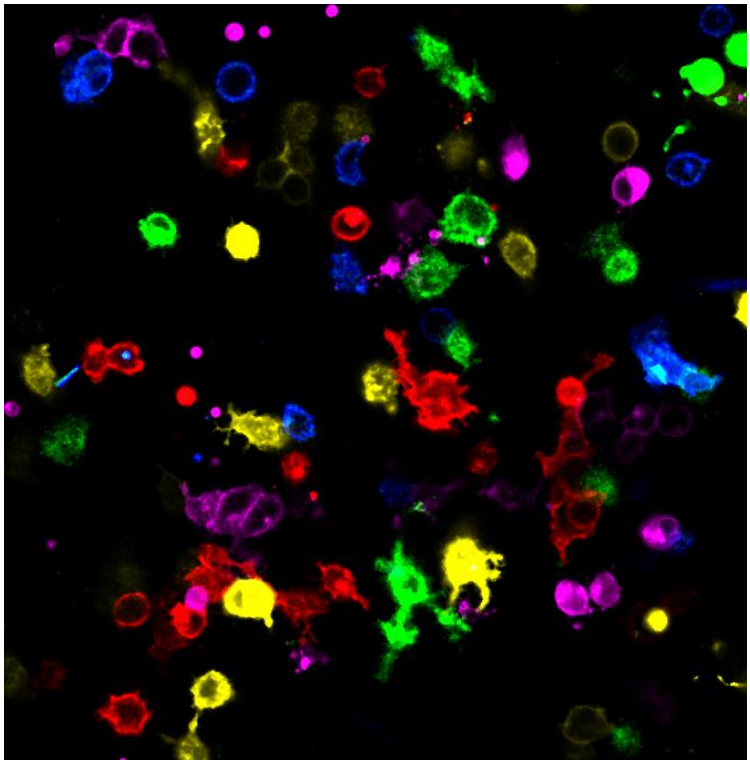
Development of new cell genetic tracing tools for the study of intratumour heterogeneity



Composite image of a mix of cells individually transfected with the different fluorescent proteins

Laura Quevedo[1]

Supervisor: Ignacio Varela[1]

[1]**Instituto de Biomedicina y Biotecnología de Cantabria (CSIC-UC), Departamento de Biología Molecular, Universidad de Cantabria, Santander, Spain.**

Máster en Biología Molecular y Biomedicina (2015-2016)

Index

1 Abstract

Intratumour heterogeneity has been observed in multiple cancers and has been postulated as a critical aspect for tumour metastasis and treatment resistance. Therefore, a further characterization of its role in cancer progression and metastasis has become essential to increase our understanding of cancer biology and to improve the treatment of cancer patients. The use of cell lineage tracing systems, combined with the use of genetically modified mouse models, could be applied in this context. Several genetic tracing systems, based in a random CRE-mediated recombination event in an allele with multiple fluorescent markers surrounded by incompatible lox sites have been created. Once this recombination occurs, the cell and its genetic descendants are permanently labelled with the same fluorescent marker. Nevertheless, these systems present several limitations like a reduced number of potential colour combinations or problems in the unique identification of the markers. Here we have identified new incompatible lox sites that, together with an efficient selection of fluorescent markers, have allowed us to design a new system that will be able to produce up to 15 different colour combinations that can be uniquely identified by confocal microscopy and FACS. This system will be combined with cancer mouse models to study the role and dynamics of intratumour heterogeneity in cancer progression.

2 Introduction

2.1 Cancer Genomics

Cancer is fundamentally a disease of the genome. It is caused by somatically acquired changes in the DNA sequence. These changes may be point mutations, copy number changes or rearrangements and may also be accompanied by epigenetic changes. These alterations might be "drivers" that confer clonal selective advantage and are therefore found in "cancer genes". Alternatively, these mutations may be "passengers" that have been carried along as a consequence of an intrinsic genome instability and offer no selective advantage to the cancer cell[1].

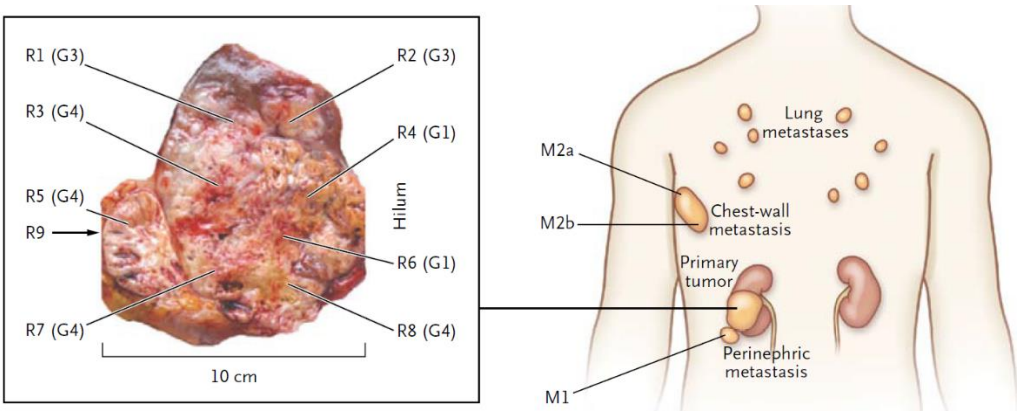
The identification of these genomic alterations is of critical importance to understand cancer biology, and the products of the newly identified cancer genes, or the molecular pathways they belong to, constitute potential targets for the design of new anti-tumour therapies. In the past years, the identification of such mutations has greatly improved the treatment of some tumours[2]. Despite all the progress made in the understanding of basic cancer biology, only a few tumour types can be successfully treated with "targeted therapies". Even when the first line of treatment is effective, most tumours acquire treatment resistance.

The limited success in treating cancer could be attributed in part to a lack of profound understanding of intratumour dynamics in cancer progression. A classical view of tumour evolution is that a main cell clone is progressively acquiring all capabilities or "hallmarks" necessary to produce an invasive tumour, so tumours arise as the result of a simple and lineal accumulation of genetic aberrations. Accordingly, metastasis is seen as the final and often fatal step. Nevertheless, this vision of tumour progression has been recently challenged with the identification of different genetic cell clones inside primary tumours Additionally, genetic analyses of circulating tumour cells and functional studies in animal models, have suggested that the dissemination of tumour cells, as well as the seeding of metastases, can be a very early event, even at a pre-malignant stage[3]. Whether this implies that the colonization of the distant tissue happens at such an early stage is still unclear. Only advanced tumours might be able to release cells capable of producing metastases, and indeed macroscopic metastases are not usually observed before the primary tumour is at an advanced stage. This last observation does, however, not necessarily reinforce the traditional view of tumour progression, as it seems that growth control strategies normally used to treat primary tumours can actually promote metastasis. In accordance with an early metastatic implantation model, it has been described evidence in cancer of early parallel evolution of primary and metastatic tumours as well as tissue specific evolutionary branches among different metastases from the same tumour[4,5]. One potential model of this metastatic tissue-specificity has been postulated many years ago as the result of specific interactions between cancer cells and the recipient tissue cells[6].

2.2 Intratumoural Heterogeneity

Taking advantage of the high-sensitivity of next-generation sequencing (NGS) technologies, it has been shown that tumours appear as complex biological structures with a hierarchical organization of different cell lineages in constant evolution, an observation that challenges the classical lineal model of tumour evolution (Figure 1)[5]. Similarly, this new view of tumour progression is difficult to reconcile with the presence of a small population of cancer stem cells, identified in some tumour types and postulated to support the growth of the tumour bulk[7]. According to this, it is plausible that one or other model is more prominent in some tumour types, whereas a combination of both is needed to explain the behaviour of others.

A)



B)

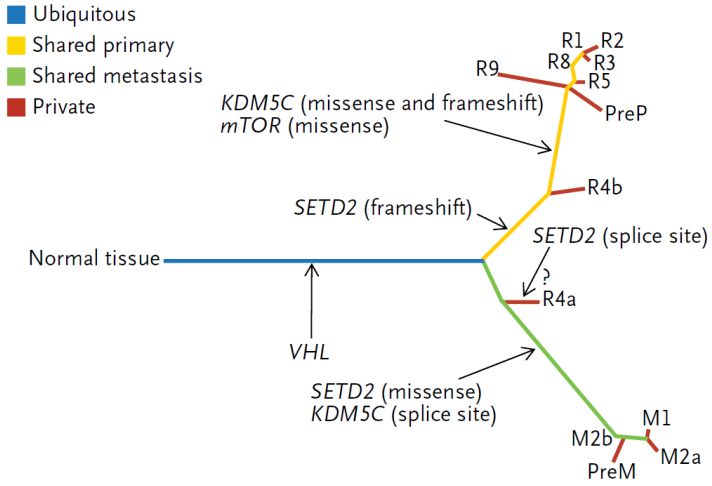


Figure 1. Genetic Intratumoural Heterogeneity and Phylogeny. A) Swanton et al., evaluated multiple tumour-biopsy samples in the same patient with metastatic renal-cell carcinoma to study intratumoural heterogeneity. Sites of core biopsies and regions harvested from nephrectomy and metastasectomy specimens are shown. They analysed nine primary tumour regions (R1 to R9), a metastasis in the perinephric fat of the nephrectomy specimen (M1), two regions of the excised chest-wall metastasis (M2a and M2b), and germline DNA. G indicates tumour grade. **B)** Phylogenetic relationships of the tumour regions. Branch lengths are proportional to the number of nonsynonymous mutations separating the branching points. Potential driver mutations were acquired by the indicated genes in the branch (arrows). Modified from Swanton et. al, 2012.

As intratumour heterogeneity has been described in several tumour types both at the transcriptional and the genetic level, it is likely a general problem not restricted to specific tumour types[8]. How this heterogeneity originates is not well understood. It could be either an ongoing process consequence of a general genetic instability or a sporadic event happening in a specific moment during tumour progression. This second scenario can derive from the occurrence of a catastrophic event that should produce a massive genetic variation in a very short period of time. "Chromothripsis" or "kataegis" are examples of this kind of events[9].

