vector with mTq2, to allow normalization over the cyan channel (Figure 3). While the tdTomato signal remained more stable upon longer (up to 60 min) fixation usually needed for wholemount tissues, mSc-I started so bright that even after quenching of the signal by fixation it should remain bright enough to easily image *in vivo*. This, together with the smaller sequence of mSc-I and the opportunity to test its performance as a genetic reporter *in vivo* for the first time, further substantiated our choice for mSc-I as our RFP.

We checked the performance of our iRFP candidate iRFP670 upon fixation by testing how the signal would quench upon fixation (Figure 3g,h). We noticed that at lower fixation times, some of the nuclear signal seemed to leak out into the cytoplasm, possibly influencing the total intensity. Overall, since visualization of the nuclear signal above background was still possible after 60 minutes fixation, we decided to continue with iRFP670 for PRIME.

Finally, the four FPs used for our PRIME allele have different evolutionary origins – mTq2 and mVen are derivatives of the original AvGFP identified in jellyfish (Prasher et al., 1992), while iRFP670 comes from a bacterial phytochromes
(Shcherbakova and Verkhusha, 2013), and mSc-I is designed synthetically (Bindels et al., 2016). This should make it possible to use unique antibodies against the different FPs should this be needed to improve detection. Unfortunately, this is not possible at present for mTq2 and mVen, since they only differ by 13 bp (Supplemental figure 2).

**Analysis of Cre/Lox recombination of the PRIME allele**

Unlike the Rosa26Confetti allele, which only uses loxP sites, we opted to include variant lox sites. Several Lox sites are available that differ in the spacer sequence and only recombine with their identical counterpart (Richier and Salecker, 2015). The set of Lox sites chosen for our Rosa26PRIME construct is the same one that has been shown to work efficiently in vivo for the Brainbow1.1 construct (Livet et al., 2007). These are LoxP, Lox2272 and LoxN. The order of Lox sites was based on the Brainbow3.2 construct (Cai et al., 2013).

We checked recombination of the PRIME construct in vitro while cloning (Figure 4c), and following targeting we also checked performance in of the Rosa26PRIME allele in ES cells (Figure 5b) while awaiting the generation of mosaic founders, germline transmission and establishment of a Rosa26PRIME colony and lineage tracing cohorts.

The final PRIME allele contains three sets of Lox sites, which are spaced apart at various lengths. By design, the colours should appear in a 1:1:1 ratio following recombination. However, we found that – at least in ES cells - the prevalence of the different fluorescent colours decreased with the distance between Lox sites (Figure 5f). Other multicolour reporters also show this phenomenon: while Rosa26Confetti is specifically designed for stochastic expression, it shows variety between the different fluorescent proteins, with the nuclear GFP frequency always being lower expressed than the expected 25% (Snippert et al., 2010a).

The preference in recombination as measured could be due to the distance between Lox sites, as well as due to differences in recombination efficiency between the different types of Lox sites present in the Rosa26PRIME construct. Whether the ratio found in the ES cells reflects the situation in vivo remains to be tested. Recombination patterns can differ between cell types and Cre line used, as illustrated for Rosa26Confetti in combination with two different driver lines in cartilage (Li et al., 2017), with the driver line Prg4CreERT2 showing all four Confetti colours after Cre/Lox recombination and the other, Col2CreERT2, resulting in only three Confetti colours and no GFP-positive cells. One difference as suggested by the authors is the difference in Cre expression levels between the two driver lines. This strongly suggests that for all multicolour reporters that are to be used for quantitative lineage tracing analyses, the recombination ratios need to be determined for each specific experimental setup (Blanpain and Simons, 2013; Lloyd-Lewis et al., 2018; Wuidart et al., 2016).
A novel multicolour reporter allele for in vivo lineage tracing

Figure 6: Recombination outcomes in homozygous Rosa26PRIME mice. Cartoon illustrating cells carrying two Rosa26PRIME alleles, that show no Cre/Lox recombination (1), show one recombined allele and one non-recombined (2, 3, 4), have recombined alleles with the same fluorescent protein (5, 6, 7) or two differently recombined alleles (8, 9, 10).

