UNIVERSITY OF ALBERTA

A novel transgenic mouse system to trace clonal stem cell lineages and estimate somatic mutation rates

by

Weonsang Simon Ro 

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Medical Sciences – Medical Genetics

Edmonton, Alberta

Spring 2006

Your file Votre référence
ISBN: 0-494-14031-3
Our file Notre référence
ISBN: 0-494-14031-3

Abstract

A novel transgenic mouse has been developed that allows clonal cell lineages to be labeled by random mutagenesis and traced in virtually any tissue. A green fluorescent cell lineage is generated by a random mutation at an enhanced green fluorescent protein (EGFP) gene containing a premature stop codon (referred to as the stop-EGFP gene) labeling individual mutants and ensuring clonality.

Using the stop-EGFP mouse, we successfully performed in vivo imaging of stem cell clonal lineages in the mouse dorsal epidermis and found cell lineages that descended from single epidermal stem cells were capable of generating multiple adjacent corneocytes, providing evidence for horizontal migration of epidermal cells between epidermal proliferative units (EPUs), in contrast to the classical EPU model. Repeated analyses of the same clonal cell lineages in the dorsal epidermis were performed over time to study the dynamics of the clonal development of particular cell lineages. The stop-EGFP system is expected to provide a novel tool for investigating clonal cell lineages of stem cells in various adult tissues as well as during development.

The stop-EGFP mouse system can also be applied to estimate rates of in vivo somatic mutation because the system allows individual mutations in a tissue, or an organ, to be identified by detecting green fluorescent colonies (i.e., clonal cell lineages descended from mutant cells) that were generated by unique mutations at the premature stop codon within the stop-EGFP gene. Using the stop-EGFP system, we estimated the rate of mutation induced by a potent mutagen, ENU (*N*-ethyl-*N*-nitrosourea) in the mouse liver and compared mutation rates among the different liver lobes. A time-course study of mutations in the liver of ENU-treated stop-EGFP mice revealed that the apparent

mutation frequencies change significantly depending on the time (following exposure) of sampling. An explicit statistical model was developed to better estimate the mutation rate induced by ENU in the liver. Comparison with vehicle-treated liver revealed a 3.4-fold increase in mutation frequency over control in ENU-treated liver, which is consistent with other previous reports and thus supports the utility of the stop-EGFP mouse as an in vivo mutation detection system.

Acknowledgements

There are a number of people without whom this thesis would not have been possible. First of all, I am indebted to my supervisor, Dr. Bruce Rannala for his continuous encouragement and intellectual contributions. Thank you for giving me your guidance when I needed it as well as allowing me the opportunity to pursue my own interests. Bruce has been a mentor who has shown me what type of scientist I want to become.

I'd like to thank members of my supervisory committee: Dr. Alan Underhill, Dr. Susan Andrew, and Dr. Linda Pilarski for their support and guidance throughout my studies. For technical support, I am grateful to Dr. Peter Dickie for the microinjection experiment, to Dr. Nick Nation for the mouse perfusion experiment and organ collection, to Dr. Xuejun Sun for skin cell imaging, to Ms. Gail Hipperson for ENU injection, and to the University of Alberta Health Sciences Laboratory Animal Services personnel for animal handling.

I am indebted to numerous people who have been kind enough to discuss with an inexperienced graduate student on a variety of subjects through emails. These include Dr. Soosan Ghazizadeh (discussion on live skin imaging), Dr. John Heddle (on mutagenicity experiments and tissue dynamics), Dr. Douglas Winton (on the *Dlb-1* system), Dr. Masaru Okabe (on "Green Mice"), and Mr. Andrew Salinger in the Dr. Monica Justice lab (on safety issues regarding ENU mutagenesis experiments). I am grateful for their kindness to respond to my never ending inquiries and their willingness to share their lab protocols when I requested. I dedicate this thesis to them.

Table of Contents

List of Tables

List of Figures

List of Symbols, Nomenclature, or Abbreviations

CI – Confidence Interval

DAPI – 4', 6-Diamidino-2-phenylindole

DLB-1 – *Dolichos biflorus*

EBFP – Enhanced Blue Fluorescent Protein

ECFP – Enhanced Cyan Fluorescent Protein

EGFP – Enhanced Green Fluorescent Protein

EYFP – Enhanced Yellow Fluorescent Protein

ENU – *N*-ethyl-*N*-nitrosourea

EPU – Epidermal Proliferative Unit

FISH – Fluorescence In Situ Hybridization

HPRT - Hypoxanthine-guanine Phosphoribosyl Transferase

LRC – Label Retaining Cells

MF – Mutation/Mutant Frequency

MLE – Maximum Likelihood Estimate

MMR – Mismatch Repair

NA – Numerical Aperture

PBS – phosphate buffered saline

PCR – Polymerase Chain Reaction

RT-PCR – Reverse Transcription-Polymerase Chain Reaction

SE – Standard Error

TA – Transit Amplifying

UV - Ultraviolet

Var - Variance

Chapter 1. Introduction

Parts of this chapter have been submitted.

Ro S. (2005) Tracing clonal lineages of stem cells in vivo: the stop-EGFP mouse. In: *Progress in Stem Cell Research* (Greer EV, ed.). New York: Nova Science Publishers (in press).

Overview

Understanding the types of daughter cells that individual precursor cells give rise to (and their distribution) is of great importance both in developmental biology and in stem cell biology. Recent studies suggest that stem cells in an adult tissue may have the capability to produce cells of even developmentally unrelated tissue types under experimentally modified conditions (Blau et al., 2001; Weissman et al., 2001). To unequivocally determine how diverse a range of cell types and how expansive a clonal cell lineage a single precursor cell (or stem cell) of a particular tissue can produce, a fate mapping technique that allows one to trace individual clonal cell lineages in a mammalian system in vivo is required. Although various mapping techniques have been developed so far, labeling of single cells is difficult to achieve and very few systems have been efficiently applied to fate mapping of single precursor cells (Clarke and Tickle, 1999; Flake, 2001).

To facilitate tracing of clonal cell lineages in a mammalian system in vivo, a novel transgenic mouse system was developed as part of my research (Chapter 2). The transgenic mouse (referred to as the stop-EGFP mouse) carries a premature stop codon-containing EGFP (enhanced green fluorescent protein) gene that acts as a target gene for mutations. Individual cells are labeled by means of random mutations and cell lineages descended from single cells that have undergone mutations express EGFP allowing clonal cell lineages to be traced in vivo via fluorescence imaging. To test the applicability of the system for investigating stem cell clonal lineages, we apply the system to trace stem cell lineages in the mouse dorsal epidermis (Chapter 3).

Another question of fundamental importance is the rate of in vivo somatic mutation in mammals. Somatic mutations play an important role in the initiation and

progression of cancer, as well as being an important factor of aging. Despite this importance, little is currently known about in vivo rates of mutation, especially in non germ-line (somatic) cells. Novel systems are therefore needed that allow one to study in vivo mutation rates in a mammalian system. Only a few mammalian systems have been developed for investigating processes of somatic mutation in vivo. These include the *lacI* and *lacZ* transgenic mouse systems (Gossen et al., 1989; Kohler et al., 1991) and the *hprt* system (Jones et al., 1985), as well as several others (Winton et al., 1988). A shortcoming of these systems is that they can only provide mutant frequency (i.e., the number of mutant cells divided by total cells counted) from which mutation rate often cannot be estimated accurately (Heddle, 1999; Thompson et al., 1998). To precisely determine the mutagenic effect of a chemical mutagen, a defective gene, or some other factor, a novel mutation detection system directed at estimating rates of mutation (rather than mutant frequency) is required.

The stop-EGFP mouse system developed in my research has considerable potential for use in estimating rates of in vivo somatic mutation in a mammalian system because the system allows individual mutations in a tissue or an organ to be identified by detecting green fluorescent colonies (i.e., clonal cell lineages descended from mutant cells) that were generated by unique mutations at the premature stop codon in the stop-EGFP gene. The ability to identify independent mutations should facilitate accurate estimation of mutation rates, which is difficult, or impossible, to achieve using other in vivo mutation detection systems. The final component of this research examines the feasibility of using the stop-EGFP mouse to estimate rates of somatic mutation in vivo (Chapter 4).

I. Green Fluorescent Protein (GFP) and Its Applications

1. GFP and its variants

GFP is a naturally fluorescent protein of 27-kDa size that emits green light of 508 nm when excited at 488 nm. The formation of its chromophore is autocatalytic, requiring no cofactors (Tsien, 1998). Its strong visible fluorescence without any cofactor or substrate allows direct detection of even single cells in live tissues (Chalfie et al., 1994; Yu et al., 2003). The protein is known to be non-cytotoxic and its expression does not perturb normal cellular processes (Chan et al., 2001; Feng et al., 2000; Okabe et al., 1997). As a result GFP has been widely applied to a variety of problems, ranging from cell lineage mapping to cellular trafficking of proteins.

GFP was originally isolated from the jellyfish *Aequorea victoria*, which will be referred to as wild-type GFP throughout the thesis. The natural form of the protein has several properties that can be undesirable for scientific purposes. First, the protein is excited at 395 nm, which falls within the range of ultraviolet (UV) light, and the imaging process can therefore result in UV-related toxicity in live cells (Basu-Modak and Tyrrell, 1993; Tyrrell, 2000). Second, the fluorescent intensity of wild-type GFP is very low (Cormack et al., 1996), which limits applications of the protein as it can be difficult to detect in some circumstances. Third, the protein exhibits poor expression in certain mammalian and plant species (Chiu et al, 1996; Yang et al., 1996). When transgenic mice were generated in which wild-type GFP was expressed under the control of a ubiquitous promoter, many tissues did not express the protein (Ikawa et al., 1998).

To increase the versatility of GFP, many variant forms of the protein have been generated by altering one or more amino acids (Cormack et al., 1996; Cubitt et al., 1995).

Most importantly, substituting phenylalanine for leucine at codon 64 and serine for threonine at codon 65 greatly improves fluorescence intensity (35 times brighter fluorescence relative to wild-type GFP) when excited at 488 nm in studies using equalized amounts of protein (Cormack et al., 1996). In addition, expression in mammalian systems has been improved by introducing 190 silent base changes to the coding sequence of the gene to systematically replace the native codons of the wild-type GFP with codons preferentially found in highly expressed human genes (Haas et al., 1996). The combination of the two codon substitutions described above and the 190 silent base substitutions leads to substantially brighter fluorescence and higher expression levels in mammalian cells (Yang et al., 1996), and the resulting protein (referred to as enhanced green fluorescent protein, "EGFP") has been widely used in a variety of contexts. In transgenic mice carrying the EGFP gene under the control of a ubiquitous promoter, all the organs exhibited strong green fluorescence (Ikawa et al., 1998; Okabe et al, 1997).

The tyrosine residue at codon 66 of GFP (equivalent to codon 67 of EGFP because of insertion of an additional codon right after the initiation codon, see Figure 2-6) is the core of the fluorophore and predominantly determines wavelengths of light for both excitation and emission (Heim and Tsien, 1996). For example, substitution of the tyrosine residue in EGFP to histidine results in major shifts of excitation and emission peaks to shorter wavelengths (380 and 440 nm, respectively) (Heim et al., 1994). The resulting protein is referred to as enhanced blue fluorescent protein (EBFP). Similarly, substitution of tyrosine at codon 67 of EGFP (or at codon 66 of GFP) to tryptophan shifts the emission peak to 475 nm and leads to cyan fluorescence emission from the protein

(Heim and Tsien, 1996). In the proteins with altered spectrums, other substitutions are usually included to increase the intensity of fluorescence. The EGFP variants and their characteristics are listed in Table 1-1 (Hadjantonakis and Nagy, 2001; Patterson et al., 2001; Tsien, 1998).

Table 1-1. Fluorescent proteins and their characteristics.

Fluorescent protein	Excitation peak (nm)	Emission peak (nm)	Relative intensity of fluorescence*	Relative bleaching time
EBFP	380	440	1	1
ECFP	433	475	28	28
EGFP	488	507	35	33
EYFP	513	527	35	12

* Information on relative intensity of fluorescence can be found at the website of Clontech, http://www.clontech.com/clontech/techinfo/faqs/LivingColors/variants.shtml. This Table was modified from Patterson et al (2001) and Tsien (1998).

2. Application of EGFP

Cell lineage studies rely on a biochemical or genetic marker to label one or more cells and identify their descendent lineages. For a marker to be suitable for use in cell lineage studies, its minimal properties should be that it is stable over several generations of cell division, that it not be secreted or transferred to other cells, and that it not perturb in any way normal cellular processes. Ideally, a marker would also allow cell fate to be

followed over some period of time in living tissues. In these regards, EGFP satisfies all the requirements for a versatile lineage marker (Zernicka-Goetz, 1999) (Table 1-2).

Table 1-2. Advantages of EGFP as a cell lineage marker.

Characteristics of EGFP	Advantages
Can be expressed in various organisms	Permanent genetic marking
Non-toxic protein	Does not perturb normal cellular processes
Very bright signal	Imaging at the single-cell level
No substrates or cofactors required	Live imaging of animal or tissue (avoids possible artifacts that might be introduced during fixation process)
Non-invasive imaging performed	Time-lapse study or repeated analyses

A cell carrying the gene encoding functional EGFP will transmit the gene to its progeny, thus the cell lineage that originated from the cell will express EGFP and be easily detected in vivo by fluorescence imaging. The first aim of my research was to develop a general method to trace clonal cell lineages using EGFP as an in vivo clonal marker (Chapter 2) and to apply the method to identify clonal lineages of stem cells in the mouse dorsal epidermis (Chapter 3) (see Figure 1-6).

In vivo mutation detection systems in which a bacterial transgene such as *lacI* or *lacZ* is inserted into the mouse genome have been widely used for mutation studies (Gossen et al., 1989; Kohler et al., 1991). One disadvantage of such systems is that the

target genes for mutations are prokaryotic and the systems therefore may not represent the characteristics of mutations one would expect to see in mammalian genes (Skopek, 1998). The EGFP gene has been genetically modified to enhance the mammalian characteristics of the gene (Yang et al., 1996). Thus, if the EGFP gene is used as a mutational target, it can better reveal mutational characteristics of typical mammalian genes. Furthermore, an EGFP-based mutation detection system could allow direct detection of independent mutations by imaging green fluorescent mutant colonies in vivo or in situ. Due to the strong fluorescence from the protein and recent advances in fluorescence imaging techniques, direct detection of single-cell mutants in the system is feasible (Chalfie et al., 1994; Yu et al., 2003), which also enhances the efficiency of the mutation detection system. The second aim of my research (Chapter 4) has been to develop a novel in vivo mutation detection system using EGFP as a mutational reporter that allows independent mutations to be directly detected in tissues of interest.

II. Stem Cell Biology and Fate Mapping

1. Stem cells, transit amplifying (TA) cells, and differentiated cells

Stem cells are undifferentiated cells with the capacity for prolonged self-renewal and are capable of producing differentiated cells of various cell types needed for periodic tissue renewal and tissue regeneration after injury (Loeffler and Potten, 1997; Watt and Hogan, 2000; Weissman, 2000). The pathway to differentiation from stem cells usually involves producing intermediate cells, called transit amplifying (TA) cells, that can rapidly divide and subsequently differentiate into mature cells of specific types (Alison et al., 2002; Loeffler and Potten, 1997) (Figure 1-1). Thus, production of TA cells facilitates the

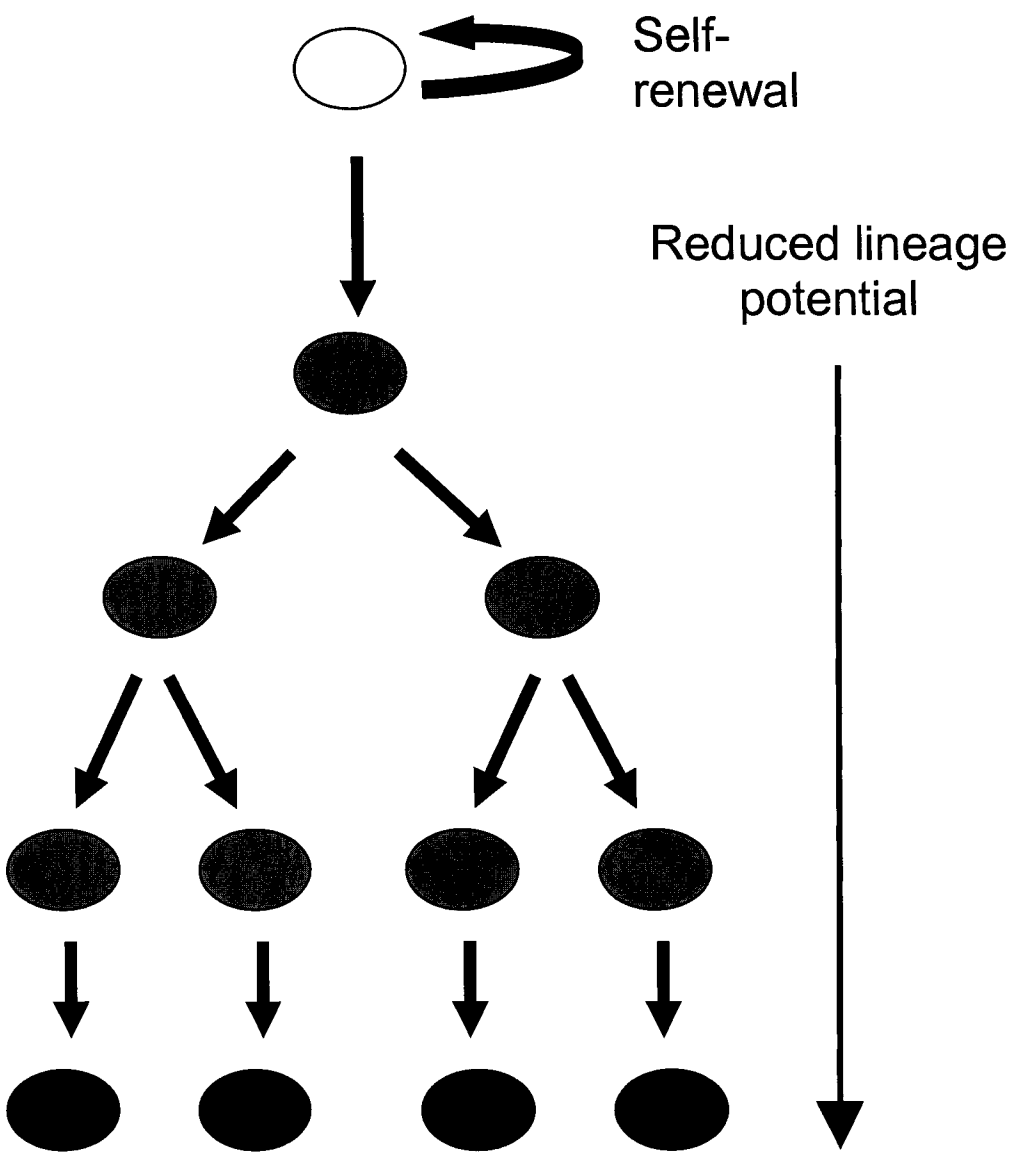


Figure 1-1. Pathway to differentiation from a stem cell.

A stem cell (open circle) can either self-renew or commit to a pathway leading to differentiation. The pathway to differentiation involves producing intermediate TA cells (gray circles), which divide and subsequently differentiate into specific types of mature cells (black circles). As TA cells divide, their lineage potentials are presumed to decrease. Adapted from Alison et al. (2002).

process of tissue renewal or regeneration, which requires production of new cells, as well as commitment of cells to specific cell lineages (Watt and Hogan, 2000). Because stem cells are the only cells capable of continuous tissue renewal, the population of stem cells must be maintained. It is still not known how stem cells maintain their numbers. Two competing models have been proposed (Loeffler and Potten, 1997; Ro and Rannala, 2001) (Figure 1-2). The deterministic model proposes that a small number of stem cells reside in a niche, each generating exactly one stem cell and one TA cell at each (asymmetrical) cell division. The daughter TA cell leaves the niche to proliferate for tissue renewal while the daughter stem cell remains in the niche; each stem cell is 'immortal' under this model. The stochastic model proposes that many stem cells exist in a niche with each stem cell division producing either two, one, or zero stem cells (and either zero, one, or two proliferating TA cells, respectively). This leads to ''drift'' in the numbers of descendents of each stem cell lineage over time. Determining whether a daughter cell of a stem cell remains a stem cell or becomes a TA cell might depend on asymmetrical segregation of cell-fate determinants or differential influences from the microenvironment (Spradling et al., 2001; Watt and Hogan, 2000).

Compared to stem cells, TA cells possess limited proliferative capacity and restricted potential for lineage commitment (Watt and Hogan, 2000). Within the population of TA cells, there could be a hierarchy of lineage potential (Alison et al., 2002). For example, early TA cells are likely more lineage-flexible than late TA cells, which have undergone several proliferation cycles (Figure 1-1). After undergoing several cell divisions, TA cells differentiate into mature cells that have functional specializations in tissues, but have no proliferative capacity.

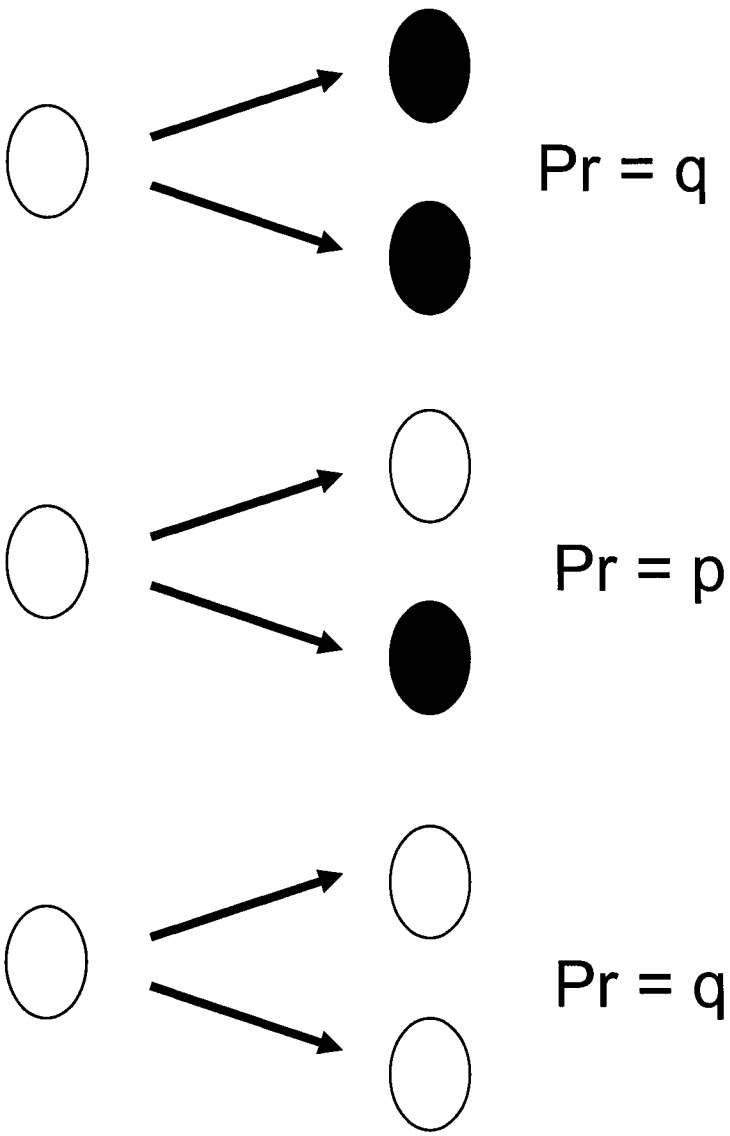


Figure 1-2. Deterministic and stochastic models of stem cell population dynamics.

Open circles represent stem cells and filled circles represent TA cells. The probability (Pr) that either two stem cells (and no TA cells), or two TA cells (and no stem cells) result from the division of a stem cell is q. The probability that one stem cell and one TA cell result is p. The deterministic model assumes that $q = 0$ and $p = 1$ and the stochastic model assumes that $p < 1$ and $q > 0$.

2. Identification and location of stem cells in adult tissues

In most tissues and organs, the identity of the stem cells in adult tissues has remained elusive and a hallmark of stem cells is the lack of expression of a specific stem cell marker (Blau et al., 2001; Flake, 2001; Robert, 2004). Even in hematopoietic cells where the stem cell population has been best characterized, isolation of stem cells using putative hematopoietic stem cell markers results in a heterogeneous cell population in which some of the cells fail to demonstrate the capacity for prolonged self-renewal and/or reconstitution of all cells of the blood (Morrison and Weissman, 1994). Furthermore, stem cells generally cannot be identified purely by morphology or location (Flake, 2001). Due to the lack of specific stem cell markers, alternative methods based on characteristics of stem cells have been proposed to identify the stem cell population. One such method is to generate label-retaining cells (LRC) (Bickenbach, 1981). Mice are repeatedly treated with a nucleotide analogue such as tritiated thymidine (^{3}H-TdR) or 5-bromo-2'-deoxyuridine (BrdU) when the tissue is in a hyperproliferative state such that all the dividing cells (i.e., both stem cells and TA cells) incorporate the label in newly synthesized DNA. The labeled cells are analyzed after several weeks or months during which the label in rapidly proliferating cells will dilute out (Bickenbach, 1981; Braun and Watt, 2004). Stem cells are believed to divide at a much slower rate than TA cells (Morrison et al., 1997; Robert, 2004) and the label in TA cells is therefore expected to dilute out during the waiting time between the exposure and monitoring of labeled cells while slowly dividing stem cells retain the label.

Application of this labeling technique to the mouse skin led to the finding of putative stem cells in the tissue where LRC were mostly restricted to the basal layer in

the epidermis and the bulge of the hair follicle (Cotsarelis et al., 1990; Mackenzie and Bickenbach, 1985; Morris et al., 1985; Morris and Potten, 1999). These LRC in the skin exhibited several stem cell attributes such as high colony-forming efficiency in vitro and the capacity to give rise to cells of various types present in the skin in vivo (Morris and Potten, 1994; Taylor et al., 2000). One problem with the method is that not all label-retaining cells will be stem cells (Braun and Watt, 2004; Potten, 2004). Some cells might not divide, for example, due to the cytotoxic damage by ^{3}H-TdR or BrdU incorporation. Furthermore, in tissues with a very slow turnover rate (e.g., liver cells), non-stem cells may also retain the label.

Putative stem cells have been identified in many adult tissues, and their locations have also been approximately described (Table 1-3). In tissues with a high turnover rate in which stem cells have been well studied, the location of stem cells is better understood (Niemann and Watt, 2002; Potten and Loeffler, 1990). In the epidermis, for example, stem cells are presumed to exist in the basal layer and basal TA cells generated from a stem cell are thought to move directly upward to terminally differentiate into corneocytes (Potten, 2004). Kinetic studies have shown that proliferation of epidermal cells is restricted to the basal layer suggesting that stem cells divide laterally with the plane of the basement membrane of extracellular matrix (i.e., along the basal layer) although a recent study by the Fuchs' group suggests that epidermal stem cells can divide perpendicularly to the membrane through asymmetric cell divisions (Lechler and Fuchs, 2005). About 10% of basal cells in the epidermis are presumed to be stem cells (Potten and Morris, 1988).

Table 1-3. Putative stem cells in adult tissues and their stem cell niches.

Tissue (organ)	Identification and/or location of putative stem cells
Epidermis	Basal layer
Small intestine	Crypt base
Large intestine	Crypt base
Liver	Terminal bile ductules (canals of Hering)
Lung	Tracheal basal cells, Clara cells, alveolar type II pneumocytes
Skeletal muscle	Satellite cells in muscle fibers
Hematopoietic tissue	Hematopoietic stem cells in bone marrow
Neural tissue	Subventricular zone of the central nervous system
Pancreas	Duct cells
Prostate	Basal cell layer

Adapted from Korbling and Estrov (2003) and Otto (2002).

Several putative markers of epidermal stem cells have been identified (Barthel and Aberdam, 2005; Morasso and Tomic-Canic, 2005). For example, epidermal stem cells have been shown to express higher levels of α6 and β_1 integrins and lower levels of the transferrin receptor CD71 when compared with TA cells (Jones and Watt, 1993; Li et al., 1998; Tani et al., 2000). The stem cells are also thought to preferentially express the cytokeratins K15 and K19 as well as β-catenin (Lyle et al., 1998; Tumbar et al., 2004; Zhu and Watt, 1999). Interestingly, studies investigating the transcriptional profile of several lineages of adult and embryonic stem cells show that integrins α6 and β_1 are highly expressed in stem cells (Ramalho-Santos, et al., 2002). Integrins have important structural and regulatory roles in the epidermis and transduce signals from the

extracellular matrix to the interior of cells (Goldfinger et al., 1999; Mainiero et al., 1996). Thus, it is speculated that integrins are involved in regulating proliferation and differentiation of epidermal cells. Recently, it has been suggested that stem cells located in the bulge of the hair follicle have the potential to maintain the epidermis (Niemann and Watt, 2002). The bulge stem cell concept has gained strong support from double-labelling studies by Taylor et al. in which the fate of the progeny of stem cells in the bulge was traced (Taylor et al., 2000). This study showed that the bulge stem cells produce daughter cells that differentiate to form the epidermis, hair follicles and sebaceous glands in response to a penetrating wound.

3. The proliferative unit

In a constantly renewing tissue, a distinct compartment of cellular proliferation is thought to exist as a unit of tissue renewal, which is maintained by a single stem cell, or a few stem cells, located within. Stem cells within a proliferative compartment (also referred to as a "proliferative unit") are thought to be responsible for the replacement of differentiated cells within the unit (Slack, 2000; Terskikh et al., 2003). Supporting this idea, it has been shown that there is a distinct proliferative unit in the epithelium of the small intestine (Bjerknes and Cheng, 1999; Griffiths et al., 1988; Schmidt et al., 1988). Likewise, there is evidence for a proliferative unit in the mouse dorsal epidermis, which is composed of a single column of corneocytes plus epidermal cells directly below the column, and maintained by a single stem cell located in the basal layer of the epidermis (Allen and Potten, 1974; Mackenzie, 1997; Potten and Morris, 1988). In other tissues with a high turnover rate, it is also expected that a proliferative compartment may exist

which is maintained by one, or a few, stem cells present in the compartment (Slack, 2000).

In response to a stimulus such as wounding, however, stem cells could provide cells beyond a proliferative unit. Following skin wounding, for example, epidermal cells appear to migrate to wounded sites beyond their proliferative compartments (Martin, 1997). As well, when hyperproliferation in the mouse epidermis was induced by chronic irradiation with 280 – 320 nm ultraviolet (UV) rays, cell lineages found in the epidermis revealed clonal expansion of basal cells to neighboring proliferative units (up to about 100 adjacent epidermal proliferative units) along the basal layer (Zhang et al., 2001). In the epithelium of the small intestine, larger clonal cell lineages beyond the normal size of the proliferative unit have been detected following treatment with cytotoxic reagents (Winton, 1997).

The precise definition of a proliferative unit in many tissues remains unresolved and an objective of my research (Chapter 3) will be to attempt to identify the size of the proliferative unit in the mouse epithelium (see Figure 1-6).

4. Lineage potential of stem cells

It is a subject of great interest in stem cell biology to determine how many different types of cells a single stem cell in a particular tissue can give rise to. In many cases, stem cells in adult tissues are multipotent in that they can produce cells belonging to a variety of cell lineages. For example, in the epithelium of the small intestine, there are four classes of mature differentiated cells and a single stem cell located in the crypt base can produce cells of all four types (Bjerknes and Cheng, 1999; Cheng and Leblond, 1974). Likewise, stem cells located in the bulge region of the hair follicle can produce not only cells of all

the hair follicle lineages, but also sebocytes and epidermal cells in response to a penetrating wound (Oshima et al., 2001; Taylor et al., 2000). However, some stem cells, such as basal stem cells of the epidermis, are unipotent so that they appear to only give rise to epidermal cells.

It has long been thought that adult mammalian stem cells are restricted to producing the cell types normally found in the tissue in which they reside. However, recent studies imply that stem cells may have the capability to produce cells of developmentally unrelated tissue types in experimentally modified conditions (Blau et al., 2001; Tosh and Slack, 2002; Weissman et al., 2001). Various studies have shown evidence for flexibility of lineage determination of adult stem cells. Transplanted bone marrow-derived stem cells can give rise to hepatocytes (Petersen et al., 1999), endothelial and myocardial cells (Lin et al., 2000), neuronal cells (Brazelton et al., 2000), and skeletal muscle cells (Ferrari et al., 1998) in adult mice. Hematopoietic stem cells appear to show a similar degree of lineage potential flexibility, such that they can give rise to hepatocytes, cardiac myocytes, and epithelial cells of the liver, gut, lung, and skin (Jackson et al., 2001; Krause et al., 2001; Lagasse et al., 2000). Furthermore, neuronal stem cells from adult mouse brain can produce blood cells, immune cells, and skeletal muscle cells (Bjornson et al., 1999; Galli et al., 2000). It is especially remarkable that cells derived from one germ layer appear to have the potential to give rise to cell types that derive from another germ layer (e.g., mesoderm-derived bone marrow cells may give rise to endoderm-derived hepatocytes).

The validity of stem cell plasticity, however, has been recently questioned (Anderson et al., 2001; Raff, 2003; Wagers and Weissman, 2004). It has been suggested

that cell-cell fusion could explain apparent lineage switching of transplanted stem cells (Raff, 2003; Wagers and Weissman, 2004). In most reported cases, claims that transplanted cells give rise to cells of a different lineage are based on detecting a particular marker of the new cell type, while retaining a genetic marker of the transplanted cells (Anderson et al., 2001). Fusion of transplanted cells and host cells of other tissue types will allow expression of both markers, which can lead to erroneous claims of lineage switching. Two independent studies have now shown that embryonic stem cells can spontaneously fuse with neuronal stem cells or bone marrow cells in culture (Terada et al., 2002; Ying et al., 2002). Confirming the in vitro studies, in vivo studies have also shown that the generation of hepatocytes from transplanted bone marrow cells was actually due to cell fusion processes rather than the differentiation of transplanted cells into hepatocytes (Vassilopoulos et al., 2003; Wang et al., 2003). Similarly, cell fusion has been implicated as the underlying mechanism for contributions of transplanted hematopoietic stem cells to injured skeletal muscle (Camargo et al., 2003).

Another problem with reports claiming the plasticity of stem cells is that a heterogeneous population of donor cells was used in most studies, thus contamination of stem cells of diverse tissue origins is possible (Wagers and Weissman, 2004; Weissman et al., 2001). For example, initial reports that skeletal muscle stem cells could produce blood cells (Jackson et al., 1999) were invalidated because it turned out that the blood cells were actually produced from hematopoietic stem cells present in the muscle, rather than skeletal muscle stem cells (McKinney-Freeman et al., 2002). Therefore, to demonstrate plasticity of stem cells, it is essential to exclude the possibility that multiple

stem cells could be contributing to the observed outcome and to use purified (ideally single) stem cells (Wagers and Weissman, 2004).

To unequivocally determine the lineage potentials of stem cells, a mapping technique is required that allows cell lineages of single stem cells to be traced in vivo. The stop-EGFP mouse system we have developed allows tracing of cell lineages that originated from single progenitor cells, thus it is expected that the system could be efficiently applied in future to address such issues in stem cell biology.

5. Embryonic stem cells and clinical potential of stem cells

At the blastocyst stage, the embryo forms an inner cell mass (ICM), which is capable of building up all three primary germ layers, the endoderm, mesoderm, and ectoderm (Gerecht-Nir and Itskovitz-Eldor, 2004; Wobus and Boheler, 2005). In the early 1980s, the mouse ICM was successfully isolated and propagated in an embryonic state under specific culture conditions, generating an embryonic stem (ES) cell line (Evans and Kaufman, 1981; Martin, 1981). ES cells are characterized by nearly unlimited self-renewal and have the developmental capacity to differentiate into cells of all somatic cell lineages in vitro as well as into germ cells. Significant efforts have been made to isolate ES cells from other species including rabbits (Graves and Moreadith, 1993), pigs (Li et al., 2003), primates (Thomson et al., 1995), and humans (Amit et al., 2000; Thomson et al., 1998).

Embryonic stem cells have a huge clinical potential as they hold the capacity to provide every type of cell and tissue in the body. A growing number of studies show that, upon transplantation, ES cell derivatives may restore the physiological function of injured organs (Brustle et al., 1999; Svendsen and Smith, 1999). In animal models of

human diseases, ES cell derivatives have shown a promising therapeutic potential (Wobus and Boheler, 2005). For example, a highly enriched population of midbrain neural stem cells derived from mouse ES cells has been found to promote partial recovery in a rat model of Parkinson disease (Kim et al., 2002). Similarly, transplantation of early differentiated cells derived from mouse ES cells significantly improved cardiac function in postinfarcted hearts (Yang et al., 2002).

Adult stem cells could also be used for treating various degenerative diseases. Hematopoietic stem cells (HSCs) have the capability to generate cells of all lineages in the hematopoietic system and have been used to treat patients with erythrocyte disorders such as sickle cell anemia and β-thalassemia (Gaziev and Lucarelli, 2003). Considering ethical issues regarding embryonic stem cell research, use of adult stem cells for therapeutic purposes could be a good alternative (Rizvi and Wong, 2005). If adult stem cells proved capable of producing cells of developmentally unrelated tissue types (see section II-4), the potential of adult stem cells for treating various degenerative diseases such as Parkinson disease would be very promising.

6. Techniques for mapping cell fates

As highlighted by the stem cell plasticity studies (section II-4), it is of great importance to determine the range of types of daughter cells that a single stem cell can give rise to in a mammalian system in vivo. As well, there has been interest in determining how expansive a clonal cell lineage a single stem cell can produce (e.g., in revealing the potential of single stem cells to provide cells to replenish other cellular compartments or proliferative units, for example). To address these issues, a fate mapping technique is required that allows one to trace clonal cell lineages in vivo. Although many different

mapping techniques have been developed, accurate labeling of single cells has proven very difficult to achieve and few systems have proven efficient for fate mapping of single precursor cells.

Labeling methods using tritiated thymidine (^{3}H-TdR) or bromodeoxyuridine (BrdU) simultaneously label large numbers of cells that are undergoing proliferation. Thus, it is unlikely that a cell lineage generated by these methods will have originated from a single progenitor cell. Another limitation is that signals from labeled cells are diluted out after multiple cell divisions (Flake, 2001; Stern and Fraser, 2001). Intracellular injection of tracers allows one to label single cells but it is hard to label small cells or cells located deep in tissue (Clarke and Tickle, 1999; Stern and Fraser, 2001). Furthermore, this method also suffers from the problem of signal dilution due to cell division. Mapping techniques using replication-deficient retroviruses that carry a marker gene can overcome the dilution problem (Sanes et al., 1986). However, the retrovirus-mediated methods can only label cells that are undergoing proliferation. As well, labeling of single cells is not easily achieved, even when using a very low titer of viruses, hindering clonal analysis of a cell lineage (Clarke and Tickle, 1999; Stern and Fraser, 2001). Laser-induced uncaging methods can allow labeling of single cells but the methods remain technically demanding and their utility as a cell fate mapping technique has not yet been widely demonstrated (Clarke and Tickle, 1999).

Transgenic mouse systems have been developed with the aim of allowing a clonal cell lineage to be traced in developmental studies. Bonnerot and Nicolas developed a transgenic mouse carrying a mutant *lacZ* gene (referred to as the *laacZ* gene) generated by placing a 289 base-pair internal duplication in the *lacZ* gene (Bonnerot and Nicolas,

1993). The duplication leads to premature termination of translation and thus the *laacZ* gene cannot produce functional protein. Spontaneous homologous recombination within the duplication recreates an open reading frame for beta-galactosidase and leads to *lacZ* expression in tissues. Thus, individual cells having undergone a revertant mutation and their descendants in the *laacZ* mouse can be identified histochemically. Similarly, DePrimo et al. modified the human placental alkaline phophatase (PLAP) gene by inserting seven repeated GC base pairs and created a transgenic mouse carrying the modified PLAP gene (DePrimo et al., 1998). Loss of one G residue from the poly (G) tract restores the open reading frame and allows expression of functional PLAP in cells which can be identified histochemically.

Given the importance of lineage potential of stem cells and cellular population dynamics to a complete understanding of stem cell biology, it is no surprise that cell fate mapping has become an intensive area of stem cell research. The ideal method for stem cell fate mapping would be a technique that allows one to label single stem cells in a tissue and subsequently trace their clonal cell lineages with minimal experimental perturbations. However, there are no molecular or phenotypic markers to identify stem cells, and thus it is virtually impossible to specifically label stem cells (Alison et al., 2002; Flake, 2001). Furthermore, even if there were a marker to identify stem cells, labeling single cells is very challenging (see above).

To trace stem cell lineages at the single-cell level, Winton et al developed a mutation-induced marker in mice heterozygous at the *Dlb-1* locus, which determines the expression of binding sites for the lectin *Dolichos biflorus* agglutinin (DBA) in the epithelium of the intestine (Winton et al., 1988). Mutation of the wild-type allele of a

Dlb-1 heterozygous mouse results in a ribbon of unstained cells on the villus (clonally expanded cells of a mutant) when the tissue is stained with DBA-peroxidase conjugate. Because random mutation allows individual cells to be labeled, tracing cell lineages that originated from single cells is possible in this system. In the study, mutations were induced using a mutagen, ENU (*N*-ethyl-*N*-nitrosourea) to randomly label epithelial cells in the small intestine, and the tissue was analyzed after several tissue turnovers to specifically detect clonal cell lineages that originated from stem cells. Because stem cells persist in constantly renewing tissues while other cells do not, cell lineages generated from mutant stem cells will remain following tissue renewal, while cell lineages that originated from non-stem cell mutants (induced by the ENU exposure) are eliminated.

The ingenious method developed by Winton et al. was successfully applied to trace stem cell clonal lineages and revealed that stem cells in the crypt base can give rise to cells of diverse cell types present in the epithelium of the small intestine (Bjerknes and Cheng, 1999). A major limitation of the *Dlb-1* system, however, is that it can only be applied to the epithelium of the small intestine. As suggested by the authors, the method can in principle be extended to other constantly renewing tissues for tracing stem cell clonal lineages if suitable markers are available for the tissues that can be obtained in a heterozygous state or can be generated in transgenic animals (Winton et al., 1988).

Applying a similar strategy of random mutagenesis for labeling individual cells, Bjerknes et al. demonstrated existence of functional multipotential stem cells in adult mouse gastric epithelium (Bjerknes and Cheng, 2002). They used transgenic mice that carry a single copy of the bacterial *lacZ* gene to label individual cells by chemical mutagenesis. Cells incurring inactivating mutations in the single *lacZ* allele will lose β-

galactosidase activity and will not stain for the enzyme. The mutated cells and their progeny that inherit the nonfunctional allele will be unstained and thus distinguishable from neighboring stained cells. Analysis for unstained cell lineages after allowing sufficient time for tissue renewal revealed stem cell clonal lineages in gastric epithelium. The system, however, has a few caveats. Adult transgenic mice carrying the bacterial *lacZ* gene frequently show variegation of transgene expression (Montoliu et al., 2000) making it hard to tell whether particular cells in an unstained stem cell lineage originated from mutated stem cells or whether they lost β-galactosidase activity simply due to transgene inactivation. Furthermore, identifying unstained cell lineages on the background of stained cells is technically challenging.

Methods using chemical mutagenesis of a transgene to label single cells for stem cell fate mapping potentially allow stem cell lineages to be traced at the single-cell level and are very promising. A transgenic mouse carrying a reporter gene (e.g., the gene encoding green fluorescent protein) that can be activated by chemical mutagenesis and easily detected should be preferred because detecting cell lineages that descended from mutated stem cells is expected to be much easier in this case.

III. Processes of DNA Mutation

1. Spontaneous mutation, mutation rate, and mutation frequency

Mutation refers to heritable changes in the genetic material (usually DNA or RNA), which arise either spontaneously or due to exposure to mutagenic agents. Spontaneous mutation results mainly from very infrequent errors occurring during DNA replication. The fidelity of DNA replication is estimated to be roughly in the range of one error per

10^{10} nucleotides synthesized in eukaryotic cells (Kunkel and Bebenek, 2000). The extremely high fidelity of DNA replication is maintained by three factors: DNA polymerases, exonucleolytic proofreading, and mismatch repair (Jiricny, 1998). DNA polymerases are highly precise enzymes, which can synthesize a double strand of DNA with an error rate of 10^{-5} per base pair (Kunkel, 1992). Should a non-complementary nucleotide be incorporated by mistake, the proofreading 3' → 5' exonuclease activity of DNA polymerases excises the terminal non-complementary nucleotide, which adds a further increase (roughly two orders of magnitude) to the fidelity of the DNA replication process (Kunkel, 1992). Mispaired bases that have escaped the exonucleolytic proofreading become substrates for mismatch repair (MMR) enzymes. MMR enzymes provide roughly two orders of magnitude increase in the fidelity of replication (Umar and Kunkel, 1996). To successfully restore the original DNA sequence, MMRs recognize base-base mismatches, and direct the process of correction of replication errors in the newly synthesized strand by a mechanism that discriminates between the parent and the daughter strand (Schofield and Hsieh, 2003).

Mutation rate can be expressed as the number of mutations per gene (or nucleotide) during a specified interval of time (e.g., number of cell divisions). Knowledge of mutation rate is often required to address underlying mutational mechanisms (Drake, 1970). In order to obtain unbiased estimates of mutation rates, information is required concering the number of independent mutational events at a target site or gene and the number of generations (or cell divisions) of a population of cells. Mutant frequency refers to the relative number of mutants in a population, which can be very different from the mutation rate especially in rapidly dividing tissues

(Thompson et al., 1998). The mutant frequency not only depends on the mutation rate, but also the number of generations of a cell population (Ono et al., 1995). Another characteristic of the mutant frequency is that it is dependent on the history of mutant cells. If a mutation occurs early in the population history, as opposed to later, the mutant cell will have more opportunity to divide and expand clonally (Heddle, 1999) (Figure 1-3). Thus even if underlying mutational mechanisms are the same (i.e., mutation rates are the same) in two populations of cells, mutant frequencies can appear to be very different due to the different histories of population dynamics.

There is currently little known about in vivo rates of mutation, especially in somatic tissues of mammals. Using our newly developed stop-EGFP transgenic mouse system in which independent mutations can be easily identified, it is attempted to estimate rates of mutation in somatic tissues in my mutation research (Chapter 4) (see Figure 1-6).

2. Induced mutation

In addition to spontaneous mutation, mutations can be induced by exposure to mutagenic agents such as ionizing irradiation or various chemicals. Ionizing radiation was the first environmental agent shown to be mutagenic (Muller, 1927). It is of high energy and can penetrate living tissues. In the process of tissue penetration, the high energy rays generate oxygen-derived free radicals (e.g., the hydroxyl radical) which cause various kinds of damage to DNA (Breimer, 1988). Although the molecular mechanisms of the mutagenic effects of ionizing radiation are not clearly understood, the radiation induces many forms of DNA damage including base alterations, DNA-protein cross-links, and single-strand or double-strand breaks (Breimer, 1988).

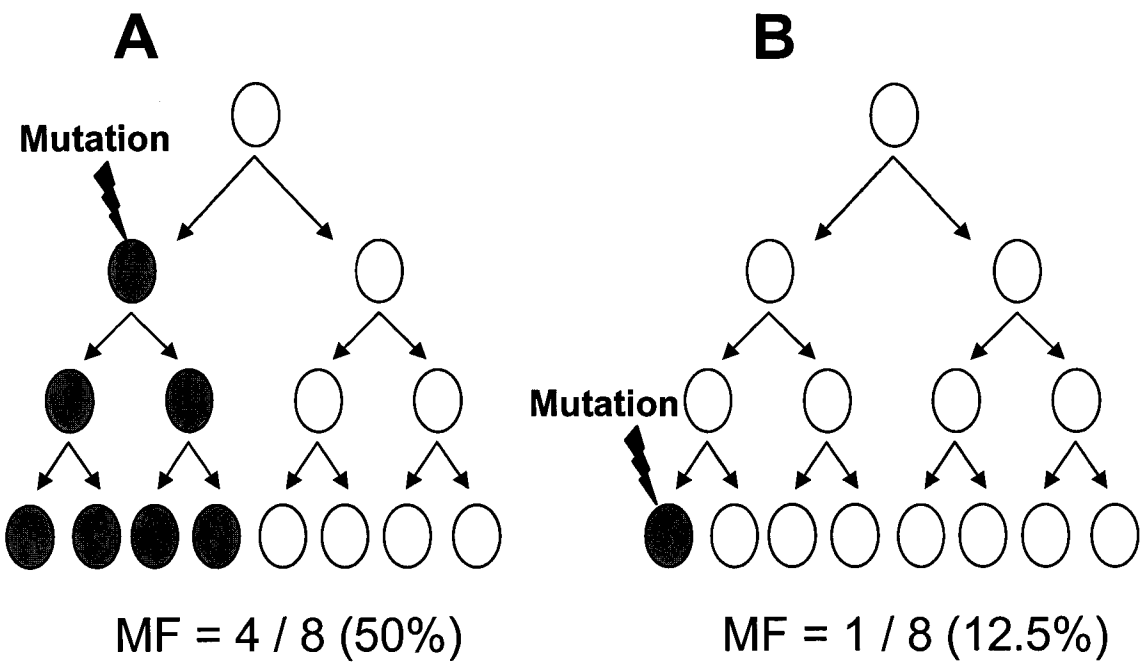


Figure 1-3. Clonal expansion of mutant cells in two exponentially growing cell populations.