The presence of intratumour heterogeneity has important clinical implications. First of all, in terms of diagnosis, single biopsies seem insufficient to capture the whole molecular heterogeneity in the primary tumours, which points into question our ability to correctly classify the molecular features of a tumour from a single simple[5]. Interestingly, Shan et al. have described the presence, inside the same primary ovarian tumour, of clones that could be categorized in different molecular groups[10]. Secondly, it has been postulated that tumour heterogeneity plays a major role in treatment resistance and metastasis, the two major causes of cancer-associated death. Thus, as a correct molecular characterization determines nowadays a preferred treatment, it is plausible that targeted therapies only affect to some cell clones inside the tumours when others are insensitive to them. This could offer an explanation to the frequent cancer relapses seen after targeted therapies. Additionally, if metastases are seeded by some of these minority clones, limited success of the same therapy in metastatic growths can be expected. Consequently, the grade of intratumour heterogeneity has been associated with worse prognosis in breast cancer[11]. Finally, specific properties developed by some cell clones, such as the ability to promote angiogenesis or tissue inflammation, could benefit the rest of clones generating a more favourable environment. In that case, it is plausible to hypothesize the existence of collaboration between the different cell clones. In contrast, the presence of highly proliferative but non-invasive clones could impede the proper growth of other more metastatic or dangerous clones. This possibility has been postulated by some authors as a new opportunity for cancer therapy[12].

In summary, intratumour heterogeneity has been described in multiple tumour types and has been postulated as a critical driver of metastasis and treatment resistance. Therefore, a further characterization of its role in cancer progression and metastasis has turned essential to increase our understanding of cancer biology and to improve the treatment of cancer patients.

2.3 Animal models and cancer

In cancer research, genetically engineered mouse models have proven to be very useful in validating gene function, identifying novel cancer genes and tumour biomarkers. These animal models have also provided fundamental insight into the molecular mechanisms of cancer and constitute excellent pre-clinical tools for the validation of novel therapeutic strategies. Despite their limitations, such as the species-specific differences to humans, there are multiple examples of rapid

translation of the knowledge acquired in mice to the treatment of patients. Additionally, new advances in mouse genetic engineering resulted in the production of more "humanized" mice that are more clinically relevant and recapitulate better the molecular, cellular and genomic events of human cancers[13].

The advances of the past decades in stem cell technology and genome engineering allow us to introduce almost any genetic alteration of interest into the mouse genome, from single base changes to large genomic rearrangements. Genetically engineered alleles can be placed under physiological or altered regulation, and their expression can be controlled in a temporal and tissue specific fashion. They can be introduced into the mouse germline or delivered to mice by different types of viruses. The combination of these genetic engineering techniques with *in vivo* imaging technologies allows researchers to monitor the location of cells expressing specific reporters. It is also possible to generate a random expression of genes and markers in only a proportion of mouse cells by developing mosaic mice[13].

2.4 Genetic tracing systems

Lineage specific genetic labelling constitutes another additional advantage of mouse models. In 2007, Livet and collaborators created a system, termed Brainbow, based on the random and inheritable expression of different fluorescent markers[14] after the action of CRE Recombinase.

CRE-loxP recombination refers to the process of site-specific recombination between two loxP sequences mediated by the CRE recombinase protein. LoxP sites are 34-base-pair long recognition sequences consisting of two 13-bp long palindromic repeats separated by an 8-bp long asymmetric core spacer sequence. The asymmetry in the core sequence gives the loxP site directionality, and the canonical loxP sequence is ATAACTTCGTATA-GCATACAT-TATACGAAGTTAT. If the lox sites face in the same direction, the sequence between the loxP sites is excised as a circular piece of DNA. One potential limitation of using loxP sites is the inability to tightly control which loxP sites recombine if more than two are present. To account for this, alternate mutant versions of the loxP site have been created, which contain a unique asymmetric spacer "NNNTANNN", where "N" indicates which bases may vary from the canonical sequence (notice the conserved 4 and 5 spacer positions). These variant lox sites undergo recombination with other sites with identical spacer sequence, but might present limited compatibility with other sites with a different spacer sequence. Among these are loxN (GtATACcT) and lox2272 (GgATACtT). Lox511 (GtATACAT) was described as incompatible at first but it was reported subsequently to be compatible with loxP. Recently, two new incompatible lox sites, loxM3 (tggTAtta) and lox M7 (tggTAttT), have been reported as incompatible with loxP[15].

Taking advantage of the CRE-lox system, the Brainbow model is composed by a variable number of different fluorescent proteins, surrounded by loxP sites, that are located downstream of

a constitutive promoter. After the action of CRE recombinase, one of these proteins is located immediately after the promoter. The combination of loxP sites that are selected by the recombination event is mainly random and once it occurs, it is permanent and inherited by all descendant cells. Dr. Clevers and co-workers generated a knock-in mice, called confetti, using one of the original constructs from Sanes' group (Brainbow 2.1)[16], which allowed them to make very interesting studies of intestinal development. A similar strategy might be useful for the study of intratumour heterogeneity, nevertheless, this strategy has several limitations in this new context. For example, the original proteins selected for the construct present quite a lot of spectra overlapping which difficult the correct identification of the four proteins and the analysis of potential protein combinations. This circumstance limits the strategy to only four potential outcomes. Sanes and co-workers, being aware of this problem, have recently developed improved constructs (like the one called Brainbow 3.2) using proteins with almost no-overlapping spectra and incompatible lox recombination sites (Figure 2)[17]. The signals from these proteins are much more easily distinguished, although they are limited to the simultaneously use of only three fluorescent proteins.

Other possible solution to this problem consists on combining the spectrometric characteristics of proteins with their cellular location. Several authors have proved the feasibility of increasing the number of potential combinations produced in the cells using more different fluorescent proteins, or with the same proteins modified to be located in different cell compartments[18,19]. This strategy could be useful in microscopic experiments but it is not practical in order to separate the population by fluorescence activated cell sorting (FACS).

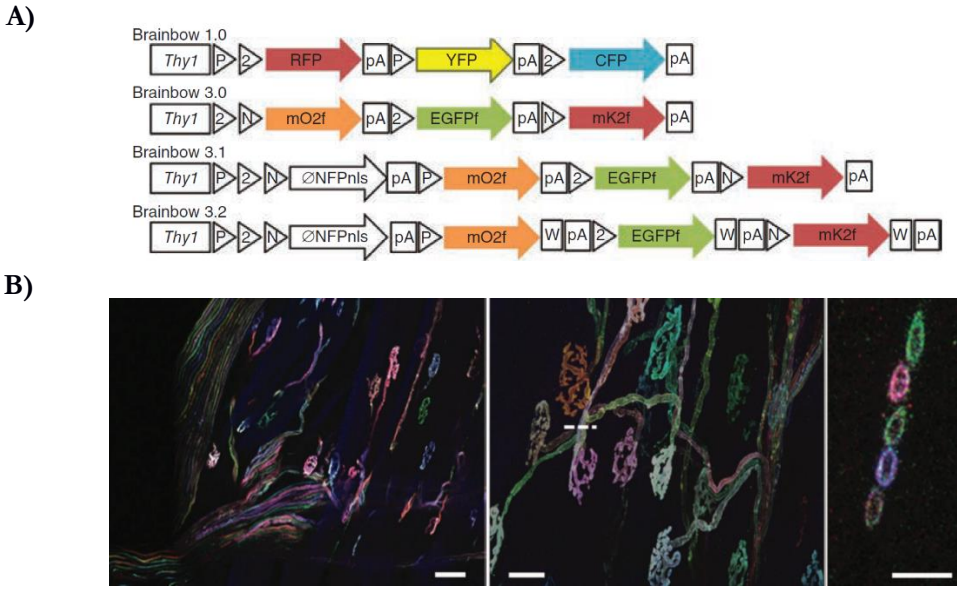


Figure 2. Brainbow constructions and examples. A) The Brainbow model is a Genetic tracing system based in a randomly CRE-mediated recombination event in an allele with multiple fluorescent markers surrounded by incompatible lox sites (P, 2, N). **B)** Terminal axons and neuromuscular junctions in extraocular muscle labelled using this system.

Finally, virus and transposon-mediated alternatives have been developed[17,19]. The system denominated RGB (Red-Green-Blue) is an example developed from lentivirus. RGB colouring is predicated on the physical principle that any spectral colour can be generated by mixing the three basic colours (red, green and blue) at different intensities[20]. Nevertheless, in these alternatives the number of alleles expressed in each cell is dependent on the distance from the viruses' injection site which complicates the analysis of the outcome. Additionally, the MAGIC (multiaddressable genome-integrative colour) system, is based on the use of transposons to induce the random integration of multiple copies of the Brainbow constructs in the mouse genome[18]. This system has the limitation that there is not control over the multiple places of integration of the genetic elements, which could have important effects on the mice.

2.5 Fluorescent proteins

Less than two decades after the introduction of green fluorescent protein (GFP) as a genetically encoded fluorescent marker, a wide spectrum of fluorescent proteins (FPs) is available (Figure 3). The original GFP is a small, stable and bright fluorescent protein that was isolated from *Aequorea victoria* jellyfish[21]. The GFP gene contains all the necessary transcriptional elements needed for the synthesis of the fluorophore, so GFP fluoresces on its own without requiring any coenzymes and substrates.

Jellyfish-derived GFP has been engineered to produce a vast number of useful green, blue, cyan and yellow mutants. T-Sapphire[22] is a green FP which excitation peak excited in the violet and YPet[23] a yellow FP based on YFP, both derive from the Enhanced Green Fluorescent Protein (EGFP), which is a brighter and more efficient folding variant of wild-type GFP. More recently, a blue FP called Sirius has been reported[24]. Sirius is a mutant of a cyan FP derived from GFP with the most blue-shifted emission spectrum to date.

Moreover, FPs from a variety of other species have also been identified, resulting in further expansion of the available colour palette, including blue, red and orange FPs. From *Entacmaea quadricolor,* Tag-BFP (blue)[25] and mKate2 (red)[26] fluorescent proteins have been isolated. Finally, mOrange2 is an orange FP that was isolated from *Discosoma* sp.[27].