Outlook for Rosa26PRIME reporter
Our newly generated Rosa26PRIME multicolour reporter mouse should be useful for wholemount imaging of complex tissues. In ES cells it performs as designed: the construct switches to one out of three different FPs when Cre is activated, and the three membrane-localized FPs nicely outline cell architecture in tissues. Moreover, the bright labelling allows for easy FACS analysis of labelled cells. Combined with an inducible Cre driver of choice, this reporter should allow unbiased analysis of stem cell fate in vivo at single cell resolution, analysing both stem cell progeny and tissue architecture in wholemount tissues by both confocal microscopy and FACS. The fluorescent proteins chosen should further allow imaging of cells and tissues without the need for labelling with antibodies.

If more colours are desired, it is possible to generate mice homozygous for the PRIME allele. Because homozygous mice have two copies of the transgene, this leads to ten different possible outcomes as illustrated in Figure 6. The only drawback for using homozygous mice, is that this would require additional breeding.
**Material and Methods**

**Generation of the PRIME construct**

All constructs were designed *in silico* using Snapgene Viewer and the online tool Benchling (benchling.com). The PRIME construct was assembled using standard cloning methods. The mTmG construct vector (Muzumdar et al., 2007) (pCA-mTmG; Addgene plasmid #26123) was used as building vector. The mTmG cassette was cut out by BamHI digestion, leaving the CAG promoter and one LoxP site. LoxN and Lox2272 were ordered as single-stranded oligos with unique flanking restriction to allow annealing and ligating into the vector. First, an oligo carrying BglII-Lox2272-BamHI was inserted at a unique BamHI restriction site. Second, the backbone was opened via BamHI/SwaI restriction and an oligo carrying BamHI-LoxN-SwaI was inserted. This oligo also carried a multiple cloning site for the following cloning steps.

The fluorescent proteins were inserted one by one, together with their correct nuclear or membrane marker sequence and flanking Lox sequence. The sequences were ordered as IDT gBlock with flanking restriction sites unique to both the vector and insert for easy ligation. Briefly, the pCAG construct was digested with BamHI/AflII and Lox2272-M-mTq2-polyA was inserted as BamHI/AflII fragment to generate pCAG-mTq2. Next, pCAG-mTq2 was digested with EcoRI/AflII and LoxN-M-mVen-polyA was inserted as EcoRI/AflII fragment to generate pCAG-mTq2-mVen. Then, pCAG-mTq2-mVen was digested with NheI/AgeI and M-mSc-I-polyA was inserted as NheI/AgeI fragment to create pCAG-mSc-I-mTq2-mVen. Lastly, pCAG-mSc-I-mTq2-mVen was digested with BamHI/XbaI and 3xNLS-iRFP670-polyA-LoxP was inserted as BamHI/XbaI fragment to create pCAG-iRFP670-mSc-I-mTq2-mVen. In between cloning steps, the construct was checked via Sanger Sequencing and tested for functional Lox recombination by co-transfecting the vector and a Cre plasmid into HeLa cells. We co-transfected with an empty pBluescript vector to get only few copies of the PRIME vector in, but almost all cells showed two, three, or even all four fluorescent colours simultaneously, meaning we introduced several copies per cell.

To complete the PRIME targeting construct, pCAG-iRFP670-mSc-I-mTq2-mVen was digested with AflII/SwaI to insert an FRT-PGK-Neo-FRT cassette to create pCAG-iRFP670-mSc-I-mTq2-mVen-Neo.

Cloned constructs were transformed into chemically competent DH5α cells and grown at 37°C. Constructs >10kb were transformed and grown in Stbl3 Chemically competent cells (Fisher) and grown at 30°C.

The PRIME construct was digested with PacI/SwaI and ligated into a linearized pROSA26-1 (Addgene plasmid #21714) digested with XbaI. Single stranded overhangs after digestion were filled with T4 polymerase before ligation. pROSA26-
1 is the original *Rosa26* targeting vector from the Soriano lab containing 1 and 4 kb homology arms (Soriano, 1999). The resulting construct is the *Rosa26*PRIME targeting construct. An extensive set of restriction digests was performed to test for a correct targeting construct, as well as sequencing of the entire vector with multiple primers (Supplementary table 1).

**Generation of *Rosa26*PRIME mice**

Stem cell targeting was performed at the NKI transgenic and knockout core facility. The targeting vector was linearized with KpnI and electroporated into B6_colA_RMCE mESCs. Briefly, individual colonies were grown on selective medium with neomycin and DT. Surviving colonies were picked and checked for correct integration of the construct by long range PCR and by Southern Blot for *Rosa26* 5’ and qPCR copy number analysis for the *Rosa26* 3’ side (Neo cassette). Correctly targeted ES*Rosa26*-PRIME/+ clones were expanded and used for blastocyst injections into Go germline blastocysts. These were scored by palpation and crossed to generate heterozygous offspring. Mice were crossed to a FLP deleter strain to remove the FRT-Neo-FRT cassette to establish heterozygous offspring in the F1 or F2 generation.