A mutation occurring early in the population history results in a large number of mutant cells (shown in A) than a late mutation (shown in B). Thus, although the numbers of mutational events are the same (or mutation rates are the same) in two populations of cells, the resulting mutant frequencies can appear to be significantly different.
A filled circle represents a mutant cell. Arrows indicate cell divisions that produce two daughter cells. MF stands for mutant frequency.

Ultraviolet (UV) rays have lower energy and penetrate only the surface layer of the cell in animals. Although UV radiation does not induce ionization, UV is known to be a potent mutagenic agent. Two major classes of UV-induced DNA lesions are cyclobutane pyrimidine dimers, which are formed by cross-linkage of two adjacent pyrimidines on the same strand of DNA, and (6-4) photoproducts (Rycyna and Alderfer, 1985; Setlow and Carrier, 1966). These UV-induced DNA alterations lead to mutations when unrepaired or misrepaired. UV radiation can also produce DNA damage indirectly by creating reactive oxygen species (Ichihashi et al., 2003).

3. Chemical mutagens

In addition to mutagenic radiation, various chemicals are known to have mutagenic effects. The first chemical mutagen to be discovered was mustard gas, an alkylating agent, which transfers alkyl groups to the bases in DNA. (Auerbach and Robson, 1946). Chemical mutagens can be classified into two groups: those that are mutagenic to both replicating and nonreplicating DNA, such as the alkylating agents and those that are mutagenic only to replicating DNA. The latter group includes base analogs, which are purines and pyrimidines with structures similar to the normal bases of DNA (Litman and Pardee, 1956).

The mutagenic base analogs have structures sufficiently similar to the normal bases so that they are metabolized and incorporated into DNA during replication, but different enough to induce base mispairing, and thus mutation (Negishi et al., 1994). The two most common base analog mutagens are 5-bromouracil and 2-aminopurine (Goodman et al., 1985). 5-bromouracil can pair with both adenine and guanine, thus can induce transition mutations (i.e., AT→GC or GC→AT). As an analog of adenine, 2-

aminopurine pairs readily with thymine during DNA replication. However, it also pairs frequently with cytosine, thus leading to transition mutations such as CG→TA or TA→CG.

Alkylating agents such as ethyl methansulfonate (EMS) and *N*-ethyl-*N*-nitrosourea (ENU) cause DNA adducts by transferring methyl or ethyl groups to the bases (Sega, 1984; Shibuya and Morimoto, 1993). Although these transferred ethyl groups alone do not constitute a mutation, their presence can cause base mispairing during DNA replication, which subsequently leads to mutations (Figure 1-4). Alkylating agents exhibit less specificity in their mutagenicity than base analogs, and induce various types of mutations including transitions, transversions, and frameshift mutations (Sega, 1984). ENU is a laboratory-synthesized potent mutagen that efficiently causes random, single base pair mutations in a wide range of organisms and in a variety of tissues (Russell et al., 1979). ENU acts directly through alkylation of nucleic acids without any metabolic processing required for its activation (Singer and Dosanjh, 1990). The ethyl group of ENU can be transferred to oxygen or nitrogen radicals at a number of reactive sites such as the N1, N3, and N7 groups of adenine; the O6, N3, and N7 of guanine; the O2, O4, and N3 of thymine; and the O2 and N3 of cytosine (Balling, 2001). These transferred ethyl groups are DNA adducts which alone do not constitute a mutation; however, their presence can cause mistaken identity of the ethylated base during DNA replication, which results in mispairing. After DNA replication, a single base pair substitution exists that is unidentifiable to cellular repair systems. The mutagen can induce all possible substitution mutations, among which AT to TA transversion is the most common type of mutation found in the mouse germline (Noveroske et al., 2000).

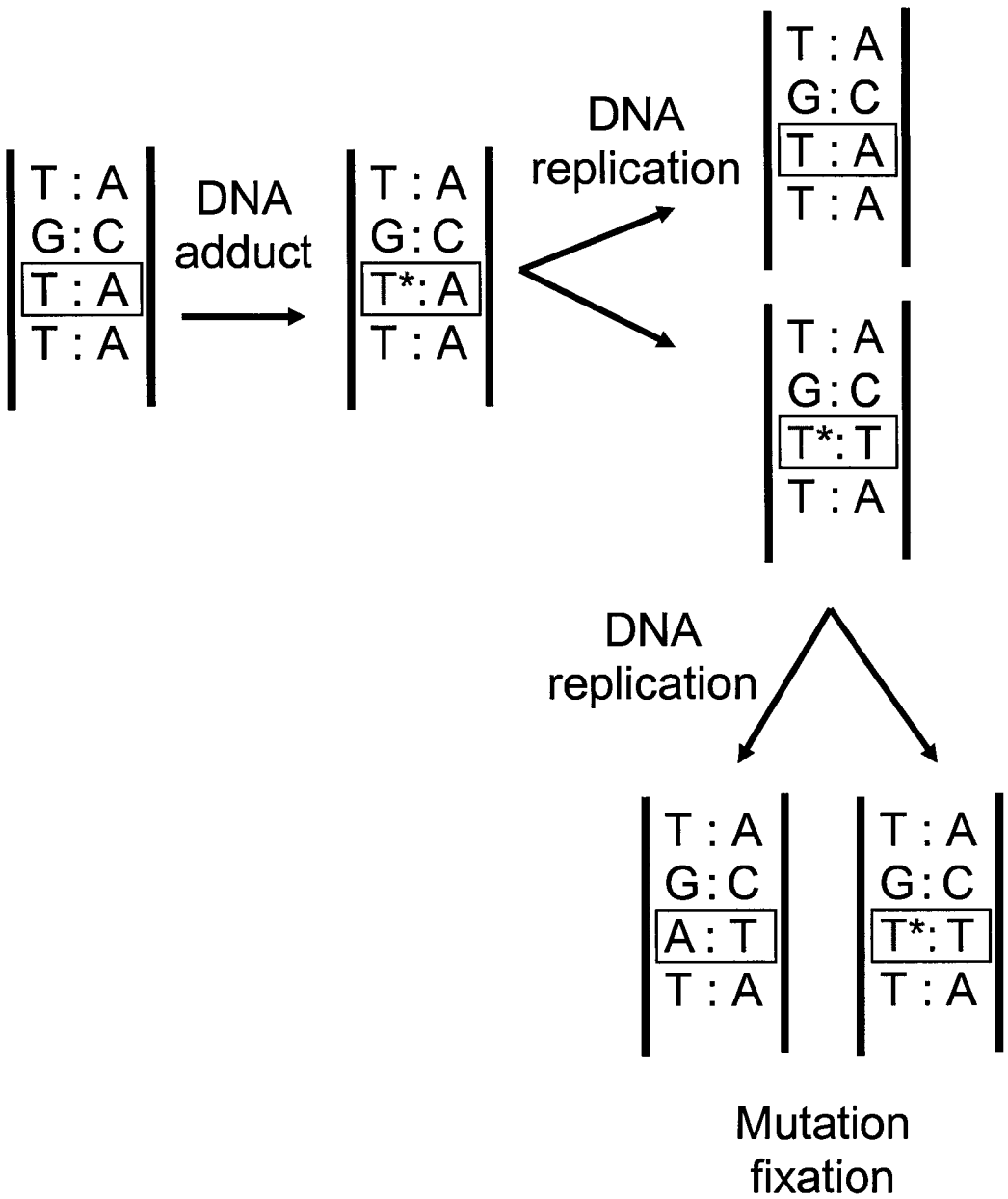


Figure 1-4. Fixation of mutation induced by an alkylating agent.

An alkylating agent causes DNA adducts by transferring methyl or ethyl groups to the bases. In this illustration a DNA adduct (indicated as an asterisk) on thymine (T) induces base mispairing with T during DNA replication. Another round of DNA replication induces base substitution to AT from original TA and the mutation is permanently fixed in DNA.

4. DNA repair systems

In response to diverse lesions that arise in DNA due to exposure to various types of mutagenic agents, specific DNA repair systems are activated to repair the DNA damages, and prevent them from being fixed as mutations. In the human, up to 130 genes have been identified that are associated with DNA repair (Wood et al., 2001). Because the genome is constantly subjected to a plethora of mutagens, diverse and systematic DNA repair systems are expected to exist to maintain the integrity of the genetic information.

In the case of helix-distorting lesions such as pyrimidine dimers induced by UV radiation, nucleotide-excision repair (NER) excises and replaces the damaged nucleotide residues (Christmann et al., 2003; Hoeijmakers, 2001). NER is achieved by dual, single-stranded incisions on either side of DNA adduct and removal of a fragment of about 30 nucleotides, followed by gap filling by DNA polymerase and ligation of the nicks by DNA ligase (Friedberg, 2001). DNA lesions induced by ionizing radiation or alkylating agents are subjected to base-excision repair (BER) whereby damaged bases are excised from the genome as free bases by a DNA glycosylase (Christmann et al., 2003; Lindahl and Wood, 1999). Both NER and BER are carried out by somewhat different mechanisms depending on whether DNA damage is located in transcriptionally active regions (transcription-coupled repair, TCR) or not (global genome repair, GGR) (Friedberg, 2003). TCR removes lesions from the transcribed strand of transcriptionally active genes where the stalled RNA polymerase is removed or displaced from the lesions preceding the repair (Svejstrup, 2002). On the contrary, GGR removes DNA lesions from the non-transcribed regions of the genome and the non-transcribed strand of transcribed regions.

The mismatch repair (MMR) system is responsible for removal of base mismatches caused by base deamination, oxidation, methylation, and replication errors (Schofield and Hsieh, 2003). The MMR machinery not only binds to base-base mismatches and insertion/deletion loops arising spontaneously during replication and homologous recombination, but also to various chemically induced DNA lesions such as alkylation-induced O^6-methylguanine paired with cytosine or thymine, UV-induced photoproducts, and 8-oxoguanine (Colussi et al., 2002; Duckett et al., 1996; Wang et al., 1999). Base mismatches or chemically modified bases are recognized by the MutSα complex which binds to the lesions and signals downstream excision steps (Schofield and Hsieh, 2003). Subsequently, the excision of the DNA strand containing the mispaired base is performed by exonuclease I and a new strand is synthesized by Polδ (Genschel et al., 2002; Longley et al., 1997).

In addition to the aforementioned DNA repair systems, which involve excision and replacement of damaged DNA, another form of repair systems exists which involves direct reversal of DNA damage (Friedberg et al., 2004). For example, alkylated lesions such as O^6-methylguanine can be repaired by O^6-methylguanine-DNA methyltransferase which directly removes the deleterious methyl group from the DNA guanine residue (Christmann et al., 2003).

DNA double-strand breaks (DSB) arise from ionizing radiation, X-rays or free radicals and are highly potent inducers of genotoxic effects (e.g., chromosomal breaks or exchanges) and cell death (Aylon and Kupiec, 2004). There are two main pathways for DSB repair, homologous recombination (HR) and non-homologous end-joining (NHEJ) (Kanaar et al., 1998). When, after replication, a second identical DNA copy is available,

HR seems to be preferred. Otherwise, cells rely on NHEJ which is more error-prone (Hoeijmakers, 2001). In the case of HR, a damaged chromosome enters into physical contact with an undamaged DNA molecule with which it shares sequence homology and which is used as a template to copy the missing information into the broken locus (Morrison and Takeda, 2000). HR starts with nucleolytic reaction of the DSB generating 3' single-stranded DNA fragments which then form two regions of heteroduplex DNA with intact homologous template (Christmann et al., 2003). Following DNA polymerization, ligation of newly synthesized DNA to the resected broken chromosome generates a joint molecule containing two Holliday junctions (Kanaar et al., 1998). The resulting structure is resolved according to the classical model of Holliday (Holliday, 1964). In contrast, NHEJ is a mechanism that is able to bring broken DNA ends together and rejoin them in the absence of long tracts of sequence homology (Hefferin and Tomkinson, 2005). NHEJ is initiated by binding of a Ku70 and Ku80 heterodimeric complex to DSB (Jeggo et al., 1992; Reeves and Sthoeger, 1989). This results in the recruitment of other proteins such as DNA protein kinase (DNA-PK), Artemis and DNA ligase IV which are involved in ligation of the broken ends (Hefferin and Tomkinson, 2005).

5. Mutations and diseases

Many diseases are caused by mutations, either in somatic cells (as in cancer), or in germ-line cells (as in inherited disorders). Mutations are also a key component in biological processes such as aging. Thus, an elevated mutation rate will increase the susceptibility to such diseases or conditions, especially cancer. The relationship between increased mutation rates and tumor susceptibility has a long history of study, showing that

increased mutations correlate with an increased incidence of tumor development (Burdette, 1955; Jackson and Loeb, 2001). It has been shown that exposure to chemical mutagens or ionizing radiation increases the incidence of tumor development (Bertram, 2000; Little, 2000) and that individuals with defective DNA repair genes are also more susceptible to tumor development (Liu et al., 1999; Papadopoulos and Lindblom, 1997).

The rate of mutation can be elevated in various ways. There are hundreds of genes, including genes coding for various kinds of DNA repair enzymes, which can influence the mutation rate (Loeb, 1991). Defects in those genes will result in an elevated mutation rate, which leads to the accumulation of mutations in the genome and finally to the development of cancer and various metabolic disorders. Exposure to chemical mutagens and ionizing radiation can also lead to elevating the mutation rate in the genome. Considering the important role that mutations play in the development of various diseases, especially cancer, it is of great importance to develop a novel mutation assay system that can precisely quantify changes in mutation rate in individuals induced by a chemical mutagen, a defective gene, or some other cause.

IV. Mutation Assay Systems

1. In vitro genotoxicity assays using bacterial systems

To uncover the important factors influencing rates of somatic mutation (and by extension, rates of cancer, etc) refined techniques are needed to estimate rates of somatic mutation in cells exposed to potential environmental, or genetic, risk factors. In vitro assays for evaluating the mutagenicity of a chemical substance are now well established, the most widely-used being the Ames *Salmonella* test (Ames et al., 1973). This in vitro

mutagenicity assay is a short-term bacterial reverse mutation assay specifically designed to detect mutagenic substances that can produce genetic damage that leads to gene mutations. The test uses *Salmonella* strains with preexisting mutations in various genes in the histidine operon, that leave the bacteria unable to grow and form colonies in the absence of histidine (Mortelmans and Zeiger, 2000). New mutations that can restore the gene's function allow the mutated cells to survive in the absence of the amino acid. By comparing the number of revertant colonies between non-treated control bacterial plates and those treated with various substances, the mutagenic effect of the substances can be measured. The Ames test is used world-wide as an initial screen to determine the mutagenic potential of substances because of advantages of time and cost effectiveness. Similarly, the *Escherichia coli* reverse mutation assay which detects trp+ revertant colonies from the bacteria with a defective biosynthesis of tryptophan, is also used as a substitute for, or as an addition to the Ames *Salmonella* test (Green and Muriel, 1976).

However, evidence has accumulated that in vitro genotoxicity results may not reflect in vivo exposures, or mutagenic mechanisms in vitro may be different from those that occur in vivo (MacGregor et al., 2000). A major limitation of prokaryotic mutagenicity assays is that metabolic enzymes present in mammalian cells that can transform a non-mutagenic into a mutagenic substance (genotoxicity) may not be present in the prokaryotic cell. Rodent liver homogenates (S-9) have been added to the Ames assay to mimic the mammalian environment, but the ability of the Ames test to identify known carcinogens (e.g., those substances known to cause cancer in mice) remains low (Gossen et al., 1994).

2. In vivo mutation assays

If in vivo mutation detection were found to be feasible, this could offer an efficient test for mutagenicity that is expected to be more sensitive than in vitro methods in evaluating potential mutagens. To this end, the mouse specific-locus test (Russell, 1951) was proposed to examine the rate of germ-line mutation at several loci influencing phenotype (coat color, etc) and this approach has been extended to many more loci using molecular markers. The method involves mating homozygous wild-type mice, treated or controls, to stock mice which are homozygously recessive at several marker loci. If a germ-line mutation at one of those loci in wild-type mice is transmitted to offspring that inherit the other recessive allele from stock mice, then the offspring will exhibit the mutant phenotype. Germ-line mutation frequency can then be estimated from the numbers of mutant versus the total offspring observed and total number of loci tested. A serious limitation of the specific-locus test is that enormous numbers of mice must be bred to detect even large changes in mutation rates. The total number of opportunities to detect mutations in this system is roughly proportional to twice the number of mice bred multiplied by the number of target loci examined.

We currently have limited knowledge of in vivo rates of mutation, especially in non germ-line cells (somatic cells) of mammals. Only a few mammalian systems have been developed for investigating in vivo mutagenicity in somatic tissues (see below) and these systems can only provide mutant frequency (number of mutant cells divided by total cells counted) (see section III-1). Theoretically, in an expanding cell population mutant frequency increases with each generation of cell division.

The *Dolichos biflorus* (*Dlb-1*) assay system (Tao et al., 1993; Winton et al.,

1988), the hypoxanthine-guanine phosphoribosyl transferase gene (*hprt*) system (Jones et al., 1985), and the *lacI* or *lacZ* transgenic mouse system (Gossen et al., 1989; Kohler et al., 1991) are the most widely used in vivo mutation assay systems. Although all three systems have been successfully used for studying the mutagenic effect of substances on somatic tissues in vivo, all the methods have serious disadvantages and limitations.

3. The *Dlb-1* system

The *Dlb-1* system quantifies somatic mutations in an endogeneous gene, *Dlb-1* in the epithelial cells of the small intestine (see section II-6). The frequencies of unstained mutant ribbons in villi are compared between controls and treated animals to test the mutagenicity of a substance (Tao et al., 1993). Using the system, the mutagenic effect of various agents has been efficiently tested in the small intestine including gamma radiation, ethylating agents, and base analogs (Cosentino and Heddle, 1999; Schmidt et al., 1990; Winton et al., 1989).

A problem with the mutation assay in the *Dlb-1* system is that small mutant colonies cannot be easily detected because it is very difficult to identify small unstained colonies on the background of stained cells. This may cause the mutant frequencies to be underestimated. Furthermore, the applicability of the *Dlb-1* system is greatly limited because it can only be applied to the epithelium of the small intestine (Reliene and Schiestl, 2003).

4. The *hprt* mutation system

The *hprt* system utilizes the endogenous hypoxanthine-guanine phosphoribosyl transferase gene (*hprt*) as a mutational target. The assay system is based on the cloning of *hprt*-negative T lymphocytes by growing them in 6-thioguanine-containing medium and

mutant frequency is calculated from counting the number of resistant colonies (Jones et al., 1985). The system has been efficiently applied for studying spontaneous and induced mutations in T lymphocyte populations in the thymus and spleen (Skopek et al., 1992; Walker et al., 1997). However, the period of time required to select the resistant colonies (7 – 10 days) is long. Another problem with the system is that a mutation in the gene might be deleterious, and thus the mutant cells might be selected against in the mouse, which could again lead to underestimates of the mutant frequency (Heddle, 1998).

5. The mouse spot test

An in vivo mutation system has also been developed for testing mutagenicity of a substance in a somatic tissue as an extension of the germ-line mouse specific locus test (Russell and Major, 1957). The method involves exposing embryos heterozygous for several coat-color genes to a potentially mutagenic agent slated for testing and then observing clones of mutant cells in adulthood. A somatic mutation in a dominant allele in a melanocyte precursor cell results in a brownish or gray color spot (a clonal lineage that originated from the mutant melanocyte precursor cell) in the black fur. The assay, however, has several shortcomings. First, color spots can arise due to other causes such as pigment cell death, misdifferentiation, or epigenetic changes (Russell, 1977; Russell and Major, 1957). Furthermore, the assay requires a large number of animals and can be applied to limited tissue types (Reliene and Schiestl, 2003).

6. The *lacZ* and *lacI* transgenic mouse mutation detection systems

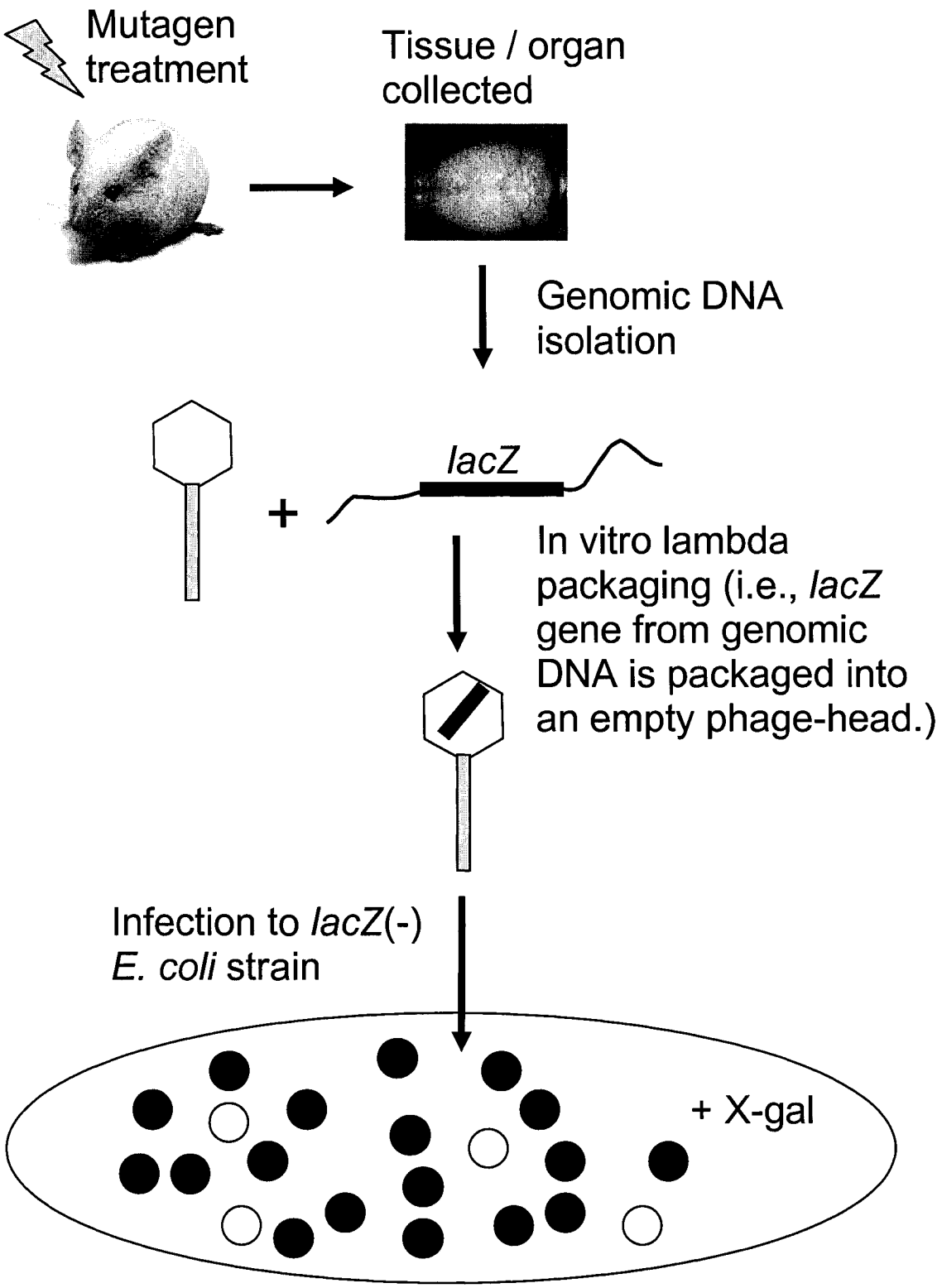
In the early 1990s, transgenic mouse models for in vivo mutation detection were developed by inserting either the *lacZ*, or *lacI*, bacterial transgene into mice (Gossen et al., 1989; Kohler et al., 1991). The advent of these mouse systems has created a new era

in the field of mutagenicity assays (Gossen et al., 1994; Nohmi et al., 2000). Since its development, the transgenic mutation assay system has been widely used for measuring mutations induced by various mutagenic agents in different tissues (Dean et al., 1999; Suzuki et al., 1999; Thybaud et al., 2003). The transgenic mouse has been commercialized as Muta^{TM}Mouse (Covance Inc., Princeton, NJ) or Big Blue® Mouse (Stratagene, La Jolla, CA) which employs the bacterial *lacZ* or *lacI* transgene, respectively as a target gene for mutations. In the mutation assay, genomic DNA of the transgenic mouse is extracted from the tissue of interest, and a reporter cassette containing the *lacZ*, or *lacI*, gene is isolated from the mouse genome and recovered as viable phage using the lambda phage packaging kit (Figure 1-5). Whether the target gene has mutated or not is determined by growing the phage on *E.coli* under suitable conditions (Dycaico et al., 1994; Nohmi et al., 2000). The most remarkable feature of the transgenic mouse mutation detection system is that any tissue can be used for the mutation assay, which allows tissue-specificity of mutagenic effects of a substance to be investigated (Suzuki et al., 1999). Another advantage of the transgenic system is that the transgenic loci in the system are genetically neutral, so that mutants are not eliminated by selective pressure in vivo (Cosentino and Heddle, 2000; MacGregor et al., 2000). Furthermore, identified mutants can be easily cloned and sequenced, providing mutation spectra which are very useful for elucidating mechanisms of mutational events (Dycaico et al., 1994; MacGregor et al., 2000).

The validity of the *lacZ*, or *lacI*, transgenic mouse system for studying somatic mutations in mammals has been questioned, however, because many features of the bacterial *lacZ* and *lacI* genes, the target genes for mutations, are typically prokaryotic

Figure 1-5. Schematic illustration of the *lacZ* transgenic mutation assay.

Tissue of interest is collected from the *lacZ* transgenic mouse treated with a mutagen. Genomic DNA is extracted from the tissue and exposed to an in vitro lambda phage packaging extract to isolate a reporter cassette containing the *lacZ* gene from the genomic DNA. The rescued phages containing the reporter gene are then plated on a *lacZ*(-) *E. coli* strain in the presence of X-gal (5-bromo-4-chloro-3-indoyl-β-D-Galactopyranoside) for selection of mutant and non-mutant *lacZ* phages. Infection of a phage containing mutant *lacZ* gene results in a colorless plaque (open circle) as opposed to a blue plaque (filled circle), which results from infection of a phage containing non-mutant *lacZ* gene. The frequency of white plaques corresponds to mutant frequency induced by the mutagen in the tissue of the mutagen-treated mouse. Other bacterial transgenic mutation systems (e.g., the *lacI* transgenic mouse system) employ similar procedures for mutation assay.



(codon usage, etc) and the systems therefore may not represent the characteristics of mutations one would expect to see in mammalian genes (Skopek, 1998). As well, the bacterial transgenes are untranscribed in the mouse and contain a high content of CpG dinucleotides which are methylated in mammalian cells to produce 5methylcytosine, a known promutagenic base (Grippo et al., 1968; Shen et al., 1994). To overcome the methylation problem, Skopek and colleagues developed a transgenic mouse carrying a modified bacterial *lacI* transgene that had a significantly reduced CpG content (Skopek et al., 1998). Furthermore, transgenic mice have been developed carrying a modified *lacI* gene that was created by re-encoding the amino acid sequence of *lacI* with mammalian codons, and is regulated by a mammalian promoter (Scrable and Stambrook, 1997). These modified bacterial transgenes have the potential to be efficiently used as an in vivo mutational target gene that is expected to better represent the mammalian characteristics of mutations than the original *lacI* gene.

7. Mutation assay systems using GFP as a reporter for mutation

Mutation detection systems using the GFP gene as a target for mutations have been developed (Cariello et al., 1998; Bachl and Olsson, 1999). A system that detects loss of the protein function due to a mutation in the GFP gene would be practically unreliable because identifying mutant (non-fluorescent) cells on the background of fluorescent cells would be very difficult. Thus, an effective GFP-based mutation detection system should rely on detection of a gain of protein function, as is induced, for example, by a revertant mutation that changes a gene encoding a non-functional protein to that encoding a functional fluorescent protein. This idea has been previously applied to a few mutation assay systems in which the function of GFP was removed by introducing a gene knock-

out mutation (e.g., a frameshift mutation, or a premature stop codon) within the coding region of the gene (Cariello et al., 1998; Bachl and Olsson, 1999). In those systems, a cell having undergone a revertant mutation that functionally eliminates a gene knock-out mutation expresses the functional GFP, and thus becomes green fluorescent. One obvious merit of such systems is that the mutation assays are very easy and time-efficient because detection of mutant cells can be directly performed using fluorescence imaging or fluorescence activated cell sorting (FACS) (Bachl and Olsson, 1999).

Cariello et al. introduced an additional nucleotide into the coding region of the GFP gene, which knocked out the gene due to a frameshift mutation (Cariello et al., 1998). Bacterial cells carrying the mutant GFP gene were treated with various mutagens that induce a frameshift mutation. The treatments yielded increased numbers of green fluorescent colonies which were generated by a revertant mutation that deleted a nucleotide within the coding region. Similarly, Bachl et al. developed a mammalian cell line mutation assay system in which the GFP reporter gene contains a premature stop codon (TAG) at codon 107 which renders the gene non-functional (Bachl and Olsson, 1999). They created the premature stop codon by substituting G for C at codon 107 (TAC) within the GFP gene. In the system, the functional GFP is only expressed when the stop codon reverts to TAC or TAT (both encoding a tyrosine) by transversion mutations. The mutation assay system was specifically designed to study the hypermutation mechanism in the immunoglobulin (Ig) gene in a mammalian cell line. To induce hypermutation at the premature stop codon within the mutant GFP gene, the Ig enhancers were included in the vector. Although the system has proven effective for the hypermutation study, its general application to mutational studies (such as genotoxicity

testing) has not been attempted thus far. In the system, only specific mutations at the premature stop codon can restore GFP function (i.e., a mutation from TAG to TAC or TAT, which encodes the original amino acid, tyrosine at codon 107). Thus, the system cannot detect all possible base substitutions such as a substitution from G to A.

Although a mutation assay system that employs detection of a revertant mutation in a mutant GFP gene facilitates the process of identifying mutations, the small number of mutational target sites reduces the efficiency of the system. With the aim of increasing the number of target sites for mutations that can induce expression of functional GFP, Dobrovolsky et al. developed a mutation detection system in which the GFP gene is placed under the control of a strong repressor (Dobrovolsky et al., 2002). The GFP is normally not expressed due to the repressor, but any mutation in the repressor gene that eliminates the repressor function can lead to expression of the GFP. Although this strategy appears elegant, the leakiness of the repressor leads to expression of the protein even in the absence of mutation, resulting in overlapping ranges of fluorescence intensity between non-mutant cells and mutant cells when analyzed using FACS. This makes it hard to distinguish mutant from non-mutant cells, and thus estimates of mutant frequency from this system are expected to be inaccurate.

V. Aims of Research (see Figure 1-6 for a flow chart)

Although various cell fate mapping techniques have been developed so far, labeling of single cells is rarely achieved. With the aim of tracing clonal cell lineages, I have developed a novel transgenic mouse (referred to as the stop-EGFP mouse) which carries a premature stop codon-containing enhanced green fluorescent protein (EGFP) gene as a

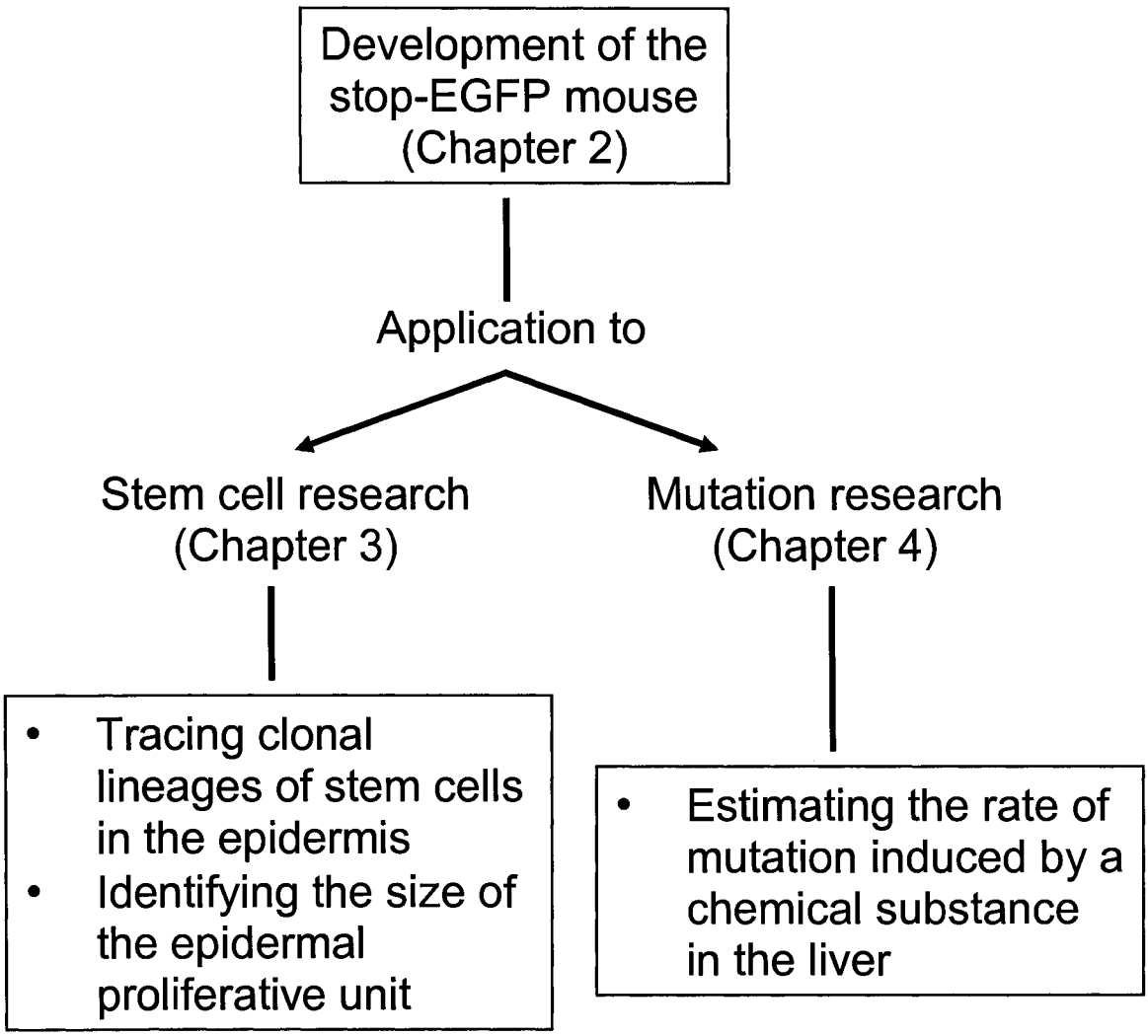


Figure 1-6. Diagram summarizing the aims of research.

The first aim of my research is to develop a novel transgenic mouse (referred to as "the stop-EGFP mouse") carrying a premature stop codon-containing enhanced green fluorescent protein (EGFP) gene (described in Chapter 2). The stop-EGFP mouse is applied to stem cell research (described in Chapter 3) and mutation research (described in Chapter 4). The objectives in stem cell research are to trace clonal cell lineages that originated from stem cells in the mouse dorsal epidermis and to identify the size of the epidermal proliferative unit that a single stem cell maintains. The objective in mutation research is to estimate the rate of mutation induced by a chemical substance in the liver and thus to evaluate the mutagenic effect of the substance.

target gene for mutations. In the stop-EGFP system, individual cells can be randomly labeled via infrequent mutations at the site of the premature stop codon in the stop-EGFP gene. A cell having undergone a mutation at the premature stop codon and its clonal descendants will express functional EGFP, thus a clonal cell lineage can be traced by following a green fluorescent colony. To test the efficiency of the stop-EGFP system for in vivo analysis of a clonal cell lineage, I have applied the system to the mouse dorsal epidermis and investigated clonal cell lineages in vivo that originated from single epidermal stem cells.

The stop-EGFP mouse can also be used to carry out in vivo mutation studies. There are several mouse mutation detection systems currently being widely used. Existing systems all have one or more undesirable properties, however. For example, some systems use an untranscribed bacterial gene as a mutational target and some provide only mutant frequency from which it is difficult to accurately estimate mutation rate. The stop-EGFP mouse system is expected to be an efficient mutation detection system because the target gene has characteristics of mammalian genes, and also because mutant colonies (i.e., independent mutations) can be directly observed in a tissue or organ using fluorescence imaging facilitating accurate estimation of mutation rates. To verify the applicability of the stop-EGFP system to mutation studies, I used the stop-EGFP mouse to estimate the rate of somatic mutations induced by a potent mutagen, ENU (*N*-ethyl-*N*-nitrosourea). The numbers of independent mutational events (i.e., the number of mutant colonies) were counted and used to estimate the rate of mutation induced by the mutagen in the liver.

The objective of this thesis is to provide evidence to support the utility of the stop-EGFP system for tracing clonal lineages of stem cells and for estimating somatic mutation rates. I attempt to verify the utility of the system by applying it to the mouse dorsal epidermis and the liver, respectively. Although the same approach using the stop-EGFP mouse could be, in principle, extended to other adult or developing tissues, verification of the utility in other tissues beyond the epidermis and liver is not provided in this thesis. Further studies in the future will be carried out to determine the extent of types of tissues where the stop-EGFP mouse can be efficiently applied for cell lineage and/or mutation studies.

VI. References

Alison MR, Poulsom R, Forbes S, Wright NA. (2002) An introduction to stem cells. *J Pathol.* 197: 419-423.

Allen TD, Potten CS. (1974) Fine-structural identification and organization of the epidermal proliferative unit. *J. Cell Sci.* 15: 291-319.

Ames BN, Lee FD, Durston WE. (1973) An improved bacterial test system for the detection and classification of mutagens and carcinogens. *Proc Natl Acad Sci U S A.* 70: 782-786.

Amit M, Carpenter MK, Inokuma MS, Chiu CP, Harris CP, Waknitz MA, Itskovitz-Eldor J, Thomson JA. (2000) Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev Biol.* 227: 271-278.

Anderson DJ, Gage FH, Weissman IL. (2001) Can stem cells cross lineage boundaries? *Nat Med.* 7: 393-395.

Auerbach C, Robson JM. (1946) Action of mustard gas on the bone marrow. *Natur.* 158: 878-879.

Aylon Y, Kupiec M. (2004) DSB repair: the yeast paradigm. *DNA Repair.* 3: 797-815.

Bachl J, Olsson C. (1999) Hypermutation targets a green fluorescent protein-encoding transgene in the presence of immunoglobulin enhancers. *Eur J Immunol.* 29: 1383-1389.

Balling R. (2001) ENU mutagenesis: analyzing gene function in mice. *Annu Rev Genomics Hum Genet.* 2: 463-492.

Barthel R, Aberdam D. (2005) Epidermal stem cells. *J Eur Acad Dermatol Venereol.* 19: 405-413.

Basu-Modak S, Tyrrell RM. (1993) Singlet oxygen: a primary effector in the ultraviolet A/near-visible light induction of the human heme oxygenase gene. *Cancer Res.* 53: 4505-4010.

Bertram JS. The molecular biology of cancer. (2000) *Mol Aspects Med.* 21: 167-223.

Bickenbach JR. (1981) Identification and behavior of label-retaining cells in oral mucosa and skin. *J Dent Res.* 60: 1611-1620.

Bjerknes M, Cheng H. (1999) Clonal analysis of mouse intestinal epithelial progenitors. *Gastroenterology.* 116: 7-14.

Bjerknes M, Cheng H. (2002) Multipotential stem cells in adult mouse gastric epithelium. *Am J Physiol Gastrointest Liver Physiol.* 283: G767-G777.

Bjornson CR, Rietze RL, Reynolds BA, Magli MC, Vescovi AL. (1999) Turning brain into blood: a hematopoietic fate adopted by adult neural stem cells in vivo. *Science.* 283: 534-537.

Blau HM, Brazelton TR, Weimann JM. (2001) The evolving concept of a stem cell: entity or function? *Cell.* 105: 829-841.

Bonnerot C, Nicolas JF. (1993) Clonal analysis in the intact mouse embryo by intragenic homologous recombination. *C R Acad Sci III.* 316: 1207-1217.

Braun KM, Watt FM. (2004) Epidermal label-retaining cells: background and recent applications. *J Investig Dermatol Symp Proc.* 9: 196-201.

Brazelton TR, Rossi FM, Keshet GI, Blau HM. (2000) From marrow to brain: expression of neuronal phenotypes in adult mice. *Science.* 290: 1775-1779.

Breimer LH. (1988) Ionizing radiation-induced mutagenesis. *Br J Cancer.* 57: 6-18.

Brustle O, Jones KN, Learish RD, Karram K, Choudhary K, Wiestler OD, Duncan ID, McKay RD. (1999) Embryonic stem cell-derived glial precursors: a source of myelinating transplants. *Science.* 285: 754-756.

Burdette WJ. (1955) The significance of mutation in relation to the origin of tumors: a review. *Cancer Res.* 15: 201-226.

Camargo FD, Green R, Capetanaki Y, Jackson KA, Goodell MA. (2003) Single hematopoietic stem cells generate skeletal muscle through myeloid intermediates. *Nat Med.* 9: 1520-1527.

Cariello NF, Narayanan S, Kwanyuen P, Muth H, Casey WM. (1998) A novel bacterial reversion and forward mutation assay based on green fluorescent protein. *Mutat Res.* 414: 95-105.

Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC. (1994) Green fluorescent protein as a marker for gene expression. *Science.* 263: 802-805.

Chan AW, Chong KY, Martinovich C, Simerly C, Schatten G. (2001) Transgenic monkeys produced by retroviral gene transfer into mature oocytes. *Science.* 291: 309-312.

Cheng H, Leblond CP. (1974) Origin, differentiation and renewal of the four main epithelial cell types in the mouse small intestine. V. Unitarian Theory of the origin of the four epithelial cell types. *Am J Anat.* 141: 537-561.

Chiu W, Niwa Y, Zeng W, Hirano T, Kobayashi H, Sheen J. (1996) Engineered GFP as a vital reporter in plants. *Curr Biol.* 6: 325-330.

Christmann M, Tomicic MT, Roos WP, Kaina B. (2003) Mechanisms of human DNA repair: an update. *Toxicology.* 193: 3-34.

Clarke JD, Tickle C. (1999) Fate maps old and new. *Nat Cell Biol.* 1: E103-E109.

Colussi C, Parlanti E, Degan P, Aquilina G, Barnes D, Macpherson P, Karran P, Crescenzi M, Dogliotti E, Bignami M. (2002) The mammalian mismatch repair pathway removes DNA 8-oxodGMP incorporated from the oxidized dNTP pool. *Curr Biol.* 12: 912-918.

Cormack BP, Valdivia RH, Falkow S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene.* 173: 33-38.

Cosentino L, Heddle JA. (1999) A comparison of the effects of diverse mutagens at the lacZ transgene and Dlb-1 locus in vivo. *Mutagenesis.* 14: 113-119.

Cosentino L, Heddle JA. (2000) Differential mutation of transgenic and endogenous loci in vivo. *Mutat Res.* 454: 1-10.

Cotsarelis G, Sun TT, Lavker RM. (1990) Label-retaining cells reside in the bulge area of pilosebaceous unit: implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell.* 61: 1329-1337.

Cubitt AB, Heim R, Adams SR, Boyd AE, Gross LA, Tsien RY. (1995) Understanding, improving and using green fluorescent proteins. *Trends Biochem Sci.* 20: 448-455.

Dean SW, Brooks TM, Burlinson B, Mirsalis J, Myhr B, Recio L, Thybaud V. (1999) Transgenic mouse mutation assay systems can play an important role in regulatory mutagenicity testing in vivo for the detection of site-of-contact mutagens. *Mutagenesis.* 14: 141-151.

DePrimo SE, Cao J, Hersh MN, Stringer JR. (1998) Use of human placental alkaline phosphatase transgenes to detect somatic mutation in mice in situ. *Methods.* 16: 49-61.

Dobrovolsky VN, McGarrity LJ, Morris SM, Heflich RH. (2002) Detection of mutation in transgenic CHO cells using green fluorescent protein as a reporter. *Mutat Res.* 518: 55-64.

Drake JW. (1970) Mutation rates. In: *The Molecular Basis of Mutation.* San Fransisco: Holden-Day, pp. 39-62.

Duckett DR, Drummond JT, Murchie AI, Reardon JT, Sancar A, Lilley DM, Modrich P. (1996) Human MutSalpha recognizes damaged DNA base pairs containing O6-methylguanine, O4-methylthymine, or the cisplatin-d(GpG) adduct. *Proc Natl Acad Sci U S A.* 93: 6443-6447.

Dycaico MJ, Provost GS, Kretz PL, Ransom SL, Moores JC, Short JM. (1994) The use of shuttle vectors for mutation analysis in transgenic mice and rats. *Mutat Res.* 307: 461-478.

Evans MJ, Kaufman MH. (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature.* 292: 154-156.

Feng G, Mellor RH, Bernstein M, Keller-Peck C, Nguyen QT, Wallace M, Nerbonne JM, Lichtman JW, Sanes JR. (2000) Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *Neuron.* 28: 41-51.

Ferrari G, Cusella-De Angelis G, Coletta M, Paolucci E, Stornaiuolo A, Cossu G, Mavilio F. (1998) Muscle regeneration by bone marrow-derived myogenic progenitors. *Science.* 279: 1528-1530.

Flake AW. (2001) Fate mapping of stem cells. In: *Stem Cell Biology* (Marshark DR, Gardner RL, Gottlieb D, eds). Cold Spring Harbor: Cold Spring Harbor Press, pp. 375-397.

Friedberg EC. (2001) How nucleotide excision repair protects against cancer. *Nat Rev Cancer.* 1: 22-33.

Friedberg EC. (2003) DNA damage and repair. *Nature.* 421: 436-440.

Friedberg EC, McDaniel LD, Schultz RA. (2004) The role of endogenous and exogenous DNA damage and mutagenesis. *Curr Opin Genet Dev.* 14: 5-10.

Galli R, Borello U, Gritti A, Minasi MG, Bjornson C, Coletta M, Mora M, De Angelis MG, Fiocco R, Cossu G, Vescovi AL. (2000) Skeletal myogenic potential of human and mouse neural stem cells. *Nat Neurosci.* 3: 986-991.

Gaziev J, Lucarelli G. (2003) Stem cell transplantation for hemoglobinopathies. *Curr Opin Pediatr.* 15: 24-31.

Genschel J, Bazemore LR, Modrich P. (2002) Human exonuclease I is required for 5' and 3' mismatch repair. *J Biol Chem.* 277: 13302-13311.

Gerecht-Nir S, Itskovitz-Eldor J. (2004) The promise of human embryonic stem cells. *Best Pract Res Clin Obstet Gynaecol.* 18: 843-852.

Goldfinger LE, Hopkinson SB, deHart GW, Collawn S, Couchman JR, Jones JC. (1999) The alpha3 laminin subunit, alpha6beta4 and alpha3beta1 integrin coordinately regulate wound healing in cultured epithelial cells and in the skin. *J Cell Sci.* 112: 2615-2629.

Goodman MF, Hopkins RL, Lasken R, Mhaskar DN. (1985) The biochemical basis of 5-bromouracil- and 2-aminopurine-induced mutagenesis. *Basic Life Sci.* 31: 409-423.

Gossen JA, de Leeuw WJ, Tan CH, Zwarthoff EC, Berends F, Lohman PH, Knook DL, Vijg J. (1989) Efficient rescue of integrated shuttle vectors from transgenic mice: a model for studying mutations in vivo. *Proc Natl Acad Sci U S A.* 86: 7971-7975.

Gossen JA, de Leeuw WJ, Vijg J. (1994) LacZ transgenic mouse models: their application in genetic toxicology. *Mutat Res.* 307: 451-459.

Graves KH, Moreadith RW. (1993) Derivation and characterization of putative pluripotential embryonic stem cells from preimplantation rabbit embryos. *Mol Reprod Dev.* 36: 424-433.

Green MH, Muriel WJ. (1976) Mutagen testing using TRP+ reversion in Escherichia coli. *Mutat Res.* 38: 3-32.

Griffiths DF, Davies SJ, Williams D, Williams GT, Williams ED. (1988) Demonstration of somatic mutation and colonic crypt clonality by X-linked enzyme histochemistry. *Nature.* 333: 461-463.

Grippo P, Iaccarino M, Parisi E, Scarano E. (1968) Methylation of DNA in developing sea urchin embryos. *J Mol Biol.* 36: 195-208.

Haas J, Park EC, Seed B. (1996) Codon usage limitation in the expression of HIV-1 envelope glycoprotein. *Curr Biol.* 6: 315-324.

Hadjantonakis AK, Nagy A. (2001) The color of mice: in the light of GFP-variant reporters. *Histochem Cell Biol.* 115: 49-58.

Heddle JA. (1998) The use of transgenic animals in research. *Environ Mol Mutagen.* 32: 110-114.

Heddle JA. (1999) On clonal expansion and its effects on mutant frequencies, mutation spectra and statistics for somatic mutations in vivo. *Mutagenesis.* 14: 257-260.

Hefferin ML, Tomkinson AE. (2005) Mechanism of DNA double-strand break repair by non-homologous end joining. *DNA Repair.* 4: 639-648.

Heim R, Prasher DC, Tsien RY. (1994) Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc Natl Acad Sci U S A.* 91: 12501-12504.

Heim R, Tsien RY. (1996) Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr Biol.* 6: 178-182.

Hoeijmakers JH. (2001) Genome maintenance mechanisms for preventing cancer. *Nature.* 411: 366-374.