The availability of a wide spectrum of FPs entails that the researchers are able to use the best possible combination of FPs that will result in minimal wavelength cross reactivity between their excitation and emission spectrum. This advantage can be used to uniquely distinguish different FPs in multiple-labelling experiments.

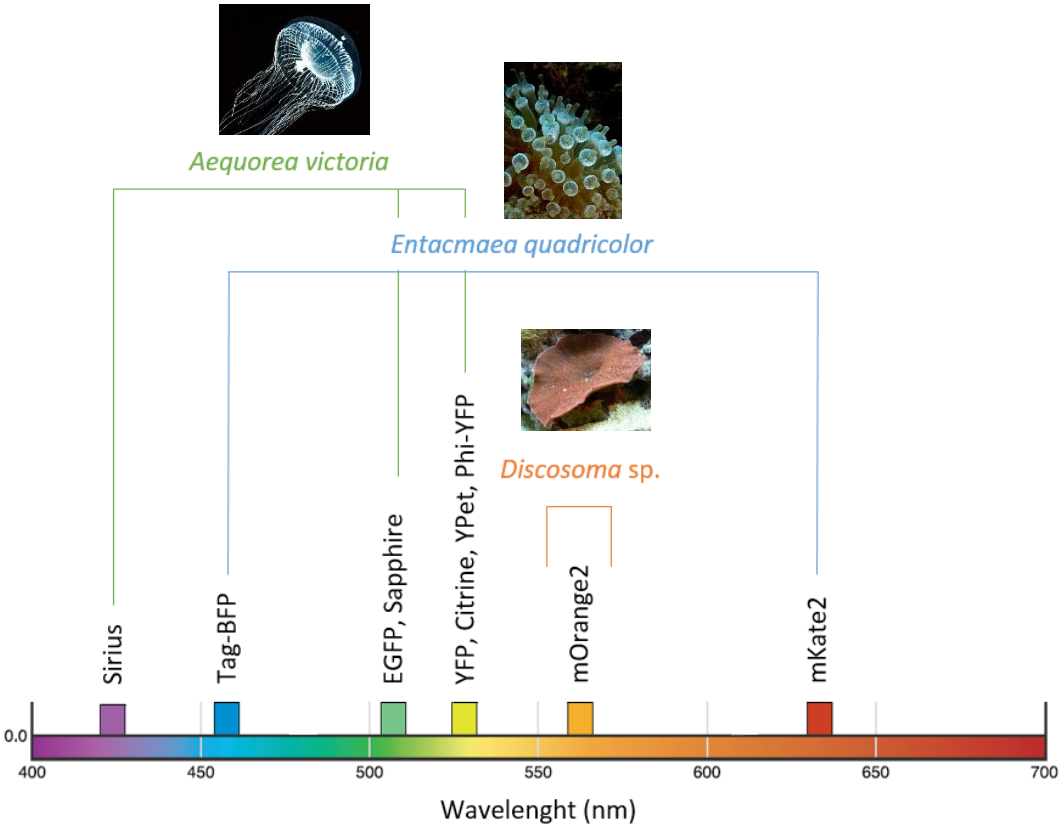


Figure 3. Spectral diversity of available fluorescent proteins and their origins.

3 Objectives

In the present project we propose to develop a new cell genetic tracing system based on the random expression of fluorescent proteins lead by the recombination between different incompatible lox sites that could improve the number of colour combinations generated by Brainbow 3.2 allele. We hypothesize that combining cancer mouse models with this cell genetic tracing tool could be very useful for the study of intratumoural heterogeneity. The concrete objectives are:

1. To select an optimal combination of fluorescent proteins with their excitation and emission spectra different enough to be able to uniquely distinguish them by confocal microscopy and to separate them by FACS.
2. To construct and validate a new system to identify new incompatible lox sites.
3. To identify new lox sites incompatible with the three lox sites (loxP, lox2272, loxN) already used in the Brainbow 3.2 allele.
4. To design a new cell genetic tracing allele based on the results obtained in the previous objectives to study intratumoural heterogeneity in cancer mouse models.

4 Methodology

4.1 Molecular cloning methods

We generated different genetic constructs to test the different fluorescent proteins and to identify new incompatible lox sites. For that, we designed primers to amplify the interest coding regions of multiple fluorescent proteins from commercial vectors acquired in Addgene repository (Table 1). The primers were designed to contain in their tail the sequences of restriction sites to allow their insertion by cohesive digestion/ligation into the desired digested expression vectors (Table 2). Typically, 15-20 units of each appropriate restriction enzyme or combination were used to digest either vectors or PCR products for 1h at 37ºC. The digested PCR products were purified with Thermo Scientific GeneJET PCR Purification Kit (#K0702). Digested vectors were purified by gel excision and the Thermo Scientific GeneJET Gel Extraction Kit (#K0692).

Table 1. Reference, name and use of vectors acquired in Addgene.

Reference	Vector	Use
45179	pThy1-Brainbow3.2	Brainbow3.2 plasmid used to amplify fWpA, loxP-lox2272-loxN, mKate and mOrange
13775	pCAG-CRE	Plasmid used to amplify and express CRE
11180	pCAG-YFP	Plasmid used to amplify YFP
54648	YPet-C1	Plasmid used to amplify YPet
54715	Citrine-C1	Plasmid used to amplify Citrine
45185	AAV-EF1a-BbTagBY	Plasmid used to amplify Tag-BFP
54412	LI D-E T-Saph + linker c1	Plasmid used to amplify t-Sapphire
51957	Sirius/pcDNA3	Plasmid used to amplify Sirius

As a general rule 1 unit of T4 DNA Ligase (Thermo Scientific #EL0014) was used to ligate the digested vector and the digested insert. We mix between 5 to 10 molecules of insert for each molecule of vector. The ligation reaction was incubated at 22ºC for 1h. After ligation, all of the ligation mix is desalted for 20 minutes. 2µl of microdialyzed ligation reaction were transformed by electroporation in DH5α competent cells, recuperated for 1h at 37ºC in LB without antibiotic and plated with the correct antibiotic according to the resistance gene contained in the vector. We analysed the transformants by PCR directly on some cells taken from the colony or by extracting the DNA (Themo Scientific GeneJET Plasmid Miniprep Kit #K0503) and performing restriction analysis. Finally, all potential constructs were sequenced using StabVida YouTube system (http://www.stabvida.com).

To clone the lox site sequences, two complementary oligonucleotides containing the lox site sequence surrounded by the cohesive sequences of two restriction sites were designed (Table 2). These oligonucleotides were phosphorylated and hybridized *in vitro*. 10 units of T4 polynucleotide kinase (T4 PNK, Thermo Scientific #EK0031) were used to phosphorylate the oligonucleotides. Subsequently, the PNK reaction was incubated at 37ºC for 30 minutes, heated at 95ºC for 15 minutes to inactivate the enzyme and to denature the DNA and incubated at 70ºC for 15 minutes to allow the hybridization of both oligonucleotides. Then, inserts were ligated to the digested vector.

Table 2. Primers designed to amplify the coding sequences of interest from different vectors.

Primers used to generate the constructs for testing the different fluorescent proteins	
Primers to clone Farnesyl-WPRE-pA from pThy1-Brainbow3.2 with restriction sites BamHi-EcoRi	
Bam-fWPREpA-F	ACGTGGATCCAGTGCACGCGTGAGTAAGC
Eco-fWPREpA-R	ACGTGAATTCCGCCTTAAGATACATTGATGAGTTT
Primers to amplify mOrange2 from pThy1-Brainbow3.2 with restriction sites HindIII-BamHI to clone in pCDNA3.1-fWPRE	
5HindmOrange2	ACGTAAGCTTGGCGTGCTAGCATAACTTCG
3BglOrange2	ACGTAGATCTCAGGGTCAGCTTGCCGTAG
Primers to amplify t-Sapphire from LI D-E T-Saph + linker c1 with restriction sites HindIII-BglII to clone in pCDNA3.1-fWPRE	
TSaphire-FHind	ACTGAAGCTTATGGTGAGCAAGGGCGAGGAGCTGTT
TSaphire-RBgl	ACTGAGATCTCTTGTACAGCTCGTCCATGC
Primers to amplify mKate from pCDNA3.1-mKate2 with restriction sites HindIII-BglII to clone in pCDNA3.1-fWPRE	
HindIII-mKateFwd	ACGTAAGCTTCGCTATGGTGAGCGAGCTGAT
BglII-mKateRev	ACGTAGATCTTGCACTTCTGTGCCCCAGTT
Primers to amplify TagBFP from AAV-EF1a-BbTagBY with restriction sites HindIII-BglII to clone in pCDNA3.1-fWPRE	
TagBFP-RevF2Bgl	ACTGAGATCTCTGTGCACTTCTGTGCCCCAGT
TagBFP-FwdHind	ACTGAAGCTTGCTCTCCTGATACCGTTCGT
Primers to amplify Ypet from YPet-C1 with restriction sites HindIII-BglII to clone in pCDNA3.1-fWPRE	
YPET-FHind	ACTGAAGCTTATGGTGAGCAAAGGCGAAG
YPET-Rbgl	ACTGAGATCTCTTATAGAGCTCGTTCATGCCCT
Primers used to generate the construct for testing incompatible lox sites based on FPs	
Primers to amplify Trilox from pTEC with restriction sites BamHI-EcoRI	
Trilox-BamHI-Fwd	ACTGAAGCTTCGGTCCTAAGGTAGCGAACC
Trilox-EcoRI-Rev	ACTGGAATTCATGGTACCGGTGGCGATATCAGATAACTTCG
Primers to amplify EGFP from pEGFP-C1 with HindIII-BamH1 extremes	
HindIII-EGFPFwd	ACTGAAGCTTTTTAGTGAACCGTCAGATCC
BamHi-EGFPRev	ACGTGAATTCCGCCTTAAGATACATTGATGAGTTT