**Cell culture**

BC44 cells (a gift from Marie-Ange Deugnier, (Deugnier et al., 1999)) were cultured in RPMI1640 Medium supplemented with 10% FBS and 5 μg/mL insulin (Sigma) at 37°C and 5% CO₂. HeLa cells were cultured in DMEM medium supplemented with Glutamax and 10% FBS at 37°C and 5% CO₂. HeLa cells were cultured at 37°C and 5% CO₂ in DMEM medium supplemented with Glutamax and 10% FBS. Both cell lines were transfected with PEI (Polyethylenimine, Polysciences, Omnilabo) dissolved in milliQ with a DNA:PEI ratio of 1:6. Medium was refreshed 4 to 6 hours after transfection.

Mouse embryonic stem cells (mESCs) were cultured in N2B27 medium with 2i + LIF at 37°C and 5% CO₂. N2B27 medium: one volume DMEM/F12 combined with one volume Neurobasal medium and 10 mM Glutamax (Invitrogen), supplemented with 0.5% N2 Supplement, 1% B27 Supplement, 0.033% BSA 7.5% solution, 50 μMβ-mercaptoethanol, 2 mM Glutamax, 100 U/ml penicillin and 100 μg/ml streptomycin (all Thermofisher). Medium was supplemented with fresh 1000 U/ml LIF (Merck) and 2i: CHIR99021 (ITK Diagnostics) 3 μM and PD325901 (Sigma) 1μM. Plates were coated with 0.1% bovine gelatin. Cells were transfected with a Cre expression plasmid (Addgene #13775) using either PEI or X-tremeGENE HP DNA (Roche) in a 1:1 ratio (1 μg DNA per 1 μl X-tremeGENE). A total of 2 independent transfection experiments were performed (experiment 1: 1 control, 4 Cre-transfections, analysed at 72 hours or 10 days after transfection; experiment 2: 1 control, 2 transfection conditions, analysed at 11 or 16 days after transfection).
Microscopy
For live cell imaging, BC44 cells were cultured on 25mm glass coverslips in 6-well plates. Right before imaging, medium was replaced with pre-warmed Microscopy Medium (137 mM NaCl, 5.4mM KCl, 1.8 mM CaCl₂, 0.8mM MgCl₂, 20mM glucose, and 20mM HEPES at PH = 7.4).

Fixing cells was done using freshly prepared 4% PFA made from 16% PFA (methanol free ampules, Thermo Scientific) diluted in PBS for 5-60 minutes at room temperature.

Images were made on a Nikon A1 confocal microscope, using a 63x oil immersion objective. The pinhole was set to 1 Airy unit. mTq2 was imaged using a 457nm laser, 457/514/561 nm dichroic mirror and bandpass (BP) 482/35 nm emission filter. SYFP2/mVen was imaged with a 514 nm laser, combined with 457/514/561 nm dichroic and BP540/30 nm emission filter. For imaging mScarlet-I, a 561nm laser was used with 405/488/561/640 nm dichroic and a 595/50 nm bandpass emission filter. For the far-red proteins, a laser of 640 nm was used, 405/488/561/640 nm dichroic and a BP700/75 nm emission filter. The photo stability experiments were performed by imaging with continuous laser exposure for 10 min with a laser power comparable with normal imaging conditions (10-60%). Images were captured every 10 seconds. For testing fixation procedures, cells were fixed in freshly made 4% paraformaldehyde for the appointed time points and washed generously with PBS. Cells were imaged the same day with the same settings for fluorescence excitation/emission.

Imaging of mESCs was performed on a Leica SP8 confocal microscope with LasX software, using a 63x oil objective. Excitation lasers used for mTurquoise2 442nm, mVenus 514nm, mScarlet-I 561nm and iRFP670 633nm. Detection was done with the Leica PMT and HyD detectors at 450-500nm for mTq2; 520-550 for mVen; 575-620nm for mSc-I and 650-720nm for iRFP670.