Holliday R. (1964) A mechanism for gene conversion in fungi. *Genet Res.* 5: 282-304.

Ichihashi M, Ueda M, Budiyanto A, Bito T, Oka M, Fukunaga M, Tsuru K, Horikawa T. (2003) UV-induced skin damage. *Toxicology.* 189: 21-39.

Ikawa M, Yamada S, Nakanishi T, Okabe M. (1998) 'Green mice' and their potential usage in biological research. *FEBS Lett.* 430: 83-87.

Jackson AL, Loeb LA. (2001) The contribution of endogenous sources of DNA damage to the multiple mutations in cancer. *Mutat Res.* 477: 7-21.

Jackson KA, Mi T, Goodell MA. (1999) Hematopoietic potential of stem cells isolated from murine skeletal muscle. *Proc Natl Acad Sci U S A.* 96: 14482-14486.

Jackson KA, Majka SM, Wang H, Pocius J, Hartley CJ, Majesky MW, Entman ML, Michael LH, Hirschi KK, Goodell MA. (2001) Regeneration of ischemic cardiac muscle and vascular endothelium by adult stem cells. *J Clin Invest.* 107: 1395-1402.

Jeggo PA, Hafezparast M, Thompson AF, Broughton BC, Kaur GP, Zdzienicka MZ, Athwal RS. (1992) Localization of a DNA repair gene (XRCC5) involved in double-strand-break rejoining to human chromosome 2. *Proc Natl Acad Sci U S A.* 89: 6423-6427.

Jiricny J. (1998) Replication errors: cha(lle)nging the genome. *EMBO J.* 17: 6427-6436.

Jones IM, Burkhart-Schultz K, Carrano AV. (1985) A method to quantify spontaneous and in vivo induced thioguanine-resistant mouse lymphocytes. *Mutat Res.* 147: 97-105.

Jones PH, Watt FM. (1993) Separation of human epidermal stem cells from transit amplifying cells on the basis of differences in integrin function and expression. *Cell.* 73: 713-724.

Kanaar R, Hoeijmakers JH, van Gent DC. (1998) Molecular mechanisms of DNA double strand break repair. *Trends Cell Biol.* 8: 483-489.

Kim JH, Auerbach JM, Rodriguez-Gomez JA, Velasco I, Gavin D, Lumelsky N, Lee SH, Nguyen J, Sanchez-Pernaute R, Bankiewicz K, McKay R. (2002) Dopamine neurons derived from embryonic stem cells function in an animal model of Parkinson's disease. *Nature.* 418: 50-56.

Kohler SW, Provost GS, Fieck A, Kretz PL, Bullock WO, Sorge JA, Putman DL, Short JM. (1991) Spectra of spontaneous and mutagen-induced mutations in the lacI gene in transgenic mice. *Proc Natl Acad Sci U S A.* 88: 7958-7962.

Korbling M, Estrov Z. (2003) Adult stem cells for tissue repair - a new therapeutic concept? *N Engl J Med.* 349: 570-582.

Krause DS, Theise ND, Collector MI, Henegariu O, Hwang S, Gardner R, Neutzel S, Sharkis SJ. (2001) Multi-organ, multi-lineage engraftment by a single bone marrow-derived stem cell. *Cell.* 105: 369-377.

Kunkel TA. (1992) DNA replication fidelity. *J Biol Chem.* 267: 18251-18254.

Kunkel TA, Bebenek K. (2000) DNA replication fidelity. *Annu Rev Biochem.* 69: 497-529.

Lagasse E, Connors H, Al-Dhalimy M, Reitsma M, Dohse M, Osborne L, Wang X, Finegold M, Weissman IL, Grompe M. (2000) Purified hematopoietic stem cells can differentiate into hepatocytes in vivo. *Nat Med.* 6: 1229-1234.

Lechler T, Fuchs E. (2005) Asymmetric cell divisions promote stratification and differentiation of mammalian skin. *Nature.* 437: 275-280.

Li A, Simmons PJ, Kaur P. (1998) Identification and isolation of candidate human keratinocyte stem cells based on cell surface phenotype. *Proc Natl Acad Sci U S A.* 95: 3902-3907.

Li M, Zhang D, Hou Y, Jiao L, Zheng X, Wang WH. (2003) Isolation and culture of embryonic stem cells from porcine blastocysts. *Mol Reprod Dev.* 65: 429-434.

Lin Y, Weisdorf DJ, Solovey A, Hebbel RP. (2000) Origins of circulating endothelial cells and endothelial outgrowth from blood. *J Clin Invest.* 105: 71-77.

Lindahl T, Wood RD. (1999) Quality control by DNA repair. *Science.* 286: 1897-1905.

Litman RM, Pardee AB. (1956) Production of bacteriophage mutants by a disturbance of deoxyribonucleic acid metabolism. *Nature.* 178: 529-531.

Little JB. (2000) Radiation carcinogenesis. *Carcinogenesis.* 21: 397-404.

Liu T, Tannergard P, Hackman P, Rubio C, Kressner U, Lindmark G, Hellgren D, Lambert B, Lindblom A. (1999) Missense mutations in hMLH1 associated with colorectal cancer. *Hum Genet.* 105: 437-441.

Loeb LA. (1991) Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res.* 51: 3075-3079.

Loeffler M, Potten CS. (1997) Stem cells and cellular pedigrees-a conceptual introduction. In: *Stem Cells* (Potten CS, ed). San Diego: Academic Press, pp. 1-27.

Longley MJ, Pierce AJ, Modrich P. (1997) DNA polymerase delta is required for human mismatch repair in vitro. *J Biol Chem.* 272: 10917-10921.

Lyle S, Christofidou-Solomidou M, Liu Y, Elder DE, Albelda S, Cotsarelis G. (1998) The C8/144B monoclonal antibody recognizes cytokeratin 15 and defines the location of human hair follicle stem cells. *J Cell Sci.* 111: 3179-3188.

MacGregor JT, Casciano D, Muller L. (2000) Strategies and testing methods for identifying mutagenic risks. *Mutat Res.* 455: 3-20.

Mackenzie IC. (1997) Retroviral transduction of murine epidermal stem cells demonstrates clonal units of epidermal structure. *J Invest Dermatol.* 109: 377-383.

Mackenzie IC, Bickenbach JR. (1985) Label-retaining keratinocytes and Langerhans cells in mouse epithelia. *Cell Tissue Res.* 242: 551-556.

Mainiero F, Pepe A, Yeon M, Ren Y, Giancotti FG. (1996) The intracellular functions of alpha6beta4 integrin are regulated by EGF. *J Cell Biol.* 134: 241-253.

Martin GR. (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A.* 78: 7634-7638.

Martin P. (1997) Wound healing--aiming for perfect skin regeneration. *Science.* 276: 75-81.

McKinney-Freeman SL, Jackson KA, Camargo FD, Ferrari G, Mavilio F, Goodell MA. (2002) Muscle-derived hematopoietic stem cells are hematopoietic in origin. *Proc Natl Acad Sci U S A.* 99: 1341-1346.

Montoliu L, Chavez S, Vidal M. (2000) Variegation associated with lacZ in transgenic animals: a warning note. *Transgenic Res.* 9: 237-239.

Morasso MI, Tomic-Canic M. (2005) Epidermal stem cells: the cradle of epidermal determination, differentiation and wound healing. *Biol Cell.* 97: 173-183.

Morris RJ, Fischer SM, Slaga TJ. (1985) Evidence that the centrally and peripherally located cells in the murine epidermal proliferative unit are two distinct cell populations. *J Invest Dermatol.* 84: 277-281.

Morris RJ, Potten CS. (1994) Slowly cycling (label-retaining) epidermal cells behave like clonogenic stem cells in vitro. *Cell Prolif.* 27: 279-289.

Morris RJ, Potten CS. (1999) Highly persistent label-retaining cells in the hair follicles of mice and their fate following induction of anagen. *J Invest Dermatol.* 112: 470-475.

Morrison C, Takeda S. (2000) Genetic analysis of homologous DNA recombination in vertebrate somatic cells. *Int J Biochem Cell Biol.* 32: 817-831.

Morrison SJ, Weissman IL. (1994) The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity.* 1: 661-673.

Morrison SJ, Shah NM, Anderson DJ. (1997) Regulatory mechanisms in stem cell biology. *Cell.* 88: 287-298.

Mortelmans K, Zeiger E. (2000) The Ames Salmonella/microsome mutagenicity assay. *Mutat Res.* 455: 29-60.

Muller HJ. (1927) Artificial transmutation of the gene. *Science.* 66: 84-87.

Negishi K, Bessho T, Hayatsu H. (1994) Nucleoside and nucleobase analog mutagens. *Mutat Res.* 318: 227-238.

Niemann C, Watt FM. (2002) Designer skin: lineage commitment in postnatal epidermis. *Trends Cell Biol.* 12: 185-192.

Nohmi T, Suzuki T, Masumura K. (2000) Recent advances in the protocols of transgenic mouse mutation assays. *Mutat Res.* 455: 191-215.

Noveroske JK, Weber JS, Justice MJ. (2000) The mutagenic action of N-ethyl-N-nitrosourea in the mouse. *Mamm Genome.* 11: 478-483.

Okabe M, Ikawa M, Kominami K, Nakanishi T, Nishimune Y. (1997) 'Green mice' as a source of ubiquitous green cells. *FEBS Lett.* 407: 313-319.

Ono T, Miyamura Y, Ikehata H, Yamanaka H, Kurishita A, Yamamoto K, Suzuki T, Nohmi T, Hayashi M, Sofuni T. (1995) Spontaneous mutant frequency of lacZ gene in spleen of transgenic mouse increases with age. *Mutat Res.* 338: 183-188.

Oshima H, Rochat A, Kedzia C, Kobayashi K, Barrandon Y. (2001) Morphogenesis and renewal of hair follicles from adult multipotent stem cells. *Cell.* 104: 233-245.

Otto WR. (2002) Lung epithelial stem cells. *J Pathol.* 197: 527-535.

Papadopoulos N, Lindblom A. (1997) Molecular basis of HNPCC: mutations of MMR genes. *Hum Mutat.* 10: 89-99.

Patterson G, Day RN, Piston D. (2001) Fluorescent protein spectra. *J Cell Sci.* 114: 837-838.

Petersen BE, Bowen WC, Patrene KD, Mars WM, Sullivan AK, Murase N, Boggs SS, Greenberger JS, Goff JP. (1999) Bone marrow as a potential source of hepatic oval cells. *Science.* 284: 1168-1170.

Potten CS. (2004) Keratinocyte stem cells, label-retaining cells and possible genome protection mechanisms. *J Investig Dermatol Symp Proc.* 9: 183-195.

Potten CS, Morris RJ. (1988) Epithelial stem cells in vivo. *J Cell Sci.* 10 (Suppl): 45-62.

Potten CS, Loeffler M. (1990) Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt. *Development.* 110: 1001-1020.

Raff M. (2003) Adult stem cell plasticity: fact or artifact? *Annu Rev Cell Dev Biol.* 19: 1-22.

Ramalho-Santos M, Yoon S, Matsuzaki Y, Mulligan RC, Melton DA. (2002) "Stemness": transcriptional profiling of embryonic and adult stem cells. *Science.* 298: 597-600.

Reeves WH, Sthoeger ZM. (1989) Molecular cloning of cDNA encoding the p70 (Ku) lupus autoantigen. *J Biol Chem.* 264: 5047-5052.

Reliene R, Schiestl RH. (2003) Mouse models for induced genetic instability at endogenous loci. *Oncogene.* 22: 7000-7010.

Rizvi AZ, Wong MH. (2005) Epithelial stem cells and their niche: there's no place like home. *Stem Cells.* 23: 150-165.

Ro S, Rannala B. (2001) Methylation patterns and mathematical models reveal dynamics of stem cell turnover in the human colon. *Proc Natl Acad Sci U S A.* 98: 10519-10521.

Robert JS. (2004) Model systems in stem cell biology. *Bioessays.* 26: 1005-1012.

Russell LB. (1977) Validation of the in vivo somatic mutation method in the mouse as a prescreen for germinal point mutations. *Arch Toxicol.* 38: 75-85.

Russell LB, Major MH. (1957) Radiation-induced presumed somatic mutations in the house mouse. *Genetics.* 42: 161-175.

Russell WL. (1951) X-ray-induced mutations in mice. *Cold Spring Harb Symp Quant Biol.* 16: 327-336.

Russell WL, Kelly EM, Hunsicker PR, Bangham JW, Maddux SC, Phipps EL. (1979) Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. *Proc Natl Acad Sci U S A.* 76: 5818-5819.

Rycyna RE, Alderfer JL. (1985) UV irradiation of nucleic acids: formation, purification and solution conformational analysis of the '6-4 lesion' of dTpdT. *Nucleic Acids Res.* 13: 5949-5963.

Sanes JR, Rubenstein JL, Nicolas JF. (1986) Use of a recombinant retrovirus to study post-implantation cell lineage in mouse embryos. *EMBO J.* 5: 3133-3142.

Schmidt GH, Winton DJ, Ponder BA. (1988) Development of the pattern of cell renewal in the crypt-villus unit of chimaeric mouse small intestine. *Development.* 103: 785-790.

Schmidt GH, O'Sullivan JF, Paul D. (1990) Ethylnitrosourea-induced mutations in vivo involving the Dolichos biflorus agglutinin receptor in mouse intestinal epithelium. *Mutat Res.* 228: 149-155.

Schofield MJ, Hsieh P. (2003) DNA mismatch repair: molecular mechanisms and biological function. *Annu Rev Microbiol.* 57: 579-608.

Scrable H, Stambrook PJ. (1997) Activation of the lac repressor in the transgenic mouse. *Genetics.* 147: 297-304.

Sega GA. (1984) A review of the genetic effects of ethyl methanesulfonate. *Mutat Res.* 134: 113-142.

Setlow RB, Carrier WL. (1966) Pyrimidine dimers in ultraviolet-irradiated DNA's. *J Mol Biol.* 17: 237-254.

Shen JC, Rideout WM 3rd, Jones PA. (1994) The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. *Nucleic Acids Res.* 22: 972-976.

Shibuya T, Morimoto K. (1993) A review of the genotoxicity of 1-ethyl-1-nitrosourea. *Mutat Res.* 297: 3-38.

Singer B, Dosanjh MK. (1990) Site-directed mutagenesis for quantitation of base-base interactions at defined sites. *Mutat Res.* 233: 45-51.

Skopek TR. (1998) Transgenic mutation models: research, testing, and reality checks. *Environ Mol Mutagen.* 32: 104-105.

Skopek TR, Walker VE, Cochrane JE, Craft TR, Cariello NF. (1992) Mutational spectrum at the Hprt locus in splenic T cells of B6C3F1 mice exposed to N-ethyl-N-nitrosourea. *Proc Natl Acad Sci U S A.* 89: 7866-7870.

Skopek TR, Marino D, Kort K, Miller J, Trumbauer M, Gopal S, Chen H. (1998) Effect of target gene CpG content on spontaneous mutation in transgenic mice. *Mutat Res.* 400: 77-88.

Slack JM. (2000) Stem cells in epithelial tissues. *Science.* 287: 1431-1433.

Spradling A, Drummond-Barbosa D, Kai T. (2001) Stem cells find their niche. *Nature.* 414: 98-104.

Stern CD, Fraser SE. (2001) Tracing the lineage of tracing cell lineages. *Nat Cell Biol.* 3: E216-E218.

Suzuki T, Itoh S, Nakajima M, Hachiya N, Hara T. (1999) Target organ and time-course in the mutagenicity of five carcinogens in MutaMouse: a summary report of the second

collaborative study of the transgenic mouse mutation assay by JEMS/MMS. *Mutat Res.* 444: 259-268.

Svejstrup JQ. (2002) Mechanisms of transcription-coupled DNA repair. *Nat Rev Mol Cell Biol.* 3: 21-29.

Svendsen CN, Smith AG. (1999) New prospects for human stem-cell therapy in the nervous system. *Trends Neurosci.* 22: 357-364.

Tani H, Morris RJ, Kaur P. (2000) Enrichment for murine keratinocyte stem cells based on cell surface phenotype. *Proc Natl Acad Sci U S A.* 97: 10960-10965.

Tao KS, Urlando C, Heddle JA. (1993) Comparison of somatic mutation in a transgenic versus host locus. *Proc Natl Acad Sci U S A.* 90: 10681-10685.

Taylor G, Lehrer MS, Jensen PJ, Sun TT, Lavker RM. (2000) Involvement of follicular stem cells in forming not only the follicle but also the epidermis. *Cell.* 102: 451-461.

Terada N, Hamazaki T, Oka M, Hoki M, Mastalerz DM, Nakano Y, Meyer EM, Morel L, Petersen BE, Scott EW. (2002) Bone marrow cells adopt the phenotype of other cells by spontaneous cell fusion. *Nature.* 416: 542-545.

Terskikh VV, Vasiliev AV, Vorotelyak EA. (2003) Structural-functional units of epidermis. *Biol Bull.* 30: 535-539.

Thompson JN Jr, Woodruff RC, Huai H. (1998) Mutation rate: a simple concept has become complex. *Environ Mol Mutagen.* 32: 292-300.

Thomson JA, Kalishman J, Golos TG, Durning M, Harris CP, Becker RA, Hearn JP. (1995) Isolation of a primate embryonic stem cell line. *Proc Natl Acad Sci U S A.* 92: 7844-7848.

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM. (1998) Embryonic stem cell lines derived from human blastocysts. *Science.* 282: 1145-1147.

Thybaud V, Dean S, Nohmi T, de Boer J, Douglas GR, Glickman BW, Gorelick NJ, Heddle JA, Heflich RH, Lambert I, Martus HJ, Mirsalis JC, Suzuki T, Yajima N. (2003) In vivo transgenic mutation assays. *Mutat Res.* 540: 141-151.

Tosh D, Slack JM. (2002) How cells change their phenotype. *Nat Rev Mol Cell Biol.* 3: 187-194.

Tsien RY. (1998) The green fluorescent protein. *Annu Rev Biochem.* 67: 509-544.

Tumbar T, Guasch G, Greco V, Blanpain C, Lowry WE, Rendl M, Fuchs E.(2004) Defining the epithelial stem cell niche in skin. *Science.* 303: 359-363.

Tyrrell RM. (2000) Role for singlet oxygen in biological effects of ultraviolet A radiation. *Methods Enzymol.* 319: 290-296.

Umar A, Kunkel TA. (1996) DNA-replication fidelity, mismatch repair and genome instability in cancer cells. *Eur J Biochem.* 238: 297-307.

Vassilopoulos G, Wang PR, Russell DW. (2003) Transplanted bone marrow regenerates liver by cell fusion. *Nature.* 422: 901-904.

Wagers AJ, Weissman IL. (2004) Plasticity of adult stem cells. *Cell.* 116: 639-648.

Walker VE, Sisk SC, Upton PB, Wong BA, Recio L. (1997) In vivo mutagenicity of ethylene oxide at the hprt locus in T-lymphocytes of B6C3F1 lacI transgenic mice following inhalation exposure. *Mutat Res.* 392: 211-222.

Wang H, Lawrence CW, Li GM, Hays JB. (1999) Specific binding of human MSH2.MSH6 mismatch-repair protein heterodimers to DNA incorporating thymine- or uracil-containing UV light photoproducts opposite mismatched bases. *J Biol Chem.* 274: 16894-16900.

Wang X, Willenbring H, Akkari Y, Torimaru Y, Foster M, Al-Dhalimy M, Lagasse E, Finegold M, Olson S, Grompe M. (2003) Cell fusion is the principal source of bone-marrow-derived hepatocytes. *Nature.* 422: 897-901.

Watt FM, Hogan BL. (2000) Out of Eden: stem cells and their niches. *Science.* 287: 1427-1430.

Weissman IL. (2000) Stem cells: units of development, units of regeneration, and units in evolution. *Cell.* 100: 157-168.

Weissman IL, Anderson DJ, Gage F. (2001) Stem and progenitor cells: origins, phenotypes, lineage commitments, and transdifferentiations. *Annu Rev Cell Dev Biol.* 17: 387-403.

Winton DJ. (1997) Intestinal stem cells and clonality. In: *The Gut as a Model in Cell and Molecular Biology* (Halter F, Winton D, Wright NA, eds). Lancaster: Kluwer Academic Publishers, pp. 3-13.

Winton DJ, Blount MA, Ponder BA. (1988) A clonal marker induced by mutation in mouse intestinal epithelium. *Nature.* 333: 463-466.

Winton DJ, Peacock JH, Ponder BA. (1989) Effect of gamma radiation at high- and low-dose rate on a novel in vivo mutation assay in mouse intestine. *Mutagenesis.* 4: 404-406.

Wobus AM, Boheler KR. (2005) Embryonic stem cells: prospects for developmental biology and cell therapy. *Physiol Rev.* 85: 635-678.

Wood RD, Mitchell M, Sgouros J, Lindahl T. (2001) Human DNA repair genes. *Science.* 291: 1284-1289.

Yang TT, Cheng L, Kain SR. (1996) Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. *Nucleic Acids Res.* 24: 4592-4593.

Yang Y, Min JY, Rana JS, Ke Q, Cai J, Chen Y, Morgan JP, Xiao YF. (2002) VEGF enhances functional improvement of postinfarcted hearts by transplantation of ESC-differentiated cells. *J Appl Physiol.* 93: 1140-1151.

Ying QL, Nichols J, Evans EP, Smith AG. (2002) Changing potency by spontaneous fusion. *Nature.* 416: 545-548.

Yu YA, Oberg K, Wang G, and Szalay AA. (2003) Visualization of molecular and cellular events with green fluorescent proteins in developing embryos: a review. *Luminescence*. 18: 1-18.

Zernicka-Goetz M. (1999) Green fluorescent protein: a new approach to understanding spatial patterning and cell fate in early mammalian development. In: *Cell Lineage and Fate Determination* (Moody SA, ed). San Diego: Academic Press, pp. 521-527.

Zhang W, Remenyik E, Zelterman D, Brash DE, Wikonkal NM. (2001) Escaping the stem cell compartment: sustained UVB exposure allows p53-mutant keratinocytes to colonize adjacent epidermal proliferating units without incurring additional mutations. *Proc Natl Acad Sci U S A*. 98: 13948-13953.

Zhu AJ, Watt FM. (1999) beta-catenin signalling modulates proliferative potential of human epidermal keratinocytes independently of intercellular adhesion. *Development*. 126: 2285-2298.

Chapter 2. Generation and Characterization of the Stop-EGFP Mouse

Parts of this chapter have been previously published.

Ro S, Rannala B. (2004) A stop-EGFP transgenic mouse to detect clonal cell lineages generated by mutation. *EMBO Rep.* 5: 914-920.

I. Introduction

A cell fate mapping technique that allows clonal cell lineages to be identified has several important advantages in addressing critical issues in stem cell biology (see sections II-3 and II-4, Chapter 1). Although many different mapping techniques have been developed, achieving accurate labeling of single cells and subsequent tracing of their clonal cell lineages remains a challenge (see section II-6, Chapter 1). To facilitate the investigation of clonal cell lineages, a novel transgenic mouse system (referred to as the stop-EGFP mouse) was developed as part of my research. The stop-EGFP mouse carries a premature stop codon-containing EGFP (enhanced green fluorescent protein) gene as a target gene for mutations. The principle of the stop-EGFP mouse system is that individual cells are labeled by means of random mutations arising at the premature stop codon within the stop-EGFP gene and cell lineages descended from the labeled (mutant) single cells express EGFP, which will allow clonal cell lineages to be traced in vivo via fluorescence imaging. Utilizing the stop-EGFP mouse system for tracing cell lineages is expected to have at least two major advantages. First, the stop-EGFP system allows one to investigate clonal cell lineages via fluorescence imaging of EGFP. Second, one can potentially investigate the dynamics of the development of a particular cell lineage by repeated in vivo imaging of the same cell lineage over time (see section I-2, Chapter 1).

The stop-EGFP system also has the potential to be applied to in vivo mutation studies in which the stop-EGFP gene is used as a reporter for mutation. In the stop-EGFP system, a mutant cell which has undergone a mutation at the premature stop codon within the stop-EGFP gene is expected to express EGFP and thus become green fluorescent. This will facilitate the process of identifying mutations in this system because detection

of green fluorescent mutant cells on the background of non-fluorescent cells can be easily and directly performed via fluorescence imaging.

The EGFP gene includes genetic modifications that confer gene mammalian characteristics, such as changes to the codon usage pattern that render it more typical of codon usage in mammalian genes (see section I-1, Chapter 1). Thus, a mutation detection system using the stop-EGFP gene as a mutational target is expected to have an important advantage over other transgenic mouse mutation detection systems using bacterial transgenes, *lacI* or *lacZ* in revealing mutational characteristics more typical of mammalian genes (Gossen et al., 1989; Kohler et al., 1991).

The stop-EGFP gene is placed under the control of a ubiquitous promoter, which will allow mRNA to be transcribed from the gene in every tissue in the stop-EGFP mouse. Transcription of the stop-EGFP gene will better reflect the gene characteristics of endogeneous genes in the mammalian system, offering another potential advantage over the *lacI* or *lacZ* transgenic mouse system in which the bacterial transgene is not transcribed. Another novel feature of the stop-EGFP system is that it can provide a 3-dimensional representation of the distribution of mutations in an organ, which allows mutation frequencies in different sectors of the organ to be directly compared.

Based on several anticipated advantages (as outlined above) and the potential versatility of the stop-EGFP system for cell lineage (and mutation) studies, the first phase of my research focused on developing the stop-EGFP transgenic mouse (see Figure 1-6). In this chapter, I will describe in detail the procedures by which the stop-EGFP system has been developed and also the characteristics of the transgenic mouse system.

II. Materials and Methods

1. Generation of the stop-EGFP and the wild-type EBFP genes

The plasmid pCX-EGFP was obtained which had been used to generate a "Green mouse" (Okabe et al., 1997). The plasmid harbors the wild-type EGFP gene under the control of the chicken beta-actin promoter. To generate a premature stop codon-containing EGFP gene from pCX-EGFP, PCR was performed using the primers STOPEGFP5 and 3EGFP (see Table 2-1 for primer sequences). PCR was performed in 50 µl reactions containing 1 unit of $Vent_R^{®}$ DNA polymerase (New England Biolabs), 0.2 mM dNTP (Roche), 1 µM primers, and 10 ng of pCX-EGFP. The reactions were performed in a thermocycler (iCycler, Bio-Rad) with an initial denaturation at 95 °C for 5 minutes, followed by 28 cycles of denaturation at 95 °C for 1 minute, annealing at 58 °C for 1 minute, extension at 72 °C for 1 minute, followed by a final extension at 72 °C for 5 minutes. The reaction products were resolved on a 1% low melting agarose gel (UltraPure™ Agarose, Invitrogen) in 1× TAE (40 mM Tris-Acetate, 1 mM EDTA, pH 8.3). A PCR product of 750 bp was isolated from the gel and purified using a centrifugal filter device (Ultrafree®-DA, Millipore). The purified PCR product (750 bp) was digested with *Eco*RI and subsequently used to replace the EGFP gene within pCX-EGFP after digestion of the plasmid with *Eco*RI. The new construct is referred to as pCX-stop-EGFP (see Figure 2-8 for the genetic map of pCX-stop-EGFP). The substitution of the stop-EGFP gene for the wild-type EGFP gene was confirmed by DNA sequencing using a LiCor-4200LR automated sequencer (LiCor Biosciences). Six nucleotides (GAGGAG) were inserted immediately upstream of the premature stop codon (TAG) to make it easy to distinguish

Table 2-1. PCR primers and their sequences.

Primer	Sequence (5' to 3')
STOPEGFP5	AGAATTCGCCACCATGGTGGAGGAGTAGAGCAAGGGCGAG-GAGCTGTTC
3EGFP	GTGAATTCTTACTTGTACAGCTC
SDMEBFP5	GTGACCACCCTGACCCACGGCGTGCAGTGCTTC
SDMEBFP3	GAAGCACTGCACGCCGTGGGTCAGGGTGGTCAC
EGFPLYS5	AGAATTCGCCACCATGGTGGAGGAGAAGAGCAAGGGCGAG-GAGCTGTTC
EGFPLEU5	AGAATTCGCCACCATGGTGGAGGAGTTGAGCAAGGGCGAG-GAGCTGTTC
EGFPTRP5	AGAATTCGCCACCATGGTGGAGGAGTGGAGCAAGGGCGAG-GAGCTGTTC
EGFPMIDDLE5	GAGGGCGACACCCTGGTGAACC
PCR100R	GCAAAGAATTCGCCACCATG
PCR100L	CCGTTTACGTCGCCGTCCAG
EBFPONLY5	TTCGCCACCATGGTGAGC
EBFPONLY3	CGTCCTTGAAGAAGATGGTC
QTEGFP5	GCTGCTGCCCGACAACC
QTEGFP3	GAACTCCAGCAGGACCATGTG

the stop-EGFP gene from the wild-type EBFP (enhanced blue fluorescent protein) gene by PCR (see text).

The wild-type EBFP gene was generated from the wild-type EGFP gene within pCX-EGFP using site-directed mutagenesis. The procedures used followed the manufacturer's guidelines (QuickChange™, Stratagene). The nucleotide C was substituted for T at the 199th nucleotide position of the EGFP gene (thus, encoding histidine rather than tyrosine at codon 67) (see Figure 2-6) using the primers SDMEBFP5 and SDMEBFP3 (see Table 2-1). The resulting construct is referred to as pCX-EBFP. Sequence change at the site was confirmed by DNA sequencing.

2. Generation of revertant forms of the stop-EGFP gene

To confirm the neutrality of the premature stop codon site, three possible revertant forms of the stop-EGFP gene, containing either AAG (encoding lysine), TTG (encoding leucine), or TGG (encoding tryptophan) at the site of the premature stop codon, were generated by PCR using pCX-stop-EGFP as a template, EGFPLYS5, EGFPLEU5, or EGFPTRP5 (see Table 2-1 for primer sequences) as a forward primer, and 3EGFP as a backward primer. The same PCR reaction conditions were used as described above (see section II-1). Each PCR product was purified from a 1% low melting agarose gel (UltraPure™ Agarose, Invitrogen) and subsequently used to replace the stop-EGFP gene within pCX-stop-EGFP after digestion of the plasmid and each PCR product with *Eco*RI (see section II-1). The resulting constructs are referred to as pCX-Lys-EGFP, pCX-Leu-EGFP, and pCX-Trp-EGFP, respectively. Sequence changes were confirmed by DNA sequencing.

3. Fluorescence intensity of proteins expressed from revertant constructs

To test the intensity of green fluorescence of protein expressed from each construct, a transient expression experiment was performed using the NIH3T3 mouse fibroblast cell line. The procedures used followed the manufacturer's guidelines (FuGENE 6 Transfection Reagent, Roche). Briefly, 22 hours before the transfection experiment, 0.25 x 10^5 cells were plated in each well of 24-well plate. On the day of the experiment, cells were 70 – 80% confluent. 24 μl of serum free media containing 1 μl of FuGENE 6 reagent was mixed with 0.33 μg of each DNA plasmid and incubated at room temperature for 30 minutes. The complex mixture was added to each well containing NIH3T3 cells. After incubation at 37 °C for 36 hours, the transfected cells were washed twice with 1× PBS (2.7 mM Potassium Cloride, 1.5 mM Potassium Phosphate Monobasic, 138 mM Sodium Cloride, 8 mM Sodium Phosphate Dibasic, pH 7.4). Transfected cells in each well were illuminated with a mercury lamp and imaged using an inverted fluorescence microscope. To quantitatively compare the intensity of fluorescence from each protein, fluorescence activated cell sorting (FACS) analysis was performed. The transfected cells were dissociated by treating of cells with trypsin and used for the FACS analysis. Flow cytometry analysis was performed using FACScan (BD Biosciences) equipped with 488 nm argon laser and standard filters for collection of green fluorescence signal. Data files were analyzed with CellQuest software (BD Biosciences).

4. Animal experiments

All experiments using live mice were performed in compliance with the recommendations of the Canadian Council on Animal Care and have been approved by the Health Sciences Animal Policy and Welfare Committee of the University of Alberta.

5. Microinjection and genomic PCR

The pCX-stop-EGFP and pCX-EBFP constructs were each digested with *Sal*I and *Bam*HI producing a DNA fragment of 3.2 kb, which then was purified (see Figure 2-8). The 3.2 kb fragment from pCX-stop-EGFP was mixed with that from pCX-EBFP with a ratio of 4 to 1, and then microinjected into the pronuclei of fertilized eggs from two different inbred strains (C57BL/6 and FVB/N) using standard techniques. The microinjection experiment was performed by staff in the transgenic animal facility at the University of Alberta.

Genotypes of mice produced were determined by genomic PCR using primers, EGFPMIDDLE5 and 3EGFP (see Table 2-1 for sequences). These primers are designed to generate PCR products of 380 bp from the transgenes. To extract genomic DNA, 1 cm of a tail tip was cut from each mouse and then incubated in a volume of 400 µl of proteinase K/SDS (Sodium Dodecyl Sulphate) buffer at 37 °C overnight (Miller et al., 1988). Protein pellet and debris were removed by a salting out procedure (Miller et al., 1988) and the supernatant was transferred to a new tube. Genomic DNA was precipitated by adding 2 volumes of room temperature absolute ethanol. Precipitated genomic DNA was washed with 70 % ethanol and resuspended in 100 µl TE buffer (10 mM Tris-Cl, 1 mM EDTA, pH 8.0). Genomic PCR was performed in a volume of 25 µl containing 1 µl of genomic DNA (prepared by the method described above), 1 unit of recombinant *Taq*

DNA polymerase (Invitrogen), 2 mM $MgCl_2$ (Invitrogen), 0.2 mM dNTP (Roche), and 1 μM primers. The reactions were performed in a thermocycler (iCycler, Bio-Rad) with an initial denaturation at 95 °C for 5 minutes, followed by 40 cycles of denaturation at 95 °C for 1 minute, annealing at 49 °C for 1 minute, extension at 72 °C for 50 seconds, followed by a final extension at 72 °C for 5 minutes. The reaction products were resolved on 1% agarose gels (UltraPure™ Agarose, Invitrogen) containing 0.5 μg/ml ethidium bromide and visualized using an ImageMaster VDS gel documentation system (Amersham Biosciences).

To determine the presence of both transgenes (i.e., the stop-EGFP gene and EBFP gene), PCR was performed using genomic DNA and primers, PCR100R and PCR100L (see Table 2-1 for sequences). These primers are designed to generate PCR products of 100 bp from the stop-EGFP gene and 91 bp from the EBFP gene. Genomic PCR was performed in a volume of 25 μl containing 1 μl of genomic DNA (prepared by the method described above), 1 unit of recombinant *Taq* DNA polymerase (Invitrogen), 2 mM $MgCl_2$ (Invitrogen), 0.2 mM dNTP (Roche), and 1 μM primers. The reactions were performed in a thermocycler (iCycler, Bio-Rad) with an initial denaturation at 95 °C for 5 minutes, followed by 42 cycles of denaturation at 95 °C for 1 minute, annealing at 52 °C for 1 minute, extension at 72 °C for 30 seconds, followed by a final extension at 72 °C for 3 minutes. The reaction products were resolved on 4% low melting agarose gels (UltraPure™ Agarose, Invitrogen) containing 1 μg/ml ethidium bromide and visualized using an ImageMaster VDS gel documentation system (Amersham Biosciences). A 10 bp DNA ladder (Life Technologies) was used as a size marker. When the presence of the EBFP gene could not be verified with PCR using these primers, the presence of the gene

was further tested using primers, EBFPONLY5 and EBFPONLY3 (Table 2-1). These primers are designed to specifically generate PCR products of 320 bp from the EBFP gene without amplifying the stop-EGFP gene (see section III-6). The reactions were performed in a thermocycler (iCycler, Bio-Rad) with an initial denaturation at 95 °C for 5 minutes, followed by 40 cycles of denaturation at 95 °C for 1 minute, annealing at 47 °C for 1 minute, extension at 72 °C for 60 seconds, followed by a final extension at 72 °C for 5 minutes. The reaction products were resolved on 1% agarose gels (UltraPure™ Agarose, Invitrogen) containing 0.5 μg/ml ethidium bromide and visualized using an ImageMaster VDS gel documentation system (Amersham Biosciences).

6. Breeding and maintenance of transgenic lines

To generate transgenic offspring and maintain transgenic lines, transgenic founders of the C57BL/6 strain were mated with C57BL/6 wild-type inbred mice and transgenic founders of the FVB/N strain were mated with BALB/c wild-type inbred mice. All mice used for breeding were kept in a VAF (Virus Antibody Free) facility and cared for by staff in Health Sciences Lab Animal Services (HSLAS). Genotypes of offspring were determined by genomic PCR using primers, EGFPMIDDLE5 and 3EGFP (see section II-5 for PCR reaction conditions).

7. Southern blot analysis and competitive PCR

To further verify transgene integration into the mouse genome, Southern blot analysis was performed. Tail tip biopsies from transgenic mice were used for extraction of genomic DNA (see section II-5). In Southern blot analysis, 10 μg of genomic DNA (size of about 6×10^9 bp in the diploid genome) per sample were digested with *Pst*I. As reference copy numbers of the transgenes, 10 pg (corresponding to one copy), 20 pg

(corresponding to two copies), and 100 pg (corresponding to ten copies) of pCX-stop-EGFP (size of about 6×10^3 bp) were digested with *Sal*I and *Bam*HI (producing a DNA fragment of 3.2 kb, see section II-5). Digested genomic DNA and reference DNA were electrophoresed on a 0.8% agarose gel. The gel was stained in a solution of 1 μg/ml ethidium bromide for 20 minutes after electrophoresis. After rinsing in deionized water, the gel was soaked in 0.25 M HCl for 10 minutes. The gel was then rinsed in deionized water and soaked in denaturing solution (1.5 M NaCl, 0.5 M NaOH) for 40 minutes with gentle rocking. The gel was then rinsed in deionized water and incubated in neutralization solution (1.5 M NaCl, 0.5 M Tris-HCl, pH 7.0) for 40 minutes with gentle rocking. The DNA was transferred to a nylon membrane (Hybond-N, Amersham Biosciences) by traditional capillary blotting overnight. After transfer, the DNA was bound to the membrane by UV cross-linking. The membrane was pre-hybridized in Church and Gilbert hybridization solution (Church and Gilbert, 1984) for one hour at 65 °C. The hybridization solution was replaced with fresh solution containing 1×10^6 cpm/ml of [α-^{32}P]dCTP-labeled *Eco*RI fragment (containing the stop-EGFP gene) of pCX-stop-EGFP (see Figure 2-8) and 100 μg/ml salmon sperm DNA (both denatured in a boiling-water bath and chilled quickly in ice water before adding to the hybridization solution). The Southern blot was probed at 65 °C overnight. After hybridization, the membrane was washed twice for 15 minutes with 2× SSC/0.1% SDS and twice for 15 minutes with 0.1× SSC/0.1% SDS at 55 °C (20× SSC: 3M NaCl, 0.3 M sodium citrate, pH 7.2). The blot was wrapped in plastic wrap and exposed to BioMax film (Kodak) at -80 °C overnight.

A competitive PCR (Gilliland et al., 1990; Diviacco et al., 1992) was performed to determine the relative ratio of the two transgenes in each transgenic line using primers, PCR100R and PCR100L. These primers are designed to generate PCR products of 100 bp from the stop-EGFP gene and 91 bp from the EBFP gene (see section II-5 for reaction conditions).

8. Real-time PCR

To quantitatively determine the total copy number of the transgenes in each transgenic line, real-time PCR was performed. Primers and probes were designed by the Primer Express 1.0 software program (Applied Biosystems). Real-time PCR analysis was performed using primers, QTEGFP5 and QTEGFP3 (Table 2-1). These primers are designed to yield products of 91 bp. The ABI Prism 7900HT sequence detection system (Applied Biosystems) was used. Reaction volumes of 7 µl were prepared in 384-well Micro-Amp optical plates (Applied Biosystems). Enzymes and reagents were from Sigma-Aldrich (Canada) and Invitrogen (Canada). Each reaction contained 1× buffer K, 3.5 mM $MgCl_2$, 200 µM dATP, dCTP, dGTP, 800 nM forward and backward primers, SYBR I Green (Molecular Probes), 1× ROX reference dye (Invitrogen), 0.03 unit/µl JumpStart *Taq* DNA polymerase (Sigma), and 20 ng genomic DNA. Each test sample was run in duplicate. PCR conditions were 94 °C for 40 seconds, then 40 cycles of 94 °C for 15 seconds, 60 °C for 1 minute. The runs were monitored via the Sequence Detection Software 2.1. After each run, a semi-log amplification plot was shown, in which the progression of the reactions along the PCR cycles was indicated by the increase in the normalized reporter fluorescence emission (Δ Rn). A threshold in the linear region of the semi-log amplification plot was established at an arbitrary value of the Δ Rn normalized

reporter fluorescence emission. For each reaction, a C_T value, which is the number of cycles necessary to reach the threshold, was identified. The absolute copy numbers of the transgenes were determined by the ratios to the copy number in line F#24 (presumed to carry one or two copies of the EBFP gene, see section III-7).

9. Reverse Transcription-PCR (RT-PCR)

To test expression of the two transgenes at the transcriptional level, reverse-transcriptional PCR (RT-PCR) was performed. RNA was extracted from various tissues and used as a substrate for reverse transcription (RT). First-strand complementary DNA (cDNA) synthesis from mRNA was performed using 5 μg total RNA and Superscript II reverse transcriptase (Invitrogen). The procedures used followed the manufacturer's guidelines. RT reactions were carried out at 42 °C for 50 minutes followed by inactivation of the reactions by heating at 70 °C for 15 minutes. The same reactions were performed without reverse transcriptase (a negative control, see below).

Subsequent PCR reactions were performed in 25 μl reactions containing 1 unit of recombinant *Taq* DNA polymerase (Invitrogen), 2 mM $MgCl_2$, 0.2 mM dNTP (Roche), 1 μM primers, and 1 μl of cDNA from first-strand reactions. Primers EGFPMIDDLE5 and 3EGFP were initially used for PCR reactions. To further verify mRNA transcription from both the stop-EGFP and EBFP genes, PCR reactions were performed using primers PCR100R and PCR100L. These primers are designed to generate PCR products of 100 bp from the stop-EGFP cDNA and 91 bp from the EBFP cDNA (see section II-5 for reaction conditions). The same PCR reactions were performed using 1 μl of the aforementioned reactions without the reverse transcription process (by not adding reverse transcriptase) and used as negative controls.

10. Western blotting

Protein extracts were prepared by homogenizing each organ (100 – 250 mg weight) in 1 ml 2× sample buffer (0.13 M Tris-HCl pH 6.8, 4% sodium dodecyl sulfate, 20% glycerol). After homogenization, each sample was transferred to a new tube. Genomic DNA was sheared by mixing up and down through a 22G needle 15 times, followed by a 26G needle 15 times. Samples were then incubated at 65 °C for 5 minutes, followed by centrifugation at 14,000 rpm for ten minutes. Supernatants were transferred to new tubes and stored at -80 °C.

Approximately, 10 μg of proteins from each sample were loaded to a well on a mini-gel and resolved on a 15% polyacrylamide gel. Pre-stained SDS-PAGE standards (Bio-Rad) were used as a size marker. After electrophoresis, proteins were transferred to a nitrocellulose membrane (Trans-Blot Transfer Medium, Bio-Rad). The blot was blocked in blocking buffer (5% powdered skim milk in rinse buffer, TBST: 25 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.05 % Tween-20) at 4 °C for 16 hours. The blot was then treated with rabbit polyclonal anti-GFP antibody (1:1000 dilution in blocking buffer) at room temperature with shaking for one and half hours. Following treatment with the primary antibody, the blot was washed with TBST four times in 10 minute intervals before treatment with horseradish peroxidase-conjugated anti-rabbit antibody (Amersham Biosciences) at room temperature with shaking for one and half hours (1:5000 dilution in blocking buffer). The blot was then washed with TBST five times in 10 minute intervals. Proteins were detected using a chemiluminescent substrate (Amersham Biosciences). The blot was exposed to BioMax film (Kodak). The anti-GFP antibody and purified GFP (a

positive control) were obtained as a generous gift from Dr. Luc Berthiaume (Department of Cell Biology, University of Alberta).

III. Results

1. The stop-EGFP system

To increase the efficiency with which individual cells are labeled by means of mutations, a premature stop codon (e.g., TAG) was introduced at a putatively neutral codon site within the EGFP gene. This allows a mutation arising at any one of the three nucleotides of the stop codon (TAG) to restore EGFP function (except in the case of a mutation from G to A at TAG, which produces another stop codon, TAA) because any substitution for the amino acid encoded by a neutral codon does not affect the protein function. That is, if a premature stop codon is located at a neutral codon site within the EGFP gene, all possible mutations from the premature stop codon **TAG** to **A**AG, **C**AG, **G**AG, T**C**G, T**G**G, T**T**G, TA**C**, and TA**T** should restore EGFP function except in the case of TA**A** (Figure 2-1). The fact that multiple copies of a transgene are usually inserted in a linear array in a transgenic mouse also increases the efficiency of this system for generating labeled cells via mutations (Goldman et al., 2004; Palmiter and Brinster, 1986). If a transgenic mouse carries 10 copies of the premature stop codon-containing EGFP gene, for example, 30 nucleotides are mutational targets (3 target nucleotides at a premature stop codon × 10 copies of the gene) which will restore EGFP function when mutated.

The mechanism by which the stop-EGFP system operates is illustrated in Figure 2-2. In cells of the stop-EGFP mouse, mRNA transcribed from the stop-EGFP gene will be translated in the cytoplasm by ribosomes which bind to the mRNA and initiate the

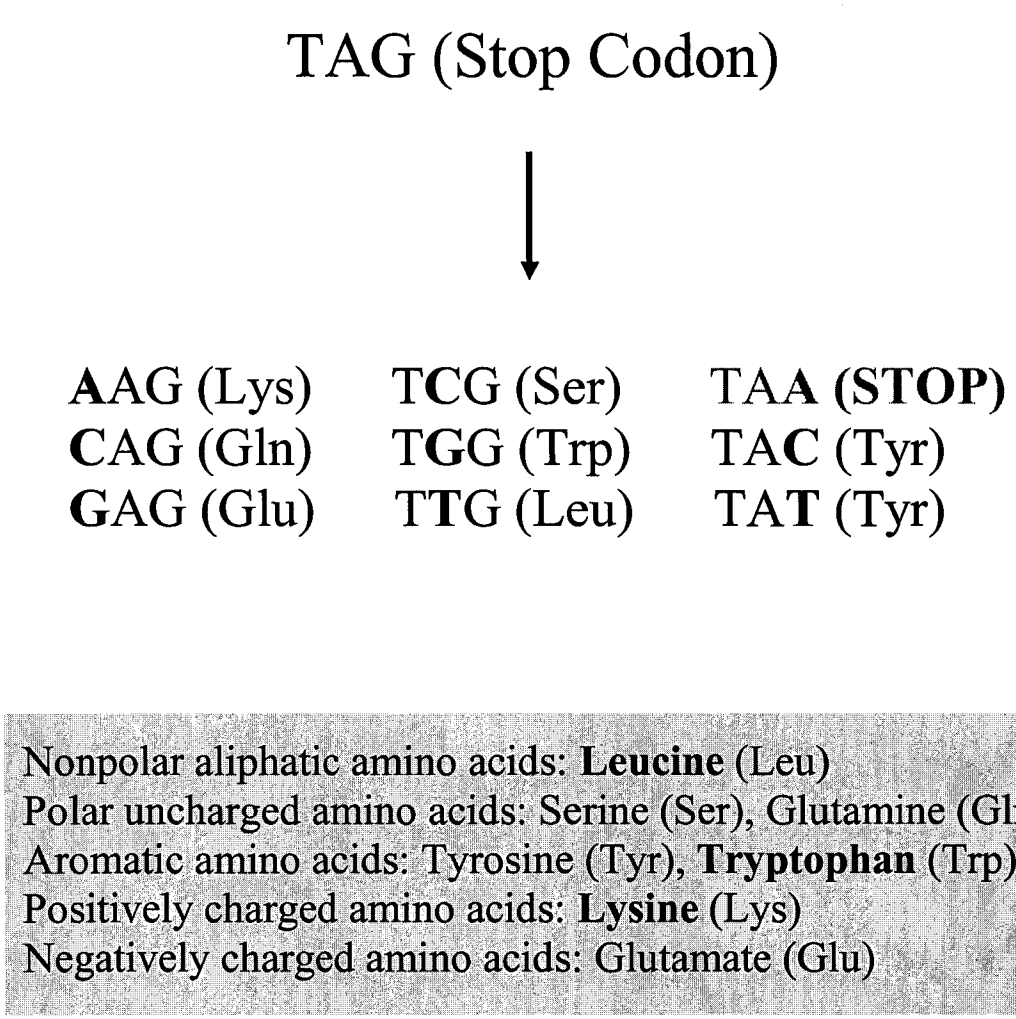


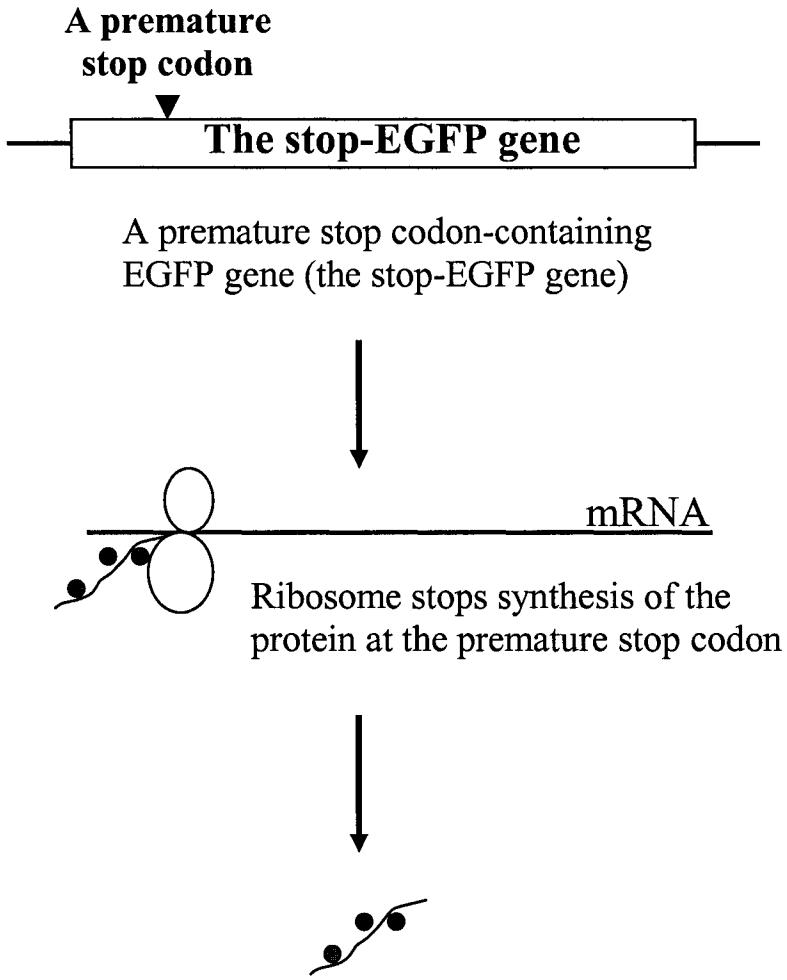
Figure 2-1. Possible substitutions from the premature stop codon, TAG and encoded amino acids from the substitutions.