Primers to amplify mCherry with EcoRI-XbaI extremes	
EcoRI-mCherryFwd	ACGTGAATTCATGGTGAGCAAGGGCGAGGAG
XbaI-mCherryRv	ACGTTCTAGACTTGTACAGCTCGTCCATGCCGCC
Primers used to generate the construct for testing incompatible lox sites based on bacteria	
Primers to amplify CRE recombinase from pCAG-CRE with NcoI and XhoI extremes	
NcoI-CRE-Fwd	ATCGCCATGGCCAATTTACTGACCGTACAC
XhoI-CRE-Rev	ATCGCTCGAGCTGATAGGCAGCCTGCACCT
Primers to amplify Kanamycin resistance with its promotor from pEGFP-C1	
BglII-Kanr-Fwd	ACTGAGATCTGAACGTGGCGAGAAAGGAAG
HindIII-Kanr-Rev	ACTGAAGCTTGGTCATTTCGAACCCCAGAG
Primers to amplify Trilox from pTEC with restriction sites HindIII-EcoRI	
Trilox_HindIII_Fwd	ACTGAAGCTTCGGTCCTAAGGTAGCGAACC
Trilox_EcoRI-AgeI_Rev	ACTGGAATTCATACCGGTGGTGGCGATATCAGATAACTTCG
Primers to clone LoxP in pET3a construct with SalI-BglII(PstI) extremes	
LoxP-SalIBglII-F	TCGACCTGCAGATAACTTCGTATAGCATACATTATACGAAGTTATA
LoxP-SalIBglII-R	GATCTATAACTTCGTATAATGTATGCTATACGAAGTTATCTGCAGG
Primers to clone Lox511 in pET3a construct with SalI-BglII(PstI) extremes	
Lox511-SalIBglII-F	TCGACCTGCAGATAACTTCGTATAGTATACATTATACGAAGTTATA
Lox511-SalIBglII-R	GATCTATAACTTCGTATAATGTATACTATACGAAGTTATCTGCAGG
Primers to clone LoxM3 in pET3a construct with SalI-BglII(PstI) extremes	
LoxM3-SalIBglII-F	TCGACCTGCAGATAACTTCGTATATGGTATTATATACGAAGTTATA
LoxM3-SalIBglII-R	GATCTATAACTTCGTATATAATACCATATACGAAGTTATCTGCAGG
Primers to clone LoxM7 in pET3a construct with SalI-BglII(PstI) extremes	
LoxM7-SalIBglII-F	TCGACCTGCAGATAACTTCGTATATTCTATCTTATACGAAGTTATA
LoxM7-SalIBglII-R	GATCTATAACTTCGTATAAGATAGAATATACGAAGTTATCTGCAGG
Primers to clone Lox2272 in pET3a construct with SalI-BglII(PstI) extremes	
Lox2272-SalIBglII-F	TCGACCTGCAGATAACTTCGTATAGGATACTTTATACGAAGTTATA
Lox2272-SalIBglII-R	GATCTATAACTTCGTATAAAGTATCCTATACGAAGTTATCTGCAGG
Primers to clone LoxN in pET3a construct with SalI-BglII(PstI) extremes	
LoxN-SalIBglII-F	TCGACCTGCAGATAACTTCGTATAGTATACCTTATACGAAGTTATA
LoxN-SalIBglII-R	GATCTATAACTTCGTATAAGGTATACTATACGAAGTTATCTGCAGG

4.2 DNA transfection into eukaryotic cell lines

Experiments of fluorescent proteins expression and compatibility evaluation of lox sites were done in the Human Embryonic Kidney 293T cell line (HEK-293T). HEK-293T cells are a specific cell line derived from human embryonic kidney cells. They are very easy to grow and transfect very readily. HEK-293T cells contains the SV40 Large T-antigen, that allows for episomal replication of transfected plasmids containing the SV40 origin of replication which allows for amplification of transfected plasmids and extended temporal expression of the desired gene products[28].

DNA was introduced into the host cells by transfection with polyethylenimine (PEI). PEI is a polycation which condenses DNA into positively charged particles that bind to anionic cell surfaces. Consequently, the DNA-PEI complex is endocytosed by the cells and the DNA released into the cytoplasm[29]. We used 1'5mg of DNA and 6µl of PEI to transfect the cells in 6-well plates (ratio DNA:PEI 1:4). For best transfection efficiency, cells were transfected at 60% of confluence.

4.3 Fluorescent proteins evaluation

To analyse the cells by confocal microscopy, we plated the cells over a glass coverslip previously washed with ethanol and coated with poly-d-lysine to promote the cells attachment. In the case of FACS analysis, cells were trypsinised and washed and resuspended in PBS. The maximum signal is reached 48h after transfection. To evaluate the fluorescent proteins, we used:

- A fluorescent confocal microscopy Leica TCS SP5 with five lasers of 405nm, 488nm, 514nm, 543nm and 549nm of wavelength. Different excitation windows were designed in order to recover uniquely as much light as possible for each of the individual proteins.
- A FACS Aria III system (Fluorescence-Activated Cell Sorting) from Becton Dickinson with the following configuration: laser 407nm (filters: 450/50, 610/20, 710/50, 780/60, 660/20, 525/50), 488nm laser (filters: 488/10, 695/40, 530/60), 561nm laser (filters: 610/20, 710/50, 780/60, 670/14 and 582/15) and 633nm laser (filters: 670/30, 780 / 60, 730/45).

Table 3 summarizes the laser and filter combinations used to register the signals for the FPs.

Table 3. Laser and filters combination of different FPs for the confocal microscopy and the sorter. Emission ranges are referred to the detection filters used to detect each protein.

	Confocal microscopy		Sorter/FACS	
Fluorescent Protein	Excitation	Emission	Excitation	Emission
Tag-BFP/Sirius	405 nm	445/25 nm	407 nm	405/50 nm
EGFP	488 nm	505/10 nm	488 nm	530/30 nm
mOrange2	543 nm	575/15 nm	561 nm	582/15 nm
YFP/Citrine/YPet/Phi-YFP	514 nm	535/10 nm	670 nm	530/30 nm
mKate2	594 nm	670/30 nm	505 nm	670/30 nm
tSapphire	405 nm	505/10 nm	407 nm	527/50 nm

4.4 *In vitro* CRE Recombination

To test the action of CRE *in vitro* we incubated 250 ng of DNA previously linearized, with the CRE recombinase (NEB #M0298S) at 37ºC for 30 minutes. The CRE recombinase reaction mix was heated at 70ºC for 10 minutes before agarose gel analysis to inactivate the enzyme. Linearized pLox 2+ was used as DNA control. It is 3625 bp in length, with a loxP site approximately 400 bp from each end. Recombination between these loxP sites produces a circular 2787 bp (which migrates at approximately 1700bp on a 0.8% agarose gel) and one 838 bp DNA fragment (Figure 4).

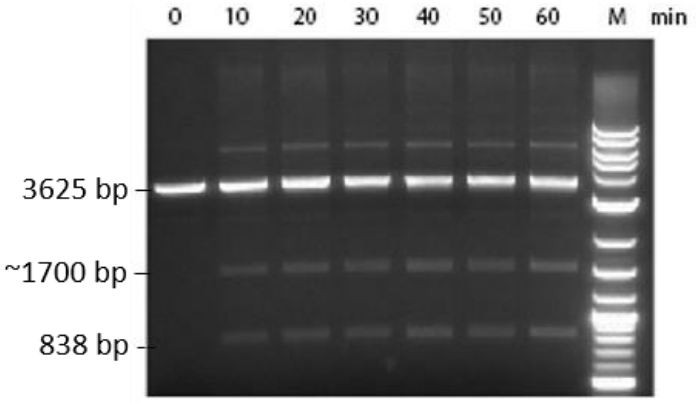


Figure 4. Agarose gel analysis of CRE Recombinase Reaction with LoxP 2+ control substrate. The reaction yields a 20-30% recombination.

4.5 Electrocompetent cells

For the generation of electrocompetent cells we grow the desired *E.Coli* strain in LB until it reaches a OD_{600nm} between 0.4 and 0.6. Subsequently, we performed two washes with cold water and two washes with a cold 10% glycerol solution keeping the cells at 4ºC during the whole procedure. Finally, cells were snap frozen in dry ice and kept at -80 ºC until used.

During this project, we prepared electrocompetent cells from DH5α, $C41^{pCDF1b\text{-}CRE}$ and $C3013^{pCDF1b\text{-}CRE}$ *Escherichia coli* strains.

4.6 IPTG induction

We induced T7 promoter expression by adding isopropylthio-β-galactoside (IPTG). IPTG is an inducer of gene expression in bacteria. It is not part of any metabolic pathway and so it is not broken down or used by the cell. This ensures that the IPTG concentration added remains constant, making it a very useful inducer.

We used the pCDF-1b vector (Novagen #71330-3) that contains a gene coding for Streptomycin resistance and the lacI gene from the lac operon that codes for the lac repressor (LacI). CRE was inserted just after the T7 promoter and the lac operator DNA sequences of

pCDF-1b. We induced CRE when OD_{600nm}=0.4. IPTG binds to the lac repressor (LacI) and induces a conformational change in the protein structure that reduces its affinity for the operator DNA sequence. In the absence of repressor, The T7 RNA polymerase can bind the T7 promoter and perform the transcription of the gene after the promoter.

For induction experiments we used C41 (a kind gift from Elena Cabezón, IBBTEC) and C3013 (NEB # C3013I) *E. coli* strains. The strain C41 have been cited in over 350 protein expression performances, they express genes cloned in any T7 vector. The strain C3013 is suitable for high level of expression control. C3013 competent cells have controlled induction of T7 RNA polymerase and, consequently, inducible control of transcription of genes downstream of the T7 promoter. Also, the promoter activity is minimum until IPTG is added.

pCDF-1b-CRE was transformed into these competent cells and plated them with streptomycin. The culture was grown at 37ºC to OD_{600nm}=0.4 at which point IPTG was added at 1mM and left growing for the desired induction time.

4.7 Protein expression in SDS-PAGE

Bacterial cells were lysed with 100µl of Lysis Buffer (2% SDS, 50mM Tris) and heated at 95ºC for 5 minutes to produce protein denaturalization. Protein extracts were quantified with the Qubit Protein Assay Kit (Thermo Scientific #Q33212). We loaded 50µg of protein on a 12% SDS-PAGE gel. We run our gels at 25mV per gel. Subsequently, we stained the gel with Coomassie Blue dye in methanol and glacial acetic acid for 45 min with agitation. Excess dye is washed out by 'destaining' overnight with tab water.