Image analysis
Confocal images and time stacks were analysed using ImageJ (FIJI). A region of interest (ROI) was made for each individual cell using the magic wand selection tool and manual setting of tolerance. For each ROI, intensity was determined using the ROI Manager “Measure” function. Background extraction was performed by determining the background at a drawn ROI on a spot without cells. Intensity values were exported to Excel for further (statistical) analysis. Graphs were made in Graphpad Prism (v6.01).

FACS analysis
To prepare cells for FACS sorting, cells were washed with pre-cooled HBSS containing 2% FBS twice before resuspension in ice-cold HBSS with 2% FBS. Cells were kept on ice until analysing and filtered using a cell strainer. Cells were analysed.
A novel multicolour reporter allele for in vivo lineage tracing

with a BD FACS AriaIII using the endogenous fluorescent signal. Analysis was performed using FlowJo. mTq2 was measured a 407nm laser and emission was filtered using a bandpass (BP) 450/50nm filter. mVenus was measured using a 488nm laser and BP530/30 nm filter. mScarlet-I was measured using a 561nm laser and BP582/15nm filter. IRFP was measured using a 633nm laser and BP660/20nm filter.

Supplementary figures

Supplementary table 1. Sequencing primers used:

<table>
<thead>
<tr>
<th>#</th>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>RVA534</td>
<td>T3 Fw</td>
<td>ATTAACCCCTCCTAAGGGGA</td>
</tr>
<tr>
<td>RVA535</td>
<td>T7 Fw</td>
<td>TAAATAGACCTAATAGGGG</td>
</tr>
<tr>
<td>RVA274</td>
<td>R26 Fw</td>
<td>AAGTCGCTCTGTTATGTTAT</td>
</tr>
<tr>
<td>RVA286</td>
<td>R26 Rv</td>
<td>TCAATGGGGGAGGTCGTTT</td>
</tr>
<tr>
<td>RVA1556</td>
<td>pCAG Fw</td>
<td>GCAACGTCTGGTATGTTG</td>
</tr>
<tr>
<td>RVA1557</td>
<td>pCAG Rv</td>
<td>GGTCCCTCAGAGGTTTCA</td>
</tr>
<tr>
<td>RVA1718</td>
<td>mSc Fw</td>
<td>GAGCGGCGGTGACTATCCGAG</td>
</tr>
<tr>
<td>RVA1719</td>
<td>mSc Rv</td>
<td>CCAGCCTGTGCTTGTTTTCGC</td>
</tr>
<tr>
<td>RVA1620</td>
<td>mTq2 Fw</td>
<td>GAGATCTTTATACGAGGTTATAAGCGGT</td>
</tr>
<tr>
<td>RVA1622</td>
<td>mVrn Fw</td>
<td>ATACGAAGGTATGAAATCCCACCA</td>
</tr>
<tr>
<td>RVA1741</td>
<td>Neo Rv</td>
<td>GCCCAGTCATGAAGCCGGG</td>
</tr>
<tr>
<td>RVA1144</td>
<td>Neo Fw</td>
<td>AATGGCGTCTACGTCTGCTGAGGAGG</td>
</tr>
<tr>
<td>forSeqPrime2</td>
<td></td>
<td>GGGCCCTGTGCTGCTGATTGG</td>
</tr>
<tr>
<td>revSeqPrime3</td>
<td></td>
<td>CACCTGTGGGAGTAGTACGCAG</td>
</tr>
<tr>
<td>forSeqPrime4</td>
<td></td>
<td>GCTGGGTGAAGCCCTGGAG</td>
</tr>
<tr>
<td>revSeqPrime5</td>
<td></td>
<td>CCTATACGAGTTGTTAGCATCCGG</td>
</tr>
<tr>
<td>forSeqPrime6</td>
<td></td>
<td>TCACTGCATCGCCATGACGG</td>
</tr>
<tr>
<td>forSeqPrime7</td>
<td></td>
<td>CGCGGCGGAATGTTCGCGG</td>
</tr>
<tr>
<td>revSeqPrime8</td>
<td></td>
<td>GTACAACCGCTCGGTGGACG</td>
</tr>
<tr>
<td>revSeqPrime9</td>
<td></td>
<td>GTGTACTGCGCAAGTGCCGAG</td>
</tr>
<tr>
<td>forSeqPrime10</td>
<td></td>
<td>CGAGGACTACCCGGTGGTG</td>
</tr>
<tr>
<td>forSeqPrime11</td>
<td></td>
<td>CCCGTGCTGGGGCGTGACG</td>
</tr>
<tr>
<td>revSeqPrime12</td>
<td></td>
<td>TTGCGGTTGGTGCGAGTACG</td>
</tr>
</tbody>
</table>