Nine DNA substitutions can be generated at the premature stop codon from which seven different amino acids can be encoded. Mutation from G to A at the third site of TAG (i.e., GC → AT transition) generates another stop codon, TAA. The seven amino acids encoded from revertant mutations at the premature stop codon can be classified five groups based on the molecular characteristics of amino acids (Lehninger et al., 1993). Amino acids, leucine (nonpolar aliphatic), tryptophan (aromatic), and lysine (positively charged) were selected to test the neutrality of the premature stop codon site (see section III-5).

Figure 2-2. Schematic representation illustrating the mechanism by which the stop-EGFP system operates.

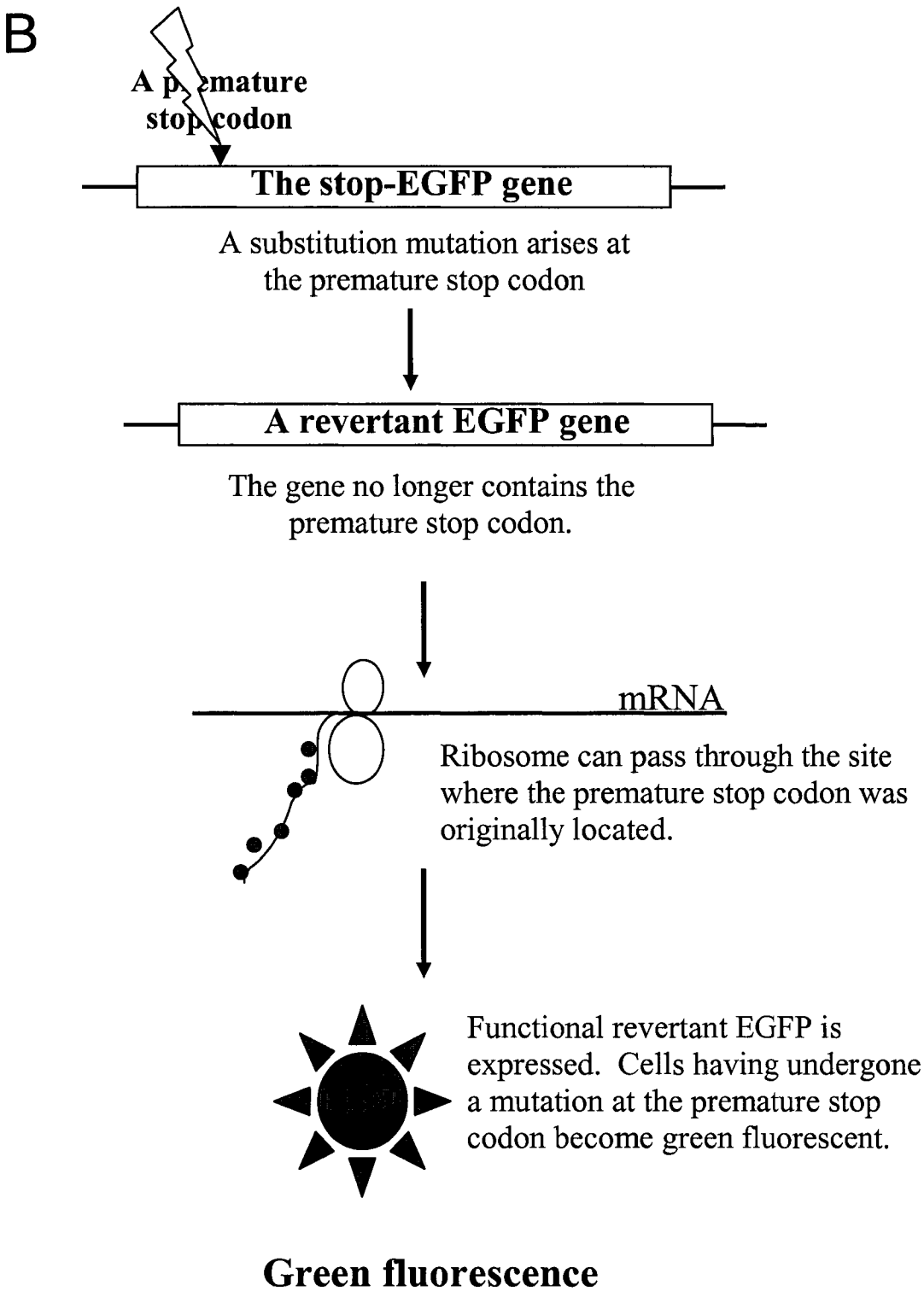
(A) In the absence of a mutation, cells containing the stop-EGFP gene will exhibit no green fluorescence when illuminated with the excitation light. (B) Cells having undergone a mutation at the premature stop codon will exhibit green fluorescence when excited with the 488 nm excitation light.

A



No green fluorescence

B



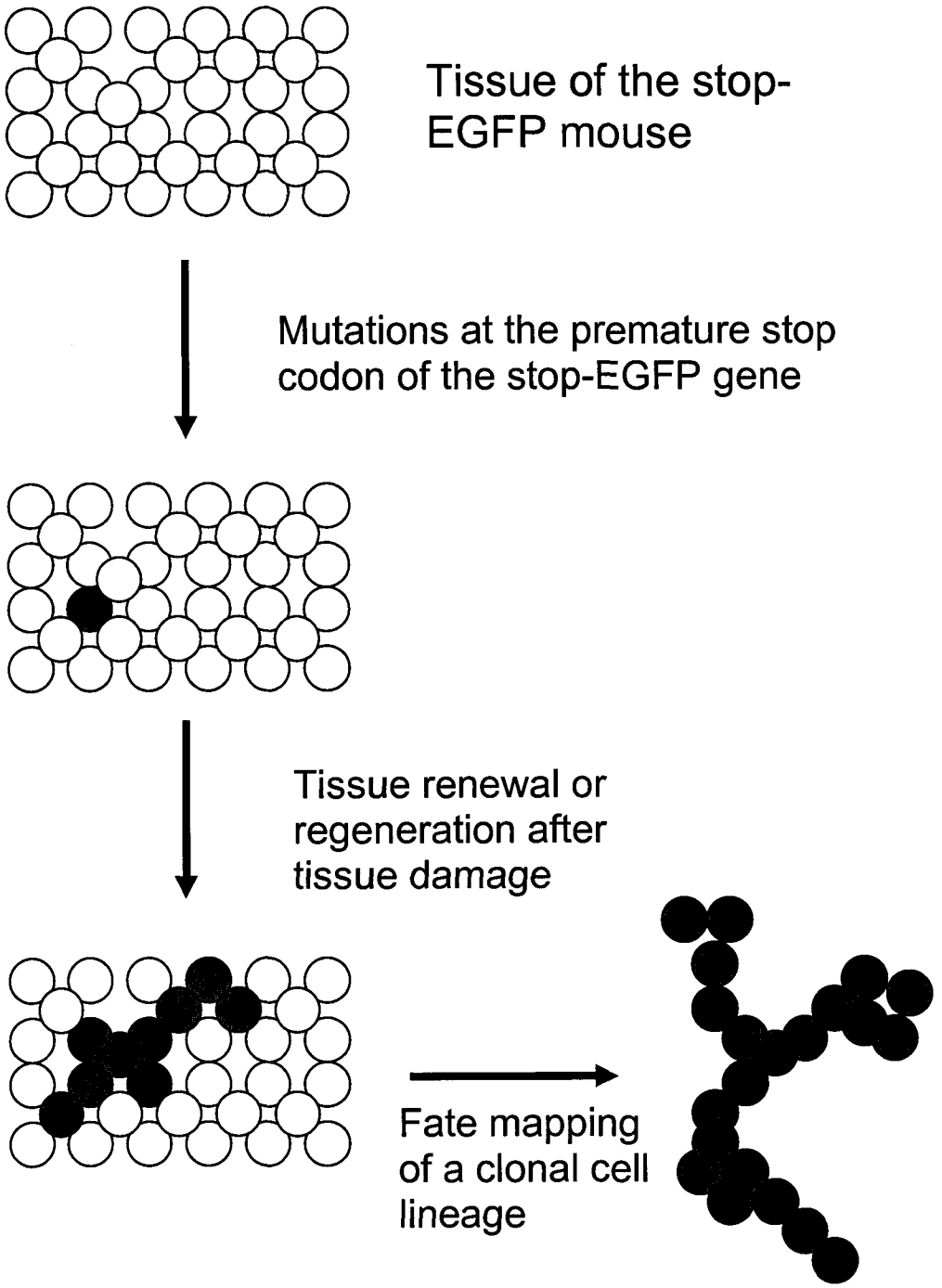
translation process at the start codon. When a ribosome encounters the premature stop codon during the elongation process, it terminates the translation process leading to the release of truncated peptides. Thus, in the absence of a mutation cells containing the stop-EGFP gene will exhibit no green fluorescence when illuminated with the excitation light (Figure 2-2A). On the other hand, if a mutation occurs at the premature stop codon, mRNA transcribed from the mutated copy of the stop-EGFP gene no longer contains the premature stop codon and this allows a ribosome to pass through the codon site where the premature stop codon was originally located, resulting in the synthesis of a full length protein. Because the premature stop codon was located at a neutral codon site, expressed protein is as functional as the wild-type EGFP, regardless of the specific amino acid encoded by the mutated site. Therefore, cells having undergone a mutation at the premature stop codon will exhibit green fluorescence when excited with the 488 nm excitation light (Figure 2-2B).

In the stop-EGFP system, random mutations at the premature stop codon of the stop-EGFP gene allow labeling of individual cells in tissue. Cells having undergone a mutation at the premature stop codon and their clonal descendants will express functional EGFP, thus clonal cell lineage can be traced by following green fluorescent colonies. For example, cell fate mapping of clonal cell lineages can be performed after tissue renewal (or tissue regeneration) (Figure 2-3).

If the stop-EGFP gene is only used as a target gene for mutations, the system is incapable of detecting a GC to AT transition mutation because the change produces another stop codon TAA from TAG (Figure 2-1). To make up for the inability of the stop-EGFP system to detect a GC to AT transition mutation at the premature stop codon,

Figure 2-3. Schematic illustration showing an example of application of the stop-EGFP system to cell fate mapping study.

Mutation at the premature stop codon within the stop-EGFP gene labels individual cells in tissue of the stop-EGFP mouse. After tissue renewal or regeneration, green fluorescent cell lineages can be imaged allowing fate mapping of clonal cell lineages.



the wild-type enhanced blue fluorescent protein (EBFP) gene was used as another mutational target. The EBFP gene differs from the EGFP gene by only a single nucleotide. The difference occurs in codon 67: CAC (encoding histidine) of the EBFP gene and TAC (encoding tyrosine) of the EGFP gene (Heim et al., 1994). An amino acid encoded by codon 67 (equivalent to codon 66 of the GFP or BFP gene, see section I-1, Chapter 1) determines the wavelengths of lights for both excitation and emission of the fluorescent proteins and substitution of a single amino acid at the codon site leads to a different type of fluorescent protein. Therefore, if a mutation changes C to T at the 199th nucleotide of the EBFP gene (i.e., GC → AT transition), a wild-type EGFP gene will be generated (i.e., a mutation from **C**AC to **T**AC at codon 67) (Figure 2-4). A transgenic system carrying both the stop-EGFP gene and the wild-type EBFP gene can therefore detect all possible point mutations arising by either transversions (AT → TA, AT → CG, GC → TA, and GC → CG) or transitions (AT → GC and GC → AT) (see Figure 2-5).

2. Identifying the site for a premature stop codon

Several factors were considered in choosing the site for a premature stop codon to be introduced. Most importantly, the site should be located in a region which is of no functional importance such that any amino acid encoded from the site does not degrade the function of EGFP. Choosing such a site will significantly increase the efficiency with which a revertant mutation is detected at the premature stop codon. Amino acids located very close to the N-terminal region of EGFP are presumed to be functionally neutral (Dopf and Horiagon, 1996; Li et al., 1997). Choosing a codon site very close to the translation initiation codon of the EGFP gene (corresponding to the N-terminal of the protein) as the location for a premature stop codon has another advantage in that the

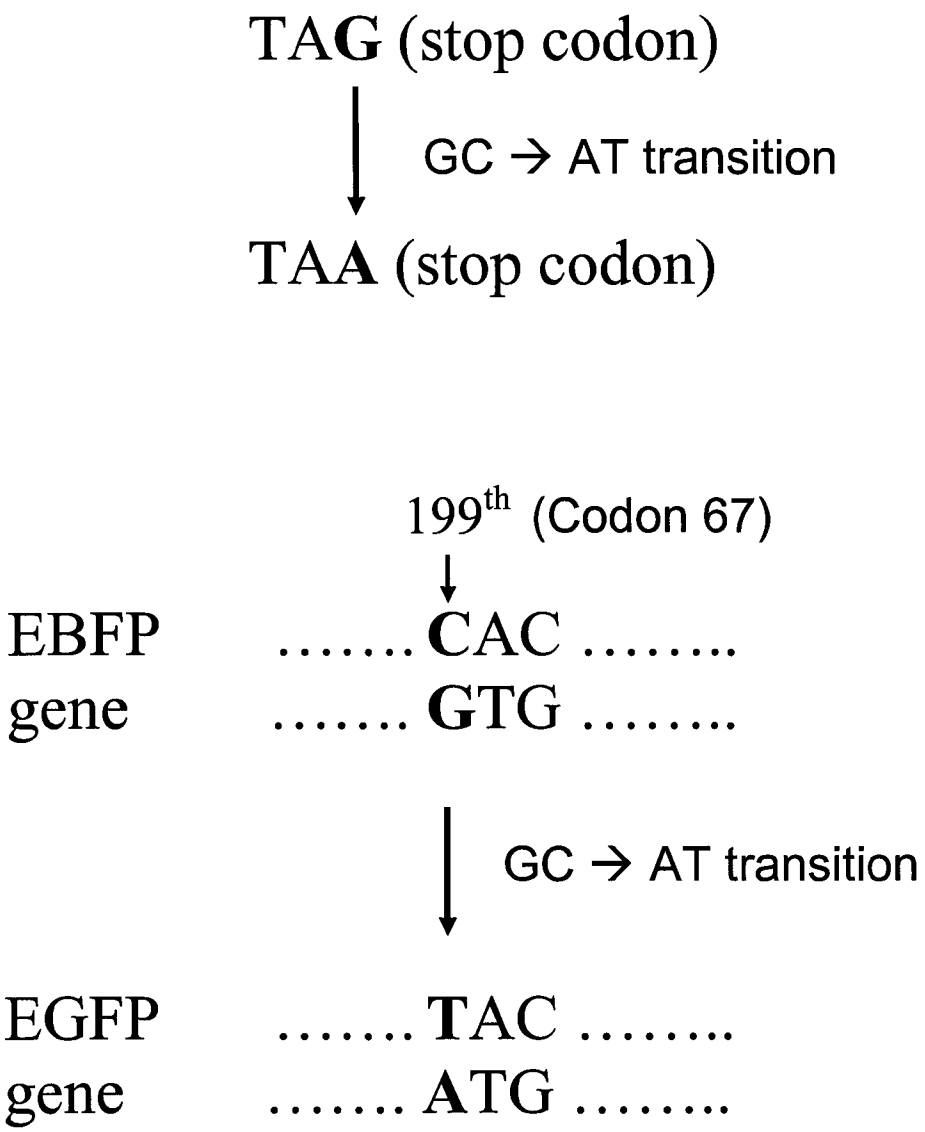


Figure 2-4. The wild-type EBFP gene which can detect a GC to AT transition mutation.

GC to AT transitional mutation at the premature stop codon does not restore EGFP function because it generates another stop codon, TAA. The wild-type EBFP gene can make up for the inability of the stop-EGFP gene to detect a GC to AT transition mutation at the premature stop codon. A mutation that changes C to T (i.e., GC to AT transition) at the 199th nucleotide of the EBFP gene generates a wild-type EGFP gene.

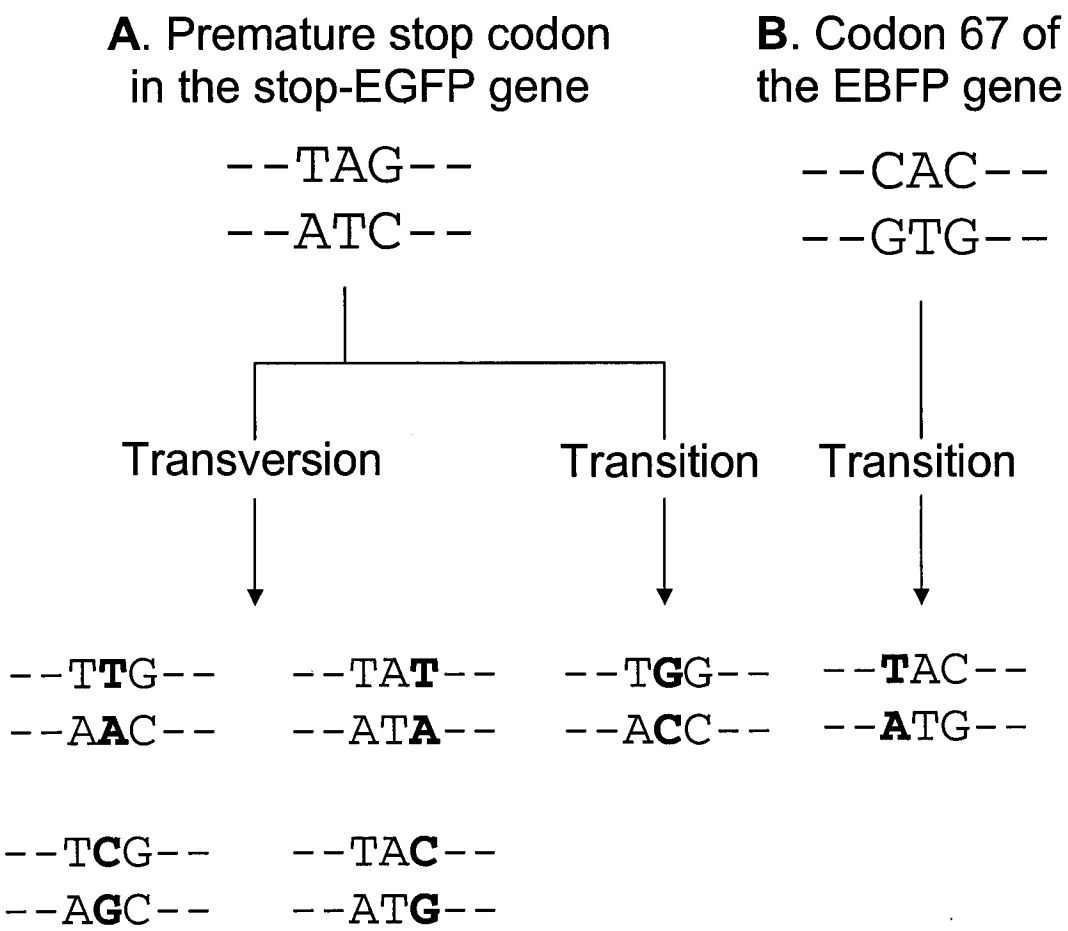


Figure 2-5. The stop-EGFP system using the wild-type EBFP gene as a supplementary target gene for mutations.

A transgenic system carrying both the stop-EGFP gene and the wild-type EBFP gene can detect all possible point mutations arising by either transversions (AT → TA, AT → CG, GC → TA, and GC → CG) or transitions (AT → GC and GC → AT). Examples of detection of all possible transition and transversion mutations in the stop-EGFP system are shown.

proximity of the stop codon to the translation initiation codon prevents cells from producing long truncated peptides which might have deleterious effects on a cell. However, one risk of introducing a premature stop codon very near the translation start codon of the EGFP gene is that translation termination at the premature stop codon might not be very efficient, which could lead to a high background level of green fluorescence from cells in the stop-EGFP mouse. It has been reported that placing a stop codon immediately after an initiation codon can lead to a low efficiency of translation termination (McCaughan et al., 1998). We chose to insert a stop codon (TAG) between the second and third codon, such that DNA sequences beyond the insertion are identical to those of the wild-type EGFP (Figure 2-6). To increase the distance between the initiation codon and the premature stop codon, six nucleotides were additionally inserted immediately upstream of the premature stop codon. Thus, 9 nucleotides (GAGGAG**TAG**) in total were inserted between the second and third codon of the EGFP gene. Another reason that nine nucleotides were inserted (instead of three) is that the insertion of additional nucleotides allows PCR to be used to verify integration of both the stop-EGFP gene and EBFP gene into the mouse genome. For example, genomic PCR using an appropriate primer set will yield a 100 bp PCR product from the stop-EGFP gene and a 91 bp product from the wild-type EBFP gene. PCR products with that length difference can be easily separated on a 4% agarose gel (see section III-6).

3. Abrogating a possible immune response to EGFP revertants

If revertant EGFP were first expressed via a mutation in the adult stop-EGFP mouse, the protein might be recognized as a foreign antigen by the immune system, and cells expressing the revertant EGFP killed by cytotoxic T lymphocytes (CTL). Although it has

Figure 2-6. DNA sequences of the EGFP gene, the stop-EGFP gene, and the EBFP gene and translated amino acid sequences.

DNA sequences of the EGFP gene and translated amino acid sequences are shown. To create the stop-EGFP gene, 9 nucleotides including a stop codon (GAGGAG**TAG**) were inserted immediately downstream of codon GTG (in bold) (see section III-2). The EBFP gene was created by a site-directed mutagenesis at codon 67 in the EGFP gene that changed T at codon 67 in the EGFP gene to C (see section III-1). When the EGFP gene was artificially created from the wild-type GFP gene to optimize gene expression in mammalian cells and increase fluorescence intensity (Cormack et al., 1996; Yang et al., 1996) an additional codon, GTG, was inserted immediately downstream of the initiation codon. Thus, codon 66 (TAC) in the wild-type GFP gene (which predominantly determines wavelengths of light for both excitation and emission) corresponds to codon 67 in the EGFP gene. Peptides of LFTGVVPIL (codon 8 – 16), NYNSHNVYI (codon 145 – 153), and HYLSTQSAL (codon 200 – 208) are three best candidates to bind to H2-K^d molecules (see section III-3 and Table 2-2).

(GAGGAG**TAG**)

ATG**GTG** AGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGAC
M **V** S K G E E **L F T G V V P I L** V E L D
1 5 10 15 20

GGCGACGTAAACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTAC
G D V N G H K F S V S G E G E G D A T Y
21 25 30 35 40

GGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACC
G K L T L K F I C T T G K L P V P W P T
41 45 50 55 60

(**CAC**)

CTCGTGACCACCCTGACC**TAC**GGCGTGCAGTGCTTCAGCCGCTACCCCGACCACATGAAG
L V T T L T **Y** G V Q C F S R Y P D H M K
61 65 70 75 80

CAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATCTTC
Q H D F F K S A M P E G Y V Q E R T I F
81 85 90 95 100

TTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTG
F K D D G N Y K T R A E V K F E G D T L
101 105 110 115 120

GTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCAC
V N R I E L K G I D F K E D G N I L G H
121 125 130 135 140

AAGCTGGAGTACAACTACAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAAC
K L E Y **N Y N S H N V Y I** M A D K Q K N
141 145 150 155 160

GGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCC
G I K V N F K I R H N I E D G S V Q L A
161 165 170 175 180

GACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAACCAC
D H Y Q Q N T P I G D G P V L L P D N **H**
181 185 190 195 200

TACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTC
Y L S T Q S A L S K D P N E K R D H M V
201 205 210 215 220

CTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA
L L E F V T A A G I T L G M D E L Y K -
221 225 230 235

been presumed in some mouse studies that EGFP elicits a very weak, or no immune response, a strong CTL response against EGFP-expressing cells has been detected in the BALB/c strain and a moderate CTL response in the C57BL/6 strain (Gambotto et al., 2000; Stripecke et al., 1999). The possibility of an immune response elicited by revertant EGFP can therefore not be neglected. Because EBFP differs from EGFP by only a single amino acid (that is encoded by codon 67), transgenic mice exposed to EBFP from conception can be expected to have no immune response to revertant EGFP.

As a host defense mechanism, the immune system identifies and eliminates cells expressing foreign proteins (due to viral infection or oncogenesis, for example). Major histocompatibility complex (MHC) class I molecules bind oligopeptide fragments derived from a cell's expressed proteins and display them on the cell surface for surveillance by the immune system (Rock and Goldberg, 1999; York and Rock, 1996). The immune system is nonreactive to MHC class I-presented peptides derived from normal cellular proteins because of self-tolerance. However, when nonnative proteins are expressed in cells, foreign peptides will be presented on the surface of the cells, which leads to the stimulation of CTL to kill the cells expressing the foreign proteins. In mouse, three types of MHC class I molecules can be generated from three MHC gene loci, *H-2K*, *H-2D*, and *H-2L* (Sher et al., 1985). In the BALB/c strain (harboring the MHC haplotype $H\text{-}2^d$) three different types of MHC class I molecules encoded by the loci are designated $H2\text{-}K^d$, $H2\text{-}D^d$, and $H2\text{-}L^d$, respectively. On the other hand, in the C57BL/6 strain (harboring the MHC haplotype $H\text{-}2^b$) only two types of MHC class I molecules, $H2\text{-}K^b$ and $H2\text{-}D^b$ can be generated because of the absence of *H-2L* gene in the strain (Sher et al., 1985).

Peptides that bind to MHC class I molecules are usually 8 – 10 amino acids long (Rock and Goldberg, 1999). Depending on a type or an allelic variant of MHC class I molecules, MHC class I-presented peptides have a characteristic pattern of amino acid residues at two or three specific positions (Rammensee et al., 1995). For example, in the BALB/c strain H2-K^d molecules preferentially bind 9 amino acids long peptides containing tyrosine at the second and isoleucine or leucine at the ninth residue from the N-terminal of the peptides (Falk et al., 1991; Rammensee et al., 1995). On the other hand, peptides binding to H2-L^d molecules usually contain proline at the second and phenylalanine or leucine at the ninth residue from the N-terminal (Corr et al., 1992).

Characteristic patterns of peptides preferentially binding to *H-2*^{d} and *H-2*^{b} haplotype MHC molecules allow one to predict peptide regions of a protein that are likely to bind to a given type of MHC class I molecules in the BALB/c and C57BL/6 strains (the genetic backgrounds of the transgenic lines we developed, see section III-6). The prediction can be facilitated using an epitope prediction program such as MAPPP (MHC class I antigenic peptide processing prediction) which is available at http://www.mpiib-berlin.mpg.de/MAPPP/ (Hakenberg et al., 2003). Using the MAPPP program, candidate peptides derived from EBFP, EGFP, and EGFP revertants were predicted that could bind to MHC class I molecules in the BALB/c and C57BL/6 strains. The three best candidates to bind to each type of MHC class I molecules in the BALB/c (*H-2*^{d}) and C57BL/6 (*H-2*^{b}) strains are listed in Table 2-2 and 2-3, respectively. For example, in the BALB/c strain peptides of LFTGVVPIL (codon 8 – 16), NYNSHNVYI (codon 145 – 153), and HYLSTQSAL (codon 200 – 208) derived from EGFP are three candidates that are most likely to bind to H2-K^d molecules (Table 2-2 and Figure 2-6).

Table 2-2. Peptides that likely bind to MHC class I molecules in the BALB/c strain.

Three best candidate peptides derived from EBFP, EGFP, and EGFP revertants that are predicted to bind to each type of MHC class I molecules in the BALB/c strain (*H-2*^{d} haplotype). An epitope prediction program, MAPPP (MHC class I antigenic peptide processing prediction available at http://www.mpiib-berlin.mpg.de/MAPPP/) was used.

i) **H2-D^d** – No epitopes predicted.

ii) **H2-K^d**

EBFP	**EGFP**	**EGFP Revertants**
HYLSTQSAL (codon 200 – 208)	HYLSTQSAL (codon 200 – 208)	HYLSTQSAL (codon 203 – 211)
NYNSHNVYI (codon 145 – 153)	NYNSHNVYI (codon 145 – 153)	NYNSHNVYI (codon 148 – 156)
LFTGVVPIL (codon 8 – 16)	LFTGVVPIL (codon 8 – 16)	LFTGVVPIL (codon 11 – 19)

(iii) **H2-L^d**

EBFP	**EGFP**	**EGFP Revertants**
TPIGDGPVL (codon 187 – 195)	TPIGDGPVL (codon 187 – 195)	TPIGDGPVL (codon 190 – 198)
DPNEKRDHM (codon 211 – 219)	DPNEKRDHM (codon 211 – 219)	DPNEKRDHM (codon 214 – 222)
MVSKGEELF (codon 1 – 9)	MVSKGEELF (codon 1 – 9)	**EX*SKGEELF** (codon 4 – 12)

X* denotes an amino acid encoded from the site of the premature stop codon, TAG after undergoing a revertant mutation (e.g., lysine, glutamine, etc). See Figure II-1 for amino acids that can be encoded by revertant mutations at the premature stop codon.

Table 2-3. Peptides that likely bind to MHC class I molecules in the C57BL/6 strain.

Three best candidate peptides derived from EBFP, EGFP, and EGFP revertants that are predicted to bind to each type of MHC class I molecules in the C57BL/6 strain (*H-2*^{b} haplotype). An epitope prediction program, MAPPP (MHC class I antigenic peptide processing prediction available at http://www.mpiib-berlin.mpg.de/MAPPP/) was used.

i) **H2-D^b**

EBFP	EGFP	EGFP Revertants
DTLVNRIEL (codon 118 – 126)	DTLVNRIEL (codon 118 – 126)	DTLVNRIEL (codon 121 – 129)
DGDVNGHKF (codon 20 – 28)	DGDVNGHKF (codon 20 – 28)	DGDVNGHKF (codon 23 – 31)
YNSHNVYIM (codon 146 – 154)	YNSHNVYIM (codon 146 – 154)	YNSHNVYIM (codon 149 – 157)

(ii) **H2-K^b**

EBFP	EGFP	EGFP Revertants
GDATYGKL (codon 36 – 43)	GDATYGKL (codon 36 – 43)	GDATYGKL (codon 39 – 46)
ATYGKLTL (codon 38 – 45)	ATYGKLTL (codon 38 – 45)	ATYGKLTL (codon 41 – 48)
GEELFTGV (codon 5 – 12)	GEELFTGV (codon 5 – 12)	GEELFTGV (codon 8 – 15)

All candidate peptides listed in Table 2-3 are located in regions where the amino acid sequences of EBFP, EGFP, and EGFP revertants are identical, strongly suggesting that EGFP revertants will be recognized as self-proteins in the C57BL/6 strain stop-EGFP mouse exposed to EBFP from conception. The differences in amino acid sequences among EBFP, EGFP, and EGFP revertants lie in two regions of the proteins; the N-terminal region (due to the 9 nucleotides inserted in the stop-EGFP gene) and the position of the amino acid encoded by codon 67 (see sections III-1 and III-2). Those two regions are not included in candidate peptides which predictably bind to MHC class I molecules in the C57BL/6 strain.

In the BALB/c strain, the peptide corresponding to amino acid residue 4 – 12 of EGFP revertants could bind to H2-L^d molecules (Table 2-2). The peptide region of EGFP revertants includes the amino acid residue that is encoded from the site of the premature stop codon after the site undergoes a revertant mutation. Because the peptide region of EGFP revertants is different from that of EBFP, it is speculated that EGFP revertants might elicit immune response via H2-L^d molecule-mediated CTL response in the stop-EGFP mouse generated using the BALB/c strain even if the mouse is exposed to EBFP from conception.

4. Construction of plasmids

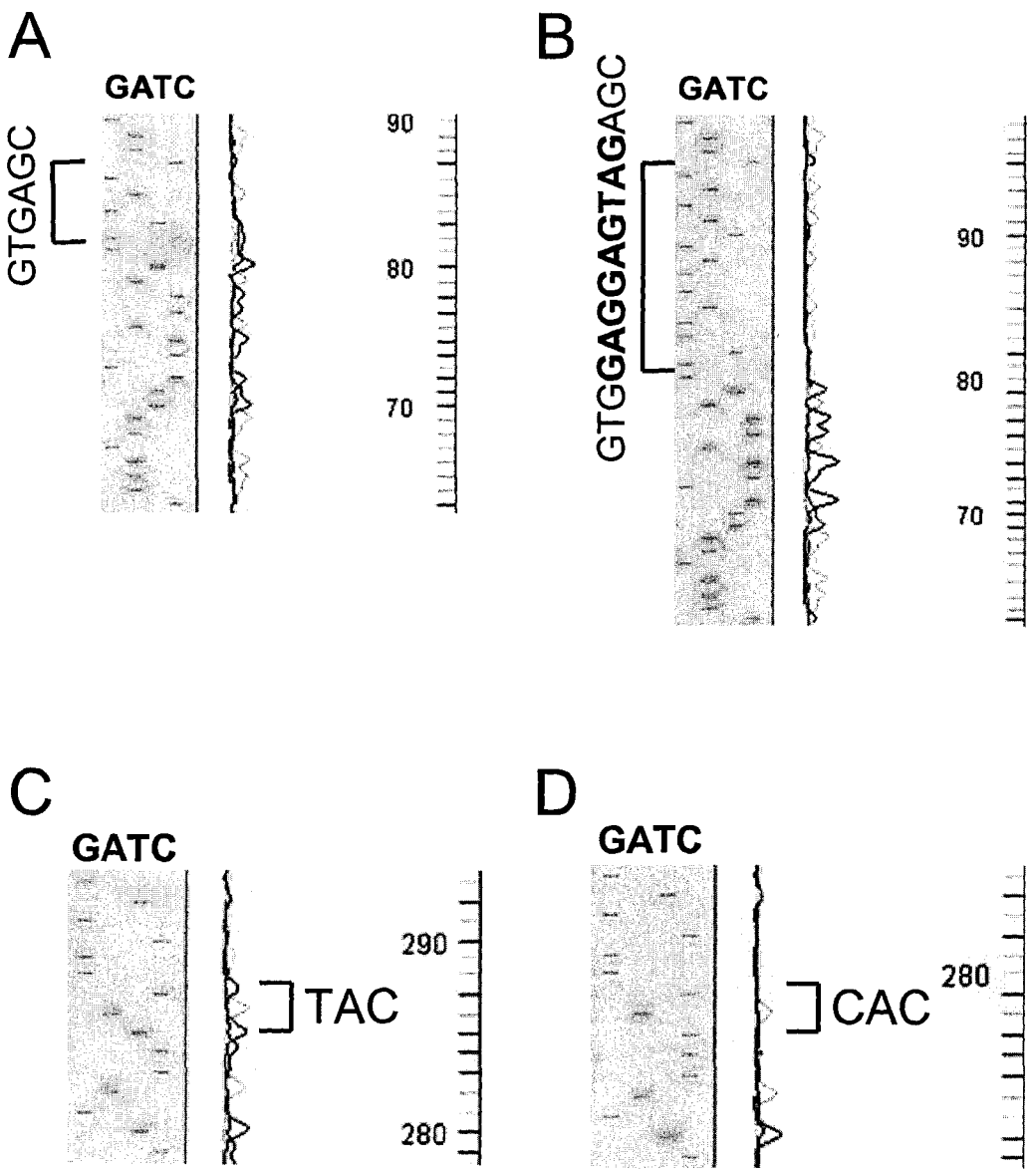
To allow the stop-EGFP system to be applied to diverse tissues, the stop-EGFP gene should be under the control of a ubiquitous promotor that can direct expression of the gene in every tissue. Okabe et al. constructed a plasmid in which the wild-type EGFP gene was placed under the controls of the β-actin promotor (referred to as pCX-EGFP) (Okabe et al., 1997). They showed that transgenic mice carrying the plasmid (referred to

as "Green mice") expressed the transgene ubiquitously by using direct imaging of diverse tissues and organs illuminated with the appropriate wavelength of light. Variegation of transgene expression was not detected in most transgenic lines when splenic and thymic T lymphocytes from mice were collected and analyzed by fluorescence activated cell sorting (FACS) (Okabe et al., 1997). We have obtained the plasmid pCX-EGFP and replaced the wild-type EGFP gene with a premature stop codon-containing EGFP gene (see section II-1). Substitution of the stop-EGFP gene for the wild-type EGFP gene in the construct was confirmed by DNA sequencing (Figure 2-7). The resulting construct is referred to as pCX-stop-EGFP (Figure 2-8).

The pCX-EBFP construct was generated from pCX-EGFP via a site-directed mutagenesis at codon 67 in the EGFP gene (equivalent to codon 66 of the GFP or BFP gene) (Figure 2-7). The mutagenesis changed T at codon 67 in the EGFP gene to C which creates a codon that encodes histidine instead of tyrosine (see section I-1, Chapter 1). A substitution at codon 146 of the EBFP gene (equivalent to codon 145 of the BFP gene) can increase the intensity of blue fluorescence from EBFP roughly 2-fold (Heim and Tsien, 1996). However, the substitution at codon 146 was not included in our construct because it could lead to a decrease in the green fluorescence intensity of revertant EGFP when a revertant EGFP gene is generated from the EBFP gene via mutation from C to T at codon 67. Detection of mutations in the stop-EGFP mouse will be carried out via imaging of green fluorescence from cells, thus maintaining intensity of green fluorescence from revertant EGFP is our main concern. As mentioned above, the main reason that the wild-type EBFP gene is utilized in the stop-EGFP system is that the EBFP

Figure 2-7. Sequencing data of the wild-type EGFP gene, the stop-EGFP gene, and the wild-type EBFP gene.

DNA sequences from the second codon of the wild-type EGFP gene (A) and the stop-EGFP gene (B) are shown. The stop-EGFP gene was generated by introducing nine nucleotides including a premature stop codon (GAGGAGTAG) between the second codon (GTG) and the third codon (AGC) of the EGFP gene. The wild-type EBFP gene was generated from the wild-type EGFP gene by changing TAC at codon 67 of the EGFP gene (C) to CAC (D) via a site-directed mutagenesis.



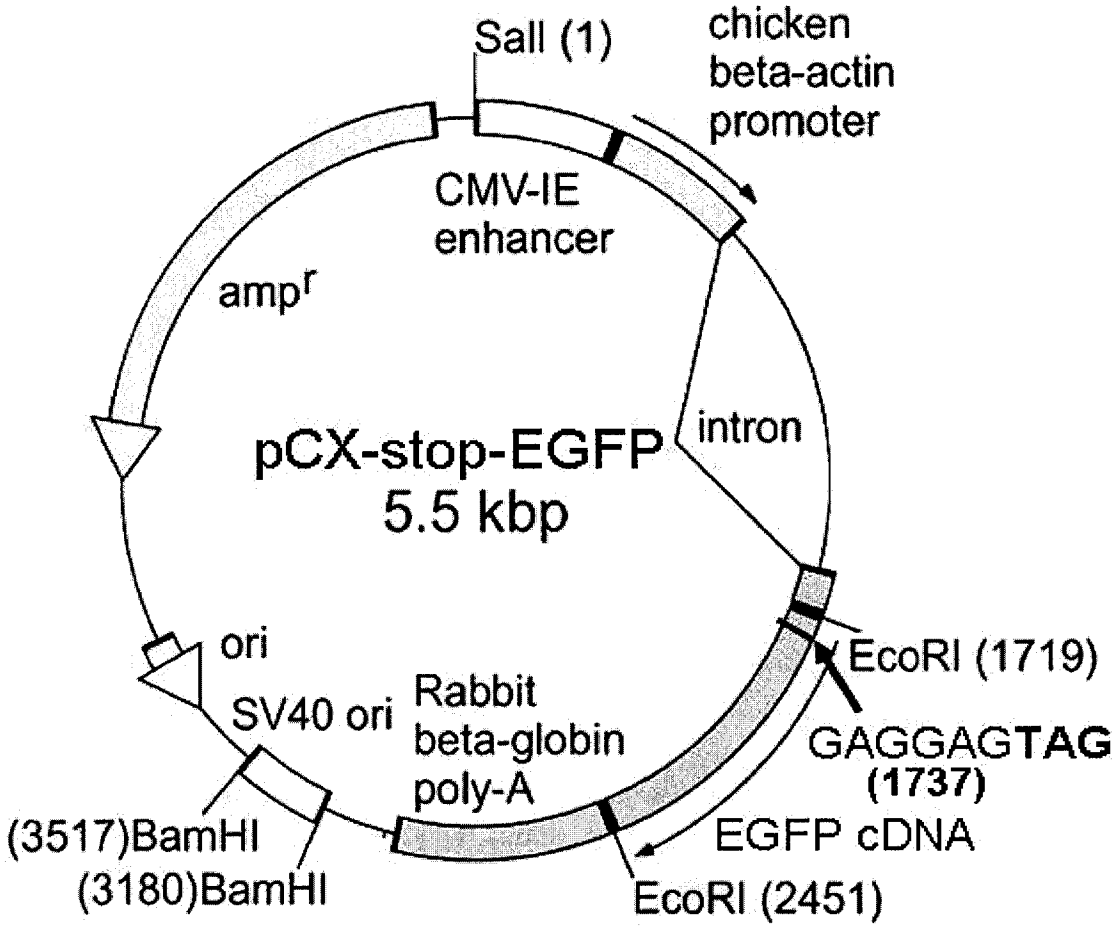


Figure 2-8. The genetic map of pCX-stop-EGFP.

Nine nucleotides including a premature stop codon (GAGGAG**TAG**) were inserted into EGFP cDNA. The plasmid was modified from pCX-EGFP (Okabe et al., 1997).

gene provides a target for GC → AT transition mutation that the stop-EGFP gene cannot detect at the premature stop codon, TAG.

5. Experimental verification of the neutrality of the premature stop codon site

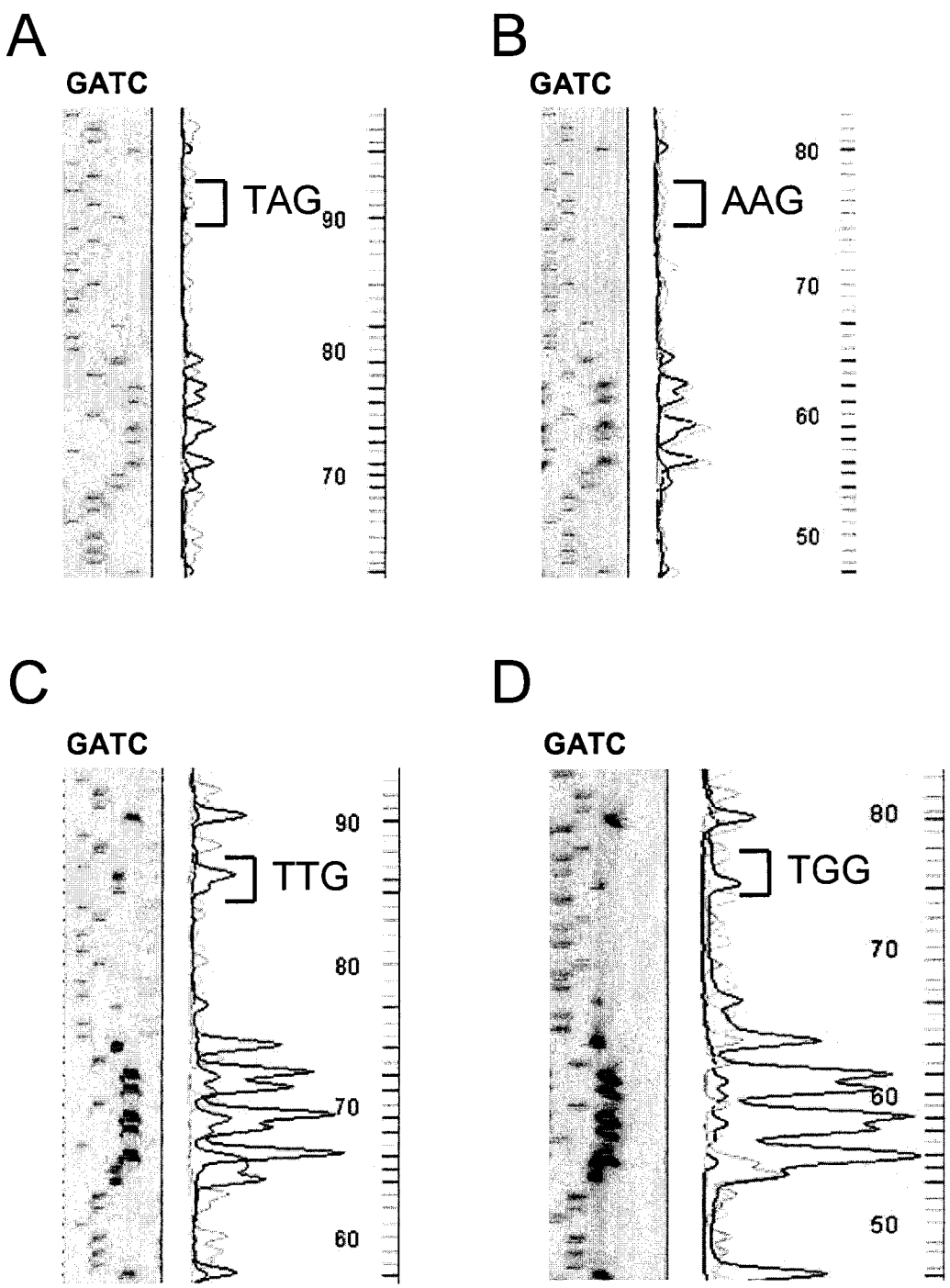
For the stop-EGFP system to function as an effective cell fate mapping and mutation detection system, the premature stop codon in the stop-EGFP gene should terminate the process of protein synthesis efficiently at the site, otherwise background green fluorescence from non-mutant cells might make it difficult to identify green fluorescent mutant cells and their descendent cell lineages. In addition, the premature stop codon should be located at a neutral codon site such that any nonsynonymous mutations at the premature stop codon will result in expression of functional EGFP revertants. Using an in vitro cell culture system in which the stop-EGFP gene and revertant forms of the gene were transiently expressed, we tested whether the premature stop codon site in the stop-EGFP gene satisfies both requirements.

To test the intensities of green fluorescence of EGFP revertants, we artificially created several revertant forms of the stop-EGFP gene using PCR (see section II-2 and Figure 2-9). Nine DNA substitutions in total can be generated at the premature stop codon from which seven different amino acids can be encoded (see Figure 2-1).

Three possible revertant forms of the stop-EGFP gene, containing AAG (encoding lysine), TTG (encoding leucine), or TGG (encoding tryptophan) in place of the premature stop codon (TAG), were selected to compare the intensities of green fluorescence of the revertant EGFP with that of the wild-type EGFP (see Figure 2-9). The three revertant forms are expected to be found most frequently when the stop-EGFP system is treated with the mutagen, ENU (*N*-ethyl-*N*-nitrosourea) which was used in our

Figure 2-9. Generation of revertant forms of the stop-EGFP gene.

Three possible revertant forms of the stop-EGFP gene, containing AAG, TTG, or TGG in place of the premature stop codon (TAG), were generated using PCR. DNA sequences are shown at the site of the premature stop codon (i.e., codon 5) for the stop-EGFP gene (A), the AAG-revertant (B), the TTG-revertant (C), and the TGG-revertant (D).



cell fate mapping (Chapter 3) and mutation studies (Chapter 4). The most common mutations induced by ENU are AT → TA transversions with the second most common being AT → GC transitions (Noveroske et al., 2000; Popp et al., 1983). Thus, if a mutation at the premature stop codon is induced by treatment with ENU, the most commonly observed types of revertant codons from TAG are expected to be AAG, TTG, CAG, and TGG.