5 Results

5.1 Identification of different fluorescent proteins

We transfected HEK293-T cells with different constructs containing the coding sequence of 10 fluorescent proteins into pCDNA3.1 eukaryotic expression vector in frame with the farnesylation signal sequence (f, which lead the location of the expressed proteins to the cell membrane), a polyadenylation signal (pA) and the woodchuck hepatitis virus post-transcriptional regulatory element (W). This strategy allows a stable membrane-bond expression of the fluorescent protein as described in the Brainbow 3.2 strategy[17]. 48 h after transfection, the cells were prepared and analysed by confocal microscopy and we generated the emission spectra for each protein by the measure at multiple wavelengths of the signal intensity. The resulting emission spectra for each protein are represented in Figure 5. According to these results, we selected five FPs (tag-BFP, tSapphire, YPet, mOrange2 and mKate2) that cover the whole light spectrum from blue to red and whose spectra show the minimum overlapping possible. tSapphire and YPet proteins show quite a lot of spectra overlapping but this will not suppose a major difficulty due to the fact that tSapphire protein is a variant of EGFP that is only excited in the violet range where YPet is not excited. In that way, combining different excitation sources, we will be able to distinguish both proteins without problems.

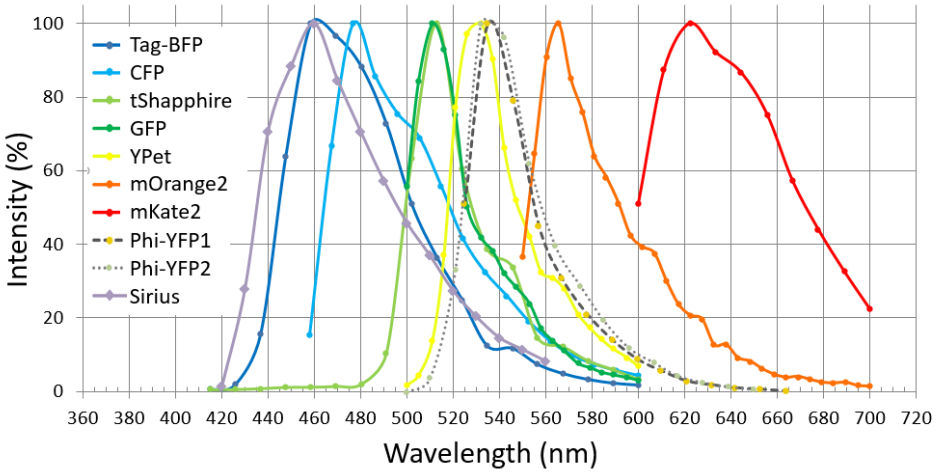


Figure 5. Emission spectra diagram of several FPs. We analysed several fluorescent proteins in the confocal microscopy in order to select an optimal combination of them.

According to the observations below, we repeated the transfection experiments and selected a combination of laser/filter combinations in the confocal microscopy to test our ability to uniquely distinguish the five selected proteins. Representative images of the signals recorded by cells individually transfected with the 5 different FPs are shown in Figure 6A. tSapphire, YPet, mOrange and mKate signals are quite clearly detected only in their respective filter combination with almost no pass-through to any other detector. Unfortunately, some tag-BFP signal is detected in the

channel that we had selected for tSapphire. This is likely due to the broad emission spectra showed by tagBFP that presents some overlapping with tSapphire spectra. Unfortunately, both proteins are excited at the same wavelength. Nevertheless, the signal percentage of tagBFP that is detected in tSapphire channel is quite low and we think we would be able to compensate it to uniquely distinguish the five FPs anyway.

As a second step, we performed FACS analysis in collaboration with the group of Pablo Menendez at the University of Barcelona on the same cell populations. As we can observe in Figure 6B, we are able to separate the cell populations individually transfected with the different FPs.

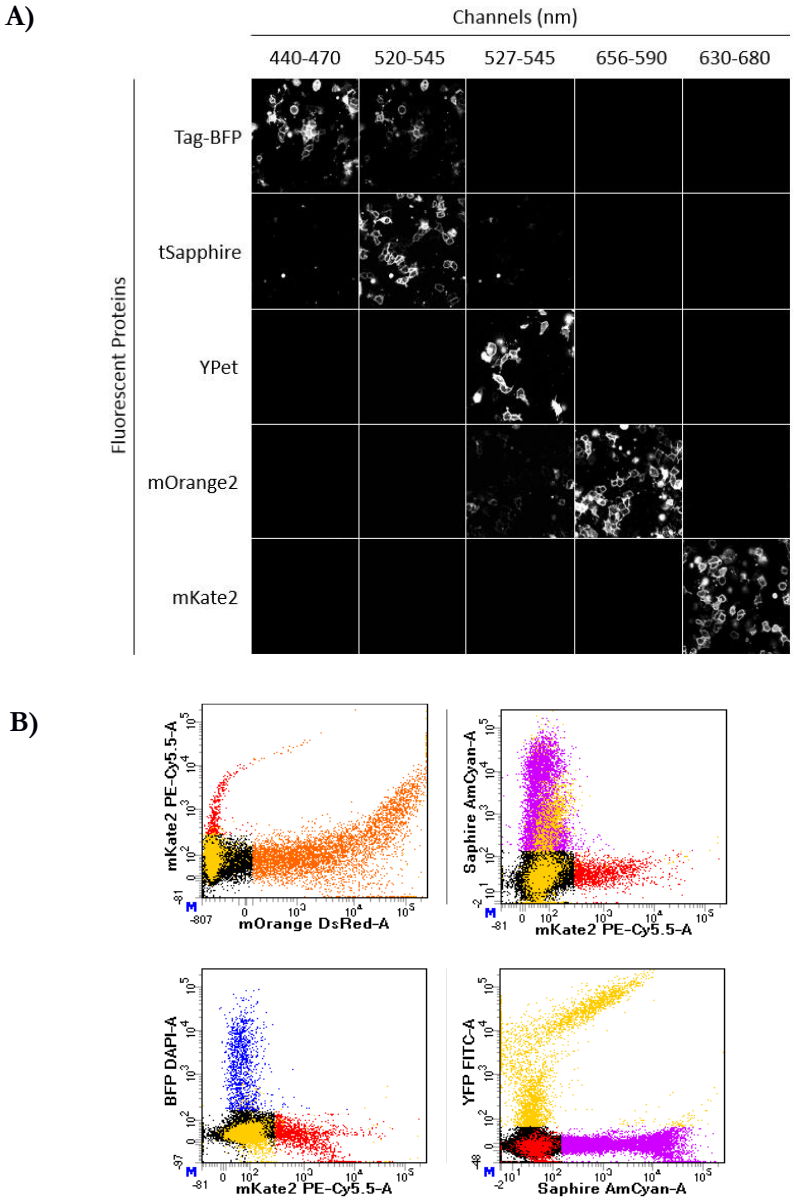


Figure 6. Confocal microscopy and FACS experiments. A) Representative images of cells transfected with the five different fluorescent proteins taken at five different emission intervals in the confocal microscopy designed to maximize the unique identification of each protein. **B)** Different dot-plots of the signal reported by each transfected population in the FACS.

5.2 Identification of new incompatible lox sites based on FPs recombination

According to the results obtained in the previous section, we would be able to include up to five different fluorescent proteins in our improved designed. In order to incorporate two new fluorescent proteins in our allele following Brainbow 3.2 strategy, we need to identify two new incompatible lox sites. Two new lox sites (loxM3 and loxM7) have been recently described as incompatible with loxP[15]. Nevertheless, to be able to use them we need to check if they are also incompatible with the other two lox sites included in the Brainbow 3.2 allele (loxN and lox2272) as well as if they are incompatible between them. For this purpose, we constructed an expression vector to test the incompatibility of lox sites that we have called pTEC. We cloned consecutively the three incompatible lox sites already described in Brainbow3.2 (loxP, lox2272 and loxN), a green fluorescent protein (EGFP) and a red fluorescent protein (mCherry) on the commercial vector pCDNA3.1. There are two unique restriction sites (BamHI and EcoRI) between both fluorescent proteins to allow the cloning of new lox sites between them keeping the reading frame. This vector produces the expression of a fusion protein containing both fluorescent proteins. After the action of CRE recombinase, if a recombination event occurs, only the expression of mCherry would be detected (Figure 7A).

We generated different constructs containing loxM3 and loxM7 sites in this position and transfected these constructs into HEK-293T cells alone or together with a CRE expressing vector (Addgene #13775). Unfortunately, in most of the experiments performed, after the action of CRE, we detected a decrease of the green signal but not a complete loss of it. A representative image of this phenomenon with loxM7 is shown in Figure 7B. This is likely the result of an incomplete recombination rate among a variable number of vector molecules present inside the same cell. An uncontrolled number of molecules inside of each cell precludes a clean and clear quantification of the recombination rate. Additionally, we also observed an unexpected decrease in mCherry signal. We hypothesize that the potential recombination can interfere with the transcription of mCherry as the sequence between the promoter and the translation start point is modified. According to that, we also observe a decrease in mCherry signal in the positive control with loxP. In any case we decided that our strategy was not optimal for the quantification of the recombination rate. To try to improve this quantification, we decided to test the recombination *in vitro*.