qPCR

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neo left</td>
<td>CTGTGCTCGAGTGTTCAC</td>
</tr>
<tr>
<td>Neo right</td>
<td>GGCAGGAGCAAGGTTGAGAT</td>
</tr>
</tbody>
</table>
Supplementary figure 1. Vector map of RosaPRIME targeting construct, created by Snapgene Software (available at Snapgene.com).
A novel multicolour reporter allele for \textit{in vivo} lineage tracing

\[
\text{mTurquoise2} \\
\text{rFP670} \\
\text{mScarlet-I} \\
\text{mVenus}
\]

\[
\text{mTurquoise2} \\
\text{rFP670} \\
\text{mScarlet-I} \\
\text{mVenus}
\]

\[
\text{mTurquoise2} \\
\text{rFP670} \\
\text{mScarlet-I} \\
\text{mVenus}
\]

\[
\text{mTurquoise2} \\
\text{rFP670} \\
\text{mScarlet-I} \\
\text{mVenus}
\]

\[
\text{mTurquoise2} \\
\text{rFP670} \\
\text{mScarlet-I} \\
\text{mVenus}
\]

\[
\text{mTurquoise2} \\
\text{rFP670} \\
\text{mScarlet-I} \\
\text{mVenus}
\]

\[
\text{mTurquoise2} \\
\text{rFP670} \\
\text{mScarlet-I} \\
\text{mVenus}
\]

\[
\text{mTurquoise2} \\
\text{rFP670} \\
\text{mScarlet-I} \\
\text{mVenus}
\]

\[
\text{mTurquoise2} \\
\text{rFP670} \\
\text{mScarlet-I} \\
\text{mVenus}
\]

\[
\text{mTurquoise2} \\
\text{rFP670} \\
\text{mScarlet-I} \\
\text{mVenus}
\]

\[
\text{mTurquoise2} \\
\text{rFP670} \\
\text{mScarlet-I} \\
\text{mVenus}
\]

\[
\text{mTurquoise2} \\
\text{rFP670} \\
\text{mScarlet-I} \\
\text{mVenus}
\]

\[
\text{mTurquoise2} \\
\text{rFP670} \\
\text{mScarlet-I} \\
\text{mVenus}
\]

\[
\text{mTurquoise2} \\
\text{rFP670} \\
\text{mScarlet-I} \\
\text{mVenus}
\]
Supplemental figure 2. Sequence alignment of PRIME fluorescent proteins. The sequence of all four fluorescent proteins in PRIME were aligned to each other, with mTurquoise2 as a reference sequence. Differences are depicted in as coloured bases. While mVenus shows only 13 bases difference with mTurquoise2, mScarlet-I and especially iRFP670 are genetically more distinct.

Supplementary sequences: full plasmid sequences of the targeting vector and Rosa26 probe are available via osf.io/43yj2.

Author contributions

RvA conceived and supervised the study. AvdM, BvdS, LK, IH and RvA conceived and designed the experiments. AvdM and BvdS cloned and tested the constructs. AvdM, BvdS performed the cell culture and imaging experiments. LK performed targeting in mESCs. AvdM and RvA analysed the data and wrote the manuscript.

Acknowledgements

We thank Joachim Goedhart for coming up with the PRIME acronym, as well as for offering his expertise with respect to FP selection, sharing YFP and mTq2 data and for input on the cloning process. We thank Katrin Wiese and Yorick van der Grift for cloning advice, Rahmen Ali, Tanya Braumuller and Fina van der Ahe for oocyte injections, the staff of the NKI and SILS animal facilities for mouse husbandry, the van Leeuwenhoek Centre for Advanced Microscopy (LCAM, Section Molecular Cytology, Swammerdam Institute for Life Sciences, University of Amsterdam) for the use of their facilities and LCAM staff for sharing their expertise and providing technical support, Amber Zeeman and Saskia de Man for help with the ES cell culture and FACS analysis and all colleagues for stimulating discussions. This research was supported by KWF Kankerbestrijding (career development award ANW 2013-6057 and project grant UVA 2015-8014, both to RvA).
A novel multicolour reporter allele for in vivo lineage tracing

References

A novel multicolour reporter allele for in vivo lineage tracing

Chapter 5