Green fluorescence from cells transiently expressing each revertant form of the stop-EGFP gene was comparable to that from cells transiently expressing the wild-type EGFP gene (Figure 2-10). The finding strongly suggests that the site of the premature stop codon is a functionally neutral codon. Green fluorescence from cells expressing the stop-EGFP gene (Figure 2-10E) was minimal and comparable to that of non-transfected cells (data not shown) and cells transfected with the wild-type EBFP gene (Figure 2-10F) demonstrating that the premature stop codon efficiently terminates protein synthesis. The data from the cell culture experiment suggests that the site we chose to introduce the premature stop codon meets the two requirements described above.

To quantitatively compare the intensity of green fluorescence from cells expressing each construct, FACS analysis was performed (Figure 2-11). No significant difference in fluorescence intensity was detected between cells expressing the wild-type EGFP gene and cells transfected with the three revertant forms of the stop-EGFP genes. Green fluorescence from those cells show a broad range of fluorescence intensity (see Figure 2-11E to H). We speculate the difference in the fluorescence intensity within the same group could result from differences in gene copy numbers among cells or differences in health among cells. Fluorescence from cells expressing the stop-EGFP

Figure 2-10. Transient expression of revertant forms of the stop-EGFP gene in the NIH3T3 cell line.

Three possible revertant forms of the stop-EGFP gene, each containing AAG (encoding lysine), TTG (encoding leucine) or TGG (encoding tryptophan) at the site of the premature stop codon, were generated and used for the transfection experiment (see text). All three revertants, AAG-revertant (A), TTG-revertant (B) and TGG-revertant (C), expressed functional proteins, which were as strongly fluorescent as the wild-type EGFP (D). The result strongly supports the neutrality of the position of the premature stop codon. Green fluorescence from cells transfected with the stop-EGFP gene (E) is comparable to that from negative controls, the EBFP gene-transfected cells (F) and non-transfected cells (data not shown).

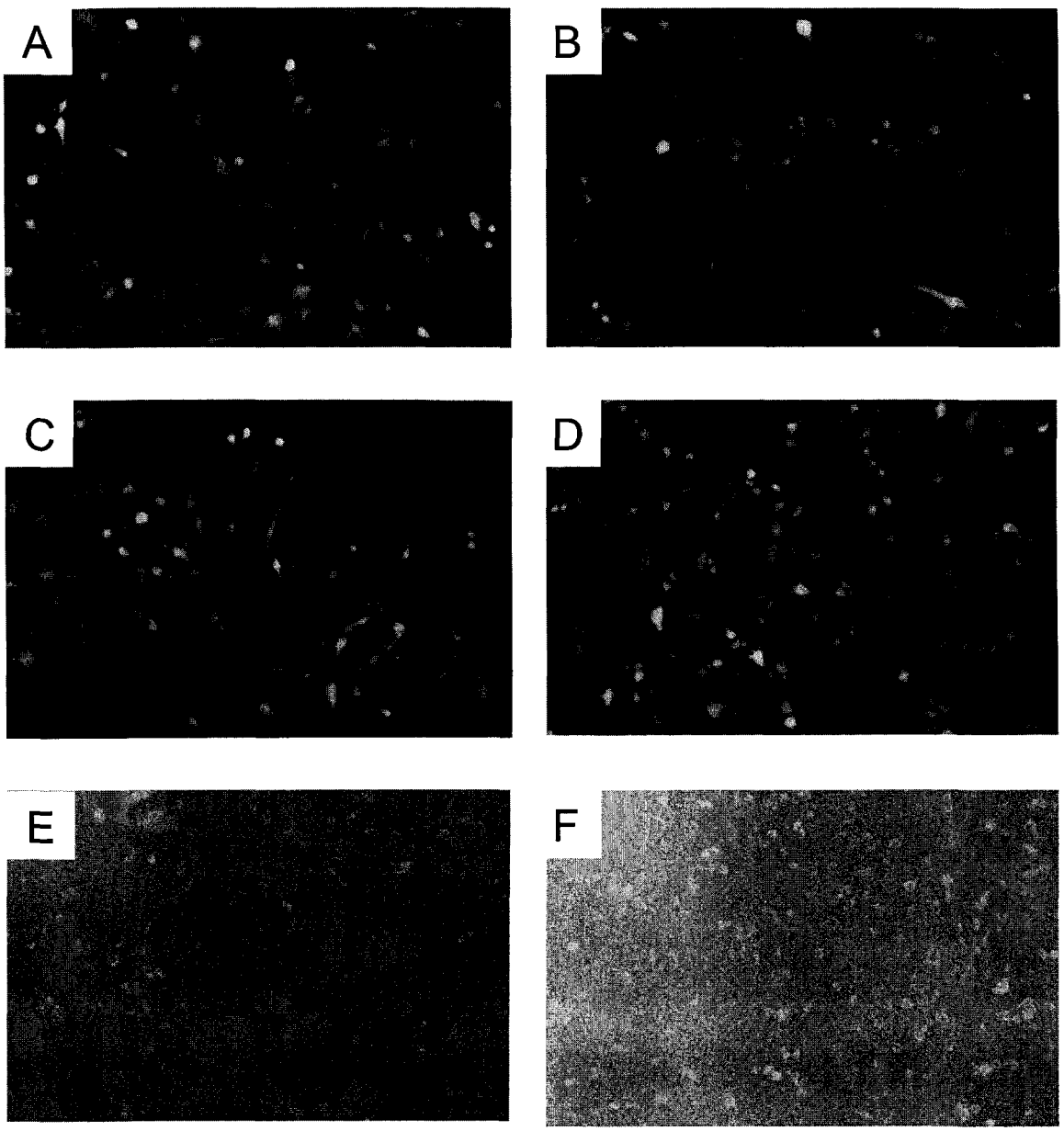
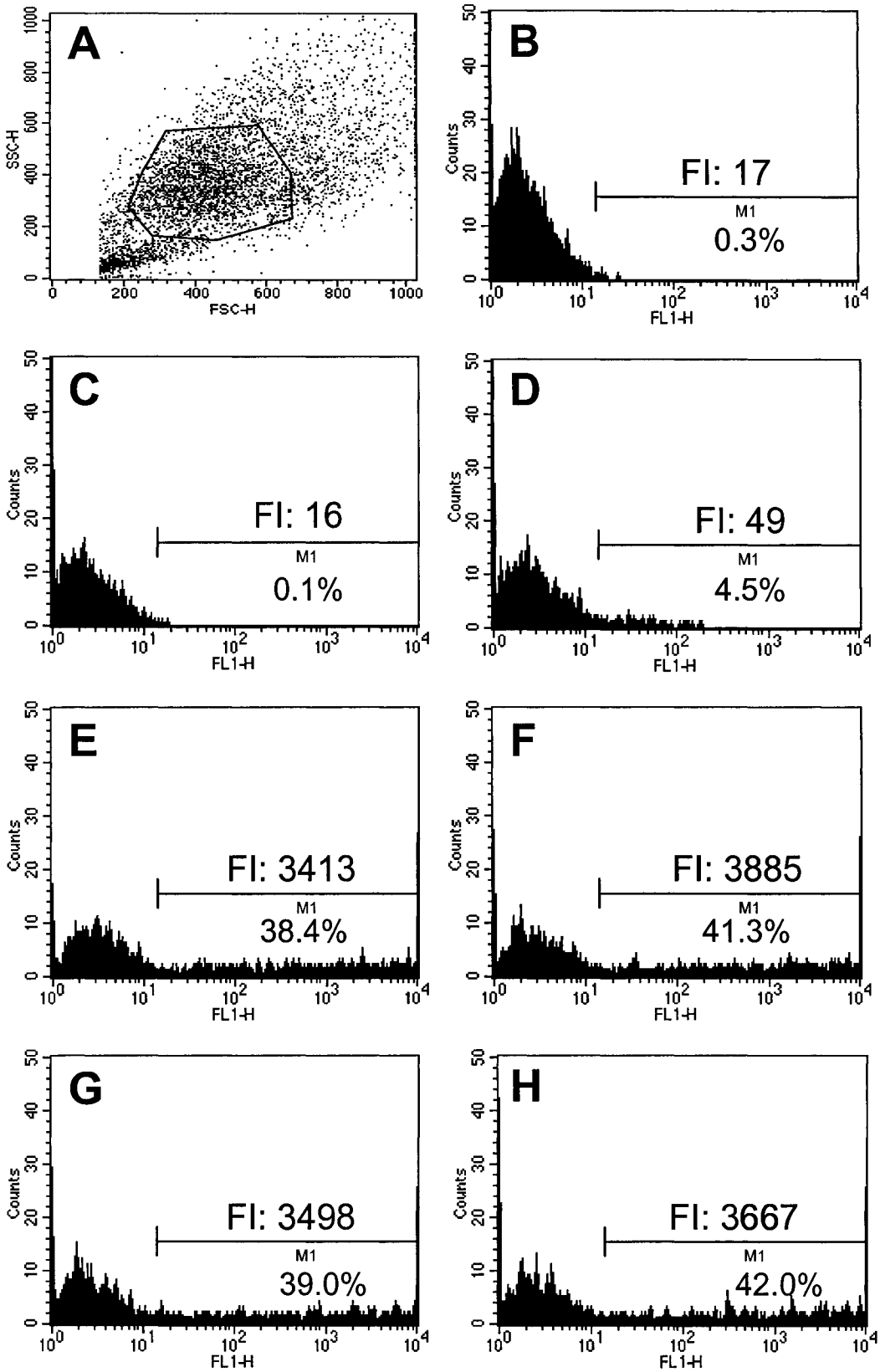


Figure 2-11. FACS analysis of cells transfected with each construct.

(A) Forward-scatter (FSC-H) vs. side-scatter (SSC-H) profile. The gate was set as indicated. (B) – (H) Green fluorescence (FL1-H) vs. cell counts. The marker M1 defines a range of fluorescence that excludes at least 99% of non-transfected cells (see Panel B). The average intensity of fluorescence from cells whose fluorescence intensities are within the range of M1 is indicated in each panel. (B) Non-transfected cells, (C) cells transfected with the wild-type EBFP gene, (D) cells transfected with the stop-EGFP gene, (E) cells transfected with the wild-type EGFP gene. (F) – (H) cells transfected with revertant forms of the stop-EGFP genes which contain AAG (Panel F), TTG (Panel G) or TGG (Panel H), respectively at the site of the premature stop codon. FI stands for fluorescent intensity. The percentage (%) of cells that the marked region (M1) contains is shown in each panel.



gene was minimal and comparable to that from non-transfected cells and cells expressing the wild-type EBFP gene (see Figure 2-11). Fluorescence beyond the background level of autofluorescence was occasionally detected from cells transfected with the stop-EGFP gene (4.5% of total cells analyzed). The fluorescent signals from those cells are likely due to inefficient termination of the translation process at the premature stop codon of the stop-EGFP gene, especially in cells carrying a high copy of the stop-EGFP gene. The mean intensity of fluorescence from those fluorescent cells, however, is about 70-fold less than that from cells expressing the wild-type EGFP gene or revertant forms of the stop-EGFP gene (Figure 2-11). Thus, it is unlikely that the background level of green fluorescence from cells expressing the stop-EGFP gene would interfere with detection of green fluorescent mutant cells in the stop-GFP system.

6. Generation of transgenic mice

Fertilized oocytes from two different inbred strains of mice (C57BL/6 and FVB/N) were used in order that transgenic mice could be generated in two different genetic backgrounds. Before the microinjection, the stop-EGFP gene and the wild-type EBFP gene were mixed as a ratio of 4 to 1 in the expectation that more copies of the stop-EGFP gene to be integrated into the mouse genome than those of the EBFP gene. Each copy of the stop-EGFP gene provides a mutational target of three nucleotides in comparison to one nucleotide in the case of the EBFP gene, thus creating a transgenic mouse with relatively more copies of the stop-EGFP gene is desirable.

In the case of the C57BL/6 genetic background, the pronuclei of 190 fertilized eggs were microinjected with the transgenes and then transplanted in oviducts of pseudopregnant foster mothers. In total, 16 mice were born, of which three appeared

transgenic when tested for the transgenes by an initial genomic PCR screening (Table 2-4 and Figure 2-12A).

Table 2-4. Generation of transgenic mice with the C57BL/6 genetic background.

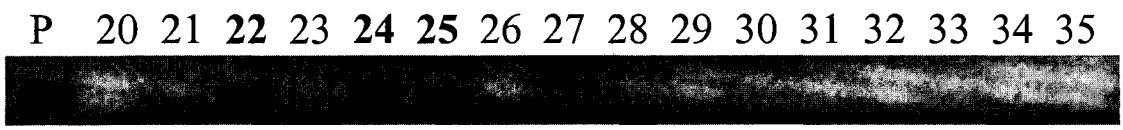
strain	# of eggs microinjected	# of births	# of transgenic founders	# of double-transgenic founders
C57BL/6	190	16	3	2

Using a primer set that specifically generates 100 bp PCR product from the stop-EGFP gene and 91 bp PCR product from the EBFP gene, transgenic mice were tested for the presence of both transgenes. PCR analysis revealed that only one transgenic mouse (#25) carried both the stop-EGFP and EBFP gene and the other two transgenic mice (#22 and 24) carried only the stop-EGFP gene (Figure 2-12B). As a positive control, plasmid pCX-stop-EGFP was mixed with pCX-EBFP with a ratio of 1 to 1 and used for PCR. Although the same copy of the two genes was used for PCR, resulting PCR products show that the stop-EGFP gene is preferentially amplified (see Figure 2-12B lane P). If the copy number of the EBFP gene were, for example, six times lower than that of the stop-EFP gene in transgenic mice, it would be hard to detect the presence of the EBFP gene by use of this PCR method. Thus, the possibility that the EBFP gene was not detected due to the sensitivity of the PCR method used cannot be ruled out. To further test for the presence of the EBFP gene, we performed an EBFP gene-specific PCR test.

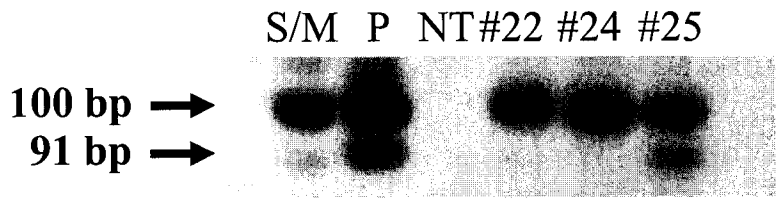
Figure 2-12. Genotyping of mice by PCR with the C57BL/6 genetic background.

(A) Transplantation of microinjected embryos produced 16 mice. PCR screening revealed that mice #22, 24, and 25 were transgenic. P denotes a positive control (plasmid DNA harboring the stop-EGFP gene, pCX-stop-EGFP). (B) Presence of both transgenes in the three transgenic founders #22, 24, and 25 was tested by PCR which generates different sizes of PCR products from the two transgenes (see Materials and Methods). S/M denotes a size marker and NT denotes a non-transgenic mouse. As a positive control, pCX-stop-EGFP (plasmid harboring the stop-EGFP gene) was mixed with pCX-EBFP (plasmid harboring the EBFP gene) with a ratio of 1 to 1 and then used for the PCR (denoted as P). (C) The presence of the EBFP gene in the transgenic founders was further tested by PCR that specifically generated PCR products from the EBFP gene without amplifying the stop-EGFP gene. S and B denote pCX-stop-EGFP and pCX-EBFP, respectively. NT denotes a non-transgenic mouse. Gels were stained in ethidium bromide. Black and white images were inverted using Adobe® Photoshop® (4.0 LE) for better resolution.

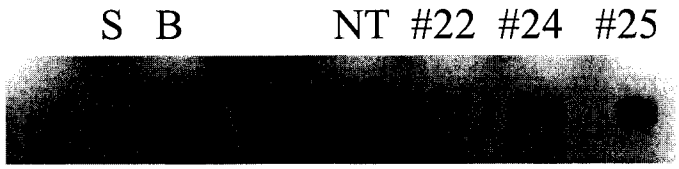
A



B



C



A primer was designed that binds to the EBFP gene but not to the stop-EGFP gene because of the 9 nucleotides insertion (GAGGAGTAG) (see Figure 2-6). PCR using the primer specifically generated PCR products from the EBFP gene without amplifying the stop-EGFP gene and verified that transgenic mouse #24 also carried the EBFP gene (Figure 2-12C). The PCR test further verified that transgenic mouse #22 did not carry the EBFP gene.

The two transgenic founders carrying both transgenes (i.e., #24 and 25 mice) were bred with C57BL/6 wild-type inbred mice to produce transgenic offspring. However, transgenic mouse #24 died during pregnancy and produced no offspring. Founder #25 produced transgenic offspring which were subsequently used for more characterizations of the transgenic line.

In the case of the FVB/N genetic background, the pronuclei of 110 fertilized eggs were microinjected with the transgenes, which produced 24 mice (Table 2-5). An initial genomic PCR screening revealed three mice were transgenic (see Figure 2-13A). To distinguish them from mice with the genetic background of the C57BL/6 strain, the three transgenic mice with the FVB/N genetic background are labeled as F#11, F#19, and F#24. When the presence of both the stop-EGFP gene and the EBFP gene was tested using the aforementioned primer set, only two transgenic founders (F#11 and F#19) carried both transgenes while one transgenic founder (F#24) carried only the EBFP gene (Figure 2-13B). Founders F#11 and F#19 were bred with wild-type BALB/c inbred mice to produce transgenic offspring. Although the FVB/N strain has been used extensively in transgenic research because of superior reproductive performance and prominent pronuclei which facilitate microinjection of transgenes (Taketo et al., 1991), the strain

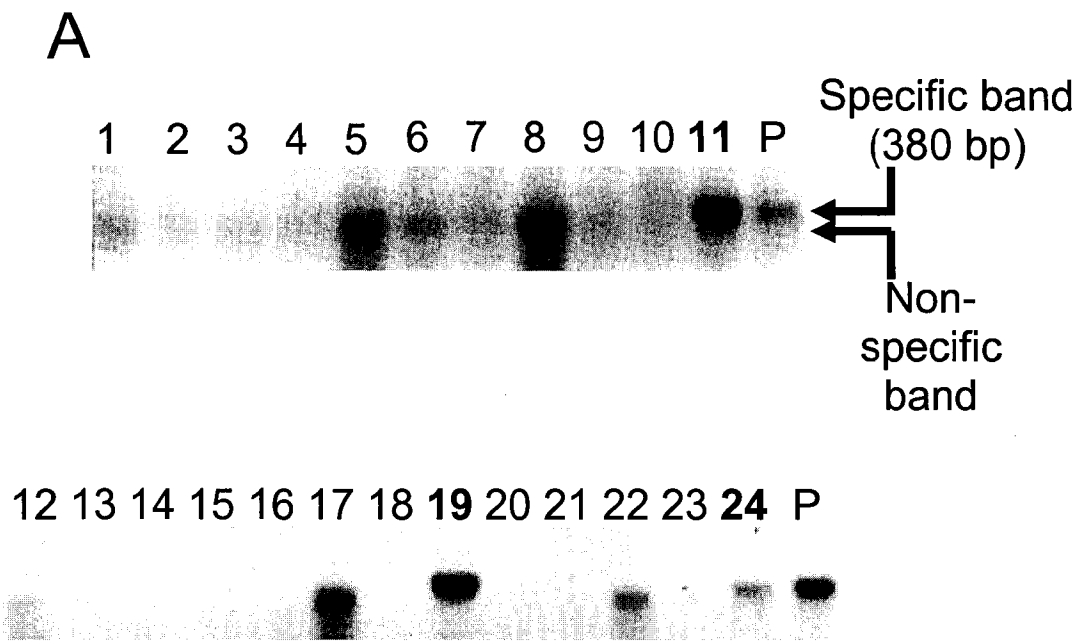
has several health issues (Goelz et al., 1998; Mahler et al., 1996). Thus, a conversion of the genetic background of the transgenic founders from FVB/N to BALB/c was undertaken by successive breeding with the latter strain.

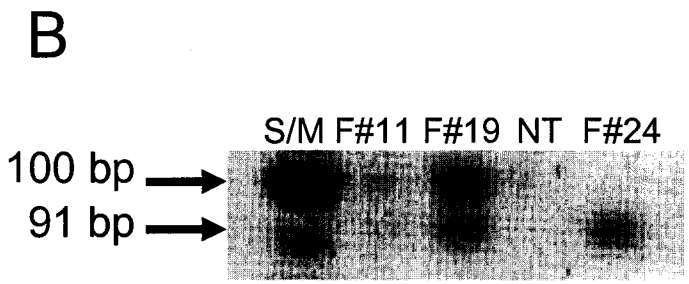
Table 2-5. Generation of transgenic mice with the FVB/N genetic background.

strain	# of eggs microinjected	# of births	# of transgenic founders	# of double-transgenic founders
FVB/N	110	24	3	2

Figure 2-13. Genotyping of mice by PCR with the FVB/N genetic background.

(A) Implantation of microinjected embryos produced 24 mice. PCR screening revealed that mice #11, 19, and 24 were transgenic. To distinguish them from mice with the genetic background of the C57BL/6 strain, the three transgenic mice with the FVB/N genetic background are labeled as F#11, F#19, and F#24. P denotes a positive control (plasmid DNA harboring the stop-EGFP gene, pCX-stop-EGFP). (B) Presence of both transgenes in the three transgenic founders F#11, F#19, and F#24 was tested by PCR analysis which can differentiate the two transgenes by producing different sizes of PCR products. S/M and NT denote a size marker and a non-transgenic mouse, respectively. Gels were stained in ethidium bromide. Black and white images were inverted using Adobe® Photoshop® (4.0 LE) for better resolution.





Founder F#19 successfully transmitted the transgenes to offspring. Founder F#11 failed to transmit the transgenes to the next generation. Among 22 offspring of founder F#11, none were found to be transgenic when the presence of the transgenes was tested by genomic PCR. It is speculated that F#11 mouse is a mosaic transgenic mouse and only a subset of cells carry the transgenes (Palmiter et al., 1984). Although a transgene is usually integrated into a mouse chromosome before the first cell division of a fertilized oocyte (Palmiter and Brinster, 1986), it occasionally happens that transgenes are integrated after cell divisions of fertilized oocytes and the germ-line cells do not carry the transgenes (Wilkie et al., 1986).

From the microinjection experiment using fertilized oocytes of two different strains and subsequent breeding of transgenic founders, two transgenic lines (#25 and F#19) were established. To determine whether the stop-EGFP gene and EBFP gene cosegregate in both transgenic lines, the presence of the two transgenes in 12 transgenic mice of line #25 and 19 transgenic mice of line F#19 was tested by genomic PCR that produces different sizes of PCR products from the two transgenes (see above). The genomic PCR showed that all transgenic offspring tested (31 mice in total) carried both transgenes. The data strongly suggests that the two transgenes were integrated into a single chromosome locus in each transgenic founder and thus segregate together. This finding is consistent with previous studies that developed transgenic mice carrying two different transgenes (Jankowsky et al., 2001; Sola et al., 1998). When a transgene is integrated into the mouse genome, multiple copies of the transgene are commonly integrated into the same site on a chromosome in a tandem, head-to-tail array (Goldman et al., 2004; Palmiter and Brinster, 1986). This is equally true for integration of different

transgenes (Goldman et al., 2004). Two different genes, especially sharing very similar sequences, are highly likely to be integrated into the same site of a chromosome.

In total, 19 mice among the 45 offspring of founder #25 and 30 mice among the 68 offspring of founder F#19 were found to be transgenic when tested for the transgenes by genomic PCR (Table 2-6). Thus, the frequency at which offspring carry the transgenes is close to 50% for both transgenic lines (42% for line #25 and 44% for line F#19, see Table 2-6) suggesting the transgenes are inherited in a dominant Mendelian fashion (Table 2-7). Transgenic mice of both lines have not shown any sign of abnormality and discomfort, which is consistent with other studies using transgenic mice carrying the EGFP gene (Feng et al., 2000; Okabe et al., 1997). The transgenic mice typically survived more than a year (data not shown).

Table 2-6. Frequency that offspring carry the transgenes.

	Average litter size	No. offspring	M/F ratio	# of Tg offspring	Tg M/F ratio	Tg frequency*
Line #25	7.5	45	22/23	19	8/11	42%
Line F#19	8.5	68	39/29	30	18/12	44%

*Tg frequency was calculated from the number of Tg offspring divided by the number of offspring. Tg denotes Transgenic and M/F denotes Male/Female.

Table 2-7. Likelihood ratio test of the frequencies of transgene transmission.

	$2\log\Delta$ (P_0=3/4)	$2\log\Delta$ (P_0=1/2)	$2\log\Delta$ (P_0=1/4)	$\chi^2_{[1]}(\alpha=0.05)$
Line #25	21.72*	1.09	6.34*	3.84
Line F#19	29.29*	0.943	11.72*	3.84

* Null hypotheses (P_0=3/4 and P_0=1/4) are rejected because the log-likelihood values are significantly higher than the $\chi^2(\alpha = 0.05)$ value.

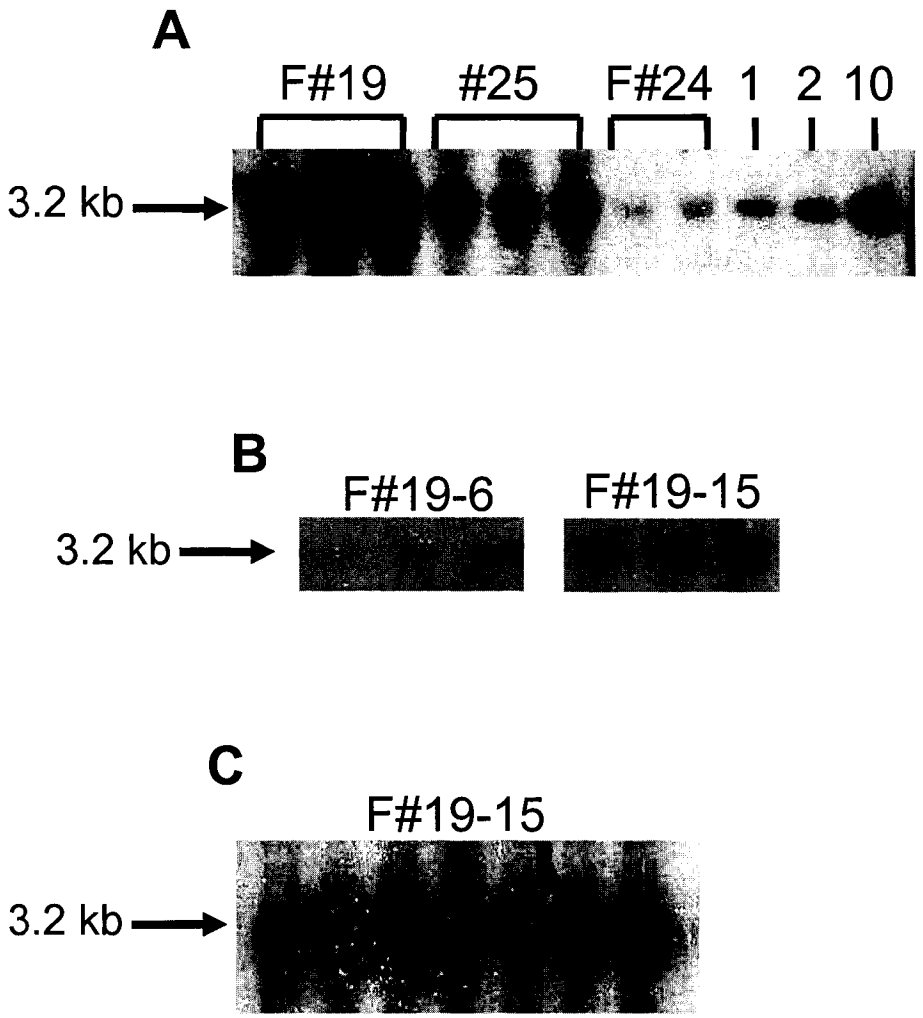
7. Southern blotting

Genomic DNA was collected from three transgenic offspring of each transgenic founder and used for Southern blot analysis to further verify transgene integration into the mouse genome and also roughly estimate the total combined copy number of both transgenes (the stop-EGFP gene and EBFP gene). The Southern blot analysis revealed that line F#19 carries more copies of the transgenes than line #25 (Figure 2-14A). Genomic DNA samples prepared from three different transgenic offspring of founder #25 all showed an identical band intensity at the expected size of 3.2 kb, implying that the transgenic offspring carry the same copy number of the transgenes.

However, the transgenic offspring of founder F#19 showed two different band intensities at the expected size of 3.2 kb (Figure 2-14A). One genomic sample showed a stronger band intensity at the size of 3.2 kb than the other two. These data suggest that some transgenic offspring of founder F#19 might carry more copies of the transgenes than others. To examine whether the pattern of different band intensities is observed extensively among transgenic offspring of founder F#19, we performed the Southern blot experiment again using more genomic DNA samples. Southern blot analysis using genomic DNA samples from 24 transgenic offspring of founder F#19 also revealed two distinct band intensities at 3.2 kb as detected in the previous experiment (6 out of the 24 genomic DNA samples showed a stronger band intensity, data not shown). Based on these observations, we speculated that two types of transgenic mice were generated from the same transgenic founder, one carrying more copies of the transgenes than the other. To establish both types of transgenic lines, we selected two transgenic offspring which carry high and low copy numbers of the transgenes (F#19-15 and F#19-6, respectively)

Figure 2-14. Southern blot analysis of genomic DNA of transgenic mice.

(A) Southern blot test shows that line F#19 and line #25 carry more than 10 copies of the transgenes. Transgenic offspring of founder F#19 showed two different band intensities at the size of 3.2 kb (the correct size of the transgenes injected). Offspring of transgenic founder F#24 (which carries only the EBFP gene, see Figure 2-13B) are expected to carry one, or two, copies of the transgene. Copy numbers of 1, 2 and 10 were used as references (see section II-7). (B) Southern blot analysis using genomic DNA collected from transgenic offspring of F#19-15 (carrying a high copy number of the transgenes) and F#19-6 mice (carrying a low copy number) shows no further variations in the copy numbers of the transgenes among offspring of the same parents. (C) Southern blot test using genomic DNA of 7 transgenic mice of line F#19-15 in successive generations. All samples tested show the same band intensity at the size of 3.2 kb.



and bred them with wild-type BALB/c inbred mice. Southern blot analysis performed using genomic DNA collected from transgenic offspring of F#19-15 and F#19-6 mice showed that there were no further variations in the copy numbers of the transgenes among transgenic offspring of the same parents (Figure 2-14B). We speculate that the transgenes integrated into a chromosome of founder F#19 had likely undergone gene rearrangement at the locus (such as transgene amplification or partial deletion of transgenes) during the first or second mitotic cell division of the fertilized oocyte (Shani, 1986) and thus at the site of transgene integration, two different copy numbers of the transgenes exist among the population of germ-line cells of founder F#19 (see Discussion). For example, transgenic offspring F#19-6 might have been generated from a germ cell carrying a lower copy number of the transgenes because of the deletion of some copies of the transgenes. As a result, we have established two transgenic sub-lines (e.g., line F#19-15 versus F#19-6) which are descended from the same transgenic founder, F#19 but carry different copy numbers of the transgenes.

Line F#19-15, which carries a high copy number of the transgenes, was subsequently used for studies of stem cell lineages and mutation research (see Chapters 3 and 4). The stop-EGFP mouse of line F#19-15 has been bred for four generations thus far and the transgenes have been consistently transmitted to subsequent generations as a Mendelian dominant trait. Additional gene rearrangement has not been detected in successive generations when the genomic DNA has been analyzed by Southern blotting (Figure 2-14C).

8. Determination of the copy numbers of the transgenes

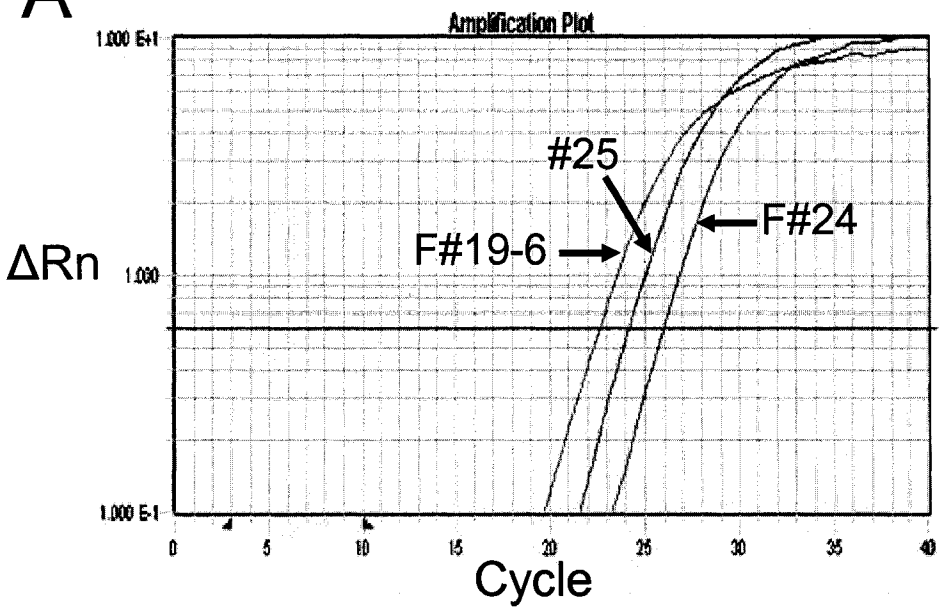
Although Southern blot analysis can roughly suggest the copy number of the transgenes, precise estimates of the copy numbers were difficult to predict simply by comparison with the references (Figure 2-14A). To better determine the total copy numbers of the transgenes, real-time PCR was performed (Figure 2-15). For a reaction containing genomic DNA of each transgenic line a C_T value, which is the number of cycles necessary to reach a threshold of reporter fluorescence emission, was identified and used to estimate the copy number of the transgenes in the transgenic line by comparing with the C_T value of the reference reaction containing genomic DNA of line F#24 (i.e., line carrying one or two copies of the transgenes, see Figure 2-14A). The estimated ratio of the copy number in each transgenic line to that in line F#24 is listed in Table 2-8. Assuming that line F#24 carries two copies of the EBFP gene (Figure 2-14A and also see Figure 2-13B), line F#19-15 is estimated to carry about 30 copies of the transgenes (the stop-EGFP gene and the EBFP gene together) and line F#19-6 is expected to carry about 20 copies of the transgenes.

Although the total copy numbers of the transgenes were estimated from Southern blotting or the quantitative PCR (i.e., real-time PCR), the individual copy numbers of the stop-EGFP gene and the EBFP gene in each transgenic line were still unknown. A competitive PCR can be performed to determine a relative ratio of the copy numbers of two genes (or mRNAs) with homologous sequences but with different sizes (Gilliland et al., 1990; Diviacco et al., 1992). Primers are designed to bind to the identical sequences of two competitor genes thus providing the same efficiency of primer binding. The advantage of a competitive PCR is that any predictable or unpredictable variable that

Figure 2-15. Amplification plots of the transgenes.

The normalized reporter fluorescence emission (Δ Rn) is plotted versus cycle number. (A) The real-time PCR shows that lines F#19-6 and #25 carry significantly higher copy numbers of the transgenes (the stop-EGFP gene and the EBFP gene together) than line F#24 which is expected to carry one or two copies of the transgenes (see Figure 2-14). (B) The PCR shows that the copy number of the transgenes in line F#19-15 is about 16 times higher than that of line F#24.

A



B

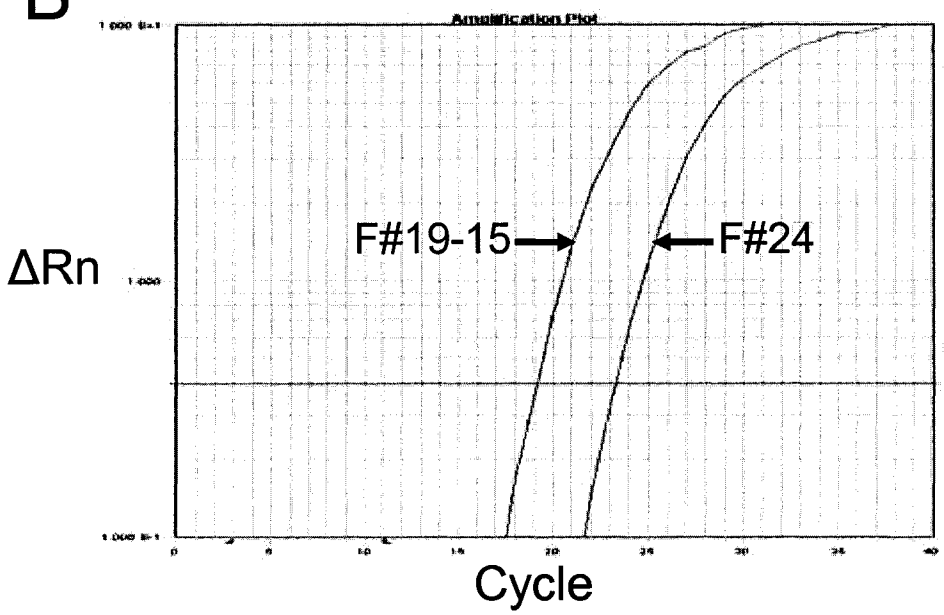


Table 2-8. Total copy number of the transgenes in each transgenic line estimated by Southern blot analysis and real-time PCR.

Transgenic line	Copy number estimated from Southern blotting*	Ratio of copy number estimated from real-time PCR**
Line F#24	1 – 2	1
Line #25	~ 10	3.88
Line F#19-6	>10	10.96
Line F#19-15	>>10	15.80

*Estimation based on the data shown in Figure 2-14A.

**The relative ratio of the copy number in each transgenic line to that in line F#24 was estimated from real-time PCR analysis.

affects PCR amplification has the same effect on both competitor genes, thus the final ratio of amplified products is expected to reflect the initial ratio of the two genes regardless of the number of amplification cycles.

Genomic DNA from each transgenic line was prepared and used for a competitive PCR to determine the ratio of the copy numbers of the two transgenes. A competitive PCR using pCX-stop-EGFP (plasmid containing the stop-EGFP gene) and pCX-EBFP (plasmid containing the EBFP gene) with a ratio of 1 to 1 showed that the stop-EGFP is more efficiently amplified than the EBFP gene and thus accurate estimation of the ratio of the copy numbers of the two transgenes could be difficult. To get a better estimation of the ratio, pCX-stop-EGFP was mixed with pCX-EBFP with ratios of 1 to 1, 4 to 3, 2 to 1, and 1 to 2 and then used for a competitive PCR to provide references for the ratios of the

copy numbers. Through comparison with the references, the ratios of the copy number of the stop-EGFP gene to that of the EBFP gene in transgenic lines F#19-15 and #25 were roughly estimated as 4 to 3 and 1 to 1, respectively (see Figure 2-16). The ratio of the copy number of the stop-EGFP gene to that of the EBFP gene in transgenic line F#19-6 appears to be close to 1 to 2. Although the stop-EGFP gene and the EBFP gene were injected with a ratio of 4 to 1, the ratio of the copies of the two transgenes integrated into the mouse genome was considerably different from the 4 to 1 ratio. Estimated copy numbers of the stop-EGFP gene and the EBFP gene in each transgenic line are listed in Table 2-9.

Table 2-9. Estimated copy numbers of the stop-EGFP gene and the EBFP gene in each transgenic line.

	F#24	#25	F#19-6	F#19-15
Estimated total copy number of the transgenes*	2	7.76 (8)	21.92 (22)	31.6 (32)
Estimated ratio of the stop-EGFP gene to the EBFP gene (see Figure 2-16)	ND**	1 : 1	1 : 2	4 : 3
Estimated copy number of the stop-EGFP gene	0	4	7	18
Estimated copy number of the EBFP gene	2	4	15	14

*Estimations are based on the real-time PCR data (see Table 2-8) assuming two copies of the transgenes in line F#24. The closest integer to the estimated copy number for each transgenic line is shown in parenthesis.

**Not determined.

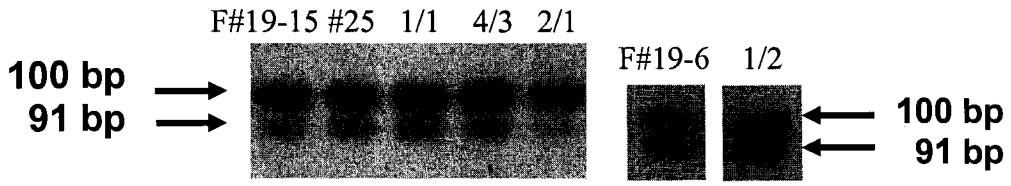


Figure 2-16. Competitive PCR analysis of genomic DNA from transgenic mice.

A competitive PCR was performed to determine the ratios of the stop-EGFP gene to the EBFP gene. The primers used generate PCR products of 100 bp from the stop-EGFP gene and 91 bp from the EBFP gene (see Materials and Methods). For references, pCX-stop-EGFP (plasmid containing the stop-EGFP gene) was mixed with pCX-EBFP (plasmid containing the EBFP gene) with ratios of 1 to 1, 4 to 3, 2 to 1, and 1 to 2 and then used for a competitive PCR (1/1, 4/3, 2/1, and 1/2 , respectively).

9. Reverse transcription-PCR (RT-PCR)

To test expression of the two transgenes at the transcriptional level, reverse transcription-PCR (RT-PCR) was performed. From various tissues of the stop-EGFP mouse (such as the kidney, skin, etc), mRNA was collected and subsequently used for reverse transcription to synthesize cDNA. The cDNA generated was then amplified by PCR. The RT-PCR experiment demonstrated that mRNA is transcribed from the transgenes in all organs tested (i.e., the liver, kidney, lung, spleen, and skin) (Figure 2-17A). To verify mRNA transcription from both the stop-EGFP and EBFP genes, PCR reactions were performed using a primer set that is expected to generate PCR products of 100 bp from the stop-EGFP cDNA and 91 bp from the EBFP cDNA (see section II-9). The RT-PCR experiment demonstrated that mRNA is transcribed from both transgenes in all organs tested (Figure 2-17B). When PCR was performed without the reverse transcription process, no bands were detected (see Figures 2-17A and B). This result shows that the RT-PCR products detected in various tissues were not generated from contaminating genomic DNA, but from mRNA transcripts.

In eukaryotic systems, mRNA transcripts containing a premature stop codon are degraded by the mechanism called nonsense-mediated mRNA decay (Maquat, 2004; Singh and Lykke-Andersen, 2003). In a mammalian system, a stop codon located more than 50 to 55 nucleotides upstream of the last exon-exon junction of mRNA is recognized as a premature stop codon and the mRNA is led to rapid decay; mRNA transcript from a gene without an intron is not targeted for nonsense-mediated mRNA decay even if the mRNA contains a premature stop codon (Maquat & Li, 2001). Because

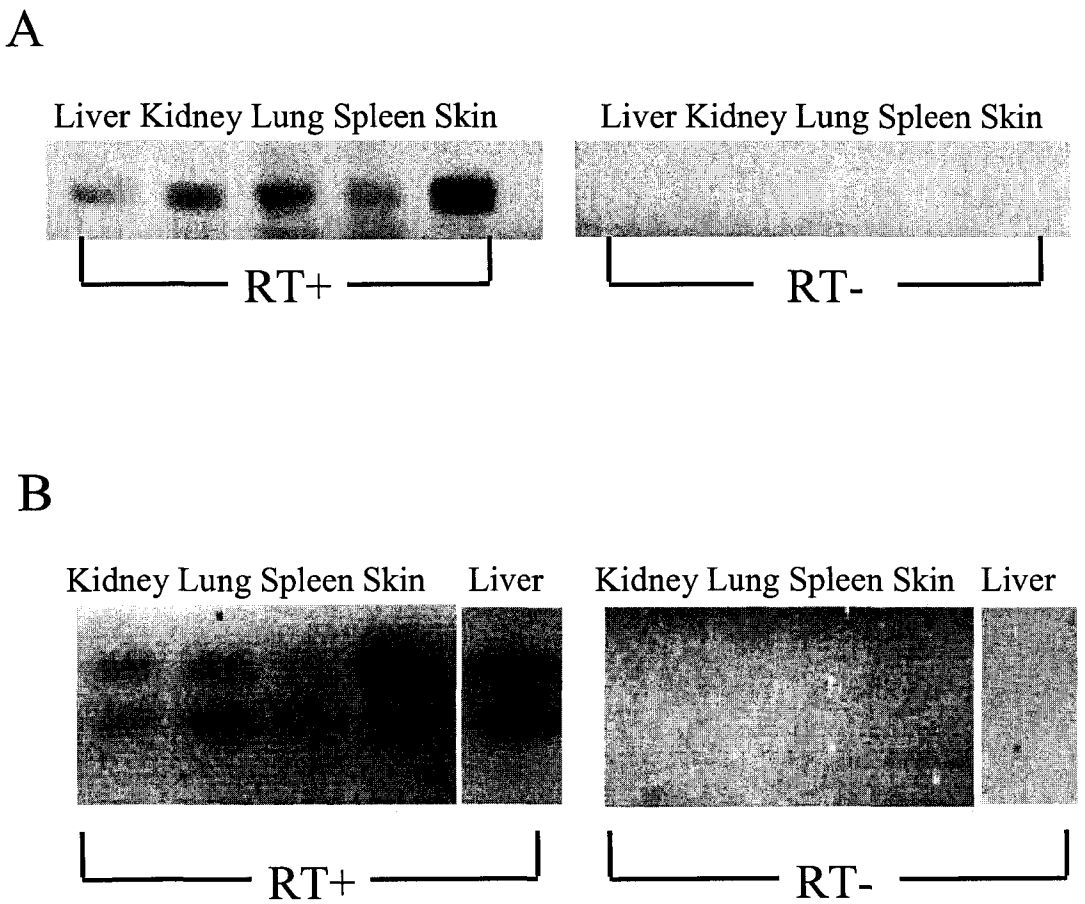


Figure 2-17. RT-PCR of mRNA collected from various tissues of the stop-EGFP mouse (line F#19-15).

(A) cDNA was synthesized from mRNA by reverse transcription and PCR was performed using the reverse transcribed cDNA. The primers used for the PCR generate PCR products of 380 bp from the cDNA of the transgenes. The RT-PCR analysis shows that the transgenes are expressed in various tissues (left panel). No bands were detected when PCR was performed without reverse transcription. (right panel). (B) PCR was also performed using a primer set that generates PCR products of 100 bp from the stop-EGFP cDNA and 91 bp from the EBFP cDNA. The RT-PCR analysis shows that both the stop-EGFP gene and the EBFP gene are expressed in the kidney, lung, spleen, skin, and liver (left panel). No bands were detected of either 100 bp or 91 bp size when PCR was performed without reverse transcription (right panel). Gels were stained in ethidium bromide. Black and white images were inverted using Adobe® Photoshop® (4.0 LE) for better resolution.

the stop-EGFP gene does not contain an intron, mRNA transcribed from the stop-EGFP gene is not degraded and thus can be detected using RT-PCR.

10. Western blotting

To test the transgene expression at the protein level, imaging of EBFP was performed using skin tissue from the stop-EGFP mouse (line F#19-15). Because the two transgenes (the stop-EGFP gene and EBFP gene) are shown to be integrated into the same site of a chromosome (see section III-6), expression of the EBFP gene can indirectly verify expression of the stop-EGFP gene.

Fresh ear skin from mice was used for direct imaging of EBFP. No apparent difference was detected in the intensity of blue fluorescence between transgenic mice and non-transgenic mice when cells were imaged using a fluorescence microscope (data not shown). We presume that the result more likely reflects inefficiency of EBFP as a gene expression marker rather than indicating a lack of protein expression from the transgene in the stop-EGFP mouse. EBFP is not an efficient expression marker due to rapid photobleaching and low fluorescence intensity and imaging of EBFP has proven to be hard (Patterson et al., 2001; Tsien, 1998). Due to the difficulty of EBFP imaging, we pursued an alternative method, Western blotting to verify protein expression from the EBFP gene.

Proteins were extracted from the spleen and kidney of the stop-EGFP mouse and analyzed using anti-GFP antibody which also binds to EBFP. EBFP-specific bands, approximately 27 kDa, were detected in both tissues of the stop-EGFP mouse (Figure 2-18). The data demonstrates that the transgenes are expressed in the stop-EGFP mouse confirming the RT-PCR data (see section III-9). Thus, once a nonsynonymous mutation

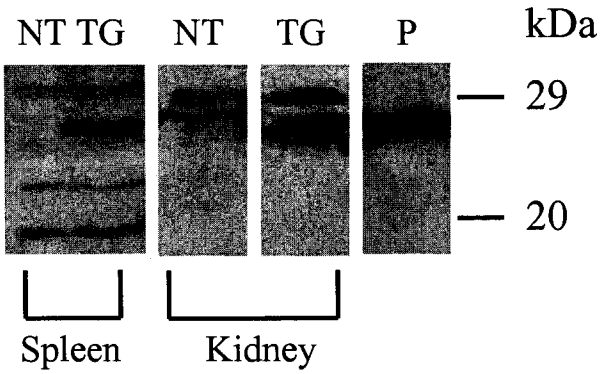


Figure 2-18. Western blot analysis of protein extracts from tissues of the stop-EGFP mouse.

Protein extracts were prepared from the spleen and kidney of the stop-EGFP mouse and its non-transgenic littermate and then used for Western blotting. EBFP-specific bands (about 27 kDa) were detected in both tissues of the stop-EGFP mouse. Purified GFP was used as a positive control (denoted as P). NT and TG denote a non-transgenic littermate and a transgenic mouse, respectively.

arises at the premature stop codon of the stop-EGFP gene in a cell of the stop-EGFP mouse, it is expected that revertant EGFP is expressed in the cell which would become green fluorescent when illuminated with the 488 nm excitation light (see section III-1).