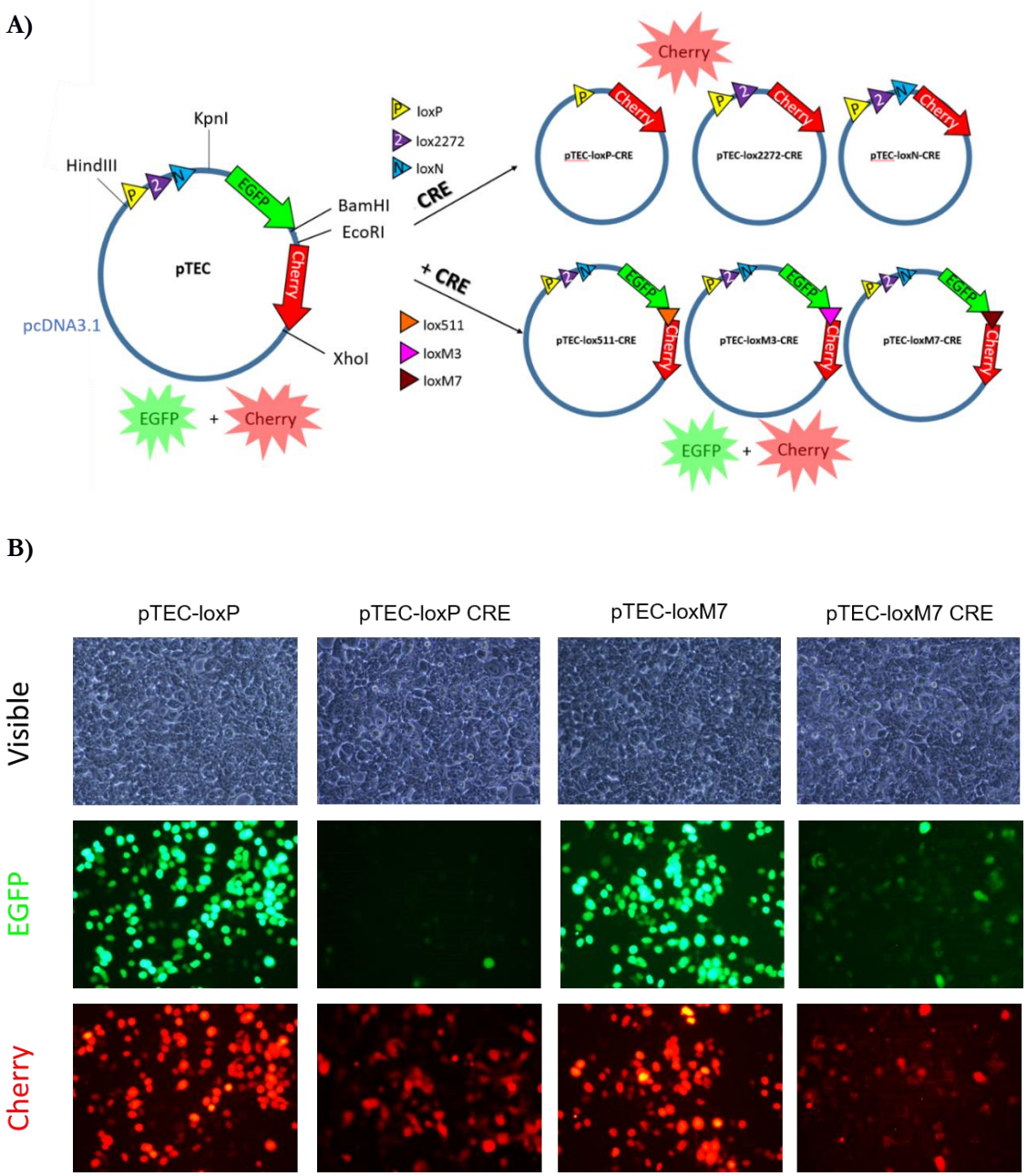


Figure 7. System for the identification of new incompatible lox sites. A. Different lox sites are introduced between EGFP and mCherry where there are two unique restriction sites (BamHI and EcoRI). By adding CRE recombinase, if the lox sites are incompatible, both fluorescent proteins would be expressed (we would see red and green fluorescence) and if they are compatible, only the expression of mCherry (red fluorescence) could be detected. **B. Lox compatibility assay.** HEK-293T cells transfected with pTEC-loxP and pTEC-loxM7 observed by fluorescence microscopy with and without CRE co-transfection. We observe a decrease of both fluorescence signals when we add CRE. Also, there is recombination in pTEC-loxP CRE.

5.3 CRE Recombination assay *in vitro*

To try to solved the problems observed in the previous experiments, we tested CRE recombination *in vitro*. We linearized the pTEC-loxP construct with a restriction enzyme (NruI). Linearized pTEC-loxP is ~7200 bp in length. It contains EGFP flanked by two lox sites. The potential recombination between these lox sites should produce a ~6450 bp DNA fragment and a circular ~750 bp (EGFP) (Figure 8A). The linearized pTEC-loxP was incubated with the CRE recombinase (NEB #M0298S) at 37ºC for 30 minutes. The CRE recombinase reaction mix was heated at 70ºC for 10 minutes before agarose gel analysis (Figure 8B). Linearized pLox 2+ was used as DNA control. It is 3625 bp in length, with a loxP site approximately 400 bp from each end. Recombination between these loxP sites produces a circular 2787 bp (which migrates at approximately 1700bp on a 0.8% agarose gel) and one 838 bp DNA fragment (Figure 8B). As it can be seen in Figure 8B, we only observed a very faint band corresponding to the recombination in both our positive control (with loxP inserted between EGFP and mCherry) and in the commercial positive control purchased with CRE recombinase. This is surely not enough sensitivity to detect low recombination rates. According to that we decided to change completely our strategy and moves to a system based in antibiotic resistance genes in bacteria.

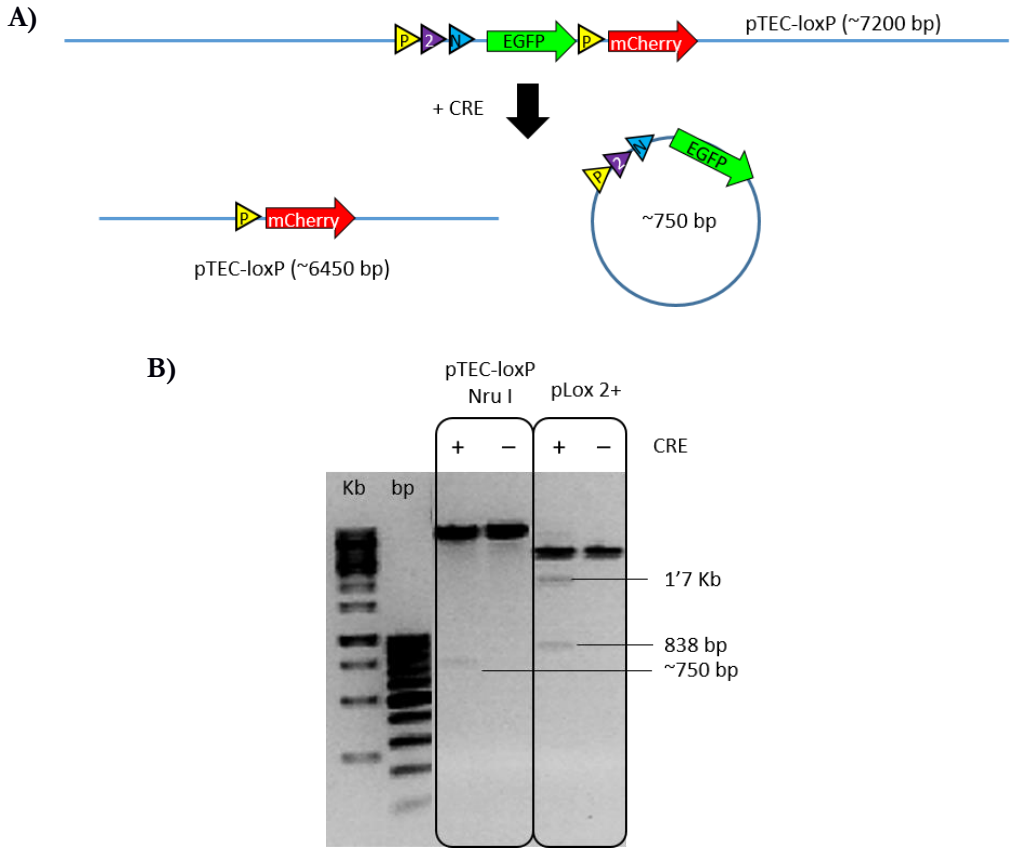


Figure 8. CRE Recombination assay *in vitro*. A) Scheme of recombination in linearized pTEC-loxP. Recombination between the loxP sites flanking EGFP produces a ~6450 bp DNA fragment and a circular ~750 bp. **B)** Agarose gel analysis of CRE Recombinase Reaction with linearized pTEC-loxP and LoxP 2+ control substrate. The reaction yields a 20-30% of recombination rate. GeneRuler 1Kb (Thermo Scientific #SM0311) and 100bp (Thermo Scientific #SM0214) DNA Ladders.

5.4 Identification of new incompatible lox sites based on bacteria

According to the results shown in the previous sections, we designed a complementary system (Figure 9) to identify new incompatible lox sites. Taking advantage of the low rate of transformation observed in bacteria, only a single molecule is introduced in each bacteria cell in most of the transformants which provides us with a big advantage for a clear quantification. According to that, we reasoned that a similar strategy of a construct that could be recombined could work better on bacteria expressing CRE recombinase after the induction with IPTG.

For that, we generated a construct in where we cloned CRE coding sequence extracted from pCAG-CRE vector (Addgene #13775) into pCDF-1b vector at the NcoI and XhoI sites. To test that the construct worked, we transformed this construct in C41 *E. coli* cells, induced the expression with IPTG at different times and performed SDS-PAGE analysis to detect CRE protein. As it can be seen in Figure 8, a band compatible with CRE size is observed even in the absence of induction. Consequently, we decided to change to another *E. coli* strain, C3013, that is supposed to have a tighter control of the T7 promoter activity in the absence of any induction. We transformed the same construct on C3013 and induced the expression of CRE with IPTG. As it can be seen in Figure 8, in this cells, the expression of CRE in the absence of IPTG is no detectable whereas the production increases very significantly after 4h of IPTG induction. Consequently, we decided to generate electrocompetent cells with the C3013 clone that is already transfected with pCDF1b-CRE.

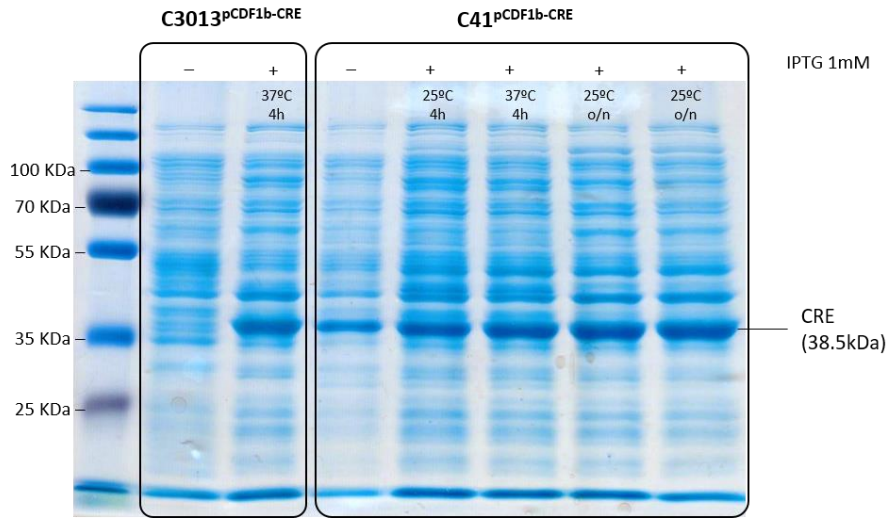


Figure 8. Cre expression in SDS-PAGE. CRE (38.5 kDa) is expressed in C3013$^{pCDF1b\text{-}CRE}$ after IPTG 1mM induction and it is C41$^{pCDF1b\text{-}CRE}$ with and without IPTG 1mM induction. We tested different incubation times and temperatures after the induction and we found out that the induction for 4h at 37ºC was optimal.

Once we generated the inducible CRE expressing bacteria, we designed a system to test lox site incompatibility (Figure 9) which consists on a plasmid with two antibiotic resistance genes (Kanamycin and Ampicillin). One of them (Kanamycin resistance gene) is flanked by lox sites (loxA

and loxB places). After a potential recombination between the two lox sites, this antibiotic resistance gene would be lost. To test the recombination rate, the different constructs were transformed into the C3013$^{PCDF1b\text{-}CRE}$ bacteria. Once transformed, CRE expression was induced for 4h to generate the potential recombination between the lox sites. Subsequently, we extracted the DNA of these bacteria in where we would obtain a mixture of plasmids, some of the recombined and others not. With this DNA we performed a second round of transformation. As only a single plasmid molecule would be introduced in each cell, we can calculate the recombination rate by parallel seeding in ampicillin and kanamycin and colony counting.

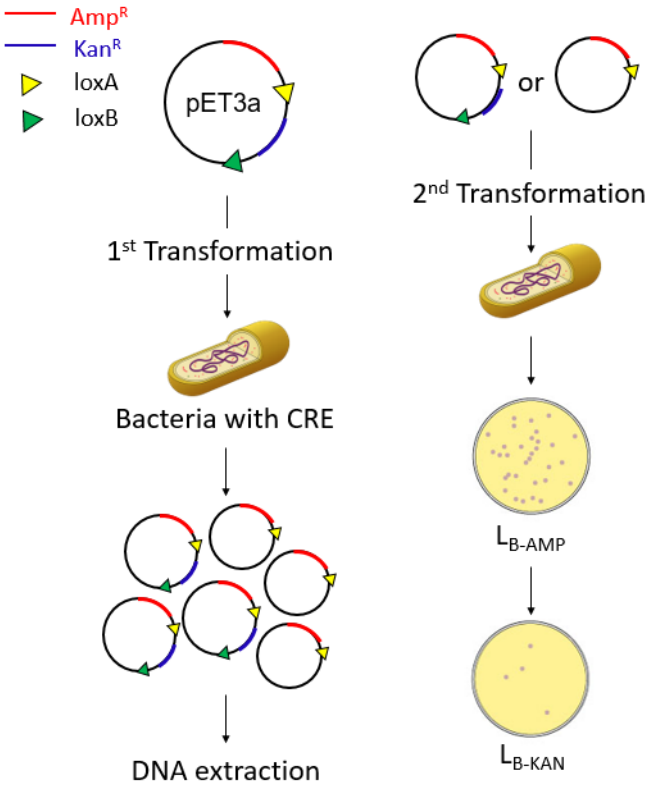


Figure 9. System to test lox site incompatibility. Scheme of the methodology used to test incompatible lox sites and to identify new ones based on a construct with two antibiotic resistance genes (Ampicillin and Kanamycin), one of them (Kanamycin resistance gene) flanked by lox sites. After transformation in CRE expressing bacteria, the number of recombinant molecules is evaluated by parallel seeding on plates with the individual antibiotics.

Firstly, in order to test that our system works, we generated two control constructs. In both of them, we cloned the three lox sites of the original model (loxP-lox2272-loxN) in one of the lox positions(loxA). Afterwards, we cloned either loxP or lox511 in loxB position. In both cases we expected a 100% recombination rate, in the first case because we have a loxP site in both lox positions, and in the other because of the described compatibility between lox511 and loxP. As can be observed in Figure 10, we observed a hundred percent of recombination in both cases so we can conclude that our system works.

Secondly, we tested the described incompatibility between the two lox sites included in the original Brainbow 3.2 allele (loxN and lox2272) with loxP. For that we cloned loxP in loxA place and either loxN or lox2272 in loxB place. We observed 0% recombination in these cases (Figure 10B).

Additionally, we tested the compatibility of loxM3 and loxM7, already described as incompatible with loxP, with the other two lox sites contained in the Brainbow 3.2 allele (loxN and lox2272). We observed a 50% of recombination in loxM7 and less than 20% of recombination in loxM3. Finally, we tested loxM3 against loxM7 in order to find out if they were incompatible between them. We observed no recombination between them (Figure 10B).

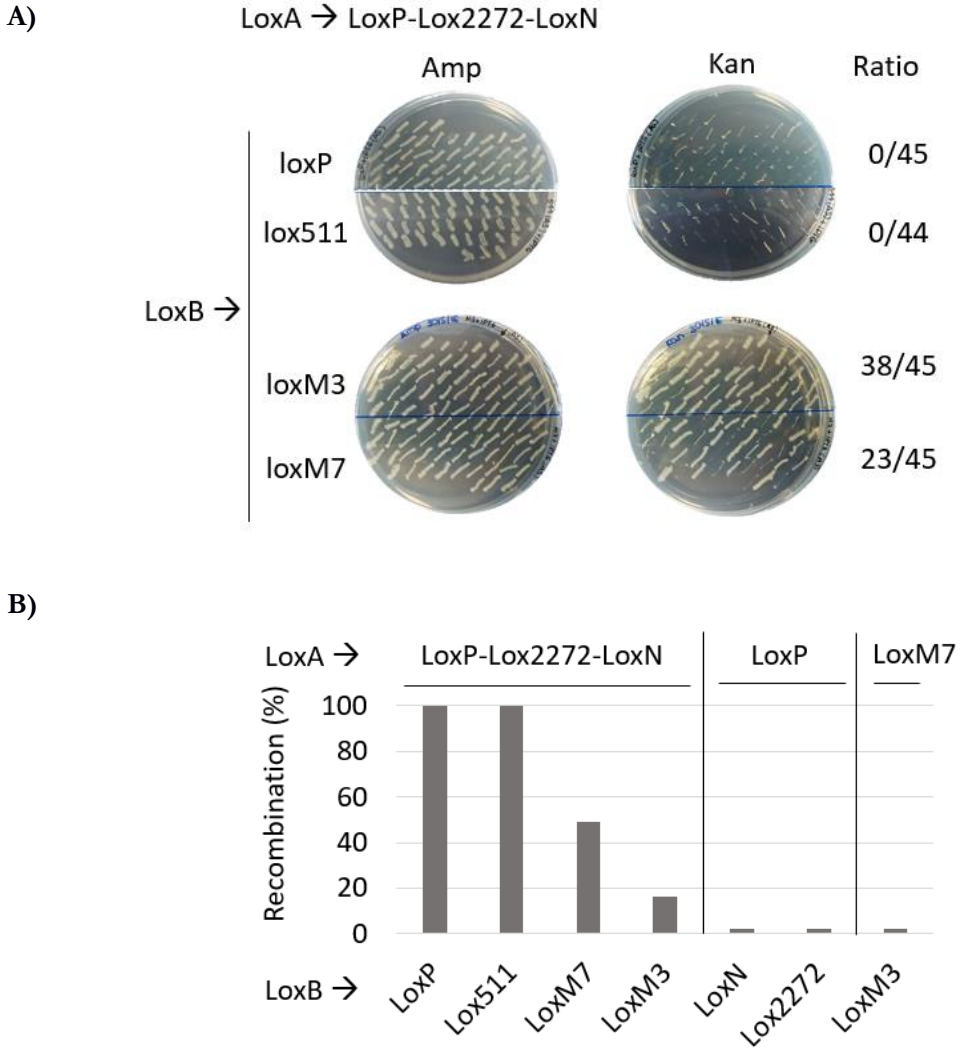


Figure 10. Identification of new incompatible lox sites based on bacteria. A) Representative images of colony growth ratio between the two plate types in each loxA-loxB combination. Amp: ampicillin. Kan: kanamycin. **B)** Diagram showing the recombination rate between different lox sites.

6 Discussion and ongoing work

Understanding the dynamics and molecular mechanisms of intratumour heterogeneity as well as its role in tumour progression could suppose a great step forward in the diagnosis and treatment of cancer patients. According to that we hypothesized that cell lineage genetic tracking systems combined with mouse models could be an essential tool to study this phenomenon. Unfortunately, available tools present several disadvantages, being the principal one the limited number of potential label combinations, that precludes their use for the study of intratumour heterogeneity.

In the present project we have identified an optimal combination of five different FPs that can be uniquely distinguished by confocal microscopy. Moreover, the signals produced by each protein can be identified and separated by FACS which opens the opportunity, in the context of intratumour heterogeneity, to study individually the different genetic clones present in a tumour.