IV. Discussion

In this chapter we have described the development of a novel transgenic mouse system for studies of cell fate mapping, as well as mutation research. In principle, the stop-EGFP transgenic mouse allows single cells to be labeled by means of random mutations and clonal cell lineages to be traced in vivo using fluorescence imaging. The system should also be applicable to studies of in vivo mutations whereby the stop-EGFP gene would serve as a target gene for mutations.

Two transgenic lines have been established (#25 and F#19) from the microinjection experiment and subsequent breeding. Southern blot analysis shows that two different transgene copy numbers exist among transgenic offspring of founder F#19 (Figure 2-14A). One possible explanation for the Southern blot data is that the transgenes were integrated into two different sites in the genome (e.g., two different chromosomes) of the F#19 founder mouse, one site carrying a higher copy number than the other. Offspring would then carry different copy numbers of the transgenes depending on which chromosomes they inherit. Alternatively, the transgenes integrated into a chromosome might have undergone gene rearrangement (e.g., gene amplification or deletion) at the locus during the first or second mitotic cell division of the fertilized oocyte. Considering that the transgenes have been transmitted to offspring of the founder mouse with the frequency of about 50% (Table 2-6), we believe gene rearrangement is the more

plausible explanation for the presence of two different copy numbers of the transgenes among the transgenic offspring of founder F#19 mouse. Had the transgenes been integrated into two different chromosomes, the probability that transgenic offspring are generated would have been closer to 75%. A likelihood ratio test rejects this hypothesis (see Table 2-7). To determine whether the transgenic loci in chromosomes for the two transgenic lines generated from the F#19 founder (i.e., lines F#19-15 and F#19-6) are identical, we are planning a fluorescence in situ hybridization (FISH) experiment in the future (Harwood et al., 2005; Matsui et al., 2002).

Although transgenes integrated into the mouse genome are in general, stable for many generations without rearrangement (Palmiter and Brinster, 1986), there have been reports of transgenes that were either rearranged, partially deleted, or amplified (Kearns et al., 2001; Shani, 1986). No additional variation in copy number of the transgenes was detected in subsequent generations. Although the transgenes had likely undergone gene rearrangement during the first or second mitotic cell division of the fertilized oocyte after microinjection, they have been stably inherited through successive generations without further gene rearrangement.

From the real-time PCR analysis, the relative ratio of the copy number in each transgenic line to that in line F#24 was estimated. Because the copy number of the transgenes in line F24 was not precisely determined for the present study, it is hard to determine the copy numbers in other lines as well. For example, line F#19-15 could carry 16 copies or 32 copies of the transgenes depending on whether line F#24 carries one copy or two copies of the transgenes (see Table 2-8). Comparison of the copy number of the transgenes with that of endogeneous genes such as hypoxanthine-guanine

phosphoribosyl transferase gene (*hprt*) and glyceraldehyde-3-phosphate dehydrogenase gene is expected to provide precise estimates of the copy number of the transgenes and will be performed in the future.

One concern with the stop-EGFP system is whether the level of protein expression from a single mutant copy of the stop-EGFP gene will be sufficient for detection under a fluorescence microscope. However, protein expression from a transgene is more likely affected by the location of the transgene in the genome than by the copy number of the transgene. Transgenic mice carrying single copies of transgenes have been developed in other studies in which a high level of protein expression from the transgenes was detected (Bronson et al., 1996; Caron et al., 2002). Furthermore, among the transgenic lines carrying the EGFP gene which have been developed by Okabe et al., one transgenic line carried only one copy of the EGFP gene but still exhibited bright green fluorescence from cells (personal communication with Dr. Okabe). Thus, although a mutation arises at the premature stop codon in only one copy of the stop-EGFP gene, the mutant copy can be expected to express revertant EGFP highly enough for detection using a fluorescence microscope.

More studies can be carried out to further characterize the stop-EGFP mouse system. For example, to test whether the stop-EGFP mouse shows variegation of transgene expression (Garrick et al., 1998), an immunohistochemical analysis using anti-EBFP antibody can be performed in various tissues, which will determine what proportion of cells in tissue express the transgene. Furthermore, more diverse tissues should be tested for transgene expression (e.g., using Western or Northern blotting) to establish ubiquitous expression of the transgenes in the stop-EGFP mouse. Another

important characterization of the transgenic system is to determine the transgenic loci in chromosomes for various stop-EGFP transgenic lines using a FISH method. These additional analyses will further increase the utility of the stop-EGFP system and will be carried out in the near future.

V. References

Bronson SK, Plaehn EG, Kluckman KD, Hagaman JR, Maeda N, Smithies O. (1996) Single-copy transgenic mice with chosen-site integration. *Proc Natl Acad Sci U S A.* 93: 9067-9072.

Caron KM, James LR, Kim HS, Morham SG, Sequeira Lopez ML, Gomez RA, Reudelhuber TL, Smithies O. (2002) A genetically clamped renin transgene for the induction of hypertension. *Proc Natl Acad Sci U S A.* 99: 8248-8252.

Church GM, Gilbert W. (1984) Genomic sequencing. *Proc Natl Acad Sci U S A.* 81: 1991-1995.

Cormack BP, Valdivia RH, Falkow S. (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene.* 173: 33-38.

Corr M, Boyd LF, Frankel SR, Kozlowski S, Padlan EA, Margulies DH. (1992) Endogenous peptides of a soluble major histocompatibility complex class I molecule, H-2Lds: sequence motif, quantitative binding, and molecular modeling of the complex. *J Exp Med.* 176: 1681-1692.

Diviacco S, Norio P, Zentilin L, Menzo S, Clementi M, Biamonti G, Riva S, Falaschi A, Giacca M. (1992) A novel procedure for quantitative polymerase chain reaction by coamplification of competitive templates. *Gene.* 122: 313-320.

Dopf J, Horiagon TM. (1996) Deletion mapping of the Aequorea victoria green fluorescent protein. *Gene.* 173: 39-44.

Falk K, Rotzschke O, Deres K, Metzger J, Jung G, Rammensee HG. (1991) Identification of naturally processed viral nonapeptides allows their quantification in infected cells and suggests an allele-specific T cell epitope forecast. *J Exp Med.* 174: 425-434.

Feng G, Mellor RH, Bernstein M, Keller-Peck C, Nguyen QT, Wallace M, Nerbonne JM, Lichtman JW, Sanes JR. (2000) Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *Neuron.* 28: 41-51.

Gambotto A, Dworacki G, Cicinnati V, Kenniston T, Steitz J, Tuting T, Robbins PD, DeLeo AB. (2000) Immunogenicity of enhanced green fluorescent protein (EGFP) in

BALB/c mice: identification of an H2-Kd-restricted CTL epitope. *Gene Ther.* 7: 2036-2040.

Garrick D, Fiering S, Martin DI, Whitelaw E. (1998) Repeat-induced gene silencing in mammals. *Nat Genet.* 18: 56-59.

Gilliland G, Perrin S, Blanchard K, Bunn HF. (1990) Analysis of cytokine mRNA and DNA: Detection and quantitation by competitive polymerase chain reaction. *Proc Natl Acad Sci U S A.* 87: 2725-2729.

Goelz MF, Mahler J, Harry J, Myers P, Clark J, Thigpen JE, Forsythe DB. (1998) Neuropathologic findings associated with seizures in FVB mice. *Lab Anim Sci.* 48: 34-37.

Goldman IL, Kadulin SG, Razin SV. (2004) Transgenic animals in medicine: integration and expression of foreign genes, theoretical and applied aspects. *Med Sci Monit.* 10: RA274-RA285.

Gossen JA, de Leeuw WJ, Tan CH, Zwarthoff EC, Berends F, Lohman PH, Knook DL, Vijg J. (1989) Efficient rescue of integrated shuttle vectors from transgenic mice: a model for studying mutations in vivo. *Proc Natl Acad Sci U S A.* 86: 7971-7975.

Hakenberg J, Nussbaum AK, Schild H, Rammensee HG, Kuttler C, Holzhutter HG, Kloetzel PM, Kaufmann SH, Mollenkopf HJ (2003). MAPPP: MHC class I antigenic peptide processing prediction. *Appl Bioinformatics.* 2: 155-158.

Harwood WA, Bilham LJ, Travella S, Salvo-Garrido H, Snape JW. (2005) Fluorescence in situ hybridization to localize transgenes in plant chromosomes. *Methods Mol Biol.* 286: 327-340.

Heim R, Prasher DC, Tsien RY. (1994) Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc Natl Acad Sci U S A.* 91: 12501-12504.

Heim R, Tsien RY. (1996) Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr Biol.* 6: 178-182.

Jankowsky JL, Slunt HH, Ratovitski T, Jenkins NA, Copeland NG, Borchelt DR. (2001) Co-expression of multiple transgenes in mouse CNS: a comparison of strategies. *Biomol Eng.* 17: 157-165.

Kearns M, Morris C, Whitelaw E. (2001) Spontaneous germline amplification and translocation of a transgene array. *Mutat Res.* 486: 125-136.

Kohler SW, Provost GS, Fieck A, Kretz PL, Bullock WO, Sorge JA, Putman DL, Short JM. (1991) Spectra of spontaneous and mutagen-induced mutations in the lacI gene in transgenic mice. *Proc Natl Acad Sci U S A.* 88: 7958-7962.

Lehninger AL, Nelson DL, Cox MM. (1993) Amino acids and peptides. In: *Principles of Biochemistry* (2nd edition). New York: Worth Publishers, pp. 111-133.

Li X, Zhang G, Ngo N, Zhao X, Kain SR, Huang CC. (1997) Deletions of the Aequorea victoria green fluorescent protein define the minimal domain required for fluorescence. *J Biol Chem.* 272: 28545-28549.

Mahler JF, Stokes W, Mann PC, Takaoka M, Maronpot RR. (1996) Spontaneous lesions in aging FVB/N mice. *Toxicol Pathol.* 24: 710-716.

Maquat LE. (2004) Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. *Nat Rev Mol Cell Biol.* 5: 89-99.

Maquat LE, Li X. (2001) Mammalian heat shock p70 and histone H4 transcripts, which derive from naturally intronless genes, are immune to nonsense-mediated decay. *RNA.* 7: 445-456.

Matsui S, Sait S, Jones CA, Nowak N, Gross KW. (2002) Rapid localization of transgenes in mouse chromosomes with a combined Spectral Karyotyping/FISH technique. *Mamm Genome.* 13: 680-685.

McCaughan KK, Poole ES, Pel HJ, Mansell JB, Mannering SA, Tate WP. (1998) Efficient in vitro translational termination in Escherichia coli is constrained by the orientations of the release factor, stop signal and peptidyl-tRNA within the termination complex. *Biol Chem.* 379: 857-866.

Miller SA, Dykes DD, Polesky HF. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16: 1215.

Noveroske JK, Weber JS, Justice MJ. (2000) The mutagenic action of N-ethyl-N-nitrosourea in the mouse. *Mamm Genome.* 11: 478-483.

Okabe M, Ikawa M, Kominami K, Nakanishi T, Nishimune Y. (1997) 'Green mice' as a source of ubiquitous green cells. *FEBS Lett.* 407: 313-319.

Palmiter RD, Wilkie TM, Chen HY, Brinster RL. (1984) Transmission distortion and mosaicism in an unusual transgenic mouse pedigree. *Cell.* 36: 869-877.

Palmiter RD, Brinster RL. (1986) Germ-line transformation of mice. *Annu Rev Genet.* 20: 465-499.

Patterson G, Day RN, Piston D. (2001) Fluorescent protein spectra. *J Cell Sci.* 114: 837-838.

Popp RA, Bailiff EG, Skow LC, Johnson FM, Lewis SE. (1983) Analysis of a mouse alpha-globin gene mutation induced by ethylnitrosourea. *Genetics.* 105: 157-167.

Rammensee HG, Friede T, Stevanoviic S. (1995) MHC ligands and peptide motifs: first listing. *Immunogenetics.* 41: 178-228.

Rock KL, Goldberg AL. (1999) Degradation of cell proteins and the generation of MHC class I-presented peptides. *Annu Rev Immunol.* 17: 739-779.

Shani M. (1986) Tissue-specific and developmentally regulated expression of a chimeric actin-globin gene in transgenic mice. *Mol Cell Biol.* 6: 2624-2631.

Sher BT, Nairn R, Coligan JE, Hood LE. (1985) DNA sequence of the mouse H-2Dd transplantation antigen gene. *Proc Natl Acad Sci U S A.* 82: 1175-1179.

Singh G, Lykke-Andersen J. (2003) New insights into the formation of active nonsense-mediated decay complexes. *Trends Biochem Sci.* 28: 464-466.

Sola I, Castilla J, Pintado B, Sanchez-Morgado JM, Whitelaw CB, Clark AJ, Enjuanes L. (1998) Transgenic mice secreting coronavirus neutralizing antibodies into the milk. *J Virol.* 72: 3762-3772.

Stripecke R, Carmen Villacres M, Skelton D, Satake N, Halene S, Kohn D. (1999) Immune response to green fluorescent protein: implications for gene therapy. *Gene Ther.* 6: 1305-1312.

Taketo M, Schroeder AC, Mobraaten LE, Gunning KB, Hanten G, Fox RR, Roderick TH, Stewart CL, Lilly F, Hansen CT, Overbeek PA. (1991) FVB/N: an inbred mouse strain preferable for transgenic analyses. *Proc Natl Acad Sci U S A.* 88: 2065-2069.

Tsien RY. (1998) The green fluorescent protein. *Annu Rev Biochem.* 67: 509-544.

Wilkie TM, Brinster RL, Palmiter RD. (1986) Germline and somatic mosaicism in transgenic mice. *Dev Biol.* 118: 9-18.

Yang TT, Cheng L, Kain SR. (1996) Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. *Nucleic Acids Res.* 24: 4592-4593.

York IA, Rock KL. (1996) Antigen processing and presentation by the class I major histocompatibility complex. *Annu Rev Immunol.* 14: 369-396.

Chapter 3. Tracing Clonal Lineages of Epidermal Stem Cells in vivo

Parts of this chapter have been previously published.

Ro S, Rannala B. (2004) A stop-EGFP transgenic mouse to detect clonal cell lineages generated by mutation. *EMBO Rep.* 5: 914-920.

Ro S. (2004) Magnifying stem cell lineages: the stop-EGFP mouse. *Cell Cycle.* 3: 1246-1249.

Ro S, Rannala B. (2005) Evidence from the stop-EGFP mouse supports a niche-sharing model of epidermal proliferative units. *Exp Dermatol.* 14: 838-843.

I. Introduction

Investigation of the distribution and the types of daughter cells that individual precursor cells give rise to is of great importance not only in developmental studies but also in stem cell biology. Recent studies imply that stem cells in an adult tissue, which have long been thought to produce only limited cell types in a given tissue, have the capability to produce even cells of developmentally unrelated tissue types in experimentally modified conditions (see section II-4, Chapter 1). However, the validity of claims on the potential of stem cells (i.e., lineage-switching potentials) has been often questioned. Most reports claiming lineage-switching of stem cells used heterogeneous population of cells for the studies, thus contamination of stem cells of diverse tissue origins is possible. To address the issues and, to a certain extent, to evaluate the clinical potential of stem cells for treating various degenerative diseases (see section II-5, Chapter 1), a novel single-cell marking technique that allows tracing of clonal lineages of stem cells in vivo is required. Such a technique could also reveal the proliferative potential of single stem cells. Investigation of distributions of clonal lineages of stem cells using the technique would allow one to study at the single-cell level how stem cells maintain the proliferative unit in a tissue under the normal steady-state condition and how stem cells could expand the proliferative unit in response to a hyperproliferative signal (see section II-3, Chapter 1).

Although various cell fate mapping techniques have been developed, labeling of single cells has proven difficult and few systems have been efficiently applied to fate mapping of single precursor cells (see section II-6, Chapter 1). Recently, we developed a transgenic mouse (referred to as the stop-EGFP mouse) in an attempt to trace clonal cell lineages in vivo (Chapter 2). The efficiency of the stop-EGFP system for in vivo analysis

of stem cell clonal lineages is demonstrated in this chapter by applying the system to the mouse dorsal epidermis.

Epidermis is composed of four layers: basal layer, spinous layer, granular layer, and cornified layer (Allen and Potten, 1974; Turksen and Troy, 1998) (Figure 3-1). In some cases, such as mouse dorsal epidermis, the spinous and granular layers are not distinct. The proliferative activity in the epidermis is restricted to the basal layer where stem cells and transit amplifying (TA) cells are presumed to be located (Potten, 1981). Basal cells move upward to the cornified layer via spinous and granular layers to replenish cells being lost at the surface of the epidermis. During the vertical migration, epidermal cells get larger, become flattened, and terminally differentiate into corneocytes (cells in the outermost cornified layer) (Figure 3-1).

In many constantly renewing tissues, such as the epidermis and the epithelium of the small intestine, a distinct compartment of cellular proliferation is thought to exist, which functions as a unit of tissue renewal (Slack, 2000). A single or a few stem cells located within a proliferative compartment (also referred to as a "proliferative unit") maintain the unit by constantly producing cells, which make up for the cells being lost (see section II-3, Chapter 1). In the mouse dorsal epidermis, it has been proposed that a proliferative unit is composed of a single column of corneocytes plus epidermal cells directly below the column (Potten, 1981; Potten and Morris, 1988).

According to the model proposed by Potten and colleagues, each epidermal proliferative unit (EPU) contains 10.6 cells on average in the basal layer (with a stem cell in the center) and surrounding TA basal cells that descend from a central stem cell move directly upward to terminally differentiate into corneocytes (Potten, 1974; Potten, 2004).

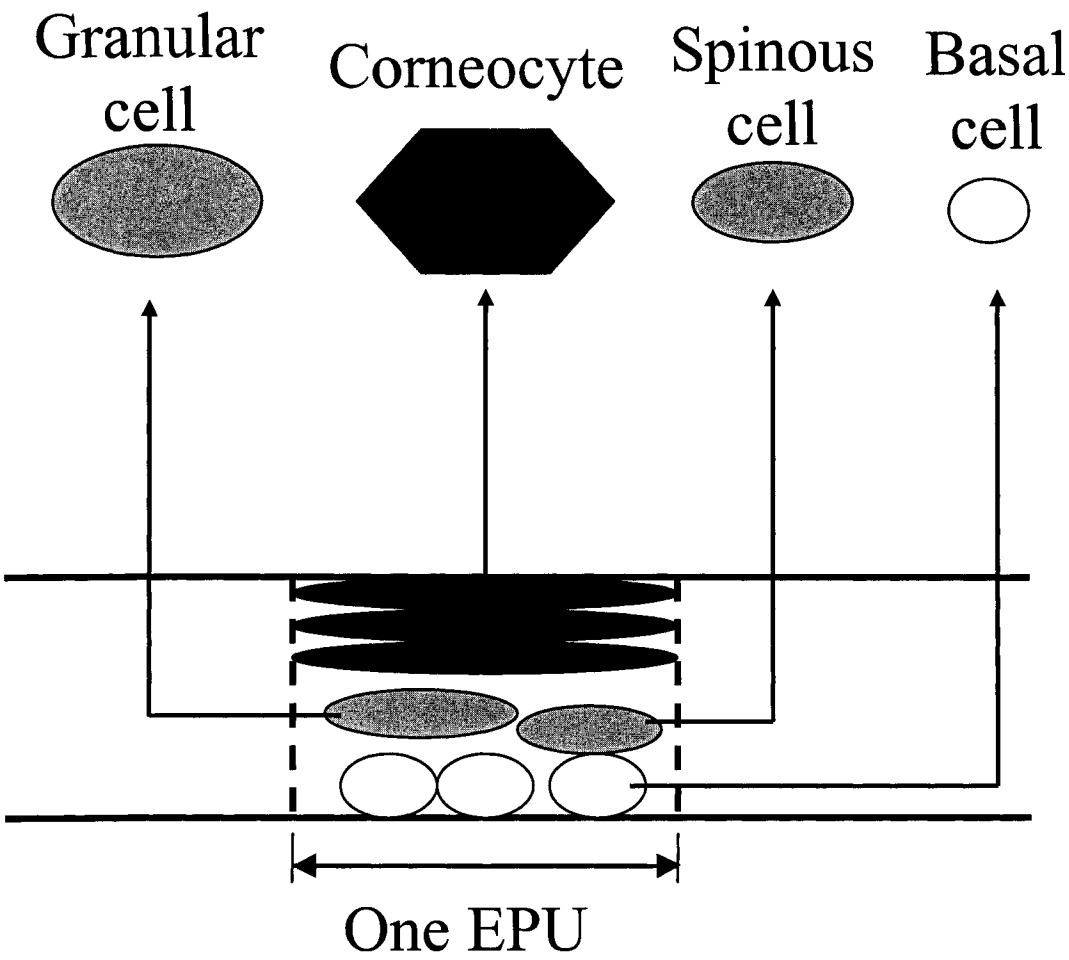


Figure 3-1. Schematic representation of the structure of the epidermis.

The epidermis consists of four layers: basal, spinous, granular, and cornified layers. In the mouse dorsal epidermis, spinous and granular layers are not distinct. Basal cells move directly upward to terminally differentiate into corneocytes via spinous and granular layers. During the vertical migration, epidermal cells get larger and become flattened. According to the classical EPU model proposed by Potten et al. (Allen and Potten, 1974; Potten and Morrison, 1988), the epidermal proliferative unit (EPU) is about 30 μm wide and consists of a single column of corneocytes plus granular, spinous, and basal cells directly below the column.

The model specifies that an EPU is a unit of cellular proliferation maintained by a single central stem cell, and thus two neighboring EPUs (or two adjacent columns of corneocytes) will be maintained by two different stem cells located within each respective EPU. Thus, the classical EPU model suggests that epidermal cells do not migrate between EPUs (Figure 3-2).

The concept of the EPU was experimentally verified in a study by Mackenzie (1997) investigating cell lineages in the mouse epidermis using retroviral transduction encoding the *lacZ* gene. In this study, it was observed that *lacZ*-labeled cell clusters derived from single stem cells corresponded to single columns of corneocytes plus epidermal cells beneath them, as suggested by the EPU model. Ultra-structural morphology and kinetic data of the epidermal tissue also have supported the EPU concept (Mackenzie, 1970; Potten, 1974; Allen and Potten, 1974). However, several recent reports bring the EPU model into question. Using chimeras of mice made from strains that express different H-2 antigens, Schmidt et al. observed that epidermal cells of both strains were located under a single column of corneocytes (i.e., within a single EPU), suggesting that epidermal cells might migrate between EPUs and thus an EPU might not be a proliferative compartment (Schmidt et al., 1987). As well, recent epidermal cell lineage studies using replication-deficient *lacZ*-carrying retrovirus suggest that a single basal stem cell might maintain multiple EPUs (Ghazizadeh and Taichman, 2001; Kameda et al., 2003). The method using retroviral transduction, however, has the limitation that cell lineages expanding to multiple EPUs might be generated from multiple stem cells which are independently labeled by the retroviral infection and thus the clonality of cell lineages detected in the studies is not guaranteed.

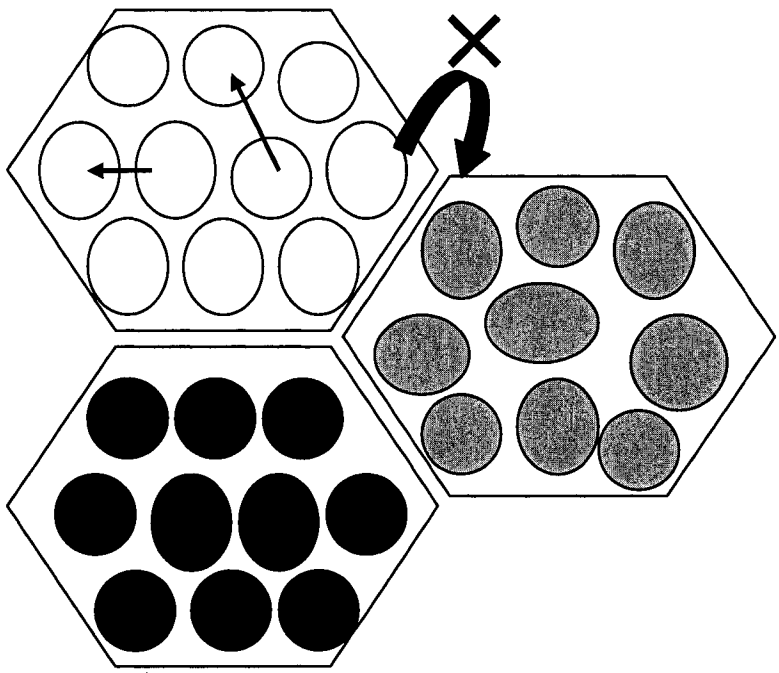


Figure 3-2. Schematic illustration of EPUs (from surface view).

Each hexagon represents the boundary of a single EPU (corresponding to the area of a single corneocyte at the surface of the epidermis). Three EPUs are shown here with basal cells overlapped. The classical EPU model specifies that epidermal cells belonging to one EPU cannot move to another EPU and migration of epidermal cells is restricted to within an EPU.

To unequivocally determine the proliferative unit that a single epidermal stem cell maintains, a cell fate mapping technique is required that allows cell lineages that originated from single cells to be identified. Because the stop-EGFP mouse is expected to allow clonal cell lineages to be identified in vivo, we applied the stop-EGFP mouse to the mouse dorsal epidermis to investigate the extent of a proliferative unit that a single stem cell can contribute to in the epidermis. To increase the efficiency with which individual cells are labeled by means of mutations at the premature stop codon within the stop-EGFP gene, we treated stop-EGFP mice with a potent mutagen ENU (*N*-ethyl-*N*-nitrosourea). Cell lineages were then investigated in vivo after several epidermal turnovers following ENU treatment to specifically identify stem cell clonal lineages in the epidermis. Because stem cells persist in constantly renewing tissues while other cells do not, cell lineages generated from green fluorescent mutant stem cells will remain following tissue renewal while cell lineages that originated from non-stem cell mutants (induced by the ENU exposure) are eliminated (see section II-6, Chapter 1). Using five ENU-treated stop-EGFP mice, we investigate stem cell clonal lineages in the epidermis.

Imaging of the skin cells of a live mouse has been successfully performed by many groups (Bussau et al., 1998; Izzo et al., 2001). Izzo et al. (2001) analyzed in vivo GFP expression in dermal fibroblasts of live mice using a laser scanning confocal microscope. Bussau et al. (1998) imaged epidermal cells of live mice after staining the epidermis with a fluorescent dye. In these studies, a laser scanning confocal microscope allowed individual cells to be imaged and detailed morphology and cellular organization of epidermal cells to be visualized. Because detection of EGFP expression in individual cells requires no additional substrates and fluorescence imaging is itself non-invasive,

even repeated in vivo analyses of a specific cell lineage over time is feasible.

In a follow-up study using 11 ENU-treated stop-EGFP mice, we successfully performed repeated imaging of the same clonal cell lineages over time in the epidermis (see section III-3). To test the effect of treatment with ENU on the proliferative structure maintained by stem cells, we performed in vivo imaging of the dorsal epidermis of 26 untreated stop-EGFP mice and investigated clonal cell lineages generated by spontaneous mutations (Ro and Rannala, 2005).

II. Materials and Methods

The handling of ENU (*N*-ethyl-*N*-nitrosourea) followed standard operating protocols developed by the Office of Environmental Health & Safety at the University of Alberta. Detailed descriptions of protocols can be downloaded from http://www.rannala.org.

1. Personal protective equipment for handling ENU

A waterproof cover-all gown (Fisher Scientific) and nitrile gloves (Fisher Scientific) were worn during all procedures involving the handling of ENU as well as a P100 particulate respirator (3M) and a splash-shield (Fisher Scientific).

2. Preparation of ENU solution

The ENU solution was prepared in a type IIB2 biosafety cabinet that exhausts 100% to the outside. A vial containing approximately 1 g of ENU powder was purchased (ISOPAC, Sigma). A 10 ml quantity of 95% ethanol was injected into a vial through the rubber injection port and ENU was dissolved by gentle shaking of the vial. A 90 ml quantity of Phosphate-Citrate (0.1 M Sodium Phosphate and 0.05 M Sodium Citrate, pH 5.0) was added to the ENU solution and mixed by inverting and shaking the vial.

3. Animal experiments

All experiments using live mice were performed in compliance with the recommendations of the Canadian Council on Animal Care and have been approved by the Health Sciences Animal Policy and Welfare Committee of the University of Alberta.

4. Injection of ENU

Mice were weighed and the optical density (OD) of the ENU solution at 398 nm was measured immediately prior to injection to precisely determine the concentration of the ENU solution. A concentration of 1 mg/ml of ENU corresponds to OD 0.72 at 398 nm. Mice were anesthetized by inhalation of isoflurane gas prior to ENU injection. ENU was administered intraperitoneally at a dose of 150 mg/kg body weight. Following the injection, mice were place in disposable cages and kept in the cages inside the biosafety cabinet for 24 hours. At the end of this 24 hour period, the outside of the disposable cage was wiped with the ENU-inactivating solution (see below) before placing the cage in the regular animal room. Four days after the ENU injection, mice were transferred from the disposable cages to regular cages and the disposable cages were incinerated.

5. Inactivation and disposal of ENU

The stability of aqueous solutions of ENU is highly dependent on pH of the solution. The half-life of ENU at pH 7.0 is 1.5 hours, and at pH 9.0 the half-life is 0.05 hours. ENU was inactivated with ENU inactivating solution (0.1 M NaOH and 20% $Na_2S_2O_3$). All the needles and materials that were in contact with ENU were rinsed with the ENU-inactivating solution and left inside the biosafety cabinet for 24 hours under fluorescent light. All personal protective equipment, disposable cages and bedding were incinerated after use.

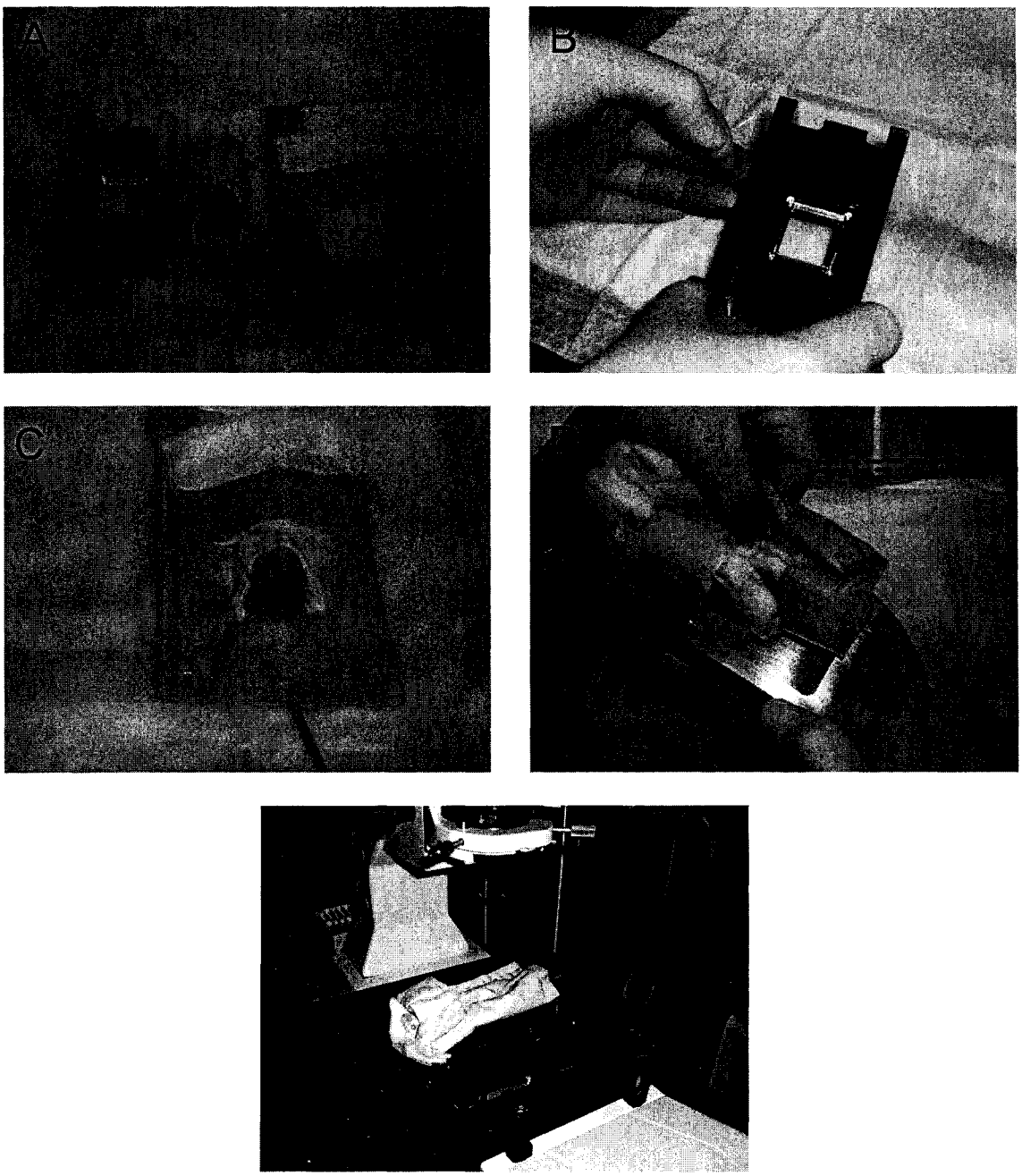
6. In vivo imaging of the mouse dorsal skin

One week before the in vivo imaging experiment, mice were anesthetized by inhalation of isoflurane gas. After the mice were sedated, the hair (2.5 cm × 2.5 cm area) in the lower part of the dorsal skin was clipped using electric clippers (Oster Golden A5, small animal clippers with a number 40 blade, Oster Professional Products). A depilatory agent (Nair, Church & Dwight Co., Inc.) was applied to the clipped area for two and one-half minutes.

On the day of the imaging experiment, mice were anesthetized with a mixture of ketamine and acepromazine administered intraperitoneally. To facilitate the in vivo imaging, a mouse restraint with a window on the bottom plate was used. The procedures for the imaging study are shown in Figure 3-3. The depilated area of the epidermis was illuminated by a 50 W mercury lamp and scanned using a Zeiss Axiovert 200M inverted microscope with a 10× F-Fluar lens (NA 0.5) and LP 520 emission filter (Carl Zeiss). With the emission filter (LP 520), EGFP-expressing cells exhibited typical green color while autofluorescence from the skin revealed yellowish color. Images of green fluorescent cells were collected with a confocal laser scanning microscope (LSM 510 NLO, software version 3.0, Carl Zeiss) mounted on the Zeiss Axiovert 200M inverted microscope with a 10× F-Fluar lens (NA 0.5). The 488 nm laser line was used to image EGFP with a band pass filter (505 – 530 nm wavelength) for detecting emissions from EGFP. In order to reduce photobleaching, laser intensity was set to minimum while it is possible to collect reasonable signal (typical acousto-optical tunable filter setting was 10 – 15%). The animals had their eyes lubricated while under the microscope. Following imaging, the animals were returned to a cage in the animal facility.

Figure 3-3. Procedures for in vivo imaging of the mouse dorsal epidermis.

(A) Mouse is anesthetized for the in vivo imaging one week after depilation. (B) A 25 mm × 25 mm cover slip is placed on the window of the mouse restraint shown in panel A. (C) To prevent the mouse from directly contacting metal surface of the restraint (and also to keep it warm), the mouse is wrapped with cotton pad. (D) The mouse is placed with the depilated dorsal skin on the cover slip of the mouse restraint. (E) The constrained mouse is placed on the microscope stage of an inverted fluorescence microscope and the depilated skin is scanned using the microscope.



III. Results

1. In vivo detection of epidermal stem cell clonal lineages

To determine the potential utility of the stop-EGFP mouse for mapping a clonal cell lineage that originated from a stem cell, we applied the stop-EGFP system to the mouse dorsal epidermis. Compared to internal organs, skin is readily accessible and relatively less invasive procedures are required for performing in vivo imaging. Furthermore, the epidermis is a constantly renewing tissue and thus stem cell lineages in the tissue can be relatively efficiently identified by performing cell lineage analysis after several tissue turnovers following labeling of cells (see section II-6, Chapter 1).

To investigate clonal cell lineages that descended from stem cells in the epidermis, we induced mutations in an 8 week old stop-EGFP mouse using a potent mutagen, ENU and performed in vivo imaging of the dorsal skin at 6 weeks post-ENU administration. The turnover rate of the epidermal tissue is estimated to be about two weeks (Potten, 1975), thus one expects approximately three epidermal turnovers during the 6 week period, so that cell lineages that originated from non-stem cell mutants (induced by the ENU exposure) can be expected to have disappeared after that period. Several studies suggest that an interval of 6 weeks after labeling is enough to eliminate epidermal cell lineages that originated from transit amplifying (TA) cells, allowing epidermal stem cell lineages to be specifically detected (Schneider et al., 2003).

For the in vivo imaging, the lower part of the dorsal skin was chosen because imaging of that skin area is least affected by movements due to the respiration and cardiac pumping. Depilation of hair is required because very strong autofluorescence from hair will prevent imaging of the epidermis. Imaging was performed at least one

week after the depilation, which provides sufficient time for the renewal of the outermost layer of the epidermis (Potten et al., 1987) which might have been damaged due to the treatment with a depilatory agent. Hair usually did not reappear until up to 12 days following depilation.

In vivo imaging of the dorsal epidermis of a stop-EGFP mouse at 6 weeks post-ENU administration revealed an epidermal stem cell clonal lineage containing three adjacent corneocytes (Figure 3-4A) (Ro and Rannala, 2004). The hexagonal shape is characteristic of the morphology of a cornified cell (corneocyte) in the outermost layer of the epidermis. This finding is counter to the classical model for stem cell clonal lineages in the mouse dorsal epidermis (Potten, 1981; Potten and Morris, 1988). The classical model specifies that an EPU is composed of a single column of corneocytes plus epidermal cells directly below the column and is a unit of cellular proliferation maintained by a single stem cell within the unit. Thus, two neighboring EPUs (or two adjacent columns of corneocytes) will be maintained by two different stem cells located within each respective EPU. In other words, under the classical EPU model if a basal stem cell has undergone a mutation at the premature stop codon of the stop-EGFP gene, only one EPU (occupying the area of a single corneocyte on the skin surface) should become green fluorescent (see illustrations in Figures 3-5A and C).

However, Figure 3-4A shows a cell lineage that descended from a single epidermal stem cell can generate at least three adjacent corneocytes (i.e., epidermal cells belonging to three adjacent EPUs). The probability that the three EGFP-expressing corneocytes we observed were generated from more than one stem cell, each having undergone independent mutation at the premature stop codon of the stop-EGFP gene, is

Figure 3-4. Green fluorescent mutant colonies in the dorsal epidermis of ENU-treated stop-EGFP mice.

(A) In vivo imaging of the dorsal epidermis of a stop-EGFP mouse at 6 weeks post-ENU administration revealed an epidermal stem cell clonal lineage containing three adjacent corneocytes. Bright fluorescent signals within the hexagonal cells are from a deeper layer (see Figure 3-6). (B) Mutations were induced in a 7-month-old stop-EGFP mouse by treatment with ENU and imaging of the dorsal skin was performed at 13 weeks post-ENU administration. The in vivo imaging experiment also revealed an epidermal stem cell clonal lineage containing three adjacent corneocytes. (C) Imaging of the same cell lineage (shown in panel B) at 20 μm below revealed green fluorescent signals of cells on the basal layer. (D) In vivo imaging of the dorsal epidermis of another stop-EGFP mouse at 18 days post-ENU administration revealed a green fluorescent mutant colony containing a single corneocyte on the skin surface. The cell lineage was not detected when the same area of the dorsal epidermis was imaged again at 6 weeks post-ENU administration. Scale bars, 20 μm.

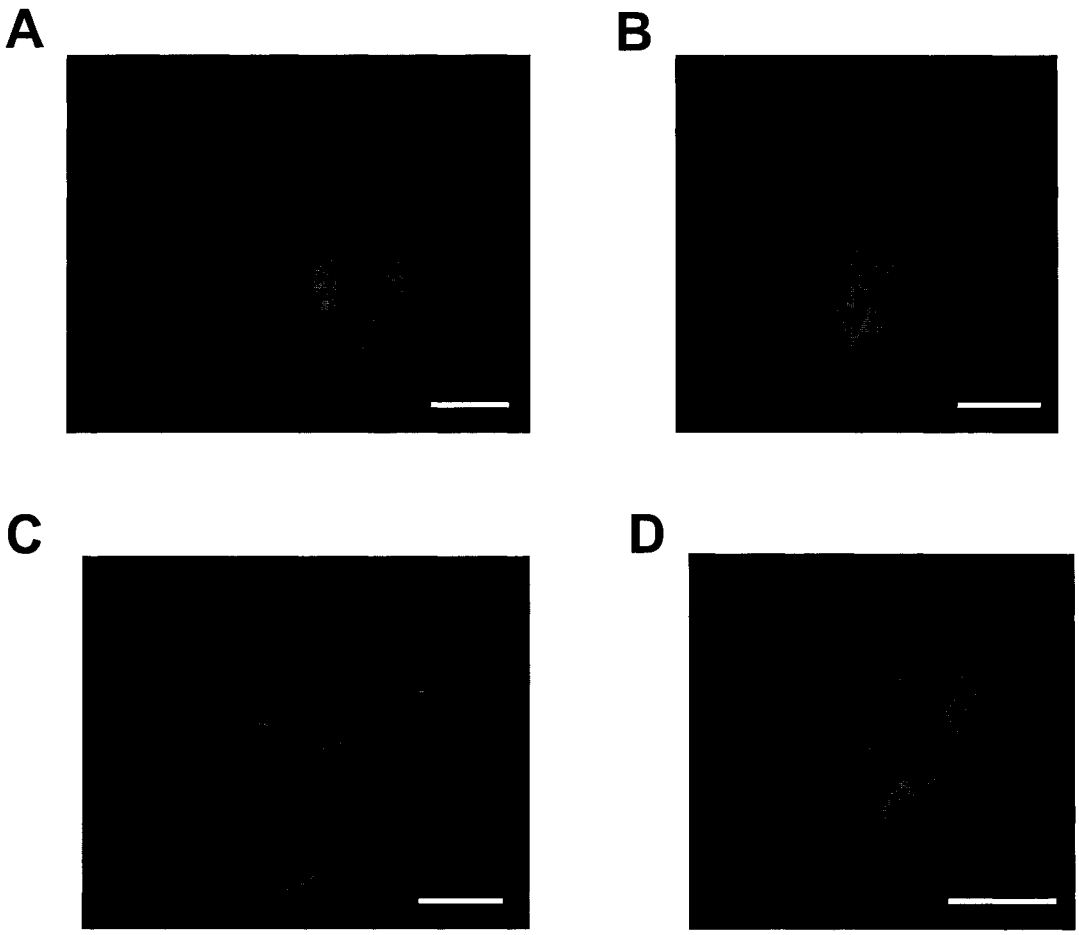
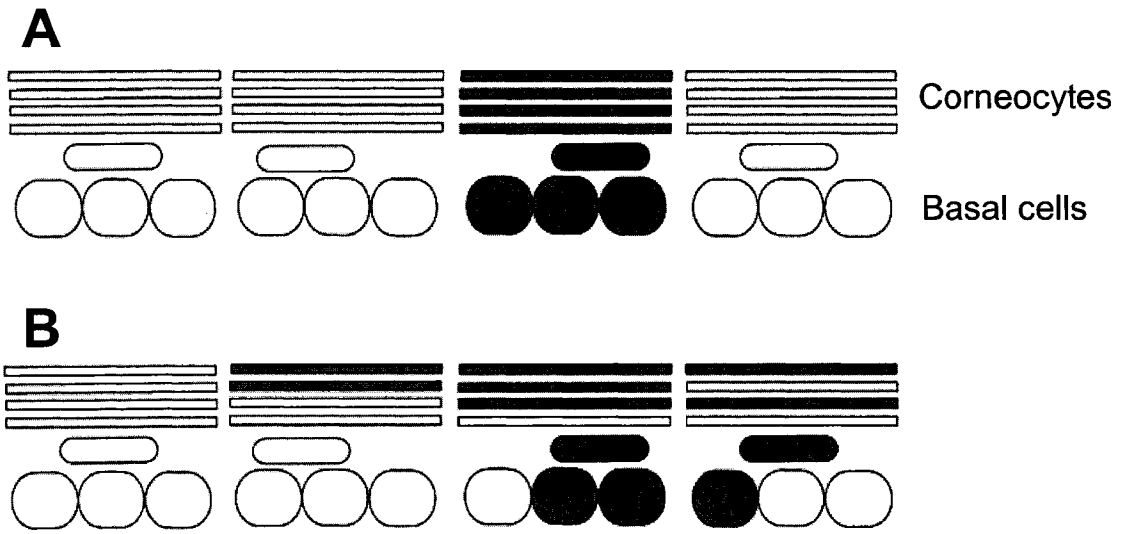


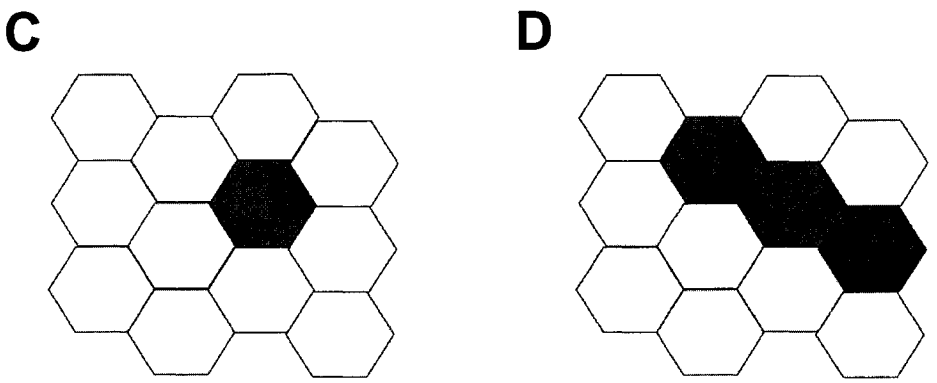
Figure 3-5. The two opposing models for a stem cell clonal lineage in the mouse dorsal epidermis.

The classical EPU model proposes that an EPU is a clonal unit of cellular proliferation. According to the EPU model, a single stem cell having undergone a mutation at the premature stop codon of the stop-EGFP gene should generate only one green fluorescent EPU (A), which occupies the area of a single corneocyte on the skin surface (C). The opposing model proposes that epidermal cells can migrate to adjacent EPUs and a single stem cell can provide epidermal cells for more than one EPU. The finding of epidermal stem cell clonal lineages containing three adjacent corneocytes (i.e., epidermal cells belonging to three adjacent EPUs) in our study strongly supports the latter model. A stem cell clonal lineage detected in our study (shown in green) is schematically represented (B), which contained three adjacent corneocytes on the skin surface (D).

Cross-section view



Surface view



virtually zero due to the low frequency of mutation which insures clonality. Thus, the finding strongly suggests that a single stem cell is capable of providing epidermal cells for at least three adjacent EPUs via horizontal migration of either basal or suprabasal cells (see illustrations in Figures 3-5B and D). Longitudinal optical sectioning of the cell lineage shown in Figure 3-4A shows a green fluorescent area of cells in the basal layer (approximately 40 μm × 30 μm) considerably smaller than that in the cornified layer (i.e., area of three corneocytes, approximately 40 μm × 90 μm) (Figure 3-6).

To demonstrate that the above finding is not restricted to a young mouse, mutations were also induced in a 7 month old stop-EGFP mouse via treatment with ENU, and in vivo imaging of the dorsal skin was performed at 13 weeks post-ENU administration (i.e., after approximately 6.5 epidermal turnovers). This imaging experiment also revealed a clonal cell lineage containing three neighboring corneocytes (Figure 3-4B) with a considerably smaller area of green fluorescent signal in the basal layer (Figure 3-4C). Because epidermal tissue is renewed every 2 weeks (Potten, 1975), it is very unlikely that this clonal cell lineage (detected at 13 weeks post-ENU administration) originated from a non-stem cell. No green fluorescent epidermal cells were detected in the dorsal skin of four untreated stop-EGFP mice suggesting that the green fluorescent cell lineages we observed were generated in adult mice exposed to ENU rather than arising due to spontaneous mutation.