To introduce two additional fluorescent proteins to the Brainbow 3.2 design, we needed two additional incompatible lox sites. In this context, we have designed and validated a new system to check the compatibility of lox sites. Using this system, we have tested the compatibility of two newly identified lox sites (loxM3 and loxM7) that have been described as incompatible with loxP. We have demonstrated that loxM3 shows a recombination rate of less than 20% with the three lox sites already included in the Brainbow 3.2 allele. In the other side, loxM7 shows a recombination rate of 50%. This recombination rate is higher than what we desired but is significantly lower than the normal recombination rate between compatible lox sites like lox511 with loxP. Additionally, we have demonstrated that it is completely incompatible with loxM3. Consider all this, we think that it is still viable to use loxM7 in our improved design as far as it is located in the last position of fluorescent markers. In that way, once it recombines, no further recombination will be possible. Nevertheless, we will continue checking additional lox sites during the time it takes us to construct the final allele in case we identify a better lox site option.

According to all the exposed results, we can design a new allele (Figure 11) with the 5FPs surrounded by incompatible lox sites. We have already started to construct this allele. With this we propose to generate a knock-in mice containing this new allele integrated in the locus *ColA1* (ColA1^{CFP5}.) In homozygous mice, the independent behaviour of each allele will give rise up to 15 different colour combinations that might be very useful for the study of the dynamics of intratumour heterogeneity.

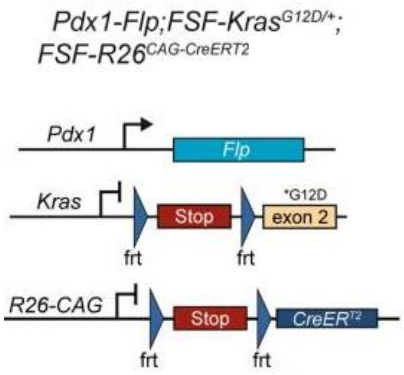
Figure 11. Final allele scheme. Representation of the new lineage genetic tracing designed allele for the study of intratumour heterogeneity in cancer mouse models.

Our system can be combined with any cancer mouse model in order to study the extent of intratumour heterogeneity in different tumour types as well as in different situations. In this context, we count already with a pancreatic cancer model generated by Drs. Roland Rad and Dieter Saur (Technical University of Munich, Germany). This model is based on the conditional expression of a specific oncogenic version of *Kras* (G12D) specifically in pancreas based in the action of a transgenic allele of FLP recombinase under a pancreas specific promoter (Pdx1). These mice develop pancreatic metastatic pancreatic ductal adenocarcinomas (PDAC) that fully recapitulates all types of lesions observed in human PDAC progression (Figure 12B). Recently, these groups have generated an extended model containing an additional knock-in allele of the CRE recombinase inducible by tamoxifen (Figure 12A)[31]. We have recently imported this model to our institute and therefore it will be ready to cross with our mice as soon as they are ready.

A major advantage of this proposed combined model is that the fluorescent labelling is activated independently of the tumour initiation events, allowing us to induce it at different progression stages of the tumour. A systematic study inducing the labelling at different times points during tumour progression will inform us first about the precise time-frame in which the clonal heterogeneity originates and how this heterogeneity is represented in metastasis. Also, we will be able to observe if this dynamic is similar in pre-malignant and metastatic lesions and to identify the clones responsible for the generation of metastases, their abundance in the primary tumours and the grade of heterogeneity of these metastases. Moreover, we could determine the phylogenetic relationship between the metastatic growths and the clones present in the primary tumours. Finally, we will establish a temporal time frame for the metastatic seeding.

Furthermore, taking advantage of the unique fluorescent tag shown by the different cell clones present in the primary tumours, we propose to purify them by FACS as explained before. This will allow us, for the first time, a precise characterization of each clone independently.

A)



B)

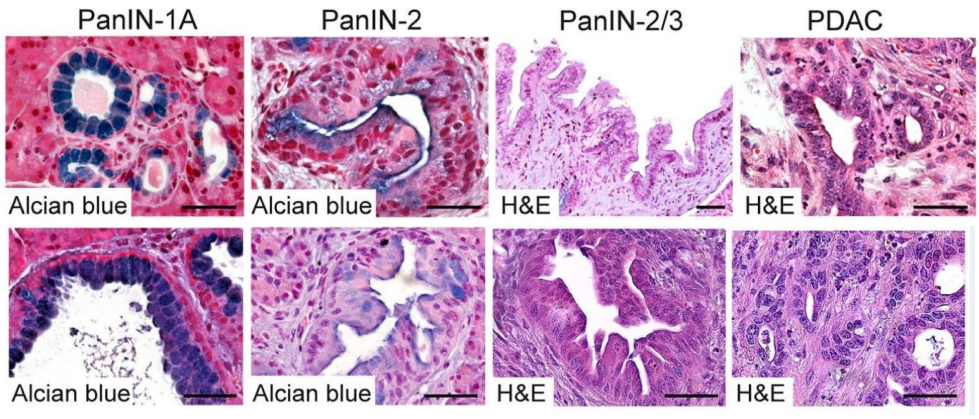


Figure 12. A) Genetic modifications of the conditional mouse model based on the oncogenic version of *Kras* (G12D) specifically in pancreas based FLP recombinase action. **B)** Representative alcian blue and hematoxylin and eosin (H&E) stained sections of different grades (1-3) of pancreatic intraepithelial neoplasia (PanIN) and invasive PDAC in the mice developed that fully recapitulates all types of lesions observed in human PDAC progression.

7 Conclusions

- We have identified a combination of five different fluorescent proteins that can be uniquely distinguished by confocal microscopy and separated by FACS.
- We have designed and validated a new system for the identification of new incompatible lox sites.
- We have tested the incompatibility of two new lox sites with the three already used by other authors (loxP, loxN and lox2272).
- With all the results above, we can design a new allele able to produce up to 15 different colour combinations that might be very useful for the study of intratumour heterogeneity in cancer mouse models.

8 References

1. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458,** 719–724 (2009).
2. Wang, Z.-Y. & Chen, Z. Acute promyelocytic leukemia: from highly fatal to highly curable. *Blood* **111,** 2505–2515 (2008).
3. Rhim, A. D. *et al.* EMT and Dissemination Precede Pancreatic Tumor Formation. *Cell* **148,** 349–361 (2012).
4. Campbell, P. J. *et al.* The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467,** 1109–1113 (2010).
5. Gerlinger, M. *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* **366,** 883–892 (2012).
6. Fidler, I. J. Timeline: The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nat. Rev. Cancer* **3,** 453–458 (2003).
7. Magee, J. A., Piskounova, E. & Morrison, S. J. Cancer Stem Cells: Impact, Heterogeneity, and Uncertainty. *Cancer Cell* **21,** 283–296 (2012).
8. Fisher, R., Pusztai, L. & Swanton, C. Cancer heterogeneity: implications for targeted therapeutics. *Br. J. Cancer* **108,** 479–485 (2013).
9. Stephens, P. J. *et al.* Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. *Cell* **144,** 27–40 (2011).
10. Bashashati, A. *et al.* Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling: Evolutionary trajectories of ovarian cancers. *J. Pathol.* **231,** 21–34 (2013).
11. Almendro, V. *et al.* Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of genetic and phenotypic cellular diversity. *Cell Rep.* **6,** 514–527 (2014).
12. Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481,** 306–313 (2012).
13. Cheon, D.-J. & Orsulic, S. Mouse Models of Cancer. *Annu. Rev. Pathol. Mech. Dis.* **6,** 95–119 (2011).
14. Livet, J. *et al.* Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450,** 56–62 (2007).
15. Sheren, J., Langer, S. J. & Leinwand, L. A. A randomized library approach to identifying functional lox site domains for the Cre recombinase. *Nucleic Acids Res.* **35,** 5464–5473 (2007).
16. Schepers, A. G. *et al.* Lineage Tracing Reveals Lgr5+ Stem Cell Activity in Mouse Intestinal Adenomas. *Science* **337,** 730–735 (2012).
17. Cai, D., Cohen, K. B., Luo, T., Lichtman, J. W. & Sanes, J. R. Improved tools for the Brainbow toolbox. *Nat. Methods* **10,** 540–547 (2013).

18. Loulier, K. *et al.* Multiplex Cell and Lineage Tracking with Combinatorial Labels. *Neuron* **81,** 505–520 (2014).
19. Malide, D., Metais, J.-Y. & Dunbar, C. E. Dynamic clonal analysis of murine hematopoietic stem and progenitor cells marked by 5 fluorescent proteins using confocal and multiphoton microscopy. *Blood* **120,** e105–e116 (2012).
20. Weber, K. *et al.* RGB marking facilitates multicolor clonal cell tracking. *Nat. Med.* **17,** 504–509 (2011).
21. Johnson, F. H. *et al.* Quantum efficiency ofCypridina luminescence, with a note on that ofAequorea. *J. Cell. Comp. Physiol.* **60,** 85–103 (1962).
22. Zapata-Hommer, O. & Griesbeck, O. Efficiently folding and circularly permuted variants of the Sapphire mutant of GFP. *BMC Biotechnol.* **3,** 5 (2003).
23. Nguyen, A. W. & Daugherty, P. S. Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat. Biotechnol.* **23,** 355–360 (2005).
24. Tomosugi, W. *et al.* An ultramarine fluorescent protein with increased photostability and pH insensitivity. *Nat. Methods* **6,** 351–353 (2009).
25. Subach, F. V. *et al.* Monomeric fluorescent timers that change color from blue to red report on cellular trafficking. *Nat. Chem. Biol.* **5,** 118–126 (2009).
26. Shcherbo, D. *et al.* Far-red fluorescent tags for protein imaging in living tissues. *Biochem. J.* **418,** 567–574 (2009).
27. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat. Biotechnol.* **22,** 1567–1572 (2004).
28. Alwine, J. C. Transient gene expression control: effects of transfected DNA stability and trans-activation by viral early proteins. *Mol. Cell. Biol.* **5,** 1034–1042 (1985).
29. Sonawane, N. D., Szoka, F. C. & Verkman, A. S. Chloride accumulation and swelling in endosomes enhances DNA transfer by polyamine-DNA polyplexes. *J. Biol. Chem.* **278,** 44826–44831 (2003).
30. Schönhuber, N. *et al.* A next-generation dual-recombinase system for time- and host-specific targeting of pancreatic cancer. *Nat. Med.* **20,** 1340–1347 (2014).