2. Imaging of the epidermis at earlier time points following exposure to ENU

Imaging of the dorsal epidermis of two ENU-treated stop-EGFP mice was performed at one week post-ENU administration in an attempt to detect more mutant cells than those found at later time points after epidermal turnovers. Considering that TA cells are about

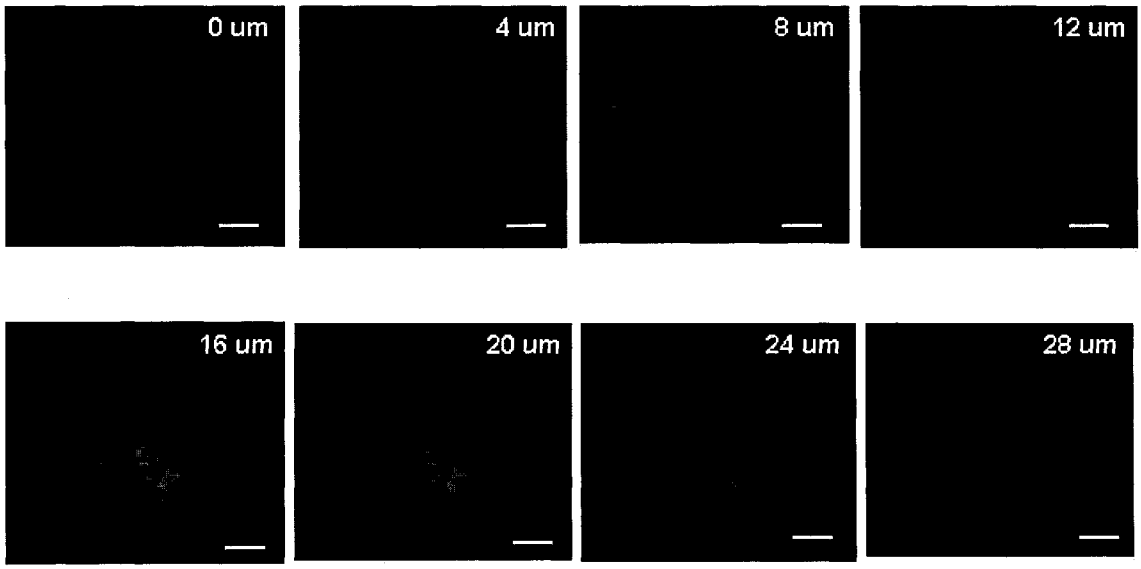


Figure 3-6. Longitudinal optical sectioning of the epidermal cell lineage shown in Figure 3-4A.

The reference depth, 0 μm, was chosen arbitrarily as a depth where a fairly bright signal was detected. The image was scanned at 4 μm intervals of depth, moving the focal plane vertically from the reference depth to successively deeper layers. The brighter fluorescent signals seen in deeper optical sections are presumed to be basal cells or suprabasal cells. Each image was taken using a 10× objective. Scale bars, 20 μm.

10 times more numerous than stem cells in the mouse epidermis (Potten, 1974) and that epidermis renews every two weeks (Potten, 1975), it is expected that many more mutant cells could be observed at one week post-ENU administration (both TA and stem cell mutants are expected to be detected) than, for example, at 6 weeks post-ENU treatment since only stem cell mutants are expected to be detected. However, no mutant cells were detected from the in vivo imaging of two stop-EGFP mice at one week post-ENU treatment. We speculate that it might take more than one week for mutant cells to express the mutant phenotype. Expression of the mutant phenotype will require several processes such as fixation of a mutation following DNA replication, transcription from a mutant copy of the stop-EGFP gene, accumulation of mutant mRNA containing no premature stop codon, and finally expression of revertant EGFP at a level sufficient for detection under a fluorescence microscope. These processes might happen slowly and thus it might be hard to detect mutant cells at an earlier time point following exposure to ENU due to insufficient expression of revertant EGFP.

At 18 days post-ENU administration, another stop-EGFP mouse was anesthetized and the dorsal skin was scanned for EGFP-expressing mutants. The in vivo imaging experiment revealed a green fluorescent mutant colony containing a single corneocyte on the skin surface (Figure 3-4D). The same area of the epidermis was imaged again at 6 weeks post-ENU administration and the green fluorescent cell lineage was not detected. Although we cannot rule out the possibility that we accidentally missed the cell lineage in the second imaging, we speculate that the cell lineage might have either originated from a transit amplifying (TA) cell, which disappeared after epidermal turnover, or represent a stem cell lineage that was lost (this might occur if stem cell turnover is a

stochastic birth-death process, for example). Kameda et al. reported similar observations (Kameda et al., 2003). They found small epidermal cell lineages at 2 – 4 weeks following infection with *lacZ*-carrying retrovirus but did not observe such small cell lineages more than 10 weeks after the retroviral infection. They suggested that these small cell lineages had originated from transit amplifying cells and not stem cells, and disappeared after tissue turnover (Kameda et al., 2003).

The epidermal cell lineage observed at 13 weeks post-ENU administration (shown in Figure 3-4B) was detected again when the same area of the skin was re-imaged at 16 weeks post-ENU administration.

3. Dynamics of epidermal proliferation studied by repeated imaging

In the first in vivo imaging experiment using five ENU-treated stop-EGFP mice, both of the two stem cell clonal lineages we detected contained three adjacent corneocytes (i.e., epidermal cells belonging to three adjacent EPUs) (see Figures 3-4A and B). The question can therefore be raised whether three EPUs constitute the actual size of the proliferative compartment that a single epidermal stem cell contributes to. To investigate whether epidermal stem cells always provide epidermal cells for three EPUs, we replicated the in vivo imaging experiment using 11 ENU-treated stop-EGFP mice. The in vivo imaging of the dorsal skin of the 11 mice performed at 8 weeks post-ENU administration (i.e., after roughly four epidermal turnovers) revealed 9 epidermal clonal cell lineages containing 1 – 4 adjacent corneocytes, showing that the number of EPUs that a single stem cell can contribute to can be larger than three. The numbers of cell lineages observed containing different numbers of adjacent corneocytes are listed in Table 3-1. The proportion of clonal cell lineages that contain multiple adjacent

corneocytes (which contradict the classical EPU model) is therefore 44% versus 56% of clonal cell lineages that contain single corneocytes (which fit the classical EPU model).

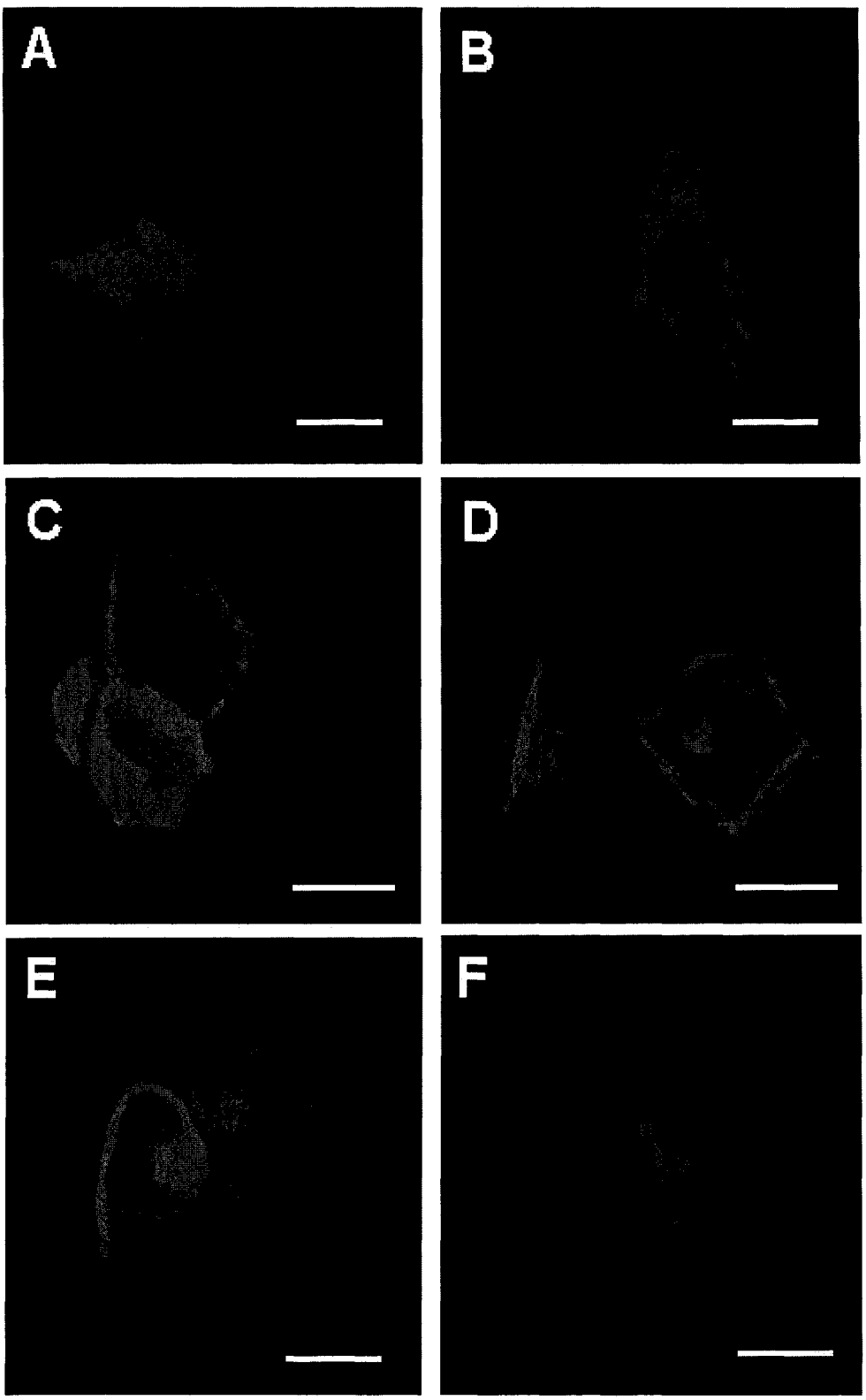
Table 3-1. In vivo imaging of the dorsal epidermis of 11 stop-EGFP mice performed at 8 weeks post-ENU administration.

	One corneocyte	Two corneocytes	Three corneocytes	Four corneocytes
Number of cell lineages found	5	1	2	1

One important advantage of using our stop-EGFP system in cell lineage studies is that the system can be used for repeated analyses of the same clonal cell lineages over time, which could allow one to study the dynamics of the clonal development of a particular cell lineage. We have performed repeated in vivo imaging of the same clonal cell lineages in the epidermis over time to investigate the process of epidermal proliferation from clonal cell lineages. Selected clonal cell lineages were re-imaged four days after they were initially detected (Figure 3-7). During the intervening four day period some epidermal cell lineages expanded while others reduced or maintain their sizes. For example, one cell lineage (Figure 3-7A) expanded from containing three adjacent corneocytes to four adjacent corneocytes (Figure 3-7B), while another cell lineage (Figure 3-7C) contracted from containing three adjacent corneocytes to two

Figure 3-7. Repeated analyses of the same cell lineages in the epidermis over time.

The first in vivo imaging was performed at 8 weeks post-ENU administration. The imaging study revealed epidermal clonal cell lineages containing multiple adjacent corneocytes (A, C, and E). Four days later, the second imaging of the same cell lineages was performed in vivo (B, D, and F, respectively). Scale bars, 20 μm.



adjacent corneocytes (Figure 3-7D). Cell lineages were also observed whose sizes did not change during the 4 day period (Figures 3-7E and F). The success of the repeated imaging experiment which followed the same clonal cell lineages over time suggests that the stop- EGFP system could be efficiently applied to study the dynamics of clonal development of particular cell lineages.

The observation that some cell lineages contracted (e.g., shrinking from containing three to two adjacent corneocytes) suggests that the apparent expansion of clonal cell lineages to multiple EPUs is not likely to be an artifact due to intercellular diffusion of the EGFP protein. If diffusion of the EGFP protein were the actual reason that expanded clonal cell lineages were observed (rather than migration of epidermal cells to adjacent EPUs), then repeated imaging might be expected to show continuous expansion of cell lineages to adjacent EPUs. Furthermore, the EGFP protein has been widely used as a versatile cell lineage marker in a variety of studies and secretion or diffusion of the protein to adjacent cells has not been thus far reported (Yu et al., 2003; Zernicka-Goetz, 1999).

4. In vivo imaging of the epidermis of non-treated stop EGFP mice

The finding of stem cell clonal lineages containing multiple adjacent corneocytes strongly suggests that a single stem cell can provide epidermal cells beyond its own EPU, which contradicts the classical EPU model. However, the question can be raised whether the stem cell lineages we observed reflect the characteristics of epidermal stem cells under the normal steady-state condition. Treatment of cells with mutagenic and cytotoxic ENU, as was done in our earlier studies, might alter the local cellular environment of

epidermis or induce hyperproliferation of epidermal cells, in which case the normal behavior of stem cells might be altered (Winton, 1997; Balling, 2001).

To test whether expanded epidermal clonal cell lineages containing multiple adjacent corneocytes can be observed in stop-EGFP mice in the absence of treatment with ENU, in vivo imaging of the dorsal epidermis was performed using untreated stop-EGFP mice. In this case, any green fluorescent clonal cell lineages that are observed will have been generated by spontaneous mutations occurring at the premature stop codon within the stop-EGFP gene. Such mutations will occur at an extremely low frequency, thus imaging of many mice is required to detect clonal cell lineages in the tissue.

In vivo imaging of the dorsal epidermis of 26 untreated stop-EGFP mice revealed four epidermal clonal cell lineages containing multiple adjacent corneocytes (see Figures 3-8A and B). These observations strongly suggest that under the normal steady-state condition, epidermal cells can migrate horizontally between EPUs, which is counter to the classical EPU model (Potten, 1974). The clonal cell lineage shown in Figure 3-8B, in particular, reveals 5 (or possibly six) adjacent corneocytes (i.e., epidermal cells belonging to five, or possibly six, adjacent EPUs) which exceeds even the numbers of adjacent corneocytes in clonal lineages found in the previous in vivo imaging studies using ENU-treated stop-EGFP mice. In those studies, the largest clonal cell lineage observed appeared to have expanded to four EPUs (Ro, 2004). The in vivo imaging experiment using 26 untreated stop-EGFP mice also revealed single columns of corneocytes which better fit the classical EPU model (Figure 3-8D and E). In total, six green fluorescent patches containing single columns of corneocytes were observed (Table 3-2). The proportion of clonal cell lineages that contain multiple adjacent corneocytes (which

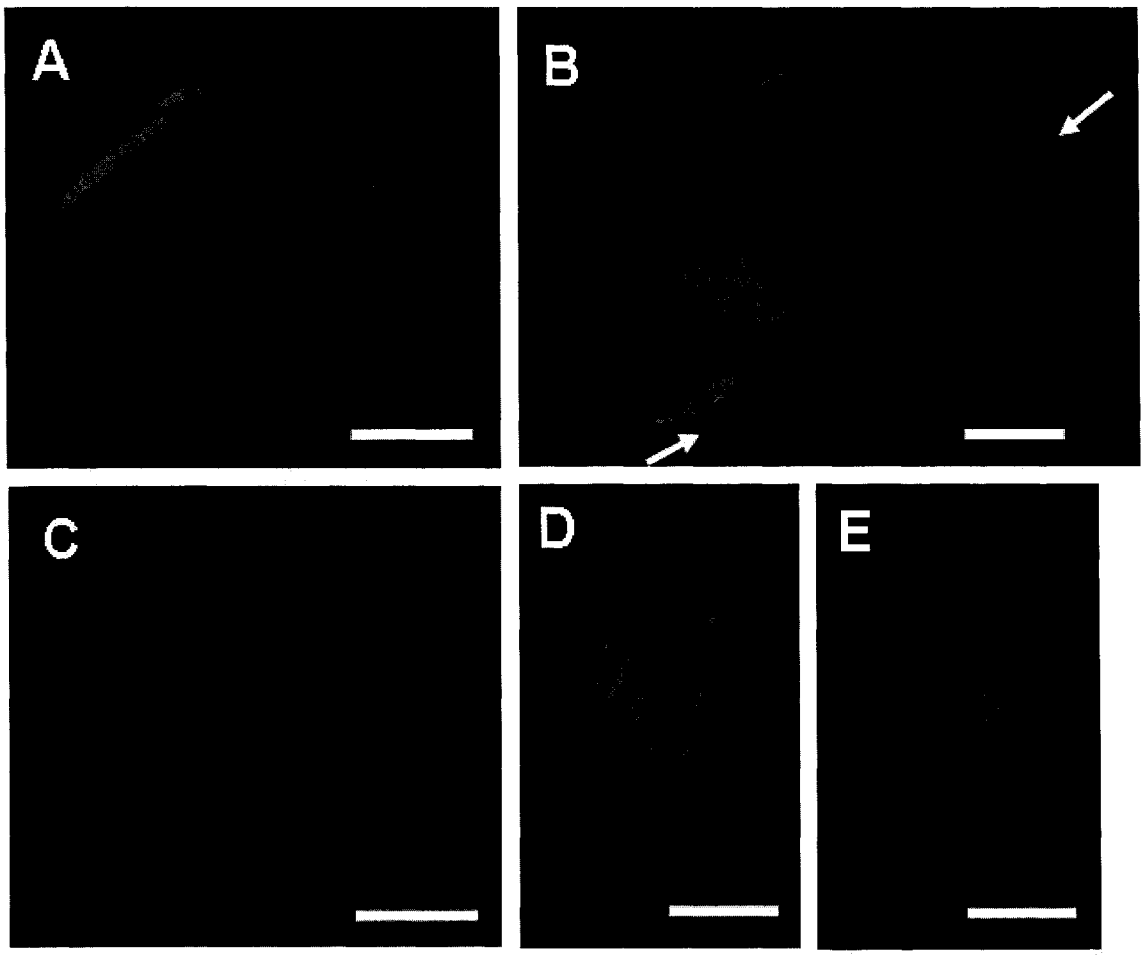


Figure 3-8. In vivo imaging of epidermal clonal cell lineages found in the dorsal epidermis of untreated stop-EGFP mice.

(A and B) The in vivo imaging experiment revealed epidermal clonal cell lineages containing multiple adjacent corneocytes. In panel B, the morphology of corneocytes located lower is not completely seen due to the line generated by skin fold (indicated by arrows). (C) Longitudinal optical sectioning detected no green fluorescent signals at deeper layers below the three corneocytes shown in panel A. (D and E) Single columns of corneocytes which better fit the EPU model were also detected. Scale bars, 20 μm.

contradict the classical EPU model) is therefore 40% versus 60% of clonal cell lineages that contain single corneocytes (which fit the classical EPU model) (see Table 3-2). Thus, we suggest that clonal expansion of cell lineages to multiple neighboring EPUs is a normal feature of epidermal cells under the steady-state condition.

Table 3-2. In vivo imaging of the dorsal epidermis of 26 untreated stop-EGFP mice.

	single corneocyte	**two corneocytes**	**three corneocytes**	**five (or six) corneocytes**
Number of cell lineages found	6	2	1	1

The clonality of epidermal cell lineages found in our in vivo imaging experiment is guaranteed because cells were labeled by means of random and very infrequent spontaneous mutations. The green fluorescent cell lineages that we observed in this study (using untreated stop-EGFP mice) are unlikely to be stem cell clonal lineages because longitudinal optical sectioning of the cell lineages detected no green fluorescent signals below the cornified layer (see Figure 3-8C). Although the clonal cell lineages containing multiple adjacent corneocytes found in this study appear to have originated from mutant TA cells, rather than mutant stem cells, the finding still implies that a single stem cell can contribute to multiple EPUs because a TA cell that has undergone a mutation must originate from a single stem cell (i.e., a TA cell can always be traced back to a single stem cell).

IV. Discussion

The results presented here support the utility of the stop-EGFP mouse as a system for investigating clonal cell lineages. The system has been successfully applied to tracing clonal cell lineages that originated from single epidermal stem cells and its utility is expected to extend to other renewing tissues such as the gastrointestinal epithelial tissues. Although various cell fate mapping techniques have been developed, tracing of clonal cell lineages is rarely achieved in general. Furthermore, in most cases existing techniques for tracing cell lineages suffer from problems such as signal dilution by cell division or perturbation of normal cellular processes. Due to the versatility of EGFP as an in vivo cell lineage marker (see section I-2, Chapter 1), the stop-EGFP mouse system has advantages over other mapping techniques in that it allows clonal cell lineages that descended from single cells to be traced in vivo and also achieves permanent labeling of cells via genetic marking with minimal invasiveness.

In vivo imaging of the dorsal epidermis of ENU-treated stop-EGFP mice (after several epidermal turnovers) revealed that epidermal cells can migrate horizontally to adjacent EPUs and a single stem cell can contribute to multiple EPUs, contradicting the classical EPU model. Nevertheless, an important question is whether the stem cell lineages we observed reflect the characteristics of epidermal stem cells under the normal steady-state condition. Treatment of cells with mutagenic and cytotoxic ENU, as was done in our earlier studies, might alter the local cellular environment of epidermis or induce hyperproliferation of epidermal cells. Observations of epidermal clonal cell lineages larger than a single EPU have been reported previously in circumstances where hyperproliferation in the mouse epidermis was induced by epidermal tissue

transplantation or by chronic irradiation with 280 – 320 nm UV (Mackenzie, 1997; Zhang et al., 2001). Cell lineages generated in the UV-induced hyperproliferative epidermis show clonal expansion of basal cells to as many as 100 adjacent EPUs (Zhang et al., 2001). Furthermore, following skin wounding, epidermal cells become hyperproliferative and migrate to wound sites beyond the EPU boundary (Christophers, 1971; Martin, 1997). Furthermore, it has been recently reported that upper follicular cells could migrate to the epidermis and differentiate into epidermal cells in response to a penetrating wound (Taylor et al., 2000).

ENU is a mutagenic and cytotoxic agent (Balling, 2001) and treatment with the mutagen might therefore be expected to induce some cell death with subsequent hyperproliferation of epidermal cells leading to an artificial expansion of stem cell clonal lineages. In the epithelium of the small intestine, for example, treatment with a mutagen led to an increase in the mean size of clonal cell lineages by comparison with those found in untreated normal tissue (Winton, 1997). Furthermore, it has been shown that application of a carcinogen can induce transient epidermal hyperplasia and affect the tissue homeostasis significantly (Elgjo, 1968; Skjaeggestad, 1964). Thus, it is possible that our previous observations that stem cell clonal lineages expanded to multiple EPUs may have been induced by treatment with ENU (Ro, 2004; Ro and Rannala, 2004).

However, our study using untreated stop-EGFP mice (see section III-4) identified clonal cell lineages containing multiple adjacent corneocytes (i.e., epidermal cells belonging to multiple EPUs) strongly suggesting that the expansion of clonal cell lineages to multiple EPUs arises under the normal steady-steady condition as well as under hyperproliferative states in the epidermis. Based on these findings, we propose an

alternative to the classical EPU model (Figure 3-9). Our "Niche-Sharing" model of EPU structure specifies that under the normal steady-state condition a single stem cell not only maintains its own EPU, but also can contribute to adjacent EPUs by providing differentiated cells. In other words, the actual proliferative compartment in the epidermis may be larger than a single EPU. In this model, several EPUs constitute a larger proliferative compartment whereby multiple stem cells of different EPUs cooperatively provide epidermal cells for the compartment as a whole (Figure 3-9B).

The characteristic morphology of large hexagonal cells detected in the outermost layer of the epidermis could verify their identity (i.e., corneocytes), and longitudinal optical sectioning using a confocal laser scanning microscope can reveal the distribution of an EGFP-labeled clonal cell lineage in the epidermis. To better reveal the distribution of a cell lineage and cell identities, an immunohistochemical assay using antibodies against cell type-specific markers can be carried out. Tissue sections analyzed immunohistochemically might also better reveal the EPU and surrounding tissue structures in the epidermis. Further investigation of EGFP-labeled clonal cell lineages expanding to multiple EPUs will be carried out in the near future using an immunohistochemical analysis.

The stop-EGFP transgenic mouse system could be applied to other tissues that undergo renewal (e.g., epithelial tissues of the gastrointestinal tract) or to tissues with regenerative capabilities. For example, if liver regeneration in an ENU-treated stop-EGFP mouse is induced following partial hepatectomy (Michalopoulos and DeFrances, 1997) or treatment of hepatotoxic chemicals (Alison et al., 1998), clonal cell lineages

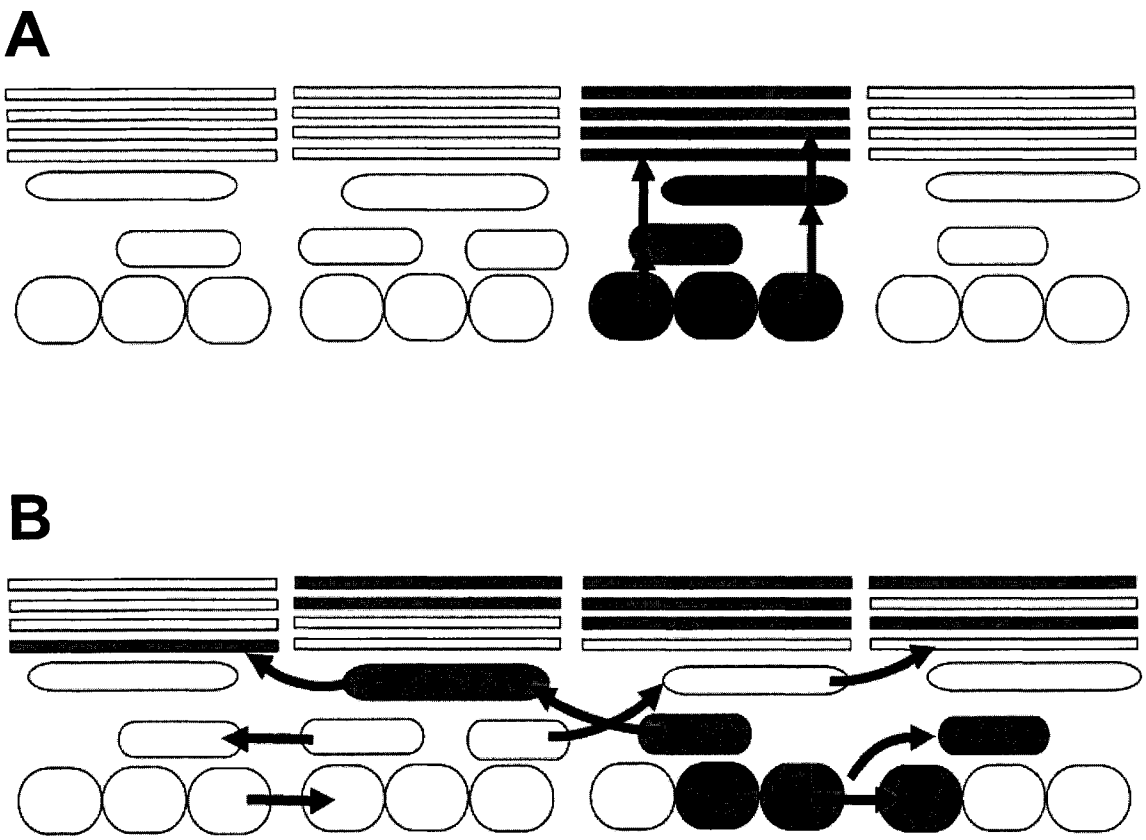


Figure 3-9. Comparison of the classical EPU model and the "Niche-Sharing" model of EPUs.

Cells in green represent a clonal cell lineage that originated from an epidermal basal stem cell under each model. (A) The classical EPU model identifies an EPU as a clonal unit of cellular proliferation and no horizontal migration between EPUs is expected. Thus, cellular migration occurs only within an EPU, mainly vertical migration from the basal layer to the cornified layer for terminal differentiation (see arrows). (B) The Niche-Sharing model proposes that epidermal cells can migrate to adjacent EPUs and thus a single stem cell can contribute to multiple EPUs. Arrows indicate horizontal migration of epidermal cells among EPUs.

could be detected in the regenerated liver that might provide insights into the liver regeneration process.

By repeated imaging of the same clonal cell lineages over time, we were able to observe the dynamics of the clonal development of epidermal cell lineages (see Figure 3-7). The success of the repeated imaging experiment supports the potential of the stop-EGFP system to be applied to various studies such as those focused on the development of clonal cell lineages, migration of cells that descended from a single cell, differentiation processes from single progenitor cells, etc.

One potential weakness of the stop-EGFP mouse is that the mutation rate at the premature stop codon within the stop-EGFP gene appears to be low even after treatment with a potent mutagen, ENU. This reduces the efficiency with which labeled cells are generated via mutation in the stop-EGFP system. Imaging of a 2.5 cm × 2.5 cm area of an ENU-treated stop-EGFP mouse reveals one or fewer stem cell lineages on average. Thus, imaging of many mice might be required to detect clonal cell lineages in a tissue of interest. Repeated treatments with ENU or another mutagen could help to increase the expected frequency of mutations at the premature stop codon. Another strategy is to create a transgenic mouse carrying a very high copy number of the stop-EGFP gene. Alternatively, a transgenic mouse carrying the wild-type EGFP gene under the control of a strong repressor could be generated in which a mutation at the repressor gene leads to expression of EGFP.

V. References

Alison M, Golding M, Lalani el-N, Sarraf C. (1998) Wound healing in the liver with particular reference to stem cells. *Philos Trans R Soc Lond B Biol Sci.* 353: 877-894.

Allen TD, Potten CS. (1974) Fine-structural identification and organization of the epidermal proliferative unit. *J Cell Sci.* 15: 291-319.

Balling R. (2001) ENU mutagenesis: analyzing gene function in mice. *Annu Rev Genomics Hum Genet.* 2: 463-492.

Bussau LJ, Vo LT, Delaney PM, Papworth GD, Barkla DH, King RG. (1998) Fibre optic confocal imaging (FOCI) of keratinocytes, blood vessels and nerves in hairless mouse skin in vivo. *J Anat.* 192: 187-194.

Christophers E. (1971) The architecture of stratum corneum after wounding. *J Invest Dermatol.* 57: 241-246.

Elgjo K. (1968) Epidermal cell population kinetics after a single application of some hyperplasia-producing substances. *Eur J Cancer.* 3: 519-530.

Ghazizadeh S, Taichman LB. (2001) Multiple classes of stem cells in cutaneous epithelium: a lineage analysis of adult mouse skin. *EMBO J.* 20: 1215-1222.

Izzo AD, Mackanos MA, Beckham JT, Jansen ED. (2001) In vivo optical imaging of expression of vascular endothelial growth factor following laser incision in skin. *Lasers Surg Med.* 29: 343-350.

Kameda T, Nakata A, Mizutani T, Terada K, Iba H, Sugiyama T. (2003) Analysis of the cellular heterogeneity in the basal layer of mouse ear epidermis: an approach from partial decomposition in vitro and retroviral cell marking in vivo. *Exp Cell Res.* 283: 167-183.

Mackenzie IC. (1970) Relationship between mitosis and the ordered structure of the stratum corneum in mouse epidermis. *Nature.* 226: 653-655.

Mackenzie IC. (1997) Retroviral transduction of murine epidermal stem cells demonstrates clonal units of epidermal structure. *J Invest Dermatol.* 109: 377-383.

Martin P. (1997) Wound healing--aiming for perfect skin regeneration. *Science*. 276: 75-81.

Michalopoulos GK, DeFrances MC. (1997) Liver regeneration. *Science*. 276: 60-66.

Potten CS. (1974) The epidermal proliferative unit: the possible role of the central basal cell. *Cell Tissue Kinet*. 7: 77-88.

Potten CS. (1975) Epidermal transit times. *Br J Dermatol*. 93: 649-658.

Potten CS. (1981) Cell replacement in epidermis (keratopoiesis) via discrete units of proliferation. *Int Rev Cytol*. 69: 271-318.

Potten CS. (2004) Keratinocyte stem cells, label-retaining cells and possible genome protection mechanisms. *J Investig Dermatol Symp Proc*. 9: 183-195.

Potten CS, Saffhill R, Maibach HI. (1987) Measurement of the transit time for cells through the epidermis and stratum corneum of the mouse and guinea-pig. *Cell Tissue Kinet*. 20: 461-472.

Potten CS, Morris RJ. (1988) Epithelial stem cells in vivo. *J Cell Sci*. 10 (Suppl): 45-62.

Ro S. (2004) Magnifying stem cell lineages: the stop-EGFP mouse. *Cell Cycle*. 3: 1246-1249.

Ro S, Rannala B. (2004) A stop-EGFP transgenic mouse to detect clonal cell lineages generated by mutation. *EMBO Rep*. 5: 914-920.

Ro S, Rannala B. (2005) Evidence from the stop-EGFP mouse supports a niche-sharing model of epidermal proliferative units. *Exp Dermatol*. 14: 838-843.

Schmidt GH, Blount MA, Ponder BA. (1987) Immunochemical demonstration of the clonal organization of chimaeric mouse epidermis. *Development*. 100: 535-541.

Schneider TE, Barland C, Alex AM, Mancianti ML, Lu Y, Cleaver JE, Lawrence HJ, Ghadially R. (2003) Measuring stem cell frequency in epidermis: a quantitative in vivo functional assay for long-term repopulating cells. *Proc Natl Acad Sci U S A.* 100: 11412-11417.

Skjaeggestad O. (1964) Experimental epidermal hyperplasia in mice: relation to carcinogenesis. *Acta Pathol Microbiol Scand.* 169 (Suppl): 1-126.

Slack JM. (2000) Stem cells in epithelial tissues. *Science.* 287: 1431-1433.

Taylor G, Lehrer MS, Jensen PJ, Sun TT, Lavker RM. (2000) Involvement of follicular stem cells in forming not only the follicle but also the epidermis. *Cell.* 102: 451-461.

Turksen K, Troy TC. (1998) Epidermal cell lineage. *Biochem Cell Biol.* 76: 889-898.

Winton DJ. (1997) Intestinal stem cells and clonality. In: *The Gut as a Model in Cell and Molecular Biology* (Halter F, Winton D, Wright NA, eds). Lancaster: Kluwer Academic Publishers, pp. 3-13.

Yu YA, Oberg K, Wang G, and Szalay AA. (2003) Visualization of molecular and cellular events with green fluorescent proteins in developing embryos: a review. *Luminescence.* 18: 1-18.

Zernicka-Goetz M. (1999) Green fluorescent protein: a new approach to understanding spatial patterning and cell fate in early mammalian development. In: *Cell Lineage and Fate Determination* (Moody SA, ed). San Diego: Academic Press, pp. 521-527.

Zhang W, Remenyik E, Zelterman D, Brash DE, Wikonkal NM. (2001) Escaping the stem cell compartment: sustained UVB exposure allows p53-mutant keratinocytes to colonize adjacent epidermal proliferating units without incurring additional mutations. *Proc Natl Acad Sci USA.* 98: 13948-13953.

Chapter 4. Inferring Somatic Mutation Rates Using the Stop-EGFP Mouse

Parts of this chapter have been previously published or submitted.

Ro S, Rannala B. (2004) A stop-EGFP transgenic mouse to detect clonal cell lineages generated by mutation. *EMBO Rep.* 5: 914-920.

Ro S, Rannala. (2005) Inferring somatic mutation rates using the stop-EGFP mouse. (manuscript submitted)

I. Introduction

Somatic mutation is a process of fundamental importance in many human diseases such as cancer; it also is a key component in biological processes such as aging (Gorbunova and Seluanov, 2005; Sedelnikova et al., 2004). Many factors influence the rates (and patterns) of somatic mutation. Some influences are environmental (e.g., exposure to chemicals and UV-radiation, see sections III-2 and III-3, Chapter 1) and others are genetic (e.g., GC content of DNA sequences, mutations in DNA repair genes, etc, see sections III-4 and III-5, Chapter 1). To uncover the important factors influencing rates of somatic mutation (and by extension, rates of cancer, etc) refined techniques are needed to estimate rates of somatic mutation in cells exposed to potential environmental, or genetic, risk factors. In vitro assays (e.g., the Ames *Salmonella* test) for evaluating the mutagenicity of a chemical substance are now well established. A limitation of prokaryotic mutagenicity assays, however, is that metabolic enzymes present in mammalian cells that can transform a non-mutagenic into a mutagenic substance (genotoxicity) may not be present in the prokaryotic cell (see section IV-1, Chapter 1). If in vivo mutation detection were found to be feasible, this could offer a potentially rapid, and inexpensive, test for mutagenicity that could be expected to be more sensitive than in vitro methods in evaluating potential carcinogens.

Mouse somatic mutation detection systems have been developed which utilized either the *Dolichos biflorus* (*Dlb-1*) locus, or the hypoxanthine-guanine phosphoribosyl transferase gene (*hprt*), as a mutational target (see sections IV-3 and IV-4, Chapter 1). In the early 1990s, transgenic mouse models for in vivo mutation detection were developed by inserting either the *lacI*, or *lacZ*, bacterial transgene into mice (see section IV-6,

Chapter 1). One shortcoming of in vivo mutation detection systems using bacterial transgenes is that they can only provide mutant frequency (i.e., the relative number of mutants in a population of cells) from which mutation rate often cannot be estimated accurately (see section III-1, Chapter 1). To precisely determine the mutagenic effect of a chemical mutagen, a defective gene, or some other factor, a novel mutation detection system directed at estimating rates of mutation (rather than mutant frequency) is desirable.

The transgenic mouse we developed carries a premature stop codon-containing enhanced green fluorescent protein gene (i.e., the stop-EGFP gene) which can function as a reporter for mutation (Ro and Rannala, 2004). In principle, a cell having undergone a mutation at the premature stop codon within the stop-EGFP gene expresses functional EGFP and thus mutant cells can be identified using fluorescence imaging. In the stop-EGFP system, independent mutations in a tissue or an organ are expected to be counted by detecting green fluorescent colonies. The ability to identify independent mutations should facilitate accurate estimation of mutation rates (see section III-1, Chapter 1).

The first aim of this chapter is to verify that the stop-EGFP system can be utilized as a novel in vivo mutation detection system with which one can quantify the number of independent mutations. For this purpose, ENU (*N*-ethyl-*N*-nitrosourea) is used as an agent to induce mutations. We treated the stop-EGFP mouse (line F#19-15) with ENU at a dose of 150 mg/kg body weight and the number of independent mutations having arisen in each lobe of the liver was counted at 5 months post-ENU administration.

Mutagenesis in live mammals is a complex process which involves DNA damage incurred by exposure to a mutagen, fixation of mutations, expression of mutant

phenotypes, etc (Heddle, 1999). Most of these processes happen slowly and maximum mutant frequency could be observed as long as several months after treatment with a mutagen depending on tissues (Douglas et al., 1995; Sun and Heddle, 1999). In the liver where tissue turnover is very slow, it has been observed that mutant frequency continuously increases over two months after treatment with ENU (Wang et al., 2004) while tissues with a fast turnover such as bone marrow reach the maximal mutant frequency less than a week (Douglas et al., 1996). It is suggested that one of the most important variables in mutagenesis experiments in vivo is the time between treatment and mutation assay (Heddle, 1999). Time-course studies have been carried out to determine the time points of the maximum mutant frequencies for different tissues and different mutation assay systems. The *Dlb-1* mutation detection system shows the maximum mutant frequency in the small intestine at about 2 weeks post-ENU administration (Sun et al., 1999). A time-course study carried out using the *hprt* mutation detection system indicated that the times (following exposure to ENU) needed to reach the maximum mutant frequency in the splenic T cells were variable depending on animal age at treatment (Walker et al., 1999). In the mammary gland and germ-cells of the *lacZ* transgenic mouse, the maximum mutant frequencies were observed at 4 and 5 weeks, respectively after ENU treatment (Douglas et al., 1995; Sun et al., 1999). Because sampling times can significantly affect the mutant frequency observed in a given tissue (Mirsalis et al., 1993), to efficiently determine the mutagenic effect of a substance it is recommended that samples are taken at the time point when the highest mutant frequency is expected to be observed (Heddle et al., 2003). The times (following exposure to a mutagen) of the maximum mutant frequency, however, are hard to know in advance

because they could be varied depending on the tissues and mutation detection systems used (Thybaud et al., 2003). Thus, multiple sampling times are recommended to identify the maximum mutant frequency in a tissue after treatment with a mutagen (Heddle et al., 2003).

In the stop-EGFP mouse treated with ENU, DNA adducts generated by exposure to the mutagen will be fixed as mutations during subsequent cell divisions. Expression of the mutant phenotype will require transcription from a mutant copy of the stop-EGFP gene, accumulation of mutant mRNA (containing no premature stop codon), and finally protein synthesis from the mutant mRNA to a level high enough for detection under a fluorescence microscope. Thus, a mutation assay carried out at an early time point might underestimate the actual mutation frequency induced by a mutagen due to insufficient expression of revertant EGFP from some mutant cells which are undetectable under a fluorescence microscope.

The second aim of this chapter is to accurately estimate the rate of somatic mutations induced by exposure to ENU taking into account the lag time until expression of the mutant phenotype using an explicit statistical model and the data from a time-course study of mutations. Twelve stop-EGFP mice (line F#19-15) were sacrificed at different time points after ENU administration, and changes in the number of mutations over time were investigated in the left caudal lobe of the liver.

Finally, we tested the mutagenic effect of ENU in the liver using the stop-EGFP mouse system. For this purpose, ten stop-EGFP mice were treated with vehicle and the mutation frequency in the control group was compared with that in the ENU-treated group.

II. Materials and Methods

(For a description of the preparation and injection of ENU, see Materials and Methods in Chapter 3)

1. Animal treatment and mutation assay in the kidney and the liver

Mice were anesthetized by inhalation of isoflurane gas prior to injection. For the ENU-treated group, ENU was administered intraperitoneally at a dose of 150 mg/kg body weight. Control mice were injected intraperitoneally with vehicle (9.5% ethanol in Phosphate-Citrate buffer, see section II-2, Chapter 3) in a volume of 20 ml/kg body weight. For mutation assay, mouse was euthanized by CO_2 asphyxiation and perfused with 10 ml of saline followed by 10 ml of 4% paraformaldehyde. The liver and kidney were removed from the mouse and were stored in 4% paraformaldehyde at 4 °C with gentle agitation for 11 hours. After fixation, the organs were transferred to PBS with 1 mM $MgCl_2$ and stored at 4 °C overnight. The organs were sectioned into slices (100 μm in thickness) using a vibratome. Each slice was transferred to a 24-well plate containing PBS with 1 mM $MgCl_2$ and stored until the imaging experiment. At least 4 hours before imaging, slices were transferred to microscopic slides for mounting. PBS and glycerol were mixed and used as a mounting media. Slices were illuminated by use of a 50 W mercury lamp and scanned using a Zeiss Axiovert 200M inverted microscope with a 10× F-Fluar lens (NA 0.5) and LP 520 emission filter (Carl Zeiss). Images of green fluorescent cells were collected with a confocal laser scanning microscope (LSM 510 NLO, software version 3.0, Carl Zeiss) mounted on the Zeiss Axiovert 200M inverted microscope with a 25× F-Fluar lens (NA 0.8). The 488 nm laser line was used to image

EGFP with a band pass filter (505 – 530 nm wavelength) for detecting emissions from EGFP.

2. DAPI (4', 6-Diamidino-2-phenylindole) staining and co-localization of nuclei

Several slices containing EGFP signals were selected and used to verify co-localization of DAPI-stained nuclei and bright EGFP signals within mutant cells. Slices were incubated for 30 minutes at room temperature in a PBS solution containing 1 μg/ml of DAPI. Slices were then briefly washed with PBS and dehydrated by storing them in 30, 50, 60, 70, 90, 100% ethanol, serially for 5 minutes at each step. Slices were then mounted in methyl salicylate. The DAPI-stained EGFP-expressing cells were imaged using a confocal laser scanning microscope (LSM 510 NLO, software version 3.0, Carl Zeiss) with an Argon laser (488 nm) to activate EGFP and a two-photon laser of 760 nm to activate DAPI-stained DNA.

3. Estimating the total number of cells in each lobe

Slices were stained with DAPI and dehydrated, and then mounted in methyl salicylate. The total number of nuclei in each slice was estimated by following methods: Randomly chosen 100 μm × 100 μm areas (at least five different areas per slice) were scanned along the z-axis and all nuclei contained in this volume were imaged. To count the total number of nuclei in the specified volume, imaging software (Imaris, version 3.0) was used (and then confirmed by manual counting). The area of each slice was measured using MetaMorph (Universal Imaging Corp.). The total number of cells in each slice was then estimated by multiplying the average number of nuclei in a unit area by the area measured. The total number of cells in each lobe was estimated by multiplying the total

cell number in a reference slice by the total number of slices of a similar size sectioned from the organ. (see section II-4 for statistical methods)

4. Statistical methods and equations to calculate total cell number, mutation frequency and mutation rate

The total cell numbers, mutation frequencies, and mutation rates can be estimated as follows using data from experiments with the stop-EGFP mouse in which an organ is sectioned and mutant cells are counted. The standard deviations and 95% confidence intervals are derived for the estimators based on standard theory for the propagation of errors (Taylor, 1997). Let A_{ij} be the area of microscopic section i from organ lobe j. Let S_j be the total number of sections from lobe j and let I_j be the number of sections from lobe j for which the area (in units of μm^2) is measured, where $S_j > I_j$. Let C_{ijk} be the number of cells counted in the k^{th} of K_{ij} areas of a specified size randomly sampled from section i of lobe j, where K_{ij} is the number of areas sampled from section i of lobe j. Let M_j be the total number of mutations detected in lobe j. The mean area of the sections from lobe j is

$$\overline{A}_j = \frac{1}{I_j}\sum_{i=1}^{I_j} A_{ij},$$

and the standard deviation is

$$\sigma_{\overline{A}_j} = \sqrt{\frac{1}{I_j}\sum_{i=1}^{I_j}(\overline{A}_j - A_{ij})^2}.$$

An estimate of the total area of lobe j is

$$A_j = \overline{A}_j \times S_j,$$

and the standard deviation is

$$\sigma_{A_j} = A_j \times \frac{\sigma_{\overline{A}_j}}{\overline{A}_j}.$$

The mean number of cells in a specified area of section i of lobe j is

$$\overline{C}_{ij} = \frac{1}{K_{ij}} \sum_{k=1}^{K_{ij}} C_{ijk},$$

and the standard deviation is

$$\sigma_{\overline{C}_{ij}} = \sqrt{\frac{1}{K_{ij}} \sum_{k=1}^{K_{ij}} (\overline{C}_{ij} - C_{ijk})^2}.$$

An estimate of the total number of cells in lobe j is

$$C_j = \overline{C}_{ij} \times \frac{1}{10^4 \mu m^2} \times A_j.$$

This equation assumes there is no variation in mean cell densities among sections within a lobe. The 95% confidence interval for the estimate of C_j is (C_j - 1.96 × σC_j, C_j + 1.96 × σC_j) where

$$\sigma_{C_j} = \left(\frac{\sigma_{A_j}}{A_j} + \frac{\sigma_{\overline{C}_{ij}}}{\overline{C}_{ij}} \right) \times C_j.$$

An estimate of the mutation frequency is $f_j = M_j / C_j$. The 95% confidence interval for the estimated mutation frequency is (f_j - 1.96 × σf_j, f_j + 1.96 × σf_j) where

$$\sigma_{f_j} = \left(\frac{\sigma_{A_j}}{A_j} + \frac{\sigma_{\overline{C}_{ij}}}{\overline{C}_{ij}} \right) \times \frac{M_j}{C_j}.$$

If interested in estimating the mutation rate r_j, the estimator $r_j = M_j / C_j$ is still used but the percent standard deviation is increased by a term equal to the inverse of the square root of M_j so that

$$\sigma_{r_j} = \left(\frac{1}{\sqrt{M_j}} + \frac{\sigma_{A_j}}{A_j} + \frac{\sigma_{\overline{C}_{ij}}}{\overline{C}_{ij}} \right) \times \frac{M_j}{C_j}.$$

III. Results

1. Mutation assay in four lobes of the liver of an ENU-treated stop-EGFP mouse

To verify that the stop-EGFP system can be utilized for quantifying the number of independent mutations having arisen in vivo, we treated the stop-EGFP mouse (line F#19-15) with ENU (*N*-ethyl-*N*-nitrosourea) at a dose of 150 mg/kg body weight. ENU was used as an agent to induce mutations in an attempt to increase the efficiency with which mutations are generated and thus to facilitate the process of counting mutations.

The liver of the stop-EGFP mouse from line F#19-15 was analyzed at 5 months post-ENU administration to detect green fluorescent mutants. In a scan of the entire liver of an ENU-treated stop-EGFP mouse, a total of 26 independent mutations were identified (Figure 4-1) demonstrating that the stop-EGFP mouse system can be used to count independent mutations having arisen in an entire organ in vivo. The study supports the utility of the stop-EGFP system as a novel in vivo mutation detection system.

The mutant cells contained what appeared to be brighter nuclei. The bright spots were confirmed to be nuclei by use of DAPI (4', 6-Diamidino-2-phenylindole) staining (Figures 4-1A and B). Most of the mutant colonies (24 mutants) displayed two nuclei. The fact that many hepatocytes are present as binucleate cells might explain the observed predominance of the two-nucleus mutants in the liver (Alison et al., 2004). Hepatocytes are normally proliferatively quiescent without induced regeneration (Sell, 2001). In this study, a mutant colony carrying more than two nuclei has not been detected at 5 months post-ENU administration and this may reflect the slow rate of cell turnover in this tissue.

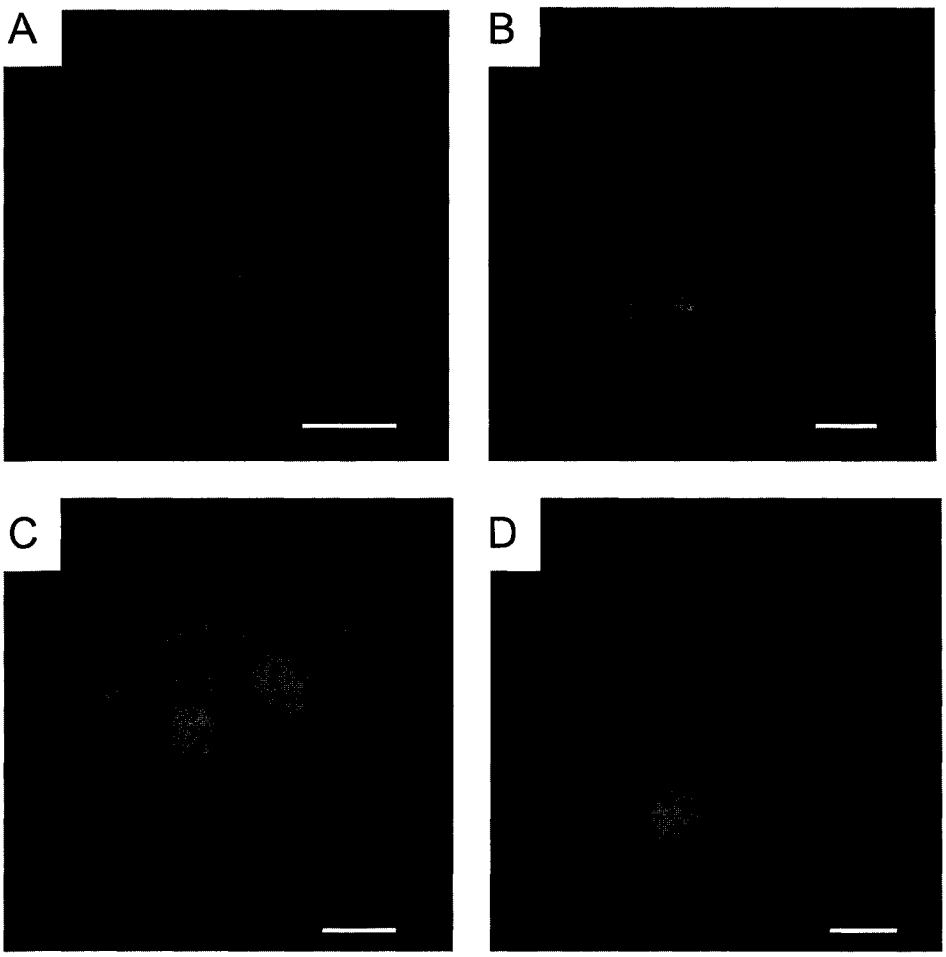


Figure 4-1. Green fluorescent mutant cells in the liver.

Panels A and B show merged images of EGFP signals and DAPI signals of mutant cells detected in the liver. The bright green fluorescent signals within these cells are co-localized with the DAPI-stained nuclei. In the liver, a total of 26 independent mutations were identified. Panels C and D show images of EGFP signals of mutant hepatocytes. Scale bars, 10 μm.

2. Estimation of the rate of mutation induced by ENU in the liver

To verify that somatic mutation rates can be easily estimated using the number of independent mutations obtained from the stop-EGFP mouse system, the liver of an ENU-exposed stop-EGFP mouse was divided into 4 lobes: left caudal, right anterior, left anterior, and right caudal. Then, the number of independent mutation events in each lobe was counted and used to estimate the rate of mutation induced in the lobe by treatment with ENU. The mutation rates were compared to test whether there are significant differences in rates of mutation induced by ENU among the lobes. There have been reports that suggest certain mutagens might be more mutagenic to some lobes than others (Lawson and Pound, 1974; Singh et al., 2001). The fact that in some cases tumors induced by treatment with a mutagen arise preferentially in specific lobes also supports the idea that each lobe might have a different response to certain mutagens (Richardson et al., 1986).

The number of mutations, and estimated total number of cells, in each lobe are shown in Table 4-1. The estimated mutation rate induced by ENU-exposure for the entire liver was $(1.87 \pm 0.27) \times 10^{-7}$ per cell. No significant differences in mutation rates were found among the different liver lobes (although mutation frequencies were significantly different in some cases). If mutations occur according to a Poisson process, the percent standard error of the estimated mutation rate is greater than the percent standard error of the mutation frequency by a factor equal to the inverse square root of the number of mutations observed. Thus, it is quite possible that mutation frequencies can be significantly different among tissues while the underlying mutation rates are identical.

Table 4-1. Observed distribution of stop-EGFP revertant mutations in four lobes of the liver.

Lobe	Mutation Count	Cell Count	Mutation Rate
Left caudal	12	$(6.61 \pm 2.32) \times 10^7$	$(1.82 \pm 3.04) \times 10^{-7}$
Right anterior	6	$(3.06 \pm 0.84) \times 10^7$	$(1.96 \pm 0.25) \times 10^{-7}$
Left anterior	1	$(2.04 \pm 0.32) \times 10^7$	$(0.49 \pm 0.25) \times 10^{-7}$
Right caudal	7	$(2.18 \pm 0.43) \times 10^7$	$(3.21 \pm 0.84) \times 10^{-7}$
All	**26**	$\mathbf{(13.89 \pm 1.99) \times 10^7}$	$\mathbf{(1.87 \pm 0.27) \times 10^{-7}}$

3. Time-course study of mutations in the liver

We investigated how mutation frequency changes over time in the liver of the stop-EGFP mouse (line F#19-15) after treatment with ENU to determine the proper sampling time and more importantly to estimate an unbiased frequency of mutation induced by the mutagen (see next section). Twelve stop-EGFP mice were treated with ENU at a dose of 150 mg/kg body weight. ENU-treated stop-EGFP mice were euthanized at 1, 3, and 7 months post-ENU administration and the number of mutations in the left caudal lobe of the liver from each mouse was counted (Table 4-2). The time course study of mutations in the liver suggests the number of mutations detected in the tissue can increase with time after treatment with ENU. We speculate the time period for one month after treatment with ENU might be too short for mutant cells in the liver to exhibit the mutant phenotype (i.e., the expression level of EGFP revertant is not high enough for detection under a fluorescence microscope). Expression of the mutant phenotype in the liver of the stop-

EGFP mouse treated with ENU will require fixation of a mutation following DNA replication, transcription from a mutant copy of the stop-EGFP gene, accumulation of mutant mRNA containing no premature stop codon, and finally protein synthesis from the mutant mRNA at a level sufficient for detection under a fluorescence microscope. These processes might take months in the liver and thus it is suggested that a mutation assay in the liver of the stop-EGFP mouse should be performed at least three months after treatment with a mutagen to get the maximum number of mutations induced by the mutagen in the tissue and thus to better determine the mutagenic effect of the mutagen.

Table 4-2. Time course study of mutation frequency induced by ENU in the left caudal lobe of the liver.

Months post-ENU treatment	# of mice analyzed	# of mutations in each mouse	# of mutations in total	Mean mutation frequency per lobe (Mean ± SE)
1	4	0, 2, 0, 0	2	0.5 ± 0.50
3	3	2, 3, 7	12	4 ± 1.53
7	5	0, 2, 1, 12, 6	21	4.2 ± 2.20

4. Inferring mutation rates induced by ENU and time to expression of the mutant phenotype

A model is presented to estimate the rate of mutation per cell induced by ENU and the average waiting time until the mutant phenotype is expressed in the liver based on the data presented in Table 4-2.

Let X_i be the total number of cells in mouse i that carry a mutation in the stop-EGFP gene as a consequence of exposure to a strong mutagen at time $T = 0$ (time is measured in units of months). Let Y_i be the number of clonal lineages of cells expressing revertant EGFP that are observed in mouse i that is scanned for mutations at time T_i after mutagen exposure. We assume that mutations in individual cells are rare events (and the total number of cells imaged is large). The process by which mutant cells arise can therefore be modeled as a Poisson process with rate μN where μ is the probability of mutation per cell and N is the total number of cells imaged. The probability distribution of X_i is then

$$\Pr(X_i \mid \mu, N) = \frac{(\mu N)^{X_i} e^{-\mu N}}{X_i!}.$$

It is assumed that a cell is equally likely to begin expressing revertant EGFP at anytime after a mutation arises, with rate λ, and once expression begins it continues indefinitely. The waiting time until expression then follows an exponential distribution,

$$p_T = \lambda e^{-\lambda T},$$

where $1/\lambda$ is the average waiting time until a mutant copy of the stop-EGFP gene is expressed (thus showing the mutant phenotype). The probability that a cell carrying a mutant copy of the stop-EGFP gene is expressing revertant EGFP at (or before) time T is then

$$q_T = \int_0^T \lambda e^{-\lambda t} dt = 1 - e^{-\lambda T}.$$

It is reasonable to assume that the time to expression is an independent random variable when considering cells with independent mutations. The probability distribution of the number of clonal lineages expressing revertant EGFP in the *i*th mouse is then

$$\Pr(Y_i \mid X_i, \lambda) = \binom{X_i}{Y_i}(1-e^{-\lambda T_i})^{Y_i} e^{-\lambda T_i (X_i - Y_i)}.$$

The actual number of mutations, X_i, is an unobserved random variable and so we instead consider the marginal distribution of Y_i,

$$\begin{aligned}\Pr(Y_i \mid \lambda, \mu, N) &= \sum_{X_i=0}^{\infty} \Pr(Y_i \mid X_i, \lambda) \Pr(X_i \mid \mu, N), \\ &= \sum_{X_i=0}^{\infty} \frac{(\mu N)^{X_i} e^{-\mu N}}{X_i!} \binom{X_i}{Y_i}(1-e^{-\lambda T_i})^{Y_i} e^{-\lambda T_i (X_i - Y_i)}, \\ &= \frac{\{\mu N(1-e^{-\lambda T_i})\}^{Y_i} e^{-\mu(1-e^{-\lambda T_i})}}{Y_i!}.\end{aligned}$$

If n mice are imaged in total, $\mathbf{Y} = \{Y_i\}$ is a vector of the counts of numbers of revertant EGFP-expressing clonal lineages per mouse, and $\mathbf{T} = \{T_i\}$ is a vector of the elapsed times since mutagen exposure for the imaged mice, the joint log-likelihood of the data is

$$\log L(\mathbf{Y} \mid \lambda, \mu, N) = \sum_{i=1}^{n} \log \Pr(Y_i \mid \lambda, \mu, N).$$

To jointly maximize this likelihood over λ and μ we need to solve the set of equations,

$$\begin{aligned}\frac{\partial \log L}{\partial \mu} &= \sum_{i=1}^{n} \left\{ \frac{Y_i}{\mu N} N - N(1-e^{-\lambda T_i}) \right\} = 0, \\ \frac{\partial \log L}{\partial \lambda} &= \sum_{i=1}^{n} \left\{ T_i e^{-\lambda T_i} \left(\frac{Y_i}{1-e^{-\lambda T_i}} - \mu N \right) \right\} = 0.\end{aligned}$$

Note that the first equation has an explicit solution,

$$\hat{\mu} = \frac{\sum_{i=1}^{n} Y_i}{N \sum_{i=1}^{n} (1 - e^{-\lambda T_i})}.$$

Substituting this solution for μ into the second equation gives an equation in one unknown,

$$\frac{1}{n} \sum_{i=1}^{n} \left(T_i e^{-\lambda T_i} \left\{ \frac{Y_i}{1 - e^{-\lambda T_i}} - \sum_{j=1}^{n} Y_j / \sum_{j=1}^{n} (1 - e^{-\lambda T_j}) \right\} \right) = 0.$$

The root of this equation in λ can be determined by a simple one-dimensional numerical search, providing maximum likelihood estimates (MLEs) of λ and μ. The variances of the MLEs can be estimated as

$$\mathrm{var}(\hat{\lambda}) \approx \left(-\frac{\partial^2 \log L}{\partial \lambda^2} \right)\Bigg|_{\lambda = \hat{\lambda}}$$

$$\mathrm{var}(\hat{\mu}) \approx \left(-\frac{\partial^2 \log L}{\partial \mu^2} \right)\Bigg|_{\mu = \hat{\mu}}$$

where

$$\frac{\partial^2 \log L}{\partial \lambda^2} = \sum_{i=1}^{n} \left\{ \frac{e^{-\lambda T_i} T_i^2 \left(\left[e^{-\lambda T_i} - 1 \right]^2 \mu N - Y_i e^{-2\lambda T_i} \right)}{\left[e^{-\lambda T_i} - 1 \right]^2} \right\},$$

$$\frac{\partial^2 \log L}{\partial \mu^2} = \frac{\mu^2}{\sum_{i=1}^{n} Y_i}.$$

Using the result that the asymptotic distribution of the MLEs is multivariate normal we use as a 95% confidence interval (CI),

$$\hat{\theta} \pm \sqrt{\mathrm{var}(\hat{\theta})} \times 1.96,$$

where θ is the parameter of interest (e.g., λ or μ). For the mouse data presented in Table 4-1 using $N = 6.6 \times 10^7$ in the left caudal liver lobe, the maximum likelihood estimates and 95% CIs are $1/\lambda = 4.1$ (2.7, 8.0) months and $\mu = 7.7 \times 10^{-8}$ (4.9×10^{-8}, 1.0×10^{-7}) per cell. Thus, the average waiting time until the mutant phenotype is expressed in the liver is about 4 months after treatment with ENU and the mutation rate in the liver of the stop-EGFP mouse treated with ENU at a dose of 150 mg/kg body weight is estimated to be about 8×10^{-8} per cell.

5. Testing the mutagenic effect of ENU in the liver

The mutagenic effect of ENU in the liver was tested by comparing mutation frequency in the left caudal liver lobe between ENU-treated stop-EGFP mice and vehicle-treated control stop-EGFP mice. Because the average waiting time until the mutant phenotype is expressed in the liver is estimated to be 4.1 months after treatment (see section III-4), we performed mutation assay of the control group at 5 months post-vehicle treatment. From imaging of the entire left caudal liver lobe of 10 vehicle-treated stop-EGFP mice, we found 12 mutations in total (i.e., 1.2 mutations per lobe, see Table 4-3). To calculate the mutation frequency induced by ENU in the liver at 5 months post-ENU treatment, the numbers of mutations found at 3 months post-ENU treatment were combined with those found at 7 months post-ENU administration (see Table 4-2). The combined mutation frequency is expected to represent the mutation frequency in the liver at about 5 months post-ENU administration. A total of 33 mutations were found in the left caudal liver lobe of 8 ENU-treated stop-EGFP mice (Table 4-3). The mutation frequency of 4.1 per lobe in ENU-treated group is significantly higher than the mutation frequency in the control group ($p < 0.05$).

Table 4-3. Mean mutation frequency per the left caudal liver lobe in control and ENU-treated mice

A. Control group (treated with vehicle only)

Months post-treatment	# of mice analyzed	# of mutations in each mouse	# of mutations in total	Mean mutation frequency per lobe (Mean ± SE)
5	10	0, 0, 0, 0, 1 1, 2, 2, 2, 4	12	1.2 ± 0.42

B. ENU-treated group (at a dose of 150 mg/kg)*

Months post-treatment	# of mice analyzed	# of mutations in each mouse	# of mutations in total	Mean mutation frequency per lobe (Mean ± SE)
5.5	8	2, 3, 7, 0, 2, 1, 12, 6	33	4.1 ± 1.41

*The numbers of mutations found at 3 and 7 months post-ENU treatment were combined to estimate the mutation frequency induced by ENU at about 5 months post-ENU administration (see Table 4-2)

Treatment of the stop-EGFP mouse with ENU at a dose of 150 mg/kg yielded a 3.4-fold increase in mutation frequency in the liver when compared with control (Figure 4-2). Previous studies using a transgenic mouse mutation detection system carrying a bacterial transgene such as *lacI* or *lacZ* have shown that mutation frequency was elevated 2-fold to 6-fold above background in the liver following treatment of mice with ENU at a

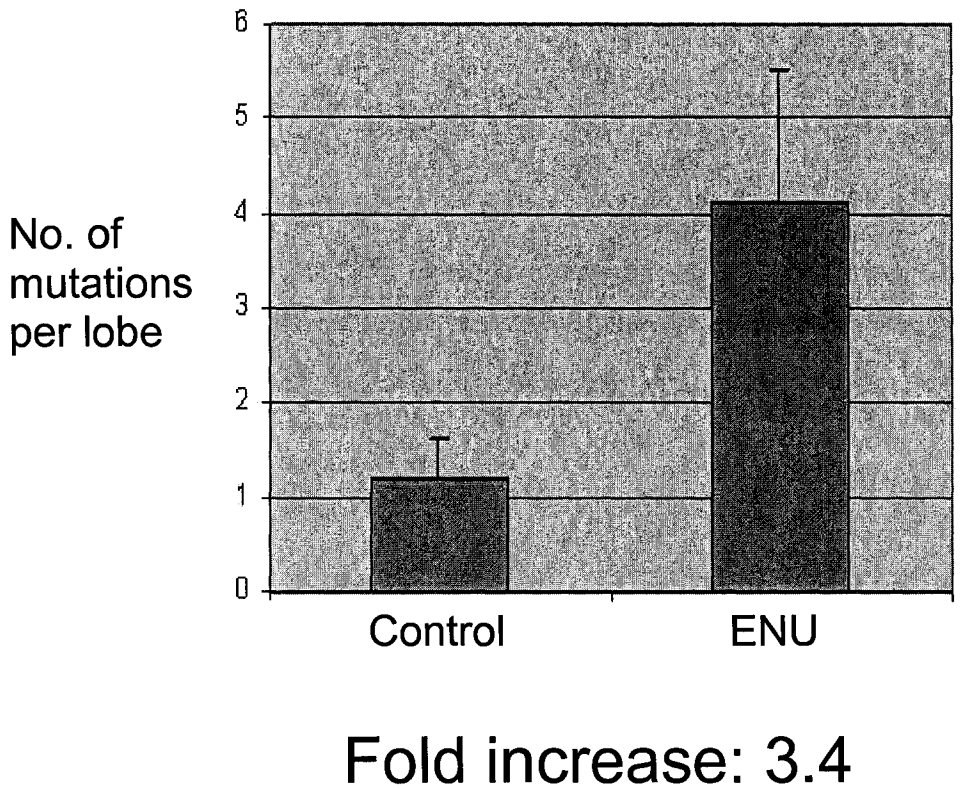


Figure 4-2. Comparison of mutation frequency between control and ENU-treated liver.

Graph indicates the mean mutation frequency per the left caudal liver lobe ± SE. The mean mutation frequency in ENU-treated liver was significantly higher than that in control liver ($p < 0.05$, *t*-test).

dose of 100 – 250 mg/kg (Table 4-4). Thus, our result showing a 3.4-fold increase in mutation frequency over control in ENU-treated liver is consistent with previous reports, verifying the applicability of the stop-EGFP mouse to identifying potential mutagens in a mammalian system.

Table 4-4. Fold increase in mutation frequency over control in ENU-treated liver in various mutation detection systems

Target gene	Dose (mg/kg)	Fold increase	References
lacI	100	4.4	Zimmer et al. (1999)
lacZ	150	2.0	Collaborative study group (1996)
lacZ	150	6.0	Mientjes et al. (1998)
cII/cI	100	2.7	Zimmer et al. (1999)
gpt	250	3.0	Yamada et al. (1999)
stop-EGFP	**150**	**3.4**	**Ro et al. (2005) unpublished data**

6. Mutation assay in the kidney

In six kidneys from ENU-treated stop-EGFP mice, five green fluorescent mutant cells were detected at five months post-ENU administration (Figure 4-3). All mutant cells detected were single cells and this may reflect the slow rate of cell turnover in this tissue. The kidney can be anatomically divided into two parts, the cortex and medulla. All mutations found in the kidney were located in the cortex area (data not shown). The

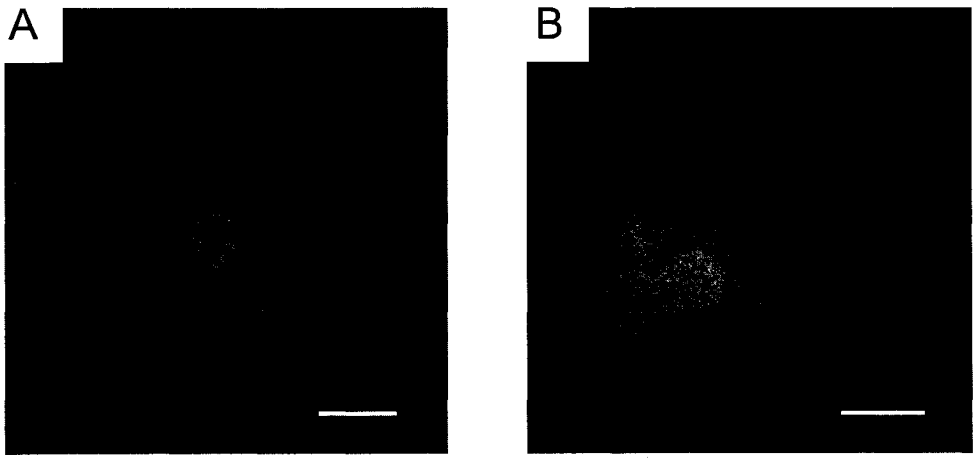


Figure 4-3. Green fluorescent mutant cells in the kidney.

Merged images of EGFP signals and DAPI signals of mutant cells detected in the kidney are shown (A and B). The bright green fluorescent signals within these cells are co-localized with the DAPI-stained nuclei. Scale bars, 10 μm.

mutation frequency was not estimated in the kidney due to the small number of mutations observed in the organ.

IV. Discussion

The stop-EGFP mouse system can be utilized as an in vivo mutation detection system by using the stop-EGFP gene as a reporter for mutation, which allows independent mutations to be counted by imaging of green fluorescent mutant colonies. Identifying independent mutations facilitates estimating mutation rate which better addresses the mutagenic effect of a potential mutagen than the mutant frequency. We successfully applied the stop-EGFP mouse to counting independent mutations having arisen in the entire liver and also to estimating the rate of mutation induced by ENU in the organ. The overall mutation rate induced by ENU-exposure for the entire liver was estimated to be about $(1.87 \pm 0.27) \times 10^{-7}$ per cell. No differences in underlying mutation rates were found among the different liver lobes. Thus the data suggests that treatment with ENU likely has the same effect on all liver lobes.

A test for a mutagenic effect is most efficient at the time (following exposure) when the highest mutant frequency is achieved and thus when a mutation assay is carried out using the *lacI*, or *lacZ*, transgenic mouse system it is recommended that samples are taken at the time of the highest mutant frequency for a given tissue. However, the time of maximum mutant frequency may be hard to predict (Thybaud et al., 2003). Sampling at an earlier time point likely underestimates mutation frequency because some mutant cells might not express the mutant phenotype yet. Sampling at a late time point after mutant frequency has reached a plateau also might be problematic. It has been suggested that the

plateau frequency can be lower than the maximum mutant frequency (Heddle et al., 2003; Nakajima et al., 1999; Wolff et al., 2001). Mutant cells might be selected against in some cases (e.g., in the *hprt* mutation detection system), which could lead to decrease in mutant frequency at a later time point (Heddle, 1998). Because mutant frequency can vary significantly depending on sampling time, it is critical to determine an estimate of mutation rate which can accurately reflect the actual mutagenic effect on cells regardless of the effect of sampling times.

Based on an explicit model developed for this purpose (see section III-4) we could accurately estimate the rate of mutation induced by ENU using the data from the time-course study with twelve ENU-treated stop-EGFP mice. The estimated mutation rate induced by treatment with ENU in the liver was 7.7×10^{-8} per cell. The average waiting time until the mutant phenotype is expressed in the liver is estimated to be 4.1 months after ENU treatment. The estimated average waiting time is slightly longer than the time of the maximum mutant frequency achieved in the liver of the transgenic mouse carrying a bacterial gene (about 2 – 3 months after ENU treatment) (Douglas et al., 1996; Wang et al., 2004). In the stop-EGFP system, to exhibit the mutant phenotype, mutant cells must express revertant EGFP at a level sufficient for detection under a fluorescence microscope. However, in the bacterial transgene-based systems such as the *lacZ*, or *lacI*, transgenic mouse, mutant phenotypes are observed in an in vitro bacterial assay system thus gene expression from the mutant gene in vivo is not required for the assay. Thus, it is reasonable to speculate that more waiting time would be required to detect a mutant phenotype from mutant cells in the stop-EGFP system than in the *lacZ*, or *lacI*, transgenic mouse.

Comparison of ENU-induced mutation frequency with the background mutation frequency in the liver was carried out using stop-EGFP mice. From the observation of 12 mutations in the left caudal liver lobe from 10 vehicle-treated stop-EGFP mice, the background mutation frequency was estimated to be 1.2 per lobe on average (see section III-5). In ENU-treated liver of stop-EGFP mice, the mean mutation frequency was estimated to be 4.1 per lobe. Our data showing a 3.4-fold increase in mutation frequency over control in ENU-treated liver are consistent with other previous reports using the *lacI*, or *lacZ*, transgenic mouse system. Thus, our study supports the utility of the stop-EGFP system as a novel in vivo mutation detection system which can identify potential mutagens. Comparison of mutation frequency per cell between ENU-treated and control liver will provide a better evaluation of the mutagenic effect of ENU in the liver than comparison of mutation frequency per lobe. After estimating the total number of cells in each lobe (see section II-3), we plan to estimate mutation frequency per cell in the liver in the near future.

V. References

Alison MR, Poulsom R, Otto WR, Vig P, Brittan M, Direkze NC, Lovell M, Fang TC, Preston SL, Wright NA. (2004) Recipes for adult stem cell plasticity: fusion cuisine or readymade? *J Clin Pathol.* 57: 113-120.

Collaborative study group. (1996) Organ variation in the mutagenicity of ethylnitrosourea in Muta mouse: results of the collaborative study on the transgenic mutation assay by JEMS/MMS. *Environ Mol Mutagen.* 28: 363-375.

Douglas GR, Jiao J, Gingerich JD, Gossen JA, Soper LM. (1995) Temporal and molecular characteristics of mutations induced by ethylnitrosourea in germ cells isolated from seminiferous tubules and in spermatozoa of lacZ transgenic mice. *Proc Natl Acad Sci U S A.* 92: 7485-7489.

Douglas GR, Jiao J, Gingerich JD, Soper LM, Gossen JA. (1996) Temporal and molecular characteristics of lacZ mutations in somatic tissues of transgenic mice. *Environ Mol Mutagen.* 28: 317-324.

Gorbunova V, Seluanov A. (2005) Making ends meet in old age: DSB repair and aging. *Mech Ageing Dev.* 126: 621-628.

Heddle JA. (1998) The use of transgenic animals in research. *Environ Mol Mutagen.* 32: 110-114.

Heddle JA. (1999) Mutant manifestation: the time factor in somatic mutagenesis. *Mutagenesis.* 14: 1-3.

Heddle JA, Martus HJ, Douglas GR. (2003) Treatment and sampling protocols for transgenic mutation assays. *Environ Mol Mutagen.* 41: 1-6.

Lawson TA, Pound AW. (1974) The different susceptibility of rat liver lobes to carbon tetrachloride and dimethylnitrosamine. *Br J Exp Pathol.* 55: 583-588.

Mientjes EJ, Luiten-Schuite A, van der Wolf E, Borsboom Y, Bergmans A, Berends F, Lohman PH, Baan RA, van Delft JH. (1998) DNA adducts, mutant frequencies, and mutation spectra in various organs of lambda lacZ mice exposed to ethylating agents. *Environ Mol Mutagen.* 31: 18-31.

Mirsalis JC, Provost GS, Matthews CD, Hamner RT, Schindler JE, O'Loughlin KG, MacGregor JT, Short JM. (1993) Induction of hepatic mutations in lacI transgenic mice. *Mutagenesis*. 8: 265-271.

Nakajima M, Kikuchi M, Saeki K, Miyata Y, Terada M, Kishida F, Yamamoto R, Furihata C, Dean SW. (1999) Mutagenicity of 4-nitroquinoline 1-oxide in the MutaMouse. *Mutat Res*. 444: 321-336.

Richardson FC, Boucheron JA, Dyroff MC, Popp JA, Swenberg JA. (1986) Biochemical and morphologic studies of heterogeneous lobe responses in hepatocarcinogenesis. *Carcinogenesis*. 7: 247-251.

Ro S, Rannala B. (2004) A stop-EGFP transgenic mouse to detect clonal cell lineages generated by mutation. *EMBO Rep*. 5: 914-920.

Sedelnikova OA, Horikawa I, Zimonjic DB, Popescu NC, Bonner WM, Barrett JC. (2004) Senescing human cells and ageing mice accumulate DNA lesions with unrepairable double-strand breaks. *Nat Cell Biol*. 6: 168-170.

Sell S. (2001) Heterogeneity and plasticity of hepatocyte lineage cells. *Hepatology*. 33: 738-750.

Singh R, Leuratti C, Josyula S, Sipowicz MA, Diwan BA, Kasprzak KS, Schut HA, Marnett LJ, Anderson LM, Shuker DE. (2001) Lobe-specific increases in malondialdehyde DNA adduct formation in the livers of mice following infection with Helicobacter hepaticus. *Carcinogenesis*. 22: 1281-1287.

Sun B, Heddle JA. (1999) The relationship between mutant frequency and time in vivo: simple predictions for any tissue, cell type, or mutagen. *Mutat Res*. 425: 179-183.

Sun B, Shima N, Heddle JA. (1999) Somatic mutation in the mammary gland: influence of time and estrus. *Mutat Res*. 427: 11-19.

Taylor JR. (1997). *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*. Sausalito: University Science Books.

Thybaud V, Dean S, Nohmi T, de Boer J, Douglas GR, Glickman BW, Gorelick NJ, Heddle JA, Heflich RH, Lambert I, Martus HJ, Mirsalis JC, Suzuki T, Yajima N. (2003) In vivo transgenic mutation assays. *Mutat Res.* 540: 141-151.

Walker VE, Jones IM, Crippen TL, Meng Q, Walker DM, Bauer MJ, Reilly AA, Tates AD, Nakamura J, Upton PB, Skopek TR. (1999) Relationships between exposure, cell loss and proliferation, and manifestation of Hprt mutant T cells following treatment of preweanling, weanling, and adult male mice with N-ethyl-N-nitrosourea. *Mutat Res.* 431: 371-388.

Wang J, Liu X, Heflich RH, Chen T. (2004) Time course of cII gene mutant manifestation in the liver, spleen, and bone marrow of N-ethyl-N-nitrosourea-treated Big Blue transgenic mice. *Toxicol Sci.* 82: 124-128.

Wolff T, Topinka J, Deml E, Oesterle D, Schwarz LR. (2001) Dose dependent induction of DNA adducts, gene mutations, and cell proliferation by the antiandrogenic drug cyproterone acetate in rat liver. *Adv Exp Med Biol.* 500: 687-696.

Yamada T, Yamamoto R, Kaneko H, Yoshitake A. (1999) Ethylnitrosourea-induced mutation and molecular analysis of transgenic mice containing the gpt shuttle vector. *Mutat Res.* 441: 59-72.

Zimmer DM, Harbach PR, Mattes WB, Aaron CS. (1999) Comparison of mutant frequencies at the transgenic lambda LacI and cII/cI loci in control and ENU-treated Big Blue mice. *Environ Mol Mutagen.* 33: 249-256.

Chapter 5. General Discussion

I. The Stop-EGFP System

1. Research outlines

In this thesis, we have developed a novel transgenic mouse system carrying a premature stop codon-containing EGFP gene (referred to as the stop-EGFP gene) as a target for mutations (Ro and Rannala, 2004). Nonsynonymous mutation at the premature stop codon in the stop-EGFP gene allows revertant EGFP to be expressed in individual mutants and their clonal descendent cells, thus clonal cell lineages can be traced by following green fluorescent colonies using fluorescence imaging.

Investigation of clonal cell lineages of progenitor cells is of great significance in stem cell research and developmental biology, and could address controversial issues such as the potential of stem cells to give rise to cells of a variety of tissue types or to provide cells for other cellular compartments under various circumstances (see sections II-3 and II-4, Chapter 1). Cell fate mapping systems based on random mutagenesis are well suited for cell lineage studies because the clonality of cell lineages generated in these systems is virtually guaranteed. A serious problem of previously developed systems using spontaneous revertant mutation to label cells for fate mapping (Bonnerot and Nicolas, 1993; DePrimo et al., 1998) is that it is difficult, or impossible, to determine the time points during development when revertant mutational events occurred, which is of critical importance in identifying the developmental origin of the cell lineages that are generated (see section II-6, Chapter 1). This can be a serious problem in applying either system to study stem cell clonal lineages in an adult renewing tissue because it is not known whether cell lineages generated in these systems have originated from embryonic cells, adult stem cells, or proliferating non-stem cells (i.e., transit amplifying cells).

The loss of function mutation-based labeling systems such as the *Dlb-1* system, and the transgenic mouse carrying a single copy of the *lacZ* gene, allow individual cells to be labeled via chemically induced mutations (Bjerknes and Cheng, 2002; Winton et al., 1988). In those systems the frequencies of spontaneous mutations are relatively low and thus cells can be labeled at specific time points via treatment with a potent mutagen. A serious disadvantage of both systems is that the labeling method via mutation relies on inactivation of the reporter genes thus both systems require that unstained cell lineages can be identified on the background of stained cells, which is technically challenging in most cases. Transgenic mouse systems carrying a reporter gene that can be activated (rather than inactivated) via chemical mutagenesis, therefore, would be preferable. In general, reactivation of a knock-out gene by revertant mutation can only be achieved via specific mutations at a specific target nucleotide, thus the efficiency with which labeled cells are generated is expected to be extremely low. Introducing a premature stop codon at a neutral codon within a reporter gene can increase the efficiency with which labeled cells are generated because a nonsynonymous mutation at any one of the three nucleotides of the premature stop codon will restore the reporter gene function. The fact that multiple copies of a transgene are usually inserted in a linear array in a transgenic mouse increases the efficiency of the system for detecting mutations.

Based on this idea, we created the stop-EGFP gene harboring a premature stop codon at a neutral codon site and the stop-EGFP mouse carrying multiple copies of the stop-EGFP gene. Due to the versatility of EGFP as a cell lineage marker, the stop-EGFP mouse system allows in vivo detection of clonal cell lineages via fluorescence imaging as well as repeated analysis of the same cell lineages over time. Because the stop-EGFP

gene can function as a reporter for mutation, the stop-EGFP system can also be utilized as an in vivo mutation detection system. The system allows independent mutations to be identified and thus has the potential to provide better estimates of mutation rates than other in vivo mutation detection systems such as the *lacI*, or *lacZ*, transgenic mouse (Gossen et al., 1989; Kohler et al., 1991).

2. Future prospects

The stop-EGFP system could be applied to other organisms. For example, transgenic fish and birds that ubiquitously express EGFP have been developed (Chapman et al., 2005; Gong et al., 2001; Higashijima et al., 1997) and presumably stop-EGFP transgenic animals could be easily created as well. If transgenic animals carrying the stop-EGFP gene were developed using other species, one might be able to investigate species difference in mutagenic response to a mutagen. One potentially interesting application is to use the stop-EGFP transgenic fish for mutagenicity testing in aquatic environments. Detection of potential mutagenic substances in aquatic environments is relevant to public health, as well as to the conservation of fish populations. When aquatic environments are contaminated with mutagenic substances, organisms are chronically exposed at a low level. Since fish mutation detection systems are thought to be more sensitive to such an exposure than rodent mutation detection systems (Winn et al., 2000), a transgenic fish mutation detection system might be efficient for use in toxicology testing in aquatic environments.

II. Application of the Stop-EGFP Mouse to Cell Fate Mapping

1. Research outcomes

In this thesis, the efficiency of the stop-EGFP mouse system as a cell fate mapping technique was evaluated by applying the system to tracing stem cell clonal lineages in the mouse dorsal epidermis. Using the system, we were able to trace clonal lineages of epidermal stem cells and found that single stem cells could contribute to a larger proliferative compartment than previously thought, which contradicts the classical model for epidermal stem cell lineages (Potten, 1974). Expansion of epidermal cell lineages to multiple EPUs has been observed previously using other lineage-tracing methods such as the retroviral transduction method (Ghazizadeh and Taichman, 2001; Kameda et al., 2003). However, due to the uncertainty of the clonality of cell lineages generated using such methods it was difficult to conclude whether a single stem cell can contribute to multiple EPUs. Using the stop-EGFP system allowed us to label individual cells by means of infrequent and random mutations and thus the clonality of resulting cell lineages is certain. Our findings from the in vivo lineage tracing study using the stop-EGFP mouse strongly support the idea that a single epidermal stem cell can contribute to multiple adjacent EPUs.

Labeling of individual cells using the stop-EGFP mouse system can be achieved via induced (by means of treatment with a mutagen) or spontaneous mutations. Although treatment with a mutagen increases the average number of clonal cell lineages generated (and thus can be more efficient), its mutagenic and cytotoxic effects may also cause alterations in the normal cellular environment. Thus, data from such studies should be interpreted with caution because cell lineages found in mutagen-treated mice might not

reflect normal cellular behavior under the steady-state condition. Thus, it is important to investigate whether cell lineages found in mutagen-treated mice can also be detected in untreated mice in which clonal cell lineages arise via spontaneous mutations.

2. Future prospects

Stem cells are the only cells remaining following tissue renewal. It is relatively easy to trace stem cell lineages in constantly renewing tissues (such as the epidermis) by labeling cells and investigating cell lineages later after several tissue turnovers. It is not possible to utilize this approach for tracing stem cell lineages in slowly renewing tissues such as in the liver. Without further knowledge on the stem cell identities, tracing stem cell lineages will be hard to achieve in those tissues. Another characteristic of stem cells is the regenerative potential to repair damaged tissues in response to mechanical wounds or infection. Stem cells are actively proliferative under such conditions and cell lineages generated during the regeneration process would likely represent stem cell lineages. If tissue regeneration were induced in tissues such as the liver of the stop-EGFP mouse after treatment with a mutagen to randomly label cells, resulting green fluorescent cell lineages might represent clonal cell lineages that originated from liver stem cells.

Although we performed in vivo imaging of clonal cell lineages only in the epidermis in this study, a similar approach could potentially be applied to other organs. In vivo imaging of cells expressing GFP has been performed in various organs (including repeated in vivo imaging of the same GFP-expressing cells over time) such as the liver, muscle and brain (Chen et al., 2000; Feng et al., 2000; Hoffman, 2002; Naumov et al., 1999). If such repeated in vivo imaging of the same tissue sites were performed over time using the stop-EGFP mouse system, one might be able to investigate the dynamics of the

clonal development of a cell lineage (e.g., in the regenerating liver). Another advantage of using GFP as a cell lineage marker is that external imaging of even single cells in internal organs can be performed (Yang et al., 2002). Thus, one might be able to investigate the dynamics of the clonal development of a cell lineage in an internal organ by repeated external imaging of the stop-EGFP mouse over time. With the current revolutionary development of GFP imaging techniques, the potential applications of the stop-EGFP mouse can be expected to increase in the future.

In addition to clonal cell lineage studies in adult tissues, the stop-EGFP system also has the potential to be efficiently applied to tracing clonal cell lineages in developmental studies. Embryos can be exposed to ENU in utero at various time points during development by exposing pregnant mice to the mutagen. In our system, cells can be labeled at a specific time point during development by treatment with a pulse of the mutagen, thus a clonal cell lineage can be generated that originated from a precursor cell present in a specific stage during development.

III. Application of the Stop-EGFP Mouse to Mutation Studies

1. Research outcomes

In this thesis, the efficiency of the stop-EGFP mouse system as an in vivo mutation detection system was evaluated by applying the system to estimate the rate of mutation induced by ENU in the liver. The stop-EGFP mouse can detect independent mutations that have arisen in a tissue, thus it has the potential to provide better estimates of mutation rates. The stop-EGFP mouse system allowed mutation rates in different sectors of the organ (e.g., different liver lobes) to be estimated and compared.

From the time-course study of mutations induced by ENU in the liver of the stop-EGFP mouse (up to seven months after ENU treatment) it was found that the apparent mutation frequencies significantly change depending on the sampling times (section III-3, Chapter 4). Expression of the mutant phenotype in cells of the stop-EGFP mouse treated with ENU depends on several processes such as transcription from a mutant copy of the stop-EGFP gene, protein synthesis from the mutant mRNA, etc. Thus, mutation assays carried out at an early time point might fail to detect some mutant cells which have not yet achieved sufficient expression of revertant EGFP from a mutant copy of the stop-EGFP gene. This could lead to skewed estimates of mutation frequency. An explicit model was developed (see section III-4, Chapter 4) to allow the rate of somatic mutations induced by a mutagen in the liver to be accurately estimated taking account of the lag time until expression of the mutant phenotype.

Comparison of mutation frequency in the left caudal liver lobe between ENU-treated and control mice enabled us to evaluate the mutagenic effect of ENU in the liver. A 3.4-fold increase in mutation frequency over control in our study was consistent with other previous reports, supporting the utility of the stop-EGFP mouse as an in vivo mutation detection system.

2. Future prospects

Similar approaches can be applied to other potential mutagens to accurately estimate the mutation rates induced by these substances. As well, several strains of knock-out mice with a deficient DNA repair gene (see section III-4, Chapter 1) have already been generated (Wijnhoven and Steeg, 2003; Winter and Gearhart, 2001). Mutation rate /or and/ mutation frequency in these strains of knock-out mice could be estimated using the

stop-EGFP system after crossing the knock-out mice with stop-EGFP mice. It would also be interesting to investigate hypermutability in cancer using the stop-EGFP mouse (Loeb et al., 2003; Nowell, 1976). Various types of tumors can be induced in stop-EGFP mice by crossing them with tumor-prone mice such as mice with a viral oncogene (Jakubczak et al., 1996), or a deficient p53 (Buettner et al., 1996) and mutation rates in tumor samples could be estimated to examine hypermutation in cancer cells.

IV. Limitations of the Current Stop-EGFP Mouse System and Future Directions

One drawback of using the stop-EGFP system for tracing green fluorescent cell lineages is that background cells and surrounding tissue structures were not easily identified. Observing background cells might reveal more detailed histological information regarding the green fluorescent clonal cell lineages that are identified. Autofluorescent signals could be used for imaging background cells, but the signals are generally weak and unreliable. Initially, we expected that the co-injected EBFP gene could help with imaging of surrounding tissue structures. However, imaging of EBFP was unsuccessful due to rapid photobleaching and low signal intensity as previously reported (Patterson et al., 2001; Tsien, 1998). To perform better histological studies of clonal cell lineages in the stop-EGFP system, one could perform fluorescence imaging of EGFP with the help of immunohistochemical methods (e.g., using cell type-specific antibodies).

Generating a stop-EGFP mouse carrying other fluorescent protein genes instead of the EBFP gene might be another solution. One promising candidate is ECFP (enhanced cyan fluorescent protein) which has relatively good intensity of fluorescence and low photobleaching (see section I-1, Chapter 1). Although the fact that the emission

peak of ECFP is somewhat close to that of EGFP could be a problem, using a proper excitation light and emission filters could better distinguish EGFP signal from ECFP signal. Considering that EYFP (enhanced yellow fluorescent protein) and ECFP have more distinct excitation and emission spectra (see Table 1-1), a combination of EYFP and ECFP might also work. For example, a transgenic mouse carrying a premature stop codon-containing EYFP gene in addition to a functional ECFP gene could be generated. This could be an ideal choice for cell fate mapping studies; clonal cell lineages expressing EYFP could then be imaged as well as surrounding tissue expressing ECFP.

Another potential weakness of the stop-EGFP mouse is that the mutation rate at the premature stop codon within the stop-EGFP gene appears to be low even after treatment with a potent mutagen, ENU. This reduces the efficiency with which labeled cells are generated via mutation in the stop-EGFP system. Thus, imaging of many mice might be required to detect clonal cell lineages in a tissue of interest. Repeated treatments with ENU or another mutagen could help to increase the expected frequency of mutations at the premature stop codon. Another strategy is to create a transgenic mouse carrying a very high copy number of the stop-EGFP gene or an inbred line homozygous for the transgene (i.e., carrying the stop-EGFP gene on both homologous chromosomes).

Although the low mutation rate in the stop-EGFP system may appear disadvantageous, the extremely low rate of spontaneous mutation in the stop-EGFP mouse can also be perceived as an advantage because it enables cells to be labeled at a specific time point by treatment with a pulse of mutagen. Therefore, imaging experiments carried out after several tissue turnovers in an adult renewing tissue of mice treated with a mutagen could allow one to specifically trace cell lineages that descended

from stem cells of the tissue. As well, treatment of an embryo with a pulse of mutagen could allow clonal cell lineages to be generated that originated specifically at a time point of interest without contamination by cell lineages generated by spontaneous mutations occurring at other time points during development.

V. References

Bjerknes M, Cheng H. (2002) Multipotential stem cells in adult mouse gastric epithelium. *Am J Physiol Gastrointest Liver Physiol.* 283: G767-G777.

Bonnerot C, Nicolas JF. (1993) Clonal analysis in the intact mouse embryo by intragenic homologous recombination. *C R Acad Sci III.* 316: 1207-1217.

Buettner VL, Hill KA, Nishino H, Schaid DJ, Frisk CS, Sommer SS. (1996) Increased mutation frequency and altered spectrum in one of four thymic lymphomas derived from tumor prone p53/Big Blue double transgenic mice. *Oncogene.* 13: 2407-2413.

Chapman SC, Lawson A, Macarthur WC, Wiese RJ, Loechel RH, Burgos-Trinidad M, Wakefield JK, Ramabhadran R, Mauch TJ, Schoenwolf GC. (2005) Ubiquitous GFP expression in transgenic chickens using a lentiviral vector. *Development.* 132: 935-940.

Chen BE, Lendvai B, Nimchinsky EA, Burbach B, Fox K, Svoboda K. (2000) Imaging high-resolution structure of GFP-expressing neurons in neocortex in vivo. *Learn Mem.* 7: 433-441.

DePrimo SE, Cao J, Hersh MN, Stringer JR. (1998) Use of human placental alkaline phosphatase transgenes to detect somatic mutation in mice in situ. *Methods.* 16: 49-61.

Feng G, Mellor RH, Bernstein M, Keller-Peck C, Nguyen QT, Wallace M, Nerbonne JM, Lichtman JW, Sanes JR. (2000) Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *Neuron.* 28: 41-51.

Ghazizadeh S, Taichman LB. (2001) Multiple classes of stem cells in cutaneous epithelium: a lineage analysis of adult mouse skin. *EMBO J.* 20: 1215-1222.

Gong Z, Ju B, Wan H. (2001) Green fluorescent protein (GFP) transgenic fish and their applications. *Genetica.* 111: 213-225.

Gossen JA, de Leeuw WJ, Tan CH, Zwarthoff EC, Berends F, Lohman PH, Knook DL, Vijg J. (1989) Efficient rescue of integrated shuttle vectors from transgenic mice: a model for studying mutations in vivo. *Proc Natl Acad Sci U S A.* 86: 7971-7975.

Higashijima S, Okamoto H, Ueno N, Hotta Y, Eguchi G. (1997) High-frequency generation of transgenic zebrafish which reliably express GFP in whole muscles or the whole body by using promoters of zebrafish origin. *Dev Biol.* 192: 289-299.

Hoffman R. (2002) Green fluorescent protein imaging of tumour growth, metastasis, and angiogenesis in mouse models. *Lancet Oncol.* 3: 546-556.

Jakubczak JL, Merlino G, French JE, Muller WJ, Paul B, Adhya S, Garges S. (1996) Analysis of genetic instability during mammary tumor progression using a novel selection-based assay for in vivo mutations in a bacteriophage lambda transgene target. *Proc Natl Acad Sci U S A.* 93: 9073-9078.

Kameda T, Nakata A, Mizutani T, Terada K, Iba H, Sugiyama T. (2003) Analysis of the cellular heterogeneity in the basal layer of mouse ear epidermis: an approach from partial decomposition in vitro and retroviral cell marking in vivo. *Exp Cell Res.* 283: 167-183.

Kohler SW, Provost GS, Fieck A, Kretz PL, Bullock WO, Sorge JA, Putman DL, Short JM. (1991) Spectra of spontaneous and mutagen-induced mutations in the lacI gene in transgenic mice. *Proc Natl Acad Sci U S A.* 88: 7958-7962.

Loeb LA, Loeb KR, Anderson JP. (2003) Multiple mutations and cancer. *Proc Natl Acad Sci U S A.* 100: 776-781.

Naumov GN, Wilson SM, MacDonald IC, Schmidt EE, Morris VL, Groom AC, Hoffman RM, Chambers AF. (1999) Cellular expression of green fluorescent protein, coupled with high-resolution in vivo videomicroscopy, to monitor steps in tumor metastasis. *J Cell Sci.* 112: 1835-1842.

Nowell PC. (1976) The clonal evolution of tumor cell populations. *Science.* 194: 23-28.

Patterson G, Day RN, Piston D. (2001) Fluorescent protein spectra. *J Cell Sci.* 114: 837-838.

Potten CS. (1974) The epidermal proliferative unit: the possible role of the central basal cell. *Cell Tissue Kinet.* 7: 77-88.

Ro S, Rannala B. (2004) A stop-EGFP transgenic mouse to detect clonal cell lineages generated by mutation. *EMBO Rep.* 5: 914-920.

Tsien RY. (1998) The green fluorescent protein. *Annu Rev Biochem.* 67: 509-544.

Wijnhoven SW, van Steeg H. (2003) Transgenic and knockout mice for DNA repair functions in carcinogenesis and mutagenesis. *Toxicology.* 193: 171-187.

Winn RN, Norris MB, Brayer KJ, Torres C, Muller SL. (2000) Detection of mutations in transgenic fish carrying a bacteriophage lambda cII transgene target. *Proc Natl Acad Sci U S A.* 97: 12655-12660.

Winter DB, Gearhart PJ. (2001) Altered spectra of hypermutation in DNA repair-deficient mice. *Philos Trans R Soc Lond B Biol Sci.* 356: 5-11.

Winton DJ, Blount MA, Ponder BAA. (1988) Clonal marker induced by mutation in mouse intestinal epithelium. *Nature.* 333: 463-466.

Yang M, Baranov E, Wang JW, Jiang P, Wang X, Sun FX, Bouvet M, Moossa AR, Penman S, Hoffman RM. (2002) Direct external imaging of nascent cancer, tumor progression, angiogenesis, and metastasis on internal organs in the fluorescent orthotopic model. Proc *Natl Acad Sci U S A.* 99: 3824-3829.